A
Summer Internship Report
On
"Cloud Computing"

(IT446 – Summer Internship - II)

Prepared by

Devanshu Parekh (19IT075)

Under the Supervision of

Prof. Hemant Yadav

Submitted to

Charotar University of Science & Technology (CHARUSAT)
for the Partial Fulfillment of the Requirements for the
Degree of Bachelor of Technology (B.Tech.)
for Semester 5

Submitted at

Accredited with Grade A by NAAC
Accredited with Grade A by KCG

SMT. KUNDANBEN DINSHA PATEL DEPARTMENT OF INFORMATION TECHNOLOGY
(NBA Accredited)
Chandubhai S. Patel Institute of Technology (CSPIT)
Faculty of Technology & Engineering (FTE), CHARUSAT
At: Changa, Dist: Anand, Pin: 388421.
July, 2022

CHARUSAT
CHAROTAR UNIVERSITY OF SCIENCE AND TECHNOLOGY

Accredited with Grade A by NAAC
Accredited with Grade A by KCG

CERTIFICATE

This is to certify that the report entitled “Cloud Computing” is a bonafied work carried out by Devanshu Parekh(19IT075) under the guidance and supervision of **Prof. Hemant Yadav** & **Mr. Rishi Raj** for the subject **Summer Internship – II (IT446)** of 6th Semester of Bachelor of Technology in **Department of Information** at Chandubhai S. Patel Institute of Technology (CSPIT), Faculty of Technology & Engineering (FTE) – CHARUSAT, Gujarat.

To the best of my knowledge and belief, this work embodies the work of candidate himself, has duly been completed, and fulfills the requirement of the ordinance relating to the B.Tech. Degree of the University and is up to the standard in respect of content, presentation and language for being referred by the examiner(s).

Under the supervision of,

Prof. Hemant Yadav
Internal Guide
Smt. Kundanben Dinsha Patel Department of Information Technology
CSPIT, FTE, CHARUSAT, Changa, Gujarat

Mr.Rishi Raj
Manager
Gugalia Fintech pvt Ltd

Dr. (Prof.) Parth Shah
Head of Department (IT)
CHARUSAT, Changa, Gujarat.

Chandubhai S. Patel Institute of Technology (CSPIT)

Faculty of Technology & Engineering (FTE), CHARUSAT

At: Changa, Ta. Petlad, Dist. Anand, Pin: 388421. Gujarat.

Table of Contents

ACKNOWLEDGEMENT

For most modern professionals, the idea of working from home used to be a luxury. As we are all familiar, the COVID-19 outbreak of 2020 dramatically increased the need to work from our homes.

I was lucky to get chance to work in such a great company under excellent guidance of Rishi sir at Gugalia Fintech. It was great to work with the seniors. I would like to thank Assistant Professor Hemant Yadav to support me in my journey as our internal guide. I gained valuable skills and knowledge during this period. I would also like to thank our HOD Mr. Parth Shah and our principal Dr. Y.P Koshta. We heartily thank you for all your support. Thank you for motivating us and guiding us through the obstacles. This time was full of excitement, curiosity and also the sincerity which doesn’t even made it look like heavy work. Thank you for all the support and guidance and getting us prepared for this valuable experience.

ABSTRACT

Cloud computing is possible because of a technology called virtualization. Virtualization allows for the creation of a simulated, digital-only "virtual" computer that behaves as if it were a physical computer with its own hardware. The technical term for such a computer is virtual machine. When properly implemented, virtual machines on the same host machine are sandboxed from one another, so they do not interact with each other at all, and the files and applications from one virtual machine are not visible to the other virtual machines even though they are on the same physical machine.

For businesses, switching to cloud computing removes some IT costs and overhead: for instance, they no longer need to update and maintain their own servers, as the cloud vendor they are using will do that. This especially makes an impact for small businesses that may not have been able to afford their own internal infrastructure but can outsource their infrastructure needs affordably via the cloud. The cloud can also make it easier for companies to operate internationally, because employees and customers can access the same files and applications from any location.

As I was fresher, they guided me through their tools and technologies that they use for virtualization. My work was to write the scripts and make the pipelines for automation.

CHAPTER 1

UNDERSTAND THE CLOUD

1.1 WHAT IS CLOUD?

"The cloud" refers to servers that are accessed over the Internet, and the software and databases that run on those servers. Cloud servers are located in data centers all over the world. By using cloud computing, users and companies do not have to manage physical servers themselves or run software applications on their own machines.

You may have heard people using terms like the cloud, cloud computing, or cloud storage. But what exactly is the cloud?

Simply put, the cloud is the Internet—more specifically, it's all of the things you can access remotely over the Internet. When something is in the cloud, it means it's stored on Internet servers instead of your computer's hard drive.

The cloud enables users to access the same files and applications from almost any device, because the computing and storage takes place on servers in a data center, instead of locally on the user device. This is why a user can log in to their Instagram account on a new phone after their old phone breaks and still find their old account in place, with all their photos, videos, and conversation history. It works the same way with cloud email providers like Gmail or Microsoft Office 365, and with cloud storage providers like Dropbox or Google Drive.

1.2 WHAT IS CLOUD COMPUTING?

Cloud Computing is a network of remote servers hosted on the internet for storing and retrieving data. The cloud provides a number of IT services such as servers, databases, software, virtual storage, and networking, among others. In layman’s terms, Cloud Computing is defined as a virtual platform that allows you to store and access your data over the internet without any limitations.

Companies that offer all the services mentioned above are called cloud providers. They provide you with the ability to store and retrieve data and run applications, managing them through configuration portals. Two of the best cloud providers available today are Amazon Web Services and Microsoft Azure.

1.3 BENEFITS OF CLOUD COMPUTING:

I. Speed
II. Cost
III. Scalability
IV. Performance
V. Better Security
VI. Flexibility

1.4 CLOUD COMPUTING WITH AWS:

Amazon Web Services (AWS) is the world’s most comprehensive and broadly adopted cloud platform, offering over 200 fully featured services from data centers globally. Millions of customers—including the fastest-growing startups, largest enterprises, and leading government agencies—are using AWS to lower costs, become more agile, and innovate faster.

AWS is an online platform that provides scalable and cost-effective cloud computing solutions.

AWS is a broadly adopted cloud platform that offers several on-demand operations like compute power, database storage, content delivery, etc., to help corporates scale and grow.

For example, Adobe creates and updates software without depending upon the IT teams. It uses its services by offering multi-terabyte operating environments for its clients. By deploying its services with Amazon services, Adobe integrated and operated its software in a simple manner.

1.5 ADVANTAGES OF USING AWS:

I. Pay as you go
II. Stop guessing capacity
III. Increase speed and agility
IV. Stop spending money running and maintaining data centers
V. Go global in minutes
VI. Data security

CHAPTER 2

INTRODUCTON TO DEVOPS

2.1 WHAT IS DEVOPS?

DevOps combines development and operations to increase the efficiency, speed, and security of software development and delivery compared to traditional processes. A more nimble software development lifecycle results in a competitive advantage for businesses and their customers.

DevOps promotes collaboration between Development and Operations team to deploy code to production faster in an automated & repeatable way. DevOps helps to increase organization speed to deliver applications and services. It also allows organizations to serve their customers better and compete more strongly in the market.

DevOps can also be defined as a sequence of development and IT operations with better communication and collaboration. DevOps has become one of the most valuable business disciplines for enterprises or organizations. With the help of DevOps, quality, and speed of the application delivery has improved to a great extent.

DevOps is nothing but a practice or methodology of making "Developers" and "Operations" folks work together. DevOps represents a change in the IT culture with a complete focus on rapid IT service delivery through the adoption of agile practices in the context of a system-oriented approach.

DevOps is all about the integration of the operations and development process. Organizations that have adopted DevOps noticed a 22% improvement in software quality and a 17% improvement in application deployment frequency and achieve a 22% hike in customer satisfaction. 19% of revenue hikes as a result of the successful DevOps implementation.

2.2 WHY DEVOPS?

I. The operation and development team worked in complete isolation.

II. After the design-build, the testing and deployment are performed respectively. That's why they consumed more time than actual build cycles.

III. Without the use of DevOps, the team members are spending a large amount of time on designing, testing, and deploying instead of building the project.

IV. Manual code deployment leads to human errors in production.

V. Coding and operation teams have their separate timelines and are not in synch, causing further delays.

2.3 ADVANTAGES OF DEVOPS:

I. DevOps is an excellent approach for the quick development and deployment of applications.

II. It responds faster to the market changes to improve business growth.

III. DevOps escalate business profit by decreasing software delivery time and transportation costs.

IV. DevOps clears the descriptive process, which gives clarity on product development and delivery.

V. It improves customer experience and satisfaction.

VI. DevOps simplifies collaboration and places all tools in the cloud for customers to access.

VII. DevOps means collective responsibility, which leads to better team engagement and productivity.

2.4 DEVOPS BEST PRACTICES:

I. Continuous Integration

II. Continuous Delivery

III. Microservices

IV. Infrastructure as Code

V. Monitoring and Logging

VI. Communication and Collaboration

CHAPTER 3

WORKING WITH AWS

3.1 STORAGE LAYER

3.1.1 INTRODUCTION TO S3:

Amazon S3 is an object storage service. It enables you to store virtually unlimited amounts of data. Data files are stored as objects. You place objects in a bucket, which you define. Every bucket must have name that is globally unique across Regions. This means that the bucket name must be unique across all AWS customer accounts. The objects you store can vary in size from 0 bytes to 5 TB. Though individual objects cannot be larger than 5 TB, you can store as much total data as you need.

Each object has five consistent characteristics. First, it has a key, which is the name that you assign to an object. You use the object key to retrieve the object. In the AWS Management Console, you can create a directory inside a bucket, and upload an object to that directory. However, in reality, Amazon S3 does not know about directories, so the key value includes the full path relative to the bucket root.

Objects also include a version ID. In a bucket, a key and version ID uniquely identify an object. You will learn more about versioning later in this module. The value of the object is the actual content that you store. It can be any sequence of bytes. Object values are immutable, which means that after you upload an object, you cannot modify the value. If you want to modify the object, you must make a change outside of Amazon S3 and then reupload the object.

Objects also include metadata, which is a set of name-value pairs you can use to store information about the object. You can assign metadata, which is referred to as user-defined metadata, to your objects in Amazon S3. Amazon S3 also assigns system-metadata to these objects, which it uses for managing objects.

3.1.2 BENEFITS OF S3:

I. Durability
II. Scalability
III. Availability
IV. Security
V. Performance

First, it provides durability, which describes the average annual expected loss of objects. 11.9s of durability means that every year, there is a **0.000000001** percent chance of losing an object. For example, if you store 10,000 objects on Amazon S3, you can expect to incur a loss of a single object once every 10,000,000 years on average. Amazon S3 redundantly stores your objects on multiple devices across multiple facilities in the Amazon S3 Region you designate. Amazon S3 is designed to sustain concurrent device failures by quickly detecting and repairing any lost redundancy. Amazon S3 also regularly verifies the integrity of your data by using checksums. Amazon S3 also provide four 9s (or 99.99 percent) of availability. Availability refers to your ability to access your data quickly, when you want it. It also provides a virtually unlimited capacity to store your data, so it is scalable. Amazon S3 has robust security settings. It provides many ways to control access to the data that you store, and also enables you to encrypt your data. Finally, Amazon S3 is highly performant, with a first-byte latency that is measured in milliseconds for most storage classes. For more information about S3 performance design patterns, see the Amazon S3 Documentation. Common approaches include using caching for frequently accessed content; configurable retry and timeout logic for objects that receive significant request traffic in a short period of time; and horizontal scaling and request parallelization for high throughput across the network.

3.1.3 AMAZON S3 USE CASES:

I. Store and distribute web content and media
II. Host static websites
III. Data store for computation and analytics
IV. Back-up and archive critical data

3.1.4 Amazon S3 costs:

With Amazon S3, you pay only for what you use. There is no minimum fee.

There are four cost components to consider when you decide which Amazon S3 storage class best fits your data profile

I. Storage pricing
II. Request and data retrieval pricing
III. Data transfer and transfer acceleration pricing
IV. Data management features pricing.

No charges for

I. Data transfer IN from the internet to Amazon S3
II. Transfer between S3 buckets or from S3 to any services in the same region
III. Tranfer OUT to Amazon CloudFront
IV. DELETE and CANCEL requests.

3.2 COMPUTE LAYER:

3.2.1: AWS COMPUTE SERVICES:

AWS offers several compute options to meet different needs. As you design the architecture to support a given type of workload, it is important that you understand the available compute options.

I. Amazon EC2
II. Amazon EBS
III. Amazon Lambda
IV. Amazon Fargate
V. Amazon Lightsail

3.2.2 AWS EC2:

Amazon Elastic Compute Cloud (Amazon EC2) enables computing in the cloud. You can use Amazon EC2 to provision virtual servers, and you can completely control the computing resources of those servers. You can obtain and start new server instances in minutes. You can quickly scale capacity both up and down as your computing requirements change. From a cost perspective, you pay only for the capacity that you use.

It is called Elastic because you can easily increase or decrease the number of servers you run to support an application automatically. You can also increase or decrease the size of existing servers.

It is called Compute because most users run servers to host running applications or process data, which require compute resources. These resources include processing power (CPU) and memory (RAM). And it is called Cloud because the EC2 instances that you run are hosted in the cloud.

3.2.3 AMAZON EC2 USE CASES:

Amazon EC2 provides virtual machines where you can host the same kinds of applications that you might run on a traditional on-premises server. Common uses for EC2 instances include web servers, application servers, database servers, and media servers. In particular, consider Amazon EC2 as a compute choice in situations where you need:

I. Complete control of your computing recourses, including operating system and processor type
II. Ability to run any type of workload
III. Options for optimizing your compute costs i.e different types of instances.

3.2.4 AMAZON EC2 PRICING:

I. On Demand Instances:

Pay for compute capacity by the second or by the hour with no long term commitments.

II. Reserved Instances:

Make a 1-year or 3-year commitment and receive a significant discount off on-demand prices.

III. Savings Plans:

Samae discounts as Reserved Instances with more flexibility in exchange for$/hour commitment.

IV. Spot Instances:

Spare Amazon Ec2 capacity at substantial savings off on-demand instance prices.

V. Dedicated Hosts:

Physical server with Amazon EC2 instance capacity fully dedicated for your use.

3.3 DATABASE LAYER:

3.3.1: DATABASE PARAMETERS:

We will often need to choose between different database types when you consider which type will best handle a particular workload. Before you choose a database, there are some key considerations that should inform your decision-making process.

The important parameters taken into consideration are

I. Scalability

II. Total storage requirements

III. Object size and type

IV. Durability

First, consider the importance of scalability. With traditional on-premises databases, scaling up capacity can be a difficult task even for experienced database administrators. It can take hours, days, or weeks. The impact to database performance while it scales up can be unpredictable, and might require downtime. However, the importance of a properly scaled database cannot be overstated. If you under provision your database, your applications might stop working. However, if you overprovision your database, you increase your upfront costs by procuring resources that you do not need, which violates the cost-optimization principle of the AWS Well-Architected Framework.

Secondly, understanding your total storage requirements is essential for choosing a database.

Third, when you choose a database for your specific workloads, consider the size and the type of the objects you must store.

Finally, fourth consideration will be durability. Data durability refers to the assurance that your data will not be lost, and data availability refers to your ability to access your data when you want to. What level of data durability and data availability do you need? If the data you will store is absolutely critical to your business, you should choose a database solution that stores multiple redundant copies of your data across multiple geographically separated physical locations. This solution will usually result in an increased cost, so it is important to balance your business needs with cost considerations.

3.3.2 RELATIONAL DATABASE:

Relational database systems are the most familiar type of databases to most people. Traditional examples include Microsoft SQL Server, Oracle Database, and MySQL.

Benefits of Relational Database are

I. Easy to use
II. Data integrity
III. Reduced data storage
IV. Common Language(i.e SQL)

Relational Database is ideal when we

I. Need to strict schema rules, ACIS compliance and data quality enforcement
II. Do not need extreme read/write capacity
III. Do not need extreme performance

3.3.3 NON-RELATIONAL DATABASE:

Non-relational databases were developed more recently, but have been around for a few decades. They play an essential role in the modern computing landscape. Examples include MongoDB, Cassandra, and Redis.

Benefits of Non-Relational Database:

I. Flexibility
II. Scalability
III. High Performance
IV. Highly Functional APIs

Non-Relational Database is ideal when

I. Database must scale horizontally to handle massive data volume
II. Data doesn’t lend itself well to traditional schemas
III. Read/write rate exceeds what can be economically supported through traditional RDBMS

CHAPTER 4
LEARNING DEVOPS

4.1 GIT & GITHUB:

Version control, also known as source control, is the practice of tracking and managing changes to software code. Version control systems are software tools that help software teams manage changes to source code over time. Git is a free and open-source distributed version control system designed to handle everything from small to very large projects with speed and efficiency. The Git feature that really makes it stand apart from nearly every other SCM out there is its branching model. Git is fast. With Git, nearly all operations are performed locally, giving it a huge speed advantage on centralized systems that constantly have to communicate with a server somewhere. One of the nicest features of any Distributed SCM, Git included, is that it's distributed. This means that instead of doing a "checkout" of the current tip of the source code, you do a "clone" of the entire repository.

4.1.1 WHY GIT:

Git is a popular version control system. It is used for tracking code changes, tracking who made changes and coding collaboration. GIT manages the projects with repositories and clone a project to work on a local copy. Also, it controls and track the changes with staging and commiting.

I. Over 70% of developers use GIT

II. Developers can work together from anywhere in the world.

III. Developers can see the full history of the project.

IV. Developers can revert to earlier versions of a project.

4.1.2 WHAT IS GITHUB:

Github makes tools that use Git and Github is the largest host of source code in the world. GitHub hosts Git repositories and provides developers with tools to ship better code through command line features, issues (threaded discussions), pull requests, code review, or the use of a collection of free and for-purchase apps in the GitHub Marketplace. With collaboration layers like the GitHub flow, a community of 15 million developers, and an ecosystem with hundreds of integrations, GitHub changes the way software is built.

GitHub builds collaboration directly into the development process. Work is organized into repositories where developers can outline requirements or direction and set expectations for team members. Then, using the GitHub flow, developers simply create a branch to work on updates, commit changes to save them, open a pull request to propose and discuss changes, and merge pull requests once everyone is on the same page.

4.2 JENKINS- THE AUTOMATION TOOL :

Jenkins is an open-source automation tool with plugins built for Continuous Integration purposes. Jenkins is used to build and test your software projects continuously making it easier for developers to integrate changes to the project, and making it easier for users to obtain a fresh build. It also allows you to continuously deliver your software by integrating with a large number of testing and deployment technologies.

With Jenkins, organizations can accelerate the software development process through automation. Jenkins integrates development life-cycle processes of all kinds, including build, document, test, package, stage, deploy, static analysis, and much more. Jenkins achieves Continuous Integration with the help of plugins. Plugins allow the integration of Various DevOps stages. If you want to integrate a particular tool, you need to install the plugins for that tools like Git, Maven 2 project, Amazon EC2, HTML publisher etc.

Advantages of Jenkins include:

I. It is an open-source tool with great community support.

II. It is easy to install.

III. It has 1000+ plugins to ease your work. If a plugin does not exist, you can code it and share it with the community.

IV. It is free of cost.

V. It is built with Java and hence, it is portable to all the major platforms.

Continuous Integration: A software development process where the changes made to software are integrated into the main code as and when a patch is ready so that the software will be always ready to be - built, tested, deployed, monitored - continuously.

Continuous Delivery: This is a Software Development Process where the continuously integrated (CI) changes will be tested & deployed continuously into a specific environment, generally through a manual release process, after all the quality checks are successful.

Continuous Deployment: A Software Development practice where the continuously integrated (CI) changes are deployed automatically into the target environment after all the quality checks are successful.

CHAPTER 5

AWS PIPELINE

5.1 AWS CODEBUILD:

AWS CodeBuild is a fully managed build service in the cloud. CodeBuild compiles your source code, runs unit tests, and produces artifacts that are ready to deploy. CodeBuild eliminates the need to provision, manage, and scale your own build servers. It provides prepackaged build environments for popular programming languages and build tools such as Apache Maven, Gradle, and more. You can also customize build environments in CodeBuild to use your own build tools. CodeBuild scales automatically to meet peak build requests.

5.2 AWS CODEPIPELINE:

AWS CodePipeline is a continuous delivery service you can use to model, visualize, and automate the steps required to release your software. You can quickly model and configure the different stages of a software release process. CodePipeline automates the steps required to release your software changes continuously.

5.3 AWS CODEDEPLOY:

AWS CodeDeploy is a fully managed deployment service that automates software deployments to a variety of compute services such as Amazon EC2, AWS Fargate, AWS Lambda, and your on-premises servers. AWS CodeDeploy makes it easier for you to rapidly release new features, helps you avoid downtime during application deployment, and handles the complexity of updating your applications. You can use AWS CodeDeploy to automate software deployments, eliminating the need for error-prone manual operations. The service scales to match your deployment needs.

5.4 BUILDING AWS PIPELINE:

Amazon services such as EC2, CodePipeline, CodeBuild, CodeDeploy are used to create a AWS Pipeline to make the deployment work easy. Deployment is the final stage of the project and with AWS, it becomes a sort of easy to deal with the pipeline.

Firstly, we had to make 2 IAM roles, one for EC2 and the other for CodeDeploy. Then, we had to launch an EC2 servers and then go to AWS CodeDeploy where we have to give the name and select the compute platform (EC2 in our case). Moving forward, select the IAM role which was earlier made and then select the Instance which is to be used and hence the application is deployed. Now, we had to make the pipeline. So move to CodePipeline and then click create Pipeline. In that, Select the Source stage (Github in our case) and then connect to the corresponding repository in which the project is updated by the developer team. Then in build stage, we can integrate Jenkins for continuous integration and automation. Next we have to select the deploy provider (AWS CodeDeploy in our case) and finally the pipeline is created and the application is deployed.

In this scenario, some scripts are also used to connect the EC2 with the project. Generally this scripts are YAML file and JSON file.

CONCLUSION

I can honestly say that my time spent interning at Gugalia Fintech was one of the important and professional period my educational life. My main motive was to grab the opportunity and work for the company and gain as much as experience I can in that period and get to learn that professionalism and also to get some industrial exposure. Not only I learnt technology skills like AWS Cloud and automation tools, I got to know how corporate world works. This experience gave me somewhat clarity of what I want to do and what I can do in my professional life after college. I consider myself very lucky to get such an opportunity to work with Gugalia Fintech. Also, thanks to Rishi Sir for the continuous guidance when needed and the to the Internal Guide Hemant Yadav Sir for lending my this opportunity.

REFERENCE

[1]. https://www.youtube.com/watch?v=Buh3GjHPmjo

[2]. https://www.youtube.com/watch?v=zkNdHv1iMgY

[3]. https://www.youtube.com/watch?v=OnI_nXa5yH4

[4]. https://aws.amazon.com/devops/what-is-devops/#:~:text=DevOps%20is%20the%20combination%20of,development%20and%20infrastructure%20management%20processes.

[5]. https://docs.aws.amazon.com/whitepapers/latest/aws-overview/six-advantages-of-cloud-computing.html

[6]. https://www.cloudflare.com/learning/cloud/what-is-cloud-migration/

[7].https://www.google.com/search?q=what+is+devops+in+software+engineering&rlz=1C1YTUH_enIN938IN938&oq=What+is+DevOps&aqs=chrome.1.0i131i433i512j0i512l9.17801j0j7&sourceid=chrome&ie=UTF-8