A

Seminar Report

On

Techniques to Secure Data on Cloud: Docker Swarm or Kubernetes?

Submitted in partial fulfillment of the requirements for the Award of the Degree

of

Master of Computer Applications

of

APJ Abdul Kalam Technological University

Submitted by

Sashwat K

Reg No: TVE17MCA042

Department of Computer Applications

COLLEGE OF ENGINEERING, TRIVANDRUM

OCTOBER 2019

DEPARTMENT OF COMPUTER APPLICATIONS

COLLEGE OF ENGINEERING TRIVANDRUM

CERTIFICATE

Certified that this Seminar report entitled, ***"Techniques to Secure Data on Cloud: Docker Swarm or Kubernetes?"*** *is the paper presented by* ***"Sashwat K" (Reg No: TVE17MCA042)*** *in partial fulfillment of the requirements for the award of the degree of Master of Computer Applications of APJ Abdul Kalam Technological University during the year 2019.*

Prof. Jose T Joseph. Prof. Sabitha S.

Co-ordinator **Head of the Department**

Acknowledgement

First and for most I thank **GOD** almighty and to my parents for the success of this seminar. I owe a sincere gratitude and heart full thanks to everyone who shared their precious time and knowledge for the successful completion of my seminar.

I would like to thank **Dr Jiji C V**, Principal, College of Engineering Trivandrum, who helped me during the entire process of work.

I am extremely grateful to **Prof.Sabitha S**, HOD, Dept of Computer Applications, for providing me with best facilities and atmosphere for the creative work guidance and encouragement.

I would like to thank my coordinator, **Prof.Jose T Joseph**, Dept of Computer Applications, who motivated me throughout the work of my seminar.

I profusely thank other Asst. Professors in the department and all other staffs of CET, for their guidance and inspirations throughout my course of study.

I owe my thanks to my friends and all others who have directly or indirectly helped me in the successful completion of this seminar. No words can express my humble gratitude to my beloved parents and relatives who have been guiding me in all walks of my journey.

Sashwat K

Abstract

In the current world with immense technological advancement, the Information Technology(IT) world is switching from physical storage to cloud storage since the "cloud" providers supply resources on demand over the Internet. Cloud computing is a successful and speedy evolving model with new features and capabilities being announced regularly. The concept of this is known as "pay as you use" which enables the firms to shift to cloud. Due to this, security of such data has become an issue. Security of cloud-based applications is one of the key concerns of cloud customers. These three principles of cloud security are Availability, Confidentiality and Integrity. One of the most efficient ways is with the help of Container Clustering. There are various ways in which container clustering can be achieved. This paper presents the study and the comparison between two such technologies, i.e. Docker Swarm and Kubernetes.

Contents

List of Figures

Chapter 1

Introduction

"A Cloud Computing model is a model which enables convenient, on demand network access to a shared pool of configurable computing resources and services. These resources include: networks, servers, storage, apps and services. These resources can be rapidly provisioned and released with minimal management effort or service provider interaction."Cloud provides portability and flexibility of online services.

Introduction continues from 1.1 to 1.5. Chapter 2 of this paper discusses the key concerns of Data Security on Cloud. Chapter 3 discusses the various Issues. Chapter 4 discusses the Various Techniques to Secure Data on Cloud. Chapter 5 to 8 discuss about container clustering using docker and its working. Chapter 9 and 10 discuss about Kubernetes and why it is better than Docker Swarm.

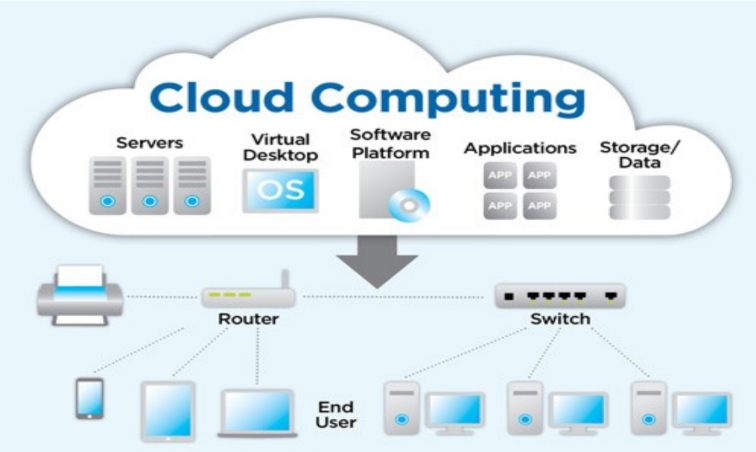

Figure 1.1: Cloud Computing

1.1 The Five Essential Characteristics

The Five essential characteristics of Cloud Computing are given below. These are vital features of this technology which define its usability and efficiency.

1. On-Demand Self Service "A consumer can one-sidedly provision computing capabilities. These capabilities include server time and network storage. They can be provisioned whenever required without having human interaction with the service providers."

2. Broad Network Access here are capabilities available over the network and can be accessed through various platforms such as mobile phones, laptops, PDAs and other electronic devices.

3. Resource Pooling The resources are amalgamated to serve multiple customers. This can be done with the help of multi-tenant models. These resources are dynamically assigned and reassigned depending on the consumer's demands. Examples of such resources include storage, Memory, Processing, Network, Bandwidth and Virtual Machines.

4. Rapid Elasticity The capabilities can be provided elastically and rapidly, sometimes also automatically, for quick scale outs, and rapid releases to quickly scale in to the customer. Such capabilities are unlimited and can be bought whenever needed and in any quantity.

5. Measured Service The use of resources in cloud systems are automatically controlled and optimised. The usage of the services can be monitored and controlled. They can also be reported providing transparency for both, the provider and consumer.

1.2 The Three Service Models

There are three service models which are implanted in Cloud Computing. They are as follows:

1. Infrastructure as a Service(IaaS) In IaaS, the consumer is provided with networks, storage, processing and various other computing resources. IaaS enables he consumer or user to deploy and execute the software. This software includes OSs and applications. The consumer is not in control of the cloud infrastructure. However, a consumer can control the OS, storage and the deployed apps. A consumer may also have partial control over selection of networking components such as host firewalls.

2. Platform as a Service (PaaS) PaaS allows the consumers to deploy their applications which are created using programming languages and various tools onto the cloud infrastructure.

The control that the consumer has is similar to that of IaaS.

3. Software as a Service (SaaS) SaaS enables the consumer to use the capabilities such as applications which is provided by the service provider which runs on the infrastructure. These applications can be accessed from various devices through a thin client interface such as a web browser like Google Chrome, Mozilla Firefox, IE, Opera etc. The control that the consumer has is similar to that of IaaS and PaaS. However, the consumer may have partial control over user specific application configuration settings.

1.3 The Four Deployment Models

The cloud can be deployed in four different ways. They are as follows:

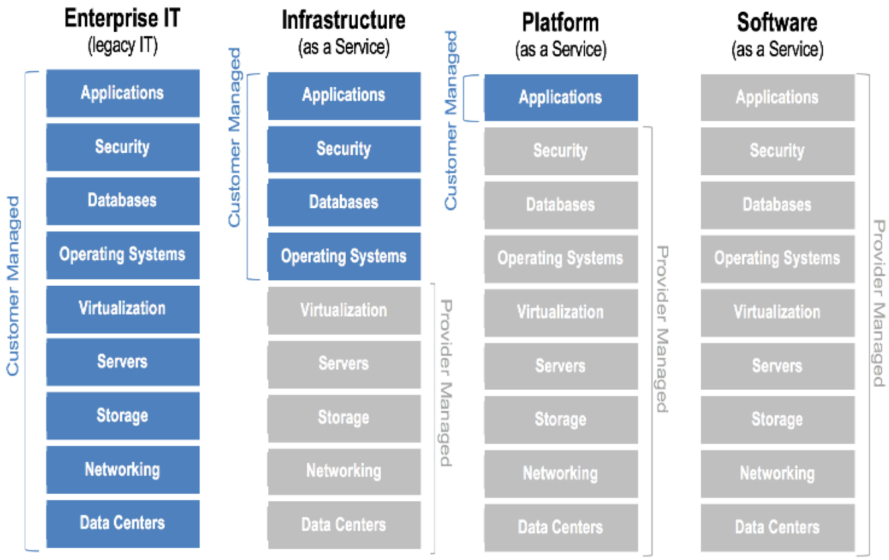

Figure 1.2: The Service Models

1. Public Cloud As the name suggests, this type of cloud is available to the general public. It is owned by organizations that sell cloud services. These service providers include Amazon's Elastic Compute Cloud(EC2), Microsoft Azure Service Platform, Sun Cloud, Google App Engine, IBM's Blue Cloud, etc.

2. Private Cloud This cloud infrastructure is operated solely for an organization. "The objective of a private cloud is not sell as-a-service offerings to external customers but to gain the benefits of cloud architecture without giving up the control of maintaining your own data centre."

A particular private cloud can be owned only by a single entity or organization.

3. Community Cloud The infrastructure of the community cloud is shared by multiple organizations. It supports a specific community that has similar or same concerns/issues. It may be managed by the organizations or a third party.

4. Hybrid Cloud A hybrid cloud is the incorporation of two or more clouds. Hybrid clouds can remain as unique entities. However, they are bound together by standardized technology that enables portability of data and applications. Users may use any of the deployment models as per their requirements and needs.

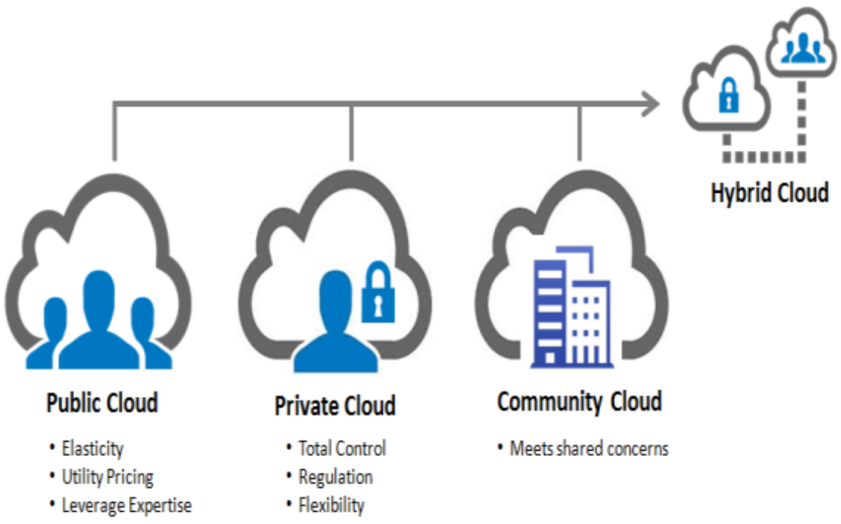

Figure 1.3: Cloud Computing Models

1.4 Pros and Cons of Computing

Every technology has its own advantages and disadvantages. The pros and cons of Cloud Computing are as given below.

1.4.1 Advantages

1. Minimized Costs
2. Higher Resource Sharing.
3. Consumption based cost.

4. Efficient power saving.
5. Faster time to deploy new services.
6. Management moves to Cloud Provider

1.4.2 Disadvantages

1. Reliability
2. Availability
3. Security and Privacy Latency and Bandwidth guarantees.
4. Absence of robust SLAs
5. Uncertainty around inter-operablity, portability and lock-in.
6. Compliance/Regulatory Laws mandate on-site ownership of data.

There are various advantages and disadvantages of Cloud Computing. Users must keep them in mind when they choose Cloud Computing for installing their applications or storing their data on Cloud servers.

Chapter 2

Data Security in Cloud Computing

Data security is one of the key concerns with Cloud Computing. Data security has three major factors to be considered. Security refers to the availability, integrity and confidentiality of data. Not providing or guaranteeing this may pose major issues for cloud vendors.

1. Availability It is one of the most problematic issue faced by the consumers. Most cloud vendors have experienced downtime of their services which have affected majority of the users of cloud services. For example, Amazon servers have faced such issues which have been said to be Denial of Service attack. A cloud server must be available at all times and if not, a particular time period should be given as notice for the services to resume.

2. Integrity This is another important feature as cloud vendor or provider must provide. Data integrity refers to the accuracy and consistency of data stored on the cloud or any database as a matter of fact. Thin clients can be used to maintain integrity of data or security of data on the client side. This is possible since thin clients use as few resources as possible and they do no store any data. By doing this, personal information such as passwords cannot be stolen.

3. Confidentiality Personal or confidential data stored on the cloud should not be accessible by anyone other than the authorised entity. For example, data from parent company X which has been stored in child company Y should not be accessible by the employees of company Y, since it is confidential data of company X.

These are the factors essential for Data Security in any kind of network platform. Guaranteeing the mentioned factors ensures the security of the consumers or customers data.

Chapter 3

Issues

There are various issues that require addressing before an organization or enterprise considers switching to the cloud computing model. These issues are Privileged User Access, Data Location, Recovery, Long- term Viability, Data Segregation and Regulatory Compliance.

The below mentioned graph shows the issues or concerns and its corresponding level/ percentage to which the consumers are affected.

Chapter 4

Various Methods of Data Security on Cloud

There are various algorithms and measures we can use to secure data on the virtualized environment of the cloud. They are Cryptography, Homomorphic Encryption, Diffie Hellman Algorithm, Rivest-Shamir-Adleman Algorithm(RSA), Container Clustering using Dockers.

Cryptography can be used to encrypt the data stored on the cloud. But encrypting a large amount of data can be very challenging and time consuming.

Other Encryption techniques such as Homomorphic Encryption and various algorithms such as RSA and DH algorithms can also be used to do the same. The major issue in using them is the excessive use of resources which eventually lead to an increased cost and time.

Out of the five methods, Container Clustering using Dockers is one of the better and more efficient options to use for securing data on the cloud.

Using dockers has its own advantage and because of this it beats all the other methods of securing data.

Chapter 5

What is a Docker? Relation between Cloud Computing and Docker

5.1 Docker

A docker is an open source enterprise, which can be used to launch any type of application/module as a light- weight container.

Due to dockers, there is independence between applications, infrastructure, developers and IT ops. This unlocks their potentials and a model is created for better collaboration and innovation.

A docker is simple, agile, secure, portable and also reduces cost.

5.2 Relation between Cloud Computing and Docker

1. Docker with PaaS: Dockers are used as fundamental units by various service providers. These include AWS Elastic Beanstalk, OpenShift or Dokku. PaaS provides automation as well as the coding environment. This should be flexible and available on demand. There shouldn't be any downtime or delay. There has been a shift from virtual machines to docker containers in many IaaS layers.

2. Docker with IaaS: Besides coding environment, there is a provision for computing platforms or a virtual server which is completely isolated from the host. It is very easy to setup a docker cluster which may have an installed webserver.

Chapter 6

What is a Container?

"A container image is a lightweight, stand-alone, executable package of a piece of software that includes everything needed to run it: code, runtime, system tools, system libraries, settings."

It is available for both Windows and Linux based applications. We can call it platform-independent since the "containerized software will always run the same, whatever be the environment."

The containers isolate the software from its environment, This helps in minimizing the conflicts between groups running different software or application on the same infrastructure.

Chapter 7

Container Clustering and GlusterFS

7.1 Container Clustering

The job of the container is to turn a number of docker engines into one, single body or entity. This means that the management and launch of containers in multiple systems works as a single body. Container Clustering can be achieved by utilizing the Swarm Manager.

7.2 GlusterFS

GlusterFS is a network file system developed by RedHat. It allows the storage of objects and files in cluster of storage servers. It stores the data in a distributed manner. This can be done either by stripping the data or replicating the data or both on multiple storage servers. GlusterFS is a scale-out network-attached file storage system. GlusterFS can be used for various applications such as cloud computing, streaming media services, and content delivery networks. GlusterFS was developed originally by Gluster, Inc. However, Gluster, Inc. was acquired by RedHat in 2011.

Chapter 8

Encryption over Cloud Model and its Process using Container

8.1 Swarm Manager

The job of a Swarm Manager is to send the plain text(unencrypted) code from the user a node which is running Docker Cryptography Container. Algorithms such as RSA and AES can be used for encryption of the file being uploaded.

After encryption, the encrypted data is sent here. SSH protocol is used to handle the transaction of the files.

After the transaction, the data is encrypted. However, it needs to be stored such that it is highly available. After the data is encrypted, it is sent back to a Gateway Manager.

The data from the Gateway Manager is then sent to another node containing Storage manager(GlusterFS) This storage manager is connected to a new node containing multiple Docker containers. Then the encrypted data is distributed. GlusterFS is used to store the container cluster. By doing this, the encrypted files will be stored on multiple docker containers, thus forming a cluster of storage. This enables us to store data in chunks. This prevents the attacker from gaining the data since even if the node is compromised, the attacker will still have to target the containers. However, each container does not contain all the data as the data was distributed in previous steps. Here, having multiple containers helps.

Similarly, we use the same process for decryption. First, the user downloads the required file, then the file is decrypted by the docker image and sent to the swarm manager, and from there the decrypted file is received.

The following diagram shows the exact working model of Container Clustering using Dockers for Securing data on Cloud Servers or Storage.

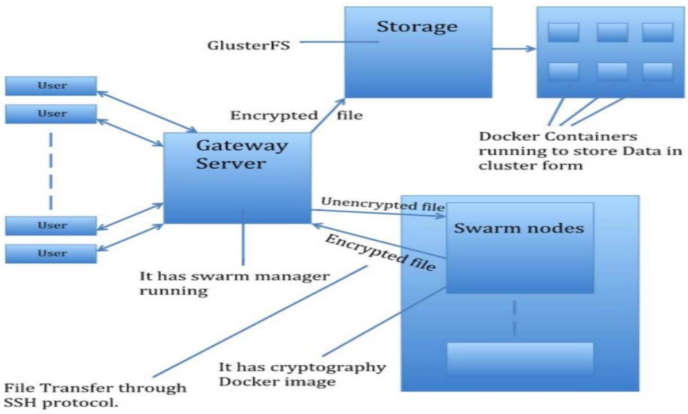

Figure 8.1: Container Clustering using Dockers Model

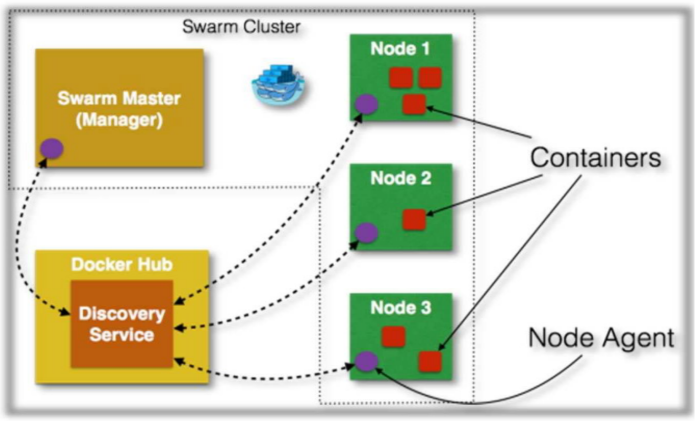

Figure 8.2: Swarm Cluster Layout

Chapter 9

Using Kubernetes Instead of Docker Swarm

1. The efficiency of the afore mentioned model can be increased by the utilization of Kubernetes rather than Docker Swarm.

2. When we compare the two container clustering tools, namely Kubernetes and Docker, Kubernetes excel in both power as well as performance. It also enables greater flexibility in container management.

3. Like Docker, Kubernetes is also an open- source system that provides similar functions such as automating deployment, scaling, and the management of containerized applications.

4. Therefore, we can use Kubernetes for optimizing the existing design to make it more efficient. It also provides fault tolerance.

9.1 Characteristics of Kubernetes

1. Planet Scale

2. Never Outgrow

3. Run Anywhere

9.2 Features of Kubernetes

1. Automatic Binpacking
2. Self-Healing
3. Horizontal Scaling
4. Storage Orchestration
5. Batch Execution
6. Automated Rollouts and RollBacks
7. Service Discovery and Load Balancing
8. Secret and Configuration Management

Chapter 10

Why choose Kubernetes over Docker Swarm

The major question is why should we use Kubernetes instead of Docker Swarm? What good will it do? The following points will compare both the technologies and answer the above questions.

Kubernetes is a much more powerful tool than Docker Swarm. It can handle containers and offer immense scalability and automation at the same time. This means it possesses greater efficiency than the Docker Swarm.

Kubernetes has an in-built library and process for monitoring and logging, which lacks in the Docker Swarm. Hence, the Docker Swarm has to use third party applications to these features.

Kubernetes overcomes the constraints of Docker and Docker API.

It is more widely deployed than Docker Swarm.

If a node fails, it is replaced so fast that you don't even come to know a node has failed unless you do a detailed troubleshoot on it.

Since we all seek after efficiency and accuracy, Kubernetes is a much better choice than Docker Swarm in almost all aspects.

The performance of Kubernetes surpasses that of the Docker Swarm.

However, there is one drawback of Kubernetes, i.e., the installation process. It is complex and time consuming. But, if the end result is so spectacular, such a drawback can be neglected.

Chapter 11

Conclusion

In recent years, cloud computing has become a very popular system to both individuals as well as establishments. However, due to this increase in popularity, the cyber threats have also increased. To prevent such threats and to protect the confidential information, certain measures have to be taken.

The technology of Containers is used to secure data being uploaded on the cloud.

In the proposed model, there is a provision of a process which encrypts and decrypts the user data. This can be achieved by launching a container for each and every user. By doing this, resources are utilized optimally and the load on the multiple servers can be balanced.

The model can be improved by using Kubernetes due to its features and more efficient behaviour in containerization and deployment. A better performance is all we sought after.

Bibliography

[1] Social GAN: Socially Acceptable Trajectories with Generative Adversarial Networks by Agrim Gupta1 , Justin Johnson1 and Alexandre Alahi.

[2] A. Alahi, K. Goel, V. Ramanathan, A. Robicquet, L. Fei-Fei, and S. Savarese. Social lstm: Human trajectory prediction in crowded spaces. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.

[3] G. Antonini, M. Bierlaire, and M. Weber. Discrete choice models of pedestrian walking behavior. Transportation Research Part B: Methodological.

[4] Generative Adversarial Networks: Introduction and Outlook Kunfeng Wang, Member, IEEE, Chao Gou, Yanjie Duan, Yilun Lin, Xinhu Zheng, and Fei-Yue Wang, Fellow, IEEE.

[5] Analyzing the Variety Loss in the Context of Probabilistic Trajectory Prediction by Luca Anthony Thiede and Pratik Prabhanjan Brahma