2024

Cloud Infrastructure and Services

Dr. Babasaheb Ambedkar Open University

Cloud Infrastructure and Services

Expert Committee

Prof. (Dr.) Nilesh K. Modi Professor and Director, School of Computer Science, Dr. Babasaheb Ambedkar Open University, Ahmedabad	(Chairman)
Prof. (Dr.) Ajay Parikh Professor and Head, Department of Computer Science Gujarat Vidyapith, Ahmedabad	(Member)
Prof. (Dr.) Satyen Parikh Dean, School of Computer Science and Application Ganpat University, Kherva, Mahesana	(Member)
M. T. Savaliya Associate Professor and Head Computer Engineering Department Vishwakarma Engineering College, Ahmedabad	(Member)
Mr. Nilesh Bokhani Assistant Professor, School of Computer Science, Dr. Babasaheb Ambedkar Open University, Ahmedabad	(Member)
Dr. Himanshu Patel Assistant Professor, School of Computer Science, Dr. Babasaheb Ambedkar Open University, Ahmedabad	(Member Secretary)

Course Writer

Dr. Jitendra Bhatia	Assistant Professor, Vishwakarma Government Engineering College, Ahmedabad
Dr. Chirag Patel	Associate Professor, Vishwakarma Government Engineering College, Ahmedabad

Content Reviewer and Editor

Prof. (Dr.) M.T.Savaliya	Professor, Vishwakarma Government Engineering College, Ahmedabad
Prof. (Dr.) Nilesh K Modi	Professor and Director, School of Computer Science, Dr. Babasaheb Ambedkar Open University, Ahmedabad

ISBN -

Printed and published by: Dr. Babasaheb Ambedkar Open University, Ahmedabad

Dr.Babasaheb Ambedkar Open University

MSCDS-205

Cloud Infrastructure and Services

Block-1: Introduction to Cloud Computing

Unit-1
Introduction to Cloud 04

Unit-2
Cloud and Other Architectures 14

Unit-3
Cloud Computing in a nutshell 25

Unit-4
Cloud Types and Models 36

Block-2: Cloud Computing Services

UNIT-1
Infrastructure as a Service & Platform as a Service 48

UNIT-2
Software as a Service and Database as a Service 60

UNIT-3
Security as a Service 74

UNIT-4
Specialized Cloud Services 87

Block-3: Application Architecture for Cloud and Cloud Deployment Techniques

UNIT-1
Introduction of Cloud Application 99

UNIT-2
Cloud Application Requirement 109

UNIT-3
Successful Cloud Implementation 118

UNIT-4
Improving Cloud database performance and Cloud Service Brokerage 125

Block-4: Risk, Security, Consequences and Cost of Cloud Computing

UNIT-1
Risk in Cloud Computing 135

UNIT-2
Data Security in the Cloud 144

UNIT-3
Application Security in the Cloud 153

UNIT-4
Costs in Cloud Computing 163

Block-1

Introduction to Cloud Computing

1

Unit 1: Introduction to Cloud

Unit Structure

1.1. Learning Objectives

1.2. Introduction to Cloud

1.3. Characteristics of Cloud Computing

1.4. Other Similar Configuration

1.5. Let us sum up

1.6. Check your Progress: Possible Answers

1.7. Further Reading

1.8. Assignment

1.9. Activities

1.10. Case studies

1.1 LEARNING OBJECTIVE

After studying this unit, student should be able to:

- Understand the basics of cloud computing
- Know the other similar configurations
- Differentiate between applications oriented versus service oriented architecture.

1.2 INTRODUCTION TO CLOUD

Cloud computing was formally put forward in the IT industry by IBM who announced its plan of cloud computing at the end of 2007. There is an increase in the need for storage and computing due to the growth of the IT industry. In general terms, the cloud refers to anything that involves delivering hosted services over the internet. Cloud Computing can be defined as the emerging technology for on-demand delivery of IT resources over the Internet with pay-as-you-go pricing. Although computer experts and manufacturers have different views about cloud computing, it's undeniably true that cloud computing has brought opportunities and challenges to the IT industry.

Cloud computing can also be defined as a large – scale distributed computing paradigm in which the cloud users or service providers are having their own private infrastructure and the different types of hosted services are provided to clients using virtual machines.

The most acceptable definition of cloud computing was introduced by the National Institute of Standards and Technology (NIST): “Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction.”

The following figure depicts the basic cloud architecture which involves cloud controller, resource controller, and various resource pools. The job of the cloud controller is to act as an interface between users’ requests and computing resources

on the cloud. Cluster controller looks after the resource management whereas resource pool consists of several computing nodes.

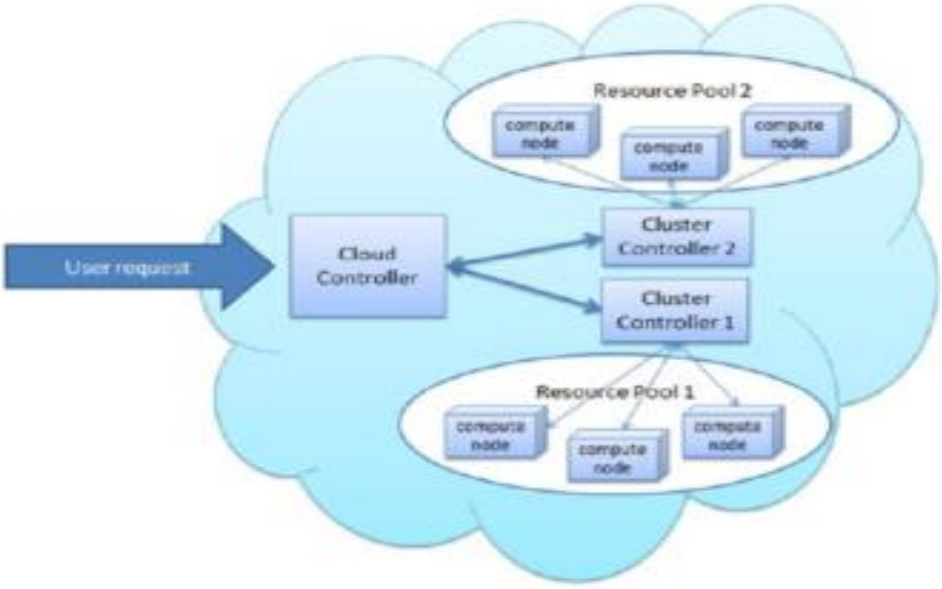


Figure: Cloud architectural view

The visualization of cloud computing includes cloud infrastructure, cloud computing, and cloud application. Similarly, the core three services of cloud computing include Infrastructure-as-a-Service (IaaS), Platform-as-a-Service (PaaS), and Software-as-a-Service (SaaS) which are accessible to the user as public, private, or hybrid cloud. Resource reservation and release of resources on demand are the key characteristics of cloud computing that helps to reduce the cost based on pay as per use model.

Cloud computing helps enterprises, governments, public and private institutions, as well as research organizations for shaping and converting their traditional computing system to demand-driven computing systems. Practical examples of such systems are as follows: a) large enterprises offloading some of their activities to cloud-based systems, b) small enterprises and startups can translate their ideas into business models quickly without excessive upfront costs, c) System/Application developers do not have to worry for dealing with the complexity of infrastructure management and scalability issues, d) End users can access the service and documents from everywhere and any device.

The pay-as-you-go basis model of cloud computing not only enables the opportunity to access IT services on demand but also introduces a new thinking paradigm about how IT services and resources should be perceived: as utilities. The public cloud, private/enterprise cloud, and hybrid cloud are the most common deployment models.

1.3 CHARACTERISTICS OF CLOUD COMPUTING

Cloud computing has some interesting characteristics that bring benefits to both cloud service users and service providers. They are On-demand self-service: The cloud computing services does not require any human administrations as users can have resources as per their need and also have control over them to release or to turn off to avoid resource wastage.

Scalability: It refers to the increase or decrease in the resource as and when necessary. Users or services may request resources more than the demand so the cloud providers should have sufficient capacity to meet customers' needs.

Multi-tenant: It refers to the users and they may be sharing the same resource (e.g. memory, CPU, etc.) with another tenant which is transparent to them. Ensuring the security aspects is the prime role of the cloud provider so that one tenant won't be able to access others' data.

Computation and storage resource: Resource provisioning, computing processes including billing, and deployment should be self-service and automated, involving much less manual processing. For example, if a machine on which our service is hosted fails, then the cloud provider should be able to tackle the failure.

Reliability: Reliability refers to the committing to uptimes of the customer service i.e., expectations of zero downtime.

Broad network access: All the cloud computing services are accessible over the standard networks and heterogeneous devices supporting various client platforms.

Resource Pooling: The computing resources of a provider are pooled together to serve multiple users using a multi-tenant model, with dynamic provisioning of different physical and virtual resources on user demand.

Rapid elasticity: It refers to the seamless scaling in and out of the cloud resources. The provisioning and releasing of the cloud services/resources takes place as per the need of the user and seems infinite to them.

Measured service: It reflects the pay-as-you-go model of utility computing and provides the accounting service of a resource. It includes monitoring, measuring, and billing transparently based on resource and service utilization.

Reduced time to market: It provides an opportunity for small organizations and start-ups without the need for large investments to start their business.

1.4 OTHER SIMILAR CONFIGURATIONS

The idea of computing services by leveraging distributed computing facilities has been around for a long time. The core technologies or similar configurations that played an important role in the realization of cloud computing are distributed systems, virtual servers, service-oriented computing, utility computing, grid computing, and cluster computing.

Distributed Systems

A distributed system refers to the collection of independent computers that appears to its users as a single computing system. The primary purpose of distributed systems is to share resources and to utilize them better. The same is applicable to cloud computing where resources like infrastructure, runtime environments, and services are rented to users. The availability of large computing facilities is one of the driving factors for cloud computing. The offering of computing capabilities as a service can be a better opportunity for utilization of the infrastructure of IT giants (Amazon, Google, Microsoft, IBM, etc.). Distributed systems often exhibit other properties such as openness, heterogeneity, scalability, transparency, concurrency, and continuous availability. To some extent, these also characterize Clouds, especially in the context of concurrency, scalability, and continuous availability.

Three major milestones which led to cloud computing are mainframe computing, cluster computing, and grid computing.

a) **Mainframes**: It refers to the large computational facilities leveraging multiple processing units and is presented as a single entity to users. Mainframes were very

powerful, highly reliable computers specialized for massive data processing like online transactions, enterprise resource planning, and other operations involving the processing of significant amounts of data. Batch processing was the main application of mainframes. Now their popularity and deployments have reduced, but evolved versions of such systems are still in use due to the reliability feature capable of tolerating failure transparently.

b) **Clusters**

More than one computer combines to form a cluster. Cluster computing follows distributed systems as its principle. Cluster computing started as a low-cost alternative to the use of mainframes and supercomputers. It exhibits very high performance but from the end-users perspective, they feel like working on a standalone system. LAN acts as the connection unit leveraged by a high-bandwidth network and controlled by specific software tools that manage them as a single system. Cluster computing was cheaper than mainframes and made available high-performance computing to a large number of groups, including larger organizations, universities, and small research labs. The computational power of commodity machines was used to solve problems previously manageable on expensive mainframes and supercomputers. Clusters could be easily extended if more computational power was required.

Some of the most widely used Cluster Computers are Petroleum Reservoir Simulation, Google Search Engine, Earthquake Simulation, Weather Forecasting.

c) **Grids**

Grid computing appeared in the early 90s as an evolution of cluster computing. Grid computing and cloud computing are conceptually similar and can be easily confused. Cloud computing is often considered the successor of Grid computing. Grid computing proposed a new approach to access large computational power, huge storage facilities, and a variety of services. The concepts are quite similar and both share the same vision of providing services to the users through sharing resources among a large pool of users. Both are based on network technology and are capable of multitasking meaning users can access a single or multiple application instances to perform different tasks. Resources are managed centrally in the cloud whereas

resources are distributed over grids.

Virtualization

Virtualization refers to the separation of the server software away from the hardware. It is one of the core technologies for cloud computing in which a virtual server includes the OS, the applications, and the storage. Virtualization allows the creation of different computing virtual environments that simulate the interface that is expected by a guest. Hardware virtualization is one of the examples that allows simulating the hardware interface expected by an operating system and different software stacks can coexists on the top of the same hardware and operate in isolation from each other. These stacks are contained inside virtual machine instances. Virtualization allows cloud computing to deliver virtual servers on demands such as Amazon EC2, Vmware vCloud, IBM Bloomix, and others.

Service Oriented Computing

Service-Oriented Computing (SOC) refers to the computing strategy that utilizes services as fundamental elements for developing applications or solutions. Service orientation is the core reference model for cloud computing systems. To build the service model, SOC relies on the Service Oriented Architecture (SOA). SOA refers to the way of reorganizing software applications and infrastructure into a set of interacting services. Services perform functions, which can be anything from simple requests to complicated business processes. Services providers offer various services implementation and provide related support. Services should be technologically neutral, loosely coupled, and support location transparency. Despite the similarities between Cloud Computing and SOA, they are not the same. We pay for the infrastructure or computing platform as an outcome while in SOA, we pay for the technology behind the service computing. In Cloud Computing, the services provided can go up and down the stack but in SOA the services are software components. Service-oriented computing presented two important concepts, which became fundamental requirements for cloud computing: *quality of service (QoS)* and *Software-as-a-Service (SaaS)*.

1.5 LET'S SUM UP

Points to ponder

- Cloud computing means storing and accessing the data and programs on remote servers that are hosted on internet instead of computer's hard drive or local server.
- SOA is a framework that allows business processes to be highlighted to deliver interoperability and rapid delivery of functionality
- Cloud computing decreases the hardware and software demand from the user's side.
- Grid computing is a network based computational model that has the ability to process large volumes of data.
- Grid is application oriented while cloud is service oriented.
- Virtualization allows the creation of different computing virtual environments that simulate the interface that is expected by a guest.
- The cloud servers are owned by infrastructure providers and are placed in physically disparate locations.

1.6 CHECK YOUR PROGRESS: POSSIBLE ANSWERS

1) What type of computing technology refers to services and applications that typically run on a distributed network through virtualized resources?

a. Distributed Computing
b. Cloud Computing
c. Soft Computing
d. Parallel Computing

2) Which one of the following options can be considered as the Cloud?

a. Hadoop
b. Intranet
c. Web Applications
d. All of the mentioned

3) Cloud computing is a kind of abstraction which is based on the notion of combining physical resources and represents them as _____resources to users.

a. Real
b. Cloud
c. Virtual
d. none of the mentioned

5) Which one of the following cloud concepts is related to sharing and pooling the resources?

a. Polymorphism
b. Virtualization
c. Abstraction
d. None of the mentioned

8) which of the following is an essential concept related to Cloud?

a. Reliability
b. Abstraction
c. Productivity
d. All of the mentioned

1.7 FURTHER READING

1. Here is an article about the cloud computing beginners:
   https://www.guru99.com/cloud-computing-for-beginners.html

1.8 ASSIGNMENTS

1) Discuss in detail, the advantages of cloud computing.
2) How does grid computing differ from cloud computing?
3) What do you mean by virtualization?
4) Which are the technologies that cloud computing relies on?
5) Discuss the role of the cloud computing in education sector?
6) Explain the various advantages of cluster computing.

1.9 ACTIVITIES

Q1. Multiple Choice Questions:

1) All cloud computing applications suffer from the inherent _____ that is intrinsic in their WAN connectivity.
   a. propagation
   b. latency
   c. noise
   d. All of the mentioned

2) Cloud computing is a _____ system and it is necessarily unidirectional in nature.
   a. stateless
   b. stateful
   c. reliable
   d. all of the above

3) Which of the following is related to service provided by Cloud ?
   a. sourcing
   b. ownership
   c. reliability
   d. none of the above

1.10 CASE STUDIES

Prepare a list of services provided by giant cloud service providers like Amazon, IBM, and Microsoft Azure etc.

Unit 2: Cloud and Other Architectures

2

Unit Structure

1.1. Learning Objectives

1.2. Cloud Layered Architecture

1.3. Peer to peer architecture

1.4. Client server architecture

1.5. Grid computing

1.6. Server Virtualization

1.5. Let us sum up

1.6. Further Reading

1.7. Assignment

1.8. Case studies

1.1 LEARNING OBJECTIVE

After studying this unit, student should be able to:

- understand the basic layered architecture of cloud
- explore various other architectures
- exploit the advantages of virtualization

1.2 CLOUD LAYERED ARCHITECTURE

The cloud layered architecture can be viewed as a collection of hosted services over the Internet. The layered architecture refers to the system architectural style that describes the physical organization of processes and components over distributed infrastructure. Cloud as a layered architecture is shown in Fig.

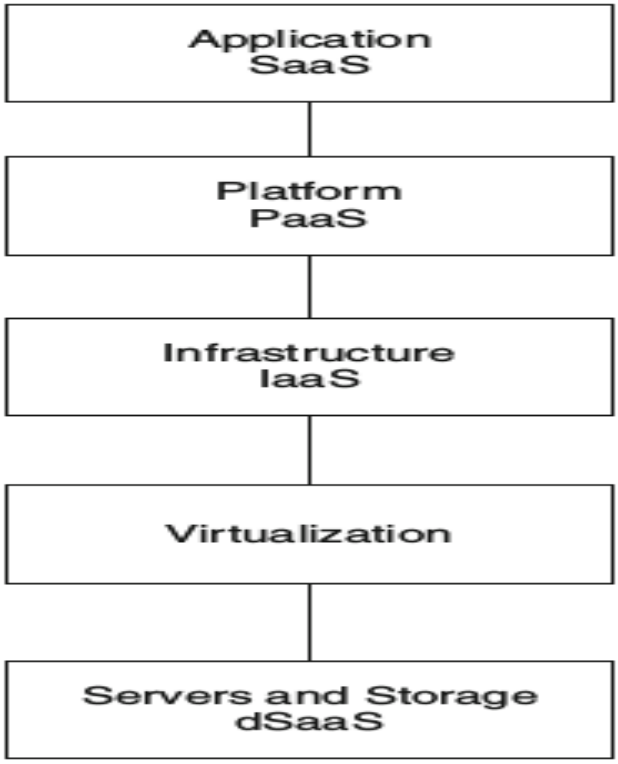


Fig: Layered Architecture of Cloud Computing

These services are divided into three main categories: infrastructure as a service (IaaS), platform as a service (PaaS), and software as a service (SaaS). IaaS provides the virtualized infrastructure for the necessary computing. PaaS as its name signifies provides the platform for creating the computing scenario on top of virtualized software. It includes operating systems and required services for a

particular application. Finally, SaaS which is on the top of a layered stack offers various deployed services through cloud computing usually includes IT services. Users can run applications remotely from the cloud. The virtualized computing platform enables guaranteed computing power and reserved bandwidth for storage and Internet access. In other words, PaaS is IaaS with a custom software stack for the given application. The data-Storage-as-a-Service (dSaaS) provides storage service for the storage.

1.3 PEER TO PEER ARCHITECTURE

A peer-to-peer architecture is a network of hosts in which resource sharing, processing, and communications control are completely decentralized. As its name signifies, this architecture introduces asymmetric architecture in which each host can act as a server or provider of certain services. Each host, called peers, plays the same role and incorporates the capabilities of both the client and server. Hence, this architecture is best suited to highly decentralized architecture. Each peer acts as a server when it receives requests from other peers and as a client when it requests to other peers for certain services. This architecture is easily scalable in terms of adding peers to existing overlay networks. The only disadvantage with this architecture is that the management of the implementation of algorithms is more complicated compared to other architectures.

Security is one of the prime concerns as in this model; users are allowed to access the resources and services available on the network on their own. All the peers are equal in terms of using and providing the services and resources. Due to its decentralized nature, P2P networks can survive attacks fairly well since there is no centralized server. There are important benefit-related dissimilarities between cloud and peer-to-peer deployments. Distinctive benefits of cloud computing are that it can be easily scaled to meet growth demands, enables access to any type of hosted applications, does not burden user-end devices, and needs to be configured with the highest levels of security. On the other hand, peer-to-peer deployments are relatively inexpensive and simple to set up and manage.

P2P architecture works best when there are lots of active peers in an active network, so new peers joining the network can easily find other peers to connect to. If a large

number of peers drop out of the network, there are still enough remaining peers to pick up the slack. If there are only a few peers, there are less resources available overall. For example, in a P2P file-sharing application, the more popular a file is, which means that lots of peers are sharing the file, the faster it can be downloaded. P2P works best if the workload is split into small chunks that can be reassembled later. This way, a large number of peers can work simultaneously on one task and each peer has less work to do. In the case of P2P file-sharing, a file can be broken down so that a peer can download many chunks of the file from different peers at the same time. Some uses of P2P architecture includes file sharing applications (such as BitTorrent, kazaa etc.), Instant messaging, Voice Communication, High Performance Computing etc.

1.4 CLIENT-SERVER ARCHITECTURE

Client-Server architecture is very popular for distributed computing in which major components involved are server and client. These both the components interact with each other over the network using a suitable communication protocol. In network protocol stack, application layer provides the various communication protocols for the client server interaction.

The client server computing works with a system of request and response. The unidirectional communication takes place when the client initiates a request to the server and after processing the request server returns a response. This architecture is suitable in many to one scenario where multiple clients are requesting for services from a designated server. The best example of this type of communication is a web server which returns the web pages to the client that requested them. It is easy to replace, upgrade or relocate the nodes in the client server modelbecause all the nodes are independent and request data only from the server.All the nodes i.e. clients and server may not be build on similar platforms yet they can easily facilitate the transfer of data. The drawback of this architecture is that if all the clients requests the data simultaneously then the server may get overloaded leading to the network congestion. The cost of maintaining and setting client server architecture are quite high.

There are two major design models for the clients i.e., Thin-Client and Fat-Client model.In the thin client model, the client has a lightweight implementation concerning data retrieval like tasks with no considerable processing and server is meant for the load of data processing. The client in fat client model is responsible for processing, transforming the data before returning it back to the user, while the server has relatively light implementation concerning data access management.

There are three major components in the client-server model i.e., presentation, application logic, and data storage which are appropriately called tiers. The two major classes exist are two-tier architecture and three-tier/ N-tier architecture. The following are the some examples which will help you to understand the concept of the client server model architecture.

- Mail servers: Email servers are used for sending and receiving emails. There are different software that allow email handling.

- File servers: File servers act as a centralized location for files. One of the daily life examples to understand this is the files that we store in Google Docs. The cloud services for Microsoft Office and Google Docs can be accessed from your devices; the files that you save from your computer, can be accessed from your phone. So, the centrally stored files can be accessed by multiple users.

- Web servers: Web servers are high-performance computers that host different websites. The server site data is requested by the client through high-speed internet.

1.5 GRID COMPUTING

Grid computing refers to a group of computers in a network that works together for providing a virtual high-performance computing platform for larger tasks. It is also considered a subset of distributed computing. All the machines on the network work under the same protocol to act as a virtual high-performance compute engine. The task that they work on may include analyzing or processing huge datasets or simulating situations that require high computing power. Computers on the network

contribute resources like processing power and storage capacity to the network.

A grid computing network comprises three types of machines of the network.

1. **Control Node:**
   It acts as an administrative node which looks after the entire network and keeps accounting of the resources in the network pool. The control node must prioritize and schedule tasks across the network. It is the control node's job to determine what resources each task will be able to access. For controlling the network and its resources a software/networking protocol is used generally known as Middleware. This is responsible for administrating the network and the control nodes are merely its executors. As a grid computing system should use only unused resources of a computer, it is the job of the control node that any provider is not overloaded with tasks.
2. **Provider:**
   The computer which contributes it's resources in the network resource pool.
3. **User:**
   It refers to the end user that uses the resources on the network

Grid computing is not centralized, as there are no servers required, except the control node which is just used for controlling instead of processing. In single grid computing network can comprised of machine with different operating systems i.e., Multiple heterogeneous machines. In Grid computing, each task is divided into smaller tasks that are deployed parallel over distributed architecture. On the other hand, cloud computing is based on network technology where every user of the cloud has its private resource that is provided by a specific provider. Both the technologies have similar characteristics of resource pooling; however, they differ in their architectures, business model, interoperability, and certain different features. Currently, grid computing is being used in various institutions to solve complex mathematical, analytical, and scientific problems. There is a major concern regarding the software of the grid, as it is still in the evolution stage and also licensing across many servers may make it restrictive for some applications.

1.6 SERVER VIRTUALIZATION

Masking server resources from the users that include the identity, number of processors, operating systems, etc., refer to server virtualization. It is one of the fundamental technologies behind cloud computing, especially regarding infrastructure-based services. Server virtualization also refers to the process of dividing a physical server into multiple unique and isolated virtual servers using a software application. Each virtual server can run its own operating systems independently. Virtualization of servers allows the creation of a customizable, secure, and isolated execution environment for running applications. The fundamental basis behind the virtualization of this technology is the ability of software and hardware—to emulate an executing environment separate from the one that hosts such programs. Moreover, virtualization technologies provide a virtual environment for not only executing applications but also for storage, memory, and networking.

For example, we can run any operating system on the top of a virtual machine, which itself is running on another or the same operating system. Virtualization provides a great opportunity to build elastically scalable systems that can provide additional capability with minimum costs. Hence, this technique is widely used to deliver customizable computing environments on-demand. It offers various key benefits like scaling server ability, lessening the operating costs, reducing server complexity, increasing application performance, and easier workload deployment. The various types of server virtualization are as follows:

1) Hypervisor: It acts as a virtual machine monitor (VMM) and exist as a layer between the operating system and hardware. It provides the necessary features and services for the multiple operating systems. Hypervisor is responsible for identifying traps, responding to privileged CPU instructions etc. A host operating system also runs on top of the hypervisor to manage and administer the virtual machines.

2) Para Virtualization: In this type of virtualization, virtual machine does not implement full isolation of OS but rather provides a different API which is utilized when OS is subjected to alteration. The guest operating system is modified and recompiled before installation into virtual machine. Due to this, performance is enhanced as the modified guest operating system communicates directly

with the hypervisor and emulation overhead is removed. Examples of Para virtualization are VMware and Xen.

3) Full Virtualization: In this type of virtualization, virtual machine permit the execution of the instructions with running of unmodified OS in an entire isolated way. It is very much similar to Paravirtualization. It can emulate the underlying hardware when necessary. Full Virtualization uses binary translation and direct approach as a technique for operations. Examples of full virtualization are VMWare ESX Server, Microsoft and Parallels systems.

4) Hardware Assisted Virtualization: As it name signifies, it requires the support of hardware rest it is similar to para and full virtualization in terms of operation. Unmodified OS can be run as the hardware support for virtualization would be used to handle hardware access requests, privileged and protected operations, and to communicate with the virtual machine.

Examples: AMD – V Pacifica and Intel VT Vanderpool provide hardware support for virtualization.

5) Kernel level virtualization: In this type of virtualization which is specialized form of server virtualization, the host operating system runs on a specially modified kernel which contains extensions designed to manage and control multiple virtual machines each containing a guest operating system. A device driver is used for communication between the main Linux kernel and the virtual machine. Examples: User – Mode Linux(UML) and Kernel Virtual Machine(KVM)
6) System Level or OS Virtualization: In this type of virtualization, OS kernel allows the existence of multiple isolated user space instances.Also called shared kernel approach as all virtual machines share a common kernel of host operating system.

   Examples: FreeVPS, Linux Vserver, and OpenVZ are some examples.

1.7 LET'S SUM UP

Points to ponder

- A peer-to-peer architecture is a network of hosts in which resource sharing, processing, and communications control are completely decentralized.
- A peer-to-peer (P2P) architecture consists of a decentralized network of peers - nodes that are both clients and servers.
- Grid computing is a distributed architecture-based computing where each task is divided into smaller tasks deployed on computer networks.
- Virtualization allows for extensive scalability, giving clients virtually limitless resources.
- Application virtualization helps a user to have remote access of an application from a server.
- Network virtualization provides a facility to create and provision virtual networks—logical switches, routers, firewalls, load balancer,
- Peer-to-peer deployments are relatively inexpensive and simple to set up and manage.
- A Hypervisor or VMM (virtual machine monitor) is a layer that exists between the operating system and hardware.
- The guest operating system is modified and recompiled before installation into the virtual machine in case of Para Virtualization.
- Full Virtualization uses binary translation and direct approach as a technique for operations.
- Paravirtualization uses hypercalls at compile time for operations.

1.8 FURTHER READING

1. Here is an article about the difference between cloud and grid computing: https://techdifferences.com/difference-between-cloud-computing-and-grid-computing.html

1.9 ASSIGNMENTS

1) What is virtualization and what are its benefits?
2) What do you mean by data virtualization?
3) What are the characteristics of virtualized environments?
4) State the difference between client server and peer to peer architectures.

5) Compare and contrast grid versus cloud computing.
6) What do you mean by Full and Para Virtualization?
7) Discuss the heterogeneity issue in grid computing.
8) Explain the Bitcoin - P2P cryptocurrency without a central monetary authority.

1.10 ACTIVITIES

Q1. Multiple Choice Questions:

1) Which of the following type of virtualization is also characteristic of cloud computing?
   a) Storage
   b) Application
   c) CPU
   d) All of the mentioned
2) Point out the correct statement.
   a) A client can request access to a cloud service from any location
   b) A cloud has multiple application instances and directs requests to an instance based on conditions
   c) Computers can be partitioned into a set of virtual machines with each machine being assigned a workload
   d) All of the mentioned
3) Which of the following has many features of that is now known as cloud computing?
   a. Web Service
   b. Softwares
   c. Internet
   d. All of the mentioned
4) Which one of the following refers to the user's part of the Cloud Computing system?
   a. back end
   b. management
   c. infrastructure
   d. Front end
5) Which one of the following is not a peer to peer (P2P) architecture?

a. Bit Torrent
b. SQL Server
c. Network Monitoring System
d. None of these

1.11 CASE STUDIES

1. Case Studies (Optional)

-Prepare documentation of various hypervisors for virtualization.

-Discuss the working of Bit Torrent File sharing architecture.

Unit 3: Cloud Computing in a nutshell

3

Unit Structure

1.1. Learning Objectives

1.2. System models

1.3. Cloud computing Layers

1.4. Desired features of cloud computing

1.5. Basic principles of cloud computing

1.6. Risks and Challenges

1.7. Let us sum up

1.8. Check your Progress: Possible Answers

1.9. Further Reading

1.8. Assignment

1.1 LEARNING OBJECTI ES

After studying this unit, student should be able to:

- understand the basics of system models for distributed systems and cloud
- understand the basic principles and desired features of cloud computing
- assess risks and challenges involved in adopting cloud computing

1.2 SYSTEM MODELS

Cloud computing system is built over a large number of autonomous computer nodes that are interconnected by LANs or WANs in a hierarchical manner. Cloud computing is a massive system which is considered as a highly scalable, and can reach web scale connectivity, either physically or logically. A cloud is a pool of virtualized computer resources. Cloud computing applies a virtualized platform with elastic resources on demand by provisioning hardware, software, and data sets dynamically.

Cloud computing can deliver any IT service through the network, most likely the Internet. These services can be consumed as a utility and are famous as utility computing concepts. These include different aspects like infrastructure for computing, platform development, application, and services. These all aspects form the system model which can be viewed as a layered stack from the hardware appliances to software systems. The layered stack for cloud computing comprises three different layers which are known as Software as a Service (SaaS), Platform as a Service (PaaS), and Infrastructure as a Service (IaaS).

1.3 CLOUD COMPUTING LAYERS

The layered stack for cloud computing comprises three different layers which are known as Software as a Service (SaaS), Platform as a Service (PaaS), and Infrastructure as a Service (IaaS).

SaaS	Cloud Applications: Social Computing, Scientific Computing, CDNs etc.
PaaS	Cloud Programming: Web 2.0, Libraries, Scripting, Mashups etc.
IaaS	Cloud Hosting: Admission Control, Pricing, SLA Management etc.
	Virtual Machine Deployment and Management
	Cloud Infrastructure and Resources

Table 1: Cloud-Computing Layers

Table 1 shows the cloud computing layered view covering the entire stack. At the bottom, the physical infrastructure and resources (like memory, processor, speed, network bandwidth, etc.) are managed by the middleware whose sole job is to provide the necessary runtime environment for applications. Amazon Web Services (AWS) is an example of IAAS. AWS provides EC2 for computing, S3 buckets for storage, etc. The hypervisor is responsible for managing the pool of resources and providing that infrastructure as a collection of virtual machines. The virtualization of CPU, memory, and device is referred to as hardware virtualization.

I. Infrastructure as a Service (IaaS):

IaaS refers to the way of availing the infrastructure having computational capabilities on demand. It offers a standardized way of acquiring the same. Infrastructure providers enable the low-level abstractions of all the physical devices. Such resources include storage facilities, processing power, networks, and virtualized servers. The Pay-as-you-go model is applicable for infrastructure utilization based on the consumption such as the amount of storage, processing power consumed over a certain duration. In this service model, customers do not need to manage infrastructure, it is up to the provider to guarantee the contracted amount of resources and availability.The user can deploy and run on multiple VMs running guest OSes on specific applications. The user does not manage or control the underlying cloud infrastructure, but can specify when to request and release the needed resources.

II. Platform as a Service (PaaS):

IaaS solutions are used for designing the system architecture but provide limited services to build the applications. Such services can be availed by cloud programming environments and tools, resulting in a new layer. This layer provides

application development platforms to the users. Users can design and develop their applications for the cloud by using appropriate APIs. This cloud programming environment is referred to as Platform as a Service. PaaS includes middleware, databases, development tools, and some runtime support such as Web 2.0 and Java. The platform includes both hardware and software integrated with specific programming interfaces. The provider supplies the API and software tools (e.g., Java, Python, Web 2.0, .NET). The user is freed from managing the cloud infrastructure. One good example of PAAS is Google App Engineer. These are the environments in which developers can develop sophisticated software with ease.

III. Software as a Service (SAAS): SAAS provider offers an actual working software application to clients. Salesforce and Github are two good examples of SAAS. They hide the underlying details of the software and just provide an interface to work on the system. Behind the scenes, the version of Software can be easily changed. The SaaS model applies to business processes, industry applications, consumer relationship management (CRM), enterprise resources planning (ERP), human resources (HR), and collaborative applications. On the customer side, there is no upfront investment in servers or software licensing. On the provider side, costs are rather low, compared with conventional hosting of user applications.

1.4 BASIC PRINCIPLES OF CLOUD COMPUTING

Following are the basic principles of cloud computing.

1) Enablement: Plan for cloud computing as a strategic enabler, rather than as an outsourcing arrangement or technical platform.

2) Cost/benefit: Evaluate the benefits of cloud acquisition based on a full understanding of the costs of cloud compared with the costs of other technology platform business solutions.

3) Enterprise risk: Take an enterprise risk management perspective to manage the adoption and use of the cloud.

4) Capability: Integrate the full extent of capabilities that cloud providers offer with internal resources to provide comprehensive technical support and delivery solutions.

5) Accountability: Manage accountabilities by clearly defining internal and provider responsibilities.

6) Trust: Make trust an essential part of cloud solutions, building trust into all business processes that depend on cloud computing.

1.5 DESIRED FEATURES OF CLOUD

On-demand self-service: A consumer can unilaterally provision computing capabilities, such as server time and network storage, as needed automatically without requiring human interaction with each service provider.

Broad network access: Capabilities are available over the network and accessed through standard mechanisms that promote use by heterogeneous thin or thick client platforms.

Resource pooling: The provider's computing resources are pooled to serve multiple consumers using a multi-tenant model, with different physical and virtual resources dynamically assigned and reassigned according to consumer demand. There is a sense of location independence in that the customer generally has no control or knowledge over the exact location of the provided resources but may be able to specify location at a higher level of abstraction (e.g., country, state or datacenter). Examples of resources include storage, processing, memory, and network bandwidth.

Rapid elasticity: Capabilities can be elastically provisioned and released, in some cases automatically, to scale rapidly outward and inward commensurate with demand. To the consumer, the capabilities available for provisioning often appear to be unlimited and can be appropriated in any quantity at any time.

Measured service: Cloud systems automatically control and optimize resource use by leveraging a metering capability at some level of abstraction appropriate to the type of service (e.g., storage, processing, bandwidth, and active user accounts). Resource usage can be monitored, controlled, and reported, providing transparency for the provider and consumer.

1.6 CHALLENGES AND RISKS IN CLOUD COMPUTING

Cloud Migration: Cloud migration is the process of moving data, applications, and other important information of an organization from its on-premises either desktops or servers to the cloud infrastructure, and this can also involve moving data between different cloud setups. Cloud migration enables all the computing capabilities that were performed earlier by devices installed on-premises. Cloud migration is a big challenge as many companies when they require migrating from on-premises to cloud or from one cloud to another, partnering with experienced cloud service providers.

Incompatibility: During moving workloads from on-premises to the cloud, the common issue is the incompatibility between on-premises infrastructure and the services which companies are going to buy from the public cloud providers. In recent years, most cloud service providers (CSPs) tried to create “connectors of sort” to make practices more standardized and homogenous.

Data Security: CSPs are responsible to provide clouds’ security, but they’re not responsible for your apps, servers, and security of data. As per CDW 2013 State of the Cloud Report, “46 percent of respondents face security of data or applications as a significant challenge.” When your CSP ensures you about the complete compliance and regulation, don’t consider it as 100% compliant and yielding. You still require encrypting and securing your data and should invest in buying a suite of tools from your CSP to protect your data from cyber-attacks. The problem with cloud computing is that the user cannot view where their data is being processed or stored. And if it is not handled correctly during cloud management or implementation, risks can happen such as data theft, leaks, breaches, compromised credentials, hacked APIs, authentication breaches, account hijacking, etc.

Lack of Expertise: With the quick advancements and improvements in cloud technologies, more and more organizations are using clouds to place their workloads. However, they face difficulties to keep up with the tools which **require** particular expertise. Organizations can deal with this challenge by providing cloud technologies training to their system administrators along with development staff.

Downtime: Businesses suppose complete data accessibility and availability when their data is stored on the cloud anytime from anywhere. The main challenge most

organizations face is they can access their data from the cloud only through an internet connection. So, poor internet connection can disrupt cloud services and higher risks of data accessibility. If you have a consistent and high-speed internet connection, you can make the most of their cloud services. But if you don't, you may face repeated downtimes, lags, and errors. It not only frustrates the users but also reduces their productivity. All these can invite bottlenecks for business operations and lead to reduced sales, revenue, and profit margins.

Bandwidth Cost: Though organizations and businesses can save money on hardware using the cloud, they have to pay extra for the bandwidth they use to access their workloads. However, it doesn't charge much for smaller apps, but data-intensive apps need more bandwidth which can cost higher.

Compliance Risks: Compliance rules are getting more stringent due to the increased cyber-attacks and data privacy issues. Regulatory bodies like HIPAA, GDPR, etc., ensure organizations comply with applicable state or federal rules and regulations to maintain data security and privacy for their business and customers.However, compliance is another big challenge for organizations adopting the cloud. The issues arise for anyone using cloud storage or backup services. When organizations move their data from on-premises to the cloud, they must comply with the local laws. For example, every healthcare institution must comply with HIPAA in the US.

Improper Access Controls and Management: Improper or inadequate cloud access controls and management can lead to various risks for an organization. Cybercriminals leverage web apps, steal credentials, perform data breaches, and whatnot. They may face access management issues if they have a large or distributed workforce. As a result of poor access controls and management, organizations can be vulnerable to attacks. And their business information and user data can be exposed. Ultimately, it can cause reputation damage and increase unnecessary expenses.

Reduced Visibility and Control: Cloud computing offers the benefit of not having to manage the infrastructure and resources like servers to keep the systems working. Although it saves time, expenses, and effort, the users end up having reduced

control and visibility into their software, systems, applications, and computing assets. As a result, organizations find it challenging to verify how efficient the security systems are due to no access to the data and security tools on the cloud platform. They also can't implement incident response because they don't have complete control over their cloud-based assets. In addition, organizations can't have complete insight into their services, data, and users to identify abnormal patterns that can lead to a breach.

Insecure APIs: Using application interfaces APIs in cloud infrastructure enables you to implement better controls for your systems and applications. They are either in-built into the mobile apps or web to allow the employees and users to access the systems.However, if the external APIs you use are insecure, it can invite a lot of trouble for you in terms of security. These issues can provide an entry point for attackers to hack into your confidential data, manipulate services, and do other harm.Insecure APIs can cause broken authentication, security misconfigurations, break function-level authorization, expose data, and mismanagement of resources and assets.

1.7 LET'S SUM UP

Points to ponder

- Cloud computing applies a virtualized platform with elastic resources on demand by provisioning hardware, software, and datasets dynamically.
- IAAS providers give low-level abstractions of physical devices.
- SAAS provider offers an actual working software application to clients.
- PaaSmodel enables the user to deploy user-built applications onto a virtualized cloud platform.
- PaaS includes middleware, databases, development tools, and some runtime support such as Web 2.0 and Java.

1.8 CHECK YOUR PROGRESS: POSSIBLE ANSWERS

Q1. Multiple Choice Questions:

1. Which of the following benefits is related to creating resources that are pooled together in a system that supports multi-tenant usage?
   a) On-demand self-service
   b) Broad network access
   c) Resource pooling
   d) All of the mentioned
2. Which of the following is the most important area of concern in cloud computing?
   a) Security
   b) Storage
   c) Scalability
   d) All of the mentioned
3. Which of the following is the best known service model?
   a) SaaS
   b) IaaS
   c) PaaS
   d) All of the mentioned
4. What is the biggest concern regarding cloud computing?
   a) Accessibility
   b) Privacy
   c) Security
   d) Both b and c
5. What aspect of cloud computing is responsible for preventing downtime and determining costs?
   a) Bandwidth cost
   b) Committing to service level agreements or SLAs
   c) Application programming interfaces
   d) Virtual private network or VPN.

6. Which organization promotes cloud computing security standards?
   a) Cloud Security Watchdog
   b) Cloud Security Alliance
   c) Cloud Security Standards Working Group

d) None of the above.

7. Which of these is not a security risk of cloud computing?
   a) Reduced visibility and control over data.
   b) Regulatory non-compliance by the cloud service provider or CSP.
   c) Data breaches by malicious co-tenants
   d) Disaster Recovery.
8. Which of the following service provider provides the least amount of built in security?
   e) SaaS
   f) PaaS
   g) IaaS
   h) All of the mentioned

1.9FURTHER READINGS

1. https://binaryterms.com/principles-of-cloud-computing.html
2. https://geekflare.com/challenges-and-risks-in-cloud-computing/
3. https://www.geeksforgeeks.org/types-of-server-virtualization-in-computer-network/

1.10 ASSIGNMENT

1) Discuss in detail, various other desirable features of cloud computing.

2) Explain how down time violates the service level agreement.

3) What do you mean by virtualization of computing resources?

4) Discuss the issues and challenges involved in IaaS.

5) Discuss the various risks involved in adoption of cloud computing?

6) What do you mean by elasticity and resource pooling in cloud computing?

1.11 CASE STUDIES

1. Case Studies (Optional)

-Explore various Amazon Web Services in Context of IaaS, PaaS, SaaS in detail.

-Assess the various risks involved in adoption of public cloud for your organization.

4

Unit 4: Cloud Types and Models

Unit Structure

1.1. Learning Objectives

1.2. Cloud Types

1.3. Private Cloud and its Components

1.4. Implementation of Private Cloud

1.5. Let’s Sum Up

1.6. Check Your Progress: Possible Answers

1.7. Further Reading

1.8. Assignment

1.9. Case Study

1.1 LEARNING OBJECTIVE

After studying this unit, student should be able to:

- understand the different types of cloud models
- understand the various components involved in private cloud implementation along with their pros and cons

1.2 CLOUD TYPES:

There are basically three types of cloud computing viz., public clouds, private clouds, and hybrid clouds. Apart from that, there exists community cloud computing in which distributed systems is created by integrating the services of different clouds to address the specific needs of an industry, a community, or a business sector. It is owned, managed, and operated by one or more organizations in the community, a third party, or a combination of them. In the community cloud, the infrastructure is shared between organizations that have shared concerns or tasks. The cloud may be managed by an organization or a third party. Sectors that use community clouds are media industry, healthcare industry, energy & core industry, scientific research etc.

Cloud computing types refer to the service deployment models that allow choosing the level of control over your information and types of services you need to provide. The cloud computing services include Infrastructure-as-a-Service (IaaS), Platforms-as-a-Service (PaaS), and Software-as-a-Service (SaaS). Choosing a cloud type or cloud service is a unique decision. Cloud type enables cloud computing i.e., the act of running workloads within that computing system. Almost every cloud includes an operating system, programming interfaces for various applications, management platforms etc., which allows sharing of scalable computing resources.

Public Cloud:

A public cloud offers infrastructure and services off-site. As it name signifies, it is open publicly for storing and accessing information over the internet based on the Pay-Per-Use model. The fundamental characteristics of public clouds are multitenancy which refers to serving multiple users instead of a single customer.

These basic compute resources are coupled with managed services like database servers, applications and security systems. The managed services are there for you to rent if you don't want the hassle of setting up and managing the whole solution. The computing resources are operated and managed by the Cloud Service Provider (CSP). It reduces the capital expenditure as all the hardware resources are available within the cloud and managed by CSP or End-users. The public cloud offers the greatest level of efficiency for the sharing of resources and service provisioning. This kind of public cloud deployment is frequently used to provide storage, testing, application hosting, and development environments. The following are the giant examples of the public cloud: Amazon elastic compute cloud (EC2), Microsoft, Google Cloud Platform, IBM SmartCloud Enterprise, Windows Azure Services Platform, Gmail, U of I Box, etc.

Advantages of Public Cloud

There are the following advantages of Public Cloud -

- Customers have no responsibility for buying or maintaining the physical components that make up their public cloud IT solutions.
- Easy to Integrate, so offers a highly flexible approach to end users/consumers.
- It is maintained by the cloud service provider.
- Public cloud is owned at a lower cost than the private and hybrid cloud.
- Location independent as its services are accessedvia the internet.
- Public cloud is possessed nearlyunlimited scalable as per the requirement of computing resources.
- It is highly reliable due to vast network of servers which ensures against failure

Disadvantages of Public Cloud

- Less security as all the resources are shared publicly.
- Internet bandwidth is a bottleneck as performance depends upon the high-speed internet network.
- The Client has no control of data i.e., lesser autonomy over servers.

- Unique business requirements may not be guaranteed sometimes i.e., offers less customization.
- Public cloud platforms may not meet industry standards, government policies or any legal requirements.
- Sudden changes by cloud provider can have dire impacts

Private Cloud:

A private cloud offers infrastructure and services on a private network. As it name signifies, they are restricted to private corporations, business, etc. Hence it is also known as corporate or internal cloud. The private clouds no longer have to be sourced from on-premise IT infrastructure as organizations are now building private clouds on rented, vendor-owned data centers located off-premises. Compared to the public cloud, it offers the greatest level of security and control. This kind of private cloud deployment does not require being online always. The services and infrastructure are always maintained on a private network and make it easier for an organization to customize its resources to meet specific IT requirements. Private cloud provides specific operations such as appropriate clustering, data replication, system monitoring, and maintenance, and disaster recovery, and other uptime services. Private clouds are often used by government bodies, financial organizations, mid to large size corporate with business-critical seeking enhanced control over their environment. The various open-source tools like Open Stack, Open Nebula, and Eucalyptus can be used for deploying the private cloud platform. Some of the Private Cloud providers include IBM, VMware, HPE, Oracle, and Dell EMC.

Advantages of Private Cloud

There are the following advantages of the Private Cloud -

- It provides a high level of security and privacy to the users.
- It provides better autonomy over the servers i.e. complete control over the infrastructure, you can tailor your cloud computing approach to your own preferences and internal processes.
- It can meet strict security, compliance, or legal requirements.
- It enables the quick allocation and deliveryof on-demand IT resources.
- No risk of sudden changes that can disrupt company operations.

Disadvantages of Private Cloud

- Requirement of expertise IT personnel to manage and leverage private cloud services.
- Private cloud is accessible within the organization which limits the operational area.
- High investment hurdle in the private cloud deployment and its maintenance.
- New operational processes are required
- Private cloud is less scalable.

Hybrid Cloud:

Hybrid cloud refers to the single IT environment from multiple environments. Hybrid cloud solutions are a blend of public and private cloud environments connected through local area networks (LANs), wide area networks (WANs), virtual private networks (VPNs), etc. It is highly recommended when 1) some data cannot be put in the cloud due to legal reasons; 2) organizations want to keep old hardware and system running locally for applications that cannot be updated. For example, the application can be deployed on the public cloud due to scalability requirements but the database can be stored on a private cloud for achieving a high level of security. The following are examples of the hybrid cloud platforms: Google Application Suite (Gmail, Google Apps, and Google Drive), Office 365 (MS Office on the Web and One Drive), etc.

Advantages of Hybrid Cloud

There are the following advantages of Hybrid Cloud -

- It leverages more flexibility and scalability compared to on premises deployment.
- It reduces time to market for delivering new products and services more quickly.
- It offers operational flexibility i.e., executing mission critical on private cloud, testing of public cloud.

- It offers flexible resources because of the public cloud and secure resources because of the private cloud.

Disadvantages of Hybrid Cloud

- Security feature is compromised compared to the private cloud.
- Reliability of the service depends on the CSP.
- Managing a hybrid cloud is complex due to heterogeneity of different models.

1.3 PRIVATE CLOUD AND ITS COMPONENTS

1) Private clouds are distributed systems that consist of cloud computing resources used exclusively by one business or organization. It makes it easier for the organization for the tailor-made resources requirements to meet specific IT goals. Private cloud implementation also includes the challenges of integrating hardware and software from several vendors. It is important to understand the key components and their expectations of the private cloud that are in demand. The components are referred to by the generic term “Cloud Infrastructure” which includes abstracted resources, hardware, storage, network resources, etc. The basic elements of cloud infrastructure are the same whether you have a private cloud, public cloud, or a combination. The components can be viewed as tools to build the cloud for hosting the services and applications. The following are the components of the cloud infrastructure.
2) Hardware: It includes networking devices like router, switches, load balancers, firewall, storage arrays, servers and backup devices.
3) Virtualization: It refers to the technology that separates IT services and functionality from the hardware. Hypervisor, a piece of software sits on the top of physical hardware and abstracts the machine resources such as computing, storage and memory.
4) Storage: Storage management is responsible for ensuring correctness of data being backed up, outdated data removal, and reliability in case of server failure in data center. Storage virtualization abstracts storage space from hardware systems and makes it available for users as cloud storage.

5) Network: It consists of physical switches, routers, cables, and other equipment and virtual networks are created on top of these network resources. Cloud allows the creation of VLANs (Virtual Local Area Networks) and these network resources are assigned static or dynamic IP addresses.

1.4 IMPLEMENTATION OF PRIVATE CLOUD

It is important to understand the key building blocks and technologies behind the implementation of private clouds as they are the ideal solution for IT leaders who want to make enterprise resources available on-demand, but can't move to the public cloud. The private cloud depends on various technologies in which understanding how virtualization works is the prime focus of how the cloud works.
The following are the key component which acts as building blocks for the implementation of the private cloud.

- Virtualization:
- Service catalog
- Self-service portal
- Support for heterogeneous systems
- Resource and workload manager
- Metering software and pay per use billing

Virtualization in the private cloud allows combining resources from physical hardware into shared pools. The addition of a management software layer leverages administrative control over the applications, platforms, and infrastructure. The final automation layer if added to replace or reduce human interaction with repeatable instructions and processes, the self-service component of the cloud is complete and that bundle of technologies is now a private cloud. An operating system like Linux is the prime need of all clouds but on the other hand, infrastructure can include a variety of bare metal, virtualization, and containers for abstracting pooling, and resource scaling across the network. Cloud infrastructure like OpenStack can help to implement our private cloud using our dedicated resources.

The following are the basic steps and important considerations to deploy a private cloud on your own hardware:

1) Hardware Setup: Select optimal hardware setup based upon the need of your application or project demands so as to provide sufficient resources to your VMs.
2) Installing management software: There is the software used for management and virtualization, sometimes called the ‘stack’, because it is in fact a host of dozens or hundreds of different programs working together as one. Some of the bigger names in management are **OpenStack, VMware vSphere, and CloudStack**. Each of these management consoles can control a variety of hypervisors, including **ESXi, Xen, and KVM**.
3) Configuring management software controller: It acts as an interface for creating and managing VMs. This can be on a completely separate server for redundancy, or on a VM running inside your private cloud.
4) Creation of VM templates and deployment of server images.
5) Create, License and Test VMs: Finally, VMs can be created (and licensed where necessary), and troubleshooting and testing of the platform can begin.

1.5 LET’S SUM UP

Points to ponder

- Private cloud deployment does not require being online always.
- The public cloud offers the greatest level of efficiency for the sharing of resources and service provisioning.
- Public cloud is highly scalable as per the requirement of computing resources.
- Private cloud provides a high level of security and privacy to the users.
- A hybrid cloud is a heterogeneous distributed system formed by combining facilities of public cloud and private cloud.
- Virtualization in private cloud allows combining resources from physical hardware in to shared pools.
- Private clouds are distributed systems that consist of cloud computing resources

used exclusively by one business or organization

- The community cloud allows systems and services to be accessible by a group of organizations.
- Community cloud is suitable for organizations that want to have a collaborative cloud with more security features than the public cloud.
- PaaS provides the runtime environment for applications, development and deployment tools, etc.

1.6 CHECK YOUR PROGRESS: POSSIBLE ANSWERS

Q1. Multiple Choice Questions

1) Which one of the following is Cloud Platform by Amazon?

a. Azure
b. AWS
c. Cloudera
d. All of the mentioned

2) Which one of the following is related to the services provided by Cloud?

a. Sourcing
b. Ownership
c. Reliability
d. PaaS

3) Which one of the following can be considered as the example of the Front-end?

a. Web Browser
b. Google Compute Engine
c. Cisco Metapod
d. Amazon Web Services

4) The Foce.com and windows Azure are examples of which of the following?

a. IaaS
b. PaaS
c. SaaS

d. Both A and B

5) Which one of the following a technology works behind the cloud computing platform?

a. Virtualization
b. SOA
c. Grid Computing
d. All of the above

6) Which one of the following is a kind of technique that allows sharing the single physical instance of an application or the resources among multiple organizations/customers?

a. Virtualization
b. Service-Oriented Architecture
c. Grid Computing
d. Utility Computing

7) Which one of the following is a type of software that supports the virtual machine?

a. Kernel
b. Hypervisor
c. VMM
d. Both B and C

8) The _____ allows systems and services to be accessible by a group of organizations.

a. Private cloud
b. Public cloud
c. Community cloud
d. Hybrid cloud

9) Which cloud allows systems and services to be accessible within an organization?

a. Public cloud
b. Private cloud

c. Community cloud

d. Hybrid cloud

10) Which one of the following is an example of Software as a Service (Saas)?

a. Google Docs

b. DropBox

c. Salesforce

d. all of the above

1.7 FURTHER READING

1. Here is an article about the public, private and hybrid cloud:

Block-2
Cloud Computing Services

https://www.redhat.com/en/topics/cloud-computing/public-cloud-vs-private-cloud-and-hybrid-cloud

1.8 ASSIGNMENTS

1) What is meant by network and storage virtualization and what are its benefits?
2) Compare and contrast public, private, and hybrid cloud.
3) Discuss the various steps involved in private cloud setup.
4) Discuss the Pros and Cons of community cloud.
5) What do you mean by licensing in cloud computing?
6) What is the relevance of VM template creation in cloud setup?
7) Explain the role of public cloud in various sectors.

1.9 CASE STUDIES

1.8 Case Studies (Optional)

a. Prepare documentation on Private Cloud Implementation by **OpenStack**.

b. Prepare a write up about the network and hardware virtualization.

Unit 1:Infrastructure as a Service & Platform as a Service

1

Unit Structure

1.1. Learning Objectives

1.2. Cloud Service Models

1.3. Infrastructure as a Service

1.4. Platform as a Service

1.5. Case Studies

1.6. Let’s Sum up

1.1 LEARNING OBJECTIVE

This chapter discusses about,

- Service Models in Cloud Environment
- Services offered under IaaS and PaaS
- Pricing models for IaaS and PaaS related services
- Major Public Cloud Service Providers and their services
- Case Study of Amazon IaaS
- Case Study of Google App Engine

1.2CLOUD SERVICE MODELS

In the last chapter, we have discussed about different cloud deployment models like: private cloud, public cloud and hybrid cloud. These deployment models are classified based upon the location of hardware resources. The Cloud computing paradigm offers different resources in the form of services. In this unit, we will discuss about different service models: Infrastructure as a Service, Platform as a Service and Software as a Service etc. These services are related with each other. The figure below shows the layered view of these services.

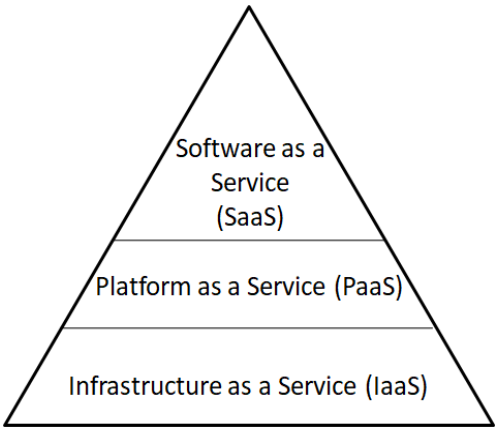


Figure-1 Layered view of Cloud Services

There are mainly two roles in cloud computing environment: Cloud Service Provider and Cloud Service User. The responsibilities of service providers and users depend upon the type of service. The figure 2 summarizes responsibilities. In case of in house on premise data centre each and every resources are managed by the user. If user moves to cloud based service model, some of the tasks are carried out

by service providers. The IaaS service provider manages all low level resources like network, physical servers, operating system etc. The users have to concentrate on their applications, run time environments, middleware and the data. If the user moves to PaaS, he or she has to just focus on application and data everything else is managed by PaaS service provider. The SaaS provides highest level of freedom in which, users just make use of services all other things in background are managed by SaaS provider. Some of the most common IaaS, PaaS and SaaS public service providers are listed in table 1

Type of the Service	Major Service Providers
IaaS	Amazon Web Services (AWS), Microsoft Azure, Google Compute Engine (GCE), DigitalOcean, Rackspace
PaaS	Google App Engine, Apache Stratos, AWS Elastic Beanstalk, Windows Azure, Heroku, Force.com, OpenShift
SaaS	Google Apps, Dropbox, Salesforce, Cisco WebEx, Concur, GoToMeeting

Table-1 Major Public Service Providers

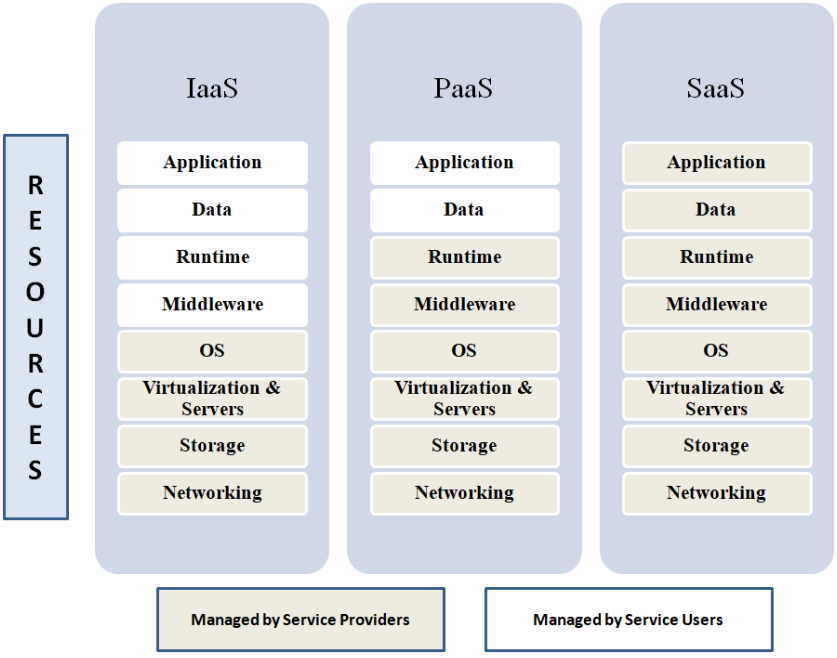


Figure-2 Responsibilities of service providers and users

1.3 INFRASTRUCTURE AS A SERVICE (IaaS)

Traditionally, companies spend huge amount to own and manage the IT related hardware infrastructure (physical resources). The infrastructure is maintained as “On premise & In-house” data centre. Instead of owning the IT resources, the cloud computing paradigm offers access to IT infrastructure in the form of services i.e. less cost of ownership. These services are called Infrastructure as a Service (IaaS). Various services offered via IaaS are:

- Storage Service
- Compute Service
- Network Service
- Cloud Management Service

➢ **Characteristics of IaaS**

- Access through rich user interface

  IaaS providers offer different interfaces to access their services. Some of the most common interfaces are Web Interface, Programming API and Command Line Interface (CLI) etc.

- Pay as you Go model

  There are various pricing models for accessing IaaS. One of the most common models is Pay as you Go. In this model, users are charged based on the amount to time the service is used and the type of service.

- Self service based on demand provisioning

  Users of IaaS can choose their desired service from rich set of available offerings. Cloud service providers offer different services via online platforms.

- Higher availability

  Infrastructure related services are offered over the Internet. Users can access their infrastructure from any part of the world via any device. This feature makes services highly available.

- Highly Reliable

  IaaS providers have huge amount of hardware resources pooled in their data centers across different locations. In case of any failure of any service, they

switch over user services on some other physical hardware in a very amount of time. This kind of feature makes IaaS very reliable.

- Scalable

  Scalability is one of the inhering characteristics of cloud computing. Generally, cloud users are not aware about the workload. They may hire infrastructure related resources by considering some initial workload. In case of higher workload, this may lead to the performance degradation. IaaS providers offer the feature of auto scaling. This enables the users to scale up or down (increase or decrease) their hired resources.

➢ **When to use IaaS?**

- When the demand of computing resources is volatile.
- When new organizations want to reduce the capital cost incurred on the purchase of hardware resources.
- When specific, trial or temporary access of hardware resources is needed.

➢ **When not to use IaaS?**

- When we want the higher level of performance. The on-premise or dedicated hosted infrastructure gives better performance than IaaS.
- When the regulatory compliance restricts the off-shoring or outsourcing of the data storage and processing.

➢ **Different services offered by IaaS Providers**

1. Storage Services

It allows users to store data on the data storage devices owned by cloud service providers. The data stored on cloud can be accessed at any time and from any place. In background, the storage service providers manage several distributed storage devices. They also maintain redundancy and consistency of data to provide reliable services. These internal complexities of storage service providers are kept hidden from users.

Some of the issues with storage service are: Dependency of Internet, Network Latency and data security. The pricing of storage service depends upon various factors like:

- Storage capacity
- Amount of data transferred to cloud
- Amount of data retrieved from cloud
- Nature of reliability and security

2. Compute Services

Generally, the IaaS providers offer compute services in the form of virtual machine instances. This service refers to the provision of various computing resources like, RAM, CPU and I/O. IaaS providers offer different configurations of compute resources in form of many instances types. Users can easily choose the type of instance from set of available options.

The price of particular instance (service) depends upon factors like: Amount of resources, Type of hardware resources, Type and version of operating system, Mode and duration of service etc. Some of the most popular pricing models are,

- On-Demand

  This type of service offers flexibility to users such that they can occupy desired compute service at any time. The users are charged on hourly basis. Users can occupy and free resources on the fly without any long term commitments. This mode is suitable for the users, who do not know their compute related requirements in advance.

- Prepaid

  It is also called the reservation based model. In this, users have to pay in advance for the specific time during which they want to access the service. IaaS services under this mode are cheaper than the on-demand mode because, service providers know the resource requirements in advance and they can do better management of their infrastructure. This mode is suitable for the users who know their compute related requirements in advance.

- Auctioned

  This model allows users to bid for required resources. The bid price is controlled by service providers on the basis of supply & demand of

resources. This mode helps service providers to improve utilization of their unutilized infrastructure. Similarly, this model is useful to users because they can get desired compute resources at cheaper rate. This mode is suitable when the resource requirement is not immediate. This model is useful for non-production environments.

3. Network Related Services

IaaS providers also offer network related services. Two most widely used IaaS network services are: Load Balancing and DNS. The Load balancing service provides a single access point to several servers owned by a particular organization. The load balancer distributes requests among several servers. Some of the most common load balancing approaches are: simple round robin, weighted round robin, dynamic round robin, least connections, fastest server etc.

DNS is hierarchical system used for resolution of names and IP addresses. A cloud provider's DNS services offer to store IP/domain-name mapping for both internal and external servers. Some of the main features under DNS service are listed below.

- Store IP/domain-name mapping for both internal and external servers. DNS management API for developers is used to automate DNS related operations such as authentication and domain-zone creation
- Extended security and access control
- Geo location-based DNS routing

4. Cloud Management Services

Cloud Management Service providers help the customers to simplify Cloud management related operations. Cloud management services (CMS) help to fill gaps between cloud management tools. CMS providers work directly with specific cloud service providers to integrate their services with a customer's existing cloud, or to create extra value added services on top of the provider's existing infrastructure.

- Cloud Value Added Management Services

The Value Added Management Services enhance existing cloud provider services by building additional features on top of it. The value added management service help customers to save time and resources by automating regular tasks which are manually performed.

- Cloud Integration Service (CIS)

  CIS provides a ways to have integration between private and public clouds. CIS offers an abstraction layer which enables transparent multi-cloud support. A CIS can automate administrative tasks like: provisioning, scaling, monitoring and configuration. CIS provides multi-cloud governance tools to allow customers to monitor logs, security events, and resource utilization from a single dashboard.

- Cloud Service Brokerages (CSB)

  The CSB aggregate the services offered by different cloud providers. It organizes these services into a easily-searchable service catalogue. CSB services can be considered as a marketplace for cloud services from where, customers can quickly identify and purchase required services.

1.4 PLATFORM AS A SERVICE (PaaS)

Platform as a Service (PaaS) is a cloud service model which provides a configurable application platform. PaaS can be considered as an abstraction layer above the hardware, operating system, and virtualization stack.

The PaaS model helps companies to reduce complexity of infrastructure and application maintenance. It allows companies to focus on core software development process and improve productivity. The figure 3 gives overview of services offered under PaaS model.

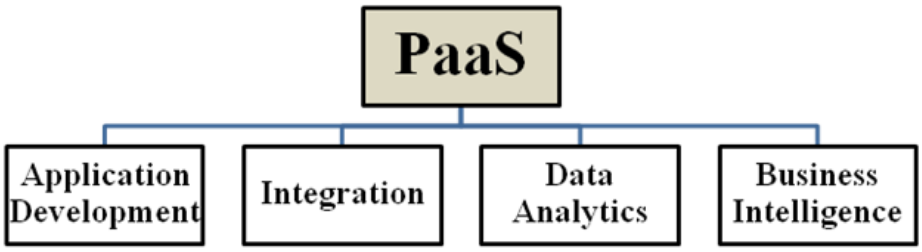


Figure-3 Services offered under PaaS Model

- **Benefits of PaaS**

PaaS provides means to simplify tasks like,

- Write, build and deployment tools for rapid application development and deployment
- Application integration with other components such as web services and databases
- Multi-tenancy, platform service that can be used by many concurrent users
- Easy Management of API
- Logging, reporting, and code instrumentation
- Security, Scalability, redundancy and auto-provisioning infrastructure to applications

- **Services offered under PaaS Model**

1. **Integration Services**

The Integration service helps companies to integrate data from multiple sources. Data may be available from different sources like: on-premise private cloud databases, public cloud databases, vendor databases, CRM systems, Messaging systems etc. The integration task becomes very challenging because of data quality, compatibility and standards.

Dell Boomi is a leader in cloud data integration. It integrates over 40 different technologies and protocols like Oracle, SAP, Siebel, Salesforce, Hadoop, QuickBooks etc. The data integration can be achieved either by writing our custom code or by using integration service.

The data integration done via custom written code approach has several drawbacks like:

- Large internal resource requirements
- Frequent changes are not supported
- Lack of business agility
- Higher maintenance cost

- Manual backups

- Custom security

Against the custom written code based approach, the integration service has several advantages like:

- Small requirements of IT infrastructure

- Easy to use

- Cost effectiveness

- Automatic update

- Built-in security

- Supports large variety of technologies and standards

2. Application Development and Deployment Services

PaaS application development and QA services help companies to improve the code quality. It also helps to accelerate software development process and better collaboration among development team.

In today's world, development teams prefer to use agile software development model instead of conventional waterfall model. The agile development method needs higher level of team collaboration and an intense continuous delivery cycle with frequent code releases, tests, and deployments. The development teams are geographically distributed across different locations. In all these situations, companies can benefit from PaaS tools that simplify and streamline team collaboration and create a faster time-to-market cycle.

PaaS development platforms like Google AppEngine and Amazon Beanstalk take a continuous delivery approach (agile development approach) to the next level by giving more support for developers through automating build, test, release, deployment, and operation processes.

In PaaS environment, users don't need to be an expert of operating system or networking tasks to manage the operations related to infrastructure. The PaaS based cloud platform performs all these tasks. In addition to routine

infrastructure operations, PaaS also does auto scaling of infrastructure based on certain application performance conditions.

3. **Data Analysis and Business Intelligence Services**

These services help companies to simplify data analysis related operations and apply statistical analysis methods to support business decision making and information retrieval. The need for data analysts and statisticians is very high. It is very expensive for companies to hire data scientists, and it also takes a long time to develop internal expertise. PaaS platforms cannot completely solve this issue, but they can certainly hide some complexity and simplify data analysis related tasks. These services are more suitable for companies which satisfies bellow conditions,

a. Data are already hosted in the cloud
b. Limitations of service are acceptable
c. Data analysis related workload is not predictable and in this case will benefit from cloud elasticity
d. Don’t have sufficient financial resources to build an internal data analysis platform
e. SLA requirements are satisfied by service provider.

Data analysis and statistical modeling related tasks require massive computational capacity which can be easily hired by IaaS model. This idea makes data analysis task cost effective.

Some of the issues with this kind of service are as below.

1. Less Control

   Companies do not have full control on public cloud data related operations such as storage, backup, network transfer and security (access control & encryption).

2. Limited Customization

Public cloud services have many customers. These services cannot fully satisfy the need of each and every customer. Typically, service providers try to offer “good-enough” solutions and sets of features that are usable by a majority of customers. If additional features are required, there is no guarantee that the provider can add them.

3. Lack of Maturity

   Many public cloud services under this category are offered and built by relatively new companies. Sometimes, these solutions may not provide the same rich set of features found in on-premise solutions like Oracle, Teradata, or SAP.

1.5 CASE STUDIES

- **Case Study 1**
  Amazon (IaaS Platform)
- **Case Study 2**
  Google App Engine (PaaS Platform)

1.6 LET’S SUM UP

Points to ponder

- Cloud service models
- Characteristics of IaaS
- When to use and not to use IaaS
- Different IaaS services
- PaaS and advantages

Unit 2:Software as a Service and Database as a Service

2

Unit Structure

2.1. Learning Objectives

2.2. Software as a Service (SaaS)

2.3. Database as a Service (DBaaS)

2.4. Let’s Sum up

2.1 LEARNING OBJECTIVE

This chapter discusses about,

- Software as a Service Model
- Characteristics, features and benefits of SaaS
- Database as a Service Model
- Case Study about Software as a Service (SalesForce.com)
- Case Study about Database as a Service

2.2 SOFTWARE AS A SERVICE (SaaS)

Software as a Service (SaaS) is a cloud computing based services delivery model in which software are provided online and on-demand basis. Two major key events which increased the popularity of the SaaS based approach are:

1. The Internet became a commercial platform.
2. For companies it was costly, complex and difficult to install, run, upgrade and manage Software.

SaaS refers to the software which is deployed on a hosted service and is accessible over the Internet. Just like other cloud delivery models as described in previous chapter, the SaaS allows companies to reduce expenses on purchase and management of software application resources.

SaaS is one of the most widely used cloud delivery model. Almost every software vendor tries to offer software over SaaS model. Currently, there are SaaS offerings in every category of software products. Some of the most widely adopted categories of application software under SaaS model are:

- Billing and Inventory Systems
- Customer Relationship Management (CRM) software
- Help desk related applications
- HR management applications
- Security related software
- Social service related software

➢ **Desired Characteristics of SaaS software**

Some of the desired characteristics which make the SaaS software commercially feasible are listed as below.

- The SaaS application should be generalized so that customers will be interested in the application.
- The SaaS application should be modular and service oriented.
- An SaaS application should include measuring and monitoring features so customers can be charged based on actual usage
- SaaS applications should have sophisticated navigation mechanism to provide ease of use
- The SaaS application must have a built-in billing service.
- SaaS applications must ensure integrity, reliability and security of customer's data
- SaaS applications should constantly provide faster releases of new features and new capabilities. This must be provided without affecting existing customers.

➢ **Features of SaaS based application software**

Some of the important features of SaaS based application software are:

a. SaaS makes the software available over the Internet.
b. SaaS allows usage or subscription based licensing.
c. SaaS applications are cost effective since they do not require any maintenance at end user side.
d. SaaS based applications are available on demand and they can be scaled up or down on demand.
e. The Software is maintained by the vendor rather than where they are running.
f. They are automatically upgraded and updated.
g. All users are running same version of the software.

- **Benefits of SaaS based Approach**

A. Minimal Software Tools

The SaaS based approach needs minimal or no software installation at client site which gives benefits like:

- No complex software package installation at client side
- Very less configurations at client side
- Software distribution cost is less
- Software management and update task becomes easy

B. Centralized Management of Data

Generally, in SaaS based application software the data is stored on Cloud. Service providers take care about reliability, security and access control of data by features like replication and encryption.

C. Multitenant Solutions

Multiple users can share a single instance of data and application by virtual isolation. Users can customize their applications based on requirements without affecting basic functionality.

D. Efficient usage of software licence

Users can have single license for multiple computers running at different locations which reduces the cost of licensing. Users do not require license servers because the software runs in the provider's infrastructure

E. Reduced Cost

SaaS platforms use multi-tenant architectures in which the same hardware platform and software is shared among multiple customers. This helps the SaaS service providers to provide services at lower costs.

- **SaaS Delivery Models**

SaaS based software can be broadly classified and delivered into three categories. These categories are discussed in this section.

- **Packaged software services**

This is the biggest area in the SaaS market. This comes in many different options like Customer Relationship Management, Supply Chain Management

and Finance Management etc. Applications in this category focus on a particular business process such as, Employee Management Process which includes functions like managing employees' benefits, salaries, leaves, and annual performance reviews etc. Some of the Companies working in the packaged software market are following:

- Netsuite, like Salesforce.com, offers a CRM foundation. Netsuite provides several modules for enterprise resource planning (ERP) application to provide financial, e-commerce, and business intelligence services.
- RightNow offers a CRM suite of products for marketing, sales, and other industry solutions.
- Constant Contact is a marketing automation platform. They partners with CRM platforms like salesforce.com. They provide solutions to automate the process of sending emails and other marketing efforts.
- Intuit provides a Financial Services Suite of products to provide accounting related services for small- and medium-sized businesses.

- **Collaborative software services**

  In current scenario the teams are located at different locations across the world which needs to collaborate. This kind of collaboration is possible due to the ubiquitous availability of the Internet. Some of the most common software which need collaborative efforts are web conferencing, document collaboration, project planning, instant messaging etc. Some of the companies working under collaborative software based approach are:

  - MicrosoftLive is one of the first collaboration as a service platform. Microsoft offers live meetings and messaging services. They also provide email server as a service.
  - GoogleApps from Google supports various collaborative applications like e-mail, document management, and instant messaging. It also publishes APIs so third party software developers can integrate with the platform.
  - Zoho is an open-source collaboration platform to provide email, document management, project management, and invoice management related services. It also supports APIs to its environment to support collaborative development.

- Cisco Webex Collaboration platform offers SaaS based unified communications as a service.

- **Enabling and management services**

  Software in this category provides support for development and the deployment of SaaS applications. They provide development tools which are needed by developers to extend SaaS applications. They also provide facilities like testing, monitoring, and measuring of applications. Major services under this category of SaaS are as described below.

  a. Testing as a service

  Testing is one of the biggest uses for cloud computing. Development companies use this service to do different types of testing like functional testing, unit testing, stress testing, performance testing etc. Many vendors offer testing as a service including HP, IBM, Sogeti etc.

  b. Monitoring and management as a service

  Sometimes, companies using SaaS need to monitor their services to check whether their desired service levels have been satisfied by SaaS providers or not. In collaborative environments, companies may need to monitor service levels of different SaaS based applications.

  c. Development as a service

  This delivery model of development infrastructure is created through one of the Platform as a Service. Some vendors like Amazon offer support service for developers.

  d. Security as a service

  Many vendors, who provide antivirus software, offer their products as a service. Some of these vendors are Symantec, McAfee, CA, and Kapersky Labs etc. Identity management is an important aspect of on premise as well as cloud services.

  e. Compliance and governance as a service

  Compliance and governance related tasks are time consuming and complicated. This kind of services provide features like patch

management, business continuity planning, discovery of records and messages, governance requirements such as SOX (Sarbanes-Oxley) in the United States and SaS 70 (Statement of Audit Standard) controls for data.

- **Limitations of SaaS Applications**
  - Not suitable for real time systems

    SaaS based applications run in cloud infrastructure. These applications are accessed by users over the Internet. The network delay makes SaaS based applications less suitable for real time systems.

  - Internet Dependency

    The Internet connectivity is must to work with these applications. The availability and speed of Internet may limit the access of these applications.

  - Data security and compliance

    Customer data is stored on cloud. It may cause security and compliance issues.

  - Lack of Interoperability and integration and Vendor locking

    SaaS based applications lack the flexibility of interoperability and integrations. These features are only possible if supported by vendors. This may lead to the vendor locking situations.

  - Less customization

    SaaS based applications are designed and developed by considering the common requirements of customers. This kind of generalized design may not provide some specific requirements of particular customer. In this case, if customization support is available than only these features can be developed.

  - Limited range of applications

SaaS based approach is relatively new and in growing phase. All the desktop based applications which are currently in use do not have SaaS based option.

2.3 DATABASE AS A SERVICE (DBaaS)

Database as a Service (DBaaS) is an architectural approach that enables IT providers to deliver database related functionality as a service to one or more consumers. The DBaaS allows customers to host their database related infrastructure in cloud and get freedom from many low level database related operations. Customers can focus more on their actual application development work.

DBaaS providers host database in cloud infrastructure and they also manage all hardware and networking related tasks. They also offer extra services like scaling, failover, backup, restore etc. Currently, DBaaS related services are available for every modern relational database like MySql, Oracle etc.

- **Why DBaaS?**

A conventional approach for database needs to allocate hardware and software resources. This requires budget and time for managing and deploying database systems. These tasks are to be repeated for development, testing and production environments.

These issues of conventional database can be overcome by DBaaS. In the case of DBaaS, the cloud-based automation provides the centralized database management application which reduces DBA related overhead. All intense and dedicated tasks like configuration, optimization, provisioning, backup, security, monitoring, networking, automation, upgrades, maintenance, etc. are performed by the DBaaS service provider's automated database management system. The figure 4 shows the usefulness of DBaaS approach over DBA. DBaaS relieves users from,

- Costly hardware/software purchases
- Compatibility checks
- Dependencies resolutions
- Backup and Cron jobs

- Hardware or OS failures

- System upgrade

- Management of Firewall rules

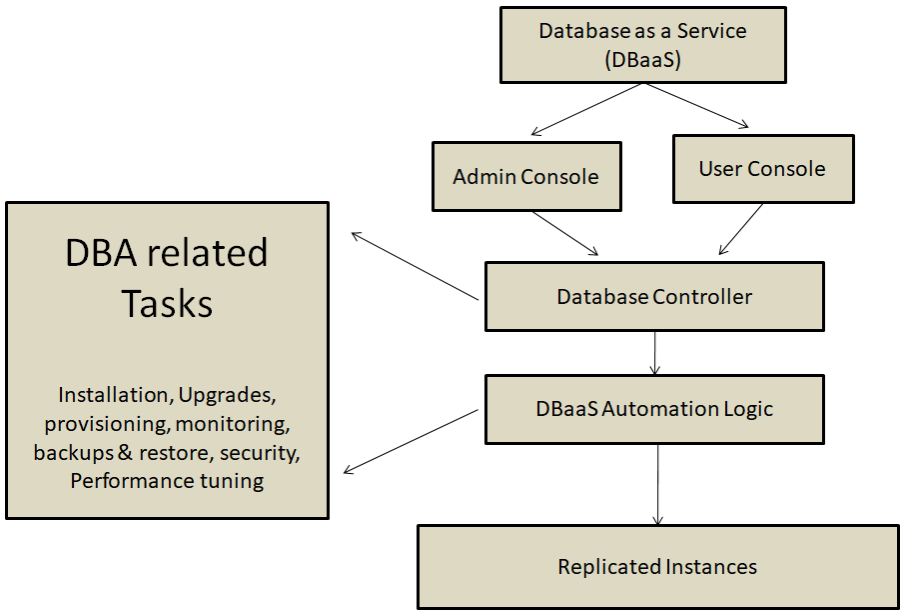


Figure- 4 DBaaS Services

The services offered by the DBaaS vendor to the end user fall into three main categories: Provisioning services, Administrative services, and Reporting services. Some of these services are optional, and others are mandatory

Provisioning Services: Some of the provisioning related services are,

- The ability to request new databases
- Choose database options as per requirements (partitioning, advanced security, Real Application Cluster)
- Add or remove computing resources (storage, CPU, network bandwidth, etc.) to existing databases.
- Database backup capability

Administrative Services: Admin services offered by DBaaS include,

- Perform on-demand database restores and recoveries
- Perform database cloning from existing database backups
- Database performance monitoring and alerting capabilities

Reporting Services: Reporting services include,

- Database performance management and performance tuning via GUI
- Resource consumption and usage reports to facilitate scaling related operations
- The ability to track provider compliance to the SLAs

➢ **Selection of correct DBaaS Vendor**

The selection of DBaaS vendor by particular organization depends on several parameters. These parameters are described in this section.

1. Location of Database

Generally organizations hire cloud based application servers to deploy their applications and database servers for data storage. Ideally, these both categories of servers should be located nearby to each other may be within same datacentre. If they are located away from each other, there will be two major problems. First one is the latency. Due to distance between application server and database server, there will be latency which will degrade overall performance of system. The other issue is security. Due to different location of application and database servers, they require network for communications. This kind of network based data transfer may cause security issues.

2. Fault Tolerance, Availability and Redundancy

If the organization is running a production level application, the database should always be available, even during the hardware failure or maintenance. Generally, DBaaS vendors provide fault tolerance and higher availability via multi node based clusters.

Before choosing DBaaS vendor, one should check the approach used by vendors for fault tolerance. The cluster nodes might be created within a same datacentre or in different datacentres at different geographic locations. Management of the cluster nodes at different geographic locations provides better protection against failures. Other parameters to consider are

- The mode of failure recovery: whether it is automatic or requires human interaction?
- Does the vender offer availability related SLA?

- How faulty components are replaced? Automatic or with human intervention.

3. Data durability via backup and restore

DBaaS vendor must have robust and reliable mechanism for backups and durability. The important things to consider are:

- Does the vendor use tools for managing backups?
- Does the vendor support point-in-time restore facility?
- Does vendor perform automatic backups and can we create custom schedule to create backups?
- Can we easily and quickly restore backups?

4. Performance Monitoring, analysis and Alerts

The Monitoring, performance analysis and alert features provide us the capabilities to check the health of database and its performance. Some of the important parameters to consider are:

- Does the vendor provide easy access to database related log files?
- Does the automatic alerts about failure are supported?
- Does the custom alerts are supported?
- Does the vendor provide real time and historical information about database performance metrics?

5. Performance and Scaling

The DBaaS vendor should provide a platform that performs well for normal workload. The similar kind of performance should be provided in case if the data volume becomes huge. The performance requirements of application should be assessed against the actual workload by simulating different kind of application workloads. Customers should not rely only on the benchmark workloads because it may not give correct picture about actual workloads. To support performance requirements, scaling operations (horizontal / vertical) are performed be vendors. In the vertical scaling, resources (RAM, CPU, Disks) are added as data volume grows. The horizontal scaling adds more nodes to the system to handle the growth in data volume.

6. Security

Data is the most important asset for any organization. The DBaaS vendor must have expertise in security and should provide tools to ensure data security against the unauthorized access. Some of the important things to consider are:

- Is the database access logged?
- Does vendor support encryption of data?
- Does SSL certificated validation based data access is supported?
- Is authentication required to connect to database?
- Can firewall be configured to limit the database access?
- Does the vendor perform security audits?
- Does the vendor follow best security practices and standards?
- Does the vendor own any security compliance certificate?

7. Support

The faster and helpful vendor support is very much important for smooth execution of database and application. DBaaS vendor must provide advises and must respond in case of any emergency. Some of the points to consider are:

- Does the support is a part of DBaaS or additional cost is required?
- Does the provider offer a SLA regarding support and response times?
- Does the provider have a good reputation for outstanding support?

- **DBaaS Conceptual Architecture**

The figure 5 [Ref oracle document] shows the conceptual architecture of DBaaS, various components of DBaaS and relationship among them. The DBaaS conceptual model shows the core capabilities which are required for the delivery of database related services.

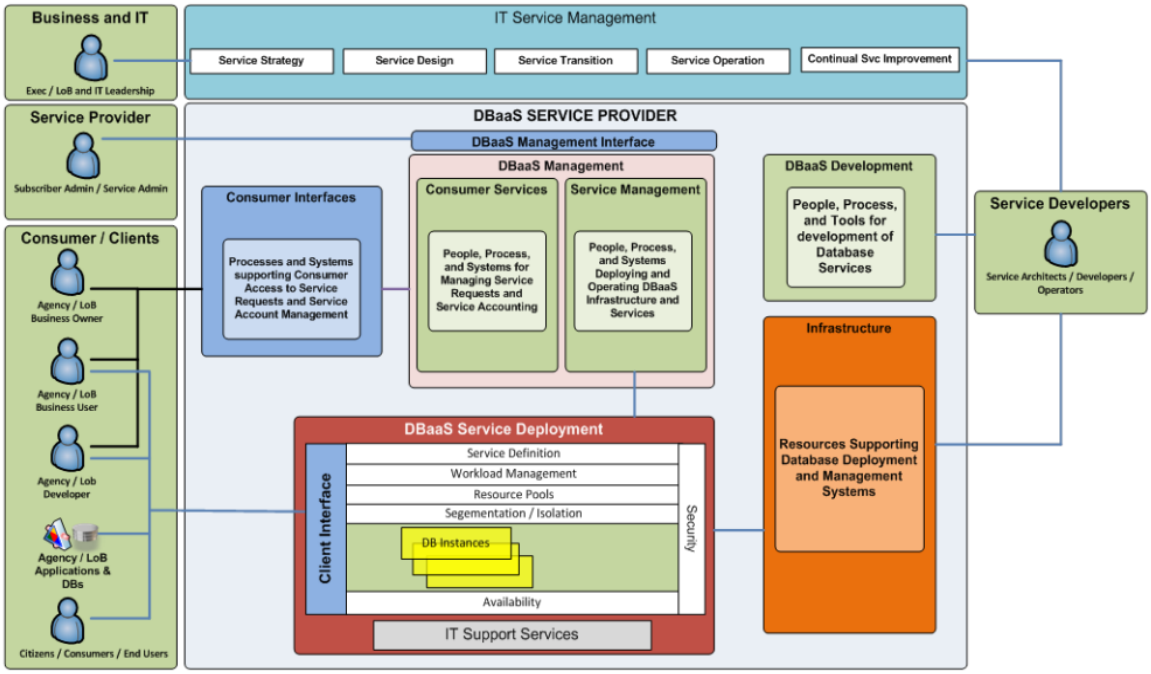


Figure-5 Conceptual architecture of DBaaS

1. DBaaS Development

DBaaS Development is required for defining the service offerings and for the management of the service catalogue. The DBaaS development is concerned with,

- The people, process and tools which are used to build the DBaaS service
- The design of the infrastructure required to provide the DBaaS
- The design and development of the system

2. DBaaS Management

It does the management of people, process and systems which are involved in organization's ability to request, manage, operate, and account for database services and their utilization. DBaaS Management has mainly two sub-capabilities:

- Subscriber Services- Supports the interactions between the subscriber and delivery of database services
- Service Management - Implements database services and manages the resources and systems supporting their delivery

3. DBaaS Service Deployment

This component deals with physical resources and their configuration to support the DBaaS. It also provides the interfaces required to manage the deployment, monitoring, and management of the services. It also deals with server configuration, networking and software which support the specific database deployment models.

4. Infrastructure

This component deals with the actual physical server resources required for systems and services for the management and deployment of the DBaaS architecture. The infrastructure includes servers, networking, software and other facilities.

5. IT Service Management

This component is used for service definition, design, operational practices and policies. The Service Management provides capabilities for defining and managing the services required for policies for design, change management, service operation & structure, the improvement framework.

6. Subscriber Interfaces

This interface provides systems and procedures to interact with the DBaaS management capabilities and subscribed database service instances.

2.4 LET'S SUM UP

Points to ponder

- Software as a Service Model
- Characteristics, features and benefits of SaaS
- Database as a Service Model
- Case Study about Software as a Service (SalesForce.com)
- Case Study about Database as a Service

Unit 3:Security as a Service 3

Unit Structure

3.1. Learning Objectives

3.2. Introduction to Security as a Service

3.3. Cloud Security Risk analysis

3.4. SECaaS Categories

3.5. Benefits Of Security As A Service

3.6. Evaluation of Cloud Security Issues

3.7. Cloud Security Standards

3.8. Let’s sum up

3.1 LEARNING OBJECTIVE

This chapter discusses about security in cloud computing model.

- Introduction to Security as a Service (SECaaS)
- Cloud Security Risk Analysis
- SECaaS Categories defined by CSA
- Cloud Security Control Mechanisms
- Benefits of SECaaS
- Evaluation of Cloud Security Issues
- Cloud Security Standards
- Case Study: SECaaS

3.2 INTRODUCTION TO SECURITY AS A SERVICE

Cloud computing paradigm offers various benefits like on-demand scalability, pay as you go, accessibility to the data and application from anywhere in the world etc. With all these benefits, one of the major issues in cloud environment is security. Though the cloud offers many benefits, the security related issues make customers reluctant about migration towards cloud.

Different types of cloud computing service models provide different levels of security services. You get the least amount of built in security with an IaaS and highest level of security with SaaS. Proxy and brokerage services can be used to separate clients from direct access to shared cloud storage. Logging, auditing, and regulatory compliance are all features that require planning in cloud computing systems.

To provide security assurance in cloud, vendors offer security related solutions in the form of services. These are called Security as a Service also known as SeCaaS. The SECaaS is a service model in which, the cloud vendors offer security solutions subscription basis. The major goal of SeCaaS is to provide security to information systems along with business objectives of the organizations. The SeCaaS mainly focuses on three pillars of information security namely: Confidentiality, Integrity and Availability. Another reason behind popularity of

SeCaaS is the lack of expertise in cyber security field. Many companies cannot find security professionals.

3.3 CLOUD SECURITY RISK ANALYSIS

Before moving to the cloud based approach, any organization should do the risk analysis. The following steps will help any organization in risk analysis.

1. Determine which resources are to be moved to cloud

   The organization should clearly identify set of functionalities (data, services, or applications) those will be shifted in cloud.

2. Determine the risk sensitivity of each resource.

   Risks regarding loss of privacy, unauthorized access, data loss, lack of availability etc. should be evaluated.

3. Determine the risk associated with the particular cloud deployment model

   Cloud types can be public, private (both external and internal), hybrid, and shared community types. With each of this type, we should consider where data and functionality will be maintained.

4. Identify, which service model will be used?

   For each different service model (IaaS, SaaS, and PaaS), the customers are responsible for security at different levels.

5. Evaluate vendor security system

   If you have selected a particular cloud service provider, you should evaluate its system and also understand how data is transferred, where it is stored, and how to move data both in and out of the cloud.

3.4SECAAS CATEGORIES

The Cloud Security Alliance (CSA) has classified SECaaS services into different categories. The figure 6 [CSA] shows the CSA cloud reference model with security boundaries.

Identity and Access Management

This category involves managing access to various enterprise resources by verifying the identity of the entities and granting the correct level of access to the entity on the basis of authorization.

Data Loss Prevention

This category includes protecting the data in the cloud in every state, i.e. data at rest and in motion

Web Security

This category involves providing real-time protection by redirecting the web-traffic to the cloud provider and then forwarding the clean traffic to the customer.

Email Security

This category provides control over outbound and inbound emails, thus protecting the customer from phishing and malicious attachments in email.

Security Assessments

These are the audits performed by third party for cloud services or assessment of the on-premises systems via cloud-provided systems.

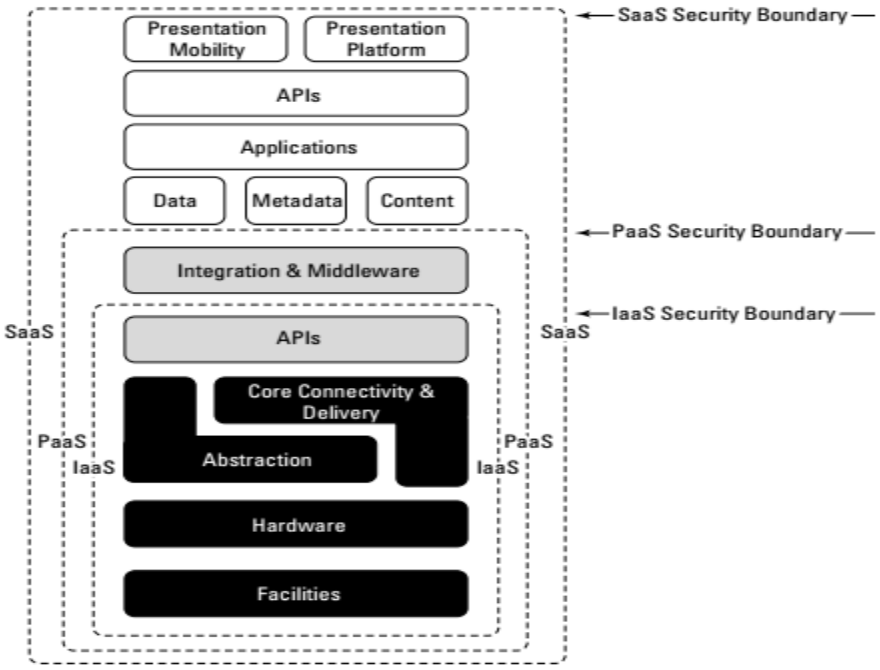


Figure-6 CSA cloud reference model [CSA]

Intrusion Management

This category includes the process of intrusion detection by prevention via anomaly based approach to respond to unusual events.

Security Information and Event Management (SIEM)

SIEM analyses and correlates the event logs related to security issues and provides real-time report and alert on the security issues that may require urgent attention.

Encryption

It is the process of providing private and public cryptographic algorithms for the security of data at rest, in motion, etc.

Business Continuity and Disaster Recovery

It is the process of ensuring the business objectives are in continuity in the event of any failures.

Network Security

It includes security provisions that allocate access and protect the underlying the network resource services.

- **Control Mechanisms for Cloud Security**

1. Detective Control

Detect and react promptly & appropriately to any security violation incident.

2. Preventive Control

Strengthen the system against security violation incident or attack by actually removing the vulnerabilities.

3. Deterrent Control

Reduce attack on cloud system; it reduces the threat level by giving a warning sign.

4. Corrective Control

Reduce the consequences of any security violation incident by controlling/limiting the damage.

3.5 BENEFITS OF SECURITY AS A SERVICE

1. The latest and most updated security tools available.

   The conventional anti-virus tools are effective and useful only if they are updated with the latest virus definitions. SeCaaS gives latest and updated tools to handle the latest threats.

2. You get access to the best security people

   Generally, there is a shortage of expert security professionals. It is very difficult to manage in-house security professional team. The SeCaaS vendors hire best security professionals. Using these services gives us access to their expertise.

3. Faster provisioning.

   The beauty of as-a-service offerings is that you can give your users access to these tools instantly. SECaaS offerings are provided on demand, so you can scale up or down as the need arises, and you can do so with speed and agility.

4. You get to focus on what's more important for your organization.

   Using a web interface or having access to a management dashboard can make it easier for your own IT team to administer and control security processes within the organization.

5. Makes in-house management simpler.

   If you have protected data, it is not enough to just keep it secure. You should know when a user accesses this data when he or she does not have any legitimate business reason to access it.

6. Save on costs

   You do not have to buy hardware or pay for software licenses. Instead, you can replace the upfront capital with variable operating expense, usually at a discounted rate compared to the upfront costs.

3.6 EVALUATION OF CLOUD SECURITY ISSUES

1. Authenticate all people who access the network.
2. Set access permissions for users to have access of the applications and data.
3. Authenticate all software and services running on any computer. Automate and authenticate software patches and configuration changes.
4. Formalize the process of requesting permission to access data or applications.
5. Monitor network activities and log unusual activity. The intruder detection system can be utilized.
6. Log all user activity and program activity and analyze it for unexpected behavior.
7. Encrypt, up to the point of use, all valuable data that needs extra protection.
8. Periodically verify the network related vulnerabilities in all software and services.

3.7 CLOUD SECURITY STANDARDS

The processes and practices required for the implementation of security in cloud are defined by standards. These standards are used to ensure secure and trusted environment with privacy and security of data in cloud.

A basic for security is to create layers of defense aka "*defense in depth*". This approach has an advantage - the security can be maintained even during the failure of any layer e.g. Use of Firewall system in combination with IDS (Intrusion Detection System). This type of layered security is desired in cloud computing environment. Traditionally, security was implemented at the endpoints. Organizations were using firewalls, IDSs, and antivirus software inside private network. Today, with managed security services offered by cloud providers, additional security can be provided inside the cloud. In this section SAML, OAuth, OpenID and SSL/TLS standards are discussed.

3.7.1 Security Assertion Markup Language (SAML)

SAML is a standard for the communication of authentication, authorization and attribute information among online partners. It is based on XML. It provides the facility to securely exchange identity and entitlements related assertions between partner organizations. SAML is built upon existing standards like: SOAP, HTTP, and XML. SAML uses HTTP as its communications protocol along with SOAP. SAML assertions and protocols are specified using XML schema. SAML uses digital signatures for authentication and message integrity. The SAML use XML-based assertions and protocols, bindings, and profiles. The SAML defines XML-based assertions, protocols, bindings, and profiles.

The SAML Core provides general syntax and semantics for SAML assertions. It also describes the protocol used to request and transmit assertions. SAML protocol refers to what is transmitted, not how it is transmitted. A SAML binding determines how SAML requests and responses map to standard messaging protocols.

The SAML has two parties involved: Asserting Party (SAML authority) and Relying party (Consumer/Requester). The asserting party is a platform or an application who relay security information. The relying party (consumer or requester) is a partner site that receives the security information. The exchanged information contains information like: subject's authentication status, access authorization etc. A subject is any entity in a particular system/domain.

SAML assertions

These are usually transferred from identity providers to service providers. The statements in assertion are used by service providers to make access control related decisions. Sample SAML assertion is below:

<saml:Assertion A...>

<Authentication>

...

</Authentication>

<Attribute>

...

</Attribute>

<Authorization>

...

</Authorization>

</saml:Assertion A>

The above assertion can be interpreted as:

> Assertion A, issued at time T by issuer I, regarding subject S, provided conditions C are valid.

SAML Protocols

A SAML protocol describes the method of packaging SAML assertions into SAML request and response elements. It also provides processing rules for SAML entities. Generally, a SAML protocol is a simple request–response based protocol. The most important type of SAML protocol request is a query. A service provider makes a query directly to an identity provider over a secure channel.

SAML Queries

There are mainly three types of SAML queries namely: Authentication query, Attribute query, and the authorization decision query. The attribute type query is mainly used. The response of an attribute query is a SAML response which contains assertion having attribute statement.

3.7.2 Open Authentication (OAuth)

OAuth provides simple and standard format for secure web application related API authorization. OAuth is a method used for publishing and interacting with protected data. For developers, OAuth provides users access to their data while protecting account credentials. The OAuth provides facility to users so that they can grant access to their information. After granting access, provider and consumers can share information without sharing the identity.

OAuth does not have features like automated discovery of endpoints, language support, XML-RPC and SOAP, OpenID integration, signing algorithms etc. It considers only fundamental aspects of the protocol,

- Establish a mechanism for exchanging a user name and password for a token

- Provide tools to protect the token

The Security and privacy are not provided by the protocol but SSL can help to get privacy.

3.7.3 Open ID

It is an open standard which provides digital identity for user authentication and access control based on single-sign-on (SSO) mechanism. The conventional log-in process (i.e., a log-in name and a password) can be replaced by allowing users to log in once and gain access to all participating systems. The OpenID is a kind of unique URL and is authenticated by the entity which has the OpenID URL. This protocol does not require any central authority to authenticate a user. The OpenID allows the parallel use of different kind of authentication methods like smart cards, biometrics, or ordinary passwords.

Sample usage of OpenID protocol:

- A user visits a web site where an OpenID based log-in form is displayed.

- The form contains only one field for the OpenID identifier (OpenID URL).

- This form is attached to OpenID client implementation library.

- User will have previously registered an OpenID URL.

- The user types this OpenID identifier into the OpenID log-in form and authentication process is completed.

The relying party requests the webpage at particular URL and reads HTML link tag to discover identity provider. There are two modes by which the relying party can communicate with the identity provider:

1. checkid_immediate

The relying party requests that the provider does not interact with the user. The communication is relayed through the user's browser without explicitly notifying the user.

2. checkid_setup

The user communicates with the provider server directly using the same web browser as is used to access the relying party site. This option is widely used on the web.

Steps to create a new OpenID session

- The relying party and the identity provider generate a shared secret handle, and the handle is stored by relying party.

- The checkid_setup process redirects relying party's user browser to the identity provider where, the user can authenticate with the provider.

- There can be different methods for authentication. Typical case can be like: The OpenID identity provider prompts the user for a password and asks whether the user trusts the relying party web site or not.

- If the user declines the identity provider’s request to trust the relying party web site, the browser is redirected to the relying party with a message indicating that authentication was rejected. The site in turn refuses to authenticate the user.

- If the user accepts the identity provider’s request to trust the relying party web site, the browser is redirected to specific page on the relying party web site along with the user’s credentials.

- The relying party must confirm that the credentials really came from the identity provider.

The relying party can be state full or stateless. The state full relying party stores the shared secret which can be used for any further communications. In case of stateless relying party, a background request using check_authentication is done for authenticating user. Once the OpenID based authentication is done, the user is considered as logged in at relying party’s system.

3.7.4 SSL / TLS

Transport Layer Security (TLS) is the successor of Secure Sockets Layer (SSL). They both are cryptographically secure protocols used to provide security and data integrity during TCP/IP based communication. These protocols perform encryption at transport layer. The TLS protocol allows client server based

communication across a network and provides protection against attacks like: eavesdropping, tampering, and message forgery.

TLS uses cryptography based mechanism at endpoints for authentication and data confidentiality. TLS uses one way authentication i.e. only the server is authenticated, because the client already knows the server's identity. In this case, the client remains unauthenticated. The browser validates server's certificate and also checks the digital signatures of the server certificate's issuing authority.

The validation process may not identify the server to the end user. For the correct identification, the end user must verify the identity related information stored in the server's certificate. The user must verify the URL, name and address details in server's certificate. Malicious web sites cannot use the valid certificate of another web site. Because, they cannot encrypt the data over transmission so that it can be decrypted using valid certificate. Only a trusted Certificate Authority (CA) can include URL in the certificate to ensure that the URL specified in the certificate is an acceptable or not.

TLS can support two directional security in which, both ends of the connection can be validated to ensure that users are communicating with the one whom they intended. This mode is known as mutual (assured) authentication. The Mutual authentication requires that both client and server must have own TLS certificate.

TLS Steps:

1. Peer negotiation for algorithm support

   The client and server negotiate about encryption, key exchange and authentication related schemes.

2. Key exchange and authentication

   The actual key exchange and authentication is performed using public key mechanisms.

3. Symmetric cipher encryption and message authentication

   The message authentication codes are made up from cryptographic hash functions and after that data transfer begins.

3.8 LET'S SUM UP

Points to ponder

- Introduction to Security as a Service
- Cloud Security Risk analysis
- SECaaS Categories
- Benefits Of Security As A Service
- Evaluation of Cloud Security Issues
- Cloud Security Standards

Unit 4:Specialized Cloud Services

4

Unit Structure

4.1. Learning Objectives

4.2. Recovery as a Service (RaaS)

4.3. Identity as a Service (IDaaS)

4.4. Storage as a Service

4.5. Communication as a Service (CaaS)

4.6. Let’s sum up

4.1 LEARNING OBJECTIVES

In this chapter, we will discuss about some of the solutions provided by cloud in the form of services.

- Recovery as a Service (RaaS)
- Identity as a Service (IDaaS)
- Storage as a Service (SAAS)
- Communications as a Service (CAAS)
- Network as a Service (NAAS)
- Monitoring as a Service (MAAS)

4.2 RECOVERY AS A SERVICE (RaaS)

Recovery as a service also known as disaster recovery as a service, is a category of cloud computing service used for protecting an application or data from any natural or human generated disasters or service disruption at one location.

RaaS provides an integrated service or solution to manage backup, archiving, disaster recovery related tasks. RaaS helps companies to recover their servers including OS, applications, and configuration, data, databases and files. In case of any disaster situation, RaaS helps to reduce downtime. Sometimes, RaaS is also known as DRaaS (Disaster Recovery as a Service). Some of the RaaS vendors are: nScaled, Geminare etc.

There are three basic models for RaaS

1. Managed

In this model, third parties take full responsibility of disaster recovery. In this option, organizations are required to work closely with RaaS providers to keep all infrastructure, application, and service changes up to date.

2. Assisted

In this model, the service provider offers services & expertise that are useful to optimize the disaster recovery process. The customer is responsible for implementing some or all of the disaster recovery plans.

3. Self Service

In this model, customers are responsible for planning, testing, and managing disaster recovery. Vendors provide backup management software, and hosts backups and virtual machines at remote locations.

Advantages of RaaS:

- Loss of critical company data can be avoided
- Protects against the loss or damage of physical infrastructure
- Offers cost effective data recovery
- Accurate and faster recovery is offered
- Offer greater flexibility for backup

Challenges of recovery of service are as below.

1. Find and maintain the minimal disaster recovery data footprint to provide necessary protection and to minimize data replication costs.

2. Continuous synchronization of the data between the production data center and the disaster recovery replicated site.

3. Establishing failover and failback processes should be as seamless and as automated as possible.

4. Allocate the human and IT resources to test the disaster recovery environment at regular intervals to make sure that it will work if and when needed

In order to meet these challenges, enterprises of all sizes often seek to implement their disaster recovery strategies by leveraging on-demand public cloud compute, network, and storage resources.

Key parameters to consider for RaaS

Reliability

RaaS service should be reliable and it should work in case of outgages of public cloud services.

Access

It is important to understand how customers will access services during disaster situations.

Assistance

It refers to the support that will be provided by service provider during normal situations and disaster situations.

4.3 IDENTITY AS A SERVICE (IDaaS)

The cloud computing paradigm allows users to access their resources from anywhere and at any time. In this scenario, the role of network is very crucial. In networked environment, proof of identity is very much important. The identity service stores user's identity related information in digital format such that, it can be further utilized for electronic transactions. Three core functions related to IDaaS are: storage of identity data, query engine and policy engine.

The transactions managed in distributed environment over public network have larger surface area for security attacks compared to private networks. Major security related issues are:

- Network traffic protection

- Resource access management and control

- Authorization based on identity

Among all, the establishment of identity is key issue. The identity provides trusted environment. IDaaS offers mechanisms for digital identity management across multiple systems. The digital identity is a set of attributes that gives reorganization to individual over the networked systems. A digital identity helps systems to identify objects and relationships among them. An identity of a person describes:

1. Biological attributes like age, race, gender, appearance etc.
2. Information about person like: Biography, personal data such as social security numbers etc.

3. Ownership details of a person like: A pattern of blood vessels in eye, fingerprints, accessible bank account, a security key, objects and possessions etc.
4. Relationships with others: Friends and Family, beliefs and values, activities and endeavors, personal selections and choices, habits and practices etc.

The identity over the network can be established by several mechanisms like, single factor, two factor, multifactor etc. In single factor based identity, a user has to provide his/her credentials in the form of user name and password. In two factor authentication mechanism, user needs to provide credentials along with hardware based key. Multi factor authentication requires extra mechanisms like biometric, OTP etc.

Example: Microsoft Office identity profile

The Microsoft maintains identity profile of MS office installation by considering many factors or attributes like:

- A 25-character software product key and product ID
- The uniquely assigned Global Unique Identifier or GUID
- PC manufacturer, CPU type and serial number
- BIOS checksum
- Network adapter and its MAC address, Display adapter
- SCSCI and IDE adapters
- RAM amount, hard drive and volume serial number
- Optical drive
- Region and language settings and user locale

From above attributes, a code is generated and stored into product registration database. If any of the above attribute changes, the reactivation of product is required.

Identity Services in Network

There are several forms of identity services used to validate website, transactions, transaction parties, clients and network related services. Certificate based identity, Ticket or token based identity and other mechanisms are used to manage trust in networked services. The protection of an identity is very challenging and complex task. Identity as a Service (IDaaS) offers different services as listed below:

- Services related to Authentication (identity verification)
- Directory services
- Identity governance
- Identity and profile management
- Policies, roles, and enforcement
- Identity provisioning (external policy administration)
- Registration
- Risk and event monitoring
- Single sign-on services

AZURE AD is example of IDaaS.

4.4 STORAGE AS A SERVICE (STaaS)

Storage as a service is a managed service in which the provider offers the access to data storage platform. This service can be delivered either on-premises (Infrastructure dedicated to a single customer), or from the public cloud (As a shared service).

Many organizations use storage as a service from public cloud for storage and backup needs. There are different cloud storage methods id STaaS like backup and restore, disaster recovery, block storage, SSD storage, object storage etc.

Backup and restore is for backing up of data to the cloud, which provides protection in case of data loss. Block storage makess provision of block storage volumes to

customers which are good for low latency input output operations. The SSD (Solid State Drive) storage is another storage type that is good for applications having intensive read/write and input output operations. Object basedstorage systems are used in data analytics. Cold storage is used to create and configure stored data quickly.

Cloud storage has a many advantages over traditional data storage. One of the most important advantages is that the user can access stored data from any location with the help of Internet. A cloud storage system maintains one or more data servers connected with the Internet. When a subscriber copies files to the server over the Internet, files are stored in data servers. When a client wants to retrieve the data, he or she can access the data server with a web-based interface, and the server then either sends the files back to the client or allows the client to access and manipulate the data itself.

The Storage as a Service means that a third-party (service provider) rents storage space to end users who do not have budget or capital. Storage as a Service is also useful when, the technical personnel to maintain storage infrastructure are not available. It is also very cost effective.

Advantages of STaaS

Major advantages of STaaS for organizations are as below:

1. The storage costs on hardware and physical storage devices are expenses are less.

2. The multiple copies of data stored in different locations offers enhanced disaster recovery.

3. STaaS offers higher level of storage level scalability using public cloud services. In this setup, users have to pay for the resources (storage) used by them.

4. STaaS offers automatic syncing of replicated files which are stored at different locations.

5. The Security is both advantage and disadvantage of STaaS. Generally, during transmission and during reset the data is encrypted for better security and protection.

Disadvantages of STaaS are:

1. Organizations may transfer business-sensitive or mission-critical data to the cloud. In this case, it is important to choose a service provider who is more reliable.

2. If data transfer bandwidth limitations are crossed, these services may be more expensive.

3. If vendors go through downtime periods organizations may not use services, which depend on mission-critical data.

4. The cloud infrastructure is owned and managed by the service provider, it is less customizable.

5. It may be difficult to shift from one service to another.

4.5 MONITORING AS A SERVICE (MaaS)

Monitoring as a service (MaaS) is one of many cloud computing delivery models which is offered as a service. It is a framework to facilitate the monitoring of other services and applications deployed in cloud.

IT infrastructure monitoring is an important part of the IT Management Policy for any organization which uses cloud based environment. Proactive monitoring improves business operations to get quicker disaster recovery, and easier capacity planning for all mission-critical applications. MaaS is a new delivery model which is suited for organizations who adopt a monitoring framework quickly with minimal investments.

Benefits of Monitoring as a Service (MaaS)

1. Use Monitoring Tool: The service providers take care of hardware setup infrastructure, monitoring tool, configuration, and alert settings on behalf of the customers. After login customers can access monitoring dashboard using an internet browser.
2. MaaS has higher availability: As the MaaS is deployed in the cloud, the monitoring dashboard should be available 24x7x365. There should be no downtimes associated with the monitoring tool.

3. Easy Integration: MaaS can generate various alerts based on specific conditions. MaaS also offers multiple levels of escalation to serve different user groups and to provide different levels of alerts.
4. MaaS is in the cloud, it works well with other cloud-based services like PaaS and SaaS. MaaS can monitor Amazon and Rackspace's cloud infrastructure. MaaS can also monitor customer's private cloud deployments.
5. The MaaS customers need not to invest in a network operations center. They need not to invest an in-house team of qualified IT engineers to run the monitoring desk.

Situations when MaaS can be utilized

1. For the small and medium level organizations, MaaS helps to manage pay per use services effectively. It helps them to reduce extra costs on servcies.
2. MaaS is quite useful to monitor various cloud based SaaS and PaaS services. MaaS can be opted as an add-on service along with SaaS and PaaS.
3. MaaS is a good to have centralized infrastructure monitoris in case when organization's infrastructure / resources are distributed at different locations.
4. If the organization has private infrastructure along with public services, MaaS offers monitoring service in hybrid environment.
5. MaaS provides good backend framework to service providers who offer multitenant hosted services to increase availability.

4.6 COMMUNICATIONS AS A SERVICE CAAS

Communications as a Service (CaaS) is an outsourced enterprise communications solution which is leased from a single vendor. These communications can include voice over IP (VoIP or Internet telephony), instant messaging (IM), collaboration and videoconference applications using fixed and mobile devices.

CaaS providers manage the hardware and software that are important for delivering Voice over IP (VoIP) for voice communication service, and other services like Instant Messaging (IM).

Features of CaaS

1. Integrated and Unified Communication

The advanced unified communication features include Chat, Multimedia conferencing, Microsoft Outlook integration, Real-time presence etc.

2. No Investment Required

It is the sole responsibility of CaaS vendor to manage hardware and software deployed to provide the communication service to their customers. The customer only has to pay for the service he is getting from the CaaS vendor, not for communication features deployed to provide communication services.

3. Flexibility & Scalability

The customer can outsource the communication services form CaaS vendors. The customers pay for what they have demanded. The customer can extend their service requirement according to their need.

4. No Risk of Obsolescence

The CaaS vendors keep on updating their hardware and software that provide communication services to meet the changing demands of the market. So the customer using the services does not have to be worried about the service obsolescence.

5. No Maintenance Cost Incurred

The customer outsourcing the CaaS service does not have to bear the cost of maintaining the equipment deployed for providing communication services.

Advantages of Communication as a Service (CaaS)

1. CaaS provides an economical way to deliver communication service to its customer by preventing them from investing in hardware and software required for delivering communication services.
2. CaaS vendor provides 24/7 service to its customers.
3. Customer receiving services from CaaS vendor do not have to indulge and invest in managing the components of CaaS.
4. CaaS vendor offers flexible service as they charge according to pay as you go basis.

5. CaaS provide scalable services as they provide service based on customers demand. CaaS provides the hosted and managed solution which offers complete communication solutions managed by a single vendor only.
6. From the customers perspective, there is no risk of service becoming obsolete as the vendors are responsible for upgrading the carrier platform.

4.7 LET'S SUM UP

Points to ponder

- Recovery as a Service (RaaS)
- Identity as a Service (IDaaS)
- Storage as a Service (SAAS)
- Communications as a Service (CAAS)
- Network as a Service (NAAS)
- Monitoring as a Service (MAAS)

Block-3
Application Architecture for Cloud and Cloud Deployment Techniques

Unit 1: Introduction of Cloud Application

1

Unit Structure

1.1. Learning Objectives

1.2. Introduction to Cloud Application

1.3. Cloud Application Requirements

1.4. Architecture for Traditional versus Cloud Applications

1.5. Assumption for Traditional and Cloud Applications

1.6. Let’s sum up

1.1 LEARNING OBJECTIVE

This chapter discusses about,

- Introduction to Cloud Application
- Cloud application requirements
- Architecture for traditional versus cloud application
- Assumption for traditional and cloud applications

1.2 INTRODUCTION TO CLOUD APPLICATION

Software application is generally built using many modules. Such modules interact with other modules and also provide interface to external application or services. Application architecture is the design or plan of such software application. It contains the details of software application components, interaction among the components and also interaction of the application with the infrastructure components. Application goal is to perform useful business task so application design also considers to automate the business task. This automation of business task helps users to manage, store and share data.

Software application designed using traditional application framework is suitable for stable demand level of the consumer. In such software application some middle-tier software or application framework is used to access database by the web server. Such traditional application framework is not much useful and suitable for application with huge variation in user demand or system load.

It is estimated that in near future all major applications will be migrated to cloud as its easy to scale as per the variation in the demand and other useful features. While migrating application to cloud environment, the migration process needs to be planned carefully. After migrating the application to the cloud, software application code needs to be changed to to suit the cloud and also to ensure better user experience.

Many organizations started migrating applications in cloud. After migration to cloud the application can provide better user experience by solving many issues for users. It may also useful to streamline the business tasks in many ways. Even the

application get migrated to cloud but if it is not modified to suit the cloud environment than it will work similar to traditional on premise application.

Time, cost and special skills in cloud-based application architecture is required to rebuild the existing application and to add and use additional features of cloud.

1.3 CLOUD APPLICATION REQUIREMENTS

For developing cloud application proper documentation of design and plan is required in order to gain advantage of cloud environment over the traditional environment. Cloud provides services like cloud-based authentication, security and replication, the newly created application must be capable to coexist with these services.

Requirements and architecture are the two major documents must be written and reviewed carefully first while working with cloud applications.

Application Requirements and Constraints can be summarized as below

- Business needs
  Organization's business may grow over a period of time and it may lead to need for more IT infrastructure and across various locations. In this case, cloud based applications are great choice to fulfill these demands.
- Organization's vision
  Organizations grow over a period of time and they have set a vision regarding their business. To achieve the vision, they always need their applications to be available across different locations with great customer satisfactions. Cloud based applications are highly scalable and can help organizations to meet their vision.
- Legal limitations when hosting in the cloud
  Organizations may have some mission critical data which contain business secrets. Different country may have different laws regarding the storage and management of such data. All legal aspects must be considered before developing cloud based applications.

- Cloud standards

  There are several standards available for various cloud services. If applications are developed by considering those standards, they will be more secure and scalable. So, applications must follow those standards.
- User of existing templates

  The template helps us to reuse some or whole part of existing functionality. While cloud application development, we should consider templates.
- Corporate policy for cloud use.

  Each corporate or business has own policy regarding the use of IT in their day to day operations. Cloud application developers must consider their policy before proceeding for any applications.

Again, application requirements can be classified as below.

1) Functional requirement
    - Purpose and objective
    - Business goals
    - User requirements
    - Required features
2) Nonfunctional requirements
    - Performance and response time
    - Security
    - Service availability in the cloud
    - Backup to other clouds
    - Extension to hybrid cloud.
    - Localization
    - Compatibility with other cloud platforms
    - Support for end user devices

Application requirements are considered for the application architecture. The architecture contains details of each section of the application. It also contains the details of how different sections interact with each other. Application architecture exists at conceptual level and detailed level. Conceptual level deals with business offerings, market products, technology growth etc. detailed level deals with designing, reuse of services and design of user interface.

Certain requirements are difficult to define on public or private cloud where the application will be hosted such as server architecture for IaaS cloud, backup, fault tolerance, data replication technique, security etc . the server architecture covers the hardware design and provide application deployment. Vertical and horizontal scalability is supported by good application server architecture.

1.4 ARCHITECTURE FOR TRADITIONAL VERSUS CLOUD APPLICATION

Architecture of the application for cloud is different from the application reside on a traditional, on- premise infrastructure.

In cloud cost is based on resource usage. Resources are available as per demand. Resources are virtualized and at multiple locations which need to be factored in the application architecture. Applications must be able to scale the resources either horizontally or vertically to use resources in highly granular manner and be cost effective to operate.

Cloud native application should have the following important properties.

- They should be designed such a way that can take maximum benefits of cloud environment. Applications must consider the concepts of distributed architecture and various modules should be designed such a way that they can be distributed and work in sync with each other.

  Scalability is another benefit of cloud environment. Application architectures should be created considering the future needs of scaling. The protocols and concepts used for application development must consider scalability feature. Generally, organizations require higher level of scalability and availability for their applications. In this situation, applications will become multi tenant i.e. hosted on different machines.

  Cloud based applications will run by different users over different platforms (operating systems and hardware). Application developers must consider platform independence as a primary requirement.

- Different from a cloud enabled or traditional monolithic application

In cloud based applications, different components or module are hosted / distributed over different machines. All these modules / components must have inherent capability to support scaling as individual modules.

- Light-weight

  Cloud application / modules may run on different virtual machines (VMs) or in containers. So, each module must be designed by considering that it will run on separate VM (virtual machine) or container. It should have capability to communicate with other module running under different containers or VMs.

- Devops/ci based deployment

  Cloud based applications may run in on-premise private cloud or it may be deployed in public cloud. During development, applications may run on development servers. But, finally they may be deployed in cloud. Generally, these deployment and change related should be done such a way that they can be automated or may require minimal human intervantions.

Some important points for cloud native versus traditional application can be summarized as below.

Factors to be considered	**Cloud native applications**	**Traditional applications**
Operating systems	Os abstraction: No need to manage the Operating systems	Os dependent: Traditional applications are Operating systems dependent.
Predictability	Predictable: Possible to predict the time to deploy the application.	Un-predictable : Testing cycle is un predictable
Capacity	Right sized capacity: easy to scale as per requirements (scale up or down)	Oversized capacity: Not possible to accurately use the capacity of the machine

Delivery model	Continuous delivery: Easy to update software. Easy to drive customer needs, Easy to do test driven development.	Waterfall : Waterfall model is used
Scaling	Automated scaling: Easy to scale up or scale down resources as per the process requirements.	Manual scaling; Scaling is manual
Recovery	Rapid recovery: In-built capability to identify erroneous system and remove that.	Slow recovery : Recovery is slower.

Cloud Application Architecture Challenges

Some of the types of challenges to consider while selection of an architecture style

1. Complexity- The complexity of the architecture should be justified for domain. Style should not be too much complex as well as too much simple.

2. Asynchronous messaging and eventual consistency

Asynchronous messaging is used to decouple services. It is also used to increase reliability (because messages can be retried). This creates challenges such as always-once semantics and eventual consistency.

3. Inter-service communication

As we decompose an application into separate services, there is a risk that communication between services will cause unacceptable latency or create network congestion.

4. Manageability

It specifies how hard is to manage, monitor, deploy and update the application.

Cloud Application Architecture Styles

Some of the most commonly used application architecture styles are:

1. N-tier

N-tier is a traditional architecture for enterprise applications. Dependencies are managed by dividing the application into layers that perform logical functions, such as presentation, business logic, and data access. N-tier is a natural fit for migrating existing applications that already use a layered architecture.

2. Web-Queue-Worker

The Web-Queue-Worker architecture is quite suitable for PaaS applications. In this style, the application has a web front end that handles HTTP requests and a back-end worker that performs CPU-intensive tasks. The front end communicates to the worker through an asynchronous message queue. Web-queue-worker is suitable for relatively simple domains with some resource-intensive tasks.

3. Micro services

If the application is more complex, we should useMicro services architecture. A micro services application is composed of many small, independent services. Each service implements a single business capability. Services are loosely coupled, communicating through API contracts.

4. CQRS

The CQRS (Command and Query Responsibility Segregation) style separates read and write operations into separate models. This isolates the parts of the system that update data from the parts that read the data. The read operations can be executed against a materialized view that is physically separate from the write database. That lets us to scale the read and write workloads independently, and optimize the materialized view for queries. CQRS is useful when it's applied to a subsystem of a larger architecture.

5. Event-Driven Architecture

Event-Driven Architectures uses apublish-subscribe (pub-sub) model, where producers publish events, and consumers subscribe to them. The producers are independent from the consumers, and consumers are independent from each other.

1.5 ASSUMPTION FOR TRADITIONAL AND CLOUD APPLICATIONS

The application architect keep using long running assumptions that worked earlier many of these conflicts in cloud area. The architect needs to know which assumptions to discard and what new to adopt. The table below shows the old and new paradigms and architecture principles.

Traditional applications depend on the conventional mode of application development and deployment architectures whereas, cloud based applications are built with the assumption that they will be deployed on highly dynamic and scalable cloud environments. Conventional applications run directly on hardware devices and are created with assumption that applications have complete access of hardware resources. Cloud computing uses the virtualized environment i.e. all hardware resources like memory, processor, network, disk are virtualized in the form of VM (Virtual machine) or Containers and applications will have to run on those virtualized resources.

Generally, traditional applications are designed to run in centralized environment. Applications are designed by considering that all computing resources are available at central location and users / customers will access application from central location. Cloud based applications are designed with the goal that various components of application are deployed in distributed manner across various sites / locations. Users or customers will try to access any resource / service from any location at any point of time.

Traditional applications make use of centralized databases for applications. For these kinds of databases, integrity management can be done by conventional integrity constraint based method. Cloud based applications make of distributed databases in which data is replicated at several locations for higher availability and fault recovery. As data is stored in distributed manner, the needs for special integrity management schemes are required for cloud based applications.

Conventional applications make use of structured data whereas cloud based applications make use of data in different format. In cloud, data may be structured, unstructured or text format. Similarly, the input / output format for conventional

applications are fixed. In cloud based applications, the format of input output data varies largely.

Sr.	Traditional assumptions	New thoughts or practice
1	Application depends on the similar infrastructure	Applications will be running on highly dynamic and hybrid cloud environment.
2	Applications can have access to device files (unchanged and static)	Application runs on the virtual resources which can move to other resource also amount of resource can vary.
3	Application exists in single location.	Applications are distributed at multiple locations.
4	For data and process integrity database can be used.	Application must be aware of the data integrity and be able to manage the integrity issues.
5	Structured data and predefine format for application.	Different media types and various data types the same field or information for application.
6	Applications will have a fixed format for input and output.	Applications must be designed around social, inter-personal communications.

1.6 LET'S SUM UP

Points to ponder

- Introduction to Cloud Application
- Cloud Application Requirements
- Architecture for Traditional versus Cloud Applications
- Assumption for Traditional and Cloud Applications

Unit 2: Cloud Application Requirement

2

Unit Structure

2.1. Learning Objectives

2.2. Introduction

2.3. Recommendations about Cloud Application

2.4. Use of Client Server Mechanism for Cloud Applications

2.5. Service Oriented Architecture (SOA) based Cloud Applications

2.6. Parallelization within Cloud Applications

2.7. Let’s sum up

2.1 LEARNING OBJECTIVE

The overall objective of this chapter is to,

- Study about requirements of cloud applications
- Role of Service Oriented Architecture (SOA) in Cloud Applications
- Architecture of parallel cloud applications

2.2 INTRODUCTION

In previous chapters, we have discussed about physical architecture of cloud computing environment. This chapter we will study about cloud application requirements, SOA and parallel cloud applications. The overall environment of cloud is quite different than traditional applications. We should architect cloud applications to leverage the benefits of cloud computing based environment.

2.3 RECOMMENDATIONS ABOUT CLOUD APPLICATION

We know that cloud computing environment is different than traditional environment. This section gives some of the recommendations for cloud application design.

1. Location independent

   Traditional applications are created to run on hardware machines at same physical location with similar,

   - Network configurations
   - Security settings
   - Regulatory domain

   In case of cloud computing environment, the application may run in datacentre of cloud service provider at any location across the globe. Application developer should also provide option of changing cloud service provider where the application is deployed. Sometimes applications are developed and tested over public cloud and finally deployed over private cloud (infrastructure). The application architect must consider the data as local or remote and the data should be able to access in parallel.

2. Consideration of resource price and utilization

   The pay per use model and scalability are major advantages of cloud based approach. The cloud users are charged according to the usage of resources. In such a scenario, the application architect should try to optimize the hardware usage by application. Similarly, the application scalability should also be considered. In some cases, the cloud service provider price changes according to the time. Some applications tasks like replication, backup, report generation etc. should be scheduled such a way that service provider's costs are lower.

3. Flexible, Dynamic and distributable

   The traditional applications are developed by considering traditional environment in which applications are designed to run on consistent set of computer hardware with stable features and behaviour. This kind of assumption may not work in cloud. The processing data may shift location at any time. Similarly, the available processing capabilities, memory and network bandwidth may very over the period of time. The application design should support heterogeneity in terms of hard resources and runtime environments.

4. Data consistency, integrity and security

   Traditional applications consider that transactions are atomic and atomicity can be achieved by applications or by some middleware software. Many times database software maintains atomicity. In case of cloud, data is distributed at different locations and in different databases. A proper mechanism to manage integrity and consistency should be developed.

   The security of data should be enhanced by automated configuration changes. Application lifecycle management should consider that it may happen that multiple versions of application run in parallel. The cloud based applications rely largely on network. The service provider and consumer have no direct control over latency and delay. Applications must be designed to sustain data integrity related issues. The application should be designed such a way that it can recover from partial failures. It must be fault tolerant i.e. it should have capability to accommodate a wide range of faults.

5. Mobile aware

   It will save effort and time if the application is designed and developed by considering access via smart phones, tablets or any other mobile devices. In

future, it may happen that application will be accessed by more and more mobile devices.

6. Applications Should be feature rich

   Traditional applications were designed to accept, preserve and process user inputs (data). It was mainly used for reporting and record keeping. In addition to these basic requirements, cloud applications should have option to communicate with other applications to support business processes. Cloud applications must include human interaction as a part of basic requirement. E.g. if we develop doctor appointment system over cloud, it should have features regarding automated email and SMS when the appointment is booked and even on the day of appointment. A better approach to build relationship between application and society is to use event driven approach.

2.4USE OF CLIENT SERVER MECHANISM FOR CLOUD APPLICATIONS

The application design approach is evolved from monolithic to modular to object oriented to service oriented. The concept of cloud computing and the evolution of powerful mobile devices, has created more options for cloud based applications. The powerful interconnected mobile devices and cloud computing have led to a new approach called Client-Cloud Architecture. In this scenario, developers prepare two applications:

a. Server Application for Cloud

   These applications are architected over private cloud, public cloud or PaaS platforms.

b. Client Application for devices

   These applications include smart phones, tablets and other connected devices and along with different operating systems like IOS, Android, Microsoft etc. The client applications can be distributed online via various app stores. The rich client applications reduces burden on server applications.

As we can see that client cloud approach have many benefits, there are several issues as well.

- Frequently changing client hardware and software configurations

- Less backup facility at client side
- Network delay and security issues
- Different setup and configuration files for type of clients

2.5SERVICE ORIENTED ARCHITECTURE (SOA) BASED CLOUD APPLICATIONS

The overall architecture of cloud based applications differs from traditional application architectures. There is a need for set of separate design principles and design patterns. The design principles and patterns are very much useful to architects and developers. They help to reduce risks, costs and time associated with application building, deploying and managing.

SOA is one of the best architectural guideline for cloud based application developments. This approach focuses on service based application development. Some of the basic characteristics of services are,

- These services are highly interoperable.
- Each service is identified and created to support business process.
- These services work as software modules.
- They can be reused
- They are independent (loosely coupled) so changes become easy
- Self-discoverable
- Services can communication with other services using well defined standards and protocols
- Services are self-contained.
- Service based application approach gives agility
- Services can be integrated (orchestrated) to achieve desired application feature
- Support cross platform and network based integration

Cloud offers all resource in the form of services. In this scenario, the Service based approach is more suitable for cloud applications. Earlier the SOA based approach was used for distributed application development and for modular application

development. Nowadays, SOA is a basic approach for SaaS based software delivery. Various patterns related to SOA are shown in figure 7.

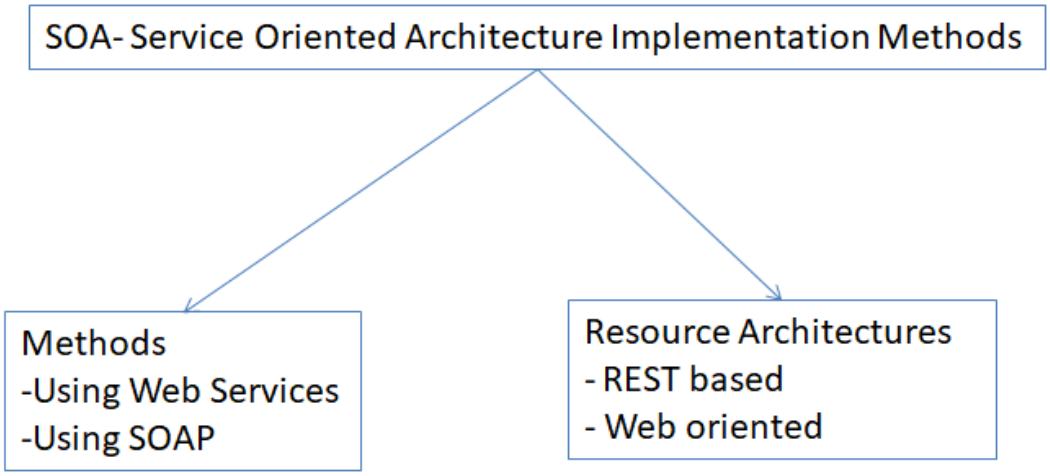


Figure-7 SOA implementation methods

The figure 7 shows overall classification of SOA implementation methods.

1. Resource Oriented SOA
   - Uses WWW standards (HTTP)
   - REST (Representational State Transfer) is used for web service implementations
   - This approach is used to support large scale applications over public cloud platforms
2. Method Oriented SOA
   - It uses SOAP (Simple Object Access Protocol) based standard. This protocol is used for data exchange over networks.
   - Support common request – response mechanism
   - WSDL (Web Service Description Language) is used for describing web services.
   - SMTP is used for message transfers
   - XML is used for configuration and exchange of information

3. Event Oriented SOA
   - It is based on asynchronous message exchange approach
   - The cloud application registers as listener and receives a message about an event once the event is generated by source program.
   - The events and messages are displayed over dashboard

- This approach allows business executives to make transaction related decisions based on up to date information.

All three approaches mentioned above enable SOA based approach to be,

- Modular: Modularity of an application helps to serve software in the form of different modules. Each module can be offered in the form of service.
- Flexible: Flexibility refers to the freedom in terms of software modules. Various modules are in the form of services. Customers can choose their required modules or services.
- Loosely coupled: Modular design mainly depends on the concept of loose coupling. Coupling refers to the communication between various modules of application. For good design, the coupling between modules should be less. If coupling is high, the dependency among modules is high or called tight coupling. If the coupling is high, design becomes less flexible and software becomes rigid. SOA uses the concept of loosely coupled service based approach.
- Discoverable

Cloud computing offers various services / resources which are made available in the form of self service. Customers can choose their desired services from set of available services. To support the self-service mechanism, services should be discoverable i.e. customers / users can discover their desired service.

The overall benefits of SOA based application development can be summarized as,

- More than 60% of SOA based projects have positive effect on organizations
- SOA based approach generate positive returns in short time span
- SOA help to improve agility of organization
- SOA can reduce cost involved in making of IT based systems
- SOA improves developer productivity

2.6 PARALLELIZATION WITHIN CLOUD APPLICATIONS

The on-demand availability of large amount of processing capability and memory force architects to use resources effectively. The scalability refers to the concept by which organizations can add / update available computing resources as per the requirements. The scalability can be in two forms either horizontal or vertical.

- Horizontal scaling for front-end and middleware components

  The horizontal scaling does the addition or removal of VM instances as per the changing workload. During horizontal scaling, organizations can manage computing resources of by adding similar type or different types of resources like virtual machine or container. Generally, all public service providers have support for horizontal scaling.

- Vertical scaling for back-end tier

  The vertical scaling reconfigures the VM by changing its resources as per the requirements. In vertical scaling, instead of adding or removing resources the capacity of virtual machine or container is changed. In this approach, VM resources are increased or decreased. Vertical scaling is not supported by all public service providers.

Horizontal and Vertical Scaling

The scaling operations can be executed either manually or automatically.

Auto scaling is the mechanism can be used to add or remove computing resources automatically and dynamically as per the application requirements.

1. Manual Scaling: IT Operations organization analyzes utilization and cost metrics periodically and scales workload resources based on peak and average demands. When a utilization threshold is crossed, monitoring tools notify your team or performance degradation is noticed and they reactively scale the workload's resources
2. Autoscaling: IT Operations leverages IaaS auto scaling capabilities to supplement manual instance selection and sizing with built-in infrastructure ability to automatically increase capacity on demand.

2.7 LET'S SUM UP

Points to ponder

- Recommendations about Cloud Application
- Use of Client Server Mechanism for Cloud Applications
- Service Oriented Architecture (SOA) based Cloud Applications
- Parallelization within Cloud Applications

Unit 3: Successful Cloud Implementation

3

Unit Structure

3.1 Learning Objectives

3.2 Introduction

3.3 Recommendation about cloud applications

3.4 Cloud Computing Networking Issues and Mitigation

3.5 Cloud deployment automation

3.6 Self-service features in cloud computing

3.7 Cloud performance

3.8 Let’s sum up

3.1 LEARNING OBJECTIVE

In this chapter we will discuss mainly about various cloud deployment techniques along with major factors which affect the successful cloud implementation.

- Successful implementation of Cloud Computing
- Network topologies in Cloud Computing Environment
- Automatic deployment in Cloud
- Performance Monitoring

3.2INTRODUCTION

Large organizations have spent long time and money to establish fault tolerant with all required features like security, availability, performance requirements etc. Cloud offers more advantages like pay per use, unlimited resources, less initial cost.

3.3 RECOMMENDATIONS ABOUT CLOUD APPLICATION

Organizations should consider following factors for before successful cloud implementation or adaption.

- Service Level Agreements (SLAs) with existing customers
- Compliance or regulatory related issues
- Which set of services are to be deployed over cloud and deliver to customers?
- Proper distribution of services i.e. which services should be deployed on cloud infrastructure and which are to be deployed on local infrastructure?

Along with above mentioned factors, some other important parameters to should e considered before selecting the service provider figure 3.3.1. These are:

1. Risk analysis must be done because in cloud environment the data and major control resides with third party. One should ensure that the cloud deployment cost should not drastically increase over the internal deployments.
2. Initial cost as well as the operational cost of cloud should be considered by considering the kind of application or workload.
3. One should also think about the emergency plan. If the provider goes out of business? What steps will be taken to keep services up. In the similar fashion the SLAs must be prepared with cloud service providers.

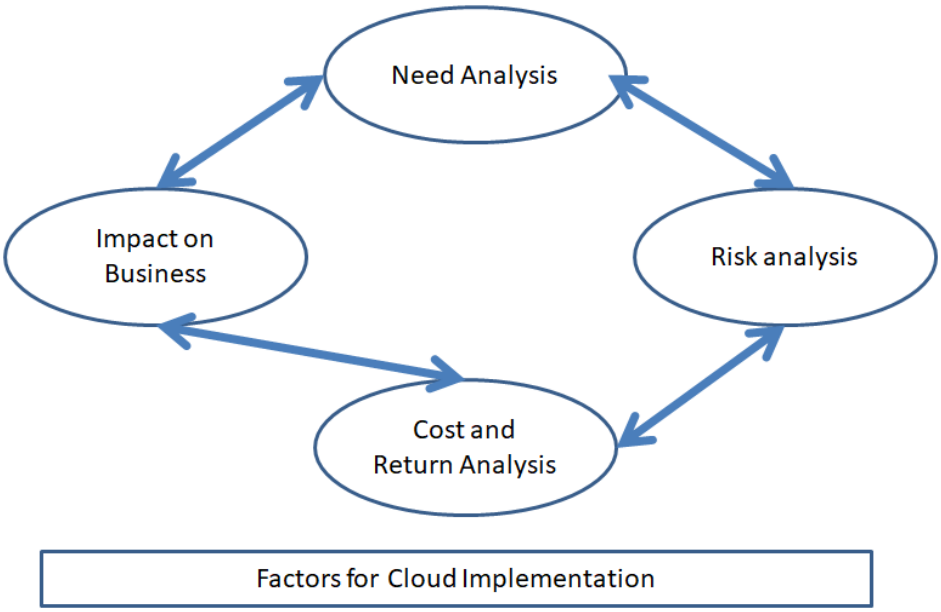


Figure 3.2.1 Factors for successful Cloud implementation

3.4 CLOUD COMPUTING NETWORKING ISSUES AND MITIGATION

During the deployment over cloud, several latency related problems may occur. The service provider and the user organization should work together to find out causes of problem and solve them.

- **Network Node related latency**

  The performance over cloud is highly affected by the latency between various network nodes. The use of optimized network solutions for cloud will help in reducing this latency effects.

- **Transport protocol latency and congestion**

  The transport layer protocol (TCP) should be optimized for latency, congestion and data loss. The optimized version gives higher performance in cloud. During network congestion and packet transmission errors, TCP uses smaller window size. This may impact the network performance and reliability. So, the TCP window size be selected such a way that we can get higher performance.

- **Number of Hops**

  Generally, cloud applications are multitier. They may contain different tiers for presentation, database and business logic. The different tiers may be deployed on different machine/node. In this situation, the data may require to traverse several intermediate nodes. Cloud service providers should reduce the latency between nodes.

Network topologies in Cloud

Cloud network topologies describe the manner in which customers access the public or private cloud resources over the Internet. The cloud network has three main components.

- Front End (User Access Layer)

  Cloud service providers offer wide range of services over web based platform. Cloud customers choose their desired services. This layer helps to provide this interface.

- Compute Layer

  This layer contains various computing devices like servers, storage devices, network devices etc.

- Network Layer

  This is the underlying layer of networking. It mainly considers layer 2 and layer 3 of network topology. The layer 2 topology is independent of location. It ensures that all servers and devices have similar addresses and common protocols. This also helps in running applications in cloud infrastructure similar to traditional environment. The layer 3 network is used to transfer packets between different clouds.

3.5 CLOUD DEPLOYMENT AUTOMATION

The automation is very much important for all type of deployments in cloud like IaaS, SaaS, PaaS, etc. The automation provides uses a highly controlled environment and secure access to applications and resources. It also improves the self service mechanism of cloud.

In cloud environment, automation is possible because the cloud has different characteristics than traditional environment. The cloud has virtualized pool of resources like memory, CPU, bandwidth, storage etc. There are many pre-defined policies to allocate resources according to the workload. The cloud has also important features like automated data backup and replication.

For IaaS, the automation is required to implement various issues regarding security, authorization, resource accessibility etc. The IaaS level automation helps to automatically allocate various resources like bandwidth, memory, storage etc.

The PaaS level automation is useful to build flexible platform which offers features like, workload optimization, security, compliance, metering (usage) etc. The SaaS automation is built to improve the user experience by offering dynamic resource allocation and protection against malware.

Role of Virtualization in automation

The virtualization plays the key role in overall cloud automation. Some important cloud automation enabling features of virtualization are:

- Virtualization helps to provide on-demand and optimized resource allocation.
- Customers can add or remove virtual machine instances as per their requirements.
- Data replication helps to recover against disasters.
- VM parameters can be tuned at runtime to improve performances.
- Security policies can be combined to provide protection.
- It also helps to ensure SLA based QOS.
- It helps to save power by automatic powering on/off physical servers.

3.6 SELF SERVICE FEATURES IN CLOUD COMPUTING

On demand and self-service based resource provisioning are two main characteristics of cloud computing. The self-service mechanism enables cloud customer to request desired service from portal. Based on request, desired resources are provided to cloud customer. The self-service feature improves application performance by automatic allocation of extra resources based on workload. Some administrative tasks can also be automated in cloud like, backup and report generations.

The self-service has some drawbacks. Using portal, customers have more freedom and ability to perform more tasks over it. This may lead to increased security risks or breaking of compliance. The self-service in cloud is prime requirement but the access must be controlled by efficient access control mechanisms.

Federated Cloud Model

A federation is an integration of various smaller units, put together to complete a range of tasks. Afederation of clouds is a mechanism used by cloud providers, where it rents and integrates resources, applications, or services from various cloud vendors to meet extensive and large-scale customer needs.

For consumers, a federated cloud offers many benefits. Various applications and unlimited resources are available from a single provider. The interaction between the multiple cloud vendors is tested by service providers. Since the services come from various providers, there is no risk of vendor lock-in.Cloud resources are distributed via many providers and hence the utilization percentage is high which helps reduce the cost for consumers.

The performance is better due to caching and multiple data copies at various provider locations. These copies can be accessed by users at various locations.Data availability within a federated cloud is also very high because being replicated to multiple sites.Security and compliance in a multi-vendor federated cloud is a primary concern. The data isphysically present in various data centres across the world.

In a federated cloud, consumers must make sure that all the cloud providers audit their environments andcomply with the regulations. They must be aware of the security policies and practices followed by theproviders and know the proposed terms in the SLA with the federated and other direct providers.

Another concern in a federated cloud is the authentication of users and authorization to use variousservices. A federated cloud requires a robust identity management system. Besides authentication,the various applications from providers require different levels of permission. There are variousindustry solutions developed using SAML and other standards, which enable identity managementapplications to link users to applications from different service providers.

3.7CLOUD PERFORMANCE

Cloud performance refers to the performance of cloud applications and servers, as well as accessspeeds of network and storage devices. It is measured primarily by the round-trip response time, whichis the time interval between a user issued command and the receipt of the result from the cloud.Performance (besides service uptime) is an important part of the cloud SLA. It can be quantified interms of the maximum response time experienced by the end user. This must be a key metric for theperformance of applications and an important SLA criterion.

Cloud performance refers to the performance of cloud applications and servers, as well as accessspeeds of network and storage devices. It is measured primarily by the round-trip response time, whichis the time interval between a user issued command and the receipt of the result from the cloud.Performance (besides service uptime) is an important part of the cloud SLA. It can be quantified interms of the maximum response time experienced by the end user. This must be a key metric for theperformance of applications and an important SLA criterion. Cloud providers have bandwidth from multiple providers and it has rarely been a cloud problem. But round-trip latency and response delay is a problem.In the cloud, ability to

control latency is very minimal because resources and users are spread over different locations across the globe.

Cloud Performance Monitoring and Tuning

There are many issues related to monitoring and tuning of cloud performance. The performance of virtual machine is difficult to trace because the resources are dynamic based on workload. All aspects are not in the control of cloud customer. The division of control depends on the type of service. E.g. in PaaS, the provider controls the hardware, network, operation system, database, other configurations etc.

The consumer Controls the applications, use of resources, database instances, application-level security and authentication for users. There are also problems with selecting the right performance management tool. Any selected tool needs to be customized and configured, to a large extent, to suit the cloud environment.

Impact of Memory on Cloud Performance

In cloud computing, memory performance and utilization is fundamental for overall performance.Large database transactions require massive amounts of memory to meet the various expected performance levels. Moreover, multi-tenancy and simultaneous user tasks put a lot of demand on memory. The coordination between different cloud services to meet a particular demand requires in- memory tasks. Jobs need to be split and assembled after processing, which increases overhead cost.

Another problem in cloud relates to memory leaks.It is a situation where a user job, database, or application does not return back the temporarily—allocated memory to the operating system even after it has been cleaned up and is no longer in use. This can be due to a bug, malware, or a deliberate user job that wants to consume all memory.

3.8 LET'S SUM UP

Points to ponder

- Recommendation about cloud applications
- Cloud Computing Networking Issues and Mitigation
- Cloud deployment automation
- Self-service features in cloud computing
- Cloud performance

Unit 4: Improving Cloud Database Performance and Cloud Service Brokerage

4

Unit Structure

4.1 Learning Objectives

4.2 Introduction to cloud databases

4.3 Cloud Database performance

4.4 Cloud service brokerage

4.5 Let’s sum up

4.1 LEARNING OBJECTIVE

In this chapter we will discuss mainly about various cloud databases and their performance. We will also discuss about Cloud service brokerage

- Cloud databases definition and types
- Improving cloud database performance
- Cloud Service brokerage

4.2INTRODUCTION

Cloud databases are similar to on premise databases, but they are built and accessed "as-a-service" from a remote cloud computing platform. There are mainly two deployment models for cloud databases.

1. Virtual machine-based independent databases

   It is more similar to on premise, except it uses a cloud-based infrastructure. Organization's IT team controls and maintains the database. We can also migrate our existing database, but it is still our responsibility to manage the database.
2. Full cloud database or database as a service (DBaaS)

   This option provides the user full maintenance of the database needs in real-time, such as scalability, management, security, and availability. It is based on the fees subscription contract. It usually includes automation, backup, scaling, and health monitoring.

Difference between Cloud Database and On-premises Database

The main differences between cloud databases and on-premise databases are:

1. Structure and design: The main difference with concerns to structure and design is the physical location of the database servers; on-premise database servers are located within own company; cloud databases reside with the company deploying them and throughout their geographic zones.

2. Network: On-premise databases use a local area network (LAN), while the cloud model requires a high-speed Internet connection.

3. Performance: Sometimes cloud databases can be slower in response time, because it requires a round-trip with every interaction, while on-premise databases have an immediate response.

4. Cost: In on premises database we have to buy hardware server and also need to manage database software licence whereas, in the cloud model, there is no need for any upfront costs for buying servers, just monthly subscription fees.

5. Customization and control: The on-premise option allows for more customization and control. However, cloud options offer plenty of control and customization configurations that are enough for most needs and requirements.

Types of Cloud Database Models

Like on-premise database model, the cloud databases model is also classified into relational and non-relational:

1. Relational cloud database: Usually programmed in structured query language (SQL). They consist of organized linked fields of rows and columns. It is used for highly consistent needs, such as banking transactions since they rely on a specific database schema.
2. Non-Relational cloud databases: Also called NoSQL databases, since they do not follow the typical table model associated with SQL databases. They save data as one document regardless of the original structure. They are more suitable for naturally unstructured information like written Internet content, audio, photos, and so forth.

Benefits of Cloud Databases

Migration to cloud databases can bring you many benefits, regardless of your deployment model. The advantages cloud databasesare:

• Access databases from anywhere your team exists and at any time.
• Enhanced data processing quality when all apps and systems are in the cloud.
• Reducing costs by eliminating servers and paying only for what resources were used.
• Enabling better business application performance by taking advantage of SaaS deployment.
• Simplifying data management by combining every data source in one place.
• Lower investment barrier for start-ups and lower financial risk involved.
• Faster deployment than setting up and configuring an on-premise database solution.

• More reliable security.
• Increasing innovation and agility possibilities due to its ease of use and speed.
• Deliver your products or software to market faster.
• Reduction of energy costs; greener and environmentally friendly.

Cloud Databases Service Providers

Major cloud computing providers also provide cloud database services. Each one has its own features that differentiate it from others. Below is a list of some of the most well-known cloud database services and providers

• Amazon Web Service (AWS): Amazon is number one in the cloud computing field and the first one to enter this industry. It is also the leader in the market of DBaaS. Its service databases are Redshift, a data warehouse, and Data Pipeline.

• Oracle Database: Oracle was well known for servers and hosting before any other company, including Amazon. It offers enterprise-scale cloud databases with many technical facilities and solutions for data migration.

• Microsoft Azure Database: A cloud computing platform to create and deploy virtual machines and offers various software applications, alongside worldwide infrastructure and comprehensive solutions, security, and ecosystems.

• Google Cloud Platform: It has a no-nonsense solution approach that businesses of all sizes have adopted. Its documentation is comprehensive, which makes it easier for IT professionals to successfully deploy it. It also supports widespread open-source software compatibility.

• IBM DB2: IBM has an amazing relational database solution and migration process. It can provide advanced management, analytics, high performance, actionable insights, data availability and reliability, and other features for any transactional or warehousing workloads. All that is supported for Linux, Unix, and Windows platforms.

• MongoDB Atlas: One of the most well-known open-source NoSQL databases with unique features like powerful scaling, sharding, and automation capabilities. It also has strengths such as a strong support community, quick Installation, flexibility, and delivering models without the need of a database administrator.

• OpenStack Database: One of the important rivals for big players. It's a database for both on-premise or cloud models. Users talk about its highly customizable and easy-to-understand (and implement) architecture. Moreover, for high-end scaling capabilities it is a great solution.

4.3 Improving Cloud Database Performance

Cloud databases offer noteworthy benefits over traditionally-hosted internal databases. Moreover, cloud vendors continue to add and improve their database offerings to make it a convincing operation for enterprises. Cloud databases have higher ease of accessibility, better replication to remote datacentres alongside automation and better elasticity.

Cloud databases have their share of woes, arising from inherent cloud problems like security, data privacy, multi-tenancy, low barrier to entry for malicious users, and reliance on a third party service provider for business-critical services.

Sharding a cloud database is another technique to improve performance. it is a process of splitting alarge database into a number of smaller databases, each being hosted on a separate Server. It helps to boost the performance of applications that require frequent and large database transactions.

Sharding also helps reduce the size of the database index, thus decreasing the time needed for searches within the database. To further improve performance and availability, prov1ders offer a horizontally-scaled server environment, where it is quick and easy to bring up more virtualmachines to meet higher workloads. Besides performance, providers focus upon improving database integrity by using database profilers.

Sharding analyzes the source database for inconsistencies in index, table relationships, or data structure. By examining the data quality and utilization pattern, it is able to point out the potential problems, if any, within a database. This improves the performance of the database.

Steps to Increase Database Performance

1. Optimize Queries

Generally, database performance issues are caused by inefficient SQL queries. One of the best ways to increase database performance is to optimize SQL queries. The query optimization takes several parameters into consideration like, whether to write a join or a

subquery, whether to use EXISTS or IN etc. The best way to optimize queries is to use a database performance analysis solution. It provides internal details during query executoin.

2. Improve Indexes

In addition to queries, the other essential element of the database is the index. Indexing can increase database performance and help optimize the duration of query execution. Indexing creates a data structure that helps keep all your data organized and makes it easier to locate information. Because it's easier to find data, indexing increases the efficiency of data retrieval and speeds up the entire process, saving both you and the system time and effort.

3. Defragment Data

Data defragmentation is one of the best approaches to increasing database performance. Over time, with so much data constantly being written to and deleted from your database, the data can become fragmented. That fragmentation can slow down the data retrieval. When we defragment data, relevant data is grouped together i.e. I/O related operations will run faster.

4. Increase Memory

The efficiency of your database can suffer significantly when enough memory is not available for the database to work correctly. Even if it seems like there is a lot of memory in total, it might not be meeting the demands of your database. A good way to figure out if you need more memory is to check how many page faults a system has. When the number of faults is high, it means hosts are either running low on or completely out of available memory. Increasing memory allocation will help boost efficiency and overall performance.

5. Strengthen CPU

A better CPU translates directly into a more efficient database. We should consider upgrading to a higher-class CPU unit if we experience issues with database performance. The more powerful your CPU is, the less strain it'll have when dealing with multiple requests and applications.

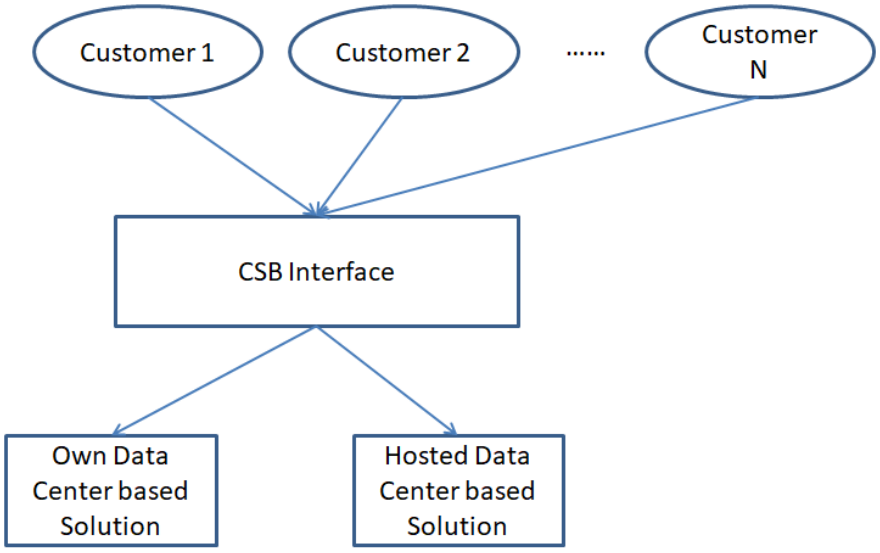
6. Review Access

Once the database hardware is working well, we need to review database access, including which applications are actually accessing your database. If one of your services or applications is suffering from poor database performance, it's important not to jump to

conclusions about which service or application is responsible for the issue. It's possible a single client is experiencing the bad performance, but it's also possible the database as a whole is having issues. We should dig into who and what is accessing the database and if it's only one service that's having an issue, drill down into its metrics to try and find the root cause.

4.4 Cloud Services Brokerage (CSB)

A Cloud Services Brokerage (CSB) is an organization that plays a role as a facilitator or inter-mediator for delivering cloud services. The CSB is usually a telecommunication or datacentre hosting service provider with a large number of customers. In the cloud provider-consumer relationship, CSBs are an optional entity involved with mediating between the two. However in the process, the CSBs also offer additional value to the provider and consumer. They help providers by relieving them of acquiring customers, billing, and enabling integrated access to multiple cloud serv1ces. The cloud consumers get integrated access to one or more cloud and value-added serv1ces such as cloud backups, SaaS, and Identity Management (IdM).



Cloud Service Broker - CSB

Cloud service brokerage provides the intermediary between cloud providers and cloud consumer that assist companies in choosing the services and offerings that best suits their needs. CSB also assist in the deployment and integration of apps across multiple clouds or provide a choice and possible cost saving function which include multiple competing services from a catalogue.

CSB also provide Value added services like migration, VM portability, and API management and normalization from cloud brokerage platforms like ComputeNext also allow end users freedom to move between platforms and keep options available at a variety of cloud vendors.
There are three primary areas in which a cloud service broker can accelerate the adoption of the cloud:
Aggregation – It enables the consumption of cloud by end users via a cloud application marketplace approved by the company
Integration – It ensures cloud applications exchange data with each other and with on-premise applications to orchestrate business processes
Customization – CSB augments cloud services with changes to data schema or enhanced security and compliance
The challenge for IT is that the cloud is relatively immature compared to on-premise enterprise software. By adding customized capabilities on top of cloud services, the enterprise can realize the benefits of cloud, while also meeting its other business objectives including data security and compliance. In particular, organizations are looking to augment the cloud and achieve the following:
Reduce risk with more robust security and compliance capabilities
Add value and visibility with analytics
Centralize functionality for audit trails and policy enforcement
Streamline the selection process of cloud services

Advantages of CSB:
1) Broader Technical Expertise
2) Lower Total Cost of Ownership – Financial Returns
3) Operational efficiencies
4) Better options in dealing with risk, compliance and governance

4.5 LET'S SUM UP

Points to ponder

- Introduction to cloud databases
- Cloud Database performance
- Cloud service brokerage

Block-4
Risk, Security, Consequences and Cost of Cloud Computing

Unit 1:Risk in Cloud Computing 1

Unit Structure

1.1. Learning Objectives

1.2. Introduction

1.3. Cloud Computing Risks And Issues

1.4. Risk Assessment And Risk Management

1.5. Let’s sum up

1.1 LEARNING OBJECTIVE

Cloud computing provides many benefits to organizations. Along with benefits, various risk factors are related to cloud based environment. In this chapter, we will discuss various risk factors like:

- Vendor Lock-in
- Resource scarcity
- Failure risks
- Network outage related risks
- Legal risks
- Risk about software and application Licensing

1.2 INTRODUCTION

Nowadays, cloud computing is used by various organizations like: banks, brokerage houses, hospitals, manufacturing companies etc. Each type of organization has its own risk level limits. They have to comply with government rules and regulations, security rules and other complacencies.

Cloud computing is multi-tenant and open for all environment which may cause various security challenges. Organizations which use cloud services must test and monitor security related threats. In some cases, it becomes very difficult and incontinent to perform security testing and monitoring. As a cloud service user, organizations need to keep the results of security tests during audit process. Another issue with cloud is about data i.e. user data is stored on remote servers. Sometimes data is replicated for security issues. It is difficult to delete all copies of data when we change service provider.

1.3 CLOUD COMPUTING RISKS AND ISSUES

In this section, various cloud related risks and issues defined by Gartner are described.

1. Regulatory Compliance

   Customers are responsible for security and integrity of their own data even in the case when the data is stored with service provider. Traditionally, service providers perform external audits and certifications. Some cloud service providers do not perform these kinds of audits or certifications whose services can only be used for some trivial functions.

2. Data Location

   Generally, cloud users do not know the physical location of their data. Cloud users ask providers if they will commit to storing and processing data in specific jurisdictions, and whether they will make a contractual commitment to obey local privacy requirements on behalf of customers.

3. Privileged User Access

   The sensitive data stored and processed outside the organization creates a risk about access of data. The outsourced services may overlook the physical, logical or personal controls of data. Cloud users must get information about the people who manage data like, information about hiring of privileged administrators and their controls.

4. Data segregation

   Generally, data in cloud is shared among customers. Cloud users must know the mechanism used to segregate data at rest. Cloud provider should provide details about used encryption algorithm. Faulty encryption may make data unusable and may also affect availability.

5. Recovery

   Cloud service provider should provide details about the data recovery in case of any disaster situation. If service provider does not maintain replicas across multiple sites, the service may not work in case of failure. Cloud users should ask service providers about the complete restoration process and the amount of time required for restoration.

6. Long term visibility

   The cloud service provider may be acquired by larger company in future. Cloud users must make sure that in such cases, data is available. Cloud users should ask service providers that how they would get your data back in the original format in case of any acquisition.

7. Investigative support

   It is very difficult to trace illegal activities in cloud environment. Cloud services are difficult to investigate because logging and data for multiple customers may be located at same place and may also be spread across different locations. Cloud service provider should provide mechanism of investigation in case of any such case.

1.4RISK ASSESSMENT AND RISK MANAGEMENT

Cloud users must do in depth analysis about risks involved in cloud environment. This will help organization to identify possible vulnerabilities and threats. Based on this analysis, one can take measures to avoid risks. The threats may cause issues like financial losses, reduced employ productivity, loss of customer loyalty etc.

Cloud users can use several mechanisms to reduce or mitigate risks. Cloud users should have updated copies of their data backup in their network. Cloud users should have plans to quickly switch between various service providers in case of any outages. The risk management includes:

- Risk identification
- Risk analysis and evaluation
- Identify risk mitigation steps
- Deploy identified steps
- Monitor and evaluate effectiveness of steps taken

➢ **Risk about vendor lock-in**

There are many steps involved when any cloud user migrates from one service provider to another. The risk of migration is always associated with cloud users. Before using any cloud services, cloud users should refer SLA (Service Level

Agreement) documents. Depending upon the type of cloud service, there are several lock-in related issues and mitigation steps.

- Infrastructure as a Service (IaaS)
  Different service providers use different mechanisms for data storage and data access. These differences make migration process difficult. To mitigate this problem, cloud users should take back up of VMs along with data and store them on local devices. Also verify that backed up VMs can be restored later.
- Platform as a Service (PaaS)

  Many times the PaaS environments are unique and proprietary to particular service provider. The standardization process is now well established. Cloud users should make sure that the development environment provided by service provider is supported by other PaaS providers. We should also use standard APIs which make easy migration among different PaaS providers.

- Software as a Service (SaaS)
  SaaS providers offer general purpose environment which can be used to develop user specific tools on particular cloud. It may happen that users data is stored in proprietary format which is not supported by other service provides. To avoid this situation, cloud users should make sure that APIs are compatible with other public clouds. One should be able to export data in standard format which can be further used on other platforms.

➢ **Risk of Not Meeting Regulatory Compliance**

Cloud service providers must certify that their service platforms are complied with regulations. The certificate form audit authorities can be used to provide assurance to cloud users. These certificates will also help cloud user during their own audits. Cloud users must ensure that their providers have obtained acceptable certificates from authorities. Some examples are

- If open source based virtualization platform is used, it may not have compliance about protection or availability.
- If the data is stored outside the country, it may violate the requirements of banks

- **Risk of Loss of Control**

In cloud based environment, infrastructure, network, security and other hardware resources are controlled by service provider. Due to this, SLA may not be much relevant to end users. Similarly, responsibilities of cloud users and service providers are different depending upon the type of service (IaaS, PaaS or SaaS). It may also happen that some of the service providers outsource some part of their datacenter services. In all these situations, the cloud users lose their control over their resources.

- **Risk of resource scarcity or poor provisioning**

One of the major benefits of cloud based environment is on demand provisioning of computing resources. In some situations, the public cloud service users may have to compete for fixed set of available resources like computing, storage or network. Sometimes, the supply of resources is insufficient for the demand. The cloud service provider may have deployed less number of physical resources. Cloud service providers use some algorithms for dynamic resource scheduling which may misbehave in some cases. There are always chances of hardware failures at service provider level. In all these circumstances, there is a risk about scarcity and poor provisioning of cloud resources.

- **Risk in Multi-Tenant Environment**

Cloud computing is multi-tenant environment which allows multiple users to access and share same physical resources. The concept of virtualization is utilized for resource sharing. One tenant may access, read, modify or delete the data associated with another tenant. Due to this situation, there is risk associated about data. This may affect the reputation of service provider and cloud users may lose their confidence.

- **Risk of Failure**

A cloud service provider maintains huge amount of hardware resources in the form of datacenter. These resources are pooled and shared among cloud users. The technology of these pooled hardware resources changes time to time. If these physical resources are not updated timely, it may affect the overall service. Along

with that, if the older physical resources (infrastructure) are not updated, the risk of failure increases. This situation impacts cloud customers.

- **Risk of failure of supply chain**

Cloud service providers acquire services from other services providers. These acquired services may include network bandwidth, monitoring services, security services etc. The cloud service provider has to follow the rules of third party service providers. Any failure or change in services of third party may adversely affect the services of cloud service provider. The failure at one provider level may cause service degradation or outage of other service providers.

- **Risk of inadequate SLA**

The Service Level Agreement (SLA) provides a mechanism by which the service users and service providers are bound to certain conditions. SLA may include details about service availability, performance, security etc. The SLA should have features so that all conditions regarding user and compliance requirements can be mentioned. The SLA document helps to set responsibilities of both service users and providers. The SLA violation may cause loss to both service provider and service user. In case of violation, the penalty is caused to service provider. Similarly, the violation may cause poor QOS (Quality of Service) to users.

Sometimes, the structure of SLA is inadequate to specify all conditions by service users. In certain situations like if the service provider is acquired by another service provider, than the SLA document should be revised which may lead to risk of incompliance.

- **Risk of malware and Internet attacks**

The cloud computing is an Internet based and self-provisioning environment. Anybody can open account form the portal or interface of service provider. This gives chance to malicious users to connect with cloud environment. They may also launch attacks to disable particular service or whole cloud function. One of the most common attacks in cloud environment is:

- Distributed Denial of Service (DDoS) attack

  In basic DOS attack, the malicious user generate dummy workload load (requests) on services such that the performance of service degrades. In some cases, the normal users are not able to use the service. The cloud environment is distributed in which attacker may use DOS attack from different locations to affect several services. It is very difficult to identify the difference between legitimate traffic and dummy traffic.

➢ **Risk of Cloud Resource Management**

Cloud computing is very dynamic environment in which users enter and exit in very dynamic nature. The efficient and effective management of shared resources will help cloud service providers to enhance QOS along with serving more user requests. The improper management of resources may be caused by:

- Unauthenticated access of cloud resources
- Virus of Malware may lead to heavy network traffic (activity)
- The compute, storage or network related resources may be unnecessarily blocked by cloud users.

➢ **Risk of physical infrastructure and network outages**

Some of the risks associated with the physical infrastructure of cloud are:

- Theft of physical resources owned by cloud users and kept at provider's datacenter.
- Damage to physical resources by a person who is able to entre datacentre without proper authentications.
- The possibilities of power disruptions or outages.
- The improper air conditioning of datacentre may lead to rise in temperature of servers and other equipments which affects overall functioning of cloud.
- Sometimes, malicious employees or intruders having privileged access may damage physical infrastructure

The outage is a situation in which a specific service is affected temporarily or permanently. In case of network outage, cloud users may not be able to access the latest of copy of data. It may also create inconsistency among replicated data.

- **Software, Application Licensing and Legal (Legislation) Risks**

The licensing of software, applications, development tools, middleware, database and operating systems is a major issue in cloud environment. Generally, licenses are categorized as below:

- User based license: It controls the number of users who can use the software or application. E.g. If we have 20 concurrent user licence with 200 named users means maximum 200 users can use the software or application but concurrently on 20 can access.
- Device based license: In this kind of license mechanism, applications / software are bound with physical hardware. This kind of software can be used by any users who have access to particular hardware machine.
- Organization wide license: This category of license allows all the users of particular organization to access software or application within the organization.

In cloud based environment, license related risks are more because

a. Cloud service provider may add or remove servers dynamically to manage user workloads
b. It is difficult to check the user count over cloud
c. If cloud user obtains license considering peak workload, the license might be unused during most of the times.

Organizations that use cloud platform are from different domains like banking, pharmaceutical, healthcare, automobile etc. Their data resides in cloud datacenter. The data is very sensitive and has to obey certain regulations regarding security, access, location, recovery, backup etc. Generally, cloud service providers are complied with these regulations. The compliance requirements must be mentioned in SLA documents. Cloud users should also get copy of compliance certificates.

1.5 LET'S SUM UP

Points to ponder

- Vendor Lock-in
- Resource scarcity
- Failure risks
- Network outage related risks
- Legal risks
- Risk about software and application Licensing

Unit 2:Data Security in the Cloud 2

Unit Structure

2.1. Learning Objectives

2.2. Introduction

2.3. Cloud Data Challenges

2.4. Let’s sum up

2.1 LEARNING OBJECTIVE

Cloud user's data is stored at remote location which is not known to users. This may lead to several data security issues. In this chapter, we will discuss about:

- Data Redundancy
- Data Backup and Recovery
- Data fragmentation and reliability
- Data integrity, confidentiality and migration
- Cloud firewall and virtual firewalls

2.2 INTRODUCTION

The data stored in cloud environment has several threats like:

- Data Security

  Data is stored at remote locations. For better security of data, encryption mechanisms can be used to achieve security.

- Data Availability and Integrity

  The problem in hardware or software at cloud service provider level may lead to unavailability of data. Multiple replicas at different geographical location can be maintained to achieve higher availability in case of any disaster or failure situations. Once, multiple copies are maintained, there is also a risk of data inconsistency is associated.

- Flexibility

  In a multi-tenant environment, some applications may have higher utilization which may affect the performance of other users. To avoid this problem, storage access speed should be decided.

- Performance

  In cloud data is stored at remote locations. Due to this, the data access may be affected by network latency. The performance may further degrade due to synchronous operations of read and write. To achieve better performance, providers should employ techniques like data caching (in memory data).

- Price and Complexity

  The price of data storage space impacts the cloud service users. The price should be low. The storage hardware at cloud service provider side may be heterogeneous which makes it very difficult and complex to manage the devices.

2.3 CLOUD DATA CHALLENGES

In this section, various challenges associated with cloud data are discussed along with some of the mitigation steps.

1. Data Redundancy

   Data stored in cloud environment is accessed by multiple users concurrently. The data is stored on heterogeneous hardware environment. The data is stored in multiple copies across different locations in synchronous manner. The data replication system should have knowledge about data location, latency, workload, backup and monitoring of data. To achieve proper data redundancy, following points should be checked.

   - Data must be consistent across all replicas
   - Different steps should be taken to improve data replication and access.
   - Load balancing about data access requests should be considered
   - Data should be recoverable in any circumstances

2. Data backup

   There are certain issues regarding data backup which are listed here:

   - When cloud users take backup of data and stores in local infrastructure, it may cause bandwidth expenses
   - Users need to place backup at safe locations and they should check the media integrity at regular intervals.
   - The backed up data must be protected against malicious users and attacks.
   - The data backup strategy should consider the restoration process before taking backup.

3. Disaster recovery

   Cloud users should consider disaster recovery mechanisms available at service provider before choosing any service provider. The strong disaster recovery mechanism provides benefits like: cost saving, scalability, quick restore, higher availability etc. Certain issues regarding disaster recovery are:

   - Initial data copy- When data size is too large, it is very difficult to make first copy of data over WAN. Cloud users should adopt manual process of data copy using tapes or hard disks.
   - Operating system support- Cloud service providers support most common operating systems like windows, Linux etc. Some of the older operating systems may not be supported by disaster recovery solutions.
   - Less bandwidth- Many disaster recovery service providers create backups using incremental approach instead of taking full backup. This saves the bandwidth, time and cost.
   - Financial Considerations- When data is small in size, it is advantageous to utilize disaster recovery mechanisms offered by cloud service providers. But, if data size is too large, the cloud users should have own mechanisms for disaster recover (it will be cost effective).

4. Data Reliability and Fragmentation

   The reliability of cloud services is very important because,

   - Cloud services are provided over WAN in which multiple vendors are engaged
   - User base is tool large
   - Accessibility and availability of cloud services is very high
   - Cloud services are offered over heterogeneous hardware and software

   As the numbers of cloud users are very high, they all might be working on different parts of data simultaneously. The data may be fragmented and stored over different locations. The data fragmentation forces us to

keep track of different parts of files. The service providers should have mechanisms to reduce data fragmentation and inconsistency.

5. Data Integration

   Some of the factors which make data integration difficult are:

   - Content distribution

     The user data is stored in different files which are distributed and fragmented across several datacentres.

   - Data exchange

     The data stored over cloud is utilized by applications running on other public or private clouds. These may cause compatibility issues among various involved cloud parties.

   - Data change rate

     The data over cloud is accessed by many users concurrently. The user actions may modify the data. The rapid change is in data is challenge for integrity and replication management.

   - Connectivity

     Cloud data can be accessed by users from any location across the world. The basic requirement is Internet connectivity and bandwidth. The connectivity also affects the latency.

   - Control over distribution

     The data control is shared among the cloud service provider and cloud user. This may create integration issues.

6. Data Transformation and Migration

   The data stored over cloud is accessed by many applications. Each application may have its own requirement about the format and structure of data. This may cause data to be converted in different formats as per the requirements of individual application. This kind of data transformation requirements may cause several types of issues.

   - Runtime issues

     Sometimes the stored data and the data required or generate by another cloud application are not compatible with each other.

   - Redundancy issues

The data transformation may create multiples copies of same data in different format. The location tracking and consistency management are major challenges.

- Implementation issues- Data transformation may be complex and expensive. To make it simple process, various standards should be adopted.

Cloud users may need to change service providers or may need to migrate data like user login, profiles and other user data to other cloud service providers. Cloud service providers should have proper format and process to make data migration easy. Some of the data migration challenges in cloud are listed below.

- Concerns about liability

  Cloud providers have maximum data value for damage claims in the SLA. These values may be less than the data value or the efforts required to fix data loss.
- Connectivity issues

  The data in cloud is accessed over the WAN (Internet). There may be faulty network links.
- Compliance issues

  Cloud service providers must comply with various regulations and legal requirements. Some of the examples of compliance are:

  FISMA – Federal Information Security Management Act
  HIPAA- Health Insurance Portability and Accountability

7. Data Security risks

Data Confidentiality and Encryption

Organizations store their data in cloud environment. These data may include financial data, medical data or any other important data. The confidentiality refers to protection of data from unintended access by users or applications. The data must be stored along with that the privacy and confidentiality must be maintained. Several encryption

based mechanisms are adopted to achieve confidentiality. The encrypted data is called cipher text. Generally, two methods are used for encryption.

- Symmetric key encryption

  In this mechanism, same key is used for encryption and decryption of data.

- Asymmetric encryption

  This mechanism is also known as public key cryptography. In this approach, two keys are maintained: private key and public key. Generally, private key is used to encrypt data and it can be decrypted using public key. The private key is kept secret at sender side and public key is shared among receivers.

Key Protection and Management

The data stored in cloud data centre is in encrypted format. The encryption mechanism can be symmetric or asymmetric. In both cases, the secure exchange of keys among stake holders is quite important. If key management mechanism is compromised, it will lead to the issues regarding data privacy and confidentiality. Generally keys can be exchanged using public key cryptography mechanisms.

Data availability and Integrity

The availability of the data refers to the idea that the data stored in cloud must be available at any location and at any time. There can be various failures or natural disasters which may affect the overall availability of data. Several mechanisms like data redundancy are used to increase availability of the data.

As the redundancy techniques are adopted to achieve higher availability, it may lead to another issue regarding the integrity of data. The data stored across all replicas must be consistent. If user changes a small amount of data, the same must be propagated across all replicas.

Management Interface for Cloud data

The Cloud Data Management Interface (CDMI) is a SNIA standard which specifies a protocol for self-provisioning, administering and accessing cloud storage. In other words we can say that, the Cloud Data Management Interface defines a functional interface by which applications will create, retrieve, update and delete data elements in the Cloud.

The client should be able to discover the capabilities of the cloud storage offerings. Client uses this interface to manage data that is placed in cloud. In addition, metadata can also be set about stored data elements through this interface.

This interface is also used by administrative and management applications to manage computing resource, accounts, security access and monitoring/billing information.

If CDMI is managed properly, users can freely move data from one cloud to another cloud vendor. CDMI standard can be applied in private, public or hybrid cloud.

Cloud Storage Gateway

Cloud storage gateway (CSG) is an appliance running in customer's premises. It provides data protection by encrypting and compressing data. CSG intercepts all the data and I/O between customer data centre and public clouds. The CSG manages all I/O between the users and the cloud storage service providers. CSGs have local cache to store data temporarily. CSG can improve performance by,

- It accelerate the I/O process using cache
- The files which are to be on cloud are first stored in CSG cache
- At regular interval, cached data is pushed in the cloud storage
- When data is fetched from storage, it is stored in cache

CSG provides us various features to change caching algorithms, pre-fetching algorithms, cache time intervals etc.

Cloud Firewall and Virtual Firewall

Cloud firewall is used to work with other cloud based applications to provide security. The cloud firewall systems differ from traditional firewall systems by following ways.

- Scalability: Firewall should scale as per the changing bandwidth of the application. Firewall should also support any changes in hardware.
- Availability: Cloud firewall offers higher availability using redundant infrastructure. It also takes care about backup mechanisms in case of failures.

Virtual firewall (VF) is a service which runs under virtualized environment. Like physical firewall, it also provides services like packet filtering, monitoring. Virtual firewall reduces the cost on physical firewall. VF can operate in two modes.

- Bridge mode- Firewall acts as physical firewall and works with physical and virtual switch to know traffic related information.
- Hypervisor mode- Firewall resides in virtualized environment.

2.4 LET'S SUM UP

Points to ponder

- Data Redundancy
- Data Backup and Recovery
- Data fragmentation and reliability
- Data integrity, confidentiality and migration
- Cloud firewall and virtual firewalls

Unit 3:Application Security in the Cloud

3

Unit Structure

3.1. Learning Objectives

3.2. Introduction

3.3. Cloud Application Software Development Lifecycle (SDLC)

3.4. Application Security

3.5. Let’s sum up

3.1 LEARNING OBJECTIVE

In this chapter, we have discussed about

- Cloud application software lifecycle
- Application security in the cloud
- Application protection

3.2 INTRODUCTION

Cloud application security is a series of defined policies, processes, controls, that govern all information exchanges in collaborative cloud environments. Protecting cloud-based applications with network and infrastructure security is not enough. Most organizations use application-level security measures. Cloud application security measures are usually implemented during software development and after applications are deployed.

Major Cloud Application Security Issues

Cloud application security issues are threats to which a cloud-based application is exposed. Some of them are:

1. Unauthorized access to application functionality or data

2. Exposed application services due to wrong configurations

3. Hijacking of user accounts due to poor encryption & identity management

4. Data leakage from insecure APIs or other infrastructure endpoints

Best Practices for Cloud application security

1. IDENTITY ACCESS MANAGEMENT
Applications in cloud should have secure measures like identity access management (IAM). IAM ensures every user is authenticated and can only access authorized data and application functionality. A holistic approach to IAM can protect cloud applications and improve the overall security posture of an organization.

2. ENCRYPTION
Implementing encryption in the right areas optimizes application performance while protecting sensitive data. In general, the three types of data encryption should be

considered- encryption in transit, encryption at rest, and encryption in use. Applying encryption for data in each of these stages can reduce the risk of cloud applications leaking sensitive data. This is crucial for achieving a high level of security and privacy that protects organizations from intellectual property theft, reputational damage, and loss of revenue.

Encryption in transit protects data as it's transmitted between cloud systems or to end-users. This includes encrypting communication between two services, whether they're internal or external, so that data cannot be intercepted by unauthorized third parties.

Encryption at rest ensures data cannot be read by unauthorized users while it is stored in the cloud. This can include multiple layers of encryption at the hardware, file, and database levels to fully protect sensitive application data from data breaches.

Encryption in use is aimed at protecting data that is currently being processed, which is often the most vulnerable data state. Keeping data in use safe involves limiting access beforehand using IAM, role based access control, digital rights protection, and more.

3. THREAT MONITORING

After applications are deployed to the cloud, it is important to continuously monitor for cyber threats in real-time. As the application security threat landscape is constantly evolving, leveraging threat intelligence data is crucial for staying ahead of malicious actors. This enables development teams to find and remediate cloud application security threats before they impact end-users.

4. DATA PRIVACY & COMPLIANCE

Along with application security, data privacy, and compliance are crucial for protecting end-users of cloud native applications. For example, compliance with

GDPR requires careful vetting of open source components, which are frequently used to speed up cloud native application development.

5. AUTOMATED SECURITY TESTING

A key part of DevSecOps is integrating automated security testing directly into the development process. By automatically scanning for vulnerabilities throughout the continuous integration and continuous delivery (CI/CD) process, development teams can ensure every new software build is secure before deploying to the cloud. This includes not only the code and open source libraries that applications rely on, but the container images and infrastructure configurations they're using for cloud deployments.

Cloud uses very strong mechanisms for the data security while data is at rest or at transit. There are two main issues regarding the security of cloud applications.

1. Cloud API and tools for development

   Cloud service providers offer many sets of APIs and application development tools to their customers for application development. Using these cloud users build and integrate their required services. These API and tools should provide protection in virtual environment. Along with this, it should be compatible with other platforms.

2. Application architecture

   Cloud applications greatly depend on other systems and services like identity management, database and encryption systems etc. The main reason behind dependency is the multi-tenancy. The applications hosted in cloud may belong to other cloud providers. The application deployment model is different because it does not follow one server per application approach.

3.3 CLOUD APPLICATION SOFTWARE DEVELOPMENT LIFECYCLE (SDLC)

Security is considered as a one of the important concern during the SDLC. When application is moved in the cloud, the security remains the major challenge. Along

with this, cloud applications have several other major issues. Cloud based applications have different environments for development and deployment. Cloud based application SDLC should have well defined trust model between these two environments. Whereas in conventional application SDLC, the trust can be ensured by isolating hosts and networks from outside infrastructure. The figure 8 shows the SDLC model for traditional internal applications.

The SDLC of cloud based application largely depends on the deployment model.

- **Application Security in IaaS**

The conventional applications ensure data security by several internal controls. The cloud based applications should consider security by different means. In IaaS based service model, service providers create virtual machines (VMs).

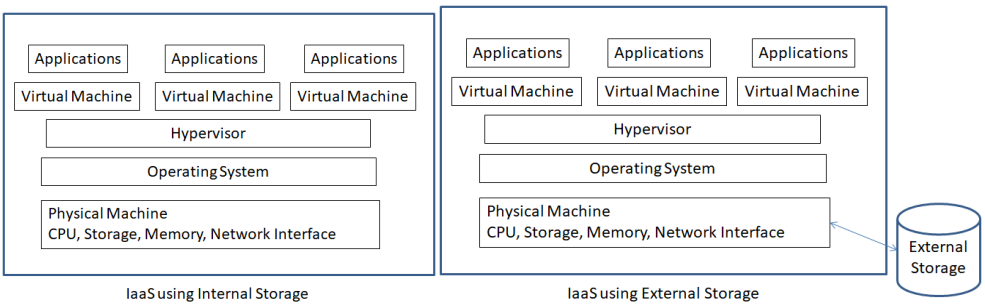


Figure-8SDLC model for traditional internal applications

When application runs under IaaS environment, it may happen that its development and initial testing may be done on internal infrastructure. But, some later SDLC phases like deployment, testing and maintenance are done on cloud infrastructure. The figure 9 show the various SDLC phases.

The cloud service provider (IaaS providers) offer set of tools to deal with security issues. These are discussed below.

- dWAF

  It facilitates to write set of rules which are applied to web applications. Based on these rules, some communication packets can be accepted, dropped by considering port number, source and destination address, protocol and other parameters.
- HIDS (Host based Intrusion Detection System)

  These systems help us to monitor and report intruders on IaaS host.

- HIPS(Host based Intrusion Prevention System)

  This system monitors each and every host of IaaS for any suspicious activities and takes steps to stop such actions. It drops packets about malicious actions. Sometimes connections are reset or complete traffic is blocked.
- Scanning programs for application security

  These programs are used to record and inspect requests – responses of web based applications. These are reviewed by operators. Using these logs, security experts can detect vulnerabilities in application architecture and code.
- Programs for analysis of source code

  These programs are designed to identify security flaws in source code. These tools are used by experts to analyze compiled codes to identify security risks.

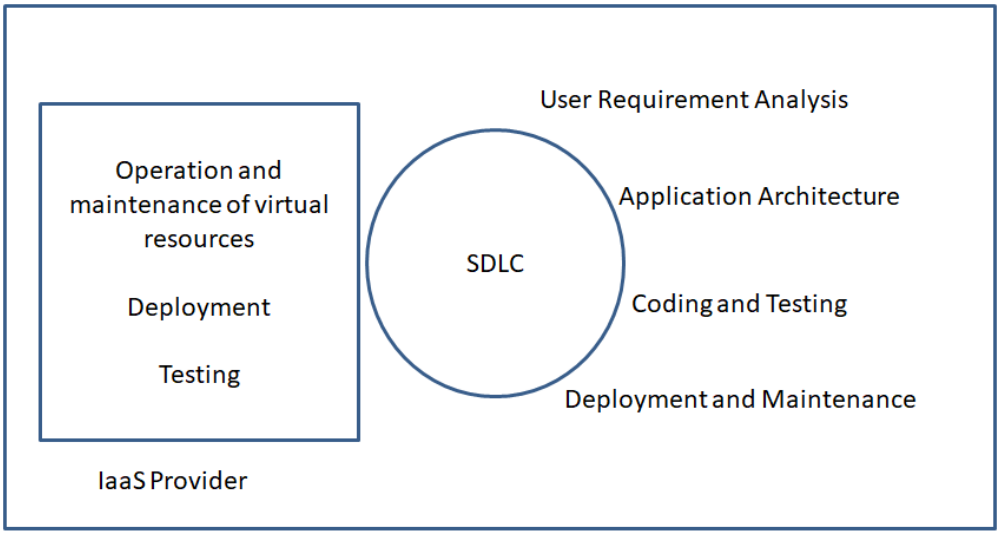


Figure-9 SDLC Phases

As from the figure 9 we can observe that, requirement analysis, coding and initial testing are done on internal customer environment. Once these phases are over, application is deployed on cloud where again some test cases are evaluated. Over IaaS the application must have tight access policy to maintain confidentiality. The security configurations in IaaS environment are similar to the internal architecture. Some security mechanisms which are in-built in intranet environment should be configured explicitly in cloud environment.

The IaaS providers offer services in the form of VMIs (Virtual Machines Images). A virtual machine image is a basic unit of deployment in IaaS environment with its own

OS (Operating System) and other configurations. Cloud users make use VM images provided by IaaS providers or by some third parties. In both cases, cloud users should perform more security configurations in VM. For IaaS, users should use hardened OS image which is used to build DMZ (Demilitarized zone) web servers. For IaaS hosts, we should install some other required OS modules. If some OS services are not required by host, these must be removed. A smaller application software stack provides very small attack surface for attackers and viruses. It makes security management very easy and efficient. The need of OS patch updates is also reduced.

Another IaaS application threat is regarding inter host communications. These communications might be done over un-trusted networks which might be used by many other users. In this situation, cloud applications should use digitally signed communication messages. If the message is digitally signed, the application can verify the origin of messages. Digitally signed message are more secure because they can be accepted or rejected based on the origin. The figure 10 shows the mechanism of digitally signed messages.

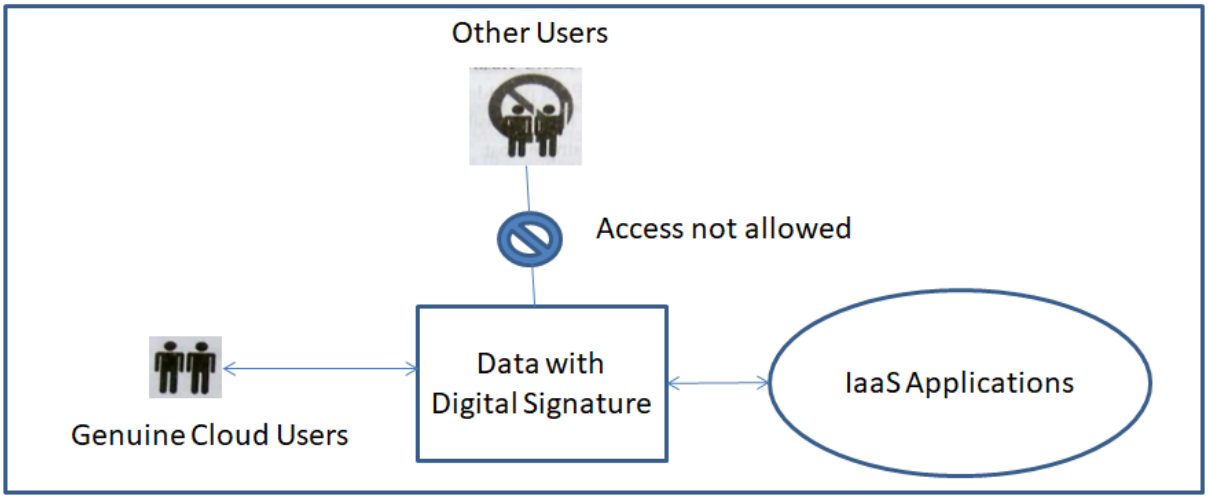


Figure-10Mechanism of digitally signed messages

Another major issue in IaaS application security is about key management. These keys can be used for authentication or encryption. These keys are to be exchanged among various users and hosts.

The data storage is again a big challenge for IaaS applications. The internal applications use the storage provided by other trusted providers. In case of cloud,

storage media is shared among customers so encryption is must for data (during rest or transit). IaaS applications must have data filtering and masking mechanisms.

- **Application Security in PaaS environment**

PaaS service providers offer hardware infrastructure, application development blocks, compilers and runtime environments. These can be used similar to the internal application development but in cloud users need to code extra security mechanisms. The security challenges in PaaS are listed below.

- Data distribution on multiple servers
  PaaS providers offer a development environment which may be spread across group of clustered hosts. The data owner cannot correctly identify the host on which data is stored. The absence of single host makes data security difficult task.
- Data distribution over different geographic locations
  PaaS providers store user data on different geographic locations to provide protection against disaster situations. It is challenging to secure data which is spread across different locations. Each datacentre might have separate copy of data and backups.
- Privileged permissions
  Application developers (programmers) have fully privileged environment during development process. During PaaS based development, developers may grant permissions to other persons. They may forget to revoke these permissions back before the actual deployment.
- Risk due to open TCP ports
  PaaS uses distributed file system mechanisms like Hadoop. Along with these, several other frameworks are also used. Each uses different ports for communication. These open ports may provide surface to DOS attackers.

It is very important to consider all above mentioned issues during application development over PaaS platforms. Once the problems or issues are identified, it is the responsibility of user to solve these issues. The PaaS vendor may offer certain security solutions along with platforms.

- **How to protect PaaS platforms**

Some of the most common ways to provide application protection over PaaS platform are,

- Use good tools to identify vulnerabilities in applications. OWSAP is a nonprofit organization to provide tools and best practices over PaaS platforms.
- All application activities should be logged and data should be stored and transferred in encrypted format. The logs should be checked on regular basis to detect security related threats.
- Generally, all API require an application key over PaaS platform. There should be a way to securely maintain these keys and other application credentials.
- One should use secure communication protocols while application development over PaaS. If web services are developed or used, they should also follow secure protocol standards. OASIS (Organization for Advancement of Structured Information Standards) is a nonprofit organization, who establishes secure communication standards. Where ever possible, Cloud applications should use SSL (Secure Socket Layer) based protocols.
- Cloud application developer should be trained about secure application architecture and best coding practices.

The figure 11shows the roles and responsibilities of PaaS service providers and users about applications.

The cloud users should adopt specific security tools and standards to ensure security during SDLC process. All software architects, developers and testers must be aware about the security features offered by PaaS provider. Some PaaS providers give details about best practices and they also give training to users about their platform.

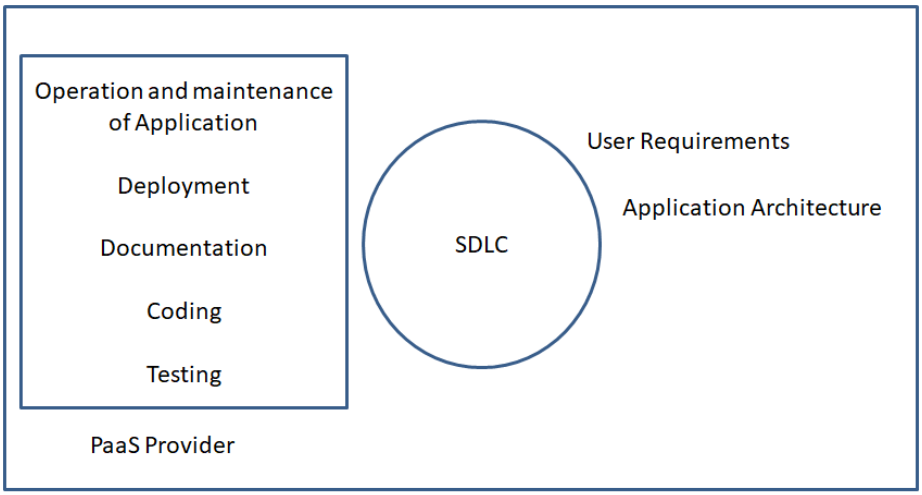


Figure-11 SDLC & PaaS service providers

3.5 LET'S SUM UP

Points to ponder

- Cloud application software lifecycle
- Application security in the cloud
- Application protection

Unit 4:Costs in Cloud Computing

4

Unit Structure

4.1 Learning Objectives

4.2 Introduction

4.3 Costs in cloud computing

4.4 Let’s sum up

4.1 LEARNING OBJECTIVES

In this chapter, we have discussed about

- TCO (Total Cost of Ownership) in Cloud environment
- Direct and Indirect costs
- Charge back models in Cloud
- Billable items
- Strategic Flexibilities in Cloud

4.2 TOTAL COST OF OWNERSHIP (TCO)

The TCO for a product or service refers to the cost calculation that adds the direct and indirect expenditure over a period of time. It helps to find final cost estimate that includes all cloud related expenses. The TCO also considers the incremental expenses and ROI.

There are certain guidelines for providing a useful insight for TCO analysis:

1. Identify various components of Cloud Cost

The bandwidth, reserved and on-demand server and storage resources, backup, use of storage, permanent IP addresses required for applications, etc. are considered. e.g. if you have reserved permanent IP addresses but are not using those, there would be a charge for not using it.

2. Identify the Combination of Cloud Services

Some applications process large amounts of data and use lots of storage space. Other applications may use more computing services. It is important to list the utilization category for each expense.

3. Identify the Variations in Utilization

If the utilization levels change a lot, the cost bracket per unit resource will vary, which will affect TCO. You must understand the way resources are used and include that in the TCO calculation.

Direct and Indirect Cloud Costs

All cost components related to a cloud can be divided into two categories: direct cost and indirect cost.

- Direct Costs

These components are those that are entirely related to a specific cloud service. For example the cost of cloud resources for a certain application, cost of the setup of a cloud service, etc.

- Indirect Costs

These components are those that are not entirely used by a certain services but supported or shared by multiple services. E.g. cost of a physical server in a private cloud that is used to host multiple cloud services. It is difficult to forecast the total cost and cloud services due to varying amounts of resource utilization. The cost of each delivery model is described as follows:

1. The cost of SaaS

The users pay a fee per billing period to the SaaS vendor. In this case, it is difficult to customize the service for different customers.

2. The cost of PaaS

In PaaS Cloud, the cost per billing period depends upon the number of developers, testers and application users and the amount of utilized resources.

3. The cost of IaaS

Like PaaS, the total cost depends on the amount of utilized resources such as compute, storage, bandwidth, etc. The consumer organizations have to incur costs for their internal IT infrastructure, management teams and monitoring services. These have to integrate with the public provider's IaaS cloud and NOC services. The TCO Includes the cost for internal IT and for public cloud.

Cost Allocations in Cloud

A cloud can provide various billable services in the form of IaaS (IT infrastructure, computer servers and storage), PaaS (an application development and runtime platform) or SaaS (subscription-based software). All such IT services delivered as cloud serv1ces typically have the following characteristics:

- A pay-per-use model so that customer costs are based on actual resource utilization.
- Minimal or no capital, one-time expenses.

- Elasticity and dynamical allocation of resources that is consumed more or less as and when needed.

Pay-per-use has historically been a foundation of several consumer businesses, electricity and phone services. It forces users to be keenly conscious of the cost of the consumption of resources, since what they pay each month is directly proportional to what they use. The keen awareness gives more efficient and selective usage and results in less overall wastage and lower cost.

In business computing, pay per use has progressively extended its acceptance The IT team and user community strives to reduce costs across infrastructure and applications. With cloud computing, pay-per use has become necessary in a multi-tenant environment. It is easy to implement and provides a range of benefits beyond just reducing expenses and cost management.

Chargeback Models for Allocation of Direct and Indirect Cost

In cloud computing, Chargeback is a term for distributing and recovering the cloud services cost from amongst consumers. In other words, consumers pay for what they use. Enterprises are increasingly using public and private clouds. In case of public clouds, the provider distributes the total cost among consumers. In case of private clouds, the IT team distributes the cost among the departments and business units.

The provider or IT team uses a charge back model to allocate the costs. The use of a Chargeback model based on the actual resource usage is a central component of the cloud architecture. A chargeback model in the cloud provides various benefits, such as:

- Correlating utilization back to c1oud consumers or corporate departments.
- Providing visibility and transparency into resource utilization and bill.
- Facilitating capacity planning, forecasting, and budgeting.
- Encouraging the use of emerging technologies, which might be priced lower than other services as an incentive. For example, thin provisioning of storage will help reduce cost.
- Providing a method for managing demand of cloud resources.

- Enabling Cloud users to know their compute footprint, thus encouraging lower consumption.
- Bringing more efficiency by increasing percentage utilization.
- Enabling enterprise IT to allocate their costs to their stakeholder business units in an easy-to-understand manner.

According to chargeback model in cloud computing, it helps users to bill the cost to internal consumers of cloud services. It means that not all expenditure falls under one department. Instead, each business unit becomes responsible for its consumption.

The effective chargeback depends on four main factors. We should consider these factors for effective implementation of chargeback in cloud.

These factors are:

- Accuracy of the chargeback model. For a chargeback model to be effective, it has to quantify the right price paid by a given business unit with utmost clarity.
- Transparency of your chargeback model. A good understanding of the cloud chargeback model helps implement effective chargeback model.
- Control on cloud consumption. Business units with a deeper control of cloud resources will embrace your chargeback model, making it more effective.
- Cost of the chargeback model - Some chargeback models are more costly to implement than others.

There are several issues which may make the overall chargeback model ineffective. These issues are

- Dislikes about the chargeback models often leads to a dispute.
- The cost of the chargeback can make it expensive to implement.
- The lack of knowledge, transparency, and understanding within business units makes it difficult to understand a chargeback.

Chargeback Methodology

- The methodology for creating a Chargeback model is straight forward. To develop chargeback model, enterprises should:
- 1. Trace and document all relevant costs.
- 2. Analyze the costs.

- 3. Identify the billable items, and for each identify the smallest unit available as a service. A unit could be, a unit of compute power, portion of datacentre or a combination of compute, network, and storage resources. This small unit becomes the atomic unit that drives the gathering, billing for customer, and reporting.
- 4. Define a pricing strategy by choosing pricing options for each billable item
- 5 Identify, integrate, and implement the tools required to collect billing data and to mediate the data into a billing solution.
- The various components of a Chargeback system are:
- a. Cost
- b. Billable Items
- c. Atomic Unit
- d. Pricing Strategy
- e. Chargeback Tools and Solution

Chargeback involves assembling and correlating billing data records. Later a pricing model is used to generate customer bills. To be able to determine a pricing model that provides business value, one must know the direct and indirect costs of providing cloud services. For example, for an IaaS service, the cost can be modeled as a fully loaded cost per physical server or per virtual machine.

Billable Items

Billable resources are items for which customers are charged. Customers will be able to purchase these items using the cloud self-service portal.

Billable Item	Resource
Virtual Machines	CPU, server memory, storage capacity, data capacity, network bandwidth etc.
Network related Services	Load balancer, Firewall
Data Services	Data encryption, data compression, backup, replication, de-duplication, data availability, data redundancy
WAN Services	VPN Connectivity, WAN Optimization, Data caching services
Security related	Isolation level, Compliance level, Certification, installed

Services	firewall, OS hardening etc.

Case Study of billable items in Alibabacloud.com

Ref:https://www.alibabacloud.com/help/en/doc-detail/107745.htm)

The table below shows the various items which are billable and the method for calculation bills. Depending of the cloud service provider, the items and method may very but the table below gives us general idea.

Billable item	Description	Billing formula
Storage space occupied by log data	The storage space is the total size of compressed log data and indexes that are created on raw log data. For example, the size of raw log data that is uploaded to Log Service is 1 GB, and indexes are created for two fields. The compression ratio is 5:1 when the raw log data is uploaded, and the size of the indexes is 0.5 GB. In this example, the storage space occupied by log data is calculated by using the following formula: 0.2 GB + 0.5 GB = 0.7 GB.	Daily fee of the storage space occupied by log data = Used storage space per day × Price per GB of log data
Storage space occupied by cold log data	After you enable the hot and cold-tiered storage feature for a Logstore, you can configure the **Hot Data Retention Period** parameter. If the retention period of hot log data exceeds the specified threshold, the hot log data is converted to cold log data. The storage space is the total size of compressed log data and indexes that are created on raw log data. For example, the size of raw log data that is uploaded to Log Service is 1 GB, and indexes are created for two fields. The compression ratio is 5:1 when the raw log data is uploaded, and the size of the indexes is 0.5 GB. In this example, the storage space occupied by cold log data is calculated by using the following formula: 0.2 GB + 0.5 GB = 0.7 GB.	Daily fee of the storage space occupied by cold log data = Used storage space per day × Price per GB of log data
Storage space	The storage space is the total size of	Daily fee of the

Billable item	Description	Billing formula
occupied by time series data	raw time series data and indexes that are created on raw time series data. For example, the size of raw time series data that is uploaded to Log Service is 1 GB, and indexes are automatically created. In this example, the storage space occupied by time series data is calculated by using the following formula: 1 GB + 1 GB = 2 GB.	storage space occupied by time series data = Used storage space per day × Price per GB of time series data
Read and write traffic	Read and write traffic includes write traffic and read traffic. Write traffic: Write traffic is calculated based on the volume of compressed data that is uploaded to Log Service. For example, 10 GB of raw data is uploaded to Log Service, and the compression ratio is 5:1. In this example, the write traffic is 2 GB. Read traffic: Read traffic is calculated based on the volume of compressed data that is transformed, shipped to AnalyticDB for MySQL or ApsaraDB for Lindorm, or consumed. For example, 10 GB of raw data is uploaded to Log Service and then shipped to ApsaraDB for Lindorm, and the compression ratio is 5:1. In this example, the write traffic is 2 GB when the raw data is uploaded to Log Service, and the read traffic is 2 GB when the uploaded data is shipped to ApsaraDB for Lindorm. **Note** When you use the consumption preview feature in the Log Service console, a small amount of read traffic is generated.	Daily fee of read and write traffic = (Write traffic per day + Read traffic per day) × Price per GB of traffic
Index traffic of log data	The index traffic is calculated based on the indexes that are created or recreated on raw log data. Indexes are created for fields. The index traffic is	Daily fee of the index traffic of log data = Index traffic per day × Price per GB of

Billable item	Description	Billing formula
	based on the lengths of indexed fields and field values. For example, 1 GB of raw log data is written to Log Service, and the full-text indexing feature is enabled. In this example, the index traffic is 1 GB. For example, 1 GB of raw log data is written to Log Service, and indexes are created for two fields whose data length is 0.5 GB. In this example, the index traffic is 0.5 GB. **Note** By default, the indexing feature is disabled. If you enable this feature, index traffic is generated, and storage space is occupied by indexes. If you create both a full-text index and a field index for a field, the fee of index traffic for the field is calculated only once. Log Service automatically creates indexes for reserved fields such as __time__ and __source__. This generates a small amount of index traffic. For more information, see Reserved fields.	index traffic
Index traffic of time series data	The index traffic is calculated based on the indexes that are created on raw time series data. Indexes are created for fields. The index traffic is based on the lengths of indexed fields and field values. When you upload time series data, Log Service automatically creates indexes on the time series data. For example, 1 GB of raw time series data is written to Log Service. In this example, the index traffic is 1 GB.	Daily fee of the index traffic of time series data = Index traffic per day × Price per GB of index traffic
Read traffic over the	If data in Log Service is read and	Daily fee of the read traffic over the

Billable item	Description	Billing formula
Internet	consumed by third-party applications, you are charged for the read traffic over the Internet. The traffic is calculated based on the data after compression. If data is transformed across regions, you are charged for the read traffic over the Internet. The traffic is calculated based on the transformed data after compression.	Internet = Read traffic over the Internet per day × Price per GB of read traffic over the Internet
Data transformation	You are charged for the transformed data before compression. The fee of the **data transformation** billable item is calculated only based on the size of transformed data. However, the data transformation feature causes other fees. When you transform data, network resources are consumed, and API operations are called to read data. You are charged for the read traffic and the number of read and write operations. If data is transmitted across regions for data transformation, read traffic over the Internet is generated.	Daily fee of data transformation = Transformed data per day × Price per GB of transformed data
Data shipping	You are charged for the shipped data before compression. You can ship data to Object Storage Service (OSS), AnalyticDB for MySQL, and ApsaraDB for Lindorm. **Note** When you ship data to AnalyticDB for MySQL or ApsaraDB for Lindorm, network resources are consumed, and API operations are called to read data. You are charged for the read traffic and the number of read and write operations.	Daily fee of data shipping = Shipped data per day × Price per GB of shipped data
Read and write operations	When you upload data to Log Service, you are charged for the number of write operations. The number of write	Daily fee of read and write operations = Number of read and

Billable item	Description	Billing formula
	operations is based on the speed at which data is generated. Log Service automatically minimizes the number of write operations. When data is transformed, shipped to AnalyticDB for MySQL or ApsaraDB for Lindorm, or consumed, Log Service reads the data in batches. You are charged for the number of read operations. **Note** The number of read and write operations includes both successful and failed operations.	write operations per day × Price per operation
Voice calls	When you receive alert notifications by using voice calls, you are charged for the number of times that alert notifications are sent. **Note** If a voice call is not answered, Log Service sends a text message. You are charged for a voice call regardless of whether the call is answered. You are not charged for the text message that is sent upon a non-answered voice call.	Daily fee of voice calls = Number of times that alert notifications are sent per day × Price per notification
Text messages	When you receive alert notifications by using text messages, you are charged for the number of times that alert notifications are sent. **Note** If a text message is longer than 70 characters, the text message is split into two messages for sending. In this case, you are charged only for one text message.	Daily fee of text messages = Number of times that alert notifications are sent per day × Price per notification
Active shards	The fee is calculated based on the number of shards in the readwrite state. You are not charged for merged or split shards. For example, you want to merge three	Daily fee of active shards = Number of shards in the readwrite state × Price per shard

Billable item	Description	Billing formula
	shards that are in the readwrite state. After you merge the shards, only one shard is in the readwrite state. On the day when you merge the shards, you are charged for three shards. On the next day, you are charged for one shard. **Note** By default, two shards are created when you create a Logstore. For more information, see Why am I charged for active shards?	
Dedicated SQL	The fee is calculated based on the CPU time that is consumed when you perform query and analysis operations. Unit: core hour. The unit indicates the fee of one core that is used for 1 hour. For more information, see Billing example of Dedicated SQL.	Daily fee of Dedicated SQL = CPU time that is calculated by hour × Price per hour

Maintaining Strategic Flexibility in a cloud

Cloud users should plan systematic migration of operations to cloud. It should be in the form of incremental steps of various business processes. Organization should migrate operations in cloud with due consideration of flexibility of users. The Strategic flexibility offers freedom to make use of on-premise hardware along with data centre hardware depending on needs. Some of the parameters to consider regarding flexibility are listed below.

1. Prepare plan for migration and training of users
2. Select suitable cloud model (private, public or hybrid)
3. Identify which business processes should be migrated over cloud
4. Make detailed plan regarding current and future needs
5. Prepare a quantitative way to analyze benefits of cloud deployment
6. Identify the cost involved in migration
7. Plan and manage the cloud resources dynamically.

Best Practices for cost management in Cloud Computing

1. Cost Analysis

There should be complete visibility about cloud services in use, the actual usage patterns and trends of usage. We should consider current cloud usage and be able to project future spending. For this we need consolidated as well granular details in the form of interactive graphical and tabular reports across multiple dimensions, time frames. This will help us to correlate data for analysis and reporting against business objectives.

2. Role based Access

   Users should have permission to actively manage the infrastructure. There should be an Enterprise level mechanism to clearly define permissions and accessibility within the platform. Identify users and roles that launched, terminated, or changed infrastructure, and what they did to take corrective action for controlling costs.

3. Cloud Inventory visibility

   According to a recent survey, 75% report, they lack visibility of their cloud resources. This lack of visibility about resources in the cloud can lead to poor management of those resources. For effective cloud cost management, there should be in-depth analysis of infrastructure. If some resources in the cloud are unused due to lack of awareness, but the organization is still paying for them, cloud costs will climb unnecessarily.

4. Automated alert and notifications

   Automated alerts and notifications regarding authorization failures, budget overruns, cost spikes, etc. help to provide higher visibility and accountability of cloud resources.

5. Budget

   We should define and allocate budgets for each Departments, centres and projects. There should be proper approval mechanisms to avoid cloud cost overrun by sending out alerts when thresholds are reached. We can use the Showback report to chargeback Departments for their cloud usage and limit the cloud cost and use of resources.

6. Policy based governance

   We should use cloud-based governance tools to track cloud usage and costs and alert administrators when the total usage for the account is higher than threshold. There should be mechanism to schedule operational hours to automatically shut down & start virtual machines. Based on event or

thresholds, we should remove unused and underutilized resources and avoid unnecessary waste. we can also do this by resizing instances so they deliver a good balancebetween performance and cost. Avoid cost overrun by using policies to terminate servers.

Cloud Cost Optimization Tools

Cloud cost management cannot easily be managed by spreadsheets and manual lists. There are many automated tools available which can retrieve metrics from APIs and give us report about cloud consumption and costs. There are two main types of tools—first party tools provided directly by the cloud provider, and third party tools from external vendors.

1. Basic Tools: All public cloud platforms provide basic cost management tools. These tools are highly integrated with the cloud platform, and are available without any special deployment effort. Some of these tools are free to use, while others are billed on a pay-per-use model.

Generally, these basic tools are the fastest way to manage cloud costs. However, these tools have some limitations. Most first-party tools are limited in their ability to identify wasted costs and maximize savings using multiple cost models. These tools are limited to one vendor. So, they are not suitable when we use multi-cloud services.

2. Specific (Third Party) Tools: Third-party cost management tools can answer the functional limitations of basic tools. They also provide multi-cloud cost management. Most of these tools are built to reduce cloud costs across a variety of cloud services and workloads, providing clear return on investment (ROI).

4.4 LET'S SUM UP

Points to ponder

- TCO (Total Cost of Ownership) in Cloud environment
- Direct and Indirect costs
- Charge back models in Cloud
- Billable items
- Strategic Flexibilities in Cloud