BHADRAK ENGINEERING SCHOOL & TECHNOLOGY(BEST), ASURALI, BHADRAK

CLOUD COMPUTING

(TH-03)

(As per the 2020-21 syllabus of the SCTE&VT, Bhubaneswar, Odisha)

SIX Semester

COMPUTER SCIENCE & ENGG.

Prepared By: *Er. Aparajita Das*

CLOUD COMPUTING

CHAPTER-WISE DISTRIBUTION OF PERIODS & MARKS

sl no	Chapter no	Name of the Chapter as per thesyllabus	No. of periods asPer the syllabus	Expected marks
1	1	Introduction To Cloud Computing	05	**15**
2	2	Cloud Computing Architecture	08	**10**
3	3	Scalability And Fault Tolerance	08	**15**
4	4	Cloud Management And Virtualization Technology	08	**15**
5	5	Virtualization	08	**10**
6	6	Cloud Security	08	**10**
7	7	Cloud Computing Security Architecture	05	**15**
8	8	Market Based Management Of Clouds	05	**10**
9	9	Hadoop	05	**10**
Total:			**60**	**110**

CHAPTER NO. -01

INTRODUCTION TO CLOUD COMPUTING

Learning Objectives:

1.1 *Historical Development*

1.2 *Vision Of Cloud Computing*

1.3 *Characteristics Of Cloud Computing*

1.4 *Cloud Computing Reference Model*

1.5 *Cloud Computing Environment*

1.6 *Cloud Service Requirements*

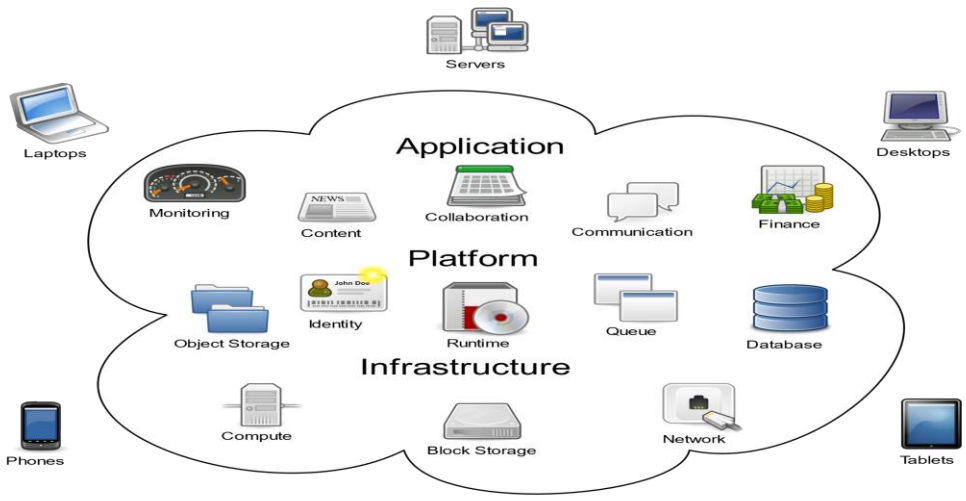
1.7 *Cloud And Dynamic Infrastructure*

1.8 *Cloud Adoption*

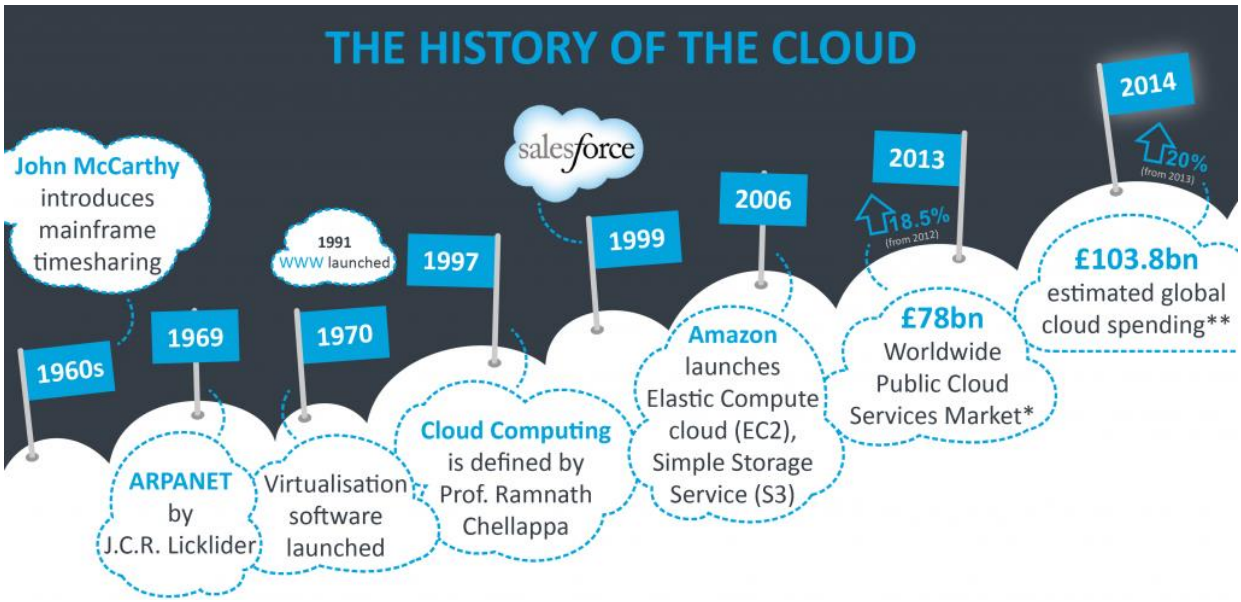
1.9 *Cloud Application*

Introduction

- Cloud computing is the on-demand availability of computer system resources, especially data storage and computing power, without direct active management by the user.
- It is generally used to describe data canters available to many users over the internet.



1.1 HISTORICAL DEVELOPMENT



- In early 1960s John Mc Carthy. Introduces mainframe timesharing system.
- In 1969 ARPANET (Advanced Research Projects Agency Network) is developed by J.C.R. licklider.
- In 1970 virtualization software launched.
- In 1997 cloud computing is defined by Prof Ramnath.
- In 1999 sales force is arrived.
- In 2003 virtual machine monitor is developed.
- In 2006, amazon expanded its cloud services.
- In 2013 World Wide Public Cloud Services come to market.
- In 2014,global business spending for infrastructure and services.

1.2 VISION OF CLOUD COMPUTING

Following are few forecasts of what we might expect in the coming future of Cloud computing:

- Cloud computing will become even more prominent in the coming years with rapid, continued growth of major global cloud data centres.
- 50% of all IT will be in the cloud within the next 5 – 10 years.
- The security and reliability of cloud computing will continue to evolve, ensuring that data will be even more secure with numerous techniques employed.
- We will not even consider ‘cloud’ as the key technology, instead we will focus on the services and applications that it enables.

1.3 CHARACTERISTICS OF CLOUD COMPUTING:

1. Resources Pooling

Cloud service provider can share resources among several clients, providing everyone with a different set of services as per their requirements. It is a multi-client strategy that can be applied to data storage services, processing services, and bandwidth provided services.

2. On-Demand Self-Service

It enables the client to constantly monitor the server uptime, abilities, and allotted network storage. This is a fundamental characteristic of Cloud Computing, and a client can likewise control the computing abilities as per his needs.

3. Easy Maintenance

The servers are effortlessly maintained, and the downtime remains low or absolutely zero sometimes. Cloud Computing powered resources undergo several updates frequently to optimize their capabilities and potential.

4. Scalability and Rapid Elasticity

A key characteristic and benefit of cloud computing is its rapid scalability. This cloud characteristic enables cost-effective running of workloads that require a vast number of servers but only for a short period. Many clients have such workloads, which can be run very cost-effectively because of the rapid scalability of Cloud Computing.

5. Economical

This cloud characteristic helps in reducing the IT expenditure of the organizations There is no covered up or additional charge which needs to be paid.

6. Measured and Reporting Service

Reporting services are one of the many cloud characteristics that make it the best choice for organizations. Measuring & reporting service is helpful for both cloud providers and their clients. It enables both the provider and the client to monitor and report what services have been used and for what purpose. This helps in monitoring billing and ensuring the optimum usage of resources.

7. Security

Cloud services create a copy of the data that is stored to prevent any form of data loss. If one server loses the data by any chance, the copy version is restored from the other server.

8. Automation

It requires the installation and deployment of virtual machines, servers, and large storage. Upon successful deployment, these resources require constant maintenance as well.

9. Resilience

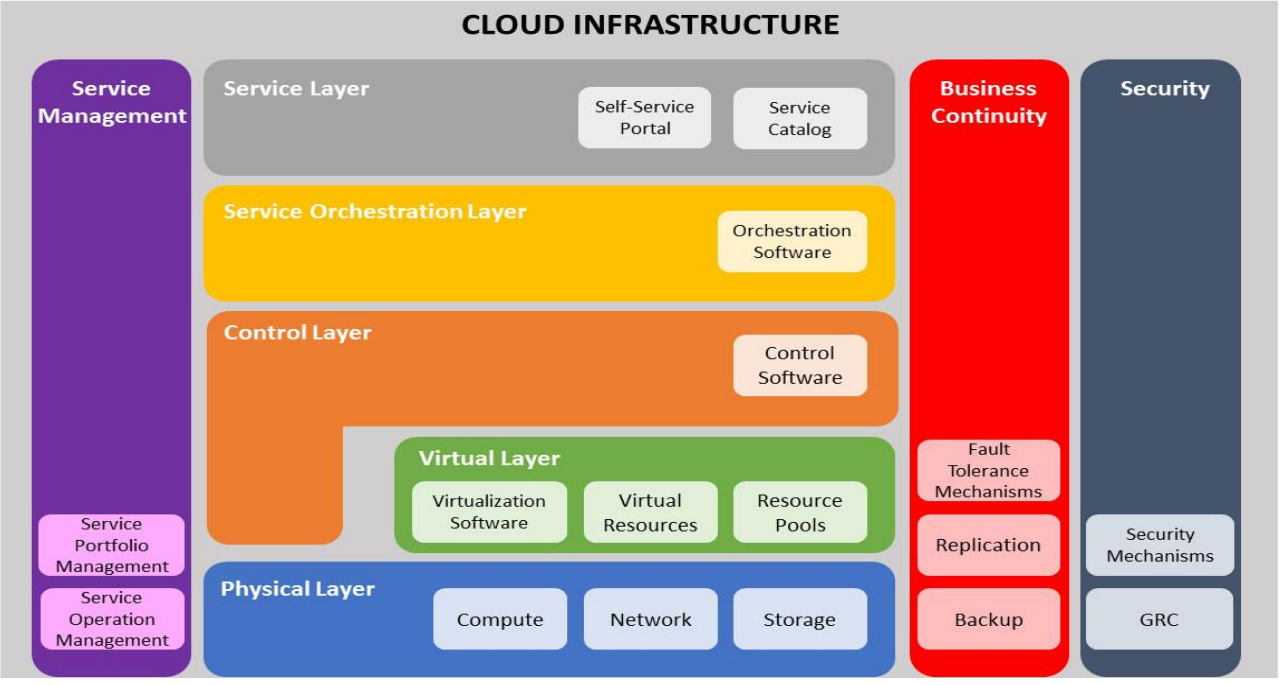
Resilience in cloud computing means the ability of the service to quickly recover from any disruption.

10. Large Network Access

Cloud providers save that large network access by monitoring and guaranteeing different measurements that reflect how clients access cloud resources and data: latency, access time, data throughput, etc.

1.4 CLOUD COMPUTING REFERENCE MODEL

- The **cloud computing reference model** is an abstract **model** that characterizes and standardizes the functions of a **cloud computing** environment by partitioning it into abstraction layers and cross-layer functions.
- The three cross-layer functions are business continuity, security, and service management.



- Beside that it consists of another five layer ie service layer,service orchestration layer,control layer, virtual layer and physical layer.
- Under service management portfolio and operation management are done.
- Fault toletance,replication and backup comes under bussiness continuity.
- Under security,security mechanism and GRC are done.

1.5 CLOUD COMPUTING ENVIRONMENT

- It is all about IT and what IT needs: different kinds of software and hardware, pay-per-use or subscription-based services offered both through the Internet and in real time.
- In this case the services are provided in a shared model.
- Here application server, database, code, mobile and PC shared the information in a cloud.
- This approach is a dream of numerous business owners who wish to get all possible IT services at one place.
- This kind of services is becoming more and more popular, as it helps enterpreneures resolve all IT challenges within one company quickly and efficiently.

1.6 CLOUD SERVICE REQUIREMENTS

1. Availability - with loss less DR

Customers want their IT services be up and available at all times. But in reality, computers sometimes fail. This implies that the service provider should have implemented a reliable

disaster recovery (DR) mechanism - where in the service provider can move the customer from one data centre to another seamlessly and the customer does not even have to know about it.

2. Portability of Data & Applications

Customers hate to be locked into a service or a platform. Ideally a cloud offering must be able to allow customers to move out their data & applications from one service provider to another - just like customers can switch from one telephone service provider to another.

3. Data Security

Security is the key concern for all customers - since the applications and the data is deciding in the public cloud, it is the responsibility of the service provider for providing adequate security. In my opinion security for customer

4. Manageability

Managing the cloud infrastructure from the customer perspective must be under the control of the customer admin. Customers of Cloud services must be able to create new accounts, must be able to provision various services, do all the user account monitoring - monitoring for end user usage, SLA breaches, data usage monitoring etc.

5. Elasticity

Customer on Cloud computing have a dynamic computing load. At times of high load, they need greater amount of computing resources available to them on demand, and when the work loads are low, the computing resources are released back to the cloud pool.

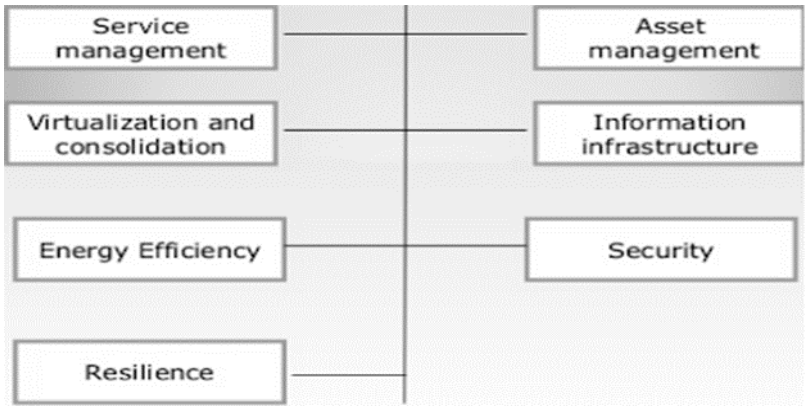
6. Federated System

This implies that each of the cloud services must have an interface with other cloud services for load sharing & application interoperability.

7. Closing Thoughts

Cloud services are still in its infancy and if cloud services were to attract large enterprise customers, then they need to do a lot more than today to address data/application portability, federated scalable system, complete end-to-end interoperability and security issues.

1.7 CLOUD AND DYNAMIC INFRASTRUCTURE



- **Service management:** This type of special facility or a functionality is provided to the cloud IT services by the cloud service providers. This facility includes visibility, automation and control to delivering the first-class IT services.
- **Asset-Management:** In this the assets or the property which is involved in providing the cloud services are getting managed.
- **Virtualization and consolidation:** Consolidation is an effort to reduce the cost of a technology by improving its operating efficiency and effectiveness. It means migrating from large number of resources to fewer one, which is done by virtualization technology.
- **Information Infrastructure:** It helps the business organizations to achieve the following: Information compliance, availability of resources retention and security objectives.
- **Energy-Efficiency:** Here the IT infrastructure or organization sustainable. It means it is not likely to damage or effect any other thing.
- **Security:** This cloud infrastructure is responsible for the risk management. Risk management Refers to the risks involved in the services which are being provided by the cloud-service providers.
- **Resilience:** This infrastructure provides the feature of resilience means the services are resilient. It means the infrastructure is safe from all sides. The IT operations will not be easily get affected.

1.8 CLOUD ADOPTION

- **Cloud adoption** is a strategy used by enterprises to improve the scalability of Internet-based database capabilities while reducing cost and risk.
- To achieve this, businesses engage in the practice of **cloud computing** or using remote servers hosted on the Internet to store, manage, and process critical data.
- A variety of industries benefit from cloud adoption, including healthcare, marketing and advertising, retail, finance, and education.
- Cloud adoption is a strategy used by enterprises to improve the scalability of Internet-based database capabilities while reducing cost and risk.

➢ Cloud adoption is a strategy used by enterprises to improve the scalability of Internet-based database capabilities while reducing cost and risk.

To achieve this, businesses engage in the practice of cloud computing or using remote servers hosted on the Internet to store, manage, and process critical data.

1.10 CLOUD APPLICATION-

- A cloud application, or cloud app, is a software program where cloud-based and local components work together.
- This model relies on remote servers for processing logic that is accessed through a web browser with a continual internet connection.
- Cloud application servers typically are located in a remote data centre operated by a third-party cloud services infrastructure provider.
- Cloud-based application tasks may encompass email, file storage and sharing, order entry, inventory management, word processing, customer relationship management (CRM), data collection, or financial accounting features.

POSSIBLE SHORT TYPE QUESTIONS WITH ANSWERS

1. Define Cloud Computing.

Ans-

- **Cloud computing** is the on-demand availability of computer system resources, especially data storage (cloud storage) and computing power, without direct active management by the user.
- cloud computing is commonly known as delivery of computing services including servers, storage, databases, and intelligence over the Internet. The term is generally used to describe data centres available to many users over the Internet.

2.What are the characteristics of cloud computing?

Ans- Resources Pooling.

- On-Demand **Self**-Service.
- Easy Maintenance.
- Large **Network** Access.
- Availability.
- Automatic System.
- Economical.

- **Security**.

3. Define Cloud Computing environment.

Ans- It is all about IT and what IT needs: different kinds of software and hardware, pay-per-use or subscription-based services offered both through the Internet**.**

4. Define virtualization.

Ans- Virtualization is the creation of virtual servers, infrastructures, devices and **computing** resources. **Virtualization** changes the hardware-software relations and is one of the foundational elements of **cloud computing** technology that helps utilize the capabilities of **cloud computing** to the full.

5. Define consolidation.

Ans- In **computing**, **consolidation** refers to when data storage or server resources are shared among multiple users and accessed by multiple applications. **Consolidation** aims to make more efficient use of computer resources and prevent servers and storage equipment from being under-utilized and taking too much space.

6. Define Cloud adoption.

Ans- Cloud adoption is a strategy used by enterprises to improve the scalability of Internet-based database capabilities while reducing cost and risk. To achieve this, businesses engage in the practice of **cloud computing** or using remote servers hosted on the Internet to store, manage, and process critical data.

7.Define Cloud application.

Ans- Cloud applications are software that users access primarily through the internet, **meaning** at least some of it is managed by a server and not users local machines.

POSSIBLE LONG TYPE QUESTIONS

1. Explain Cloud Computing reference model.
2. Explain the characteristics of cloud computing.
3. Explain dynamic infrastructure.
4. Write short note on
    a. Cloud Service requirement
    b. Cloud Computing environment

CHAPTER NO.- 02

CLOUD COMPUTING ARCHITECTURE

Learning Objectives:

2.1 Introduction

2.2 Cloud reference model

2.3 Types of cloud

2.4 interoperability in cloud computing

2.5 Cloud computing Interoperability use cases

2.6 Role of standards in Cloud Computing environment

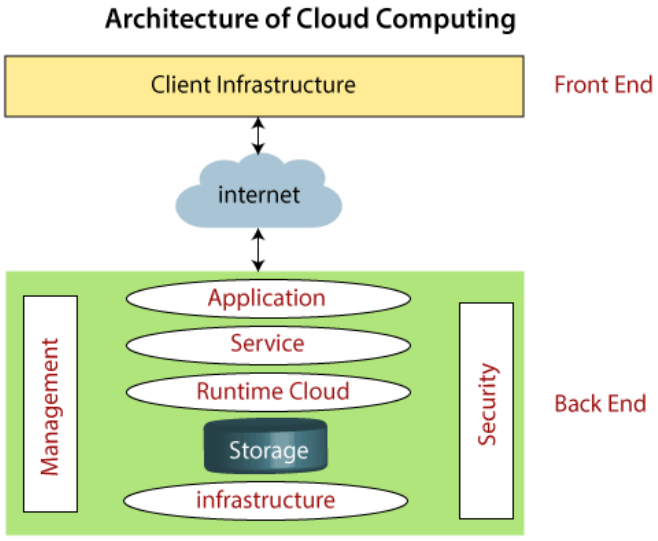
2.1 INTRODUCTION-

- Cloud computing technology is used by both small and large organizations to store the information in cloud and access it from anywhere at any time using the internet connection.
- Cloud computing architecture is a combination of service-oriented architecture and event-driven architecture.
- Cloud computing architecture is divided into the following two parts –

A. Front End
B. Back End

Front End

- The front end is used by the client.
- It contains client-side interfaces and applications that are required to access the cloud computing platforms.
- The front end includes web servers (including Chrome, Firefox, internet explorer, etc.), thin & fat clients, tablets, and mobile devices.



Back End

- The back end is used by the service provider.
- It manages all the resources that are required to provide cloud computing services.
- It includes a huge amount of data storage, security mechanism, virtual machines, deploying models, servers, traffic control mechanisms, etc.

2.2 CLOUD REFERENCE MODEL

- To achieve the potential of cloud computing, there is a need to have a standard cloud reference model for the software architects, software engineers, security experts and businesses, since it provides a fundamental reference point for the development of cloud computing.
- The Cloud Reference Model brings order to this cloud landscape.
- This figure appearing here also illustrates various cloud providers and their technologies within the available cloud service models in the market.

2.3 TYPES OF CLOUD

There are four types of cloud:

1. Public cloud.
2. Private cloud.
3. Hybrid cloud.
4. Community cloud.

1. **Public cloud:**
   - Public cloud is managed by third parties which provide cloud services over the internet to public, these services are available as pay-as-you-go billing mode.
   - They offer solutions for minimizing IT infrastructure costs and act as a good option for handling peak loads on the local infrastructure.
   - They are a goto option for small enterprises, which are able to start their businesses without large upfront investments by completely relying on public infrastructure for their IT needs.
   - A fundamental characteristic of public clouds is multitenancy. A public cloud is meant to serve multiple users, not a single customer. A user requires a virtual computing environment that is separated, and most likely isolated, from other users.
2. **Private Cloud**

   - Individuals/organizations that choose Private Cloud gets dedicated infrastructure that is not shared by any other individual/organization.

   - The security and control level is highest while using a private network. The costs are born by an individual/organization and are not shared with any other individual/organization. Management of Private Cloud is taken care of by the user and the CSP does not provide any Cloud management services.

3. **Hybrid Cloud**

   - This Cloud deployment model includes the characteristics of Public Cloud and Private Cloud. Hybrid Cloud allows the sharing of data and applications between Public and Private Cloud environments.

   - Organizations mainly use Hybrid Cloud when their On-Premise infrastructure needs more scalability, so they make use of scalability on Public Cloud to meet fluctuating business demands.

   - Organizations can keep their sensitive data on their Private Cloud when reaping the power of the Public Cloud.

4. **Community Cloud**

   - A Community Cloud is a Cloud infrastructure that is shared by users of the same industry or by those who have common goals.

   - This Cloud infrastructure is built after understanding the computing needs of a community as there are many factors including compliances and security policies which need to be included in the community Cloud infrastructure.

2.4 INTEROPERABILITY IN CLOUD COMPUTING:

Interoperability means enabling the cloud ecosystem so that multiple cloud platforms can exchange information.

- It gives the ability to the customers to use the same or similar management tools, re-use server images and other software within a variety of cloud computing providers and platforms.
- It refers to portability, i.e., the ability to move a system from one platform to another.
- Standards are necessary to consolidate efforts in a technology domain and to enable interoperability in any technology domain.

• Service modelling: Open-SCA (service composition and interaction), USDL/SoaML/CloudML

(multi-view services), EMML (mashups)

• Service interfaces: OCCI (infrastructure management), CIMI (infrastructure management), EC2

(de-facto standard), TOSCA (portability), CDMI (data management)

2.5 CLOUD COMPUTING INTEROPERABILITY USE CASES

1. **Workload migratio**n. A workload that executes in one cloud provider can be uploaded to another cloud provider. Some standardization efforts that support this use case are Amazon Machine Image (AMI), Open Virtualization Framework (OVF), and Virtual Hard Disk (VHD).
2. **Data migration**. Data that resides in one cloud provider can be moved to another cloud provider.

   Eg- Cloud Data Management Interface (CDMI).

   It supports data- and storage-management interfaces that use SOAP and REST.

3. **User authentication**. A user who has established an identity with a cloud provider can use the same identity with another cloud provider.

   Eg-Amazon Web Services Identity Access Management (AWS IAM), OAuth, OpenID, and WS-Security.

4. **Workload management**. Custom tools developed for cloud workload management can be used to manage multiple cloud resources from different vendors.

   Even though most environments provide a form of management console or command-line tools, they also provide APIs based on REST or SOAP.

2.6 ROLE OF STANDARDS IN CLOUD COMPUTING ENVIRONMENT

1. 20BInfrastructure as a Service (IaaS) 13
2. BPlatform as a Service (PaaS) 14
3. BSoftware as a Service (SaaS) 14
4. 23BDo Standards Make Sense Beyond IaaS? 15
5. 24BCan Existing Standards Support Cloud Interoperability Instead of Portability, or Do Clouds Require New Standards.

POSSIBLE SHORT TYPE QUESTIONS WITH ANSWERS

1. Define cloud computing architecture and its type?

Ans - Cloud computing architecture is a combination of service-oriented architecture and event-driven architecture.

Cloud computing architecture is divided into the following two parts –

Front End

Back End

2. What are the types of cloud?

Ans- There are four types of cloud:

Public cloud, Private cloud, Hybrid cloud, Community cloud.

3.Define Public cloud?

Ans- Public cloud are managed by third parties which provide cloud services over the internet to public, these services are available as pay-as-you-go billing mode.

4.Define Private cloud?

Ans - Individuals/organizations that choose Private Cloud gets dedicated infrastructure that is not shared by any other individual/organization. The security and control level is highest while using a private network. The costs are born by an individual/organization and are not shared with any other individual/organization.

5.Define Hybrid cloud?

Ans -This Cloud deployment model includes the characteristics of Public Cloud and Private Cloud. Hybrid Cloud allows the sharing of data and applications between Public and Private Cloud environments.

6. Define Community cloud?

Ans - A Community Cloud is a Cloud infrastructure that is shared by users of the same industry or by those who have common goals.

7. **What is interoperability in cloud computing?**

Ans- Interoperability means enabling the cloud ecosystem so that multiple cloud platforms can exchange information.

POSSIBLE LONG TYPE QUESTIONS

5. Define cloud computing architecture and explain its type.
6. Explain cloud reference model.
7. Explain the types of cloud.
8. What are the cloud computing interoperability use cases?

CHAPTER NO.-03
SCALABILITY AND FAULT TOLERANCE

Learning Objectives:

3.1 Introduction

3.2 Scalability and Fault Tolerance

3.3 Cloud solutions

3.4 Cloud Ecosystem

3.5 Cloud Business process management

3. 6 Portability and Interoperability

3. 7 Cloud Service management

3.8 Cloud Offerings

3.9 testing under control

3.10 Cloud service control

3.11 Virtual desktop Infrastructure

3.1 INTRODUCTION

It is a challenging task for the cloud providers to develop such high scalable and fault tolerance systems who can get managed and at the same time they will provide a competitive performance.

3.2 SCALABILITY AND FAULT TOLERANCE

Fault tolerance – The management system must automatically detect and recover from application and resource failures.

Scalability – The infrastructure must scale to hundreds or thousands of resources.

- Cloud Fault Tolerance is tolerating the faults by the cloud that are done by mistake by the user.

- Here the scaling is beyond the limits, it means we can't even imagine what will be the limit.
- Cloud middleware is designed on the principle of scalability along different dimensions in mind e.g.: - performance, size and load.
- The cloud middleware manages a huge number of resources and users which depends on the cloud to obtain that they can't obtain within the premises without affording the administrative and maintenance costs.

3.3 CLOUD SOLUTIONS

- A cloud-based solution refers to on-demand services, computer networks, storage, applications or resources accessed via the internet and through another provider's shared cloud computing infrastructure.
- It can enable companies to focus on revenue driving initiative rather than time consuming, non-core business tasks.
- The ability to access cloud-based solutions from anywhere with an internet connection paired with the widespread adoption of smartphones and faster mobile networks.
- The user able to access cloud- based solutions from anywhere and anytime.

Benefits:

It increased capacity, scalability, functionality and reduced maintenance and cost for computer infrastructure or in- house staff.

3.4 CLOUD ECOSYSTEM:

- Cloud ecosystem is a term used to describe the complex system of interdependent components that work together to enable cloud services.
- The centre of a cloud ecosystem is a public cloud provider. It might be an IaaS provider such as Amazon Web Services (AWS) or a SaaS vendor such as Salesforce.
- There is no vendor lock-in in the cloud ecosystem.

For example

AWS is the centre of its own ecosystem, but it's also a part of the Salesforce ecosystem. Salesforce runs a number of its services on AWS's infrastructure.

3.5 CLOUD BUSINESS PROCESS MANAGEMENT

- Cloud business process process management is usually a platform-as-a-service solution that lets you create workflows and optimise them.
- Without having to install a single Mb of software on your office hardware, you can use these cloud-based software solutions to streamline and optimise everyday business activities.

- In business, as in life, you have the option to transform and grow proactively or react to pressing industry demands after it's already too late to get a competitive advantage.
- While business process management applications aren't new, technological advancement now presents you with the opportunity to move to cloud business process management software.

3. 6 PORTABILITY AND INTEROPERABILITY

- **Interoperability** means the ability of two cloud systems to talk to another, i.e., to exchange messages and information in a way that both can understand.
- **Data portability** means the ability to move data (files, documents, database tables, etc.) from one cloud system to another, and have that data usable in the other system.
- **Application Portability** means the ability to move executable software from one cloud system to another, and be able to run it correctly in the destination system.

3. 7 CLOUD SERVICE MANAGEMENT

- Service Management in the Cloud era» ITSM (Information Technology Service Management) must expand service management methodologies to include managing cloud services – CSM (Cloud Service Management)
- The landscape of how we deliver IT services is rapidly changing; from an on premise or traditional datacentre service delivery, to IT services being delivered by cloud service providers.
- Service Management must be redesigned and include new methodologies in how we manage these new cloud services.
- There is huge potential both for the service provider and the end user by adopting the processes in cloud computing in to service management.

3.8 CLOUD OFFERINGS

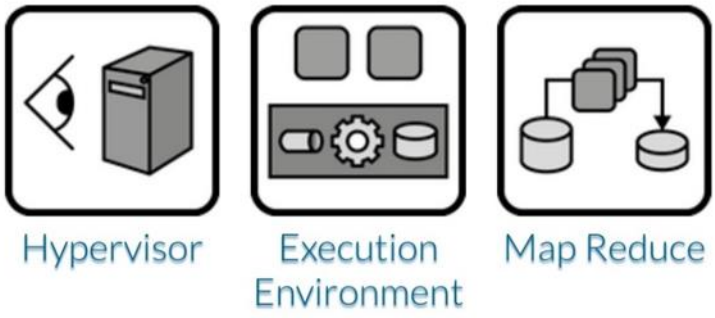
- Patterns of this category cover different functionality found in clouds regarding the functionality they provide to customers and the behavior they display.

Cloud Environment

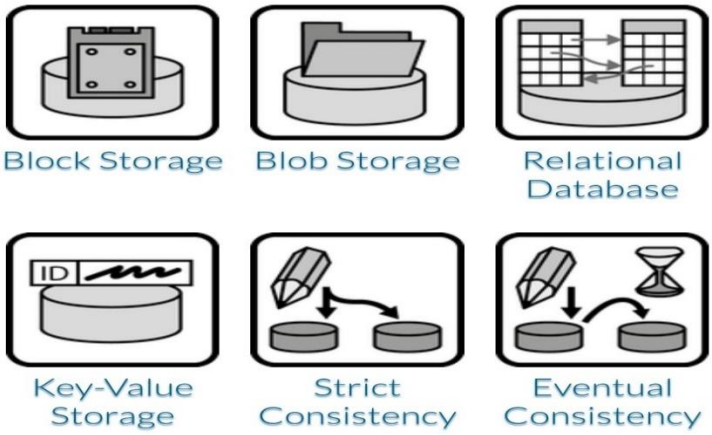
Patterns of this category descibe the hosting environments of cloud in detail and refer to other offerings composed to form these environments.

Processing Offering

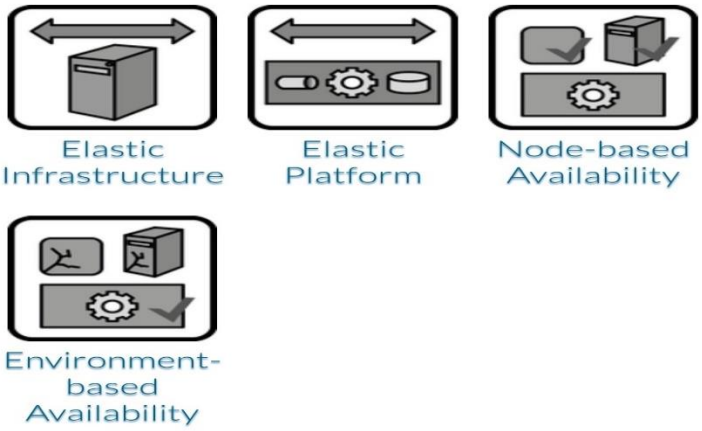
Patterns of this category describe how computation can be performed in the cloud



Storage Offering

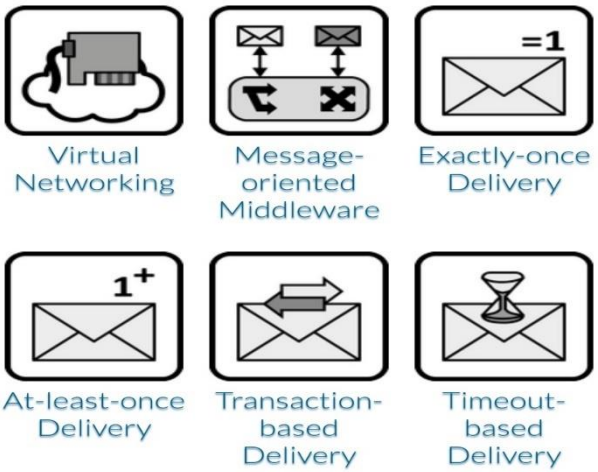


Patterns of this category describe how data can be stored in the cloud



Communication Offering

Patterns of this category describe how data can be exchanged in the cloud.



3.9 TESTING UNDER CONTROL

- Cloud testing is a subset of software testing in which simulated, real-world Web traffic is used to test cloud-based Web applications.
- Cloud testing also verifies and validates specific cloud functions, including redundancy
- A number of small to medium-sized IT organizations have migrated to cloud solutions. As a result, cloud testing has become necessary to validate functional system and business requirements.
- In addition to cloud experience, cloud testing engineers require the knowledge of different types of testing and tools.

3.10 CLOUD SERVICE CONTROL

- **Cloud Service Controls** allow customers to address threats such as data theft, accidental data loss, and excessive access to data stored in Google Cloud multi-tenant **services**.
- It enables clients to tightly **control** what entities can access what **services** in order to reduce both intentional and unintentional losses.
- It provides an additional layer of security defense for Google cloud services that is independent of identity and access management.

3.11 VIRTUAL DESKTOP INFRASTRUCTURE

- Virtual desktop infrastructure (VDI) is defined as the hosting of desktop environments on a central server.
- It is a form of desktop virtualization, as the specific desktop images run within virtual machines (VMs) and are delivered to end clients over a network.
- Those endpoints may be PCs or other devices, like tablets or thin client terminals.

POSSIBLE SHORT TYPE QUESTIONS WITH ANSWERS

i. **Define Fault tolerance?**

Ans -The management system must automatically detect and recover from application and resource failures.

ii. **Define Scalability?**

Ans - The infrastructure must scale to hundreds or thousands of resources.

iii. Define Cloud service control?

Ans -Cloud Service Controls allow customers to address threats such as data theft, accidental data loss, and excessive access to data stored in Google Cloud multi-tenant **services.**

iv. Define Cloud Service management?

Ans - Service Management in the Cloud era» ITSM (Information Technology Service Management) must expand service management methodologies to include managing cloud services – CSM (Cloud Service Management)

v. **What are the benefits and limitations of VDI?**

Ans - VDI supports enhanced user mobility and remote access, as a standardized desktop can be reached from almost any approved and compatible endpoint in any location. For workers who are frequently on the go and need to pull up a virtual desktop containing a full range of virtual apps and data, VDI is like having an office available on-demand. In that regard, it fits right into their digital workspace workflows that already feature similar, regular consumption of cloud, web and mobile apps across multiple contexts, especially if it's persistent VDI.

vi. Define Interoperability.

Ans - Interoperability *means the ability of two cloud systems to* talk to another*, i.e., to exchange messages and information in a way that both can understand.*

vii. **Define Portability.**

Ans -Data portability means the ability to move data (files, documents, database tables, etc.) from one cloud system to another, and have that data usable in the other system.

POSSIBLE LONG TYPE QUESTIONS

1.Explain cloud offerings.

2. Explain cloud solution.
3.Explain the need of cloud services control and management.

CHAPTER NO.-04
CLOUD MANAGEMENT AND VIRTUALISATION TECHNOLOGY

Learning Objectives:

4.1 Create a virtualised Architecture

4.2 Data Centre

4.3 Resiliency

4.4 Agility

4.5 Cisco Data Centre Network Architecture

4.6 Storage

4.7 Provisioning

4.8 Asset Management

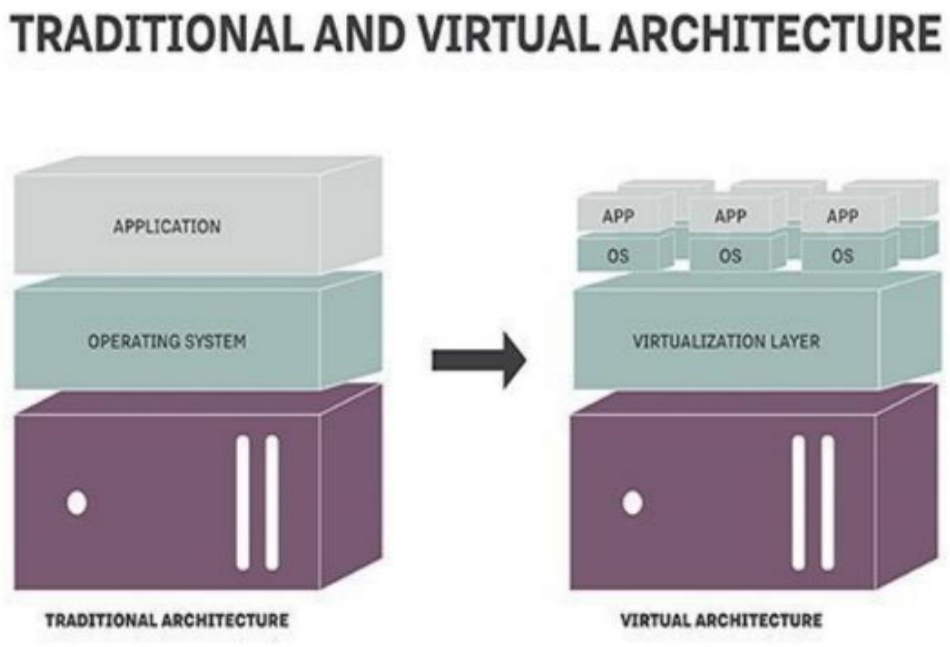
4.9 Concept of Map Reduce

4.1 CREATE A VIRTUALISED ARCHITECTURE

- A virtualization architecture is a conceptual model specifying the arrangement and interrelationships of the particular components involved in delivering a virtual – rather than physical – version of something, such as an operating system (OS), a server, a storage device or network resources.
- Virtualization is commonly hypervisor-based. The hypervisor isolates operating systems and applications from the underlying computer hardware so the host machine can run

multiple virtual machines (VM) as guests that share the system's physical compute resources, such as processor cycles, memory space, network bandwidth and so on..

- Type 1 hypervisors, sometimes called bare-metal hypervisors, run directly on top of the host system hardware. Bare-metal hypervisors offer high availability and resource management.
- Their direct access to system hardware enables better performance, scalability and stability. Examples of type 1 hypervisors include Microsoft Hyper-V.
- A type 2 hypervisor, also known as a hosted hypervisor, is installed on top of the host operating system, rather than sitting directly on top of the hardware as the type 1 hypervisor does. Each guest OS or VM runs above the hypervisor.
- The convenience of a known host OS can ease system configuration and management tasks. However, the addition of a host OS layer can potentially limit performance and expose possible OS security flaws.
- Examples of type 2 hypervisors include VMware Workstation, Virtual PC and Oracle VM VirtualBox.
- The main alternative to hypervisor-based virtualization is containerization. Operating system virtualization, for example, is a container-based kernel virtualization method.
- OS virtualization is similar to partitioning.
- In this architecture, an operating system is adapted so it functions as multiple, discrete



systems, making it possible to deploy and run distributed applications without launching an entire VM for each one.

4.2 DATA CENTRE

- Data centre is a building, dedicated space within a building, or a group of buildings used to house computer systems and associated components, such as telecommunications and storage systems.
- Since IT operations are crucial for business continuity, it generally includes redundant or backup components and infrastructure for power supply, data communication connections, environmental controls (e.g. air conditioning, fire suppression) and various security devices.
- A large data centre is an industrial-scale operation using as much electricity as a small town.

4.3 RESILIENCY

- Resiliency is the ability to handle failures gracefully and recover the whole system. This is a huge challenge for services and applications where the components compete for resources, and depend on other internal or external components/ services that fail, or may rely on defective software.
- Resiliency is the ability of a server, network, storage system, or an entire data centre, to recover quickly and continue operating even when there has been an equipment failure, power outage or other disruption.
- Data centre resiliency is a planned part of a facility's architecture and is usually associated with other disaster planning and data centre disaster-recovery considerations such as data protection. The adjective *resilient* means "having the ability to spring back."
- Data centre resiliency is often achieved through the use of redundant components, subsystems, systems or facilities.
- When one element fails or experiences a disruption, the redundant element takes over seamlessly and continues to support computing services to the user base. Ideally, users of a resilient system never know that a disruption has even occurred.

For example

- If an ordinary server's power supply fails, the server fails -- and all of the workloads on that server become unavailable until the server is repaired and restarted (or the workloads can be restarted on another suitable server).
- If the server incorporates a redundant power supply, the backup supply keeps the server running until a technician can replace the failed power supply.
- Techniques, such as server clustering, support redundant workloads on multiple physical servers. When one server in the cluster fails, another node takes over with its redundant workloads.
- The same concept holds true all the way up to entire data centre facilities. For example, an organization may power its data centre with two separate utility feeds from different utility providers so that a backup provider is available when the first utility provider fails.
- As another example, organizations that support hot sites can support data centre collocation–shifting an entire operation from one facility to another in response to any kind of local disruption or regional disaster.
- The resiliency techniques employed in a data centre can vary with the importance of the respective workloads. Organizations with mission-critical workloads will utilize more resiliency techniques at more levels within the data centre, because the cost of *not* preserving critical computing services is typically costlier during a prolonged service outage.
- For example, critical business services, such as transaction processing software or database systems, may be designed with comprehensive data centre resiliency, including clustering, snapshots and off-site redundancy.
- Conversely, nonessential workloads that can tolerate some level of disruption may receive little resiliency or simply remain offline until they can be restored.

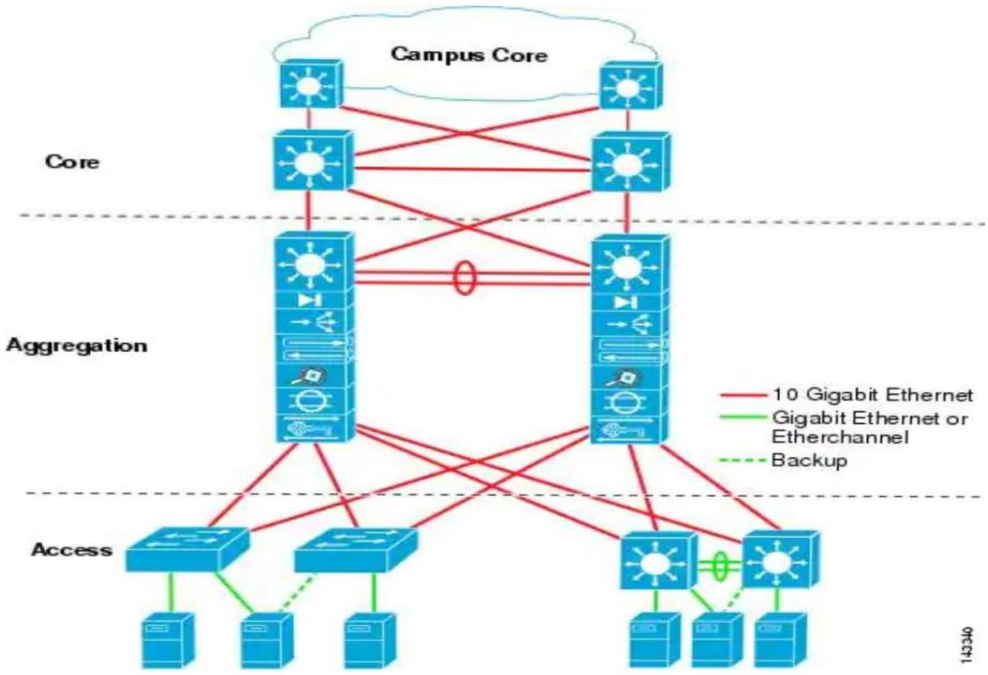
4.4 AGILITY

A key benefit often discussed about cloud computing is how it enables agility. As agility may be defined as "the power of moving quickly and easily; nimbleness" it's easy to see how this rapid provisioning is referred to advancing agility.

4.5 CISCO DATA CENTER NETWORK ARCHITECTURE

It can be grouped into four key areas: Enterprises are starting to use low-latency interconnects to support parallel and tightly coupled applications that provide financial modelling, fluid dynamics and data mining.

Basic Layered Design



- **Core layer**—Provides the high-speed packet switching backplane for all flows going in and out of the data centre. The core layer provides connectivity to multiple aggregation modules and provides a resilient Layer 3 routed fabric with no single point of failure. The core layer runs an interior routing protocol, such as OSPF or EIGRP, and load balances traffic between the campus core and aggregation layers using Cisco Express Forwarding-based hashing algorithms.
- **Aggregation layer modules**—Provide important functions, such as service module integration, Layer 2 domain definitions, spanning tree processing, and default gateway redundancy. Server-to-server multi-tier traffic flows through the aggregation layer and can use services, such as firewall and server load balancing, to optimize and secure applications. The smaller icons within the aggregation layer switch in represent the integrated service modules. These modules provide services, such as content switching, firewall, SSL offload, intrusion detection, network analysis, and more.
- **Access layer**—Where the servers physically attach to the network. The server components consist of 1RU servers, blade servers with integral switches, blade servers with pass-through cabling, clustered servers, and mainframes with OSA adapters. The access layer network infrastructure consists of modular switches, fixed configuration 1 or 2RU switches, and integral blade server switches. Switches provide both Layer 2 and Layer 3 topologies, fulfilling the various server broadcast domain or administrative requirements.

4.6 STORAGE

- Data centre storage primarily refers to the devices, equipment and software technologies that enable data and application storage within a data centre facility. This includes: Hard disk drives, tape drives and other forms of internal and external storage.
- Data Centre storage primarily refers to the devices, equipment and software technologies that enable data and application storage within a data centre facility. This includes:
- Hard disk drives, tape drives and other forms of internal and external storage

Storage and backup management software utilities.

- External storage facilities/solutions such as cloud or remote storage

Storage networking technologies such as storage area networks (SAN), network attached storage (NAS), RAID and more.

4.7 PROVISIONING

- Cloud provisioning is the allocation of a cloud provider's resources and services to a customer.
- Cloud provisioning is a key feature of the cloud computing model, relating to how a customer procures cloud services and resources from a cloud provider.
- The growing catalogue of cloud services that customers can provision includes infrastructure as a service (IaaS), software as a service (SaaS) and platform as a service (PaaS) in public or private cloud environments.

Types of cloud provisioning

- The cloud provisioning process can be conducted using one of three delivery models.
- Each delivery model differs depending on the kinds of resources or services an organization purchases, how and when the cloud provider delivers those resources or services, and how the customer pays for them.
- The three models are advanced provisioning, dynamic provisioning and user self-provisioning. With advanced provisioning, the customer signs a formal contract of service with the cloud provider.
- The provider then prepares the agreed-upon resources or services for the customer and delivers them. The customer is charged a flat fee or is billed on a monthly basis.
- With dynamic provisioning, cloud resources are deployed flexibly to match a customer's fluctuating demands.
- Cloud deployments typically scale up to accommodate spikes in usage and scale down when demands decrease.

- The customer is billed on a pay-per-use basis. When dynamic provisioning is used to create a hybrid cloud environment, it is sometimes referred to as cloud bursting.

4.8 ASSET MANAGEMENT

- Cloud asset management (CAM) is a component of cloud management services focused exclusively on the management of a business's physical cloud environment, such as the products or services they use.
- Put simply, CAM keeps track of every aspect of your cloud estate, managing the maintenance, compliance, upgrading, and disposal of cloud assets.
- By ensuring these processes run smoothly, companies can reap the benefits of their cloud infrastructure while only spending what they need.

4.9 CONCEPT OF MAP REDUCE

- MapReduce is a software framework for processing (large1) data sets in a distributed fashion over a several machines.
- The core idea behind MapReduce is mapping your data set into a collection of <key, value> pairs, and then reducing overall pairs with the same key.
- MapReduce is a programming model and an associated implementation for processing and generating big data sets with a parallel, distributed algorithm on a cluster.
- A MapReduce program is composed of a map procedure, which performs filtering and sorting (such as sorting students by first name into queues, one queue for each name), and a reduce method, which performs a summary operation (such as counting the number of students in each queue, yielding name frequencies).
- The "MapReduce System" (also called "infrastructure" or "framework") orchestrates the processing by marshalling the distributed servers, running the various tasks in parallel, managing all communications and data transfers between the various parts of the system, and providing for redundancy and fault tolerance.
- MapReduce libraries have been written in many programming languages, with different levels of optimization.
- A popular open-source implementation that has support for distributed shuffles is part of Apache Hadoop.
- The name MapReduce originally referred to the proprietary Google technology, but has since been genericized.

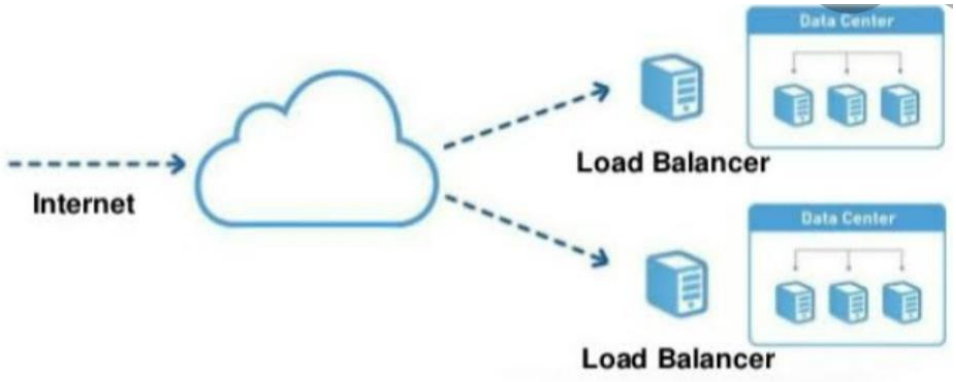
- By 2014, Google was no longer using MapReduce as their primary big data processing model, and development on Apache Mahout had moved on to more capable and less disk-oriented mechanisms that incorporated full map and reduce capabilities.

4.10 CLOUD GOVERANCE

- Cloud governance is a set of rules you create, monitor, and amend as necessary in order to control costs, improve efficiency, and eliminate security risks.
- There may be other areas of your cloud operations that require governance, but these will become apparent when you first start pulling together the components that will eventually form your rules of governance.

4.11 LOAD BALANCING

7. Cloud load balancing is a type of load balancing that is performed in cloud computing.
8. Cloud load balancing is the process of distributing workloads across multiple computing resources.
9. Cloud load balancing reduces costs associated with document management systems and maximizes availability of resources.
10. It is a type of load balancing and not to be confused with Domain Name System (DNS) load balancing.
11. While DNS load balancing uses software or hardware to perform the function, cloud load balancing uses services offered by various computer network companies.



Importance of load balancing-

- Cloud computing brings advantages in "cost, flexibility and availability of service users."
- Those advantages drive the demand for Cloud services.
- The demand raises technical issues in Service Oriented Architectures and Internet of Services (IoS)-style applications, such as high availability and scalability.
- As a major concern in these issues, load balancing allows cloud computing to "scale up to increasing demands" by efficiently allocating dynamic local workload evenly across all nodes.

4.12 HIGH AVAILABILITY

- **High availability** is a quality of **computing** infrastructure that allows it to continue functioning, even when some of its components fail.
- Highly available systems guarantee a certain percentage of uptime—for example, a system that has 99.9% uptime will be down only 0.1% of the time—0.365 days or 8.76 hours per year.
- **High Availability** is a non-functional factor that provide uninterrupted IT services to the customer from ONE data centre. Every Infrastructure layer of an Application Architecture has more than one similar device and runtime software products.
- For example: 2+ web servers, 2+ application servers, 2+ databases, 2+ load balancers, 2+ firewalls for one application.
- The failure of one device in the above flow doesn't affect the end customer.
- An Availability Zone in the Cloud is a Data-centre.
- A Region in the Cloud is a geographical area that consists of more than one Availability Zone.

4.13 DISASTER RECOVERY

- **Cloud disaster recovery** is a **cloud computing** service which allows for storing and recovering system data on a remote **cloud**-based platform.
- Internet connectivity with sufficient bandwidth to enable remote access to the secondary data centre.
- It is another non-functional factor that provide uninterrupted IT services on demand basis. In this model, the IT business have TWO data-centres.

- Primary and Standby (Second data-centre that is geographically separated). The application flow at both data centres is usually identical like above.
- There is a possibility for the primary data centre failure due to flood, power failures, hurricanes, and other unexpected issues.
- In that case, the second data centre (Standby) will start serving the end user. There are different DR models and you may want to refer some online articles to know about them.
- Cloud Computing is a different animal.
- HA and DR differences exists for Cloud Providers however it shouldn't be a concern for Cloud Consumers.

POSSIBLE SHORT TYPE QUESTIONS WITH ANSWERS

1. **Define virtualization architecture?**

Ans- A virtualization architecture is a conceptual model specifying the arrangement and interrelationships of the particular components involved in delivering a virtual – rather than physical – version of something, such as an operating system (OS), a server, a storage device or network resources.

2.What are the Importance of load balancing?

Ans-Cloud computing brings advantages in "cost, flexibility and availability of service users. Those advantages drive the demand for Cloud services. The demand raises technical issues in Service Oriented Architectures and Internet of Services (IoS)-style applications, such as high availability and scalability.

3. **What is Disaster Recovery**?

Ans -Cloud disaster recovery is a **cloud computing** service which allows for storing and recovering system data on a remote **cloud**-based platform.

4. **What is Cloud Governance?**

Ans - Cloud governance is a set of rules you create, monitor, and amend as necessary in order to control costs, improve efficiency, and eliminate security risks. There may be other areas of your cloud operations that require governance, but these will become apparent when you first start pulling together the components that will eventually form your rules of governance.

5. Define Asset Management.

Ans - Cloud asset management (CAM) is a component of cloud management services focused exclusively on the management of a business's physical cloud environment, such as the products or services they use.

6. Define Cloud Provisioning.

Ans - Cloud provisioning is the allocation of a cloud provider's resources and services to a customer. Cloud provisioning is a key feature of the cloud computing model, relating to how a customer procures cloud services and resources from a cloud provider.

7. Define Agility.

Ans - A key benefit often discussed about cloud computing is how it enables agility. As agility may be defined as "the power of moving quickly and easily; nimbleness" it's easy to see how this rapid provisioning is referred to advancing agility.

8. Define Data Centre.

Ans - Data centre is a building, dedicated space within a building, or a group of buildings used to house computer systems and associated components, such as telecommunications and storage systems.

POSSIBLE LONG TYPE QUESTIONS

1. Explain virtualization architecture?
2. What is the importance of resilience?
3. Write short note on
    a. Load balance
    b. Cloud governance
    c. Data centre
    d. Map reduces

CHAPTER NO.-05

VIRTUALISATION

Learning Objectives:

5.1 *Virtualization*

5.2 *Network Virtualization*

5.3 *Desktop and application Virtualisation*

5.4 *Desktop as a service*

5.5 *Local desktop Virtualisation*

5.6 *Virtualisation Benefits*

5.7 *Server Virtualisation*

5.8 *Block and File level Storage Virtualisation*

5.9 *Virtual Machine Monitor*

5.10 *Infrastructure Requirements*

5.11 *VLAN and VSAN*

5.1 VIRTUALIZATION

- It is the creation of virtual servers, infrastructures, devices and computing resources.
- Virtualization changes the hardware-software relations and is one of the foundational elements of cloud computing technology that helps utilize the capabilities of cloud computing to the full.

- Virtualization techniques allow companies to turn virtual their networks, storage, servers, data, desktops and applications.

5.2 NETWORK VIRTUALIZATION

- Network virtualization in cloud computing is a method of combining the available resources in a network by splitting up the available bandwidth into different channels, each being separate and distinguished.
- They can be either assigned to a particular server or device or stay unassigned completely all in real time.
- The idea is that the technology disguises the true complexity of the network by separating it into parts that are easy to manage, much like your segmented hard drive makes it easier for you to manage files.

5.2 DESKTOP AND APPLICATION VIRTUALISATION

Desktop virtualization

- As compared to other types of virtualizations in cloud computing, this model enables you to emulate a workstation load, rather than a server.
- This allows the user to access the desktop remotely.
- Since the workstation is essentially running in a data centre server, access to it can be both more secure and portable.

Application Virtualization

- Software virtualization in cloud computing abstracts the application layer, separating it from the operating system.
- This way the application can run in an encapsulated form without being dependent upon the operating system underneath.
- In addition to providing a level of isolation, an application created for one OS can run on a completely different operating system.

5.4 DESKTOP AS A SERVICE

- Desktop as a Service (DaaS) is a cloud computing offering where a service provider delivers virtual desktops to end users over the Internet, licensed with a per-user subscription.
- The provider takes care of backend management for small businesses that find creating their own virtual desktop infrastructure to be too expensive or resource-consuming.
- This management typically includes maintenance, back-up, updates, and data storage. Cloud service providers may also handle security and applications for the desktop, or users may manage these service aspects individually.
- There are two kinds of desktops are available in DaaS—persistent and non-persistent.

Persistent desktop: Users have the ability to customize and save a desktop so it will look the same way each time a particular user logs on. Persistent desktops require more storage than non-persistent desktops, which can make them more expensive.

Non-persistent desktop: Desktops are wiped each time the user logs out—they are merely a way to access shared cloud services.

Advantages of Desktop as a Service

Desktop as a Service offers some clear advantages over a traditional desktop model. Deploying or decommissioning active end users with DaaS is much faster and less expensive.

Faster deployment and decommissioning of active end users: The desktop is already configured, it just needs to be connected to a new device. For seasonal businesses that consistently experience spikes and drops in demand or employees, DaaS can save a lot of time and money.

Reduced downtime for IT support: Desktop as a Service also allows companies to provide remote IT support to their employees, reducing downtime.

Cost savings: Because the devices that run DaaS require much less computing power than a traditional desktop machine or laptop, they are less expensive and use less power.

Increased device flexibility: DaaS runs on a variety of operating systems and device types, which supports the trend of users bringing their own devices into the office and shifts the burden of supporting the desktop on all of those devices to the cloud service provider**.**

Enhanced security: Because the data is stored in the data centre with DaaS, security risks are considerably lower. If a laptop or mobile device is stolen, it can simply be disconnected from the service. Since none of the data lives on that stolen device, the risk of a thief accessing sensitive data is minimal. Security patches and updates are also easier to install in a DaaS environment because all of the desktops can be updated simultaneously from a remote location.

5.5 LOCAL DESKTOP VIRTUALISATION

- Local desktop virtualization is well suited for environments where continuous network connectivity cannot be assumed and where application resource requirements can be better met by using local system resources.
- However, local desktop virtualization implementations do not always allow applications developed for one system architecture to run on another.
- For example, it is possible to use local desktop virtualization to run Windows 7.

5.6 OVIRTUALISATION BENEFITS

Protection from System Failures

- Technology is always at the risk of crashing down at the wrong time. Businesses can tolerate a few glitches, but if your developer is working on an important application that needs to be finished immediately, the last thing you could wish for is a system crash.
- To counter this risk, virtualization lets you open the same work on another device. Store all your backup data through virtualization on cloud services or virtual networks and get easy access to it from any device.
- Apart from that, there are usually two servers working side-by-side keeping all your data accessible.
- If one faces any problem, the other is always available to avoid any interruption.

2. Hassle-free Transfer of Data

- You can easily transfer data from physical storage to a virtual server, and vice versa. Administrators don’t have to waste time digging out hard drives to find data.
- With a dedicated server and storage, it’s quite easy to locate the required files and transfer them within no time.
- You'll realize virtualization's actual worth when you’ll have to transfer data over a long-distance. You also have the choice of getting a virtual disk space.
- If you don’t need much space, you can opt for a thin-provisioned virtual disk.

3. Firewall and Security

- Security is a major aspect IT professionals have to focus on. However, with virtual firewalls, access to your data is restricted at much lower costs as compared to traditional methods.
- Through virtualization, you get protected by a virtual switch that protects all your data and applications from harmful malware, viruses, and other cyber threats.

- You are allotted the firewall feature for network virtualization to create segments within the system. Server virtualization storage on cloud services will save you from the risks of having your data get lost or corrupted.
- Cloud services are also encrypted with high-end protocols that protect your data from other various threats. So it's a good idea to virtualize all your storage and then create a backup on a server that you can store on cloud services.
- However, in order to ensure that you do this correctly, it's preferable to first go through a cloud computing online course, to avoid making any errors.

4. Smoother IT Operations

- Virtual networks help IT professionals become efficient and agile at work. These networks are easy to operate and process faster, reducing the effort and time required to work on them.
- Before virtual networks were introduced in the digital world, it would take days and weeks for technical workers to maintain and install devices and software on physical servers.
- Apart from the operations, visualization has also benefited IT support teams in solving technical problems in physical systems.
- As all the data is available on a virtual server, technicians don't have to waste time recovering it from crashed or corrupted devices.
- Learn all the skills behind virtualization with cloud training online, and become a successful technician.

5. Cost-Effective Strategy

- Virtualization is a great way to reduce operational costs. With all the data stored on virtual servers or clouds, there's hardly a need for physical systems or hardware, thus allowing businesses to witness a vast reduction in wastage, electricity bills, and maintenance costs.
- Virtualization also helps companies save a significant amount of space which can be utilized to increase the operations of a profitable department.
- This cost-effective strategy is both a profitability and productivity booster!

5.7 SERVER VIRTUALISATION

- Server virtualization is the process of dividing a physical server into multiple unique and isolated virtual servers by means of a software application. Each virtual server can run its own operating systems independently.

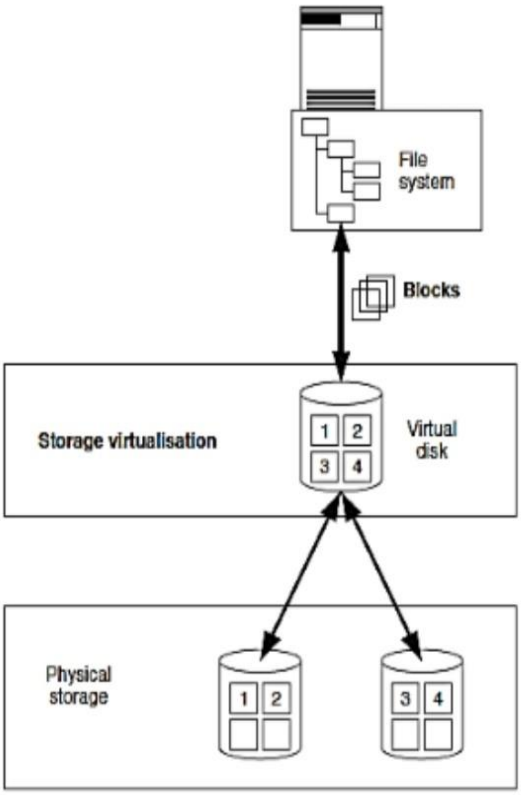
Advantages

Cost Reduction: Server virtualization reduces cost because less hardware is required.

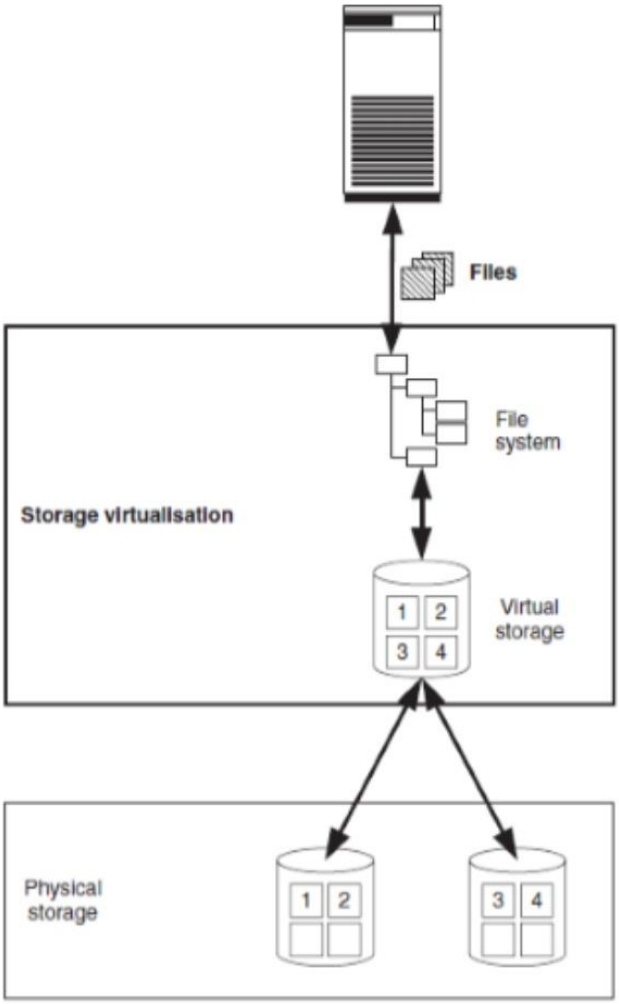
Independent Restart: Each server can be rebooted independently and that reboot won't affect the working of other virtual servers.

5.8 BLOCK AND FILE LEVEL STORAGE VIRTUALISATION

- Virtualisation on block level means that storage capacity is made available to the operating system or the applications in the form of virtual disks.
- In virtualisation on block level the task of file system management is the responsibility of the operating system or the applications
- The task of the virtualisation entity is to map these virtual blocks to the physical blocks of the real storage devices



- Virtualisation on file level means that the virtualisation entity provides virtual storage to the operating systems or applications in the form of files and directories
- The applications work with files instead of blocks and the conversion of the files to virtual blocks is performed by the virtualisation entity itself (This means, the task of file system management is performed by the virtualisation entity, unlike in block level which is done by OS or application).
- The physical blocks are presented in the form of a virtual file system and not in the form of virtual blocks.



5.9 VIRTUAL MACHINE MONITOR

- A Virtual Machine Monitor (VMM) is a software program that enables the creation, management and governance of virtual machines (VM) and manages the operation of a virtualized environment on top of a physical host machine.
- VMM is also known as Virtual Machine Manager and Hypervisor. However, the provided architectural implementation and services differ by vendor product.
- VMM is the primary software behind virtualization environments and implementations. When installed over a host machine, VMM facilitates the creation of VMs, each with separate operating systems (OS) and applications.
- VMM manages the backend operation of these VMs by allocating the necessary computing, memory, storage and other input/output (I/O) resources.

- VMM also provides a centralized interface for managing the entire operation, status and availability of VMs that are installed over a single host or spread across different and interconnected hosts.

5.10 INFRASTRUCTURE REQUIREMENTS

- In the proposed ontology, infrastructure requirements define the capabilities, features or qualities that are necessary (or desired) for an infrastructure on which to execute the application.
- Resource requirements describe the specifications of resources, such as hardware, software and operating system.
- Defining exactly what a cloud infrastructure is can be broad and complex. But when it comes down to it, a cloud-based infrastructure has several key components, including, but not limited to a combination of:

- Servers

- Software

- Network devices, and

- Other storage resources

It is these components, all of which are necessary to create applications that are then accessed via the cloud. These apps can be retrieved remotely over the internet, telecom services, WANs (wide area networks), and other network means.

Types:

1. Computing: The computing portion of the infrastructure is delivered by server racks in order to deliver cloud services for various services and partners.

2. Networking: To transfer data externally as well as between computer and storage systems, this part of the infrastructure relies on routers and switches.

3. Storage: A cloud infrastructure will likely need considerable storage often using a combination of hard disks and flash storage.

5.11 VLAN AND VSAN

VLAN (Virtual Local Area Network)

- A **virtual LAN** (**VLAN**) is any broadcast domain that is partitioned and isolated in a computer network at the data link layer .

- VLANs allow network administrators to group hosts together even if the hosts are not directly connected to the same network switch.

- Because VLAN membership can be configured through software, this can greatly simplify network design and deployment.

- Without VLANs, grouping hosts according to their resource needs the labour of relocating nodes or rewiring data links.

- VLANs allow devices that must be kept separate to share the cabling of a physical network and yet be prevented from directly interacting with one another.

- VLAN can be used to separate traffic within a business based on individual users or groups of users or their roles (e.g., network administrators), or based on traffic characteristics (e.g., low-priority traffic prevented from impinging on the rest of the network's functioning).

VSAN (Virtual Storage Area Network)

- A virtual storage area network (virtual SAN, VSAN or VSAN) is a logical representation of a physical storage area network (SAN).
- A VSAN abstracts the storage-related operations from the physical storage layer, and provides shared storage access to the applications and virtual machines by combining the servers' local storage over a network into a single or multiple storage pools.
- The use of VSANs allows the isolation of traffic within specific portions of the network. If a problem occurs in one VSAN, that problem can be handled with a minimum of disruption to the rest of the network. VSANs can also be configured separately and independently**.**
- VSAN is a logical partition in a storage area network. It allows traffic to be isolated within specific partitions of a storage area network.
- The use of multiple VSAN can make a system easier configure. It uses a software-defined approach that creates shared storage for virtual machines.

POSSIBLE SHORT TYPE QUESTIONS WITH ANSWERS

a. Define virtualization?

Ans - It is the creation of virtual servers, infrastructures, devices and computing resources. Virtualization changes the hardware-software relations and is one of the foundational elements of cloud computing technology that helps utilize the capabilities of cloud computing to the full. Virtualization techniques allow companies to turn virtual their networks, storage, servers, data, desktops and applications.

b. What is network virtualization?

Ans - Network virtualization in cloud computing is a method of combining the available resources in a network by splitting up the available bandwidth into different channels, each being separate and distinguished.

c. Define VMM.

Ans - A Virtual Machine Monitor (VMM) is a software program that enables the creation, management and governance of virtual machines (VM) and manages the operation of a virtualized environment on top of a physical host machine.

d. What are the advantages of server virtualization?

Ans - Advantages

Cost Reduction: Server virtualization reduces cost because less hardware is required.

Independent Restart: Each server can be rebooted independently and that reboot won't affect the working of other virtual servers.

e. What is Local desktop Virtualisation?

Ans -Local desktop virtualization is well suited for environments where continuous network connectivity cannot be assumed and where application resource requirements can be better met by using local system resources.

f. What is Application Virtualization?

Ans - Software virtualization in cloud computing abstracts the application layer, separating it from the operating system. This way the application can run in an encapsulated form without being dependent upon the operating system underneath.

g. What is Desktop as a service?

Ans - Desktop as a Service (DaaS) is a cloud computing offering where a service provider delivers virtual desktops to end users over the Internet, licensed with a per-user subscription.

POSSIBLE LONG TYPE QUESTIONS

1. What is virtualization and explain its types?
2. Difference between VLAN and VSAN?
3. Explain Block and File level Storage Virtualisation.
4. Explain the benefits of virtualization.
5. Explain the Advantages of Desktop as a Service?
6. Write short notes on:
    a. VMM
    b. Virtualization

CHAPTER NO.-06

CLOUD SECURITY

Learning Objectives:

6.1 *Cloud Security Fundamentals*

6.2 *Cloud Security Services*

6.3 *Design Principles*

6.4 *Secure Cloud software requirements*

6.4 *Secure Cloud software requirements*

6.5 *Policy implementation*

6.6 *Cloud Computing Security Challenges*

6.1 CLOUD SECURITY FUNDAMENTALS

- **Cloud Security** is defending the confidentiality(C), integrity(I) and availability(A) of enterprise assets (data, application, infrastructure), using **cloud** services, from an outside or inside threat.
- Cloud Security is using effective guardrails to ensure company assets (data, application, infrastructure) using cloud services can function as expected and respond to unexpected threats.

For the security folks, Cloud Security is defending the confidentiality(C), integrity(I) and availability(A) of enterprise assets (data, application, infrastructure), using cloud services, from an outside or inside threat.

For the non-security background, the above-mentioned CIA are the three triads of Information security. There are others considerations in the mix too e.g. Authentication, Authorisation etc. but, trust me, CIA is the most commonly used one to explain the risk around a threat.

6.2 CLOUD SECURITY SERVICES

Cloud security is a set of control-based safeguards and technology protection designed to protect resources stored online from leakage, theft, or data loss. ... **Security** applications operate as software in the **cloud** using a Software as a **Service** (SaaS) model.

- Identity and access.
- Data loss prevention.
- Web **security**.
- E-mail **security**.
- **Security** assessment.

- Intrusion management.
- **Security** information and managing events.
- Encryption.

1.Identity and access

- You are provided with control for secured management of identities and access. It includes people, processes and systems used for managing access to your enterprise resources. It is managed by making sure that the identity of the user is verified and the access rights are provided at the correct level.

2. Data loss prevention

- This service offers protection of data by providing you with pre-installed data loss prevention software, along with a set of rules deployed.

3. Web security

- Web security is provided as an additional protection against malware from entering the enterprise through web browsing and other such activities. This cloud service is provided either by installing a software or an appliance or through the cloud by redirecting your web traffic over to the cloud provider.

4. E-mail security

- It provides control over the in-bound and out-bound e-mails to protect your organization from malicious attachments and phishing. This cloud service helps enforce corporate policies such as acceptable use, spam and in providing business continuity options. One of the solution adopted by many cloud e-mail security services is digital signatures, which allows identification and non-repudiation.

5. Security assessment

- There are various tools implemented for the users of the SaaS delivery model, such as variant elasticity, low administration overhead, negligible setup time and pay-per use with low investment in the initial stage.

6. Intrusion management

- It is the process that uses pattern recognition for detection and reaction to events that are statistically unusual and unexpected.
  It may also require reconfiguration of your system components in real time so as to prevent an intrusion.

7. Security information and managing events

- Your system gathers information related to log and events. This information is used in correlating and analysing, to provide you with real time reporting and alerts on events that require intervention.

8. Encryption

- There are typical algorithms that are computationally difficult or nearly impossible to break.

9. Disaster management

- This cloud service helps in continuing your business and managing disasters by providing flexibility and reliable failover for services that are required in case of service interruptions.

10. Network security

- The network security services provide you with address security controls, which in a cloud environment is generally provided through virtual devices.

6.3 DESIGN PRINCIPLES

Establish the context before designing a system.
Make compromise difficult.
Make disruption difficult.
Make compromise detection easier.

Reduce the impact of compromise.

Security Design Principles

- Least Privilege
- Fail-Safe Defaults
- Economy of Mechanism
- Complete Mediation
- Open Design
- Separation Privilege
- Least Common Mechanism
- Psychological Acceptability
- Defence in Depth

6.4 SECURE CLOUD SOFTWARE REQUIREMENTS

- 1: Top-of-the-Line Perimeter Firewall.
- 2: Intrusion Detection Systems with Event Logging.
- 3: Internal Firewalls for Individual Applications, and Databases.
- 4: Data-at-Rest Encryption.
- 5: Tier IV Data Centres with Strong Physical **Security**.

Software-as-a-Service (SaaS) is probably the most well-known application for **cloud computing**. Essentially, SaaS products distribute data online, and are accessible from a browser on any device, which allows those companies to continue to host the **software**.

6.5 POLICY IMPLEMENTATION:

It involves translating the goals and objectives of a **policy** into an action. ... Some practical strategies are suggested to overcome **implementation** performance and concludes with the proposition that **implementation** failure is also due to lack of theoretical sophistication.

How to Develop and Implement a New Company Policy

- **Step** 1: Identify the Need for a **Policy**.
- **Step** 2: Determine **Policy** Content.
- **Step** 3: Obtain Stakeholder Support.
- **Step** 4: Communicate with Employees.
- **Step** 5: Update and Revise the **Policy**.

6.6 CLOUD COMPUTING SECURITY CHALLENGES

- Cloud computing is a term used to describe the use of hardware and software delivered via network (usually the Internet).
- The term comes from the use of cloud shaped symbol that represents abstraction of rather complex infrastructure that enables the work of software, hardware, computation and remote services.
- By using these types of services, businesses usually “rent” the capabilities of larger set of applications, reducing the need to buy, maintain or upgrade the software and infrastructure.
- End users access cloud-based applications usually through web browser or desktop/mobile application, while the data and computation are stored on remote servers (cloud).

POSSIBLE SHORT TYPE QUESTIONS WITH ANSWERS

1. Define cloud security fundamental.

Ans-

This involves using layers of **security** technologies and business practices to protect data and infrastructure against threats in multiple ways. With appropriate encryption mechanisms, data stored in the **cloud** can be protected even if access is gained by malicious or unauthorized personnel.

2.Define web security.

Ans –

Web security is provided as an additional protection against malware from entering the enterprise through web browsing and other such activities. This cloud service is provided either by installing a software or an appliance or through the cloud by redirecting your web traffic over to the cloud provider.

3.What is the function of network security?

Ans –

The network security services provide you with address security controls, which in a cloud environment is generally provided through virtual devices.

4.What is the function of e-mail security?

Ans –

It provides control over the in-bound and out-bound e-mails to protect your organization from malicious attachments and phishing.

This cloud service helps enforce corporate policies such as acceptable use, spam and in providing business continuity options.

5.Define Data loss prevention.

Ans –This service offers protection of data by providing you with pre-installed data loss prevention software, along with a set of rules deployed.

POSSIBLE LONG TYPE QUESTIONS

1. Explain cloud security services.
2. How to Develop and Implement a New Company Policy?
3. What are the Cloud Computing Security Challenges?

CHAPTER NO.-07
CLOUD COMPUTING
SECURITY ARCHITECTURE

Learning Objectives:

7.1 *Architectural Considerations*
7.2 *Information Classification*
7.3 *Virtual Private Networks*
7.4 *Public Key and encryption key management*
7.5 *Digital certificates*
7.6 *Key management*
7.7 *Memory card*
7.8 *Implementing Identity Management*
7.9 *Controls and Autonomic System*

7.1 ARCHITECTURAL CONSIDERATIONS

- Cloud security architecture is a strategy designed to secure and view an enterprise's data and collaboration applications in the cloud through the lens of shared responsibility with cloud providers.
- Cloud-enabled innovation is becoming a competitive requirement.
- As more enterprises seek to accelerate their business by shifting data and infrastructure to the cloud, security has become a higher priority.
- Operations and development teams are finding new uses for cloud services, and companies are searching for strategies to gain speed and agility.
- Enterprises must remain competitive by adding new collaborative capabilities and increasing operational efficiency in the cloud – while also saving money and resources.
- Security and risk management professionals are left with a patchwork of controls at the device, network, and cloud – with significant gaps in visibility to their data.

7.2 INFORMATION CLASSIFICATION

- Information classification is a process in which organisations assess the data that they hold and the level of protection it should be given.

- Organisations usually classify information in terms of confidentiality – i.e. who is granted access to see it.

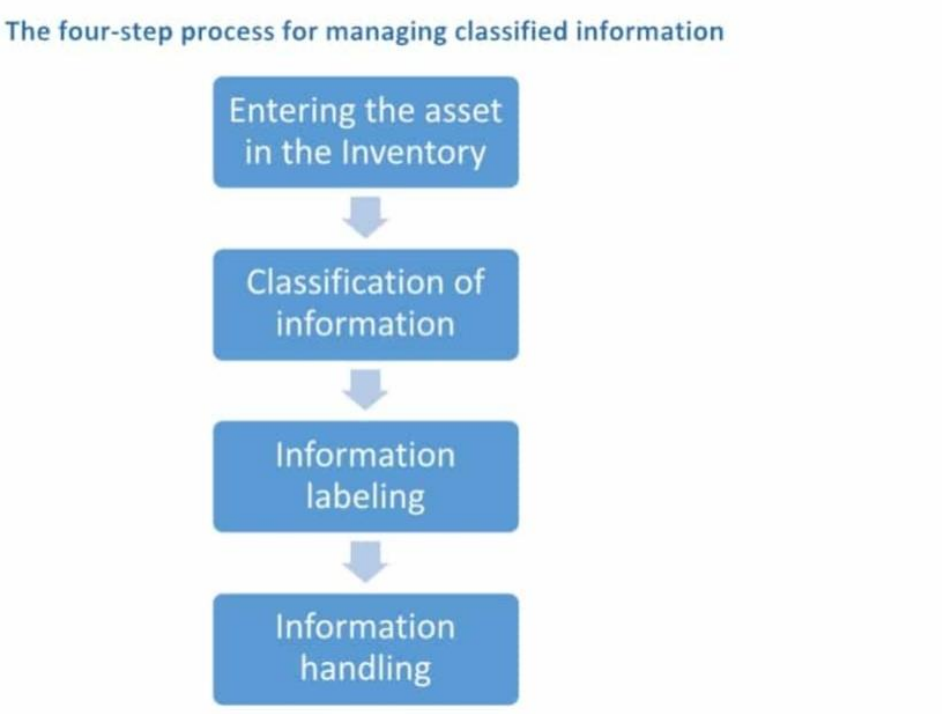
Classification of information

1.Confidential (top confidentiality level)

2.Restricted (medium confidentiality level)

3.Internal use (lowest level of confidentiality)

4.Public (everyone can see the information)



This means that:

(1) the information should be entered in the Inventory of Assets (control A.8.1.1 of ISO 27001),

(2) it should be classified (A.8.2.1), (3) then it should be labelled (A.8.2.2), and finally

(4) it should be handled in a secure way (A.8.2.3).

7.3 VIRTUAL PRIVATE NETWORKS

- A **virtual private network**, or **VPN**, is an encrypted connection over the Internet from a device to a **network**. The encrypted connection helps ensure that sensitive data is safely transmitted.
- It prevents unauthorized people from eavesdropping on the traffic and allows the user to conduct work remotely.
- The term virtual private network (abbreviated VPN) describes any technology that can encapsulate and transmit network data, typically Internet Protocol data, over another network.
- Such a system enables users to access network resources that may otherwise be inaccessible from the public internet.
- VPNs are frequently used in the information technology sector to provide access to resources for users that are not physically connected to an organization's network, such as telecommuting workers.
- VPNs are so named because they may be used to provide virtual (as opposed to physical) access to a private network.

7.4 PUBLIC KEY AND ENCRYPTION KEY MANAGEMENT

A **public-key** infrastructure is a type of **key management** system that uses hierarchical digital certificates to provide authentication, and **public keys** to provide encryption. PKIs are used in World Wide Web traffic, commonly in the form of SSL and TLS.

Distribution of public key

- Public-key encryption helps address key distribution problems.
- It has two aspects
- Distribution of public keys.
- Use of public-key encryption to distribute secret keys.

Encryption Key management

- Encryption is a process that uses algorithms to encode data as ciphertext.
- This ciphertext can only be made meaningful again, if the person or application accessing the data has the data encryption keys necessary to decode the ciphertext.
- So, if the data is stolen or accidentally shared, it is protected because it is indecipherable, thanks to data encryption.

Controlling and maintaining data encryption keys is an essential part of any data encryption strategy, because, with the encryption keys, a cybercriminal can return encrypted data to its original unencrypted state. An encryption key management system includes generation, exchange, storage, use, destruction and replacement of encryption keys.

- Can be considered as using one of:
- Public announcement
- Publicity available directory
- Public-key authority
- Public-key certificates

Types-

1.An HSM or other hardware key management appliance, which provides the highest level of physical security
2. A key management virtual appliance
3. Key management software, which can run either on a dedicated server or within a virtual/cloud server
4. Key Management Software as a Service (SaaS)

7.5 DIGITAL CERTIFICATES

Digital certificates are electronic credentials that bind the identity of the **certificate** owner to a pair of electronic encryption keys, (one public and one private), that can be used to encrypt and sign information digitally.

- All the receiver would know is that a valid key pair was used.
- The main purpose of the digital certificate is to ensure that the public key contained in the certificate belongs to the entity to which the certificate was issued, in other words, to

verify that a person sending a message is who he or she claims to be, and to then provide the message receiver with the means to encode a reply back to the sender.

- A Certificate Authority or CA then is a commonly trusted third party that is relied upon to verify the matching of public keys to identity, e-mail name, or other such information.
- Digital Certificates can be used for a variety of electronic transactions including e-mail, electronic commerce, groupware and electronic funds transfers.

7.6 KEY MANAGEMENT

- Key management refers to management of cryptographic keys in a cryptosystem. This includes dealing with the generation, exchange, storage, use, crypto-shredding (destruction) and replacement of keys. It includes cryptographic protocol design, key servers, user procedures, and other relevant protocols.
- Key management concerns keys at the user level, either between users or systems. This is in contrast to key scheduling, which typically refers to the internal handling of keys within the operation of a cipher.
- Successful key management is critical to the security of a cryptosystem. It is the more challenging side of cryptography in a sense that it involves aspects of social engineering such as system policy, user training, organizational and departmental interactions, and coordination between all of these elements, in contrast to pure mathematical practices that can be automated.

Use of key management

- Key management servers (KMS) are used to administer the full lifecycle of cryptographic keys and protect them from loss or misuse. KMS solutions, and other key management technology, ultimately control the generation, usage, **storage**, archival, and deletion of encryption keys.
- Encryption **key management** is crucial to preventing unauthorized access to sensitive information—if **keys** are compromised, entire systems and data can be compromised and rendered unusable until the situation is resolved. Different industries have different **requirements** for **key management**.

7.7 MEMORY CARD

- A memory card or memory cartridge is an electronic data storage device used for storing digital information, typically using flash memory.
- These are commonly used in portable electronic devices, such as digital cameras, mobile phones, laptop computers, tablets, PDAs, portable media players, video game consoles, synthesizers, electronic keyboards and digital pianos, and allow adding, as the card is usually contained within the device rather than protruding like USB flash drives.
- Secure Digital, officially abbreviated as SD, is a proprietary non-volatile memory card format developed by the SD Association (SDA) for use in portable devices.
- Memory cards or SD cards are small storage devices that are used to store the data backups such as the text, the pictures, audio, video, they are more compact and portable than CDs or DVD, and they can hold more data than a CD.

Advantages

1. Memory cards are reliable because they have no moving parts.
2. Memory cards have a non-volatile memory.
3. Memory cards are very portable, they can be used in small devices, lightweight and low power easily.
4. Memory cards come in all sorts of sizes.
5. Memory cards are used in various devices such as cameras, computers or mobile phones.
6. Memory card consumes very little power.

Disadvantages

1. Memory cards can easily break, they can be lost, misplaced or crushed easily, they can be affected by electronic corruption, and they make all the unreadable card, they are more expensive than CD or DVD, the metal part can be net or damaged if treated roughly broken.
2. Memory cards have prices and rewrite the boundaries, there is a finite amount of information that can be erased and written to memory cards.

7.8 IMPLEMENTING IDENTITY MANAGEMENT

- Identity management (ID management) is the organizational process for ensuring that individuals have the appropriate access to technology resources.
- More specifically, this includes the identifying, authentication and authorization of a person, or persons, to have access to applications, systems or networks.

- This is done by associating user rights and restrictions with established identities. Managed identities can also refer to software processes that need access to organizational systems.
- Identity management can be considered an essential component for security.
- The main goal of identity management is to ensure that only authenticated users are granted access to the specific applications, systems or IT environments for which they are authorized.
- This includes control over user provisioning and the process of onboarding new users such as employees, partners, clients and other stakeholders.
- Identity management also includes control over the process of authorizing system or network permissions for existing users and the offboarding of users who are no longer authorized to access organization systems.

Importance of identity management

- Identity management is an important part of the enterprise security plan, as it is linked to both the security and productivity of the organization.

- An identity and access management (IAM) system can provide a framework with the policies and technology needed to support the management of identities.

7.9 CONTROLS AND AUTONOMIC SYSTEM

Control system

- **Cloud management** is how administrators **control**—and orchestrate—all the products and services that operate in a **cloud**: the users and access **control**, data, applications, and services.

- A control system is a system of devices that manages, commands, directs or regulates the behaviour of other devices to achieve a desired result.

Automatic system

- Automation is the use of technology to perform tasks with reduced human assistance.

- Automation helps you accelerate processes and scale environments, as well as build continuous integration, continuous delivery, and continuous deployment (CI/CD) workflows.

- There are many kinds of automation, including IT automation, business automation, robotic process automation, industrial automation, artificial intelligence, machine learning, and deep learning.

- Hybrid and multicloud environments add an additional layer of complexity to infrastructure, network, application, and user administration.

- IT teams need to manage both on-site and cloud-based environments, often using specialized management tools for each.

- As a result, it can be nearly impossible to effectively maintain, track, scale, and secure resources and applications by hand.

- Automation can unite hybrid and multicloud management under a single set of processes and policies to improve consistency, scalability, and speed.

Examples of automation services from public cloud providers include:

- AWS Config, AWS CloudFormation, AWS EC2 Systems Manager;
- Microsoft Azure Resource Manager, Azure **Automation**;
- Google **Cloud** Composer, **Cloud** Deployment Manager; and.
- IBM **Cloud** Orchestrator.

Autonomic computing is a **computer's** ability to manage itself automatically through adaptive technologies that further **computing** capabilities and cut down on the time required by **computer** professionals to resolve system difficulties and other maintenance such as software updates.

POSSIBLE SHORT TYPE QUESTIONS WITH ANSWERS

1. Define cloud security architecture.

Ans –

Cloud security architecture is a strategy designed to secure and view an enterprise's data and collaboration applications in the cloud through the lens of shared responsibility with cloud providers.

2. **What is the use of key management?**

Ans –

Key management servers (KMS) are used to administer the full lifecycle of cryptographic keys and protect them from loss or misuse.

KMS solutions, and other key management technology, ultimately control the generation, usage, **storage**, archival, and deletion of encryption keys.

3. What are the advantages of memory card?

Ans-

- Memory cards are reliable because they have no moving parts.
- Memory cards have a non-volatile memory.
- Memory cards are very portable, they can be used in small devices, lightweight and low power easily.

4. What are the disadvantages of memory card?

Ans-

- Memory cards can easily break, they can be lost, misplaced or crushed easily, they can be affected by electronic corruption, and they make all the unreadable card, they are more expensive than CD or DVD, the metal part can be net or damaged if treated roughly broken.

5. Define digital certificate.

Ans –

Digital certificates are electronic credentials that bind the identity of the **certificate** owner to a pair of electronic encryption keys, (one public and one private), that can be used to encrypt and sign information digitally.

6. Define public key encryption.

Ans-

- A **public-key** infrastructure is a type of **key management** system that uses hierarchical digital certificates to provide authentication, and **public keys** to provide encryption. PKIs are used in World Wide Web traffic, commonly in the form of SSL and TLS.

7. What is VPN?

Ans-

- A **virtual private network**, or **VPN**, is an encrypted connection over the Internet from a device to a **network**.
- The encrypted connection helps ensure that sensitive data is safely transmitted.

8. What is the classification of information?

Ans –

Classification of information

1.Confidential (top confidentiality level)

2.Restricted (medium confidentiality level)

3.Internal use (lowest level of confidentiality)

4.Public (everyone can see the information)

POSSIBLE LONG TYPE QUESTIONS

2. Explain the key management technique.
3. Explain advantages and disadvantages of memory card.
4. What do you mean by automatic system?
5. Explain the implementation of identity management?
6. Write short notes on

.6.1 VPN

.6.2 Encryption

.6.3 Digital certificate

CHAPTER NO.- 08

MARKET BASED MANAGEMENT OF CLOUDS

Learning Objectives:

8.1 *Cloud Information security vendors*
8.2 *Cloud Federation, characterization*
8.3 *Cloud Federation stack*
8.4 *Third Party Cloud service*
8.5 *Case study*

8.1 CLOUD INFORMATION SECURITY VENDORS

- **Cloud security** is the protection of data stored online via **cloud** computing platforms from theft, leakage, and deletion.
- Methods of providing **cloud security** include firewalls, penetration testing, obfuscation, tokenization, virtual private networks (VPN), and avoiding public internet connections.
- Microsoft, IBM, and Amazon are the top **companies** that are popular for their cloud and other services. They are also the provider of **cybersecurity** services.

8.2 CLOUD FEDERATION, CHARACTRIZATION

- Cloud Federation refers to the unionization of software, infrastructure and platform services from disparate networks that can be accessed by a client via the internet.
- It is important to note that federated cloud computing services still rely on the existence of physical data centres.

Benefits of a Cloud Federation

- Increased security and control.
- Reduction in IT costs to support growing **federation** infrastructure.

- Eliminates federated application deployments.
- Audit and compliance using our **Cloud** Reporting solution that tracks all activity in pretty charts/graphs/etc.

The four centric of the federated cloud are customer, business, provider, service. The federated cloud architecture and mechanism are designed prioritizing the customer.

8.3 CLOUD FEDERATION STACK

- Cloud federation requires one provider to wholesale or rent computing resources to another cloud provider.
- Those resources become a temporary or permanent extension of the buyer's cloud computing environment, depending on the specific federation agreement between providers.
- Cloud federation offers two substantial benefits to cloud providers.
- First, it allows providers to earn revenue from computing resources that would otherwise be idle or underutilized.
- Second, cloud federation enables cloud providers to expand their geographic footprints and accommodate sudden spikes in demand without having to build new points-of-presence (POPs).

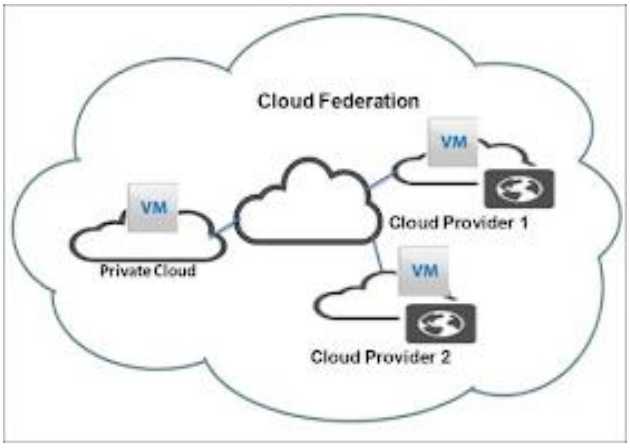


Fig. Cloud Federation Stack

8.4 THIRD PARTY CLOUD SERVICE

- A **cloud service** provider is a **third-party** company offering a **cloud**-based platform, infrastructure, application or **storage services**.
- Much like a homeowner would pay for a utility such as electricity or gas, companies typically have to pay only for the amount of **cloud services** they use, as business demands require.
- While these third-party cloud services may be easier to use, more full-featured and sometimes cheaper than their public cloud counterparts, there are drawbacks to using them.
- They are their own point of application failure. As a result, IT teams could suffer an outage both due to the cloud provider as well as their third-party service provider.
- This is especially a concern for critical services like authentication.

Advantages

1. Maintenance and support
2. Skilled company with all the resources
3. More secure
4. less cost

Disadvantages

1. Lack of control
2. Potential cost drawback

8.5 CASE STUDY

1. **Google app engine**

    - Google App Engine is a cloud computing platform as a service for developing and hosting web applications in Google-managed data centres.
    - Applications are sandboxed and run across multiple servers. **Microsoft agene.**

2. **Amazon Web Services**

    - **AWS** is made up of so many different cloud computing products and **services**.
    - The highly profitable **Amazon** division provides servers, storage, networking, remote computing, email, mobile development, and security.

3. **Hadoop (develop by Apache)**

- Apache Hadoop software is an open-source framework that allows for the distributed storage and processing of large datasets across clusters of computers using simple programming models.
- In this way, Hadoop can efficiently store and process large datasets ranging in size from gigabytes to petabytes of data.

4. **Aneka**

- **Aneka** is an Application Platform-as-a-Service (**Aneka** PaaS) for Cloud Computing.
- It acts as a framework for building customized applications and deploying them on either public or private Clouds.

POSSIBLE SHORT TYPE QUESTIONS WITH ANSWERS

1. What is the need of cloud security?

Ans-

- **Cloud security** is the protection of data stored online via **cloud** computing platforms from theft, leakage, and deletion.
- Methods of providing **cloud security** include firewalls, penetration testing, obfuscation, tokenization, virtual private networks (VPN), and avoiding public internet connections.

2. Define cloud federation?

Ans-

- Cloud Federation refers to the unionization of software, infrastructure and platform services from disparate networks that can be accessed by a client via the internet.

3. **What are the advantages of third-party cloud service provider?**

Ans-

1. Maintenance and support
2. Skilled company with all the resources
3. More secure
4. Less cost

4. **What is the benefit of cloud federation stack?**

Ans-

- It allows providers to earn revenue from computing resources that would otherwise be idle or underutilized.
- Cloud federation enables cloud providers to expand their geographic footprints and accommodate sudden spikes in demand without having to build new points-of-presence (POPs).

5. What is the benefit of cloud federation?

Ans-

- Increased security and control.
- Reduction in IT costs to support growing **federation** infrastructure.
- Eliminates federated application deployments.

POSSIBLE LONG TYPE QUESTIONS

1. Explain cloud federation and its benefits.
2. What is the working of third-party cloud services? Give some advantages and disadvantages?
3. Explain case study of Market Based Management of Clouds?

CHAPTER NO.-09

HADOOP

Learning Objectives:

9.1 *Introduction*

9.2 *Data source*

9.3 *Data storage and Analysis*

9.4 *Comparison with other system*

9.1 INTRODUCTION

- **Hadoop** is an open-source software framework for storing data and running applications on clusters of commodity hardware.
- It provides massive storage for any kind of data, enormous processing power and the ability to handle virtually limitless concurrent tasks or jobs.
- Apache **Hadoop** software is an open-source framework that allows for the distributed storage and processing of large datasets across clusters of **computers** using simple programming model.
- ln this way, **Hadoop** can efficiently store and process large datasets ranging in size from gigabytes to petabytes of data.

9.2 DATA SOURCE

- To perform data subset, masking, and discovery operations, you must import source metadata into the TDM repository.
- You can import sources from the PowerCenter repository or from a source database.
- To perform data generation operations, you must import target metadata into the TDM repository.
- When you create a project, add one or more sources to the project.
- You can add more than one type of source to the project.

example

You can add a flat file source and a relational source to the project. You can create constraints to create relationships between the sources and apply filter criteria for data subset and data masking.

9.3 DATA STORAGE AND ANALYSIS

Data storage-

- HDFS exposes a file system namespace and allows user data to be stored in files.
- Internally, a file is split into one or more blocks and these blocks are stored in a set of Data Nodes.
- The NameNode executes file system namespace operations like opening, closing, and renaming files and directories.

Format of data storage

- Text/CSV. A plain text file or CSV is the most common format both outside and within the **Hadoop** ecosystem.
- Sequence File. The Sequence File format stores the **data** in binary format.
- Avro. Avro is a row-based storage format.
- Parquet.
- RCFile (Record Columnar File)
- ORC (Optimized Row Columnar)

Data analysis

- Hadoop is an open-source software framework that provides for processing of large data sets across clusters of computers using simple programming models.
- Hadoop is designed to scale up from single servers to thousands of machines.

Analysing Big Data with Hadoop

- **Big Data** is unwieldy because of its vast size, and needs tools to efficiently process and extract meaningful results from it.
- **Big Data** is a term used to refer to a huge collection of **data** that comprises both structured **data** found in traditional databases and unstructured **data** like text documents, video and audio.

9.4 COMPARISON WITH OTHER SYSTEM

- Unlike RDBMS, Hadoop is not a database, but rather a distributed file system that can store and process a massive amount of data clusters across computers.

RDBMS	Hadoop
1. Traditional row-column based databases, basically used for data storage, manipulation and retrieval.	1. An open-source software used for storing data and running applications or processes concurrently.
2. In this structured data is mostly processed.	2. In this both structured and unstructured data is processed.
3. It is best suited for OLTP environment.	3. It is best suited for BIG data.
4. It is less scalable than Hadoop.	4. It is highly scalable.
5. Data normalization is required in RDBMS.	5. Data normalization is not required in Hadoop.
6. It stores transformed and aggregated data.	6. It stores huge volume of data.
7. It has no latency in response.	7. It has some latency in response.
8. The data schema of RDBMS is static type.	8. The data schema of Hadoop is dynamic type.
9. High data integrity available.	9. Low data integrity available than RDBMS.
10. Cost is applicable for licensed software.	10. Free of cost, as it is an open-source software.

POSSIBLE SHORT TYPE QUESTIONS WITH ANSWERS

1. Define Hadoop?

Ans – Apache **Hadoop** software is an open-source framework that allows for the distributed storage and processing of large datasets across clusters of **computers** using simple programming model.

2. Define Data analysis in Hadoop?

Ans-

- Hadoop is an open-source software framework that provides for processing of large data sets across clusters of computers using simple programming models.
- Hadoop is designed to scale up from single servers to thousands of machines.

3. How is data stored in Hadoop?

Ans-

- HDFS exposes a file system namespace and allows user data to be stored in files.
- Internally, a file is split into one or more blocks and these blocks are stored in a set of Data Nodes.
- The Name Node executes file system namespace operations like opening, closing, and renaming files and directories.

4. How Hadoop is different from another database?

Ans-

- Hadoop is not a database, but rather a distributed file system that can store and process a massive amount of data clusters across computers.
- RDBMS is a structured database approach in which data is stored in rows and columns which can be updated with SQL and presented in different tables.

POSSIBLE LONG TYPE QUESTIONS

1. Define Hadoop. Explain how data are stored and analysis in it.
2. Explain how hadoop is differ from RDBMS.
3. Explain the data source in hadoop.