

1

Introduction

Important Points to Remember

- Cloud computing is a technology that delivers many kinds of resources as services, mainly over the internet
- Cluster computing is a group of linked computers, working together closely thus in many respects forming a single computer
- SOA define as a loosely-coupled architecture designed to meet the business needs of the organization
- Five essential characteristics of Cloud Computing are on demand self-service, Broad network access, Resource Pooling, Rapid Elasticity and Measured service
- NIST five essential characteristics of Cloud Computing are as On demand self-service, Broad network access, Resource pooling, Rapid Elasticity and Measured service
- SaaS provider dispose the applied software unified on their server, the user can subscribe applied software service from the manufacturer through Internet
- PaaS takes develop environment as a service to supply. This layer provides a platform for creating applications.
- Cloud enabling technologies are Broadband networks and internet architecture, Data center technology, Virtualization technology, Web technology and Multitenant technology
- Cloud service is any service made available to users on demand via the Internet from a cloud computing provider's servers as opposed to being provided from a company's own on-premises servers.
- Dynamic infrastructure platform is a system that provides computing resources, particularly servers, storage, and networking, in a way that they can be programmatically allocated and managed.

- Q.1 Define distributed system. What are the major milestones to lead cloud computing ?**
- Ans. : A distributed system is a collection of independent computers that appears to its users a single coherent system.
- A distributed system can consist of any number of possible configurations, such as mainframes, personal computers, workstations, minicomputers, and so on.
 - The major milestones have led to cloud computing are mainframes computing, cluster computing, and grid computing

Q.2 Differentiate between distributed computing and cloud computing. [RGPV : Dec-15, Marks 3]

Ans. : Cloud computing is a technology that delivers many kinds of resources as services, mainly over the internet, while distributed computing is the concept of using a distributed system consisting of many self-governed nodes to solve a very large problem.

- Cloud computing is basically a sales and distribution model for various types of resources over the internet, while distributed computing can be identified as a type of computing which uses a group of machines to work as a single unit to solve a large scale problem.

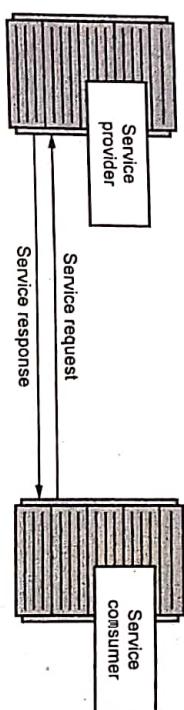


Fig. Q.5.1

- Web services currently provide the main enabling technique for Service Oriented Architecture. The Web Service technique can function both as a middleware and a modeling and management tool for composed business processes.
- Web-based applications that dynamically interact with other Web applications using open standards that include XML, UDDI and SOAP.
- SOA and cloud computing are complementary activities, and both will play important roles in IT planning for senior leadership teams for years to come.
- Cloud computing and SOA can be pursued independently, or concurrently, where cloud computing platform and storage service offerings can provide a value-added underpinning for SOA efforts.

Q.8 What are the properties of cloud computing ? [RGPV : June-16, Marks 2]

OR What are the essential characteristics of cloud computing ? [RGPV : Dec-15,16, Marks 2]

Ans. : Five essential characteristics of Cloud Computing are on demand self-service, Broad network access, Resource Pooling, Rapid Elasticity and Measured service.

Q.9 Define NIST definition of cloud computing:
Ans. : NIST definition of cloud : Cloud computing is a pay-per-use model for enabling available, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, services) that can be rapidly provisioned and released with minimal management effort or service-provider interaction.

1.2 : Vision of Cloud Computing

Q.7 What is trusted cloud computing ?

[RGPV : Dec-16, Marks 2]

Ans. : Clients of cloud computing services currently have no means of verifying the confidentiality and

Q.10 Discuss the vision of cloud computing ?

[RGPV : June-16, Marks 7]

Ans. : • In the near future we can image that it will be possible to find the solution that matches with our requirements by simply entering one request in a

Cloud Computing

- **Broad Network Access :** Cloud capabilities are available over the network and accessed through standard mechanism that promote use by heterogeneous client such as mobile phone, laptop.
- **Measured service :** Cloud systems automatically control and optimize resource use by leveraging a metering capability at some level of abstraction appropriate to the type of service. Resource usage can be monitored, controlled, and reported, providing transparency for both the provider and consumer of the utilized service.

Q.13 Write brief note on cloud computing challenges. [ST][RGV : June-15, Marks 7]

OR Discuss the different barriers of cloud computing. [ST][RGV : Dec-13, Marks 7]

Ans. :

1. Increased Security Vulnerabilities
2. Reduced Operational Governance Control
3. Limited Portability Between Cloud Providers
4. Multi-Regional Compliance and Legal Issues

- Use of cloud for business purpose means that the responsibility over data security becomes shared with the cloud provider. Organization extends their trust boundary to cloud consumer to external cloud.
- It is clear that the security issue has played the most important role in hindering cloud computing acceptance.
- Without doubt, putting your data, running your software on someone else's hard disk using someone else's CPU appears daunting to many.
- Well-known security issues such as data loss, phishing, pose serious threats to organization's data and software.

Q.14 Discuss the advantages and disadvantages of cloud computing. Also write down the limitation of cloud computing. [ST][RGV : Dec-17, Marks 7]

Ans. :

- 1. Lower computer costs : Since applications run in the cloud, not on the desktop PC, your desktop PC does not need the processing power or hard disk space demanded by traditional desktop software.

Limitation of cloud computing :

1. Peripherals : Peripheral devices like printers or scanners might not work with cloud. Many of them require software to be installed locally. Networked peripherals have lesser problems.
2. Integration : Integrating internal applications with those on cloud can be complex and in some cases not viable.
3. Generic : Public cloud offerings are very generic and offer multi-tenancy service which all organizations might not be comfortable with.

1.3 : Cloud Computing Reference Model

Q.15 Explain cloud computing reference model. [ST][RGV : Dec-16, Marks 7]

Ans. : Fig. Q15.1 shows cloud computing reference model.

Software as a Service

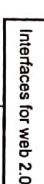
- SaaS provider dispose the applied software unified on their server, the user can subscribe applied software service from the manufacturer through Internet.
- Provider supply software pattern through Browser, and charge according to the quantity of software

Infrastructure As A Service (IaaS) :

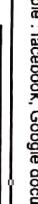
- In this layer, servers, network devices, and storage disks are made available to organizations as services on a need-to basis.
- IaaS takes infrastructure which is made of many servers as a measurement service to the customers.

Platform as a Service (PaaS) :

- Google App Engine is the representative product through their server and Internet.
- With PaaS, developers can build Web applications without installing any tools on their computers and then deploy those applications without any specialized systems administration skills.



- Scientific applications, end user applications
- Social networking
- Example : facebook, Google documents, Flickr



- Runtime environment for applications
- Development and data processing platforms
- Examples : Windows Azure, Hadoop, Aneka



- Virtualized servers
- Storage and networking
- Examples : Amazon EC2, S3, vCloud

Fig. Q.15.1 : Cloud computing reference model

- It integrates memory and I/O devices, storage and computing ability into a virtual resources pool, and provides storage resources and virtualization service for the whole industry.
- This is a way of hosted hardware, and the customer pays when they use the hardware.
- For example, Amazon Web Service and IBM Blue Cloud all rent the infrastructure as a service.

Q.16 Explain difference between IaaS, PaaS and SaaS.

Ans. :

IaaS	PaaS	SaaS
IaaS gives users automated and scalable environments	PaaS provides a framework for quickly developing and deploying applications	SaaS makes applications available through the internet.
Amazon Web Services, for example, offers IaaS through the Elastic Compute Cloud, or EC2	Google Cloud Platform provides another PaaS option in App Engine	SaaS applications such as Gmail, Dropbox, Salesforce, or Netflix
In IaaS, infrastructure as a service.	In PaaS, platform as a service	In SaaS, software as a service
Virtual platform on which required operating environment and application deployed	Operating environment included	Operating environment largely irrelevant, fully functional application provided
IaaS is a cloud service that provides basic computing infrastructure, servers, storage, and networking resources. In other words, IaaS is a virtual data center.	PaaS refers to cloud platforms that provide runtime environments for developing, testing, and managing applications	SaaS allows people to use cloud-based web applications.

Q.17 What makes cloud computing so interesting to IT stakeholders and research practitioners ? Explain it by listing essential cloud computing environment and cloud service requirements.

[RGPU : Dec-14, Marks-7]

Ans. : Cloud computing so interesting to IT stakeholders and research practitioners because of following reasons :

- Capacity planning :** Storage capacity is one of the main reasons for organization using cloud. Capacity planning is an unavoidable responsibility for most IT organizations. Future demands from business need to be planned for and accommodated. This can be very challenging because this involves estimating the usage and specially usage fluctuations over time. So there is constant need to balance peak usage requirements without unnecessarily over-spending on on-premise IT infrastructure.
- Cost reduction and operating overhead :** For any organization, initial investment of cloud is huge. The growth of IT environments often corresponds to the assessment of their maximum usage requirements. This can make the support of new and expanded business automation an ever-increasing investment.

- Organizational agility :** From cloud perspective IT organizations, the IT resources needs to be more available and/or reliable than previously thought. The ability for an IT organization to be able to respond to these changes in capacity or availability helps to increase an organizational agility.

Major IaaS providers include Amazon Web Services, Microsoft Azure, and Google Compute Engine.	Examples of PaaS services are available on a pay-for-what-you-use model	Examples of SaaS services are available with a pay-as-you-go pricing model.
Amazon Web Services, Heroku and Google App Engine.	PaaS solutions are available with a pay-as-you-go pricing model.	SaaS services are usually available with a pay-as-you-go pricing model.

Unit 10 Make Learning less Scary / Impressive

Action is the reason or fact of choosing to take action up follow we something.

Introduction

Cloud Computing

Introduction

Cloud Computing

1-10

Introduction

Cloud Computing

1-9

Introduction

increased demands. This is one of the most valuable and predominant feature of cloud computing. Through scalability you can scale up your data storage capacity or scale it down to meet the demands of your growing business.

- Archiving : it is a cost-effective, cloud-based service that moves unimportant email and file data out of your data centre.

1.6 : Cloud and Dynamic Infrastructure

Q.22 What is a Dynamic Infrastructure Platform?

Ans. : A dynamic infrastructure platform is a system that provides computing resources, particularly servers, storage, and networking, in a way that they can be programmatically allocated and managed.

Q.23 Explain cloud and dynamic infrastructure.

[RGPV : Dec-14, Marks 7]

Ans. : • Cloud dynamic infrastructure depends on an architecture that contains following:

1. Asset management : provide maximum value of important business and IT assets over their life cycle with industry tailored asset management solution.
2. Service management : this facility includes visibility, automation and control to delivering the first class IT service.
3. Virtualization and Consolidation : Resource are getting utilize more and more efficiently and operating cost of the system getting down.
4. Energy efficiency : IT infrastructure and organization sustainable. It is not likely to damage or effect any other things
5. Information infrastructure : It helps to business organization to achieve information compliance, availability, retention and security.
6. Security : Cloud infrastructure is responsible for risk management, customizing the governance.
7. Resilience : infrastructure is safe from all side

[RGPV : Dec-14, Marks 7]

Ans. : Infrastructure as a Service (IaaS)

• Examples : Amazon EC2, Rackspace Mosso, GoGrid

IaaS Server Types :

1. Physical server : Actual hardware is allocated for the customer's dedicated use.
2. Dedicated virtual server : The customer is allocated a virtual server, which runs on a

Ans. : Infrastructure as a Service (IaaS)

- IaaS gives the storage room likeness to the in-house datacenter stood out from various organizations sorts.
- Center datacenter framework segments are capable, servers (registering units), the system itself, and administration apparatuses for foundation upkeep and checking.

• Each of these parts has made a different market specialty. While some little organizations have practical experience in just a single of these IaaS cloud specialties, vast cloud suppliers like Amazon or Right Scale have offerings over all IaaS territories.

- Fig. Q.24.1 shows IaaS.
- 1. Client
- 2. Computer
- 3. Data storage device
- 4. Physical server

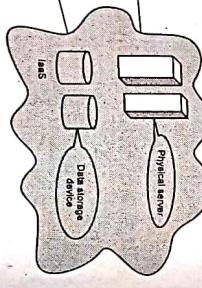


Fig. Q.24.1 IaaS

• It offers the hardware so that your organization can put whatever they want onto it. Rather than purchase servers, software, racks, and having to pay for the datacenter space for them, the service provider rents those resources.

1. Server space
2. Network equipment
3. Memory
4. CPU cycles
5. Storage space
6. Again, the customer is not managing cloud infrastructure, but in this case, the customer does control operating systems, deployed applications, storage, and sometimes-certain networking components

[RGPV : June-15, Dec-14, Marks 7]

Rudiments.

Q.25 What is cloud adoption?

Ans. : Cloud Adoption is a strategic move by organisations of reducing cost, mitigating risk and achieving scalability of data base capabilities. Cloud adoption may be up to various degrees in an organisation, depending on the depth of adoption. In fact the depth of adoption yields insight into the maturity of best practices, enterprise-ready cloud services availability.

- 1. Self service : Self service is facility provided by cloud to consumers.
- 2. This support the account- owners signing up and being able to use the purchased capacity

Q.26 Discuss the benefits of cloud adoption and rudiments.

Ans. : Benefits of this strategy include the ability to store privileged data in the private cloud while leveraging the technological resources from the public cloud to run applications relying on private data.

Q.27 Write a brief notes on cloud adoption and rudiments.

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.

• Marketing and Advertising : In an industry dependent on social media, as well as the quick creation and publishing of customer-relevant content, agencies are using hybrid cloud adoption strategies to deliver critical client messages to their local and worldwide audiences.

• Retail : A successful e-commerce strategy requires a sound Internet strategy. With the help of cloud adoption, Internet-based retail is able to effectively

Cloud rudiments :

- Resource aggregation and integration : Cloud solution integrates or aggregates the information of virtualization management. Physical server provisioning, system management.
- After that the integrated information will be sent into a central logical view.
- Application services : Here app services states that the services related to a particular software.
- The application instances represents the agreement between service provider and the consumer to use services on On-Demand basis.

1-10

Introduction

Cloud Computing

1-9

Introduction

Cloud Computing

1-10

Introduction

Cloud Computing

1-9

Introduction

Cloud Computing



Cloud Computing

1-11

market to customers and save their product data for less money.

- Finance : Efficient expense management, human resources, and customer communications are three of the most important business needs of today's finance organizations. For these reasons, financial services institutions are now placing their email platforms and marketing tools in the cloud

Q.29 Illustrate the cloud adoption discussing its several policies. [RGV : Dec-17, Marks 7]

- Ans. • Assessment : Executives and IT decision makers must assess the opportunities and challenges of employing a cloud computing strategy in their marketplace. In addition to researching popular vendors within their industry, business leaders and their technology teams should glean data on the challenges and successes of past adopters in their space.
- Planning : Once organizations do their research, they must plan their specific cloud strategy. IT leaders should choose platforms and services that are well-known to their industry and quick to market. They should also decide between a public, private, or hybrid cloud.
- Adoption : During the adoption phase, IT leaders should develop risk mitigation strategies. They

Introduction

1-12

should also have an expert understanding of their servers, software, and data stores for the future reiteration and scalability of their strategy.

- Optimization : By meeting regularly with their executive team, IT departments can discuss lessons learned in their cloud computing strategy and create new and improved solutions for further processes and tasks.

1.8 : Overview of Cloud Applications

Q.30 How satellite image processing uses cloud computing ? [RGV : Dec-16, Marks 3]

- Ans. • Satellite image processing plays a vital role for research and developments in Astronomy, Remote Sensing, GIS, Agriculture, Monitoring, Disaster Management and many other fields of study. Satellite images are recorded in digital forms and then processed by the computers to extract information.
- Satellite remote sensing generates hundreds of giga-bytes of raw images that need to be further processed to become the basis of several different GIS products. Fig. Q.30.1 shows cloud environment for satellite data processing.
- This process requires both I/O and compute-intensive tasks. Large images need to be moved from a ground station's local storage to

Introduction

1-12

compute facilities, where several transformations and corrections are applied.

- Cloud computing provides the appropriate infrastructure to support such application scenarios.
- Satellite data processing consists of various methods to correct the radiometric errors and geometric distortions in the basic data generated by the sensor; this data is termed as Level-0.

• The procedures like georeferencing and registration applied on the Level-0 data to generate the products such as:

1. Level 1 - Radio metrically corrected and geometrically corrected only for earth rotation (Browse product)
2. Level 2 - Both radiometric and geometrically corrected (Standard product)

Q.31 How cloud computing can help in solving Gene Expression Data Analysis ? Explain in details. [RGV : Dec-15, Marks 7]

- Ans. • Gene expression technology using DNA microarrays, allows for the monitoring of the expression levels of thousands of genes at once. As a direct result of recent advances in DNA microarray technology, it is now feasible to obtain gene expression profiles of tissue samples at relatively low costs.

• Gene expression profiles provide important insights into, and further our understanding of, biological processes. As such, they are key tools used in medical diagnosis, treatment and drug design.

• Cloud-CoXCS is a machine learning classification system for gene expression datasets on the Cloud infrastructure.

• It is composed of three components: CoXCS, Aneka, and Offspring.

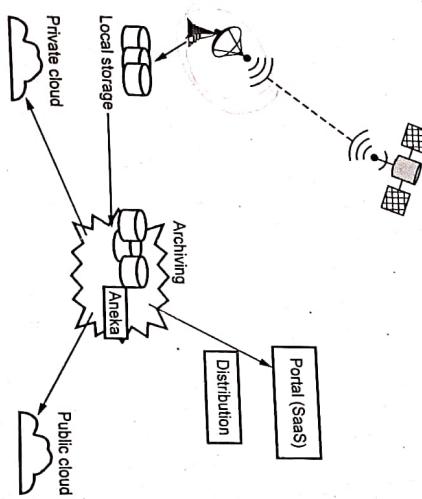


Fig. Q.30.1

Cloud Computing

1-12

• The model uses a modified covering operator and crossover operators, which improves the generation of new classifiers during the evolutionary process.

- After a fixed number of iterations, selected classifiers from each of the independent populations are transferred to a different population. The evolutionary cycle is then repeated.

W.E.F. 20
CSE)
DING SY

• Offspring : Offspring is a software tool that allows scientists and developers to quickly prototype distributed applications.

Q.32 Explain the steps in implementing cloud computing based applications.

[RGV : Dec-15, Marks 7]

- Ans. :
1. Define your project - Some applications and infrastructures should never be put on a cloud. Decide what you want to move to the cloud and whether or not it's feasible.
 2. Select the platform - Choose a platform that is fast, easy and safe to deploy. Ensure you have a flexible platform that scales to support your evolving business model and future growth.
 3. Understand security policies - Many service providers believe that data security is your responsibility, not theirs. Make sure you have a clear understanding of who is responsible and ensure that the right resources are in place.

4. Select your cloud computing service provider - Partner with a service provider that has success with businesses similar to yours and knows your technology.

5. Determine service level agreements - In addition to uptime, be very clear with your service provider when it comes to SLAs and exactly what they do and do not cover, such as data availability or data protection.
6. Understand who owns recovery - Outages will happen, so know in advance if you or your service provider is responsible for recovery.

7. Migrate in phases - Roll out a phased migration that allows you to gradually increase the load and gives you time to fine tune and minimize risks while maintaining business continuity.
8. Think ahead - Your business requirements can change at any time, so choose a cloud solution that allows you to move between on-premise and cloud as needed, and one that allows you to move to a different cloud service provider if necessary.

- Creating and implementing a cloud strategy takes time, energy and effort. It's important to choose the right cloud strategy that can help you open up new market opportunities, grow your business and increase customer loyalty.

Q.33 How does the cloud computing help to reduce the times to market applications and to cut down capital expenses?

[RGPV : June-15, Marks 7]

Ans. :

- Cloud computing is often cheaper and less labor-intensive for companies too. There is no need to buy and install expensive software because it's already installed online remotely and you run it from there, not to mention the fact that many cloud computing applications are offered free of charge. What that means ? Cloud computing reduces capital expenditures and operational overhead.
- Cloud solutions offer virtually unlimited storage space compared to server and hard drive limits. Needing more storage space does not cause issues with server upgrades and equipment. Usually all you need to do is increase your monthly fee slightly for more data storage.
- With cloud computing, you subscribe to the software, rather than buying it outright. This means that you only need to pay for it when you need it, and it also offers flexibility: it can be quickly and easily scaled up and down according to demand. Plus, online storage and back-up means their data can't be lost or destroyed.
- What that means ? Cloud computing provides greater business flexibility through an on-demand, pay-as-you-go model that scales with your business.

- Cloud backup solutions play an essential part of a comprehensive backup, recovery and business continuity strategy.
- In the event of a disaster, be it natural or technical, businesses must be confident that their data was protected, and that it can be restored in a timely manner to ensure smooth running of the business without any disruption. When companies experience downtime, they stand to lose a great deal of money.

- Cloud computing allows you to better control your downtime. By migrating to the cloud your organization can benefit from a massive pool of latest IT resources, as well as redundant computer servers, without the worry of a large upfront investment or in-house tech know-how.

Q.34 List the different cloud application available in the market ? Briefly explain the scenario/situations of "when to not use clouds".

[RGPV : Dec-14, Marks 7]

Ans. : Different cloud application available in the market :

1. Healthcare
2. Data center for storage
3. Social networking
4. Media application

5.Image processing

Scenario/situations of "when to not use clouds":

- In some cases cloud providers are more expensive than on-premise systems.
- Using the Internet can cause network latency with some cloud applications
- Security is largely immature, and currently requires specialized expertise.
- Much of the technology is proprietary, and thus can cause lock-in.
- Compliance issues could raise the risks of using cloud computing.
- Data privacy issues could arise, if your cloud provider seeks to monetize the data in their system

Q.35 What do you mean by Social Network Analysis ? How Cloud computing can help in this problem ?

[RGPV : Dec-15, Marks 7]

- Social network analysis (SNA) is an important and valuable tool for knowledge extraction from massive and unstructured data. Social network provides a powerful abstraction of the structure and dynamics of diverse kinds of inter-personal connection and interaction.
- Facebook is a social networking service and website that connects people with other people, and share data between people. A user can create a personal profile, add other users as friends, exchange data, create and join common interest communities.
- Twitter is a social networking and microblogging service. The users of Twitter can exchange text-based posts called tweets. A tweet is a maximum 140 characters long but can be augmented by pictures or audio recording. The main concept of Twitter was to build a social network formed by friends and followers. Friends are people who you follow, followers are those who follow you.

- The role of social networks in labor markets deserves attention for at least two reasons: first, because of the central role networks play in disseminating information about job openings they place a critical role in determining whether labor markets function efficiently; and second, because network structure ends up having implications for things like human capital investment as well as inequality.
- Social Network Analysis (SNA) primarily focuses on applying analytic techniques to the relationships between individuals and groups, and investigating how those relationships can be used to infer additional information about the individuals and groups.
- SNA is used in a variety of domains. For example, business consultants use SNA to identify the effective relationships between workers that enable work to get done; these relationships often differ from connections seen in an organizational chart.

Q.36 What is Customer relationship management ?

Ans. : Customer Relationship Management (CRM) is a strategy for managing all your company's relationships and interactions with your customers and potential customers. It helps you improve your profitability.

- focus on individuals who had close ties to Hussein.
- Brief the application of cloud computing in the field of social networking and satellite image processing.

[RGPV : Dec-17, Marks 7]

W.E.F. (GSE) ADINGS

- Artist communities : These network communities are specifically composed to enable artists, musicians, or celebrities to personalize and intensify their contact with their existing fans as well as enabling contact among community members.
- Sport communities : These are network communities for special interests and activities of athletes and sport fans. People can find friends, celebrate their passion, and exchange ideas.
- Health communities : These are dedicated to the needs of actors concerned about health issues.
- Congresses and event communities : These are customized to support all preparations necessary for congresses and events, as well as all processes thereafter.
- Alumni communities : After completing their studies, alumni can find fellow students, stay in touch, and foster friendships

Also refer Q* 30

Q.37 What is Customer relationship management ?

- Ans. : Customer Relationship Management (CRM) is a strategy for managing all your company's relationships and interactions with your customers and potential customers. It helps you improve your profitability.

END... ↗

2

Cloud Computing Architecture

Cloud Computing

2-2

Cloud Computing Architecture

- Q.3 List and explain cloud deployment models.**
Ans.: Cloud deployment models are refers to the location and management of the clouds' infrastructure. Deployment models are defined by the ownership and control of architectural design and the degree of available customization. Cloud deployment models are private, public and community clouds.

- d. Cloud broker :** Cloud broker is an entity that manages the use, performance and delivery of cloud services, and negotiates relationships between cloud providers and cloud consumers.
- e. Cloud carrier :** It is a class of cloud that integrates wide area networks (WAN) and other attributes of communications service providers' carrier grade networks to enable the deployment of highly demanding applications in the cloud.

- Q.4 What is API ?**
Ans.: An application program interface (API) is a set of routines, protocols, and tools for building software applications. Basically, an API specifies how software components should interact.

- Cloud analytics can refer to any data analytics or business intelligence process that is carried out in collaboration with a cloud service provider.
- Virtual Desktop Infrastructure is a term that refers to using a virtualized desktop that is hosted on a virtual machine that lives on a server.
- Amazon Web Services (AWS) is a cloud computing platform from Amazon that provides customers with a wide array of cloud services.

- Important Points to Remember**
- Cloud deployment models are refers to the location and management of the clouds' infrastructure.
 - 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.
 - Cloud application interoperability addresses the application components, whether they are deployed as IaaS, PaaS, or SaaS.
 - Vendor lock-in is a situation in which a customer using a product or service cannot easily transition to a competitor's product or service.
 - Public cloud is built over the Internet and can be accessed by any user who has paid for the service.
 - A private cloud is built within the domain of an intranet owned by a single organization.
 - Scalability refers to the idea of a system in which every application or piece of infrastructure can be expanded to handle increased load.
 - Cloud stability : The ability to scale on demand constitutes one of the most attractive features of cloud computing.
 - Cloud computing ecosystem are business process, application services, platform services and Infrastructure services.
 - BPM is not a one-time task, but rather an ongoing activity that involves persistent process re-engineering.

- Q.1 What is cloud reference model ? What are the applications of this models ?**
[RGPV : Dec-15, Marks 2]

Ans.: 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.

- OR What are the fundamental components introduced in cloud reference model ? Explain.**
[RGPV : Dec-14, Marks 7]

Ans.: Fundamental components introduced in cloud reference model are as follows :

- a. **Cloud provider :** A person, organization, or entity responsible for making a service available to interested parties. When assuming the role of cloud provider, an organization is responsible for making cloud services available to cloud consumers, as per agreed upon service level agreement (SLA) guarantees. Cloud provider have their own IT resources.
- b. **Cloud consumer :** A person or organization that maintains a business relationship with, and uses service from, Cloud Providers. The cloud consumer uses a cloud service consumer to access a cloud service.
- c. **Cloud auditor :** Cloud auditor is a party that can perform an independent examination of cloud service controls with the intent to express an opinion thereon. A cloud auditor can evaluate the services provided by a cloud provider in terms of security controls, privacy impact, performance, etc.

- Q.2 What is pod and availability zone ?**
Ans. : Workloads support a certain number of users. When the workload reaches the limit of largest virtual machine instance possible, a copy or clone of the instance is required. A group of users within a particular instance is called a pod.

- Sizing limitation of pod need to be considered when building large cloud-based application. Pcs are aggregated into pools within IaaS region or site called an availability zone.

- Information exchange and service discovery requires standard protocols to realize interoperable platforms.

- Cloud management interoperability targets the management aspects between various cloud services deployed on SaaS, PaaS, or IaaS levels.

- Each provider realizes different cloud features and interfaces to manage them, so the clients would prefer to have a unique approach and generic off-the-shelf system management offered via standard interfaces.

- The Open Virtualization Format (OVF) is an attempt to provide a common format for storing the information and metadata describing a virtual machine image.

- Even though the OVF provides a full specification for packaging and distributing virtual machine images in completely platform-independent fashion, it is supported by few vendors that use it to import static virtual machine images.

Q.7 What is "VENDORLOCK-IN" concept ? Explain it using cloud interoperability and standards. Also discuss Open Virtualization Format (OVF) ?

[RGPV : June-15, Marks 7]

- Ans. :**
- Vendor lock-in is a situation in which a customer using a product or service cannot easily transition to a competitor's product or service.
 - Vendor lock-in is usually the result of proprietary technologies that are incompatible with those of competitors. However, it can also be caused by inefficient processes or contract constraints, among other things.
 - Vendor lock-in can prevent a customer from switching to another competitor's solution, or when this is possible, it happens at considerable conversion cost and requires significant amounts of time.
 - This can occur either because the customer wants to find a more suitable solution for customer needs or because the vendor is no longer able to provide the required service.
 - The presence of standards that are actually implemented and adopted in the cloud computing

community could give room for interoperability and then lessen the risks resulting from vendor lock-in.

Open Virtualization Format (OVF) :

- OVF is an open standard, specified by the Distributed Management Task Force (DMTF), for packaging and distributing a virtual appliance consisting of one or more virtual machines (VMs).
- An OVF Package is composed of metadata and file elements that describe virtual machines, plus additional information that is important to the deployment and operation of the applications in the OVF package. Its file extension is .ovf.
- An OVF Package always includes a descriptor file (*.ovf) and may also include a number of other files.

File type	Description
Descriptor	The descriptor specifies the virtual hardware requirements of the service and can also include other information such as descriptions of virtual disks, the service itself, and guest operating systems, a license agreement (EULA), instructions to start and stop VMs in the appliance, and instructions to install the service. The descriptor file extension is .ovf.

Manifest	The manifest is an SHA-1 digest of every file in the package, allowing the package contents to be verified by detecting any corruption. The manifest file extension is .mf.
Signature	The signature is the digest of the manifest signed with the public key from the X.509 certificate included in the package, and allows the package author to be verified. The signature file extension is .cert.

Virtual disks	OVF does not specify a disk image format. An OVF package includes files comprising virtual disks in the format defined by the virtualization product that exported the virtual disks. XenServer produces OVF packages with disk images in Dynamic VHD format; VMware products and VirtualBox produce OVF packages with virtual disks in Stream-Optimized VMDDK format.
---------------	--

Benefits of SaaS

- Streamlined administration
- Automated update and patch management services
- Data compatibility
- Facilitated enterprise-wide collaboration
- Global accessibility
- Familiarity with the WWW
- Smaller staff IT
- Customization
- Better marketing
- Web reliability
- Security secure sockets layer
- More bandwidth.

2. Platform as a Service (PaaS)

- Platform as a service is another application delivery model and also known as cloud-ware. Supplies all the resources required to build applications and services completely from the internet, without having to download or install software.

- Services include : Application design, development, testing, deployment, and hosting, team collaboration, web service integration, database integration, security, scalability, storage, state management, and versioning.
- PaaS is closely related to SaaS but delivers a platform from which to work rather than an application to work with.

- Q.13 What is community cloud ?**
- Ans. : A private cloud is built within the domain of an intranet owned by a single organization. Therefore, it is client owned and managed, and its access is limited to the owning clients and their partners.

- Q.14 What do you understand by SaaS ?**
- [RGPV : June-16, Marks 2]**

Ans. : Software-as-a-Service (SaaS) is a software delivery model that provides access to applications through the Internet as a Web-based service. It provides a means to free users from complex hardware and software management by offloading such tasks to third parties, which build applications accessible to multiple users through a Web browser.

2.2 : Types of Clouds**Q.11 What is public cloud ?**

Ans. : Public cloud is built over the Internet and can be accessed by any user who has paid for the service. Public clouds are owned by service providers and are accessible through a subscription.

- Q.12 What is private clouds ?**
- Ans. : A private cloud is built within the domain of an intranet owned by a single organization.

Ans. : The cloud infrastructure is shared by several organizations and supports a specific community that has shared concerns (e.g. mission, security requirements, policy, or compliance considerations). It may be managed by the organizations or a third party and may exist on-premises or off-premises.

be managed by the organization or a third party and may exist on-premises or off-premises.

Private cloud benefits :

1. Fewer security concerns as existing data center security stays in place.
2. IT organization retains control over data center.

Private cloud risks :

1. High investment hurdle in private cloud implementation, along with purchases of new hardware and software.
2. New operational processes are required; old processes not all suitable for private cloud.
- 3) **Community Cloud :** The cloud infrastructure is shared by several organizations and supports a specific community that has shared concerns (e.g. mission, security requirements, policy, or compliance considerations). It may be managed by the organizations or a third party and may exist on-premises or off-premises.
- 4) **Hybrid Cloud :** The cloud infrastructure is a composition of two or more clouds (private, community or public) that remain unique entities but are bound together by standardized or proprietary technology that enables data and application portability (e.g., cloud bursting for load-balancing between clouds).

Hybrid cloud benefits :

1. Operational flexibility : Run mission critical on private cloud, dev/test on public cloud.
2. Scalability : Run peak and bursty workloads on the public cloud.

Hybrid cloud risks :

1. Hybrid clouds are still being developed; not many in real use.
2. Control of security between private and public clouds, some of same concerns as in public cloud.

2.3 : Cloud Interoperability and Standards, Scalability and Fault Tolerance

Q.17 What do you mean by term scalability in terms of cloud ?

[RGPV : June-16, Marks 2]

- Ans. : The ability to scale on demand constitutes one of the most attractive features of cloud computing. Scalability refers to the idea of a system in which every application or piece of infrastructure can be expanded to handle increased load.

Q.18 How cloud computing provides scalability and fault tolerance ?

[RGPV : Dec-13, Marks 2]

Ans. : Fault tolerance is the process of finding faults and failures in a system. If a fault occurs or there is a hardware failure or software failure then also the system should work properly.

Q.19 Explain the key steps in cloud implementing planning process with example.

[RGPV : Dec-13, Marks 7]

Ans. : 1. Set up a cloud computing committee with representatives from all departments as well as upper management. The committee's responsibilities include the following :

- Developing a clear, detailed cloud computing acquisition and use plan.
 - Providing cloud computing awareness for key decision makers and users.
 - Conducting a basic cost/benefit analysis and calculating an ROI for the cloud computing acquisition.
 - Overseeing enforcement of the cloud computing policy.
2. Define the organization's needs. A clear definition of needs will assist the organization to decide on the following :
 - Does the organization need grid computing ?
 - Does the organization need SaaS ?
 - Does the organization need utility (on-demand) computing ?
3. Force software updates and work with the cloud computing provider(s) in order to spot unusual activities.
 4. Examine the providers of cloud platforms and match their offerings against your needs, not all



1 CPU / 1 GB RAM
2 CPU / 2 GB RAM

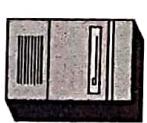


Fig. Q.21.2 : Vertical scaling

Q.22 Explain difference between horizontal and vertical scaling.

Ans. :

Horizontal scaling	Vertical scaling
In horizontal scaling, we build to the minimum requirements and then use monitoring and automation to scale it out.	Vertical scaling is where we estimate what we think the maximum requirements will be and add additional capacity beyond this to cover for any potential miscalculations and future expansion.
Cost migration is low.	Cost migration is low.
Upgrading downtime low.	Upgrading downtime is high.
Need load balance and gateway.	No coordination overhead.
Not limited by hardware capacity.	Limited by hardware capacity.
In horizontal scaling, resource of cluster is available.	All resources are in single host.

2.4 : Cloud Solutions : Cloud Ecosystem

Q.23 What is cloud ecosystem ?

[RGPV : Dec-16, Marks 2]

- Q.24 Write a brief notes on cloud ecosystem along with the examples.**
- [RGPV : Dec-14, June-15, Marks 7]
- Ans. : Fig. Q.24.1 shows cloud computing ecosystem.

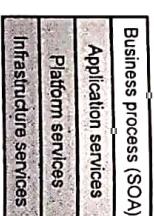


Fig. Q.24.1

- The center 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.
- Radiating out from the center of the cloud are software companies that use the provider's anchor platform, as well as consultants and companies that have formed strategic alliances with the anchor provider.
- There is no vendor lock-in because these companies overlap, making the ecosystem more complex.

- For example, AWS is the center 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, and Salesforce customers can gain access, through devices called connectors, to pieces of AWS, such as its Simple Storage Service (S3).
- A robust ecosystem provides a cloud provider's customers with an easy way to find and purchase business applications and respond to changing business needs.

<p>Cloud Computing</p> <p>Q.27 What is Virtual Desktop Infrastructure ? [RGV : Dec-16, Marks 2]</p> <p>Ans. : Virtual Desktop Infrastructure is a term that refers to using a virtualized desktop that is hosted on a Virtual machine that lives on a server.</p> <p>Q.28 What are the three key components of virtual desktop infrastructure ?</p> <p>Ans. : Three key components of virtual desktop infrastructure are host, connection broker and end points.</p> <p>Q.29 What is cloud analytics ?</p> <p>Ans. : Cloud analytics is a type of cloud service model where data analysis and related services are performed on a public or private cloud. Cloud analytics can refer to any data analytics or business intelligence process that is carried out in collaboration with a cloud service provider.</p> <p>Q.30 What do you understand by cloud analytics ?</p> <p>Also explain how it works. [RGV : Dec-13, Marks 7]</p> <p>OR Explain about cloud analytics using suitable example. [RGV : Dec-15, Marks 7]</p> <p>Ans. : • Cloud analytics is primarily a cloud-enabled solution that allows an organization or individual to perform business analysis or intelligence procedures.</p> <ul style="list-style-type: none"> • Basic categories of cloud analytics can be identified: <ol style="list-style-type: none"> 1. Storage as a service : This cloud service tends to be quite basic and corresponds directly to the storage layer. IBM Softlayer, Object Storage software and the Amazon Simple Storage Service (S3) are examples of this kind of service. 2. Data as a service : This service provides more sophisticated access to the data than the storage 		<p>layer and management systems includes relational database (RDBMSs), NoSQL databases, and other data access protocols and running tools are able to discover critical activities and bottlenecks, creating greater business value.</p> <p>2.6 : Cloud Offerings: Cloud Analytics, Testing Under Control, Virtual Desktop Infrastructure</p>
<p>Cloud Computing Architecture</p> <p>Q.31 Explain in brief cloud desktop infrastructure in brief with its services. [RGV : Dec-13, Marks 7]</p> <p>OR Explain in details about Virtual Desktop Infrastructure. [RGV : Dec-15, Marks 7]</p> <p>OR What do you mean by Virtual Desktop Infrastructure ? Write down its advantages and disadvantages. [RGV : Dec-17, Marks 7]</p> <p>Ans. : Virtual Desktop Infrastructure (VDI) :</p> <ul style="list-style-type: none"> • Virtual Desktop Infrastructure is a term that refers to using a virtualized desktop that is hosted on a Virtual machine that lives on a server. • Virtual Desktop Infrastructure is built on VMware virtual infrastructure hardware. Desktop virtual machines are hosted on server hardware running VMware Virtual Infrastructure software. These desktop environments are then accessed remotely 		<p>from a thick or thin client via a remote desktop protocol such as RDP.</p> <ul style="list-style-type: none"> • In other words, there is no physical desktop computer in front of you, but rather a terminal or remote desktop connection to a virtual machine that is hosted, secured and fully customizable per user basis. <ul style="list-style-type: none"> • VDI allows businesses and IT departments to streamline the cost of equipment by utilizing a single server to host multiple desktops for end-users while delivering a secure platform that is accessible to users from virtually anywhere they are. • Many firms use this Virtual Desktop Infrastructure for applications and business functions that additionally cut the cost of ownership and use down tremendously. <p>• Here are some reasons why VDI is beneficial :</p> <ol style="list-style-type: none"> 1. Save money on licensing and individual workstations/PCs by using thin clients. 2. Fully secured virtual environment that is fully monitored and managed. 3. Centralized management and backups. 4. Secure remote access from anywhere in the world. 5. Cost reduction for multiple software licenses. <p>Disadvantages :</p> <ul style="list-style-type: none"> • If an individual requires different applications from the other users, they will require a completely different image, without changing the applications for other users. • A substantial initial outlay is required for the main server hardware, storage and network infrastructure. This might not be feasible for some smaller businesses. • Administrators, savvy to the limitations, problem solving and installation of VDIs will either have to be brought in or existing IT staff given the relevant training. • If a problem occurs, this will generally affect all users, rather than being able to isolate problems if operating systems run off individual PCs.
<p>Cloud Computing</p> <p>Q.32 What is Amazon Web Services ? [RGV : Dec-13, Marks 7]</p> <p>Ans. : • AWS consists of many cloud services that you can use in combinations tailored to your business or organizational needs.</p> <ul style="list-style-type: none"> • With Amazon Web Services you will find a complete cloud platform ready to use for virtually any workload. • The user requests to the server by the method such as E-mail either to register or to transfer the domain. • Your request which includes all information will be sent to Amazon API Gateway restful service. • API Gateway will transfer the collected user information to an AWS Lambda function. • AWS Lambda function will generate an e-mail and forward it to the 3rd party mail server using Amazon SES. • Components of Amazon Web Service Architecture are Amazon API Gateway, AWS Lambda, Amazon Simple Email Service. • API Gateway is a front-door to access data, business logic and functionality. API Gateway will provide a restful API endpoint for our AWS Lambda function. 		<p>Cloud Computing Architecture</p> <p>Q.33 What is AWS ecosystem ?</p> <p>Ans. : • AWS ecosystem is made up of three subsystems :</p> <ol style="list-style-type: none"> 1. AWS computing services provided by Amazon. 2. Computing services provided by third parties that operate on AWS. 3. Complete applications offered by third parties that run on AWS. <p>Q.34 Explain the services provided by Amazon infrastructure cloud from user perspective. [RGV : Dec-13, Marks 7]</p> <p>Ans. : • AWS consists of many cloud services that you can use in combinations tailored to your business or organizational needs.</p> <ul style="list-style-type: none"> • With Amazon Web Services you will find a complete cloud platform ready to use for virtually any workload. • The user requests to the server by the method such as E-mail either to register or to transfer the domain. • Your request which includes all information will be sent to Amazon API Gateway restful service. • API Gateway will transfer the collected user information to an AWS Lambda function. • AWS Lambda function will generate an e-mail and forward it to the 3rd party mail server using Amazon SES. • Components of Amazon Web Service Architecture are Amazon API Gateway, AWS Lambda, Amazon Simple Email Service. • API Gateway is a front-door to access data, business logic and functionality. API Gateway will provide a restful API endpoint for our AWS Lambda function.

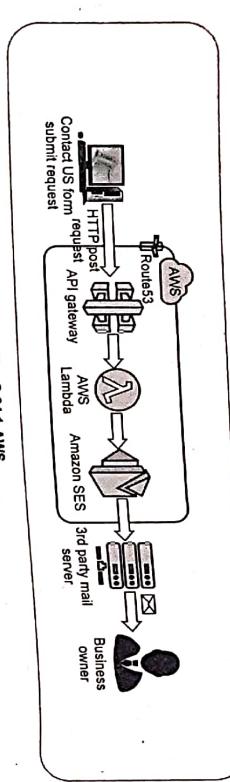


Fig. Q.34.1 AWS

- API works at small as well as large-scale and helps developers to manage, spectate, create and provide security to the APIs.

• AWS Lambda is a compute service that runs your back-end code and responds to events such as object uploads to Amazon S3 bucket, Dynamo DB or in-app activity. The Lambda function will get all the information from a user through API Gateway.

• Amazon Simple Email Service helps us to send E-mail with minimal setup and maximum deliverability. It is integrated with AWS management console so that you can monitor your sending activity. Amazon Simple Email Service helps us by monitoring insecurity.

Q.35 How testing under cloud can be performed ? [RGV : Dec-16, Marks 3]

OR How testing under cloud can be performed ? Explain it by taking service based model of cloud computing under consideration.

[RGV : Dec-14, June-15, Marks 7]

Ans. : Cloud testing, also called cloud-based testing is the assessment of a Web application's performance, reliability, scalability and security in a third-party's cloud computing environment.

- Compared to a traditional on-premises environment, cloud testing offers users pay-per-use pricing, flexibility and reduced time-to-market.
- The test processes and technologies used to perform functional testing against cloud-based applications are not significantly different than traditional in-house applications, but awareness of the non-functional risks around the cloud is critical to success.

- For example, if testing involves production data then appropriate security and data integrity processes and procedures need to be in place and validated before functional testing can begin.
- In the cloud environment, any application can be subjected to the following types of testing :
 1. Functional testing to ensure that software meets functional requirements.
 2. Non-functional testing to ensure the quality of service.
 3. Ability testing to show whether users will receive application services from the cloud environment on-demand.

- Functional testing : Functional software testing checks all the features and functions of software and its interaction with hardware. For conducting functional testing, testers can use such tools as Rapise, Sauce Labs, and TimeShiftX.
- Non-functional testing : Non-functional testing is also known as performance testing, as it allows you to check the non-functional aspects of software like its performance, usability, and reliability. For conducting this type of testing, you can use cloud-based tools such as CloudTest, AppPerfect, CloudFestoo, and AppLoader.
- Ability testing : Ability testing is necessary to verify whether users really receive application services on demand.
- Cloud testing focuses on the core components like Application : It covers testing of functions, end-to-end business workflows, data security, browser compatibility, etc.

3

Cloud Management and Virtualization Technology

Important Points to Remember

- Cloud provisioning is the allocation of a cloud provider's resources and services to a customer.
 - MapReduce is a programming model that simultaneously processes and analyzes huge datasets logically into separate clusters.
 - Disaster Recovery Plan is a plan designed to recover all the vital business processes during a disaster within a limited amount of time. This plan has all the procedures required to handle the emergency situations.
 - RPO is the amount of time that passes that's "acceptable" to you during an emergency situation.
 - RTO is the amount of time it takes you to restore regular business processes after a natural disaster or emergency situation.
 - Virtualization means running multiple machines on a single hardware.
 - Storage virtualization is a system administration practice that allows decoupling the physical organization of the hardware from its logical representation.
 - Hyper-V is an infrastructure virtualization solution.
 - File level storage is seen and deployed in Network Attached Storage (NAS) systems.
 - Block level storage is seen and deployed in Storage Area Network (SAN) storage.
 - A hypervisor or virtual machine monitor (VMM) is computer software, firmware or hardware that creates and runs virtual machines.
 - A Virtual Storage Area Network (VSAN) is a logical partition in a physical Storage Area Network (SAN).
 - A cloud storage provider is an organization that offers organizations and individuals the ability to place and retain data in an off-site storage system. Customers can lease cloud storage capacity per month or on demand.
- Q.1 What are the benefits of cloud based resilience ? [KGPV : June-16, Marks 3]**
- Ans. : Benefits of cloud based resilience are as follows :
1. Security
 2. Virtualization
 3. Managed cloud services
 4. Advanced technology
- A managed cloud services approach can increase business resilience in a security-rich, cost-effective, scalable way. The Cloud model improves the reliability of recovery operations with improved service levels at reduced costs
- Security : With top-tier cloud service providers, your organization's applications and data will often be more secure than on your own organization's premises.
- Virtualization : These benefits can be especially important to the many organizations already embarking on server virtualization plans. However, virtualizing only part of the server environment can make recovery more difficult for those portions of the environment that remain in their traditional physical states.
- Q.2 What is cloud provisioning ?**
- Ans. : Cloud provisioning is the allocation of a cloud provider's resources and services to a customer. It 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 catalog

of cloud services that customers can provision includes infrastructure as a service, software as a service and platform as a service, in public or private cloud environments.

Provisioning is the process by which a resource is prepared for use, reserved, accessed, used and then released when the transaction is completed.

Q.3 Discuss the benefits, goals and characteristics of provisioning.

[RGPV : Dec-13, Marks 7, Dec-15, Marks 2]

Ans. : • It 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 catalog of cloud services that customers can provision includes infrastructure as a service, software as a service and platform as a service, in public or private cloud environments.

Benefits of Provisioning :

1. Continuous improvement of activity which relay on process measurement.
2. There is isolation of install, configure, build and customize task.
3. Ability to measure progress of all work
4. Assembly line approach to provisioning

Characteristics of provisioning :

1. Missing and incorrect information is resolved.
2. Product being deliver to customer before provisioning.

- The characteristics of a provision are that it is a liability where there is uncertainty as to either the timing of settlement or the amount to be settled.
- When measuring a provision, the amount to be recognized should be the best estimate of the consideration required to settle the present obligation at the end of the reporting period.
- The fact that it is difficult to measure a provision and that estimates have to be used does not mean that the provision is not reliably measurable

Q.4 Explain Cloud Provisioning along with its types.

[RGPV : June-16, Marks 7]

Ans. : Types of cloud provisioning are as follows :

1. Dynamic provisioning : Cloud resources are deployed flexibly to match a customer's fluctuating demands. The 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.
2. Self provisioning : With user self-provisioning also called cloud self-service, the customer buys resources from the cloud provider through a web interface or portal. This usually involves creating a user account and paying for resources with a credit card.
3. Advance provisioning : Customer undertake contract with the provider for the required services.

3.2 : Concepts of Map Reduce

Q.5 Write properties of Map Reduce.

[RGPV : June-16, Marks 2]

Ans. : Properties of Map Reduce :

- Very large scale data.
- MapReduce allows for distributed processing of the map and reduction operations.
- MapReduce provides analytical capabilities for analyzing huge volumes of complex data.
- Number of maps tasks and reduce tasks are configurable.
- Map and reduce operations are typically performed by the same physical processor.

Q.6 What is MapReduce ?

Ans. : MapReduce is a programming model that simultaneously processes and analyzes huge datasets logically into separate clusters.

While Map sorts the data, Reduce segregates it into logical clusters, thus removing 'bad' data and retaining the necessary information

- The fact that it is difficult to measure a provision and that estimates have to be used does not mean that the provision is not reliably measurable

Q.7 Explain the concept of Map Reduce in cloud management.

[RGPV : Dec-16, Marks 7]

OR Example with suitable example map reduce model.

Q.8 Explain Map-reduce programming methodology. How it helps in implementing complex problems using cloud computing?

[RGV : Dec.-15, Marks 7]

Ans.: • The computation takes a set of input key-value pairs, and produces a set of output key-value pairs. The user of the MapReduce library expresses the computation as two functions: Map and Reduce.

- Map, written by the user, takes an input pair and produces a set of intermediate key/value pairs. MapReduce library groups together all intermediate values associated with the same intermediate key "I" and passes them to the Reduce function.
- The Reduce function, also written by the user, accepts an intermediate key I and a set of values for that key. It merges together these values to form a possibly smaller set of values.

• Fig. Q.8.1 shows the overall flow of a MapReduce operation.

• When the user program calls the MapReduce function, the following sequence of actions occurs :

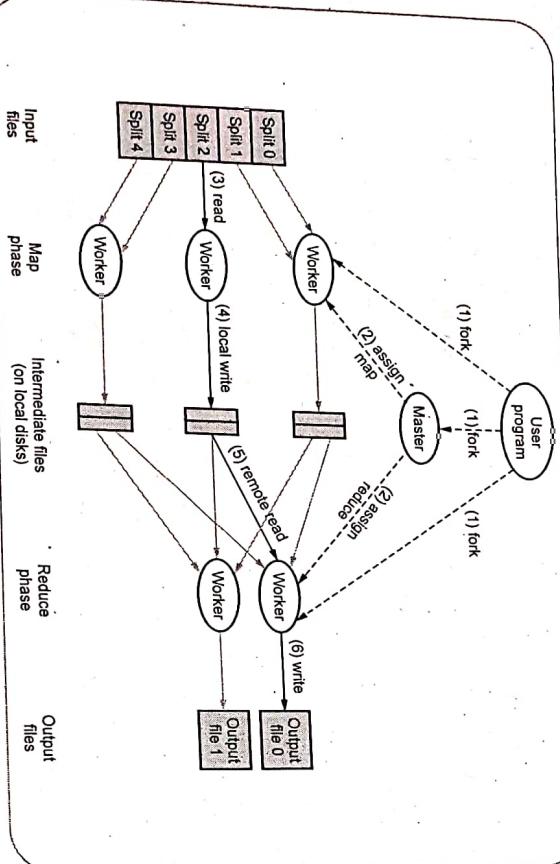


Fig. Q.8.1 MapReduce execution

5. When a reduce worker is notified by the master about these locations, it uses remote procedure calls to read the buffered data from the local disks of the map workers. When a reduce worker has read all intermediate data, it sorts it by the intermediate keys so that all occurrences of the same key are grouped together.

- 6. The reduce worker iterates over the sorted contents of the corresponding input split. Reduce function is appended to a final output file for this reduce partition.
- 7. When all map tasks and reduce tasks have been completed, the master wakes up the user program. At this point the MapReduce call in the user program returns back to the user code.

Q.9 Explain data flow in the MapReduce programming model.

Ans.: • MapReduce : A programming model to facilitate the development and execution of distributed tasks.

• The programmer defines the program logic as two functions :

- Map transforms the input into key-value pairs to process
- Reduce aggregates the list of values for each key

• The MapReduce environment takes in charge distribution aspects. A complex program can be decomposed as a succession of Map and Reduce tasks. Higher-level languages (Pig, Hive, etc.) help with writing distributed applications.

• MapReduce is a parallel programming model especially dedicated for complex and distributed computations, which has been derived from the functional paradigm.

- In general, MapReduce processing is composed of two consecutive stages, which for most problems are repeated iteratively: the map and the reduce phase. Fig. Q.9.1 shows data flow in the MapReduce programming model.

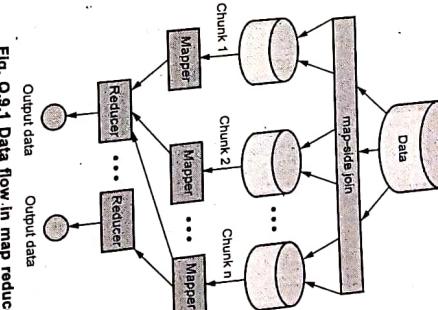


Fig. Q.9.1 Data flow in map reduce

• Each chunk may be processed by only one mapper. Once the data is processed by mappers, they can emit View the MathML source <key,value> pairs to the reduce phase.

- Before the reduce phase the pairs are sorted and collected according to the View the MathML sourcekey values, therefore each reducer gets the list of values related to a given View the MathML sourcekey. The consolidated output of the reduce phase is saved into the distributed file system.

Q.10 How MapReduce processes big data ? Explain with diagram.

Ans.: • With HDFS, we are able to distribute the data so that data is stored on hundreds of nodes instead of a single large machine.

- Mapreduce provides the framework for highly parallel processing of data across clusters of commodity hardware. Fig. Q.10.1 shows MapReduce data processing.

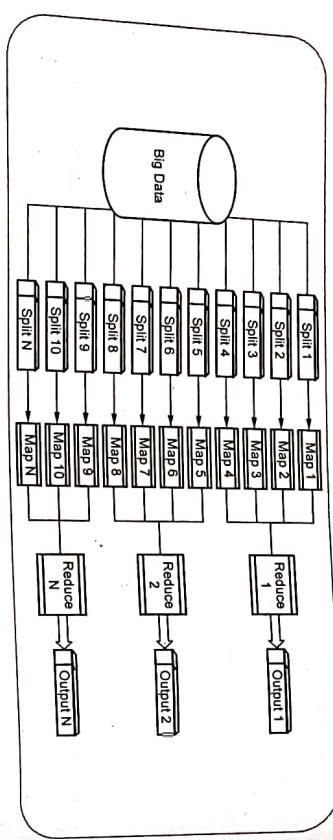


Fig. Q.10.1

3.3 : Cloud Governance, High Availability and Disaster Recovery

- It removes the complicated programming part from the programmers and moves into the framework. Programmers can write simple programs to make use of the parallel processing.
- The framework splits the data into smaller chunks that are processed in parallel on cluster of machines by programs called mappers.
- Data locality is achieved by mapreduce by working closely with HDFS. When you specify the file system as HDFS for mapreduce, it automatically schedules the mappers on the same node as where the block of data exists.
- Mapreduce can get the blocks from HDFS and process them. The final output from Mapreduce also can be stored in HDFS file system. However, the intermediate files between mappers and reducers are not stored in HDFS and are stored on the local file system of the mappers.

Q.11 What Disaster recovery plan.

Ans. : Disaster Recovery Plan is a plan designed to recover all the vital business processes during a disaster with in a limited amount of time. This plan has all the procedures required to handle the emergency situations.

Q.12 What happens if you Ignore RTO and RPO ?

Ans. : • If you ignore RTO, the amount of time it takes for you to get up and running again after a disaster and you won't be able to manage downtime properly in the future.

- If you ignore RPO, the amount of downtime that's acceptable during a disaster and you won't meet your business continuity objectives. That could jeopardize your business reputation. In fact, you could lose customers if you experience too much downtime. Consumers won't be able to purchase products from your online store or contact you will queries.

Q.13 Write a brief note on high availability and disaster recovery in cloud computing. Also explain RTO and RPO.

Ans. : A highly available infrastructure component or IT system is described as "fault tolerant" or having the ability to "fail over." An example of high availability at the component level is adding redundant power supplies.

• Disaster recovery can, and often does, include high availability in the technology design. Often this configuration takes the form of implementing highly available clustered servers for an application within a production datacenter and having backup hardware in the recovery datacenter.

• With data from the production server backed up or replicated to the recovery datacenter, systems are both protected from component failures at the production datacenter and can be recovered during a disaster at the recovery data center.

• The ultimate combination of high availability and disaster recovery occurs when servers are configured as "active-active" or in a "continental cluster" across geographically diverse datacenters.

• In this case clustered servers for an IT application reside in two different datacenters connected by a load balancer and a very low latency data connection.

• Data between the two servers is synchronously replicated and both systems are "active" at the same time. Should one datacenter be impacted by a disaster, the server in the second internet data center picks up the full load of the application and continues on uninterrupted.

• Add to this the people, processes and documentation necessary to manage and respond to a datacenter incident and the high availability configuration is now incorporated within the disaster recovery program.

• Recovery Point Objective (RPO) and Recovery Time Objective (RTO) are two of the most important parameters of a disaster recovery or data protection plan. These are objectives which can guide enterprises to choose an optimal data backup plan.

• Recovery Point Objective (RPO) describes the interval of time that might pass during a disruption before the quantity of data lost during that period exceeds the Business Continuity Plan's maximum allowable threshold or "tolerance."

• RPO is the amount of time that passes that's "acceptable" to you during an emergency situation. You should define this amount of time in your business continuity plan. If you exceed this period of time, you could jeopardize your business.

• The Recovery Time Objective (RTO) is the duration of time and a service level within which a business process must be restored after a disaster in order to avoid unacceptable consequences associated with a break in continuity.

• RTO is the amount of time it takes you to restore regular business processes after a natural disaster or emergency situation.

Q.14 Explain disaster recovery plan.

Ans. : • Disaster recovery plan is a plan designed to recover all the vital business processes during a disaster with in a limited amount of time. This plan has all the procedures required to handle the emergency situations.

• A disaster recovery process should have provable recovery capability and hence it provides the most efficient method to be adopted immediately after a disaster occurs.

• Mostly the DRP has technology oriented methodologies and concentrates on getting the systems up as soon as possible, within a reasonable amount of time Recovery Time Objective (RTO) and Recovery Point Objective (RPO).

• RTO and RPO are the recovery time objective and recovery point objective, which are the targets of DRP.

• The most successful disaster recovery strategy is the one that will never be implemented; therefore, risk avoidance is a critical element in the disaster recovery process.

• To minimize the impact of a disaster, companies invest time and resources to plan and prepare, to train employees and to document and update processes.

• The amount of investment for DR planning for a particular system can vary dramatically depending on the cost of a potential outage.

- Q.15 What is virtualization?**
[GPV : Dec-16, Marks 2]

Ans.: Virtualization is an abstraction layer that decouples the physical hardware from the operating system to deliver greater IT resource utilization and flexibility. It allows multiple virtual machines, with heterogeneous operating systems to run in isolation, side-by-side on the same physical machine.

- Virtualization means running multiple machines on a single hardware. The "Real" hardware invisible to operating system. OS only sees an abstracted out picture. Only Virtual Machine Monitor (VMM) talks to hardware.

Q.16 What is networking virtualization?

[GPV : June-16, Marks 2]

Ans.: Network virtualization refers to the technology that enables partitioning or aggregating a collection of network resources and presenting them to various users in a way that each user experiences an isolated and unique view of the physical network.

Q.17 What are the benefits of virtualization in the context of cloud computing?

Ans.: 1. It is possible to achieve a more efficient use of resources

2. Portability and self-containment also contribute to reducing the costs of maintenance.

3. A virtual execution environment can be configured as a sandbox, thus preventing any harmful operation.

Q.18 List disadvantages of virtualization.

- Ans.: 1. Performance degradation
2. Virtualization can sometimes lead to an inefficient use of the host
3. Virtualization opens the door to a new and unexpected form of phishing

Q.19 What is operating system level virtualization?

Ans.: In essence, virtualization transforms physical hardware into virtual machines, while cloud

3.4 : Virtualization

[GPV : Dec-16, Marks 2]

Ans.: Operating-system-level virtualization is a server-virtualization method where the kernel of an operating system allows for multiple isolated user-space instances, instead of just one. Such software containers.

Q.20 What are hardware virtualization techniques?

Ans.: This technology allows simulating the hardware interface expected by an operating system. Hardware virtualization allows the coexistence of different software stacks on top of the same hardware. These stacks are contained inside virtual machine instances, which operate in complete isolation from each other.

Q.21 What is application server virtualization?

Ans.: Application server virtualization abstracts a collection of application servers that provide the same services as a single virtual application server by using load-balancing strategies and providing a high-availability infrastructure for the services hosted in the application server.

Q.22 Why operating system level virtualization is required?

Ans.: Operating system level virtualization provides feasible solution for hardware level virtualization issue. It inserts a virtualization layer inside an operating system to partition a machine's physical resources.

• It enables multiple isolated VMs within a single operating system kernel. This kind of VM is often called a virtual execution environment (VEx), Virtual Private System (VPS), or simply container.

• From the user's point of view, virtual execution environment look like real servers.

• This means a virtual execution environment has its own set of processes, file system, user accounts, network interfaces with IP addresses, routing tables, firewall rules etc.

Q.23 How are virtualization and cloud computing interrelated? Explain virtualization in cloud.

[GPV : Dec-17, Marks 7]

computing is a kind of service that is used to store the transformed data.

• Virtualization is also known as a component of cloud computing. If you plan to get a cloud computing service, you'll need a virtualization software program first to create virtual networks and servers that you can upload on the private cloud.

• Private cloud computing doesn't rely on a single technology, but virtualization is one of the best options. Virtualization makes a great contribution to the core concepts of cloud computing.

• Hardware virtualization is an enabling factor for solutions in the Infrastructure-as-a-Service (IaaS) market segment, while Programming language virtualization is a technology leveraged in Platform-as-a-Service (PaaS) offerings.

In both cases, the capability of offering a customizable and sandboxed environment

constituted an attractive business opportunity for companies featuring a large computing infrastructure that was able to sustain and process huge workloads.

• Moreover, virtualization also allows isolation and a finer control, thus simplifying the leasing of services and their accountability on the vendor side.

• Besides being an enabler for computation on demand, virtualization also gives the opportunity to design more efficient computing systems by means of consolidation, which is performed transparently to cloud computing service users. Since virtualization allows us to create isolated and controllable environments, it is possible to serve these environments with the same resource without them interfering with each other.

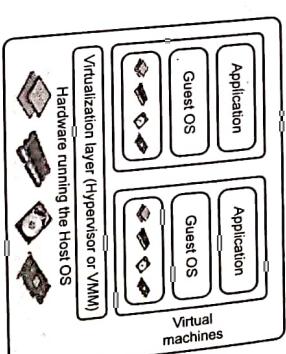
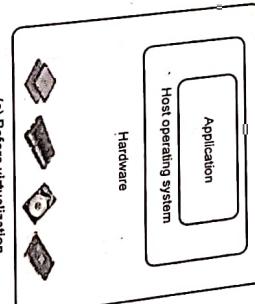


Fig. Q.23.1



• Fig. Q.23.1 shows before and after virtualization.

Q.24 How does an unauthorized access can be detected by the help of virtualization techniques?

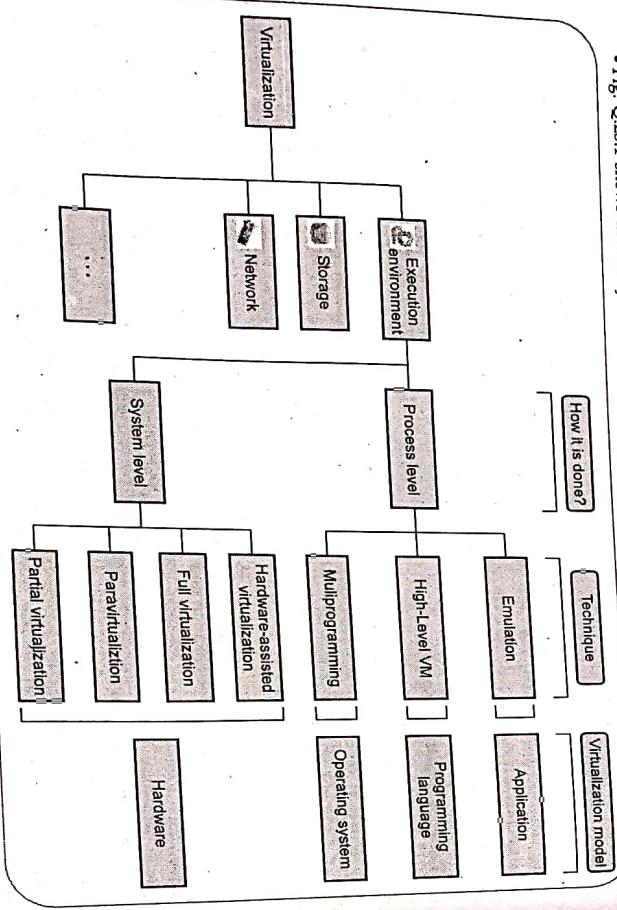
Ans.: • The solutions based on virtualization infrastructure aim to solve security vulnerabilities by creating secure gateways in the virtualization infrastructure.

- Security on virtual layer is achieved by securing how VMs and hypervisors talk to each other in a virtual network. In order to take full advantage of the virtualization infrastructure, Virtual Private Networks (VPNs) are commonly created to manage different levels of authority in VMs.
- Because of the virtual nature of the network, features such as monitoring, access controls, integrity, encryption, authentication and transportability of VMs can be implemented directly into the network.
- This will solve many of the vulnerabilities present in a virtualization as the security on the virtual layer will isolate different virtual management networks and bring ease to deployment and operation of VMs across different authorities or data centers.
- Security on physical layer is the design of the structure of the physical systems that brings about security in a virtualized environment.
- It allows the system to ensure that at least the physical layer will not be compromised easily through other means. The structure of the data center or the cloud also plays an important role.
- How the machines that are running the VMs interconnected physically can determine the possible security measures that can be used.
- Also, routine inspection for hardware failures and outdated systems is part of the security on the physical infrastructure that plays a large role in determining how secure the virtualized environment is.

Q.25 Explain the characteristics and types of virtualization in cloud computing.

Ans. : Types of virtualization in cloud computing :

• Fig. Q.25.1 shows taxonomy of virtualization techniques.



- Execution virtualization includes all techniques that aim to emulate an execution environment that is separate from the one hosting the virtualization directly on top of the hardware by the operating system, an application, or libraries dynamically or statically linked to an application image.
- Operating-system-level virtualization is a server-virtualization method where the kernel of an operating system allows for multiple isolated user-space instances, instead of just one. Such instances, which are sometimes called containers and software containers.
- Programming-language-level virtualization is mostly used to achieve ease of deployment of applications, managed execution, and portability across different platforms and operating systems. It consists of a virtual machine executing the byte code of a program, which is the result of the compilation process. Compilers implemented and used this technology to produce a binary format representing the machine code for an abstract architecture.

Platform virtualization

- The creation of a virtual machine using a combination of hardware and software is referred to as platform virtualization. Platform virtualization is performed on a given hardware platform by "host" software, which creates a simulated computer environment for its "guest" software.
- The "guest" software, which is often itself a complete operating system, runs just as if it were installed on a stand-alone hardware platform. Typically, many such virtual machines are simulated on a given physical machine.
- For the "guest" system to function, the simulation must be robust enough to support all the guest system's external interfaces, which may include hardware drivers.

Resource virtualization

- The basic concept of platform virtualization was later extended to the virtualization of specific system resources, such as storage volumes, name spaces and network resources.

- Q.26 What is need of storage virtualization?**

Ans. : Discuss the three ways in which it is structured ?

- [RGPV : Dec.-17, Marks 7]**
- Storage virtualization is a system administration practice that allows decoupling the physical organization of the hardware from its logical representation.

- Using this technique, users do not have to be worried about the specific location of their data, which can be identified using a logical path.

- If this virtualization implements in IT environment it will improve the management of the storage. As each and everything will properly store and manage there won't be any congestion and the task will perform quickly.
- There will be very less downtime as the storage availability is better. All these problems eliminate with the help of an automated management system.
- Storage virtualization will provide better storage utilization as storing most information in a particular place can cause loss of data, congestion and any other problems. So, properly dividing storage and storing data can be useful.
- Storage virtualization is structured in following ways :
 1. Hardware assisted virtualization
 2. Kernel level virtualization
 3. Hypervisor virtualization
 4. Para-virtualization
 5. Full virtualization

Fig. Q.25.1 Taxonomy of virtualization

- Full device emulation : This process emulates well-known and real-world devices. All the functions of a device or bus infrastructure such as device enumeration, identification, interrupts etc. are replicated in the software which itself is located in the VMX and acts as a virtual device. The I/O requests are trapped in the VMX accordingly.
- Para-virtualization : This method of I/O VZ is taken up since software emulation runs slower than the hardware it emulates. In para-VZ, the frontend driver runs in Domain-U; it manages the requests of the guest OS. The backend driver runs in Domain-0 and is responsible for managing the real I/O devices. This methodology (para) gives more performance but has a higher CPU overhead.

Q.27 What is need of virtualization? Discuss the architecture of hyper-V and discuss its use in cloud computing.

Ans. : • Hyper-V is an infrastructure virtualization solution.

- Hyper-V supports multiple and concurrent execution of guest operating systems by means of partitions.

• Fig. Q.27.1 shows block diagram architecture of hyper-V.

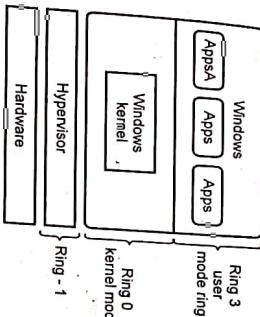


Fig. Q.27.1 Diagram of hyper-V

- The instructions access is divided by four privileged levels in the processor called rings. The most privileged level is Ring 0, with direct access to the hardware and where the Windows Kernel sits.
- Ring 3 is responsible for hosting the user level, where most common applications run and with the least privileged access.

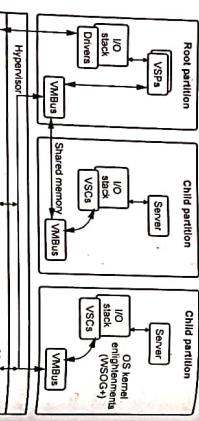


Fig. Q.27.1 Architecture of hyper-V

- Child partitions also do not have direct access to other hardware resources and are presented a virtual view of the resources, as virtual devices (VDev's).
- Requests to the virtual devices are redirected either via the VMBus or the hypervisor to the devices in the parent partition, which handles the requests.
- Virtualization Service Provider (VSP) resides in the root partition and provide synthetic device support to child partitions over the Virtual Machine Bus (VMBus).
- Virtualization Infrastructure Driver (VID) provides partition management services, virtual processor management services and memory management services for partitions.
- Virtualization Service Client (VSC) : A synthetic device instance that resides in a child partition.

- Hyper-V supports isolation in terms of a partition. A partition is a logical unit of isolation, supported by the hypervisor, in which operating systems execute. The Microsoft hypervisor must have at least one parent or root partition, running at least one virtual machine.
- The virtualization management stack runs in the parent partition and has direct access to hardware devices. The root partition then creates the child partitions which host the guest operating systems. A root partition creates child partitions using the Hyper-V Application Programming Interface (API).
- Partitions do not have access to the physical processor, nor do they handle the processor interrupts.

- VMBus : it is channel-based communication mechanism used for inter-partition communication and device enumeration on systems with multiple active virtualized partitions. The VMBus is installed with Hyper-V Integration Services.
- Virtual Machine Management Service (VMMs) is responsible for managing the state of all virtual machines in child partitions.

Q.28 Write in brief about block and file level storage virtualization.

Ans. : • File level storage is seen and deployed in Network Attached Storage (NAS) systems.

• File-level storage is the predominant storage technology used on hard drives, Network-Attached Storage (NAS) systems and similar storage systems.

• File-level storage stands in contrast to block-level storage in that individual files and folders can be accessed and managed by the storage system, whereas the smaller storage blocks that make up the files and folders cannot be directly controlled.

• Block level storage is seen and deployed in Storage Area Network (SAN) storage.

• Each block in a block-level storage system can be controlled as an individual hard drive and the blocks are managed by a server operating system.

• Block-level storage protocols like iSCSI, Fibre Channel and FCoE (Fibre Channel over Ethernet) are utilized to make the storage blocks visible and accessible by the server-based operating system.

This is in contrast with file-level storage, in which the storage drives are configured with a storage protocol like NFS or SMB/CIFS.

Q.29 What do you mean by server virtualization? Explain using example.

Ans. : • The ability to run multiple operating systems on a single physical system and share the underlying hardware resources. Virtual machines

- provide a layer of abstraction between the OS and the underlying hardware.
- Creating multiple logical server OS instances on one physical piece of hardware. All hardware drivers are virtualized and same virtual hardware regardless of physical hardware.
- Each virtual machine is completely independent of the others and doesn't realize it's virtualized.
- Requirements of server virtualization :

1. **Consolidation** : It is common practice to dedicate each server to a single application. If several applications only use a small amount of processing power, the network administrator can combine several machines into one server running multiple virtual environments.
2. **Redundancy** : Redundancy refers to running the same application on multiple servers. It's a safety measure, if a server fails for any reason, another server running the same application can take its place.

3. **Legacy hardware** : Server hardware will eventually become obsolete and switching from one system to another can be difficult. In order to continue offering the services provided by these outdated systems, sometimes called legacy systems a network administrator could create a virtual version of the hardware on modern servers.
4. **Migration** : Migration refers to moving a server environment from one place to another. With the right hardware and software, it's possible to move a virtual server from one physical machine in a network to another. Originally, this was possible only if both physical machines ran on the same hardware, operating system and processor. It's possible now to migrate virtual servers from one physical machine to another even if both machines have different processors, but only if the processors come from the same manufacturer.

- Virtualization allows multiple operating system instances to run concurrently on a single computer; it is a means of separating hardware from a single operating system. Each "guest" OS is managed by a Virtual Machine Monitor (VMM), also known as a hypervisor.

Cloud Computing 3 - 16 **Cloud Management and Virtualization Technology**

Ans. :-

- In essence, hypervisors create a virtualization layer that separates CPU / Processors, RAM and other physical resources from the virtual machines you create.
- In order to create virtual instances, you need a management console on another machine. Using the console, you connect to the hypervisor on the server to manage your virtual environment.
- A management console can be web-based or a separate software package you install on the desired machine you will use for remote management.
- Depending on what functionalities you need, the license cost for management consoles changes substantially. One of the actions you can perform include moving virtual machines between physical servers, manually or automatically, according to the resource needs of a VM at a given point.
- This migration happens without any impact to the end users. The same goes if a piece of hardware or a whole server fails. Properly configured management software moves virtual machines to a working server as soon as an issue arises. Once the problem is taken care of, the restoration procedure also happens automatically and seamlessly.
- Hardware level virtualization inserts a layer between real hardware and traditional OS. This layer is commonly called the Virtual Machine Monitor (VMM) and it manages the hardware resources of a computing system.
- Three requirements for a VMM.

 - VMM should provide an environment for programs which is essentially identical to the original machine.
 - The programs run in this environment should show, at worst, only minor decreases in speed.
 - VMM should be in complete control of the system resources. Any program run under a VMM should exhibit a function identical to that which it runs on the original machine directly.

Q.34 Enlist and explain some of the common pitfalls that come with virtualization.
[RGPV : Dec-13, June-15, Marks 7]

Cloud Computing 3 - 17 **Cloud Management and Virtualization Technology**

Ans. :-

 - Virtualization can sometime lead to an inefficient use of the host.
 - The Network : Virtualizing servers before making sure that the network infrastructure can handle it is a risky undertaking. After the process is complete the network will be under a lot more strain and making sure that it has what it takes to sustain the added traffic is critical.
 - Performance degradation : Performance is definitely one of the major concerns in using virtualization technology. Since virtualization interposes an abstraction layer between the guest and the host, the guest can experience increased latencies.
 - Security holes and new threats : Virtualization opens the door to a new and unexpected form of phishing. The capability of emulating a host in a completely transparent manner led the way to malicious programs that are designed to extract sensitive information from the guest.
 - Misplacing applications : A virtualized infrastructure is a more complex than a traditional one and with a number of applications deployed, losing track of applications is a distinct possibility. Within a physical server infrastructure keeping track of all the apps and the machines running them isn't a difficult task.
 - Mismatching Servers : This aspect is commonly overlooked especially by smaller companies that don't invest sufficient funds in their IT infrastructure and prefer to build it from several bits and pieces. This usually leads to simultaneous virtualization of servers that come with different chip technology.

Q.35 Explain benefits of virtualization.
[RGPV : June-16, Marks 7]

Ans. :-

 - Data center and energy-efficiency savings : As companies reduce the size of their hardware and server footprint, they lower their energy consumption
 - Operational expenditure savings : Once servers are virtualized, your IT staff can greatly reduce the ongoing administration and management of manual.
 - Reduced costs : One of the biggest advantages associated with using virtualization in your IT infrastructure is the ability to cut costs.

Q.37 What is full virtualization? Explain host based virtualization.
Ans. :- Full Virtualization doesn't need to modify the host OS; it relies upon binary translation to trap and to virtualize certain sensitive instructions.

 - Fig. Q.37.1 shows full virtualization.

Fig. Q.37.1 : Full virtualization

VMware Workstation applies full virtualization, which uses binary translation to automatically modify x86 software on-the-fly to replace critical instructions.

 - Normal instructions can run directly on the host OS. This is done to increase the performance overhead - normal instructions are carried out in the normal manner, but the difficult and precise executions are first discovered using a trap and executed in a virtual manner.
 - This is done to improve the security of the system and also to increase the performance.

Host based virtualization :

 - Virtualization implemented in a host computer rather than in a storage subsystem or storage appliance.
 - Virtualization can be implemented either in host computers, in storage subsystems or storage appliances, or in specific virtualization appliances in the storage interconnect fabric.
 - The guest OS are installed and run on top of the virtualization layer. Dedicated applications may run on the VMs. Certainly, some other applications can also run with the host OS directly.

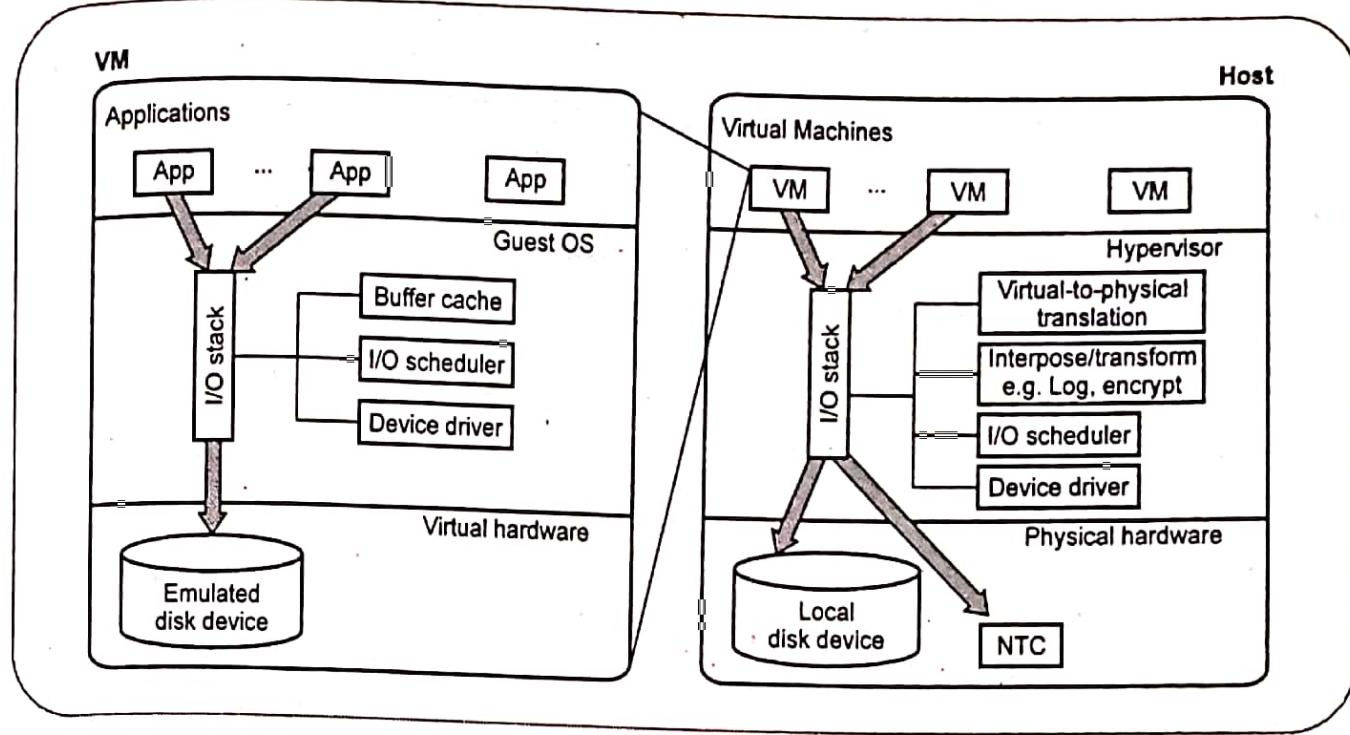


Fig. Q.40.1 : I/O virtualization

1. **Full Device Emulation :** This process emulates well-known and real-world devices. All the functions of a device or bus infrastructure such as device enumeration, identification, interrupts etc. are replicated in the software, which itself is located in the VMM and acts as a virtual device. The I/O requests are trapped in the VMM accordingly.
2. **Para-virtualization :** This method of I/O VZ is taken up since software emulation runs slower than the hardware it emulates. In para-VZ, the frontend driver runs in Domain-U; it manages the requests of the guest OS. The backend driver runs in Domain-0 and is responsible for managing the real I/O devices. This methodology (para) gives more performance but has a higher CPU overhead.
3. **Direct I/O virtualization :** This lets the VM access devices directly; achieves high performance with lower costs. Currently, it is used only for the mainframes.
- I/O virtualization features facilitate offloading of multi-core packet processing to network adapters as well as direct assignment of virtual machines to virtual functions, including disk I/O.
- Examples include Virtual Machine Device Queues (VMDQ), Single Root I/O Virtualization.

Q.41 What are the live migration steps ? Explain.

Ans. : • Steps 0 and 1 : Start migration automatically and checkout load balances and server consolidation.

- Step 2 : Transfer memory (transfer the memory data + recopy any data that is changed during the process). This goes on iteratively till changed memory is small enough to be handled directly.
- Step 3 : Suspend the VM and copy the last portion of the data.
- Steps 4 and 5 : Commit and activate the new host. Here, all the data is recovered, and the VM is started from exactly the place where it was suspended, but on the new host.

Q.42 Write short note on File System Migration.

Ans. : • To support VM migration from one cluster to another, a consistent and location-dependent view of the file system is available on all hosts.

- Each VM is provided with its own virtual disk to which the file system is mapped to. The contents of the VM can be transmitted across the cluster by inter-connections (mapping) between the hosts.
- But migration of an entire host is not advisable due to cost and security problems. We can also provide a global file system across all host machines where a VM can be located.

- This methodology removes the need of copying files from one machine to another, all files on all machines can be accessed through network.
- It should be noted here that the actual files are not mapped or copied. The VMM accesses only the local file system of a machine and the original/modifies files are stored at their respective systems only.
- This decoupling improves security and performance but increases the overhead of the VMM - every file has to be stored in virtual disks in its local files.

Q.43 What is pre copy and post copy of live migration ?

Ans. : In pre copy, which is mainly used in live migration, all memory pages are first transferred; it then copies the modified pages in the last round iteratively. Here, performance 'degradation' will occur because migration will be encountering dirty pages all around in the network before getting to the right destination. The iterations could also increase, causing another problem. To encounter these problems, check-pointing/recovery process is used at different positions to take care of the above problems and increase the performance.

In post-copy, all memory pages are transferred only once during the migration process. The threshold time allocated for migration is reduced. But the downtime is higher than that in pre-copy.

3.5 : Virtual LAN (VLAN) and Virtual SAN (VSAN)

Q.44 What is VSAN ? [RGPV : Dec.-16, Marks 2]

Ans. : A Virtual Storage Area Network (VSAN) is a logical partition in a physical Storage Area Network (SAN). VSANS allow traffic to be isolated within specific portions of a storage area network so that if a problem occurs in one logical partition, it can be handled with a minimum of disruption to the rest of the network.

Q.45 Explain Virtual LAN (VLAN) and Virtual SAN (VSAN) in cloud computing. [RGPV : Dec.-15, Marks 7]

- Ans. : Virtual LAN (VLAN) :**
 - Virtual local area network define as a local area network configured by software, not by physical wiring.
 - A VLAN is a switched network that is logically segmented on an organizational basis, by function, project teams or applications rather than on a physical or geographical basis.
 - VLAN can be thought of as a broadcast domain that exists within defined set of switches.
 - A VLAN consists of a number of end systems either hosts or network equipment connected by a single bridging domain.
 - The bridging domain is supported on various pieces of network equipment; for example, LAN switches that operate bridging protocols between them with a separate bridge group for each VLAN.
 - VLANs are created to provide the segmentation services traditionally provided by routers in LAN configurations.
 - VLANs address scalability, security and network management.
 - A LAN can be divided into several logical LANs called VLANs. Each VLAN is a work group in the organization. The group membership in VLANs is defined by software, not hardware. All members belonging to a VLAN can receive broadcast messages sent to that particular VLAN.
 - Stations are configured in following ways.
 - Manual 2. Semiautomatic 3. Automatic
- Manual Configuration :** In this configuration, the network administrator uses the VLAN software to manually assign the stations into different VLANs at setup. Later moving from one VLAN to another is also done manually. It is a logical configuration. The administrator types the port numbers, IP addresses or other characteristics, using the VLAN software.
- Automatic Configuration :** In this configuration the stations are automatically connected and disconnected from a VLAN using criteria defined by the administrator.

Storage Area NIW

Q.46 Write difference between VLAN and VSAN. [RGPV : Dec.-14, Marks 7, June-15, Marks 7]

Ans. : Virtual Local Area Network (VLAN) is the partitioning of a single network into multiple, distinct domains. It's used to simplify network design, allowing the grouping of hosts with common requirements, regardless of physical location.

- The VLAN ID is a number that represents that particular VLAN.

The concept of naming a VLAN is used as it provides a layer of abstraction which is helpful in situations like where you need to update servers.

- You create a VLAN which is visible in either one fabric interconnect or both fabric interconnect. It is also possible to create separate VLANs with same name on both of the fabric interconnects but with different VLAN IDs. You can also create more than one named VLANs with the same VLAN ID.

VSAN :

- VSAN can be created on either one of the fabric interconnects or both of the fabric interconnects.
- A VSAN is identified with a unique ID, which is a number, and is also assigned a name.
- VSANs can also be created with same name but different IDs on both of the fabric interconnects.

Q.47 What is difference between process virtual machines, host VMMS and native VMMS ? [RGPV : Dec.-13, Marks 7]

Ans. : Process Virtual Machines :

- By using a process VM, a guest program developed for a computer (ISA and OS) other than the user's host system can be used in the same way as all other programs in the host system.
- Process virtual machine provides a platform for the execution of a single program (process). Example: Linux process, Java VM, .NET VM.
- Also known as Application VM.
- Virtualization below the API or ABI, providing virtual resources to a single process executed on a machine. It is created for the process alone, destroyed when process finishes.

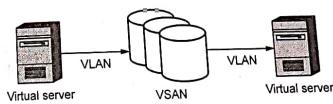


Fig. Q.45.1

- 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.

- Host VMM :**
- A host virtual machine is the server component of a Virtual Machine (VM), the underlying hardware that provides computing resources to support a particular guest Virtual Machine (guest VM).

Fig. Q.47.1 shows Host VMM.

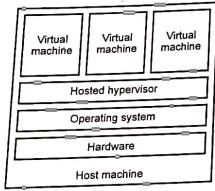


Fig. Q.47.1

- The host virtual machine and the guest virtual machine are the two components that make up a virtual machine. The guest VM is an independent instance of an operating system and associated software and information.
- The host VM is the hardware that provides it with computing resources such as processing power, memory, disk and network I/O (input/output), and so on.
- A Virtual Machine Monitor (VMM) or hypervisor intermediates between the host and guest VM, isolating individual guest VMs from one another and making it possible for a host to support multiple guests running different operating systems.
- VMM is computer software, firmware or hardware that creates and runs virtual machines. A computer on which a hypervisor runs one or more virtual machines is called a host machine, and each virtual machine is called a guest machine.

Native VMM :

- Native hypervisors run directly on the host machine and share out resources (such as memory and devices) between guest machines.

Fig. Q.47.2 shows native VMM.

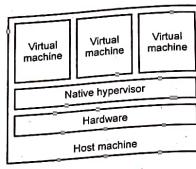


Fig. Q.47.2

Q.48 Describe various cloud storage providers ? Also write their merits and demerits of cloud storage.

[RGPV : Dec.-17, Marks 7]

Ans. : A cloud storage provider is an organization that offers organizations and individuals the ability to place and retain data in an off-site storage system. Customers can lease cloud storage capacity per month or on demand.

- A cloud storage provider hosts a customer's data in its own data center, providing fee-based computing networking and storage infrastructure. Both individual and corporate customers can get unlimited storage capacity on a provider's servers at a low per-gigabyte price.
- Top rated cloud storage providers include : Google Drive, Google Cloud Storage, Dropbox, DropBox Business, Citrix ShareFile, Amazon Cloud Drive, and Box.
- Other cloud storage providers on the TrustMap include : Egnyte, OneDrive, ownCloud, Apple iCloud, MediaFire, Cubby and SugarSync

- Advantages of using a cloud storage provider are cost control, elasticity and self-service provisioning.

Characteristic	Public cloud storage	Private cloud storage	Hybrid cloud storage
Scalability	Very high	Limited	Very high
Security	Good, but depends on the security measures of the service provider	Most secure, as all storage is on-premises	Very secure, integration options add an additional layer of security
Performance	Low to medium	Very good	Good, as active content is cached on-premises
Reliability	Medium; depends on internet connectivity and service provider availability	High, as all equipment is on-premises	Medium to high, as cached content is kept on-premises, but also depends on connectivity and service provider availability
Cost	Very good, pay-as-you-go model and no need for on-premises storage infrastructure	Good, but requires on-premises resources, such as data center space, electricity and cooling	Improved, since it allows moving some storage resources to a pay-as-you-go model

END...ES

4

Cloud Security

Important Points to Remember

- Cloud computing security consists of a set of policies, controls, procedures and technologies that work together to protect cloud-based systems, data and infrastructure.
- The data confidentiality, authentication, and access control issues in cloud computing could be addressed by increasing the cloud reliability and trustworthiness.
- Data integrity in the cloud system means preserving information integrity. The data should not be lost or modified by unauthorized users.
- Cloud Security Services are Authentication, Authorization, Auditing and Accountability.
- Cloud security design principles are Least Privilege, Separation of Duties, Defense in Depth, fail Safe, Economy of Mechanism and Complete Mediation.
- Cloud security is a shared responsibility of the cloud provider and customer.
- A multi-tenant cloud is a cloud computing architecture that allows customers to share computing resources in a public or private cloud.
- Risk analysis of the applications for both security and privacy are performed; threat models are built and maintained.
- Cloud Computing architecture patterns which explicitly mitigate threats such as from the Open Security Architecture are used.

4.1 : Cloud Information Security Fundamentals

Q.1 Define cloud security.

Cloud computing security consists of a set of policies, controls, procedures and technologies that work together to protect cloud-based systems, data and infrastructure. These security measures are configured to protect data, support regulatory compliance and

protect customers' privacy as well as setting authentication rules for individual users and devices.

Q.2 Discuss different attributes of cloud security. [RGPV : June-16, marks-3]

Ans.: Different attributes of cloud security are as follows :

- Confidentiality, Privacy and trust
Confidentiality gets compromised if an unauthorized person is able to access a message. Trust is an important aspect of decision making for Internet applications and particularly influences the specification of security policy. It implies depth and assurance of confidence based on some evidence.
- Data security : Data security becomes particularly serious in the cloud computing environment, because data are scattered in different machines and storage devices including servers, PCs, and various mobile devices such as wireless sensor networks and smart phones.
- Enterprise Cloud services Security : Enterprise Cloud Computing refers to a computing environment residing behind a firewall that delivers software, infrastructure and platform services to an enterprise

Q.3 Explain the cloud information security fundamental introduced by cloud security management. [RGPV : Dec-14, June-15, Marks-7]

Ans.: • Cloud security is important for both business and personal users. Everyone wants to know that their information is safe and secure and businesses have legal obligations to keep client data secure, with certain sectors having more stringent rules about data storage

• Fig. Q.3.1 shows organization of data security and privacy in cloud computing.

(4 - 1)

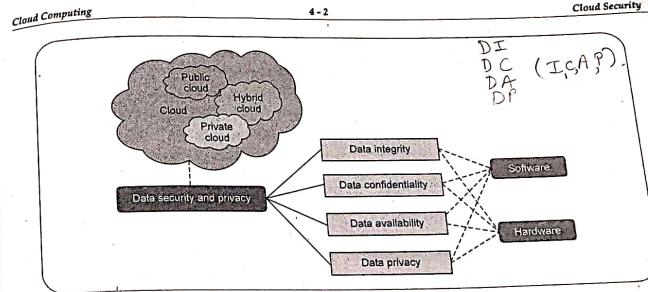


Fig. Q.3.1

• Cloud Information security fundamental introduced by cloud security management are Confidentiality, Integrity, and Availability.

1. Data Confidentiality :

- Data confidentiality is important for users to store their private or confidential data in the cloud. Authentication and access control strategies are used to ensure data confidentiality.

• The data confidentiality, authentication, and access control issues in cloud computing could be addressed by increasing the cloud reliability and trustworthiness.

• Because the users do not trust the cloud providers and cloud storage service providers are virtually impossible to eliminate potential insider threat, it is very dangerous for users to store their sensitive data in cloud storage directly.

• Simple encryption is faced with the key management problem and cannot support complex requirements such as query, parallel modification, and fine-grained authorization

2. Data Integrity :

- Data integrity in the cloud system means preserving information integrity. The data should not be lost or modified by unauthorized users.

• Data integrity in the cloud system means preserving information integrity. The data should not be lost or modified by unauthorized users.

• Data integrity is the basis to provide cloud computing service such as SaaS, PaaS, and IaaS.

• Besides data storage of large-scaled data, cloud computing environment usually provides data processing service. Data integrity can be obtained by techniques such as RAID-like strategies and digital signature.

3 Data Availability :

• Data availability means the following : when accidents such as hard disk damage, IDC fire, and network failures occur, the extent that user's data can be used or recovered and how the users verify their data by techniques rather than depending on the credit guarantee by the cloud service provider alone.

• The cloud service provider should ensure the data security, particularly data confidentiality and integrity. The cloud provider should share all such concerns with the client and build trust relationship in this connection. The cloud vendor should provide guarantees of data safety and explain jurisdiction of local laws to the clients.

Q.4 Describe the top threats identified by Cloud Security Alliance (CSA) of cloud computing. [RGPV : Dec-13, Marks-7]

Ans.: Top threats identified by Cloud Security Alliance (CSA) of cloud computing are as follows:

1. Insecure Interfaces and APIs : Cloud Computing providers expose a set of software interfaces or APIs that customers use to manage and interact with cloud services. Provisioning, management, orchestration, and monitoring are all performed using these interfaces.

- Remediation : Analyze the security model of cloud provider interfaces.

2. Malicious Insiders : The threat of a malicious insider is well-known to most organizations. This threat is amplified for consumers of cloud services by the convergence of IT services and customers under a single management domain, combined with a general lack of transparency into provider process and procedure.

- Remediation : Determine security breach notification processes.

3. Shared Technology Issues : IaaS vendors deliver their services in a scalable way by sharing infrastructure.

- Remediation : Implement security best practices for installation/configuration. Monitor environment for unauthorized changes/activity.

4. Data Loss or Leakage : There are many ways to compromise data. Deletion or alteration of records without a backup of the original content is an obvious example. Unlinking a record from a larger context may render it unrecoverable, as can storage on unreliable media. Loss of an encoding key may result in effective destruction.

- Remediation : Implement strong API access control. Encrypt and protect integrity of data in transit.

5. Account or Service-Hijacking : Attack methods such as phishing, fraud, and exploitation of software vulnerabilities still achieve results. Credentials and passwords are often reused, which amplifies the impact of such attacks.

- Remediation : Prohibit the sharing of account credentials between users and services. Leverage

4.2 : Cloud Security Services, Design Principles

Q.7 Discuss the different cloud security services. [RGPV : Dec-15, Marks-7]

Ans.: Cloud security services are authentication, authorization, auditing and accountability.

Q.8 Explain the categories of security services provided for information over the cloud. [RGPV : June-16, Marks-7]

OR Discuss the different cloud security services. [RGPV : Dec-14, Marks-7]

Ans.: Cloud Security Services are Authentication, Authorization, Auditing and Accountability.

1. Authentication :

- Authentication is the act of confirming the truth of an attribute of a single piece of data claimed true by an entity.

- Authentication is verifying the identity of a user, process, or device, often as a prerequisite to allowing access to resources in an information system.

2. Authorization :

- Authorization is the function of specifying access rights/privileges to resources related to information security and computer security in general and to access control in particular.

- Authorization determines what the user can access and what he cannot access.

3. Accountability :

- This is the process that keeps track of a user's activity while attached to a system; the trail included the amount of time attached, the resources accessed, and how much data transferred.

- Accounting data is used for trending, detecting breaches, and forensic investigating. Keeping track of users and their activities serves many purposes.

- For example, tracing back to events leading up to a cyber security incident can prove very valuable to a forensics analysis and investigation case.

4. Auditing :

- Cloud security audit can help by assessing and prioritizing risks, evaluating current controls, identifying the gaps in existing cloud security strategy and programs, and making recommendations tied to business priorities.

- Functions performed by IT auditors:

- a. Backup controls
- b. Data center security
- c. System development standards
- d. System and transaction controls
- e. Contingency plan

Q.9 Write a brief notes on cloud security design principles. [RGPV : June-15, Marks-7]

Ans.: Cloud security design principles are as follows:

- a) Least Privilege
- b) Separation of Duties
- c) Defense in Depth
- d) Fail Safe
- e) Economy of Mechanism
- f) Complete Mediation

g) Open Design

h) Least Common Mechanism

i) Psychological Acceptability

j) Weakest Link

k) Leveraging Existing Components

- Least Privilege : The principle of least privilege simply states that people or automated tools should be able to access only what they need to do their jobs, and no more.

- Defense in depth is an acknowledgment that almost any security control can fail, either because an attacker is sufficiently determined or because of a problem with the way that security control is implemented. With defense in depth, you create multiple layers of overlapping security controls so that if one fails, the one behind it can still catch the attackers.

- Separation of Duties (SoD) is an increasingly common concept in internal controls that essentially requires more than one person to complete a transaction or task in an effort to reduce fraud.

Cloud Computing

4-5

- Fail Safe : When cloud system fails it should fail to a state where the security of the system and its data are not compromised.
- Economy mechanism : Complexity is the enemy of security. It's just too easy to screw things up in a complicated system, both from a design perspective and from an implementation perspective. Developers may create user interfaces made to be as simple as possible by eliminating buttons and dialog boxes that may potentially confuse the user.
- Complete Mediation : Every access and every object should be checked, every time. Make sure your access control system is thorough and designed to work in the multi-threaded world we all inhabit today. Make sure that if permissions change on the fly in your system, that access is systematically rechecked. Don't cache results that grant authority or wield authority.
- Open design : open access cloud system design offers a more secure authentication method.
- Least common mechanism : It enhances the least possible sharing of common security mechanism.
- Psychological Acceptability : If your security mechanisms are too annoying and painful, your users will go to great length to circumvent or avoid them. Make sure that your security system is as secure as it needs to be, but no more
- Weakest link : It is important to identify the weakest mechanisms in the security chain and layers of defense, and improve them so that risks to the system are mitigated to an acceptable level

4.3 : Secure Cloud Software Requirements, Policy Implementation

Q.10 How security policies are implemented on cloud computing ? ES² [RGPV : Dec-15, Marks-2]

Ans. : Cloud security is a shared responsibility of the cloud provider and customer.

Step 1 : Perform data classification (Statement of Sensitivity);

Step 2 : Perform Threat Risk Assessment on the solution;

- Step 3 :** Address threats/risks identified by implementing the proper controls;
- Step 4 :** Continuously monitor and periodically audit systems and services

Q.11 Explain the requirements of secure cloud software. ES² [RGPV : Dec-16, Marks-7]

Ans. : Requirements of secure cloud software are as follows :

1. **Secure Development Practices** : it includes data handling, Code Practices, Language Options, Input Validation and Content Injection, Physical Security of the System
2. **Approaches to Cloud Software Requirements Engineering** : A Resource Perspective on Cloud Software Security Requirements, Goal-Oriented Software Security Requirements and Monitoring Internal and External Requirements
3. **Cloud Security Policy Implementation**, and Decomposition includes implementation Issues, decomposing Critical Security Issues into Secure Cloud Software Requirements (Confidentiality, Integrity, Availability), Authentication and Identification, Authorization, Auditing)

• Also refer Q.10.

Q.12 What is Identity and Access Management ?

Ans. : • The components and policies necessary to control and track user identities and access privileges for IT resources, environments and systems

• Comprised of authentication, authorization, user management and credential management

Implementation mechanisms

- Authentication : Username and password pair; typical user authentication credentials managed by IAM. Additional mechanism : digital signature, digital certificates, biometric hardware (fingerprint reader), specialized software (voice recognition), locking user accounts to registered IP/MAC address, etc.
- Authorization : Access controls based on relationships between identities, access control rights and IT resource availability
- User management : Administrative capabilities including creating new user identities & access

DECODE

Cloud Computing

4-6

- group, resetting passwords, defining password policies and managing privileges
- Credential management : Credential management including establishing identities and access control rules for defined user account which mitigates the threat of insufficient authorization

Q.13 What is Single Sign-On (SSO) ?

Ans. : • A mechanism enabling one cloud service consumer to be authenticated by a security broker which establishes a security context that is persisted while the cloud service consumer accesses other cloud services or cloud-based IT resources in order for the cloud service consumer not to re-authenticate itself with every subsequent request

Implementation mechanisms

- Not a trivial job at all to propagate the authentication and authorization information for a cloud service consumer across multiple cloud services, especially with a numerous cloud services or cloud-based IT resources to be invoked as part of the same overall runtime activity
- SSO (or security broker) mechanism to enable mutually independent cloud services and IT resources to generate and circulate runtime authentication and authorization credentials (security token) in order to allow the credentials provided by the cloud service consumer at its login time to be valid through out the duration of the same session

• Security brokerage mechanism is especially useful when a cloud service consumer needs to access cloud services residing on different clouds.

- Not to counter security threats directly, but to enhance the usability of cloud-based environments for access and management of distributed IT resources and solutions without violating security policies

4.4 Cloud Computing Security Challenges

Q.14 What is multitenancy issue in cloud computing ? ES² [RGPV : Dec-16, Marks-2]

Ans. : A multi-tenant cloud is a cloud computing architecture that allows customers to share computing

TECHNICAL PUBLICATIONS - An up thrust for knowledge

resources in a public or private cloud. Each tenant's data is isolated and remains invisible to other tenants.

Q.15 Discuss the problem associated with cloud computing. ES² [RGPV : Dec-16, Marks-2]

Ans. : • Problem associated with cloud computing are security, integration and interoperability, governance and regulatory compliance.

- Some governments or enterprises may need to enforce strict limits on the spatial and temporal existence of data. For example, a government might want to keep the data of its citizens within the country and for an exact duration

Q.16 List few cloud computing security challenges. ES² [RGPV : June-16, Marks-2]

OR List few major challenges in securing cloud computing. ES² [RGPV : Dec-15, Marks-3]

Ans. : Cloud computing security challenges fall into three broad categories:

1. Data Protection : Securing your data both at rest and in transit
 2. User Authentication : Limiting access to data and monitoring who accesses the data
 3. Disaster and Data Breach : Contingency Planning
- Data Protection: Data needs to be encrypted at all times, with clearly defined roles when it comes to who will be managing the encryption keys.
 - User Authentication : Data resting in the cloud needs to be accessible only by those authorized to do so, making it critical to both restrict and monitor who will be accessing the company's data through the cloud. In order to ensure the integrity of user authentication, companies need to be able to view data access logs and audit trails to verify that only authorized users are accessing the data.
 - Contingency Planning : With the cloud serving as a single centralized repository for a company's mission-critical data, the risks of having that data compromised due to a data breach or temporarily made unavailable due to a natural disaster are real concerns.

repository : a place where things may be stored.

wrapping : a future event or circumstance which is remote but can't be predicted surely.

Q.17 List and explain various cloud computing security challenges. [RGPV : June-15, Marks 7]

OR What are cloud security challenges ? How is a security provided to data at various stages in context of cloud? [RGPV : Dec-17, Marks 7]

Ans. : Security challenges for Cloud Service Customers :

1. Ambiguity in responsibility : A CSC uses services based on different service categories as well as different deployment models. If the responsibilities are not clearly defined in any of these cases then it may result in inconsistency or may leave an open gate for attacks.
2. Loss of trust : Because of the abstraction of the security implementation details between a CSC and a CSP, it is difficult for a CSC to get details of the security mechanisms that the CSP has implemented to keep the cloud data secure.
3. Loss of governance : When the CSC uses cloud services, it has to move its data onto the cloud and has to provide certain privileges to the CSP for handling the data in the cloud. This may result in misconfiguration or an attack due to the abstraction of the CSP's cloud practices and due to the privileges that need to be given to the CSP.
4. Loss of privacy : CSC's privacy may be violated due to leakage of private information while the CSP is processing CSC's private data or using the private information for a purpose that the CSP and CSC haven't agreed upon.
5. Cloud service provider lock-in : This issue arises if a CSP doesn't abide by the standard functions or frameworks of cloud computing and hence makes it difficult for a CSC using its services to migrate to any other CSP. The use of non-standard functions and cloud framework makes the CSP non-inter-operable with other CSPs and also leaves CSC open to security attacks.
6. Misappropriation of intellectual property : A CSC may face this challenge due to the possibility that a CSC's data on the cloud might leak to third parties that are using the same CSP for their cloud services. This leakage may violate

the CSC's copyrights and may result in the disclosure of CSC's private data.

7. Loss of software integrity : A CSC encounters this challenge due to the fact that its software is running in the cloud once it is given to the CSP. It is possible that this software might be tampered with or might be affected while the software is running in the CSP and is not in CSC's control, resulting in CSC's loss over its software.

Q.18 Explain the risk from multi-tenancy with respect to various cloud environments. [RGPV : Dec-13, Mark 7]

Ans. : • The multitenant application design was created to enable multiple users (tenants) to access the same application logic simultaneously.

• Each tenant has its own view of the application that it uses, administers, and customizes as a dedicated instance of the software while remaining unaware of other tenants that are using the same application

• Risks in multi-tenancy environment :

1. Inadequate Logical Security Controls : Physical resources are shared between multiple tenants. That means dependence on logical segregation to ensure that one tenant deliberately cannot interfere with the security of the other tenants.
2. Malicious or Ignorant Tenants : If the provider has weaker logical controls between tenants, a malicious or an ignorant tenant may reduce the security posture of other tenants.
3. Shared Services can become single point of failure.
4. Performance Risks : One tenant's heavy use of the service may impact the quality of service provided to other tenants.
5. Uncoordinated Change Controls and Misconfigurations : When multiple tenants are sharing the underlying infrastructure all changes needs to be well coordinated and tested.

Q.19 What is multitenant technology ? What is role of tenants ? Explain benefit of multitenant technology.

Ans. : • A multi-tenant cloud is a cloud computing architecture that allows customers to share computing resources in a public or private cloud. Each tenant's data is isolated and remains invisible to other tenants

Ans. : • Common characteristics of multitenant applications are as follows :

1. Usage Isolation - The usage behavior of one tenant does not affect the application availability and performance of other tenants.
2. Data Security - Tenants cannot access data that belongs to other tenants.
3. Recovery - Backup and restore procedures are separately executed for the data of each tenant.
4. Application Upgrade - Tenants are not negatively affected by the synchronous upgrading of shared software artifacts.
5. Scalability - The application can scale to accommodate increases in usage by existing tenants and/or increases in the number of tenants.
6. Metered Usage - Tenants are charged only for the application processing and features that are actually consumed.
7. Data Tier Isolation - Tenants can have individual databases, tables, and schemas isolated from other tenants.

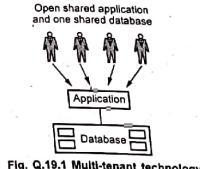


Fig. Q.19.1 Multi-tenant technology

2. Business Process : Tenants can customize the rules, logic, and workflows of the business processes that are implemented in the application.

3. Data Model : Tenants can extend the data schema of the application to include, exclude, or rename fields in the application data structures.

4. Access Control : Tenants can independently control the access rights for users and groups.

• Benefits of a Multitenancy technology :

1. Costs savings : It yields tremendous economy of scale for the provider so he can offer the service at a lower cost to customers.
2. Improved quality, user satisfaction, and customer retention : a multitenant application is one large community hosted by the provider which can gather operational information from the collective user population and make frequent, incremental improvements to the service that benefit the entire user community at once.
3. Improved security : most current enterprise security models are perimeter-based, making them vulnerable to inside attacks

Q.20 List and explain common characteristics of multitenant applications.

4.5 Virtualization Security Management

Q.22 What do you understand by virtualization security management ? [RGPV : June-16, Marks 2]

Ans. : Virtualization security is the collective measures, procedures and processes that ensure the protection of a virtualization infrastructure / environment. It addresses the security issues faced by the components of a virtualization environment and

methods through which it can be mitigated or prevented

Q.23 Write short note : Virtualization security Management. [RGPV : Dec-17, Marks 4]

OR Explain virtualization security management in cloud computing. [RGPV : Dec-15, Marks 7]

OR Explain virtualization security management.

[RGPV : Dec-14, Marks 7]

Ans. : • Different virtualization approaches can be applied to various system layers including hardware, desktop, operating system, software, memory, storage, data and network.

• Full virtualization is a form of hardware virtualization that involves complete abstraction of underlying hardware and provides better operational efficiency by putting more work load on each physical system.

• Full virtualization can be categorized into two forms: bare metal virtualization and hosted virtualization.

• Bare metal approach is mostly used for server virtualization in large computing systems like Cloud computing as it provides better performance, more robustness and agility.

• Hypervisor is the underlying component of all these architectures. It is a new layer which needs to be protected

• Unauthorized person should not have physical access to the virtualization hardware of the system.

• In order to protect VMs from unauthorized access by Cloud administrators, each VM can be assigned access control that can only be set through Hypervisor.

• The three core principles of access control namely identification, authentication and authorization will restrict admin access from unauthorized data and system components.

• Hypersafe is a system that maintains code integrity of the Hypervisor. It extends the hypervisor implementation and prevents its code modification by locking down the write-protected memory

- It secures the Hypervisor against the control-flow hijacking attacks by protecting its code from unauthorized access.

• Organizations using virtualization must have a policy to manage the creation, usage, storage and deletion of images. Images must be scanned for the detecting viruses, worms, spyware and rootkits that hide themselves from security software running in guest OS.

• When VMs are to be migrated from one physical machine to another, Cloud admin must recheck and ensure that all data is removed from previous or broken disks.

• To protect the backup VM images cryptographic techniques such as encryption may be employed to encrypt all backup images. If any VM is deleted then its backup must also be removed from system.

• Checkpoint attacks can be prevented by encrypting the checkpoint files. Another mechanism to provide security to Checkpoints is SPARC.

• **SPARC** is a mechanism designed to deal with security and privacy issues resulting from VM checkpoint. SPARC enables users to select applications that they want to check-point so sensitive applications and processes can not be check pointed.

4.6 Cloud Computing Security Architecture

Q.24 Explain the cloud computing security architecture using suitable block diagram.

[RGPV : Dec-15, Marks 7]

Ans. : • Fig. Q.24.1 shows cloud computing security architecture.

• This infrastructure provides the storage and networking components to cloud networking. It relies heavily on Application Programming Interfaces (APIs) to allow enterprises to manage and interact with the cloud.

• While all cloud architecture models require performance management tools and strategy, the security architecture varies based on the type of cloud model : Software-as-a-Service (SaaS), Infrastructure-as-a-service (IaaS), or Platform-as-a-Service model (PaaS).

UNIT - V

5

Market Based Management of Clouds

Important Points to Remember

- Market oriented cloud computing is the presence of a virtual market place where IT Service are traded and dynamically.
- Cloud marketplace is an online storefront operated by a cloud service provider.
- Main components that implement a MOCC system are directory, auctioneer and bank.
- A federated cloud also called cloud federation is the deployment and management of multiple external and internal cloud computing services to match business needs.
- Eucalyptus is an open-source software platform that provides the platform for private cloud computing implementation on computer clusters.

5.1 : Market Based Management of Clouds

Q.1 What is market oriented cloud computing ?

Ans. : Market oriented cloud computing is the presence of a virtual market place where IT Service are traded and dynamically.

Q.2 What is Cloud Marketplace ?

Ans. : Cloud marketplace is an online storefront operated by a cloud service provider. It provides services to customers. Customers can access software applications and services. A marketplace typically provides customers with native cloud applications and approved apps created by third-party developers.

Q.3 What is market oriented cloud computing? *M OCC*
List the main components that implement a MOCC system. [SSPU : Dec-14, Marks 7]

OR What are the main components that implements a MOCC based system
[SSPU : Dec-14, Marks 7]

1. Directory : It is market directory which contains a listing of all the published services that are available in the cloud marketplace.
2. Auctioneer : The auctioneer is in charge of keeping track of the running auctions in the market place and of verifying that the auctions for services are properly conducted and that malicious market players are prevented from performing illegal activities.

OR Explain the main components that implements a MOCC based system [SSPU : June-15, Marks 7]

- Ans. : • Main components that implement a MOCC system are directory, auctioneer and bank.
- Fig. Q.3.1 shows market oriented cloud computing.

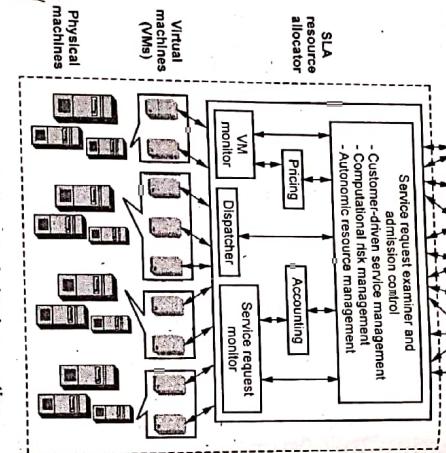


Fig. Q.3.1 : market oriented cloud computing

5.2 : Federated Clouds/Inter Cloud

Q.4 Discuss market based management of clouds. [SSPU : June-16, Marks 3]

OR What is market based management of clouds? Explain in details. [SSPU : Dec-15, Marks 7]

- Ans. : • Marketplace of a typical Cloud is composed of varied types of services it offers, such as content delivery to enterprises and end-users, computing and storage.

- The users of the cloud interact with its market by making requests for acquiring resources as per need or through applications that leverages the Cloud environment.
- Currently, Cloud Providers are facilitating the users to access the Cloud services. The communication happens through a market maker also called as meta-broker component which is responsible for provisioning of the best service taking into consideration budget constraints of the user.

- The client of cloud broker is often embedded within the applications directly or can be seen as a separate tool.
- The client communicates with the market maker by specifying the required QoS parameters through establishing Service Level Agreement.
- Among the raised requests, the meta-broker selects the best request option among the cloud providers. Such interaction is made possible via standardized brokering services or through interfaces exposed by provider.

Q.5 What is Federation in cloud? [MCC]

- Ans. : A federated cloud also called cloud federation is the deployment and management of multiple external and internal cloud computing services to match business needs. A federation is the union of several smaller parts that perform a common action.

(1) Director of Marketing
(2) Banker

(3) Randa

Market Based Management of Clouds

Q.6 What is Inter cloud ?

- Ans. : Mesh of cloud is called as inter cloud, which is interconnected using open standards to provide a universal acceptance.

- Intercloud Architecture Framework (ICAF) provides a framework to support provisioning of cloud based project oriented infrastructures on-demand and distributed virtualised applications mobility.

- Each cloud should be able to work and offer its services without any dependence with other clouds.
- Resources, services and data are shared through the intercloud architecture.
- 2. The intercloud architecture is scalable and able to add new clouds.

1. Resources, services and data are shared through the intercloud architecture.
3. The availability of the resources, services and data should not depend on the customer's applications.
4. The architecture should be able to provide better load balancing capabilities.

- The main objective of intercloud is to create an open interface to ease the exchange of data from one cloud to another. The connections are established between one or more clouds for this systematic exchange of data.
- The main objective of intercloud is to create an open interface to ease the exchange of data from one cloud to another. The connections are established between one or more clouds for this systematic exchange of data.

- Ans. : Interoperability in cloud federation is achieved by using following standards and protocols :
1. Open Cloud Manifesto : first step towards the realizations of a cloud inter-operability platform. The Cloud Computing Manifesto is a manifesto containing a "public declaration of principles and intentions" for cloud computing providers and vendors, annotated as "a call to action for the worldwide cloud community" and "dedicated belief that the cloud should be open". It follows the earlier development of the Cloud Computing Bill of Rights which addresses similar issues from the users' point of view.

2. Distributed Management Task Force (DMTF) : The DMTF has made a significant contribution toward the realization of an interoperable cloud environment. With specific reference to cloud

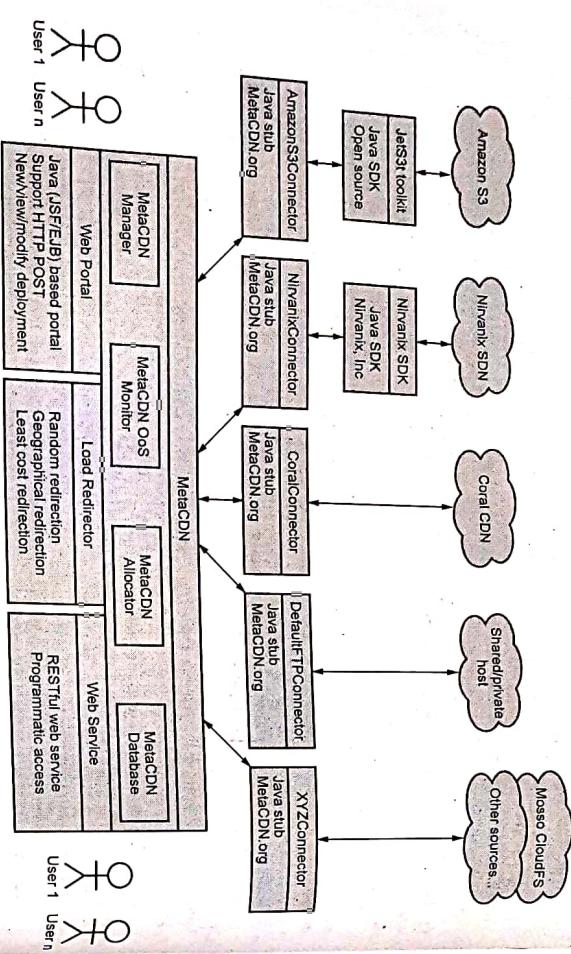


Fig. Q.14.1 Metacdn

credentials for any cloud storage or other provider they have an account with.

- Once this simple step has been performed, they can utilize the MetaCDN system to intelligently deploy content onto storage providers according to their performance requirements and budget limitations.
- The web portal is most suited for small or ad-hoc deployments, and is especially useful for less technically inclined content creators.
- The second method of accessing the MetaCDN service is via RESTful Web Services.
- These Web Services expose all of the functionality of the MetaCDN system.
- This access method is most suited for customers with more complex and frequently changing content delivery needs, allowing them to integrate the MetaCDN service in their own origin web sites and content creation workflows.

- The goal of Eucalyptus is to allow sites with existing clusters and server infrastructure to host a cloud that is interface-compatible with Amazon's AWS and the Sun Cloud open API.
- Eucalyptus cloud computing architecture is highly scalable because of its distributed nature. The Cloud level of the computing architecture is comprised of only two components and while used by many users, the transactions at each component are typically small.

Q.15 What Is Third Party Cloud Services ? Give some examples [SPPU : Dec-14, June-15, Marks 7]

- The Node level may have many components, but each component only supports a few users, even though the transactions are larger. This distributed cloud architecture is flexible enough to support businesses of any size.

Ans. : • SpotCloud marketplace platform provides an easy way to sell unused cloud capacity. Cloud providers can use SpotCloud to clear out unused capacity and sell computing inventory that would otherwise go unsold, enabling increased utilization and revenue, without undermining their standard pricing

- SpotCloud bridges many disparate regional cloud providers, allowing buyers to find the best cloud providers at the best price.
- SpotCloud serves as the central place to discover and buy computing capacity, based on performance and location parameters, through a simple and easy to use web dashboard and API.

Q.16 Explain conceptual representation of components within Eucalyptus system. [SPPU : Dec-13, Marks 7]

- Eucalyptus stands for Elastic Utility Computing Architecture for Linking Your Programs to Useful Systems.
- It is an open-source software framework that provides the platform for private cloud computing implementation on computer clusters.
- Eucalyptus implements infrastructure as a service (IaaS) methodology for solutions in private and hybrid clouds.
- Eucalyptus provides a platform for a single interface so that users can calculate the resources available in private clouds and the resources available externally in public cloud services.

- The MetaCDN service has a number of core components that contain the logic and management layers required to encapsulate the functionality of different upstream storage providers and present a consistent, unified view of the services available to end-users.
- These components include the MetaCDN Allocator, which selects the optimal providers to deploy content to, and performs the actual physical deployment.
- The MetaCDN QoS monitor tracks the current and historical performance of participating storage providers, and the MetaCDN Manager tracks each user's current deployment and performs various housekeeping tasks.
- The MetaCDN Database stores crucial information needed by the MetaCDN portal, ensuring reliable and persistent operation of the system.

- The goal of Eucalyptus is to allow sites with existing clusters and server infrastructure to host a cloud that is interface-compatible with Amazon's AWS and the Sun Cloud open API.
- Eucalyptus cloud computing architecture is highly scalable because of its distributed nature. The Cloud level of the computing architecture is comprised of only two components and while used by many users, the transactions at each component are typically small.

Ans. : • Eucalyptus cloud platform pools together existing virtualized infrastructure to create cloud resources for infrastructure as a service, network as a service and storage as a service.

- The Node level may have many components, but each component only supports a few users, even though the transactions are larger. This distributed cloud architecture is flexible enough to support businesses of any size.

Q.15 What Is Third Party Cloud Services ? Give some examples [SPPU : Dec-14, June-15, Marks 7]

- The Node level may have many components, but each component only supports a few users, even though the transactions are larger. This distributed cloud architecture is flexible enough to support businesses of any size.

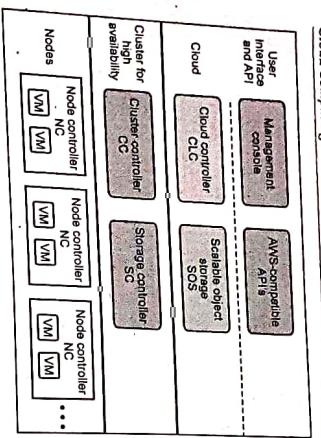


Fig.Q.16.1 Eucalyptus software architecture

2. Instances: When we run the image and use it, it becomes an instance. The controller will decide how much memory to allocate and provide all other resources.

3. Networking: The Eucalyptus network is divided into three modes:
- Managed mode: In this mode, it assigns a local network of instances, which includes security groups and IP addresses.
 - System mode: In this mode, it assigns a MAC address and attaches the instances network interface to the physical network through the NC's bridge.
 - Static mode: In this mode, it assigns IP addresses to instances.

Q.17 List major third party cloud services. What are the applications of these services?

[SPU : Dec-15, Marks 7]

Ans.: • Major third party cloud services are MetaCDN and SocioCloud.

• MetaCDN provides users with the high-level services of a CDN for content distribution. It offers video trans-coding, streaming video and web accelerator services.

• SocioCloud is used to provide restriction to users. Each user will get a unique identity. All identities can be grouped and managed by access control.

5. Eucalyptus elastic block storage (EBS) provides block-level storage volumes, which we can attach to an instance.

6. Auto scaling and load balancing is used to automatically create or destroy instances or services based on requirements. CloudWatch provides different metrics for measurement.

• Cloud controller: Cloud controller (CLC) is the main controller, which manages the entire cloud platform. It provides a Web and EC2 compatible interface. All the incoming requests come through the Cloud controller.

- Walrus: This is similar to AWS S3. It provides persistent storage to all the instances. It can contain any kind of data like application data, volume or image snapshots.
- Cluster controller: This is the heart of the cluster within a Eucalyptus cloud. It manages VM (instance) execution and service level agreements. It communicates with the storage and network controller.
- Storage controller: Storage controller (SC) is similar to AWS EBS. It provides block level storage to instances and snapshots within a cluster. If an instance wants persistent storage outside of storage, then it must pass through Walrus. The storage controller doesn't handle this kind of request.
- Node controller: NC (node controller) hosts all the instances and manages their end points. There is no limit to the number of NCs, in the Eucalyptus cloud. It takes images and also caches from Walrus and creates instances. One should manage the number of NCs used as it affects the performance.

Q.18 What is Azure queues? [SPU : Dec-16, Marks 2]

Ans.: Azure Queue storage is a service for storing large numbers of messages that can be accessed from anywhere in the world via authenticated calls using HTTP or HTTPS. A single queue message can be up to 64 KB in size, and a queue can contain millions of messages, up to the total capacity limit of a storage account.

Q.19 How virtualization employed in Azure? [SPU : Dec-16, Marks 2]

Ans.: Azure is a virtualized infrastructure to which a set of additional enterprise services has been layered on top, including a virtualization service called Azure AppFabric that creates an application hosting environment. AppFabric is a cloud-enabled version of the .NET Framework.

Q.20 Discuss in brief Aneka for cloud computing. [SPU : June-16, marks 2]

Ans.: • Aneka is a software platform for developing cloud computing applications.

• Aneka is a platform and a framework for developing distributed applications on the Cloud. It harnesses the spare CPU cycles of a heterogeneous network of desktop PCs and servers or datacenters on demand.

• Aneka provides developers with a rich set of APIs for transparently exploiting such resources and expressing the business logic of applications by using the preferred programming abstractions.

• Google App Engine offers users the ability to build and host web applications on Google's infrastructure.

• Google App Engine offers users the ability to build and host web applications on Google's infrastructure.

Q.21 Discuss the cloud features of Aneka.

[SPU : Dec-16, Marks 3]

- For service consumers, it acts as a market directory where they can browse and compare different IaaS offerings and select the most appropriate solution for them.

5.5 : Case study: Google App Engine, Microsoft Azure, Hadoop , Amazon , Aneka

- At the application level, a set of different components and tools are provided to:
 - simplify the development of applications (SDK);
 - porting existing applications to the Cloud; and
 - monitoring and managing the Aneka Cloud.
- Aneka provides APIs and tools that enable applications to be virtualized over a heterogeneous network.

Q.22 List the major feature of Google App Engine. Which kind of problems can be solved using Google App Engine? [SPU : Dec-15, Marks 2]

Ans.: Major feature of Google App Engine:

- Automatic scaling and load balancing
- Authentication using Google Accounts API
- Provides dynamic web services based on common standards
- Integration with other Google cloud services and API
- Support persistent storage, with query access sorting and transaction management features

Q.23 What are the major functionalities of Hadoop API? Explain. [SPU : Dec-15, Marks 2]

Ans.: • Hadoop provides scalable analytics.

- It is possible to store the historical data longer
- It can handle unstructured data.

• Hadoop permits to capture new or more data.

Q.24 Describe the major cloud features of Google applications engine. [SPU : Dec-13, Marks 7]

- OR Explain the major cloud features of Google application engine.** [SPPU : Dec-16, Marks 7]

Ans. :- This is more a web interface for a development environment that offers a one stop facility for design, development and deployment Java and Python-based applications in Java, Go and Python.

- Google offers the same reliability, availability and scalability at par with Google's own applications.
- Interface is software programming based.
- Comprehensive programming platform irrespective of the size (small or large)

- Signature features : Templates and appspot, excellent monitoring and management console.
- The Google App Engine environment includes the following features

- Dynamic web serving, with full support for common web technologies.
- Persistent storage with queries, sorting, and transactions.
- Automatic scaling and load balancing.
- APIs for authenticating users and sending email using Google accounts
- A fully featured local development environment that simulates Google App Engine on your computer.

Q.25 Write brief notes on Hadoop and Aneka

[SPPU : June-15, Marks 7]

Ans. : Hadoop : Refer Q.26.

Aneka :

- Aneka is a software platform and a framework for developing distributed applications on the cloud.
- Aneka is a market oriented Cloud development and management platform with rapid application development and workload distribution capabilities.
- Fig. Q.25.1 shows Aneka framework architecture.
- The container is the building block of the middleware and represents the runtime environment for executing applications; it contains the core functionalities of the system and is built up from an extensible collection of services that allow administrators to customize the Aneka cloud.

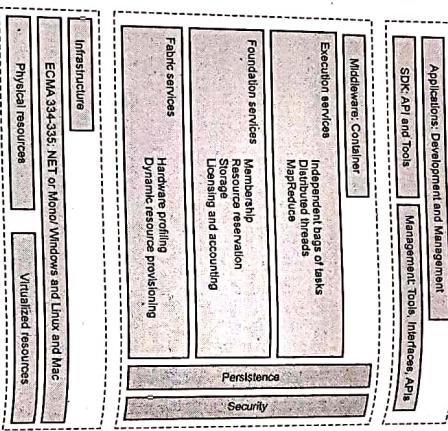


Fig. Q.25.1 : Aneka framework architecture

- There are three classes of services that characterize the container:
- Execution Services: They are responsible for scheduling and executing applications. Each of the programming models supported by Aneka defines specialized implementations of these services for managing the execution of a unit of work defined in the model.
- Foundation Services. These are the core management services of the Aneka container. They are in charge of metering applications, allocating resources for execution, managing the collection of available nodes, and keeping the services registry updated.
- Fabric Services : They constitute the lowest level of the services stack of Aneka and provide access to the resources managed by the cloud

- HDFS is Distributed File System (HDFS) and the MapReduce engine.
- HDFS is designed to run on commodity hardware, which is highly fault-tolerant and scalable. Fig. Q.26.1 shows HDFS architecture.
- Hadoop Distributed File System is a block-structured file system where each file is divided into blocks of a pre-determined size. These blocks are stored across a cluster of one or several machines.
- Apache Hadoop HDFS Architecture follows a Master/Slave Architecture, where a cluster

- Applications managed by the Aneka container can be dynamically mapped to heterogeneous resources, which can grow or shrink according to the application's needs. This elasticity is achieved by means of the resource provisioning framework, which is composed primarily of services built into the Aneka fabric layer.

Ans. :- Hadoop is an open-source software framework that supports data-intensive distributed processing of large datasets in real-time applications. Hadoop provides the basic platform for big data processing. The hadoop architecture have mainly two parts: Hadoop Distributed File System (HDFS) and the MapReduce engine.

- HDFS is Distributed File System (HDFS) architecture.
- DataNodes are the slave nodes in HDFS. Unlike NameNode, DataNode is a commodity hardware, that is, a non-expensive system which is not of high quality or high-availability. The DataNode is a block server that stores the data in the local file ext3 or ext4.
- Journal is the modification log of image, which is available in local hosts native file system. Journal is updated for every client transaction.
- Checkpoint is persistent record of the image, which is also stored on local hosts native file system to enable recovery. NameNode is not allowed to update or modify Checkpoint file.
- Administrator or Checkpoint Node can demand to create new checkpoint file on startup, or restart.

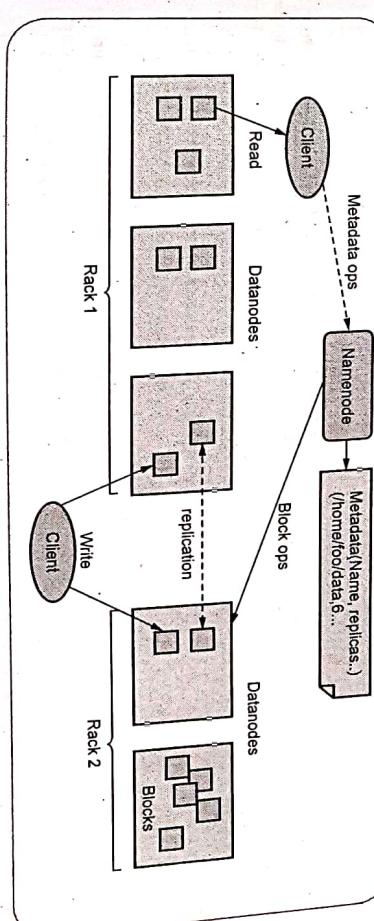


Fig. Q.26.1 Hadoop architecture

- Q.27 Write short notes on the following:**
- Hadoop
 - Microsoft Azure

[SPU : June-16, Marks 7]

- i) Microsoft Azure :** Windows Azure is Microsoft's application platform for the public Cloud. Applications can be deployed on to Azure in various models.

- Windows Azure is used to :

- Build a web application that runs and stores its data in Microsoft datacenters.
 - Store data while the applications that consume this data run on premise (outside the public Cloud).
 - Create virtual machines to develop and test, or run SharePoint and other out-of-the-box applications.
 - Develop massively scalable applications with many users.
 - Offer a wide range of services.
- Microsoft Azure is a cloud computing service created by Microsoft for building, testing, deploying, and managing applications and services through a global network of Microsoft-managed data centers.

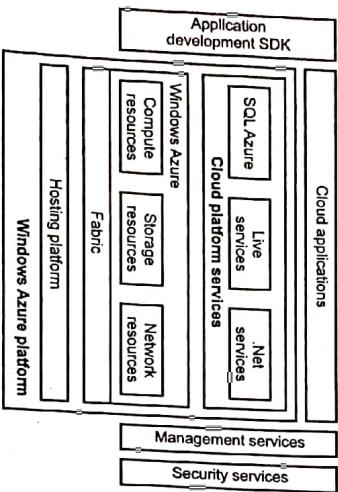


Fig. Q.27.1 Windows Azure platform architecture

- Fig. Q.27.1 shows Windows Azure platform architecture.
- Windows Azure provides resources and services for consumers. For example, hardware is abstracted and exposed as compute resources.
- Physical storage is abstracted as storage resources and exposed through very well defined interfaces. A common Windows Fabric abstracts the hardware and the software and exposes virtual compute and storage resources.
- Each instance of an application is automatically managed and monitored for availability and scalability.
- If an application goes down, the Fabric is notified and a new instance of the application is created.
- Because virtualization is a key element in cloud computing, no assumption must be made on the state of the underlying hardware hosting the application.

- Azure has three components : compute, storage and fabric

- Q.28 Write short note on "Hadoop"**

[SPU : Dec-17, marks 7]

Ans. : Refer Q.26.

- Q.29 Explain Aneka Integration of private and public cloud in details.**

[SPU : Dec-17, marks 7]

- Ans. : Applications managed by the Aneka container can be dynamically mapped to heterogeneous resources, which can grow or shrink according to the application's needs.

- This elasticity is achieved by means of the resource provisioning framework, which is composed primarily of services built into the Aneka fabric layer.
- Fig. Q.29.1 shows an overview of Aneka resource provisioning over private and public clouds.

- This is a typical scenario that a medium or large enterprise may encounter; it combines privately owned resources with public rented resources to dynamically increase the resource capacity to a larger scale.
- Reserved resources are provisioned in advance by paying a low, one-time fee and mostly suited for

- Private resources identify computing and storage elements kept in the premises that share similar internal security and administrative policies.
- Aneka identifies two types of private resources :

- Static resources are constituted by existing physical workstations and servers that may be idle for a certain period of time. Their membership to the Aneka cloud is manually configured by administrators and does not change over time.
- Dynamic resources are mostly represented by virtual instances that join and leave the Aneka cloud and are controlled by resource pool managers that provision and release them when needed.

- Q.30 List the key advantages of Aneka over other GRID or Cluster based workload distribution.**

Ans. : Key advantages are as follows :

- Rapid deployment tools and framework.
- Ability to harness multiple virtual and/or physical machines for accelerating application result.
- Provisioning based on QoS/SLA.
- Support of multiple programming and application environments.

- Public resources reside outside the boundaries of the enterprise and are provisioned by establishing a service-level agreement with the external provider.
- Even in this case we can identify two classes: on-demand and reserved resources.

- Built on-top of Microsoft .NET framework, with support for Linux environments through Mono.

- Q.31 What is Google App Engine ?**

- Ans. : Google App Engine is a way to write your own web applications and have them hosted on Google servers. It enables developers to build their web applications on the same scalable system that power Google applications.

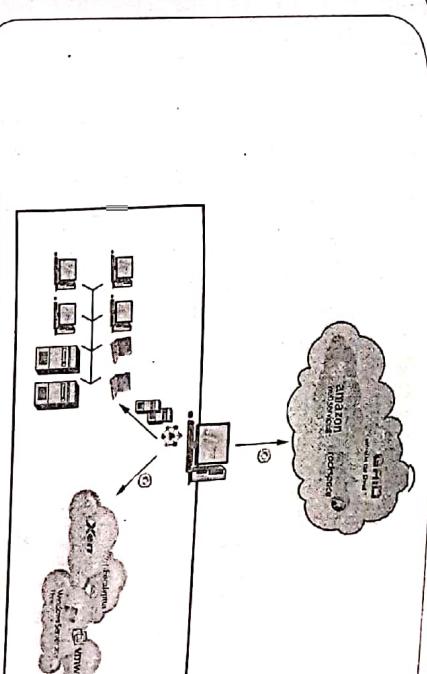


Fig. Q.29.1

long-term usage. These resources are actually the same as static resources, and no automation is needed in the resource provisioning service to manage them.

Ans. : Key advantages are as follows :

- Simultaneous support of multiple run-time environments.
- Built on-top of Microsoft .NET framework, with support for Linux environments through Mono.

- Q.31 What is Google App Engine ?**

- Ans. : Google App Engine is a way to write your own web applications and have them hosted on Google servers. It enables developers to build their web applications on the same scalable system that power Google applications.

- An app is a piece of software which can run on the computer, internet, phone or any other electronic device. Google refers to their online services as Apps. They also sell a specific suite of services known as Google Apps.
- Google's providing both SaaS and PaaS solutions in cloud computing. Some of the example for SaaS solutions including Google Apps which including Gmail, Doc, etc., and PaaS includes Google App Engine.

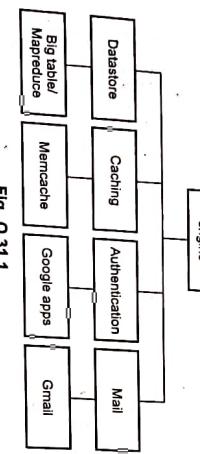


Fig. Q.31.1

- The APP Engine offers a number of services that enable you to perform several common operations when managing your application. The following APIs are available to access these services :

- Mail : Using the mail API, the developers can send email messages.
- Memcache : The Memcache service gives the users the benefit of working efficiently by providing high retrieval speed, even when multiple users access the same application at the same instance of time.
- Image Manipulation : The Image service allows you to manipulate images of your application. With the use of this API, you can resize, crop, rotate and flip images in JPEG and PNG formats.
- In the PaaS space Google is a key player. APP Engine is a platform to create, store and run applications on Google's servers using development languages as java and python.
- App Engine includes tools for managing the data store, monitoring the site and its resource consumption, and debugging and logging. A user can serve the app from his own domain name using Google Apps.

- Q.32 Explain block placement in Hadoop? Why do required block placement?**
- Ans. • For a large cluster, it may not be practical to connect all nodes in a flat topology. A common practice is to spread the nodes across multiple racks.

- Nodes of a rack share a switch, and rack switches are connected by one or more core switches. Communication between two nodes in different racks has to go through multiple switches.
- In most cases, network bandwidth between nodes in the same rack is greater than network bandwidth between nodes in different racks.

- Fig. Q.32.1 shows cluster with two racks, each of which contains three nodes.

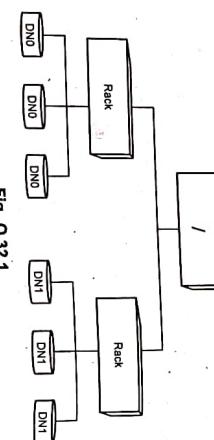


Fig. Q.32.1

1. Mail : Using the mail API, the developers can send email messages.

2. Memcache : The Memcache service gives the users the benefit of working efficiently by providing high retrieval speed, even when multiple users access the same application at the same instance of time.

3. Image Manipulation : The Image service allows you to manipulate images of your application. With the use of this API, you can resize, crop, rotate and flip images in JPEG and PNG formats.

- The default HDFS block placement policy provides a tradeoff between minimizing the write cost, and maximizing read bandwidth.
- Ans. • When a new block is created, HDFS places the first replica on the node where the writer is located. The second and the third replicas are placed on two different nodes in a different rack.

- The choice to place the second and third replicas on a different rack better distributes the block replicas for a single file across the cluster. If the first two replicas were placed on the same rack, for any file, at any one node and no more than two replicas are placed in the same rack, if possible.

- After all target nodes are selected, nodes are organized as a pipeline in the order of their proximity to the first replica. Data are pushed to nodes in this order.

- For reading, the NameNode first checks if the client's host is located in the cluster. If yes, block locations are returned to the client in the order of closeness to the reader. The block is read from DataNodes in this preference order.

- Q.33 Explain file system namespace in HDFS.**
- Ans. • HDFS supports a traditional hierarchical file organization in which a user or an application can create directories and store files inside them. The file system namespace hierarchy is similar to most other existing file systems; you can create, rename, relocate, and remove files.

- HDFS also supports third-party file systems such as CloudStore and Amazon Simple Storage Service.

- In the DataNode, the NameNode runs the configured script to decide which rack the node belongs to.
- If no such a script is configured, the NameNode assumes that all the nodes belong to a default single rack.

- placement makes HDFS unique from most other distributed file systems, and is facilitated by a rack-aware replica placement policy that uses network bandwidth efficiently.

- Data organization : One of the main goals of HDFS is to support large files. The size of a typical HDFS block is 64 MB. Therefore, each HDFS file consists of one or more 64 MB blocks. HDFS tries to place different nodes in a different rack.

- HDFS applications need a write-once-read-many access model for files. A file once created, written, and closed need not be changed.

- HDFS has a master/slave architecture. An HDFS cluster consists of a single NameNode, a master server that manages the file system namespace and regulates access to files by clients.

- HDFS supports a traditional hierarchical file organization. A user or an application can create directories and store files inside these directories. Files in HDFS are write-once and have strictly one writer at any time.

- Q.34 What is Hadoop ecosystem ?**
- Ans. • The Hadoop ecosystem refers to the various components of the Apache Hadoop software library, as well as to the accessories and tools provided by the Apache Software Foundation for these types of software projects, and to the ways that they work together. Hadoop is a Java-based framework that is extremely popular for handling and analyzing large sets of data.

- The idea of a Hadoop ecosystem involves the use of different parts of the core Hadoop set such as MapReduce, a framework for handling vast amounts of data, and the Hadoop Distributed File System (HDFS), a sophisticated file-handling system. There is also YARN, a Hadoop resource manager.

- Data replication : HDFS replicates file blocks for fault tolerance. An application can specify the number of replicas of a file at the time it is created, and this number can be changed any time after that. The name node makes all decisions concerning block replication.
- These include Apache Hive, a data analysis tool; Apache Spark, a general engine for processing big data; Apache Pig, a data flow language; HBase, a

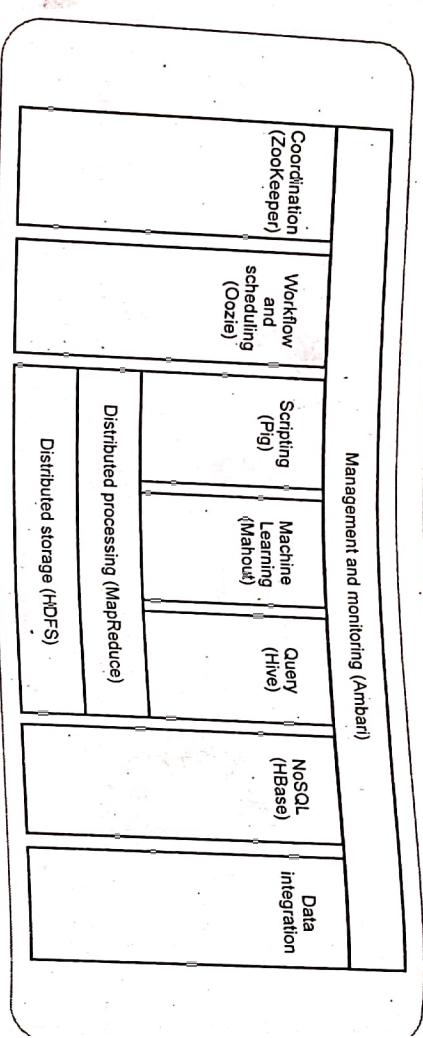


Fig. Q.34.1 : Apache Hadoop ecosystem

database tool; and also Ambari, which can be considered as a Hadoop ecosystem manager, as it helps to administer the use of these various Apache resources together.

- Fig. Q.34.1 shows Apache Hadoop ecosystem.

END.

- Mahout - A library of scalable machine-learning algorithms, implemented on top of Apache Hadoop and using the MapReduce paradigm.

- Hadoop HDFS - Distributed storage layer for Hadoop.
- Yarn Hadoop - Resource management layer introduced in Hadoop 2.x.
- Hadoop Map-Reduce - Parallel processing layer for Hadoop.

- HBase - It is a column-oriented database that runs on top of HDFS. It is a NoSQL database which does not understand the structured query. For sparse data set, it suits well.
- Hive - Apache Hive is a data warehousing infrastructure based on Hadoop and it enables easy data summarization, using SQL queries.
- Pig - It is a top-level scripting language. As we use it with Hadoop, Pig enables writing complex data processing without Java programming.
- Oozie - It is a Java Web application used to schedule Apache Hadoop jobs. It combines multiple jobs sequentially into one logical unit of work.
- Zookeeper - A centralized service for maintaining configuration information, naming, providing distributed synchronization, and providing group services.