

CLLOUD COMPUTING (SUBJECT CODE - P16CS41)

Preface - This course material for Cloud Computing has been prepared in the form of questions and answers to help the students in preparing for semester exams easily. It covers most of the contents from all the five units of the Cloud Computing syllabus of Bharathidasan University, Tiruchirappalli.

Author: Dr. S. Chellammal
Address: Assistant Professor, Department of Computer Science
Bharathidasan University Constituent Arts & Science College
Navalurkuttapattu
Tiruchirappalli-620027

Date uploaded: 26.05.2020

SYLLABUS

Objective:

To provide understanding the concepts & technologies associated with Cloud Computing.

UNIT I FOUNDATIONS: Introduction to Cloud Computing:

Cloud Computing in a Nutshell – Roots of Cloud Computing – Layers and types of Clouds – Desired features of a Cloud – Cloud Infrastructure Management – Challenges and Risks – Migrating into a Cloud: - Introduction – Broad Approaches – The Seven step model – Enriching the ‘Integration as a Services’ Paradigm for the Cloud Era: - Introduction – The Evolution of SaaS – The Challenges of SaaS Paradigm – Approaching the SaaS Integration Enigma – New Integration Scenarios – The Integration Methodologies – SaaS Integration Services – The Enterprise Cloud Computing Paradigm: - Introduction – Background – Issues – Transition Challenges – The Cloud Supply Chain

UNIT II INFRASTRUCTURE AS A SERVICE: Virtual Machine Provisioning and Migration Services

Introduction – Background – Manageability – Migration Services – Management of Virtual Machines for Cloud Infrastructures: - Anatomy of Cloud Infrastructures – Distributed Management of Virtual Infrastructures – Scheduling techniques for Advance Reservation of Capacity – Enhancing Cloud Computing Environments Using a Cluster as a Service: - Introduction – Related Work – RVWS Design – The Logical Design – Secure Distributed Data Storage in Cloud Computing: - Introduction – Cloud Storage from LANs to WANs – Technologies for Data Security – Challenges

UNIT III PLATFORM AND SOFTWARE AS SERVICE (PAAS/IAAS) Aneka Integration of Private and Public Clouds

Introduction– Technologies and Tools – Aneka Cloud Platform - Aneka Resource Provisioning Service – Hybrid Cloud Implementation – CometCloud: An Autonomic Cloud Engine: - Introduction – CometCloud – Architecture – Autonomic Behavior of CometCloud – Overview of CometCloud-based Applications – Implementation and Evaluation

UNIT IV PLATFORM AND SOFTWARE AS SERVICE (PAAS/IAAS) TSystems Cloudbased Solutions for Business Applications:

Introduction – Enterprise Demand of Cloud Computing – Dynamic ICT Service – Importance of Quality and Security in Clouds – Dynamic Data Centre Producing Business-ready; Dynamic ICT Services – The MapReduce Programming Model and Implementations: - Introduction – MapReduce Programming Model – MapReduce implementations for the Cloud

UNIT V MONITORING AND MANAGEMENT: An Architecture for Federated Cloud Computing

Introduction – A typical Use case – The Basic Principles of Cloud Computing – A Federated Cloud Computing Model – Security Considerations – Service Providers Perspective of SLA Management in Cloud Computing: - Traditional Approaches to SLO Management – Types of SLA – Life Cycle of SLA – SLA Management in Cloud –Automated Policy-based Management – Performance Prediction for HPC on Clouds: - Introduction – Background – Grid and Cloud – Performance related issues of HPC in the Cloud

Text Book: Rajkumar Buyya, James Broberg, Andrzej Goscinsky, “Cloud Computing Principles and Paradigms”, Wiley India Pvt. Ltd., 2011.

UNIT – I

1. What is cloud computing?

Cloud computing is the delivery of on-demand computing services -- from applications to storage and processing power -- typically over the internet and on a pay-as-you-go basis.

2. Why do we need cloud computing?

Rather than owning their own computing infrastructure or data centers, companies can rent access to anything from applications to storage from a cloud service provider. One benefit of using cloud computing services is that firms can avoid the upfront cost and complexity of owning and maintaining their own IT infrastructure, and instead simply pay for what they use, when they use it.

3. List out different service models or delivery models of cloud computing.

- Infrastructure-as-a-Service (IaaS)
- Platform-as-a-Service (PaaS)
- Software-as-a-Service (SaaS)

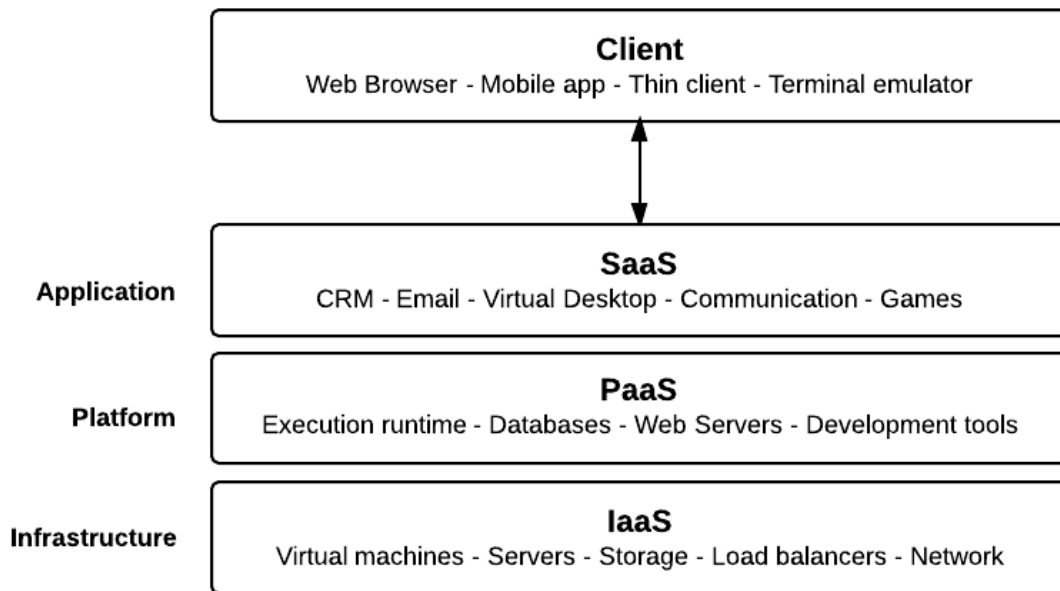
4. Explain different service models of cloud computing

Infrastructure-as-a-Service (IaaS) - refers to the fundamental building blocks of computing that can be rented: physical or virtual servers, storage and networking. This is attractive to companies that want to build applications from the very ground up and want to control nearly all the elements themselves, but it does require firms to have the technical skills to be able to orchestrate services at that level. Research by Oracle found that two thirds of IaaS users said using online infrastructure makes it easier to innovate, had cut their time to deploy new applications and services and had significantly cut on-going maintenance costs.

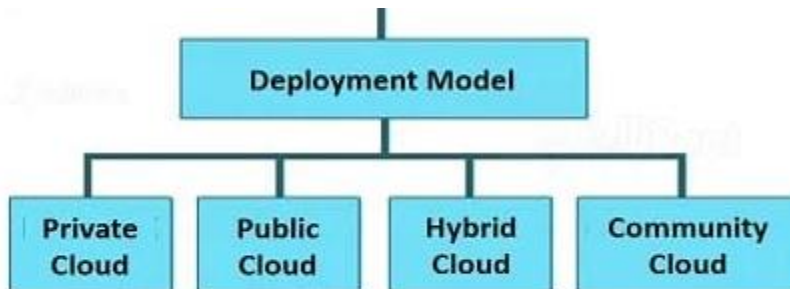
Platform-as-a-Service (PaaS) is the next layer up -- as well as the underlying storage, networking, and virtual servers this will also include the tools and software that developers need to build applications on top of: that could include middleware, database management, operating systems, and development tools.

Software-as-a-Service (SaaS) is the delivery of applications-as-a-service, probably the version of cloud computing that most people are used to on a day-to-day basis. The underlying hardware and operating system is irrelevant to the end user, who will access the service via a web browser or app; it is often bought on a per-seat or per-user basis.

Diagram to show different service models



5. Explain different deployment models of cloud computing



Public Cloud

This type of cloud services is provided on a network for public use. Customers have no control over the location of the infrastructure. It is based on a shared cost model for all the users, or in the form of a licensing policy such as pay per user. Public deployment models in the cloud are perfect for organizations with growing and fluctuating demands. It is also popular among businesses of all sizes for their web applications, webmail, and storage of non-sensitive data.

Private Cloud

It is a cloud-based infrastructure used by stand-alone organizations. It offers greater control over security. The data is backed up by a firewall and internally, and can be hosted internally or externally. Private clouds are perfect for organizations that have high-security requirements, high management demands, and availability requirements.

Hybrid Cloud

This model incorporates the best of both private and public clouds, but each can remain as separate entities. Further, as part of this deployment of cloud computing model, the internal, or external providers can provide resources. A hybrid cloud is ideal for scalability, flexibility, and security. A perfect example of this scenario would be that of an organization who uses the private cloud to secure their data and interacts with its customers using the public cloud.

Community Cloud

It is a mutually shared model between organizations that belong to a particular community such as banks, government organizations, or commercial enterprises. Community members generally share similar issues of privacy, performance, and security. This type of deployment model of cloud computing is managed and hosted internally or by a third-party vendor.

6. List out features of cloud computing

Following are the characteristics/features of Cloud Computing:

- Great Availability of Resources
- On-demand Self-service
- Easy Maintenance
- Large Network Access
- Availability
- Automatic System
- Economical
- Security
- Pay as you go

Great Availability of Resources -

This service is made to serve multiple customers and this is done with the help of the multi-tenant model. There are many physical and virtual resources provided which can modify as per the customer's demand.

On-Demand Self-Service -

It is one of the important and valuable features of cloud computing as the user can continuously monitor the server uptime, capabilities, and allotted network storage. With this feature, the user can also monitor the computing capabilities.

Easy Maintenance

The servers are easily maintained and the downtime is very low; in some cases, there is no downtime. The cloud computing comes up with an update every time by gradually making it better. The updates are more compatible with the devices and perform faster than older ones along with the bugs which are fixed.

Large Network Access

The user can access the data of the cloud or upload the data to the cloud from anywhere just with the help of a device and an internet connection. These capabilities are available all over the network and accessed with the help of internet.

Availability

The capabilities of the cloud can be modified as needed and can extend a lot. It analyzes the storage usage and allows the user to buy extra storage if needed for a very small amount. This service is available anytime and can be accessed from anywhere.

Automation

Cloud computing automatically analyzes the data needed and supports a metering capability at some level of services. This usage can monitor, control, and report, providing transparency for the host as well as the customer.

Economical

It is the one-time investment as the company (host) has to buy the storage and a small part of it can provide to the many companies which save the host from monthly or yearly costs only the amount which spends on the basic maintenance and few more expenses.

Security

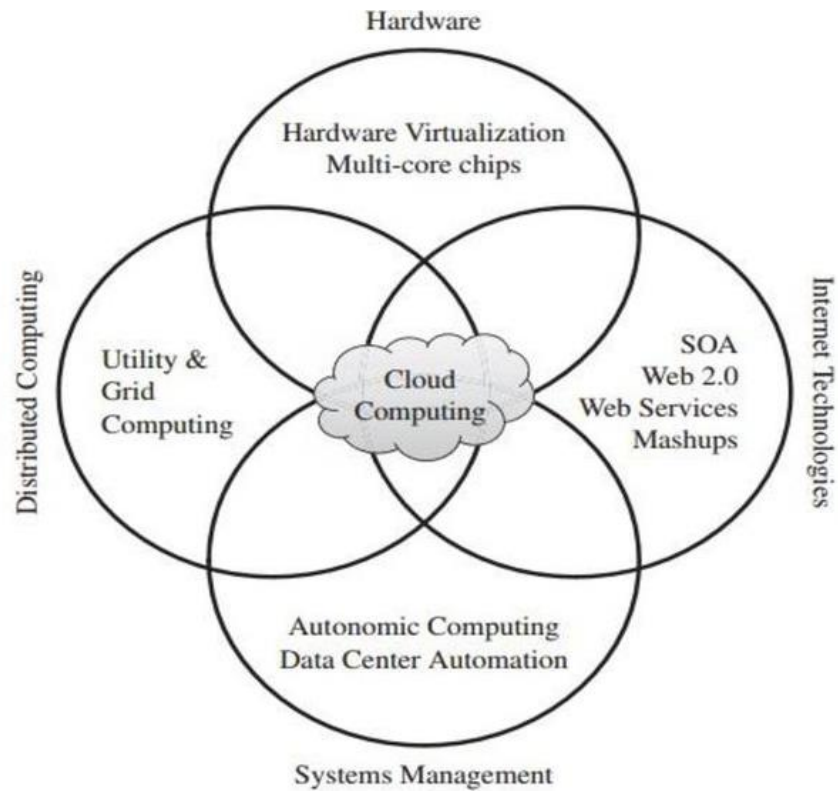
This is one of the best features of cloud computing. It creates a snapshot of the data stored so that the data may not get lost even if one of the server damages. The data stores within the storage devices which cannot hack and utilize by any other person. The storage service is quick and reliable which can access from anywhere just with the help of a device and internet connection.

Pay-as-You-Go

In cloud computing, the user has to pay only for the service or the space they have utilized. There is no hidden or extra charge which is to be paid. The service is economical and most of the time some space allows for free.

7. Explain the basic roots of cloud computing. (or) Briefly describe key enablers for cloud computing .

- (i) Mainframe to cloud
- ii) SOA, Web Services, Web 2.0 and Mashups
- (iii) Grid Computing
- (iv) Utility Computing
- (v)Hardware Virtualization
- (vi)Virtual Appliance and OVF
- (vii) Autonomic Computing



(i) From mainframe to cloud

Currently experiencing a switch in the IT world, from in-house generated computing power into utility-supplied computing resources delivered over the Internet as Web services. Computing delivered as a utility can be defined as “on demand delivery of infrastructure, applications, and business processes in a security-rich, shared, scalability based computer environment over the Internet for a fee”

(ii) SOA, Web Services, Web 2.0 and Mashups

Web services can glue together applications running on different messaging product platforms, enabling information from one application to be made available to others, and enabling internal applications to be made available over the Internet.

(iii) Grid Computing

Grid computing is the collection of computer resources from multiple locations to reach a common goal. The grid can be thought of as a distributed system with non-interactive workloads that involve a large number of files.

(iv) Utility Computing

Utility computing is a service provisioning model in which a service provider makes computing resources and infrastructure management available to the customer as needed, and charges them for specific usage rather than a flat rate.

(v) Hardware Virtualization

Hardware virtualization allows running multiple operating systems and software stacks on a single physical platform. 3 basic capabilities related to management of workload: isolation, Consolidation and Migration

References

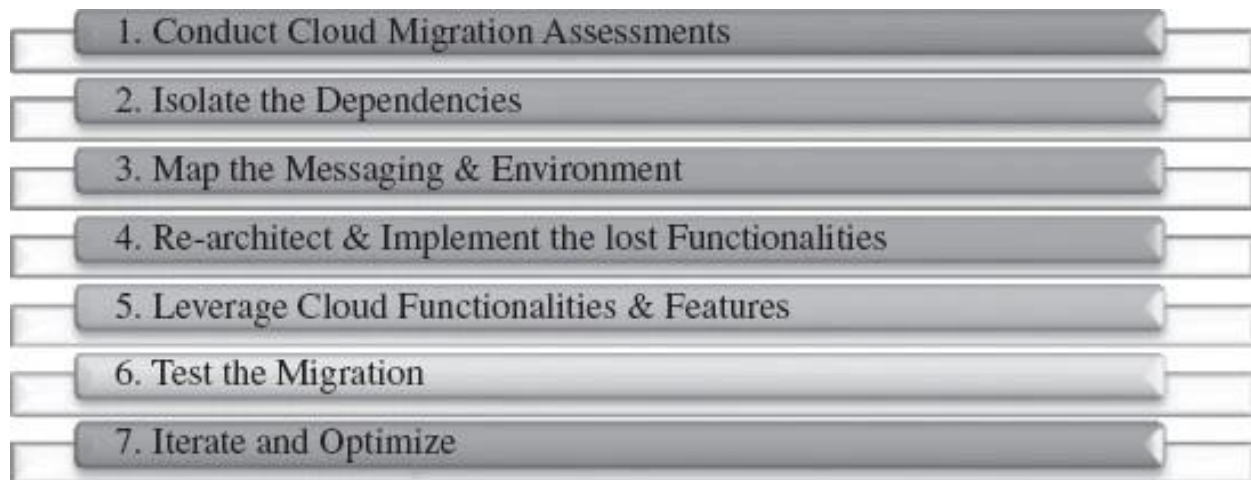
(vi) Virtual Appliance and OVF(open virtual format)

An application combined with the environment needed to run it (operating system, libraries, compilers, databases, application containers, and so forth) is referred to as a “virtual appliance.”

(vii) Autonomic Computing

The increasing complexity of computing systems has motivated research on autonomic computing, which seeks to improve systems by decreasing human involvement in their operation

8. Explain seven step model of migration into a cloud .



Step – 1 Cloud migration assessments comprise assessments to understand the issues involved in the specific case of migration at the application level or the code, the design, the architecture, or usage levels. These assessments are about the cost of migration as well as about the ROI that can be achieved in the case of production version.

Step 2- isolating all systemic and environmental dependencies of the enterprise application components within the captive data center

Step 3- generating the mapping constructs between what shall possibly remain in the local captive data center and what goes onto the cloud.




Step -4 substantial part of the enterprise application needs to be rearchitected, redesigned, and re implemented on the cloud.

Step 5- We leverage the intrinsic features of the cloud computing service to augment our enterprise application in its own small ways.

Step 6 - we validate and test the new form of the enterprise application with an extensive test suite that comprises testing the components of the enterprise application on the cloud as well

Step 7-Test results could be positive or mixed.In the latter case, we iterate and optimize as appropriate. After several such optimizing iterations, the migration is deemed successful

9. Explain layers and types of cloud computing in detail.

Service Class	Main Access & Management Tool	Service content
 SaaS	Web Browser	Cloud Applications Social networks, Office suites, CRM, Video processing
 PaaS	Cloud Development Environment	Cloud Platform Programming languages, Frameworks, Mashups editors, Structured data
 IaaS	Virtual Infrastructure Manager	Cloud Infrastructure Compute Servers, Data Storage, Firewall, Load Balancer

Layers and Types of Cloud Computing

Cloud computing services are divided into three classes, according to the abstraction level of the capability provided and the service model of providers, namely:

- (1) Infrastructure as a Service (IaaS),
- (2) Platform as a Service (PaaS),
- (3) Software as a Service (SaaS).

The reference model explains the role of each layer in an integrated architecture. A core middleware manages physical resources and the VMs deployed on top of them; in addition, it provides the required features (e.g., accounting and billing) to offer multi-tenant pay-as-you-go services. Cloud development environments are built on top of infrastructure services to offer application development and deployment capabilities. In this level, various programming models, libraries, APIs, and mashup editors enable the creation of a range of business, Web, and scientific applications. Once deployed in the cloud, these applications can be consumed by end users.

Infrastructure as a Service Offering virtualized resources (computation, storage, and communication) on demand is known as Infrastructure as a Service (IaaS). Infrastructure services are considered to be the bottom layer of cloud computing systems. Amazon web services offers IaaS, which in the case of EC2 service means offering VMs with a software stack that can be customized similar to how an ordinary physical server would be customized. Users are given privileges to perform numerous activities to the server, such as: starting and stopping it, customizing it by installing software packages, attaching virtual disks to it, and configuring access permissions and firewall rules

Platform as a Service - It offers a higher level of abstraction to make a cloud easily programmable, known as Platform as a Service (PaaS). It provides Integrated Development Environment (IDE) including data security, backup and recovery, application hosting, and scalable architecture. A Cloud platform offers an environment on which developers create and deploy applications and do not necessarily need to know how many processors or how much memory that applications will be using.

Software as a Service - Applications reside on the top of the cloud stack. Services in this layer are accessed through Web Portals. Consumers are increasingly shifting from locally installed computer systems to online software services that offer the same functionality.

Three deployment types of cloud computing (a) public cloud, (b) private cloud, and (c) hybrid cloud public cloud

Public Cloud

This type of cloud services is provided on a network for public use. Customers have no control over the location of the infrastructure. It is based on a shared cost model for all the users, or in the form of a licensing policy such as pay per user. Public deployment models in the cloud are perfect for organizations with growing and fluctuating demands. It is also popular among businesses of all sizes for their web applications, webmail, and storage of non-sensitive data.

Private Cloud

It is a cloud-based infrastructure used by stand-alone organizations. It offers greater control over security. The data is backed up by a firewall and internally, and can be hosted internally or externally. Private clouds are perfect for organizations that have high-security requirements, high management demands, and availability requirements.

Hybrid Cloud

This model incorporates the best of both private and public clouds, but each can remain as separate entities. Further, as part of this deployment of cloud computing model, the internal, or external providers can provide resources. A hybrid cloud is ideal for scalability, flexibility, and security. A perfect example of this scenario would be that of an organization who uses the private cloud to secure their data and interacts with its customers using the public cloud.

Community Cloud

It is a mutually shared model between organizations that belong to a particular community such as banks, government organizations, or commercial enterprises. Community members generally share similar issues of privacy, performance, and security. This type of deployment model of cloud computing is managed and hosted internally or by a third-party vendor.

10. List out benefits of cloud computing

- Lower Computational Costs
- Improved Performance
- Reduced Software Costs
- Instant Software updates
- Unlimited storage capacity
- Increased Data Reliability
- Universal Document Access
- Latest version availability
- Easier Group Collaboration/ Sharing
- Device Independence
- Lower computer costs
- Improved performance:
- Reduced software costs:
- Unlimited storage capacity- Cloud computing offers virtually limitless storage.
- Increased data reliability:
- Universal document access:
- Latest version availability:
- Easier group collaboration:
- Device independence.

11. List out the disadvantages of cloud computing

- Requires constant Internet Connection
- Does not work well with low speed connection
- Stored data might not be Secured
- Stored data can be lost
- Features might be limited
- Requires a constant Internet connection:

- Cloud computing is impossible if you cannot connect to the Internet
- Does not work well with low-speed connections:
- Similarly, a low-speed Internet connection, such as that found with dial-up services, makes cloud computing painful at best and often impossible.
- Stored data might not be secure:
- With cloud computing, all your data is stored on the cloud. The questions is How secure is the cloud?
- Can unauthorized users gain access to your confidential data?
- Stored data can be lost

12. What is the Security management in terms of Cloud Computing?

- **Identity management**- It provides access to the authorization of application services.
- **Access control permission**- It provides users to have complete controlling access of another user too who is entering into the same cloud environment.
- **Authentication and Authorization**- It provides access to only the authorized and authenticated personnel to securely access the data and applications.

13. Mention what is the difference between elasticity and scalability in cloud computing?

Scalability is a typical characteristic of cloud computing which is used to handle the escalating workload by escalating in proportion to the amount of resource capacity. By the use of scalability, the architecture provides resources on requirement BA is resources as and when the requirement is being raised by the traffic. On the other hand, Elasticity is a characteristic that provides for the concept of commissioning and decommissioning of the huge amount of resource capacity dynamically. It is usually measured by the speed by which the resources are coming on demand and the usage of those resources.

14. Before cloud migration what are the essential things to be taken in concern by users?

- Compliance
- Loss of data
- Data storage
- Business continuity
- Uptime
- Data integrity in cloud computing

15. Compare public, private and hybrid cloud storage

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

16. What is cloud migration?

Cloud migration is the process of moving data, applications or other business elements to a cloud computing environment.

17. What are the main benefits of migrating to the cloud?’

- **Scalability:** Cloud computing can scale up to support larger workloads and greater numbers of users far more easily than on-premises infrastructure, which requires companies to purchase and set up additional physical servers, networking equipment, or software licenses.
- **Cost:** Companies that move to the cloud often vastly reduce the amount they spend on IT operations, since the cloud providers handle maintenance and upgrades. Instead of keeping things up and running, companies can focus more resources on their biggest business needs – developing new products or improving existing ones.

- **Performance:** For some businesses, moving to the cloud can enable them to improve performance and the overall user experience for their customers. If their application or website is hosted in cloud data centers instead of in various on-premises servers, then data will not have to travel as far to reach the users, reducing latency.
- **Flexibility:** Users, whether they're employees or customers, can access the cloud services and data they need from anywhere. This makes it easier for a business to expand into new territories, offer their services to international audiences, and let their employees work flexibly.

18. List out the challenges associated with cloud migration

- **Migrating large databases** - Often, databases will need to move to a different platform altogether in order to function in the cloud. Moving a database is difficult, especially if there are large amounts of data involved. Some cloud providers actually offer physical data transfer methods, such as loading data onto a hardware appliance and then shipping the appliance to the cloud provider, for massive databases that would take too long to transfer via the Internet. Data can also be transferred over the Internet. Regardless of the method, data migration often takes significant time
- **Data integrity:** After data is transferred, the next step is making sure data is intact and secure, and is not leaked during the process.
- **Continued operation:** A business needs to ensure that its current systems remain operational and available throughout the migration. They will need to have some overlap between on-premises and cloud to ensure continuous service; for instance, it's necessary to make a copy of all data in the cloud before shutting down an existing database. Businesses typically need to move a little bit at a time instead of all at once.

19. List out different migration strategies or migration techniques

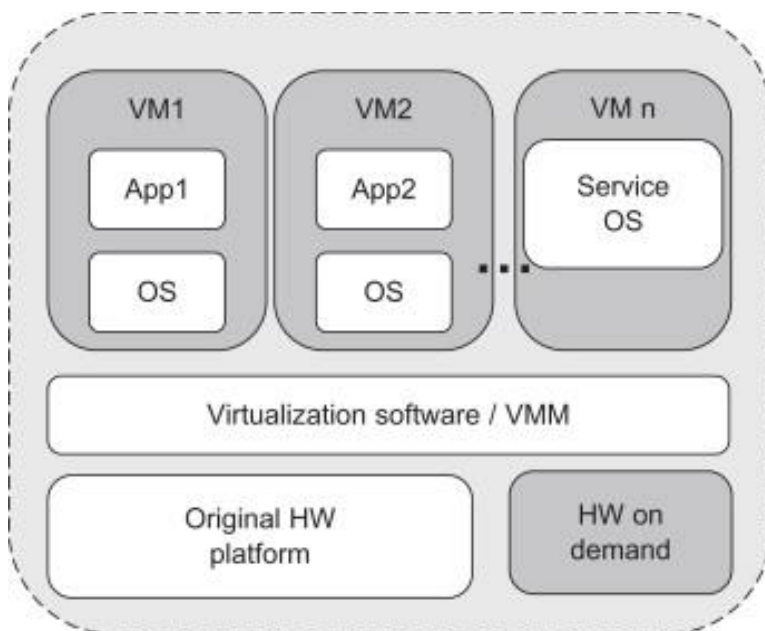
- **Rehost** - Rehosting can be thought of as "the same thing, but on cloud servers". Companies that choose this strategy will select an IaaS (Infrastructure-as-a-Service) provider and recreate their application architecture on that infrastructure.
- **Refactor** - Companies that choose to refactor will reuse already existing code and frameworks, but run their applications on a PaaS (Platform-as-a-Service) provider's platform – instead of on IaaS, as in rehosting.
- **Revise** - This strategy involves partially rewriting or expanding the code base, then deploying it by either rehosting or refactoring (see above).
- **Rebuild** - To "rebuild" means rewriting and re-architecting the application from the ground up on a PaaS provider's platform. This can be a labor intensive process, but it also enables developers to take advantage of modern features from PaaS vendors.
- **Replace** - Businesses can also opt to discard their old applications altogether and switch to already-built SaaS (Software-as-a-Service) applications from third-party vendors.

UNIT II

1. What is the requirement of virtualization platform in implementing cloud?

- The requirement of virtualization platform in implementing cloud is to
- Manage the service level policies
- Cloud Operating System
- Virtualization platforms helps to keep the backend level and user level concepts different from each other

2. Explain virtualization in detail.



- **Virtualization** is the "creation of a virtual (rather than actual) version of something, such as a server, a desktop, a storage device, an operating system or network resources".
- In other words, Virtualization is a technique, which allows to share a single physical instance of a resource or an application among multiple customers and organizations. It does by assigning a logical name to a physical storage and providing a pointer to that physical resource when demanded.
- Creation of a virtual machine over existing operating system and hardware is known as Hardware Virtualization. A Virtual machine provides an environment that is logically separated from the underlying hardware.
- The machine on which the virtual machine is going to create is known as Host Machine and that virtual machine is referred as a Guest Machine

3. Describe different types of virtualization

- Hardware Virtualization.
- Operating system Virtualization.
- Server Virtualization.
- Storage Virtualization.
-

Hardware Virtualization

When the virtual machine software or virtual machine manager (VMM) is directly installed on the hardware system is known as hardware virtualization.

The main job of hypervisor is to control and monitoring the processor, memory and other hardware resources.

After virtualization of hardware system we can install different operating system on it and run different applications on those OS.

Usage

Hardware virtualization is mainly done for the server platforms, because controlling virtual machines is much easier than controlling a physical server.

Operating System Virtualization

When the virtual machine software or virtual machine manager (VMM) is installed on the Host operating system instead of directly on the hardware system is known as operating system virtualization.

Usage

Operating System Virtualization is mainly used for testing the applications on different platforms of OS.

Server Virtualization

When the virtual machine software or virtual machine manager (VMM) is directly installed on the Server system is known as server virtualization.

Usage

Server virtualization is done because a single physical server can be divided into multiple servers on the demand basis and for balancing the load.

Storage Virtualization

Storage virtualization is the process of grouping the physical storage from multiple network storage devices so that it looks like a single storage device.

Storage virtualization is also implemented by using software applications.

Usage

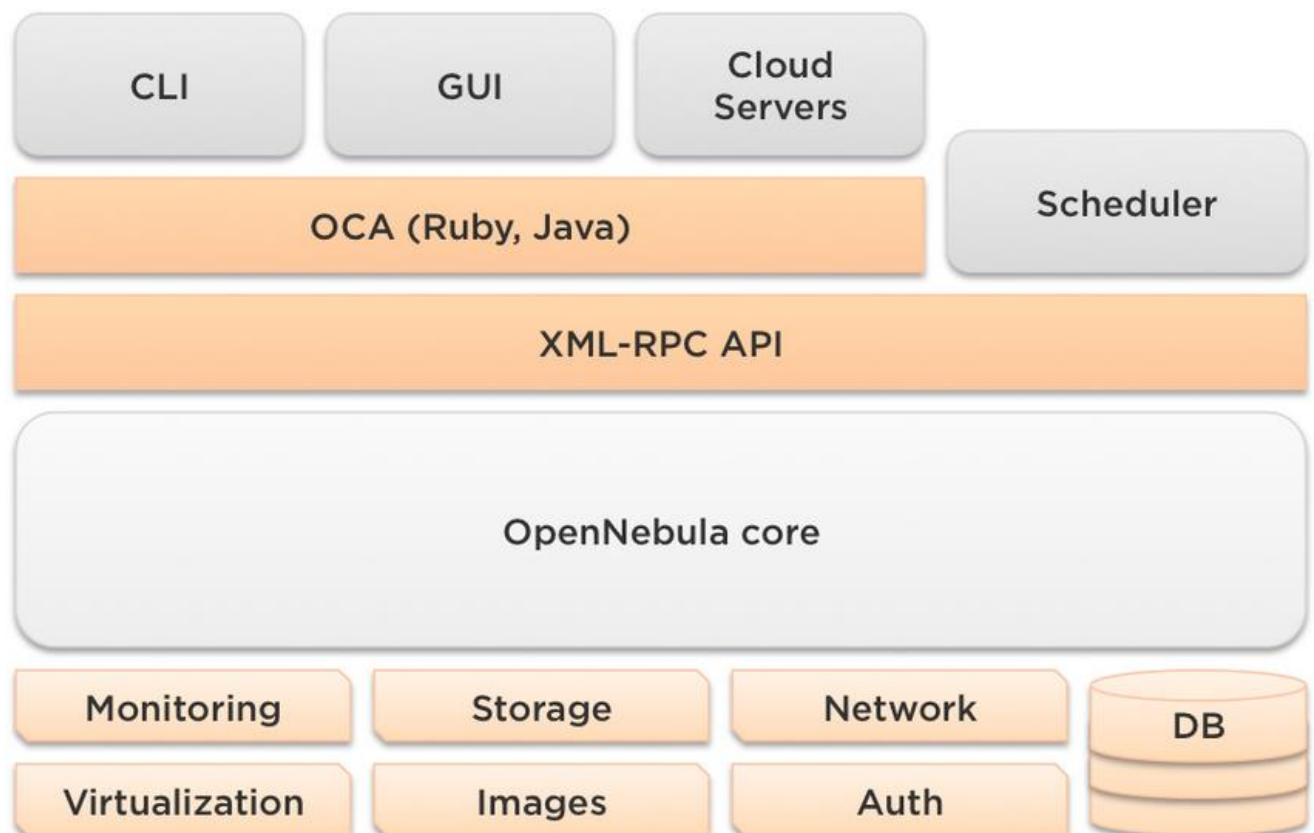
Storage virtualization is mainly done for back-up and recovery purposes.

4. What is OpenNebula?

- OpenNebula is an open-source management platform to build IaaS private, public and hybrid clouds
- OpenNebula is a cloud computing platform for managing heterogeneous distributed data center infrastructures. The OpenNebula platform manages a data center's virtual infrastructure to build private, public and hybrid implementations of infrastructure as a service

Following are the objectives of OpenNebula

- Deploy in a easier way a powerful and customizable virtualized cloud.
- Support to users and developers with options of cloud and users interfaces, which may create an elevated level of customization and components.
- Grant software stable and reliable.
- Improve the data center management tools and the quality of services.
- Avoid errors in the project(poor code).



OpenNebula

5. Explain different scheduling techniques or leasing models for advance reservation of resources.

- Scheduler that can efficiently support advance reservations efficiently by using the suspend/resume/migrate capability of VMs, but minimizing the overhead of using VMs. The fundamental resource provisioning abstraction in Haizea is the lease, with three types of lease currently supported:
- Advanced reservation leases, where the resources must be available at a specific time.
- Best-effort leases, where resources are provisioned as soon as possible and requests are placed on a queue if necessary.
- Immediate leases, where resources are provisioned when requested or not at all.
- The scheduling component of Haizea uses classical backfilling algorithms extended to allow best-effort leases to be preempted if resources have to be freed up for advance reservation requests.
- Best-effort leases are scheduled using a queue. When a best-effort lease is requested, the lease request is placed at the end of the queue, which is periodically evaluated using a backfilling algorithm.
- For advance reservation, the scheduler uses Earlier Deadline First(EDF) algorithm

6. Give an overview above Cluster as a Service(CaaS).

The purpose of the CaaS Technology is to ease the publication, discovery, selection, and use of existing computational clusters. The exposure of a cluster via a Web service is intricate and comprises several services running on top of a physical cluster.

A typical cluster is comprised of three elements: nodes, data storage, and middleware.

The middleware virtualizes the cluster into a single system image.

The components that manage the allocation of jobs to nodes (scheduler) and that monitor the activity of the cluster (monitor).

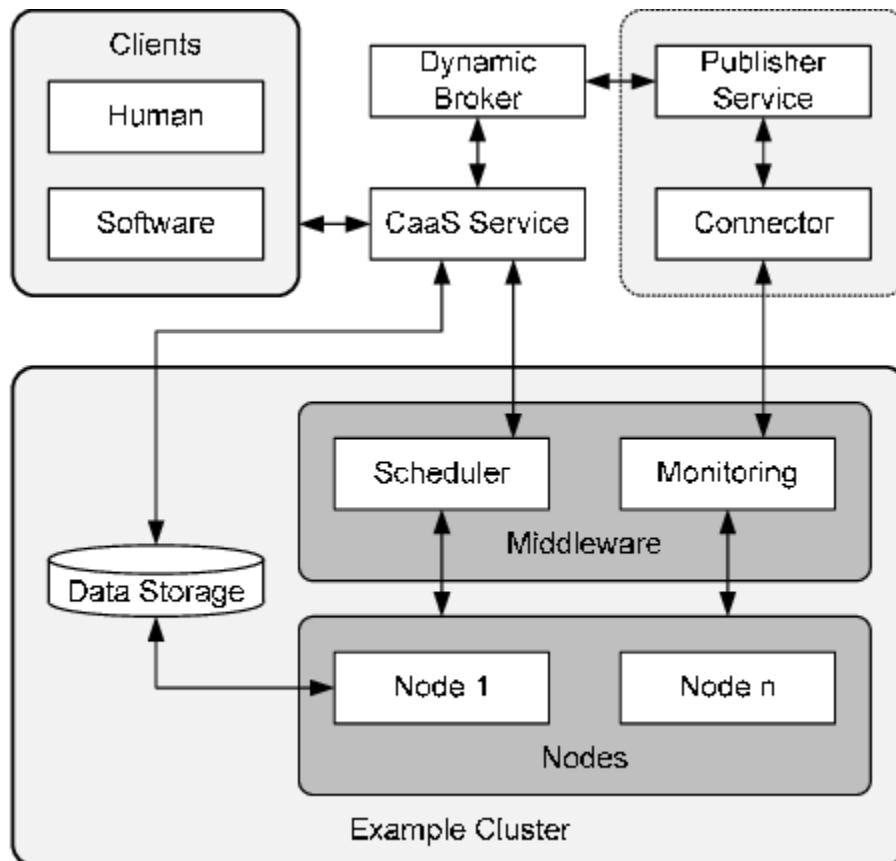
As time progresses, the amount of free memory, disk space, and CPU usage of each cluster node changes. Information about how quickly the scheduler can take a job and start it on the cluster also is vital in choosing a cluster.

To make information about the cluster publishable, a Publisher Web service and Connector were created using the RVWS framework.

The purpose of the publisher Web service was to expose the dynamic attributes of the cluster via the stateful WSDL document.

The Publisher service is published to the Dynamic Broker so clients can easily discover the cluster.

The role of the CaaS Service is to (i) provide easy and intuitive file transfer tools so clients can upload jobs and download results and (ii) offer an easy to use interface for clients to monitor their jobs.



Complete Cluster as a Service

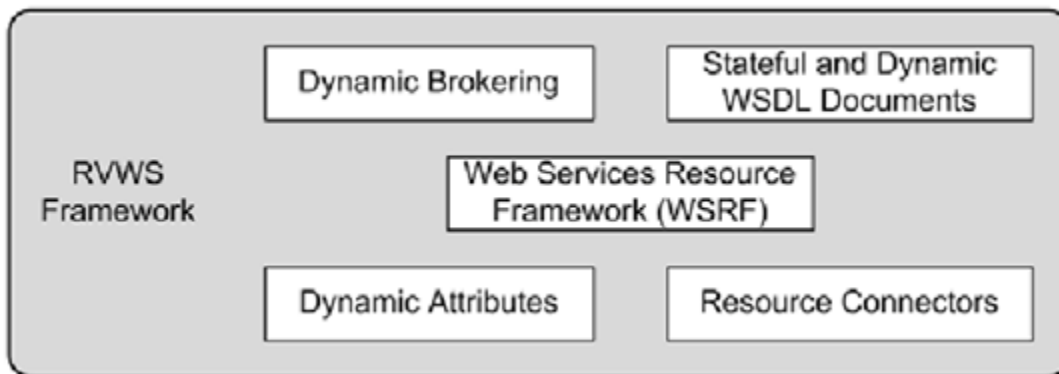
7. What is RVWS framework? Briefly explain RVWS design (or) Explain the elements of RVWS.

RVWS stands for Resources Via web services (RVWS) framework.

The elements of RVWS is as shown in the diagram RVWS framework exposes resources as stateful web services and supports publication using WSDL documents with dynamically changing resource state and characteristics, and engages attributed naming to select required by clients services.

There are two categories of dynamic attributes addressed in RVWS: state and characteristics. State attributes cover the current activity of the service and its resources thus indicating if a given service is ready for client requests. Characteristic attributes cover the operational and physical

limitations of the service, the resources behind it, quality of service (QoS), price and even information about the providers of the services, thus indicating if the service is appropriate for the client



8. List out the advantages of private cloud computing infrastructure.

Private cloud infrastructure is a dedicated infrastructure provided to one single organization or client.

- **Controls:** Better controls for data, users and information assets.
- **Cost:** Initial investment for hardware is very high in case of an on-premise infrastructure.
- **Security:** The cloud belongs to a single client. Hence, the infrastructure and systems can be configured to provide high levels of security.
- **Superior Performance:** Normally private clouds are deployed inside the firewall of the organization's intranet which ensures efficiency and good network performance.
- **Easy Customization:** The hardware and other resources can be customized easily by the company.
- **Compliance:** Compliance is achieved easily in private clouds.

9. What is "EUCALYPTUS" in cloud computing?

EUCALYPTUS is an acronym that stands for Elastic Utility Computing Architecture For Linking Your Program To Useful Systems. It is used to execute clusters in cloud computing platform.

10. Explain what is the use of "EUCALYPTUS" in cloud computing?

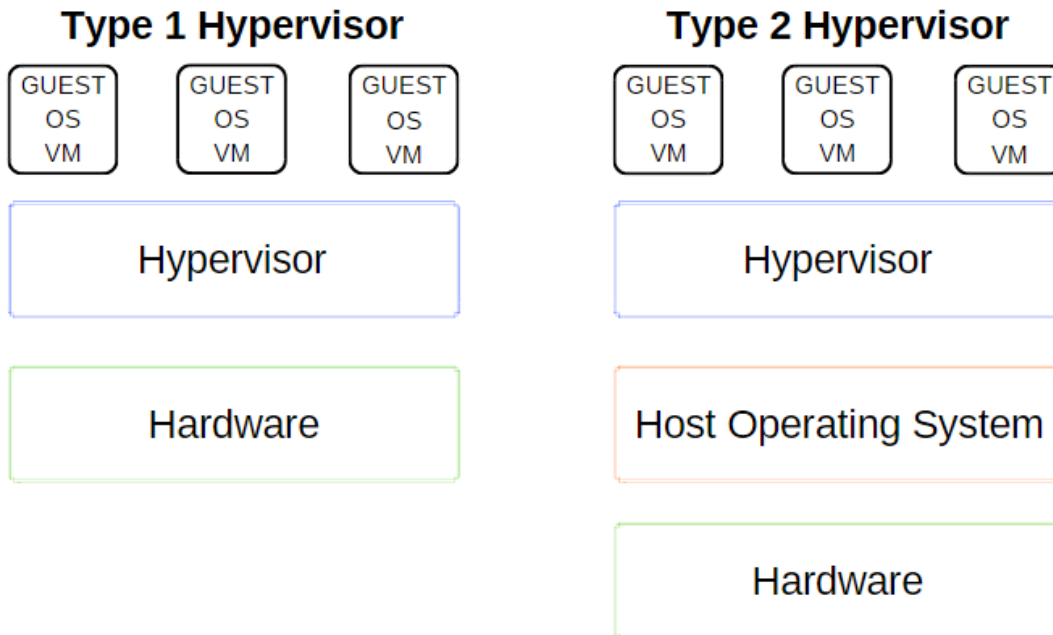
"Eucalyptus" is an open source software infrastructure in cloud computing, which is used to implement clusters in cloud computing platform. It is used to build public, hybrid and private clouds. It has the ability to produce your own data center into a private cloud and allows you to use its functionality to many other organizations.

11. Mention what is Hypervisor in cloud computing and their types?

Hypervisor is a Virtual Machine Monitor which manages resources for virtual machines. There are mainly two types of hypervisors

Type 1: The guest Vm runs directly over the host hardware, eg Xen, VmWare ESXI

Type 2: The guest Vm runs over hardware through a host OS, eg Kvm, oracle virtualbox



12. What is the use of API's in cloud services?

API's (Application Programming Interface) is very useful in cloud platforms

It eliminates the need to write the fully fledged programs

It provides the instructions to make communication between one or more applications

It allows easy creation of applications and link the cloud services with other systems

13. Mention the name of some large cloud providers and databases

- Google bigtable
- Amazon simpleDB
- Cloud based SQL

14. What are the security laws which are implemented to secure data in a cloud ?

The security laws which are implemented to secure data in cloud are

- Processing: Control the data that is being processed correctly and completely in an application
- File: It manages and control the data being manipulated in any of the file
- Output reconciliation: It controls the data which has to be reconciled from input to output
- Input Validation: Control the input data
- Security and Backup: It provides security and backup it also controls the security breaches logs

UNIT III

1. What is Aneka?

Aneka is a .NET-based application development Platform-as-a-Service (PaaS), which offers a runtime environment and a set of APIs that enable developers to build customized applications by using multiple programming models such as Task Programming, Thread Programming and MapReduce Programming, which can leverage the compute resources on either public or private Cloud

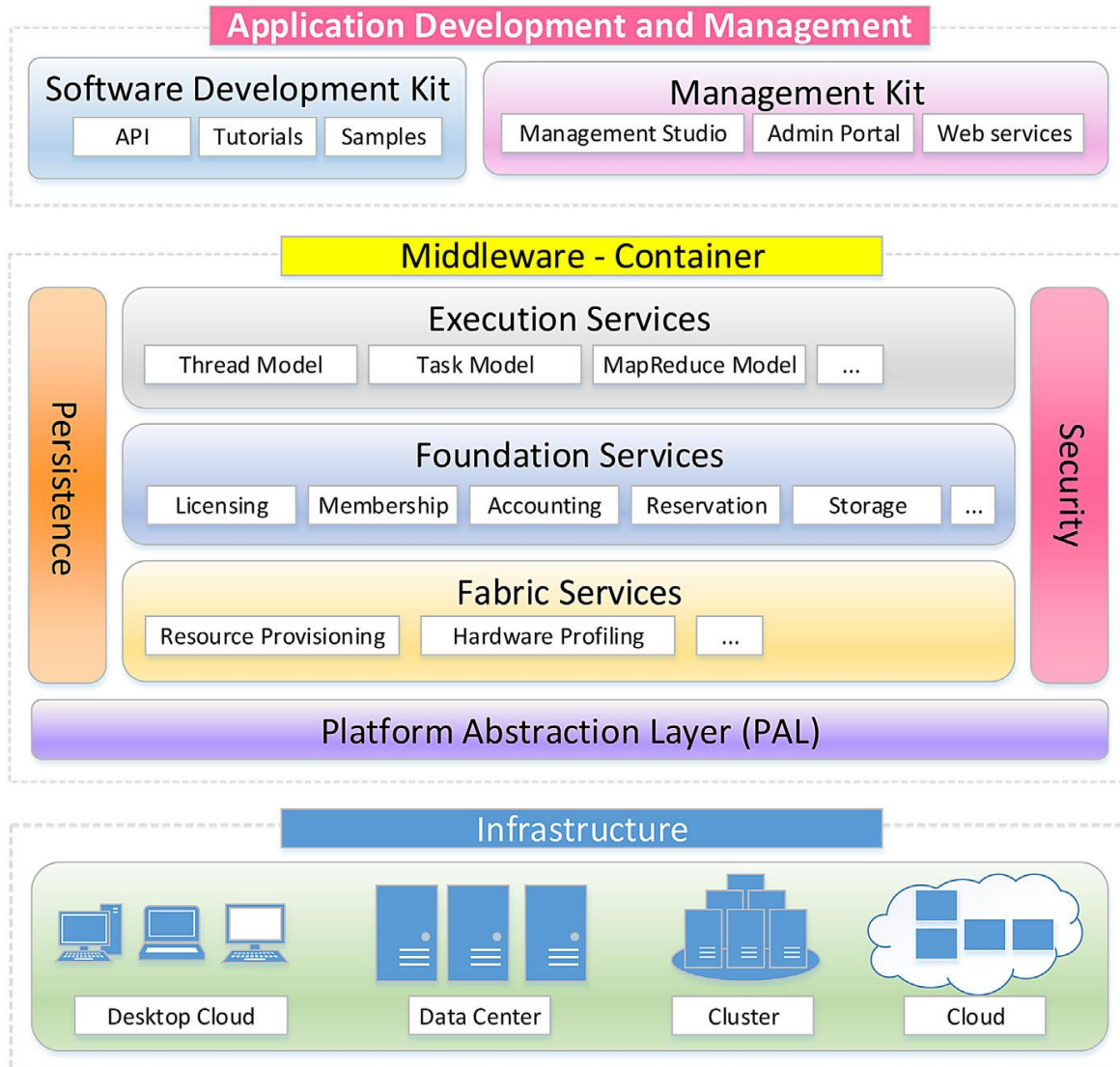
One of key characteristics of Aneka PaaS is to support provisioning of resources on public Clouds such as Windows Azure, Amazon EC2, and GoGrid, while also harnessing private Cloud resources ranging from desktops and clusters, to virtual datacentres when needed to boost the performance of applications.

2. Explain the layered architecture of Aneka platform

The architecture of Aneka cloud platform is shown below.

The above diagram provides a layered view of the Aneka components. Aneka provides a runtime environment for executing applications by leveraging heterogeneous resources on the underlying infrastructure built on the top of computing nodes employed from network of workstations, clusters, grids, and data centers. In other words, the infrastructure layer is a collection of nodes hosting components of Aneka middleware.

The middleware provides a collection of services for interactions with the Aneka cloud. The container represents the unit of deployment of Aneka clouds and the runtime environment for services. The core functionalities residing in the Platform Abstraction Layer (PAL) constitute the basic services that are used to control the infrastructure of Aneka clouds. It provides a uniform interface for management and configuration of nodes and the containers instances deployed on them in the infrastructure layer. Middleware is composed of two major components representing the building blocks of Aneka clouds: the Aneka Daemon and Aneka Container. Each node hosts the Aneka daemon and one or more Aneka container instances. The daemon is a management component controlling the container instances installed on the particular node. A node forms the infrastructure layer running the Aneka master container which plays the role of resource manager and application scheduler. Nodes running Aneka worker containers are responsible for processing and executing work units of the applications. In addition, each container provides a messaging channel for accessing features of different services provided by the container.



There are three classes of services characterizing the container:

1. **Execution services:** are responsible for scheduling and executing applications. Specialized implementations of these services are defined for execution of work units of each programming model supported by Aneka.
2. **Foundation services:** are in-charge of metering applications, allocating resources, managing the collection of available nodes, and keeping the services registry updated.
3. **Fabric services:** provide access to the physical and virtualized resources managed by the Aneka cloud. The Resource Provisioning Service (RPS) enables horizontal scaling out and allows for elastic and dynamic growth and shrinkage of the Aneka cloud to meet Quality of Service (QoS) requirements of applications.

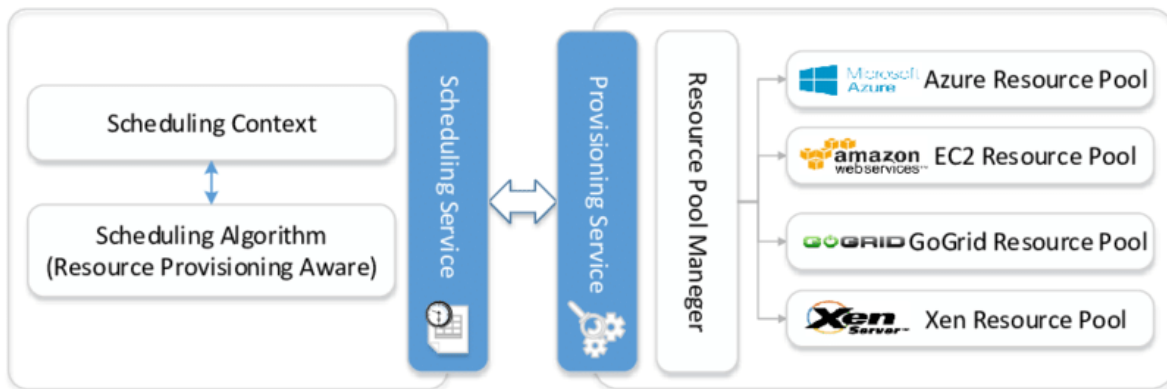
The services of the middleware are accessible through a set of interfaces and tools in the development and management layer. The Software Development Kit (SDK) embodies a collection of abstractions and APIs for definition of applications and leveraging existing programming models. The Management Kit contains a collection of tools for management, monitoring, and administration of Aneka clouds. All the management functions of the Aneka cloud are made accessible through the Management Studio, a comprehensive graphical environment providing a global view of the cloud for administrators.

3 . Explain Aneka resource provisioning in detail

Dynamic provisioning is the ability to dynamically acquire resources and integrate them into existing infrastructures and software systems. In the most common case, resources are Virtual Machines (VMs) acquired from an Infrastructure-as-a-Service (IaaS) cloud provider. Dynamic provisioning in Aneka happens as part of the Fabric Services by offering provisioning services for allocating virtual nodes from public cloud providers to complement local resources. This is mainly achieved as a result of the interaction between two services: the Scheduling Service and the Resource Provisioning Service. The former triggers on-demand provisioning requests based on the system status and the requirements of applications, while the latter is responsible for interacting with IaaS providers to instantiate VMs and deploy Aneka containers to meet the requests.

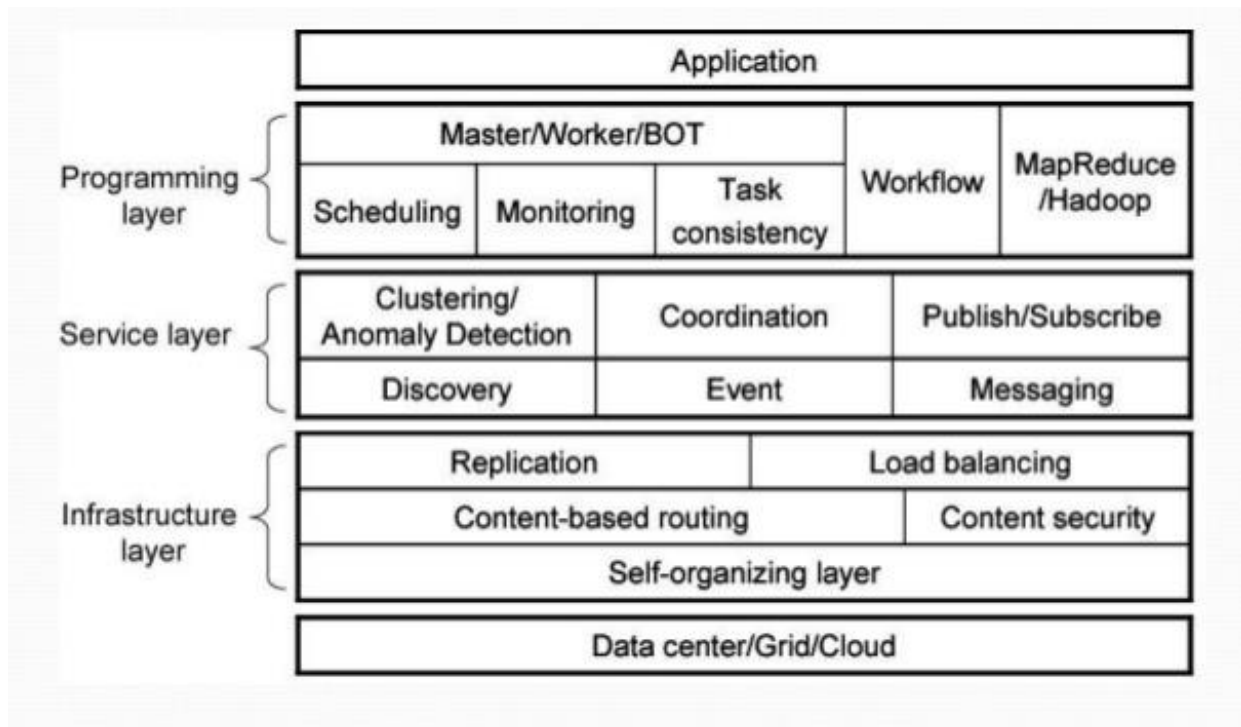
Execution of applications in Aneka happens through allocating tasks to the available set of resources in a dynamic fashion using the existing scheduling algorithms. Scheduling algorithms might be designed to leverage dynamic provisioning to cope with the application or system requirements. The scheduling algorithm makes decisions regarding when and how many resource allocations must take place to meet the application QoS requirements.

Aneka supports interactions with different resource providers, e.g., Amazon Elastic Computer Cloud (EC2), Microsoft Azure, XenServer, and GoGrid, using its dedicated provider-specific resource pool component. The main operations performed by this component are the translation of provisioning requests into provider specific requests, controlling the life cycle of VMs, and shutting them down when they are no longer needed. The life cycle of resource pools and redirecting provisioning requests, their release, or directing queries to the appropriate pool is the responsibility of the pool manager component. The pool manager also notifies the provisioning service when a dynamic resource is activated and terminated. The diagram given below illustrates a schematic overview of Aneka's dynamic provisioning mechanism



4. Explain in detail the architecture of comet cloud

Comet Cloud is an autonomic computing engine for cloud and Grid environments. It is based on the Comet [2] decentralized coordination substrate, and supports highly heterogeneous and dynamic cloud/Grid infrastructures, integration of public/private clouds and autonomic cloudbursts. A schematic overview of the architecture is given below.



Comet Cloud is composed of

- a programming layer,
- service layer, and
- Infrastructure layer.

Infrastructure layer

- The infrastructure layer uses the Chord self-organizing overlay, and the Squid information discovery and content-based routing substrate build on top of Chord.
- The routing engine supports flexible content-based routing and complex querying using partial keywords, wildcards, or ranges.
- It also guarantees that all peer nodes with data elements that match a query/message will be located.
- This layer also provides replication and load balancing services, and handles dynamic joins and leaves of nodes as well as node failures.

Service layer

- This layer provides a range of services to supports autonomies at the programming and application level.
- Asynchronous (publish/subscribe) messaging and event services are also provided by this layer.

Programming layer

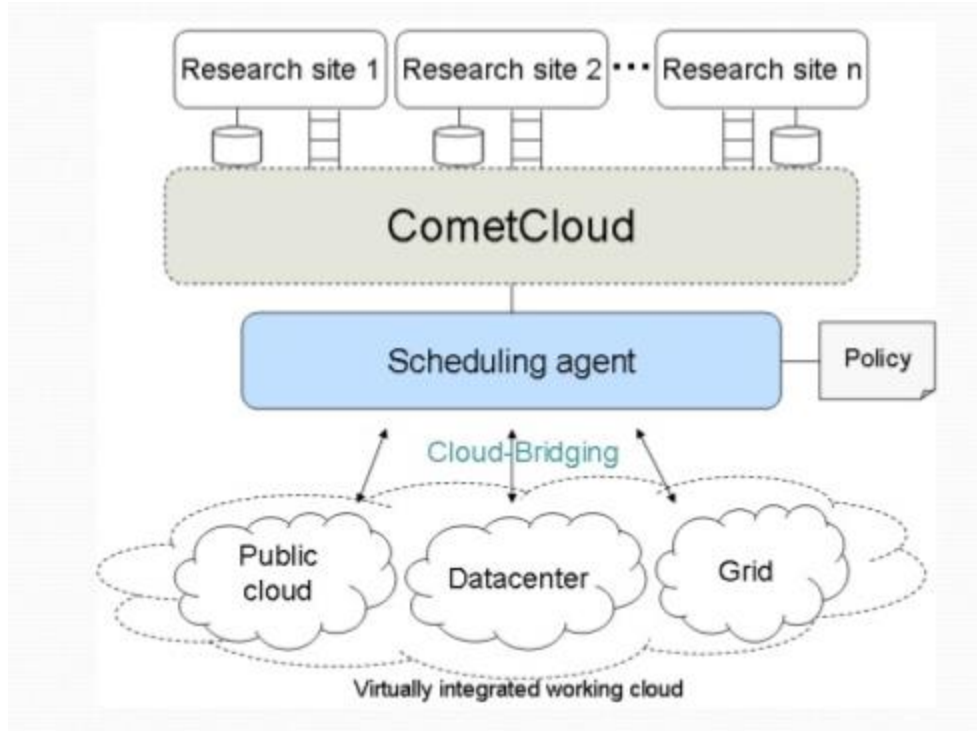
- This layer provides the basic framework for application development and management. It supports a range of paradigms including the master/worker/BOT.
- Masters generate tasks and workers consume them. Masters and workers can communicate via virtual shared space or using a direct connection.
- Scheduling and monitoring of tasks are supported by the application framework.
- The task consistency service handles lost/failed tasks.

5. Explain autonomic cloud bridging in Comet cloud

Autonomic cloud bridging is meant to connect CometCloud and a virtual cloud which consists of public cloud, data center, and grid by the dynamic needs of the application. The clouds in the virtual cloud are heterogeneous and have different types of resources and cost policies, besides, the performance of each cloud can change over time by the number of current users.

The scheduling agent manages autonomic cloudbursts over the virtual cloud, and there can be one or more scheduling agents. A scheduling agent is located at a robust/secure master site. If multiple collaborating research groups work together and each group requires generating tasks with its own data and managing the virtual cloud by its own policy, then it can have a separate scheduling agent in its master site. The requests for tasks generated by the different sites are logged in the CometCloud virtual shared space that spans master nodes at each of the sites. These tasks are then consumed by workers, which may run on local computational nodes at the site, a shared data center, and a grid or on a public cloud infrastructure.

A scheduling agent manages QoS constraints and autonomic cloudbursts of its site according to the defined policy. The workers can access the space using appropriate credentials, access authorized tasks, and return results back to the appropriate master indicated in the task itself.



A scheduling agent manages autonomic cloudbridging and guarantees QoS within user policies. Autonomic cloudburst is represented by changing resource provisioning not to violate defined policy.

There are three types of policies.

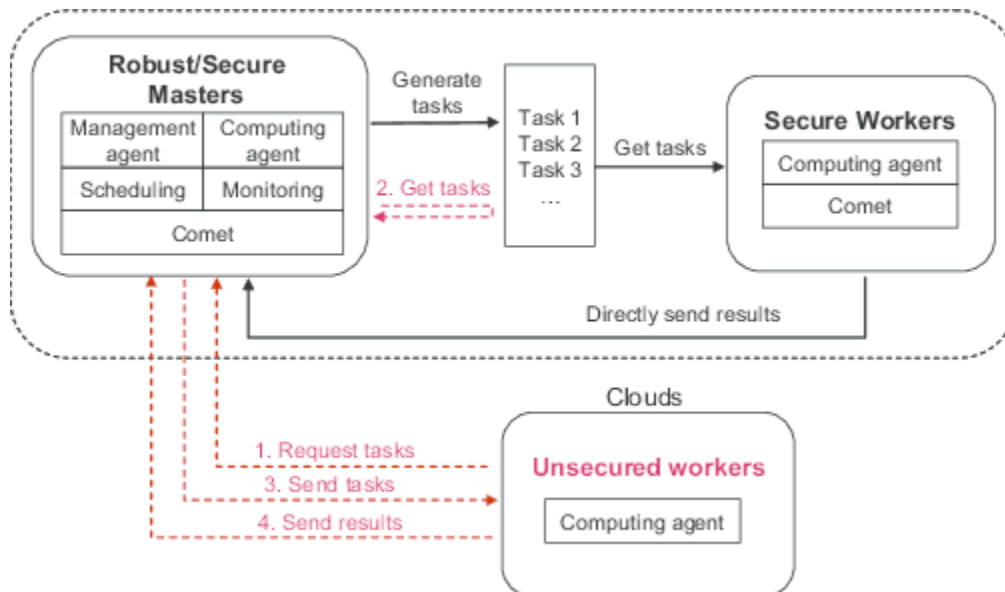
- **Deadline-Based.** When an application needs to be completed as soon as possible, assuming an adequate budget, the maximum required workers are allocated for the job.
- **Budget-Based.** When a budget is enforced on the application, the number of workers allocated must ensure that the budget is not violated.
- **Workload-Based.** When the application workload changes, the number of workers explicitly defined by the application is allocated or released.

6. Explain autonomic cloud bursting in Comet cloud

The goal of autonomic cloudbursts is to seamlessly and securely integrate private enterprise clouds and data centers with public utility clouds on-demand, to provide an abstraction of resizable computing capacity that is driven by user-defined high-level policies. It enables the dynamic deployment of application components, which typically run on internal organizational compute resources, onto a public cloud (i.e., cloudburst) to address dynamic workloads, spikes in demands, economic/budgetary issues, and other extreme requirements. Furthermore, given the increasing application and infrastructure scales, as well as their cooling, operation and

management costs, typical over-provisioning strategies are no longer feasible. Autonomic cloudbursts can leverage utility clouds to provide on-demand scale-out and scale-in capabilities based on a range of metrics.

The overall support for autonomic cloudbursts in CometCloud is given in the following diagram



Comet Cloud considers three types of clouds based on perceived security/trust and assigns capabilities accordingly. The first is a highly trusted, robust and secure cloud, usually composed of trusted/secure nodes within an enterprise, which is typically used to host masters and other key (management, scheduling, monitoring) roles. These nodes are also used to store state. In most applications, the privacy and integrity of critical data must be maintained, and as a result, tasks involving critical data should be limited to cloud nodes that have required credentials. The second type of cloud is one composed of nodes with such credentials, i.e., the cloud of secure workers. A privileged Comet space may span these two clouds and may contain critical data, tasks and other aspects of the application-logic/workflow. The final type of a cloud consists of casual workers. These workers are not part of the space but can access the space through the proxy to obtain (possibly encrypted) work units as long as they present required credentials and these credentials also define the nature of the access and type of data that can be accessed. Note that while nodes can be added or deleted from any of these clouds, autonomic cloudbursts primarily target worker nodes, and specifically worker nodes that do not host the Comet space as they are less expensive to add and delete.

7. List out the motivations of autonomic cloud bursting

Load dynamics: Application workloads can vary significantly. This includes the number of application tasks as well the computational requirements of a task. The computational environment must dynamically grow (or shrink) in response to these dynamics while still maintaining strict deadlines.

- **Accuracy of the analytics:** The required accuracy of risk analytics depends on a number of highly dynamic market parameters, and has a direct impact on the computational demand. The computational environment must be able to dynamically adapt to satisfy the accuracy requirements while still maintaining strict deadlines.

- **Collaboration of different groups:** Different groups can run the same application with different data sets policies. Here, policy means user's SLA bounded by their condition such as time frame, budgets and economic models.

- **Economics:** Application tasks can have very heterogeneous and dynamic priorities, and must be assigned resources and scheduled accordingly. Budgets and economic models can be used to dynamically provision computational resources based on the priority and criticality of the application task.

- **Failures:** Due to the strict deadlines involved, failures can be disastrous. The computation must be able to manage failures without impacting application quality of service, including deadlines and accuracies.

8. Give an overview of two Comet cloud applications

(i) Value at Risk (VaR)

(ii) Image registration

Value at Risk

A VaR calculation should be completed within the limited time and the computational requirements for the calculation can change significantly. Besides, the requirement for additional computation happens irregularly. Hence, for VaR we will focus on how autonomic cloudbursts work for dynamically changing workloads.

Image registration – It is the process to determine the linear/nonlinear mapping T between two images of the same object or similar objects which acquired at different time, or from different perspectives. Besides, because a set of image registration methods are used by different (geographically distributed) research groups to process their locally stored data, jobs can be injected from multiple sites. Another distinguished difference between two applications is that data size of image registration is much larger than that of VaR. In case of 3D image, image size is usually a few tens of mega bytes. Hence, image data should be separated from its task tuple and instead, it locates on a separate storage server and its location is indicated in the task tuple. For image registration, because it usually needs to be completed as soon as possible within budget limit. Comet Cloud uses budget based policy

UNIT IV

1. List out the benefits of dynamic ICT services?

- Standardization – reduced cost and increased flexibility
- Automation
- Modularization
- Integrated creation of ICT services

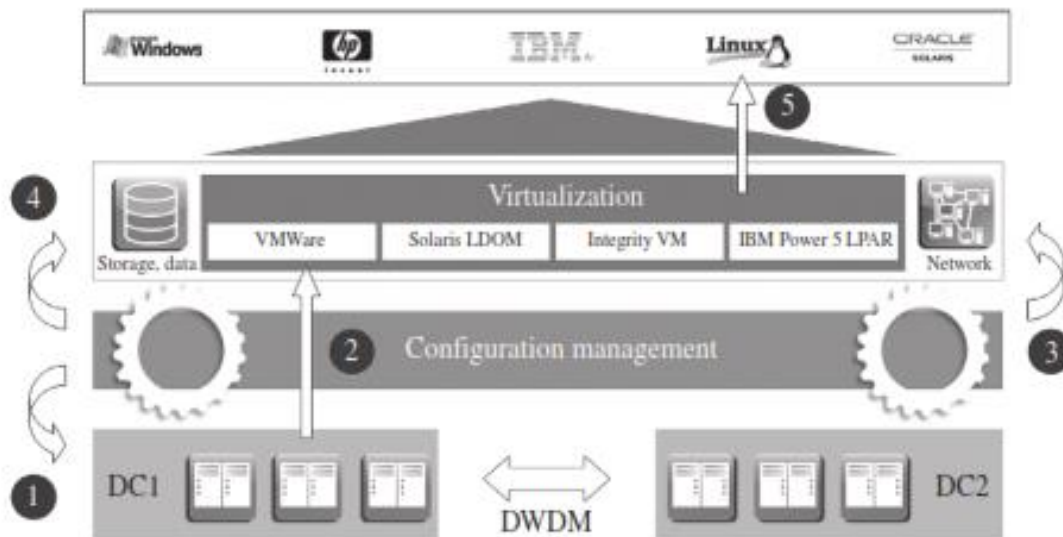
2. What is the use of T Systems?

T Systems is used to create dynamic and flexible ICT services by providing

- Standardization of customer's environment
- Technical consolidation
- Separation of logical and physical entities with the help of virtualization
- Process automation

3. Explain in detail T Systems cloud based solution for business applications (or) for dynamic ICT

T System's core modules are (i) Computing (ii) Storage



Computing module

The computing pool is based on server farms located in different data centers. Logical server systems are created automatically at these farms. The server systems comply with predefined standards. They are equipped with the network interface cards required for communications and

integration with storage systems. No internal hard drives or direct-attached storage systems are deployed.

The configuration management database (CMDB) plays a key role in computing resource pools (Figure 11.3). This selects and configures the required physical server (1). Once a server has been selected from the pool, virtualization technology is selected in line with the relevant application and the demands it has to meet (2). At the same time, the configuration requirements are sent to the network configuration management system (3) and to the storage configuration management system (4). Once all the necessary elements are in place, the storage systems are mounted on the servers, after which the operating-system images are booted (5)

Operating systems are provided in the form of images stored on a central storage system.

Storage module

The necessary storage is provided and configured in much the same way as the computing resources. IP-based storage systems are deployed. To reduce hardware-configuration effort, the computing systems use neither SAN nor direct-attached storage.

Using fiber-channel (FC) cards in the servers and deploying an FC network increases overall system complexity substantially. The IP storage systems are linked via Gbit Ethernet. Storage is automatically allocated to the server systems that require it.

Storage resources are located in different fire zones as well as in different data centers, preventing data loss in the event of a disaster

4. What is MapReduce?

MapReduce is a processing technique and a program model for distributed computing based on java. The MapReduce algorithm contains two important tasks, namely Map and Reduce. Map takes a set of data and converts it into another set of data, where individual elements are broken down into tuples (key/value pairs).

5. List out different phases of MapReduce.

A Map-Reduce job is divided into four simple phases,

1. Map phase

Map function operates on a single record at a time. On each input of key-value pair (LongWritable key, Text value) MapReduce framework will call map function with key and value as arguments

2. Combine phase

The combiner is the process of applying a reducer logic early on an output from a single map process. Mappers output is collected into an in memory buffer. MapReduce framework sorts this buffer

3. Shuffle phase

In the shuffle phase, MapReduce partitions data and sends it to a reducer. Each mapper sends a partition to each reducer.

4. Reduce phase

In reduce phase, each reducer copies its input partition from the output of each mapper. After copying all parts, the reducer first merges these parts and sorts all input records by key. In the Reduce phase, a reduce function is executed only once for each key found in the sorted output.

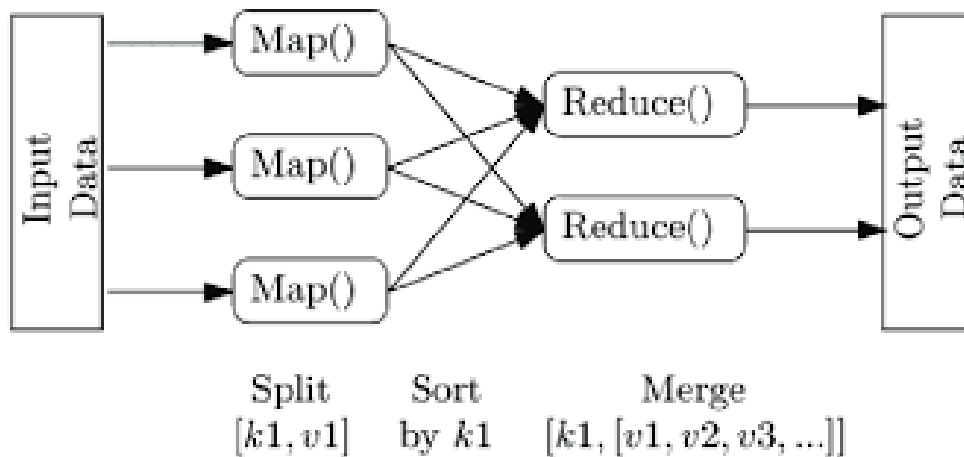
6. Explain MapReduce programming model

MapReduce is a programming framework that allows us to perform distributed and parallel processing on large data sets in a distributed environment.

MapReduce consists of two distinct tasks – Map and Reduce.

$$\begin{array}{lcl} \text{map}(k1, v1) & \rightarrow & \text{list}(k2, v2) \\ \text{reduce}(k2, \text{list}(v2)) & \rightarrow & \text{list}(v2) \end{array}$$

In the above model, the map() function is run in parallel against an input list of key(k1) value(v1) pairs. The map() implementation can perform any type of computation, to produce key(k2) and intermediate value(v2). The reduce() function does not have to keep track of different keys. The reduce() function, with the k2, v2 input, is able to perform a number of processes such as aggregation, sort, transformation, etc.



As the name MapReduce suggests, the reducer phase takes place after the mapper phase has been completed.

So, the first is the map job, where a block of data is read and processed to produce key-value pairs as intermediate outputs.

The output of a Mapper or map job (key-value pairs) is input to the Reducer.

The reducer receives the key-value pair from multiple map jobs.

Then, the reducer aggregates those intermediate data tuples (intermediate key-value pair) into a smaller set of tuples or key-value pairs which is the final output.

Let us understand more about MapReduce and its components. MapReduce majorly has the following three Classes. They are,

Mapper Class

The first stage in Data Processing using MapReduce is the Mapper Class. Here, RecordReader processes each Input record and generates the respective key-value pair. Hadoop's Mapper store saves this intermediate data into the local disk.

Input Split

It is the logical representation of data. It represents a block of work that contains a single map task in the MapReduce Program.

RecordReader

It interacts with the Input split and converts the obtained data in the form of Key-Value Pairs.

Reducer Class

The Intermediate output generated from the mapper is fed to the reducer which processes it and generates the final output which is then saved in the HDFS.

Driver Class

The major component in a MapReduce job is a Driver Class. It is responsible for setting up a MapReduce Job to run-in Hadoop. We specify the names of Mapper and Reducer Classes along with data types and their respective job names.

UNIT V

1. What is federated cloud?

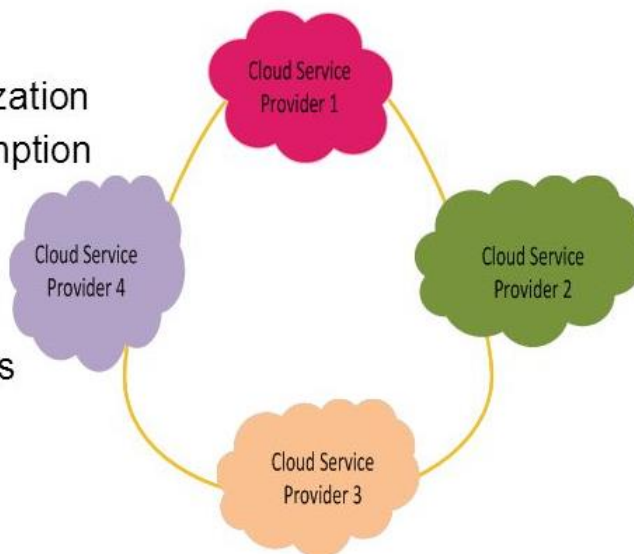
A federated cloud (also called cloud federation) is the deployment and management of multiple external and internal cloud computing services to match business needs. A federation is the union of several smaller parts that perform a common action.

Cloud federation is the practice of interconnecting the cloud computing environments of two or more service providers for the purpose of load balancing traffic and accommodating spikes in demand.

Different CSPs join together to form a federation

Benefits include:

- Maximize resource utilization
- Minimize power consumption
- Load balancing
- Cloud bursting
- Global Unity
- Expand Cloud provider's geographic footprints



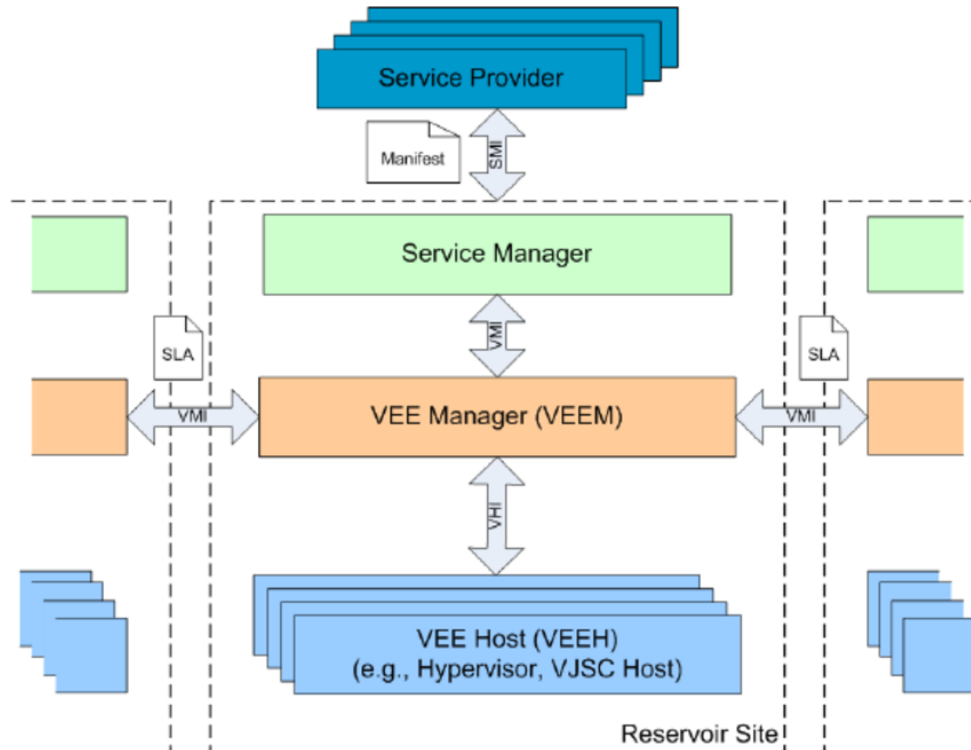
2. What is RESERVOIR?

The Reservoir project is motivated by the vision of implementing an architecture that would enable providers of cloud infrastructure to dynamically partner with each other to create a seemingly infinite pool of IT resources while fully preserving their individual autonomy in making technological and business management decisions.

3. Explain in detail the architecture of RESERVOIR

The essence of the RESERVOIR Service Cloud is to effectively manage a service specified as a collection of virtual execution environments (VEEs). A VEE is an abstraction representing both virtual machines running on a generic hypervisor infrastructure, as well as any application component that can be run (and/or migrated) on a leased infrastructure (e.g., Web applications on Google's App Engine, a Java based OSGi bundle). A Service Cloud, such as RESERVOIR, operates by acting as a platform for running virtualized applications in VEEs, which have been deployed on behalf of a service provider.

The service provider defines the details and requirements of the application in a Service Definition Manifest. This is done by specifying which virtual machine images are required to run, as well as specifications for (i) Elasticity Rules or performance objectives, which determine how the application will scale across a Cloud, and (ii) Service Level Agreement (SLA) Rules, which determine how and if the Cloud site is providing the right level of service to the application. Within each Service Cloud site there is a Service Manager (SM) and a VEE Manager (VEEM) which together provide all the necessary management functionality for both the services and the infrastructure. These management components of a Cloud system are shown in Figure 1 and are presented in more detail.



The Service Manager (SM) is the component responsible for accepting the Service Definition Manifest and the raw VEE images from the service provider. It is then responsible for the instantiation of the service application by requesting the creation and configuration of executable VEEs for each service component in the manifest. In addition, it is the Service Manager that is responsible for (i) evaluating and executing the elasticity rules and (ii) ensuring SLA compliance, by monitoring the execution of the service applications in real-time. Elasticity of a service is done by adjusting the application capacity, either by adding or removing service components and/or changing the resource requirements of a particular component according to the load and measurable application behavior.

The Virtual Execution Environment Manager (VEEM) is the component responsible for the placement of VEEs into VEE hosts (VEEHs). The VEEM receives requests from the Service Manager to create VEEs, to adjust resources allocated to VEEs, and to also finds the best placement for these VEEs in order to satisfy a given set of constraints. The role of the VEEM is

to optimize a site and its main task is to place and move the VEEs anywhere, even on remote sites, as long as the placement is done within the constraints set in the Manifest, including specifications of VEE affinity, VEE anti-affinity, security, and cost.

The Virtual Execution Environment Host (VEEH) is a resource that can host a certain type of VEEs. For example one type of a VEEH can be a physical machine with the Xen hypervisor

These three main components of the Service Cloud architecture interact with each other using specific interfaces, namely SMI (service management interface), VMI (VEE management interface), and VHI (VEE host interface), within a site and also use the VMI interface for site-to-site federation via the VEEM

4. What is SLO?

Service Level Objective (SLO) serves as a benchmark for indicators, parameters or metrics defined with specific service level targets.

A service level objective (SLO) is a key element of a service level agreement (SLA) between a service provider and a customer. SLOs are agreed as a means of measuring the performance of the Service Provider and are outlined as a way of avoiding disputes between the two parties based on misunderstanding

5. What is SLA? Briefly explain SLA life cycle

A cloud SLA (cloud service-level agreement) is an agreement or a contract between a cloud service provider and a customer that ensures a minimum level of service is maintained.

Each SLA goes through a sequence of steps starting from identification of terms and conditions, activation and monitoring of the stated terms and conditions, and eventual termination of contract once the hosting relationship ceases to exist. Such a sequence of steps is called SLA life cycle and consists of the following five phases:

1. Contract definition
2. Publishing and discovery
3. Negotiation
4. Operationalization
5. De-commissioning

Contract Definition- Generally, service providers define a set of service offerings and corresponding SLAs using standard templates. These service offerings form a catalog. Individual SLAs for enterprises can be derived by customizing these base SLA templates.

Publication and Discovery. Service provider advertises these base service offerings through standard publication media, and the customers should be able to locate the service provider by searching the catalog. The customers can search different competitive offerings and shortlist a few that fulfill their requirements for further negotiation.

Negotiation. Once the customer has discovered a service provider who can meet their application hosting need, the SLA terms and conditions needs to be mutually agreed upon before signing the agreement for hosting the application. At the end of this phase, the SLA is mutually agreed by both customer and provider and is eventually signed off.

Operationalization. SLA operation consists of SLA monitoring, SLA accounting, and SLA enforcement. SLA monitoring involves measuring parameter values and calculating the metrics defined as a part of SLA and determining the deviations. On identifying the deviations, the concerned parties are notified. SLA accounting involves capturing and archiving the SLA adherence for compliance.

De-commissioning. SLA decommissioning involves termination of all activities performed under a particular SLA when the hosting relationship between the service provider and the service consumer has ended. SLA specifies the terms and conditions of contract termination and specifies situations under which the relationship between a service provider and a service consumer can be considered to be legally ended.

6. Explain in detail the traditional approaches to SLO management

There are two techniques.

- (i) Load balancing
- (ii) Admission control

Load balancing

The objective of a load balancing is to distribute the incoming requests onto a set of physical machines, each hosting a replica of an application, so that the load on the machines is equally distributed.

The load balancing algorithm executes on a physical machine that interfaces with the clients. This physical machine, also called the front-end node, receives the incoming requests and distributes these requests to different physical machines for further execution.

This set of physical machines is responsible for serving the incoming requests and are known as the back-end nodes.

Typically, the algorithm executing on the front-end node is agnostic to the nature of the request. This means that the front-end node is neither aware of the type of client from which the request originates nor aware of the category (e.g., browsing, selling, payment, etc.) to which the request belongs to. This category of load balancing algorithms is known as class-agnostic.

There is a second category of load balancing algorithms that is known as class-aware. With class-aware load balancing and requests distribution, the front-end node must additionally inspect the type of client making the request and/or the type of service requested before deciding which back-end node should service the request.

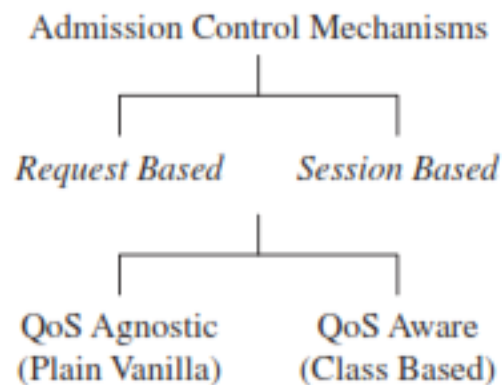
Admission Control

The objective of admission control mechanisms, therefore, is to police the incoming requests and identify a subset of incoming requests that can be admitted into the system when the system faces overload situations. Admission control algorithms are divided into two types.

(1) request-based algorithms and

(2) session-based algorithms.

- Request-based admission control algorithms reject new requests if the servers are running to their capacity.
- session-based admission control mechanisms try to ensure that longer sessions are completed and any new sessions are rejected. Accordingly, once a session is admitted into the server, all future requests belonging to that session are admitted as well.



7. Explain SLO management in Cloud

SLA management of applications hosted on cloud platforms involves five phases.

1. Feasibility
2. On-boarding
3. Pre-production
4. Production
5. Termination

Feasibility Analysis

Service Provider conducts the feasibility study of hosting an application on their cloud platforms.

This study involves three kinds of feasibility:

- (1) technical feasibility,
- (2) infrastructure feasibility, and
- (3) financial feasibility.

The technical feasibility of an application implies determining the following:

1. Ability of an application to scale out.
 2. Compatibility of the application with the cloud platform being used within the provider's data center.
 3. The need and availability of a specific hardware and software required for hosting and running of the application.
 4. Preliminary information about the application performance and whether they can be met by the provider
- Performing the infrastructure feasibility involves determining the availability of infrastructural resources in sufficient quantity so that the projected demands of the application can be met. The financial feasibility study involves determining the approximate cost to be incurred by the.

On-Boarding of Application

On-boarding activity consists of the following steps:

- Packing of the application for deploying on physical or virtual environments
- The packaged application is executed directly on the physical servers to capture and analyze the application performance characteristics.
- The application is executed on a virtualized platform and the application performance characteristics are noted again.
- Based on the measured performance characteristics, different possible SLAs are identified.
- Policies are created

Preproduction

- Once the determination of policies is completed as discussed in previous phase, the application is hosted in a simulated production environment.

Production

In this phase, the application is made accessible to its end users under the agreed SLA.

Termination

When the customer wishes to withdraw the hosted application and does not wish to continue to avail the services of the service provider for managing the hosting of its application, the termination activity is initiated

8. Explain different types of SLA. Give the structure of different SLAs

Infrastructure SLA. The infrastructure provider manages and offers guarantees on availability of the infrastructure, namely, server machine, power, network connectivity, and so on. Enterprises manage themselves, their applications that are deployed on these server machines. The machines are leased to the customers and are isolated from machines of other customers.

Key elements of Infrastructure SLA (or structure of Infrastructure SLA)

<i>Hardware availability</i>	● 99% uptime in a calendar month
<i>Power availability</i>	● 99.99% of the time in a calendar month
<i>Data center network availability</i>	● 99.99% of the time in a calendar month
<i>Backbone network availability</i>	● 99.999% of the time in a calendar month
<i>Service credit for unavailability</i>	● Refund of service credit prorated on downtime period
<i>Outage notification guarantee</i>	● Notification of customer within 1 hr of complete downtime
<i>Internet latency guarantee</i>	● When latency is measured at 5-min intervals to an upstream provider, the average doesn't exceed 60 msec
<i>Packet loss guarantee</i>	● Shall not exceed 1% in a calendar month

Application SLA. In the application co-location hosting model, the server capacity is available to the applications based solely on their resource demands. Hence, the service providers are flexible in allocating and de-allocating computing resources among the co-located applications. Therefore, the service providers are also responsible for ensuring to meet their customer's application SLOs.

Key elements of Application SLA/Structure of Application SLA

<i>Service-level parameter metric</i>	● Web site response time (e.g., max of 3.5 sec per user request)
<i>Function</i>	● Latency of web server (WS) (e.g., max of 0.2 sec per request)
	● Latency of DB (e.g., max of 0.5 sec per query)
<i>Measurement directive</i>	● Average latency of WS = (latency of web server 1 + latency of web server 2) / 2
	● Web site response time = Average latency of web server + latency of database
<i>Service-level objective</i>	● DB latency available via http://mgmtserver/em/latency
	● WS latency available via http://mgmtserver/ws/instanceno/latency
<i>Penalty</i>	● Service assurance
	● web site latency < 1 sec when concurrent connection < 1000
	● 1000 USD for every minute while the SLO was breached

9. What is cloud security?

Cloud security is the protection of data stored online via cloud computing platforms from theft, leakage, and deletion. Methods of providing cloud security include firewalls, penetration testing,

obfuscation, tokenization, virtual private networks (VPN), and avoiding public internet connections.

10. What are security issues in cloud computing?

- Distributed-Denial-of-Service Attacks. ...
- Shared Cloud Computing Services. ...
- Employee Negligence. ...
- Data Loss and Inadequate Data Backups. ...
- Phishing and Social Engineering Attacks. ...
- System Vulnerabilities.

11. Why cloud security is important?

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.

12. What does Data Security mean?

Data security refers to protective digital privacy measures that are applied to prevent unauthorized access to computers, databases and websites. Data security also protects data from corruption. ... Data security is also known as information security

13. Explain in detail, the key concepts of data security in cloud

There are 8 key concepts

(1) Privacy Protection

Consumer's data should be protected from unauthorized access regardless

(2) Preserve Data Integrity

Data integrity can be defined as protecting data from unauthorized modification or deletion. This is easy in a single database, because there is only one way in or out of the database. But in the cloud, especially a multicloud environment, it gets difficult.

Because of the large number of data sources and means to access, authorization becomes crucial in assuring that only authorized entities can interact with data. This means stricter means of access, like two-factor authorization, and logging to see who accessed what. Another potential means of security is a trusted platform module (TPM) for remote data checks.

(3) data availability

Data should be always available for use

(4) Data privacy

Many providers may store data on servers not physically located in a region as the data owner and the laws may be different.

(5) Encryption

Encryption is the means for which data privacy is protected and insured and encryption is done via key-based algorithms and the keys are stored by the cloud provider.

Virtually every cloud storage provider encrypts the data while it is in transfer.

Many cloud services offer key management solutions that allow you to control access because the encryption keys are in your hands. This may prove to be a better or at least more reassuring risk because you are in control over who has the keys. Again, this should be spelled out in the SLA.

(6) Threats

Cloud service providers have a variety of security tools and policies in place but problems still happen, usually originating in human error.

- **Data breaches:** This can happen any number of ways, from the usual means – a hacked account or a lost password/laptop – to means unique to the cloud. For example, it is possible for a user on one virtual machine to listen for the signal that an encryption key has arrived on another VM on the same host. It's called the "side channel timing exposure," and it means the victim's security credentials in the hands of someone else.
- **Data loss:** While the chance of data loss is minimal short of someone logging in and erasing everything, it is possible. You can mitigate this by insuring your applications and data are distributed across several zones and you backup your data using off-site storage.
- **Hijacked accounts:** All it takes is one lost notebook for someone to get into your cloud provider. Secure, tough passwords and two-factor authentication can prevent this. It also helps to have policies that look for and alert to unusual activity, like copying mass amounts of data or deleting it.
- **Cryptojacking:** Cryptojacking is the act of surreptitiously taking over a computer to farm cryptocurrency, which is a very compute-intensive process. Cryptojacking spiked in 2017 and 2018 and the cloud was a popular target because there is more compute resources available. Monitoring for unusual compute activity is the key way to stop this.

(7) Data Security and Staff

Insider related incidents were caused by careless employees or contractors.

(8) Contractual Data Security

The SLA should include a description of the services to be provided and their expected levels of service and reliability, along with a definition of the metrics by which the services are measured,

the obligations and responsibilities of each party, remedies or penalties for failure to meet those metrics, and rules for how to add or remove metrics.

There are multiple checkmarks for a SLA.

- Specifics of services provided, such as uptime and response to failure.
- Definitions of measurement standards and methods, reporting processes, and a resolution process.
- An indemnification clause protecting the customer from third-party litigation resulting from a service level breach.

14. Briefly describe how security responsibilities are shared between consumer and provider. (or) cloud security strategies

Cloud security involves the procedures and technology that secure cloud computing environments against both external and insider cybersecurity threats. Cloud security and security management best practices designed to prevent unauthorized access are required to keep data and applications in the cloud secure from current and emerging cybersecurity threats.

In each public cloud service type, the cloud provider and cloud customer share different levels of responsibility for security. By service type, these are:

- Software-as-a-service (SaaS) — Customers are responsible for securing their data and user access.
- Platform-as-a-service (PaaS) — Customers are responsible for securing their data, user access, and applications.
- Infrastructure-as-a-service (IaaS) — Customers are responsible for securing their data, user access, applications, operating systems, and virtual network traffic.

Within all types of public cloud services, customers are responsible for securing their data and controlling who can access that data.

The sharing of responsibilities between provider and consumer is shown

Responsibility	On-Prem	IaaS	PaaS	SaaS
Data classification & accountability	Cloud Customer	Cloud Customer	Cloud Customer	Cloud Customer
Client & end-point protection	Cloud Customer	Cloud Customer	Cloud Customer	Cloud Customer / Cloud Provider
Identity & access management	Cloud Customer	Cloud Customer	Cloud Customer / Cloud Provider	Cloud Customer / Cloud Provider
Application level controls	Cloud Customer	Cloud Customer	Cloud Customer / Cloud Provider	Cloud Provider
Network controls	Cloud Customer	Cloud Customer / Cloud Provider	Cloud Provider	Cloud Provider
Host infrastructure	Cloud Customer	Cloud Customer / Cloud Provider	Cloud Provider	Cloud Provider
Physical security	Cloud Customer	Cloud Provider	Cloud Provider	Cloud Provider

■ Cloud Customer ■ Cloud Provider

References

1. <https://www.bluepiit.com/blog/different-types-of-cloud-computing-service-models/>
2. <https://www.javatpoint.com/virtualization-in-cloud-computing>
3. Ade Nadjaran Toosi, Richard O.Sinnott, Rajkumar Buyya, Resource provisioning for data-intensive applications with deadline constraints on hybrid clouds using Aneka, Future Generation Computer Systems Volume 79, Part 2, February 2018, Pages 765-775
4. Rajkumar Buyya, James Broberg, Andrzej Goscinski, Cloud Computing: Principles and Paradigms, Wiley, 2011
5. <https://www.cloudflare.com/learning/cloud/what-is-cloud-migration/>