UNIT 1
Cloud Computing

Cloud Computing and its Characteristics

Cloud Computing is defined as storing and accessing of data and computing services over the internet. It doesn't store any data on your personal computer. It is the on-demand availability of computer services like servers, data storage, networking, databases, etc. The main purpose of cloud computing is to give access to data centers to many users. Users can also access data from a remote server.Whenever you travel through a bus or train, you take a ticket for your destination and hold back to your seat till you reach your destination. Likewise other passengers also takes ticket and travel in the same bus with you and it hardly bothers you where they go. When your stop comes you get off the bus thanking the driver. Cloud computing is just like that bus, carrying data and information for different users and allows to use its service with minimal cost.

The term "Cloud" came from a network design that was used by network engineers to represent the location of various network devices and there inter-connection. The shape of this network design was like a cloud.

Why Cloud Computing?

With increase in computer and Mobile user's, data storage has become a priority in all fields. Large and small scale businesses today thrive on their data & they spent a huge amount of money to maintain this data. It requires a strong IT support and a storage hub. Not all businesses can afford high cost of in-house IT infrastructure and back up support services. For them Cloud Computing is a cheaper solution. Perhaps its efficiency in storing data, computation and less maintenance cost has succeeded to attract even bigger businesses as well.

Cloud computing decreases the hardware and software demand from the user's side. The only thing that user must be able to run is the cloud computing systems interface software, which can be as simple as Web browser, and the Cloud network takes care of the rest. We all have experienced cloud computing at some instant of time, some of the popular cloud services we have used or we are still using are mail services like gmail, hotmail or yahoo etc.

While accessing e-mail service our data is stored on cloud server and not on our computer. The technology and infrastructure behind the cloud is invisible. It is less important whether cloud services are based on HTTP or other specific technologies as far as it is user friendly and functional. An individual user can connect to cloud system from his/her own devices like desktop, laptop or mobile.

Cloud computing harnesses small business effectively having limited resources, it gives small businesses access to the technologies that previously were out of their reach. Cloud computing helps small businesses to convert their maintenance cost into profit. Let's see how?

In an in-house IT server, you have to pay a lot of attention and ensure that there are no flaws into the system so that it runs smoothly. And in case of any technical glitch you are completely responsible; it will seek a lot of attention, time and money for repair. Whereas, in cloud computing, the service provider takes the complete responsibility of the complication and the technical faults.

Benefits of Cloud Computing

The potential for cost saving is the major reason of cloud services adoption by many organizations. Cloud computing gives the freedom to use services as per the requirement and pay only for what you use. Due to cloud computing it has become possible to run IT operations as a outsourced unit without much in-house resources.

Following are the benefits of cloud computing:

1. Lower IT infrastructure and computer costs for users

2. Improved performance
3. Fewer Maintenance issues
4. Instant software updates
5. Improved compatibility between Operating systems
6. Backup and recovery
7. Performance and Scalability
8. Increased storage capacity
9. Increase data safety

History of cloud computing

On the basis of above computing, there was emerged of cloud computing concepts that later implemented.

At around in 1961, John MacCharty suggested in a speech at MIT that computing can be sold like a utility, just like a water or electricity. It was a brilliant idea, but like all brilliant ideas, it was ahead if its time, as for the next few decades, despite interest in the model, the technology simply was not ready for it.

But of course time has passed and the technology caught that idea and after few years we mentioned that:

In 1999, **Salesforce.com** started delivering of applications to users using a simple website. The applications were delivered to enterprises over the Internet, and this way the dream of computing sold as utility were true.

In 2002, **Amazon** started Amazon Web Services, providing services like storage, computation and even human intelligence. However, only starting with the launch of the Elastic Compute Cloud in 2006 a truly commercial service open to everybody existed.

In 2009, **Google Apps** also started to provide cloud computing enterprise applications.

Of course, all the big players are present in the cloud computing evolution, some were earlier, some were later. In 2009, **Microsoft** launched Windows Azure, and companies like Oracle and HP have all joined the game. This proves that today, cloud computing has become mainstream.

Applications

Cloud Computing Applications

Cloud Computing can run every programs and software as a normal computer can run. It can also provide us with numerous applications which are free of cost. So, let's start elaborating these Cloud Computing applications one by one:

i. Storing File Online

Cloud Computing provides a benefit to store and access the software with the help of internet connection to the Cloud. The interface provided is very easy to operate and is economical too.

ii. Video Making and Editing Software

There are many software available which can access with the help of the cloud. This software helps to create and modify the videos. The videos create or modify are stored in the cloud itself and we can access anytime.

iii. File Converters

There are many applications which utilize to change to format of the file such that from HTML to pdf and so on. This software is available at cloud and access from anywhere with the help of internet connection.

iv. Anti-Virus Applications

There is software which is stored in the cloud and from there they fix the system. All the viruses and the malware are detected and analyzed by the software and the system is fixed. They also come up with a feature of downloading the software.

v. E-commerce Application

With the help of e-commerce application in the cloud, user and e-business allow responding quickly to the opportunities which are emerging. It also allows the user to respond quickly to the market opportunities and the challenges. Business tycoons focus on the usage of cloud computing without keeping time in the mind. Cloud-based e-commerce applications allow the companies, business leaders to evaluate new opportunities and making things done with the minimum amount possible.

vi. Business Process

Business management applications are based on the cloud service provider. The business utilizes the cloud computing to store the necessary data and all the relevant information. This information can be anything such as the personal data of the customer, analyzed records, and many more.

vii. Backup and Recovery

The cloud computing can be used as a backup option in which we can store the files, information, and the data. This data is stored will be protected and provided much security. When the data is lost the user can recover the data which he/she has stored in the cloud.

3. Cloud Computing Use Cases

After studying Cloud Computing Applications, its time to explore.

i. Private Cloud and Hybrid Cloud

There are situations where the firms are searching for the ways through which they can find a way to access the applications they **intend to deploy into their environment through the use of a cloud.** This leads to the fact that providing the facilities without the initial investment will be rendered useless and the workload testing fails.

ii. Big Data Analysis

Cloud Computing can store a tremendous amount of data which can also help Big Data. **Big Data,** a large amount of data (structured or unstructured) is analyzed for further analysis or for decision making in the business.

iii. Disaster Recovery

Disaster Recovery is one of the major benefits which gathers from Cloud Computing. It provides an economical way from the disaster recovery as there is a solution which provides a faster recovery from the congested different physical locations. The traditional DR sites can cost much of the amount which has fixed assets, tough productions, and a much higher cost.

iv. Iaas and PaaS

While using **Infrastructure as a service** there is a pay as you go through the scheme available. It benefits the companies and organizations by cutting the cost of investing to maintain the IT infrastructure. Moreover, there is an instance where the companies using **Platform as a Service** searching to increase the speed of development on a ready-to-use platform to deploy applications.
So, this was all about Cloud Computing Applications and its use cases. Hope you found this helpful.

Cloud Computing has provided many solutions which are useful for companies as well as individuals. The Cloud Computing helps by providing the solutions in the minimum cost possible. Cloud Computing has many examples which can be in the field of everything such as messaging apps, audio, and video service.

Grid Computing Vs Cloud Computing

When we switch on the fan or any electric device, we are less concern about the power supply from where it comes and how it is generated. The power supply or electricity that we receives at our home travels through a chain of network, which includes power stations, transformers, power lines and transmission stations. These components together make a 'Power Grid'. Likewise, 'Grid Computing' is an infrastructure that links computing resources such as PCs, servers, workstations and storage elements and provides the mechanism required to access them.

Grid Computing is a middle ware to co-ordinate disparate IT resources across a network, allowing them to function as whole. It is more often used in scientific research and in universities for educational purpose. For example, a group of architect students working on a different project requires a specific designing tool and a software for designing purpose but only couple of them got access to this designing tool, the problem is how they can make this tool available to rest of the students. To make available for other students they will put this designing tool on campus network, now the grid will connect all these computers in campus network and allow student to use designing tool required for their project from anywhere.

Cloud computing and Grid computing is often confused, though there functions are almost similar there approach for their functionality is different. Let see how they operate-

• Cloud computing works more as a service provider for utilizing computer resource	• Grid computing uses the available resource and interconnected computer systems to accomplish a common goal
• Cloud computing is a centralized model	• Grid computing is a decentralized model, where the computation could occur over many administrative model
• Cloud is a collection of computers usually owned by a single party. •	• A grid is a collection of computers which is owned by a multiple parties in multiple locations and connected together so that users can share the combined power of resources
• Cloud offers more services all most all the services like web hosting, DB (Data Base) support and much more	• Grid provides limited services
• Cloud computing is typically provided within a single organization (eg :	• Grid computing federates the resources located within different organization.

Amazon)	

Utility Computing Vs Cloud Computing

In our previous conversation in “Grid Computing” we have seen how electricity is supplied to our house, also we do know that to keep electricity supply we have to pay the bill. Utility Computing is just like that, we use electricity at home as per our requirement and pay the bill accordingly likewise you will use the services for the computing and pay as per the use this is known as ‘Utility computing’. Utility computing is a good source for small scale usage, it can be done in any server environment and requires Cloud Computing.

Utility computing is the process of providing service through an on-demand, pay per use billing method. The customer or client has access to a virtually unlimited supply of computing solutions over a virtual private network or over the internet, which can be sourced and used whenever it’s required. Based on the concept of utility computing , grid computing, cloud computing and managed IT services are based.

Through utility computing small businesses with limited budget can easily use software like CRM (Customer Relationship Management) without investing heavily on infrastructure to maintain their clientele base.

• Utility computing refers to the ability to charge the offered services, and charge customers for exact usage	• Cloud Computing also works like utility computing, you pay only for what you use but Cloud Computing might be cheaper, as such, Cloud based app can be up and running in days or weeks.
• Utility computing users want to be in control of the geographical location of the infrastructure	• In cloud computing, provider is in complete control of cloud computing services and infrastructure
• Utility computing is more favorable when performance and selection infrastructure is critical	• Cloud computing is great and easy to use when the selection infrastructure and performance is not critical
• Utility computing is a good choice for less resource demanding	• Cloud computing is a good choice for high resource demanding
• Utility computing refers to a business model	• Cloud computing refers to the underlying IT architecture

UNIT 2

Service and Deployment Models

The cloud" refers to servers that are accessed over the Internet, and the software and databases that run on those servers. Cloud servers are located in data centers all over the world. By using cloud computing, users and companies don't have to manage physical servers themselves or run software applications on their own machines.

Types of Clouds

There are four different cloud models that you can subscribe according to business needs:

1. Public Cloud: This type of cloud is used usually for B2C (Business to Consumer) type interactions. Here the computing resource is owned, governed and operated by government, an academic or business organization.
2. Private Cloud: Here, computing resources are deployed for one particular organization. This method is more used for intra-business interactions. Where the computing resources can be governed, owned and operated by the same organization.
3. Hybrid Cloud: This type of cloud can be used for both type of interactions - B2B (Business to Business) or B2C (Business to Consumer). This deployment method is called hybrid cloud as the computing resources are bound together by different clouds.

Now we discuss all in detail

1)Public clouds are the most common way of deploying cloud computing. The cloud resources (like servers and storage) are owned and operated by a third-party cloud service provider and delivered over the Internet. Microsoft Azure is an example of a public cloud. With a public cloud, all hardware, software, and other supporting infrastructure is owned and managed by the cloud provider. In a public cloud, you share the same hardware, storage, and network devices with other organizations or cloud “tenants.” You access services and manage your account using a web browser. Public cloud deployments are frequently used to provide web-based email, online office applications, storage, and testing and development environments.

Advantages of public clouds:

- Lower costs—no need to purchase hardware or software, and you pay only for the service you use.
- No maintenance—your service provider provides the maintenance.
- Near-unlimited scalability—on-demand resources are available to meet your business needs.
- High reliability—a vast network of servers ensures against failure.

2)Private cloud

A private cloud consists of computing resources used exclusively by one business or organization. The private cloud can be physically located at your organization’s on-site data center, or it can be hosted by a third-party service provider. But in a private cloud, the services and infrastructure are always maintained on a private network and the hardware and software are dedicated solely to your organization. In this way, a private cloud can make it easier for an organization to customize its resources to meet specific IT requirements. Private clouds are often used by government agencies, financial institutions, any other mid- to large-size organizations with business-critical operations seeking enhanced control over their environment.

Advantages of a private clouds:

- More flexibility—your organization can customize its cloud environment to meet specific business needs.
- Improved security—resources are not shared with others, so higher levels of control and security are possible.
- High scalability—private clouds still afford the scalability and efficiency of a public cloud.

3)Hybrid cloud?

Often called “the best of both worlds,” hybrid clouds combine on-premises infrastructure, or private clouds, with public clouds so organizations can reap the advantages of both. In a hybrid cloud, data and applications can move between private and public clouds for greater flexibility and more deployment options. For instance, you can use the public cloud for high-volume, lower-security needs such as web-based email, and the private cloud (or other on-premises infrastructure) for sensitive, business-critical operations like financial reporting. In a hybrid cloud, “cloud bursting” is also an option. This is when an application or resource runs in the private cloud until there is a spike in demand (such as seasonal event like online shopping or tax filing), at which point the organization can “burst through” to the public cloud to tap into additional computing resources.

Advantages of hybrid clouds:

- Control—your organization can maintain a private infrastructure for sensitive assets.
- Flexibility—you can take advantage of additional resources in the public cloud when you need them.
- Cost-effectiveness—with the ability to scale to the public cloud, you pay for extra computing power only when needed.
- Ease—transitioning to the cloud doesn’t have to be overwhelming because you can migrate gradually—phasing in workloads over time.

Cloud Computing Services

The three major Cloud Computing Offerings are

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

1)Infrastructure As A Service (IAAS)

IaaS (Infrastructure as a Service)

Infrastructure as a service (IaaS) is also known as Hardware as a service (HaaS).IaaS customers pay on a per-use basis, typically by the hour, week or month. Some providers also charge customers based on the amount of virtual machine space they use.It simply provides the underlying operating systems, security, networking, and servers for developing such applications, services, and for deploying development tools, databases IaaS (Infrastructure As A Service) is one of the fundamental service model of cloud computing alongside PaaS(Platform as a Service). It provides access to computing resources in a virtualized environment “the cloud” on internet. It provides computing infrastructure like virtual server space, network connections, bandwidth, load balancers and IP addresses. The pool of hardware resource is extracted from multiple servers and networks usually distributed across numerous data centers. This provides redundancy and reliability to IaaS.

IaaS(Infrastructure as a service) is a complete package for computing. For small scale businesses who are looking for cutting cost on IT infrastructure, IaaS is one of the solutions. Annually a lot of money is spent in maintenance and buying new components like hard-drives, network connections, external storage device etc. which a business owner could have saved for other expenses by using IaaS.Infrastructure as a Service (IAAS) is a form of cloud computing that provides virtualized computing resources over the internet. In an IAAS model, a third party provider hosts hardware, software, servers, storage and other infrastructure components on the behalf of its users. IAAS providers also host users' applications and handle tasks including system maintenance backup and resiliency planning.IAAS platforms offer highly scalable resources that can be adjusted on-demand which makes it a well-suited for workloads that are temporary, experimental or change unexpectedly. Other characteristics of IAAS environments include the automation of administrative tasks, dynamic scaling, desktop virtualization and policy-based services. Other characteristics of IAAS include the automation of administrative tasks, dynamic scaling, desktop virtualization and policy-based services.

Advantages of IaaS :

1. Cost Effective : Eliminates capital expense and reduces ongoing cost and IaaS customers pay on a per use basis, typically by the hour, week or month.
2. Website hosting : Running websites using IaaS can be less expensive than traditional web hosting.
3. Security : The IaaS Cloud Provider may provide better security than your existing software.
4. Maintainence : There is no need to manage the underlying data center or the introduction of new releases of the development or underlying software. This is all handled by the IaaS Cloud Provider.The various companies providing Infrastructure as a service are Amazon web services, Bluestack, IBM, Openstack, Rackspace and Vmware

2)Platform As A Service (PAAS)

Platform as a service, is referred as PaaS, it provides a platform and environment to allow developers to build applications and services. This service is hosted in the cloud and accessed by the users via internet.To understand in a simple terms, let compare this with painting a picture, where you are provided with paint colors, different paint brushes and paper by your school teacher and you just have to draw a beautiful picture using those tools. PaaS services are constantly updated & new features added. Software developers, web developers and business can benefit from PaaS. It provides platform to support application development. It includes software support and management services, storage, networking, deploying, testing, collaborating, hosting and maintaining applications.Platform as a Service (PAAS) is a cloud computing model that delivers applications over the internet. In a PAAS model, a cloud provider delivers hardware and software tools, usually those needed for application development, to its users as a service. A PAAS provider hosts the hardware and software on its own infrastructure. As a result, PAAS frees users from having to install in-house hardware and software to develop or run a new application.

Some of the main characteristics of PAAS are :

- Scalability and auto-provisioning of the underlying infrastructure.
- Security and redundancy.
- Build and deployment tools for rapid application management and deployment.
- Integration with other infrastructure components such as web services, databases, and LDAP.
- Multi-tenancy, platform service that can be used by many concurrent users.
- Logging, reporting, and code instrumentation.

Advantages of PaaS :

1 Simple and convenient for users : It provides much of the infrastructure and other IT services, which users can access anywhere via a web browser
2 Cost Effective : It charges for the services provided on a per-use basis thus eliminating the expenses one may have for on-premises hardware and software
3 Efficiently managing the lifecycle : It is designed to support the complete web application lifecycle: building, testing, deploying, managing and updating.
4 Efficiency : It allows for higher-level programming with reduced complexity thus, the overall development of the application can be more effective
The various companies providing Platform as a service are Amazon Web services, Salesforce, Windows Azure, Google App Engine, cloud Bess and IBM smart cloud

3 Software As A Service (SAAS)

SaaS or software as a service is a software distribution model in which applications are hosted by a vendor or service provider and made available to customers over a network (internet). SaaS is becoming an increasingly prevalent delivery model as underlying technologies that supports Service Oriented Architecture (SOA) or Web Services. Through internet this service is available to users anywhere in the world.Traditionaly, software application needed to be purchased upfront &then installed it onto your computer. SaaS users on the other hand, instead of purchasing the software subscribes to it, usually on monthly basis via internet.Anyone who needs an access to a particular piece of software can be subscribe as a user, whether it is one or two people or every thousands of employees in a corporation. SaaS is compatible with all internet enabled devices.Many important tasks like accounting, sales, invoicing and planning all can be performed using SaaS.Software as a Service(SAAS) is a software distribution model in which applications are hosted by a vendor or service provider and made available to customers over a network, typically the Internet. SAAS has become an increasingly prevalent delivery model as underlying technologies that support Web services and service-oriented architecture (SOA) mature and new development approaches, such as Ajax, become popular. SAAS is closely related to the ASP (Application service provider) and on-demand computing software delivery models. IDC identifies two slightly different delivery models for SAAS namely the hosted application model and the software development model.

Some of the Features of SAAS model are:

- Easier administration.
- automatic updates and patch management.
- compatibility: all users will have the same version of software.
- easier collaboration, for the same reason.
- global accessibility.

Advantages of SaaS :
1. Cost Effective : Pay only for what you use
2 Reduced time : Users can run most SaaS apps directly from their web browser without needing to download and install any software.This reduces the time spent in installation and configuration, and can reduce the issues that can get in the way of the software deployment
3 Accessibility : We can Access app data from anywhere.
4 Automatic updates : Rather than purchasing new software, customers rely on a SaaS provider to automatically perform the updates
5 Scalability : It allows the users to access the services and features on demand
The various companies providing software as a service are Cloud9 Analytics, Salesforce.com, Cloud Switch, Microsoft Office 365, Eloqua, dropBox and Cloud Tran

Cloud service providers:

A cloud service provider, or CSP, is a company that offers some component of cloud computing -- typically infrastructure as a service (IaaS), software as a service (SaaS) or platform as a service (PaaS) -- to other businesses or individuals.

Types of cloud service providers

Customers will purchase an increasing variety of services from cloud service providers today. As mentioned above, the most common categories of cloud-based services include IaaS, SaaS and PaaS.

IaaS providers. In the IaaS model, the cloud service provider delivers infrastructure components that would otherwise exist in an on-premises data center. These components could consist of servers, storage and networking as well as the virtualization layer, which the IaaS provider hosts in its own data center. Cloud service providers may also complement their IaaS products with services such as monitoring, security, load balancing and storage resiliency.

SaaS providers. SaaS vendors currently offer a wide array of business technologies, such as productivity suites, customer relationship management (CRM) software and human resources management (HRM) software, all of which the SaaS vendor hosts and provides over the internet. Many traditional software vendors now sell cloud-based options of their on-premises software products.

PaaS providers. The third type of cloud service provider, PaaS vendors, offers cloud infrastructure and services that users can access to perform various functions. PaaS products are commonly used in software development. In comparison to an IaaS provider, PaaS providers will add more of the application stack, such as operating systems and middleware, to the underlying infrastructure.

Amazon Web Services (AWS) is a giant that enables companies across the globe to develop their digital infrastructure entirely or partially using the cloud. Amazon was the first major cloud provider, with the 2006 offering of Amazon Simple Storage Service (Amazon S3). The growing cloud market since saw rapid development of Amazon's cloud platform as well as Microsoft's Azure platform and the Google Cloud Platform, and the three vendors continue to jockey for the lead on a variety of cloud fronts today. The vendors are currently engaged in developing cloud-based services around emerging technologies such as machine learning, artificial intelligence (AI) and containerization.

Unit 3
Service Level Agreement (SLA) Management

Service Level Management (SLM) deals with negotiating, agreeing and documenting existing services with some level of policies.Service Level Manager is the process owner of this process.

SLM deals with following two kinds of agreements −

- Service Level Agreement (SLA)
- Operational Level Agreement (OLA)

Service Level Agreement (SLA)

It is agreed document assuring the warranty with regard to level of service quality delivered by the service provider. It is between service provider and the customer. A Service Level Agreement (SLA) is the service contract component between a service provider and customer. A SLA provides specific and measurable aspects related to service offerings. For example, SLAs are often included in signed agreements between Internet service providers (ISP) and customers.SLA is also known as an operating level agreement (OLA) when used in an organization without an established or formal provider-customer relationship. Service Level Agreement (SLA)Adopted in the late 1980s, SLAs are currently used by most industries and markets. By nature, SLAs define service output but defer methodology to the service provider's discretion. Specific metrics vary by industry and SLA purpose.A Service Level Agreement (SLA) is the service contract component between a service provider and customer. A SLA provides specific and measurable aspects related to service offerings. For example, SLAs are often included in signed agreements between Internet service providers (ISP) and customers.SLA is also known as an operating level agreement (OLA) when used in an organization without an established or formal provider-customer relationship.It Adopted in the late 1980s, SLAs are currently used by most industries and markets. By nature, SLAs define service output but defer methodology to the service provider's discretion. Specific metrics vary by industry and SLA purpose. A service level agreement (SLA) is a contract between a business and its customer outlining the details that the two parties have agreed to in a transaction. The types of SLAs that an organization can use depends on many significant aspects. While some are targeted at individual customer groups, others discuss issues relevant to entire companies. This is because the needs of one user differ from those of another. Below is a list of the types of SLAs used by businesses today, and how each one is utilized for specific situations A Service Level Agreement (SLA) is the bond for performance negotiated between the cloud services provider and the client. Earlier, in cloud computing all Service Level Agreements were negotiated between a client and the service consumer. Nowadays, with the initiation of large utility-like cloud computing providers, most Service Level Agreements are standardized until a client becomes a large consumer of cloud services

SLAs features include:

- Specific details and scope of provided services, including priorities, responsibilities and guarantees
- Specific, expected and measurable services at minimum or target levels
- Informal or legally binding
- Descriptive tracking and reporting guidelines
- Detailed problem management procedures
- Detailed fees and expenses
- Customer duties and responsibilities
- Disaster recovery procedures
- Agreement termination clauses

Types of SLA

1 Customer-based SLA
2 Service-based SLA
3 Multilevel SLA

Customer-based SLA

This type of agreement is used for individual customers and comprises all relevant services that a client may need, while leveraging only one contract. It contains details regarding the type and quality of service that has been agreed upon. For example, a telecommunication service includes voice calls, messaging and internet services, but that all exists under a single contract.

Service-based SLA

This SLA is a contract that includes one identical type of service for all of its customers. Because the service is limited to one unchanging standard, it is more straightforward and convenient for vendors. For example,

using a service-based agreement regarding an IT helpdesk would mean that the same service is valid for all end-users that sign the service-based SLA.

Multi-level SLA

This agreement is customized according to the needs of the end-user company. It allows the user to integrate several conditions into the same system to create a more suitable service. It addresses contracts at the following levels:

Few Service Level Agreements are enforceable as contracts, but mostly are agreements or contracts which are more along the lines of an Operating Level Agreement (OLA) and may not have the restriction of law. It is fine to have an attorney review the documents before making a major agreement to the cloud service provider.
Service Level Agreements usually specify some parameters which are mentioned below:

1. Availability of the Service (uptime)
2. Latency or the response time
3. Service components reliability
4. Each party accountability
5. Warranties

In any case, if a cloud service provider fails to meet the stated targets of minimums then the provider has to pay the penalty to the cloud service consumer as per the agreement. So, Service Level Agreements are like insurance policies in which the corporation has to pay as per the agreements if any casualty occurs.
Microsoft publishes the Service Level Agreements linked with the Windows Azure Platform components, which is demonstrative of industry practice for cloud service vendors. Each individual component has its own Service Level Agreements. Below are two major Service Level Agreements (SLA) described:

1. Windows Azure SLA –
   Window Azure has different SLA's for compute and storage. For compute, there is a guarantee that when a client deploys two or more role instances in separate fault and upgrade domains, client's internet facing roles will have external connectivity minimum 99.95% of the time. Moreover, all of the role instances of the client are monitored and there is guarantee of detection 99.9% of the time when a role instance's process is not runs and initiates properly.
2. SQL Azure SLA –
   SQL Azure clients will have connectivity between the database and internet gateway of SQL Azure. SQL Azure will handle a "Monthly Availability" of 99.9% within a month. Monthly Availability Proportion for a particular tenant database is the ratio of the time the database was available to customers to the total time in a month. Time is measured in some intervals of minutes in a 30-day monthly cycle. Availability is always remunerated for a complete month. A portion of time is marked as unavailable if the customer's attempts to connect to a database are denied by the SQL Azure gateway.

Service Level Agreements are based on the usage model. Frequently, cloud providers charge their pay-as-per-use resources at a premium and deploy standards Service Level Agreements only for that purpose. Clients can also subscribe at different levels that guarantees access to a particular amount of purchased resources. The Service Level Agreements (SLAs) attached to a subscription many times offer various terms and conditions. If client requires access to a particular level of resources, then the client need to subscribe to a service. A usage model may not deliver that level of access under peak load condition.

Operational Level Agreement (OLA)

Unlike SLA it is agreement within the organization.

Here are the objectives of SLM −

1	Define, document, agree, monitor, measure, report, and review the level of IT service provided.
2	Provide and improve the relationship and communication with the business and customers.

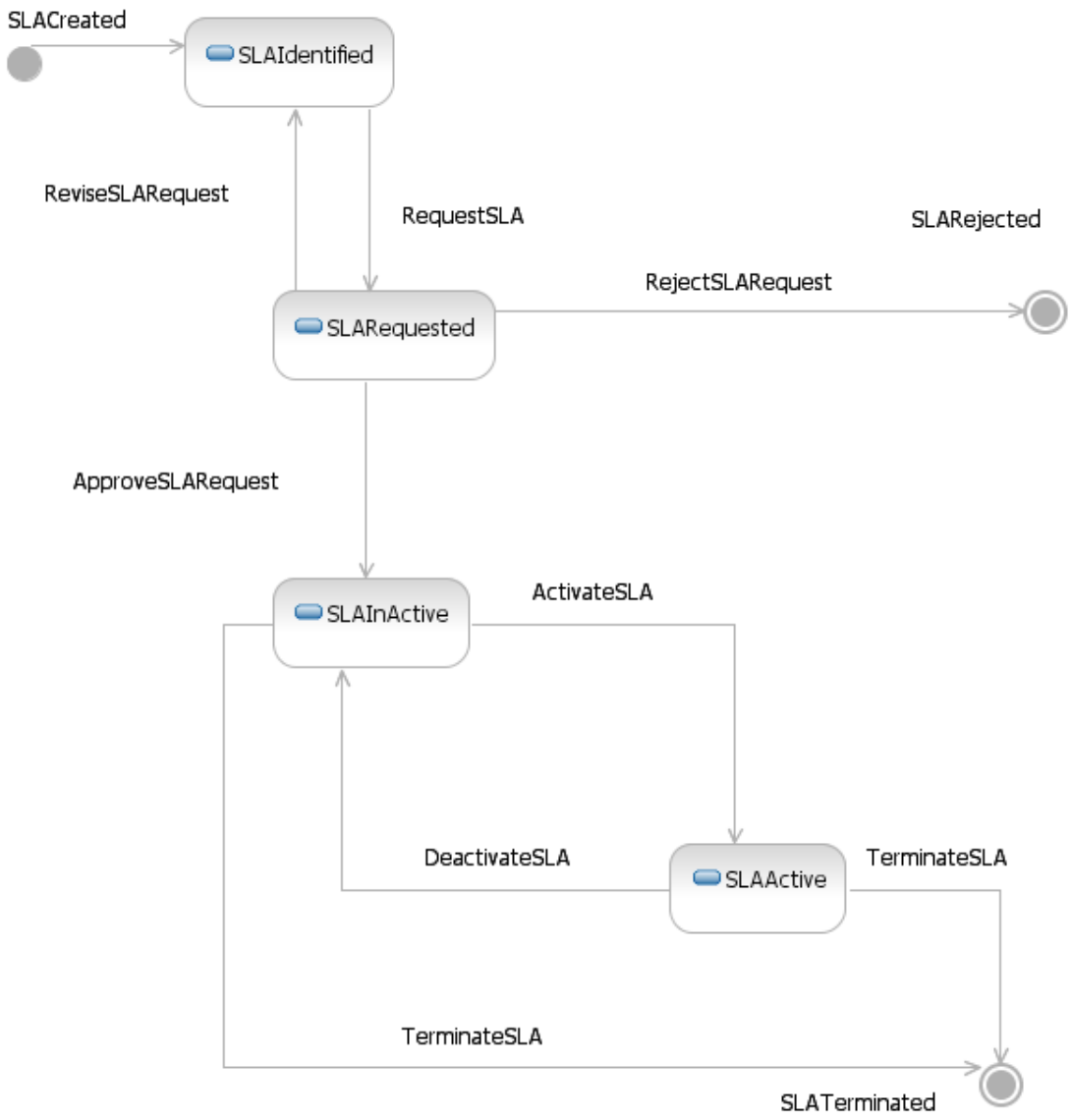
3	Ensure that specific and measurable targets are developed for all IT services.
4	Monitor and improve customer satisfaction with the quality of service delivered
5	Ensure that IT and customers have a clear and unambiguous expectation of the level of service to be delivered

The following table describes the states of the service level agreement lifecycle, and, for each state, names the transition that moves a service level agreement forward to that state.

Service level agreement lifecycle.

Transition	State	Description
(Initial state)	SLA identified	This state is entered as soon as a consumer, represented by a capability version, requests a dependency on a service version or other capability version that offers the service level definition (SLD) that they require.
Request SLA	SLA requested	The agreed endpoints relationship target has been selected together with details of the required SLA properties and policies. The provider of the selected SLD must approve the request, reject it or ask for it to be revised.
Approve SLA request	SLA inactive	The development team that want to consume the service can continue their development based on the consumption of this specific SLA, but they do not yet have authorization to access any endpoints.
Revise SLA request	SLA identified	As part of the negotiation of an SLA, the service provider requests a rework of the details of the SLA by the service consumer. This is done by moving the SLA back into the identified state, ready for a resubmission.
Activate SLA	SLA active	All the approved endpoints associated with the SLD, that are online, can be invoked using the terms of the SLA. There might be situations where the SLA is deactivated, in which case the SLA enters the SLA inactive state and any further interactions are blocked until it is reactivated.
Deactivate SLA	SLA inactive	For operational issues, the SLA is temporarily suspended by moving it back to the inactive state. Once the operational issues have been removed, the SLA can be reactivated.
Terminate SLA	SLA terminated	No interactions from this SLA are permitted.

Diagram of the service level agreement lifecycle



Unit 4
Virtualization Concepts

Virtualization and Cloud Computing

Virtualization is the key to unlock the Cloud system, what makes virtualization so important for the cloud is that it decouples the software from the hardware. For example, PC's can use virtual memory to borrow extra memory from the hard disk. Usually hard disk has a lot more space than memory. Although virtual disks are

slower than real memory, if managed properly the substitution works perfectly. Likewise, there is software which can imitate an entire computer, which means 1 computer can perform the functions equals to 20 computers.Virtualization as a concept has existed since the 1960s, IBM Cambridge Scientific Center began the development of CP-40, and a virtual time bound operating system. It was not till 2000 cloud computing virtualization started to be incorporated for better utilization of the resources. Virtualization in the world of cloud computing is considered a scalable technology that will transform the IT industry The main enabling technology for Cloud Computing is Virtualization. Virtualization is a partitioning of single physical server into multiple logical servers. Once the physical server is divided, each logical server behaves like a physical server and can run an operating system and applications independently. Many popular companies's like VmWare and Microsoft provide virtualization services, where instead of using your personal PC for storage and computation, you use their virtual server. They are fast, cost-effective and less time consuming. Virtualization in cloud computing allows the user to run multiple applications and operating systems on the same system. For enterprises that tend to focus on better utilization of resources that will be leading to increases in efficiency and reduction in cost, the virtual machine offers an opportunity. Virtualization is a process of creating a better environment for the user with an existing server with a required set of programs. Using virtualization enterprises can provide various services through the same server for all users.The virtual environment is an underlying platform that can support the combination of different operating systems, networks or application servers. Virtualization can also be used to create virtual storage devices and different hardware equipment's. The most common example of Virtualization is a partition of your hard drive; though it's a single hardware a user can access each of them differently without affecting the other drives.For software developers and testers virtualization comes very handy, as it allows developer to write code that runs in many different environments and more importantly to test that code.

Virtualization is mainly used for three main purposes 1) Network Virtualization 2) Server Virtualization 3) Storage Virtualization

A)Network Virtualization: It is a method of combining the available resources in a network by splitting up the available bandwidth into channels, each of which is independent from the others and each channel is independent of others and can be assigned to a specific server or device in real time.

B)Storage Virtualization: It is the pooling of physical storage from multiple network storage devices into what appears to be a single storage device that is managed from a central console. Storage virtualization is commonly used in storage area networks (SANs).

C)Server Virtualization: Server virtualization is the masking of server resources like processors, RAM, operating system etc, from server users. The intention of server virtualization is to increase the resource sharing and reduce the burden and complexity of computation from users.

.

What are the various types of virtualization and how do they work will help in understanding modern virtualization that affects the enterprises?

A)HARDWARE VIRTUALIZATION:

Server virtualization or hardware-assisted virtualization is also known as hardware virtualization. The underlying concept that is applied in the hardware virtualization is the creation of multiple virtual servers that run on a single physical server. The virtual server can function as an independent machine serving totally different purposes with the different operating system. Hardware virtualization is an abstract computing platform that also hides the physical characteristics of the system. Hypervisor acts as a support technology that handles the resources allotment for hardware virtualization.

The main advantage of the hardware virtualization is maximized utilization of the hardware and application uptime through servers.

TYPES OF HARDWARE VIRTUALIZATION

1. Full Virtualization: The total underlying software is virtualized and doesn't require any modifications to support the system.
2. Emulation Virtualization: The virtual machine simulates the hardware and becomes independent of the machine. The guest operating system doesn't require any modifications to support the system.
3. Para-virtualization: The guest software runs on the isolated domain and the hardware is not simulated. The virtual machine offers a special API that can be used for modifying the guest operating system.

B)SOFTWARE VIRTUALIZATION:

Software virtualization revolves on the primary concept that develops multiple virtual environments on the server. Using the software virtualization an enterprise can create a complete environment and supported by the hardware. A user can run different operating systems using any one of the operating systems for the underlying technology support.

TYPES OF THE SOFTWARE VIRTUALIZATION TO SUPPORT THE SYSTEM:

1. Application Virtualization: Hosting an individual application in a virtual environment that can be managed separately with the underlying native operating system.
2. Service Virtualization: A particular application can host on the server for specific purpose and service.
3. Operating System Virtualization: Hosting multiple operating systems on the native operating system.

BENEFITS OF VIRTUALIZATION:

Virtualization will improve disaster recovery, software testing, and workload distribution. This reduces the hardware costs, saves energy and reducing the physical size of the data storage.

Virtualization improves the flexibility that enables the implementation of Infrastructure as a Service (IaaS) for better utilization of resources. Using the IaaS we can virtualize the hardware, software, servers, storage, and other infrastructure components. Using the virtualization a user can create a scalable model without over-loading the infrastructure. Creation of temporary or experimental virtual workload management can help us bring changes into management without affecting the current operations.

Effective management of resources for better productivity of the employees the enterprise can reduce the risk of data loss. Virtualization reduces the risk of data access, eliminating the need for an extra layer of security.

Virtualization is an imperative concept for the enterprises that seek to turnaround the resource utilization. With implementation, resource utilization and usage challenge have been simplified with virtual machines that can be scaled according to the requirements. To know more about the recent development of Virtualization

Hypervisor

Hypervisor is a form of virtualization software used in Cloud hosting to divide and allocate the resources on various pieces of hardware.The program which provide partitioning, isolation or abstraction is called virtualization hypervisor. Hypervisor is a hardware virtualization technique that allows multiple guest operating systems (OS) to run on a single host system at the same time. A hypervisor is sometimes also called a virtual machine manager(VMM).

Types of Hypervisor –

TYPE-1 Hypervisor:

Hypervisor runs directly on underlying host system.It is also known as “Native Hypervisor” or “Bare metal hypervisor”.It dose not require any base server operating system.It has direct access to hardware resources.Examples of Type 1 hypervisors include VMware ESXi, Citrix XenServer and Microsoft Hyper-V hypervisor.

TYPE-2 Hypervisor:

A hypervisor (or virtual machine monitor, VMM, virtualizer) 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.
A Host operating system runs on underlying host system. It is also known as ‘Hosted Hypervisor”. Basically a software installed on an operating system. Hypervisor asks operating system to make hardware calls. Example of Type 2 hypervisor include VMware Player or Parallels Desktop. Hosted hypervisors are often found on endpoints like PCs.

The following factors should be examined before choosing a suitable hypervisor:

1. Understand your needs: The company and its applications are the reason for the data center (and your job). Besides your company’s needs, you (and your co-workers in IT) also have your own needs.Needs for a virtualization hypervisor are:
a. Flexibility
b. Scalability
c. Usability
d. Availability

e. Reliability
f. Efficiency
g. Reliable support

2. The cost of a hypervisor: For many buyers, the toughest part of choosing a hypervisor is striking the right balance between cost and functionality. While a number of entry-level solutions are free, or practically free, the prices at the opposite end of the market can be staggering. Licensing frameworks also vary, so it's important to be aware of exactly what you're getting for your money.
3. Virtual machine performance: Virtual systems should meet or exceed the performance of their physical counterparts, at least in relation to the applications within each server. Everything beyond meeting this benchmark is profit.
4. Ecosystem: It's tempting to overlook the role of a hypervisor's ecosystem – that is, the availability of documentation, support, training, third-party developers and consultancies, and so on – in determining whether or not a solution is cost-effective in the long term.
5. Test for yourself: You can gain basic experience from your existing desktop or laptop. You can run both VMware vSphere and Microsoft Hyper-V in either VMware Workstation or VMware Fusion to create a nice virtual learning and testing environment.

Unit 5
Cloud Security

Cloud security is the technology and best practices designed to protect data and information within a cloud architecture. It's a critical component of any IT infrastructure strategy that uses the cloud. Cloud security ensures data privacy and compliance around data stored in the cloud.More and more businesses are taking advantage of the cloud for at least part, if not all of their IT infrastructure. Private clouds, public

clouds and hybrid cloud combinations of both private and public cloud platforms have all grown in popularity, as well as the use of multiple public clouds in a multi-cloud strategy. Because of cloud computing's distributed and dynamic nature, there are unique considerations when it comes to securing data within the cloud Security in cloud computing is a major concern. Data in cloud should be stored in encrypted form. To restrict client from accessing the shared data directly, proxy and brokerage services should be employed. Since all the data is transferred using Internet, data security is of major concern in the cloud. Here are key mechanisms for protecting data.

- Access Control
- Auditing
- Authentication
- Authorization

Security Planning

Before deploying a particular resource to cloud, one should need to analyze several aspects of the resource such as:

- Select resource that needs to move to the cloud and analyze its sensitivity to risk.
- Consider cloud service models such as IaaS, PaaS, and SaaS. These models require customer to be responsible for security at different levels of service.
- Consider the cloud type to be used such as public, private, community or hybrid.
- Understand the cloud service provider's system about data storage and its transfer into and out of the cloud.

Security for Cloud Computing

While using cloud computing, the major issue that concerns the users is about its security. Some countries government may decide to search through data without necessarily notifying the data owner, depending on where the data resides, which is not appreciated and is considered as a privacy breach (Example Prism Program by USA).To provide security for systems, networks and data cloud computing service providers have joined hands with TCG (Trusted Computing Group) which is non-profit organization which regularly releases a set of specifications to secure hardware, create self-encrypting drives and improve network security. It protects the data from root kits and malware. As computing has expanded to different devices like hard disk drives and mobile phones, TCG has extended the security measures to include these devices. It provides ability to create a unified data protection policy across all clouds.Some of the trusted cloud services are Amazon, Box.net, Gmail and many others.

Privacy Concern & Cloud Computing

Privacy present a strong barrier for users to adapt into Cloud Computing systems

There are certain measures which can improve privacy in cloud computing.

1. The administrative staff of the cloud computing service could theoretically monitor the data moving in memory before it is stored in disk.To keep the confidentiality of a data, administrative and legal controls should prevent this from happening.
2. The other way for increasing the privacy is to keep the data encrypted at the cloud storage site, preventing unauthorized access through the internet; even cloud vendor can't access the data either.

Legal Issues in Cloud Computing

With the explosive growth of innovations in the Information Technology industry, the Legal provisions are currently lagging behind and desperately looking for ways to cope up with the never-seen-before advancements.

Cloud computing, being one of such recent advancements, have raised a number of legal issues including privacy and data security, contracting issues, issues relating to the location of the data, and business considerations.

The abovementioned issues are the primary ones faced by almost all the nations across the globe. However, when it comes to the Indian scenario, a number of additional complicated issues are faced by India owing to lack of awareness and lack of resources. With the 'Digital India' initiative in the news, it is obvious that more and more individuals and organisations will be using online services and infrastructure via the Cloud in the near future; and it is, therefore, necessary to analyse our position thereon and discuss whether our legal system is ready for such a revolutionary change.

The legal issues that frequently arise in the cloud are wide-ranging. However, attempting a broad generalisation, mainly four types of issues arise therein:

1. Privacy of data and data security
2. Issues relating to contractual relation between the cloud service provider and the customer
3. Complex jurisdictional issues, or issues relating to the location of the data and the set of laws applicable
4. Commercial as well as business considerations

Privacy and data security issues:

Seemingly, the main privacy/data security issue relating to the cloud is 'data breach'. Data breach may be in the generic sense defined as the loss of unencrypted electronically stored personal information (Buyya, Broberg, & Goscinski, 2015). A data breach can cause loss to both the provider as well as the customer in numerous ways; with identity theft and chances of debit/credit card fraud to the customer, and financial harm, loss of customer, loss of reputation, potential lawsuits et cetera for the provider.The American law requires data breach notification to be issued of affected persons in such case of a data breach. Almost all the states in the United States now require notification of affected persons upon the occurrence of a data breach.Talking about the Indian scenario, most of the providers are seen to attempt at lessening their risk liability in case of a data breach scenario. However, as more sensitive information is entering the cloud every passing day, businesses and corporations have started negotiating the contracts so as to insert terms that expand the contractual obligations of the providers.Problem arises when the data is subject to more than one jurisdictions, and the jurisdictions have different laws regarding data privacy. For example, the European Union Data Privacy Directive clearly states that 'Data cannot leave the EU unless it goes to a country that ensures an "adequate level of protection".' Now, although such statement makes the EU provisions easily enforceable, but it restricts the data movement thereby reducing the data efficiency.

1.Contracting Issues:

Clearly, licensing agreements are fundamentally different from Service agreements. Cloud essentially, in all its permutations (IaaS, PaaS, SaaS), is a service, and therefore is governed by a Service agreement instead of a Licensing agreement.

However, the main issue regarding the Cloud Service agreements is 'contract of adhesion'. Owing to the limited expansion of Cloud Services in India, most of the time the 'Click-wrap agreement' model is used, causing the contract to be one of the contract of adhesion. It leaves no or little scope for negotiation on the part of the user/customer.

With the expansion of the Cloud computing, gradually the negotiation power of the large corporation will cause the Cloud Contracts to be standard and negotiated ones. However, at an individual level, this is still a far destination.Legal provisions clearly cannot force the cloud providers to have a negotiating session with each and every customer. However, legal provisions may be made to ensure that the liability and risk responsibility clauses follow a standard pattern which compensates the user for the lack of negotiation during the formation of the contract.

2.Jurisdictional Issues:

Jurisdiction is the authority of a court to judge acts committed in a certain territory. Jurisdiction in case of legal issues relating to the Cloud services becomes difficult and critical because of the features of Cloud like 'Virtualization', and 'Multi-tenancy'.While virtualization ensures the requirement of less hardware and consumption of less power thereby ensuring computing efficiency, it also on the other hand makes it difficult for the cloud user or the cloud provider to know what information is housed on various machines at any given time.Multi-tenancy refers to the ability of a cloud provider to deliver services to many individuals or organisations from a single shared software. The risk with this is that it makes it highly possible that the data of one user may be accessed in an unauthorised manner by another user since the data of various users are only virtually separated and not physically. Also, it makes it difficult to back up and restore data.The cloud enables a great deal of flexibility in data location, which ensures maximum efficiency in data usage and accessibility. However, it creates a number of legal issues as well. It makes it quite possible a scenario that the same data may be stored in multiple locations at a given time. Now, if the multiple locations are subject to different jurisdiction and different legal system, there arises a possibility that there may be conflicting legal provisions regarding data in the two aforementioned different locations. This gives rise to most of the jurisdictional issues in Cloud computing.Also, laws relating to confidentiality and Government access to data are different across different nations. While the Indian laws manage to strike a balance between national security and individual privacy, most of the nations do not prefer a balance and have adopted a biased view on this. Problem of conflict of laws arises herein, in such cases.

3.Commercial and Business Considerations:

Other commercial and business considerations like the urge to minimize risk, maintain data integrity, accessibility and availability of data as well as Service level Agreements have also significantly shaped the present as well as future of Cloud Computing in India. It also creates a number of foreseeable as well as unforeseeable issues that needs to be addressed by dedicated legislations therefore.It is an accepted truth that Law always lags behind technical innovations, and the complexities of the Cloud innovations and related Cloud Services like Software as a Service (SaaS), Platform as a Service (PaaS), and Infrastructure as a Service (IaaS) will force the law and legislations to catch up in order for an effective legal system that provides legal remedies to prevent and redress the resultant harms.Raising awareness, ensuring universal access to information, and resource mobilizing are complimentary solutions that'll never go wrong for the Indian scenario in order to add to the effectiveness of an effective legal system.

Unit 6

Cloud Storage

Cloud storage is a cloud computing model that stores data on the Internet through a cloud computing provider who manages and operates data storage as a service. It's delivered on demand with just-in-time capacity and costs, and eliminates buying and managing your own data storage infrastructure.

- Unlimited storage with no minimum object size
- Worldwide accessibility and worldwide storage locations
- Low latency (time to first byte is typically tens of milliseconds)
- High durability
- Geo-redundancy if the data is stored in a multi-region or dual-region

Advantages of Cloud Storage

1. CostPurchasing physical storage can be expensive. Without the need for hardware cloud storage is exceptionally cheaper per GB than using external drives.
2. Accessibility
   Using the cloud for storage gives you access to your files from anywhere that has an internet connection.
3. Recovery
   In the event of a hard drive failure or other hardware malfunction, you can access your files on the cloud. It acts as a backup solution for your local storage on physical drives.
4. Syncing and Updating
   When you are working with cloud storage, every time you make changes to a file it will be synced and updated across all of your devices that you access the cloud from.
5. Security
   Cloud storage providers add additional layers of security to their services. Since there are many people with files stored on the cloud, these providers go to added lengths to make sure your files don't get accessed by someone who shouldn't

Disadvantages of Cloud Storage

1. Internet Connection
   Cloud based storage is dependent on having an internet connection. If you are on a slow network you may have issues accessing your storage. In the event you find yourself somewhere without internet, you won't be able to access your files.
2. Costs
   There are additional costs for uploading and downloading files from the cloud. These can quickly add up if you are trying to access lots of files often.
3. Hard Drives
   Cloud storage is supposed to eliminate our dependency on hard drives right? Well some business cloud storage providers require physical hard drives as well.
4. Support
   Support for cloud storage isn't the best, especially if you are using a free version of a cloud provider. Many providers refer you to a knowledge base or FAQs.
5. Privacy
   When you use a cloud provider, your data is no longer on your physical storage. So who is responsible for making sure that data is secure? That's a gray area that is still being figured out.

Storage as a service (SaaS) is a business model in which a company leases or rents its storage infrastructure to another company or individuals to store data. Small companies and individuals often find this to be a convenient methodology for managing backups, and providing cost savings in personnel, hardware and physical space.A company providing SaaS may be called a storage service provider (SSP). Storage as a service can also be referred to as hosted storage.Storage as a Service (SaaS)As an alternative to storing magnetic tapes

offsite in a vault, IT administrators are meeting their storage and backup needs by service level agreements (SLAs) with an SaaS provider, usually on a cost-per-gigabyte-stored and cost-per-data-transferred basis. The client transfers the data meant for storage to the service provider on a set schedule over the SaaS provider's wide area network or over the Internet. The storage provider provides the client with the software required to access their stored data. Clients use the software to perform standard tasks associated with storage, including data transfers and data backups. Corrupted or lost company data can easily be restored.Storage as a service is prevalent among small to mid-sized businesses, as no initial budget is required to set up hard drives, servers and IT staff. SaaS is also marketed as an excellent technique to mitigate risks in disaster recovery by providing long-term data storage and enhancing business stability.

Storage area networks (SANs)

Storage area networks (SANs) are the most common storage networking architecture used by enterprises for business-critical applications that need to deliver high throughput and low latency. A rapidly growing portion of SAN deployments leverages all-flash storage to gain its high performance, consistent low latency, and lower total cost when compared to spinning disk. By storing data in centralized shared storage, SANs enable organizations to apply consistent methodologies and tools for security, data protection, and disaster recovery.A SAN is block-based storage, leveraging a high-speed architecture that connects servers to their logical disk units (LUNs). A LUN is a range of blocks provisioned from a pool of shared storage and presented to the server as a logical disk. The server partitions and formats those blocks—typically with a file system—so that it can store data on the LUN just as it would on local disk storage.SANs make up about two-thirds of the total networked storage market. They are designed to remove single points of failure, making SANs highly available and resilient. A well-designed SAN can easily withstand multiple component or device failures.

Types of SAN

The most common SAN protocols are:

- Fibre Channel Protocol (FCP). The most widely used SAN or block protocol, deployed in 70% to 80% of the total SAN market. FCP uses Fibre Channel transport protocols with embedded SCSI commands.
- Internet Small Computer System Interface (iSCSI). The next largest SAN or block protocol, with approximately 10% to 15% of the market. iSCSI encapsulates SCSI commands inside an Ethernet frame and then uses an IP Ethernet network for transport.
- Fibre Channel over Ethernet (FCoE). FCoE is less than 5% of the SAN market place. It is similar to iSCSI, since it encapsulates an FC frame inside an Ethernet datagram. Then like iSCSI, it uses an IP Ethernet network for transport.
- Non-Volatile Memory Express over Fibre Channel (FC-NVMe). NVMe is an interface protocol for accessing flash storage via a PCI Express (PCIe) bus. Unlike traditional all-flash architectures, which are limited to a single, serial command queue, NVMe supports tens of thousands of parallel queues, each with the ability to support tens of thousands of concurrent commands.

Unit 7

Scheduling in Cloud

In Computing, scheduling is the method by which work is assigned to resources that complete the work. The work may be virtual computation elements such as threads, processes or data flows, which are in turn scheduled onto hardware resources such as processors, network links or expansion cards.A scheduler is what carries out the scheduling activity. Schedulers are often implemented so they keep all computer resources busy (as in load balancing), allow multiple users to share system resources effectively, or to achieve a target quality of service. Scheduling is fundamental to computation itself, and an intrinsic part of the execution model of a computer system; the concept of scheduling makes it possible to have computer multitasking with a single central processing unit (CPU) A scheduler may aim at one or more goals, for example: maximizing throughput (the total amount of work completed per time unit); minimizing wait time (time from work becoming ready until the first point it begins execution); minimizing latency or response time .In real-time environments, such as embedded systems for automatic control in industry (for example robotics), the scheduler also must ensure that processes can meet deadlines; this is crucial for keeping the system stable. Scheduled tasks can also be distributed to remote devices across a network and managed through an administrative back end.

Scheduling disciplines

Scheduling disciplines are algorithms used for distributing resources among parties which simultaneously and asynchronously request them. Scheduling disciplines are used in routers (to handle packet traffic) as well as in operating systems (to share CPU time among both threads and processes), disk drives (I/O scheduling), printers (print spooler), most embedded systems, etc.The main purposes of scheduling algorithms are to minimize resource starvation and to ensure fairness amongst the parties utilizing the resources. Scheduling deals with the problem of deciding which of the outstanding requests is to be allocated resources. There are many different scheduling algorithms.

A) First in, first out (FIFO), also known as first come, first served (FCFS), is the simplest scheduling algorithm. FIFO simply queues processes in the order that they arrive in the ready queue. This is commonly used for a task queue, for example as illustrated in this section.

- Since context switches only occur upon process termination, and no reorganization of the process queue is required, scheduling overhead is minimal.
- Throughput can be low, because long processes can be holding the CPU, causing the short processes to wait for a long time (known as the convoy effect).
- No starvation, because each process gets chance to be executed after a definite time.
- Turnaround time, waiting time and response time depends on the order of their arrival and can be high for the same reasons above.
- No prioritization occurs, thus this system has trouble meeting process deadlines.
- The lack of prioritization means that as long as every process eventually completes, there is no starvation. In an environment where some processes might not complete, there can be starvation.
- It is based on queuing.

B)Priority scheduling

Earliest deadline first (EDF) or least time to go is a dynamic scheduling algorithm used in real-time operating systems to place processes in a priority queue. Whenever a scheduling event occurs (a task finishes, new task is released, etc.), the queue will be searched for the process closest to its deadline, which will be the next to be scheduled for execution.

C)Shortest remaining time first

Similar to shortest job first (SJF). With this strategy the scheduler arranges processes with the least estimated processing time remaining to be next in the queue. This requires advanced knowledge or estimations about the time required for a process to complete.

- If a shorter process arrives during another process' execution, the currently running process is interrupted (known as preemption), dividing that process into two separate computing blocks. This creates excess overhead through additional context switching. The scheduler must also place each incoming process into a specific place in the queue, creating additional overhead.
- This algorithm is designed for maximum throughput in most scenarios.
- Waiting time and response time increase as the process's computational requirements increase. Since turnaround time is based on waiting time plus processing time, longer processes are significantly affected by this. Overall waiting time is smaller than FIFO, however since no process has to wait for the termination of the longest process.
- No particular attention is given to deadlines, the programmer can only attempt to make processes with deadlines as short as possible.
- Starvation is possible, especially in a busy system with many small processes being run.
- To use this policy we should have at least two processes of different priority

D)Fixed priority pre-emptive scheduling

The operating system assigns a fixed priority rank to every process, and the scheduler arranges the processes in the ready queue in order of their priority. Lower-priority processes get interrupted by incoming higher-priority processes.

- Overhead is not minimal, nor is it significant.
- FPPS has no particular advantage in terms of throughput over FIFO scheduling.
- If the number of rankings is limited, it can be characterized as a collection of FIFO queues, one for each priority ranking. Processes in lower-priority queues are selected only when all of the higher-priority queues are empty.
- Waiting time and response time depend on the priority of the process. Higher-priority processes have smaller waiting and response times.
- Deadlines can be met by giving processes with deadlines a higher priority.
- Starvation of lower-priority processes is possible with large numbers of high-priority processes queuing for CPU time.

E) Round-robin scheduling

The scheduler assigns a fixed time unit per process, and cycles through them. If process completes within that time-slice it gets terminated otherwise it is rescheduled after giving a chance to all other processes.

- RR scheduling involves extensive overhead, especially with a small time unit.
- Balanced throughput between FCFS/ FIFO and SJF/SRTF, shorter jobs are completed faster than in FIFO and longer processes are completed faster than in SJF.
- Good average response time, waiting time is dependent on number of processes, and not average process length.
- Because of high waiting times, deadlines are rarely met in a pure RR system.
- Starvation can never occur, since no priority is given. Order of time unit allocation is based upon process arrival time, similar to FIFO.
- If Time-Slice is large it becomes FCFS /FIFO or if it is short then it becomes SJF/SRTF.

F)Multilevel queue scheduling

This is used for situations in which processes are easily divided into different groups. For example, a common division is made between foreground (interactive) processes and background (batch) processes. These two types of processes have different response-time requirements and so may have different scheduling needs. It is very useful for shared memory problems.

.

Scheduling problems

There are several scheduling problems in which the goal is to decide which job goes to which station at what time

- Job shop scheduling – there are n jobs and m identical stations. Each job should be executed on a single machine. This is usually regarded as an online problem.
- Open-shop scheduling – there are n jobs and m different stations. Each job should spend some time at each station, in a free order.
- Flow shop scheduling – there are n jobs and m different stations. Each job should spend some time at each station, in a pre-determined order.