**Govt. Polytechnic, Loharu**

**E-Contents**

**SUBJECT: Cloud Computing**

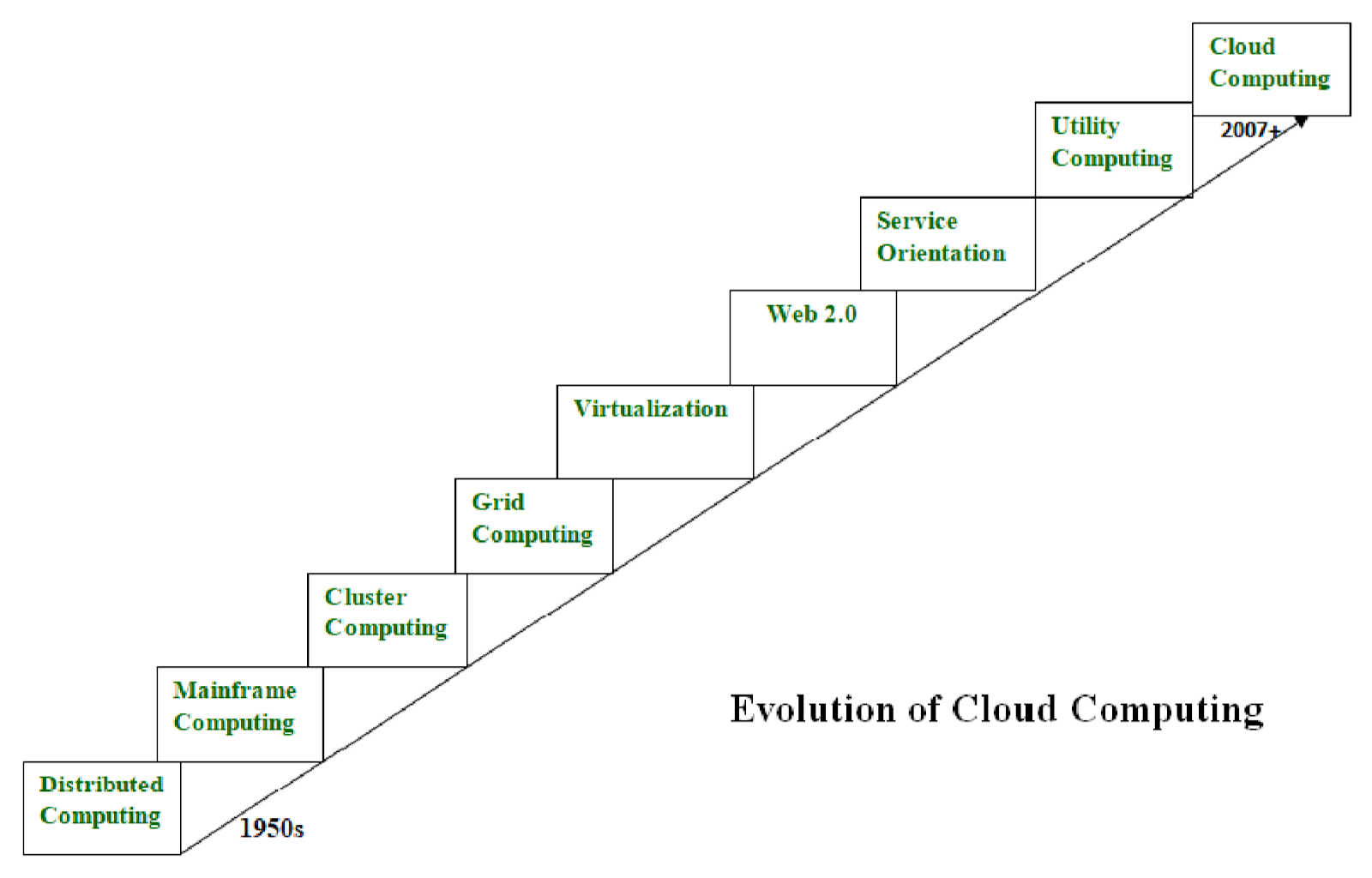
**BRANCH: Computer Engineering**

**SEM: 5th**

**Unit 1 Introduction**

**Evolution of Cloud Computing**

Cloud computing is all about renting computing services. This idea first came in the 1950s. In making cloud computing what it is today, the various technologies played a vital role. These are as shown below



1. **Distributed Systems:**

It is a composition of multiple independent systems but all of them are depicted as a single entity to the users. The purpose of distributed systems is to share resources and also use them effectively and efficiently. Distributed systems possess characteristics such as scalability, concurrency, continuous availability, heterogeneity, and independence in failures. But the main problem with this system was that all the systems were required to be present at the same geographical location. Thus to solve this problem, distributed computing led to three more types of computing and they wereMainframe computing, cluster computing, and grid computing.

1. **Mainframe computing:**

Mainframes which first came into existence in 1951 are highly powerful and reliable computing machines. These are responsible for handling large data such as massive input-output operations. Even today these are used for bulk processing tasks such as online transactions etc. These systems have almost no downtime with high fault tolerance. After distributed computing, these increased the processing capabilities of the system. But these were very expensive. To reduce this cost, cluster computing came as an alternative to mainframe technology.

1. **Cluster computing:**

In 1980s, cluster computing came as an alternative to mainframe computing. Each machine in the cluster was connected to each other by a network with high bandwidth. These were cheaper than those mainframe systems. These were equally capable of high computations. Also, new nodes could easily be added to the cluster if it was required. Thus, the problem of the cost was solved to some extent but the problem related to geographical restrictions still pertained. To solve this, the concept of grid computing was introduced.

1. **Grid computing:**

In 1990s, the concept of grid computing was introduced. It means that different systems were placed at entirely different geographical locations and these all were connected via the internet. These systems belonged to different organizations and thus the grid consisted of heterogeneous nodes. Although it solved some problems but new problems emerged as the distance between the nodes increased. The main problem which was encountered was the low availability of high bandwidth connectivity and with it other network associated issues. Thus cloud computing is often referred to as “Successor of grid computing”.

1. **Virtualization:**

It was introduced nearly 40 years back. It refers to the process of creating a virtual layer over the hardware which allows the user to run multiple instances simultaneously on the hardware. It is a key technology used in cloud computing. It is the base on which major cloud computing services such as Amazon EC2, VMware vCloud, etc work on. Hardware virtualization is still one of the most common types of virtualization.

**6     Web 2.0:**

It is the interface through which the cloud computing services interact with the clients. It is because of Web 2.0 that we have interactive and dynamic web pages. It also increases flexibility among web pages. Popular examples of web 2.0 include Google Maps, Facebook, Twitter, etc. Needless to say, social media is possible because of this technology only. In gained major popularity in 2004.

**7    Service orientation:**

It acts as a reference model for cloud computing. It supports low-cost, flexible, and evolvable applications. Two important concepts were introduced in this computing model. These were Quality of Service (QoS) which also includes the SLA (Service Level Agreement) and Software as a Service (SaaS).

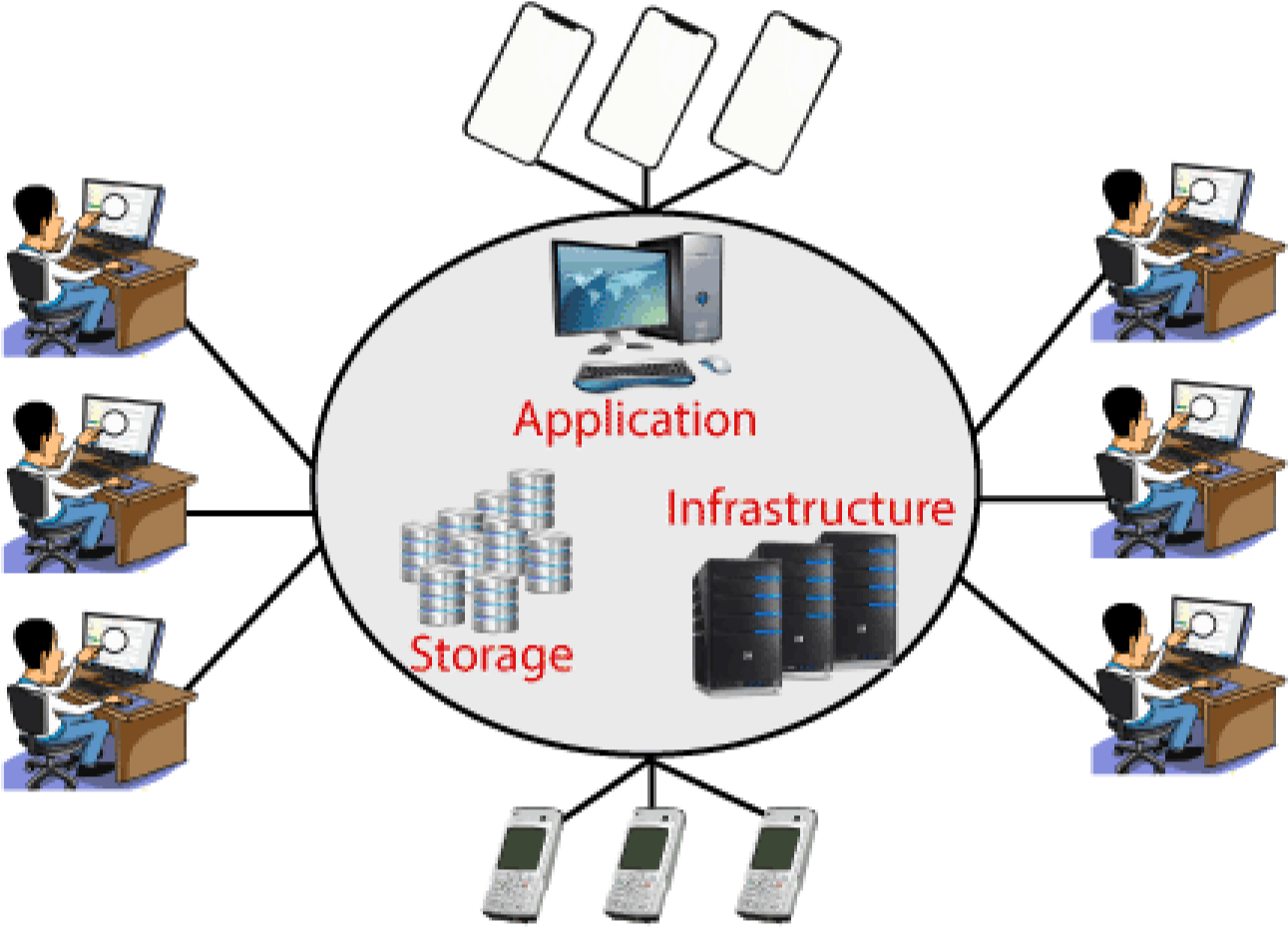
**8    Utility computing:**

It is a computing model that defines service provisioning techniques for services such as compute services along with other major services such as storage, infrastructure, etc which are provisioned on a pay-per-use basis.

Thus, the above technologies contributed to the making of cloud computing.

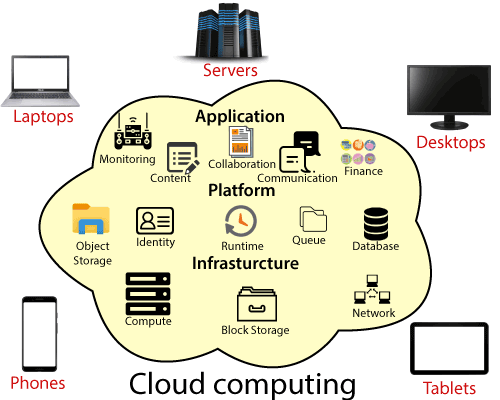
**Cloud Computing Overview**

Cloud Computing is the delivery of computing services such as servers, storage, databases, networking, software, analytics, intelligence, and more, over the Cloud (Internet).



Cloud Computing provides an alternative to the on-premises datacentre. With an onpremises datacentre, we have to manage everything, such as purchasing and installing hardware, virtualization, installing the operating system, and any other required applications, setting up the network, configuring the firewall, and setting up storage for data. After doing all the set-up, we become responsible for maintaining it through its entire lifecycle.

But if we choose Cloud Computing, a cloud vendor is responsible for the hardware purchase and maintenance. They also provide a wide variety of software and platform as a service. We can take any required services on rent. The cloud computing services will be charged based on usage.



The cloud environment provides an easily accessible online portal that makes handy for the user to manage the compute, storage, network, and application resources. Some cloud service providers are in the following figure.



**Main Characteristics of Cloud Computing**

The five major characteristics of the Cloud Computing are as follows:  
1. **On-demand self-service** - The service of the cloud is available round theclock and provides computing capabilities on-demand of the userautomatically.  
2. **Broad network access** - Users can access the services via differentmodes as the heterogeneous thin and thick client platforms.  
3. **Resource pooling** - The feature of multi-tenancy where users areassigned resources dynamically, based on demands.  
4. **Rapid elasticity**– The service is flexible and can be scaled up or down tosuit the business requirements. Resources and programs can be usedbased on requirement and the user is billed only for the usage.  
5. **Measured service** – Usage metering is available and you pay only forwhat you use. You need not pay for any infrastructure that you do notuse.

**Applications of Cloud Computing**

1. Online File storage
2. Photo editing software
3. Digital video software
4. Twitter-related applications
5. Creating image-album
6. Web application for antivirus
7. Word processing application
8. Presentation software
9. Finding a way on the map
10. E-commerce software
11. Miscellaneous applications

1. **Online File storage**

MediaFire, megaupload, hotfile, 4Shared, rapidshare, yourfilehost are such examples which are used to host files including documents, images, presentation, videos, etc. The interface is easy to use, and users can upload and download files from these sites. Here users can utilize 200GB of storage space and a file size of 2GB. The charge for the premium version of these cloud storage application is an average of $9.

**2 .Photo editing software**

Picnik, Pixlr, etc. are popular free online photo editing software. This online software has features such as cropping of the image, resizing, rotation based on degrees, special effects, addition and editing features are also included in a GUI (Graphical User Interface) format. Some of them offer paint tools and other adjustment features. The brightness and contrast can also be editable, and users can layer the images. In the case of Pixlr, though it has various high-level, complex features, still it's easy to use.

**3. Digital Video Software**

Hulu is a free application for videos that are found online for free. Cloud users can download popular movies, television shows, and documentaries and view them on their web-browser. Hulu is a joint venture of three firms viz. - Fox Entertainment Group, NBC Universal and ABC Inc. There are other popular video sites like - WatchMoviesOnline, the most famous YouTube, Google video, etc.

**4. Twitter-Related Applications**

One example is bit.ly which converts long URL into a short small-sized unique URL. When a user clicks that small unique URL, it redirects the user to that real website. Sometimes it seems harmful as hackers can put malicious attachments or programs with it which can further affect the user. Ly made a partnership with Twitter, to allow twitter users to use shortened URLs. There is also another site name Twitpic which allows the user to upload pictures to be linked from twitter. It uses twitters login, creates shorterened URLs that can be invoked from twitters microblogging.

**5. Creating Image Album**

Some of the examples are [flickr](http://www.flickr.com/), [photobucket](http://www.photobucket.com/), [webshots](http://www.webshots.com/), [imagebam](http://www.imagebam.com/) and [ziddu](http://www.ziddu.com/) that allows users to host images on the web. These sites are a part of the cloud that allows users to organize images into albums and create slideshows for free.

**6. Web Application for Antivirus**

One example is Cloud Antivirus, this application on the cloud is provided by Panda Security - a Spanish company which provides functionality to keep the virus away from a clean system and also detects and fix a system infested with malware or other forms of computer viruses. It has been rated as the best free antivirus application by PC World. This also includes the feature to download it into the systems, and the checking of malware is done by sending the information of the file to the data-center of the cloud.

**7. Presentation Software**

Sliderocket is an online free application to create a presentation. It allows importing of Microsoft's PowerPoint presentations. Since it is a web-based cloud application, the presentations can be accessed from anywhere within the globe. But the free version doesn't allow cloud users to edit presentations offline.

**8. Word Processing Application**

Writeboard is another online word processing and document editing application. It has a unique feature that multiple users can access the same document using this application, edit that document and save the document after editing, but the document will have different versions. But it doesn't allow importing the word files.

**9. Finding a Way on the Map**

Another area where cloud applications became worth popular was finding directions and locations on the web. The leading sites are [mapquest](http://www.mapquest.com/), [Google Maps](http://maps.google.com/maps), and [Yahoo Maps](http://maps.yahoo.com/). They are the most useful free online application that helped millions of users in various ways by showing direction and paths and helped people get to their destinations over the last decade.

1. **E-Commerce Software**

Cloud based e-application allows users and e-business to respond quickly to market opportunities & challenges the modern e-commerce is facing. It became for business tycoons to focus on the usage of cloud computing without considering the time and effort involved in implementing a reliable solution. Whatever cloud computing solution they select, the free online application will need to access customer data, product data, fulfilment systems and other operational systems to support e-commerce. Cloud-based e-commerce application provides IT firms, and business leaders evaluate new opportunities without a huge amount of upfront investment.

**11. Miscellaneous Applications**

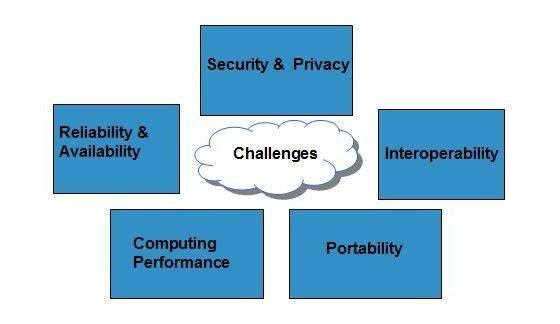
One of the 1st utilization of free SaaS applications is to check for the status of packages & items. Applications such as UPS, FedEx, US Postal Service, etc. provide free tracking of packages online. Another application name - [XE](http://www.xe.com/) provides services online from foreign exchange tools.

So every users and reader must need to know the benefits that cloud computing gives by providing free applications for users. This can reduce cost OS storage and buy paid software to do different tasks.

**Benefits of cloud computing**

o**Cost:** It reduces the huge capital costs of buying hardware and software.  
o**Speed:** Resources can be accessed in minutes, typically within a fewclicks.  
o**Scalability:** We can increase or decrease the requirement of resourcesaccording to the business requirements.  
o**Productivity:** While using cloud computing, we put less operationaleffort. We do not need to apply patching, as well as no need to maintainhardware and software. So, in this way, the IT team can be moreproductive and focus on achieving business goals.  
o**Reliability:** Backup and recovery of data are less expensive and very fastfor business continuity.  
o**Security:** Many cloud vendors offer a broad set of policies, technologies,and controls that strengthen our data security

**Cloud Computing Challenges**

Cloud computing, an emergent technology, has placed many challenges in different aspects of data and information handling. Some of these are shown in the following diagram:

**Security and Privacy**

Security and Privacy of information is the biggest challenge to cloud computing. Security and privacy issues can be overcome by employing encryption, security hardware and security applications.

**Portability**

This is another challenge to cloud computing that applications should easily be migrated from one cloud provider to another. There must not be vendor lock-in. However, it is not yet made possible because each of the cloud provider uses different standard languages for their platforms.

**Interoperability**

It means the application on one platform should be able to incorporate services from the other platforms. It is made possible via web services, but developing such web services is very complex.

**Computing Performance**

Data intensive applications on cloud requires high network bandwidth, which results in high cost. Low bandwidth does not meet the desired computing performance of cloud application.

**Reliability and Availability**

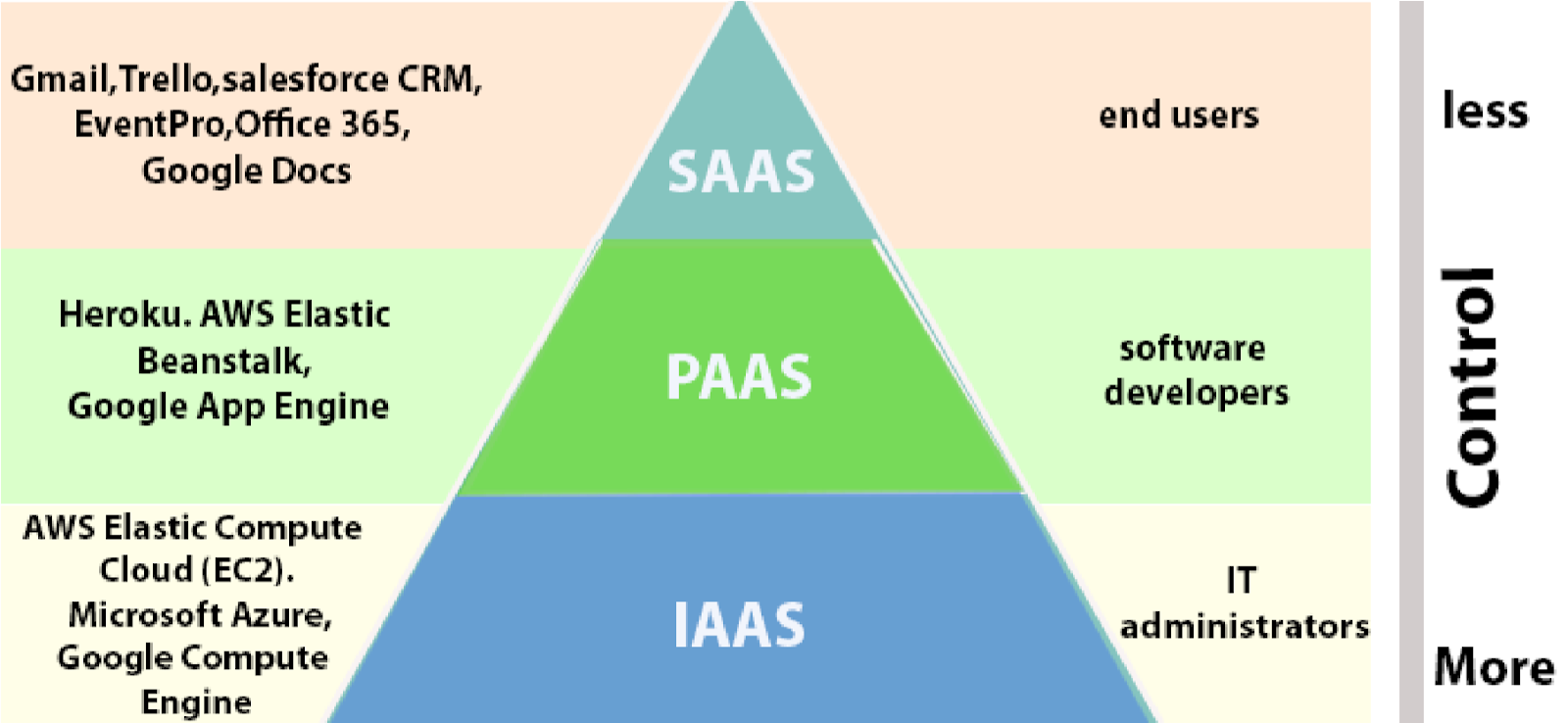
It is necessary for cloud systems to be reliable and robust because most of the businesses are now becoming dependent on services provided by third-party.

**Unit 2: Cloud Computing Service Models**

**Various Cloud Computing Service model are as:**

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

**Types of Cloud Services**



1. **Infrastructure as a Service (IaaS):** In IaaS, we can rent IT infrastructures like servers and virtual machines (VMs), storage, networks, operating systems from a cloud service vendor. We can create VM running Windows or Linux and install anything we want on it. Using

IaaS, we don’t need to care about the hardware or virtualization software, but other than that, we do have to manage everything else. Using IaaS, we get maximum flexibility, but still, we need to put more effort into maintenance.

1. **Platform as a Service (PaaS):** This service provides an on-demand environment for developing, testing, delivering, and managing software applications. The developer is responsible for the application, and the PaaS vendor provides the ability to deploy and run it. Using PaaS, the flexibility gets reduce, but the management of the environment is taken care of by the cloud vendors.
2. **Software as a Service (SaaS):** It provides a centrally hosted and managed software services to the end-users. It delivers software over the internet, on-demand, and typically on a subscription basis. E.g., Microsoft One Drive, Dropbox, WordPress, Office 365, and Amazon Kindle. SaaS is used to minimize the operational cost to the maximum extent.

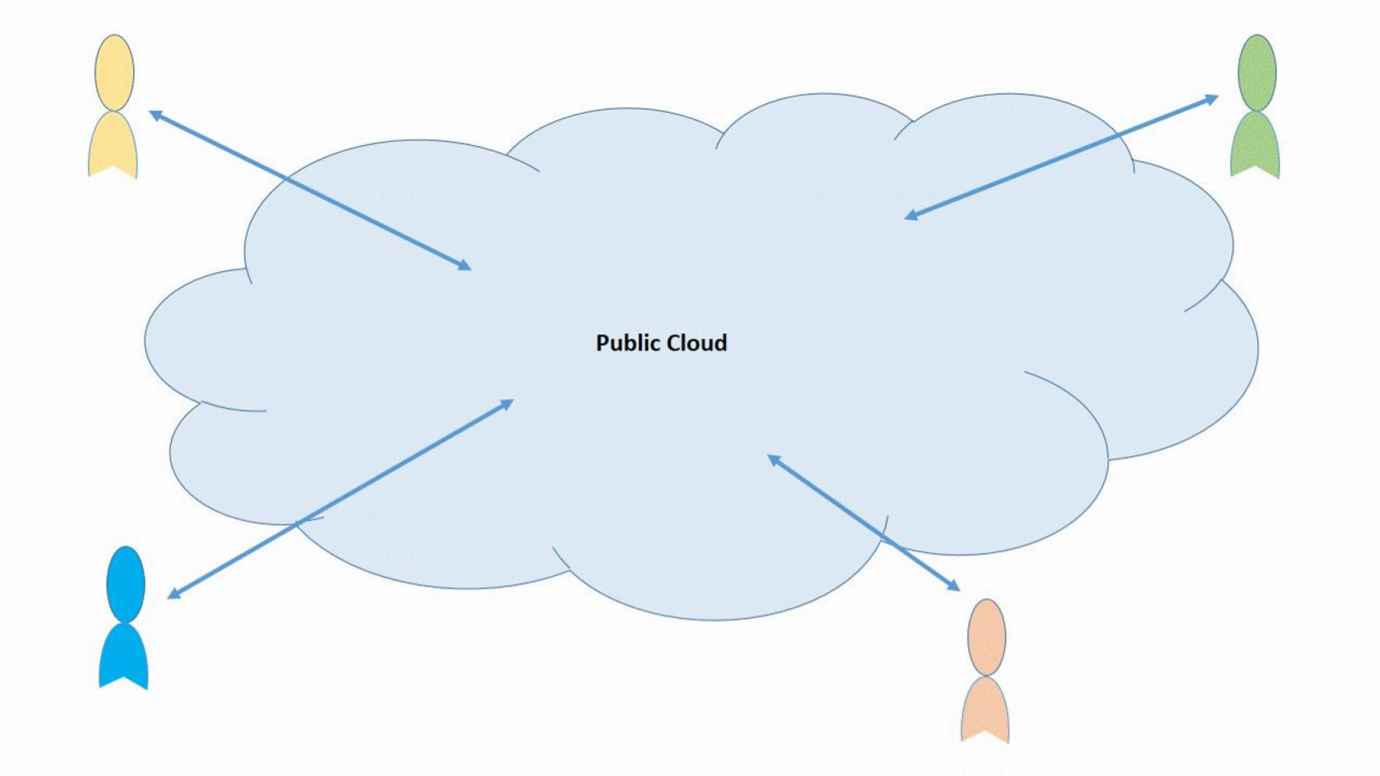
**Cloud Computing Deployment Models**

The cloud services can be deployed in different methods. The deployment model is based on the service model, organizational structure, location, user base, and so on. The four most commonly used deployment models are as follows:

1. **Public Cloud**
2. **Private Cloud**
3. **Community Cloud**
4. **Hybrid Cloud**

**1.Public Cloud**

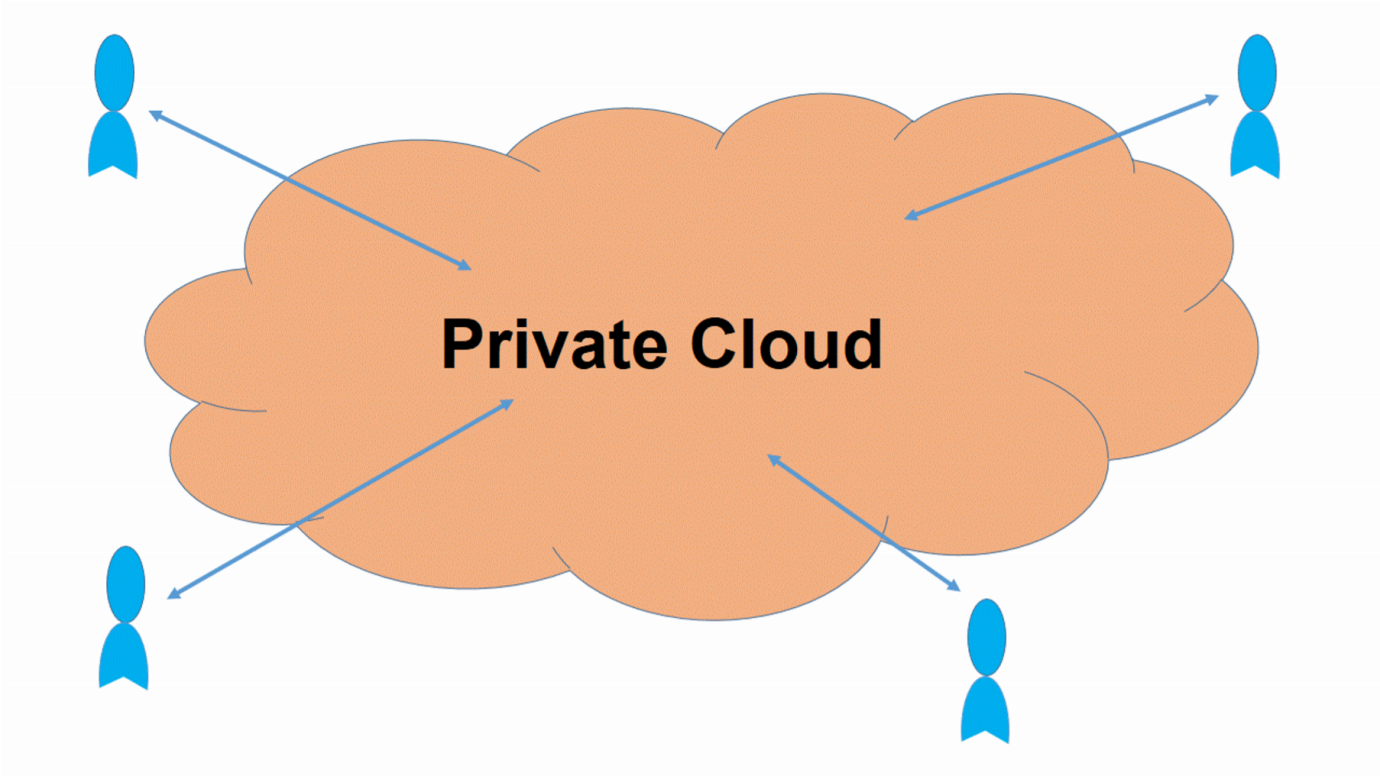
In this model, the infrastructure is accessible to the public and it is owned by a vendor, who offers the services of the cloud to the users. The cloud vendor shares the cloud resources with the end users. The resource pool is huge and the services are shared by lots of users. The services of this cloud model can be free or available for nominal charges. Google uses a public cloud deployment model. With this model, users need not purchase any infrastructure but can use that of the vendor. A drawback of the public cloud model is that it poses a security threat. If you have very confidential data running in your network, it is not safe to use the public cloud model.



**Fig2.2.1: Public Cloud**

**1.Private Cloud**

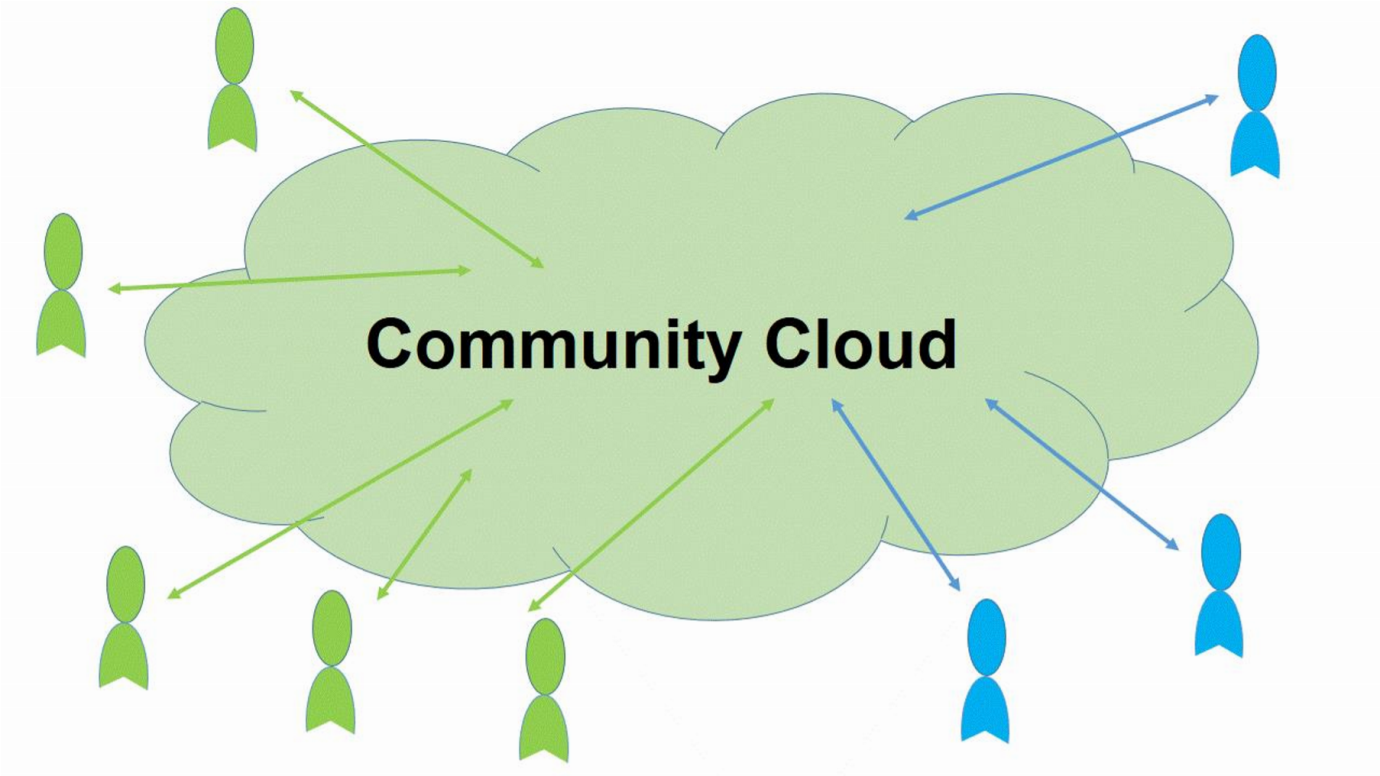
As the name suggests, this would be a privately owned cloud. Here, the user or organization owns the cloud and only the user or employees of the company have access to the cloud, thereby making data and transactions secure. There is more control over resources when compared to the Public Cloud model. The Private cloud model uses the Virtualization solution and the data centers belong to the company. The major advantage of this model is the security and the control that the users have over the resources and application. However, the drawback is that more financial investment is required and the offering is not as big scale as that of a public cloud model.



**Fig2.2.2: Private Cloud**

1. **Community Cloud**

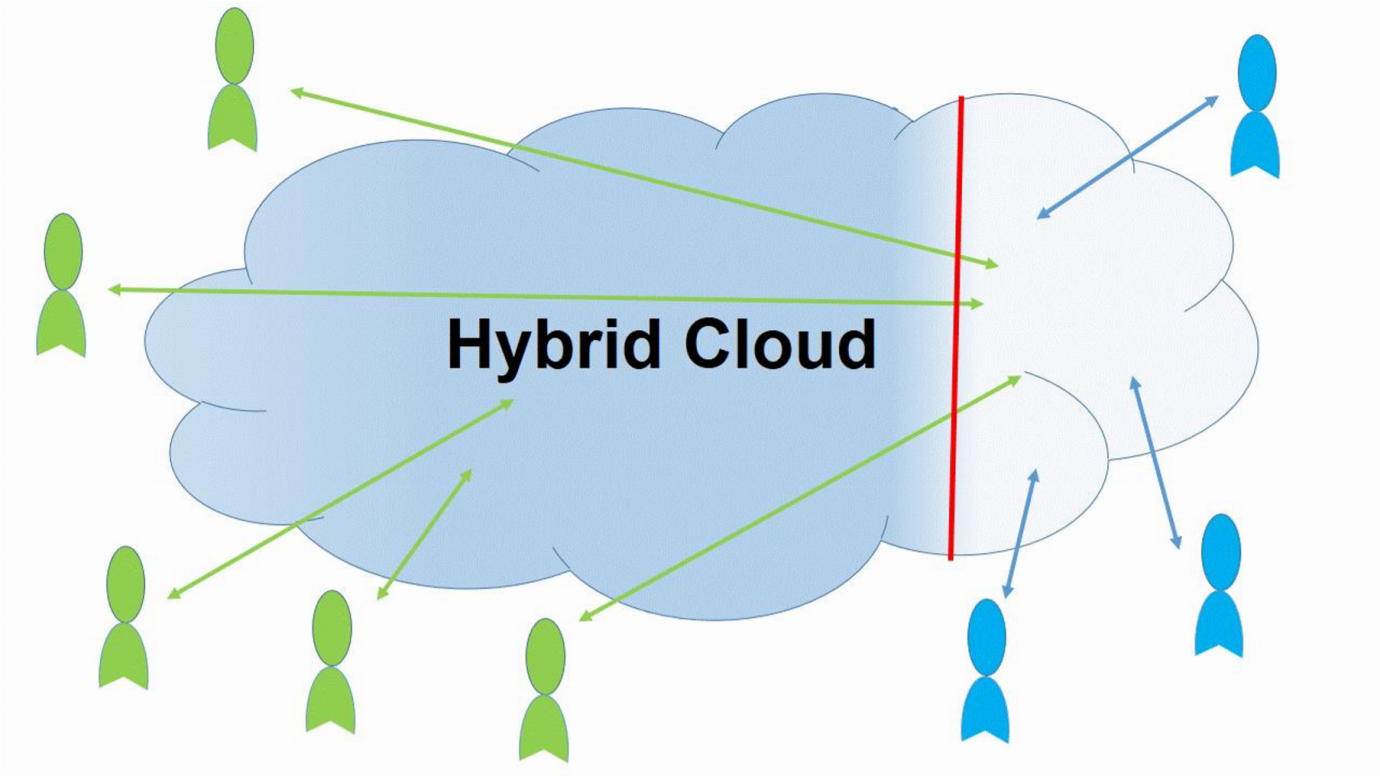
In the Community Cloud model, the infrastructure is owned jointly by different organizations. The organizations may have a similar set of requirements, policies, and customer base. So, they can combine the offerings and make the customer base even bigger. Duplication of same or similar applications and resources are avoided. This model helps reduce the costs, which would otherwise be higher if the organization deploys the Private Cloud model. This is again a classification of the Private Cloud, as it is available to only a certain group of users.



**Fig2.2.3: Community Cloud**

1. **Hybrid Cloud**

The Hybrid Cloud deployment model comprises of two or more clouds. This can be a combination of the other three cloud types – public, private, or community. The hybrid deployment is complex compared to the other three owing to the execution and management tasks involved. An example scenario of this model can be where an organization is on the private cloud but there are load spikes which the private cloud cannot handle. For this the organization depends on the public cloud to support the load. The shift from the private to the public cloud and back will be seamless to the end user



**Fig2.2.4: Hybrid Cloud**

**Major Cloud Service Providers**

* **Kamatera**
* **Softchoice**
* **Prolifics**
* **ScienceSoft**
* **phoenixNAP**
* **Cloudways**
* **pCloud**
* **Amazon Web Services**
* **Microsoft Azure**
* **Google Cloud Platform**
* **Adobe**
* **VMware**
* **IBM Cloud**
* **Rackspace**
* **Red Hat**
* **Salesforce**
* **Oracle Cloud**
* **SAP**
* **Verizon Cloud**
* **Navisite**
* **Dropbox**
* **Egnyte**

The following table summarizes the top 3 key players and their offerings in the cloud computing world:

|  |  |  |  |
| --- | --- | --- | --- |
|  | **AWS** | **Azure** | **Google Cloud** |
| **Company** | **AWS Inc.** | **Microsoft** | **Google** |
| **Launch year** | 2006 | 2010 | 2008 |
| **Geographical Regions** | 25 | 54 | 21 |
| **Availability Zones** | 78 | 140 (countries) | 61 |
| **Key offerings** | Compute, storage, database, analytics, networking, machine learning, and AI, mobile, developer tools, IoT, security, enterprise applications, blockchain. | Compute, storage, mobile, data management,  messaging, media services, CDN, machine learning and AI, developer tools, security, blockchain, functions, IoT. | Compute, storage, databases, networking, big data, cloud AI, management tools, Identity and security,  IoT, API platform |
| **Compliance Certificates** | 46 | 90 |  |
| **Annual Revenue** | $33 billion | $35 billion | $8 billion |

Brief of Cloud Computing Deployment Models are as:

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

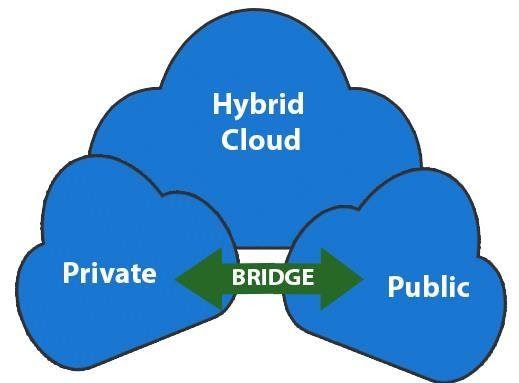


Fig: Types of Cloud Computing

1. **Public Cloud:** The cloud resources that are owned and operated by a third-party cloud service provider are termed as public clouds. It delivers computing resources such as servers, software, and storage over the internet
2. **Private Cloud:** The cloud computing resources that are exclusively used inside a single business or organization are termed as a private cloud. A private cloud may physically be located on the company’s on-site data centre or hosted by a third-party service provider.
3. **Hybrid Cloud:** It is the combination of public and private clouds, which is bounded together by technology that allows data applications to be

shared between them. Hybrid cloud provides flexibility and more deployment options to the business.

**Unit-3 Service-level agreement (SLA)**

**Overview of SLA**

SLAs are a critical component of any vendor contract. Beyond listing expectations of service type and quality, an SLA provides remedies when requirements aren't met.

A service-level agreement (SLA) is nothing more than a type of contract between two parties. In the context of managed IT services (in which SLAs most frequently appear), SLAs dictate the quality and type of service that will be provided to the client in exchange for a fee. SLAs also provide the remedy, such as a reduced fee structure, that will apply in the case of a service outage.

So, for example, if the contract specifies 99.9999 percent uptime and that is not met, the customer would have the right to reduce its bill by an agreed-on percentage. SLAs are important because they set the tone for the relationship between the parties and will govern if and when things break down. A "good" SLA is a balance between being thorough and clear on one side, while not being overly onerous on the service provider on the other.

**Why it’s vital to have an SLA?**

Without service level agreement, it is not clear what will occur on the off chance that one of the parties does not hold up to their end of the deal. For instance, we should think about that a telecom supplier’s service level target is to answer all assistance work area calls within 5 seconds, and the calls are just being replied in 5 minutes. They can without much of a stretch say that they never guaranteed that the calls will be replied in 5 seconds if there was no SLA set up. Having a service level agreement takes into account precision about what the administration level targets are just as what occurs if the required targets are not met. With an SLA set up, the two parties are ensured.

**The primary segments of an SLA**

A service level agreement states what the two parties need to accomplish with their understanding alongside a layout of the duties of each party incorporating expected productivity with performance measures. A service level agreement, as a rule, has a term period that is indicated in the agreement. Every single service that is incorporated into the agreement is depicted and can likewise contain insights about systems for checking the exhibition of the services and also as crisis-management processes.



In case you have cracked an ITIL (Information Technology Infrastructure Library) Training you would not need to ask what service level agreement (SLA) is. In any case, on the off chance that you have not completed an ITIL online course yet, we will help you out. An SLA is an agreement between an IT Service provider and a consumer.

If you have cleared ITIL Training (Information Technology Infrastructure Library) Training, you will not ask about service level agreement (Service Level Agreement). This is an agreement made between a consumer and an IT service provider. There are three different types of SLA and it is important to know these in detail before drafting the agreement.

For example, you are a consumer of a bank and the bank gives services to you. An SLA among you and the bank depicts the services provided and the service levels at which they will be given. For instance, the bank will enable you to withdraw cash from an ATM and the transaction will last no longer than 10 seconds. That is a case of an SLA and it is a section of service level administration.

There are three types of [SLA](https://vakilsearch.com/service-level-agreement) that can be filed. Before explaining ITIL service level necessities and concurring on the service levels through SLA, the most suitable SLA structure must be planned. Depending upon the administrationsthat will be given to a client or business, an SLA structure is planned that fits the reason. SLAs are generally marked amid the ITIL service configuration phase of the ITIL lifecycle.

[**Service level SLA**](https://vakilsearch.com/master-service-agreement)

The principal sort of SLA structure is the service based SLA. An administration based SLA covers one service for all clients. Let’s believe that the IT specialist organisation provides client question administration to numerous clients. In a service based SLA, the service level of the consumer query service will be the same for all clients that will utilise this administration.

For example, if the account division and the HR office are two clients that will utilise this administration, same SLA will be legitimate between the IT service provider and these two offices since it is an administration based SLA.

**Client based SLA**

The second sort of SLA structure is the client based SLA. A client based SLA is concurrence with one client, covering every one of the administrations utilised by this client. Let us think about the connection between you and your telecom administrator. You utilise the voice service, SMS administrations, date service, and a few different administrations of the telecom service. For every one of these administrations, you have just one contract among you and the telecom administrator. Likewise, if the IT specialist co-op gives a few administrations to the business and the clients, and if all the service levels are archived in one SLA for the provided administrations, it will be a client based SLA.

**Multi-level SLA**

The third and the last kind of service level agreement is the multi-level SLA. In multi-level SLA, parts of SLA are characterised by the association of the client utilising some sort of legacy within general definitions with pertinence for every single subordinate dimension. This SLA centres on the association of the client. All administrations and their inter-relationships with subordinate administrations are utilised when characterising the multi-level administration level agreement structure.

Sustaining [service level agreements](https://vakilsearch.com/service-level-agreement) are a piece of SL management. Each time an administration change, or the administration level focus of an administration change, the service level agreement should be checked on and modified. The new SLA needs to mirror the progressions made to the administration or the administration level targets. In this way, the administration of service levelagreement is a significant piece of ITIL interminable administration improvement.

**SLA Life Cycle**

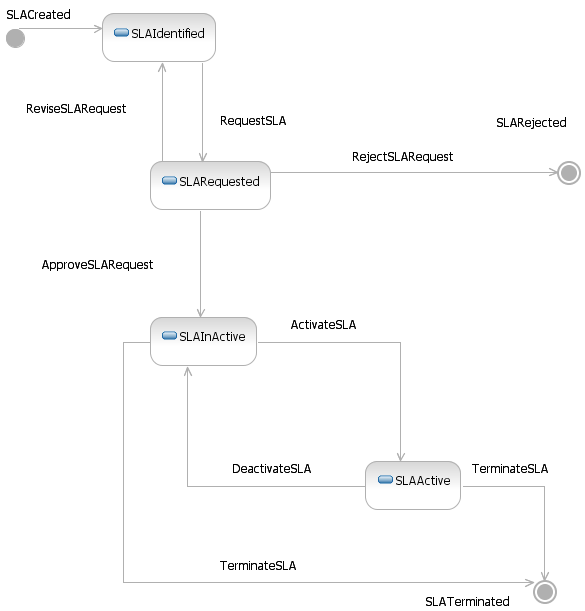
You use the service level agreement (SLA) lifecycle to govern a service level agreement from being initially identified, through to being activated, and, eventually, terminated when it is no longer required.

A [Service level agreement](https://www.ibm.com/support/knowledgecenter/SSWLGF_8.0.0/com.ibm.sr.doc/rwsr_gep_service_level_agreement.html?view=kc) object can pass through the service level agreement lifecycle.

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.

|  |  |  |
| --- | --- | --- |
| **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**



* [**Service level agreement lifecycle: state and transition URIs**](https://www.ibm.com/support/knowledgecenter/SSWLGF_8.0.0/com.ibm.sr.doc/rwsr_sla_lc_uris.html?view=kc)

Each of the states and transitions in the deploy and manage phases of the service level definition lifecycle are identified by a URI.

**SLA Management Process**

**Step 1. Meet Prerequisites for SLAs**

Experts believe that there are a few prerequisites that the organization must meet in order to come up with a spot on the service level agreement. Here are a few prerequisites you may want to consider.

* The organization must have a service oriented culture
* All IT activities must be driven by customer/business initiatives
* The organization must fully commit to the SLA process and contract

**Step 2. Determine the Parties Involved In the SLA**

You must choose wisely when it comes to choosing that parties must be involved in the SLA so that it serves as a platform for coming up with new service level goals. Some of the possible goals are:

* Create and meet reactive support business objectives
* Provide the top level availability by clearly defining proactive SLAs
* Promote or sell a service

**Step 3. Determine Service Elements**

Typical service level agreements usually have many different components including support level, how it will be measured along with overall budget concerns and escalation path for SLA reconciliation. Having proactive service definitions and reactive goals are a must for high availability environments. Some of them are listed below.

* Onsite support business hours and procedures for off-support timings
* Priority definitions like the type of problem, maximum time for solving the problem and escalation processes
* Products and services that are to be supported and ranked in order of business criticality
* Support level issues pertaining to geographic and business units
* Goals for improving help desk service
* Network error detection and service response
* Funding the implemented SLA
* Procedures for conflict resolution
* Measurement of network availability and reporting

**Step 4. Define the SLA Required For Each Group**

Primary support service level agreements must include functional group representation, critical business units, server operations, networking operations and application support groups. Such groups are usually on business needs as well as the roles they play in the support process.

**Step 5. Negotiate the SLA**

The last step of creating an SLA is final negotiation and sign off. SLA negotiation is further split up into the following steps.

* Review of the draft
* Negotiation of the contents
* Edit and revise the document
* Obtain the final approval

**Step 6. Measure and Monitor SLA Conformance**

To ensure long-term consistency and results, measuring SLA conformance and reporting results becomes an important aspect that one must pay due attention to. It is recommended that all major components of the SLA must be measurable, and a measurement method should be put in place prior to implementing the SLA. Problems can then be solved by holding meetings with the customers quarterly or as and when problems arise.

**EXAMPLES OF SLA**

* Backbone internet providers usually create and state their service level agreement on their official website.
* A WSLA or Web Service Level Agreement is usually put in place so as to monitor web services. Authors also get a chance to specify their performance metrics associated with a web service application.
* Cloud computing services also provide a service based agreement instead of a customer based agreement. It contains important details such as measuring, monitoring and reporting related to the cloud’s performance.

**Unit-4 Virtualization Concepts**

**Overview**

Virtualization is a technology that helps us to install different Operating Systems on a hardware. They are completely separated and independent from each other. Virtualization hides the physical characteristics of computing resources from their users, their applications or end users. This includes making a single physical resource (such as a server, an operating system, an application or a storage device) appear to function as multiple virtual resources. It can also include making multiple physical resources (such as storage devices or servers) appear as a single virtual resource...”

Virtualization is often −

* The creation of many virtual resources from one physical resource.
* The creation of one virtual resource from one or more physical resource.

Types of Virtualization

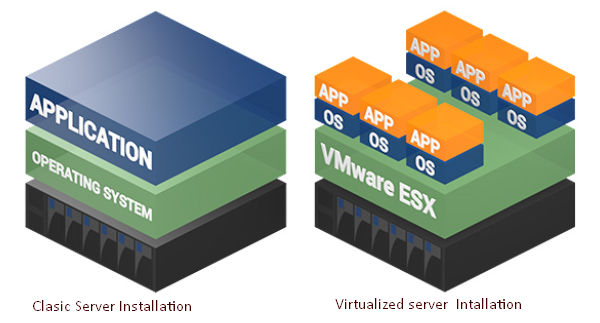
Today the term virtualization is widely applied to a number of concepts, some of which are described below −

* Server Virtualization
* Client & Desktop Virtualization
* Services and Applications Virtualization
* Network Virtualization
* Storage Virtualization

Let us now discuss each of these in detail.

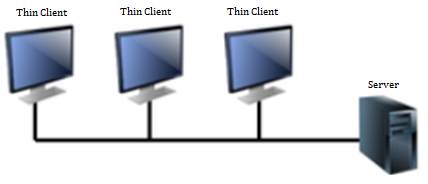
Server Virtualization

It is virtualizing your server infrastructure where you do not have to use any more physical servers for different purposes.



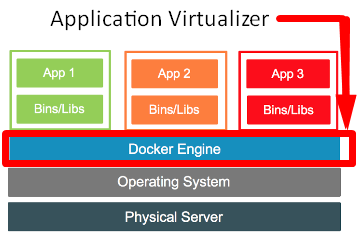
**Client & Desktop Virtualization**

This is similar to server virtualization, but this time is on the user’s site where you virtualize their desktops. We change their desktops with thin clients and by utilizing the datacentre resources.



Services and Applications Virtualization

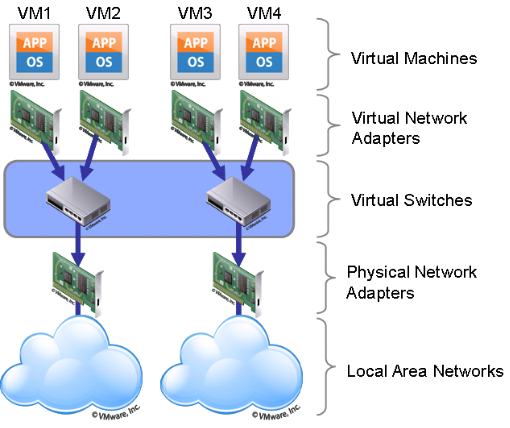
The virtualization technology isolates applications from the underlying operating system and from other applications, in order to increase compatibility and manageability. For example – Docker can be used for that purpose.



Network Virtualization

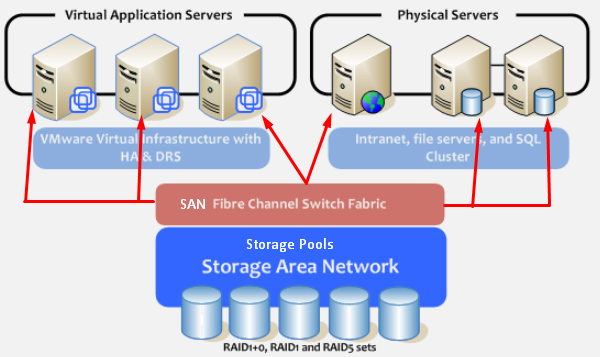
It is a part of virtualization infrastructure, which is used especially if you are going to visualize your servers. It helps you in creating multiple switching, Vlans, NAT-ing, etc.

The following illustration shows the VMware schema −



Storage Virtualization

This is widely used in datacentres where you have a big storage and it helps you to create, delete, allocated storage to different hardware. This allocation is done through network connection. The leader on storage is SAN. A schematic illustration is given below −



Understanding Different Types of Hypervisors

A hypervisor is a thin software layer that intercepts operating system calls to the hardware. It is also called as the **Virtual Machine Monitor** (VMM). It creates a virtual platform on the host computer, on top of which multiple guest operating systems are executed and monitored.

Hypervisors are two types −

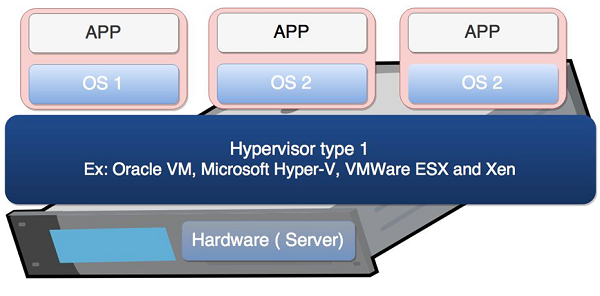
* Native of Bare Metal Hypervisor and
* Hosted Hypervisor

Let us now discuss both of these in detail.

Native or Bare Metal Hypervisor

Native hypervisors are software systems that run directly on the host's hardware to control the hardware and to monitor the **Guest Operating Systems**. The guest operating system runs on a separate level above the hypervisor. All of them have a Virtual Machine Manager.

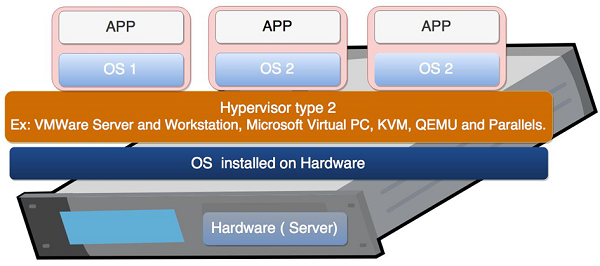
Examples of this virtual machine architecture are **Oracle VM, Microsoft Hyper-V, VMWare ESX** and **Xen**.



Hosted Hypervisor

Hosted hypervisors are designed to run within a traditional operating system. In other words, a hosted hypervisor adds a distinct software layer on top of the host operating system. While, the guest operating system becomes a third software level above the hardware.

A well-known example of a hosted hypervisor is **Oracle VM VirtualBox**. Others include **VMWare Server and Workstation, Microsoft Virtual PC, KVM, QEMU** and **Parallels**.



Understanding Local Virtualization and Cloud

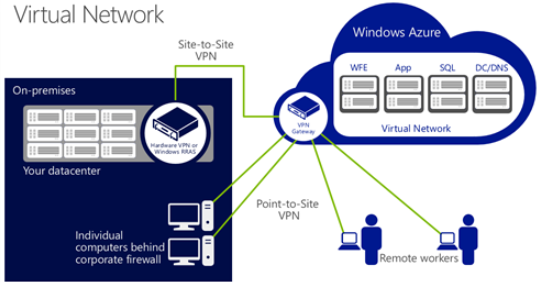
Virtualization is one of the fundamental technologies that makes cloud-computing work. However, virtualization is not cloud computing. Cloud computing is a service that different providers offer to you based on some costs.

In enterprise networks, virtualization and cloud computing are often used together to build a public or private cloud infrastructure. In small businesses, each technology will be deployed separately to gain measurable benefits. In different ways, virtualization and cloud computing can help you keep your equipment spending to a minimum and get the best possible use from the equipment you already have.

As mentioned before, virtualization software allows one physical server to run several individual computing environments. In practice, it is like getting multiple servers for each physical server you buy. This technology is fundamental to cloud computing. Cloud providers have large data centers full of servers to power their cloud offerings, but they are not able to devote a single server to each customer. Thus, they virtually partition the data on the server, enabling each client to work with a separate “virtual” instance (which can be a private network, servers farm, etc.) of the same software.

Small businesses are most likely to adopt cloud computing by subscribing to a cloud-based service. The largest providers of cloud computing are **Microsoft with Azure** and **Amazon**.

The following illustration is provided by Microsoft where you can understand how utilizing extra infrastructure for your business without the need to spend extra money helps. You can have the on-premises base infrastructure, while on cloud you can have all your services, which are based on Virtualized technology.



**Benefits of Virtualization**

**1. It is cheaper.**  
 Because virtualization doesn’t require actual hardware components to be used or installed, IT infrastructures find it to be a cheaper system to implement. There is no longer a need to dedicate large areas of space and huge monetary investments to create an on-site resource. You just purchase the license or the access from a third-party provider and begin to work, just as if the hardware were installed locally.

**2. It keeps costs predictable.**  
 Because third-party providers typically provide virtualization options, individuals and corporations can have predictable costs for their information technology needs. For example: the cost of a Dell PowerEdge T330 Tower Server, at the time of writing, is $1,279 direct from the manufacturer. In comparison, services provided by Bluehost Web Hosting can be a slow as $2.95 per month.

**3. It reduces the workload.**  
 Most virtualization providers automatically update their hardware and software that will be utilized. Instead of sending people to do these updates locally, they are installed by the third-party provider. This allows local IT professionals to focus on other tasks and saves even more money for individuals or corporations.

**4. It offers a better uptime.**  
 Thanks to virtualization technologies, uptime has improved dramatically. Some providers offer an uptime that is 99.9999%. Even budget-friendly providers offer uptime at 99.99% today.

**5. It allows for faster deployment of resources.**  
 Resource provisioning is fast and simple when virtualization is being used. There is no longer a need to set up physical machines, create local networks, or install other information technology components. As long as there is at least one point of access to the virtual environment, it can be spread to the rest of the organization.

**6. It promotes digital entrepreneurship.**  
 Before virtualization occurred on a large scale, digital entrepreneurship was virtually impossible for the average person. Thanks to the various platforms, servers, and storage devices that are available today, almost anyone can start their own side hustle or become a business owner. Sites like Fiverr and UpWork make it possible for anyone to set a shingle and begin finding some work to do.

**7. It provides energy savings.**  
 For most individuals and corporations, virtualization is an energy-efficient system. Because there aren’t local hardware or software options being utilized, energy consumption rates can be lowered. Instead of paying for the cooling costs of a data center and the operational costs of equipment, funds can be used for other operational expenditures over time to improve virtualization’s overall ROI.

**What is hypervisor?**

The hypervisor is a key to enable virtualization. In its simpler form, the hypervisor is specialized firmware or software, or both, installed on single hardware that would allow you to host several virtual machines. It allows physical hardware to be shared across several virtual machines. A computer on which hypervisor runs one or more virtual machines is called a host machine. The virtual machine is called a guest machine. Basically, the hypervisor allows the physical host machine to run various guest machines. This helps in achieving maximum benefits from computing resources such as memory, network bandwidth, and CPU cycles.

**Advantages of Hypervisors**

* Though virtual machines operate on the same physical hardware, they are separated from each other. This also depicts that if one virtual machine undergoes a crash, error, or a malware attack, it doesn't affect the other virtual machines.
* Another benefit is that virtual machines are very mobile as they don't depend on the underlying hardware. Since they are not linked to physical hardware, switching between local or remote virtualized servers gets a lot easier as compared to traditional applications.

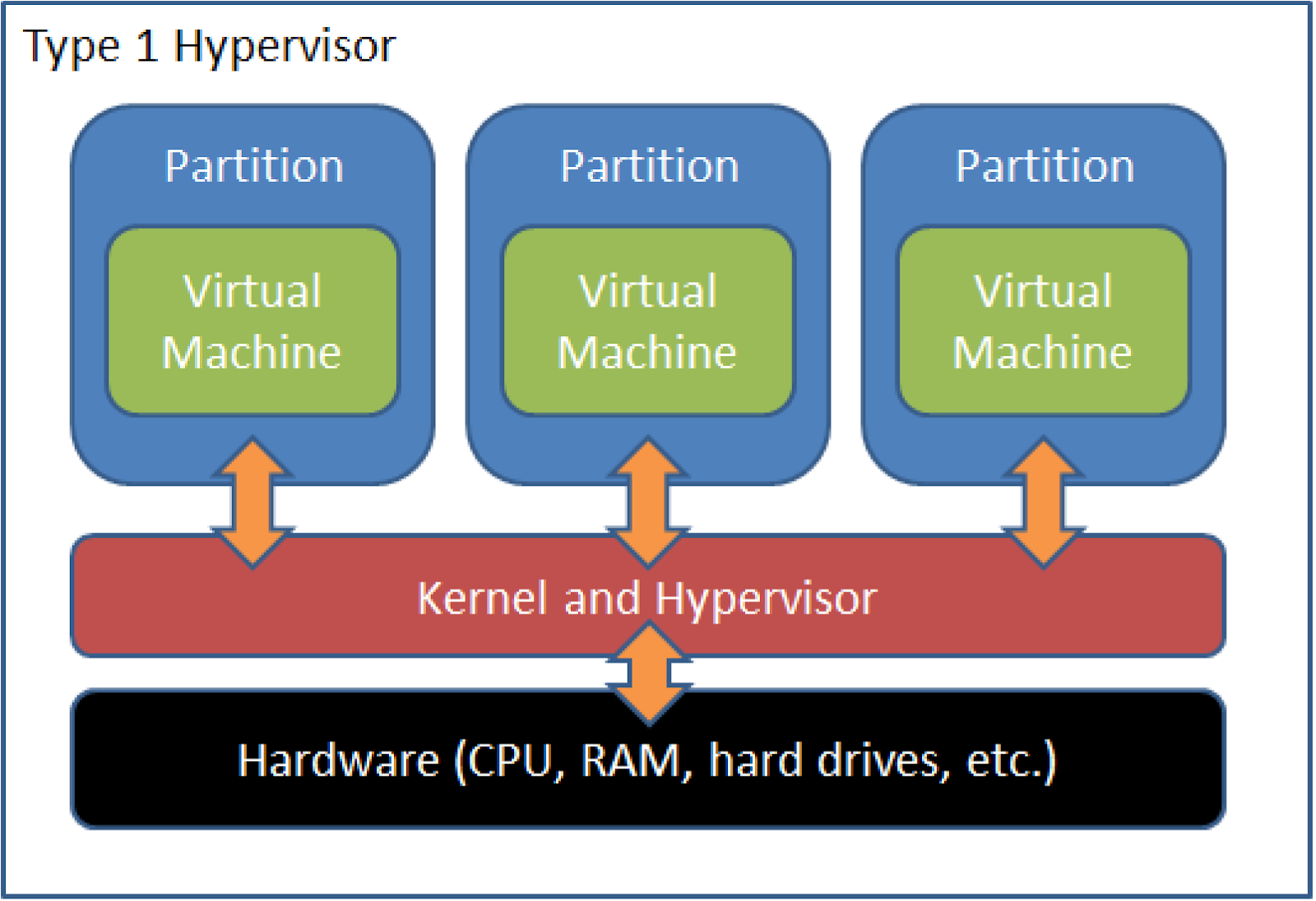
**Types of Hypervisor In Cloud Computing**

There are two main types of hypervisor in cloud computing.

**Type I Hypervisor**

A type I hypervisor operates directly on the host's hardware to monitor hardware and guest virtual machines, and it's referred to as the bare metal. Usually, they don't require the installation of software ahead of time. Instead, you can install right onto the hardware. This type of hypervisor tends to be powerful and requires a great deal of expertise to function it well. In addition, Type I hypervisor are more complex and have certain hardware requirements to run adequately. Due to this, it is mostly chosen by IT operations and data center computing.

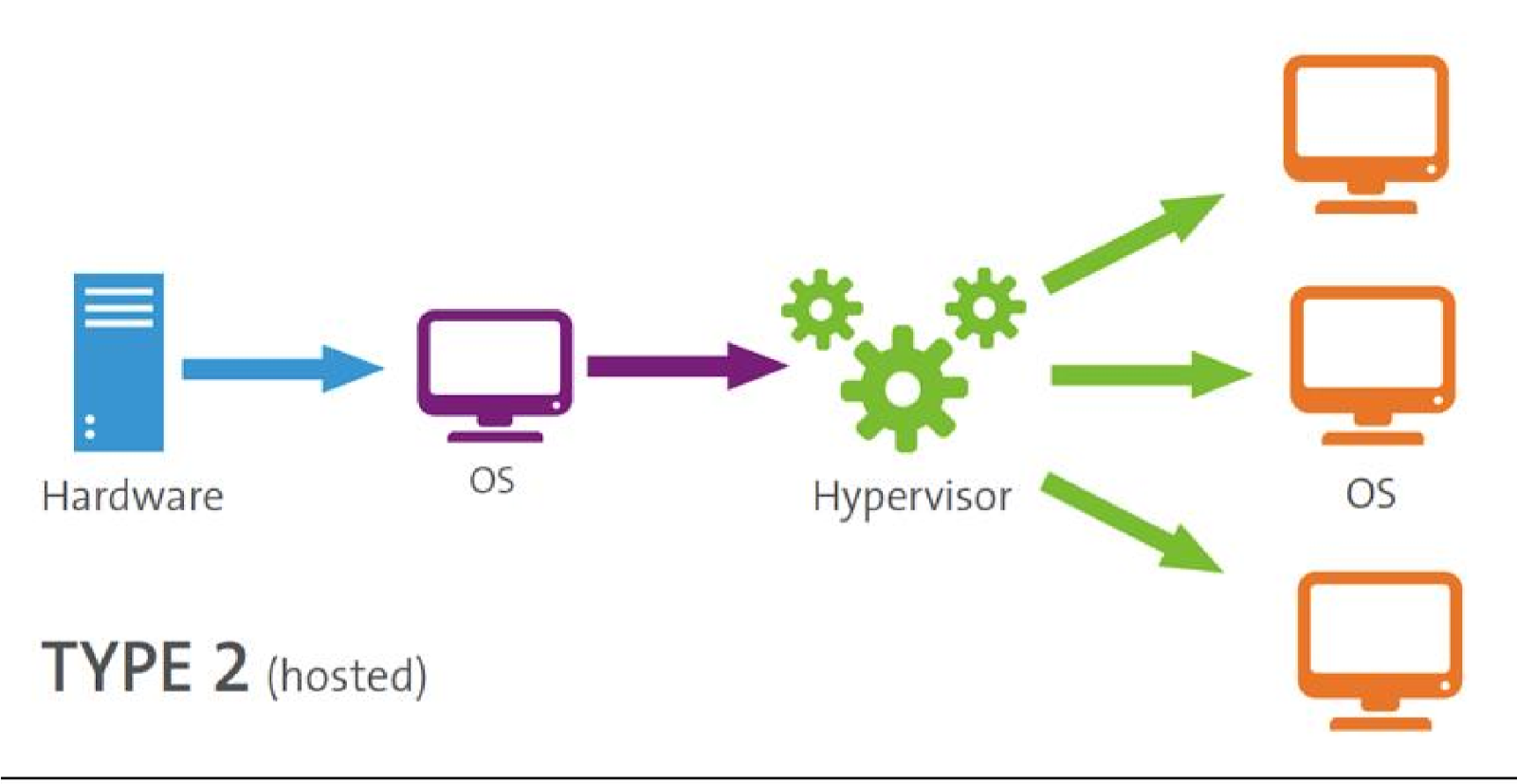
Examples of type I hypervisors include Xen, Oracle VM Server for SPARC, Oracle VM Server for x86, Microsoft Hyper-V and VMware’s ESX/ESXi.



**Type II Hypervisor**

It's also called a hosted hypervisor because it is usually installed onto an existing operating system. They are not much capable to run more complex virtual tasks. People use it for basic development, testing, and emulation. If there is any security flaw found inside the host OS, it can potentially compromise all of virtual machines running. This is why type II hypervisors cannot be used for data center computing. They are designed for end-user systems where security is a lesser concern. For instance, developers could use type II Hypervisor to launch virtual machines in order to test software product before their release.

A few examples are Virtual box, VMware workstation, fusion.



**Benefits of hypervisors**

There are several benefits to using a hypervisor that hosts multiple virtual machines:

* **Speed:** Hypervisors allow virtual machines to be created instantly, unlike bare-metal servers. This makes it easier to provision resources as needed for dynamic workloads.
* **Efficiency:** Hypervisors that run several virtual machines on one physical machine’s resources also allow for more efficient utilization of one physical server. It is more cost- and energy-efficient to run several virtual machines on one physical machine than to run multiple underutilized physical machines for the same task.
* **Flexibility:** Bare-metal hypervisors allow operating systems and their associated applications to run on a variety of hardware types because the hypervisor separates the OS from the underlying hardware, so the software no longer relies on specific hardware devices or drivers.
* **Portability:** Hypervisors allow multiple operating systems to reside on the same physical server (host machine). Because the virtual machines that the hypervisor runs are independent from the physical machine, they are portable. IT teams can shift workloads and allocate networking, memory, storage and processing resources across multiple servers as needed, moving from machine to machine or platform to platform. When an application needs more processing power, the virtualization software allows it to seamlessly access additional machines.

**Unit-5 Cloud Security**

**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.

Security Planning

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

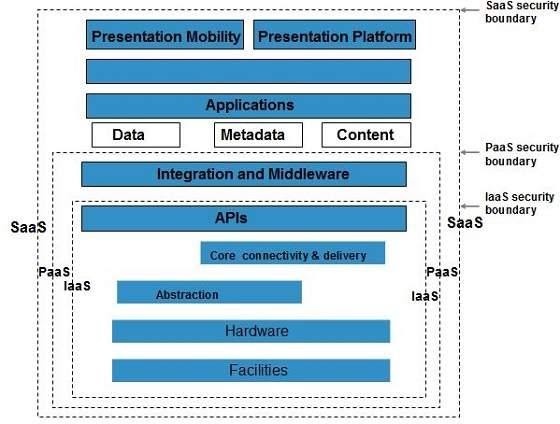
* Select resource that needs to move to the cloud and analyze its sensitivity to risk.
* Consider cloud service models such as **IaaS, PaaS,** and **SaaS.** These models require customer to be responsible for security at different levels of service.
* Consider the cloud type to be used such as **public,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

The risk in cloud deployment mainly depends upon the service models and cloud types.

Understanding Security of Cloud

Security Boundaries

A particular service model defines the boundary between the responsibilities of service provider and customer. **Cloud Security Alliance (CSA)** stack model defines the boundaries between each service model and shows how different functional units relate to each other. The following diagram shows the **CSA stack model:**



Key Points to CSA Model

* IaaS is the most basic level of service with PaaS and SaaS next two above levels of services.
* Moving upwards, each of the service inherits capabilities and security concerns of the model beneath.
* IaaS provides the infrastructure, PaaS provides platform development environment, and SaaS provides operating environment.
* IaaS has the least level of integrated functionalities and integrated security while SaaS has the most.
* This model describes the security boundaries at which cloud service provider's responsibilities end and the customer's responsibilities begin.
* Any security mechanism below the security boundary must be built into the system and should be maintained by the customer.

Although each service model has security mechanism, the security needs also depend upon where these services are located, in private, public, hybrid or community cloud.

Understanding Data Security

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

All of the service models should incorporate security mechanism operating in all above-mentioned areas.

Isolated Access to Data

Since data stored in cloud can be accessed from anywhere, we must have a mechanism to isolate data and protect it from client’s direct access.

**Brokered Cloud Storage Access** is an approach for isolating storage in the cloud. In this approach, two services are created:

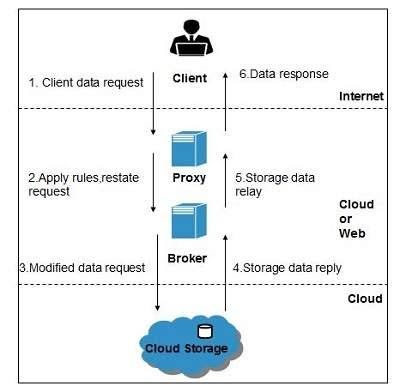
* A broker with full access to storage but no access to client.
* A proxy with no access to storage but access to both client and broker.

Working Of Brokered Cloud Storage Access System

When the client issues request to access data:

* The client data request goes to the external service interface of proxy.
* The proxy forwards the request to the broker.
* The broker requests the data from cloud storage system.
* The cloud storage system returns the data to the broker.
* The broker returns the data to proxy.
* Finally the proxy sends the data to the client.

All of the above steps are shown in the following diagram:



Encryption

Encryption helps to protect data from being compromised. It protects data that is being transferred as well as data stored in the cloud. Although encryption helps to protect data from any unauthorized access, it does not prevent data loss.

**Legal Issues around Cloud Computing**

Cloud computing is bringing amazing advantages and benefits companies. But it also brings some challenges. There are several legal issues that must be taken into consideration when moving into the cloud. Let’s see which are the most challenging legal issues around cloud computing and how to overcome them.

**Security procedures**

The majority of companies which implemented cloud solutions and services do not have security procedures in place. Also, they lack measures to approve or evaluate cloud applications. When adopting the BYOD trend for example, organizations needed these security procedures more than ever. General data security trainings, multiple levels of security, rigorous procedures to use one’s own device and to transfer or copy data are some of the options available to protect data in organizations. The bottom line is that security procedures must be established according to every company’s objectives and work flow.

**Third party access issues**

Third-party involvement could be a risk. All third parties using a multi-tenant shared cloud are using the same administration interface, so make sure multi-factor authentication and enhanced security is present.  Also, look for [HIPAA](https://www.rickscloud.com/how-to-build-hipaa-compliant-cloud-applications/)compliant providers – a business associate agreement (BAA) with third-party vendor who access Protected Health Information (PHI) is necessary to ensure privacy and security requirements. A partnership with a HIPAA solutions provider that signs a BAA is an efficient method to make sure this this goes smoothly and everything is secure. And don’t forget to read carefully the terms and conditions before signing up for a cloud based services.

**Intellectual Property Rights**

Intellectual Property Rights differ from one country to another, so it is not very clear what intellectual property laws will apply in the cloud computing environment. Make sure you are aware of the regulations and rights from the country you store your intellectual work. The provider you choose should know how to protect intellectual property it stores and how to avoid potential infringement pitfalls.

**Confidential data theft attacks**

Data stored in the cloud might be compromised or breached. Therefore, most cloud computing providers also offer the customer different [levels](https://www.rickscloud.com/top-myths-about-cloud-computing/)of security protection, which allows for more enhanced security. Encryption might seem to have failed in protecting data from theft attacks, but other methods have been discovered and implemented, including monitoring data access in the cloud to detect abnormal data access patterns. The customer has to understand the cloud provider’s disclosure policy and how quickly the breach would be disclosed to them. Most of the U.S. states have security breach disclosure laws requiring the provider to inform the customers when their data has been compromised.

Many of these legal issues and the methods to inform about them or to solve them should be mentioned in the [Service Level Agreement](https://www.rickscloud.com/working-on-a-cloud-software-service-level-agreement/). It is essential to understand all the terms of the cloud’s provider and to consider the needs and objectives of the enterprise before signing an agreement.

**Unit-6 Cloud Storage**

**What is cloud storage?**

Cloud storage is a service which lets you store data by transferring it over the Internet or another network to an offsite storage system maintained by a third party. There are hundreds of different cloud storage systems which include personal storage which holds and/or backs up emails, pictures, videos and other personal files of an individual, to enterprise storage which lets businesses use cloud storage as a commercially-supported remote backup solution where the company can securely transfer and store data files or share them between locations.

Storage systems are typically scalable to suit an individual’s or organisation’s data storage needs, accessible from any location and are application-agnostic for accessibility from any device. Businesses can select from three main models: a [public cloud](https://azure.microsoft.com/en-in/overview/what-is-a-public-cloud/) storage service which is suitable for unstructured data, a [private cloud](https://azure.microsoft.com/en-in/overview/what-is-a-private-cloud/) storage service which can be protected behind a company firewall for more control over data and a [hybrid cloud](https://azure.microsoft.com/en-in/overview/what-is-hybrid-cloud-computing/) storage service which blends public and private cloud services together for increased flexibility.

**Cloud Storage Benefits**

The [benefits](https://www.enterprisestorageforum.com/cloud-storage/on-premise-vs-cloud-storage.html) of cloud storage are similar to the other benefits of cloud computing, and they include the following:

* **Low costs** Because of economies of scale, public cloud vendors can offer extremely [low prices](https://www.enterprisestorageforum.com/storage-management/cost-savings-data-protection-driving-enterprises-to-cloud-based-data-storage.html) on storage. Public cloud storage services eliminate the need for organizations to buy and configure their own hardware, and they allow organizations to convert capital expenses (capex) into operational expenses (opex), which looks good on their financial reports.
* **Simplified management** When organizations use a cloud storage service, their IT personnel no longer need to configure, deploy and maintain the physical storage hardware. That can further reduce ongoing expenses.
* **Speed of deployment** Users can set up new cloud storage services and begin using them within just a few minutes. By comparison, it may take weeks or months to deploy new storage hardware in an enterprise data center.
* **Scalability** When they store data on-premises, organizations have to forecast their needs far into the future and purchase excess capacity in order to have enough space as their data continues to grow. But with cloud storage, more resources are automatically available as the organization needs them, with no need for overprovisioning.
* **Availability** While public cloud outages sometimes occur, in general, the leading public cloud vendors offer better uptime guarantees than most enterprises are able to achieve in their own data centers.
* **Security** People have different opinions about whether it's safer for enterprises to keep their sensitive data in their own data centers or store it in a public cloud. However, there is certainly an argument to be made that the large cloud vendors with their big budgets and knowledgeable are better equipped to protect against constantly evolving threats.

**Cloud Storage Challenges**

While the benefits of [migrating data](https://www.enterprisestorageforum.com/cloud-storage/managing-cloud-storage-migration.html) to a cloud storage service can be substantial, cloud storage also poses some potential [risks and pitfalls](https://www.enterprisestorageforum.com/cloud-storage/cloud-storage-pros-and-cons.html). And in many cases, the very benefits cloud storage offers are also possible challenges.

* **Security** While some people believe that the public cloud is [more secure](https://www.enterprisestorageforum.com/cloud-storage/cloud-storage-security.html) than in-house storage, many IT managers simply aren't comfortable trusting sensitive data to someone else's control. In the Cloudian survey, 62 percent of organizations listed security as one of their top cloud storage concerns, making it the most common cloud storage management challenge.
* **Compliance** In some cases, regulations require organizations to keep data in-house. In fact, 59 percent of survey respondents told Cloudian that they had data that cannot be migrated to the public cloud, and on average, 47 percent of their data must stay local.
* **Cost** While [lower costs](https://www.enterprisestorageforum.com/storage-management/cost-savings-data-protection-driving-enterprises-to-cloud-based-data-storage.html) are a big driver for cloud migration, the cloud isn't always cheaper than on-premises storage, especially if organizations already have legacy storage appliances that they could continue using for many more years. Also, many organizations that use the public cloud find it difficult to optimize costs. According to the Cloudian report, 55 percent of organizations expressed concerns about cloud storage costs.
* **Cloud storage management complexity** Forty percent of those surveyed by Cloudian expressed concerns about cloud storage management. While the public cloud eliminates the need to manage hardware, it can make it more difficult to make sure that the organization is enforcing its data management policies and following best practices.
* **Interoperability** With so many organizations adopting hybrid cloud storage strategies, interoperability with on-premise infrastructure is a key challenge for many enterprises. In fact, 40 percent of respondents in the Cloudian surveyed mentioned interoperability as a top concern.
* **Vendor lock-in** Once an organization begins using one cloud storage vendor, moving that data to a different vendor or bringing it back in-house becomes a costly and complicated operation. As a result, IDC found that nearly 20 percent of organizations had vendor lock-in concerns related to public cloud storage.
* **Connectivity** Inside their own data centers, organizations have very fast connections that allow them to access data quickly. But in the public cloud, they are often forced to rely on the public Internet, which is [much slower](https://www.enterprisestorageforum.com/storage-services/4-options-to-make-your-cloud-storage-faster.html) and can result in noticeable latency for users. In the Cloudian survey, 33 percent of organizations said they felt they needed to keep some data on premises because of concerns about performance, and in the DataCore survey, 32 percent cited performance as an issue.
* **Migration** Connectivity is also a concern when it comes to cloud storage migration. Transferring several terabytes of data to a cloud storage vendor over standard Internet connections might take weeks. In addition, organizations going through a cloud migration almost invariably run into problems in getting their cloud storage to function the same way that their on-premise storage did.

**What is SAN?**

Short for **s**torage **a**rea **n**etwork, **SAN** is a high-speed network of storage devices that also connects those storage devices with servers. It provides [block-level storage](https://www.webopedia.com/TERM/B/block-level-storage.html) that can be accessed by the applications running on any networked servers. SAN storage devices can include tape libraries and disk-based devices, like [RAID](https://www.webopedia.com/TERM/R/RAID.html) hardware.

The main functions of a storage area network (SAN) includes the following:

* A high-speed network of storage devices.
* Connects the storage devices with servers.
* Can be accessed by applications on networked servers.
* Particularly helpful in backup and disaster recovery.
* Uses networking protocols to span longer distances geographically.
* SAN can also simplify some management tasks.
* Offers flexibility, availability and performance.

**Unit 7 Scheduling in cloud**

Cloud computing is a new technology derived from grid computing and distributed computing and refers to using computing resources (hardware, software, and platforms) as a service and provided to beneficiaries on demand through the Internet [1]. It is the first technology that uses the concept of commercial implementation of computer science with public users [2]. It relies on sharing resources among users through the use of the virtualization technique. High performance can be provided by a cloud computing, based on distributing workloads across all resources fairly and effectively to get less waiting time, execution time, maximum throughput, and exploitation of resources effectively. Still, there are many challenges prevalent in cloud computing, Task scheduling and load balance are the biggest yet because it is considered the main factors that control other performance criteria such as availability, scalability, and power consumption.

The concept of scheduling in cloud computing refers to the technique of mapping a set of jobs to a set of virtual machines (VMs) or allocating VMs to run on the available resources in order to fulfil users’ demands. The aim of using scheduling techniques in cloud environment is to improve system throughput and load balance, maximize the resource utilization, save energy, reduce costs, and minimize the total processing time. Therefore, the scheduler should consider the virtualized resources and users’ required constraints to get efficient matching between jobs and resources. Each scheduling technique should be based on one or more strategies.

Scheduling problem is the problem of matching elements from different sets, which is formally expressed as a triple (E, S, O), where

• E is the set of examples, each of which is an instance of problem.

• S is the set of feasible solutions for the example.

• O is the object of the problem.

Scheduling problem can be further classified into two categories depending on object O: optimization problem and decision problem. An optimization problem requires finding the best solution among all the feasible solutions in set S. Different from optimization, the aim of decision problem is relatively easy. For a specified feasible solution s ∈ S, problem needs a positive or negative answer to whether the object O is achieved. Clearly, optimization problem is harder than decision problem, because the specified solution only compares with one threshold solution in decision problem, instead of all feasible solutions in optimization problem.

**Tasks scheduling algorithms overview**

Tasks scheduling algorithms are defined as the mechanism used to select the resources to execute tasks to get less waiting and execution time.

**Scheduling levels**

In the cloud computing environment there are two levels of scheduling algorithms:

First level: in host level where a set of policies to distribute VMs in host.

Second level: in VM level where a set of policies to distribute tasks to VM.

In this research we focus on VM level to scheduling tasks, we selected task scheduling algorithms as a research field because it is the biggest challenge in cloud computing and the main factor that controls the performance criteria such as (execution time, response time, waiting time, network, bandwidth, services cost) for all tasks and controlling other factors that can affect performance such as power consumption, availability, scalability, storage capacity, buffer capacity, disk capacity, and number of users.

**Tasks scheduling algorithms definition and advantages**

Tasks scheduling algorithms are defined as a set of rules and policies used to assign tasks to the suitable resources (CPU, memory, and bandwidth) to get the highest level possible of performance and resources utilization.

**Task scheduling algorithms advantages**

Manage cloud computing performance and QoS.

Manage the memory and CPU.

The good scheduling algorithms maximizing resources utilization while minimizing the total task execution time.

Improving fairness for all tasks.

Increasing the number of successfully completed tasks.

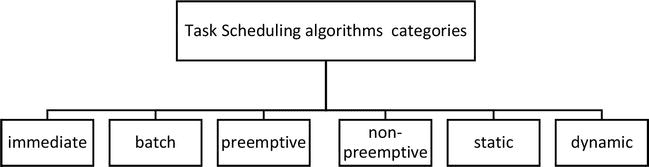
Scheduling tasks on a real-time system.

Achieving a high system throughput.

Improving load balance.

**Tasks scheduling algorithms classifications**

Tasks scheduling algorithms classified as in [Figure](https://www.intechopen.com/books/scheduling-problems-new-applications-and-trends/types-of-task-scheduling-algorithms-in-cloud-computing-environment#F1) given below



**Immediate scheduling**: when new tasks arrive, they are scheduled to VMs directly.

**Batch scheduling**: tasks are grouped into a batch before being sent; this type is also called mapping events.

**Static scheduling**: is considered very simple compared to dynamic scheduling; it is based on prior information of the global state of the system. It does not take into account the current state of VMs and then divides all traffic equivalently among all VMs in a similar manner such as round robin (RR) and random scheduling algorithms.

**Dynamic scheduling**: takes into account the current state of VMs and does not require prior information of the global state of the system and distribute the tasks according to the capacity of all available VMs [4, 5, 6].

**Preemptive scheduling**: each task is interrupted during execution and can be moved to another resource to complete execution [6].

**Non-preemptive scheduling**: VMs are not re-allocated to new tasks until finishing execution of the scheduled task

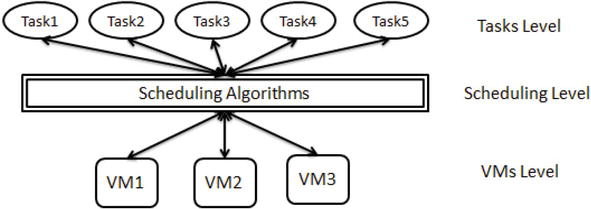
**Task scheduling system in cloud computing**

The task scheduling system in cloud computing passes through three levels [7].

The first task level: is a set of tasks (Cloudlets) that is sent by cloud users, which are required for execution.

The second scheduling level: is responsible for mapping tasks to suitable resources to get highest resource utilization with minimum makespan. The makespan is the overall completion time for all tasks from the beginning to the end [7].

The third VMs level: is a set of (VMs) which are used to execute the tasks as in Figure



Static tasks scheduling algorithms in cloud computing environment

**FCFS**

FCFS: the order of tasks in task list is based on their arriving time then assigned to VMs [3].

**Advantages**

Most popular and simplest scheduling algorithm.

Fairer than other simple scheduling algorithms.

Depend on FIFO rule in scheduling task.

Less complexity than other scheduling algorithms.

**Disadvantages**

Tasks have high waiting time.

Not give any priority to tasks. That means when we have large tasks in the begin tasks list, all tasks must wait a long time until the large tasks to finish.

Resources are not consumed in an optimal manner.

In order to measure the performance achieved by this method, we will be testing them and then measuring its impact on (fairness, ET, TWT, and TFT).

**SJF**

Tasks are sorted based on their priority. Priority is given to tasks based on tasks lengths and begins from (smallest task ≡ highest priority).

**Advantages**

Wait time is lower than FCFS.

SJF has minimum average waiting time among all tasks scheduling algorithms.

**Disadvantages**

Unfairness to some tasks when tasks are assigned to VM, due to the long tasks tending to be left waiting in the task list while small tasks are assigned to VM.

Taking long execution time and TFT.

**MAX-MIN**

In MAX-MIN tasks are sorted based on the completion time of tasks; long tasks that take more completion time have the highest priority. Then assigned to the VM with minimum overall execution time in VMs list.

**Advantages**

Working to exploit the available resources in an efficient manner.

This algorithm has better performance than the FCFS, SJF, and MIN-MIN algorithm.

**Disadvantages**

Increase waiting time to small and medium tasks; if we have six long tasks, in MAX-MIN scheduling algorithm they will take priority in six VMs in VM list, and short tasks must be waiting until the large tasks finish.

Unfairness to some or most small and medium tasks when tasks are assigned to VM.

**Round Robin scheduling algorithm (RR)**

Round Robin algorithm is considered as one of the simplest, conventional and most used scheduling algorithms which works exceptionally better for timesharing systems. It distributes the load equally to all the resources. It works very similar way in cloud computing as it does in process scheduling. The working includes a circular queue and a fixed time unit called quantum. Each individual job’s execution takes place only within this quantum. The whole process goes like this: first of all the first process from the queue is picked by CPU scheduler after which it sets a timer to interrupt after one quantum and then finally the process is dispatched. In case if the job does not achieve in one allotted quantum, it returns to the queue and wait for the following round. The massive benefit is that jobs are executed in a sequence turn wise and there is need to wait for the previous job to get complete. Therefore, there is no starvation issue. But the dark side is that if the workload is heavy and queue is fully loaded, it takes a lot of time to accomplish all the jobs and moreover, a perfectly suitable time quantum is hard to decide.

**Priority-based Job Scheduling Algorithm**

In order to reduce the makespan time, another Priority based Job Scheduling (PJSC) Algorithm for cloud computing has been proposed by Ghanbari.S, which is pre-emptive in nature where each process in the system is based on the priority and priority is allowed to run. The highest priority job can run first whereas lower priority job can be made to wait. Equal-Priority processes are scheduled in FCFS order. The drawback of this algorithm is starvation of a process.

**RASA Task Scheduling Algorithm**

RASA is a new scheduling algorithm composed of two traditional techniques- Max-min and min-min. A min-min strategy is used to execute small tasks before large tasks and Max- Min strategy is applied to avoid the delays in large tasks execution. Both the ways are used for tasks and alternative exchange ends up in the consecutive execution of a tiny low and an outsized task on totally different resources, therefore, ignoring the waiting time of the tiny tasks in Max-min rule and therefore the waiting time of the massive tasks in the Min-min algorithm.

**Real-Time Scheduling**

There are emerging classes of applications that can benefit from increasing timing guarantee of cloud services. These mission critical applications typically have deadline requirements, and any delay is considered as failure for the whole deployment. For instance, traffic control centres periodically collect the state of roads by sensor devices. Database updates recent information before next data reports are submitted. If anyone consults the control centre about traffic problems, a real-time decision should be responded to help operators choose appropriate control actions. So more flexible, transparent and trust-worthy service agreement between cloud providers and users is needed in future. Given the above analysis, the ability to satisfy timing constraints of such real-time applications plays a significant role in cloud environment.

A real-time scheduler must ensure that processes meet deadlines, regardless of system load or make span. Priority is applied to the scheduling of these periodic tasks with deadlines. Every task in priority scheduling is given a priority through some policy, so that scheduler assigns tasks to resources according to priorities. Based on the policy for assigning priority, real-time scheduling is classified into two types: fixed priority strategy and dynamic priority strategy. Fixed priority scheduling is that all instances of one task have the same priority. Dynamic priority assignment is more efficient than the fixed manner, since it can fully utilized the processor for the most pressing tasks. The priorities change with time, varying from one request to another or even during the same request.

These contents have been compiled from the study material available free on various websites.