

# Cloud Computing

## Chapter 1 : Basics of Cloud Computing

### Q. 1 What is Cloud computing?

Ans. :

In Computer science, Cloud computing describes a type of outsourcing of computer services, similar to electricity, in which electricity supply is outsourced. Users can simply pay for it. They do not need to worry where the electricity come from, how it is made, or transported. Every month, they pay for what they consumed. The idea behind Cloud computing is similar. The user can simply use storage, computing power, or specially crafted development environments, without having to worry how these works internally.

Cloud computing is usually Internet-based computing. The cloud is a metaphor for the Internet on how the internet is described in computer network diagrams; which means it is an abstraction hiding the complex infrastructure of the internet. It is a style of computing in which IT-related capabilities are provided "as a service", allowing users to access technology enabled services from the Internet ("in the cloud") without knowledge of control over the technologies behind these servers.

Cloud computing is an information technology (IT) paradigm, a model for enabling all over access to shared pools of configurable resources (such as computer networks, servers, storage, applications and services), which can be rapidly provisioned with minimal management effort over the Internet. Cloud computing allows users and enterprises with various computing capabilities to store and process data either in a privately-owned cloud, or on a third-party server located in a data center, thus making data-accessing mechanisms more efficient and reliable.

Cloud computing relies on sharing of resources to achieve consistency and economy of scale, similar to a utility. Cloud computing allows companies to avoid or minimize up-front IT infrastructure costs. As well, third-party clouds enable organizations to focus on their core businesses instead of expending resources on computer infrastructure and maintenance. Cloud computing allows enterprises to get their applications up and run faster, with improved manageability and less maintenance, and that it enables IT teams to more rapidly adjust resources to meet fluctuating and unpredictable business demand. Cloud providers typically use a "pay-as-you-go" model. This could lead to unexpectedly high charges if administrators are not familiarized with cloud-pricing models.

In 2009 the availability of high-capacity networks, low-cost computers and storage devices as well as the widespread adoption of hardware virtualization, service-oriented architecture, and autonomic and utility computing led to a growth in Cloud computing. Companies can scale up as computing needs increase

and then scale down again when demands decrease. In 2013 it was reported that Cloud computing had become a highly demanded service or utility due to the advantages of high computing power, cheap cost of services, high performance, scalability, and accessibility as well as availability.

### Q. 2 Define Cloud computing as per NIST.

Ans. :

"Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction."

### Q. 3 Explain the term On-premise.

Ans. :

1. Cloud is a distinct and remotely accessible environment. Hence cloud provides an option for the IT resources deployment on client side.
2. When an IT resource is hosted in an IT enterprise within its organizational boundary, it is considered to be located at the premises of that enterprise. It is known as on-premise.
3. The term on-premise can be stated as "on the premises of an IT enterprise environment which is not cloud-based". The on-premise IT resource cannot be cloud based.
4. Important points to be noted :
  - (i) A cloud-based IT resource can be accessed and communicated by an on-premise IT resource.
  - (ii) It is possible to move the on-premise IT resource on cloud by converting it into cloud-based IT resource.
  - (iii) Sometimes, few redundant IT resources may exist in both on-premise as well as cloud based environments.

### Q. 4 Explain the types of Scaling.

Ans. :

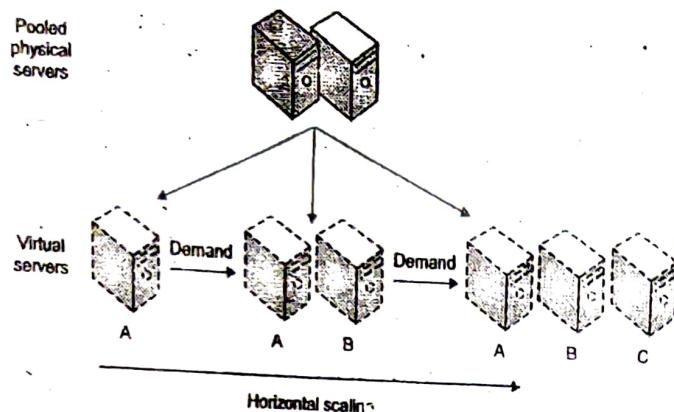
Scaling is the ability of IT resource to handle continues increasing or decreasing demands of usage.

There are two types of scaling

#### 1. Horizontal Scaling

In this type of scaling, same type of IT resources are allocated or released as per the requirement. The process of allocating the IT resources horizontally is known as scaling out. The process of

releasing the IT resources horizontally is known as scaling in. Usually in cloud environments, horizontal scaling is considered as a common form of scaling.



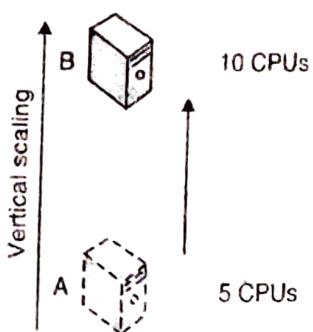
**Fig. 1.1 : Horizontal scaling**

In Fig. 1.1 IT resource 'Virtual Server A' is scaled out by adding other same types of virtual servers say B and C.

## 2. Vertical Scaling

In this type of scaling, an existing IT resource is upgraded or completely replaced by another IT resource with higher or lower capacity. The process of up gradation or complete replacement of an IT resource by another IT resource with higher capacity is known as **scaling up**. The process of up down gradation or complete replacement of an IT resource by another IT resource with lower capacity is known as **scaling down**.

In cloud environments, vertical scaling is generally not much used due to more **downtime** which is required for scaling up or down. In Fig. 1.2, IT resource is a virtual server which has initially five CPUs is scaled up by replacing it with a more powerful IT resource having ten CPUs for data storage.



**Fig. 1.2 : Vertical Scaling**

## Q. 5 Differentiate between Horizontal and vertical scaling.

**Ans. :**

### Difference between Horizontal and vertical scaling

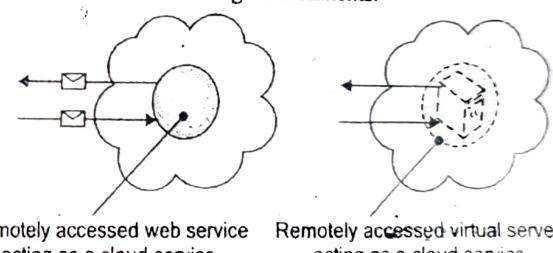
Parameter	Horizontal Scaling	Vertical Scaling
Cost	Less expensive because of the use of commodity hardware components.	More expensive because of the need of specialized servers.

Parameter	Horizontal Scaling	Vertical Scaling
Availability	Usually all the time, IT resources are available instantly.	Not all the time, IT resources are available instantly.
Additional setup	Resource replication and automated scaling.	Additional setup is normally needed.
Additional resources	Additional IT resources are required.	No additional IT resources are required.
Limitation	It is not limited by the capacity of hardware.	It is limited by the capacity of hardware.

## Q. 6 Explain the term Cloud Service.

**Ans. :**

For security reason, some IT resources which reside within cloud are made unavailable for remote access. For example, a database or a physical server which exists within a cloud may accessible only to IT resources which are present in the same cloud. A cloud service can be considered as any IT resource which can be accessed remotely via a cloud. It is totally different and broad concept than the other IT fields. A cloud service can be a simple Web-based software application which is invoked with the help of messaging protocol or a virtual server (remote access point) for administrative tools for big environments.



**Fig. 1.3 : Cloud Service**

## Q. 7 Explain cloud computing applications.

**Ans. :** Applications of cloud computing

### 1. Online File Storage

There are number of examples of sites like MediaFire, megaupload, hotfile, 4Shared, rapidshare, yourfilehost which help to host files such as documents, images, presentation, videos, etc. The interfaces provided are very simple to use, and these sites help users to upload and download files. 200GB of storage space is given to users with file size of 2GB.

### 2. Photo Editing Software

There are some famous online photo editing software such as Picnik, Pixlr, etc. which are absolutely free. Such software provides features like image cropping, changing size of image, rotation depending upon degrees, special effects, addition as well as editing features. Paint tools and other adjustment features are also provided by some of them. It is possible to edit the brightness

and contrast, and users can layer the images. For this purpose Pixlr, provides various high-level, complex features which are easy to use.

### 3. Digital Video Software

A free application named Hulu is available for videos which are free on the web. It is possible for cloud users to download popular movies, television shows as well as documentaries and enjoy them on the web-browser like Google Chrome. Hulu is a joint venture of three firms viz. - Fox Entertainment Group, NBC Universal and ABC Inc. There are also some other video sites such as WatchMoviesOnline, YouTube, Google video, etc.

### 4. Twitter-Related Applications

There are number of twitter-related applications like bit.ly which can convert lengthy URL in a short unique URL. User is redirected to the real website after clicking this URL. But in some cases it may be dangerous as hackers can add malicious attachments or programs with it. There is an agreement of Ly with Twitter, to let twitter users for using shortened URLs. One more site named Twitpic is present which lets the user to upload pictures to be linked from twitter. For this purpose it uses twitter's login, generates small URLs which can be called from twitter's microblogging.

### 5. Creating Image Album

To host images on the web, number of applications are available such as flickr, photobucket, webshots, imagebam and ziddu. These sites are under the roof of cloud which allows users to arrange images into albums as well as create slideshows for free.

### 6. Web Application for Antivirus

Cloud Antivirus is an example of web application for antivirus. Panda Security - a Spanish company is creator of this application which offers functionality to save the system from virus and also detects and fixes a system which is already infected with malware or any different forms of computer viruses. PC World rates this application as the best free antivirus application. This application can be downloaded into the systems, and the inspection of malware can be done by sending the information regarding the file to the cloud data-center.

### 7. Presentation Software

For presentation purpose, an online free application is available named Slidrocket. User can import Microsoft's PowerPoint presentations using this application. It is possible to access the presentations from anywhere in the world as it is a web-based cloud application. But in the free version there is no facility to edit presentations offline for the cloud users.

### 8. Word Processing Application

Writeboard is an online word processing as well as document editing application. This application allows multiple users to access a document and simultaneously edit and save the document, but the document will have different versions. This application does not provide facility to import the word files.

### 9. Finding a Way on the Map

Searching directions as well as locations on the web is one more important area where cloud applications are useful. There are number of famous sites such as mapquest, Google Maps, and Yahoo Maps. These are considered as the most useful free online applications which are useful for millions of users in several ways by giving information about direction and paths and help users to reach their destinations.

### 10. E-Commerce Software

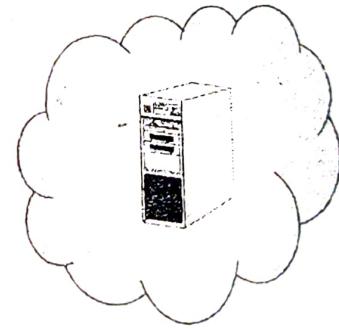
Cloud based e-application helps users as well as e-business to provide instant response to market opportunities and challenges which are faced by e-commerce. Irrespective of selection of computing solution, the free online application must be able to access customer data, product data, fulfillment systems as well as several other operational systems so as to support e-commerce. The e-commerce applications which are Cloud-based offer IT firms, and business leaders to evaluate new opportunities without the need of big amount of upfront investment.

### Q.8 Explain involvement of Cloud Computing in an organization.

**Ans.:**

Cloud computing cannot be considered as one-size-fits-all affair. As the hardware and software configuration used in an enterprise is different from the enterprise's regular affair; the needs of cloud computing will also be different.

### When You Can Use Cloud Computing



Compute clouds allows you to access applications maintained on a providers equipment

Fig. 1.4 : Compute Cloud

There are number of factors which decides whether to use cloud computing or not :

1. Cost/benefit ratio
2. Speed of delivery
3. How much capacity will be used
4. Whether the data is regulated
5. The corporate and IT structure of organization

There may be occasions when the requirement of organization is an ideal match for cloud computing. But sometimes cloud computing may not be a good match to the requirements of

organization. Here we will see both for what organization can use clouds, and when organization should steer clear of them.

## Scenarios

Cloud computing has 3 major implementations. The way organizations are using cloud computing is considered as rather different at a granular level, but the uses normally fall into one of these three specified solutions.

### Compute Clouds

An access is given to highly scalable, inexpensive, on-demand computing resources by the compute cloud which execute the code given to them.

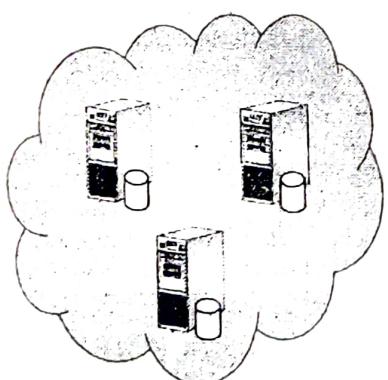
There are three examples of compute clouds :

1. Amazon's EC2
2. Google App Engine
3. Berkeley Open Infrastructure for Network Computing (BOINC)

Compute clouds are considered as maximum flexible regarding their offerings and it is possible to use them for sundry purposes; it just depends on the access of application needed by the user. These applications are suitable for any organization irrespective of its size, but it has been seen that the big organizations might be at a disadvantage since the standard management, monitoring, and governance capabilities are not offered by those applications which are necessary for the organizations.

### Cloud Storage

Cloud storage was one of the initial offerings by cloud and it still remains a popular solution. Cloud storage is considered as a big world. Nowadays there are more than hundred vendors offering cloud storage. For maintaining files off-site, it is a perfect solution. In cloud environment security and cost are the important aspects and vary greatly among vendor to vendor. Currently, Amazon's S3 is considered as best.



Cloud storage allows you to store your data on a vendor's equipment

Fig 1.5 : Compute Storage

## Cloud Applications

There is difference between Cloud applications and compute clouds in that Cloud applications uses software applications which are basically rely on cloud infrastructure. Cloud applications are versions of SaaS (Software as a Service) and contain things like web applications which are sent to users with the help of a browser or application such as Microsoft Online Services.

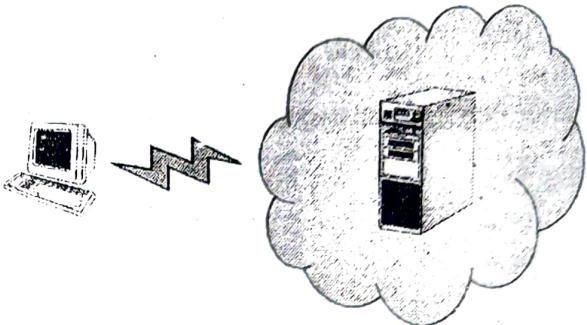


Fig 1.6 : Compute Cloud

The important feature of cloud applications is that they usually eliminate the requirement to install and run the application on computer of user, which minimizes the burden regarding software maintenance, ongoing operation as well as support.

Some cloud applications include :

1. Peer-to-peer computing such as BitTorrent, Skype etc.
2. Web applications such as MySpace, YouTube, etc.
3. SaaS such as Google Apps
4. Software plus services such as Microsoft Online Services.

### Q. 9 Define advantages of Cloud computing.

**Ans. : Advantages of cloud**

1. **Public Cloud** : In this cloud, general public can access systems and services easily.
2. **Private Cloud** : In this cloud the systems and services are accessible within the organization only.

### Q. 10 Explain the term Reduced Investment and Proportional Cost.

**Ans. :**

Shopping Malls purchase products in bulk with lower prices from companies and make them available to general public in lower prices. Using the same logics, the public cloud providers purchase IT resources in bulk with lower prices and make them available to consumer IT companies with attractive low priced leasing packages. This opens the door for various small scale organizations to gain access to costly strong infrastructure without any need to purchase it by them.

For the consumer organizations, the main goal of taking cloud service is to reduce the up-front IT investments, for example hardware and software purchases and also the ownership costs. Use of cloud replaces the anticipated higher capital expenditures by the lower operational expenditure.

This helps to eliminate or minimize the up-front financial costs and allows enterprise owners to start small and accordingly increase IT resource allocation as per the requirement. The saved cost can be redirected to the core business investment. In its most basic form, chance to minimize costs are derived from the process of deployment and operation of large-scale data centres by the various major providers of cloud.

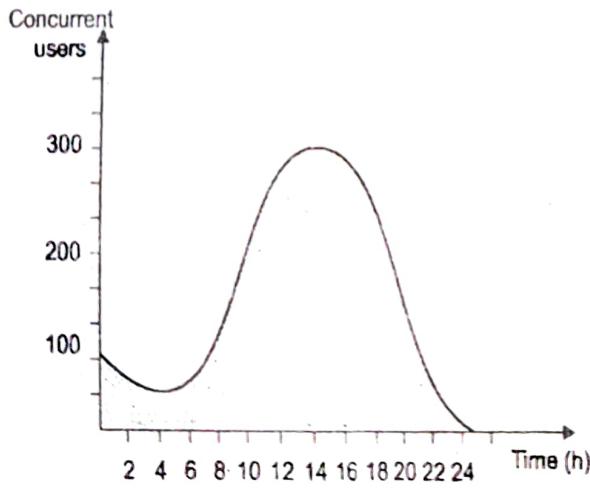
Usually these data centres are deployed in such destinations where availability of IT professionals, network bandwidth, and offices are available at low cost which helps to support both capital and operational savings. The same principle is applied to operating systems, middleware or platform software, and application software. Pooled IT resources are made available which can be utilized by the various cloud consumers to implement maximum possible utilization.

#### **Q. 11 Explain the term Increased Scalability.**

**Ans. :**

Cloud provides pools of IT resources with the support of technologies which can control them collectively. The IT resources can be instantly and dynamically allocated as per the requirement. This allows the cloud consumers to scale their cloud-based IT resources to manage the fluctuations in processing and peaks these resources automatically or manually. In the same manner, it is possible to release cloud-based IT resources automatically or manually if there is decrease in processing demand.

A simple example of fluctuations in the usage demand in one day is provided in Fig. 1.7. The built-in feature of clouds to provide flexible option of scaling of IT resources gives effective and proportional cost benefit.



**Fig. 1.7 : Fluctuations in usage demand**

Besides the financial benefit, it also increases the ability of business to fulfil the unpredictable usage demands of IT resources which avoids a potential loss which may occur because of lack of this ability.

#### **Q. 12 Explain the term "Increased Availability and Reliability".**

**Ans. :**

The benefit of business is completely depends upon the availability and reliability of IT resources. If any IT resource is not available then its unavailability for customers may affect the revenue generation of the business. Also runtime failure of any IT resource during high level usage period may lead to same loss. These issues affect the service provided to the customers.

The main advantages of cloud based environment is its intrinsic ability to give extensive support for increasing the availability of IT resource to decrease or even remove outages. For improving its reliability so as to decrease the effect of various runtime failure conditions :

- A cloud based IT resource can be accessible for longer periods of time. For example near about 22 hours in a day. A cloud based IT resource having increased reliability is able to better avoid and recover in failure condition. The cloud environments have modular architecture which provides a strong failover support to enhance the reliability.

#### **Q. 13 Explain the term "Increased security Vulnerability".**

**Ans. :**

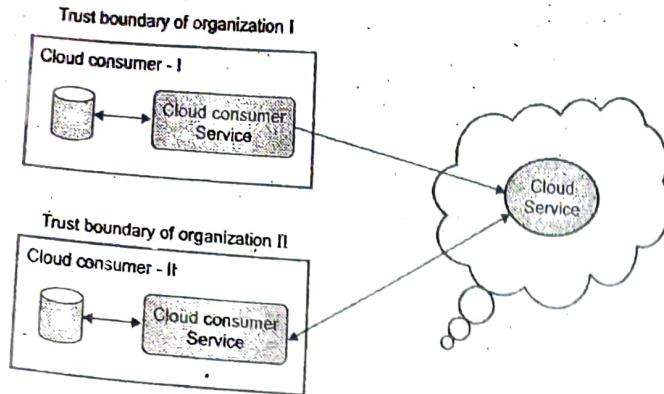
When important business data is shifted to cloud, the concern of security get arise. The usage of various IT resources remotely, needs a strong trust by the cloud consumer to involve the external cloud. It is little bit complicated to establish such a strong security architecture without introducing vulnerabilities, unless both the cloud consumers as well as providers support the same or compatible security frameworks which is mostly difficult with public clouds. Another aspect regarding security is the privileged access of cloud provider to business data of cloud consumer. The level of security is now restricted to the security controls and policies implemented by both the cloud consumer as well as provider. As the cloud provides sharing feature, the security of consumer data is in trouble.

The increased exposure of consumer data may provide some other malicious issues to cloud consumers which may be human or automated with more opportunities to attack IT resources and steal or damage consumer's important business data.

#### **Q. 14 Explain the term "Reduced Operational Control".**

**Ans. :**

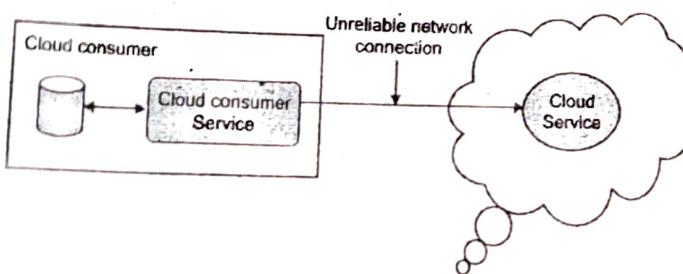
In general there is level of governance control allotted to cloud consumers. This level is always lower than the on-premise IT resources. This low level of governance control may include some risks associated with the operative method of cloud provider, and also outside connections which are necessary for the communication of cloud provider and consumer.



**Fig. 1.8 : Reduced Operational Governance Control**

Consider the following examples :

1. If the cloud provider is not reliable, then he may not maintain the guarantee which is made in the SLA (Service-level agreement,) which has been published for its cloud services. This affects the cloud consumer solutions which are dependent upon these cloud services.
2. If the geographic distance between the cloud consumer and cloud provider is longer, then there is need of additional network hops which may lead to fluctuating latency and potential bandwidth constraints.
3. The latter scenario is shown in Fig. 1.9.



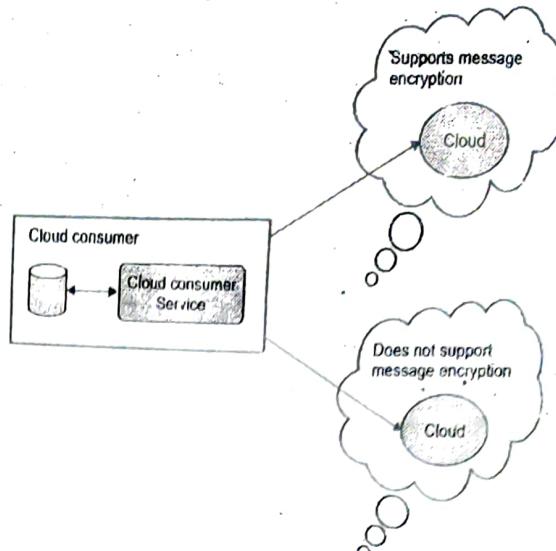
**Fig. 1.9 : An unreliable network connection**

**Q. 15 Explain the term “Limited Portability between Cloud Providers”.**

**Ans. :**

In the industry of Cloud computing, there is lack of established industry standards. Hence the public clouds are usually proprietary to large extents. Some cloud consumers have customized built solutions with dependencies on such types of proprietary environments. In such case it becomes very difficult for a cloud consumer to move from one cloud provider to another.

Portability is considered as a measure which helps to determine the effect of moving cloud consumer IT resources and data from one cloud provider to another. A cloud consumer's application may have low level of portability in moving to another cloud as second cloud provider does not support the same environment as of the first.



**Fig. 1.10 : Limited Portability between Cloud Providers**

**Q. 16 Explain the term “Multi-Regional Regulatory and Legal Issues”.**

**Ans. :**

Third-party cloud providers always like to establish such data centers which are affordable or convenient as per the geographical locations. Sometimes the cloud consumers do not know the actual physical locations of public clouds from which they are accessing the IT resources. This can pose serious legal issues regarding the industry or local government rules and regulations that state policies of data privacy and storage.

Another important legal issue is regarding the accessibility and disclosure of data. Different countries have different laws regarding the disclosure to extent of data. The cloud consumer has responsibility of the security, integrity, and storage of their own data, even though it is held by cloud provider.

**Q. 17 Elaborate characteristics of cloud computing.**

**Dec. 16**

**Ans. :**

Cloud computing exhibits the following key characteristics

### 1. Agility

For organizations it may be improved, as Cloud computing may increase users flexibility with re-provisioning, adding, or expanding technological infrastructure resources.

### 2. Cost reductions

They are claimed by cloud providers. A public-cloud delivery model converts capital expenditures (e.g., buying servers) to operational expenditure.

### 3. Device and location independence

It enables users to access systems using a web browser regardless of their location or what device they use (e.g., PC, mobile phone). As infrastructure is off-site (typically provided by a third-party) and accessed via the Internet, users can connect to it from anywhere.

**4. Maintenance**

The Maintenance of Cloud computing applications is easier, because they do not need to be installed on each user's computer and can be accessed from different places (e.g., different work locations, while travelling, etc.).

**5. Multitenancy**

It enables sharing of resources and costs across a large pool of users thus allowing for centralization of infrastructure in locations with lower costs (such as real estate, electricity, etc.)

**6. Performance**

It is monitored by IT experts from the service provider, and consistent and loosely coupled architectures are constructed using web services as the system interface.

**7. Resource pooling**

It is the process of combining resources to serve multiple consumers using a multi-tenant model with different physical and virtual resources dynamically assigned and reassigned according to user demand.

**8. Productivity**

It may be increased when multiple users can work on the same data simultaneously, rather than waiting for it to be saved and emailed. Time may be saved as information does not need to be re-entered when fields are matched, nor do users need to install application software upgrades to their computer.

**9. Reliability**

It improves with the use of multiple redundant sites, which makes well-designed Cloud computing suitable for business continuity and disaster recovery.

**10. Scalability and elasticity**

Scalability and elasticity via dynamic ("on-demand") provisioning of resources on a fine-grained, self-service basis.

**11. Security**

Security can improve due to centralization of data.

**Q. 18 Write a short note on SaaS.****Ans. : SaaS :**

This model is basically used to provide software application as a service to the end users. It refers to any specific software which has been deployed on a host server and can be accessed through Internet. There are various types of SaaS applications available which are listed below :

1. Billing and invoicing system
2. Help desk applications

**3. Customer Relationship Management (CRM) applications****4. Human Resource (HR) solutions**

It is not necessary that all the SaaS applications should be customizable, for example Microsoft Office Suite. For customized application development, SaaS provides Application Programming Interface (API) to the user.

**Q. 19 Write characteristics of SaaS model.****Ans. :**

SaaS makes available various types of software over the Internet. Vendor has the responsibility to maintain the software. The license of the software is available either subscription based or on usage based. And on the basis of recurring it is charged. User do not have any responsibility of maintenance of the software, hence it is cost effective.

They are anytime anywhere available on demand. They can be scaled up or down on demand. Automatically up gradation and updation is available. SaaS provides shared data model. Hence, several users can share single instance of infrastructure at a time. There is no need of hard coding for functionality for any specific user.

**Q. 20 Explain advantages of SaaS model.****Ans. : Advantages of SaaS****1. Modest software tools**

For the SaaS application deployment, there is no need of software installation on client side. This results in the following benefits :

- (i) No need of any complex software packages to be installed at client side.
- (ii) Zero risk of configuration at client side.
- (iii) Low distribution cost.

**2. Efficient use of software licenses**

Single license copy can be used on multiple computers which may be at different locations. This lowers the licensing cost. As well, there is no need of license servers since the software runs in the infrastructure of service provider.

**3. Centralized management and data**

The data is saved centrally by the cloud provider. This helps to manage the data effectively. But sometimes to maintain the redundancy and reliability, the data may be in decentralized manner.

**4. Platform responsibilities managed by providers**

All platform responsibilities are performed by the cloud provider. These responsibilities includes backups of data, maintenance of the system, security, hardware updates, power management, etc. No need for the customer to bother about them.

## 5. Multitenant solutions

Multitenant solutions enable several users to share single instance of multiple resources in virtual isolation. It is possible for the customer to customize the application without disturbing the core functionality.

### Q. 21 List disadvantages of SaaS model.

Ans. :

1. Browser-based risks.
2. Network dependence.
3. Lack of portability between SaaS clouds.

### Q. 22 Write a note on multitenant nature of SaaS solutions.

Ans. :

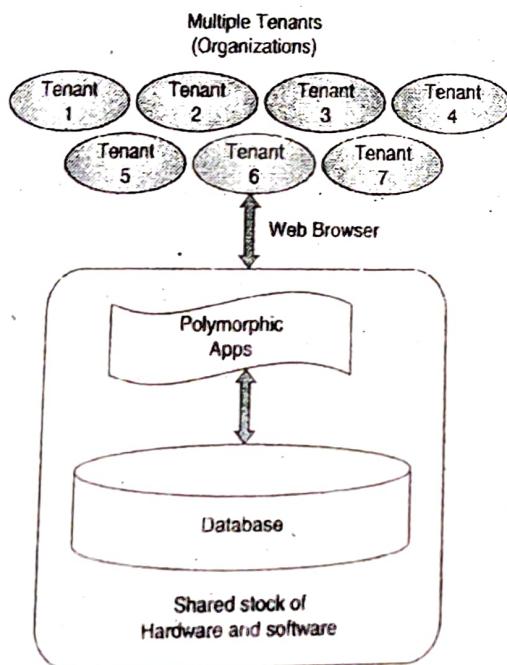


Fig. 1.11 : A Multi-tenant application

It enables the ability to independently update the core system, modify the core application, and customize tenant-specific components with virtually no risk affecting others. In SaaS environments it's a nightmare to manage a vast, frequently changing set of actual database structures on behalf of each application and tenant. It is suggested to use "virtual" database structures using a set of Meta-data, data, and pivot tables, as shown in Fig. 1.12.

In order to achieve cost efficiencies in delivering same applications to various sets of users it is a vital and obvious choice that an increasing number of applications are Multi-tenant instead of single-tenant. A Multi-tenant application should be able to satisfy the needs of multiple sub-organizations or sections within the organization (multiple tenants), using the single, shared stack of software and hardware resources and staff needed to manage.

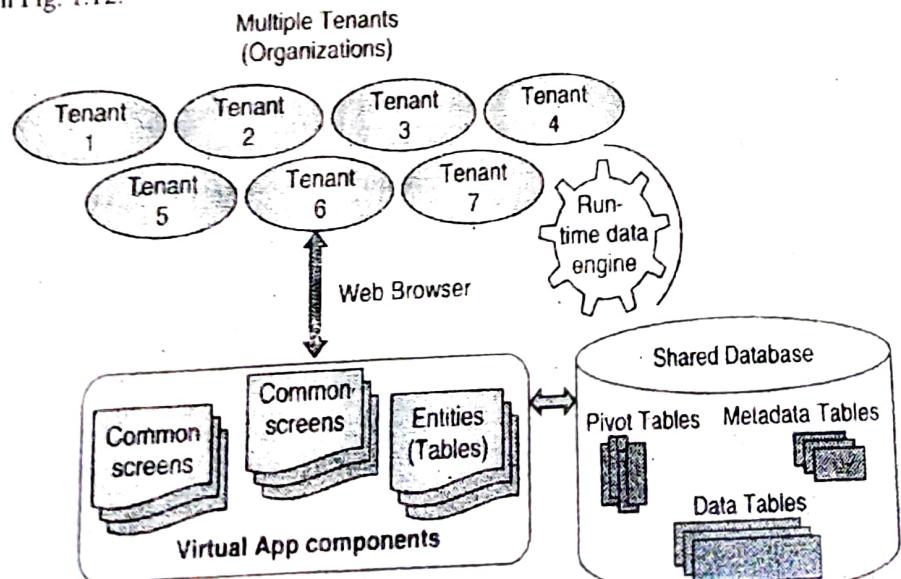
A Multi-tenant application shares a single stack of resources both hardware and software to cater multiple tenants (Organizations, sub-organizations , sections etc)

### Meta-Data-Driven Reference Architecture

A traditional static application is not capable of addressing unique challenges of multi-tenancy. A Multi-tenant application should be dynamic in nature, or polymorphic to fulfil the specific expectations of various tenants and their users. Application components need to be generated at runtime from meta-data i.e. data about the application itself.

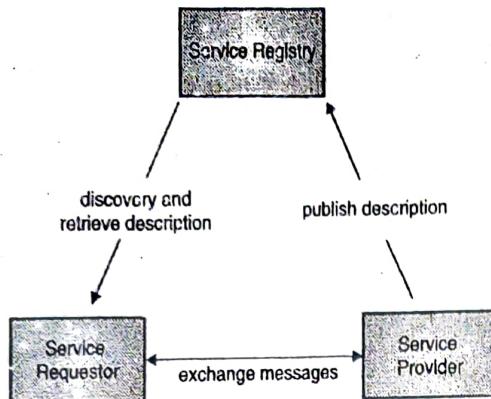
It should be a well-defined meta-data driven architecture (Fig. 1.11), with a separate component for :

1. Runtime application data
2. Meta-data that describes the base functionality of an application
3. Meta-data that corresponds to each tenant specific data and customizations



### **Q. 23 Explain SOA in detail.**

**Ans. : SOA :**



**Fig. 1.13 : SOA Architecture**

A service-oriented architecture is basically a group of services that interact with each other. The interaction can engage either transfer of data or synchronization of one or more services into particular action. SOA can be referred as a framework which designs software systems with the help of software architecture that observes each and every unit as services.

- 1. Service :** A service is defined as a function which is definite and independent of framework or condition of other services.
- 2. Architecture :** Architecture is defined as a basic management of a system represented in its elements, their correlations with each other and to the surroundings, and the rules directing the plan and development.

SOA is defined as architecture useful for constructing business applications as a set of freely joined black-box units arranged to convey a well-defined level of service by connecting different business processes together. It is a structure for designing software systems with the help of a software architecture that observes all units as services. A service is a task which is precise, self-sufficient, and does not require any context or state of other services. Each SOA consists of three building blocks which are as follows :

#### **1. Service provider**

It generates a web service and gives its information to the service registry. Every provider discusses upon a lot of questions like which service is to be discovered, which service is more important, precautions taken for services, which service is easily available, what should be the price of service etc. The provider also decides the category of the service that must be selected for a particular broker service and arrangement of trading partner agreements that are necessary for service usage.

#### **2. Service broker, service registry or service repository**

The main purpose of service registry is to provide the information about the available web service to every possible

requester. The service registry that employs the broker chooses the range of the broker. Generally public brokers are accessible all over but private brokers are only accessible to an inadequate quantity of public.

#### **3. Service requester/consumer**

It establishes access in the broker registry with the help of different operations and then combines them to the service provider so as to invoke one of its web services. Any service the service-consumers require must be taken from brokers, combine it with particular service and then applied. They can use several services if the service offers several services. The service consumer-supplier association is ruled by a consistent service contract that consists of three parts which are follows :

1. Business
2. Functional and
3. Technical

Service composition models have two comprehensive and advanced architectural techniques which are follows :

1. Choreography and
2. Orchestration.

Minor level enterprise combination models that are not limited to a specific architectural technique persist to be applicable and suitable in SOA design.

### **Q. 24 Explain the PaaS model.**

**Ans. :**

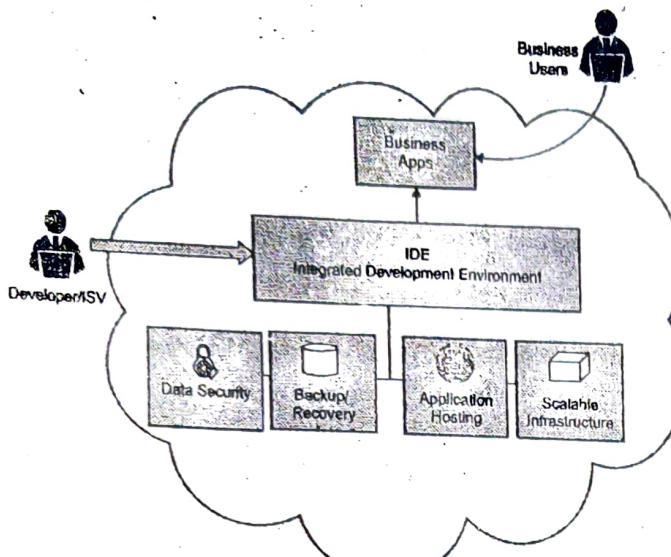
Platform-as-a-Service provides a ready-made runtime environment for different types of applications. It also offers various tools regarding development and deployment which are necessary to develop applications. PaaS provides a very important feature of point-and-click tools which is very useful for the non-developers to create various types of web applications as per their requirement.

Examples of PaaS offering vendors :

- (i) App Engine of Google
- (ii) App Engine Force.com

Developer can log on to these websites to take help of built-in API for the creation of web-based applications. The main disadvantage of using PaaS is that, the developer gets locked in with a specific vendor. For example if an application is created in Python using API of Google and App Engine of Google, then it will now work in that environment only.

Fig. 1.14 shows that how PaaS provides an API and development tools to the developers and how it guides the user to access the business applications.



**Fig. 1.14 : PaaS**

#### **Q. 25 Explain advantages of PaaS model**

**Ans. :**

##### **Advantages of PaaS**

###### **1. Low Overhead of Administration**

As administration is the responsibility of cloud provider, it is not headache of customer.

###### **2. Less Ownership Cost**

For the customer there is no need to purchase expensive hardware, servers, data storage and power.

###### **3. Scalable solutions**

It is very simple and flexible to scale the available resources up or down automatically, depending on their requirement.

###### **4. Latest system software**

It is the responsibility of the cloud provider to maintain and keep advanced and latest software versions.

---

#### **Q. 26 Explain the disadvantages of PaaS model.**

**Ans. :**

##### **Disadvantages of PaaS**

###### **1. Lack of portability**

Lack of portability between PaaS clouds.

###### **2. Resource constraints**

Event based processor scheduling which constitutes resource constraints on the applications, i.e. time limitation to answer a request.

###### **3. Dependency on network**

PaaS applications are completely reliant on network, hence they have to explicitly use cryptography and manage the upcoming security exposures.

---

#### **Q. 27 Explain the IaaS model.**

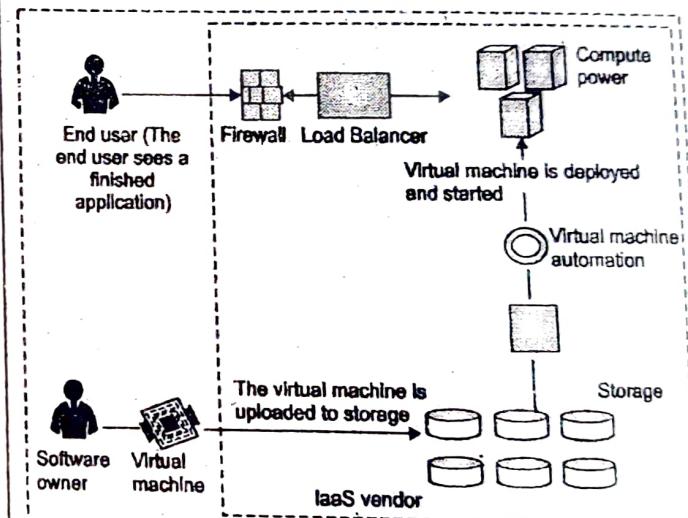
**Ans. :**

Infrastructure-as-a-Service is a model which provides access to primary resources like physical machines, virtual machines, virtual storage, etc.

In addition to these resources, the IaaS also provides following facilities :

- (i) Virtual machine disk storage
- (ii) Software bundles
- (iii) Virtual local area network (VLANS)
- (iv) IP addresses
- (v) Load balancers

The concept of server virtualization is used to make available these resources to the user. The extent of access rights gives feeling of owner of these resources to the end user.



**Fig. 1.15 : IaaS**

---

#### **Q. 28 Explain advantages of IaaS.**

**Ans. :**

IaaS helps the cloud provider to freely place the infrastructure over the Internet in a cost-effective manner. Through administrative level access to VMs, a complete control of the computing resources is provided.

Using the administrative access, the customer can access computing resources in the following way:

- (i) If the end user wants to run the virtual machine or store data on cloud based server, then he issues an administrative command to cloud provider.
- (ii) Administrative command is given by the customer to virtual machines to start the service of web server or to install new applications.

**The renting of computer hardware is very flexible as well as efficient**

Number of IT resources like virtual machines, IP addresses, storage devices, monitoring services, bandwidth, firewalls, etc. made available to the customers on lease. The payment of the service is depending upon the time for which the resources are retained by the customer. Also using the administrative access to virtual machines, number of software can be executed by the customer.

**Portability and interoperability is provided with legacy applications.**

A legacy can be maintained between applications and workloads within IaaS clouds. For example the network applications like web or email server which generally run on customer side server can be run on VMs in IaaS cloud.

**Q. 29 Explain disadvantages of IaaS.**

**Ans. :**

**Disadvantages of IaaS**

**1. Compatibility with legacy security vulnerabilities**

As the legacy software of customer is run in provider's infrastructure, the software get exposes to all of the security vulnerabilities.

**2. Virtual Machine sprawl**

IaaS gives permission to customer to make use of virtual machines in running, suspended and off state modes; hence the Virtual Machine may be out-of-date with respect to security updates.

**3. Robustness of VM-level isolation**

IaaS provides an environment in isolated form to the customers through the software hypervisor.

The Hypervisor is an application layer which has hardware support for virtualization to divide a physical computer system into several virtual machines.

**4. Data erase practices**

The virtual machines use the common disk resources offered by the cloud provider. After releasing the resource by customer, the data may remain on the virtual machine.

**Q. 30 Compare SaaS, PaaS and IaaS.**

**Ans. : Comparison between SaaS, PaaS and IaaS.**

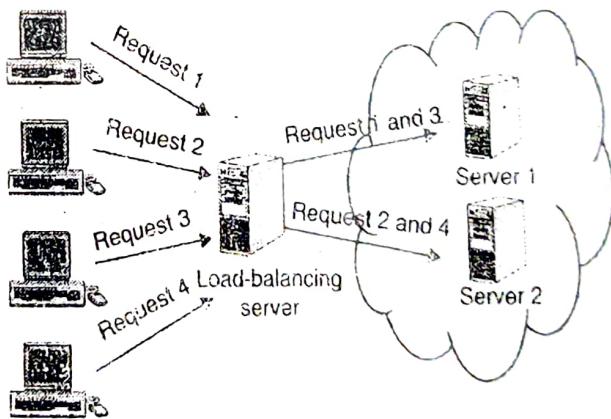
Parameter	SaaS	PaaS	IaaS
Level of control for consumer	As per usage and corresponding configuration	limited level of administrative	full level of administrative

Parameter	SaaS	PaaS	IaaS
Functionality for consumer	access to front-end user interface	moderate level of administrative control on available IT resources	full access to the IT resources
Common Cloud Consumer Activities	uses and configures the cloud service	development, deployment, testing and management of cloud services	sets up and configures bare infrastructure and do the installation, management, and monitoring of any needed software
Common Cloud Provider Activities	Does the implementation, management, and maintenance of cloud service monitors usage by the cloud consumers	pre-configures the platform and make provisions of underlying infrastructure, middleware, and other necessary IT resources, monitors usage by cloud consumers	Do the provisions and management of the networking, storage, and hosting, monitors usage by the cloud consumers

**Q. 31 How to improve performance through load balancing?**

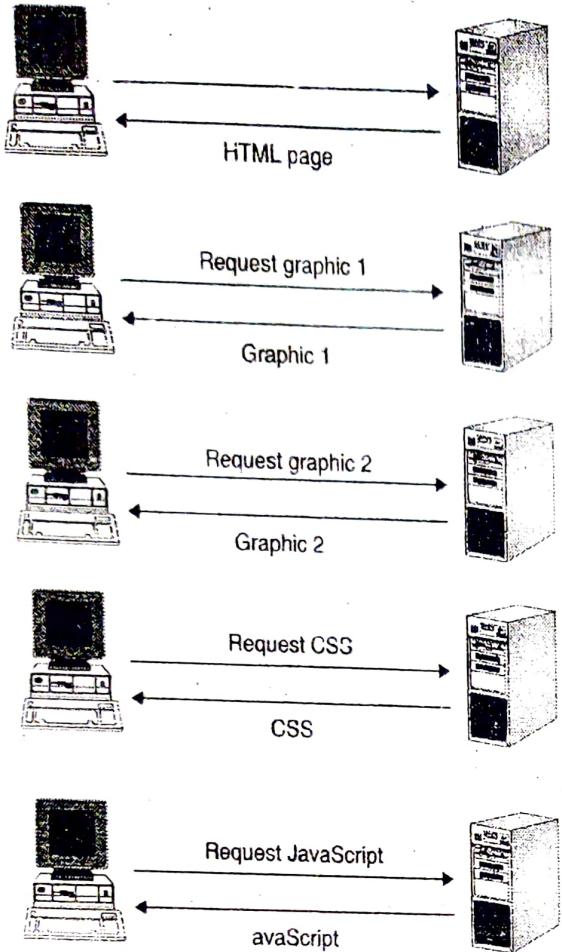
**Ans. :**

In the world of web, the several sites experience a huge network traffic.



**Fig. 1.16 : Load balancing takes help of a server to route traffic to other servers which share the workload.**

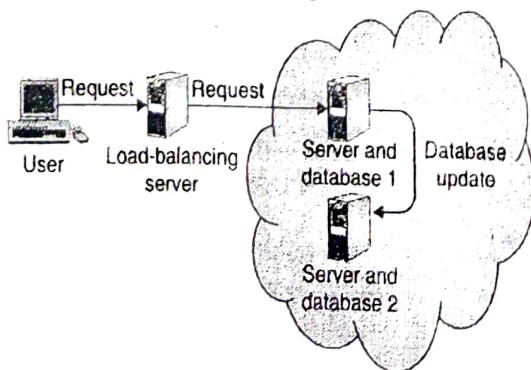
There are number of famous sites such as Google, Yahoo!, Amazon which experience millions of user hits per day. A technique known as **load balancing** is used by those sites to manage the huge web requests. It helps to share the requests across multiple servers. Simply the browser working at client side send request of an HTML page first and then the associated **graphics, CSS, and JavaScript files** from the web server, as shown in Fig. 1.17.



**Fig. 1.17 : A client makes several requests through browser to a server so as to download the HTML, CSS, JavaScript, and page graphics.**

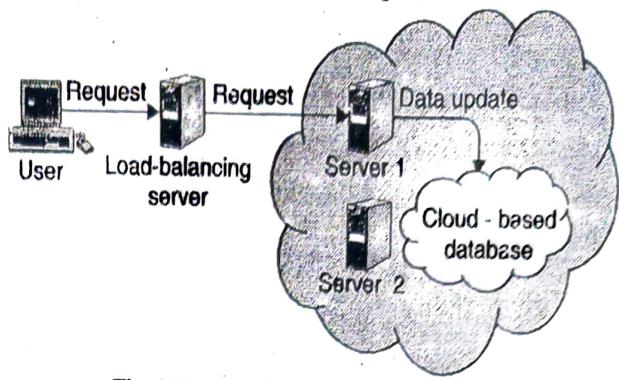
When the demand regarding the server increases in huge manner, the organization has choice of placing a load-balancing server with two or more servers to which the web requests are distributed by the load balancer. For simple applications the load balancing is considered as quite straightforward since either server may able to handle all requests.

When there is complexity in server-based applications regarding various aspects such as accessing data within a database, then the developers has to provide shared access to the database. To avoid any single point of failure, companies make replicas of the database on more than one server. The database software further should do the process of synchronization of data updates across the systems, as illustrated in Fig. 1.18.



**Fig. 1.18 : Load-balanced systems for redundancy of data, often make several replicas of databases on different servers**

All the databases, in turn, will send data updates to other for maintaining data synchronization among the servers. As another option, the organization may select to simplify the solution with the help of cloud-based database or a cloud-based network-attached storage (NAS) device, as illustrated in Fig. 1.19.



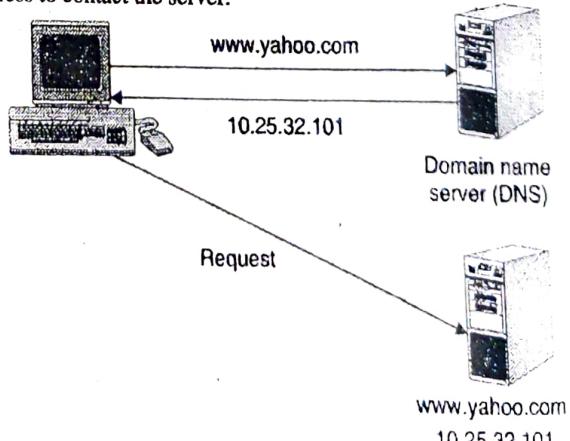
**Fig. 1.19 : Cloud-based NAS device and a cloud-based database to manage the database replication as well as load balancing**

In this manner, there is no need to the applications to worry about the data synchronization and replication since this work is handled within the cloud.

#### Taking a Closer Look at Load Balancing

Now to exactly understand the concept of load balancing, we will consider the processing which occurs when a client/user visits a website like Yahoo! Initially user enters the domain name, such as [www.yahoo.com](http://www.yahoo.com) in the address bar of browser.

Further the web browser of user sends the domain name to a specialized server on the web known as a DNS (Domain Name Server), which, in turn, returns the site's IP (Internet Protocol) address. As illustrated in Fig. 1.20, the browser in then uses the IP address to contact the server.



**Fig. 1.20 : On the web, to access server, the web browsers use the IP address they receive from a DNS**

When the mechanism of load balancing is used by website, the IP address which is returned by the Domain Name Server might correspond to the load-balancing server. When a browser request is

received by the load balancer, it just forwards the request to one of the servers in the round-robin mechanism.

If the requests on the website increases, more servers can be introduced, to which the load balancer can distribute requests. With the help of same technique, most of the IaaS solutions offer on-demand scaling and load balancing.

### **Q. 32 What are the advantages of IaaS solutions ?**

**Ans. :**

IaaS solution has following advantages :

- (i) No need of an expensive as well as staff-intensive data center
- (ii) Hardware scalability is easy and flexible
- (iii) Reduced hardware cost
- (iv) On-demand, pay-as-you-go scalability
- (v) Need of less IT staff
- (vi) Ad hoc test environments suitability
- (vii) Absolute system administration and management

### **Q. 33 Write note on Identity management - as - a - service (IDaaS).**

**Dec. 16**

**Ans. :**

Nowadays in most of the organizations, to perform number of tasks users have to log in to several different systems. In those organizations some systems may be based on cloud, some may be local server based, and some may be accessible using different devices. In handling multiple servers, the challenge is that users have to remember as well as manage more than one username and password combinations.

Further, whenever any employee leaves the company, it is responsibility of the IT staff to coordinate with the HR department to make sure that all the accounts of the respective user have been disabled. ID management (User identity management) is considered as tedious, time consuming, and expensive. Since last decade, organizations have started to emerge to provide identity (or identification) as a service (IDaaS), or cloud-based ID management.

## **Chapter 2 : Data Storage and Security in Cloud**

### **Q.1 Write a note on cloud File System with architectures.**

**Ans. :**

In computing, a file system controls how data is stored and retrieved. Without a file system, information placed in a storage medium would be one large body of data with no way to tell where one piece of information stops and the next begins. By separating the data into pieces and giving each piece a name, the information is easily isolated and identified. The structure and logic rules used to manage the groups of information and their names are called as "file system".

Cloud supports distributed file system which is a file system that allows many clients to have access to data and supports operations (create, delete, modify, read, write) on that data. Each data file may be partitioned into several parts called chunks. Each chunk may be stored on different remote machines facilitating the parallel execution of applications. Typically, data is stored in files in a hierarchical tree, where the nodes represent directories. There are several ways to share files in a distributed architecture : each solution must be suitable for a certain type of application, depending on how complex the application is.

Meanwhile, the security of the system must be ensured. Confidentiality, availability and integrity are the main keys for a secure system. Users can share computing resources through the Internet, thanks to cloud computing which is typically characterized by scalable and elastic resources such as physical servers, applications and any services that are virtualized and

allocated dynamically. Synchronization is required to make sure that all devices are up-to-date. Distributed file systems enable many big, medium, and small enterprises to store and access their remote data as they do with local data, facilitating the use of variable resources.

### **Q. 2 Explain architecture of GFS.**

**Ans. :**

GFS is enhanced for Google's core data storage and usage needs (primarily the search engine), which can generate enormous amounts of data that must be retained. Google File System grew out of an earlier Google effort, "BigFiles", developed by Larry Page and Sergey Brin in the early days of Google, while it was still located in Stanford. Files are divided into fixed-size chunks of 64 megabytes, similar to clusters or sectors in regular file systems, which are only extremely rarely overwritten, or shrunk; files are usually appended to or read.

It is also designed and optimized to run on Google's computing clusters, dense nodes which consist of cheap "commodity" computers, which means precautions must be taken against the high failure rate of individual nodes and the subsequent data loss. Other design decisions select for high data throughputs, even when it comes at the cost of latency. A GFS cluster consists of multiple nodes. These nodes are divided into two types : one Master node and a large number of Chunkservers.

Each file is divided into fixed-size chunks. Chunk servers store these chunks. Each chunk is assigned a unique 64-bit label by

the master node at the time of creation, and logical mappings of files to constituent chunks are maintained.

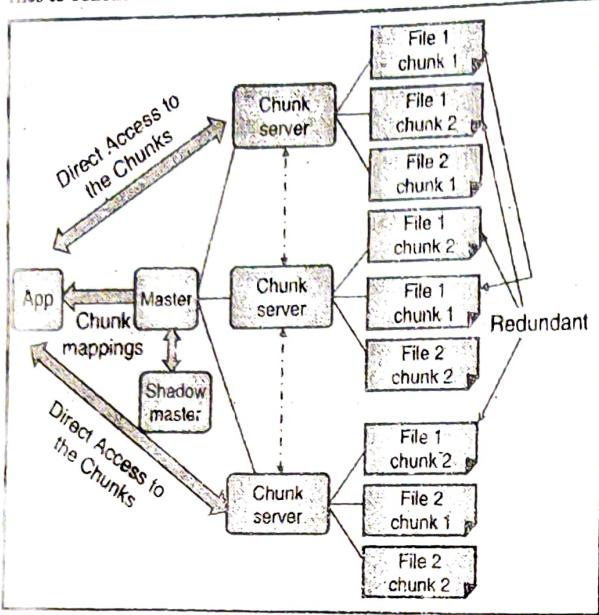


Fig. 2.1 : Google File System

Each chunk is replicated several times throughout the network. At default, it is replicated three times, but this is configurable. Files which are in high demand may have a higher replication factor, while files for which the application client uses strict storage optimizations may be replicated less than three times - in order to cope with quick garbage cleaning policies.

The Master server does not usually store the actual chunks, but rather all the metadata associated with the chunks, such as the tables mapping the 64-bit labels to chunk locations and the files they make up, the locations of the copies of the chunks, what processes are reading or writing to a particular chunk, or taking a "snapshot" of the chunk pursuant to replicate it (usually at the instigation of the Master server, when, due to node failures, the number of copies of a chunk has fallen beneath the set number).

All this metadata is kept current by the Master server periodically receiving updates from each chunk server ("Heart-beat messages"). Permissions for modifications are handled by a system of time-limited, expiring "leases", where the Master server grants permission to a process for a finite period of time during which no other process will be granted permission by the Master server to modify the chunk. The modifying chunkserver, which is always the primary chunk holder, then propagates the changes to the chunkservers with the backup copies.

The changes are not saved until all chunkservers acknowledge, thus guaranteeing the completion and atomicity of the operation. Programs access the chunks by first querying the Master server for the locations of the desired chunks; if the chunks are not being operated on (i.e. no outstanding leases exist), the Master replies with the locations, and the program then contacts and receives the data from the chunkserver directly. Unlike most other file systems, GFS is not implemented in the kernel of an operating system, but is instead provided as a userspace library.

## Performance

Deciding from benchmarking results, when used with relatively small number of servers (15), the file system achieves reading performance comparable to that of a single disk (80-100 MB/s), but has a reduced write performance (30 MB/s), and is relatively slow (5 MB/s) in appending data to existing files.

The authors present no results on random seek time. As the master node is not directly involved in data reading (the data are passed from the chunk server directly to the reading client), the read rate increases significantly with the number of chunk servers, achieving 583 MB/s for 342 nodes. Aggregating a large number of servers also allows big capacity, while it is somewhat reduced by storing data in three independent locations (to provide redundancy).

### Q. 3 Explain various components of Hadoop.

**Ans. :**

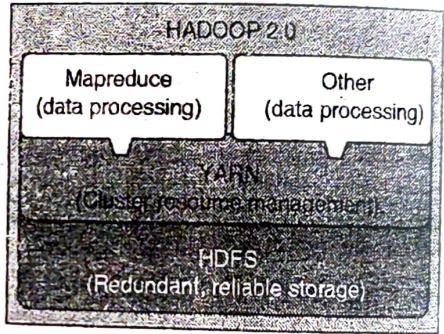


Fig. 2.2

- HDFS :** HDFS stands for Hadoop Distributed File System. It states that the files will be broken into blocks and stored in nodes over the distributed architecture. It provides high-throughput access to application data.
- Yarn :** Yarn stands for "Yet another Resource Negotiator". It is used for job scheduling and managing the cluster (multiple nodes).
- Map Reduce :** This is YARN-based system for parallel processing of large data set using key value pair. The Map task takes input data and converts it into a data set which can be computed in Key value pair.
- Hadoop Common :** These Java libraries and utilities are used to start Hadoop. These are used by other Hadoop modules. These libraries provide file system and OS level abstractions.

### Q. 4 Write a note on MapReduce.

**Ans. :**

MapReduce is an important part of Hadoop. It is a software framework which is used to write applications easily to process huge amount of data (multi-terabyte data-sets) simultaneously on large clusters (thousands of nodes) in reliable, fault-tolerant manner. MapReduce basically refers to two tasks performed by the Hadoop programs. One is map and another is reduce.

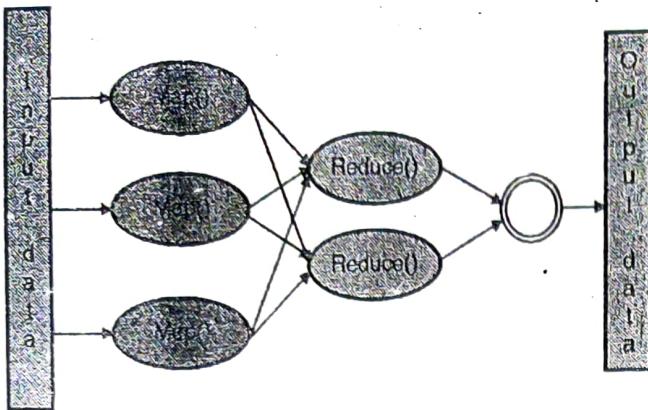


Fig. 2.3

Hadoop programs perform following two tasks on **MapReduce**

- The Map Task :** This is the first task, which takes a set of data and converts it into another set of data in which individual elements are broken down into tuples (key/value pairs).
- The reduce :** This job takes the output of previously executed map task as input and combines those data tuples into a smaller set of tuples. As the sequence of the name MapReduce implies, the reduce job is always performed after the map job.

In MapReduce framework, the data of input and output is stored in a file system. The framework handles the scheduling of all the tasks, monitoring these tasks and if fails, re-executes them. The major advantage of MapReduce is that it is easy to scale data processing over multiple computing nodes. Under the MapReduce model, the data processing primitives are called mappers and reducers. The MapReduce framework consists of single master JobTracker and one slave TaskTracker per cluster-node.

**Master JobTracker :** The tasks under master are :

- Managing the resources.
- Tracking consumption and availability of resources.
- Scheduling the jobs component tasks on the slaves.
- Monitoring the tasks and re-executing the failed tasks.

**The slaves TaskTracker :** It execute the tasks as per the directions of the master and provide task-status information to the master periodically.

The JobTracker is very important in Hadoop MapReduce service. If JobTracker goes down, all running tasks get halted.

#### Q. 5 What is HDFS? Explain in detail.

Dec. 16

**Ans. :**

The HDFS is the primary storage system used by Hadoop applications. HDFS is a distributed file system and a framework provided by Hadoop for the analysis and transformation of huge data sets which uses the MapReduce paradigm. The HDFS is based on Google File System (GFS). It provides high-performance access

to data across Hadoop clusters (thousands of computers), HDFS has become a key tool for managing pools of big data and supporting big data analytics applications.

HDFS is usually deployed on commodity hardware of low-cost where the possibility of server failures is common. The file system is designed to be highly fault-tolerant. The HDFS facilitates the rapid transfer of data between different computer nodes and enables Hadoop systems to preceeed its execution even if one or more nodes get failed. That decreases the risk of catastrophic failure, even in the event that numerous nodes fail.

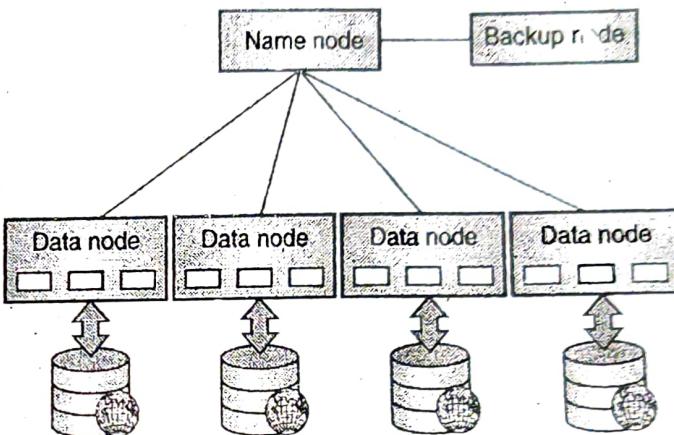


Fig. 2.4

The architecture used by HDFS is known as master/slave architecture. NameNode which manages the metadata of file system and DataNode which stores the actual data.

The HDFS namespace is a hierarchy of files and directories. Inodes are used to represent these file and directories. Inodes are used to record attributes such as permissions, modification and access times etc. The file content is split into large blocks and each block of the file is independently replicated at multiple DataNodes.

The tree structure of namespace is maintained by the NameNode. It maps the blocks to DataNodes. In a cluster there may be hundreds of DataNodes and thousands of HDFS clients per cluster, as number of application tasks can be executed by each DataNode simultaneously.

#### Advantages of HDFS

- High scalability.
- Low limitation.
- Open source.
- Low cost.

#### Disadvantages of HDFS

- Still rough – means software under active development.
- Programming model is very restrictive.
- Cluster management is high.

**Q. 6 Write a short note on HBase.**

May 16

**Ans. :**

HBase is the database that stores huge amounts of data and access the data in a random manner. HBase built on top of the Hadoop file system. It is a distributed column-oriented database. It is an open-source project and is horizontally scalable. HBase leverages the fault tolerance provided by the Hadoop File System (HDFS). HBase is a part of the Hadoop ecosystem that provides random real-time read/write access to data in the Hadoop File System. HBase is used to store data in HDFS which can be easily read and write randomly.

**Features of HBase**

1. Strictly consistent reads and writes.
2. Automatic and configurable sharding of tables
3. Linear and modular scalability.
4. Block cache and Bloom Filters for real-time queries.
5. Automatic failover support between RegionServers.
6. Easy to use Java API for client access.
7. Query predicate push down via server side Filters
8. Convenient base classes for backing Hadoop MapReduce jobs with Apache HBase tables.

**Q. 7 Discuss Hbase data model.****Ans. :**

HBase stores data in table format. Tables are combinations of rows and columns.

**HBase Data Model Terminology**

1. **Table** : An HBase table consists of multiple rows.
2. **Row** : In HBase the row contains a row key and one or more columns with values associated with them. Records are stored in sorted format depending upon the row key. The main focus while storing data is that the related rows should be stored near to each other.
3. **Column** : In HBase column contains a column family and a column qualifier. These are delimited by a colon(:) character.

**Column Family**

Column families contain set of columns and their values. The column family has set of storage properties, such as whether its values should be cached in memory, how its data is compressed or its row keys are encoded, and others. The column families are same in all the tables even though do not have any value. At the time of table creation, the column families are specified.

- (i) **Column Qualifier** : To provide the index for a given piece of data, a column qualifier is added to a column family. Column qualifiers are mutable and different column qualifiers may be assigned to different rows.

(ii) **Cell** : Cell contains a value and a timestamp, which represents the value's version. A cell is a combination of row, column family, and column qualifier.

(iii) **Timestamp** : A timestamp is the identifier for a given version of a value. It is written alongside each value. On the RegionServer, the timestamp is by default the time when the data was written. It is also possible to specify different timestamp value while adding data into the cell.

**Q. 8 Write a note on Bigtable.****Ans. :**

Cloud Bigtable is a sparsely populated table that can scale to billions of rows and thousands of columns, enabling user to store terabytes or even petabytes of data. A single value in each row is indexed; this value is known as the row key. Cloud Bigtable is ideal for storing very large amounts of single-keyed data with very low latency. It supports high read and write throughput at low latency, and it is an ideal data source for MapReduce operations.

Cloud Bigtable is exposed to applications through multiple client libraries, including a supported extension to the Apache HBase library for Java. As a result, it integrates with the existing Apache ecosystem of open-source Big Data software. Cloud Bigtable's powerful back-end servers offer several key advantages over a self-managed HBase installation :

1. **Incredible scalability** : Cloud Bigtable scales in direct proportion to the number of machines in the cluster. A self-managed HBase installation has a design bottleneck that limits the performance after a certain QPS (Queries per second) is reached. Cloud Bigtable does not have this bottleneck, and so user can scale the cluster up to handle more queries.
2. **Simple administration** : Cloud Bigtable handles upgrades and restarts transparently, and it automatically maintains high data durability. To replicate the data, simply add a second cluster to the instance, and replication starts automatically. No more managing masters or regions; user has to just design table schemas, and Cloud Bigtable will handle the rest for user.
3. **Cluster resizing without downtime** : User can increase the size of a Cloud Bigtable cluster for few hours to handle a large load, then reduce the cluster's size again, all without any downtime. After user changes a cluster's size, it typically takes just a few minutes under load for Cloud Bigtable to balance performance across all of the nodes in the cluster.

Cloud Bigtable is ideal for applications that need very high throughput and scalability for non-structured key/value data, where each value is typically no larger than 10 MB. Cloud Bigtable also excels as a storage engine for batch MapReduce operations, stream processing/analytics, and machine-learning applications.

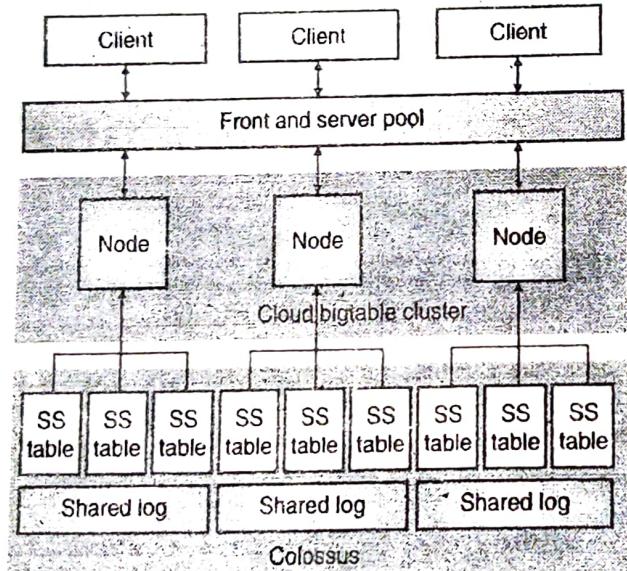
User can use Cloud Bigtable to store and query all of the following types of data :

1. **Time-series data**, such as CPU and memory usage over time for multiple servers.
2. **Marketing data**, such as purchase histories and customer preferences.
3. **Financial data**, such as transaction histories, stock prices, and currency exchange rates.
4. **Internet of Things data**, such as usage reports from energy meters and home appliances.
6. **Graph data**, such as information about how users are connected to one another.

#### **Q. 9 Explain Bigtable architecture with suitable diagram.**

**Ans. :**

The following diagram shows a simplified version of Cloud Bigtable's overall architecture :



**Fig. 2.5 : Cloud Bigtable Architecture**

As the Fig.2.5 illustrates, all client requests go through a front-end server before they are sent to a Cloud Bigtable node. The nodes are organized into a Cloud Bigtable cluster, which belongs to a Cloud Bigtable instance, a container for the cluster. Each node in the cluster handles a subset of the requests to the cluster.

By adding nodes to a cluster, user can increase the number of simultaneous requests which the cluster can handle, as well as the maximum throughput for the entire cluster. If user enables replication by adding a second cluster, he/she can also send different types of traffic to different clusters, and can fail over to one cluster if the other cluster becomes unavailable.

A Cloud Bigtable table is sharded into blocks of contiguous rows, called *tablets*, to help balance the workload of queries. (Tablets are similar to HBase regions). Tablets are stored on Colossus, Google's file system, in SSTable format. An SSTable

provides a persistent, ordered immutable map from keys to values, where both keys and values are arbitrary byte strings.

Each tablet is associated with a specific Cloud Bigtable node. In addition to the SSTable files, all writes are stored in Colossus's shared log as soon as they are acknowledged by Cloud Bigtable, providing increased durability. Importantly, data is never stored in Cloud Bigtable nodes themselves; each node has pointers to a set of tablets that are stored on Colossus. As a result :

Rebalancing tablets from one node to another is very fast, because the actual data is not copied. Cloud Bigtable simply updates the pointers for each node. Recovery from the failure of a Cloud Bigtable node is very fast, because only metadata needs to be migrated to the replacement node. When a Cloud Bigtable node fails, no data is lost.

#### **Q. 10 Write a note on Load Balancing in Bigtable.**

**Ans. :**

Each Cloud Bigtable zone is managed by a master process, which balances workload and data volume within clusters. The master splits busier/larger tablets in half and merges less-accessed/smaller tablets together, redistributing them between nodes as needed. If a certain tablet gets a spike of traffic, the master will split the tablet in two, then move one of the new tablets to another node. Cloud Bigtable manages all of the splitting, merging, and rebalancing automatically, saving users the effort of manually administering their tablets.

To get the best write performance from Cloud Bigtable, it's important to distribute writes as evenly as possible across nodes. One way to achieve this goal is by using row keys that do not follow a predictable order. For example, usernames tend to be distributed more or less evenly throughout the alphabet, so including a username at the start of the row key will tend to distribute writes evenly.

At the same time, it's useful to group related rows so they are adjacent to one another, which makes it much more efficient to read several rows at the same time. For example, if user is storing different types of weather data over time, his/her row key might be the location where the data was collected, followed by a timestamp (for example, WashingtonDC#201803061617).

This type of row key would group all of the data from one location into a contiguous range of rows. For other locations, the row would start with a different identifier; with many locations collecting data at the same rate, writes would still be spread evenly across tablets.

#### **Q. 11 Explain the term "Dynamo Cloud".**

**Ans. :**

Dynamo is the name given to a set of techniques that when taken together can form a highly available key-value structured storage system or a distributed data store. It has properties of both databases and distributed hash tables (DHTs). Amazon DynamoDB

is "built on the principles of Dynamo" and is a hosted service within the AWS infrastructure. It was created to help address some scalability issues that Amazon.com website experienced. By 2007, it was used in Amazon Web Services, such as its Simple Storage Service.

### Principles

- Incremental scalability** : Dynamo should be able to scale out one storage host (or "node") at a time, with minimal impact on both operators of the system and the system itself.
- Symmetry** : Every node in Dynamo should have the same set of responsibilities as its peers; there should be no distinguished node or nodes that take special roles or extra set of responsibilities.
- Decentralization** : An extension of symmetry, the design should favour decentralized peer-to-peer techniques over centralized control.
- Heterogeneity** : The system needs to be able to exploit heterogeneity in the infrastructure it runs on. e.g. the work distribution must be proportional to the capabilities of the individual servers. This is essential in adding new nodes with higher capacity without having to upgrade all hosts at once.

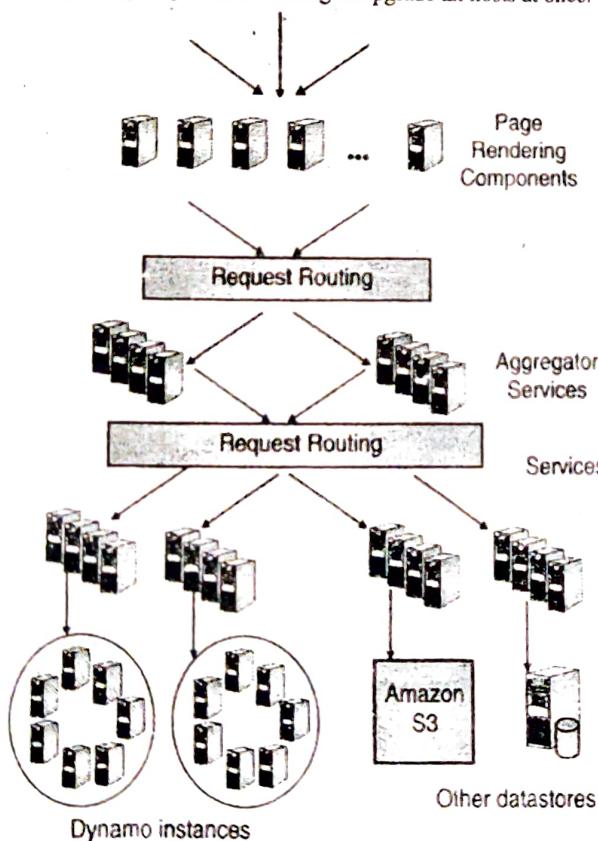


Fig. 2.6 : Amazon Dynamo

### Q. 12 Explain Cloud storage.

**Ans. :**

There are number of benefits of Cloud storage over traditional data storage. The data kept on cloud is accessible from any location having Internet access. There is no need of using the same computer to access data and no need to carry physical storage devices. Also, if any company has branch offices, cloud makes it easy to access data from any branch. There are several cloud

storage systems available in market in which few are very particular in what they do. Some focus on specific market and store only email or digital pictures, while some can store any type of data. Some providers are comparatively very small, while some are huge and are able to fill an entire warehouse.

In the cloud storage system, at the most basic level, there is need of only single data server which is connected to the Internet. Files are stored by the subscriber on the server. A subscriber copies files to the server over the Internet, in which records are stored. Whenever client needs the data, he or she has access to the respective data server with a web-based interface. The client can get the data from the server directly or get permission from server to access as well as manipulate the data.

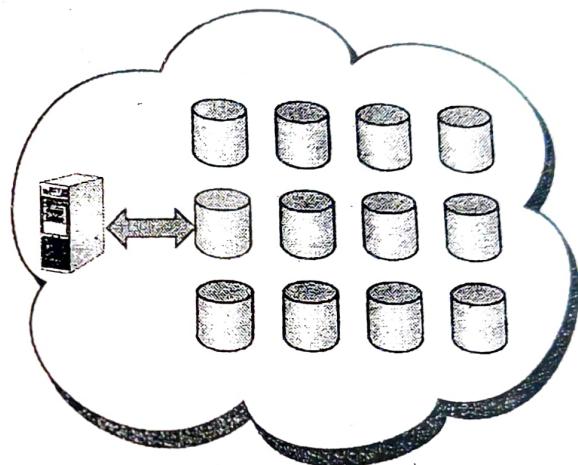


Fig. 2.7 : More commodity hard drives can be added by the provider to enhance capacity

More specifically, on the other hand, several data servers are used by the cloud storage systems. As there is need of maintenance or repair to the servers, the data must be stored on more than one machines enabling redundancy. In the absence of redundancy, it is difficult for cloud storage systems to guarantee clients that their data will be accessible at any given time.

Number of systems uses different power supplies so as to store the same data on servers which helps the clients to access their data in case of failure of power supply also. Number of clients prefer cloud storage not as they've run out of room locally, but for the purpose of safety. If there is any bad incidence with their building occurs, then they haven't lost all their data.

### Q. 13 Explain security in cloud storage.

**Ans. :**

For security purpose most systems prefer combination of techniques :

- Encryption** : Information is encoded with the help of a complex algorithm. Encryption key is used to decode the encrypted files. Despite the fact that it is probable to crack encrypted data, it is very difficult for most of the hackers since they do not have access to the amount of computer processing power which is required to crack the code.

- 2. Authentication processes :** It needs a user to generate a name and password.
- 3. Authorization practices :** Lists are created by the client regarding the people who should be able to access data stored on the cloud system. Number of times there is multiple levels of authorization.

For example, there may be restricted access to a front-line employee for the data stored on the cloud and the access right of head of the IT department will be full and have free access to everything.

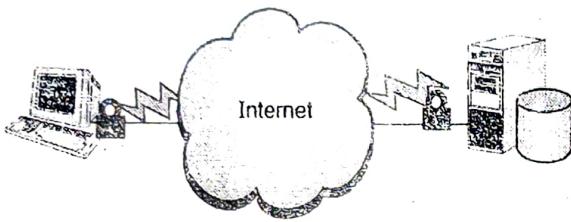


Fig. 2.8

**Encryption and authentication :** security measures to keep data safe on a cloud storage. But in spite of all these measures, there are still concerns that the information present on a remote system is vulnerable. One threat is always there that a hacker may find a way to enter into the secure system to access the data. One another aspect is that the employees who are disgruntled may provide harm to the data with their access credentials.

#### Q. 14 Explain reliability in cloud storage.

**Ans. :**

Reliability is another important aspect. Unreliable storage system becomes a liability. Everyone will be unhappy to save data on an unstable system, as well users also keen to trust an organization which is financially unstable. Redundancy is used by most of the cloud storage providers to take care of reliability, but there is always possibility that the whole system may get crashed and leave clients without access to their own data.

Reputation is considered as the most important factor to cloud storage providers. If in case there is a perception that the provider is unreliable, getting clients will be hard. It is difficult for unreliable cloud storage providers to stay in the field, as there are large number of players in the market.

#### Q. 15 Explain services offered by Amazon S3.

**Ans. :**

Amazon S3 stands for Amazon Simple Storage Service. It is a Infrastructure as a Service (IaaS) solution which is developed by AWS. Amazon S3 provides highly-scalable, secured and low-latency data storage from the cloud. The most important functionality of Amazon S3 interface is of data storage and its retrieval from anywhere and at any time. The Amazon S3 has flexible back-end infrastructure which avoids the requirement of in house storage and also provides unlimited or add-as-you-grow form of storage which is billed depending upon the usage.

#### Amazon S3 features and benefits include

Huge capacity for data and object storage is provided by Amazon S3 for maximum of data types in different formats. The size of data may vary in the range of 1B - 5 TB. To reduce latency Amazon S3 provides Reduced Redundancy Storage (RRS). The data is stored in regionally segregated buckets. Hence the number of resources are saved and increases the application efficiency for the subscribers which may be at different locations. It provides strong security of regionally stored data through authentication.

#### S3 API and competing services

Competing services have got an effective rise because of the Amazon S3 and tooling based on the S3 API. The standard programming interface is used by these services; however underlying technologies and supporting business models helps to differentiate them.

Competing service providers are enabled by the cloud storage standard (like electrical and networking standards) to design their services and provide following benefits :

- (i) A set of rules are provided to increase competition. It encourages market entry of smaller companies.
- (ii) The innovations of developers, cloud storage & tool vendors, are encouraged which helps them to concentrate on improvement of quality of product and services rather than focusing on compatibility.
- (iii) The economies of scale in implementation are allowed. (i.e., if a service provider finds an outage or as tools of client outgrow and require faster OS or tools, the swapping the solutions can be done).
- (iv) In response to demands of the market place, timely solutions for delivering functionality are provided.

#### Examples of competing S3 compliant storage implementations include

- (i) Pure Storage's FlashBlade
- (ii) Minio Object Storage
- (iii) NooBaa Hybrid Storage
- (iv) Scality RING

#### Q. 16 What are the general security disadvantages of cloud based solutions?

**Ans. :**

#### Disadvantages of cloud based solution

1. Country or jurisdiction
  - The location of cloud-based resources is sometimes considered as confusing aspect. If location of hosting of cloud-based resources is a remote country, then one must be concerned with the laws as well as the government stability of particular country. Also if the cloud resources reside in more than one state, there may be questions of jurisdiction regarding the legal matter.

## 2. Multitenant risks

Multitenant solutions are used by number of cloud-based service providers that means same resources like database can be used by multiple customers. Because of this, one company's data might expose to another company due to any application error. Also if there is sharing of data storage device, there is possibility that data of one company may expose to another company.

## 3. Malicious insiders

In spite of best human-relations effort of a cloud solution provider, some problems may arise because of malicious employees. Based on the role of such malicious employee, a company's cloud-based data may be at risk.

## 4. Vendor lock in

Based on the way of storing data of an organization by the cloud-based solution provider, it may be complicated for the organization to replace provider later in the situation of violation of a service-level agreement or any other problem.

## 5. Risk of the cloud-based provider failing

Organizations which depend on cloud-based providers always are at risk that the cloud service provider could fail. Some organizations ask for a source code escrow agreement, because of which a copy of the provider's source code is present with a third-party company. If there is incident that the provider fail, the organization can gain access to the respective source code, using which they can re-host the solution.

## Q. 17 Write a note on business continuity and disaster recovery.

Ans. :

For years, in an organization a primary job of the IT staff was to take care of availability of several computing resources such as applications, files, phone systems, etc. This responsibility is not only on a day-to-day basis, but also in the unexpected situations of a crisis or any natural disaster.

Business continuity is related to the policies, procedures, as well as actions which are taken by a company on primary basis to ensure the availability of several crucial business functions to employees, clients, and any other important stakeholders. Disaster recovery is nothing but the respective steps an organization will take to restore operations as well as data in the situation of a disaster such as fire, flood, or storm like.

## Q. 18 What are data security risks? How will you mitigate these risks?

Dec. 16

Ans. :

IT staff of an organization must be able to anticipate and prepare for a various types of system threats.

### Types of system threats

#### Threat : Disk Failure

Disk drives are the mechanical devices which in due course wear out and fail. Also some other types of threats like fire, flood, theft, or power surges are reasons for loss of disk-based data.

#### Traditional Risk Mitigation for Disk Failure

- (i) The initial and basic risk mitigation for the situation of disk failure is to ready with up-to-date disk backups. If at any occasion a disk fails, the organization has option to simply replace the disk and restore the backup.
- (ii) That implies that the reason of the disk failure (fire, smoke, flood, or theft) did not also make damage to the disk backup. To avoid this problem, most of the organizations keep their disk backups at an off-site storage facility.

#### RAID-Protected Storage

Nowadays most of the data centers use a redundant array of independent (or inexpensive) disks (RAID) for the purpose of minimizing the impact of disk failure.

## Chapter 3 : Virtualization

## Q. 1 Explain characteristics of virtualized environments in detail.

Dec. 16, April 17

Ans. :

In computing, **virtualization** refers to the act of creating a virtual (rather than actual) version of something, including virtual computer hardware platforms, storage devices, and computer network resources. Virtualization which is considered as a state-of-art technology is now a great advantage for an IT industry which allows performing more with less.

The most important goal of virtualization is to centralize the important as well as administrative tasks with parallel management

of scalability and work load. There are several reasons which show the need of virtualization.

### 1. Resource optimization

- (i) Now a day, the computer resources are very powerful with excess capacity. The hardware and allocating parts of it can be virtualized depending upon the actual requirements of users and applications. This leads to much more effective use of available computing power, storage space as well as the network bandwidth.
- (ii) In this case the computers no longer required to be idle or performing lower than their capabilities.

(iii) Isolated, constrained and test environments are made available to software developers by the Virtual machines. Instead of buying dedicated physical hardware, virtual machines can be generated on the available hardware.

## 2. Consolidation

- (i) Usually individual computers are dedicated to a single application. If small amount of processing power is used by the several applications, then it is possible to consolidate multiple computers into single server executing several virtual environments.
- (ii) In case of organizations which needs thousands of servers, consolidation can effectively minimize the requirement of floor space, HVAC, A/C power as well as the co-location resources. This leads to significant decrease in cost of ownership, as less physical servers, floor as well as rack space are needed.

## 3. Maximizing Uptime

- (i) Agility is the capacity to respond to frequently changing requirements in a quick and flexible manner.
- (ii) Virtualization provides new openings to data centre administration, which allow users to have.
- (iii) Assured uptime of servers and applications; very fast disaster recovery in case of large scale failures.
- (iv) Template images helps to deploy new virtual machines as well as aggregated pools of virtual machines
- (v) Elasticity ; On demand provision of resources.
- (vi) Reconfiguration of computing environments which are executing without affecting the users.

## 4. Automatically Protect Applications from Server Failure

- (i) Virtualization helps to implement redundancy without any need to buy additional hardware. Redundancy, is the mechanism of running single application on several servers, is a safety measure: if there is failure of server.
- (ii) Another server which is executing the same application takes the control by itself, which avoids the interruption in service. In Virtual machines, such type of redundancy works in two ways :
  - (a) In case of failure of one virtual system, another virtual system takes over.
  - (b) Against physical hardware failure, better protection can be provided by running the redundant virtual machines on different physical hardware.

## 5. Easily Migrate Workloads as Needs Change

- (i) Migration means shifting the server environment to another location. In virtualization, it is easy to move a virtual machine from one physical machine to another physical machine.

- (ii) In case of physical servers, this migration is possible only if both physical machines are running on the common hardware, operating system and processor.
- (iii) In the virtual world, the migration of server is possible in between physical hosts which may have completely different hardware configurations.
- (iv) The purpose of migration is to improve reliability as well as availability: in case of failure in the hardware, it is possible to move the guest system c to a healthy server with limited downtime, if any.
- (v) It is also helpful in the situation when a virtual machine is required to scale beyond the physical capabilities of the existing host.

## 6. Save resources and money

- (i) Profit is the most important aspect of any organization. Saving expenses can also be termed as profit. Virtualization minimizes or may completely omit the need of maintaining servers, desktop and data storage machines.
- (ii) Hence all the costing which may be required for the maintenance of these multiple resources is directly cut down.

## 7. Simplified management of data centre

Just by managing few resources, virtualization helps to handle huge amount of data. There is no need to manage multiple hard drives and data cables.

## 8. Increased IT productivity and efficiency

Now without any hassle as well as wariness of IT management for big firms, all the advancements in technical and software areas can be implemented with more efficiency.

## 9. Easy disaster recovery

If any event regarding system crash occurs, then with virtualization it is very simple to detect the loophole because of the simplification of network or data system in just 2-3 layers instead of a complicated web of hard wires and connection.

## 10. Enhanced Security

As the level of complexity is dropped, the security level enhancement is easy to implement more effectively.

## Q. 2 Discuss undesirable effects of virtualization.

**Ans. :**

### 1. It can be expensive

One of the important things in case of virtualization is the cost. Virtualization can be quite a pricey investment upfront. For installing the virtual server one-time heavy investment is required which is more than the cost of a traditional server.

**2. Virtualization might not be compatible with other servers and applications**

This is another drawback of virtualization. It may be possible that not all servers and applications are virtualization friendly. This can be a big headache when an investment has already been made on number of servers or if the software used in the business process is not an upgraded version which allows for virtualization.

**3. It needs training to network administrators**

Another drawback of virtualization is the necessity to spend for training purpose of the network administrator which is an additional expense.

**4. It still has limitations**

It is not possible that all applications or servers can work within an environment of virtualization. Hence there is need of hybrid environment to function properly. Since not all the vendors support virtualization and few may stop supporting it later on, there is always a level of uncertainty to implement this system completely.

**5. It creates a security risk**

In virtualization system, sometimes the maintaining security of data gets difficult.

**6. It creates an availability issue**

It is little bit uncertain that every time there will be availability of resources, hence at crucial point if resources not get available, then it affect work efficiency as well as the credential of company.

**7. It creates a scalability issue.**

Sudden demand or access of large number of virtualized resources by a growing business may affect the availability of resources for other small businesses.

**8. It requires several links in a chain that must work together cohesively**

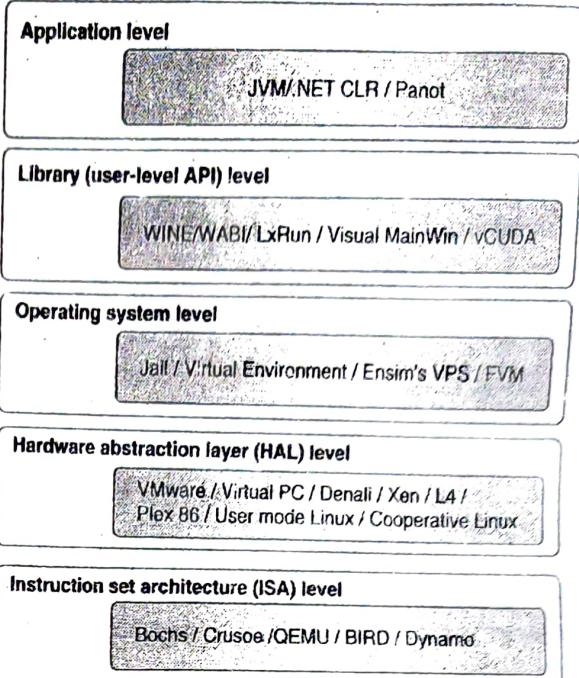
There is need of proper functioning of all the links in a chain of virtualization otherwise simple task like saving file may also become tedious for user.

**Q. 3 Explain different levels of virtualization.**

**Ans. :**

Usually the virtualization layers includes following levels :

- (i) Instruction set architecture (ISA) level
- (ii) Hardware level
- (iii) Operating system level
- (iv) Library support level
- (v) Application level.



**Fig. 3.1 : Implementation Levels of Virtualization**

**1. Instruction Set Architecture Level (ISA)**

In the ISA level, the given ISA is emulated by the ISA of the host machine to carry out the virtualization. For example, we can execute MIPS binary code on an x86-based host machine by the use of ISA emulation. This approach helps to execute plenty of legacy binary code which is written for various processors on any given new hardware host machine. The virtual ISAs are generated on any hardware machine by the Instruction set emulation. Code interpretation is the basic mechanism for emulation. The source code is interpreted one by one by the interpreter program. To perform its function, there may be need of hundreds of native target instructions for a single source instruction. Hence this process is relatively slow.

Dynamic binary translation is required for the better performance. The basic blocks of dynamic source instructions are translated to target instructions. For increasing the efficiency, it is also possible to extend the basic block to program traces or super blocks. There is need of binary translation and optimization for the emulation of instruction set. In this way, virtual instruction set architecture (V-ISA) needs inclusion of a processor-specific software translation layer to the compiler.

**2. Hardware Abstraction Level**

Hardware-level virtualization process is carried out right on top of the bare hardware. With the help of this approach, two things can be achieved; first a virtual hardware environment is generated for a VM and second underlying hardware can be managed through virtualization. The main aim of this approach is to virtualized different resources of a computer like its processors, memory, and I/O devices. It can lead to up gradation of hardware utilization rate by the multiple users simultaneously.

### 3. Operating System Level

This approach basically refers to an abstraction layer among traditional OS and the user applications. Isolated containers are created on a single physical server by the Operating System level virtualization. It also creates OS instances which are used to utilize the hardware and software available in data centres.

The behaviour is just like as they are real servers. The Operating System level virtualization is generally used in the creation of virtual hosting environments for the purpose of allocating hardware resources among a large number of mutually distrusting users. Rarely it can also be used for consolidation of server hardware by shifting services on separate hosts into various types of containers or VMs on one server.

### 4. Library Support Level

Mostly applications use APIs which are exported by user-level libraries instead of lengthy system calls by the OS. As there are well-documented APIs in most of the systems, they can be used as an interface for virtualization. Virtualization using library interfaces is achievable by controlling the communication link between applications and the rest of a system. This is done with the help of API hooks. This approach is implemented by the software tool WINE for the support of Windows applications on top of UNIX hosts.

### 5. User-Application Level

In the User-Application Level approach, an application is virtualized as a VM. Usually an application executes as a process on traditional OS. Hence the User-Application Level virtualization is also called process-level virtualization. In User-Application Level, the virtualization layer resides as an application program on top of the OS.

This layer exports an abstraction of a VM which can execute applications written and compiled for specific abstract machine definition. Any program which is written in the High Level Languages and compiled for this VM can be executed on it. Two good examples are; Microsoft .NET CLR and Java Virtual Machine (JVM).

### Q. 4 Explain Xen Architecture.

**Ans. :**

Xen is an open source hypervisor program which is developed by Cambridge University. Xen hypervisor is of type microkernel, which is used to separate the policy from the mechanism. All the mechanisms are implemented by the Xen hypervisor, and policy handling is the responsibility of Domain 0 as shown in Fig. 3.2.

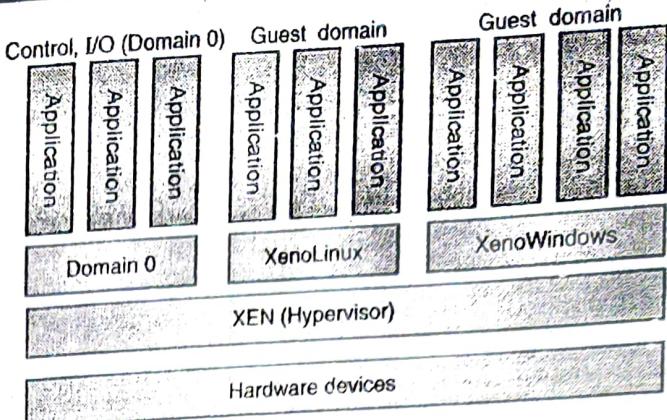


Fig. 3.2 : Xen Architecture

In Xen, no device drivers are included natively. Just a mechanism is provided by Xen with the help of which the guest OS can have direct access to the physical devices. This helps to keep the size of Xen hypervisor effectively small. A virtual environment is provided by Xen between the hardware and the OS.

Now a day there are number of vendors who are in process of developing commercial Xen hypervisors. Some of them are Citrix XenServer and Oracle VM. The most important components of a Xen system are considered as the hypervisor, kernel, and applications. These three components should be organized in proper manner. Like different virtualization systems, number of guest operating systems can be executed on top of the hypervisor. However, not all guest OSes are created with similar functionality, and one in particular situation can control the others.

The guest OS, which can control other OSes is known as Domain while the others are known as Domain U. Domain 0 is considered as a privileged guest OS of Xen. Initially Domain 0 is loaded when Xen boots without the availability of any file system drivers. The basic aim behind design of Domain 0 is to access hardware directly as well as manage devices. Hence allocating and mapping hardware resources for the guest domains (Domain-U) is responsibility of Domain 0.

### Q. 5 What is Binary translation with full virtualization ?

**Ans. :**

Hardware virtualization can be classified into two categories based on the implementation technologies: full virtualization and host-based virtualization. Usually the hypervisor as well as VMM approaches are considered as full virtualization. In full virtualization, there is no need to modify the host OS. It depends upon the binary translation for the process of trapping and to virtualized the execution specific, sensitive, non-virtualizable instructions. There are number of non-critical and critical instructions in guest OSes and their applications. In a host-based system, both of the host OS as well as the guest OS are used. A virtualization software layer is established in between the host OS and guest OS.

## Full Virtualization

In full virtualization, the non-critical instructions execute on the hardware directly whereas for emulation by software, critical instructions are revealed and replaced with traps into the VMM. Only critical instructions are usually trapped into the VMM because there is possibility of large performance overhead because of binary translation.

The hardware is controlled by the critical instructions, as well as these instructions may threaten the security of the system. This does not happen with non-critical instructions. Hence executing the non-critical instructions on hardware enhances the efficiency and ensures system security.

### Binary Translation of Guest OS Requests using a VMM

VMware and many other software companies have implemented the Binary Translation. As shown in Fig. 3.3, VMware has placed the VMM at Ring 0 and the guest OS at Ring 1. The instruction stream is scanned by the VMM and it identifies the instructions which are privileged, control- and behavior-sensitive. After identification of instructions, VMM trap them. The VMM then emulates the behavior of those instructions. This emulation is implemented by the method - binary translation.

Hence binary translation and direct execution are combined by full virtualization. The guest OS is absolutely separated from the underlying hardware. As a result, the guest OS is not aware of that it is being virtualized. The binary translation is a time-consuming process, hence the performance of full virtualization may not be ideal. It is a big challenge to fully virtualized I/O-intensive applications. Binary translation utilizes code cache to save translated hot instructions for the improvement of performance, but the cost of memory usage also get increased.

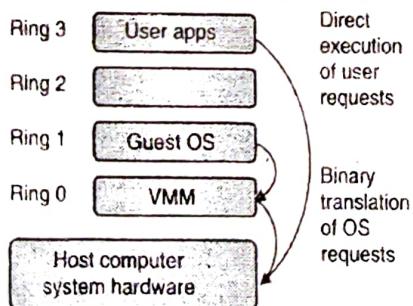


Fig. 3.3 : Binary Translation of Guest OS Requests Using a VMM

### Q. 6 Define hypervisor.

**Ans. :**

A hypervisor or virtual machine monitor (VMM) is computer software, firmware or hardware that creates and runs virtual machines.

### Q. 7 Explain virtualization techniques in cloud computing.

**Ans. :**

For supporting the concept of virtualization, processors like x86 uses a special running mode and instructions, called as hardware-assisted virtualization. The Virtual Machine Manager

(VMM) and guest OS are executed in different modes. The VMM traps all the sensitive instructions of the guest OS and its applications. Hardware completes the mode switching to the save processor states.

### Q. 8 Write a note on Hardware Support for virtualization.

**Ans. :**

Now a day, multiple processes can be executed simultaneously on modern operating systems. If protection mechanism is not provided in a processor, then the hardware will get accessed by the instructions of different processes. This may lead to system crash. Hence all processors provide minimum two modes; user mode and supervisor mode, to make sure control over the access of critical hardware. Instructions which executes in supervisor mode are known as privileged instructions. All other instructions are known as unprivileged instructions.

In a virtualized environment, it is very complicated to execute OSes and applications properly since there are multiple layers in the machine stack. Now a day number of hardware virtualization products are available in the market. The VMware Workstation is a VM software suite for x86 and x86-64 computers. This software suite is used to set up many x86 and x86-64 virtual computers. These VMS can be used concurrently with the host operating system.

### Example : Hardware Support for Virtualization in the Intel x86 Processor

As the software dependent virtualization techniques are considered as difficult and also generates performance overhead, a hardware-assist technique is provided by the Intel to ease the process of virtualization and improve performance. Fig. 3.4 shows the entire overview of Intel's full virtualization techniques. Intel provides VT-x or VT-i technique for processor virtualization. Automatically all the sensitive instructions are trapped by this enhancement. Intel provides EPT for memory virtualization. To improve performance it translates the virtual address to the physical addresses. VT-d and VT-c are offered by Intel for I/O virtualization.

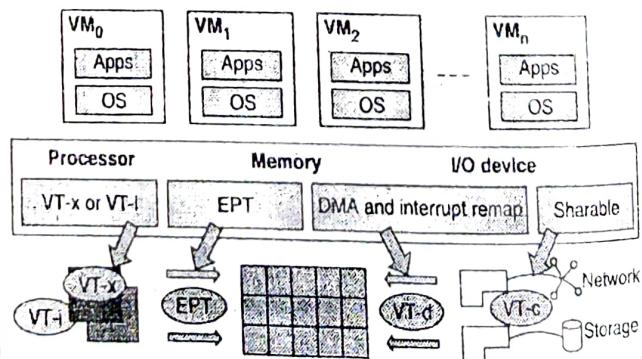


Fig. 3.4 : Hardware Support for Virtualization in the Intel x86 Processor

**Q. 9 Write a note on CPU virtualization.****Ans. :**

VM can be considered as a duplicate of an existing system (computer) in which maximum of the VM instructions are usually executed on the host processor in native mode. Thus the instructions of VM which are unprivileged, execute directly on the host machine to provide higher efficiency.

Care should be taken while handling other critical instructions to maintain correctness and stability. These critical instructions are divided into three categories:

- (i) Privileged instructions
- (ii) Control-sensitive instructions
- (iii) Behavior-sensitive instructions.

Privileged instructions are executed in a privileged mode and if executed outside this mode, then get trapped. Control-sensitive instructions are used to make change in the configuration of resources used. Behavior-sensitive instructions show different behaviors which depends upon the configuration of resources and also the load as well as store operations over the virtual memory. A CPU architecture can be considered as virtualizable on one condition that if it gives support to the ability to execute the VM's privileged as well as unprivileged instructions in the user mode of CPU when the VMM is running in supervisor mode.

There is possibility of trapping the instructions in VMM when privileged instructions containing control sensitive and behavior sensitive instructions of a VM get executed. In such case the VM works as a mediator for the access of hardware from various VMs to make sure the correctness and stability of the whole system. However, it is not possible to virtualize all the CPU architectures. For example RISC CPU architectures are easily virtualized since all control sensitive as well as behavior sensitive instructions are of type privileged instructions. On the other hand, x86 CPU architectures cannot be virtualized easily because near about 10 sensitive instructions in it are not privileged instructions. These instructions cannot be trapped when executed in virtualization.

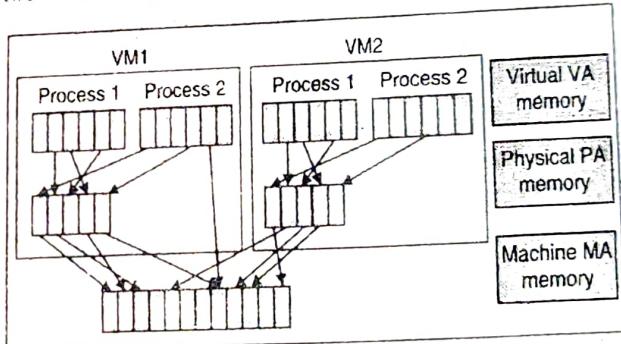
**Hardware-Assisted CPU Virtualization**

The full or para-virtualization is complicated; hence this technique is used to simplify virtualization. An additional mode known as privilege mode level or Ring-1 is added by the Intel and AMD to x86 processors. Hence operating systems can be executed on Ring 0 while the hypervisor can be executed on Ring 1. The automatic trapping of all privileged as well as sensitive instructions is done in the hypervisor. This technique helps to avoid the complexity of implementation of binary translation of full virtualization. It also allows the OS to run in VMs without need of any modification.

**Q. 10 Write a note on Memory Virtualization.****Ans. :**

Memory virtualization is a same process as of the virtual memory support which is provided by modern operating systems. In a traditional execution environment, the mappings of virtual memory to machine memory were maintained by the OS with the help of page tables. This was a one step process of mapping from virtual memory to machine memory. For the optimization of virtual memory performance, now a day all modern x86 CPUs contains MMU (Memory Management Unit) and TLB (Translation Look aside Buffer).

In the virtualization of memory, the physical system memory in RAM is shared and dynamically allocated to the VM's physical memory. That means the guest OS and the VMM maintains the two-stage mapping process; first virtual memory to physical memory and then in next phase physical memory to machine memory. The MMU virtualization which is transparent to the guest OS should also be supported. The mapping of virtual addresses to the physical memory addresses of VMs is further controlled by the guest OS. But the actual machine memory is not directly accessible to guest OS. The mapping of guest physical memory to the actual machine memory is responsibility of the VMM. Fig. 3.5 shows the two-level memory mapping procedure.



**Fig. 3.5 : Two level memory mapping procedure**

As each page table in the guest Operating System has a separate page table in the VMM related to it, the VMM page table is known as shadow page table. One more layer of indirection to virtual memory is added by the nested page tables. The virtual-to-physical translations are handled by the MMU as defined by the OS. Another set of page tables defined by the hypervisor translates the physical memory addresses to machine addresses.

A set of page tables for every process is maintained by modern operating systems, hence the shadow page tables will get flooded. It leads to high performance overhead as well as cost of memory. For virtual memory to machine memory address translation, shadow page tables are used by the VMware. Processors take help of TLB hardware for the process of mapping the virtual memory with the machine memory which helps to avoid the two levels of translation on each access. When the virtual memory is changed into physical memory by the guest OS, the shadow page tables are updated by the VMM for enabling of direct lookup.

## Q. 11 Write a note on I/O Virtualization.

**Ans. :**

I/O virtualization is used to manage the routing of various types of I/O requests in between virtual devices and the shared physical hardware.

There are three ways to implement I/O virtualization

- Full device emulation
- Para-virtualization
- Direct I/O.

Full device emulation is considered as a first approach for I/O virtualization. Generally, well-known, real-world devices are emulated by this approach. All types of functions of a device or bus infrastructure like identification, interrupts, device enumeration are simulated in software. This software works as a virtual device and placed in the VMM. The I/O access requests of the guest OS which communicates with the different types of I/O devices get trapped in the VMM. The full device emulation approach is shown in Fig. 3.6.

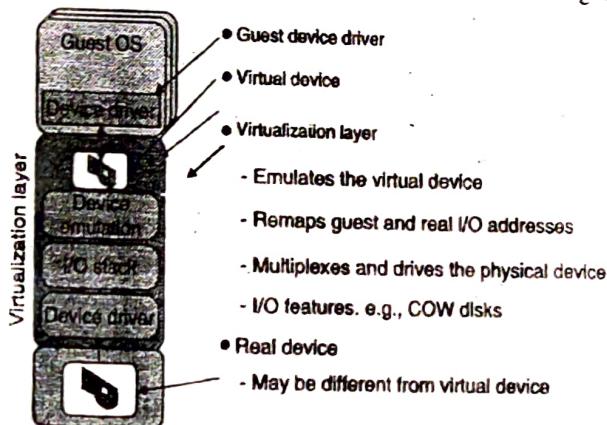


Fig. 3.6 : Device emulation for I/O virtualization

Multiple VMs can share common hardware device simultaneously. The execution of software emulation is slower than the hardware which is emulated by it. In Xen, usually the para-virtualization method of I/O virtualization is used. It is also called as the "split-driver-model" which contains frontend and backend driver. The execution of frontend driver is done in Domain U while execution of backend driver is done in Domain 0. Both of these drivers communicate with each other through block of shared memory.

The responsibility of the frontend driver is to manage the I/O requests of the guest OSes while the backend driver manages the actual I/O devices and also handles the I/O data of various VMs. The device performance of para-I/O-virtualization is better than full device emulation, but it has higher CPU overhead. The VM can access devices directly by the direct I/O virtualization. Close-to-native performance can be achieved by it by avoiding high CPU costs.

## Q. 12 Explain in brief Virtual Clusters and Resource Management.

**Ans. :**

A physical cluster is considered as a group of physical machines called as servers which are interconnected with each

other with the help of physical network such as a LAN. In this section we are going to learn virtual clusters, their properties and their potential applications.

There are three critical design issues of virtual clusters :

- Live migration of VMs
- Memory and file migrations
- Dynamic deployment of virtual clusters

In the traditional VM, the process of writing configuration information or specifying the configuration sources is done manually. When in a network, multiple VMs are joined. When more VMs join a network, an ineffective configuration leads to problems with overloading or underutilization.

## Q. 13 Write a note on Physical Vs Virtual Clusters.

**Ans. :**

In Virtual clusters, the VMs are installed at servers which are distributed from one or more physical clusters. These VMs are connected with each other logically using virtual network throughout the several physical networks. Fig. 3.7 demonstrates the concepts of virtual clusters along with physical clusters. All the virtual clusters are formed with the help of physical machines or a VM which is basically hosted by more than one physical cluster. The boundaries of virtual cluster are displayed as distinct boundaries.

VMs are provided to virtual cluster dynamically to achieve following properties :

- The virtual cluster nodes may be of two types; physical or virtual machines. It is possible to deploy more than one VMs executing with dissimilar OSes on the same physical node.
- The resources in the physical machine are managed by the host OS. The VM executes with guest OS which is probably different from the host OS.
- To improve the features like server utilization and flexibility of an application, VM combines various functionalities on the same server.

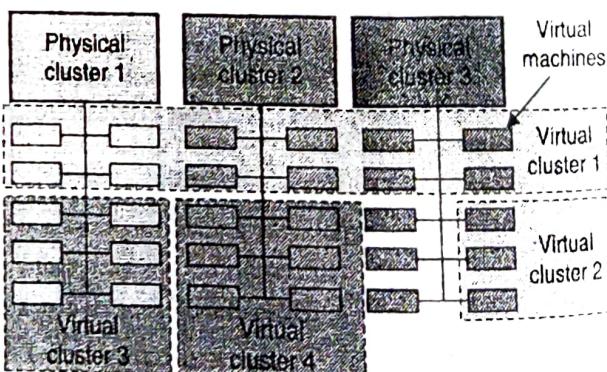


Fig. 3.7 : Cloud platform with physical and virtual clusters

It is necessary to promote features like distributed parallelism, disaster recovery as well as fault tolerance. For this purpose, VMs can be colonized (replicated) in multiple servers. Number of nodes (size) of a virtual cluster can be changed (grow

or shrink) dynamically. The failure of VMs is never leads to pull down the host system while failure of any physical nodes may lead to disability of VMs installed on it.

System virtualization is a widely used concept, hence The VMs should be effectively managed which are running on a group of physical computing nodes which are called as virtual clusters and as a result build a virtualized computing environment with higher performance. For this purpose number of things should be handled with care such as virtual cluster deployment, monitoring and management of large number of clusters, scheduling of resource, load balancing, server consolidation, fault tolerance etc. In Fig. 3.7, different virtual clusters are referred by different node colors.

In a virtual cluster system, it is necessary to store the huge number of VM images in efficient manner. Fig. 3.8 shows the concept of a virtual cluster which is based on application partitioning or customization. In the Fig. 3.8, nodes in different virtual clusters are represented by different colors. Storing the large number of VM images in the system efficiently is difficult.

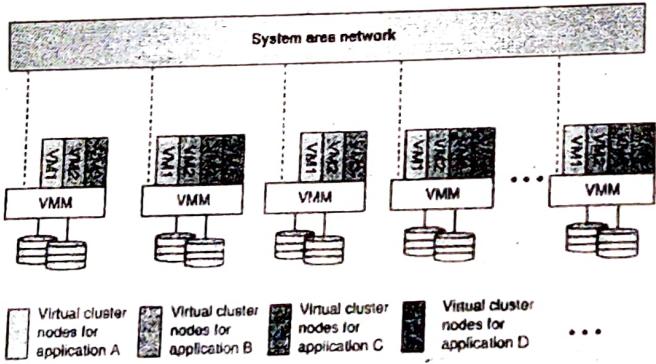


Fig. 3.8 : Virtual cluster based on application partitioning

#### Q. 14 Explain the Live VM migrations steps.

**Ans. :**

When there are mixed nodes of host and guest systems in a cluster, the usual method of operation is to execute the whole thing on the physical machine. In the situation of failure of one VM, another VM on a different node can replace its role until both the VMs are executing with same OS. That means a physical node can fail over to a VM on another host. This is vary from physical to physical failover of physical cluster in a traditional system.

The advantage is that it provides improved failover flexibility. The main disadvantage is that the role of a VM get stopped when its residing host node fails. However, the extent of this problem can be reduced with VM life migration. Fig. 3.9 shows the process of life migration of a VM in between host A to host B. the VM state file is copied by the migration from the storage area to the host machine.

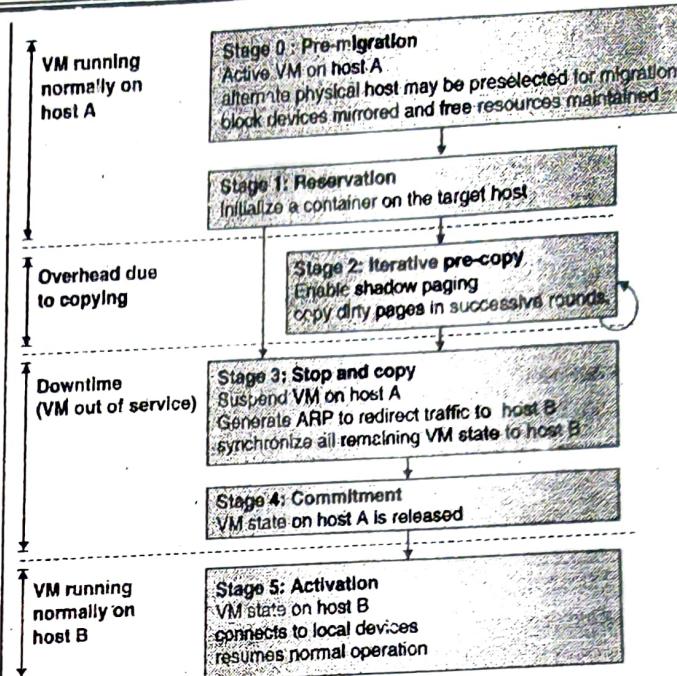


Fig. 3.9 : Live migration process of a VM from one host to another

Virtual clusters can be managed by four ways :

- First, a guest-based manager can be used with the help of which the cluster manager resides on a guest system. In such case, a virtual cluster is formed by multiple VMs. A cluster manager can be built on the host systems.
- The guest systems are supervised by the host-based manager and can be restarted on another physical machine.
- Third way for managing a virtual cluster is to utilize an independent cluster manager on both the host and guest systems.
- Finally, an integrated cluster can be used on the guest and host systems.

This indicates that the manager must be designed to distinguish between virtualized and physical resources.

A VM can be in one of the following four states :

- An inactive state which is usually described by the virtualization platform under which the VM is disabled.
- An active state which refers to a VM and instantiated at the virtualization platform for performing the real task.
- A paused state regarding to an instantiated VM which has been disabled to process a task or paused in a waiting state.

A VM is considered in the suspended state when its machine file and virtual resources are stored again in the disk.

The process of migrating a VM from one machine to another consists of the following steps :

#### Steps 0 and 1 : Start migration

This step consists of preparations for the process of migration, containing determination of the migrating VM and the destination host.

**Steps 2 : Transfer memory**

The entire execution state of VM is stored in memory, the process of sending the VM's memory to the destination node guarantees consistency in the service provided by the VM. All of the memory data is transferred.

**Step 3: Suspend the VM and copy the last portion of the data**

When the last bit of memory data is transferred, the migrating VM's execution is suspended.

**Steps 4 and 5 : Commit and activate the new host**

After finishing of the data copying on the destination host, the states are reloaded by the VM and the execution of programs is recovered, and the service provided by this VM continues.

**Q. 15 Write a note on memory migration.**

**Ans. :**

In the process of VM migration, memory migration is an important aspect. There are number of ways available for this migration. The techniques used for this process depends upon the characteristics of application or workloads which is supported by the guest OS. In the migration of a system to another physical node, the amount of memory to be migrated may be huge. Hence this migration should be done in an efficient manner.

The ISR (Internet Suspend-Resume) technique is used to exploit temporal locality since there may be considerable overlap of memory states in the suspended as well as resumed instances of a VM. For this purpose, all files in the file system are represented as a tree of small sub-files. Both the suspended and resumed VM instances have the copy of this tree. The tree-based representation of files provides an advantage that the caching makes sure the transmission of only modified files. The ISR technique generally used in such situations where there is no need of the migration of live machines.

**Q. 16 Write a note on File Network Migration.**

**Ans. :**

A VM which is in migrating state should always maintain all open network connections without depending upon the forwarding mechanisms on the original host or on support from various mechanisms regarding mobility or redirection. Each VM should be assigned a virtual IP address acknowledged to other entities. This enables the remote systems to locate as well as interact with a VM. This virtual IP address can be vary from IP address of the host machine. Each VM can also be provided with a distinct virtual MAC address.

The VMM preserve a mapping of the virtual IP address and MAC addresses with their consequent VMs. Usually, a migrating VM contains all the protocol states and also carries its IP address with it. The meaning of live migration is moving a VM among physical nodes without breaking the OS environment and applications. Now a day, this capability is widely used to provide

proficient online system maintenance, re-configuration, load balancing, and proactive fault tolerance. It provides number of expected features to fulfill requirements for computing resources in advanced computing systems such as server consolidation, ease of management and performance isolation.

**Q. 17 Explain Virtualization for Data Center Automation.**

**Ans. :**

The concept of Data centers becomes widely popular now a day. Numbers of IT companies are utilizing their resources for building new data centers. The IT giants like Google, Yahoo!, Amazon, Microsoft, HP, Apple, and IBM have invested billions of dollars in the construction and automation of data centers. In the concept of Data-center automation, the various resources available in large volumes in the Data centers like hardware, software, and database are allocated dynamically to millions of Internet users at the same time, with guaranteed QoS and effectiveness in the cost.

Virtualization is moving in the direction of increasing mobility, minimizing planned downtime (for maintenance), and rising the number of virtual clients. The most recent virtualization development denotes HA (high availability), backup services, increases in client bases as well as workload balancing. In this section we are going to learn server consolidation, virtual storage, OS support, and trust management in automated data-center designs.

**Q. 18 Write a note on Consolidation in Data Center.**

**Ans. :**

In data centers, at different times, a huge number of heterogeneous workloads can execute on servers.

These workloads can be divided into two types :

- (i) Chatty workloads
- (ii) Non-interactive workloads

The behavior of Chatty workloads is somewhat fluctuating means it may burst at some point while may be in silent state at some other point. Video web service like youtube is an example of this. The number of visitors increases in holidays and night times while few people use it during the day hours. Non-interactive workloads are not dependent upon the people's efforts or interactions for the purpose of progress after they are submitted. High-performance computing can be considered as an example of this.

At different stages, the necessities for resources of such workloads are dramatically differing. However, to ensure that the workload will always be able to handle with all types of demand levels, sufficient resources are statically allocated to workload so as to satisfy the peak demands. The granularity of resource optimization is concentrated on the CPU, memory as well as network interfaces. Hence usually most servers in data centers may be underutilized. There is wastage of huge amount of hardware, power, space and management cost of such underutilized servers.

To improve the less utility ratio of such type of hardware resources, under server consolidation the number of physical servers can be decreased. There are various techniques for server consolidation like centralized and physical consolidation. Among these techniques, the most powerful technique is virtualization-based server consolidation. The resource management of Data centers should be optimized.

In Server virtualization, the main focus is on allocation of smaller resource rather than a physical machine. Usually the resource management complexity is increased by the use of VMs. Hence there is a challenge in improvement of resource utilization and assurance of QoS in data centers.

#### **Q. 19 Write a note on Virtual Storage Management.**

**Ans. :**

Previously, "Storage Virtualization" was widely used to illustrate operations like aggregation and repartitioning of disks for the use of physical machines. In system virtualization, the virtual storage is nothing but the storage which is managed by VMMs and guest OSes. In this environment, the stored data is basically classified into two categories ; VM images and application data. In virtual environments, the VM images are important while the application data is similar to data in traditional OS environments.

In system virtualization, encapsulation and isolation are mainly the two significant aspects. In VMs the traditional operating systems with its executing applications can be encapsulated. In virtualization single operating system can be run in which multiple applications can be executed. In system virtualization, more than one VMs can be run on physical machine which are completely isolated. To gain the encapsulation as well as isolation, the system software and the hardware platform are rapidly updated. However there may lagging in storage. This is the main bottleneck of VM deployment.

In virtualization environments, a virtualization layer is placed among the hardware and traditional operating systems or sometimes modifications are done in traditional operating system to support virtualization. Because of this the storage operations becomes complicated. In contradiction to this, the storage management of the guest OS operated in such a manner that it is operating on actual hard disk while the hard disk cannot be accessed by the guest OS directly. Number of guest OSes wants the access of the hard disk when on a single physical machine multiple VMs are running. Therefore, storage management regarding the underlying VMM is comparatively more complex than that of the traditional OSes (guest OSes).

#### **Q. 20 Explain OCC.**

**Ans. :**

The Open Commons Consortium (OCC- formerly the Open Cloud Consortium) is a non-profit venture which provides cloud computing and data commons resources to support "scientific, environmental, medical and health care research. OCC manages and operates resources including the Open Science Data Cloud

(OSDC), which is a multi-petabyte scientific data sharing resource. The consortium is based in Chicago, Illinois, and is managed by the Center for Computational Science Research. The OCC was among six partners engaged by the Global Lambda Integrated Facility to establish a testbed for a 100 Gbps data transmission capability.

The OCC is divided into Working Groups which include :

- a. **The Open Science Data Cloud :** This is a working group that manages and operates the Open Science Data Cloud (OSDC), which is a petabyte scale science cloud for researchers to manage, analyze and share their data.

Individual researchers may apply for accounts to analyze data hosted by the OSDC. Research projects with TB-scale datasets are encouraged to join the OSDC and contribute towards its infrastructure.

- b. **Project Matsu :** Project Matsu is collaboration between the NASA Goddard Space Flight Center and the Open Commons Consortium to develop open source technology for cloud-based processing of satellite imagery to support the earth science research community as well as human assisted disaster relief.

- c. **The Open Cloud Testbed :** This working group manages and operates the Open Cloud Testbed. The Open Cloud Testbed (OCT) is a geographically distributed cloud testbed spanning four data centers and connected with 10G and 100G network connections.

The OCT is used to develop new cloud computing software and infrastructure.

#### **1. The Biomedical Data Commons**

The Biomedical Data Commons (BDC) is cloud-based infrastructure that provides secure, compliant cloud services for managing and analyzing genomic data, electronic medical records (EMR), medical images, and other PHI data. It provides resources to researchers so that they can more easily make discoveries from large complex controlled access datasets. The BDC provides resources to the institutions in the BDC Working Group. It is an example of what is sometimes called condominium model of sharing research infrastructure in which the research infrastructure is operated by a consortium of educational and research organizations and provides resources to the consortium.

#### **2. NOAA Data Alliance Working Group**

The OCC National Oceanographic and Atmospheric Administration (NOAA) Data Alliance Working Group supports and manages the NOAA data commons and the surrounding community interested in the open redistribution of NOAA datasets.

In 2015, the OCC was accepted into the Matter healthcare community at Chicago's historic Merchandise Mart. Matter is a community healthcare entrepreneurs and industry leaders working together in a shared space to individually and collectively fuel the

future of healthcare innovation. In 2015, the OCC announced collaboration with the National Oceanic and Atmospheric Administration (NOAA) to help release their vast stores of environmental data to the general public. This effort is managed by the OCC's NOAA data alliance working group.

#### **Q. 21 Explain Open Virtualization Format.**

**Ans. :**

Open Virtualization Format (OVF) is an open standard for packaging and distributing virtual appliances or, more generally, software to be run in virtual machines. The standard describes an "open, secure, portable, efficient and extensible format for the packaging and distribution of software to be run in virtual machines". The OVF standard is not tied to any particular hypervisor or instruction set architecture. The unit of packaging and distribution is a so-called OVF Package which may contain one or more virtual systems each of which can be deployed to a virtual machine.

#### **Q. 22 Explain Standards for Application Developers.**

**Ans. :**

The standards for application developer are provided to guarantee uniform, consistent as well as high quality software solutions. For programmers, the programming standards are always very much important. It is estimated that near about 80% of the total costing of an application is spent on its maintenance. Also the maintenance person and developers are always different. Programming standards are basically used to enhance the readability of the application, which helps the programmer in understanding of new code in a fast and detail manner.

In an application it is very important that the source code must be well packaged and fulfills are the industry standards when compared to the products of competitors. All the developers should conform to standards to work them properly. In the following sections, we are going to see some application standards which are generally used in the Internet in browsers, for transferring data, and securing data.

#### **Q. 23 Write a note on Solution Stacks : LAMP and LAPP.**

**Ans. :**

LAMP stands for

- (i) Linux, (ii) Apache,
- (iii) MySQL, and (iv) PHP (or Perl or Python.)

LAMP is a widely used open source solution generally used to execute dynamic web sites and servers. LAMP is basically used to develop as well as deploy advanced web applications which require high-performance and reliability. All the mentioned technologies are used in combination for creating web server infrastructure or programming environment for the development of software.

Initially at the time of creation of these technologies, the creators do not have any intention to use them in combination, but later on the LAMP combination become very much popular because of number of features like open source, cost effectiveness and easy distribution of components. Now a day almost in all Linux distributions, the LAMP combination is provided in bundle. In the combined form, all these technologies represent a solution stack that supports various types of application servers.

#### **LAPP**

LAPP stands for

- (i) Linux (ii) Apache
- (iii) PostgreSQL (iv) PHP(or Perl or Python)

The LAPP stack also is an open source platform which helps to execute dynamic web sites and servers. As compared to LAMP, the LAPP is considered as more powerful combination. The technologies under LAPP are mature and provide very strong foundation to develop and deploy web applications with high performance. For Apache2 and PostgreSQL the LAPP combination provides SSL, PHP, Python, and Perl support. In LAPP an administration front end for PostgreSQL is available. It also provides web based administration modules for the configuration of Apache2 as well as PHP. By default the encryption of PostgreSQL password is also enabled.

The reliability of PostgreSQL user depends upon whether he is connected over local UNIX sockets or not. As compared to LAMP, the LAPP is considered as more secure. Both the stacks LAMP and LAPP are good but the selection out of them by the developer depends upon application purpose and security concerns at the time of use by the user.

#### **Q. 24 Write a note on Syndication.**

**Ans. :**

**Content syndication** is mechanism through which the new content and headlines are conveniently made accessible to the citizens through Really Simple Syndication (RSS) and other syndication standards which are available online. Now a day governments of various countries make available more and more information online.

The new or timely content is often buried because of the complexity in browsing the websites. The new content can be presented dynamically which only can be considered as the first step. The standards for messaging headlines and content can be shared with the help of syndication standards like RSS. This helps the government to keep control on the window of content throughout the web sites which selects to display the government's headlines.

Citizens can use the "newsreaders" to aggregate and display headlines with the help of standalone applications or as part of their web sites. From multiple agency web sites, it is possible to aggregate and combine headlines automatically by the portals.

Because of it, the value of distributed effort can be shared, which is obviously more sustainable. Automatic aggregation of Press releases is possible from various systems if all of them offer RSS feed with content which are tagged with similar metadata.

#### **Q. 25 Explain Atom and Atom Publishing Protocol.**

**Ans. :**

The term Atom is basically applies to a group of related standards. The Atom Syndication Format is nothing but the XML language which is used for web feeds, while the concept of Atom Publishing Protocol (AtomPub or APP) leads to a simple HTTP-based protocol which is used to create and update web resources which are also called as web feeds. The software programs can use web feeds to check for updates which are published on a web site. A site owner can use special software like content management system for the purpose of providing web feed. This software can publish list of currently available articles or content in a very standard and machine readable format.

The downloading of feed can be implemented by the websites which syndicate content from the feed. For this purpose, another option available is feed reader programs where user can subscribe to feeds and view their content. In a feed, lots of contents are available like entries which may be headlines, articles with full-text, excerpts, summaries, and links to various types of content of web site with metadata. Atom can be considered as an alternative option to RSS. This format is developed by Ben Trott. According to him, there are some limitations and flaws in Atom format like lack of ongoing innovation and its need to always stay backward compatible.

The weblog community uses the Web feeds to share headlines of the recent entries or their full text, and also the attached multimedia files. The other web sites can include the

weblog's *syndicated* headline or "headline and short summary" feeds under various usage regarding agreements. Now a day the Atom and other web syndication formats are used for number of purposes such as journalism, online marketing, "bug reporting", or number of other activities which involve periodic updates or publications.

It is possible to export an entire blog or part of it through Atom for the purpose of backup or for importing it into another blogging system. An application called as a feed reader or aggregator is used to check web pages on behalf of user and show any updated articles that it gets. Finding web feeds on major or small web sites is common. In some websites facility is given to user to select option from RSS or Atom-formatted web feeds. While some websites offer only either RSS or Atom. Sometimes the web feeds are in Atom format in particular blog and wiki sites. Client-side readers and aggregators can be developed as standalone programs or also as extensions to existing applications like web browsers.

Now a day the various types of Browsers are shifting toward integrated feed reader methods. These types of programs are available for different types of operating systems. There is no need of software installation for Web-based feed readers and news aggregators and user's feeds can be made available on any computer with web access. Sometimes, the web feeds are syndicated into news feeds by some aggregators. E.g. gathering all cricket regarding items from several sports feeds and provide as a new cricket feed. Number of search engines like Google is available which offer search functionality over content published through these web feeds.

## **Chapter 4 : Amazon Web Services**

#### **Q. 1 Write a note on Services offered by Amazon.**

**Ans. :**

Amazon.com has provided a subsidiary known as **Amazon Web Services (AWS)**. AWS provides cloud computing platforms for various elements like individuals, companies and governments on demand basis. AWS gives a free trial option for one year. AWS has a full-fledged virtual cluster of computers which it makes available to subscribers via internet. The virtual computers of AWS provide most of the features of real computers such as hardware, software (OS, web servers, databases, CRM etc) and networking.

Console I/O (keyboard, display, and mouse) is also virtualized by the AWS which lets the subscribers to have connection with AWS system with the help of a modern browser. In virtual computer, the browser can be considered as a window which helps subscribers to login, configure and utilize the virtual systems just like real physical computers. The subscribers can provide internet-based services to their customers using AWS system.

The charges are depending upon the various elements like usage, hardware, Operating System, software, features regarding networking. There are various options for subscribers as per their need like only one virtual computer or a cluster of VMs. It is responsibility of Amazon to manage and upgrade the system as well as provide industry-standard security. Since 2016 Amazon provides 70 services including computing, storage, networking, database, analytics, application services, deployment, management, mobile, developer tools, and tools for the Internet of Things.

EC2 (Amazon Elastic Compute Cloud) and S3 (Amazon Simple Storage Service) are the two most popular services. Most of the AWS services are not directly exposed to the subscribers; rather their functionality is made available with the help of various APIs. The AWS services are accessed over HTTP, by means of the REST architectural style and SOAP (Simple Object Access Protocol) protocol. The basic aim of Amazon to build AWS is that users should get large scale computing capacity in a quick and cost effective manner as compare to building an actual physical server.

farm. Virtual machines can share the computing resources safely and flexibly. The IaaS model of cloud is adopted by Amazon for the purpose of providing public cloud services.

Fig. 4.1 shows the AWS architecture.

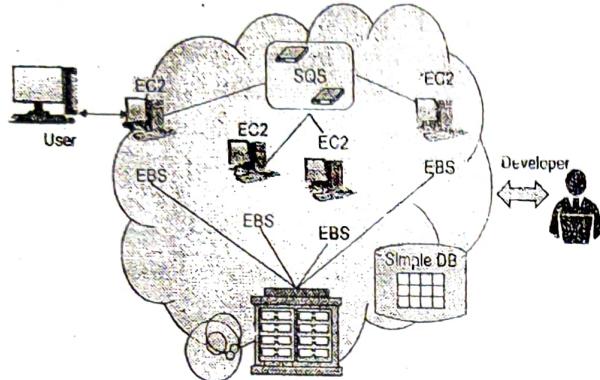


Fig. 4.1 : AWS Architecture

EC2 (Amazon Elastic Compute Cloud) provides virtualized platforms for the host Virtual Machines where the cloud application can be executed. Simple Storage Service (S3) provides the storage service which is OOP based for users. Block storage interface is provided by the Elastic Block Service (EBS), which supports the traditional applications. SQS (Simple Queue Service) is used to guarantee a reliable message service in between two different processes. It is possible to keep the message reliably even if the execution of processes is stopped. The objects can be retrieved by the user with the help of SOAP by using browsers or applications which support the SOAP standards.

Queuing and notification services like SQS and SNS are also provided by Amazon. These services are implemented in the AWS cloud. In cloud, the brokering systems are generally executed in efficient manner which provides office support of smart-phones and tablets. For development of cloud applications, Amazon provides more flexible cloud computing platform. Amazon cloud provides platform for small and medium-size companies. Through AWS these companies can provide service to large numbers of Internet users and also can generate revenue through those paid services. The incoming application traffic is automatically distributed by the ELB across multiple Amazon EC2 instances. It helps user to avoid the nodes which are currently non-operating and equalize load on functioning images.

CloudWatch enables the autoscaling as well as ELB to monitor the running instances. CloudWatch is a web service which is used to monitor AWS cloud resources. CloudWatch helps user to improve resource utilization and overall performance of operation with demand patterns. It also enable user to concentrate on CPU utilization, disk reads / writes and most importantly the network traffic. A RDS (Relational Database Service) is provided by Amazon with messaging interface. The capability of Elastic MapReduce is same as of the Hadoop running on the basic EC2 offering. It is possible to transfer large volumes of data to and from EC2 with the help of shipping of physical disks by the AWS Import/Export.

A content distribution network is implemented by the Amazon CloudFront. Amazon provides an easy billing and account management service known as Amazon DevPay. This service helps user to sell the applications which are built and executed on the top of AWS. For payment by customers, Amazon provides commercial system known as FPS (Flexible Payments Service) which gives an easy payment gateway. The customer information which is already available on Amazon like login credentials, address for shipping, and payment ways, can be used by the customers for payment of services. Amazon provides another service known as FWS to access fulfilment capabilities of Amazon using a simple web service interface. On behalf of customer, merchants can send order information to Amazon.

## Q. 2 Explain various features of EC2.

Ans. : Features of EC2

### 1. Persistent Storage

There are two options for the storage of its boot disk or "root device" in EC2.

- (a) "Instance-store" disk as a root device
- (b) Use an EBS volume as a root device.

The first option "Instance-store" is a temporary form of storage. In case of rebooting of an EC2 instance, the data is survived but the stoppage or termination of the instance leads to loss of data. The second option EBS (Amazon Elastic Block Store) gives raw block devices. The Amazon EC2 instances can be attached with these raw block devices. Then these block devices can be utilized as of raw block devices. EBS also supports various types of advanced storage features such as snap-shutting and cloning. The maximum size of EBS volumes is 1TB.

In EBS failure of any single component does not lead to loss of data as the Volumes of EBS are built on replicated storage. Amazon introduces EBS in 2008 publicly. The act of EBS volumes is similar to hard drives of a real server. The persistence storage provided by the EBS is independent of the lifetime of the EC2 instance. Amazon's disk arrays back the EBS which appeared as block devices to the OS. Device can be used by the OS as per need. In general the loading of file system is implemented and the volume works as a hard drive. In another use, two or more EBS volumes can be combined to creates RAID arrays which helps to increase the speed and as well as reliability of EBS. The storage volumes in the range of 1GB to 1TB can be set up and managed by the users. Snapshots from a GUI tool or the API are supported by the volumes.

From instances, the attachment or detachment of EBS volumes can be done in running mode or when moved from one instance to another. S3 stands for Simple Storage Service which is a storage system where the data is accessible to EC2 instances, or directly over the network. This communication is done over HTTP. In the same region, there is no charge for communications between

EC2 instances and S3 storage. Normal rates are applied when S3 data is accessed in different regions. For Linux also S3 file-systems are available. This system mounts a remote S3 file store on an EC2 image just same as of the local storage.

## 2. Elastic IP addresses

In traditional data centres, there is a feature called as static IP address which resembles with Amazon's elastic IP address feature. The mapping of elastic IP address with any virtual machine instance can be done without the help of network administrator. It indicates that an Elastic IP Address is concerned with the account and not to the instance of virtual machine. It remains unless and until it is not removed explicitly, and also be associated with the account even if there is no instance associated with it.

## 3. Amazon CloudWatch

Amazon CloudWatch is a web service which is basically designed to provide real-time monitoring service for the customers of Amazon's EC2 on the utilization of various resources like CPU, disk and network. CloudWatch requires additional software on the instance to provide memory, disk space, or load average metrics. Example scripts is provided by the Amazon for Linux instances.

The management console of AWS is used to aggregate and provide the data. If customer wants to monitor the EC2 resources with the help of their own enterprise monitoring software, then command line tools and Web API's can be used to access data. The auto-scaling feature is enabled by the metrics collected by Amazon CloudWatch which helps to add remove EC2 instances dynamically. The number of monitoring instances is considered for the charges to customer.

## 4. Automated scaling

This feature of EC2 enables the automatic adaption of computing capacity to site traffic. The two auto scaling mechanisms; schedule-based (e.g. time-of-the-day) and rule-based (e.g. CPU utilization thresholds) are provides feature like simple to use and efficiency. But there is an important problem that VMs may require lot of time to be ready to use. This is problematic when there is a time binding. Image size, VM type, data center locations are factors responsible for VM start-up times. However, one potential problem is that VMs may take up to several minutes to be ready to use, which are not suitable for time critical applications. The VM start-up times are dependent on image size, VM type, data center locations, etc.

## 5. Reliability

Availability Zones are engineered by the Amazon to make EC2 more fault-tolerant. The Availability zones have different infrastructures. Higher availability is achieved by the applications which execute on multiple availability zones. The geographical location of instances can be controlled by the users through EC2. It helps for optimization of latency and high levels of redundancy. For example, downtime can be minimized server instances can be set up in more than one zone which are insulated from each other for most causes of failure such that one backs up the other.

## Q. 3 Explain steps to configure server for EC2.

### Ans. :

Following steps should be performed to get set up to use Amazon EC2 :

### 1. Sign Up for AWS

When you sign up for Amazon Web Services (AWS), your AWS account is automatically signed up for all services in AWS, including Amazon EC2. You are charged only for the services that you use. With Amazon EC2, you pay only for what you use. If you are a new AWS customer, you can get started with Amazon EC2 for free. If you have an AWS account already, skip to the next task. If you don't have an AWS account, use the following procedure to create one.

### To create an AWS account

Open <https://aws.amazon.com/>, and then choose Create an AWS Account.

Follow the online instructions.

- (i) Part of the sign-up procedure involves receiving a phone call and entering a PIN using the phone keypad.
- (ii) Note your AWS account number, because you'll need it for the next task.

### 2. Create an IAM User

Services in AWS, such as Amazon EC2, require that you provide credentials when you access them, so that the service can determine whether you have permission to access its resources. The console requires your password. You can create access keys for your AWS account to access the command line interface or API. However, it is not recommended that you access AWS using the credentials for your AWS account; it better that you use AWS Identity and Access Management (IAM) instead.

Create an IAM user, and then add the user to an IAM group with administrative permissions or grant these user administrative permissions. You can then access AWS using a special URL and the credentials for the IAM user. If you signed up for AWS but have not created an IAM user for yourself, you can create one using the IAM console.

To create an IAM user for yourself and add the user to an Administrators group :

1. Use your AWS account email address and password to sign in as the AWS account root user to the IAM console at <https://console.aws.amazon.com/iam/>.
2. In the navigation pane of the console, choose *Users*, and then choose *Add user*.
3. For User name, type **Administrator**.
4. Select the check box next to **AWS Management Console access**, select **Custom password**, and then type the new user's password in the text box. You can optionally select **Require**

- password reset* to force the user to create a new password the next time the user signs in.
5. Choose Next: *Permissions*.
  6. On the Set permissions page, choose *Add user to group*.
  7. Choose Create group.
  8. In the Create group dialog box, for Group name type **Administrators**.
  9. For Filter policies select the check box for AWS managed - job function.
  10. In the policy list, select the check box for Administrate or Access. Then choose Create group.
  11. Back in the list of groups, select the check box for your new group. Choose *Refresh* if necessary to see the group in the list.
  12. Choose Next: *Review* to see the list of group memberships to be added to the new user. When you are ready to proceed, choose *Create user*.

You can use this same process to create more groups and users, and to give your users access to your AWS account resources. To sign in as this new IAM user, sign out of the AWS console, then use the following URL, where *your\_aws\_account\_id* is your AWS account number without the hyphens (for example, if your AWS account number is 1234-5678-9012, your AWS account ID is 123456789012) :

[https://your\\_aws\\_account\\_id.signin.aws.amazon.com/console/](https://your_aws_account_id.signin.aws.amazon.com/console/)

Enter the IAM user name (not your email address) and password that you just created. When you're signed in, the navigation bar displays "*your\_user\_name* @ *your\_aws\_account\_id*". If you don't want the URL for your sign-in page to contain your AWS account ID, you can create an account alias. From the IAM console, choose **Dashboard** in the navigation pane. From the dashboard, choose **Customize** and enter an alias such as your company name. To sign in after you create an account alias, use the following URL :

[https://your\\_account\\_alias.signin.aws.amazon.com/console/](https://your_account_alias.signin.aws.amazon.com/console/)

To verify the sign-in link for IAM users for your account, open the IAM console and check under **IAM users sign-in** link on the dashboard.

### 3. Create a Key Pair

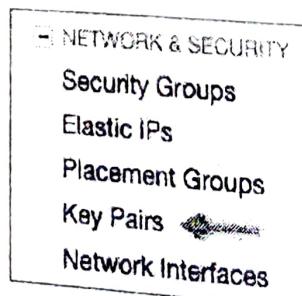
AWS uses public-key cryptography to secure the login information for your instance. A Linux instance has no password; you use a key pair to log in to your instance securely. You specify the name of the key pair when you launch your instance, then provide the private key when you log in using SSH. If you haven't created a key pair already, you can create one using the Amazon EC2 console. Note that if you plan to launch instances in multiple regions, you'll need to create a key pair in each region.

### To create a key pair

1. Sign in to AWS using the URL that you created in the previous section
2. From the AWS dashboard, choose **EC2** to open the Amazon EC2 console.
3. From the navigation bar, select a *region* for the key pair. You can select any region that's available to you, regardless of your location. However, key pairs are specific to a region; for example, if you plan to launch an instance in the US East (Ohio) Region, you must create a key pair for the instance in the US East (Ohio) Region.



4. In the navigation pane, under **NETWORK & SECURITY**, choose **Key Pairs**.



5. Choose **Create Key Pair**.
6. Enter a name for the new key pair in the Key pair name field of the Create Key Pair dialog box, and then choose **Create**. Use a name that is easy for you to remember, such as your

IAM user name, followed by -key-pair, plus the region name. For example, me-key-pair-useast2.

7. The private key file is automatically downloaded by your browser. The base file name is the name you specified as the name of your key pair, and the file name extension is .pem. Save the private key file in a safe place.
8. If you will use an SSH client on a Mac or Linux computer to connect to your Linux instance, use the following command to set the permissions of your private key file so that only you can read it.

`chmod 400 your_user_name-key-pair-region_name.pem`

If you do not set these permissions, then you cannot connect to your instance using this key pair.

#### 4. Create a Virtual Private Cloud (VPC)

Amazon VPC enables you to launch AWS resources into a virtual network that you've defined. If you have a default VPC, you can skip this section and move to the next task, Create a Security Group. To determine whether you have a default VPC, see Supported Platforms in the Amazon EC2 Console. Otherwise, you can create a nondefault VPC in your account using the steps below.

To create a nondefault VPC

1. Open the Amazon VPC console at <https://console.aws.amazon.com/vpc/>.
2. From the navigation bar, select a *region* for the VPC. VPCs are specific to a region, so you should select the same region in which you created your key pair.
3. On the VPC dashboard, choose Launch VPC Wizard.
4. **On the Step 1 :** Select a *VPC Configuration page*, ensure that VPC with a Single Public Subnet is selected, and choose *Select*.
5. **On the Step 2 :** VPC with a Single Public Subnet page, enter a friendly name for your VPC in the VPC name field. Leave the other default configuration settings, and choose *Create VPC*. On the confirmation page, choose *OK*.

#### 5. Create a Security Group

Security groups act as a firewall for associated instances, controlling both inbound and outbound traffic at the instance level. You must add rules to a security group that enable you to connect to your instance from your IP address using SSH. You can also add rules that allow inbound and outbound HTTP and HTTPS access from anywhere.

#### Q. 4 Write a note on Virtual Amazon Cloud.

Ans. :

Amazon Virtual Private Cloud (Amazon VPC) enables developers to establish a virtual network for the purpose of launching resources in a section of the Amazon Web Services cloud which is totally isolated. AWS users has facility to connect

to Amazon VPC using an Internet gateway, an on-premises datacentre via the Hardware Virtual Private Network (VPN) Connection Tool or via several type of AWS tools as well as other vendor VPCs. Better granular control of the cloud network is possible with the help of Amazon VPC that provides an additional layer of security for the purpose of workloads and data.

It is easy for users to define network configurations like IP address range and route tables, and handle network gateways and subnets; subnets are minor independent parts of the entire network. With the help of Amazon VPC, developer is able to generate security groups to fix limits on inbound as well as outbound traffic to Amazon Elastic Compute Cloud (EC2) instances and network access control lists to allow or deny traffic to subnets. While generating an EC2 instance, users can set an IP address to the particular instance until it is in the IP address range of the related subnet, not assigned to any other interface and not held in reserve by Amazon.

VPC routers are used to set communication between instances in various subnets. Routers also enable communication between subnets, Internet gateways and virtual private gateways. Internet gateways enable communication between instances and EC2 resources which reside outside of a VPC and in different regions. AWS provides various options to link EC2 instances in the environment of VPC to Amazon Simple Storage Service. AWS offers a "Start VPC Wizard," that provides four options of basic network architectures which dictate whether the category of subnets is public or private and whether the user likes to get access through the Hardware VPN.

VPCs have ability to span more than one AZ (Availability Zones), but it is necessary that the subnets are in one AZ. User may additionally prefer the deployment of Amazon CloudWatch and Auto Scaling within an Amazon VPC for the purpose of monitoring resources and let them to meet spikes in workload demand. When there is provision of EC2 resources by the AWS users first, for their launching default VPC is considered if the subnet ID is not specified. The depth of security as well as network control provided by default VPCs is similar to regular Amazon VPCs, but one important difference is that they enable user to create and manage resources with the help of AWS Management Console, EC2-Classic command line or API.

The size of each VPC is restricted by AWS; user does not have facility to change the size after the VPC has been created. Also there is limit of Amazon VPC of two hundred subnets per VPC, all of them are able to support at least fourteen IP addresses. AWS has further restrictions per account / per region, containing restrictive number of VPCs to 5, the number of Elastic IP address is to 5, the number of Internet gateways per VPC to 1, the number of virtual private gateways to 5 and the number of customer gateways to fifty.

**Q. 5 Which are the key storage options provided by Amazon?**

**Ans. :**

Amazon provides following key storage options :

**1. S3 (Simple Storage Service)**

This service Provides secure, durable, highly-scalable object storage to the developers as well as IT teams. Simple to use, with easy web service interface to store and access any quantity of data from any location on the web.

**2. CloudFront**

It is considered as a content delivery web service. It collaborates with other AWS products to provide developers as well as businesses a simple way to provide content to end users with minimum latency, enhanced data transfer speeds, and without any minimum usage commitments.

**3. (EFS) Elastic File System**

It is a file storage service for Amazon EC2 instances. EFS is simple for use and offer an easier interface which allows to generate and configure file systems quickly and easily. In EFS, storage capacity is elastic, growing and shrinking automatically as files are added and removed, so the applications get storage whenever they need.

**4. Glacier**

Glacier is a secure, durable as well as very low-cost storage service for data archiving and long-term backup. To keep costs low, Amazon Glacier is optimized for rarely accessed data where a retrieval time of several hours is suitable.

**5. Import/Export Snowball**

It is a petabyte-scale data transport solution that uses secure appliances to transfer large amounts of data in and out of AWS. Using snowball addresses common challenges with large-scale data transfers includes high network cost, long transfer times, and security concerns. Transferring data with Snowball is simple, fast, secure, and cost wise can be as little as one-fifth the cost of high-speed internet.

**6. Storage Gateway**

It is a service connecting an on-premises software appliance with cloud-based storage to provide seamless and secure integration between organizations' on-premises IT environment and AWS's storage infrastructure. The service allows storing data securely in the AWS cloud for scalable and cost-effective.

**Q. 6 Write a note on Amazon EBS with its advantages.**

**Ans. :**

Amazon Elastic Block Store (Amazon EBS) provides persistent block storage volumes for use with Amazon EC2 instances in the AWS Cloud. Each Amazon EBS volume is automatically replicated within its Availability Zone to protect user

from component failure, offering high availability and durability. Amazon EBS volumes offer the consistent and low-latency performance needed to run the workloads. With Amazon EBS, user can scale the usage up or down within minutes – all while paying a low price for only what user provision.

Amazon EBS is designed for application workloads that benefit from fine tuning for performance, cost and capacity. Typical use cases include Big Data analytics engines (like the Hadoop/HDFS ecosystem and Amazon EMR clusters), relational and NoSQL databases (like Microsoft SQL Server and MySQL or Cassandra and MongoDB), stream and log processing applications (like Kafka and Splunk), and data warehousing applications (like Vertica and Teradata).

**Q. 7 How to create an Elastic Block Store Volume?**

**Ans. :**

User can create an Amazon EBS volume that he/she can then attach to any EC2 instance within the same Availability Zone. User can choose to create an encrypted EBS volume, but encrypted volumes can only be attached to selected instance types. User can use IAM policies to enforce encryption on new volumes. User can also create and attach EBS volumes when he/she launch instances by specifying a block device mapping.

User can restore volumes from previously created snapshots. User can apply tags to EBS volumes at the time of creation. With tagging, user can simplify tracking of his/her Amazon EC2 resource inventory. Tagging on creation can be combined with an IAM policy to enforce tagging on new volumes. If User is creating a volume for a high-performance storage scenario, he/she should make sure to use a Provisioned IOPS SSD (io1) volume and attach it to an instance with enough bandwidth to support the application, such as an EBS-optimized instance or an instance with 10-Gigabit network connectivity. The same advice holds for Throughput Optimized HDD (st1) and Cold HDD (sc1) volumes.

New EBS volumes receive their maximum performance the moment that they are available and do not require initialization (formerly known as pre-warming). However, storage blocks on volumes that were restored from snapshots must be initialized (pulled down from Amazon S3 and written to the volume) before user can access the block. This preliminary action takes time and can cause a significant increase in the latency of an I/O operation the first time each block is accessed. For most of the applications, amortizing this cost over the lifetime of the volume is acceptable. Performance is restored after the data is accessed once.

**Steps to create an EBS volume using the console**

1. Open the Amazon EC2 console at <https://console.aws.amazon.com/ec2/>.
2. From the navigation bar, select the region in which you would like to create your volume. This choice is important because some Amazon EC2 resources can be shared between

- regions, while others can't. In the navigation pane, choose **ELASTIC BLOCK STORE, Volumes**.
3. Choose Create Volume.
  4. For Volume Type, choose a *volume type*.
  5. For Size (GiB), type the size of the volume.
  6. With a Provisioned IOPS SSD volume, for IOPS, type the maximum number of input/output operations per second (IOPS) that the volume should support.
  7. For Availability Zone, choose the *Availability Zone* in which you want to create the volume. EBS volumes can only be attached to EC2 instances within the same Availability Zone. (Optional) To create an encrypted volume, select the *Encrypted* box and choose the *master key* you want to use when encrypting the volume.
  8. You can choose the default master key for your account, or you can choose any customer master key (CMK) that you have previously created using the AWS Key Management Service.
  9. Available keys are visible in the Master Key menu, or you can paste the full ARN of any key to which you have access.
  10. (Optional) Choose *Create additional tags* to add tags to the volume. For each tag, provide a tag key and a tag value.
  11. Choose Create Volume.

#### **Q. 8 Write a note on "Adding an EBS Volume to an Instance".**

**Ans. :**

You can attach an available EBS volume to one of your instances that is in the same Availability Zone as the volume.

#### **Prerequisites**

Determine how many volumes you can attach to your instance. If a volume is encrypted, it can only be attached to an instance that supports Amazon EBS encryption.

If a volume has an AWS Marketplace product code :

- (i) The volume can only be attached to a stopped instance.
- (ii) You must be subscribed to the AWS Marketplace code that is on the volume.
- (iii) The configuration (instance type, operating system) of the instance must support that specific AWS Marketplace code. For example, you cannot take a volume from a Windows instance and attach it to a Linux instance.
- (iv) AWS Marketplace product codes are copied from the volume to the instance.

#### **To attach an EBS volume to an instance using the console**

1. Open the Amazon EC2 console at <https://console.aws.amazon.com/ec2/>.
2. In the navigation pane, choose Elastic Block Store Volumes.
3. Select an available volume and choose *Actions, Attach Volume*.

4. For Instance, start typing the name or ID of the instance. Select the instance from the list of options (only instances that are in the same Availability Zone as the volume are displayed).
5. For Device, you can keep the suggested device name, or type a different supported device name.
6. Choose *Attach*.
7. Connect to your instance and mount the volume.

#### **To attach an EBS volume to an instance using the command line**

You can use one of the following commands.

- (i) *Attach-volume* (AWS CLI)
- (ii) *Add-EC2Volume* (AWS Tools for Windows PowerShell)

#### **Q. 9 Explain "Snapshotting an EBS Volume and Increasing Performance".**

**Ans. :**

User can back up the data on his/her Amazon EBS volumes to Amazon S3 by taking point-in-time snapshots. Snapshots are *incremental* backups, which means that only the blocks on the device that have changed after the most recent snapshot are saved. This minimizes the time required to create the snapshot and saves storage costs by not duplicating data. When user delete a snapshot, only the data unique to that snapshot is removed. Each snapshot contains all of the information needed to restore user's data (from the moment when the snapshot was taken) to a new EBS volume.

When user creates an EBS volume based on a snapshot, the new volume begins as an exact replica of the original volume that was used to create the snapshot. The replicated volume loads data lazily in the background so that user can begin using it immediately. If user access data that hasn't been loaded yet, the volume immediately downloads the requested data from Amazon S3, and then continues loading the rest of the volume's data in the background. User can track the status of EBS snapshots through CloudWatch Events.

#### **How Incremental Snapshots Work ?**

Here we are going to see how an EBS snapshot captures the state of a volume at a point in time, and also how successive snapshots of a changing volume create a history of those changes. In the Fig. 4.2, Volume 1 is shown at three points in time. A snapshot is taken of each of these three volume states.

- (i) In State 1, the volume has 10 GiB of data. Because Snap A is the first snapshot taken of the volume, the entire 10 GiB of data must be copied.
- (ii) In State 2, the volume still contains 10 GiB of data, but 4 GiB have changed. Snap B needs to copy and store only the 4 GiB that changed after Snap A was taken. The other 6 GiB of unchanged data, which are already copied and stored in Snap A, are referenced by Snap B rather than (again) copied. This is indicated by the dashed arrow.
- (iii) In State 3, 2 GiB of data have been added to the volume, for a total of 12 GiB. Snap C needs to copy the 2 GiB that were

added after Snap B was taken. As shown by the dashed arrows, Snap C also references the 4 GiB of data stored in Snap B, and the 6 GiB of data stored in Snap A. The total storage required for the three snapshots is 16 GiB.

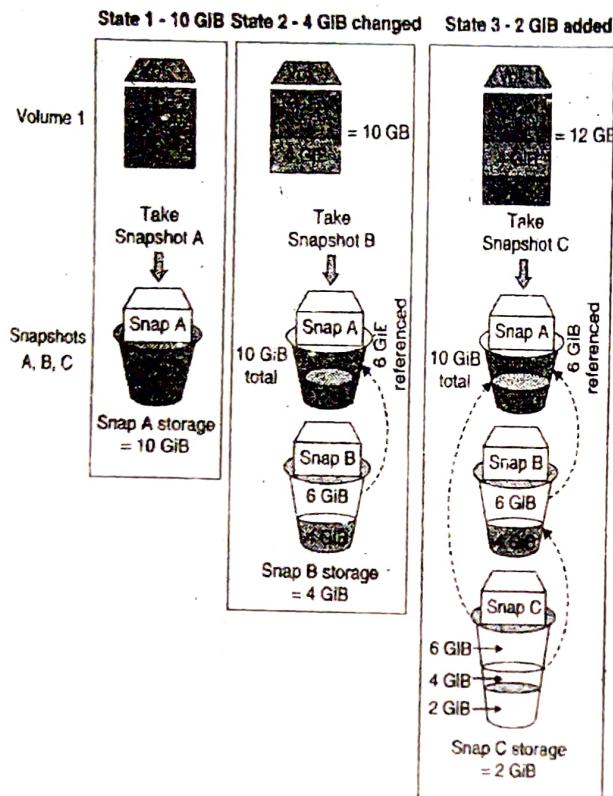


Fig. 4.2 : Relations among Multiple Snapshots of a Volume

**Q. 10 Explain the steps to create an Amazon S3 Bucket and managing associated objects.**

**Ans. :**

#### (A) Creating Bucket

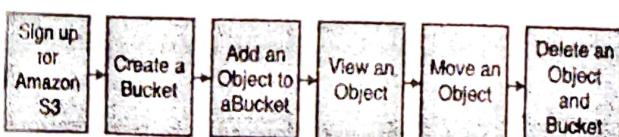
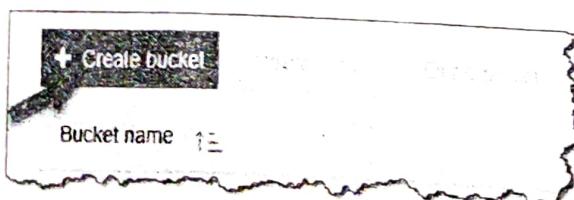


Fig. 4.3: Creating Bucket

Following steps are performed to create an Amazon S3 bucket :

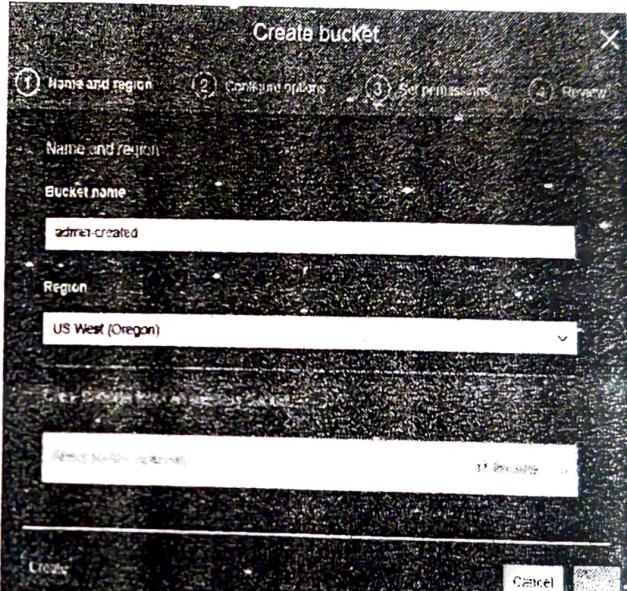
1. Sign in to the AWS Management Console and open the Amazon S3 console at <https://console.aws.amazon.com/s3/>.
2. Choose Create bucket.



3. In the Bucket name field, type a unique DNS-compliant name for your new bucket. (The example screen shot uses the bucket name admin-created. You cannot use this name

because S3 bucket names must be unique.) Create your own bucket name using the follow naming guidelines :

- i) The name must be unique across all existing bucket names in Amazon S3.
  - ii) After you create the bucket you cannot change the name, so choose wisely.
  - iii) Choose a bucket name that reflects the objects in the bucket because the bucket name is visible in the URL that points to the objects that you're going to put in your bucket.
4. For Region, choose respective *region* where you want the bucket to reside.
  5. Choose *Create*.

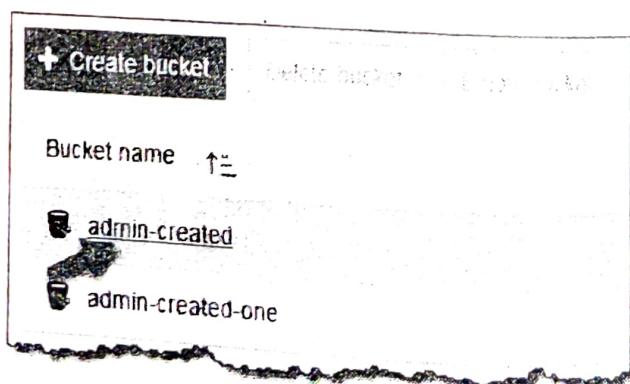


#### B) Adding a object to Bucket

Now that you've created a bucket, you're ready to add an object to it. An object can be any kind of file: a text file, a photo, a video, and so on.

Following are the steps to add an object in bucket :

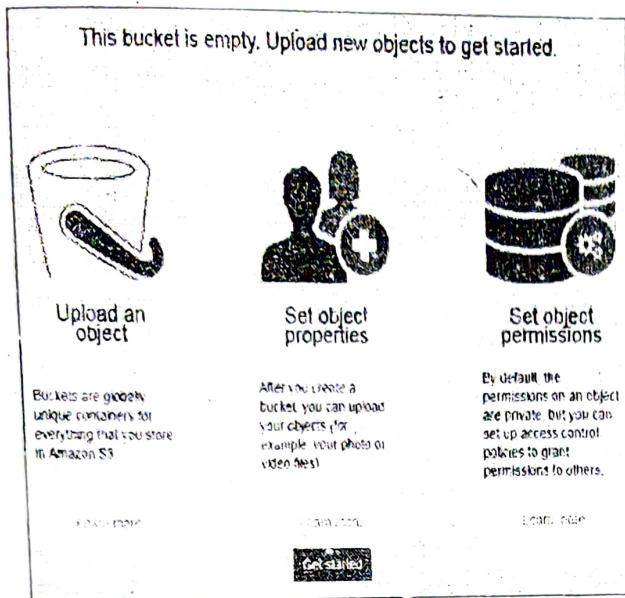
1. In the **Bucket name** list, choose the name of the bucket that you want to upload your object to.



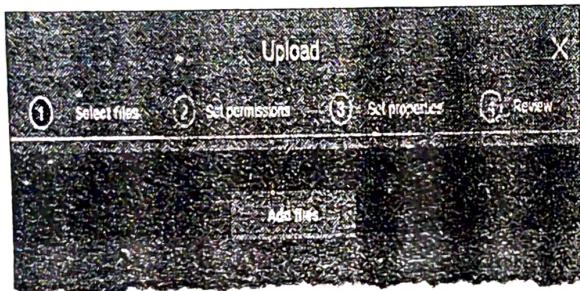
2. Choose Upload.



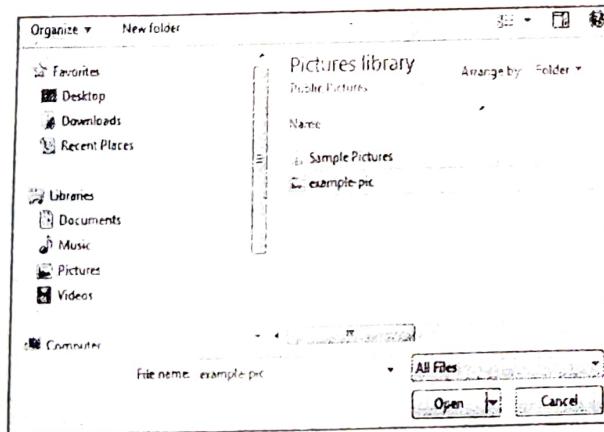
(a) Or you can choose Get started.



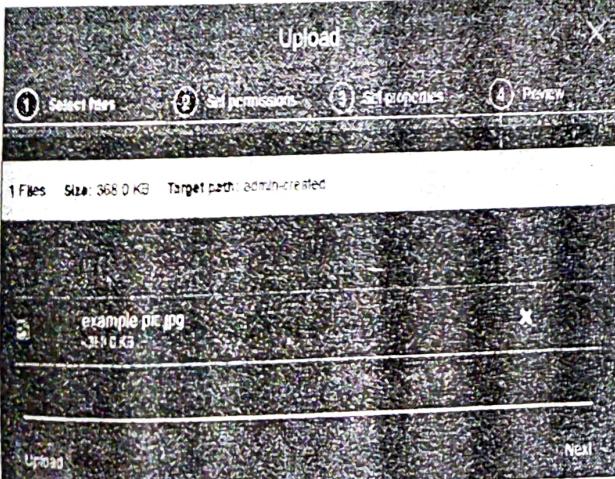
3. In the Upload dialog box, choose Add files to choose the file to upload.



4. Choose a file to upload, and then choose Open.



5. Choose Upload.

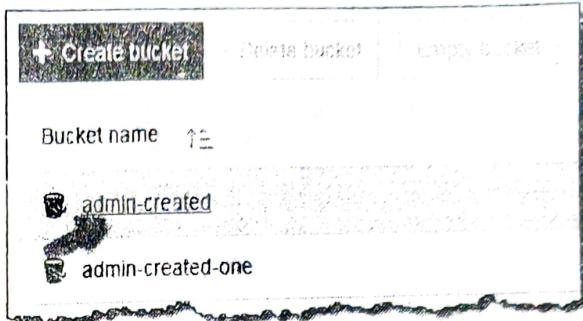


(c) View an Object

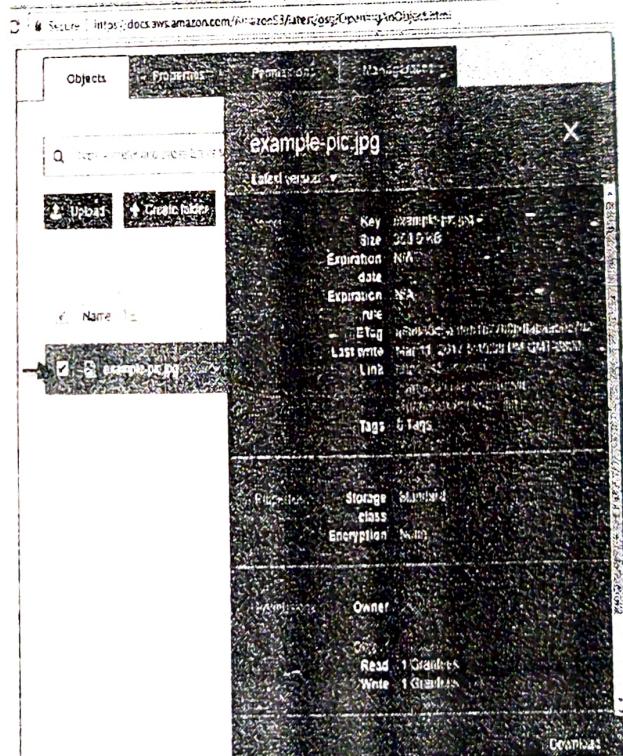
Now that you've added an object to a bucket, you can view information about your object and download the object to your local computer.

Following are the steps to view an object :

1. In the Bucket name list, choose the name of the bucket that you created.



2. In the Name list, select the check box next to the object that you uploaded, and then choose Download on the object overview panel.

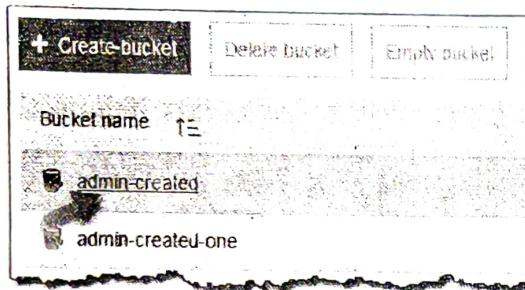


(d) Move an Object

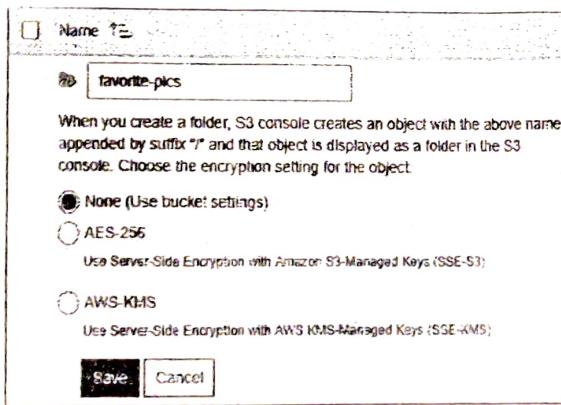
So far you've added an object to a bucket and downloaded the object. Now we create a folder and move the object into the folder by copying and pasting the object.

Following are the steps to move an object :

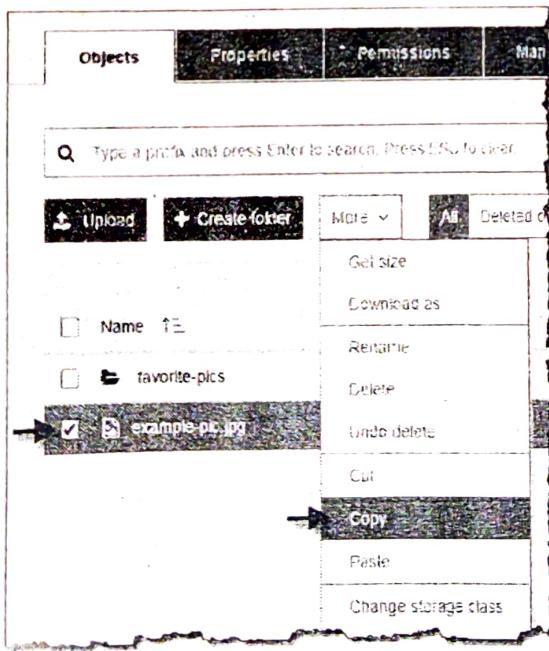
- In the **Bucket name** list, choose the name of the bucket that you created.



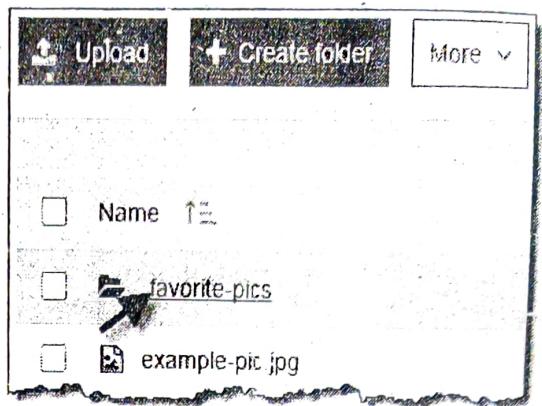
- Choose **Create Folder**, type favorite-pics for the folder name, choose **None** for the encryption setting for the folder object and then choose **Save**.



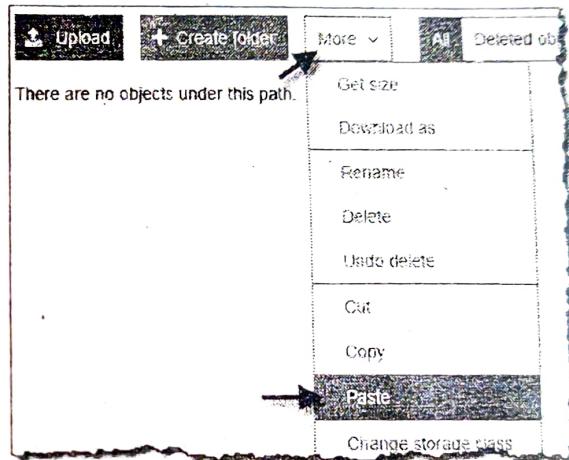
- In the Name list, select the check box next to the object that you want to copy, choose More, and then choose Copy.



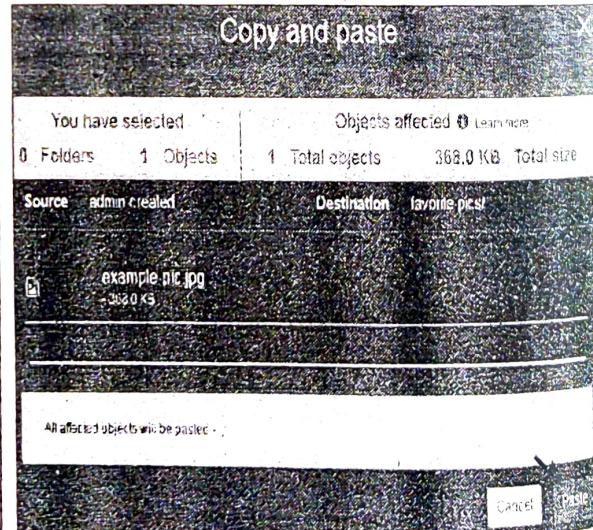
- In the Name list, choose the name of the folder favorite-pics.



- Choose **More**, and then choose **Paste**.



- Choose **Paste**.



#### (e) Delete an Object and Bucket

If you no longer need to store the object that you uploaded and made a copy, you should delete the objects to prevent further charges. You can delete the objects individually. Or you can empty a bucket, which deletes all the objects in the bucket without

deleting the bucket. You can also delete a bucket and all the objects contained in the bucket.

However, if you want to continue to use the same bucket name, don't delete the bucket. It is recommended that you empty the bucket and keep it. After a bucket is deleted, the name becomes available to reuse, but the name might not be available to reuse for various reasons. For example, it might take some time before the name can be reused and some other account could create a bucket with that name.

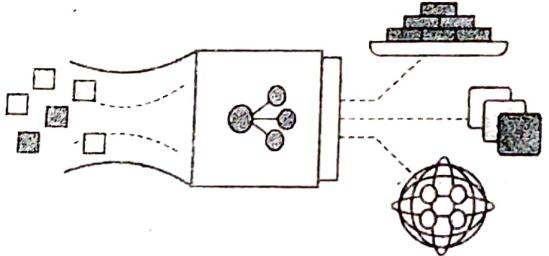
#### **Q. 11 Explain Elastic Load Balancer.**

**Ans. :**

Elastic Load Balancing distributes incoming application or network traffic across multiple targets, such as Amazon EC2 instances, containers, and IP addresses, in multiple Availability Zones. Elastic Load Balancing scales the load balancer as traffic to the application changes over time, and can scale to the vast majority of workloads automatically. Elastic Load Balancing offers three types of load balancers that all feature the high availability, automatic scaling, and robust security necessary to make the applications fault tolerant.

#### **Types of Elastic Load Balancer**

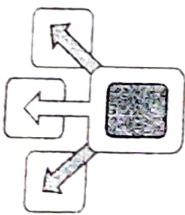
##### **1. Application Load Balancer**



**Fig. 4.4 : Application Load Balancer**

Application Load Balancer is best suited for load balancing of HTTP and HTTPS traffic and provides advanced request routing targeted at the delivery of modern application architectures, including micro-services and containers. Operating at the individual request level, Application Load Balancer routes traffic to targets within Amazon Virtual Private Cloud (Amazon VPC) based on the content of the request.

##### **2. Network Load Balancer**

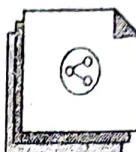


**Fig. 4.5 : Network Load Balancer**

Network Load Balancer is best suited for load balancing of TCP traffic where extreme performance is required. Operating at

the connection level, Network Load Balancer routes traffic to targets within Amazon Virtual Private Cloud (Amazon VPC) and is capable of handling millions of requests per second while maintaining ultra-low latencies. Network Load Balancer is also optimized to handle sudden and volatile traffic patterns.

#### **3. Classic Load Balancer**



**Fig. 4.6 : Classic Load Balancer**

Classic Load Balancer provides basic load balancing across multiple Amazon EC2 instances and operates at both the request level and connection level. Classic Load Balancer is intended for applications that were built within the EC2-Classic network.

#### **Q. 12 What are the advantages of Elastic Load Balancer ?**

**Ans. :**

#### **Advantages of Elastic Load Balance**

##### **1. Highly Available**

Elastic Load Balancing automatically distributes incoming traffic across multiple targets – Amazon EC2 instances, containers, and IP addresses – in multiple Availability Zones and ensures only healthy targets receive traffic. Elastic Load Balancing can also load balance across a Region, routing traffic to healthy targets in different Availability Zones.

##### **2. Secure**

Elastic Load Balancing works with Amazon Virtual Private Cloud (VPC) to provide robust security features, including integrated certificate management, user-authentication, and SSL decryption. Together, they give the flexibility to centrally manage SSL settings and offload CPU intensive workloads from user's applications.

##### **3. Elastic**

Elastic Load Balancing is capable of handling rapid changes in network traffic patterns. Additionally, deep integration with Auto Scaling ensures sufficient application capacity to meet varying levels of application load without requiring manual intervention.

##### **4. Flexible**

Elastic Load Balancing also allows using IP addresses to route requests to application targets. This offers flexibility in how user virtualizes application targets, allowing hosting more applications on the same instance. This also enables these applications to have individual security groups and use

the same network port to further simplify inter-application communication in micro-services based architecture.

### 5. Robust Monitoring and Auditing

Elastic Load Balancing allows monitoring the applications and their performance in real time with Amazon CloudWatch metrics, logging, and request tracing. This improves visibility in the behaviour of user's applications, uncovering issues and identifying performance bottlenecks in the application stack at the granularity of an individual request.

### 6. Hybrid Load Balancing

Elastic Load Balancing offers ability to load balance across AWS and on-premises resources using the same load balancer. This makes it easy for the user to migrate, burst, or failover on-premises applications to the cloud.

#### Q. 13 Write steps to create and verify Elastic Load Balancer.

**Ans. :**

For creating and verifying Elastic Load Balancer, following steps must be performed :

##### Step 1 : Select a Load Balancer Type

Elastic Load Balancing supports three types of load balancers : Application Load Balancers, Network Load Balancers, and Classic Load Balancers.

##### To create a Classic Load Balancer

1. Open the Amazon EC2 console at <https://console.aws.amazon.com/ec2/>.
2. On the navigation bar, choose a region for load balancer. Be sure to select the same region that is selected for EC2 instances.
3. On the navigation pane, under *LOAD BALANCING*, choose *Load Balancers*.
4. Choose Create Load Balancer.
5. For Classic Load Balancer, choose *Create*.

##### Step 2 : Define Your Load Balancer

You must provide a basic configuration for your load balancer, such as a name, a network, and a listener. A *listener* is a process that checks for connection requests. It is configured with a protocol and a port for front-end (client to load balancer) connections and a protocol and a port for back-end (load balancer to instance) connections. Here, we configure a listener that accepts HTTP requests on port 80 and sends them to instances on port 80 using HTTP.

##### To define the load balancer and listener

1. For Load Balancer type a name.
2. The name of Classic Load Balancer must be unique within set of Classic Load Balancers for the region, can have a

maximum of 32 characters, can contain only alphanumeric characters and hyphens, and must not begin or end with a hyphen.

3. To Create LB inside, select the same *network* that is selected for instances: EC2-Classic or a specific VPC.

[Default VPC] If default VPC is selected and would like to choose the subnets for load balancer, select *Enable advanced VPC configuration*.

4. Leave the default listener configuration.

Load Balancer name:	my-load-balancer		
Create LB Inside:	My Default VPC (172.31.0.0/16)		
Create an internal load balancer:	<input checked="" type="checkbox"/>		
Enable advanced VPC configuration:	<input checked="" type="checkbox"/>		
Listener Configuration:			
Load Balancer Protocol	Load Balancer Port	Instance Protocol	Instance Port
HTTP	80	HTTP	80

5. [EC2-VPC] For Available subnets, select at least one available public subnet using its add icon.

The subnet is moved under selected subnets. To improve the availability of load balancer, select more than one public subnet.

##### Available subnets

Actions	Availability Zone	Subnet ID	Subnet CIDR	Name
	us-west-2c	subnet-cb653da2	10.0.1.0/24	
	us-west-2c	subnet-c9663da0	10.0.0.0/24	

##### Selected subnets

Actions	Availability Zone	Subnet ID	Subnet CIDR	Name
	us-west-2a	subnet-e4b3493	10.0.2.0/24	
	us-west-2b	subnet-5264e837	10.0.3.0/24	

6. Choose *Next: Assign Security Groups*.

##### Step 3 : Assign Security Groups to Your Load Balancer in a VPC

If you selected a VPC as your network, you must assign your load balancer a security group that allows inbound traffic to the ports that you specified for your load balancer and the health checks for your load balancer.

##### To assign security group to your load balancer

1. On the Assign Security Groups page, select *Create a new security group*.
2. Type a name and description for your security group, or leave the default name and description. This new security group

contains a rule that allows traffic to the port that you configured your load balancer to use.

Assign a security group:	<input type="radio"/> Create a new security group <input type="radio"/> Select an existing security group		
Security group name:	my-load-balancer-group		
Description:	for my load balancer		
Type	Protocol	Port Range	Source
HTTP	TCP	80	Anywhere

3. Choose *Next : Configure Security Settings*.
4. Choose *Next : Configure Health Check* to continue to the next step.

#### Step 4 : Configure Health Checks for Your EC2 Instances

Elastic Load Balancing automatically checks the health of the EC2 instances for your load balancer. If Elastic Load Balancing finds an unhealthy instance, it stops sending traffic to the instance and reroutes traffic to healthy instances. In this step, you customize the health checks for your load balancer.

##### To configure health checks for your instances

1. On the Configure Health Check page, leave Ping Protocol set to HTTP and Ping Port set to 80.
2. For Ping Path, replace the default value with a single forward slash (""). This tells Elastic Load Balancing to send health check queries to the default home page for your web server, such as index.html.

Ping Protocol	HTTP
Ping Port	80
Ping Path	/

3. For Advanced Details, leave the default values.
4. Choose *Next: Add EC2 Instances*.

#### Step 5 : Register EC2 Instances with Your Load Balancer

Your load balancer distributes traffic between the instances that are registered to it.

##### To register EC2 instances with your load balancer

1. On the Add EC2 Instances page, select the instances to register with your load balancer.
2. Leave cross-zone load balancing and connection draining enabled.
3. Choose *Next : Add Tags*.

Alternatively, you can register instances with your load balancer later on using the following options :

Select running instances after you create the load balancer. Set up Auto Scaling to register the instances automatically when it launches them.

#### Step 6 : Tag Your Load Balancer (Optional)

You can tag your load balancer, or continue to the next step. Note that you can tag your load balancer later on; To add tags to your load balancer :

1. On the *Add Tags* page, specify a key and a value for the tag.
2. To add another tag, choose *Create Tag* and specify a key and a value for the tag.
3. After you are finished adding tags, choose *Review and Create*.

#### Step 7 : Create and Verify Your Load Balancer

Before you create the load balancer, review the settings that you selected. After creating the load balancer, you can verify that it's sending traffic to your EC2 instances.

##### To create and test your load balancer

1. On the *Review* page, choose *Create*.
2. After you are notified that your load balancer was created, choose *Close*.
3. Select your new load balancer.
4. On the Description tab, check the Status row. If it indicates that some of your instances are not in service, its probably because they are still in the registration process.
5. After at least one of your EC2 instances is in service, you can test your load balancer. Copy the string from DNS name (for example, my-load-balancer-1234567890.us-west-2.elb.amazonaws.com) and paste it into the address field of an Internet-connected web browser. If your load balancer is working, you see the default page of your server.

#### Step 8 : Delete Your Load Balancer (Optional)

As soon as your load balancer becomes available, you are billed for each hour or partial hour that you keep it running. When you no longer need a load balancer, you can delete it. As soon as the load balancer is deleted, you stop incurring charges for it. Note that deleting a load balancer does not affect the instances registered with the load balancer.

## Chapter 5 : Ubiquitous Clouds and the Internet of Things

**Q. 1 Write a short note on IBM Cloud Projects.**

**Ans. :**

Initially IBM was just a hardware/software and system company, but now it gradually changing into a computing services company. The platforms provided by IBM regarding Cloud are most probably developed with IBM server clusters and are supported by IBM WebSphere. The fundamental products of IBM; the z-series and p-series servers are upgraded with virtualization capabilities. Since 2007 cloud-related R&D projects are being launched by the IBM. IBM earlier has implemented on-demand and autonomous computing with the help of SOA (service-oriented architecture). This work is extended by the newly launched projects.

These projects are extended from IBM's previous work for on demand and autonomous computing using service-oriented architecture (SOA). A virtualized cloud system is provided by the IBM for IaaS services. The management complexity is simplified by this system by keeping a large resource pool together. This provides application flexibility and competent resource deployment with the help of dynamic server and storage.

An advanced tool is developed by the IBM known as TSAM (Tivoli Service Automation Manager) which rapidly implements design and deployment and manages service processes. To manage private clouds, one more platform is available known as WCA (WebSphere CloudBurst Appliance). For application service development, SaaS cloud is offered by IBM sLotusLive. It provides various types of services such as an online conference service, coordinated office management, and e-mail services.

**Q. 2 Write a short note on Cloud System from SGI.**

**Ans. :**

SGI provides services regarding high performance computing. Cyclone : a large-scale on-demand cloud computing service is announced by SGI in February 2010 for the HPC (high-performance computing) applications. Cyclone provides both of SaaS as well as IaaS models. Technical applications which are prepackaged are offered by the SaaS model which accommodates variety of domains. For example, CFD (Computational Fluid Dynamics) and finite element analysis can be performed by the system in airplane as well as automobile industries.

SGI's fastest Altix servers and ICE clusters can be accessed by the users with the help of the IaaS model. For customized user applications SGI experts manage and optimize these servers and clusters. Nowadays organizations have to face lot much of problems, the huge amount of data to manage, process and deliver within time. To solve these problems, most of the organizations expand their IT infrastructure. But it leads to increase in costs and less throughput than expectation due to data bottlenecks. For such problems SGI has designed a solution *Altix ICE integrated blade cluster*. This new solution increases the efficiency, provide easy scaling to fulfill the increasing demands without any compromise in quality.

**Q. 3 Write a short note on Force.com Cloud.**

**Ans. :**

Mostly for CRM applications, to provide online solutions to SaaS, Salesforce.com was established in 1999. Initially third-party cloud platforms are used by the Salesforce.com to promote its software services. The subscribers of Salesforce.com were drastically increasing day by day.

**The Force.com Cloud from Salesforce.com**

For the support of SaaS and IaaS applications, the Force.com platform is offered by Salesforce.com. The external developers can create add-on applications which can be easily integrated into the applications which are hosted by Salesforce.com. The main focus is on enterprise users in business computation applications. Salesforce.com makes the CRM easy for subscribers to apply in cloud services. For simplification of business development cycle, Force.com provides Apex, a programming language and the Visualforce.

Force.com design AppExchange to allow several users to communicate and work in coordinating manner. Application services are mostly in the fields of CRM databases, development and customization of applications. In security concern, some mechanisms are provided by Salesforce.com to protect data integrity. It also provides control mechanisms to assure administrative level security and security recording.

In June 2010 a new tool known as Chatter is provided by the company which is considered as "Facebook for the enterprise". For the purpose of releasing application innovations, packaging service is provided to users. The CRM applications of users can be customized in the Force.com platform. The Force.com platform also provides SOAP web service APIs. With devices such as BlackBerry, iPhone, and Windows mobile, subscribers can get the mobile support. The SaaS and PaaS services of Salesforce.com are compatible with more than a dozen international languages.

**Q. 4 Write a short note on Large scale private Clouds on NASA and CERN.**

**Ans. :**

In the United States and European Union, two large-scale private clouds were in the phase of construction in 2010. NASA (National Aeronautics and Space Administration) develops the US cloud Nebula to run climate models on its remote systems. It removes the necessity of supercomputers to thousands of NASA users at their local sites. Further NASA developed complex cost effective weather models around its data centers using Nebula.

European Organization for Nuclear Research, known as CERN is an international organization whose purpose is to operate



the world's largest particle physics laboratory in Geneva, Switzerland. The acronym CERN originally stood, in French, for Conseil Européen pour la Recherche Nucléaire (European Council for Nuclear Research), which was a provisional council for setting up the laboratory. In Geneva, the CERN built the EU cloud. This cloud is considered as a huge private cloud basically built for the purpose of distributing data, applications, and computing resources to millions of users throughout the world. CERN provides huge data sets and throughput using global workforce and with fixed cost.

#### **Q. 5 Write a short note Cloud Mashups.**

**Ans. :**

In the field of developing web applications, a mashup is considered as a web page or application which integrates data, presentations, or functionality from multiple sources for the purpose of creating a new service. Mashup provides the important functionality of combining virtualization and aggregation. The reason of wide spread of cloud is its ability to provide resource allocation dynamically from a pool of VM resources. It is possible to employ the resource pool for both parallel and distributed computing as per the documentations of Amazon EC2 and S3 services.

In contrast the GAE (Google App Engine) offers web-based services for storage of files, e-mailing, and messaging. The functionality of AWS and GAE is different and also they also can complement each other. All these concerns lead to an idea of mashing up different clouds. This mashing can dynamically build an inter-cloud or cloud of clouds. We are going to see the mashing up of GAE and AWS clouds to get the required features such as application agility as well as performance scalability.

In fact new startups get a cost-effective solution which do not have budget to invest in private cloud or create their own data center with hardware and software infrastructure. The widely used paradigms like MapReduce on AWS should have been supported by cloud mashup and yet be controllable with the help of GAE using the web interfaces which are easy to operate.

#### **Q. 6 Write a short note on Cloudlets.**

**Ans. :**

Cloudlets is low-cost infrastructure which enables the cloud computing mechanism with the help of mobile devices. The idea of Cloudlets concept is proposed by the researchers at Carnegie Mellon University, Microsoft, AT & T, and Lancaster University. A resource-rich portal is offered by the Cloudlet to enhance mobile devices with advanced capabilities to access distant clouds. It is expected that the portal should be trustworthy and take help of VMs to explore applications which are location-aware cloud. The internet cloud can be accessible through Cloudlet in easy and cost effective manner.

#### **Q. 7 Differentiate between cloudlets and cloud.**

**Ans. : Difference between cloudlets and cloud**

Parameter	Cloudlets	Clouds
State	Soft state only	Hard as well as soft state
Management	Mostly Self-managed. There little or near about no professional management	Professionally managed, require 24x7 operator
Environment	At business premises "Data center in a box"	Machine room with uninterrupted power conditioning and cooling
Ownership	Decentralized ownership by local business	Centralized ownership by big giants like Amazon, Yahoo, etc
Network	LAN latency/bandwidth	Internet latency/bandwidth
Sharing	Small number of users at a time	Large number of users at a time

#### **Q. 8 Write a short note on FutureGrid.**

**Ans. :**

Fig. 5.1 shows the FutureGrid. This project is funded by NSF as an experimental component of TeraGrid and designed around the concept of Grid'5000.

FutureGrid project is led by Indiana University in partnership with following universities :

- (i) Purdue University (HTC hardware),
- (ii) San Diego Supercomputer Center at the University of California,
- (iii) San Diego (hardware, INCA, monitoring),
- (iv) University of Chicago/Argonne National Labs (hardware, Nimbus),
- (v) University of Florida (hardware, VINE, education, and outreach),
- (vi) University of Southern California Information Science Institute (Pegasus workflow software to manage experiments),
- (vii) University of Tennessee, Knoxville (benchmarking),
- (viii) University of Texas at Austin/Texas Advanced Computing Center (hardware, portal),
- (ix) University of Virginia (interoperability, grid software), and
- (x) Center for Information Services and GWT-TUD from Technische Universität Dresden (VAMPIR software).

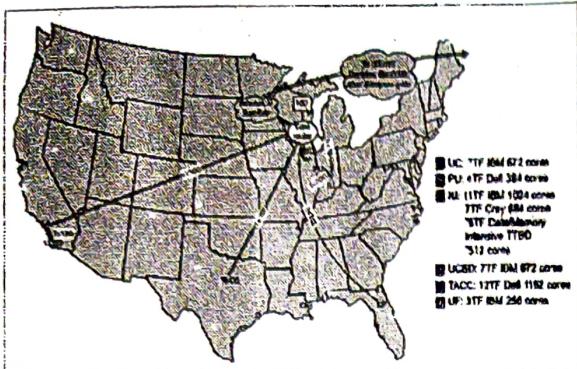


Fig. 5.1 : FutureGrid

Even FutureGrid is an infrastructure project, system development is significantly included in it to support its goals. FutureGrid is not considered as production system, but instead an environment which supports a flexible development and testing platform for the users of middleware and application concerning with issues regarding interoperability, functionality, and performance. With the help of FutureGrid, the researchers can submit experiment plan to conduct experiments which will be further executed through a sophisticated workflow engine, conserving the provenance and state regarding information essential to allow reproducibility.

In the testbed the elements involve are : group of heterogeneous computing systems which is existing in distributed manner, system to manage data which will store both metadata and

a software images library, and a devoted network which will allow isolatable as well as secure experiments. The testbed will provide support for environments which are VM-based, and also for native operating systems for the experiments which are basically focus on maximizing performance by minimizing overheads. The existing open source software packages are integrated by the project partners to establish an user friendly software environment supporting the processes such as instantiation, execution, and recording of grid as well as cloud computing experiments.

Observing the behavior and utility of approaches regarding cloud computing is an important objective of FutureGrid. The overhead of cloud mechanism can be calculated by sending request for associated experiments on both virtual as well as bare-metal systems. Seven distinct systems (clusters) are provided by FutureGrid. These systems include near about 5,000 cores of diverse optimizations. To allow security isolation FutureGrid offers a dedicated network. In cloud, as the environments are built on top of hypervisors, a high level flexibility is available. FutureGrid provides software dynamically on demand onto "bare metal." For test environment of user, there are variety of different images available, from which user has option to select. The Fig. 5.2 shows the software stack of FutureGrid project.

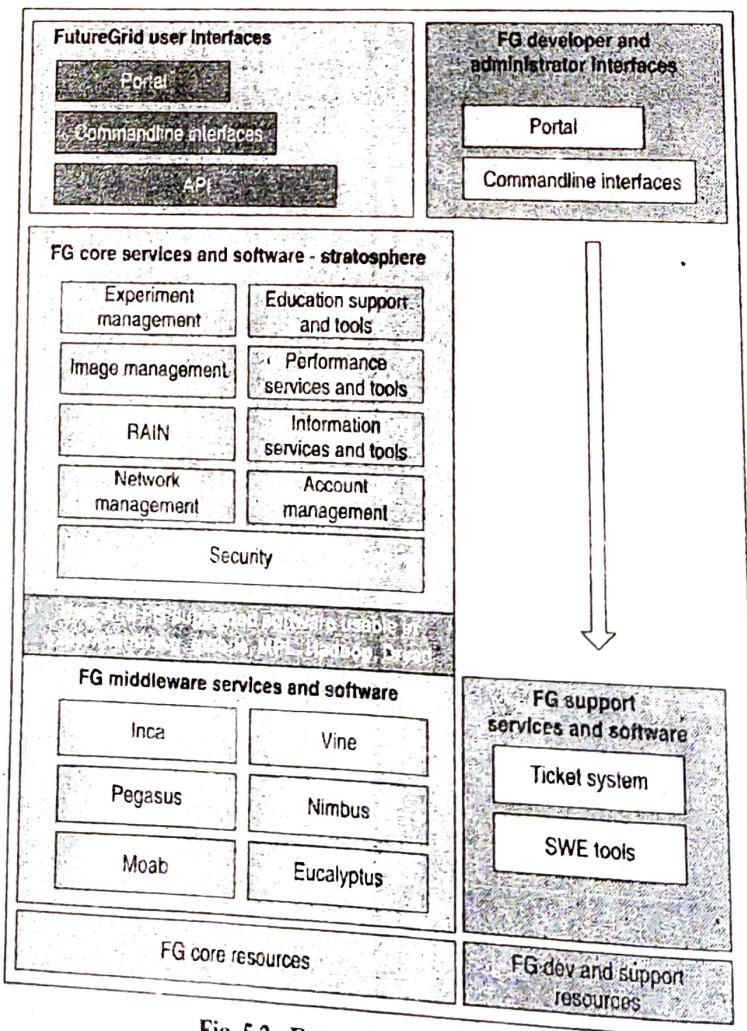


Fig. 5.2 : FutureGrid software Stack

Key components included in the FutureGrid project are :

- Portal interfaces
- The Experiment Manager, a newly added feature which helps in reproducibility of execution scenarios
- VINE intercloud networking : It is virtual clusters developed with the help of virtual networks
- Feature of generating image and repository
- RAIN (Runtime Adaptable InsertioN) Service for scheduling and deploying images
- Performance library
- Security regarding features such as isolated network and authentication and authorization.

#### **Q. 9 Explain Grid' 5000 and Magellan.**

**Ans. : Grid' 5000**

The Grid'5000 is a France based distributed system which links 9 sites in France and one from Brazil (Porto Alegre). Over fifteen hundred nodes, there are above five thousand cores in Grid'5000. It has been proved useful in various types of projects regarding computer science research containing cloud as well as grid computing and green IT involving both of the software systems and performance goals. For users Grid'5000 aims to provide a platform which will be greatly reconfigurable and manageable.

#### **Magellan**

Magellan is established at two sites of U.S. Department of Energy. These are 1) Argonne and 2) NERSC (National Energy Research Scientific Computing Center) in California. It is the considered as largest in all of the research clouds. At both of the sites, there is medium-sized cluster and storage related facilities. In NERSC system there are above seven hundred nodes and near about six thousand cores. At Argonne, there are above five hundred nodes and four thousand cores.

A wide variety of research issues are addressed by the Magellan system such as :

- For data-intensive applications, solid state storage is tested. The solid state storage also known as flash storage provides significantly high bandwidth and I/O operation rate (IOPS), and minimized latency, particularly for read-intensive applications.
- Exploring the applications' portability to cloud systems especially concerning MapReduce.
- Offering portals on the clouds so as to provide simple access to applications, databases, or automated workflows.
- Testing various types of substitute resourcing models. E.g., where availability is very much important , a virtual private cluster could assure access to specific research groups for particular time duration.
- Searching for cloud computing, the best suited scientific applications and user communities.

- Understanding the issues regarding the deployment process and support necessary to build large science clouds. Checking the cost effectiveness of science clouds. Determining the maximum use of commercial clouds.
- Understanding how current cloud software fulfils the requirements of science and whether utility will be increased if the current cloud software is extended.
- Determining support of cloud computing to data-intensive scientific applications.
- Determining the security challenges for a virtualized cloud environment.

#### **Q. 10 Explain following**

- Open Cirrus**
- Open Cloud TestBed**
- Science clouds**

**Ans. :**

#### **(i) Open Cirrus**

This is collaboration of 14 partners exploring cloud computing. There are above six hundred nodes a following four sites :

- KIT (Karlsruhe Institute of Technology)
- University of Illinois at Urbana Champaign
- HP Labs in Palo Alto, California
- BigData cluster at Intel Research in Pittsburgh

The testbed of Open Cirrus supports research into the design, provisioning, and management of services at a global, multi-data center scale. Basically to support research into all aspects of service and data-center management, the testbed of Open Cirrus is designed. Additionally, the objective of Open Cirrus is to promote a collaborative community around the testbed, providing methods for tools sharing, lessons, and best practices, and methods to benchmark and compare various approaches to service management at data-center scale.

#### **(ii) Open Cloud Testbed**

The OCC (Open Cloud Consortium) is an organization of member-driven type which provides support for the setting up the standards for cloud computing and frameworks for intercommunication in between clouds, set benchmarks for cloud computing. A testbed is also managed by the OCC for cloud computing. This testbed is known as Open Cloud Testbed. For the support to scientific research, the cloud computing infrastructure is operated by OCC, known as Open Science Data Cloud, with an emphasis on data-intensive computing.

There are various sites of OCC around the world. Some of them are as follows :

- Chicago (USA)
- California (La Jolla, California)
- John Hopkins University (USA)



### (iii) Science Clouds

Science Clouds is an open cloud federation which offers the service of compute cycles in the cloud environment for the research activities of scientific communities exploring the use of Nimbus.

The Science Clouds has sites in Europe and the United States :

- (i) Nimbus Cloud at the University of Chicago
- (ii) Stratus at the University of Florida
- (iii) Wispy at Purdue University

### Q. 11 Differentiate DISC and Supercomputers.

**Ans. : Difference between DISC and Supercomputers**

Parameter	Conventional Supercomputers	Data-Intensive Scalable Computing (DISC)
System architecture	An HPC (high-performance computing) system used to retrieve data from remote sites and brought the data into the system for execution. Data movement overhead.	An HTC (High Throughput Systems) data-center cluster used to gather data and maintain it. For fast retrieval, computation is collocated with storage. No data movement overhead.
Programming models	Programs are written at low level and are machine dependent. Usage of software tools is less. Specialists are required for optimization.	Programs are machine independent. Optimization is done using runtime system controls via load balancing, etc.
System access and use model	When resources are in ready state, main machine for batch processing. Offline visualization is used at remote site.	Communicative access with priority control and user involvement over multiple users simultaneously.
Reliability issues	Brittle systems with which latest checkpoint is recovered. Need to bring down the system for the process of maintenance.	Error detection and recovery is very flexible. Redundancy techniques are used to enter elegant degraded operations failure occurs.

### Q. 12 Write a note on Performance Metrics for HPC/HTC Systems.

**Ans. : Basic Performance Attributes**

The performance models which are applied for the evaluation of MPPs, clusters in data-center, and virtualized clouds might be extremely diverse. In the cloud computing paradigm, to manage the layered structure, the performance metrics of IaaS (Infrastructure as a Service) works as a basis to model PaaS (Platform as a Service) performance. In the same way, the attributes relevant to model PaaS performance works as a foundation for the evaluation of SaaS performance. It is expectation that all the layers of computing services should be covered by a good model.

As the model applicable to variety of cloud platforms which are under various workload distributions, it is considered as generic. In the cloud services, for assessing QoS, the influence on the top layer (Software as a Service) must be transparent to the lower layers (Platform as a Service and Infrastructure as a Service). Multiple users should be able to access cloud services simultaneously. Hence for assessing distributed system performance, multitasking is a necessary.

#### Performance metrics

##### 1. System Throughput and Efficiency

Usually the total jobs which can be performed per unit of time are measured by the system throughput of a distributed system. The throughput measure is attributed to various important aspects which impacts the overall execution time ( $T_{total}$ ) of all the jobs which are processed in a specified time window. Initially, we have to inspect all component times in  $T_{total}$ .

There is need of initiation time for each submitted job to acquire resources. In this overhead time some more needed time slots are added such as boot time related to all machine instances and the scheduling time of jobs given by user. The horizontal scalability is limited by the type of the application.

Usually there are five components to which the overhead can be attributed :

- (i) Infrastructure initiation delay,
- (ii) Resource provisioning delay
- (iii) Interjob communication delay
- (iv) OS overhead
- (v) Loading application software overhead.

The overheads regarding OS and software remain constant in the process of executing different jobs. But depending upon the problem size and system management policies, the remaining three factors may vary. To simplify the analysis, we will make a single overhead time  $kT_0$  by grouping all five time factors. Here the  $T_0$  is their sum calculated over number of jobs executing in a fixed time interval. The  $k$  is a constant and platform dependent coefficient which may differ for different system configurations.

Now consider  $T_e(n, m)$  be the efficient execution time to finish  $n$  number of independent jobs in a cloud platform, where  $m$  stands for the total machine instances in a provided system

configuration. In case of physical servers, the machine instances are cluster nodes while for virtualized cloud they are VM counts. The total time of job execution will be computed as

$$T_{\text{total}} = kT_0 + T_e(n, m).$$

Now the system throughput ( $\pi$ ) is defined by:

$$\pi = n/T_{\text{total}} = n / [kT_0 + T_e(n, m)]$$

In an ideal situation, the overhead  $kT_0$  is considered to be very small or just zero, compared with the magnitude of the execution time  $T_e$ . Therefore, the formula of measuring ideal throughput is  $n/T_e(n, m)$ .

Normalized throughput is used to define the system efficiency ( $\alpha$ ) as follows :

$$\alpha = \pi / [n/T_e(n, m)]$$

$$= T_e(n, m) / [kT_0 + T_e(n, m)]$$

In the system, the effective utilization of all provisioned resources is represented by System efficiency. Remember that the arguments  $n$  and  $m$  within expression  $T_e(n, m)$  differ depending upon the workload (total users), system size ( $m$ ), the concurrent and distributed computing model applied.

## 2. Multitasking Scalability

In multitasking number of tasks are handled by the system simultaneously. The system services should have ability to scale in both directions : horizontally regarding the machine or cluster size and vertically from applications to middleware, runtime and Operating System support, and hardware. For providing elastic resources it is very important to have ability to scale resources dynamically. Horizontal scalability refers to increase the same type of cloud resources.

Vertical scalability refers to increase in the performance by the addition of more resources along the service layers. To simplify the concept, we will focus on horizontal scalability. Yet, we assume both scale-up as well as scale-down situations. Occasionally, the efficiency of scaling down could be more than scaling up. Ability of scaling up and down on demand is considered as the potential benefit of cloud system.

Now the multitasking scalability ( $\beta$ ) can be defined as follows. Remember that  $m$  is total of machine instances and  $n$  is the number of user jobs to be processed. Usually, we have  $n \gg m$ , because the value of  $n$  could be in millions while the value of  $m$  will be in hundreds. The upper bound on scalability is approximately denoted by the ratio  $n/m$ , which is considered as the average of total tasks (user jobs) running per machine instance, when all the overhead times  $T_0$  are not considered.

The scalability is inversely proportional to the overhead time. The magnitude of  $m$  can be increased to approach  $n$  to avail complete scalability. In a real life server cluster, achieving a scalability  $\beta$  even up to 20 percent is considered as very good.

$$\begin{aligned} \beta &= (n/m)\alpha \\ &= (n/m)T_e(n, m) / [kT_0 + T_e(n, m)]d \end{aligned}$$

## 3. System Availability

The percentage of time the system is up and running smoothly is referred as System availability ( $\gamma$ ). This percentage reflects the impacts of downtime when sudden failures occur and maintenance which is well scheduled for the software upgrades. The definition of System availability can be written with the help of ratio of mean time to failure (MTTF) to the sum of MTTF and mean time to repair (MTTR).

That definition is applicable for all the systems such as clusters, MPPs, grids, and clouds. The performance level of a system is specified by the HA (High availability) controlled downtime with the help of quick recovery from failure. There is another concept known as CA (Continuous Availability). It refers to an operational level with almost no failure. Usually the providers are responsible for system availability. The availability is maintained sufficiently high up to 99.999 percent of uptime by the providers.

## 4. Security Index

There are several elements which affects the availability. These are architecture of platform, applied service models, vulnerability regarding system, and recover capacity to network attacks. The large company such as Google has hundreds of data centers and near about 50 lac servers. In the platform there are server cluster, GFS, and data centers.

The security of cloud is attributed with various elements as follows :

- (i) User confidentiality
- (ii) Data integrity
- (iii) Access control
- (iv) Firewalls
- (v) IDSe
- (vi) Defense capability against viruses or worm attacks
- (vii) Reputation systems
- (viii) Copyright protection
- (ix) Data lock-in
- (x) APIs
- (xi) Data-center security policies
- (xii) Trust negotiation
- (xiii) Security auditing services

For the security of elastic resources, the reputation system must be established to protect scattered resource sites and data centers. The auditing should be carried out periodically to maintain the resource site security index ( $\delta$ ) and user-access records.

## 5. Cost Effectiveness

For a given system, the cost effectiveness is the estimation of an efficient scale of economy. A rough estimation is shown in following cost model for clouds. Here costs by both of cloud service provider as well as the data-center owner are considered.

The data-center owner must be paid by the provider for used resources. The cost of cloud usage is not dependent upon the cloud service model applied, rather the hours used are considered.

Now here the hourly charge is denoted by  $c$  and number of service hours is denoted by  $h$ . the amount which provider will get from use is calculated as :  $\text{Cost}_{\text{p}} = hc$ . The total amount which the data center will get from provider is calculated as :  $\text{Cost}_{\text{d}} = h(d/u)$ . Here the hourly charge by the data center is denoted by  $d$  and utilization rate of the data center resources is denoted by  $u$ . As anticipated by the Berkeley Group, this rate is usually in the range of 0.6 to 0.8. For assessing the profit margin, the Service provider measures the cost-effectiveness ( $\mu$ ) of the cloud service.

$$\begin{aligned}\mu &= (\text{Cost}_{\text{p}} - \text{Cost}_{\text{d}}) / \text{Cost}_{\text{d}} \\ &= (hc - hd/u) / (hd/u) = cu/d - 1\end{aligned}$$

Considering example, using Amazon EC2, say  $c = \$3/\text{hour}$ ,  $d = \$1/\text{hour}$ , and  $u = 60\%$  . It indicates that if the data center resources are 60 % utilized, then provider has a profit margin of  $\mu = 80\%$ .

#### Q. 13 Write a note on IoT for Ubiquitous computing.

**Ans. :**

The IoT is considered as a natural extension of the Internet. The foundation of the IoT is RFID. This technology is used to search tagged objects and mobile devices with the help of browsing an IP address or using the database entry. In the nineteenth century, computers follow the commands given by user. In the twentieth century, computers start to think. In the twenty-first century, computers eager to study how to recognize with the help of sensing and understanding. The IoT system helps to effectively decrease the chances of running out of stock of a company or products

wastage, as all the associated parties knows the needed product exactly.

This demands that all objects on Earth be radio-tagged is increasing day by day. The identification and management of all tagged objects by the computers is done in same way as of humans do. The IoT mechanism definitely will improve the quality of life and make the world more clean and secure, convenient, and enjoyable.

#### Q. 14 What is Ubiquitous computing?

**Ans. :**

Ubiquitous computing is considered as a post-desktop model regarding communication between person and computer system where the processing of data is integrated within daily objects and activities. Number of pervasive devices are used simultaneously by the people for daily activities. They even unaware about the presence of the interactive devices. The concept is easy but difficult to implement.

The systems design and engineering and user interfaces are the integral parts of IoT. Even the latest models of human-computer interaction such as command-line, menu-driven, or GUI-based, are not appropriate and adequate to fulfill demands of ubiquitous computing. The natural IoT paradigm suitable for the world of ubiquitous computing has not emerged up till now. Modern devices provide support for ubiquitous computing such as mobiles, desktops, sensor networks, RFID tags, smart cards, GPS (Global Positioning System) devices, etc.

#### Q. 15 Explain architecture of IoT with suitable diagram.

**Ans. :**

The Fig. 5.2 shows the event-driven three-layer architecture of IoT.

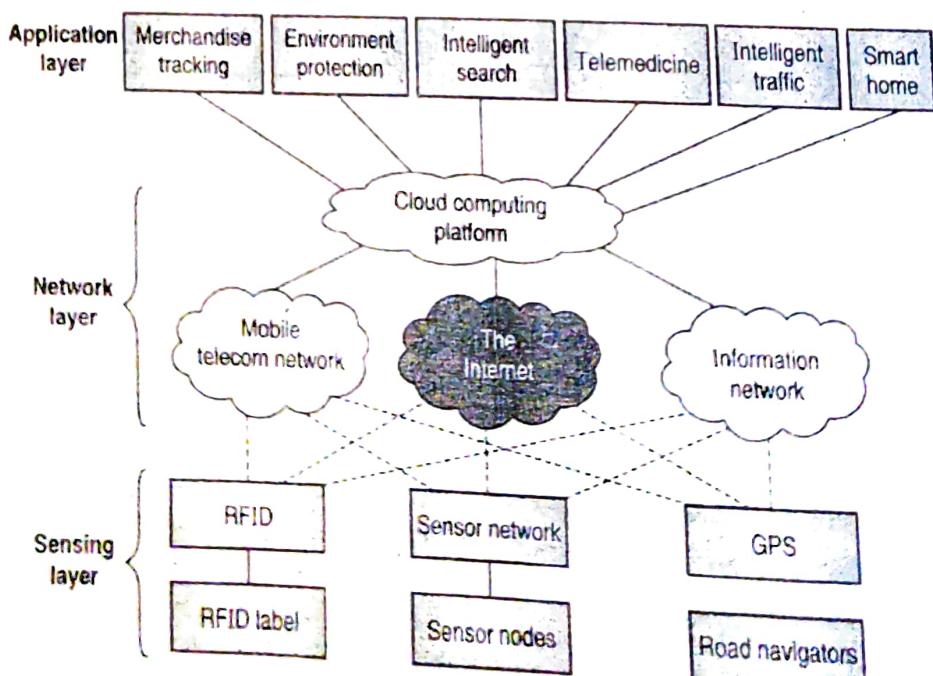


Fig. 5.2 : Architecture of IoT



The topmost layer of this architecture is formed by driven applications. Several types of sensing devices are represented by the bottom layers such as RFID tags, ZigBee or other types of sensors, and road-mapping GPS navigators. The connection of sensing devices may be local or wide in the form of RFID networks, sensor networks, and GPSes. Signals or data gathered at these sensing devices are connected to the applications with the help of the cloud computing platforms at the middle layer.

At the middle layer, over the mobile networks, the Internet backbone, and several networks of information, the signal processing clouds are established. The raw data is collected with the help of large number of sensors and filters. With the help of various types of compute and storage clouds and grids the data is processed and transformed into information and knowledge formats. A decision-making system for intelligence applications is put together with the help of sensed information.

#### Q. 16 Write a short note on RFID tag.

**Ans. :**

For monitoring or tracking of an object, RFID is applied with electronic labels or RFID tags. The application of tagging is possible to various types of objects like merchandise, tools, cell-phones, PCs, animals, or human being. The objective is to use radio waves or sensing signals to identify and track the objects. It is also possible to read some tags even from the distance of tens or hundreds of meters away using wireless reader.

Usually there are minimum two major parts of most of the RFID tags : first is the integrated circuit used for the purpose of storing and processing the information, modularization and de-modularization of a RF (Radio Frequency) signal, and other specific functions. The second part is an antenna which is used to receive and transmit the radio signals.

#### Q. 17 Explain RFID Tags and Device components.

**Ans. :**

Usually there are three types of RFID tags :

**Active RFID** - These tags contains battery and transmit signals autonomously

**Passive RFID tags** - These tags do not have battery and need an external source to handle signal transmission

**Battery-assisted passive RFID** - These tags need an external source to wake up the battery. But their capability of forward link is significantly higher.

Depending upon the utilized radio frequency, the passive RFID tags perform operation with from LF (Low Frequency) to HF (High Frequency), UHF (Ultra High Frequency), and the microwave range.

There are three major components of RFID hardware based on functionality :

**RFID Tag** : It is a small silicon chip which is attached to a small antenna.

**Reader Antenna** : This antenna radiates the energy and further capture the signal returned by the tag. It is possible to

integrate it with a handheld reader device or connect to the reader with the help of cable.

**Reader** : This is considered as a device station which can communicate with the tags. One or more antennae are supported by the reader. Similar to electronic barcode, signals can be detected by the reader device even without a line of sight.

Multiple objects can be identified simultaneously by some of the RFID readers. Also few RFID tag-reader architectures provide support for the security regarding features like requiring a human operator to enter a challenge code before the process of decoding an ID. The different elements of RFID devices may vary such as sizes, requirements regarding power, operating frequencies, storage (rewritable and nonvolatile), and software intelligence. The operational range of RFID devices is from a few centimetres to hundreds of meters. The operational range of larger devices which have internal power source is longer compared to smaller devices without internal power source.

#### Q. 18 Explain working of RFID.

**Ans. :**

New innovations in RFID bring out some new concepts such as active, semi-active, and passive RFID tags. 2 KB of data can be stored in these tags. They contain microchip, antenna, and battery for active and semi-passive tags. Plastic, silicon, or glass is used to enclose the components of tag. Data saved in the microchip waits to be read. RFID reader's antenna provides electromagnetic energy to antenna of tag. The tag uses power from its internal battery or from electromagnetic field of user. Using this power the radio signals are returned back to the reader by the tag.

The radio signals of tag are picked up by the reader the frequencies as meaningful data are interpreted. To couple schemes in RFID tags, two methods are used : inductively coupled or capacitively coupled. Fig. 5.3 shows the operations of RFID between the RFID tags (e-labels), reading-writing devices, and backend computers. As the RFID tags are expensive and bulky, their use is limited. Mostly they are used by big enterprises, shipping firms, and service companies. However, the use of RFID tags will make our daily life or work greatly easier, much convenient, and accessible to visible or invisible objects around us.

These tags can be fit on near about all the products or objects. However, in operating at shorter ranges, the RFID is considered as more effective. The standards and regulating the use of RFID is defined by various agencies such as ISO (International Organization for Standardization), IEC (International Electrotechnical Commission), ASTM (American Society for Testing Materials) International etc. There are various industrial applications of RFID such as merchandise tracking and supply chain management. The inventory control efficiency and supply line management accuracy can be improved with the help of RFID.

To power the circuits, internal batteries are used by the active and semi-passive RFID tags. The active tag has its own battery for the purpose of broadcasting radio waves to a reader while the semi-passive tag depends on the reader for the supplement of its power for broadcasting. If the to boost range of a tag, repeater batteries are used then for reading over 30 to 100 meters, active and semi-passive tags are reserved. For example with provided radio wand, a customer can find shoes of perfect size in the shelf. Then after selection, customer can use RFID setting to pay through cell phone.

Passive RFID tags depend completely on the reader as their power source. These tags are read maximum in the range of 20 feet. Their production cost is less, that means for less expensive merchandise, they can be used. These tags are disposable. In railway car, active RFID tags can be used while passive tag can be used in a bottle of shampoo. The cost of RFID tags is also affected by the data storage.

There are three storage types:

- Read-write
- Read-only
- Write-once-read-many (WORM)

The data of read-write tag can be added or overwritten. It is not possible to add or overwrite Read-only tags. The data of WORM tags can be added once, but cannot be overwritten.

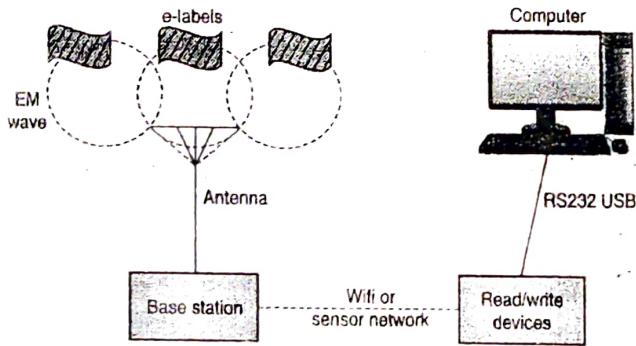


Fig. 5.3 : RFID Working

#### Q. 19 Explain Sensor Networks and ZigBee Technology.

**Ans. :**

#### Wireless Sensor Networks

A wireless sensor network is considered as a group of dedicated transducers having a communications infrastructure with basic aim of monitoring and recording conditions at different locations.

The monitoring of following parameters is usually done :

- Intensity of vibration
- Pressure
- Voltage of Power-line
- Chemical concentrations,
- Levels of Pollutant

(vi) Humidity

(vii) Direction and speed of wind

(viii) Intensity of Illumination

(ix) Intensity of Sound,

(x) Vital body functions.

There are several detection stations in sensor network which are called as sensor nodes. These nodes are small, lightweight as well as portable. In all the sensor nodes the equipments like transducer, microcomputer, transceiver, and power source are present. Based on sensed data, electrical signals are generated by the transducer. The sensor output is processed and stored by the microcomputer.

The central computer sends commands to transceiver which may be hardwired or wireless. The transceiver then transmits data to the central computer. Electric utility or a battery provides power to each sensor node. The cost and sizes of sensor nodes may vary which depends upon the energy, speed of computational, bandwidth and memory.

#### ZigBee Technology

The word "ZigBee" is related with the behavior of honeybees at the time when come back to the beehive. As per IEEE 802.15.4 standard, ZigBee is a high level communication protocol which uses the sensors of small size, low-power and are radio-based. For example, ZigBee devices are used in WHANs (wireless home area networks) in wireless light switches as well as consumer electronics managed by short-range radio. As compare to Bluetooth or WiFi, the ZigBee technology is easy to use and cost effective.

The RF (Radio Frequency) used in ZigBee leads to low data rates, extended battery life, and safe networking. As the technology is cost effective, it is possible to deploy it in wireless control and monitoring applications. A longer life with small batteries is possible due to low power consumption. Higher availability and an extended operating range is provided by the mesh networking.

Three types of ZigBee devices are available as follows :

- ZigBee coordinator (ZC) :** This is considered as the most efficient ZigBee device which serve as the coordinator or the root of a ZigBee network. Coordinator is the device which starts the network, hence exactly one coordinator is present in each network. The coordinator stores information regarding network.
- ZigBee Router (ZR) :** This device can work as an intermediary router and can pass data within end devices.
- ZigBee End Device (ZED) :** This device has only that much of functionality which is required to communicates with the parent node such as the coordinator or the router. The memory requirement of ZED is very less; hence it is cost effective for manufacture as compared to ZR or ZC.

The conformation of IEEE 802.15.4 Wireless Personal Area Network (WPAN) standard is must for ZigBee devices. The standard indicates the lower protocol layers: physical layer and the MAC(Media Access Control) at the data link layer. Fig. 5.4 shows a typical ZigBee network, direct-sequence spread spectrum coding is used by the radio whose management is done by the digital stream into the modulator.

The data rate of ZigBee is :

- (i) In the 2.4 GHz band - 250 Kb/second per channel
- (ii) In the 915 MHz band - 40 Kb/second per channel
- (iii) In the 868 MHz band - 20 Kb/second.
- (iv) The range of transmission is 10 - 75 meters.

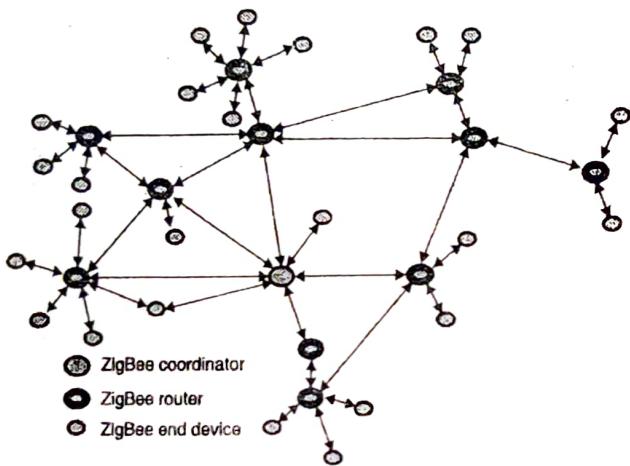


Fig. 5.4 : ZigBee Network

#### Q. 20 Write a note on GPS.

**Ans. :**

People and machines can use the LBS (Location-based service) to search things and find out their physical whereabouts. Sensors are little bit helpful but does not satisfy practical requirements of geo-location. Hence there was a need of creating a global positioning and navigation system. The U.S. Air Force develop GPS in 1973. The European Union, Russia, and China also develop the similar systems. Other signals and cellular towers are used to enhance the GPS. Generally, the GPS is augmented by other signals and cellular towers. Timing signals are broadcasted by the fixed or orbiting satellite transmitters and response is given by the receiving device to trace the location of moving objects.

Relative locations of things can be detected by radar, lidar, and sonar with the help of electromagnetic, optical, and acoustic signals. It is also possible to locate some objects with help from satellites, by attaching active GPS devices to reveal their position by radio, light, and/or sound. For various purposes such as providing reliable positioning, navigation, and timing services, a degraded GPS has been made available for general public since 1994. For anybody having a GPS receiver, the system will present perfect location as well as time information for infinite users in any weather condition, anytime, anywhere in the world.

#### Q. 21 Write a note on Smart Buildings.

**Ans. :**

The lives can be enhanced with the help of IoT in the aspects of comfort, convenience, and security. It can also help to cut the costs of energy and environmental impacts. The IoT can be used in the construction of smart buildings in various types of areas such as residential, commercial, industrial as well as government.

It involves the smart system of alarm, controlling access, management of indoor climate, elevators, and so forth. It is possible to integrate IT into electricity, various appliances and building furniture. A smart building can be anything such as a shopping mall or a house, a hotel or office tower. In Smart buildings, there is requirement of tracking and regulating the heat, air conditioning, lighting, as well as some environmental changes. Some other aspects can also be kept under the roof of smart building such as security system, fire and elevator operations. Smart building technologies concentrates on setting more comprehensive monitoring and sensing "awareness" to buildings.

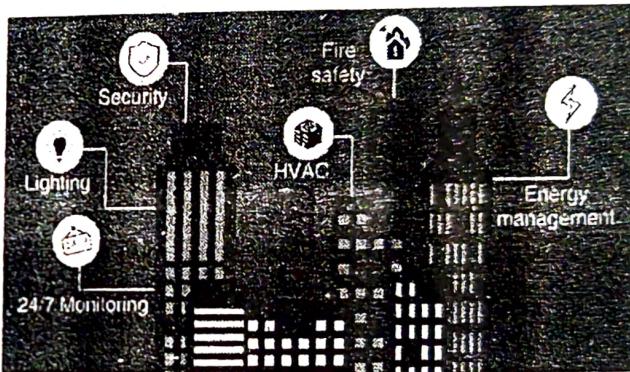


Fig. 5.5 : Smart Buildings

#### Q. 22 Write a note on Smart Power Grid.

**Ans. :**

Nowadays, in India number of power companies are in the way to upgrade their power management and distribution systems. Several sensors which are located at individual homes can gather data which is sent through a network to main stations which can apply critical power management and return control signals to the grid so as to save energy. The concept of smart grid is possible by the application of different elements such as sensing, measurement, and control devices to various processes like electricity production, transmission, distribution, and consumption.

#### Example : A Smart Power Grid Supporting the Internet of Things

A smart grid involves the intelligent monitoring system which can keep the track of the entire electricity which is flowing in the system. Smart meters, a digital upgrading of existing utility meters, tracking of the energy usage in real time so that the information about the total utilized energy will be available to both the consumer and the utility company. The charges will be applicable depending upon timing. That means in peak hours the charges will be more.

Fig. 5.6 shows IoT support of a smart power grid. For example, when the cost of power is less, smart grid can turn on some specific home appliances like washing machines. In peak hours it will turn off some selected appliances. The users will be able to see energy usage remotely with the help of smart meter and make real-time decisions regarding the consumption of energy. The appliances such as refrigerator or air conditioning could be turned off remotely while residents are away from home.

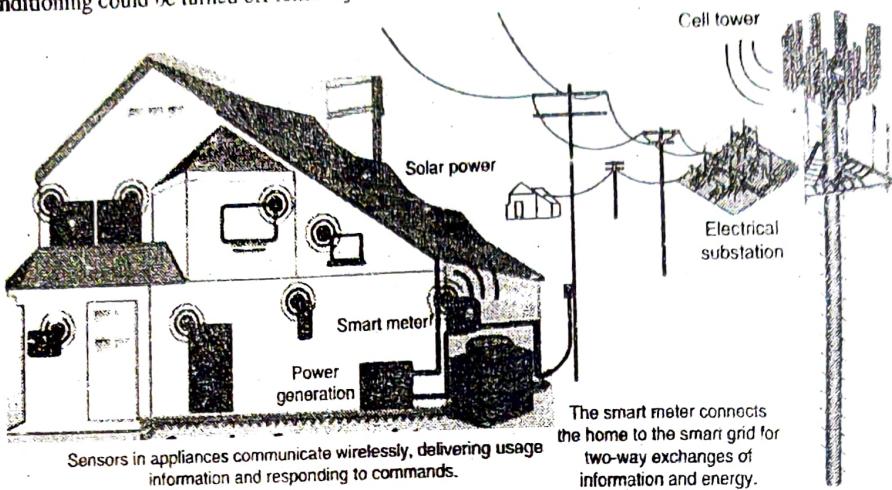


Fig. 5.6 : Smart Power grid

#### Q. 23 Write a note on supply chain management.

**Ans. :**

IoT supports Supply chain management in an effective manner. The idea is to control an entire network of associated enterprises or partners concerned with product manufacturing, delivery, and services needed by end customers. Number of times changes are required to different stakeholders such as suppliers, logistics providers, customers in a supply chain.

These changes may have significant impact on the supply chain infrastructure, starting from the foundation layers of setting the electronic interaction between the various trading partners to the effectively complicated configuration of the related processes and the setting up of workflows which are necessary to a speedy production process. A supply line merges the processes, tools, delivery options and methodologies to assist associated partners to work in a sequence to handle business with effective efficiency and delivery speed.

There are number of factors because of which the speed of the supply chain increase such as impact of global competition, fluctuations in costs, rising and falling of oil prices, short product life cycles, extended specialization, etc. Hence the cooperative organizations must work in mutual understandings. A supply chain is well-organized network of services that get materials, convert the materials to finished products, and at the end distributes those finished products to consumers.

#### Example : Supply Chain Management Aided by the Internet of Things

Supply chain management is a method followed by the enterprises to make sure that their supply chain is well-organized and cost-effective. Fig. 5.7 we can observe the supply chain

consumer products involving production and sales. In supply chain, there are number of entities such as product suppliers, distribution centers, communication links, and cloud data centers, quantity of retail stores, corporate headquarters, and bank payments.

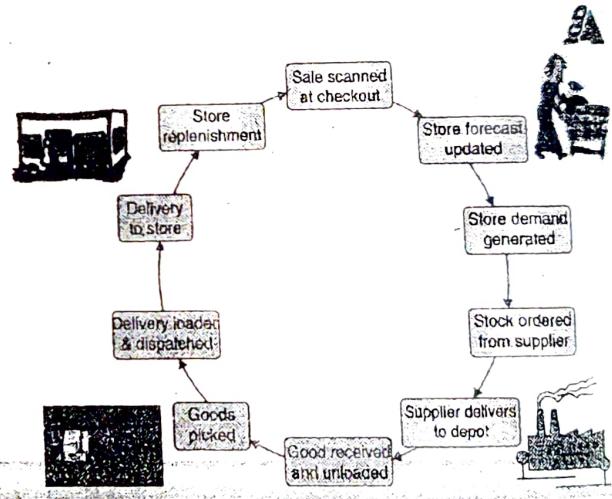


Fig. 5.7 : Supply chain management in manufacturing

All these stakeholders are connected by satellite, Internet, wired and wireless networks. It is possible to keep the Sensors, RFID tags, and GPS devices at any place in the supply chain. This promotes the online business, e-commerce, or mobile transactions.

There are five major stages in supply chain management:

- Stage 1: Planning and Coordination
- Stage 2: Material and Equipment Supplies
- Stage 3: Manufacturing and Testing
- Stage 4: Delivery of Products
- Stage 5: After-Sale Service and Returns

#### Q. 24 Explain the Cyber-Physical System.

**Ans. :**

Nowadays the computers become pervasive as well as ubiquitous. These computer devices are embedded in remote of TV, mobiles, elevators, windshield wipers, traffic signals etc. People are now handy with all these devices. In current era, by the introduction of mobiles, GPS navigators, and PCs, embedded systems are now transforming into a new category of intelligent systems, called as cyber physical systems.

A **cyber-physical system (CPS)** is an embedded system which meshes the computing process with the physical world in the form of an interactive as well as intelligent system.

Various computer and TV game systems contains the CPSes. In the fields such as automotive, aerospace, robotics, production, war training, health care and consumer goods industries, cursor-based CPSes are found. In a complete CPS, there are computers, network screens, and intelligent control of physical processes with humans interaction in case of feedback loop. The CPS handles the man-machine interactions. In Fig. 5.8, we can observe the abstraction architecture of a typical CPS. The coordination between the computations of system and physical elements us handled by the CPS.

Virtual reality systems mainly concentrate on the computational elements rather than strong connection within the computational and physical elements. To overcome this shortcoming, a CPS must be designed in the form of network of communicating elements with the actual input and output of physical form rather than just as a stand-alone device. The concepts of robotics and sensor networks are the base of this idea. It is expected that the CPS community should improve the linking between computational and physical elements. The design of CPS is mainly focused on adaptability, functionality, reliability, safety, and usability, autonomy, efficiency.

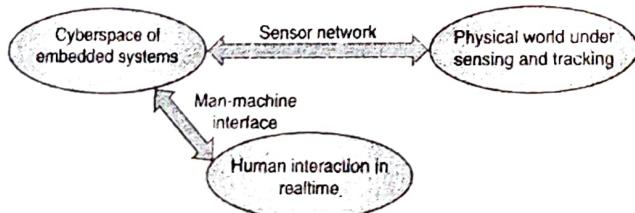


Fig. 5.8 : CPS components

The three components shown in Fig 5.8 interact with each other. To connect the three components strongly, sensor networks and man-machine interfaces are established. There is equal importance of embedded computing, sensed physical world information, and real time responses. To improve the potential of CPSes in different dimensions, new advances help effectively. One of them is human interference at the correct moment, for example collision prevention in driving a vehicle.

In robotic surgery, doctors require precision, nano-level manufacturing is necessary to workers. It is important to design CPSes to work on place of humans in dangerous or inaccessible environments like battlefields, fire, and earthquake. There is need

of close coordination in air traffic control and battlefield and improved efficiency of human capabilities in the fields such as health care monitoring and life-saving operations. Users demand close coordination in air traffic control and war fighting and enhanced efficiency and augmentation of human capabilities in health care monitoring and life-saving operations. Some other types of CPSes involve autonomous automotive systems, automatic pilot avionics process control systems, distributed robotics, and medical monitoring.

#### Q. 25 Explain benefits of social network.

**Ans. : Benefits of social network**

1. **High return visit rate**

Users will return to the social network community regularly. This will gives the great opportunity for great page impressions and a large advertising inventory.

2. **User loyalty**

Users connect to their friends and will not leave them easily. They do not move over to a competing social network. As an alternative, they show high customer loyalty.

3. **Virtual growth**

Members request their friends to the social network community. This is effective marketing at a low cost, and the OSN raise by itself.

4. **Business model**

With a social network community you can earn revenues via the subscriptions to premium content in addition to advertising revenues.

#### Q. 26 Explain graph properties of social network.

**Ans. :**

Social networks play a significant role in solving the problems, running an organization, and the degree to which individuals succeed in accomplishing their objectives. A social network is basically a map of all of the related ties among all the actor nodes. The Fig. 5.9 shows an example of social network graph. The black dot represents nodes and the lines represent the link between the nodes under specified relationships.

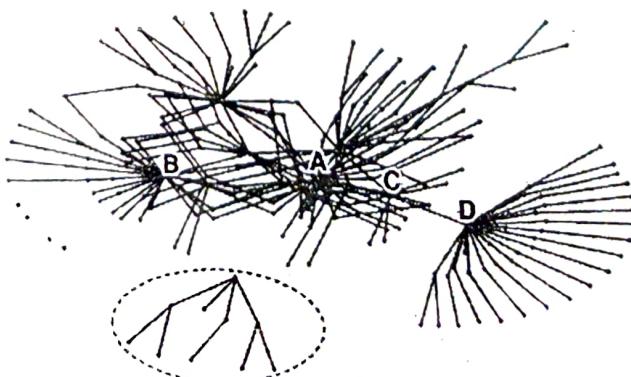


Fig. 5.9 : A graph representation of an example social network

**Q. 27 Explain closeness and cohesion.****Ans. :**

Closeness is defined as degree of an individual which is close to all other remaining individuals in a network. It has the ability to access information via the network members. Cohesion is defined as the degree to which actors are linked directly to one another through cohesive bonds. Groups are known as "cliques" if each individual is directly coupled to every other individual.

**Q. 28 Explain social circles or clusters?****Ans. :**

It can refer to some structured groups. The social circles or clusters are loosely coupled if there is less stringency of direct contact or as structurally cohesive blocks and if there is strong stringency of direct contact or as structurally cohesive blocks then they are tightly coupled.

**Q. 29 Explain functionality of Facebook.****Ans. :**

Facebook requires improvement in terms of privacy. Facebook provides very strict access control, due to which it is more secure to use than Twitter. On the other hand, one must be aware that all personal data is stored on the provider's servers, consisting information related to education and career and also the social links and personal messages. Users are wishing to share more information because they like to communicate with their real-life friends.

Facebook makes use of a network with authentication to authenticate users. A network can be considered as subset of the users on Facebook with some authentications. For example, to connect a college network, one may have to register with a college e-mail address, or follow some other authentication method approved by the network administrator. The friends of this user could recognize the user by his profile in the network.

**Table 5.1 : Functionality of Facebook Features**

Sr. No.	Function	Implementation
1.	Profile page	Combined profile: profile picture, bio information, friends list, user's activity log, public message board, other components selectively displayed
2.	Social graph traversal	Access through user's friend list on profile pages, with access control

Sr. No.	Function	Implementation
3.	Communication tools	Internal e-mail-like message; send and receive private messages between friends; instant messaging: accessed on the web page, or via third-party client; public message board: "Wall," with access control; update status: a short message, like micro-blogging, with access control
4.	Share information	Photo album: built-in, with access control; links: post links to outside URL, will come out on the activity log; videos: embedded outside videos on profile page
5.	Access control	Every item on the profile page can be set to one of four access control levels: only me, only friends, friends of friends, or everyone
6.	API applications	Games, calendars, mobile clients

**Q. 30 Explain Twitter.****Ans. :**

In 2006, the Twitter has been launched by Jack Dorsey and Evan Williams. It is a microblogging service. The message which is sent through the Twitter contains only 140 characters not more than that. Twitter just doubled the character limit for tweets to 280. This idea of Twitter is invented from Short Message Service (SMS) technology over mobile devices such as smartphones.

A user makes use of SMS to send a message to Twitter, and the message will be forwarded to lots of other Twitter users that are related to the sender. Twitter was built to share thoughts or ideas instantaneously. But increasingly major TV networks and newspapers also started to use Twitter to give updated news to the public. Twitter has become a media platform in which ordinary users can be in the front line of breaking news.

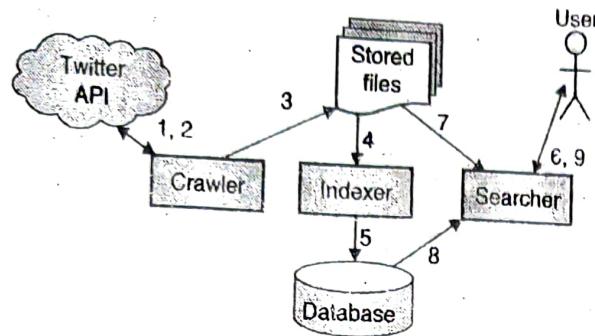
**Q. 31 Explain architecture of Twitter.****Ans. :**

The Twitter system contains three components as follows:

1. Crawler
2. Indexer
3. Searcher

Each of these components varies in the function that it performs. They are used collectively with the help of balanced coordination to form a complete system. The Fig. 5.10 shows the Twitter access architecture in 8 steps. Twitter.com started as a status sharing site that people accessed with the help of their

mobile phones. As of late 2017, Twitter has 330 million monthly active users.



1. The crawler sends a request to the Twitter API.
2. The Twitter API responds with the Twitter data.
3. The crawler stores the data into files.
4. The indexer reads the stored files as input.
5. The indexer stores the data in a database as output.
6. Users send a search query to the searcher.
7. The searcher reads the data from the database and calculates a rank.
8. The searcher reads associated raw tweets.

**Fig. 5.10 : Twitter architecture and access control sequence**

#### Q. 32 Explain functionality of Twitter.

Ans. :

#### Functions of Twitter and their implementations

Sr. No.	Function	Implementation
1.	Profile page	Simple profile : profile picture, bio information, followers list, following list, user's message timeline.
2.	Social graph traversal	Access through the following and follower's lists on users' profile pages, without access control.
3.	Communication Tools	Internal e-mail : send private direct messages to followers; instant messaging: not really instant, can be implemented by direct messages; public message board: message timeline, with access control.
4.	Share information	Photo album : third party, such as Twitpic, but no merge access control; links: post links to outside URL, will display as a message on timeline.

Sr. No.	Function	Implementation
5.	Access control	Message timeline on profile page can be set access control in two levels: public or private.
6.	API applications	PC and mobile clients, automatic news feed, photo sharing.

The Twitter functions, two of which are access control and API applications. The key to attractiveness of the Twitter is its simplicity and openness. The core service of Twitter is sending messages to a group of people. All other uses are added by third-party developers. Twitter offers open API, so third-party developers can make powerful clients or special applications for Twitter.

Twitter cannot be considered as a fully trusted network. Consider two types of accounts: public and protected accounts. Anyone can see updates of public accounts. For protected accounts, only the users that have been permitted can see the updates. To avoid identity confusion, Twitter selects accounts from celebrities to offer verification. The verified notice is posted on verified accounts. Twitter offers two new features, Twitter Lists and Location, to support this verification process.

#### Q. 33 Explain applications of Twitter.

Ans. :

#### Applications of twitter

1. **Fast News Release :** Twitter is becoming the fastest way to broadcast breaking news. Huge user collaboration has given Twitter a clear edge over most news centres. Some news centres have set up Twitter accounts to give confidence to users to spread breaking news. For example, CNN maintains several official Twitter accounts with millions of followers. During the periods of election in different countries, Twitter attracted more concentration than news centres. Information on Twitter may consist of photos, links, and videos that might receive much faster than arrive through traditional news media.
2. **Alert Systems :** Twitter offers a system that can connect residents of a city virtually w/ no cost. It also enhance the capabilities of an alert system by inputting more user-generated data. Some cities have already selected Twitter to alert their residents.

## Chapter 6 : Future of Cloud Computing

### Q. 1 Write a note on Location Aware Appliances.

Ans. :

Location-aware applications are used to provide online content to users specifically depending upon their geographical locations. Nowadays there are various advanced technologies uses cellular phone infrastructure, wireless access points or GPS to determine the physical location of users (through users' electronic gadgets like cell-phones or laptops). Then user has option to select whether he/she wants to share the information of not.

After getting the location information, various resources are provided to user by the location-aware applications. These resources include the information about restaurants, movie theatres, shops, traffic status and number of such things in specific area. Location-aware applications provide an incredibly concentrated marketing potential for retailers. They also offer enhanced social connectivity and better environmental awareness, giving users with a location-specific filter for online data.

To capture the location there are various tools available such as browser plug-ins installed in devices like mobiles or other Web-enabled devices. Also to detect the physical location of the user, Mobile phone towers, wireless access points, GPS satellites or a combination of these can be used. When there is concern of access points and cell towers, the location information is gathered depending upon the connectivity to the independent connection point. This data is then mapped and logged into databases where constant updation is done.

When a user having compatible device (mobile / laptop) opts for a location-based service then the related information is send to location-aware applications, which has responsibility to provide resources according to the location where the user is right now. In another aspect, the location-aware application can also send this data to another location-aware application or social media applications. It is user's choice that which application should get the data and how much.

### Q. 2 What are the advantages of Location Aware Appliances ?

Ans. :

Advantages are provided by a location-aware application :

Provides an reasonable implementation without the need of resorting to extra hardware such as those required for GPS-centered systems. Provides location awareness in such buildings or locations where even the GPS cannot work. Builds a fully customized map by providing convenience to define user-specific locations.

### Q. 3 List the applications of Location Aware Appliances.

Ans. :

Applications of location aware appliances are as follows

- (i) Fleet management
- (ii) Tourism
- (iv) Electronic queuing
- (v) Finding the nearest area of interest
- (vi) Finding the nearest social media contact
- (vii) With the help of cloud and location-aware solutions, it is possible to track not only the packages shipped, but also cars which are stolen, lost luggage or cell phones, missing pets, and many more.

### Q. 4 What are Intelligent Fabrics and Paints?

Ans. :

Intelligent Fabrics

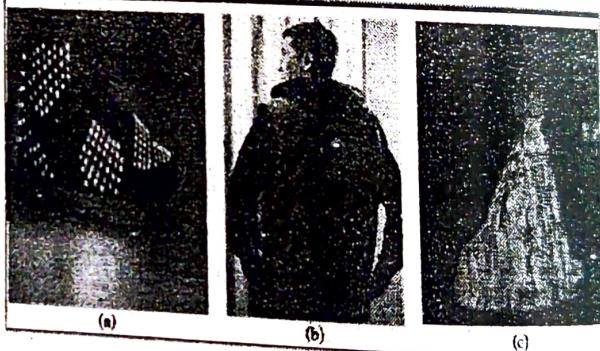


Fig. 6.1 : Intelligent Fabrics, Paints and More

Various devices can be connected to the cloud from any location and at any time. It opens the door to a wide range of cutting-edge applications. The devices like electric meters and parking meters which are read by utilities can be connected to web and reports are generated.

In the modern era, the intelligence is created in the fabrics of our clothes, bedding, and furniture. These intelligent fabrics will provide a large number of services as follows :

- (a) Automatically adjusts room temperature as per the body temperature of a person.
- (b) When a person entered or leaves any premises, the room will be notified automatically to control the lights, music, and other devices.
- (c) Monitor various types of body functions like blood pressure, blood sugar levels, stress, etc. and notify the person and adjust the environment as per the report of those functions.
  - (i) Notify others when an elderly person has fallen.

- (ii) Provide prevention against mosquitoes and other insects.

### Paints

Nowadays, new types of paints have been developed which can change their form depending upon the environmental conditions. In the countries where ice-fall is common, such paints are developed for roads which can change color to indicate the presence of ice. In the future, it is possible that the intelligent paint may notify driving conditions back to the cloud.

### Q.5 Explain future of Cloud TV.

**Ans.:**

Nowadays we observe that several companies like Hulu are trying to change the way consumers watch TV. The higher level bandwidth is available. It will definitely bypass the use of DVDs. People can watch TV shows on demand in their homes, in their cars (Hopefully only backseats), and on airplanes. The advanced technologies in combination with cloud will make it possible to convert any flat surface as a TV screen. Furthermore, users will be able to communicate with content, possibly changing the outcome of a story in real time.

### Predicting Cloud TV's Future

Cisco's Internet Business Solutions Group (IBSG) has developed a presentation on future Cloud based TV :

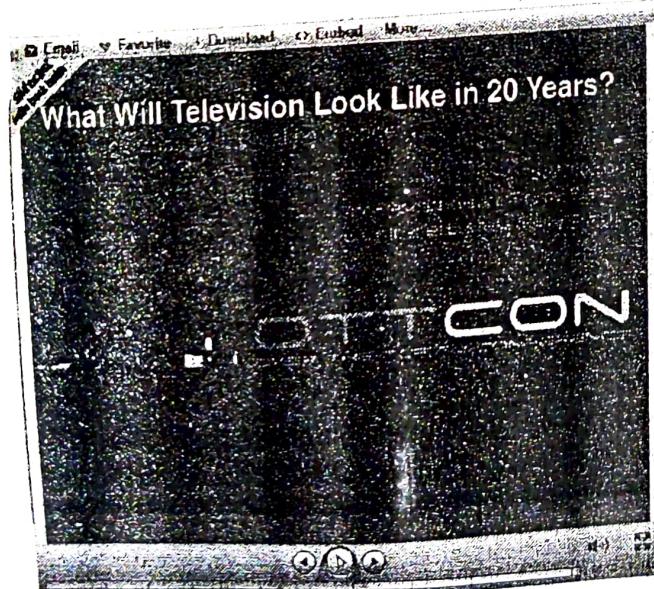


Fig. 6.2: The Future of Cloud TV

### Q.6 What is Home-Based cloud Computing?

**Ans.:**

Nowadays, wireless network capabilities are available in the most of the households which helps the people to connect to the Web and access sites and content as per their need. People can get on-demand personalized technology solutions by the innovation of smart devices, intelligent fabrics, and widely used RFID (Radio Frequency Identification) devices. Just imagine when a student enters in room, the music will play louder (or gentle), light may dim, the laptop may start particular Skype connections.

Now when parent enters the room, the music will play with lower volume, lights will turn on, and the Skype page might change to another useful application. It indicates that, family members will use the various cloud-based devices which help them to customize the environments and experiences. In such type of environments, the family members will expect the entire mechanism to work within the home. Outsiders (neighbours) should not get the cloud devices generated signals.

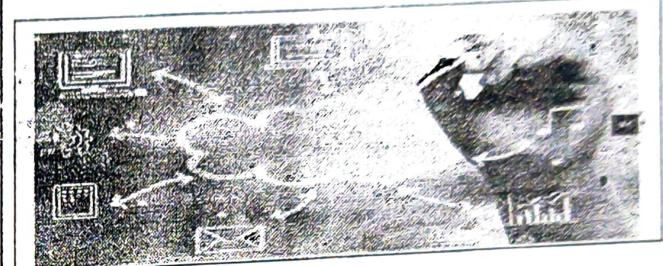


Fig. 6.2 : Home-Based Cloud Computing

This can be implemented by the mechanism which will encrypt the signals within the home. In future there will be cloud-based, in-home devices which can store family files, maintain settings of appliances, download and store movies and TV shows, and more.

### Q.7 What is Mobile Cloud and how HTML5 will drive the mobile applications?

**Ans.:**

There are two interrelated statements :

1. The cloud is going to drive mobile applications.
2. Mobile applications will drive the growth of the cloud.

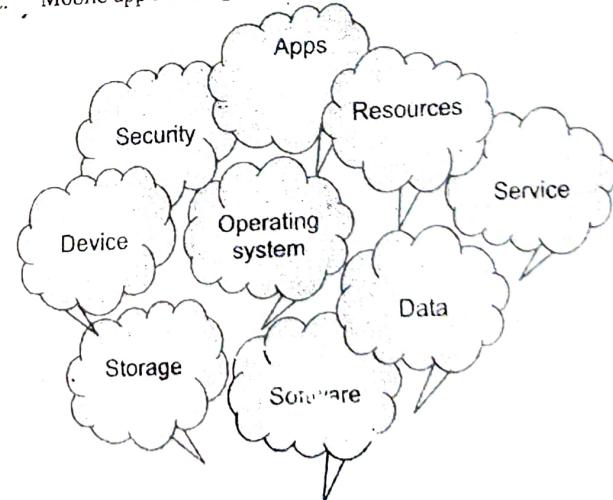


Fig. 6.3 : Cloud and Mobile

In Fig. 6.3, there will be explosive growth of Cloud-based mobile applications. Cisco is considered as a big giant in hardware and network technology that drives the cloud. Cisco's IBSG (Internet Business Solutions Group) is a strategic consulting group whose responsibility is to examine the new and emerging technologies that utilize the Web.

### How HTML5 Will Drive Mobile Applications ?

In current era, the mobile applications are considered as one of the fastest growing IT market segments. There are multiple options for mobile developers. First simple HTML-based site can be developed which can be displayed on both computer and a mobile device. Second, separate sites can be developed for computers and mobiles. Third there may be necessity to implement a computer specific page, an iPhone centric page, and pages for other devices such as Android.

The initial problem which may have to face by the developers is that most of the computer-based websites use Flash-based applications, and Flash is not compatible with mobile devices. In such situation, it is difficult for developers to build a single website that works on all devices. Or rather, this was impossible before the advent of HTML5. Across the Web, the contents which are viewed by the user in the browser eventually become in the form of HTML (Hypertext Markup Language).

HTML5 is the 20th anniversary update release of HTML, which supports all the capabilities available in the previous version of HTML through Flash-based development. Using HTML5, it is possible for developers to create multimedia content which will be suitable for all the devices. As a result, we can say that HTML5 will further drive the success of mobile applications.

### Q. 8 Write note on autonomic Cloud Engine.

**Ans. :**

Usually Clouds have highly dynamic demands for resources with highly heterogeneous and dynamic workloads. For example, the workloads which are linked with the applications may be dynamic in regards to two elements; how many tasks the application should process and the computation necessities of each task. Furthermore, it is possible that different types of applications may have different and dynamic quality of service (QoS) necessities; for example, for one application, there may be necessity of high throughput while another may not (due to limited budget).

There are various factors affecting the performance of cloud such as varying loads, failures, network conditions, etc., resulting in different "QoS" to the application. Integrating the public cloud platforms with existing grids and data centers can enable the features like on-demand scale-up, scale-down, and scale-out. Sometimes users like to try the resources in their private cloud

before going to public cloud. User also may have preferences for specific cloud or even they may want the combination of multiple clouds. However this type of integration and interoperability is near about nontrivial.

CometCloud is an autonomic cloud engine whose basic aim is to realize a virtual computational cloud with computing capability which may be resizable, that can merge local computational environments with public cloud services on-demand and facilitate various types of mechanisms to fulfill the applications' requirements. Particularly CometCloud support the policy-based autonomic cloudbridging and cloudbursting.

1. **Autonomic cloudbridging** – it integrates local computational environments such as data centers, grids, etc with public cloud services provided by cloud service providers like Amazon EC2, Eucalyptus, etc.
2. **Autonomic cloudbursting** – It enables the feature of dynamic application scale-out to handle the dynamic workloads, spikes in demands as well as some intense requirements.

### Q. 9 Explain the cometCloud Architecture with diagram.

**Ans. :**

CometCloud supports dynamic cloud or grid infrastructures which are extremely heterogeneous, integration of public and private clouds. The base of CometCloud is a peer-to-peer substrate which involves enterprise data centers, grids, and clouds. Resources are made available as per the demand in its peer-to-peer overlay which helps to provide services to applications. Conceptually, there are three layers in CometCloud ; programming layer, a service layer, and an infrastructure layer. The Squid information discovery scheme is adapted by the CometCloud which helps to map the information space with the dynamic set of peer nodes.

It results in a structure which preserves the locality and is a semantic distributed hash table. The locality of content is maintained and content-based queries can be solved with the help of flexible content descriptors in various types of forms like keywords, partial keywords, and wildcards in an effective manner. The Squid helps Comet to establish a tuple-based coordination space abstraction which can be accessed in mutual understanding by all system peers without the need of location information of tuples and host identifiers. Transient spaces are also provided by the CometCloud which enables applications to explicitly exploit context locality.

## CometCloud Layered Abstractions

In Fig. 6.4 we can observe a schematic overview of the CometCloud architecture.

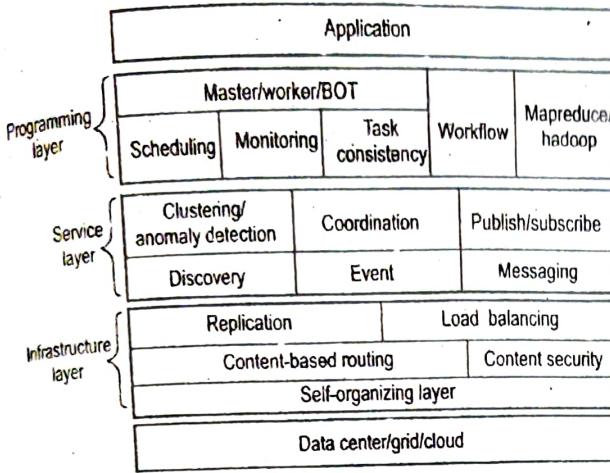


Fig. 6.4 : Cometcloud Architecture for Autonomic Cloudbursts

The Chord self-organizing overlay is used by the **infrastructure layer**, and on the top of the Chord, the Squid information discovery and content-based routing substrate are built. Partial keywords, wildcards, or ranges are used by the routing engine so as to support flexible content-based routing and complex querying. It also guarantees the placement of all peer nodes with data elements which matches a query or message. There are different roles for the nodes which provide resources in the overlay, and depending upon their credentials and capabilities, access privileges are granted to them.

Other features like replication and load balancing services are also provided by this layer. The dynamic joins, leaves of nodes and situations like node failure are handled by this layer. The replica of successor node's state is maintained by all the nodes, whenever the successor notify about the changes, they are reflected in the replica by the nodes. Every node notify to its predecessor about the changes. Whenever a new node is added, then the predecessor of new node updates its replica to state of new node, and the successor gives its state to the newly joining node.

When a new node joins, for the purpose of balancing the load, it should be redistributed among all the nodes. For the support to the mechanism of autonomics at the programming and application level several types of services are provided by the **service layer**. Linda-like tuple space coordination model is supported by the service layer. This layer also provides other features like virtual shared-space abstraction and associative access primitives.

The basic coordination primitives are listed below :

- out (ts, t) : This is a non-blocking process through which the tuple t is inserted into space ts
- in (ts, f) : This is a blocking process through which tuple matching template f is removed from the space ts and returns it.
- rd (ts, f) : This is a blocking process which a tuple t matching template f is returned from the space ts. Here the tuple is not removed from the space.

The out is used to insert a tuple into the space while both in and rd are used to read a tuple from the space. We can observe that the in removes the tuple after read while the rd only reads the tuple without removing it. The basic framework for the development as well as management of application is provided by the **programming layer**. A wide range of paradigms is supported by the programming layer such as master/worker/BOT. Tasks generation is the responsibility of masters while workers consume them. Virtual shared space or with the help of direct connection, both masters and workers set communication between them.

Application framework supports the important tasks of scheduling as well as monitoring. Despite of the fact that infrastructure layer provides the replication; there is possibility of loss of a task due to network congestion. In such situation as there is no failure, it is not possible for infrastructure level replication to handle it. This situation is then handled by the masters. For example for predefined time it can wait for result of each task and if not getting the result back, it can regenerate the lost task. In any case if duplicate results are received by the masters then it accepts the first one and ignores the further duplicate results.

### Q. 10 Write a note on Comet Space.

Ans. :

In Comet, a tuple is a simple XML string, in which tuple's tag is the first element then comes an ordered list of elements which contains the tuple's fields. Each field has name value pair. For a tuple the tag, names of fields, and values should be actual data and may have wildcards ("\*") for a template tuple. This is considered as a lightweight format which represents the information for lots of applications and also has support for rich matching relationships. Information exchange in distributed heterogeneous environments becomes possible by this format due to the cross-platform nature of XML.

When a tuple in Comet matches a template tuple exactly or approximately, then it can be retrieved. For exact matching it is necessary that there should not be any wildcard in specification of the tag and field names of the template tuple. But in highly dynamic environments, such strict matching pattern must be relaxed as applications may not know the exact structures of tuples.

Comet gives flexibility of retrieving tuple without complete structure information with the help of approximate matching where there is need of only the tag of the template tuple through a keyword or a partial keyword. Examples are shown below. Here, tuple (a) tagged "student" has fields "rollno, name, class" with values "101, Kunal, TE" can be retrieved using tuple template (b) or (c).

(A)

```
<student>
<rollno> 101 </rollno>
<name> Kunal </name>
<class> TE </class>
</student>
```

(B)

```
<student>
<rollno> 101 </rollno>
<name> Kunal </name>
<class> * </class>
</student>
```

(C)

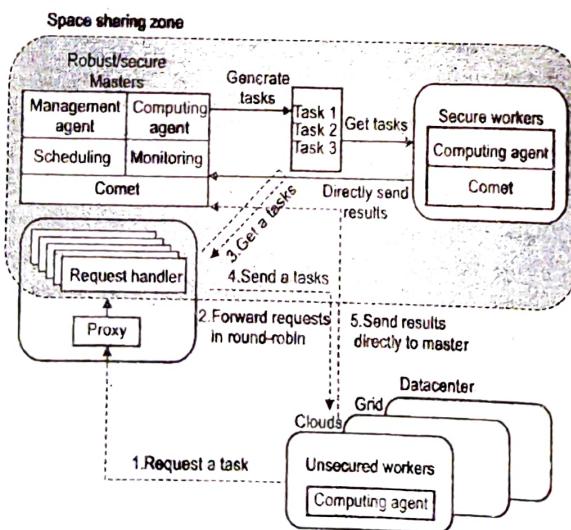
```
<student>
<rol*> 101 </rol*>
<*>
<*>
</student>
```

### Q. 11 Write a note on Cloudbursting.

**Ans. :**

Autonomic cloudbursts integrate clouds and data centres of private enterprise with public utility clouds in seamless and secure manner on demand basis. This facilitates the feature of resizable computing capacity. It helps to deploy the application components (which are currently running in local environment) dynamically onto public cloud for the purpose of handling dynamic workloads, spikes in demands as well as some other excessive necessities.

The increasing demands of application and infrastructure scales and their management and operational costs are no more feasible for the organizations. Here Autonomic cloudbursts comes in picture which enable utility clouds to supply on-demand scale-out as well as scale-in capabilities depending upon the range of metrics.



**Fig. 6.5 : Autonomic Cloudbursts using Cometcloud**

Fig. 6.5 shows the autonomic cloudbursts in CometCloud. Depending upon perceived security or trust there are three types of clouds.

The first is a greatly trusted, robust as well as secure cloud. In general it contains secure nodes within an enterprise. It is generally used to host masters and other key roles such as management, scheduling, monitoring. These nodes are also used to store states. Most of the times, there is need to maintain the privacy

and integrity of critical data. Hence the tasks which contain the critical data should be assigned only to cloud nodes which have required credentials.

The second type of cloud considered as a cloud of secure workers with nodes with such credentials. These two clouds (first and second) can be span by a privileged Comet space, and may consist of critical data as well as tasks.

In last type of cloud, there are casual workers. These workers have rights to access the space through the proxy even though they are not part of space. In all these clouds, it is possible to add or remove the node as per necessity. If workload is dynamically growing, then there may be need of scale-up to handle the increasing need of computing capability. In such situation, the secure worker is targeted to scale up by the autonomic cloudbursts. But if there is requirement of only computing capability, then unsecured workers are added.

### Q. 12 What are the key motivations of autonomic cloudbursts ?

**Ans. : Key motivations for autonomic cloudbursts**

#### 1. Load Dynamics

There is significant variation in Application workloads. It involves the count of application tasks and their computational requirements. In response to these dynamics there must be dynamic growth in computational environment. Also the deadlines should be maintained strictly.

#### 2. Accuracy of the Analytics

The necessary accuracy of risk analytics based on various highly dynamic market parameters and has a direct impact on the computational demand. The computational environment should have ability to adapt dynamically so as to fulfill the accuracy needs while still taking care of strict deadlines.

#### 3. Collaboration of Different Groups

Same application can be executed by different groups with dissimilar dataset policies. Here, the meaning of policy is the SLA of users bounded by their required condition like time frame, budgets, and economic models.

#### 4. Economics

In applications, there may be dissimilar tasks and may have different dynamic priorities based on which the resources are assigned and scheduling is done. To provide computational resources based on the priority and criticality of the application task, various types of Budgets and economic models are used.

#### 5. Failures

Because of the involvement of strict deadlines, failures can be disastrous. The computation must have capability to handle failures by avoiding effects on application quality of service as well as deadlines and accuracies.

### Q. 13 Write a note on Cloudbridging.

**Ans. :**

Autonomic cloudbridging is the mechanism by means of which we can connect CometCloud with virtual cloud which contains public cloud, data center, and grid based on the applications' dynamic requirements. In virtual cloud, every cloud has dissimilar types of resources as well as cost policies. The number of current users affects the performance of each cloud. Hence depending upon the varying environment of clouds and resource needs of applications, the elements like types of used clouds, the count of nodes in each cloud, and resource types of nodes should be decided.

Fig. 6.6 shows working of the CometCloud-based autonomic cloudbridging. Autonomic cloudbursts are managed by the scheduling agents. There may be one or multiple scheduling agents. Robust and secure master site is the location of scheduling agent. If more than one associated research groups work jointly and there is need to each group to generate tasks with their own data and has their policies to manage virtual cloud, then separate scheduling agents are made available in their master sites.

The requests for all of the tasks created by the dissimilar sites are logged in the CometCloud virtual shared space where the master nodes are spanned at all the sites. The consumption of these tasks is done by the workers, which executes on the local computational nodes at the site, a shared data center, and a grid or on a public cloud infrastructure. As per the defined policy, QoS constraints and autonomic cloudbursts of its site are managed by the scheduling agent. The space is accessible to workers with the help of appropriate credentials. The workers then access authorized tasks, perform them and finally return the results to the suitable master which is mentioned in the task itself.

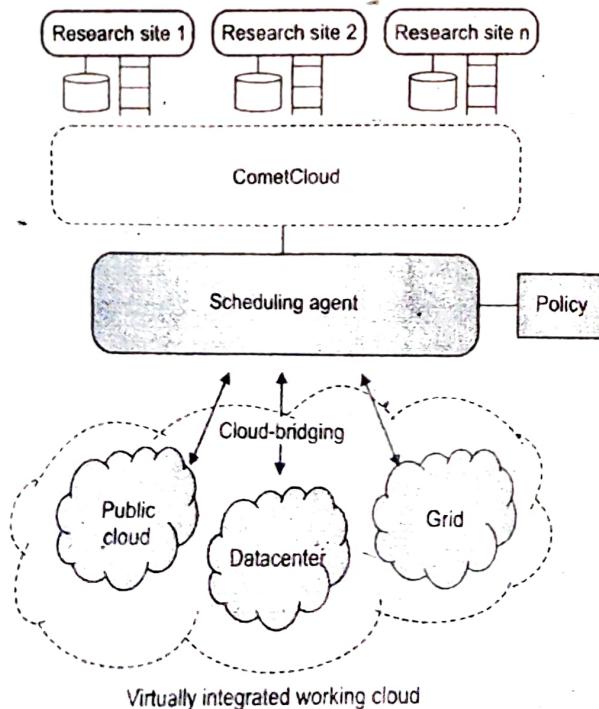


Fig. 6.6 : Autonomic Cloudbridging

Autonomic cloudbridging is managed by the scheduling agent and it assures the QoS within user policies. Changing resource provisioning represents the autonomic cloudburst so as to avoid violation of defined policy. Autonomic cloudburst is represented by changing resource provisioning not to violate defined policy.

We consider three types of policies :

- Deadline-Based** : When there is need of urgent completion of application, then by considering adequate budget, maximum possible workers are allocated for the job.
- Budget-Based** : If there is budget constraint, then while allocating the number of workers, the budget should be considered.
- Workload-Based** : The allocation or release of workers depends upon the changing workload.

### Q. 14 What is cloud Media ?

**Ans. :**

Cloud media is, a cloud in which the owner of that particular cloud can store multimedia content. Multimedia signaling protocols are used to access the media content and further can be streamed to client through their computers, tablets, cars and smartphones. The streaming protocols (For example TCP/IP, UDP, RTP, HTTP etc.) are used to share the media content in between different clouds. In the process of media content streaming, various sub processes are involved like loading data, buffering data, writing codes, mixing content, rating and rendering over the service providers. The cloud based application perform other tasks like profiling, packetizing, tokenizing of media contents and stream it to client system.

The multi-screen experience is offered by the cloud. It makes avail the media content by hiding its location details to maintain the security. On pay per use basis, cloud media provide facility to save, purchase, access and share the media content. Nowadays, music is the main focus of cloud media, but another contents are also equally important like video, image, e-books, games, mails, apps etc. Cloud media provides a platform for integrating the media contents and also works as an interface for the purpose of sharing the media contents to number of smart devices such as computers, smartphones etc.

### Q. 15 Explain different types of cloud media services.

**Ans. :**

An effective environment is provided by the cloud for the transparent access, storing and secure processing of data. The transfer of media content from local device to cloud is done by an efficient data distribution model which comparatively more secure than traditional model. Cloud effectively reduces or completely avoids the storage challenges of local devices such as computers, smartphones, tablets etc.

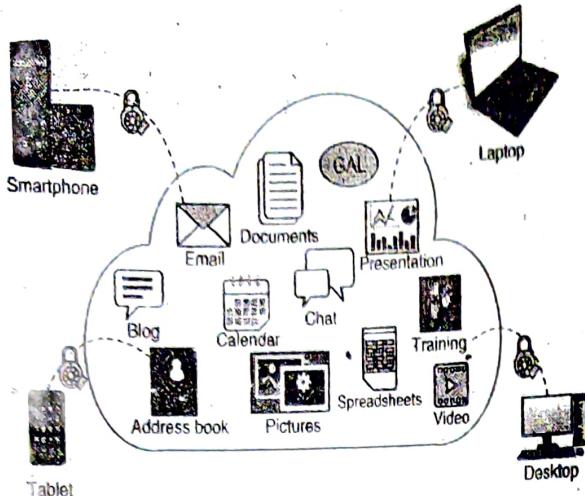


Fig. 6.7 : Cloud media services

The various cloud media services are,

### Cloud media services

#### 1. Cloud Gaming

One of the main uses of smart devices and computers is to play video games. But because of high requirement of power and computation speed, it becomes difficult to play number of 3D games on those devices. Cloud is the best solution through a dedicated service. Even on low end devices, user is able to play any kind of video game.

#### 2. Experiencing Multi-screens

In concern with multiscreens in devices, the users have high expectations. Users are able to access the media content of the cloud with the help of multiple devices because there is huge support of cloud to various types of videos, audio codecs, aspect ratio and various screen sizes on pay per use basis.

#### 3. Image Processing

Nowadays Image Processing is used in various types of fields. Investigation and medical field are the important examples. Image Processing involves techniques such as segmentation, denoising, fusion, compression etc. The software available in market for Image Processing are bigger in size which affects the computation of the local system. This problem can be solved by processing images by cloud based services.

#### Q. 16 Explain the advantages of cloud media.

**Ans. :**

#### Advantages of cloud media

##### 1. Cost

Cost effective services are provided by the Cloud media to cloud service provider via efficient multiplexing of various types of media contents such as audio, video, image by offering a common infrastructure. It reduces the cost of physically infrastructure on site and thus reduces the overall cost.

#### 2. Upgradable

The up-gradation of cloud media is the responsibility of cloud provider. Hence it is maintained properly.

#### 3. Compatibility

The cloud media is compatible with all the types of client devices like computers, smart-phones, cars tablets etc.

#### 4. Consistent

There is consistency in cloud media service like distribution of media contents to the users of other clouds with the help of streaming protocols such as TCP/IP, UDP, RTP etc.

#### 5. Green Technology

Optimized data centers are used by the cloud media computation for the operations like processing, distributing or sharing the media contents to the users. In case of traditional computation there is requirement of more energy consumption for this work.

#### 6. Storage

There are number of bases with huge capacity in cloud for the purpose of storing the content.

#### Q. 17 What are the challenges of cloud media?

**Ans. :**

#### Challenges of Cloud Media

##### 1. More confusion

Selecting suitable cloud is very confusing for the users as there are number of pay per use and free clouds are offered by the cloud service providers.

##### 2. Inside attacks

The employees of cloud Service Provider may harm or steal the media content.

##### 3. Legal and piracy difficulties

The legal standards in cloud are not very good as it is a new technology. There are number of legal troubles in storing media content in the cloud which are outside the country. Sometimes the extent to access of media contents is limited.

##### 4. Migration

Moving the media content from one cloud to another might becomes difficult if both has different infrastructures and environments.

##### 5. Challenges over standards

The different cloud service providers has different own conditions and security features which may create issues in interoperability in the near future.

#### **Q. 18 Write note on Energy Aware cloud Computing.**

**Ans. :**

Because of the adoption of Cloud Computing technology widely, it is significantly come ahead that the cloud providers has to face a big issue of maintaining infrastructure due to the related high operational costs like energy consumption. In general consideration, a data center may require energy about 100 times more than in-house organization of same size. Hence it is an important aspect to manage the power consumption of servers.

As the less power consumption of servers decreases the heat generated by these servers. This in turn decreases the requirement of cooling resources which consume huge energy. This will save more energy. To improve the energy efficiency there is a requirement of making the whole stack regarding the Cloud Computing more energy-aware. In cloud environment, there are plenty of Cloud applications which have different requirements of resources. Few of them are data-intensive while few are computing intensive.

Hence the energy consumption of applications is different. The characteristics of Cloud applications are inherited from the properties of Cloud Computing. It is generally assumed that the applications should be able to support the various characteristics such as Isolate State, Distribution, Elasticity, Automated management, loose coupling. All these characteristics reflect the patterns of Cloud Computing environments. The applications can experience different patterns of workloads based on the behavior of users, submitted tasks and the required resources.

These workloads can be categorized as static workload that has equal utilization of resources over time, periodic workload that has repeating peak utilization at interval time, once-in-a-lifetime workload that has a peak utilization once over time, unpredictable workload that has a frequent and random peak utilization over time, and continuously changing workload that has a utilisation increases or decrease continuously over time. These different types of application workloads can have different impact of energy consumption depending on usage of the resources component.

Energy-aware profiling is studied in order to understand how the energy is consumed by the infrastructure components, like CPUs, when the application is in operation. Thus, the output measurements of energy-aware profiling and energy efficiency metrics will be combined to form KPIs (Key Performance Indicators) for the running application. Also, these KPIs will be further analyzed and used to facilitate the decision-making of application developers with better energy-aware programming.

#### **Q. 19 What is Energy Efficiency in Cloud computing ?**

**Ans. :**

For the Cloud Computing stack, energy efficiency has been extensively studied in the literature and has focused on a large number of different topics, like virtualization, requirement engineering, programming models, and resource management. In

terms of virtualization, there are different approaches for allowing resource utilization, server consolidation and live migration of virtual machines, which all can offer significant energy and costs savings.

With the advancement of software-intensive systems for self-adaptive systems to meet the growing needs for autonomic computing, requirements engineering for self-adaptive software systems ensuring energy aspects has received less attention; as that can be justified with the challenges to encounter when dealing with uncertainties associated with the operating environment. Optimizing energy efficiency at different layers of Cloud stack is considered significantly important, as argued by Djemame. They therefore have proposed a Cloud architecture that addresses energy efficiency at all layers of the Cloud stack and throughout the whole Cloud application lifecycle.

In terms of programming models, there are a number of platforms used for the development and deployment of Cloud applications and services, like Hadoop, Windows Azure, Microsoft Daytona, Twister, Manjrasoft Aneka, and Google App Engine. Yet, these platforms lack consideration for energy efficiency.

#### **Q. 20 Explain Jungle Computing in detail.**

**Ans. :**

##### **Jungle Computing System**

The basic intention of Grid Computing is to present competent and transparent wall-socket computing on the top of distributed set of resources. Up till now various other types of distributed computing paradigms are come in market like peer-to-peer computing, volunteer computing and the most advanced Cloud Computing. The various types of targets of grid computing are allocated by these paradigms to provide controlled, distributed resources to the users with minimum exertions.

The new advanced distributed computing paradigms provides various types of resources for innovation in computing research, which helps to incorporate the different elements like stand-alone systems, clusters, grids and clouds, etc. As the clusters, grids as well as clouds contains multi-core processors and many-core 'addons', it becomes hard for developers to program and use them. When desire of users for speed and scalability is considered, a single high-performance computing environment is not sufficient.

Day by day the complexity of the high-performance computing environment is increasing drastically which makes it difficult to select suitable option. The emergence of various types of advanced core technologies such as GPUs and supercomputers on chip added the complexity in these environments. Multiple varied platforms and systems can be used in High-performance

computing concurrently. In technical and diverse areas, there is need of a platform having scalability as well as quick processing.

Consumers require various types of resources in addition to cluster, grid or cloud. Hence for speedy processing and high scalability, there is need of diverse collections of assets. Jungle Computing is considered as a circulated computing paradigm which is basically evolved in practice instead of developing from theory. A Jungle Computing System provides all the types of computing assets which includes clusters, clouds, grids, desktop grids, supercomputers and various types of stand-alone appliances as well as wireless devices.

Jungle computing is generally concerned with the set of diverse, allocated and highly non-uniform functionalities of computer systems which helps to achieve the peak performance. Jungle Computing Systems is the combination of heterogeneous, hierarchical as well as allocated resources.

#### **Q. 21 Explain Docker in detail.**

**Ans. :**

Docker is a container management service. The main functionality involved in Docker is represented by the keywords : develop, ship and run anywhere. The entire concept of Docker is for developers to simply develop applications, ship them into containers and further deploy them anywhere. Docker was released in March 2013 and since then, in modern world development Docker becomes the buzzword particularly in the field of Agile-based projects. Even though Docker is not tremendously complicated, it is a powerful technology.

#### **Q. 22 Explain traditional as well as Docker deployment workflow.**

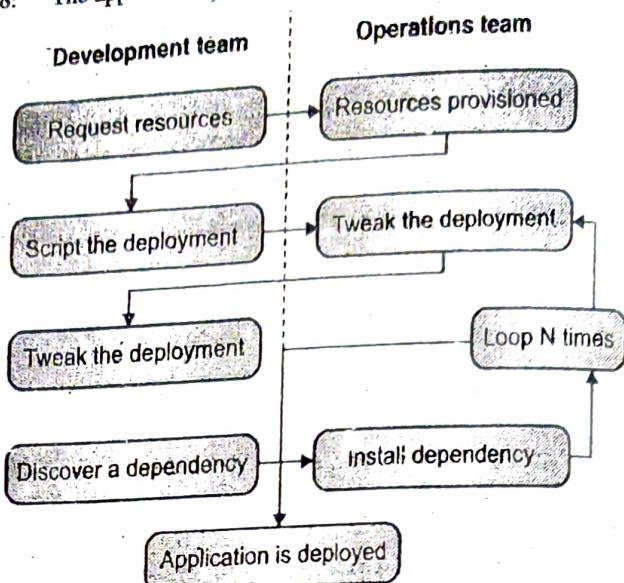
**Ans. :**

Docker can simplify both workflows and communication, and that usually begins with the basic deployment process. First we will understand the traditional deployment process of an application.

Traditionally, there are following phases in the process of getting an application to production.

1. Resources are requested by the application developers from operations engineers.
2. The operations engineers make provision of resources and hand over to developers.
3. Developers script and tool their deployment.
4. The deployment is repetitively tweaked by the operations engineers and developers.
5. Developers discover some more application dependencies.
6. The additional requirements are installed by the operations engineers.
7. The step 5 and 6 are repeated for some more times.

#### **8. The application is deployed.**



**Fig. 6.8 : A traditional deployment workflow (without Docker)**

In traditional system, the deployment process of a complex new system may require near about a week to deploy. This is time consuming, even though efforts are taken to overcome some of the barriers, it often needs a lot of work and proper communication between teams of people. This process is expensive and also technically challenging. It is also possible that it can limit the innovation which can be carried out by the development teams in the future. If the process of deployment is hard and required more time and resources from another team, then every required thing is built in the application by the developers so as to avoid suffering the new deployment penalty.

Docker provides a clear separation of responsibilities and encapsulation of dependencies, which helps to boost the productivity. It also provides the developers more fine-grained control by putting them in control of everything, down to the distribution of Operating System on which application is executed. The basic approach of Docker as a company is "batteries included but removable". It indicates that they want that their tools should consist of everything which is required by most of the people to complete a job. Also the part of those tools should be interchangeable so as to support the custom solutions.

Docker uses the image repository as a hand-off point which allows to separate the responsibility of building the application from the deployment and operation of the container. The meaning of this is; an application can be build by the developer team with environments, and move this package of application and its dependencies to production. As the look of all such packages is same, a standard tooling is built or installed by the operations engineers for the purpose of deploying and running the application.

The cycle is described in Fig. 6.9 :

1. Docker image is built by the developers and shipped to the registry.
2. Configuration details are provided by the Operations engineers to the container and they also make provision of resources.
3. Deployment is triggered by the developers.

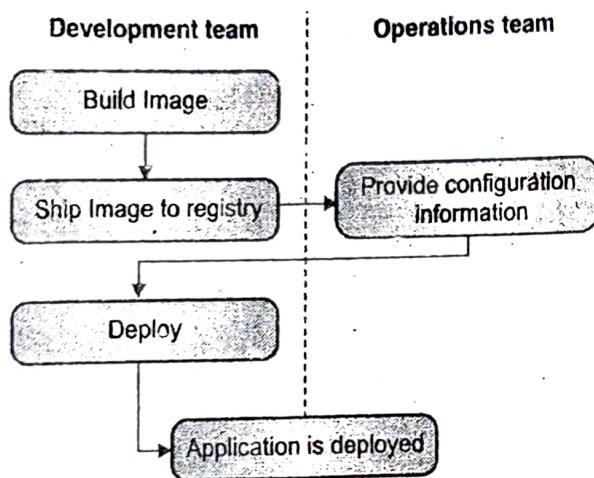


Fig. 6.9 : A Docker deployment workflow

This simplification is possible since, in Docker all of the issues are discovered regarding dependency are during the development and test cycles. Meanwhile, the application is all set for first deployment, that work is done. In general, there is no need of plenty of handovers between the development and operations teams.

#### Q. 23 Explain the client server architecture of Docker.

Ans. :

Docker is a tremendously powerful technology, but still having a simple architecture with client/server model.

#### Client/Server Model

Usually there are two main components in Docker: the client and the server/daemon. In some cases there is a third component known as registry. This registry holds the information regarding Docker images and metadata about those images. Running and managing containers is the responsibility of server while clients are used to give instructions to the server what to do. It is possible to execute the Docker daemon on any number of servers in the infrastructure and those multiple servers can be addressed by a single client.

The entire communication is handled by the Clients, but the Docker servers can directly communicate to image registries when instructed by the clients. This is responsibility of clients to instruct the servers about working and server should concentrate on hosting containerized applications.

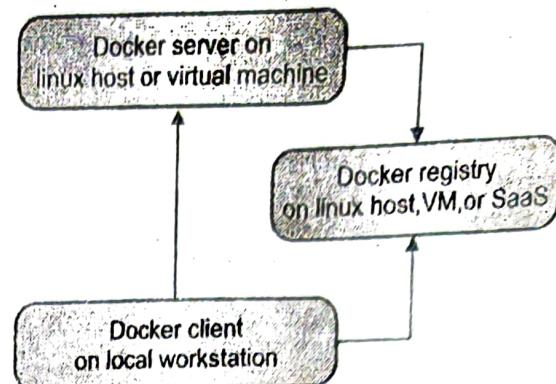


Fig 6.10 : Docker client / Server Model

The structure of Docker is little bit different from some other client/server technologies. Docker uses same binary for client and server executables instead of having separate components. When Docker is installed, both the components get available, but only on supported Linux host, the server can be launched. The launching the Docker server/daemon is very easy, same as of executing Docker using the -d command-line argument, which instructs it to work like a daemon and listen for incoming connections. Usually there is one Docker daemon running for each Docker host which has ability to manage a number of containers. Then docker command-line tool client can be used to communicate with server.

#### Q. 24 Write a note on "Robust tooling in Docker".

Ans. :

There are various reasons behind the wide adoption of Docker, in which the most important is its simple and powerful tooling. Since the initial release by Docker, it has been expanding ever wider. This tooling supports both generation of Docker images as well as basic deployment to individual Docker daemons. It also provides the entire functionality required to manage a remote Docker server.

An orchestration toolset is launched by the Docker including Compose, Machine, and Swarm, which ensures cohesive deployment mechanism across several environments. Docker supports easy accumulation of additional tooling in any language with the help of command-line tool and a remote web API.

#### Q. 25 What is Docker Command Line Tool ?

Ans. :

The Docker command-line tool is the main interface which is preferred by most of the users. This is considered as a Go application which compiles and executed on all the types of major architectures and OSes. The command-line tool is basically provided as a part of the main Docker distribution on number of platforms and also can compile directly from the Go source.

Some of the functionalities of Docker command-line tool are :

- (i) Build a container image.

- (ii) Push the images from the Docker daemon to registry or pull them from a registry to a Docker daemon.
- (iii) Starts a container on a Docker server either in the foreground or background.
- (iv) Access the Docker logs from a remote server.
- (v) Starts a command-line shell inside a running container on a remote server.

#### **Q. 26 Write a note on Docker API.**

**Ans. :**

Docker daemon has a remote API just like many other pieces of modern software. This API is used by the Docker command-line tool for communication with the daemon. The API can also be used directly by the external tooling as it is documented and public. It provides various features like mapping deployed Docker containers with servers, automating deployment, and provides distributed schedulers.

Even though the beginners will not like to interact directly to the Docker API, but it's most likely that they will gradually find the API to be a good integration point for this tooling. Nowadays in many popular languages, robust implementations of Docker API libraries have begun to appear. Most of the functionalities of the Docker command-line tooling are supported easily by the API.

#### **Q. 27 Write a note on "Docker are containers not virtual machines".**

**Ans. :**

Docker can be considered as containers not as virtual machines. They are very lightweight wrappers around a single Unix process. In the actual implementation, this process may spawn others, but instead there is a statically compiled binary inside the container. Containers are also considered as ephemeral that means they may come and go more readily as compared to a virtual machine. Virtual machines are considered as; by design a stand-in for actual hardware. Virtual machines have long life and they abstract the real server.

Even in the environment cloud where the organizations usually spin virtual machines up and down depending upon the demand, they generally have a running lifespan of days or more. On the contrary, a specific container might exist for months, or it is also possible to create it for a minute and destroyed as soon as work is done.

#### **Q. 28 Write a note on building application in Docker.**

**Ans. :**

Application building is really a tedious task in many organizations where only few people are aware of all the concerns regarding creation of a well-formed, shippable artifact. Docker solves most of the issues regarding getting a right build. A standardized tool configuration and tool set is provided by the Docker for builds. This makes it easy for people to understand how to build an application and get new builds up and running. A build flag is provided in the Docker command-line tool, which consumes a Dockerfile and generates a Docker image. A new layer in the image is created by each and every command in a Dockerfile which makes it easy to understand the working of build through Dockerfile.

The most advantageous part of this standardization is that an engineer who has handled the Dockerfile can dive right in and make changes in the build of various different applications. As the Docker image is considered as a standardized artifact, all the components behind the build are same irrespective of the language used, the OS distribution, or the number of required layers. Most of the times, the Docker builds are only single result of the Docker build command and create only container image. The Dockerfile contains the entire logic about the build, hence for the development team, it becomes easy to create standard build jobs.

#### **Q. 29 Explain Testing in Docker workflow.**

**Ans. :**

In Docker, there is no any built-in framework for testing. Docker containers are used for testing purpose. There are many forms of testing from unit testing to complete integration testing. Docker facilitates better testing with assurance that the only those products will be shipped which has passed the testing. The Docker SHA for the container or a custom tag is used to confirm that there is consistent shipping of the same version of the application which is successfully tested. The container tests all of the dependencies with application.

When unit test against a container image is successful, then it is sure that there will not be any problem regarding the versioning of an underlying library at deployment time. This is difficult in most of the other technologies. Another benefit of using Docker containers is that in sites where there are several applications which communicate with each other remotely using API, it is easy for developers of one application to develop against a version of another service which is currently tagged for their expected environment such as production or staging. There is no need for developers of each team to have knowledge about the working and deployment of other service.