UNIT I

1 Fundamentals of Cloud Computing

Syllabus

Origins and Influences, Basic Concepts and Terminology, Goals and Benefits, Risks and Challenges, Roles and Boundaries, Cloud Characteristics, Cloud Delivery Models, Cloud Deployment Models, Federated Cloud / Intercloud, Types of Clouds.

Contents

1.1 Origins and Influences

- Idea of cloud computing was introduced by computer scientist John McCarthy publicly in 1961. Then in 1969, Leonard Kleinrock, a chief scientist of the ARPANET project comments about Internet.
- The general public has been leveraging forms of Internet-based computer utilities since the mid-1990s through various incarnations of search engines, e-mail services, open publishing platforms and other types of social media.
- Though consumer-centric, these services popularized and validated core concepts that form the basis of modern-day cloud computing. The Salesforce.com provides remote service from 1990 to organizations.
- Amazon launched its web services in 2002 and it provides services to organizations for storage and remote computing. Cloud computing definition as per Gartner "a style of computing in which scalable and elastic IT-enabled capabilities are delivered as a service to external customers using Internet technologies".
- In 2008, Gartner's original definition of cloud was changed. In the definition, "massively scalable" was used instead of "scalable and elastic."
- **NIST definition of cloud :** Cloud computing is a pay-per-use model for enabling available, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, services) that can be rapidly provisioned and released with minimal management effort or service-provider interaction.
- The above cloud definition was published by NIST in 2009, followed by a revised version after further review and industry input that was published in September of 2011.
- Cloud computing refers to a variety of services available over the Internet that deliver compute functionality on the service provider's infrastructure.
- Its environment (infrastructure) may actually be hosted on either a grid or utility computing environment, but that doesn't matter to a service user.

1.1.1 Business Drivers of Cloud Computing

1. **Capacity planning :** Storage capacity is one of the main reasons for organization using cloud. Capacity planning is an unavoidable responsibility for most IT organizations. Future demands from business need to be planned for and accommodated. This can be very challenging because this involves estimating the usage and specially usage fluctuations over time. So, there is constant need to balance peak usage requirements without unnecessarily over-spending on-premise IT infrastructure.

2. **Cost reduction and operating overhead :** For any organization, initial investment of cloud is huge. The growth of IT environments often corresponds to the assessment of their maximum usage requirements. This can make the support of new and expanded business automations an ever-increasing investment.
3. **Organizational agility :** From cloud perspective IT organizations, the IT resources needs to be more available and/or reliable than previously thought. The ability for an IT organization to be able to respond to these changes in capacity or availability helps to increase an organizational agility.

1.1.2 Technology Innovations

- It is pre-existing technologies considered to be the primary influences on cloud computing.

1. Grid computing technology :

- In 1990, grid computing was started. Grid is a hardware and software infrastructure that provides dependable, consistent, pervasive and inexpensive access to high-end computational facilities. It is "pay-as-you-go" pricing model.
- Grid computing is a distributed computing system where a group of computers are connected to create and work as one large virtual computing power, storage, database, application and service.
- Grid computing environment is a necessity to many end-users who cannot afford huge computational resources, both hardware and software.
- Therefore, any large corporate body or government organization having a large geographical spread will be essentially required to set up at least some kind of grid computing environment, so that the expensive resources of their grid can be shared and effectively utilized by all the end users.
- Grid computing is based on a middleware layer that is deployed on computing resources. These IT resources participate in a grid pool that implements a series of workload distribution and coordination functions.

2. Clustering technology :

- Cluster is a group of linked computers, working together closely thus in many respects forming a single computer. Clusters are usually deployed to improve performance and availability over that of a single computer, while typically being much more cost-effective than single computers of comparable speed or availability.
- Clustering allows us to run applications on several parallel servers. The load is distributed across different servers and even if any of the servers fails, the application is still accessible via other cluster nodes.

- Clustering is crucial for scalable enterprise applications, as user can improve performance by simply adding more nodes to the cluster.

3. Virtualization technology :

- Virtualization is an already established and proven technology that has enabled IT organizations to repeatedly leverage physical servers for wide, concurrent usage.
- Virtualization is an abstraction layer that decouples the physical hardware from the operating system to deliver greater IT resource utilization and flexibility.
- It allows multiple virtual machines, with heterogeneous operating systems to run in isolation, side-by-side on the same physical machine. Virtualization is an absolute key technology in modern cloud computing environments.
- As cloud computing evolved, a generation of modern virtualization technologies emerged to overcome the performance, reliability and scalability limitations of traditional virtualization platforms.

1.2 Basic Concepts and Terminology

SPPU : April-18, May-19, Dec.-18, March-20

- Cloud computing refer to a variety of services available over the Internet that deliver compute functionality on the service provider's infrastructure.
- Its environment (infrastructure) may actually be hosted on either a grid or utility computing environment, but that doesn't matter to a service user.
- Cloud computing is a general term used to describe a new class of network based computing that takes place over the Internet, basically a step on from utility computing.
- In other words, this is a collection/group of integrated and networked hardware, software and Internet infrastructure (called a platform).
- Cloud computing refers to applications and services that run on a distributed network using virtualized resources and accessed by common Internet protocols and networking standards.
- Fig. 1.2.1 shows cloud symbol. It denotes cloud boundary.
- Using the Internet for communication and transport provides hardware, software and networking services to clients.
- These platforms hide the complexity and details of the underlying infrastructure from users and applications by providing very simple graphical interface or API.

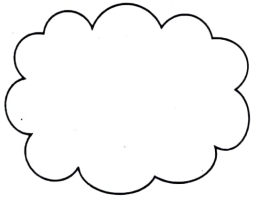


Fig. 1.2.1 Cloud symbol

- In addition, the platform provides on demand services that are always on anywhere, anytime and anyplace. Pay for use and as needed.
- The hardware and software services are available to the general public, enterprises, corporations and business markets.

1.2.1 IT Resources

- IT resources are of two types : Software based and hardware based.
- Software based resources are virtual server, custom software program and hardware based means physical server and networking devices.
- IT resources include server, virtual server, storage device, networking device, services and software programs.
- An on-premise IT resource can access and interact with a cloud-based IT resource.

1.2.2 Scaling

- Scaling is the capability of a system, network or process to handle a growing amount of work or its potential to be enlarged to accommodate that growth. For IT resources, scaling represents the ability of the IT resource to handle increased or decreased usage demands.
- One of the key aspects that made cloud popular is scalability, that means you can increase or decrease your resources at any given time.

1. Horizontal scaling :

- It is scaling out and scaling in. The allocating or releasing of IT resources that are of the same type is referred to as horizontal scaling.
- Horizontal scaling, means increasing the number of nodes in the cluster, reduces the responsibilities of each member node by spreading the keyspace wider and providing additional end-points for client connections.
- Fig. 1.2.2 shows horizontal scaling.

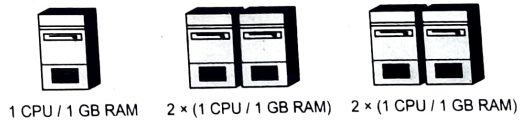


Fig. 1.2.2 Horizontal scaling

- Horizontal scaling affords the ability to scale wider to deal with traffic. It is the ability to connect multiple hardware or software entities, such as servers, so that they work as a single logical unit.

- The horizontal allocation of resources is referred to as scaling out and the horizontal releasing of resources is referred to as scaling in.

2. Vertical scaling :

- Vertical scaling can essentially resize your server with no change to your code. It is the ability to increase the capacity of existing hardware or software by adding resources. Vertical scaling usually means upgrade of server hardware.
- Vertical scaling is limited by the fact that you can only get as big as the size of the server. Fig. 1.2.3 shows vertical scaling.
- The replacing of an IT resource with another that has a higher capacity is referred to as scaling up and the replacing an IT resource with another that has a lower capacity is considered scaling down.
- Vertical scaling is much more used in small and middle-sized companies and in applications and products of middle-range.

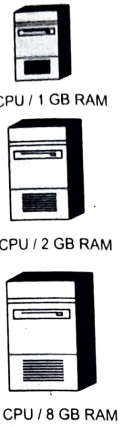


Fig. 1.2.3 Vertical scaling

1.2.3 Difference between Horizontal and Vertical Scaling

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

1.2.4 Cloud Components

- Cloud computing solutions are made up of several elements. Fig. 1.2.4 shows cloud components.

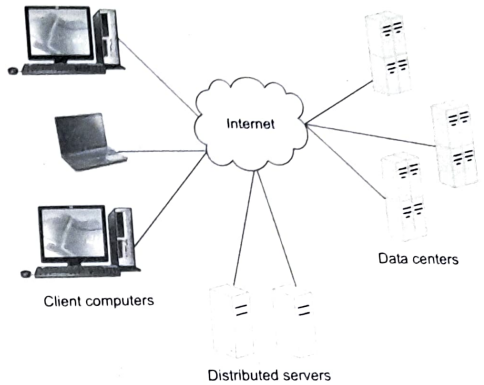


Fig. 1.2.4 Cloud components

1. **Clients :** Mobile, terminals or regular computers.
2. **Benefits :** Lower hardware costs, lower IT costs, security, data security, less power consumption, ease of repair or replacement, less noise.
3. **Data centers :** Collection of servers where the application to subscribe is housed. It could be a large room in the basement of your building or a room full of servers on the other side of the world.
4. **Virtualizing servers :** Software can be installed allowing multiple instances of virtual servers to be used and a dozen virtual servers can run on one physical server.
5. **Distributed servers :** Servers don't all have to be housed in the same location. It can be in geographically disparate locations. If something were to happen at one site, causing a failure, the service would still be accessed through another site. If the cloud needs more hardware, they can add them at another site.

1.2.5 Cloud Service and Consumer

- Cloud service is any service made available to users on demand via the internet from a cloud computing provider's servers as opposed to being provided from a company's own on-premises servers.

- A cloud service can exist as a simple web-based software program with a technical interface invoked via the use of a message protocol or as a remote access point for adminstrative tools or larger environments and other IT resources.
- The organization that provides cloud-based IT resoucrs is cloud provider. Cloud providers normally own the IT resources for lease by cloud consumers and could also resell IT resources leased from other providers.

Cloud consumer

- A cloud consumer is an organization that has a formal contract or arrangement with a cloud provider to use IT resources made available by the cloud provider.
- The cloud consumer uses a cloud service consumer to access a cloud service.

Review Questions

1. *What is the difference between horizontal scaling and vertical scaling ?*

SPPU : April-18 In Sem, May-19 End Sem, Marks 4

2. *Define cloud computing. Explain different types of cloud computing.*

SPPU : Dec.-18 End Sem, Marks 8

3. *What is cloud computing ? Explain advantages and disadvantages of cloud computing.*

SPPU : March-20, In Sem, Marks-5

1.3 Goals and Benefits

SPPU : May-18, Dec.-19

Pros of cloud computing :

1. **Lower computer costs :** Since applications run in the cloud, not on the desktop PC, your desktop PC does not need the processing power or hard disk space demanded by traditional desktop software.
2. **Improved performance :** Computers in a cloud computing system boot and run faster because they have fewer programs and processes loaded into memory.
3. **Reduced software costs :** Instead of purchasing expensive software applications, you can get most of what you need for free.
4. **Instant software updates :** When you access a web-based application, you get the latest version - without needing to pay for or download an upgrade.
5. **Improved document format compatibility :** You do not have to worry about the documents you create on your machine being compatible with other user's applications or operating systems.
6. **Unlimited storage capacity :** Cloud computing offers virtually limitless storage.

7. **Increased data reliability :** Unlike desktop computing, in which if a hard disk crashes and destroy all your valuable data, a computer crashing in the cloud should not affect the storage of your data.
8. **Universal document access :** All your documents are instantly available from wherever you are.
9. **Latest version availability :** The cloud always hosts the latest version of your documents; as long as you are connected, you are not in danger of having an outdated version.
10. **Easier group collaboration :** Sharing documents leads directly to better collaboration.
11. **Device independence :** Move to a portable device and your applications and documents are still available.

Cons of cloud computing :

1. It requires a constant Internet connection : Cloud computing is impossible if you cannot connect to the Internet.
2. Features might be limited.
3. Stored data might not be secure : With cloud computing, all your data is stored on the cloud.
4. Does not work well with low-speed connections.

Review Questions

1. *Explain advantages and limitations of cloud computing in brief.*

SPPU : May-18 End Sem, Marks 6

2. *Explain advantages and disadvantages of cloud computing.*

SPPU : Dec.-19 End Sem, Marks 5

1.4 Risks and Challenges

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

- Use of cloud for business purpose means that the responsibility over data security becomes shared with the cloud provider. Organization extends their trust boundary to cloud consumer to external cloud.

- It is clear that the security issue has played the most important role in hindering cloud computing acceptance.
- Without doubt, putting your data, running your software on someone else's hard disk using someone else's CPU appears daunting to many.
- Well-known security issues such as data loss, phishing, pose serious threats to organization's data and software.

1.5 Roles and Boundaries

- Organizations and humans can assume different types of predefined roles depending on how they relate to and/or interact with a cloud and its hosted IT resources. The cloud computing defines these roles and identifies their main interactions.

1. Cloud provider :

- A person, organization or entity responsible for making a service available to interested parties. When assuming the role of cloud provider, an organization is responsible for making cloud services available to cloud consumers, as per agreed upon Service Level Agreement (SLA) guarantees. Cloud providers have their own IT resources.
- Fig. 1.5.1 shows cloud provider.

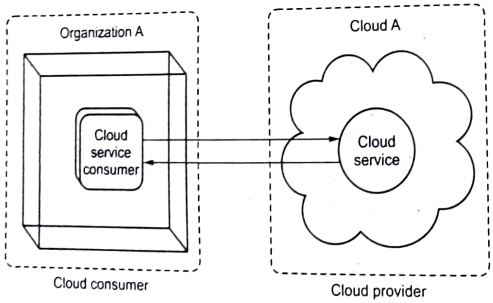


Fig. 1.5.1 Cloud service and cloud service consumer

- A cloud provider would have a significant number of roles responsible for the management of its cloud resources, including those responsible for setting up, onboarding, configuring and supporting cloud services for its consumers.

2. Cloud consumer :

- A person or organization that maintains a business relationship with and uses service from, cloud providers. The cloud consumer uses a cloud service consumer to access a cloud service.
- Anyone who purchases a cloud service is a consumer and within the consumer there could be an array of roles responsible for configuring and managing the resources from the cloud provider depending on the services obtained.

3. Cloud service owner :

- The person or organization that legally owns a cloud service is called a cloud service owner. The cloud service owner can be the cloud consumer or the cloud provider that owns the cloud within which the cloud service resides.
- Fig. 1.5.2 shows cloud service owner.

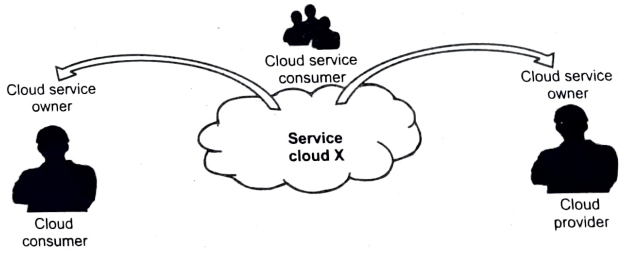


Fig. 1.5.2 Cloud service owner

- The reason a cloud service owner is not called a cloud resource owner is because the cloud service owner role only applies to cloud services.

4. Resource administrator :

- Cloud resource administrator is the person or organization responsible for administering a cloud-based IT resource. The cloud consumer or cloud provider or even third-party organization could be a cloud resource administrator.
- For example, a cloud service owner can contract a cloud resource administrator to administer a cloud service.

5. Cloud auditor :

- Cloud auditor is a party that can conduct independent assessment of cloud services, information system operations, performance and security of the cloud implementation. Generally, cloud auditors are categorized based on intent.

- For the most part, their focus is on risk and compliance, especially around information security. Other auditors can provide advisory services especially to consumers looking to cut down their bills or raise the level of efficiency in the resources consumed.

6. Cloud broker :

- Cloud broker is any entity that manages the use, performance, and delivery of cloud services and negotiates relationships between cloud providers and cloud consumers.
- Cloud brokers support consumers to get value for money by playing the advisory role especially for consumers who have a hybrid mix of resources from multiple providers.

7. Cloud carrier :

- Cloud carrier is an intermediary that provides connectivity and transport of cloud services from cloud providers to cloud consumers.
- Most ISPs have taken the role of cloud carriers as they provide the requisite bandwidth needed to connect consumers with providers as well as capabilities that support the connectivity.

8. Trust boundary :

- Logical perimeter that typically spans beyond physical boundaries to represent the extent to which IT resources are trusted. Fig. 1.5.3 shows trust boundary.

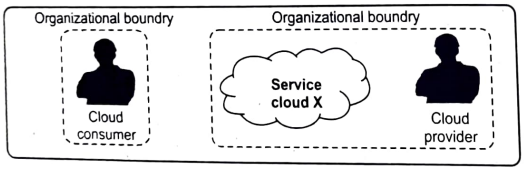


Fig. 1.5.3 Trust boundary

- When analysing cloud environments, the trust boundary is most frequently associated with the trust issued by the organization acting as the cloud consumer.

1.6 Cloud Characteristics

SPPU : April-18,19, Dec.-18, March-20

1. **On-demand self-service :** A consumer can unilaterally provision computing capabilities, such as server time and network storage, as needed without requiring human interaction with each service's provider.

2. **Ubiquitous network access :** Capabilities are available over the network and accessed through standard mechanisms that promote use by heterogeneous thin or thick client platforms.
3. **Location-independent resource pooling :** The provider's computing resources are pooled to serve all consumers using a multi-tenant model, with different physical and virtual resources dynamically assigned and reassigned according to consumer demand.
4. **Rapid elasticity :** Capabilities can be rapidly and elastically provisioned to quickly scale up, and rapidly released to quickly scale down.
5. **Pay per use :** Capabilities are charged using a metered, fee-for-service or advertising-based billing model to promote optimization of resource use.

Review Questions

1. *State and explain characteristics of cloud computing.*

SPPU : April-18 In Sem, Dec.-18 End Sem, Marks 6, March-20, In Sem, Marks 5

2. *Enlist and explain in brief any six characteristics of cloud computing.*

SPPU : April-19 In Sem, Marks 6

1.7 Cloud Delivery Models

SPPU : April-18,19, Dec.-18,19, May-19

- Service models describe the type of service that the service provider is offering. The best-known service models are software as a service, platform as a service, and Infrastructure as a service.
- The service models build on one another and define what a vendor must manage and what the client's responsibility is.
- Service models : This consists of the particular types of services that you can access on a cloud computing platform.
- Cloud service is any service made available to users on demand via the Internet from a cloud computing provider's servers as opposed to being provided from a company's own on-premises servers.
- Cloud services are designed to provide easy, scalable access to applications, resources and services and are fully managed by a cloud services provider.
- A cloud service can exist as a simple web-based software program with a technical interface invoked via the use of a messaging protocol or as a remote access point for administrative tools or larger environments and other IT resources.

- The organization that provides cloud-based IT resources is the cloud provider. Cloud providers normally own the IT resources for lease by cloud consumers and could also resell IT resources leased from other providers.
- Cloud computing, often described as a stack, has a broad range of services built on top of one another under the name cloud.
- Fig. 1.7.1 shows cloud computing stack.
- Flavors of cloud computing is as follows;

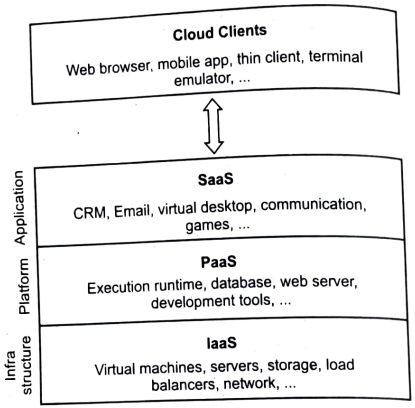


Fig. 1.7.1 Cloud computing stack

1. SaaS applications are designed for end-users, delivered over the web.
2. PaaS is the set of tools and services designed to make coding and deploying those applications quick and efficient.
3. IaaS is the hardware and software that powers it all - servers, storage, networks, operating systems.

1.7.1 Software as a Service (SaaS)

- Model in which an application is hosted as a service to customers who access it via the Internet.
- The provider does all the patching and upgrades as well as keeping the infrastructure running.
- The traditional model of software distribution, in which software is purchased for and installed on personal computers, is referred to as product.
- In this model, the user, client or consumer runs an application from a cloud infrastructure. Through an interface such as a web browser, the client or user may access this application from a variety of devices.
- The complete application is offered as on demand service. This saves the client from having to invest in any software licenses or servers up front and can save

the provider money since they are maintaining and providing only a single application.

- In this model, the client does not manage cloud infrastructure, networks or servers, storage or operating systems. Even, Microsoft, Google and Zoho offer SaaS.
- The SaaS concept can be defined as providing robust "web-based, on-demand software, storage and various applications" to organizations.
- The SaaS model has emerged as an alternative to traditional one-time licensing for providing and maintaining the software needed by knowledge workers within organizations.
- Fig. 1.7.2 shows SaaS.

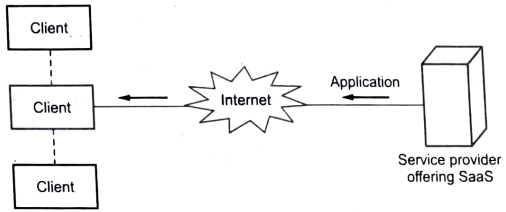


Fig 1.7.2 SaaS

Characteristics of SaaS :

1. Software applications or services are stored remotely.
2. A user can then access these services or software applications via the Internet.
3. In most cases, a user does not have to install anything onto their host machine, all they require is a web browser to access these services and in some cases, a browser may require additional plug-in/add-on for certain services.
4. Network-based management and access to commercially available software from central locations rather than at each customer's site, enabling customers to access applications remotely via the Internet.
5. Application delivery from a one-to-many model, as opposed to a traditional one-to- one model.

Benefits of SaaS :

1. You only pay for what you use.
2. Easier administration and invoicing.

3. Automatic updates and patch management.
4. Compatibility : All users have access to the same version of software.
5. Easier collaboration.
6. It support automated update and patch management services.

1.7.2 Platform as a Service (PaaS)

- Platform as a service is another application delivery model and also known as **cloud-ware**. Supplies all the resources required to build applications and services completely from the Internet, without having to download or install software.
- Services include : Application design, development, testing, deployment and hosting, team collaboration, web service integration, database integration, security, scalability, storage, state management and versioning.
- PaaS is closely related to SaaS but delivers a platform from which to work rather than an application to work with.
- This model involves software encapsulated and offered as a service, from which higher levels of service may then be built. The user, customer or client in this model is the one building applications which then run on the provider's infrastructure.
- This in turn provides customers and clients with the capability to deploy applications onto the cloud infrastructure using programming tools and languages, which the provider supports.
- The customer still does not manage the framework, network, servers or operating system, but has control over deployed applications and sometimes over the hosting environment itself.
- Some examples of Platform as a Service include Google's App Engine or Force.com
- PaaS consists of following components :
  1. Browser based development studio.
  2. Pay contrary to billing.
  3. Management and supervising tools.
  4. Seamless deployment to host run time environment.
- **Characteristics of PaaS :**
  1. It support multi-tenant architecture.
  2. It support for development of group collaboration.
  3. PaaS systems can be deployed as public cloud services or as private cloud services.

4. Provision of runtime environments. Typically each runtime environment supports either one or a small set of programming languages and frameworks.
5. Support for custom applications. Support for the development, deployment and operation of custom applications.
6. Preconfigured capabilities. Many PaaS systems are characterized by capabilities that are preconfigured by the provider, with a minimum of configuration available to developers and customer operations staff.
7. Support for porting existing applications. While many PaaS systems are primarily designed to support "born on the cloud" applications.
8. Security is an important characteristic in PaaS. It needs to provide authentication and authorization to differentiate the access rights of different users.

Benefits of Paas :

1. Scalability including rapid allocation and deallocation of resources with a pay-as-you-use model.
2. Reduced capital expenditure.
3. Reduced lead times with on-demand availability of resources.
4. Self-service with reduced administration costs.
5. Reduced skill requirements.
6. Support of team collaboration.
7. Ability to add new users quickly.

1.7.3 Infrastructure as a Service (IaaS)

- IaaS gives the storage room likeness to the in-house datacenter stood out from various organizations sorts.
- Center datacenter framework segments are capacity, servers (registering units), the system itself, and administration apparatuses for foundation upkeep and checking.
- Each of these parts has made a different market specialty. While some little organizations have practical experience in just a single of these IaaS cloud specialties, vast cloud suppliers like Amazon or Right Scale have offerings over all IaaS territories.
- Fig. 1.7.3 shows IaaS.

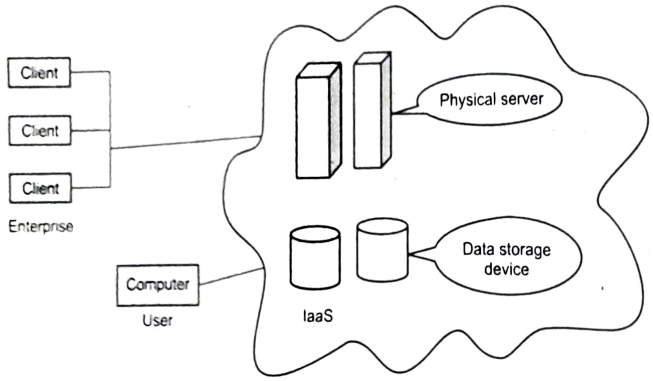


Fig. 1.7.3 IaaS

- It offers the hardware so that your organization can put whatever they want onto it. Rather than purchase servers, software, racks and having to pay for the datacenter space for them, the service provider rents those resources :
  1. Server space 2. Network equipment
  3. Memory 4. CPU cycles 5. Storage space
- Again, the customer is not managing cloud infrastructure, but in this case, the customer does control operating systems, deployed applications, storage and sometimes-certain networking components.
- Examples : Amazon EC2, Rackspace Mosso, GoGrid
- IaaS server types :
  1. **Physical server :** Actual hardware is allocated for the customer's dedicated use.
  2. **Dedicated virtual server :** The customer is allocated a virtual server, which runs on a physical server that may or may not have other virtual servers.
  3. **Shared virtual server :** The customer can access a virtual server on a device that may be shared with other customers.

Advantages of IaaS :

1. Elimination of an expensive and staff-intensive data center.
2. Ease of hardware scalability.
3. Reduced hardware costs.
4. On-demand, pay as you go scalability.

5. Reduction of IT staff.
6. Suitability for ad hoc test environments.
7. Allows complete system administration and management.
8. Support multiple tenants.

1.7.4 Difference between IaaS, PaaS and SaaS

IaaS	PaaS	SaaS
IaaS gives users automated and scalable environments.	PaaS provides a framework for quickly developing and deploying applications.	SaaS makes applications available through the internet.
Amazon Web Services, for example, offers IaaS through the Elastic Compute Cloud or EC2.	Google Cloud Platform provides another PaaS option in App Engine.	SaaS applications such as Gmail, Dropbox, Salesforce or Netflix.
In Iaas, infrastructure as a service.	In Paas, platform as a service.	In SaaS, software as a service
Virtual platform on which required operating environment and application deployed.	Operating environment was included.	Operating environment largely irrelevant, fully functional application provided.
IaaS is a cloud service that provides basic computing infrastructure : Servers, storage, and networking resources. In other words, IaaS is a virtual data center.	PaaS refers to cloud platforms that provide runtime environments for developing, testing and managing applications.	SaaS allows people to use cloud-based web applications.
Major IaaS providers include Amazon Web Services, Microsoft Azure and Google Compute Engine.	Examples of PaaS services are Heroku and Google App Engine.	email services such as Gmail and Hotmail are examples of cloud-based SaaS services.
IaaS services are available on a pay-for-what-you-use model.	PaaS solutions are available with a pay-as-you-go pricing model.	SaaS services are usually available with a pay-as-you-go pricing model.
Used by IT administrator.	Used by software developers.	Used by end user.

Review Questions

1. *Explain any two cloud delivery models,* **SPPU : April-18 In Sem, Marks 4**
2. *Explain benefits of IaaS.* **SPPU : Dec.-18 End Sem, Marks 6**
3. *Explain cloud delivery models with example.* **SPPU : April-19 In Sem, Marks 6**

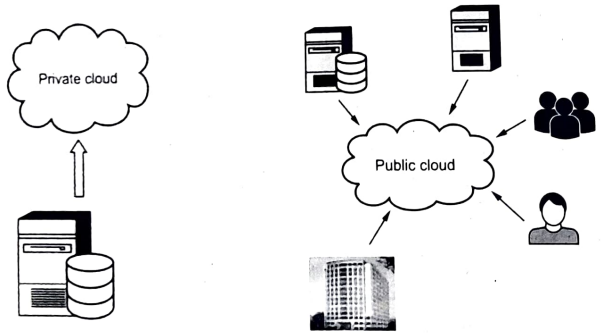
4. *Compare and contrast IaaS, SaaS, PaaS related to consumer activities and provider activities.* **SPPU : May-19 End Sem, Marks 6**

5. *Compare different cloud delivery models.* **SPPU : Dec.-19 End Sem Marks 5**

1.8 Cloud Deployment Models

SPPU : May-18, Dec.-18, 19, March-20

- Cloud deployment models are refers to the location and management of the cloud's infrastructure.
- Deployment models are defined by the ownership and control of architectural design and the degree of available customization. Cloud deployment models are private public and community clouds.
- Fig. 1.8.1 shows cloud deployment model.



(a) Private cloud (b) Public cloud

Fig. 1.8.1 Cloud deployment model

1. Public cloud :

- The cloud infrastructure is made available to the general public or a large industry group and is owned by an organization selling cloud services.
- Public cloud is a huge data centre that offers the same services to all its users. The services are accessible for everyone and much used for the consumer segment.
- Examples of public services are Facebook, Google and LinkedIn.
- **Public cloud benefits :**
  a) Low investment hurdle : Pay for what user use.
  b) Good test/development environment for applications that scale to many servers.

- **Public cloud risks :**
  a) Security concerns : Multi-tenancy and transfers over the Internet.
  b) IT organization may react negatively to loss of control over data center function.

2. Private cloud :

- The cloud infrastructure is operated solely for a single organization. It may be managed by the organization or a third party and may exist on-premises or off-premises.
- **Private cloud benefits :**
  a) Fewer security concerns as existing data center security stays in place.
  b) IT organization retains control over data center.
- **Private cloud risks :**
  a) High investment hurdle in private cloud implementation, along with purchases of new hardware and software.
  b) New operational processes are required; old processes not all suitable for private cloud.

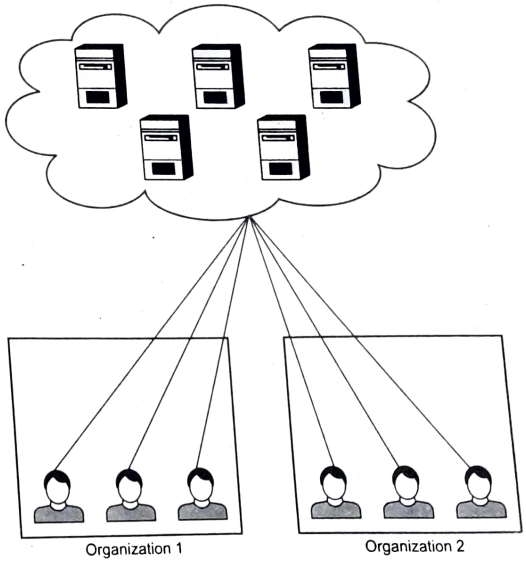


Fig. 1.8.2 Community cloud

3. Community cloud :

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

4. Hybrid cloud :

- The cloud infrastructure is a composition of two or more clouds (private, community or public) that remain unique entities but are bound together by standardized or proprietary technology that enables data and application portability (e.g., cloud bursting for load-balancing between clouds).
- **Hybrid cloud benefits :**
  a) Operational flexibility : Run mission critical on private cloud, dev/test on public cloud.
  b) Scalability : Run peak and bursty workloads on the public cloud.
- **Hybrid cloud risks :**
  a) Hybrid clouds are still being developed; not many in real use.
  b) Control of security between private and public clouds, some of same concerns as in public cloud.

1.8.1 Difference between Public and Private Cloud

Public cloud	Private cloud
Public cloud infrastructure is offered via web applications and also as web services over Internet to the public.	Private cloud infrastructure is dedicated to a single organization.
Support multiple customer.	Support dedicated customer.
Full utilized of infrastructure.	Does not utilize shared infrastructure.
Security is low as compared to private cloud.	High level of security.
Low cost	High cost
Azure, Amazon Web Services, Google App Engine and Force.com are a few examples of public clouds.	An example of the Private Cloud is NIRIX's one Server with dedicated servers.

Review Questions

1. *Compare private cloud versus public cloud.*
   SPPU : May-18 End Sem, Marks 4, Dec.-19, End Sem, Marks 5

2. *Explain different types of cloud deployment models.* **SPPU : Dec.-18 End Sem, Marks 8**
3. *Differentiate between deployment models : Private, public and hybrid*
   SPPU : March-20, In Sem, Marks 5

1.9 Federated Cloud / Intercloud

- The inter-cloud is an interconnected global "cloud of clouds". Intercloud Architecture Framework (ICAF) provides a framework to support provisioning of cloud based project oriented infrastructures on-demand and distributed virtualized applications mobility. Each cloud should be able to work and offer its services without any dependence with other clouds.
- The main objective of intercloud is to create an open interface to ease the exchange of data from one cloud to another. The connections are established between one or more clouds for this systematic exchange of data.
  a) Resources, services and data are shared through the intercloud architecture.
  b) The intercloud architecture is scalable and able to add new clouds.
  c) The availability of the resources, services and data should not depend on the customer's applications.
  d) The architecture should be able to provide better load balancing capabilities.

Need of inter-cloud

- The limitations of cloud are that they have limited physical resources. If a cloud has exhausted all the computational and storage resources, it cannot provide service to the clients. The inter-cloud addresses such situations where each cloud would use the computational, storage or any kind of resource of the infrastructures of other clouds.
- The inter-cloud environment provides benefits like diverse geographical locations, better application resilience and avoiding vendor lock-in to the cloud client. Benefits for the cloud provider are expand-on-demand and better Service Level Agreements (SLA) to the cloud client.

Types of inter-cloud :

1. Federation clouds

- A federation cloud is an inter-cloud where a set of cloud providers willingly interconnect their cloud infrastructures in order to share resources among each other.
- The cloud providers in the federation voluntarily collaborate to exchange resources. This type of inter-cloud is suitable for collaboration of governmental clouds or private cloud portfolios.
- Types of federation clouds are peer to peer and centralized clouds.

2. multi-cloud

- In a multi-cloud, a client or service uses multiple independent clouds. A multi-cloud environment has no volunteer interconnection and sharing of the cloud service providers' infrastructures.
- Managing resource provisioning and scheduling is the responsibility of client or their representatives. This approach is used to utilize resources from both governmental clouds and private cloud portfolios.
- Types of multi-cloud are services and libraries.

1.10 Multiple Choice Questions

Q.1 Point out the wrong statement :

a Abstraction enables the key benefit of cloud computing : Shared, ubiquitous access.

b Virtualization assigns a logical name for a physical resource and then provides a pointer to that physical resource when a request is made.

c All cloud computing applications combine their resources into pools that can be assigned on demand to users.

d All of the mentioned.

Q.2 Point out the wrong statement :

a The massive scale of cloud computing systems was enabled by the popularization of the Internet.

b Soft computing represents a real paradigm shift in the way in which systems are deployed.

c Cloud computing makes the long-held dream of utility computing possible with a pay-as-you-go, infinitely scalable, universally available system.

d All of the mentioned.

Q.3 Which of the following is essential concept related to cloud ?

a Reliability

b Productivity

c Abstraction

Q.4 Point

UNIT II

2 Cloud-Enabling Technology and Virtualization

Syllabus

Cloud-Enabling Technology : *Broadband Networks and Internet Architecture, Data Center Technology Virtualization Technology, Web Technology, Multitenant Technology, Service Technology.*

Implementation Levels of Virtualization, *Virtualization Structures/Tools and Mechanisms, Types of Hypervisors, Virtualization of CPU, Memory, and I/O Devices, Virtual Clusters and Resource Management, Virtualization for Data-Center Automation.*

Contents

2.1 Cloud - Enabling Technology

SPPU : April-18, Dec.-18,19, May-19, March-20

- Cloud - Enabling technologies are as follows :
  1. Broadband networks and internet architecture
  2. Data center technology
  3. Virtualization technology
  4. Web technology
  5. Multitenant technology

2.1.1 Broadband Networks and Internet Architecture

- All clouds must be connected to a network. Internet's largest backbone networks, established and deployed by ISP's, are interconnected by core routers.
- Cloud consumers have the option of accessing the cloud using only private and dedicated network links in LANs, although most clouds are Internet-enabled.
- Cloud platform generally grow with internet connectivity and service quality.
  1. Internet Service Providers (ISP's)
  2. Connectionless Packet Switching (Datagram Networks)
  3. Router-Based Interconnectivity
  4. Technical and Business Considerations.

1. Internet Service Providers

- An ISP is a company that provides its customers access to the internet and other web services. In addition to maintaining a direct line to the internet, the company usually maintains web servers. Almost all ISPs offer email and web browsing capabilities. They also offer varying degrees of user support, usually in the form of an email address or customer support hotline.
- Fig. 2.1.1 shows messages travel over dynamic network routes in this ISP internetworking configuration. (See Fig. 2.1.1 on next page)
- Internet Corporation for Assigned Names and Numbers (ICANN) is an non-profit corporation that is responsible for allocating IP addresses and managing the domain name system.
- Government and regulatory laws dictate the service provisioning conditions for organizations and ISPs both within and outside of national borders. Certain realms of the internet still require the demarcation of national jurisdiction and legal boundaries.
- The Internet's topology has become a dynamic and complex aggregate of ISPs that are highly interconnected via its core protocols.

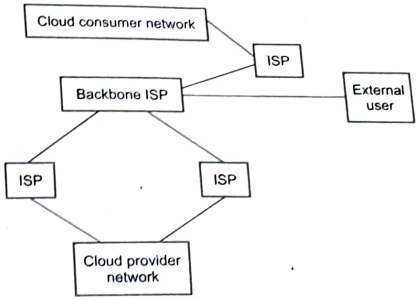


Fig. 2.1.1 ISP

- Smaller branches extend from these major nodes of interconnection, branching outwards through smaller networks until eventually reaching every internet-enabled electronic device. Worldwide connectivity is enabled through a hierarchical topology composed of Tiers 1, 2 and 3.

Tier 1	• Large ISP's directly connected to internet backbone.
	• Network connectivity without paying IP transits.
	• Connected to international gateway.
	• Example : VSNL, Reliance
Tier 2	• Medium size ISP's having peers with some networks.
	• Pays IP transits to reach some parts of the network.
Tier 3	• Local ISP's buying services from Tier 1 and Tier 2 ISP's.
	• No backbone.
	• Focused only on retail market.

- The communication links and routers of the internet and ISP networks are IT resources that are distributed among countless traffic generation paths.
- Two fundamental components used to construct the internetworking architecture are connectionless packet switching (datagram networks) and router - based interconnectivity.

2. Connectionless Packet Switching (Datagram Networks) :

- Data flow between end to end is through limited size packet. It passed through network switches and routers, then queued and forwarded from one intermediary node to the next.

- Each packet carries the information, such as the Internet Protocol (IP) or Media Access Control (MAC) address, to be processed and routed at every source, intermediary, and destination node.

3. Router-Based Interconnectivity :

- A router is a networking device that is connected to multiple networks through which it forwards packets. Router maintains the information like network topology, source address and destination address and other information. Using this information, it forwards the packet.
- Routers also manage network traffic and select the most efficient hop for packet delivery.
- Fig. 2.1.2 shows the basic mechanics of internetworking. The depicted router receives and forwards packets from multiple data flows.

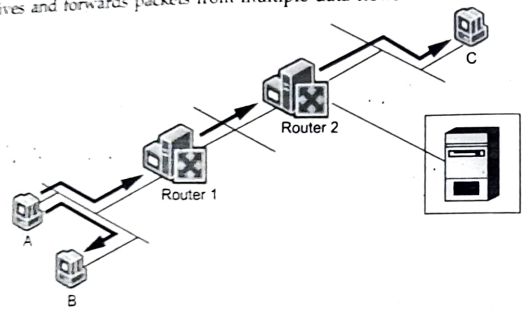


Fig. 2.1.2 Basic mechanics of internetworking

- The communication path that connects a cloud consumer with its cloud provider may involve multiple ISP networks. The internet's mesh structure connects internet hosts (endpoint systems) using multiple alternative network routes that are determined at runtime.
- Communication can therefore be sustained even during simultaneous network failures, although using multiple network paths can cause routing.
- Fig. 2.1.3 shows generic view of the internet reference model and protocol stack. (See Fig. 2.1.3 on next page)
- The internet architecture, which is also sometimes called the TCP/IP architecture after its two main protocols. The TCP/IP reference model is a set of protocols that allow communication across multiple diverse networks.
- Application layer : Application layer includes all process and services that use the transport layer to deliver data. Protocols such as HTTP, SMTP for email,

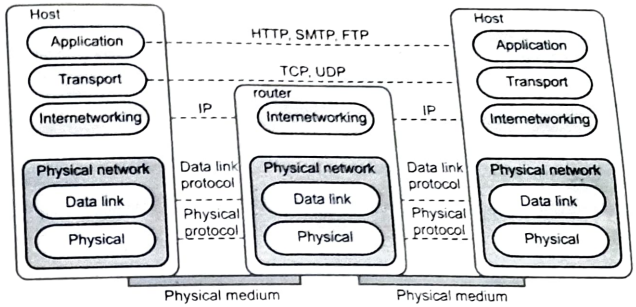


Fig. 2.1.3 Internet reference model and protocol stack

BitTorrent for P2P and SIP for IP telephony use transport layer protocols to standardize and enable specific data packet transferring methods over the internet.

- Transport layer : Application programs send data to the transport layer protocols TCP and UDP. An application is designed to choose either TCP or UDP based on the services it needs. The transport layer provides peer entities on the source and destination hosts to carry on a conversation.
- Physical network : It is responsible for accepting and transmitting IP datagrams. This layer may consist of a device driver in the operating system and the corresponding network interface card in the machine. IP packets are transmitted through underlying physical networks that connect adjacement nodes, such as Ethernet, ATM network, and the 3G mobile.

2.1.2 Comparison between On-Premise and Cloud-Based IT Resources

Sr. No.	On-Premise IT Resources	Cloud-Based IT Resources
1.	Internal end-user devices access corporate IT services through the corporate network.	Internal end-user devices access corporate IT services through an internet connection.
2.	Internal users access corporate IT services through the corporate internet connection while roaming in external networks.	Internal users access corporate IT services while roaming in external networks through the cloud provider's internet connection.
3.	External users access corporate IT services through the corporate internet connection.	External users access corporate IT services through the cloud provider's internet connection.
4.	Security is critical, and you know your assets and your people.	Security is complex, expensive and must be maintained 24 × 7.

2.1.3 Data Center Technology

- Data centers are buildings where multiple servers and communication gear are co-located because of their common environmental requirements and physical security needs and for ease of maintenance. Data centers are specialized environments that safeguard company's most valuable equipment and intellectual property
- Data centers support the following things :
  1. Processing of users business transactions
  2. Hosting of company website
  3. Process and store intellectual property
  4. Maintain financial records
  5. Route electronic mails
- The data center infrastructure is central to the IT architecture, from which all content is sourced or passes through. Proper planning of the data center infrastructure design is critical, and Performance, resiliency, and scalability need to be carefully considered.
- Data center equipment's environmental conditions should fall within the ranges.
- Data center uses five core elements for processing. These elements are application, database, network, storage array, operating system and server.
- The main purpose of a data center is running the applications that handle the core business and operational data of the organization. Data centers are the facilities that will house the equipment in order to secure, store and exchange data.
  1. **User Application** : It is a computer program. Computation is performed in data center. Application may includes order processing, salary calculation etc. It uses operating system and data base for processing.
  2. **Database** : A Database Management System (DBMS) is a software package designed to define, manipulate, retrieve and manage data in a database. A DBMS always provides data independence. Any change in storage mechanism and formats are performed without modifying the entire application. Software for organizing and serving data to users, managing physical storage of media and virtual resources.
  3. **Network** : It provides communication between client and server. Network resources refer to the telecommunication networks like intranets, extranets, and the internet. These resources facilitate the flow of communication in the organization. Networks consist of both physical devices such as network cards, router, hubs and cables, and software such as operating systems, web servers, data servers and application servers.

4. **Storage Array** : It is a device which stores the data persistently for later use.
5. **OS and Server** : OS provides platform for processing.

- Fig. 2.1.4 shows example of sales order processing system.

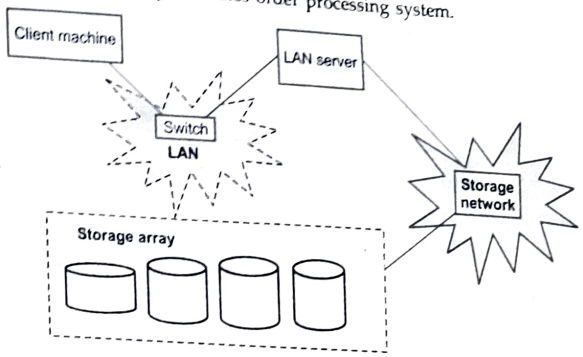


Fig. 2.1.4 Sales order processing system

- On client machine, required application software is installed. Customer can place the order through this software. Client machine is connected with server by using local area network. Required database is installed on the server.
- DBMS uses the server operating system for reading and updating database. Database is store on the secondary storage device in the storage array.
- Storage network provides the communication link between the server and the storage array and transports the read or writes commands between them. The storage array, after receiving the read or write commands from the server, performs the necessary operations to store the data on physical disks.
- A data center is a specialized IT infrastructure that houses centralized IT resources, such as servers, databases and software systems.
- Data center IT hardware is typically comprises of standardized commodity servers of increased computing power and storage capacity, while storage system technologies include disk arrays and storage virtualization. Technologies used to increase storage capacity include DAS, SAN, and NAS.
- Data centers are typically comprised of the following technologies and components :
  1. Virtualization
  2. Standardization and modularity
  3. Automation

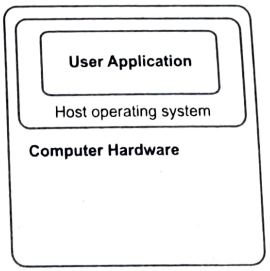
4. Remote operation and management
5. High availiability
6. Security-aware design, operation and management
7. Facilities
8. Computing hardware
9. Storage hardware
10. Network hardware
11. Technical and business considerations

- Data centers requires extensive network hardwork in order to enable multiple levels of connectivity.
- For networking infrastructure, the data center is down into five network subsystems : Carrier and external network interconnection, web - tier load balacing and acceleration.
- Carrier and external networks innterconnection : It consists of backbone routers, firewall and VPN gateways. Backbone routers provide routing between external WAN connections and data center LAN.
- Web - tier load balancing and acclearation : It contains web acceleration device such as XML pre - processors, encryption/decryption appliances, and layer 7 switching devices that perform content - aware routing.
- LAN fabric : It contains an internal LAN and provides high - performance and redundant connectivity for all of the data center's network - enabled IT resources.
- SAN fabric : It provide connectivity between servers and storage systems, the SAN fabric is ususally implemented with Fibre Channel (FC), Fibre Channel over Ethernet (FCoE) and InfiniBand network swicthes.
- NAS gateways : It supplies attachment points for NAS-based storage devices and implements protocol conversion hardware that facilitates data transmission between SAN and NAS devices.

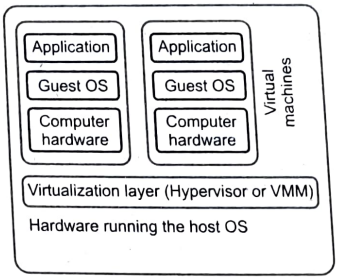
2.1.4 Virtualization Technology

- Virtualization is a broad term that refers to the abstraction of resources across many aspects of computing. For our purposes : One physical machine to support multiple virtual machines that run in parallel.
- Virtualization is a frame work or methodology of dividing the resources of computer into multiple execution environments.
- Virtualization is an abstraction layer that decouples the physical hardware from the operating system to deliver greater IT resource utilization and flexibility.

- It allows multiple virtual machines, with heterogeneous operating systems to run in isolation, side-by-side on the same physical machine.
- Fig. 2.1.5 shows before and after virtualization.



(a) : Before virtualization



(b) After virtualization

Fig. 2.1.5

- Virtualization means running multiple machines on a single hardware. The "Real" hardware invisible to operating system. OS only sees an abstracted out picture. Only Virtual Machine Monitor (VMM) talks to hardware.
- It is "a technique for hiding the physical characteristics of computing resources from the way in which other systems, applications or end users interact with those resources.
- This includes making a single physical resource appear to function as multiple logical resources; or it can include making multiple physical resources appear as a single logical resource."
- It is divided into two main categories :
  1. Platform virtualization involves the simulation of virtual machines.

2. Resource virtualization involves the simulation of combined, fragmented or simplified resources.

- Fig. 2.1.6 shows taxonomy of virtualization.

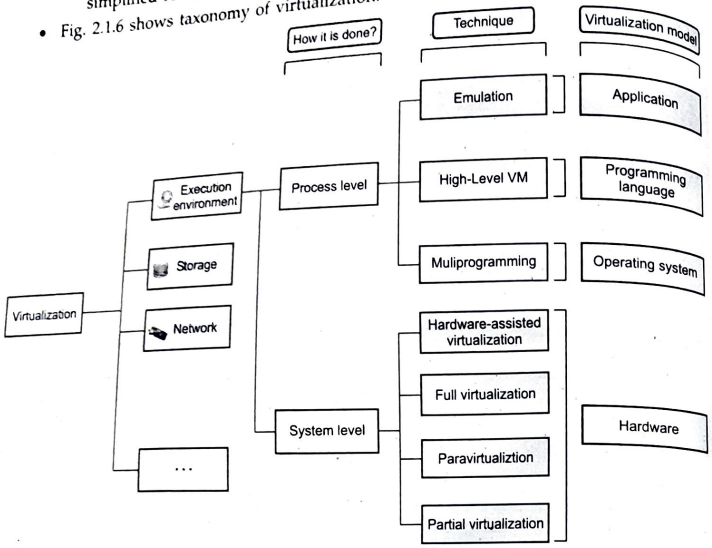


Fig. 2.1.6 Taxonomy of virtualization

- Virtualization is mainly used to emulate execution environment, storage and network. Execution environment classified into two types : process level and system level.
- Process level is implemented on top of an existing operating system.
- System level is implemented directly on hardware and do not or minimum requirement of existing operating system.

2.1.4.1 Advantages and Disadvantages

a) Pros

1. Data center and energy-efficiency savings : As companies reduce the size of their hardware and server footprint, they lower their energy consumption.
2. Operational expenditure savings : Once servers are virtualized, your IT staff can greatly reduce the ongoing administration and management of manual work.
3. Reduced costs : It reduced cost of IT infrastructure.

4. Data does not leak across virtual machine.
5. Virtual machine is completed isolated from host machine and other virtual machine.
6. Simplifies resource management by pooling and sharing resources.
7. Significantly reduce downtime.
8. Improved performance of IT resources.

b) Cons

1. Not all hardware or software can be virtualized.
2. Not all servers are applications are specifically designed to be virtualization - friendly.

2.1.5 Web Technology

- The World Wide Web (WWW) is an evolving system for publishing and accessing resources and services across the internet. Web is an open system. Its operations are based on freely published communication standards and documents standards. The Web is one with respect to the types of 'resource' that can be published and shared on.
- Fig. 2.1.7 shows the web servers and web browsers.

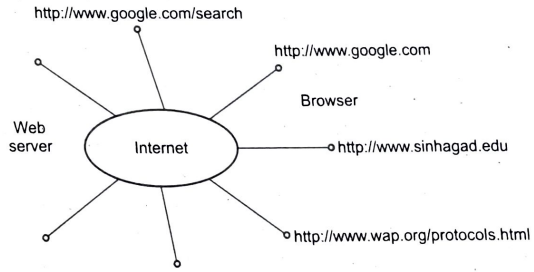


Fig. 2.1.7 Web servers and browsers

Key feature :

1. Web provides a hypertext structure among the documents that it stores.
2. The documents contain links i.e. references to other documents or resources. The structures of links can be arbitrarily complex and the set of resources that can be added is unlimited.

- The main standard components of Web :
  1. HyperText Markup Language (HTML)
  2. Uniform Resource Locators (URLs)
  3. HyperText Transfer Protocol (HTTP)
- HTML specifies the contents and layout of web pages. The content contains text, table, form, image, links, information for search engine, etc. The layout is in the form of text format, background and frame. HTML is also used to specify links and which resources are associated with them.
- URL identifies a resource to let browser find it. HTTP URL is mostly widely used today. An HTTP URL has two main jobs to do :
  1. To identify which web server maintains the resource
  2. To identify which of the resources at that server is required.
- HTTP defines a standard rule by which browsers and any other types of client interact with web servers. Main features are -
  1. Request-reply interaction
  2. Content types may or may not be handled by browser-using plug-in or external helper.
  3. One resource per request so several requests can be made concurrently.
  4. Simple access control : Any user with network connectivity to a web server can access any of its published resources.
- Three fundamental elements comprise the technology architecture of the Web :
  1. Uniform Resource Locators (URL) : A standard syntax used for creating identifiers that point to web-based resources, the URL is often structured using a logical network location.
  2. Hypertext Transfer Protocol (HTTP) : This is the primary communications protocol used to exchange content and data throughout the World Wide Web. URLs are typically transmitted via HTTP.
  3. Markup Languages (HTML, XML) : Markup languages provide a lightweight means of expressing Web - centric data and metadata. The two primary markup language are HTML and XML.

2.1.5.1 URL and HTTP

URL :

- The Uniform Resource Locator (URL) is a standard for specifying any kind of information on the Internet. Each URL uniquely identifies a page of information by giving the name of a remote computer, a server on that computer and a specific page of information available from the server.

- Fig. 2.1.8 shows how the URL encodes the information.

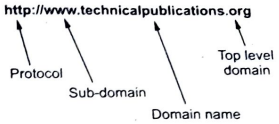


Fig. 2.1.8 URL

- URL has three parts :
  1. The protocol
  2. DNS name of the machine where the page is located.
  3. File name containing the page.
- The protocol is the client-server program used to retrieve the document. Host is the computer on which the information is located. The URL can optionally contain the port number of the server. File name gives where the information is located.
- The URL sends users to a specific resource online such as video, webpage, or other resources. When user search any query on Google, it will display the multiple URLs of the resource that are all related to your search query. The displayed URLs are the hyperlink to access the webpages.

HTTP :

- Hyper Text Transfer Protocol (HTTP) is standard web transfer protocol. HTTP is the set of rules governing the format and content of the conversation between a web client and server.
- The HTTP protocol is a request/response protocol based on the client / server - based architecture where web browsers, robots and search engines, etc. act like HTTP clients and the web server acts as a server.
- HTTP uses internet media types in the content-type and accept header fields in order to provide open and extensible data typing and type negotiation.
- HTTP uses language tags within the accept-language and content-language fields. It supports the proxy servers. A proxy server is a computer that keeps copies of responds to recent requests.
- The HTTP client sends a request to the proxy server. The proxy server checks its cache. If the response is not stored in the cache, the proxy server sends the request to the corresponding server.
- HTTP connections are of two types : Persistent HTTP and Non-persistent HTTP

Web Application :

- Web application is an application that is invoked with a web browser over the internet. Web applications can be defined as applications that are accessed over a network such as the internet or an Intranet. It is utilizing web browser technologies to accomplish one or more tasks over a network, typically through a web browser.
- Fig. 2.1.9 shows web application model.

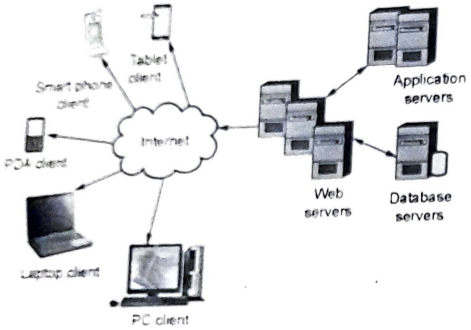


Fig. 2.1.9 Web application model

- Common web applications include webmail, online retail sales, online auctions, wikis and many other functions. Web applications software and database reside on a central server rather than being installed on the desktop system and is accessed over a network.
- Web applications commonly use a combination of server-side script (ASP, PHP, etc.) and client-side script (HTML, Javascript, etc.) to develop the applications.
- The client-side script deals with the presentation of the information while the server-side script deals with all the hard stuff like storing and retrieving the information.

2.1.5.2 Web Services

In 2000, the W3C accepted a submission for the Simple Object Access Protocol (SOAP). This XML-based messaging format established a transmission framework for inter-application communication via HTTP.

- As a vendor-neutral technology, SOAP provided an attractive alternative.

(see Fig. 2.1.10 on next page)

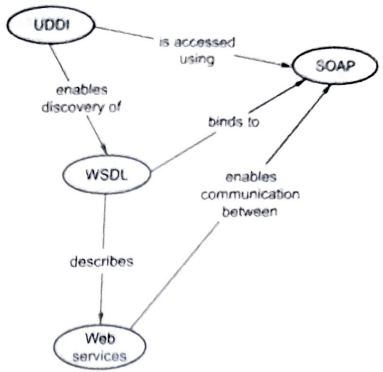


Fig. 2.1.10

- During the following year, the W3C published the WSDL specification. Another implementation of XML, this standard supplied a language for describing the interface of web services.
- Further supplemented by the Universal Discovery, and Integration (UDDI) specification that provided a standard mechanism for the dynamic discovery of services platform had been established.
- Since then, Web services have been adopted by vendors and manufacturers at a remarkable pace. Industry-wide support furthered the popularity and importance of this platform and of service-oriented design principles. This led to the creation of a second generation of web services specifications.
- Web services currently provide the main enabling technique for Service Oriented Architecture. The Web service technique can function both as a middelware and a modeling and management tool for composed business processes.
- Web-based application that dynamically interact with other Web applications using open standards that include XML, UDDI and SOAP.
- Web services protocols and standards are the technology that promotes the sharing and distribution of information and business data.
- A protocol is a standard method for transmitting data through a network. There are many specialized protocols to accommodate the many kinds of data that might be transmitted.
- Web services publish the details of their functions and interfaces, but they keep their implementaton detail confidential.

- Thus a client and a service that support common communication protocols can interact regardless of the platforms on which they run, or the programming language in which are written. This makes web services particularly applicable to a distribute heterogeneous environment.
- Key specifications used by web services are :
  1. XML schemas convey the data syntax and semantics for various application domain, such as business-to-business transaction, medical records and production status reports.
  2. SOAP is a simple XML-based protocol to let applications exchange information over HTTP. SOAP is a lightweight protocol for exhange of information in a decentralized, distributed environment.
  3. WSDL (Web Services Description Language) is an XML - based language for describing web services and how to access them.

2.1.6 Difference between Cloud Application and Web Application

Sr. No.	Cloud application	Web application
1.	All cloud applications are web applications.	Not all web applications are cloud applications.
2.	Inherently scalable	Limited by scalability
3.	Very high uptime.	Limited by availability
4.	Multi - tenancy solution.	Isolated - tenancy solution
5.	The provided application is standardized for all customers.	Each customer uses its own instance of the application.
6.	User data and business process store in a multiple replicated data centers.	User data and business process store in single data center.
7.	The cloud applications can be installed on a public cloud or a private cloud and accessed there.	The web applications can be installed on internet or intranet and accessed there.

2.1.7 Multitenant Technology

- A multi - tenant cloud is a cloud computing architecture that allows customers to share computing resources in a public or private cloud. Each tenant's data is isolated and remains invisible to other tenants.
- It allows multiple users to work in a software environment at the same time, each with their own separate user interface, resources and services. The multitenant application design was created to enable multiple users (tenants) to access the same application logic simultaneously.

- Multitenancy can describe hardware or software architectures in which multiple systems, applications, or data from different enterprises are hosted on the same physical hardware.
- Multitenant applications typically include a level of customization for tenants, such as customizing the look and feel of the application or allowing the tenant to decide on specific access control permissions and restrictions for users.
- "Tenants" is a term for a group of users or software applications that all share access to the hardware through the underlying software. Multiple tenants on a server all share the memory, which is dynamically allocated and cleaned up as needed. They also share access to system resources, such as the network controller.
- Fig. 2.1.11 shows multi-tenant technology.

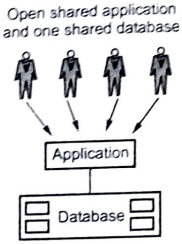


Fig. 2.1.11 Multi-tenant technology

- Multi-tenant architecture is to offer shared tenancy on public cloud providers like Amazon Web Services, Microsoft Azure and Google Cloud.
- Tenants can individually customize features of the application, such as :
  1. **User Interface :** Tenants can define a specialized look for their application interface.
  2. **Business Process :** Tenants can customize the rules, logic, and workflows of the business processes that are implemented in the application.
  3. **Data Model :** Tenants can extend the data schema of the application to include, exclude, or rename fields in the application data structures.
  4. **Access Control :** Tenants can independently control the access rights for users and groups.
- Common characteristics of multitenant applications are as follows :
  1. **Usage Isolation -** The usage behaviour of one tenant does not affect the application availability and performance of other tenants.

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

Benefits of a Multitenancy technology :

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

2.1.8 Service Technology

- Following core technology are used in web services :
  1. **Web Service Description Language (WSDL) :** This markup language is used to create a WSDL definition that defines the API of a Web service, including its individual operations and each operation's input and output messages.
  2. **XML Schema Definition Language (XML Schema) :** Messages exchaned by web services must be expressed using XML. XML schemas are created to defined the data structure of the XML-based input and output messages exchanged by web services.
  3. **SOAP (Simple Object Access Protocol) :** It define a common messaging format used for request and response exhanged by web services. SOAP messages are comprosed of body and header sections.

4. **Universal Description, Discovery and Integration (UDDI) :** This standard regulates service registries in which WSDL defintions can be published as part of service catalog for discovery purposes.

REST Service :

- REST means REpresentational State Transfer. REST is a term coined by ROY Fielding to describe an architecture style of networked systems.
- REST is a set of design criteria and not the physical structure (architecture) of the system. It is not tied to the 'web' i.e. doesn't depend on the mechanics of HTTP.
- Client-Server : A pull based interaction style(Client request data from servers as and when needed).
- Stateless : Each request from client to server must contain all the information necessary to understand the request and cannot take advantage of any stored context on the server.
- Cache : To improve network efficiency, responses must be capable of being labeled as cachable or non-cacheable.
- Uniform interface : All resources are accessed with a generic interface (e.g. HTTP, GET, POST, PUT, DELETE).
- Named resources : The system is comprised of resources which are named using a URL.
- Interonnected resource representations : The representations of the resources are interconnected using URLs, thereby enabling a client to prgress from one state to another.
- Characteristics of a REST based network :
  1. **Client Server :** A pull-based interaction styke (client request data from servers as and when needed)
  2. **Stateless :** Each request from client to server must contain all the information necessary to understand the request, and cannot take advantage of any stored context on the server.
  3. **Cache :** To improve network efficiency, reponses must be capable of being labeled as cacheable or non-cacheable.
  4. **Uniform interface :** All resources are accessed with a generic interface (e.g. HTTP GET, POST, PUT, DELETE).
  5. **Named resources :** The system is comprised of resources which are named using a URL.

6. **Interconnected resource representations :** The representations of the resources are interconnected using URLs, thereby enabling a client to progress from one state to another.

Benefits of REST

1. REST allows a greater variety of data formats, whereas SOAP only allows XML.
2. REST provides superior performance, particularly through caching for information that's not altered and dynamic.
3. It is the protocol used most often for major services such as Yahoo, Ebay, Amazon and even Google.
4. REST is generally faster and uses less bandwidth.
5. It's also easier to integrate with existing websiites with no need to refactor site infrastructure.

Review Questions

1. *Draw and explain relation of first-generation web service technologies.* **SPPU : April-18 In Sem, Marks 6**
2. *State and describe any four cloud enabling technologies.* **SPPU : Dec.-18 End Sem, Marks 8**
3. *Write short note on multitenant technology.* **SPPU : Dec.-19 End Sem, Marks 4**
4. *Write short note on : Web technology.* **SPPU : March-20, In Sem, Marks 4**
5. *List cloud enabling technologies. Explain any two in details.* **SPPU : May-19 End Sem, Marks 4**

2.2 Implementation Levels of Virtualization **SPPU : April-18, 19, March-20**

- Virtualization is implemented at various levels :
  1. Instruction set architecture level
  2. Hardware abstraction level
  3. Operating system level
  4. Library support level
  5. User application level

2.2.1 Instruction Set Architecture Level

- The definition of the storage resources and the instructions that manipulate data are documented in what is referred to as Instruction Set Architecture (ISA).

- ISA view of a machine corresponds to the machine and assembly language levels. For example, MIPS binary code can run on an x86-based host machine with the help of ISA emulation.
- Instruction set emulation leads to virtual ISAs created on any hardware machine. The basic emulation method is through code interpretation. An interpreter program interprets the source instructions to target instructions one by one.
- The key to virtualize a CPU lies in the execution of the guest instruction, including both system-level and user-level instructions virtualizing a CPU can achieved in one two ways :
  1. Emulation : The only processor vitalization mechanism available when the ISA of the guest is different from the ISA of the host.
  2. Direct native execution : Possible only if the ISA of the host is identical to the ISA of the guest.
- Fig. 2.2.1 shows ISA emulation.

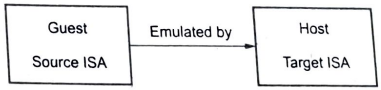


Fig. 2.2.1 ISA emulation

- Emulation is the process of implementing the interface and functionality of one system (or subsystem) on a system (or subsystem) having different interface and functionality.
- In other words, emulation allows a machine implementing one ISA (the target), to reproduce the behavior of a software compiled for another ISA (the source). Emulation can be carried out using :
  1. Interpretation
  2. Binary traslation

2.2.2 Hardware Abstraction Level

- This type of virtualization is performed right on top of the bare hardware. On the hand, this approach generates a virtual hardware environment for a VM. On the other hand, the process manages the underlying hardware through virtualization.
- The idea is to virtualizes a computer's resources, such as processors memory, and I/O devices. The intention is to upgrade the hardware utilization rate by multiple users concurrently.
- The Xen hypervisor has been applied to virtualize x86-based machines to run Linux or other guest OS applications.

2.2.3 Operating System Level Virtualization

- Operating-system-level virtualization is a server-virtualization method where the kernel of an operating system allows for multiple isolated user-space instances, instead of just one. Such instances, which are sometimes called containers and software containers.
- This refers to an abstraction layer between traditional OS and user applications.
- This type of virtualization creates isolated containers on a single physical server and the OS instances to utilize the hard-ware and software in data centers.
- Containers behave like real servers. With containers you can create a portable, consistent operating environment for development, testing, and deployment.
- This virtualization creates virtual hosting environments to allocates hardware resources among a large number of mutually distrusting users.
- Operating-system-level virtualization usually imposes little to no overhead, because programs in virtual partitions use the operating system's normal system call interface and do not need to be subjected to emulation or be run in an intermediate virtual machine.
- Operating system-level virtualization is not as flexible as other virtualization approaches since it cannot host a guest operating system different from the host one, or a different guest kernel.
- Instead of trying to run an entire guest OS, container virtualization isolates the guests, but doesn't try to virtualize the hardware. Instead, you have containers for each virtual environment.
- With container-based technologies, you'll need a patched kernel and user tools to run the virtual environments. The kernel provides process isolation and performs resource management.

Why operating system level virtualization is required ?

- Operating system level virtualization provides feasible solution for hardware level virtualization issue. It inserts a virtualization layer inside an operating system to partition a machine's physical resources.
- It enables multiple isolated VMs within a single operating system kernel. This kind of VM is often called a virtual execution environment (VE), Virtual Private System (VPS), or simply container.
- From the user's point of view, virtual execution environment look like real servers.

- This means a virtual execution environment has its own set of processes, file system, user accounts, network interfaces with IP addresses, routing tables, firewall rules etc.
- Although VEs can be customized for different people, they share the same operating system kernel. Therefore, OS-level virtualization is also called single-OS image virtualization.

Challenges to cloud computing in OS level virtualization ?

- Cloud computing is transforming the computing landscape by shifting the hardware and staffing costs of managing a computational center to third parties.
- Cloud computing has at least two challenges :
  1. The ability to use a variable number of physical machines and virtual machine instances depending on the needs of a problem. For example, a task may need only a single CPU during some phases of execution but may need hundreds of CPUs at other times.
  2. It is related to slow operation of instantiating new virtual machine. Currently, new virtual machines originate either as fresh boots or as replicates of a template VM, unaware of the current application state. Therefore, to better support cloud computing, a large amount of research and development should be done.

Advantages of OS virtualization :

1. OS virtualization provide least overhead among all types of virtualization solution.
2. They offer highest performance and highest density of virtual environment.
3. Low resource requirements.
4. High Scalability.

Disadvantage of OS virtualization :

1. They support only one operating system as base and guest OS in a single server.
2. It supports library level virtualization.

2.2.4 Library Support Level

- Library-level virtualization is also known as user-level Application Binary Interface (ABI).
- This type of virtualization can create execution environments for running alien programs on a platform rather than creating a VM to run the entire operating system.
- It is done by API call interception and remapping.

- Virtualization with library interfaces is possible by controlling the communication link between applications and the rest of a system through API hooks.
- Example : Wine, WAB, LxRun, Visual MainWin
- Advantage : It has very low implementation effort
- Shortcoming and limitation : Poor application flexibility and isolation.

2.2.5 User Application Level

- Virtualization at the application level virtualizes an application as a VM. On a traditional OS, an application often runs as a process. Therefore, application-level virtualization is also known as process-level virtualization.
- A fully virtualized application is not installed in the traditional sense, although it is still executed as if it were. The application behaves at runtime like it is directly interfacing with the original operating system and all the resources managed by it, but can be isolated to varying degrees.
- Full application virtualization requires a virtualization layer. Application virtualization layers replace part of the runtime environment normally provided by the operating system.
- The layer intercepts all disk operations of virtualized applications and transparently redirects them to a virtualized location, often a single file.
- The application remains unaware that it accesses a virtual resource instead of a physical one. Since the application is now working with one file instead of many files spread throughout the system, it becomes easy to run the application on a different computer and previously incompatible applications can be run side-by-side.
- The most popular approach is to deploy High Level Language (HLL) VMs. Here the virtualization layer sits as an application program on top of the operating system, and the layer exports an abstraction of a VM that can run programs written and compiled to a particular abstract machine definition. Any program written in the HLL and compiled for this VM will be able to run on it.
- **Benefits :**
  1. Application virtualization uses fewer resources than a separate virtual machine.
  2. Application virtualization also enables simplified operating system migrations.
  3. Applications can be transferred to removable media or between computers without the need of installing them, becoming portable software.
- **Limitations :**
  1. Not all computer programs can be virtualized
  2. Lower performance

Review Questions

1. *Explain different abstraction levels of virtualization.* **SPPU : April-18 In Sem, Marks 6**
2. *Draw the diagram of 'Two level memory mapping' with reference to memory virtualization.* **SPPU : April-19 In Sem, Marks 3**
3. *Explain levels of virualization.* **SPPU : April-19 In Sem, Marks 4**
4. *Draw and explain implementation level of virtualization.* **SPPU : April-19 In Sem, Marks 6**

2.3 Virtualization Structures/Tools and Mechanisms

SPPU : May-18,19, April-19, March-20

2.3.1 XEN Architecture

- Xen is a type 1 hypervisor that creates logical pools of system resources so that many virtual machines can share the same physical resources.
- Xen is a hypervisor that runs directly on the system hardware. It inserts a virtualization layer between the system hardware and the virtual machines, turning the system hardware into a pool of logical computing resources that Xen can dynamically allocate to any guest operating system.
- Fig. 2.3.1 shows Xen architecture.

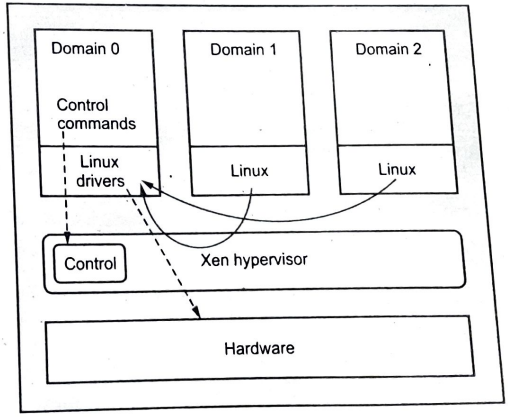


Fig. 2.3.1 Xen architecture

- The operating systems running in virtual machines interact with the virtual resources as if they were physical resources. Xen provides a virtual environment located between the hardware and the OS.
- Xen doesn't include any device drivers; it provides a mechanism by which a guest-OS can have direct access to the physical devices.
- The core components of Xen are the hypervisor, kernel and applications. Many guest operating systems can run on the top of the hypervisor; but it should be noted that one of these guest OS controls the others.
- This guest OS with the control ability is called Domain 0 , the others are called Domain U. Domain 0 is first loaded when the system boots and can access the hardware directly and manage devices by allocating the hardware resources for the guest domains (Domain U).
- The Control Domain (or Domain 0) is a specialized Virtual Machine that has special privileges like the capability to access the hardware directly, handles all access to the system's I/O functions and interacts with the other Virtual Machines.
- It also exposes a control interface to the outside world, through which the system is controlled. The Xen Project hypervisor is not usable without Domain 0, which is the first VM started by the system.

2.3.2 VMware ESXi

- VMware ESXi (formerly ESX) is an enterprise-class, type-1 hypervisor developed by VMware for deploying and serving virtual computers.
- As a type-1 hypervisor, ESXi is not a software application that is installed on an operating system; instead, it includes and integrates vital OS components, such as a kernel.
- VMware Inc. developed ESX and ESXi as bare metal embedded hypervisors, which means that they run directly on server hardware and do not require the installation of an additional underlying operating system.
- This virtualization software creates and runs its own kernel, which is run after a Linux kernel bootstraps the hardware. The resulting service is a microkernel, which has three interfaces : Hardware, Guest system and Console operating system.
- VMware ESXi server is computer virtualization software developed by VMware Inc. The ESXi server is an advanced, smaller-footprint version of the VMware ESX server.

Components of ESXi

- Fig. 2.3.2 shows architecture of ESXi.

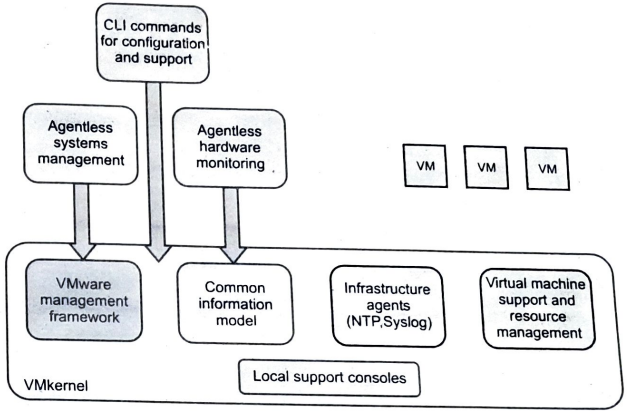


Fig. 2.3.2 ESXi architecture

- The VMware ESXi architecture comprises the underlying operating system, called VMkernel, and processes that run on top of it.
- VMkernel provides means for running all processes on the system, including management applications and agents as well as virtual machines.
- It has control of all hardware devices on the server, and manages resources for the applications. The main processes that run on top of VMkernel are :
  1. Direct Console User Interface (DCUI) : The low-level configuration and management interface, accessible through the console of the server, used primarily for initial basic configuration.
  2. The virtual machine monitor, which is the process that provides the execution environment for a virtual machine, as well as a helper process known as VMX. Each running virtual machine has its own VMM and VMX process.
  3. Various agents used to enable high-level VMware infrastructure management from remote applications.
  4. The Common Information Model (CIM) system : CIM is the interface that enables hardware-level management from remote applications via a set of standard APIs.

2.4 Hypervisors

SPPU : April-18, Dec.-19

- In computing, a hypervisor is a virtualization platform that allows multiple operating systems to run on a host computer at the same time. The term usually refers to an implementation using full virtualization.
- A hypervisor is a software layer installed on the physical hardware, which allows splitting the physical machine into many virtual machines. This allows multiple operating systems to be run simultaneously on the same physical hardware.
- The operating system installed on the virtual machine is called a guest OS, and is sometimes also called an instance. The hardware the hypervisor runs on is called the host machine.
- A hypervisor management console, which is also called a Virtual Machine Manager (VMM), is computer software that enables easy management of virtual machines.
- Hypervisors are currently classified in two types : type 1 and type 2

2.4.1 Type 1

- Type 1 hypervisor is software that runs directly on a given hardware platform. A "guest" operating system thus runs at the second level above the hardware.
- Fig. 2.4.1 shows Type 1 hypervisor.

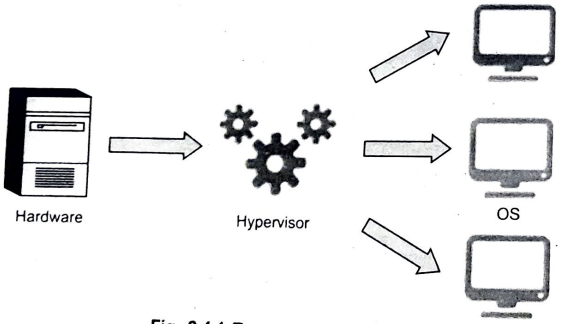


Fig. 2.4.1 Type 1 hypervisor

- Type 1 VMs have no host operating system because they are installed on a bare system. An operating system running on a Type 1 VM is a full virtualization because it is a complete simulation of the hardware that it is running on.
- Type 1 hypervisor is also called a native or bare-metal hypervisor that is installed directly on the hardware, which splits the hardware into several virtual machines where we can install guest operating systems.

- Virtual machine management software helps to manage this hypervisor, which allows guest OSes to be moved automatically between physical servers based on current resources requirements.
- It is completely independent from the Operating System.
- The hypervisor is small as its main task is sharing and managing hardware resources between different operating systems.
- A major advantage is that any problems in one virtual machine or guest operating system do not affect the other guest operating systems running on the hypervisor.

2.4.2 Type 2 Hypervisor

- This is also known as Hosted Hypervisor.
- In this case, the hypervisor is installed on an operating system and then supports other operating systems above it.
- It is completely dependent on host Operating System for its operations. Fig. 2.4.2 shows type 2 hypervisor.

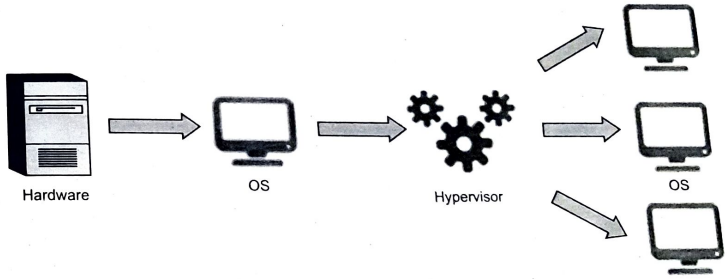


Fig. 2.4.2 Type 2 Hypervisor

- While having a base operating system allows better specification of policies, any problems in the base operating system affects the entire system as well even if the hypervisor running above the base OS is secure.
- Type 2 hypervisors don't support over/dynamic allocation of RAM, so care is required when allocating resources to virtual machines.
- This is why we call type 2 hypervisors hosted hypervisors. As opposed to type 1 hypervisors that run directly on the hardware, hosted hypervisors have one software layer underneath. What we have in this case is :
  1. A physical machine.

2. An operating system installed on the hardware (Windows, Linux, MacOS).
3. A type 2 hypervisor software within that operating system.
4. The actual instances of guest virtual machines.

- Type 2 hypervisors are typically found in environments with a small number of servers. Type 2 hypervisors are convenient for testing new software and research projects.

2.4.3 Para-virtualization

- Para-virtualization is a type of virtualization in which a guest operating system (OS) is recompiled, installed inside a virtual machine (VM), and operated on top of a hypervisor program running on the host OS.
- Para-virtualization refers to communication between the guest OS and the hypervisor to improve performance and efficiency.
- Para-virtualization involves modifying the OS kernel to replace non-virtualizable instructions with hyper-calls that communicate directly with the virtualization layer hypervisor.
- The hypervisor also provides hyper-call interfaces for other critical kernel operations such as memory management, interrupt handling and time keeping.
- Fig. 2.4.3 shows para-virtualization architecture.

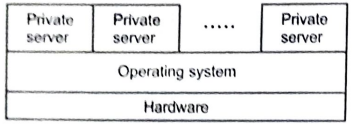


Fig. 2.4.3 Para-virtualization architecture

- In Para-virtualization, the virtual machine does not necessarily simulate hardware, but instead offers a special API that can only be used by modifying the "guest" OS. This system call to the hypervisor is called a "hypercall" in Xen.
- Xen is an open source para-virtualization solution that requires modifications to the guest operating systems but achieves near native performance by collaborating with the hypervisor.
- Microsoft Virtual PC is a para-virtualization virtual machine approach. User-mode Linux (UML) is another para-virtualization solution that is open source.
- Each guest operating system executes as a process of the host operating system. Cooperative Linux, is a virtualization solution that allows two operating systems to cooperatively share the underlying hardware.

- Linux-V server is an operating system-level virtualization solution for GNU/Linux systems with secure isolation of independent guest servers.
- The Linux KVM is virtualization technology that has been integrated into the mainline Linux kernel . Runs as a single kernel loadable module, a Linux kernel running on virtualization-capable hardware is able to act as a hypervisor and support unmodified Linux and Windows guest operating systems.
- Para-virtualization shares the process with the guest operating system.

Problems with para-virtualization

1. Para-virtualized systems won't run on native hardware
2. There are many different para-virtualization systems that use different commands, etc.

- The main difference between full virtualization and paravirtualization in Cloud is that full virtualization allows multiple guest operating systems to execute on a host operating system independently while paravirtualization allows multiple guest operating systems to run on host operating systems while communicating.

2.4.4 Difference between Type 1 and Type 2 Hypervisor

Type 1 Hypervisor	Type 2 Hypervisor
This is also known as Bare Metal or Embedded or Native Hypervisor	This is also known as Hosted Hypervisor
It is completely independent from the Operating System	It is completely dependent on host Operating System for its operations
It works directly on the hardware of the host and can monitor operating systems that run above the hypervisor	In this case, the hypervisor is installed on an operating system and then supports other operating systems above it
It support hardware virtualization	It support OS virtualization
Examples : VMware ESXi Server and Microsoft Hyper-V	Examples : Vmware Workstation, Microsoft Virtual PC, Oracle Virtual Box
Higher performance and scalability because of being bare metal type	Low performance as a result of host operating system overhead

Review Questions

1. *Explain different types of hypervisors with example.* **SPPU : April-18 In Sem, Marks 6**
2. *Explain and differentiate types of hypervisor.* **SPPU : Dec.-19 End Sem, Marks 5**
3. *Explain in brief about para-virtualization.* **SPPU : April-18, In Sem, Marks 4**

2.5 Full Virtualization

SPPU : April-19

- Full Virtualization doesn't need to modify the host OS; it relies upon binary translation to trap and to virtualize certain sensitive instructions.
- Fig. 2.5.1 shows full virtualization.

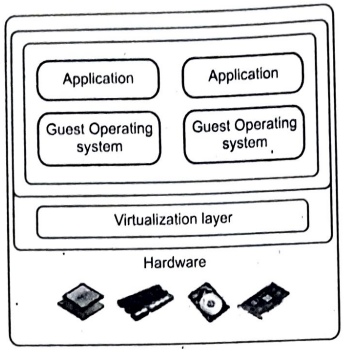


Fig. 2.5.1 Full virtualization

- VMware Workstation applies full virtualization, which uses binary translation to automatically modify x86 software on-the-fly to replace critical instructions
- Normal instructions can run directly on the host OS. This is done to increase the performance overhead - normal instructions are carried out in the normal manner, but the difficult and precise executions are first discovered using a trap and executed in a virtual manner.
- This is done to improve the security of the system and also to increase the performance.

Host based virtualization :

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

- **Advantages** of host-based architecture :
  1. The user can install this VM architecture without modifying the host OS.
  2. The host-based approach appeals to many host machine configurations.

2.5.1 Memory Virtualization

- Memory virtualization features allow abstraction isolation and monitoring of memory on a per Virtual Machine (VM) basis. These features may also make live migration of VMs possible, add to fault tolerance, and enhance security.
- Example features include Direct Memory Access (DMA) remapping and Extended Page Tables (EPT), including their extensions: accessed and dirty bits, and fast switching of EPT contexts.
- The VMkernel manages all machine memory. The VMkernel dedicates part of this managed machine memory for its own use. The rest is available for use by virtual machines.
- Virtual machines use machine memory for two purposes : each virtual machine requires its own memory and the VMM requires some memory and a dynamic overhead memory for its code and data.
- The virtual memory space is divided into blocks, typically 4 KB, called pages. The physical memory is also divided into blocks, also typically 4 KB.
- When physical memory is full, the data for virtual pages that are not present in physical memory are stored on disk. ESX/ESXi also provides support for large pages.
- The VMM is responsible for mapping the guest physical memory to the actual machine memory.
- Each page table of a guest OS has a page table allocated for it in the VMM. The page table in the VMM which handles all these is called a shadow page table.
- As it can be seen all this process is nested and inter-connected at different levels through the concerned address.
- If any change occurs in the virtual memory page table or TLB, the shadow page table in the VMM is updated accordingly.

2.5.2 I/O Virtualization

- I/O Virtualization involves managing of the routing of I/O requests between virtual devices and shared physical hardware.
- The there are three ways to implement this are full device emulation, para-VZ and direct I/O.

- I/O virtualization features facilitate offloading of multi-core packet processing to network adapters as well as direct assignment of virtual machines to virtual functions, including disk I/O.
- Examples include Virtual Machine Device Queues (VMDQ), Single Root I/O Virtualization.
- Fig. 2.5.2 shows I/O virtualization.

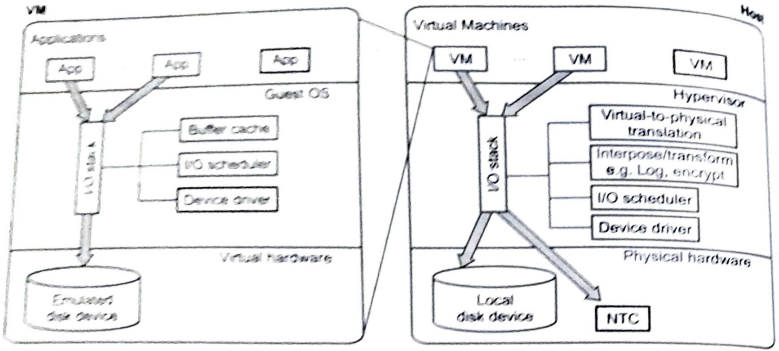


Fig. 2.5.2 I/O virtualization

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

2.5.3 Difference between Full and Para Virtualization

Sr. No.	Full Virtualization	Para Virtualization
1.	Full Virtualization relies upon binary translation to trap and to virtualize certain sensitive instructions. Example : VMware	Para-Virtualization refers to communication between the guest OS and the hypervisor to improve performance and efficiency. Example : Xen architecture
2.	Full Virtualization doesn't need to modify the host OS.	Para-Virtualization involves modification of OS kernel.
3.	Normal instructions can run directly on the host OS.	Para-virtualized systems won't run on native hardware.
4.	Full Virtualization uses binary translation and direct execution.	Para-Virtualization uses hyper - calls.
5.	Performance is good.	Performance is better in certain cases.
6.	Guest software does not require any modification since the underlying hardware is fully simulated.	Hardware is not simulated and the guest software run their own isolated domains.

2.5.4 Virtualization of CPU

- Certain processors such as Intel VT provide hardware assistance for CPU virtualization.
- When using this assistance, the guest can use a separate mode of execution called guest mode. The guest code, whether application code or privileged code, runs in the guest mode.
- On certain events, the processor exits out of guest mode and enters root mode. The hypervisor executes in the root mode, determines the reason for the exit, takes any required actions, and restarts the guest in guest mode.
- When you use hardware assistance for virtualization, there is no need to translate the code. As a result, system calls or trap-intensive workloads run very close to native speed.
- Some workloads, such as those involving updates to page tables, lead to a large number of exits from guest mode to root mode. Depending on the number of such exits and total time spent in exits, this can slow down execution significantly.
- CPU virtualization features enable faithful abstraction of the full prowess of Intel CPU to a virtual machine.
- All software in the VM can run without any performance, as if it was running natively on a dedicated CPU. Live migration from one Intel CPU generation to another, as well as nested virtualization, is possible.

2.5.5 Binary Translation with Full Virtualization

- This approach relies on binary translation to trap and to virtualize certain sensitive and non-virtualizable instructions with new sequences of instructions that have the intended effect on the virtual hardware. Meanwhile, user level code is directly executed on the processor for high performance virtualization.
- Fig. 2.5.3 shows full virtualization with binary translation.

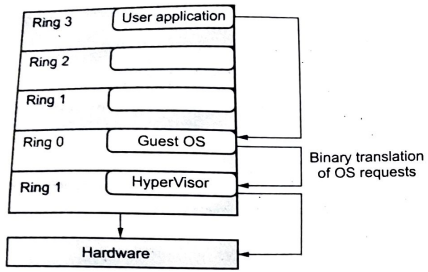


Fig. 2.5.3

- This combination of binary translation and direct execution provides full virtualization as the guest OS is completely decoupled from the underlying hardware by the virtualization layer.
- The guest OS is not aware it is being virtualized and requires no modification.
- The hypervisor translates all operating system instructions at run-time on the fly and caches the results for future use, while user level instructions run unmodified at native speed.
- VMware's virtualization products such as VMWare ESXi and Microsoft Virtual Server are examples of full virtualization.
- The performance of full virtualization may not be ideal because it involves binary translation at run-time which is time consuming and can incur a large performance overhead.

Review Question

1. *Explain binary translation with full virtualization.* **SPPU : April-19 In Sem, Marks 4**

2.6 Virtual Clusters and Resource Management

SPPU : May-18

- A computer cluster is a set of connected computers (nodes) that work together as if they are a single machine. All processor machines share resources such as a common home directory and have a software such as a common home directory and havea software such as a Message Passing Interface (MPI) implmentation installed to allow programs to be run across all nodes simultaneously.
- Computer clusters are often used for cost-effective High performance Computing (HPC) and High Availability (HA) by businesses of all sizes. A computer cluster help to solve complex operations more efficiently with much faster porcessing speed, better data intergrity than a single computer and they only used for mission-critical applications.

Characteristics Virtual Cluster :

1. Virtual machine or physical machine is used as virtual cluster nodes. Multiple VM running with different types of OS can be deployed on the same physical node.
2. Virtual machine runs with guest operating system. Host OS and VM OS are different but it manages the resources in the physical machine.
3. Virtual machine can be replicated in multiple servers and it support distributed parallelism, fault tolerance, and disaster recovery.
4. Number of nodes of a virtual cluster may change accordingly.
5. If virtual machine failes, it can not affect the host machine.

- **Virtual cluster is managed by four ways :**
  1. We can use a guest-based manager, by which the cluster manager resides inside a guest OS. Ex: A Linux cluster can run different guest operating systems on top of the Xen hypervisor.
  2. We can bring out a host-based manager which itself is a cluster manager on the host systems. Ex : VMware HA (High Availability) system that can restart a guest system after failure.
  3. An independent cluster manager, which can be used on both the host and the guest - making the infrastructure complex.
  4. Finally, we might also use an integrated cluster (manager), on the guest and host operating systems; here the manager must clearly distinguish between physical and virtual resources.

Virtual machine states :

1. Inactive State : This is defined by the VZ platform, under which the VM is not enabled.

2. Active State : This refers to a VM that has been instantiated at the VZ platform to perform a task.
3. Paused State: VM has been instantiated but disabled temporarily to process a task or is in a waiting state itself.
4. Suspended State : A VM enters this state if its machine file and virtual resources are stored back to the disk

Live migration steps :

- Steps 0 and 1 : Start migration automatically and checkout load balances and server consolidation.
- Step 2 : Transfer memory (transfer the memory data + recopy any data that is changed during the process). This goes on iteratively till changed memory is small enough to be handled directly.
- Step 3 : Suspend the VM and copy the last portion of the data.
- Steps 4 and 5 : Commit and activate the new host. Here, all the data is recovered, and the VM is started from exactly the place where it was suspended, but on the new host

File System Migration :

- To support VM migration from one cluster to another, a consistent and location-dependent view of the file system is available on all hosts.
- Each VM is provided with its own virtual disk to which the file system is mapped to. The contents of the VM can be transmitted across the cluster by inter-connections (mapping) between the hosts.
- But migration of an entire host is not advisable due to cost and security problems. We can also provide a global file system across all host machines where a VM can be located.
- This methodology removes the need of copying files from one machine to another, all files on all machines can be accessed through network.
- It should be noted here that the actual files are not mapped or copied. The VMM accesses only the local file system of a machine and the original/modified files are stored at their respective systems only.
- This decoupling improves security and performance but increases the overhead of the VMM – every file has to be stored in virtual disks in its local files.

Pre copy and post copy of live migration :

- In pre copy, which is manly used in live migration, all memory pages are first transferred; it then copies the modified pages in the last round iteratively.

- Here, performance 'degradation' will occur because migration will be encountering dirty pages all around in the network before getting to the right destination.
- The iterations could also increase, causing another problem. To encounter these problems, check-pointing/recovery process is used at different positions to take care of the above problems and increase the performance.
- In post-copy, all memory pages are transferred only once during the migration process. The threshold time allocated for migration is reduced. But the downtime is higher than that in pre-copy.

Review Question

1. *What is live VM migration ? Write down the steps required for live VM migration.*

SPPU : May-18 End Sem, Marks 6

2.7 Virtualization for Data-Center Automation

2.7.1 Server Consolidation in Data Centers

- The heterogeneous workloads in the data center is divided into two categories : chatty workloads and noninteractive workloads.
- Chatty workloads may burst at some point and return to a silent state at some other point. For example, video services can be used by a lot of people at night and few people use it during the day.
- Noninteractive workloads do not require people's efforts to make progress after they are submitted. Server consolidation is an approach to improve the low utility ratio of hardware resources by reducing the number of physical servers.
- The use of VMs increases resource management complexity.
- It enhances hardware utilization. Many underutilized servers are consolidated into fewer servers to enhance resource utilization. Consolidation also facilitates backup services and disaster recovery.
- In a virtual environment, the images of the guest OSes and their applications are readily cloned and reused.
- Total cost of ownership is reduced
- Improves availability and business continuity
- Automation of data-center operations includes resource scheduling, architectural support, power management, automatic or autonomic resource management, performance of analytical models, and so on.

- In virtualized data centers, an efficient, on-demand, fine-grained scheduler is one of the key factors to improve resource utilization.
- Dynamic CPU allocation is based on VM utilization and application-level QoS metrics.
- One method considers both CPU and memory flowing as well as automatically adjusting resource overhead based on varying workloads in hosted services.
- Another scheme uses a two-level resource management system to handle the complexity involved. A local controller at the VM level and a global controller at the server level are designed.
- Three resource managers are as follows :
  1. Instance Manager controls the execution, inspection, and terminating of VM instances on the host where it runs.
  2. Group Manager collects an information about schedules VM execution on specific instance managers and it manages virtual instance network.
  3. Cloud Manager is the entry-point into the cloud for users and administrators. It queries node managers for information about resources, makes scheduling decisions, and implements them by making requests to group managers,

2.7.2 Trust Management in Virtualized Data Centers

- Virtual machine in the host machine entirely encapsulates the state of the guest operating system running inside it.
- Encapsulated machine state can be copied and shared over the network and removed like a normal file, which proposes a challenge to VM security.
- In general, a VMM can provide secure isolation and a VM accesses hardware resources through the control of the VMM, so the VMM is the base of the security of a virtual system.
- Normally, one VM is taken as a management VM to have some privileges such as creating, suspending, resuming, or deleting a VM.

1.VM-Based Intrusion Detection

- Intrusion detection is a set of techniques and methods that are used to detect suspicious activity both at the network and host level. Intrusion Detection System is software, hardware or combination of both used to detect intruder activity.
- Fig. 2.7.1 shows IDS. (See Fig. 2.7.1 on next page)
- A lightweight intrusion detection system can easily be deployed on most any node of a network, with minimal disruption to operations. Snort is a libpcap based packet sniffer and logger that can be used as a lightweight network intrusion detection system.

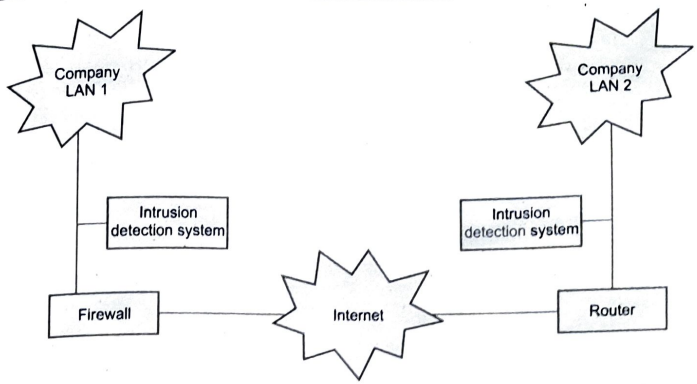


Fig. 2.7.1 IDS

- IDSs serve three essential security functions; monitor, detect and respond to unauthorized activity.
- Functions of intrusion detection systems
  1. IDS monitor and do analysis of user and system activity.
  2. Auditing of system configurations and vulnerabilities.
  3. Assessing the integrity of critical system and data files.
  4. Recognition of activity patterns reflecting known attacks.
  5. Statistical analysis for abnormal activity patterns.

Benefits of intrusion detection

1. Improving integrity of other parts of the information security infrastructure.
2. Improved system monitoring.
3. Tracing user activity from the point of entry to point of exit or impact.
4. Recognizing and reporting alterations to data files.
5. Spotting errors of system configuration and sometimes correcting them.
6. Recognizing specific types of attack and alerting appropriate staff for defensive responses.
7. Keeping system management personnel up to date on recent corrections to programs.
8. Allowing non-expert staff to contribute to system security.
9. Providing guidelines in establishing information security policies.

Limitations of IDS

1. Detect attack only after they have entered the network.
2. Cannot expect to detect all malicious activity at all-time handling alert to trigge false positive or false negative alarm.
3. Cannot integrated with filtering rules security to stop traffic from attacking

2.8 Multiple Choice Questions

Q.1 Which of the following network resources can be load balanced ?

[a] Connection through intelligent switches
[b] DNS
[c] Storage resources
[d] All of these

Q.2 Each guest OS is managed by a Virtual Machine Monitor also known as _____.

[a] server
[b] hypervisor
[c] storage
[d] none

Q.3 _____ is the process of making logical components of resources over the existing physical resources.

[a] Virtualization
[b] Cloud computing
[c] Storage
[d] Loading

Q.4 Which of the following are types of server virtualization _____ ?

[a] full virtualization
[b] para-virtualization
[c] OS level virtualization
[d] All of these

Q.5 Which of the following type of computing ?