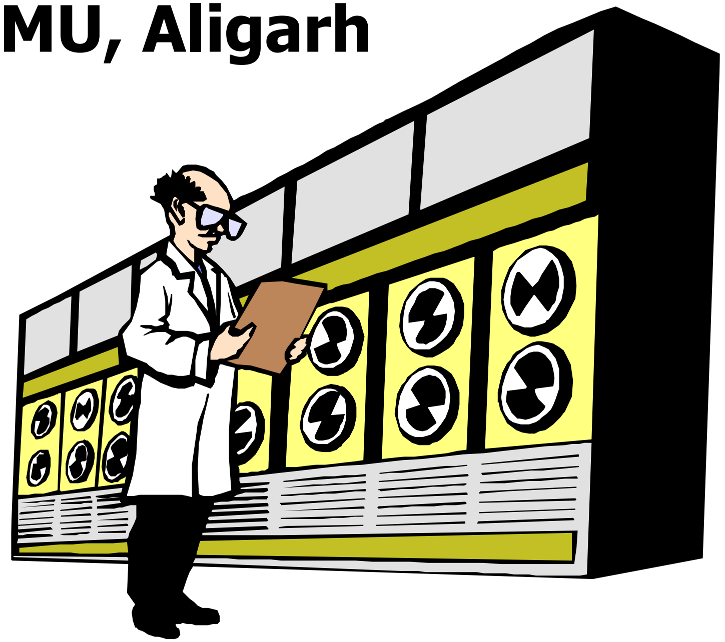
Introduction to Cloud Computing

Unit- I

Assistant Prof. M. Shahid
Department of Commerce, AMU, Aligarh



Content

- Cloud Computing
- Characteristics of Cloud Computing
- Advantages of Cloud Computing
- Challenges in Cloud Computing
- Cloud Service Delivery Models
- Cloud Services
- Cloud Deployment Models
- Cloud Operating System (OS)
- Federated Clouds

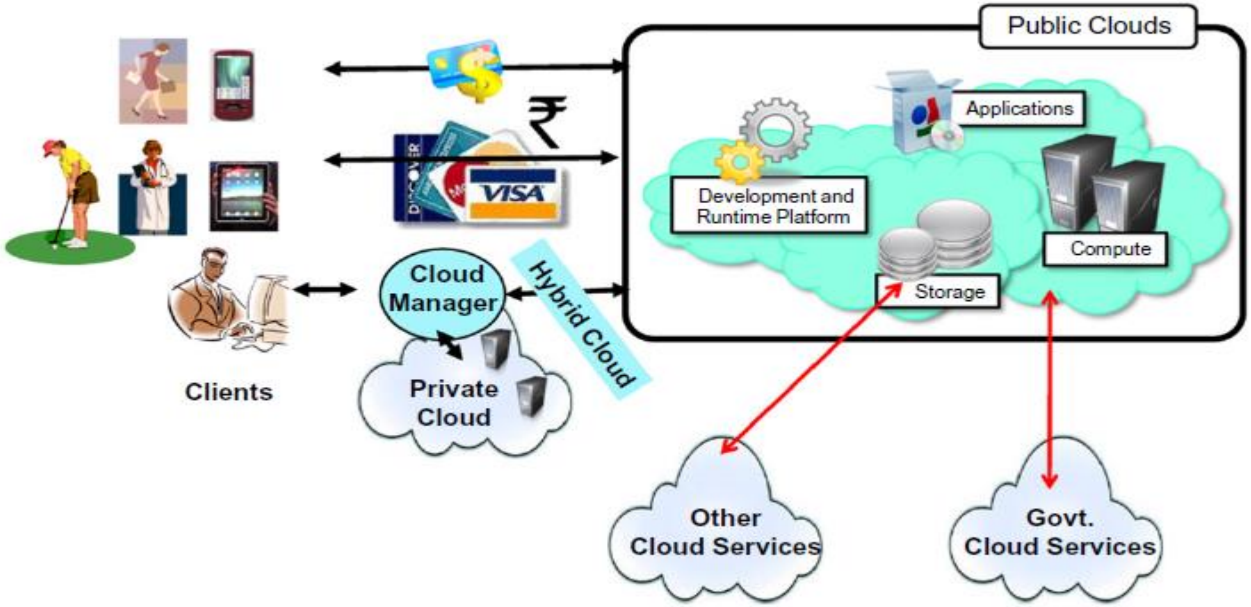
Definition of Cloud…

- Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction (NIST).
- Cloud computing refers to both the applications delivered as services over the Internet and the hardware and system software in the datacenters that provide those services (Armbrust et. al.)
- A cloud is a type of parallel and distributed system consisting of a collection of interconnected and virtualized computers that are dynamically provisioned and presented as one or more unified computing resources based on service-level agreements established through negotiation between the service provider and consumers (Raj Kumar Buyya).

Cloud Computing Scenario..

Manjrasoft

Subscription - Oriented Cloud Services:
X{compute, apps, data, ..}
as a Service (..aaS)



Public Clouds

Applications

Development and Runtime Platform

Storage

Compute

Clients

Cloud Manager

Private Cloud

Hybrid Cloud

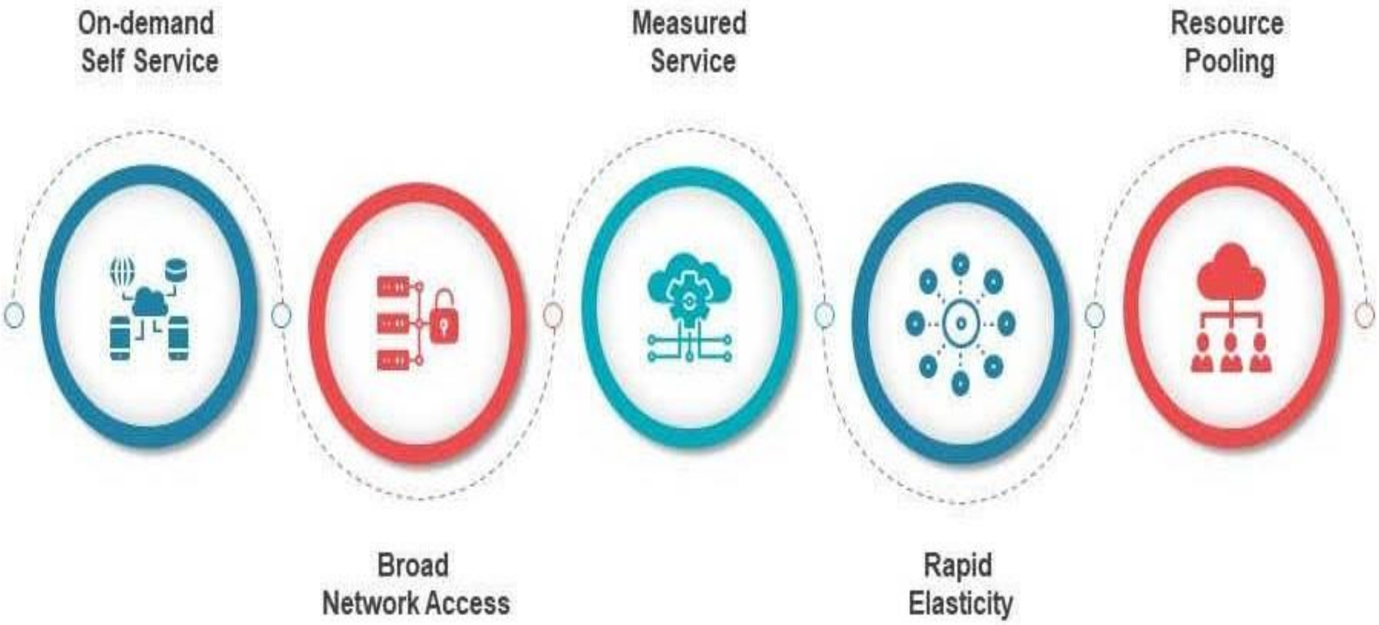
Other Cloud Services

Govt. Cloud Services

Definition of Cloud…

- **Cloud** is a type of Internet-based computing infrastructure that provides anything (shared hardware resources, s/w, data,…) as a Service on demand at Pay per use basis.
- Regardless of the kind of service, cloud computing services provide users with a series of functions including:
  - Email
  - Storage, backup, and data retrieval
  - Creating and testing apps
  - Analyzing data
  - Audio and video streaming
  - Delivering software on demand

Characteristics of Cloud



Characteristic of Cloud..

Here are the some main characteristics that cloud computing offers businesses today.

- **On-demand capabilities**
- **Device and location independence (through web browser)**
- **Resource pooling**
- **Nice pricing**
- **Measured Service**
- **Rapid elasticity: (peak-load capacity**) via dynamic provisioning
- **Pay you Go model: measured based service charge**
- **Virtualization**
- **Reliability** improves
- **Performance** is monitored by providers
- Broad Network access

Characteristics of Cloud Computing

- **On-demand self-service:** On-demand self service refers to the service provided by cloud computing vendors that enables the provision of cloud resources on demand whenever they are required.
- **Broad network access:** Broad network access refers to resources hosted in a private cloud network that are available for access from a wide range of devices. These resources are also accessible from a wide range of locations that offer online access.
- **Resource pooling:** The provider's computing resources are pooled to serve multiple consumers using a multi-tenant model.
- **Rapid elasticity:** Rapid elasticity is a term for scalable provisioning, or the ability to provide scalable services.
- **Measured service:** Resource usage can be monitored, controlled and reported, providing transparency for the provider and consumer.

Advantages of Cloud Computing

- ***No up-front investment:***
  - Pay-as you-go pricing model.
  - No need to invest in the infrastructure.
  - Resources are rented from the according to needs.
- ***Minimum operating cost:***
  - Resources are allocated and de-allocated on demand.
  - No need to provide capacities according to peak load.
  - Resources can be released to save on operating costs when service demand is low.

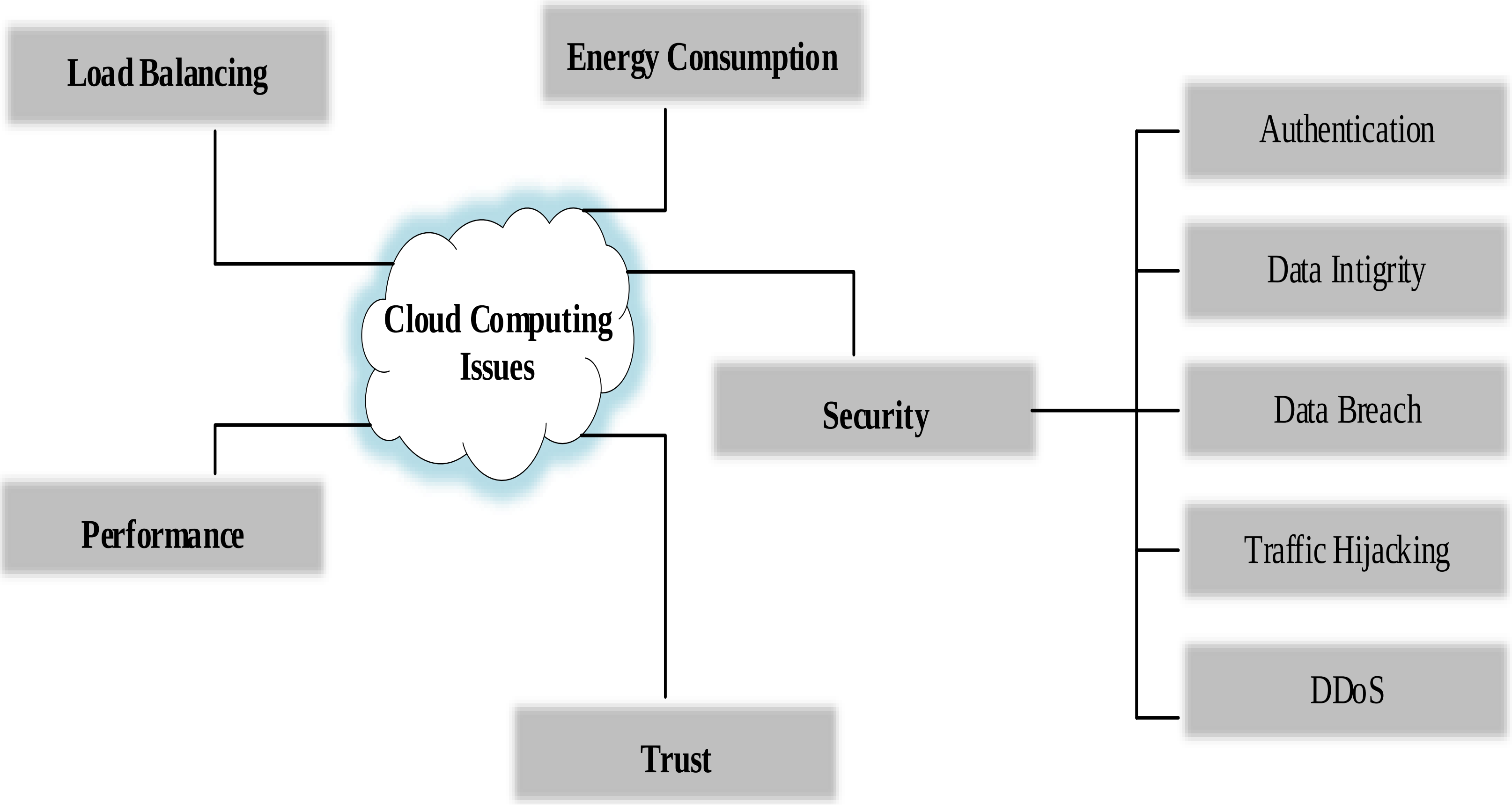
Advantages of Cloud Computing

- ***High Scalability:***
  - Infrastructure providers pool large amount of resources and make them easily accessible.
  - Service providers can easily expand its service to large scales to handle rapid increase in service demands
- ***Easy Access:***
  - Services hosted in the cloud are generally web-based.
  - Accessible through devices with Internet connections.
  - Devices: desktop, laptop, cell phones and PDAs.

Advantages of Cloud Computing

- ***Reducing business risks and maintenance expenses***
  - Outsource the service infrastructure to the cloud
  - Service providers shift business risks (such as hardware failures) to infrastructure providers
  - Infrastructure providers have better expertise and are better equipped for managing these risks.
  - A service provider can cut down the hardware maintenance and the staff training costs.

Challenges in Cloud Computing



Cloud Computing Issues

Load Balancing

Energy Consumption

Security

Authentication

Data Intigrity

Data Breach

Traffic Hijacking

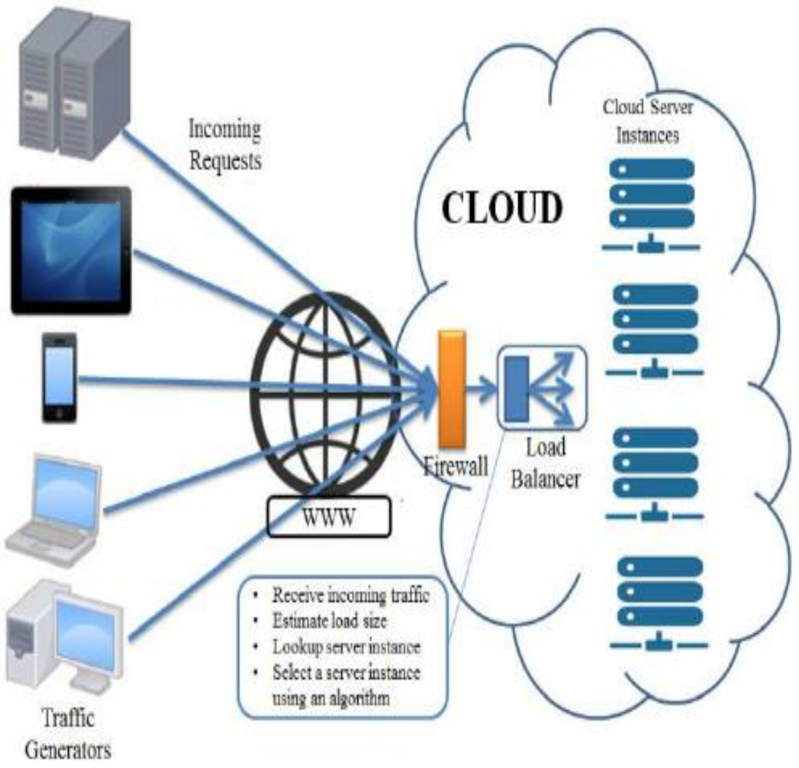
DDoS

Performance

Trust

Load Balancing

- Load balancing in cloud computing is the process in which workloads and computing resources are distributed across more than one servers. The workload is divided among two or more servers, network interfaces, hard drives and other computing resources which result in better utilization and system response time.
- Load Balancing helps in achieving Green computing by reducing energy consumption and carbon emission.



Energy Consumption

- The growth of using cloud computing technology has a remarkable rise in energy consumption in cloud infrastructure. The huge energy consumption has become a very crucial issue in cloud computing.
- The pace of data-centre expansion around the world raises energy issues and calls for innovative ways to make data-centre energy sustainable. Cloud infrastructure should be designed as a power-efficient infrastructure

Performance

- Communication delay makes cloud performance very worst.
- Cloud provider’s inability to scale up infrastructure as rising in customer’s demands can be a cause of performance failure.

Trust

- Most people worried about the capabilities of cloud computing systems.
- Lack of knowledge, lack of control and ownership over data, data loss, less transparency evolves trust issue in the cloud environment.

Security

- ***Authentication and access management:***
  - Mechanism to verify a user’s identity when he wants to access cloud resources.
- ***Data Integrity:***
  - Data stored or transit in the cloud environment will not be distorted by any unauthorized parties.
- ***Data Breach:***
  - exposes confidential information, potentially allowing sensitive information to be stolen and published in an unsecured or illegal venue.

Security

- ***Traffic Hijacking:***
  - misleading of internet traffic towards a malicious third party which can be harmful to cloud users.
- ***Integrity of resources***
- ***Authentication***
- ***Privacy Loss***
- ***Denial of Services:***
  - by sending malicious traffic into the cloud system, hackers attempt to disable cloud services, preventing cloud users from accessing cloud services.

Evolution of Cloud

Benefits

- Cost & management
  - Economies of scale, “out-sourced” resource management
- Reduced Time to deployment
  - Ease of assembly, works “out of the box”
- Performance
- Scaling
  - On demand provisioning, co-locate data and compute
- Reliability
  - Massive, redundant, shared resources
- Sustainability
  - Hardware not owned

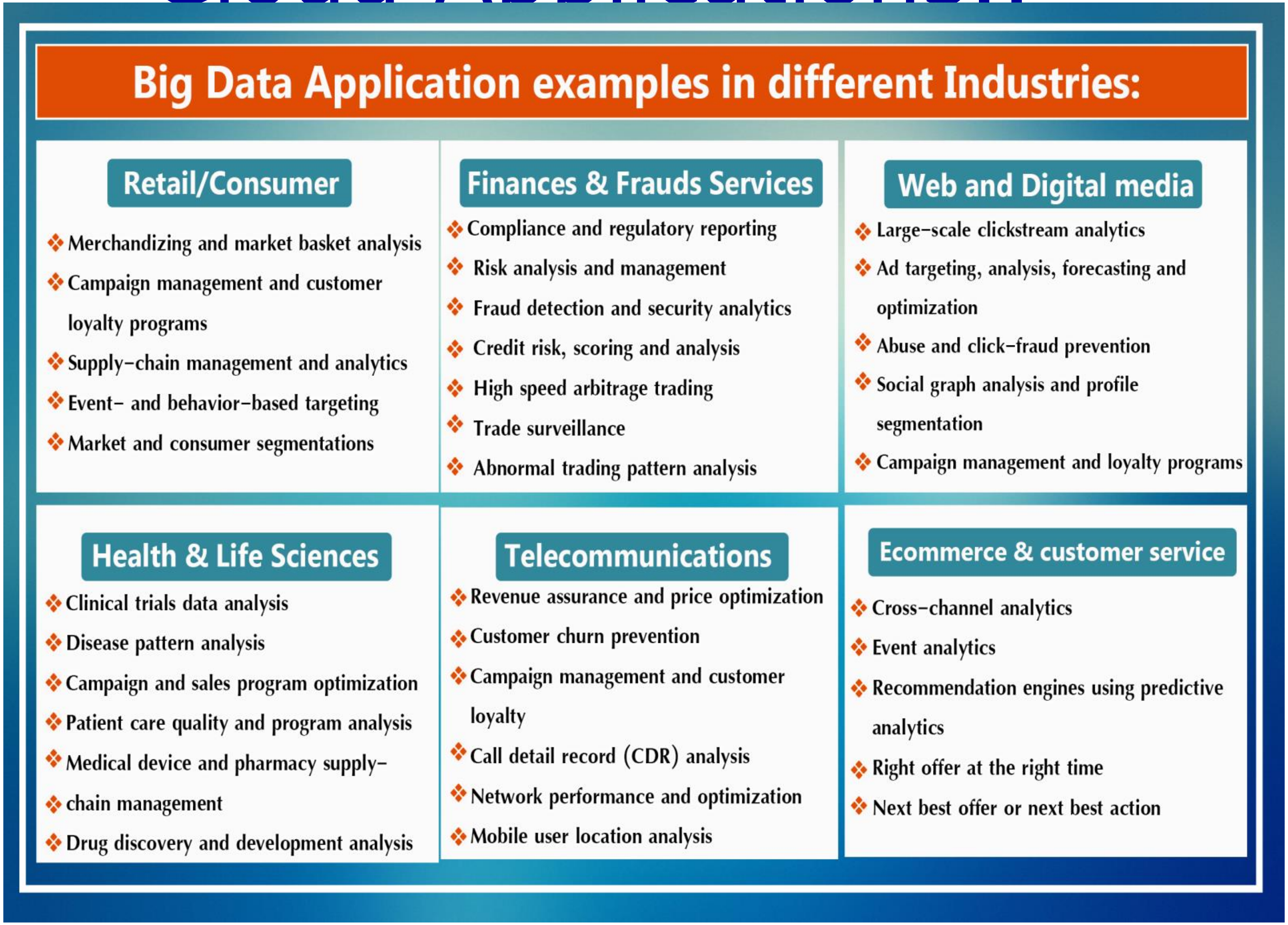
Cloud examples

- Amazon Elastic Compute Cloud
- Google App Engine, Microsoft Azure
- GoGrid, AppNexus
- Elastic Compute Cloud – EC2 (IaaS)
- Simple Storage Service – S3 (IaaS)
- Elastic Block Storage – EBS (IaaS)
- SimpleDB (SDB) (PaaS)
- Simple Queue Service – SQS (PaaS)
- CloudFront (S3 based Content Delivery PaaS)
- Consistent AWS Web Services API

Cloud Applications

- Create cloud-native applications
- Store, back up and recover data
- Stream audio and video
- Deliver software on demand
- Test and build applications
- Analyse data

Cloud Applications..



Big Data Application examples in different Industries:

Retail/Consumer

- Merchandizing and market basket analysis
- Campaign management and customer loyalty programs
- Supply-chain management and analytics
- Event- and behavior-based targeting
- Market and consumer segmentations

Finances & Frauds Services

- Compliance and regulatory reporting
- Risk analysis and management
- Fraud detection and security analytics
- Credit risk, scoring and analysis
- High speed arbitrage trading
- Trade surveillance
- Abnormal trading pattern analysis

Web and Digital media

- Large-scale clickstream analytics
- Ad targeting, analysis, forecasting and optimization
- Abuse and click-fraud prevention
- Social graph analysis and profile segmentation
- Campaign management and loyalty programs

Health & Life Sciences

- Clinical trials data analysis
- Disease pattern analysis
- Campaign and sales program optimization
- Patient care quality and program analysis
- Medical device and pharmacy supply-
- chain management
- Drug discovery and development analysis

Telecommunications

- Revenue assurance and price optimization
- Customer churn prevention
- Campaign management and customer loyalty
- Call detail record (CDR) analysis
- Network performance and optimization
- Mobile user location analysis

Ecommerce & customer service

- Cross-channel analytics
- Event analytics
- Recommendation engines using predictive analytics
- Right offer at the right time
- Next best offer or next best action

Enterprise cloud paradigm

- Enterprises will place stringent requirements on cloud providers to pave the way for more widespread adoption of cloud computing
- Enterprise cloud computing is the alignment of a cloud computing model with an organization's business objectives and processes
- Objectives may be (profit, return on investment, reduction of operations costs).

Enterprise cloud

- An enterprise cloud is **a unified IT operating environment that melds private cloud, public cloud, and distributed cloud**, providing a single point of control for managing infrastructure and applications in any cloud.

Cloud Service Provider

- A cloud service provider is a third-party company offering a cloud-based platform, infrastructure, application, or storage services. Some are as:
    - Amazon Web Services (AWS)
    - Microsoft Azure.
    - Google Cloud (GCP—formerly Google Cloud Platform)
    - IBM Cloud (formerly SoftLayer)
    - Oracle Cloud.
    - Alibaba Cloud
    - RedHat.

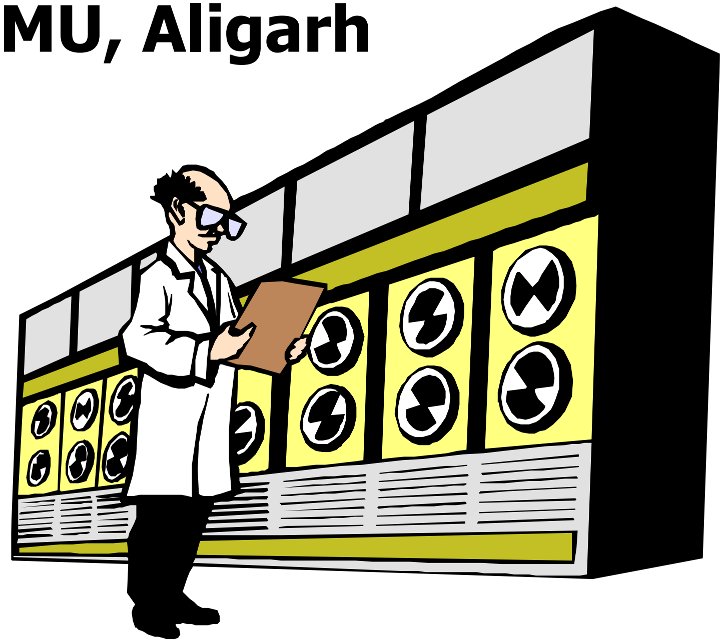
Significance of Cloud Service Provider

- Better control over service levels
- Performance
- Maintenance
- Disaster recovery and Management
- Fast response times
- Vendor interfacing

Introduction to Cloud Computing

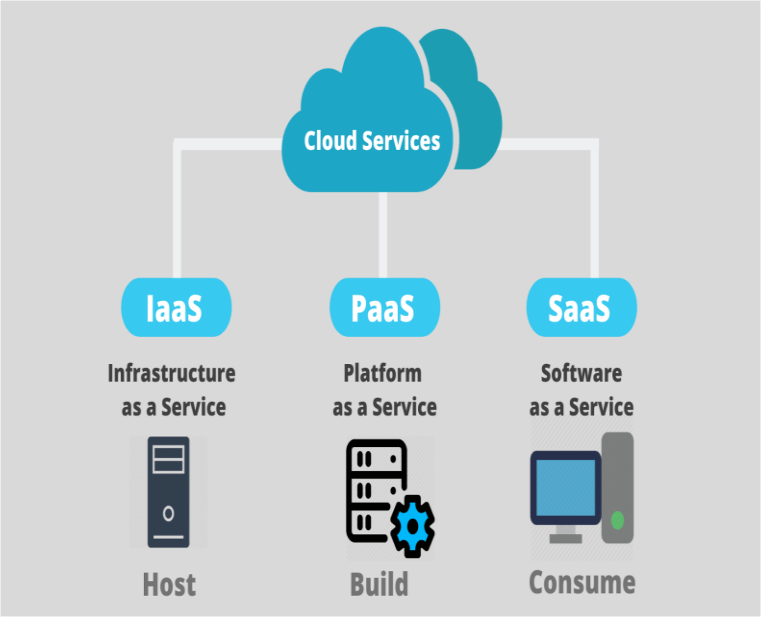
Unit- II

Assistant Prof. M. Shahid
Department of Commerce, AMU, Aligarh



Cloud Service Delivery Models

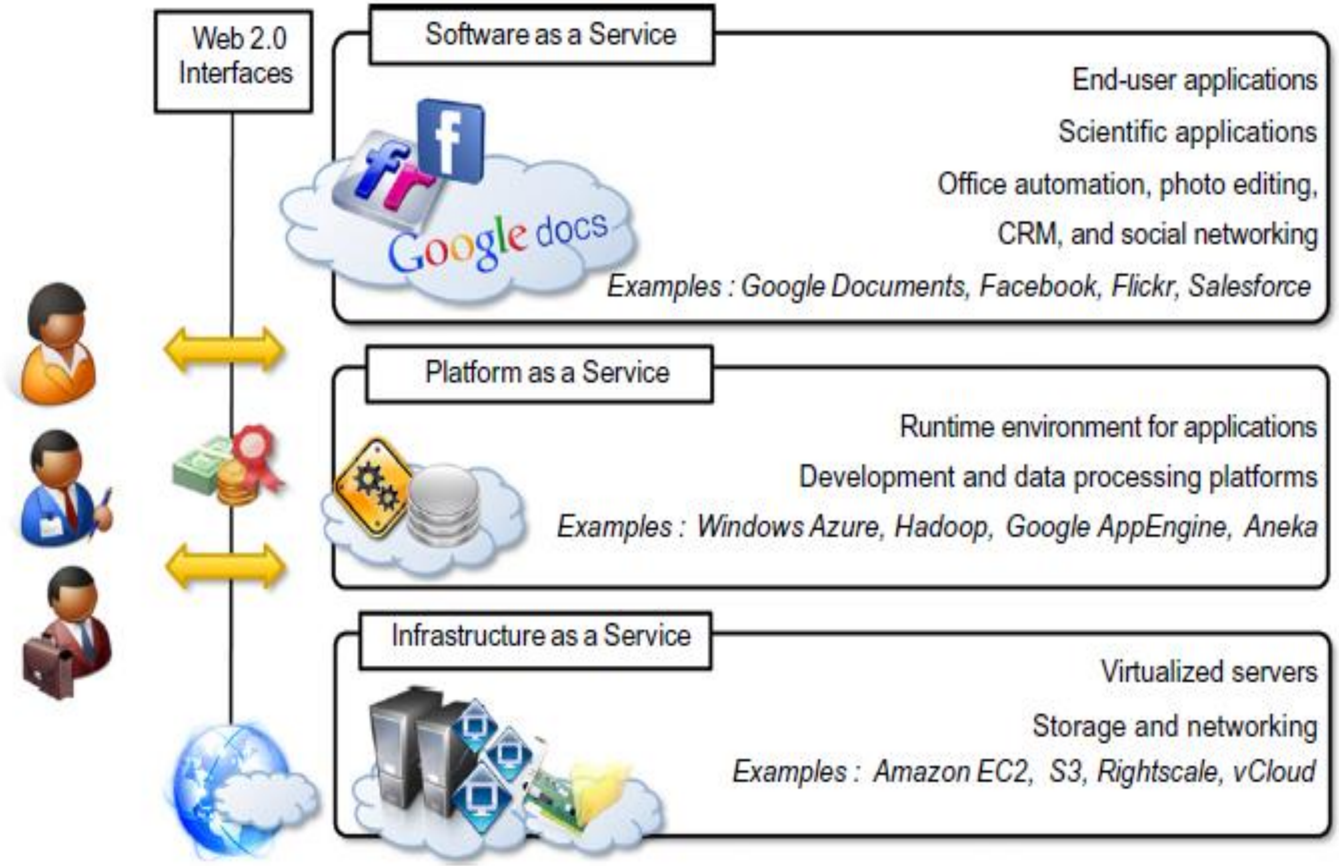
- Three main types of Delivery Models:
  - IaaS ***(Infrastructure as a Service)***
  - PaaS ***(Platform as a Service)***
  - SaaS ***(Software as a Service)***



Everything as a Service

- Utility computing = Infrastructure as a Service (IaaS)
  - Why buy machines when you can rent cycles?
  - Examples: Amazon’s EC2, Rackspace
- Platform as a Service (PaaS)
  - Give me nice API and take care of the maintenance, upgrades, …
  - Example: Google App Engine
- Software as a Service (SaaS)
  - Just run it for me!
  - Example: Gmail, Salesforce

Service Models



Software as a Service

End-user applications

Scientific applications

Office automation, photo editing, CRM, and social networking

Examples : Google Documents, Facebook, Flickr, Salesforce

Platform as a Service

Runtime environment for applications

Development and data processing platforms

Examples : Windows Azure, Hadoop, Google AppEngine, Aneka

Infrastructure as a Service

Virtualized servers

Storage and networking

Examples : Amazon EC2, S3, Rightscale, vCloud

IaaS ***(Infrastructure as a Service)***

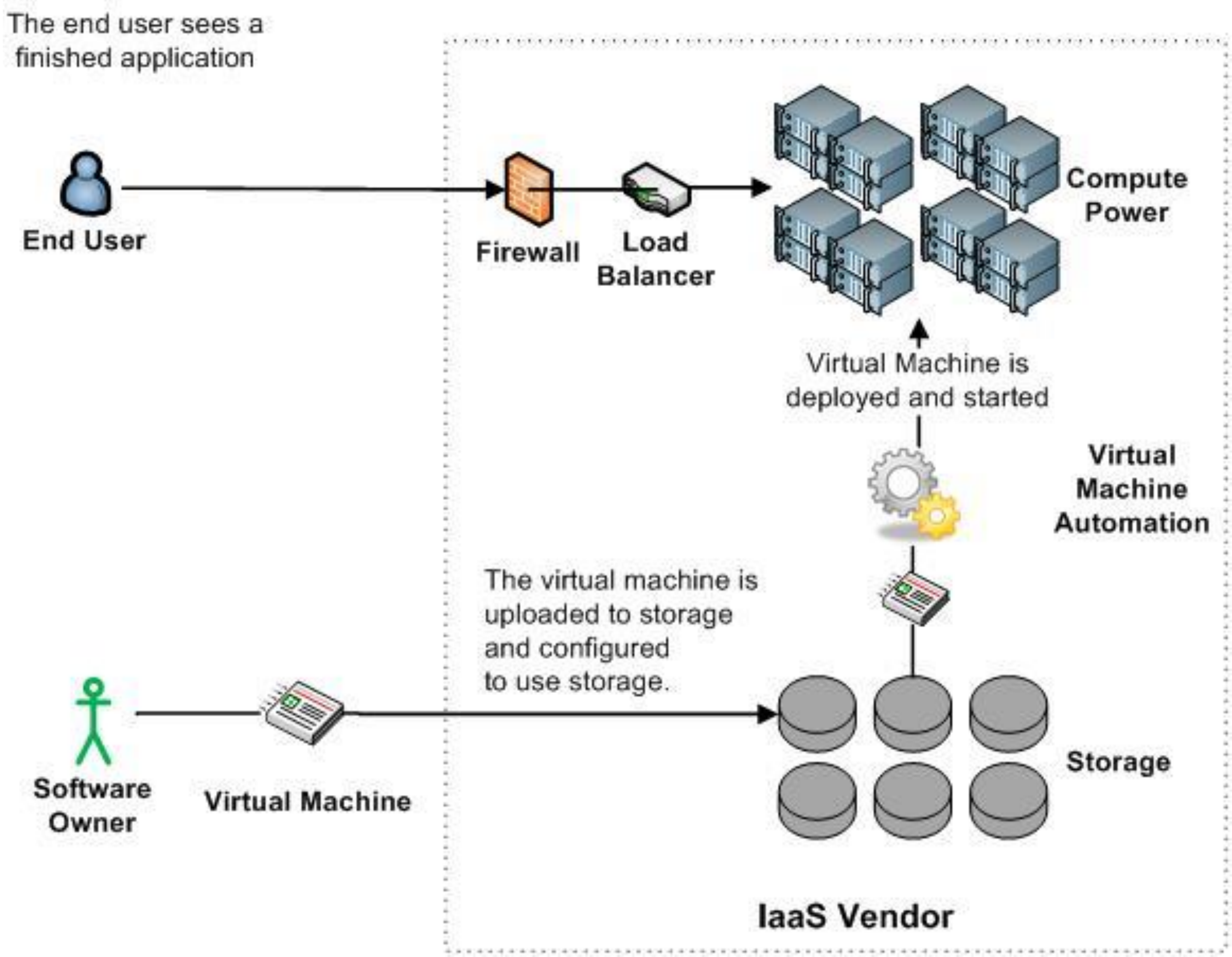
- Infrastructure of server, software and network equipment is provided as an on-demand service by the cloud provider.

  e.g: Amazon, Zimony etc.

Advantages:

- Used to avoid buying, housing and managing the basic network and software infrastructure component.

Infrastructure as a Service (IaaS)


The end user sees a finished application
End User
Firewall
Load Balancer
Compute Power
Virtual Machine is deployed and started
Virtual Machine Automation
The virtual machine is uploaded to storage and configured to use storage.
Software Owner
Virtual Machine
Storage
IaaS Vendor

PaaS ***(Platform as a Service)***

- It is delivery of computing platform over the web where user can create and install their application as their requirement.
- Configuration of computing platform and server is managed by vendor or cloud provider

e.g: Google App Engine, Window Azure

Advantages:

- Elimination of network dependencies

SaaS ***(Software as a Service)***

- Cloud computing deliver a SaaS where user do not need to manage installation and configuration of any hardware or software.
- All the installation and configuration services are managed by vendor or cloud provider.

e.g. Google Docs, Google Online Office etc.

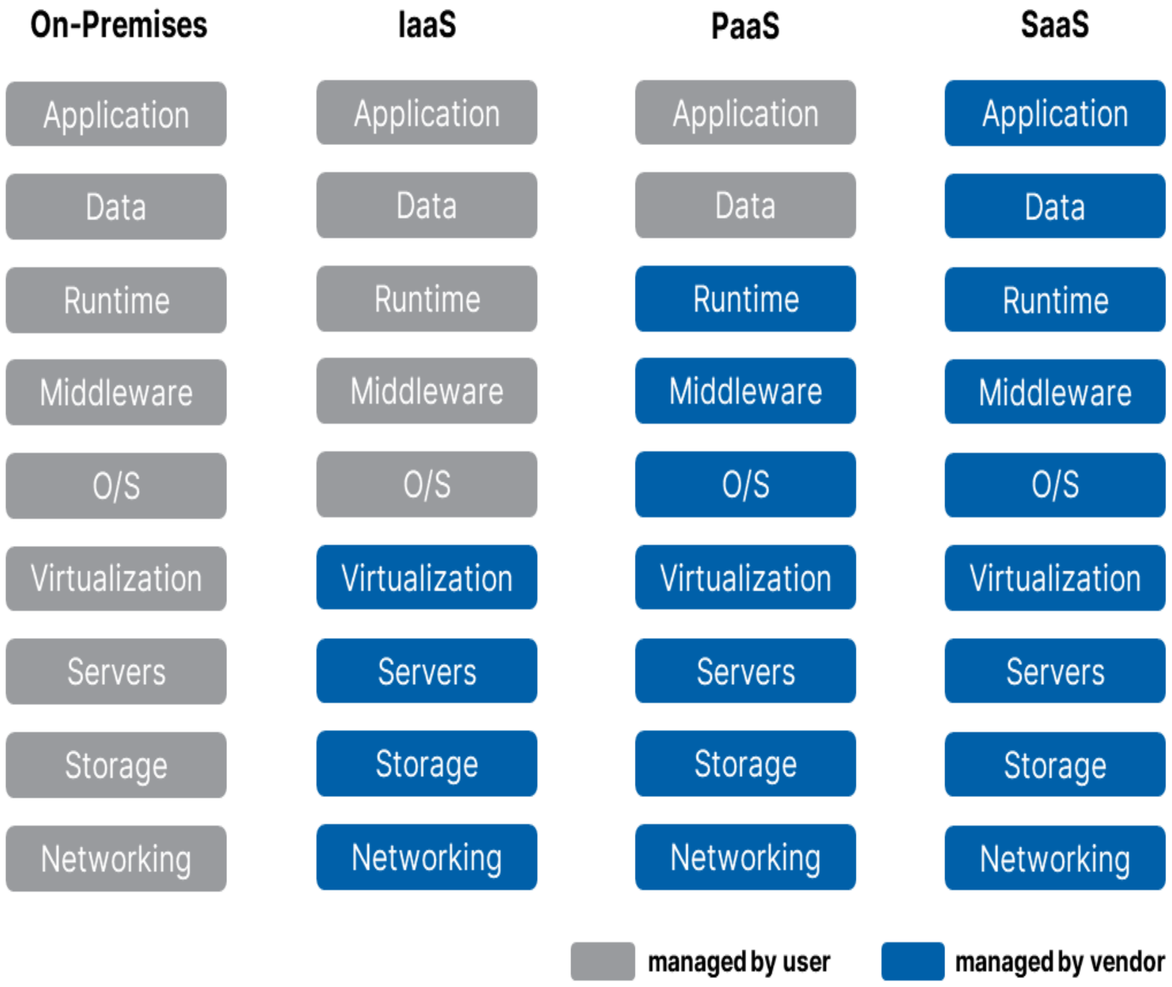
Advantages:

- Reduced up-front cost.
- Potential for reduced lifetime cost
- Elimination of licensing risk

Service Model..

- Infrastructure as a service (IaaS)
  - Offering hardware related services using the principles of cloud computing. These could include storage services (database or disk storage) or virtual servers.
  - Amazon EC2, Amazon S3, Rackspace Cloud Servers and Flexiscale.
- Platform as a Service (PaaS)
  - Offering a development platform on the cloud.
  - Google's Application Engine, Microsofts Azure
- Software as a service (SaaS)
  - Including a complete software offering on the cloud. Users can access a software application hosted by the cloud vendor on pay-per-use basis. This is a well-established sector.
  - Salesforce.coms' offering in the online Customer Relationship Management (CRM) space, Googles gmail and Microsofts hotmail, Google docs.

Service Management in IaaS, PaaS & SaaS

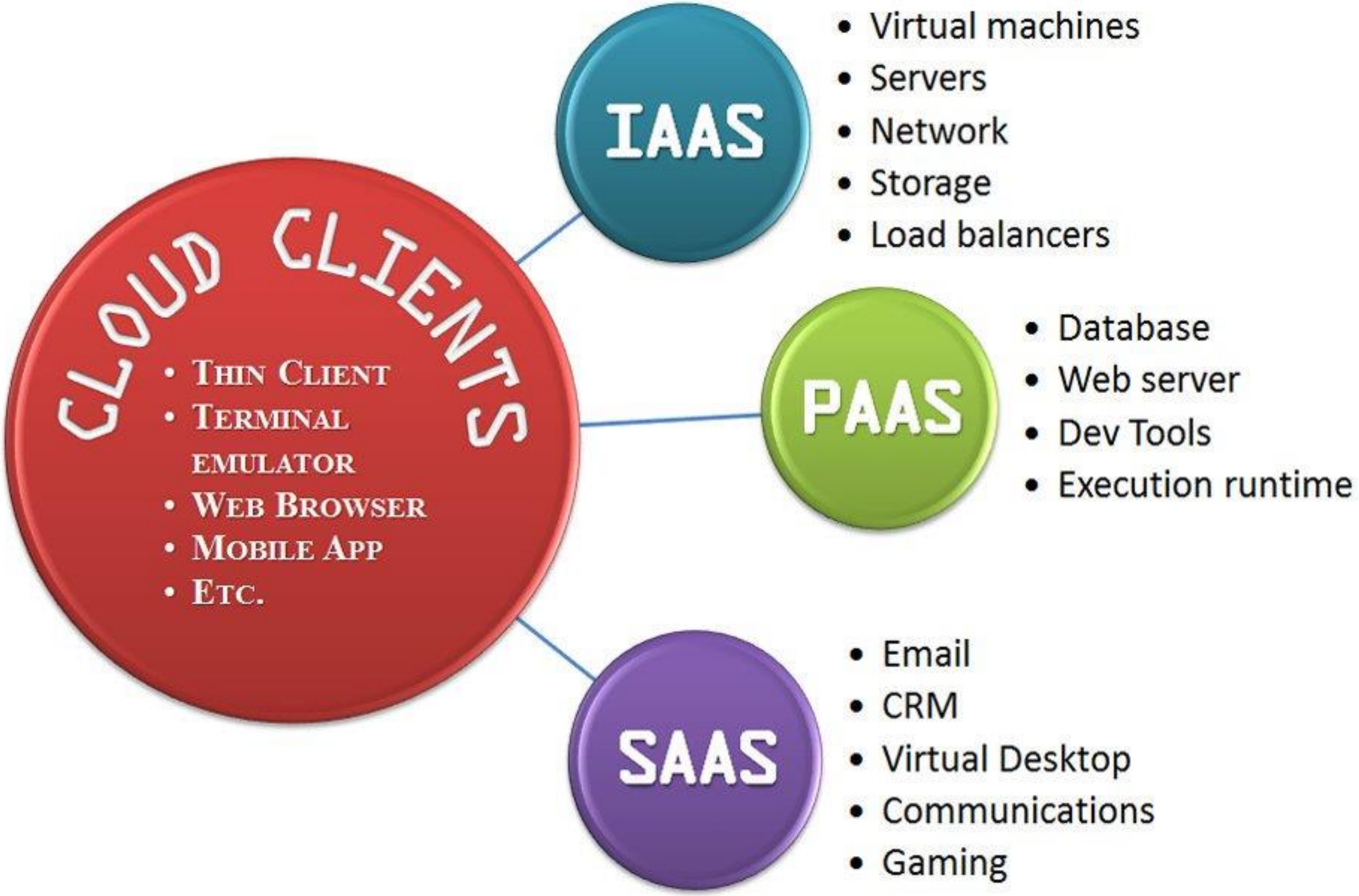


On-Premises	IaaS	PaaS	SaaS
▨ Application	▨ Application	▨ Application	🟦 Application
▨ Data	▨ Data	▨ Data	🟦 Data
▨ Runtime	▨ Runtime	🟦 Runtime	🟦 Runtime
▨ Middleware	▨ Middleware	🟦 Middleware	🟦 Middleware
▨ O/S	▨ O/S	🟦 O/S	🟦 O/S
▨ Virtualization	🟦 Virtualization	🟦 Virtualization	🟦 Virtualization
▨ Servers	🟦 Servers	🟦 Servers	🟦 Servers
▨ Storage	🟦 Storage	🟦 Storage	🟦 Storage
▨ Networking	🟦 Networking	🟦 Networking	🟦 Networking

▨ managed by user

🟦 managed by vendor

Cloud Services



Types of Cloud

Cloud Deployment Models

Public/Internet Clouds

*Third-party, multitenant cloud infrastructure and services

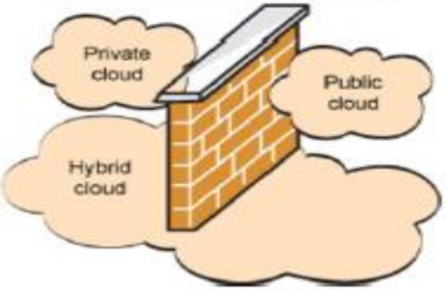
*Available on a subscription basis to all

Private/Enterprise Clouds

*A public cloud model within a company's own datacenter/infrastructure for internal and/or partners' use

Hybrid/Inter Clouds

* Mixed use of private and public clouds; leasing public cloud services when private cloud capacity is insufficient



Types of Cloud..

- **Public Cloud**: Computing infrastructure is hosted at the vendor’s premises.
- **Private Cloud**: Computing architecture is dedicated to the customer and is not shared with other organizations.
- **Hybrid Cloud**: Organizations host some critical, secure applications in private clouds. The not so critical applications are hosted in the public cloud

Cloud examples

- Amazon Elastic Compute Cloud
- Google App Engine, Microsoft Azure
- GoGrid, AppNexus
- Elastic Compute Cloud – EC2 (IaaS)
- Simple Storage Service – S3 (IaaS)
- Elastic Block Storage – EBS (IaaS)
- SimpleDB (SDB) (PaaS)
- Simple Queue Service – SQS (PaaS)
- CloudFront (S3 based Content Delivery PaaS)
- Consistent AWS Web Services API

Public Cloud

- Public cloud allows users to access the cloud publicly via interfaces using web browsers.
- Users need to pay only for the time duration they use the services i.e pay-per-use.

Advantages:

- Low cost
- High scalability
- Higher degree of elasticity
- Reliable and flexible.

Disadvantages:

- Low secure and less customizable.

Private Cloud

- Especially architects and accessible for one individual enterprise only. It is restricted only to an area.
- In most cases, private cloud infrastructure is owned and maintained by enterprises itself.

Advantages:

- Good performance
- High reliability
- More secure

Disadvantages:

Low elasticity, Expensive and required large space

Community Cloud

- Architect to share delivered resources for several organisations which have mutual concerns such as security, privacy, mission, policy, compliances, etc
- Managed by some organizations of the community or may be managed by an outer organization.

Advantages:

Low expenses

High security

Good for handling large spikes in workload

Disadvantages:

Limited control, No cloud bursting support and Low control over cloud resources.

Hybrid Cloud

- infrastructure is the intermixing of both private cloud and public cloud
- facilitate the services of both types of clouds (Public & Private) for efficient work.
- widely used infrastructure in the business-oriented sector

Advantages:

Low capital expenses

provide high security for sensitive data

Disadvantages:

More complexity

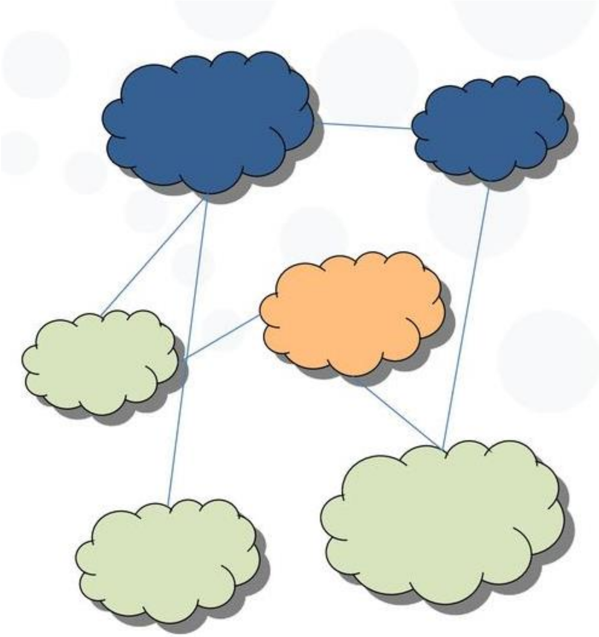
Compatibility issues

Cloud Operating System (OS)

- Cloud operating system manages, operates and executes the process of virtual infrastructure like virtual machines, virtual servers and hardware and software resources within the cloud environment.
- Cloud OS aims to provide an expressive set of resource management options and metrics to applications to facilitate programming in the Cloud
- Functionalities varies depending on the virtual environment.
- Can be installed and used along with other operating systems or can operate as a stand- alone operating system.
    - Eg: Google Chrome, Microsoft Azure

Federated Clouds

- A seamless environment formed by connecting the cloud environment of two or more cloud service provider using a common standard.
- It integrates heterogeneous cloud environment to scale up the resources and services for the users.

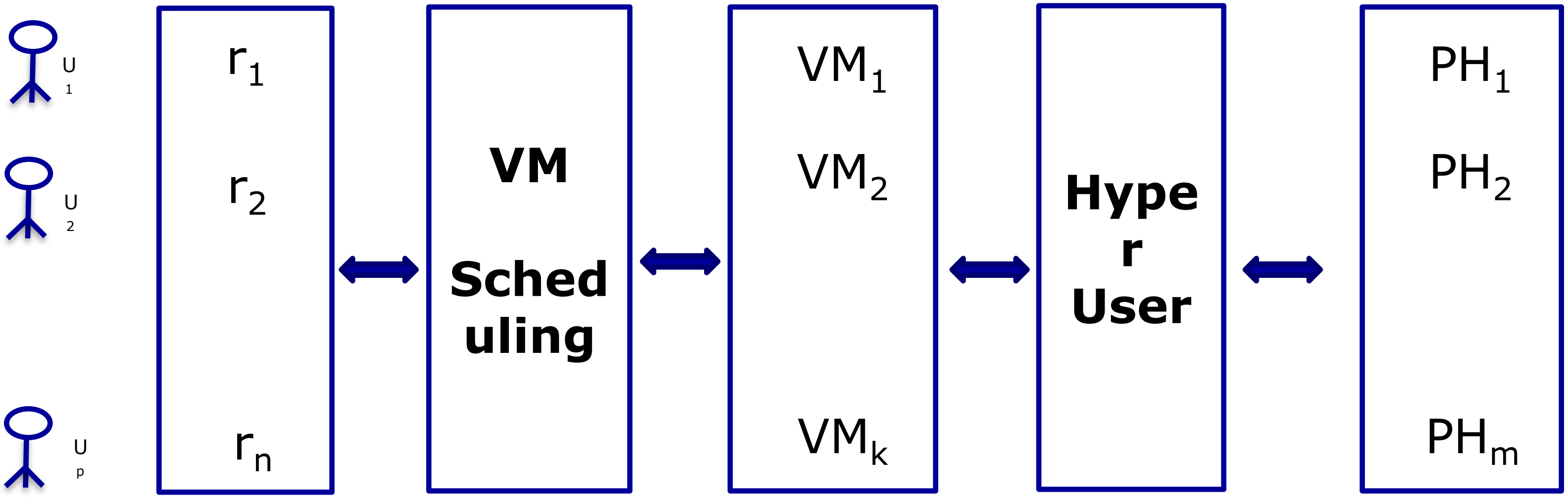


Properties of Federated Cloud

- Users can interact with the architecture either centrally or in a decentralized manner.
- Can be practiced with various niches like commercial and non-commercial.
- Can be monitored in two ways.
  - MaaS (Monitoring as a Service) provides information that aids in tracking contracted services to the user.
  - Global monitoring aids in maintaining the federated cloud.
- Providers in federated clouds publish their offers to a central entity. users cam interact with central entity to negotiate the prices and offers.
- Marketing objects like infrastructure, software, and platform have to pass through federation when consumed in the federated cloud.

Benefits of Federated Cloud

- Low Energy Consumption
- highly reliable
- minimize the time and cost of providers due to dynamic scalability.
- Provide easy scaling up of resources.



U_1

U_2

U_p

r_1

r_2

r_n

VM

Scheduling

VM_1

VM_2

VM_k

Hyper User

PH_1

PH_2

PH_m

Security
Unit- III

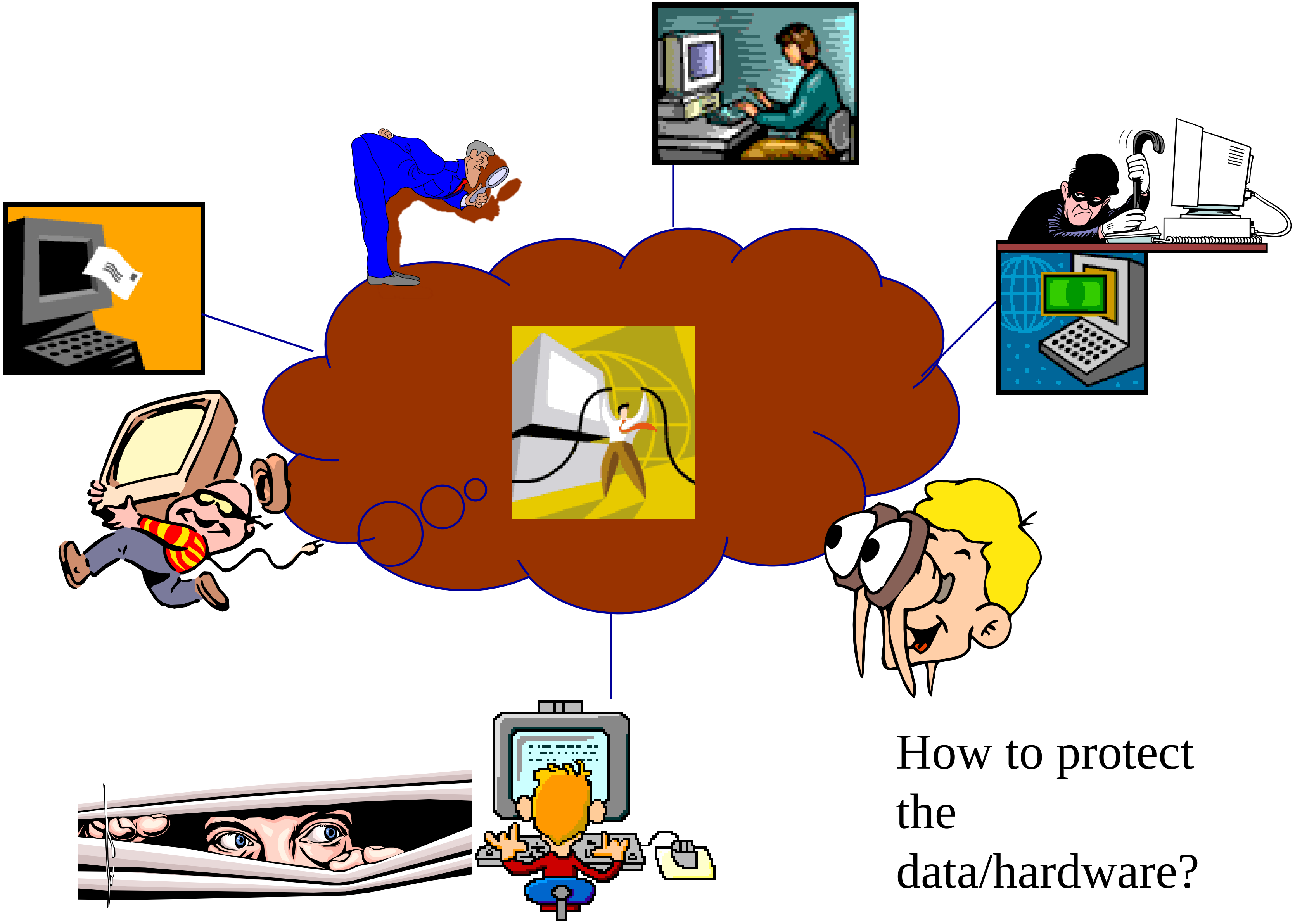
Presented By:

Dr Mohammad Shahid
Dept. of Commerce, AMU, Aligarh

In This Lecture

- Security
- Security Objectives
- Security Threats/Attacks
- Security Measures

Basic scenario of Security



How to protect
the
data/hardware?

Security

It is protection mechanism against unauthorized use, access, disclosure, modify and destruction of resources.

Information
Services

To prevent theft of or damage to the hardware

To prevent theft of or damage to the information

To prevent disruption of service

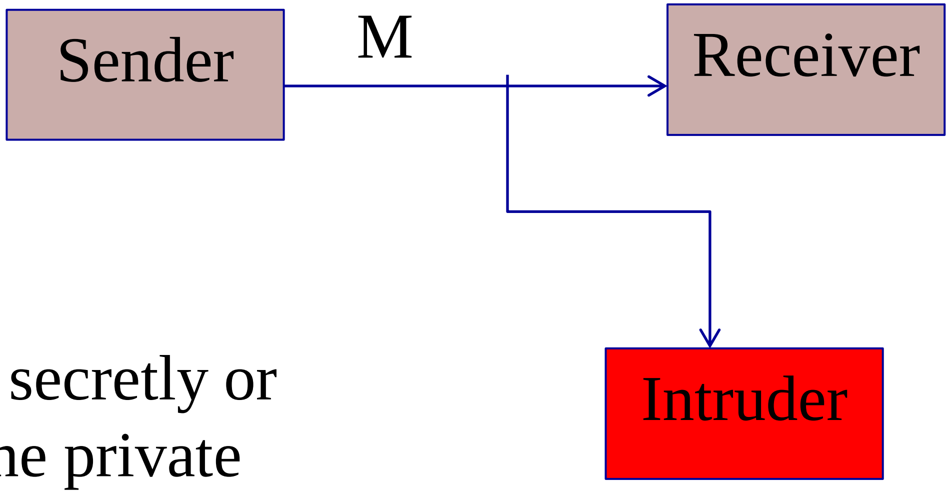
Security Objectives

- Confidentiality
  - unauthorized disclosure of data
- Integrity
  - unauthorized modification
- Availability
  - Easy access of information at any time
- Authentication
  - receiver must be sure of senders identity
- Nonrepudiation
  - Can’t deny action after fact
- Entity authentication
  - User is verified prior to access to the resources

Security Threats/Attacks

1. Interception:

Confidentiality
Passive Attack



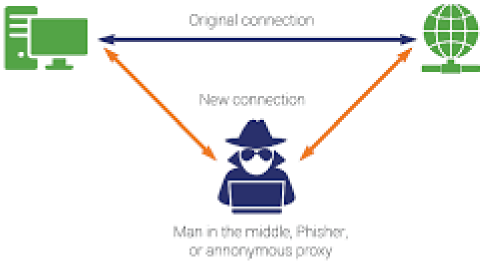
Sender

M

Receiver

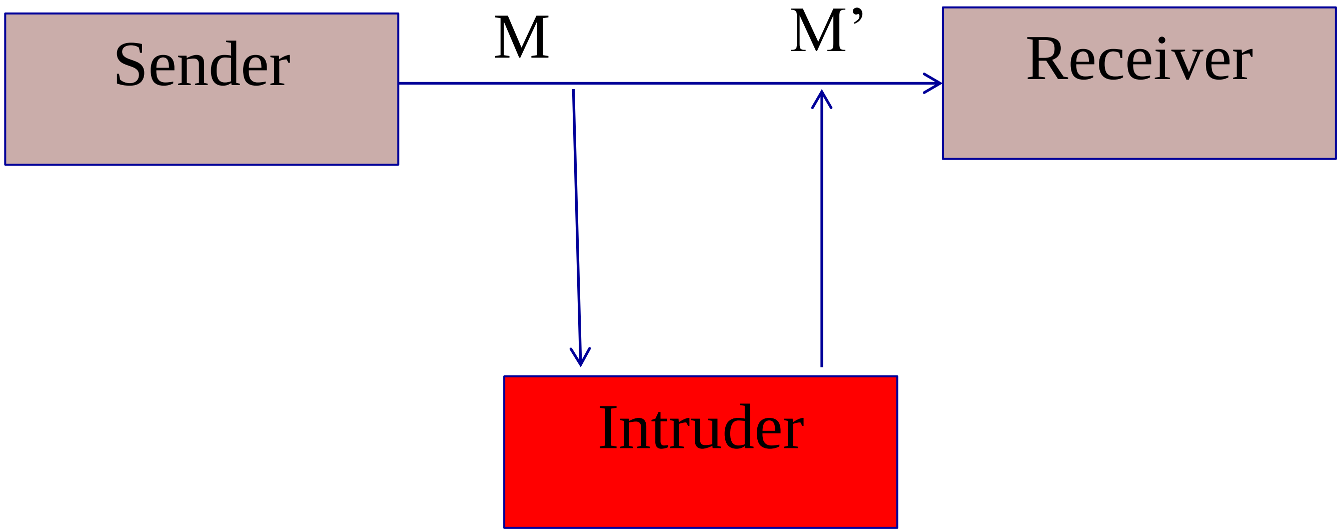
Intruder

Eavesdropping: Act of secretly or stealthily listening to the private communications without consent



Security Threats/Attacks..

2. Modification/Alteration:

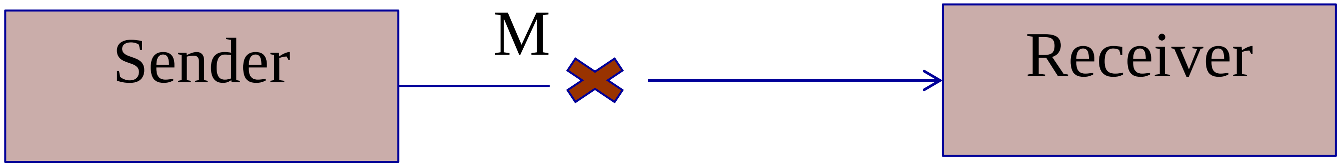


Integrity

Active Attack

Security Threats/Attacks..

3. Interruption

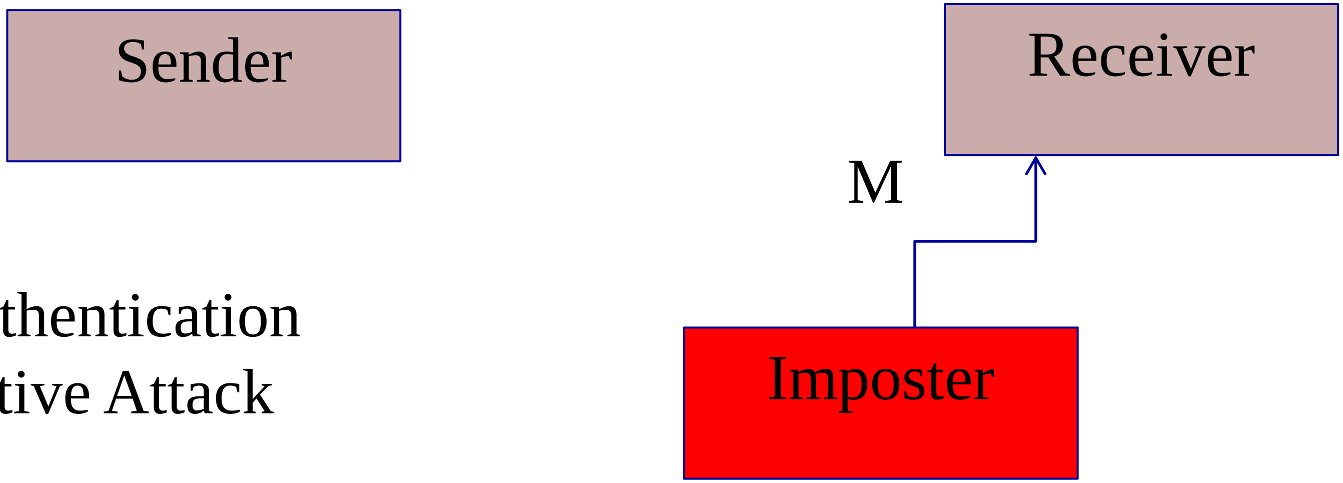


Availability

Active Attack

Security Threats/Attacks..

4. Fabrication



Sender

Receiver

M

Imposter

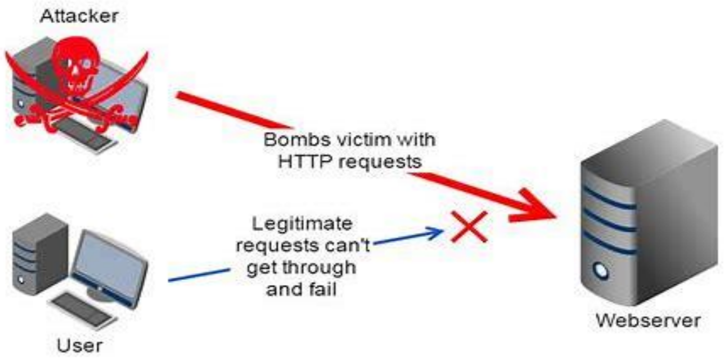
Authentication
Active Attack

Spoofing: Obtain someone's personal information by pretending to be a legitimate entity.

Security Threats/Attacks..

- ❑ DoS Attack

Flooding of request on server to interrupt its availability


Attacker
Bombs victim with HTTP requests
Legitimate requests can't get through and fail
User
Webserver

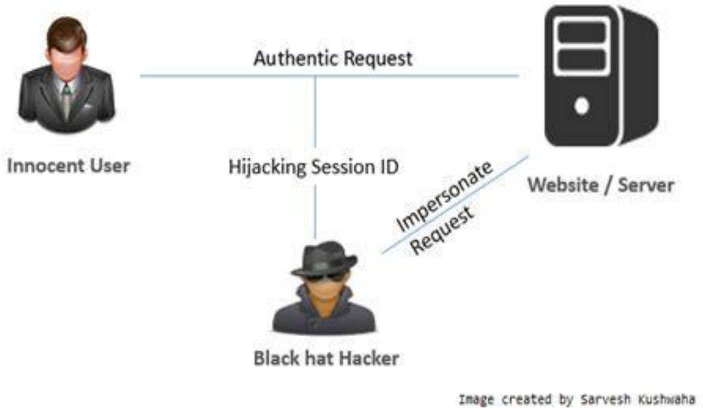
- ❑ Pishing Attack

Stealing secrets by fake entities i.e. email, website etc.

Security Threats/Attacks..

- ❑ Highjack attack

Gaining control over system



- ❑ Sniffing

Monitoring and capturing all the packets passing through a given network The message may be altered.

Security Threats/Attacks..

Malwares attacks

- Virus (parasitic & infectious for files)
- WORM (self-replicating)
- Ransom ware (infects files and demands ransom)
- Keylogger (record & steal the keystrokes)
- Spyware (gain control over cameras etc.)
- Trojan Horse (Non self-replicating, pretending harmless, invoke bot or ransom ware)
- Adware (unwanted adds)

Security Measures

Cryptography
Hashing
Digital Signature
Backups
Firewall
Anti Virus
Passwords

What is Cryptography?

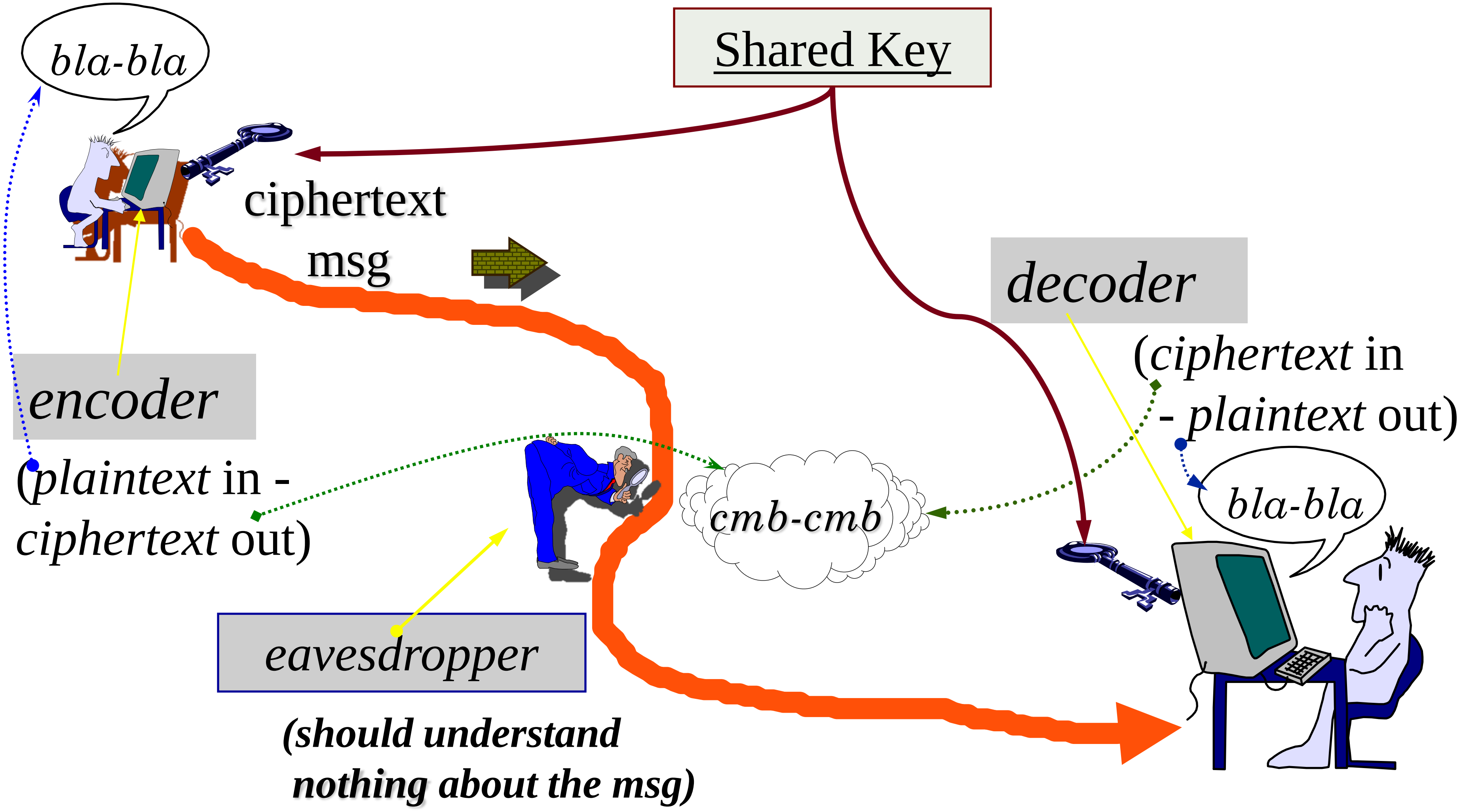
- *It all started with*
  - **Encryption / Decryption**

“ATTACK AT MIDNIGHT” - *plaintext*

↓

“BUUBDL BU NJEOJHIU” - *ciphertext*

Encryption / Decryption



Shared Key

bla-bla

ciphertext msg

encoder

(*plaintext* in - *ciphertext* out)

eavesdropper

(should understand nothing about the msg)

cmb-cmb

decoder

(*ciphertext* in - *plaintext* out)

bla-bla

Firewall

- A firewall is a network security system that monitors and controls the incoming and outgoing network traffic based on predetermined security rules. Has two types
  - Packet filter firewall
  - Application level proxy server

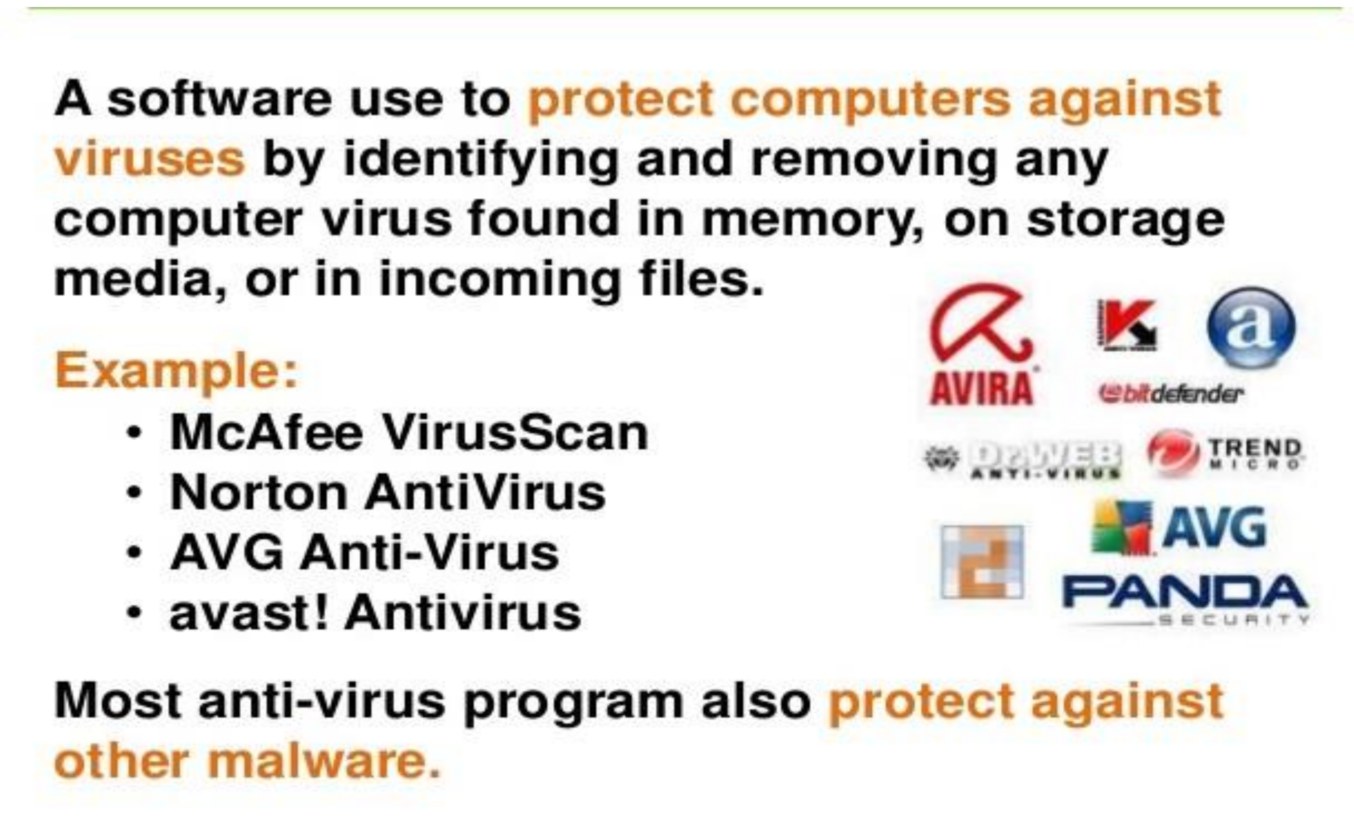
Packet filter firewall

- Checks IP address of incoming packet and rejects anything that does not match the Access Control Lists of trusted addresses (prone to IP spoofing: one host claims to have the IP address of another)
- Filter based on :
  - Protocols,
  - TTL Values,
  - Network block of the originator,

Application level proxy

- Examines the application used for each individual IP packet (e.g., HTTP, FTP) to verify its authenticity.
- A proxy server may act as a firewall by responding to input packets in the manner of an application, while blocking other packets.
- Can restrict or prevent networked computer worms and Trojans.

Anti Virus



A software use to protect computers against viruses by identifying and removing any computer virus found in memory, on storage media, or in incoming files.

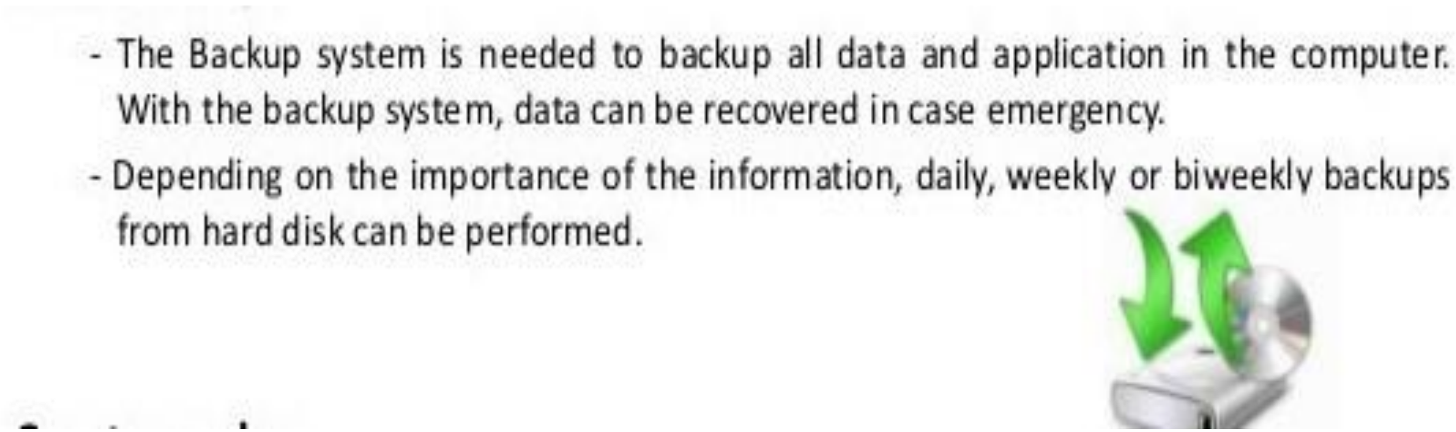
Example:

- McAfee VirusScan
- Norton AntiVirus
- AVG Anti-Virus
- avast! Antivirus

Most anti-virus program also protect against other malware.

AVIRA
bitdefender
Dr.WEB ANTI-VIRUS
TREND MICRO
AVG
PANDA SECURITY

Data Backup



- The Backup system is needed to backup all data and application in the computer. With the backup system, data can be recovered in case emergency.
- Depending on the importance of the information, daily, weekly or biweekly backups from hard disk can be performed.

Entity Authentication

Passwords
Biometrics
Cards

Fixed passwords
OTP

FP
Voice
Face
Retina

Smart
Credit
Debit

Unit IV
Management and Case Study

Change Management in the Cloud Age

In the cloud age, change management refers to the processes and tools that organizations use to manage and implement changes to their cloud-based systems and applications. This can include things like updating software, adding or removing users, or making changes to security settings. Effective change management in the cloud is important because it helps organizations ensure that their systems remain stable and secure and that any changes made to the system do not disrupt business operations. This can help organizations avoid downtime and other problems that can result from unplanned or poorly-managed changes. To effectively manage changes in the cloud, organizations should establish clear processes and procedures for requesting, reviewing, and implementing changes. They should also have tools in place to track and monitor changes and to roll back any changes that cause problems. Additionally, organizations should ensure that all members of the IT team are trained on how to use the change management processes and tools and that they understand the importance of following these processes.

SLA Management in the Cloud Computing

SLA management in cloud computing refers to the processes and tools that organizations use to manage and monitor their service level agreements (SLAs) with cloud service providers. SLAs are contracts that outline the level of service that a cloud provider will provide to an organization, including things like uptime, response times, and security measures. Effective SLA management in the cloud is important because it helps organizations ensure that they are receiving the level of service that they have contracted for and that any problems or issues with the service are addressed promptly. This can help organizations avoid downtime and other problems that can result from poor service. To effectively manage SLAs in the cloud, organizations should establish clear processes and procedures for monitoring and enforcing their SLAs. They should also have tools in place to track and monitor the performance of their cloud service providers and to raise any issues or concerns with the provider. Additionally, organizations should ensure that all members of the IT team are trained on how to use the SLA management processes and tools and that they understand the importance of following these processes.

Legal and Ethical Issues in Cloud Computing

There are several legal and ethical issues that organizations need to consider when using cloud computing. These include:

- **Data Privacy and Security:** Organizations need to ensure that they are compliant with relevant laws and regulations regarding the storage, access, and use of personal data in the cloud. This includes things like the General Data Protection Regulation (GDPR) in Europe and the California Consumer Privacy Act (CCPA) in the United States.
- **Data Ownership and Control:** Organizations need to be clear about who owns the data that is stored in the cloud, and who has the right to access and use it. This is particularly important for sensitive or confidential data, such as personal or financial information.
- **Data Location and Jurisdiction:** Organizations need to be aware of where their data is stored in the cloud, and what laws and regulations apply to that location. For example, data stored in a different country may be subject to different privacy laws than data stored in the organization's home country.

- **Service Level Agreements (SLAs):** Organizations need to carefully review and understand the SLAs that they have with their cloud service providers, and ensure that they are in compliance with these agreements. This includes things like uptime, response times, and security measures.
- **Interoperability and Vendor Lock-in**: Organizations need to be aware of the potential risks of vendor lock-in when using cloud services, and ensure that they have the ability to switch to a different provider if necessary. This can be achieved through the use of open standards and interoperable technologies.

Overall, it is important for organizations to carefully consider the legal and ethical implications of using cloud computing, and to have policies and procedures in place to ensure compliance and protect the interests of their stakeholders.

Energy Consumption Issues in Cloud Computing

One of the main issues with cloud computing is the amount of energy that is required to power and cool the large data centers that host cloud services. These data centers consume vast amounts of electricity to run servers, storage devices, networking equipment, and other infrastructure, and this energy consumption has a significant impact on the environment. There are several reasons why energy consumption is a concern in cloud computing.

Firstly, the energy required to run data centers is typically generated from fossil fuels, which contribute to greenhouse gas emissions and climate change.

Secondly, the process of generating and transmitting electricity also has environmental impacts, such as air and water pollution. To address these issues, many organizations are taking steps to reduce their energy consumption in the cloud. This can include things like using more efficient servers and other hardware, implementing power management technologies, and using renewable energy sources to power data centers. Additionally, some organizations are adopting "green" cloud services, which are designed to be more environmentally friendly.

Overall, it is important for organizations to consider the environmental impact of their use of cloud computing, and to take steps to reduce their energy consumption and minimize their impact on the environment.

Key Challenges in Maintaining High-Scale Information in Cloud Computing

Maintaining high-scale information in cloud computing can be challenging for several reasons. These challenges include:

- **Data Storage and Management:** Storing large amounts of data in the cloud can be difficult, and organizations need to have effective tools and processes in place to manage this data. This can include things like data replication, backup and recovery, and data security.
- **Data Access and Performance:** Ensuring that users can access large amounts of data quickly and efficiently can be challenging, particularly when dealing with high volumes of data or a large number of users. This can require the use of specialized technologies and techniques, such as caching and load balancing.
- **Data Integration and Interoperability:** Maintaining high-scale information in the cloud often involves integrating data from multiple sources, which can be difficult due to differences in data formats, structures, and schemas. Organizations need to have effective data integration and interoperability solutions in place to ensure that their data is consistent and can be accessed and used by different systems and applications.
- **Data Governance and Compliance:** Maintaining high-scale information in the cloud also requires effective governance and compliance processes, to ensure that the data is accurate, complete, and secure and that it complies with relevant laws and regulations.

Overall, maintaining high-scale information in the cloud requires a combination of effective technology, processes, and people, and organizations need to carefully plan and manage these factors to ensure the success of their cloud initiatives.

Amazon Web Service

Amazon Web Services (AWS) is a cloud computing platform offered by *Amazon.com*. It provides a wide range of services, including computing, storage, database, networking, analytics, machine learning, security, and more. AWS allows organizations to access these services on a pay-as-you-go basis, without the need to invest in hardware or software. AWS is widely used by organizations of all sizes, across a range of industries. It is known for its scalability, reliability, and flexibility, and is often considered the market leader in cloud computing. AWS also offers a range of tools and services to help organizations manage and optimize their use of the platform, including training and support. Overall, AWS is a powerful and popular cloud computing platform that offers a wide range of services and features to help organizations build and operate their applications and services in the cloud.

Best Practices in Architecting Cloud Applications in the AWS Cloud

There are several best practices to consider when architecting a cloud application in the AWS cloud. Some of these include: Identifying the right AWS services and features for your specific application needs. This will help ensure that your application is scalable, highly available, and cost-effective. Designing your application to be resilient and fault-tolerant. This means using multiple Availability Zones and/or Regions, as well as implementing self-healing mechanisms to automatically recover from failures. Implementing security best practices, including using IAM to control access to AWS resources, encrypting data at rest and in transit, and regularly monitoring and auditing your application. Using automation to manage and deploy your application, such as using AWS Cloud Formation or the AWS CLI to automate the provisioning and management of AWS resources. Monitoring and optimizing the performance and cost of your application, including using AWS tools like Amazon Cloud Watch and AWS Trusted Advisor to monitor performance and identify opportunities for cost savings. Adopting a DevOps approach to continuously improve and iterate on your application, including using AWS services like AWS Code Pipeline and AWS Code Build for continuous integration and delivery. By following these best practices, you can help ensure that your cloud application is well-architected and able to meet the demands of your users and business.

Open-Source Cloud Service

An open-source cloud service is a cloud computing service that uses open-source software and is typically delivered using a pay-as-you-go or subscription-based pricing model. Some examples of open-source cloud services include OpenStack, CloudFoundry, and OpenShift. These services provide many of the same benefits as other cloud services, such as scalability, flexibility, and cost-efficiency, but with the added advantage of being able to access and modify the underlying source code. This can be beneficial for organizations that want to have more control over the technology they use, as well as those that want to contribute to and improve the open-source software ecosystem

VM Migration Explanation

VM (Virtual Machine) migration is the process of moving a virtual machine from one host or physical server to another. This can be done for various reasons, such as to upgrade the hardware of the host, to balance workloads across multiple hosts, or to move the virtual machine to a different location.

There are several methods for migrating a VM, including live migration and offline migration. Live migration allows the VM to be moved while it is still running, without any downtime or disruption to the applications or services running inside the VM. Offline migration involves shutting down the VM, transferring its disk files to the new host, and then starting it up again on the new host.

To perform a VM migration, you will need a hypervisor (such as VMware vSphere or Microsoft Hyper-V) that supports the migration of VMs and the necessary tools and utilities to manage the migration process. You may also need to consider factors such as network connectivity, storage resources, and security when planning the migration.

There are two main methods for migrating a virtual machine: live migration and offline migration.

Live migration: Live migration allows the virtual machine to be moved from one host to another while it is still running, without any downtime or disruption to the applications or services running inside the VM. This is possible because the hypervisor can temporarily pause the VM's execution on the source host, transfer its state and memory contents to the destination host, and then resume execution on the destination host. Live migration is typically used when it is important to minimize downtime or disruption to the VM's workloads.

Offline migration: Offline migration involves shutting down the virtual machine, transferring its disk files to the new host, and then starting it up again on the new host. This requires some downtime while the VM is shut down and its disk files are transferred, but it can be a simpler and more straightforward process than live migration. Offline migration is typically used when the workloads running inside the VM can tolerate some downtime, or when the source and destination hosts are not compatible with live migration.

VM transferring and task transferring

Task transferring refers to the process of moving a task or workload from one device or system to another. This can be done for various reasons, such as to distribute workloads across multiple devices to improve performance or to move a task to a device that is better suited to handle it. Task transferring can be done manually or automatically, depending on the needs and capabilities of the devices involved.

What is load balancing ?

Load balancing is the method of distributing network traffic equally across a pool of resources that support an application. Modern applications must process millions of users simultaneously and return the correct text, videos, images, and other data to each user in a fast and reliable manner. To handle such high volumes of traffic, most applications have many resource servers with duplicate data between them. A load balancer is a device that sits between the user and the server group and acts as an invisible facilitator, ensuring that all resource servers are used equally.

What are the types of load balancing ?

We can classify load balancing into three main categories depending on what the load balancer checks in the client request to redirect the traffic.

1. Application load balancing

Complex modern applications have several server farms with multiple servers dedicated to a single application function. Application load balancers look at the request content, such as HTTP headers or SSL session IDs, to redirect traffic.

For example, an ecommerce application has a product directory, shopping cart, and checkout functions. The application load balancer sends requests for browsing products to servers that contain images and videos but do not need to maintain open connections. By comparison, it sends shopping cart requests to servers that can maintain many client connections and save cart data for a long time.

2. Network load balancing

Network load balancers examine IP addresses and other network information to redirect traffic optimally. They track the source of the application traffic and can assign a static IP address to several servers. Network load balancers use the static and dynamic load balancing algorithms described earlier to balance server load.

3. Global server load balancing

Global server load balancing occurs across several geographically distributed servers. For example, companies can have servers in multiple data centers, in different countries, and in third-party cloud providers around the globe. In this case, local load balancers manage the application load within a region or zone. They attempt to redirect traffic to a server destination that is geographically closer to the client. They might redirect traffic to servers outside the client’s geographic zone only in case of server failure.

What are the types of load balancing technology ?

Load balancers are one of two types: hardware load balancer and software load balancer.

1. Hardware load balancers

A hardware-based load balancer is a hardware appliance that can securely process and redirect gigabytes of traffic to hundreds of different servers. You can store it in your data centers and use virtualization to create multiple digital or virtual load balancers that you can centrally manage.

2. Software load balancers

Software-based load balancers are applications that perform all load-balancing functions. You can install them on any server or access them as a fully managed third-party service.

A centralized load-balancing algorithm is a type of load-balancing algorithm that uses a central controller or server to distribute workload among a group of servers or resources. The central controller is responsible for receiving requests from clients, determining the appropriate server or resource to handle the request, and forwarding the request to that server or resource.

3. Hybrid Load Balancers

There are several different types of **centralized load-balancing** algorithms, including:

i. **Round-robin:** In this algorithm, the central controller cycles through the available servers or resources, sending each new request to the next server in the list. This ensures that each server or resource receives an equal number of requests unless one of the servers becomes unavailable.

ii. **Osmosis load balancing:** It is a type of load balancing algorithm that aims to evenly distribute workload among a group of servers or resources by using a feedback mechanism to adjust the load on each server or resource based on their current workload and capacity. The algorithm adjusts the load on each server or resource in a way that is analogous to the way that osmosis, a biological process, works.

iii. Max-min load balancing: It is a type of load balancing algorithm that aims to evenly distribute workload among a group of servers or resources by assigning the maximum amount of work to the least loaded server or resource, and the minimum amount of work to the most loaded server or resource. The algorithm continually adjusts the workload on each server or resource to try to keep all of the servers or resources as evenly loaded as possible.

iv. LBA, or Load Balancing Algorithm: It is a general term that refers to any algorithm that is used to distribute workload among a group of servers or resources. There are many different types of load-balancing algorithms, including centralized algorithms (such as round-robin and least connections) and decentralized algorithms (such as DNS-based load balancing and anycast load balancing). The specific algorithm that is used will depend on the needs of the system and the resources that are available.

Decentralized load-balancing algorithms are used to distribute workloads among a group of servers or other computing resources in a distributed system. There are many different algorithms that can be used for this purpose, and they differ in terms of their approach and the specific characteristics they are designed to optimize.

RILB (Randomized Incremental Load Balancing) and SILB (Self-organizing Incremental Load Balancing).

i. **RILB** is a decentralized load-balancing algorithm that uses randomization to distribute load among a group of servers or nodes in a distributed system. It works by randomly selecting a server from the pool of available servers and assigning a new request to it. If the selected server is already heavily loaded, the request is forwarded to another server. This process is repeated until a server is found that can handle the request.

ii. SILB is a self-organizing version of RILB that aims to improve the efficiency of the load-balancing process by using a decentralized, adaptive approach. It works by continuously monitoring the workload of each server in the system and adjusting the load distribution accordingly. If a server becomes overloaded, it can request additional resources from other servers in the system to help balance the load.

Both RILB and SILB are designed to be scalable and fault-tolerant, making them suitable for use in distributed systems with a large number of servers or nodes. However, they may not be as effective as more sophisticated load-balancing algorithms in certain situations, such as when the workload is highly variable or the servers have different processing capabilities.

Energy Issues in Cloud Computing

There are several energy issues in cloud computing that can arise due to cloud servers running for longer periods of time. For example:

- **Sustainability:** The energy required to power and cool cloud servers can have significant environmental impacts, especially if the servers are running for long periods of time. This can be a concern from a sustainability perspective, as it may contribute to carbon emissions and other environmental problems.
- **Environmental issues:** The energy consumption of cloud servers can also contribute to air pollution and other environmental problems, as it may require the use of fossil fuels or other non-renewable energy sources.
- **Increased functional cost:** The energy required to run cloud servers for extended periods of time can also increase the functional cost of the servers. This can be a concern for cloud computing providers, as it may impact their profitability.
- **Reduced profit margins:** The increased functional cost of running cloud servers for longer periods of time can also lead to reduced profit margins for cloud computing providers. This can be a concern for businesses that rely on cloud computing services, as it may impact their bottom line.

How to handle energy issues:

There are several strategies that can be used to address the energy issues in cloud computing that can arise due to cloud servers running for longer periods of time:

- **Use renewable energy sources:** One way to address the energy issues associated with cloud computing is to use renewable energy sources to power and cool the servers. This can help to reduce the environmental impact of cloud computing and improve sustainability.
- **Develop more energy-efficient hardware and software: Another** approach is to focus on the development of more energy-efficient hardware and software for use in cloud computing. This can help to reduce the energy consumption of cloud servers and improve their efficiency.

- **Implement energy-efficient design principles in data centers:** Data centers can be designed and operated in ways that minimize their energy consumption and environmental impact. This can involve the use of energy-efficient hardware, the implementation of energy-efficient design principles, and the use of renewable energy sources to power the data center.
- **Use power management tools:** Cloud computing providers can also use power management tools to optimize the energy usage of their servers. This can involve adjusting the power settings of the servers, using power management software, or using power-saving modes to reduce energy consumption when the servers are not in use.
- **Use load balancing strategies:** Load balancing strategies can be used to distribute workloads among multiple servers, which can help to reduce the energy consumption of individual servers by ensuring that they are not overloaded. This can also help to improve the overall efficiency of the cloud computing system.

SLA (Service Level Agreement) is a contract between a cloud computing provider and a customer that defines the terms and conditions of the service being provided. SLAs typically specify the level of service that the provider will deliver, including the availability and performance of the service, and the terms under which the service will be provided.

In the context of cloud computing, an SLA might include details such as:

- The uptime and availability of the service
- The performance and response time of the service
- The level of support provided by the provider
- The terms of service, including any restrictions or limitations on the use of the service
- The terms of billing and payment

SLAs are important for cloud computing customers because they provide a clear understanding of the level of service that they can expect to receive, as well as the terms and conditions under which the service will be provided. They can also provide a basis for dispute resolution if there are any issues with the service.

- **Reliability:** the uptime and availability of the service. This can help to ensure that the service is consistently available for use.
- **Responsiveness:** the performance and response time of the service. This can help to ensure that the service performs well and responds promptly to user requests.
- **Availability:** the percentage of time that the service is expected to be available for use. This can help to ensure that the service is consistently available for use.
- **Accountability:** the terms of service and any restrictions or limitations on the use of the service. This can help to ensure that the provider is transparent about the terms of the service and that customers understand their rights and responsibilities.
- **Warranties:** the provider will be held responsible for any issues with the service. This can help to protect customers if there are any issues with the service.