LECTURENOTES

ON

CLOUD COMPUTING

(FOR 6TH SEMESTER CSE)

SUBMITED BY:

MR AMARDEEP DAS

ASST. PROF(CSE),CVRP

Introduction To Cloud Computing

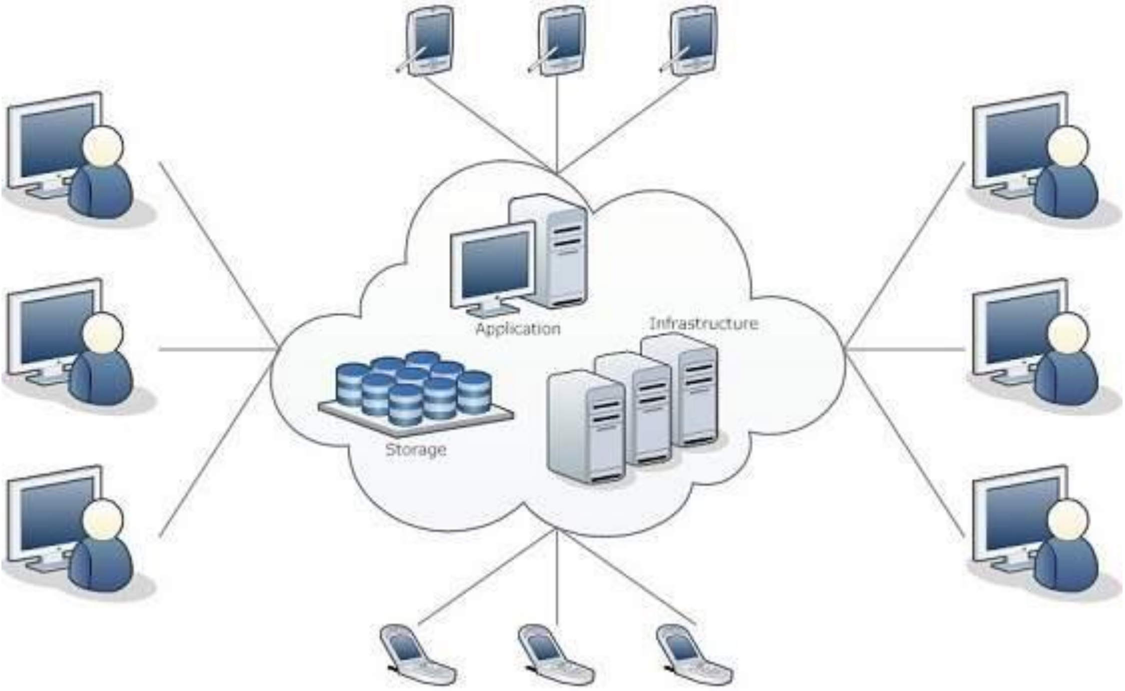
What is Cloud?

The term **Cloud** refers to a **Network** or **Internet.** In other words, we can say that Cloud is something, which is present at remote location. Cloud can provide services over public and private networks, i.e., WAN, LAN or VPN.

Applications such as e-mail, web conferencing, customer relationship management (CRM) execute on cloud.

What is Cloud Computing?

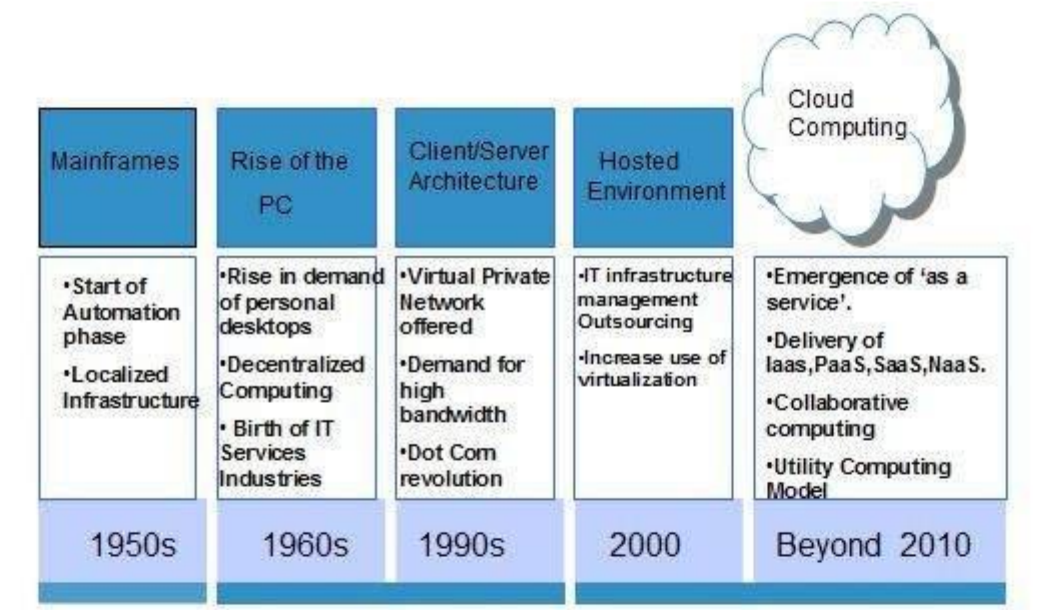
Cloud Computing refers to **manipulating, configuring,** and **accessing** the hardware and software resources remotely. It offers online data storage, infrastructure, and application.



Cloud computing offers **platform independency,** as the software is not required to be installed locally on the PC. Hence, the Cloud Computing is making our business applications **mobile** and **collaborative.**

History of Cloud Computing

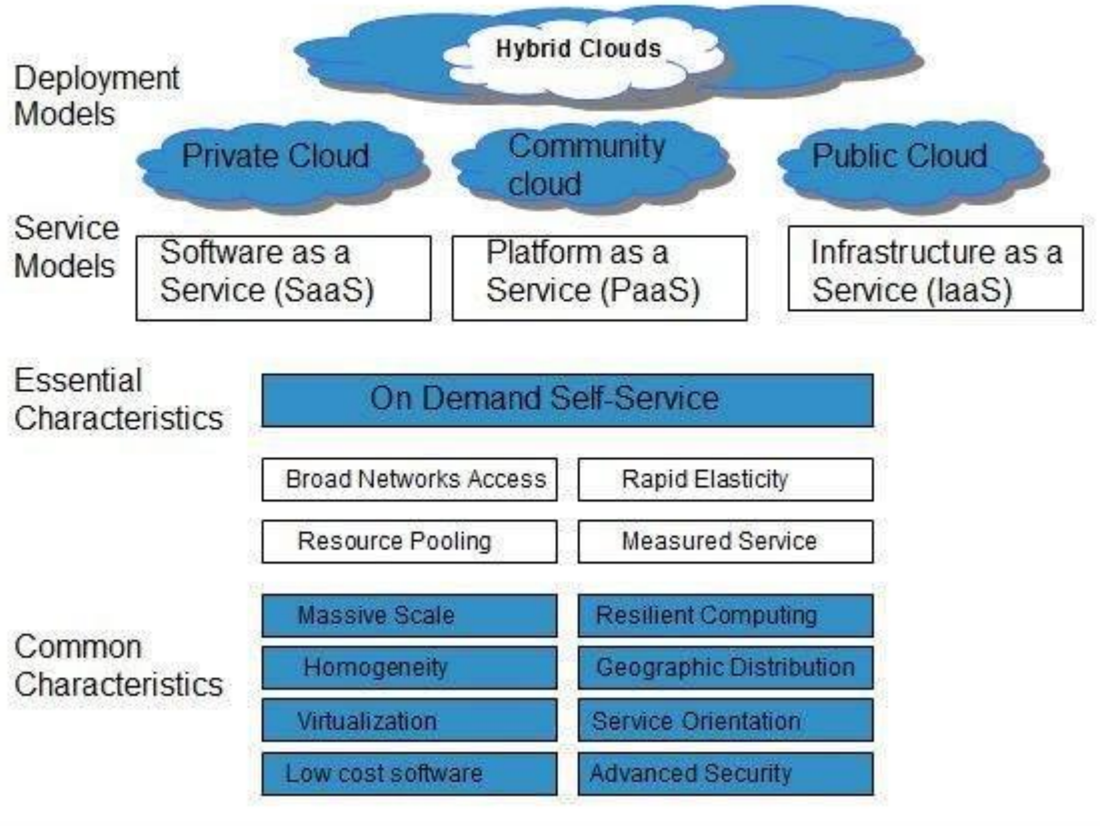
The concept of **Cloud Computing** came into existence in the year 1950 with implementation of mainframe computers, accessible via **thin/static clients.** Since then, cloud computing has been evolved from static clients to dynamic ones and from software to services. The following diagram explains the evolution of cloud computing:



Mainframes
Rise of the PC
Client/Server Architecture
Hosted Environment
Cloud Computing
•Start of Automation phase
•Localized Infrastructure
•Rise in demand of personal desktops
•Decentralized Computing
• Birth of IT Services Industries
•Virtual Private Network offered
•Demand for high bandwidth
•Dot Com revolution
•IT infrastructure management Outsourcing
•Increase use of virtualization
•Emergence of 'as a service'.
•Delivery of Iaas,PaaS,SaaS,NaaS.
•Collaborative computing
•Utility Computing Model
1950s
1960s
1990s
2000
Beyond 2010

Characteristics of Cloud Computing

There are four key characteristics of cloud computing. They are shown in the following diagram:



On Demand Self Service

Cloud Computing allows the users to use web services and resources on demand. One can logon to a website at any time and use them.

Broad Network Access

Since cloud computing is completely web based, it can be accessed from anywhere and at any time.

Resource Pooling

Cloud computing allows multiple tenants to share a pool of resources. One can share single physical instance of hardware, database and basic infrastructure.

Rapid Elasticity

It is very easy to scale the resources vertically or horizontally at any time. Scaling of resources means the ability of resources to deal with increasing or decreasing demand.

The resources being used by customers at any given point of time are automatically monitored.

Measured Service

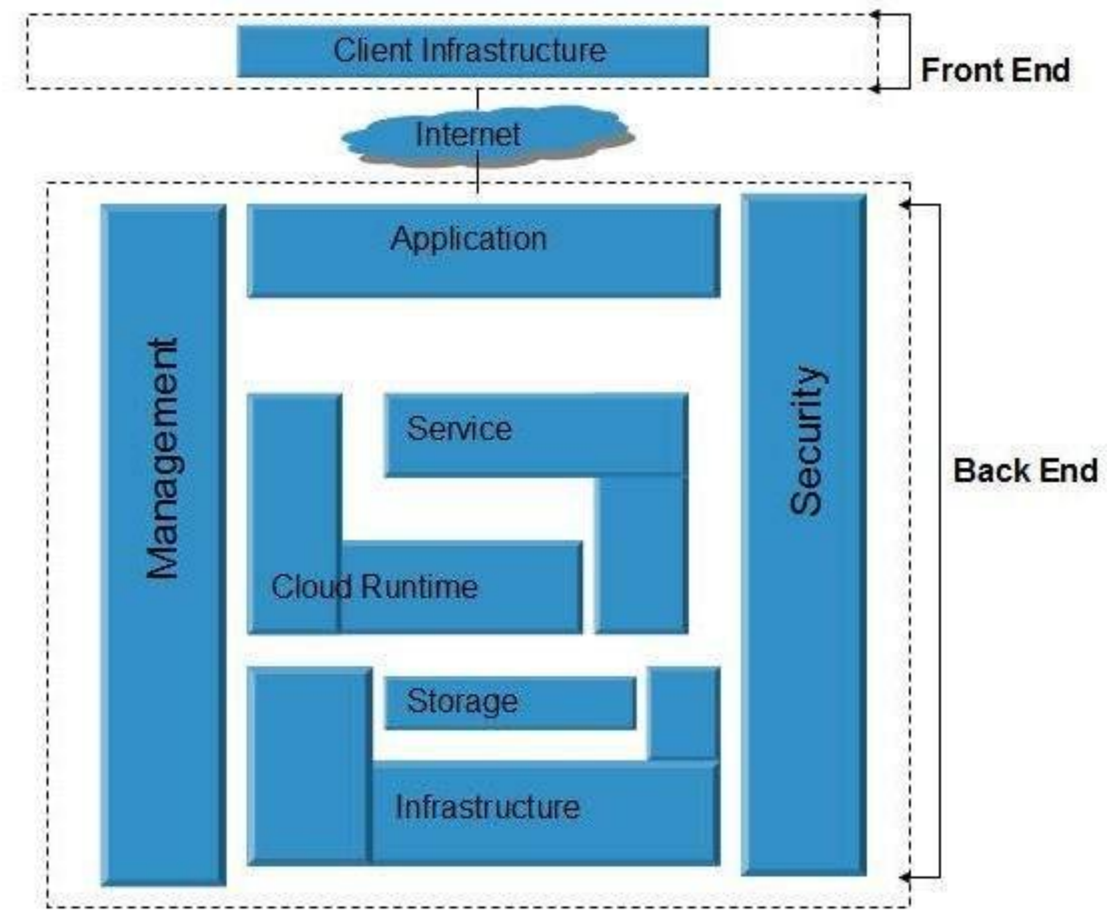
In this service cloud provider controls and monitors all the aspects of cloud service. Resource optimization, billing, and capacity planning etc. depend on it.

Cloud Computing Architecture

Cloud Computing architecture comprises of many cloud components, which are loosely coupled. We can broadly divide the cloud architecture into two parts:

- Front End
- Back End

Each of the ends is connected through a network, usually Internet. The following diagram shows the graphical view of cloud computing architecture:



Front End

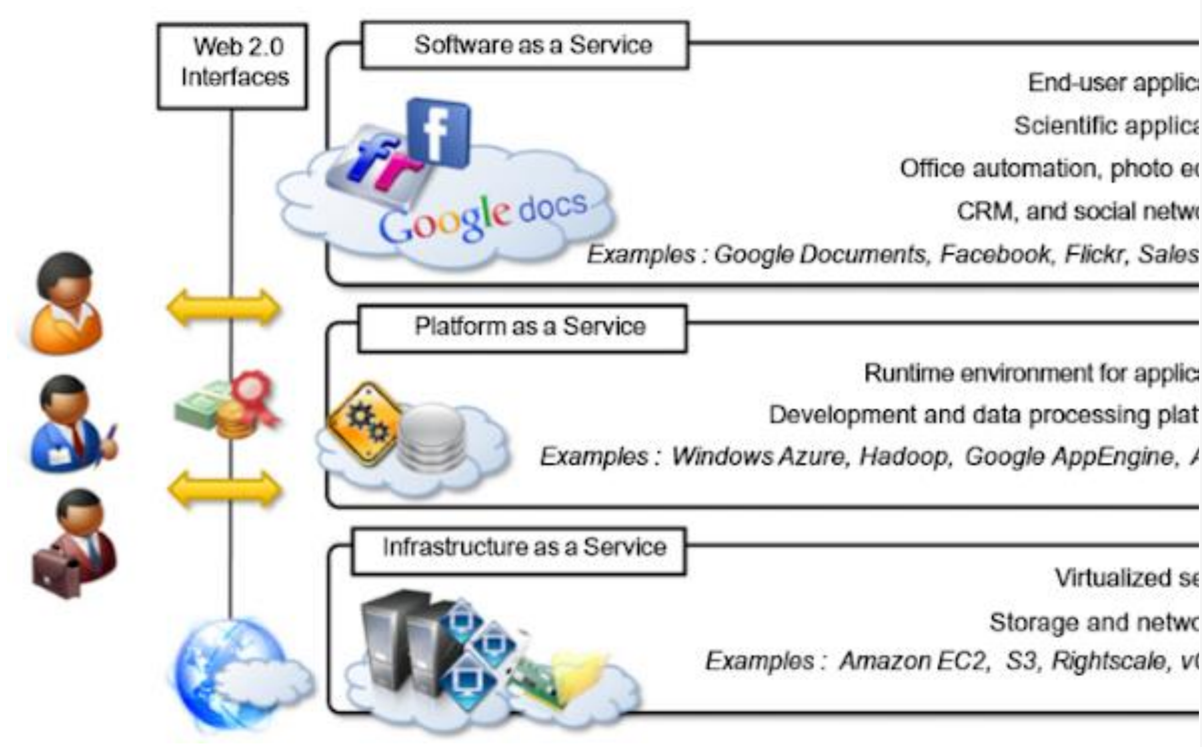
The **front end** refers to the client part of cloud computing system. It consists of interfaces and applications that are required to access the cloud computing platforms, Example - Web Browser.

Back End

The **back End** refers to the cloud itself. It consists of all the resources required to provide cloud computing services. It comprises of huge data storage, virtual machines, security mechanism, services, deployment models, servers, etc.

Cloud Reference Model

The reference model for cloud computing is an abstract model that characterizes and standardizes a cloud computing environment by partitioning it into abstraction layers and cross-layer functions.


Web 2.0 Interfaces
Software as a Service
End-user applic
Scientific applica
Office automation, photo ed
CRM, and social netw
Examples : Google Documents, Facebook, Flickr, Sales
Platform as a Service
Runtime environment for applic
Development and data processing plat
Examples : Windows Azure, Hadoop, Google AppEngine, A
Infrastructure as a Service
Virtualized se
Storage and netwo
Examples : Amazon EC2, S3, Rightscale, v

From the book of Sir Rajkumar Buyya
Cloud computing reference model

If we look in to the reference model as seen in above image we will find classification of Cloud Computing services:

Infrastructure-as-a-Service (IaaS),

Platform-as-a-Service (PaaS), and

Software-as-a-Service (SaaS).

Web 2.0

1. Infrastructure as a service (IaaS) is a cloud computing offering in which a vendor provides users access to computing resources such as servers, storage and networking. To read more about IaaS click here.

2. Platform as a service (PaaS) is a cloud computing offering that provides users with a cloud environment in which they can develop, manage and deliver applications. To read more about PaaS click here.

3. Software as a service (SaaS) is a cloud computing offering that provides users with access to a vendor’s cloud-based software. Users do not install applications on their local devices. Instead, the applications reside on a remote cloud network accessed through the web or an API. Through the application, users can store and analyze data and collaborate on projects. To read more about SaaS click here.

4. Web 2.0 is the term used to describe a variety of web sites and applications that allow anyone to create and share online information or material they have created. A key element of the technology is that it allows people to create, share, collaborate & communicate.

Types of Clouds

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

Private Cloud: Here, computing resources are deployed for one particular organization. This method is more used for intra-business interactions. Where the computing resources can be governed, owned and operated by the same organization.

Community Cloud: Here, computing resources are provided for a community and organizations.

Public Cloud: This type of cloud is used usually for B2C (Business to Consumer) type interactions. Here the computing resource is owned, governed and operated by government, an academic or business organization.

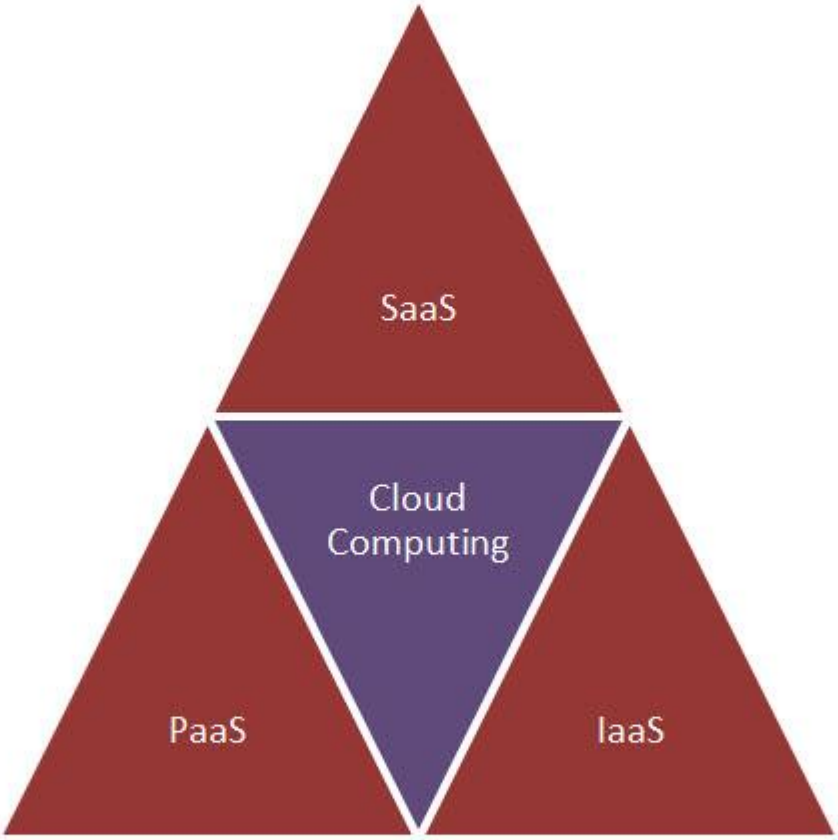
Hybrid Cloud: This type of cloud can be used for both type of interactions - B2B (Business to Business) or B2C (Business to Consumer). This deployment method is called hybrid cloud as the computing resources are bound together by different clouds.

Cloud Computing Services

The three major Cloud Computing Offerings are

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

Different business use some or all of these components according to their requirement.



SaaS (Software as a Service)

SaaS or software as a service is a software distribution model in which applications are hosted by a vendor or service provider and made available to customers over a network (internet). SaaS is becoming an increasingly prevalent delivery model as underlying technologies that supports **Service Oriented Architecture (SOA) or Web Services**. Through internet this service is available to users anywhere in the world.

Traditionaly, software application needed to be purchased upfront &then installed it onto your computer. SaaS users on the other hand, instead of purchasing the software subscribes to it, usually on monthly basisvia internet.

Anyone who needs an access to a particular piece of software can be subscribe as a user, whether it is one or two people or every thousands of employees in a corporation. SaaS is compatible with all internet enabled devices.

Many important tasks like accounting, sales, invoicing and planning all can be performed using SaaS.

PaaS (Platform as a Service)

Platform as a service, is referred as PaaS, it provides a platform and environment to allow developers to build applications and services. This service is hosted in the cloud and accessed by the users via internet.

To understand in a simple terms, let compare this with painting a picture, where you are provided with paint colors, different paint brushes and paper by your school teacher and you just have to draw a beautiful picture using those tools.

Platform Computing can be compared to your painting class where the teacher gives you paints, brushes etc as a platform to create your painting

PaaS services are constantly updated & new features added. Software developers, web developers and business can benefit from PaaS. It provides platform to support application development. It includes software support and management services, storage, networking, deploying, testing, collaborating, hosting and maintaining applications.

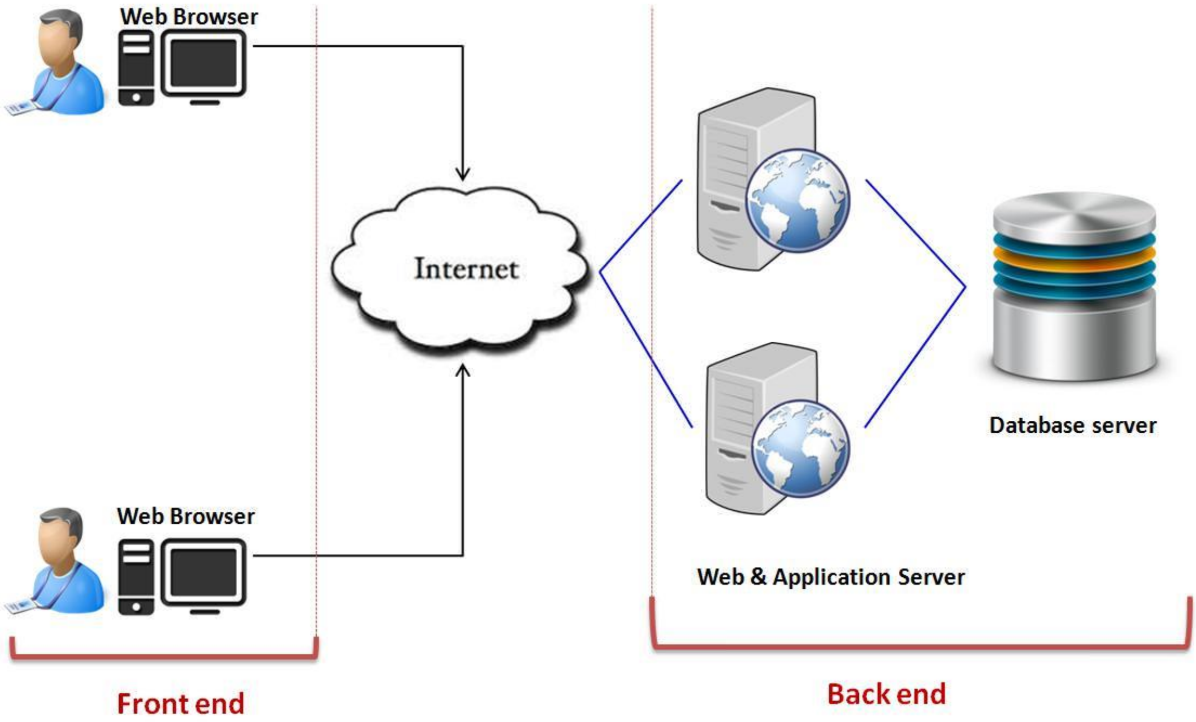
IaaS (Infrastructure as a Service)

IaaS (Infrastructure As A Service) is one of the fundamental service model of cloud computing alongside PaaS(Platform as a Service). It provides access to computing resources in a virtualized environment “the cloud” on internet. It provides computing infrastructure like virtual server space, network connections, bandwidth, load balancers and IP addresses. The pool of hardware resource is extracted from multiple servers and networks usually distributed across numerous data centers. This provides redundancy and reliability to IaaS.

IaaS(Infrastructure as a service) is a complete package for computing. For small scale businesses who are looking for cutting cost on IT infrastructure, IaaS is one of the solutions. Annually a lot of money is spent in maintenance and buying new components like hard-drives, network connections, external storage device etc. which a business owner could have saved for other expenses by using IaaS.

What is Cloud Computing Architecture?

Let’s have a look into Cloud Computing and see what Cloud Computing is made of. Cloud computing comprises of two components front end and back end. Front end consist client part of cloud computing system. It comprise of interfaces and applications that are required to access the cloud computing platform.



While back end refers to the cloud itself, it comprises of the resources that are required for cloud computing services. It consists of virtual machines, servers, data storage, security mechanism etc. It is under providers control.

Cloud computing distributes the file system that spreads over multiple hard disks and machines. Data is never stored in one place only and in case one unit fails the other will take over automatically. The user disk space is allocated on the distributed file system, while another important component is algorithm for resource allocation. Cloud computing is a strong distributed environment and it heavily depends upon strong algorithm.

Cloud Scalability and Fault Tolerance

Cloud Scalability

In cloud computing, cloud scalability refers to the ability to increase or reduce IT resources as required to meet evolving demands. One of the hallmarks of the cloud and the key factor of its burgeoning popularity with companies is scalability.

Using existing cloud computing technology, data storage space, processing power and networking can all be escalated. Better still, scaling, usually with little or no interruption or downtime, can be achieved rapidly and easily. Third-party cloud providers now have the entire infrastructure in place; in the past, the process could take weeks or months to scale with on-site physical infrastructure and entail enormous costs.

1.1 How to achieve cloud scalability?

To set up a personalized, scalable cloud solution via a public cloud, private cloud, or hybrid cloud, businesses have several options.

In cloud computing, two specific forms of scalability exist vertical and horizontal scaling.

We can add or subtract power to an existing cloud server memory upgrade, storage, or computing power with vertical scaling, also known as "scaling up" or "scaling down". This generally indicates that scaling has an upper limit based on the scaling capability of the server or machine; scaling above that also includes downtime.

We can add more resources like servers to our system using horizontal scalability to spread the workload across computers, which in turn improves efficiency and storage space. For companies with high-availability services that need limited downtime, horizontal scaling is essential.

2. Cloud Fault Tolerance

In cloud computing, fault tolerance is conceptually the same as in private or hosted environments. In other words, it means the infrastructure's ability to continue to provide service/services to underlying applications even when one or more component fails. To continue to work through failure or repair, we do not need to configure certain facilities for our infrastructure to use.

2.1 Objectives of Fault Tolerance in Cloud Computing

The fault-tolerant system uses backup components that take the place of failed components automatically, ensuring no service loss. They include:

- **Hardware systems**
  Hardware systems can be backed up using identical or equivalent systems. For instance, using an identical server running in parallel, with all operations mirrored to the backup server, a server can be made fault-tolerant.
- **Software systems**
  Software systems can be backed up using software instances. For example, it is possible to continuously replicate a database with customer information on another machine and operations can be mechanically redirected to another database in case a primary database goes down.
- **Power sources**
  Power sources use alternative sources using fault-tolerant. In many instances, organizations have power generators that can be used in case the electricity fails.
  Similarly, using redundancy, any system or component that is a single point of failure can be made fault-tolerant.
- **Security Breach Occurrences**
  Owing to security failures, there are many explanations about why fault tolerance exists. The server's hacking adversely affects the server and results in a leak of data. Ransomware, phishing, virus attack, etc. are other explanations for the need for fault tolerance in the form of security violations.

Key principles behind Cloud Computing Device Fault Tolerance

- **Replication**
  For every operation, the fault-tolerant system operates on the principle of running many other replicates. Therefore, if one aspect of the device goes wrong, it has other instances that can be put to keep it going instead. For example, a database of clusters that has 3 servers with the

same information on each of them. All the acts are written on each of them, such as adding data, upgrading, and deleting. The redundant servers will be in inactive mode unless and until the availability of them is requested by any fault tolerance scheme.

- **Redundancy**
  If any part of the system fails or moves to a downstate, then it is necessary to have backup systems. For example, due to some hardware faults, a website programmer that has MS SQL as its database can fail in between. In the redundancy principle, a server works with an emergency database comprising many backup resources.

Cloud business process process management is usually a platform-as-a-service solution that lets you create workflows and optimise them. Without having to install a single Mb of software on your office hardware, you can use these cloud-based software solutions to streamline and optimise everyday business activities.

Cloud Computing Portability and Interoperability – Cloud Portability and Interoperability

Portability and interoperability relate to the ability to build systems from re-usable components that will work together "out of the box".

A particular concern for cloud computing is cloud on-boarding – the deployment or migration of systems to a cloud service or set of cloud services. A common scenario is that some components cannot be moved to the cloud; for example, because of requirements for the enterprise to have complete control over personal data. On-boarding requires portability of those components that can be moved to the cloud, and interoperability with them of components that remain on in-house systems.

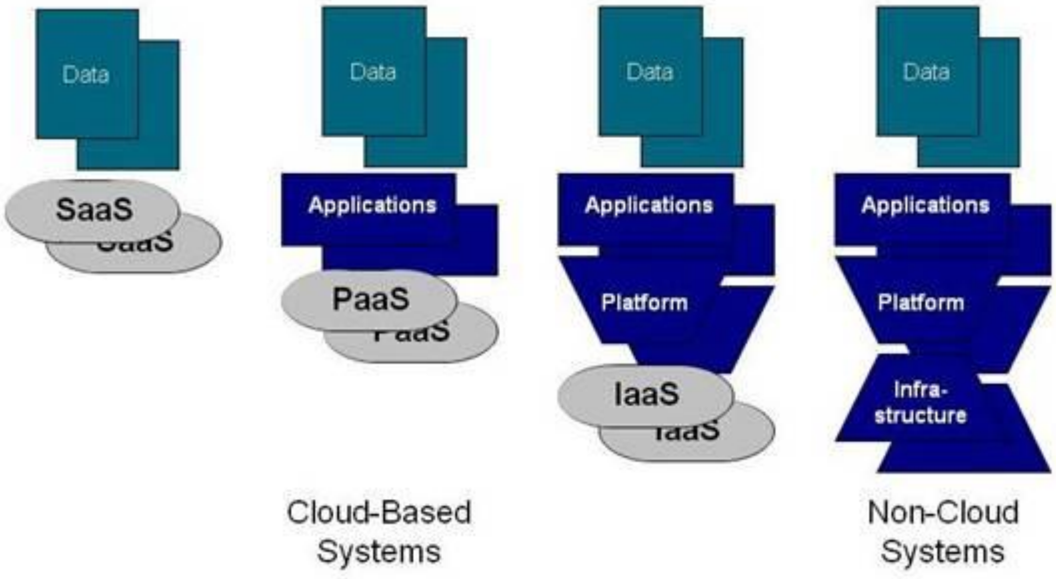
The Important Categories of Cloud Computing Portability and Interoperability

A system that involves cloud computing typically includes data, application, platform, and infrastructure components, where:

- Data is the machine-processable representation of information, held in computer storage.

- Applications are software programs that perform functions related to business problems.
- Platforms are programs that support the applications and perform generic functions that are not business-related.
- Infrastructure is a collection of physical computation, storage, and communication resources.

The application, platform, and infrastructure components can be as in traditional enterprise computing, or they can be cloud resources that are (respectively) software application programs (SaaS), software application platforms (PaaS), and virtual processors and data stores (IaaS).



Data, Applications, Platforms, and Infrastructure

Non-cloud systems include mainframes, minicomputers, personal computers, and mobile devices owned and used by enterprises and individuals.

Data components interoperate via application components rather than directly. There are no “data interoperability” interfaces.

The cloud computing portability and interoperability categories to consider are thus:

- Data Portability
- Application Portability
- Platform Portability
- Application Interoperability

- Platform Interoperability
- Management Interoperability
- Publication and Acquisition Interoperability

Data Portability

- Data portability enables re-use of data components across different applications.
- Suppose that an enterprise uses a SaaS product for Customer Relations Management (CRM), for example, and the commercial terms for use of that product become unattractive compared with other SaaS products or with use of an in-house CRM solution. The customer data held by the SaaS product may be crucial to the enterprise's operation. How easy will it be to move that data to another CRM solution?
- In many cases, it will be very difficult. The structure of the data is often designed to fit a particular form of application processing, and a significant transformation is needed to produce data that can be handled by a different product.
- This is no different from the difficulty of moving data between different products in a traditional environment. But, in a traditional environment, the customer is more often able to do nothing; to stay with an old version of a product, for example, rather than upgrading to a newer, more expensive one. With SaaS, the vendor can more easily force the customer to pay more or lose the service altogether.
- Cloud introduces no new technical problems, but its different commercial arrangements can make the old technical problems much more serious.

Application Portability

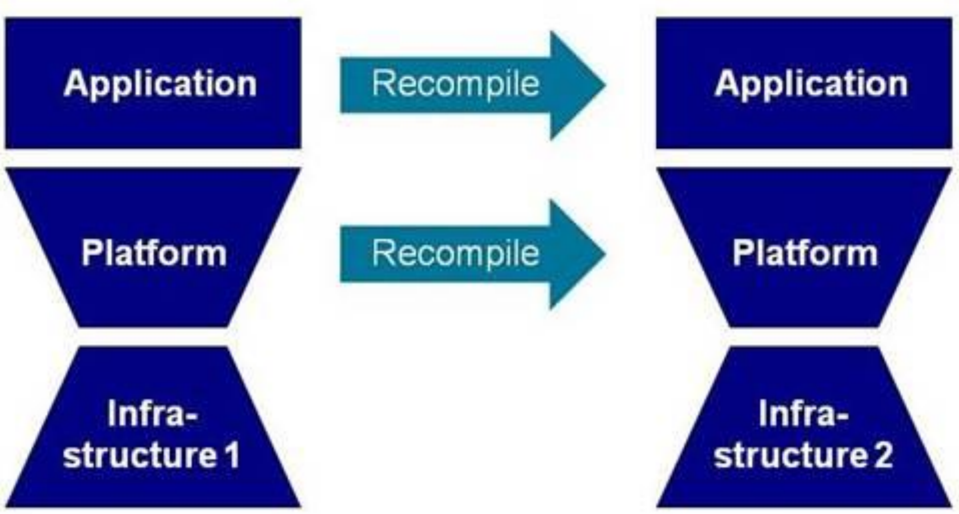
Application portability enables the re-use of application components across cloud PaaS services and traditional computing platforms.

Platform Portability

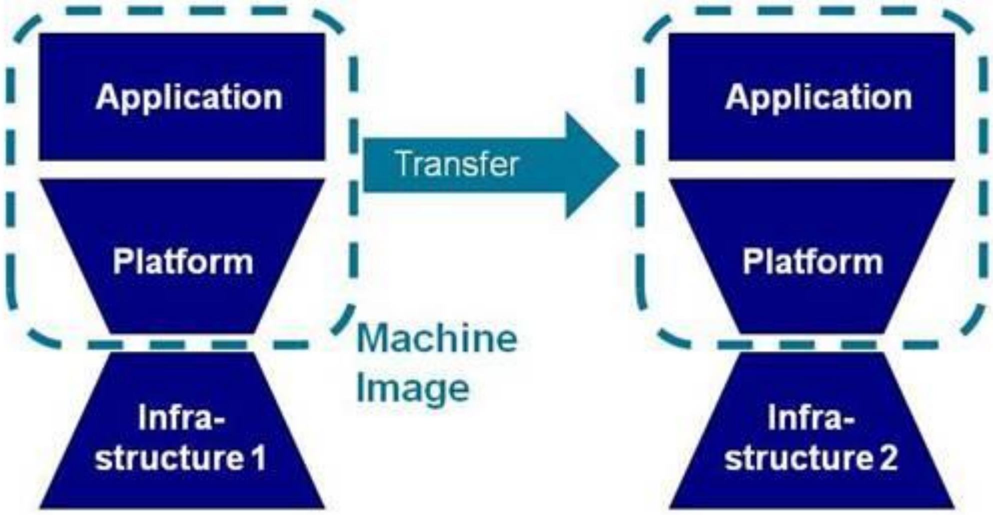
There are two kinds of platform portability:

- Re-use of platform components across cloud IaaS services and non-cloud infrastructure – platform source portability
- Re-use of bundles containing applications and data with their supporting platforms – machine image portability

- The UNIX operating system provides an example of platform source portability. It is mostly written in the C programming language, and can be implemented on different hardware by re-compiling it and re-writing a few small hardware-dependent sections that are not coded in C. Some other operating systems can be ported in a similar way. This is the traditional approach to platform portability. It enables applications portability because applications that use the standard operating system interface can similarly be re-compiled and run on systems that have different hardware. It is illustrated in Platform Source Portability.



Platform Source Portability



Machine Image Portability

Application Interoperability

Application interoperability is interoperability between application components deployed as SaaS, as applications using PaaS, as applications on platforms using IaaS, in a traditional enterprise IT environment, or on client devices. An application component may be a complete monolithic application, or a part of a distributed application.

. The design approach must address:

- Management of “system of record” sources
- Management of data at rest and data in transit across domains that may be under control of a cloud service consumer or provider
- Data visibility and transparency

Platform Interoperability

- Platform interoperability is interoperability between platform components, which may be deployed as PaaS, as platforms on IaaS, in a traditional enterprise IT environment, or on client devices.
- Platform interoperability requires standard protocols for service discovery and information exchange. As discussed above, these indirectly enable interoperability of the applications that use the platforms. Application interoperability cannot be achieved without platform interoperability.

Management Interoperability

- Management interoperability is interoperability between cloud services (SaaS, PaaS, or IaaS) and programs concerned with the implementation of on-demand self-service.
- As cloud computing grows, enterprises will want to manage cloud services together with their in-house systems, using generic off-the-shelf systems management products. This can only be achieved if cloud services have standard interfaces.
- These interoperability interfaces may provide the same functionality as the management interfaces mentioned under Application Portability.

Virtual desktop Infrastructure

What is VDI (Virtual Desktop Infrastructure)?

- Virtual desktop infrastructure (VDI) is a technology that refers to the use of virtual machines to provide and manage virtual desktops. VDI hosts desktop environments on a centralized server and deploys them to end-users on request.

How does VDI work?

In VDI, a hypervisor segments servers into virtual machines that in turn host virtual desktops, which users access remotely from their devices. Users can access these virtual desktops from any device or location, and all processing is done on the host server. Users connect to their desktop instances through a connection broker, which is a software-based gateway that acts as an intermediary between the user and the server.

VDI can be either persistent or nonpersistent. Each type offers different benefits:

With persistent VDI, a user connects to the same desktop each time, and users are able to personalize the desktop for their needs since changes are saved even after the connection is reset. In other words, desktops in a persistent VDI environment act exactly like a personal physical desktop.

In contrast, nonpersistent VDI, where users connect to generic desktops and no changes are saved, is usually simpler and cheaper, since there is no need to maintain customized desktops between sessions. Nonpersistent VDI is often used in organizations with a lot of task workers, or employees who perform a limited set of repetitive tasks and don’t need a customized desktop.

Why VDI?

VDI offers a number of advantages, such as user mobility, ease of access, flexibility and greater security. In the past, its high-performance requirements made it costly and challenging to deploy on legacy systems, which posed a barrier for many businesses. However, the rise in enterprise adoption of hyperconverged infrastructure (HCI) offers a solution that provides scalability and high performance at a lower cost.

What are the benefits of VDI?

Although VDI's complexity means that it isn't necessarily the right choice for every organization, it offers a number of benefits for organizations that do use it. Some of these benefits include:

- Remote access: VDI users can connect to their virtual desktop from any location or device, making it easy for employees to access all their files and applications and work remotely from anywhere in the world.
- Cost savings: Since processing is done on the server, the hardware requirements for end devices are much lower. Users can access their virtual desktops from older devices, thin clients, or even tablets, reducing the need for IT to purchase new and expensive hardware.
- Security: In a VDI environment, data lives on the server rather than the end client device. This serves to protect data if an endpoint device is ever stolen or compromised.
- Centralized management: VDI's centralized format allows IT to easily patch, update or configure all the virtual desktops in a system.

What is VDI used for?

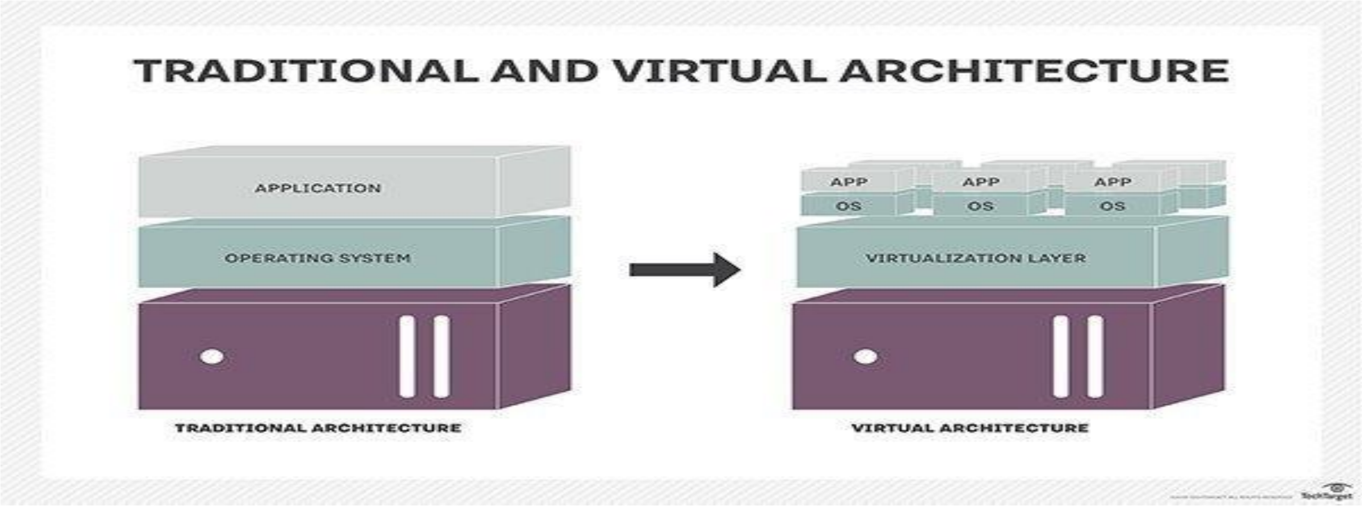
Although VDI can be used in all sorts of environments, there are a number of use cases that are uniquely suited for VDI, including:

- Remote work: Since VDI makes virtual desktops easy to deploy and update from a centralized location, an increasing number of companies are implementing it for remote workers.
- Bring your own device (BYOD): VDI is an ideal solution for environments that allow or require employees to use their own devices. Since processing is done on a centralized server, VDI allows the use of a wider range of devices. It also offers better security, since data lives on the server and is not retained on the end client device.
- Task or shift work: Nonpersistent VDI is particularly well suited to organizations such as call centers that have a large number of employees who use the same software to perform limited tasks.

Cloud Management and Virtualisation Technology

Create a virtualised Architecture

A virtualization architecture is a conceptual model specifying the arrangement and interrelationships of the particular components involved in delivering a virtual -- rather than physical -- version of something, such as an operating system (OS), a server, a storage device or network resources.



Data Centre

- A data center is a centralized location that houses the computing and network equipment required to collect, store, process, and distribute large amounts of data.

- Data center facilities are also critical in providing access to the large amounts of data stored there for employees running daily operations, applications, and other processes in a cloud computing environment.

- Modern data centers consist of a variety of infrastructure components, including servers and network connectivity equipment, that permit access to the server storage over the Internet

Resilience

Resiliency is the ability of a server, network, storage system, or an entire data center, to recover quickly and continue operating even when there has been an equipment failure, power outage or other disruption. ... When one server in the cluster fails, another node takes over. with its redundant workloads

Agility

- In the cloud computing context, agility often refers to the ability to rapidly develop, test and launch software applications that drive business growth.
- Cloud Agility allows them to focus on other issues such as security, monitoring and analysis, instead of provisioning and maintaining the resources.

Cloud storage

- Cloud storage allows you to save data and files in an off-site location that you access either through the public internet or a dedicated private network connection. ... Computer hard drives can only store a finite amount of data.
- When users run out of storage, they need to transfer files to an external storage device.

MapReduce

- MapReduce is a programming model introduced by Google for processing and generating large data sets on clusters of computers.
- Google first formulated the framework for the purpose of serving Google's Web page indexing, and the new framework replaced earlier indexing algorithms. Beginner developers find the MapReduce framework beneficial because library routines can be used to create parallel programs without any worries about infra-cluster communication, task monitoring or failure handling processes.
- MapReduce runs on a large cluster of commodity machines and is highly scalable. It has several forms of implementation provided by multiple programming languages, like Java, C# and C++.

Cloud Goverance

- Cloud Governance is a set of rules. It applies specific policies or principles to the use of cloud computing services.This model aims to secure applications and data even if located distantly. The best Cloud

Governance solutions include People, Processes, and Technology.
It basically refers to the decision making processes, criteria, and policies involved in the planning, architecture, acquisition, deployment, operation, architecture, acquisition, implementation, operation, and management of a Cloud computing capability.

- Cloud Governance best practices help to optimize the organization's:
- Operations: Doing it efficiently
- Risk and compliance: Doing it securely
- Financial: Doing more with less

Load Balancing

- Cloud load balancing is defined as the method of splitting workloads and computing properties in a cloud computing.
- It enables enterprise to manage workload demands or application demands by distributing resources among numerous computers, networks or servers.
- Cloud load balancing includes holding the circulation of workload traffic and demands that exist over the Internet.

Load balancing solutions can be categorized into two types –

1. Software-based load balancers: Software-based load balancers run on standard hardware (desktop, PCs) and standard operating systems.
2. Hardware-based load balancer: Hardware-based load balancers are dedicated boxes which include Application Specific Integrated Circuits (ASICs) adapted for a particular use. ASICs allows high speed promoting of network traffic and are frequently used for transport-level load balancing because hardware-based load balancing is faster in comparison to software solution.

Major Examples of Load Balancers –

1. Direct Routing Requesting Dispatching Technique: This approach of request dispatching is like to the one implemented in IBM's Net Dispatcher. A real server and load balancer share the virtual IP address. In this, load balancer takes an interface constructed with the virtual IP address that accepts request packets and it directly routes the packet to the selected servers.
2. Dispatcher-Based Load Balancing Cluster: A dispatcher does smart load balancing by utilizing server availability, workload, capability and other user-defined criteria to regulate where to send a TCP/IP request. The

dispatcher module of a load balancer can split HTTP requests among various nodes in a cluster. The dispatcher splits the load among many servers in a cluster so the services of various nodes seem like a virtual service on an only IP address; consumers interrelate as if it were a solo server, without having an information about the back-end infrastructure.

3. Linux Virtual Load Balancer: It is an opensource enhanced load balancing solution used to build extremely scalable and extremely available network services such as HTTP, POP3, FTP, SMTP, media and caching and Voice Over Internet Protocol (VoIP). It is simple and powerful product made for load balancing and fail-over. The load balancer itself is the primary entry point of server cluster systems and can execute Internet Protocol Virtual Server (IPVS), which implements transport-layer load balancing in the Linux kernel also known as Layer-4 switching.

<u>High Availability</u>

- High Availability in the cloud is achieved by creating clusters.
- A high availability cluster is a group of servers that act as a single server to provide continuous uptime.
- These servers will have access to the same shared storage for data, so if a server is unavailable, the other servers pick up the load.

Four Steps to Achieving High Availability in the Cloud

- Build for server failure. Instances in the cloud — just as in a typical data center — are ephemeral. ...
- Build for zone failure. ...
- Build for cloud failure. ...
- Automate and test everything.

<u>How does cloud disaster recovery work?</u>

Cloud disaster recovery takes a very different approach than traditional DR.

Instead of loading the servers with the OS and application software and patching to the last configuration used in production, cloud disaster recovery encapsulates the entire server, which includes the operating system, applications, patches, and data into a single software bundle or virtual server.

The virtual server is then copied or backed up to an offsite data center or spun up on a virtual host in minutes. Since the virtual server is not dependent on hardware, the operating system, applications, patches, and data can be migrated from one data center to another much faster than traditional DR approaches.

Cloud Security

Step 1: Determine where sensitive data lives, and prioritize integrations that increase visibility. ...

Step 2: Configure **cloud** platforms to maximize the **security** of their architecture. ...

Step 3: Monitor the **cloud** through integration.

Data Centre

- A data center is a centralized location that houses the computing and network equipment required to collect, store, process, and distribute large amounts of data.
- Data center facilities are also critical in providing access to the large amounts of data stored there for employees running daily operations, applications, and other processes in a cloud computing environment.
- Modern data centers consist of a variety of infrastructure components, including servers and network connectivity equipment, that permit access to the server storage over the Internet.

Resilience

Resiliency is the ability of a server, network, storage system, or an entire data center, to recover quickly and continue operating even when there has been an equipment failure, power outage or other disruption. ... When one server in the cluster fails, another node takes over. with its_redundant workloads

Agility

- In the cloud computing context, agility often refers to the ability to rapidly develop, test and launch software applications that drive business growth.
- Cloud Agility allows them to focus on other issues such as security, monitoring and analysis, instead of provisioning and maintaining the resources.

Cloud storage

- Cloud storage allows you to save data and files in an off-site location that you access either through the public internet or a dedicated private network connection. ... Computer hard drives can only store a finite amount of data.
- When users run out of storage, they need to transfer files to an external storage device.

MapReduce

- MapReduce is a programming model introduced by Google for processing and generating large data sets on clusters of computers.
- Google first formulated the framework for the purpose of serving Google's Web page indexing, and the new framework replaced earlier indexing algorithms. Beginner developers find the MapReduce framework beneficial because library routines can be used to create parallel programs without any worries about infra-cluster communication, task monitoring or failure handling processes.

- MapReduce runs on a large cluster of commodity machines and is highly scalable. It has several forms of implementation provided by multiple programming languages, like Java, C# and C++.
- Advertisement

Cloud Goverance

- Cloud Governance is a set of rules. It applies specific policies or principles to the use of cloud computing services.This model aims to secure applications and data even if located distantly. The best Cloud Governance solutions include People, Processes, and Technology. It basically refers to the decision making processes, criteria, and policies involved in the planning, architecture, acquisition, deployment, operation, architecture, acquisition, implementation, operation, and management of a Cloud computing capability.
- Cloud Governance best practices help to optimize the organization's:
- Operations: Doing it efficiently
- Risk and compliance: Doing it securely
- Financial: Doing more with less

Load Balancing

- Cloud load balancing is defined as the method of splitting workloads and computing properties in a cloud computing.

- It enables enterprise to manage workload demands or application demands by distributing resources among numerous computers, networks or servers.
- Cloud load balancing includes holding the circulation of workload traffic and demands that exist over the Internet.
- There are two elementary solutions to overcome the problem of overloading on the servers-
- First is a single-server solution in which the server is upgraded to a higher performance server. However, the new server may also be overloaded soon, demanding another upgrade. Moreover, the upgrading process is arduous and expensive.
- Second is a multiple-server solution in which a scalable service system on a cluster of servers is built. That’s why it is more cost effective as well as more scalable to build a server cluster system for network services.
- Load balancing is beneficial with almost any type of service, like HTTP, SMTP, DNS, FTP, and POP/IMAP. It also rises reliability through redundancy.

High Availability

- High Availability in the cloud is achieved by creating clusters.
- A high availability cluster is a group of servers that act as a single server to provide continuous uptime.

- These servers will have access to the same shared storage for data, so if a server is unavailable, the other servers pick up the load.

Four Steps to Achieving High Availability in the Cloud

- Build for server failure. Instances in the cloud — just as in a typical data center — are ephemeral. ...
- Build for zone failure. ...
- Build for cloud failure. ...
- Automate and test everything.

Security in cloud computing is a major concern. Data in cloud should be stored in encrypted form. To restrict client from accessing the shared data directly, proxy and brokerage services should be employed.

Security Planning

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

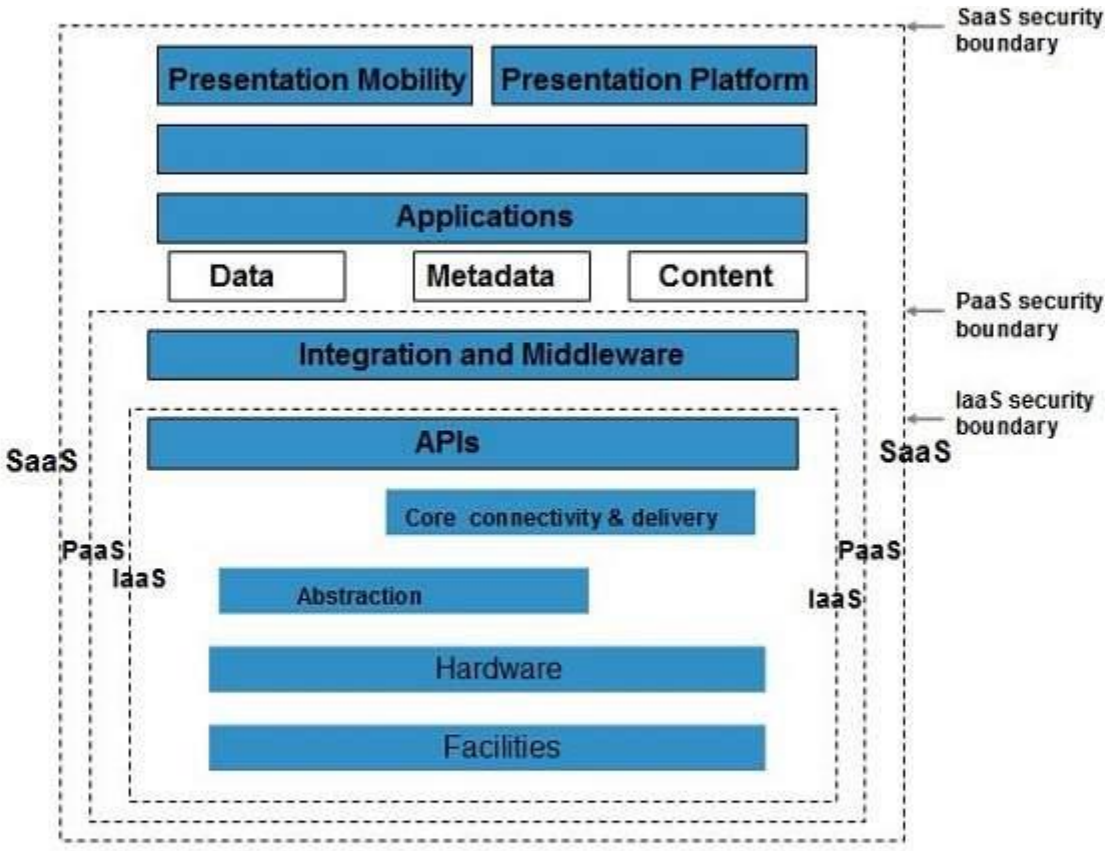
- Select resource that needs to move to the cloud and analyze its sensitivity to risk.
- Consider cloud service models such as **IaaS, PaaS,** and **SaaS.** These models require customer to be responsible for security at different levels of service.
- Consider the cloud type to be used such as **public,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

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

Understanding Security of Cloud

Security Boundaries

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



Key Points to CSA Model

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

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

Understanding Data Security

Since all the data is transferred using Internet, data security is of major concern in the cloud. Here are key mechanisms for protecting data.

- Access Control
- Auditing
- Authentication
- Authorization

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

Isolated Access to Data

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

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

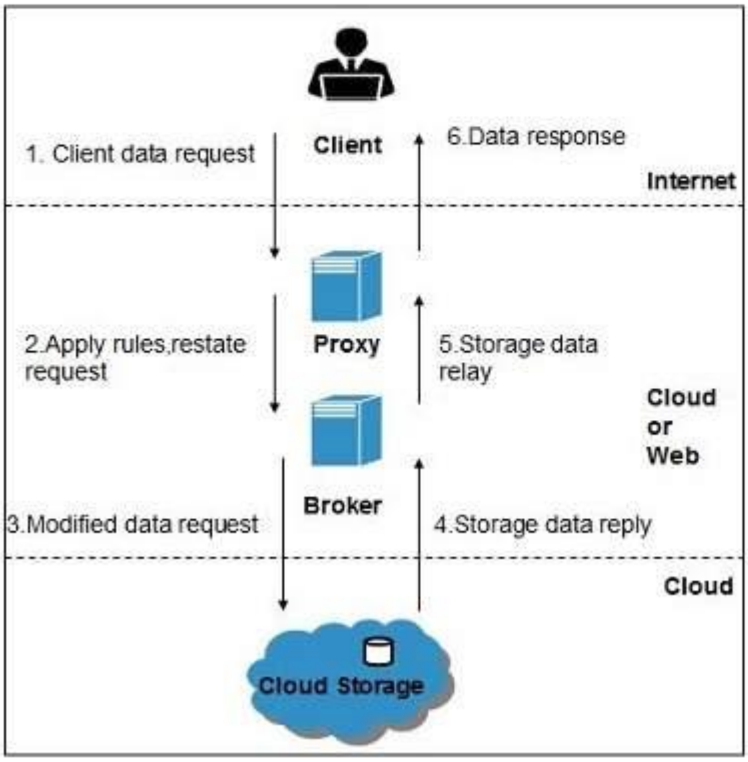
- A broker with full access to storage but no access to client.
- A proxy with no access to storage but access to both client and broker.

Working Of Brokered Cloud Storage Access System

When the client issues request to access data:

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

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



Encryption

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

Cloud Computing Security Architecture

- Cloud security architecture describes all the hardware and technologies designed to protect data, workloads, and systems within cloud platforms.
- Developing a strategy for cloud security architecture should begin during the blueprint and design process and should be integrated into cloud platforms from the ground up.
- Cloud architects will focus entirely on performance first and then attempt to bolt security on after the fact.

Cloud Security Core Capabilities

Secure cloud computing architecture encompasses three core capabilities: confidentiality, integrity, and availability.

Confidentiality

- It is the ability to keep information secret and unreadable to the people who shouldn't have access to that data, such as attackers or people inside an organization without the proper access level.

- Confidentiality also includes privacy and trust, or when a business pledges secrecy in handling their customers' data.

Integrity

- It is the idea that the systems and applications are exactly what you expect them to be, and function exactly as you expect them to function.

- If a system or application has been compromised to produce an unknown, unexpected, or misleading output, this can lead to losses.

Availability

- It is the third capability and is generally the least considered by cloud architects.
- Availability speaks to denial-of-service (DoS) attacks.
- Perhaps an attacker can't see or change your data.
- But if an attacker can make systems unavailable to you or your customers, then you can't carry out tasks that are essential to maintain your business.

Information classification

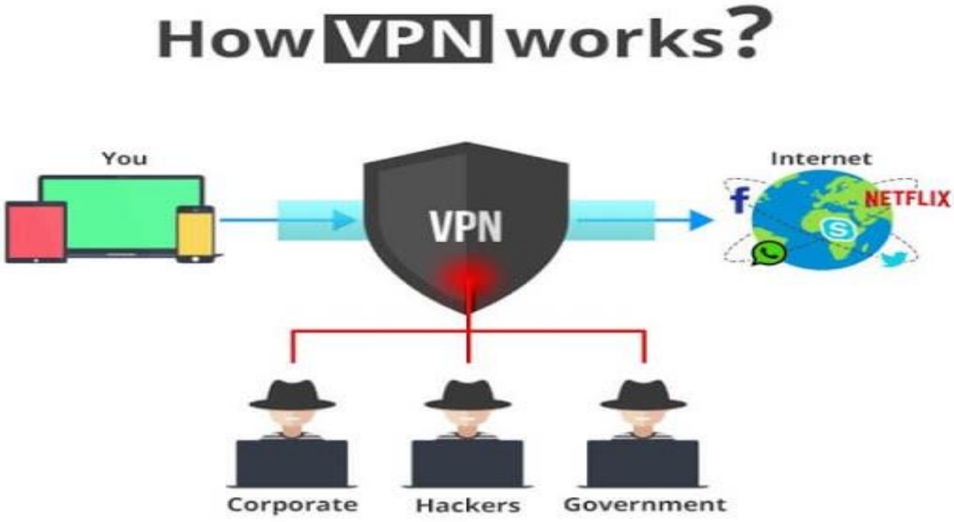
Information classification is a process in which organisations assess the data that they hold and the level of protection it should be given.

- Organisations usually classify information in terms of confidentiality – i.e. who is granted access to see it.

Virtual Private Networks

- VPN stands for virtual private network.
- A virtual private network (VPN) is a technology that creates a safe and encrypted connection over a less secure network, such as the internet.
- Virtual Private network is a way to extend a private network using a public network such as internet.

- The name only suggests that it is Virtual "private network" i.e. user can be the part of local network sitting at a remote location.
- It makes use of tunneling protocols to establish a secure connection.



Key Management in Cryptography

Key Management:
In cryptography it is a very tedious task to distribute the public and private key between sender and receiver.

If key is known to the third party (forger/eavesdropper) then the whole security mechanism becomes worthless.

So, there comes the need to secure the exchange of keys.

There are 2 aspects for Key Management:

- Distribution of public keys.
- Use of public-key encryption to distribute secret.

Distribution of Public Key:

Public key can be distributed in 4 ways:

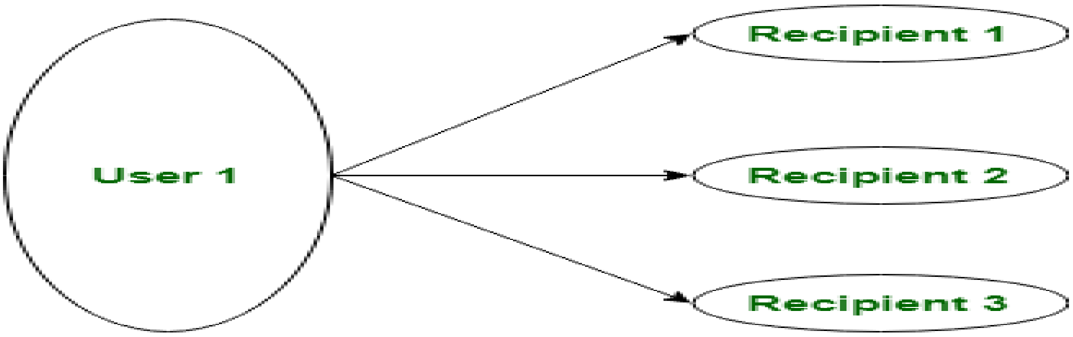
- Public announcement
- Publicly available directory
- Public-key authority

Public-key certificates

Public Announcement

Here the public key is broadcasted to everyone.

- ❖ Major weakness of this method is forgery.
- ❖ Anyone can create a key claiming to be someone else and broadcast it.



Public Key Announcement

Publicly Available Directory:

In this type, the public key is stored at a public directory.

- Directories are trusted here, with properties like Participant Registration, access and allow to modify values at any time, contains entries like {name, public-key}.
- Directories can be accessed electronically still vulnerable to forgery or tampering.

Public Key Authority:

- It is similar to the directory but, improve security by tightening control over distribution of keys from directory.
- It requires users to know public key for the directory.
- Whenever the keys are needed, a real-time access to directory is made by the user to obtain any desired public key securely.

Public Certification:

- This time authority provides a certificate (which binds identity to the public key) to allow key exchange without real-time access to the public authority each time.
- The certificate is accompanied with some other info such as period of validity, rights of use etc.
- All of this content is signed by the trusted Public-Key or Certificate Authority (CA) and it can be verified by anyone possessing the authority’s public-key.

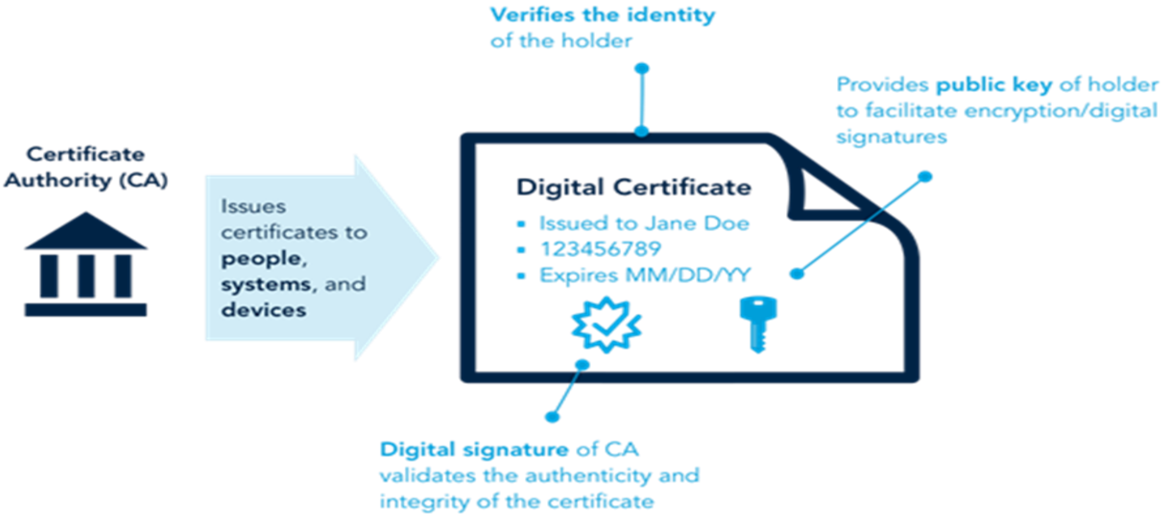
Digital certificates

- Digital certificate is issued by a trusted third party which proves sender's identity to the receiver and receiver’s identity to the sender.
- A digital certificate is a certificate issued by a Certificate Authority (CA) to verify the identity of the certificate holder.

- The CA issues an encrypted digital certificate containing the applicant’s public key and a variety of other identification information.
- Digital certificate is used to attach public key with a particular individual or an entity.

Digital certificate contains:-

- Name of certificate holder.
- Serial number which is used to uniquely identify a certificate, the individual or the entity identified by the certificate
- Expiration dates.
- Copy of certificate holder's public key.(used for decrypting messages and digital signatures)
- Digital Signature of the certificate issuing authority.
- Digital ceritifcate is also sent with the digital signature and the message.


Verifies the identity of the holder
Provides public key of holder to facilitate encryption/digital signatures
Certificate Authority (CA)
Issues certificates to people, systems, and devices
Digital Certificate
Issued to Jane Doe
123456789
Expires MM/DD/YY
Digital signature of CA validates the authenticity and integrity of the certificate

Digital Signature

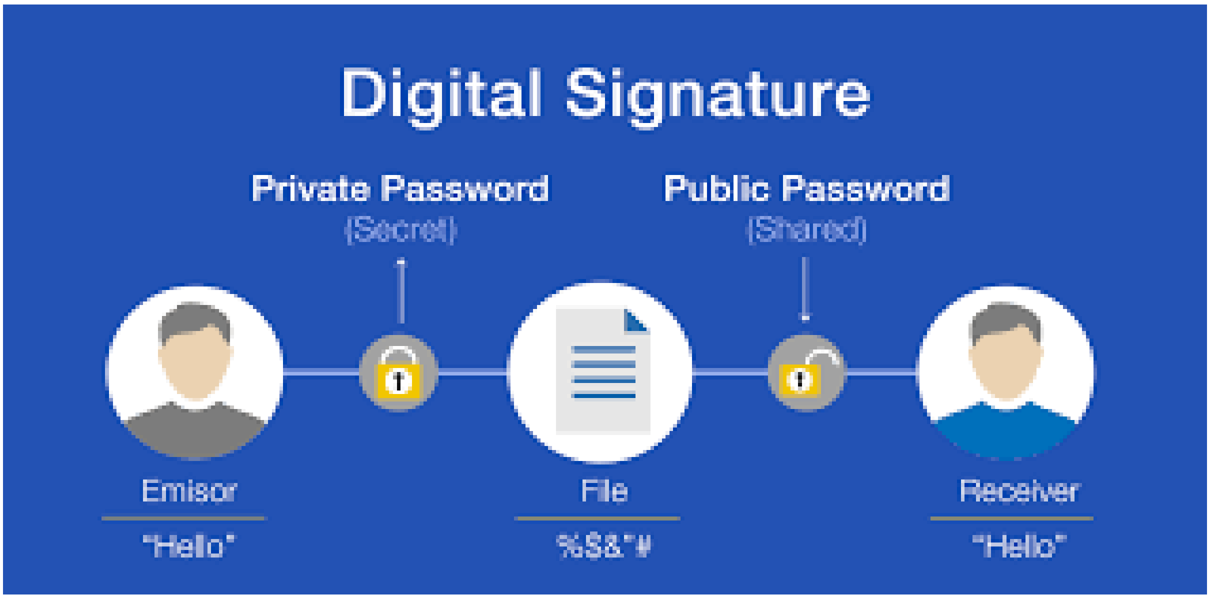
- A digital signature is a mathematical technique used to validate the authenticity and integrity of a message, software or digital document.

Key Generation Algorithms :

- Digital signature are electronic signatures, which assures that the message was sent by a particular sender.
- While performing digital transactions authenticity and integrity should be assured, otherwise the data can be altered or someone can also act as if he was the sender and expect a reply.

Signing Algorithms:

- To create a digital signature, signing algorithms like email programs create a one-way hash of the electronic data which is to be signed.
- The signing algorithm then encrypts the hash value using the private key (signature key).
- This encrypted hash along with other information like the hashing algorithm is the digital signature.
- **Signature Verification Algorithms :**
- Verifier receives Digital Signature along with the data.
- It then uses Verification algorithm to process on the digital signature and the public key (verification key) and generates some value.



- **The steps followed in creating digital signature are :**
- Message digest is computed by applying hash function on the message and then message digest is encrypted using private key of sender to form the digital signature. (digital signature = encryption (private key of sender, message digest) and message digest = message digest algorithm(message)).
- Digital signature is then transmitted with the message.(message + digital signature is transmitted)
- Receiver decrypts the digital signature using the public key of sender.(This assures authenticity,as only sender has his private key so only sender can encrypt using his private key which can thus be decrypted by sender’s public key).
- The receiver now has the message digest.
- The receiver can compute the message digest from the message (actual message is sent with the digital signature).

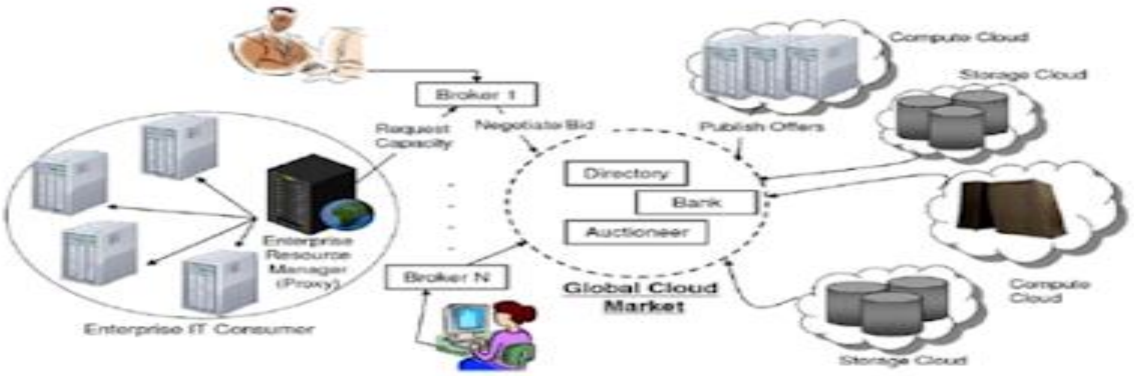
- The message digest computed by receiver and the message digest (got by decryption on digital signature) need to be same for ensuring integrity.

Feature	**Digital Signature**	**Digital Certificate**
Basics / Definition	**Digital signature is like a fingerprint or an attachment to a digital document that ensures its authenticity and integrity.**	**Digital certificate is a file that ensures holder’s identity and provides security.**
Process / Steps	**Hashed value of original message is encrypted with sender’s secret key to generate the digital signature.**	**It is generated by CA (Certifying Authority) that involves four steps: Key Generation, Registration, Verification, Creation.**
Security Services	**Authenticity of Sender, integrity of the document and non-repudiation.**	**It provides security and authenticity of certificate holder.**

Standard	**It follows Digital Signature Standard (DSS).**	**It follows X.509 Standard Format**

Market Based Management of Clouds

The Real Potential of cloud computing resides is the fact that it actually facilitates the establishment of a market for trading IT utilities.



There are three major components of cloud exchange are:-

Directory:-the market directory contains a listing of all the published services

that are available in the cloud marketplace.

Auctioneer:-the auctioneer is in charge of keeping track of the running auctions in

the market place and of verifying that the Aauctions for services are properly

conducted and thatmalicious market players are prevented from performing

illegal activities.

- Bank:-the bank is the component that takes care of the financial aspect of all the

operations happening in the virtual market place.

Cloud Information security vendors

10 vendors that offer top cloud security tools.

- CloudPassage
- FireEye
- LaceWork
- McAfee
- Palo Alto Networks
- Qualys
- Symantec
- Tenable
- Trend Micro
- VMware

Third Party Cloud service

- **Third-party cloud service providers, such as Azure, AWS, and Google Cloud Platform (GCP), calculate these costs for you and provide the technology.**
- **While it might be challenging to optimize costs or hidden costs may exist, the service providers do most of the hard work and then charge a subscription fee**

HADOOP TECHNOLOGY

CONTENTS

✓What is hadoop Technology??

✓Why hadoop?

✓Developers of hadoop Technology

✓Famous hadoop users

✓Hadoop Features

✓Hadoop Architectures

✓Core-Components of Hadoop

✓Hadoop High Level Architechture

✓Hadoop cluster

CONTENTS…

✓What is HDFS

✓HDFS – Name Node features:

✓HDFS-name node architecture

✓HDFS-data node

✓Hadoop MAPREDUCE

✓Benefits of Hadoop…

✓Conclusion

✓Reference

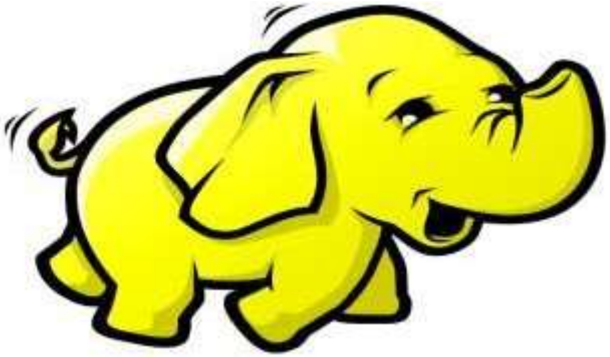
HADOOP TECHNOLOGY

What is Hadoop Technology??

- The most well known technology used for Big Data is Hadoop.
- It is actually a large scale batch data processing system

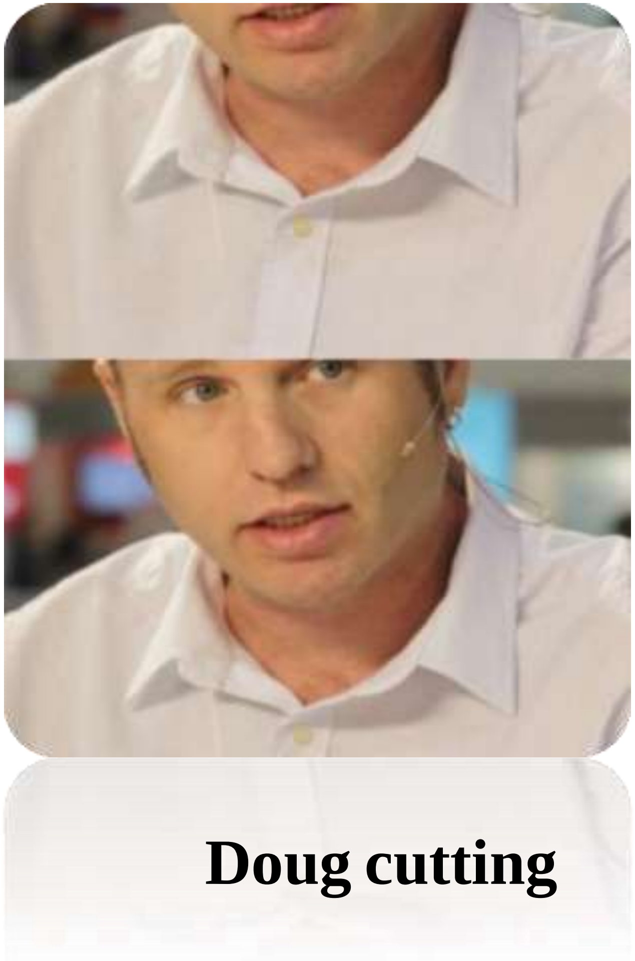
Why Hadoop ??

- Distributed cluster system
- Platform for massively scalable applications
- Enables parallel data processing



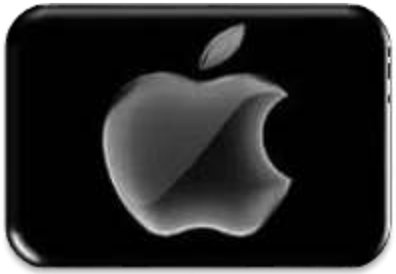
Developers of Hadoop Technology:

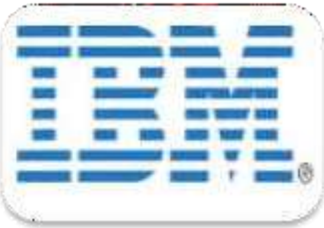
Michael j. cafarella

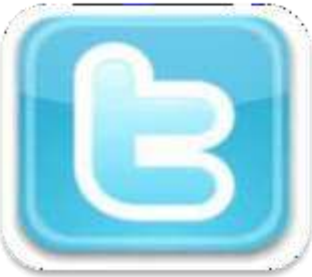


Doug cutting

Famous Hadoop users



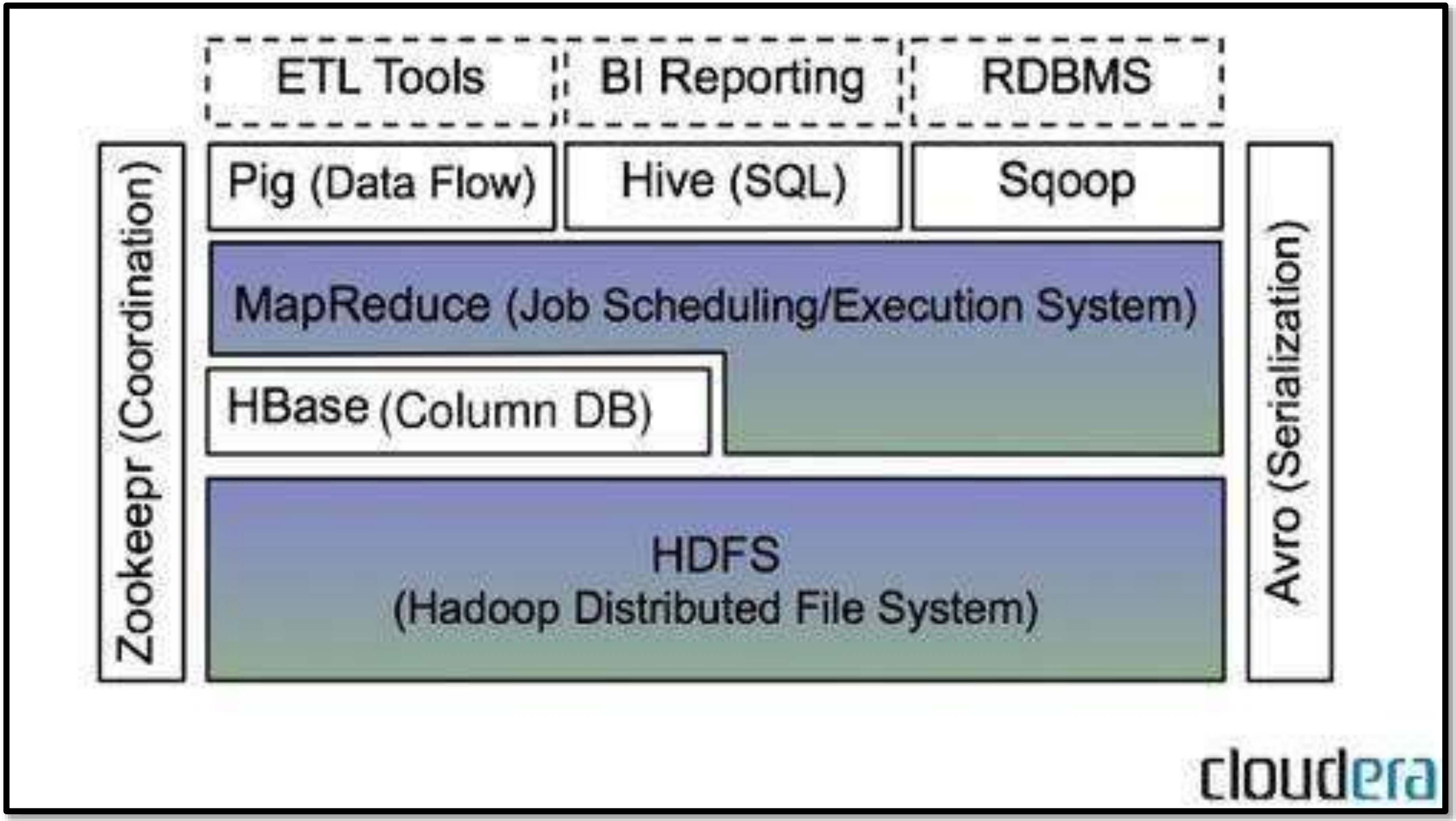




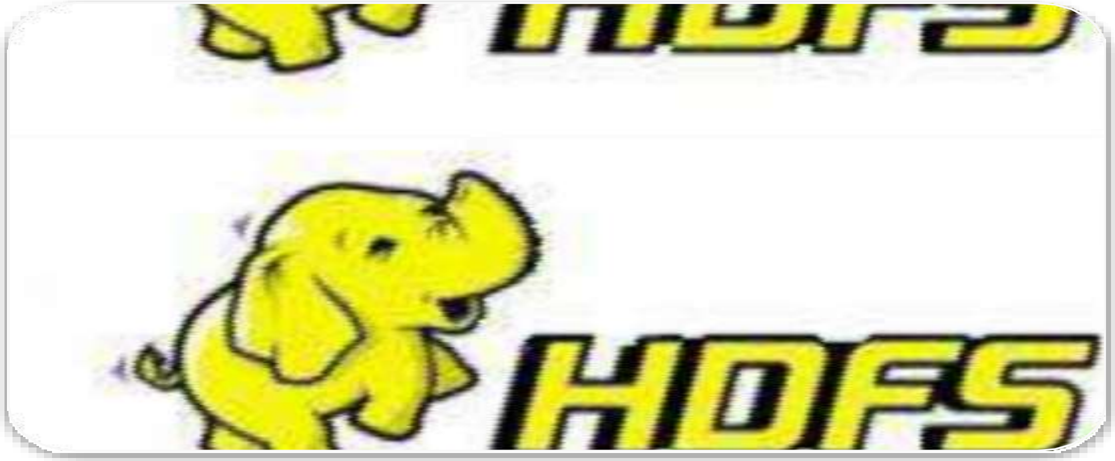
Hadoop Features

- Hadoop provides access to the file systems
- The Hadoop Common package contains the necessary JAR files and scripts
- The package also provides source code, documentation and a contribution section that includes projects from the Hadoop Community.

HADOOP ARCHITECTURE



Core-Components of Hadoop:



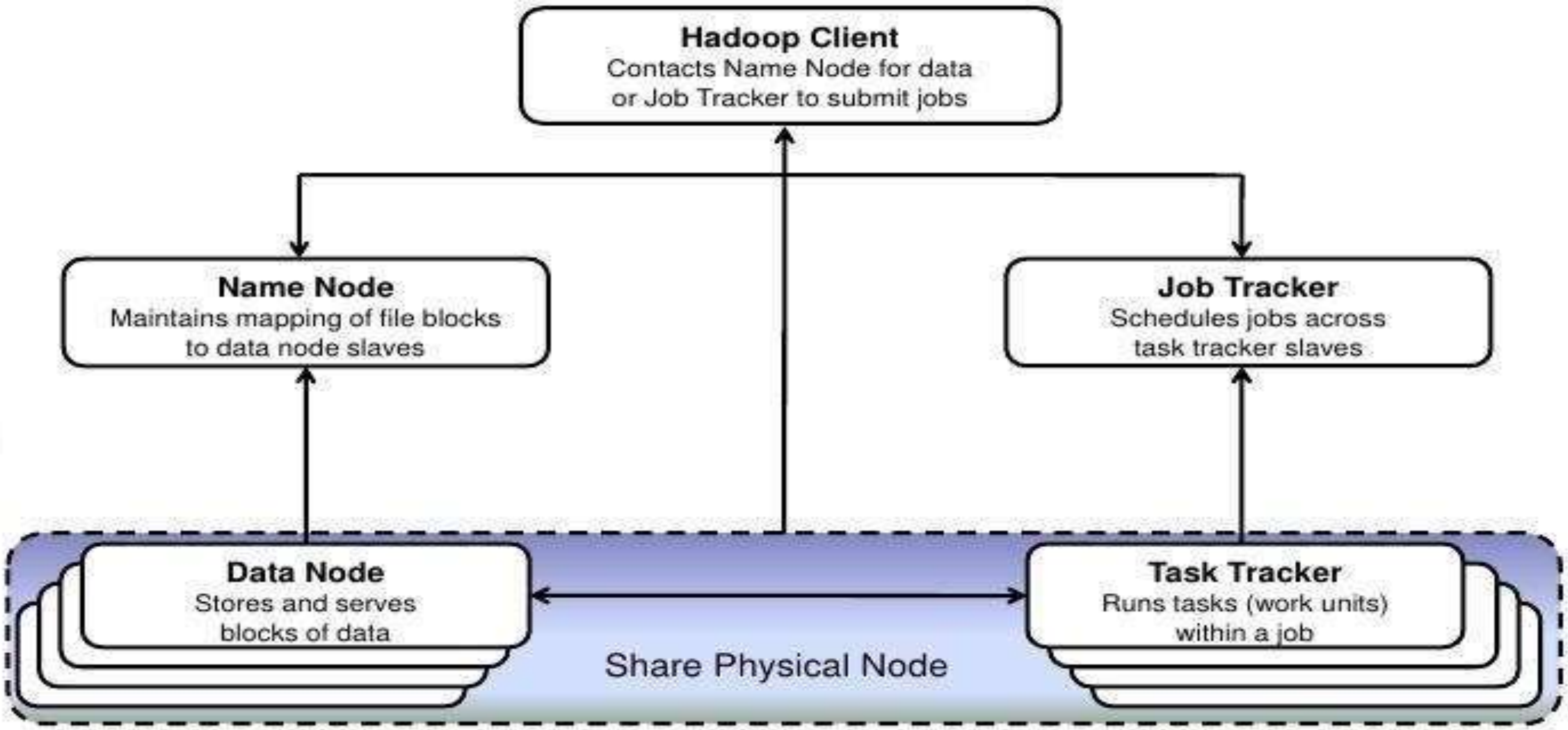
Hadoop distributive file system.

Map reduce.

What is HDFS ?

- Distributed file system
- Traditional hierarchical file organization
- Single namespace for the entire cluster
- Write-once-read-many access model
- Aware of the network topology

Hadoop High Level Architechture



Hadoop Client
Contacts Name Node for data
or Job Tracker to submit jobs

Name Node
Maintains mapping of file blocks
to data node slaves

Job Tracker
Schedules jobs across
task tracker slaves

Data Node
Stores and serves
blocks of data

Task Tracker
Runs tasks (work units)
within a job

Share Physical Node

Hadoop cluster

- A Small Hadoop Cluster Include a single master & multiple worker nodes

Master node:

Data Node

Job Tracker

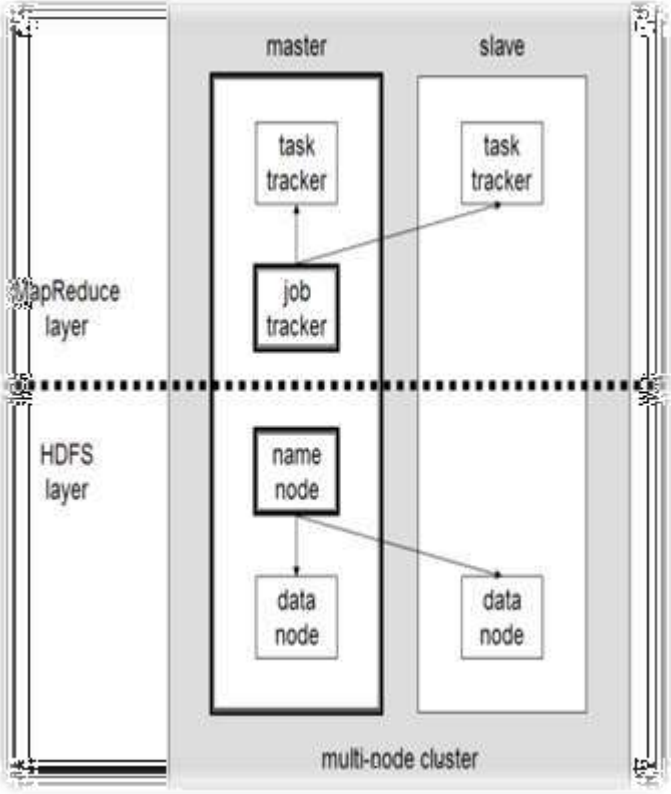
Task Tracker

Name Node

Slave node:

Data Node

Task Tracke

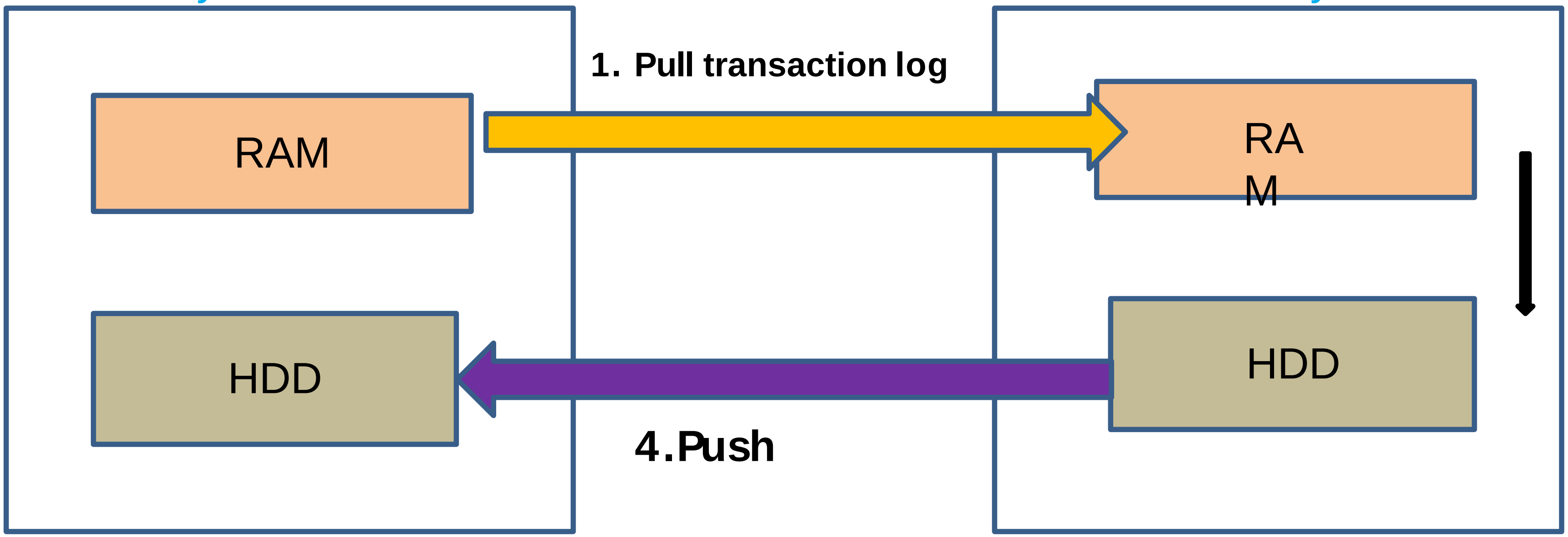


HDFS – Name Node Features

Metadata in main memory:

- List of files
- List of blocks for each file
- List of Data Nodes for each block
- File attributes
- Creation time
- Records every change in the metadata

HDFS-name node architecture



2. Merge changes

Primary name-node

Secondary name node

1. Pull transaction log

RAM

RA
M

HDD

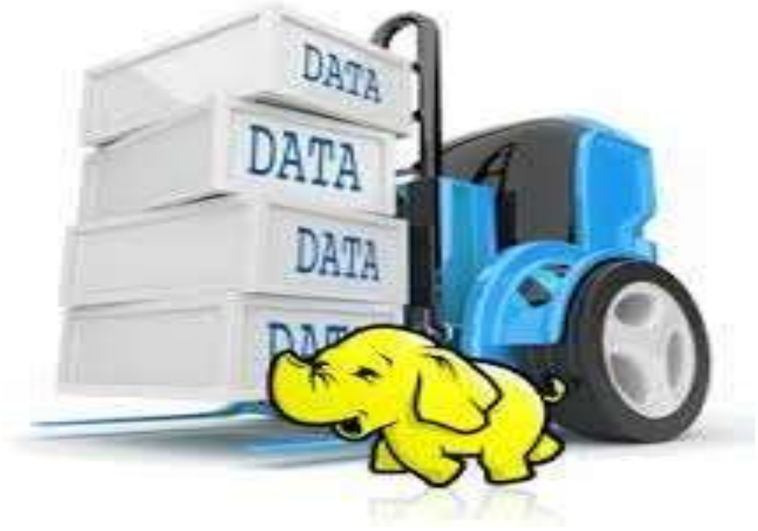
HDD

4.Push

3.Store to HDD

HDFS-Data node

- Block Server Stores data in the local file system
- Periodic validation of checksums
- Periodically sends a report of all existing blocks to the Name Node



Hadoop MAPREDUCE

Map reduce implementation:

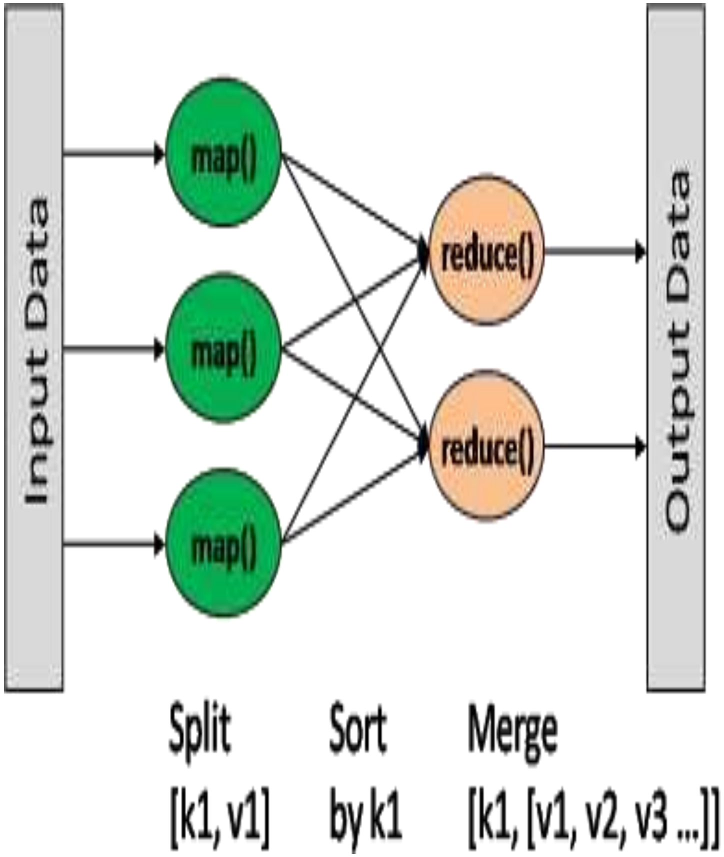
Job Tracker:

Splitting into map and reduce tasks

Scheduling tasks on a cluster node

Task Tracker:

Runs Map Reduce tasks periodically



Benefits of Hadoop…

- Cost Saving and efficient and reliable data processing
- Provides an economically scalable solution
- Storing and processing of large amount of data
- Data grid operating system
- It is deployed on industry standard servers rather than expensive specialized data storage systems.
- Parallel processing of huge amounts of data across inexpensive, industry-standard servers.

CONCLUSION

Why commodity hw ?

✓because cheaper

✓designed to tolerate faults

Why HDFS ?

✓network bandwidth vs seek latency

Why Map reduce programming model?

✓parallel programming

✓large data sets

✓moving computation to data

✓single compute + data cluster

REFERENCES

- **Apache Hadoop!**

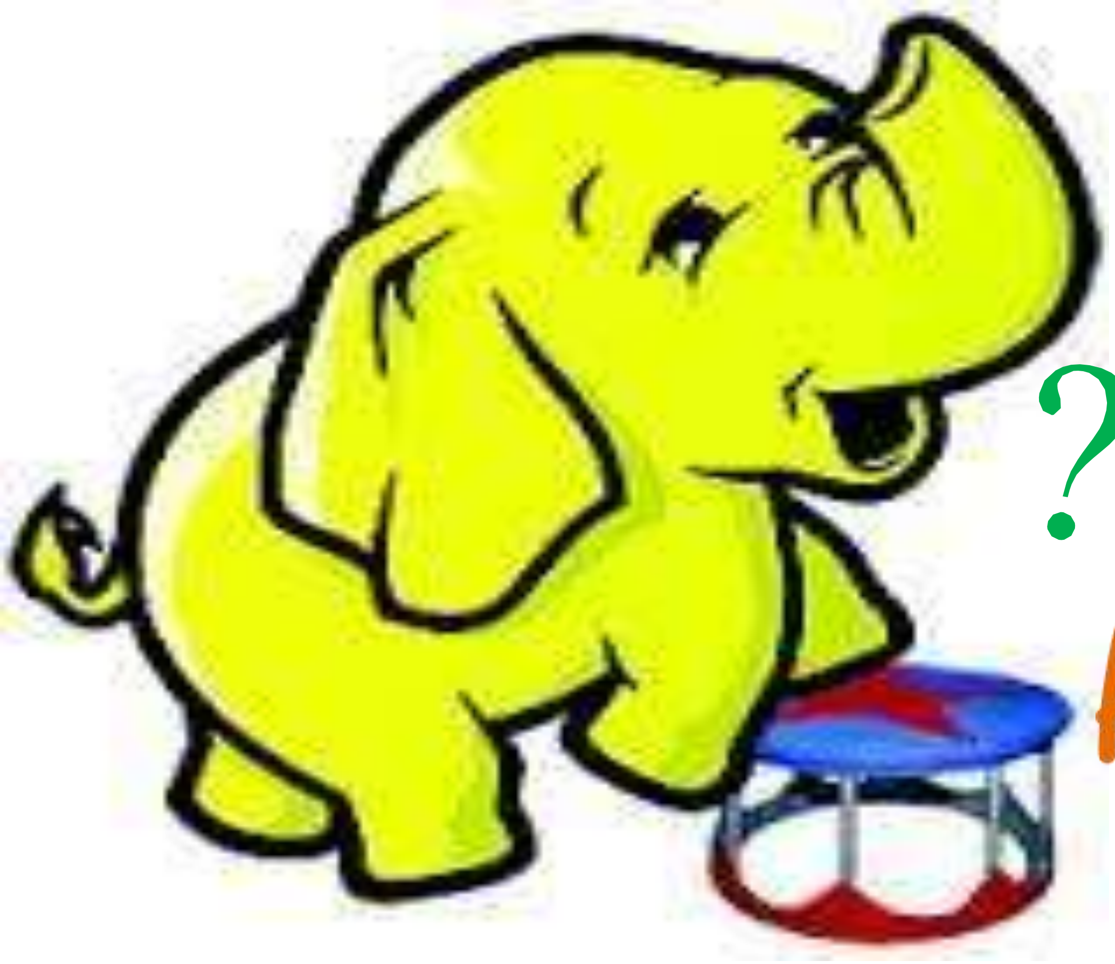
(http://hadoop.apache.org)

- **Hadoop on Wikipedia**

(http://en.wikipedia.org/wiki/Hadoop)

- **Cloudera - Apache Hadoop for the Enterprise**

(http://www.cloudera.com



Any Queries

Thank you