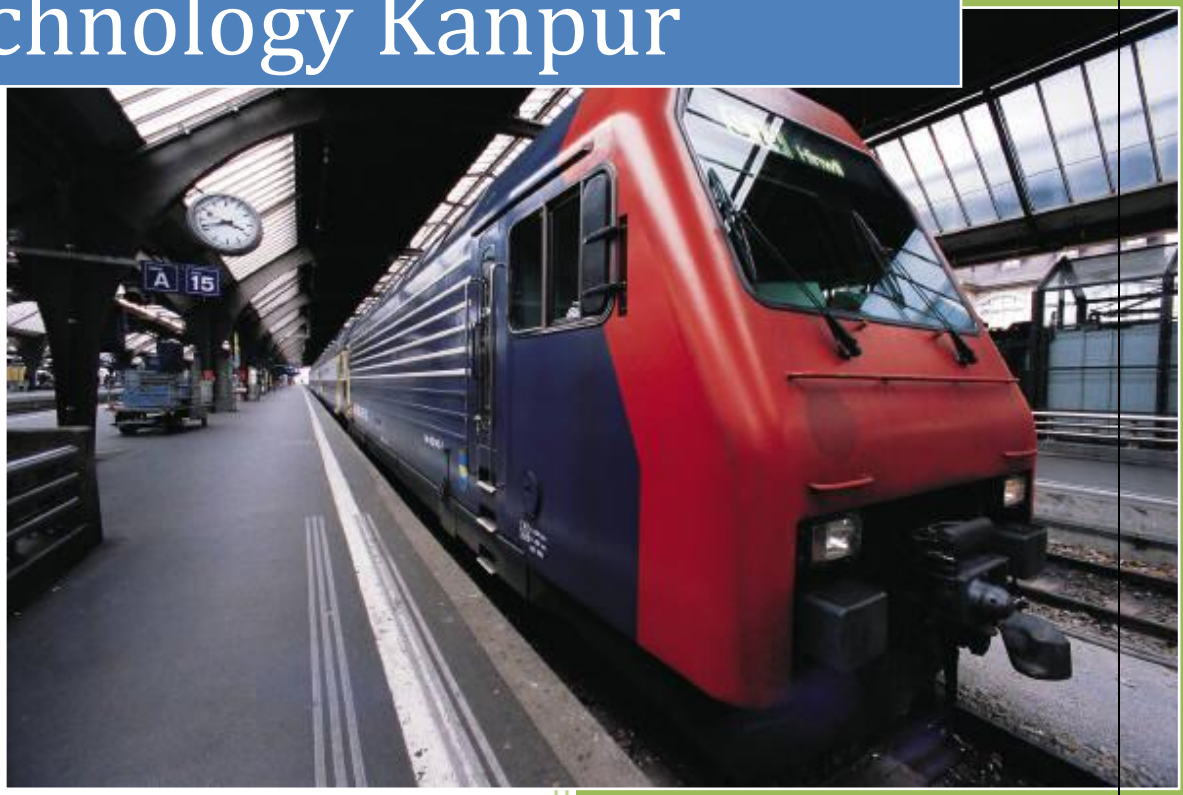
Cloud Computing, KOE081

2023

Vision Institute Of Technology Kanpur



By Kaptan Yadav

HOD, Electrical & Electronics Engineering

2/6/2023

KOE081: CLOUD COMPUTING

Unit-1

Introduction: Cloud Computing – Definition of Cloud – Evolution of Cloud Computing – Underlying Principles of Parallel and Distributed, History of Cloud Computing - Cloud Architecture **- Types of Clouds** - Business models around Clouds – Major Players in Cloud Computing, issues in Clouds - Eucalyptus - Nimbus - Open Nebula, CloudSim.

Unit-2

Types of Cloud services: Software as a Service, Platform as a Service –Infrastructure as a Service - Database as a Service - Monitoring as a Service –Communication as services. Service providers- Google, Amazon, Microsoft Azure, IBM, Sales force.

Unit-3

Collaborating Using Cloud Services:

Email Communication over the Cloud - CRM Management – Project Management-Event Management - Task Management – Calendar - Schedules - Word Processing – Presentation – Spreadsheet - Databases – Desktop - Social Networks and Groupware.

Unit-4

Virtualization for Cloud: Need for Virtualization – Pros and cons of Virtualization – Types of Virtualization –System VM, Process VM, Virtual Machine monitor – Virtual machine properties - Interpretation and binary translation, HLL VM - supervisors – Xen, KVM, VMware,Virtual Box, Hyper-V.

Unit-5

Security, Standards and Applications:

Security in Clouds: Cloud security challenges – Software as a Service Security, Common Standards: The Open Cloud Consortium – The Distributed management Task Force – Standards for application Developers – Standards for Messaging – Standards for Security, End user access to cloud computing, Mobile Internet devices and the cloud. Hadoop – MapReduce – Virtual Box — Google App Engine – Programming Environment for Google App Engine

Unit- 5th

Common Standards in Cloud Computing : Common Standards in Cloud Computing refer to the widely accepted guidelines and protocols that govern the use of cloud computing technologies. These standards help ensure that cloud services are compatible with each other, secure, and reliable.

1. Working Groups
2. Open Cloud Consortium
3. The Distributed Management Task Force
4. Standards for Application Developers
5. Standards for Messaging
6. Standards for Security

Working Groups : A working group is an assembled, cooperative collaboration of researchers working on new research activities that would be difficult for any one member to develop alone.

A working group can exist for anywhere between a few months to many years. Working groups generally strive to create an informational document a standard, or find some resolution for problems related to a system or network. Working groups are sometimes also referred to as task groups or technical advisory groups.

Open Cloud Consortium(OCC) : The Open Cloud Consortium (OCC) is A not for profit. It Manages and operates cloud computing infrastructure to support scientific, medical, health care and environmental research. OCC members span the globe and include over 10 universities, over 15 companies, and over 5 government agencies and national laboratories. The OCC is organized into several different working groups.

The OCC Mission

- The purpose of the Open Cloud Consortium is to support the development of standards for cloud computing and to develop a framework for interoperability among various clouds.
- The OCC supports the development of benchmarks for cloud computing.
- Manages cloud computing testbeds, such as the Open Cloud Testbed, to improve cloud computing software and services.
- Develops reference implementations, benchmarks and standards, such as the MalStone Benchmark, to improve the state of the art of cloud computing.
- Sponsors workshops and other events related to cloud computing to educate the community

The Distributed Management Task Force (DMTF)

- DMTF enables more effective management of millions of IT systems worldwide by bringing the IT industry together to collaborate on the development, validation and promotion of systems management standards.
- The group spans the industry with 160 member companies and organizations, and more than 4,000 active participants crossing 43 countries.
- The DMTF board of directors is led by 16 innovative, industry-leading technology companies.
- DMTF management standards are critical to enabling management interoperability among multi vendor systems, tools and solutions within the enterprise.
- The DMTF started the Virtualization Management Initiative (VMAN).
- The Open Virtualization Format (OVF) is a fairly new standard that has emerged within the VMAN Initiative.
- Benefits of VMAN are
  * Lowering the IT learning curve, and
  * Lowering complexity for vendors implementing their solutions

Standards for Application Developers: The purpose of application development standards is to ensure uniform, consistent, high-quality software solutions. Programming standards help to improve the readability of the software, allowing developers to understand new code more quickly and thoroughly.
Commonly used application standards are available for the Internet in browsers, for transferring data, sending messages, and securing data.
Standards for Messaging : A message is a unit of information that is moved from one place to another.
Most common messaging standards used in the cloud are
1. Simple Message Transfer Protocol (SMTP)
2. Post Office Protocol (POP)
3. Internet Messaging Access Protocol (IMAP)
4. Syndication (Atom, Atom Publishing Protocol and RSS)
5. Communications (HTTP, SIMPLE, and XMPP)

1. Simple Message Transfer Protocol (SMTP) : Simple Message Transfer Protocol is arguably the most important protocol in use today for basic messaging. Before SMTP was created, email messages were sent using File Transfer Protocol (FTP).

2. The FTP protocol was designed to transmit files, not messages, so it did not provide any means for recipients to identify the sender or for the sender to designate an intended recipient.
3. SMTP was designed so that sender and recipient information could be transmitted with the message.
4. SMTP is a two-way protocol that usually operates using TCP (Transmission Control Protocol) port 25

2. Post Office Protocol (POP) :

1. SMTP can be used both to send and receive messages, but the client must have a constant connection to the host to receive SMTP messages.
2. The Post Office Protocol (POP) was introduced to circumvent this situation.
3. POP is a lightweight protocol whose single purpose is to download messages from a server. This allows a server to store messages until a client connects and requests them.
4. Once the client connects, POP servers begin to download the messages and subsequently delete them from the server (a default setting) in order to make room for more messages.

3. Internet Messaging Access Protocol (IMAP) :

1. Once mail messages are downloaded with POP, they are automatically deleted from the server when the download process has finished.
2. Many businesses have compulsory compliance guidelines that require saving messages. It also becomes a problem if users move from computer to computer or use mobile networking, since their messages do not automatically move where they go.
3. To get around these problems, a standard called Internet Messaging Access Protocol was created. IMAP allows messages to be kept on the server but viewed and manipulated (usually via a browser) as though they were stored locally.

Standards for Security : Security standards define the processes,procedures, and practices necessary for implementing a secure environment that provides privacy and security of confidential information in a cloud environment.

Security protocols, used in the cloud are

1.Security Assertion Markup Language (SAML)
2.Open Authentication (Oauth)
3.OpenID

4.SSL/TLS

1.Security Assertion Markup Language (SAML) : SAML is an XML-based standard for communicating authentication, authorization, and attribute information among online partners. It allows businesses to securely send assertions between partner organizations regarding the identity and entitlements of a principal.

● SAML allows a user to log on once for affiliated but separate Web sites. SAML is designed for business-to-business (B2B) and business-to-consumer (B2C) transactions.

● SAML is built on a number of existing standards, namely, SOAP, HTTP, and XML. SAML relies on HTTP as its communications protocol and specifies the use of SOAP.

● Most SAML transactions are expressed in a standardized form of XML. SAML assertions and protocols are specified using XML schema.

2.Open Authentication (Oauth)

OAuth is an open protocol, initiated by Blaine Cook and Chris Messina, to allow secure API authorization in a simple, standardized method for various types of web applications.

● OAuth is a method for publishing and interacting with protected data.

● OAuth provides users access to their data while protecting account credentials.

● OAuth by itself provides no privacy at all and depends on other protocols such as SSL to accomplish that.

3.OpenID

OpenID is an open, decentralized standard for user authentication and access control that allows users to log onto many services using the same digital identity.

● It is a single-sign-on (SSO) method of access control.

● It replaces the common log-in process (i.e., a log-in name and a password) by allowing users to log in once and gain access to resources across participating systems.

● An OpenID is in the form of a unique URL and is authenticated by the entity hosting the OpenID URL.

Hadoop : Hadoop is an open source framework from Apache and is used to store process and analyze data which are very huge in volume. Hadoop is written in. It is used for batch/offline processing.It is being used by Facebook, Yahoo, Google, Twitter, LinkedIn and many more. Moreover it can be scaled up just by adding nodes in the cluster.

Instead of using one large computer to store and process the data, Hadoop allows clustering multiple computers to analyze massive datasets in parallel more quickly.

Modules of Hadoop :

HDFS: Hadoop Distributed File System. Google published its paper GFS and on the basis of that HDFS was developed. It states that the files will be broken into blocks and stored in nodes over the distributed architecture.

Yarn: Yet another Resource Negotiator is used for job scheduling and manage the cluster.

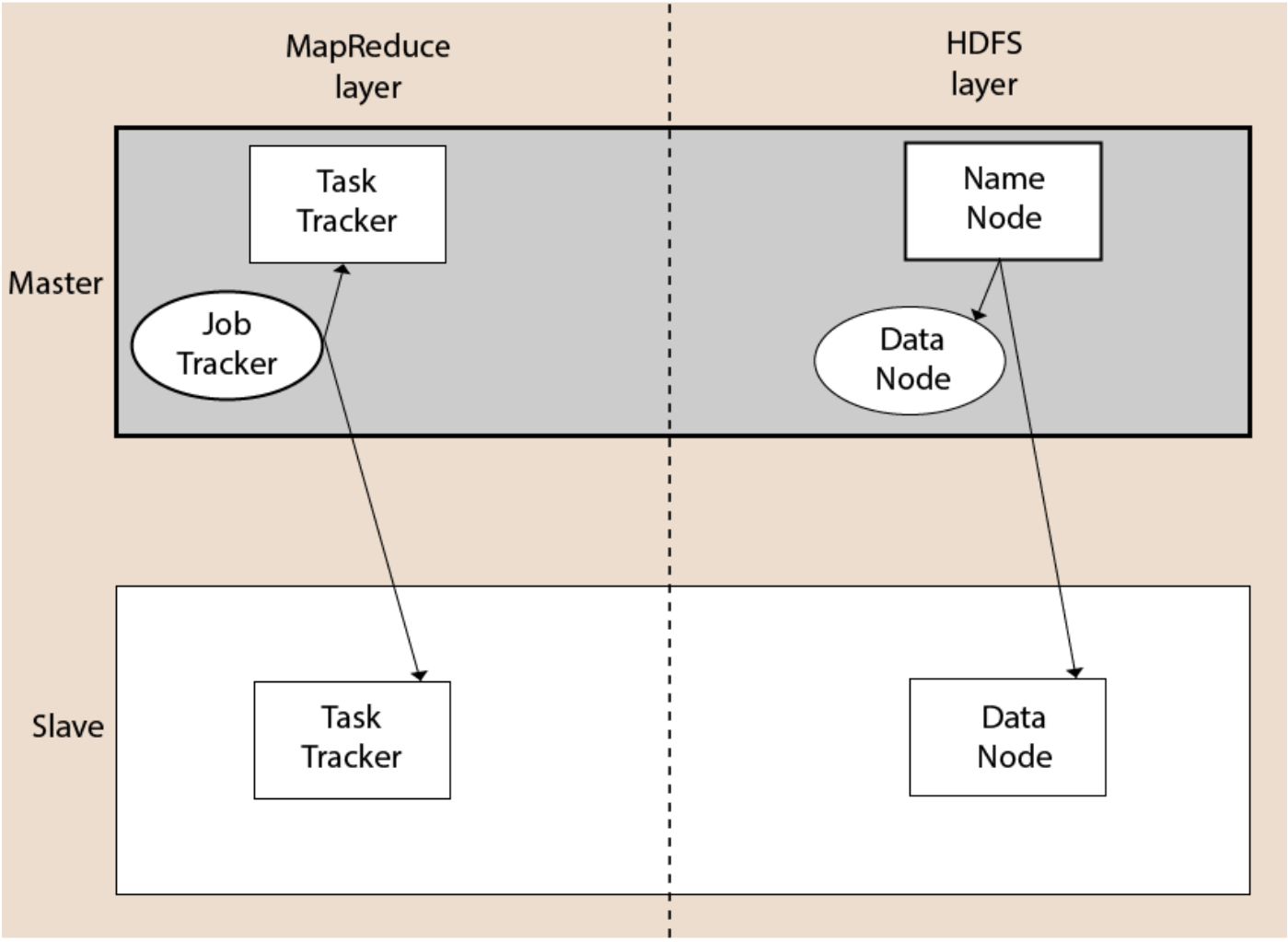
Map Reduce: This is a framework which helps Java programs to do the parallel computation on data using key value pair. The Map task takes input data and converts it into a data set which can be computed in Key value pair. The output of Map task is consumed by reduce task and then the out of reducer gives the desired result.

Hadoop Common: These Java libraries are used to start Hadoop and are used by other Hadoop modules.

Hadoop Architecture :

The Hadoop architecture is a package of the file system, MapReduce engine and the HDFS (Hadoop Distributed File System). The MapReduce engine can be MapReduce/MR1 or YARN/MR2.

A Hadoop cluster consists of a single master and multiple slave nodes. The master node includes Job Tracker, Task Tracker, NameNode, and DataNode whereas the slave node includes DataNode and TaskTracker.



Hadoop Distributed File System : The Hadoop Distributed File System (HDFS) is a distributed file system for Hadoop. It contains a master/slave architecture. This architecture consist of a single NameNode performs the role of master, and multiple DataNodes performs the role of a slave.

Both NameNode and DataNode are capable enough to run on commodity machines. The Java language is used to develop HDFS. So any machine that supports Java language can easily run the NameNode and DataNode software.

NameNode :

- It is a single master server exist in the HDFS cluster.
- As it is a single node, it may become the reason of single point failure.
- It manages the file system namespace by executing an operation like the opening, renaming and closing the files.
- It simplifies the architecture of the system.

DataNode :

- The HDFS cluster contains multiple DataNodes.
- Each DataNode contains multiple data blocks.

- These data blocks are used to store data.
- It is the responsibility of DataNode to read and write requests from the file system's clients.
- It performs block creation, deletion, and replication upon instruction from the NameNode.

Job Tracker :

- The role of Job Tracker is to accept the MapReduce jobs from client and process the data by using NameNode.
- In response, NameNode provides metadata to Job Tracker**.**

Task Tracker :

- It works as a slave node for Job Tracker.
- It receives task and code from Job Tracker and applies that code on the file. This process can also be called as a Mapper.

Advantages of Hadoop

- **Fast:** In HDFS the data distributed over the cluster and are mapped which helps in faster retrieval. Even the tools to process the data are often on the same servers, thus reducing the processing time. It is able to process terabytes of data in minutes and Peta bytes in hours.
- **Scalable:** Hadoop cluster can be extended by just adding nodes in the cluster.
- Cost Effective: Hadoop is open source and uses commodity hardware to store data so it really cost effective as compared to traditional relational database management system.
- **Resilient to failure**: HDFS has the property with which it can replicate data over the network, so if one node is down or some other network failure happens, then Hadoop takes the other copy of data and use it. Normally, data are replicated thrice but the replication factor is configurable.

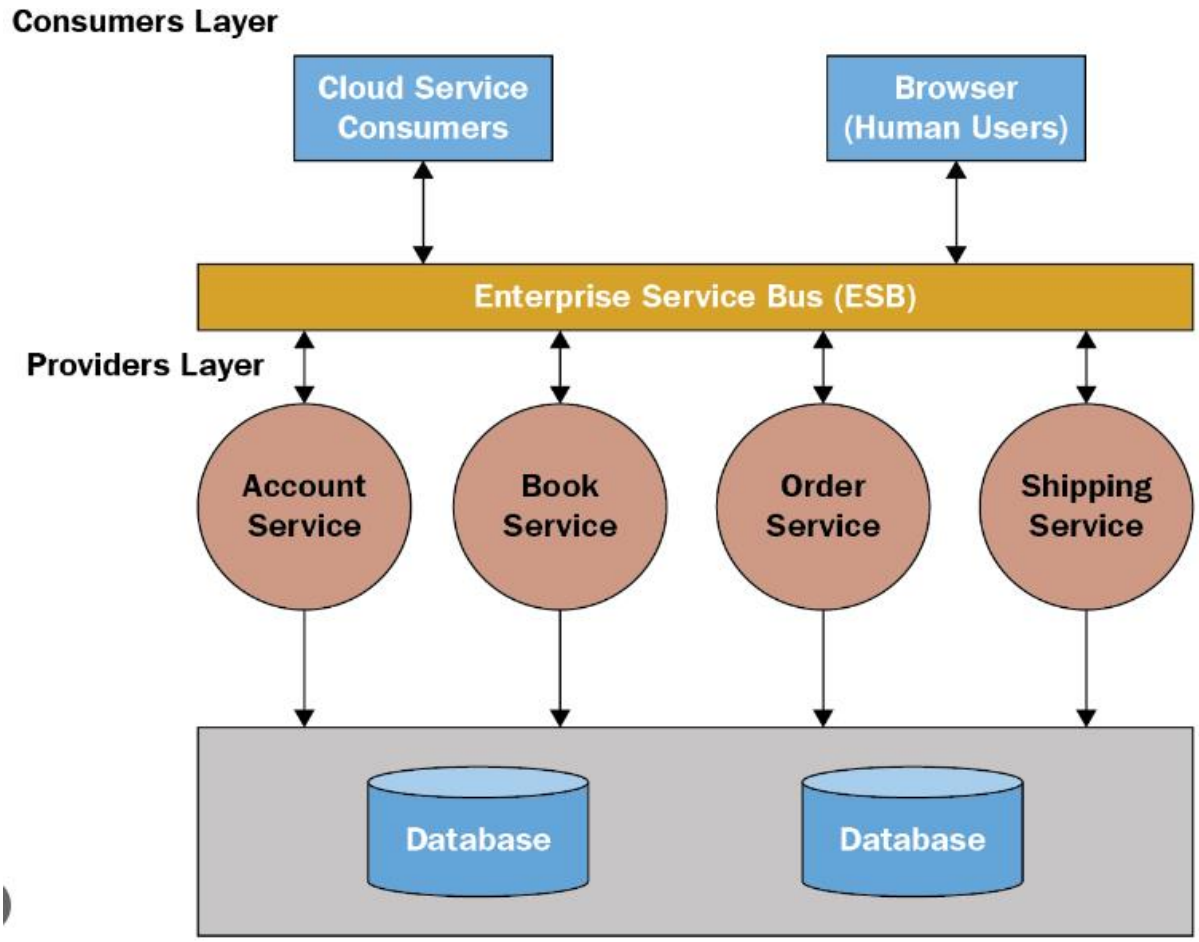
Cloud Computing Technologies

1) **Virtualization:**
2) **Service-Oriented Architecture (SOA) :** Service-Oriented Architecture (SOA) allows organizations to access on-demand cloud-based computing solutions according to the change of business needs. It can work without or with cloud computing. The advantages of using SOA is that it is easy to maintain, platform independent, and highly scalable.
   Service Provider and Service consumer are the two major roles within SOA**.**

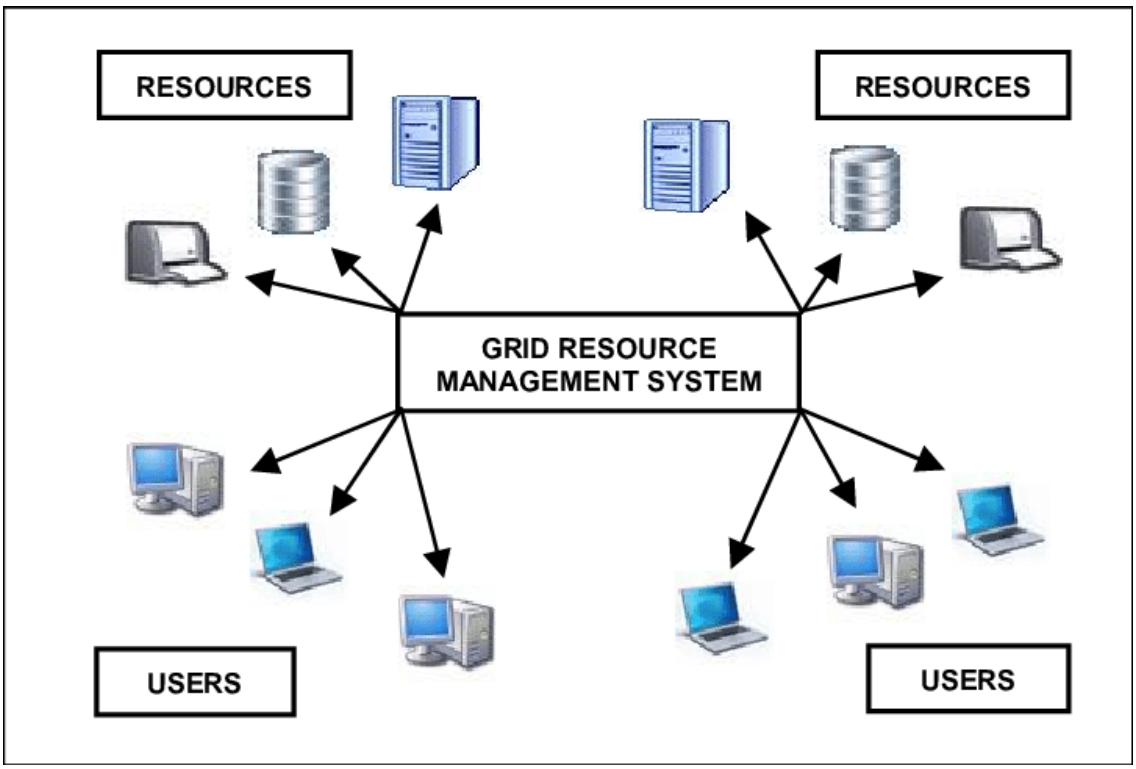
   Applications of Service-Oriented Architecture :

   ✓ It is used in the healthcare industry.
   ✓ It is used to create many mobile applications and games.
   ✓ In the air force, SOA infrastructure is used to deploy situational awareness systems.
   Situational awareness systems*: The situational awareness system combines data from various sources, including radar, cameras, sensors, and communication systems, and*

presents it in a user-friendly format to the pilot or operator. This allows them to make informed decisions and respond to changing conditions quickly and effectively.



3) **Grid Computing:** Grid computing is a way of connecting many computers together to work as a single system, sharing their resources like processing power, storage, and software applications. It allows people and organizations to solve complex problems by breaking them into smaller tasks and distributing them across many computers, working together as a grid.
For Example: Think of it like a team of people working on a project, each person doing a specific task, but they work together to achieve a common goal. Similarly, in grid computing, each computer performs a specific task and together they complete a larger task, such as scientific simulations, data analysis, or weather forecasting.



Advantages Of Grid Computing

- **Efficiency:** With grid computing, you can break down a complex task into multiple subtasks. Multiple computers can work on the subtasks together, making grid computing an efficient computational solution.
- **Cost:** Grid computing works with existing hardware, which means you can reuse existing computers. You can save costs while accessing your excess computational resources. You can also cost-effectively access resources from the cloud.
- **Flexibility:** Grid computing is not constrained to a specific building or location. You can set up a grid computing network that spans several regions. This allows researchers in different countries to work collaboratively with the same supercomputing power.

Applications:

- **Financial services**: Financial institutions use grid computing primarily to solve problems involving risk management. By harnessing the combined computing powers in the grid, they can shorten the duration of forecasting portfolio changes in volatile markets.
- **Gaming:** The gaming industry uses grid computing to provide additional computational resources for game developers. The grid computing system splits large tasks, such as creating in-game designs, and allocates them to multiple machines. This results in a faster way for the game developers**.**
- **Entertainment:** Some movies have complex special effects that require a powerful computer to create. The special effects designers use grid computing to speed up the production timeline. They have grid-supported software that shares computational resources to render the special-effect graphics.
- **Engineering:** Engineers use grid computing to perform simulations, create models, and analyze designs.

Components Of Grid Computing: In grid computing, a network of computers works together to perform the same task. The following are the components of a grid computing network.

- **Nodes:** The computers or servers on a grid computing network are called nodes. Each node offers unused computing resources such as CPU, memory, and storage to the grid network. At the same time, you can also use the nodes to perform other unrelated tasks. There is no limit to the number of nodes in grid computing. There are three main types of nodes: control, provider, and user nodes.
- **Grid Middleware:** Grid middleware is a specialized software application that connects computing resources in grid operations with high-level applications. For example, it handles your request for additional processing power from the grid computing system. It controls the user sharing of available resources to prevent overwhelming the grid computers. The grid middleware also provides security to prevent misuse of resources in grid computing.

Grid Computing VS Cloud Computing

Criteria	Grid Computing	Cloud Computing
User Management	Decentralised management	Centralised management
Dependancy	Other computer picks up the work whenever the computer stops	Totally dependent on internet
Operation	Operates within a corporate network	Can also operate through the internet
Accessibility	Through Grid middleware	Through standard Web protocols
Domains	Multiple Domains	Single Domain
Scalability	Normal	High
Architecture	Distributed computing architecture	Client-server architecture
Virtualization	Data and computing resources	Hardware and software platforms
Computation	Maximum computing	On-demand

Corporate Network: *A corporate network is a group of interconnected computers and other devices that are used by employees of a company to communicate, share resources, and collaborate on projects. A corporate network can be a local area network (LAN), which connects computers and devices within a physical office or building, or a wide area network (WAN), which connects computers and devices across different locations, such as branch offices or remote workers.*

Standard Web Protocols: *Standard web protocols are a set of rules and guidelines that define how data is transmitted and communicated over the internet. These protocols enable different devices and systems to communicate with each other in a standardized and efficient way*

Some of the most commonly used standard web protocols include:

- ***Hypertext Transfer Protocol (HTTP):*** *This protocol is used to transfer data over the World Wide Web (WWW). It defines how web pages and other online resources are requested and served.*
- ***File Transfer Protocol (FTP):*** *This protocol is used to transfer files between computers on a network. It allows users to upload and download files to and from a remote server.*

- ***Simple Mail Transfer Protocol (SMTP):*** *This protocol is used to send and receive email messages over the internet. It defines how email messages are transmitted between mail servers and clients.*
- ***Transmission Control Protocol/Internet Protocol (TCP/IP****): This is a suite of protocols that define how data is transmitted over the internet. It enables different devices and systems to communicate with each other in a standardized way.*

What is distributed computing?

Distributed computing refers to a computing system where software components are shared among a group of networked computers. However, users who use the software will see a single coherent interface. For example, a web search engine is a distributed computing system. It allows you to search a specific website by sending the request to several servers.

Distributed computing compared to grid computing

Distributed computing aims to achieve a single goal at any one time. In contrast, grid computing does not act cohesively but allocates resources on its network for multiple related subtasks. A grid computing network might consist of several distributed computing systems**.**