UNIT 3

FUNDAMENTAL CONCEPT OF DISTRIBUTED SYSTEM

COURSE CONTAINS

3.1 Introduction to Distributed Systems

3.2 Distributed Systems over Centralized Systems

3.3 Client-Server System

3.4 Peer-to-peer System

3.5 Examples of Distributed Systems

3.6 Main Characteristics of Distributed System

3.7 Advantages and Disadvantages of Distributed System

COURSE CONTAINS…….

3.8 Design Goals of Distributed Systems

3.9 Main Problems of Distributed Systems

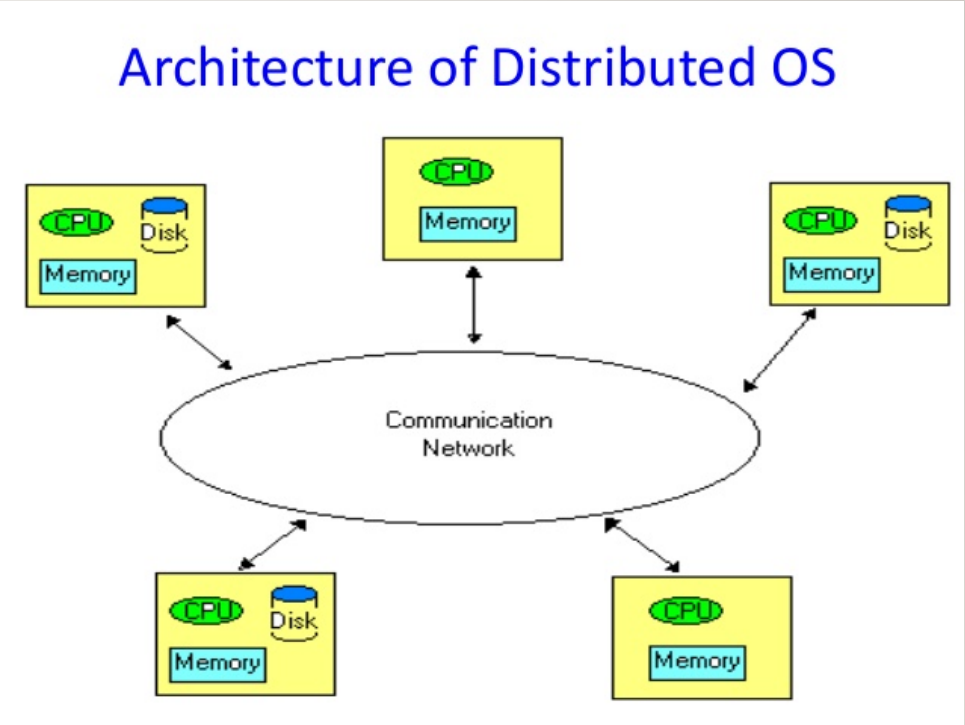
3.10 Models of Distributed Systems

3.11 Resource Sharing and the Web Challenges

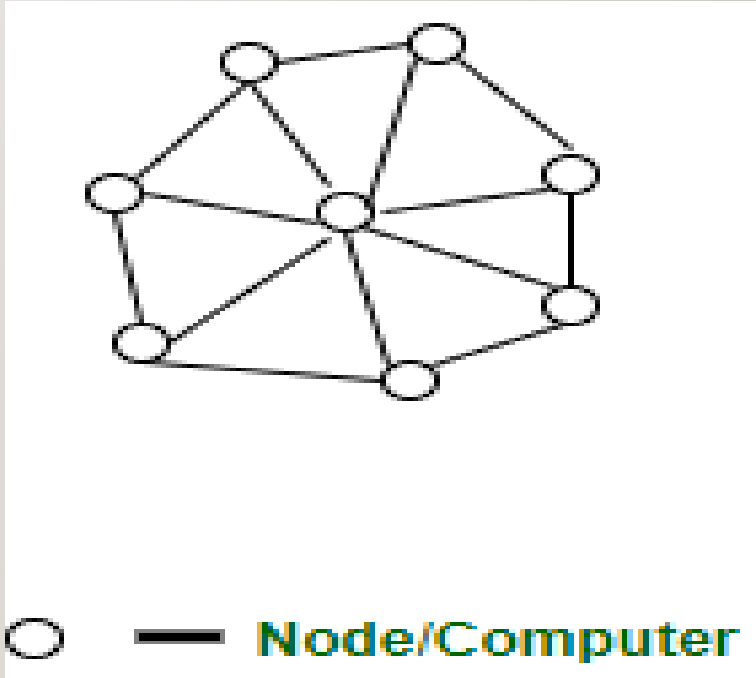
3.12 Types of Distributed System: Grid , Cluster , Cloud

3.1 INTRODUCTION TO DISTRIBUTED SYSTEMS

- A distributed system is a **collection of independent computers** that appear to the users of the system as a **single computer (single coherent system).**
- A distributed system is a collection of autonomous computers and they are communicate and coordinate with each other only by **passing message.**
- A distributed system contains multiple nodes that are physically separate but **linked together using the network.**
- Each computer consists of a single CPU , memory , peripherals and some terminals.



- In distributed system a large number computer connected by high speed network.
- These independent, autonomous computers are linked by a network to share information, communicate, and exchange information easily.



Components of Distributed System are:

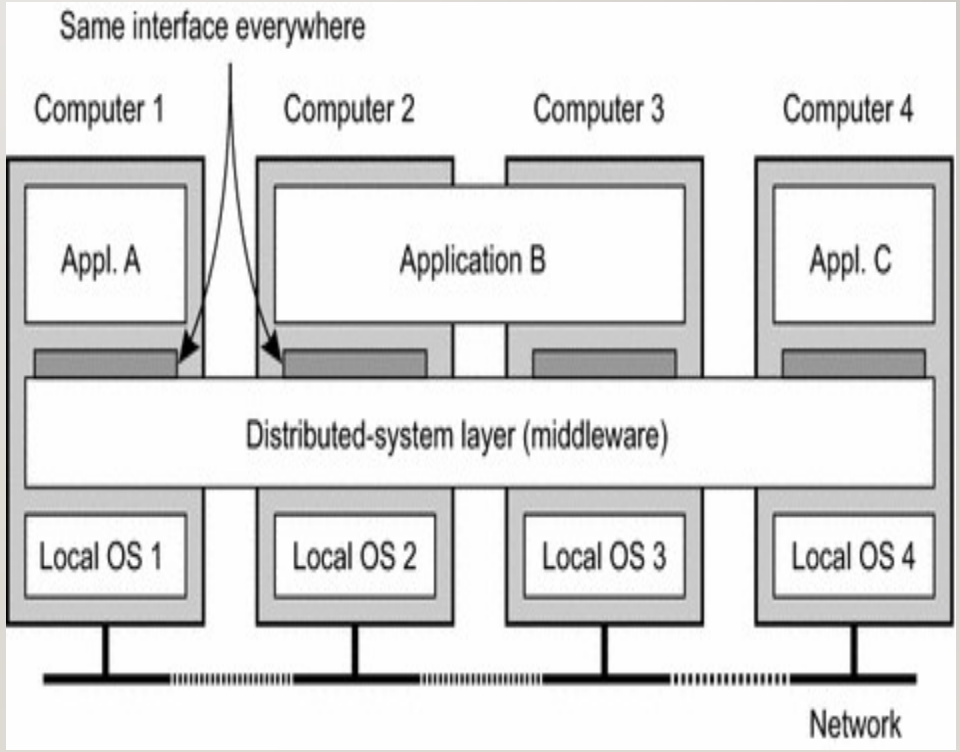
Components of Distributed System are,

- Node (Computer, Mobile, etc.)
- Communication link (Cables, Wi-Fi, etc.)

- Distributed system is collection of loosely coupled processor i.e a computer is independent to each other.
- Sharing of resources is a main motivation for constructing distributed systems.
- In Distributed System, jobs are distributed among several processor such that the calculation is distributed to multiple computers. Example: when you have a large amount of data then you can divide it and send each part to particular computers which will make the calculations for their part. i.e load distribution

ORGANIZATION OF DISTRIBUTED SYSTEM

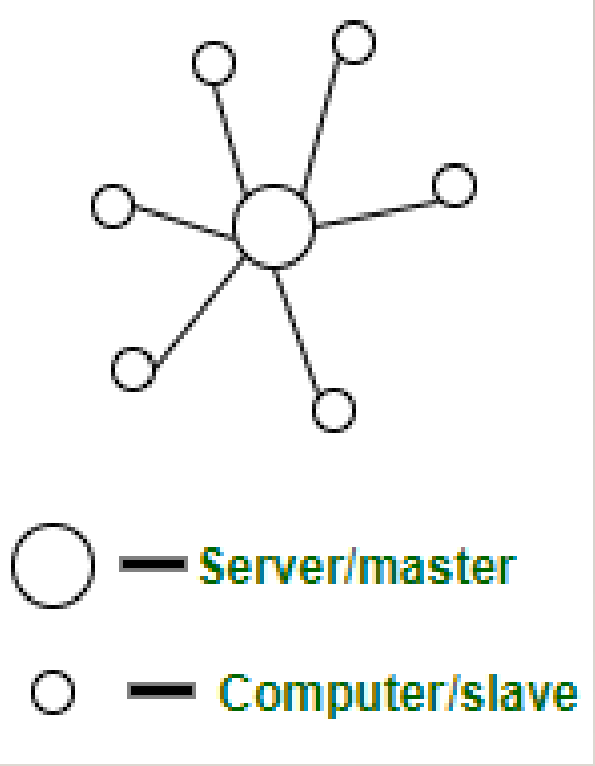
- A distributed system organized as middleware .
- The middleware layer extends over multiple machines , and offers each application the same interface.
- The different processors have their own local memory. They use a distribution middleware. They help in sharing different resources and capabilities to provide users with a single and integrated coherent network.



3.2 DISTRIBUTED SYSTEMS OVER CENTRALIZED SYSTEMS

Centralized System: State stored on a single computer

- A centralized system have a single server that handles all the major processing i.e all calculations are done on one particular computer (system). Example: you have a dedicated server for calculating data.
- Centralized systems are systems that use client/server architecture where one or more client nodes are directly connected to a central server. This is the most commonly used type of system in many organizations where client sends a request to a company server and receives the response.
- Centralized system is a master-slave based **system** where all information and commands are passed through the central unit (master).



- **Example –**Wikipedia. Consider a massive server to which we send our requests and the server responds with the article that we requested. Suppose we enter the search term ‘junk food’ in the Wikipedia search bar. This search term is sent as a request to the Wikipedia servers (mostly located in Virginia, U.S.A) which then responds back with the articles based on relevance. In this situation, we are the client node , wikipedia servers are central server.
- Centralized computing: Early computing was performed on a single processor. Uniprocessor computing can be called Centralized computing.
- In a centralized system, all nodes are connected to a central network owner or “server”. The server stores data, which other nodes can access,.

Components of Centralized System are

1. Node (Computer, Mobile, etc.)
2. Server
3. Communication link (Cables, Wi-Fi, etc.)

CHARACTERISTICS OF CENTRALIZED SYSTEM

1.Presence of a global clock: As the entire system consists of a central node(a server/ a master) and many client nodes(a computer/ a slave), all client nodes sync up with the global clock(the clock of the central node).

2.One single central unit: One single central unit which serves/coordinates all the other nodes in the system.

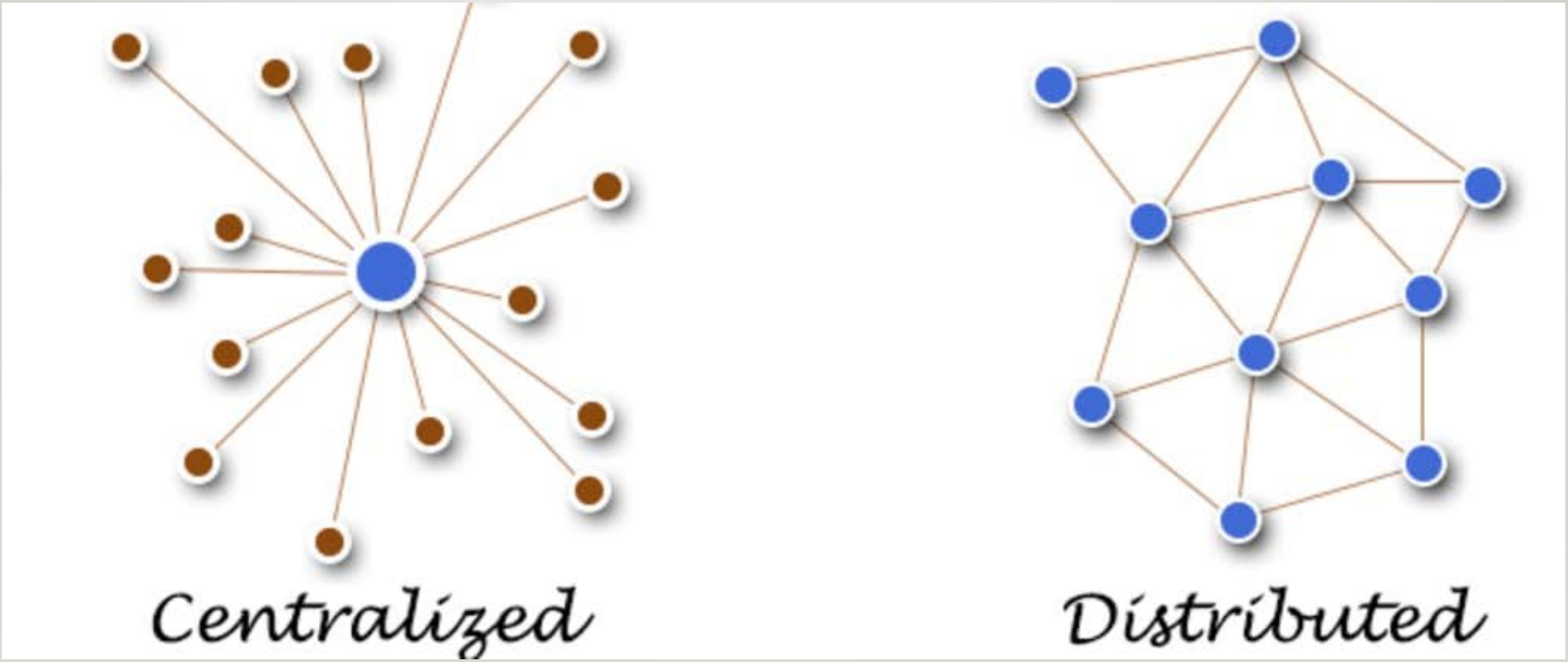
3.Dependent failure of components: Central node failure causes entire system to fail. This makes sense because when the server is down, no other entity is there to send/receive response/requests.

ADVANTAGES OF CENTRALIZED SYSTEM

1. A centralized system is **easy to set up** and can be **developed quickly.**
2. As centralized systems consists of less implementation costs than distributed system, and single server handles all the major processing so they are **easy to maintain.**
3. Organization **better controlled** and **managed better security** as all data and processing task are performed and kept at the central location.
4. **Quick updates are possible** – Only one machine to update.

DISADVANTAGES OF CENTRALIZED SYSTEM –

1. **Highly dependent** on the network connectivity – System can fail if the nodes lose connectivity as there is only one central node.
2. **Less possibility of data backup**. If the server node fails and there is no backup, you lose the data straight away
3. **Longer access times** to data for users who are far from the server.
4. The system **performance** for each user **decrease** when many users try to attach simultaneously.
5. The **scalability** of the mainframe systems is **extremely low.**


Centralized
Distributed

Advantages of Distributed Systems over Centralized Systems:

1. Economics:

A Collection of microprocessors offer a better price/performance than mainframes . Although distributed systems consists of high implementation costs, they are cost effective in the long run compared to centralized system.

2. Speed:

A distributed system may have more total computing power than a mainframe. Example: 10,000 CPU chips, each running at 50 MIPS(million instruction per second).Not possible to build 500,000 MIPS single processor since it would require 0.002nsec instruction cycle. Enhanced performance through load distributing.

Advantages of Distributed Systems over Centralized Systems:

3. Inherent distribution:
Some applications involve spatially separated machines. Example : Computerized worldwide Airline reservation ,Computerized banking system in which a customer can deposit/withdraw money from his/her account from any branch of the bank.

4. Reliability:
If one machine crashes, the system as a whole can still survive . Higher availability and improved reliability

5. Incremental growth:
Computing power can be added in small increments

DIFFERENCE

CENTRALIZED SYSTEM

1. State stored on a single computer
2. Dependence : If the main system fails , everything stops.
3. No data or application redundancy as the whole system is managed by the main computer at central location.
4. Easier to understand

DISTRIBUTED SYSTEM

1. State stored over multiple computer
2. Independence : no harm to other unit as one fails.
3. Redundancy of data and application are likely.
4. More complex

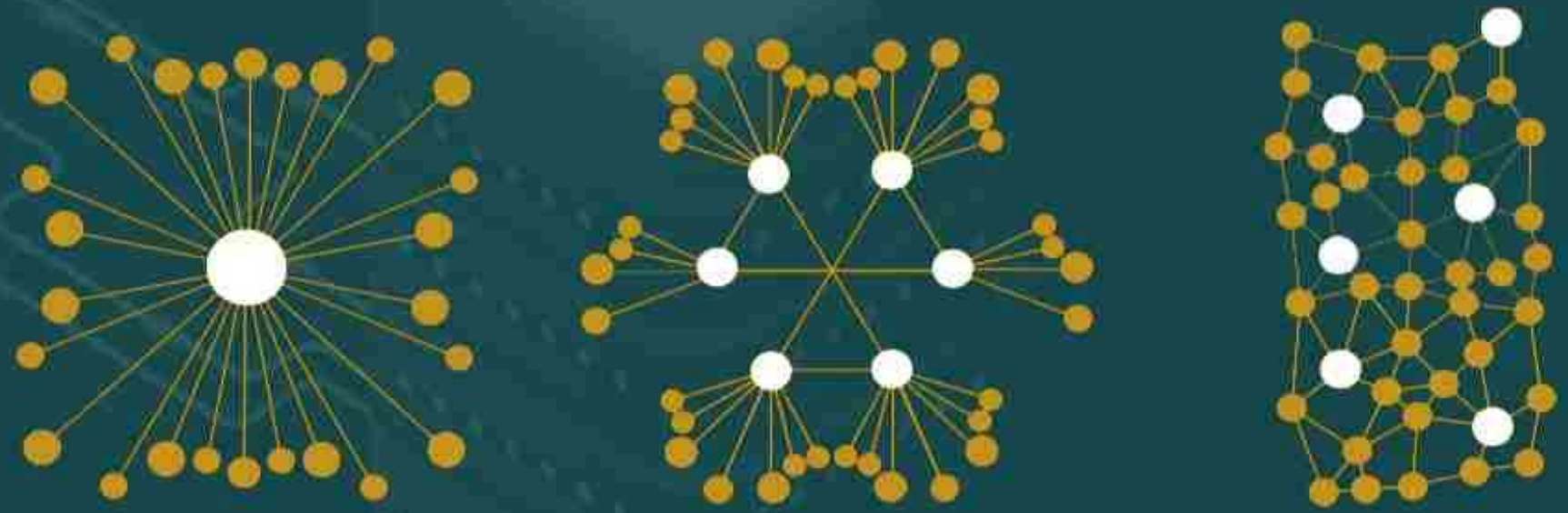
CENTRALIZED SYSTEM

5. Centralized systems have non-autonomous components.
6. better security as all data and processing task are performed and kept at the central location.
7. Centralized systems are often build using homogeneous technology.

DISTRIBUTED SYSTEM

5. Distributed systems have autonomous components.
6. As all data are distributed, data security is the major issues.
7. Distributed systems may be built using heterogeneous technology.

Comparison – Centralized, Decentralized and Distributed Systems



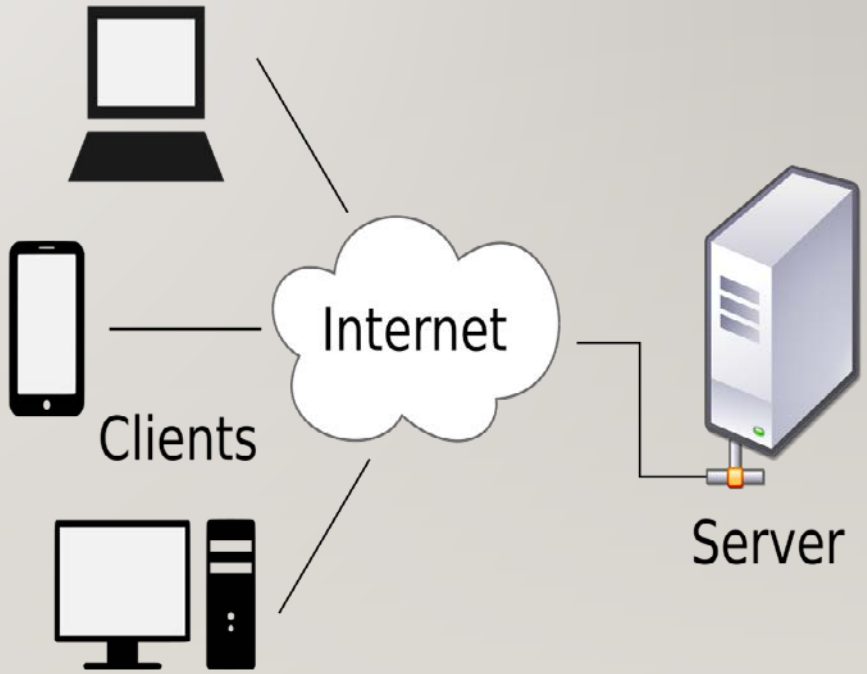
Centralized Systems

Decentralized Systems

Distributed Systems

3.3 CLIENT-SERVER SYSTEM

- Client-server system is a computing system that is composed of two logical parts: the **clients** requests a resource or services and the **server** provides that resource or services.
- A server may serve multiple clients at the same time while a client is in contact with only one server.
- Both the client and server usually communicate via a computer network or the Internet but sometimes they may reside in the same system
- In a client server system, there are clients and servers. A client can be a device or a program. It helps the end users to access the web. Example : servers include web servers, mail servers, and file servers. Each of these servers provide resources to **client** devices, such as desktop computers, laptops, tablets, and smartphones.



CLIENT-SERVER SYSTEM….

- Client/server systems are not limited to traditional computers. An example is an automated teller mac hine (ATM) network. Customers typically use ATMs as clients to interface to a server that manages all of the accounts for a bank. This server may in turn work with servers of other banks (such as when wi thdrawing money at a bank at which the user does not have an account). The ATMs provide a user inte rface and the servers provide services, such as checking on account balances and transferring money b etween accounts.

CHARACTERISTICS OF CLIENT SERVER SYSTEM

Some of the characteristics for client server computing are as follows:

- The client server computing works with a system of request and response. The client sends a request to the server and the server responds with the desired information.
- The client and server should follow a common communication protocol so they can easily interact with each other. All the communication protocols are available at the application layer.
- A server can only accommodate a limited number of client requests at a time. So it uses a system based to priority to respond to the requests.
- An example of a client server computing system is a web server. It returns the web pages to the clients that requested them.

ADVANTAGES OF CLIENT SERVER COMPUTING

The different advantages of client server computing are −

1. **Centralization :** access, resources, and integrity of the data are controlled by the dedicated server so that a program or unauthorized client cannot damage the system.
2. **Security:** All the required data is concentrated in a single place i.e. the server. So it is easy to protect the data and provide authorization and authentication.
3. **Proper Management :** All the files are stored at the same place. In this way, management of files becomes easy. Also it becomes easier to find files.

ADVANTAGES OF CLIENT SERVER COMPUTING

4. **Back-up and Recovery possible :** As all the data is stored on server its easy to make a back-up of it. Also, in case of some break-down if data is lost, it can be recovered easily and efficiently

5. **Scalability**: We can increase the capacity of clients and servers separately. Any element can be increased or enhanced at any time, or we can add new nodes to the network (clients or servers).

6. **Accessibility :**The server need not be located physically close to the clients. Yet the data can be accessed efficiently.

7. It is easy to replace, upgrade or relocate the nodes in the client server model because all the nodes are independent and request data only from the server.

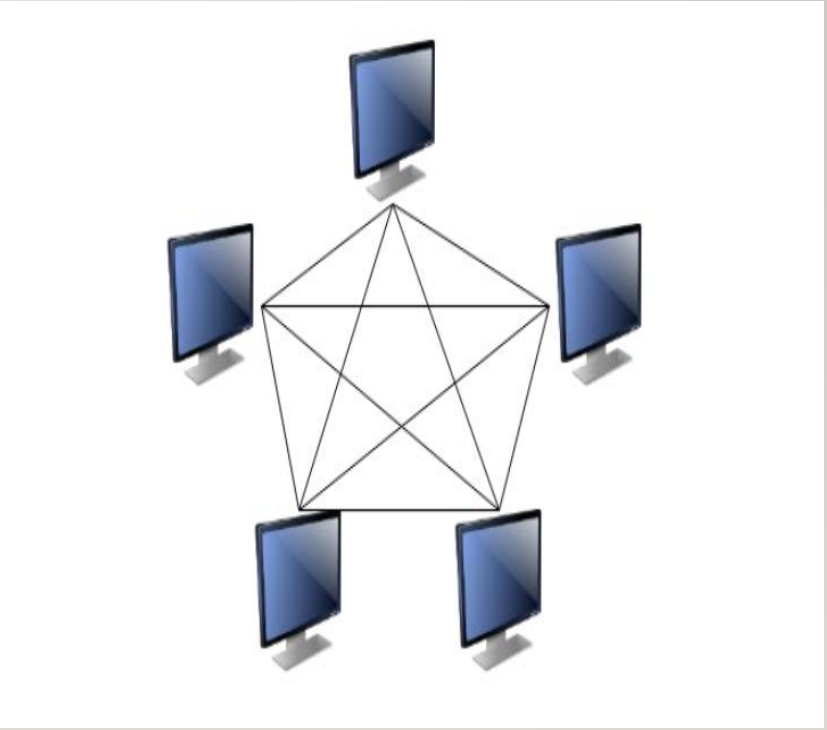
DISADVANTAGES OF CLIENT SERVER COMPUTING

The different disadvantages of client server computing are −

1. **Traffic Congestion** : The primary disadvantage of client server system is the traffic congestion it undergoes. If too many clients make request from the same server, it will result in crashes or slowing down of the connection. An overloaded server creates many problems in accessing information.
2. **Robustness:** As we all know client server system are centralized. If the main serve fails for any reason, then the whole system will be disrupted. Therefore, client server system lacks on the side of robustness.
3. **Cost:** The cost of setting and maintaining a client server model are quite high
4. **Maintenance:** When the servers are implemented, it is going to work non-stop. Which means it must be given proper attention. If there are any problems, it must be resolved immediately without any delay. Hence, there should be a specialized network manager appointed to maintain the server.

3.4 PEER-TO-PEER SYSTEM

- Peer to peer (P2P) is a type of architecture in which nodes are interconnected with each other **through the internet** and share resource with each other without the need of central server.
- In a peer to peer system, there is no specific client or a server. A device can send and receive data directly with each other. Each node can either be a client or a server. It can request or provide services accordingly.
- Peer to peer computing is a distributed application architecture that partitions tasks or workloads between peers. All the tasks are equally divided between all the nodes. **A node is also called a peer.**
- A diagram to better understand peer to peer computing is as follows −



CHARACTERISTICS OF PEER TO PEER COMPUTING

The different characteristics of peer to peer system are as follows −

1. Peer to peer networks are usually formed by groups of computers or node. These computers store their data using individual security but also share data with all the other nodes.
2. The nodes in peer to peer system both use resources and provide resources. So, if the nodes increase, then the resource sharing capacity of the peer to peer network increases.
3. Since nodes in peer to peer system act as both clients and servers, it is difficult to provide adequate security for the nodes. This can lead to denial of service attacks.
4. Most modern operating systems such as Windows and Mac OS contain software to implement peer to peer system.

ADVANTAGES OF PEER TO PEER COMPUTING

Some advantages of peer to peer computing are as follows −

1. **Cost** :The overall cost of building and maintaining a peer to peer system is relatively inexpensive. The setup cost has been greatly reduced due to the fact that there is no central configuration.
2. **Reliability**: Peer to Peer system is not dependent on a centralized system, which means that the connected computers can function independently with each other. Even if one part of the network fails, it will not disrupt other parts. Only the user will not be able to access those files
3. **Scalability** :Peer to Peer system has one of the best scalability features. Even if there are extra nodes added, the performance of the network will remain the same.
4. **Resource Sharing :** In Peer to Peer system , the resources are shared equally among all the users. The connected devices can provide and consume resources at the same time.
5. **Server Requirement :** In peer to peer system, each connected computers acts as a server and a workstation. Therefore, there is no need to use a dedicated server. All the authorized users can use their respective client computer to access the required files. This can lead to saving more overhead costs.

DISADVANTAGES OF PEER TO PEER COMPUTING

Some disadvantages of peer to peer computing are as follows −

1. **Decentralization :** Peer to Peer networking lacks the feature of centralization. There is no central server, thus files are stored on individual machines. The entire network accessibility is not in the hands of a single person. This makes it more challenging for the users to locate and find files.
2. It is difficult to backup the data as it is stored in different computer systems and there is no central server.
3. It is difficult to provide overall security in the peer to peer network as each system is independent and contains its own data

DIFFERENCE BETWEEN CLIENT SERVER COMPUTING AND PEER TO PEER COMPUTING

CLIENT SERVER SYSTEM

1. A distributed application structure based on resource or service providers called servers and service requester called clients.
2. Client request for service and server responds with a service.
3. A centralized system.

PEER TO PEER SYSTEM

1. A distributed application architecture that partitions tasks or workloads between peers.
2. Each node can request for services and provide services.
3. A decentralized system.

CLIENT SERVER SYSTEM

4. Clients depend on the server, failure in the server will disrupt the functioning of all clients.
5. Access time for a service is higher.
6. More stable and secure.

PEER TO PEER SYSTEM

4. Reliable as there are multiple service providing nodes.
5. Service requesting node does not need to wait long.
6. Comparatively less stable.

3.5 EXAMPLES OF DISTRIBUTED SYSTEMS

Some of the examples of distributed system are listed below:

1. Network
2. Telecommunication networks
3. World Wide Web(WWW)
4. Distributed artificial intelligence
5. Distributed Real-time Systems
6. Distributed Database Systems

EXAMPLES OF DISTRIBUTED SYSTEMS….

1. Network

- The earliest example of a distributed system happened in the 1970s when ethernet was invented and LAN (local area networks) were created. For the first time computers would be able to send messages to other systems with a local IP address.
- Peer-to-peer networks evolved and e-mail and then the **Internet** as we know it continue to be the biggest, ever growing example of distributed systems.
- As the internet changed from IPv4 to IPv6, distributed systems have evolved from "LAN" based to "Internet" based. **The internet is a vast interconnected collection of computer networks of many different types.**

EXAMPLES OF DISTRIBUTED SYSTEMS…….

2. Telecommunication networks:

- Telephone networks and Cellular networks- Land line and cellular.

3. World Wide Web(WWW)

- WWW is the biggest example of distributed system- information, resource sharing.
- Nowadays, millions of people use the World Wide Web for different purposes such as access information stored on Web servers situated anywhere on the glove, email, reading news, online shopping, downloading music or play games.

4. Distributed artificial intelligence

- Distributed Artificial Intelligence is a way to use large scale computing power and parallel processing to learn and process very large data sets using multi-agents.

5. Distributed Real-time Systems

Many industries use real-time systems that are distributed locally and globally.

- Aircraft control systems -Airlines use flight control systems,
- manufacturing plants use automation control systems,
- logistics and e-commerce companies use real-time tracking systems.

6. Distributed Database Systems

A distributed database is a database that is located over multiple servers and/or physical locations. The data can either be replicated or duplicated across systems.

- Most popular applications use a distributed database and need to be aware of the homogenous or heterogenous nature of the distributed database system.
- A homogenous distributed database means that each system has the **same** database management system and data model. They are easier to manage and scale performance by adding new nodes and locations.
- Heterogenous distributed databases **allow for multiple** data models, different database management systems. Gateways are used to translate the data between nodes and usually happen as a result of merging applications and systems.

And also

Facebook and google use distributed system extensively

- Massive scale
- Fast enough
- Very reliable

Google search system. Each request is worked upon by hundreds of computers which crawl the web and return the relevant results. To the user, the Google appears to be one system, but it actually is multiple computers working together to accomplish one single task (return the results to the search query).

APPLICATION OF DISTRIBUTED SYSTEM

Finance and commerce	eCommerce e.g. Amazon and eBay , PayPal , online banking and trading
The information society	Web information and search engines , ebooks , Wikipedia, social networking : Facebook
Creative industries and entertainment	Online gaming , music and film in the home , user-generated content, e.g. YouTube
Healthcare	Health informatics, on online patient records, monitoring patients
Education	E-learning , virtual learning environment
Transport and logistics	GPS in route finding systems, map services : Google Maps

3.6 MAIN CHARACTERISTICS OF DISTRIBUTED SYSTEM

1. Resource sharing
2. Heterogeneity
3. Openness
4. Scalability
5. Transparency
6. Concurrency
7. Fault Tolerance
8. Lack of a global clock

MAIN CHARACTERISTICS OF DISTRIBUTED SYSTEM….

1. Resource Sharing

- Resource sharing means that the existing resources in a distributed system can be accessed or remotely accessed across multiple computers in the system.
- Computers in distributed systems shares resources like
  - → hardware (disks and printers),
  - → software (files, windows and data objects) and
  - → data.
- Hardware resources are shared for reductions in cost and convenience.
- Data is shared for consistency and exchange of information.
- Resources are managed by a software module known as a resource manager.

2. Heterogeneity

- In distributed systems components can have **variety and differences** in Networks, Computer hardware, Operating systems, Programming languages and implementations by different developers.
- Heterogeneity (that is, variety and difference) applies to all of the following:
  - → Hardware devices: computers, tablets, mobile phones, embedded devices, etc.
  - → Operating System: Ms Windows, Linux, Mac, Unix, etc.
  - → Network: Local network, the Internet, wireless network, satellite links, etc.
  - → Programming languages: Java, C/C++, Python, PHP, etc.
  - → Different roles of software developers, designers, system managers

3. Openness

- Openness is concerned with extensions and improvements of distributed systems , because it will fulfill the future needs and update. Example – facebook .
- The distributed system must be open in terms of Hardware and Software.
- In order to make a distributed system open,

a. A detailed and well-defined interface of components must be published.

b. The new component must be easily integrated with existing components

4. Scalability

- A system is said to be scalable if it can handle the **addition of users and resources without suffering a noticeable loss of performance** or increase in administrative complexity.
- Mostly we scale the distributed system by adding more computers in the network.
- System should work efficiently with an increasing number of users.
- System performance should increase with inclusion of additional resources.
- Distributed systems are made on default to be scalable. Whenever there is an increase in workload, users can add more workstations. There is no need to upgrade a single system. Moreover, no any restrictions are placed on the number of machines. Which means that, these machines will be able to handle high demand workload easily.

5. Transparency

- Transparency is defined as the **hiding of the complexity of a system from user and the application programmer.**
- It hides the fact that the processes and resources are physically distributed across multiple computer and appear to the users of the system as a **single computer**
- **Aim :** to make certain aspects of distribution invisible to the application programmer so that they need only be concerned with the design of their particular application.
- Transparency is of various forms like access, location , migration , relocation , replication , concurrency, failure transparency etc..

FORMS OF TRANSPARENCY

Transparency	Description
Access	Hide differences in data representation and how a resources is accessed.
Location	Hide where a resource is located
Migration	Hide that a resource may move from one place to another.
Relocation	Hide that a resource may be moved to another location while in use.
Replication	Hide the multiple copies from users.
Concurrency	Hide the fact that the resource that one user are using is also use by other user at the same time.
Failure	Hide the failure from the users

6. Concurrency

- Concurrency is a property of a system representing the fact that **multiple activities are executed at the same time**.
- Components in distributed systems are executed in concurrent processes.
- Concurrency reduces the latency and increases the throughput of the distributed system.

7. Fault Tolerance

- In a distributed system hardware, software, network anything can fail.
- Fault tolerance enable a system to continue to operate , even in the presence of faults.
- The system must be designed in such a way that it is available all the time even after something has failed.

8. Lack of a global clock: The most important one is that in a distributed system, the different nodes maintain their own time using local clocks and their time values may not be same for the different nodes. I.e. there is no global clock within the system .

3.7 ADVANTAGES OF DISTRIBUTED SYSTEM

Some advantages of Distributed Systems are as follows :

1. Reliability, high fault tolerance:

The distributed systems are far more reliable than single systems in terms of failures. Failure of one node does not lead to the failure of the entire distributed system. Other nodes can still communicate with each other

2. Scalability:

More nodes can easily be added to the distributed system i.e. it can be scaled as required.

3.Flexibility:

It makes it easy to install, implement and debug new services.

4.Efficiency:

Distributed systems are made to be efficient in every aspect since they posses multiple computers. Each of these computers could work independently to solve problems. This not only considered to be efficient, it significantly saves time of the user.

5. Resource sharing :

All the nodes in the distributed system are connected to each other. So nodes can easily share data and resources with other nodes.

6. Fast calculation speed:

A distributed computer system can have the computing power of multiple computers, making it faster than other systems.

7. Openness:

Since it is an open system, it can be accessed both locally and remotely.

8. High performance:

Distributed systems are extremely efficient because work loads can be broken up and sent to multiple machines.

9. Cost Effective :

Although distributed systems consists of high implementation costs, they are cost effective in the long run. The initial cost is higher than a traditional system, but because of their scalability, they quickly become more cost effective

3.7 DISADVANTAGES OF DISTRIBUTED SYSTEM

Some disadvantages of Distributed Systems are as follows :

1. Startup Cost:

- Compared to a single system, implementing and maintaining a complex system architecture of a distributed system is significantly higher.
- The infrastructure used in a distributed system makes it expensive.

2. Security

- Distributed systems always comes with security risks since it contains open system characteristics.
- The data of the user is stored in different workstations. Thus, the user needs to make sure that their data is secured in each of these computers. Moreover, unlike in a centralized computing system, it is not an easy task to manage data access in a distributed system.

3. Complexity

- The difficulty involved in implementation, maintenance and troubleshooting makes distributed system a complex strategy.
- Besides hardware complexity, distributed system posses difficulty in software too. The software used in distributed system needs to be well attentive when handling communication and security.

4. Overloading

- Overloading is a common problem faced by a distributed system. Overloading may occur in the network if all the nodes of the distributed system try to send data at once.

5. Network Errors

- Distributed systems are prone to network errors which results in communication breakdown. The information and data may fail to be delivered or not in the correct sequence.

3.8 DESIGN GOALS OF DISTRIBUTED SYSTEMS

The design goals of distributed systems are:

1. Resource Sharing :

- The main goal of a distributed system is to make it easy for users to access remote resources and to share resources with other users. Examples: printers, files etc .
- A distributed system make it easier for users to exchange information.

2. Openness :

- Openness is an important goal of distributed system which concerned with **extensions and improvements** of distributed systems.
- Open distributed system must be flexible making it easy to configure and add new components without affecting existing components.

3. Transparency:.

- The most important goal of distributed is to hide the user from the fact that the processes and resources are physically distributed across multiple computer.
- A distributed system that can portray itself as a single system is said to be transparent.

4. Scalability :

- A system is described as scalable if it remains effective when there is a significant increase in the number of resources and the number of users. Relatively easy to expand or scale

5**. Availability** :

- Another goal of distributed system is that it is continuously available even though certain parts may be temporarily out of order

3.9 MAIN PROBLEMS OF DISTRIBUTED SYSTEMS

Some of the issues or challenges of distributed system are mentioned below:

1. Scalability
2. Security and Privacy
3. Transparency
4. Concurrency
5. Performance
6. Flexibility
7. Failure Handling

1. Scalability:

- A system is said to be scalable if it can handle the addition of users and resources without suffering a noticeable loss of performance or increase in administrative complexity.
- But, this always does not happen because of the problems in the followings.

a) Size- growth in number of users and resources makes a distributed system overloaded because it must process too many user request.

b) Geography- increase in geography(distance between the nodes) generally results in greater communication delays and the potential for communication failure.

c) Administration- As the size of distributed systems increases, many of the system needs to be controlled. Problem associated is administrative mess.

2. Security and Privacy

- Security is one of the main issue in distributed system. Since distributed systems deal with sensitive data and information , So the system must have a strong security and privacy measurement.
- Protection of distributed system assets including base resources , storage , communications and user-interface I/O as well as higher-level composites of these resources like files , messages, display windows and more complex objects are important issues in distributed system.
- Security for information resources has three components: confidentiality, integrity, and availability.

3. Transparency

- Complete transparency is not always desirable due to the trade-offs with performance and scalability as well as the problems that can be caused while confusing local and remote operations.
- Is not always possible because of natural limitations on how fast communication take place in WAN.

4. Concurrency

- Both services and applications provide resources that can be shared by different clients in a distributed system.
- Therefore there is a possibility that several clients will attempt to access a shared resources at the same time which affect the system performance.
- Example : Online Auction

A data structure that records bids for an auction may be accessed very frequently when it gets close to the deadline time.

- Each resource (servers, object in applications…) must be designed to be safe in a concurrent environment.

5. Performance

- Any system should aim for maximum performance.
- But in the case of distributed systems this is a particularly interesting challenge, since it directly conflicts with some other desirable properties like
  - Transparency
  - Security
  - Dependability
  - Scalability can easily be detrimental to performance.

6. Flexibility

- A flexible distributed system can be configured to provide exactly the services that a user or programmer needs.
- But for that kind of flexibility a system should have the number of key properties.

  - Extensibility
  - Openness
  - Interoperability

Which is very difficult to achieve because it contradicts with the other desirable properties as well.

7. Failure Handling

- When some faults occur in hardware and the software program, it may produce incorrect results or they may stop before they have completed the intended computation so corrective measures should to implemented to handle this case.
- Failure handling is difficult in distributed systems because the failure is partial i.e., some components fail while others continue to function.
- It is difficult to notice and recover a failure in real time is an issue.

3.10 MODELS OF DISTRIBUTED SYSTEMS

There are two types of models and they are listed below:

1. Architectural Models
2. Fundamental Models

1. ARCHITECTURAL MODELS

- Architectural models deals with organization of component across the network of computers and their relationship.
- Architectural model consider
  - → Organization of components
  - → Placement of components across a network of computers
  - → Relationship between components.
- Simply, architectural model describes responsibilities distributed between system components and how are these components placed.

Common architecture for distributed system are:

A. Client-server model and

B. Peer-to-Peer model

Client-server and peer-to-peer can be considered as basic models.

Other

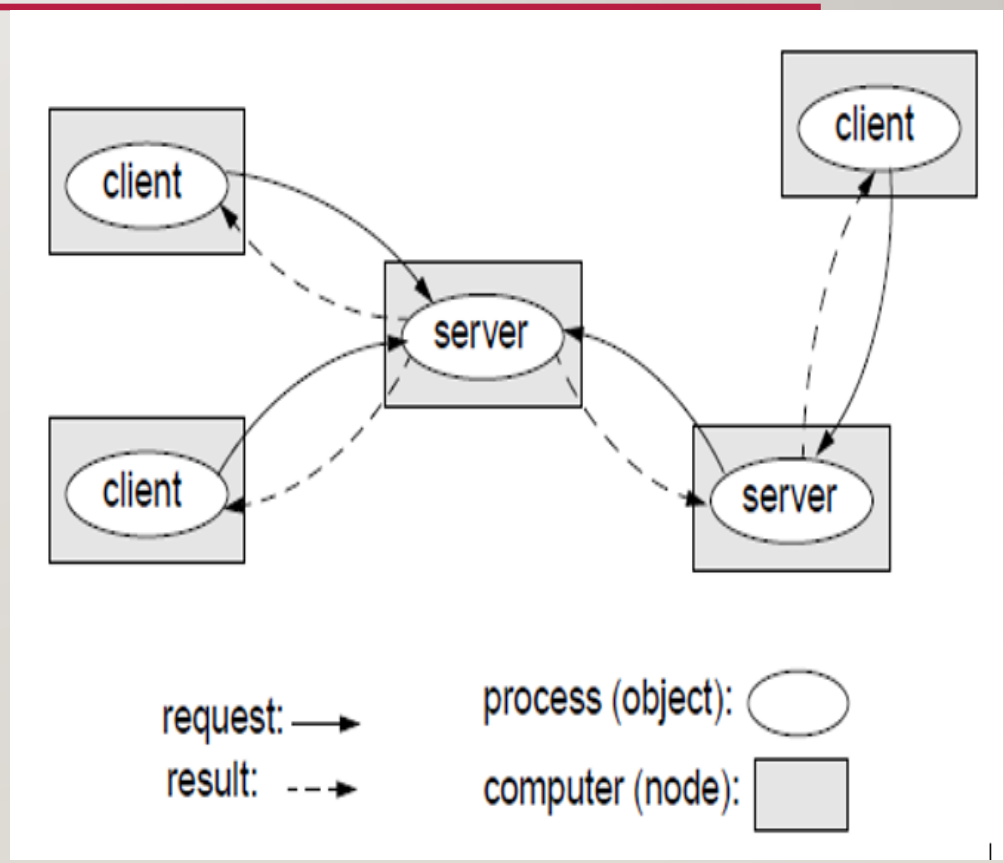
-Proxy server

-Mobile code and Mobile agents

- Mobile devices

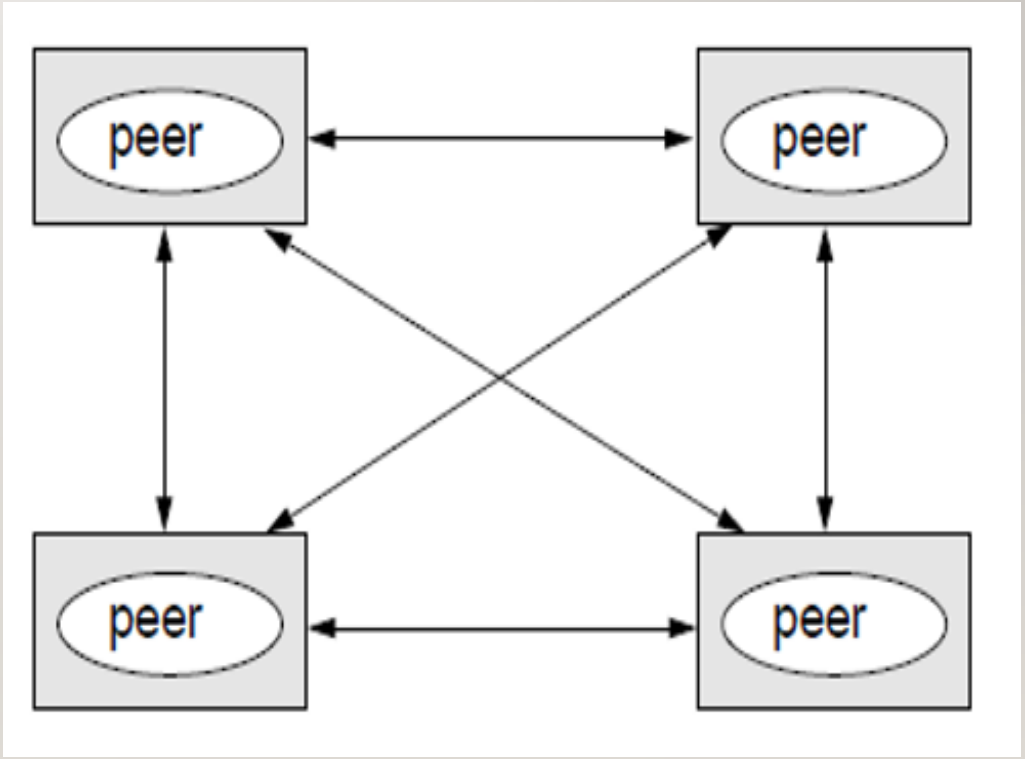
A. CLIENT-SERVER MODEL

- The system is structured as a set of processes(object) called clients, and the servers.
- the client sends a request (invocation) message to the server asking for some service.
- the server return a result to client as per requested . Server is responsible for accepting, processing, and replying to requests send by a client.
- basic web follows the client-server model. Your browser is the client. It requests web pages from a server (e.g., google.com), waits for results, and displays them for the user.
- A server can itself request services from other servers; in this new relation, the server itself acts like a client.



B. PEER-TO-PEER MODEL.

- Peer to peer (P2P) model is a type of architecture in which computer(nodes) are interconnected with each other and share resource with each other without the need of central server.
- In a peer to peer system, there is no specific client or a server. A node can send and receive data directly with each other.
- Each node can either be a client or a server. It can request or provide services accordingly.
- Peer to peer computing is a distributed application architecture that partitions tasks or workloads between peers. All the tasks are equally divided between all the nodes. **A node is also called a peer.**

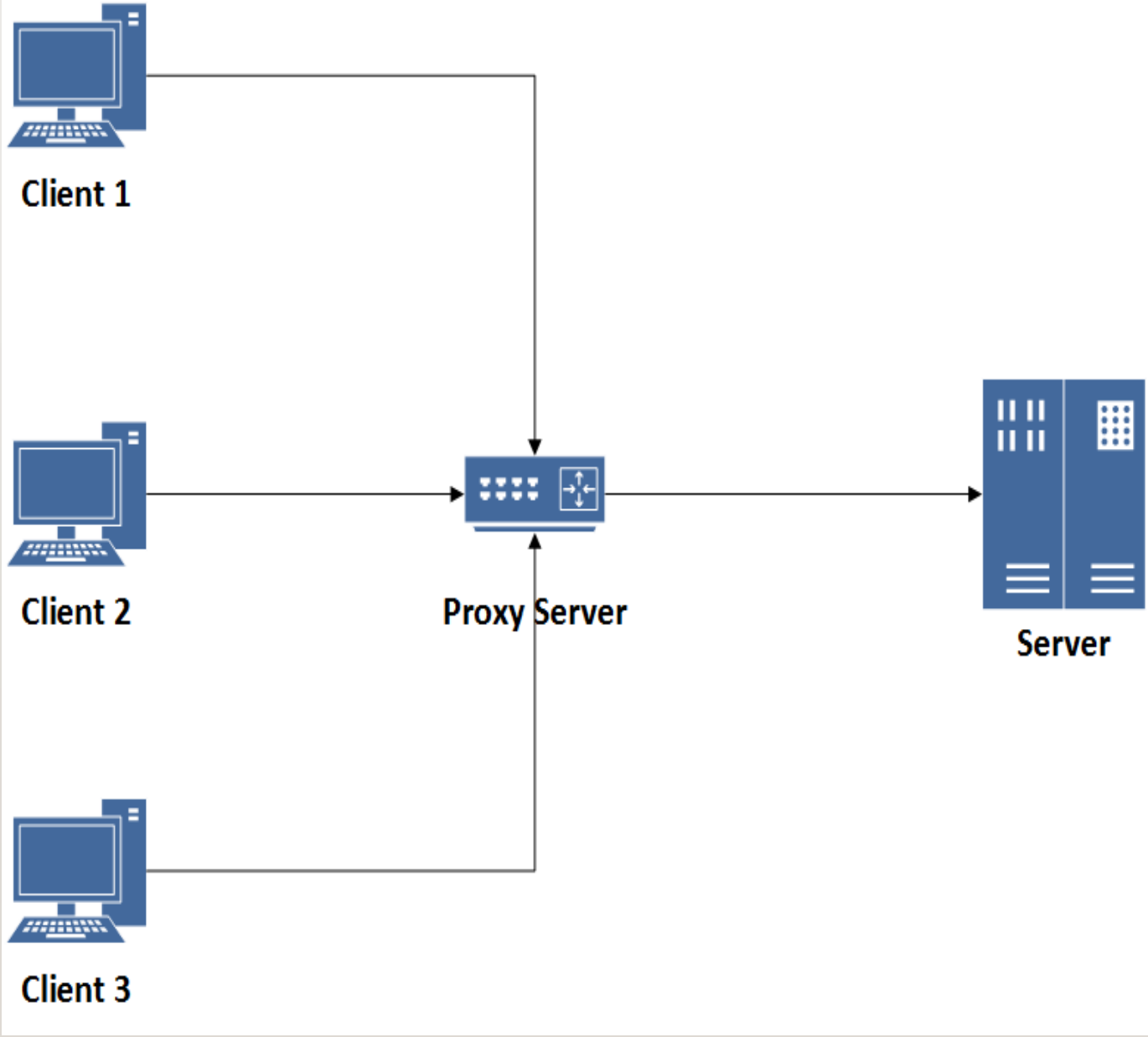


- Proxy server

- In computer networking, a proxy server is a server application that acts as an intermediary between a client requesting a resource and the server providing that resource.

- Instead of connecting directly to a server that can fulfill a requested resource such as a file or web page, the client directs the request to the proxy server, which evaluates the request and performs the required network transactions.

- This serves as a method to simplify or control the complexity of the request, or provide additional benefits such as load balancing, privacy, or security.



-Mobile Code and Mobile Agents

Mobile Code

➢The previous models assume that the client/server/peer entities exchange data. The mobile code model assumes that components may exchange code. code sent from one computer to another and run at the destination

➢An example of this is Java Applets. When your browser downloads and applet, it downloads some Java code that it then runs locally. The big issue with this model is that it introduces security risks.

Mobile agents

➢Mobile agent: a running program that travels from one computer to another carrying out a task on someone's behalf.

➢ A mobile agent is a complete program, **code + data**, that can work (relatively) independently.

➢Typical tasks:

- Collect information
- Install/maintain software on computers

➢Attention: potential security risk (like mobile code)

- Mobile Devices

➢There is an increasing need to develop distributed systems that can run devices such as cell phones, cameras, and MP3 players.
➢Unlike traditional distributed computing entities, which communicate over the Internet or standard local area networks, these devices often communicate via wireless technologies such as Bluetooth or other low bandwidth and/or short range mechanisms.
➢As a result, the geographic location of the devices impacts system design. Moreover, mobile systems must take care to consider the battery constraints of the participating devices. System design for mobile ad hoc networks (MANETs), sensor networks, and delay/disruption tolerant networks (DTNs) is a very active area of research.

2. FUNDAMENTAL MODELS

- Based on the some fundamental properties , such as **characteristics, failures, and security.**
- Fundamental models take an abstract perspective in order to describe solutions to individual issues faced by most distributed systems.
- Fundamental models are addressed by three models:

A) Interaction Model – Issues dealing with the interaction of process such as **performance and timing of events.**

B) Failure Model –discusses the failure of **processes** and **communication channels.**

C) Security Model – possible threats to processes and communication channels.

A) INTERACTION MODEL

- Deals with what are the computation occurs within processes and how it is interacted.
- The processes interact by passing messages resulting in:

→ communication(information flow) and

→coordination (synchronization and ordering of activities) between processes.

- The interaction model deals with **performance and reflect the facts that communication takes place with delays**
- The following characteristics of communication channels impact the performance of the system:

→ Latency: The delay between the start of a message's transmission from one process and the beginning of its receipt by another is referred to as latency.

→ Bandwidth: The bandwidth of a computer network is the total amount of information that can be transmitted over it in a given time.

→ Jitter: Jitter is the variation in the time taken to deliver a series of messages.

- Two variants of the interaction model:

I) Synchronous distributed systems

II) Asynchronous distributed systems

I) Synchronous distributed systems: there will be a time limit or known bound time for message delivery . Synchronous DS has on:

- Process is executing in a known lower and upper bounded time.
- Transmitted messages are received within a known bounded time.
- Drift rates between local clocks have a known bound.

II) Asynchronous Distributed system: there will be no time limit or unknown bound time for message delivery . Asynchronous distributed system has :

- No bound on process execution time (nothing can be assumed about speed, load, reliability of computers).
- No bound on message transmission delays (nothing can be assumed about speed, load, reliability of interconnections)
- No bounds on drift rates between local clocks.

B) FAILURE MODEL

- Failure model defines and classifies the faults.
- Failure can occur both in processes and communication channels. The reason can be both software and hardware.
- Failure models are needed in order to build systems with predictable behavior in case of faults (systems which are fault tolerant).
- It is important to understand the kinds of failures that may occur in a system. Following are some of the type of failure that may occur

1. Omission Failure
2. Arbitrary Failure
3. Timing Failure

1) Omission failure :

- **Omission failure is occur** when a process or communication channel **fails** to perform actions that it is expected to do.

❑ **Process Omission failure**

- Fail-stop : A process halts and remains halted. Other processes can detect that the process has failed.
- Crash: A process halts and remains halted. Other processes may not be able to detect this state.

❑ **Communication Omission failure**

- Send-omission: A process completes a send, but the message is not put in its outgoing message buffer.
- Receive-omission: A message is put in a process's incoming message buffer, but that process does not receive it.

2) Arbitrary (Byzantine) failures:

- Process or channel exhibits arbitrary(unpredictable) behavior

- **Arbitrary process failure:** Arbitrarily omits(cut or delete) intended processing steps or executed unintended processing steps . Results may not come at all or may come but carry wrong values.

- **Arbitrary channel failures :** messages may be corrupted, duplicated, delivered out of order, non-existent messages may be delivered.

3) Timing failure:

- Timing faults can occur in synchronous distributed systems, where time limits are set to process execution, communications, and clock drifts. A timing fault results in any of these time limits being exceeded .

- Clock failure : the process’s local clock exceeds the bound on its drift rate from real time
- Process- time failure : Process’s response lies outside specified time interval
- Channel-time failure : message transmission time exceed the budgeted bound.

C) SECURITY MODEL

- Security model gives the precise specification of the faults that can be exhibited by processes and communication channel.
- There are several potential threats a system designer need be aware of:

- Threats to processes - An attacker may send a request or response using a false identity (spoofing).
- Threats to communication channels - An attacker may eavesdrop (listen to messages) or inject new messages into a communication channel. An attacker can also save messages and replay them later.
- Denial of service - An attacker may overload a server by making excessive requests.

- Security threats can be defeated by

- **Cryptography and**
- **Authentication**

Communication entities can use a shared secret (key) to ensure that they are communicating with one another and to encrypt their messages so that they cannot be read by attackers.

3.11 RESOURCE SHARING AND THE WEB CHALLENGES

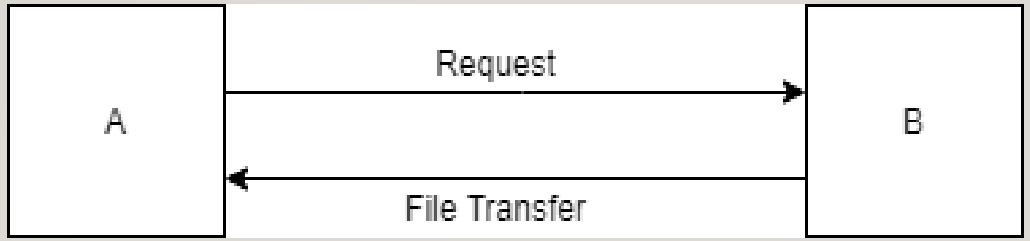
Resource Sharing

- A distributed system make it easier for users to **access** remote resources and to share resources with others users. Examples: Printers ,files, web pages etc.
- A distributed system should also make it easier for users to exchange information.
- The resources of distributed system are made available to users in following ways:

1. Data migration
2. Computation migration

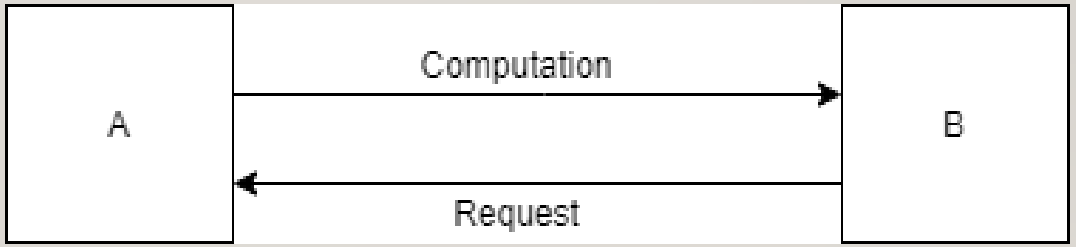
1) Data migration-

- Data is brought to the location of computation that needs access to it by distributed system.
- Simply a data is requested from source A to destination B and a data is migrated from destination B to source A as per request.



2) Computation migration

- In computation migration computation is transfer rather than the data across the system.
- Sometime data may be of large size which consume more time to transfer , in that case only computation is migrate from source A to destination B .
- Destination B perform an operation on that computation provide a result to source A by destination B as shown in figure.



- Helps in computation speed , Load balancing

Web Challenges

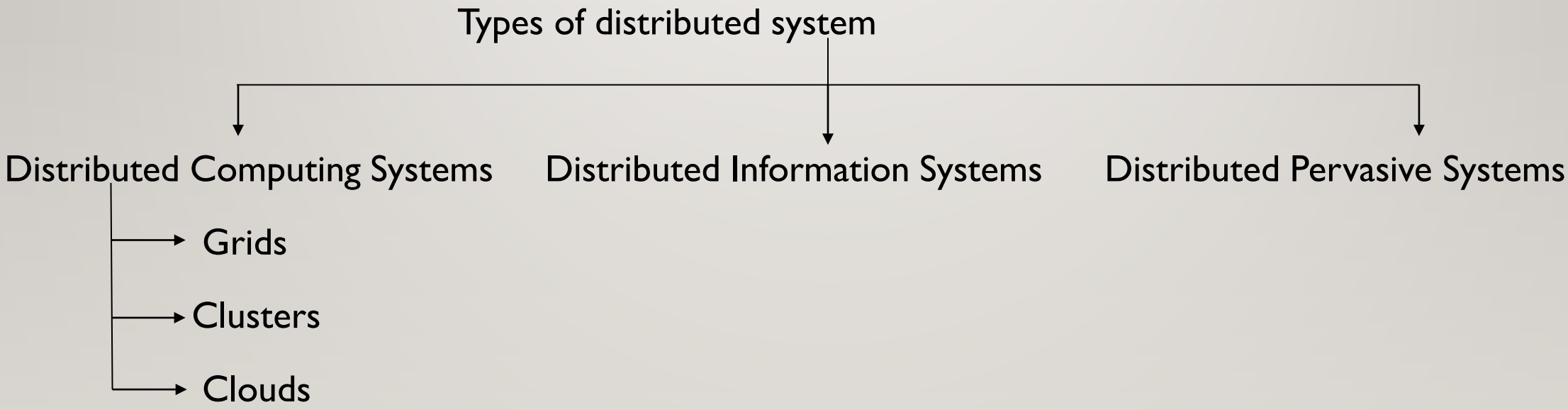
Some of the web challenges for designing the distributed system are:

1. Scalability
2. Heterogeneity
3. Security
4. Handling of failure

3.12 TYPES OF DISTRIBUTED SYSTEM: GRID , CLUSTER , CLOUD

1. Distributed Computing Systems
    a. Grids
    b. Clusters
    c. Clouds
2. Distributed Information Systems
3. Distributed Pervasive Systems

TYPES OF DISTRIBUTED SYSTEM



TYPES OF DISTRIBUTED SYSTEM

There are three kinds of distributed computing systems with the following goals:

1. Distributed Information Systems: distribute information across different servers via multiple communication models . Transaction processing system

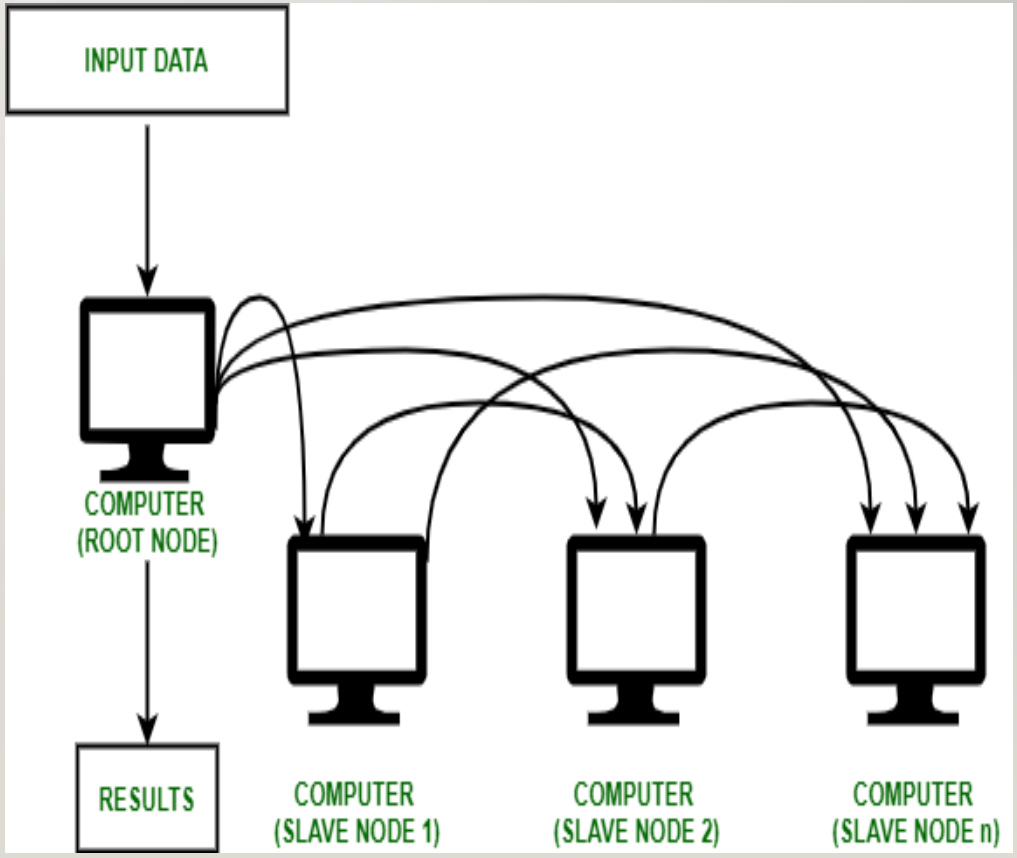
2. Distributed Pervasive Systems: use embedded computer devices (i.e. ECG monitors, sensors, mobile devices) . Home systems, sensor networks , Health care systems.

3. Distributed Computing Systems: computers in a network communicate via message passing . The distributed system uses group of computer that share a common computation problem among them so as to generate an efficient results in short time span.

A. CLUSTER COMPUTING

Cluster → group of something that are similar so the system is homogeneous.

- Cluster Computing is a type of computing in which **a group of computers are linked together through a fast Local Area Network (LAN)** so that they can **act as a single entity**.
- A collection of similar processor(PCs, workstations) running the same operating system , connected by a high speed LAN.
- It is the technique in which group of computers worked together in order to take advantage of the parallel processing power of those computers.
- Clusters are typically used for High Availability (HA) for greater reliability or High Performance Computing to provide greater computational power than a single computer can provide.
- Figure shows how cluster works



WHY CLUSTER COMPUTING IS IMPORTANT?

- Performing a complex task where more resources is needed.
- Fault Tolerance(Computational power is always available)
- Enhanced performance than that of mainframe computer
- Load Balancing
- Many organizations and IT companies are implementing cluster computing to augment their scalability, availability, processing speed and resource management at economic prices.

TYPES OF COMPUTER CLUSTERS

1. High Performance (HP) Clusters
2. Load-balancing clusters
3. High Availability (HA) or Failover Clusters

1. High performance (HP) clusters

HP clusters use computer clusters and supercomputers to solve complex and highly advanced computational problems.

- They are used to performing functions that need nodes to communicate as they perform their jobs.
- They are designed to take benefit of the parallel processing power of several nodes.
- Scientific , military , engineering apps e. g –weather modelling

2. Load-balancing clusters :

- Here workload is equally distributed across multiple computers or a computer clusters. This prevents any single node from receiving a disproportionate amount of task.
- This type of distribution is generally used in a web-hosting environment.

3. High Availability (HA) or Failover Clusters :

- A group of clusters which ensure to maintain very high availability.
- HA clusters are designed to maintain redundant nodes that can act as backup systems in case any failure occurs..

ADVANTAGES OF CLUSTER COMPUTING

Cluster computing offers a wide array of benefits. Some of these include the following

1. **Cost-Effectiveness**

Compared to highly stable and more storage mainframe computers these form of cluster computing systems are considered to be much more cost-effective and cheaper.

2. High Performance :

The systems offer better and enhanced performance than that of mainframe computer networks

3. Availability :

The other nodes will be active when one node gets failed and will function as a proxy for the failed node. This makes sure for enhanced availability.

4. Processing speed

The cluster computing systems offer the same processing speed as that of mainframe computers and the speed is also equal to supercomputers.

5. Scalable

Resources can be added to the clusters accordingly.

6. Expandability

Computer clusters can be expanded easily by adding additional computers to the network. Cluster computing is capable of combining several additional resources or the networks to the existing computer system.

DISADVANTAGES OF CLUSTER COMPUTING

1. Problem in finding fault :

It is difficult to find which component has a fault.

2. More space is needed :

Infrastructure may increase as more servers are needed to manage and monitor

3. All system should be homogeneous.

APPLICATIONS OF CLUSTER COMPUTING

Some of the application of cluster computing are :

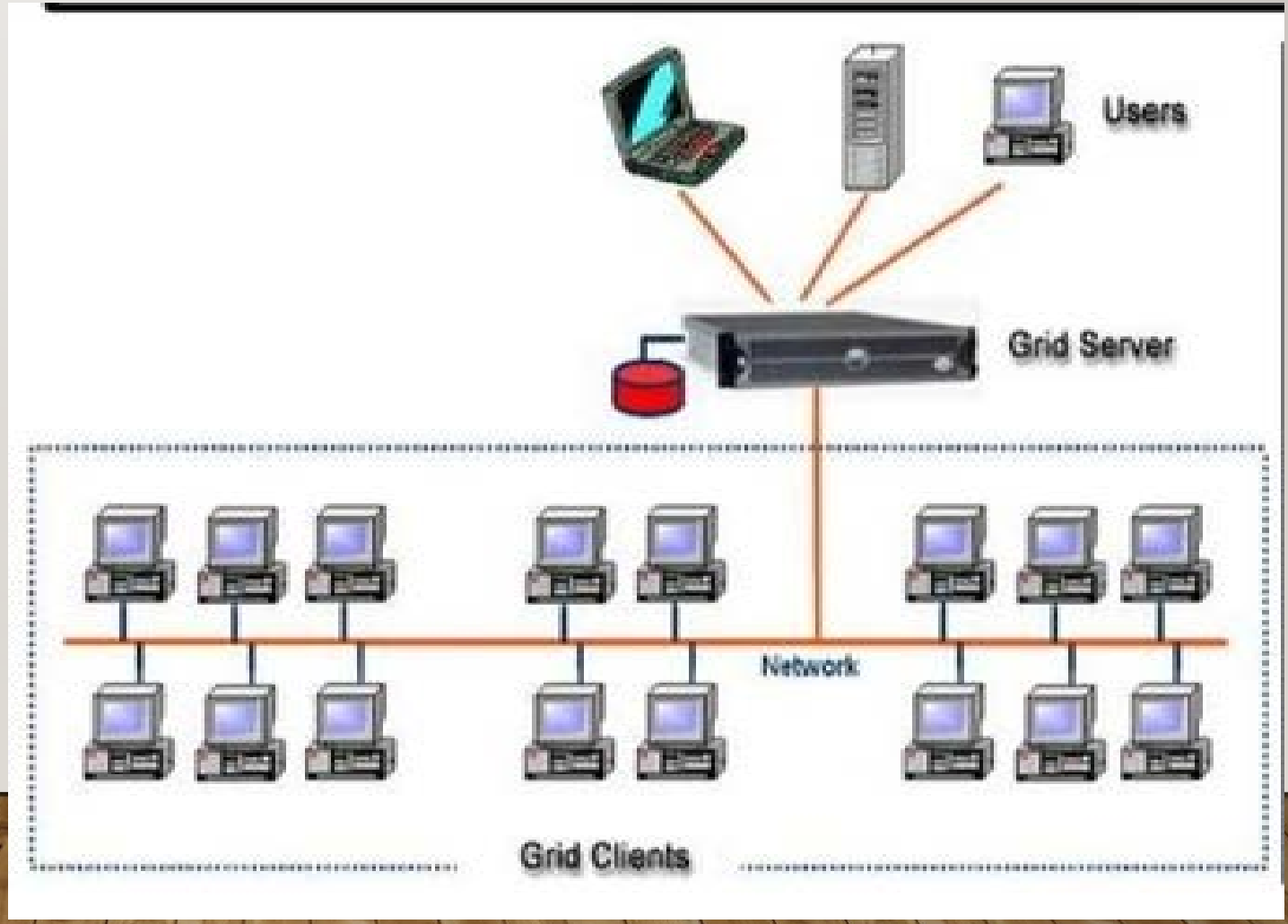
- Weather forecasting.
- Image Rendering.
- Various e-commerce applications(Amazon, eBay).
- Earthquake Simulation.
- Database Applications(Oracle on cluster)
- Assist to solve complex computational problems.

B. GRID COMPUTING

- It is a collection of computer resources from multiple locations to reach a common goal.
- Actually Grid computing is a network in which there are many computers and they can share their resources with each other using this network.
- They can share process power, memory , data storage and many resources over the internet.
- Highly heterogeneous with respect to hardware , software , network , security etc.
- The main objective of Grid Computing is to create a network which is very comfort to share their resources with every other computer , as a result all computer can work on a task together as a unit.
- Another objective is , by using grid computing a large numbers of file can be distributed evenly to all the systems that is connected to this network.

HOW GRID COMPUTING WORKS

- Grid computing system requires User , Grid server , Grid Clients.
- One of the user give a task to grid server, grid server divide the given task to the grid clients , grid client perform the task independently and return the result to the grid server . Grid server merge that result and return to the user.
- Middleware is a software that binds all the resources together.
- Every computer can share their resource by using some middleware and internet . An central computer and a server control whole the system and their sharing process.
- In this way no computer remain idle because they share their tasks with all other computers.
- On the other hand they works together like a supercomputer. As a result it is easy to complete a task.



ADVANTAGES OF GRID COMPUTING

1. Computer resources can be share and utilize in a more efficient way.
2. Multiple heterogeneous machines i.e. machines with different Operating Systems can use a single grid computing network.
3. Tasks can be performed parallelly across various physical locations .
4. It can solve larger and complex problems in a short time.
5. No computer be idle in this system.
6. Unused computing power of many computers are used to perform complex problems using grid computing.

DISADVANTAGES OF GRID COMPUTING

- Large resource distribution is more complex.
- Need for interoperability when different groups want to share resource.
- Non-interactive job submission.
- Technical challenges how to distribute resources and fix problem , allocate jobs to the grid , and coordinate between different middleware used by grid clients.

APPLICATION OF GRID COMPUTING

- Telecommunication Organizations
- Government Offices
- Multinational Companies
- Financial Organizations
- Industrial research
- Engineering research

C. CLOUD COMPUTING

- Cloud computing is the use of remote server on the internet to store, manage and process data rather than local server.
- Simply , cloud computing is the delivery or distribution of computing services/resources including servers, data storage, databases, networking and software over the Internet (“the cloud”) to offer faster innovation, flexible resources, and economies of scale.
- Cloud users can store files, data, and applications on remote servers and access this data with the help of the Internet. This helps in a way that the user doesn’t have to be in a precise location to acquire access to the data and can retrieve the data from anywhere

HOW DOES CLOUD COMPUTING WORK?

- Rather than owning their own computing infrastructure or data centers, companies can rent access to anything from applications to storage from a cloud service provider.
- One benefit of using cloud computing services is that firms can avoid the upfront cost and complexity of owning and maintaining their own IT infrastructure, and instead simply pay for what they use, when they use it.

CLOUD COMPUTING DEPLOYMENTS

- A single type of cloud computing may not be accurate for various kinds of businesses. That is why there are many different clouds out there, each offering different kinds of deployments and services to help provide us with a precise solution for our needs.

1. Public cloud - Public clouds are owned and maintained by third-party cloud service provider, which deliver their computing resources like servers, databases, and storage over the Internet. As the name suggests , public cloud is open to public. We just use these services and control our account using a web browser provided by them.

2.Private cloud - The private cloud , in contrast to its public counterparts , is not available to the public but is built specifically for a single organization to fits its needs.

3. Hybrid Cloud : A hybrid cloud uses elements from both the public cloud and private cloud and combines them, allowing data and applications to move between private and public clouds . This model gives business greater flexibility and more deployment options.

CLOUD COMPUTING SERVICES

Cloud computing is a system primarily comprised of three services: software-as-a-service (SaaS), infrastructure-as-a-service (IaaS), and platform-as-a-service (PaaS).

1. Software-as-a-service (SaaS)

Software as a service is a method for delivering software applications over the Internet, on demand and typically on a subscription basis. With SaaS, cloud providers host and manage the software application and underlying infrastructure, and handle any maintenance, like software upgrades and security patching. Users connect to the application over the Internet, usually with a web browser on their phone, tablet, or PC.

2. Infrastructure-as-a-service (IaaS)

The most basic category of cloud computing services. With IaaS, you rent IT infrastructure—servers and virtual machines (VMs), storage, networks, operating systems—from a cloud provider on a pay-as-you-go basis

3. Platform-as-a-service (PaaS)

Platform as a service refers to cloud computing services that supply an on-demand environment for developing, testing, delivering, and managing software applications. PaaS is designed to make it easier for developers to quickly create web or mobile apps, without worrying about setting up or managing the underlying infrastructure of servers, storage, network, and databases needed for development

SIMPLE EXAMPLES OF CLOUD COMPUTING

1. **Email** : web-based email services like Gmail and Hotmail deliver a cloud computing service: users can access their email ”in the cloud” from any computer with a browser and internet connection , regardless of what kind of hardware is on that particular computer. The emails are hosted on Google’s and Microsoft’s server, rather than being stored locally on the client computer.
2. **Office Productivity Software:** Office 365 , Google docs this software allow us to keep and edit documents online . The documents will be accessible anywhere and we can share the documents and collaborate on them. Multiple people can work in the same documents simultaneously.
3. **Storage :** One Drive , Google Drive , iCloud and Drop Box.

ADVANTAGES OF CLOUD COMPUTING

- **Cost -** Cloud computing eliminates the capital expense of buying hardware and software and setting up and running on-site datacenters—the racks of servers, the round-the-clock electricity for power and cooling, and the IT experts for managing the infrastructure. It simply lower the computer costs.
- Scalability - Organizations can scale up as requirements rise and scale down as requirements fall. This reduces the investment cost in infrastructure.
- Cloud services provide an opportunity to use modified applications to numerous customers at a time
- Improved performance
- Easy to manage and upgraded
- Unlimited storage capacity
- Instant software updates
- Flexibility and reliable
- Location independent

DISADVANTAGES OF CLOUD COMPUTING

- Requires good internet speed with good bandwidth and does not work well with low-speed connections.
- Security- Stored data might not be secure and can be lost
- Featured might be limited : For example , you can do a lot more with Microsoft PowerPoint than with Google Presentation's web-based offering.

APPLICATION OF CLOUD COMPUTING

1.Art Application- Cloud computing offers various art applications for quickly and easily design attractive cards, booklets, and images.

2. Business Application –Today, every organization requires the cloud business application to grow their business. It also ensures that business applications are 24*7 available to users. PayPal , Slack

3. Education Applications- Cloud computing in the education sector becomes very popular. It offers various online distance learning platforms and student information portals to the students. Google Apps for Education ,Chromebooks for Education

4. Entertainment Applications - Online games , Video Conferencing Apps

5 . Social Applications - Social cloud applications allow a large number of users to connect with each other using social networking applications such as Facebook, Twitter, Linkedln etc

DIFFERENCES

Parameter	Cluster Computing	Grid Computing	Cloud Computing
Definition	It is a form of computing in which a group of computers are linked together so that they can act like a single entity.	It is collection of computer resources from multiple locations to reach a common goal.	It is a type of computing that relies on sharing computing resources rather than having local server or personal devices to handle applications.
Homo/Heterogeneity	Nodes must be homogeneous i.e. they should have same type of hardware and operating system.	Nodes may have different Operating systems and hardwares . Machines can be homogeneous or heterogeneous.	Heterogeneous
Size	Small to Medium	Large	Small to Large

Parameter	Cluster	Grid	Cloud
Security	It offers very high level of security.	Grid offers high level of security but does not reach the level of cluster computing	Cloud offers security level lower than cluster and grid.
Resource Management	It is centralized (locally managed)	It is distributed.	Cloud is both centralized and distributed.
Location of computers	All computers are physically in the same location.	Computers may be located at a huge distance from one another or distributed geographically all over the globe.	Location does not matter.

Parameter	Cluster	Grid	Cloud
purpose	Aggregation of resources	Segregation of resources	Consolidation of Resources
Processes	Same processes run on all computers over the cluster at the same time.	Task is divided into sub task, each is assigned to an idle CPU so they all run concurrently.	Depends on service provisioning and the requesting clients.
Management	Easy	Difficult	Difficult
Implementation	Easy	Difficult	Difficult – need to be done by the host.
Internet	No internet access is required	Required	Required