

## UNIT I

### 1. What is meant distributed system?

1. We define a distributed system as a collection of autonomous computers linked by a network, with software designed to produce an integrated computing facility.

2. A system in which hardware or software components located at networked computers communicate and coordinate their actions only by message passing.

3. A collection of two or more independent computers which coordinate their processing through the exchange of synchronous or asynchronous message passing.

4. A collection of independent computers that appear to the users of the system as a single computers.

### 2. What are the significance of distributed system?

a. Concurrency of computers.

b. No global clock.

c. Independent failures.

### 3. Why we do you need distributed system?

a. **Functional distribution:** Computers have different functional capabilities (i.e., sharing of resources with specific functionalities).

b. **Load distribution/balancing:** Assign tasks to processors such that the overall system performance is optimized.

c. **Replication of processing power:** Independent processors working on the same task.

d. Distributed system consisting of collections of microcomputers may have processing powers that no supercomputer will ever achieve.

e. **Physical separation:** Systems that rely on the fact that computers are physically separated (e.g., to satisfy reliability requirements).

f. **Economics:** Collections of microprocessors offer a better price/performance ratio than large mainframes. mainframes: 10 times faster, 1000 times as expensive.

### 4. Examples of distributed system?

a. Internet

b. Intranet

c. Mobile and ubiquitous computing.

### 5. What is meant by location aware computing?

Mobile computing is the performance of computing tasks while the users are on the move and away from their residence intranet but still provided with access to resources via the devices they carry with them. They can continue to access the intranet, they can continue to access resources in their home intranet, and there is increasing provision for users to utilize resources such as printers that are conveniently nearby as they move around. This is known as location aware computing.

### 6. What are the two type of resource sharing?

a. **Hardware sharing:** Printers, plotters and large disks and other peripherals are shared to reduce costs.

b. **Data sharing is important in many applications:**

1. Software developers working in a team need to access each other's code and share the same development tools.

2. Many commercial applications enable users to access shared data objects in a single active database.

3. The rapidly growing area of group-ware tools enables users to cooperate with in a network.

**7. List the importance of data sharing?**

- Software developers working in a team need to access each other's code and share the same development tools.
- Many commercial applications enable users to access shared data objects in a single active database.
- The rapidly growing area of group-ware tools enables users to cooperate with in a network.

**8. Write the technological components of web?**

- HTML
- HTTP-request-reply protocol
- URL's

**9. List the distributed systems challenges?**

- a. Heterogeneity:** standards and protocols; middleware; virtual machine;
- b. Openness:** publication of services; notification of interfaces;
- c. Security:** firewalls; encryption;
- d. Scalability:** replication; caching; multiple servers;
- e. Failure Handling,** failure tolerance; recover/roll-back; redundancy;
- f. Concurrency,** concurrency control to ensure data consistency.
- g. Transparency.** Middleware; location transparent naming; anonymity

**10. What are the three components of security?**

Security for information resources has three components:

- **Confidentiality:** production against disclosure to unauthorized individuals.
- **Integrity:** production against or corruption.
- **Availability:** production against interference with the means to access the resources.

**11. What is the use of firewall?**

A firewall can be used to form a barrier around an intranet to protect it from outside users but does not deal with ensuring the appropriate use of resources by users within the intranet.

**12. What are the security challenges? List them.**

- a. **Denial of service attacks:** Another security problem is that the user may wish to disrupt a service for some reason. This can be achieved by bombarding the service with such a large number of pointless requests that the serious users are unable to use it. This is called a denial of service attack and there are many on well known web services.
- b. **Security of mobile code:** Mobile codes needed to be handled with care. PC users sometimes send executable files as email attachments to be run by the recipient, but a recipient will not be able to run it.

**13. List the challenges to be considered for designing scalable distributed system?**

- Controlling the cost of physical resources
- Controlling the performance loss

- Preventing software resources running out
- Avoiding performance bottlenecks.

**14. What are the types of transparencies?**

- Access transparency:** enables local and remote resources to be accessed using identical operations.
- Location transparency:** enables resources to be accessed without knowledge of their location.
- Concurrency transparency:** enables several processes to operate concurrently using shared resources without interference between them.
- Replication transparency:** enables multiple instances of resources to be used to increase reliability and performance without knowledge of the replicas by users or application programmers.
- Failure transparency:** enables the concealment of faults, allowing users and application programs to complete their tasks despite the failure of hardware or software components.
- Mobility transparency:** allows the movement of resources and clients within a system without affecting the operation of users or programs.
- Performance transparency:** allows the system to be reconfigures to implement performance as loads vary.
- Scaling transparency:** allows the system and applications to expand in scale without change to the system structure or the application algorithms.
- Access and location transparency together provide network transparency.

**15. What are the failures detected in DS?**

**Masking failures:** Some detected failures can be hidden or made less severe. Examples of hiding failures:

1. Messages can be retransmitted when they fail to arrive
2. File data can be written to a pair of disks that if one is corrupted, the other may still be correct.

**Tolerating failures:** Most of the services in the Internet do exhibit failures. It would not be practical for them to detect and hide all the failures occur in such network. Their clients are designed to tolerate failures, which generally involve the users in that.

**Recovery from failures:** involves the design of software so that the state permanent data can be rolled back after a server has crashed.

**16. List the key design goals of DS?**

- a. Performance
- b. Reliability
- c. Scalability
- d. Consistency
- e. Security

**17. List the technical design goals of DS?**

- a. Naming
- b. Communication
- c. Software structure
- d. Workload allocation

e. Maintenance of consistency.

**18 .What is the use of multicast?**

- a. **Locating an object:** A process multicasts a message containing a name of a resource to a group of server processes. Only the process that holds the resource responds to the message.
- b. **Fault tolerance:** A process multicasts its request to a group of identical server processes. The group of servers can continue to provide their service even if one of its members fails.
- c. **Multiple update:** Used for example in video conferencing with multiple participants.

**19. Write the models used in workload allocation?**

The following four models for workload allocation are presented here.

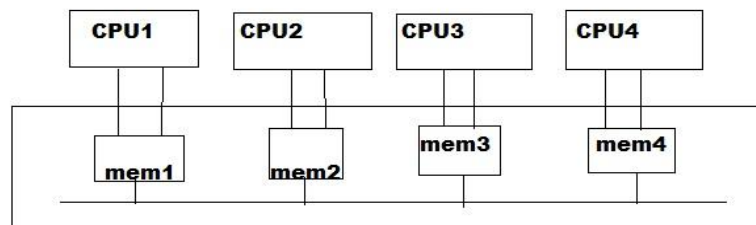
- Processor pool model
- Shared memory multiprocessors
- Parallel virtual machines
- Distributed shared memory.

**20.What is meant by PVM?**

PVM is an integrated set of software tools and libraries that emulates a general-purpose, flexible heterogeneous concurrent computing framework on interconnected computers of varied architecture. The overall objective of the PVM system is to enable a collection of computers to be used cooperatively for concurrent or parallel computation.

**21.What is meant by DSM?**

**Distributed shared memory (DSM):** DSM provides a global shared address space across the different machines on a cluster. The shared address space distinguishes it from packages such as PVM that provide a message passing interface between machines. There is a growing consensus in the parallel computing community that a shared memory interface is more desirable from the application programmer's viewpoint, allowing him to focus on algorithmic development rather than on managing communication.



**22. List the types of consistencies in DS?**

- Update consistency
- Replication consistency
- Cache consistency
- Failure consistency
- Clock consistency

**23. List the user requirements used in design of DS?**

- **Functionality:** what should the system do for the users.
- **Quality of service:** issue containing performance, reliability and security.
- **Reconfigurability:** the need to accommodate changes without causing disruption of the actual service.

**24. List the main types of architectural model?**

**a. Software architecture.**

**b. System architecture.**

- Client server model
- Services provided by multiple servers.
- Proxy servers and cache.
- Peer processes.

**25. Enumerate the factors to be considered for variations in client server model?**

The factors considered for several variations on the client server model.

- The use of multiple servers and caches to increase performance and flexibility.
- The use of mobile code and mobile agents.
- User's need for low cost components with limited hardware resources that are simple to manage.
- The requirements to add remove mobile devices in a convenient manner.

**26. What is meant by thin clients?**

- This refers to software layer that supports a window based user interface on a computer that is local to the user while executing application programs on a remote computer.
- It has the low management and hardware costs, but it runs the application code of user's computer in the computer server, which is the powerful computer that has the capacity to run large numbers of applications simultaneously. It can be multiprocessor or cluster computer running a multiprocessor version of OS such as UNIX or windows NT.

**27. What is meant by x-11 window system?**

- a) The X-11 window system is a process that manages the display and interactive input devices (keyboard, mouse) of the computer on which it runs. It provides an extensive library of procedures (the X-11 protocol) for displaying and modifying graphical objects in windows as well as the creation and manipulation of windows.
- b) The X-11 system is referred to as a window server process. The clients of the X-11 server are the application programs that the user is currently interacting with.

- c) The client programs communicate with the server by invoking operation in the X-11 protocol, these include operations to draw text and graphical objects in windows.

**28.Enumerate the key features of spontaneous networking?**

- a. **Easy connection to a local network:** Wireless links avoid the need for pre-installed cabling and avoid the inconvenience and reliability issues surrounding plugs and sockets.
- b. **Easy integration with local services:** Devices are able to find themselves inserted into existing networks of devices discover automatically what services are provided there, with no special configuration actions by the user.

**29.List the design issues to be considered for spontaneous networking?**

- a. **Limited connectivity:** Users are not always connected as they move around. They are irregularly disconnected from wireless network as they travel through tunnels by train. They may also be totally disconnected for longer periods of time in regions where wireless connectivity ceases or it is too expensive to remain connected.
- b. **Security and privacy:** Many vulnerable security issues arises due to the attempt of wireless connections in unsupervised way. Some systems track the physical locations of users as they move around and this may threaten the user's privacy. This facility enables users to access their home intranet while on the move may expose data that is supposed to remain behind the intranet firewall or it may open up the intranet to attacks from outside.
- c. **Discovery services:** Spontaneous networking requires client processes running on portable devices and other appliances to access services on the networks to which they are connected. Here the clients discover what services are available in the network to which they are connected and to investigate their properties. The purpose of a discovery service is to accept and store details of services that are available on the network and to respond to queries from clients about them.

**30.What is the purpose solved by fundamental model?**

- In general, such a fundamental model should contain only the essential ingredients that we need to consider understanding and reasoning about some aspects of a system's behaviour. The purpose of such a model is:
- To make explicit all the relevant assumptions about the system we are modelling.
- To make generalization concerning what is possible or impossible, given those assumptions. The guarantees are our assumptions clear and explicit, we can hope to prove system properties using mathematical techniques. These properties will then hold for any system meeting our assumptions.

**31.How the fundamental models are categorized?**

- a. Interaction
- b. Failure
- c. Security

**32.List out the characteristics of performance of DS?**

The following performance characteristics relating to latency, bandwidth and jitter.



- a. **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.
- b. **Bandwidth:** The bandwidth of a computer network is the total amount of information that can be transmitted over it in a given time. When large number of communication channels are using the same network, they have to share the available bandwidth.
- c. **Jitter.** Jitter is the variation in the time taken to deliver a series of messages. Jitter is relevant to multimedia data. For example, if consecutive samples of audio data are played with differing time intervals, the sound will be badly distorted.

### 33. What is synchronous DS?

- 1) The time to execute each step of a process has known lower and upper bounds.
- 2) Each message transmitted over a channel is received within a known bounded time.
- 3) Each process has a local clock whose drift rate from real time has a known bound.
- 4) It is possible to suggest likely upper and lower bounds for process execution time, message delay and clock drift rates in a distributed system, but it is difficult to arrive at realistic values and to provide guarantees of the chosen values.
- 5) In a synchronous system it is possible to use timeouts, for example to detect the failure of a process.

### 34. What is asynchronous DS?

1. Many distributed systems, such as the Intranet, qualify as asynchronous system.
2. An asynchronous distributed system is one in which there are no bounds on:
  1. **Process execution speeds**-for example, one process step may take only a picoseconds and another a century; all that can be said is that each step may take an arbitrarily long time.
  2. **Message transmission delays**-for example, one message from process A to process B may be delivered in negligible time and another may take several years. In other words, a message may be received after an arbitrarily long time.
  3. **Clock drift rates**- again, the drift rate of a clock is arbitrary.

### 35. What is omission failure?

The faults classified as omission failures refer to cases when a process or communication channel fails to perform actions that it is supposed to do.

### 36. What is meant by arbitrary failure?

- 1) The term arbitrary or Byzantine failure is used to describe the worst possible failure semantics, in which any type of error may occur. For example, a process may set wrong values in its data items, or it may return a wrong value in response to an invocation.
- 2) An arbitrary failure of a process is one in which it arbitrarily omits intended processing steps to take unintended processing steps. Arbitrary failures in processes cannot be detected by seeing whether the process responds to invocations, because it might arbitrarily omit to reply.

### 37. List out the characteristics of networks hidden by stream abstraction?

- a) **Message sizes:** The application can choose how much data it writes to a stream or reads from it. It may deal in very small or very large sets of data. The underlying implementation of a TCP stream decides how much data to collect before transmitting it as one or more IP packets.
- b) **Lost messages:** The TCP protocol uses an acknowledgement scheme. As an example of a simple scheme (which is not used in TCP), the sending end keeps a record of each IP packet sent and receiving end acknowledges all the arrivals. If the sender does not receive an acknowledgement within a timeout, it retransmits the message.
- c) **Flow control:** The TCP protocol attempts to match the speeds of the processes that read from and write to a stream. If the writer is too fast the reader, then it is blocked until the reader has consumed sufficient data.
- d) **Message duplication and ordering:** Message identifiers are associated with each IP packet, which enables the recipient to detect and reject duplicates, or to reorder message that do not arrive in sender order.
- e) **Message destinations:** A pair of communication processes establishes a connection before they can communicate over a stream. Once a connection is established, the processes simply read from and write to the stream without needing to use Internet addresses and ports. Establishing a connection involves a connect request from client to server followed by an accept request from server to client before any communication can take place.

**38. List the issues related to stream communication?**

- a. **Matching of data items:** Two communicating processes need to agree as to the contents of the data transmitted over a stream. For example, if one process writes an int followed by a double to a stream, then the reader at the other end must read an int followed by a double. When a pair of processes does not cooperate correctly in their use of a stream, the reading process may experience errors when interpreting the data or may block due to insufficient data in the stream.
- b. **Blocking:** The data written to a stream is kept in a queue at the destination socket. When a process attempts to read data from an input channel, it will get data from the queue or it will block until data becomes available. The process that writes data to a stream may be blocked by the TCP flow-control mechanism if the socket at the other end is queuing as much data as the protocol allows.
- c. **Threads:** When a server accepts a connection, it generally creates a new thread in which to communicate with the new client. The advantage of using a separate thread for each client is that the sever can block when waiting for input without delaying other clients.
- d. **Failure model:** To satisfy the integrity to property of reliable communication, TCP streams use checksums to detect and reject corrupt packets and sequence numbers to detect and reject duplicate packets. For the deal with lost packets. Therefore, messages are guaranteed to be delivered ever when some of the underlying packets are lost.

**39. What is marshalling and unmarshalling?**

- An agreed standard for the representation of data structures and primitive values is called an external and data representation



- **Marshalling** is the process of taking a collection of data items and assembling them into a form suitable for transmission in a message.
- **Unmarshalling** is the process of disassembling them on arrival to produce an equivalent collection of data items at the destination. Thus marshalling consists of the translation of structured data items and primitive values into an external data representation. Similarly, unmarshalling consists of the generation of primitive values from their external data representation and the rebuilding of the data structures.

**40. What are the approaches used in data representation?**

- a. **CORBA's common data representation**, which is concerned with an external representation for the structured and primitive types that can be passed as the arguments and results of remote method invocations in CORBA. It can be used by a variety of programming languages.
- b. **Java's object serialization**, which is concerned with the flattening and external data representation of any single object or tree of objects that may need to be transmitted in a message or stored on a disk. It is for use only by java.
- c. **XML (Extensible Markup Language)**, which defines a textual format for representing structured data. It was originally intended for documents containing textual self-describing structured data-for example documents accessible on the Web- but it is now also used to represent the data sent in message exchanged by clients and servers in web services.

**16 MARK QUESTIONS**

**1. Explain the need of Distributed systems its characteristics with example**

- Fault tolerance
- Consistency
- Security
- Reliability
- Concurrent transactions

**2. Explain how resource sharing is done in the web**

- Patterns of resource sharing vary widely in their scope and in how closely users work together. At one extreme, a search engine on the Web provides a facility to users throughout the world, users who need never come into contact with one another directly. At the other extreme, in computer-supported cooperative working(CSCW), a group of users who cooperate directly share resources such as documents in a small, closed group. The pattern of sharing and the geographic distribution of particular users determines what mechanisms the system must supply to coordinate users' actions.

- Types of resources:
  - Hardware resources
  - Data
  - Service

**3. Explain the challenges to be considered in the design of DS**

- Heterogeneity
- Openness
- Security
- Scalability
- Failure handling
- Concurrency
- Transparency
- Quality of service

**4. What are the trends in Distributed systems.**

- Distributed systems are undergoing a period of significant change and this can be traced back to a number of influential trends:
  - the emergence of pervasive networking technology;
  - the emergence of ubiquitous computing coupled with the desire to support user mobility in distributed systems;
  - the increasing demand for multimedia services;
  - the view of distributed systems as a utility.
- Pervasive networking and the modern Internet
- Mobile and ubiquitous computing
- Distributed multimedia systems
- Distributed computing as a utility