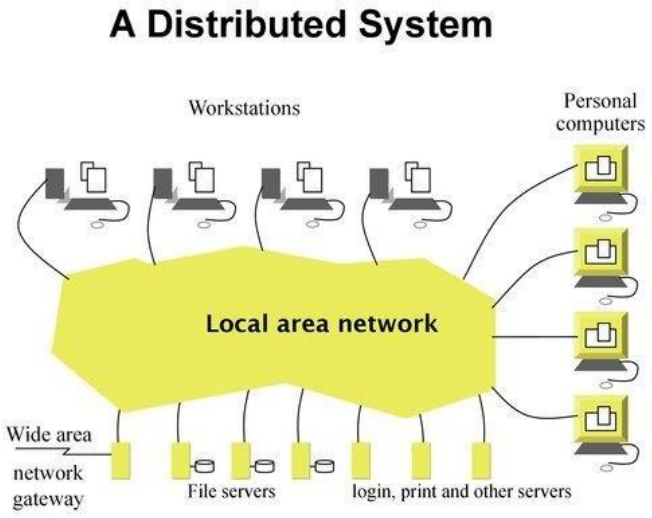
Unit: 1 Introduction to Distributed System

Define Distributed Operating System and Explain Goals of Distributed System.

- A distributed system is a collection of independent computers that appear to the users of the system as a single computer.
- Using high performance computers connected by equally high speed communication links, it is possible to build a single system consisting of multiple computer and using it as a single consolidated system.
- In such a system, multiple resources work together to deliver the required processing speed and the operating system takes care of maintaining the system and overall maintenance.
- In distributed system computers are not independent but interconnected by a high speed network.
- It means that many computers, be they workstation or desktop systems linked together, can do work of a high performance supercomputer.
- If user thinks that the entire interlinked system is a single and unified computer, there has to be software envelope that joins multiple operating system together and offers a seamless interface to user.



The following are the main goals of distributed systems:

- The relative simplicity of the software - each processor has a dedicated function.
- Incremental growth - if we need 10 percent more computing power, we just add 10 percent more processors.
- Reliability and availability - a few parts of the system can be down without disturbing people using the other parts.
- Openness: Offer services according to standard rules that describe the syntax and semantics of those services.

Explain Advantage of Distributed system over centralized system.

Economics

- A quarter century ago, according to Grosch's law: the computing power of a CPU is proportional to the square of its price.

- By paying twice as much you could get four times the performance.
- This observation fit the mainframe technology of its time.
- With microprocessor technology, Grosch's law no longer holds.
- For a few hundred dollars you can get the CPU chip that can execute more instructions per second than one of the largest 1980s mainframe.

Speed

- A collection of microprocessors cannot only give a better price/performance ratio than a single mainframe, but also give an absolute performance that no mainframe can achieve at any price.
- For example, with current technology it is possible to build a system from 10,000 modern CPU chips, each of which runs at 50 MIPS (Millions of Instructions Per Second), for a total performance of 500,000 MIPS.
- For a single processor (i.e., CPU) to achieve this, it would have to execute an instruction in 0.002 nano sec (2 picosecond).

Inherent Distribution

- Many institutions have applications which are more suitable for distributed computation.
- Assume that there is large company buying and selling goods from different countries.
- Its offices situated in those countries are geographically diverse.
- If a company wishes unified computing system, it should implement a distributed computing system.
- Consider the global employee database of such a multinational company.
- Branch offices would create local database and then link them for global viewing.

Reliability

- By distributing the workload over many machines, a single chip failure will bring down at most one machine, leaving the rest intact.
- Ideally, if 5 percent of the machines are down at any moment, the system should be able to continue to work with a 5 percent loss in performance.
- For critical applications, such as control of nuclear reactors or aircraft, using a distributed system to achieve high reliability may be the dominant consideration.

Incremental Growth

- A company will buy a mainframe with the intention of doing all its work on it.
- If the company expands and the workload grows, at a certain point the mainframe will no longer be adequate.
- The only solutions are either to replace the mainframe with a larger one (if it exists) or to add a second mainframe.
- Both of these can cause damage on the company's operations.
  In contrast, with a distributed system, it may be possible simply to add more processors to the system, thus allowing it to expand gradually as the need arises.

Explain the Advantage of Distributed system over Independent PCs.

Data sharing

- Many users need to share data.
- For example, airline reservation clerks need access to the master data base of flights and existing reservations.
- Giving each clerk his own private copy of the entire data base would not work, since nobody would know which seats the other clerks had already sold.

- Shared data are absolutely essential for this and many other applications, so the machines must be interconnected.

Resource Sharing

- Expensive peripherals, such as color laser printers, phototypesetters, and massive archival storage devices (e.g., optical jukeboxes), can also be shared

Communication

- For many people, electronic mail has numerous attractions over paper mail, telephone, and FAX.
- It is much faster than paper mail, does not require both parties to be available at the same time as does the telephone, and unlike FAX, produces documents that can be edited, rearranged, stored in the computer, and manipulated with text processing programs.

Flexibility

- A distributed system is more flexible than giving each user an isolated personal computer.
- One way is to give each person a personal computer and connect them all with a LAN, this is not the only possibility.
- Another one is to have a mixture of personal and shared computers, perhaps of different sizes, and let jobs run on the most appropriate one, rather than always on the owner's machine.
- In this way, the workload can be spread over the computers more effectively, and the loss of a few machines may be compensated for by letting people run their jobs elsewhere.

Classify the Network operating system and Distributed operating system

Network Operating system	Distributed Operating system
A network operating system is made up of software and associated protocols that allow a set of computer network to be used together.	A distributed operating system is an ordinary centralized operating system but runs on multiple independent CPUs.
Environment users are aware of multiplicity of machines.	Environment users are not aware of multiplicity of machines.
Control over file placement is done manually by the user.	It can be done automatically by the system itself.
Performance is badly affected if certain part of the hardware starts malfunctioning.	It is more reliable or fault tolerant i.e. distributed operating system performs even if certain part of the hardware starts malfunctioning.
Remote resources are accessed by either logging into the desired remote machine or transferring data from the remote machine to user's own machines.	Users access remote resources in the same manner as they access local resources.

Classify Distributed operating system.

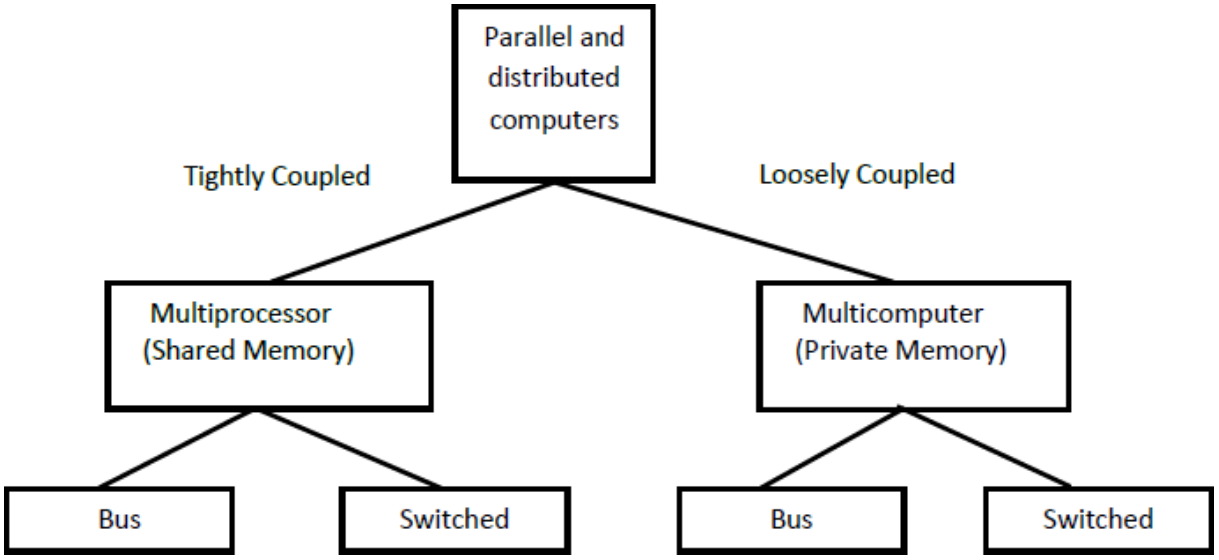


Figure – Classification of Distributed System

Hardware Concepts

- Even though all distributed systems consist of multiple CPUs, there are several different ways the hardware can be organized, especially in terms of how they are interconnected and how they communicate.
- Various classification schemes for multiple CPU computer systems have been proposed over the years.
- According to Flynn, classification can be done based on, the number of instruction streams and number of data streams.

SISD

- A computer with a single instruction stream and a single data stream is called SISD.
- All traditional uniprocessor computers (i.e., those having only one CPU) fall in this category, from personal computers to large mainframes.

SIMD

- The next category is SIMD, single instruction stream, multiple data stream.
- This type refers to array processors with one instruction unit that fetches an instruction, and then commands many data units to carry it out in parallel, each with its own data.
- These machines are useful for computations that repeat the same calculation on many sets of data, for example, adding up all the elements of 64 independent vectors.
- Some supercomputers are SIMD.

MISD

- The next category is MISD, multiple instruction stream, single data stream, no known computers fit this model.

MIMD

- Finally, comes MIMD, which essentially means a group of independent computers, each with its own program counter, program, and data.
- All distributed systems are MIMD.
  We divide all MIMD computers into two groups: those that have shared memory, usually called multiprocessors, and those that do not, sometimes called multicomputer.

Software Concepts

- The image that a system presents to its users, and how they think about the system, is largely determined by the operating system software, not the hardware.
- There are basic two types of operating system namely (1) Tightly coupled operating system and (2) loosely couple operating system; for multiprocessor and multicomputer.
- Loosely-coupled software allows machines and users of a distributed system to be fundamentally independent of one another.
- Consider a group of personal computers, each of which has its own CPU, its own memory, its own hard disk, and its own operating system, but which share some resources, such as laser printers and data bases, over a LAN.
- This system is loosely coupled, since the individual machines are clearly distinguishable, each with its own job to do.
- If the network should go down for some reason, the individual machines can still continue to run to a considerable degree, although some functionality may be lost.
- For tightly coupled system consider a multiprocessor dedicated to running a single chess program in parallel.
- Each CPU is assigned a board to evaluate, and it spends its time examining that board and all the boards that can be generated from it.
- When the evaluation is finished, the CPU reports back the results and is given a new board to work on.
- The software for this system, both the application program and the operating system required to support it, is clearly much more tightly coupled than in our previous example.

Tightly Coupled System

- Distributed Operating system is a tightly coupled software on loosely coupled hardware.
- The goal of such a system is to create the illusion in the minds of the users that the entire network of computers is a single timesharing system, rather than a collection of distinct machines.

Characteristics:

There is a single, global inter process communication mechanism so that any process can talk to any other process.

- There is a global protection scheme.
- Process management is also uniform everywhere.
- How processes are created, destroyed, started, and stopped does not vary from machine to machine.
- The file system looks same from everywhere.
- Every file is visible at every location, subject to protection and security constraints.
- Each kernel has considerable control over its own local resources.
- For example, if swapping or paging is used, the kernel on each CPU is the logical place to determine what to do swap or page.