UNIT 6

CLASSIFICATION OF PARALLEL SYSTEMS

Flynn's Taxonomy

A taxonomy first introduced by Flynn [FLYN72] is still the most common way of categorizing systems with parallel processing capability.Flynn proposed the following categories of computer systems:

• Single instruction, single data (SISD) stream: A single processor executes a single instruction stream to operate on data stored in a single memory. Uniprocessors fall into this category.

• Single instruction,multiple data (SIMD) stream: A single machine instruction controls the simultaneous execution of a number of processing elements on a lockstep basis. Each processing element has an associated data memory, so that each instruction is executed on a different set of data by the different processors. Vector and array processors fall into this category, and are discussed in Section 18.7.

• Multiple instruction, single data (MISD) stream: A sequence of data is transmitted to a set of processors,each of which executes a different instruction sequence.This structure is not commercially implemented.

• Multiple instruction,multiple data (MIMD) stream:A set of processors simultaneously execute different instruction sequences on different data sets.SMPs, clusters,and NUMA systems fit into this category.

With the MIMD organization,the processors are general purpose;each is able to process all of the instructions necessary to perform the appropriate data transformation. MIMDs can be further subdivided by the means in which the processors communicate (Figure 17.1). If the processors share a common memory, then each processor accesses programs and data stored in the shared memory,and processors

communicate with each other via that memory.The most common form of such system is known as a symmetric multiprocessor (SMP), In an SMP,multiple processors share a single memory or pool of memory by means of a shared bus or other interconnection mechanism; a distinguishing feature is that the memory access time to any region of memory is approximately the same for each processor.A more recent development is the nonuniform memory access (NUMA)organization, As the name suggests,the memory access time to different regions of memory may differ for a NUMA processor. A collection of independent uniprocessors or SMPs may be interconnected to form a **cluster**.Communication among the computers is either via fixed paths or via some network facility.

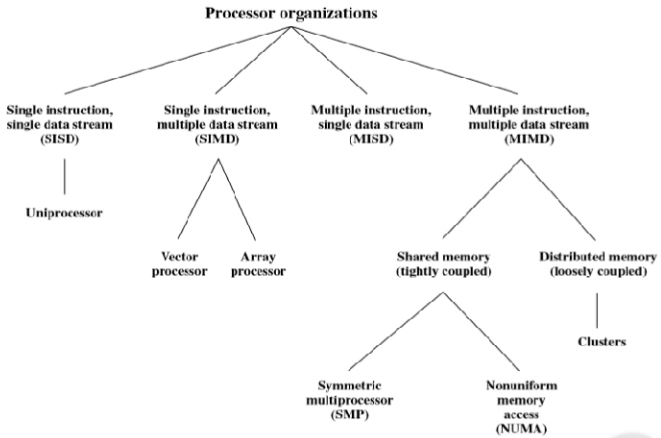


Fig: A Taxonomy of Parallel Processor Architectures

Clusters

An important and relatively recent development computer system design is clustering.Clustering is an alternative to symmetric multiprocessing as an approach to providing high performance and high availability and is particularly attractive for server applications.

We can define a cluster as a group of interconnected, whole computers working together as a unified computing resource that can create the illusion of being one machine.The term whole computer means a system that can run on its own,apart from the cluster;in the literature,each computer in a cluster is typically referred to as a node.

Four benefits that can be achieved with clustering.These can also be thought of as objectives or design requirements:

• Absolute scalability: It is possible to create large clusters that far surpass the power of even the largest standalone machines.A cluster can have tens, hundreds,or even thousands of machines,each of which is a multiprocessor.

• Incremental scalability: A cluster is configured in such a way that it is possible to add new systems to the cluster in small increments.Thus,a user can start out with a modest system and expand it as needs grow, without having to go through a major upgrade in which an existing small system is replaced with a larger system.

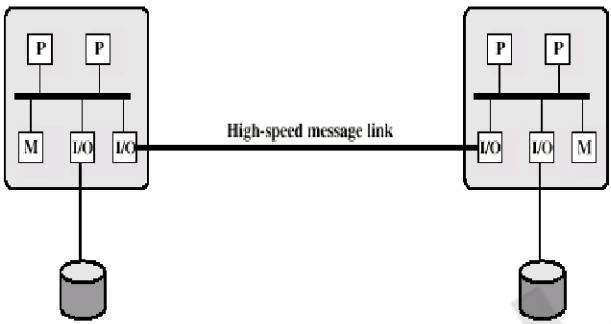
• High availability:Because each node in a cluster is a standalone computer,the failure of one node does not mean loss of service.In many products,fault tolerance is handled automatically in software.

• Superior price/performance:By using commodity building blocks,it is possible to put together a cluster with equal or greater computing power than a single large machine,at much lower cost.

Cluster Configuration

Clusters are classified in a number of different ways. Perhaps the simplest classification is based on whether the computers in a cluster share access to the same disks.

[1].Cluster Configurations - Standby Server, No Shared Disk



[2]. Cluster Configurations - Shared Disk

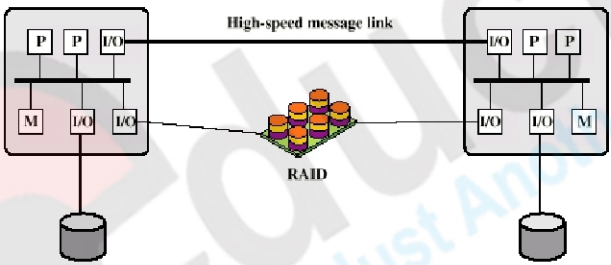


Figure [1] shows a two-node cluster in which the only interconnection is by means of a high-speed link that can be used for message exchange to coordinate cluster activity.The link can be a LAN that is shared with other computers that are not part of the cluster or the link can be a dedicated interconnection facility.

Figure[2] is a shared-disk cluster. In this case, there generally is still a message link between nodes.In addition,there is a disk subsystem that is directly linked to multiple computers within the cluster.In this figure,the common disk subsystem is a RAID system. The use of RAID or some similar redundant disk technology is common in clusters so that the high availability achieved by the presence of multiple computers is not compromised by a shared disk that is a single point of failure.

Clustering Methods:

- Passive standby
- Active secondary
  - Separate servers

- Servers connected to disks (one variation is shared nothing)
- Servers share disks

Clustering Method	Description	Benefits	Limitations
Passive Standby	A secondary server takes over in case of primary server failure	Easy to implement	High cost because the secondary server is unavailable for other processing tasks
Active Standby	A secondary server is also used for processing	Reduced cost because secondary server can be used for processing	Increased complexity
Separate Servers	Separate servers have its own disks. Data are continuously copied from primary to secondary server	High availability	High network and server overhead due to copying operations
Servers Connected to the disks	Servers are cabled to the same disks but each server owns its disks. If one server fails, its disk are taken over by the other server	Reduced network and server overhead due to elimination of copying operations	Usually requires disk mirroring or RAID technology to compensate for risk of disk failure
Servers Share disks	Multiple servers simultaneously share access to disks.	Low network and server overhead. Reduced risk of downtime caused by disk failure.	Require lock manager software. Usually used with disk mirroring or RAID technology

Three classifications of clustering can be identified:**separate servers,shared nothing,and shared disk.**

In one approach to clustering,each computer is a **separate server** with its own disks and there are no disks shared between systems .This arrangement provides high performance as well as high availability.In this case,some type of management or scheduling software is needed to assign incoming client requests to servers so that the load is balanced and high utilization is achieved.It is

desirable to have a failover capability,which means that if a computer fails while executing an application,another computer in the cluster can pick up and complete the application.For this to happen,data must constantly be copied among systems so that each system has access to the current data of the other systems.The overhead of this data exchange ensures high availability at the cost of a performance penalty.

To reduce the communications overhead, most clusters now consist of servers connected to common disks . In one variation on this approach,called **shared nothing**,the common disks are partitioned into volumes,and each volume is owned by a single computer.

If that computer fails,the cluster must be reconfigured so that some other computer has ownership of the volumes of the failed computer. It is also possible to have multiple computers share the same disks at the same time (called the **shared disk** approach),so that each computer has access to all of the volumes on all of the disks.This approach requires the use of some type of locking facility to ensure that data can only be accessed by one computer at a time

Operating System Design Issues

- **Failure management**
  - Highly available cluster offers a high probability that all resources will be in service
    - No guarantee about the state of partially executed transactions if failure occurs
  - Fault-tolerant cluster ensures that all resources are always available

A highly available cluster offers a high probability that all resources will be in service. If a failure occurs,such as a system goes down or a disk volume is lost,then the queries in progress are lost.Any lost query, if retried, will be serviced by a different computer in the cluster. However, the cluster operating system makes no guarantee about the state of partially executed transactions.This would need to be handled at the application level.

A fault-tolerant cluster ensures that all resources are always available.This is achieved by the use of redundant shared disks and mechanisms for backing out uncommitted transactions and committing completed transactions

- **Load balancing**
  - A cluster requires an effective capability for balancing the load among available computers.This includes the requirement that the cluster be incrementally scalable.When a new computer is added to the cluster,the load-balancing facility should automatically include this computer in scheduling applications.Middleware mechanisms need to recognize that services can appear on different members of the cluster and may migrate from one member to another.

- **Parallelizing Computation**

  Three general approaches to the problem:

  - Parallelizing compiler

    parallelizing compiler determines, at compile time, which parts of an application can be executed in parallel.These are then split off to be assigned to different computers in the cluster.Performance depends on the nature of the problem and how well the compiler is designed. In general,such compilers are difficult to develop.

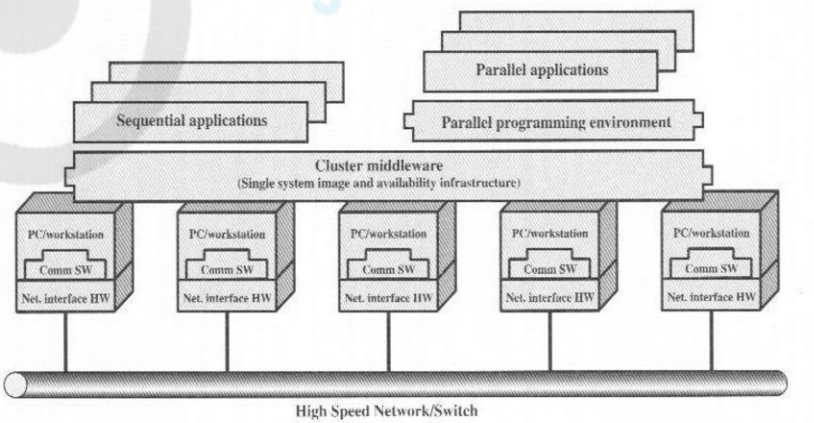
  - Parallelized application

    In this approach,the programmer writes the application from the outset to run on a cluster, and uses message passing to move data, as required, between cluster nodes. This places a high burden on the programmer but may be the best approach for exploiting clusters for some applications.

  - Parametric computing

    This approach can be used if the essence of the application is an algorithm or program that must be executed a large number of times, each time with a different set of starting conditions or parameters. A good example is a simulation model,which will run a large number of different scenarios and then develop statistical summaries of the results. For this approach to be effective, parametric processing tools are needed to organize, run,and manage the jobs in an effective manner

Cluster computer architecture



- Cluster middleware services and functions

• Single entry point: A user logs onto the cluster rather than to an individual computer.

• Single file hierarchy: The user sees a single hierarchy of file directories under the same root directory.

• Single control point: There is a default workstation used for cluster management and control.

• Single virtual networking: Any node can access any other point in the cluster, even though the actual cluster configuration may consist of multiple interconnected networks.There is a single virtual network operation.

• Single memory space: Distributed shared memory enables programs to share variables.

• Single job-management system: Under a cluster job scheduler,a user can submit a job without specifying the host computer to execute the job.

• Single user interface: A common graphic interface supports all users, regardless of the workstation from which they enter the cluster.

• Single I/O space:Any node can remotely access any I/O peripheral or disk device without knowledge of its physical location.

• Single process space: A uniform process-identification scheme is used. A process on any node can create or communicate with any other process on a remote node.

• Checkpointing:This function periodically saves the process state and intermediate computing results,to allow rollback recovery after a failure.

• Process migration: This function enables load balancing.

The last four items on the preceding list enhance the availability of the cluster. The remaining items are concerned with providing a single system image

Clusters Compared to SMP

Both clusters and symmetric multiprocessors provide a configuration with multiple processors to support high-demand applications.

Both solutions are commercially available,although SMP schemes have been around far longer.

The main strength of the SMP approach is that an SMP is easier to manage and configure than a cluster.

The SMP is much closer to the original single-processor model for which nearly all applications are written.

The principal change required in going from a uniprocessor to an SMP is to the scheduler function.

Another benefit of the SMP is that it usually takes up less physical space and draws less power than a comparable cluster.

A final important benefit is that the SMP products are well established and stable.

Over the long run,however,the advantages of the cluster approach are likely to result in clusters dominating the high-performance server market.

Clusters are far superior to SMPs in terms of incremental and absolute scalability.

Clusters are also superior in terms of availability,because all components of the system can readily be made highly redundant.

NUMA

Nonuniform memory access (NUMA): All processors have access to all parts of main memory using loads and stores.The memory access time of a processor differs depending on which region of main memory is accessed.The last statement is true for all processors; however, for different processors, which memory regions are slower and which are faster differ**.**

Cache-coherent NUMA (CC-NUMA): A NUMA system in which cache coherence is maintained among the caches of the various processors.

A NUMA system without cache coherence is more or less equivalent to a cluster. The commercial products that have received much attention recently are CC-NUMA systems, which are quite distinct from both SMPs and clusters. Usually, but unfortunately not always,such systems are in fact referred to in the commercial literature as CC-NUMA systems.This section is concerned only with CC-NUMA systems.