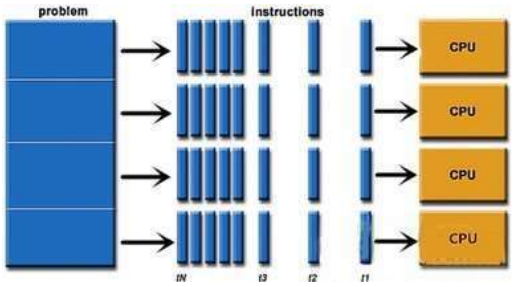
Computer Clusters



Parallel and Distributed Computing

Computer Clusters

- A computer cluster is a ***set of computers*** that ***work together*** so that they can be viewed as a ***single system***.
- Unlike ***grid computers*** (where each node set to perform a **different task/ application**), ***computer clusters*** have ***each node set to perform the same task***, ***controlled*** and ***scheduled*** by ***software***.

Computer Clusters cont...

Sun Microsystems Solaris computer cluster

Computer Clusters cont...

- The components of a cluster are usually connected to each other through ***fast local area networks***, with each **node** (computer used as a server) ***running its own instance of an operating system***.

Computer Clusters cont...

- In most circumstances, all of the nodes use the ***same hardware*** and the ***same operating system***, although in some setups (e.g. using ***Open Source Cluster Application Resources*** (OSCAR)), ***different operating systems*** can be used on each computer, or ***different hardware***.
- **OSCAR** is a Linux-based ***software installation*** for ***high-performance cluster computing***.

Computer Clusters cont...

- Clusters are usually deployed ***to improve performance*** and ***availability*** over that of a ***single computer***, while typically being much more ***cost-effective*** than single computers of **comparable speed** or **availability**.

Computer Clusters cont...

- Computer clusters emerged as a result of convergence of a number of computing trends including the availability of ***low-cost microprocessors***, ***high-speed networks***, and ***software*** for ***high-performance distributed computing***.

Computer Clusters cont...

- Prior to the advent of clusters, single unit ***fault tolerant*** ***mainframes*** with ***modular redundancy*** were employed; but the lower upfront cost of clusters, and increased speed of network fabric has favoured the adoption of clusters.
- In contrast to ***high-reliability mainframes*** ***clusters are*** ***cheaper*** to **scale out**, but also have ***increased complexity*** in ***error handling***.

Computer Clusters cont...
Challenges

- One of the challenges in the use of a computer cluster is the ***cost of administrating*** it which can at times be ***as high as*** the ***cost of administrating N independent machines***, ***if the cluster has N nodes***.

- Some other challenges are discussed here:

Computer Clusters cont…
Challenges (Task scheduling)

- When a large multi-user cluster needs to access ***very large amounts of data***, task ***scheduling becomes a challenge***.
- In a ***heterogeneous CPU-GPU cluster*** with a ***complex application environment***, the performance of each job depends on the characteristics of the underlying cluster.

Computer Clusters cont…
Challenges (Task scheduling)

- Therefore, ***mapping tasks*** onto **CPU cores** and **GPU devices** provides significant challenges.

Computer Clusters cont…
Challenges (Node Failure Management)

- When a node in a cluster fails, strategies such as ***fencing*** may be employed to keep the ***rest of the system operational***.
- Fencing is the process of ***isolating a node*** or ***protecting shared resources*** when a node ***appears to be malfunctioning***.
- There are two classes of fencing methods; one ***disables a node itself***, and the other ***disallows access to resources such as shared disks***.

Computer Clusters cont... Implementation

- Linux supports various **cluster software**; for application clustering, there is distcc, and MPICH.
- **Linux Virtual Server**, *Linux-HA* - director-based clusters that allow incoming requests for services to be distributed across multiple cluster nodes.

Computer Clusters cont... Implementation

- MOSIX, LinuxPMI, Kerrighed, OpenSSI are **full-blown clusters** integrated into the kernel that provide for automatic process migration among ***homogeneous nodes***.
- OpenSSI, openMosix and Kerrighed are single-system image implementations.

Computer Clusters cont... Implementation

- Microsoft Windows computer cluster **Server 2003** based on the Windows Server platform provides pieces for High Performance Computing like the ***Job Scheduler***, ***MSMPI library*** and ***management tools***.