

Clustered Computing

Robert Tompkins



Topics discussed

- Define parallel computing
 - Paradigms and purposes
 - High availability
 - PVM/MPI
 - Mosix
 - Beowulf 2
 - NFS
 - Lustre
-
-

Paradigms and purposes

- Clustered computing serves different purpose, and uses different software to accomplish that goal
 - High availability for systems that need to be up 99.99% of the time
 - High performance for scientific computing
 - High demand/workload for webservers and databases
-
-

High availability Redhat and heartbeats

- Sometimes it is important to have a machine running 99.99% of the time
 - Redhat and other HA distributions allow you to have two or more machines doing the exact same transactions
 - When one system goes down, the backup system detects it and automatically takes over its ip address and transactions continue
-
-

PVM/MPI

- **Parallel Virtual Machine** shares the memory and cpu across all of the nodes
- **MPI** starts a process on each of the nodes that can communicate with all of other other nodes, similar multithreaded programming

Mosix

- **Mosix** allows you to migrate a process from one machine to another
- The entire memory space is moved from one machine to another
- Processes can migrate to machines that are least used



Beowulf 2

- **Beowulf 2** is a reimplementation of Beowulf with Lightweight kernels
- This allows more of the processor to be used
- Also decrease the management requirements and required resources
- The machine only needs memory and a cpu

NFS/LUSTRE

- All clusters need a shared file system
 - Originally most used **NFS** or **AFS** to mount the home directories
 - **PVFS** was an attempt to distribute writes across the nodes, the same as raid 0 and raid 1
 - **Lustre** is a commercial/freeware filesystem that does raid 0,1,5 and drastically increases the write and read speeds
-
-