Linux Clusters

Jim Phillips, John Stone
Theoretical Biophysics Group

- What is a cluster good for?
- Applications / How to use it
- Hardware
- System Software

Why Clusters

- Cheap alternative to "Big Iron"
- In-house development platform for "Big Iron" code
- Built to task (buy only what you need)
- Built from COTS components
- Runs COTS software (Linux/MPI/PVM)
- Lower yearly maintenance costs
- Single hw/sw failure does not take down entire facility
- Re-deploy as desktops or "throw away"

Why Not Clusters

- Non-parallelizable or tightly coupled parallelism needed
- Significant existing codebase, cost too high to port
- No source code for application(s)
- No local expertise (Don’t know Unix?)
- No vendor hand holding
- Massive I/O or memory requirements

How to Use a Cluster

Serial Jobs:

- Large number of queued serial jobs
- Cluster is used for increased throughput
- Run standard applications, no changes needed
- Example: Pixar renderfarm

Parallel Jobs:

- Small number of highly parallel jobs
- Cluster is used for decreased turnaround time
- Run parallelized applications, changes required
- Example: TB Linux cluster

Hardware

CPUs:

- Alpha (64-bit), Intel/AMD (32-bit), Sparc (64-bit)
- Integer / Floating Point balance

Nodes:

- Symmetric Multiprocessor vs. Uniprocessor
- System memory bandwidth
- SPECfp vs. SPECratefp
- Build to suit the task, don’t buy from "Uncle Bob"

Network:

- 100baseT /w Hub
- 100baseT /w Switch
- Myrinet / Giganet / SCI

Summary:

- More (slow) nodes: harder to utilize, harder to manage
- Less (fast) nodes: easier to utilize, easier to manage

System Software

Operating System:

- SMP: Do threads work? Does kernel scale?
- Performance of TCP/IP, NFS, YP etc.
- Security: Are you going to get cracked?

Compilers:

- Spend money, get good compilers
- Real compilers give 10-30% performance increase

Network Configuration:

- Firewall or part of normal environment, not both!
- Parallel communication: TCP, UDP, Myrinet-like?
- System services: NIS, DNS, NFS, AFS etc.
- Do system services fight with parallel communication?

Queuing System:

- DQS, PBS, Custom
- How to deal with 1-CPU jobs on SMP nodes