So You Want to Deploy a Production Cluster...

Dana Brunson - Oklahoma State University

Jeff Pummill - University of Arkansas

Words to live by

"It depends..."

--H. Neeman

Environmental Considerations

- Power
- Cooling

Power Usage Examples

Machine	# of racks	Theortical (Gflops)	Power (kilowatts)
Cimarron (OSU-mini)	0.5	896	5.75
Star (UARK)	8	13,364	73.37
Roadrunner	296	1,375,780	2,345.50

Relevant Links

http://www.dell.com/calc

http://www.sun.com/servers/x64/x2250/calc/index.jsp

http://www-03.ibm.com/systems/bladecenter/resources/powerconfig/index.html

Intended Use

- R & D Cluster
- Production Cluster

Quote

"In theory, there is no difference between theory and practice.
But in practice, there is!"

-- anonymous

Hardware Choices

- Whiteboxes
- Commodity Servers
- True Supercomputers
- Hybrid Hardware

Cluster Software Stacks

- Free: ROCKS, OSCAR, or xCAT
- Commercial: OCS or Rocks+
- Alternately: Roll-Yer-Own

Relevant Links

http://www.rocksclusters.org
http://oscar/openclustergroup.org
http://xcat.sourceforge.net/
http://my.platform.com/products/platform-ocs
http://clustercorp.com/rocksplus/index.html
http://debianclusters.cs.uni.edu/index.php/Main_Page

Filesystem Choice

- NFS
- PVFS2
- Lustre
- Panasas

HPC File System Articles

HPC File Systems by Jeff Layton

http://www.linux-mag.com/id/4169 (part 1)

http://www.linux-mag.com/id/4181 (part 2)

http://www.linux-mag.com/id/4358 (part 3)

Relevant Links

http://nfs.sourceforge.net/nfs-howto

http://www.pvfs.org/

http://wiki.lustre.org

http://www.panasas.com/

Quote

We stand at a crossroads. One path leads to despair, the other to destruction. Let's hope we make the right choice.

--Woody Allen

Interconnect Options

- Gigabit Ethernet
- Infiniband / Myrinet
- 10Gig Ethernet

Latency & Bandwidth comparison

Interconnect	Latency (microseconds)	Bandwidth (MBps)	N/2 (Bytes)
GigE	~29-120	~125	~8,000
GigE: GAMMA	~9.5 (MPI)	~125	~7,600
10 GigE: Chelsio (Copper)	9.6	~862	~100,000+
Infiniband: Mellanox Infinihost (PCI-X)	4.1	760	512
Infiniband: Mellanox Infinihost III EX SDR	2.6	938	480
Infiniband: Mellanox Infinihost III EX DDR	2.25	1502	480

data from http://www.linux-mag.com/id/3507

HPC Network Articles

HPC Networks by Jeff Layton

http://www.linux-mag.com/id/3507 (part 1)

http://www.linux-mag.com/id/4146 (part 2)

Relevant Links

http://en.wikipedia.org/wiki/Gigabit_ethernet
http://www.qlogic.com/Products/HPC_products_landingpage.aspx
http://www.myri.com/
http://en.wikipedia.org/wiki/10_Gigabit_Ethernet

Schedulers / Resource Managers

- Honor System
- Free: Torque / SGE / Slurm
- Commercial: LSF / MOAB

Relevant Links

http://www.clusterresources.com/pages/products/torque-resource-manager.php

http://www.sun.com/software/gridware/

https://computing.llnl.gov/linux/slurm/

http://www.platform.com/Products/platform-lsf

http://www.clusterresources.com/pages/products/moab-cluster-suite.php

Software Applications

- Serial vs Parallel
- Open Source vs Commercial

Cluster Tools

- Modules
- IPMI
- screen
- Commercial Remote Access Tools
- Ganglia

Data Management

- Backups
- User Quotas
- Aging Scripts

Quote

I think I might believe what I just said!

-- Bill Camp