Diskless Computing

Frank Roybal, NMHU

Nickolas Torres, UNM

Amanda Muñoz, NMT

Computer System, Cluster, and Networking Summer Institute

Mentors: David Kennel, Andree Jacobson, Carolyn Connor

Cluster Background

- **Increasing Cluster Size**
  - Roadrunner has around 3000 compute nodes
- **Reliability Issues**
  - Hard Drives
  - Other Components
- **Move to Diskless Cluster**
  - Cost Savings
  - Increased Reliability

Choosing a Diskless System Management Solutions

- Perceus
- XCAT2

Disk Full Clusters

- Nodes are installed and configured using a kickstart script
- Nodes are booted from their hard drives
- Changes in configuration need to be done on each node

Perceus and XCAT2 Diskless Clusters

- Using a diskless management solution a bootable image is generate
- Changes in configuration can be made on the image and will be applied next time a node is rebooted
- Nodes are then provisioned with the bootable image using PXEBoot
- Perceus is simple to use and good for smaller cluster
- XCAT2 is more complicated to set up and will work well with larger clusters
- Hardware support issues

Results Part 1

Feature Comparison	Perceus	XCAT2	Disk Full
Documentation	7	5	9
Hard Drive Required	No	No	Yes
Node Installation Method	PXE Boot	PXE Boot	Kickstart
Reliability	8	8	6
Hardware Compatibility	5	7	10
Personal Preference	8	7	8
Time Required	6	7	4
Scalability	5	7	8
Feature Rating	39	41	45

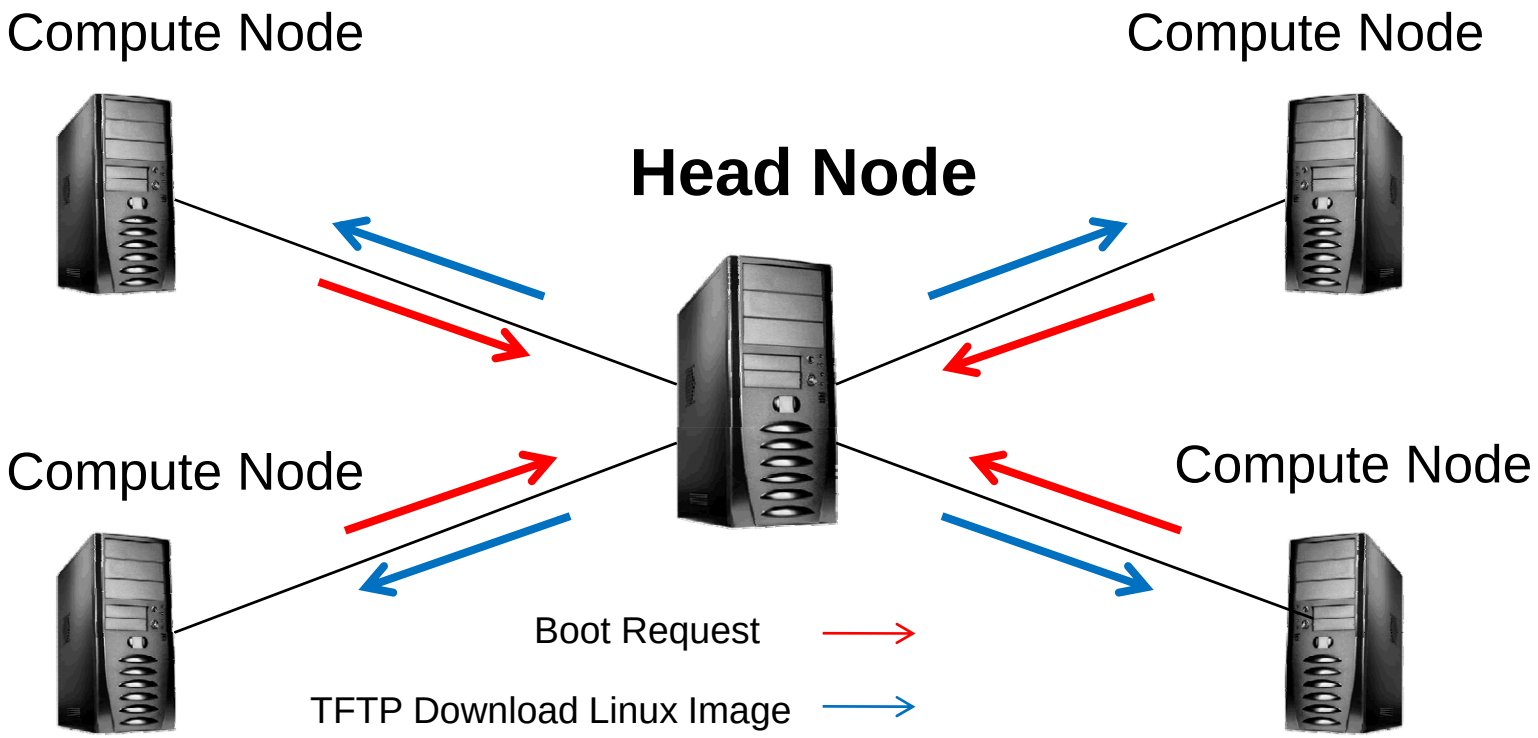
Scale: 1 = Poor 10 = Excellent

Results Part 2

Difficulty of :	Perceus	XCAT2	Disk Full
System Installation	8	7	8
System Interface	9	5	8
Adding Nodes	10	7	7
Replacing Nodes	7	7	6
Node Configuration	9	7	6
Changing Node Image	9	7	6
Scheduler Implementation	3	6	7
Infiniband Installation	7	7	7
MPI Job Execution	8	8	8
Difficulty Rating	70	61	63
Overall Rating	109	102	108

Scale: 1 = Difficult 10 = Easy

How a Diskless Cluster Works



Gluster

- **Gluster is a cluster file system that can integrate many storage devices into one large file system**
  - In our case 2.25 tera-bytes using the unused hard drives in the compute nodes
  - Can be scaled up to several peta-bytes
- **It is capable of dealing with disk failure**
  - Gluster can be configured to repair itself if a node must be eliminated
- **Gluster is a network file system**
  - Works via InfiniBand or TCP/IP

GlusterFS (continued)

- **The head node is the only node with a diskfull Operating System install**
  - Compute nodes are booted using the Perceus diskless infrastructure
  - Configuration files are installed in the image Perceus boots on nodes
  - All compute nodes now have access to a 2.25 tera-byte file system
    - Again, the hard drives are used SOLELY for data storage
- **Gluster is very well documented**

Conclusion

- **Disk full cluster are simpler to set up however they are hard to maintain**
  - Reliability
- **Diskless cluster are the future of high performance computing**
- **Perceus works well for managing smaller diskless clusters**
  - Does not work as well for larger clusters
- **XCAT2 is harder to implement but is designed to work well with large scale clusters**

Contacts

- **Nickolas Torres**
  - ntorres1@unm.edu
- **Frank Roybal**
  - froybal5@student.nmhu.edu
- **Amanda Muñoz**
  - amunoz@nmt.edu