



# Resource Management on a Mixed Processor Linux Cluster

Haibo Wang
Mississippi Center for Supercomputing Research

May 20, CUG SUMMIT 2002

# MCSR Linux Beowulf Cluster

16 Nodes

July, 2000

56 Nodes

July, 2001

# MCSR Linux Beowulf Cluster Configuration

July 2000

1 Gateway Node (Server)

16 Compute Nodes

(16 Pentium III)

July 2001

1 Gateway Node

3 I/O Nodes (Pentium IV)

53 Compute Nodes

(16 Pentium III and 37 Pentium IV)

# Applications and Users Information

Applications/Software: MPICH, NWCHEM, GAMESS, MPQC, GAUSSIAN, GA

User Information:

- 30 Research Accounts
- 20 General User Accounts
- 16 Class Accounts
- 4 Universities
- 7 Research Groups

# Resource Management Issues

I/O Performance Management:
- I/O Hardware
- File Systems
- Ratio of Compute to I/O Node

Job Scheduling:
- PBS vs PBS Pro
- Application Specified Scripts

Application Specified Resource Management:
- Resource Usage
- Scratch / Checkpoint

Kernel Parameter Tuning

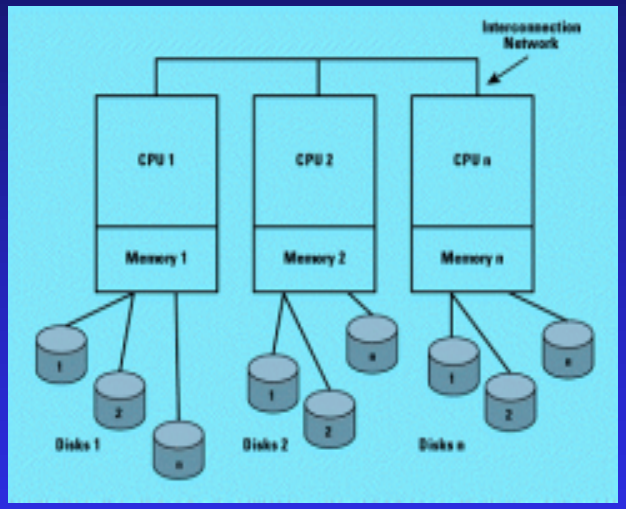
# I/O Performance Management

Sufficient I/O Hardware
- Fast System
- High Quality Hardware
- RAID

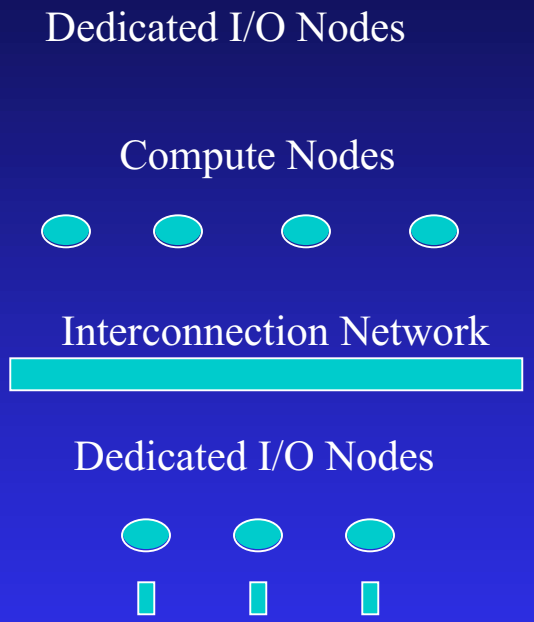
Faster File System
- NFS to NFS + RAID
- XFS
- PVFS

# Why Dedicated I/O Node

Shared I/O & Compute Nodes




Dedicated I/O Nodes



# I/O Performance Management

Ratio of Compute to I/O Node

User Per I/O Node
Job Per I/O Node
How Many?

# I/O Performance Management

Our Approaches to Improve I/O Performance

1. Define User Type
2. Define Job Type
3. Configure I/O Nodes
4. Use Optimized I/O Implementation

# Job Scheduling

From Open PBS to PBS Pro

- Grouping Users
- Grouping Nodes
- Grouping Jobs
- Job Priority
- Deleting Jobs On Dead Nodes
- Allow and Deny Access
- Preemptive Job Scheduling

# Job Scheduling

Application Specified Scripts:

- Load Balancing
- Improving Reliability
- Optimizing Resource Usage
- Debugging User's Input

# Application Resource Management

MPI Based Applications:

- NWCHEM
- MPQC
- GAMESS
- GA

Other Application:

- Parallel Gaussian

Scratch /Checkpoint Files

# Kernel Parameter Tuning

SHMMAX

| | |
|---|---|
| Default Value: | 64MB |
| Physical Memory: | 512MB/node |
| System and Application Usage: | 250MB/node |
| New Value: | 256MB |

# Conclusion

Improving I/O Performance

More Application Specified PBS Scripts

More Complicated Node/User/Job Grouping

Fast File System

Future Upgrading

# Contact Information

Haibo Wang

Chwang@olemiss.edu

http://www.mcsr.olemiss.edu