

Data Storage and Movement on Farms

First “how to ...?” session

Data Storage Options

- Home areas
 - Limited size: 17G on fnsfo, shared by all users
 - Used for: sources, executables, libraries, control
 - Globally visible (currently via NFS)
 - Not for data
 - Statically allocated to groups/users
 - Permanent storage, backups
- Staging areas
 - Large capacity: 5-30 G * 10-50 disks
 - Used for: data, executables, libraries, control
 - Globally visible via NFS
 - Can be used for data storage, not for remote writing
 - Accessible through NFS or rcp/scp, etc.
 - Statically allocated to groups
 - Permanent storage, no backups

Data Storage Options

- Scratch disks on workers
 - Average capacity: 8-30 G * 1-4 disks on each node
 - Used for: data, executables, libraries
 - Visible locally, direct access
 - Can be used for scratch data, must be cleaned when process exits
 - Dynamically allocated to batch processes
 - Short-term storage – as long as the process runs
- Workers disk storage
 - Very large capacity 8-30 G * 1-4 disks * 50 nodes
 - Should be used for intermediate data
 - Globally visible
 - Accessible through “disk farm”
 - Allocated using quotas
 - Short-term parking – between stages of production

Data Movement

- Between tape and staging disk
 - OCS + tape read/write
 - MSS (fmss, enstore)

- Between staging disk and remote sites
 - ftp, rcp, scp

- Between worker nodes
 - ftp, rcp, scp

- rcp or scp ?
 - Identical semantics
 - scp encrypts data
 - Good for sensitive data
 - Needs a "lot" of CPU

Data Movement

Between worker nodes and central (I/O) node

- Star topology: 50 nodes – 100 processes vs. 4 CPUs of I/O node
- NFS
 - As easy as “cp”, open()...
 - No transaction concept – can not refuse
 - Very delicate, fails if NFS server node is loaded
 - May overload NFS (central) node
 - NFS + data = **volatile explosive !**
- AFS – may be a better choice, but not without price
- rcp/scp
 - Better choice for data – transaction-based, can be controlled
 - As fast as NFS
- fcp – controlled gate to rcp/scp
 - Limits number of simultaneous transactions
 - Blocks the client until disk space is available

Conclusion: do's and dont's

- Do not store data in home area
- Do not transfer data over NFS
- Do not use significant amount (>0.5G) of scratch disk without allocating it
- Use staging areas or disk farm
- Use rcp/scp through fcp
- Allocate scratch space using
PROC_RESOURCES = disk:5