Lattice Gauge Computing User Environment

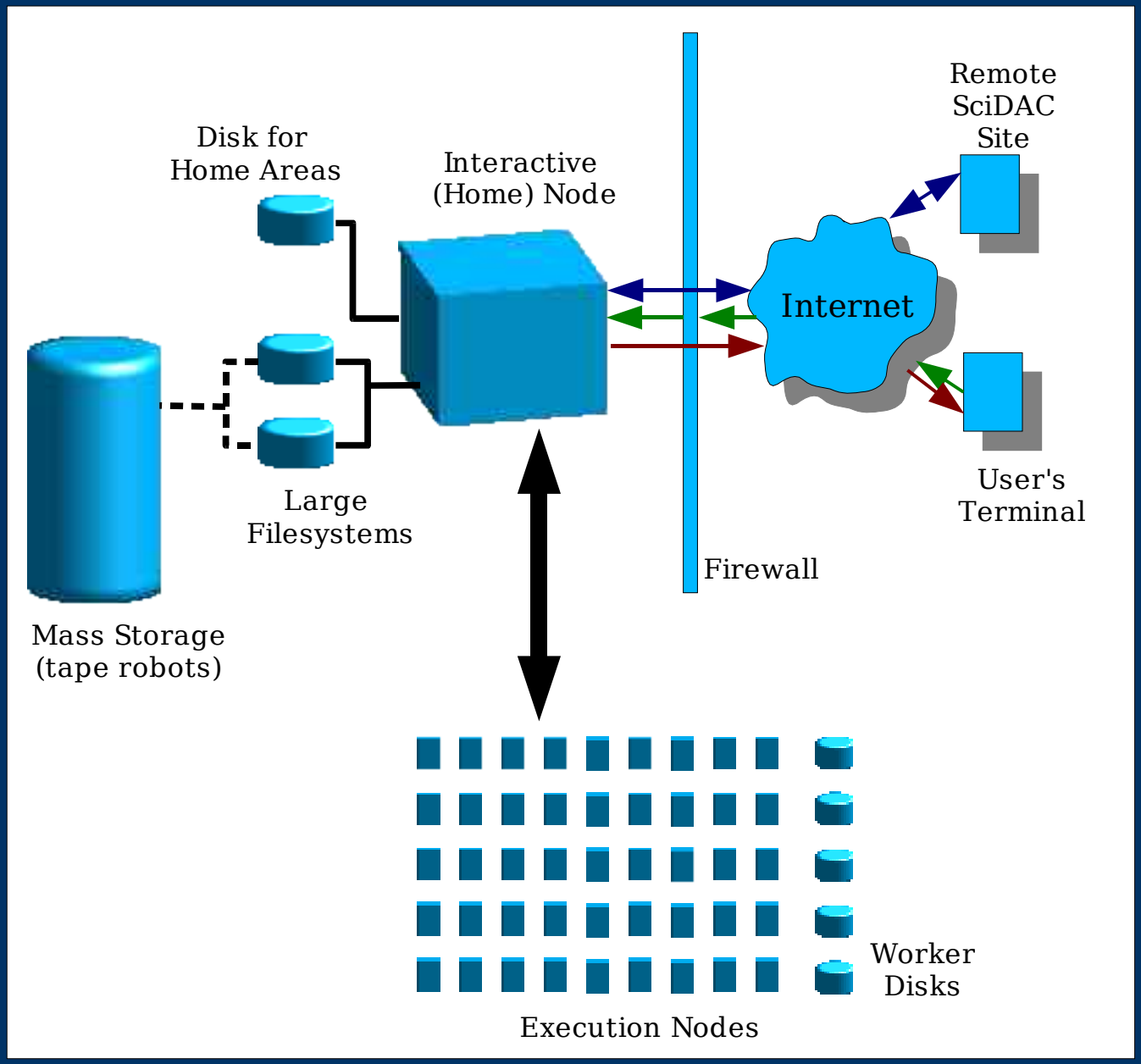
Don Holmgren
Fermilab
SciDac Collaboration All-Hands Meeting
March 26, 2004

http://lqcd.fnal.gov/runtime.pdf

User Environment

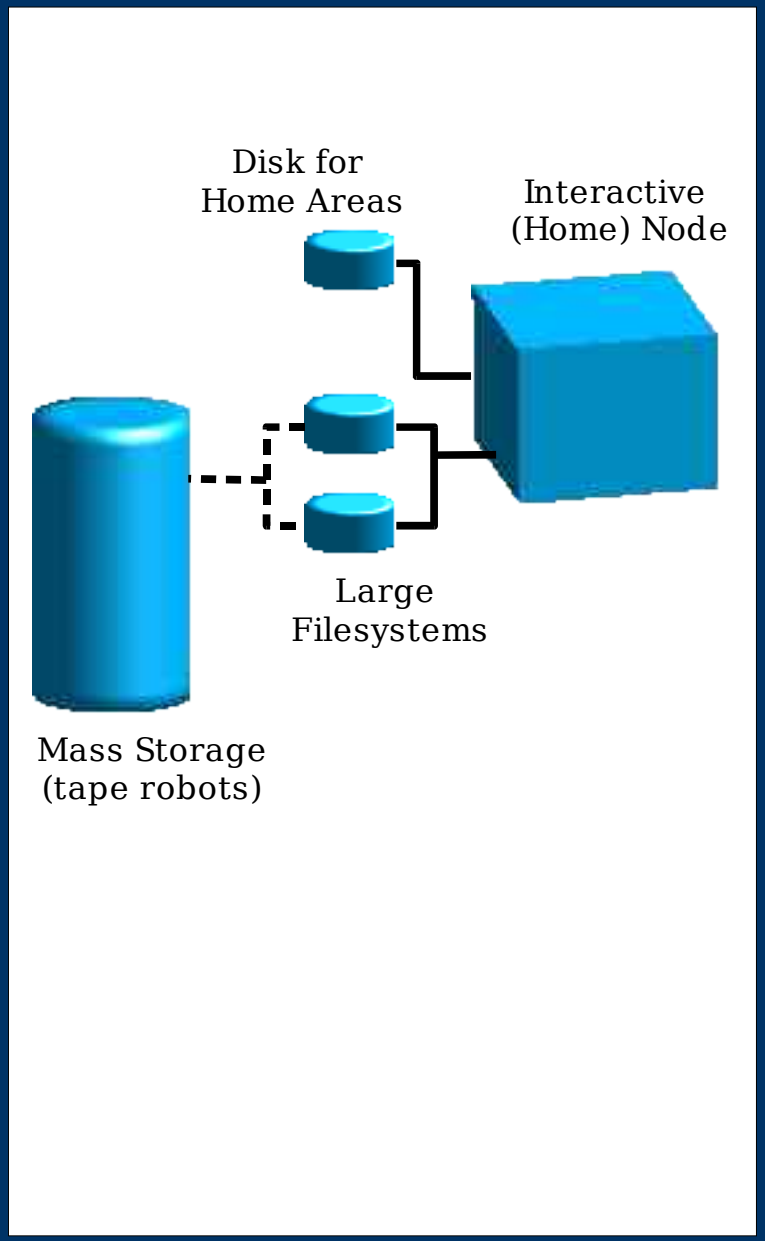
- The goal: Establish a uniform user (“runtime”) environment at each of the SciDAC lattice gauge computing sites.
- This implies standardization of:
  - the interactive environment
    - the view from user's terminal
  - the batch environment
    - how jobs are queued and manipulated
    - the view from job scripts
  - the execution environment
    - the view from execution nodes
  - the storage environment
    - how to access local and remote storage
- Scripts and makefiles which conform to the specifications should run without modification at each site

Architecture of a Generic Lattice Facility



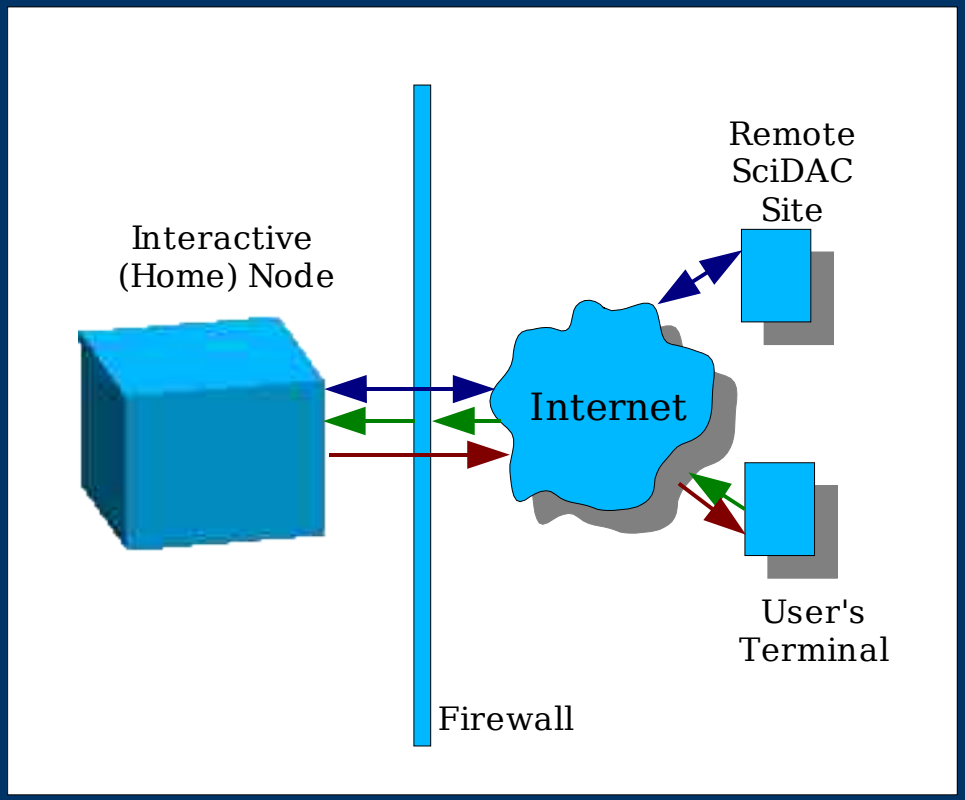
- Interactive node:
  - runs batch system
  - the access point from the internet
  - attached to large filesystems
  - protected by firewall
- Execution nodes:
  - each may have local scratch disk (clusters) or may share one of many NFS servers (QCDOC)

File Systems



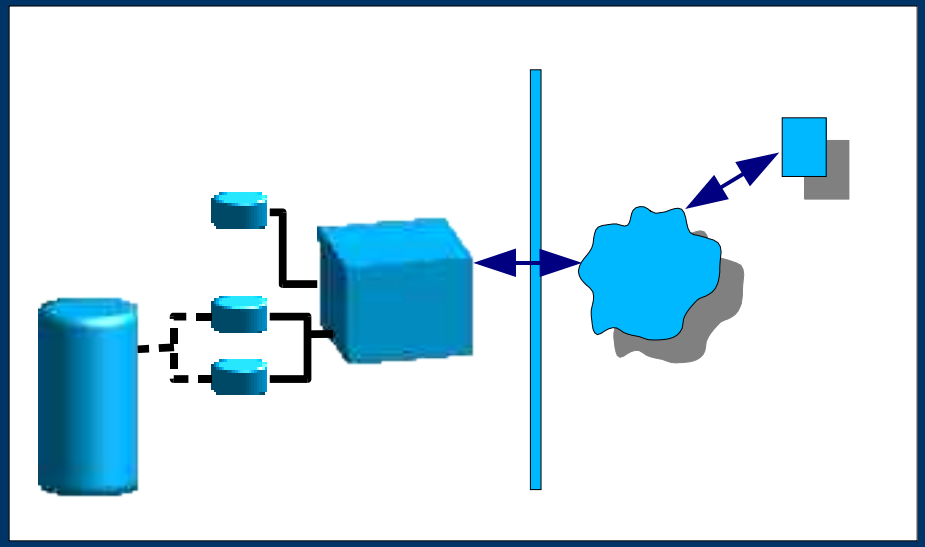
- home directories
  - backed-up frequently, with quota
  - /home/<user-name> and $HOME
- large filesystems
  - not backed up, likely RAID'ed
  - possibly backed by tertiary (*i.e.*, tape) storage
  - always space to write output of active jobs (100 GBytes this year)
  - possibly with auto-migrate to tape or another site
    - controlled by file attributes or by location in filesystem
  - commands to move large files (> 2GB) to/from compute nodes
  - /cache/projectA
  - /cache/users/<user-name>
- no maximum file size, but files over 2 GB may not work with all utilities

Interactive Environment



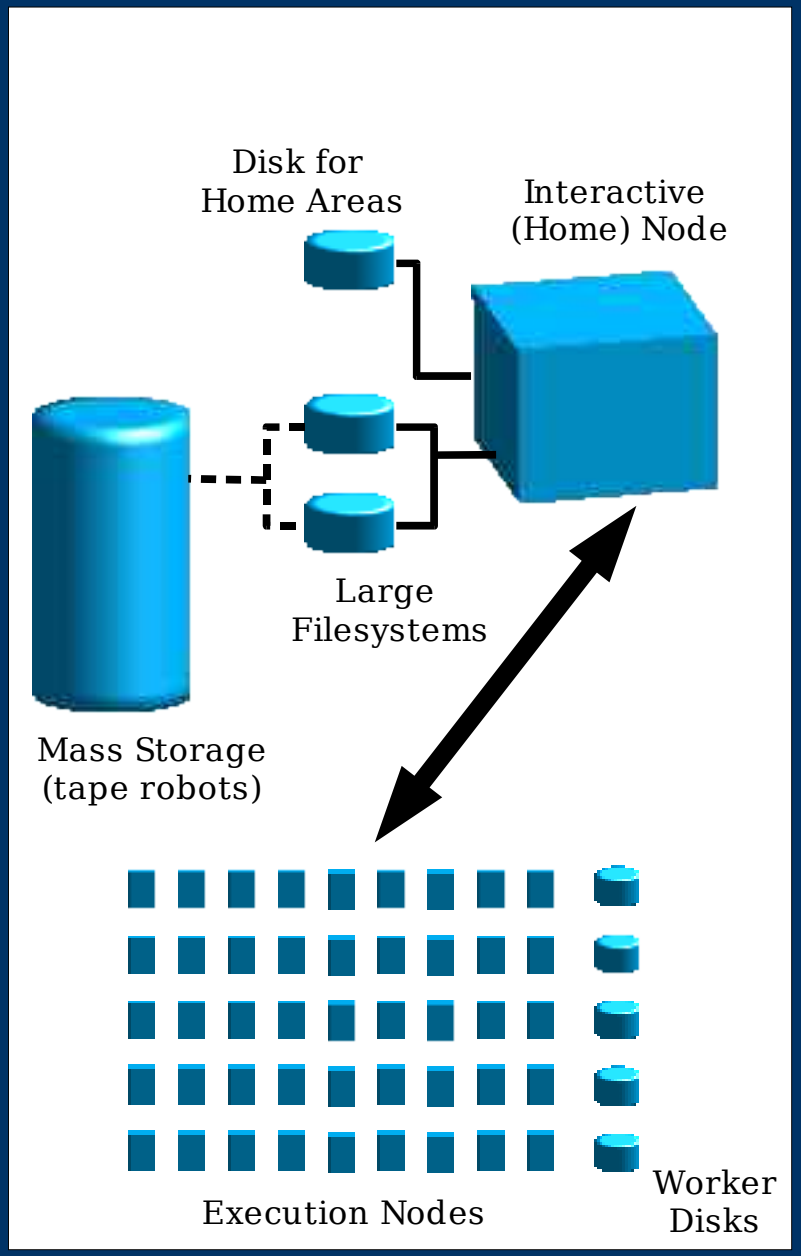
- Standard UNIX shell
  - bash, csh/tcsh, others
- DOE mandated firewall
  - JLAB, BNL: first login to firewall node, then login to interactive node
  - FNAL: Kerberized access direct to interactive node
- Moving files to/from remote sites
  - may require 2 hops
  - via GRID tools, single hop transfers between SciDAC sites (maybe not in 2004)
  - single hop to FNAL from behind firewall at JLAB and BNL via kerberized clients

Interactive Environment



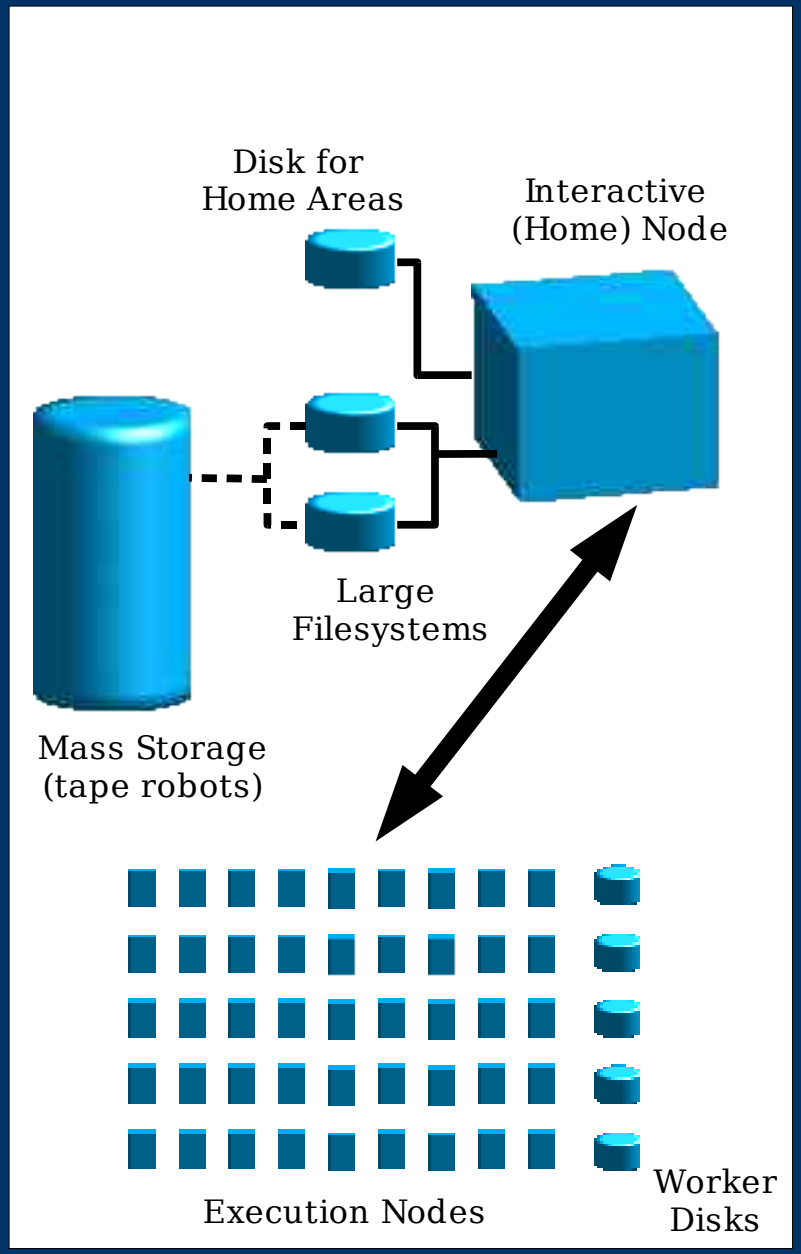
- Large file system is mounted
- Data grid commands to:
  - fetch by global file name (GFN)
  - push / publish a new file
  - request a copy of a file to move to this or a remote site
- Standardized environment variables (*eg* for makefiles)
  - final names to be determined
  - manually set by user, or via a “setup” utility
  - examples:
    - $MPI_DIR, $QMP_DIR, $QIO_DIR, $QDP_DIR

Batch Environment



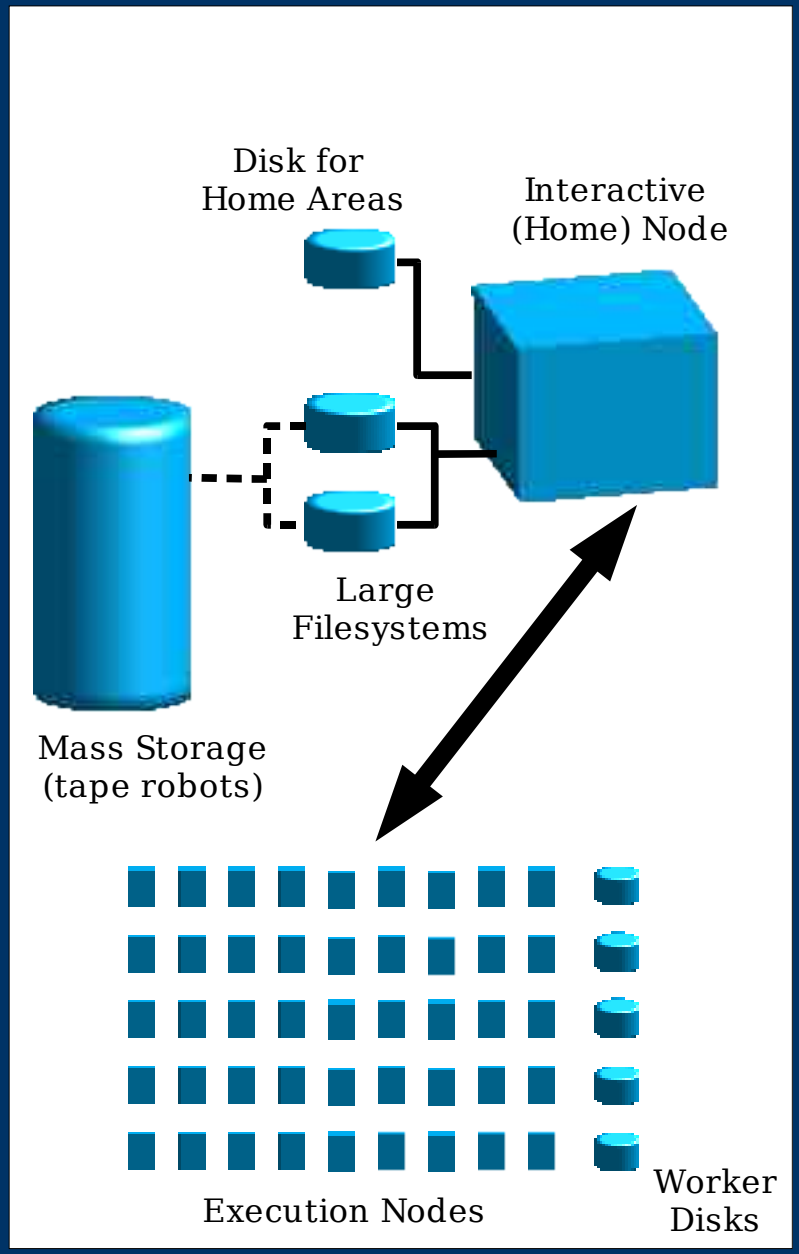
- Standardized commands to submit and control jobs, based on PBS:
  - qsub, qstat, qdel, qalter, qhold, qrls
  - queue names and policies TBD
- Time limits
  - upper bound set so that normal outages and failures will cause no more than a 5% efficiency loss/year
  - limit this year is 24 hours
- stdout, stderr (treated separately)
  - node 0 spools to host
  - optionally capture from all nodes
  - all can be seen in real time from interactive node
  - redirectable to a file visible to interactive user
- Exit status of completed jobs can be queried

Batch Environment



- Batch jobs can submit batch jobs
- Project accounting:
  - qsub has a switch to specify a project charge code
  - batch system will reject jobs submitted with expired or overspent charge codes
  - users may retrieve account balance for each charge code
- Job scripts
  - qrun is used to launch executable on all nodes and is responsible for distributing executable image
  - executables which read from stdin will read from temporary files generated at job launch from the qrun invocation and distributed to execution nodes

Batch Environment



- Large filesystem may not be accessible via open() from execution nodes
  - qcp command will move files from large filesystem to local disks or NFS attached disks on execution nodes
  - user job scripts are responsible for staging files

For Further Information

- Current runtime environment specification is available from http://lqcd.fnal.gov/RunTimeEnv.html
- This presentation will be updated to reflect changes to the specification, and will available from http://lqcd.fnal.gov/runtime.pdf
- We would greatly benefit from your feedback
  - what have we overlooked?
  - what are we making too difficult for the users?
- Remember, we are specifying a minimum set of requirements for a portable environment
  - makefiles, build scripts, and run scripts which use this minimal set will be portable across the sites
  - the individual sites will no doubt have additional features available which you can use at the expense of loss of portability