Alliance HPC Resources at

Lawrence Livermore National Laboratory

Blaise Barney, LLNL
Tom Spelce, LLNL

University of Michigan, Ann Arbor
October 15, 2009

UCRL-PRES-208318, LLNL-PRES-404021, UCRL-TR-223612, LLNL-PRES-405061, UCRL-WEB-148749

Overview

- Who is the CRT?
- Organization
- LLNL HPC Platforms
- Accounts & Access
- Software Environment
- File Systems
- Running Jobs
- Queues and DATs
- Support, Documentation & Training
- Alliance Communications
- Usage Stats
- The Future
- Gratuitous Plug for SC09 Booth

Who Is The CRT?

- The Computer Resource Team (CRT) is the component of the PSAAP program that connects Alliance researchers to the High Performance Computing (HPC) resources required to perform their work.

- The CRT is comprised of a representative from each NNSA Lab who is familiar with their lab's computing resources, personnel and policies. The following individuals serve on the CRT:
  - Blaise Barney, LLNL
  - Rob Cunningham, LANL
  - TBD, SNL

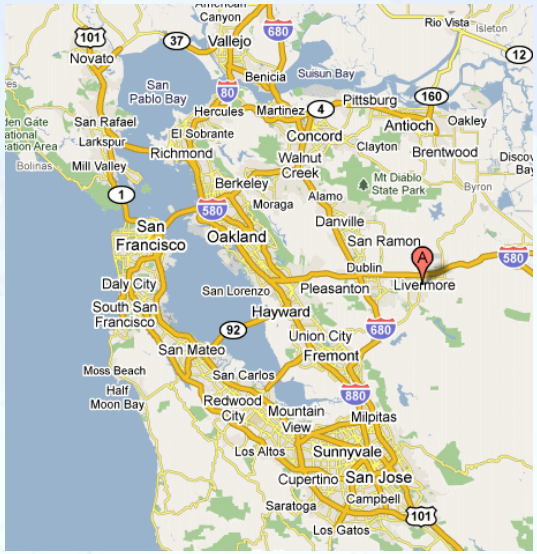
- Our primary purpose is to provide assistance and guidance in all aspects related to the use of HPC resources located at LANL, LLNL, and Sandia.

What Does The CRT Do For You?

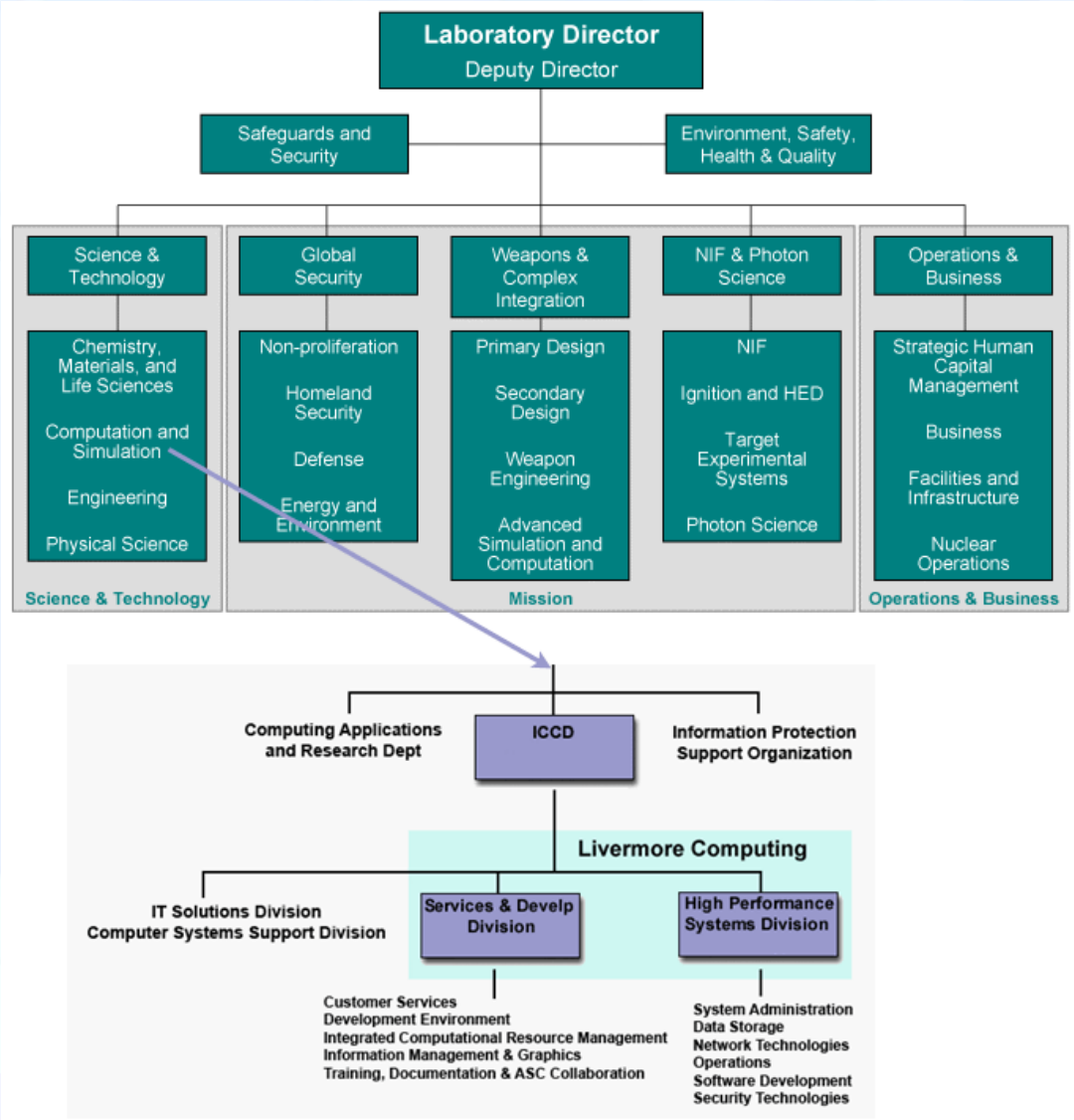
- Assist with the establishment and use of computer accounts
- Assist with accessing compute resources
- Provide HPC user documentation
- Provide technical support and referral to in-depth consulting
- Conduct monthly telecons to keep Alliance users up-to-date with account, access, policy, scheduling and technical issues, and to address issues with HPC platform usage
- Interface with other groups within the Labs, such as management, networking, system administration, storage, customer support, etc., to facilitate the effective support of Alliance users
- Track and facilitate the resolution of problems reported to each Labs' customer support “hotline”
- Provide training opportunities
- Collect and distribute monthly machine usage statistics
- Schedule and support special/dedicated runs
- Maintain a balance of machine usage between the Alliances
- Conduct Alliance visits to discuss HPC resources, user issues and to offer technical consultation and/or training
- Showcase Alliance research in the NNSA/ASC research exhibit booth at the annual SC conference

Lawrence Livermore National Laboratory (LLNL)

- Located in Livermore, CA
- Established in 1952
- 7,000+ employees and contractors
- Managed by Lawrence Livermore National Security, LLC (LLNS)
  - Bechtel National, University of California, Babcock and Wilcox, the Washington Division of URS Corporation, and Battelle
- Budget @ 1.4 billion



Organization



LLNL HPC Platforms - Summary

- IBM BG/L and BG/P
- AMD Opteron
- IBM Power5
- Dell Intel Xeon

System Category	TFLOP/s	% of Total
Unclassified	**559.0**	**43%**
Capability	362.1	65%
Capacity	145.7	26%
Serial	4.0	1%
Visualization	1.2	0%
Collaboration	46.1	8%
Classified	**729.0**	**57%**
Capability	460.4	63%
Capacity	258.0	35%
Serial	8.1	1%
Visualization	2.5	0%

System	Top500 Rank	Program	Manufacture/ Model	OS	Interconnect	Nodes	Cores	Memory (GB)	Peak TFLOP/s
Unclassified Network (OCF)									**559.0**
uBG/L	4*	ASC+M&IC	IBM BG/L	Linux	IBM	40,960	81,920	21,504	229.4
Hera (TLCC)	TBD	ASC+M&IC	Appro	Linux	IB DDR	864	13,824	27,648	127.2
Coastal	TBD	M&IC	Dell	Linux	IB DDR	1,152	9,216	13,824	88.5
Hyperion	TBD	ASC	Dell	Linux	IB DDR	576	4,608	4,608	46.1
Atlas (Peloton)	74	M&IC	Appro	Linux	IB DDR	1,152	9,216	18,432	44.2
Zeus (Peloton)		M&IC	Appro	Linux	IB DDR	288	2,304	4,608	11.1
uPurple		ASC	IBM SP	AIX	Federation	108	864	3,456	6.6
Yana		M&IC	Appro	Linux	N/A	80	640	1,600	3.1
Prism		ASC	GraphStream	Linux	IB SDR	128	256	2,048	1.2
Hive		M&IC	Appro	Linux	IB DDR	6	96	1,088	0.9
Ebert		M&IC	Appro	Linux	N/A	6	96	192	0.9
Classified Network (SCF)									**729.0**
Dawn	TBD	ASC	IBM BG/P	Linux	IBM	36,864	147,456	147,456	501.4
BlueGene/L	4*	ASC	IBM BG/L	Linux	IBM	65,536	131,072	49,152	367.0
Juno (TLCC)	TBD	ASC	Appro	Linux	IB DDR	1,152	18,432	36,864	162.2
Purple	33	ASC	IBM SP	AIX	Federation	1,532	12,288	49,152	93.4
Eos (TLCC)	TBD	ASC	Appro	Linux	IB DDR	288	4,608	9,216	40.6
Minos (Peloton)	99	ASC	Appro	Linux	IB DDR	864	6,912	13,824	33.2
Rhea (Peloton)	236	ASC	Appro	Linux	IB SDR	576	4,608	9,216	22.1
Hopi		ASC	Appro	Linux	N/A	112	1,152	2,896	7.6
Gauss		ASC	GraphStream	Linux	IB SDR	256	512	3,072	2.5
Tempest		ASC	IBM	AIX	N/A	12	84	672	0.6

LLNL Alliance Compute Resources
Hera

- Hera Configuration
  - AMD quad-core quad-socket Opteron @2.3 GHz
  - 127.2 TFlop system
  - 864 nodes; 16 cpus/node; 13,824 cpus total
  - 32 GB memory per node
  - 4X DDR Infiniband
    - 20 Gbits/sec rating
    - 2 - 2.5 Gbytes/sec non-block pt2pt MPI in practice
  - 64-bit architecture; CHAOS OS
  - Lustre parallel I/O file systems
- Usage
  - ASC and Institutional resource
  - Alliance allocation @20% of the machine
  - DAT runs

LLNL Alliance Compute Resources UBGL

- UBGL = Unclassified BGL
- UBGL Configuration
  - IBM PowerPC 440 @ 700 MHz
  - 229.4 TFlop system
  - 40,960 nodes; 2 cpus/node; 81,920 cpus total
  - 512 MB memory per node
  - Multiple networks
    - Torus for pt2pt MPI (2.1 GB/sec)
    - Tree for MPI collectives (2.1 GB/sec)
    - Barrier
  - 32-bit architecture; lightweight kernel OS
  - 690 TB Lustre parallel I/O file system
- Usage
  - ASC and Institutional resource
  - Alliance allocation @20% of the machine

Alliance Use of UBGL

- LLNL usage model will encourage “capability” rather than “capacity” computing
- Looking for users who can routinely run jobs using 8K, 16K and 32K nodes
- Potential for full system 40K node runs
- Debugging queues are also available to run smaller jobs in the range of 128 to 4K nodes
- Limited number of users over a given period – selected by their ability to utilize the machine as a capability platform
- Interested? Please talk to your LLNL Computer Resource Team (CRT) representative: Blaise Barney 925-422-2578 blaiseb@llnl.gov
- Note: this resource can be selected as usual through the Sarape account request website

Account Requests

- Begin with Sarape online: sarape.sandia.gov
  - Most fields are self explanatory
  - Select LLNL and then "hera" or "ubgl" from pull-down menu
  - Additional hints for completing the form can be found at: asc.llnl.gov/alliances/quickguideLLNLLANL.pdf
- Coordinated and approved through the Univ. Michigan Alliance POC - Quentin Stout.
  - Account request is then processed by LLNL.
- Account processing for non-US citizens requires additional time and “paperwork” - allow 30-90 days (plan ahead)
- Following approval, you will have two "identities" at LLNL
  - Institutional (lab-wide) Official User Name/ID (OUN) - smith25
  - Livermore Computing machine login name - jjsmith
  - Sometimes they can be the same

Accounts to Date

- Seven Michigan user accounts on LLNL's Hera cluster:
  - Bruce Fryxell
  - Daryl Hawkins
  - Ryan McClarren
  - Eric Myra
  - Ben Torralva
  - Gabor Toth
  - Bartholomeus Van der Holst
- No UBGL accounts yet

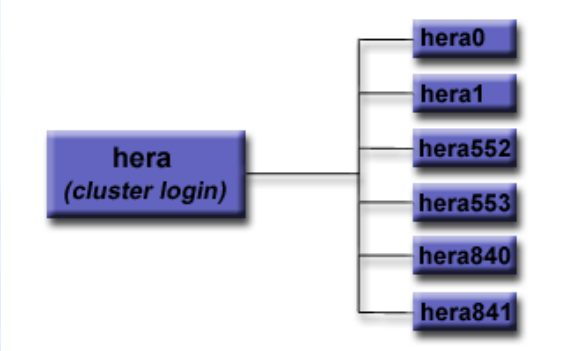
Access

- One Time Password (OTP) token
  - Provided when your account request is approved
  - 2-factor authentication: static PIN + OTP passcode
  - Used for all LC machines and some web pages
  - More information & help: access.llnl.gov/otp

- Virtual Private Network - Collaborator (VPN-C)
  - VPN-C account provided along with your machine account. Required for all remote machine access.
  - You can install the software on your local machine or use the SSL VPN web interface (***recommended***)
  - Software, setup instructions and help available at access.llnl.gov/vpn_access
  - Windows, Mac and Linux software available, but no support for Linux. If problems - use web interface.
  - Authentication to VPN client is your OUN (smith25) and PIN + OTP passcode

Access

- SSH to LLNL machine after VPN is running on your local machine
  - ssh hera.llnl.gov ssh ubgl.llnl.gov
  - login with your LC machine name and PIN + OTP passcode



- Some internal web pages require similar authentication
- Automatic HPSS archival storage account
  - "ftp storage" from an LLNL machine
- File transfers
  - Use scp to/from LLNL machine
  - ftp initiated from within LLNL may also work (depends upon your local site permitting outside ftp connections)

Software Environment

- LC's Development Environment Group (DEG) – comprised of @ 15 staff responsible for the installation and support of software needed by most code development teams:
  - Compilers (computing.llnl.gov/code/compilers.html)
  - Debuggers
  - MPI libraries
  - Profiling, tracing and performance analysis tools
  - Parallel I/O libraries
  - Misc. utilities, some math libraries
  - See computing.llnl.gov/?set=code&page=software_tools

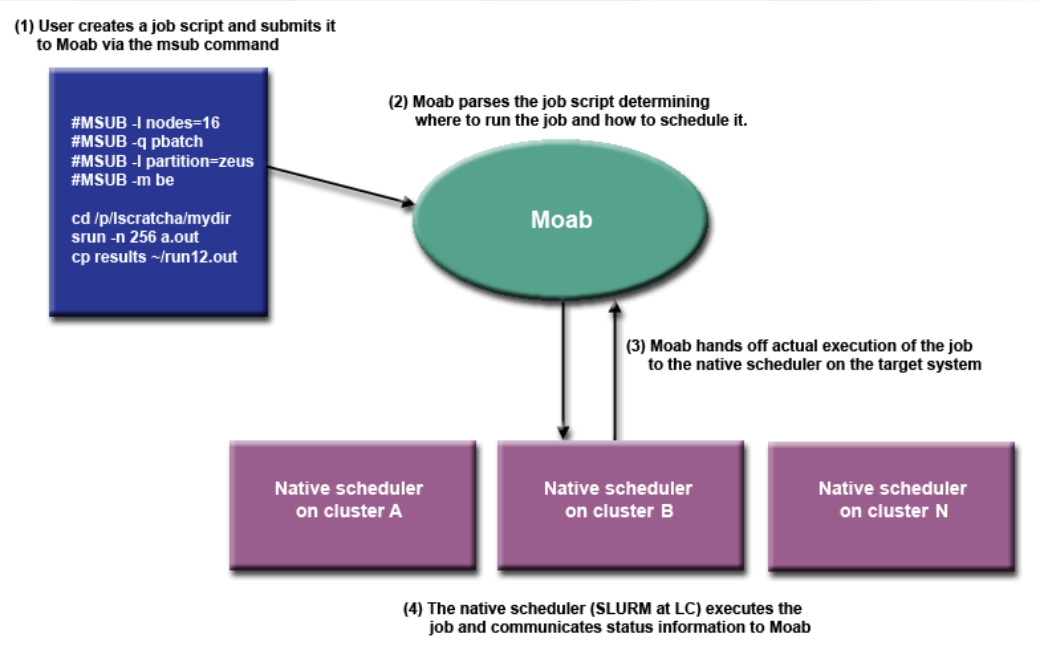
- dotkit
  - Many of LC's software packages are managed by "dotkit".
  - To see packages: `use -l`
  - To load a package: `use packagename`
  - More info: computing.llnl.gov/?set=jobs&page=dotkit

Software Environment

- CHAOS operating system on all Linux clusters
  - Based on Red Hat but differs in the following areas:
    - Modified kernel - to support high performance hardware, the Lustre file system, and other Livermore requirements.
    - New packages - support added for cluster monitoring, system installation, power/console management, parallel job launch, resource management, compilers, etc.
    - Modified packages - a few Red Hat packages are modified to implement timely bug fixes or enable them to work with CHAOS packages.
  - More info: computing.llnl.gov/linux/projects.html#chaos
- /usr/gapps
  - File space provided to code development teams that permits group/world sharing
  - Managed/supported by the individual code teams
- Visualization group – provides support for viz software and hardware, consulting, other viz related services
  - More info: computing.llnl.gov/vis/

Software Environment

- SLURM and Moab Batch System
  - SLURM - Simple Linux Utility for Resource Management. Native resource manager, and instance of which runs on each LC cluster. More info: computing.llnl.gov/linux/slurm/documentation.html
  - Moab - Workload Manager - top level batch scheduler that "sits on top" of multiple clusters managed by their local SLURM resource manager. More info: computing.llnl.gov/tutorials/moab



File Systems Summary

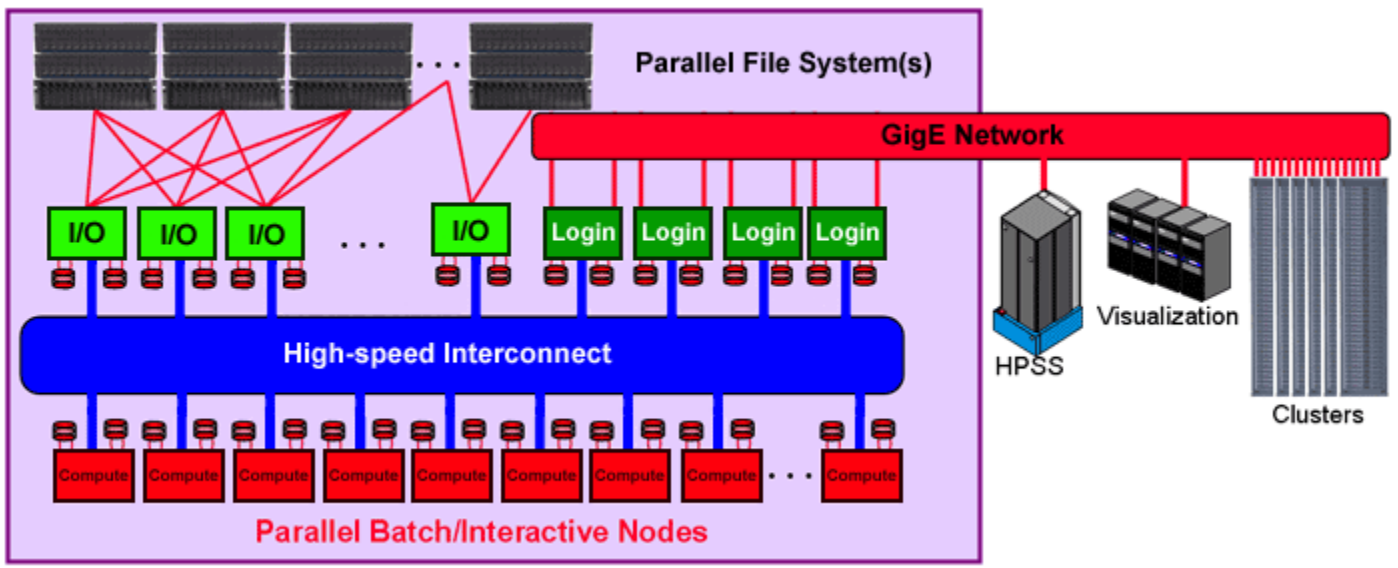
File System	Description	Backup?	Purge?
Home directory plus special `.snapshot` directory	NFS mounted from all machines 16GB quota per user. Online backups past 2 days: cd to `.snapshot`	YES	NO
/usr/tmp /var/tmp /tmp	Temporary file space local to a single node Varies in size – multiple GB	NO	YES
/nfs/tmp2	Shared temporary file space across all platforms 134 TB; User quota in effect (400 GB)	NO	YES
Parallel file systems /p/lscratch{a,b,c,d}	Very large parallel file systems specifically for parallel I/O. No quota. Multiple file systems in 700TB - 1.1PB range	NO	YES
HPSS Storage	Virtually unlimited; accessible from all clusters	NO	NO

Running Jobs

- Types of nodes
  - Login nodes: where you end up after you ssh to hera or ubgl
    - Shared by multiple users
    - Primarily used for interactive work such as editing files, submitting batch jobs, compiling, running GUIs, etc.
    - Interactive use exclusively - login only nodes do not permit any batch jobs.
    - DO NOT run production jobs on login nodes! It will impact other users.
  - Debug (pdebug) nodes
    - Intended for short running, small, interactive jobs
    - Not shared - owned exclusively by the running job
    - Can be logged into while your job is running
  - Batch (pbatch) nodes
    - Intended for all production work
    - Comprise the majority of all nodes on a given cluster
    - Not shared - owned exclusively by the running job
    - Can be logged into while your job is running - useful for debugging
    - Jobs must be submitted via Moab
  - Service and I/O nodes: not available directly to users

Running Jobs

- Typical LC machine configuration



- Most jobs require submission to Moab via a job control script. See computing.llnl.gov/tutorials/moab for details. Example on next slide.
- Interactive (pdebug) jobs can be run without Moab on hera. Example:

```
srun -n64 -ppdebug a.out
```

- Job limits: news job.lim.hera news job.lim.ubgl

Example Moab Job Script

```
#!/bin/csh
##### These lines are for Moab
#MSUB -l nodes=16
#MSUB -l partition=hera
#MSUB -l walltime=2:00:00
#MSUB -q pbatch
#MSUB -m be
#MSUB -V
#MSUB -o p/lscratchb/joeuser/par_solve/myjob.out

##### These are run-time shell commands
date
cd /p/lscratchb/joeuser/par_solve

##### Note that srun command is required to
##### launch parallel jobs
srun -n 256 a.out
echo 'Done‘
```

See the LC Moab tutorial for details: computing.llnl.gov/tutorials/moab

Summary of Batch System Commands

Command	Description	Source	Scope
msub	Submit a job control script to the batch system. Many options used to specify job requirements and behavior.	Moab	Job
srun	Launch a parallel job from within a Moab job script or interactively	SLURM	Job
mshow	Display running, idle and blocked jobs	Moab	Grid
showq	Display running, idle and blocked jobs	Moab	Grid
mjstat	Display queue summary and running jobs	LC	Cluster
squeue	Display running jobs. Numerous options.	SLURM	Cluster
checkjob	Display detailed information about a single job	Moab	Job
mdiag -j	Display running, idle and blocked jobs	Moab	Grid
sacct -j	Display information about a running job, including multiple job steps	SLURM	Job
sinfo	Display a concise summary of queues and running jobs	SLURM	Cluster
ju	Display a concise summary of queues and running jobs	LC	Cluster
mjobctl -h	Place a queued job on user hold	Moab	Job
mjobctl -u	Release a user held job	Moab	Job
mjobctl -c	Cancel a running or queued job	Moab	Job
canceljob	Cancel a running or queued job	Moab	Job
scancel	Cancel a running or queued job	SLURM	Job
mjobctl -m	Change a job's parameters. Multiple options.	Moab	Job
mshare	Display bank/account allocations, usage statistics, and priorities.	Moab	Grid
mdiag -u	Display a user's bank/account information	Moab	Grid
sreport	Report usage information for a cluster, bank, individual, date range, and more.	SLURM	Grid
mdiag -f	Display fair-share scheduler statistics	Moab	Grid
showstart	Guesstimate when a job will start	Moab	Job
showbf	Show what batch resources are immediately available	Moab	Grid
showstate	Display a text map of jobs and which nodes they are running on	Moab	Grid
smap	Display a text map of jobs and which nodes they are running on	SLURM	Cluster

Queues and DATs

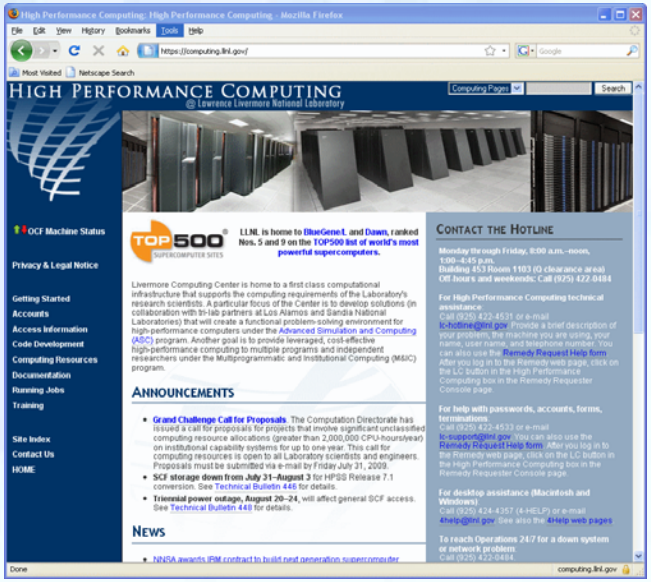
- Queues
  - Each machine has its own unique queue structure
  - Specific limits based on job size, duration, and interactive vs. batch
  - To view queue information, use the command:
    `news job.lim.`*`machinename`*
  - Moab batch scheduler across clusters; SLURM is the resource manager local to each cluster.
  - "Fair share" algorithm employed for resource delivery
    - Use more than your allocation and your service decreases
    - Use less, and it increases
    - Available cycles don't go to waste
- DATs
  - Dedicated Application Time
  - Scheduled runs that exceed normal queue limits
  - Usually allocated on weekends
  - To request: computing.llnl.gov/?set=forms&page=ASC_dat_form
  - Coordinated through the Livermore Computing (LC) Hotline

Support, Documentation & Training

- Livermore Computing Hotline
  - Issues concerning use of LC's supercomputers
  - Email, phone (925-422-4531) and walk-in during normal business hours: Mon-Fri, 8:00 am–noon, 1:00–4:45 pm Pacific time.
  - Accounts/passwords: lc-support@llnl.gov
  - Technical questions: lc-hotline@llnl.gov
  - Referral to subject experts and vendors
  - Problem tracking/database system
  - Operations support during off-hours
- Machine status email lists, technical bulletins
- 4HELP Institutional Help Desk
  - Desktop issues, VPN, network, OTP tokens
  - 4help@llnl.gov 925-424-4357 (4-HELP)

Support, Documentation & Training

- Documentation: computing.llnl.gov
  - Most of what users need to know is available online via LC web pages Recommended starting points:
    - computing.llnl.gov
- Training: computing.llnl.gov/training
  - Online tutorials
  - Local workshops, seminars
  - Remote workshops
- Recommended for new (and even old) users:
  - computing.llnl.gov/tutorials/lc_resources
  - computing.llnl.gov/tutorials/moab
  - computing.llnl.gov/tutorials/linux_clusters

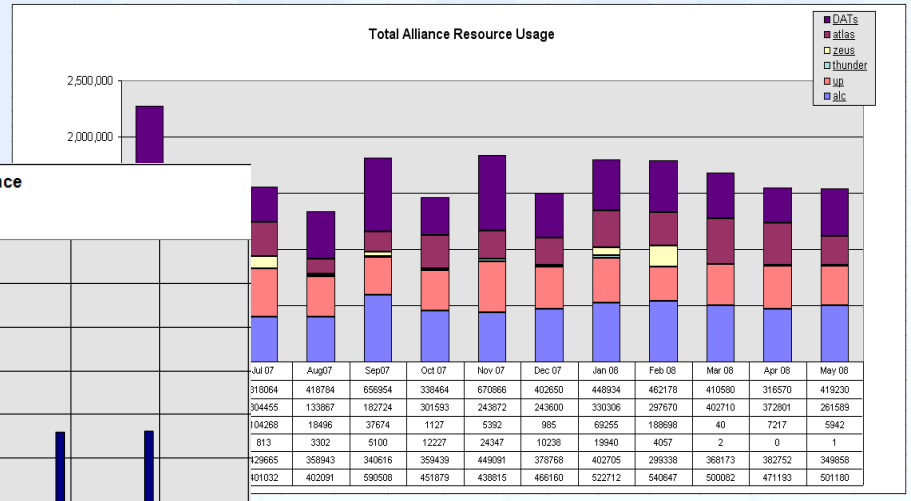


Alliance Communications

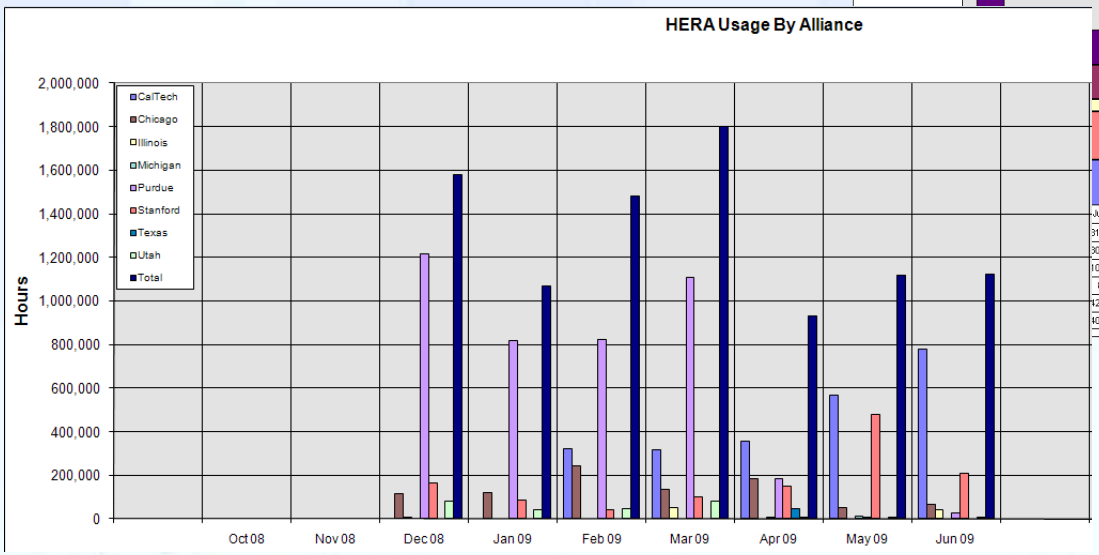
- Monthly telecons and email list (asap-crt@lanl.gov)
  - Toll-free number: 866-914-3976 code: 187522#
  - Active participation by all 8 Alliances, LLNL and LANL
  - Forum for discussion/questions on user topics such as accounts, access, technical issues, machine schedules, etc.
  - Minutes and usage stats are distributed via our email list to all Alliances, ASC HQ and various staff & managers within the Labs
  - Let us know if you want anyone else at your Center added to our list - initially it includes only your POC and PI
- LC Hotline
  - User support as described previously
- Your LLNL POC
  - Blaise Barney blaiseb@llnl.gov 925-422-2578

Usage Stats

- Sent monthly with the telecon minutes
  - By Lab, machine, user



Jul 07	Aug07	Sep07	Oct 07	Nov 07	Dec 07	Jan 08	Feb 08	Mar 08	Apr 08	May 08
318064	418784	656954	338464	670866	402650	448934	462178	410580	316570	419230
304455	133867	182724	301593	243872	243600	330306	297670	402710	372801	261589
104268	18496	37674	1127	5392	985	69255	188698	40	7217	5942
813	3302	5100	12227	24347	10238	19940	4057	2	0	1
429665	358943	340616	359439	449091	378768	402705	299338	368173	382752	349858
401032	402091	590508	451879	438815	466160	522712	540647	500082	471193	501180



The Future

- Sequoia procurement
  - Awarded to IBM in Feb 2009. See computing.llnl.gov/tutorials/bgp/pressReleaseSequoiaLLNL.pdf
  - 20 Petaflop system
    - 98,304 compute nodes
    - 1.6 million processors
    - 1.6 petabytes of memory
  - 0.5 petaflop early delivery BG/P system (Dawn)
  - @ $215 million award over a 2008 - 2015 time span
  - Too early to say much more right now

SC09 NNSA/ASC Research Exhibit Booth

- Opportunity to showcase Alliance research
- 3D theater, demo stations, collaboration area, viz clips…
- Sign up at www.lanl.gov/conferences/ascsc09/pdfs/sc09participationform.pdf

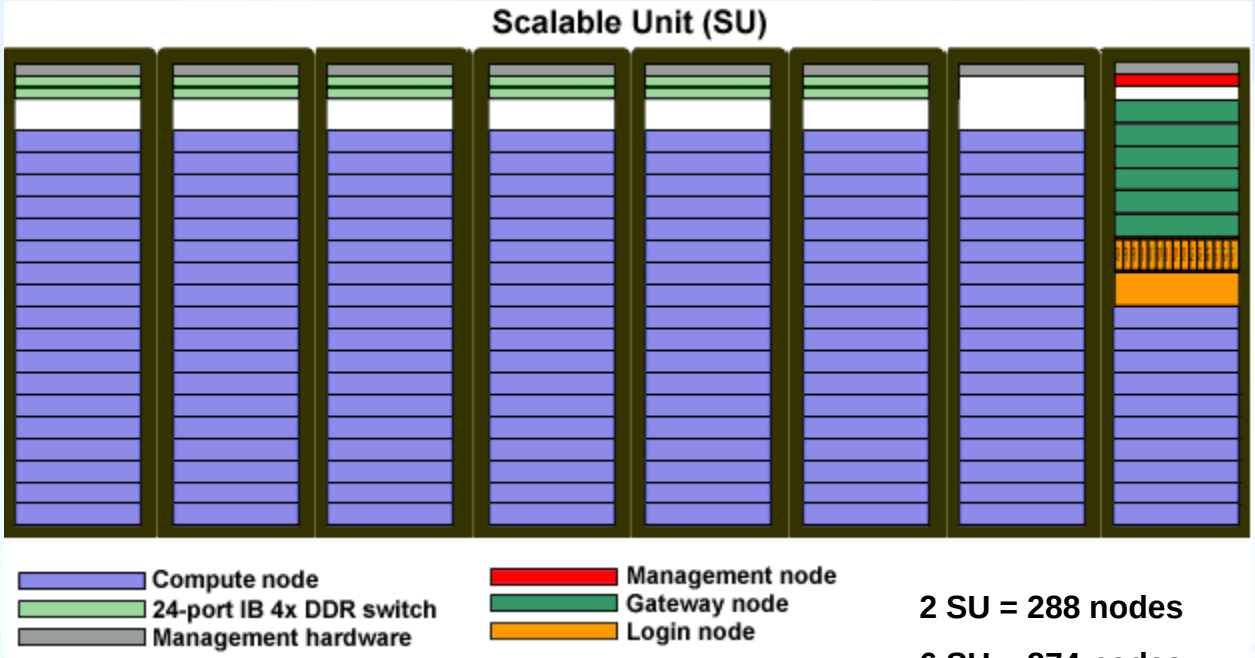
Questions?

Using LLNL's Hera System

computing.llnl.gov/tutorials/linux_clusters
computing.llnl.gov/tutorials/moab
computing.llnl.gov/tutorials/lc_resources

TLCC Hardware Overview

- Hera is one of the Tri-labs' (LLNL, LANL, Sandia) TLCC systems
- TLCC = Tri-lab Linux Capacity Cluster
  - Based on the "Scalable Unit" (SU) design
  - Quad-core, quad-socket AMD Opteron with Infiniband interconnect
  - Leveraged Tri-lab procurement to provide a common, standardized, scalable architecture and reduce costs
  - Includes a common OS and software stack



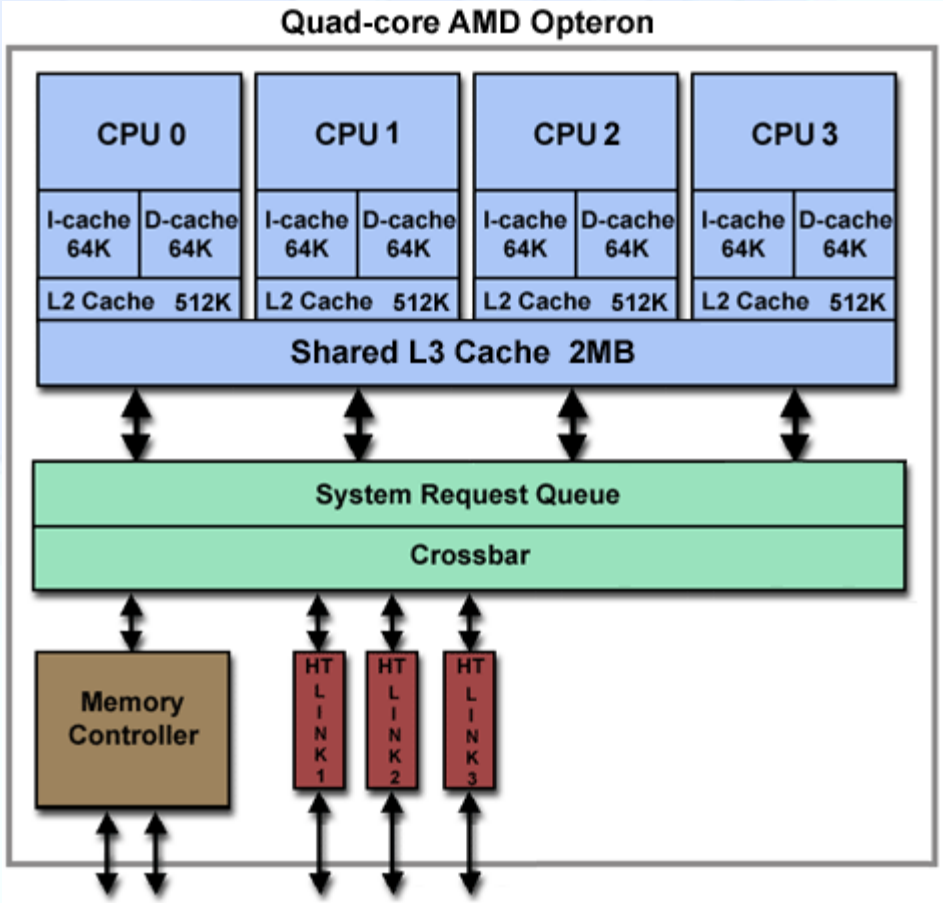
- 144 nodes (136 compute, 1 login, 1 management, 6 gateway)
- Twelve 24-port Infiniband 4X DDR switches
- Miscellaneous management hardware
- Frames sufficient to house all of the hardware
- Additionally, one 288-port second level switch is also needed for every 2 SUs in a multi-SU cluster (not shown)

2 SU = 288 nodes **4 SU = 576 nodes**

6 SU = 874 nodes **8 SU = 1152 nodes**

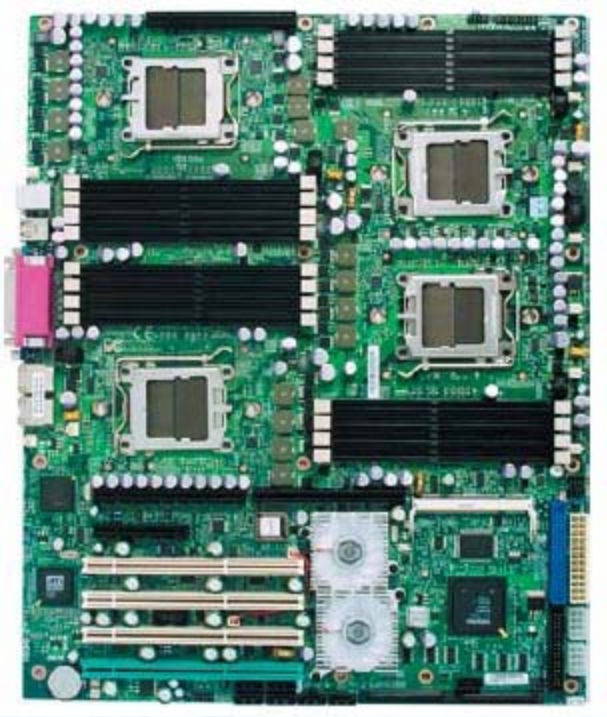
TLCC Hardware Overview

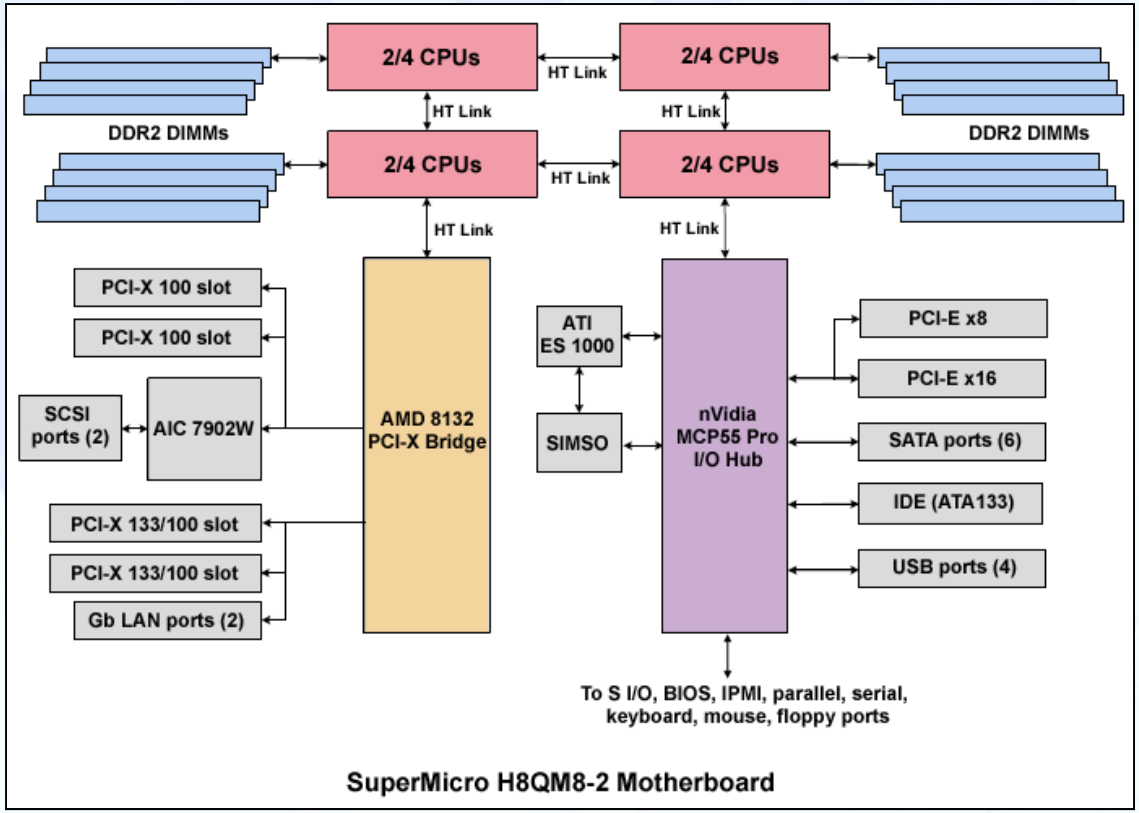
- AMD Opteron CPU @ 2.2 – 2.3 GHz
- 64-bit, “little endian”
- 64 KB L1 data cache
- 64 KB L1 instruction cache
- 512 KB dedicated L2 cache
- 2 MB shared L3 cache
- On chip memory controller @ 10.7 GB/s
- 3 HyperTransport links to I/O, other chipsets, off-chip CPUs @ 8 GB/s per link
- Enhanced branch prediction capabilities
- Speculative / out-of-order execution
- Superscalar
- Pipelined architecture with full support for Intel’s SIMD vectorization instructions (SSE, SSE2, SSE3...)



TLCC Hardware Overview

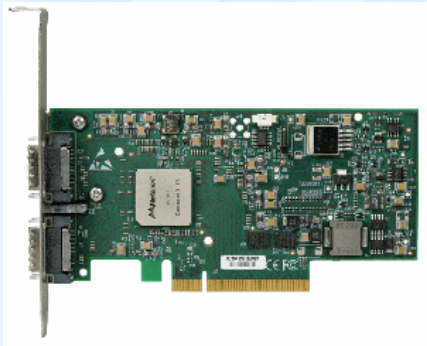
SuperMicro H8QM8-2 motherboard





TLCC Hardware Overview

- Infiniband Host Channel Adapter (HCA) Card:
  - One per node; provides the interface between a node and a two-stage network
  - Copper cabling connects each adapter card to a 24-port first stage switch
  - Mellanox ConnectX IBA 4x DDR (double data rate) PCIe 8x

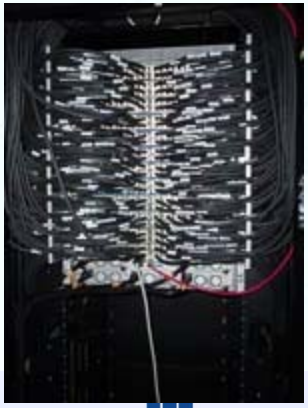


Mellanox HCA

- 24-port Infiniband Switch: First stage
  - 12 ports connect to HCA adapters in nodes and 12 ports connect to a second stage switch
  - 20 Gbits/sec per port
  - 21" x 17" x 1U, rack mounted

Six 24-port Voltaire switches

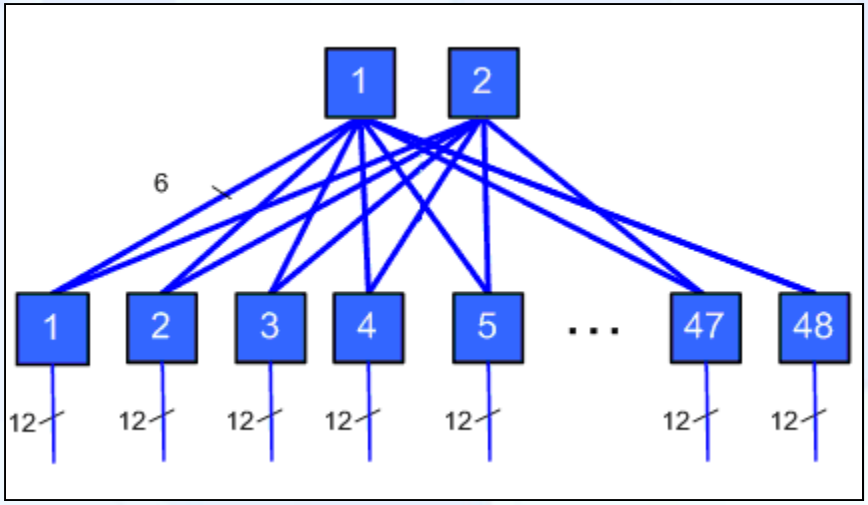
- 288-port Infiniband Switch: Second stage
  - All ports connect to a first stage switch
  - 20 Gbits/sec per port
  - 26" x 18" x 23"; holds up to 12 line boards (24 ports each)



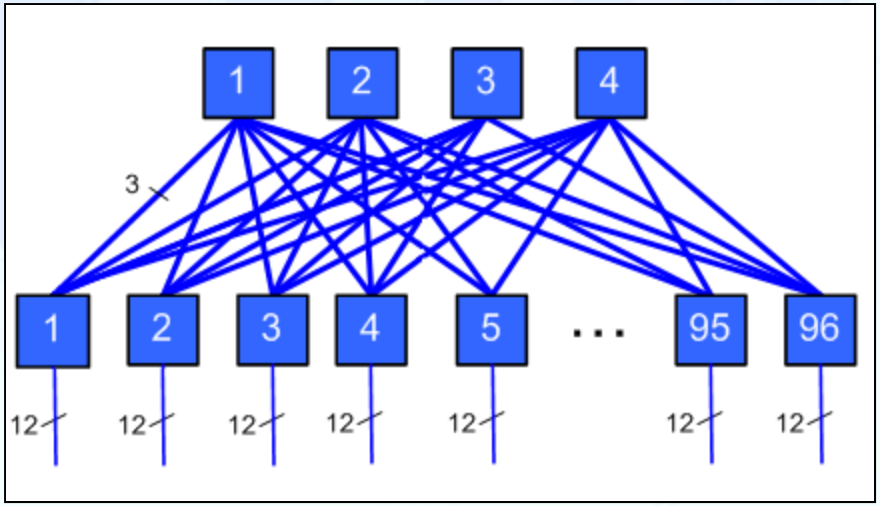
288-port Voltaire switch – front & back

TLCC Hardware Overview

- Infiniband switch topology:
  - Two-stage, federated, bidirectional, fat-tree
  - The number of 288-port, second stage switches depends upon the number of scalable units (SUs) that comprise the cluster. One second stage switch is used for each 2 SUs.



4 SU – 576 nodes configuration



8 SU – 1152 nodes configuration

- Actual performance - on heavily loaded LLNL clusters
  - MPI Latency: ~2-3 us
  - MPI Bandwidth: ~2.4 GB/sec

LLNL's Hera TLCC Cluster

- Hera Configuration
  - AMD quad-core quad-socket Opteron @2.3 GHz
  - 127.2 TFlop system
  - 864 nodes; 16 cpus/node; 13,824 cpus total
  - 32 GB memory per node
  - 4X DDR Infiniband
    - 20 Gbits/sec rating
    - 2 - 2.5 Gbytes/sec non-block pt2pt MPI in practice
  - 64-bit architecture; CHAOS OS
  - Lustre parallel I/O file systems

- Usage
  - ASC and Institutional resource
  - Alliance allocation @20% of the machine
  - DAT runs

TLCC Software Summary

- CHAOS operating system
  - Red Hat modified for/by Tri-labs
- Moab workload scheduler (batch)
- Compilers - C/C++/Fortran
  - Intel
  - PGI
  - Pathscale
  - GNU
- Math libraries
  - Intel's MKL
  - AMD's ACML
  - Other available libs vary by lab
- MPI
  - MVAPICH at LLNL
  - OpenMPI at LANL
- Debuggers
  - TotalView is the preferred debugger
  - Others are available (varies between labs)
- Tools – vary between labs
  - TAU common to Tri-labs

TLCC Compilers

Compiler		Serial Command	Parallel Command
Intel	C	icc	mpiicc
	C++	icpc	mpiicpc
	Fortran	ifort	mpiifort
PathScale	C	pathcc	mpipathcc
	C++	pathCC	mpipathCC
	Fortran	pathf90	mpipathf90
PGI	C	pgcc	mpipgcc
	C++	pgCC	mpipgCC
	Fortran	pgf77 pgf90	mpipgf77 mpipgf90
GNU	C	gcc	mpicc mpigcc
	C++	g++	mpiCC mpig++
	Fortran	g77 gfortran	mpif77 mpif90 mpigfortran

- Available compilers are the same at the Tri-labs, but versions may vary
- Many options for each compiler – consult the vendor documentation
- For the parallel compiler commands, the underlying MPI library varies between the Tri-labs
- All parallel compiler commands are scripts that call the base vendor compiler

Running Jobs on LLNL's Hera Cluster

- As with most of LLNL's clusters, there is a small interactive "pdebug" queue and a large batch "pbatch" ("p" stands for "parallel")

```
hera552% mjstat
Scheduling pool data:
-------------------------------------------------------------
Pool         Memory  Cpus  Total Usable   Free  Other Traits
-------------------------------------------------------------
pbatch*    30000Mb    16    796    794      2
pdebug     30000Mb    16     16     16     13
```

- pdebug
  - Small, short jobs, testing, debugging, prototyping, non-production
  - Launched with the "srun" command from your shell command line:

    `srun -n32 -ppdebug a.out args`

    `srun -N4 -ppdebug a.out args`
- pbatch
  - Small to large jobs
  - Short to long time
  - Production runs

Running Jobs on LLNL's Hera Cluster

- Job limits can be viewed by the command
  `news job.lim.hera`
- Limits differ by queue
- Limits also differ between prime and non-prime time
- It is particularly important that your batch jobs choose an appropriate time/node limit. They could sit in the idle queue forever otherwise...

```
================================= job.lim.hera =================================
   SUMMARY OF INTERACTIVE AND BATCH JOB LIMITS ON HERA
   ---------------------------------------------------------------------------
   There are 816 compute nodes, all of which have AMD Opteron Quad Core
   processors.
   There are 2 node scheduling pools:
   pdebug -  256 cores (16 nodes), only interactive use permitted.
   pbatch - 12,800 cores (800 nodes), only batch use permitted.
   Atlas limits:
                           Max nodes/job   Max runtime
   -------------------------------------------------------
   pdebug Prime Time(*)          16           30 minutes
   pdebug Non-prime Time(*)      16          120 minutes
   pbatch - weekdays             256           16 hours
   pbatch - weekends             256           24 hours
   -------------------------------------------------------
   (*) Prime time is from 6 AM to 6 PM weekdays. The non-prime time limit
        of 120 minutes starts on 2/26/07.

   Weekend pbatch limits start on Friday at 1800. They are adjusted back
   starting Sunday evening and are set to the weekday limit at 0800 Monday.
   Weekend limits on long weekends start at 1800 the evening before the long
   weekend and end at 0800 the morning after the long weekend.
   Weekend limits on a holiday in the middle of a week start at 1800 the day
   before the holiday and end at 0800 the morning after the holiday.
   Please limit standby jobs to 12 hours maximum.
   Please limit the use of pdebug to the above limits (16 nodes prime time,
   32 nodes otherwise) on a PER USER basis, not a PER JOB basis, to allow
   other users access.  Using more than the posted limit PER USER can
   result in job removal without notice.
   HARDWARE:
   Nodes are AMD Opteron Quad Core with 16 cores (2.3 GHz) & 32 GiB memory
   SCHEDULING:
   Hera jobs are scheduled through Moab.
```

Running Jobs on LLNL's Hera Cluster

Now we'll switch to the Moab tutorial....

computing.llnl.gov/tutorials/moab