

BLUE WATERS

SUSTAINED PETASCALE COMPUTING

4/15/13

Blue Waters User Monthly Teleconference



GREAT LAKES CONSORTIUM
FOR PETASCALE COMPUTATION

CRAY®

Agenda

Topics for this teleconference call:

- Current System Usage
- XK Upgrade
- Resource Management and Job Scheduling
- Storage
- Recommendations for application resiliency
- Request for publications, presentations, ...
- Upcoming Events
- Topics for future Blue Waters Users telecons/webinars.
- Open Discussion

System Usage

- Currently, Blue Waters is showing good job turn-around time.
- We encourage teams to avoid waiting to use allocation.
- We prefer teams to even out usage over an allocation year.
- We are looking into several options to avoid end of allocation period congestion.
- We have not completed our assessment but we have discussed something like this:
 - End of Q1, if more than 90% of the allocation is left, then 5% is returned.
 - End of Q2, if more than 60% of the allocation is left, then 25% is returned.
 - End of Q3, if more than 35% of the allocation is left, then 60% is returned.
- Let us know if you expect to have substantial, clustered or “burst” usage.

XK Upgrade

- Additional 12 cabinets of XK blades
 - Increase to 4,224 XK nodes
- Increase of overall torus dimension to 24x24x24
- Change in “shape” of XK region of torus
 - Better aggregate bandwidth
- Timeline
- Impact on access

Resource Management and Job Scheduling

- Upgrade to Torque 4.x and Moab 7.x
 - Better scalability, improved HA configuration, numerous bug fixes
 - Resolves pressing issues: hostlist
- Job scheduling priorities (not fully implemented)

QOS	Priority (1 highest)	Charge Factor	Preemptable
Debug/Interactive	1	1	No
High	2	2	No
Regular	3	1	No
Low	4	.5	Yes

- Running jobs limit: 20 per user and 100 per group
- Jobs that accrue priority limit: 20 per user
- Changes coming to the daily schedule
 - Limit draining due to full system jobs
 - Adjusting policies to improve transition to overnight
- Avoid excessive calls to monitor queue status (watch -n ...)

Storage Update

- HPSS
 - Nearline system up and running
 - Files from Nearline_gap being transferred by Storage Group
 - Moab generic resource hpss to denote jobs doing file staging/transfers.

- Quotas

- Online
- Nearline

Filesystem	Size (PB)	OSTs	Quota Policy	Quota Size (TB)
HOME	2.2	144	user	1
SCRATCH	22.0	1440	group	500
PROJECTS	2.2	144	group	5

- Structure of HOME and PROJECTS same as above

- Purging

- 2 week file aging
- For quota issues contact:
help+bw@ncsa.illinois.edu

System	Quota Policy	Quota Size (TB)
HOME	user	5
PROJECTS	group	50

Application Resiliency Suggestions

- Create job scripts that anticipate a node failure
 - Ask for a few extra nodes.
 - Restart from the last checkpoint file.
 - Use the aprun relaunch option “-R int” where int is the number of failed PEs the application can tolerate.
- Checkpoint/Restart considerations
 - Mean time between compute node failures
 - $MTBF \sim 8 \times (25712/nodecnt)$ hours.
 - Need to factor in time for checkpoint.
 - Other factors that cause job interruption

Request for publications, presentations, ...

- We would like keep current on products that result from time on Blue Waters.
- Appreciate updates sooner than annual reports.
- NSF PRAC teams to send information to PoCs.
- See the [Share Results](#) section of the portal.

Upcoming Events

- Next Blue Waters User webinar May 20th
- Blue Waters PRAC Symposium May 21 – 22
- New User Workshop May 22 – 23.
- See the [Workshops](#) portal page for more information.

- Topics for future teleconferences / webinars
 - XK / Kepler features and tuning
 - Advanced topics on running jobs and aprun
 - VisIt
- Open Discussion