

BLUE WATERS

SUSTAINED PETASCALE COMPUTING

8/14/16

Blue Waters User Monthly Teleconference



GREAT LAKES CONSORTIUM
FOR PETASCALE COMPUTATION

CRAY®

Agenda

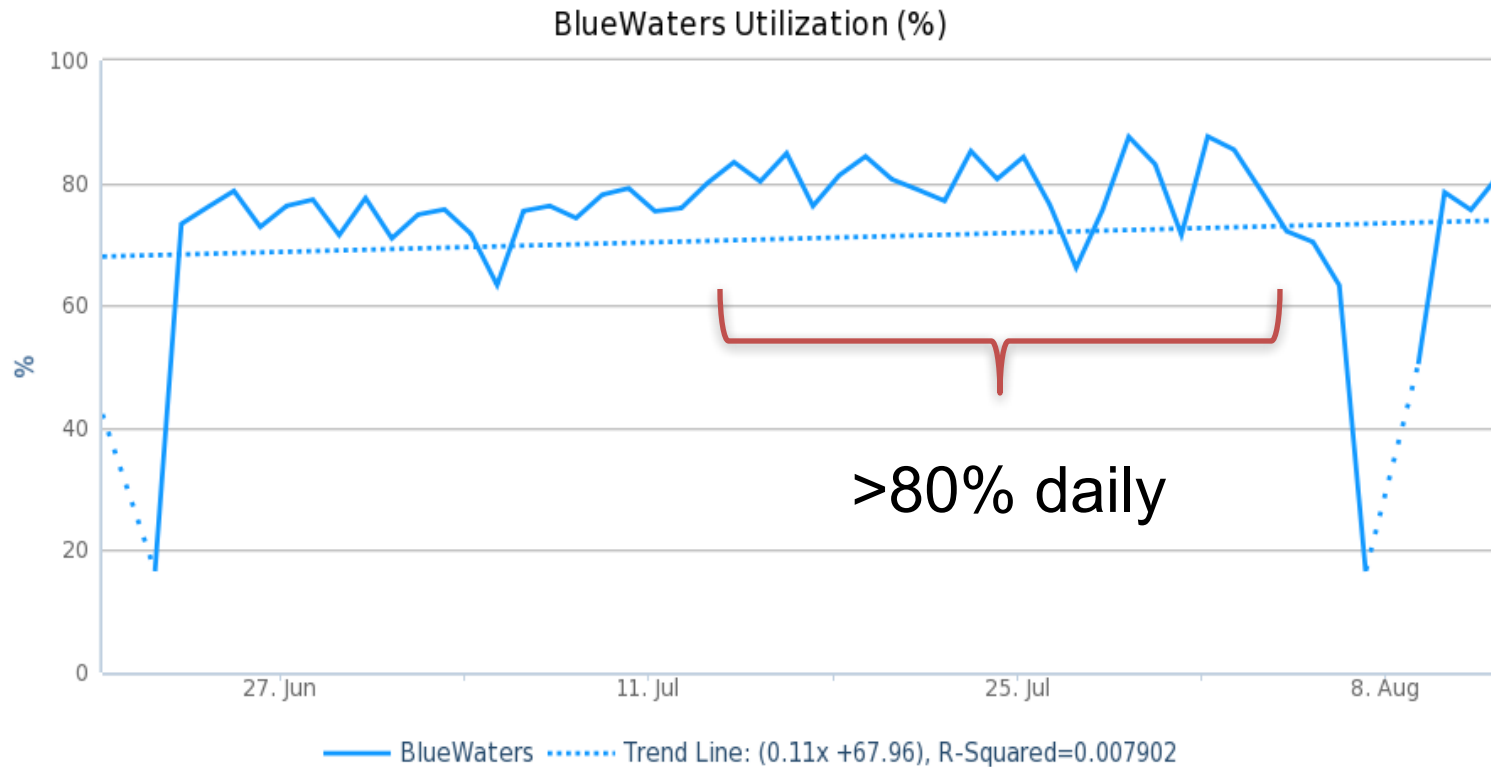
- Symposium
- File System Upgrade Status
- Maintenance Changes
- Utilization
- Recent Events
- Opportunities
- Workflow Workshop follow-up
- PUBLICATIONS!

File System Upgrade Status

- Home & Projects file system upgrade complete. Final sync during recent maintenance.
 - soft and hard quotas to be enforced this week for home; quotas for projects to be enforced next week.
- Scratch moved to upgraded ½ file system during recent maintenance.
 - A few users had to be resynced.
 - Vendor on site this week to upgrade other ½ of scratch file system.
- Final service interruption to merge two halves of scratch.

System Utilization

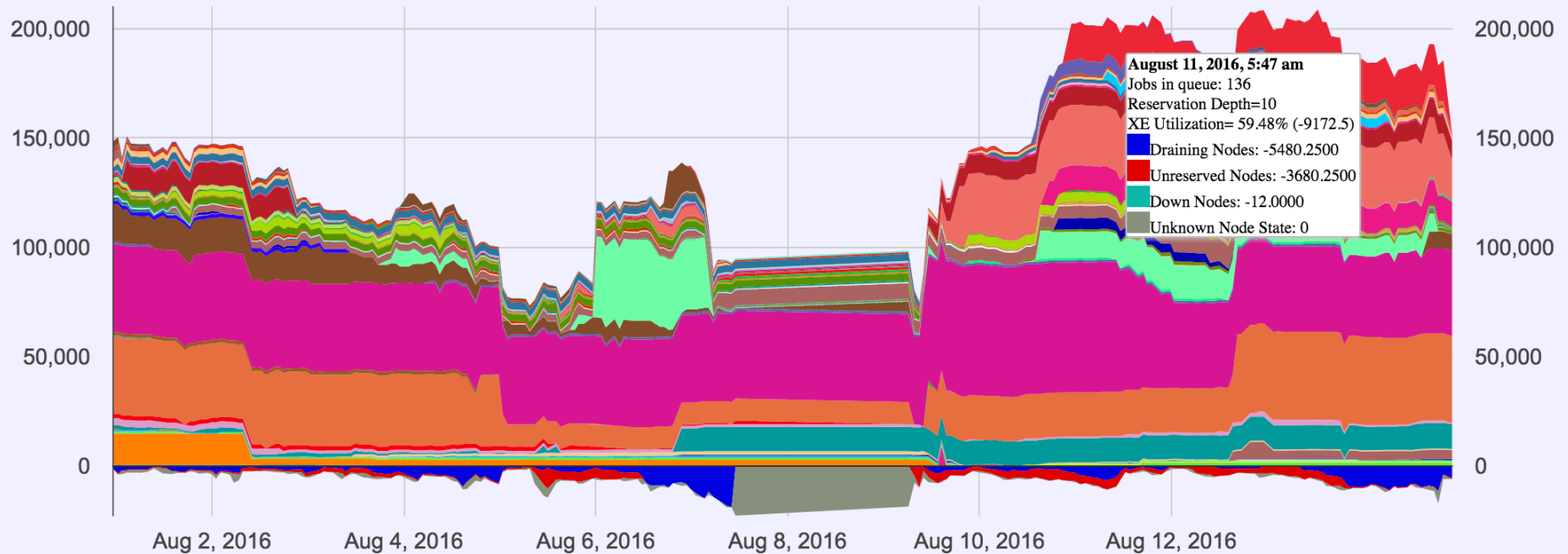
- Utilization since last BW User Call (June 20)



Workload backlog

Blue Waters xe Frontlog/Backlog

■ Draining Nodes ■ Unreserved Nodes ■ Down Nodes ■ Unknown State ■ aadelson ■ ackerman ◀ 1/26 ▶



- Vertical axis in units of nodes.
- Plenty of jobs and varied node-hour mix

Why isn't my job running

- Check [system status page](#) for utilization.
- Check backfill at above url or on system
 - `showbf -p bwsched -f xe`
- Check top jobs
 - `showq -i`
 - Ordered by priority. Jobs with * have a reservation on nodes.
- Check start times of jobs with reservations using `showres`.

```
> showbf -f xe -p bwsched
```

Partition	Tasks	Nodes	Duration	StartOffset	StartDate	Geometry
bwsched	12032	376	2:26:17	00:00:00	14:21:45_08/14	4x6x8
bwsched	7168	224	4:09:44	00:00:00	14:21:45_08/14	8x2x7
bwsched	1536	48	5:56:17	00:00:00	14:21:45_08/14	3x2x4
bwsched	1536	48	6:01:17	00:00:00	14:21:45_08/14	3x2x4
bwsched	1024	32	6:06:17	00:00:00	14:21:45_08/14	1x2x8
bwsched	640	20	INFINITY	00:00:00	14:21:45_08/14	1x2x5

```
> showq -i | grep -v xk | grep \*
```

Job	Tasks	Nodes	Priority	to	User	CPUs	Start	Priority	State	Time	Time
5207970*	10321211	99.0	99.0	to	dtoussai	baea	4096	3:30:00	normal	Wed Jul 20	07:08:18
5276564*	10125901	99.0	99.0	to	yeung	jmo	16512	00:30:00	high	Wed Aug 10	20:59:04
5207973*	9587740	99.0	99.0	to	dtoussai	baea	4096	3:30:00	normal	Wed Jul 20	07:08:27
5275602*	7617711	99.0	99.0	to	guo2	jna	8000	00:30:00	high	Wed Aug 10	14:14:31
5271992*	7307965	99.0	99.0	to	pinelli	jno	9216	00:05:00	high	Tue Aug 9	13:44:27
5273294*	7200637	99.0	99.0	to	pinelli	jno	18432	00:05:00	high	Tue Aug 9	17:05:50
5246922*	5689508	99.0	99.0	to	dtoussai	baea	4096	3:30:00	normal	Sun Jul 31	01:51:04
5246923*	5642442	99.0	99.0	to	dtoussai	baea	4096	3:30:00	normal	Sun Jul 31	01:51:12
5269231*	5637764	4.9	4.9	to	clay1	jmo	8340	2:00:00:00	high	Sat Aug 6	19:06:02
5246924*	5600588	99.0	99.0	to	dtoussai	baea	4096	3:30:00	normal	Sun Jul 31	01:51:19

```
> showres 5207970 ...
```

Job	CPUs	Start	End	Start	End	Time	Time
5207970	Job I	00:33:05	4:03:05	3:30:00	4096/131072	Sun Aug 14	14:54:44
5276564	Job I	6:06:23	6:36:23	00:30:00	16512/528384	Sun Aug 14	20:28:02
5207973	Job I	2:26:23	5:56:23	3:30:00	4096/131072	Sun Aug 14	16:48:02
5275602	Job I	4:09:50	4:39:50	00:30:00	8000/256000	Sun Aug 14	18:31:29
5271992	Job I	5:56:23	6:01:23	00:05:00	9216/147456	Sun Aug 14	20:18:02
5273294	Job I	6:01:23	6:06:23	00:05:00	18432/147456	Sun Aug 14	20:23:02
5246922	Job I	6:36:23	10:06:23	3:30:00	4096/131072	Sun Aug 14	20:58:02
5246923	Job I	6:36:23	10:06:23	3:30:00	4096/131072	Sun Aug 14	20:58:02
5269231	Job I	10:06:22	2:10:06:22	2:00:00:00	8340/266880	Mon Aug 15	00:28:02
5246924	Job I	2:10:06:22	2:13:36:22	3:30:00	4096/131072	Wed Aug 17	00:28:02

Recent Events

- Last week – Login node instability.
- 8/5 – Cabinet cooling failure caused cabinet EPO. Later an extended metadata server failover took place.
- 7/23 – MDS issue. Failover needed.
- 7/21 – Known lustre issue, failover cleared issue but many nodes marked down.

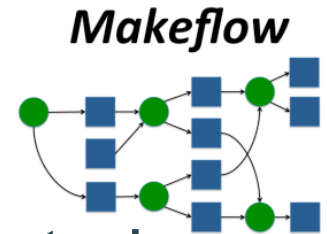
Changes

- Wrapper to qsub that looks for aprun. Emits warning if not found in submitted script.
- Re-enable XALT this week. Look for MOTD and portal blog entry.
- Portal login credential expiration increased to 1 week.
- Portal “Manage Users” to require additional login check.

Virtual School events

- Blue Waters Virtual Course Announcement
 - Introduction to High Performance Computing
 - Autumn 2016
 - Contact: Steve Gordon (sgordon@osc.edu)
 - <https://www.osc.edu/~sgordon/IntroHPC>
 - Instructor Dr. David E. Keyes

Workflow Workshop



- Over 180 registrants from 10 sites participated.
- Material available from [workflow website](#).
- Videos to be available in 2 to 3 weeks.
- We are looking to connect teams with appropriate workloads with workflow developers.



RADICAL


Copernicus



Review of Best Practices

- Improper use of login nodes
 - Use compute nodes for all production workloads.
- Avoid excessive calling of job scheduling commands
 - Unintentional denial of service may result otherwise.
- MOM node use should be limited to aprun launch.
 - All other commands can be run on compute nodes via aprun.
- Bundling of Jobs
 - Independent jobs bundled from 2 node to 32 nodes.
 - Avoid excessive, single nodes jobs.
 - Use a workflow.
- Small files usage
 - Use directory hierarchies, less than 10,000 files per directory.
 - Avoid many writers to same directory.
 - Tar up files before transferring to Nearline.

Request for Science Successes

- We need to be current on products that result from time on Blue Waters such as:
 - Publications, Preprints (e.g. [arXiv.org](https://arxiv.org) ), Presentations.
 - Very interested in data product sharing.
- Appreciate updates sooner than annual reports.
 - Send to gbauer@illinois.edu
- NSF PRAC teams send information to PoCs.
- See the [Share Results](#) section of the portal as well.
- **Be sure to include [proper acknowledgment](#)**
 - Blue Waters - National Science Foundation (ACI 1238993)
 - NSF PRAC – OCI award number