

OLCF Best Practices (and More)



Bill Renaud
OLCF User Assistance Group



Overview

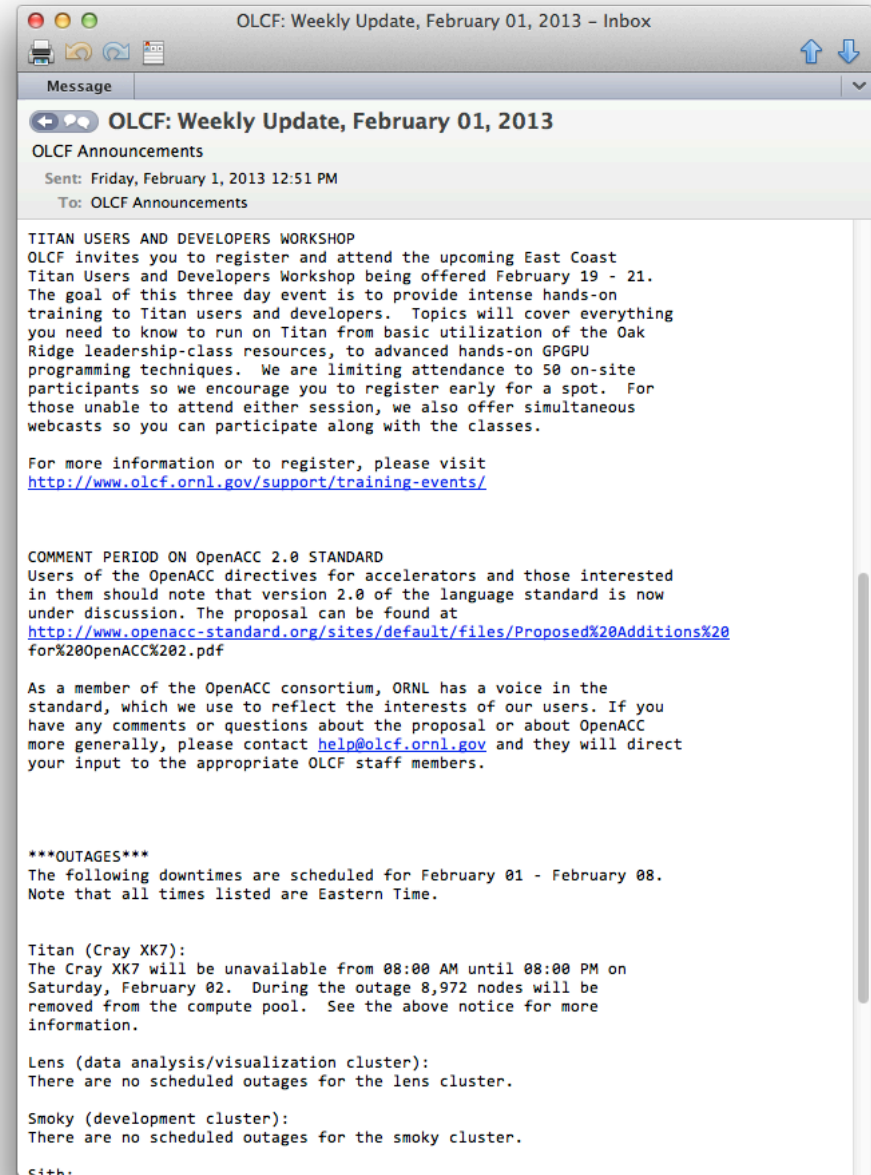
- This presentation covers some helpful information for users of OLCF
 - Staying informed
 - Some aspects of system usage that may differ from your past experience
 - Some common error messages
 - Common questions/Other tips on using the systems
- This is by no means an all-inclusive presentation
- Feel free to ask questions

Staying Informed

- OLCF provides multiple layers of user notifications about system status and downtimes
 - OLCF Weekly Update
 - OLCF Status Page (<https://users.nccs.gov/statuspages/summary>)
 - Status indicators on olcf.ornl.gov
 - Opt-in email lists
 - Smartphone Apps
 - Twitter (@OLCFStatus)
- A summary of these items can also be found at http://www.olcf.ornl.gov/kb_articles/communications-to-users/

Staying Informed-Weekly Update

- Sent weekly (Thursday/Friday)
- Contains several items
 - Announcements about upcoming training
 - Announcements about upcoming system changes
 - Planned outages for the next week
- **All OLCF users should receive this email**
 - Let us know (help@olcf.ornl.gov) if you're not receiving it!



Staying Informed-System Status

- Automated scripts parse logs from our monitoring software and make an educated guess as to system state
- This status is then sent to multiple destinations: websites, Twitter, smartphone apps, and email lists
- While this is fairly accurate, it is a fully automated process
 - Thus, there is a possibility of both false positives and false negatives
 - We do take some measures to mitigate this

System Status: www.olcf.ornl.gov

The image shows a screenshot of the Oak Ridge Leadership Computing Facility (OLCF) website. The main navigation bar includes links for HOME, ABOUT OLCF, LEADERSHIP SCIENCE, COMPUTING RESOURCES, CENTER PROJECTS, and MEDIA CENTER. A red box highlights the system status section at the top, which lists the following systems and their last update times:

System	Status	Last Update
jaguarpf	Down	8/21/12 08:05 am
jaguarpfext	Up	2/24/12 01:50 pm
hpss	Up	8/8/12 11:20 am
lens	Up	6/27/12 08:55 am
smoky	Up	8/15/12 09:30 am

The KnowledgeBase section on the left provides a search bar and a list of articles by resource. The Computing Resources section on the right features a large image of the OLCF building and a detailed description of the facility's capabilities. A red box highlights the SYSTEM STATUS section on the right, which provides a more detailed view of the system status:

System	Status	Last Update
JAGUARPF	Down	since August 21 08:05 AM
JAGUARPFEXT	Up	since February 24 01:50 PM
HPSS	Up	since August 8 11:20 AM
LENS	Up	since June 27 08:55 AM
SMOKY	Up	since August 15 09:30 AM

The SYSTEM STATUS section also includes a 'GET SUPPORT' link and a 'Help' button. The 'GET SUPPORT' section provides contact information for OLCF, including a phone number (865) 241-6536 and an email address (help@olcf.ornl.gov).

System Status: users.nccs.gov

<https://users.nccs.gov/statuspages/summary>

Project Pages – National Center for Computational Sciences (NCCS)

users.nccs.gov/statuspages/summary

NCCS.GOV
NATIONAL CENTER FOR COMPUTATIONAL SCIENCES

Support | Contact Us

Search

Home You are here: [Home](#) » [Summary](#)

NCCS System Status

Summary

Jaguar Lens Smoky Frost HPSS Spider

Site Summary

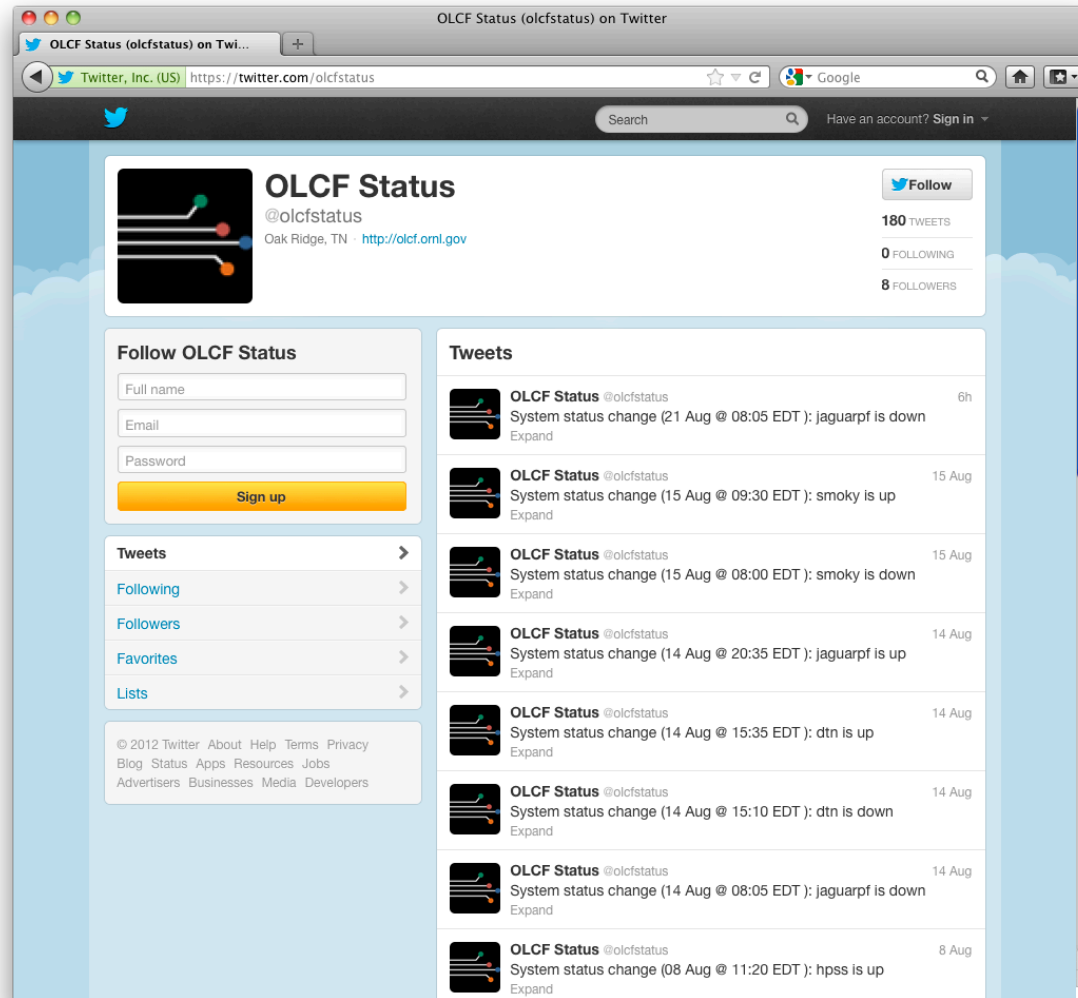
Page Last Updated:
21 Aug 2012 at 13:25 EDT

Current Status

System	State	Last State Change	Comments
Jaguar	⬇️	21 Aug 2012 08:05 EDT	Jaguarpf is currently down.
	⬆️	24 Feb 2012 13:50 EST	Jaguarpf-ext is currently up.
Lens	⬆️	27 Jun 2012 08:55 EDT	Lens is currently up.
Smoky	⬆️	15 Aug 2012 09:30 EDT	Smoky is currently up.
Frost	⬆️	20 Jun 2012 07:00 EDT	Frost is currently up.
HPSS	⬆️	08 Aug 2012 11:20 EDT	HPSS is currently up.

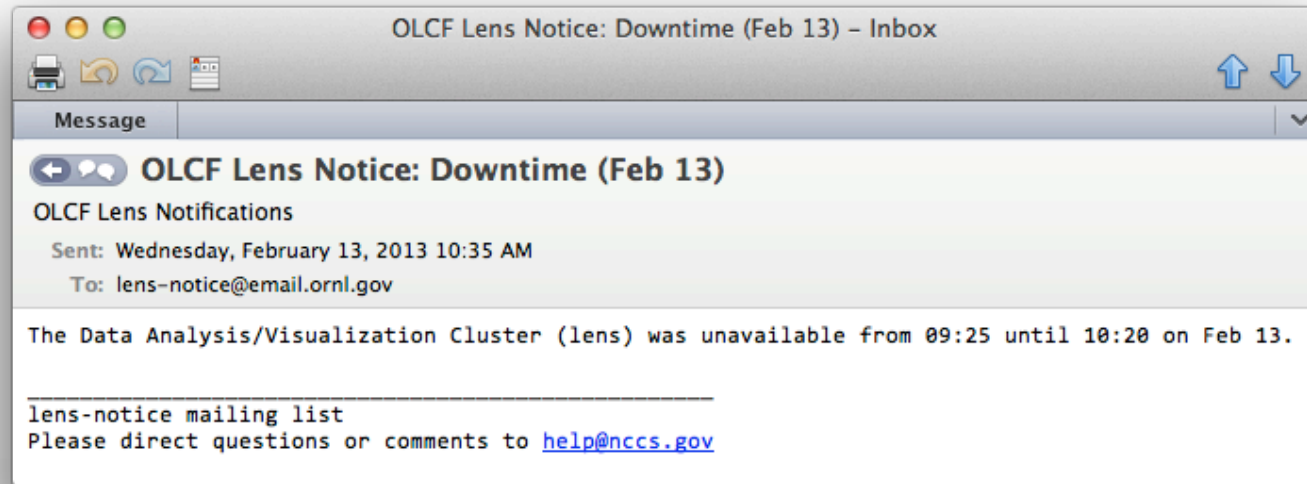
System Status: Twitter

@OLCFStatus



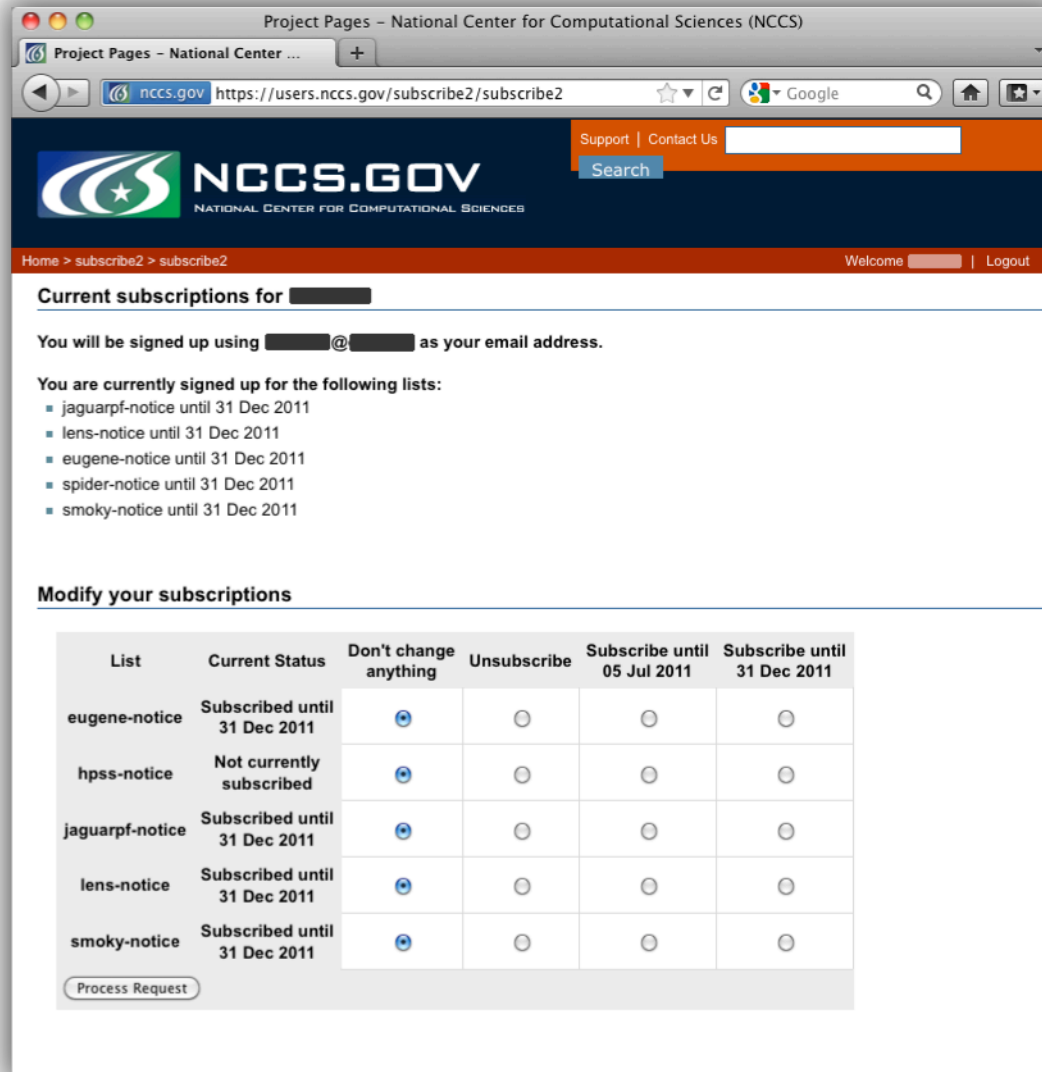
System Status: Opt-In Email Lists

- We also send status up/down notices via email
 - These are available on an opt-in basis
 - See http://www.olcf.ornl.gov/kb_articles/system-notification-lists/
 - Subscribe only to lists of systems of interest to you
- Other notices are sent to these lists, so you may want to sign up



System Status: Opt-In Email Lists

<http://users.nccs.gov/subscribe>



The screenshot shows a web browser window with the address bar displaying `https://users.nccs.gov/subscribe2/subscribe2`. The page header includes the NCCS.GOV logo and navigation links for Support, Contact Us, and a Search bar. A breadcrumb trail shows `Home > subscribe2 > subscribe2`. The main content area is titled "Current subscriptions for [redacted]" and states: "You will be signed up using [redacted]@[redacted] as your email address." Below this, it lists the current subscriptions: jaguarpf-notice, lens-notice, eugene-notice, spider-notice, and smoky-notice, all with expiration dates of 31 Dec 2011. A section titled "Modify your subscriptions" contains a table with options to change the subscription status or unsubscribe for each list.

List	Current Status	Don't change anything	Unsubscribe	Subscribe until 05 Jul 2011	Subscribe until 31 Dec 2011
eugene-notice	Subscribed until 31 Dec 2011	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
hpss-notice	Not currently subscribed	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
jaguarpf-notice	Subscribed until 31 Dec 2011	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
lens-notice	Subscribed until 31 Dec 2011	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
smoky-notice	Subscribed until 31 Dec 2011	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

System Status: Smartphone Apps

- System status apps have been developed
 - Android: “OLCF StatusApp”
 - iPhone: “OLCF System Status”
- Choose systems to monitor
- Automated notifications of system changes
- Usage instructions on olcf.ornl.gov
- ***Currently unavailable due to app maintenance***



OLCF StatusApp

OLCF Status Summary

Data Last Updated	Last Upd
15:30 on 21 Aug 2012	15:30 on 21 Aug 2012
↓ jaguarpf	down since 08:05 on 21 Aug 2012
↑ jaguarpfext	up since 13:50 on 24 Feb 2012
↑ lens	up since 08:55 on 27 Jun 2012
↑ smoky	up since 09:30 on 15 Aug 2012
↑ frost	up since 07:00 on 20 Jun 2012
↓ hpss	down since 15:15 on 21 Aug 2012
↑ widow2	up since 22:25 on 19 Jun 2012
↑ dtn	up since 15:35 on 14 Aug 2012

help@olcf.ornl.gov (865) 241-6536



Systems

↑ jaguarpf	up since: Oct 10, 2011 5:10 PM
↑ widow1	up since: Sep 28, 2011 4:50 PM
↑ widow2	up since: Sep 28, 2011 4:50 PM
↑ widow3	up since: Sep 28, 2011 4:50 PM
↑ lens	up since: Sep 28, 2011 4:50 PM
↑ jaguarext	up since: Sep 28, 2011 4:50 PM
↑ hpss	up since: Sep 30, 2011 1:00 PM
↑ dtn	up since: Sep 28, 2011 4:50 PM

Updated 10/11/11 3:45 PM

Using the Systems at OLCF

- Software
- Compiling
- Common Error Messages
- Common Questions
- Other Tips/Best Practices

Finding Software

- Some software is part of the default environment
 - Basic commands
 - Text editing utilities
- Software is typically managed via the ‘modules’ utility
 - Software is actually installed in `/sw`
 - To list available software, use `module avail`
 - To use a package, use `module load`
 - More information is available on the OLCF website
- “Important” items, such as compilers, are also available via modules

Software Installation/Updating

- We are moving to a model of updates to software packages at certain intervals (or in the case of major revisions)
 - This means not all minor versions will be installed
 - We'll move towards adding build instructions on the website so that you can build minor revisions/slightly different versions
- Look for information on the OLCF website and via Weekly Update emails

Software Installation

- You are free to install software in your directories (including your project directory)
 - Subject to terms of license agreements, export control laws, etc.
- If you think a piece of software would be of general interest, you might ask us to install it for general use
 - Preferred method:
<http://www.olcf.ornl.gov/support/software/software-request/>, but email to help@olcf.ornl.gov works, too.
 - This will be reviewed by our software council

Compiling for the XK7

- The compilers on the XT/XE/XK line of systems may differ (significantly) from your previous experience
- Combination of `?-asyncpe` and `PrgEnv-?` modules
 - `?-asyncpe` provides compiler wrapper scripts
 - `PrgEnv-?` loads modules for compilers, math libraries, MPI, etc.
- Regardless of actual compiler being used (PGI, Intel, GNU), invoke with `cc`, `CC`, or `ftn`
- MPI, math, and scientific libraries included automatically
 - No `-lmpi`, `-lscalapack`, etc.
 - This can be challenging when dealing with some build processes

Compiling for the XK7

- You are actually cross-compiling...processors (& instruction sets) differ between login and compute nodes
 - *It is very important to realize this...utilities like “configure” often depend on being run on the target architecture, so they can be challenging to use on the XK7*
- Compiling for login/batch nodes is occasionally necessary
- There are three ways to do this
 - `module swap xtpe-interlagos xtpe-target-native`
 - Add `–target=native` to `cc/CC/ftn`
 - Call the compilers directly (e.g. `pgcc`, `pgf90`, `ifort`, `gcc`)

Common Runtime Errors

- `Illegal Instruction`
 - A code was compiled for the compute nodes but executed on login nodes
- `request exceeds max nodes alloc`
 - The number of cores required to satisfy the `aprun` command exceeds the number requested
 - Node allocation should be more intuitive now that we're using `#PBS -l nodes=n` to request resources (as compared to `#PBS -l size=n`)
 - Also happens when your request is correct, but at launch time a node is discovered to be down
 - Quick fix: request “extra” nodes & include logic to restart in your batch script

Common Runtime Errors

- relocation truncated to fit: R_X86_64_PC32
 - The static memory used by your code exceeds what's allowed by the memory model you're using
 - Only the “small” memory model is available (static size \geq 2GB)
 - Solution: use dynamic memory allocation to the greatest extent possible

Common Runtime Errors

- `aprun: [NID 94]Exec /tmp/work/userid/a.out failed: chdir /autofs/nal_home/userid No such file or directory`
 - When you run ‘aprun’, you must do so from a directory visible to compute nodes (/tmp/work or /tmp/proj)
 - As a general rule, lustre directories are visible and others are not
 - When the job starts, it will attempt to ‘cd’ into the working directory from which you submitted aprun
 - In the example above, ‘aprun’ was executed from the user’s home directory
 - Any files used by the processes running on compute nodes (input, output, .so, etc) must be in directories visible to the compute node.
 - This is not the case for executables...they can be in your home directory

Common Questions

- *Is my data backed up?*

- NFS directories: Yes, to an extent. Take a look at
/ccs/home/.snapshot/

```
$ ls /ccs/home/.snapshot  
hourly.0 hourly.1 hourly.2 hourly.3 hourly.4 hourly.5  
nccsfiler3(0151729160)_home.1 nightly.0 nightly.1
```

- Lustre directories: No
- HPSS: No.

While you might use it as a backup of your directories, HPSS itself is not backed up. It's a good idea to have another level of backup at some other site if possible.

Common Questions

- *What project am I on, and what's its allocation?*
 - Use showproj to list your projects
 - Use showusage to display utilization
 - Both commands have a “help” option...run them with `–h` for usage info

```
$ showproj
```

```
brenaud is a member of the following project(s) on jaguarpf:  
  stf007
```

```
$ showusage
```

```
jaguar usage in CPU hours:
```

Project	Allocation	Project Totals		brenaud Usage
		Usage	Remaining	
stf007	600001	562227.60	37773.40	12968.42
stf007del	500000	0.00	500000.00	0.00

Common Questions

- *What happens when my project overruns its allocation?*
 - Most importantly, we do **not** disable the project...jobs simply run at lower priority
 - Jobs submitted to projects slightly over allocation (100-125%) receive a 30-day priority reduction
 - Jobs submitted to projects well over allocation (>125%) receive a 365-day priority reduction
 - This allows a degree of “fairshare” while still allowing people to run when the system is quiescent

Common Questions

- *My project has lost X hours due to system issues...can I get that time reimbursed?*
 - Since we don't disable projects for going over allocation, we also don't deal with refunds *per se*
 - If many jobs are affected, the priority reduction can be delayed. This is basically a refund but is much easier to manage.

Common Questions

- *I changed permissions on `/tmp/work/$USER`, but they changed back...why?*
 - Permissions in the lustre filesystem are controlled by settings in our accounts database
 - These settings only affect the top-level permission
 - Permissions are automatically (re-)set regularly
 - Most users can request they be changed
 - Send email to help@olcf.ornl.gov
 - Note that you need to email us to change them “back”
 - Of course, you can always just `chmod` everything under the top-level directory
 - We can't change permissions on directories associated with sensitive data

Common Questions

- ***Can I get a priority boost? ...a higher walltime limit? ... purge exemption? ...larger home directory quota ...etc?***
 - Perhaps...
 - See <https://www.olcf.ornl.gov/support/documents-forms/>
(Forms to Request Changes to Computers, Jobs or Accounts)
 - Once submitted, forms are sent to the Resource Utilization Council for approval
 - Make request well in advance...it can be difficult to make last-minute changes
 - If requesting job priority, be sure the job is submitted (the queue may move more quickly than expected, eliminating the need for the request)

Common Questions

- *Where do I archive my data? Is there a mass storage system available?*
 - HPSS
 - Accessed via `hsi` & `htar`
 - See the “Storage @ OLCF” presentation for full details

Common Questions

- *How do I transfer my data? If the main system is down, can I access my data?*
 - The Data Transfer Nodes are the preferred place for data movement (both internal and external)
 - `dtm01.ccs.ornl.gov` and `dtm02.ccs.ornl.gov`
 - Almost always available, even during outages of major resources
 - All users should have access

Data Storage Practices

- HPSS is the proper location for long-term storage
- Project areas (NFS and lustre) offer a common area for shared data files, executables, but should not be considered long-term storage
 - Need to keep an eye on disk usage
 - Should still be backed up
- User scratch areas are intended for use during computations
 - Regularly purged
 - Store files to HPSS as soon as practicable
 - File cleanup is important

Lustre/Scratch Filesystem

- Remember that the current version of lustre has a single metadata server
- Codes may need revisions to prevent opening large numbers of files simultaneously
- Compiling in lustre can be slow
 - When possible, compile in NFS
- There is a purge, but if you delete files when not needed it will help

Dealing With the Scratch Purge-Conditional Transfers

- Many codes use files from previous iterations of the code, and those files may get deleted by the scratch purge
- This can present challenges
 - Pulling from HPSS every time is inefficient
 - Multiple scripts (“data in place”, “data not in place”) are cumbersome
 - touch isn’t always ideal
- Conditional transfers help with this

```
#!/bin/bash
...
if [[ ! -a /tmp/work/brenaud/some_important_file ]]; then
    hsi -q get /home/brenaud/data/some_important_file
fi

aprun -n 4096 ./a.out
...
```

Interacting with HPSS

- HPSS is a somewhat complex system
- HPSS prefers a small number of large files and not a large number of small files-`htar` is your friend in this regard
 - `htar` is (much) faster than a `tar` followed by `hsi put`
 - Limited disk space is no problem...data is streamed directly to HPSS so there is no “intermediate” local storage
- Running many transfers at a time can be problematic
 - Multiple transfers may not give you parallelism
 - Limiting the number of per-user transfers helps the system operate more efficiently (& therefore can be more efficient for you)
- Usage examples are on the OLCF web site

Running Jobs at OLCF

- Batch job information is available on the OLCF Web Site
- Since we are designated as a “leadership-class” facility, queuing policy heavily favors large jobs
- Remember that jobs are charged based on what’s made unavailable to others and not what you use
 - The system cannot allocate a single node to multiple `aprun` instances (much less multiple users)
- Requests for high priority/quick turnaround are considered
 - Submit your job ASAP...it may start more quickly than expected
 - Allow plenty of lead time when making a request...discussion may be necessary prior to a decision on approval

Running Jobs at OLCF

- From a user perspective, Titan has three major parts
 - The system proper
 - External login nodes
 - MOAB server
- Often, only the system proper is affected by outages
 - External login nodes and the MOAB server node remain up
 - This means you can compile/submit jobs/etc while titan is down
 - Jobs will be queued and will run when the system returns

Debugging/Optimization at OLCF

- Several software tools are provided for debugging and optimizing your applications
 - DDT
 - Vampir
 - CrayPAT
- Information on these tools is available on the web; you can also contact the OLCF User Assistance Center if you have questions

Support Best Practices

- Send as many error messages as possible
 - Or, place them all in a file and direct us to it
- When sending code, create a .tar file & direct us to it
 - Include all files necessary to build/run the code
 - More efficient than sending through email
- When possible, reduce error to a small reproducer code
 - We can assist with this
 - If the error has to go to the vendor, they'll want this
- Send new issues in new tickets, not replies to old ones

Finally...

- We're here to help you
- Questions/comments/etc. can be sent to the OLCF User Assistance Center
 - 9AM – 5PM Eastern, Monday-Friday exclusive of ORNL holidays
 - help@olcf.ornl.gov
 - (865) 241-6536