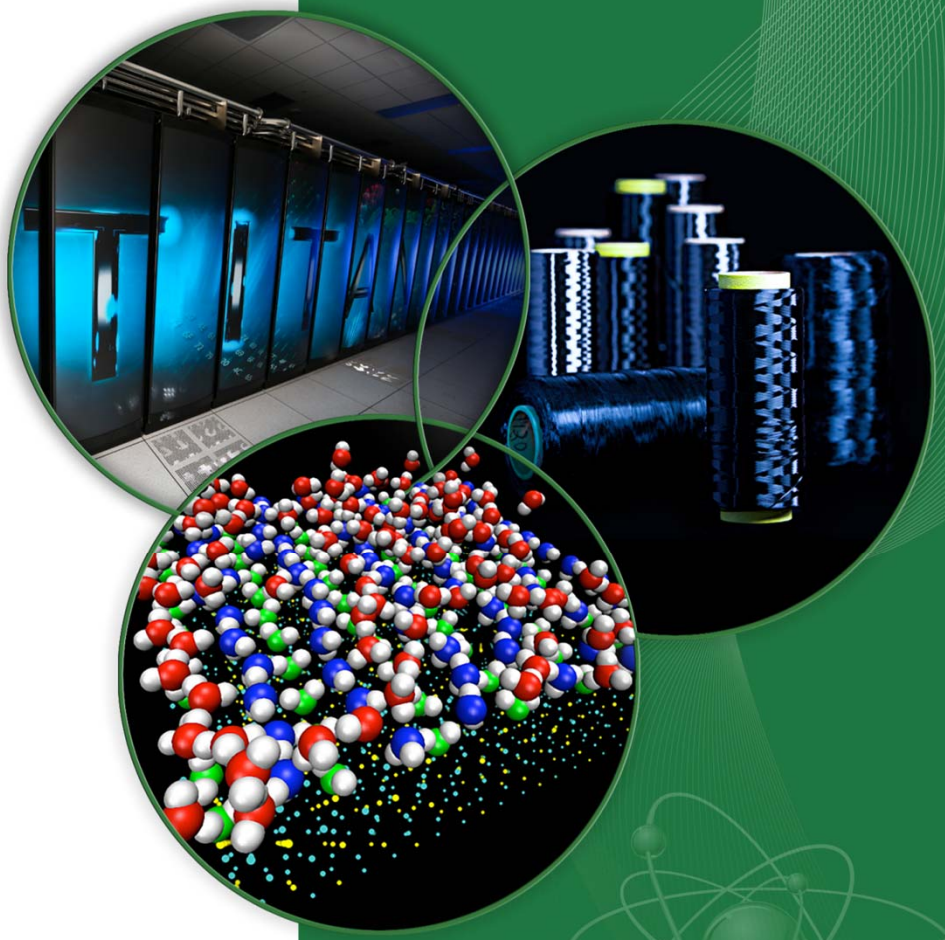
Year in Review: User Survey Results

Bill Renaud

Oak Ridge Leadership Computing Facility



ORNL is managed by UT-Battelle
for the US Department of Energy

OAK RIDGE National Laboratory

First…Thank You!

- Thanks to everyone who completed a survey
- The user survey is vitally important to us
  - Required by DOE
  - More importantly, it allows us to know what we're doing right and what needs improvement

Selected Highlights

- Once again, handled by ORISE
- 367 respondents (just under 30%)
- Good mix of INCITE, ALCC and DD projects
- Good mix of new and previous users
  - Over 2/3 had been users less than 2 years
- Overall satisfaction remains high (4.41/5.0)
  - Only 5% gave an overall rating of “neutral” (3.0/5.0)
  - No “dissatisfied” or “very dissatisfied” ratings

What You Like

- Both the people and the hardware
- Large job turnaround
- OLCF communications

What Improvements You'd Like To See

- Queue policy
- Faster account turnaround
- Fewer outages
- Training improvements
- Better communication of announcements/system status
- Disk/storage issues
  - Small quotas
  - Data access/sharing
  - Purge timeline

How We're Responding

- Staff members review results and then meet to identify how to address concerns
- Surveys have guided us to several system enhancements and policy changes
  - Scheduled DTNs
  - Filesystem updates ($WORLDWORK, etc.)
  - HPSS enhancements
  - NetApp quota increase
  - Cross-system workflows
  - New dashboard (soon)
- The next few presenters will expand on these

Finally, …

- Once again, thank you for completing the survey
- The 2014 User Survey is due out in early October
  - Please join in…help us help you!
  - Your opinion *DOES* matter to us
  - We'd love to have a 100% response rate

Year in Review: Systems - Eos and Rhea

ORNL is managed by UT-Battelle
for the US Department of Energy

OAK RIDGE
National Laboratory

Eos

- Cray XC-30
- 744 Nodes
- Prototyping for non-accelerated codes
- Available to INCITE, ALCC users
- Available to DD users upon request

Eos

- Aries Interconnect with Dragonfly network topology
  - Dragonfly is a “low-diameter” topology
  - Maximum “hops” between any two nodes is minimized
- 64GB RAM in each node
- Intel Xeon E5-2670 features 16 physical cores
  - 32 virtual CPUs / node via HyperThreading
- Shares the center-wide Atlas filesystem

Rhea

- 512 node commodity cluster
- Available for any Pre- and Post-processing work
  - Grid preparation
  - Visualization
  - Data reduction
  - And more
- Available to all OLCF users

Rhea

- Dual Intel Xeon E5-2650 packages in each node
  - 16 physical cores
  - 32 virtual CPUs via HyperThreading
- 64GB of RAM
- 4X FDR Infiniband interconnect
- Shares the center-wide Atlas filesystem
- Ideal for supporting automatic pre-/post-processing workflows

Cross-System Workflows

- OLCF now provides cross-system batch submission
- For example, Titan jobs can
  - submit viz tasks to Rhea
  - submit transfer tasks to the DTNs
- And vice-versa
  - DTNs can start simulations on Titan after data is unloaded from HPSS
  - Rhea can start simulations on Titan after input decks have been generated

Year in Review: Data Management At OLCF

Suzanne Parete-Koon

Oak Ridge Leadership Computing Facility

ORNL is managed by UT-Battelle
for the US Department of Energy

Filesystem Upgrade

- The Spider center-wide filesystem was upgraded in January.
  - 32 petabytes of disk space. Aggregate speed of 1 TB/s, however, this bandwidth is shared so the effective speed is much lower per user.
- After the initial upgrade users may have noticed sluggish performance
- These problems have largely been mitigated, by moving Titan's Lustre client back to version 1.8.6, and by patches to the Lustre server however, if you continue to see issues please report them to help@olcf.ornl.gov
- The 1.8 Lustre clients running on Titan do not support stripe counts greater than 160.

Directory Structure for 2014

	Member Work	Project Work	World Work
Description	Scratch area	Scratch Area for Sharing data within a project	Scratch Area for sharing data between projects.
Location	$MEMBERWORK	$PROJWORK	$WORLDWORK
Quota	10 TB	100 TB	10TB
Purge	14 days	90 days	14 days
Access	May alter permissions to share with project	All project members have access	All OLCF users can access

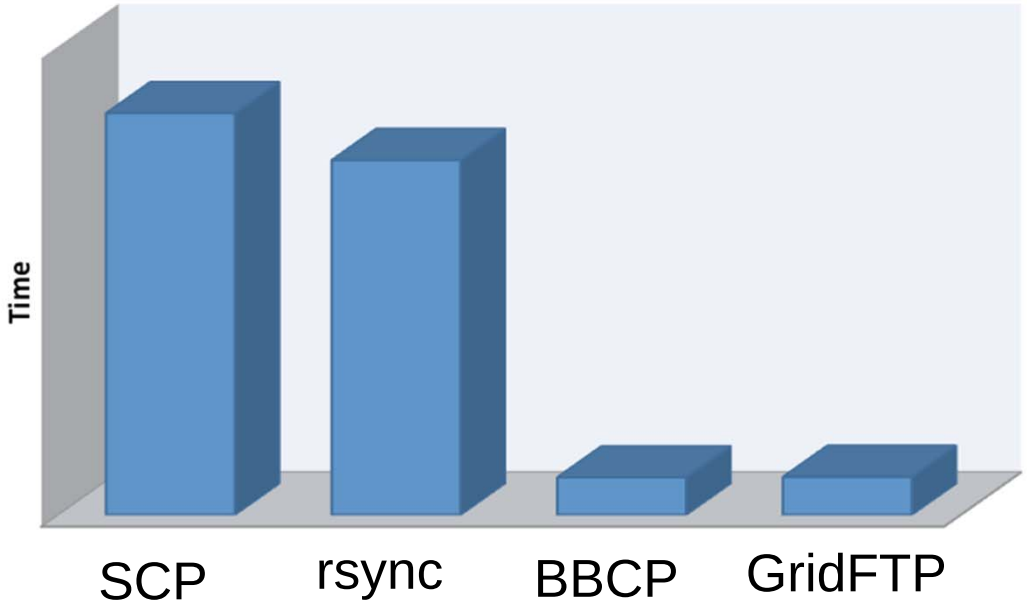
Data Transfer Nodes

- Data transfer nodes should be used for all remote and local data transfers.
- 2 interactive data transfer nodes
  - dtn.ccs.ornl.gov
- 14 Batch scheduled data transfer nodes
  - 2 batch nodes dedicated to hpss transfers.
  - nodes do not suffer from contention
  - dtn jobs may be launched from Titan, Rhea or Eos.

Data Transfer Software

New in 2014: A Supported Globus Endpoint olcf#dtn

- Globus and globus_url_copy require an Open Science Grid Certificate
- Speeds for tools:



4 parallel streams:

- bbcp –s4
- GridFTP –p4

Data Management Users Guide

- We have organized a New Data Management User Guide
  - Data management policy
  - Directory Structures of the filesystems
  - Data transfer

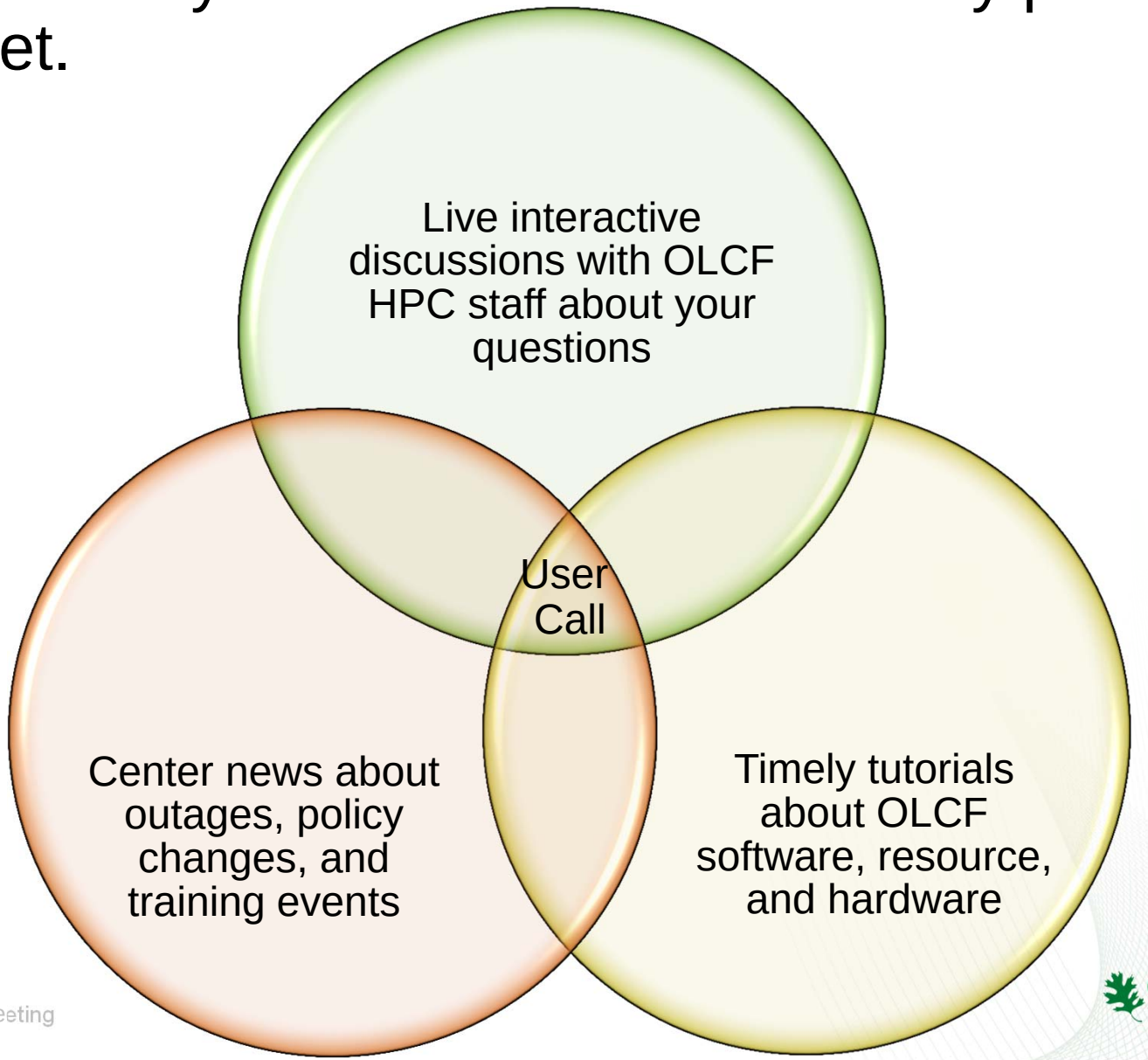
  Look for this icon on the systems guide page:

HPSS Improvements

- In the next year the OLCF will bring many new resources to the HPSS archive
  - The addition of a larger disk cache will mean fewer retrievals from tape
  - 40 Gigabit/s Ethernet will speed up staging from tape to disk cache
  - New tape drive technology will speed reads and writes to tape
- Upcoming upgrades to HPSS software will increased performance to the database which should help storing and retrieving files

OLCF User Call

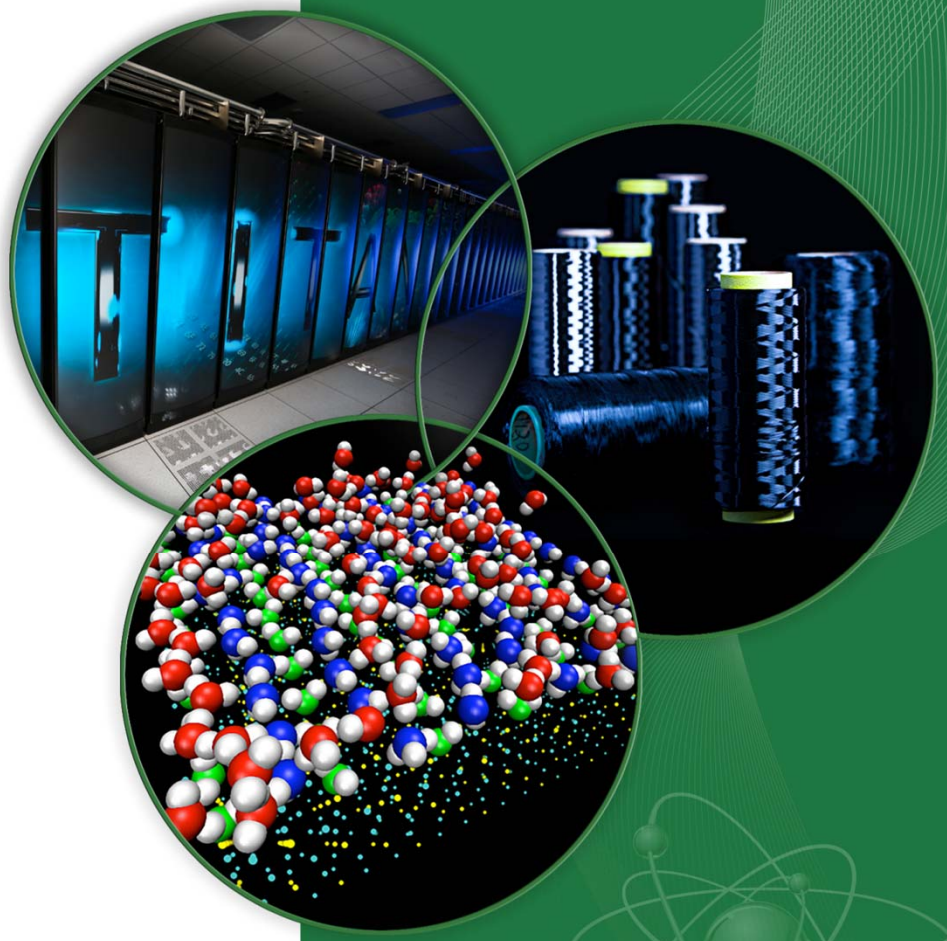
- First Tuesday of the Month. Delivered by phone and Internet.



Year in Review: Accelerator Support

Adam Simpson

Oak Ridge Leadership Computing Facility

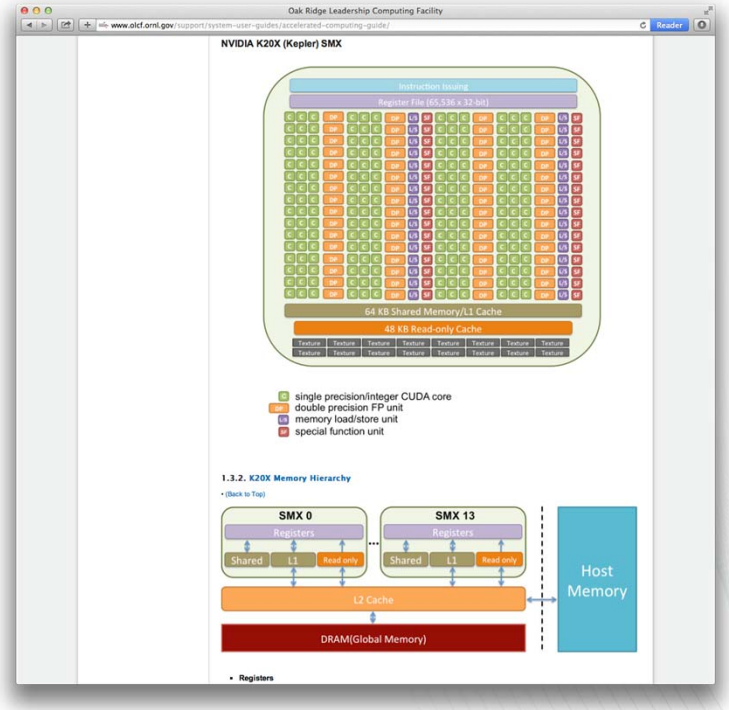


ORNL is managed by UT-Battelle
for the US Department of Energy

OAK RIDGE National Laboratory

Accelerated Computing User Guide

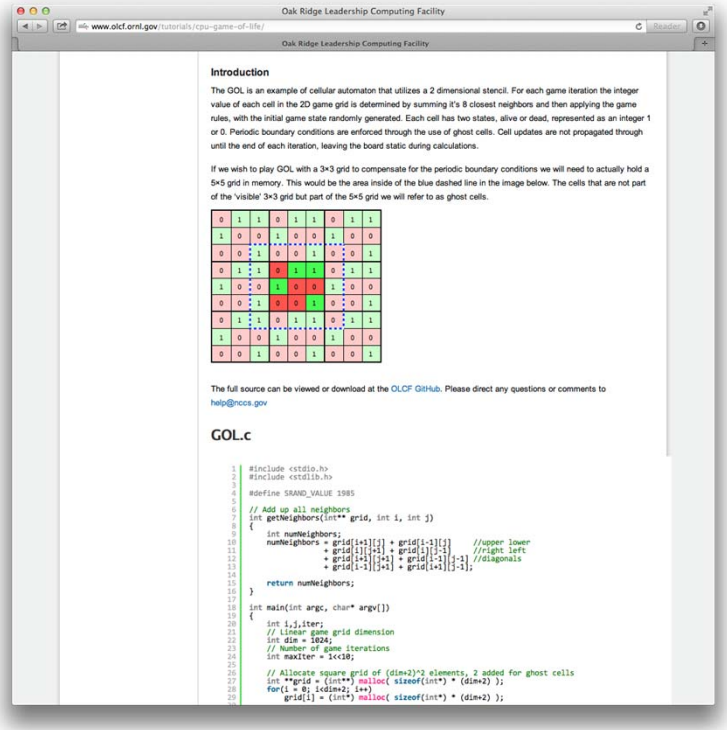
- User survey indicated preference for online documentation
- Created Accelerated Computing User Guide
  - Comprehensive guide for GPU computing on Titan

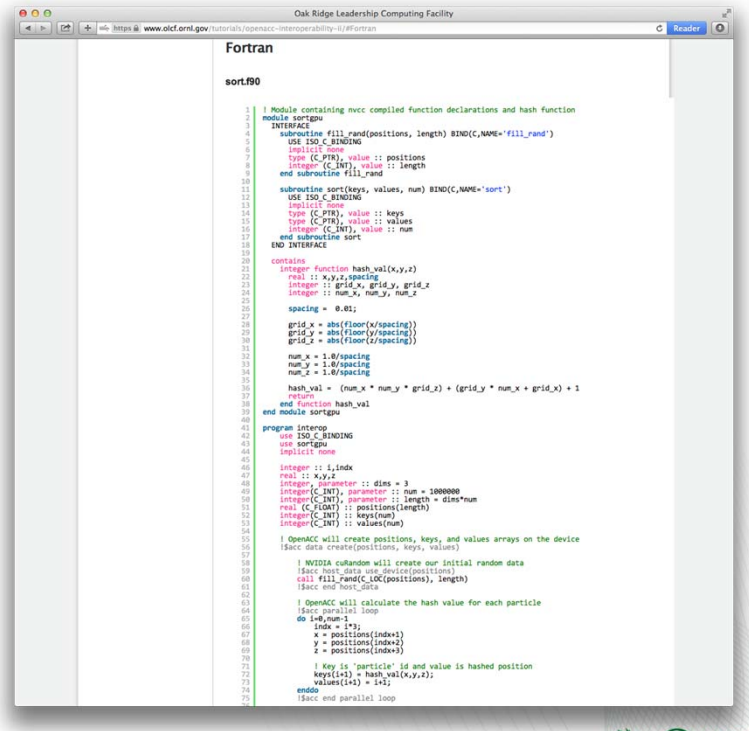


OAK RIDGE National Laboratory

Tutorials

- Programing and use examples
  - C and Fortran
  - Highlight HPC specific features and functions
- Several more to be released soon





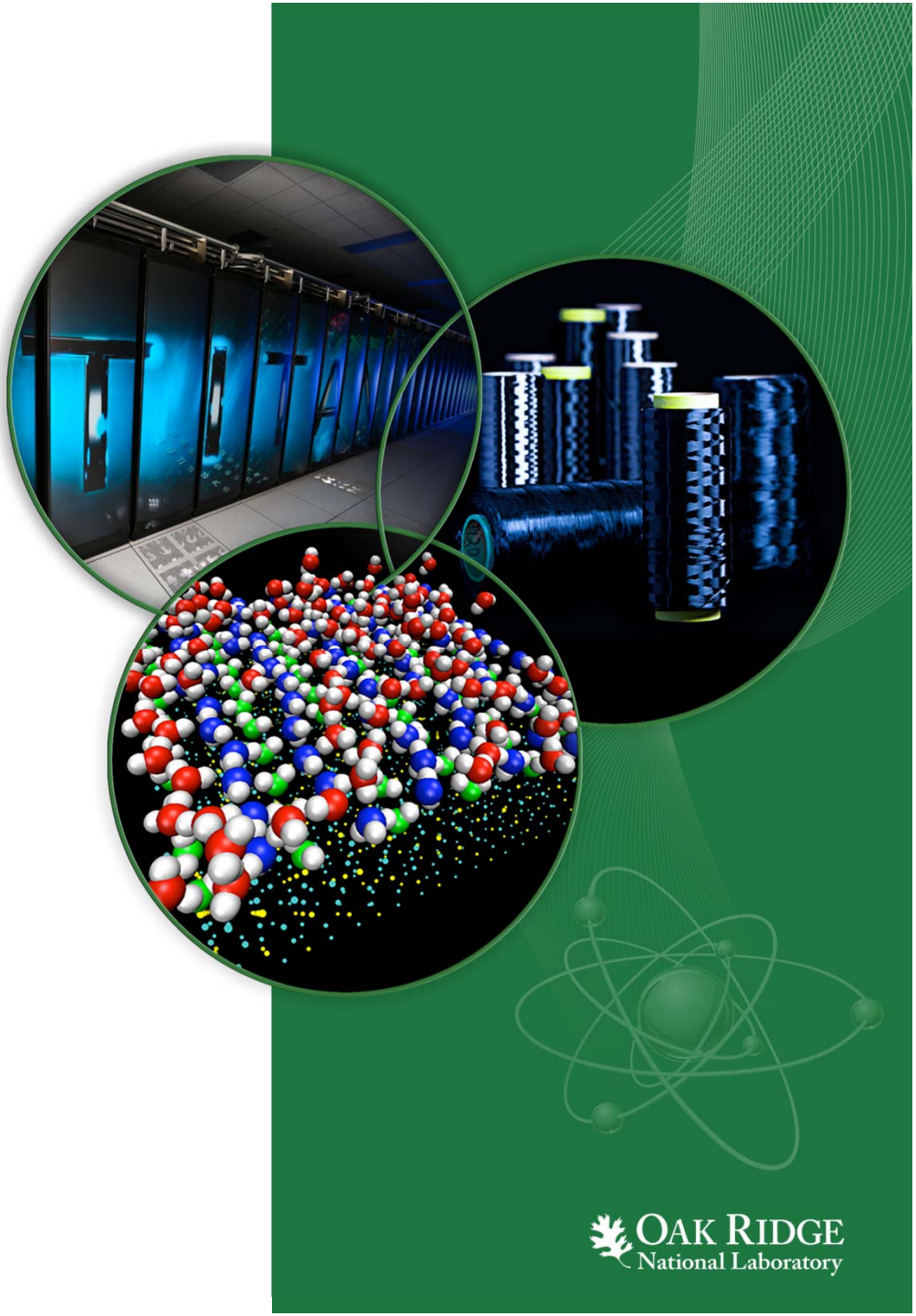
User Feedback

- User feedback determines documentation topics and delivery format
- Must hear from users to provide relevant HPC documentation
- Email any suggestions to help@olcf.ornl.gov

Year in Review: OpenMP/OpenACC and Training Updates

Fernanda Foertter

Oak Ridge Leadership Computing Facility



ORNL is managed by UT-Battelle
for the US Department of Energy

OpenMP™

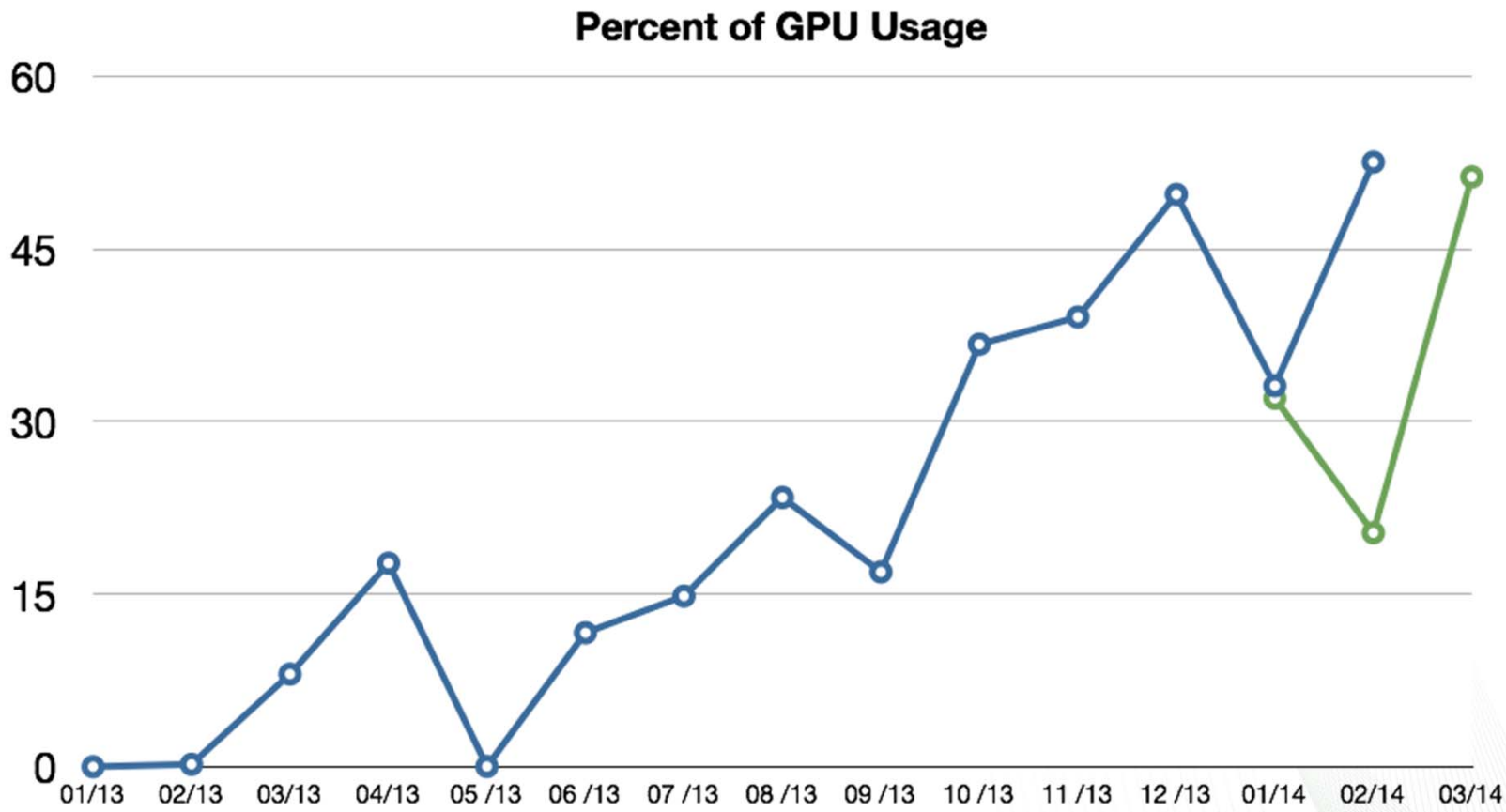
OpenACC

Directives for Accelerators

OpenMP and OpenACC activities

- Weekly calls (2h)
  - Recommendation on standards, syntax, features, behavior
- User use cases
- Bugs, behavior
- Communication with compiler developers

Increasing Usage of GPUs



As measured by ALTD against linked libraries

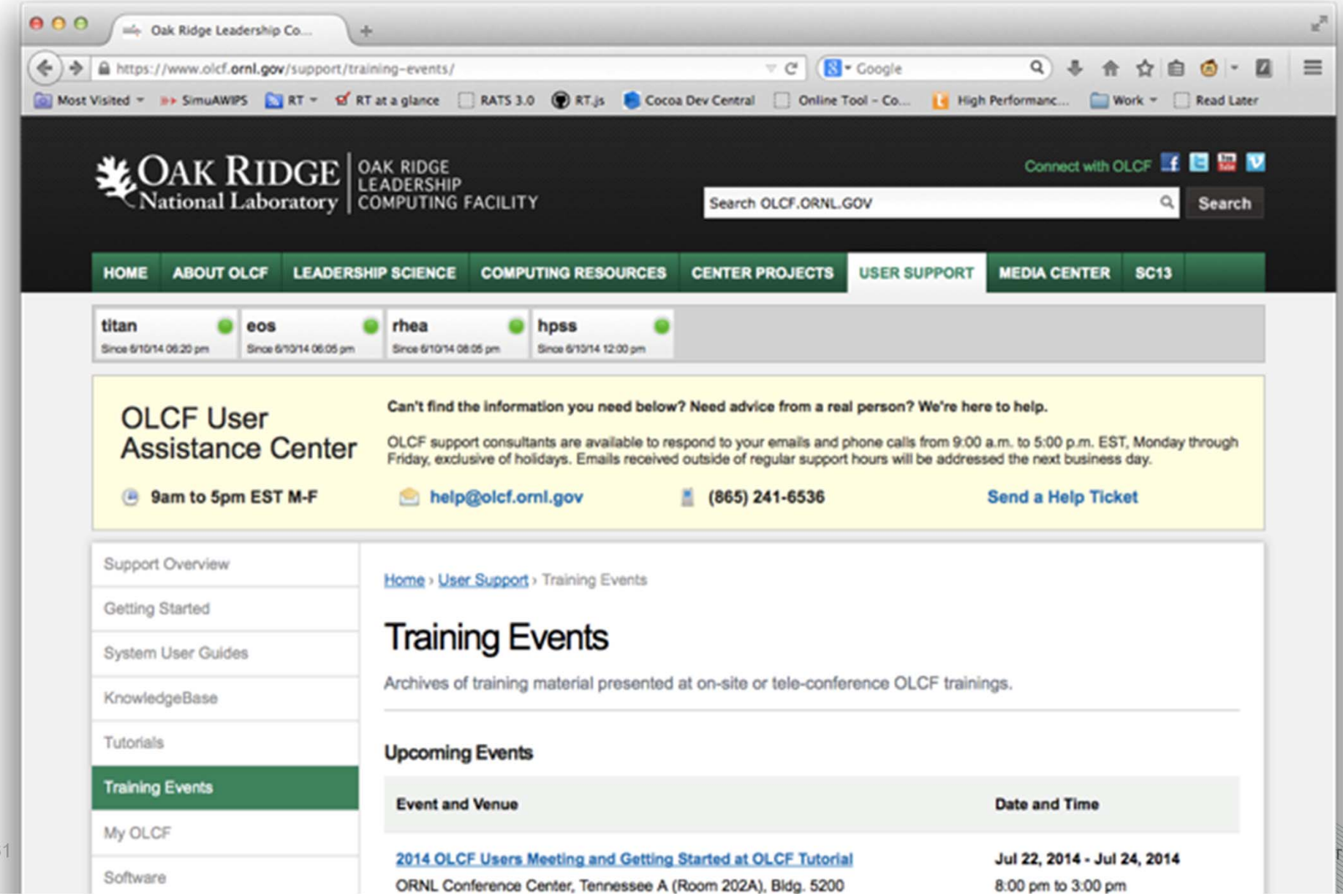
Many of these run on Titan

Growing OpenACC Activity

Weather / Climate / Ocean	Chemistry	Physics	Oil & Gas	Astronomy	Fluid Dynamics
CAM-SE	CASTEP	**CloverLeaf**	AWP-ODC	CHIMERA	FVCF-NIP
COSMO (Physics)	CoMD	GENE	EMGS ELAN	IUmd	HemeLB
DYNAMICO	GAMESS CCSD(T)	GTC	MOT-TDVIE	RAMSES	NEK5000
FIM	GAUSSIAN	LULESH	**Seismic CPML**	X-ECHO	PALM GPU
Harmonie	MiniMD	MiniGHOST	SEISMO		UPACS
HBM	ONETEP	**S3D**	TeraP		ZFS
ICON	Quantum Espresso				
NICAM	SMMP				
NEMO GYRE					
NIM					
NTSU Snow Simulator					
OLAM					
PALM-GPU					
ROMS					
SCALE-LES					
WRF					

OpenACC
Directives for Accelerators

HPC User Assistance Team at OLCF



Training Focus This Past Year

- More webcasts
- More specific topics
- More in-depth topics
- Shorter topics
- Repository (github)
- Open to the public

To come: short screencasts

Upcoming Training Activities

OpenACC Hack-a-thon	Oct 27, 2014 - Oct 30, 2014
Compiler Directives Weekly Lunch Webinars Bldg. 5200, Room 219	**Jun 9, 2014 - Jun 9, 2014** 12:00 pm to 1:30 pm
OpenACC Tutorials JICS Auditorium, Building 5100	**Jul 15, 2013 - Jul 18, 2013** 9:00 am to 2:00 pm

- Fernanda Foertter
- ccs-training@email.ornl.gov

@hpcprogrammer

(unofficial!)

Year in Review: Software - Tools

Chris Fuson

Oak Ridge Leadership Computing Facility

ORNL is managed by UT-Battelle
for the US Department of Energy

Vendors and Software Developers

- Accounts on Titan and internal development systems
- Allows and speeds the process
  - verify
  - build
  - system specific features
  - work bugs
  - support

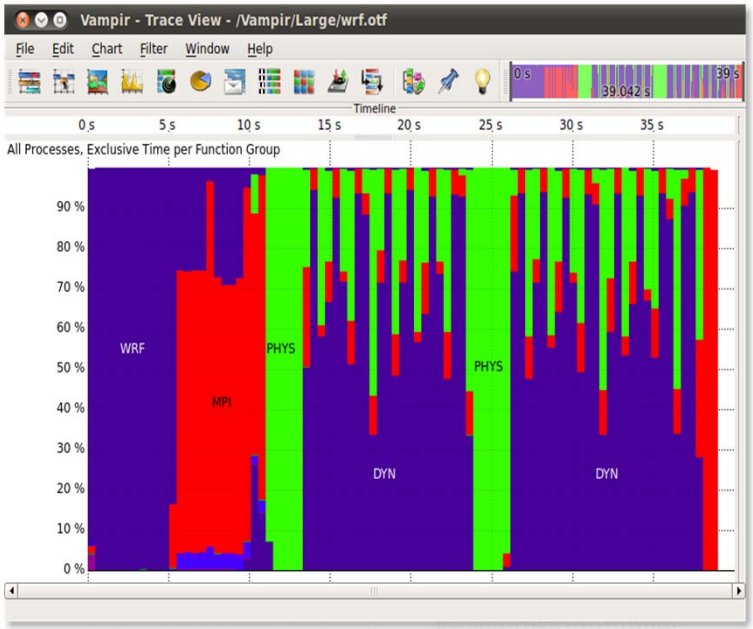
- Examples
  - **Nick Forrington**
    - *DDT*
  - **Frank Winkler**
    - *Vampir*
  - **Scott Atchley**
    - *aprun layout*

New Features in DDT

- Logbook view
  - Logs the user's interaction with DDT (e.g. stacks, locals, tracepoints)
  - Enabling comparative debugging and repeatability.
- Version control integration (Git / Mercurial / Subversion)
  - Shows annotation information against source code
  - Add breakpoints and tracepoints to track changes in a particular revision
- VisIt
  - Visual preview of how multi-dimensional & distributed arrays will be displayed.
  - Visualize multiple arrays at a vispoint
- More information in user guide
  - Press F1 while running DDT
  - http://content.allinea.com/downloads/userguide.pdf

New Features in Vampir

- Score-P (Successor of VampirTrace)
  - Common instrumentation and measurement infrastructure for parallel codes
  - Supports a number of performance analysis tools
    - Periscope, Tau, Cube, Scalasca, Vampir
  - Available on Titan, Eos and Rhea
    - module load scorep

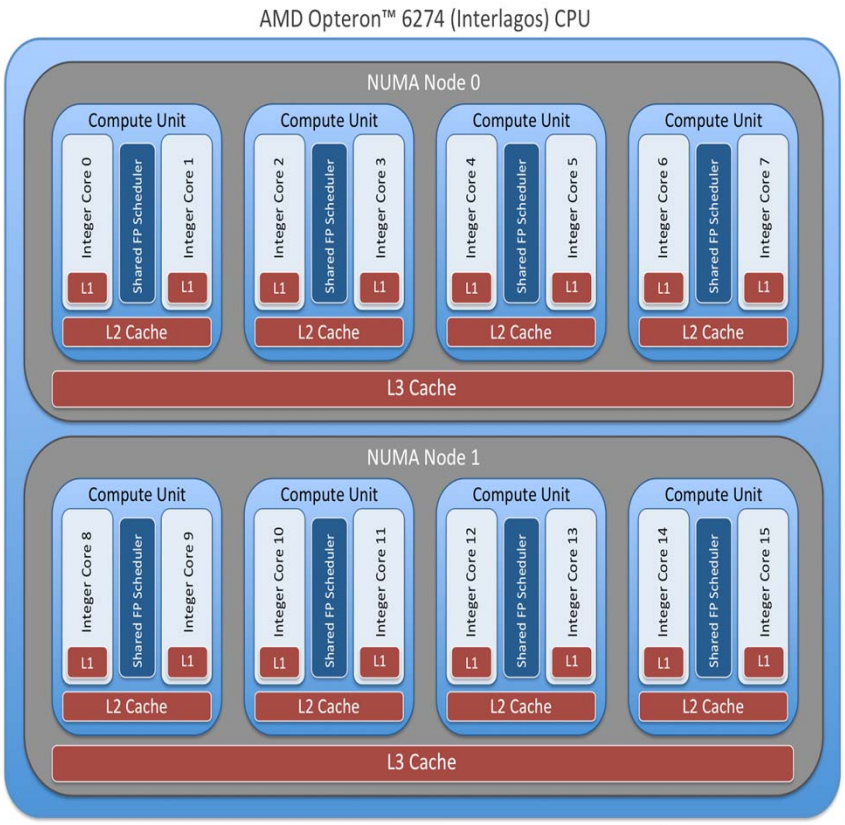


- Vampir 8.3.0
  - New Summary Timeline chart
  - Available on Titan, Eos and Rhea
    - module load vampir

- www.olcf.ornl.gov/kb_articles/software-vampir/
- www.olcf.ornl.gov/kb_articles/software-scorep/

Aprun - Floating-Point Contention

- Titan Node
  - Each Bulldozer compute unit:
    - 2 Integer Cores
    - 1 Floating Point Unit (FPU)
- By default, aprun will:
  - place 16 processes per node
  - place the processes sequentially starting at core 0
- Floating point intensive codes
  - Can see 2x speed-up by not sharing FPU
  - $ aprun -n8 -S4 -j1 ./a.out



OAK RIDGE National Laboratory

Aprun - Floating-Point Contention

- Noticed many codes requesting partial nodes but not spreading tasks ideally

$ aprun –n8 ./a.out

NUMA 0								NUMA 1							
Core 0	Core 1	Core 2	Core 3	Core 4	Core 5	Core 6	Core 7	Core 0	Core 1	Core 2	Core 3	Core 4	Core 5	Core 6	Core 7
0	1	2	3	4	5	6	7								

NUMA 0								NUMA 1							
Core 0	Core 1	Core 2	Core 3	Core 4	Core 5	Core 6	Core 7	Core 0	Core 1	Core 2	Core 3	Core 4	Core 5	Core 6	Core 7
0	1	2	3					4	5	6	7				

$ aprun -n8 -S8 ./a.out

- One task per compute unit: **-j1**

$ aprun -n8 –S4 –j1 ./a.out

NUMA 0								NUMA 1							
Core 0	Core 1	Core 2	Core 3	Core 4	Core 5	Core 6	Core 7	Core 0	Core 1	Core 2	Core 3	Core 4	Core 5	Core 6	Core 7
0		1		2		3		4		5		6		7	

Aprun - Floating-Point Contention

- Wrap aprun to check partial node requests
- Warning message:

```
$ aprun -n32 -S4 ./a.out
---------------------------------------------------------------------------
APRUN usage: requested less processes than cores (-S 4) without
using -j 1 to avoid floating-point unit contention
```

- Aprun still executed, only message
  - Set the environment variable APRUN_USAGE_QUIET to supress these messages.
- www.olcf.ornl.gov/support/system-user-guides/titan-user-guide/#2941