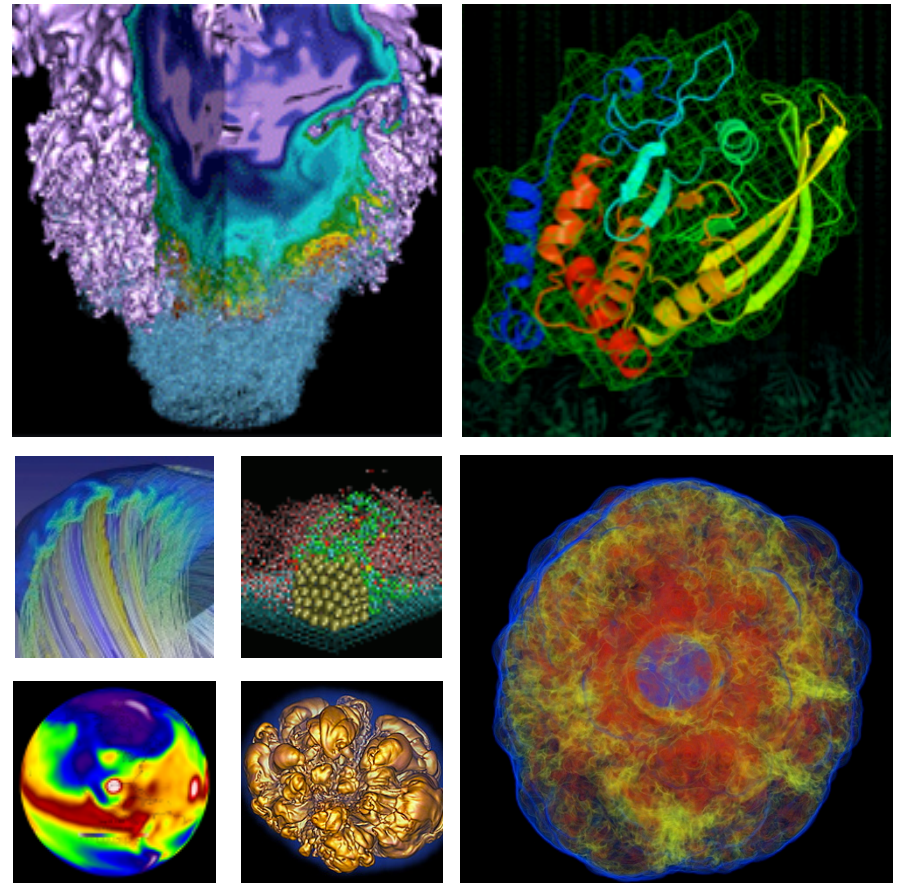


Computational Systems Update



Jay Srinivasan
Group Lead, Computational
Systems
NUG – Feb 2015



NERSC - 2014

Hopper: 1.3PF, 212 TB RAM



Cray XE6, 150K Cores

Edison: 2.5PF, 357 TB RAM



Cray XC30, 130K Cores

Sponsored Compute Systems
Carver, PDSF, JGI, KBASE, HEP
8 x FDR IB

Vis & Analytics, Data Transfer Nodes,
Adv. Arch., Science Gateways

2.2 PB Local
Scratch
70 GB/s

48 GB/s

6.4 PB Local
Scratch
140 GB/s

80 GB/s

80 GB/s

50 GB/s

5 GB/s

12 GB/s

/global/
scratch

4 PB

/project

5 PB

/home

250 TB

HPSS

45 PB stored, 240 PB
capacity, 40 years of
community data

Ethernet & IB Fabric

Science Friendly Security
Production Monitoring

Power Efficiency

WAN

2 x 10 Gb

1 x 100 Gb

Science Data
Network



NERSC - 2014

Computational Systems Group

Hopper: 1.3PF, 212 TB RAM

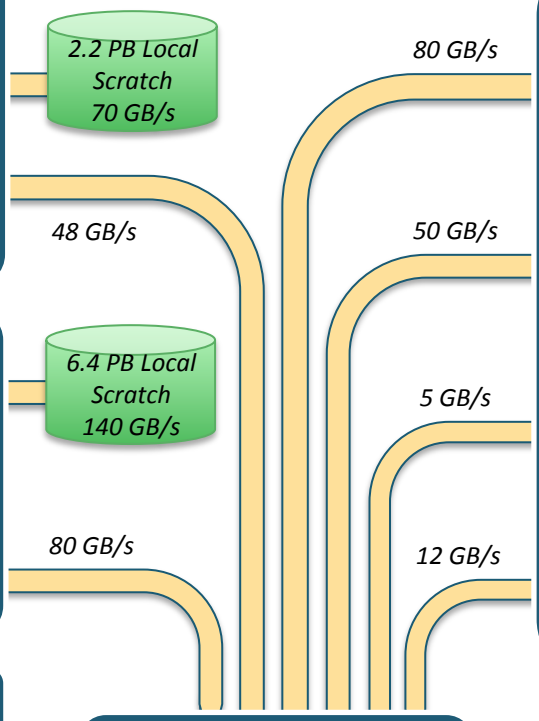
 Cray XE6, 156K Cores

Edison: 2.5PF, 57 TB RAM

 Cray XE30, 130K Cores

Integrated Compute Systems
 Carver, PDSF, JGI, KBASE, HEP
 8 x FDR IB

**Vis & Analytics, Data Transfer Nodes,
 Adv. Arch., Science Gateways**



Ethernet & IB Fabric
 Science Friendly Security
 Production Monitoring
 Power Efficiency
WAN

2 x 10 Gb
1 x 100 Gb
 Science Data Network



ESnet
 Energy Sciences Network

- **Hopper (XE-6) and Edison (XC-30) are the largest computational resources that NERSC provides**
- **Hopper – 2014**
 - O/S, Systems S/W has been upgraded to CLE-4.2
 - Improvements to system resilience and DVS performance, upgrade paths for Lustre and O/S
 - Filesystem Improvements:
 - Scratch filesystems have been upgraded to the Lustre 2.4
 - GPFS upgraded
 - Workload Manager has been upgraded
 - Serial Queue implemented

- **Hopper – Future**

- Upgrade O/S to CLE-5.2 (same level as Edison)
- Minimal changes after that – on the path to retirement
- Hopper retires at OSF – shutdown once Edison move to CRT is complete

- **Edison – 2014**

- O/S upgraded to CLE-5.2
 - Allows for future upgrade paths and newer O/S releases, improved system, ALPS, and DVS resiliency, improved DVS performance
- Filesystem improvements
 - Newer Lustre releases
 - Stability improvements to all Scratch filesystems

- **Edison – 2014 (contd.)**

- ALL memory on Edison replaced to fix H/W issue – we are now running at full performance (2 3-day outages)
- Workload Manager upgrades
 - Improved stability and features.
- Serial queue

- **Edison – Future**

- Move to CRT in Fall 2015
- Change WLM to SLURM (see Cori Phase 1) after the move

- **Carver – 2014 (contd.)**
 - Upgraded O/S to SL6 – continued to provide SL5 environment to users via CHOS
 - RAS improvements
 - Ease of management of the system
 - Workload manager upgrade
 - Major stability and performance improvement
 - Pre-emptive queue support
- **Carver – Future**
 - Will be retired at OSF in Fall 2015 (Sept 30)

- **Other activities**

- Continue supporting the HEP and NP community with PDSF – expanded to over 2600 cores
 - Supporting ALICE, ATLAS and Dayabay, LUX ...
 - Improvements to Networking and storage
- Expanded Genepool to over 8000 cores
 - Users have access to SL5, SL6 and Debian environments
- Supporting the Materials Project on Carver
- Provided a Intel Xeon Phi testbed (Babbage)
- Deployed the SciDB cluster
- Dirac (GPU cluster) retired in Dec 2014

- **PDSF, Genepool, Babbage, and the Materials Project systems will move to CRT in late Summer 2015**

- **Move to CRT**

- Edison, PDSF, Genepool, Babbage and the MatGen system will move to CRT
- Expanded capability for Genepool and PDSF to be installed in CRT
 - Fully FDR-IB connected systems
 - High B/W access to NERSC Global Filesystems

- **Cori Phase 1 System**

- Expected deployment in Summer 2015 (in CRT)
- Haswell-based system with Aries network

Cori Phase 1



- **System will be very similar to Edison**
- **Intended to replace Hopper and provide data-intensive capabilities**
- **Access to the Cori Filesystem when installed**
- **Will be fully integrated into Cori when Cori is deployed (2016)**
- **Includes**
 - Burst Buffer capabilities
 - WLM support for Data intensive workflows

- **New Workload Manager**
 - We will be using SLURM (tested extensively over the last year)
 - SLURM provides all the same functionality as Torque/Moab – with a few differences in implementation
 - SLURM is open source, and extensible
 - SLURM will perform better at handling the mixed Data-HPC needs of the Cori Ph1/Cori system
 - Simple transition from Torque/Moab to SLURM
 - Plan to give users access to test instances of SLURM soon
- **Cori (KNL nodes) to be deployed in mid-2016**
 - Phase 1 nodes and BB functionality will be integrated into Cori
 - Details of the Cori configuration wrt. Phase 1 nodes and BB will be forthcoming.

Systems Summary

System status	Date
PDSF/Genepool/Mtrls. Move to CRT	Jun. 2015
Carver Retires	Sept. 30, 2015
Cori Phase 1 available for users to run	Sept. 2015
Edison goes down to move to CRT	Oct. 2015
Edison comes back up in CRT	Nov. 2015
Hopper Retires	Dec. 2015
Cori Phase 2 Delivered	June 2016

Acknowledgements

- **The Systems Group:**

- Current: James Botts, Scott Burrow, Tina Butler, Tina Declerck, Doug Jacobsen, David Paul, Iwona Sakrejda, Bhupender Thakur, Cary Whitney
- Alumni: Nick Cardo, Ilya Malinov, Larry Pezzaglia





Thank you.