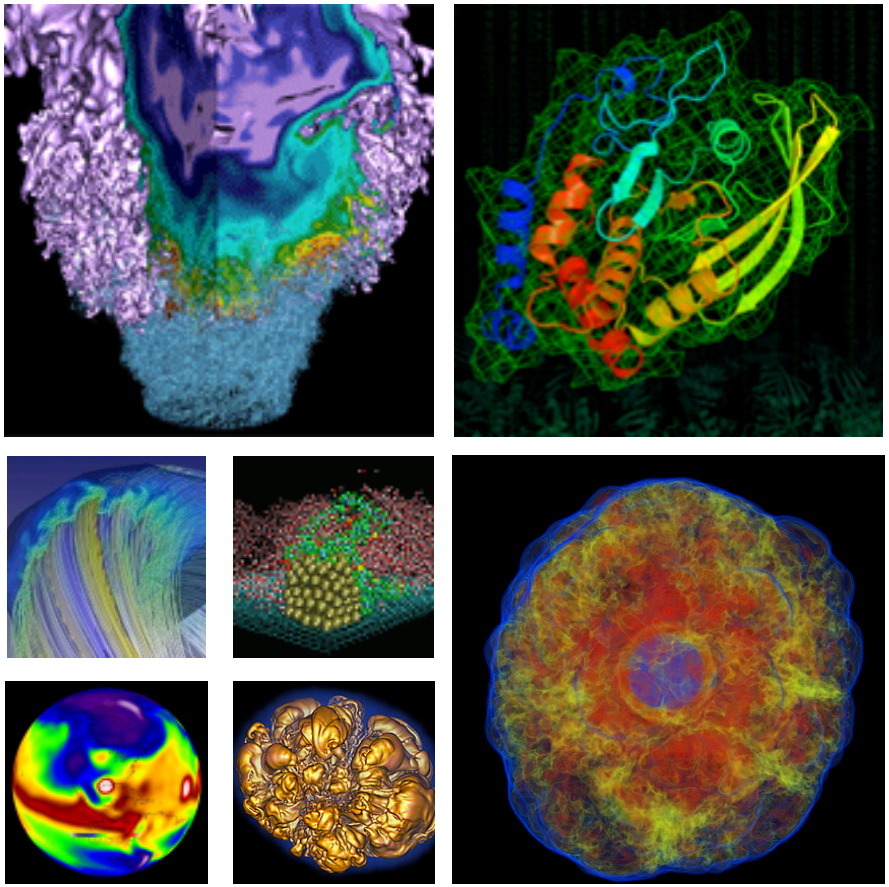
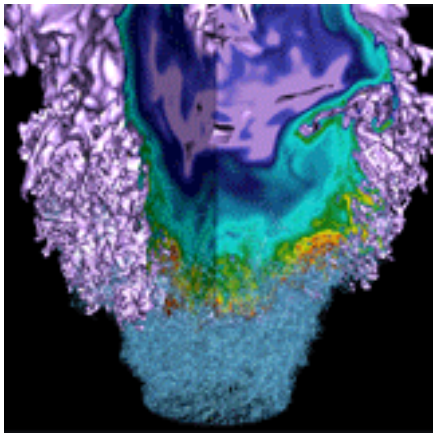
NERSC's 10 year plan

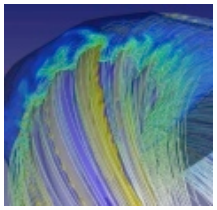


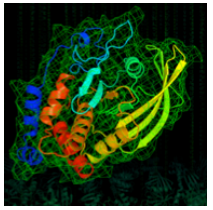
Sudip Dosanjh
Director

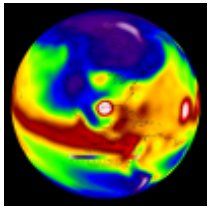
October 8, 2013

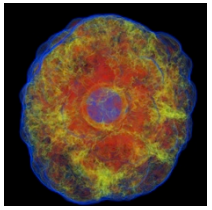
NERSC Overview

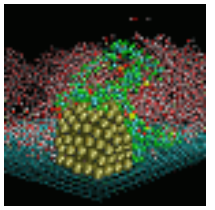




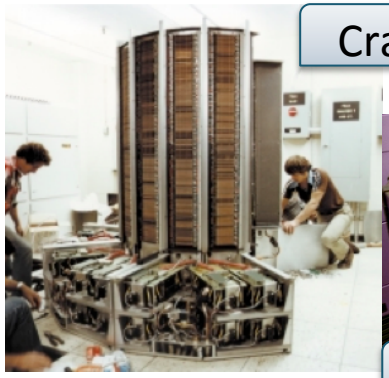








NERSC History



Cray 1 - 1978

Cray 2 – 1985

Cray T3E Mcurie - 1996

IBM Power3 Seaborg - 2001

1974	Founded at Livermore to support fusion research with a CDC system
1978	Cray 1 installed
1983	Expanded to support today's DOE Office of Science
1986	ESnet established at NERSC
1994	Cray T3D MPP testbed
1994 - 2000	Transitioned users from vector processing to MPP
1996	Moved to Berkeley Lab
1996	PDSF data intensive computing system for nuclear and high energy physics
1999	HPSS becomes mass storage platform
2006	Facility wide filesystem
2010	Collaboration with JGI

U.S. DEPARTMENT OF ENERGY | Office of Science

BERKELEY LAB

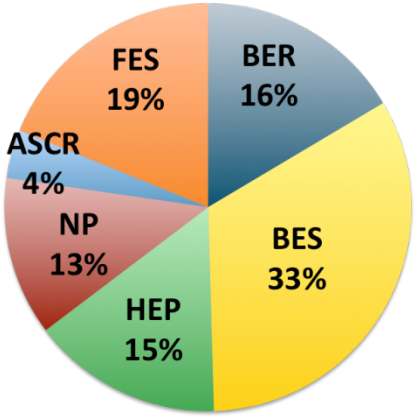
NERSC collaborates with computer companies to deploy advanced HPC and data resources

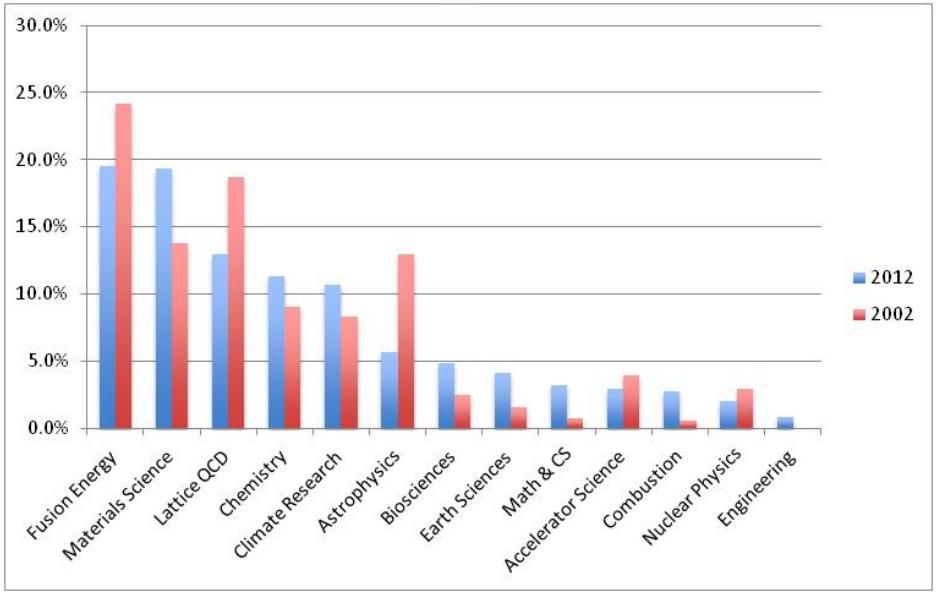
- Hopper (N6) and Cielo (ACES) were the first Cray petascale systems with a Gemini interconnect
- Edison (N7) is the first Cray petascale system with Intel processors, Aries interconnect and Dragonfly topology (serial #1)
- N8 and Trinity (ACES) are being jointly designed as on-ramps to exascale
- Architected and deployed data platforms including the largest DOE system focused on genomics
- One of the first facility-wide filesystems

We employ experts in high performance computing, computer systems engineering, data, storage and networking

We directly support DOE's science mission

- **We are the primary computing facility for DOE Office of Science**
- **DOE SC allocates the vast majority of the computing and storage resources at NERSC**
  - Six program offices allocate their base allocations and they submit proposals for overtargets
  - Deputy Director of Science prioritizes overtarget requests
- **Usage shifts as DOE priorities change**



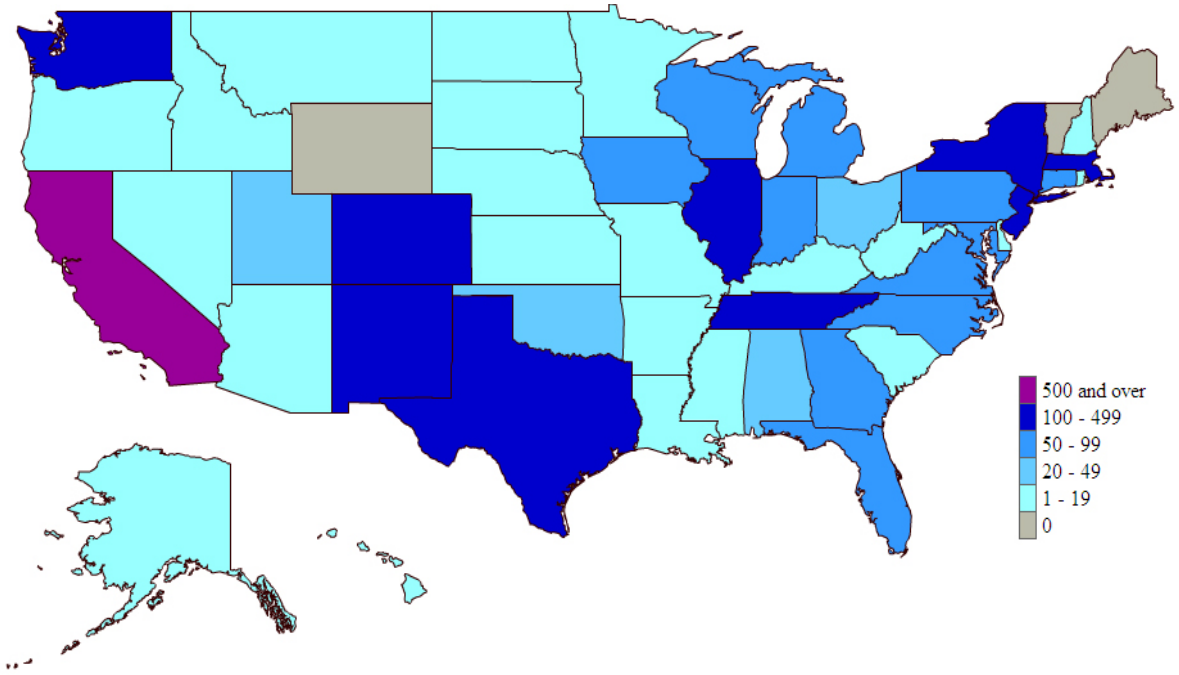


We focus on the scientific impact of our users

- 1500 journal publications per year
- 10 journal cover stories per year on average
- Simulations at NERSC were key to two Nobel Prizes (2007 and 2011)
- Supernova 2011fe was caught within hours of its explosion in 2011, and telescopes from around the world were redirected to it the same night
- Data resources and services at NERSC played important roles in one of Science Magazine's Top Ten Breakthroughs of 2012 — the measurement of the Θ_{13} neutrino weak mixing angle
- MIT researchers developed a new approach for desalinating sea water using sheets of graphene, a one-atom-thick form of the element carbon. Smithsonian Magazine's fifth "Surprising Scientific Milestone of 2012."
- Four of Science Magazine's insights of the last decade (three in genomics, one related to cosmic microwave background)

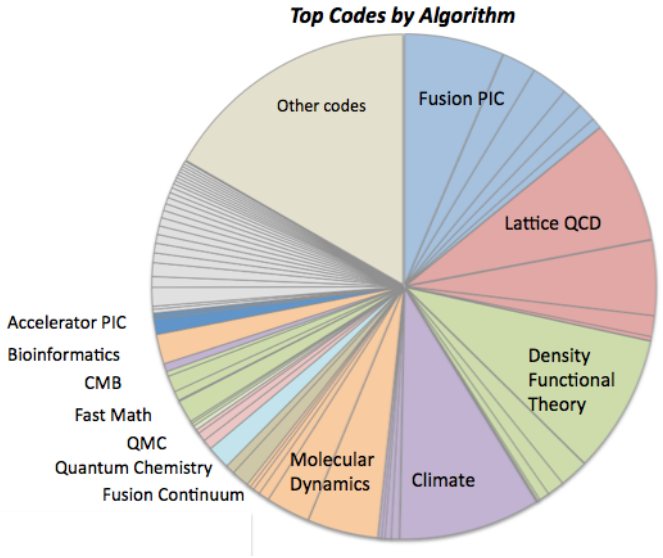
We support a broad user base

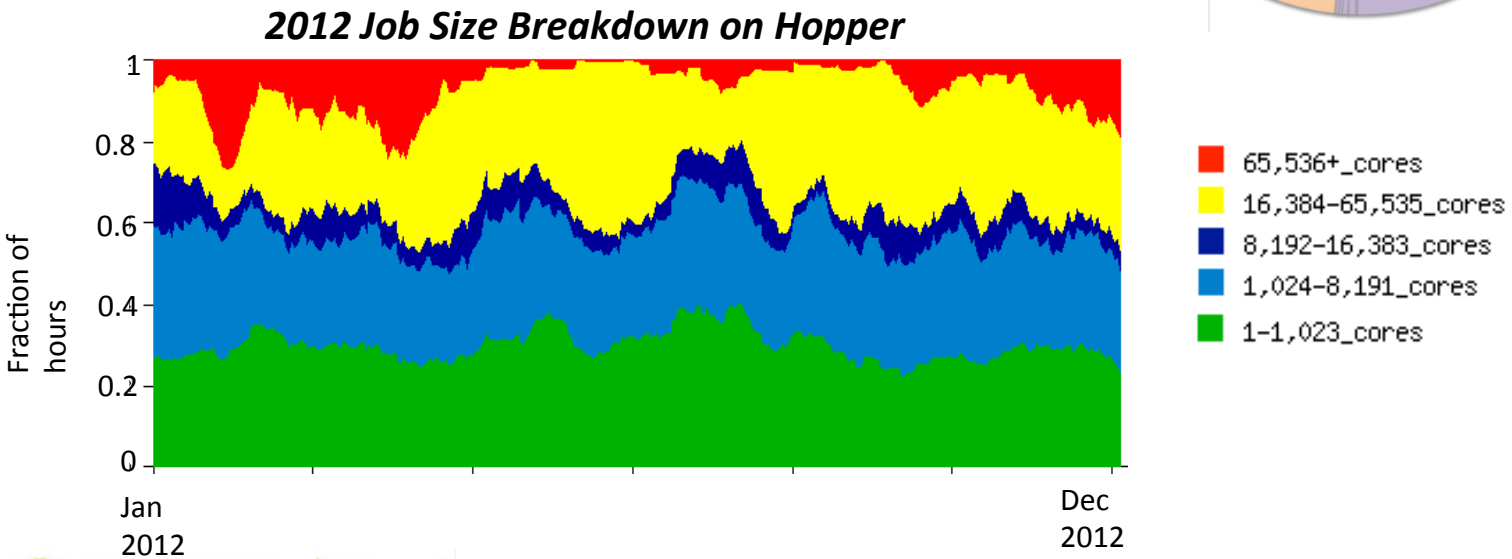
- **4500 users, and we typically add 350 per year**
- **Geographically distributed: 47 states as well as multinational projects**



We support a diverse workload

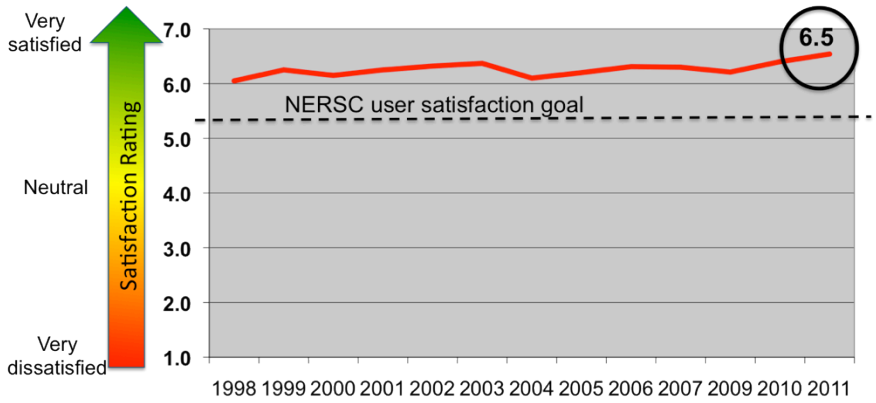
- **Many codes (600+) and algorithms**
- **Computing at scale and at high volume**

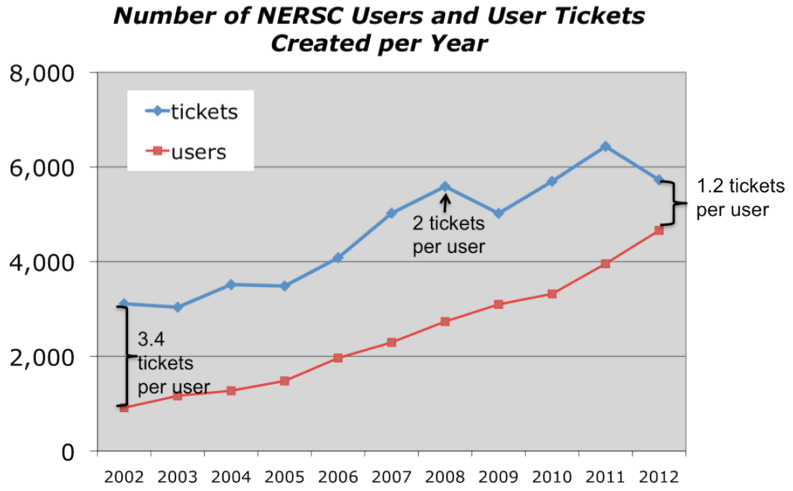




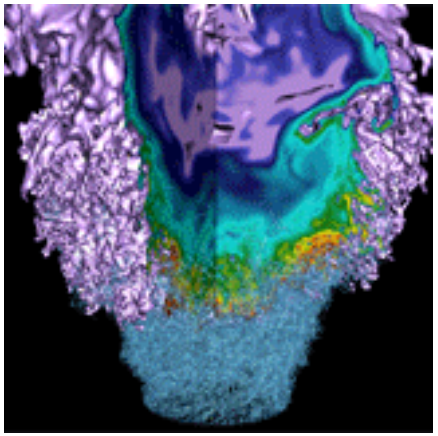
Our operational priority is providing highly available HPC resources backed by exceptional user support

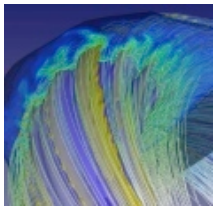
- **We maintain a very high availability of resources (>90%)**
  - One large HPC system is available at all times to run large-scale simulations and solve high throughput problems
- **Our goal is to maximize the productivity of our users**
  - One-on-one consulting
  - Training (e.g., webinars)
  - Extensive use of web pages
  - We solve or have a path to solve 80% of user tickets within three business days

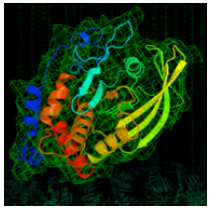


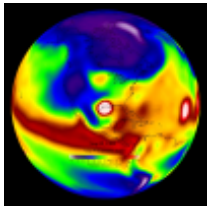


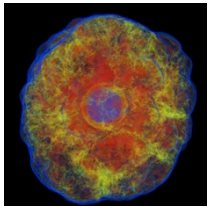
NERSC Today

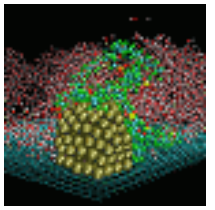




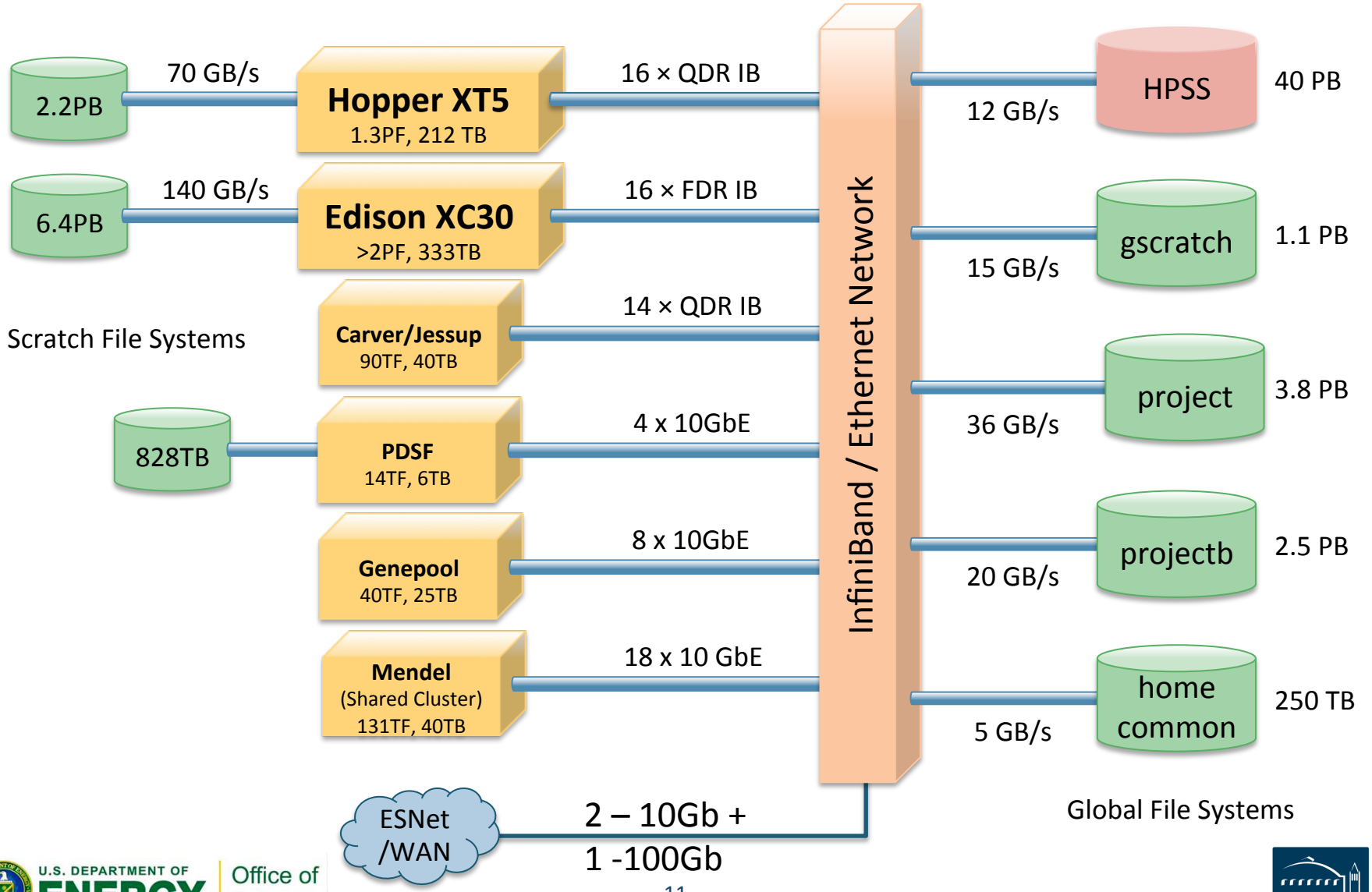








NERSC Systems



	Edison	Mira	Titan	Hopper
Peak Flops (PF)	2.4	10.0	5.26 (CPU) 21.8 (GPU)	1.29
CPU cores	124,800	786,432	299,008 (CPU) 18,688 (GPU's)	152,408
Frequency (GHz)	2.4	1.6	2.2 (CPU) 0.7 (GPU)	2.1
Memory (TB)	333	786	598 (CPU) 112 (GPU)	217
Memory/node (GB)	64	16	32 (CPU) 6 (GPU)	32
Memory BW* (TB/s)	530.4	1406	614 (CPU) 3,270 (GPU)	331
Memory BW/ node* (GB/s)	102	29	33 (CPU) 175 (GPU)	52
Filesystem	6.4 PB 140 GB/s	35 PB 240 GB/s	10 PB 240 GB/s	2 PB 70 GB/s
Peak Bisection BW (TB/s)	11.0	24.6	11.2	5.1
Peak Bisection BW/node (GB/s)	2.12	0.50	0.60	0.80
Sq ft	1200	~1500	4352	1956
Power (MW Linpack)	2.10	3.95	8.21	2.91

* STREAM

The Computational Research and Theory (CRT) building will be the home for NERSC-8

- **Four story, 140,000 GSF**
  - Two 20Ksf office floors, 300 offices
  - 20K -> 29Ksf HPC floor
  - Mechanical floor
  - 42 MW to building, 12.5 initially provisioned
- **Natural air conditioning (PUE<1.1)**

NERSC-8 Mission Need

The Department of Energy Office of Science requires an HPC system to support the rapidly increasing computational demands of the entire spectrum of DOE SC computational research.

- Provide a significant increase in computational capabilities, at least 10 times the sustained performance of the Hopper system on a set of representative DOE benchmarks
- Delivery in the 2015/2016 time frame
- Provide high bandwidth access to existing data stored by continuing research projects.
- Platform needs to begin to transition users to more energy-efficient many-core architectures.

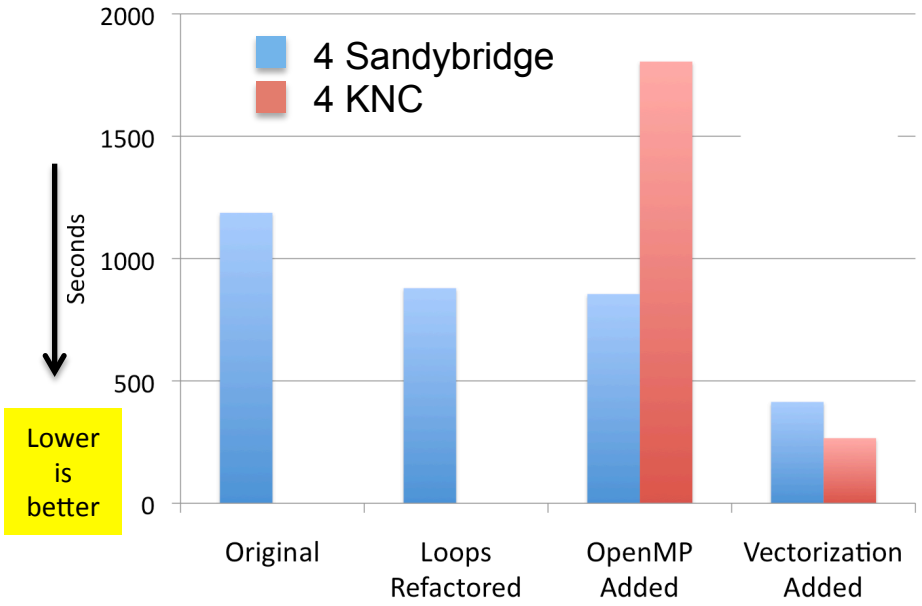
Although architecture for NERSC-8 is not yet known, trend is toward manycore processors

- **Regardless of chip vendor chosen for NERSC-8, users will need to modify applications to achieve performance**
- **Multiple levels of code modification may be necessary**
  - Expose more on-node parallelism in applications
  - Increase application vectorization capabilities
  - For co-processor architectures, locality directives must be added

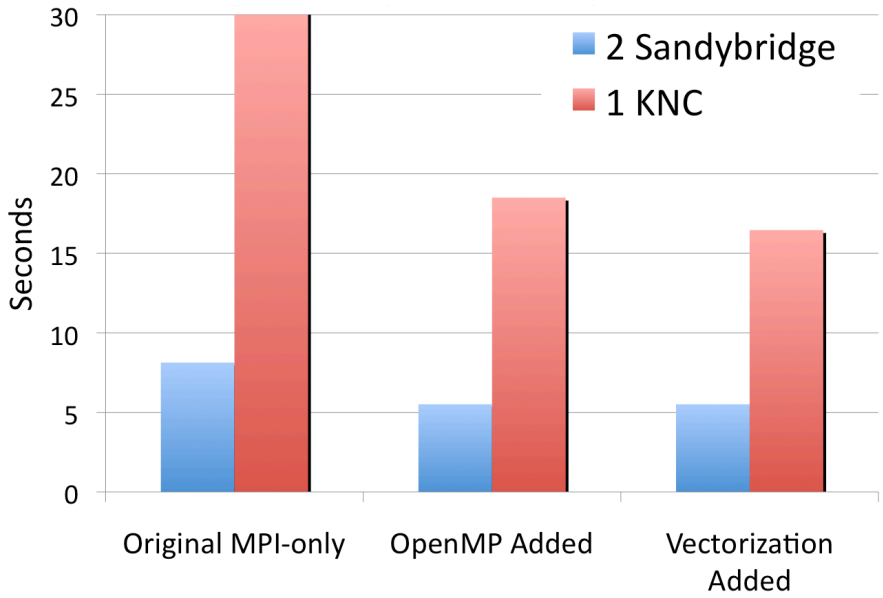
Application Readiness team is examining KNC

Some applications are well suited to the Knight's architecture, while others will need significant changes to achieve good performance.

Berkeley GW Kernel Performance on Knight's Corner (KNC)



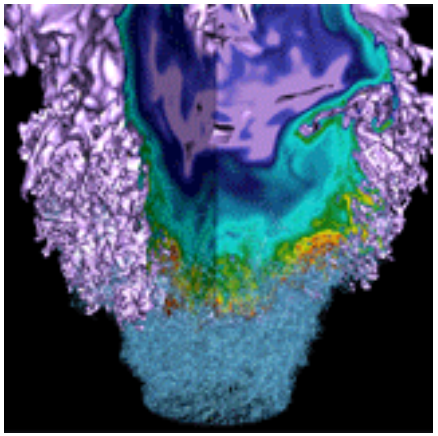
CSU Atmospheric Model Multigrid Solver on Knight's Corner (KNC)

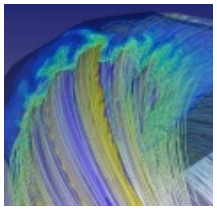


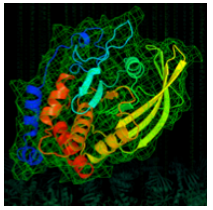
- BerkeleyGW kernel is example of code that can benefit from manycore architecture.
- Early prototype KNC hardware roughly equals performance of Sandybridge processor
- Optimizations for KNC improve performance on Sandybridge

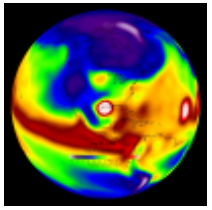
- Despite improvements from adding OpenMP and vectorization, this multigrid solver will need further restructuring to run on optimally on KNC

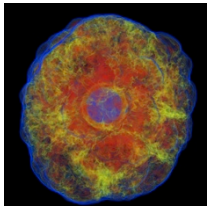
Forecasting

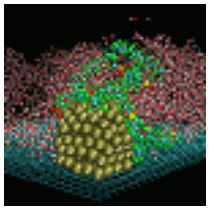






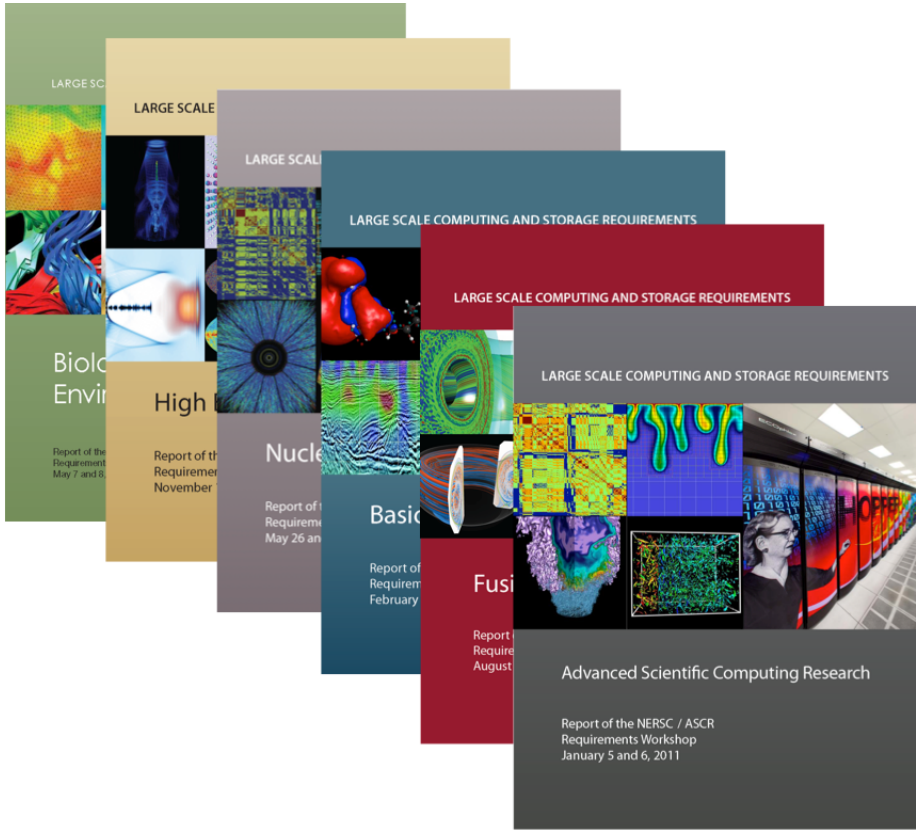






Requirements with six program offices

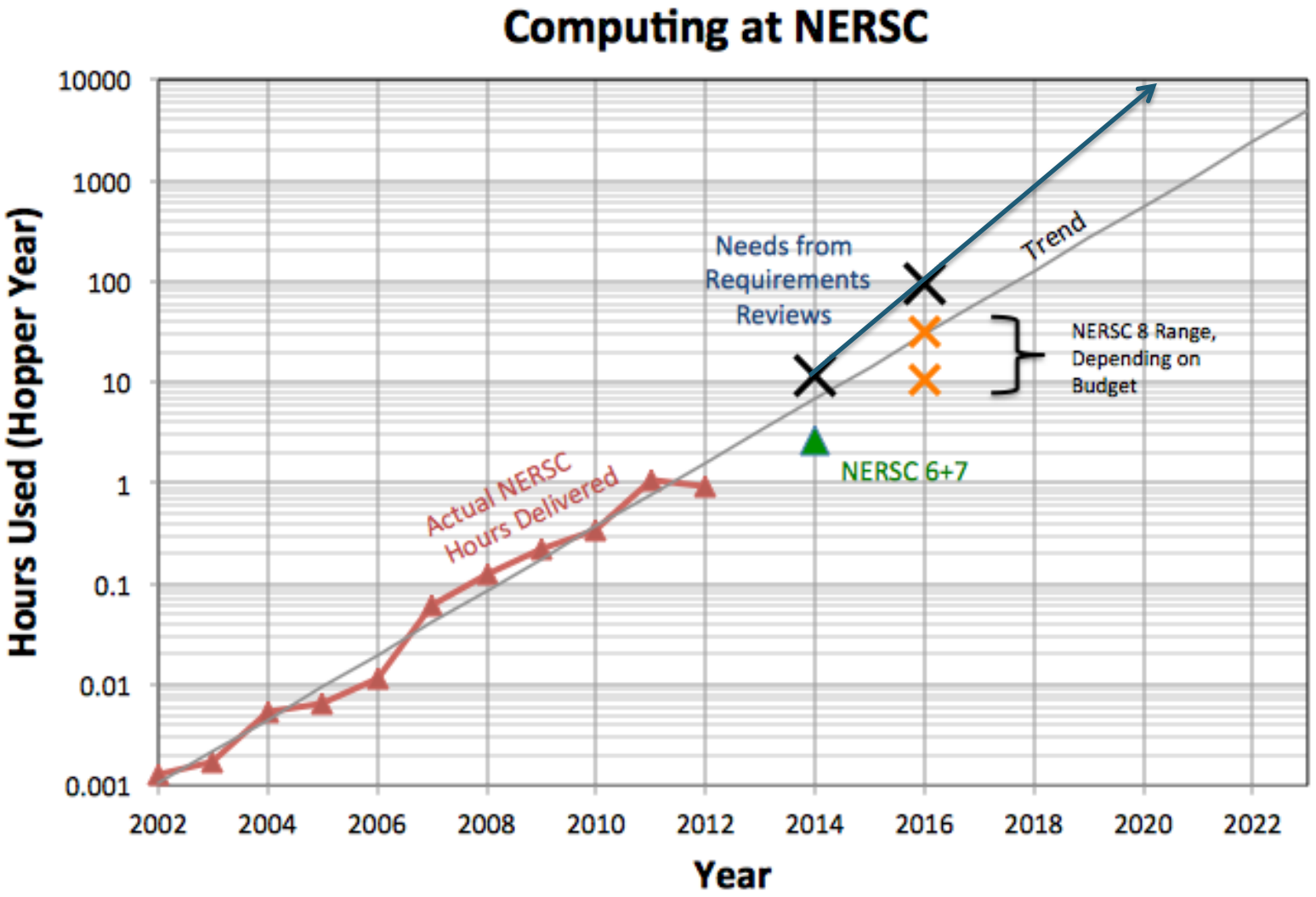
- Reviews with six program offices every three years
- Program managers invite representative set of users (typically represent >50% of usage)
- Identify science goals and representative use cases
- Based on use cases, work with users to estimate requirements
- Re-scale estimates to account for users not at the meeting (based on current usage)
- Aggregate results across the six offices
- Validate against information from in-depth collaborations, NERSC User Group meetings, user surveys



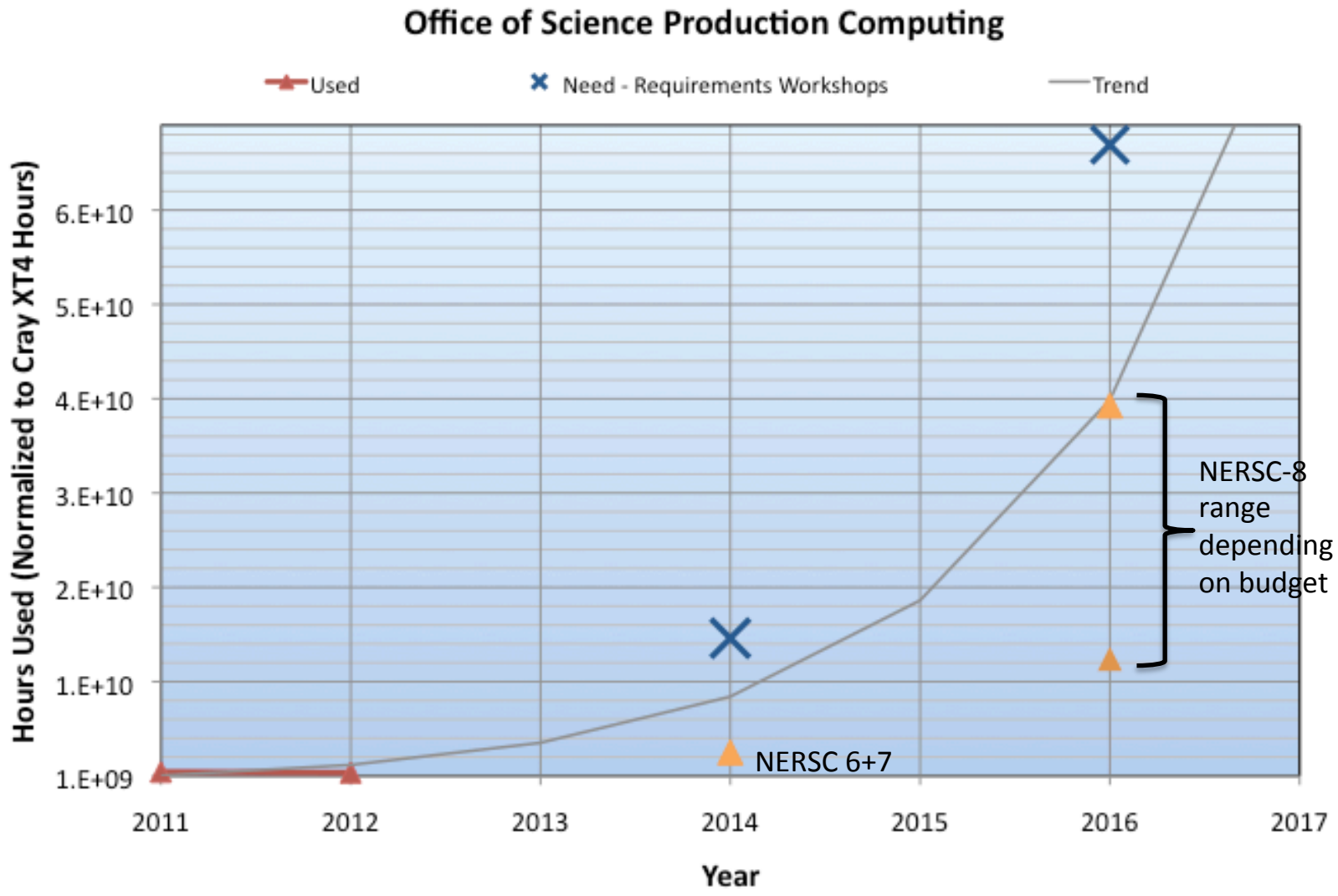
Tends to underestimate need because we are missing future users

http://www.nersc.gov/science/requirements-reviews/final-reports/

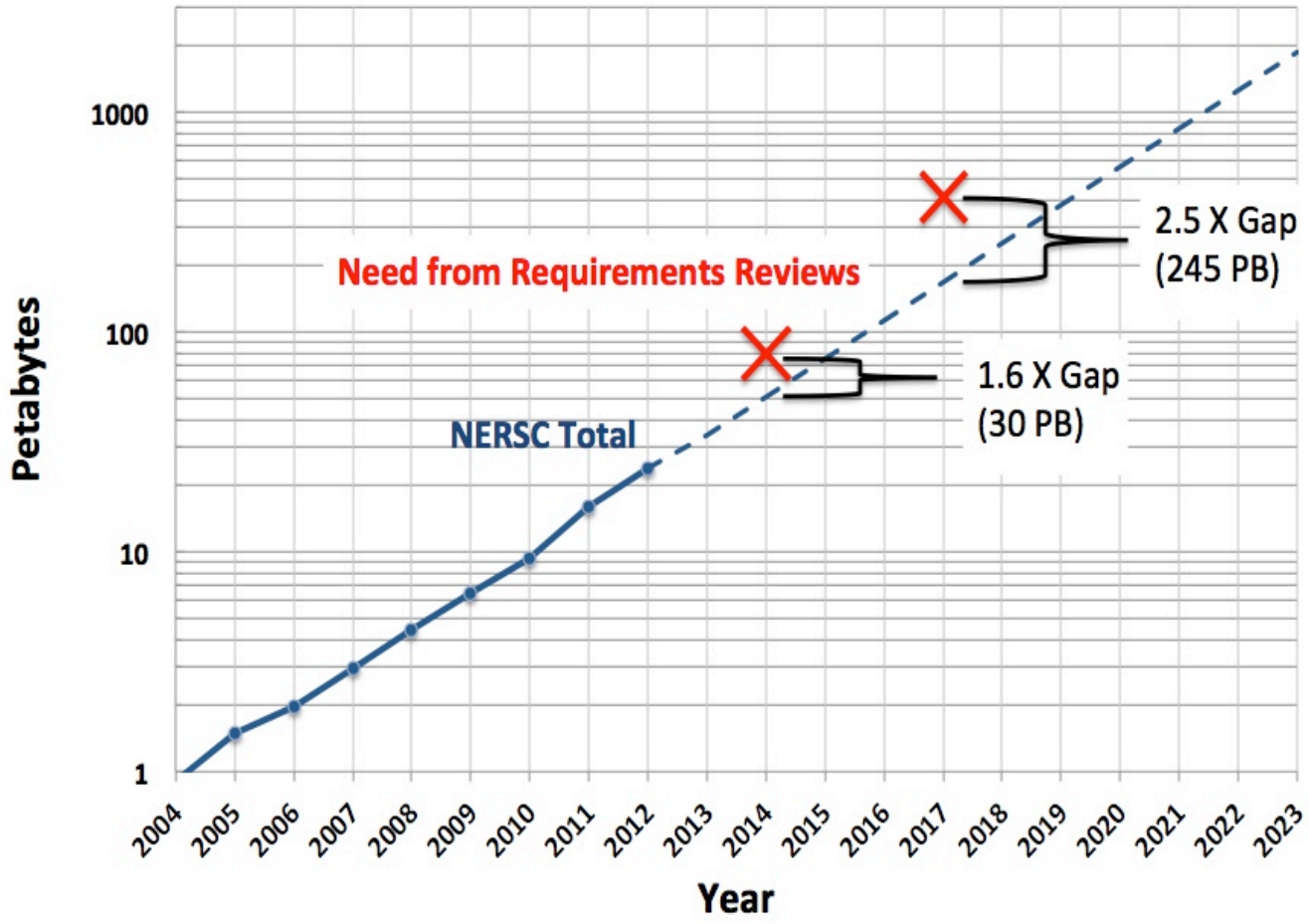
Keeping up with user needs will be a challenge



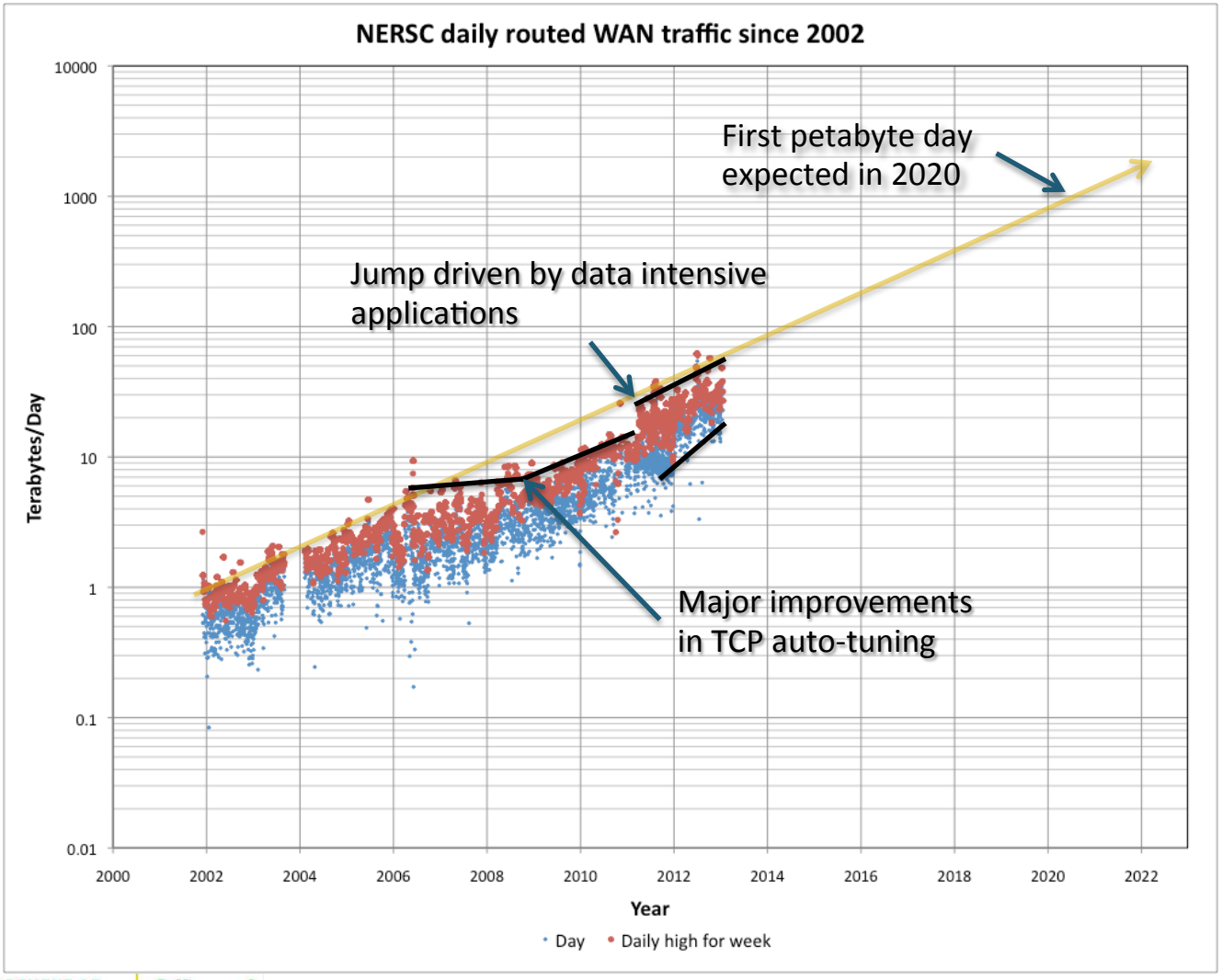
Keeping up with user needs will be a challenge (cont.)



Future archival storage needs

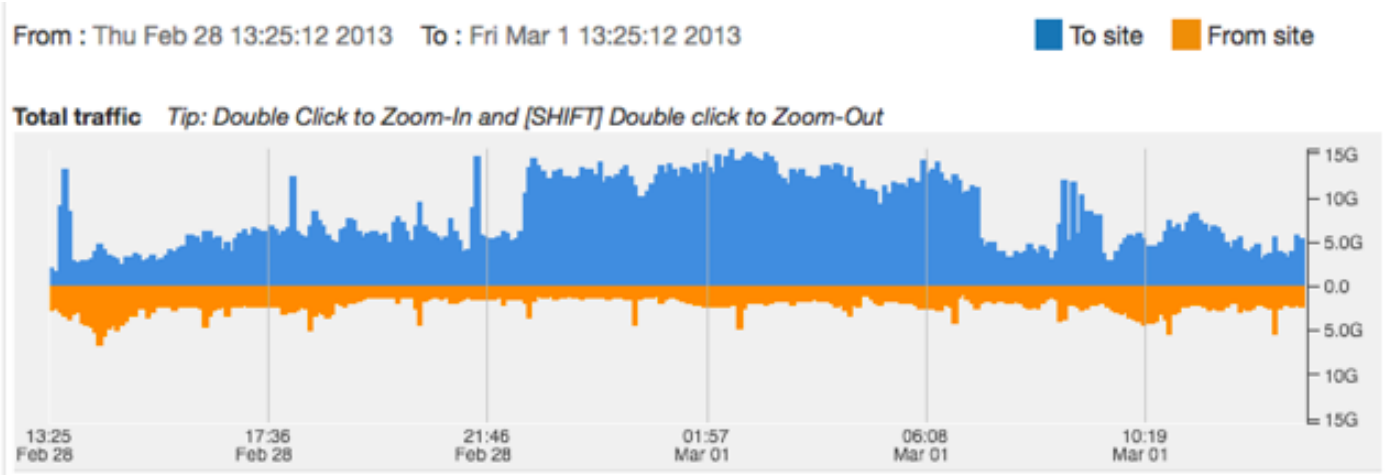


Exponentially increasing data traffic

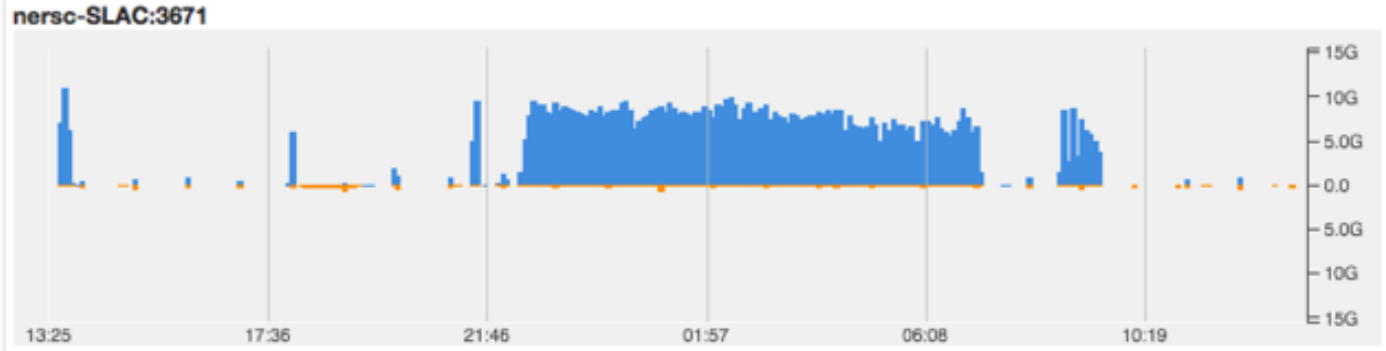


Cross Bay Data Transfer

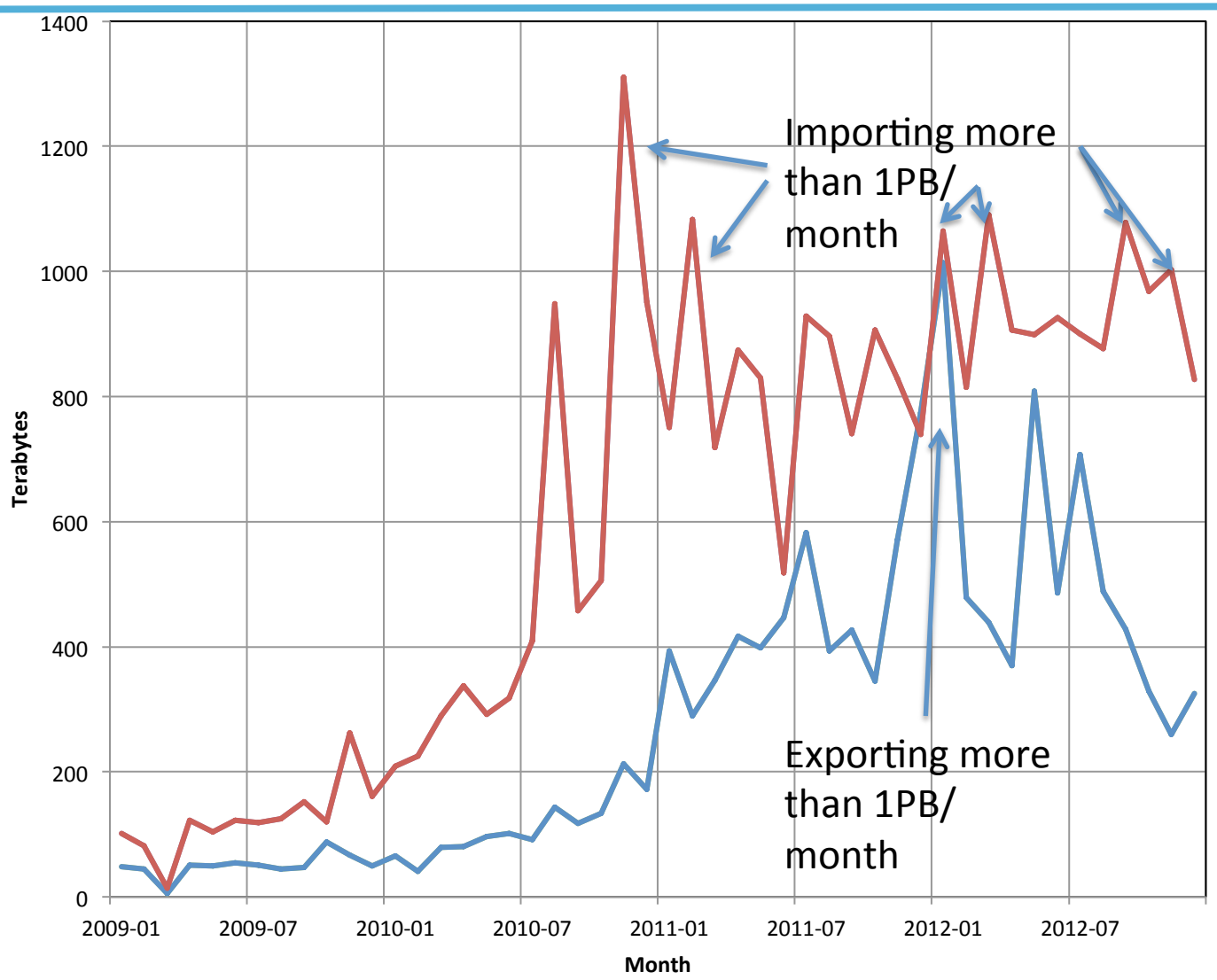
All NERSC Traffic



Photosystem II X-Ray Study

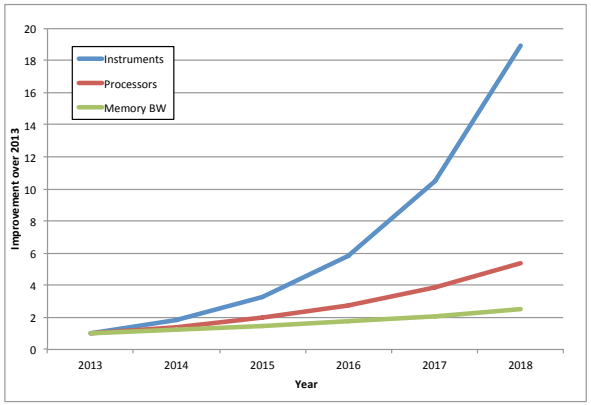


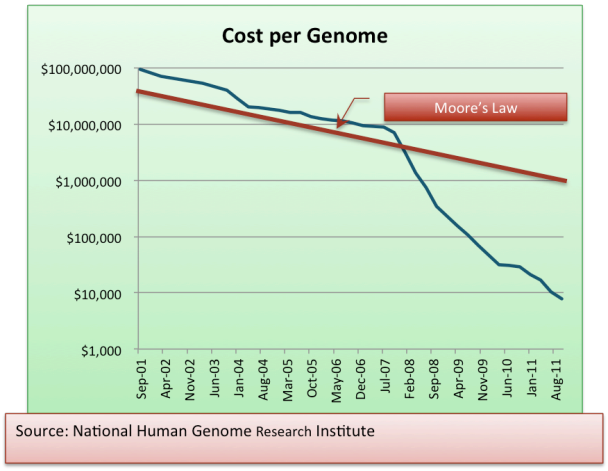
NERSC users import more data than they export!

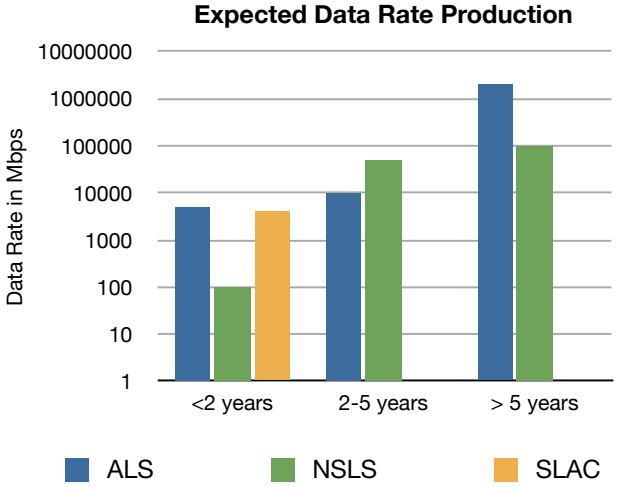


DOE experimental facilities are also facing extreme data challenges

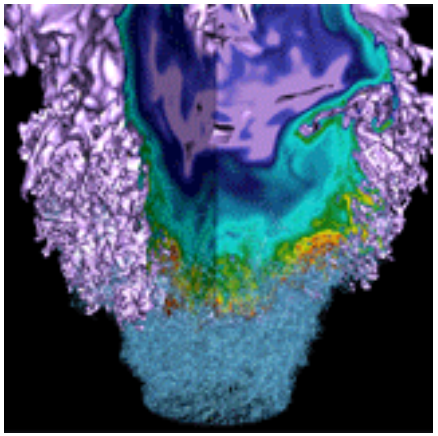
- **The observational dataset for the Large Synoptic Survey Telescope will be ~100 PB**
- **The Daya Bay project will require simulations which will use over 128 PB of aggregate memory**
- **By 2017 ATLAS/CMS will have generated 190 PB**
- **Light Source Data Projections:**
  - 2009: 65 TB/yr
  - 2011: 312 TB/yr
  - 2013: 1.9 PB /yr
  - EB in 2021?
  - NGLS is expected to generate data at a terabit per second



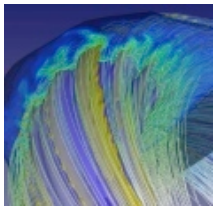
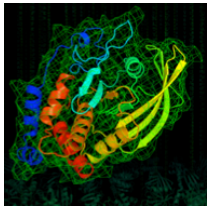
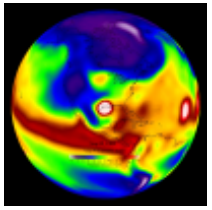
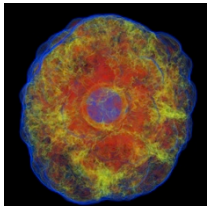
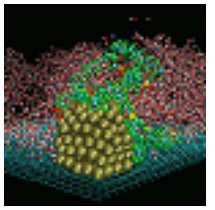




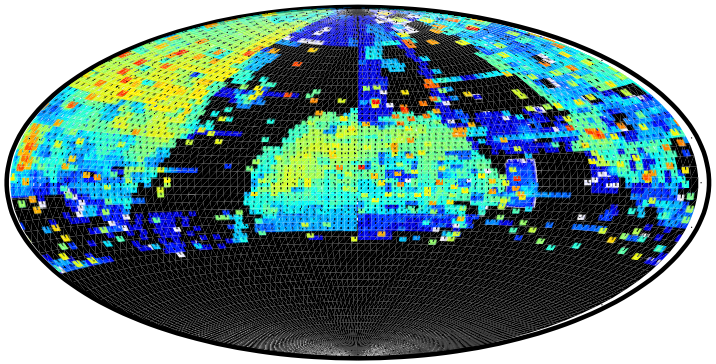
Data Analysis is Playing a Key Role in Scientific Discovery

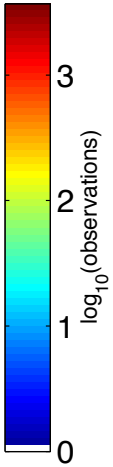


NERSC

Astrophysics





Palomar Transient Factory: Discovered over 2000 spectroscopically confirmed supernovae in the last 5 years, including the youngest and closest Type Ia supernova in past 40 years.

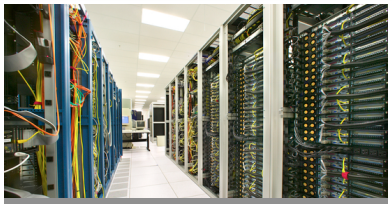
PI: Shri Kulkarni (Caltech)

67 refereed publications to-date including 2 in *Science Magazine* and 4 *Nature* articles. Processing pipeline runs on NERSC's Carver nightly and makes heavy use of the Science Gateway Nodes to share the data among the collaboration.

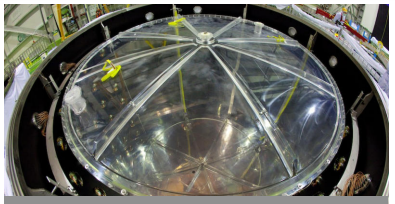
Solving the Puzzle of the Neutrino

- **HPC and ESnet vital in the measurement of the important "θ_{13}" neutrino parameter.**
  - Last and most elusive piece of a longstanding puzzle: why neutrinos appear to vanish as they travel
  - The result affords new understanding of fundamental physics; may eventually help solve the riddle of matter-antimatter asymmetry in the universe.
- **HPC for simulation / analysis; HPSS and data transfer capabilities; NGF and Science Gateways for distributing results**
  - All the raw, simulated, and derived data are analyzed and archived at a single site
  - => Investment in experimental physics requires investment in HPC.
- **One of Science Magazine's Top-Ten Breakthroughs of 2012**

The Daya Bay experiment counts antineutrinos at three detectors (shown in yellow) near the nuclear reactors and calculates how many would reach the detectors if there were no oscillation. transformation.



NERSC's PDSF cluster



Daya Bay detectors

U.S. DEPARTMENT OF ENERGY | Office of Science

HEP

Science | AAAS

PI: Kam-Biu Luk (LBNL)

BERKELEY LAB

The Planck Mission

- **A European Space Agency (+NASA) satellite mission to measure the temperature and polarization of the Cosmic Microwave Background.**
  - The echo of the Big Bang: primordial photons have seen it all.
  - Fluctuations encode all of fundamental physics & cosmology.
  - Planck results assumed by all Dark Energy experiments.

- **Realizing the full scientific potential of Planck requires very significant computing resources**
  - Tiny signal (μK - nK) requires huge data volume for sufficient S/N
  - 72 detectors sampling at 30-180Hz for 2.5 years => 10^{12} samples.
  - Analysis depends critically on Monte Carlo methods
    - Simulate and analyze 10^4 realizations of the entire mission!

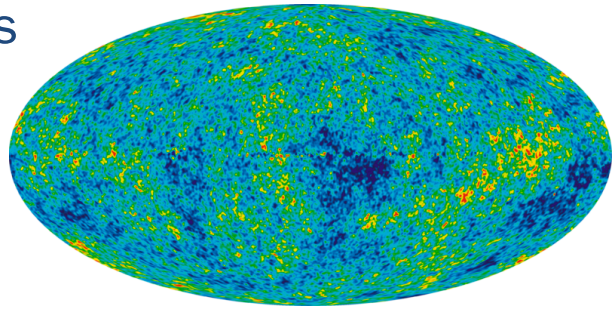
Planck At NERSC

- **NERSC has been a critical Planck resource for over a decade.**
  - Open access: 100+ data analysts from 20 nations & 3 continents.
  - State-of-the-art: new top-5 system every 2-3 years.
  - Reliable roadmap: 6 HPC generations, 1000x capability growth.
- **NASA/DOE MoU**
  - Uniquely guarantees access to NERSC through lifetime of Planck
    - essential for mission-class projects
  - Allows Planck to located dedicated resources on NERSC floor
    - initially a dedicated cluster; now a supercomputer cabinet
  - Strong symbiotic relationship between Planck & NERSC
    - beta-testing new hardware (NGF) & methods (pseudo-user)

March 21st 2013 – First Planck Results

Simulations for missing data

BAO maps, like CMB maps (shown) will miss data behind Milky Way. Required millions of map-making hours at NERSC.



Interpreting data

[redacted]

The tight coupling between modeling & simulation and experimental data analysis will be the norm as science complexity grows.

The New York Times

NEW YORK, FRIDAY, MARCH 22, 2013

PRESIDENT URGES ISRAELIS TO PUSH EFFORT FOR PEACE

APPEAL AIMED AT YOUNG

In Jerusalem, He Eases Stance on Settlement Halt Before Talks

The Cosmos, Back in the Day

Bronx Inspector, Secretly Taped, Suggests Race Is a Factor in Stops

Once Few, Women Hold More Power in Senate

Fast-Growing Brokerage Firm Often Tangles With Regulators

Mood Darkens in Cyprus Ahead of Bailout Deadline

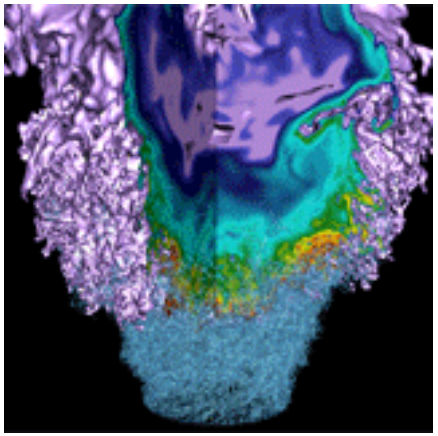
O Revelations! Letters, Once Banned, Flesh Out Willa Cather

Cartier

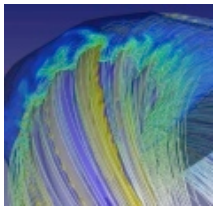
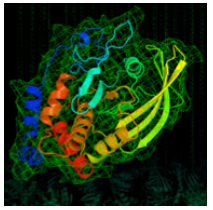
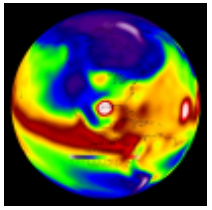
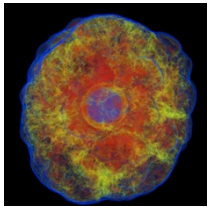
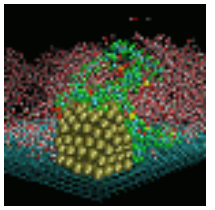
U.S. DEPARTMENT OF ENERGY | Office of Science

BERKELEY LAB
Lawrence Berkeley National Laboratory

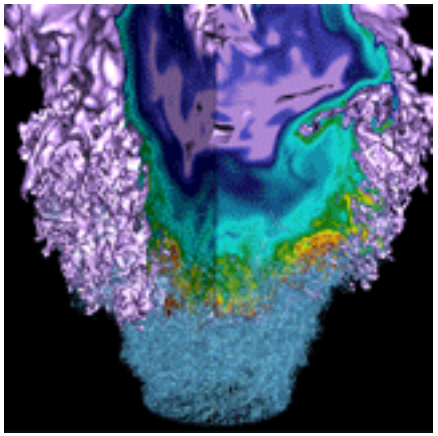
Exascale and Big Data Face the same Computing Challenges



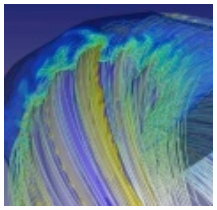
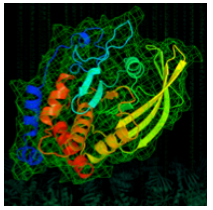
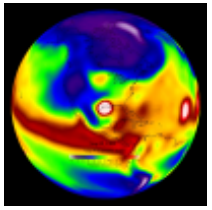
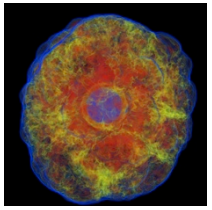
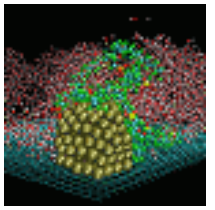
NERSC

Data Deluge at Experimental Facilities

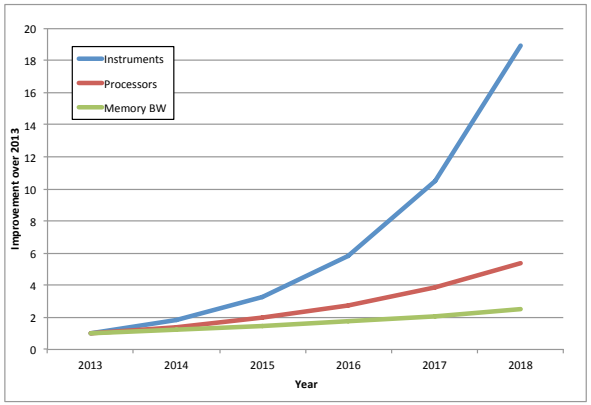


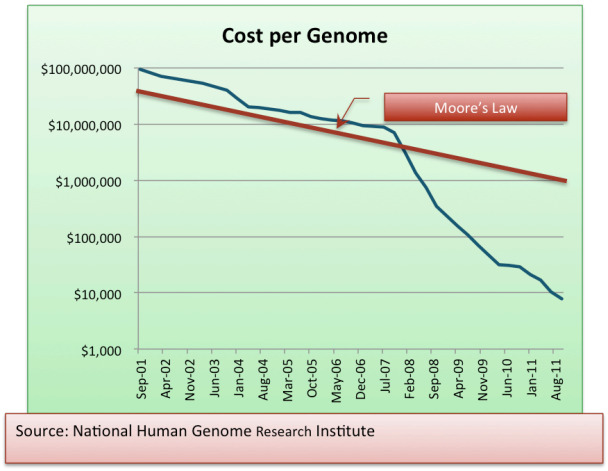
NERSC

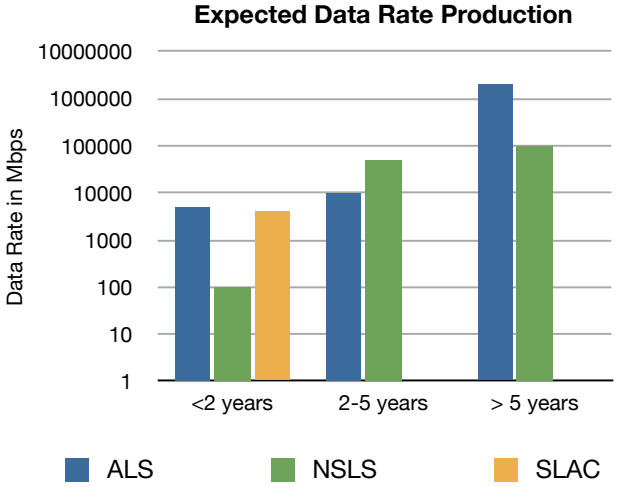






DOE experimental facilities are also facing extreme data challenges

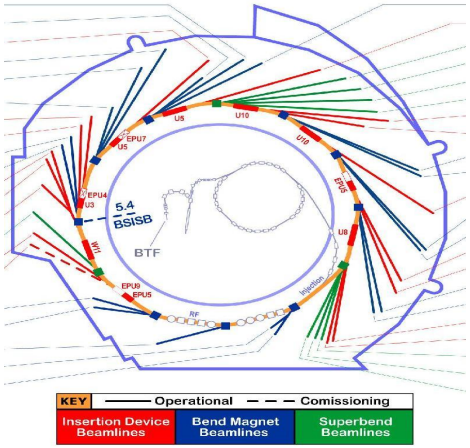
- **The observational dataset for the Large Synoptic Survey Telescope will be ~100 PB**
- **The Daya Bay project will require simulations which will use over 128 PB of aggregate memory**
- **By 2017 ATLAS/CMS will have generated 190 PB**
- **Light Source Data Projections:**
  - 2009: 65 TB/yr
  - 2011: 312 TB/yr
  - 2013: 1.9 PB /yr
  - EB in 2021?
  - NGLS is expected to generate data at a terabit per second



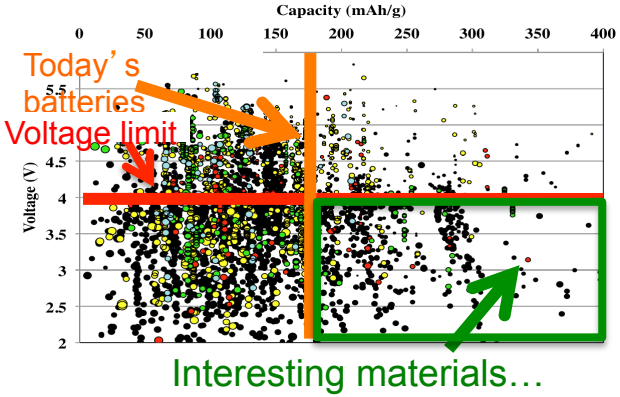




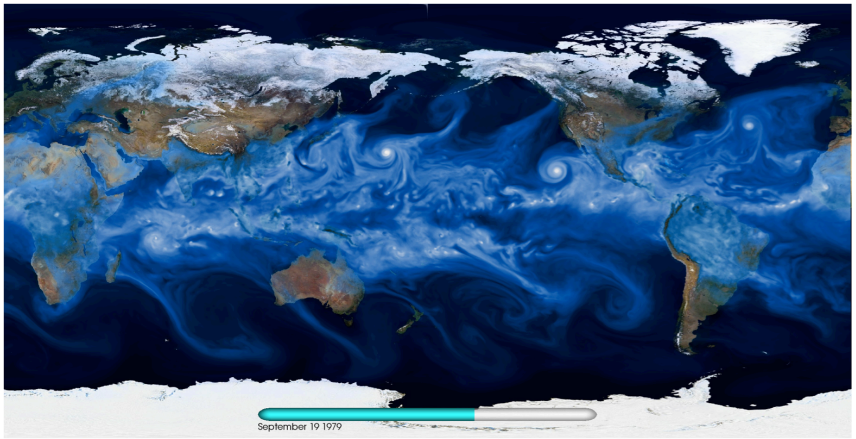
NERSC has an expanding set of data science engagements



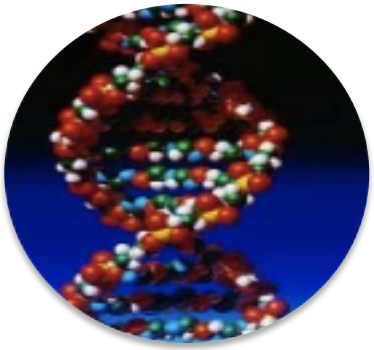
Data Pipeline for ALS



Materials genome

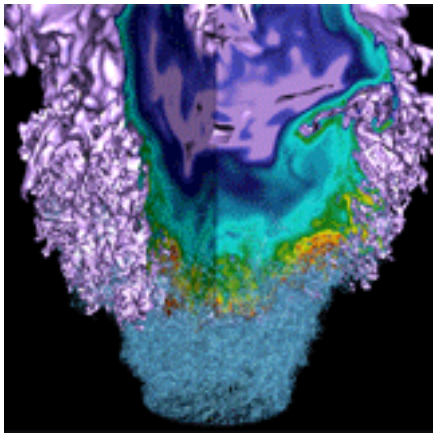


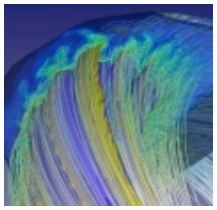
Fusion of simulations and measurements

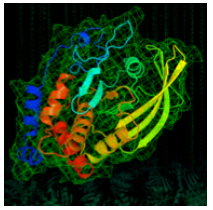


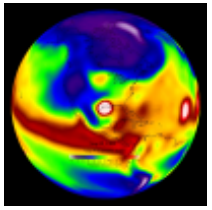
JGI Database now has over 3 Billion Genes

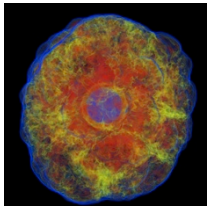
NERSC Strategy

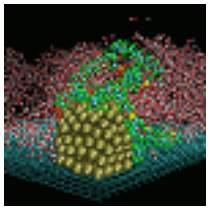








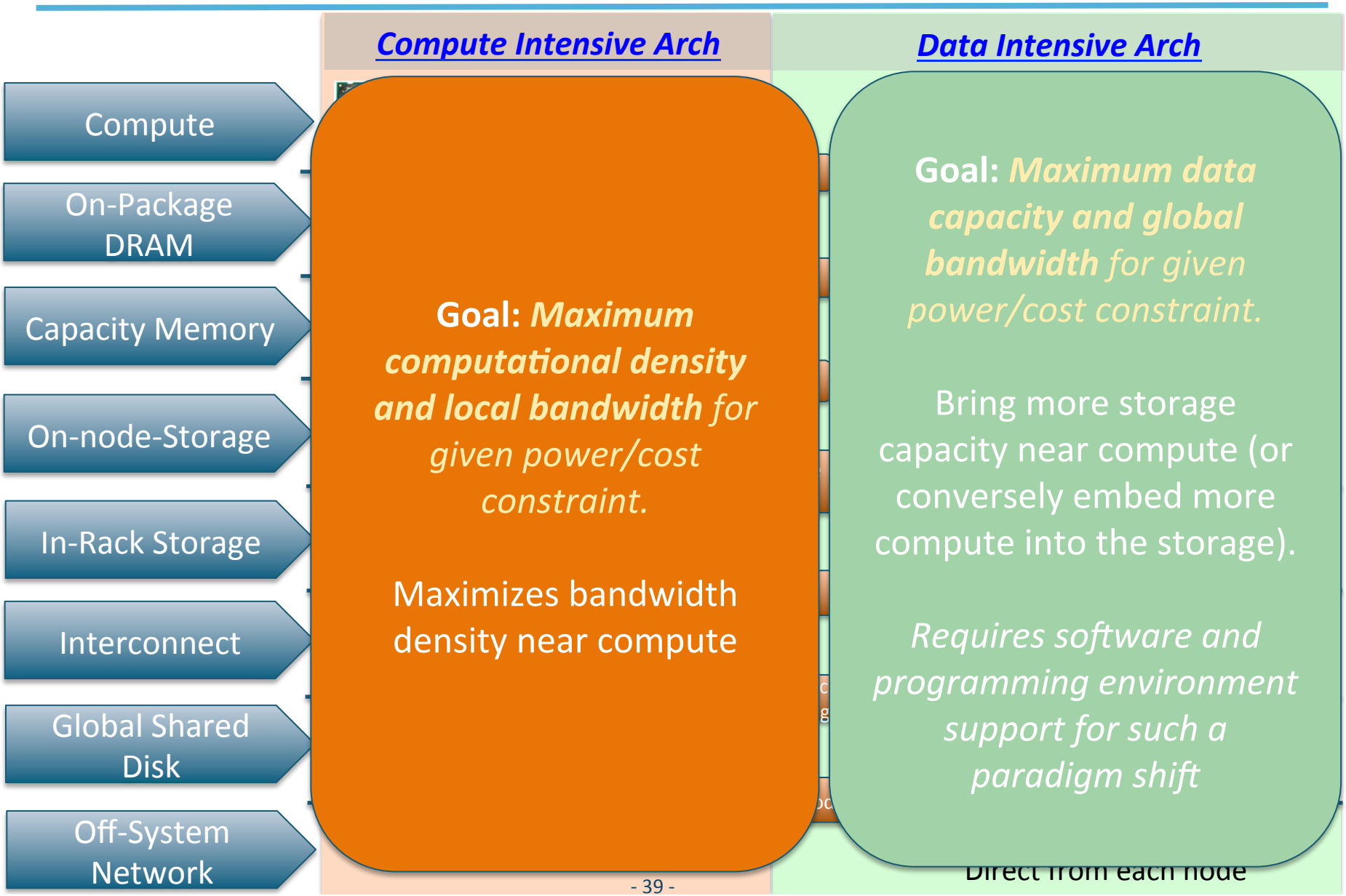




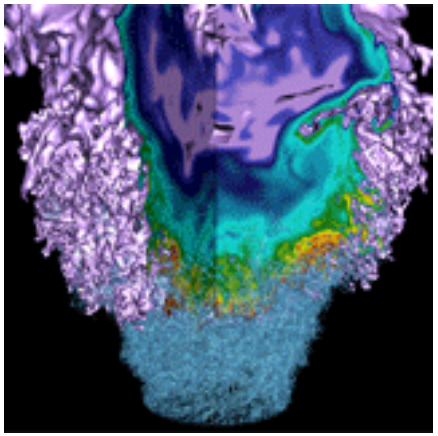
Strategic Objectives

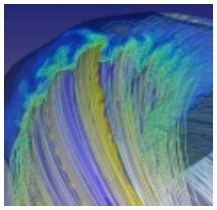
- **Meet the ever-growing computing and data needs of our users by**
  - providing usable exascale computing and storage systems
  - transitioning SC codes to execute effectively on manycore architectures
  - influencing the computer industry to ensure that future systems meet the mission needs of SC
- **Increase the productivity, usability, and impact of DOE's user facilities by providing comprehensive data systems and services to store, analyze, manage, and share data from those facilities**

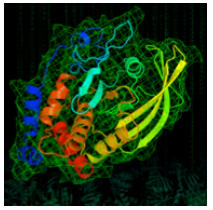
Unique data-centric resources will be needed

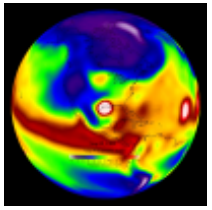


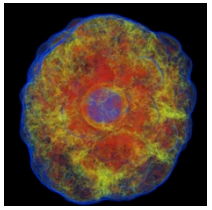
NERSC System Plan

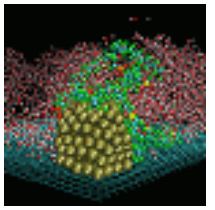












Projections of Installed Capacity