# NERSC
## Strategic Plan for FY2014–2023


BERKELEY LAB

NERSC

U.S. DEPARTMENT OF ENERGY
Office of Science

# National Energy Research Scientific Computing Center (NERSC) Strategic Plan for FY2014–2023

May 1, 2013

This work was supported by the Director, Office of Science, Office of Advanced Scientific Computing Research of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

# Table of Contents

1. Facility Overview ........ 1
2. Mission, Goals, Drivers, and Strategic Connections ........ 2
2.1 Vision and Mission Statements ........ 2
2.2 Facility Strategic Goals ........ 3
2.3 Connections with DOE and ASCR Strategic Plans ........ 5
2.4 Science Drivers ........ 6
3. Facility SWOT and Dependencies ........ 8
3.1 and 3.2 Strengths, Weaknesses, Opportunities, and Threats ........ 8
3.3 Risks, Dependencies, and Critical Gaps ........ 8
4. Science and Community Needs ........ 9
4.1 User Community Description ........ 9
4.2 User Community Engagement ........ 11
4.2.1 Consulting and Account Support ........ 11
4.2.2 Deep Engagements ........ 12
4.2.3 User Survey ........ 13
4.2.4 NERSC Users Group (NUG) ........ 13
4.2.5 Training ........ 13
4.3 Requirements Gathering ........ 14
4.3.1 Data Gathering ........ 14
4.3.2 Data Forecasting ........ 15
5. Center Operational Priorities ........ 15
5.1 Center Priorities ........ 15
5.2 Inter-Center and Inter-Facility Collaboration ........ 16
5.3 Connections with National Laboratory Plans ........ 17
6. Center Strategy for the Future ........ 17
6.1 Initiatives ........ 18
6.1.1 Transition the SC Workload to Energy-Efficient Architectures ........ 18
6.1.2 Ensure That Future Systems Meet SC Mission Needs ........ 19
6.1.3 Address the Most Demanding Data Challenges from DOE Facilities ........ 20
6.2 System Software and Libraries ........ 22
6.3 Technology Strategy ........ 23
6.3.1 Acquisition Strategies ........ 24
6.3.2 Partnership Strategies ........ 24

6.4 Power, Space, Cooling ........ 25
6.5 Summary Roadmap and Timeline ........ 26
6.5.1 Computational Systems ........ 26
6.5.2 Storage and Network Systems ........ 28
7. Human Resources ........ 28
7.1 Current Workforce ........ 28
7.2 Future Workforce ........ 29
7.3 Obstacles and Strategies ........ 29
8. Conclusion ........ 30
Appendix: Current NERSC Staffing ........ 31

# 1. Facility Overview

NERSC is the high-end scientific production computing facility for DOE's Office of Science. With more than 4,500 users from universities, national laboratories, and industry, NERSC supports the largest and most diverse research community of any computing facility within the DOE complex. NERSC provides large-scale, state-of-the-art computing, storage, and networking for DOE's unclassified research programs in high energy physics, biological and environmental sciences, basic energy sciences, nuclear physics, fusion energy sciences, mathematics, and computational and computer science.

NERSC was founded in 1974 at Lawrence Livermore National Laboratory and moved to Berkeley in 1996. Over its 16-year history at Berkeley Lab, NERSC has developed an outstanding reputation for providing both high-end computing systems and comprehensive scientific client services. NERSC successfully helped users transition from vector systems to massively parallel systems, and is now helping with the transition to multicore architectures. Results of the NERSC User Surveys show a high level of satisfaction with NERSC systems and services, and that translates into high scientific productivity: over the last five years, our users co-authored an average of 1,500 journal publications per year based on the computations performed at NERSC, making it DOE's most scientifically productive supercomputing user facility. On average, NERSC users publish ten journal cover stories per year. Two Nobel prizes have been awarded for research that included significant contributions from NERSC: the Physics Prize to Saul Perlmutter for discovering the accelerating expansion of the universe, and the Peace Prize to the team that produced the Intergovernmental Panel on Climate Change's Fourth Assessment Report.

Data and storage are playing an increasingly important role at NERSC. Data resources at NERSC played a major role in one of Science Magazine's Top Ten Breakthroughs of 2012 — the measurement of the $\Theta_{13}$ neutrino mixing angle. The international Planck collaboration's creation of the most detailed maps yet of the cosmic microwave background, based on trillions of observations, depended on both the supercomputers and the data resources at NERSC. In addition to large-scale high performance systems, NERSC deploys a variety of data-centric systems for its users. Since 1996, the High Energy and Nuclear Physics communities and NERSC have deployed large clusters, called PDSF, to help scientists analyze large data sets. In 2010 the Joint Genome Institute (JGI) and NERSC entered into a partnership to deploy the Genepool data cluster, enabling researchers to sequence and analyze over 50 terabases (approximately 20,000 human genomes) this year alone.

NERSC's role within the Office of Science is unique:

- NERSC directly supports DOE's science mission and is the Office of Science's primary computing facility. DOE program managers across the office allocate most of the center's computing resources. Each office receives a base allocation that it can use to advance its strategy. Offices request over-targets that are prioritized by the DOE Deputy Director of Science. NERSC usage directly reflects DOE's strategic priorities. Figure 1 shows that relative usage by the applied sciences has grown in the last ten years.
- Future computing and data requirements are derived directly from program needs. Requirement reviews are held with each program office. Program offices invite a representative subset of their scientists to participate. Based on information from these reviews, NERSC works with the program managers to estimate their overall need in a particular year. It is very clear from these reviews that it will be extremely challenging to meet the computing needs of our users in the next ten years.
- NERSC supports the largest number of users of any computing facility within the DOE, and our net growth is 350 users per year. In addition to one-on-one consulting services and in-depth collaborations, we must deploy scalable methods to effectively support all of our users. For example, we make extensive use of training and the worldwide web.
- Our primary focus is maximizing the scientific productivity of our users. To achieve high productivity, we balance computing, storage, and networking resources at our center. In addition,

we deploy flexible resources to meet the wide-ranging needs of our users. For example, projects often need to run capability simulations that span a significant fraction of our largest resource as well as high volumes of smaller simulations.

- NERSC has served as a data center for DOE scientists. For the last four years, we have been a net importer of data. About a petabyte of data is typically transferred to NERSC every month for storage, analysis, and sharing.

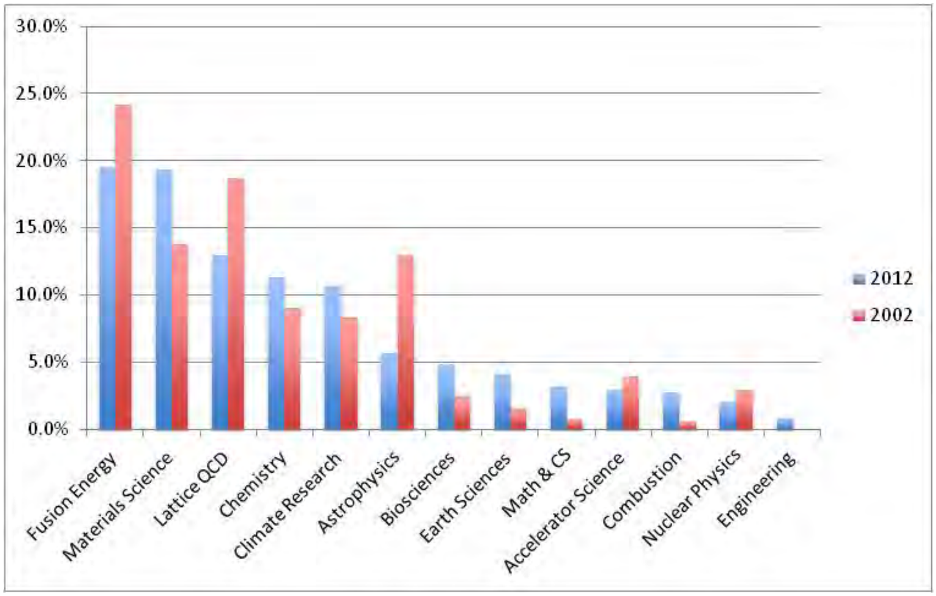



**Figure 1.** Changes in percent of hours used by science areas from 2002 to 2012 to support evolution of DOE's science mission. DOE program managers allocate most of the computing and data resources at NERSC.

In order to meet the future computing and storage needs of our users, NERSC will be moving to the Computational Research and Theory (CRT) facility in early 2015, where it will be co-located with ESnet and Berkeley Lab's Computational Research Division. CRT will be a highly energy-efficient, state-of-the-art computing facility that can provide over 40 MW of power and 30,000 square feet of space for computing and storage. The additional power and space within CRT will allow NERSC to deploy pre-exascale and exascale systems that are usable by the broad scientific community, and to meet the exponentially growing data needs of DOE.

# 2. Mission, Goals, Drivers, and Strategic Connections

## 2.1 Vision and Mission Statements

**NERSC's mission** is to accelerate scientific discovery at the DOE Office of Science (SC) through high performance computing and extreme data analysis.

**NERSC's vision** is to advance DOE Office of Science's mission by:

- Meeting 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, and influencing the computer industry to ensure that future systems meet the mission needs of SC.
- Increasing the productivity, usability, and impact of SC's data intensive science by providing comprehensive data systems and services to store, analyze, manage, and share data.

## 2.2 Facility Strategic Goals

NERSC requirement reviews with SC program offices have documented the science need for greatly expanded computing and storage capabilities. **The aggregate computing needs of SC science teams at NERSC are well into the exascale regime by the end of the decade (Figure 2).** Science teams need to run simulations at hundreds of petaflops, and they need to run thousands to millions of petascale simulations. These computing needs will also greatly stress our ability to store, manage, move, and analyze data. Future archival storage requirements are shown in Figure 3, and they greatly outpace NERSC's traditional growth rate.

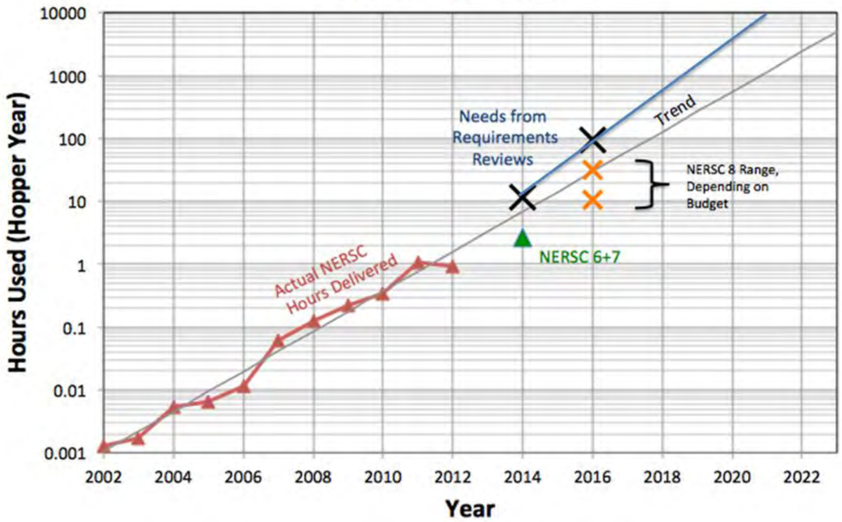



**Figure 2.** Projection of NERSC computing capability and the computational needs for 2014 and 2016 identified in the requirements reviews NERSC holds with Office of Science program offices. The plot is normalized to Hopper, which was NERSC's first petascale system. User needs will reach exascale in 2018.

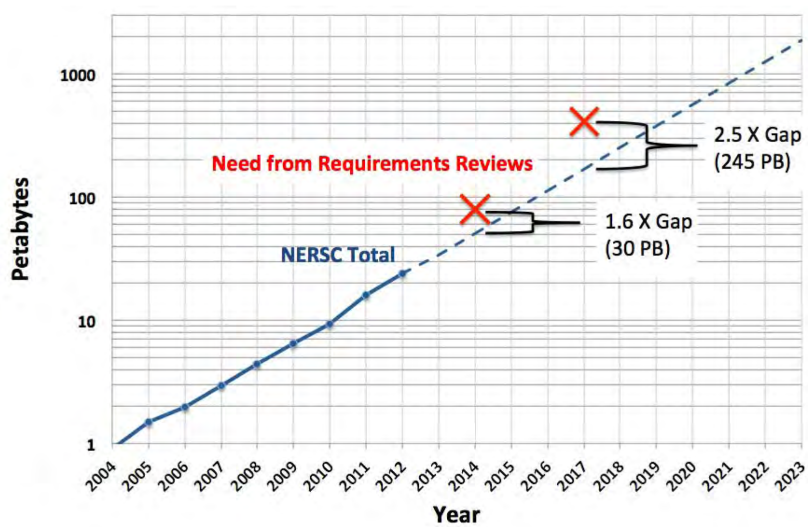



**Figure 3.** Projection of NERSC storage capacity and the storage needs for 2014 and 2017 identified in the requirements reviews. User needs will reach an exabyte by 2021.

The challenges with reaching exascale computing are well documented. Breakthroughs are needed in energy efficiency to field exascale systems that are affordable from a power standpoint. Applications will need to change to execute effectively on energy-efficient architectures. This technology transition will impact all scales of computing, because the technology disruption is in processors and memory, which are the building blocks of any computer system.

NERSC's strategic goals in this area are:

- Deploy usable exascale systems
- Transition SC science codes to energy-efficient architectures
- Influence the computer industry so that future systems meet the mission needs of SC.

NERSC's systems must be usable by a broad community, so they must be general purpose. Our pre-exascale systems will be NERSC-8 in 2016 and NERSC-9 in 2019. We anticipate deploying our first exascale system, NERSC-10, in 2022. We plan to provide a smooth progression to exascale for our users. Our programming models strategy is described further in Section 6.

Transitioning our science codes to energy-efficient architectures will be a challenge, and NERSC will build on its success in transitioning users from vector supercomputers to massively parallel systems in the 1990s. We are deploying testbeds that are representative of important technology trends. We will perform in-depth collaborations with science teams to implement their codes or kernels from their codes on these testbeds. These engagements will begin transitioning a dozen codes and will help NERSC evaluate technology choices. Based on this experience, as well as on collaborations with co-design centers and ASCR research projects, we will develop training and education programs for our vast community of over 4,500 researchers, including online training modules specific to code migration to manycore architectures. We will make highly tuned libraries available to our users to help with this migration. In addition, we will expand our exceptional user services to provide in-depth algorithm consulting on demand.

It is also imperative that we influence the computer companies to ensure that future products meet the mission needs of SC. NERSC's strategy for influencing industry involves:

- Partnering with Los Alamos and Sandia on procurements in 2016 (NERSC-8) and 2019 (NERSC-9). The NERSC-8/Trinity procurement is well under way. The larger size of this combined procurement has greatly increased the interest of computer companies.
- Provide industry with greater information on NERSC's workload through new and innovative instrumentation, measurement, and analysis.
- Actively engage with industry through DOE's Fast Forward and Design Forward programs.
- Leverage the Berkeley/Sandia Computer Architecture Laboratory (CAL) that has been established by ASCR.
- Serve as a conduit for information flow between computer companies and our user community.

Another strategic objective is to increase the productivity, usability, and impact of **DOE's experimental user facilities and other data-intensive science** by providing comprehensive data systems and services to store, analyze, manage, and share data. These facilities are seeing an explosion in data because detectors are improving at rates much faster than Moore's Law. Figure 4 shows that data rates at light sources will reach terabytes per second. The Next Generation Light Source (NGLS) will need to process exabyte data sets in 2021–2023. The High Energy Physics community will need to move, store, and analyze 100 petabyte datasets within 5–10 years. As mentioned earlier, NERSC has a very successful partnership with the Joint Genome Institute. Figure 5 shows the precipitous drop in the cost of sequencing a genome, which has led to petabytes of sequencing data from the JGI; new technologies are emerging which promise to continue this trend.

Our strategic goals related to data include:

- Enabling science teams with the nation's largest data-intensive challenges to rely on NERSC to the same degree they already do for modeling and simulation.
- Addressing the data growth from DOE experimental facilities with high-bandwidth, high-volume, and low-response-time data systems.

- Making NERSC a home for DOE SC data, and enabling the entire scientific community to easily access, search, and collaborate on the data stored at NERSC.
- Developing advanced algorithms, capabilities, and services that significantly lower the barrier to data-intensive computing and enable the broad DOE SC community to tackle the most demanding data-intensive problems.

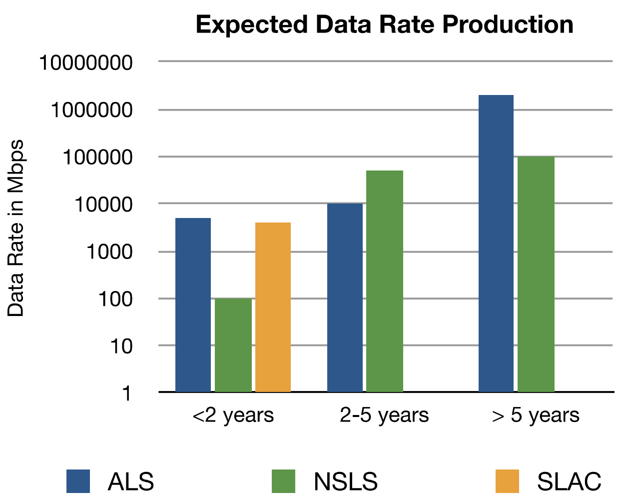



Figure 4. Data rates from experimental facilities are expected to reach terabytes per second.

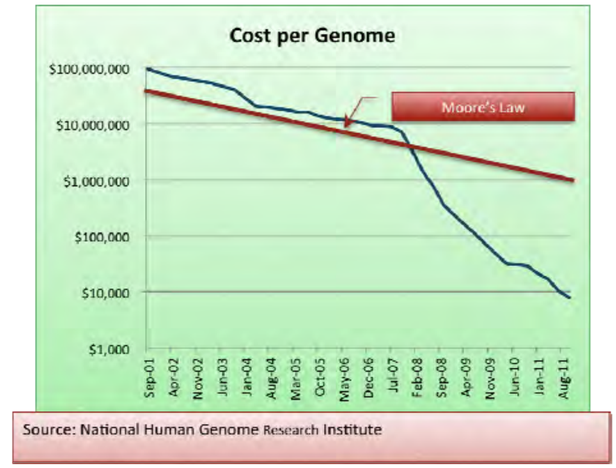



Figure 5. The cost per genome is falling at a much greater pace than Moore's Law.

To manage the high data velocity, NERSC will partner with ESnet to deploy advanced networking technologies. As part of our data strategy, we will enhance NERSC-7 for data in 2014 by adding large memory visualization/analysis nodes, adding a flash-based burst buffer or node local storage, and deploying a throughput partition for fast turnaround of jobs. We will then consolidate our data-centric capability into large systems that are deployed in 2017 and 2020, NERSC Data-1 and NERSC Data-2, respectively. The data systems will provide more memory relative to compute capability, a tighter integration of storage, and high connectivity connections to facilities and compute capabilities. They will also support a wide range of software services and databases.

User services and software are perhaps an even larger challenge than hardware, especially due to the variety and variability of scientific data. Making progress will require close collaborations with the CS research community to develop and deploy new methods for extreme data analysis. Our experience with JGI and the HEP community has been that a deep data engagement with a science community requires eight to ten staff, divided between user services and system management.

## 2.3 Connections with DOE and ASCR Strategic Plans

The first two goals of the DOE 2011 Strategic Plan are: (1) Catalyze the timely, material, and efficient transformation of the nation's energy system and secure U.S. leadership in clean energy technologies, and (2) maintain a vibrant U.S. effort in science and engineering as a cornerstone of our economic prosperity with clear leadership in strategic areas. The first goal encompasses several of NERSC's science drivers (described in the following section), which involve support for Energy Frontier Research Centers, Energy Innovation Hubs, and Energy Systems Simulation, as described in the DOE Strategic Plan. The second goal, providing leadership in science and engineering, is the motivation for NERSC's proposed initiatives in exascale code development, exascale system co-design, and comprehensive support for extreme data (all described in section 6.1).

NERSC's science drivers and initiatives also reflect the goals of the DOE SC Strategic Plan of 2004: (1) Advance the Basic Sciences for Energy Independence, (2) Harness the Power of Our Living World, (3) Bring the Power of the Stars to Earth, (4) Explore the Fundamental Interactions of Energy, Matter, Time, and Space, (5) Explore Nuclear Matter — from Quarks to Stars, (6) Deliver Computing for the Frontiers of Science, and (7) Provide the Resource Foundations that Enable Great Science.

The current plan contributes to many goals from ASCR's ten-year plan, *Simulation and Modeling at the Exascale for Energy, Ecological Sustainability and Global Security,* including (1) developing new science-driven computer architectures and algorithms that will be energy-efficient, extremely scalable, and tied to the needs of scientific computing at all scales; and (2) developing scalable data systems and storage architectures needed to accelerate discovery from large-scale experiments and enable verification and validation of the results of pioneering applications.

## 2.4 Science Drivers

NERSC is the primary provider of computing and data resources, as well as user services, for DOE's Office of Science. Our science drivers span virtually all domains in the physical sciences, bioenergy, bioremediation, and computational science. The defining feature of NERSC's workload is the astonishing breadth of science objectives. Some of the principal drivers and potential outcomes are:

- **Solar energy:** Understand the processes by which biological systems use solar energy, and design materials to efficiently convert solar energy to usable electricity or fuels.
- **Energy-efficient lighting:** Design highly efficient energy conversion processes for enhanced light emission in solid-state lighting devices.
- **Energy storage:** Design efficient and low-cost batteries and hydrogen storage devices (Figure 6).
- **Fusion science:** Understand and control instabilities that limit efficiency in fusion energy reactors.
- **Fundamental particles and interactions:** Supercomputer simulations are crucial to support DOE's "targeted outcomes" involving new facilities and experiments in quantum chromodynamics, the theory of the nuclear force, and the unification of the forces of nature.
- **Accelerator design and development:** Design cost- and energy-efficient particle accelerators, both for exploring the fundamental nature of the universe as well as for biomedical diagnoses and treatment.
- **Astrophysics:** Support DOE efforts in unraveling the origins of the universe, the nature of dark energy and matter, and astrophysical production of exotic nuclei, as outlined in the 2011 DOE Strategic Plan. Another notable outcome will be the design and validation of a new generation of world-class telescopes and satellites (LSST, DES, CMBPol, BigBOSS).
- **Climate science:** Develop methods to predict extreme events. Support policy makers by providing accurate and precise understanding of relationships between climate change and Earth's ecosystems. Incorporate into predictive simulations the micro- and meso-scale physics that drive the Earth's climate, and reconstruct the Earth's climate history from sparse and incomplete data sets.

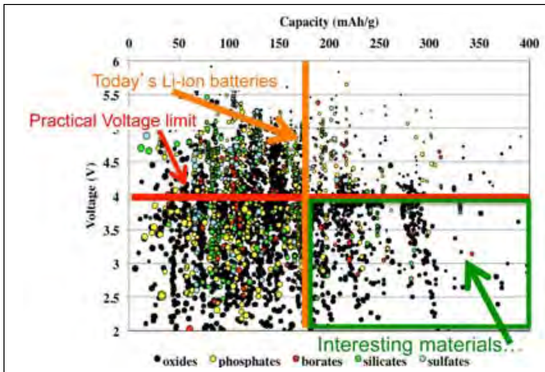



Figure 6. Ab initio simulation of over 25,000 potential lithium-ion battery compounds. The green area shows interesting new materials that are stable and thus worth exploring. The Materials Project replaces trial and error and educated guesses with a systematic approach to design.

- **Biology:** Create predictive, systems-level understanding of complex biological systems in support of DOE missions in energy, the environment, and carbon sequestration. Understand how cells communicate, how proteins are created from DNA sequences, and how enzymes operate.
- **Bioinformatics:** Develop genome-scale microbial and plant system technologies for energy, the environment, and carbon sequestration.
- **Materials Science:** Identify or design novel materials for energy capture, conversion, and storage; energy-efficient devices; carbon sequestration; bioremediation; desalination; radically faster computers; and information storage and communication devices.
- **Computer Science:** Develop new high-efficiency multi-physics simulation methods and codes to support research in combustion, porous media flow, astrophysics, cosmology, and engineering.

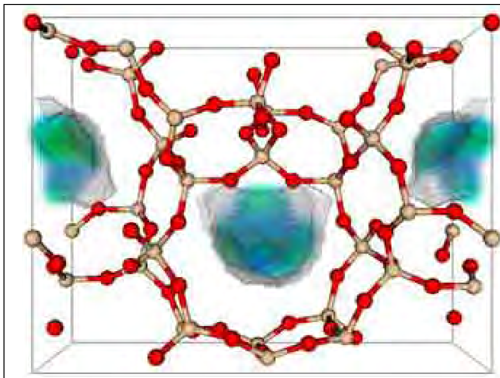


**Figure 7.** One of the 50 best zeolite structures for capturing carbon dioxide. The red balls are oxygen, the tan balls are silicon. The blue-green area is where carbon dioxide prefers to nestle when it adsorbs.

NERSC is deeply engaged with emerging DOE initiatives, providing computing and data capabilities to several DOE Energy Innovation Hubs and Energy Frontier Research Centers, including the Joint Center for Energy Storage Research (JCESR), the Joint Center for Artificial Photosynthesis (JCAP), the Carbon Capture Materials EFRC (Figure 7), and the Center for Nanoscale Control of Geologic $CO_2$.

A rapidly emerging driver is **accelerating scientific discovery at DOE's experimental facilities** such as accelerators, telescopes, light sources, and sequencing facilities. The facility requirements for computing are more data focused and lead to new sets of requirements for systems, services, and expertise. NERSC has a long history in working with these communities through partnerships with the HEP/NP community, the Plank project, and more recently the Joint Genome Institute. NERSC is seeing a rapid increase in the requirements from these communities, often as a direct result of the growth rates in data from instruments. In many cases, these growth rates have out paced Moore's law in recent years and are expected to continue to outpace Moore's law in the coming decade. For example, the cost of sequencing dropped 1000x in the three-year period from 2007 to 2010, compared to approximately a 4x improvement in computing based on Moore's Law. Instruments at DOE light sources are exhibiting similar trends. These rapid improvements strain the ability of facilities to capture and store data. However, the data growth only captures part of the overall challenges. These communities are also looking for new ways to analyze, share, and curate data. For these reasons, the needs of the facilities extend beyond resources alone, but call for an orchestrated solution that brings together the user facilities and their end users, computer science researchers, and ASCR computing facilities.

NERSC's strategy is to broadly advance DOE's science mission through computing and extreme data technologies. U.S. leadership in science increasingly depends on advances in computing and the computing sciences. The breakthroughs and outcomes mentioned above will only be possible through exascale simulations and large volumes of petascale simulations. In order to meet the computational needs of our large community of users and to support a wide range of applications, we will deploy general purpose pre-exascale and exascale supercomputers that are broadly usable. In order to meet the growing data needs of DOE experimental facilities and computing, we will enhance NERSC's data capabilities. Our strategy for achieving these goals is discussed further in Section 6.

# 3. Facility SWOT and Dependencies

## 3.1 and 3.2 Strengths, Weaknesses, Opportunities, and Threats

NERSC's SWOT analysis is presented in Table 1.

**Table 1. NERSC SWOT analysis.**

| | |
|---|---|
| **Strengths**<br>• Unrivaled focus on scientific productivity through computation. NERSC has 40 years of experience providing supercomputer systems and support services in support of DOE's science mission.<br>• Prolific and high impact science output.<br>• Established capability to efficiently support a large, diverse community of scientists.<br>• Open access.<br>• Direct support of all the SC program offices.<br>• Many years of practical experience in data-centric and high-throughput computing.<br>• Highly energy-efficient facility with low-cost, clean power sources. | **Weaknesses**<br>• Power, space, and cooling limits at the Oakland Scientific Facility (OSF) limit our ability to grow. This was the prime driver behind the decision to construct the CRT building.<br>• Remote location. NERSC is located approximately 30 minutes from the main LBNL campus, making it a challenge to collaborate with ESnet and and the Computational Research Division (CRD). Support from the Lab's facilities and other organizations is limited. |
| **Opportunities**<br>• Berkeley Lab has a unique combination of need and capability for data-intensive computing. Among others, the Advanced Light Source (ALS) and the planned Next Generation Light Source (NGLS) are seeking assistance managing the large amounts of data generated by their experiments. NERSC (supplying computational skills and resources), ESnet (providing high-speed network connectivity), and CRD (providing math and algorithm skills) can work together with these user facilities to demonstrate a vertically integrated data program.<br>• NERSC faces the challenge and opportunity to move its large user community and their applications to the coming generation of energy-efficient, manycore systems.<br>• Key roles in DOE's Fast Forward and Design Forward Programs.<br>• Collaboration with LANL/Sandia on future procurements. | **Threats**<br>• Rapid growth of user needs in an era of flat budgets. The growth rate of user computational needs exceeds historical norms, and is higher than the rates of improvements in system performance and power utilization. The consequence is that budgets and power must grow to keep up with demand.<br>• Changing technical landscape for HPC. The basic architectural and programming models for HPC have remained stable for nearly two decades, but are now undergoing a sea change as the industry turns to hybrid and manycore systems for future generation systems.<br>• Reduced vendor and technology options.<br>• Growing disconnect between processor peak performance and scientific productivity. |

## 3.3 Risks, Dependencies, and Critical Gaps

The key challenge for NERSC is to keep up with an accelerating growth of user computational needs that is outpacing not only resources, but also expected improvements in system performance. To do so, we need:

- A robust and timely exascale program that delivers breakthrough improvements in energy efficiency (flops/watt), programmability, and price/performance.

- A comprehensive effort to ensure that applications are ready to efficiently exploit exascale technology systems.
- Adequate funding for systems, staff, and facilities.

Extreme data needs of SC experimental facilities and other programs will place new demands on NERSC. Our experience with data-intensive computing identifies a number of critical needs:

- Up-front and ongoing partnering between science groups and NERSC to plan strategies to effectively manage, store, and process their data.
- Development of technologies (such as burst buffers) to cost-effectively provide increased bandwidth and lowered latency.
- Tools for data management and curation.

NERSC will move from OSF to CRT in 2015. To avoid an extended center outage, we plan a series of incremental moves in which systems and storage will be located at both sites. Transparent access to data from either site will require:

- ESnet support for networks between OSF and CRT: 400 Gbit/sec of bandwidth between the buildings is required.

# 4. Science and Community Needs

## 4.1 User Community Description

In 2012 NERSC supported 4,659 scientific users from universities, national laboratories, and industry, working on 734 projects with allocations of NERSC resources. Our users came from across the United States, with 47 states represented (Figure 8), in addition to a number of international projects. NERSC is truly a broad-reaching national resource. In 2012, 58.3% of our users were from universities, and 32.1% were from DOE labs (14.8% were from Berkeley Lab). The remainder were from industry, nonprofits, and non-DOE laboratories.

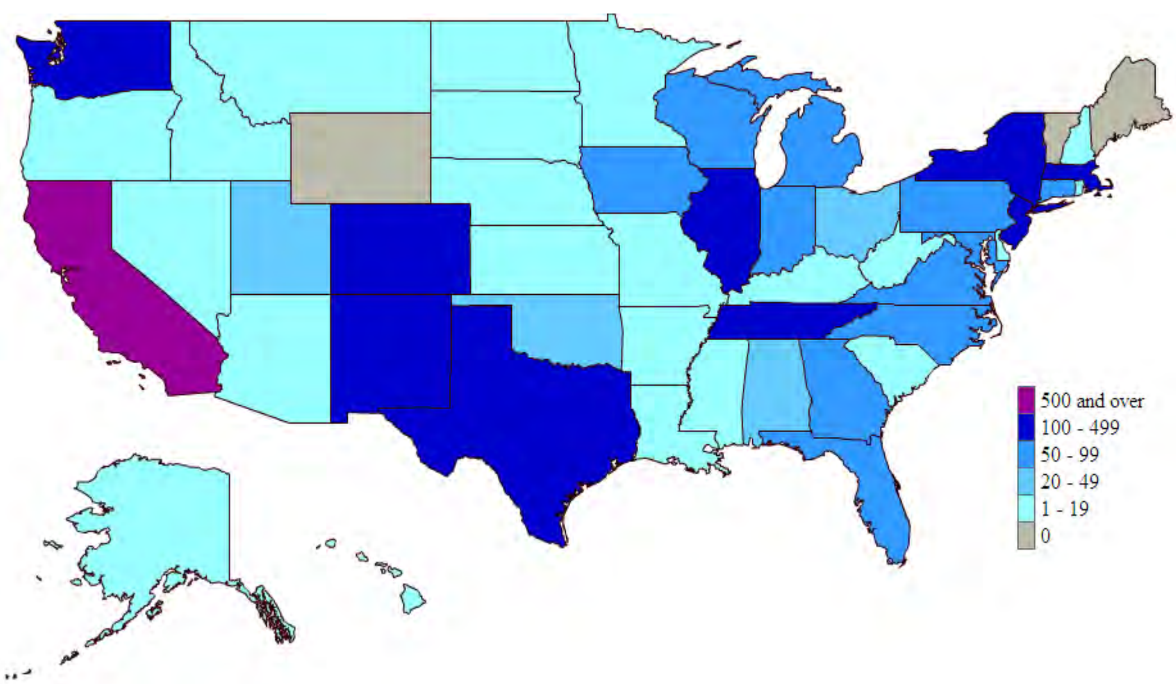



**Figure 8.** Number of current NERSC users per state.

The number of NERSC science users has been growing almost 350 per year since 2008, and we expect this growth rate to continue. We have observed a usage trend in the past ten years towards sciences with more potential for applied research applications (e.g., materials science, biosciences). We expect this trend to continue as well (see Figure 1).

NERSC's users span the entire Office of Science portfolio. Usage by office is shown in Figure 9. Figure 10 shows application usage at NERSC from largest to smallest. Although there are over 600 application codes run at NERSC, the workload is unevenly distributed and highly concentrated. Ten key applications make up approximately 50% of the workload, while hundreds of applications make up 10% of the workload.

In addition to application diversity, DOE user applications are also diverse in terms of algorithms. Figure 11 groups the applications from Figure 10 by algorithm area.

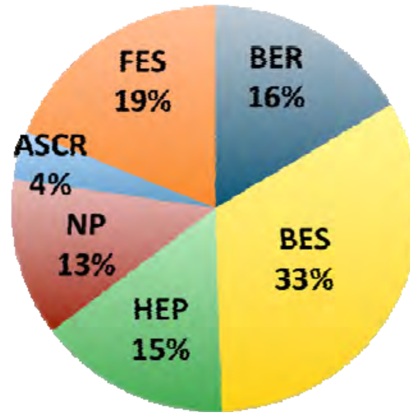



**Figure 9.** NERSC usage in 2012 by DOE SC office.

NERSC's users need to run at many scales (Figure 12), including capability simulations that span all or most of the machines, and high volumes of smaller simulations for uncertainty quantification, scientific exploration, optimization, and statistical aggregation.

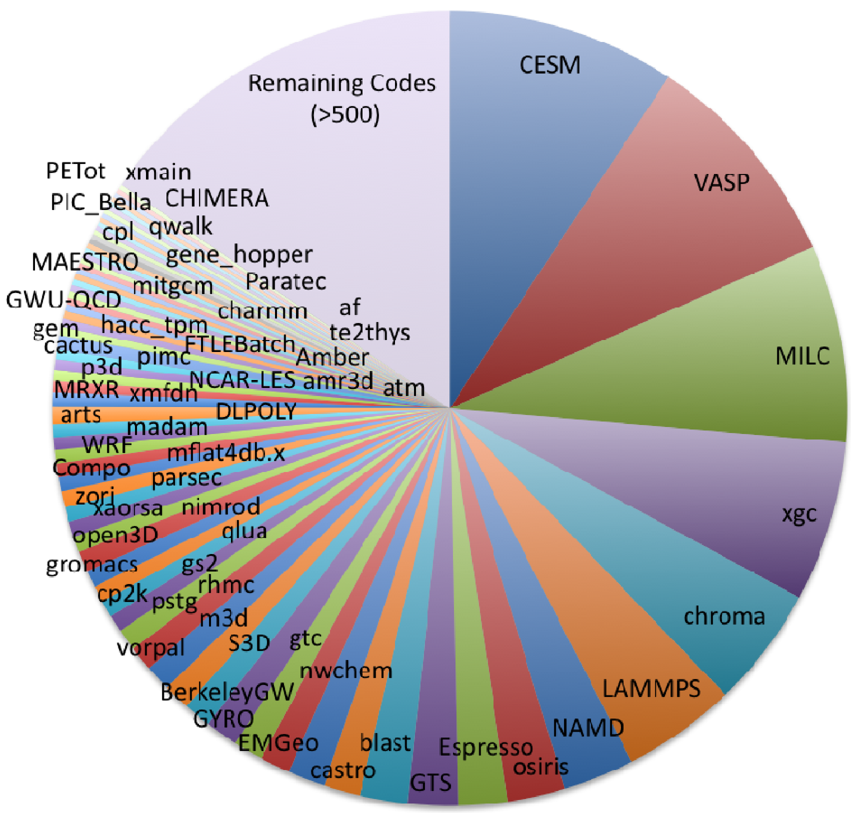



**Figure 10.** Application usage at NERSC in 2012 from largest to smallest.

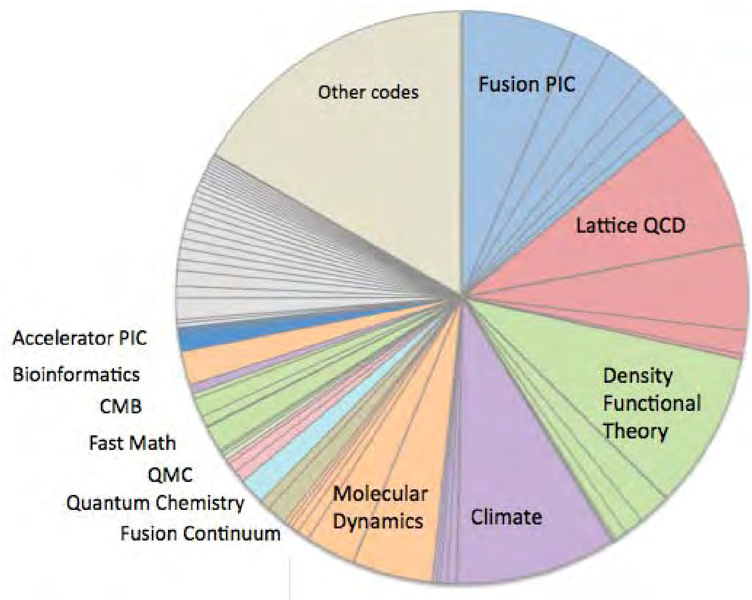



**Figure 11.** Codes grouped by algorithm type.

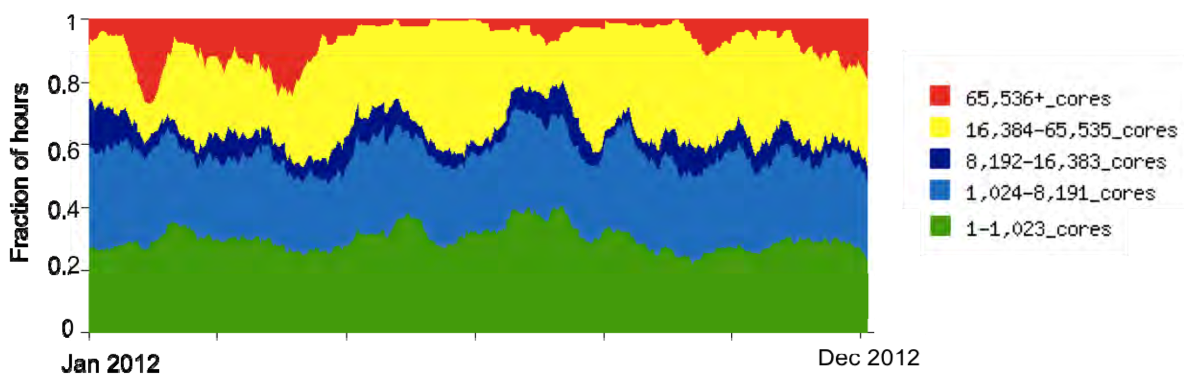



**Figure 12.** Job size breakdown on Hopper.

## 4.2 User Community Engagement

NERSC offers a wide range of user support services. The User Services Group provides the direct interface between NERSC and its user community. The services and engagements through User Services are many-fold: from resetting passwords, to one-on-one consulting, to running Office of Science-wide requirements reviews, and to coordinating its active user group (NUG). NERSC's website features a prominent "For Users" top-level section that makes it easy for users to find information; and our training activities provide additional opportunities for communicating with users.

### 4.2.1 Consulting and Account Support

NERSC's consultants and account support staff are available to users via email, an online web interface, and on the phone during business hours (8:00–5:00 Pacific Time). Basic account support (password resets, resetting login failures) is available via the NERSC operations staff 24 x 7, 365 days per year.

NERSC's 12 consultants include 9 experts in high performance computing, 6 with Ph.D. degrees. NERSC's two account support personnel each have 10+ years of experience. When users contact

NERSC, they immediately are in contact with highly trained HPC specialists, who can usually solve issues directly, or immediately route the request to the appropriate systems engineers. NERSC's policy is to respond to all inquiries within four business hours, and either solve the problem or communicate a work plan to the user within three business days for all reported incidents. The latter is a metric we report to DOE.

When users require more extensive assistance — such as programming or optimization help — they are routed to a consultant with the most appropriate skill set, who will engage in a longer-term collaboration with the user. Of course, with thousands of users and only a handful of consultants, extensive one-on-one assistance is often not possible. NERSC's strategy is to deliver assistance at scale by heavily relying on extensive web documentation and an active training program. Likewise, many account support services that were once labor intensive (e.g., password resets, updating contact information) have been converted to self-service operations via the NERSC Information Management (NIM) web interface (nim.nersc.gov).

NERSC's trouble ticket interface (help.nersc.gov) allows users to submit and track their tickets at any time via the web. This interface is also used to track incidents internally and to interface with certain vendors. Popularity of this web interface has grown to the point that 30% of reported incidents in the second half of 2012 were initiated through the web interface, and only 2% started as a phone contact (58% were email, and 11% were created by NERSC staff proactively for users). NERSC innovations in user support have allowed us to handle an ever growing number of users and user tickets with extremely high user satisfaction on the annual user survey.

When resolution of a user issue requires interaction with NERSC system engineers (e.g., tuning network settings between NERSC and a remote site, or optimizing file transfers) or other HPC experts (e.g., data analytics or science gateway staff), the ticketing system is used to directly connect the user and the expert, who work together to implement a solution.

### 4.2.2 Deep Engagements

NERSC provides special assistance to strategic projects by making special staff assignments for limited periods. These deep engagements often have high payoffs. For example, NERSC was awarded a 2011 Scientific Computing Award for HPC Innovations Excellence for enabling the 20th Century Reanalysis Project. NERSC staff provided significant support that was needed to successfully transfer all of the dataset to NERSC, collaborated with the science team on science gateway software to construct a portal that provides access to the data, which currently hosts the master copy of the entire dataset.

Other deep engagements leverage NERSC's extensive HPC technology expertise. For example, we have developed tools to help users create science gateways (web portals). These tools have contributed to NERSC hosting a number of prominent gateways for science teams, including those for the Materials Project, Deep Sky, the Coherent X-Ray Imaging Data Bank, and the Daya Bay Neutrino Detector. Work is under way to set up workflows to connect NERSC with the Advanced Light Source at Berkeley Lab and with users of the LINAC Coherent Light Source at SLAC.

The Data Analytics Team often spearhead some of our deep engagements as we endeavor to help our users get the most science out of their simulations or data analyses. A recent example involved the analysis of a leading plasma physics simulation with over a trillion particles on 120,000 Hopper cores. This unprecedented simulation produced 30 TB of data per time step. Here the team worked with both NERSC and Cray staff to optimize the performance of HDF5 and MPI-IO to obtain peak I/O rates on 144 Lustre object storage targets (OSTs). In addition, leveraging our ties to LBNL's Computational Research Division, we were able to apply FastBit-based indexing/querying to the dataset. This allowed the scientists to query the indexed one trillion particle dataset in approximately three seconds.

In 2012 NERSC initiated a two-year **Data Pilot Program** which focuses on data-centric projects with the aim of advancing their science by providing access to significant computational and storage resources. NERSC allocated over a petabyte of storage, 20 million core-hours for analysis, special resources such as large memory systems, and access to high-performance flash storage. Eight projects were selected to participate in the pilot program. These pilot projects applied to NERSC not for computer time but for a range of data science capabilities. The projects also are assisting NERSC in investigating new data methods and understanding their usefulness to NERSC science stakeholders. The selected projects and related big data technology areas are shown in Table 2.

**Table 2. Projects Selected for NERSC's Data Pilot Program.**

| Project | PI | Data Technologies |
|---|---|---|
| High Throughput Computational Screening of Energy Materials | Gerbrand Ceder (MIT) | FLASH, high-throughput queueing, gateways |
| Analysis and Serving of Data from Large-Scale Cosmological Simulations | Salman Habib (ANL) | Databases, gateways |
| Interactive Real-Time Analysis of Hybrid Kinetic-MHD Simulations with NIMROD | Charlson Kim (U Wash) | H5Part, FLASH, FastBit |
| Next-Generation Genome-Scale In Silico Modeling: The Unification of Metabolism | Josh Lerman (UCSD) | Databases, gateways |
| Data Processing for the Daya Bay Reactor Neutrino Experiment's Search for $\Theta_{13}$ | Craig Tull (LBNL) | WAN data pipes, gateways |
| Transforming X-Ray Science Toward Data-Centrism | Amedeo Perrazoo (SLAC) | WAN data pipes |
| Data Globe | John Wu (LBNL) | FastBit, gateways |
| Integrating Compression with Parallel I/O for Ultra-Large Climate Data Sets | Jian Yin (PNNL) | MPI-IO and FLASH |

During 2012 these projects have accomplished a great deal of data science discovery. Among those are the achievements of 700 MB/sec WAN data transfers, new all-to-all data set comparisons using flash storage, new data sets made available to the science community, and one pilot's work was awarded Runner-Up AAAS Breakthrough of the Year 2012 (see http://www.nersc.gov/news-publications/news/science-news/2013/nersc-contributes-to-science-magazine-s-breakthroughs-of-the-year/).

### 4.2.3 User Survey

NERSC conducts an annual User Survey, in which users are asked roughly 90 questions to evaluate the NERSC systems and services. This is an important undertaking as it allows a cross check on input gathered from science-area-focused requirements gathering processes and provides individual as opposed to community consensus feedback. The User Survey also provides specific feedback on individual systems and services so that NERSC can identify and correct problem areas that may be impacting user productivity. Most importantly, an overall satisfaction metric gauges how well NERSC is serving its users. NERSC prides itself on maintaining a rating of above 6.0 out of 7.0 for the past thirteen years (Figure 13), with 6.5/7.0 in 2011 (the last reporting period).

### 4.2.4 NERSC Users Group (NUG)

The NERSC Users Group (NUG) has active participation at monthly calls, which use videoconferencing to enhance the user experience. NUG also has an annual face-to-face meeting with NERSC staff to discuss NERSC's future plans and to review the previous year. NERSC holds a training session at this meeting, typically on new systems, new programming methods, or new software.

### 4.2.5 Training

NERSC has an active training program, with classes available to its users in person and via the web. NERSC conducted eight training events in 2012, on topics ranging from "Getting Started at NERSC" to three-day advanced courses.

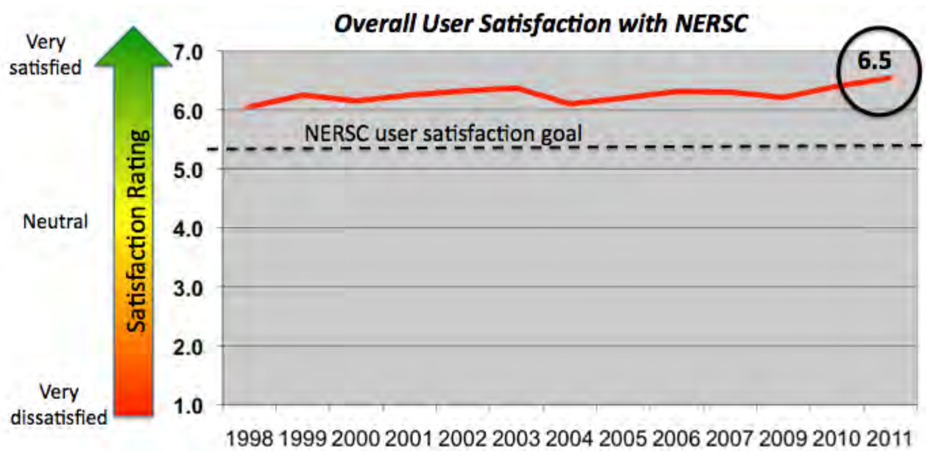



**Figure 13.** Overall user satisfaction with NERSC has remained high over the years.

## 4.3 Requirements Gathering

In support of its mission and to maintain its reputation as a center keenly focused on satisfying the needs of its scientific community, NERSC continually gathers requirements from its user community, from the DOE Office of Science program offices, and from other key HPC or domain experts.

### 4.3.1 Data Gathering

As the primary provider of high performance scientific computing, storage, and services for the entire DOE Office of Science, NERSC must support the mission of all six Office of Science program offices as well as satisfy the operational needs of the scientists who use NERSC's resources. To insure that it can meet these needs, NERSC regularly gathers requirements from leading scientists in each research area as well as from its 4,500 users. Requirements from the user community are obtained through regular meetings with the NERSC Users Group (NUG), an annual User Survey, allocations requests (ERCAP), one-on-one conversations with users, participation at community conferences and meetings, and a series of Computing and Storage Requirements Reviews with each program office within the Office of Science.

The requirements reviews are key to NERSC's strategy for gathering community requirements. This ongoing series of reviews, one for each program office, brings together DOE program managers, science leaders, and NERSC staff to derive future HPC needs for scientific communities. The results of the reviews include community requirements for computing, storage, and services five years out. Each review report also contains a number of significant observations, topics of keen interest to the review participants. These results help DOE and NERSC plan for future systems and HPC service offerings.

One round of six reviews has been completed, and the findings have been published (available at http://www.nersc.gov/science/hpc-requirements-reviews/). The first round of reviews engaged 12 DOE managers, 85 leading scientists, and 68 NERSC projects, representing 62 percent of all NERSC hours used. The second round of reviews, targeting needs for 2017, is under way and has generated even more participation from within DOE. The scope is expanding beyond the traditional NERSC user bases to include new areas of science that are for the first time experiencing extreme computing and data needs.

NERSC holds regular monthly teleconferences with the NERSC Users Group (NUG) and an annual face-to-face meeting. In both the monthly and annual meetings, NERSC users are invited to discuss issues of interest or concern and make their needs known. NERSC users elect a 21-member Executive Committee (NUGEX) that plays an active role in making suggestions for improving NERSC systems, services, policies, and procedures (e.g., queue configurations). NUGEX members are given invitations to attend the requirements reviews in their area of research.

NERSC collects satisfaction, importance, and usefulness scores on various systems and services via its annual User Survey. Users can also leave comments on the survey. The survey responses help guide NERSC's choices for future systems and services offerings. In addition, the survey usually contains a special section on a relevant topic to help NERSC plan for the future. The 2012 survey's special section is on readiness for using future architectures and data requirements.

Another important vehicle for collecting requirements is the ERCAP allocations application process. When applying to use NERSC resources, projects are required to give detailed descriptions of their codes, data requirements, and software needs for the coming year.

### 4.3.2 Data Forecasting

Projected needs from the reviews form a basis for users' future needs at NERSC. The requirements stated in the review reports are derived from DOE managers' and scientists' research goals, guided by expectations for available resources (funding) and technology advances. NERSC plays an important role in arriving at these quantitative needs by providing historical perspectives (e.g., Moore's Law) and technology projections.

The specific forecasting method is as follows: projects represented at the requirements reviews will have used a certain fraction of all NERSC hours or bytes stored within that office for the current (or just completed) allocation year. Their aggregate projected needs are then scaled up by that fraction to arrive at a total need for the office under consideration. For example, at the first HEP review, 85% of the 2009 HEP usage at NERSC was represented. Their total requirements were then scaled up by a factor of $1/0.85 = 1.18$ to arrive at a need for all of HEP.

These quoted numerical requirements largely represent the needs of the traditional NERSC simulation community for computing and storage at NERSC. They do not reflect the needs in areas of science that have emerging HPC needs, nor do they reflect the requirements from DOE experimental facilities, which may be considerable, especially for data.

Requirements for software and services are also captured in the review reports. Responses to the annual user survey provide a cross check on the review reports as well as individual user perspectives.

# 5. Center Operational Priorities

## 5.1 Center Priorities

As the primary computing center for the Office of Science, NERSC supports a diverse and growing scientific workload. **Our fundamental center priority is to optimize the computational productivity of our users.** To do so, we provide scalable and reliable systems, backed by exceptional user support. Because DOE selects most of the science areas, problems, and users that run at NERSC, our systems and services must remain flexible so that we can quickly adapt to our users' evolving needs.

In order to maximize the available computational resources available to users, we procure systems on the basis of price/performance — sustained performance over life cycle cost — instead of peak performance or performance on narrow micro-benchmarks such as HPL. Sustained performance is estimated by a set of benchmarks (Sustained System Performance or SSP) that provide representative coverage of applications in various science domains and algorithmic methods — a proven method for determining how efficiently a system will execute real scientific applications. These benchmarks have evolved over the years as user needs and programming models have changed. We intend to develop new, sustained performance measures tailored for extreme data systems.

NERSC works to reduce ongoing operational costs so that a higher fraction of our budget is available to meet practically unlimited computational and storage needs. We have continuously improved energy efficiency at OSF, and CRT is a model for energy-efficient design.

We develop internal tools and processes to allow us to manage larger, more complex, and more differentiated systems with a fixed staff. To deliver excellent user services to a continually increasing set of users, we focus on efforts that are scalable (e.g., training).

User needs scale from large numbers of low-concurrency jobs to individual high-core-count jobs that occupy most of a machine. Even within single projects, computational needs may vary over the life of the project (from code development to production simulation) or involve different workflow steps with varying computational profiles. NERSC could provide separate systems for different classes of applications; however, by allocating most of our budget to large, flexible systems, we can configure the system dynamically to support throughput workloads or simulations at much larger scales than we could if we divided our resources. We are able to keep utilization well above 90% by backfilling nodes not being used by larger jobs.

Where these large systems do not meet user requirements, we provide midrange systems and other specialized systems. Much of our current capability to support data-intensive workloads comes from this fraction. We also invest in small testbed systems to evaluate promising technologies such as hybrid and manycore architectures, and new computational models such as Hadoop.

In order to meet the data needs of users, NERSC allocates a growing proportion of its budget (now ~10%) to provide storage at ever-increasing capacities and capabilities. A multi-tier strategy is used to optimize costs and performance. System-specific scratch or node-local file systems provide running jobs with temporary storage at the highest bandwidths. A center-wide file system (NERSC Global Filesystem, or NGF) provides medium-term storage for active projects, and enables high-bandwidth data sharing among our various systems. Long-term archival storage is provided with tape libraries. Tapes are now and will continue to be the most cost- and energy-efficient method for long-term storage. NERSC currently holds 40 years of scientific data in its archive. The center-wide file system and tape archives are procured using an "evergreen" strategy in which annual procurements are used to expand capacity and refresh technology.

Ensuring that users can make effective use of the provided computational assets is critical for scientific productivity. NERSC is a recognized leader in user support, as shown in the results of our annual user survey. Our approach focuses on directly helping the DOE scientific community become more productive in their end-to-end workflow, including computational and data management. We provide expert assistance and advice on subjects such as parallel computing, compilers, libraries, programming models, MPI, OpenMP, I/O, performance analysis, debugging, use of HPC software, data analytics and visualization, and workflows.

NERSC uses various control metrics for assessing and improving its operations. Key metrics include: (1) response time for user problem reports, (2) number of computational hours and amount of storage delivered, (3) system scheduled and overall availability, (4) system utilization, (5) queue wait times and backlogs, (6) percent of jobs running at or above a target concurrency, and (7) the overall user satisfaction rating from the User Survey. Many of these metrics are reported to ASCR during the annual Operational Assessment, and NERSC has consistently exceeded target goals.

## 5.2 Inter-Center and Inter-Facility Collaboration

NERSC recognizes its role in the broader family of DOE centers and facilities. Seamless interoperability between DOE resources is a guiding goal, as is the exploration of community best practices in HPC. NERSC's strategy has collaboration to those ends built into it. NERSC works closely with other facilities from the DOE Office of Science arena and beyond to build productive partnerships. Examples of such collaborations include the NERSC-8/Trinity procurement, the data transfer working group (which works to optimize data transfers primarily between ANL/ORNL/LBNL), hosting a DOE-wide and international workshop on best practices (power management, file and archive system), membership on the National User Facility Organization (NUFO) Steering Committee, and our deep bonds with ESnet.

DOE user facilities are not static entities. As needs shift over time and technology, DOE facilities must remap their capabilities and methods to best meet mission goals. The year 2011 saw a major shift in genomic science from wet chemistry to data analysis. In 2013 NERSC now provides the core computing and data capabilities for the Joint Genome Institute and its users. We expect other such rebalancing toward computing and data to be part of the future of DOE facilities.

## 5.3 Connections with National Laboratory Plans

NERSC's strategic plan aligns closely with LBNL's lab-wide strategy. The focal point of LBNL's strategy is "Carbon Cycle 2.0," which aims to create a sustainable carbon cycle by developing new energy sources, reducing energy consumption, and improving carbon capture. This strategy is supported by three pillars: nanoscience, biology, and computing. NERSC's strategy is connected with all three pillars, since nanoscience and biology are heavily driven by experimental facilities like the JGI, ALS, and in the future, NGLS. The data initiative described in Section 6.1.3 is directly targeted at addressing the data challenges faced by those facilities. The computing pillar of the lab strategy highlights the role for simulation and modeling, the increasing importance of data, and ESnet's central role in connecting facilities with scientists.

# 6. Center Strategy for the Future

A critical strategic objective is meeting 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, and influencing the computer industry to ensure that future systems meet the mission needs of SC.

**The aggregate computing needs of SC science teams at NERSC are well into the exascale regime by the end of the decade.** Users need to run simulations at hundreds of petaflops, and they need to run thousands to millions of petascale simulations to achieve their goals. We will deploy pre-exascale systems in 2016 (NERSC-8) and 2019 (NERSC-9). We anticipate deploying our first exascale system, NERSC-10, in 2022. Our goal is to provide our users a consistent path to exascale, particularly in the evolution of the programming model. We are asking bidders on NERSC-8 to describe how their proposed architecture aligns with their plans for exascale. We will continue to support MPI into the foreseeable future so that our codes can execute, albeit at less than optimal performance, on future systems with little or no modifications. We anticipate that many of our codes will transition to MPI for interprocessor communication plus OpenMP for on-node concurrency in the NERSC-8 time frame. NERSC-8 will also support another programming model (MPI+X, where X is yet to be determined), which could provide even greater performance. Even though only a few users will directly take advantage of this programming model, many will benefit from the availability of optimized libraries.

**Longer term, Berkeley Lab is willing to lead a multinational effort to converge on the next programming model.** The requirements for broad adoption of a new programming model are performance (at least 10-50X improvement), portability, durability (users are only willing to rewrite their codes once) and ubiquity/availability/openness**.** We will

- leverage research efforts (XStack, XTune, DEGAS) for advanced programming constructs
- assess existing models through participation in standards bodies (OMP, MPI, Fortran, C/C++) and assess emerging languages
- engage co-design community of vendors & HPC for cross-cutting solutions
- share results and build consensus

In order to ensure that users are ready to take advantage of NERSC-8 as soon as it arrives, we are launching an initiative (Section 6.1.1) to begin transitioning the SC workload to energy-efficient architectures. We are also launching a longer-term initiative to ensure that future systems meet SC mission needs (Section 6.1.2).

Another critical strategic objective is increasing the productivity, usability, and impact of DOE's user facilities by providing comprehensive data systems and services to store, analyze, manage, and share data. We will begin enhancing our data capabilities in 2014, and **we will deploy data systems, storage, advanced networking, and enhanced user services so that scientists can store, move, share, and analyze exabytes of data early next decade.** We have launched an initiative focused on increasing our engagement with DOE experimental facilities (section 6.1.3). We plan to enhance the data capabilities of NERSC-8 by increasing the size of the disk system, adding visualization/analysis nodes, and adding a burst buffer. We will deploy new data-centric systems in 2017 (NERSC Data-1) and 2020 (NERSC Data-2).

# 6.1 Initiatives

## 6.1.1 Transition the SC Workload to Energy-Efficient Architectures

NERSC is the primary computing facility for SC, and over 600 applications run at NERSC. A large fraction of these codes must transition to new architectures. Clock speeds are expected to stall near 1 GHz, and our code teams will realize no performance growth unless they take much greater advantage of on-chip concurrency. Making this transition as smoothly and effectively as possible will allow our science teams to meet their goals in the 2015–2023 timeframe. NERSC plans to collaborate proactively with our users to make sure they are prepared to effectively use our first pre-exascale system, NERSC-8, when it arrives in early 2016.

Our strategy for enabling this code transition is as follows:

- We will deploy testbeds to gain experience with new technologies and to better understand emerging programming models and potential tradeoffs. In particular, we will deploy a testbed representative of the NERSC-8 architecture as soon as it is determined.
- We will have in-depth collaborations with selected users and application teams to begin transitioning their codes to our testbeds and to NERSC-8. We will choose partnerships based on level of interest, expected application usage, and algorithmic diversity.
- We will develop training and online resources to help the rest of our users based on our in-depth collaborations and on collaborations with co-design centers and ASCR research projects. We will leverage our existing training vehicles, such as online webinars, as much as possible.
- We will add consultants with an algorithms background who can help users when they have questions about improving the performance of key code kernels.

NERSC has assembled an Architecture Evaluation and Application Readiness team with approximately 12 people to help transition applications. Unfortunately, because of other operational duties, each team member can spend only 10–20% of their time devoted to this effort. NERSC needs to increase the number of staff devoted to the Application Readiness effort and proposes adding an Exascale Postdoc Program of six, in addition to an increase of three consultants and one systems engineer (Table 3). These new staff will work directly with application teams to improve application performance, and will be responsible for developing training modules and online resources as well.

Table 3. Additional Staffing for Transitioning the SC Workload Initiative

| Roles | Effort (FTEs) | Added capability |
|---|---|---|
| Application consultants | 3 | Assistance with transitioning users to new architectures; exploration of performance issues with new architectures. |
| Postdoctoral associates | 6 | New Exascale Postdoc Program. Application-specific expertise with areas to be transitioned to using new architectures. |
| Systems engineer | 1 | Deployment of testbed systems; exploration of issues with putting new architectures into production in a reliable and stable fashion. |

### Risks

The risks associated with this initiative fall into two categories: risks associated with undertaking this initiative and those associated with NOT undertaking this initiative. The risks in undertaking the initiative arise from problems and delays with transitioning to the new architectures, lack of a clear consensus on how to use the new architectures effectively, and staffing concerns. Risks associated with not undertaking the initiative include lowered productivity for science efforts, not achieving SC science goals, fragmented computing methodologies, and large swaths of scientists being unable to take advantage of technological advances in computing. We believe that proceeding is critical for advancing DOE's science mission, and we will mitigate the risks to the extent possible. Supporting the initiatives will be an important priority in future hiring decisions at NERSC.

## 6.1.2 Ensure That Future Systems Meet SC Mission Needs

The previous initiative was focused on reacting to what the computer industry gives us in the next ten years. However, we also want to influence the computer industry as much as possible, because the broader market might drive companies in a direction that does not fully meet the mission needs of SC. To achieve greater influence, we plan to:

- **Partner with Los Alamos and Sandia on procurements in 2016 and 2019.** The larger size of these procurements will give us greater leverage with industry. In addition, these procurements are expected to fund Non-Recurring Engineering which will provide enhancements that benefit DOE applications. SC science applications and their requirements have been included in the Request for Proposals through the involvement of NERSC in the partnership.
- **Provide industry with greater information on NERSC's workload through new and innovative instrumentation, measurement, and analysis.** To complement DOE co-design centers, NERSC will focus on the aggregate needs of SC. This is especially appropriate since we are trying to meet the aggregate computing and data needs of SC's program offices during the next 10 years. NERSC has historically procured systems based on the sustained performance of key applications. It will be difficult to sustain this approach because architectures are changing much more dramatically than in the past. In fact, leading microprocessor companies do not have tools to predict the performance of SC applications on their future chips. However, they are very interested in developing a better understanding of our workload, and we have an opportunity to influence their designs by providing them more detailed information on our aggregate requirements for memory capacity, memory bandwidth, interconnect bandwidth, etc. It is very unlikely that a microprocessor or memory company will make an architectural change for one application; however, they might for a representative HPC workload.
- **Actively engage with industry through DOE's Fast Forward and Design Forward programs.** NERSC has partnered with the other DOE laboratories to launch and manage DOE's Fast Forward program, which is funding processor, memory, and storage efforts; and we are currently filling several leadership positions. In addition, NERSC is responsible for the Design Forward procurement on behalf of DOE and the other laboratories.
- **Leverage the Berkeley/Sandia Computer Architecture Laboratory (CAL) that has been established by ASCR.** A key goal of CAL is partnering with computer companies to develop proxy architectures that are representative of future systems. The intent is to make these proxy architectures open and non-proprietary, so that we can engage a broad user community like NERSC's in co-design discussions.
- **Serve as a conduit for information flow between computer companies and our user community.** We will use existing venues like the requirements reviews and the NERSC Users Group meetings to present future proxy architectures and get feedback. We will also share results on testbeds at these venues and then provide consolidated feedback to the companies.

To accomplish these goals, we need to add three staff members to our existing efforts. Two will focus on collaborations with computer companies, and one will be responsible for workload aggregation (Table 4).

Table 4. Additional Staffing for Future Systems Initiative

| Roles | Effort (FTEs) | Added capability |
|---|---|---|
| Systems analysts | 2 | Co-design efforts; application porting. |
| Workload analysis specialist | 1 | Deeper understanding of workload in aggregate. |

### 6.1.3 Address the Most Demanding Data Challenges from DOE Facilities

DOE user facilities are facing extreme data challenges. A majority of Advanced Light Source (ALS) users analyze only part of the data they collect and rarely search for patterns across multiple data sets. In the future, data volumes and velocities at these facilities will be even higher, because many scientific instruments are improving at rates much faster than Moore's Law. Within a decade, data sets will reach an exabyte in size, and data velocities will be measured in terabytes per second.

Meeting the challenges of extreme data is not simply a question of technology. Our experience with data-intensive projects shows that early, deep, and continuing engagement with the science teams is critical to ensure a workable and efficient solution. Domain scientists must work with software engineers, mathematicians, and computer/network architects to craft complex workflows. These projects are typified by the need to collect, analyze, organize, and publish data. These steps are custom to the project and often involve deploying systems that have more in common with real-time or e-commerce than traditional HPC. New skills, training, and services will be required. Success will be people intensive.

Key elements of NERSC's data strategy are:

- **Partner with DOE experimental facilities to identify requirements and create early success.** Our early work with ALS has identified requirements and goals to transform synchrotron data analysis toward scalable solutions required by modern beamlines. NERSC pilot projects have shown automated data pipelining, indexing, search, and tape archive to be a good fit with current and future ALS needs. Our initiative seeks to implement scalable solutions to such requirements along the entire data analysis pipeline using NERSC computing and data resources. Extending this work to an inter-facility national framework for light sources serves the interests of all DOE facilities and their users by increasing the efficiency and impact of these instruments.
- **Develop and deploy new data resources and capabilities.** We will accelerate NERSC's traditional storage growth rate to meet rapidly increasing user requirements for capacity and bandwidth. We will expand our existing science gateways and data transfer systems to facilitate high-volume data intake and publishing; implement remote, redundant tape copies to provide increased protection for valuable data; and deploy automated data management techniques to efficiently utilize storage. We plan to enhance the data processing capabilities of NERSC-7 in 2014 by adding large memory visualization/analysis nodes, adding a flash-based burst buffer or node local storage, and deploying a throughput partition for fast turnaround of jobs. We will deploy new follow-on data-centric systems in 2017 (NERSC Data-1) and 2020 (NERSC Data-2).
- **Provide expertise and services for extreme data.** Data-intensive workloads require additional layers of services beyond the traditional HPC services. Storage needs and workflow logic which are database driven, the use of structured and unstructured scalable object stores, software solutions for codes which traverse massive data sets for search or analysis, sophisticated web-based gateways to interact with data, and the growing needs to move beyond archive toward accessible scientific data curation all point toward a new class of HPC services that are needed. This model contrasts and complements the traditional approach to providing high performance

computing cycles, where users log into a shell, compile monolithic applications, and submit batch jobs.

- **Leverage ESnet and ASCR research to create end-to-end solutions.** We will deploy high speed networking between our data resources and nearby DOE Facilities in collaboration with ESnet. We will use these networks in new software-adaptable ways to raise the bar in what scientists imagine possible in terms of data access, making data quickly available where it will have the highest impact. Our plan emphasizes developing scalable software and algorithms for dealing with the high volume, the high velocity, and the high variability of data. We plan to leverage ASCR research and bring useful software packages from research to a production setting.

In order to support our data-centric efforts, we plan to increase our staff in key areas and launch a Data Postdocs Program. Table 5 shows a summary of staffing needs.

Table 5. Staffing for NERSC's Data Initiative

| Roles | FTEs | Added NERSC Capabilities |
|---|---|---|
| Systems administration | 4 | • Reliable, high performance, world-class data systems.<br>• Tuning file systems, developing HPSS, deploying data migration facility. |
| User services consultants, programmers, and software engineers | 8 | • Consultants for data-centric application engagement.<br>• Topical expertise in X-ray, image analysis, and astro.<br>• Workflow, high-throughput computing, and data format specialists.<br>• Data pipeline and data API developers.<br>• Web gateway and application programmers.<br>• Bring deep statistics expertise to consulting rotation.<br>• Software engineer dedicated to hardening research software into well packaged apps & libraries. |
| Endstation computationalists | 3 | • Tactical data planning in situ with small teams<br>• Boots on the ground at DOE facilities (BES beamlines, benchtops, and detector labs). |
| NERSC Data Postdocs | 10 | • Repeat Petascale Postdoc Program success.<br>• Bring new applications to HPC; staff development. |

## Risks

We summarize some of the high-level risks, impact, and mitigation in Table 6. However, NERSC believes the greatest risk comes from ASCR failing to provide DOE SC with computing science research, services, and resources required to meet its growing data needs. Given the magnitude of DOE SC investments in experimental facilities, even modest improvements in efficiency, turnaround, and productivity can have large benefits.

Table 6. Risks Associated with NERSC's Data Initiative

| Risk | Impact | Mitigation |
|---|---|---|
| Scientists aren't ready to leverage advanced capabilities. | Resources go unused or are heavily used by only a small community, limiting the full scientific impact. | User Services staff engage and collaborate with scientists to assist them in using the new services and scaling applications. |
| Resources and services are misaligned with needs. | Scientists aren't able to port applications or workloads to the provided resources. | Leverage the requirements review process to periodically gather requirements and guide changes in services and systems. |
| Consolidation of resources leads to solutions and policies that inhibit individual groups. | Users aren't able to work effectively due to competing requirements. | Establish a committee composed of representatives from the user community to steer policy decisions. |

## 6.2 System Software and Libraries

The computational environment to support the systems at NERSC includes vendor-supplied operating systems, batch and programming environments, self-supported operating systems and programming environments, as well as third-party products installed by NERSC. Table 7 summarizes the software stack currently relevant to users at NERSC.

Table 7. Major System Software, Applications, and Libraries at NERSC

| Software Area | Application Area | Representative Software | Source |
|---|---|---|---|
| **Operating Systems** | | Linux, Cray Linux | Open source, vendor |
| **Batch Systems** | | Torque, Moab, Grid Engine | Vendor |
| **Storage Systems** | | Lustre, GPFS, HPSS | Open source, vendor, ASCR/NNSA/vendor collaboration |
| **Communication Stack** | | OFED, uGNI, Aries | Open source, vendor |
| **Applications** | Math | Mathematica, Matlab | Vendor |
| | Chemistry | NWChem, Gaussian | Open source, vendor |
| | Material science | VASP, PARATEC | Vendor, open source |
| **Compilers/Languages** | | GCC, Intel, Cray Compiler/Fortran, C, C++, PGAS, Python | Open source, vendor |
| **Development Tools** | Version control | CVS, SVN | Open source |
| | Debugging | DDT, GDB | Vendor, open source |
| | Profiling | IPM, Darshan, CrayPat | ASCR, vendor |
| **Programming Libraries** | Math | ACML, LAPACK, LibSci | ASCR, vendor |
| | I/O | HDF, NetCDF | Open source, NSF |
| | Graphics | NCAR | NSF |
| | Communication | MPI | Open source, vendor |
| **Analytics/Visualization** | | AVS/Express, R, Visit | Vendor, open source, ASCR |

There are several well-known problems that can be anticipated with using this software stack to support using newer architectures. Primary among these will be issues arising from increased concurrency, larger scale, and the need to manage locality, power, and possibly resilience. The gaps that are currently present in the software stack are indicated by how the software stack deals (or does not deal) with these issues.

NERSC aims to provide users with a stable software environment while allowing for periodic technology refreshes and the ability for users to construct and customize their workflows to maximize their productivity. Some of the open source software packages used at NERSC are supplied and supported by vendors, and some are supported internally at NERSC. As we consider the introduction of new architectures and programming paradigms, it is expected that this model of software support will continue. NERSC has gained considerable experience in supporting elements of the open source software stack, and we will leverage this experience going forward for the initiatives proposed in Section 6.1.

Under both NERSC exascale initiatives, we expect to partner with vendors and software developers and to work closely with our users to ensure the software areas listed above are ready for effective use in newer systems and with newer computing methodologies. When using testbeds, we expect that as operating system and runtime software develop, we will utilize them on the testbed systems to help evaluate their production readiness. We also expect to leverage the considerable expertise at Berkeley Lab in relevant areas (profiling, resilience, PGAS languages, etc.).

As the recent *Exascale Operating Systems and Runtime Software Report*[1] indicates, there is substantial agreement among the community of vendors, academic researchers, and DOE labs that a community based collaborative effort will be needed to address software gaps in moving to future architectures. The initiatives proposed above take a step in that direction by ensuring that we will participate in such a collaborative process between vendors and other DOE research efforts, always using the needs of our users as a guide.

## 6.3 Technology Strategy

NERSC's technology strategy will be based upon implementing science-driven solutions to meet the needs of our user community. In order to achieve this, our strategy we will be guided by the following principles:

- **Open competition for best solutions.** Open procurements enable us to gain value-for-money and ensures that NERSC's vision reaches the broadest possible audience. They also allow us to mitigate risk and not introduce a dependency upon the success of a single company or piece of technology as much as possible. We do plan to have earlier procurements so that we can have a greater influence on design.
- **Focus on the performance of a broad range of applications, not benchmarks.** Users care about application performance, not TOP500 position.
- **Smooth progression to exascale from a user's point of view.** Application code lifetimes are significantly greater than the lifetime of any one machine; it is therefore essential to have evolutionary changes to the programming model in order to maximize the productivity of our users. We seek solutions that do not require the use of proprietary languages or assembly-level coding in order to reach reasonable application performance.
- **Support for legacy code, albeit at less than optimal performance.** Solutions that allow current MPI-only applications to recompile and run with minimal effort, even with less than optimal performance, are essential to lower the barrier to entry to new architectures.
- **Performance portability — reasonable performance with the MPI+OpenMP programming model.** We seek solutions that do not require significant amounts of architecture-specific optimization.

[1] http://science.energy.gov/~/media/ascr/pdf/research/cs/Exascale%20Workshop/ExaOSR-Report-Final.pdf

- **Engage co-design centers and leverage fast forward and design forward.** The challenges facing the HPC community are significant and a broad national effort focused on software and hardware innovation is needed. NERSC will partner with early exascale efforts to take advantage of breakthroughs and to advocate for our users' needs.
- **General-purpose architectures are needed in order to support a wide range of applications, both large-scale simulations and high volumes of smaller simulations.** Our requirements reviews determined that some users have the need to run many different applications at different scales to achieve their scientific objectives. The demand for high volumes of smaller jobs is due to increased demand for uncertainty quantification and other ensemble-based approaches.
- **Separate data-centric resources will be needed.** Our analysis and experience shows that data-oriented systems need to be optimized for maximum data capacity, throughput, and global network bandwidth. They will be heavy on both memory capacity and I/O bandwidth (to both local and global file systems). By contrast, we expect our compute-oriented systems to be optimized for maximal computational density and local bandwidth, as we expect these applications to be able to exploit locality to a greater degree.
- **Consolidation of data resources is essential.** This allows us to gain significant efficiencies and address elasticity of demand from different user communities. It will also allow for greater overall capabilities, for example, a consolidated data system would provide significantly larger I/O bandwidth to each user than a series of clusters for each program.

### 6.3.1 Acquisition Strategies

A detailed consideration of these guiding principles and a desire to maximize the useful lifetime of our platforms has led us to expect to trigger a deployment of major computational systems every three years, beginning in early 2016. This three-year cadence was derived from our determination of user demand, and it will still allow us to partner with Los Alamos/Sandia on major procurements in 2016 and 2019. We have validated this strategy against industry roadmaps. Specifically, the trigger for each procurement will be determined by a careful analysis of the technology options available in the timeframe; in order to maximize the useful lifetime of a piece of technology before obsolescence, it is essential that we deploy it near the beginning of its availability. This strategy will ensure NERSC continues to field cutting edge technologies that maximize the scientific output of our users.

We also propose to field dedicated data systems every three years, beginning in 2017. To meet the needs before then, we are going to augment NERSC-7 to increase its data-oriented capabilities. We will increase the amount of storage, deploy a flash-based file system, and increase the analysis and visualization capabilities with additional large memory nodes.

We plan to continue our successful testbed strategy. By deploying small resources that contain early generations or prototypes of key technologies, we are able to understand in great depth the strengths and weaknesses of each new approach. So far, NERSC has successfully fielded testbeds to evaluate GPUs, flash disks, and Hadoop, and we will soon install an Intel MIC system. This ensures that we are as informed as possible about the applicability of any new technology, and helps our users prepare for technology changes.

### 6.3.2 Partnership Strategies

In order to ensure the success of these deployments in the current climate of disruptive technological change, we expect that NERSC will need to engage in a number of key partnerships. Recognizing this, we are planning to begin our procurement processes earlier than in the past. For NERSC-8 and NERSC-9, we will start the procurement process earlier so that vendor selection is two years before the machine is due to be deployed, and three years before for NERSC-10. Especially important in the near term will be the collaboration with the selected vendor for NERSC-8 and the Los Alamos/Sandia team.

In the longer term, partnerships with processor, interconnect, and storage vendors will all be essential to the successful outcome of our overall strategy. Specifically, we plan to provide as much detailed information about our workload to vendors as possible and engage in the co-design process, to ensure the success of both the exascale and data initiatives. We expect to engage in these partnerships with as many vendors as are interested, in order to minimize our risk and to ensure that a vibrant HPC ecosystem continues to exist. We have already begun this process as part of NERSC and LBNL's participation in the Co-Design, Fast-Forward, and Design-Forward efforts.

As described in Section 6.1, NERSC will undertake a co-design initiative with the specific goal of ensuring that exascale era technologies will meet the needs of DOE's science mission. We will also need to fund non-recurring engineering to advance key system integration technologies. For the NERSC-8/ Trinity procurement we identified these areas for potential investment as advanced power management and the development of a "burst buffer" flash-based file system, and plan to include them in the final RFP.

We also expect to continue our role of intellectual leadership within the HPC community. LBNL staff are involved in HPC-oriented projects with many other federal agencies, such as DOD, DARPA, and NSF, and we will work to ensure that these inform and enhance our abilities to meet our strategic goals.

## 6.4 Power, Space, Cooling

NERSC is currently located in the Oakland Scientific Facility (OSF) in Oakland, CA, but will move to a new facility located on the main LBNL campus in Berkeley in early CY2015. The Computational Research and Theory Facility (CRT) will provide space and power for NERSC to expand its support of Office of Science programs. The location will increase cooperation with the Computational Research Division and the Scientific Networking Division (ESnet), and will provide opportunities for joint projects with UC Berkeley.

The CRT is funded primarily by the University of California, with a construction budget of $125M. Data center provisioning (up to 12.5 MW) is funded by DOE with $19.8M of additional funding (allocated from the NERSC budget). CRT is designed for modular expansion of power and cooling subsystems, and the floor space can be expanded by approximately 50% by finishing a shelled space. Budget permitting, NERSC plans to undertake facilities upgrades to support its systems roadmap, as shown in Table 8.

**Table 8. NERSC Facilities Capacity**

| | OSF | CRT (move-in) | CRT 2016/17 | CRT 2019 | CRT (capability)* |
|---|---|---|---|---|---|
| Total Power | 11 MW | 12.5 MW | 20 MW | 35 MW | 42 MW |
| Max System Power / Cooling Capacity | 9 MW | 10 MW | 17.5 MW | 30 MW | 38 MW |
| PUE | 1.3 | 1.1 | 1.1 | 1.1 | 1.1 |
| Space | 19 Ksf | 20 Ksf | 20 Ksf | 20 Ksf | 29.9 Ksf |

*CRT has been provisioned with electrical feeders from the main LBNL substation having a capacity of 42 MW. Further expansion is possible with additional site modifications.

CRT will provide the lowest operational costs of any ASCR or NNSA supercomputing facility. CRT is designed to utilize "free" air and water cooling (Figure 14). By leveraging the benign Bay Area climate, 74°F air and 75°F water can be provided year-round without mechanical cooling. This will provide exceptional energy efficiency, with a PUE of less than 1.1. In addition, the LBNL site has access to WAPA power, which is lower cost (about half the price of commercial power at OSF) and very clean (primarily hydro and nuclear with some natural gas).

NERSC will move to CRT in CY2015. The Edison system, storage systems, and midrange systems will be migrated to CRT over the year. The Hopper system will stay at OSF until retirement in FY2016. Some tape libraries will remain at OSF indefinitely and provide for redundant physical copies for improved disaster recovery. NERSC has options to continue to use some or all of OSF through 2025 for future applications.

The CRT data center will house systems for NERSC, PDSF (HEP/NP), JGI (BER), the Kbase project (BER), the Materials Project, and potentially other systems sponsored directly by Office of Science program offices. Facility and power costs are shared according to agreements with the program offices, with the concurrence of ASCR. ESnet will also house systems in CRT. NERSC and ESnet share operations staff and the control room space.

NERSC will provision power based on sustained, not peak, consumption. Sustained power typically has been ~80% of peak, and trends suggest that the difference will increase as processors utilize more complex approaches to power optimization. Power capping technology now becoming available will enable us to ensure that peak surges do not occur and overload the center. By planning for sustained power only, we can avoid the cost of provisioning power that will not be used most of the time, and can deliver more computational capacity within a fixed power budget. (A consequence is that it may not be possible to measure TOP500 HPL performance for our systems.)

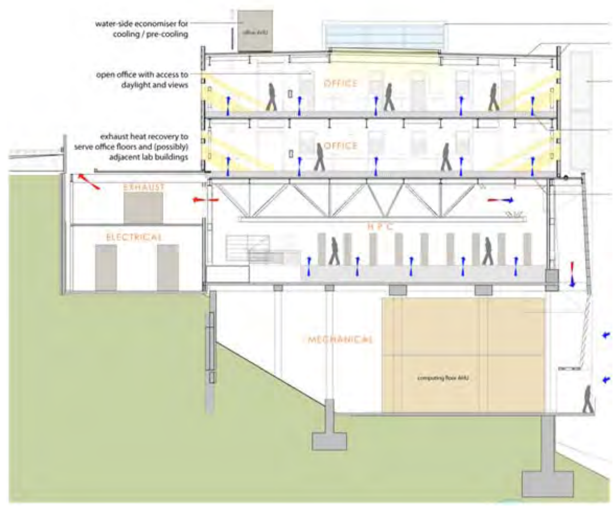



Figure 14. Computer rendition of the CRT Facility free air cooling approach.

## 6.5 Summary Roadmap and Timeline

### 6.5.1 Computational Systems

NERSC plans to deploy large-scale computational systems every three years, beginning with NERSC-8 in 2016. Each system has an expected lifetime of six years, but may be extended by midlife upgrades. Adjustments to this timeline will be made to minimize the chance of deploying stale technology. Likewise, NERSC plans to deploy extreme data systems on a three-year schedule beginning in 2017. Major deployments will thus alternate between computational and data systems.

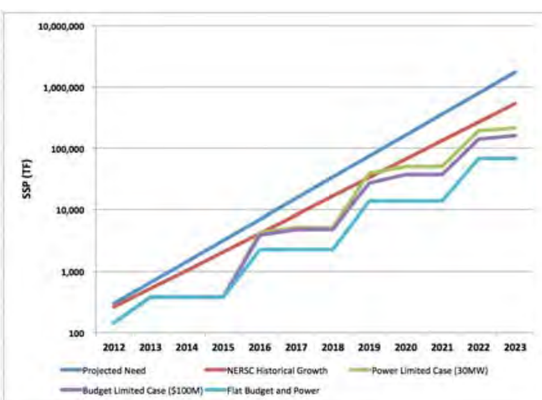
The systems roadmap is shown in Table 9. The most significant technology transition in this roadmap occurs with NERSC-8, which introduces hybrid or manycore architecture, and will require significant enhancements to applications to efficiently use the technology.

There is much uncertainty about budgets and the speed with which the exascale program will advance. Consequently, it is difficult to accurately predict cost or performance of these systems. The table assumes that a DOE exascale program is funded and 30 MW of facility power is available (per ASCR guidance), but with no budgetary constraints. In practice, NERSC will seek to maximize sustained performance given available technology, budget, and power at the time of acquisition.

**Table 9. NERSC Systems Roadmap**

| | NERSC-8 (2016) | NERSC Data-1 (2017) | NERSC-9 (2019) | NERSC Data-2 (2020) | NERSC-10 (2022) | NERSC Data-3 (2023) |
|---|---|---|---|---|---|---|
| System Average Power (MW) | 6 | 2 | 11 | 2 | 11 | 2 |
| Linpack FLOPS (PF) | 50 | 10 | 320 | 80 | 1,120 | 170 |
| Sustained Perf (SSP) / Hopper | 28 | 6 | 238 | 73 | 1,026 | 154 |
| Memory (PB) | 2 | 3 | 10 | 7 | 30 | 20 |
| System Dedicated Storage (PB) | 50 | 60 | 200 | 140 | 610 | 450 |
| I/O Bandwidth (TB/s) | 0.9 | 2 | 4 | 3 | 12 | 9 |
| System Space (sf) | 5,250 | 2,500 | 4,600 | 1,700 | 3,900 | 1,700 |

Since there will be multiple systems in operation at one time, aggregate capacity available is higher than that of a single system. Figure 15 shows installed capacity over time, including computational and data systems. In addition, the figure shows several scenarios in order to characterize uncertainty: facility power limited (30 MW as in the table above), a system procurement budget twice recent procurements ($100M), and a flat budget. These are compared with the projected need and NERSC's historical trendline.



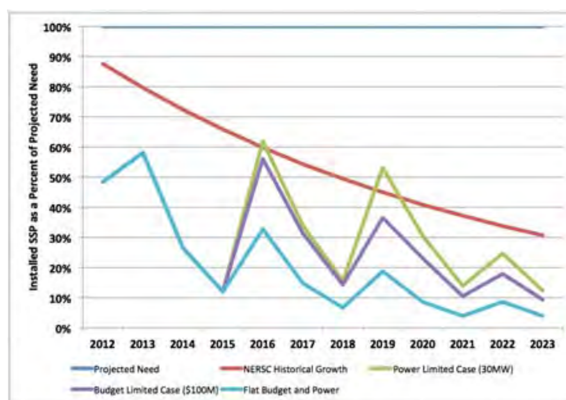




**Figure 15.** Projections of installed performance and percentage of need fulfilled.

A substantial gap exists between need and ability to deliver. In part, this can be solved with increased budgets, but technology acceleration is also crucial. The exascale program is not aggressive enough to provide systems with cost/performance and power efficiency (performance/watt) that will enable NERSC (at any realistic budget level) to keep up with aggregate demand for computational resources.

### 6.5.2 Storage and Network Systems

Storage at NERSC currently includes local scratch file systems for HPC systems to minimize contention between systems for high bandwidth storage and center-wide shared file systems called the NERSC Global Filesystem (NGF) to minimize data movement within the Center and provide maximum capacity for large data collections and collaborations. We expect to shift toward exclusively global over the next five to ten years with both capacity and bandwidth-oriented tiers. Archival storage for long-term and high-volume data storage demands will continue to be provided using HPSS.

The primary hardware used in current storage systems are magnetic disk and tape to meet capacity and bandwidth needs. Recently NAND flash technologies are augmenting storage system capabilities, mostly to improve bandwidth and decrease latency in existing storage systems. Disruptive storage technologies are in development; deployment may begin as early as 2015, but will certainly impact the production storage system architecture by 2017.

The archival storage system (HPSS) will be required to support a capacity above 10 exabytes by 2023 (see Figure 3 in Section 2.2). The ratio of disk to tape in the center has been increasing; in 2012, it was approximately 11%. Therefore, the center-wide storage system (NGF) capacity will need to increase to a capacity of well over 1 exabyte by 2023. This does not include system scratch storage. To operate efficiently, both disk and tape systems need reserve capacity. Additional tape storage is also required for disk backup and dual copies of portions of the user data archive. Total projected need and provisioned capacity are shown in Table 10.

Storage and data transfer accounts for 50–75% of network utilization (WAN and LAN) at NERSC. There is a direct correlation between the network capacity and storage capacity. For this reason, NERSC uses a continual and early adoption methodology to deploying new networking and storage technologies. If these ratios are maintained, we will need terabit Ethernet networking capability by the 2017–2020 timeframe.

**Table 10. Projected Storage Need and Provisioned Capacity.**

| | Projected (PB) | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | 2019 | 2020 | 2021 | 2022 | 2023 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| NGF | Need | 3 | 5 | 9 | 15 | 27 | 50 | 90 | 150 | 270 | 470 | 830 | 1,480 |
| | Capacity | 8 | 14 | 22 | 36 | 60 | 100 | 165 | 270 | 440 | 725 | 1,190 | 1,965 |
| HPSS | Need | 24 | 42 | 74 | 129 | 225 | 395 | 690 | 1,210 | 2,115 | 3,695 | 6,470 | 11,315 |
| | Capacity | 32 | 57 | 100 | 210 | 370 | 635 | 1,115 | 1,955 | 3,430 | 6,020 | 10,570 | 18,560 |

# 7. Human Resources

## 7.1 Current Workforce

NERSC currently has 72 employees that are funded from the NERSC program and overhead, and another 20 funded by other DOE program offices, research programs, and laboratory overhead. A staffing chart is shown in the Appendix.

The major gaps between the present workforce and what is necessary to deliver on the mission today are in the following areas:

- **HPC Consulting:** HPC consulting FTEs have actually been reduced from ten in the mid 1990s supporting 2,000 users, to nine today supporting 4,500 users. To deliver excellent user services to all users, we focus on efforts that "float all boats," relying especially on training and web documentation. We are short on staff in both these areas and seek to add an additional two FTEs.
- **Technology Integration:** As user workloads become more diverse, there is a need to write middleware to support workflows, databases, and data formats. This area of applied research is not currently represented in NERSC's FTE count. An infusion of two FTEs is needed for these activities.
- **Science Gateways:** JGI represents a data-intensive user community that NERSC supports with direct funding from JGI. Three JGI FTEs are devoted to creating and maintaining science gateway (web portal) infrastructure for this user community. Past NERSC efforts to create gateways for the astrophysics community also required three FTEs during the development phase. There is a growing interest in creating these portals, and NERSC currently devotes 1.5 FTEs to this effort. An infusion of 1.5 FTEs is needed to bring the total to three.
- **Storage Systems:** Storage is the most rapidly growing area at NERSC, driven by the increasing demands of science programs for data. The current staff is straining to keep up with the combined effort to maintain existing systems, install new capacity, and evaluate new technologies such as flash storage, burst buffers, and object-oriented file systems. In order to handle the current workload, NERSC has had to withdraw one FTE from the joint national laboratory/IBM consortium which drives development of HPSS. We seek to add an additional three FTEs in this area.

## 7.2 Future Workforce

Staffing gaps were identified for our major initiatives in Section 6.1.

## 7.3 Obstacles and Strategies

NERSC is fortunate to be located in the Bay Area and has access to a vast pool of qualified talent. Berkeley Lab and NERSC are a place where HPC workers aspire to be, despite fierce competition for technical talent in the area.

A significant challenge has been developing new leadership. NERSC has a number of senior workers, and we aggressively take advantage of Berkeley Lab's Leadership Development Program to train future leaders. Several graduates from this program have already become group leads and department heads. The leadership program teaches skills and provides hands-on training in:

- Building high performing teams
- Strategic and scenario planning
- Innovation
- Leadership styles, presence, communication, and rapport building
- Negotiating and conflict resolution.

Another challenge is the increasing overhead cost for staff to cover benefits such as medical care and pensions. Even without growth, staff cost is anticipated to rise by approximately one-third through 2023, and could seriously impact available budget for system acquisition and other initiatives.

# 8. Conclusion

NERSC is the primary computing facility for the Office of Science (SC). Our focus is maximizing the scientific productivity of our users, and the results can be seen in NERSC's very high scientific impact based on publications and awards. Our users typically publish 1,500 journal articles and 10 journal cover stories annually based on computations and data stored at NERSC. We contributed significantly to two Nobel Prizes and to one of Science Magazine's Top Ten Breakthroughs of 2012. Another important metric for productivity is user satisfaction, which we maintain at a very high level every year.

NERSC's role within the Office of Science is unique:

- We directly support DOE's science mission. SC program managers and the DOE Deputy Director of Science allocate most of the resources at NERSC.
- Future computing, storage, and networking requirements are derived from reviews with all six SC program offices.
- We support the largest user community within the DOE. We currently have 4,500 users with a net growth of 350 users per year.
- We are a data center for DOE. For the last four years, we have been a net importer of data. During a typical month, users transfer about a petabyte of data to NERSC for storage, analysis, and sharing.

To maintain our unique role and productivity in a changing scientific and technological environment, NERSC has two 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, and influencing the computer industry to ensure that future systems meet the mission needs of SC.
- Increase the productivity, usability, and impact of DOE's scientific user facilities by providing comprehensive data systems and services to store, analyze, manage, and share data from those facilities.

**The aggregate computing needs of SC science teams at NERSC are well into the exascale regime by the end of the decade.** Science teams need to run simulations at hundreds of petaflops, and they need to run thousands to millions of petascale simulations. We will deploy pre-exascale systems in 2016 (NERSC-8) and 2019 (NERSC-9). We anticipate deploying our first exascale system, NERSC-10, in 2022. We will begin enhancing our data capabilities starting in 2014, and **we will deploy data systems, storage, advanced networking, and enhanced user services so that current users and DOE experimental facilities can move and process exabytes of data early in the next decade.**

This plan will be reviewed and revised periodically with input from the regularly scheduled SC program office requirements reviews, the annual NERSC Policy Board meetings, the NERSC Users Group meetings, and ongoing discussions with ASCR.

It is hard to predict ten years into the future. There is much uncertainty about budgets and the speed with which the exascale program will advance. Consequently, it is difficult to accurately predict cost or performance of future systems today. While the roadmap described here will almost certainly need to be revisited every few years, our strategy of staggered system procurements provides the flexibility to revise our plans if necessary, and our testbed work will allow us to validate our approach before committing to it. In short, we think that our vision for the next ten years provides a sound basis for planning and action.

# Appendix: Current NERSC Staffing

| Group/Area | Program & Overhead FTEs | Other Funded FTEs |
|---|---|---|
| **Center Leadership** | **3.2** | |
| **Center Administration** | **6** | |
| Project/Risk Management | 1 | |
| Administrative support | 3 | |
| Budget Analyst | 1 | |
| Procurement | 1 | |
| **Facilities** | **2** | |
| **Science and Report Writing** | **1.5** | |
| **User Services** | **11** | **3** |
| HPC Consulting | 9 | |
| Account Support | 2 | |
| JGI Consulting | | 2 |
| PDSF Consulting | | 1 |
| **Analytics and Visualization** | **3.5** | |
| **Outreach, Software & Programming** | **5** | **2** |
| Science Gateways & Outreach | 2.5 | |
| HPC Infrastructure Programming | 2.5 | |
| Research Grant Funded | | 2 |
| **Advanced Technologies & Workload Analysis** | **4** | |
| **Postdoc Program** | | **3** |
| **Computational Systems** | **9** | **2** |
| Programmatic Systems | 9 | |
| JGI Computational Systems | | 1 |
| PDSF Systems | | 1 |
| **Storage Systems** | **8** | **3** |
| File Systems | 4.5 | |
| Mass Storage | 3.5 | |
| JGI Storage & Database Systems | | 3 |
| **Networking, Security & Servers** | **10** | **3** |
| Networking | 3.5 | |
| Security | 3 | |
| Server Systems | 2.5 | |
| Desktop Support | 1 | |
| JGI Servers | | 3 |
| **Technology Integration** | **1** | **1** |
| JGI Applications Support | | 1 |
| **Operations Technology (24 x 7)** | **7.5** | **3.5** |
| **TOTALS** | **71.7** | **20.5** |

## DISCLAIMER

Find out more at: **www.nersc.gov.**
For more details and the latest news, visit **www.lbl.gov.**