NERSC 2011

High Performance Computing Facility Operational Assessment

for the National Energy Research Scientific Computing Center

April 11, 2012

Revised May 9, 2012

Ernest Orlando Lawrence Berkeley National Laboratory
1 Cyclotron Road, Berkeley, CA 94720-8148

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.

Contents

Executive Summary/Background 1

Communications with Users and Key Stakeholders 2

New NERSC Website 2

Communications with DOE Program Managers 2

Communications with NERSC Users 4

Communications with Vendors 4

Science Requirements Workshops 4

Summary of 2011 Metrics 4

Responses to Recommendations from the August 22, 2011 OA Review 10

1 User Support Results 16

1.1 User Support Metrics 18

1.1.1 Metric: Key Average Scores Are Satisfactory 20

1.1.2 Metric: Key Average New User Scores Are Satisfactory 20

1.1.3 Metric: Improvement on at Least Half of Below-Satisfactory Scores on the 2010 User Survey 21

1.1.4 Key Average Scores for Large, Medium, and Small Users 22

1.1.5 Supplemental Information from the 2011 Survey 22

1.2 Problem Resolution Metrics 25

1.3 User Support and Outreach 26

1.3.1 Providing Rapid, Quality Front-Line User Support 26

1.3.2 Compiler Comparisons 28

1.3.3 NERSC Users Learn How to Use Hopper's Multi-Core Nodes Effectively 28

1.3.4 Facilitating Statistical Computing and Graphics Generation with Parallel R 30

1.3.5 Visualizing Electronic Structures from Quantum Monte Carlo Simulations 30

1.3.6 Improving Access to X-Windows Based Tools and Applications 31

1.3.7 Improving the Efficiency of a User Code 32

1.3.8 Improving Account Support 32

1.3.9 Providing Comprehensive Support for an ALCC Project 32

1.3.10 Software Support 33

2 Business Results 35

2.1 Resource Availability 39

2.1.1 Scheduled Availability 42

2.1.2 Overall Availability 42

2.1.3 System Mean Time To Interrupt (MTTI) ... 43
2.1.4 System Mean Time To Failure (MTTF) ... 44
2.2 Resource Utilization ... 45
2.2.1 Total System Utilization ... 46
2.2.2 HPSS Resources ... 46
2.2.3 Global File System Resources ... 46
2.2.4 Network Resources ... 49
2.3 Capability Utilization ... 51

3 Strategic Results ... 54
3.1 Science Output ... 54
3.2 Scientific Accomplishments ... 54
3.2.1 Solar System's Magnetic Edge More Turbulent Than Expected ... 55
3.2.2 LED Lighting Breakthrough from Materials Science ... 55
3.2.3 Greenhouse Gas Remediation ... 56
3.2.4 Supernova Caught in the Act ... 58
3.2.5 A Better Way to Identify Extreme Weather Events in Climate Models ... 59
3.2.6 The Speculative ^{14}F Nucleus: Simulation Precedes Experiment ... 59
3.2.7 Heaviest Antimatter Nucleus Detected ... 60
3.2.8 Promise for Solar Energy on Demand ... 61
3.2.9 AMR for a Clearer Picture of Carbon Sequestration ... 62
3.2.10 Climate Studies Earn NERSC and PI an HPC Innovation Award ... 62
3.2.11 Understanding Catalyzed Growth of Carbon Nanotubes ... 63
3.2.12 A Selection of Papers Featured as Journal Covers (2010-2011) ... 64
3.3 Allocation of NERSC Director's Reserve of Computer Time ... 67

4 Innovations ... 71
4.1 Innovations to Improve Usability of HPC systems ... 71
4.1.1 Supporting Diverse Workloads on HPC Platforms ... 71
4.1.2 Increasing Scientific Productivity by Tracking Data ... 73
4.1.3 Safeguarding Data with Parallel Incremental Backups ... 74
4.1.4 Data Transfer Nodes Facilitate Scientific Collaborations ... 75
4.2 Innovations to Provide New Computational Models to Users ... 76
4.2.1 Provide GPU Testbed and Evaluation ... 76
4.3 Innovations to Increase NERSC Center Efficiency ... 78

5 Risk Management ... 79
5.1 Risk Management Strategy ... 79

5.2 Current Risk Status 79
5.3 Major Risks Tracked for the Current Year (2011) 80
5.3.1 Budget Insufficient to Meet DOE (Allocation) Commitments 80
5.3.2 Hopper Phase 2 Stability 81
5.3.3 Loss of Key Personnel (Temporary or Permanent) 82
5.3.4 Disruptive Vendor Changes: Oracle Acquisition of Sun Microsystems 82
5.3.5 OSF Computer Room Insufficient in the Short Term 83
5.3.6 Budget Uncertainties 84
5.3.7 Computer Room Safety 84
5.4 Risks That Occurred and the Effectiveness of Their Mitigations 85
5.5 Retired Risks During the Current Year 86
5.6 Major New or Re-Characterized Risks Since the Last Review 86
5.6.1 Annual Budget Plan under Continuing Resolution 87
5.6.2 Facility Impacts on Tape System Reliability 87
5.6.3 CRT Facility Occupancy Delayed 88
5.6.4 Disruptive Vendor Changes: Oracle & Tape (HPSS) 89
5.6.5 NERSC Fails to Meet IPv6 Compliance for NERSC Systems and/or Networks 90
5.6.6 Disruptive Vendor Changes: Oracle Lustre 90
5.6.7 Failure of Largest System 91
5.7 Major Risks Next Year, with Mitigations 92

6 Summary of the Proposed Metric Values for Future OARs 93
6.1 Rationale and Use of Proposed Metrics and Targets 95
6.1.1 User Survey Metrics 95
6.1.2 Problem Resolution Metrics 95
6.1.3 Availability Metrics and Reported Numbers 96
6.1.4 Capability Metric 96

Executive Summary/Background

The National Energy Research Scientific Computing Center (NERSC) is the high-end scientific computing facility for the Department of Energy's Office of Science (DOE SC). With more than 4,000 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 for DOE'S unclassified research programs in alternative energy sources, climate change, energy efficiency, environmental science, and other fundamental science areas within the DOE mission.

Because NERSC is the DOE SC's mission computing facility, DOE program managers across SC allocate most of the center's computing resources. In 2011, the six programmatic offices allocated 80 percent of NERSC computing time, while DOE's ASCR Leadership Computing Challenge allocated 10 percent. The remaining 10 percent was allocated through the NERSC Initiative for Scientific Exploration (NISE) program. In Allocation Year 2011 (January 12, 2011–January 9, 2012), 1.075 billion NERSC computer hours were allocated in total.

NERSC's primary mission is to accelerate the pace of scientific discovery by providing high-performance computing, information, data, and communications services to the DOE Office of Science community. In 2011, our users co-authored more than 1,500 refereed papers based on the computations performed at NERSC. Discoveries ranged from the tiny to the vast: computational prediction of the structure of the highly unstable fluorine-14 nucleus led to experiments that achieved the first observation of this isotope, while simulations of the magnetic edge of our solar system found unexpectedly high levels of turbulence. NERSC's integrated resources and services enabled the earliest-ever discovery of a supernova—within hours of its explosion—providing new information about supernova explosion dynamics. Addressing critical energy and climate issues, NERSC users achieved breakthroughs in LED lighting, solar thermal fuels, carbon capture and sequestration, and identification of extreme weather events in climate model datasets.

Over its 15-year history at Berkeley Lab, the NERSC program 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-based architectures. Results of the NERSC User Survey show a high level of satisfaction with NERSC systems and services; in fact, the Overall Satisfaction with NERSC score for 2011 was the highest ever recorded in the 13 years the survey has been in its current form. Expanded training efforts and improved data analysis and visualization support contributed to these favorable results. User support and outreach highlights included:

- Rapid, quality front-line user support
- A comparative study of compiler performance, resulting in new tutorials and performance tips
- A series of tutorials, talks, and presentations to educate users on how to use the 24-core Hopper nodes effectively
- Facilitating statistical computing and graphics generation with Parallel R
- Visualizing electronic structures from quantum Monte Carlo simulations
- Improving access to X-Windows based tools and applications
- Collaborating with users to improve code efficiency
- Improved account support and software support

NERSC developed and implemented several innovations in 2011 that improve the usability of HPC systems, scientific productivity, and center efficiency. These innovations include:

- A task farmer that simplifies running high-throughput workloads
- Tools to create a virtual private cluster (VPC) within a parallel allocation
- A file cacher to improve IO efficiency

- A dynamic database the continually collects statistics on data movement and storage, helping to identify immediate bottlenecks as well as longer term trends
- Safeguarding data with parallel incremental backups
- The Dirac GPU testbed
- Improved machine room monitoring

Communications with Users and Key Stakeholders

NERSC regularly communicates with and solicits feedback from its key stakeholders — the Department of Energy program managers, NERSC users, and vendor partners — through a number of channels detailed below. NERSC also communicates with other DOE HPC centers as needed to address issues of common interest and find solutions to problems. For example, NERSC led the effort to develop common definitions and formulas across the three centers for our operational assessment reports.

New NERSC Website

NERSC's new, modern website was designed to improve how NERSC communicates with its users, DOE managers, peers in the HPC community, and the general public. The website has prominent sections for Live Status, For Users, News, Events, Systems, and About NERSC (Figure ES-1).

The new website contains presentations from training classes, video recordings of those classes, popular presentations given by NERSC staff, and workshop reports. The 2011 NERSC user survey gives one example of the users' favorable response:

> *"NERSC's website has an enormous amount of good information and Google frequently takes me to it when I am looking for generic answers. The new website has been very easy to navigate."*

Figure ES-1. NERSC's new home page provides easy access to a wealth of information.

Communications with DOE Program Managers

NERSC holds weekly calls with both DOE headquarters and local site office managers. In addition, other NERSC projects are managed through close communication between DOE and NERSC. NERSC communicates science and technology news and innovations to DOE through frequent updates to the News & Publications page of the NERSC web site, quarterly science highlights, and the annual report. In 2011, NERSC

created a new web site that includes additional pages targeted at DOE program managers, including an online live "dashboard" (Figure ES-2) and an expanded News section (Figure ES-3).

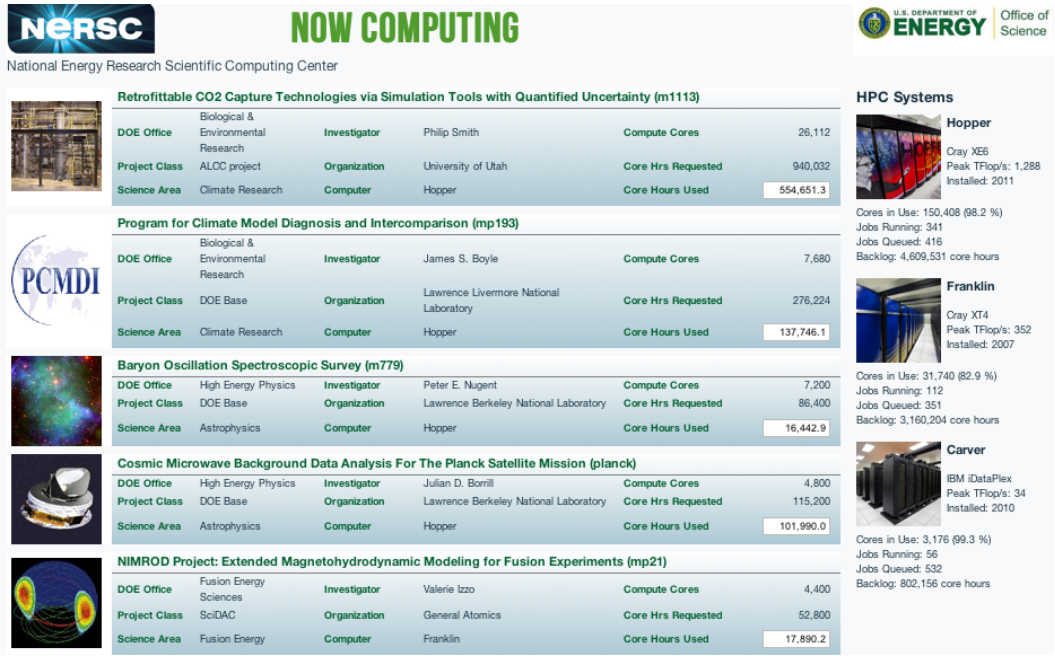


Figure ES-2. NERSC's web "dashboard."

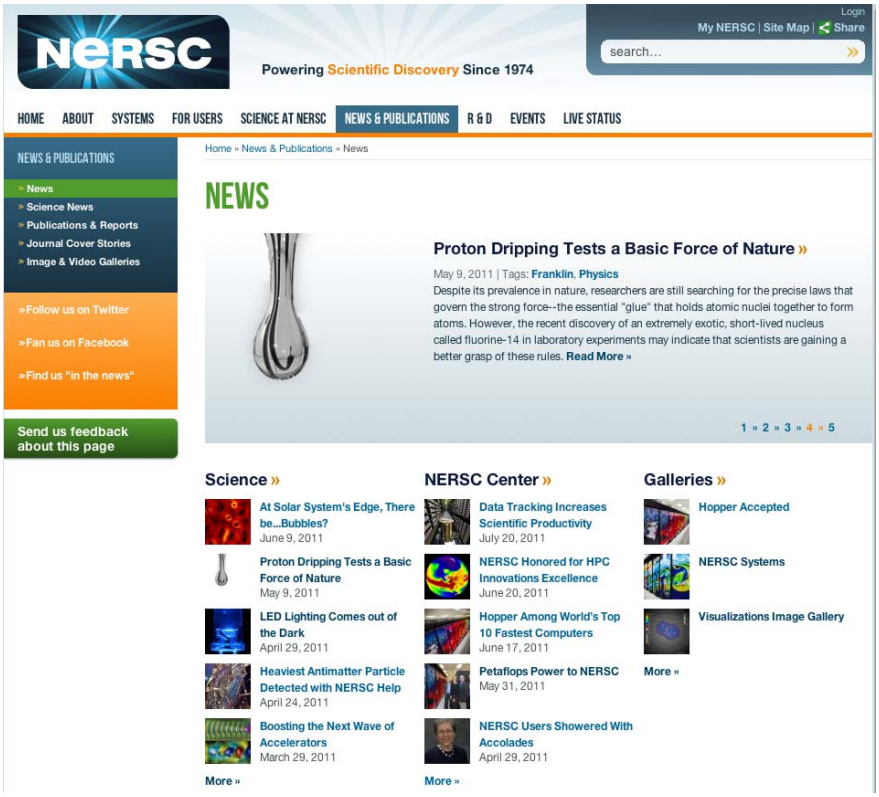


Figure ES-3. NERSC's new News webpage.

Communications with NERSC Users

NERSC communicates with users in a variety of ways. First, the NERSC Users Group (NUG) has monthly calls. 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. The October 2010 NUG meeting featured three days of training for the new Hopper system.

NERSC's new website features a prominent "For Users" top-level section that makes it easy for users to find information. NERSC's expanded training efforts in 2011 opened additional opportunities for communicating with users. NERSC also conducts the annual User Survey, in which users are asked roughly 90 questions to evaluate the NERSC systems and services. NERSC staff host and participate in workshops and conferences where many NERSC users are present. Finally, NERSC staffers talk with users daily when they call or submit tickets online. These informal discussions often lead to procedural changes and system optimizations.

Communications with Vendors

The NERSC Computational Systems Group has weekly calls with Cray to discuss any ongoing issues with the Hopper and Franklin systems and to plan for future activities on these systems. These meetings are held more frequently if needed. The NERSC Storage Systems Group has regular calls with Oracle Storage through the Large Tape Users Group to provide specific requirements and features to improve tape library and drive design. NERSC staff also participate in both the Cray Users Group and the IBM SPXXL user group to communicate specific software and hardware requirements for using the vendors' systems at scale.

Science Requirements Workshops

In 2009 NERSC and DOE initiated a set of scientific computing requirements workshops with key NERSC users in each of the six offices in the DOE Office of Science. The BER and HEP workshops were conducted in 2009; the BES and FES workshops in 2010; ASCR and NP in 2011. The workshops elicit information from program managers and key scientists about the current and future computational requirements for each science area managed by the particular program office. These requirements provide input to the NERSC architecture and planning processes, and help ensure that NERSC continues to provide world-class support for scientific discovery for DOE scientists and their collaborators. Three findings from the BER workshop are typical:

- Without a significant increase in available resources, scientific progress will be acutely limited.
- Many projects have mission-critical time constraints that demand guaranteed throughput on powerful resources.
- To exploit many-core (10K–1M) computers, scientists request access to testbed machines and assistance from NERSC in choosing effective programming models.

Summary of 2011 Metrics

In consultation with the DOE program sponsor, a series of metrics and targets were identified to assess NERSC's operational performance in calendar year 2011. The metrics are associated with a series of questions posed to reviewers of the center. The 2010 and 2011 metrics, target values, projections (July 1–December 31, 2011) presented in the first August 22, 2011 OAR, and actual results as of December 31, 2011 are summarized below. NERSC's actual 2011 performance exceeded all targeted performance metrics, as seen in the table below.

Table ES-1. Summary of 2011 Metrics (1 of 5)

Metric Description	2010 Target	2010 Actual	2011 Target	2011 Actual	Projected July 1-Dec. 31, 2011	Actual July 1-Dec. 31, 2011
User Survey Metrics	**Key average scores on the user survey are satisfactory (> 5.25 / 7)**					
Overall Satisfaction with NERSC: Average Score on User Survey	> 5.25	6.40	> 5.25	6.54	N/A	N/A
Average of All High-Level ("Overall Satisfaction") Scores on the User Survey	> 5.25	6.33	> 5.25	6.39	N/A	N/A
Average of All User Survey Scores	> 5.25	6.16	> 5.25	6.29	N/A	N/A
Average of All User Support (Services) Survey Scores	> 5.25	6.24	> 5.25	6.37	N/A	N/A
Annual user survey results show improvement in at least one-half of the questions that scored below satisfactory (5.25/7) in the previous year	Two scores were < 5.25 in 2009. At least one of these must be ≥ 5.25 in 2010.	One score was > 5.25 in 2010. The other no longer applied since Bassi had been decommissioned.	Two scores were < 5.25 in 2010. At least one of these must be ≥ 5.25 in 2011.	The two scores that were < 5.25 in 2010 were both > 5.25 in 2011.	N/A	N/A
Consulting Overall Average Score on User Survey	> 5.25	6.64	> 5.25	6.58	N/A	N/A
Consulting Response Time Average Score on User Survey	> 5.25	6.60	> 5.25	6.57	N/A	N/A
Quality of Technical Advice Average Score on User Survey	> 5.25	6.57	> 5.25	6.53	N/A	N/A
Consulting Time to Solution Average Score on User Survey	> 5.25	6.40	> 5.25	6.40	N/A	N/A
Problem Resolution Metric						
Eighty percent of user problems are addressed within three working days (resolved or user informed how problem will be handled)	≥ 80%	84.50%	≥ 80%	86.9%	85%	86.1%

Table ES-1. Summary of 2011 Metrics (2 of 5)

Metric Description	2010 Target	2010 Actual	2011 Target	2011 Actual	Projected July 1–Dec. 31, 2011	Actual July 1–Dec. 31, 2011
Scheduled Availability Metrics	**Scheduled availability is ≥ 90% for systems in their first year of production, ≥ 93% in their second year, ≥ 95% thereafter.**					
Franklin Scheduled Availability	≥ 93%	95.1%	≥ 95%	96.4%	95%	96.7%
Hopper Phase I Scheduled Availability	≥ 90%	98.2%	≥ 93%	99.6%	N/A	N/A
Hopper Phase II Scheduled Availability	N/A	N/A	≥ 90%	98.7%	97%	99.0%
Carver Scheduled Availability	≥ 90%	98.6%	≥ 93%	99.9%	98%	99.9%
Euclid Scheduled Availability	≥ 90%	97.9%	≥ 93%	99.9%	97%	99.9%
HPSS User System Scheduled Availability	≥ 95%	99.9%	≥ 95%	99.8%	98%	99.7%
Global Filesystem Scheduled Availability	≥ 95%	99.7%	≥ 95%	99.6%	98%	99.9%
Overall Availability Metrics	**Overall availability is ≥ 80% for systems in their first year of production and 85% thereafter.**					
Franklin Overall Availability	N/A	93.0%	≥ 85%	95.5%	92%	96.1%
Hopper Phase I Overall Availability	N/A	96.6%	≥ 85%	99.6%	N/A	N/A
Hopper Phase II Overall Availability	N/A	N/A	≥ 80%	97.6%	95%	98.3%
Carver Overall Availability	N/A	96.4%	≥ 85%	98.8%	97%	98.7%
Euclid Overall Availability	N/A	95.7%	≥ 85%	99.4%	97%	99.4%
HPSS User System Overall Availability	N/A	99.4%	≥ 85%	98.9%	98%	98.8%
Global Filesystem Overall Availability	N/A	98.7%	≥ 85%	99.0%	97%	99.6%

Table ES-1. Summary of 2011 Metrics (3 of 5)

Metric Description	2010 Target	2010 Actual	2011 Target	2011 Actual	Projected July 1–Dec. 31, 2011	Actual July 1–Dec. 31, 2011
Mean Time To Interrupt (MTTI) Results	**These are reported numbers, not metrics.**					
Franklin MTTI	N/A	4:09:47	N/A	5:04:53	N/A	5:21:23
Hopper Phase I MTTI	N/A	13:10:28	N/A	22:10:00	N/A	N/A
Hopper Phase II MTTI	N/A	N/A	N/A	7:15:51	N/A	10:15:22
Carver MTTI	N/A	15:17:57	N/A	30:01:15	N/A	25:22:33
Euclid MTTI	N/A	10:04:45	N/A	10:23:54	N/A	10:18:14
HPSS User System MTTI	N/A	11:04:40	N/A	10:22:32	N/A	11:08:43
Global Filesystem MTTI	N/A	72:05:25	N/A	40:03:39	N/A	45:19:46
Mean Time To Failure (MTTF) Results	**These are reported numbers, not metrics**					
Franklin MTTF	N/A	5:22:00	N/A	6:06:56	N/A	7:03:05
Hopper Phase I MTTF	N/A	21:12:24	N/A	22:10:00	N/A	N/A
Hopper Phase II MTTF	N/A	N/A	N/A	11:05:16	N/A	15:04:60
Carver MTTF	N/A	26:22:53	N/A	121:15:05	N/A	91:23:02
Euclid MTTF	N/A	17:08:24	N/A	17:08:33	N/A	22:23:40
HPSS User System MTTF	N/A	45:13:51	N/A	60:19:07	N/A	91:23:40
Global Filesystem MTTF	N/A	121:13:31	N/A	72:22:31	N/A	91:23:15

Table ES-1. Summary of 2011 Metrics (4 of 5)

Allocations and Usage Results	These are reported numbers, not metrics			
Allocation group	**2010 Allocated**	**2010 Hours Used**	**2011 Allocated**	**2011 Hours Used**
ASCR	11,973,000	20,971,000	47,450,000	48,479,000
BER	44,725,000	58,959,000	166,694,000	183,285,000
BES	75,175,000	127,118,000	272,127,000	380,409,000
FES	44,712,000	64,140,000	172,694,000	239,564,000
HEP	37,327,000	78,227,000	143,124,000	211,975,000
NP	26,088,000	58,588,000	95,911,000	178,717,000
ALCC	30,000,000	4,472,000	88,500,000	67,986,000
NERSC Director's Reserve (NISE)	30,000,000	43,453,000	88,500,000	114,065,000
Totals	300,000,000	455,928,000	1,075,000,000	1,424,480,000

Large System Utilization	Percent of up time the system's computational nodes run batch jobs: reported numbers, not metrics					
	2010 Target	**2010 Actual**	**2011 Target**	**2011 Actual**	**Projected July 1–Dec. 31, 2011**	**Actual July 1–Dec. 31, 2011**
Franklin Utilization	N/A	96.4%	N/A	93.7%	N/A	92.3%
Hopper Utilization	N/A	N/A	N/A	87.2%	N/A	87.7%

Capability Metric	On NERSC's largest machine in production, at least N percent of the cycles should be used by jobs running on 1/8th or more of the processors, where N is determined annually by the NERSC Program Manager. This is an OMB Fiscal Year metric.					
	FY 2010 Target	**Calendar 2010 Actual**	**FY 2011 Target**	**FY 2011 Actual**	**4Q FY 2011 Projected**	**4Q FY 2011 Actual**
Cycles used by jobs running on 4,096 or more of Franklin's cores	≥ 30%	58.3%	≥ 35%	49.0%	40%	38.0%

Table ES-1. Summary of 2011 Metrics (5 of 5)

Science Output	These are reported numbers, not metrics.					
	2010 Target	2010 Actual	2011 Target	2011 Actual	Projected July 1–Dec. 31, 2011	Actual July 1–Dec. 31, 2011
Number of refereed publications (published or in press) based on using (at least in part) NERSC resources	N/A	1,407	N/A	1,528	N/A	N/A

Responses to Recommendations from the August 22, 2011 OA Review

In August 2011, NERSC delivered a Facility Operational Assessment to the DOE program sponsor that documented the 2011 calendar year-to-date operational activities of the center. Recommendations provided by reviewers of that report, NERSC actions, and DOE ASCR comments are given in Table ES-2 below.

Table ES-2. Responses to Recommendations from the Previous 2011 OA Review (1 of 6)

Section 2. Are the processes for supporting the customers (users), resolving problems, and communicating with key stakeholders and Outreach effective?		
Recommendation	**NERSC Response**	**DOE Program Manager Response**
Include the average of all user support questions on user survey in the OA report.	NERSC has provided the user support score average in the revised 2011 OA report. This is the average of all "services" questions covering HPC consulting, account support, visualization and analytics support, web interfaces, allocations, and training. In the 2011 survey there were 21 such topics, and their average score was 6.37 out of 7.	
The report should list the response rate target that was negotiated with the DOE program manager and provide the statistical analysis described in the guidance.	In the revised 2011 OA report, NERSC has reported the response rate targets which were negotiated with the DOE program manager. For statistical analysis, NERSC uses two measures: (1) standard deviation and (2) t-test at the 95% confidence level to measure the significance of changes in scores between survey years. Both measures are reported in the survey results on the web; only the t-test was reported in the initial 2011 OA report. For the 2012 OA report, NERSC will include both, subject to ASCR management approval.	
The user survey response rate should also be reported as a percentage of all users surveyed.	NERSC invites all users with registered accounts ("authorized users") to take the annual survey, and we have reported this number in the revised 2011 OA report. NERSC would like to continue to be inclusive and maximize the number of users invited to take the survey, but we do not believe that the response rate based on the number of users invited to take the survey is the best way to judge participation. Rather, we believe the most meaningful reflection of the response rate is to report the response rates for large, medium, and small MPP users as well as the percentage of total hours used by those who answered the survey.	Disagree. The current guidance says: "The Facility will negotiate the target response rates for the user survey with their DOE Program Manager. The Leadership Computing Facilities will include sufficient demographic information such that the report can describe results by INCITE, ALCC, and Discretionary allocations. NERSC will report on larger and smaller users, where larger users are defined as those using more than 0.3% of total NERSC allocation hours." This method is more meaningful than the recommendation to PM.

Table ES-2. Responses to Recommendations from the Previous 2011 OA Review (2 of 6)

Section 2. Are the processes for supporting the customers (users), resolving problems, and communicating with key stakeholders and Outreach effective? (continued)		
Recommendation	**NERSC Response**	**DOE Program Manager Response**
The date the survey was sent and the time period of the survey should be provided.	NERSC agrees. The online survey was available to all NERSC users from June 6, 2011 to July 1, 2011. We have included this information in the revised 2011 OA report.	

Section 3. Is NERSC maximizing the use of its HPC systems and other resources consistent with its mission?		
Recommendation	**NERSC Response**	**DOE Program Manager Response**
NERSC should adopt the same availability metrics as the other centers (80%/85% for the first year and 90%/95% thereafter.	NERSC understands the request, but it would mean that NERSC would have to lower its availability targets. NERSC Is following current DOE guidance and will continue to do so.	
NERSC considers a job using 1/8, or 12.5%, rounded down to a power of two, to be a capability job. Using that definition, the NERSC metric is 35% of core hours must be capability. NERSC should use metrics more in line with those used by other facilities for their capability platforms.	NERSC disagrees and follows guidance from the NERSC Program Manager. NERSC's role is different from the role of the Leadership Centers. NERSC's goal is to maximize scientific output, not machine or job size. Furthermore, NERSC does not have sufficient computing capability to allow DOE/SC to provide very large allocations to any one user. Since NERSC must meet the needs of all DOE SC users and their wide range of computing needs, it makes sense that NERSC's capability target is lower than the capability target at the Leadership Centers. NERSC analyzes and projects job usage data every year, and together with its Program Manager sets the appropriate capability target for the next year.	
NERSC should provide CY 2011 capability projections, as outlined in the overview section on usage-related performance metrics in the OAR guidance.	NERSC disagrees. Capability usage is hard to project. It can be quite variable because it is constrained by remaining allocation for the largest projects. For example, the four FY11 quarterly values for Franklin's capability metric were 52.1, 55.6, 50.5, and 38.0. In July, when we would have had to make the prediction, we didn't yet know how much this value would change.	

Table ES-2. Responses to Recommendations from the Previous 2011 OA Review (3 of 6)

Section 3. Is NERSC maximizing the use of its HPC systems and other resources consistent with its mission? (continued)		
Recommendation	**NERSC Response**	**DOE Program Manager Response**
NERSC should describe how the projected values for CY2011 were determined.	This no longer applies moving forward, since we will be reporting actual data and not need to use projected data. When we made the projections on the original August 2011 OA report, projected values for availability metrics were derived from an analysis of past outages combined with anticipated maintenances and outages. Known upgrade plans for the remainder of the reporting period were also taken into consideration. Projected values for usage were proportional to the resources used between March and June. January and February were excluded because Hopper was still being tested by NERSC staff and Cray, and these months were not representative of usage by the NERSC users.	

Section 4. Is NERSC enabling scientific achievements consistent with the Department of Energy strategic goals 3.1 and/or 3.2?		
Recommendation	**NERSC Response**	**DOE Program Manager Response**
NERSC should consider tracking its publications for five years and provide the citations for these papers.	NERSC will follow ASCR management direction.	Disagee. OLCF has been asked to develop an automated publication tracking system that can be used by all ASCR Scientific User Facilities, including NERSC.

Section 5. Are the costs for the upcoming year reasonable to achieve the needed performance?		
Recommendation	**NERSC Response**	**DOE Program Manager Response**
NERSC should combine Tables 5.1, 5.2, and 5.3 for a more efficient and useful financial snapshot.	NERSC can change the financial tables to whatever format and whatever guidance we receive from ASCR. Tables 5-1 and 5-2, as presented, support the narrative requested by DOE. The final table, 5-3, was intended as an overall summary of FY11 and FY12 budget figures. Note: Moving forward, the financial section will only be required for onsite reviews.	

Table ES-2. Responses to Recommendations from the Previous 2011 OA Review (4 of 6)

Section 5. Are the costs for the upcoming year reasonable to achieve the needed performance? (continued)		
Recommendation	**NERSC Response**	**DOE Program Manager Response**
NERSC should provide a schedule with planned lease payments outstanding for current and future fiscal years until each lease is paid off. This will be useful to help show future liabilities (claims against future funds) and funds available for all other operating costs. The information is consistent with that provided by companies as part of their formal financial statements in accordance with GAAP.	NERSC will follow ASCR guidance. NERSC annual liabilities on leases can easily be included in the report. Currently, NERSC's total liability is $23,493K: $14,003 in FY12 and $9,490K in FY13. There are no additional lease liabilities.	Disagree. That level of detailed information was presented to DOE in the 2011 Budget Deep Dive.

Section 6. What innovations have been implemented that have improved NERSC's operations?		
Recommendation	**NERSC Response**	**DOE Program Manager Response**
NERSC has developed a number of technologies of potential relevance to other HPC facilities. Where possible, these technologies should be made available to the broader community through an open-source licensing model that is industry friendly (such as a BSD style license).	NERSC agrees that open distribution of such technologies to other facilities is advantageous to the community and will do so where possible.	
NERSC should provide innovations and best practices identified in the area of user services.	NERSC agrees and will report innovations and best practices, if any occur, identified in the area of user services in the 2012 OA report.	

Section 7. Is NERSC effectively managing risk?		
Recommendation	**NERSC Response**	**DOE Program Manager Response**
NERSC should develop triggers for their risks.	All NERSC risks have triggers.	

Table ES-2. Responses to Recommendations from the Previous 2011 OA Review (5 of 6)

Section 7. Is NERSC effectively managing risk? (continued)		
Recommendation	**NERSC Response**	**DOE Program Manager Response**
NERSC should consider developing additional mitigations for their sole high risk "Budget Insufficient to Meet DOE (Allocation) Commitments" (risk 7.3.1). For example, preparation of contingency plans in case of a reduced budget, austerity measures that can be implemented quickly, etc. Similar to the mitigations for risk 7.6.1. Communication as the mitigation can only go so far.	NERSC agrees that additional mitigations should have been explicitly stated for this risk. Although not shown in the OA report, additional mitigations are active.	
NERSC should provide their impact and probability matrices in their report.	NERSC agrees and will show impact and probability matrices in future OA reports.	

Section 9. Are the performance metrics for the next year proposed by NERSC reasonable?		
Recommendation	**NERSC Response**	**DOE Program Manager Response**
NERSC should use the availability metric targets used by OLCF and ALCF: 80% overall, 85% scheduled the first year, 90% overall, 95% schedule from the end of the first year on, or explain why their metric should be different.	NERSC understands the request, but it would mean that NERSC would have to lower its availability targets. NERSC Is following current DOE guidance and will continue to do so.	Disagree. PM approved. The three centers use the agreed-upon definitions and formulas to compute availability.
NERSC should add a survey response rate metric to their set of User Results metrics.	NERSC will report the response rate target that was negotiated with the DOE program manager. For the 2012 user survey, NERSC proposes to use 0.2% of total usage to define large MPP users. Medium MPP users would be those using between 0.02% and 0.2% of the resources, and small MPP users those using less than 0.02%. The target response rates proposed are 50% for large users, 25% for medium users, and 10% for small users. This proposal needs to be reviewed by NERSC's Program Manager.	

Table ES-2. Responses to Recommendations from the Previous 2011 OA Review (6 of 6)

Section 9. Are the performance metrics for the next year proposed by NERSC reasonable? (continued)		
Recommendation	**NERSC Response**	**DOE Program Manager Response**
NERSC capability metric targets are very low compared to the Leadership Computing Facilities. While NERSC is not an LCF and there is no expectation of a large amount of capability jobs running on the system, a capability job at less than 10% of the cores on the system is not really a capability job. Perhaps the percentage of cycles given to capability jobs could be reduced to 10% of the cycles run on the system and the size of the capability job be increased to be the same as OLCF and ALCF (20% of the cores on the system). Another option would be to remove the target and just have a reportable.	NERSC disagrees and follows guidance from the NERSC Program Manager. NERSC's role is different from the role of the Leadership Centers. NERSC's goal is to maximize scientific output, not machine or job size. Furthermore, NERSC does not have sufficient computing capability to allow DOE/SC to provide very large allocations to any one user. Since NERSC must meet the needs of all DOE SC users and their wide range of computing needs, it makes sense that NERSC's capability target is lower than the capability target at the leadership centers. NERSC analyzes and projects job usage data every year, and together with its Program Manager sets the appropriate capability target for the next year.	Disagree. NERSC was the only center with the PMM reporting responsibility since 2004 to OMB. As such, NERSC's PMM or capability metric has been scrutinized, and has been approved by ASCR AD and NERSC Program Manager as appropriate. The metric is reportable to OMB.
Section 10. Overall comments and/or recommendations		
Recommendation	**NERSC Response**	**DOE Program Manager Response**
NERSC should provide a copy of the previous year's peer reviews to the facilities performing this year's review. Without context and the ability to see the findings, comments, and recommendations, it is difficult to truly determine if the facility responded appropriately to the previous year's OA reviews.	NERSC will follow ASCR management direction.	Disagree. Follow the OAR Guidance from the ASCR Facilities Division.

1 User Support Results

Charge question 1: Are the processes for supporting the customers, resolving problems, and communicating with key stakeholders, and outreach effective?

> *NERSC response: Yes, NERSC has implemented effective processes for supporting customers, resolving problems, and communicating with key stakeholders. NERSC continues its tradition of making its users and their science requirements the drivers for NERSC system acquisitions, system management policies, and service activities. This includes both long-term strategic plans and day-to-day operations. To measure the effectiveness of these processes, the yearly user survey measures customer satisfaction, user problem reports are effectively managed through resolution, and NERSC uses feedback from its stakeholders to improve its services.*

The effectiveness of NERSC's processes for supporting customers, resolving problems, and communicating with key stakeholders is demonstrated by the metrics in Table 1-1.

NERSC developed its annual user survey in house in 1998. The 1998 survey contained 30 questions, developed by a combination of NERSC management and User Services consulting staff. Based on responses to this first survey NERSC refined its set of questions, and the format used from 1999 on has been quite stable. Individual questions might change from year to year, based on the hardware on the floor, but the format has remained consistent.

The survey is conducted on the web, and consists of about 100 satisfaction questions that are scored numerically. In addition, we solicit free-form feedback from the users. All authorized users (those with registered accounts who have signed the computer policy use form) are invited by email to participate. The survey tends to be open for about a month and users are reminded of the survey several times during that month. In recent years, and in order to increase the response rate, individual emails are sent to the larger users who have not yet responded.

The survey uses a seven-point rating scale, where "1" is "very dissatisfied" and "7" indicates "very satisfied." For each question the average score and standard deviation are computed. For questions that span previous surveys, the change in scoring is tested for significance using the t test at the 90% confidence level. In the survey results published on the NERSC website significant increases in satisfaction are in blue; significant decreases in satisfaction are shown in red.

In late 2007 NERSC contracted with Barry S. Sapolsky, Ph.D., President of Sapolsky Research, Inc. and Director of the Communication Research Center at Florida State University to review our survey process. His report from February 27, 2008, is titled Review of the NERSC Annual User Satisfaction Survey. Highlights of this review are:

- The survey questionnaire design is functional, attractive and on a par with most Web-based customer satisfaction surveys. The web-based survey approach is acceptable and should be continued.
- The seven-point satisfaction scale used in the NERSC User Scale is widely used. It is easily understood by respondents and provides a range of responses including "neutral." The scale should be continued in future NERSC User Surveys.
- The low response rate must be increased to insure valid and generalizable results. NERSC should focus its attention on higher usage individuals.
- The questionnaire is of reasonable length, based on the rough timings for completing the 2006 version. Suggested improvements to the questionnaire may reduce the time required to complete the survey.

Table 1-1. Summary of User Support Results

Metric Description	2010 Target	2010 Actual	2011 Target	2011 Actual	Projected July 1-Dec. 31, 2011	Actual July 1-Dec. 31, 2011
User Survey Metrics	**Key average scores on the user survey are satisfactory (> 5.25/7)**					
Overall Satisfaction with NERSC Average Score on User Survey	> 5.25	6.40	> 5.25	6.54	N/A	N/A
Average of All High-Level ("Overall Satisfaction") Scores on the User Survey	> 5.25	6.33	> 5.25	6.39	N/A	N/A
Average of All User Survey Scores	> 5.25	6.16	> 5.25	6.29	N/A	N/A
Average of All User Support (Services) Survey Scores	> 5.25	6.24	> 5.25	6.37	N/A	N/A
Consulting Overall Average Score on User Survey	> 5.25	6.64	> 5.25	6.58	N/A	N/A
Annual user survey results show improvement in at least one-half of the questions that scored below satisfactory (5.25/7) in the previous year	Two scores were < 5.25 in 2009. At least one of these must be ≥ 5.25 in 2010.	One score was > 5.25 in 2010. The other no longer applied since Bassi had been decommissioned.	Two scores were < 5.25 in 2010. At least one of these must be ≥ 5.25 in 2011.	The two scores that scored < 5.25 in 2010 were both > 5.25 in 2011.	N/A	N/A
Consulting Response Time Average Score on User Survey	> 5.25	6.60	> 5.25	6.57	N/A	N/A
Quality of Technical Advice Average Score on User Survey	> 5.25	6.57	> 5.25	6.53	N/A	N/A
Consulting Time to Solution Average Score on User Survey	> 5.25	6.40	> 5.25	6.40	N/A	N/A
Problem Resolution Metric						
Eighty percent of user problems are addressed within three working days (resolved or user informed how problem will be handled)	≥ 80%	84.50%	≥ 80%	86.9%	85%	86.1%

- NERSC provides a comprehensive summary of past survey results on its Website. The results are clearly presented and well organized. The presentation of comments from the open-ended text boxes are helpful, as is the section "Survey Results Lead to Changes at NERSC." Rarely do organizations provide evidence of follow-through based on feedback from users/customers. NERSC is to be commended for doing so.

A number of small suggestions were made in this analysis and all surveys since 2007 have incorporated these suggestions. We also followed the recommendation to increase participation rates among groups of higher usage users. We define MPP respondents as "large", "medium" or "small" based on their usage. We target a 50% response rate for large users, a 25% response rate for medium users and a 10% response rate for small users.

The 2011 user survey was conducted from June 6–30, 2011. The results are posted at https://www.nersc.gov/news-publications/publications-reports/user-surveys/2010-2011-user-survey-results/, where reviewers can see the statistical analysis performed for each question. Survey participation for 2010 and 2011 is summarized in Table 1-2.

Table 1-2. Survey Participation

	2010 Number of Users Queried	2010 Target Response Percent	2010 Actual Response Percent	2010 Number of Responses	2011 Number of Users Queried	2011 Target Response Percent	2011 Actual Response Percent	2011 Number of Responses
Sent to Large-Scale MPP Users	126	> 50%	77.8%	98	78	> 50%	76.9%	60
Sent to Medium-Scale MPP Users	479	> 25%	30.9%	148	344	> 25%	38.1%	131
Sent to Small-Scale MPP Users	948	> 10%	8.8%	83	1,064	> 10%	14.0%	149
Sent to All Authorized Users	3,533	N/A	11.2%	395	4,129	N/A	10.0%	411

1.1 User Support Metrics

NERSC's users are very satisfied with NERSC, as indicated by the 2011 results of the 13th annual survey of NERSC users. The high-level results, including the highest overall satisfaction score ever, are shown in Table 1-3.

Table 1-3. User Support Metrics

Metric Description	2010 Target	2010 Actual	2011 Target	2011 Actual	Projected July 1-Dec. 31, 2011	Actual July 1-Dec. 31, 2011
User Survey Metrics	**Key average scores on the user survey are satisfactory (> 5.25 / 7)**					
Overall Satisfaction with NERSC Average Score on User Survey	> 5.25	6.40	> 5.25	6.54	N/A	N/A
Average of All High-Level ("Overall Satisfaction") Scores on the User Survey	> 5.25	6.33	> 5.25	6.39	N/A	N/A
Average of All User Survey Scores	> 5.25	6.16	> 5.25	6.29	N/A	N/A
Average of All User Support (Services) Survey Scores	> 5.25	6.24	> 5.25	6.37	N/A	N/A
Annual user survey results show improvement in at least one-half of the questions that scored below satisfactory (5.25/7) in the previous year	Two scores were < 5.25 in 2009. At least one of these must be ≥ 5.25 in 2010.	One score was > 5.25 in 2010. The other no longer applied since Bassi had been decommissioned.	Two scores were < 5.25 in 2010. At least one of these must be ≥ 5.25 in 2011.	Both scores were > 5.25 in 2011.	N/A	N/A

1.1.1 Metric: Key Average Scores Are Satisfactory

In 2011, overall user satisfaction with NERSC improved significantly over 2010. As shown in Table 1-4, NERSC exceeded every target.

Table 1-4. NERSC Users Survey: Key Average Satisfaction Scores

All Users	2010 Target	2010 Actual	2011 Target	2011 Actual	2012 Minimum Satisfactory Target
Overall satisfaction with NERSC (one question)	5.25	6.40	5.25	6.54	5.25
Average of high-level ratings ("overall satisfaction" questions)	5.25	6.33	5.25	6.39	5.25
Average of all survey questions	5.25	6.16	5.25	6.29	5.25

Overall Satisfaction with NERSC Reaches New Heights

NERSC's users Overall Satisfaction score for 2011 was the highest ever recorded in the 13 years the survey has been in its current form. In 2011 NERSC increased its consulting staff and put a large, stable Hopper system into production.

Scientists indicated their satisfaction with NERSC on the user survey. As one user put it:

> *"NERSC keeps the machines up & running, sets up accounts quickly, proactively contacts us when allocations are running low. Web pages are very useful, help line & support is friendly & reliable. As long as jobs move quickly through the queues, everyone is happy!"*

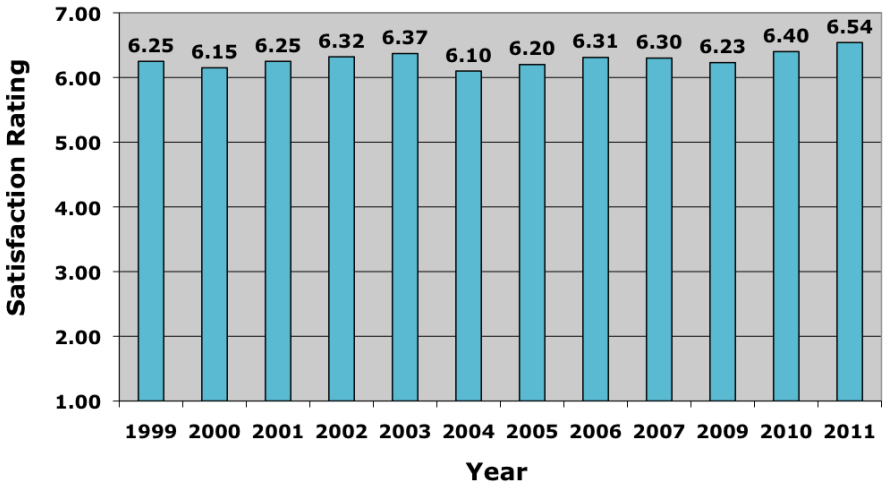


Figure 1-1. NERSC's overall satisfaction rating, while always extraordinary, was the highest ever in 2011. The scale is 7 (most satisfied) to 1 (least satisfied).

1.1.2 Metric: Key Average New User Scores Are Satisfactory

As shown in Table 1-5, NERSC has exceeded the target for key new user survey scores. NERSC users of one year or less are considered as new for the purposes of reporting on the survey. Note that for 2012 (with the goal to align operational assessment metrics across the DOE centers) new user survey scores are not part of the proposed metrics.

Table 1-5. NERSC Users Survey: Key Average Scores—New Users

New Users	2009	2010	2011	2012 Minimum Satisfactory Target
Overall satisfaction with NERSC (one question)	6.21	6.20	6.42	5.25
Average of high-level ratings ("overall satisfaction" questions)	6.07	6.19	6.32	5.25
Average of all survey questions	6.03	5.96	6.34	5.25

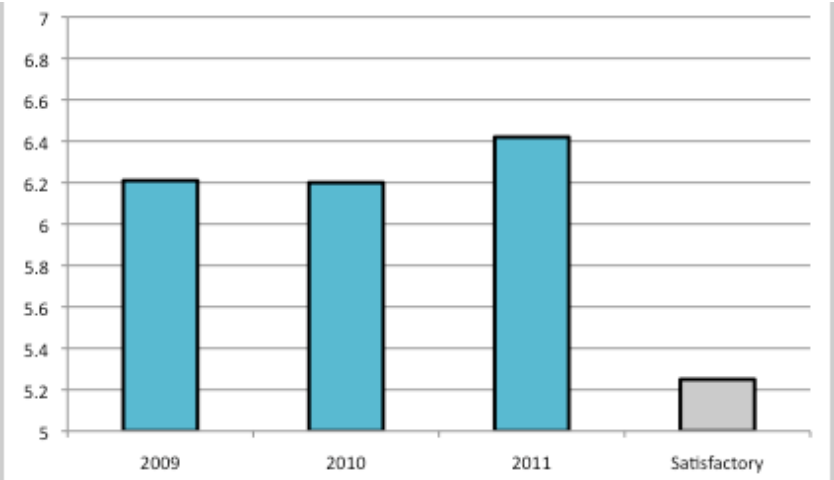


Figure 1-2. New users' overall satisfaction with NERSC.

1.1.3 Metric: Improvement on at Least Half of Below-Satisfactory Scores on the 2010 User Survey

Two survey questions scored below the minimum satisfactory score of 5.25 in 2010: Hopper: Batch Wait Time (5.19) and Franklin: Batch Wait Time (4.87).

In early 2011 NERSC opened user access to the Cray XE6 Hopper system, which replaced the much smaller Cray XT5 Hopper machine. The new petaflop system has proved very stable and productive, greatly improving job turnaround time on the Cray systems, Hopper and Franklin (Table 1-6).

Table 1-6. Improvements in Average Wait Time for Hopper and Franklin Batch Queues

System	May 2010 Average Wait Hours*	May 2011 Average Wait Hours*	Change from 2010
Hopper: Batch Wait Time	16.3	13.2	-2.1 Hours
Franklin: Batch Wait Time	14.0	9.3	-4.7 Hours

*Average wait in NERSC regular queues.

These improvements are reflected in the scores on the 2011 user survey (Table 1-7).

Table 1-7. Improvements in Below Satisfactory Scores from 2010

System	2010 Score	2011 Score	Change from 2010
Hopper: Batch Wait Time	5.19	5.86	+0.67
Franklin: Batch Wait Time	4.87	5.33	+0.46

1.1.4 Key Average Scores for Large, Medium, and Small Users

NERSC's largest users expressed the highest satisfaction with the center, according to the 2011 user survey (Table 1-8). The 60 largest respondents, who accounted for almost 60 percent of all hours used through June 2011, gave NERSC an average overall satisfaction rating of 6.80. Large users were defined as those users who were on pace (as of June 2011) to use more than 0.3% of 2011 total allocated hours, or about 3M hours each. Those projected to use 200K to 3M hours were classified as medium users, and those under 200K were tagged small users. The overall satisfaction score was a bit lower for the medium and small users, but still very high at 6.53 and 6.47, respectively. The scale is 7 (most satisfied) to 1 (least satisfied).

Table 1-8. NERSC User Survey: Key Average Scores for Large, Medium, and Small Users

	Large Users	Medium Users	Small Users	All Users with MPP Usage*
Overall satisfaction with NERSC (one question)	6.80	6.53	6.47	6.55
Average of high-level ratings ("overall satisfaction" questions)	6.55	6.42	6.19	6.40
Average of all survey questions	6.32	6.31	6.24	6.29
Number of respondents	60	131	149	340
Aggregate Percent of All Hours Used at NERSC (through 6/2011)	59.3%	11.1%	0.6%	71%

*Usage on Hopper, Franklin, or Carver.

1.1.5 Supplemental Information from the 2011 Survey

Highest and Most Improved Scores

The top three scores on the 2011 User Survey were Global Homes: Reliability (Data integrity), HPSS: Reliability (Data integrity), and HPSS Uptime (Table 1-9). These are areas where NERSC has excelled in the past.

Table 1-9. User Survey: Areas of Greatest Satisfaction

Category	2011 Score
Global Homes: Reliability (Data integrity)	6.79
HPSS: Reliability (Data integrity)	6.71
HPSS: Uptime	6.70

Most significant improvements from 2010 were related to the Hopper transition from a small Cray XT5 to a large Cray XE6 (No. 5 on the November 2010 TOP500 list) and NERSC training initiatives. The improvement in training scores was a direct result of NERSC's response to recommendations from the 2010 Operational Assessment.

The two lowest scores on the 2010 survey — Hopper and Franklin batch wait times — improved significantly in 2011, thanks to the new Hopper system (Table 1-10).

Table 1-10. User Survey: Areas of Most Significant Improvement

Category	2011 Score	Change from 2010
Hopper Batch Wait Time	5.86	+0.67
NERSC Training Classes	5.90	+0.51
Hopper: Ability to Run Interactively	6.11	+0.49
Franklin Batch Wait Time	5.33	+0.46

Users Least Satisfied with Batch Wait Times

Only one score on the 2011 survey was below the minimum satisfactory score of 5.25, Carver: Batch wait time. But even that lowest score, 5.16, was significantly higher than the lowest score from 2010, 4.86. The second lowest score in 2011 was Franklin: Batch wait time. These two scores stood alone at the bottom of the satisfaction ratings, 0.37 points below the next-lowest score (Table 1-11).

Table 1-11. User Survey: Areas of Least Satisfaction

Category	2011 Score
Carver: Batch wait time	5.16
Franklin: Batch wait time	5.33

Carver is a relatively small system (4,000 cores compared to 38,000 for Franklin and 150,000 on Hopper), but is in high demand due to its off-the-shelf Linux environment (ease of use) and relatively fast single-processor (core) performance (1.5 times that of Hopper). NERSC plans to increase the size of Carver by 60 percent, adding 240 nodes (1,920 cores) from the Magellan cloud research testbed when the Magellan project concludes in late 2011.

Franklin wait times, while still scoring near the bottom, nonetheless improved greatly (+0.46) from 2010 as many users shifted onto the large Hopper system.

NERSC Users Welcome Expanded Training Efforts

As a direct result of a 2010 Operational Assessment recommendation, NERSC renewed its training efforts in FY2011. In additional to its traditional training during the annual NERSC Users Group (NUG) Meeting, NERSC conducted a two-day workshop for Cray XE6 users at its facility in Oakland, joining with members of the Cielo team from Los Alamos National Laboratory and staff from Cray, Inc. Both the NUG training and the XE6 training were concurrently broadcast over the web. In addition, NERSC held a number of web-based training events (webinars) through 2010–2011. In all, NERSC put on eight events for its users from July 1, 2010 to June 30, 2011, with an aggregate attendance of about 375. In addition, NERSC staff contributed to lectures and

tutorials at UC Berkeley, at SC10, and at the Astrosim Summer School on Computational Astrophysics, Nicolaus Copernicus University, Torun, Poland.

NERSC's users responded positively to the training classes as indicated by the satisfaction score increase of 0.51 points on the 2011 User Survey. Additional surveys were conducted after each class, with 97.8% of respondents indicating that the training was "useful to me."

All the presentation materials from the classes are available on the NERSC website, and video recordings of most sessions are also online.

Statistically Significant Decrease Driven by Demand for Mid-Scale Cluster Computing

NERSC's Carver system, a fast Linux cluster with a high-speed network, is popular among its users. The only statistically significant decrease in satisfaction on the 2011 survey was for Carver batch queue wait times (Table 1-12).

Table 1-12. User Survey: Statistically Significant Decrease in Satisfaction

Category	2011 Score	Change from 2010
Carver: Batch wait time	5.16	-0.65

The decrease from 2010 caused the score to be below the satisfactory score level. In late 2011 NERSC increased the size of the Carver system from 3,200 to 8,032 cores with hardware from the Magellan project. This should help alleviate some of the congestion.

Data Analysis and Visualization Improvements in 2011

NERSC replaced its data analysis/visualization machine, DaVinci, with a new Sun Sunfire platform, Euclid, in June 2010. User satisfaction with the new system was evident in the user survey results, with Euclid scoring a 6.10/7 satisfaction score, an increase of 0.27 points over DaVinci's 2010 rating.

In addition to new hardware, NERSC has hired a new consultant to enhance NERSC's visualization and data analysis software and services. A significant accomplishment was the robust implementation of an NX server that enabled remote X-Windows based graphical software. NERSC aggressively publicized this new service to its users and held training sessions. As a result, users' satisfaction as measured by the 2011 user survey increased significantly on the following items (Table 1-13).

Table 1-13. Significant Data Analysis and Visualization Changes 2010-2011

Category	2011 Score	Change from 2010
Ability to Perform Data Analysis	6.13	+0.40
Data Analysis and Visualization Assistance	6.07	+0.40
Data Analysis Software	5.74	+0.27

1.2 Problem Resolution Metrics

NERSC tracks user requests for assistance using RightNow's trouble ticket system. Each ticket is managed by one consultant who is responsible for following the request from the initial contact through resolution. The NERSC consultants address a wide range of user needs, from installing software to optimizing and debugging their applications, porting codes, improving data transfer speeds, providing advice on algorithms, and helping solve science problems. In calendar year 2010, NERSC users submitted almost 5,000 question or problem tickets. For the one-year period from July 1, 2010 to July 1, 2011, NERSC's consultants addressed 5,969 requests for assistance (Figure 1-3).

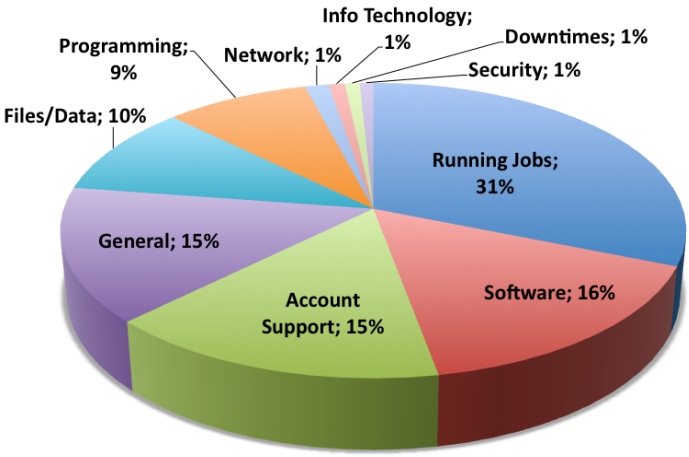


Figure 1-3. User requests for help by category, July 1, 2010 to July 1, 2011. Most user tickets concerned running jobs, software, account support, and general inquires that did not fall into a specific predefined category.

NERSC consultants strive to provide, fast, efficient, and high-quality responses to users. The consultants are a highly skilled group, with over half having PhDs in a science domain. Their science expertise is invaluable in bridging the gap between user questions and HPC hardware.

Table 1-14 shows that satisfaction with NERSC consulting is excellent. Across all questions — response time, quality of technical advice, and time to solution — the consulting scores are well above satisfactory. The high consulting scores are not only a reflection of a single group within NERSC, but are a measure of NERSC's responsiveness to user needs as a whole. Consultants collaborate with other NERSC staff members to solve many user problems.

Table 1-14. User Survey: Key HPC Consulting Average Scores

Metric Description	2010 Target	2010 Actual	2011 Target	2011 Actual	Projected July 1-Dec. 31, 2011	Actual July 1-Dec. 31, 2011
Consulting Overall Average Score on User Survey	> 5.25	6.64	> 5.25	6.58	N/A	N/A
Consulting Response Time Average Score on User Survey	> 5.25	6.60	> 5.25	6.57	N/A	N/A
Quality of Technical Advice Average Score on User Survey	> 5.25	6.57	> 5.25	6.53	N/A	N/A
Consulting Time to Solution Average Score on User Survey	> 5.25	6.40	> 5.25	6.40	N/A	N/A
Problem Resolution Metric						
Eighty percent of user problems are addressed within three working days (resolved or user informed how problem will be handled)	≥ 80%	84.5%	≥ 80%	86.9%	85%	86.1%

1.3 User Support and Outreach

With over 550 projects using more than 500 different codes, NERSC is constantly working to improve the productivity of all 4,000 scientists and engineers who use the center. NERSC closely collaborates with users to improve their productivity and the performance of their applications. Software installed or optimized for one user or group is installed for the benefit of the entire NERSC community of users. Tools that are found to be useful for one group become supported by NERSC for all users.

This section presents a few examples of our how NERSC's service-oriented efforts have benefitted its users.

1.3.1 Providing Rapid, Quality Front-Line User Support

The heart of NERSC's support services is its Consulting and Account Support staff. NERSC provides rapid responses and advice to hundreds of users each month. Relatively simple questions are answered immediately, while more complex and challenging problems develop into longer-term collaborations. In both cases — and everything in between — NERSC users benefit from NERSC technical staff's deep knowledge of high performance computing to overcome obstacles that otherwise might hold up research for months.

NERSC's consultants are on duty 8 a.m. to 5 p.m. Pacific Time during standard work days, immediately fielding phone calls, emails, and questions submitted through the web. During off hours, NERSC's Computer Operations and Networking staff is available 24x7 for password resets and system status inquiries.

The consultants work at the help desk for 4½-hour shifts on a rotating basis. Users appreciate instant access to HPC professionals because it allows them to overcome problems — sometimes simple ones — that could

otherwise immediately block their work. Typical issues involve network connection and authentication problems, errors in batch scripts, quota issues, and allocation issues.

In addition to the outstanding satisfaction scores documented in previous sections of this report, users offered many comments on NERSC's services in response to the question "What does NERSC do best? How does NERSC distinguish itself from other computing centers you have used?" Some of the comments:

> *"NERSC's people are very talented and are on par with the staff at the best supercomputing centers in the world. I currently run on 8 petaflop/s worth of supercomputers on two continents, and I count NERSC among the top 3 places I use."*
>
> *"NERSC consulting has been helpful with environment variables for my memory-intensive large runs."*
>
> *"Great user support, software support is exhaustive on mature systems"*
>
> *"NERSC has superb technical support staff. I'll never forget the experience of calling at 3am on a weekend for a password reset and getting a staff member right away who was much more awake than I was."*
>
> *"The consulting is good and they always follow through. With webinars the training has become easily available."*
>
> *"User Services and Viz/Analysis support are outstanding."*
>
> *"Web pages are very useful, help line & support is friendly & reliable."*
>
> *"Beyond simply furnishing computer time and resources, I have found that NERSC's consulting services are extremely helpful. Whenever I have run into a problem, the help desk has jumped in to investigate and almost always solved the issue within a day or two. For a large computing facility, this is a fabulous service. In addition, we recently had a couple of folks from NERSC come down to interview our group about our current and future computing needs to use in planning the next big cluster. This sort of user-centric planning is no doubt part of what makes NERSC great."*
>
> *"The staff is the most consistently helpful of any HPC systems I have ever used."*
>
> *"When I eventually start up my own research group, getting resources at NERSC will be a top priority as the environment is extremely conductive to scientific research."*

There were some negative comments and suggestions in response to the question "What can NERSC do to make you more productive?":

> *"One problem I have had is regarding visualization software for my field. The software has been installed, but there is not someone to help with troubleshooting."*
>
> *"The amount of email coming from NERSC is a bit excessive."*
>
> *"Often it is hard to get complex problems worked out with the consultants."*

"Have someone on call for software/environment problems on weekends (especially long weekends)."

"Some tickets are not resolved for a long time. If a solution is not obvious the attitude is to avoid solving the problem."

1.3.2 Compiler Comparisons

Compilers can be said to be the most common software used by scientists on HPC systems. Every application that runs on the NERSC systems is making a use of a compiler. Despite the obvious importance of compilers, users often don't get enough guidance on how to choose the best one for their application or what compiler flags to use. The Hopper system currently has four compilers installed: Portland Group, Cray, Pathscale, and GNU. (Additionally, due to large demand, NERSC has recently purchased the Intel compilers for the Hopper system as well.) This past year NERSC did a study comparing compiler performance and compiler flags for different benchmarks. The intention of the study is to educate users on the available compilers and encourage them to try a new compiler for their application.

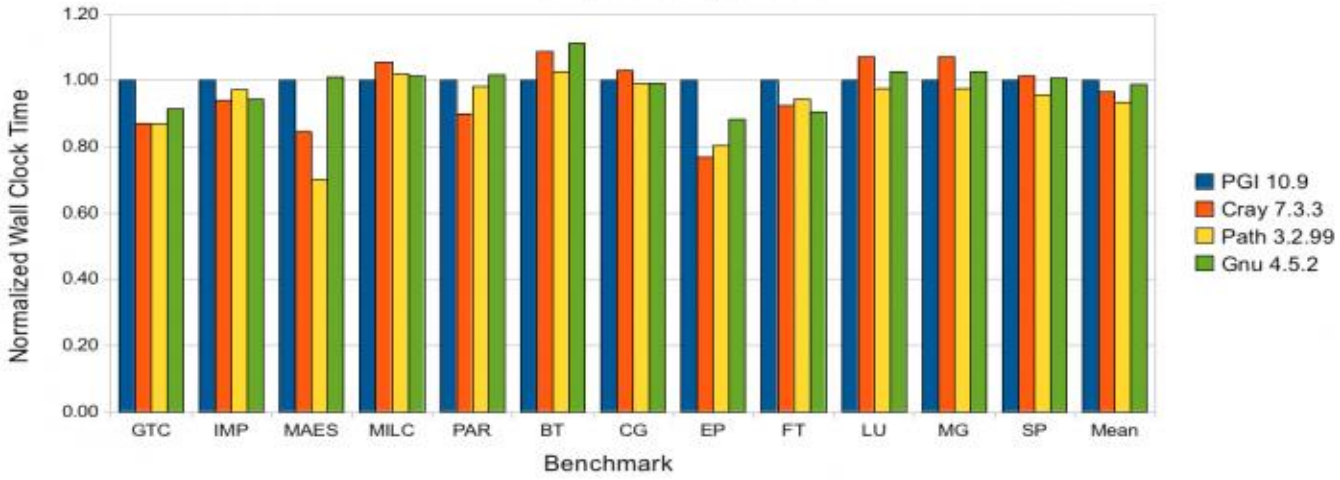


Figure 1-4. Compiler comparisons for different benchmarks.

Figure 1-4 shows the performance of micro-benchmarks and five full applications using the four compilers available on the Hopper system. The Portland Group compilers are the default on the Hopper system and are stable, reliable compilers that show solid performance for many applications. As the graph shows, however, a number of applications perform 20 and even 30 percent faster using alternative compilers. The MAESTRO code, for example, shows roughly an 18% performance improvement with the Cray compiler and over 30% improvement with the Pathscale compiler. The complete study is available to users on the NERSC website: http://www.nersc.gov/users/computational-systems/hopper/performance-and-optimization/compiler-comparisons/.

Changing compilers is often the simplest step a user can take to improve an application's performance. Additionally, the large number of NERSC users makes it impossible to work with each scientist one-on-one for code optimization, and so having accessible tutorials and performance tips on the website is key.

1.3.3 NERSC Users Learn How to Use Hopper's Multi-Core Nodes Effectively

A combined team of NERSC and Cray staff has formed a Cray Center of Excellence to examine programming models beyond pure MPI on the 24-core Hopper nodes. One of the key findings with hybrid MPI-OpenMP codes is that although OpenMP may not alter the performance of an application very much, using OpenMP can

dramatically decrease memory usage, allowing larger problems to be addressed. The core change in architecture of Hopper from that of earlier systems is the increasing number of cores per node and the decreasing amount of memory available per core. The majority of NERSC users run pure MPI applications. To ease the transition to systems with less memory available per task, NERSC has studied a number of benchmark codes and put together a series of tutorials, talks, and presentations to educate users on how to use the 24-core Hopper nodes effectively.

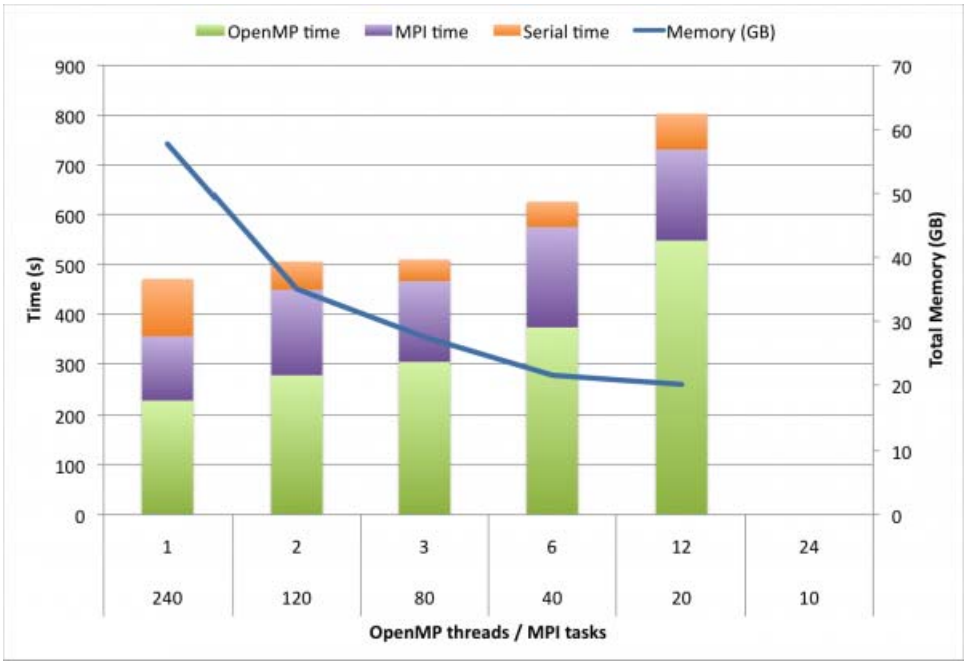


Figure 1-5. Performance of the fvCAM benchmark on Hopper as a function of MPI tasks and OpenMP threads — in other words, the performance of the code using a fixed number of nodes and differing amounts of MPI tasks and OpenMP threads. As one decreases the number of MPI tasks and increases the number of OpenMP threads, the overall memory usage decreases significantly. For example, using three threads and 80 MPI tasks, the performance decreases by only 6% compared to using single-threaded MPI tasks, whereas the memory requirement is reduced by almost 50%. Using six threads incurs a substantial performance penalty, in principle because of a growth in the time taken in OpenMP parallelized parts of the code. For 12 threads, there are also NUMA effects which add to the performance decrease.

As with the compiler comparison study, to reach a large number of users, NERSC creates online tutorials and presents optimization tips at workshops and conferences. The OpenMP study of CAM, GTC, Paratec, and PMAMR is available on the NERSC website: https://www.nersc.gov/users/computational-systems/hopper/performance-and-optimization/using-openmp-effectively-on-hopper/.

An example of a user application that has made use of the NERSC recommendations for running applications is Pourous Media Adaptive Mesh Refinement (PMAMR), an application being used to model carbon sequestration and contaminant transport as part of the Advanced Simulation Capability for Environmental Management (ASCEM) project at Berkeley Lab. The goal of the ASCEM project is to better understand and quantify flow and contaminant transport behavior in complex geological systems. The PMAMR code is built on the BoxLib framework and was previously parallelized using MPI. In this work, hybrid MPI/OpenMP programming was added to the code to see what kind of performance gains could be achieved.

Figure 1-6 shows the results of adding OpenMP. The hybrid approach yields a net speedup of 2.6× compared to the MPI-only version, with substantially reduced memory usage, thus decreasing the time to solution by increasing the rate at which the calculation runs.

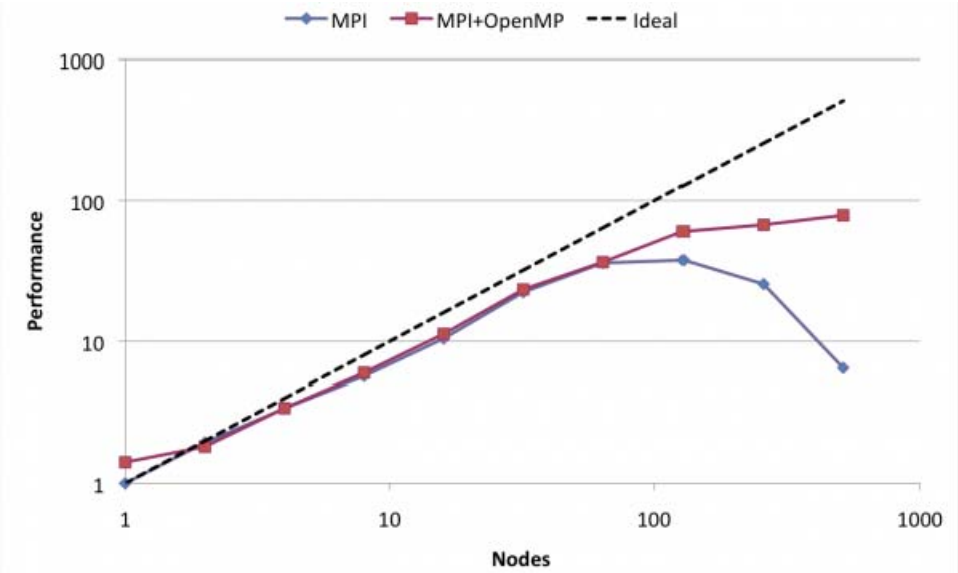


Figure 1-6. Results of adding OpenMP to the PMAMR code.

1.3.4 Facilitating Statistical Computing and Graphics Generation with Parallel R

R is an integrated suite of software facilities for statistical computing, data analysis, and graphics. NERSC has provided R training at User Group meetings, maintains an on-line R tutorial and links to extensive documentation, and provides in-depth consulting services.

For example, David Romps of the UC Berkeley Department of Geology and Geophysics has been working with the RNetCDF and ncdf libraries, which are required for creation, access, and analysis of array-oriented scientific data in the NetCDF file format. Daniela Ushizima of the Analytics Group worked with Romps to customize R software installations on various NERSC machines.

Romps says, "Without the Data Analytics Team's help, I would not be able to run my cloud-resolving large-eddy simulations on Hopper. I use R to write the NetCDF initialization files (initial value data, forcing terms, etc.) that are read in by my LES model, Das Atmospherische Modell (DAM). I also use R+NetCDF to analyze the resulting NetCDF output and to generate figures for manuscripts. With the large number of cores available to me on Hopper, I was able to run simulations on 12,288 processors at a resolution that I could not dream of on the other machines available to me. These simulations performed on Hopper will be part of a paper soon to be submitted to the Journal of Atmospheric Sciences, which is tentatively titled 'Convective momentum transport and its parameterization.'"

1.3.5 Visualizing Electronic Structures from Quantum Monte Carlo Simulations

Prabhat of the NERSC Analytics Group has supported Michael Philpott of UC Berkeley in visualization and analysis of quantum Monte Carlo simulations of electronic structures. Philpott was interested in studying the magnetic properties of graphene as it was selectively modified by adding a hydrogen atom. Depending on the placement of the atom in various geometric locations of the structure, radically different magnetic properties were revealed by computer simulations. Prabhat assisted Philpott in making sense of the simulation output by producing visuals that would immediately and effectively convey the resulting spatial patterns of spin and

bonding. The resulting work, titled "Magnetism and Bonding in Graphene Nanodots with H modified Interior, Edge and Apex," was recently accepted by the Journal of Chemical Physics.

1.3.6 Improving Access to X-Windows Based Tools and Applications

X-Windows is a display system that enables many popular tools to have a graphical user interface for ease of use and interactivity. However, X uses is a very chatty protocol that is very sensitive to network latency. That makes applications of every moderate complexity difficult or impossible to use outside a local area network (LAN). But virtually every NERSC user accesses the center over a wide area network (WAN), which previously severely restricted their access to debuggers, performance tools, and data analysis and visualization applications.

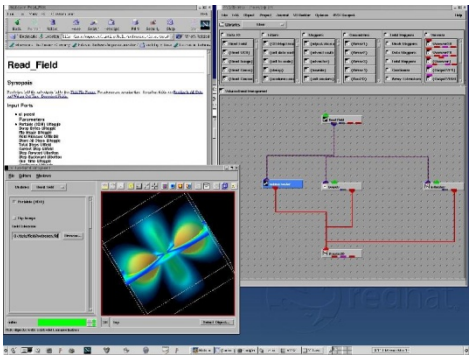


Figure 1-7. AVS/Express is an X-Windows based application that is enabled for remote users by the NX X-Windows accelerator.

In early 2011 NERSC put into production a service that enables excellent performance for X applications from virtually any geographic location. Very quickly after installation of a server and an enterprise version of a software package called NX, NERSC users were taking full advantage of the service. NERSC had been experimenting with NX for more than a year before putting it into full production.

This service opens an entirely new realm of visualization applications to all NERSC users (Figure 1-7). In addition, users can remotely employ GUI-heavy state-of-the-art debuggers like DDT and Totalview.

In the first weeks of service, more than 200 different NERSC users had connected to the NX server (Figure 1-8). At any given time, about 30 users had active sessions during June 2011.

Figure 1-8. NERSC users have connected to the NX server from around the world, enabling them to use X applications that were previously unusable.

User comments from the 2011 user survey included:

> "*The new NX is amazing, it is already more than I hoped to have.*"

"The new NX server is wonderful; it changed my way of working."

"The NX Client for visualization software is critical to make doing analysis at NERSC tractable."

1.3.7 Improving the Efficiency of a User Code

In addition to creating online tutorials and working with users one on one, NERSC consultants rapidly solve problems and collaborate with users to improve the efficiency of user codes. Collaborations with users can last a few hours or a few months, depending on the type of question the user has. One example from many is described.

A group of users from the University of Utah and Argonne National Laboratory were testing MPI2's RMA (Remote Memory Access), often referred to as MPI's one-sided communication. Their code remaps local arrays to a global array, partitioned among a selected group of MPI aggregator tasks, using MPI's RMA functions. The users reported that the code ran fine up to 2048 MPI tasks, but when 4096 tasks were used on Hopper, the code failed. A NERSC consultant extracted the communication information from more than 3000 lines of code to create a simple test case that reproduced the problem and verified that the users had not made an error. After examining the communication pattern, the consultant suspected that there may be insufficient internal MPI buffer space available. In particular, he noticed that many messages were being aggregated on PE 0. During the remapping of data to a global array, all the MPI messages were directed to PE 0 during the first 12 out of 16 aggregation steps. Based on this observation, the consultant suggested a workaround where an additional synchronization was introduced so that the data sent up to that point could be cleared from PE 0's buffers. This approach was tested successfully on 4096 cores. The consultant also filed a report with Cray to get the vendor's view on the problem and to determine if the issue should be filed as a bug in the implementation of the MPI library.

This example of a common interaction between a user and a NERSC consultant occurred over the period of a week. The user reported a problem running an application, and the consultant responded quickly with a workaround that allowed the user to continue running simulations. In the meantime, the consultant communicated with a vendor to analyze the problem and consider developing a longer-term solution.

1.3.8 Improving Account Support

NERSC has built a reputation as the computer center you can call on the phone and speak to someone who can help you right away. While continuing that tradition in 2010–2011, we looked for ways to increase what users can do to help themselves via web interfaces. NERSC users may now reset their expired or forgotten passwords themselves by answering a set of predetermined security questions. NERSC also made it easier for first-time users to get up and running with their accounts. The initial account passwords which had been distributed by phone are now handled through secure links in an email. Another step in making services easier for web-savvy users was to allow scanned or photographed versions of the computer use policy form to be uploaded via the web. Previously only faxes were accepted. Streamlined self-service web interfaces for NERSC users have added value to the center and reduced the number of emails, tickets, and staff hours required to help users be productive. In addition to being a convenience for NERSC users, these efficiency improvements help us provide quality service to a large numbers of users. If users find a self-service interface to account management useful, then NERSC staff have more time available to address problems for which human expertise is required.

1.3.9 Providing Comprehensive Support for an ALCC Project

The ASCR Leadership Computing Challenge (ALCC) project "A Multi-Decadal Reforecast Data Set to Improve Weather Forecasts for Renewable Energy" is an example of a project that uses a wide range of support services. Led by Thomas Hamill of NOAA, the project was awarded 14.5 million hours and to date has used over 8.5 million hours on Franklin and Hopper.

A reforecast is a retrospective forecast of the weather. Since numerical weather predictions are affected both by the chaotic growth of errors during the forecast and by model deficiencies, the use of direct model output for making weather-related decisions can lead to poor decisions. However, model forecast errors can be diagnosed from past forecasts and statistically corrected in post-processing, dramatically improving the forecast accuracy (Figure 1-9).

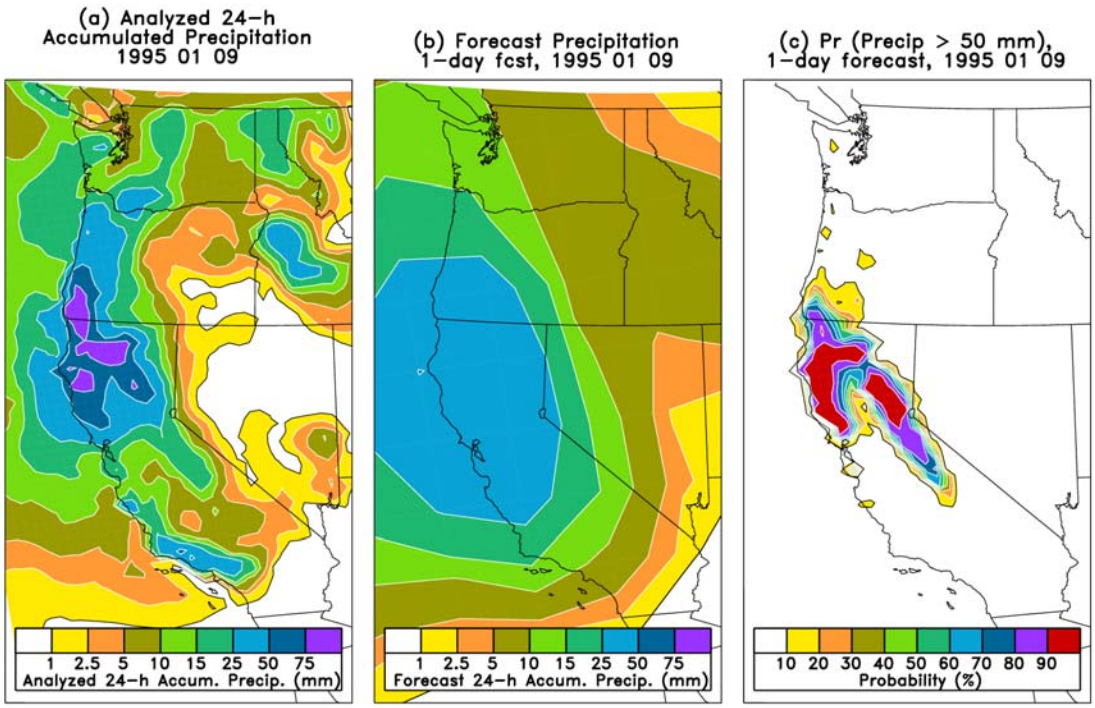


Figure 1-9. (a) 24-hour observed precipitation amounts for 9 January 1995; (b) average 1-day precipitation forecasts; (c) today's forecast calibrated with old reforecasts and precipitation analyses.

This project will generate a next-generation 30-year reforecast data set, using the same modeling system that will be operational for the National Weather Service next year. The reforecast data set will spur the development of novel longer-lead weather forecast applications for renewable energy (wind, solar, hydro), as well as improved probabilistic forecasts of rare extreme events.

NERSC provided significant support in optimizing transfer of the initial data from the National Climatic Data Center in Asheville, NC, to NERSC, and from NERSC to NOAA in Boulder. NERSC's HPSS now serves as the master storage site for the full dataset.

NERSC consultants rewrote part of the weather modeling code to enable porting from NOAA's IBMs to NERSC's Cray systems; provided tools to code developers to help them debug a memory problem; and improved throughput by assisting in redesigning the post-processing script and by providing a queue boost. After a slow start because of the code porting problem, the project is now on track to finish by December 2011.

1.3.10 Software Support

NERSC compiles and supports a large number of software packages for our users, totaling over 13.5 million lines of source code. These software packages include debugging tools, analysis and post-processing software, and third-party applications scientists use to run their primary simulations. In the past few years, NERSC has seen an

increase in the number of scientists using third-party applications. This is primarily due to the relative increase in the proportion of materials science and chemistry time allocated to NERSC projects and the popularity of third-party applications in these science areas.

Instead of each user installing his or her own version of a third-party software package, NERSC installs and supports optimized versions of many applications, allowing users to focus on their research rather than on installing and supporting software. Figure 1-10 shows the growth in the number of users running third-party applications. In 2010 there were over 495 unique users of third-party applications, and by June 30, 2011, there were already 492 scientists running third-party applications. This is before counting the typical surge of new graduate students and post-docs starting the new school year in September.

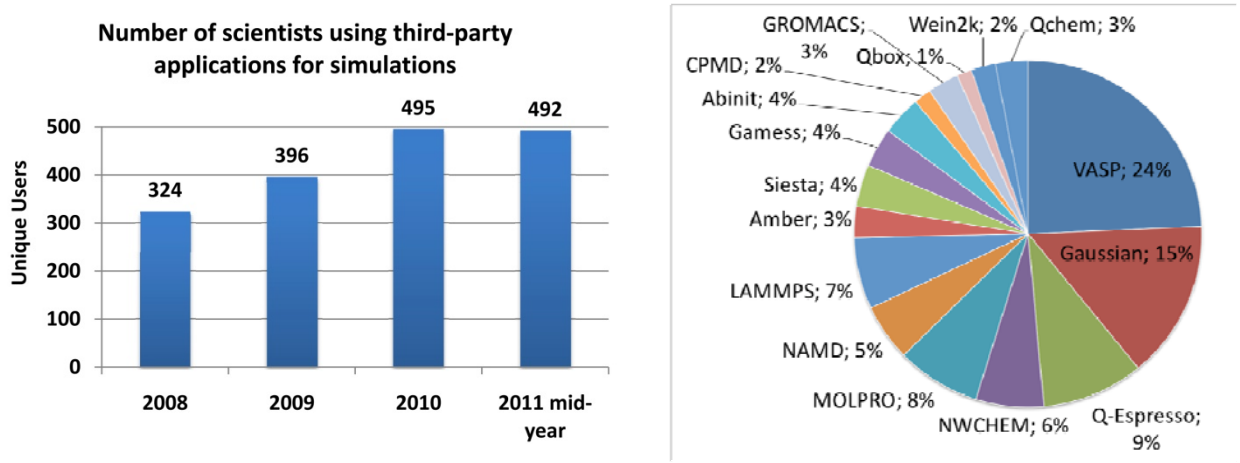


Figure 1-10. Third-party applications running at NERSC: number, growth since 2008, and types.

2 Business Results

Charge question 2: Is the facility maximizing the use of its HPC systems and other resources consistent with its mission?

> *NERSC Response: Yes, NERSC supports over 4,000 scientific researchers across a wide variety of scientific disciplines. The computational resources provided by NERSC are a key component to advancing their research. The metrics reported in this section show NERSC's commitment to providing computational resources for the advancement of science.*

NERSC's success in maximizing its HPC systems and other resources is summarized in Table 2-1.

Table 2-1. Business Results Metrics (1 of 3)

Scheduled Availability Metrics	Scheduled availability is ≥ 90% for systems in their first year of production, ≥ 93% in their second year, ≥ 95% thereafter.					
Metric Description	**2010 Target**	**2010 Actual**	**2011 Target**	**2011 Actual**	**Projected July 1–Dec. 31, 2011**	**Actual July 1–Dec. 31, 2011**
Franklin Scheduled Availability	≥ 93%	95.1%	≥ 95%	96.4%	95%	96.7%
Hopper-Phase-I Scheduled Availability	≥ 90%	98.2%	≥ 93%	99.6%	N/A	N/A
Hopper-Phase-II Scheduled Availability	N/A	N/A	≥ 90%	98.7%	97%	99.0%
Carver Scheduled Availability	≥ 90%	98.6%	≥ 93%	99.9%	98%	99.9%
Euclid Scheduled Availability	≥ 90%	97.9%	≥ 93%	99.9%	97%	99.9%
HPSS User System Scheduled Availability	≥ 95%	99.9%	≥ 95%	99.8%	98%	99.7%
Global Filesystem Scheduled Availability	≥ 95%	99.7%	≥ 95%	99.6%	98%	99.9%
Overall Availability Metrics	**Overall availability is ≥ 80% for systems in their first year of production, and 85% thereafter.**					
Metric Description	**2010 Target**	**2010 Actual**	**2011 Target**	**2011 Actual**	**Projected July 1–Dec. 31, 2011**	**Actual July 1–Dec. 31, 2011**
Franklin Overall Availability	N/A	93.0%	≥ 85%	95.5%	92%	96.1%
Hopper Phase I Overall Availability	N/A	96.6%	≥ 85%	99.6%	N/A	N/A
Hopper Phase II Overall Availability	N/A	N/A	≥ 80%	97.6%	95%	98.3%
Carver Overall Availability	N/A	96.4%	≥ 85%	98.8%	97%	98.7%
Euclid Overall Availability	N/A	95.7%	≥ 85%	99.4%	97%	99.4%
HPSS User System Overall Availability	N/A	99.4%	≥ 85%	98.9%	98%	98.8%
Global Filesystem Overall Availability	N/A	98.7%	≥ 85%	99.0%	97%	99.6%

Table 2-1. Business Results Metrics (2 of 3)

Mean Time To Interrupt (MTTI) Results	These are reported numbers, not metrics					
Description	2010 Target	2010 Actual (d:h:m)	2011 Target	2011 Actual (d:h:m)	Projected July 1–Dec. 31, 2011	Actual July 1–Dec. 31, 2011
Franklin MTTI	N/A	4:09:47	N/A	5:04:53	N/A	5:21:23
Hopper Phase I MTTI	N/A	13:10:28	N/A	22:10:00	N/A	N/A
Hopper Phase II MTTI	N/A	N/A	N/A	7:15:51	N/A	10:15:22
Carver MTTI	N/A	15:17:57	N/A	30:01:15	N/A	25:22:33
Euclid MTTI	N/A	10:04:45	N/A	10:23:54	N/A	10:18:14
HPSS User System MTTI	N/A	11:04:40	N/A	10:22:32	N/A	11:08:43
Global Filesystem MTTI	N/A	72:05:25	N/A	40:03:39	N/A	45:19:46

Mean Time To Failure (MTTF) Results	These are reported numbers, not metrics					
Description	2010 Target	2010 Actual (d:h:m)	2011 Target	2011 Actual (d:h:m)	Projected July 1–Dec. 31, 2011	Actual July 1–Dec. 31, 2011
Franklin MTTF	N/A	5:22:00	N/A	6:06:56	N/A	7:03:05
Hopper Phase I MTTF	N/A	21:12:24	N/A	22:10:00	N/A	N/A
Hopper Phase II MTTF	N/A	N/A	N/A	11:05:16	N/A	15:04:60
Carver MTTF	N/A	26:22:53	N/A	121:15:05	N/A	91:23:02
Euclid MTTF	N/A	17:08:24	N/A	17:08:33	N/A	22:23:40
HPSS User System MTTF	N/A	45:13:51	N/A	60:19:07	N/A	91:23:40
Global Filesystem MTTF	N/A	121:13:31	N/A	72:22:31	N/A	91:23:15

Table 2-1. Business Results Metrics (3 of 3)

Allocations and Usage Results	These are reported numbers, not metrics			
Allocation group	**2010 Allocated**	**2010 Hours Used**	**2011 Allocated**	**2011 Hours Used**
ASCR	11,973,000	20,971,000	47,450,000	48,479,000
BER	44,725,000	58,959,000	166,694,000	183,285,000
BES	75,175,000	127,118,000	272,127,000	380,409,000
FES	44,712,000	64,140,000	172,694,000	239,564,000
HEP	37,327,000	78,227,000	143,124,000	211,975,000
NP	26,088,000	58,588,000	95,911,000	178,717,000
ALCC	30,000,000	4,472,000	88,500,000	67,986,000
NERSC Director's Reserve	30,000,000	43,453,000	88,500,000	114,065,000
Totals	300,000,000	455,928,000	1,075,000,000	1,424,480,000

Large System Utilization	Percent of up time the system's computational nodes run batch jobs. These are reported numbers, not metrics					
	2010 Target	**2010 Actual**	**2011 Target**	**2011 Actual**	**Projected July 1–Dec. 31, 2011**	**Actual July 1–Dec. 31, 2011**
Franklin Utilization	N/A	96.4%	N/A	93.7%	N/A	92.3%
Hopper Utilization	N/A	N/A	N/A	87.2%	N/A	87.7%

Capability Metric	On NERSC's largest machine in production, at least *N* percent of the cycles should be used by jobs running on 1/8th or more of the processors, where *N* is determined annually by the NERSC Program Manager. This is an OMB Fiscal Year metric.					
	FY 2010 Target	**Calendar 2010 Actual**	**FY 2011 Target**	**FY 2011 Actual**	**4Q FY 2011 Projected**	**4Q FY 2011 Actual**
Cycles used by jobs running on 4,096 or more of Franklin's cores	≥ 30%	58.3%	≥ 35%	49.0%	40%	38.0%

NERSC provides the following computational resources for scientific research:

Table 2-2. NERSC Computational Systems as of July 2011

System Name	Type	CPU		Computational Pool					Node Interconnect
		Type	GHz	Nodes	SMP size	Total Cores	Aggregate Memory	Avg. Memory per core	
Hopper1	Cray XT5	AMD Opteron	2.4	664	8	5,312	10.6 TB	2.0 GB	SeaStar
Hopper	Cray XE6	AMD Magny Cours	2.1	6,384	24	153,216	211.8 TB	1.4 GB	Gemini
Franklin	Cray XT4	AMD Opteron	2.3	9,592	4	38,368	74.9 TB	2.0 GB	SeaStar
Carver	IBM iDataPlex	Intel Xeon	2.7	400	8	3,200	11.3 TB	3.6 GB	QDR InfiniBand
Euclid	Sun Sunfire	AMD Opteron	2.6	1	48	48	512 GB	10.7 GB	QDR InfiniBand

The metrics in this section are based on length of time the computational resource has been in production. The NERSC production computational systems entered into production according to the following list:

- Hopper: May 1, 2011
- Hopper1: March 1, 2010, decommissioned February 14, 2011
- Franklin: October 26, 2007, upgraded April 27, 2009
- Carver: May 2, 2010
- Euclid: May 17, 2010

The start date used for computing statistics will be either the initial production date or the date of the last upgrade to the computational resource. These are the dates used to determine if a system's metrics are based on the first year of service.

2.1 Resource Availability

From the 2011 Facility Metric Updates Guidance:
The Facility reports four availability statistics for each HPC computational and storage system:

- Scheduled Availability
- Overall Availability
- Mean Time to Interrupt (MTTI)
- Mean Time to Failure (MTTF)

Note that the MTTI and MTTF are not metrics but reported numbers.

Resource availability metrics are summarized in Table 2-3.

Table 2-3. Resource Availability Metrics (1 of 2)

Scheduled Availability Metrics	Scheduled availability is ≥ 90% for systems in their first year of production, ≥ 93% in their second year, ≥ 95% thereafter.					
Metric Description	**2010 Target**	**2010 Actual**	**2011 Target**	**2011 Actual**	**Projected July 1–Dec. 31, 2011**	**Actual July 1–Dec. 31, 2011**
Franklin Scheduled Availability	≥ 93%	95.1%	≥ 95%	96.4%	95%	96.7%
Hopper Phase I Scheduled Availability	≥ 90%	98.2%	≥ 93%	99.6%	N/A	N/A
Hopper Phase II Scheduled Availability	N/A	N/A	≥ 90%	98.7%	97%	99.0%
Carver Scheduled Availability	≥ 90%	98.6%	≥ 93%	99.9%	98%	99.9%
Euclid Scheduled Availability	≥ 90%	97.9%	≥ 93%	99.9%	97%	99.9%
HPSS User System Scheduled Availability	≥ 95%	99.9%	≥ 95%	99.8%	98%	99.7%
Global Filesystem Scheduled Availability	≥ 95%	99.7%	≥ 95%	99.6%	98%	99.9%

Overall Availability Metrics	Overall availability is ≥ 80% for systems in their first year of production, and 85% thereafter.					
Metric Description	**2010 Target**	**2010 Actual**	**2011 Target**	**2011 Actual**	**Projected July 1–Dec. 31, 2011**	**Actual July 1–Dec. 31, 2011**
Franklin Overall Availability	N/A	93.0%	≥ 85%	95.5%	92%	96.1%
Hopper Phase I Overall Availability	N/A	96.6%	≥ 85%	99.6%	N/A	N/A
Hopper Phase II Overall Availability	N/A	N/A	≥ 80%	97.6%	95%	98.3%
Carver Overall Availability	N/A	96.4%	≥ 85%	98.8%	97%	98.7%
Euclid Overall Availability	N/A	95.7%	≥ 85%	99.4%	97%	99.4%
HPSS User System Overall Availability	N/A	99.4%	≥ 85%	98.9%	98%	98.8%
Global Filesystem Overall Availability	N/A	98.7%	≥ 85%	99.0%	97%	99.6%

Table 2-3. Resource Availability Metrics (2 of 2)

Mean Time To Interrupt (MTTI) Results	These are reported numbers, not metrics					
Description	2010 Target	2010 Actual (d:h:m)	2011 Target	2011 Actual (d:h:m)	Projected July 1–Dec. 31, 2011	Actual July 1–Dec. 31, 2011
Franklin MTTI	N/A	4:09:47	N/A	5:04:53	N/A	5:21:23
Hopper Phase I MTTI	N/A	13:10:28	N/A	22:10:00	N/A	N/A
Hopper Phase II MTTI	N/A	N/A	N/A	7:15:51	N/A	10:15:22
Carver MTTI	N/A	15:17:57	N/A	30:01:15	N/A	25:22:33
Euclid MTTI	N/A	10:04:45	N/A	10:23:54	N/A	10:18:14
HPSS User System MTTI	N/A	11:04:40	N/A	10:22:32	N/A	11:08:43
Global Filesystem MTTI	N/A	72:05:25	N/A	40:03:39	N/A	45:19:46

Mean Time To Failure (MTTF) Results	These are reported numbers, not metrics					
Description	2010 Target	2010 Actual (d:h:m)	2011 Target	2011 Actual (d:h:m)	Projected July 1–Dec. 31, 2011	Actual July 1–Dec. 31, 2011
Franklin MTTF	N/A	5:22:00	N/A	6:06:56	N/A	7:03:05
Hopper Phase I MTTF	N/A	21:12:24	N/A	22:10:00	N/A	N/A
Hopper Phase II MTTF	N/A	N/A	N/A	11:05:16	N/A	15:04:60
Carver MTTF	N/A	26:22:53	N/A	121:15:05	N/A	91:23:02
Euclid MTTF	N/A	17:08:24	N/A	17:08:33	N/A	22:23:40
HPSS User System MTTF	N/A	45:13:51	N/A	60:19:07	N/A	91:23:40
Global Filesystem MTTF	N/A	121:13:31	N/A	72:22:31	N/A	91:23:15

2.1.1 Scheduled Availability

Scheduled Availability is the percentage of time a designated level of resource is available to users, excluding scheduled downtime for maintenance and upgrades. To be considered a scheduled outage, the user community must be notified of the need for a maintenance event window no less than 24 hours in advance of the outage (emergency fixes). Users will be notified of regularly scheduled maintenance in advance, on a schedule that provides sufficient notification, and no less than 72 hours prior to the event, and preferably as much as seven calendar days prior. If that regularly scheduled maintenance is not needed, users will be informed of the cancellation of that maintenance event in a timely manner. Any interruption of service that does not meet the minimum notification window is categorized as an unscheduled outage.

A downtime or outage is any event (hardware, software, human, environmental) that disrupts full service to the users. It is computed from the time when user computational processing halts to the restoration of computation (see formula below).

$$SA = \left(\frac{time\ in\ period - time\ unavailable\ due\ to\ outages\ in\ period}{time\ in\ period - time\ unavailable\ due\ to\ scheduled\ outages\ in\ period}\right) \times 100$$

As shown in Table 2-4 below, NERSC has exceeded scheduled availability targets for 2010 and 2011.

Table 2-4. Scheduled Availability Metrics

Scheduled Availability Metrics	Scheduled availability is ≥ 90% for systems in their first year of production, ≥ 93% in their second year, ≥ 95% thereafter.					
Metric Description	2010 Target	2010 Actual	2011 Target	2011 Actual	Projected July 1–Dec. 31, 2011	Actual July 1–Dec. 31, 2011
Franklin Scheduled Availability	93%	95.1%	95%	96.4%	95%	96.7%
Hopper-Phase-I Scheduled Availability	90%	98.2%	93%	99.6%	N/A	N/A
Hopper-Phase-II Scheduled Availability	N/A	N/A	90%	98.7%	97%	99.0%
Carver Scheduled Availability	90%	98.6%	93%	99.9%	98%	99.9%
Euclid Scheduled Availability	90%	97.9%	93%	99.9%	97%	99.9%
HPSS User System Scheduled Availability	95%	99.9%	95%	99.8%	98%	99.7%
Global Filesystem Scheduled Availability	95%	99.7%	95%	99.6%	98%	99.9%

2.1.2 Overall Availability

In the past NERSC reported and monitored Overall Availability. For the first time in 2011, NERSC has targeted metrics to meet for Overall Availability (note that these were not in the 2011 metrics included in the 2010 OA report because this was not known at the time). NERSC's goal is to perform continuous improvements to provide maximum overall availability.

Overall Availability is the percentage of time a designated level of resource is available to users.

$$OA = \left(\frac{time\ in\ period - time\ unavailable\ due\ to\ outages\ in\ period}{time\ in\ period}\right) \times 100$$

Table 2-5 shows the Overall Availability metrics for each of the production systems for the calendar years 2010 and 2011.

Table 2-5. Overall Availability Metrics

Overall Availability Metrics	**Overall availability is ≥ 80% for systems in their first year of production, and 85% thereafter.**					
Metric Description	**2010 Target**	**2010 Actual**	**2011 Target**	**2011 Actual**	**Projected July 1–Dec. 31, 2011**	**Actual July 1–Dec. 31, 2011**
Franklin Overall Availability	N/A	93.0%	85%	95.5%	92%	96.1%
Hopper Phase I Overall Availability	N/A	96.6%	85%	99.6%	N/A	N/A
Hopper Phase II Overall Availability	N/A	N/A	80%	97.6%	95%	98.3%
Carver Overall Availability	N/A	96.4%	85%	98.8%	97%	98.7%
Euclid Overall Availability	N/A	95.7%	85%	99.4%	97%	99.4%
HPSS User System Overall Availability	N/A	99.4%	85%	98.9%	98%	98.8%
Global Filesystem Overall Availability	N/A	98.7%	85%	99.0%	97%	99.6%

2.1.3 System Mean Time To Interrupt (MTTI)

Mean Time To Interrupt (MTTI) is the time, on average, to any outage on the system, whether unscheduled or scheduled. Also known as MTBI (Mean Time Between Interrupt).

The MTTI for a system is calculated with the following formula:

$$MTTI = \left(\frac{time\ in\ period - (duration\ of\ scheduled\ outages\ + duration\ of\ unscheduled\ outages)}{number\ of\ scheduled\ outages\ + number\ of\ unscheduled\ outages\ + 1}\right)$$

The reporting of MTTI is based on actual system events. There are no targets or projections defined by the Guidance.

Table 2-6 shows a two year comparison of MTTI for NERSC systems in 2010 and 2011.

Table 2-6. Mean Time To Interrupt (MTTI) Results
(These are reported numbers, not metrics)

Description	2010 Target	2010 Actual (d:h:m)	2011 Target	2011 Actual (d:h:m)	Projected July 1-Dec. 31, 2011	Actual July 1-Dec. 31, 2011
Franklin MTTI	N/A	4:09:47	N/A	5:04:53	N/A	5:21:23
Hopper Phase I MTTI	N/A	13:10:28	N/A	22:10:00	N/A	N/A
Hopper Phase II MTTI	N/A	N/A	N/A	7:15:51	N/A	10:15:22
Carver MTTI	N/A	15:17:57	N/A	30:01:15	N/A	25:22:33
Euclid MTTI	N/A	10:04:45	N/A	10:23:54	N/A	10:18:14
HPSS User System MTTI	N/A	11:04:40	N/A	10:22:32	N/A	11:08:43
Global Filesystem MTTI	N/A	72:05:25	N/A	40:03:39	N/A	45:19:46

This table shows a decrease in MTTI for NGF in 2011 as compared to 2010. Several factors come into play that affect the MTTI. Due to its increased size, in 2011 NERSC implemented a routine quarterly maintenance schedule for NGF, which results in NGF being scheduled offline more frequently than in the past. Additionally, a series of upgrades to NGF required the file systems be scheduled down.

2.1.4 System Mean Time To Failure (MTTF)

Mean Time To Failure (MTTF) is the time, on average, to an unscheduled outage on the system. The MTTF is calculated by the following formula:

$$MTTF = \left(\frac{time\ in\ period - duration\ of\ unscheduled\ outages}{number\ of\ unscheduled\ outages\ + 1}\right)$$

The reporting of MTTF is based on actual system events. There are no targets or projections defined by the Guidance.

Table 2-7 shows a two year comparison of MTTF for NERSC systems in 2010 and 2011.

Table 2-7. Mean Time To Failure (MTTF) Results
(These are reported numbers, not metrics)

Description	2010 Target	2010 Actual (d:h:m)	2011 Target	2011 Actual (d:h:m)	Projected July 1-Dec. 31, 2011	Actual July 1-Dec 31, 2011
Franklin MTTF	N/A	5:22:00	N/A	6:06:56	N/A	7:03:05
Hopper Phase I MTTF	N/A	21:12:24	N/A	22:10:00	N/A	N/A
Hopper Phase II MTTF	N/A	N/A	N/A	11:05:16	N/A	15:04:60
Carver MTTF	N/A	26:22:53	N/A	121:15:05	N/A	91:23:02
Euclid MTTF	N/A	17:08:24	N/A	17:08:33	N/A	22:23:40
HPSS User System MTTF	N/A	45:13:51	N/A	60:19:07	N/A	91:23:40
Global Filesystem MTTF	N/A	121:13:31	N/A	72:22:31	N/A	91:23:15

This table shows a decrease in MTTF for NGF in 2011 as compared to 2010. NGF upgrades in 2011 involved installation of new hardware on which all new data is stored; and some old data has also been migrated to the new hardware. Most unscheduled NGF outages were caused by faulty hardware-level microcode, but there have been some actual hardware malfunctions as well. We continue to work with the vendor to improve their hardware, and it has stabilized in the last two months.

The table also shows that Carver is serving as a consistently stable platform.

2.2 Resource Utilization

Allocations of NERSC computer time are made on a yearly basis. The NERSC allocation year runs from the second Tuesday in January through the Monday before the second Tuesday of the following year. Computer time is characterized as "DOE time" or "NERSC overhead time."

Eighty percent of the DOE time is allocated by program managers in the six DOE Offices of Science. Applicants need to be part of a research project funded by the DOE Office of Science or show that their work meets the DOE mission.

Ten percent of DOE time is allocated through the ASCR Leadership Computing Challenge (ALCC) program described at http://science.energy.gov/ascr/facilities/alcc/.

The final 10 percent of DOE time is the NERSC Director's reserve, allocated through the NERSC Initiative for Scientific Exploration (NISE) program described at https://www.nersc.gov/users/accounts/allocations/nise/. NISE applications are open to all NERSC users. They are reviewed by the NERSC staff, and awards are made to the projects judged to have the highest likelihood of meeting the goals of the program.

NERSC overhead time is used to support the activities of the NERSC staff, its vendors, and a small number of guests. Startup and educational awards come from overhead time. Finally, overhead time is used to make refunds when user jobs encounter system errors.

Table 2-8 shows how DOE time was allocated and used in 2010 and 2011. Usage for 2011 includes free pre-production time on Hopper in January through April 2011.

Table 2-8. Allocations and Usage Results
(These are reported numbers, not metrics)

Allocation group	2010 Allocated	2010 Hours Used	2011 Allocated	2011 Hours Used
ASCR	11,973,000	20,971,000	47,450,000	48,479,000
BER	44,725,000	58,959,000	166,694,000	183,285,000
BES	75,175,000	127,118,000	272,127,000	380,409,000
FES	44,712,000	64,140,000	172,694,000	239,564,000
HEP	37,327,000	78,227,000	143,124,000	211,975,000
NP	26,088,000	58,588,000	95,911,000	178,717,000
ALCC	30,000,000	4,472,000	88,500,000	67,986,000
NERSC Director's Reserve	30,000,000	43,453,000	88,500,000	114,065,000
Totals	300,000,000	455,928,000	1,075,000,000	1,424,480,000

2.2.1 Total System Utilization

Total System Utilization is the percent of time that the system's computational nodes run user jobs. No adjustment is made to exclude any user, group, staff, or vendor. Total System Utilization is calculated with the following formula:

$$TSU = \left(\frac{core\ hours\ used\ in\ period}{core\ hours\ available\ in\ period}\right)$$

Table 2-9 shows a two year comparison for Total System Utilization for NERSC's MPP systems in 2010 and 2011.

Table 2-9. Large System Utilization

	Percent of up time the system's computational nodes run batch jobs. These are reported numbers, not metrics.					
	2010 Target	2010 Actual	2011 Target	2011 Actual	Projected July 1-Dec. 31, 2011	Actual July 1-Dec. 31, 2011
Franklin Utilization	N/A	96.4%	N/A	93.7%	N/A	92.3%
Hopper Utilization	N/A	N/A	N/A	87.2%	N/A	87.7%

2.2.2 HPSS Resources

NERSC collects HPSS Storage Resource Units (SRUs) in its NERSC Information Management (NIM) system. HPSS SRUs are computed based on the number of files stored, space used, and amount transferred to and from the mass storage system.

Table 2-10. HPSS Resources Allocated and Delivered to DOE Allocated Scientific Users 2010-2011

Allocation Year	AY 2010		AY 2011	
	Allocated	Used	Allocated	Used
HPSS Storage Resource Units (SRUs)	80 million	60.2 million (9.9 PB)	135 million	109.5 million (16.4 PB)

As outlined in Table 2-10, users tend to use fewer SRUs than allocated. This is explained by NERSC's approach to storage management, which enables users to store increasing amounts of data at decreasing cost. The largest contributing factor to the SRU is the amount transferred to and from the mass storage system.

Whereas NERSC provisions sufficient bandwidth to handle the largest transfers, actual bandwidth consumption is variable — usually saturated preceding scratch file system purging as users are moving mass amounts of data to the mass storage system. The number of SRUs allocated (especially to provision bandwidth) is therefore always larger than the number of SRUs used.

2.2.3 Global File System Resources

The global file system resources at NERSC are the Global Home, Global Scratch, and Project file systems. The key benefit to presenting global storage assets is to minimize data movement between various storage resources

internal to the center. Global home is the resource used on login to all NERSC computational systems and provides each user with 40 GB of space. It aids in presenting a common user environment across distinct systems. Global Scratch is used as a large quota (20 TB per user) high bandwidth file system generally for job output. Project provides a 4 TB default quota and is used to share data among multiple people or between systems at NERSC.

Figure 2-1 shows the Global Home growth rate from January 1, 2010 through December 31, 2010. The monthly growth rate was around 7%. NERSC addressed this high growth rate by using programmatic funds to increase the capacity of the file system in April 2010. Figure 2-2 shows the growth characteristics of the Global Home file system for January 1, 2011 through July 31, 2011 (updated statistics are not available). In 2010 the Global Home file system grew at around 7% per month.

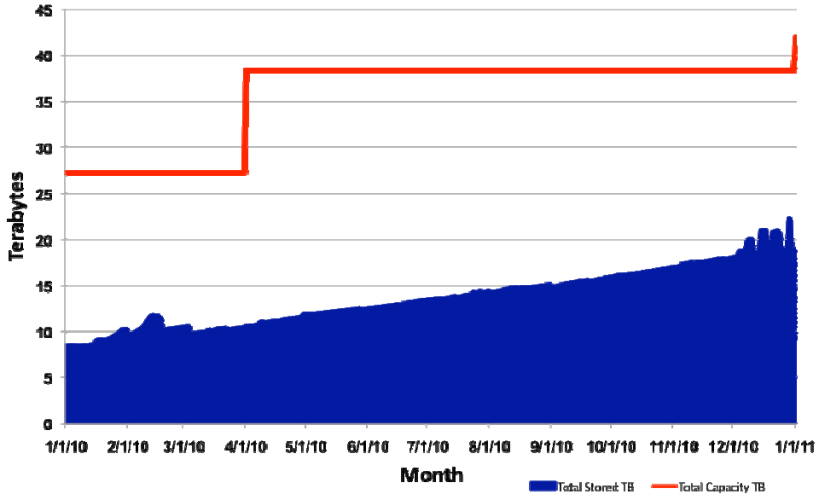


Figure 2-1. Global Home capacity utilization, January–December 2010

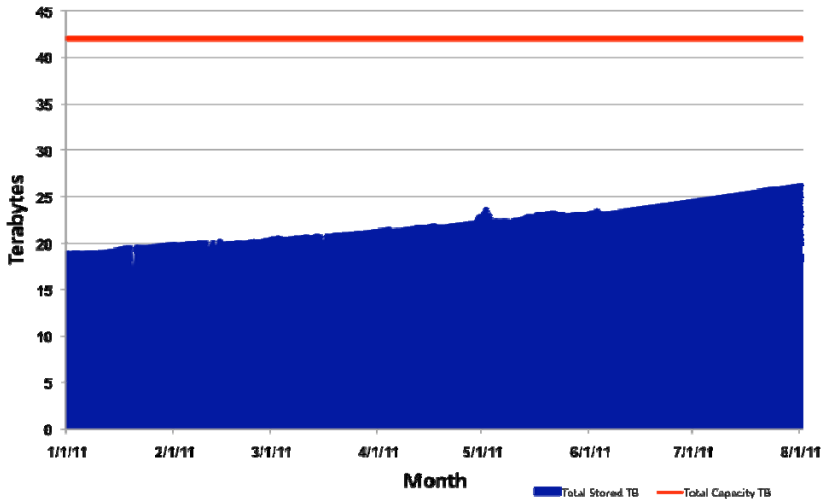


Figure 2-2. Global Home capacity utilization, January–August 2011.

Given the average monthly growth of Global Home of 7% in 2010, Figure 2-3 shows projected growth through the end of 2011.

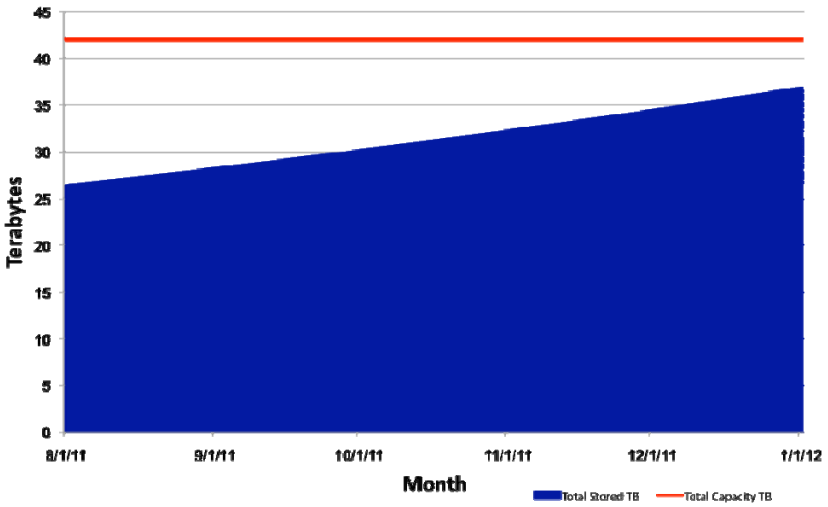


Figure 2-3. Global Home projected capacity utilization, August–December 2011.

Global Scratch is a high-bandwidth, centralized, large file system shared across a number of the computational systems. It provides a default of 20 TB scratch storage per user and serves as the only storage for the Carver system. The Global Scratch resource is too new to provide meaningful growth or utilization statistics.

The Project file system is a shared resource mounted on every major computational system at NERSC. It is allocated on a request-only basis and enables a group of users to share data in their project directory both among themselves and between systems. It has been a key to enabling the pace of scientific discovery for numerous projects since its inception in 2006 by allowing users to produce output on one machine and analyze it on another, or switch between different systems easily using the same input data sets.

Figure 2-4 shows capacity utilization statistics for the Project file system from January 1, 2010 through December 31, 2010. This file system grows at 8% per month on average.

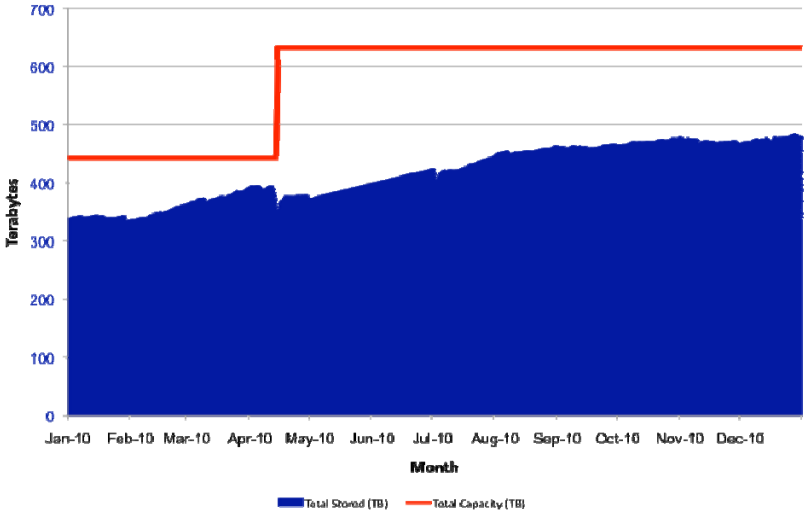


Figure 2-4. Project file system capacity utilization, January–December 2010.

Due to recent requests for space increases in the Project file system, NERSC significantly increased capacity in April 2011 (see Figure 2-5 below). The final increase to 1.6 PB in July 2011 represents the total increase after

completing the new configuration. This will benefit users who have recently requested large (30–50 TB) allocations per project in the file system.

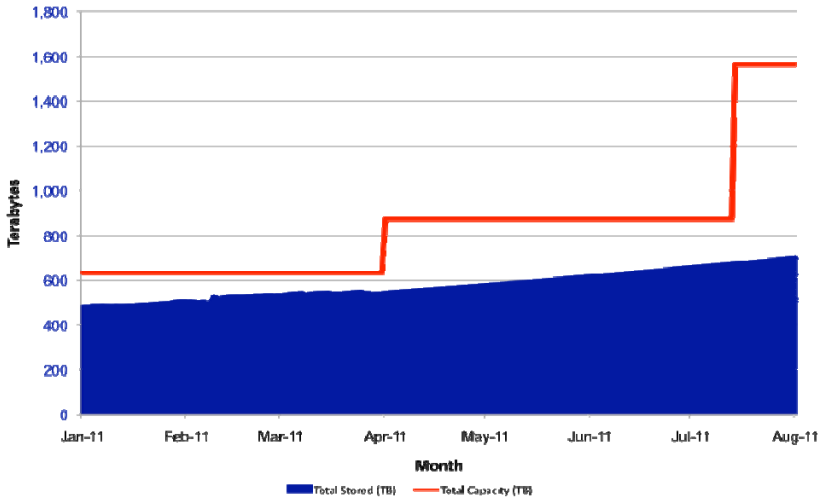


Figure 2-5. Capacity utilization for Project file system, January–August 2011.

Figure 2-6 below shows the projected capacity utilization for the Project file system for the remainder of 2011 (August 1, 2011 through December 31, 2011). Utilization increase is based on the 2010 monthly growth rate of 8%.

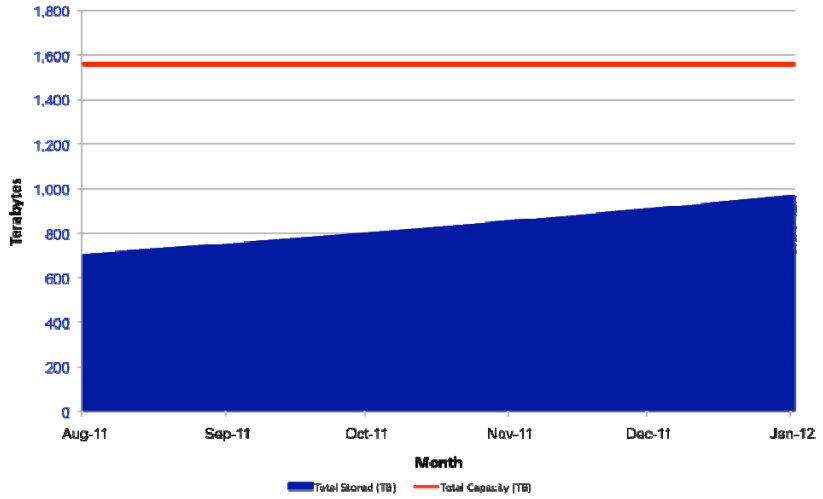


Figure 2-6. Projected capacity utilization for Project file system, August–December 2011.

2.2.4 Network Resources

Data transfers consume more network bandwidth than any other function within the NERSC network, especially transfers to and from the HPSS archival storage system. Moving to shared or global file systems across NERSC's computational systems has improved available network capacity for HPSS transfers and helps prevent congestion. This improves the user experience in minimizing or avoiding data movement between file systems and improves the available network bandwidth from a given computational system to the HPSS system.

The 2010 network traffic (Figure 2-7) shows the constant influx of data to NERSC, where the amount of data coming into NERSC exceeds the amount of data leaving NERSC. An exception is the spike in outgoing traffic in November 2010, when large datasets were exported to support demonstrations at the SC'10 conference.

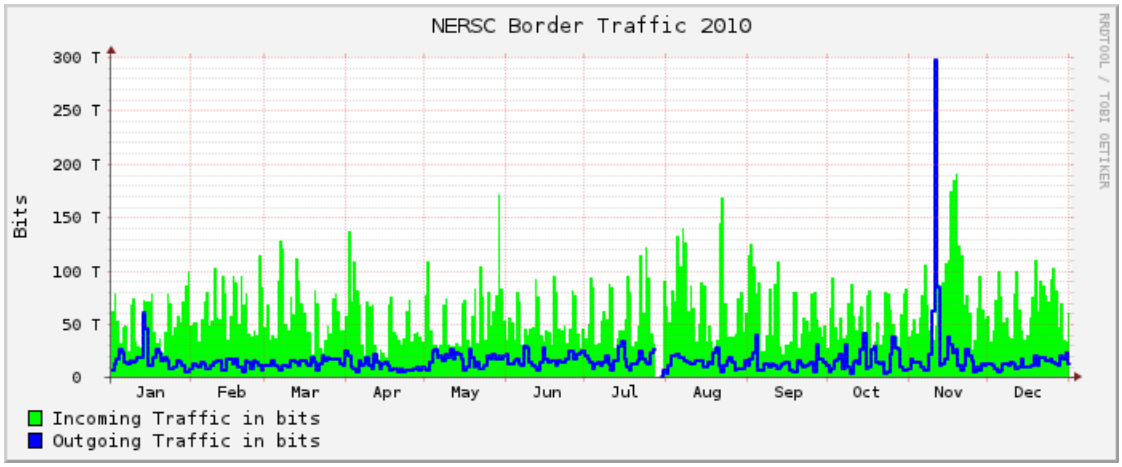


Figure 2-7. NERSC network traffic, 2010.

In 2011, NERSC continued the trend of being a net data importer, meaning that more data came into NERSC than went out to other destinations. Over the last year, NERSC's average bandwidth has seen an almost 300% increase in data transferred into the system, as seen in Figure 2-8.

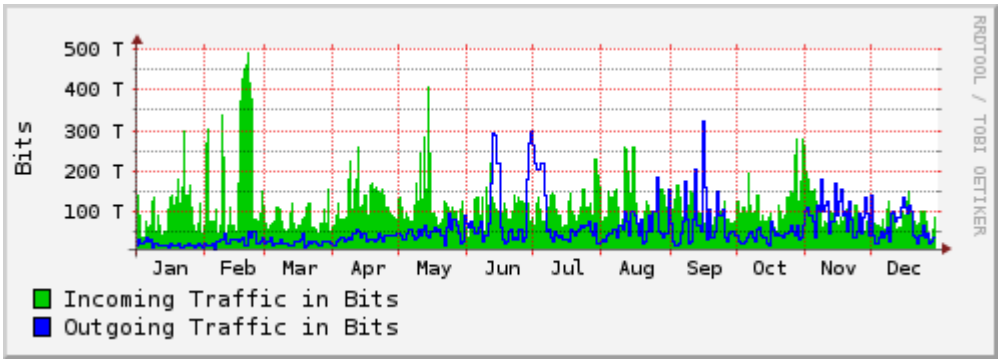


Figure 2-8. NERSC network traffic, 2011.

Utilization justified a significant network capacity increase this year. Major parts of the network were upgraded to provide improved bandwidth and accessibility to various subnets. In looking at Figure 2-8, we can see that NERSC sustained several periods of where we had greater than 200 terabits of traffic in a day, and one day with a maximum of 500 terabits. It is worth noting that in 2010, NERSC's maximum was only 200 terabits.

NERSC also utilizes a separate link to ESnet to provide data-rich large science projects a dedicated path into NERSC. Utilization on this link, the Science Data Network, has been steadily growing (Figure 2-9) and now consumes almost as much network bandwidth as NERSC's production link of just a year ago. Much of the outgoing traffic in January and February 2011 was going to file systems located at the Joint Genome Institute (JGI) in Walnut Creek, CA; in March these file systems were moved to NERSC.

NERSC continues to partner with ESnet in providing quality network capabilities to users outside of NERSC as well. On several occasions, this required significant collaboration with ESnet and other facility network administrators to tune and otherwise improve network performance from the user's home site and resource to NERSC resources (see section 6.1.4, Data Transfer Nodes Facilitate Scientific Collaborations).

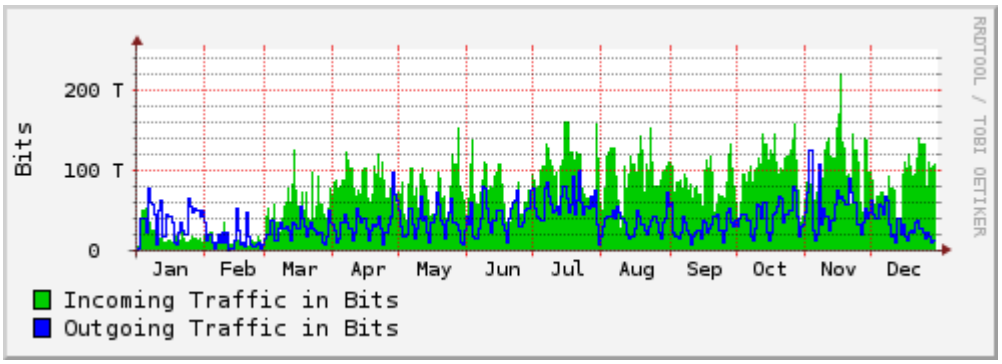


Figure 2-9. NERSC/ESnet Science Data Network traffic, 2011.

2.3 Capability Utilization

Batch jobs exceeding specific core counts for the individual systems are classified as capability jobs. An agreement is made with the DOE Program Manager to set these core counts, and progress against this agreement is reported here. The metric is defined as 1/8th the computational core count., rounded down to the nearest power of 2 (4,096 cores on Franklin for FY 2011 and 16,384 cores on Hopper for FY 2012). Since this is an OMB 300 metric, it is set for fiscal rather than calendar years.

Table 2-11. Capability Utilization Metrics

Capability Metric	On NERSC's largest machine in production, at least *N* percent of the cycles should be used by jobs running on 1/8th or more of the processors, where *N* is determined annually by the NERSC Program Manager. This is an OMB Fiscal Year metric.					
	FY 2010 Target	Calendar 2010 Actual	FY 2011 Target	FY 2011 Actual	4Q FY 2011 Projected	4Q FY 2011 Actual
Cycles used by jobs running on 4,096 or more of Franklin's cores	≥ 30%	58.3%	≥ 35%	49.0%	40%	38.0%

The reason the fourth quarter of FY 2011 is significantly lower than the previous results is that large jobs started moving from Franklin to Hopper after Hopper went into production in May 2011.

Figure 2-10 shows the rate of capability usage in 2011 on Franklin based on seven-day averages. Figure 2-11 shows job usage at different core counts for Franklin in 2011 based on seven-day averages.

Figure 2-12 shows the rate of capability usage in 2011 on Hopper based on seven-day averages. It is interesting to note that there was a significant amount of capability use through Hopper's free period (which ended April 30, 2011), and that capability use dropped drastically once charging started. To address this, in June NERSC started boosting and discounting (by 25%) jobs at the 16K core level. In July NERSC introduced a Hopper scaling incentive program (including profiling and scaling advice and reimbursements for capability jobs). In October NERSC increased the discount rate for jobs at the 16K core level to 40%.

Figure 2-13 shows job usage at different core counts for Hopper in 2011 based on seven-day averages.

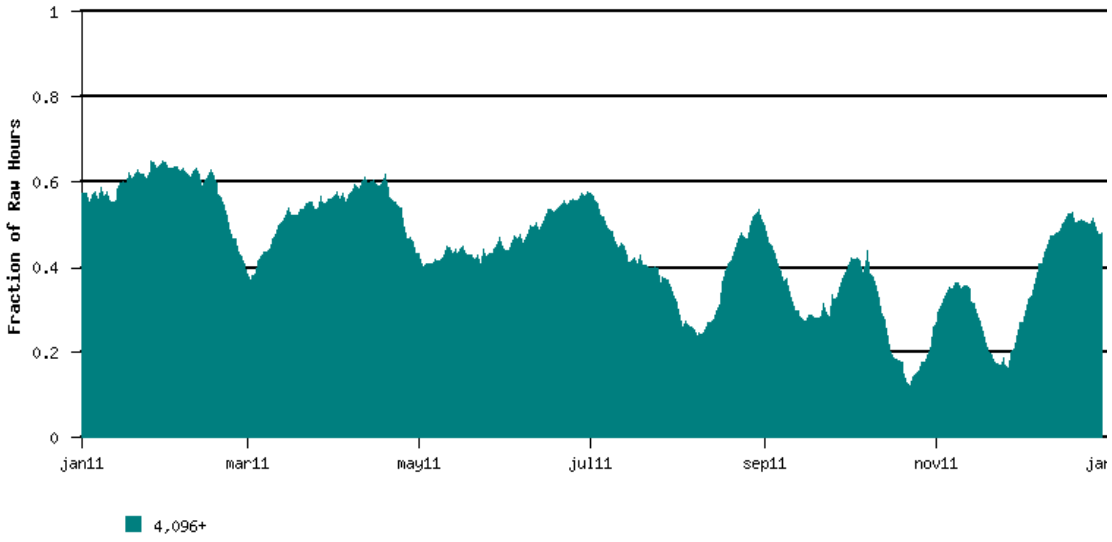


Figure 2-10. Franklin capability utilization for 2011 based on seven-day averages.

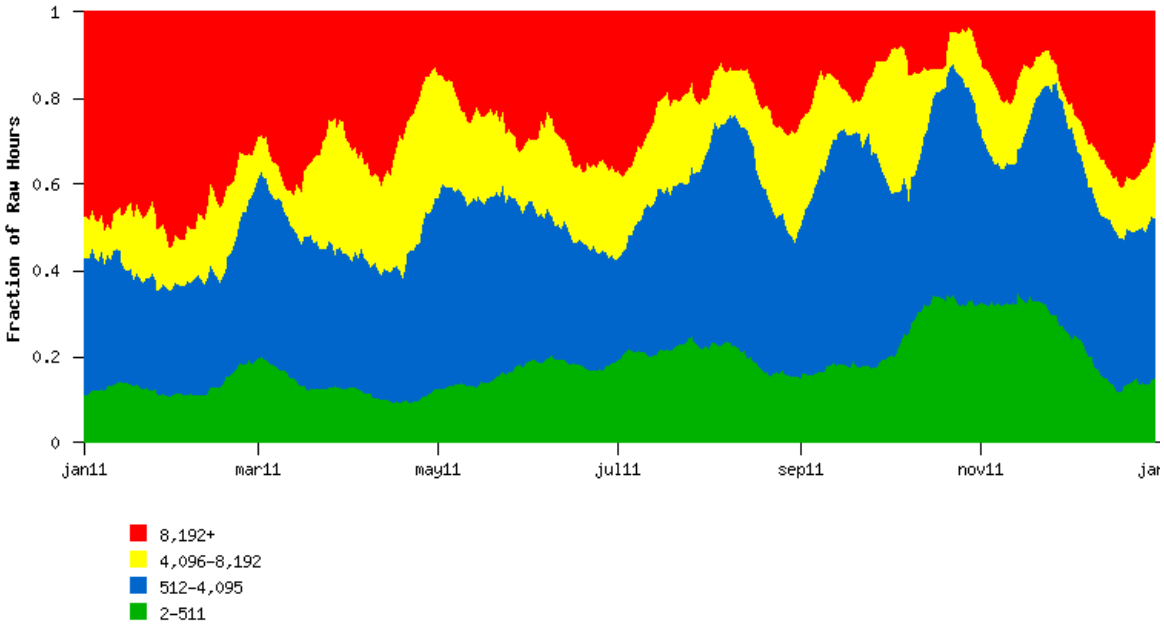


Figure 2-11. Franklin job usage at different core counts for 2011 based on seven-day averages.

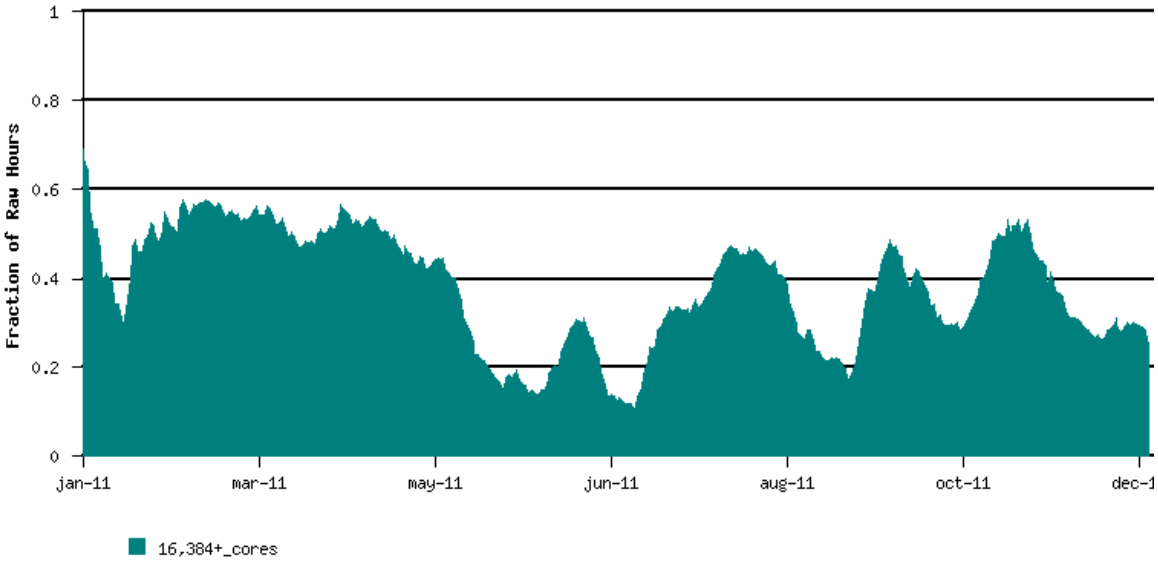


Figure 2-12. Capability usage on Hopper in 2011 based on seven-day averages.

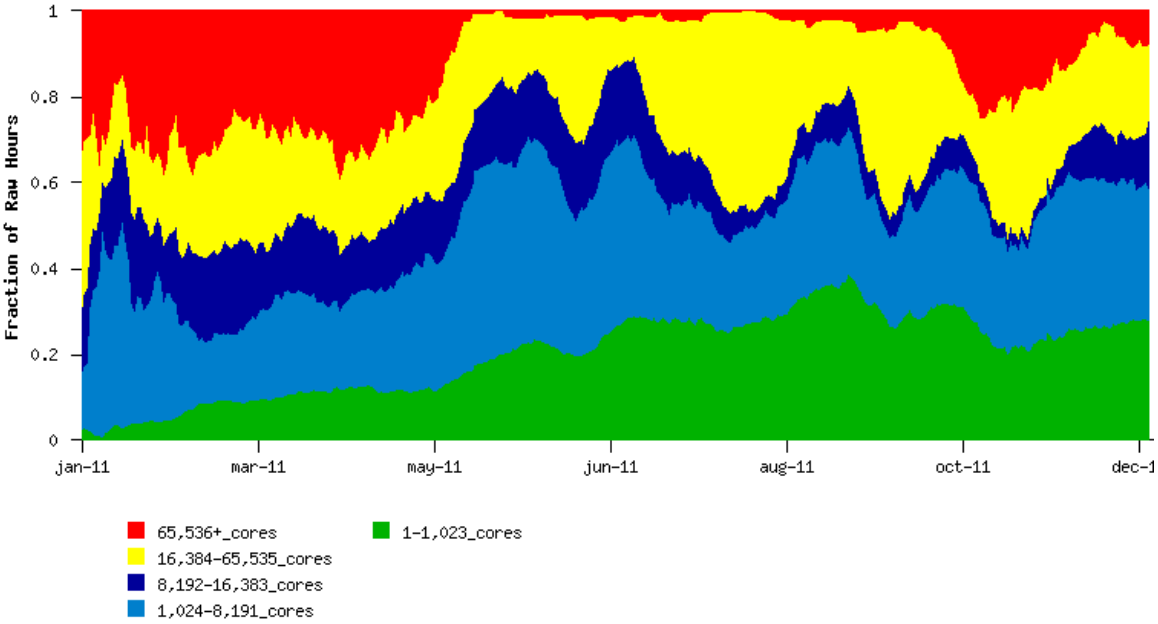


Figure 2-13. Hopper job usage at different core counts for 2011 based on seven-day averages.

3 Strategic Results

Charge question 3: Is NERSC enabling scientific achievements consistent with the Department of Energy strategic goals 3.1 and/or 3.2?

NERSC Response: Yes, NERSC is meeting DOE Strategic goals 3.1 and 3.2, and we will substantiate our assessments in terms of the following topics per the Operational Assessment Guidance:

- *Science Output*
- *Scientific Accomplishments*
- *Allocation of Facility Director's Reserve computer time*

3.1 Science Output

NERSC tracks and reports the number of refereed publications written annually based on using (at least in part) NERSC resources. In previous Operational Assessments, NERSC reported the number of refereed publications published or submitted. Starting with the 2011 Operational Assessment NERSC reports the number of refereed publications published or in press (Table 3-1).

Table 3-1. Number of Refereed Publications Based on NERSC Resources

Science Output	2010 Refereed Publications*	2011 Refereed Publications*
Number of refereed publications (published or in press) based on using (at least in part) NERSC resources	1,407	1,528

*As reported in allocation renewals

NERSC researchers reported 16 cover stories in 2010 and eight in 2011. They are displayed on the web at http://www.nersc.gov/news-publications/journal-cover-stories/SortList/Date/DESC.

In 2011, NERSC spread the word about our science accomplishments through:

- The 2010 NERSC Annual Report
- Monthly NERSC Newsletter: 24 news articles and 50 shorter news items in 12 issues
- Quarterly Science Highlight collections
- Social Media, including the NERSC Facebook Fan Page, LBNL CS Facebook Group and NERSC Twitter (retweeted by CS, ESnet and LBNL)
- Making News: 64 news articles (Google News Alert) and 483 mentions (Google Blog Search)

Science articles can be found on the NERSC web site: https://www.nersc.gov/news-publications/science-news/

3.2 Scientific Accomplishments

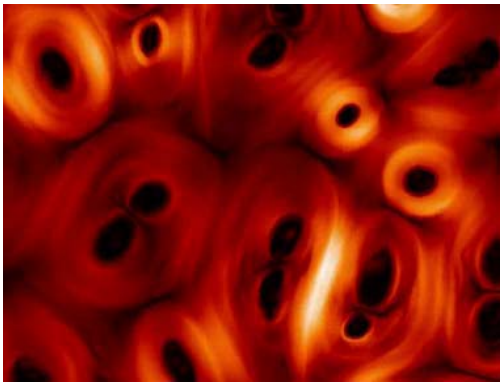
With more than 4,000 users, 550-plus projects, and over 1,500 peer-reviewed published papers reported in 2011, NERSC users' scientific accomplishments could not be fully covered in this report. Instead, we present and briefly describe a representative sample below.

3.2.1 Solar System's Magnetic Edge More Turbulent Than Expected

Principal Investigator: James Drake (University of Maryland)

NERSC Project: Turbulence, Transport and Magnetic Reconnection in High Temperature Plasma (FES)

Objective: Use detailed simulation of plasma turbulence and magnetic reconnection to understand conditions at the solar system's outer boundary.



A frame from a visualization showing how magnetic bubbles about 100 million miles across form at the edge of our solar system. Scientists believe that this work may change our view of how the sun interacts with particles, fields, and gases from other stars.

Implications: The heliosheath, the boundary region between our solar system and the rest of the Milky Way Galaxy, shields the solar system from some of the harmful effects of intergalactic cosmic rays. The heliosheath represents the furthest reach of the highly charged particles emitted by the sun, and the enormous magnetic field they create deflects most of the dangerous cosmic rays. All current models of this region of space assume that this magnetic field is laminar (smooth); but if it is not, as Voyager spacecraft data suggest, this shield may not be as effective a barrier as had been thought. A better understanding of what cosmic rays are and how they get into the solar system may be required.

Accomplishments: Three-dimensional magnetohydrodynamics (MHD) simulations done at NERSC with very high numerical resolution have determined that the edge of the solar system is a turbulent place. Far from being a smooth sheet, it is filled with a roiling sea of huge magnetic bubbles that act as particle traps. This finding shakes up prevailing views of the solar system's outer reaches and shows that this region, once thought to be a relatively firm shield against interstellar particles such as galactic cosmic rays, is actually more of a porous membrane. The importance of this result is also the demonstration that rapid magnetic field reconnection — responsible for phenomena that drive solar flares and aurora, among other astrophysical events — is ubiquitous in space. This is the first glimpse of it in the heliosheath.

NERSC Resources:
2010: 2.1 M hours used by repo mp217; 9,000 jobs on Bassi, Carver, Franklin, and Hopper; up to 11,972 cores.
2011 through July: 3.1 M hours used; 4,200 jobs on Carver, Franklin, and Hopper; up to 49,152 cores.
NERSC 2011 NISE award provided part of the time.

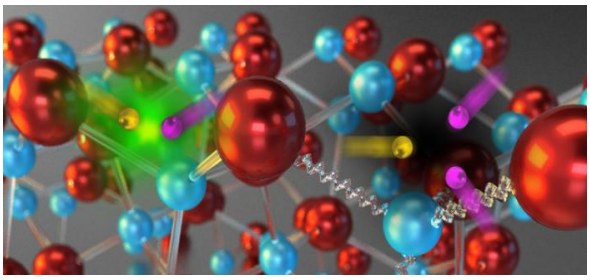
Full Story: http://www.nersc.gov/news-publications/science-news/2011/magnetic-bubbles/

3.2.2 LED Lighting Breakthrough from Materials Science

Principal Investigators: E. Kioupakis, C. Van de Walle (UC Santa Barbara)

Project: First-Principles Modeling of Nitride and Oxide Materials for Applications in Electronic and Optoelectronic Devices (BES)

Objective: Obtain an atomistic understanding of the dramatic efficiency loss (droop) in Light-Emitting Diodes (LEDs).



Direct electron-hole recombination, indicated with the green glow on the left, almost always results in photon emission; whereas in indirect recombination with free-carrier absorption (right), little photon emission occurs. Simulations at NERSC suggest this mechanism as the dominant source of LED droop.

Implications: LEDs hold enormous potential as energy-efficient, non-toxic, long-lasting replacements for incandescent bulbs and compact florescent lights. However, LED lights are still expensive because of the dramatic drop in efficiency they exhibit at the high intensities needed for full room illumination. The origin of this so-called "LRD droop" is still an issue of active research and hot debate. Finding its origin could help make lower-cost, more efficient lighting more widely available.

Accomplishments: Using a new quantum mechanical approach, researchers found that LED droop can be attributed to Auger recombination, a process that occurs in semiconductors in which three charge carriers interact without giving off light. Most importantly, they also discovered that indirect Auger effects, which are assisted by carrier scattering due to lattice vibrations and the random distribution of the alloy atoms, form the dominant contribution to the Auger recombination rate in nitride materials. Now that Auger recombination has been established as the root of the problem, the scientists say they can focus on creative approaches to suppressing or circumventing the loss mechanism. LED droop can't be eliminated because Auger effects are intrinsic, but it could be minimized by using thicker quantum wells in LEDs or by growing devices along nonpolar or semipolar directions to keep carrier density low.

NERSC Resources:
2010: 1.3 M hours used by repo m934; 36,000 jobs on Franklin, Carver, and Hopper; up to 9,412 cores.
2011 through July: 4.5 M hours used; 25,750 jobs on Franklin, Carver, and Hopper; up to 16,392 cores.
NERSC 2010 and 2011 NISE awards provided part of the time.

Full Story: http://www.nersc.gov/news-publications/science-news/2011/led-lighting-comes-out-of-the-dark/

3.2.3 Greenhouse Gas Remediation

Carbon dioxide (CO_2) gas is considered to be present in only trace proportions in our atmosphere but it has a leading role in the cast of greenhouse gases, with a thermal radiative effect nearly three times as large as the next biggest contributor. Energy related processes are the biggest sources of atmospheric CO_2, especially the burning of fossil fuels and the production of hydrogen from methane. Since both human-caused CO_2 concentrations and global average temperatures have been increasing steadily since the mid-20th century, it could very well be that our energy future depends on our ability to effectively remove CO_2 from the atmosphere. But there are significant technological challenges.

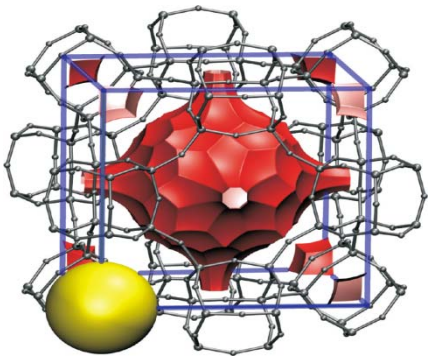
Several NERSC projects associated with Energy Frontier Research Centers (EFRCs) are seeking to surmount those challenges by simulating some of the complicated processes related to carbon capture and underground sequestration. The goal is to remove greenhouse gases from the atmosphere and prevent them from reentering by storing stable geologic formations. Two example projects are described below.

1. Speeding Up Materials Evaluation for Carbon Capture and Sequestration

Principal Investigators: B. Smit, M. Haranczyk, J. Sethian (LBNL and UC Berkeley)

Project: Computational Characterization of Porous Materials for the Center for Gas Separations Relevant to Clean Energy Technologies EFRC (BES)

Objective: The vision of this EFRC is development of new methods to create, understand, and predict novel materials with optimal molecular properties for carbon capture and sequestration. Separation of CO_2 from other gases is a key technological bottleneck in carbon capture. Highly porous materials are attractive choices for gas separation and are already widely used in the chemical industry; zeolites, complicated microporous aluminosilicate minerals, are an example. The number of possible zeolite structures has been estimated to be more than 2.5 million, although only about 190 structures have been synthesized to date. Complete characterization of an entire database of hypothetical structures would be out of the question even with today's supercomputers.



Schematic of a candidate porous material showing a suitable gas capture area (center) and unsuitable areas (corners).

The specific goal, then, is to develop alternative approaches for screening large databases of these complex structures so that only those predicted to exhibit properties of interest are subjected to follow-on characterizations that use more accurate but also more expensive molecular simulation methodology.

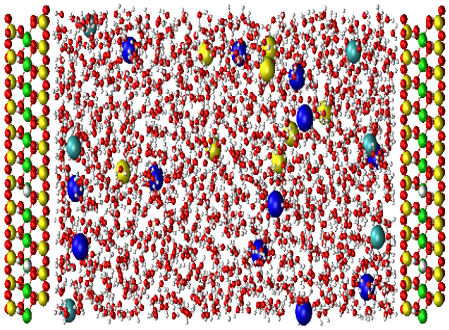
Implications: Finding the right material for separation of CO_2 from gas mixtures will significantly reduce the parasitic costs associated with this technology.

Accomplishments: NERSC resources have been used to develop an automated method to bypass a manual, time-consuming, visual analysis of void spaces in porous materials. This will enable high-throughput, unsupervised molecular simulations, and it means that NERSC resources can now be used at a larger scale to characterize many more materials. This work is also related to the BES/ASCR SciDAC SAP project "Knowledge Guided Screening Tools for identification of Porous Materials for CO_2 Separations."

NERSC Resources:
2010: 0.2 M hours used by repo m979; 1,400 jobs on Bassi, Franklin, Carver, and Hopper; up to 256 cores.
2011 through July: 0.6 M hours used; 800 jobs on Franklin, Carver, and Hopper; up to 336 cores.

Full Story: http://pubs.acs.org/doi/abs/10.1021/ci900451v

2. Nanoscale Modeling for Geologic Carbon Sequestration



Snapshot of a NERSC molecular dynamics simulation showing brine confined in a nanopore between two clay surfaces.

Principal Investigators: G. Sposito, I. Bourg (LBNL); L. Nielsen (UC Berkeley)

Project: Clay Mineral Surface Geochemistry Studies at the Center for Nanoscale Control of Geologic CO_2 (BES)

Objective: A portion of the research at this LBNL Energy Frontiers Research Center seeks to investigate molecular-scale processes relevant to the sequestration of CO_2 in geologic formations.

Implications: The performance of CO_2 repositories depends strongly on the ability of nanoporous clay seals to prevent CO_2 escape and on the efficiency with which the reservoir pore space can be filled with CO_2 . Our need to predict and optimize the behavior of CO_2 in geologic media depends on our ability to gain insight into mineral-water-CO_2 interfacial phenomena via molecular-scale simulation.

Accomplishments: Current investigations have focused on the structure of the electric double layer (EDL) on flat charged surfaces (smectite basal surfaces) in aqueous brines and on the CO_2-brine interfacial tension and solubility properties under conditions relevant to CO_2 sequestration. Molecular dynamics simulations have already yielded an unprecedentedly detailed view of water/ion distribution & dynamics near clay surfaces. Other studies will focus on molecular simulations of water, CO_2, and brine near solid surfaces, especially the nanopore between two clay layers. They will seek to know if CO_2 is absorbed at the water surface, how fast water and CO_2 molecules transfer across the interface, does an electric double layer form at the interface, and if so, would it affect water and CO_2 structure and dynamics by making the fluid more viscous.

NERSC Resources:
2010: 0.7 M hours used by repo mp47; 1,500 jobs on Bassi, Franklin, Carver, and Hopper; up to 2,048 cores.
2011 through July: 0.9 M hours used; 1,120 jobs on Franklin, Carver, and Hopper; up to 19,656 cores.
NERSC 2011 NISE award provided part of the time.

Full Story: http://esd.lbl.gov/research/facilities/cncgc/research/nanopore_processes.html

3.2.4 Supernova Caught in the Act

NERSC Analytics Group Lead and Palomar Transient Factory Principal Investigator Peter Nugent discusses the discovery of PTF11kly, the nearest Type 1a supernova seen in decades, on the PBS Newshour. The inset shows three images of the pinwheel galaxy in which it was discovered.

Principal Investigators: Peter Nugent (LBNL)

NERSC Project: Palomar Transient Factory (HEP)

Objective: Process, analyze and make available data from the fully automated Palomar Transient Sky wide-field survey to expose rare and fleeting cosmic events in a systematic way.

Implications: This is the first survey dedicated solely to finding transient events. Analysis at NERSC is fast enough to reveal transients as data are collected. By September 2011 the survey had already uncovered more than 1300 supernovae. Supernovae (SN) have a profound influence upon many diverse areas of astrophysics. However, our lack of complete understanding of many aspects of SN characteristics is due in part to technical challenges associated with locating SN in numbers required to create statistically meaningful samples.

Accomplishments: The earliest-ever detection of a supernova, made possible by NERSC and ESnet, happened in September 2011. It was essentially observed *within hours* of its explosion and will allow a rare glimpse at outer layers that contain hints about what kind of star exploded. The last time a supernova of this type occurred so close was 1972; before that 1937, 1898 and 1572. Images of this supernova contained information about its so-called early-time light curve, including strong features from unburned material (carbon and high-velocity oxygen).

NERSC Resources: NERSC data transfer nodes accept telescope data (300 GB/night) and archive to NERSC HPSS archives. Data (~150 TB now) are also resident on the NERSC Global Filesystem. The image subtraction pipeline runs on the NERSC Carver (iDataPlex) system, processing the images and detecting new transients. NERSC Science Gateway nodes serve the DeepSky database to astronomers worldwide.

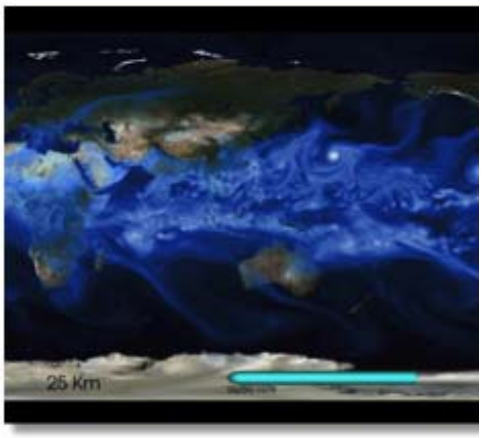
Full Story: http://www.nersc.gov/news-publications/news/science-news/2011/closest-type-ia-supernova-in-decades-solves-a-cosmic-mystery/

3.2.5 A Better Way to Identify Extreme Weather Events in Climate Models

Principal Investigators: James Boyle (LLNL), Michael Wehner (LBNL), and Prabhat (LBNL)

NERSC Projects: Program for Climate Model Diagnosis and Intercomparison, PCMDI (BER) and High Performance Visualization, Analytics, and I/O (ASCR)

Objective: Evaluation of low-probability, high-consequence (extreme) events in climate simulations using very high horizontal resolution climate models to capture events that are not seen in coarser simulations.



A one-year portion of a recent high-resolution atmospheric simulation that was conducted at NERSC.

Implications: Quantifying how well models can capture these extreme events is an outstanding question in climate change research. But more sophisticated, higher-resolution simulations produce more data, making it even more difficult to quantify rare but potent events. What's needed is an automated way to quickly comb through a climate simulation's huge dataset and tally the weather events that spell big trouble.

Accomplishments: New state-of-the art data mining, image processing, and topological analysis techniques have proven helpful in detecting extreme events, such as hurricanes, that might be buried in massive amounts of simulation output data. For example, 100 TB of output data from a 27-year atmospheric simulation were searched in two hours at NERSC. (For comparison, at the time the simulation was run, the entire Library of Congress contained only about 2.5 times as much data.) Another pattern-detection algorithm was developed to sift through large datasets for unusual weather phenomena derived from structures called atmospheric rivers that can transport significant amounts of water over long distances in very narrow bands.

NERSC Resources: The VisIt visualization package used in this work is supported by members of the NERSC Analytics Group, who presented results of these studies at the annual meeting of the American Geophysical Union in San Francisco in December 2011.

Full Story: http://newscenter.lbl.gov/feature-stories/2011/12/06/id-extreme-weather/

3.2.6 The Speculative ^{14}F Nucleus: Simulation Precedes Experiment

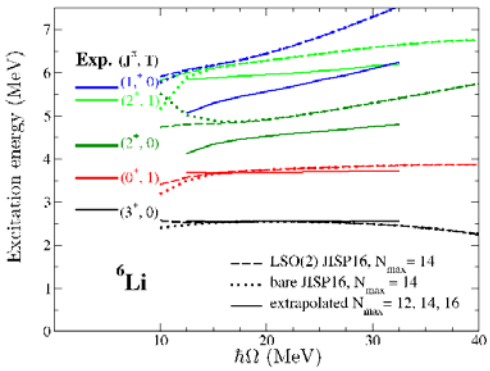
Principal Investigators: P. Maris and J. P. Vary (Iowa State University)

Project: Structure and Reactions of Hadrons and Nuclei. Building a Universal Nuclear Energy Density Functional (NP)

Objective: Develop and use *ab initio* predictive power to test nuclear theory by evaluating properties of exotic nuclei.

Implications: Many fundamental properties of nuclei are poorly understood today, and examining proton-rich light nuclei that are far from the "valley of stability" is a most demanding test of existing theory. Quantitative understanding will help create a predictive capability for key properties of short-lived or undiscovered nuclei that

could be important for improved energy sources or new medical diagnostic and therapeutic applications. The work will also provide important guidance for DOE-supported experiments.



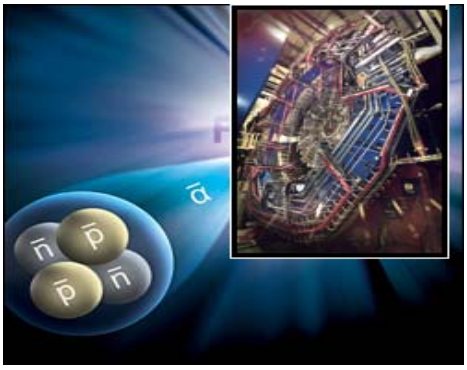
Comparison of calculated and experimental values for the spectrum of the 6Li nucleus, done at NERSC to validate an approach to extrapolating the total eigenenergies in the calculation of the speculative 14F nucleus. The experimental values are on the far left.

Accomplishments: NERSC and other DOE resources have been used to obtain converged ground state energies and nearly converged excited states of several light nuclei using realistic nuclear interactions, with results that compare favorably with experiment. This proves that the methods developed are accurate and that they preserve the predictive power of a microscopic theory with no adjustable parameters. NERSC platforms were critical in the development and optimization of this Many Fermion Dynamics–nuclear (MFDn) software package. A recent prediction of the structure of the highly unstable ^{14}F nucleus was supported by calculations at NERSC that applied the method to the ^{6}Li nucleus. These predictions served as motivation for experiments that ultimately achieved the first observation of fluorine-14.

NERSC Resources:
2010: 3.4 M hours used by repo m94; 4,900 jobs on Franklin and Hopper; up to 96,720 cores.
2011 through July: 13 M hours used; 3,420 jobs on Franklin, Carver, and Hopper; up to 129,168 cores.

Full Story: https://www.nersc.gov/news-publications/science-news/2011/proton-dripping-tests-a-basic-force-of-nature/

3.2.7 Heaviest Antimatter Nucleus Detected



A rendering of an antihelium-4 nucleus emerging from a Relativistic Heavy Ion Collider collision with the STAR detector shown in the inset.

Principal Investigators: G. Odyniec (LBNL) and R. J. Porter (NERSC)

Project: STAR Detector Simulations and Data Analysis (NP)

Objective: Operate an end-user facility supporting simulation, analysis, and archive of raw and reconstructed data for the STAR (Solenoidal Tracker At RHIC) detector at Brookhaven National Laboratory, an experiment designed to study the collision of heavy nuclei at very high energy.

Implications: High-energy STAR collisions create an energy density similar to that of the universe microseconds after the Big Bang. The STAR experiment is helping to answer the question of whether there are new states of matter that exist at those energies. Good understanding of detector response and the ability to rapidly and accurately analyze detector tracks of the collisions is essential for making progress in understanding the nature of matter and antimatter.

Accomplishments: NERSC's PDSF, a farm of Linux computers dedicated to nuclear and high-energy physics problems, continued to provide about one-half of the STAR computing capacity. NERSC's HPSS was also critical to physicists performing analysis of STAR data that produced results reported in 13 new publications, with seven

additional ones under review. Recent analysis of STAR data detected the antimatter partner of the helium nucleus: antihelium-4. This new particle, also known as the anti-alpha, is the heaviest antinucleus ever detected, and barring any new breakthroughs in accelerator technology, is likely to be heaviest stable antinucleus observed for decades to come. The discovery required sifting through data for half a trillion charged particles and resulted in detection of 18 antihelium-4 nucleus counts in 10^9 recorded gold/gold collisions with a calculated misidentification rate at the 10^{-11} level. The finding was reported in the journal Nature in April 2011.

NERSC Resources: At the start of 2011, STAR had about 800 TB of data on NERSC's HPSS. Steady production from embedding adds about 3 TB/month. STAR has consumed about one-half of PDSF cycles and also uses NERSC Globus facilities for data transfer.

Full Story: https://www.nersc.gov/news-publications/science-news/2011/heaviest-antimatter-particle-detected-with-nersc-help/

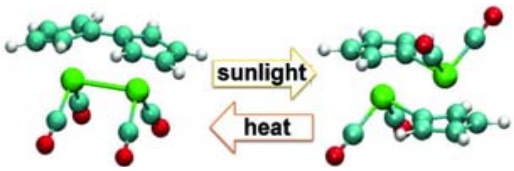
3.2.8 Promise for Solar Energy on Demand

Principal Investigator: J. Grossman (MIT)

NERSC Project: First Principles-Based Design of Novel Solar Thermal Fuels (BES)

Objective: Investigate electronic structure of novel materials to harvest solar power.

Implications: Trapping solar energy in chemical form would vastly improve its storage and transport — as long as it can be done reversibly. An attractive method might involve photoconversion of a suitable molecule to a higher energy isomer, which could later release the stored energy as heat. Electronic structure computations must be carried out to understand how a prospective molecule could do this reversibly.



A model of a molecule that reversibly changes it structure when it absorbs light. Red = oxygen; blue = carbon; green = iron; white = hydrogen. Simulations at NERSC revealed the mechanism for the heat-releasing step (2->1) and suggest substantial promise for creating batteries from molecules that can reversibly store solar energy as heat.

Accomplishments: NERSC researchers have performed first-principles density functional theory calculations on such a highly promising solar fuel candidate. They provided detailed atom-scale awareness of key conversion and storage mechanisms limiting its performance. Their findings indicate substantial promise for this class of molecule for both energy capture and storage.

NERSC Resources:
2010: 1.1 M hours used by repo m655; 1,640 jobs on Franklin, Carver, and Hopper; up to 640 cores.
2011 through July: 1.5 M hours used; 3,420 jobs on Franklin, Carver, and Hopper; up to 8,192 cores.
Uses the NERSC-provided version of the Quantum Espresso electronic structure code.

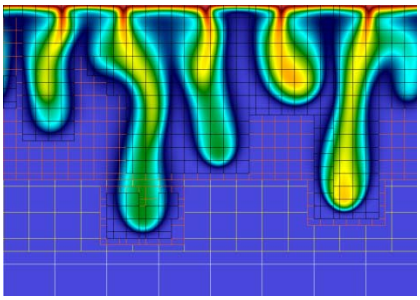
Full Story: http://www.nersc.gov/news-publications/science-news/2011/rechargeable-heat-battery-s-secret-revealed/

3.2.9 AMR for a Clearer Picture of Carbon Sequestration

Principal Investigators: G. Pau, J. Bell (LBNL)

NERSC Project: Simulation and Analysis of Reacting Flows (ASCR)

Objective: Apply a proven, parallel, structured-grid, adaptive mesh refinement (AMR) method to porous media flow to address challenges in simulating diffusive mixing when CO_2 is injected into saline aquifers for carbon sequestration.



A snapshot of CO_2 concentration after onset of convection showing an AMR grid with three levels of refinement.

Implications: Carbon sequestration is a geoengineering technique suggested as a possible mitigation strategy for global warming. High-resolution simulations providing accurate treatments of multiphase, multicomponent flow with a large number of reactive species must be carried out to accurately determine geologic storage capability across the huge length and time scales of interest.

Accomplishments: The Porous Media AMR code uses a proven block-structured framework that provides reuse of core technologies to reduce development efforts, good parallel scaling behavior, and significant gain in resolution compared to existing codes. Initial detailed 2- and 3-dimensional simulations show that the method captures the diffusion-convection process at space and time scales of interest for geological storage of CO_2 and has shown that mass flux behavior is convection-dominated, reaching a stabilized state that approaches a constant value. The code has been recently extended to include compressible flow.

NERSC Resources: This work is a subset of the work carried out in repo mp111. It used 2 M hours on Franklin (using up to 8,192 cores) and Hopper (using up to 12,000 cores).

Full Story: http://www.nersc.gov/news-publications/science-news/2011/clearer-picture-of-carbon-sequestration/

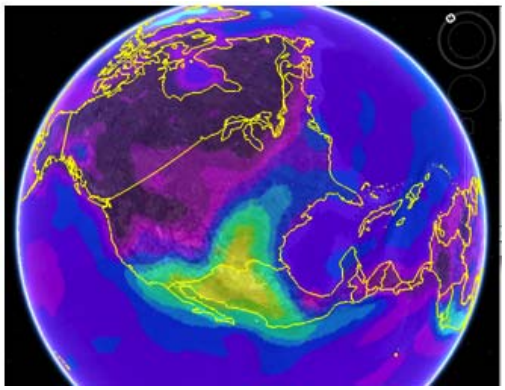
Sample Publication: Advances in Water Resources, 33(4):443-455, 2010.

3.2.10 Climate Studies Earn NERSC and PI an HPC Innovation Award

Principal Investigators: Gilbert Compo (University of Colorado)

NERSC Project: Surface Input Reanalysis for Climate Applications (Allocated by BER)

Objective: Comprehensive reanalysis of all global weather events from 1850 onward for a more detailed record of past weather variability than is currently available and also explain historically important extreme weather events.



Relative humidity for 1920–1929.

Implications: This project addresses objectives of the BER Climate Change Research Division to accurately predict any global and regional climate change induced by increasing atmospheric concentrations of aerosols and greenhouse gases. It will help

explain climate variations that may have misinformed early-century policy decisions and provide an important validation check on climate models being used to make 21st-century climate projections.

Accomplishments: The weather maps resulting from this work provide the first estimates of global tropospheric variability at a quality level similar to today's three-day weather forecasts but dating back more than 100 years. This is of enormous value to climate scientists as it will allow them to view current weather patterns in a historical perspective and determine if current weather extremes, such as the Northeast U.S. 2011 blizzard, are becoming more so. It has provided missing information and key insight on historically important extreme weather events such as the 1930s Dust Bowl, the deadly 1922 "Knickerbocker storm," and a variety of El Niño episodes. NERSC and the project's PI were one of nine winners of the first HPC Innovation Excellence Awards, given by the International Data Corporation to organizations achieving an important, quantifiable achievement with the help of high performance computing.

NERSC Resources:
2010: 4.1 M hours used by repo m958; 3,000 jobs on Franklin and Hopper; up to 8,640 cores.
2011 through July: 1.6 M hours used; 147 jobs on Franklin and Hopper; up to 3,360 cores.
NERSC 2010 and 2011 NISE awards provided part of the time.
Selected fields from their dataset are made available via a NERSC science gateway:
http://portal.nersc.gov/project/20C_Reanalysis/

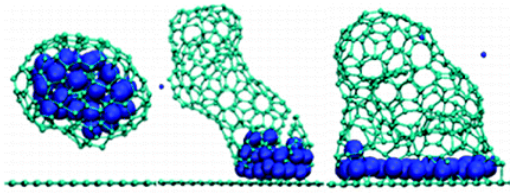
Full Story: http://www.nersc.gov/news-publications/science-news/2011/climate-time-machine/

3.2.11 Understanding Catalyzed Growth of Carbon Nanotubes

Principal Investigators: Perla Balbuena (Texas A&M)

NERSC Project: Catalyzed Growth of Carbon Nanotubes (BES)

Objective: Use custom simulation methods for the computational design of novel materials and processes, especially single-wall carbon nanotubes and their metal-catalyzed growth.



Snapshots from a molecular dynamics simulation of the vapor deposition growth of a SWCNT on a 15-atom metal catalyst (dark blue) deposited on a flat metal-carbide surface. The tube starts out as a "cap" attached to the metal cluster and grows outward.

Implications: Single-wall carbon nanotubes (SWCNT) are complex functional materials likely to become essential in new energy technologies. These lightweight materials have extraordinary strength — 100 times that of steel but only one-sixth the mass density — and can be synthesized as metals with conductivity superior to copper. There are multiple challenges, however; the most important is producing nanotubes in sufficient quantity, at low cost, and in controllable configurations. Simulation plays a key role in understanding the factors that control their catalyzed manufacture.

Accomplishments: Reactive classical molecular dynamics and first-principles density functional theory (DFT) simulations have been carried out with the objective of developing a thorough analysis of the metallic catalyst behavior during SWCNT synthesis on metal carbide surface substrates. Investigations have revealed how the chemical nature of the substrate and the attachment of the metal cluster to the substrate are crucial influences on the structure of the carbon networks synthesized on the catalyst surface. Recent studies have emphasized how to control nanotube chirality (the property representing the "handedness" of the tube structure). There are three basic kinds of nanotube structures, called zigzag, armchair, and chiral, and the three differ in mechanical and conductive properties. Balbuena's entirely computational research is providing significant new insights into the nucleation and growth of carbon nanotubes of specific chiralities. This is very important for the development of a

selective synthesis process, not yet available experimentally, and could have a huge impact on many applications of SWCNTs.

NERSC Resources:
2010: 0.5 M hours used by repo m249; 3,200 jobs on Bassi, Franklin, Carver, and Hopper; up to 256 cores.
2011 through July: 18,000 hours used; 5 jobs on Carver; up to 240 cores.
Uses the NERSC-provided versions of Gaussian, CPMD, and VASP.

Full Story: http://research.che.tamu.edu/orgs/groups/Balbuena/research_1.htm

3.2.12 A Selection of Papers Featured as Journal Covers (2010-2011)

Hierarchically porous graphene as a lithium-air battery electrode.
Jie Xiao et al., PNNL and Princeton University
NanoLetters, November 2011

Architectural simulation for exascale hardware/software co-design
John Shalf, Curtis Jansseny, and Dan Quinlan, LBNL, LLNL, SNL
IEEE Computer, November 2011

Moving protons with pendant amines: Proton mobility in a nickel catalyst for oxidation of hydrogen
M. O'Hagan, W. J. Shaw, S. Raugei, Shentan Chen, J. Y. Yang, U. J. Kilgore, Daniel L. DuBois, and R. M. Bullock, PNNL
Journal of the American Chemical Society, September 2011

Does nitric acid dissociate at the aqueous solution surface?
Tanza Lewis, Bernd Winter, Abraham C. Stern, Marcel D. Baer, Christopher J. Mundy, Douglas J. Tobias, and John C. Hemminger, UC Irvine, HZBME Berlin, PNNL
Journal of Physical Chemistry C, September 2011

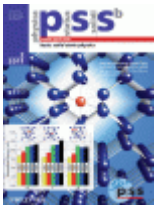
Dielectric nanodroplets: Structure, stability, thermodynamics, shape transitions and electrocrystallization in applied electric fields
W. D. Luedtke, Jianping Gao, and Uzi Landman, Georgia Institute of Technology
Journal of Physical Chemistry C, September 2011

Gyrokinetic particle-in-cell optimization on emerging multi- and manycore platforms
K. Madduri, Eun-Jin Im, K. Z. Ibrahim, S. Williams, S. Ethier, L. Oliker, LBNL, Kookmin University, Seoul, PPPL
Parallel Computing, September 2011

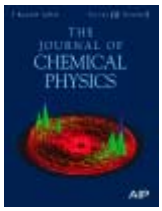
A new computational tool, Systematic Screening of Conformers (SSC), generates automatic and systematic searches of conformational spaces of molecules.
Sanliang Ling and Maciej Gutowski, Heriot-Watt University, Edinburgh
Journal of Computational Chemistry, July 2011.

A longstanding debate regarding the surface potential of water has been resolved.
Christopher J. Mundy, PNNL
Journal of Physical Chemistry B, March 2011.

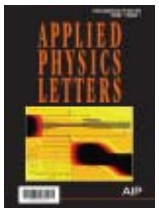
Simulations may lead the way toward rational design of polymer membranes, key components of fuel cells.
R. Davanathan, M. Dupuis, PNNL
Journal of Physical Chemistry, November 2010.



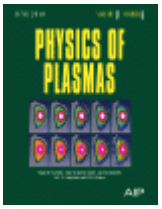
Quantum Monte Carlo methods present an accurate tool for the calculation of point defect formation energies in solids.
William D. Parker, John W. Wilkins, and Richard G. Hennig, Ohio State University and Cornell University
Physica Status Solidi (b), August 2010



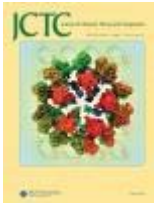
Quantum dynamics of a reaction important to both combustion and ozone depletion were revealed via simulations at NERSC.
H. Guo, University of New Mexico.
Journal of Chemical Physics, August 2010.



Quantum transport simulations may lead to discovery of nano-devices that can act as transistors.
S. Salahuddin, UC Berkeley
Applied Physics Letters, July 2010.



Simulations reveal the mechanism behind a plasma instability that can damage fusion reactors.
L. E. Sugiyama and H. R. Strauss, MIT and HRS Fusion
Physics of Plasmas, June 2010

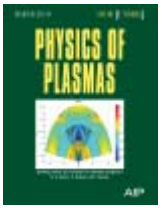


A new numerical approach for solving surface charges in interacting proteins realizes better accuracy, more flexible memory management, and at reduced cost.
Eng-Hui Yap and Teresa Head-Gordon, UC Berkeley, UC San Francisco, and LBNL
Journal of Chemical Theory and Computation, June 2010

Simulations convincingly demonstrate that certain nanofluids undergo a phase transition into a solid that would render them useless as lubricants.
P. Cummings, Vanderbilt University
AIChE Journal, April 2010.

Simulations help validate a new experimental technique that allows researchers to "see" individual atoms.
S. Pennycook, ORNL.
Nature, March 2010

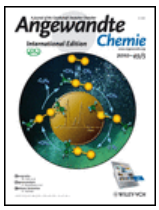
The Dynameomics database of protein folding pathways allows biophysical questions to be addressed in ways that were previously cumbersome or impossible.
Valerie Daggett, University of Washington
Structure, March 2010.



An innovative, computationally efficient method resolves velocity space dynamics in plasma fusion simulations.
M. Barnes, W. Dorland, and T. Tatsuno, University of Oxford and University of Maryland
Physics of Plasmas, March 2010

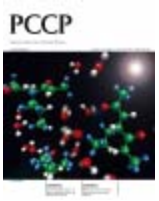
Simulations of cyclones in the warm Pliocene era may have implications for future climate.
A. V. Fedorov, C. M. Brierley and K. Emanuel, Yale and MIT
Nature, February 2010

The AMOEBA force field offers a significant improvement for more accurate simulation of structural and thermodynamic properties of small protein-like fragments.
J. W. Ponder, Washington University
Journal of Physical Chemistry B, February 2010.



First-principles simulations performed on Franklin reveal the mechanism of a full catalytic cycle for low-temperature activation of methane on small gold clusters.
U. Landman, Georgia Tech
Angewandte Chemi, International Edition, January 2010.

First principles simulations of the bonding, vibrational, and electronic properties of metal ion hydration shells provide critical insights into chemical processes.
Stuart A. Bogatko, Eric J. Bylaska and John H. Weare, UC San Diego and PNNL
Journal of Physical Chemistry A, January 2010



Simulations highlight the differences in the two co-solvent solutions important to studies of protein folding and functioning.
Margaret E.Johnson, Cecile Malardier-Jugroot and Teresa Head-Gordon, UC Berkeley, UC San Francisco, LBNL, and Royal Military College of Canada
Physical Chemistry Chemical Physics, January 2010

Simulations captured the detailed structure of a lean hydrogen flame on a laboratory-scale low-swirl burner.
J. Bell, M. Day, et al., LBNL
SciDAC Review, January 2010

3.3 Allocation of NERSC Director's Reserve of Computer Time

The NERSC director receives 10 percent of the total allocation of computer time. In 2010 almost all of this time was made available to NERSC's Initiative for Scientific Exploration (NISE) program. A very small amount (less than 5 percent) was kept in reserve in order to provide refunds to users whose jobs fail due to system failures.

NISE provides a transparent way to allocate the NERSC Director's reserve.

Any project exploring new scientific research areas, programming models, or algorithms may apply for NISE time. NISE research areas include:

- A new research area not covered by the existing ERCAP proposal, sometimes a tangential research project or a tightly coupled, supplemental research initiative.
- New programming techniques that take advantage of multicore compute nodes by using OpenMP, Threads, UPC or CAF. This could include modifying existing codes, creating new applications, or testing the performance and scalability of multicore programming techniques.
- Developing new algorithms that increase the ability to perform science (e.g., run at higher scale or incorporate new physics).

NISE time is distributed to projects that submit a NISE request and are judged by NERSC's science group to be the best projects to meet the above goals.

In 2011, 54 projects received awards totaling 88.5 million hours. Table 3-2 (on the following three pages) lists the 2011 NISE awardees.

Table 3-2. 2011 NISE Awards (1 of 3)

Investigator	NERSC Repo	Hours Awarded	DOE Office	Project Title
Dmitri Babikov, Marquette University	m409	1,450,000	BES Chemistry	New potential energy surface for ozone molecule
Connor Balance, Auburn University	m41	600,000	Fusion Energy	Hybrid OpenMP/MPI approach to R-matrix scattering
Amitava Bhattacharjee, University of New Hampshire	m148	1,000,000	Fusion Energy	Global Effects on the Dynamics of Plasmoids and Flux Ropes during Magnetic Reconnection
Ian Bourg, Berkeley Lab	mp47	300,000	BES Geosciences	MD simulations of liquid water films at the boundary between gas hydrate and mineral surfaces
Shane Canon, Berkeley Lab	kbase	100,000	BER Biosciences	Systems Biology Knowledge Base
Pierre Carrier, University of Minnesota	m684	80,000	ASCR Computer Science	Sparse Lanczos and GMRES solvers in co-array fortran (CAF)
Paul Cassak, West Virginia University	m866	200,000	Fusion Energy	Numerical Study of Mode Conversion and Ion Heating in Laboratory-Relevant Plasmas
Choong-Seock Chang, Princeton Plasma Physics Lab	m499	4,000,000	Fusion Energy	Large Scale Numerical Study of Bootstrap Current in Edge Pedestal Plasma
Wai-Yim Ching, University of Missouri	mp250	7,650,000	BES Materials Science	Ab initio simulation of mechanical properties of bulk metallic glass and Mo-bases alloys
Gil Compo, University of Colorado	m958	7,000,000	BER Climate Research	Improving algorithms and resolution for the Ocean-Atmosphere Reanalysis for Climate Applications OARCA (1850-2012)
Alan Condron, University of Massachusetts Amherst	m1068	1,000,000	BER Climate Research	High resolution modeling to identify the location of the meltwater flood responsible for the Younger Dryas cold episode
Thomas Devereaux, Stanford Linear Accelerator Center	m772	1,000,000	BES Materials Science	Numerical Investigations of Symmetry Breaking, Magnetism and the Pseudogap in the Three-Orbital Hubbard Model
James Drake, University of Maryland	mp217	500,000	Fusion Energy	Breaking magnetic field lines with turbulence
Huazhi Fang, Pennsylvania State University	m1215	500,000	BES Materials Science	Development of CALPHAD type approach to thermal conductivity aided by first-principles phonon approach
Keith Gilmore, Berkeley Lab	m1141	800,000	BES Materials Science	An efficient real-space approach to calculating Auger spectra
James Glimm, Stony Brook University	m1289	500,000	ASCR Applied Math	Surface Instabilities and Subgrid Scale Models in Computational Fluid Dynamics
Gary Grest, Sandia National Lab	m627	3,500,000	BES Materials Science	Self-Healing of Polymer Films
H. Guo, University of New Mexico	m627	300,000	BES Chemistry	Computational Studies of Methanol Steam Reforming
Michael Helle, Naval Research Lab	m965	1,000,000	HEP Accelerator Physics	Laser Acceleration of Ions Using Few Times Critical Density Gas Targets
Ivaylo Ivanov, Georgia State University	m1254	960,000	BER Biosciences	An Integrative Strategy to Model Complex Biological Assemblies

Table 3-2. 2011 NISE Awards (2 of 3)

Investigator	NERSC Repo	Hours Awarded	DOE Office	Project Title
Janet Jansson, Berkeley Lab	m1317	1,000,000	BER Biosciences	Unveiling Microbial Carbon Cycling Processes in Key U.S. Soils using "omics"
Puru Jena, Virginia Commonwealth University	m847	100,000	BES Materials Science	Ab initio studies on graphenes and superhalogens
Michael Kaplan, Desert Research Institute	m965	750,000	BER Climate Research	Tall Tower Wind Energy Monitoring and Numerical Model Validation in Southern Nevada
Noel Keen, Berkeley Lab	m1116	100,000	BER Environmental Science	Threading EnergyPlus using OpenMP
Emmanouil Kioupakis, University of California Santa Barbara	m934	1,000,000	BES Materials Science	Phonon-assisted optical absorption in indirect-band-gap semiconductors
Mal Soon Lee, Michigan State University	m968	400,000	BES Materials Science	Deep defect states and valence skipping in narrow band gap semiconductors
Mathan Lewis, California Institute of Technology	jcap	5,000,000	BES Materials Science	Material Simulations in Joint Center for Artificial Photosynthesis (JCAP)
Xiaoyi Li, United Technologies Corp Research Center	m945	550,000	ASCR Computer Science	High fidelity study of multiscale multiphase lubricant flow in high-speed gear systems
Jian Liu, University of California Berkeley	mp14	1,350,000	BES Chemistry	Semiclassical approaches for clean energy resources
Deyu Lu, Brookhaven National Lab	m1288	350,000	BES Materials Science	First Principles Characterization of van der Waals Dispersion Interactions
David Mikkelsen, Princeton Plasma Physics Lab	m64	3,000,000	Fusion Energy	Simulating ETG Plasma Turbulence with Impurities
Thomas Miller, California Institute of Technology	m822	750,000	BES Chemistry	Coupled electronic and nuclear dynamics in enzymes and photocatalysts
Thomas Miller, California Institute of Technology	m822	10,000,000	BES Chemistry	Proton-coupled electron transfer in photocatalysis and protein translocation in biosynthesis: Bridging lengthscales and timescales in molecular simulation
Benson Muite, University of Michigan	m1237	200,000	BES Materials Science	Three dimensional models for martensitic phase transformations
Victor Ovchinnikov, Massachusetts Institute of Technology	gc8	1,000,000	BER Biosciences	Molecular simulations of the release of phosphate and ADP from F1-ATPase
Edward Patton, National Center for Atmospheric Research	m917	5,500,000	BER Climate Research	Turbulence over complex terrain: a wind energy perspective

Table 3-2. 2011 NISE Awards (3 of 3)

Investigator	NERSC Repo	Hours Awarded	DOE Office	Project Title
Dmitry Pekurovsky, University of California San Diego	m1243	350,000	ASCR Computer Science	Performance enhancements for three-dimensional fast fourier transforms library P3DFFT
Kristin Persson, Berkeley Lab	matgen	2,500,000	BES Materials Science	Accelerated Materials Design Towards the Materials Genome
Mitch Pindzola, Auburn University	m41	3,000,000	Fusion Energy	Triple Photoionization of the Li atom
Mauro Pivi, Stanford Linear Accelerator Center	mp93	300,000	HEP Accelerator Physics	Implementation of C-MAD parallel code to include new physics and capabilities for future particle colliders
Tomasz Plewa, Florida State University	m461	1,000,000	NP Astrophysics	Morphology of Young Core-Collapse Supernova Remnants from Multi-Physics Three-Dimensional Simulations
Joel Primack, University of California Santa Cruz	mp363	1,500,000	HEP Astrophysics	Mass Production of High-Resolution Galaxy Simulations
Joe Prusa, Iowa State University	m612	400,000	BER Climate Research	Towards high performance urban canopy flows
Jason Riedy, Georgia Institute of Technology	mp156	250,000	ASCR Applied Math	Evaluating PGAS scientific graph analysis codes on the Gemini interconnect
David Romps, Berkeley Lab	m1196	1,000,000	BER Climate Research	Interactions of Clouds, Convection, and Climate
Sayeef Salahuddin, University of California Berkeley	m946	750,000	BES Materials Science	Simulation of Inelastic decoherence in electronic transport through nanoscale structures
Grady Schofield, University of Texas at Austin	m1291	1,000,000	BES Materials Science	Improving scalability of an eigensolver for the Kohn-Sham electronic structure problem
ShunLi Shang, Pennsylvania State University	m679	500,000	BES Materials Science	Phonon, thermodynamic and diffusion properties of lithium-ion batteries from first-principles calculations
Brian Van Straalen, Berkeley Lab	m1055	470,000	ASCR Computer Science	CACHE novel architecture research
Brian Van Straalen, Berkeley Lab	m1270	1,200,000	ASCR Computer Science	High performance simulation of reactive transport processes at the pore scale
Yi Wang, Pennsylvania State University	m891	1,250,000	BES Materials Science	First-principles calculations of multiferroic oxides
Michael Wehner, Berkeley Lab	mp193	5,600,000	BER Climate Research	High resolution integration of CAM5 for extreme weather research
Nicholas Zabaras, Cornell University	m1182	500,000	ASCR Applied Math	Theoretical investigation on the properties of energy materials
Qianfan Zhang, Stanford University	m1131	400,000	BES Materials Science	Modeling and Simulation of High Dimensional Stochastic Multiscale PDE Systems at the Exascale

4 Innovations

Charge question 4: What innovations have been implemented that have improved NERSC's operations?

NERSC Response: While we cannot cover everything our staff create and implement to improve facility operations, we've outlined a few highlights and organized them by benefit:

- *Improve usability of HPC systems*
- *Provide new computational models to users*
- *Increase NERSC center efficiency.*

4.1 Innovations to Improve Usability of HPC systems

4.1.1 Supporting Diverse Workloads on HPC Platforms

The HPC platforms at centers like NERSC provide incredible computing capability, however many users with growing computing have encountered barriers when trying to take advantage of these systems. Genomics is a clear example of this. Improvements in sequencing technology have led to year of year improvements of 3x or more. This has led to a commensurate increase in the computing requirements for this domain. This rapid growth in computing requirements was one reason that the Joint Genome Institute turned to NERSC to help meet its computing needs. Since HPC systems have not been specifically architected to match this workload, genomics researchers have typically avoided HPC systems in the past. However, the incredible scale of systems like Hopper coupled with the rapidly growing computing needs of genomics have forced a reevaluation. NERSC has worked with users in the genomics community to understand their workloads and developed tools to help lower the barrier to entry for these users. A few examples of the innovations we have developed include a task farmer, tools to create a private virtual cluster, and file caching utilities for data intensive applications.

Task Farmer

Many of the workloads in the genomics community, as well as other data intensive communities, are high-throughput oriented. Unfortunately, the typical scheduling policies as well as the run-time systems on typical HPC systems are not well suited for these workloads. To address this shortcoming, NERSC has developed a task farmer that simplifies running high-throughput workloads. The task farmer provides several critical capabilities, including load balancing, fault tolerance, and aggregation of output. The goal of the design is to make it trivial to run serial applications at scale. The task farmer uses a client-server architecture. The clients are launched in parallel as part of a parallel job submission. This helps address many of the scheduling policies constraints, since the throughput workload is scheduled as a single parallel job. The task farmer tracks the execution of tasks and automatically reschedules tasks that fail to complete. The framework also maintains a checkpoint of job progress and can be easily restarted after a failure or running out of wall time. The framework can also automatically collect output from the tasks and aggregate it into a single set of output files. This avoids the need to create output files for each task execution, which can lead to a large number of small files that are inefficient for the file system and difficult to manage.

The task farmer was initially developed to simplify performing alignment of sequence data using BLAST. The task farmer has been used to perform critical BLAST analysis for a variety of metagenomic and microbial data sets. The framework has since been generalized and used for other genomic analysis including hidden Markov model search (HMMER), assembly, and data reorganization.

Virtual Private Clusters

While the task farmer is useful for running many workloads, there is still a need for a more flexible tool for complex workflows. Workloads that have complex dependencies or need to dynamically generate new job steps based on the results from previous steps are more aligned with the capabilities of traditional schedulers and resource managers like Sun Grid Engine (SGE) or Maui Torque. However, making policy changes to an HPC scheduler typically requires careful review and may result in a negative impact for other workloads. To help address this situation, NERSC has developed tools to create a virtual private cluster (VPC) within a parallel allocation. So far NERSC has implemented a prototype with Sun Grid Engine, and the tool is called MySGE (currently in beta testing). Using this method, a user requests a large parallel allocation from the system scheduler (Moab in the case of Hopper). When the parallel job is started, the VPC system starts a personal SGE scheduler for the user and starts up the execution daemons on the allocated compute nodes. The user can then submit the workload to the personal scheduler, and these jobs will be executed on the allocated compute nodes. The jobs can include large job arrays, large numbers of serial steps, and job dependencies which are difficult to efficiently run on a typical HPC system. Furthermore, all of the services for the VPC run as the users, so MySGE doesn't require any special privileges to run. The user even has the ability to change the scheduling policies and configuration within the VPC. This means the user can create special queues or change job priorities for jobs runs in the VPC.

MySGE is ideal for users who have already implemented tools to submit and monitor jobs for throughput environments and would like to easily execute these workflows on an HPC system. MySGE also allows users to dynamically allocate additional nodes to the VPC, allowing the system to be adjusted based on the compute needs. MySGE is still under development but has already been used by several genomics groups; broader deployment will be evaluated next year. NERSC is also working on an implementation that will use the Torque resource manager which is called MyTorque.

File Cacher

One challenge to scaling up some data intensive workloads can come from reading-in large input or reference data sets. For example, when BLAST runs, the user specifies a reference database that is used to compare and align against. This reference database can be several gigabytes in size and is accessed via a memory-mapped file descriptor. At scale, this can place a large load on the file system, since the data is typically read in page size chunks (4 kilobytes) by all the compute cores in parallel. To help address this, NERSC developed a file caching mechanism. The cacher starts by reading-in the input data prior to starting the application. Since the cacher can read the input data in large blocks, it can achieve very good IO efficiency. The data is stored in a shared memory segment. A shared library is preloaded when executing the application. The shared library intercepts IO calls to the cached files and reads the data directly from the shared memory space. Using this method, read-only data can be read-in at close to peak bandwidth of the system. This reduces the startup costs by over an order of magnitude at scale and significantly reduces the stress on the file system.

Conclusion

NERSC has developed tools like the task farmer, MySGE, and the file cacher to lower the barrier of entry and enable new user communities to easily make use of the incredible computing capabilities at NERSC. NERSC is investigating other methods to support these types of workloads on HPC platforms. For example, NERSC is evaluating Cray's Cluster Compute Mode (CCM) to see what role it can play for high-throughput, data intensive workloads. Given the growing need for computational cycles from genomics and other data-intensive workloads, extending the capabilities of HPC platforms to these communities will be critical to supporting scientific discovery for all of DOE.

4.1.2 Increasing Scientific Productivity by Tracking Data

New experimental instruments, supercomputers, and networks are contributing to record levels of scientific productivity. To effectively meet the increasing scientific demand for data storage systems and services, NERSC's staff must first understand how data moves within the facility. Until recently, the process of obtaining these insights was extremely tedious because the statistics came from multiple sources, including network router statistics, client and server transfer logs, storage and accounting reports—all saved as very large, independently formatted text files.

Now a dynamic database created by the NERSC Storage Systems Group continually collects statistics from all of these sources and compiles them into a single, searchable repository. The system also automatically generates daily email reports and graphs that illustrate how data moves in and out of the facility's HPSS archival storage system, which is the largest repository of scientific data at the center.

The daily reports help the Storage Systems Group understand the frequency, amount, and method of data movement between the archive and other systems in the center. Because 50 percent of all data movement activities within the center involve HPSS, this capability allows them to identify user bottlenecks and gives them an indication of how we should invest in storage solutions.

Strategy for Planning Hardware and Software

When all the Storage Systems Group had were text files of various daily logs and reports, identifying specific events was like looking for a needle in a haystack, and quantifying trends proved to be extremely time consuming. But thanks to the new 50 gigabyte database, which contains several years of historical information with all the current information, a quick query will allow any analyst to instantly find a historical event—like when a new piece of hardware was installed—and quantify its usefulness within seconds.

For instance, one query showed that NERSC's Cray XT4 (Franklin) and XE6 (Hopper) systems contributed to the largest amount of data movement within the storage archive last year, which was not surprising. But the tool also revealed that NERSC's Data Transfer Nodes (DTNs)—which are dedicated servers for performing transfers between local storage resources like HPSS and the NERSC Global Filesystem, and wide area resources like Leadership Computing Facilities at Argonne and Oak Ridge—are major data movement systems for HPSS. As a result, the Storage Systems Group determined that the center should invest in more DTNs to aid users. Currently, those systems are being configured for installation.

Another aspect of the database that has been especially useful for storage analysts is the client software statistics, which look at the software used to transfer data to HPSS over time. These statistics have been critical in directing NERSC's efforts to improve software solutions in support of user data movement over time.

A query of the HPSS software used between May 1, 2010 and 2011 shows that the vast majority of data transfers to or from HPSS used the HSI client, which provides a UNIX-like interface into the system, or the HTAR client, which is similar to UNIX tar and is recommended for archiving small files. Due to the high usage of HSI on the Franklin and pre-production Hopper systems, the center made specific improvements in the way this client was deployed on Hopper. These improvements allowed users to achieve twice the bandwidth to HPSS for multi-gigabyte sized files.

Speeding Up Data Access for Users

For many data centers, tape libraries represent a cost, energy, and space efficient solution for storing ever-increasing amounts of scientific data. NERSC houses four SL8500 libraries, each composed of four library storage modules (LSMs), and each module contains about 2,500 tape cartridges and a collection of drives to read them. In total the center has about 40,000 tapes for user data archives and HPSS backups.

When a scientist makes a request to retrieve data from the tape library, a robot locates the tape where the requested information is stored, grabs it, and drops it into a tape drive that will read it to the user. This entire process occurs within a few seconds, unless all of the drives in that particular LSM are full. In that case, the robot will look for an available drive in another LSM.

Movements between LSMs are relatively slow because the cartridge either has to be deposited in an elevator and moved to another LSM within the same library, or go through a pass-through port into another library. The user will observe this as slower access time to their data. This really becomes a problem when the user is requesting data spread across numerous cartridges in LSMs with unavailable tape drives.

The new database allows staff to easily identify and monitor such events by including a section in the daily report that shows cartridge movements between LSMs. These reports allow storage analysts to track cartridge movements across LSMs daily, and determine if re-arranging cartridges and drive locations will speed up data access for the user.

Now that the usefulness of this database is becoming increasingly apparent, the Storage Systems Group would like to see the database grow to include statistics from the center's largest disk storage repository, the NERSC Global File system (NGF). The team would also like to design a graphic user interface (GUI) and analytics framework on top of the database so that all NERSC staff can use this information on demand for troubleshooting and decision-making.

4.1.3 Safeguarding Data with Parallel Incremental Backups

To ensure that users' data is always available when they need it, NERSC is extremely vigilant about regularly backing up persistent user data (that is, data that does not change from session to session) to tape. For many years, these backups were done with fairly basic mechanisms like *dump*, *tar,* and *cpio*. But as the center transitioned to a shared *project* file system—which is mounted on all major NERSC systems to facilitate data sharing—it became increasingly clear that these methods could not keep up. That's when Matthew Andrews of the NERSC Storage Systems Group started developing the Parallel Incremental Backup System (PIBS).

The group realized that they needed to have daily backup to ensure that users have current and relevant backup data, and that any system used to do this needed to be fast and scalable to meet the demands of future growth.

The center-wide *project* file system at NERSC aims to provide well-performing, persistent, no-purge storage to projects where scientists are sharing both code and data. In 2006 when the system was initially installed, it had 80 TB capacity with only 20 TB of data stored. At this stage the traditional backups were still feasible.

However, just one year later the system grew to 180 TB capacity with 60 TB of data stored. At this point, Andrews and his colleagues in the Storage Systems Group had developed a few scripts to split up jobs by project, and then run them in parallel to back up files onto tape cartridges. This method was quite effective for small- to mid-range projects, but larger projects still took weeks to back up.

This is when they saw a need for a custom backup system for NERSC, and started developing the parallel backup system called PIBS. A single installation of the PIBS software can back up multiple file systems. Within PIBS, each file system is subdivided into logical subtrees or "backup sets." The purpose of these sets is to determine its priority in the daily backup schedule, and whether the system will do a full or incremental backup.

Multiple instances of a schedule type allow full backups of different directories at different frequencies, or on different days, thus ensuring that the backup is current and minimizes disruption to users. After all, backing up files should not take up half of the system's bandwidth all day.

Every day the file system metadata is scanned to determine what files have changed and need to be backed up. A set of daemons, or background programs, will then write *.tar* files using the lists generated in the metadata scan, and transfer daemons store the *.tar* files to HPSS.

Although the system was just fully implemented at the end of last year, the storage systems group has already achieved some major successes. This past February, when the Storage Systems Group upgraded to a new *project* file system, they were able to leverage the scalability of the new software by using it to restore 500 TB of data in just seven days with no disruption to users. While staff members set up the new file system, users continued to save to the existing *project* file system while PIBS continued to back up the new data to HPSS. When the new system was up and running, the NERSC team used the file system backup in HPSS to restore the data.

Recent improvements to the PIBS software should enable data to be restored at a rate of over 100 TB per day using 10 Oracle T10KB tape drives.

This system is truly an innovation from the NERSC Storage Systems Group. It used to take weeks to copy from an old to new *project* system, and the group was competing with users for bandwidth to do it. With PIBS they got the new system up and running in three days with near zero impact to users.

4.1.4 Data Transfer Nodes Facilitate Scientific Collaborations

The ability to reliably move and share data around the globe is essential to scientific collaboration, that's why three Department of Energy (DOE) Scientific Computing Centers—the Argonne and Oak Ridge Leadership Computing Facilities and NERSC—have teamed up to focus on optimizing wide area network (WAN) transfers.

This ongoing effort began several years ago when each site deployed dedicated transfer nodes (DTNs), optimized for carrying data between the DOE facilities. Today, engineers from each site continue to meet regularly with DOE's Energy Sciences Network (ESnet) staff to develop strategies for optimizing bandwidth performance. This collaboration is called the Data Transfer Working Group, and their effort is having a huge impact on scientific research.

For example, Chesley McColl, Associate Scientist at the National Oceanic and Atmospheric Administration's (NOAA's) Earth System Research Laboratory, typically computes at one facility and analyzes data at another. He recently moved an 80 terabyte dataset from the high performance storage system at Oak Ridge to NERSC in just two months. Without the DTNs this would have taken as long as six months.

Because the DTNs mounted on NERSC's filesystems proved so useful to other researchers, the PDSF user community decided to mount one on their cluster too. PDSF is a networked distributed computing cluster designed primarily to meet the detector simulation and data analysis requirements of physics, astrophysics and nuclear science collaborations. Although PDSF is housed at NERSC, it is different from the center's other systems because it does not rely on allocations. Instead, each collaboration or group receives shares in the batch system proportional to their contributions to shared resources like compute nodes or staffing. NERSC's DTNs are important benchmarks for the PDSF transfer nodes. Running WAN transfer tests on both sets of nodes and comparing them ensures that the PDSF hardware is working well.

Because the perfSONAR network monitoring applications are deployed on all of the DTNs, users can now identify "choke points" in their data movements and work with NERSC staff to fix them. In the past, there have been issues with WAN transfers to PDSF, and finding the problem was like finding a needle in a haystack. But with the DTNs' well defined and maintained configuration, if there is a problem, staff can look at the transfer and system logs, identify the problem, and fix it.

This past year, NERSC also enabled Globus Online on their DTNs. This hosted service automates tasks associated with moving files between sites, or "endpoints," including retrying failed transfers, recovering from faults automatically whenever possible, and reporting status. So users simply log into the system, choose a start and end point, hit transfer, and the system does the rest.

Many scientists are familiar with tools like scp and GridFTP, but with Globus Online NERSC can offer a much simpler and faster method for moving data. Globus Online actually makes web-based data syncing an easy, nearly trivial process, so users don't have to be IT or middleware experts to move their files.

4.2 Innovations to Provide New Computational Models to Users

4.2.1 Provide GPU Testbed and Evaluation

In April 2010 NERSC fielded a 50-node GPU cluster called Dirac in collaboration with the Computational Research Division at Berkeley Lab, with funding from the DOE ASCR Computer Science Research Testbeds program. The system includes a complete set of development tools for hybrid computing, including CUDA, OpenCL, and PGI GPGPU-targeted compilers. The Dirac testbed provides the opportunity to engage with the NERSC user community to answer the following questions:

- What parts of the NERSC workload will benefit from GPU acceleration?
- What portions of the workload see no benefit, or insufficient benefit to justify the investment?
- Do GPUs represent the future of HPC platforms, or a feature that will be beneficial to a subset of the community?

Over 100 NERSC users have requested and been given access to the Dirac cluster. With their help, we have collected a broad range of performance data that support the idea that GPUs in their current configuration will be an important way to augment the effectiveness of NERSC systems for a subset of the user base.

ICCS Tests Two Scientific Codes on Dirac

Perhaps the largest collaboration using Dirac is the International Center for Computational Science (ICCS), a collaboration between Berkeley Lab, the University of California–Berkeley, NERSC, the University of Heidelberg, the National Astronomical Observatories of the Chinese Academy of Science, and Nagasaki University. ICCS explores emerging hardware devices, programming models, and algorithm techniques to develop effective and optimal tools for enabling scientific discovery and growth. A forthcoming special issue of the International Journal of High Performance Computing Applications will contain papers (some based on Dirac data) from the ICCS Manycore and Accelerator-Based High-Performance Scientific Computing Workshop held in Berkeley in January 2011.

ICCS researchers have used Dirac and other GPU clusters for rigorous testing of an adaptive mesh refinement code called GAMER used for astrophysics and other science domains, and the *N*-body code phi-GPU, which is used to simulate dense star clusters with many binaries and galactic nuclei with supermassive black holes, in which correlations between distant particles cannot be neglected. The results played a significant role in developing metrics for workloads and respective speedups.

The ICCS phi-GPU users won the 2011 PRACE (Partnership for Advanced Computing in Europe) Award for the paper "Astrophysical Particle Simulations with Large Custom GPU Clusters on Three Continents."[1] In this paper,

[1] R. Spurzem, P. Berczik, T. Hamada, K. Nitadori, G. Marcus, A. Kugel, R. Manner, I. Berentzen, J. Fiestas, R. Banerjee, and R. Klessen, "Astrophysical Particle Simulations with Large Custom GPU Clusters on Three Continents," Computer Science **26** (3–4**)**, 145–151 (2011).

they present direct astrophysical *N*-body simulations with up to 6 million bodies using the parallel MPI-CUDA code phi-GPU on Dirac and large GPU clusters in Beijing and Heidelberg, with different kinds of GPU hardware. They reached about one-third of the peak performance for this code in a real application scenario with hierarchically blocked timesteps and a core-halo density structure of a stellar system.

The ICCS GAMER users demonstrated significant speedups in a disparate class of scientific applications on Dirac and two GPU clusters in Beijing. By extensively reusing the extendable single codebase, they mitigated the impediments of significant code rewrites. They discovered load balance issues within GAMER and implemented remedies that will make the code more efficient and robust. And they demonstrated that the energy footprint of many-core devices individually may be high but is extremely efficient in the phase-space of flops per watt. This research will be presented at the SC11 conference.[2]

Other Results from Dirac

More than a dozen papers so far have been based at least in part on computations done on Dirac. Here are a few highlights:

- A study by researchers from Berkeley Lab's Computational Research Division, NERSC, UC Berkeley, and Princeton Plasma Physics Laboratory optimized particle-in-cell (PIC) kernels from a gyrokinetic fusion application across a broad range of emerging multicore designs (Figure 4-1). This included the first study on tuning gyrokinetic (PIC) algorithms for graphics processors. They achieved significant performance improvements of 1.3–4.7x on these complex kernels, despite the inherent challenges of data dependency and locality.[3]
- Researchers from Argonne National Laboratory, testing coupled cluster equations for electronic correlation in small- to medium-sized molecules, found that the GPU-accelerated algorithm readily achieves a factor of 4 to 5 speedup relative to the multithreaded CPU algorithm on same-generation hardware.[4]
- The CUDA Data Parallel Primitives Library (CUDPP),[5] a project of the University of California–Davis and NVIDIA Corporation, has also been a major user of Dirac. CUDPP is a library of primitives that are important building blocks for a wide variety of data-parallel algorithms, including sorting, stream compaction, and building data structures such as trees and summed-area tables.

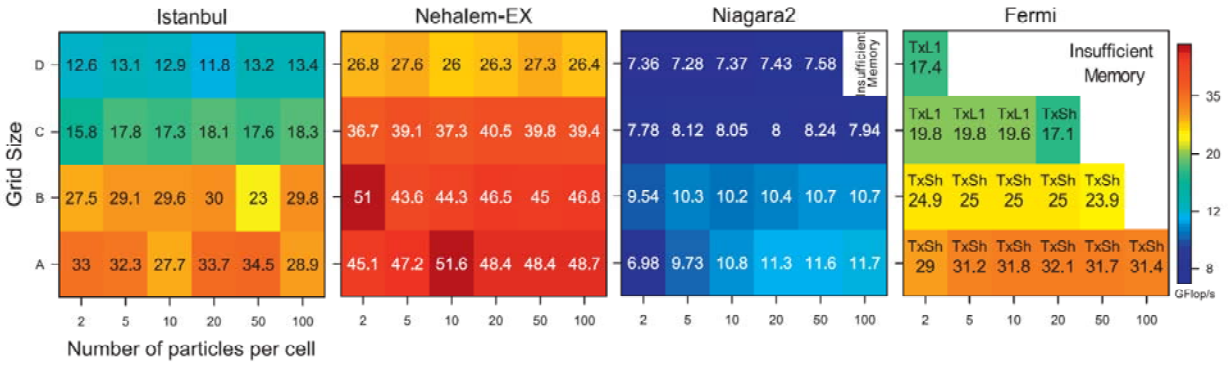


Figure 4-1. SMP/GPU push performance in Gflop/s (450 flop/s per particle per iteration) achieved by the best implementation on each system for varying grid sizes and particle densities.

[2] Hemant Shukla, Hsi-Yu Schive, Tak-Pong Woo, and Tzihong Chiueh, "Multi-Science Applications with Single Codebase — GAMER — for Massively Parallel Architectures," Proceedings of SC11 (accepted).
[3] Kamesh Madduri, Eun-Jin Im, Khaled Ibrahim, Samuel Williams, Stéphane Ethier, and Leonid Oliker, "Gyrokinetic particle-in-cell optimization on emerging multi- and manycore platforms," Parallel Computing, in press.
[4] A. E. DePrince and J. R. Hammond, "Coupled Cluster Theory on Graphics Processing Units I. The Coupled Cluster Doubles Method," Journal of Chemical Theory and Computation **7**, 1287 (2011).
[5] http://code.google.com/p/cudpp/

4.3 Innovations to Increase NERSC Center Efficiency

In 2010 NERSC started instrumenting its machine room with state-of-the-art wireless monitoring technology to gather information on variables important to machine room operation, including air temperature, pressure, and humidity, and chilled water flow and temperature. With the addition of new sensors for Hopper and Carver/Magellan, a total of 1,036 sensors were installed.

In 2011, NERSC added 475 electrical circuit meters to collect power usage data at four different levels of the OSF electrical infrastructure (Figure 4-2).

Cray supercomputers also have their own built-in sensors to monitor power, temperature, and air pressure at various points in the cabinets.

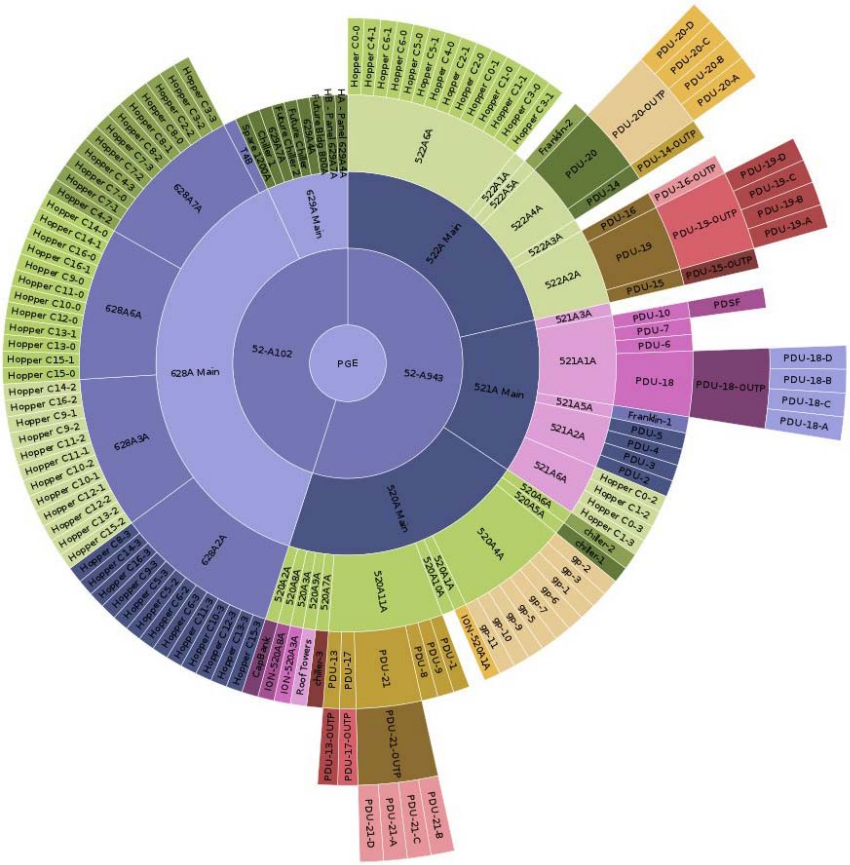


Figure 4-2. Electrical circuit meters provide data on electrical power flow and distribution throughout the NERSC center, all the way from the utility company's 9 MW power line to the computers. This diagram shows the first 175 meters installed; a total of 475 meters are now in operation.

NERSC's Outreach, Software, and Programming Group created database systems that collect and archive all of this environmental, power, and Cray data every five minutes; and they developed a web-based interface that provides a single view of the data.

This integrated monitoring infrastructure provides both real-time information to Operations staff and trend information that can be used for optimizing performance—for example, optimizing airflow management. This data is also being used by NERSC staff working with data center efficiency experts from Berkeley Lab's Environmental Energy and Technologies Division to analyze and optimize power usage. Future machine room design modifications will also take advantage of this data.

5 Risk Management

Charge question 5: Is NERSC effectively managing risk?

> *NERSC Response: Yes. NERSC has adopted a proven risk management strategy and incorporated these practices in both its normal operations and large projects, such as acquisition of large systems and deployment of new technologies. The risk management process, updated in March 2011, includes a documented risk management plan, regular risk reviews, and risk tracking via a commercial risk management register (RiskRadar). Risk management is an iterative and systematic process, building upon experience over time.*

5.1 Risk Management Strategy

The risk management plan describes how risks are identified, ranked, and monitored. Risk management is an iterative and systematic process, which includes identifying the root cause, likelihood, and impact for each risk. In addition, risks are assigned owners and mitigation and contingency actions are created depending on the severity of the risk. During risk meetings, risks are reviewed and updated, changes to risks are explained, new risks are added if needed, and old risks are retired. These changes are implemented in the Risk Register.

Beyond the formal risk process, risk management is an integral part of the daily operations of the center. Changes to the system, user policies and technology investments are all evaluated for risks, and mitigation and contingency strategies are developed where needed.

In a continuing effort to improve risk management at NERSC, a new position, NERSC Risk Manager, reporting to the Deputy Director, was established to provide risk management oversight at NERSC. Various duties include:

- Lead NERSC Division risk management.
- Coordinate risks mitigation and response planning.
- Manage one-on-one meetings with individual Risk Owners and review/update current risk status and info.
- Serve as the subject matter expert for risk management.
- Manage the risk reporting process, including providing inputs to DOE such as the annual Operational Assessment Review.
- Survey risk practices at other high performance computing centers and make recommendations for improvements in software and procedures for effective usage at NERSC.
- Manage the implementation and maintenance of the program.
- Lead cross-group activities for managing risks that involve multiple groups, and devise mitigations.
- Establish risk mitigation project schedules and budget estimates.
- Schedule and lead periodic risk management sessions and report findings to management.
- Maintain and update the NERSC Risk Management Plan and Risk Register.

5.2 Current Risk Status

The year 2011 has been exciting on the risk management front. The NERSC-6 Hopper system went into general user availability, entered and completed system availability with final system acceptance on April 15, 2011. This culmination of activities also brought the successful completion of the project. NERSC-6 closeout activities included closing (retiring) appropriate risks and transferring remaining steady state risks, not retired, to the NERSC Program. Another major activity was the 3 MW upgrade that was successfully completed and the corresponding risk retired.

As of December 2011, NERSC has 45 active risks, of which one is rated as a high risk, nine as medium risks, and 35 as low. NERSC's highest risk, Budget Insufficient to Meet DOE Commitments, which would result from a significant decrease in funding relative to the OMB 300 plan, did not occur in 2011.

5.3 Major Risks Tracked for the Current Year (2011)

Here are the major risks that were tracked in 2011. More details about specific risks will appear in the following sections.

Risk	Section	Risk Rating	Status
Budget Insufficient to Meet DOE Commitments	5.3.1	High	Did not occur.
Hopper Phase 2 Stability	5.3.2	Medium then Low	Rating lowered to Low.
Loss of Key Personnel (Temporary or permanent)	5.3.3	Medium	Two Group Lead changes occurred.
Disruptive Vendor Changes — Oracle acquisition of SUN	5.3.4	Medium	Risk split into two separate risks (Tape: 5.6.4 & Lustre: 5.6.6).
OSF Computer Room Insufficient	5.3.5	Medium	Retired.
Budget Uncertainties	5.3.6	Medium then Low	Retired and replaced by two new risks: 5.3.1 & 5.6.1.
Computer Room Safety	5.3.7	Medium then Low	Rating lowered to Low.

5.3.1 Budget Insufficient to Meet DOE (Allocation) Commitments

Probability			
High			
Med			X
Low			
	Low	Med	High

Impact

Owner: Kathy Yelick, NERSC Director

Trigger: DOE informs NERSC of budget reduction

Strategy: Accept

Description

Every budget year brings the risk of a reduction in budget. A significant decrease in funding relative to the OMB 300 plan for NERSC may interrupt or halt improvements to the resources that NERSC provides.

Mitigation

Since the occurrence of this risk is controlled by DOE, the risk is accepted. Mitigation strategies are implemented by DOE.

Contingency

Leadership in DOE's Office of Science is made aware of the wide-ranging potential impacts of funding cuts. Communicate as often as possible. Reduce NERSC services and system purchases, stop maintenance payments or decommission systems earlier than planned.

Status

Risk rating remains high pending FY2012 budget process.

5.3.2 Hopper Phase 2 Stability

Probability	Low	Med	High
High			
Med			
Low		X	

Impact

Owner: Jay Srinivasan, Computational Systems Group Lead

Trigger: Inability to meet contractual and DOE performance and availability metrics

Strategy: Mitigate

Description

Hopper Phase 2 is one of the first Cray XE6 systems deployed at large-scale to a computing facility. The XE6 uses the new Cray Gemini interconnect and a new node design that features the latest AMD Magny-Cours Opteron chips. A new version of the Cray Linux Environment, CLE 3.1, has been built in order to expose the full functionality of this new hardware to user applications. Taken together, there is risk that—despite a rigorous acceptance test—the system could be less stable than other NERSC systems as it enters early science-use and then production.

Mitigation

NERSC has been in close communication with the vendor throughout the development of the XE6 and has required monthly updates on the features, testing, and schedule of the new hardware and software. NERSC has developed an acceptance test and ongoing contractual metrics based on:

- performance of actual scientific simulation codes
- component and system-wide failure rates
- acceptable level of jobs completing.

Cray must meet these metrics or provide additional or alternative equipment.

A test-and-development system with the same internal and external components has been deployed as a testbed for hardware and software changes.

Contingency

The XE6 Gemini interconnect could be replaced by the previous design, called Seastar. This could be run in production until Gemini is stabilized.

Status

Lowered risk probability to low (from medium) given successful system acceptance on April 15, 2011. System stability has been very good.

5.3.3 Loss of Key Personnel (Temporary or Permanent)

Probability			
High			
Med		X	
Low			
	Low	Med	High

Impact

Owner: Kathy Yelick, NERSC Director

Trigger: Retirement, resignation, transfer, external competition

Strategy: Monitor

Description

NERSC has a highly trained but not overly large staff. Sudden staff departures (including illness) of key personnel may cause a disruption of services and resources provided by NERSC to the DOE researchers.

Reasons include: competition with industry and other labs; inability to give raises under current DOE policy.

Mitigation

To encourage retention of key personnel, managers meet with direct reports to discuss career goals and develop organizational structure that allows for growth of individuals; recognize major accomplishments through awards and promotions. Managers ensure cross-training of NERSC center staff so that essential tasks can be carried out by at least two people.

Contingency

Reassign job responsibilities to cover all the critical or major roles from the designated backup staff. Hire into key positions as needed.

Status

Two Group Lead changes, the Computational Systems Group Lead and the Operations Group Lead, occurred this year, and the positions were filled with minimal disruption. All affected critical or major job responsibilities were covered by existing staff while the positions were being competed. Other open positions have been filled expediently.

5.3.4 Disruptive Vendor Changes: Oracle Acquisition of Sun Microsystems

Probability			
High			
Med		X	
Low			
	Low	Med	High

Impact

Owner: Jason Hick, Storage Systems Group Lead, and Jay Srinivasan, Computational Systems Group Lead

Trigger: Notice of discontinued products and/or support

Strategy: Mitigate

Description

Oracle finalized the acquisition of Sun Microsystems in February 2010. The acquisition increases risk for delivering production quality systems to our users in two important areas: the Lustre scratch file systems that we currently have on Franklin and Hopper, and the tape infrastructure that comprises the HPSS archival storage systems.

Mitigation

When procuring the next generation system, NERSC negotiates many contractual options with the selected vendor primarily to meet certain guaranteed levels of performance. Vendors are selected on their ability to deliver hardware and services for the anticipated life-cycle of the resource.

Contingency

If a vendor should cease business, NERSC can operate and maintain the resource with in-house expertise while alternatives can be procured or developed.

Status

This risk was split into two different risks to focus more on the specifics of each risk: Tape/HPSS (5.6.4) and Lustre (5.6.6) development and support.

5.3.5 OSF Computer Room Insufficient in the Short Term

Probability			
High			
Med			
Low			X
	Low	Med	High

Impact

Owner: Howard Walter, NERSC Deputy

Trigger: OSF power, cooling or space is insufficient for expected systems

Strategy: Mitigate

Description

The NERSC Oakland Scientific Facility (OSF) computer room space, power and/or cooling become inadequate to support future DOE needs.

Mitigation

Constrain procurements to fit within available facility resources; use more energy-efficient methods to cool and power the computer room; procure new systems, which use energy-efficient technologies; and/or add additional capacity to the OSF.

Contingency

NERSC develops a plan to expand the OSF computer room space, power, or cooling or develops a plan to seek a new site elsewhere.

Status

Risk was retired when the OSF's power and cooling capacity were successfully upgraded to 9 MW and 1,100 tons of additional cooling capacity. A new risk, CRT Facility Occupancy Delayed (5.6.3), has been created to address future DOE needs.

5.3.6 Budget Uncertainties

Probability			
High			
Med			
Low		X	
	Low	Med	High

Impact

Owner: Kathy Yelick, NERSC Director

Trigger: Monthly financial forecast predicts Budget Authority will be less than expected costs

Strategy: Mitigate

Description

Every budget year brings the risk of a Continuing Resolution of undetermined length. Cash flow for this time period needs to be planned carefully. In addition, power costs for the upcoming year cannot be predicted precisely, due both to the possibility of price variation and to the range of estimates for new systems.

Mitigation

Cash flow planning for a Continuing Resolution (CR).

Contingency

Delay some acquisitions. Carry forward additional funds. Plan larger procurements for later in the year. Notify DOE of cash flow issues

Status

Late in 2010 this risk was rated as medium given the budget impasse and the CR process. Early in 2011, it was determined at a risk review that the Budget Uncertainties risk was too general since the risk co-mingled short-term CR cash flow risks with longer-term electrical power costs. This risk was retired and two new risks were created. The Unplanned Electricity Cost Increases risk describes electrical power costs. The Annual Budget Plan under Continuing Resolution risk addresses the CR risk and is described in 5.6.1.

5.3.7 Computer Room Safety

Probability			
High			
Med			
Low	X		
	Low	Med	High

Impact

Owner: Betsy MacGowan, NERSC Safety Coordinator

Trigger: Near-miss or injury

Strategy: Mitigate

Description

The OSF computer room has unique safety concerns for people working there. High electrical power, mechanical cooling equipment, heavy racks, seismic events, and noise all contribute to a potentially hazardous environment.

Mitigation

Safety systems are periodically tested, and preventive maintenance is scheduled. Safety training is required for anyone working in the room, and the required training is annually reviewed by the LBNL safety division. Annual earthquake drills are conducted. The NERSC Safety Coordinator reviews and updates training and procedures.

Safety meetings are regularly scheduled and attended by senior management. Individuals are given Spot Awards for reporting significant near-miss incidents. Safety issues are reviewed by NERSC management and the LBNL safety division, and corrective actions are tracked for completion. Those with access to the room are trained in safety issues, including seismic safety, using an online training, updated with new risks such as the noise issue.

Contingency

Specialized safety training and staff meetings to address specific issues.

Status

This risk continues to be actively worked and monitored. The Emergency Power Off system was tested in 2010 to verify that all power panels were de-energized. An online computer room safety course was created, and every individual with computer room access is required to take it annually. In addition, the LBNL security badge system has been modified to deny computer room access to anyone who has not completed the course in the last year.

5.4 Risks That Occurred and the Effectiveness of Their Mitigations

Risk	Section	Risk Rating	Main Mitigations
Budget Uncertainties	5.3.1	Medium then Low	This risk was rated Medium for most of the year due to the prolonged budget and CR process for FY11. NERSC's cash flow was adjusted to minimize impact to NERSC users and systems. The rating was changed to Low once FY11 funding was finalized.
Hopper Phase 2 Stability	5.3.2	Medium then Low	This risk was rated Medium due to Gemini interconnect uncertainties and pre-acceptance risk. By using a mitigation strategy of monitoring SSP, benchmark performance, and job success rates, NERSC was able to successfully identify bugs causing periods of performance slowdowns and job failures. Acceptance was delayed until Cray fixed issues that were identified, with eventual successful system acceptance on April 15, 2011.
Disruptive Vendor Changes: Oracle acquisition of SUN	5.3.4	Medium	This risk was complicated to manage given there were multiple issues to deal with after the merger of Oracle and Sun. This Oracle risk was eventually split into two issues dealing with development and support for both Tape/HPSS & Lustre. See 5.6.4 & 5.6.6 for additional details.

5.5 Retired Risks During the Current Year

Retired Risk	Section	Risk Rating	Info
OSF Computer Room Insufficient in the Short Term	5.3.5	Medium	The OSF power and cooling project was successfully completed in FY2011.
Staff Safety Incident during the 3 MW Power Upgrade	N/A	Medium	This risk was retired with the successful completion of the 3 MW upgrade. No injuries occurred.
Budget Shortages	N/A	Low	This risk originated from the NERSC-6 Project and was transferred to the NERSC Program. Risk retired when new budget risks were created. New risks: 5.3.1 & 5.6.1.
Facility Power, Cooling and Floor Space Limits Remediation Choices	N/A	Low	Retired upon NERSC-6 acceptance.
Budget Uncertainties	5.3.6	Medium then Low	Retired and replaced by two new risks: 5.3.1 & 5.6.1.

5.6 Major New or Re-Characterized Risks Since the Last Review

Risk	Section	Risk Rating	Main Mitigations
Annual Budget Plan under Continuing Resolution	5.6.1	Medium	Carry forward additional funds. Plan larger procurements for later in the year. Discuss cash flow analysis at the monthly senior management financial meeting and inform DOE of cash flow issues in a timely manner.
Facility Impacts on Tape System Reliability	5.6.2	Medium	Regularly clean the computer room. Minimize and supervise activities in the immediate vicinity of the tape libraries.
CRT Facility Occupancy Delayed	5.6.3	Medium	Monthly CRT project meetings to provide early detection of schedule slippage.
Disruptive Vendor Changes — Oracle Tape (HPSS)	5.6.4	Low	Self maintain; rely on other third parties to support tape libraries; evaluate other tape libraries.
NERSC Fails to Meet IPv6 Compliance for NERSC Systems and/or Networks	5.6.5	Medium then Low	All NERSC procurements since 2008 have required IPv6 compliance; major hardware and software vendors support IPv6 semantics in network devices and operating systems.
Disruptive Vendor Changes — Oracle Lustre	5.6.6	High then Low	Rely on Cray to support Lustre; rely on Data Virtualization Service (DVS) to use alternate file systems if needed; evaluate open source Lustre.
Failure of largest system	5.6.7	Low	Availability and Mean Time to Repair are continuously tracked. Dual filesystems and redundant login nodes are configured. Hardware and software changes are tested on a test system before being installed on the production system.

5.6.1 Annual Budget Plan under Continuing Resolution

Probability			
High			
Med		X	
Low			
	Low	Med	High

Impact

Owner: Kathy Yelick, NERSC Director

Trigger: Financial forecast predicts Budget Authority will be less than expected costs

Strategy: Mitigate

Description

Budget authority under Continuing Resolution is reduced or prorated to prior FY level or below and is insufficient for current FY planned obligations.

Mitigation

- Carry forward additional fiscal year funds.
- Plan large procurements later in the year and delay where possible.
- Discuss cash flow analysis at the monthly senior management financial meeting and inform DOE of cash flow issues in a timely manner.
- Leases are planned for a 36-month period, consistent with our longer-term plan to buy a major system every three years. There are no major early payments in the plan that might be impacted by an extended continuing resolution. The early payments are planned in July, which puts them beyond the reasonable range of even an extended continuing resolution.

Contingency

Reduce or postpone planned obligations (reduce center-balance items, planned maintenance, etc.).

Status

As of 12/31/2011, Budget Authority is only 33% of FY2012 guidance.

5.6.2 Facility Impacts on Tape System Reliability

Probability			
High			
Med			
Low			X
	Low	Med	High

Impact

Owner: Jason Hick, Systems Storage Group Lead

Trigger: Tape errors exceed observed norm

Strategy: Mitigate

Description

Due to various facility activities, dirt and dust could have an adverse effect on the tape system. Given the long-term nature of the tape system, steps to mitigate the effect of facility activities are considered.

Mitigation

Institute a program to regularly inspect/clean the machine room.

Monitor to identify and remove suspect media and drives.
Minimize and supervise activities in the immediate vicinity of the tape libraries.

Consider protective coverings and ventilation (e.g., a bubble) around the tape libraries during activities where debris/contamination can occur.

Contingency

Engage with environmental cleanup companies and Oracle Professional Services to contain the libraries from potential contamination (e.g., identify the source and mitigate against it directly), and clean up the contaminant.

Status

Machine room cleaning each quarter to ensure we minimize airborne particulates has been initiated. We are in the process of installing humidity, air pressure, and temperature sensors on the tape libraries. These will help to ensure we meet the humidity and temperature recommendations and requirements for tape storage. In addition, we have taken several different particulate readings in and around the tape libraries in the past year.

5.6.3 CRT Facility Occupancy Delayed

Probability			
High			
Med		X	
Low			
	Low	Med	High

Impact

Owner: Howard Walter, NERSC Deputy

Trigger: Significant schedule slippage

Strategy: Mitigate

Description

The Computational Research and Theory (CRT) building will provide the increased power and energy-efficient cooling required for future generations of high performance computing systems. CRT construction has started and is scheduled for completion in the second half of 2014; delays could adversely affect NERSC systems planned for 2015.

Mitigation

Monthly CRT project meetings to provide early detection of schedule slippage.

Contingency

Upgrade OSF power and cooling. Procure lower-power systems and replace less efficient older systems in the OSF. Occupy CRT computer room before office floors are completed.

Status

The first phase of the CRT project, site preparation and underground utilities, has started and is on schedule. The main construction will begin in April 2012, and building occupancy is estimated for fall 2014.

PG&E has confirmed up to 3 MW of new power is available at OSF.

5.6.4 Disruptive Vendor Changes: Oracle & Tape (HPSS)

Probability			
High			
Med			
Low	X		
	Low	Med	High

Impact

Owner: Jason Hick, Storage Systems Group Lead

Trigger: Oracle refuses support or unreasonably increases pricing

Strategy: Mitigate

Description

The tape library manager, tape libraries, and tape drives in the archival storage system are StorageTek products now owned by Oracle. The primary risks are in affordability of tape products and support, and continued interoperability of their products with NERSC infrastructure.

Mitigation

To mitigate risk with the tape subsystem, NERSC has two initiatives:

1. **Executive involvement in the Large Tape Users Group.** This enables us to have private dialogues with Oracle tape engineering and support to ensure we have input into the current and future direction. Support issues are addressed outside of the normal support process and we ensure they are familiar with our infrastructure needs.

2. **Evaluation of competitive alternatives.** We are currently evaluating a variety of other tape products to ensure that we minimize disruption to users in the event we cannot afford the Oracle solution or their technology requires significant infrastructure changes.

The above mitigation steps minimize the impact and likelihood of Oracle tape products failing to work in the NERSC environment. The mitigation makes this risk low.

Contingency

If a vendor should cease business, NERSC could use a third-party service/support provider to maintain the Oracle hardware. In the worst case, NERSC can operate and maintain the resource with in-house expertise while parts are available, or until alternatives can be procured or developed.

Status

NERSC continues to be actively involved in with executive involvement in the Large Tape Users Group, and NERSC is also evaluating new tape libraries available. Our mitigation steps and what has occurred over the past year and a half have reduced the risk of the Oracle merger affecting their tape solutions. They are listening to customers and developing software/hardware products that improve our environments.

5.6.5 NERSC Fails to Meet IPv6 Compliance for NERSC Systems and/or Networks

Probability	Low	Med	High
High			
Med			
Low	X		

Impact

Owner: Jason Lee, Networking

Trigger: Inability to meet DOE IPv6 requirements

Strategy: Mitigate

Description

DOE has mandated that all public services such as web and email be IPv6 compliant by the end of FY12. An IPv6 compliant product or system must be able to receive, process, and transmit or forward (as appropriate) IPv6 packets and should interoperate with other systems and protocols on both IPv4 and IPv6 modes of operation.

Mitigation

NERSC/DOE has a requirement for IPv6 and interoperability with IPv4, and this requirement has been included in all NERSC procurements since 2008. Procure and test IPv6 systems, which provide public services, early in FY12.

Contingency

NERSC has in place hardware that can translate between IPv6 and IPv4 packets (NAT). This hardware will be placed into service in the event that not all public services can be converted to native IPv6.

Status

NERSC currently has several IPv6 allocations. NERSC's web infrastructure is currently IPv6 enabled and has been since world IPv6 day (June 8, 2011).

5.6.6 Disruptive Vendor Changes: Oracle Lustre

Probability	Low	Med	High
High	X		
Med			
Low			

Impact

Owner: Jeff Broughton, Systems Department Head

Trigger: Oracle stops support and/or makes pricing infeasible

Strategy: Mitigate

Description

Lustre is a third-party product that is vital to the successful operations of both Franklin and Hopper. With the corporate transition of ownership of Lustre from Sun to Oracle, there are potential concerns that Oracle will make changes that negatively impact Cray and NERSC's use of Lustre for the NERSC systems. The possible changes include: stopping development, dropping support, unfavorable pricing, licensing terms that prevent resellers like Cray from bundling Lustre with their operating system. Oracle has announced there will be no further development for Lustre versions 1.6 and 1.8. Further, support for Lustre 1.6 ended June 2010, and support for Lustre 1.8 ended June 2011. Oracle will focus on development of Lustre 2.0 on Oracle supported hardware. In

February 2011, Oracle announced that it would not pursue additional software releases for Lustre and that it would eventually terminate third-party support.

Mitigation

Three independent community development organizations have emerged to further Lustre development, and multiple companies have announced support for Lustre.

Contingency

If a vendor should cease business, NERSC can always operate and maintain the resource with in-house expertise while alternatives can be procured or developed.

- Rely on Cray or other third parties to support Lustre
- Rely on Data Virtualization Service (DVS) to use alternate file systems if needed
- Evaluate open-source Lustre

Status

As of spring 2011, the three community organizations have merged efforts, further reducing the likelihood that Lustre will be orphaned or that code bases will be split. The primary concern at this point is that commercially viable support organizations emerge. NERSC continues to be actively involved in the Lustre community.

5.6.7 Failure of Largest System

Probability			
High			
Med			
Low		X	
	Low	Med	High
	Impact		

Owner: Jeff Broughton, Systems Department Head

Trigger: Inability to meet contractual and DOE performance and availability metrics

Strategy: Mitigate

Description

The largest production computer system (NERSC-6) fails to operate adequately. Planning has begun for the next large production system (NERSC-7). The older system, NERSC-5, will be retired in 2012. When NERSC-5 is retired, NERSC-6 will be the only large production system until NERSC-7 comes online.

Mitigation

Availability and Mean Time to Repair are continuously tracked. Dual filesystems and redundant login nodes are configured. Hardware and software changes are tested on a test system before being installed on the production system.

Contingency

Reconfigure NERSC-6 as multiple smaller systems. Use redundant file systems and/or login nodes. Revert to previous software versions and/or hardware configurations. Direct NERSC users to the ALCF and the OLCF.

Status

The previous largest system, Franklin, will be retired on April 30, 2012.

5.7 Major Risks Next Year, with Mitigations

These are the risks we expect to be the most actively tracked in 2012.

Risk	Section	Risk Rating	Main Mitigations
Budget Insufficient to Meet DOE (Allocation) Commitments	5.3.1	High	Since the occurrence of this risk is controlled by DOE, the risk is accepted. Mitigation strategies are implemented by DOE.
Loss of Key Personnel (Temporary or Permanent)	5.3.3	Medium	Frequent communication between managers and staff. Cross-training.
CRT Facility Occupancy Delayed	5.6.3	Medium	Monthly construction review meetings.
Annual Budget Plan under Continuing Resolution	5.6.1	Medium	Plan cash flow for a prolonged Continuing Resolution. Monitor spending and communicate with DOE.
Computer Room Safety	5.3.7	Low	Test safety systems and perform preventive maintenance. Review and update required safety training and procedures.
NERSC Fails to meet IPv6 compliance for NERSC systems and/or networks	5.6.5	Low	Procure and test IPv6 systems, which provide public services early in FY12.
Failure of largest system	5.6.7	Low	Availability and Mean Time to Repair are continuously tracked. Dual filesystems and redundant login nodes are configured. Hardware and software changes are tested on a test system before being installed on the production system

6 Summary of the Proposed Metric Values for Future OARs

Charge question 6: Are the performance metrics for the next year proposed by the facility reasonable?

NERSC Response: See detailed response below.

The proposed metrics for future OARs are outlined in Table 6-1, followed by a discussion of the rationale and use of the proposed metrics and targets.

Table 6-1. Proposed Metrics for Future OARs (1 of 2)

Metric Description	2011 Target	2011 Actual	2012 Target	Proposed 2013 Target
User Survey Metrics	**Key average scores on the user survey are satisfactory (> 5.25 / 7)**			
Overall Satisfaction with NERSC Average Score on User Survey	> 5.25	6.54	> 5.25	> 5.25
Average of All High-Level ("Overall Satisfaction") Scores on the User Survey	> 5.25	6.39	> 5.25	> 5.25
Average of All User Survey Scores	> 5.25	6.29	> 5.25	> 5.25
Average of All User Support (Services) Survey Scores	> 5.25	6.37	> 5.25	> 5.25
Annual user survey results show improvement in at least one-half of the questions that scored below satisfactory (5.25/7) in the previous year	Two scores were < 5.25 in 2010. At least one of these must be ≥ 5.25 in 2011.	The two scores that scored < 5.25 in 2010 were both > 5.25 in 2011.	Results will show improvement in at least one half of questions that scored below satisfactory (5.25) in the previous period.	Results will show improvement in at least one half of questions that scored below satisfactory (5.25) in the previous period.
Consulting Overall Average Score on User Survey	> 5.25	6.58	> 5.25	> 5.25
Consulting Response Time Average Score on User Survey	> 5.25	6.57	> 5.25	> 5.25
Quality of Technical Advice Average Score on User Survey	> 5.25	6.53	> 5.25	> 5.25
Consulting Time to Solution Average Score on User Survey	> 5.25	6.40	> 5.25	> 5.25
Problem Resolution Metric				
Eighty percent of user problems are addressed within three working days (resolved or user informed how problem will be handled)	≥ 80%	86.9%	≥ 80%	≥ 80%

Table 6-1. Proposed Metrics for Future OARs (2 of 2)

Metric Description	2011 Target	2011 Actual	2012 Target	Proposed 2013 Target
Scheduled Availability Metrics	**Scheduled availability is ≥ 90% for systems in their first year of production, and 95% thereafter.**			
Franklin Scheduled Availability	≥ 95%	96.4%	≥ 95%	N/A
Hopper Scheduled Availability	≥ 90%	98.7%	≥ 95%	≥ 95%
Carver Scheduled Availability	≥ 93%	99.9%	≥ 95%	≥ 95%
Euclid Scheduled Availability	≥ 93%	99.9%	≥ 95%	≥ 95%
HPSS User System Scheduled Availability	≥ 95%	99.8%	≥ 95%	≥ 95%
Global Filesystem Scheduled Availability	≥ 95%	99.6%	≥ 95%	≥ 95%
2012 Overall Availability Metrics	**Overall availability is ≥ 85% for systems in their first year of production, and 90% thereafter.**			
Hopper Overalll Availability	≥ 80%	97.6%	≥ 90%	≥ 90%
Carver Overall Availability	≥ 85%	98.8%	≥ 90%	≥ 90%
Franklin Overall Availability	≥ 85%	95.5%	≥ 90%	N/A
Euclid Overall Availability	≥ 85%	99.4%	≥ 90%	≥ 90%
HPSS User System Overall Availability	≥ 85%	98.9%	≥ 90%	≥ 90%
Global Filesystem Overall Availability	≥ 85%	99.0%	≥ 90%	≥ 90%
Capability Metric				
On NERSC's largest machine in production, at least *N* percent of the cycles should be used by jobs running on 1/8th or more of the processors, where *N* is determined annually by the NERSC Program Manager.	On Franklin, at least 35% of cycles should be used by jobs running on 4,096 or more processor cores.	49.0%	On Hopper, at least 25% of cycles should be used by jobs running on 16,384 or more processor cores.	On Hopper, at least 25% of cycles should be used by jobs running on 16,384 or more processor cores.

6.1 Rationale and Use of Proposed Metrics and Targets

6.1.1 User Survey Metrics

The annual NERSC user surveys provide feedback about every aspect of NERSC's operation, help us judge the quality of our services, give DOE information on how well NERSC is doing, and point us to areas where we can improve.

About 100 satisfaction questions are rated on a 7-point scale. For each question, users are asked "How satisfied are you?" They select from the following menu of answers:

Menu choice	Corresponding Numeric Score
Very Satisfied	7
Mostly Satisfied	6
Somewhat Satisfied	5
Neutral	4
Somewhat Dissatisfied	3
Mostly Dissatisfied	2
Very Dissatisfied	1

NERSC and its Program Manager have agreed that a satisfactory score is 5.25, a bit above the "Somewhat Satisfied" level. Target average scores for key questions or groups of questions on the survey must be at or above this satisfactory level. NERSC pays particular attention to scores that are not satisfactory, and describes in the OA report the actions that are proposed to address them.

Each year NERSC assesses the areas that the users raise the most concerns about in the comments section. We incorporate this feedback into our planning process and evaluate potential changes that can address these areas of concern.

NERSC also uses the survey as a way to communicate with users, and sends individual responses to users if warranted.

6.1.2 Problem Resolution Metrics

The metric "Eighty percent of user problems are addressed within three working days" is one way of measuring responsiveness to the problems users submit to us. NERSC receives all sorts of problem reports from users, from the immediately resolved (e.g., password reset request) to ones that may take a few days to resolve (e.g., very large disk quota requests, which need to be carefully assessed) to longer-term requests (e.g., new software may take weeks to install; code bug reports may take days to assess whether it is a user or software problem, and if it is a software problem, may take the vendor months to fix). The "80% addressed metric" attempts to be a fair assessment of how responsive NERSC is in responding to this broad range of user problems.

NERSC has other ways to ensure that user questions are addressed promptly. Escalation rules are built into the problem ticket system so that staff can receive alerts about tickets that need a response. We also ask about problem resolution satisfaction on the user survey.

6.1.3 Availability Metrics and Reported Numbers

Availability metrics allow us to set expectations for system availability and to track how well we are doing in meeting these expectations. They are monitored and briefly reviewed weekly by on-call personnel to identify trends and identify remedies (e.g., software upgrades to eliminate bugs, changes in procedures, hardware component replacements). In addition, the availability metrics are reviewed several times a year for improving our system maintenance plans. These reviews help us determine whether downtimes are effective, current system architecture is adequate, or operational expenses need adjustment. This includes discussions with the system vendors to help tune the system architecture and to ensure that vendor deliverables are being met.

Scheduled availability is a measure of system reliability from the user's point of view. It tracks availability, taking into account only unscheduled outages. Overall availability is a measure of how available the system is overall to users; it takes into account both scheduled and unscheduled outages.

A new system often needs a shakeout period to address issues that occur early in its lifecycle, e.g., delivery of faulty hardware or new software that shows problems only when many users and codes are running. Although much of this is detected and addressed during pre-production, it is expected that some software problems will not be caught during pre-production. Thus systems in their first year tend to have more scheduled maintenances after software and hardware installation. This is why both scheduled and overall availability targets are lower for systems in their first year of operation than for subsequent years.

For 2012 NERSC has the following availability targets:

- Scheduled availability is ≥ 90% for systems in their first year of production, and 95% thereafter.
- Overall availability is ≥ 80% for systems in their first year of production, and 95% thereafter.

Mean Time To Interrupt (MTTI) and Mean Time To Failure (MTTF) results are reported numbers, not metrics with targets. MTTI tells us how long users can expect the system to stay up. It is often dominated by the frequency of scheduled downs. On computational systems, it also tells us what the maximum wall time limits can be for batch jobs. MTTF is a measure of system and environment stability.

6.1.4 Capability Metric

The "capability percent" is a measure of what percent of total batch job time is being used to run jobs at scale. NERSC and its Program Manger have to carefully assess the needs of a broad range of users when defining what "large scale" means and when setting the target percent of job time that should be at scale. NERSC serves many "capability users," whose codes can efficiently run at scale and who have enough allocation to run at scale. However, NERSC does not have sufficient computing capability to allow DOE SC to provide very large allocations to more than several dozen of the 600 or so allocations it makes yearly. Since NERSC must meet the needs of all DOE SC users and their wide range of computing needs, it makes sense that NERSC's capability size (about $1/8^{th}$ of machine size)and target percent (usually in the 25% to 35% range) are lower than those at the leadership centers. NERSC analyzes and projects job usage data every year, and together with its Program Manager sets the appropriate capability target for the next year.

DISCLAIMER

This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor The Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or The Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof or The Regents of the University of California.