NERSC 2016: Extreme Computation and Data for Science[1]

William T.C. Kramer
NERSC Facility
kramer@nersc.gov

Overview

By the year 2016, scientific computing will see a continued exponential increase in computational power. Simulations at or near exascale level will be conceivable in a growing number of scientific frontiers. HPC facilities in 2016 will be dealing with an exponential increase in experimental and simulation data. Scientific discovery will increasingly be based on the creation, maintenance, and analysis of exa- to zetabyte data repositories that need to be stored, accessed, analyzed, processed, shared, and understood. Analytics (the techniques and technology in data analysis, visualization, analytics, networking, and collaboration tools) will be essential in data-rich scientific applications.

As DOE's *Keystone* high performance computing and stroage facility with the mission to accelerate the pace of scientific discovery by providing high performance computing, information, data, and communications resources for *all* open applied and basic science and engineering sponsored by the DOE Office of Science, NERSC 2016 will provide unique resources and assistance to the open science community to enable community members to make effective use of exascale HPC resources and data.

To meet the computational challenge, NERSC will field a series of early production systems that increase performance for science. To meet the data challenges, NERSC envisions hosting community data repositories with integrated information management and analytics tools and services, as well as new exascale storage technologies. To meet the ever increasing electrical power demands of ultra scale computers NERSC will explore new, tightly coupled hardware/software designs that emulate the achievements of the low-power embedded computing market to result in computers that meet the computational needs of scientific applications while showing a dramatic increase in energy efficiency.

This paper describes NERSC's approach to achieving this vision for two if the three areas — Computing and Data. NERSC's approach to ultra efficient computing is documented in other papers[1] and not discussed further here.

NERSC Computational Services

The Keystone Facility

The need for increases in computational resources between today and 2016 is well documented in the DOE Greenbook,[2] the SCaLeS Report,[3] and the E3 Report.[4] The scientific requirements go beyond the traditional Office of Science work that NERSC has supported for the past 38 years, adding untouched areas of life science, energy resources (nuclear, biofuels, and renewable), energy efficiency, climate management, nanotechnology, and knowledge discovery. Simulations will grow in complexity, spatial resolution, timescales, ensemble sizes, and data assimilation (Figure 1). The computational needs are far beyond what can be supplied today by NERSC alone, NERSC combined with the other DOE centers, or even an augmented NERSC with other DOE centers.

[1] 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.

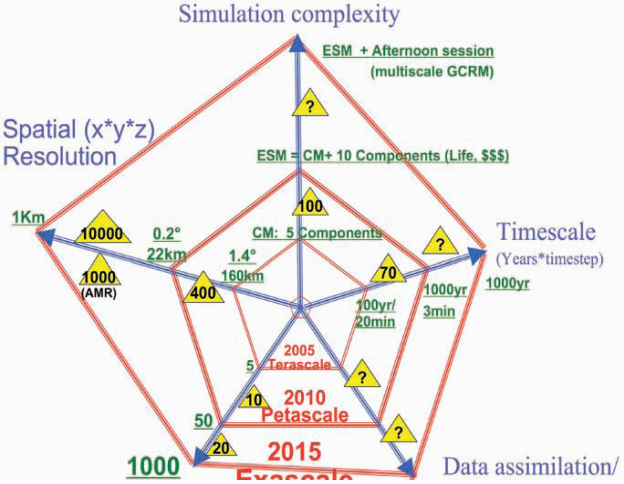

Figure 1. Investment of exascale and petascale computational resources in several aspects of a simulation: spatial resolution, simulation complexity, ensemble size, etc. Each red pentagon represents a balanced investment at a compute scale for climate science. (Image from E3 Report.[4])

NERSC 2016 will strive to exceed the vision of the DOE *Facilities for the Future of Science* plan. Priority Number 7 in that plan states:

> … NERSC … will … deploy a capability designed to meet the needs of an integrated science environment combining experiment, simulation, and theory by facilitating access to computing and data resources, as well as to large DOE experimental instruments. NERSC will concentrate its resources on supporting scientific challenge teams, with the goal of bridging the software gap between currently achievable and peak performance on the new terascale platforms.[5]

Rather than terascale platforms, NERSC 2016 will be operating near-exascale platforms. The Number 2 DOE facilities priority states:

> The USSCC [Ultra Scale Scientific Computing Capability, now known as the Leadership Computing Facilities or LCFs], located at multiple sites, will increase by a factor of 100 the computing capability available to support open … scientific research—reducing from years to days the time required to simulate complex systems, such as the chemistry of a combustion engine, or weather and climate—and providing much finer resolution.[6]

NERSC will meet this goal by supporting selected leadership computing projects represented today as the DOE INCITE program which was first proposed by and implemented at NERSC, and by closely cooperating with other DOE facilities to provide long-term data, software, and scaling support.

NERSC's role in this environment is to provide exceptional resources and quality of service for

- *high-impact* computing (defined below)
- *broad-impact* computing for the diversity of DOE mission science
- efficient and transparent *access* to and *management of* simulation and experimental *data*
- integrated data *analysis* tools and platforms
- *integrated support* for SciDAC and other science community-developed tools
- *outreach* to new HPC user communities.

Of course, scientists need more than floating point operations (flops). They need balanced computational system and service architectures that achieve high marks on the *PERCU* metrics:

- Performance — How fast will a system process work if everything is working well?

- Effectiveness — What is the likelihood that users can get the system to do their work when they need it?
- Reliability — How often is the system's available to do work and operate correctly?
- Consistency — How often will the system process users' work at the optimal performance?
- Usability — How much effort is it for users to get the systems to go as fast as possible?

A Diverse Scientific Workload

NERSC 2016 must support a diverse scientific workload that requires large-scale computation, data, and network resources. Many users need resources spanning three or more orders of magnitude. NERSC 2016 must continue to support both *high-impact* and *broad-impact* science workloads. High-impact work is ultrascale workflows/applications that require 20–100% of the largest resources at any given time. Broad-impact work is science that runs at scale, with high throughput, using 1–20% of the resources at a given time. NERSC expects to support between 10 and 20 high-impact science projects and from 200 to 250 broad-impact projects. In addition to broad-impact and high-impact science, NERSC 2016 will support three to five INCITE-like *"breakthrough science*" projects a year. These projects receive preferred processing and services from NERSC in order to meet their milestones.

Balanced System Architecture

The driving force for Exa-scale computational systems is changing, as expressed Table 1 – The Computational Common Wisdom.

Table 1
Computational Common Wisdom

Old	New
1. Performance Per Processor increase so improving science just waits for Moore's Law	1. Performance increases as the number of cores increases and mutli-core impacts all work.
2. Most of the systems cores are in processors.	2. Processors are essentially free – memory and TCO dictate costs.
3. Performance is limited by bandwidth	3. Performance is limited by latency
4. Science projects can use one system for every step of their workflow.	4. Job steps are best run on systems with the most appropriate balance.

Today, NERSC has about ~12,000 nodes and will soon be over 40,000 cores – in 5-6 hybrid[7] architectures. We have ~1 Petabtyes of on-line storage and 4-5 Petabytes of near-line storage in active use. Codes run from 500 to 16,000 cores with multiple I/O Modes - one file per core; one file per application - therefore one file for many cores and many core and many files - all from many processors with a fair amount of random I/O. NERSC services 3,100 uses, 350 projects, 700-900 major codes. In the past fours years, we have successfully helped the science community increase in scalability by 10X for "low hanging" applications.

NERSC 2016 will very likely have the following characteristics. Physically, we will have ~100,000 to 150,000 nodes and ~10,000,000 cores - mostly in a hybrid architectures. There will be 2 to 10 discrete systems with 1.5 PB/s aggregate BW (.5-2 GByes per core), and 10-100 Exabtyes of on-line storage. All systems and users have fair and equivalent access. There will be 100-1,000 Exabytes to near-line storage.

Exa-scale codes will run from 100,000-1,000,000 cores with more often than not one file from many, many cores using file parallel I/O format libraries such as HDF5. There will probably be 4,000 uses, 400 projects, 1,200 major codes, all of which will need to increase 10-100x in scalability from today.

	Multi-physics, multi-scale	Dense linear algebra (DLA)	Sparse linear algebra (SLA)	Spectral methods (FFTs) (SM-FFT)	N-body methods (N-Body)	Structured grids (S-Grids)	Unstructured grids (U-Grids)	Data Intensive (Map Reduce)
Nanoscience	X	X	X	X	X	X		
Chemistry	X	X	X	X	X			
Fusion	X	X	X			X	X	X
Combustion	X		X			X	X	X
Astrophysics	X	X	X	X	X	X	X	X
Biology	X	X					X	X
Nuclear		X	X		X			X
System balance implications	General-purpose balanced system	High-speed CPU, high flop/s rate	High perform-ance memory	High inter-connect bisection bandwidth	High perform-ance memory	High-speed CPU, high flop/s rate	Irregular data and control flow	High storage and network bandwidth/ low latency

Figure 0-1: Characteristics of scientific discipline codes.

While it is possible to create specialized system solutions for a single algorithmic approach, Table 2 shows it is not feasible to segregate system balance by discipline, since different workflow steps and/or approaches within a discipline require different system balance. Often these different algorithms exist in the same codes. Likewise, a system balanced for just one computational approach cannot fully serve a discipline area alone.

Table 2 shows the general characteristics of the discipline codes and their implications for system balance. Since NERSC will support a diverse workload as indicated in Table 2, resources can be provisioned in one of two ways. First, provide general-purpose systems that are optimized to do well with the entire or a large segment of the workload. Second, provide a small number of specialized systems that are very efficient at particular algorithms and in aggregate support the entire diverse workload. NERSC is open to either approach or one that balances a portfolio of general and special systems that overall provides a very efficient facility for its entire workload space. In either case, a key requirement is easing the burden on scientists to move data between systems. In all likelihood, NERSC will have some large general-purpose systems and a few specialized systems for breakthrough science areas or specific algorithmic needs.

NERSC's general large-scale systems will provide an appropriate relationship between sustained performance, usable memory, and usable disk space. The general-purpose system ratio of usable memory to *sustained performance* is between 2 and 4 bytes/flop. The general-purpose system ratio of *usable*

disk to *sustained performance* is between 20 and 70 bytes/flop. The interconnect performance and latency is sufficient to provide cost-effective sustained performance on a representative workload.

Regardless of whether the general-purpose or the specialized system approach is taken, a key system architecture component that will make scientists succeed will be having a high performance, parallel, facility-wide file system tightly integrated with NERSC large-scale systems. Balanced system architectures also include:

- High performance local-area network
- Wide-area network interfaces matching or exceeding ESnet backbone speeds
- Archival storage with
  - large-scale near line data repository
  - online data cache
- Data focused systems
  - community data services: "Google for Science" (described below)
  - visualization and analysis
- Servers and specialized systems
- Infrastructure special-arrangement systems
- Cyber security
- Advanced concept systems.

NERSC 2016 will be open to deploying systems from multiple vendors, with major systems arriving at three-year intervals. At least two major systems (NERSCN and NERSC N+1) will be on the floor at any given time, one providing a stable platform while the next-generation system is brought into production. There are a number of reasons why a multi-system, possibly multi-vendor approach benefits NERSC users and DOE:

- It makes NERSC a fair broker to engage the entire vendor community in improving systems. Vendors are more likely to engage with NERSC if there is a fair chance their technology will be included in the facility in the foreseeable future. NERSC has demonstrated that it is possible to foster long-term, in-depth relationships with multiple vendors at the same time.
- It is likely the strengths of different vendor systems provide optimal platforms for different parts of the NERSC workload.
- It mitigates risk for NERSC and DOE in that the facility is less prone to experiencing a severe impact if one vendor has a problem with their technology or business roadmaps.

While open to multiple vendors at any given time it is not a goal in its own right. NERSC will always evaluate the alternatives for each major system enhancement independently, with the objective to provision the systems that provide the best overall value for the scientific workload

Balanced Service Architecture

NERSC will provide and support a robust code development and tuning environment, including advanced scientific and data libraries, code management tools, debugging and performance interfaces and tools, web services and selected applications that span many application projects.

NERSC will provide flexible and high-quality support services across the entire range of science and resources. This includes providing:

- expert assistance and advice on parallel computing, compilers, libraries, programming models, MPI, I/O, performance analysis, debugging, and HPC software
- expert large-scale system management, including expertise in HPC storage, operating systems, scheduling, distributed computing, networking, and security.

Because NERSC does not pick the science areas, problems, or users that run at NERSC, these services will be *flexible* so the expertise is valuable regardless of science discipline or code. Service architecture has to scale to new usages, new systems, and new projects.

Metrics have always been an important tool for assessing and improving NERSC's operations, and the facility will continue to use metrics to ensure that it is meeting the scientific needs of the United States. Control metrics will be:

1. number of computational hours delivered
2. system reliability and availability
3. customer service — survey and problem resolution
4. security of NERSC systems.

NERSC's Data Services

Virtually all fields of scientific endeavor base hypothesis testing on data analysis. Scientific disciplines vary in how they produce data (via observation or simulation), in how they manage data (storage, retrieval, archiving, indexing, summaries, sharing across the science team), and in how they analyze data and communicate results. It is widely agreed that one of the primary bottlenecks in modern science is managing and discovering knowledge in light of a deluge of data resulting from increasing computational capacity and the increasing fidelity of scientific observational instruments.[8] Further, as data becomes too large to move, we are evolving towards a model where data-intensive services are centrally located.[9] These services span a diverse set of activities that form the basis of the future NERSC Data services, including but not limited to: community-oriented data repositories; browsing, exploration, and analysis capabilities that operate on the centrally located community repositories; and providing and maintaining the centrally located hardware and software infrastructure that enables these capabilities.

Areas of Data Science

Today, a number of science areas are already struggling with the deluge of data. It is useful to look at few examples.

- The US Effort for the Climate Science, during the latest IPCC 4th Assessment Report took about 100 FTE years of effort, and studied approximate 11,000 model years at ~160 KM resolution. This produced 110 TBs of primary data. The Exscale goal for climate is 1.5 km models, 24 TB per day and a total of 1 PB of on-line primary storage per year.
- Energy Science is posed for rapid expansion – in areas such as alternative energy, efficiency, combustion, fusion and fission. An example of just one experiment ITER – that will produce 1 TB of primary data per test shot, and plans to do 2,000 test shots per year.

Many other data deluge projects are planned or under way, including ESG, LHC, JDEM/SNAP, Planck, SciDAC Computational Astrophysics Consortium, and JGI. Failure to act decisively to address the data needs of science will cost potentially tens of millions of dollars in duplicated effort when scientific staff set up and administer their own clusters for doing community-based data management and analysis.

Easy Access to Data Accelerates Science

The value of accessing massive datasets with powerful analytic tools was illustrated in the 2005 National Institute of Standards and Technology (NIST) Open Machine

Translation Evaluation, which involved academic, government, and commercial participants from all over the world. Although it was Google's first time competing, their translation system achieved the highest scores in both Arabic- and Chinese-to-English translation, outperforming sophisticated rules-based systems developed by expert linguists.[11] Google used statistical learning techniques to build its translation models, feeding the machines billions of words of text, including matching pairs of human-translated documents.[12] In this case, Google, with more data, beat others with more expertise.

Similar results can be expected from applying advanced analytics tools to massive scientific datasets. Indeed, several projects at NERSC. One such project is the cosmic microwave background data analysis for the Planck satellite mission. The satellite will be launched in summer 2008, but the data production pipeline is already in place at NERSC (Figure 2). Access to both raw and processed data will be provided through a web portal for a remote community of thousands of users.

These examples can be summarized as:

- Community data repositories, with information management and analytics for data shared across communities
- Storage of large amounts of data for distributed communities of researchers
  - Storage/retrieval/sharing/searching
  - Analytics: analysis, visualization
  - Production workflow
- Production-quality information management and analytics software infrastructure.

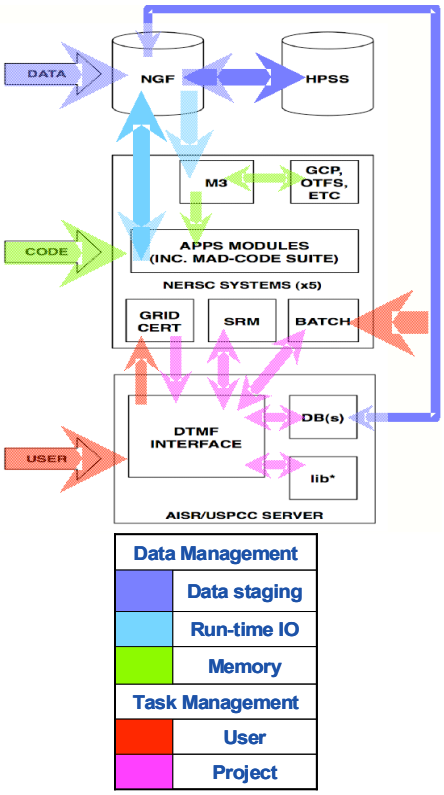

Figure 2. Planck production pipeline courtesy of Julian Borrill (LBNL)

NERSC Data Program Elements

To address the changing needs of its user communities, the NERSC Data services will include:

- The next-generation mass storage system
- Production infrastructure for data
  - Hardware: computational platforms
  - Software for data management, analysis/analytics, and interfaces between integrated data components
- Development or adaptation of reusable, broad-impact tools
  - Such as The Scientists Data Cart and Metadata Wizards discussed below
  - Analogous to Google Earth or Microsoft SharePoint
- Focused data project support
  - Consulting expertise in scientific data management, analytics, visualization, workflow management, etc.

NERSC Data Storage

NERSC is a founding development partner in the High Performance Storage System (HPSS) project.[13] HPSS is software that manages petabytes of data on disk and robotic tape libraries. It provides highly flexible and scalable hierarchical storage management that keeps recently used data on disk and less recently used data on tape. HPSS uses cluster, LAN and/or SAN technology as well as online and near line storage to aggregate capacity and performance of many computers, disks, and tape drives into a single virtual storage system of exceptional size and versatility.

While HPSS has been invaluable as a mass storage platform, it is now 15 years old and likely be very challenged to evolve to meet future science needs at the exa and zeta-byte scale. The current ways of storing data in global filesystems and archival storage systems will probably not scale to exascale. NERSC 2016 Data Services will evolve to address the new common wisdom for storage shown in Table 3

Table 3
Storage Common Wisdom

Old	New
1. Users have a small number of large files.	1. Large numbers of small files dominate performance.
2. Files are the lowest level unit of storage.	2. Objects are the lowest unit of storage.
3. Systems need to cause users "pain" to move their files from place to place in order to limit data growth.	3. It is more productive to systems and users to let systems manage the placement of files.
4. Users have all the files they need in each place they compute.	4. Users have data in many places and need to move the data frequently, even within a facility.
5. One system is sufficient for all the steps to a workflow.	5. Job steps are best run on systems with the most appropriate balance.

The open science community needs to initiate the collaborative development of what might be called EXA-HPSS — a next-generation mass storage system to serve science'ss future needs. This system must be energy efficient, scalable, closely integrated with parallel filesystems and on-line data, and designed for the requirements of new data profiles (e.g., the increasing importance of metadata). Archival storage needs to go beyond file-based access to support a broader set of data storage and retrieval operations and more user-friendly functionality.

With decades of experience serving a large communities of science users, NERSC envisions being one of the leaders for the specification, design, and research effort for a next-generation mass storage system, and to participate in R&D of an interface to support efficient use of the system. LBNL's Scientific Data Management Group is already initiating research into an energy-smart disk-based mass storage system, envisioned as an energy-efficient, low-latency, scalable mass storage system with a three-level hierarchy (compared to HPSS's two-level hierarchy). This effort takes aim at these facts: (1) disk storage accounts for about 27% of the total energy consumption in the nation's data centers; (2) the cost of powering the nation's data centers is about $4 billion per year and is expected to increase at a rate of about 25% per year; (3) research towards reducing power consumption at these centers is vital and mandated by legislation.[14] There is a body of work from DOE's Scientific Data Management SciDAC program that shows commercial RDBMS systems are not adequate to meet the needs of large, data-intensive science activities.[15]

NERSC Data Production Infrastructure

The NERSC Data production infrastructure will consist of computational platforms for high-capacity and high-throughput interactive analytics, a single namespace supported by ubiquitous, parallel and highly performant access from systems, high-capacity and energy-efficient mass storage, high performance intra- and inter-networking capability, and a robust collection of software tools. Software tools will include applications and libraries for data management, analysis, visualization, and exploration, as well as applications and libraries enabling scientific community access, e.g., web portal infrastructure, a new data archive interface, etc.

A good analogy for this infrastructure is Google and Yahoo, where significant investment in computational and software infrastructure enables the retrieval of data most relevant to a query from a variety of sources and presents it quickly in an easily comprehensible, usable, productive and navigable form. NERSC's long-term vision is to provide this type of on-demand capability to all users and stakeholders. The resulting solutions span a diverse range: community-centric data repositories and analysis, portal-based interfaces to data and computation, high performance and production-quality visual analytics pipelines/workflows and systems.

NERSC's Data Tools

"Google for Science" may become the next paradigm for scientific analytics if one considers the powerful capabilities that search engines put in the hands of anyone with Internet access. One of the keys tools that makes these capabilities possible is MapReduce, a programming model and an associated implementation for generating and processing large data sets.[16] MapReduce is used to regenerate Google's index of the World Wide Web as well as perform a wide variety of analytic tasks — more than ten thousand applications to date. The basic steps in MapReduce are:

- read a large quantity of data
- *map* the data: extract interesting items
- shuffle and sort
- *reduce:* aggregate and transform the selected data
- write the results.

MapReduce features that suggest it could be used very productively in scientific analytics, but it is a much different operational paradigm than HPC has used to date[2]. Its functions can be applied to numeric, image, or text data, e.g., simulations, telescopic images, or genomic data. Its simple, extensible interface allows for domain-specific analysis and leverages domain-independent infrastructure. It makes efficient use of wide area bandwidth by shipping functions to the raw data and returning filtered information. It hides messy details, such as parallelization, load balancing, and machine failures, in the MapReduce runtime library, allowing programmers who have no experience with distributed or parallel systems to exploit large amounts of resources easily.

Focused Data Projects

Exa-scale NERSC program will provide integrated, production-quality analytics pipelines for experimental and computational science projects in the following areas:

- **Data formats and models.** High performance, parallel data I/O libraries will optimize data storage, retrieval, and exchange on NERSC and other

[2] Mapreduce currently expects different paradigms for task scheduling, file systems, cyber security and resources management.

platforms. NERSC will evaluate these technologies, participate in efforts to create an improved technology, and directly consult with science projects to deploy these technologies.

- **Community-wide data access.** Science projects need straightforward, unfettered, yet authenticated and authorized access to their community data regardless of location across multiple sites. Access to data could potentially be based on files, like the current approach familiar to users of HPSS and other typical filesystems; or based on "objects," where the result of a "data gather" operation is performed by an agent on the user's behalf and later made available to the user.
- **Data filtering and processing.** Often raw data undergoes additional processing (e.g., gap filling, filtering, etc.) before being ready for downstream use by consumers. NERSC will assist its users in areas such as knowledge of the best algorithms for filtering/processing and their deployment on parallel machines, and assistance in deploying these algorithms in scientific workflows.
- **Data exploration.** Science users want to be able to quickly and easily explore their data, either with a traditional application that reads files and displays results, or through a web-based application that interfaces to back-end infrastructure to access and process data, then displays results.
- **Data analysis.** These activities include generating statistical summaries and analysis, supervised and unsupervised classification and clustering, curve fitting, and so forth. While many contemporary applications provide integrated data analysis capabilities, some science projects will want to run standalone analysis tools on data collections offline as part of a workflow to produce derived data for later analysis.
- **Data visualization.** The role of visualization and visual data analysis in the scientific process is well established.
- **Workflow management.** The goal of workflow management is to automate specific sets of tasks that are repeated many times and thus simplify execution and avoid typical human errors and increase scientific productivity that often occur when repetitive tasks are performed.
- **Interfaces and usability.** Recent production analytics workflows like Sunfall[18] show the dramatic increase in scientific productivity that results from careful attention to the combination of highly capable analytics software and highly effective interfaces to those software tools. Hence the objective is to increase scientific productivity for data-intensive activities through well designed and engineered interfaces.

Two Use Case Examples

In brief, two concepts illustrated with use cases that the Exa-scale scientist might use once a NESC-2016 infrastructure exists. The first is labeled *The Scientist's Data Cart* that is a data search and retrieval "Data Shopping" function providing fast, free form, search and query for complex tasks. Searches such as "Simulations using CCSM 3.0 with a resolution of T85 or greater run between January and May 2007" or "Supernova simulations for Type Ia US between January and May 2007" would be common.

Search results would be determined by user roles and results would put data "objects" into the user's "data cart". Carts will have suites of actions such as download, compare,

move for computation, visualize. Data carts would also have defined APIs so users can plug in their own actions for analyzing and manipulating the data. Similar functions will exist on the computational platforms in the form of shell scripts and tools for a complete environment.

Another conceptual use case is the *Meta Data Wizard* to assist data creation, importation and categorization. The Wizard would have default meta data defined including a set of data attributes (i.e. Originator, Source of data {simulation code, experiment, observation…), Sharing role {public, collaborative scientist, system-wide, project only, creator defined access control list, none), Formatting {flat, HDF5, netCDF…), etc.). Computational systems will have the ability to automatically annotate simulation data with data attributes. The metadata would be permanently associated with the file or object. Hopefully, standard methods would be used so data created at one site can be used at other sites

The *Metadata wizard* would annotate data whenever metadata is not already associated using both web and command line based. The Wizard would extract metadata automatically, and/or inform users that metadata needs to be supplied when an automated process is insufficient. Site specific context and application specific translation may be feasible.

Summary

NERSC is one of the world's largest, open, unclassified supercomputer centers. NERSC 2016 will continue providing some of the largest computational resources, but will also be one of the largest "data centers" in science for exa-scale computing , satisfying the production needs for computational science through:

- innovative approaches to cost-effective and energy-efficient exascale science
- comprehensive solutions to exascale data needs
- highly effective systems, outstanding support, and exceptional infrastructure for exascale science.

Exa-scale Computing will be accompanied with a transformation in many aspects of scientific computation where simulation is just as important as ever, but large scale data is just as important as simulation. Data is now intimately intertwined with simulation to give insight. Strictly commercially targeted system designs alone will likely not meet the scientific needs at the Exa-scale because the properties of such systems currently being built are not sufficient to support the expansion of scientific insight at the exa-scale. Futhermore, the technology to get the data in and out of subcomponents of large scale systems is lagging the increase in performance of such systems.

HPC will have an even greater impact on science and life at the Exa Scale if these challenges can be solved. Innovation must make key contributions since productivity and competitiveness are at stake. NERSC will be a leader in this innovation and the evolution to effective Exa-scale Science.

Acknowledgement

The vision expressed in this paper is an variation of a vision provided to DOE in September, 2008 in an internal, proprietary LBNL report LBID-2615 authored by Deborah A. Agarwal, Michael J. Banda, E. Wes Bethel, John A. Hules, William T. C. Kramer, Juan C. Meza, Leonid Oliker, John M. Shalf, Horst D. Simon, David Skinner, Francesca Verdier, Howard A. Walter, Michael F. Wehner, Katherine A. Yelick all of the NERSC and Computational Research Divisions at LBNL.

References

[1] M. Wehner, L. Oliker, J. Shalf, "Towards Ultra-High Resolution Models of Climate and Weather", International Journal of High Performance Computing Applications (IJHPCA), April, 2008.

[2] S. C. Jardin, ed., "DOE Greenbook: Needs and Directions in High Performance Computing for the Office of Science. A Report from the NERSC User Group." PPPL-4090/LBNL-58927, June 2005; http://www.nersc.gov/news/greenbook/2005greenbook.pdf.

[3] David E. Keyes, Phillip Colella, Thom H. Dunning, Jr., and William D. Gropp, eds., *A Science-Based Case for Large-Scale Simulation ("The SCaLeS Report"),* Washington, D.C.: DOE Office of Science, Vol. 1, July 30, 2003; Vol. 2, September 19, 2004; http://www.pnl.gov/scales/.

[4] H. D. Simon, T. Zacharia, R. Stevens, et al., "Modeling and Simulation at the Exascale for Energy and the Environment," Department of Energy Technical Report (2007); http://www.mcs.anl.gov/~insley/E3/E3-draft-2007-08-09.pdf.

[5] *Facilities for the Future of Science: A Twenty-Year Outlook,* U.S. Department of Energy Office of Science (November 2003), p. 21; http://www.sc.doe.gov/Scientific_User_Facilities/History/20-Year-Outlook-screen.pdf.

[6] Ibid., p. 15.

[7] Supercomputing: getting Up to Speed, a report of the National Research Council, 2006.

[8] Richard P. Mount, ed., The Office of Science Data-Management Challenge: Report from the DOE Office of Science Data-Management Workshops, March–May 2004; http://www.sc.doe.gov/ascr/ProgramDocuments/Final-report-v26.pdf.

[9] Gordon Bell, Jim Gray, and Alex Szaley, "Petascale Computational Systems," IEEE Computer 39(1), January 2006.

[11] NIST 2005 Machine Translation Evaluation Official Results, August 1, 2005, http://www.nist.gov/speech/tests/mt/2005/doc/mt05eval_official_results_release_20050801_v3.html.

[12] Bill Softky, "How Google translates without understanding," The Register, May 15, 2007, http://www.theregister.co.uk/2007/05/15/google_translation/print.html.

[13] http://www.hpss-collaboration.org/hpss/index.jsp

[14] Public law 109-431, Dec. 20, 2006, 109th Congress, http://www.uptimeinstitute.org/symp_pdf/Public Law 109-431.pdf.

[15] K. Wu, W.-M. Zhang, V. Perevoztchikov, J. Laurent, and A. Shoshani, "Grid Collector: Using an Event Catalog to Speed Up User Analysis in a Distributed Environment," presented at Computing in High Energy and Nuclear Physics (CHEP) 2004, Interlaken, Switzerland, September 2004; http://www.osti.gov/bridge/servlets/purl/882078-E3rSLU/882078.PDF.

[16] Jeffrey Dean and Sanjay Ghemawat, "MapReduce: Simplified Data Processing on Large Clusters," Proc. OSDI'04: Sixth Symposium on Operating System Design and Implementation, San Francisco, CA, December, 2004; http://labs.google.com/papers/mapreduce-osdi04.pdf.

[18] C. Aragon, S. Bailey, S. Poon, K. Runge, and R. Thomas, "Sunfall: A Collaborative Visual Analytics System for Astrophysics," IEEE Visual Analytics Science and Technology Conference, Sacramento, CA, Oct. 30–Nov. 1, 2007; http://vis.lbl.gov/Publications/2007/Sunfall_VAST07.pdf (abstract), http://vis.lbl.gov/Publications/2007/Sunfall_VAST07_poster.pdf (poster).