EOS
EOS, TRANSACTIONS, AMERICAN GEOPHYSICAL UNION
VOLUME 92 NUMBER 13
29 MARCH 2011
PAGES 109–116

Collaborative Supercomputing for Global Change Science

PAGES 109–110

There is increasing pressure on the science community not only to understand how recent and projected changes in climate will affect Earth's global environment and the natural resources on which society depends but also to design solutions to mitigate or cope with the likely impacts. Responding to this multidimensional challenge requires new tools and research frameworks that assist scientists in collaborating to rapidly investigate complex interdisciplinary science questions of critical societal importance. One such collaborative research framework, within the NASA Earth sciences program, is the NASA Earth Exchange (NEX). NEX combines state-of-the-art supercomputing, Earth system modeling, remote sensing data from NASA and other agencies, and a scientific social networking platform to deliver a complete work environment. In this platform, users can explore and analyze large Earth science data sets, run modeling codes, collaborate on new or existing projects, and share results within or among communities (see Figure S1 in the online supplement to this *Eos* issue (http://www.agu.org/eos_elec)).

Over the past 2 decades, climate science has evolved into a multidisciplinary field fusing observations from a tremendous constellation of subsurface, surface, air, and space-borne sensors with intensive computer modeling [*NASA*, 2010]. Collaborations among climate scientists, biologists, economists, and policy makers address a spectrum of societal impacts resulting from climate variability and change. Except for a few well-funded efforts that support climate informatics, the process of capturing, curating, analyzing, and visualizing climate data takes place ad hoc. Collaboration and sharing of tools and knowledge across the community remain an afterthought, mainly in the form of publication of the results. The exponential increase in the volume of data poses a serious challenge to climate science, and new approaches are needed to minimize potentially crippling inefficiencies and ensure continued scientific progress in the field. Of course, this predicament is not unique to climate science. In *The Fourth Paradigm: Data-Intensive Scientific Discovery* [*Hey et al.*, 2009], a tribute to the ideas of pioneering computer scientist James Gray, authors from various disciplines emphasize that further advances in the new era of data- and computer-intensive science depend on how well researchers collaborate with one another and with technologists in the areas of database architectures, workflow management, visualization, and cloud-computing technologies.

The NEX software framework is built upon the Terrestrial Observation and Prediction System (TOPS) [*Nemani et al.*, 2009], a data assimilation and modeling framework developed at NASA. TOPS provides many of the software libraries and tools required by NEX to automate the processing of satellite and climate data and manage workflows for data analysis and modeling studies. Providing data, software, and large-scale computing power together reduces the need for redundantly downloading data, developing data preprocessing software tools, building standard modeling and analysis codes, and expanding local computer infrastructures. Reducing this extra effort accelerates fundamental research and the development of new applications and potentially reduces overall costs associated with scientific endeavors.

Additionally, the social networking platform provides a forum for NEX members to connect and then initiate the exchange of complete work environments that may contain data sets, results, algorithms, and codes. Using a permissions control mechanism, each member may point other members to their NEX work environment and allow collaborators to evaluate or extend their work. Complete NEX user environments may be shared without the transfer of large volumes of data or the porting of complex codes, making NEX an ideal platform for building upon and exchanging research and fostering innovation.

NEX Resources

NEX leverages the NASA Advanced Supercomputing (NAS) facility at NASA Ames Research Center, Moffett Field, Calif., which has up to 100,000 Intel cores with a peak performance of 1.009 petaflops for the computer infrastructure. An expandable "sandbox" cluster is dedicated to prototype development, basic analysis and visualization, work environment creation, and collaborative outreach. Information on how to access NEX can be found in the online supplement.

NEX hosts a large collection of data from satellite sensors such as advanced very high resolution radiometer (AVHRR), Moderate Resolution Imaging Spectroradiometer (MODIS), and Landsat, as well as global data sets of surface weather records, topography, soils, land cover, global climate simulations, and regionally downscaled climate projections. These data sets are held together with a collection of modeling codes (climate, weather, ecosystem, hydrology), data processing utilities, database management systems, and analysis and visualization toolboxes. Software components are provided in the customizable work environments, which utilize hardware virtualization, enabling each user to create a virtual machine that is populated with specific data sets, models, and tools aimed at maximizing the ease of use and portability. These virtual work environments can be migrated between computing resources to ensure efficient use of available high-end computing resources. Using enhancements to the Red Hat Makara cloud application platform tools developed under this project, the custom environment can also be saved as a complete "snapshot" of a research project, archived, searched, and reloaded at any time in the future to re-create the results of an analysis or be a starting point for further scientific advancements.

NEX fosters collaboration through a Web portal as well as via audiovisual conferencing. The portal (http://nex.arc.nasa.gov) provides an intuitive way for researchers to collaborate on projects and contribute knowledge to the NEX platform. The scientific social networking platform, similar to many of the social platforms found on the Web today, gives scientists and researchers a meeting ground to coordinate the sharing of software tools, algorithms, and data sets so that the knowledge and utilities can be reused and built upon by other members. Adobe® Connect™ Web conferencing tools are integrated into the social networking platform, allowing members within a collaborative project to conduct webinars, share documents and presentations, and work together on a daily basis as a fully distributed virtual team. Finally, a 128-screen hyperwall, capable of rendering one quarter of a billion pixels, is available to NEX members for scientific visualization and as a data exploration environment.

NEX Expected Outcomes

Whether it is hundreds of researchers downloading MODIS data from Distributed Active Archive Centers (DAAC) and writing redundant code to analyze data, or dozens of cities around the world each conducting its own analysis of vulnerability to climate change, the problem is the same: There is very little sharing of software code or workflows [*Barnes*, 2010]. Allowing NEX members to instantiate work environments on a large computing infrastructure that is directly connected to large data archives may significantly reduce both the costs and time associated with research and development efforts by relieving researchers of redundantly retrieving and integrating data sets and building modeling analysis codes.

For example, in early 2010 a team of researchers from around the country came together on NEX to mosaic and atmospherically correct 9000 Landsat thematic mapper scenes to retrieve the leaf area index of vegetation globally at 30-meter resolution. The entire processing of nearly 340 billion pixels took less than 30 minutes on NEX, allowing the team to experiment with new algorithms and products in the space of a few days.

The NEX environment, with its support for broader community participation, will benefit community scientific efforts such as community modeling, where complex modeling codes are designed, built, and tested; model intercomparison studies, where various models are tested against each other and/or against observations; production and maintenance of long-term data records, which are vital for monitoring climate and ecosystems; and national assessments, in which scientists from a variety of disciplines work together to analyze, summarize, and report on the current state of affairs, for example, as in the case of climate impacts.

An additional feature of NEX is to link results published in scientific literature with data and computing to promote collaboration and ensure reproducible results in the current data-intensive era. When researchers are willing to share workflows along with their publications, tracing a publication to its data (including relevant literature on the data and analysis techniques) has the potential to allow other scientists to verify, modify, or enhance the workflows and quickly improve or extend published results. Open workflows squarely address the issue of scientific integrity, which has become an important component of conveying results from climate science and maintaining public confidence in scientific analyses of climate change.

As the development of NEX continues, it will lower the barrier of entry to data- and computer-intensive science and provide a mechanism for continuous engagement among members of the global change science community.

References

Barnes, N. (2010), Publish your computer code: It is good enough, *Nature, 467,* 753, doi:10.1038/467753a.

Hey, T., S. Tansley, and K. Tolle (Eds.) (2009), *The Fourth Paradigm: Data-Intensive Scientific Discovery*, 286 pp., Microsoft Res., Redmond, Wash.

NASA (2010), Responding to the challenge of climate and environmental change: NASA's plan for a climate-centric architecture for Earth observations and applications from space, report, 48 pp., Washington, D. C.

Nemani, R., H. Hashimoto, P. Votava, F. Melton, W. Wang, A. Michaelis, L. Mutch, C. Milesi, S. Hiatt, and M. White (2009), Monitoring and forecasting ecosystem dynamics using the Terrestrial Observation and Prediction System (TOPS), *Remote Sens. Environ., 113,* 1497–1509, doi:10.1016/j.rse.2008.06.017.

—R. Nemani, NASA Ames Research Center, Moffett Field, Calif.; E-mail: rama.nemani@nasa.gov; and P. Votava, A. Michaelis, F. Melton, and C. Milesi, California State University, Monterey Bay, Seaside

---

Great Lakes Literacy Principles

PAGES 109–110

Lakes Superior, Huron, Michigan, Ontario, and Erie together form North America's Great Lakes, a region that contains 20% of the world's fresh surface water and is home to roughly one quarter of the U.S. population (Figure 1). Supporting a \$4 billion sport fishing industry, plus \$16 billion annually in boating, 1.5 million U.S. jobs, and \$62 billion in annual wages directly, the Great Lakes form the backbone of a regional economy that is vital to the United States as a whole (see http://www.miseagrant.umich.edu/downloads/economy/11-708-Great-Lakes-Jobs.pdf). Yet the grandeur and importance of this freshwater resource are little understood, not only by people in the rest of the country but also by many in the region itself.

To help address this lack of knowledge, the Centers for Ocean Sciences Education Excellence (COSEE) Great Lakes, supported by the U.S. National Science Foundation and the National Oceanic and Atmospheric Administration, developed literacy principles for the Great Lakes to serve as a guide for education of students and the public. These "Great Lakes Literacy Principles" represent an understanding of the Great Lakes' influences on society and society's influences on the Great Lakes.

Eight Key Facts to Guide Public Understanding

The Great Lakes literacy effort had its origins in Ocean Literacy (http://www.oceanliteracy.net), a movement by hundreds of scientists and educators who developed a concise framework for conveying the most important ocean science concepts that all students in grades K-12 should know.

With support from Ocean Literacy proponents in COSEE California, COSEE Great Lakes leaders examined the Ocean Literacy and fledgling Lake Erie Literacy Principles and Concepts (www.ohiodnr.com/LakeErieLiteracy/) and drafted a baseline set appropriate for the region. About 80 scientists and educators from around the Great Lakes region reviewed and edited the list through multiple drafts, and the Great Lakes Literacy Principles emerged.

The Great Lakes Literacy Principles are as follows:

1. The Great Lakes, bodies of fresh water with many features, are connected to each other and to the world ocean.
2. Natural forces formed the Great Lakes; the lakes continue to shape the features of their watershed.
3. The Great Lakes influence local and regional weather and climate.
4. Water makes Earth habitable; fresh water sustains life on land.
5. The Great Lakes support a broad diversity of life and ecosystems.
6. The Great Lakes and humans in their watersheds are inextricably interconnected.
7. Much remains to be learned about the Great Lakes.
8. The Great Lakes are socially, economically, and environmentally significant to the region, the nation, and the planet.

These principles are designed to be crisp and quotable. Numbering provides shorthand for comparison with similar principles of ocean literacy and for easy discussion among users.

Using the Literacy Principles: Lesson Plans, Textbooks, and Tests

To enable use of the Great Lakes Literacy Principles, Ohio Sea Grant hosts an Internet resource (http://greatlakesliteracy.net) for Great Lakes literacy subject matter, curricula, and other resources. At this site, fundamental concepts within each principle are defined, accompanied by relevant examples and a description of each principle's scope. Educators and scientists can contribute annotated links and other resources useful for Great Lakes literacy—site managers are particularly interested in information that compares the Great Lakes to equivalent characteristics of the ocean. The site currently links to lesson plans,

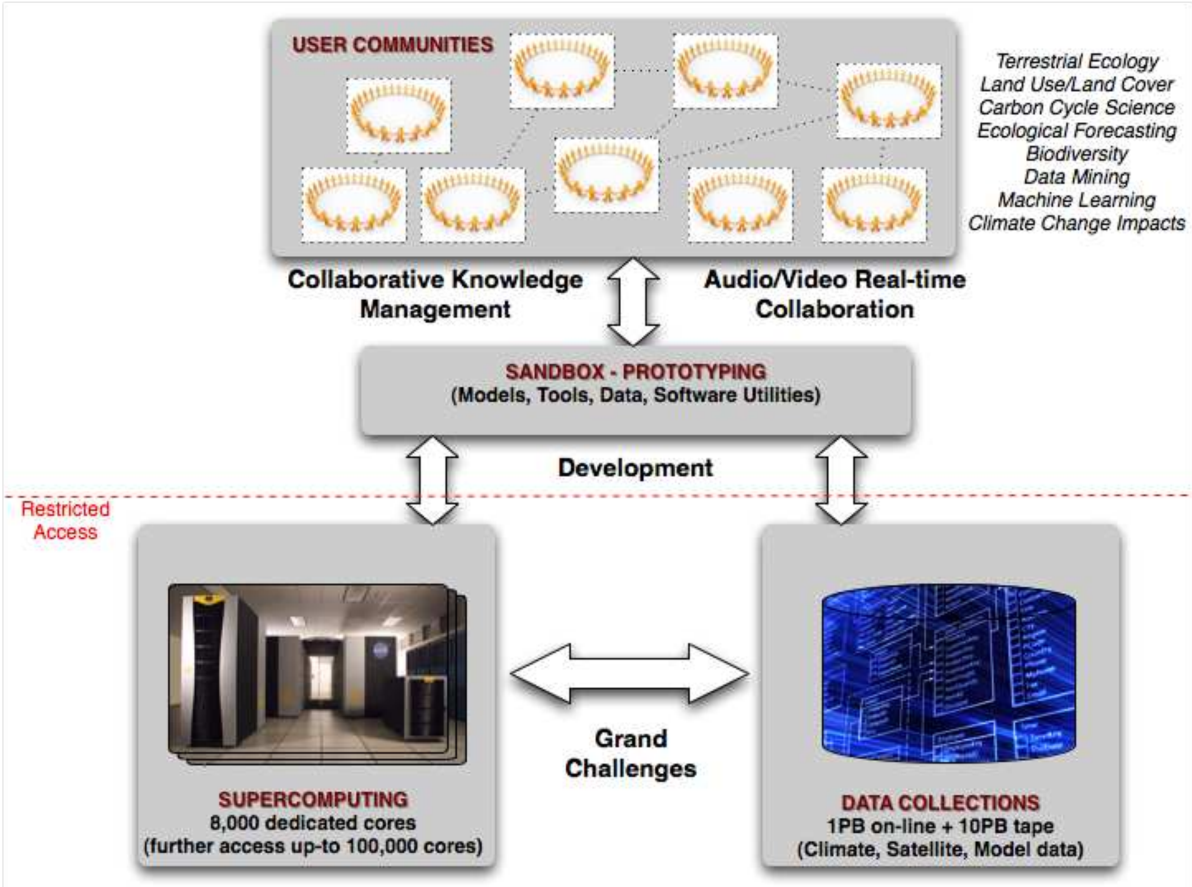

Figure S1: Overview architecture of the NEX system.

Access to NEX

Access to NEX is a three-step process. The first step involves registration and a request to access the NEX web portal using OpenID. Once the request is approved, members have access to the scientific social networking portal where they can interact with existing members via discussions and "wikis" and build project pages for their research. Users have complete control over what items are in their project, and whether these items are shared with selected project members or with the wider Earth science community. The second step, access to NEX data and computing resources, requires a formal application to the NAS facility (www.nas.nasa.gov). For US researchers this may take up to two weeks. Once approved, members will have access to a "sandbox" with 48 AMD Opteron cores that can be used for prototype development, basic analysis and visualization, work environment creation and collaborative outreach. Step three, necessary only for members that need access to high performance computing and storage for large data sets, requires a brief proposal to NEX that will be reviewed by members of a selection committee. Proposals to NEX should address NASA research and applied science priorities, as defined within the NASA report

Responding to the challenge of climate and environmental change (NASA, 2010), and the NASA Research Opportunities in Space and Earth Sciences (ROSES) solicitations.

Initial development of NEX was funded through the American Recovery and Reinvestment Act of 2009.

Acknowledgments: The NEX project was conceived and implemented collaboratively among scientists from NASA's Earth and computational science programs. We would like to acknowledge the support of Dr. Woody Turner, program manager for ecological forecasting at NASA/HQ, Dr. Tsengdar Lee, program manager for scientific computing at NASA/HQ, Mr. Steve Hipskind, Division Chief, Earth Sciences at NASA Ames Research Center, Dr. Rupak Biswas, Division Chief, NASA Advanced Supercomputing (NAS) Division, Mr. William Thigpen, manager at NAS, and other members of the NEX team (Ashok Srivatsava, Elizabeth Foughty, Jennifer Dungan, Weile Wang, Hirofumi Hashimoto, Samuel Hiatt, and Sangram Ganguly).