

# BER





## EXASCALE REQUIREMENTS REVIEW

An Office of Science review sponsored jointly by Advanced Scientific Computing Research and Biological and Environmental Research

March 28–31, 2016 Rockville, Maryland



### **BIOLOGICAL AND ENVIRONMENTAL RESEARCH**



#### **Exascale Requirements Reviews:** Overview

During 2015 and 2016, the U.S. Department of Energy (DOE) conducted Exascale Requirements Reviews for each of its six Office of Science (SC) program offices. The goal of the reviews was to help ensure the ability of DOE's Advanced Scientific Computing Research (ASCR) facilities to support SC mission science in the exascale age. The reviews brought together scientists, planners, and experts to identify:

- 1. **Grand Challenges:** forefront scientific challenges and opportunities that could benefit from exascale computing over the next decade.
- 2. **Priority Research Directions:** how new high-performance computing (HPC) capabilities will be used to advance boundaries at the various scientific frontiers.
- 3. **Computing Requirements:** how to maximize the potential for exascale computing to advance scientific discovery.

DOE program managers are using the review reports to guide strategic planning and investments for the 2020–2025 time frame.

#### **BER Grand Challenges**

Biological and Environmental Research (BER) programs span two divisions: the Biological Systems Science Division (BSSD) and the Climate and Environmental Sciences Division (CESD).

#### Major BSSD goals are to pursue and understand:

- Plant and microbial biology for sustainable biofuels and bioproducts.
- Genome biology for advancing the fundamental understanding of complex biological processes essential to DOE missions.
- Biological processes as they relate to ecosystem function.
- Enabling computational, visualization, and characterization capabilities.
- New technologies and approaches to study plant and microbial cells' processes.
- Plant, microbe, and microbial community biology through DOE user facilities' capabilities.

#### Major CESD goals are to pursue and understand:

- Next-generation, integrated models of the human-Earth system.
- Processes of atmospheric systems and terrestrial ecosystems.
- Coupled biogeochemical processes in complex subsurface environments.
- The frontiers of climate and environmental science through user facilities like the Atmospheric Radiation Measurement (ARM), Environmental Molecular Sciences Laboratory (EMSL), and other BER community resources.
- Science gaps to address DOE's most pressing energy and environmental challenges.

## Answering BER Challenges in the Exascale Age

Participants at the BER Exascale Requirements Review identified these broadly grouped findings that would directly affect the BER mission need:

- Scalable data processing, data analysis, machine learning, discrete algorithms, and multiscale multiphysical simulation are crucial for advancement of biological and environmental systems science.
- Innovations in representation, search, and visualization of large-scale, heterogeneous, ontologically rich primary and derived biological and contextual data (e.g., abiotic environmental information) are crucial for input to and validation of these methods.
- New architectures, data transport protocols, software libraries, and languages are necessary to create a platform for community tool development and use supporting interactive and seamless interoperation of both mid- and large-scale cluster resources and enterprise-class computing environments.
- Algorithms are needed for Earth system processes such as atmospheric dynamics and clouds, oceans, tracer transport, coastal processes, and land that scale effectively on advanced computer architectures.
- Capability is needed for large ensembles, together with methods to effectively capture statistical information on Earth systems and climate variability beyond brute-force ensembles.
- Fusion of model simulations and observational data must take place for better model initialization, uncertainty analysis, validation, and tuning.
- Earth system model complexity requires exascale systems built with powerful general purpose nodes with large amounts of highbandwidth memory.
- Creation of the necessary system components requires a workforce trained deeply not only in the core computational, data scientific, mathematical, and natural scientific disciplines that underlie the above technologies but in how to co-design and develop tools that support open-community development and research.

#### **Priority Research Directions**

The BER Exascale Requirements Review focused on 11 areas for which an exascale ecosystem can be transformative.

#### BSSD

#### Multiscale Biophysical Simulation from Molecules to Cells

The activities of cellular and organismal populations can be harnessed to produce energy (including from renewables) or to mitigate energy production processes. Highly data-aware systems that scale and integrate data into large, multiscale modeling codes will help accelerate discovery and application of biological knowledge.









#### Mapping Sequence to Models

"Disruptive" experimental technologies are driving the need for extraordinary computational innovations, such as constantly updated taxonomic and gene family phylogenetic trees, open-access publications, and data in large-scale functional genomic resources.

#### Microbes to the Environment

BSSD and CESD researchers incorporate predictions and data into integrated models of biomes and their functions and then propagate their effects to the Earth system scale. New algorithms will help researchers infer largescale networks; select models; train on microand geospatially resolved functional data; and simulate reactive transport models.

#### Biological Big Data Challenges

The rates and scales of measuring sequence data; molecular structure, abundance, and activities; and population numbers are increasing rapidly. Innovations are needed in data transport, management, and processing; knowledge representation; machine learning; and mixed mechanistic and statistical simulation.

#### CESD

#### Atmospheric Simulation and Data Assimilation within the Earth System

Improved representation of clouds and aerosolcloud and land-atmosphere interactions will enable more accurate simulation of their roles in climate sensitivity. Better Earth system models can realistically model droughts, floods, and other low-frequency, high-impact events.

#### Terrestrial and Subsurface Research

Within 10 years, some land processes in Earth system models will be resolved in 1-km (or finer) grids, but many will still operate at subgrid scales. Process knowledge will need to migrate from process-resolving to processparameterized scales.

#### Oceans and Cryospheric Research

Challenges include projecting ice-free Arctic conditions, changes in the ocean's heat and carbon uptake, and sea-level rise (SLR). Higherresolution models could better represent oceanice sheet interactions that contribute to SLR.

#### Earth System Models

Large ensembles of Earth system model simulations and high spatial resolutions (~1 km) will help project extreme weather events, requiring better statistical and ensemble methods, increased model and process resolution, and improved uncertainty quantification methods.

#### Integrated Assessment and Impacts-Adaptation-Vulnerability Modeling

Historically focused on the effects of human activity on climate systems, integrated assessment models are increasingly coupled to other models, requiring flexible, modular, and extensible infrastructures.

#### Transforming Science through Exascale Capabilities: Model-Data Fusion and Testbeds

Exascale computing would enable the use of high-resolution simulations and multiyear hindcasts, testing of computationally expensive parameterizations in high-resolution simulations, and more frequent use of instrument simulators.

#### Transforming Science through Exascale Capabilities: Algorithms and Computational Science

Given the uncertainty concerning future machine architectures (DOE is studying two at present), code will need to be portable across two or more exascale systems, bridging compiler, performance, and scientific portability.

networking services provided by ESNet; all

of these facilities operate under the direction

## The Exascale Requirements Review reports and supporting materials can be found at http://exascaleage.org

DOE's HPC centers are based at Argonne National Laboratory (ALCF), Lawrence Berkeley National Laboratory (NERSC), and Oak Ridge National Laboratory (OLCF), with







of ASCR.

