



U.S. DEPARTMENT OF
ENERGY

Office of
Science

Charge to ASCAC

August 14, 2012

Dr. Daniel Hitchcock

Associate Director for Advanced Scientific Computing Research (ASCR)

U.S. Department of Energy

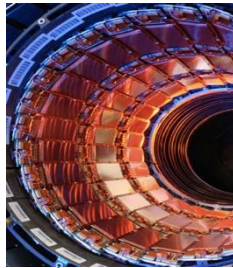
<http://science.energy.gov>

Data Explosion is Occurring Everywhere in DOE



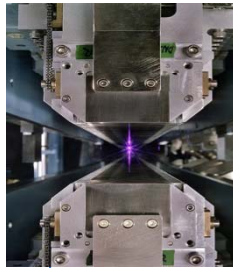
Genomics

- Sequencer data volume increasing 12x over the next 3 years
- Sequencer cost decreasing by 10x over same time period



High Energy Physics

- LHC experiments produce & distribute petabytes of data/year
- Peak data rates increase 3-5x over 5 years



Light Sources

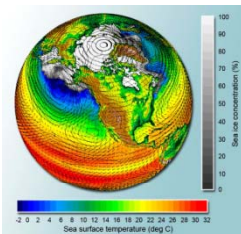
- Many detectors on a Moore's Law curve
- Data volumes rendering previous operational models obsolete

LCLS: 18 TB/day, order of magnitude increase next decade

Upgraded ALS: 10 TB/hour, order magnitude increase next decade

SNS: 90 TB/day at full operation

NSLS-II: initially 100 TB/day, full ops increases by factor of 5 in



Climate

- By 2020, climate data expected to be hundreds of exabytes or more
- Significant challenges in data management, analysis, and networks



Dimensions of the Data Problem

- All of the exascale hardware trends impact data-intensive science (in many cases more than compute-intensive applications) **Square Kilometer Array in Australia needs 100 MW for compute infrastructure;**
- Leverages investments in exascale to maximize impact on the Science missions;
- Data from instruments still on 18-24 month doubling because detectors on CMOS feature size path;
- Significant hardware infrastructure needed to support this, which probably will not be replicated at users' home institution (i.e. launching a petabyte file transfer at a user's laptop is not friendly).



Current ASCR Data Activities

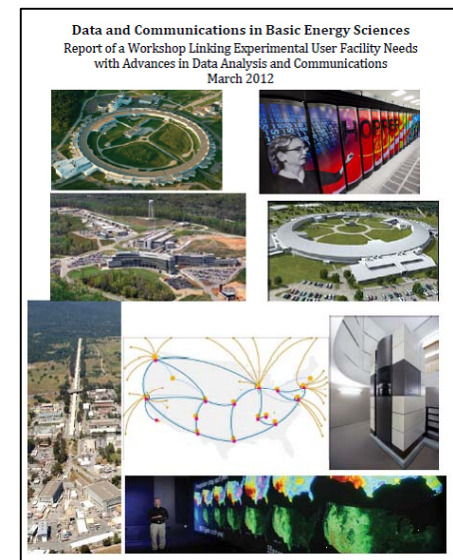
- SciDAC SDAV;
- Mathematics of Large Data;
- Data Intensive Codesign planned for 2013;
- Data Portfolio in Computer Science; and
- Workshops

Data and Communications in the Basic Energy Sciences

GEANT4 Workshop with HEP

Esnet Requirements workshops

NERSC Requirements workshops



Charge Letter



Department of Energy
Office of Science
Washington, DC 20585

Office of the Director

July 25, 2012

2

Professor Roscoe Giles, ASCAC Chair
Department of Electrical & Computer Engineering
Boston University
8 St. Mary's Street
Boston, MA 02215

Dear Professor Giles:

Thank you for the recent Advanced Scientific Computing Advisory Committee (ASCAC) report on the Computational Sciences Graduate Fellowship. The report was thorough, informative and very timely.

Overcoming the challenges of managing data rates and movement of data in an exascale computing environment will likely require significant research investments. In addition to the challenges and opportunities of exascale computing, the Office of Science is facing related challenges from data-intensive research activities, such as the growing volumes of data generated at our next generation scientific user facilities and by the new genomics-based technologies that are enabling a revolution in systems biology research. The Linac Coherent Light Source, for example, currently generates several petabytes of data each year and the National Synchrotron Light Source II, currently under construction and scheduled to begin operations later this decade, is expected to generate hundreds of petabytes of data each year. In order to maximize the return on our limited federal resources, we need to understand the similarities among and differences between these data challenges and the potential to leverage research investments to address issues spanning both exascale and data-intensive science.

By this letter, I am charging the ASCAC to assemble a subcommittee to examine the potential synergies between the challenges of data-intensive science and exascale. The subcommittee should take into account the Department's mission needs, which define the Office of Science's unique role in data-intensive science vis-a-vis other agencies. The subcommittee should specifically address what investments are most likely to positively impact both our exascale goals and our data-intensive science research programs, including data management at our next generation facilities.

I would appreciate the committee's preliminary comments by November 2012 and a final report by March 30, 2013. I appreciate ASCAC's willingness to undertake this important activity.

If you have any questions regarding this matter, please contact either Daniel Hitchcock, the Associate Director of the Office of Science for ASCR or Christine Chalk, the Designated Federal Official for the ASCAC.

Sincerely,

W. F. Brinkman
Director, Office of Science



U.S. DEPARTMENT OF
ENERGY

Office of
Science

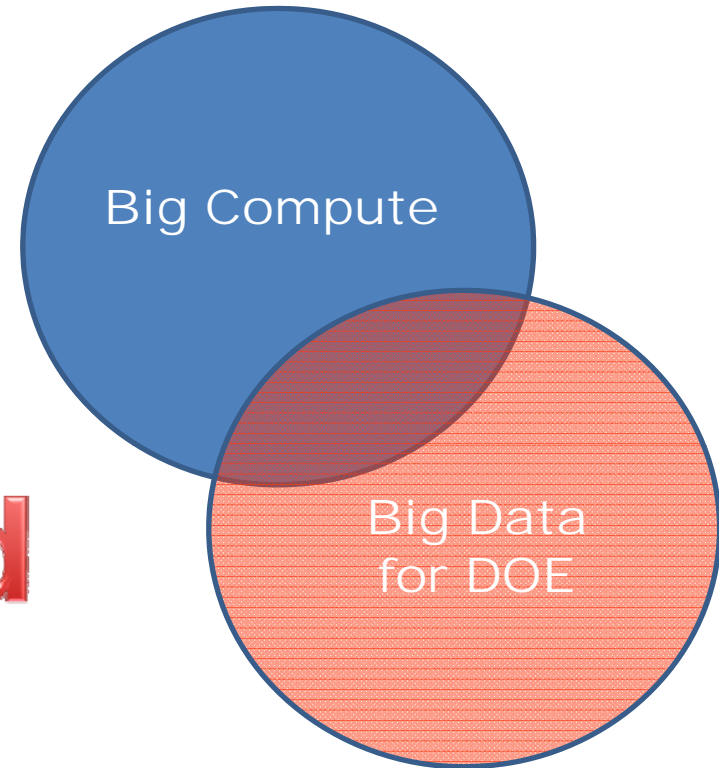
ASCAC August 14, 2012

5

In a Nutshell



And



What part of Big Data apple should/must DOE bite off?

What is intersection of research needed for exascale compute and exabyte data?

