



ADVANCING DISCOVERY IN SCIENCE AND ENGINEERING

The Role of Basic Computing Research

Rapid advances in networking and information technologies, coupled with progress in all fields of scientific inquiry, are enabling researchers to collect and analyze data in magnitudes never before thought possible.

This revolution in “big data” is transforming the way we live, allowing us to infer new knowledge and make

better decisions about some of our biggest societal challenges. | Spring 2011

What is “eScience”?

Discoveries in all fields of science and engineering depend on the collection and analysis of large quantities of data. At the core of this work, computer scientists are developing computational approaches that infer knowledge from large, complex and oftentimes noisy data sets – knowledge that was previously undiscoverable.

This new approach to science – called data-enabled science, or eScience – provides scientists with the ability to:

- collect and manage more data than ever before
- use algorithmic approaches to extract meaning from the data
- provide the right information to the right decision-maker at the right time
- use mathematical and computational algorithms, rather than humans, to guide the scientific process (hypothesis generation, measurement, and evaluation)

The Tools of eScience ...



Sensor Networks

Sophisticated networks of real-time sensors wirelessly linked to the Internet are allowing researchers to instrument, observe and respond to phenomena in the natural environment, as well as in physical and cyber infrastructures.



Data Visualizations

Approaches for presenting large volumes of data in clear, concise, and logical displays are allowing us to quickly identify important elements within these data – often in order to make time-sensitive decisions.



Databases

Advanced information technologies are facilitating the storage of increasingly massive amounts of data in secure and easily accessible formats.



Machine Learning

New and advanced machines are evolving in their ability to learn new knowledge over time.



Data Mining

Computer algorithms are enabling us to mine massive data sets quickly and inexpensively, looking for meaningful patterns and extracting new knowledge (including novel hypotheses) about topics as diverse as health, energy, transportation, education and national defense.



Cloud Computing

The ability to utilize the shared resources of many computers linked to one another through the Internet is making it possible to perform all of the above tasks – including storing, transmitting, and accessing enormous quantities of data – much more quickly and at lower costs.



Collecting and Managing Big Data

Advances in computing research are improving how we are able to collect and store data.

The Data Tsunami

Domain-specific scientists have been able to create new experimental platforms – such as gene sequencers, particle accelerators, and high-end telescopes – to generate data in high-throughput fashion. Similarly, all of us are contributing to a “data tsunami” through social networks, mobile phones, and even electronic health records. The result is an unprecedented amount of data becoming available for research – from gigabytes to terabytes to petabytes (about 1,000 terabytes!) – that must be appropriately stored to facilitate later retrieval, analysis and knowledge extraction.

Innovative Use of Data Collection

Scientists today are collecting computerized data streams from a wide variety of diverse locations to better understand:

- Traffic flow on highways
- Moisture in vineyards
- Energy use in homes
- Conditions inside volcanoes
- Violent hurricanes
- Bridge stresses over time
- Trends in human disease



Managing the Data

- Research into advanced computer processors is allowing scientists to capture and store more data than ever thought possible. These processors analyze data similar to a human brain, rather than with computer code. This allows for computer learning that predicts and infers trends in data.



Geophysical Sensing Solution

- Networking and information technology researchers worldwide are using millions of sensors to collect and store geophysical data for energy exploration. An advanced wireless sensing system using networks, storage capabilities, and computation and analysis tools is acquiring extremely high-resolution seismic data and providing a better picture of oil and gas resources. Ultimately, this system is enabling more environmentally friendly exploration.



DID YOU KNOW?

- More technical data have been collected in the past year alone than in all previous years. ¹
- The amount of data scientists and researchers are collecting doubles every year. ¹



Extracting Knowledge from Scientific Data

Computing researchers are facilitating analysis, visualization and understanding of large amounts of data in new ways.

The Essence of eScience

Great strides have been made in developing machine learning algorithms that automatically “learn” knowledge from historical and scientific data. Predictive models can now suggest several outcomes – along with probabilities for each possible outcome – when the system of interest is complex and the solution uncertain. Graphical tools provide visual displays that are easily understood and that highlight meaningful elements of data that may not otherwise be discernable. These advances are increasingly used by scientists to unearth new knowledge in many disciplines, especially when the underlying data sets are so large, heterogeneous, and noisy that they cannot be processed by a single human brain.

Did You Know?

- By 2015, the world will generate the equivalent of almost 93 million Libraries of Congress.²
- Nearly 40 exabytes (an exabyte is 1.074 billion gigabytes) of unique new information will be generated worldwide this year.³
- In 1999, the total volume of information generated was two exabytes. The Internet currently handles one exabyte of data every hour.⁴

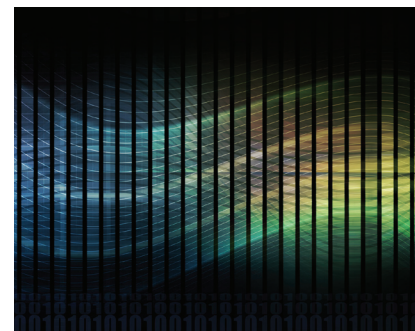
Artificial Intelligence Discoveries

- Scientists are developing computers and robots that collect data and find new discoveries faster than ever imagined. Robots with advanced artificial intelligence systems are discovering breakthroughs in fields such as gene sequencing and biology. This knowledge is resulting in new pharmaceutical cures to combat deadly diseases worldwide without the need for human assistance.



Visualizing Trends

- Computer scientists are collecting data on demographics, finances, crime, health and illness to determine societal trends. For example, using three-dimensional images, interactive maps and other visualization tools, scientists are tracking population changes and determining patterns in disease outbreaks.



Facilitating Growth through Agency Investment

Continued forward progress in scientific advancement is essential to our nation's leadership – and to a broad spectrum of Federal agencies' missions. This progress requires sustained and long-term investment in the tools and applications of eScience, particularly in a way that fosters close partnerships between disciplinary scientists and computer scientists.

National Science Foundation

- Fully fund the Cyber-enabled Discovery and Innovation (CDI) program at levels originally envisioned by the NSF. This investment would provide \$250 million in funding for large-scale, long-term projects with multidisciplinary teams comprised of senior investigators, graduate students and senior personnel. CDI is the key to advancing data-driven discovery and the tools and techniques that enable it.

National Institutes of Health

- Create and invest in a CDI-like program that specifically unites teams of clinicians with computer scientists. For example, pathologists studying tumor samples for protein expression signatures would benefit from data mining approaches. This program should fund projects of similar size and duration as NSF's CDI initiative.

Department of Energy

- The Advanced Research Projects Agency- Energy (ARPA-E) should commit at least \$150 million to eScience that supports the agency's mission of developing a smart-energy infrastructure. A key component of next-generation power systems involves mining large quantities of data to optimize the generation, transmission, and delivery of energy. This program should be forged by teams of power engineers and computer scientists.

Federal Investment Must Support:

- Fundamental research that advances eScience
- Highly collaborative, multi-disciplinary groups of researchers
- Data sharing through incentive-based funding opportunities
- Communication through workshops and conferences
- Stronger programs for education and outreach for and about the data > knowledge > action pipeline

The Need for Computing Research

Computing research has led to breakthrough technologies that have solved some of the world's biggest challenges. Most of the revolutionary technological advances of the last 50 years were pioneered at U.S. universities through Federal research grants. We have a unique opportunity to address grand societal challenges through additional Federal support in key areas of basic networking and information technology research.

Economic Development – Every billion-dollar sub-sector of the IT industry bears the stamp of Federal support for basic research. U.S. preeminence in science and technology has long been the engine of job creation and the source of global economic leadership.

Scientific Advancement – Innovations in networking and information technologies have led researchers to develop new tools that expand the breadth of many scientific disciplines – ranging from the mapping of the human brain to studying issues of climate change to analyzing massive amounts of astronomical data to better understand our universe.

Improve Daily Life – Computing research is improving areas as diverse as healthcare, transportation, energy and education. The development and distribution of these technologies will allow people to live safer lives, conserve natural resources, receive personalized education and beyond.

Technologies Developed from Government-Funded Computing Research

- The Internet
- Google
- Global Positioning Systems (GPS)
- Smart Phones
- Home Security Systems
- Doppler Weather Radar
- Health Monitoring Devices



CITATIONS

¹ Lee Hotz, Robert. "A Data Deluge Swamps Science Historians." *The Wall Street Journal*. (August 28, 2009).

² Duffy Marsan, Carolyn. "Data Deluge." *Nextgov.com*. (August 23, 2010).

³ Gartner Webinar. "Technology Trends You Can't Afford to Ignore." (July 1, 2009)

⁴ Mehlman, Bruce. "Bring On The Exaflood!; Broadband Needs a Boost." *The Washington Post*. (May, 24, 2007).

FOR MORE INFORMATION

Computing Community Consortium

Computing Research Association

1828 L Street, NW, Suite 800 | Washington, DC 20036-4632

(202) 234-2111 | www.cra.org/ccc

The Computing Community Consortium (CCC) is a standing committee of the Computing Research Association (CRA) funded through a cooperative agreement between CRA and the U.S. National Science Foundation. The CCC seeks to mobilize the computing research community to debate and articulate long-term research challenges.

