Data-Intensive Computing at NSF
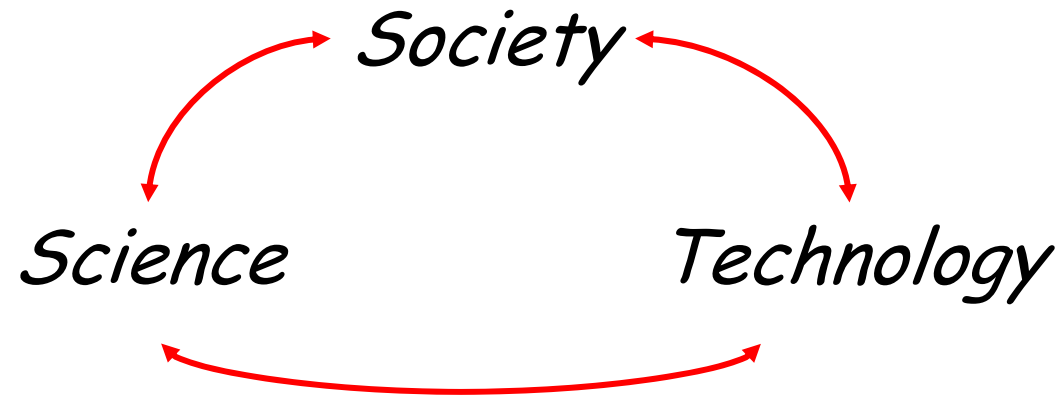
Jeannette M. Wing

Assistant Director
Computer and Information Science and Engineering Directorate

Thanks to the NSF team: Dan Atkins, Debbie Crawford, *Haym Hirsh, Jim French, Stephen Meacham, …*

Data-Intensive Computing Symposium, Yahoo!, Sunnyvale, CA
March 26, 2008

Drivers of Computing

Data-Intensive Computing
Sample Research Questions

Science

- What are the fundamental capabilities and limitations of this paradigm?
- What new programming abstractions (including models, languages, algorithms) can accentuate these fundamental capabilities?
- What are meaningful metrics of performance and QoS?

Engineering

- How can we automatically manage the hardware and software of these systems at scale?
- How can we provide security and privacy for simultaneous mutually untrusted users, for both processing and data?
- How can we reduce these systems' power consumption?

Users

- What (new) applications can best exploit this computing paradigm?

Data-Intensive Computing Challenge

(How) can this new computing paradigm be used

- to solve problems unsolvable any other way,
- to solve old problems in simpler or more efficient ways, or
- to enable new applications?

Search is the killer app. Is there more to data-intensive computing?

NSF's Interest in Data-Intensive Computing

- Broad interest, (potentially) long-term

- CISE
  - Cross-directorate: CCF, CNS, IIS
  - Short-term: CluE (more later)
  - Longer-term: Look for cross-cutting theme in FY09 solicitation

- NSF
  - Potentially cross-foundational, e.g., via Cyber-enabled Discovery and Innovation (CDI); CISE, OCI, MPS, ENG, …
  - Why? Scientists are drowning in data!
    - Aside: Opportunity for CISE PIs to work with other scientists and engineers—they have tons of data.

CISE's Immediate Goal

To provide the broad academic community:

- Access to large-scale computing cluster
- Access to massive data sets

- To support research from applications to the machine
- To support education for new courses, labwork, projects

CluE: Cluster Exploratory

- Google+IBM cluster software and services
  - Same as Academic Computing Cluster provided for six universities (announced last October)

- Seed program by NSF
  - $5M will fund SGERs and regular awards
  - Solicitation about to be released
  - Jim French (IIS Program Director)

- Hope: CluE will be a wild success and community interest and demand will be high

Academia-Industry-Government Partnership

- Win-win-win for all

- New model for NSF
  - CISE is breaking new ground at NSF (in many ways)

- NSF/CISE welcomes
  - Other corporations to participate in Data-Intensive Computing effort and other efforts in the future
  - This and other new models of A-I-G partnerships

Thank you!