Big Data and HPC: Merger or Takeover?

Andrew A. Chien
William Eckhardt Professor of Computer Science
University of Chicago and
Director, CERES Center for Unstoppable Computing
May 27, 2016

THE UNIVERSITY OF CHICAGO

CERES
Center for Unstoppable Computing

Argonne NATIONAL LABORATORY

Big Data Applications and Users

- Dramatically richer sources of data...
- Dramatically more errors in the data...
- Broad range of users drawn from non-engineering and physical science (less-technical) fields:
  - business, marketing, finance, public policy, social sciences, law, campaign managers, public health, health care delivery, hotel management, conference management, bus scheduling, apartment rental, retail stocking, general logistics

- => higher level programming languages, flexible interactive shells, easily scalable interfaces, large shared libraries, etc.
- => escape from low-level programming

HPC has much to Contribute

- Range of high performance hardware technologies
  - IB/RDMA, Low latency networks, System integration
- Deep understanding of architecture, system architecture, and storage systems
- Deep performance knowledge, methodologies, culture
- Scalability experience
- => enable rich, high performance implementations

The best of Both Worlds?

Big Data

- Large data
- Inefficient
- Varied applications
- Scalable
- Loose integration
- High Level Programming

High Performance Computing

- Batch scheduled
- Efficient/High Performance
- Low level Programming
- Global Sharing (gigantic memory)