Gaps Between Big Computing and Big Data: Dynamic Graphs and Knowledge Understanding

Thomas Sterling

Professor of Informatics and Computing

Chief Scientist and Associate Director
Center for Research in Extreme Scale Technologies (CREST)
Pervasive Technology Institute at Indiana University

Fellow, Sandia National Laboratories CSRI

March 26, 2013

CENTER FOR RESEARCH
IN EXTREME SCALE
TECHNOLOGIES

INDIANA UNIVERSITY
Pervasive Technology Institute

Graphs – Pervasive, Dynamic, Unstructured

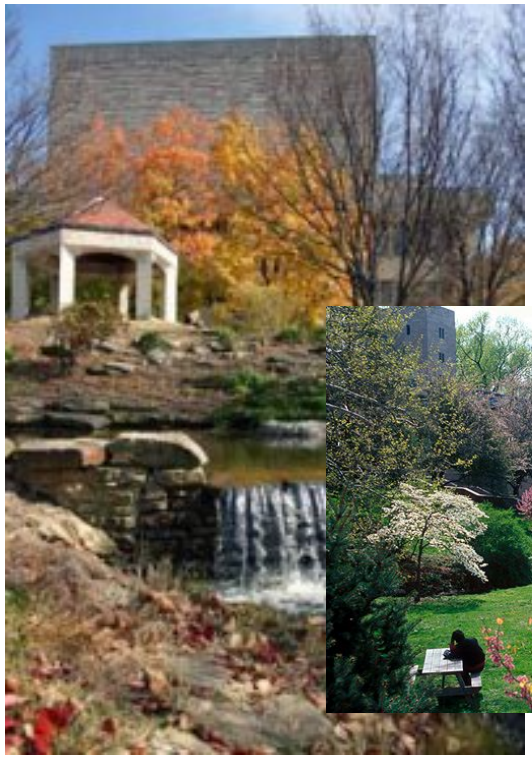
- Historically, big compute (HPC):
  - emphasis on floating point operations
  - HPC memory hierarchy assumes matrices & vectors
    - Rapid indexing through arrays
    - Static structures
    - Exploits spatial and temporal locality to minimize latency and power
- Future Big data (science or informatics):
  - limited in arithmetic operations
  - HPC Graph processing
    - Global data structures
    - Unstructured
    - Dynamic in form
  - Parallelism, synchronization amorphous structure dependent
  - Logical locality

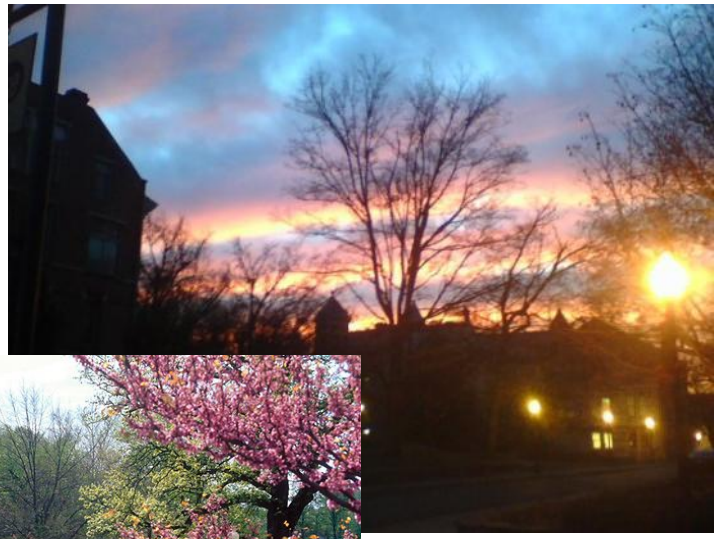
HPGC – High Performance Graph Computing

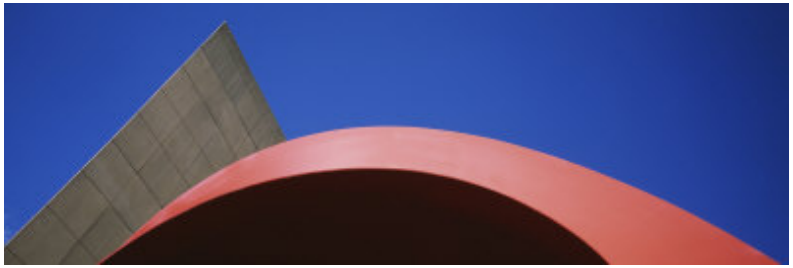
- Concurrency of operation
  - Meta-data structure parallelism
  - Data-directed parallelism discovery
  - Embedded synchronization for decentralized control in the presence of conflicting actions
  - Conventional storage sequentializes parallelism via degree of links
- Low data reuse – high operation repetition
- Suggested transitions
  - Execution model derived to serve new modalities and co-design
  - Message-driven computing to move work to data
  - Powerful synchronization embedded in data
  - Multi threaded
  - Global address space and first class objects

Understanding Knowledge for Knowledge Understanding

- Goal: declarative interrogatives, decisions, discovery, & control
- Computers use data as actionable for operation
- People use knowledge as actionable for understanding
- The gap between "data processing" & "knowledge understanding"
- How to get from Tycho's Mars data to Newton's inverse square law Knowledge (with Galileo's experiments and Kepler's laws)
- But – we confuse knowledge with data correlations and patterns
- Knowledge is about semantics based on absolute foundations
- Machine understanding through knowledge management is ultimate
- Not big data → Smart Knowledge







Thank You