Supporting Data-Intensive Computing at Extreme Scales

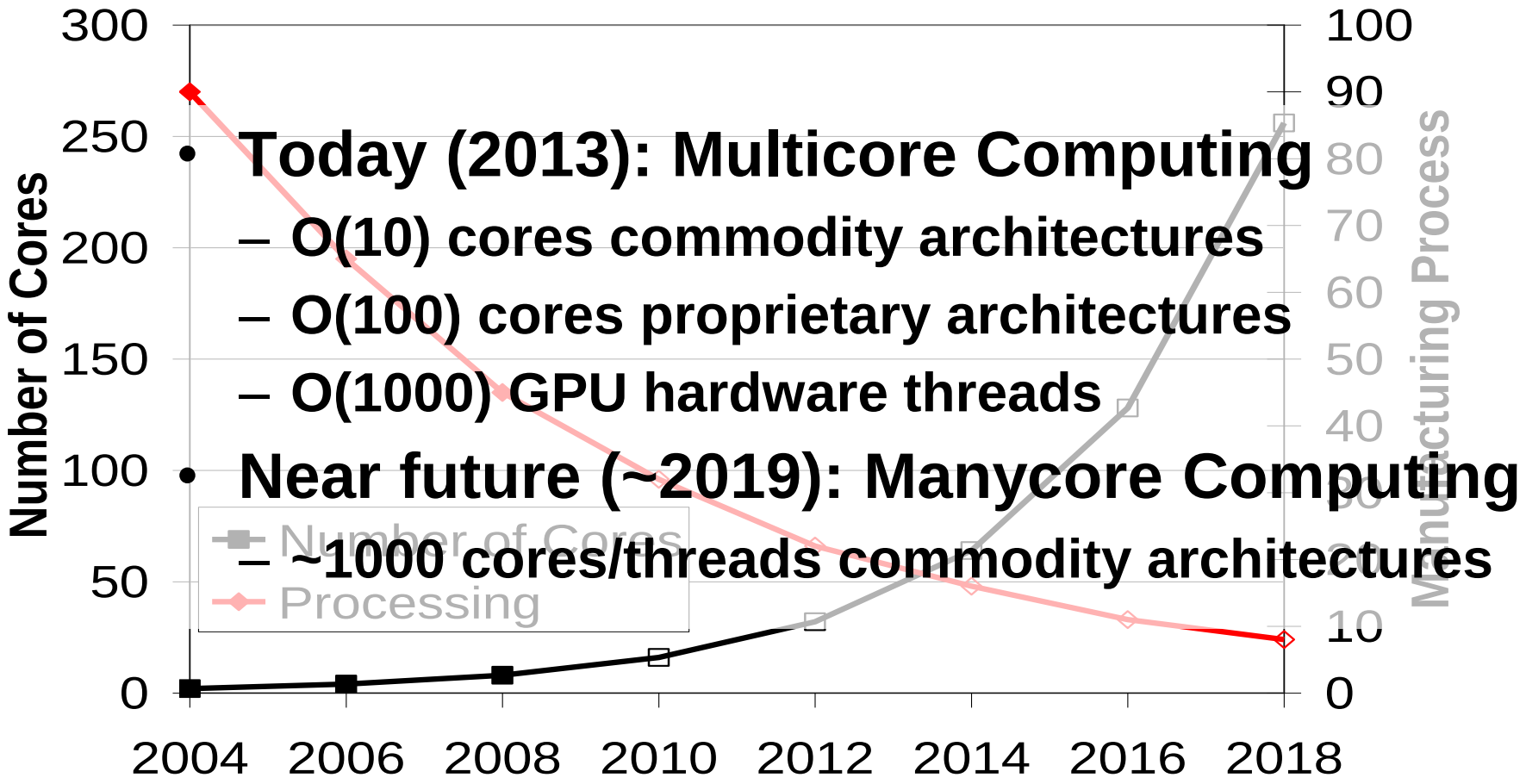
Ioan Raicu
Computer Science Department, Illinois Institute of Technology
Math and Computer Science Division, Argonne National Laboratory

March 12th, 2013
IIT Research Forum – Big Data

DataSys: Data-Intensive Distributed Systems Laboratory

- **Research Focus**
  - Emphasize designing, implementing, and evaluating systems, protocols, and middleware with the goal of supporting **data-intensive applications on extreme scale distributed systems**, from many-core systems, clusters, grids, clouds, and supercomputers
- **People**
  - Dr. Ioan Raicu (Director)
  - 5 PhD Students
  - 3 MS Students
  - 5 UG Students
- **Contact**
  - http://datasys.cs.iit.edu/
  - iraicu@cs.iit.edu

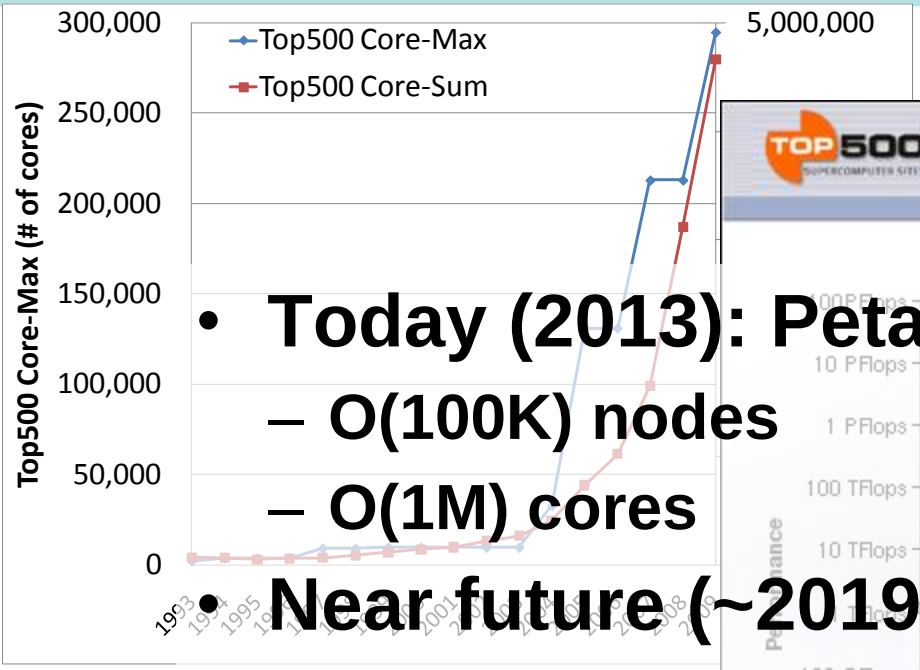
Manycore Computing

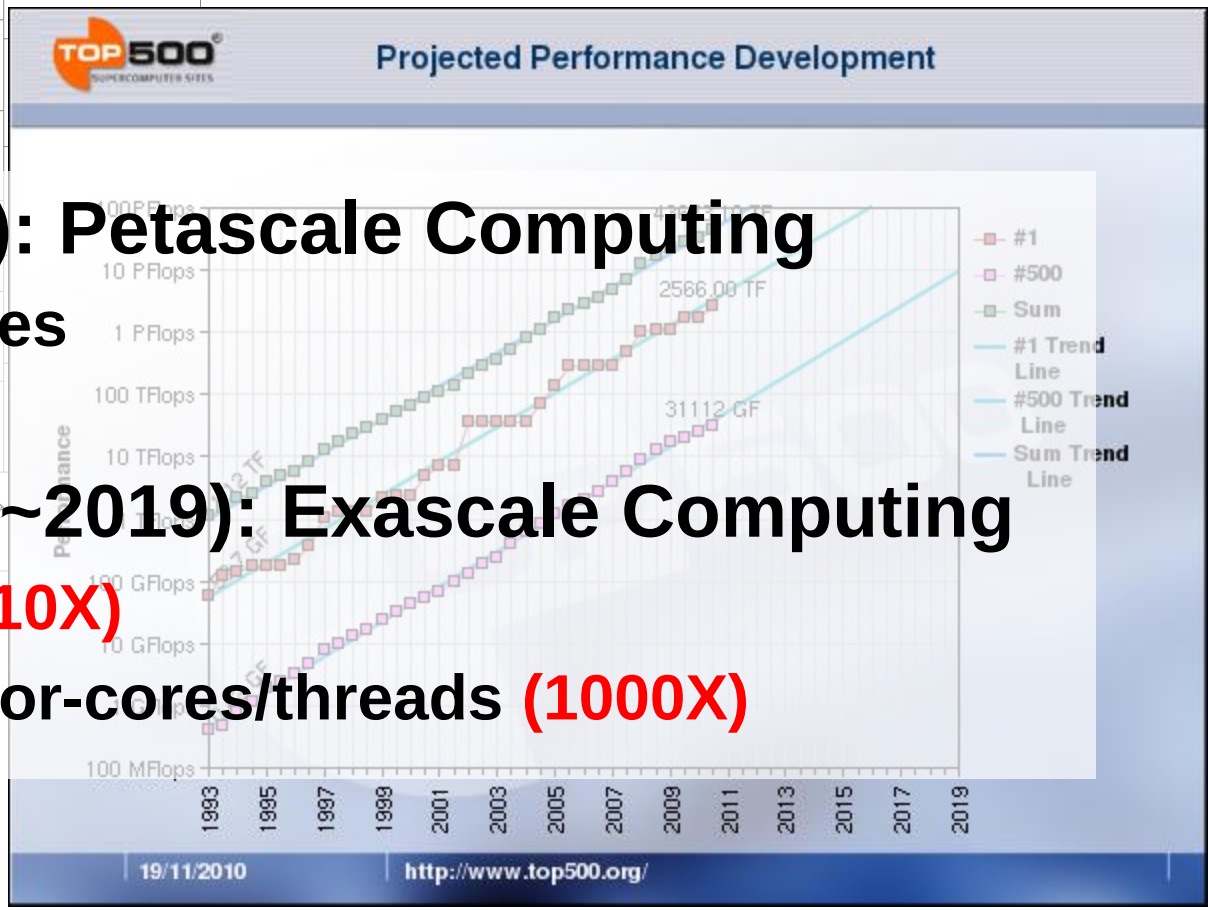


- **Today (2013): Multicore Computing**
  - **O(10) cores commodity architectures**
  - **O(100) cores proprietary architectures**
  - **O(1000) GPU hardware threads**
- **Near future (~2019): Manycore Computing**
  - **~1000 cores/threads commodity architectures**

Pat Helland, Microsoft, The Irresistible Forces Meet the Movable Objects, November 9th 2007

Exascale Computing





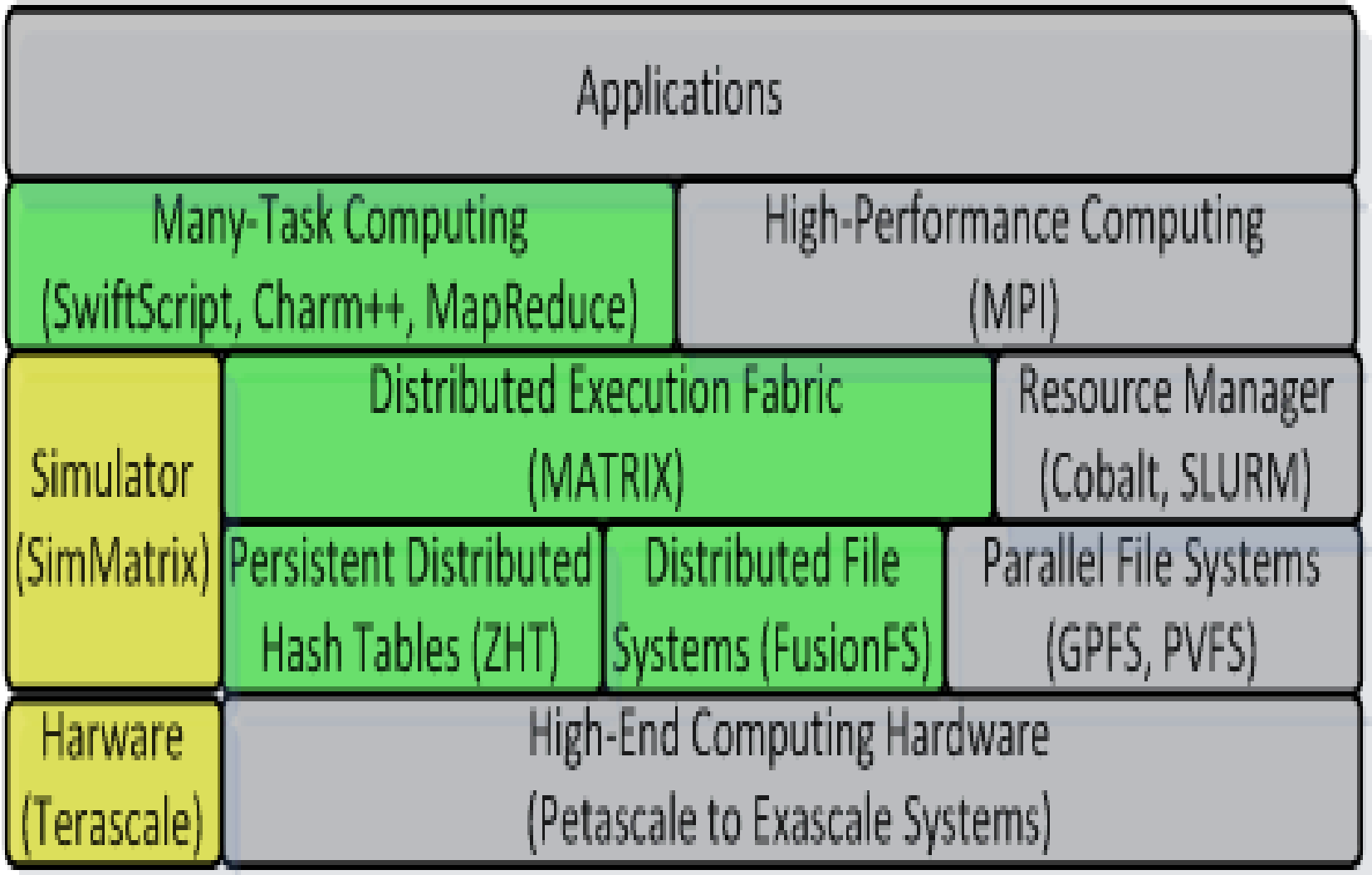
- **Today (2013): Petascale Computing**
  - **O(100K) nodes**
  - **O(1M) cores**
- **Near future (~2019): Exascale Computing**
  - **~1M nodes (10X)**
  - **~1B processor-cores/threads (1000X)**

Top500 Projected Development,
http://www.top500.org/lists/2010/11/performance_development

Projects

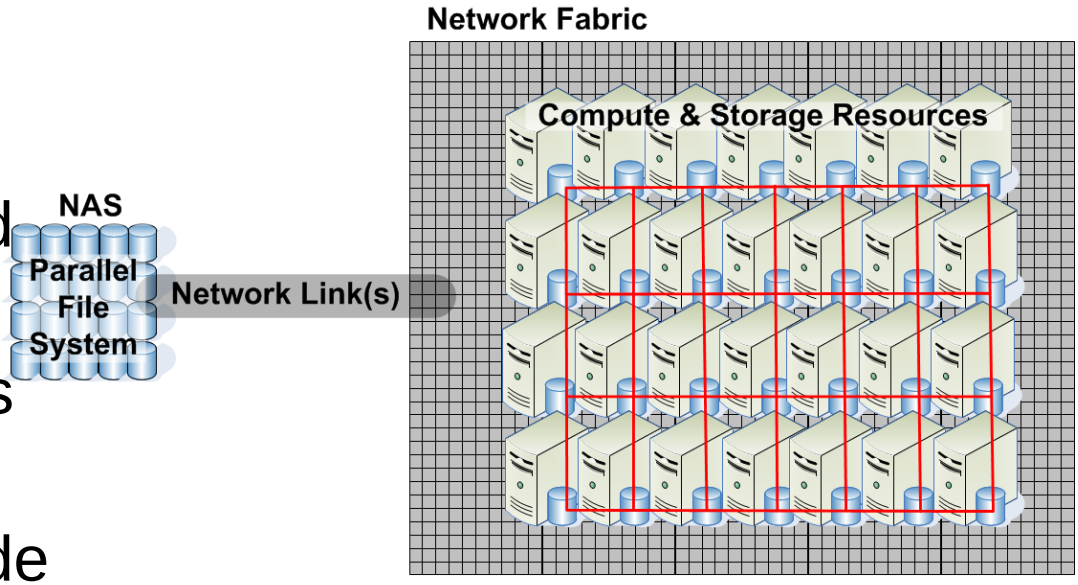
- **Computing**
  - **Many-Task Computing**
    - SimMatrix: Simulator for MAny-Task computing execution fabRIc at eXascales (PDF)
    - MATRIX: MAny-Task computing execution fabRIc at eXascales (PDF)
    - Falkon: Fast and Light-weight tasK executiON framework (PDF)
    - Swift: Fast, Reliable, Loosely Coupled Parallel Computation (PDF)
  - **High-Performance Computing**
    - RXSim: Exploring Reliability of Exascale Systems through Simulations
  - **Cloud Computing**
    - Optimizing Cloud Infrastructure for Scientific Computing Applications
  - **Many-Core Computing**
    - ManyCoreSim: Scheduling Direct Acyclic Graphs on Massively Parallel Processors (PDF)
    - GeMTC: Virtualizing GPUs to Support MTC Applications (PDF)
- **Storage**
  - FusionFS: Fusion distributed File System (PDF)
  - PAFS: Provenance-Aware Distributed File System (PDF)
  - HyCache: A Hybrid User-Level File System with SSD Caching (PDF)
  - ZHT: Zero-Hop Distributed Hash Table (PDF)
  - NoVoHT: Non-Volatile Hash Table (PDF)

Proposed Software Stack in Large-Scale Distributed Systems

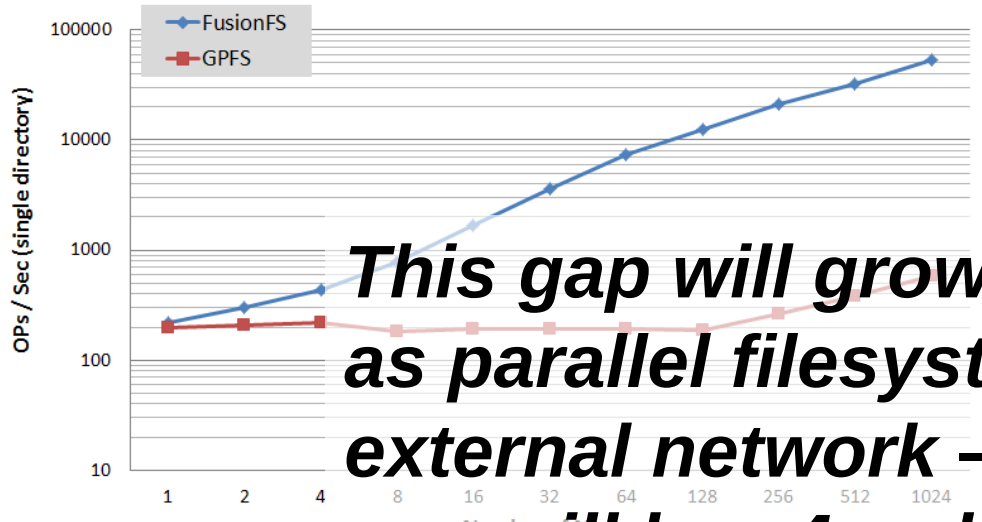


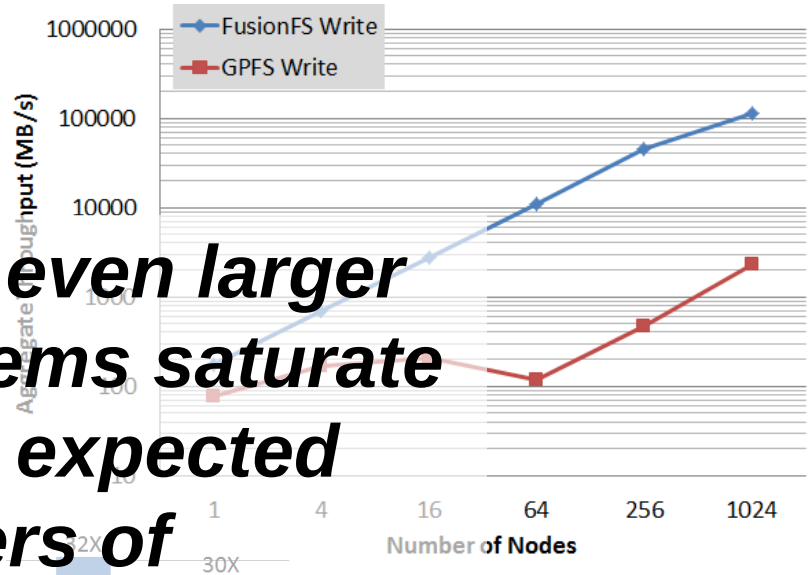
FusionFS Project
NSF CAREER

- A distributed file system co-locating storage and computations, while supporting POSIX
- Everything is decentralized and distributed
- Aims for millions of servers and clients scales
- Aims at orders of magnitude higher performance than current state of the art parallel file systems



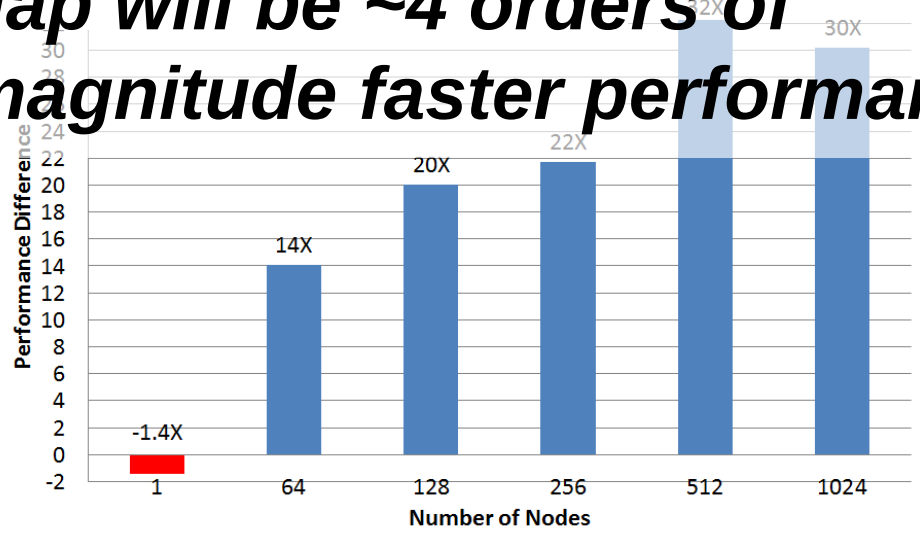
FusionFS Project
NSF CAREER





This gap will grow even larger as parallel filesystems saturate external network – expected gap will be ~4 orders of magnitude faster performance

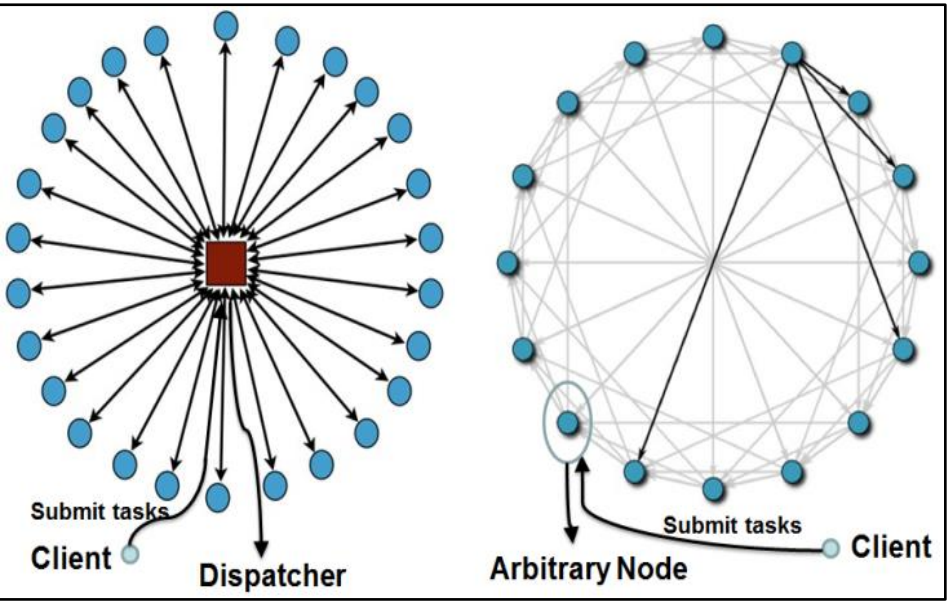
^ ~2 orders of magnitude faster metadata

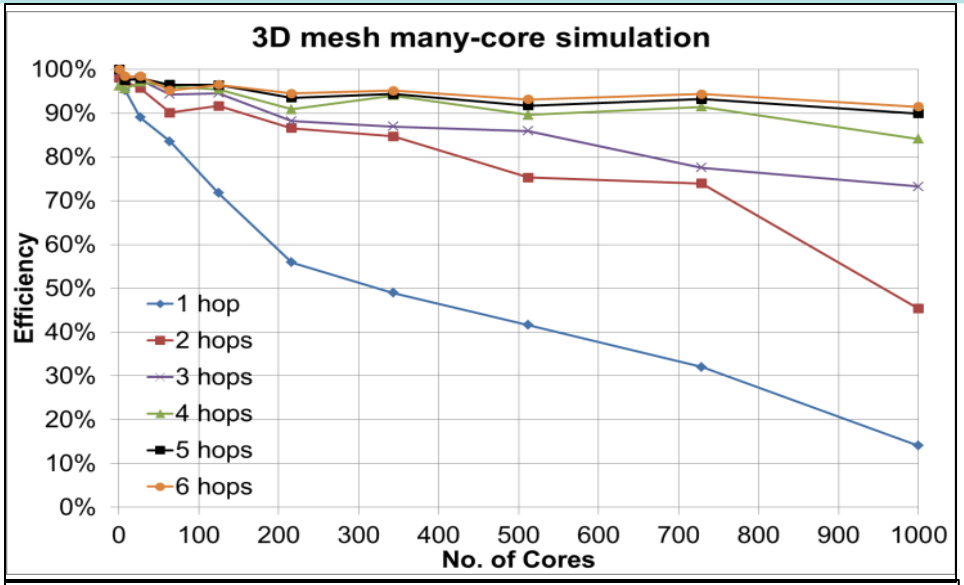


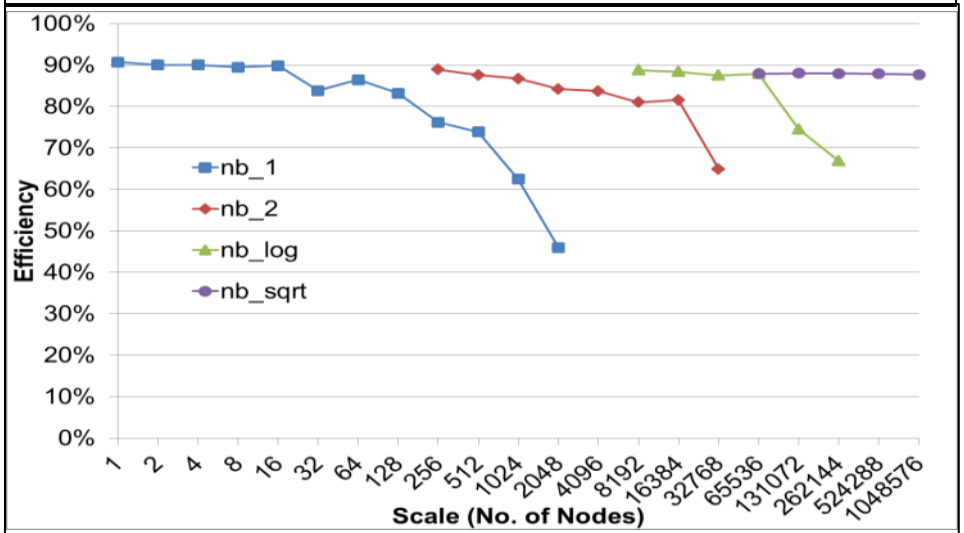
^ ~1.5 order of magnitude faster I/O

< ~1.5 order of magnitude faster runtime for real application

Many-Task Computing







Collaborators

- **National Laboratories**
  - **ANL**: Kamil Iskra, Rob Ross, Mike Wilde, Snir Marc, Pete Beckman, Justin M. Wozniak
  - **FNAL:** Gabriele Garzoglio
  - **LANL**: Mike Lang
  - **ORNL**: Arthur Barney Maccabe
  - **LBL**: Lavanya Ramakrishnan
- **Industry**
  - **Cleversafe:** Chris Gladwin
  - **EMC:** John Bent
  - **Accenture Technology Laboratory:** Teresa Tung
  - **Microsoft:** Roger Barga
  - **SchedMD:** Morris Jette
- **Academia**
  - **UChicago:** Ian Foster, Tanu Malik, Zhao Zhang
  - **University of Electronic Science and Technology:** Yong Zhao
  - **JHU:** Alex Szalay
  - **SUNY:** Tevfik Kosar
  - **USC:** Carl Kesselman

Funding ($)

- **NSF 2011 – 2015: $466K**
  - "*Avoiding Achilles' Heel in Exascale Computing with Distributed File Systems*", NSF CAREER
- **IIT 2013: $15K**
  - **"***Towards the Support for Many-Task Computing on Many-Core Computing Platforms***",** IIT STARR Fellowship
- **DOE 2013: $75K***
  - "Investigation of Distributed Systems for HPC System Services", DOE LANL
- **Fermi 2011 – 2013: $84K***
  - "Networking and Distributed Systems in High-Energy Physics", DOE FNAL
- **Amazon 2011 - 2013: $18K***
  - **"***Distributed Systems Research on the Amazon Cloud Infrastructure***",** Amazon

Funding (Time)

- **DOE 2011 – 2013: 450K hours**
  - "*FusionFS: Distributed File Systems for Exascale Computing*", DOE ANL ALCF; 450,000 hours on the IBM BlueGene/P
- **XSEDE 2013: 200K hours**
  - "*Many-Task Computing with Many-Core Accelerators on XSEDE*", NSF XSEDE; 200K hours on XSEDE
- **GLCPC 2013: 6M hours**
  - "*Implicitly-parallel functional dataflow for productive hybrid programming on Blue Waters*", Great Lakes Consortium for Petascale Computation (GLCPC); 6M hours on the Blue Waters Supercomputer
- **NICS 2013: 320K hours**
  - "Many-Task Computing with Many-Core Accelerators on Beacon", National Institute for Computational Sciences (NICS); 320K hours on the Beacon system

More Information

- More information:
  - http://www.cs.iit.edu/~iraicu/
  - http://datasys.cs.iit.edu/
- Contact:
  - iraicu@cs.iit.edu
- Questions?