

# High Performance Computing (HPC) Networks for Apache Hadoop

Tong Liu

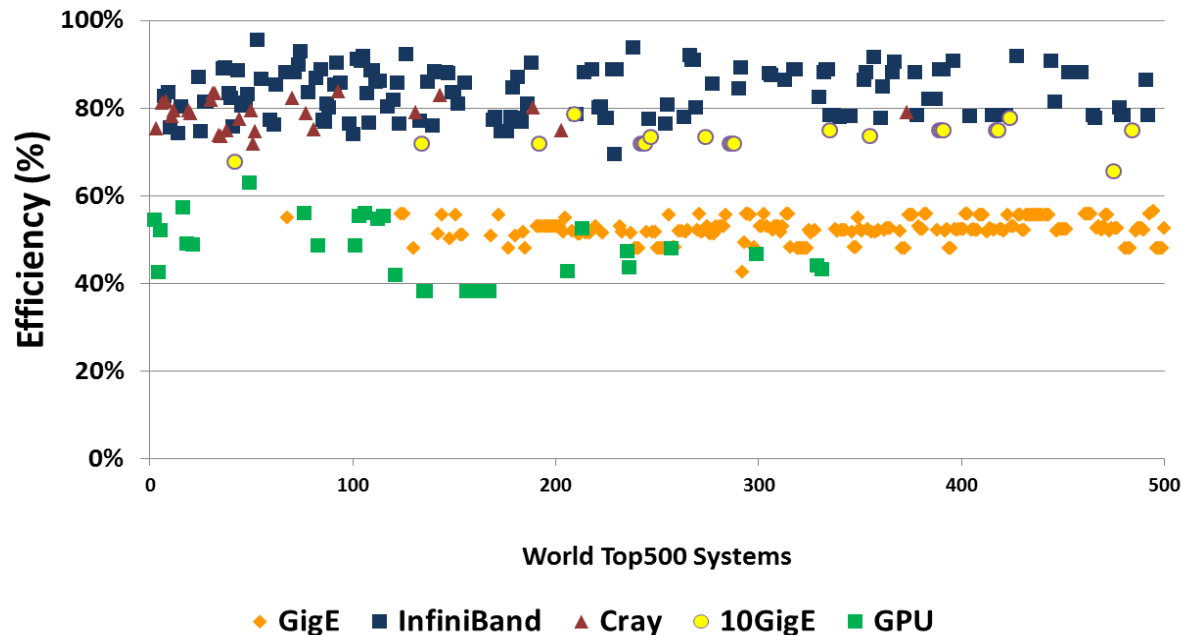
Director of Technical Computing  
The Workshop on Big Data Benchmarking  
May 2012



# Where Big Data Meets Performance

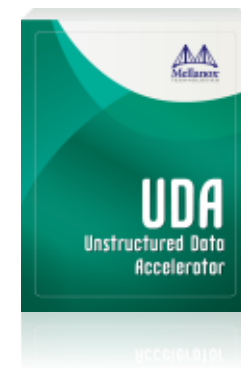
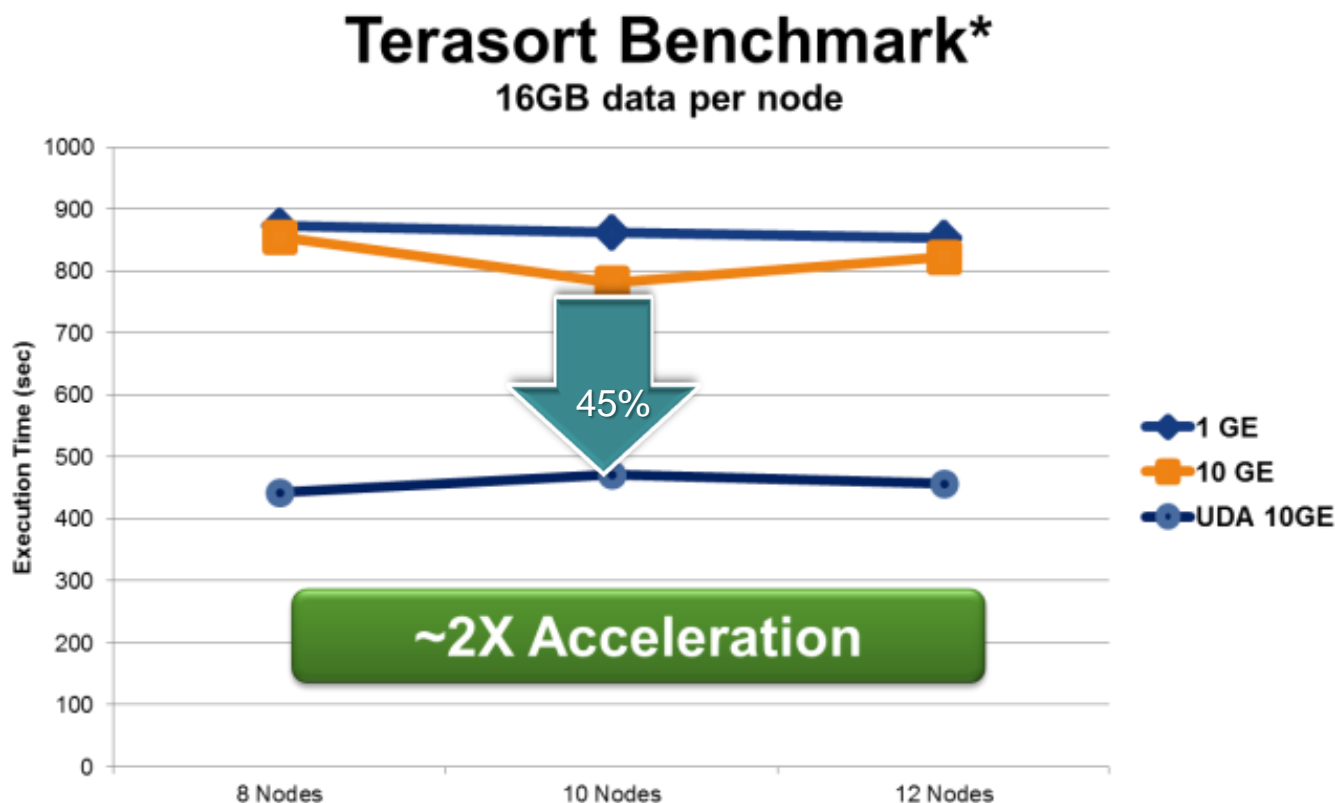
- Leverage HPC Practices to Achieve Higher Server Efficiency
  - Reduce overall Hardware Cost and Energy Footprint
- Enable Centralize Storage Systems for Hadoop
  - Bring Hadoop faster to the Enterprise and Government Markets

World Leading Compute Systems Efficiency Comparison



# Take Advantage of RDMA Programming

- Guarantee Near Linear Scalability and Performance
- Create Big Data Frameworks in a Modular Environment



- **Small Footprint, Lower Cost, Better Performance**
  - Using Only 5 servers (Westmere)
    - Benchmark used: 1 Terabyte, Terasort Benchmark
    - Result: Terasort completed in less than 27 minutes!
  - We know we can push it Faster!
- **Help Building a High Performance Community Test bench**



- **We are looking for collaboration in the community to enable**
  - RDMA based HDFS®
  - RDMA capable distributed Data Bases