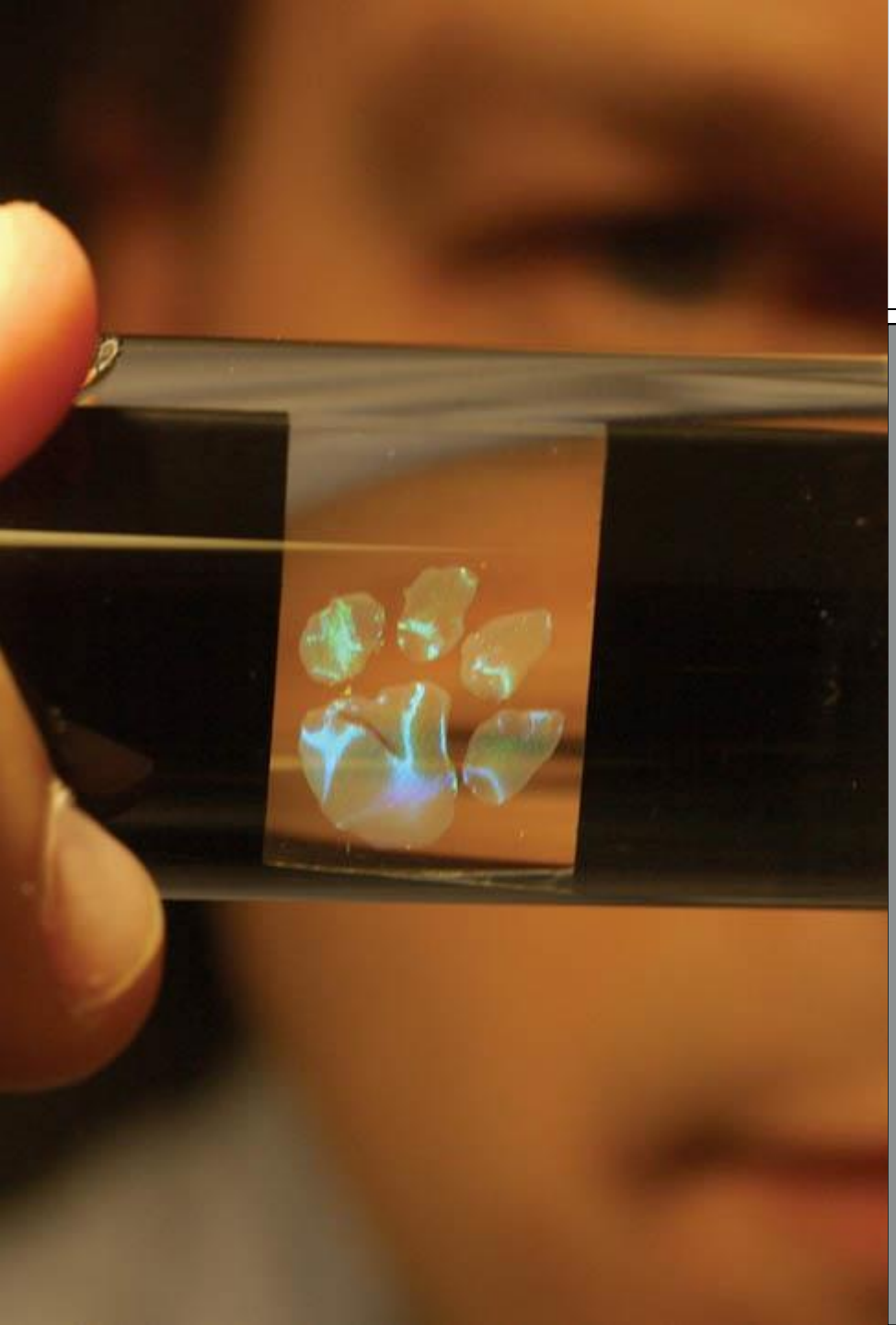


CLEMSON UNIVERSITY

Performance Measurement on Scale-up and Scale-out Hadoop with Remote and Local File Systems

Zhuozhao Li and Haiying Shen

Dept. of Electrical and Computer Engineering
Clemson University, SC, USA

Outline

- Introduction of motivations
- System configuration
- Measurement study
- Conclusion and remark

Introduction

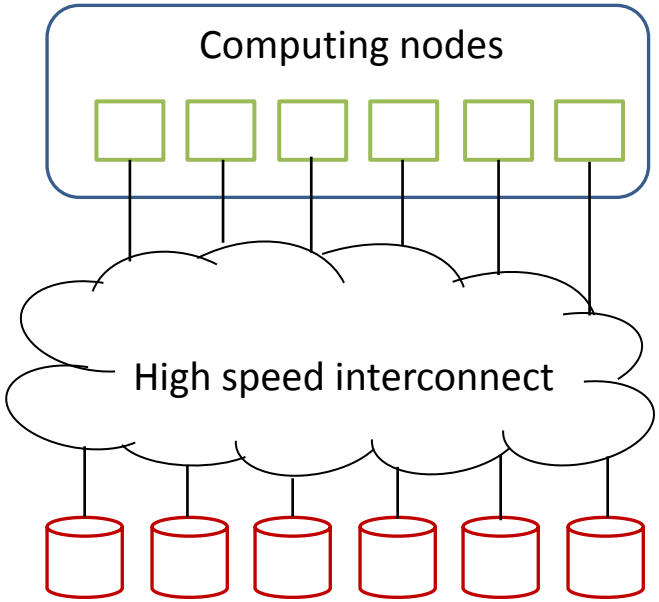
- MapReduce
  - distributed, parallel
  - data-intensive application
  - a cluster of computing nodes
- Hadoop
  - data analytic clusters
  - Facebook and Yahoo

Introduction

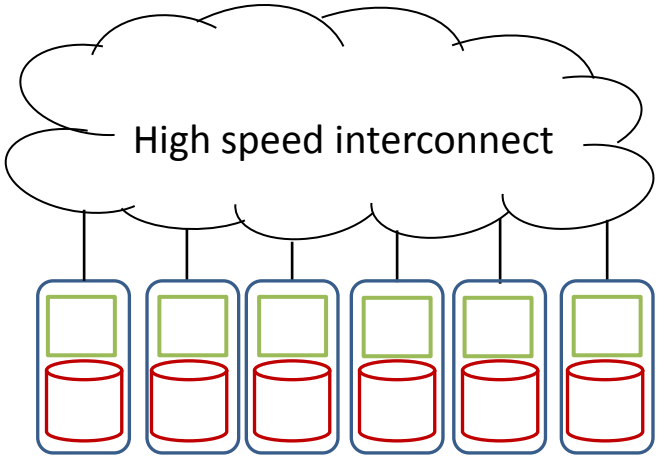
- High-performance computing (HPC) clusters are widely adopted to support CPU-intensive applications.
- HPC clusters also need to process data-intensive workloads.
- Many high-performance computing (HPC) sites extended their clusters to support Hadoop MapReduce.
- However, several settings are different between HPC and traditional data analytic clusters.

Introduction

- File systems?
  - HDFS and HPC remote file system



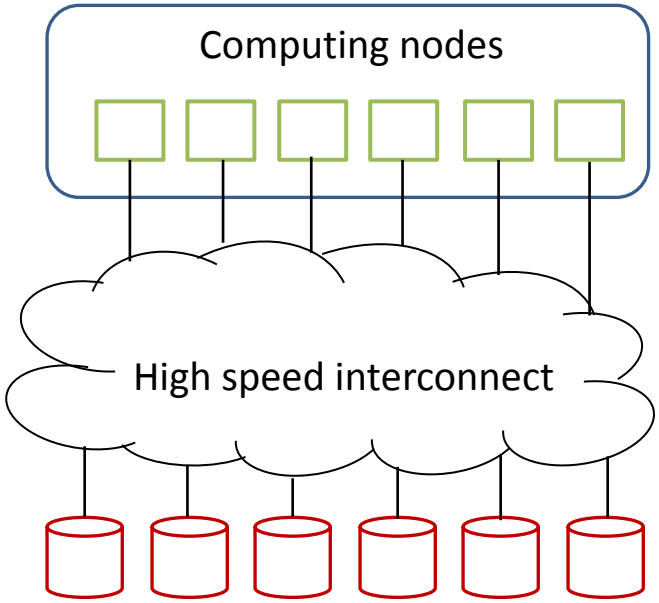
(a) A typical HPC cluster



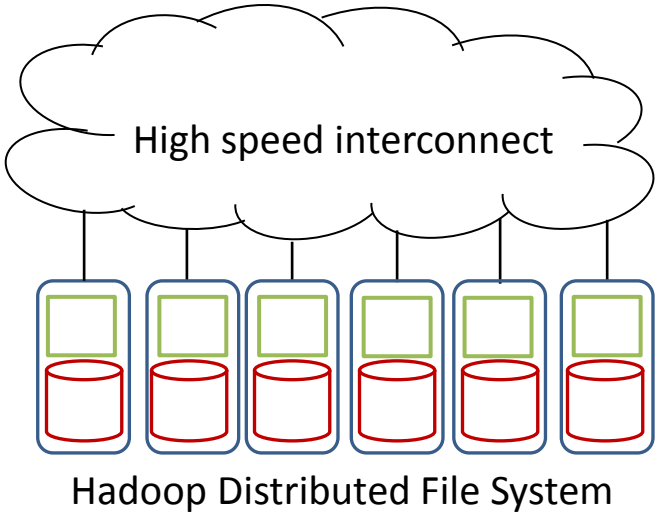
(b) A Hadoop cluster

Introduction

- Clemson Palmetto HPC cluster successfully configured Hadoop by replacing the local HDFS with the remote Orange File System (OFS).



(a) A typical HPC cluster



(b) A Hadoop cluster

Introduction

- Types of machines?
  - A large amount of scale-out machines
  - A few scale-up machines
- Scale-up and scale-out
  - Scale-up: adding more resources to the nodes of a system, typically the processors and RAM
  - Scale-out: adding more nodes with few processors and RAM to a system

Goal

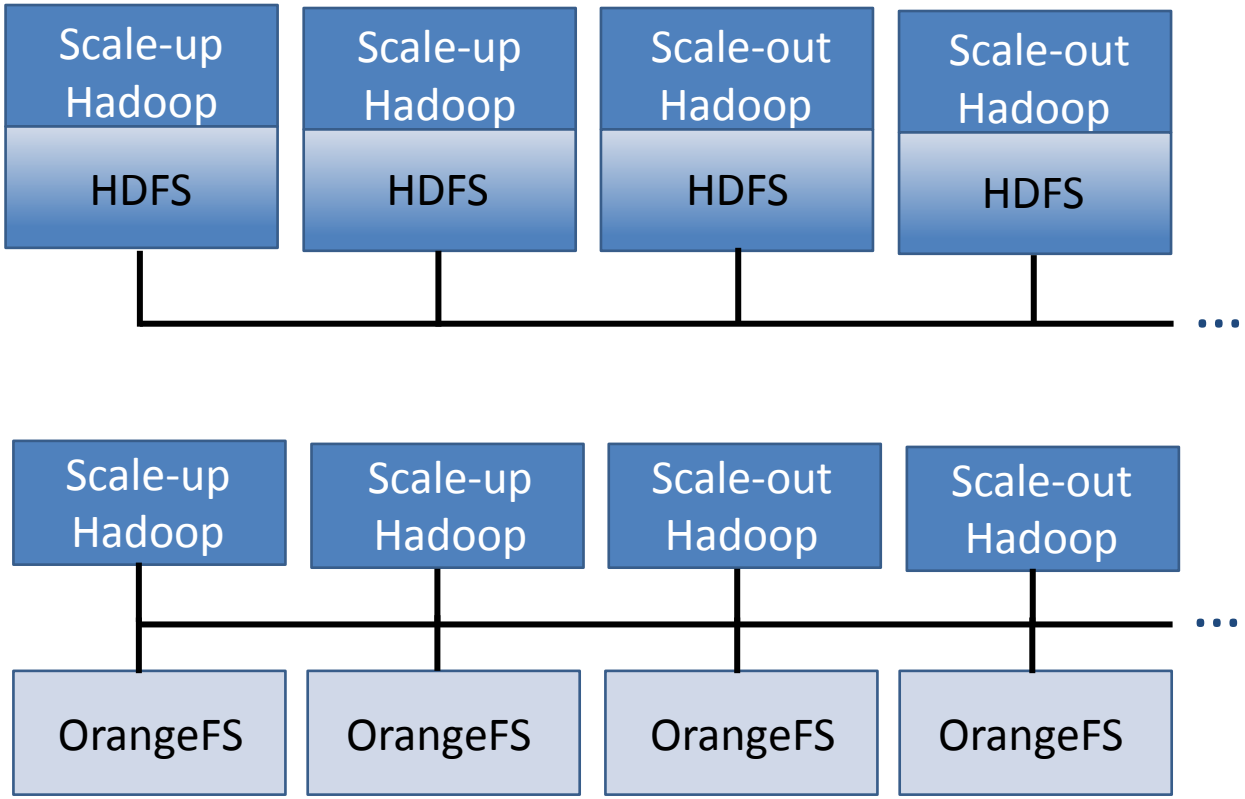
- Real MapReduce workload
  - A real world workload consists of many different types of applications with different job characteristics (data-intensive, CPU-intensive, I/O-intensive)[1].

- We are interested in selecting the best platforms for different types of applications.

DIFFERENT PLATFORMS.

	Scale-up	Scale-out
OFS	up-OFS	out-OFS
HDFS	up-HDFS	out-HDFS

[1] Y. Chen, A. Ganapathi, R. Griffith, and R. Katz. The Case for Evaluating MapReduce Performance Using Workload Suites. In Proc. of MASCOTS, 2011


Scale-up Hadoop
HDFS
Scale-up Hadoop
HDFS
Scale-out Hadoop
HDFS
Scale-out Hadoop
HDFS
...
Scale-up Hadoop
Scale-up Hadoop
Scale-out Hadoop
Scale-out Hadoop
...
OrangeFS
OrangeFS
OrangeFS
OrangeFS

Measurement Setting

- Clemson Palmetto HPC Cluster
- Comparison
  - 2 scale-up machines, 24-cores processor, 505GB RAM
  - 12 scale-out machines, 8-cores processor, 16GB RAM
  - Similar price cost (according to market investigation)
- Hadoop 1.2.1
- HDFS
- Remote file system (OrangeFS), a parallel file system
- RAM drive of scale-up machines
  - Half of the RAM serves as RAMdisk
  - Used to store shuffle data
  - Improve the performance of shuffle stage
- Block sizes 128MB

	Scale-up	Scale-out
OFS	up-OFS	out-OFS
HDFS	up-HDFS	out-HDFS

Measurement Application

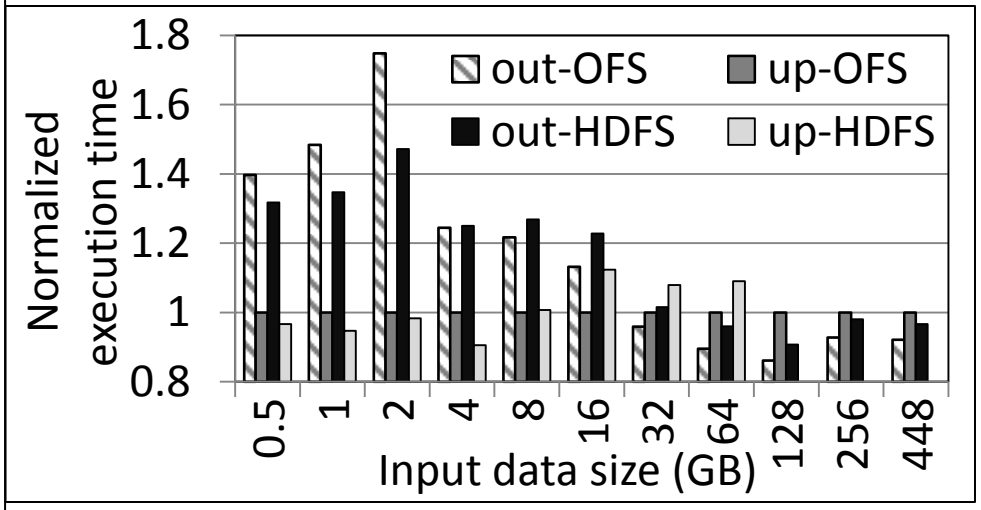
- Data-intensive application
  - A large amount of I/O read/write and a few amount of computation
  - WordCount, Grep
  - Input data generated from BigdataBench [1]
- I/O-intensive application
  - Purely consists of I/O read/write
  - Read test of TestDFSIO
- CPU-intensive application
  - A large amount of computation such as iterative computation
  - PiEstimator

[1] L. Wang, J. Zhan, C. Luo, Y. Zhu, Q. Yang, Y. He, W. Gao, Z. Jia, Y. Shi, S. Zhang, et al. Bigdatabench: A big data benchmark suite from internet services. In Proc. of HPCA, 2014

Measurement Application

- Metrics
  - **Execution time**
  - Map phase duration
  - Shuffle phase duration
  - Reduce phase duration

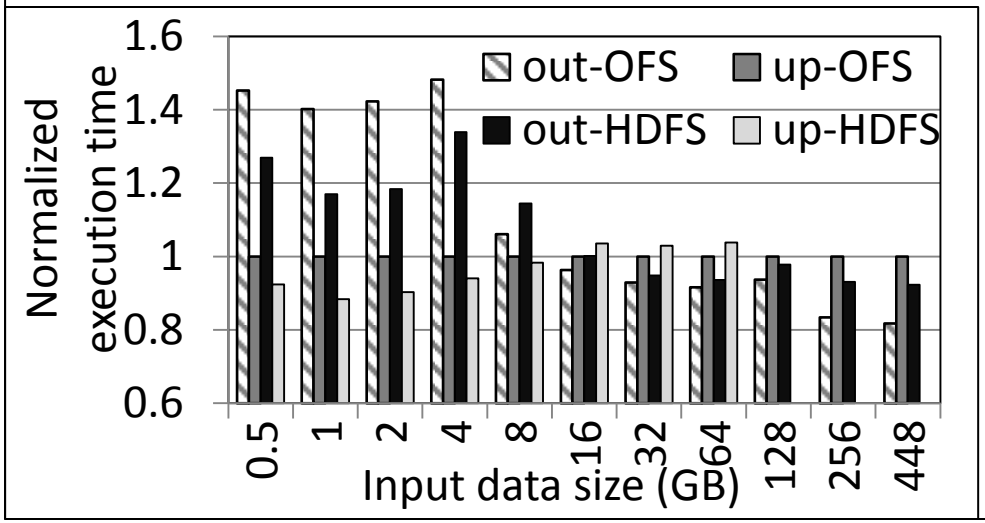
Measurement Analysis



Execution time of WordCount

Small input size (<32GB):
up-HDFS>up-OFS>out-HDFS>out-OFS

Large input size (>=32GB):
out-OFS>out-HDFS>up-OFS>up-HDFS



Execution time of Grep

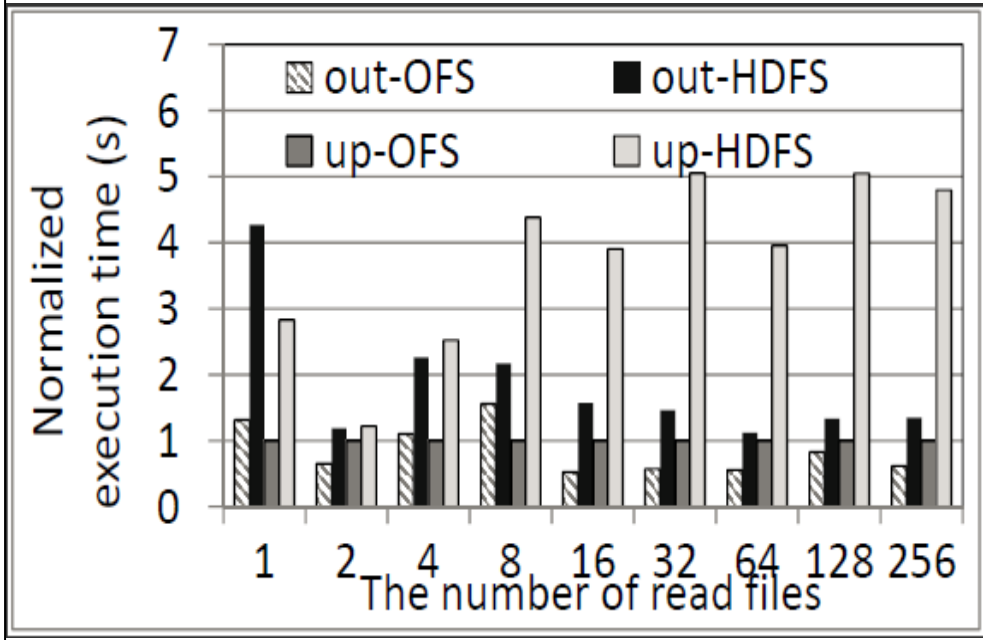
Small input size (<16GB):
up-HDFS>up-OFS>out-HDFS>out-OFS

Large input size (>=16GB):
out-OFS>out-HDFS>up-OFS>up-HDFS

Measurement Analysis

- Scale-up or scale-out
  - Small, scale-up
  - Large, scale-out
  - Powerful CPU for scale-up, RAM disks, but fewer CPU cores
  - More CPU cores
- Local or remote file system
  - Small, local
  - Large, remote
  - Latency non-negligible

Measurement Analysis



Total size = 80 GB

Execution time of read test of TESTDFSIO

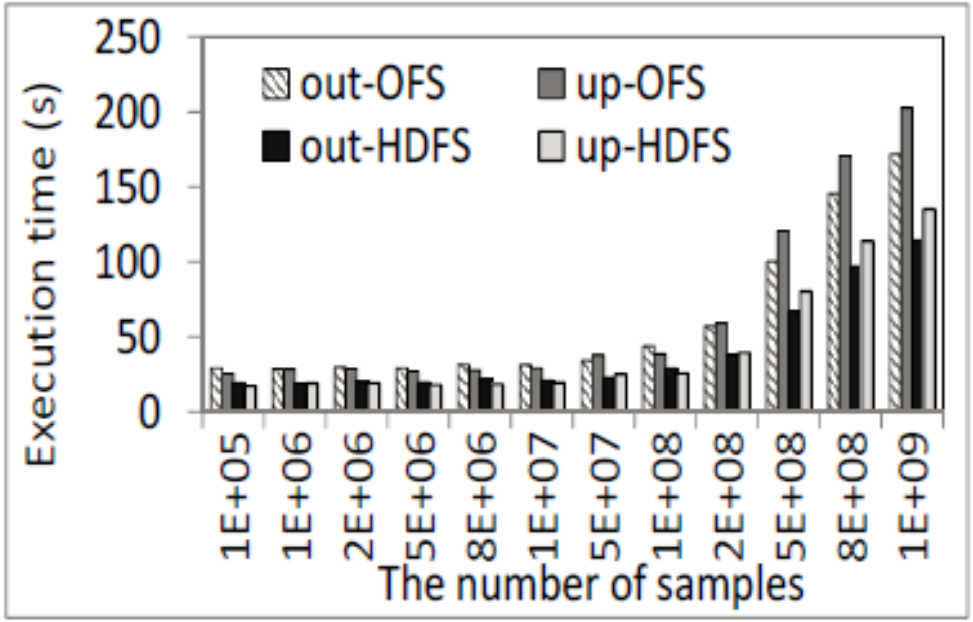
Number of read files is small (<16GB):
up-OFS>out-OFS>up-HDFS>out-HDFS

Number of read files is large (>=16GB):
out-OFS>up-OFS>out-HDFS>up-HDFS

Measurement Analysis

- Scale-up or scale-out
  - Number of files, small, scale-up
  - Number of files, large, scale-out
  - Number of disks
- Local or remote file system
  - OFS better

Measurement Analysis



Number of mappers = 80

Execution time of PiEstimator

Amount of computation is small:
up-HDFS>out-HDFS>up-OFS>out-OFS

Amount of computation is large:
out-HDFS>up-HDFS>out-OFS>up-OFS

Measurement Analysis

- Scale-up or scale-out
  - Amount of computation, small, scale-up
  - Amount of computation, large, scale-out
  - Hit rate
- Local or remote file system
  - HDFS better

Discussion

- We expect that this gives a guidance to users on how to select the best platforms
  - selecting machines
  - selecting file systems
  - Not necessary to maintain the same in different HPC clusters

- Clouds, e.g., EC2
  - data is stored in a dedicated storage (e.g., Amazon S3)
  - multiple types of machines are available to rent

Conclusion

- Conducted performance measurement study of data-intensive, I/O-intensive and CPU-intensive applications on four HPC-based Hadoop platforms
- Expect that our measurement results can help users to select the most appropriate platforms for different applications with different characteristics
- Future Work
  - The same situations occurs in Clouds architecture. We plan to investigate in Clouds.

Thank you!

Questions & Comments?

Zhuozhao Li

zhuozhl@clemson.edu

Ph.D. Candidate

Pervasive Communication Laboratory

Clemson University