High-Performance Big Data Processing Tools for Neuroscience

Lightning Talk at ACNN (Sept. 2016)

Xiaoyi Lu

The Ohio State University

luxi@cse.ohio-state.edu

http://web.cse.ohio-state.edu/~luxi/

THE OHIO STATE UNIVERSITY

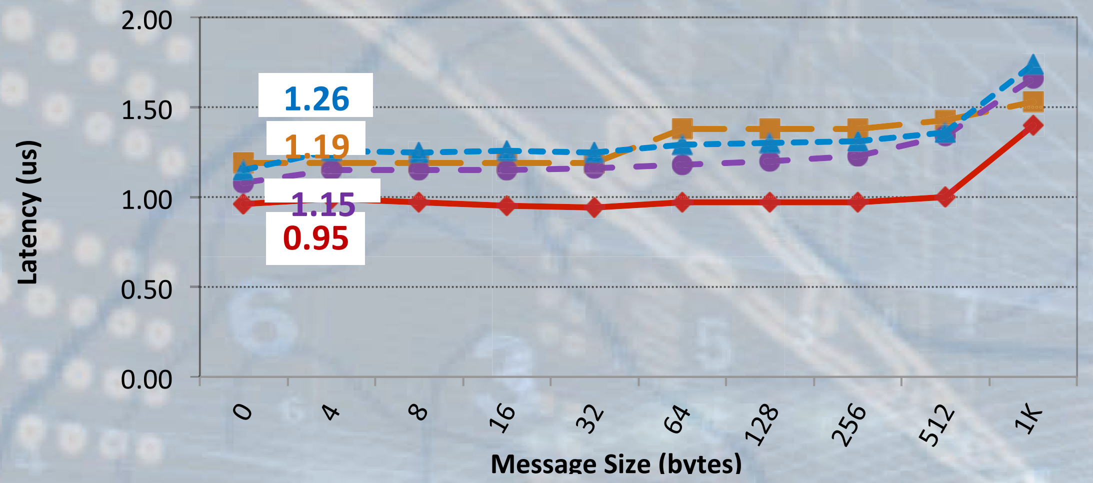
Two Major Middleware Categories to Handle Big Data Processing on Modern Clusters

- HPC Middleware
  - Message Passing Interface (MPI), including MPI + OpenMP, is the Dominant Programming Model
  - Many discussions towards Partitioned Global Address Space (PGAS)
    - UPC, OpenSHMEM, CAF, etc.
  - Hybrid Programming: MPI + PGAS (OpenSHMEM, UPC)
- Big Data Middleware
  - Focuses on large data and data analysis
  - Hadoop (HDFS, HBase, MapReduce)
  - Spark is emerging for in-memory computing
  - Memcached is also used for Web 2.0

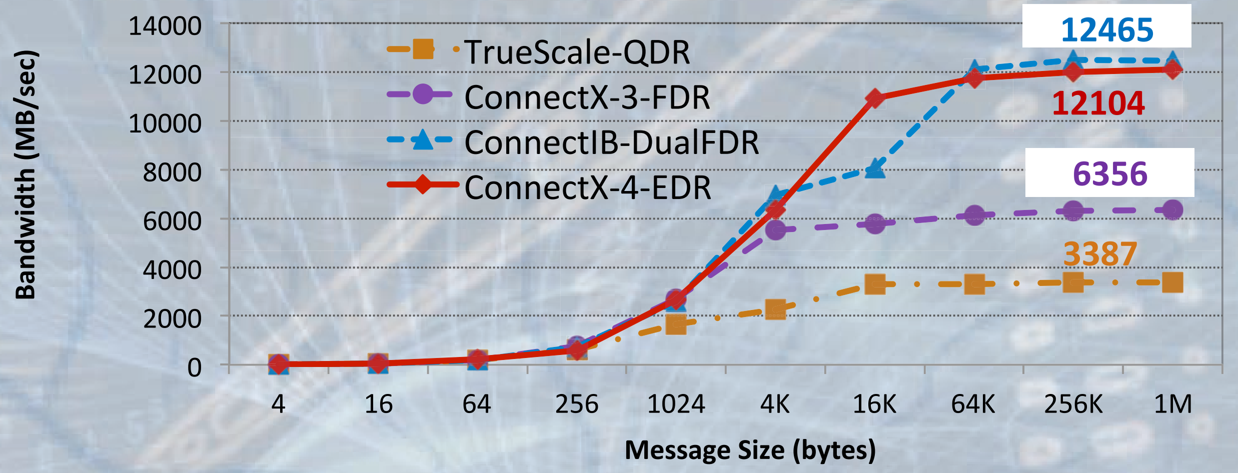
Overview of the MVAPICH2 Project

- High Performance open-source MPI Library for InfiniBand, Omni-Path, Ethernet/iWARP, and RDMA over Converged Ethernet (RoCE)
  - MVAPICH (MPI-1), MVAPICH2 (MPI-2.2 and MPI-3.0), Started in 2001, First version available in 2002
  - MVAPICH2-X (MPI + PGAS), Available since 2011
  - Support for GPGPUs (MVAPICH2-GDR) and MIC (MVAPICH2-MIC), Available since 2014
  - Support for Virtualization (MVAPICH2-Virt), Available since 2015
  - Support for Energy-Awareness (MVAPICH2-EA), Available since 2015
  - Support for InfiniBand Network Analysis and Monitoring (OSU INAM) since 2015
  - **Used by more than 2,650 organizations in 81 countries**
  - **More than 389,000 (> 0.38 million) downloads from the OSU site directly**
  - Empowering many TOP500 clusters (Jun '16 ranking)
    - 12th ranked 519,640-core cluster (Stampede) at TACC
    - 15th ranked 185,344-core cluster (Pleiades) at NASA
    - 31st ranked 76,032-core cluster (Tsubame 2.5) at Tokyo Institute of Technology and many others
  - Available with software stacks of many vendors and Linux Distros (RedHat and SuSE)
  - http://mvapich.cse.ohio-state.edu
- Empowering Top500 systems for over a decade
  - System-X from Virginia Tech (3rd in Nov 2003, 2,200 processors, 12.25 TFlops) ->
  - Stampede at TACC (12th in Jun'16, 462,462 cores, 5.168 Plops)

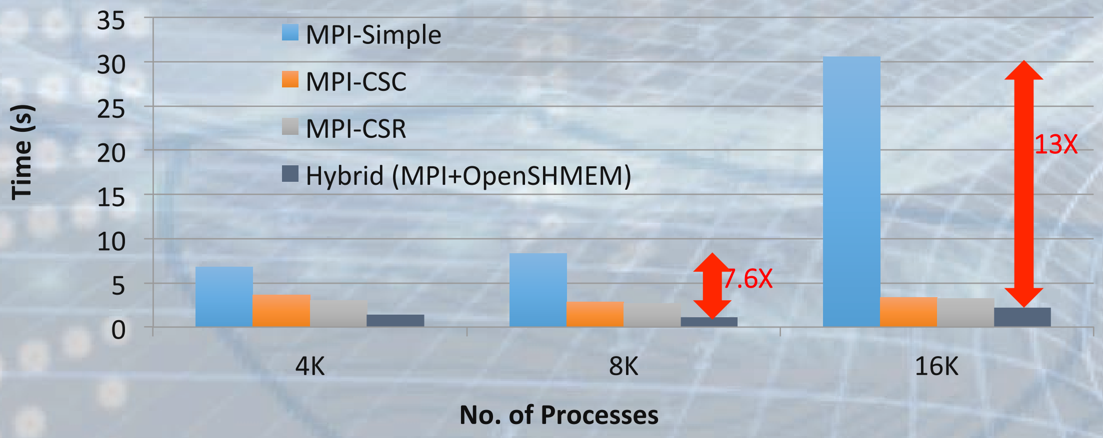
Performance Benefits of MVAPICH2 on OMB, Graph500, and Sort



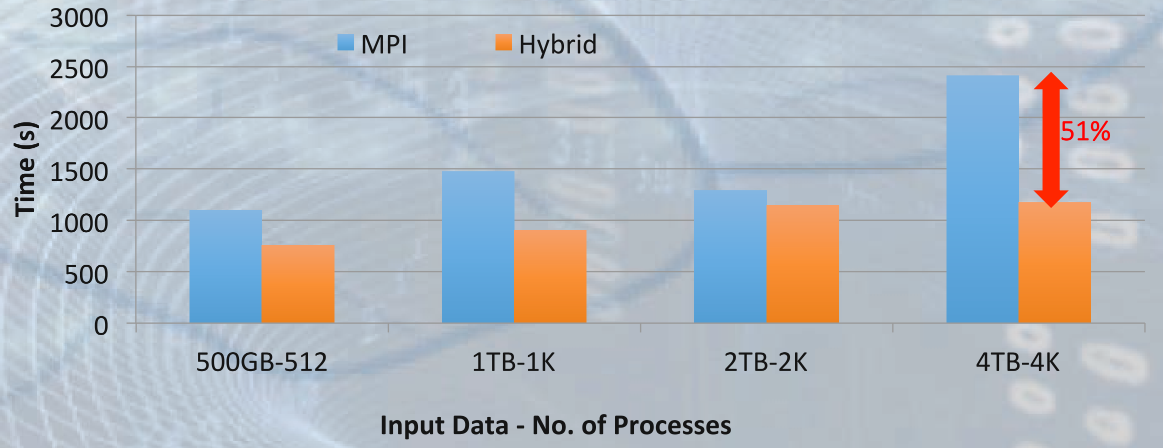
OSU MPI Micro-benchmark Latency



OSU MPI Micro-benchmark Bandwidth



Graph500 Execution Time



Sort Execution Time

The High-Performance Big Data (HiBD) Project

- **http://hibd.cse.ohio-state.edu**
- RDMA for Apache Spark
- RDMA for Apache Hadoop 2.x (RDMA-Hadoop-2.x)
  - Plugins for Apache, Hortonworks (HDP) and Cloudera (CDH) Hadoop distributions
- RDMA for Apache HBase
- RDMA for Memcached (RDMA-Memcached)
- RDMA for Apache Hadoop 1.x (RDMA-Hadoop)
- OSU HiBD-Benchmarks (OHB)
  - HDFS, Memcached, and HBase Micro-benchmarks
- Users Base: 190 organizations from 26 countries
- More than 17,800 downloads from the project site
- RDMA for Impala (upcoming)

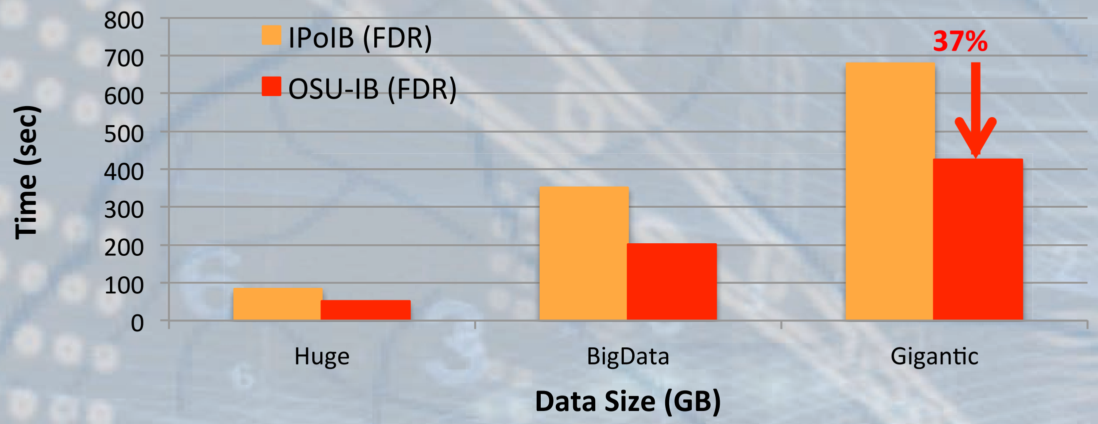
Available for InfiniBand and RoCE

HiBD High-Performance Big Data

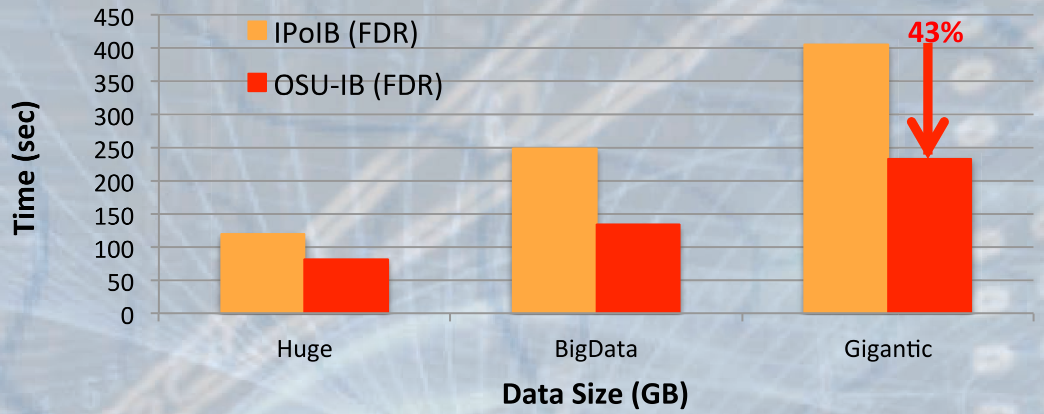
Network Based Computing Laboratory

THE OHIO STATE UNIVERSITY

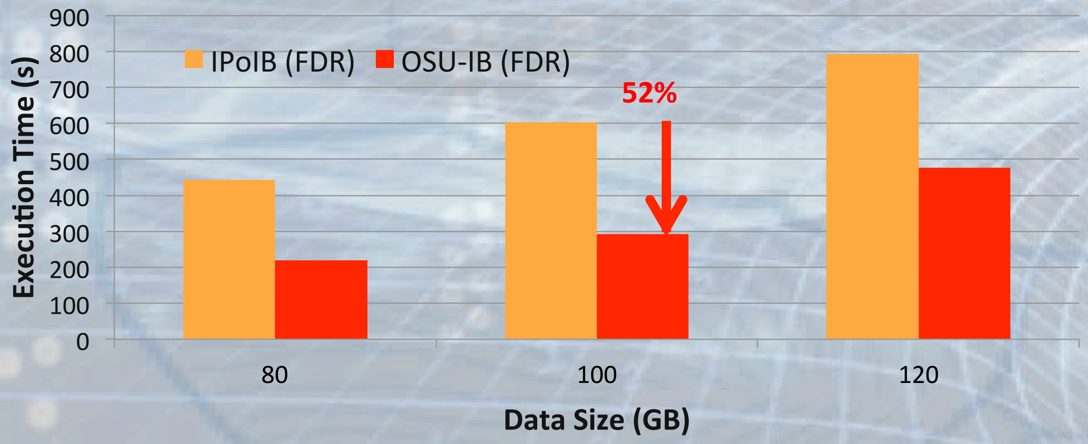
Performance Benefits of HiBD on PageRank, Sort, and TeraSort



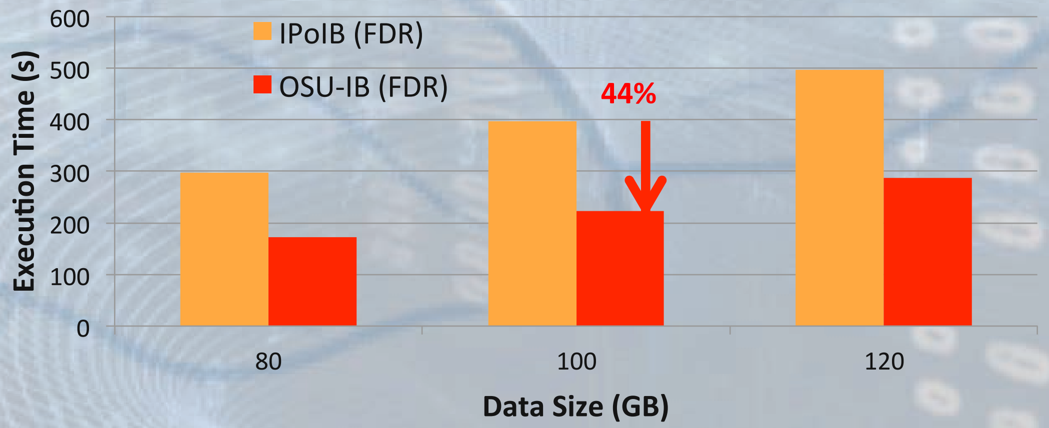
SDSC Comet 32 Worker Nodes, 768 cores
PageRank Total Time



SDSC Comet 64 Worker Nodes, 1536 cores
PageRank Total Time



TACC Stampede 32 Worker Nodes,
Sort Total Time



TACC Stampede 32 Worker Nodes,
TeraSort Total Time

Word Count in Hadoop!

```
public class WordCount {
 public static class Map extends Mapper<LongWritable, Text, Text, IntWritable> {
    private final static IntWritable one = new IntWritable(1);
    private Text word = new Text();
    public void map(LongWritable key, Text valu   Contex  ontext) throws IOException, InterruptedException {
        String line = value.toString()
        StringTokenizer tokenizer = 
        while (tokenizer.hasMoreToken
            word.set(tokenizer.nex        ()
            context.write(word, one);
        }
    }
 }
 public static class Reduce extends Reducer<Text, IntWritable, Text, IntWritable> {
    public void reduce(Text key, Iterator<IntWritable> values, Context context)
      throws IOException, Inte  uptedException {
        int     = 0;
        whil
        
        }
        context     it    y,     IntWr   ble(sum));
    }
 }
}
```

Productive

Scalable

Fault-
Tolerant

Example: Word Count in Spark!

```
val fi           extFile("hdfs:
val cou          le.flatMap(line =>                " "))
                   .map(word => (word, 1))
                   .reduceByKey(_ + _)
counts.saveAsTextFile("hdfs://...")
```

Productive

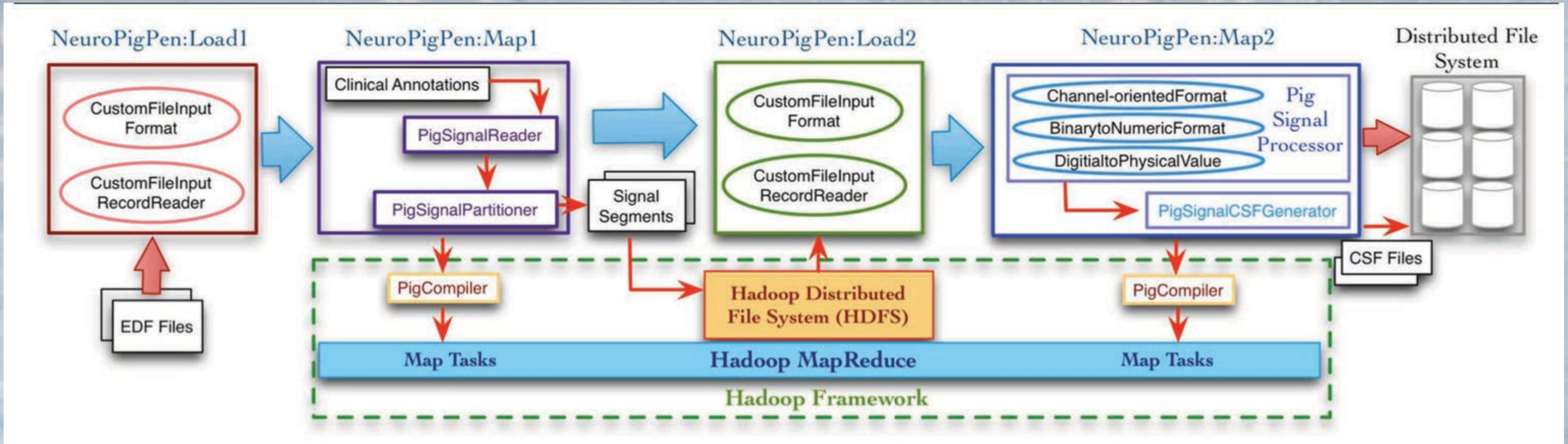
High-Performance

Scalable

Fault-Tolerant

A Case Study with NeuroPigPen

- **NeuroPigPen Toolkit[1]:** a data management toolkit using Hadoop Pig for processing electrophysiological signals in Neuroscience applications
  - Dr. Sahoo, etc., Case Western Reserve University



The data processing workflow supported by the NeuroPigPen toolkit modules consists of multiple steps with EDF files as input and CSF files as output. The Load functions in the toolkit extend the Hadoop FileInputFormat and FileInputRecordReader classes to support signal data. The Map functions in the toolkit are automatically compiled into MapReduce tasks by the Apache Pig compiler. The intermediate and final results are stored in Hadoop Distributed File System (HDFS), which provides a reliable and scalable storage platform for signal data.

1. Sahoo SS, Wei A, Valdez J, Wang L, Zonjy B, Tatsuoka C, Loparo KA, Lhatoo SD. NeuroPigPen: a Data Management Toolkit using Hadoop Pig for Processing Electrophysiological Signals in Neuroscience Applications, http://www.ncbi.nlm.nih.gov/pubmed/27375472

Conclusion

- Overview of MVAPICH2 project and its benefits on Big Data applications
- Overview of HiBD project: accelerating Big Data processing middleware (e.g., Spark, Hadoop, Memcached) by taking advantage of HPC technologies, such as InfiniBand/RDMA, SSD, etc.
- Programming examples with Hadoop and Spark
- A case study with NeuroPigPen
- Soliciting collaboration with other groups
  - Interested in exploring BigData processing requirements from Neuroscience researchers
  - We can explore how to collaborate further
- More details will be presented during our breakout session
  - Overview of Hadoop and Spark programming models
  - Overview of Hadoop and Spark architectures
  - Demo of RDMA-Hadoop on NSF supported Chameleon Cloud

Thank You!

luxi@cse.ohio-state.edu

Network-Based Computing Laboratory
http://nowlab.cse.ohio-state.edu/

The MVAPICH2 Project
http://mvapich.cse.ohio-state.edu/

The High-Performance Big Data Project
http://hibd.cse.ohio-state.edu/

Computational Neuroscience Network (ACNN)

http://www.NeuroscienceNetwork.org/ACNN_Workshop_2016.html