



# Pedal to the Metal: Accelerating Spark\* with Silicon Innovation

**Ziya Ma**

Intel Corporation



DATA IS THE

“Analytics will be the #1 workload  
in the data center by 2020.”

# THE MULTI LAYER VISION

APPLICATIONS



ANALYTICS PLATFORM

Spark

TAP

DATA PLATFORM



INFRASTRUCTURE



Machine Learning

Multi-layered, fully-optimized algorithms

Performance and Security



# ACCELERATE SPARK\* WITH SILICON INNOVATION

## INTEL® XEON® + FPGA INTEGRATED CUSTOM ACCELERATION

**2X** lower latency

**2X** higher bandwidth

## 3D XPOINT™ DIMMS NEW CLASS OF NON-VOLATILE MEMORY

**4X** memory capacity

**1/2** the cost of DDR

## INTEL® OMNI-PATH ARCHITECTURE LEADERSHIP VS. INFINIBAND EDR

**10%** performance advantage

**60%** lower system power

**20%** system cost savings

\*Other names and brands may be claimed as the property of others.  
-intel internal estimates

# HARDWARE OPTIMIZATION FOR PERFORMANCE IMPROVEMENT

**7X<sup>1</sup>** for Spark\* thru hardware migration

**2X<sup>2</sup>** for MLlib\* thru Intel Math Kernel Library

**1.7X<sup>3</sup>** Spark thru hierarchical storage

**1.35X<sup>4</sup>** Spark shuffle RPC encryption for Terasort and

**1.12X** for Big Bench\*



\*Other names and brands may be claimed as the property of others

1. Tested by Intel for Cloud Service Provider real life workloads like nweight

2. Tested by Intel using micro workload and Intel MKL

3. Tested by Intel and partner using spark performance thru hierarchical storage

4. Tested by Intel using Spark shuffle encryption performance for terasort and BigBench



# OPEN SOURCE COMMITMENT

## SPARK CONTRIBUTIONS

- Spark core/Tungsten\*
- Graph X\*
- Spark SQL\*
- Spark Streaming\*
- Spark R\*
- Mllib\*

## SPARK COMMUNITY SUPPORT

- Performance Benchmark
- Big Bench\*
- Spark Meet-up



## SPARK-OPTIMIZED SOFTWARE

- Trusted Analytics Platform (TAP)
- Intel Math Kernel Library
- Intel Data Analytics Library
- Machine Learning – 10X~70 scalability, 8X training cycle reduction

\*Other names and brands may be claimed as the property of others  
-Intel internal estimates

# SPARK USE CASES

Making Solutions Easier to Deploy



Other names and brands may be claimed as the property of others.

# MOVING FORWARD



Enable Internet-of-Things & Cloud Analytics



Scale out Machine and Deep Learning



Optimize Machine and Deep Learning  
Using Hardware and Math Kernels





# Thank You

**[Software.intel.com/bigdata](https://software.intel.com/bigdata)**

**Intel booth: G3**

<http://www.intel.com/performance>