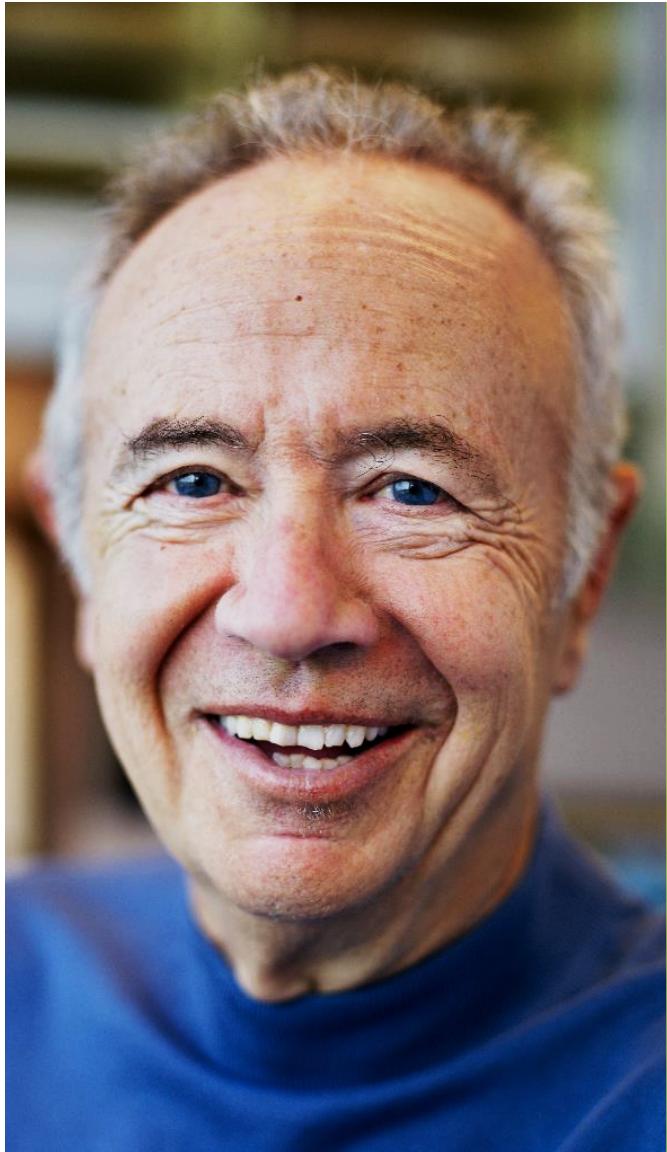




Pedal to the Metal: Accelerating Spark* with Silicon Innovation

Ziya Ma
Intel Corporation



DATA IS THE **GAME CHANGER**

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

THE MULTI LAYER VISION

APPLICATIONS



ANALYTICS PLATFORM

Spark

TAP

DATA PLATFORM

hadoop

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

½ 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

7X¹ for Spark* thru hardware migration

2X² for MLlib* thru Intel Math Kernel Library

1.7X³ Spark thru hierarchical storage

1.35X⁴ 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

Intel booth: G3

LEGAL DISCLAIMERS

- Intel technologies' features and benefits depend on system configuration and may require enabled hardware, software or service activation. Learn more at intel.com, or from the OEM or retailer.
- No computer system can be absolutely secure.
- Tests document performance of components on a particular test, in specific systems. Differences in hardware, software, or configuration will affect actual performance. Consult other sources of information to evaluate performance as you consider your purchase. For more complete information about performance and benchmark results, visit <http://www.intel.com/performance>.

§ Configurations:

Spark 7x performance tested by Intel for Cloud Service Provider real life workloads like nweight. Configuration:

Before: Intel® Xeon® Processor E5-2680 v2 @ 2.80GHz *2 (10 cores, 20 threads); DDR3 1600MHz 192GB; SATA HDD 500GB - 7200RPM; 1Gbps NIC
After: Intel® Xeon® Processor E5-2699 v3 @ 2.30GHz *2 (18 cores, 36 threads); DDR4 2133 MHz 256GB; Intel SSD DC P3700 Series 1.6 TB; 10Gbps NIC

Spark MLlib 2x (micro workload) performance tested using Intel MKL (validated in Intel lab) Configuration:

Micros: Intel® Xeon® processor E5-2697 v2 @ 2.70GHz * 2 (12 cores, 24 threads); 190GB memory

1.7x spark performance thru hierachical storage (validated at customer site by Intel and customer together). Configuration:

Before: 4 nodes each with Intel® Xeon® Processor E5-2650 v3 @ 2.30GHz *2 (10 cores, 20 threads); DDR4 2133 MHz 128GB; 7x HDD, 10Gbps NIC
After: 4 nodes each with Intel® Xeon® Processor E5-2650 v3 @ 2.30GHz *2 (10 cores, 20 threads); DDR4 2133 MHz 128GB; 7x HDD + 1x Intel SSD DC P3700 Series 1.6TB, 10Gbps NIC

1.35X Spark shuffle RPC encryption performance for terasort and 1.12x performance for bigbench (both in Intel lab) Configuration:

Terasort (same for before and after): 3 nodes each with Intel® Xeon® Processor E5-2699 v3 @ 2.30GHz *2 (18 cores, 36 threads); 128GB; 4x SSD; 10Gbps NIC
BigBench (same for before and after): 6 nodes each with Intel® Xeon® Processor E5-2699 v3 @ 2.30GHz *2 (18 cores, 36 threads); 256GB; 1x SSD; 8x SATA HDD 3TB, 10Gbps NIC

Intel, the Intel logo, Xeon, 3D Xpoint are trademarks of Intel Corporation in the U.S. and/or other countries.

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

© 2016 Intel Corporation