TECHNOLOGY BRIEF

Data Center
Big Data Processing

Optimizing Apache Spark* to Maximize Workload Throughput

Apache Spark* throughput doubled and runtime was reduced by 40% with Intel® Optane™ SSD DC P4800X and Intel® Memory Drive Technology.

Executive Summary

Apache Spark* is a popular data processing engine designed to execute advanced analytics on very large data sets which are common in today's enterprise use cases involving Cloud-based services, IoT, and Machine Learning. Spark implements a general-purpose clustered computing framework that can ingest and process real-time streams of very large data, enabling instantaneous event- and exception-handling, analytics, and decision-making for responsive user interaction.

To enable Spark's high performance for different workloads (e.g. machine-learning applications), in-memory data storage capabilities are built right in. Consequently, Spark significantly outperforms the alternative big data processing technologies. However, Spark's in-memory capabilities are limited by the memory available in the server; it is common for computing resources to be idle during the execution of a Spark job, even though the system's memory is saturated. To mitigate this limitation, Spark's distributed architecture can run on a cluster of nodes, thus taking advantage of the memory available across all nodes. While employing additional nodes would solve the server DRAM capacity problem, it does so at an increased cost. DRAM is not only expensive, it also requires that the operator provision additional servers in order to procure the additional memory.

Intel® Memory Drive Technology is a software-defined memory (SDM) technology, which combined with an Intel® Optane™ SSD, expands the system's memory. This combination of Intel® Optane™ SSD with Intel Memory Drive Technology alleviates those memory limitations that are inherent to Spark applications, by making more memory available to the operating system and to Spark jobs, transparently. To demonstrate this feature, Intel used a readily available Spark benchmark named TeraSort.[1] The initial value of Intel Memory Drive Technology demonstrated by this benchmark is increased utilization for higher performance.

With this memory extension approach, system memory is larger (with the addition of Intel Memory Drive Technology), and more of the system's compute capacity is harnessed (by running more Spark executors). This benchmark demonstrates that on a system with identical size of memory and computing power, Spark job throughput can be doubled by adding Intel Memory Drive Technology software. The alternative to adding Intel Memory Drive Technology is adding more DRAM to the system. As shown in the benchmark results in Figure 3, adding more DRAM provides only a slight increase in performance, while the added DRAM is much more expensive than the Intel Memory Drive Technology alternative.[2]

The goal of this technology brief is to compare both of these alternatives side by side, identify the performance gains, then contrast them with the total cost of ownership (TCO) gains.

Author

Ravi Durgavajhala
SSD Solutions Architect

Benchmark Methodology

TeraSort* is a popular benchmark that measures the amount of time it takes to sort one terabyte of randomly distributed data on a given computer system. It started as a frequently used method to measure MapReduce* performance of an Apache Hadoop* cluster, and there are variations of it for use with Spark. Incoming data must be sorted before it can be analyzed or manipulated, making the sort performance crucial – which explains the popularity of this benchmark suite.

System Configuration

Table 1 describes the system configurations for the three different scenarios tested. The three configurations include: baseline DRAM configuration; baseline plus Intel Memory Drive Technology to increase the memory capacity; and a comparison with an increase in DRAM only.

Table 1: Comparison Configurations

Baseline (128GB DRAM)	Alternative 1 (128GB DRAM + 2x Intel® Optane™ SSD DC P4800X/Intel® Memory Drive Technology)	Alternative 2 (768GB DRAM)
• Server based on two Intel® Xeon® processors E5-2699 v4 (22 core, 3.60 GHz with Intel® Turbo Boost Technology) • Hyper threading turned off • 128GB system memory (DRAM only) • Six Intel® SSD Data Center S3500 Series (SATA) of 1.6TB for storage • Red Hat Enterprise Linux 7.3* • Hortonworks Data Platform 2.4* • Intel® Memory Drive Technology 8.1.1145.22 • Spark 1.6.2 • Oracle Java 8 Update 60*	• Server based on two Intel® Xeon® processors E5-2699 v4 (22 core, 3.60 GHz with Intel® Turbo Boost Technology) • Hyperthreading turned off • Total system memory 768GB (128GB DRAM + 2 x 320GB Intel® Optane™ SSD DC P4800X) • Six Intel® SSD Data Center S3500 Series (SATA) of 1.6TB for storage • Red Hat Enterprise Linux 7.3* • Hortonworks Data Platform 2.4* • Intel® Memory Drive Technology 8.1.1145.22 • Spark 1.6.2 • Oracle Java 8 Update 60*	• Server based on two Intel® Xeon® processors E5-2699 v4 (22 core, 3.60 GHz with Intel® Turbo Boost Technology) • Hyper threading turned off • Total system memory 768GB (DRAM only) • Six Intel® SSD Data Center S3500 Series (SATA) of 1.6TB for storage • Red Hat Enterprise Linux 7.3* • Hortonworks Data Platform 2.4* • Intel® Memory Drive Technology 8.1.1145.22 • Spark 1.6.2 • Oracle Java 8 Update 60*

Testing Approach

Test data was generated at four sizes: 100GB, 250GB, 500GB, and 1TB, using three different executor counts.

Figure 1: Software Stack

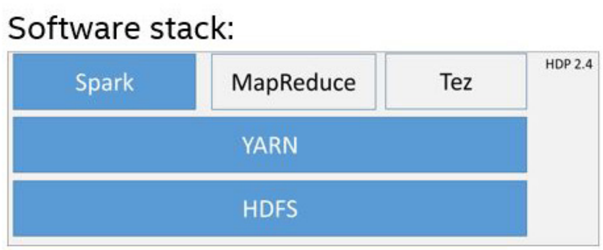
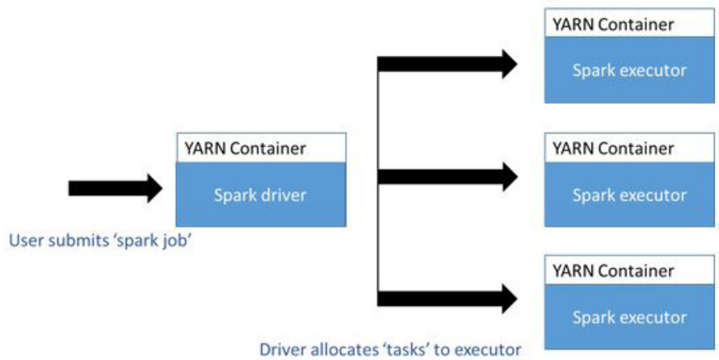
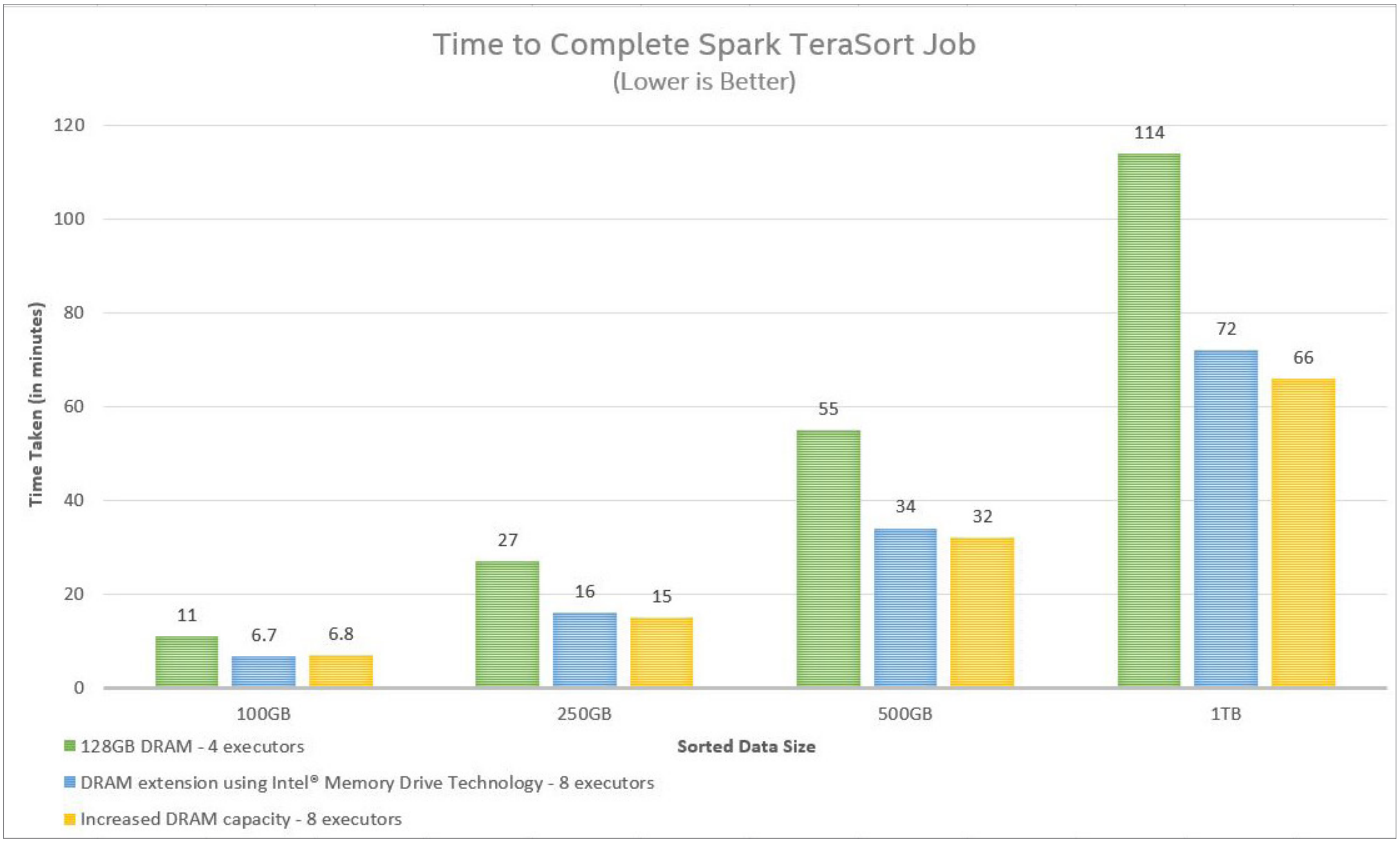


Figure 2: Spark Executor Processes



Spark driver and executors are JVM (Java virtual machine) processes. Cores and memory used by Spark executors are configurable; 7.5GB for the Spark driver and 21GB for the Spark executor were used in these tests.

Figure 3: Benchmark Results



Conclusion

Testing indicates that by adding two Intel® Optane™ SSD DC P4800X with Intel® Memory Drive Technology to a single server node running a Spark-based TeraSort workload, throughput can be doubled, while runtime is reduced by up to 40%. By adding extra DRAM to the system, a slight performance gain can be realized. However, this slight performance gain of up to 6% comes at approximately 50% higher cost. Comparing the cost of Intel Memory Drive Technology software (approximately half the cost of DRAM at the time of this writing[3]) and high capacity (Intel Memory Drive Technology enables the addition of 1280-3200 GB of system memory in a dual-socket node[4]) – Intel Memory Drive Technology effectively leads in TCO.

To learn more about Apache Spark, visit http://spark.apache.org

For more information, visit intel.com/datacenterssd

1. See http://sortbenchmark.org; Intel used this implementation for this paper: https://github.com/ehiggs/spark-terasort.
2. Up to 6% performance improvement, as shown in the Figure 3 in this document
3. Approximate DRAM cost is $10/GB, compared to $5.06/GB for Intel® Optane™ SSD DC P4800X with Intel Memory Drive Technology. Source: www.newegg.com 11/27/17
4. See Table 3, *Maximum Software-defined Memory (SDM) capacity for Intel Optane SSDs and Intel® Xeon® E5 v3/v4 processors*, in the Intel® Memory Drive Technology Set-up and Configuration Guide, for details.