Cisco Publishes First-Ever Industry-Standard Benchmark Results for Big Data Systems

Performance Brief
January 2015

Highlights

First-Ever Results
- Cisco extends its performance leadership into big data by publishing the first TPCx-HS benchmark results.
- Being the first to publish demonstrates not only the performance and scalability but also the enterprise readiness of Cisco UCS® Integrated Infrastructure for Big Data, which is already in use across major industry vertical markets

Industry-Leading Performance with MapR
- The MapR Distribution including Apache Hadoop uniquely overcomes the architectural bottlenecks in the standard Hadoop Distributed File System (HDFS) to achieve the high level of performance shown in this benchmark.
- Performance is increased and costs are reduced by the MapR distributed metadata architecture, enabling a superior level of scalability and supporting the creation of thousands of files per second, and up to one trillion files in a single cluster.

Linear Scalability
- The benchmark results show linear scalability at the scale factors tested, demonstrating how you can scale your Cisco UCS big data deployments as your business needs grow.

Complete Solution from Cisco
- The solution provides industry-leading scalability and manageability with all hardware and software components from Cisco.
- Cisco offers management tools and switching products that help you scale to thousands of servers in a single data center or distributed around the world, all with the same automated provisioning.

Cisco extends performance leadership to big data with TPC Express Benchmark HS results on Cisco UCS® Integrated Infrastructure for Big Data.

The capability to make business decisions based on massive amounts of data is increasingly important. But without objective evidence to compare big data implementations, you could be left without guidance when making purchasing decisions. By posting the first certified results ever on the new TPC Express Benchmark HS (TPCx-HS), Cisco demonstrates the real-world performance and scalability of Cisco UCS Integrated Infrastructure for Big Data (Cisco UCS Common Platform Architecture [CPA] for Big Data Version 2) for enterprise Hadoop deployments.

First Industry-Standard Benchmark for Big Data Systems

The new TPCx-HS benchmark was developed by the Transaction Processing Council (TPC) in response to the growing need for a vendor-neutral way to evaluate the performance and price-to-performance ratio of big data implementations. TPCx-HS provides an objective measure of hardware, operating system, and commercial software distributions compatible with the Apache Hadoop Distributed File System (HDFS) API. The benchmark models continuous system availability, 24 hours a day, 7 days a week. The TPCx-HS benchmark helps assess a broad range of system topologies (including clustered and nonclustered) and implementation methodologies in a technically rigorous, directly comparable, vendor-neutral manner.

Industry's First Ever TPCx-HS Publication

Cisco is the first vendor to publish TPCx-HS benchmark results. This achievement demonstrates not only the enterprise readiness of Cisco UCS for big data but also Cisco's commitment to industry standards. Table 1 describes three benchmark results using Cisco UCS C240 M3 Rack Servers running Red Hat Enterprise Linux and MapR M5 Enterprise Edition.

Table 1. TPCx-HS Benchmark Results

System	Scale	Performance	Price/Performance	Availability
Cisco UCS CPA for Big Data v2 with 16 Cisco UCS C240 M3 Rack Servers	1 TB	5.07 HSph@1TB[1]	US$121,231.76/HSph@1TB	January 8, 2015
	3 TB	5.10 HSph@3TB[2]	US$120,518.63/HSph@3TB	
	10 TB	5.77 HSph@10TB[3]	US$106,524.00/HSph@10TB	

1. See http://www.tpc.org/5500. 2. See http://www.tpc.org/5501. 3. See http://www.tpc.org/5502.

Configuration

The benchmark configuration uses the Cisco UCS CPA for Big Data v2 with 16 Cisco UCS C240 M3 Rack Servers interconnected using two Cisco UCS 6296 96-Port Fabric Interconnects with embedded management using Cisco UCS Manager:

- **The Cisco UCS C240 M3 Rack Server** is a 2-rack-unit (2RU) server designed for a wide range of computing, I/O, and storage-capacity demands. Powered by two Intel® Xeon® processor E5-2600 series CPUs, the server supports up to 768 GB of main memory and up to 24 Small Form Factor (SFF) disk drives. The server uses Cisco UCS virtual interface cards (VICs) to provide high-bandwidth and low-latency cluster connectivity with support for up to 256 virtual devices. The tested configuration consists of 16 Cisco UCS C240 M3 servers each with two Intel Xeon processor E5-2660 v2 CPUs, 256 GB of memory, and 24 1-TB 7200-rpm SFF disk drives.

- **Cisco UCS 6296 96-Port Fabric Interconnects** provide high-bandwidth, low-latency connectivity for servers and centralized management for all connected devices with Cisco UCS Manager. These devices provide active-active redundancy, high performance, and exceptional scalability for the large number of nodes typical in big data clusters. Cisco UCS Manager enables rapid and consistent server integration using Cisco UCS service profiles which automate ongoing system maintenance activities such as firmware update operations across the entire cluster as a single operation, advanced monitoring, and option to raise alarms and send notifications about the health of the entire cluster.

Cisco UCS Integrated Infrastructure for Big Data

Cisco UCS Integrated Infrastructure for Big Data can scale to up to 160 servers in a single Cisco UCS domain. Scale beyond 160 servers can be accomplished by interconnecting multiple Cisco UCS domains using Cisco Nexus® 7000 or 9000 Series Switches. This capability increases scalability to thousands of servers and hundreds of petabytes of storage, all managed from a single pane using Cisco UCS Central Software in a data center or distributed globally.

MapR M5 Enterprise Edition

The MapR Distribution including Apache Hadoop was the Hadoop platform used in this test. The MapR Distribution delivers the architectural innovations required to overcome the typical computing and storage bottlenecks and achieve a high level of performance.

Conclusion

The new TPCx-HS benchmark is designed to provide verifiable performance and price/performance metrics for big data systems. By publishing the first set of results, Cisco demonstrates the performance and price/performance ratio for real-life representative workloads, establishing a new trend in the world of performance benchmarking

For More Information

- Read the Cisco blog on the TPCx-HS publication.

- Read the Cisco blog on the TPCx-HS benchmark.

- View the TPC Press-Release on the new TPCx-HS benchmark.

- Review the TPCx-HS benchmark specification.

- Learn more about Cisco UCS performance.

Disclosures

The Transaction Processing Performance Council (TPC) is a nonprofit corporation founded to define transaction processing and database benchmarks, and to disseminate objective and verifiable performance data to the industry. TPC membership includes major hardware and software companies. The performance results described in this document are derived from detailed benchmark results available as of January 8, 2015, at http://www.tpc.org/tpcx-hs/results/tpcxhs_perf_results.asp