Cisco Launches the First Flash-Enhanced Solution for Hadoop with a Leading Yahoo Cloud Serving Benchmark Result

Performance Brief
October 2013

Highlights

Leading Scalability for Big Data Workloads
- Cisco UCS® Common Platform Architecture (CPA) v2 for Big Data offers industry-leading performance, scalability, and cost built on Cisco Unified Computing System™ (Cisco UCS) fundamentals.

Transparent Flash Acceleration
- As an intelligent and transparent cache accelerator, the Cisco® Nytro MegaRAID 200-GB controller can significantly improve application performance and can also host the operating system, freeing disk drives for data and storing frequently accessed files.

Rapid and Consistent Deployment
- Integrated with the Cisco UCS CPA v2 for Big Data, Cisco UCS Solution Accelerator Paks deliver exceptional performance, agility, and scale for big data deployments.

A Choice of Hadoop Distributions
- Cisco UCS supports leading Hadoop distributions, including those from Cloudera, HortonWorks, Intel, MapR, and Pivotal.

Cisco UCS Common Platform Architecture (CPA) Version 2 (v2) for Big Data with the MapR Distribution for Apache Hadoop is a high-performance solution for enterprise Hadoop deployments.

While big data is becoming a major part of the enterprise application ecosystem, performance and scalability remain critical to success. To address this need, Cisco recently announced support for transparent storage caching with the Cisco UCS CPA v2 for Big Data. In partnership with MapR, the solution achieves performance, scalability, and consistent and extremely low latency for Hadoop workloads.

Unleashing the MapR Distribution for Apache Hadoop

As a Hadoop technology leader, MapR offers a comprehensive Hadoop platform that is fully optimized for performance and scalability. MapR delivers unprecedented performance across various Hadoop applications, including MapReduce and NoSQL, as well as search and streaming applications. With a host of technical innovations, MapR makes Hadoop easy, dependable, and fast.

Cisco UCS CPA v2 for Big Data

Cisco Unified Computing System (Cisco UCS) has been highly successful in unifying computing, networking, management, virtualization, and storage access resources into a single integrated architecture. Cisco UCS CPA for Big Data extends the strengths of Cisco UCS: It offers computing and network scalability, performance, management, and monitoring that yield essential operational simplification, modularity, risk reduction, and lower total cost of ownership (TCO). Integral Cisco UCS fabric interconnects offer a common management plane for scale-out designs with both single-rack and multirack form factors. Cisco UCS virtual interface cards (VICs) enable unified fabric management and direct SAN access. Now Cisco UCS CPA v2 for Big Data offers a choice of infrastructure options, including "balanced", "high capacity", and "high capacity with flash" to support a range of workload needs.

Yahoo Cloud Serving Benchmark

With a wealth of available solutions, it can be difficult to decide which system is right for a given application. The Yahoo Cloud Serving Benchmark (YCSB) was

designed with a framework and a common set of workloads for evaluating the performance of Hadoop clusters. The assigned core workloads yield a well-rounded picture of system performance.

Industry-Leading Performance with the Cisco Nytro MegaRAID Controller

To evaluate flash-enhanced performance, Cisco engineers recently deployed a YCSB configuration[1] against a Cisco UCS CPA v2 for Big Data–based cluster of 16 nodes running MapR. The implementation employed the Cisco Nytro MegaRAID 200-GB controller for flash acceleration. Based on LSI's Nytro technology, the card is the industry's only enterprise-class flash storage solution with built-in support for direct-attached storage and advanced data protection to accelerate big data applications. Each node in the cluster consisted of a Cisco UCS C240 M3 Rack Server with two Intel® Xeon® E5-2660 v2 processors, 256 GB of memory, a Cisco Nytro MegaRAID 200-GB controller, and 12 4-TB large form factor (LFF) disk drives.

As shown in Figure 1, the YCSB results demonstrated linearity across the 4-, 8-, and 16-node tests. The Load test consistently showed an average latency
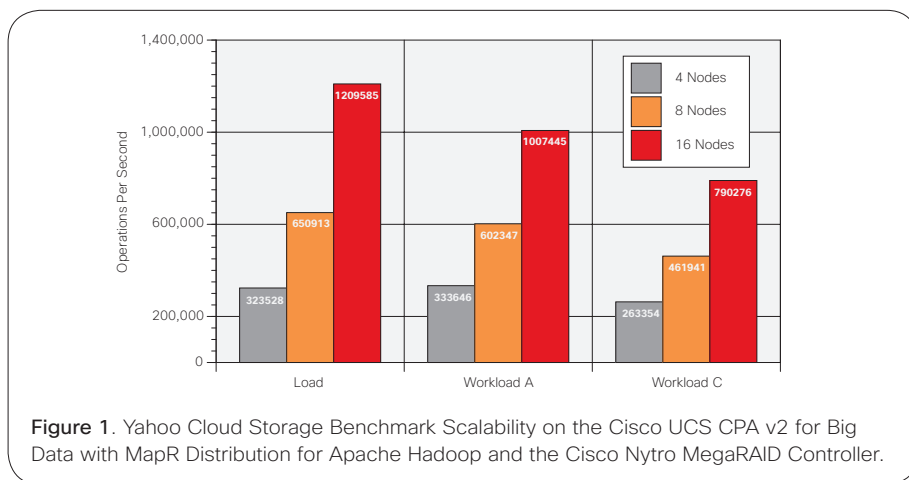

Figure 1. Yahoo Cloud Storage Benchmark Scalability on the Cisco UCS CPA v2 for Big Data with MapR Distribution for Apache Hadoop and the Cisco Nytro MegaRAID Controller.

of 0.42 milliseconds. The update-heavy Workload A showed an average update latency of 0.03 ms. The read-only Workload C showed an average read latency of 1.40 ms.

Conclusion

Cisco UCS CPA v2 for Big Data is an excellent platform for enterprise-class big data deployments. Leading flash technology and the high-performance MapR Distribution easily combine with this well-designed architecture. Together these technologies deliver significant performance advantages and lower cost of ownership, along with essential scaling capabilities across a broad range of Hadoop-based applications.

For More Information

For more information about big data, please and Cisco UCS performance, please visit:

· blogs.cisco.com/datacenter/cpav2
· www.cisco.com/go/bigdata
· www.cisco.com/go/bigdata_design
· www.cisco.com/go/ucsatwork

1. A YCSB record size of 400 bytes (10 fields of 40 bytes each) was chosen to present a more substantial workload to the cluster.