Editorial contacts

Terri Molini, HP
+1 415 480 4355
terri.molini@hp.com

Erin Dendinger
Burson Marsteller for HP
+1 312 596 3419
erin.dendinger@bm.com

www.hp.com/go/newsroom

Fact Sheet

HP Unleashes the Power of Hadoop

Industry's first enterprise-ready Hadoop solution

Overview

Organizations are experiencing dramatic growth in the variety, velocity and volume of information generated by their organizations. This leaves many enterprises struggling with how to manage vast and diverse quantities of traditional structured data as well as a new category of content consisting of large unstructured data sets of information. Together these are referred to as: Big Data. Enterprises realize the potential for insight from collected information to drive fast, actionable insight, but they need data management and analysis tools that are effective with today's technologies.

To help address these challenges, many organizations are implementing Apache™ Hadoop.™ Hadoop is an open-source-distributed data-processing technology that takes advantage of large clusters of industry-standard servers to create a single, highly available environment capable of storing and managing petabytes of information.

While the Hadoop platform offers many advantages, the technology needs an enterprise-ready analytics platform together with robust management tools to assist organizations in the set-up, deployment and optimization of their Hadoop-clustered environments. Only HP takes open source Hadoop software, combines it with the industry's most-advanced Converged Infrastructure including management tools to deliver massive scale-out data processing and analysis with built-in resiliency and fault tolerance.

By leveraging years of experience in designing, integrating and managing large clusters of compute nodes, HP is a leading provider of solutions and services for Hadoop environments. It is designed and optimized to work in extreme scale-out computing environments enabling clients to scale up to thousands of nodes for seamless growth. The new HP AppSystem for Apache Hadoop is a turnkey appliance that simplifies setup, provisioning, deployment and optimization of Hadoop clusters.

HP achieves No. 1 performance benchmark for Hadoop

Based on results of the industry-standard Apache Hadoop Terasort benchmark, which is designed to demonstrate real world big data workloads, the HP Apache Hadoop solution is the first to deliver industry-leading performance for a 10-terabyte (TB) dataset processed in 5,128 seconds (approximately 1.5 hours). Built on HP Converged Infrastructure consisting of an 18-node HP ProLiant Generation 8 (Gen8) DL380 cluster and HP Networking, HP solutions proved to be 3.8 times and 2.6 times faster than Oracle and SGI Hadoop offerings, respectively.[1]

Store and optimize without the complexity

The solution combines HP Converged Infrastructure and HP Insight Cluster Manager software enhances system management control and resiliency by providing tools that enable clients to move from test environments to full production. HP offers advanced administrative capabilities at the node, rack or cluster level. Additionally, clients can improve total cluster performance in real time with 3-D monitoring technology that quickly locates workload bottlenecks and delivers instant feedback. This enables organizations to spend less time determining the root cause of performance issues and more time gaining valuable insights for their enterprise.

With HP AppSystem for Apache Hadoop, clients also can:

- Reduce the cost and complexity of implementing a Hadoop cluster by rapidly deploying clusters of any size with push-button simplicity.[2]
- Analyze data on a massive scale to gain critical insight from unstructured and semi-structured data through seamless integration with HP Vertica Analytics Platform.
- Leverage industry-leading scale and performance by supporting up to five times the memory per node than our competitors with HP ProLiant Gen8 DL380 servers.[3]
- Increase network traffic performance with HP Networking 5830 switches, while providing failover redundancy in the event of system or component failure.
- Easily integrate the Hadoop platform with existing data center environments by leveraging infrastructure technologies based upon industry-standard servers.

Hadoop reference architectures set clients up for success

In addition, HP is the first to offer a family of reference architectures, tools and white papers to support the top three Hadoop distributions: Cloudera, Hortonworks and MapR.

The HP reference architectures for Apache Hadoop are pretested, preconfigured and optimized for the type of workloads typically deployed on Hadoop clusters. With HP reference architectures, clients can maintain high availability and reduce risks associated with implementing and managing Big Data within Hadoop clusters, enabling clients to remain focused on strategic data analysis.

HP services help clients plan, deploy and support Big Data environments

HP also introduced new services that provide clients with guidance and roadmaps for leveraging Big Data and Hadoop.

The HP Big Data Strategy Workshop helps clients reduce risk and accelerate decision making by providing a deep understanding of Big Data challenges and available solutions. Clients learn how to align corporate IT and business goals to identify critical success factors, as well as methods for evolving their IT infrastructure to handle growing volumes of information. HP works with clients to architect a business intelligence platform that extracts value from their data and delivers insight for marketable opportunities. The workshop also provides insights into when and how data should be archived, secured and protected.

The HP Roadmap Service for Apache Hadoop helps organizations to determine size and plan the deployment of the Hadoop platform. Taking best practices, experience and organizational considerations into account, the service develops a roadmap that helps drive the successful planning and deployment for Hadoop including risk and mitigation strategies. The service applies proven HP Reference Architecture for Apache Hadoop tailored to the needs and evolving requirements of the client.

HP Always On Support Services are available for the new HP AppSystem for Apache Hadoop and reference architectures covering the HP components; Hadoop support is available from individual Hadoop providers.

Availability

- The HP reference architectures for Apache Hadoop are scheduled to be available this month.[4] Pricing varies according to location and implementation.
- The HP AppSystem for Apache Hadoop solution, pre-loaded with Cloudera software is expected in Q4 2012. Pricing varies according to location and implementation.
- The HP Big Data Strategy Workshop and HP Roadmap Service for Apache Hadoop are currently available through HP Technology Consulting. Pricing varies according to location and implementation.

(1) As the first vendor to submit performance results for the 10TB Terasort benchmark, an 18-node cluster of HP ProLiant Gen8 DL380 servers sorted the 10TB data set in 5128 seconds, a rate of 1.99 gigabytes per second; it sorted the 100 gigabyte data set in 55 seconds at a rate of 1.82 gigabytes (GB) per second. On a per node basis, the HP ProLiant Gen8 DL380 was 3.8 times faster than Oracle's 2010 100GB result and 2.6 times faster than SGI's 100GB 2011 result. Hardware Configuration: 18 HP ProLiant DL380 Gen8 servers; Dual 6 core Intel® E5-2667 2.9GHz processors; 64 GB memory; 16 x 1 TB SAS 7.2K disks per node; 4 x 1GB Ethernet. Software Configuration: Red Hat Enterprise Linux 6.2; Java Platform, Standard Edition, JDK 6 Update 29-b11.
(2) Based on internal testing, deployed 800 nodes in 30 minutes.
(3) HP ProLiant Gen8 DL380P with 768GB memory versus Oracle Sun Fire X4270 M2 with 144GB memory.
(4) The HP Reference Architecture for Hortonworks Data Platform is expected in Q4 2012.

Intel is a trademark of Intel Corporation in the U.S. and other countries. Oracle and Java are registered trademarks of Oracle and/or its affiliates. Apache and Hadoop are either registered trademarks or trademarks of the Apache Software Foundation in the United States and/or other countries.