Brochure

Big Data, big platform, big benefits

HP Reference Architecture for

Hortonworks Data Platform

Capitalizing on a new information architecture

Companies today are actively investigating innovative ways to exploit the explosion of Big Data—structured, unstructured, semi-structured, machine generated, graph analyses, and complex statistical modeling—using data from within and outside their enterprises. The goal is to use the information to increase business responsiveness and capitalize on new business opportunities.

Many organizations find the best way to handle Big Data is to deploy a scale-out computing model. And today, HP is the leader in scale-out computing and platforms for data management—including Apache™ Hadoop™.

Did you know?

• Hadoop can process more than 20 petabytes of arbitrarily structured data per cluster.
• The largest Hadoop cluster is roughly 4000 nodes.
• The Hadoop Distributed File System is uniquely designed to reliably store and manage large volumes of data.

Supporting data-intensive distributed apps

An open-source information architecture, Hadoop enables enterprises to make better business decisions by processing vast amounts of semi-structured and unstructured data, and running multiple analyses on Hadoop-based clusters.

In many cases, however, enterprises lack the expertise required to deploy, configure, manage, and scale Hadoop clusters for optimal performance and resource utilization.

To address these challenges, HP designed a series of Reference Architectures for Hadoop that reduce the time, cost, and risk of implementing a Hadoop system. The Reference Architectures for Hadoop include the following components:

• **HP ProLiant Gen8 servers** – Powered by the HP ProActive Insight architecture, HP ProLiant DL360p and DL380e Gen8 servers continuously analyze thousands of system parameters to optimize application performance, proactively improve uptime, and provide insight into every aspect of your IT infrastructure. Built with over 150 design innovations, HP ProLiant DL360p and DL380e Gen8 servers help you eliminate common problems that can cause downtime and data loss. Innovations include Integrated Lifecycle Automation, Dynamic Workload Acceleration, Automated Energy Optimization, and Proactive Service.

• **HP Insight Cluster Management Utility (CMU)** – An efficient, robust tool for deploying and managing Linux-based nodes in large clusters, HP Insight CMU enables push-button scale-out deployment, remote cluster management, and real-time and time-series monitoring.

• **End-to-end Hadoop management** – Incorporating solutions from best-in-class HP partners

• **HP Networking** – By changing the rules of networking with the power of convergence, HP Networking offers greater security, bandwidth capacity, and speed for your Hadoop deployment.

To create an end-to-end Big Data management engine, you can integrate the Vertica 5.1 Analytics Platform with the HP Reference Architectures for Hadoop. With several customers analyzing more than a petabyte of data in real time, Vertica is proven to deliver outstanding results in the most demanding business intelligence and analytic environments.

HP solution advantage

Rather than offer you an assortment of piece-parts, the HP Reference Architectures for Hadoop enable you to order and implement scalable Hadoop clusters of varying sizes quickly and with low risk. These proven designs include:

- Various solution sizes—single and multi-rack configurations—pre-tested, pre-configured, and optimized enabling you to go from pilot to production.
- Deployment monitoring and management expertise as HP has over a decade of experience helping customers manage extremely large compute clusters.
- Consulting expertise to help you determine strategies for data management and scalability, and to remove the guesswork from planning, implementation, and management.

HP Reference Architectures for Hadoop provide a cross-cutting platform for low-cost, batch-oriented analytics across many vertical industries including:

- Risk and fraud detection models for financial services organizations and government agencies
- Customer profiles and product recommendation systems for retail
- Textual analysis of social data for consumer packaged goods
- Web crawling and information mining for healthcare

HP Technology Consulting at your side

Developing your solution is a journey that can be accomplished better with the HP Roadmap Service for Hadoop. HP Technology Consulting professionals can help you develop a common vision; understand sources of data and its sensitivity; identify organizational inhibitors; understand risks and mitigation strategies; and develop a roadmap that drives the successful planning, deployment, and support for Hadoop within your IT organization. We leverage our experience, best practices, and proven Reference Architectures—all tailored to the specific needs and drivers of your roadmap.

Best-in-class partners for Hadoop management

To deliver Hadoop-based solutions that provide the greatest value to our customers, HP partners with best-in-class software vendors. For each partner, HP offers a Reference Architecture for Hadoop—including one for Hortonworks Data Platform.

As a massively scalable, 100% open source platform for storing, processing, and analyzing large volumes of data, the Hortonworks Data Platform can handle data from many sources and formats in a quick, easy, cost-effective manner. Providing a stable and highly extensible platform, Hortonworks Data Platform makes it easier to integrate Apache Hadoop with your existing data architectures, and maximize the value of the data flowing through your business.

Hortonworks Data Platform is ideal for organizations that prefer a 100% Apache-licensed open source platform that consists of the essential Apache Hadoop projects. It is recommended for any organization that has encountered difficulty installing and integrating Hadoop projects downloaded directly from Apache, or organizations that want to avoid getting locked into a proprietary vendor's platform. The Hortonworks Data Platform is also ideal for solution providers that want to integrate or extend their solutions for Apache Hadoop—thanks to the openness and extensibility of the platform.

Inside the Reference Architecture

Using Hortonworks Data Platform and the HP ProLiant DL360p and DL380e Gen8 server platform, HP and Hortonworks jointly designed this Reference Architecture, optimized for the Hortonworks distribution. This customizable Reference Architecture features:

- HP Networking A5830 Switch, providing 1 GB of deep packet buffering for increased network performance
- Intel® Socket R C600 Chipset for increased CPU performance
- HP 420i Smart Array Controller and SAS disks for increased I/O performance

In addition to the features described above, the Reference Architecture also includes:

- HP FlexNet Networking Technology, providing enhanced flexibility and speed for your new networking infrastructure
- HP iLO Management Engine with Active Health System, offering a complete set of embedded management features for HP power/cooling, Agentless Management, Active Health System, and Intelligent Provisioning
- HP management, including HP Insight Manager and HP Insight CMU for node-level and cluster-level management
- HP Service Pack for ProLiant, easing firmware management

The recommended HP software, servers, and switches and their respective configurations have been carefully tested with a variety of I/O, CPU, network-bound, and memory-bound workloads to ensure the best possible designs—supporting optimum performance for Map/Reduce and HBase functions.

HP – bringing it all together

For your HP solution for Hadoop with Hortonworks Data Platform, one hardware platform offers all the flexibility, scalability, and resiliency you need: HP Converged Infrastructure—where all IT resources are pooled and shared enterprise-wide. By overcoming the inflexibility and high costs created by IT sprawl, HP Converged Infrastructure solutions enable you to shift more resources to innovation and strategic initiatives. The HP approach and solutions have been designed to help make the data center simpler, more efficient, and less expensive to operate. This way, your IT infrastructure is positioned to deliver any workload, anywhere, anytime to accelerate the time to application and business value.

Leverage the power of your Big Data

Contact your HP representative today. Arrange for an assessment of your IT environment. Discover how HP Reference Architectures can help you leverage the power of Big Data and Apache Hadoop. For more information, visit:

hp.com/go/cisystems

vertica.com

hortonworks.com

Get connected
hp.com/go/getconnected

Get the insider view on tech trends, support alerts, and HP solutions.

Share with colleagues