nimblestorage

Accelerating Hortonworks with Nimble Storage

Nimble Storage's Adaptive Flash accelerates Hortonworks® Data Platform (HDP) deployments. Adaptive Flash arrays deliver speed and capacity, are easy to deploy and manage, and offer integrated data protection.

The IT landscape has changed. The amount of unstructured data being created in the form data-generating sensor equipped and networked devices, both machine- and human-generated online transactions, and other sources of unstructured and semi-structured data is 10 times greater and growing 10 times faster than structured business data. Within this exploding data repository are important insights for companies able to analyze and act on it quickly.

The Challenge

Today's businesses must process massive amounts of structured, semi-structured, and unstructured data at unprecedented speed, and with a finer level of granularity than ever before. To do that, they need an advanced platform. Products like Hortonworks Data Platform (HDP) are being rapidly adopted by organizations using a variety of data handling techniques from batch through interactive to real-time. At the same time, they are looking for benefits in various functional areas like governance, integration, security and operations. Nimble's Adaptive Flash solutions are ideal for these organizations that are using sophisticated HDP implementations, delivering the performance they need at price they can afford.

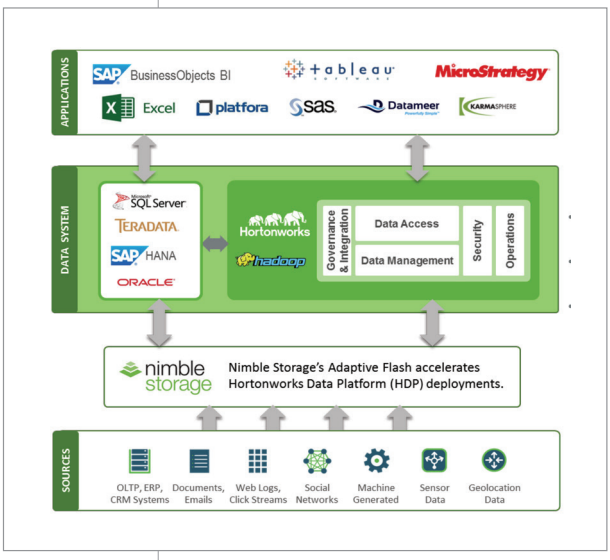
HDP: A Complete Enterprise Apache Hadoop

Apache™ Hadoop® is an open-source software framework for storage, and the processing of large-scale data sets on clusters of commodity hardware (servers.) HDP is an open source software provider to develop, distribute and support an Apache Hadoop platform explicitly architected, built and tested for enterprise-grade deployments. It was designed to meet the needs of today's modern data architecture: It allows enterprises to capture, store, and process vast quantities of data cost-efficiently to unleash a multitude of analytic opportunities and innovations.

HDP has been optimized for multiple workloads from batch through interactive to real-time with improvements for governance, integration, security and operations.

Hadoop and ETL

Extract, transform and load (ETL) systems play a critical role in gathering and loading data sets from various sources on scale-out server clusters. ETL systems are commonly used to integrate data from multiple applications, typically developed and supported by different vendors like Hortonworks, or hosted on separate computer hardware.



Solution Overview

Nimble Storage's Adaptive Flash is a powerful and flexible platform that delivers actionable intelligence with lightning speed for Hortonworks Data Platform environment. Nimble's Adaptive Flash solutions offer:

- **Cost-Efficient Performance** for near real-time business insights.
- **Efficient and Affordable Capacity** for data growth.
- **Independent and Dynamic Scaling** of capacity and/or performance.
- **Integrated Data Protection and Disaster Recovery** with frequent and fast backups.

Hortonworks

The Problem with Scale-Out Server Clusters

Organizations must maintain service level agreements (SLAs) even as they grow at a rapid pace. Most meet growth requirements by adding servers, a practice that presents the following challenges:

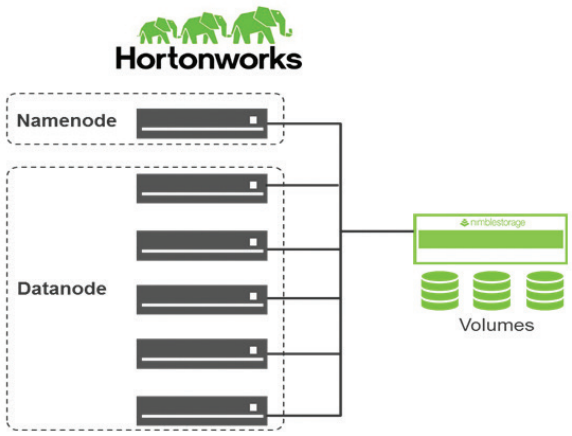
- Organizations processing terabytes of data soon realize they must analyze it quickly to gain relevant insights, and adding servers won't deliver the necessary speed.
- Hadoop clusters add considerable cost and complexity; clusters of greater than 6 servers are difficult to manage.
- Data warehouses, typically assembled from a variety of data sources with different formats and purposes, add unnecessary processing cycles to servers.
- Extra servers complicate data protection, basic three-way mirroring is not cost-effective, and site recovery is disruptive and time consuming.

IT professionals must build stand-alone grids of client computers running Hadoop to analyze data with MapReduce, a processing paradigm for data-intensive computational analysis. But before data can be analyzed, it must be moved from one or more storage systems to a Hadoop machine. Importing this data can be a complicated and time-consuming: It requires running HDFS copy operations or using specialized Hadoop connectors. And, it is repetitive: As the data expands or changes, it must be again loaded into the machines on the grid.

A Better Approach with Nimble Storage

Nimble's Adaptive Flash separates data from compute, delivering the unique performance required by Enterprise Apache Hadoop. Nimble cost-effectively speeds transitions to data-focused modern data architecture, while offering ease and flexibility.

Nimble streamlines analytics workflows, eliminating the need to extract the data from storage systems so that it can be loaded onto a traditional Hadoop environment. With Nimble, data no longer needs to be exported after it has been analyzed — a further simplification of the analytics workflow. And, application developers do not have move data onto the compute grid, directing the compute function to where data resides; i.e. the storage.



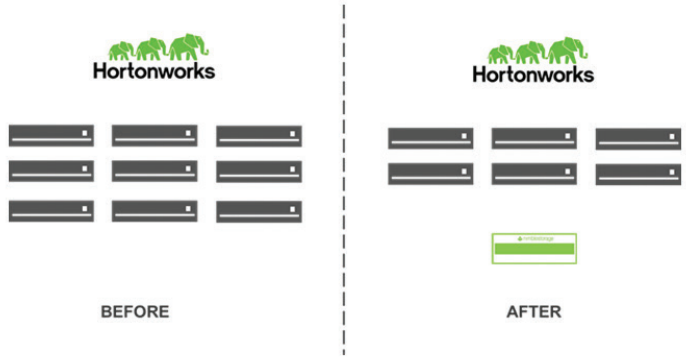
Nimble and HDP deliver cost-effective performance and capacity in systems that are easy to deploy, manage and scale, and provide integrated data protection.

Cost-Effective Performance

Nimble arrays deliver adaptive performance for HDP environments. At the core of every Nimble array is its patented Cache Accelerated Sequential Layout (CASL™) architecture, designed from the ground up to leverage the random read performance of flash. CASL delivers the performance demanded by HDP's NameNodes, a massively parallel processing (MPP) database that sits atop of the Hadoop platform. NameNodes handles millions of small interactive queries, a read-only I/O pattern that taxes the resources of traditional storage systems.

Capacity Efficiency

CASL incorporates innovative efficiency features such as inline variable-block compression, offering a 2x improvement compared to traditional storage solutions. Given the range of data types (random vs. sequential) and data sizes (8k-32K block) in HDP environments, post compression is an inefficient way to reduce capacity. Nimble's variable-length compression offers greater space reduction, and



improves performance. Furthermore, with InfoSight™, Nimble's innovative approach to management using cloud-based analytics, Nimble's solutions have achieved greater than "five nines" availability — which gives customers confidence to reduce Hadoop data copies from 3 to 2.

Ease of Management

Nimble arrays are easy to deploy and manage, able to be set up by an IT generalist. In addition, Nimble solutions free server capacity by moving data to disk, using a fraction of the capacity required by server-only HDP environment.

InfoSight is Nimble's innovative approach to management using cloud-based analytics to deliver true operational efficiency across all storage activities. InfoSight monitors all Nimble arrays, collectively and individually, from the cloud, automatically gathering millions of data points per array each day and making sense of them in real-time. Using that data, InfoSight identifies potential problems — and offers solutions — long before they can bring systems down.

Improved Scalability

Nimble solutions allow independent and non-disruptive scaling of storage performance and capacity, reducing the need to add server nodes (compute). With Nimble, enhancing performance is as easy as upgrading controllers; active data can be accommodated by simply adding SSDs; and storage capacity can be scaled to hundreds of terabytes (TBs) by adding disk shelves — resulting in a very flexible HDP environment.

Data Protection

Nimble offers maximum data protection with frequent snapshots, consistent backups, fast restores, and efficient replication for disaster recovery. Nimble solutions deliver simple and affordable data recovery, efficiently protecting months of stored data, easing the burden on server and network resources by eliminating the need for 3-way mirroring, the default for Hadoop installations. In addition, Nimble efficiently replicates data to another array by transferring compressed, block-level changes only. These remote copies can be activated even when the primary array is unavailable, so that data recovery is both simple and affordable.

Conclusion

Deployed as part of comprehensive analytics solution, Nimble delivers actionable intelligence with lightning speed. Nimble solutions are easy to deploy and use, reducing total cost of ownership (TCO), and providing an exceptional return on investment (ROI) for Hortonworks Data Platform environments.

Hortonworks CERTIFIED TECHNOLOGY PARTNER

nimblestorage

211 River Oaks Parkway, San Jose, CA 95134
Phone: 408-432-9600; 877-364-6253
Email: info@nimblestorage.com
www.nimblestorage.com

vmware READY · Windows Server 2012 Certified · cisco Compatible · CITRIX ready · ORACLE Gold Partner