NetApp®

Solution Brief

Hortonworks Data Platform and NetApp Open Solution for Hadoop

The Challenge

The emergence of big data is now well known with more data coming from sources like machines, sensors, videos, blogs, social media, and mobile devices. Businesses are collecting more data than ever and consumers are producing more data than ever. Big data is about opportunity and a vital part is having a strategy around analytic tools so businesses can find value in the structured and unstructured data to create differentiated customer benefits and gain competitive advantage. Hadoop is a powerful and essential tool for big data and has extended business analytics to go beyond traditional BI tools. However there are challenges with Hadoop as there are with any new technology. Some of these are:

- Can take significant time to become productive with Hadoop
- Not easy to use or operate with existing staff or skillsets
- Hadoop users typically create different data types and sharing diverse data can be difficult as there is no common schema between Hadoop Projects Pig, Hive or MapReduce

- Concerns about Hadoop reliability: if the NameNode fails, Hadoop goes down so addressing this single point of failure is important
- Standard Hadoop uses 3x software mirroring or making three copies of the data, requiring additional storage and reducing cluster throughput

Hortonworks has partnered with NetApp to address these challenges. The Hortonworks Data Platform (HDP) powered by Apache Hadoop, delivers a comprehensive Hadoop-based software stack and NetApp Open Solution for Hadoop delivers the infrastructure for running HDP.

The Solution

NetApp Open Solution for Hadoop with HDP delivers enterprise grade Hadoop that's easy to setup, deploy and manage. The components are:

FAS series provides high reliability of the Name Node with reduced single points of failure, faster recovery time, and metadata Name Node protection with hardware RAID.
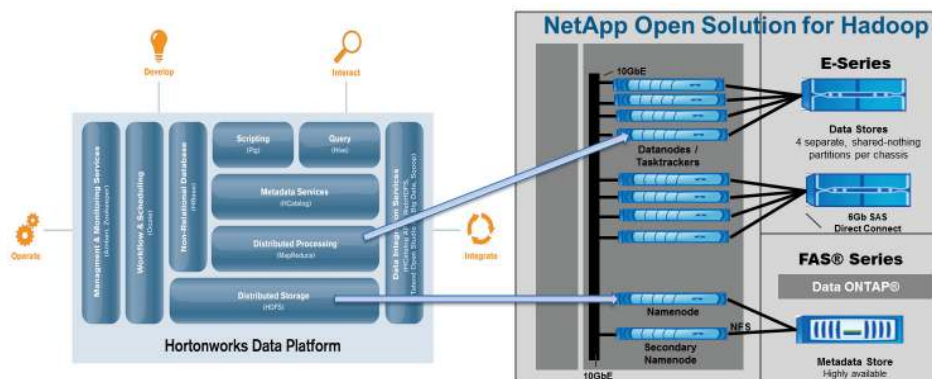
Figure 1) Hortonworks Data Platform connecting to NetApp Open Solution for Hadoop.

NetApp E-series provides increased performance and bandwidth for Hadoop clusters along with higher storage efficiency and scalability.

Hortonworks Data Platform provides the Apache Hadoop core components (HDFS and MapReduce) and essential Hadoop components (HBase, Hive, Pig, HCatalog, Zookeeper) so data in any format at any scale can be stored, processed and analyzed.

HDP makes Hadoop even more open and extensible, by providing APIs to work with Ambari, Squoop, Oozie, and Mahout.

Extract Value from Any Kind of Data—Structured, Semistructured, Unstructured

HDP ingests and persistently stores data in any format, aggregates and refines a wide variety of data from multiple formats into one, and performs complete data analysis while leaving the source data in place for later reprocessing. HCatalog Metadata catalog management allows you to treat Hadoop as structured as you need.

High Reliable Hadoop with Ease of Use in Your Enterprise

Hortonworks and NetApp's joint solutions solve two of the biggest challenges of implementing Hadoop in the enterprise: simplifying its operation and making it

SERVICE COMPONENTS	
Hortonworks Products	**NetApp Products**
Hortonworks Development Platform (HDP)	NetApp FAS running Data ONTAP
HDP Universe	NetApp E-series with SANtricity software

reliable. HDP delivers centralized cluster management and enterprise class data integration capabilities to simplify Hadoop use and NetApp delivers enterprise reliable storage to run Hadoop clusters with improved failure handling and recovery, excellent Name Node protection, and double parity RAID 6 on a Data ONTAP storage platform operating system.

Broad Support, Services and Training from the Experts

Hortonworks support organization is backed by leading contributors to the Apache Hadoop open source projects as Hortonworks engineers have contributed more than half of all code to core Hadoop and are uniquely qualified to address technical issues. Hortonworks delivers full life-cycle support to ensure customer success at every step including development, proof-of-concept and staging. Hortonworks ensures patches and bug fixes are contributed back to Hadoop and included in future editions so they are leading vendor to provide

forward-compatibility. NetApp support can provide non-stop operations for the Hadoop cluster, professional services and seamless third party vendor support.

About Hortonworks

Hortonworks is a leading commercial vendor of Hadoop, and Hortonworks Data Platform provides an open and stable foundation to build big data solutions and enable a partner ecosystem. Hortonworks makes Hadoop easier to manage, and provides unmatched training support.

About NetApp

NetApp creates innovative storage and data management solutions that deliver outstanding cost efficiency and accelerate business breakthroughs. Discover our passion for helping companies around the world go further, faster at *www.netapp.com*.

Go further, faster ®

www.netapp.com