DELLEMC

Hortonworks

DELL EMC HORTONWORKS HADOOP SOLUTION

A solution to simplify the design and architecture of Hadoop DAS environments

ACCELERATE TIME TO VALUE

The digital transformation is causing churn, uncertainty and disruption for many business leaders, who must act as the pressure is increasing from all directions. Big data and analytics will be at the core to enable this transformation, with Hadoop being a foundational component of a big data and analytics solution stack.

While the Hadoop platform offers unique value for organizations seeking to capitalize on big data, it can be a challenge to implement. At Dell EMC, we give you the option to buy your Hadoop ecosystem or build it out yourself. We recognize that the upfront work to build it yourself can be time-consuming and labor intensive. One way we help is to address those challenges by providing certified Hadoop solutions and the expertise needed to accelerate your time to value.

This goal of faster time to value is at the heart of the Dell EMC Hortonworks Hadoop Solution. It simplifies the architecture, design, configuration and deployment of Hadoop environments. Dell EMC engineers have validated and certified Hortonworks Data Platform (HDP) 2.5 on Dell EMC PowerEdge™ R730xd servers and Dell EMC Networking S4048ON and S3048ON, allowing your organization to build a Hadoop cluster without the guesswork. You can leverage the Dell EMC solution to streamline the front-end work, from server configuration to network setup to running HDP 2.5 on a certified solution.

Scales from 5 to 252 nodes, 3.8 PB

Hortonworks™ Data Platform

Dell OpenManage™ / iDRAC with Lifecycle Controller

Pod Network
2x Dell Networking S4048 10GbE Pod Switches
1x S3124 iDRAC Switch
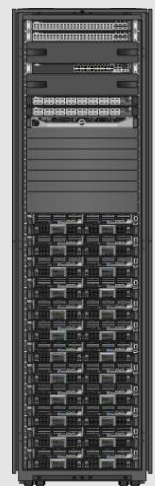
Cluster Network
2x Dell Networking S6000 40GbE Cluster Switches

Infrastructure Nodes
1x Dell PowerEdge™ R630 Admin Node
3x PowerEdge R730xd Name Nodes
1x PowerEdge R730xd Edge Node

Data Nodes
10x PowerEdge R730xd with 3.5 Drives – 48 TB or 10x PowerEdge R730xd with 2.5" Drives – 24TB

Decrease time to value

Hardware and expertise matter when building a Hadoop environment. Dell EMC does all the hard upfront work, allowing your organization to focus on delivering deeper insights and enhanced data-driven decision making. Dell EMC simplifies the architecture, design, planning and configuration for you, saving your organization valuable time and resources.

In addition, you have the assurance that Dell EMC engineers have certified the server and network configuration running Hortonworks HDP 2.5. Dell EMC develops an architecture document that provides guidance and know-how to help you successfully build a Hadoop cluster from bare-metal hardware.

REDUCE THE RISK

Up to 70 percent of Hadoop deployments will fail to meet cost savings and revenue generation objectives due to skills and integration challenges.* Dell EMC helps your organization reduce this risk. We have been building Hadoop architectures since 2011. You can leverage our expertise to help fill the skills gap and build an architecture that will meet the needs of the business.

Dell EMC certifies the Hadoop architecture by running and validating HDP 2.5 on Dell EMC PowerEdge R730xd servers and Dell EMC Networking S4048ON and S3048ON, allowing your organization to build a Hadoop cluster without the guesswork.

* Gartner. Market Guide for Hadoop Distributions. January 6, 2015.

LEVERAGE PROVEN EXPERTISE

You can gain hands-on experience with the Dell EMC Hortonworks Hadoop Solution in a Dell EMC Customer Solution Center. These state-of-the-art technical labs enable Dell EMC customers to explore, architect, validate and build solutions, from the data center to the edge of the network, to drive toward targeted business outcomes.

After your exploration of the technologies, Dell EMC Services makes getting started easy. Options include custom solution design, hardware and software deployment, ongoing support and training. With Dell EMC, you have the assurance that your Hortonworks solution is backed by expert hardware and software support that can be tailored to your specific needs.

Dell EMC PowerEdge servers

Maximize server-based storage flexibility and performance with the remarkable Dell EMC PowerEdge R730xd server, part of the new 13th generation of PowerEdge servers. The R730xd offers an optimal balance of storage utilization, performance and cost with an optional in-server hybrid storage configuration that can support tiering and capacity for up to 28 drives in a 2S/2U system, including up to 18 x 1.8-inch SATA SSDs.

Dell Networking S-Series 10GbE switches

Deploy modern workloads and applications designed for the open networking era with an optimized data center top-of-rack (ToR) networking solution.

Key features:

- Delivers low latency, superb performance and high density with hardware and software redundancy

- Offers Active Fabric designs using S- or Z-Series core switches to create a two-tier, 1/10/40 GbE data center network architecture

- Provides an ideal solution for applications in high-performance data center and computing environments
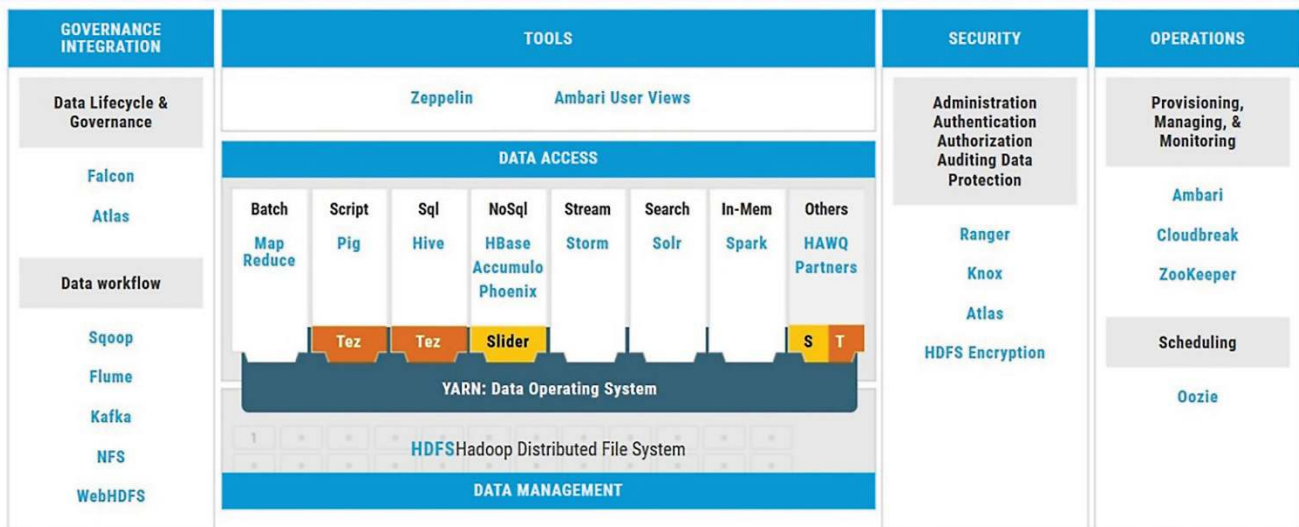
HORTONWORKS DATA PLATFORM (HDP)

Hortonworks is a leading commercial vendor for the Apache Hadoop platform. Hortonworks Data Platform, the company's distribution of Apache Hadoop, provides an open and stable foundation for enterprises and a growing ecosystem to build and deploy big data solutions.

HDP is a secure, enterprise-ready, open source Hadoop distribution based on a centralized architecture (YARN). HDP addresses the complete needs of data-at-rest, powers real-time customer applications and delivers robust analytics that accelerate decision making and innovation.

HDP is a key component of the Hortonworks Connected Data Platform, which helps organizations create actionable intelligence to transform their businesses. Whether it's data-in-motion, data-at-rest or modern data applications, HDP can power the future of data for any organization and any line of business with a variety of use cases, including data discovery, single view, predictive analytics and enterprise data warehouse optimization.

Hortonworks HDP 2.5

For security administrators and data stewards

- **Classification-based Policy**. Assign access to data assets based on reusable metadata tags such as PCI or PII

- **Location-based Policy.** Customize entitlements based on geography. A user trying to access the same data from different locations would be subject to unique geographical context.

- **Data Expiry-based Policy.** Assign expiration dates to data tag to automatically deny users access to the tagged data after the expiration date.

- **Prohibition-based Policy.** Define security policy that restricts combining two data sets to help avoid privacy violations.

- **Row Level Security and Dynamic Data Masking.** Restrict row access and anonymize sensitive data in real-time in Hive based on user characteristics and runtime context.

For Hadoop operators

- **Role-Based Access Control.** Apache Ambari 2.4 includes additional cluster operational roles to provide more granular division of control for cluster operations.

- **Log Search (Technical Preview).** Automatically configures the collection of cluster operational metrics to aid with analysis and troubleshooting by including a new Log Search service.

- **Customizable Cluster Alerts.** Tailor HDP to fit with your enterprise monitoring environment by configuring a set of predefined alerts that seamlessly integrates with your existing enterprise monitoring tools.

- **Activity Reporting and Visualization.** Activity Reporting and Visualization in Hortonworks SmartSense 1.3 (available separately) helps Hadoop operators understand how their cluster operates.

Server Architecture	Master Nodes	General Purpose Data Node	High Performance Data Node	High Capacity Data Node
R730xd Server	R730xd	R730xd	R730xd	R730xd
Processor	2 X Intel E5-2650 v4 2.2GHz (12 Core)	2 X Intel E5-2650 v4 2.2GHz (12 Core)	2 X Intel E5-2690 v4 2.6GHz (14 Core)	2 X Intel E5-2650 v4 2.2GHz (12 Core)
Memory	256GB	256GB	256GB	256GB
Network Card	Intel X520 Dual Port 10Gbe	Intel X520 Dual Port 10Gbe	Intel X520 Dual Port 10Gbe	Intel X520 Dual Port 10Gbe
Hard Drive Controller	H730	H730	H730	H730
Hard Drives	8 x 1TB 7.2K RPM SAS 12Gbps 2 x 600GB 10K RPM SAS 12Gbps (Flex Bay)	12 x 4TB 7.2K RPM SAS 12Gbps 2 x 600GB 10K RPM SAS 12Gbps (Flex Bay)	20 x 1.2TB 10K RPM SAS 12Gbps HDFS Tier 2 x 800GB Intel S3710 SSD Scratch Space 2 x 800GB Intel S3710 SSD HDFS SSD Tier 2 x 600GB 10K RPM SAS 12Gbps (Flex Bay)	12 x 4TB 7.2K RPM SAS 12Gbps 4 x 4TB 7.2K RPM SAS 12Gbps HDFS (Mid Bay) 2 x 600GB 10K RPM SAS 12Gbps (Flex Bay)

RAID Layout	Operating System- 2 HDs, RAID 1	Operating System- 2 HDs, RAID 1	Operating System- 2 HDs, RAID 1	Operating System- 2 HDs, RAID 1
	Zookeeper Journal- 1 HD, Non-Raid or RAID 0	HDFS Data- 12 or 16 HDs, Non-RAID or RAID 0	HDFS Data- 12 or 16 HDs, Non-RAID or RAID 0	HDFS Data- 12 or 16 HDs, Non-RAID or RAID 0
	NameNode Journal- 1 HD, Non-Raid or RAID 0			
	HDFS Metadata- 2 HDs, RAID 1			

Cluster Network	**Cluster Data Network**	**BMC Network**	**Edge Network**	
Network Switch	S4048ON	S3048ON	S4048ON	
Connection	Bonded 10GbE Dual Top of Rack	1GbE Dedicated Switch Per Rack	Bonded 10GbE, optionally bonded Direct to Edge Network or via aggregation switch	

Hortonworks HDP 2.5		
Data Management	Hadoop and YARN 2.7.3	
Data Access	PIG 0.16.0 Hive 1.21 Tez 0.7.0 Solr 5.2.2 Spark 1.6.2 Zeppelin 0.6.0	Slider 0.91.0 HBase 1.1.2 Phoenix 4.7.0 Accumulo 1.7.0 Storm 1.0.1
Governance and Integration	Falcon 0.10.0 Atlas 0.7.0 Sqoop 1.4.6	Flume 1.5.2 Kafka 0.10.0
Operations	Ambari 2.4.1 Cloudbreak 1.3.0	Zookeeper 3.4.6 Oozie 4.2.0
Security	Knox 0.9.0 Ranger 0.6.0	

THE USE CASE FOR A DATA LAKE

Dell EMC continues to keep customer use cases first as they offer robust Data Lake opportunities using the Hortonworks Data Platform (HDP) together with Dell EMC Isilon scale-out NAS storage running the OneFS operating system. Dell EMC provides a fully tested and complete solution encompassing enterprise grade Apache Hadoop and related projects to form a Data Lake for maximizing the value of data. The combination of HDP and Isilon shared storage helps organizations speed time to insights, improve storage utilization, eliminate islands or silos of storage, and lower storage management costs of migration, security and protection.

To learn more, visit Dell.com/Hadoop, Dell.com/BigData, or EMC.com/BigData

DATA SHEET