

# DELL EMC CLUDERA HADOOP SOLUTION

Streamline the architecture, design, planning and configuration of DAS Hadoop environments

## ACCELERATE TIME TO VALUE

The Internet of Everything is transforming the way every business operates. It is ruthlessly redrawing industries and reinventing our future. This digital transformation will cause disruption, and business leaders must act. Pressure is increasing in all directions.

Who is driving digital transformation? Lines of business, competitors, customers, and C-suite executives. Business leaders must invest in the digital transformation to stay relevant. Leaders have agreed the following digital business attributes are imperatives for success:

- Predictively spot new opportunities
- Deliver unique and personalized experiences
- Innovate in agile ways
- Operate in real time

Big data and analytics will be at the core of these strategies, and the Apache™ Hadoop® platform will be a major driver. Hadoop is an innovative technology that offers unique business value for organizations seeking to capitalize on big data. Nevertheless, at times organizations struggle to make Hadoop projects successful. Oftentimes they are hamstrung by a lack of Hadoop expertise and skillsets, and they end up spending too much time and effort on the front-end work before they can get to value of the use cases for their data.

Dell EMC, in partnership with Cloudera, has helped organizations solve the skills gap by providing expert guidance and knowledge to streamline the architecture, design, planning and configuration of Hadoop environments. Today, this expert guidance and knowledge is embodied in the Dell EMC Cloudera Hadoop Solution.

### Dell EMC Hadoop expertise

Dell started custom designing and building the first Hadoop server platforms for the largest Web 2.0 companies in 2009 through the Dell Data Center Solutions Group (DCS), created in 2008 as a startup within Dell. Dell has been talking to the leading innovators in big data since then and has turned that experience into big data expertise that enables Dell EMC to make Hadoop as simple as possible for customers. Dell was the first server provider to build server hardware optimized to run customers' demanding Hadoop workloads.

The Dell DCS FS12 server was the first purpose-built platform for Hadoop. As Hadoop workloads evolved, the Dell PowerEdge™ C2100 server was created. The PowerEdge C2100 platform would eventually lead to the development of the Dell PowerEdge™ R720xd server. In 2014, the next-generation Hadoop platform was introduced — the Dell PowerEdge R730xd server — to provide an ideal, flexible building block for Hadoop workloads. Now we add the PowerEdge FX2, a 2U hybrid rack-based computing platform that combines the density and efficiencies of blades with the simplicity and cost benefits of rack-based systems.

### Scales from 5 to 252 nodes, 3.8 PB

- Cloudera™ Enterprise
- 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 FC630 Name Nodes  
1x PowerEdge FC630 Edge Node
- Data Nodes**  
20x PowerEdge FC630 / FD332 – 32 TB or  
40x PowerEdge FC430 / FD332 – 16 TB



## DELL EMC CLUDERA HADOOP SOLUTION

The Dell EMC Cloudera Hadoop Solution based on Version 5.9 of the Cloudera distribution of Hadoop (CDH 5.9) will be the 18th and 19th iterations of our proven architectures based on expert engineering and full validation. Dell EMC engineers have validated and certified CDH 5.9 on Dell EMC PowerEdge R730xd servers with Intel® Xeon® processors and also on the PowerEdge FX2 FC630 server nodes with FD332 storage blocks and Networking S4048ON and S3048ON.

The result is two solutions that allow organizations to build Hadoop clusters without the guesswork. You can leverage the Dell EMC solution to streamline the entire process, from bare-metal server configuration to network setup to running CDH 5.9 on a certified architecture.

### ACCELERATE TIME TO VALUE

Infrastructure and expertise matter when building a Hadoop DAS environment. You get both when you leverage the Dell EMC validated solutions. They allow you to focus on using Hadoop to gain deeper insights and enhance data-driven decision making, rather than designing a solution from scratch.

You can simultaneously reduce the risk that comes with technology projects and solve the skills gap by leveraging the Dell EMC Cloudera Hadoop Solutions. Dell EMC engineers have certified and validated the full solutions, starting from the server BIOS and firmware settings to the network configuration, running Cloudera CDH 5.9.

Dell EMC develops an architecture document that provides the guidance to help you build a Hadoop DAS cluster from bare metal hardware. With the expertise and experience of Dell EMC to take you through the steps, you have the outline to help you to save valuable time and resources — as you accelerate time to value.

### REDUCE THE RISK

According to Gartner, through 2018, 70 percent of Hadoop deployments will fail to meet cost savings and revenue-generation objectives due to skills and integration challenges.<sup>1</sup> The high failure rate has left many organizations wary of adopting Hadoop, yet the demands of the business will continue to require big data technologies. With this knowledge, Dell EMC recognized the need to assist and enable organizations to lower the risk using a tested, validated and certified plan for implementing Hadoop.

Dell EMC has been building Hadoop DAS architectures for more than six years, and you can leverage that experience and expertise to fill the skills gap and build an architecture that meets the needs of your business, all while reducing the risks that come with technology projects.

### CONTROL COSTS

The Dell EMC Cloudera Hadoop Solution has proven value. To understand the value of the Dell EMC Cloudera Hadoop Solution for organizations, Forrester conducted a Total Economic Impact™ Study through one-on-one interviews with four Dell EMC Hadoop customers. Based on the TEI model, Forrester produced the risk-adjusted results that showed in three years the Dell EMC Cloudera Hadoop Solution delivered an ROI of \$1.97 return per \$1.00 invested, a net present value of \$5.6 million, and payback within six months.

### DELL POWEREDGE SERVERS

Maximize server-based storage flexibility and performance with the Intel-based Dell PowerEdge R730xd server, part of the 13th generation of Dell EMC PowerEdge servers. The PowerEdge R730xd server 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 1.8-inch SATA SSDs.

Scale workloads quickly, as needed, adding resources incrementally without the expense and inefficiency of overprovisioning. The groundbreaking FX architecture combines efficient management with flexible data center building blocks to optimize Hadoop workloads.

Combine various modules with the FX2 chassis, a compact 2U rack-based enclosure that shares cooling, power, networking and PCI expansion slots, to quickly deploy the exact combination of compute, storage and networking resources you need to best fulfill your business requirements.

FX2 solution components include:

- Server blocks at the heart of the FX converged architecture powered by the latest Intel® Xeon® processors
- FC630 server nodes — 2-socket, half-width 1U workhorse server blocks ideal for a wide variety of business applications
- FD332 storage blocks — flexible, high-density, half-width 1U storage modules that enable you to rapidly scale direct attached storage (DAS) in your FX-based infrastructures

<sup>1</sup> Gartner. "Market Guide for Hadoop Distributions." January 6, 2015.

## 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.

Among other benefit, this switch series:

- 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

## DELL EMC DATA LAKE

Dell EMC continues to keep customer use cases first with robust Data Lake opportunities using Cloudera Enterprise together with Dell EMC scale-out NAS Isilon storage running the OneFS operating system. The result is 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 Cloudera, Dell EMC PowerEdge servers and Isilon shared storage helps your organization speed time to insights, improve storage utilization, eliminate islands or silos of storage, and lower storage management costs of migration, security and protection.

## DELL SERVICES

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 Cloudera solution is backed by expert hardware and software support that can be tailored to your specific needs.

## CLOUDERA ENTERPRISE

The right technology is key for turning your data into real business value. Powered by Apache Hadoop, Cloudera

Enterprise is a fast, easy and secure modern data platform. From analytics to data science, anyone can now get results from any data and across any environment — all within a single, scalable platform.

When you make Cloudera Enterprise the center of your business, you open up limitless possibilities with your data. Whether you're powering data engineering and data science workloads, building an operational or analytic database, or looking to bring them all together in an enterprise data hub, Cloudera has the right platform to fit your needs.

Cloudera Enterprise delivers:

- **Data engineering** — Bring your data engineers and data scientists together to build real-time pipelines, speed data processing, and develop and train data models.
- **Analytic database** — Modernize your IT architecture to enable ELT and high-performance SQL analytics for reporting, exploration and self-service business intelligence.
- **Operational database** — Build data-driven applications that deliver real-time insights for monitoring and detection, as well as streaming applications like the Internet of Things and model scoring and serving.

## CLOUDERA CDH 5.9

Cloudera CDH 5.9 is at once:

- **Fast for business** — From analytics to data science and everything in between, Cloudera delivers the performance you need to unlock the potential of unlimited data.
- **Easy to manage** — Focus on results, not fighting fires. Cloudera provides the operations that keep mission-critical applications up and running — especially at scale.
- **Secure without compromise** — Meet your most stringent data security and compliance needs without sacrificing business agility and innovation. Cloudera provides a comprehensive, integrated approach to data security and governance.

## TECHNICAL SPECIFICATIONS

Server Architecture	Infrastructure Nodes 3.5"	General Purpose Data Node 2.5"	Data Node 3.5"	Data Node 2.5"
R730XD Server	R730XD	R730XD	R730XD	R730XD
Processor	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)	2 X Intel E5-2690 v4 2.6GHz (14 Core)
Memory	128GB	128GB	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)	8 x 1.2TB 10K RPM SAS 12Gbps 2 x 600GB 10K RPM SAS 12Gbps (Flex Bay)	12 x 4TB 7.2K RPM SAS 6Gbps 2 x 600GB 10K RPM SAS 12Gbps	24 x 1.2TB 10K RPM SAS 12Gbps 2 x 600GB 10K RPM SAS 12Gbps
RAID Layout	Operating System- 2 HDs, RAID 1 Data- 8 HDs, RAID 10	Operating System- 2 HDs, RAID 1 Data- 8 HDs, RAID	Operating System- 2 HDs, RAID 1 JBOD- 12HDs, Non-RAID	Operating System- 2 HDs, RAID 1 JBOD- 12HDs, Non-RAID
<b>Cluster Network</b>	<b>Cluster Data Network</b>	<b>BMC Network</b>	<b>Edge Network</b>	
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	

Infrastructure Nodes		Data Node	
FX2 Server	FC630	FC630	
Processor	2 X Intel E5-2650 v4 2.2GHz (12 Core)	2 X Intel E5-2680 v4 2.4GHz (14 Core)	
Memory	128GB	256GB	
Network Card	Intel X710 Quad Port 10Gbe	Intel X710 Quad Port 10Gbe	
Hard Drive Controller	H730	H730	
Storage	FD332 10 x 1TB 7.2K RPM SAS 12Gbps	On-board Storage 2 x 400GB SSD SATA 6Gbps PERC S130 FD332 16 x 2TB 7.2K RPM SAS 12Gbps	
RAID Layout	Operating System- 2 HDs, RAID 1 Data- 8 HDs, RAID 10	Operating System- 2 HDs, RAID 1 JBOD- 8 HDs, NON-RAID	
<b>Cluster Network</b>	<b>Cluster Data Network</b>	<b>BMC Network</b>	<b>Edge Network</b>
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

Cloudera CDH 5.9	
Accumulo	1.7.2
Impala	2.7
Kafka	2.0
Navigator	2.8
Sentry	Cloudera Manager 5.X & Higher
Spark	1.6

To learn more, visit:

[Dell.com/Hadoop](http://Dell.com/Hadoop) | [Dell.com/BigData](http://Dell.com/BigData) | [EMC.com/BigData](http://EMC.com/BigData)

Dell is a trademark of Dell Inc. and EMC is a trademark of EMC Corp. Dell Technologies is a trademark of Dell Inc. Copyright © 2016 Dell Inc. or its subsidiaries. All Rights Reserved. Dell, EMC, and other trademarks are trademarks of Dell Inc. or its subsidiaries. Other trademarks may be trademarks of their respective owners. Intel, Xeon and the Intel logo are trademarks of Intel Corporation in the U.S. and/or other countries.

November 2016 | Rev 1.0