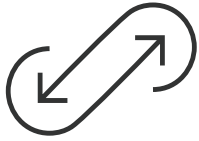


Dell EMC Ready Solutions for Hadoop

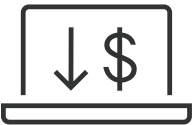
Leverage optimized Hadoop solutions designed to address data analytics requirements, reduce costs and deliver outstanding performance

Table of Contents

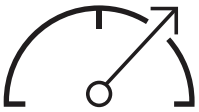
Get the power of Hadoop faster, with less risk.	2
Hadoop use cases.	3
Are you facing these challenges?.	4
Dell EMC Ready Solutions for Hadoop	4
Choose your Hadoop distribution.	5
Cloudera Hadoop	5
Dell EMC machine learning with Hadoop	5
Hortonworks Hadoop	5
Configuration details.	6
Shared storage Hadoop vs. distributed storage Hadoop	8
Why Dell EMC?	8
Complete your solution with Dell EMC services and financing	9
Dell EMC Professional Services	9
Dell Financial Services.	10
Dell EMC Customer Solution Centers.	10
Find out more today.	10



Leverage an optimized solution



Reduce cost



Deliver outstanding performance

Get the power of Hadoop faster, with less risk

Digital transformation is causing churn, uncertainty and disruption for many business leaders who need to act quickly as pressure increases from all directions. Big data and analytics are at the core of this transformation. However, without the right tools for big data analytics, you'll have to conduct multiple, limited data analyses and then find a way to synthesize the results, involving a lot of manual effort and subjective analysis. With Apache® Hadoop® as a foundational component of the big data and analytics solution stack, you gain the ability to quickly process large sets of disparate data for a more comprehensive view of your customers, operations, opportunities, risks, and more.

Despite these tantalizing benefits, many organizations struggle — either to begin their data analytics journey or to make Hadoop projects successful once they've begun. They are often impeded by a lack of Hadoop expertise and end up spending too much time and effort on front-end work before they can get to the results of an operational solution.

Expertise and infrastructure matter when building a Hadoop environment. That's why Dell EMC has teamed up with industry leaders — such as Intel®, Hortonworks®, Cloudera® and Syncsort® — to remove the uncertainty and barriers that may dissuade you from deploying Hadoop.

Cost-effective, future-ready Dell EMC Ready Solutions for Hadoop can be easy-to-implement solutions that help you efficiently harness Hadoop and the power of data analytics to drive competitive advantage.

Leverage an optimized solution

Dell EMC Ready Solutions for Hadoop are developed jointly with leading Hadoop distributions, and are based on extensive customer experience with real-world Hadoop production installations. These solutions offer the hardware, software, resources and services needed to deploy and manage Hadoop in a production environment.

Reduce costs

Dell EMC Ready Solutions for Hadoop offer compelling total cost of ownership (TCO) benefits by using cost-optimized, industry-standard Dell EMC servers and storage to decrease the cost to store and process large data sets versus traditional business intelligence (BI) and analytics solutions.

Deliver outstanding performance

Dell EMC Ready Solutions for Hadoop have been engineered and certified to work together and provide known performance parameters and deployment methods — ensuring that users can get excellent performance and minimal risk with solution deployment.

Hadoop use cases

The use cases for Hadoop are very diverse, but there are common patterns across industries and verticals. Here is a sampling of possible use cases utilizing Dell EMC Ready Solutions for Hadoop.

Operational efficiency use cases			
Data warehouse augmentation	Log aggregation and analytics	Dual storage and active archive	Archive-intensive and tiered Hadoop
Reduces total cost of ownership (TCO) and increases return on investment (ROI)	Secures your enterprise	Reduces TCO and eases compliance	Provides enterprise storage features for storage-centric Hadoop workloads with large capacity requirements
<ul style="list-style-type: none"> • Offload extract, transform, load (ETL) workloads • Reduce licensing costs • Enhance data accessibility • Enable better data exploration and analytics • Manage performance more effectively 	<ul style="list-style-type: none"> • Prevent security breaches and threats • Detect operational anomalies • Increase infrastructure efficiency and automation 	<ul style="list-style-type: none"> • Lower data storage costs while maintaining accessibility • Ease compliance and reporting • Streamline inquiry processes • Enjoy business operations improvement 	<ul style="list-style-type: none"> • Lowers costs for active archive • Use for long-term tiered storage for regulatory compliance • Get multi-protocol support for storage consolidation

Business transformation use cases				
Marketing	Finance	Healthcare	Pharmaceutical	Manufacturing
Anticipating customer needs	Reducing risk and detecting fraud	Improving patient care and reducing costs	Ensuring regulatory compliance and validation	Achieving continuous process improvement
<ul style="list-style-type: none"> • Customer 360 insight • Customer retention • Customer segmentation • Customer loyalty • New product/service launch 	<ul style="list-style-type: none"> • Credit scoring • Customer analytics • Fraud detection • Risk management • Sarbanes-Oxley Act (SOX) compliance 	<ul style="list-style-type: none"> • Quality of care • Patient safety • Risk mitigation • Fraud detection • Claims management 	<ul style="list-style-type: none"> • Biomedical analytics • Stability and shelf life • Primary research • FDA compliance manufacturing 	<ul style="list-style-type: none"> • Product quality • Customer insight • Demand forecasting • Improved operations

Ready Solutions for Hadoop yield results¹

2X

ROI

\$4.1M

of additional profit

\$15M

of savings from data storage and offload

\$1.4M

in savings from reduced fraud

30%

improvement in data science productivity

20%

reduction in fraud

<6

months payback

\$540K

savings in system administration efficiency per year

Are you facing these challenges?

“We’re struggling to make our analytics projects successful.”

Many businesses have trouble getting started with analytics solutions or making sure projects are successful once they’re completed. Lack of in-house Hadoop expertise is common, and can lead to project delays and configuration problems. For nearly a decade, Dell EMC has helped organizations solve the Hadoop 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 Ready Solutions for Hadoop, helping make Hadoop projects easier to start and more successful once completed.

“We need to reduce costs across the data center.”

IT budgets are typically constrained, which can make it difficult to free up resources for new Hadoop projects. Cost-optimized, industry-standard servers and storage are at the core of Dell EMC Ready Solutions for Hadoop. This helps decrease the cost of your total analytics solution. Dell EMC Services help shorten project timelines and can reduce production issues that might affect costs and ROI down the road.

“We’re not sure how to provide the blazing performance that makes Hadoop projects successful.”

Data is critical to every aspect of running a modern business. Teams need fast, concurrent access to data from every corner of the business to delight customers, outpace the competition, secure the enterprise, and maintain regulatory compliance. But testing and tuning Hadoop can be difficult and time-consuming. Dell EMC Ready Solutions for Hadoop have been engineered and certified so that users get outstanding performance with minimal intervention from IT.

Dell EMC Ready Solutions for Hadoop

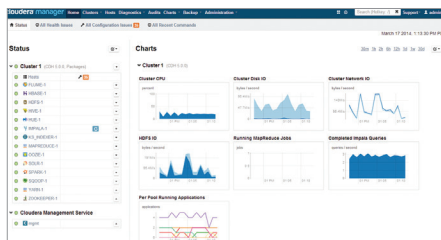
Dell EMC Ready Solutions for Hadoop are integrated Hadoop solutions, designed to address data analytics requirements, reduce costs and deliver outstanding performance.

Dell EMC has been working with the leading innovators in big data since 2008, and started designing and building custom Hadoop solutions in 2009. With our deep roots in data analytics solutions and Hadoop — and other leading partners in data analytics — Dell EMC has the expertise, tools and solutions needed to drive successful, flexible and scalable Hadoop deployments.

¹ Forrester Research study commissioned by Dell EMC, “[The Total Economic Impact of the Dell EMC Ready Bundle for Cloudera Hadoop](#),” May 2018.

Choose your Hadoop distribution

Dell EMC Ready Solutions support both Cloudera Enterprise and the Hortonworks Data Platform®, so you can choose the right Hadoop solution for your needs. Both offer optional Syncsort ETL processing, allowing you to capitalize on the advantages of Hadoop while making better use of existing enterprise data warehouse (EDW) investments.



Cloudera software

Cloudera Hadoop

A fast, easy and secure modern data solution

The Cloudera Hadoop design delivers the core elements of Apache Hadoop, including scalable storage and distributed computing, within a solution based on Cloudera Enterprise software and Dell EMC hardware delivered with optional services.

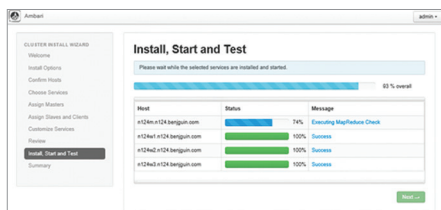
- **Fast for business:** From analytics to data science and everything in between, Cloudera delivers the performance you need to unlock the potential of unlimited data.
- **Makes Hadoop easy to manage:** With Cloudera Manager, automated wizards let you quickly deploy your cluster, no matter what the scale or the deployment environment. Through a centralized interface, your operations team can easily tune configurations and resourcing, manage a wide range of user roles for cross-departmental, self-service access, and even manage multiple clusters for multi-tenant environments.
- **Secure without compromise:** Meet stringent data security and compliance needs without sacrificing business agility. Cloudera provides an integrated approach to data security and governance.

Dell EMC machine learning with Hadoop

Dell EMC machine learning with Hadoop builds on the power of tested and proven Dell EMC Ready Solutions for Hadoop, created in partnership with Cloudera. The design includes an optimized stack along with data science and framework optimization, so you can get up and running quickly.

If you're running Apache Spark®, you can take advantage of the BigDL distributed deep learning library for Apache Spark, for a faster path to enabling machine learning and to making deep learning more accessible. BigDL integrates into the Apache Hadoop frameworks and tools, enabling deep learning on the same Hadoop/Spark cluster where data is stored.

See the Ready Solutions for AI solution overview at dell EMC.com/readyforai.



Hortonworks software

Hortonworks Hadoop

An open and stable foundation with a growing ecosystem to build and deploy big data solutions

Hortonworks Data Platform makes business intelligence on Hadoop a reality, combining a fast in-memory SQL engine to create data marts with an online analytical processing (OLAP) cube engine that enables huge data set queries in seconds. Dell EMC has worked with Hortonworks to develop a well-designed and scalable solution for Hortonworks Hadoop.

- **Open:** Hortonworks is committed to a 100% open approach to software development that spurs innovation.
- **Interoperable:** Its 100% open-source architecture enables Hortonworks Data Platform to interoperate with a broad range of BI and other applications.
- **Enterprise ready:** Hortonworks is built for enterprises, providing consistent operations with centralized management and monitoring of clusters through a single pane of glass.

Configuration details

Dell EMC Ready Solutions for Hadoop offer a variety of configurations to meet your needs.²

Dell EMC Ready Solutions for Hadoop configuration options			
	Direct-attached storage		Isilon Scale-out NAS
	Rack servers	Modular infrastructure	Modular infrastructure
Cloudera Hadoop	<ul style="list-style-type: none"> Cloudera Enterprise 5.12 PowerEdge R740xd or R640 Servers 	<ul style="list-style-type: none"> Cloudera Enterprise 5.12 PowerEdge FX2 FC630 Servers 	
Hortonworks Hadoop	<ul style="list-style-type: none"> Hortonworks Data Platform 2.6 PowerEdge R730xd Servers 	<ul style="list-style-type: none"> Hortonworks Data Platform 2.6 PowerEdge FX2 FC630 Servers 	<ul style="list-style-type: none"> Hortonworks Data Platform 2.5 PowerEdge FX2 FC630 Servers Isilon X410 Storage
Quickstart	<ul style="list-style-type: none"> Cloudera Enterprise 5.12 PowerEdge R740xd Servers 		

Dell EMC Ready Solutions for Hadoop configuration characteristics						
	PowerEdge server					Isilon Scale-out NAS
	R740xd - 3.5"	R640	R730xd - 2.5"	R730xd - 3.5"	FX2 FC630	FX2 FC630
Scalability	Up to 288 nodes		Up to 252 nodes			<ul style="list-style-type: none"> Up to 168 compute Nodes Up to 84 storage nodes
Raw storage	64TB / node	24TB / node	24TB / node	48TB / node	23TB / node	102TB / Isilon X410 Storage
Maximum HDFS storage	6PB	2.3PB	1.9PB	3.8PB	1.8PB	8.6PB
Processors	Dual Intel Xeon® Gold 6140 2.3GHz, 18C/36T	Dual Intel Xeon Gold 6136 3.0G, 12C/24T	Dual Intel Xeon E5-2690 v4 2.6GHz, 14C/28T	Dual Intel Xeon E5-2650 v4 2.2GHz, 14C/28T	Dual Intel Xeon E5-2680 v4 2.4GHz, 14C/28T	Dual Intel Xeon E5-2680 v4 2.4GHz 14C/28T
iDRAC network	S3048-ON					
Pod network	Z9100 25GbE		S4048-ON 10GbE			
Cluster aggregation network	Z9100 100GbE		S6010-ON 40GbE			
PowerEdge Server infrastructure nodes	<ul style="list-style-type: none"> 3x R640 name 1x R640 edge 	<ul style="list-style-type: none"> 3x R640 name 1x R640 edge 	<ul style="list-style-type: none"> 3x R730xd name 1x R730xd edge 	<ul style="list-style-type: none"> 3x R730xd name 1x R730xd edge 	<ul style="list-style-type: none"> 3x FC630 name 1x FC630 edge 	4x FC630 master
PowerEdge Server worker nodes	R740xd - 3.5"	R640 - 12x 2.5"	R730xd - 2.5" chassis	R730xd - 3.5" chassis	FC630 + FD332	FC630 + FD332

² The quantity and configuration of each system depends on the specific size and use case.

- **Dell EMC PowerEdge R740xd Server:** Maximum storage performance and scalability ideal for software-defined storage. With up to 24 NVMe drives, the R740xd ensures application performance scales to meet company demands.
- **Dell EMC PowerEdge R640 Server:** Get scalable computing and storage in a 1U, 2-socket platform with an ideal mix of performance, cost and density for most data centers.
- **Dell EMC PowerEdge R730xd Server:** Impressive processor performance, a large memory footprint, extensive input/output (I/O) options and a choice of dense, high-performance storage or low-cost, high-capacity storage.
- **Dell EMC PowerEdge FX2 Server:** Modular server, storage and networking blocks are neatly combined in a compact, converged 2U rack chassis to redefine data center agility.
- **Dell EMC PowerEdge FC630 Server:** The 2-socket, half-width 1U workhorse server blocks are ideal for a wide variety of business applications.
- **Dell EMC Networking S3048-ON 1GbE Switch:** 1GbE top-of-rack (TOR) switch with an industry-hardened operating system (OS) and support for open networking, providing freedom to run third-party operating systems.
- **Dell EMC Networking S4048-ON 10/40GbE Switch:** TOR, high-density 1U switch, with 48 10GbE uplinks, offers ultra-low latency and line-rate performance.
- **Dell EMC Networking S6010-ON 10/40GbE Switch:** Disaggregated hardware and software TOR networking solution that empowers you to deploy modern workloads and applications designed for the open networking era.
- **Dell EMC FD332 Storage:** Flexible, high-density, half-width 1U storage modules enable you to rapidly scale direct-attached storage (DAS).
- **Dell EMC Isilon X-Series:** The Isilon X-Series enables companies to store and analyze large volumes of data (> 100TB) economically. This is especially relevant for companies that need to store and analyze historical data. The OneFS file system software helps consolidate IT workloads by managing both Hadoop data and enterprise IT data — accessible through several file and object interfaces — in one place. Isilon also delivers enterprise-grade file management capabilities such as data protection, data tiering, security, and data-at-rest encryption (DARE) with self-encrypting drive (SED) options to make it easier for companies to conform to regulatory requirements and data management policies.

Dell EMC marketing analytics team

Realized US\$4 million in OpEx savings¹⁰

Texas Advanced Computing Center (TACC)

Gained up to 50X performance improvement¹¹

Dell EMC

40% increased automation of standard sales reports¹²

60% faster response times for prescriptive and predictive analysis¹²

Shared storage Hadoop vs. distributed storage Hadoop

It's a testament to Hadoop's flexibility that it supports multiple deployment models accounting for varying budget, performance, capacity and density requirements. Dell EMC Isilon is a shared-storage model where the persistent file system data for Hadoop is stored in an Isilon network-attached storage (NAS) cluster versus in the distributed model where data is spread across the local storage of the Hadoop nodes themselves.

These two approaches offer varying advantages:

Shared-storage Hadoop	Distributed-storage Hadoop
Single copy of data for IT workloads and analytics	Massive (hundreds of petabytes)
Reduced data center footprint (storage density)	Ability to use commodity servers
Enterprise file management: data protection, security, storage tiering	Linear scaling
Independent scaling of storage and compute	Flexible replica model

Why Dell EMC?

Dell EMC holds leadership positions in some of the largest growth categories in the IT infrastructure business, and that means you can confidently source your IT needs from one provider — Dell EMC.

- #1 converged and hyper-converged infrastructure³
- #1 in traditional and all-flash storage⁴
- #1 virtualized data center infrastructure⁵
- #1 cloud IT infrastructure⁶
- #1 server virtualization and cloud systems management software (VMware®)⁷
- #1 in data protection⁸
- #1 in software-defined storage⁹

³ IDC WW Quarterly Converged Systems Tracker, Q4 2017, April 2018, Vendor Revenue.

⁴ IDC WW Quarterly Enterprise Storage Systems Tracker, September 2017, Vendor Revenue — EMC Q2 2017.

⁵ Dell EMC Annual Report, 2015.

⁶ IDC WW Quarterly Cloud IT Infrastructure Tracker, April 2017, Vendor Revenue — EMC Q4 2016.

⁷ IDC WW Virtual Machine and Cloud System Market Shares 2016, July 2017.

⁸ Dell EMC Pulse, "Gartner Recognizes EMC as a Leader in the 2016 Data Center Backup and Recovery Software Magic Quadrant," June 2016.

⁹ IDC WW Semiannual Software Tracker, 2H2016, April 2017.

¹⁰ Dell EMC case study, "An Efficient Way to Discover New Customer Insights," January 2017.

¹¹ Dell EMC case study video, "Dell EMC Drives Big Data Solutions at TACC," November 2016.

¹² Dell EMC case study, "Unlocking data's value for better insights and decisions," May 2015.

“We’ve completely redesigned how we capture, store and provision data with the new Dell Hadoop cluster. We can gather larger amounts of data, and our analysts and statisticians can mine that data in ways they couldn’t before.”¹³

— Tony Giordano, Executive Vice President of the Technology Solutions Group, Merkle, United States

“Addressing exhausted enterprise data capacity can cost up to \$800,000 per terabyte of data. But with Hadoop’s extreme scalability, adding terabytes can cost as little as \$5,000 using MetaScale’s big data appliances based on Dell PowerEdge Servers.”¹⁴

— Ankur Gupta, General Manager, MetaScale, United States

Complete your solution with Dell EMC services and financing

Dell EMC Professional Services

Solutions customized for your needs

- [Dell EMC Big Data Vision Workshop](#) focuses on Hadoop for business leaders. We have a unique methodology to identify and prioritize a single analytics use case with a combination of implementation feasibility and business value. It’s a three-week engagement that applies research, interviews and data science expertise and techniques to the organization — culminating in a one-day workshop to identify and agree on the analytics use case and path forward. This approach sets us apart from the “bring in a bunch of technology and see what it can do” approach that’s pushed by many vendors.
- [Dell EMC Hadoop Consulting](#) is delivered by certified Hadoop experts to help you get the business value of data analytics using Hadoop. The services include a data analytics assessment, workshop, testing, proofs of concept and production implementation. These Hadoop experts help determine where Hadoop is a good fit for your organization. They also help you build your own internal team of Hadoop experts through knowledge transfer at each step.
- [Dell EMC Jumpstart Services](#) offer best-practice guidance, hands-on labs, roadmap planning and knowledge transfer on Hadoop installations, so you can get from install to productivity with the skills and knowledge to get the greatest value from your big data solution.
- [Dell EMC Hadoop Health Check Service](#) reviews your current data technologies and processes, and makes recommendations for tools, testing and operational practices.
- [Big Data Technology Assessment Services](#) assess Isilon customers with Hadoop installations on the current state of your Hadoop environment and make recommendations with no implementation of changes.
- [Big Data Technology Implementation Services](#) help current and prospective Isilon customers with Hadoop installations of one to three clusters and up to hundreds of nodes. Involves installing and configuring Hadoop to work with an Isilon storage cluster, including security.
- [Dell EMC Hadoop on Isilon Implementation Starter Kit](#) is ideal if you’re interested in installing a proof of concept, or developing or testing Hadoop clusters using best practices to leverage Isilon for HDFS storage.

Deployment assistance when you need it

Dell EMC offers a broad menu of custom installation and implementation services for our Hadoop solutions. Dell EMC Services include on-site hardware and operating system software installation, optional rack integration at a Dell EMC facility and validation of the installed solution. Hadoop software deployment is a custom project that is delivered based on your needs. Dell EMC takes care of project management, from order drop to your acceptance.

Support always on for you

[Dell EMC ProSupport](#) offers a single point of accountability from experts with solution-specific training, along with premium hardware and software support available 24x7x365. ProSupport also includes collaborative software support for Cloudera and Hortonworks software. Additionally, ProSupport includes next-business-day, on-site service with four- and eight-hour parts and labor response options, and escalation management with customer-set severity level options.

¹³ Dell EMC case study, “A powerful new foundation for creating customer campaigns,” May 2015.

¹⁴ Dell EMC case study, “Accelerating big data ROI with Hadoop,” April 2015.

Dell Financial Services

- Leasing and financing solutions are available throughout the U.S., Canada and Europe.
- Dell Financial Services can finance the total technology purchase.
- Efficient electronic quoting and online contracts offer an efficient customer experience.

Learn more about [Dell Financial Services](#).

Dell EMC Customer Solution Centers

Our global network of 21 dedicated [Dell EMC Customer Solution Centers](#) are trusted environments where world-class IT experts collaborate with customers and prospects to share best practices; facilitate in-depth discussions of effective business strategies using briefings, workshops, or proofs-of-concept (PoCs); and help businesses become more successful and competitive. Dell EMC Customer Solution Centers reduce the risks associated with new technology investments and can help improve speed of implementation.

Find out more today

Don't wait to harness the benefits of Hadoop on an optimized solution designed from the ground up to address data analytics requirements, reduce costs and deliver outstanding performance for deep data mining and analytics. Contact your Dell EMC representative to find out more today.

Our solution partners

cloudera®



syncsort

DELLEMC

Copyright © 2018 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 the property of their respective owners Published in the USA 07/18 Solution overview DELL-EMC-SO-HADOOP-USLET-105

Apache®, Hadoop® and Spark® are either registered trademarks or trademarks of the Apache Software Foundation in the United States and/or other countries. Intel® and Xeon® are trademarks of Intel Corporation in the U.S. and other countries. Cloudera® is a trademark or trade dress of Cloudera. Hortonworks® and Hortonworks Data Platform® are trademarks of Hortonworks, Inc in the U.S. and other countries. Syncsort® is the property of Syncsort in the United States and/or other countries. VMware® products are covered by one or more patents listed at <http://vmware.com/go/patents>. VMware® is a registered trademark or trademark of VMware, Inc in the United States and/or other jurisdictions.

Dell EMC believes the information in this document is accurate as of its publication date. The information is subject to change without notice.