Dell EMC Ready Bundles for Hadoop

Leverage end-to-end Hadoop systems designed to address data analytics requirements, reduce costs and optimize performance

Table of Contents

Get the power of Hadoop faster, with less risk

The 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, with Apache® Hadoop® as a foundational component of the big data and analytics solution stack. Unlike traditional systems, Hadoop enables multiple types of analytic workloads to run — on the same data, at the same time, at massive scale — on industry-standard hardware.

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 Bundles for Hadoop are easy-to-implement solutions that help you efficiently harness the Hadoop platform and the power of data analytics to drive competitive advantage.

Leverage an end-to-end solution

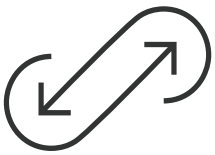
Dell EMC Ready Bundles for Hadoop are developed jointly with leading Hadoop distributions, and are based on extensive customer experience with real-world Hadoop production installations. These solutions include the hardware, software, resources and services needed to deploy and manage Hadoop in a production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

Reduce costs

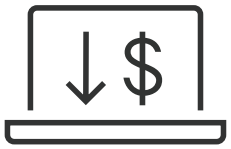
Dell EMC Ready Bundles for Hadoop offer compelling total cost of ownership (TCO) benefits by using cost-optimized, industry-standard Dell EMC servers and storage to decrease the cost to store and process large data sets versus traditional business intelligence (BI) and analytics solutions.

Optimize performance

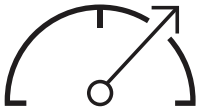
Dell EMC Ready Bundles for Hadoop have been engineered and certified to work together and provide known performance parameters and deployment methods — ensuring that users get excellent performance with minimal testing prior to purchase and minimal risk with solution deployment.



Leverage an end-to-end solution



Reduce costs



Optimize performance

What is your business goal?

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 a Dell EMC Ready Bundle for Hadoop.

Operational efficiency use cases			
Dell EMC 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
• Offload extract, transform, load (ETL) workloads • Reduce licensing costs • Enhance data accessibility • Enable better data exploration and analytics • Manage performance more effectively	• Prevent security breaches and threats • Detect operational anomalies • Increase infrastructure efficiency and automation	• Lower data storage costs while maintaining accessibility • Ease compliance and reporting • Streamline inquiry processes • Enjoy business operations improvement	• 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
• Customer 360 insight • Customer retention • Customer segmentation • Customer loyalty • New product / service launch	• Credit scoring • Customer analytics • Fraud detection • Risk management • Sarbanes-Oxley Act (SOX) compliance	• Quality of care • Patient safety • Risk mitigation • Fraud detection • Claims management	• Biomedical analytics • Stability and shelf life • Primary research • FDA compliance manufacturing	• Product quality • Customer insight • Demand forecasting • Improved operations

Dell EMC Ready Bundles for Hadoop

Dell EMC Ready Bundles for Hadoop are integrated Hadoop systems, designed to address data analytics requirements, reduce costs and optimize performance.

Dell EMC has been working with the leading innovators in big data since 2008, and started designing and building custom Hadoop server platforms 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.

Configuration details

Dell EMC Ready Bundles for Hadoop offer a variety of configurations to meet your needs.[1]

	Dell EMC Hadoop solution on PowerEdge R730xd Server	Dell EMC Hadoop solution on PowerEdge FC630 Server	Dell EMC ETL offload Hadoop solution on PowerEdge R730xd Server	Dell EMC ETL offload Hadoop solution on PowerEdge FC630 Server	Dell EMC QuickStart for Hadoop
Use cases	Active archive/customer 360-degree	Active archive/customer 360-degree	ETL offload	ETL offload	Testing and proofs of concept (POCs)
Sizing	Scales from 5 to 252 nodes, 3.8PB	Scales from 5 to 252 nodes, 3.8PB up to 16PB with Dell EMC Isilon X410 Storage	Scales from 5 to 252 nodes, 3.8PB	Scales from 5 to 252 nodes, 3.8PB up to 16PB with Dell EMC Isilon X410 Storage	48TB
Software	Cloudera Enterprise or Hortonworks Data Platform and OpenManage/iDRAC with Lifecycle Controller	Cloudera Enterprise or Hortonworks Data Platform and OpenManage/iDRAC with Lifecycle Controller	Cloudera Enterprise and Syncsort DMX-h and SILQ™ or Hortonworks Data Platform and OpenManage/iDRAC with Lifecycle Controller	Cloudera Enterprise and Syncsort DMX-h and SILQ or Hortonworks Data Platform and OpenManage/iDRAC with Lifecycle Controller	Cloudera Basic Edition and OpenManage/iDRAC with Lifecycle Controller
Pod network	2 x Networking S4048-ON 10GbE pod switches 1 x S3048-ON iDRAC switch	2 x Networking S4048-ON 10GbE pod switches 1 x S3048-ON iDRAC switch	2 x Networking S4048-ON 10GbE pod switches 1 x S3048-ON iDRAC switch	2 x Networking S4048-ON 10GbE pod switches 1 x S3048-ON iDRAC switc	1 x Networking S3048-ON switch 1 x Networking S4048-ON switch
Cluster aggregation network	2 x Networking S6010-ON 40GbE cluster switches	2 x Networking S6010-ON 40GbE cluster switches	2 x Networking S6010-ON 40GbE cluster switches	2 x Networking S6010-ON 40GbE cluster switches	
Infrastructure nodes	3 x PowerEdge R730xd Server name nodes 1 x PowerEdge R730xd Server edge node	3 x PowerEdge FC630 Server name nodes 1 x PowerEdge FC630 Server edge node	3 x PowerEdge R730xd Server name nodes 1 x PowerEdge R730xd Server edge node	3 x PowerEdge FC630 Server name nodes 1 x PowerEdge FC630 Server edge node	2 x PowerEdge R730xd Servers
Worker nodes	10 x PowerEdge R730xd Servers with 3 5" Drives — 48TB or 10 x PowerEdge R730xd Servers with 2 5" Drives — 24TB	10 x PowerEdge FC630 Servers 10 x PowerEdge FD332 Storage — 32TB	10 x PowerEdge R730xd Servers with 3 5" Drives — 48TB or 10 x PowerEdge R730xd Servers with 2 5" Drives — 24TB	10 x PowerEdge FC630 Servers 10 x PowerEdge FD332 Storage — 32TB	3 x PowerEdge R730xd Servers with 3 5" Drives — 48TB or 3 x PowerEdge R730xd Servers with 2 5" Drives — 24TB
Storage[2]		5 x Dell EMC Isilon X410 Storage with 102TB HDD/ 3.2TB SSD/ 256 GB 2x Mellanox® QDR InfiniBand® Switch 8 ports		5 x Dell EMC Isilon X410 Storage with 102TB HDD/ 3.2TB SSD/ 256 GB 2x Mellanox® QDR InfiniBand® Switch 8 ports	

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.

[1] The quantity and configuration of each system depends on the specific size and use case.

[2] Isilon storage option available only with Hortonworks solutions.

Dell EMC Networking S4048-ON 10/40GbE switch: Top-of-rack, high-density 1U switch with 48 10GbE uplinks. It 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 Nodes: The Isilon X-Series Nodes deliver a high-capacity density Hadoop system where storage can scale much faster than compute, to enable 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 within Isilon enables companies to consolidate their IT workloads by managing both Hadoop data and enterprise IT data — accessible through several file and object interfaces — all in one place. Isilon also delivers several 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.

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. The Dell EMC Isilon solution is a shared storage model where the persistent filesystem data for Hadoop is stored in an Isilon 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 (100s+ of petabytes)
Reduced datacenter footprint (storage density)	Ability to use commodity platforms
Enterprise file management: data protection, security, storage tiering, etc.	Linear scaling
Independent scaling of storage and compute	Flexible replica model

Merkle

Cut costs by **60%**[10]

Texas Advanced Computing Center (TACC)

Gained up to **50X performance improvement**[11]

Dell EMC

40% increased automation of standard sales reports[12]

60% faster response times for prescriptive and predictive analysis[12]

Choose your Hadoop distribution

Dell EMC Ready Bundles support both Cloudera Enterprise and the Hortonworks Data Platform (HDP), so you can choose the right Hadoop solution for your needs.

Dell EMC Ready Bundle for Cloudera Hadoop

Powered by Apache Hadoop, Cloudera Enterprise is a fast, easy and secure modern data platform.

- **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 stringent data security and compliance needs without sacrificing business agility. Cloudera provides an integrated approach to data security and governance.

Dell EMC Ready Bundle for Hortonworks Hadoop

The Hortonworks Data Platform provides an open and stable foundation for enterprises, and a growing ecosystem to build and deploy big data solutions.

- **Open:** Hortonworks is committed to a 100% open approach to software development that spurs innovation.
- **Interoperable:** Its 100% open source architecture enables HDP to interoperate with a broad range of business intelligence 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.

Why Dell EMC for Hadoop?

The combination of Dell and EMC brings together two industry-leading companies with strong reputations for value and innovation. Dell EMC holds leadership positions in some of the biggest and 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 infrastructure[3]
- #1 in traditional and all-flash storage[4]
- #1 virtualized data center infrastructure[5]
- #1 cloud IT infrastructure[6]
- #1 server virtualization and cloud systems management software (VMware®)[7]
- #1 in data protection[8]
- #1 in software-defined storage[9]

[3] IDC WW Quarterly Converged Systems Tracker, June 2016, Vendor Revenue—EMC FY 2015.

[4] IDC WW Quarterly Enterprise Storage Systems Tracker, June 2016, Vendor Revenue—EMC FY 2015.

[5] Dell EMC Annual Report, 2015.

[6] IDC WW Quarterly Cloud IT Infrastructure Tracker, Q1 June 2016, Vendor Revenue—EMC FY 2015.

[7] IDC WW Virtual Machine and Cloud System Market Shares 2015, July 2016.

[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 white paper, "Software-Defined Storage: A Pervasive Approach to IT Transformation Driven by the 3rd Platform," November 2015.

[10] Dell EMC case study, "Merkle creates its big data insight solution with Dell Hadoop clusters," May 2015.

[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.

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.

Deployment assistance when customers 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.

[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 EMC Financial Services

Let the wealth of leasing and financing options from Dell EMC Financial Services help you find opportunities when your organization faces decisions regarding capital expenditures, operating expenditures and cash flow. Dell EMC offers a wide range of payment options to make it easier than ever to meet your needs.

Learn more about Dell EMC Financial Services.

Find out more today

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

Our solution partners

cloudera

Hortonworks

Contact us

To learn more, visit dell.com/hadoop or contact your local representative or authorized reseller.

DELL EMC