DELLTechnologies

Validated Designs for Hadoop

Leverage an optimized solution, reduce costs and deliver outstanding performance

Table of Contents

Get the power of Hadoop faster, with less risk.

The digital universe is large and growing fast. Analysts predict that within the next five years we will be creating 175 zettabytes (ZB), or 175 trillion gigabytes (GB), of new data every year. That's roughly 175 times more digital bits than there are stars in the known universe.[1]

Historically, to handle growing data needs, organizations have placed enterprise data warehouse (EDW) platforms at the center of their business intelligence (BI) operations. However, these legacy solutions can be expensive and may create a competitive disadvantage. Many are falling behind without the proper analytics capabilities to drive actionable business insights.

With Apache® Hadoop® as the foundation of a analytics solution stack, you can quickly process large sets of disparate data for a more comprehensive view of your customers, operations, opportunities, risks and more. However, expertise and infrastructure matter when building a Hadoop environment. That's why Dell Technologies works with industry leaders like Cloudera® to remove the uncertainty and barriers that may dissuade you from deploying Hadoop. Validated Designs for Hadoop provide a faster way to harness the power of analytics to drive competitive advantage.

Dell Technologies has what you need.

Expertise and guidance

Technology is evolving quickly, so your team may not have time to design, deploy and manage solution stacks optimized for Hadoop. While advanced analytics and artificial intelligence (AI) might seem like the latest IT trend, Dell Technologies has been a leader in the advanced computing space for over a decade, with proven products, solutions and expertise. Dell has a team of analytics, AI and high performance computing (HPC) experts dedicated to staying on the cutting edge, testing new technologies and tuning solutions to your applications to help you keep pace with this constantly evolving landscape.

Validated Designs for Analytics

The data-driven age is dramatically reshaping industries and reinventing the future. Leveraging vast amounts of data from diverse sources is both critical and transformational. Mastering analytics holds tremendous potential for dramatically growing revenue and controlling costs. But it's often difficult know where to begin. That's where Dell Technologies can help. Validated Designs for Analytics can help unlock the value of data. The breadth of the Dell Technologies portfolio makes it easy to find a solution that's right for where you are on your analytics journey. And, together with our partners, Dell Technologies also offers consulting, installation, implementation, support and education services for analytics.

Solutions customized for your environment

Dell Technologies uniquely provides an extensive portfolio of technologies to deliver the advanced computing solutions that underpin successful analytics, AI and HPC implementations. With years of experience and an ecosystem of curated technology and service partners, Dell Technologies provides innovative solutions, workstations, servers, networking, storage and services that reduce complexity and enable you to capitalize on a universe of data.

> "Our partnership with Dell Technologies allows us to take advantage of the full breadth and depth of their compute, storage, networking and security solutions."
>
> — David J. Brzozowski Jr, Chief Technology Officer, Medacist[2]

[1] Forbes, 6 Predictions About Data In 2020 And The Coming Decade, January 2020.

[2] Dell Technologies case study, Medacist Advances Healthcare Analytics with AI running on Dell EMC PowerEdge and PowerScale, January 2021.

Hadoop use cases

The use cases for Hadoop are diverse, but there are common patterns across industries and verticals. Here is a sampling of possible use cases utilizing Validated Designs for Hadoop.

Operational efficiency

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 the enterprise.	Reduces TCO and eases compliance.	Provides enterprise storage features for storagecentric 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.	• Lower costs for active archive. • Use for long-term tiered storage for regulatory compliance. • Get multi-protocol support for storage consolidation.

Business transformation

Marketing	Finance	Healthcare	Pharmaceutical	Manufacturing
Anticipate customer needs.	Reduce risk and detect fraud.	Improve patient care and reduce costs.	Ensure regulatory compliance and validation.	Achieve 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

Do any of these challenges sound familiar?

"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 inhouse Hadoop expertise can lead to project delays and configuration problems. For over a decade, Dell Technologies has helped organizations by providing expert guidance to streamline the design, implementation and optimization of Hadoop architectures.

"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. However, data storage in Hadoop is dramatically less expensive than legacy storage in data warehouses. For example, RealPage® was able to reduce physical storage requirements by two-thirds for their Hadoop data lake with Dell EMC PowerEdge servers running in a VMware® environment with Dell EMC Isilon network-attached storage (NAS).[3]

"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 timeconsuming. Validated Designs for Hadoop have been engineered to deliver outstanding performance.

Validated Designs for Hadoop

Dell Technologies has invested to create a portfolio of Validated Designs to simplify the configuration, deployment and management of Hadoop clusters. These trusted designs have been optimized, tested and tuned for a variety of key Hadoop use cases. They include the servers, storage, networking, software and services that have been proven in our labs and in customer deployments to meet workload requirements and customer outcomes. The modular solution building blocks provide a customized yet validated approach for deploying new clusters and scaling or upgrading existing environments.

Validated Designs for Hadoop have been jointly engineered to optimize investments, reduce costs and deliver outstanding performance.

Leverage an optimized solution.

Validated Designs for Hadoop are jointly engineered with Cloudera. These solutions are tested, tuned and optimized so you can realize Hadoop benefits faster than implementing on your own. Based on over a decade of Dell Technologies experience with enterprise Hadoop installations, Validated Designs for Hadoop can be delivered as integrated solutions, with all the hardware, software and services to quickly get Hadoop into production.

Reduce costs.

By extracting, transforming and loading data from legacy data warehouse environments into Hadoop (ETL offload), you can realize significant hardware, software and administrative cost savings. In fact, with Validated Designs for Hadoop, customers report using one-third fewer servers and two-thirds less physical storage along with avoiding staff retraining costs.[3]

"Dell solved our data lake challenge with PowerEdge servers running in a VMware environment with Dell EMC Isilon network-attached storage. This gave us the power and throughput we needed, while reducing physical storage by two-thirds. We also didn't have to retrain our IT team to deal with bare metal."

— Barry Carter,
Chief Information Officer,
RealPage[3]

"Using Dell EMC PowerEdge Servers, we are able to increase our platform capacity in a matter of weeks versus a matter of months."

— Jun Chen,
Senior Vice President of
Technology Services, Epsilon[5]

Deliver outstanding performance.

With Validated Designs for Hadoop, queries and analytics that usually take days can be completed in minutes, and certain jobs can be completed near instantaneously thanks to the parallel processing power of Hadoop.[2] Another customer saw a 50X performance increase.[4]

Customer success stories

Medacist

5 minutes	**Millions of dollars**	**13% savings**
Instead of 24 hours for delivering analytics results	Saved due to upholding 99.99% uptime SLAs	Using Dell EMC PowerEdge for a Hadoop cluster running AI

Read the case study: Medacist Advances Healthcare Analytics with AI running on Dell EMC PowerEdge and PowerScale.

RealPage

5.3 billion	**1/3 fewer**	**2/3 less**
Transactions handled daily	Servers required	Physical storage required

Read the case study: RealPage Boosts Property Management Performance and Results with Analytics.

Informatica®

5 billion	**40**	**360-degree view**
Records managed	Different source systems brought together in a single view	Of customer drives micro-segmentation and hyper-personalization

Read the case study: Informatica: Building a 360 customer view with a modern data platform.

- The University of Cambridge UK Science Cloud uses Hadoop to expand the boundaries of AI.
- Mastercard® applies 1.9 million rules to 165 million transactions per hour in a matter of milliseconds.
- The Texas Advanced Computing Center (TACC) saw up to a 50X performance increase.
- Epsilon® sends billions of emails daily, with campaign adjustments in real time.

Read more customer stories.

[4] Dell Technologies case study, TACC Powers Wrangler HPC with Dell EMC Servers and Flash Storage, accessed February 2021.

[5] Dell Technologies case study, Epsilon, accessed February 2021.
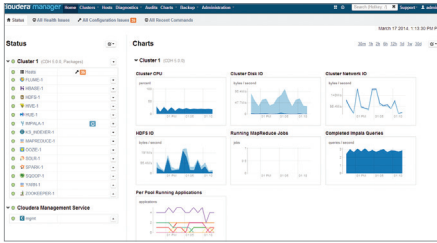
Cloudera software

Technical specifications

Validated Designs for Hadoop

Dell Technologies offers multiple Validate Designs for Hadoop, so you can choose the right Hadoop solution for your needs.

Validated Designs for Hadoop
Address analytics requirements, reduce costs and deliver outstanding performance

Many organizations struggle to begin their analytics journey or to make Apache Hadoop projects successful. Many are constrained by a lack of Hadoop expertise and end up spending too much time and effort on the front-end work. Expertise and infrastructure matter when building a Hadoop environment. That's why Dell Technologies has teamed up with industry leaders such as Cloudera® to remove the uncertainty and barriers around deploying Hadoop. Cost effective, future-ready Validated Designs for Hadoop are easy to implement solutions that help efficiently harness Hadoop and the power of analytics to drive competitive advantage.

Configuration options				
	Dell EMC PowerEdge servers			**PowerScale Isilon Scale-out NAS**
	R740xd — 3.5"	R740xd — 2.5"	R640	
Scalability	Up to 288 nodes	Up to 252 nodes	Up to 288 nodes	Up 168 compute nodes Up to 84 storage nodes
Raw storage	64TB / node	24TB / node	24TB / node	102TB / Isilon H600 storage
Processors	Dual Intel Xeon Gold 6140 2.3GHz, 18C/36T		Dual Intel Xeon Gold 6136 3.0G, 12C/24T	Dual Intel Xeon E5 2680 v4 2.4GHz 14C/28T
	Dell EMC PowerSwitch networking			
iDRAC network	S3048-ON			
Pod network	Z9100 25GbE	S4048-ON 10GbE	Z9100 25GbE	S4048-ON 10GbE
Cluster aggregation network	Z9100 100GbE	S6010-ON 40GbE	Z9100 100GbE	S6010-ON 40GbE

Cloudera Private Cloud Base on Dell EMC Infrastructure
Multi-function analytics and data management platform for on-premises IT environments

Most organizations understand the importance of extracting insights from data. However, the considerations and requirements for data management are constantly evolving. Cloudera Data Platform (CDP) Private Cloud Base is a scalable and customizable Hadoop platform for securely running many types of analytics workloads. When paired with Dell EMC infrastructure, the potential use cases that can be addressed through full-featured data platform are nearly limitless.

Configuration options					
Dell EMC PowerEdge servers				**Dell EMC PowerSwitch networking**	**Software**
Master nodes	**Utility node**	**Edge node**	**Worker nodes**		
3x R640	1x R640	1x R640	3x R740xd	S5248F-ON Z9264F-ON	Cloudera CDP Private Cloud Base

Cloudera Private Cloud Experiences on Dell EMC infrastructure
Run modern analytics services with cloud-like scalability and agility on-premises

Extracting insights from data is a critical activity for the business to survive and thrive in our data-driven era. But many IT organizations are challenged to provide the resources for advanced analytics as quickly as business users demand them. Deploying Cloudera CDP Private Cloud Experiences in tandem with CDP Private Cloud Base brings cloud-native, self-service analytic experiences to your customer's private data center. With CDP Private Cloud Experiences, users can rapidly provision and deploy analytics services — such as data warehouse, machine learning, and traditional workloads — through the management console, and easily scale them up or down as required.

Configuration options				
Dell EMC PowerEdge servers			**Dell EMC PowerSwitch networking**	**Software**
Master nodes	**Worker nodes**	**Container services node**		
3x R640 (minimum)	3x R640 (minimum)	1x R640 (minimum)	S3148-ON S5248F-ON Z9263F-ON	• Cloudera CDP Private Cloud Base • Red Hat Open Shift Container Platform • Kubernetes

Shared storage vs. distributed storage in 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 PowerScale is a shared storage model where the persistent file system 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 (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

Solution highlights

- **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 PowerSwitch Networking:** Today's analytics workloads call for new thinking about network architecture. Based on open standards, Dell EMC Networking frees the data center from outdated, proprietary approaches. Our future-ready networking technology helps improve network performance, lower networking costs and provide the flexibility to adopt new innovations. Help your customers take control of their network's future and learn how the Dell Technologies strategy for open networking can dramatically transform their businesses.
- **Dell EMC Isilon H600 hybrid scale-out NAS** brings the performance of 120 SAS drives, SSD caching and built-in multiprotocol capabilities. Management stays simple as you grow with the single file system, single volume architecture and automation features of Isilon OneFS.

Services and financing

Dell Technologies is there every step of the way, linking people, processes and technology to accelerate innovation and enable optimal business outcomes.

- Analytics Consulting Services help customers create a competitive advantage. Our expert consultants work with companies at all stages of analytics to help plan, implement and optimize solutions that enable customers to unlock data capital and support advanced techniques, such as AI and machine learning (ML).
- Deployment Services help streamline complexity and bring new IT investments online as quickly as possible. Customers can leverage our 30+ years of experience for efficient and reliable solution deployment to accelerate adoption and ROI while freeing IT staff for more strategic work.
- Support Services driven by AI and deep learning will change the way customers think about support with smart, ground-breaking technology backed by experts to help maximize productivity, uptime and convenience. Customers can experience more than fast problem resolution — our AI engine proactively detects and prevents issues before they impact performance.
- Payment Solutions from Dell Financial Services help customers maximize IT budgets and get the technology required sooner. Our portfolio includes traditional leasing and financing options, as well as advanced flexible consumption products.
- Dell Technologies On Demand offers a simple approach that gives customers a wide range of consumption models, payment solutions and services so they can optimize for a variety of factors while realizing more predictable outcomes.
- Managed Services can help reduce the cost, complexity and risk of managing IT so customers can focus resources on digital innovation and transformation while our experts help optimize IT operations and investments.
- Residency Services provide the expertise needed to drive effective IT transformation and keep IT infrastructure running at its peak. Resident experts work tirelessly to address challenges and requirements, with the ability to adjust as priorities shift.

"With Dell, we get world-class support, so we can avoid the finger-pointing you get with competing vendors. This is a key to our relationship."

— Barry Carter,
Chief Information Officer,
RealPage[3]

Why choose Dell Technologies

Were committed to advancing analytics, HPC and AI, and we've dedicated a great deal of resources toward that goal.

- Schedule customers an executive briefing where they can collaborate on ways to reach their business goals.
- Dell Technologies Customer Solution Centers are staffed with computer scientists, engineers and subject matter experts in a variety of disciplines.
- We are committed to providing customers with choice. We want our customers to get what they need and have a great experience working with us. If we don't have what they need, we'll tell them who does. We believe in being open, and we publish our performance results.
- Dell Technologies is the only company in the world with a portfolio that spans from workstations to supercomputers, including servers, networking, storage, software and services.
- Because Dell Technologies offers such a wide selection of solutions, we can act as your trusted advisor without trying to sell you a one-size-fits-all approach to your problem. That range of solutions has also given us the expertise to understand a broad spectrum of challenges and how to address them.

Customer Solution Centers

Our global network of dedicated Dell Technologies Customer Solution Centers are trusted environments where world-class IT experts collaborate with you to share best practices, facilitate in-depth discussions of effective business strategies and help your business become more successful and competitive. Dell Technologies Customer Solution Centers reduce the risks associated with new technology investments and can help improve speed of implementation.

AI Experience Zones

Curious about AI and what it can do for your business? Run demos, try proofs of concept and pilot software in Singapore, Seoul, Sydney, Bangalore and other Customer Solution Centers. Dell Technologies experts are available to collaborate and share best practices as you explore the latest technology, and get the information and hands-on experience you need for your advanced computing workloads.

HPC & AI Innovation Lab

The Dell Technologies HPC & AI Innovation Lab in Austin, Texas, is the flagship innovation center. Housed in a 13,000-square-foot data center, it gives you access to thousands of Dell EMC servers, three powerful HPC clusters, and sophisticated storage and network systems. It's staffed by a dedicated group of computer scientists, engineers and subject matter experts who actively partner and collaborate with you and other members of the HPC community. The team engineers HPC and AI solutions, tests new and emerging technologies, and shares expertise including performance results and best practices.

HPC & AI Centers of Excellence

As analytics, HPC and AI converge and the technology evolves, Dell Technologies worldwide HPC & AI Centers of Excellence provide thought leadership, test new technologies and share best practices. They maintain local industry partnerships and have direct access to Dell Technologies and other technology creators to incorporate your feedback and needs into their roadmaps. Through collaboration, Dell Technologies HPC & AI Centers of Excellence provide a network of resources based on the wide-ranging know-how and experience in the community.

Proven results

Dell Technologies holds leadership positions in some of the biggest and largest-growth categories in the IT infrastructure business, and that means customers can confidently source information technology needs from Dell Technologies.

- #1 in servers[6]
- #1 in converged and hyperconverged infrastructure (HCI)[7]
- #1 in storage[8]
- #1 cloud IT infrastructure[9]

See Dell Technologies Key Facts.

Take the next step, today.

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

---

[6] IDC WW Quarterly Server Tracker, Units & Revenue, September 2021.

[7] IDC WW Quarterly Converged Systems Tracker, Vendor Revenue, March 2021.

[8] IDC WW Quarterly Enterprise Storage Systems Tracker, Vendor Revenue, September 2021.

[9] IDC WW Quarterly Cloud IT Infrastructure Tracker, Vendor Revenue, July 2021.

Contact us

To learn more, visit
DellTechnologies.com/Analytics
or contact your local representative
or authorized reseller.

DELLTechnologies