DELL EMC

Big Data and Analytics

DELL EMC HADOOP SOLUTIONS

Helping Organizations Capitalize on the Digital Transformation

The digital transformation: a disruptive opportunity

Across virtually all industries, the digital transformation is disrupting legacy business models and creating new and exciting opportunities. At the heart of this transformation is explosive data growth.

A forecast from IDC estimates that the digital universe — the data we create and copy annually — is doubling in size every two years, and by 2020 will reach 44 zettabytes, or 44 trillion gigabytes.[1] To put this into perspective, one gigabyte of data can supply you with enough music to listen for 16 hours, but just one zettabyte would keep you busy listening to music for nearly 2 billion years.[2]

Thanks to the rise of the Internet of Things, social media and next-generation applications, most organizations now contribute to the growth of the digital universe by generating and capturing a huge amount of data every day. This data can potentially transform the way a business operates — for those who have the systems and solutions in place to turn raw data into valuable business insights.

The truth is, capitalizing on data isn't always easy or straightforward, and some companies are faring better than others in the race to become a digital business. But either way, the revolution is already here — ready or not.

1 IDC. "The Digital Universe of Opportunities: Rich Data and the Increasing Value of the Internet of Things." April 2014.

2 Based on a mobile encoding rate of 1 megabyte per minute and a recording length of 4 minutes, a petabyte of songs would play for 2,000 years and a zettabyte (1 million times larger than a petabyte) would play for 2 billion years.

Why do some companies struggle, while others thrive?

Some businesses clearly aren't ready to capitalize on digital transformation. Some see a chaotic, uncertain future ahead.

Consider these findings from a study that probed the impact of the digital transformation on businesses.[3] Among business leaders:

- 48 percent don't know what their industry will look like in three years
- 78 percent feel threatened by digital startups
- 45 percent fear they may become obsolete in three to five years

Digital transformation efforts are being hampered by a wide range of issues — particularly a lack of budget and resources. The lack of resources often points to a bigger issue: Digital transformation is not a central strategic pillar for many organizations.

3 Vanson Bourne. Digital Business Research Survey, 2016.

Capitalizing on digital transformation

Forward-looking business leaders know they must invest in the digital transformation to stay relevant. Research by Vanson Bourne found that business leaders agree that certain digital business attributes are imperatives for success:

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

For those organizations who embrace the digital revolution, bigger data equates to bigger opportunities. A few examples:

- An online retailer can draw on many sources of data to create a directed preference engine for online shoppers.
- A financial services company can improve risk assessment models using multiple data inputs.
- A healthcare provider can leverage big data for genome processing and narrowly targeted therapies.

To realize these and a myriad of other opportunities, organizations need to gather insights from vast volumes of varied data types, many in real time, and for these tasks they need more than traditional, structured systems and tools. Big data analysis requires a powerful and scalable infrastructure with best-of-breed software solutions. And this is where Apache™ Hadoop® platforms, technologies and solutions come into play.

The Hadoop value proposition

What is Hadoop?

Hadoop is an open source platform that is designed to store and process large datasets in a distributed computing environment. It has two main sub-projects: Hadoop Distributed File System (HDFS) for data storage and MapReduce for data processing. Hadoop breaks down large datasets across inexpensive commodity servers to process the data in parallel.

What makes Hadoop unique?

Organizations turn to Hadoop for both business and technology advantages.

At a business level, Hadoop offers a compelling value proposition from a total cost of ownership standpoint. Hadoop uses industry-standard servers and storage, decreasing the cost to store and process huge datasets versus traditional existing business intelligence (BI) and analytics solutions. In addition, cost efficiencies can be achieved via a data lake using scale-out NAS, such as Dell EMC Isilon.

From a technology standpoint, Hadoop offers:

- **Scalability** — Organizations can add unlimited nodes to a Hadoop cluster and process the data in parallel and at scale.
- **A high degree of fault tolerance** — Hadoop's distributed architecture can direct jobs between alternate nodes in the unlikely event of a node fault, reducing risk and ensuring uptime.
- **Unmatched flexibility and consolidation** — Organizations can store all data types from any data source on a Hadoop cluster within a single file system (HDFS) using either scale out NAS storage or direct attached storage, without the traditional BI requirement of transforming the data to a well-known format, or schema.
- **Future-ready components** — Apache Spark, a key component of real-time analytics in a Hadoop environment, provides an open source, parallel data processing framework that complements the capabilities of Apache Hadoop.

Moving forward with Dell EMC

Overcoming the barriers

If your organization is on the path to digital transformation and Hadoop-based solutions, Dell EMC is an ideal partner. Dell EMC is uniquely positioned to help you overcome the top roadblocks in the route to digital transformation.

What is the ROI on Hadoop investments?

Dell EMC and Intel® commissioned Forrester Consulting to conduct a Total Economic Impact™ (TEI) study to examine the potential ROI enterprises may realize by deploying the Dell EMC Cloudera® Hadoop Solution, accelerated by Intel. The goal was to provide a framework to help organizations evaluate the potential financial impact of deploying the Hadoop platform in a big data solution.

Based on interviews with organizations using the Dell EMC Cloudera Hadoop Solution, the study identified these three-year risk-adjusted results:

- ROI: 97 percent—meaning for every $1.00 spent, $1.97 is returned
- Net present value (NPV) per node: $58,542
- Payback: Six months
- Revenue growth: 20 – 30 percent

To review the study results in more depth, download the Forrester Total Economic Impact study.

Research by Vanson Bourne identified these top barriers as:

1. Insufficient budget and resources
2. Lack of executive support
3. Inadequate expertise and skills
4. Technologies that can't work at the speed of business
5. Data privacy and security concerns

To enable organizations to overcome their budgetary and resource limitations and build executive support for Hadoop deployments, Dell EMC documents the value of solutions via proofs of concept in Dell EMC Customer Solution Centers and a QuickStart solution that enables a predictable Hadoop deployment on a small scale. Dell EMC also provides third-party reports that document the ROI on Hadoop deployments.

To help organizations overcome their expertise and skills shortages and accelerate the deployment of Hadoop environments that grow with the business, Dell EMC provides a portfolio of validated reference architectures and scalable solutions for Hadoop deployments. Dell EMC backs these offerings with a wide range of professional consulting and support services.

And to help organizations address their data privacy and security concerns, Dell EMC solutions incorporate proven Hadoop software distributions that include a wide range of enterprise-class security features.

Leveraging technologies, solutions and services

If your organization is on this path to the future, Dell EMC is an ideal partner for your journey. Dell EMC can help you build a plan and strategy for your digital transformation journey and identify your ideal technologies and solutions.

Dell EMC has a robust big data and analytics portfolio that provides choice in technologies, consumption models and analytical tools. This portfolio allows Dell EMC customers to either build with reference architectures and validated systems or buy turnkey engineered systems with white glove service and lifecycle support.

Regardless of the chosen deployment model, Dell EMC helps organizations achieve their digital transformation imperatives by gaining deeper insights from all their data, reducing the total cost of ownership and business risk for greater and more predictable ROI and improving productivity for end users by providing flexibility in tools and applications.

The measurable advantages of the experience and expertise that Dell EMC brings are clear. With a comprehensive portfolio of data infrastructure that can integrate, store, manage, analyze, protect and enable analytics and application development, as well as an all-data integration approach capable of incorporating all relevant data from any source, including the cloud, Dell EMC delivers advanced analytic solutions that empower end users with deeper insights and enhanced data-driven decision making.

The Dell EMC Hadoop portfolio

There is no single "right answer" for success with data analytics. It is a journey of continual growth. Every organization's data is unique, and must be treated as such. Solutions that are perfect for one company may not address the needs of another.

With this thought in mind, Dell EMC offers a wide range of products and solutions to address diverse big data and analytics challenges — from starter bundles and validated reference architectures to integrated appliances and engineered solutions, or even completely customized solutions for your specific environment.

Many of these solutions leverage the Apache® Hadoop™ platform, analytic software for Hadoop and best-of-breed technologies from Hadoop partners, along with Dell EMC

The Dell EMC difference

Here are some of the ways that Dell EMC can help your organization take control of your data:

- Begin your big data journey with a collaborative partner and leader in big data solutions that can help you identify use cases and refine your goals.
- Simplify and accelerate Hadoop deployments by integrating with and augmenting your existing legacy relational databases and traditional data warehouse environments using secure, tested and validated reference architectures.
- Increase your confidence with complete end-to-end implementation services — from installation and configuration to data science and use case development to support and managed services — from Dell EMC or our Partners around the world.

DELL EMC

servers, storage and networking. These solutions can also include services from Dell EMC or our extensive network of global partners for solution design, installation and configuration.

Getting started with Hadoop

Dell EMC QuickStart for Cloudera Hadoop

Dell EMC QuickStart for Cloudera Hadoop provides an easy and affordable way for organizations to test a fully-supported big data Hadoop solution. Dell EMC QuickStart for Cloudera Hadoop includes the hardware, the software — including Cloudera Basic Edition — and the services needed to deliver a Hadoop cluster to begin a proof of concept.

Taking the first steps in a Hadoop big data solution with Dell EMC QuickStart allows organizations to quickly deploy their clusters to realize revenue, increase customer loyalty and pinpoint effective strategies to address the business and technical demands of a big data solution implementation.

Dell EMC Cloudera Syncsort® ETL Offload Hadoop Solution

An ideal initial use case for Hadoop includes the offload of extract-transform-and-load (ETL) processing from Enterprise Data Warehouse (EDW) systems. As traditional EDWs hit their limits, many organizations are deploying the open source Hadoop platform as a complement to their existing EDWs and then offloading data warehouse processing functions to Hadoop. With a robust offload solution, your organization can accelerate ETL processing, work easily with a wide range of new data sources and formats and modernize systems to make better use of your existing EDW investments.

Dell EMC offers a proven solution for organizations interested in ETL offload: the Dell EMC Cloudera Syncsort ETL Offload Hadoop Solution, accelerated by Intel. This groundbreaking solution allows organizations to capitalize on the unique technical and cost advantages of the Hadoop platform while making better use of existing EDW investments.

The Dell EMC Cloudera Syncsort ETL Offload Hadoop Solution delivers a use-case driven Hadoop reference architecture to guide data warehouse optimization efforts. This solution can incorporate:

- Leading Hadoop distributions from Dell EMC partners, including Cloudera and Hortonworks®
- A rich framework and toolset for ETL offload from Syncsort
- Dell EMC PowerEdge™ R series servers with Intel® Xeon® processors
- Dell EMC networking components
- Optional consulting and integration services

The value of these solutions was underscored in a research study by Principled Technologies, a technology testing and analysis firm. Principled Technologies found that the Dell EMC Cloudera Syncsort ETL Offload Solution can help organizations drive operational efficiency by completing Hadoop ETL jobs faster, simplifying the Hadoop ETL design process and saving thousands of dollars on Hadoop ETL jobs.

In specific terms, the study determined:

- A Dell EMC Cloudera Syncsort ETL Offload Hadoop Solution fully implemented by an entry-level employee could reduce data warehouse administrative costs by 76 percent.[4]
- ETL jobs created by an entry-level technician using the Dell EMC Cloudera Syncsort ETL Offload Hadoop Solution ran up to 60 percent faster than a solution created by a Hadoop expert using open source tools.[5]
- A Dell EMC Cloudera Syncsort ETL Offload Hadoop Solution enables less-experienced users to develop and design complex data transformation jobs up to 54 percent faster.[6]

4 Principled Technologies. "Cost Advantages of Hadoop ETL Offload with the Intel Processor-Powered Dell | Cloudera | Syncsort Solution." July 2015.

5 Principled Technologies. "Performance Advantages of Hadoop ETL Offload with the Intel Processor-Powered Dell | Cloudera | Syncsort Solution." July 2015.

6 Principled Technologies. "Design Advantages of Hadoop ETL Offload with the Intel Processor-Powered Dell | Cloudera | Syncsort Solution." July 2015.

Dell EMC Data Lake

Another option for getting started with Hadoop is to deploy a Dell EMC Data Lake using the Hadoop platform in conjunction with Dell EMC scale-out NAS Isilon storage running the OneFS operating system. The Data Lake gives you one single system to capture, store, analyze, protect and manage your data. Everything is faster without the need to move and copy data across silos. The Data Lake approach can help your organization speed time to insights, improve storage utilization, eliminate islands or silos of storage and lower the storage management costs of migration, security and protection.

Deploying a broader Hadoop environment

Dell EMC Cloudera Hadoop Solution

The Dell EMC Cloudera Hadoop Solution provides a proven, end-to-end solution based on Dell EMC PowerEdge™ servers, Dell EMC Networking and Cloudera Enterprise software. Powered by Apache Hadoop, Cloudera Enterprise is a fast, easy and secure modern data platform. Cloudera Enterprise is thoroughly tested and validated to integrate with a wide range of operating systems, hardware, databases, data warehouses, business intelligence and ETL systems. This broad compatibility enables your organization to take advantage of Hadoop while leveraging your existing tools and resources.

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 Internet of Things and model scoring and serving.

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.

Dell EMC Hortonworks Hadoop Solution

The 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.

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

With Dell EMC, you have the confidence that comes with experience we've gained 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 — while reducing risk.

Dell EMC Hadoop service offerings

Hadoop solution installation and implementation

Dell EMC offers a broad menu of installation and implementation services for Hadoop solutions. These services include onsite hardware and software installation, optional rack integration at a Dell EMC facility and validation of the installed solution. Dell EMC can take care of the complete project management, from order drop to customer acceptance.

Consulting services for Hadoop and data analytics

Dell EMC Hadoop Consulting

Dell EMC Hadoop Consulting is a best-in-class service delivered by certified Cloudera Hadoop experts to help organizations fully exploit the business value of big data using Hadoop. The services may include a big data assessment, workshop, proof-of-concept or production implementation. These Hadoop experts help determine where Hadoop is a good fit for an organization. They also help Dell EMC customers build their own teams of Hadoop experts through knowledge transfer at each step.

Dell EMC Services for data analytics

Dell EMC Services offerings start with strategic consulting and deployment services to help business leaders prioritize where and how to apply data and analytics to drive their business initiatives, and to help technology leaders understand how to build big data and analytics capabilities into their IT architectures and processes.

The Big Data Vision Workshop employs a unique methodology to identify and prioritize a single analytics use case with the best combination of implementation feasibility and business value. It's a three-week engagement that applies research, interviews, data science expertise and techniques to your organization — culminating in a one-day workshop to identify and agree on the best analytics use case and path forward to solving a business problem. From here, customers can move on to a Proof of Value service, to prove out the potential ROI of a proposed analytics solution.

Technology leaders who seek to expand analytics capabilities to meet growing demands for better/faster data and analytics start with the Big Data Technology Advisory service, a three-week engagement to identify strategic capabilities for data, determine a future-state architecture, perform a gap analysis and build a deployment roadmap. They can then opt for a Proof of Technology service to pilot their solution and confirm the desired capabilities before moving into production.

At this point, the customer can choose the Big Data Technical Implementation to deploy their use case as a custom solution that is designed to support that one use case, or deploy a custom data lake-based platform that can support additional use cases. Alternatively, the customer can choose to implement an analytics use case by deploying a packaged, pre-engineered solution, such as the Analytic Insights Module, with included support. Dell EMC Services offers comprehensive consulting engagements to develop and implement an analytics use case into production, including solution architecture, data ingestion, data science services and application integration.

Dell EMC ProSupport

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 support for Cloudera Enterprise software.

Additionally, ProSupport includes next-business-day onsite service with four- and eight-hour parts and labor response options and escalation management with customer-set severity level options.

Work with the leaders in big data

Every organization has unique needs and unique requirements for big data and analytics solutions. What's more, every organization has existing infrastructure to maintain. At Dell EMC, we understand this need, and work actively to help our customers implement future-ready solutions that bridge the gap between traditional IT and new IT — solutions that transform with the IT organization, so IT can transform the business.

With its deep roots in big data solutions and the Hadoop platform and its leading partners in big data analytics, Dell EMC has the knowledge, expertise, tools and solutions needed to drive successful, flexible and scalable Hadoop deployments. Dell EMC big data solutions help organizations store, analyze and gain valuable insights from their data to build competitive advantages, protect their businesses and drive measurable growth.

Ultimately, Dell EMC big data solutions can help all sizes of organizations efficiently harness the Hadoop platform and the power of big data to differentiate the business.

Let's get started

Gets hand-on experience in a Dell EMC Customer Solution Center

As you consider your path forward, you can gain hands-on experience with Hadoop in a Dell EMC Customer Solution Center.

Dell EMC offers a global network of dedicated state-of-the-art technical labs that enable customers to architect, validate and build solutions, from the data center to the edge of the network. Our Customer Solution Center experts work collaboratively with organizations to explore, test and prove end-to-end solutions that help drive business outcomes and create competitive advantage.

The combination of dedicated labs, teams and infrastructure provides IT teams with a complete experience, allowing them to talk to the experts, understand the relevant technologies and get hands-on knowledge of Dell EMC solutions. And through a remote infrastructure, organizations can engage through the cloud, enabling virtual access for globally distributed teams.

Ultimately, through your Customer Solution Center experience, you'll gain valuable experiences and insights—so you can move forward with confidence.

Learn more visit:
Dell.com/Hadoop and EMC.com/BigData

For questions, please contact us at:
Hadoop@Dell.com

Share

DELL EMC

intel