



Dell Big Data Solutions

Dell Apache Hadoop Solutions for Big Data

Unlock business critical insights from data of any type and size



Data is the new oil

Data growth is exploding, and analyzing large datasets—big data—has become the next frontier for innovation, competition and productivity. In many ways, data is the new oil—waiting to be mined, transformed and put into production to fuel countless opportunities.

Consider this finding: 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. That's nearly as many digital bits as there are stars in the universe.⁴ Anymore, a single project can yield a once-unfathomable amount of data. For example, since its launch in 1990, the Hubble Space Telescope has collected more than 45 terabytes of data.⁵

This vast amount of data creates new business opportunities. A few examples: An online retailer can draw on big 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 near 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 the Apache Hadoop platforms, technologies and solutions from Dell, Cloudera and Intel come into play.

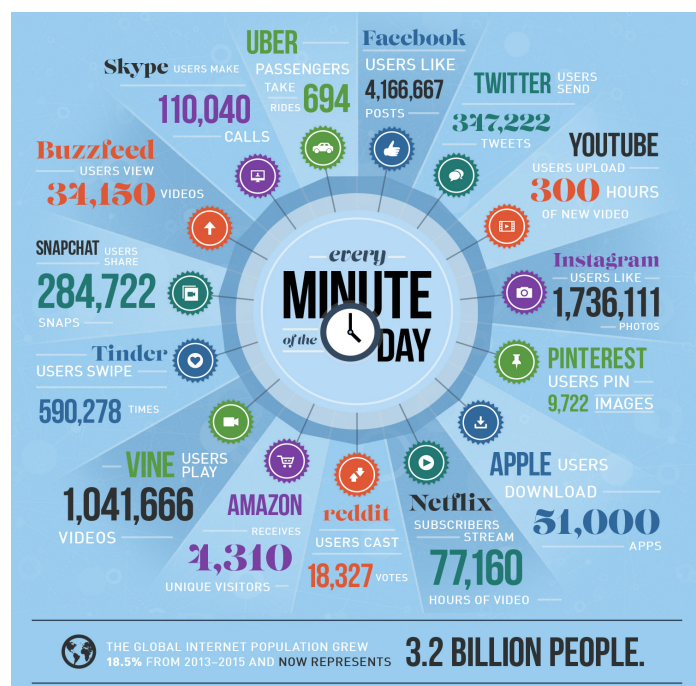


Figure 1. Every Minute of Every Day³

The Dell Difference

Since 2009, Dell has invested heavily in technologies, solutions and partnerships that help organizations capitalize on the Hadoop platform and ever-growing volumes of data. Here are some of the ways that Dell continues to help organizations take control of their data:

- Simplify and accelerate Hadoop deployments by integrating with existing legacy relational database environments using a tested and validated reference architecture
- Implement Hadoop with confidence with enterprise-class software distributions, including CDH, Cloudera's open source software distribution.
- Derive actionable business insights with an extensive range of Hadoop-ready business analytic software solutions.
- Protect the business with enterprise-class security features in the evolving Hadoop platform.
- Leverage complete end-to-end implementation services—from installation to configuration and support services—from the Dell Support organization.
- Scale economically with Intel-based Dell™ PowerEdge™ servers that offer the performance, energy efficiency and density needed for large-scale environments.

⁴ IDC. "The Digital Universe of Opportunities: Rich Data and the Increasing Value of the Internet of Things." April 2014.

⁵ Space.com. "Hubble Space Telescope: Pictures, Facts & History," by Nola Taylor Redd, April 24, 2013.

⁶ Domo. [Data Never Sleeps 3.0](#), August 13, 2015.

Dell solutions for big data start with Apache Hadoop

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: HDFS for data storage and MapReduce for data processing. Hadoop breaks down large datasets across inexpensive commodity servers and storage to process the data in parallel.

What makes Hadoop unique?

- Reduced cost—Hadoop uses commodity servers and storage, decreasing the cost to store and process huge datasets.
- Scalability—Organizations can add unlimited nodes to a Hadoop cluster and process the data in parallel.
- A high degree of fault tolerance—Hadoop can direct jobs to other nodes, reducing the risks associated with node faults.
- Unmatched flexibility and consolidation—Organizations can store all data types from any data source on a Hadoop cluster within one file system (HDFS).
- 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. Spark is included in CDH, Cloudera's open source software distribution, to make it easy to develop fast, unified big data applications that combine batch, streaming and interactive analytics on all types of data.

What is the ROI on Hadoop investments?

Dell and Intel commissioned Forrester Consulting to conduct a Total Economic Impact™ (TEI) study to examine the potential ROI enterprises may realize by deploying Dell | Cloudera Apache Hadoop Solutions, 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 Dell | Cloudera Apache Hadoop Solutions, the study identified these three-year risk-adjusted results:

- ROI: 97%
- Net present value (NPV) per node: \$58,542
- Payback: Six months
- Revenue growth: 20% – 30%

To review the study results in more depth, download the [Forrester Total Economic Impact study](#).

Dell | Cloudera Hadoop solutions, accelerated by Intel

Together with Cloudera and Intel, Dell provides integrated, end-to-end Hadoop solutions that are designed from the ground up to address big data requirements, reduce development costs and optimize performance for deep data mining and analytics.

These solutions, based on validated reference architectures, incorporate analytic software for Hadoop, best-of-breed technologies from Hadoop partners, Dell™ PowerEdge™ servers and networking components, and Dell Professional Services for solution design, installation and configuration.

The portfolio of Dell | Cloudera Hadoop solutions, accelerated by Intel, includes the following offerings:

Dell QuickStart for Cloudera Hadoop

Dell QuickStart for Cloudera Hadoop provides an easy and affordable way for organizations to test a fully-supported big data Hadoop solution. Dell 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 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 | Cloudera Apache Hadoop 5.5 Solution

Dell | Cloudera Apache Hadoop 5.5 Solution is an enterprise-ready Hadoop solution built on Dell servers with Cloudera software and backed by optional managed Hadoop services. The solution is based on the Cloudera Enterprise 5.5 release,

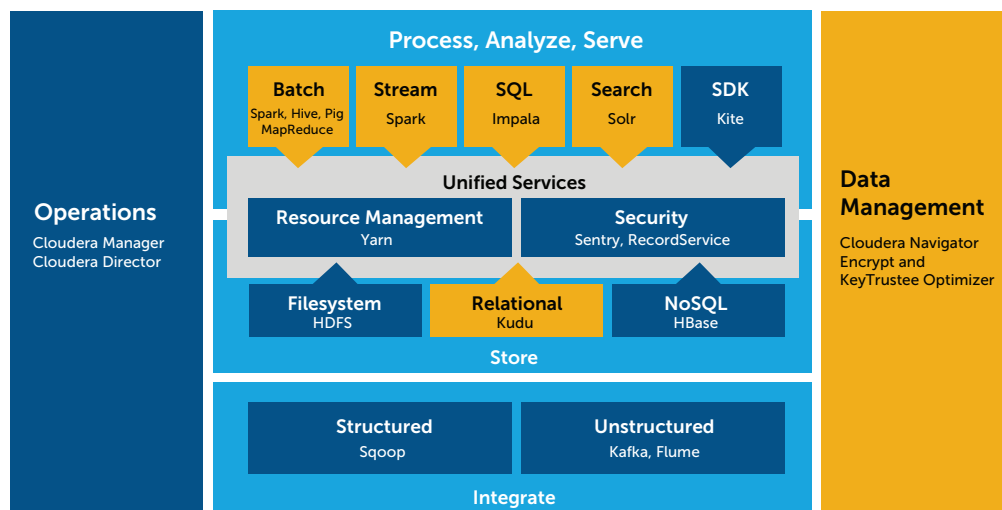


Figure 2. Cloudera Apache Hadoop for the enterprise



which continues to improve the performance, security and functionality of analytics on Hadoop.

Cloudera Enterprise 5.5 includes the limited beta release of Cloudera Navigator Optimizer for improved workload performance and efficiency. Cloudera Navigator Optimizer instantly analyzes existing workloads, providing visibility into which ones are the most critical, which data is accessed most, and how it is being used. It then automatically turns this information into a full optimization strategy for fast success with Hadoop. Another integrated part of CDH supported with Cloudera Enterprise is Spark, the open standard for flexible in-memory data processing for batch, real-time and advanced analytics.

Dell's tested and validated reference architectures include Dell PowerEdge servers and Dell Networking combined with Cloudera Enterprise. Cloudera Enterprise is thoroughly tested and validated to integrate with a wide range of operating systems, hardware, databases, data warehouses, business intelligence, and extract-transform-and-load (ETL) systems. This broad compatibility helps organizations take advantage of Hadoop while leveraging their existing tools and resources.

Ultimately, the Dell | Cloudera solution includes everything an organization needs to tackle big data challenges, including software, hardware, networking and services.

Dell | Cloudera | Syncsort Data Warehouse Optimization – ETL Offload Reference Architecture

As traditional Enterprise Data Warehouses (EDWs) hit their limits, many forward-looking 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, an 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 existing EDW investments.

That's the idea behind the Dell | Cloudera | Syncsort Data Warehouse Optimization – ETL Offload Reference Architecture, 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 | Cloudera | Syncsort Data Warehouse Optimization – ETL Offload solution delivers a use-case driven Hadoop reference architecture to guide data warehouse optimization efforts. The architecture brings together:

- The industry-leading Hadoop distribution from Cloudera
- A rich framework and toolset for ETL offload from Syncsort
- Dell™ PowerEdge™ R series servers with Intel® Xeon® processors
- Dell networking components
- Optional consulting and integration services

The value of this reference architecture was underscored in a research study by Principled Technologies, a technology testing and analysis firm. Principled Technologies found that the Dell | 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 firm determined:

- A Dell | Cloudera | Syncsort solution for Hadoop 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 | Cloudera | Syncsort solution for Hadoop ran up to 60 percent faster than a solution created by a Hadoop expert using open source tools.⁵
- A Dell | Cloudera | Syncsort solution for Hadoop enables less-experienced users to develop and design complex data transformation jobs up to 54% faster.⁶

Dell software suite

The Dell software portfolio offers an unprecedented mix of leading-edge technologies for data analytics, integration and management. Collectively, these software products cover the end-to-end needs of a data analytics environment.

Dell Statistica Big Data Analytics

Dell Statistica provides the most comprehensive array of data analysis, data management, data visualization and data mining solutions. Techniques include a wide selection of predictive modeling, clustering, classification and exploratory procedures in one software platform.

Dell Boomi

Dell Boomi integration platform as a service (iPaaS), AtomSphere and Dell Boomi MDM harness the power of Salesforce Analytics Cloud to enable every business user to instantly and confidently gain key business insights from their data, from multiple sources, from any device.

Dell Toad™ Data Point

Dell Toad Data Point is a cross-platform self-service data-integration tool that simplifies data access, data preparation and provisioning. It provides nearly limitless data connectivity, desktop data integration, visual query building and workflow automation.

Dell Toad Intelligence Central

Dell Toad Intelligence Central allows users to increase collaboration and streamline data sharing. It provides a centralized repository for accurate, up-to-date information to make data provisioning more efficient, while greatly improving productivity.

Dell SharePlex

Dell SharePlex allows users to achieve three core IT initiatives: scalability, availability and near real-time data integration with an easy, affordable, impact-free, all-in-one database replication solution.

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 Hadoop solution installation and implementation

Dell offers a broad menu of installation and implementation services for Hadoop solutions, ranging from the initial steps into Hadoop with the Dell QuickStart for Cloudera Hadoop proof-of-concept bundle, to the Dell | Cloudera validated Reference Architecture, to the Dell | Cloudera | Syncsort Data Warehouse Optimization—ETL Offload Reference Architecture.

Dell services include onsite hardware and software installation, optional rack integration at a Dell facility and validation of the installed solution. In all of the Dell | Cloudera Hadoop solutions, Dell takes care of the complete project management, from order drop to customer acceptance.

Dell Hadoop Consulting

Dell 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 include a big data assessment, workshop, proof of concept and production implementation. These Hadoop experts help determine where Hadoop is a good fit for an organization. They also help Dell customers build their own teams of Hadoop experts through knowledge transfer at each step.

Dell Customer Solution Centers

Dell 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 solutions. And through a remote infrastructure, organizations can engage through the cloud, enabling virtual access for globally distributed teams.

Dell ProSupport

Dell 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.



To learn more visit: Dell.com/Hadoop and Dell.com/BigData

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

To complete Dell's approach to big data

Adding to the Dell | Cloudera Hadoop solutions and the Dell Software Suite, Dell enables and accelerates the big data journey with the reference architectures and engineered solutions that complete the Dell Big Data and Analytics portfolio.

Reference architectures

Dell offers tested and certified reference architectures for small, medium and large data analytics deployments, drawing on technologies from Dell and strategic partner industry leaders. These reference architectures include all the hardware and software needed to contain costs and increase the scalability of a solution for a complete big data environment.

Dell DWFT Reference Architectures for SQL Server

Dell Data Warehouse Fast Track (DWFT) Reference Architectures for SQL Server deliver innovation to ramp up data warehousing and analytics performance, including advances in server and networking hardware as well as leading-edge feature-rich SAN storage.

Engineered solutions

The following engineered solutions from Dell bring together the components of complete data analytics environments in turnkey appliances that are easy to order and quick to deploy.

Dell Microsoft Appliance Analytics Platform System (APS) Appliance

Dell delivers a solution for real-time management of relational (SQL) and non-relational (Hadoop) data.

Dell SAP HANA Appliance

Dell delivers an in-memory analytics appliance engineered by Dell and certified by SAP to ensure non-disruptive, modular scalability from two to 18 nodes, without requiring replacement or reconfiguration.

Work with the leaders in big data

When organizations work with Dell for their Hadoop projects, they have the confidence that comes with a partner who has provided innovative server platforms for big data since 2009 and was an early leader in Hadoop implementations. Together with Cloudera and Intel, Dell has designed custom Hadoop platforms since 2009 and published its first Hadoop reference architecture on Cloudera in 2011.

With its deep roots in big data solutions and the Hadoop platform, and its leading partners in big data analytics, Dell has the knowledge, expertise, tools and solutions needed to drive successful, flexible and scalable Hadoop deployments. Dell 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 big data solutions can help all sizes of organizations efficiently harness the Hadoop platform and the power of big data to differentiate the business.

