



Handling Hadoop Together: Dell | Cloudera | Intel

Three leaders. One voice. Infinite possibilities.

A Dell Big Data White Paper



The big data opportunity

To consider the scope of today's big data challenges, it can help to begin with the big picture: the massive amounts of data we are creating on a daily basis. By one estimate, the digital universe is doubling in size every two years and will multiply 10-fold between 2013 and 2020—rising from 4.4 trillion gigabytes to 44 trillion gigabytes.¹

This nonstop data deluge is both a challenge and an opportunity for private enterprises and public organizations. In a blog post, Dell Chairman and CEO Michael Dell summarized today's data dilemma in these terms:

"Organizations of all sizes are trying to figure out how to use all of this data to deliver a better customer experience and build new business models. Consumers are struggling to balance a desire for automated, personalized services with the need for safety. Governments are pressured to

*use all available data in support of national security, but not at the expense of citizens' right to privacy. And underlying it all is the realization that data, if managed, secured and leveraged properly, is the pathway to progress and economic success."*²

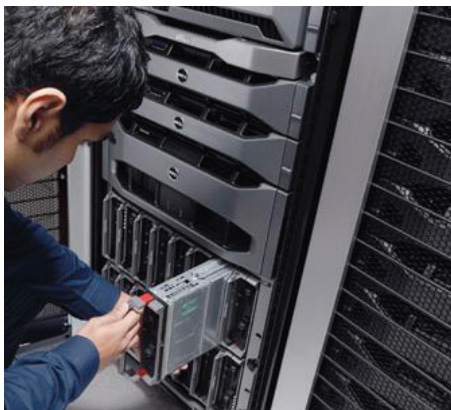
Clearly, the opportunity is ripe for organizations that have the systems in place to take advantage of big data. That's the idea behind the Hadoop solutions delivered together by the partnership of Dell, Cloudera and Intel—three leaders in big data.

The power of partnership

For organizations trying to extract value out of massive amounts of structured and unstructured data, the open source Hadoop data management, analysis, storage and processing system offers compelling benefits. It can store any kind of data from any source, inexpensively and at very large scale, and it can do sophisticated analysis on that data easily and quickly.

¹ "The Digital Universe of Opportunities: Rich Data and the Increasing Value of the Internet of Things," EMC Digital Universe study with research and analysis by IDC. April 2014.

² ["A New Gold Rush Is On. Who Will Strike It Rich?"](#) Blog post by Michael Dell. April 2, 2014.



A powerful partnership

This Dell | Cloudera | Intel combination brings together:

- Consistent direction and proven technology for each part of the Hadoop solution stack
- Expertise that stabilizes and accelerates the maturation of your end-to-end analytics solutions

- Innovation-focused roadmaps to help you integrate, secure and analyze data for new insights
- A commitment to leverage and contribute innovation to the open-source community to help ensure the success of Hadoop as an enterprise data hub

That's all the upside of the story. The other side of this story is one of IT challenges. For all its benefits, Hadoop is not the most seamless and simple solution to design, configure and deploy. To realize the full potential of the platform, you need the expertise and solutions of technology partners who are heavily invested in Hadoop. That's the case with the Dell, Cloudera and Intel partnership.

Together as leaders in big data with a tight engineering partnership, Dell, Cloudera and Intel are uniquely positioned to deliver end-to-end Hadoop solutions that enable your organization to collect and mine growing amounts of data, manage that data, and analyze it to gain valuable business insights and competitive advantages—all in an affordable manner.

A few examples of this close partnership:

- In 2009, Dell worked closely with Intel to custom design a new generation of servers to run Hadoop in some of the world's largest search engines.
- Both Dell and Intel work closely with Cloudera, the leading provider of Hadoop-based software and services. Since 2011, the three companies have built pre-validated reference architectures that accelerate time to value for Hadoop deployments.
- Dell, Cloudera and Intel deliver the experience, innovation and expertise that comes with working with organizations around the world to deliver high-performance Hadoop environments with the latest enterprise-class security features.

Ultimately, this strategic partnership yields robust Hadoop solutions that together deliver compelling business and IT benefits.

Diverse use cases

Hadoop software supports a wide range of use cases for storing, processing and analyzing data. Here's a look at some of the value you can leverage from Hadoop to gain greater results in your journey to realize the full potential of your big data projects and infrastructure.

Cost-effective active archive: Create one massively scalable platform, an enterprise data hub, to store any amount or any types of data, in its original form, for as long as desired or required.

Accelerated connection of datasets:

Collect, manage and transform data to identify emerging opportunities and gather insights to build, protect and grow your organization.

Data warehouse optimization: Optimize your enterprise data warehouse (EDW) with robust security, disciplined governance and built-in data protection.

ETL offload: Move the heavy lifting of data extract-transform-load (ETL) processing out of your EDW and into a cost-effective and flexible Hadoop environment.

Agile data exploration: Run a variety of enterprise workloads—including batch processing, interactive SQL, enterprise search and advanced analytics—to accelerate data insight discovery.

Data science: Apply scientific methods to extend big data analytics well beyond the realm of traditional business intelligence tools and analytic engines.

Converged analytics: Distribute and scale analytics dashboards and tools to meet the needs of executives and analysts.

Realizing the benefits: case studies

The power of the combination of technologies from Dell, Cloudera and Intel is evident in the accomplishments of organizations that are putting solutions to work to capitalize on big data and realize tangible business and IT benefits. Let's look at some of these case studies.

Dell SupportAssist

Using big data to transform IT service and support

The Dell SupportAssist team relies on a Dell | Cloudera | Intel big data analytics solution to transform the service and support experience for more than 25 million Dell customers. Drawing on data collected from the Internet of Things and analyzed in real time, the proactive big data solution enables tech support teams to start troubleshooting system issues before customers even know they are facing issues.

A Dell SupportAssist strategist notes: "With the SupportAssist intelligence engine and the Hadoop technologies behind that, we're able to run analytics at the speed of thought, which is sub-three-second response times."

[Watch the Dell SupportAssist video.](#)



MetaScale

Accelerating big data ROI with Hadoop

MetaScale's clients need cost-effective ways to derive intelligence from enormous data stores while radically reducing costs. To meet this need, the global consulting firm partners with Dell to provide Hadoop big data solutions, training and support. These solutions help the firm's clients speed processing, improve decision support and realize major cost reductions.

In recognizing Dell as a key partner, a MetaScale executive notes: "Powerful yet highly cost-effective hardware, like our big data appliances, which incorporate Dell PowerEdge servers, is what makes the Hadoop cluster work."

[Read the MetaScale case study.](#)

National Center for Supercomputing Applications

Merging big data with high-performance computing to speed innovation

Through its Private Sector Program, the National Center for Supercomputing Applications provides supercomputing capabilities and related services to manufacturing and engineering companies. The NCSA focuses on helping organizations address their compute and big data challenges in ways that make them more competitive.

To carry out this mission, the NCSA works with Dell, Intel and Cloudera to implement Hadoop solutions for managing and analyzing enormous amounts of data with leading-edge computational capabilities. The organization has used four generations of Dell PowerEdge servers, each with Intel processors. With the latest generation of Intel processors, the NCSA realized performance increases of 20 – 30 percent over the previous generation.

[Watch the NCSA video.](#)

San Diego Supercomputer Center

Transforming scientific research with big data analytics

The San Diego Supercomputer Center (SDSC) supplies HPC and big data resources to the national research community. The organization's newest HPC resource is a petascale supercomputer named Comet. Based on Dell PowerEdge servers with Intel® Xeon® processors, Comet is designed to transform advanced scientific computing by expanding access and capacity among traditional as well as non-traditional research domains.

"Dell and Intel came to us with the right processors, and they came to us with a willingness to innovate on the design," says the SDSC's manager of HPC systems. "So we ended up with a large-scale system that serves a huge diversity of research."

[Watch the SDSC video.](#)

Siemens PLM Software

Searching billions of records in seconds

To enable global-brand owners to manage product performance and the customer experience, Siemens PLM Software needed to collect, manage, search and analyze vast amounts of diverse data types, and it sought the right software and hardware infrastructure to support this effort.

To meet this need, the organization worked with Dell and Cloudera to build a software solution on top of the Cloudera® Distribution of Hadoop® (CDH) platform running on a cluster of Intel-based Dell PowerEdge servers. The resulting solution gives Siemens PLM Software customers total product data visibility throughout their entire supply chains. The solution, which scales to support 1 billion new records every month, helps companies detect and pinpoint emerging issues within the supply chain, save millions of dollars, boost productivity, and improve both product performance and the customer experience.

[Read the Siemens PLM Software case study.](#)

Translational Genomics Research Institute

Improving patient outcomes with personalized medicine

To help fight cancer and other diseases, the Translational Genomics Research Institute (TGen) needs extremely scalable, reliable and available HPC nodes to develop personalized medical treatments. With this goal in mind, TGen tuned its system for Genomics I/O demands by scaling its existing Dell HPC cluster to include more servers, storage and networking bandwidth so that researchers can get the IT resources they need faster without having to depend on shared systems.

With the optimized cluster, which incorporates additional Dell PowerEdge servers with Intel® Xeon® processors, researchers can create targeted treatments in less time and help improve outcomes for more patients. "Last month, for the first time ever, our researchers logged 1 million CPU hours on our Dell HPC cluster — significantly increasing the amount and speed of research," notes a TGen vice president. "There's no way we would have been able to handle that volume of computation before, even if we used a university's HPC cluster along with ours."

[Watch the TGen video.](#)

Tulane University

Enabling discovery with high-performance computing and big data

Tulane University partnered with Dell and Intel to create a new supercomputer, called Cypress. The Dell architecture represents a leading-edge design for the convergence of big data analytics and high-performance computing into one platform. The platform is based on Dell PowerEdge servers with Intel® Xeon® processors, a Dell Networking fabric switch and Cloudera software.

"As Tulane was looking at how to build an architecture, Dell became clearly the leading partner for us," notes the university's chief technology officer. "Dell leveraged their relationship with Intel,



who in turn leveraged their relationship with Cloudera.” The resulting Cypress supercomputer—which allows Tulane to do big data analytics using Hadoop in an HPC environment—now helps Tulane accelerate discovery and attract new faculty and funding.

[Watch the Tulane video.](#)

University of Cambridge

Putting cloud and big data to work for scientific and business innovation

The University of Cambridge consistently ranks as one of the world’s top universities. It works with small and midsize businesses in the U.K. to lower barriers of entry to markets and provide them with cutting-edge cloud-based compute power. Access to these HPC resources helps companies compete in tomorrow’s knowledge economy.

To help carry out this mission, the university works closely with Dell on the development of cloud and big data solutions. “We have a long-standing and close technical partnership with Dell, where Dell helps us develop tomorrow’s solutions,” notes the university’s deputy director for research and institutional services. “Dell is a good partner for us because Dell’s direction is the same direction we are moving in.”

[Watch the University of Cambridge video.](#)

Select your path forward

To help your organization capitalize on Hadoop, Dell offers a choice of solutions built on the engineering partnership of Dell, Intel and Cloudera. In addition, Dell, Cloudera and Intel can work with your organization to develop a proof of content in a Dell Customer Solution Center.

Here’s a look at some of these ways to get stated down the path to a Hadoop big data deployment:

Dell QuickStart for Cloudera Hadoop

Dell QuickStart for Cloudera Hadoop offers an easy entry point for your organization to begin managing and analyzing data. It’s an all-in-one system designed to reduce the complexity of deploying, configuring and managing Hadoop systems. QuickStart includes the hardware, software and services you need to deliver a Hadoop cluster that enables your organization to quickly engage in Hadoop testing, development and proof of concept work and is delivered at an affordable price point.

Dell Hadoop Reference Architectures in the Dell | Cloudera | Intel Apache Hadoop Solution

For highly scalable Hadoop deployments, Dell offers optimized reference architectures that provide baseline configurations for Hadoop clusters that leverage Dell PowerEdge servers, Intel® Xeon® processors, Dell networking, and Cloudera Enterprise Distribution for Hadoop software. These configurations have been developed jointly by Dell, Cloudera and Intel, and are based on extensive customer experience with real-world Hadoop production installations. They provide a roadmap for deploying your Hadoop environment and for scaling from a few to hundreds of data nodes.

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

This reference architecture helps your organization lower data transformation costs and build operational efficiencies while laying a robust, cost-effective, secure and scalable foundation for managing data and maturing into advanced data analytics. Jointly designed by Dell, Cloudera, Intel and Syncsort, this tested and validated reference architecture outlines the end-to-end components for a complete extract transform-load (ETL) offload solution.

Dell Customer Solution Centers

Located in key sites around the globe, Dell Customer Solution Centers give you the opportunity to experience Dell solutions and technology in a dedicated, hands-on environment equipped with state-of-the-art labs and teams of solution experts. To date, approximately 750 organizations have used these technical centers to investigate next-generation, scale-out computing technologies, including Hadoop.

[Visit the website](#)

Make your journey with Dell, Cloudera and Intel

Around the world, organizations large and small are seeking ways to uncover and act on opportunities hidden in data, regardless of the type of data and where it’s located, inside or outside the organization. In many cases, this search leads to Hadoop.

The open source Hadoop software platform gives forward-looking organizations the ability to store and analyze data more affordably than ever before. With its power and flexibility, Hadoop offers a natural complement to existing data warehousing infrastructure.

If your organization is on this path, Dell, Cloudera and Intel are your ideal technology partners. Dell has worked with Hadoop since 2008, Cloudera is the leading provider of Hadoop-based software and services and Intel is world leader in computing innovation. This combination of technology leaders delivers the benefits of proven architectures for Hadoop deployments, processors purpose-built for big data, powerful analytics engines, and technologies to enable enterprise-class security in Hadoop environments.

Together, Dell, Cloudera and Intel are ideally positioned to deliver end-to-end Hadoop solutions that enable your organization to manage and mine growing amounts of data to gain valuable insights and competitive advantages.

To learn more, visit Dell.com/Hadoop | Dell.com/BigData

