



Turning enterprise data into business value

Dell increases business insights, improves services and saves \$4 million in one year by augmenting its enterprise data warehouse using ETL offload with Cloudera Hadoop

Informed decisions are better decisions. That's why big data and business intelligence solutions are the panacea for improving efficiency and service. For years, Dell had been using an enterprise data warehouse (EDW) to create reports, but as the EDW grew in size, reports took longer to create. The EDW's limitations became more of an issue when staff needed to analyze unstructured big data from new sources such as social media to gain a 360-degree view of customers.

Although scaling the EDW was the obvious answer, doing so would just be building on an expensive and limited solution. Replacing the EDW wasn't an option either. It provided an important and reliable central data store for global employees, and they knew how to use the built-in analytical tools. Engineers recognized that the main culprit behind slowing report performance was increasing volumes of extract, transform and load (ETL) processes, which ingest data from disparate

sources and convert it into a common format for analysis. So the first step in solving the big data challenge was finding an affordable solution to accelerate ETL processes.

Getting the best of both worlds: Familiar tools and big data

Engineers came up with a solution by building a five-node Cloudera™ Hadoop® data hub — or data reservoir — to manage ETL processes for the EDW. Hadoop can manage any kind of data. Plus, by choosing the Cloudera version of Apache™ Hadoop, engineers gained tools for simplifying adoption, and they could engage Cloudera Professional Services for design guidance. Armando Acosta, Hadoop planning product manager and subject matter expert at Dell, says, "We solved the skills gap and realized faster time to value by using Cloudera. We didn't have to learn Hadoop MapReduce, Java, Pig, Hive and Sqoop. We could even use existing SQL processes so that we didn't have to write new code."

"We have a three-dimensional view of the business because we can ingest and save any kind of data with Hadoop, including information we previously discarded, so we can now build better reports and predictive analytics."

Attila Finta, IT Director of Enterprise Business Intelligence Architecture and Engineering, Dell

Customer profile

Company	Dell
Industry	Technology
Country	United States
Employees	100,000
Website	www.dell.com

Business need

Decision makers needed to create reports faster and build them with more types of information including data feeds from public sources. Any new IT had to be easy and affordable to implement.

Solution

The company deployed an enterprise data hub, also known as a data reservoir, using Cloudera Hadoop to offload extract, transform and load (ETL) data from the existing enterprise data warehouse.

Benefits

- Gains a 360-degree view of customers and operational processes
- Saves \$4 million in one year
- Cuts the cost of a terabyte 50x — by \$49,000
- Makes use of existing IT and skill sets
- Facilitates innovation with big data

Solutions at a glance

- Analytics & Business Intelligence
- Application & Data Integration
- Big Data
- Servers

Gaining faster, 360-degree views and saving millions in the process

After a successful proof of concept, Dell moved its Cloudera hub into production. It now spans 64 Dell PowerEdge R730xd servers with Intel® Xeon® processors. Deepak Gattala, big data architect at Dell, says, "Reporting is faster today because we now distribute ETL workloads across multiple systems, rather than having one server do it all."

Business employees also have a 360-degree view of customer experience and requirements. In addition, marketing employees no longer have to pay third parties to analyze social media posts. Jeff Weidner, marketing data science architect at Dell, says, "In the first year alone, we realized \$4 million in OPEX savings just by bringing in two projects that were initially outsourced."

Making big data feasible by dropping the cost of a terabyte by 50x

Dell now has an affordable big data solution that can support new requirements. "We brought the cost of storing a terabyte of data down to as low as \$1,000 with our Hadoop solution," says Acosta. "To store a terabyte in a traditional database costs up to \$50,000."

Scalability is no longer an issue either. Spike White, senior Linux engineer at Dell, explains, "Because of the work we did during design and testing, it's cake adding extra nodes to our Hadoop cluster. We just drop them in the cabinet, no problem."

Customers drive \$5.6 million in new revenue with the solution

To help customers realize similar benefits, Dell, Intel and Cloudera are on the eighteenth version of the Intel-powered Dell | Cloudera Apache Hadoop Solution Reference Architecture. Forrester studied the stories and results of four companies that implemented the solution. "The customers achieved a TCO of 97 percent, a return on investment within six months and a net-present value of \$5.6 million over three years by using the Dell | Cloudera Apache Hadoop Solution," says Acosta. "And from there, once customers get comfortable with the technology, they start building new use cases, like we did at Dell, to transform their business."

Improving the bottom line and facilitating new offerings

The increased insight as well as the ability to add new data sources for analysis boosts business efficiency and agility, so Dell is better positioned to reach its goals. Attila Finta, IT director of enterprise business intelligence architecture and engineering at Dell, says, "We have a three-dimensional view of the business because we can ingest and save any kind of data with Hadoop, including information we previously discarded, so we can now build better reports and predictive analytics." For example, engineers can now store and analyze log reports from supported systems to identify potential issues and remedy them before they create an outage. Dell is also testing the use of

Products & Services

Hardware

Dell | Cloudera® ETL Offload Reference Architecture

Dell PowerEdge R730xd rack servers with Intel® Xeon® processors

Partner

Cloudera Hadoop

Cloudera Professional Services

Apache Spark™, a tool included in the Cloudera distribution of Hadoop, to analyze live data streams in near-real time to create new insights and product offerings for use cases that include the Internet of Things.

Intel Inside®. Powerful Productivity Outside.

View all Dell case studies at Dell.com/CustomerStories

