



LEVERAGING DATA ANALYTICS TO GAIN COMPETITIVE ADVANTAGE IN YOUR INDUSTRY

Unlock the value of your data with analytics
solutions from Dell EMC

ABSTRACT

To unlock the value of their data, organizations around the world are turning to data analytics, data lake for analytics, and the Apache™ Hadoop® platform for data collection, management, and analysis. In this paper, we highlight examples of these data solutions in a wide range of industry-specific and cross-industry use cases. These use cases are based on the collective experiences of Dell EMC and our partners Intel, Cloudera, and Hortonworks.

March 2017

The information in this publication is provided “as is.” Dell Inc. makes no representations or warranties of any kind with respect to the information in this publication, and specifically disclaims implied warranties of merchantability or fitness for a particular purpose.

Use, copying, and distribution of any software described in this publication requires an applicable software license.

Copyright © 2017 Dell Inc. or its subsidiaries. All Rights Reserved. Dell, EMC, and other trademarks are trademarks of Dell Inc. or its subsidiaries. Other trademarks may be the property of their respective owners. Published in the USA 3/17, White Paper.

Dell EMC believes the information in this document is accurate as of its publication date. The information is subject to change without notice.

TABLE OF CONTENTS

- THE BUSINESS CASE FOR A DATA ANALYTICS SOLUTION4**
 - Hadoop.....4
 - Data lake for analytics4
 - Turnkey analytics solutions4
- ADDRESSING COMMON CROSS-INDUSTRY DATA CHALLENGES5**
 - Data warehouse optimization5
 - Data lake5
 - Active archive5
 - Log aggregation5
 - Agile data exploration.....5
 - Self-service analytics.....5
- INDUSTRY-SPECIFIC USE CASES6**
 - Financial services.....6
 - Government6
 - Healthcare.....7
 - Manufacturing.....7
 - Media8
 - Oil and gas exploration and energy utilities8
 - Retail9
 - Technology.....10
 - Telecommunications10
 - Universities and research institutions11
- BROAD BENEFITS 12**
- MAKE YOUR DATA ANALYTICS JOURNEY WITH DELL EMC AND INTEL 12**
- LET’S GET STARTED..... 12**

THE BUSINESS CASE FOR A DATA ANALYTICS SOLUTION

Across all industries, segments, and geographies, data is being created, captured, and consumed at unprecedented rates. With this data explosion comes the need for new, highly scalable platforms for data collection, management, and analysis. The ultimate goal is to unlock the business value that is inherent in data, and capitalize on the digital transformation that is now under way.

For many organizations, the search for a data analytics solution leads to the Apache Hadoop platform, data lake for analytics, and turnkey analytics solutions, such as the Dell EMC Analytic Insights Module. Let's look at each of these enablers of data analytics solutions.

HADOOP

The open source Hadoop platform emerged through multiple efforts among the world's largest Internet companies, which needed a solution to capture and analyze the massive amounts of data they generate. Unlike earlier platforms, Hadoop can store any kind of data in its native format—structured, unstructured, or semi-structured—and be used to perform a wide variety of analyses and transformations on that data.

Hadoop enables organizations to store and analyze petabytes of data. As the amount of data in a cluster grows, an organization can add new servers with local storage and continue expansion. Both robust and reliable, Hadoop handles hardware and system failures with minimal interruption for data analyses. Better still, Hadoop runs on clusters of industry-standard servers. In a Hadoop environment, each server has local compute and storage resources, and each has the flexibility to be configured with an optimal mix of CPU, memory, and storage capacity to meet the application requirements.

Ultimately, Hadoop makes it possible to conduct the types of analysis that would be impractical using other database or data warehousing technologies available today. Along the way, Hadoop can help organizations reduce operational costs and extract more business value from data—provided they find the right use cases for the platform.

DATA LAKE FOR ANALYTICS

Organizations are beginning to understand the value of the data lake. The data lake is a single shared system to capture, store, analyze, protect, and manage data. With all data in one central repository, things happen faster because there is no need to move and copy data across silos. A data lake gives you the capability of in-place analytics, which speeds up time to insights, simplifies the IT infrastructure, and reduces costs.

A data lake is based on the Isilon scale-out network-attached storage (NAS) platform. Isilon enables you to store, manage, protect, and analyze your semi-structured and unstructured data with a powerful platform that stays simple, no matter how large your data environment is. With Dell EMC Isilon scale-out NAS, you have massive room for growth—to more than 68 petabytes of capacity per cluster.

Dell EMC Isilon is the first and only scale-out NAS platform with native Hadoop Distributed File System (HDFS) support. When you use Hadoop with Dell EMC Isilon NAS, there is no need for data ingestion, copying, or migration. You can run data analytics in place, without moving data to a dedicated Hadoop infrastructure. The Isilon platform is also an ideal foundation for a data lake that requires the ability to scale out while supporting traditional and next-generation applications and workloads.

TURNKEY ANALYTICS SOLUTIONS

Another option for capitalizing on data is to deploy a turnkey analytics solution, such as the Analytic Insights Module from Dell EMC. This quick-to-deploy solution delivers a self-service analytics experience for rapidly transforming data into actionable insights with high business value, eliminating the months it could take to build your own system. The module provides capabilities you need to gather, analyze, and act on data insights to monetize new digital business opportunities.

Delivered on Native Hybrid Cloud, Analytic Insights Module is engineered to combine self-service data analytics with cloud native application development into a single cloud platform. It delivers flexible, deep data awareness, features for fast time-to-value, and an integrated data lake for analytics from Dell EMC. You gain a single view of all your discovered, indexed and curated data sets, both within the data lake and selected external sources.

Analytic Insights Module provides a role-based, self-service experience for data analysts to focus on rapidly generating insights with high business value while IT teams work to ensure corporate governance by applying policies to the environment. Analytic Insights Module makes it easy to share data and analytics services with the application development framework for collaborating with developers on data-driven applications, services, and business processes.

ADDRESSING COMMON CROSS-INDUSTRY DATA CHALLENGES

While some solutions for data collection, management, and analysis are tied very closely to the needs of specific industries, others span all industries. Either way, many of these solutions incorporate the open source Apache Hadoop platform, which addresses key challenges associated with storing, managing, and processing large amounts of data in diverse formats.

Let's look at some common cross-industry use cases for data solutions.

DATA WAREHOUSE OPTIMIZATION

The extract-transfer-load (ETL) process can create bottlenecks in enterprise data warehouses (EDWs). A few heavy jobs can bog down an enterprise data warehouse, and more processing means less query capacity. This processing work can be offloaded to a Hadoop environment to reduce CPU utilization for heavy jobs and to accelerate complex ETL processes.

Hadoop is designed for extreme parallel data processing. When used in conjunction with a tool like Syncsort, Hadoop can help you speed up ETL processes while reducing costs in comparison to running ETL jobs in a traditional data warehouse.

In addition, Hadoop can serve as an efficient staging and ETL source to complement your existing EDW. Using Hadoop as an enterprise data hub (EDH) to complement your EDW can drive significant cost savings and other benefits. The goal here isn't to replace your EDW. Rather it is to move certain data, workloads, and processes from your existing systems into Hadoop to gain new capabilities and cost economies.

DATA LAKE

The fundamental concept of a data lake is to have a single store for all data. In an enterprise, data ranges from raw data from various sources, like systems, operations, and documents, to transformed data that is used for various tasks including reporting, visualization, analytics, and machine learning. Typically, the data lake consolidates semi-structured and unstructured data, and enables the organization to analyze it all holistically.

ACTIVE ARCHIVE

When used as the data-landing zone for ETL offload, Hadoop stores and manages data in its native format, creating the foundation for a natural archive repository. Hadoop can serve as a single environment for managing all of your data, in any format, at any volume, for as long as you like. While providing cost-effective data archiving, a Hadoop environment can enable broad organizational access to varied data sets for ad-hoc analysis.

LOG AGGREGATION

Aggregating and storing logs in a Hadoop environment enables real-time and in-depth analysis, and to an extreme scale. These capabilities make Hadoop an ideal platform for rapidly growing log aggregation. Hadoop excels at high-speed ingestion and is an excellent place to both store and aggregate log data. Both batch and real-time analytics are possible.

AGILE DATA EXPLORATION

Hadoop can serve as a staging area that allows users to capture and store new datasets or datasets that have not yet been placed in an enterprise data warehouse. With technologies such as Hive, Impala, Search, and Spark, the platform extends the data access funnel to all users in the organization. Users can combine, compose, and explore the data to gain new insights.

SELF-SERVICE ANALYTICS

Businesses recognize the power of data-driven applications to impact customer behavior and provide insights into business opportunities, yet many organizations struggle with how to extract business value from their data. Turnkey analytics solutions, such as the Dell EMC Analytic Insights Module, enable organizations to quickly deploy a solution to aggregate and analyze data to gain valuable business insights. The Analytic Insights Module delivers all of the software, hardware, and services necessary to stand up an environment for both self-service data analytics and cloud native application development in a matter of days.

INDUSTRY-SPECIFIC USE CASES

Let's turn to a sampling of specific applications of the Hadoop platform, data lake for analytics, and turnkey analytics solutions. These use cases, which span a wide range of industries, are based on the experiences of Dell EMC, which has worked on hundreds of Hadoop deployments, along with our partners Intel, Cloudera, and Hortonworks.

FINANCIAL SERVICES

Unlocking the value of data

To thrive in uncertain and fluctuating markets, upsell customers, combat fraud, maintain compliance with regulations, and meet other critical goals, financial services firms need to unlock the value of data. With goals like these in mind, firms are turning to solutions that consolidate data traditionally managed at the departmental level to create EDWs and enable self-service analytics.

Use cases in financial service

- Fraud prevention in credits and payments
- Risk modeling in investment banking
- Enhanced customer experiences for building brand loyalty
- Cross-selling and upselling in retail banking
- Introduction of new financial products
- Insurance policy personalization
- Mortgage lending portfolio valuation
- Streamlined compliance with evolving financial regulations

Solution example

- The Dell EMC Analytic Insights Module enables financial services firms to analyze data faster and move from discovery to actionable insights in the shortest possible time. It provides a data analytics experience that delivers the agility, scalability, and processing power to drive better business decisions.
- The Dell EMC Isilon along with Hadoop-based data platforms helps financial services firms analyze risk exposure more holistically comply with regulatory mandates, and perform enterprise-level analytics without jeopardizing regulatory or compliance mandates.

GOVERNMENT

Empowering the information-driven agency

Government agencies need to leverage data to improve productivity and derive new insights while managing risk and costs. This is a challenging proposition for many public sector entities because budgets are limited and conventional data management architectures can't meet the technical requirements necessary to analyze diverse and petabyte-scale datasets.

Use cases in government

- Data and application consolidation
- Security intelligence and fraud detection
- Data fusion and analytics for real-time and archival data

Solution examples

- Dell EMC Analytic Insights Module is an ideal solution for agencies seeking to extract value from data in an efficient manner. It combines self-service data analytics and cloud-native application development into a single cloud platform that enables organizations to rapidly transform data into actionable insights.
- A data lake based on Dell EMC Isilon provides a secure and cost-efficient storage for a Hadoop-based enterprise data hub that allows public sector entities to store and analyze petabytes of data in various formats and from various sources, while enabling the confidence that comes with centralized oversight and security.

HEALTHCARE

Improving quality and affordability

In conventional IT environments, clinical, operational, and financial data are managed in data silos. Meanwhile, with the movement from paper-based to electronic health records, and with the increase in usage of machines and medical devices that produce steady streams of data, the volume of data that healthcare institutions capture and analyze has skyrocketed, while the variety of that data has grown.

To overcome these challenges, healthcare providers need a fast, easy, and cost-effective solution to integrate multiple, large data sets with agile, iterative processes to uncover trends, patterns, relationships, correlations, and discoveries that can truly impact integrated patient care. At the same time, clinicians and researchers need a compliant and secure workbench that protects the security and privacy of patient data.

Use cases in healthcare

- Quality of care optimization to improve outcomes and the patient experience
- Clinical quality and cost analysis
- Genome processing and DNA population health management
- Detection of fraud and suspicious transactions
- Clinical and researcher workbench
- Enhanced compliance with end-to-end data governance

Solution examples

- The Dell EMC Analytic Insights Module provides enhanced tools to integrate multiple, large data sets with processes for more collaborative patient care delivery and clinical research. It enables organizations to gather data across the healthcare spectrum into a single organized resource that provides self-service access to all relevant data sets, and includes features that help ensure that data is secure and compliant.
- Dell EMC Isilon enables healthcare organizations to process and manage an ever-larger influx of data in a secure and cost-effective manner to improve quality and affordability. They can leverage the Isilon platform with their custom analytics applications to bring together large volumes of detailed data from diverse sources, in a variety of formats, and consolidate it into a single flexible and scalable system for long-term storage and analysis.

MANUFACTURING

Generating product and process insights

Manufacturers collect an enormous amount of data pertaining to the production of product components, the post-production performance of products, and manufacturing and supply chain processes. Today, in the era of the Internet of Things, manufacturers' data management challenges are growing in scope as products continually generate data related to their performance, functionality, and quality. The challenge is not only to capture all this data, but to manage and analyze it to generate product and process insights. Hadoop provides an ideal solution to these challenges.

Use cases in manufacturing

- Proactive quality assurance
- Analysis of demand for new products and services
- Product research guided by machine-generated data
- Detection of supply chain issues
- Identification of cross-sell and upsell opportunities
- Identification of opportunities to develop new value-added services

Solution examples

- A large industrial manufacturer deployed a solution based on the Hadoop and Dell EMC Isilon storage for predictive analytics of component failure. The solution continuously ingests and analyzes component sensor data from the field to enable proactive maintenance and help the company improve the uptime of its remote equipment.

MEDIA

Enabling targeted advertising

To help customers maximize the impact of their advertising spend, cable TV and digital media companies need to extract insights from diverse data sources and huge data sets. This data is often held in disparate silos, making aggregation and analysis difficult and time-consuming.

Use cases in media

- Aggregation of data from database silos for analysis
- Accelerated responsiveness to customer requests
- Streamlined decision-making with simplified analytics results
- Increased storage scalability to enable new market opportunities

Solution examples

- Data lake for analytics based on Dell EMC Isilon scale-out storage and Hadoop enable media companies to streamline data aggregation from multiple sources and deliver analytics insights in less time to help their customers target advertising for the highest impact. By providing direct data integration with Hadoop, Isilon allows companies to eliminate time-consuming overnight data importation processes.
- Dell EMC Analytic Insights Module empowers data analytics teams at media companies with a fast and efficient self-service user experience, while giving them the freedom to use their analytics and visualization tools of choice. It allows them to collaborate and share their analytics across the organization for reuse, avoiding the “orphaned analytics” phenomenon.

OIL AND GAS EXPLORATION AND ENERGY UTILITIES

Leveraging research and operational data

Oil and gas companies need to accelerate discovery and extend the life of existing fields by improving recovery efficiency. This requires a huge amount of data to improve well production and minimize non-productive time (NPT). For electric utilities, the flood of data from smart meters and smart grid technologies holds the key to improving distribution and pricing models based on demand. The challenge is that legacy data warehouse systems are ill-equipped to handle big data. This results in multiple, small data silos that can require months to produce useful intelligence.

Use cases in oil and gas exploration

- Horizontal drilling enablement and optimization
- Seismic data processing
- Predicting where best to drill next
- Determining which leases to sell and which sections to acquire
- Predictive maintenance to improve production uptime
- Enhanced commodity pricing to increase revenue

Solution example

With its parallel processing capabilities and extreme scalability, Hadoop, running on Dell EMC servers, can help oil and gas exploration companies make more efficient use of massive volumes of data. This is one of the keys to finding and extracting energy resources in a cost-effective manner.

Leveraging data analytics in energy utilities

Energy utilities now capture and store large amounts of data from advanced metering infrastructure, smart appliances, interactive user applications, and sensors. They also make use of historical home energy data, weather data, and social media data, along with disparate other types of information.

The data collection part is relatively easy. The real challenge is to consolidate and analyze this diverse range of data to answer granular questions about energy usage, predict fluctuations in demand, identify conservation opportunities, and achieve various other goals of a progressive utility. Analytics solutions are ideally suited for these challenges.

Use cases in energy utilities

- Generate energy reports for individual consumers
- Compare energy usages among different users
- Predict sudden and temporary shifts in demand
- Gain an actionable 360-degree customer view
- Prevent fraud and energy theft

Solution example

The Dell EMC Analytic Insights Module arms utilities with a solution to pool, enrich, analyze, and act upon their data in real time. It integrates data from diverse utility operations and customer information systems into a single organized resource that data analysts can leverage to shrink the time required to produce actionable insights.

RETAIL

Cashing in on data

To compete in the age of the Internet storefront, large retailers need scalable data management systems that integrate online and offline data so they can better understand their customers and improve the efficiency of their operations. In particular, retailers now need to connect and process data in many formats from disparate systems and sources, including the social media sites that consumers interact with.

Use cases in retail

- Enablement of a 360-degree customer view
- Generation of personalized offers

- Enablement of first in-basket analysis
- Merchandising and supply chain analysis
- Isolation of products and mixes indicative of larger baskets
- Event correlation to store traffic
- Single customer identity across all operational systems
- Loyalty program management
- Customer churn analysis

Solution examples

- A global online retailer deployed an analytics solution based on a Hadoop cluster and Dell EMC Isilon storage nodes to harness market intelligence. The solution gives the company the ability to correlate retail data about customer purchases, marketing campaign results, and online browsing behavior, and to generate individualized customer profiles.
- A retail giant needed much more cost-effective ways to derive intelligence from enormous data stores while radically reducing costs. The company found its solution in a Hadoop environment running on hundreds of Dell EMC server nodes. While driving savings in terms of both time and money, the solution supports diverse use cases, including a recommendation engine, ETL mainframe replacement, supply chain analysis, a central data repository, price setting, and logistics planning.

TECHNOLOGY

Gaining greater value from data

Hadoop was inspired by papers published by Google and driven by the need to store, process, and analyze massive amounts of data. Today, technology companies are challenged to gain ever-greater value from the data deluge they face on a daily basis. With this goal in mind, organizations continue to come up with new ways to leverage the Hadoop platform.

Use cases in technology

- Improving search quality through analysis of search results
- Geospatial, image, and video processing
- Cyber security and fraud detection
- IP-TV subscriber recommendation engine
- House-holding and matching data across social networking and computing applications

Solution example

A software developer needed to gain a better understanding of how its customers were using its software. The company found its answer in a highly scalable, high performance data analysis platform built on Hadoop and Dell EMC hardware. The solution helped the company improve the design of its software based on customer use while identifying new revenue-generating opportunities.

TELECOMMUNICATIONS

Gaining business-driven insights

Communications service providers are some of the biggest collectors of data today. Analysis of that data is one of the keys to identifying and understanding network capacity trends, reducing infrastructure costs, increasing average-revenue-per-user, preventing churn, expanding service offerings, and improving the customer experience.

In traditional environments, customer information is captured in different systems. This fragmentation makes it difficult for retailers to analyze data in a holistic manner to gain a consolidated view of the customer. Silos of data trapped in disparate systems cannot be easily accessed or analyzed to identify market opportunities, target new service offerings, and meet other business goals.

Use cases in telecommunications

- Enablement of an actionable 360-degree customer view
- Identification of affinity strengths between services and products
- Identification of new market opportunities for targeted services
- Collaborative planning and forecasting
- Network capacity planning, trending, and management
- Prioritization of network infrastructure based on demand
- Customer churn prevention
- Identification of subscriber quality of service issues
- Enablement of emerging IoT technology
- Research and development guided by machine-generated data
- Bandwidth-hog identification

Solution examples

- The Dell EMC Analytic Insights Module enables service providers to integrate information from across telecommunications value chain into a single organized resource. It helps data analysts shrink the time dedicated to discovery and insights from months to days using their preferred analytics tools while giving business executives powerful ammunition to monetize new opportunities.
- Hadoop allows service providers to combine information from different systems quickly and efficiently to enable large-scale data processing and analysis.

UNIVERSITIES AND RESEARCH INSTITUTIONS

Marrying data analytics and high-performance computing

Hadoop, running on Dell EMC servers, and high-performance computing (HPC) have changed the game for large-scale data analysis. Yet the two technologies are built for different purposes. HPC grew out of a need for large-scale computational speed and high performance for scientific research. Hadoop was born from the need to process large volumes of data in the Web 2.0 space. The challenge for universities is to find complementary ways to leverage Hadoop and HPC resources to accelerate time to insight on scientific investigations.

Use cases in universities and research institutions

- Data analytics in HPC clusters
- Analyzing huge genomic datasets
- Enabling a 360-degree view of research subjects
- Data explorations to identify unpredicted correlations

Solution example

A U.S. university deployed Hadoop, running on Dell EMC, in conjunction with a high-performance file system to enable big data analytics in an HPC environment. The solution allows the university's researchers to move easily between data analytics and HPC capabilities to accelerate throughput for a wide range of data-intensive workloads.

BROAD BENEFITS

Regardless of the use case, today's solutions for data collection, management, and analysis offer compelling benefits for organizations that want to extract value out of huge amounts of structured, unstructured, and semi-structured data. With the right solutions in place, your organization can use and store any kind of data, from any source, in its native format, and perform a wide variety of analyses and transformations on that data.

In more specific terms, Dell EMC data analytics solutions enable your organization to:

- Use and store any data in its native format without forcing transformation
- Control costs with open source software that runs on industry-standard hardware
- Work with industry leaders to enable a fully supported solution
- Contain the rising costs and challenges of data management
- Leverage a global user and developer community that spans industries
- Scale up quickly to meet your evolving data storage and processing needs

MAKE YOUR DATA ANALYTICS JOURNEY WITH DELL EMC AND INTEL

Regardless of where you are in your data analytics journey, you can look to Dell EMC for expertise and the benefits of the growing portfolio of Dell EMC data analytics solutions accelerated by Intel. These solutions provide end-to-end scalable infrastructure, leveraging open source technologies. They allow you to simultaneously store and process large datasets in a distributed environment for data mining and analysis on all types of data—structured, semi-structured, and unstructured—and to do it all in an affordable manner.

When you partner with Dell EMC and Intel for your data analytics exploration and deployment, you have the confidence that comes with organizations that have worked with data analytics, data lakes, and the Hadoop platform for years. Both Dell EMC and Intel maintain a close working relationship with leaders in Hadoop-based software and services, including Cloudera and Hortonworks.

You can also look to Dell EMC for other components of a complete data analytics solution, including unique software products for data integration including Dell EMC Boomi. It enables you to synchronize data between mission-critical applications—on-premises and in the cloud—without the costs of procuring appliances, maintaining software, or generating custom codes.

Ultimately, Dell EMC, together with partners Intel, Cloudera, and Hortonworks, has what it takes to help you gain hands-on experience with data analytics through a proof of concept. You can then take your solution to a full production environment guided by proven reference architectures, enabled by validated solutions, and supported by the Dell EMC Professional Services organization.

LET'S GET STARTED

In your explorations of data analytics opportunities, you can leverage the resources of a [Dell EMC Customer Solution Center](#). Located in key sites around the globe, these technical centers give you the opportunity to experience Dell EMC solutions and technologies in a dedicated, hands-on environment equipped with state-of-the-art labs and teams of solution experts.

To learn more, visit DellEMC.com/BigData