

A Forrester Total Economic Impact™
Study Commissioned By Dell EMC
May 2018, Updated August 2018

The Total Economic Impact™ Of The Dell EMC Ready Solutions For Hadoop

Cost Savings And Business Benefits
Enabled By Dell EMC's Ready Solutions

Table Of Contents

Executive Summary	1
Key Findings	1
TEI Framework And Methodology	3
Customer Journey	4
Interviewed Organizations	4
Key Drivers	4
Solution Requirements	5
Key Results	5
Composite Organization	6
Analysis Of Benefits	7
Legacy Hardware Savings	7
Profit From New Hadoop-Enabled Business	8
Fraud Reduction	10
Administrative Cost Savings	11
Improved Data Scientist Productivity	12
Unquantified Benefits	13
Flexibility	14
Analysis Of Costs	15
Hardware Costs	15
Software License Costs	16
Implementation Costs	16
Hadoop Operational Costs	17
Financial Summary	18
Dell EMC Ready Solutions: Overview	19
Appendix A: Total Economic Impact	21
Appendix B: Endnotes	22

Project Director:
Benjamin Brown
Sean McCormick

ABOUT FORRESTER CONSULTING

Forrester Consulting provides independent and objective research-based consulting to help leaders succeed in their organizations. Ranging in scope from a short strategy session to custom projects, Forrester's Consulting services connect you directly with research analysts who apply expert insight to your specific business challenges. For more information, visit forrester.com/consulting.

© 2018, Forrester Research, Inc. All rights reserved. Unauthorized reproduction is strictly prohibited. Information is based on best available resources. Opinions reflect judgment at the time and are subject to change. Forrester®, Technographics®, Forrester Wave, RoleView, TechRadar, and Total Economic Impact are trademarks of Forrester Research, Inc. All other trademarks are the property of their respective companies. For additional information, go to forrester.com.

Executive Summary

Key Benefits

(Three-year present value)



Legacy hardware savings for data storage and offload:

\$15 million



Profit from new Hadoop-enabled business:

\$4,090,158



Reduced fraud:

\$1,386,560



Improved data scientist productivity:

30%

Dell EMC provides integrated Ready Solutions that help customers quickly deploy Hadoop solutions and realize their benefits, such as running advanced data analytics, reducing costs, and accelerating new revenue streams. Dell EMC commissioned Forrester Consulting to conduct a Total Economic Impact™ (TEI) study and examine the potential return on investment (ROI) enterprises may realize by deploying the Ready Solutions for Hadoop. The purpose of this study is to provide readers with a framework to evaluate the potential financial impact of Dell EMC's Ready Solutions for Hadoop on their organizations.

To better understand the benefits, costs, and risks associated with this investment, Forrester interviewed six customers with years of experience using the Ready Solutions. Since organizations' requirements vary, Dell EMC offers two different Hadoop distributions: Cloudera Enterprise and Hortonworks Data Platform. This study focuses on the Ready Solutions for Cloudera Hadoop, powered by Intel.

Traditionally, organizations implemented enterprise data warehouse platforms as a centerpiece of their business intelligence. However, these legacy solutions are expensive and may now lead to a competitive disadvantage. Interviewed organizations were concerned they were falling behind competition without the proper big data capabilities to derive business insights and act. With the Dell EMC Ready Solutions for Cloudera Hadoop, organizations created cost-effective data repositories, enabled new business capabilities, and improved data scientist efficiency by streamlining the reporting process. Interviewees explained that analytics ran faster with the Ready Solutions, reducing delays in deriving insights and improving overall productivity for data scientists. Furthermore, organizations utilized Hadoop for cost savings measures such as fraud prevention, reducing fraud by up to 20%. Overall, with the Dell EMC Ready Solutions for Cloudera Hadoop, customers streamlined deployment to increase time-to-value.

Key Findings

Quantified benefits. The interviewed organizations experienced the following risk-adjusted present value (PV) benefits:

- › **Data offload and storage saved \$15 million in hardware costs.** Organizations reduced dependency on traditional data warehouses by building out data lakes, repositories of structured and unstructured data, and running extract, transform, and load (ETL offload) in Hadoop. Storage costs were nearly one-twelfth of the legacy platforms, which cost at least \$21,660 per terabyte.
- › **New and improved products and services leveraging Hadoop drove \$4.1 million in additional profit.** Deploying Hadoop quickly and getting to market faster accelerated the benefits of the new capabilities and insights provided by Hadoop. These new powerful analytics enabled businesses to improve their product offerings and build out new services and products.
- › **Reduced fraud enabled \$1.4 million in savings.** Organizations leveraged technologies such as Hadoop and Spark to create more accurate machine models to detect fraud in real time. By directing the analytical power of Hadoop toward monitoring transactions, organizations could reduce fraud by up to 20%.



ROI
218%



Benefits PV
\$21.8 million



NPV
\$14.9 million



Payback
<6 months

› **Improved system administration efficiency saved up to \$540,000 per year.** The Ready Solutions simplified operations and maintenance, and administrators could support much larger sets of data in the same or less time than with traditional data warehouses.

› **Data scientists improved productivity by 30% for analytics and reporting, driving up to \$337,000 in annual savings.** Running queries and analytics on large sets of data was much faster with the Ready Solutions, when previously, these same reports might take days to run.

Unquantified benefits. The interviewed organizations experienced the following benefits, which are not quantified for this study:

› **Tightened data security, identifying and remediating almost 100 million attacks and achieving PCI compliance.** By using Hadoop to create machine learning models, the organizations could more accurately remediate threats to enable security analysts to be more efficient, reducing the risk and costs of breaches.

› **Optimized system capacity by 30% using Hadoop analytics.** With Hadoop, clusters can identify usage patterns as well as the types of queries being run. This helps organizations tune their processes, reduce response times, and free up capacity, leading to potential cost savings.

› **Improved ability to meet service-level agreements (SLAs) by 8%.** Some organizations had seasonal spikes in demand, and many data scientists would have to wait to submit workloads. With Hadoop, organizations could better manage demand, reducing the overall failure rate from 10% to 2%.

Costs. The interviewed organizations experienced the following risk-adjusted present value costs, attributed to a single composite organization called *The Representative Organization*, described later in this study:

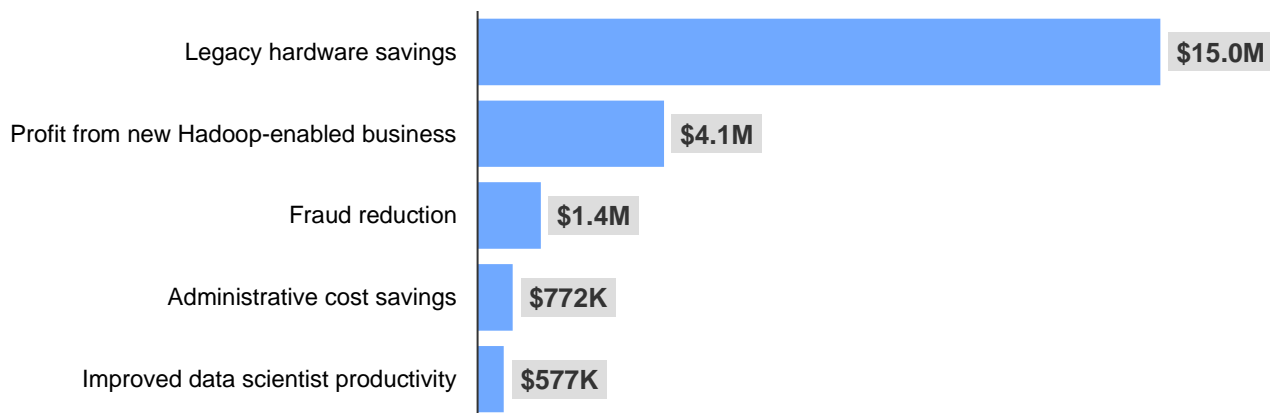
› **Hardware costs of \$1,608 per terabyte (including triple redundancy) to build out the Hadoop platform.** The composite organization spent a total of \$4.4 million for its 150-node deployment over three years.

› **Cloudera software license costs of \$1,400 per node.** The composite organization accumulated three-year licensing costs of \$296,471.

› **Implementation costs equivalent to four full-time equivalents (FTEs) for six months.** Labor plus professional services from Dell EMC equated to \$369,321 in costs.

› **System administration costs ranging from \$646,800 to \$950,400 per year.** The composite organization required up to two administrators and four data engineers to support the Hadoop environment.

Benefits (Three-Year)



The TEI methodology helps companies demonstrate, justify, and realize the tangible value of IT initiatives to both senior management and other key business stakeholders.

TEI Framework And Methodology

From the information provided in the interviews, Forrester has constructed a Total Economic Impact™ (TEI) framework for those organizations considering implementing the Dell EMC Ready Solutions for Cloudera Hadoop.

The objective of the framework is to identify the cost, benefit, flexibility, and risk factors that affect the investment decision. Forrester took a multistep approach to evaluate the impact that the Dell EMC Ready Solutions for Cloudera Hadoop can have on an organization:



DUE DILIGENCE

Interviewed Dell EMC stakeholders and Forrester analysts to gather data relative to the Dell EMC Ready Solutions for Cloudera Hadoop.



CUSTOMER INTERVIEWS

Interviewed six organizations using the Dell EMC Ready Solutions for Cloudera Hadoop to obtain data with respect to costs, benefits, and risks.



COMPOSITE ORGANIZATION

Designed a composite organization based on characteristics of the interviewed organizations.



FINANCIAL MODEL FRAMEWORK

Constructed a financial model representative of the interviews using the TEI methodology and risk-adjusted the financial model based on issues and concerns of the interviewed organizations.



CASE STUDY

Employed four fundamental elements of TEI in modeling the Dell EMC Ready Solutions for Cloudera Hadoop's impact: benefits, costs, flexibility, and risks. Given the increasing sophistication that enterprises have regarding ROI analyses related to IT investments, Forrester's TEI methodology serves to provide a complete picture of the total economic impact of purchase decisions. Please see Appendix A for additional information on the TEI methodology.

DISCLOSURES

Readers should be aware of the following:

This study is commissioned by Dell EMC and delivered by Forrester Consulting. It is not meant to be used as a competitive analysis.

Forrester makes no assumptions as to the potential ROI that other organizations will receive. Forrester strongly advises that readers use their own estimates within the framework provided in the report to determine the appropriateness of an investment in the Dell EMC Ready Solutions for Hadoop.

Dell EMC reviewed and provided feedback to Forrester, but Forrester maintains editorial control over the study and its findings and does not accept changes to the study that contradict Forrester's findings or obscure the meaning of the study.

Dell EMC provided the customer names for the interviews but did not participate in the interviews.

Customer Journey

BEFORE AND AFTER THE READY SOLUTIONS INVESTMENT

Interviewed Organizations

For this study, Forrester conducted interviews with six customers using Dell EMC's Ready Solutions for Cloudera Hadoop. Interviewed customers include the following:

INDUSTRY	REGION	SIZE	INTERVIEWEE	INFRASTRUCTURE
Marketing analytics	Global	\$8M revenue 50 employees	VP, database and technology	8 nodes 400 terabytes
Information technology	USA	\$20M revenue	Chief architect	40 nodes 150 terabytes
Retail	Global	\$25B+ revenue 100K+ employees	Architect	640 nodes 5 petabytes
Digital media services	USA	100 employees	Principal architect	65 nodes 1.8 petabytes
Information technology	Global	\$25B+ revenue 100K+ employees	Business intelligence architect	120 nodes 2 petabytes
Financial services	Global	\$1 to \$5B revenue 10K to 20K employees	VP, big data	1,000+ nodes 20+ petabytes

Key Drivers

Interviewed customers noted a variety of business challenges or goals that led to their investment in the Dell EMC Ready Solutions for Hadoop.

- › **Data storage in traditional data warehouses was expensive.** All interviewees took advantage of significant cost savings for both hardware and administration by moving a portion of their data to the Hadoop data lake, a data repository with both structured and unstructured data, via ETL offload.
- › **Running analytical queries was excessively slow.** Workloads took anywhere from two to eight days to process, wasting productivity and hampering data scientists' analysis capabilities.
- › **Organizations lacked the capability to leverage big data for business insights.** These organizations felt they were falling behind their competitors without an effective platform to store and analyze data, risking their ability to meet their customers' needs. They desired to both improve their current offerings (improving features, functionality, speed, and availability) and build new businesses leveraging the capabilities of Hadoop, Spark, and Spark Streaming.
 - One retail organization primarily leveraged Hadoop to learn more about its customers to better serve them.
 - A marketing analytics organization leveraged Hadoop to launch a new business that would not otherwise have been feasible, driving millions in additional annual revenue.
 - A manufacturing execution systems company utilized Hadoop to track product quality through root-cause analysis.

"Hadoop was an excellent solution for us because you could start small and as your data grows, you can get bigger."

*Principal architect,
digital media services*



Solution Requirements

Organizations started their big data journeys by downloading a free distribution of Hadoop but didn't have the expertise internally to determine how to deploy and use it effectively. Hiring Hadoop experts also proved difficult, as resources in this space were scarce. Interviewed organizations knew they needed:

- › A Hadoop cluster to support big data analytics and machine learning.
- › A partner that could help deploy Hadoop with best practices to ensure smooth deployment and management.
- › The ability to quickly and easily scale up to multiple petabytes of both structured and unstructured data.
- › An optimized infrastructure architecture for performance and advanced analytics based on its desired use cases.

The interviewed organizations selected the Ready Solutions due to its relative ease and speed of deployment, its reference architecture that is optimized for performance, and its ability to easily scale to meet ever-increasing data volumes and use cases.

Key Results

The interviews revealed several key results from the investment in the Dell EMC Ready Solutions for Cloudera Hadoop. This investment:

- › **Reduced the dependency on and costs of legacy environments.** Data storage in Hadoop was dramatically less expensive than legacy storage in data warehouses. By extracting, transforming, and loading data from their legacy environments into Hadoop (ETL offload), organizations realized significant hardware, software, and administrative cost savings.
- › **Improved time-to-value.** Dell EMC's Ready Solutions for Cloudera Hadoop, developed in collaboration with Cloudera and accelerated by Intel, made implementation of the Hadoop environment relatively quick and easy. Organizations speculated that if they had tried to implement on their own, it would have taken six to 12 months longer to hire the expertise, figure out the correct configurations, and deploy the platform. They further speculated that the learning curve would have been much more difficult after deployment, with more errors and mistakes, which they avoided thanks to the Ready Solutions and their partnership with Dell EMC and Cloudera. The VP of big data for a financial services company framed the challenge: "It used to take 12 to 18 months to architect a new solution with software, hardware, PoC, and vendor selection. That 12- or 18-month cycle isn't affordable anymore. How do we cut that down into a six-month cycle or less?"
- › **Enabled new business services.** Hadoop's strengths are in the platform's ability to store and process large amounts of structured and unstructured data. A data lake, or large storage repository of data in its native format, was created by customers to incorporate new voluminous sources like social media, production data, or detailed usage data like transactions and logs. Additionally, this new storage capability allowed organizations to keep data spanning more than six months, enabling data scientists to develop new analysis and services for customers when previously they had to turn away customers or limit analysis due to capacity and processing constraints.

"The easiest part of this entire project was working with [Dell EMC]."

VP, database and technology marketing analytics



"Dell EMC's Ready Solutions provide all the possible capabilities you're going to use for data analytics problems in one infrastructure. You can run all your workloads, distribute your resources, use reporting, and take advantage of features like mass storage with redundancy, batch processing, the in-memory engine, stream processing, and machine learning."

Chief architect, information technology



"We have gone all-in on Hadoop. Reducing total cost of ownership was our first and foremost use case, but we're now finding that moving from our legacy systems into Hadoop is a tremendous boost to business because we have a far greater speed-to-market with analytics."

VP of big data, financial services



- › **Reduced time for business intelligence.** Hadoop's processing speed, combined with capabilities for machine learning and AI leveraging Spark and other tools built into the architecture, enabled organizations to run reports faster and to include near real-time data. Previously, loading data could take days, meaning that analysis was running with data that was, at best, from the prior day and, at worst, weeks old. Slow processing time further delayed reporting: It could take hours or even days to return results.
 - The retail organization interviewed explained how Hadoop's processing speed was crucial in generating reports for business leaders. With its previous business intelligence solution, the CEO would have to wait 10 minutes for a critical report — and the data was a day old. Not only was this frustrating, but it delayed the CEO from gaining insights into the business in a timely manner. With Hadoop, those reports now ran in under 10 seconds with near real-time results. Not only did this alleviate the CEO's frustration, but it also helped to gain high-level support in expanding the usage of Hadoop throughout the organization.
- › **Delivered high-level performance.** The Dell EMC Ready Solutions for Cloudera Hadoop enabled organizations to gain high performance from the initial deployment. One principal architect said: "Because they [Dell EMC] did the due diligence, because they partnered with Cloudera, because they understood what works, what types of workloads are optimized, and what are good use cases for different hardware configurations, we didn't have to be experts at hardware. That was huge." Organizations realized additional performance benefits in using Dell EMC's OpenManage Server Administrator (OMSA), which helped to analyze hardware requiring large changes to the nodes or in troubleshooting. Put succinctly in an interview, "It makes management of the environment much easier."

"Dell EMC is more business focused versus just technology focused. That helps us to say, 'Here's what we want to accomplish,' so Dell EMC can help us find the technology to meet those goals."

*VP of big data,
financial services*



"The Dell EMC and Cloudera partnership provides hardware that's specifically designed and optimized for the Hadoop's use cases and challenges. They work in conjunction to make sure the distribution runs well on the hardware, and they certify the hardware so if you run into any trouble, you have a hardware and software team that partner with you to solve the problem."

*Chief architect,
information technology*



Composite Organization

Based on the interviews, Forrester constructed a TEI framework, a composite company, and an associated ROI analysis that illustrates the areas financially affected. The composite organization, called *The Representative Organization*, is representative of the six companies that Forrester interviewed and is used to present the aggregate financial analysis. The composite organization that Forrester synthesized from the customer interviews has the following characteristics:

Description of composite. *The Representative Organization* is a US-based B2C retailer that sells products directly to consumers in stores and online. It has \$200 million in annual revenue and 1,000 employees. Big data and Hadoop are critical to growing and staying competitive, while also protecting its security posture and reducing fraud.

Deployment characteristics. At initial implementation, it had one cluster with 24 data nodes and three infrastructure nodes, which tripled in size to 72 data nodes in Year 2. A new cluster of 24 data nodes and three infrastructure nodes was rolled out in Year 3, along with further growth of the initial cluster to 120 data nodes for a total Hadoop deployment of 144 data nodes and six infrastructure nodes in Year 3. Its data requirements of 512 terabytes (TB) therefore grew to 1,536 TB in Year 2 and 3,072 TB in Year 3. Four FTEs supported the deployment in Year 1, growing to six FTEs in Year 3.



Key assumptions

- › B2C product retailer
- › \$200M annual revenue
- › 1,000 employees
- › 27-node initial deployment grows to 150 nodes
- › 512 TB data storage increases to 3,072 TB
- › Four to six FTEs manage the Hadoop deployment

Analysis Of Benefits

QUANTIFIED BENEFIT DATA AS APPLIED TO THE COMPOSITE

Total Benefits

REF.	BENEFIT	YEAR 1	YEAR 2	YEAR 3	TOTAL	PRESENT VALUE
Atr	Legacy hardware savings	\$3,111,459	\$6,222,918	\$9,334,377	\$18,668,754	\$14,984,562
Btr	Profit from new Hadoop-enabled business	\$400,000	\$1,600,000	\$3,200,000	\$5,200,000	\$4,090,158
Ctr	Fraud reduction	\$346,758	\$558,195	\$811,920	\$1,716,873	\$1,386,560
Dtr	Administrative cost savings	\$108,000	\$324,000	\$540,000	\$972,000	\$771,660
Etr	Improved data scientist productivity	\$126,360	\$252,720	\$336,960	\$716,040	\$576,895
	Total benefits (risk-adjusted)	\$4,092,577	\$8,957,833	\$14,223,257	\$27,273,667	\$21,809,835

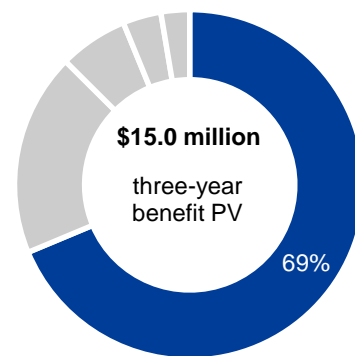
Legacy Hardware Savings

All customers interviewed for this study found significant value in the reduced cost of data storage using Hadoop. Organizations could extract, transform, and load (ETL offload) data from their data warehouses to a data lake in Hadoop. These interviewees explained that storage in Hadoop was a fraction of the cost of their legacy systems, even including Hadoop's need for triple redundancy.

- › The information technology company interviewed for this study stated that its legacy environment was 20x more expensive than the equivalent storage in its Hadoop cluster. This organization moved almost 700 terabytes from its legacy environment using ETL offload in Hadoop.
- › A financial services company reduced costs by at least 67% via ETL offload from its data warehouse to Hadoop, as the VP of big data described: "It costs us a mean of \$120K to support a terabyte of data in our legacy system. By moving to Hadoop with the Dell EMC Ready Solutions, we've reduced the cost to less than \$40K per terabyte. When you start doing that over petabytes of data, you can see why we have invested all-in on Hadoop. We can support new businesses while also reducing the footprint and the dollar size for our total cost."

The Representative Organization leveraged ETL offload to reduce its database cost per terabyte. Prior to Hadoop, *The Representative Organization* had a relational database and data warehouse platform serving its needs. With the Dell EMC Ready Solutions for Cloudera Hadoop, the overall cost per terabyte (including triple redundancy requirements) was one-twelfth the cost of its legacy platform. As a result, *The Representative Organization* moved data off its warehouse and avoided growing its legacy system as the business grew. However, much of this new data in Hadoop would likely not have been collected, stored, and analyzed in the legacy system at all as it was used for new business with Hadoop. Forrester therefore assumed that two-thirds of this new data demand would not have existed in the legacy environment, leaving only one-third to be avoided.

The table above shows the total of all benefits across the areas listed below, as well as present values (PVs) discounted at 10%. Over three years, the composite organization expects risk-adjusted total benefits to be a PV of more than \$21.8 million.



Legacy hardware savings:
69% of total benefits

"We can support new businesses while also reducing the footprint and the dollar size for our total cost."

VP of big data,
financial services



The Representative Organization's legacy database cost per terabyte was \$21,660. Given the size of the Hadoop environment excluding both triple redundancy and the two-thirds of data that only existed due to new business powered by Hadoop, The Representative Organization was able to avoid the cost of purchasing 169, 338, and 507 terabytes of legacy database storage in years 1, 2, and 3, respectively.

The interviewed organizations provided a broad range of legacy costs per terabyte. Further, some organizations eliminated or heavily downsized legacy warehouses while others simply avoided scaling up the environment further by instead growing the Hadoop environment. To account for these risks, Forrester adjusted this benefit downward by 15%, yielding a three-year risk-adjusted total PV of \$14,984,562.

Impact risk is the risk that the business or technology needs of the organization may not be met by the investment, resulting in lower overall total benefits. The greater the uncertainty, the wider the potential range of outcomes for benefit estimates.

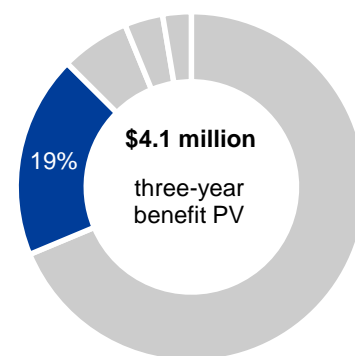
Legacy Hardware Savings: Calculation Table

REF.	METRIC	CALC.	Year 1	Year 2	Year 3
A1	Legacy database cost per terabyte	12x cost of Hadoop	\$21,660	\$21,660	\$21,660
A2	Terabytes of new Hadoop database storage, including triple redundancy	F4*64	1,536 <i>*Note value from initial period</i>	3,072	4,608
A3	Terabytes of new Hadoop storage, excluding triple redundancy	A2/3	512	1,024	1,536
A4	Percentage of Hadoop storage offloaded from legacy storage		33%	33%	33%
A5	Terabytes of legacy storage avoided with Hadoop ETL offload	A3*A4	169	338	507
At	Legacy hardware savings	A1*A5	\$3,660,540	\$7,321,080	\$10,981,620
	Risk adjustment	↓15%			
Atr	Legacy hardware savings (risk-adjusted)		\$3,111,459	\$6,222,918	\$9,334,377

Profit From New Hadoop-Enabled Business

Hadoop created a new way to store, manage, access, and analyze data for organizations and their customers. Insights powered new businesses and improved existing offerings. The VP of database and technology at a marketing analytics company shared, "In one year, business has doubled; without Hadoop, our business would not have survived."

- Streaming data made it possible to use data in near real time, enabling organizations to gain a live understanding of what was going on in the market. They could then react quickly to these insights to better serve their customers and drive increased sales. As the VP of big data for a financial services company described, "Hadoop helps you understand the right data to measure, then watch activity in real time to make virtually instant decisions and recommendations."
- One manufacturer recouped \$15 to \$25 million of revenue annually with Hadoop. It had many different parts and vendors that went into its finished product, and when the final products failed inspection, historically it would take up to three weeks to identify the issue — causing production delays and lost sales. With Hadoop, the customer could now identify the faulty part in hours and work with its vendor to get new working parts as quickly as possible and reduce losses.



Profit from new Hadoop-enabled business: **19%** of total benefits

- › Improving the speed of reporting drove sales for the financial services company, as the VP of big data described: “Moving from legacy is a tremendous boost to business, and we now have a far greater speed-to-market. Instead of coming out once a month, our reports can come out once per week.”
- › The chief architect for an information technology company described how streaming is essential for the internet of things (IoT) to load data fast, process it, make instant decisions, and then store it for later usage. “Spark is a great platform for machine learning and handling data that is in transition. There are many use cases where this comes into play, such as IoT, so you get a huge benefit from having the entire ecosystem bundled together.”

Interviewees also highlighted faster and easier deployment as a key benefit of the Dell EMC Ready Solutions for Cloudera Hadoop. Without having the in-house expertise in Hadoop, it could have taken six to 12 months longer to build and test a Hadoop solution and get into production. Organizations realized further value thanks to the reduced effort required to scale the deployment after launch to accommodate business growth. The Dell EMC Ready Solutions provided not only a faster time-to-value, but a proven configuration that optimized performance for workloads.

- › A financial services company interviewed for this study described how Dell EMC reduced the design time by at least 33% and the deployment time by about 50% — ultimately launching in six months instead of the expected 12 to 18 months.
- › One principal architect described the faster time-to-market enabled by the Dell EMC partnership with Cloudera: “We were able to get to production a lot quicker, and we didn’t have to go hire a hardware consultant or have a consultant come up and help us solve the problem and answer those questions that we didn’t know how to answer.” The architect continued, “Because [Dell EMC] did the due diligence, because they partnered with Cloudera and Intel, and because they understood what in fact works, what types of workloads are optimized, and what are good use cases for different hardware configurations, we didn’t have to be experts at hardware. That was huge.”

The Representative Organization utilized Cloudera Hadoop to offer new products and services and leveraged customer insights to improve and better target its existing offerings to customers. This helped the organization access new markets and customers and increase conversion, acquisition, retention, and average order value, leading to increased revenue. Better customer insights enabled improved suggestions and marketing, increasing the lifetime value of customers.

- › *The Representative Organization* realized additional incremental revenues of \$20 million and \$40 million in Year 2 and Year 3, respectively, thanks to Dell EMC and Cloudera Hadoop.
- › It would have taken *The Representative Organization* at least six months longer to deploy Hadoop on its own. Therefore, in Year 1, it earned an additional \$5 million in incremental revenue that was only possible due to this faster time-to-market.
- › Incremental revenue was burdened by an estimated 10% operating margin to account for the costs of goods sold.

Organizations will leverage Hadoop in varying ways for different use

“The value is that we were able to do analytics on data that we could never do analytics on before.”

*Architect,
retail*



“With the Dell EMC Ready Solutions, we’ve reduced the design time by at least a third and halved the time for ‘rack and stack’ because they are already preconfigured.”

*VP of big data,
financial services*



“We were able to get to production a lot quicker [with Dell EMC’s Ready Solutions], and we didn’t have to go hire a hardware consultant or have a consultant come up and help us solve the problem and answer questions that we didn’t know how to answer.”

*Principal architect,
digital media services*



The Dell EMC Ready Solutions for Hadoop reduced time-to-market by at least six months.

cases. While interviewees pointed to new business ventures and improved offerings thanks to Hadoop, the financial impact varied widely. Further, external factors can impact sales and diminish the incremental revenue increases from Hadoop-enabled business. To account for this variation and unpredictability, Forrester adjusted this benefit downward by 20%, yielding a three-year risk-adjusted total PV of \$4,090,158.

Profit From New Hadoop-Enabled Business: Calculation Table

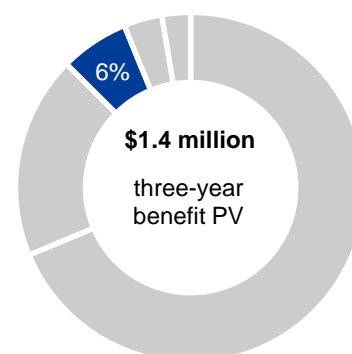
REF.	METRIC	CALC.	YEAR 1	YEAR 2	YEAR 3
B1	Revenue from new Hadoop-enabled services			\$20,000,000	\$40,000,000
B2	Increased revenue from new Hadoop-enabled services due to faster time-to-market		\$5,000,000		
B3	Operating margin rate		10%	10%	10%
Bt	Profit from new Hadoop-enabled business	$(B1+B2)*B3$	\$500,000	\$2,000,000	\$4,000,000
	Risk adjustment	↓20%			
Btr	Profit from new Hadoop-enabled business (risk-adjusted)		\$400,000	\$1,600,000	\$3,200,000

Fraud Reduction

Organizations leveraged powerful technologies like Hadoop and Spark on their clusters to conduct real-time analysis with machine learning for data-intensive use cases, such as IoT management, manufacturing, log analysis, security, and transaction monitoring. One key use case identified by customers was real-time fraud identification and prevention. Businesses in industries such as retail, eCommerce, and financial services lose substantial amounts of revenue to fraud every year, and losses are particularly pronounced for those that conduct their business digitally. LexisNexis identified that the average company loses 1.99% of revenue annually with a range from 1.58% for retailers to 2.39% for financial services companies.¹

- › The interviewed financial services company described how, with the Ready Solutions, it could monitor transactions and follow the money trail using data such as geolocation, facial recognition, and web addresses. Its VP of big data described: “You need to stop the money flow in fraudulent transactions, but you only have seconds to make sure it’s not transferred. Literally seconds, that’s how fast these guys move. Hadoop enables us to tackle this challenge, and every dollar of fraud we stop goes directly to the bottom line. That’s why moving this analysis as close to the transaction as possible is very important.”

Fraud has a major impact on business profitability, and every dollar prevented goes straight to a business’ bottom line. *The Representative Organization* leveraged the power of its Hadoop deployment to monitor transactions and block a higher rate of suspected fraud than previously possible. The organization reduced fraudulent transactions by 10% initially; as it gathered more data and improved its models, it increased this reduction to 15% in Year 2 and 20% in Year 3. Therefore, the business had an effective reduction in the percentage of revenues lost to fraud from 1.99% to 1.79% in Year 1, 1.69% in Year 2, and 1.59% in Year 3. The organization avoided losing 0.2%, 0.3%, and 0.4% of total revenues of \$205 million, \$220 million, and \$240 million in years 1, 2, and 3, respectively.



Fraud reduction: 6% of total benefits

“You need to stop the money flow in fraudulent transactions, but you only have seconds to make sure it’s not transferred. Literally seconds, that’s how fast these guys move. Hadoop enables us to tackle this challenge, and every dollar of fraud we stop goes directly to the bottom line. That’s why moving this analysis as close to the transaction as possible is very important.”

VP of big data,
financial services



Fraud reduction may not be relevant for all customers, and for each organization, the impact may vary based on industry and the other fraud prevention systems in place. Organizations may not achieve reductions as significant as those modeled in this calculation through their usage of Hadoop, as success hinges on the data fed into the system and the machine learning models themselves; therefore, the design of the architecture created by engineers is essential to a significant reduction. To compensate, Forrester adjusted this benefit downward by 15%, yielding a three-year risk-adjusted total PV of \$1,386,560.

Fraud Reduction: Calculation Table

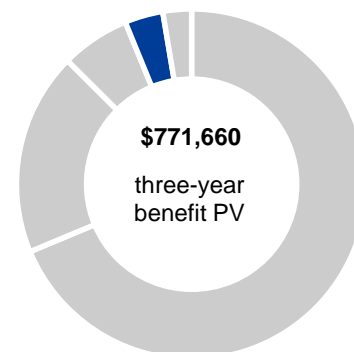
REF.	METRIC	CALC.	YEAR 1	YEAR 2	YEAR 3
C1	Annual business revenue	\$200M+B1+B2	\$205,000,000	\$220,000,000	\$240,000,000
C2	Average percentage of revenue lost annually due to fraud	LexisNexis 2017 True Cost of Fraud Study	1.99%	1.99%	1.99%
C3	Percentage reduction in fraud with Hadoop		10%	15%	20%
Ct	Fraud reduction	C1*C2*C3	\$407,950	\$656,700	\$955,200
	Risk adjustment	↓15%			
Ctr	Fraud reduction (risk-adjusted)		\$346,758	\$558,195	\$811,920

Administrative Cost Savings

Organizations realized increased operational efficiency from the implementation of the Dell EMC Ready Solutions for Cloudera Hadoop. Administrators could now support much larger data sets in less time; as data shifted into the Hadoop environment and the legacy platform was repurposed, these administrators could take on other value-add tasks for the business. The chief architect of an information technology company explained one area where efficiencies were gained: “We can do ETL on a much larger data set much faster by distributing it across the cluster with parallel processing. Days of processing time can turn into hours with Hadoop.”

The Representative Organization repurposed one FTE in Year 1, three FTEs in Year 2, and five FTEs in Year 3 thanks to operational efficiency gains. Forrester calculated this savings using an average fully burdened annual salary of \$120,000 for administrators.

Interviewed organizations provided a broad range of operational efficiency depending upon their legacy platform, use case, and size of deployment. To compensate, Forrester adjusted this benefit downward by 10%, yielding a three-year risk-adjusted total PV of \$771,660.



**Administrative cost savings:
3% of total benefits**

Administrative Cost Savings: Calculation Table

REF.	METRIC	CALC.	YEAR 1	YEAR 2	YEAR 3
D1	Number of administrative FTEs reprovisioned		1	3	5
D2	Average fully loaded annual salary for administrators		\$120,000	\$120,000	\$120,000
Dt	Administrative cost savings	D1*D2	\$120,000	\$360,000	\$600,000
	Risk adjustment	↓10%			
Dtr	Administrative cost savings (risk-adjusted)		\$108,000	\$324,000	\$540,000

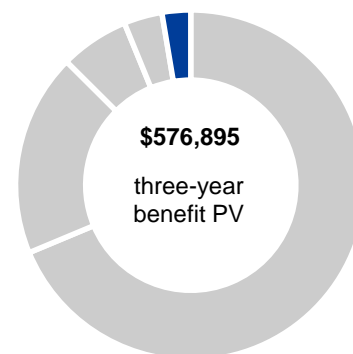
Improved Data Scientist Productivity

Running workloads using technologies such as Spark within the Hadoop environment was significantly faster than other forms of data analysis conducted on legacy platforms. Queries and analytics that took days could be completed in hours or less, and certain jobs could be completed near instantaneously thanks to Hadoop's parallel processing power.

- › For one financial services company, major analytics reports that used to take two to three weeks could be completed in two to three days. Certain queries could even be completed virtually instantly in Hadoop that would have taken days to run previously. Data scientists ultimately improved productivity by 30% thanks to Hadoop.
- › One interviewee described how Hadoop sheds light on what data has been collected: "You can build a data lake, but you might not know what is in there. With other databases, it could take a week just to figure out what you have, and it becomes a data swamp. Hadoop helps out tremendously to figure out what is in your data lake with just a few simple queries. When you look at the data lake, it comes to, 'Yes, we have it all, but what do I do now that we know what we have? Which one of these pieces of data make a difference?'"
- › One information technology company improved the scripting efficiency of a 23-developer team by 40% with Hadoop.
- › The chief architect of an information technology company described, "Running Spark with Hadoop underneath across a large cluster lets you distribute workloads and execute more iterations of model training to build a better model much, much faster." The architect continued: "We ran one machine learning job on a single workstation and it took three or four days. We then optimized it for Spark, ran it on a five-node cluster, and we were able to complete the same job in just a couple of hours."

Forrester modeled this benefit for *The Representative Organization* based on the following assumptions:

- › Twelve data scientists supporting the business could analyze data on the Hadoop platform in Year 1 and Year 2. As a new cluster was launched in Year 3, an additional four data scientists could run workloads on the new cluster.
- › Data scientists using Hadoop improved their productivity by 15% in Year 1 as they learned the solution and by 30% in Year 2 and Year 3.
- › With a base of 2,080 total labor hours per data scientist, per year, the organization saved 312 hours per FTE in Year 1 and 624 hours per FTE in years 2 and 3. The organization therefore saved a total of 3,744 labor hours in Year 1, 7,488 hours in Year 2, and 9,984 hours in Year 3.
- › Not all time saved was necessarily devoted to value-add tasks. The organization recaptured 50% of the time saved for additional value.
- › Data scientists earned an average fully loaded salary of \$75 per hour. Productivity improvements for data scientists will vary depending on the types and complexity of workloads they run, the amount of those workloads that are run in Hadoop, and the time required to analyze data in their legacy environments. The number of data scientists using Hadoop and their average salary will impact the value of improved productivity. To compensate, Forrester adjusted this benefit downward by 10%, yielding a three-year risk-adjusted total PV of \$576,895.



Improved data scientist productivity: **3%** of total benefits



30% improvement in data scientist productivity

"The ability to add a wide variety of data without having to 'structure' it into a data warehouse has been really beneficial. Hadoop helps us profile it to create better libraries and catalogs and helps us create new reports in two or three days instead of two or three weeks."

*VP of big data,
financial services*



Improved Data Scientist Productivity: Calculation Table

REF.	METRIC	CALC.	YEAR 1	YEAR 2	YEAR 3
E1	Number of data scientists who leverage Hadoop		12	12	16
E2	Productivity improvement for data scientists with Hadoop		15%	30%	30%
E3	Hours of labor saved per scientist for structuring data	2,080*E2	312	624	624
E4	Total hours of data scientist labor saved	E1*E3	3,744	7,488	9,984
E5	Percentage of time recaptured for productivity		50%	50%	50%
E6	Hours of data scientist labor recaptured for productivity	E4*E5	1,872	3,744	4,992
E7	Average fully loaded hourly salary for data scientists		\$75	\$75	\$75
Et	Improved data scientist productivity	E6*E7	\$140,400	\$280,800	\$374,400
	Risk adjustment	↓10%			
Etr	Improved data scientist productivity (risk-adjusted)		\$126,360	\$252,720	\$336,960

Unquantified Benefits

The interviewed organizations cited several other benefits that could not be quantified in this study, including the following:

- Optimized system capacity by 30% with a Hadoop research cluster.** One organization leveraged Hadoop to build a research cluster to identify usage patterns and the types of queries being run and to understand system failures and response times. It compiled the information to improve response times by categorizing data and optimizing the nodes. This tuning process allowed it to reduce faceting times and free up 30% of additional capacity.
- Improved system availability and performance, which reduced customer support hours by 25% and improved employee productivity.** One interviewed information technology company reduced the mean-time-to-resolve issues by 50% thanks to its intelligence engine that used Hadoop for log analysis. Improvements led to both reduced support calls and reduced troubleshooting effort, ultimately reducing customer support hours by 25%.
- Improved ability to meet SLAs by 8%.** The information technology company interviewed for this study described how its system experienced major spikes due to seasonal events and how it struggled to meet data scientist workloads, as all the data scientists would wait until the evening to submit their workloads. With Hadoop, the organization could better plan and clear our warehouse space — enabling it to reduce a 10% failure rate to only 2%.
- Tightened data security, identifying and remediating almost 100 million attacks and achieving PCI compliance.** One financial services company explained how it used machine learning with Hadoop to identify and, when possible, automatically remediate threats. For those threats that remained, Hadoop provided key insights and enabled security analysts to efficiently remediate the threats. Further, by identifying patterns, analysts could even go on the hunt and disrupt the attackers and their network. Not only did this reduce the risk and cost of breaches, but it helped protect the company's brand reputation and therefore its market share.

“We use Hadoop to monitor our systems in real time to make sure we don't get a breach. We've fought tens of millions of attempts, and the ability to detect and fight them is essential to protecting our brand reputation. Ensuring that our customers feel safe and secure is a big part of getting our market share.”

*VP of big data,
financial services*



Flexibility

The value of flexibility is clearly unique to each customer, and the measure of its value varies from organization to organization. There are multiple scenarios in which a customer might choose to implement the Dell EMC Ready Solutions and later realize additional uses and business opportunities, including:

- › Implement the Dell EMC Ready Solutions for Cloudera Hadoop for a specific, predetermined use case, then expand the deployment to tackle additional use cases as they become apparent.
- › Leverage other technologies such as Spark and Spark Streaming that can come as part of the Cloudera distribution with Dell EMC's Ready Solutions for Cloudera Hadoop.

Flexibility would also be quantified when evaluated as part of a specific project (described in more detail in Appendix A).

Flexibility, as defined by TEI, represents an investment in additional capacity or capability that could be turned into business benefit for a future additional investment. This provides an organization with the "right" or the ability to engage in future initiatives but not the obligation to do so.

Analysis Of Costs

QUANTIFIED COST DATA AS APPLIED TO THE COMPOSITE

Total Costs

REF.	COST	INITIAL	YEAR 1	YEAR 2	YEAR 3	TOTAL	PRESENT VALUE
Ftr	Hardware costs	\$930,888	\$0	\$1,729,022	\$2,659,910	\$5,319,821	\$4,358,262
Gtr	Software license costs	\$39,690	\$0	\$110,250	\$220,500	\$370,440	\$296,471
Htr	Implementation costs	\$369,321	\$0	\$0	\$0	\$369,321	\$369,321
ltr	Hadoop operational costs	\$0	\$646,800	\$646,800	\$950,400	\$2,244,000	\$1,836,595
	Total costs (risk-adjusted)	\$1,339,899	\$646,800	\$2,486,072	\$3,830,810	\$8,303,581	\$6,860,649

Hardware Costs

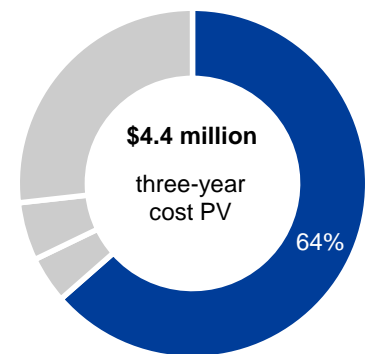
The Representative Organization purchased the Dell EMC Ready Solutions for Cloudera Hadoop. Hadoop architecture requires triple redundancy in data storage; with an estimated capacity of 64 terabytes per node, each node has an operable storage of approximately 21.3 terabytes.

- › The organization purchased infrastructure nodes at a cost of \$21,072 each and data nodes at a cost of \$34,306 each.
- › In Year 1, *The Representative Organization* deployed an initial cluster of three infrastructure nodes and 24 data nodes.
- › As data needs grew with the new business demand, so did the number of nodes required, with the initial cluster growing by 48 data nodes in Year 2 to a total of 72 data nodes and three infrastructure nodes.
- › In Year 3, *The Representative Organization* rolled out a new cluster for added use cases with three infrastructure nodes and 24 data nodes, while also expanding its initial cluster by another 48 data nodes.

Over three years, *The Representative Organization* acquired a total of six master nodes and 144 data nodes, with a total storage capacity of 3,072 terabytes (3.1 petabytes) of operable storage.

Hardware costs are based on Dell EMC's retail pricing, which considers average discounts and packaging. The cost per node varies from organization to organization and even within an organization's deployment, as different nodes may require varying memory, storage, and processing power depending upon their use cases. To account for these risks, Forrester adjusted this cost upward by 5%, yielding a three-year risk-adjusted total PV of \$4,358,262.

The table above shows the total of all costs across the areas listed below, as well as present values (PVs) discounted at 10%. Over three years, the composite organization expects risk-adjusted total costs to be a PV of almost \$6.9 million.



**Hardware costs:
64% of total benefits**

Implementation risk is the risk that a proposed investment may deviate from the original or expected requirements, resulting in higher costs than anticipated. The greater the uncertainty, the wider the potential range of outcomes for cost estimates.

Hardware Costs: Calculation Table

REF.	METRIC	CALC.	INITIAL	YEAR 1	YEAR 2	YEAR 3
F1	Cost per infrastructure node		\$21,072	\$21,072	\$21,072	\$21,072
F2	Number of new infrastructure nodes		3	0	0	3
F3	Cost per data node	64 TB per node	\$34,306	\$34,306	\$34,306	\$34,306
F4	Number of new data nodes		24	0	48	72
Ft	Hardware costs	$F1 \times F2 + F3 \times F4$	\$886,560	\$0	\$1,646,688	\$2,533,248
	Risk adjustment	↑5%				
Ftr	Hardware costs (risk-adjusted)		\$930,888	\$0	\$1,729,022	\$2,659,910

Software License Costs

The organization incurred software licensing fees for the Cloudera distribution of Hadoop during the initial implementation period and in subsequent years. The license cost was priced at approximately \$1,400 per node. *The Representative Organization* incurred this cost for each node as it grew from 27 initial nodes to 75 nodes and 150 nodes in years 2 and 3, respectively.

Software costs vary from organization to organization, considering different licensing agreements and other discounts. To account for these risks, Forrester adjusted this cost upward by 5%, yielding a three-year risk-adjusted total PV of \$296,471.

Software License Costs: Calculation Table

REF.	METRIC	CALC.	INITIAL	YEAR 1	YEAR 2	YEAR 3
G1	Annual Cloudera license cost per node		\$1,400	\$0	\$1,400	\$1,400
G2	Number of new infrastructure nodes	F2	3	0	0	3
G3	Total number of infrastructure nodes		3	3	3	6
G4	Number of new data nodes	F4	24	0	48	72
G5	Total number of data nodes	$G4 + G5^{[PY]}$	24	24	72	144
G6	Total number of nodes	$G3 + G5$	27	27	75	150
Gt	Software license costs	$G1 \times G6$	\$37,800	\$0	\$105,000	\$210,000
	Risk adjustment	↑5%				
Gtr	Software license costs (risk-adjusted)		\$39,690	\$0	\$110,250	\$220,500

Implementation Costs

Initial implementation costs included Dell EMC Professional Services in the amount of \$23,746, along with four internal FTEs at an average rate of \$75 per hour. The implementation of the original Hadoop cluster took six months in total, costing \$312,000 in labor.

Implementation costs and timelines can vary from organization to organization; to compensate, Forrester adjusted this cost upward by 10%, yielding a three-year risk-adjusted total PV of \$369,321.

Implementation Costs: Calculation Table

REF.	METRIC	CALC.	INITIAL	YEAR 1	YEAR 2	YEAR 3
H1	Dell EMC professional services		\$23,746			
H2	Total implementation time in weeks		26			
H3	Number of FTEs involved in implementation		4			
H4	Average fully loaded hourly salary for data scientists and developers	E7	\$75			
H5	Internal labor cost	$H2*40*H3*H4$	\$312,000			
Ht	Implementation costs	$H1+H5$	\$335,746	\$0	\$0	\$0
	Risk adjustment	↑10%				
Htr	Implementation costs (risk-adjusted)		\$369,321	\$0	\$0	\$0

Hadoop Operational Costs

The Representative Organization needed to hire one administrator and three data engineers to manage and support the initial deployment of Hadoop. The business hired another administrator and data engineer as it added a new cluster in the third year. The fully loaded annual salary was \$120,000 for administrators and \$156,000 for data engineers.

The number of resources required and the mix between internal versus external resources and onshore versus offshore resources can change from organization to organization. Certain organizations may require more or fewer resources depending on the number of clusters, their size, and the complexity of the data and workloads being run on the system. To compensate, Forrester adjusted this cost upward by 10%, yielding a three-year risk-adjusted total PV of \$1,836,595.



Four to six FTEs
spend 100% of their time
on ongoing
management.

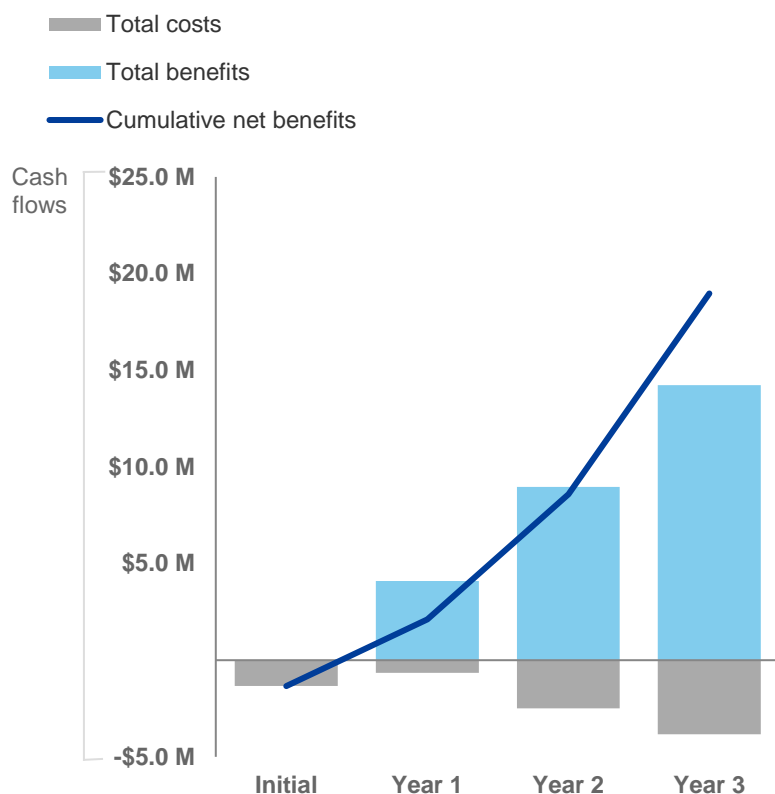
Hadoop Operational Costs: Calculation Table

REF.	METRIC	CALC.	INITIAL	YEAR 1	YEAR 2	YEAR 3
I1	Number of administrators to manage Hadoop system			1	1	2
I2	Average fully loaded annual salary for Hadoop administrators	D2		\$120,000	\$120,000	\$120,000
I3	Number of data engineers to manage Hadoop system			3	3	4
I4	Average fully loaded annual salary for Hadoop data engineers	$H4*2,080$		\$156,000	\$156,000	\$156,000
It	Hadoop operational costs	$I1*I2+I3*I4$	\$0	\$588,000	\$588,000	\$864,000
	Risk adjustment	↑10%				
Itr	Hadoop operational costs (risk-adjusted)		\$0	\$646,800	\$646,800	\$950,400

Financial Summary

CONSOLIDATED THREE-YEAR RISK-ADJUSTED METRICS

Cash Flow Chart (Risk-Adjusted)



The financial results calculated in the Benefits and Costs sections can be used to determine the ROI, NPV, and payback period for the composite organization's investment. Forrester assumes a yearly discount rate of 10% for this analysis.



These risk-adjusted ROI, NPV, and payback period values are determined by applying risk-adjustment factors to the unadjusted results in each Benefit and Cost section.

Cash Flow Table (Risk-Adjusted)

	INITIAL	YEAR 1	YEAR 2	YEAR 3	TOTAL	PRESENT VALUE
Total costs	(\$1,339,899)	(\$646,800)	(\$2,486,072)	(\$3,830,810)	(\$8,303,581)	(\$6,860,649)
Total benefits	\$0	\$4,092,577	\$8,957,833	\$14,223,257	\$27,273,667	\$21,809,835
Net benefits	(\$1,339,899)	\$3,445,777	\$6,471,761	\$10,392,447	\$18,970,085	\$14,949,186
ROI						218%
Payback period						<6 months

Dell EMC Ready Solutions: Overview

The following information is provided by Dell EMC. Forrester has not validated any claims and does not endorse Dell EMC or its offerings.

Digital transformation is causing churn, uncertainty, and disruption for many business leaders who need to act quickly as pressure increases from all directions. Big data and analytics are at the core of this transformation. However, without the right tools for big data analytics, you'll have to conduct multiple, limited data analyses and then find a way to synthesize the results, involving a lot of manual effort and subjective analysis.

With Apache® Hadoop as a foundational component of the big data and analytics solution stack, you gain the ability to quickly process large sets of disparate data for a more comprehensive view of your customers, operations, opportunities, risks, and more. Despite these tantalizing benefits, many organizations struggle — either to begin their data analytics journey or to make Hadoop projects successful once they've begun. They are often impeded by a lack of Hadoop expertise and end up spending too much time and effort on front-end work before they can get to the results of an operational solution.

Expertise and infrastructure matter when building a Hadoop environment. That's why Dell EMC has teamed up with industry leaders — such as Intel®, Hortonworks®, Cloudera®, and Syncsort® — to remove the uncertainty and barriers that may dissuade you from deploying Hadoop.

The cost-effective, future-ready Dell EMC Ready Solutions for Hadoop is an easy-to-implement solution designed to help you efficiently harness Hadoop and the power of data analytics to drive competitive advantage.

- › **Leverage an optimized solution.** The Dell EMC Ready Solutions for Hadoop is developed jointly with leading Hadoop distributions and is based on extensive customer experience with real-world Hadoop production installations. This solution offers the hardware, software, resources, and services needed to deploy and manage Hadoop in a production environment.
- › **Reduce costs.** The Dell EMC Ready Solutions for Hadoop offers compelling total cost of ownership (TCO) benefits by using cost-optimized Dell EMC servers, networking, and storage to decrease the cost to process and store large data sets versus traditional business intelligence (BI) and analytics solutions.
- › **Deliver outstanding performance.** The Dell EMC Ready Solutions for Hadoop has been engineered and certified to provide known performance parameters and deployment methods, for excellent performance and minimal risk.

Dell EMC Ready Solutions For Hadoop

The Dell EMC Ready Solutions for Hadoop is designed to address data analytics requirements, reduce costs, and deliver outstanding performance. Dell EMC has been working with the leading innovators in big data since 2008. With deep roots in data analytics solutions and Hadoop — and other leading partners in data analytics — Dell EMC has the expertise, tools, and solutions needed to drive successful, flexible, and scalable Hadoop deployments.

Choose Your Hadoop Distribution

Dell EMC supports both Cloudera Enterprise and the Hortonworks Data Platform®, so you can choose the right Hadoop solution for your needs. Both offer optional Syncsort ETL processing, allowing you to capitalize on the advantages of Hadoop while making better use of existing enterprise data warehouse (EDW) investments.

Configuration details

Dell EMC Ready Bundles for Hadoop offer a variety of configurations to meet your needs.²

Dell EMC Ready Bundle for Hadoop configuration options			
	Direct-attached storage		Isilon shared storage
	Rack servers	Modular infrastructure	Modular infrastructure
Dell EMC Ready Bundle for Cloudera Hadoop	<ul style="list-style-type: none"> Cloudera Enterprise 5.12 PowerEdge R740xd or R640 Servers 	<ul style="list-style-type: none"> Cloudera Enterprise 5.12 PowerEdge FX2 FC630 Servers 	
Dell EMC Ready Bundle for Hortonworks Hadoop	<ul style="list-style-type: none"> Hortonworks Data Platform 2.6 PowerEdge R730xd Servers 	<ul style="list-style-type: none"> Hortonworks Data Platform 2.6 PowerEdge FX2 FC630 Servers 	<ul style="list-style-type: none"> Hortonworks Data Platform 2.5 PowerEdge FX2 FC630 Servers Isilon X410 Storage
Dell EMC Quickstart for Hadoop	<ul style="list-style-type: none"> Cloudera Enterprise 5.12 PowerEdge R740xd Servers 		

Solution Use Case Summary

The Dell EMC Ready Solutions for Hadoop, accelerated by Intel, is designed to address the following use cases:

Use case	Description
Big data analytics	Ability to query in real time at the speed of thought on petabyte scale unstructured and semistructured data using HBase and Hive.
ETL offload	Offload the extract, transform, load (ETL) process from a relational management database or enterprise data warehouse into a Hadoop cluster.
Data warehouse optimization	Augment the traditional relational management database or enterprise data warehouse with Hadoop. Hadoop acts as single data hub for all data types.
Data storage	Collect and store unstructured and semistructured data in a secure, fault-resilient scalable data store that can be organized and sorted for indexing and analysis.
Batch processing of unstructured data	Ability to batch-process (index, analyze, etc.) tens to hundreds of petabytes of unstructured and semistructured data.
Data archive	Active archival of medium-term (12 to 36 months) data from EDW/DBMS to expedite access, increase data retention time, or meet data retention policies or compliance requirements.
Integration with data warehouse	Extract, transfer, and load data in and out of Hadoop into separate DBMS for advanced analytics.
Big data visualization	Capture, index, and visualize unstructured and semistructured big data in real time.
Search and predictive analytics	Crawl, extract, index, and transform semistructured and unstructured data for search and predictive analytics.

For additional information visit: DellEMC.com/Hadoop.

Appendix A: Total Economic Impact

Total Economic Impact is a methodology developed by Forrester Research that enhances a company's technology decision-making processes and assists vendors in communicating the value proposition of their products and services to clients. The TEI methodology helps companies demonstrate, justify, and realize the tangible value of IT initiatives to both senior management and other key business stakeholders.

Total Economic Impact Approach



Benefits represent the value delivered to the business by the product. The TEI methodology places equal weight on the measure of benefits and the measure of costs, allowing for a full examination of the effect of the technology on the entire organization.



Costs consider all expenses necessary to deliver the proposed value, or benefits, of the product. The cost category within TEI captures incremental costs over the existing environment for ongoing costs associated with the solution.



Flexibility represents the strategic value that can be obtained for some future additional investment building on top of the initial investment already made. Having the ability to capture that benefit has a PV that can be estimated.



Risks measure the uncertainty of benefit and cost estimates given: 1) the likelihood that estimates will meet original projections and 2) the likelihood that estimates will be tracked over time. TEI risk factors are based on "triangular distribution."

The initial investment column contains costs incurred at "time 0" or at the beginning of Year 1 that are not discounted. All other cash flows are discounted using the discount rate at the end of the year. PV calculations are calculated for each total cost and benefit estimate. NPV calculations in the summary tables are the sum of the initial investment and the discounted cash flows in each year. Sums and present value calculations of the Total Benefits, Total Costs, and Cash Flow tables may not exactly add up, as some rounding may occur.



Present value (PV)

The present or current value of (discounted) cost and benefit estimates given at an interest rate (the discount rate). The PV of costs and benefits feed into the total NPV of cash flows.



Net present value (NPV)

The present or current value of (discounted) future net cash flows given an interest rate (the discount rate). A positive project NPV normally indicates that the investment should be made, unless other projects have higher NPVs.



Return on investment (ROI)

A project's expected return in percentage terms. ROI is calculated by dividing net benefits (benefits less costs) by costs.



Discount rate

The interest rate used in cash flow analysis to take into account the time value of money. Organizations typically use discount rates between 8% and 16%.



Payback period

The breakeven point for an investment. This is the point in time at which net benefits (benefits minus costs) equal initial investment or cost.

Appendix B: Endnotes

¹ Source: "LexisNexis® Risk Solutions 2017 True Cost of FraudSM Study," LexisNexis Risk Solutions, October 2017 (<https://risk.lexisnexis.com/insights-resources/research/2017-tcof>).