

A Forrester Total Economic
Impact™ Study
Commissioned By
Dell and Intel

Project Director:
Sean McCormick
November 2015

The Total Economic Impact™ Of The Dell | Cloudera Apache Hadoop Solution, Accelerated By Intel

Cost Savings And Business Benefits
Enabled By The Dell | Cloudera
Apache Hadoop Solution, Accelerated
By Intel

Table Of Contents

Executive Summary	3
Disclosures	5
TEI Framework and Methodology	6
Analysis	7
Financial Summary	20
Dell Cloudera Apache Hadoop Reference Architecture, Accelerated By Intel: Overview	21
Appendix A: Composite Organization Description	23
Appendix B: Total Economic Impact™ Overview	24
Appendix C: Forrester And The Age Of The Customer	25
Appendix D: Glossary	26
Appendix E: Endnotes	26

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.

© 2015, 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 www.forrester.com.

Executive Summary

Dell and Intel commissioned Forrester Consulting to conduct a Total Economic Impact™ (TEI) study to examine the potential return on investment (ROI) enterprises may realize by deploying Dell | Cloudera Apache Hadoop Solutions, accelerated by Intel. The purpose of this study is to provide readers with a framework to evaluate the potential financial impact of the Hadoop solution on their organizations.

To better understand the benefits, costs, and risks associated with an implementation of a Dell | Cloudera Apache Hadoop Solution, accelerated by Intel, Forrester interviewed several customers with multiple years of experience using Hadoop. Since many companies have different requirements and needs, Dell offers multiple Hadoop solutions. The Dell QuickStart for Cloudera Hadoop is a starter bundle that delivers an all-in-one solution to enable organizations to test their proof of concept using Hadoop. The ability to test a proof of concept to determine how Hadoop will be of value to them gives organizations just starting out with Hadoop an opportunity to learn as much as they can while testing their use case. Additionally, Dell, together with Cloudera and Intel, offers two reference architectures. One is an extract, transform, and load (ETL) offload reference architecture that helps companies implement a first use case using ETL offload to augment expensive and capacity-constrained traditional relational databases with an offload to Hadoop, where they realize value in data transformation from multiple data sources. And, for this study, we will focus on the value customers are achieving when deploying and implementing Hadoop using the blueprint that is the Dell | Cloudera Apache Hadoop Reference Architecture, accelerated by Intel.

Traditionally, customers had implemented enterprise data warehouse platforms as a centerpiece of their business intelligence solution. Organizations know they need to adopt big data solutions; however, few had the in-house expertise, knowledge base, or resources to do so because Hadoop is an emerging open source technology. These limitations led to prolonged implementation cycles due to configuration issues and platform instability. With the Dell | Cloudera Apache Hadoop Reference Architecture, customers are able to streamline their deployment to increase their time-to-value. Said one big data principal architect, “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.” They went on to say that Dell’s reference architecture sped up their time-to-value by six months compared with deploying on their own.

DELL HADOOP SOLUTIONS ACCELERATE NEW PRODUCT REVENUE

Our interviews with four existing customers detail the subsequent financial analysis, which found that a composite organization based on these interviewed organizations experienced the risk-adjusted ROI, benefits, and costs shown in Figure 1.¹ See Appendix A for a description of the composite organization.

The composite organization analysis points to benefits of \$11.4 million versus implementation and ongoing support costs of \$5.8 million, adding up to a net present value (NPV) of \$5.6 million.

This translates to benefits of more than \$119,035 per node, implementation and support costs of \$60,493 per node, and an NPV of \$58,542 per node. With Dell | Cloudera Apache Hadoop Solutions, accelerated by Intel, performance, operations, and analytics in the Hadoop environment are optimized to help organizations focus on growing business offerings rather than

Dell | Cloudera Hadoop Solutions increase time-to-value, provide analytical insights not previously available, and create new business opportunities.

The costs and benefits for a composite organization starting with 24 nodes and 300 terabytes and growing to 96 nodes and 1.4 petabytes of data over three years, based on customer interviews, include:

- **Initial investment costs: \$991,377.**
- **Average Annual costs: \$2,023,049.**
- **Total cost savings and benefits: \$11,427,408.**

A 24-node 300-terabyte deployment with no growth:

- **Initial investment costs: \$991,377.**
- **Average Annual costs: \$780,422.**
- **Total cost savings and benefits: \$5,017,751.**

wrestling with the configuration, deployment, and management of the Hadoop cluster. In the composite organization, Hadoop enabled the processing of more data than previously possible with its legacy data warehouse and in a fraction of the time, generating incremental sales of \$10 million to \$20 million per year in new services. One VP said, "In one year business has doubled; without Hadoop, our business would not have survived."

FIGURE 1
Financial Summary Showing Three-Year Risk-Adjusted Results



*A 97% ROI equates to \$1.97 of benefit for every \$1.00 spent.
Source: Forrester Research, Inc.

› **Benefits.** The composite organization experienced the following risk-adjusted benefits that represent those experienced by the interviewed companies:

- **Faster time-to-value, leading to \$900,000 of increased margin on new client services.** Through the use of Dell's reference architecture, the composite organization was able to deploy six months quicker, allowing it to earn \$5 million in revenue in the first year at a 20% margin rate.
- **Revenue growth from new client services led to \$3.7 million of incremental income over three years.** Utilizing Hadoop, the composite organization was able to offer new analytical services that previously weren't economical or possible. These new services generated \$10 million to \$20 million of revenue annually.
- **Avoided infrastructure costs of \$6.4 million over three years.** With the cost per terabyte (TB) to operate Hadoop being one-twelfth of our composite's legacy relational database costs, the composite organization was able to save \$6,000 per terabyte purchased.
- **Repurposed legacy database full-time equivalents (FTEs), reducing costs by \$450,135 over three years.** Coinciding with the avoidance in legacy database hardware, the composite organization was able to repurpose five FTEs in three years.

› **Costs.** The composite organization experienced the following risk-adjusted costs:

- **Initial Cloudera Enterprise software license costs of \$180,634.** In addition to the initial license cost, a recurring annual cost of \$7,168 per node for ongoing maintenance and support was recognized.
- **Hardware costs of \$555 per terabyte to build out the Hadoop platform.** Dell's Hadoop optimized servers accelerated by Intel cost the composite organization \$2.2 million over three years.
- **Support costs equivalent to four FTEs to implement Hadoop.** The composite organization required four incremental FTEs for a three-month period, equating to \$191,121.
- **Ongoing operation costs of \$2.4 million over three years.** The composite organization required eight FTEs composed of two Hadoop admins, two developers, and four data scientists to support the Hadoop environment and new business services.

Disclosures

The reader should be aware of the following:

- › The study is co-commissioned by Dell and Intel 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 Dell | Cloudera Apache Hadoop Solutions, accelerated by Intel.
- › Dell and Intel 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 provided the customer names for the interviews but did not participate in the interviews.

TEI Framework and Methodology

INTRODUCTION

From the information provided in the interviews, Forrester has constructed a Total Economic Impact (TEI) framework for those organizations considering implementing the Dell | Cloudera Apache Hadoop Solution, accelerated by Intel. The objective of the framework is to identify the cost, benefit, flexibility, and risk factors that affect the investment decision, to help organizations understand how to take advantage of specific benefits, reduce costs, and improve the overall business goals of winning, serving, and retaining customers.

APPROACH AND METHODOLOGY

Forrester took a multistep approach to evaluate the impact that the Dell | Cloudera Apache Hadoop Solution, accelerated by Intel, can have on an organization (see Figure 2). Specifically, we:

- › Interviewed Dell marketing, sales, and/or consulting personnel, along with Forrester analysts, to gather data relative to Dell's Hadoop solutions and the marketplace for Hadoop.
- › Interviewed four organizations currently using the Dell | Cloudera Apache Hadoop Solution, accelerated by Intel, to obtain data with respect to costs, benefits, and risks.
- › Designed a composite organization based on characteristics of the interviewed organizations (see Appendix A).
- › Constructed a financial model representative of the interviews using the TEI methodology. The financial model is populated with the cost and benefit data obtained from the interviews as applied to the composite organization.
- › Risk-adjusted the financial model based on issues and concerns the interviewed organizations highlighted in interviews. Risk adjustment is a key part of the TEI methodology. While interviewed organizations provided cost and benefit estimates, some categories included a broad range of responses or had a number of outside forces that might have affected the results. For that reason, some cost and benefit totals have been risk-adjusted and are detailed in each relevant section.

Forrester employed four fundamental elements of TEI in modeling the Dell | Cloudera Apache Hadoop Solution, accelerated by Intel's service: 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 [B] for additional information on the TEI methodology.

FIGURE 2
TEI Approach



Source: Forrester Research, Inc.

Analysis

COMPOSITE ORGANIZATION

For this study, Forrester conducted a total of four interviews with representatives from the following companies, which are Dell customers based in the US:

- › A data and analytics marketing organization focused on providing customer insights, predictive models, and analytics. After implementing Hadoop, it was able to create new services and products for customers that previously wouldn't have been feasible. This small business-to-business (B2B) organization employs approximately 45 employees and processes 7 billion to 10 billion transactions for customers.
- › A manufacturing execution systems company that utilizes its Hadoop platform to help customers track product quality through root cause analysis. It has 150-plus terabytes of data running across 43 nodes.
- › A retail organization with over \$30 billion in annual revenue and 200,000 employees. It is using Hadoop to learn more about its customers in order to better serve them. It has 10 to 15 Hadoop clusters, with the production cluster having 640 nodes with 5 petabytes of data.
- › A digital media services company that specializes in programmatic solutions. It has 65 nodes in production supporting 1.8 petabytes of data and exclusively utilizes Dell hardware accelerated by Intel.

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 that Forrester synthesized from these results — let's call it *The Representative Organization* — represents an organization with the following characteristics:

- › It is a US-based B2B data and analytics services company.
- › It has \$50 million in annual revenue.
- › It has 325 employees.
- › At initial implementation it had 24 nodes, growing to 48 nodes in Year 2 and 96 in Year 3.
- › Its data requirements of 300 terabytes grew to nearly 700 terabytes in Year 2 and 1,500 terabytes in Year 3.
- › Four to eight FTEs from the internal IT department support infrastructure.
- › Big data and Hadoop are critical to growing and staying competitive.

“The easiest part of this entire project was working with Dell.”

~ VP of database and technology, data and analytics marketing company

INTERVIEW HIGHLIGHTS

The Representative Organization faced similar challenges that many of the interviewed Hadoop customers had faced in their big data journeys. Initially, *The Representative Organization* knew it needed to move into the big data space to stay competitive in its industry and continue to meet its customers' needs but wasn't quite sure how. It did not have the platform to analyze and productize the insights from big data.

The Representative Organization started its big data journey by downloading a free distribution of Hadoop but didn't have the expertise internally to determine how to deploy and use it effectively. To hire Hadoop experts also proved difficult, as resources in this space were scarce. *The Representative Organization* knew it needed:

- › A Hadoop cluster to support big data analytics.
- › The ability to grow into multiple petabytes of both structured and unstructured data.
- › An optimized infrastructure architecture for performance and advanced analytics.

The Representative Organization selected the Dell | Cloudera Apache Hadoop Solution, accelerated by Intel, for its ability to provide an reference architecture that is optimized for performance and can easily scale to meet exponentially increasing data volumes to support new client analytics services.

The interviews revealed that with the Dell | Cloudera Apache Hadoop Solution, accelerated by Intel, customers:

- › **Improved time-to-value.** Dell's proven reference architecture made implementation of the Hadoop environment quick and easy. Organizations speculated that if they had tried to implement on their own, it would have taken six months longer to hire the expertise, figure out the correct configurations, and deploy the platform.
- › **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 within *The Representative Organization* that enabled it to incorporate new voluminous sources like social media. Additionally, this new storage capability allowed the organization to keep more than six months of its customers' data, which enabled its data scientists to develop new analysis and services for customers. Previously, it had to turn away customers due to the data capacity constraints. One customer's example enabled it to recoup \$15 million to \$25 million of revenue on an annual basis. The customer, a manufacturer, had many different parts and many different vendors that went into its finished good. When the final products failed inspection, historically it would take up to three weeks to identify the issue causing delays in production and lost sales. With *The Representative Organization's* Hadoop solution, the customer could now identify the faulty part in hours and work with its vendor to not only get new working parts but recover some losses as well.
- › **Reduced time for business intelligence.** With Hadoop's ability to process large amounts of data quickly, the retail organization we interviewed explained how crucial this was in generating reports for business leaders. With its previous business intelligence solution, the CEO would have to wait 10 minutes for a critical report to return results. 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. 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. Our composite organization was able to provide better customer response times, helping to retain customers and grow its business.

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

~ Principal architect, MES company

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

~ Architect, Fortune 500 retailer

- › **Delivered Dell and Intel hardware optimized to offer high-level performance.** The Dell | Cloudera Apache Hadoop Solution, accelerated by Intel, has enabled organizations to gain high performance from the initial deployment. Said one principal architect, “Because they [Dell] did the due diligence and because they partnered with Cloudera and because they understood what in fact works, and 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 and that was huge.” Additional performance benefits were realized in using Dell OpenManager 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.”

BENEFITS

The composite organization created for this study, *The Representative Organization*, experienced a number of quantified benefits in this case study:

- › Faster time-to-value.
- › Significant revenue increases from new client services.
- › Measureable savings on legacy hardware.
- › Increased operational cost savings.

Another important benefit mentioned by *The Representative Organization* was the ability to utilize Hadoop internally to better understand its customers. To do this, it built a research cluster to identify usage patterns and types of queries being run and to understand system failures and response times. It was then able to compile the information to help 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.



Faster Time-To-Value

The Representative Organization indicated that a key benefit delivered when using the reference architecture to implement the Dell Hadoop Solution was faster and easier deployment. Without having the in-house expertise in Hadoop, it could have taken six months or more to build and test a Hadoop solution and get into production. The Dell | Cloudera Apache Hadoop Reference Architecture provided not only a faster time-to-value but was a proven configuration that optimized performance for *The Representative Organization*.

Once in production, Hadoop was able to deliver \$10 million to \$20 million in new client services revenue per year. Being able to accelerate the time to realize that revenue by six months allowed *The Representative Organization* to capture an extra \$800,000 per month, or \$5 million of revenue in the first year of use. With a 20% profit margin rate, this equated to \$1 million of incremental income. See Table 1 for the detailed calculation.

Interviewed organizations provided a broad range of initial in-house expertise and margin rates. To compensate, this benefit was risk-adjusted and reduced by 10%. The risk-adjusted total benefit resulting from quicker time-to-value in the first year was \$900,000, or about \$9,375 per node. See the section on Risks for more detail.

TABLE 1
Faster Time-To-Value

Ref.	Metric	Calculation	Initial	Year 1	Year 2	Year 3
A1	Time-to-value increase	Months		6		
A2	Average new services revenue			\$10,000,000		
A3	Profit margin rate			20%		
At	Faster time-to-value	$(A1/12)*A2*A3$	\$0	\$1,000,000	\$0	\$0
	Risk adjustment	↓ 10%				
Atr	Faster time-to-value margin (risk-adjusted)		\$0	\$900,000	\$0	\$0

Source: Forrester Research, Inc.



Revenue From New Business

Hadoop created a new way to store, manage, access, and analyze data for *The Representative Organization* and its customers. *The Representative Organization* could now accept jobs that previously had to be turned down because of timing requirements and data storage needs. This helped the organization access new markets and customers, ultimately leading to an increase in revenue and income. Additionally, this new homogenous Dell environment allowed *The Representative Organization* to easily scale with the growth of its new business.

Through increasing customer satisfaction, offering faster processing times and unlimited data storage, *The Representative Organization* was able to realize \$10 million of incremental revenue in Year 2, growing to \$20 million in Year 3. Applying its 20% profit margin rate, it realized \$2 million in Year 2 and \$4 million in Year 3 in incremental income.

Interviewed organizations provided a broad range of revenue from new business examples, since there are a variety of outside forces that might also have an impact on this. To compensate, this benefit was risk-adjusted and reduced by 20%. The risk-adjusted total benefit resulting from new business revenue over the three years was \$3,726,521, or about \$38,818 per node. See the section on Risks for more detail.

TABLE 2
Revenue From New Business

Ref.	Metric	Calculation	Initial	Year 1	Year 2	Year 3
B1	Average new business revenue				\$10,000,000	\$20,000,000
B2	Profit margin rate				20%	20%
Bt	Income from new business	B1*B2	\$0	\$0	\$2,000,000	\$4,000,000
	Risk adjustment	↓ 20%				
Btr	Revenue from new business (risk-adjusted)		\$0	\$0	\$1,600,000	\$3,200,000

Source: Forrester Research, Inc.



Legacy Hardware Savings

The composite organization, *The Representative Organization*, indicated that another key benefit from the Hadoop implementation was a reduction in its database cost per terabyte. Prior to Hadoop, *The Representative Organization* had a relational database and data warehouse platform serving its needs. With Dell's Hadoop solution, the overall cost per terabyte (including triple redundancy requirements) was one-twelfth the cost of its legacy platform. As a result, *The Representative Organization* was able to avoid growing its legacy system as new business was being generated. However, in that new business, data needs were delivered in Hadoop, and we assumed that two-thirds of that new data demand would not have existed in the legacy environment leaving only one-third to be avoided (C3). This avoidance is captured in Table 3

The Representative Organization's legacy database cost per terabyte was \$18,500. Given the size of the Hadoop environment excluding triple redundancy, *The Representative Organization* was able to avoid 304 TB, 384 TB, and 768 TB of database storage in years 1, 2, and 3, respectively. Assuming one-third of the new data demand was organic and not driven by Hadoop, the overall legacy hardware avoidance was \$7,147,299 over three years.

The interviewed organizations provided a broad range of legacy costs per terabyte. Since there are many variables that might also have an impact on this benefit, to compensate, it was risk-adjusted and reduced by 10%. The risk-adjusted total benefit resulting from legacy hardware savings over the three years was \$6,432,569, or about \$67,006 per node. See the section on Risks for more detail.

TABLE 3
Legacy Hardware Savings

Ref.	Metric	Calculation	Initial	Year 1	Year 2	Year 3
C1	Legacy database cost per terabyte			\$18,500	\$18,500	\$18,500
C2	Hadoop terabytes (excludes double replication)			304	384	768
C3	Legacy database avoidance percentage			33%	33%	33%
Ct	Legacy hardware savings	$C1 * C2 * C3$	\$0	\$1,855,920	\$2,344,320	\$4,688,640
	Risk adjustment	↓ 10%				
Ctr	Legacy hardware savings (risk-adjusted)		\$0	\$1,670,328	\$2,109,888	\$4,219,776

Source: Forrester Research, Inc.



Operational Cost Savings

Another benefit *The Representative Organization* realized was increased operational efficiency from the implementation of the Dell | Cloudera Apache Hadoop Solution. Admins could now support much larger data sets requiring much less time. Our composite organization was able to repurpose one FTE in Year 1; three FTEs in Year 2; and five FTEs in Year 3. This savings was realized as data shifted into the Hadoop environment and the legacy platform was repurposed. On average, an admin FTE was paid \$70,000 per year, equating to \$500,150 of savings over three years. See Table 4 for the detailed calculation.

Interviewed organizations provided a broad range of operational efficiency. To compensate, this benefit was risk-adjusted and reduced by 10%. The risk-adjusted total benefit resulting from legacy hardware savings over the three years was \$450,135, or about \$4,689 per node. See the section on Risks for more detail.

TABLE 4
Operational Cost Savings

Ref.	Metric	Calculation	Initial	Year 1	Year 2	Year 3
D1	Number of FTEs repurposed			1	3	5
D2	Average cost per FTE			70,000	70,000	70,000
Dt	Operational cost savings	D1*D2	\$0	\$70,000	\$210,000	\$350,000
	Risk adjustment	↓ 10%				
Dtr	Operational cost savings (risk-adjusted)		\$0	\$63,000	\$189,000	\$315,000

Source: Forrester Research, Inc.

Total Benefits

Table 5 shows the total of all benefits across the four areas listed above, 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 \$11 million, or \$119,035 per node.

TABLE 5
Total Benefits (Risk-Adjusted)

Ref.	Benefit Category	Initial	Year 1	Year 2	Year 3	Total	Present Value
Atr	Faster time-to-value	\$0	\$900,000	\$0	\$0	\$900,000	\$818,182
Btr	Revenue from new business	\$0	\$0	\$1,600,000	\$3,200,000	\$4,800,000	\$3,726,521
Ctr	Legacy hardware savings	\$0	\$1,670,328	\$2,109,888	\$4,219,776	\$7,999,992	\$6,432,569
Dtr	Operational cost savings	\$0	\$63,000	\$189,000	\$315,000	\$567,000	\$450,135
	Total benefits (risk-adjusted)	\$0	\$2,633,328	\$3,898,888	\$7,734,776	\$14,266,992	\$11,427,408

Source: Forrester Research, Inc.

COSTS

The composite organization, *The Representative Organization*, experienced a number of costs associated with Dell's Hadoop solution, including:

- › **Software license cost.** The cost to license the Cloudera Hadoop distribution.
- › **Hardware cost.** The cost of the nodes and cabling.
- › **Implementation cost.** Dell professional services to help implement Hadoop.
- › **Hadoop operational cost.** Incremental FTEs to support the Hadoop environment.

These represent the mix of internal and external costs experienced by the *The Representative Organization* for initial planning, implementation, and ongoing maintenance associated with the solution. Please note that the license and hardware costs in this study represent list prices provided by Dell and do not take into consideration licensing agreements or other discounts that may apply.



Software Licensing Cost

Software licensing fees for Cloudera distribution of Hadoop were incurred during the initial implementation period and in subsequent years. The license cost is priced by node and is approximately \$7,168 per node. With 24 nodes initially, and growing to 48 and 96 in years 2 and 3, respectively, *The Representative Organization* is paying \$973,383 over three years.

Software costs vary from organization to organization, considering different licensing agreements and other discounts. To compensate, this cost was risk-adjusted up by 5%. The risk-adjusted cost of software over the three years was \$1,022,052. See the section on Risks for more detail.

TABLE 6
Software Licensing Cost

Ref.	Metric	Calculation	Initial	Year 1	Year 2	Year 3
E1	Cloudera license cost per node		\$7,168		\$7,168	\$7,168
E2	Initial infrastructure nodes		5		5	5
E3	Data nodes		19		43	91
E4	Total cumulative nodes	E2+E3	24		48	96
Et	Software license cost	E1*E4	\$172,032	\$0	\$344,064	\$688,128
	Risk adjustment	↑ 5%				
Etr	Software license cost (risk-adjusted)		\$180,634	\$0	\$361,267	\$722,534

Source: Forrester Research, Inc.



Hardware Cost

The Representative Organization worked with Dell to purchase the 730xd hardware for its Hadoop nodes. Initially, five infrastructure nodes were required at a cost of \$16,723 each, and an additional 19 data nodes were purchased at \$26,658 each. It's helpful to note that Hadoop architecture requires triple redundancy in data storage with an estimated capacity of 48 terabytes per node. Therefore, the initial deployment had 304 operable terabytes of storage. As data needs grew with the new business demand, so did the number of nodes required. Over three years, *The Representative Organization* acquired a total of 96 nodes with a total storage capacity of 4,608 terabytes or 1.5 petabytes of operable storage. Over three years, the composite organization paid \$2,080,241.

For hardware costs, we used Dell's retail pricing, which doesn't take into account any discounts or packaging. Since these costs may vary from organization to organization, they were risk-adjusted up by 5%. The risk-adjusted cost of hardware over the three years was \$2,184,253, or \$22,753 per node. See the section on Risks for more detail.

TABLE 7
Hardware Cost

Ref.	Metric	Calculation	Initial	Year 1	Year 2	Year 3
F1	Initial infrastructure nodes	5 nodes @ \$16,723	\$83,615			
F2	Data node cost per terabyte	\$26,658 per data node	\$555		\$555	\$555
F3	Number of terabytes	48 terabytes per node	912		1,152	2,304
Ft	Hardware cost	$F1+F2 \cdot F3$	\$590,117	\$0	\$639,792	\$1,279,584
	Risk adjustment	↑ 5%				
Ftr	Hardware cost (risk-adjusted)		\$619,623	\$0	\$671,782	\$1,343,563

Source: Forrester Research, Inc.



Implementation Cost

Initial implementation costs included Dell 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 three months in total and cost \$173,746 overall.

Implementation costs and timelines can vary from organization to organization; to compensate, this cost was risk-adjusted up by 10%. The risk-adjusted cost of implementing Hadoop was \$191,121. See the section on Risks for more detail.

TABLE 8
Implementation Cost

Ref.	Metric	Calculation	Initial	Year 1	Year 2	Year 3
G1	Dell professional services		\$23,746			
G2	Number of internal FTEs		4			
G3	Hourly rate per FTE		\$75			
G4	Hours		500			
Gt	Implementation cost	$G1+G2*G3*G4$	\$173,746	\$0	\$0	\$0
	Risk adjustment	↑ 10%				
Gtr	Implementation cost (risk-adjusted)		\$191,121	\$0	\$0	\$0

Source: Forrester Research, Inc.



Hadoop Operational Cost

The Representative Organization needed to hire one admin, one developer, and two data scientists to support the initial deployment of Hadoop. As business grew in Year 2, it hired another developer and data scientist. By Year 3, it had to hire one more admin and one more data scientist, totaling two admins, two developers, and four data scientists in the third year. The average fully loaded hourly rate for these resources was \$75 per hour, equating to \$600,000 of costs in Year 1; \$900,000 in Year 2; and \$1,200,000 in Year 3. See Table 9 for calculation details.

The number of resources required and the mix between internal versus external resources and onshore versus offshore resources can change from organization to organization. To compensate, this cost was risk-adjusted up by 10%. The risk-adjusted cost of operating Hadoop was \$2,409,917 over three years. See the section on Risks for more detail.

TABLE 9
Hadoop Operational Cost

Ref.	Metric	Calculation	Initial	Year 1	Year 2	Year 3
H1	Number of FTEs			4	6	8
H2	Hourly rate per FTE			\$75	\$75	\$75
H3	hours per year			2,000	2,000	2,000
Ht	Hadoop operational cost	$H1*H2*H3$	\$0	\$600,000	\$900,000	\$1,200,000
	Risk adjustment	↑ 10%				
Htr	Hadoop operational cost (risk-adjusted)		\$0	\$660,000	\$990,000	\$1,320,000

Source: Forrester Research, Inc.

Total Costs

Table 10 shows the total of all costs as well as associated present values, discounted at 10%. Over three years, the composite organization's total present value of costs was a little more than \$5.8 million, or \$60,493 per node.

TABLE 10
Total Costs (Risk-Adjusted)

Ref.	Cost Category	Initial	Year 1	Year 2	Year 3	Total	Present Value
Etr	Software license cost	\$180,634	\$0	\$361,267	\$722,534	\$1,264,435	\$1,022,052
Ftr	Hardware costs	\$619,623	\$0	\$671,782	\$1,343,563	\$2,634,968	\$2,184,253
Gtr	Implementation costs	\$191,121	\$0	\$0	\$0	\$191,121	\$191,121
Htr	Hadoop operational cost	\$0	\$660,000	\$990,000	\$1,320,000	\$2,970,000	\$2,409,917
Total costs (risk-adjusted)		\$991,377	\$660,000	\$2,023,049	\$3,386,098	\$7,060,523	\$5,807,343

Source: Forrester Research, Inc.

FLEXIBILITY

Flexibility, as defined by TEI, represents an investment in additional capacity or capability that could be turned into business benefit for some 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. There are multiple scenarios in which a customer might choose to implement the Dell | Cloudera Apache Hadoop Reference Architecture, accelerated by Intel, and later realize additional uses and business opportunities. Flexibility would also be quantified when evaluated as part of a specific project (described in more detail in Appendix B).

For organizations wanting to adopt the Dell | Cloudera Apache Hadoop Reference Architecture and solve data integration and transformation constraints in their legacy relational database or enterprise data warehouse, Dell offers another alternative. The Dell | Cloudera | Syncsort Data Warehouse Optimization — ETL Offload solution, which includes Syncsort DMX-h, is a reference architecture that helps customers augment their legacy data warehouse by providing an initial use case for running ETL jobs in Cloudera Enterprise Hadoop. Organizations that have done this recognized additional infrastructure and licensing cost avoidance, improved SLAs for business reporting in legacy data warehouses, and simplified ongoing ETL operations. One interviewed customer said, “Eighty percent of our workload is transforming the data into the format that the end user wanted.” With the Dell | Cloudera | Syncsort Data Warehouse Optimization — ETL Offload solution, these jobs can now be completed in a fraction of the time while freeing up expensive capacity in legacy databases. This interviewee went on to say, “Although we didn’t retire the mainframe, we reduced MIPS from the mainframe and provided the business with faster results.”

RISKS

Forrester defines two types of risk associated with this analysis: “implementation risk” and “impact risk.” Implementation risk is the risk that a proposed investment in the Dell | Cloudera Apache Hadoop Reference Architecture, accelerated by Intel, may deviate from the original or expected requirements, resulting in higher costs than anticipated. Impact risk refers to the

risk that the business or technology needs of the organization may not be met by the investment in the Dell | Cloudera Apache Hadoop Reference Architecture, accelerated by Intel, resulting in lower overall total benefits. The greater the uncertainty, the wider the potential range of outcomes for cost and benefit estimates.

TABLE 11
Benefit And Cost Risk Adjustments

Benefits	Adjustment
Faster time-to-value	↓ 10%
Revenue from new business	↓ 20%
Legacy hardware savings	↓ 10%
Operational cost savings	↓ 10%
Costs	Adjustment
Software license cost	↑ 5%
Hardware cost	↑ 5%
Implementation cost	↑ 10%
Hadoop operational cost	↑ 10%

Source: Forrester Research, Inc.

Quantitatively capturing implementation risk and impact risk by directly adjusting the financial estimates results provides more meaningful and accurate estimates and a more accurate projection of the ROI. In general, risks affect costs by raising the original estimates, and they affect benefits by reducing the original estimates. The risk-adjusted numbers should be taken as “realistic” expectations since they represent the expected values considering risk.

The following impact risks that affect benefits are identified as part of the analysis:

- › Faster time-to-value may be shorter or longer based on a range of in-house expertise and knowledge within an organization. Additionally, the complexity of the use cases required for deployment can also lengthen the time-to-value.
- › Revenue from new business could be different from organization to organization based on many outside variables including product quality, customer service, and pricing. Additionally, there are many external economic factors that might increase or decrease the magnitude of revenue growth realized.
- › Legacy hardware savings for many companies may differ based on the amount of inorganic versus organic growth, the capabilities and configuration of the legacy platform, and vendor pricing of legacy hardware.
- › Operational cost savings may differ based on the organization’s skill sets and expertise.

The following implementation risks that affect costs are identified as part of this analysis:

- › License costs can change based on contract terms, deal size, and other discounts that may apply.

- › Hardware costs utilize Dell's retail price and could be affected by contract terms, deal size, and other discounts that may apply.
- › Implementation costs and resources may fluctuate based on complexity, size, and length of implementation.
- › Hadoop operational costs might differ based on support needs and analytical requirements.

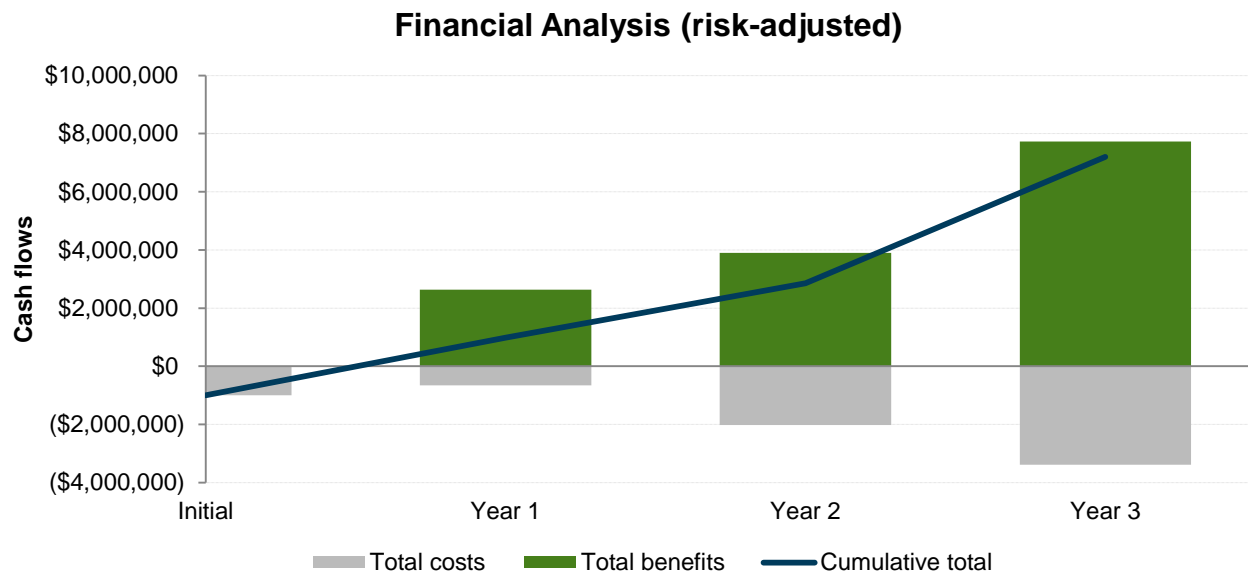
Table 11 shows the values used to adjust for risk and uncertainty in the cost and benefit estimates for the composite organization, *The Representative Organization*. Readers are urged to apply their own risk ranges based on their own degree of confidence in the cost and benefit estimates.

Financial Summary

The financial results calculated in the Benefits and Costs sections can be used to determine the ROI, NPV, and payback period for the *The Representative Organization's* investment in the Dell | Cloudera Apache Hadoop Reference Architecture, accelerated by Intel.

Table 12 below shows the risk-adjusted ROI, NPV, and payback period values. These values are determined by applying the risk-adjustment values from Table 11 in the Risks section to the unadjusted results in each relevant cost and benefit section.

FIGURE 3
Cash Flow Chart (Risk-Adjusted)



Source: Forrester Research, Inc.

TABLE 12
Cash Flow (Risk-Adjusted)

Summary	Initial	Year 1	Year 2	Year 3	Total	Present Value
Total costs	(\$991,377)	(\$660,000)	(\$2,023,049)	(\$3,386,098)	(\$7,060,523)	(\$5,807,343)
Total benefits	\$0	\$2,633,328	\$3,898,888	\$7,734,776	\$14,266,992	\$11,427,408
Total	(\$991,377)	\$1,973,328	\$1,875,839	\$4,348,678	\$7,206,469	\$5,620,064
ROI						97%
Payback period						Six months

Source: Forrester Research, Inc.

Dell | Cloudera Apache Hadoop Reference Architecture, Accelerated By Intel: Overview

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

DELL | CLOUDERA APACHE HADOOP SOLUTION: OVERVIEW

With the explosive growth in data volumes and complexity, organizations of all sizes are turning to the open source Apache Hadoop platform to store, process, and generate value from their data. Hadoop solutions are not just about being able to capture data; they are also about being able to work with the many new and different varieties of unstructured data — social media data, sensor data, machine-generated data, and more.

The Dell™ | Cloudera™ Apache™ Hadoop® Solution, accelerated by Intel, was jointly designed by Dell, Cloudera, and Intel to lower the barriers to adoption for organizations considering Hadoop. This end-to-end solution approach reduces time to value compared to an Open Source do-it-yourself approach.

There are many advantages to using Hadoop, particularly in scalability, flexibility and economics. And, without guidelines, as with any open source technology, it presents a unique set of challenges when deployed into production. Installing, configuring, and running a production Hadoop cluster involves multiple considerations, including:

- › The appropriate Hadoop software distribution and extensions
- › Monitoring and management software
- › Allocation of Hadoop services to physical nodes
- › Selection of appropriate server hardware
- › Design of the network fabric
- › Sizing and scalability
- › Performance

These considerations are complicated by the need to understand the type of workloads that will be running on the cluster, the fast-moving pace of the core Hadoop project and the challenges of managing a system designed to scale to thousands of nodes in a single instance.

To address the challenges associated with Hadoop implementations, Dell, Cloudera and Intel deliver a tested, validated and proven reference architecture that outlines the design of an end-to-end Hadoop solution for organizations who need to tackle big-data challenges for production deployments. The solution includes components that span the entire solution stack including:

- › Optimized server configurations using the Dell PowerEdge R730xd Server and Intel Xeon Processors
- › Optimized network infrastructure based on Dell Force 10 Network Switches
- › Cloudera Distribution for Apache Hadoop
- › Detailed reference architecture guide
- › Detailed deployment guide and deployment tools.

SOLUTION USE CASE SUMMARY

The Dell | Cloudera Apache Hadoop Solution, 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 semi structured 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 semi-structured 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 semi- structured data.
Data archive	Active archival of medium-term (12–36 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 semi structured big data in real time.
Search and predictive analytics	Crawl, extract, index and transform semi structured and unstructured data for search and predictive analytics.

SOLUTION COMPONENTS

The following figure illustrates the primary components in the Dell | Cloudera Apache Hadoop Solution.

Sitting on top of these storage layers are four complementary access layers providing data processing, in-memory processing, data query and data search:

› **Data processing:** MapReduce is the core processing framework in the Hadoop system, and provides a massively parallel data processing framework inspired by Google's MapReduce papers.

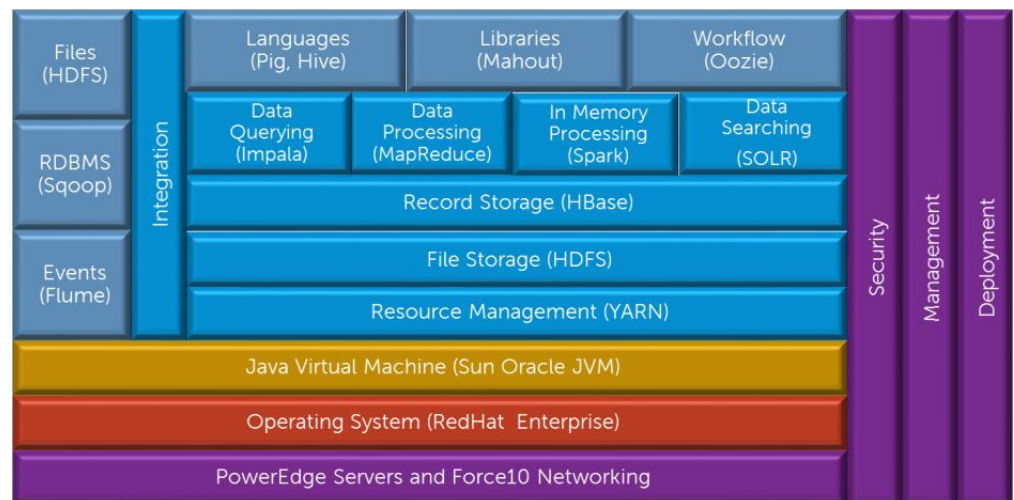
› **In-memory processing:** Another processing framework is the real-time, in-memory processing framework called Spark.

› **Data query:** The Data Query layer provides real-time query access to data using Cloudera Impala.

› **Data search:** The Data Search layer provides real-time search of indexed data using Apache SOLR Cloud technology.

All four of these layers can be used simultaneously or independently, depending on the workload and challenges being solved.

For additional information visit: Dell.com/Hadoop



Appendix A: Composite Organization Description

For this TEI study, Forrester has created a composite organization to illustrate the quantifiable benefits and costs of implementing the Dell | Cloudera Apache Hadoop Reference Architecture, accelerated by Intel. The composite company is intended to represent a data and analytics services organization and is based on characteristics of the interviewed customers. For this study we have named the composite organization *The Representative Organization*. It has 325 employees and generates \$50 million of revenue annually.

The data needs of the composite company were 300 TB, 700 TB, and 1,500 TB in years 1, 2, and 3, respectively. Based on the hardware configuration, this equated to 24 nodes in Year 1 and grew to 96 nodes by Year 3.

In purchasing the Dell | Cloudera Apache Hadoop Reference Architecture, accelerated by Intel, the composite company has the following objectives:

- › Create a platform that could be easily scaled and grow with the business.
- › Optimize architecture and increase performance.
- › Gain the ability to manage multiple petabytes of unstructured data.
- › Maintain customer relevance and stay competitive.

FRAMEWORK ASSUMPTIONS

The discount rate used in the PV and NPV calculations is 10%, and the time horizon used for the financial modeling is three years. Organizations typically use discount rates between 8% and 16% based on their current environment. Readers are urged to consult with their respective company's finance department to determine the most appropriate discount rate to use within their own organizations.

Appendix B: Total Economic Impact™ Overview

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. TEI assists technology vendors in winning, serving, and retaining customers.

The TEI methodology consists of four components to evaluate investment value: benefits, costs, flexibility, and risks.

BENEFITS

Benefits represent the value delivered to the user organization — IT and/or business units — by the proposed product or project. Often, product or project justification exercises focus just on IT cost and cost reduction, leaving little room to analyze the effect of the technology on the entire organization. The TEI methodology and the resulting financial model place 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. Calculation of benefit estimates involves a clear dialogue with the user organization to understand the specific value that is created. In addition, Forrester also requires that there be a clear line of accountability established between the measurement and justification of benefit estimates after the project has been completed. This ensures that benefit estimates tie back directly to the bottom line.

COSTS

Costs represent the investment necessary to capture the value, or benefits, of the proposed project. IT or the business units may incur costs in the form of fully burdened labor, subcontractors, or materials. Costs consider all the investments and expenses necessary to deliver the proposed value. In addition, the cost category within TEI captures any incremental costs over the existing environment for ongoing costs associated with the solution. All costs must be tied to the benefits that are created.

FLEXIBILITY

Within the TEI methodology, direct benefits represent one part of the investment value. While direct benefits can typically be the primary way to justify a project, Forrester believes that organizations should be able to measure the strategic value of an investment. Flexibility represents the value that can be obtained for some future additional investment building on top of the initial investment already made. For instance, an investment in an enterprisewide upgrade of an office productivity suite can potentially increase standardization (to increase efficiency) and reduce licensing costs. However, an embedded collaboration feature may translate to greater worker productivity if activated. The collaboration can only be used with additional investment in training at some future point. However, having the ability to capture that benefit has a PV that can be estimated. The flexibility component of TEI captures that value.

RISKS

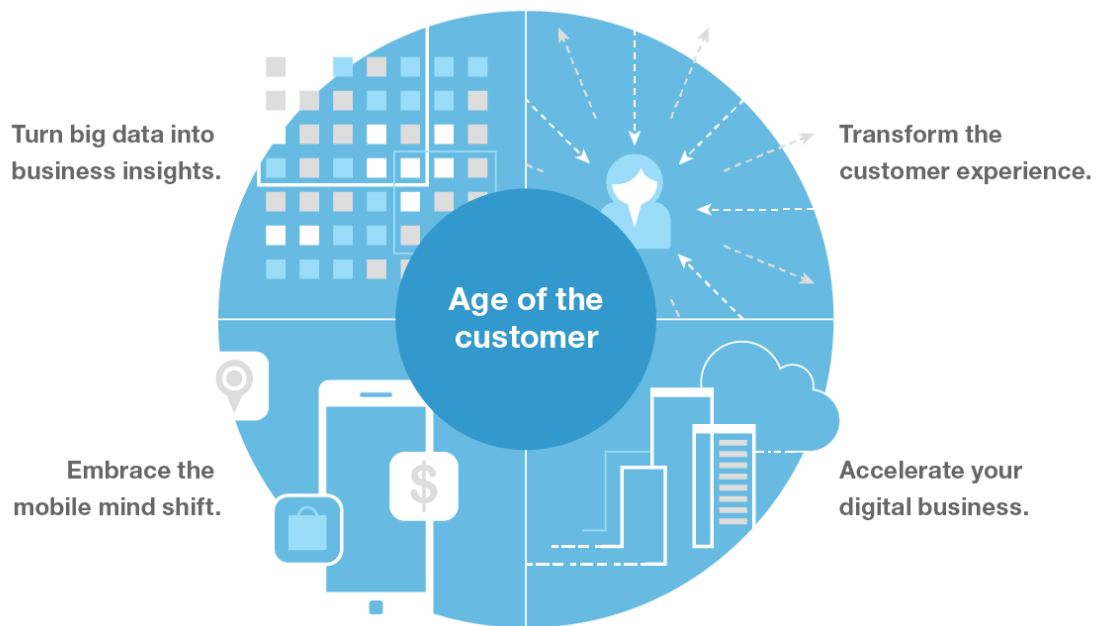
Risks measure the uncertainty of benefit and cost estimates contained within the investment. Uncertainty is measured in two ways: 1) the likelihood that the cost and benefit estimates will meet the original projections and 2) the likelihood that the estimates will be measured and tracked over time. TEI risk factors are based on a probability density function known as "triangular distribution" to the values entered. At a minimum, three values are calculated to estimate the risk factor around each cost and benefit.

Appendix C: Forrester And The Age Of The Customer

Your technology-empowered customers now know more than you do about your products and services, pricing, and reputation. Your competitors can copy or undermine the moves you take to compete. The only way to win, serve, and retain customers is to become customer-obsessed.

A customer-obsessed enterprise focuses its strategy, energy, and budget on processes that enhance knowledge of and engagement with customers and prioritizes these over maintaining traditional competitive barriers.

CMOs and CIOs must work together to create this companywide transformation.



Forrester has a four-part blueprint for strategy in the age of the customer, including the following imperatives to help establish new competitive advantages:



Transform the customer experience to gain sustainable competitive advantage.



Accelerate your digital business with new technology strategies that fuel business growth.



Embrace the mobile mind shift by giving customers what they want, when they want it.



Turn (big) data into business insights through innovative analytics.

Appendix D: Glossary

Discount rate: The interest rate used in cash flow analysis to take into account the time value of money. Companies set their own discount rate based on their business and investment environment. Forrester assumes a yearly discount rate of 10% for this analysis. Organizations typically use discount rates between 8% and 16% based on their current environment. Readers are urged to consult their respective organizations to determine the most appropriate discount rate to use in their own environment.

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.

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.

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.

Return on investment (ROI): A measure of a project's expected return in percentage terms. ROI is calculated by dividing net benefits (benefits minus costs) by costs. A 100% ROI means the return or benefits are 2x larger than the costs.

A NOTE ON CASH FLOW TABLES

The following is a note on the cash flow tables used in this study (see the example table below). The initial investment column contains costs incurred at "time 0" or at the beginning of Year 1. Those costs are not discounted. All other cash flows in years 1 through 3 are discounted using the discount rate (shown in the Framework Assumptions section) at the end of the year. PV calculations are calculated for each total cost and benefit estimate. NPV calculations are not calculated until 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.

TABLE [EXAMPLE]
Example Table

Ref.	Metric	Calculation	Year 1	Year 2	Year 3

Source: Forrester Research, Inc.

Appendix E: Endnotes

¹ Forrester risk-adjusts the summary financial metrics to take into account the potential uncertainty of the cost and benefit estimates. For more information, see the section on Risks.