



Research®

Advisory

**451
ADVISORS**

FEB 2016

The Cloud-Based Approach to Achieving Business Value From Big Data

Based on Research Commissioned by





ABOUT 451 RESEARCH

451 Research is a preeminent information technology research and advisory company. With a core focus on technology innovation and market disruption, we provide essential insight for leaders of the digital economy. More than 100 analysts and consultants deliver that insight via syndicated research, advisory services and live events to over 1,000 client organizations in North America, Europe and around the world. Founded in 2000 and headquartered in New York, 451 Research is a division of The 451 Group.

ABOUT THIS RESEARCH

451 conducted six in depth interviews with Big Data users whose primary deployment is cloud based. These users were sourced from customer lists provided by Amazon Web Services (AWS). 451 Research analyzed the results of these interviews and, combined with our ongoing customer and market analysis, created this report.

© 2015 451 Research, LLC and/or its Affiliates. All Rights Reserved. Reproduction and distribution of this publication, in whole or in part, in any form without prior written permission is forbidden. The terms of use regarding distribution, both internally and externally, shall be governed by the terms laid out in your Service Agreement with 451 Research and/or its Affiliates. The information contained herein has been obtained from sources believed to be reliable. 451 Research disclaims all warranties as to the accuracy, completeness or adequacy of such information. Although 451 Research may discuss legal issues related to the information technology business, 451 Research does not provide legal advice or services and their research should not be construed or used as such. 451 Research shall have no liability for errors, omissions or inadequacies in the information contained herein or for interpretations thereof. The reader assumes sole responsibility for the selection of these materials to achieve its intended results. The opinions expressed herein are subject to change without notice.

NEW YORK

20 West 37th Street
3rd Floor
New York, NY 10018
P 212-505-3030
F 212-505-2630

SAN FRANCISCO

140 Geary Street
9th Floor
San Francisco, CA 94108
P 415-989-1555
F 415-989-1558

LONDON

37-41 Gower Street
London, UK WC1E 6HH
P +44 (0)20 7299 7765
F +44 (0)20 7299 7799

BOSTON

1 Liberty Square,
5th Floor
Boston, MA 02109
P 617-261-0699
F 617-261-0688

Executive Summary

Ongoing developments in cloud computing are rapidly moving the promise of deriving business value from big data into reality. With billions of devices globally already streaming data, forward-leaning companies have begun leveraging the cloud to harvest big benefits from this data storm. These benefits include faster time to market of products and services; lower total cost of ownership (TCO) and reduced management overhead costs compared to on-premises infrastructure; and more agile application development.

451 Research conducted in-depth interviews with six organizations to examine their cloud-based big-data adoption efforts and the various tools and services they use. These interviews unearthed several important factors that can be useful to other organizations seeking to generate actionable business insights from their data. Included among them are:

- A broad array of organizations can realize the benefits of big data. In this report alone, the enterprises interviewed represent a variety of industries: a mobile technology analytics platform provider; a mobile application platform provider; a financial services regulator; a technology consultancy; a marketing strategy firm; and a mainstream financial services firm.
- Underpinning all these early successful efforts to gain business value from big data is the ability to capture, store and process cloud data far less expensively than could ever be done in on-premises environments. This has allowed users to take advantage of advanced analytics techniques, especially the ability to analyze unstructured data to exploit the companies' stored information.
- Remarkably, after relatively brief forays into cloud-based utilization of big data, some organizations find that this move is not just incremental to the business, but rather, it becomes the heart of the business.
- Benefits of cloud-based big data are often measurable. In one organization, data queries to the cloud showed an impressive 400-fold improvement over on-premises-based queries. Another company witnessed the time to execute a business-critical risk-threat analysis drop from 6-9 months to a week or less - a 98% improvement over on-premises systems.

The interviewees also cited potential challenges and roadblocks to cloud-based big data. While the cloud significantly lowers the cost and other barriers for big-data processing (such as the time and effort to deploy the required infrastructure), big-data cloud users still face some of the same challenges found in on-premises deployments. The challenge most often mentioned is the lack of personnel with the appropriate skill sets in this nascent area. To get the most value out of cloud-based big data quickly, developing a training plan is recommended.

5 Ways the Cloud Delivers for Big-Data Users

All six enterprises that contributed to this report are using the cloud for big-data processing to increase competitiveness. They are doing so by providing products or services to their customers that would not have been possible or affordable using traditional data-processing and analysis techniques or deployment models.

The study identified five fundamental ways that the combination of cloud computing and big-data analytics are delivering business value:

- 1. Faster time to market:** Query times for one organization improved by 4,000% over on-premises systems, enabling far more rapid and informed decision-making while shrinking product and service development time.
- 2. Lower TCO:** >50% cost savings over on-premises environments.
- 3. Reduced management overhead:** >50% reduction in operating costs, freeing up valuable IT staff time for more strategic, business-focused development tasks.
- 4. Improved developer agility:** From concept to full production deployment in 24 hours.
- 5. New revenue opportunities:** Faster, easier access to analytics is helping to uncover new revenue opportunities in minutes, not days.

INTRODUCTION

451 Research defines big data as the realization of new business intelligence insights driven by the fact that it is now more economically feasible to store and process data that was previously ignored due to the cost and functional limitations of traditional data management technologies. This includes existing data sources such as Web, network and server log data, as well as new data sources: machine-generated and social media data.

Technology historian George Dyson put it more succinctly when he stated, “Big data is what happened when the cost of keeping information became less than the cost of throwing it away.” Whether this statement is true is perhaps open to debate. What is not debatable is that the cost of storing and processing data has declined dramatically in recent years, enabling enterprises to take advantage of more advanced data-analytics techniques to deliver greater business intelligence and competitive advantage.

Cloud computing has had a significant role to play in driving down the cost of storing and processing data. However, this is just one of the potential benefits of the cloud, alongside developer/business agility, faster time to adoption for emerging technologies, high availability and reduced infrastructure configuration/management overhead. In recent years, there has been an explosion in terms of the number of approaches for storing, processing and analyzing data in cloud and hosted services.

Cloud adoption is widespread, and the emergence of big data is driving more enterprises to look for new approaches to generate insight from data that was previously ignored or simply could not be effectively or efficiently processed. Cloud services provide an alternative to on-premises infrastructure for the storage and processing of data sets both large and small. Furthermore, these services can lower the barriers to adoption by reducing the need for large capital investments, enabling pay-as-you-go models and seamlessly facilitating global expansion. As a result, enterprises are able to reduce the cost and time involved in configuring new big-data processing and analytics platforms.

This report provides an overview of the available offerings and adoption trends, driven by a customer perspective generated from direct interaction with enterprise stakeholders and decision-makers. 451 Research conducted six in-depth interviews with enterprises that are utilizing cloud services to:

- Fulfill their big-data objectives, identifying business objectives and success metrics.
- Determine current barriers to meeting those objectives.
- Identify the role cloud service providers can play in helping meet those objectives.

Table of Contents



CLOUD-BASED BIG DATA: BUSINESS VALUE RAISED BY THE POWER OF 2	1
BUSINESS OBJECTIVES FOR BIG DATA.	1

USERS AND USE CASES	3
DATA-DRIVEN APPLICATIONS AND SERVICES	3
BUSINESS REPORTING AND ADVANCED ANALYTICS	3
DATA SCIENCE	4

CHALLENGES – TECHNICAL, PROCESS, BUSINESS	5
<i>Potential Challenges to Big Data Projects:</i>	6

MEASURING SUCCESS	7
FASTER TIME TO MARKET	7
TOTAL COST OF OWNERSHIP	8
BUSINESS PERFORMANCE	9
AGILITY	9

EXPECTATIONS	10
---------------------	-----------

Cloud-Based Big Data: Business Value Raised by the Power of 2

While adoption of cloud computing is historically associated with cutting-edge startups, 451 Research has found that enterprises across all industries are investigating and adopting cloud services. This report provides a snapshot of a group of companies that have already adopted cloud services for their big-data initiatives – storing, processing and analyzing data that was previously ignored due to incompatibility with traditional on-premises infrastructure, as well as traditional data storage and processing techniques.

This incompatibility is due to a number of factors:

- Relational databases were not designed for storing and processing unstructured data.
- Relational databases are unsuited to processing large-scale data volumes.
- Relational databases are better suited to processing predefined queries, rather than enabling data analysts and data scientists to explore data sets in search of new insight.

However, as noted above, we believe the primary factor driving adoption of emerging data-processing technologies is an economic one: even where it might be technically possible for traditional architectures to store and process big data, the cost of doing so at scale is prohibitive.

Cloud services have the potential to lower the barriers to adoption for these big-data-processing platforms by removing the requirement to build, configure and manage on-premises infrastructure. This enables developers to move from proof of concept to production more quickly and to provide data analysts and scientists with quicker time to business insight.

The combination of cloud services and new approaches to storing, processing and visualizing data have provided enterprises with an opportunity to drive business change and fulfill emerging business objectives. However, adoption of data processing in the cloud is still in the early stages, and the interviewees in this study also highlighted their intentions to expand their cloud-based data processing – providing a glimpse into the potential direction of big-data processing on cloud services.

Interviewees in the study used a wide range of AWS offerings, such as Amazon S3 for cloud storage, Amazon Redshift for data warehousing and Amazon's Elastic Map Reduce (EMR) – a Hadoop-as-a-service offering that is compatible and interoperable with Hive, Presto and other Hadoop-related projects. Meanwhile some respondents were also running Hadoop distributions on cloud infrastructure. Amazon's DynamoDB NoSQL database and Relational Database Service were also either being used in production or test and development, along with other cloud services.

BUSINESS OBJECTIVES FOR BIG DATA

As noted above, one of the key trends driving the adoption of big-data processing – both on-premises and in the cloud – is the ability to take advantage of inexpensive compute resources to perform data processing and analytics at a scale that was previously impossible due to cost and complexity.

There are a variety of business opportunities that are enabled by big-data processing, but they largely fall into two key areas:

- **Improved operational efficiencies** through the migration of existing data-processing workloads to less expensive and more powerful platforms. For example, many enterprises are taking advantage of the low storage costs associated with Hadoop to migrate extract, transform and load (ETL) workloads and historical data from data warehouses to Hadoop, with the aim of lowering the costs of data processing and storage while also freeing up data warehousing capacity for more valuable analytic workloads.
- **Enabling and supporting new business initiatives** by taking advantage of previously unused data and emerging analysis techniques – via clickstream analysis, social data analysis, the Internet of Things, predictive analytics and machine learning, for example – to gain new business insight. Business case examples include real-time customer service, fraud analysis, and targeted marketing and advertising.

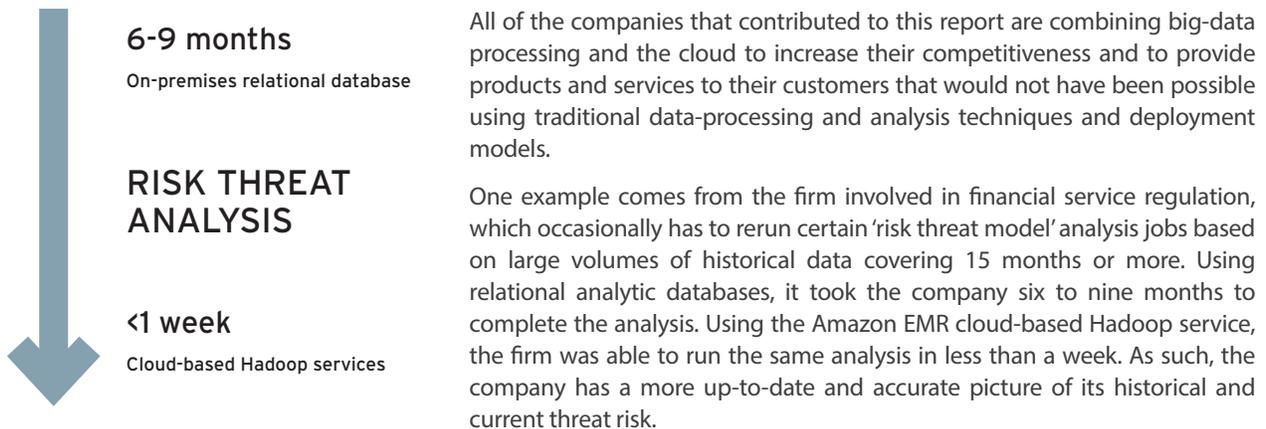
As such, companies that are successfully taking advantage of big-data processing in the cloud are not simply enjoying incremental improvements. The benefits enabled by cloud-based big-data processing quickly become the heart of the business – enabling new applications and business processes through which the company can potentially gain competitive advantage.

The companies interviewed for this report unanimously described their cloud-based big-data processing initiatives as mission-critical in driving change for the business thanks to a variety of metrics, including improved time to market, lower data-processing costs, increased customer insight and improved customer services.

“If anything, our entire existence as a company is a measurement of our results of using the big-data initiatives.”
 – Mobile platform provider

A clear example of this came from the financial services firm, which is using Amazon Redshift as its primary data warehouse for its billing system and related reporting and analytics. The applications, and therefore the underlying Amazon Redshift data warehouse, are mission-critical in enabling the financial services firm to operate successfully.

The same is true for another of our interviewees, a provider of a mobile technology analytics platform that was designed and deployed entirely in the cloud. Without big-data cloud services, the company simply would not exist.



The same company has also seen benefits in querying indexed data for business analysts. Some of those queries are running on a couple of zettabytes of data. With the company’s data warehouse environment, such queries could take up to four hours to complete. That environment has now transitioned to Hadoop and HBase on Amazon EC2 with substantial performance improvements:

“Those same queries using HBase on the cloud are now coming back in ten to twelve seconds, so that’s a 400-fold improvement. Not 400% quicker, a 400-fold improvement.”
 – Financial services regulator



4-hour on-premises data warehouse query reduced to 10-12 seconds with Hadoop and HBase on IaaS

Users and Use Cases

Enterprises are employing a range of use cases for big-data cloud services. Several key examples emerged from our interviews, including providing applications and services directly to customers; performing business reporting and analytics; and exploring data with data scientists to drive new insight and business opportunities.

DATA-DRIVEN APPLICATIONS AND SERVICES

Providing big-data-driven applications and services directly to customers is a primary use case for the marketing strategy firm. The firm today is making use of big-data cloud services to deliver insight to its customers in terms of how their marketing dollars are being spent, and with what results, especially when it comes to competitive analysis.

Naturally, the technology consultancy is also focused on deriving insight to meet the demand of clients based on what their customers were doing. Specific examples of projects taking advantage of big-data cloud services include transaction analysis and loyalty card analysis, as well as the generation of personalized offerings.

The mobile technology analytics platform provider is similarly concerned with serving data to its clients, but in its case, it is providing mobile application developers with insights into their mobile applications.

BUSINESS REPORTING AND ADVANCED ANALYTICS

Cloud-based big data offers a choice of starting points. External audiences can be served with reporting and analytics as easily as internal customers.

For example, while the financial services firm is providing some data to clients to enable them to do competitive analysis compared to their peers, its primary use case is more internally focused, involving the search for illegal trading patterns. As such, the core users are business analysts and financial analysts with domain expertise using reporting and visualization tools.

In addition to providing clients with data-driven services, the marketing strategy firm also has its own business analysts querying the same data via internal reporting and trend-analysis tools so that it can better serve its customers.

Meanwhile, the financial services regulator is running three primary internal workloads in the cloud for business analytics: batch analytics, interactive analytics and summary reporting. The fourth use case – ad hoc analytics – is currently still run on the on-premises data-warehousing infrastructure, but a plan is in place to migrate ad hoc analytics to the cloud.

Delivering an environment to provide this functionality in the cloud is a work in progress. However, the financial services regulator intends to develop the tools that will enable its data scientists to also take advantage of data stored in the cloud based on the advantages it has seen from its batch analytics, interactive analytics and summary reporting use cases.

The end users of those on-premises ad hoc analytics tools are considered data scientists and are likely to use a combination of analysis tools and techniques, including statistical analysis, programmatic analysis and SQL-on-Hadoop.

DATA SCIENCE

Other interviewees are already using big-data cloud services to drive data science. The marketing strategy firm has a team of data scientists doing modeling and predictive analysis on data in the cloud to identify new sources of data and competitive insights that can be rolled back into the products and services sold to its customers, including predictive analytics services.

“There’s a very small pool of data scientists that have free rein ...and run wild, but then if anything is going to become something more than an experiment, we formalize the process.” – *Marketing strategy firm*

Interviewees are using a range of cloud-based big-data services, as indicated in our introduction. The common denominator, though, is the Amazon S3 cloud storage service, which provides some evidence for 451 Research’s belief that data processing and analytics gravitate toward the data, and the increased use of data storage services will be a leading driver of increased cloud-based data processing.

451 Research believes that this concept of data gravity is a primary force influencing the adoption of big-data cloud services because processing resources will migrate to the platform that stores the most data (or perhaps the most important data).

Challenges – Technical, Process, Business

There are multiple challenges associated with big-data projects, whether in the cloud or on-premises. It is striking, however, that the challenges highlighted by the interviewees are largely not specific to the cloud. The conclusion we draw from this is the cloud itself does not represent a major challenge when it comes to big-data deployments.

The shortage of data scientists is a significant challenge across the board, whether on-premises or in the cloud – although cloud less so than on-premises because cloud is set up to handle some of the complexity. For all big-data practitioners, finding candidates who can not only use tools such as Python and R to analyze data but also can apply those skills to create prototypes of new products and services has proved to be difficult.

Hiring people with business analytics skills is also a challenge – ensuring they come with the domain knowledge to apply their technical skills to the business problem at hand. Again, this is a problem highlighted by both financial services firms and the consulting firm, which highlights the need to identify use cases to take advantage of the technology.

Other challenges include integration with external data sources, specifically via APIs, as well as integration with internal systems. This represents a key challenge for the mobile analytics platform provider because it was recently acquired and now is part of a larger company. Because the company has become more successful, another cloud-specific challenge is getting a handle on potential costs and balancing the cost of success – trying to double the number of customers without doubling the amount of infrastructure (and the cost of managing it) required to support those customers.

Data governance is another business-process challenge, particularly for the marketing strategy firm in light of regional regulations related to data governance. Data governance was also raised as a process issue by the consulting firm, along with security – perennial challenges for any company and of particular concern for organizations migrating existing applications to cloud services. These big-data compliance concerns affect both on-premises and cloud-based instances.

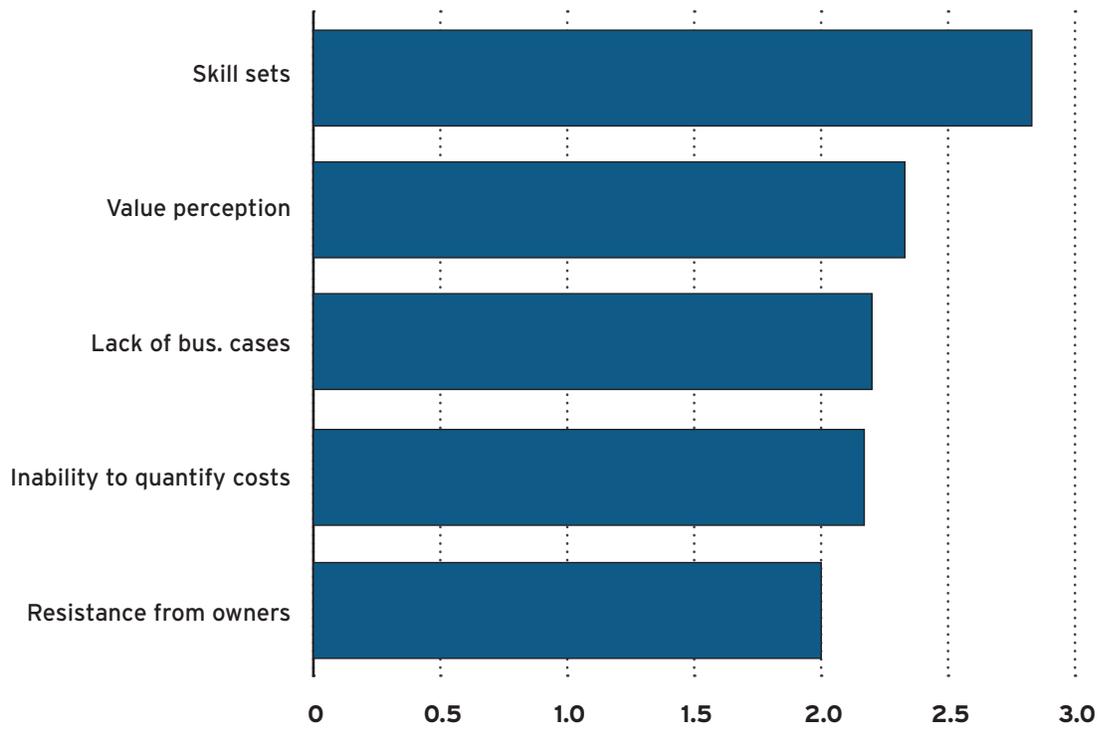
“I think the explosion of data and accessibility in data flying around far surpasses the actual functional practitioners in the governance and management piece. I think we do a pretty good job, but we can always be doing better.” – *Marketing strategy firm*

The financial services firm, not surprisingly, subjected its cloud provider to a multi-month security review to ensure that it had the confidence to migrate data and services to the cloud. Although the cloud provider passed the security review, not all of the financial services firm’s data resides in the cloud. Due to data-governance regulations, some data sets are still stored on-premises.

Resistance to change from established business owners is also a potential challenge raised by both the financial services firm and the consulting provider, although the former found that it could be quickly mitigated with evidence of the TCO advantages.

Potential Challenges to Big Data Projects:

Rated on a scale of 1-5, from least serious to most critical



Measuring Success

Given that companies are making use of big data in the cloud to achieve new business objectives, how do they measure the success of a big-data project – and more specifically, a big-data project in the cloud? The answers varied throughout the organizations based on role. **Interviewees reported a variety of potential benefits, including cost savings, faster time to market, developer agility and revenue generation.**

FASTER TIME TO MARKET

Faster time to market is often the driver for adopting big-data cloud services for startups and emerging industries, enabling them to move from concept to production without the need to design, procure, configure and maintain on-premises infrastructure. This is the primary benefit identified by the mobile technology analytics platform provider, with the additional benefit that scalability within the cloud is enabling the company to be more responsive to peaks in demand. Another benefit identified by the company is the attractiveness of the cloud to engineers and developers, which has enabled it to hire and retain talent.

The mobile application platform provider also cites time to market as its primary benefit from adopting cloud big-data services, having been able to develop and launch its offering more quickly thanks to the availability of out-of-the-box services with built-in scalability.

“Faster time to market was by far the most important aspect. Being able to leverage these out-of-the-box, hands-on services that have built-in scalability and reasonable cost allowed us to create our service much more efficiently.” – *Mobile application platform provider*

The marketing strategy firm has also seen multiple benefits – cost savings, faster time to market, developer agility and revenue generation. Their relative importance changed depending on who was responding: the finance department, the marketing specialists or the IT department. For the marketing strategy firm’s CTO, however, the primary benefit is faster time to market.

“We could have an idea in the morning and then be developing it in the afternoon if need be. We’ve had things where we’ve turned around a full solution in 24 hours. It was unheard of before.” – *Marketing strategy firm*



From concept to production deployment

TOTAL COST OF OWNERSHIP

Larger and established big-data users tend to be drawn to cloud services by the opportunity to reduce costs and improve operational efficiencies. Perhaps not surprisingly, the financial services firm focused its attention on cost-reduction benefits, achieved by migrating from an on-premises data warehouse to a data warehouse service at roughly 40% of the cost.

“Basically, given that total cost of ownership is one of our biggest driving factors, the business’s perception of that value is key to them embracing this whole thing.”

– Financial services firm

“We definitely realized a substantial cost reduction when we moved that warehouse from on-premises [to the cloud]. It was 57% savings.” – Financial services firm



On-premises data warehouse

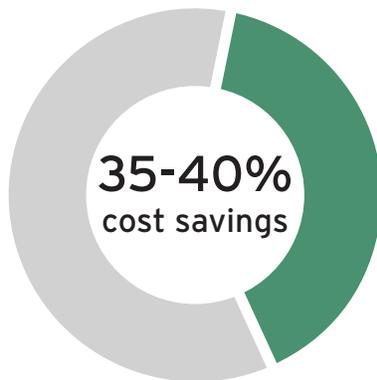
57%
COST SAVINGS

Cloud-based data warehouse service

The financial services regulator also has a strong focus on cost; it uses a variety of key performance indicators (KPIs) to measure its success with its cloud services. These KPIs include whether the project is fit for purpose; whether it was delivered on schedule; and whether it was delivered on budget.

Moving to the cloud has provided some advantages in that regard.

The company discovered via an initial costing project that it could save 35-40% of the cost of a comparative on-premises environment. Additionally, the company has found that by constantly evolving its cloud services to take advantage of innovative new services and technologies, it is able to fulfill its goal of reducing costs by 12-14% annually.



Compared to a comparative in-house environment

“A year and a half ago, we installed this HBase environment” [using a Hadoop distribution]. “We’re looking at replacing that now with Amazon EMR and running HBase on Amazon S3. We had another application that was using these bigger ephemeral systems, these HS1 systems [dense storage and high sequential read and write performance,] and within a year, we replaced them with D2s [dense storage instances and high sequential read and write performance, designed specifically for very large data sets] because it gave greater performance at 40% less cost. We’ve got this one interactive application that’s based upon Hive. It’s been in production for eight months. We’re looking at replacing it with Presto today.” – Financial services regulator

Big data is so intrinsic to the financial services firm’s business that it has applied more generic business indicators to measure the success of its project, using aspects such as month-over-month revenue and market share. As for the system itself, cost is a significant indicator, as is the ability to produce reports in a timely manner to measure the success of the business.

“When we did the numbers and showed that a cloud-based system was going to cost less than half the legacy on-premises system, then the business said immediately, ‘Well, I want that. Just make sure that it’s going to work and my data is going to be secure.’ Once we were able to check all those boxes, that was a no-brainer.”

– Financial services firm

BUSINESS PERFORMANCE

For the marketing firm, the key performance indicators are more specific to the system – cost, uptime, application performance and security. And, these were viewed within the context of the larger concern of how the business as a whole is performing.

Similarly, while the mobile analytics platform provider measures multiple indicators related to the cloud service – performance and relative cost are the key benefits – the most important indicators are related to the business itself. Here, the KPIs include how many customers are using the service; how much customer data it is handling; and how much value it can deliver to customers.

The consultancy has also seen multiple advantages. While saving money is the most obvious upside, faster time to market and revenue-generation improvements have also stood out as significant benefits. Meanwhile, the consulting firm also zeroed in on customer-related indicators such as new revenue generation, new customer generation and customer retention.

The mobile application platform provider also estimates that it has saved costs in terms of avoiding having to hire people to manage and maintain the services themselves.

AGILITY

For the mobile application development platform provider, speeding time to market and the ability to modify and expand its platform in an agile manner is key to making a decision between whether an application should be deployed on-premises or in the cloud.

While cost is often the primary driver for established industries and vendors, as noted above, agility often becomes a greater benefit that can help drive broader adoption by enabling and supporting new business initiatives. In addition to cost considerations, agility is also a key factor for the financial services firm, specifically the ease with which it can provision new compute and storage resources in the cloud compared to the paperwork and hurdles required to provision server hardware on-premises.

“If I want to provision a bunch of hardware [in the cloud], I can do that right now. If I want to provision a bunch of hardware in our datacenter, it is a multi-month extravaganza of paperwork and phone calls.”
 – *Financial services firm*

Agility is also cited as a benefit by the marketing strategy firm, which notes that improved agility is a benefit for both the operations team and the business analysts.

“I think the biggest beneficiary is the IT infrastructure ops guys, because we have nothing that’s on-prem anymore...”the fact that an analyst or a data scientist could, within an hour of asking, have a complete environment set up means the business users are benefactors too!” – *Marketing strategy firm*

Expectations

Not surprisingly, given they are using a combination of big-data processing and the cloud to increase their competitiveness and deliver new services, the interviewees are unanimous in seeing big data as a critical component of the products and services they provide. They also noted that their investment in big data is expected to grow in the next 12-24 months. The unlimited capacity, agility and lower costs that come with hosting big data in the cloud are critical components of enabling that growth, as illustrated by the visions they shared with us about their plans:

Unlimited Capacity

"I think in the past 10 years, people have figured out how to store and gather as much data as possible... people are now figuring out how to utilize that data more and more, so the demand for capacity is only going to continue." – *Mobile analytics platform provider.*

Agility

"The more data that we have and the more rapidly we can do analytics on that data, the more rapidly we can produce new data products that we could sell. You can't manage what you can't measure. I really think that sums it up." – *Financial services firm*

Lower TCO

"We're moving more systems into [the cloud], and we're collecting more data. So while it does cost us more money to collect more data, it is an investment that is definitely paying dividends, and one that the business is clearly backing. They are actively asking us to add more to that data warehouse. As we move systems into [the cloud], we can retire the internal costs for that, which we can then use to save money and spend more on big data." – *Financial services firm*
