



# The State of Big Data Infrastructure:

Benchmarking global Big Data users to drive future performance

April 2015





## Contents

---

<b>Key findings</b>	<b>3</b>
<b>Big Data projects have arrived</b>	<b>4</b>
Level of adoption	4
Complexity of current adoption	4
Drivers and benefits	5
<b>Swimming in data</b>	<b>7</b>
Volume of data is increasing	7
Data sources	8
<b>Infrastructure and Big Data environment</b>	<b>9</b>
Infrastructure	9
Environment	10
Overcoming infrastructure and environment problems	12
<b>Conclusion</b>	<b>13</b>
<b>Appendix</b>	<b>14</b>
Aims of the research	14
Research Scope	14

# Key findings

---

## Big data projects are being widely adopted

- Over half (55%) of respondents' organizations have already implemented Big Data projects
- A further 29% are planning to implement within the next year
- Organizations that have fully implemented Big Data projects have implemented an average of four different ones

## The main drivers for Big Data projects are customer related

- Three in five (60%) respondents say that improving customer experience is driving their organizations' need for Big Data projects
- Over half (54%) say that it is the desire to get new customers

## Benefits of Big Data projects are already being experienced or are anticipated

- Nine in ten (90%) respondents' organizations are either already experiencing or anticipate seeing more effective targeted marketing and selling campaigns due to Big Data projects
- 88% are seeing or anticipate seeing an increase in revenue

## Obstacles to achieving success for Big Data projects

- 92% admit that there are major obstacles to implementing Big Data projects in their organization, for example:
  - their organization's existing infrastructure is not sufficient
  - or there is organizational complexity
- Nine in ten (91%) respondents' organizations are prevented from including data from both mainframe and non-mainframe sources in their Big Data solutions

- Almost all (98%) respondents admit that major investments are required to allow their organizations' Big Data projects to work well

## The majority of Big Data projects are independent projects

- 44% of respondents say that integrating individual projects across functions/departments is a major priority for their organizations
- 35% of respondents say that their organizations' Big Data projects are not integrated across departments
- A further one in five (21%) have a project limited to a single department or area

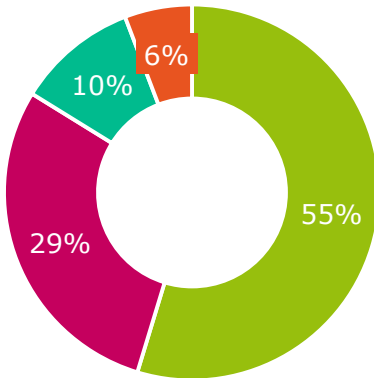


# Big Data projects have arrived

## Level of adoption

Over half (55%) of respondents' organizations have already implemented Big Data projects. An additional three in ten (29%) are planning to start implementing within the next year. Although a relatively new approach, many organizations are already conducting Big Data projects, and only the minority are looking to do this in the distant future.

### Big Data project adoption plan



- Yes
- Not yet, but we plan to in the next year
- Not yet, but we plan to in one to two years
- Not yet, but we plan to in two to three years

**Figure 1:** "Has your company implemented a Big Data project(s)?" (1000 respondents) Respondents whose organization is either planning on implementing beyond three years, or is not planning were excluded from the survey

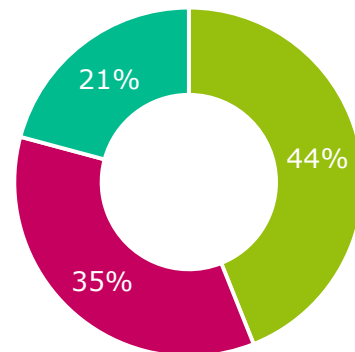
Existing Big Data projects are costing organizations 18% of their overall IT budget, on average. This is expected to increase to 25% in three years' time. Big Data projects are becoming more costly. Even those organizations planning to implement Big Data projects are already spending an average of 15% of their IT budget on this area. Organizations are heavily investing in Big Data projects suggesting that they expect Big Data projects to become an important part of their IT infrastructure/environment.

## Complexity of current adoption

On average organizations have already fully implemented four Big Data projects. Organizations that are implementing Big Data are not relying on just one Big Data project to satisfy their needs. This could be because organizations are unable to analyze all of their data in a single Big Data project because it is too complex to do so.

44% of respondents' organizations' Big Data projects are an integrated initiative across the company. But this is not the case for the majority. Over a third (35%) of respondents say that their organizations' Big Data projects are multiple departmental projects that are not integrated. This is likely to be adding another layer of complexity for organizations. A further one in five (21%) have a project limited to a single department or area.

### Type of Big Data projects



- Integrated initiative across the company
- Multiple departmental projects, not integrated
- Project limited to a single department or area

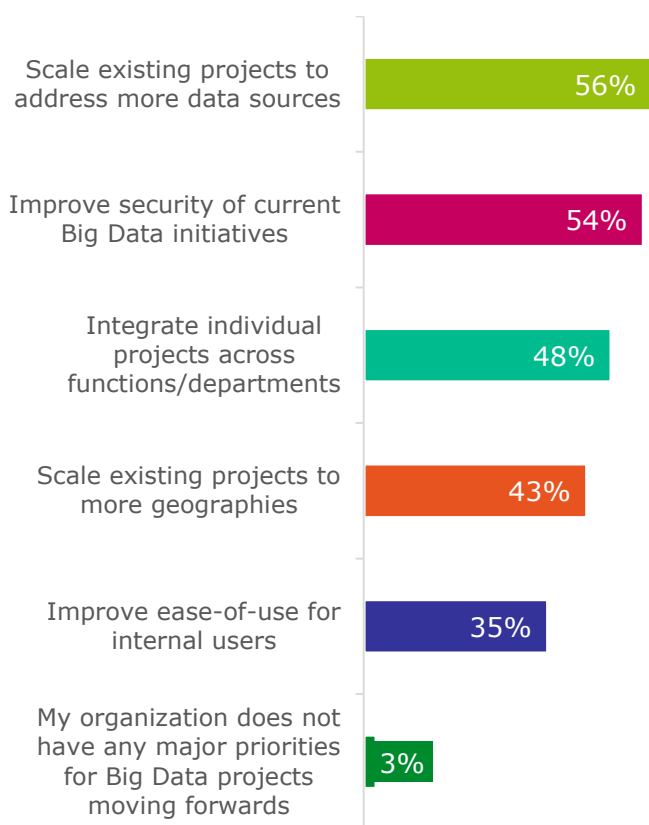
**Figure 2:** "What is the scope for your company's Big Data project(s)?" (1000 respondents)

Not all organizations are approaching Big Data projects in the same way, suggesting that there is no best practice framework for organizations to follow. Consequently, implementing successful Big Data analysis is unlikely to be simple. Those who are not integrating their multiple departmental projects currently may find it very complex to integrate them across the organization in the future.

Integrating individual projects across functions/departments is a major priority for 48% of respondents' organizations, despite the fact that many are not integrating across departments. Organizations want to integrate but they are struggling to do so from the start.

Two in five (40%) respondents whose organizations' Big Data projects are not integrated say that integrating is a major priority. Rather than follow in the footsteps of other organizations, these organizations could overcome the complex nature of integrating from the start of their process by working with good vendors. At present, many organizations could be failing to overcome complexities by going it alone. Other priorities include improving security (54%) and allowing for further data sources (56%). Both of which serve to add another layer of complexity.

### Priorities



**Figure 3:** "What are your organization's major priorities for your Big Data projects moving forward?" (1000 respondents)

It is not just integrating that is complex. According to surveyed users/technology managers, the average analytics infrastructure has 38 databases connected to it, including 38 clusters with an average of 50 nodes per cluster. With many databases, clusters and nodes per

cluster, this too can add to the complexity of Big Data projects. (Please note, these averages are excluding the 4% outliers who stated 1000 or more clusters, nodes or databases.)

### Drivers and benefits

The biggest drivers for Big Data projects are customer related. Three in five (60%) respondents say that improving customer experience is driving their organizations' need for Big Data projects. Over half (54%) also say that they are driven by the need to get new customers. However, customer focus is not the only driver. 46% of respondents say that driving top-line revenue growth is also key for their organizations.

### Driving Big Data projects



**Figure 4:** "What is driving your company's need for your Big Data project(s)?" (1000 respondents)

Most organizations are driven to implement Big Data projects to improve customer experience or volume but other drivers also play a role. Even within organizations, their desire for Big Data projects can vary. Users of the information generated through Big Data and analytics solutions are most likely to be from the finance (59%), operations (58%), marketing (57%) or sales (56%) departments. The finance department are likely to be looking at how the projects can drive top-line revenue growth, and the sales and marketing interest relates to the

biggest drivers for the projects being customer related.

Regardless of what drives organizations or departments to embark on Big Data projects, these projects need to produce measurable success. Over half (54%) of surveyed senior managers say that their organization will measure the success of their Big Data projects primarily through external business factors. This may be expected, as the main driver for many Big Data projects is customer related.

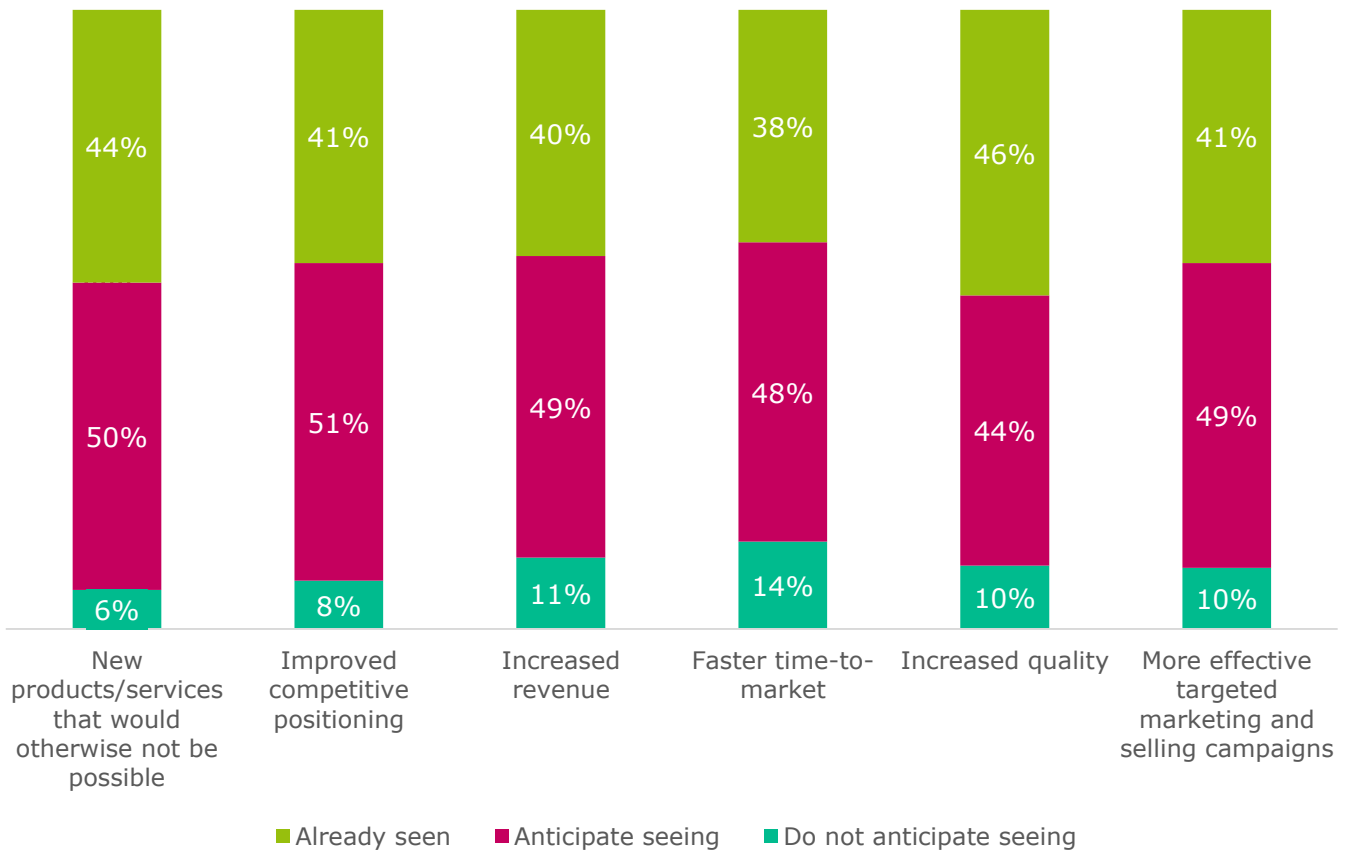
Organizations are measuring the success of their Big Data projects and many are already yielding benefits. The vast majority of senior manager respondents are already experiencing, or

anticipate seeing, benefits from their organizations' Big Data projects.

Of the 41% who are already experiencing more effective targeted marketing and selling campaigns they have experienced a 21% improvement on average in this area due to their Big Data projects. This explains why the sales and marketing departments are avid users of the Big Data output.

Organizations that invest in Big Data projects are likely to experience significant benefits, benefits that have the potential to exceed the initial investment required. For instance, an 18% increase, on average, in revenue.

### Benefits of Big Data projects



**Figure 5:** "What benefits have you seen or do you anticipate seeing from your Big Data Projects?" Asked of senior managers (538 respondents)

# Swimming in data

## Volume of data is increasing

The amount of data coming into and through organizations has increased by an average of 16% in the last two years. This is set to rise by a further 24% in the next two years. The more data organizations have at their disposal, the more likely they are to need to implement Big Data projects. But if that investment is implemented poorly, it would create a more complicated Big Data environment.

Organizations therefore need to invest in ensuring that their Big Data approach is appropriate and can cope with the expected increase in data volumes in the future. Many organizations are likely to need help managing all of this extra data.

The majority (80%) of respondents whose organizations already have Big Data projects estimate that their Big Data tools are analyzing more than 2PBs of data. Similarly, the majority (65%) of those who have not yet implemented a Big Data project estimate that the amount is going to be more than 2PBs.

However, it is a smaller proportion of those who have not yet implemented Big Data projects who estimate this, suggesting that they could be underestimating the amount of data that there will be going through their Big Data tools. Organizations with more data are more likely to have already implemented their projects due to having so much data.

### Big Data volumes

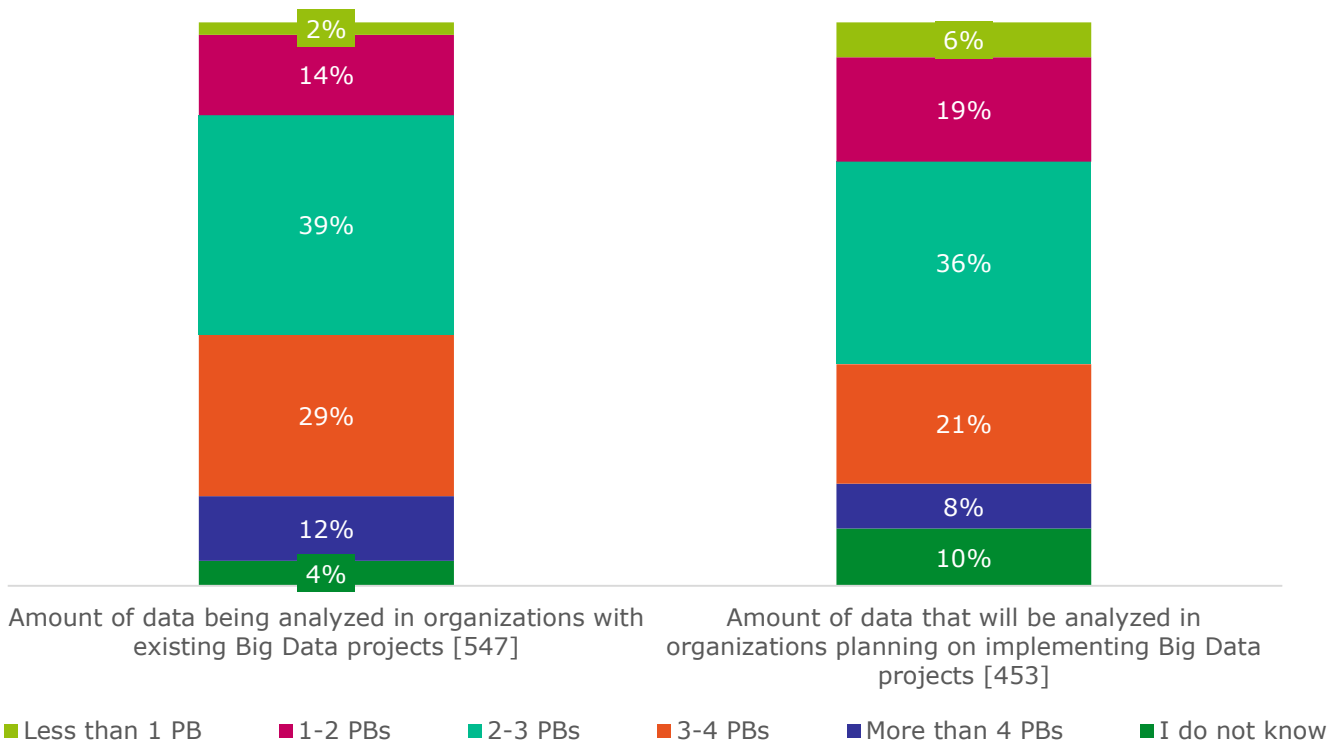


Figure 6: Analysis of the amount of data that is or will be analyzed with Big Data tools ([X] respondents)

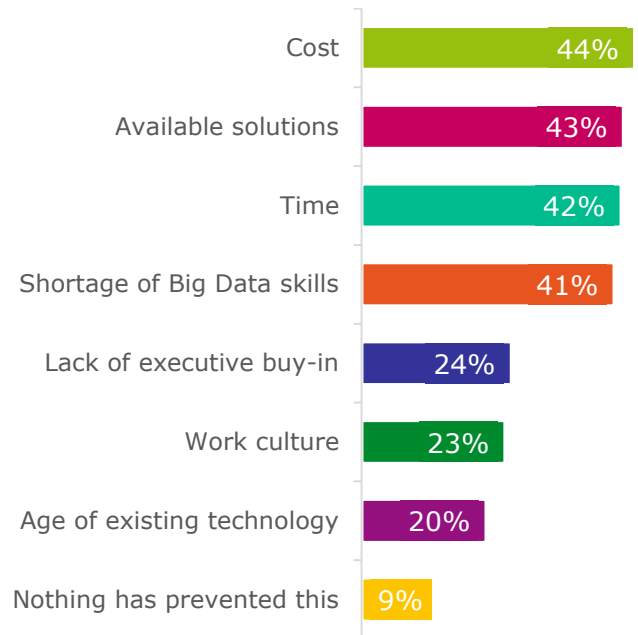
## Data sources

Almost all (94%) users/technology manager respondents' organizations already do, or plan to, analyze structured data with their Big Data analysis tools. Almost half are doing so, or plan to, on a weekly basis at least. Now organizations also want to include structured data with their unstructured data. This will help organizations to get a complete view of the findings across ALL of their data sources. However, this is another source of complexity.

Both mainframe and non-mainframe data sources are, or plan to be, included within 94% of respondents' organizations' Big Data projects. Of those who have already implemented a Big Data project, seven in ten (70%) include data from both sources, with most (29%) of the remainder planning on doing so.

But including data from both sources is not easy. Of the 6% of respondents whose organizations are not planning on including data from both mainframe and non-mainframe sources, the majority (86%) claim that there are a variety of reasons preventing them. If obstacles can be overcome, organizations would want to include data from both sources. But for many, this is proving too complex for them to consider doing so. These organizations need to seek help from vendors to overcome the complexities.

### Preventions to using mainframe and non-mainframe data sources



**Figure 7:** "Have any of the following prevented your organization from moving forward with including data from both mainframe and non-mainframe sources in your Big Data and analytics solution?" (1000 respondents)

Ineffective solutions could have a cost implication. One of the most common barriers respondents' organizations experience when including data from both mainframe and non-mainframe sources is cost (44%), as it is complex to do it costs a lot of money. This is not helped by the fact that there are not many easily available solutions (43%). Due to this complicated environment, there is a lack of internal Big Data skills (41%) to combat challenges. Investment is needed for most businesses in these areas to allow Big Data projects to work well, or external support might be the most beneficial long-term answer.



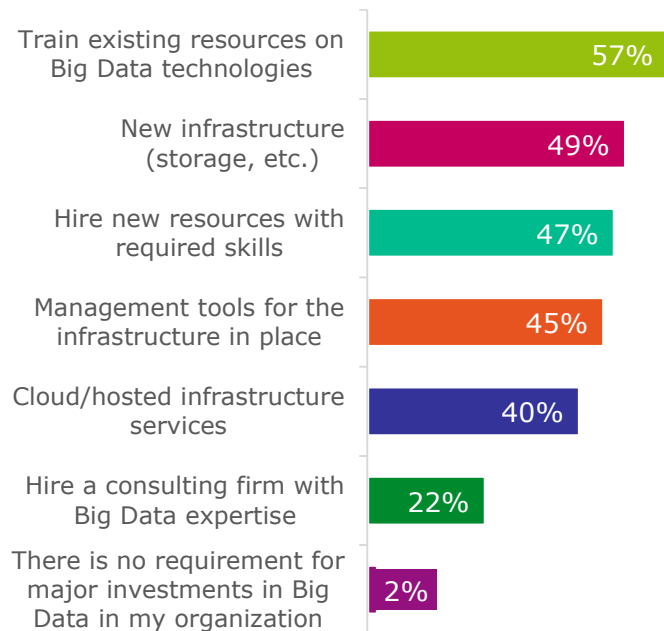


# Infrastructure and Big Data environment

## Infrastructure

Major investments are needed to allow Big Data projects to work well, according to respondents. This includes investing in people, including training existing resources (57%) or hiring new staff with the required skills already (47%). But two fifths of respondents report that their organizations also need to invest in new infrastructure (49%), management tools for infrastructure already in place (45%) or cloud/hosted infrastructure services (40%). Organizations are likely to require investment in both personnel and infrastructure if they are to utilize their Big Data projects fully. Despite the fact that so many say that there are cost/skills shortages currently, their decision makers are also aware that they would benefit from implementing data from multiple sources. Decision makers will be looking to improve their current systems as soon as possible to achieve the next step in their Big Data project's vision

### Investments needed



**Figure 8:** "Which of the following areas require major investments to allow your organization's Big Data projects to work well?" (1000 respondents)

Big Data infrastructures are currently deployed in the cloud for three quarters (74%) of respondents' organizations, either solely (41%) or as a hybrid with on-premise (33%). However,

three in five (59%) deploy their Big Data infrastructure on-premise to some extent.

As many have previously said that their Big Data projects include structured data sources, it is likely that many are using databases containing sensitive information (such as financial data etc.) which are still kept on-site. They are server hugging. They do not want to let go of control due to security concerns, fear of implications if data was lost, deleted or accessed. Keeping the infrastructure on-premise gives a sense of security.

On-premise use is expected to decline slightly in the future, with around half (48%) of all respondents' organizations not planning to deploy their Big Data infrastructure on-premise in the future. As cloud continues to be widely adopted and concerns over security subside, organizations will be more likely to put their data into the cloud confidently.

Before organizations can think to the future, they have to ensure that their current solutions are working efficiently. Over nine in ten (92%) respondents report there being major obstacles with implementing Big Data projects in their organizations. With many projects in place, and many departments running their own projects, an integrated Big Data solution is proving difficult. Around a third (32%) of respondents admit that their organization's existing infrastructure is not sufficient and that it is a major obstacle for them to implement Big Data projects. This might explain why around half (49%) have said that major investments are needed in new infrastructure.

## Big Data project obstacles



**Figure 9:** "What are the major obstacles to implementing Big Data projects in your organization?" (1000 respondents)

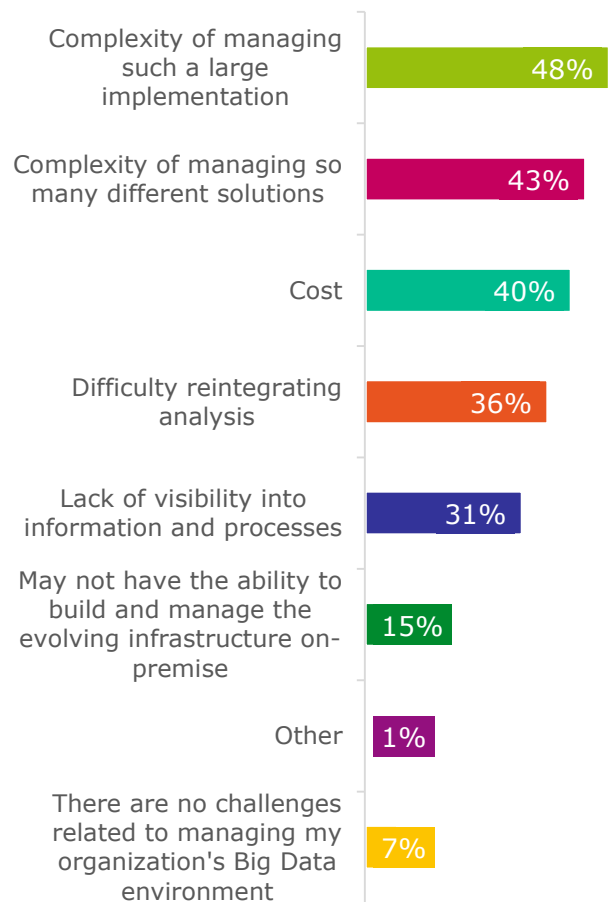
Over a quarter (27%) of respondents state that organizational complexity is a major obstacle. Making sure data is held in the right form in terms of security or compliance is a concern for 26%. Big Data projects are meant to allow organizations to see all information and process, but 25% say that they lack visibility, suggesting they cannot implement Big Data projects appropriately.

Over nine in ten (92%) respondents also anticipate major challenges with managing their organization's infrastructure to support current and future Big Data initiatives. A quarter (25%) see the lack of ability to build the evolving infrastructure on-premise as a major challenge. One possible reason why so many organizations are using the cloud to deploy at least part of their Big Data infrastructure.

## Environment

Over nine in ten (93%) respondents anticipate Big Data environment challenges. The most anticipated challenge is the complexity of managing such a large implementation (48%). But many are implementing multiple projects, not just a single implementation. Many organizations, possibly overwhelmed by how complex it is, are doing smaller projects as a result. But in the process, businesses might be making things even more complex.

### Environment challenges



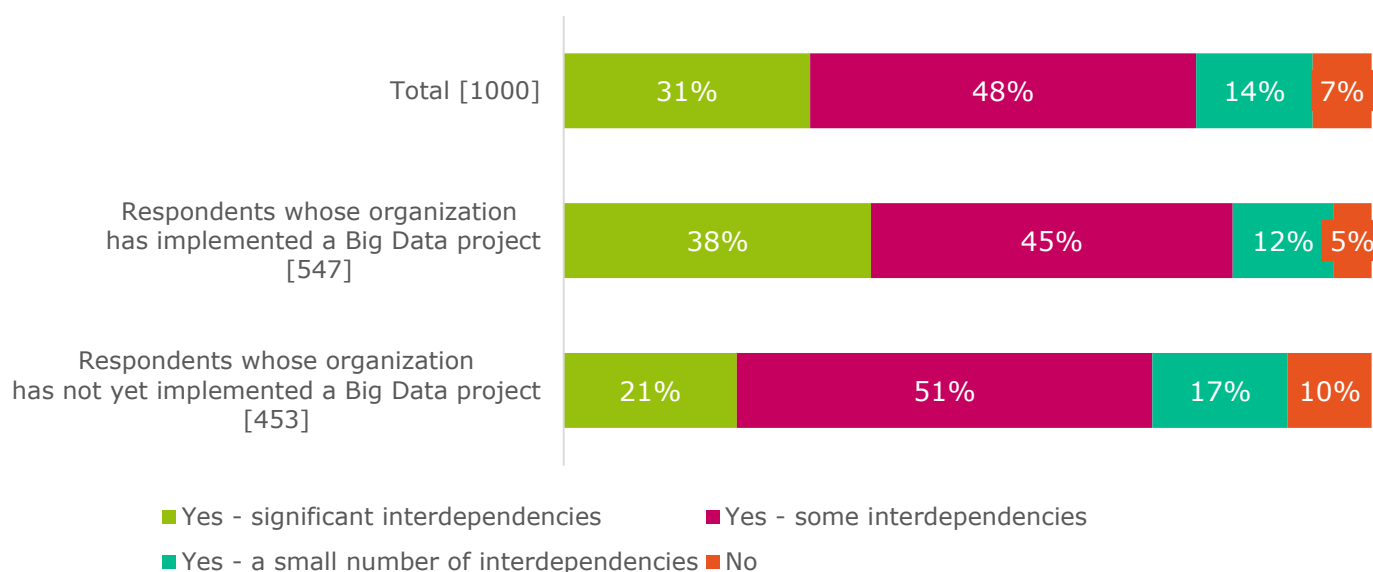
**Figure 10:** "What are or do you anticipate to be, your organization's greatest challenges related to managing your organization's Big Data environment?" (1000 respondents)

Another challenge felt by over two in five (43%) respondents is a concern about the complexity of managing so many different solutions. With a lack of skills in the marketplace, organizations are likely to need assistance from vendors to help manage some of these complexities.

The vast majority (93%) of respondents say that there are interdependencies between their Big Data processes and other processes outside of the Big Data environment. This could require efficient coordination within the departments to limit complexities. Most organizations have isolated Big Data projects, which will hinder the organization's communication and collaboration rather than reduce complexities.

Around two in five (38%) respondents whose organizations have implemented a Big Data project say that these interdependencies are significant. Whereas only one in five (21%) of those who have not yet implemented a project think that they will be significant. These organizations may be underestimating how significant the interdependencies will be and may implement Big Data solutions that make things more complex as a result.

## Interdependencies



**Figure 11:** "Are there interdependencies between your Big Data processes and other processes outside of the Big Data environment?" Analyzed by whether the respondents' organization has implemented a Big Data project yet or not ([X] respondents)

Another potential complexity around the Big Data environment is that organizations need to adhere to certain auditing standards. Almost all (97%) respondents' organizations that have implemented a Big Data project have to adhere to auditing standards. Over three in five (63%) adhere to data compliance checks or data retention periods. These auditing standards need to be taken into account when Big Data projects are set up, or even when they are amended, to ensure organizations do not fall foul of them and bring difficulties upon themselves.

Similarly, certain regulations or authorities affect the majority (87%) of respondents' organizations in the way that they manage data. To avoid further complexities, these standards need to be incorporated when Big Data projects are set up or amended. If not, organizations may come to regret their decisions.



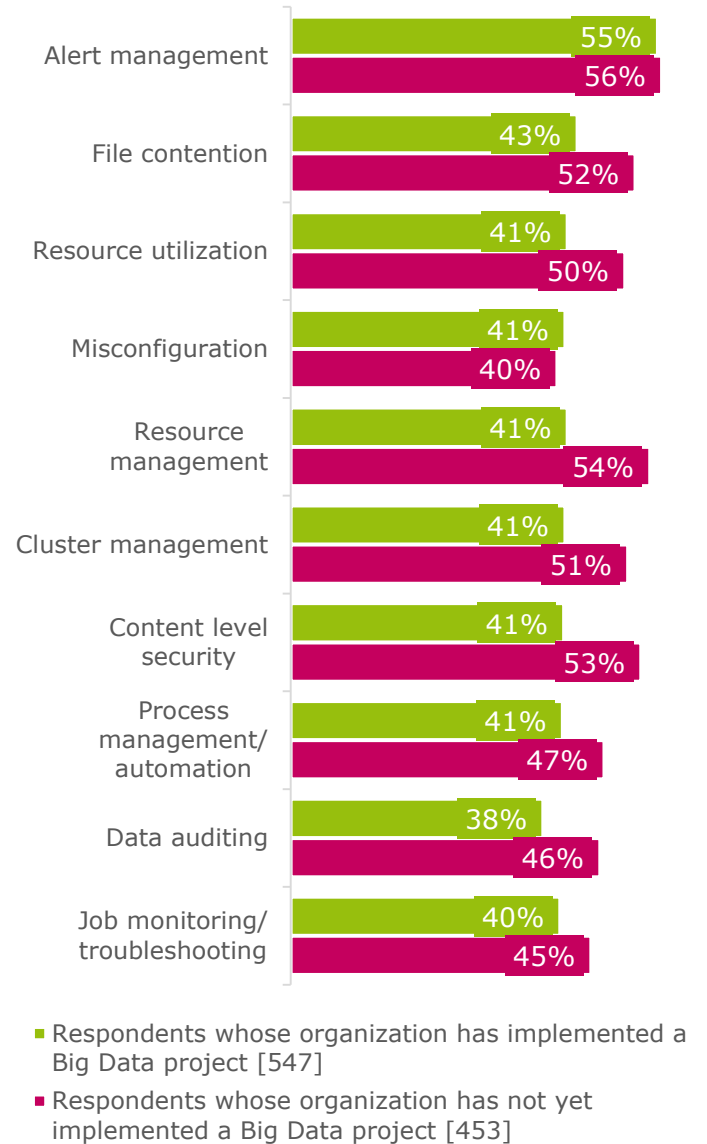
## Overcoming infrastructure and environment problems

To create a more effective Big Data infrastructure management solution, over half (55%) of respondents believe an alert management capability would assist their organization. Respondents whose organizations have not yet implemented Big Data projects are the most likely to think that capabilities such as resource management (54%), content level security (53) or file contention (52%) would assist their organization.

Implementing effective Big Data projects involves innovative technology, which can be obtained through the support of leading vendors.



### Capabilities to assist effective Big Data projects



**Figure 12:** Analysis of respondents who believe these capabilities would best assist their organization in creating a more effective Big Data infrastructure management solution, by whether they have already implemented a Big Data project or not ([X] respondents)

# Conclusion

---

In order to remain competitive, or gain a competitive advantage, organizations must figure out how to harness Big Data successfully. As spend on Big Data projects is expected to increase from 18% of the IT budget to 25% in the next three years, more organizations will be taking on their first Big Data projects and others will be expanding the scope of their existing portfolio.

While benefits are being experienced and are expected from Big Data projects, there are complexities that organizations need to consider. The amount of data organizations have access to is expected to continue to increase. The more data there is in different forms (both structured and unstructured); the more complex processes involved to ensure that analysis of the data is comprehensive, meaningful, and useful.

Organizations that do not create Big Data projects in an organized, future-looking way may find that their solutions become more challenging and less appropriate to use over time.

Organizations should not shy away from investing in the necessary infrastructure and solutions to support the vast amount of differing types of data that Big Data projects currently use now, and decision makers intend to use in the future. Especially given the significant benefits that successful use of Big Data can yield. For instance, an increased revenue of 18%.

While there may be complexities experienced when managing such large implementations, leading vendors will be able to help alleviate these complexities and work together with organizations to ensure the benefits outweigh these concerns. Few organizations that go alone will implement the robust solutions required in today's and importantly tomorrow's Big Data environment.



# Appendix

---

## Aims of the research

For decades, businesses have been analyzing their structured data to gain strategic insight about their business performance. Advances in technology in recent years have allowed organizations to take a step further with analyzing their data. They are now pulling in semi-structured and unstructured data into the mix to conduct Big Data analysis.

As Big Data analysis is a relatively new approach, some organizations might not be doing this as effectively as possible and some businesses might be doing better than others. It is likely that many organizations are struggling to utilize all of the data available to them effectively. This will prevent them from the many advantages of Big Data projects; they could be falling behind their competitors.

This research was commissioned to identify:

- The types of Big Data projects organizations have or are planning for
- What benefits organizations are experiencing or anticipate seeing as a result of their Big Data projects
- What complexities organizations experience when adopting or enhancing Big Data projects
- The amount of data organizations have, and the different sources of data used in Big Data projects
- The challenges and complexities organizations face with their infrastructure and environment in relation to Big Data projects

## Research Scope

CA commissioned independent technology market research specialist Vanson Bourne to undertake the research upon which this report is based.

A total of 1000 senior managers and users/technology managers in IT were interviewed in February 2015. The respondents were from organizations with an annual revenue or budget of more than \$500m and from a range of sectors, including retail; financial services; banking; telecoms; insurance; and government.

Interviews were conducted in the following countries:

- Brazil (100 interviews)
- US and Canada (200 interviews)
- France (100 interviews)
- Germany (100 interviews)
- Italy (100 interviews)
- UK (100 interviews)
- Australia (75 interviews)
- China (75 interviews)
- India (75 interviews)
- Japan (75 interviews)

Respondents' organizations had to either have already implemented a Big Data project, or plan to be doing so within three years. Interviews were conducted online using a rigorous multi-level screening process to ensure that only suitable candidates were given the opportunity to participate.



---

### **About CA:**

CA Technologies (NASDAQ: CA) creates software that fuels transformation for companies and enables them to seize the opportunities of the application economy. Software is at the heart of every business in every industry. From planning to development to management and security, CA is working with companies worldwide to change the way we live, transact, and communicate – across mobile, private and public cloud, distributed and mainframe environments. Learn more at [ca.com](http://ca.com).

Copyright © 2015 CA. All rights reserved. All trademarks, trade names, service marks and logos referenced herein belong to their respective companies.

### **About Vanson Bourne:**

Vanson Bourne is an independent specialist in market research for the technology sector. Our reputation for robust and credible research-based analysis, is founded upon rigorous research principles and our ability to seek the opinions of senior decision makers across technical and business functions, in all business sectors and all major markets. For more information, visit [www.vansonbourne.com](http://www.vansonbourne.com)

---