



Peer Research

Big Data Analytics

Intel's 2013 IT Manager Survey on How Organizations Are Using Big Data

Why You Should Read This Document

This report describes key findings from a survey of 200 IT professionals in the United States about big data analytics. These findings can help you plan your own projects and give perspective on what the results mean for the IT industry, including:

- Big data is a top strategic priority for IT managers.
- As might be expected, updating data center infrastructure also ranks as a top priority for our survey group.
- Adoption of tools such as the Apache Hadoop* framework continues to rise. Over half of respondents have already deployed or are currently implementing a Hadoop* distribution.
- Half of those with a Hadoop distribution under way use an internal private cloud.
- Interest in using a third-party cloud service provider to analyze data sets is growing.



- The leading ways respondents are using big data today relate to understanding staffing levels and productivity. Generating competitive intelligence also tops the list.
- In the future, it's anticipated that big data will be used to help improve operational efficiency and identify new sources of revenue.
- When it comes to big data, analytics is clearly ranked as a leading challenge.
- The learning curve to understanding big data appears to have peaked for IT managers and business units making big data requests. However, now finding the right skill set to do big data analytics has emerged as a top concern, and a shortage of skilled workers is one thing keeping IT managers up at night.
- Security remains a top concern, as does the rate of data growth.

As an update to a 2012 Intel survey, this report also shows a year-over-year evolution in the understanding of big data analytics, and demonstrates the progress made in implementing big data analytics projects in the enterprise.



Contents

- 4 About This Report
- 5 Big Data: The Big Picture
- 8 Getting Clear on Big Data
- 12 What Big Data Solutions Are Being Used?
- 18 Big Data Challenges
- 22 Conclusions: Making Sense of Our Results
- 23 Appendix: Methodology and Audience

About This Report

Across all types of industries, the proliferation of data continues at breakneck speed, and the need to turn information into insight has never been greater. As interest in using big data for greater insight grows, companies that seek to stay competitive in the future will need a plan for addressing big data in their enterprise.

We surveyed 200 IT managers in large U.S.-based companies to learn their approach to and understanding

of big data analytics. We asked them what technologies they have adopted, what challenges they face, and what kind of insight they were expecting to gain.

Results of our survey are detailed in this report. Our goal is to provide IT managers with data that can serve as a benchmark for how your peers are addressing big data in the enterprise and offer insight IT managers can use as they tackle their own big data analytics strategy and projects.

Executive Summary

Big data offers companies the opportunity to gain richer, deeper, and more accurate insights into customers, partners, and business and grow their competitive advantage. Results from the survey indicate that companies are rounding the learning curve for big data. Not only is big data a top strategic priority for large organizations, but most enterprise companies already have a formal big data analytics strategy in place. IT managers are confident in their understanding of big data, and the requests they receive from their constituents indicate that business units have a good grasp of their big data needs.

Similar to our 2012 big data analytics survey, adoption of tools such as the Apache Hadoop* framework continues to rise. Many organizations have selected the type of Hadoop* implementation they will use (a commercial version or building their own open-source solution), and increasingly, that solution includes a cloud implementation. Most importantly, organizations have identified the purposes for using big data analytics to gather insight, as well as how they expect to employ big data through 2016.

Also building on our 2012 survey, respondents identified a number of challenges and obstacles to big data analytics. Security concerns still pose a challenge to our respondents, as does the ability to manage the volume, velocity, and variety of big data. Newly emerging is concern over the shortage of skilled data scientists and data professionals available to help make sense of big data.

Key findings include:

- Big data is a top strategic priority for IT managers.
- As might be expected, updating data center infrastructure also ranks as a top priority for our survey group.
- Adoption of tools such as the Apache Hadoop framework continues to rise. Over half of respondents have already deployed or are currently implementing a Hadoop distribution.
- Half of those with a Hadoop distribution under way use an internal private cloud.
- Interest in using a third-party cloud service provider to analyze data sets is also growing.
- The leading ways respondents are using big data today relate to understanding staffing levels and productivity. Generating competitive intelligence also tops the list.
- In the future, it's anticipated that big data will be used to help improve operational efficiency and identify new sources of revenue.
- When it comes to big data, analytics is clearly ranked as a leading challenge.
- The learning curve to understanding big data appears to have peaked for IT managers and business units making big data requests. However, now finding the right skill set to do big data analytics has emerged as a top concern, and a shortage of skilled workers is one thing keeping IT managers up at night.
- Security remains a top concern, as does the rate of data growth.







Big Data: The Big Picture

It's big and it's everywhere. No doubt you've felt the impact of data growth within your own organization. So it should come as no surprise that data is exploding at a phenomenal rate, with worldwide growth predicted to reach 8 zettabytes by 2015. As the vast amount of structured and unstructured

data continues to expand, so does interest in harnessing the value of that information. As we approached completing an "updated for 2013" peer research study around big data analytics, we again sought out companies already handling large volumes of data, similar to our 2012 study.

Big Data Leads as a Strategic Priority

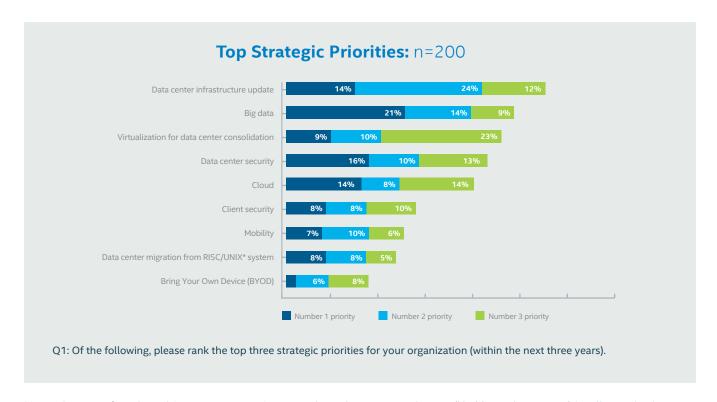
We asked respondents to pinpoint the top three strategic priorities for their organization within the next three years. Big data and data center infrastructure updates emerge as the most important priorities for our survey group.

Additionally, big data is mentioned as the number one priority by 21 percent of all respondents. As might be expected, this is followed closely by foundational priorities such as data center security and infrastructure updates as a top strategic priority. Cloud also ranks highly as a top priority.

Not surprisingly, larger companies (\$100M+) are three times more likely than smaller organizations (<\$100M) to rank big data as their top priority (34 percent vs. 11 percent).

Key Finding

Big data and data center infrastructure updates are hand-in-hand top strategic priorities for our survey group.



¹ Source: "Big Data Infographic and Gartner 2012 Top 10 Strategic Tech Trends." Business Analytics 3.0 (blog) (November 11, 2011). http://practicalanalytics.wordpress.com/2011/11/11/big-data-infographic-and-gartner-2012-top-10-strategic-tech-trends/







Who Has a Formal Big Data Analytics Strategy, and Are They Processing Unstructured Data?

As mentioned, our 2013 findings provide a keen sense that our survey group is making significant strides in tackling big data projects. In fact, most of our respondents have a strategy in place for dealing with big data analytics. And the majority of those without a formal strategy intend to have one in place within the next six months.

Further, we asked 2013 respondents to specify their timeline for putting a big data strategy in place. This yielded an additional insight: The vast majority of organizations that are planning to build a big data strategy will have done so within a year, with a small group tackling big data within two years. Only 2 percent of our survey group has no timeline for developing a big data analytics strategy.

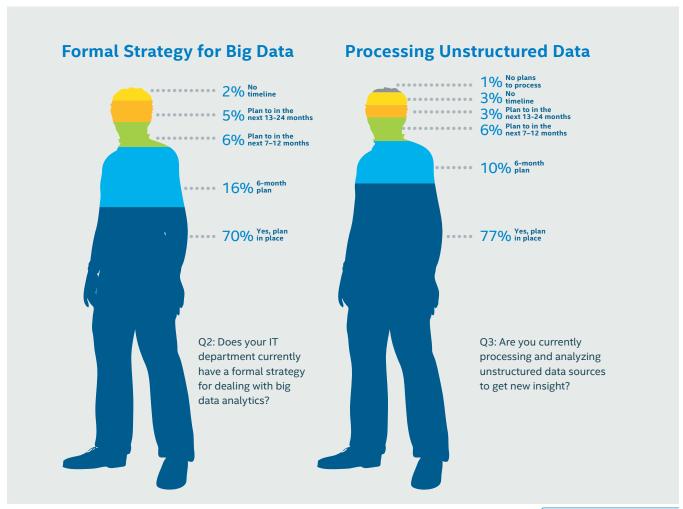
Additionally, three in four respondents are currently processing both structured and unstructured data; again, a large number of those who are not currently processing both types of data plan to do so within the next six months.

Interestingly, those in smaller companies (500 to 999 employees) are significantly more likely than larger organizations (1,000+ employees) to report having a formal strategy in place (78 percent vs. 64 percent). While it's possible that this can be attributed to the ability of smaller companies to act faster and more nimbly, the survey questions posed did not ascertain that insight specifically.

And finally, while the percentage of organizations with a formal plan in place is down slightly compared to our 2012 survey (75 percent in 2012 vs. 70 percent in 2013), the number of organizations that are not planning to address big data has also declined (5 percent in 2012 vs. 2 percent in 2013).

Key Finding

Three of four respondents are currently processing both structured and unstructured data.





What Are the Leading Types of Unstructured Data Anticipated in the Next 12 to 18 Months?

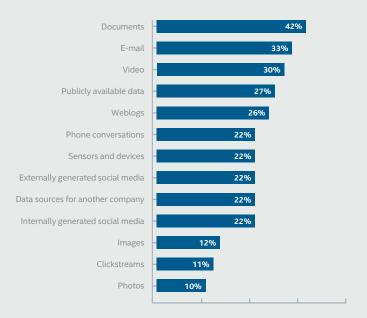
Documents, e-mail, and video are expected to be the top sources of unstructured data to be analyzed in the next 12 to 18 months. Of those three data sources, documents and e-mail topped the list in 2012, as well.

Our survey indicates that larger companies are more likely to process video (36 percent vs. 21 percent) and internally generated company or employee social media (28 percent vs. 15 percent).

Finally, and not surprisingly, it's clear from our respondents that interest in a wide variety of unstructured data, including sensor and device data, is prevalent.

Projected Primary Unstructured Data Sources (Next 12 to 18 Months): n=200

Up to three mentions



Q4: Among the following unstructured or semistructured data sources, please select the three you believe will be the most important for your organization to analyze for insights in the next 12 to 18 months.







Getting Clear on Big Data

As with any technology, there is a learning curve, and big data is no different. The good news: Our survey shows that IT managers are rounding that curve, and are actively fielding big data requests and working on big data projects within their organizations.

Are We Understanding Big Data Better Today?

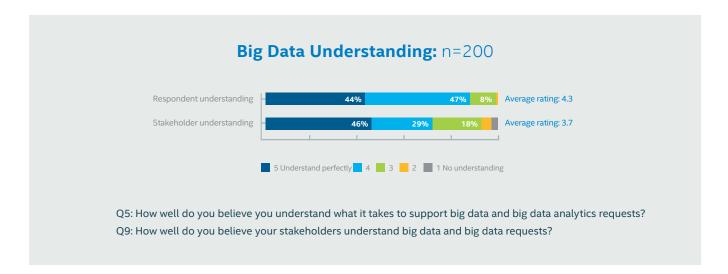
Overall, IT managers are confident with their big data knowledge. Ninety-one percent indicated that they have a strong understanding of what it takes to support big data. This represents a continued growth in understanding, compared to our 2012 survey, in which 80 percent of respondents had a strong understanding.

Survey respondents believe that their stakeholders are pretty clear on big data, too, both in understanding and requests.

Not surprisingly, those with a formal big data strategy in place are significantly more likely to say both they (55 percent vs. 18 percent) and their stakeholders (59 percent vs. 12 percent) "understand perfectly" big data and big data requests.

Key Finding

IT managers surveyed believe that they—and their stakeholders making requests—have a strong understanding of big data.



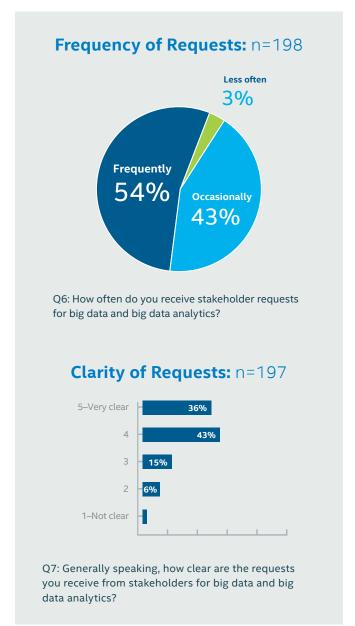




Are Big Data Requests Clear?

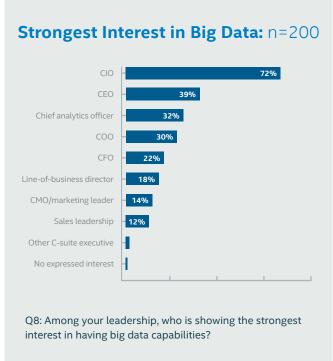
Requests from stakeholders for big data analytics are fairly common. In fact, over half of our survey group says they receive requests frequently, and nearly all (97 percent) receive them at least occasionally.

Fortunately, these requests tend to be fairly clear (79 percent), given the stakeholders' understanding of big data.



No Question: The C-Suite Is Interested in Big Data

Survey respondents indicated that within their organizations, there is strong interest in big data analytics being expressed from executive management. Among executive leadership, the CIO shows the strongest interest in having big data capabilities (72 percent), followed by the CEO (39 percent), the chief analytics officer (CAO) (32 percent), and the COO (30 percent). Such widespread interest from the C-suite supports the continued growth in the number of organizations who already have, or will soon have, a big data strategy in place. (See questions 2 and 3.)







How Is Big Data Being Used?

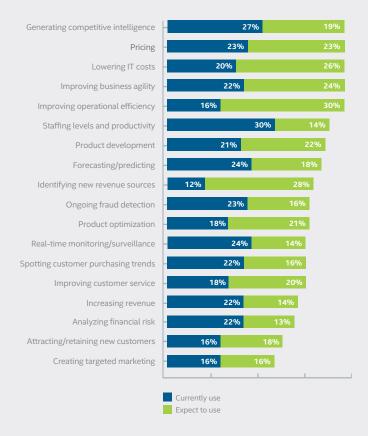
Today, big data is being used to gather multiple types of insight. Among our respondents, generating competitive intelligence and determining staffing levels and productivity are the most frequently cited current uses of big data analytics. Use of big data to help gain actionable insight for pricing, fraud detection, monitoring and surveillance, forecasting, and pricing also registered as prominent current focal points.

By 2016, the expectation is that big data will be used most often to help improve operational efficiency (30 percent) and identify new revenue sources (28 percent), as well as provide insight to other areas of business, such as lowering IT costs, improving business agility, and improving pricing.

Key Finding

Currently, big data is used most frequently to generate competitive intelligence and determine staffing levels.

Big Data Use/Expected Use: n=200



Q10: For the big data purposes listed above, please indicate for which your company currently uses big data, and for which you expect to use big data by 2016.







What's the Ratio of Real-Time to Batch Processing?

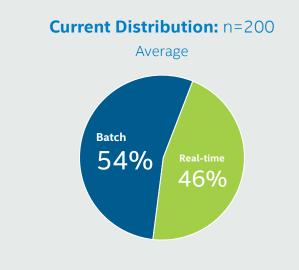
Current levels of data being processed batch versus real time are fairly evenly split, with a slight edge to batch processing. By 2016, respondents expect this ratio to be reversed, with an average of 57 percent of processing being done in real time.

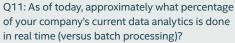
Results compare somewhat similarly to 2012, although, curiously, more data is processed today (54 percent) in batch compared to 2012 (50 percent), and less real-time processing is anticipated by 2016 (57 percent) than it was in 2012 by 2015 (63 percent). This year-over-year trend from 2012 to 2013 showing a slight decline in the move toward real-time data processing could suggest a couple of points:

- As big data projects get under way, IT managers are more clearly seeing when and how they need to use which type of processing for what purposes.
- Batch processing is still a workhorse in today's data center (and is still expected to be so by 2016).

Key Finding

IT managers are developing a clearer sense of the mix of batch and real-time processing they need today, as well as the mix they may require in the future.





2016 Expected Distribution: n=200



Q12: By 2016, approximately what percentage of your company's current data analytics do you expect to be done in real time (versus batch processing)?





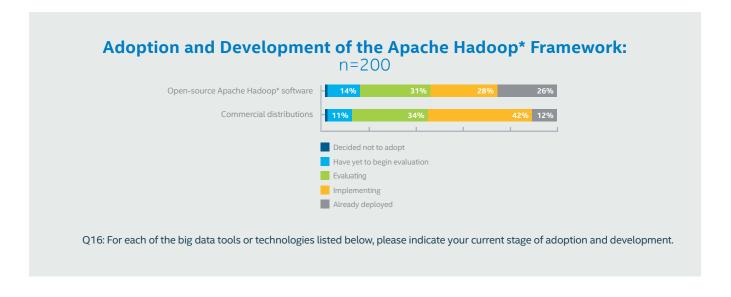
What Big Data Solutions Are Being Used?

As IT managers tackle big data analytics projects, they face numerous technology considerations. Is a commercial distribution of the Hadoop framework a good option, or are they comfortable building their own open-source

distribution? What additional data center infrastructure needs crop up with big data? And is a private, public, or hybrid cloud the right choice? Our snapshot of IT organizations points to a variety of ways to address big data.

What Types of Apache Hadoop* Implementations Are Organizations Activating?

Roughly a quarter (26 percent) of respondents have deployed open-source Hadoop software, double the number of those who have deployed a commercial distribution of the Hadoop framework (12 percent).



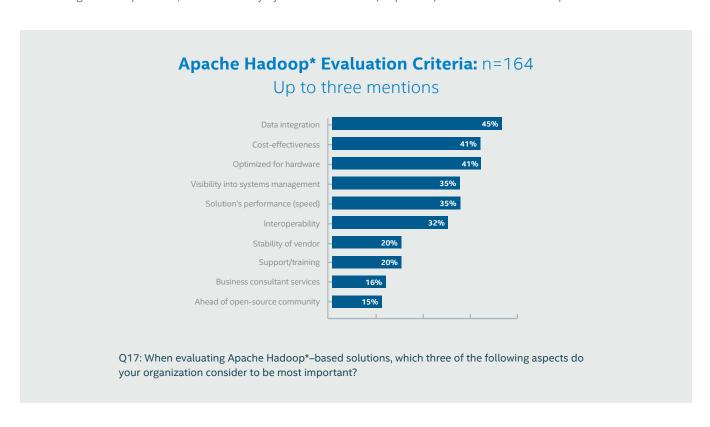






What Evaluation Criteria Is Most Important?

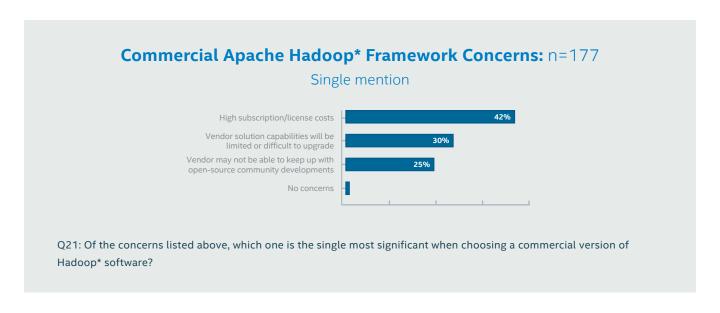
As they evaluate Hadoop-based solutions, our sample group of companies tends to prioritize features related to operability, efficiency, and cost implications over items such as stability of vendor, support, or consultant services. At 45 percent, data integration rates the highest in importance, followed closely by cost-effectiveness (41 percent) and solutions that are optimized for the hardware.





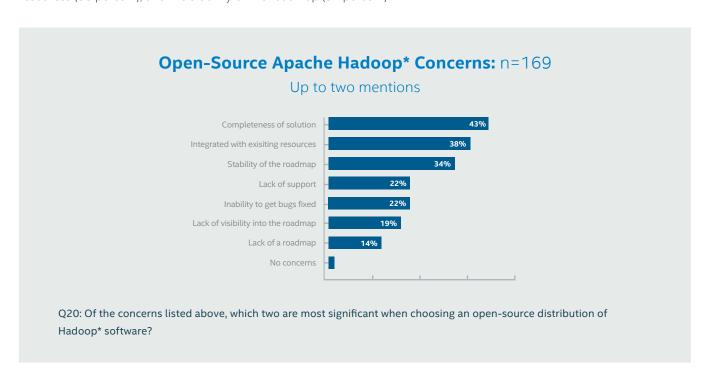
What Barriers Exist to Using a Commercial Distribution of the Hadoop* Framework?

As our survey group evaluates commercial distributions of the Hadoop framework, cost surfaces as a predominant concern. A high subscription or license cost of the solution was listed as the most significant factor when choosing a commercial version of Hadoop software. Concerns over the ability to upgrade the solution followed at 30 percent.



What Concerns Do IT Managers Have about an Open-Source **Apache Hadoop Solution?**

Of respondents reaching the evaluation stage or beyond for open-source Apache Hadoop software, the greatest concerns relate to the completeness of the solution (43 percent), the ability to integrate an open-source solution with existing resources (38 percent), and the stability of the roadmap (34 percent).

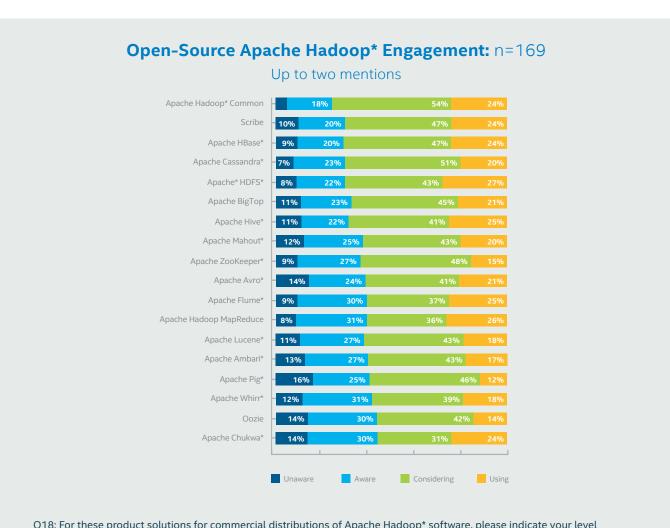






Which Hadoop Components Are Favored?

Hadoop component use is evenly distributed across the survey group; no clear leader or laggard emerges. Usage and consideration of open-source Apache Hadoop product solutions is relatively the same across all tested solutions, meaning that respondents do not identify any particular solution as either having a clear market share or lacking in terms of exposure.



Q18: For these product solutions for commercial distributions of Apache Hadoop* software, please indicate your level of awareness or engagement.





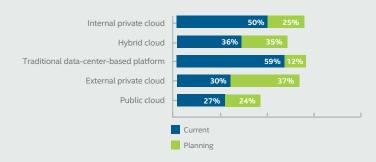
How Are Organizations **Delivering Big Data?**

When it comes to big data analytics, cloud use is fairly widespread. While 59 percent are delivering big data analytics to their organization via a traditional data-centerbased platform, half are delivering big data via an internal private cloud. Additionally, 36 percent are using a hybrid cloud, 30 percent have deployed an external private cloud, and 27 percent are using a public cloud.

As use of the cloud continues to grow, the use of traditional data-center-based platforms is expected to wane, with only 12 percent of respondents planning to deliver big data this way in the future.

Similar to other peer research conducted by Intel, public cloud trails as an option, with possible reasons for this related to security concerns and the desire to maintain better control of data center resources by keeping them in house.

Delivering Big Data Analytics: n=200



Q22A: How are you currently delivering big data analytics to your organization?

Q22B: How are you planning on delivering big data analytics to your organization in the future?

Note: For questions 22A and 22B, survey respondents were able to choose more than one option.





Is There Interest in Shifting Big Data Processing to Third-Party **Cloud Providers?**

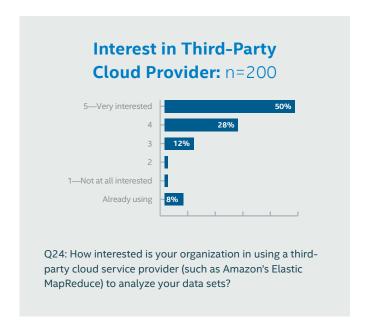
When it comes to considering the use of third-party cloud services to analyze data sets, the majority of respondents (78 percent) show strong interest. This represents slight growth from our 2012 survey, in which 71 percent showed strong interest. Those with a formal IT strategy in place for big data are significantly more likely to say they are very interested in a third-party cloud service.

Overall, those already using a third-party cloud provider to analyze their data sets remained constant from 2012 to 2013, at 8 percent of those surveyed. Larger organizations are significantly more likely to already be using a thirdparty cloud service provider for big data analytics (11 percent vs. 4 percent).

Key Finding

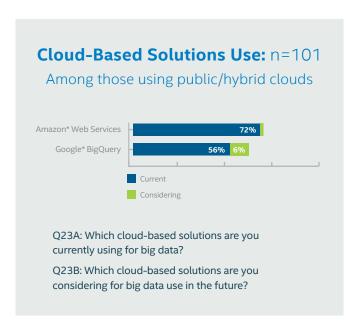
Compared to 2012, interest in using a third-party cloud service provider to process big data is growing.

Also, large organizations are more likely to already be using a third-party cloud service provider.



What Types of Cloud-Based Solutions Are IT Managers Using?

For those using public or hybrid clouds, Amazon* Web Services (72 percent) has a strong share in terms of respondents' cloud-based solutions used for big data.







Big Data Challenges

Big data analytics represents a new frontier and gamechanging opportunities for organizations that are able to use it to gain competitive advantage. By processing huge volumes of real-time data from various sources, organizations can make time-sensitive decisions faster than ever before,

monitor emerging trends, course-correct rapidly, and jump on new business opportunities. Along the way, challenges and obstacles can arise. Our survey group cited a few hurdles as they work to reach the point where they can gain insight from their big data efforts.

What Do Your Peers Cite as Their Big Data Challenges?

The following keyword density map details the greatest challenges respondents associate with big data, with the relative size denoting the frequency of mentions. The most commonly mentioned challenges center around data security (17 percent), data storage (14 percent), and data analytics (11 percent).

Also mentioned were data management challenges such as data compression and organization, infrastructure challenges such as network congestion, processing power, and operational challenges such as cost and time restraints.



Q13: In your mind, what is the greatest challenge associated with big data?



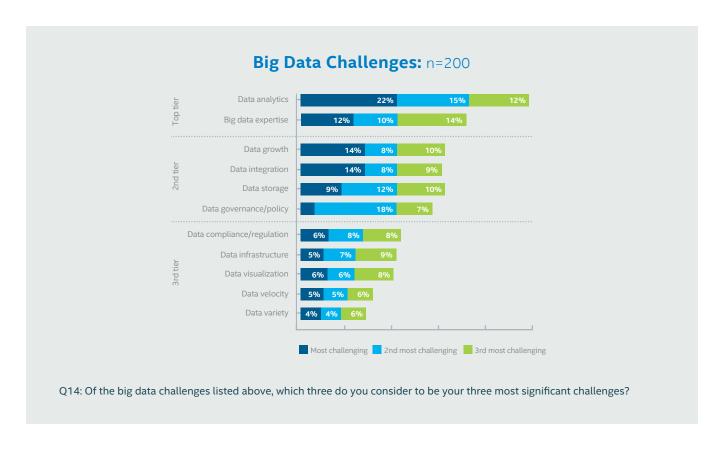




What Are the Top Challenges Today?

When presented with a list of potential big data challenges, respondents overwhelmingly selected "data analytics" as the most significant, followed by "big data expertise within the company."

While our 2013 survey's top two answers were not tested last year, other challenges were rated similarly year over year, in terms of their relative positions, as "data growth" topped the list in 2012.







What Obstacles Do Respondents Face with Big Data?

Among our survey respondents, a clear top tier of obstacles emerges when it comes to big data analytics, most notably a shortage of skilled data science and analysis professionals and security concerns.

Compared to 2012, no two attributes have seen as much change as these two. While security concerns was by far the top mention in 2012 (32 percent "greatest obstacle," 27 percent second and third), a shortage of skilled professionals was a middling concern. This has now garnered double the amount of "greatest obstacle" mentions among larger companies of 1,000+ employees (11 percent to 22 percent).

A shortage of skilled professionals appears to be trending as a major concern, as respondents also rated "data expertise within the company" as a top big data concern.

Key Finding

Concern over a shortage of skilled data science professionals emerged as a top obstacle, followed closely by security challenges.









The Emergence of the Data Scientist

Data science is an emerging field. Demand is high, and finding skilled personnel can be a challenge.

Data scientists are responsible for modeling complex business problems, discovering business insights, and identifying opportunities. They bring to the job:

- Skills for integrating and preparing large, varied data sets
- Advanced analytics and modeling skills to reveal and understand hidden relationships
- Business knowledge to apply context
- Communication skills to present results

A data scientist may reside in IT or the business unit, but either way, he or she is your new best friend and collaborator for planning and implementing big data analytics projects.







Conclusions: Making Sense of Our Results

Big data continues to be top of mind for IT. Organizations know the importance of being able to gain deeper, richer insights that can speed decision making and identify new strategic initiatives, and they are moving forward with big data analytics projects at a strong pace.

Our benchmark of 200 IT professionals again shows that big data is a leading strategic priority. Not only do many companies already have a strategy in place for addressing big data, they are actively fielding big data requests from a user base that increasingly knows how it wants to use big data analytics to gather insight. There is also clear growth in understanding, requests, and the use of big data analytics, when compared to a little over one year ago with our previous survey group.

Important to the ability to process vast amounts of structured and unstructured data quickly and efficiently, both commercial and open-source deployments of the Hadoop framework are proliferating in larger companies. And as these companies activate big data, the cloud is a prominent solution for them—mainly a private cloud, but that, too, is evolving.

Challenges and obstacles that IT professionals face are evolving as well. Where last year, security and the rate of data growth topped the list, today, a shortage of skilled professionals has emerged as a real concern for our survey group. Still, the value of big data is clear, and organizations are moving forward in spite of these challenges.

And they're using big data in a variety of ways—to evaluate staffing levels and productivity, generate competitive intelligence, improve pricing and lower IT costs, and more. And those uses will expand over time as well, to include understanding ways to improve operational efficiency, identify new revenue resources, and enhance product development, to name a few.

We provided the information in this report to help you learn from the experience of your peers as you plan and implement big data analytics. For more information about Intel and big data analytics, visit http://intel.com/bigdata.

Learn More about How Intel Can Support Your Big Data Analytics Needs

Intel's vision for distributed analytics and our leadership role in the development of standards and best practices are helping to move the industry forward so that organizations can realize the full promise of big data analytics. Count on Intel for the technology, guidance, and vision to make big data work for you.

We offer planning guides, peer research, use cases,

industry analyst insights, and live events to help you get started. We also share practical guidance to help you get the most out of your Apache Hadoop* deployment through performance optimization, assistance in selecting the best hardware and software, and configuration and tuning tips, including an Intel® Cloud Builders reference architecture for Apache Hadoop implementations. Find







Appendix: Methodology and Audience

Respondents were screened to ensure that they:

- Perform a wide range of IT-related job functions.
- Work in a company of 500 or more employees.
- Work in a company with at least 100 physical servers that store at least 10 TB of data.
- Are currently involved in a big data analytics project (support or use).
- Currently have, or plan to have, a formal strategy in place for dealing with big data analytics.
- Work in a U.S. company—the only region targeted for this research.

Being an Intel customer was not a consideration for inclusion in the survey. Quotas for company size and industry were enforced to ensure a representative sample.

Respondent Profile Information

Physical Servers	n=200
100–199 physical servers	58%
200 or more physical servers	42%

Virtualized Servers	n=169
Average % virtualized	47%

Weekly Data Generation	n=200
Less than 500 GB	14%
Between 500 GB and 10 TB	20%
Between 10 TB and 100 TB	14%
Between 100 TB and 500 TB	26%
Between 500 TB and 1 PB	22%
More than 1 PB	2%
Unsure	2%





Stored Data	n=200
10–19 TB	20%
20–49 TB	41%
50 TB or more	39%

Job Role	n=200
IT manager	32%
IT director	20%
CIO	15%
Chief technology officer (CTO)	10%
Senior IT manager	7%
Manager of IT operations	6%
VP of IT	6%
Other	3%

Reports to: (multiple mention)	n=200
CIO	49%
Chief technology officer (CTO)	40%
VP of IT	38%
Other	4%

IT Specialists Reporting	n=200
2	2%
3–5	14%
6–9	20%
10–19	42%
20 or more	20%



Responsibilities (multiple mention)	n=200
Evaluating, recommending, or making decisions for vendor selection for key components of server technology strategy such as operating systems, hardware, and applications	100%
Leading a team of IT specialists to make detailed recommendations, and decision making to support business intelligence initiatives	97%
Participating in strategic planning, implementation, and maintenance of server technology	94%
Supporting infrastructure for large amounts of corporate transaction data and any business intelligence initiatives	94%
Working with most senior IT management (CIO, chief technology officer, etc.) to set strategic IT direction for the company	93%
"Hands-on" implementation responsibilities	84%

Worldwide Locations	n=200
1 location	2%
2–4 locations	8%
5–9 locations	19%
10–14 locations	25%
15–19 locations	16%
20 or more locations	30%
Unsure	<1%

Annual Revenue	n=200
\$1M-\$3.9M	2%
\$4M-\$9.9M	4%
\$10M-\$49.9M	30%
\$50M-\$99.9M	16%
\$100M or more	46%
Unsure	2%





Company Size	n=200
500–999 employees	41%
1,000+ employees	59%

Industry	n=200
Manufacturing	16%
Transportation and logistics	15%
Financial services	14%
Retail	14%
Healthcare	8%
Computer-related business/service	8%
Professional services	6%
Construction	6%
Education	4%
Telecommunications	4%
Wholesale and distribution	2%
Utilities	1%
Others	3%





More from the Intel® IT Center

Peer Research: Big Data Analytics is brought to you by the Intel® IT Center, Intel® program for IT professionals. The Intel IT Center is designed to provide straightforward, fluff-free information to help IT pros implement strategic projects on their agenda, including virtualization, data center design, cloud, and client and infrastructure security. Visit the Intel IT Center for:

- · Planning guides, peer research, and solution spotlights to help you implement key projects
- · Real-world case studies that show how your peers have tackled the same challenges you face
- Information on how Intel's own IT organization is implementing cloud, virtualization, security, and other strategic initiatives
- Information on events where you can hear from Intel product experts as well as from Intel's own IT professionals Learn more at intel.com/ITCenter.

Share with Colleagues









Legal

This paper is for informational purposes only. THIS DOCUMENT IS PROVIDED "AS IS" WITH NO WARRANTIES WHATSOEVER, INCLUDING ANY WARRANTY OF MERCHANTABILITY, NONINFRINGEMENT, FITNESS FOR ANY PARTICULAR PURPOSE, OR ANY WARRANTY OTHERWISE ARISING OUT OF ANY PROPOSAL, SPECIFICATION, OR SAMPLE. Intel disclaims all liability, including liability for infringement of any property rights, relating to use of this information. No license, express or implied, by estoppel or otherwise, to any intellectual property rights is granted herein.

Copyright © 2014 Intel Corporation. All rights reserved. Intel, the Intel logo, and the Look Inside. logo are trademarks of Intel Corporation in the U.S. and/or other countries.

*Other names and brands may be claimed as the property of others.

