Intel IT Center

AUGUST 2012

Peer Research
Big Data Analytics

Intel's 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 about big data analytics that can help you plan your own projects, as well as a perspective on what these results mean for the IT industry, including:

- Many IT managers consider big data analytics projects one of the most important imperatives for their organization.
- Adoption of big data analytics tools such as the Apache Hadoop* framework and commercial distributions of Hadoop* is growing, with 25 percent of our survey group having already implemented these technologies and another 20 percent being in the process of deploying at the time of the survey.
- IT managers currently use both batch and real-time delivery of analytics equally, with a 50-50 split, but expect that to increase to two-thirds using real-time processing and delivery by 2015.
- Security is a huge concern for IT managers and a barrier to moving and sharing data between their Hadoop environment and external partners like analytics service providers.
- IT managers place high value on getting help in three areas: deploying data visualization methods for business users, services to manage data from sensors and devices at the edge of networks, and software built for big data frameworks like Hadoop (along with deployment guides and other practical guidance on how to get up and running smoothly).

Intel IT Center

AUGUST 2012

Peer Research

Big Data Analytics

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

Sponsors of Tomorrow.™ intel

Contents

About This Report

Big data analytics is one of the great new frontiers of IT. Data is exploding so fast and the promise of deeper insights is so compelling that IT managers are highly motivated to turn big data into an asset they can manage and exploit for their organizations. Emerging technologies such as the Hadoop* framework and MapReduce offer new and exciting ways to process and transform big data—defined as complex, unstructured, or large amounts of data—into meaningful insights, but also require IT to deploy infrastructure differently to support the distributed processing requirements and real-time demands of big data analytics.

The aim of this report is to provide benchmark data you can use for your own big data analytics planning.

We surveyed 200 IT managers[1] in large companies to find out how they were approaching big data analytics. We asked them about the challenges they face and the technologies they have adopted to capture and analyze new sources of semistructured and unstructured data. We also wanted to measure interest in software and infrastructure designed for big data analytics and determine which tools IT managers deemed most helpful in moving forward with big data projects in the future.

The results of our survey are detailed in this report. The goal is to provide you with data that can serve as a benchmark for how your peers are approaching big data analytics for use in your own IT planning efforts.

1 Our survey was conducted May 10 to 18, 2012. The sample group included 200 IT managers from U.S. companies with 1,000 or more employees. They work in a company with at least 100 physical servers.

Executive Summary

Big data is here. The results of this survey clearly indicate a great deal of excitement and activity around planning and implementing big data environments. The IT managers surveyed for this report represent large companies currently involved in big data analytics.[2] They come from organizations generating a median of 300 terabytes (TB) of data weekly and represent a mix of verticals. They are being asked to manage a wide range of data sources, and most have already gone beyond structured data to begin processing unstructured and semistructured data. Almost half are implementing or have already implemented big data tools and technologies.

Our group identified a number of challenges and obstacles to successful big data implementations. Security is called out loud and clear, an issue that is most problematic when it comes to worries about data security and privacy related to the use of third-party cloud service providers. However, when it comes to big data analytics, companies face a number of different challenges, including infrastructure and data governance and policy issues. Despite this, IT managers expressed genuine excitement about big data analytics and consider it a top IT priority for their organizations.

Key findings from our study include:

"Highly engaged" IT managers in big data analytics share important characteristics. One-third (33 percent) of companies surveyed are working with very large amounts of data. This group is "highly engaged" with big data analytics, and individuals share a number of important characteristics, including:

- Working with 500 TB or more of data per week
- Prioritizing big data analytics—very high importance is attached to improving these capabilities
- Working from a formalized strategy for big data
- Currently processing unstructured data sources

The top data source continues to be business transactions—but that's changing. While data comes from a variety of sources—both structured and unstructured—most IT managers cited business transactions in a database as the top source. However, general business documents, e-mail, sensor or device data, and imaging data—all unstructured data—were in the top five.

Most IT managers are processing unstructured data of some type. Some 84 percent are currently analyzing unstructured data, and 44 percent of those that aren't expect to do so in the next 12 to 18 months.

Adoption of big data tools and technologies continues to grow. Slightly less than half are implementing or have already adopted big data tools or technologies, with one-third reporting they are in the evaluation stage.

IT managers predict that most big data analytics will be in real time by 2015. While batch versus real-time data analytics is currently split 50-50, respondents predict that by 2015, nearly two-thirds (63 percent) of all analytics will be done in real time.

Key Finding:

A third of our survey group can be viewed as "highly engaged" in big data analytics.

2 Our survey defines big data analytics as data sets whose size and variety is beyond the ability of typical database software to capture, store, manage, and analyze—in other words, data that is high volume, complex, or semistructured or unstructured.

Three challenges are considered equally problematic. Data growth, infrastructure, and governance and policy concerns are three equally difficult challenges faced by IT organizations who want to implement big data analytics solutions.

Security is a big deal when it comes to concerns about third-party cloud solutions. Ninety-five percent of IT managers who reported little to no interest in the use of third-party cloud solutions ranked data security and privacy as the top obstacle to moving their data externally to cloud service providers. Related to this same issue, IT managers are bothered by the lack of data security standards and specific technology designed to keep customers' data private.

Most IT managers say they understand big data demands. IT managers have been doing their homework, with most (80 percent) saying that they have a good understanding of the scope of the big data demands from business users and analytics experts.

Interest is very high in learning about how to deploy big data infrastructure. The majority of our sample group (80 percent) has plans to adopt infrastructure and software that will support high data volumes, data variety and complexity, and velocity issues. IT managers indicate very high interest in learning about that technology as well as how to deploy and adopt such infrastructure for performance and cost.

Key Finding:

IT managers predict that two-thirds of big data analytics will be in real time in just three years.

Big Data and IT Today

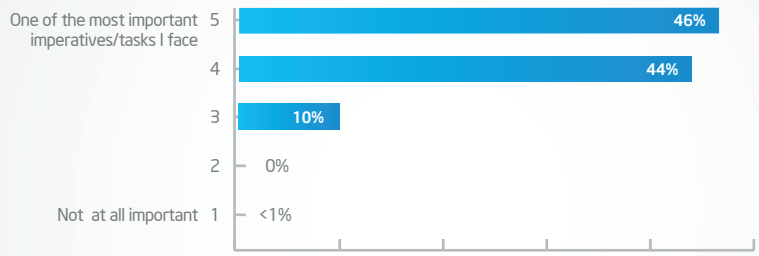
Big data is growing fast, but so is interest on the part of IT managers to harness this relatively untapped resource and drive value for their organizations. Our research looked specifically for companies handling large volumes of data. We wanted to know from those already under intense data management pressures where they were in their approach to big data.

Big Data: The Most Important Tasks IT Managers Face

We found out that most of our survey group sees big data analytics as a high priority in their IT environments. Ninety percent rated improving data analytics capabilities as *very important*—and almost half (46 percent) rated improving data analytics capabilities as *one of the most important imperatives/tasks I face*.

Those rating the importance of improving capabilities with a "5" are significantly more likely than others to be asked to analyze predominately web-based data (search indexing, weblogs, social media, and clickstreams). They are also more likely to have already deployed each tested type of big data tools or technologies (the Apache Hadoop* framework, commercial distributions of the Hadoop framework, and other NoSQL databases).

Improving Data Analytics Capabilities (*n=200*)

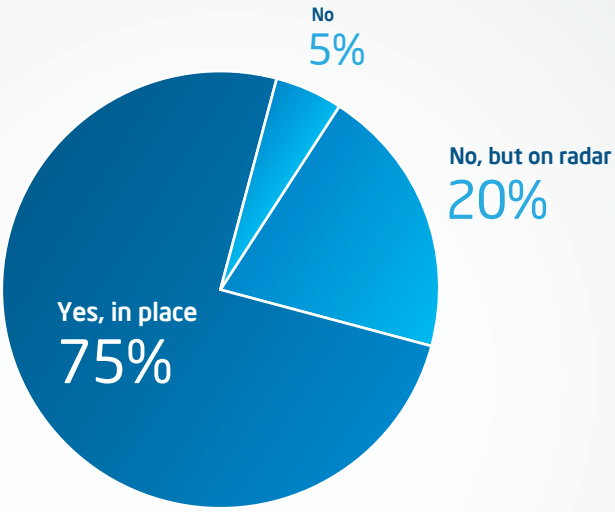


Q: *How would you rate the importance of improving your data analytics capabilities for your business?*

Formal Big Data Analytics Strategy: The Place to Start

Our survey group is clear that a formal strategy for handling big data should be in place in the organization. Three-quarters of the surveyed companies have a formal strategy for dealing with big data analytics. These companies demonstrate much greater data needs and demands, generating roughly six times more weekly data than those without a formal strategy. They are also analyzing data from a large variety of data sources (both structured and unstructured). This same group is significantly more likely than those without a strategy to be in the process of implementing or to have already deployed open-source Apache Hadoop frameworks, commercial distributions of the Hadoop framework, and other NoSQL databases.

Formal Big Data Strategy (*n=200*)



Q: *Does your IT department currently have a formal strategy for dealing with big data analytics?*

Varied Data Types: Business Transactions Still Top Source

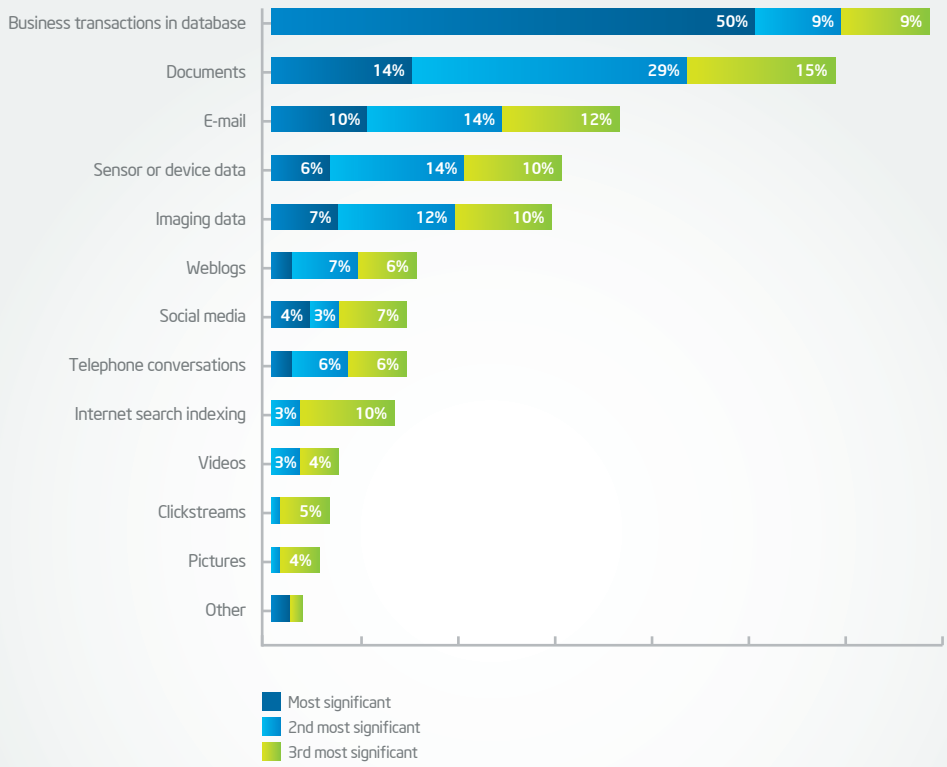
Companies are being asked to analyze a broad range of data. As you can see in the table below, between 42 percent and 84 percent are being asked to analyze data types ranging from clickstreams and telephone conversations to imaging data and business documents.

Data Source	Total % Asked to Analyze
Documents	84%
Business transactions in database	82%
E-mail	74%
Imaging data	68%
Sensor or device data	57%
Internet search indexing	57%
Weblogs	55%
Social media	54%
Telephone conversations	52%
Videos	52%
Pictures	46%
Clickstreams	42%
Other	2%

Q: *Which of the following data sources are you being asked to analyze and glean insights from?*

When asked to rank their top three data sources, IT managers placed business transactions at the top of the list, traditional structured data that continues to be of critical importance to companies despite the buzz about unstructured data types. However, the top five also includes four semistructured or unstructured sources: documents, e-mail, sensor or device data, and imaging data. Our "highly engaged" group of IT managers is more likely than others to be asked to analyze nearly all of the data sources listed.

Top Three Sources of Data Analysis (*n=200*)

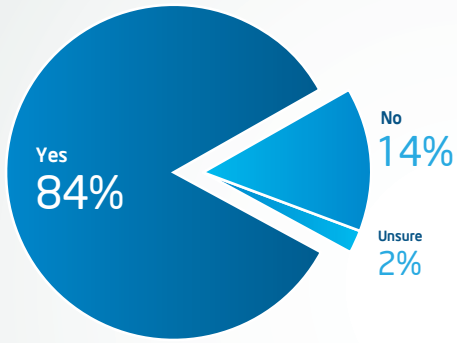


Q: *Of the data sources you're asked to analyze, which are the top three sources?*

Unstructured and Semistructured Data Analytics Growing

More than four in five (84 percent) are currently processing and analyzing unstructured data sources , including weblogs, social media, e-mail, photos, and video. Of those who are, or expect to be, analyzing unstructured data, documents and e-mail are mentioned as the most important for their organization to analyze in the next 12 to 18 months. Social media, weblogs, sensors and devices, and images make up a second tier in data sources expected to be important in the next 12 to 18 months.

Currently Analyzing Unstructured Data (*n=200*)



Among those not currently analyzing unstructured data:

- 44 percent expect to be asked in the next 12–18 months.
- 12 percent say they will not be asked to do so.
- 44% are unsure whether they will be asked.

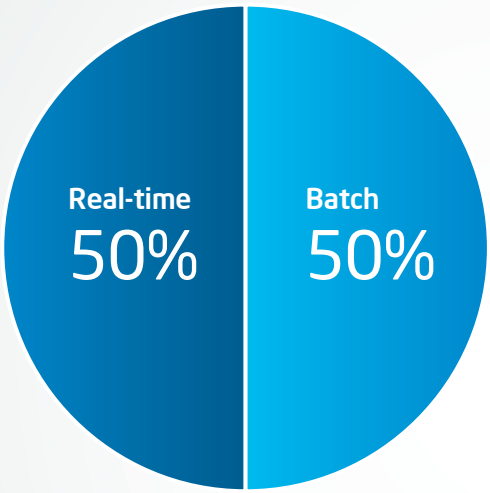
Q: *Are you currently processing and analyzing unstructured data sources to get new insights?*

Q: *Will you be asked to process and analyze unstructured data sources in the next 12 to 18 months?*

Analytics Delivery: Moving from Batch to Real-Time

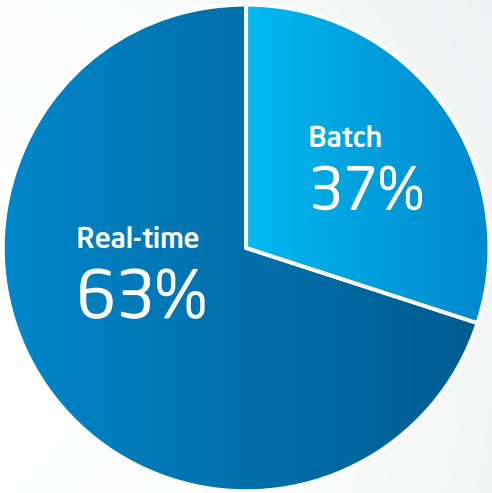
IT managers are using both batch and real-time delivery of analytics equally, with a 50-50 split. However, that's changing—companies are expecting to do almost two-thirds of their analytics in real time by 2015. Our highly engaged IT managers are already conducting significantly more of their data analytics in real time and plan to do even more.

Current Data Analytics
(*n=200*)



Q: *What percentage of your company's current data analytics is done in real time, and what percentage is done via batch processing?*

2015 Data Analytics Projections
(*n=200*)



Q: *What do you predict will be your company's percentage of real-time versus batch data analytics in 2015?*

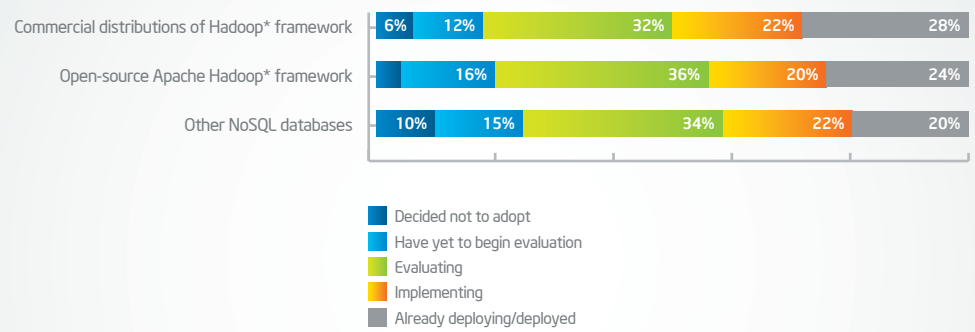
What's Holding IT Back?

Our snapshot of IT organizations under big data pressure shows a solid percentage of companies that have implemented or are in the process of implementing big data technologies. What's holding the rest of them back? IT managers cite a number of challenges and obstacles to overcome first before they can step up their big data game.

Adoption of Big Data Analytics Is Growing

Across our entire sample group of companies, approximately 25 percent have implemented open-source versions of the Apache Hadoop framework or commercial distributions. Based on our findings, we expect adoption of these tools to continue to grow in large companies. At the time of the study, another 20 percent were also in the process of implementing these big data tools. Few have decided not to adopt these tools, and one-third are currently evaluating them for the future.

Adoption of Big Data Tools (*n=200*)



Q: *For each of the big data tools or technologies listed, please indicate your current stage of adoption and development.*

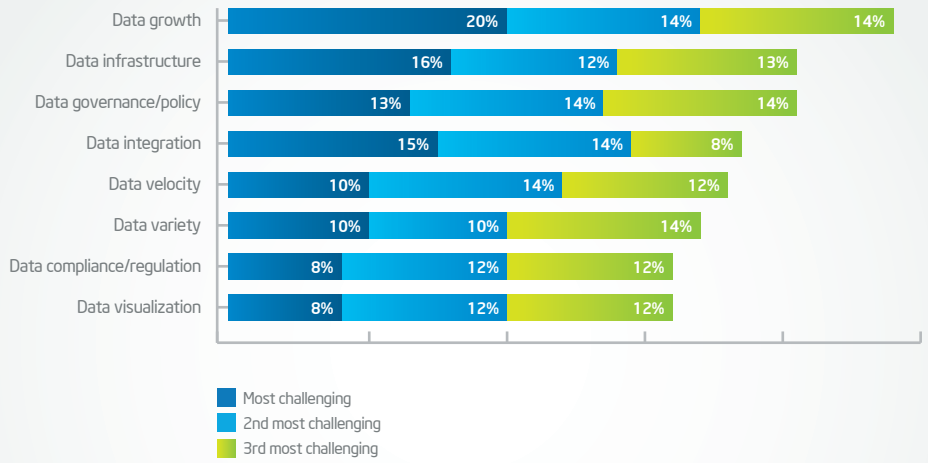
Several Challenges and Obstacles with Big Data

When we asked IT managers about their top big data challenges, we found that most were rated similarly in terms of their significance. This indicates that respondents face more than one challenge with big data. The top challenge cited was not surprising: data growth and the cost and effort required to contain or store it. Forty-eight percent cited this as one of their top three challenges. However, this is closely followed by:

- Concern about data infrastructure and the ability of their data center to provide the scalability, low latency, and performance needed to handle big data projects (41 percent)
- Data governance and policy challenges for defining the data that will be stored, analyzed, and accessed, along with determining its relevance (41 percent)

Only slightly less, ranging from 27 to 32 percent, chose their top three challenges among the remaining five. The results are clear. IT managers face significant big data challenges from several directions.

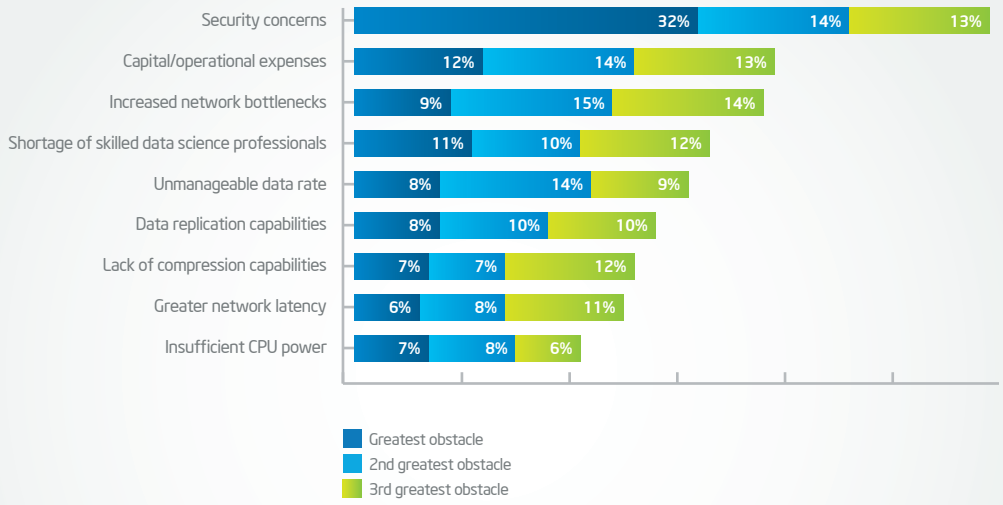
Big Data Challenges (*n=200*)



Q: *Of the big data challenges listed, which do you consider to be your three most significant challenges?*

When asked about specific obstacles to big data analytics, IT managers see security trumping them all. We believe this is driven by IT managers who desire to work with third-party cloud service providers but at the same time worry about those providers' abilities to protect their data. When asked about third-party cloud service providers, an overwhelming number identified this as an issue. Interestingly, concerns about security are consistent across company segments regardless of weekly data volume, number of physical servers, or revenue.

Obstacles to Big Data Analytics (*n=200*)



Q: *Of the obstacles listed above, which do you consider to be the three most significant as they relate to big data analytics?*

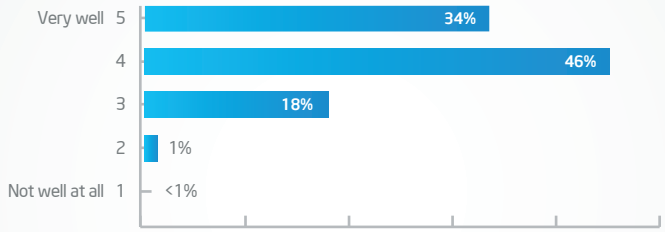
Moving Forward with Big Data Analytics

What do IT managers need to move forward? Interestingly, the IT managers in our study are comfortable that they understand the big data analytics needs of their organizations. They have plans to adopt infrastructure to fulfill those needs, and they show high interest in using third-party cloud providers. IT managers are looking for support in a number of different areas, including innovative solutions for specific applications and practical tools to help with deployment. Development of industry standards for data security and privacy as well as interoperability are also cited as important.

The Scope of User Demands for Big Data: IT Managers Get It

When asked whether they understand the scope of demand for big data from business users and data analytics experts in their organizations, 80 percent of our sample thought they did. The vast majority of that group (89 percent) also have plans to adopt infrastructure and software to support big data requirements, including data volume, velocity, and variety.

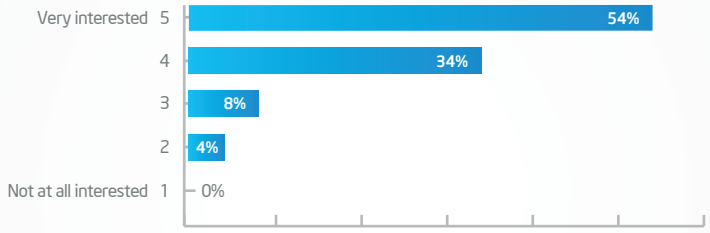
Understand User/Expert Demands (*n=200*)



Q: *As of today, how well do you understand the scope of big data demands from your business users and data analytics experts?*

IT managers show very high interest in learning more about infrastructure and software built to support big data analytics.

Interest in Software for Big Data Analytics (*n=200*)

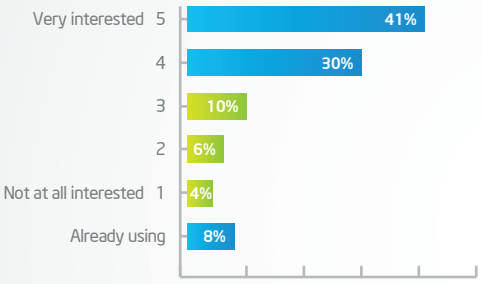


Q: *How interested would you be in learning more about infrastructure and software that is built for big data analytics, and how to deploy and optimize that infrastructure for performance and cost?*

Third-Party Cloud Providers for Big Data Analytics

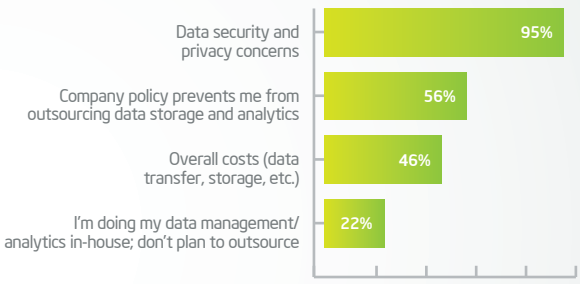
Although 71 percent of those surveyed are interested in third-party cloud service providers for analysis of their data sets, only 8 percent are already using this approach. For those showing little to no interest, security and privacy concerns are by far the most significant (95 percent) barrier to moving data to a third-party vendor for processing, analysis, and storage

Interest in Third-Party Cloud Provider (*n=200*)



Q: *How interested is your organization in using a third-party cloud service provider (such as Amazon's Elastic MapReduce) to analyze your data sets?*

Concerns with Third-Party Cloud Vendors (*n=41*)

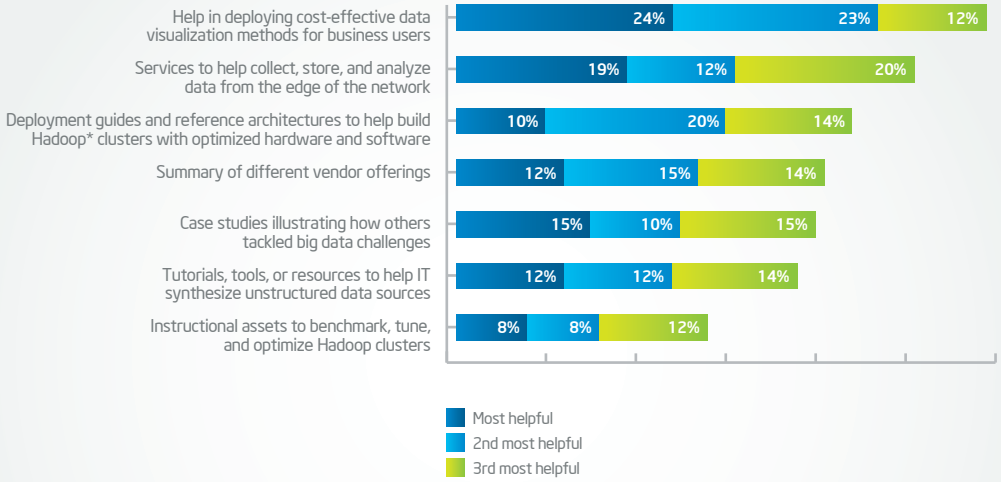


Q: *What are the biggest concerns and/or roadblocks in moving your data to a third-party provider for processing, analysis, and storage?*

High-Value Help: Three Areas

Among the ideas tested, respondents would find the most value in help deploying cost-effective data visualization methods as well as services to collect, store, and analyze data coming from sensors or other embedded devices on the edge of the network. IT managers are also interested in assets such as deployment guides, reference architectures, and other content that would help them build high-performing Hadoop clusters.

Help in Moving Forward with Big Data Projects (*n=200*)

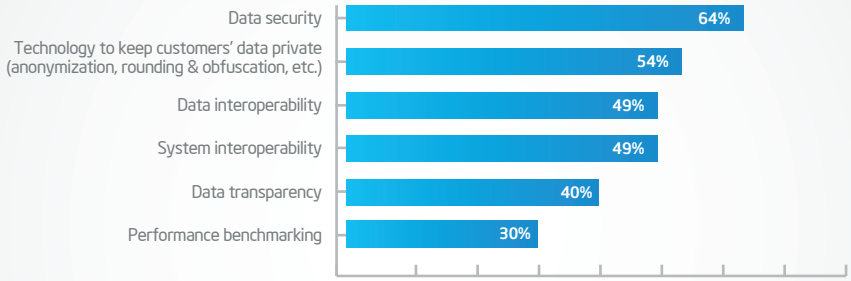


Q: *Of the ideas listed, which three would you find most helpful in moving forward with big data projects?*

Industry Standards Needed: Security, Privacy, Interoperability

We asked IT managers what standards they would like to see addressed for big data analytics. Data security is the number one issue, most likely due to the desire to work with third-party cloud service providers. Closely behind data security is a desire for standards related to data privacy, including anonymization, rounding, and obfuscation as well as interoperability for data and systems.

Addressing Industry Standards (*n=200*)



Q: *What standards for big data analytics are you most interested in having the industry address?*

Your Peers Speak Out

We asked those surveyed for the single most important piece of advice they could offer to another IT manager just starting the process of planning a big data analytics project, and then we grouped the answers under similar categories. Experienced IT managers stressed that it was critical to do thorough research and detailed planning. Due diligence is required for big data analytics. They would remind IT managers to prioritize security and plan for future needs, especially in terms of their data center capabilities. Plus, they would caution IT managers to expect some frustrations and to bring a healthy dose of patience to their big data analytics undertakings. The following keyword density map shows the occurrence of specific advice categories. It's followed by examples of verbatim responses from the IT managers in this survey.

IT Manager Advice for Big Data

Word Cloud Showing Frequency of Advice

Q: *In your experience evaluating, planning, and/or implementing big data analytics projects in your IT environment, what is the single most important piece of advice you could offer to another IT manager just starting the process?*

Due diligence required:

"Research, research, and do more research. This is not something you should jump into without researching how it will affect or help your company, what vendors can help you, and what products or services are best for you."

"Clearly identify the project objectives, with the business managers and technical staff signing off on each step."

"Take time and digest the information that is out there, and if possible, speak with other firms in the industry to gauge where they are headed."

Don't forget security and privacy:

"Don't underestimate the importance of security."

"Ensure that your data remains private at all times. There's no excuse for any type of susceptibility to data getting into the wrong hands."

Plan for the future:

"Ensure that a sound infrastructure is in place: top-quality network and SAN [storage area network] fabric switches, current-level servers and storage, and support agreements current for everything."

"Be prepared at all times to review and renew your data center capability. Just when you thought you have caught up, you need to move up again."

"Be agile … The goalposts are changing fast!"

Don't give up:

"Don't give up. Make sure you understand all your resources, and plan accordingly."

"Be patient, 'cause there's lots to learn."

Conclusions: Making Sense of Our Results

Big data is already here, and organizations are under considerable pressure to understand it, mine it, and exploit it for richer, deeper insights. Our benchmark report of 200 IT professionals confirms what other studies have shown: The strategic value of big data analytics is compelling enough for organizations to move forward despite challenges and concerns.

Our study deliberately surveyed larger companies that generate large volumes of data—a median of 300 TB per week. Managing big data is already top of mind for this group, who can give us valuable insight into what IT managers are actually planning and implementing now and where they see their organizations going with big data in the future.

Insights into Big Data

Most of our respondents are beyond the early evaluation stages with big data analytics, so their experiences go beyond speculation. They are already handling a variety of unstructured data types from a broad range of sources, and a substantial number have implemented big data technologies like the Hadoop framework. While structured data mining continues to be critically important to organizations, our survey group expects to be managing more and more unstructured and semistructured data—especially images and data from sensors and devices. Similarly, they predict that real-time analytics will grow to be the majority of their workload by 2015—three years from now.

We found that IT managers put a great deal of importance in having a formal strategy for big data. This makes sense, since the issue of data governance and policy for defining what data is stored, analyzed, and accessed was called out as one of the three top challenges they face. Worry about whether their infrastructure is capable of supporting demands for scalability, low latency, and performance in big data environments was another. Help understanding how data center infrastructure needs to evolve to support and take advantage of big data implementations is one way to address this concern.

Also, security is called out as a significant obstacle to implementing big data analytics projects—in particular for those projects that require moving sensitive data to third-party cloud providers. No matter how we asked the question, system and data security and privacy rose to the top as issues in our survey group. While organizations are handling more and more big data, we conclude that without significant improvements in security and privacy technologies for big data environments, IT managers will be cautious about moving as much data as they would like into external big data analytics environments, whether that's in a centralized cloud or a distributed architecture housed off-site.

Next Steps for IT

What's the industry to do? Our survey provides us with some clear answers. IT managers are highly motivated and interested to learn about infrastructure and software built for big data analytics. The survey group identified three areas that would help them move forward with big data analytics projects. These include a focus on:

- Cost-effective deployment of data visualization methods
- Services to help collect, store, and analyze data from intelligent systems and sensor and device data at the edge of the network (one of the fastest growing sources of big data)
- Software built for big data environments like the Hadoop framework, along with deployment guides, reference architectures, and other practical guidance on how to get up and running smoothly

Development of standards for big data environments is another critical area for industry focus. Industry groups must work on defining usage model requirements and best practices to support:

- Data security in big data environments
- Data privacy, including anonymization, rounding, and obfuscation
- Interoperability for data and systems, including standards to make big data frameworks cloud ready

The result of these industry efforts will be to enable organizations to take best advantage of infrastructure for big data analytics at both the hardware and software levels. With enterprises today already utilizing some components of a big data environment—intelligent systems and devices, visualization software, and cloud services, for example—the Hadoop framework and other big data tools must develop in a direction that supports interoperability, usability, and manageability while addressing concerns about capital expenses.

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 intel.com/bigdata.

Big Data Analytics: Count on Intel for Technology, Guidance, and Vision

Big data is a game changer. Intel can help organizations get the most from their implementations with platforms that deliver the exceptional performance, low latency, and high throughput needed to handle large data sets and transform them into deep insights. Plus, we offer practical guidance to help deploy big data environments more quickly and with lower risk. At the same time, 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.

Learn more about how Intel can support your big data analytics initiatives at intel.com/bigdata.

Appendix: Methodology and Audience

Responses to this survey were gathered via an online questionnaire; 200 surveys were received between May 10 and May 18, 2012. A sample size of 200 has a maximum sampling variability of ±6.9 percent at the 95 percent confidence level.

Respondents were screened to ensure that they:

- Perform a wide range of IT-related job functions.
- Work in a company of 1,000 or more employees.
- Work in a company with at least 100 physical servers that stores at least 10 TB of data.
- Are currently involved in big data analytics (support or use).
- Represent what Intel terms a "tech enthusiast" company—that is, a company that considers IT to be a driver of their business success.
- 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 Installed	n=200
100–199 servers	38%
200 or more servers	62%

% of Servers Virtualized	n=200
Less than 25%	12%
25%–49%	22%
50%–74%	34%
75% or more	16%
Unsure	16%

Total Stored Data	n=200
10 TB–19 TB	6%
20 TB–49 TB	26%
50 TB or more	68%

Data Generated Weekly	n=200
500 GB–9 TB	15%
10 TB–99 TB	18%
100 TB–499 TB	30%
500 TB–1 PB	30%
More than 1 PB	2%
Unsure	4%

Worldwide Locations	n=200
1 location	2%
2–4 locations	8%
5–9 locations	12%
10–14 locations	14%
15–19 locations	12%
20 or more locations	51%
Unsure	2%

Annual Revenue	n=200
Less than $500,000	1%
$500,000–$0.9M	–
$1M–$3.9M	2%
$4M–$9.9M	2%
$10M–$49.9M	10%
$50M–$99.9M	12%
$100M or more	66%
Unsure	6%

Industry	n=200
Computer-related business/service	15%
Financial services	14%
Manufacturing	14%
Healthcare	8%
Professional services	8%
Retail	7%
Education	6%
Government	6%
Telecommunications	6%
Wholesale & distribution	4%
Construction	3%
Transportation & logistics	3%
Media & entertainment	2%
Utilities	2%
Agriculture, forestry & fishing	<1%
Others	2%

More from the Intel® IT Center

Peer Research: Big Data Analytics is brought to you by the Intel® IT Center, Intel's program for IT professionals. The Intel IT Center is designed to provide straightforward, fluff-free, unbiased 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 vendor 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

0812/RF/ME/PDF-USA 327825-001

Sponsors of Tomorrow.™