VENTANA RESEARCH

Big Data Analytics

Assessing the Revolution in Big Data and Business Analytics

White Paper

Sponsors

NUEVORA
Outthink. Outperform.

pentaho

SmartData Collective
THE WORLD'S BEST THINKERS ON BUSINESS INTELLIGENCE AND DATA ANALYTICS

February 2014

Ventana Research performed this research to determine attitudes toward and utilization of big data analytics. This document is based on our research and analysis of information provided by organizations that we deemed qualified to participate in this benchmark research.

This research was designed to investigate big data analytics practices and needs and potential benefits. It is not intended for use outside of this context and does not imply that organizations are guaranteed success by relying on these results to improve big data analytics. Moreover, gaining the most benefit from big data analytics requires an assessment of your organization's unique needs to identify gaps and priorities for improvement.

The full report with detailed analysis is available for purchase. We can provide detailed insights on this benchmark research and advice on its relevance through the Ventana On-Demand research and advisory service. Assessment Services based on this benchmark research also are available.

We certify that Ventana Research wrote and edited this report independently, that the analysis contained herein is a faithful representation of our evaluation based on our experience with and knowledge of big data and analytics, and that the analysis and conclusions are entirely our own.

Ventana Research
2603 Camino Ramon, Suite 200
San Ramon, CA 94583-9137
info@ventanaresearch.com
(925) 242-2579
www.ventanaresearch.com

Ventana Research

Table of Contents

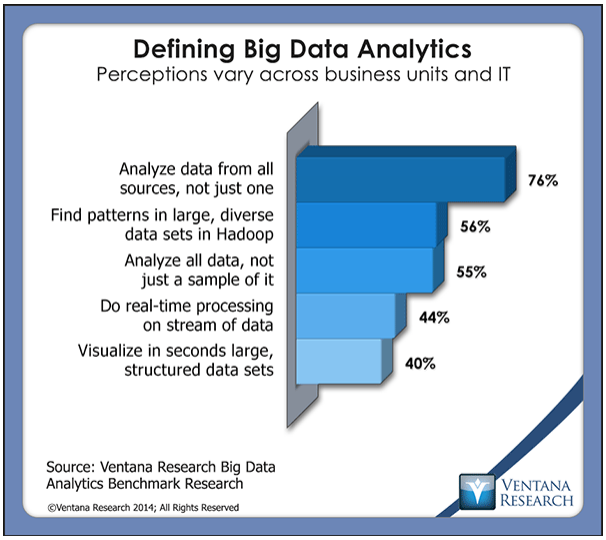
Executive Summary

Big data is of immediate relevance to the situation many organizations face today. They need to store, process and use data in significantly greater volumes than ever before. The data comes from a larger number of sources and often differs in kind from traditional business data, ranging from transactional records stored in databases to free-form text comments from websites and social media. And it needs to be processed much faster than before.

Embracing big data and applying analytics can help organizations resolve many data-dependent management and operational issues encountered by both business units and the IT groups that support them. Among them are recommending better offers to customers, preventing fraud, ensuring security, optimizing networks and forecasting demand.

Big data analytics is a set of processes that includes accessing data sources, producing an analytical data set, applying analytic processes and methods to the data and presenting the analytic results of these processes. Yet how big data works and what benefits it can deliver still aren't clear to many business people.

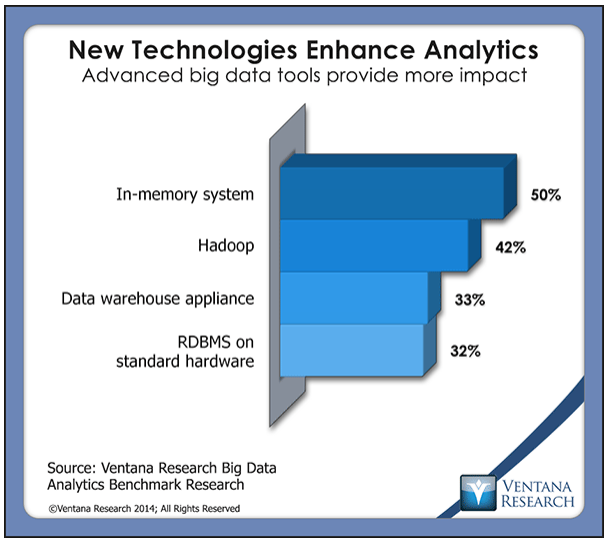
Ventana Research undertook this benchmark research to illuminate these and other issues. We set out to determine the experiences, attitudes, requirements and future plans of organizations that adopt or are considering use of big data analytics and to identify the best practices of organizations that are most mature in it. We set out to examine both the commonalities and the qualities specific to major industry sectors and across sizes of organizations. We explored how organizations perform big data analytics, what they analyze, issues they encounter in the process and the information technology they use.



The research uncovered much uncertainty and confusion. Asked

about the capabilities of big data analytics, majorities of participants chose three definitions: analyzing data from all sources rather than just one, finding patterns in large and diverse data sets using Hadoop (an original big data technology) and analyzing all of the data rather than just a sample of it. Although these are not necessarily mutually exclusive (and we allowed multiple responses), the degree to which participants see big data analytics in different ways is significant. Another finding underscored the likelihood of confusion: Asked about perceptions within their organization, the largest percentage (44%) responded that among those involved in making business technology decisions there are many different opinions about the meaning of big data analytics terminology; only 9 percent said there is total agreement on its meaning.

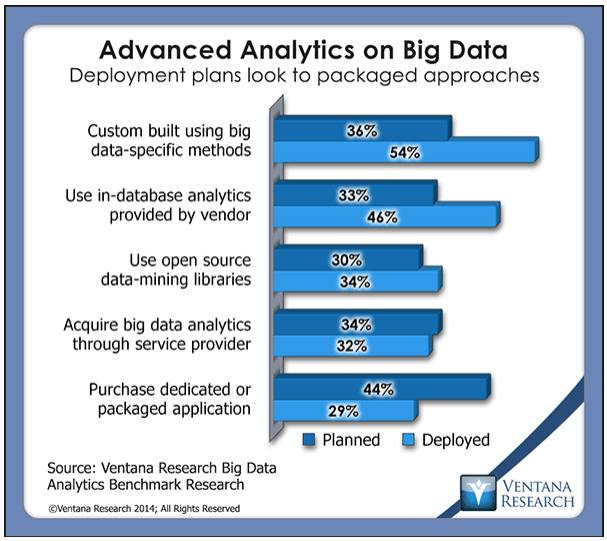
The substantial presence of differing opinions suggests widespread familiarity with the topic, and the research confirms this. Almost half (47%) of participants said that big data analytics is very important. Half use it now. And more than half said they are satisfied or very satisfied with their big data analytics efforts. Further analysis shows that using advanced technology tools, such as Hadoop, database appliances or in-memory systems, correlates with satisfaction more often than does the use of more conventional tools such as an RDBMS on standard hardware. In other words, the research shows, those using analytics tools designed for big data perform better. In particular, users of in-memory systems and Hadoop most often reported significant improvement in the results of their activities and processes from using big data analytics.



Such tools, however, are not yet widespread. Currently just three in 10 organizations (31%) use Hadoop to generate and work with big data analytics, and even fewer (17%) use an in-memory database. By far the most common tool (used by three out of four organizations) is business intelligence technology for query, reporting and analysis. About half use specialized analytic databases, spreadsheets and a relational database management system (RDBMS). Among those three

only the first is built for analytics, and not necessarily on the scale of big data. More than one-third (35%) use custom-built systems.

The custom approach is most evident in organizations that rely on analytics applied to big data. The largest percentage (54%) did so using big data-specific language and interfaces. Others used in-database analytics provided by a vendor or open source data-mining libraries or acquired big data analytics through a service provider. The lowest percentage (29%) purchased a dedicated or packaged application.



This balance is shifting, however. Among those intending to deploy advanced capabilities in the future, the largest number plan to purchase a dedicated or packaged application (44%), while about one-third will use a custom build or acquire the analytics through a service provider.

Successfully applying big data analytics requires technical expertise somewhere in the organization. Among those not satisfied with their current process of creating big data analytics, two-thirds said it is because there aren't enough skilled resources. Packaged applications typically are designed to accommodate less technically adept users than more complex tools and can be a way to close the skills gap.

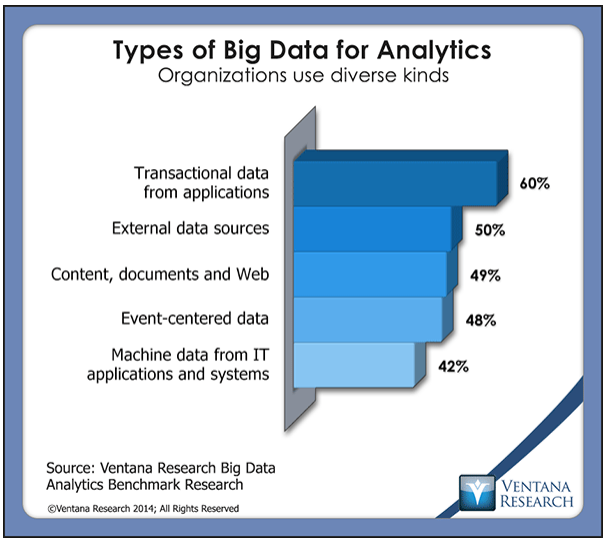
Given the complexity of big data analytics, we recommend involving those who design, deploy and apply the analytics in decisions for purchasing software. IT organizations are most often involved; also involved roughly 20 percent of the time are cross-functional teams, data scientists or data miners, and line-of-business analysts. But the involvement of these latter appears to be of highest value; analysis shows that organizations that use specialized roles such as data scientist or data miner report greater improvement from using big data analytics than others do.

The research makes clear that collaboration is important to big data analytics. One in four ranked communication and knowledge sharing,

which is a prominent aspect of collaboration, first among benefits realized from deploying it, and one-third cited collaborating on review of analytics as a critical feature. We therefore expect to see more embedding of collaborative capabilities into big data analytics products.

As with the lack of adequate skills, two out of five organizations said that data quality and information management issues are a barrier to improvement in big data analytics. Quality and consistency of data also is the part of the process most often cited (by 56%) as presenting challenges; validating data from a business perspective ranked second on this list. Quality issues can undermine confidence in data, and only one in five organizations in this research said they are very confident in the information being generated by big data analytics.

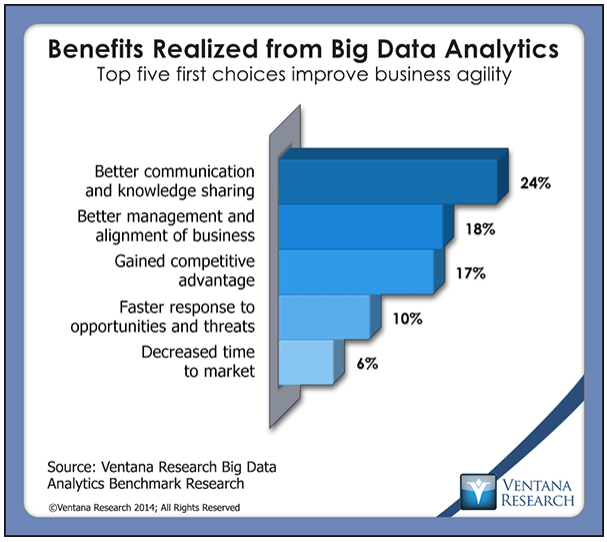
Closely related to data quality is data integration. This is a particular issue with big data because so much must be collected and combined from so many sources. More participants in this research said they are not satisfied with their organization's integration of information for creating big data analytics than are satisfied (47% vs. 40%). Nearly half said that optimizing information is important to their big data analytics efforts. Assuring decision-makers that the information they use is reliable is an essential element of gaining acceptance of big data analytics and support for investments in it.



The research makes clear that information management on this large scale can be daunting. As noted, most participants define big data analytics as analyzing data from all sources of information. They range from conventional structured transactional data to unstructured content such as documents and Web pages to event-centric data. Half of organizations analyze data from external sources, especially cloud computing applications and social media. Larger companies do so more often than smaller ones; for example, two-thirds of very large ones by number of employees use cloud-based data.

The variety of departments interested in big data analytics and of the information they use is matched by the variety of uses to which they put the analytics. Three of the four applications most often mentioned are directed toward customers and sales: enabling cross-selling and up-selling, understanding the customer better and optimizing pricing. These are especially popular with companies in the Services and Finance, Insurance and Real Estate industries. The four mentioned next most often are more internal: to optimize IT operations, analyze fraud in transactions, enable preventive maintenance and improve website usability.

The most frequent benefits from deploying big data analytics, each cited by half of participants, show a similar bifurcation: Communication and knowledge sharing is important to operations and performance while gaining competitive advantage enhances business performance. The next four benefits are split between operational (better management and alignment of business and improved efficiency in business processes) and results-oriented (faster response to opportunities and threats and improved customer experience and satisfaction). When sorted only by first choices, the order shifts somewhat, but the same priorities remain. Thus we conclude that big data analytics has something to offer almost all parts of an organization.



As yet, however, adoption and use of these tools is uneven. Three in 10 have used big data analytics technology for more than a year, and 20 percent more began to do so in the last 12 months. Equal numbers (23% each) will begin using it within 12 months or intend to use it but don't know when they'll start. Along similar lines, about one-third each said either they plan to change the way they assess and select this technology in the next 12 to 18 months, don't plan to change or don't know whether they will change.

Regarding criteria for selecting software to design and deploy big data analytics there was more unanimity: 63 percent of all participants said usability is very important, and more than 90 percent said functionality and reliability are important or very important. Added to the emphasis on collaborative capabilities, this indicates a demand for tools that many people can use without difficulty.

For the time being, big data analytics is likely to remain challenging for many organizations. To enter the business mainstream, vendors will have to design products that satisfy these criteria and be ready to provide assistance in deployment and training. To organizations seeking ways to handle the masses of data unceasingly coming their way, we recommend defining the topic clearly, determining how they can use the technology most profitably and studying successful deployments of others that have similar businesses and needs.

Key Insights

This benchmark research yielded the following important general findings and key insights regarding big data analytics. (We discuss performance levels in the Performance Index portion of the full research report; the actual questions asked in our survey are in an appendix to the research report. Specifics of organization sizes are in the appendix "About This Benchmark Research.")

Organizations are advancing in using big data analytics.

The research shows organizations are aware of and using big data analytics. Almost half (47%) said that big data analytics is very important, and half use this technology now. These findings indicate that this is no longer a market only for early adopters. More than half said they are satisfied (42%) or very satisfied (14%) with their big data analytics efforts today. Satisfaction is understandably more prevalent in organizations that have been using the technology for more than a year (66% satisfied or very satisfied) than in those that implemented it within the past 12 months (44%).

> Those that use conventional technology tools for big data analytics are satisfied less often than those that use more advanced tools such as Hadoop, database appliances or in-memory systems.

One driver of satisfaction is advanced technology. Those that use conventional technology tools for big data analytics, such as a relational database management system (RDBMS) on standard hardware or flat files, are satisfied less often (on average 56% satisfied or very satisfied) than those that use more advanced tools such as Hadoop, database appliances or in-memory systems (64% on average). We conclude that as big data analytics advances, the shortcomings of conventional analytic approaches become more apparent.

Organizations emphasize the "big" aspects of big data analytics.

Asked about what big data analytics means in terms of capabilities, the largest percentage of participants in this research said it involves analyzing data from all sources rather than just one (76%), while for 55 percent it means analyzing all of the data rather than just a sample of it. (We allowed multiple responses.) More than half (56%) told us they view it as finding patterns in large and diverse data sets in Hadoop, which indicates the strong influence this original big data technology still has. A second tier of percentages emphasized timeliness as an aspect of big data: doing real-time processing on streams of data (44%), visualizing large structured data sets in seconds (40%) or doing real-time scoring against a database record (36%). Thus the research identifies the primary characteristic of this technology as analyzing large amounts of data from many data sources. In addition, the importance placed on Hadoop reflects a focus on analyzing anomalous data. Secondarily, big data analytics is about fast processing.

> The research identifies the primary characteristic of this technology as analyzing large amounts of data from many data sources. Secondarily, big data analytics is about fast processing.

Communication and knowledge sharing is an important benefit of these initiatives.

Among organizations planning to deploy big data analytics, the most often mentioned anticipated benefits are faster response to opportunities and threats (57%), improving efficiency (57%), improving the customer experience (48%) and gaining competitive advantage (43%). However, once the big data analytics system has moved into production, the benefits most often mentioned are better communication and knowledge sharing (51%), gaining competitive advantage (51%), improved efficiency in business processes (49%) and improved customer experience and satisfaction (46%).

While the last three realized benefits of deployment are predictable, it's noteworthy that communication and knowledge sharing becomes

one of the two most often mentioned. This benefit of big data analytics initiatives is not a priority before deployment.

In our view, one reason why communication and knowledge sharing is seen more often as a key benefit after deployment is that companies often lack agreement on big data analytics terminology. Participants from fewer than half (44%) of organizations said that the people making business technology decisions mostly agree or completely agree on the meaning of big data analytics; the same number said there are many different opinions about its meaning.

A variety of issues hold back organizations in big data analytics.

Most organizations that do not plan to implement big data analytics or don't know when they will do so said that is because it poses security risks (39%), promotes saving data that will not be used (32%) and will have no positive impact on the business (29%). Organizations that have deployed big data analytics gave other reasons for being held back from making improvements in it: difficulty in identifying return on investment and defining the business case (50%), lack of skills (39%) and data quality and information management issues (39%). We conclude that organizations that do not plan on rolling out big data analytics simply don't see the value of doing so, while those that have deployed the technology focus on realizing return on investment and using the tools fully.

Currently, only one in five organizations (21%) in this research are very confident in the information being generated by big data analytics.

Integration and data quality are also problematic areas. While three in five organizations are satisfied overall with the process of creating big data analytics, the percentage drops to 40 percent for satisfaction with the integration of information as a step in the process. The quality and consistency of data (for 56%) and validating data from a business perspective (39%) are also issues. Quality issues can undermine an organization's confidence in the data it uses to make decisions. Currently, only one in five organizations (21%) in this research are very confident in the information being generated by big data analytics. Most will need to address their big data analytics processes,

particularly those having to do with data quality, to realize full value from their initiatives.

Many organizations lack skills for big data analytics.

As technology advances, custom-developed approaches are giving way to packaged software, which can be easier to deploy and require less sophisticated skills. Most organizations in the research that have already implemented systems did so with custom builds using big data-specific languages and interfaces (54%), in-database analytics provided by a software vendor (34%) and open source data mining libraries (32%). For those still intending to deploy these capabilities, the largest percentage plan to purchase a dedicated or packaged application (44%), use a custom build (36%) or acquire them through a service provider (34%). The trend toward standardized commercial products has important implications for the skills needed for a big data analytics initiative; in addition to reducing the need for expertise in-house, in many cases it can also change the nature of the needed skills. For instance, using SQL skills for exploratory analytics or R language skills for advanced analytics can be faster and more direct than trying to implement those same analytic processes by writing custom code.

> The most important training needs are for applying analytics to business problems, the use of big data analytics tools, analytics concepts and techniques, and visualizing big data.

The new skills that the research suggests are needed to take advantage of the technologies are in the areas of statistics, mathematics and visualization. These findings are consistent with our examination of training needs, which identified as most important training for applying analytics to business problems (54%), the use of big data analytics tools (53%), analytics concepts and techniques (46%) and visualizing big data (41%).

Given the complexity of building and deploying big data analytic models, having dedicated skilled people to build and deploy the models is critical for success. Organizations that use specialized roles such as data scientist or data miner report improvement in organizational activities from using big data analytics more often (88%) than do those that are led by IT (73%) or lines of business (78%).

A variety of data sources are used to feed big data analytics.

In our view, a primary purpose of big data is to make as much data as possible available for analysis. Research participants agree; the most common definition of big data analytics (by 76% of organizations) is analyzing data from all sources of information. More than one-third of research participants named nine different types of big data they analyze. The most common types are transactional data from applications (60%), external sources (50%), content such as documents and Web pages (49%) and event-centric data (48%).

> The most common definition of big data analytics is analyzing data from all sources of information. More than one-third of research participants analyze nine different types of big data.

Accessing and integrating data sources from beyond the company firewall is a key enabler of big data analytics. The most widely used external data sources are cloud computing applications (by 54%) and social media data (46%); five additional data sources, including Internet, consumer, market and government sources, are virtually tied in third position (with 39% to 42% each). The research shows that the importance of external data sources varied with size of organization. Thus, for example, the inclusion of cloud computing data grew with the organization, from less than half (46%) of small companies to more than two-thirds (69%) of very large ones.

The most common uses of big data analytics involve customers and sales.

We asked participants to identify from among 25 uses for big data analytics those that they are personally involved with, and three of the four most often mentioned involve customers and sales: enabling cross-selling and up-selling (36%), understanding the customer better (32%) and optimizing pricing (28%). Optimizing IT operations ranked fifth (24%) and not surprisingly was the one most often chosen by those in IT roles (76%).

Industries differ significantly in their focuses for big data analytics. For example, identifying cross-sell and up-sell opportunities tops the list in Services (43%) and Finance, Insurance and Real Estate (FIRE, 39%), while enabling preventive maintenance is most important for Government (50%) and Manufacturing (23%). Analysis by size of business reveals subtle differences in priority, although supporting cross-selling and up-selling is cited most often across the board. For very large companies, the second-highest priority is improvement of forecast models (36%); for large and midsize companies, second is price optimization (38% and 31%, respectively); and for small companies, second is gaining a better understanding of the customer (37%). We conclude that the best place to begin to develop a business case for big data analytics is with top-line-driven project objectives, but big data analytics represents a breadth of applications that are specific to an organization's function, industry and size.

The best place to begin to develop a business case for big data analytics is with top-line-driven project objectives, but big data analytics represents a breadth of applications that are specific to an organization's function, industry and size.

Users of in-memory systems and Hadoop reported the most significant improvement.

Among current users of big data analytics, the most significant improvement was reported by those using in-memory systems (50%), Hadoop (42%) and real-time dashboards (39%).

In-memory technology can enhance versatility in multiple areas of big data analytics, including data preparation, exploratory analytics, predictive analytics and data visualization. It also significantly speeds up the process of rendering dashboards. Users of this technology said that real-time analytics (65%) is an important benefit of in-memory technology, and they also named as benefits being able to do analysis without having to define a data model first (44%), to mash up multiple data sets (39%) and to marry operational and analytic systems (38%). As in-memory systems are deployed more widely and the price of memory continues to fall, we anticipate an increase in the use of in-memory technology.

Hadoop also offers advantages in big data analytics. The longest-established Hadoop implementations, Apache (48%) and Cloudera (42%), remain the most widely used; four other providers occupy a second tier, each in about one-quarter of organizations. To implement analysis in Hadoop, the most organizations favor standard SQL approaches (52%), custom access via the HIVE ODBC connector (42%) and utilizing open source analytic libraries (39%).The predominance of standard SQL approaches reflects the skills that are currently available in organizations as well as the need to access Hadoop in a more timely manner than ODBC connectors can do. As standardized SQL approaches mature, we anticipate that this method of access will take market share from both native approaches and customized HIVE access. Furthermore, Hadoop 2.0, which is still new but can run multiple workloads simultaneously, should promote more advanced analytic software and streaming analytics on this platform.

Timely model updating based on business need produces better outcomes.

Most organizations regularly update the analytic models they use for big data analysis. Seven in 10 do it weekly or more often, and half (54%) do it daily or more frequently. Further analysis shows those that update their models based on business need have more significant improvement in outcomes than those that do not (73% vs. 50%). Furthermore, those that implement a champion or challenger approach to model updates, in which a better performing model automatically replaces a less well performing model, are more satisfied than others with their big data analytics efforts and see more significant impacts from big data analytics. However, currently only 19 percent take this approach. We conclude that best practices in big data analytic modeling include frequent updates done by some sort of automated process.

> Organizations that update their models based on business need have more significant improvement in outcomes than those that do not (73% vs. 50%).

Model scoring is a critical but not well understood aspect of analytics.

The scoring of databases is an established process, but nonetheless still a complex and somewhat arcane area, as reflected by the large number (30%) of participants who said they do not know how their organization performs scoring of its analytic models. Setting those responses aside, we find that 44 percent do some sort of SQL translation by hand coding (16%) or PMML model passing (28%). The majority (62%) currently run the process directly in the database by embedding the analytic processes (34%) or implementing a user-defined function natively in the database (21%). These findings reflect advanced environments among organizations that have implemented big data analytics.

Timely scoring is one benefit of in-database scoring and correlates with tight architectural integration for big data analytics. While only 23 percent of all organizations said big data analytics is well integrated into their architecture, 41 percent of those that score models more often than weekly said this, as did 71 percent of those that score in real time. We conclude that organizations that are looking for a tighter architectural fit for big data analytics should focus on timely model scoring and in-database approaches.

Many departments and functional roles have a say in selecting big data analytics.

Given the complexity of big data analytics, it is wise to involve those who design and deploy the analytics in the purchase decisions for software. The roles most often involved in designing and deploying big data analytics systems are IT organizations (37%), cross-functional teams (22%), data scientists or data miners (20%) and line-of-business analysts (15%). In our view all these roles should take part in selecting software, but in particular data scientists and data miners should be involved. Organizations that use specialized roles such as these show greater improvement in organizational activities from using big data analytics than others do. Departments that are planning most

> Given the complexity of big data analytics, it is wise to involve those who design and deploy the analytics in the purchase decisions for software .

often to deploy big data analytics are accounting and finance (39%), sales (34%), IT (34%), marketing (31%) and business development (30%). They, too, should have representatives if their group will use the product. A majority (52%) of participants will look to their business intelligence systems providers to supply big data analytics, though significant percentages will turn to a specialized analytics provider (35%) or their database provider (34%). Each of these providers has particular strengths, and it is incumbent on the organization to understand their capabilities. Big data analytics encompasses databases, analytics and applications, and it's important to have a provider that understands all these dimensions.

When evaluating big data analytics products and vendors, nine out of 10 organizations said that usability, functionality and reliability are important or very important criteria. More than three-fourths (78%) said that collaboration is important to big data analytics, whereas mobile access is important to fewer than half (46%). Coupled with the finding that communication and knowledge sharing is the number-one benefit of big data analytics already implemented, it is clear that buyers are aware of the collaborative imperative when choosing a big data analytics product and even more so once the system is implemented. For this reason, we expect embedding of collaborative capabilities into big data analytics products to become more common.

10 Best Practice Recommendations

This benchmark research reveals significant new insights into the evolving nature and use of big data analytics processes and systems. For organizations considering how to optimize their analytics and insights derived from them, we offer the following recommendations.

1. Examine the experiences of early adopters of big data analytics.

The research shows that big data analytics is important or very important for four out of five organizations, but only about half (56%) are satisfied with their current efforts. We find satisfaction 50 percent more often among those that have had implementations for more than a year (66%) than those that began in the last 12 months (44%). Similarly, those using advanced tools are satisfied more often than those using legacy systems such as relational databases. Therefore, prepare to invest time in your own big data analytics deployment, and consider using tools that are designed to smooth the process and deliver specific results. Insist that vendors provide customer references that include best practices, ideally from companies in your type of industry or of your size and complexity, and ask them how they can accelerate and ease deployment. Learning from them can help your organization avoid pitfalls and shorten the time to value for its deployment.

2. Define big data analytics clearly and establish communications among all stakeholders.

In nearly half of organizations participating in the research there are many different opinions about the meaning of big data analytics. Two of the three definitions most mentioned are that big data analytics involves analyzing data from all sources rather than just one (76%) and analyzing all of the data rather than just a sample of it (55%). We advise focusing your efforts, at least in beginning, on analyzing large amounts of data from many data sources. The research also shows that communication and knowledge sharing (ranked first by 24% and mentioned by 51%) is the most-often realized benefit of big data analytics, so from the beginning work to ensure that communication channels are in place, there is consensus on what big data analytics is and will do, and outcome expectations

are clearly defined. Doing so can help your organization achieve other benefits cited as important by research participants; topping that list are gaining competitive advantage (mentioned by 51%), making business processes more efficient (49%) and improving the customer experience and satisfaction (46%).

3. Determine where you can profit most from big data analytics.

Identify those business activities where application of big data analytics can produce important benefits. Three of the ones most often cited in the research are related to customers and revenue: enabling cross-selling and up-selling (36%), understanding the customer better (32%) and optimizing pricing (28%). Such a results-based focus may appeal to decision-makers considering an initiative. For example, recommendation engines have multiple uses including cross-selling and up-selling, which should be a priority for marketing. Improving forecasting (30%) and pricing should appeal to those in operations and finance. And optimizing operations through analysis of data such as log files is important for IT; one in four said they use big data analytics to optimize IT operations. Our analysis shows differing priorities based on organizations' sizes and industries. Carefully consider your competitive environment and organizational needs and build a business case based on them.

4. Address concerns that can hold back improvement and effectiveness.

The research identifies a variety of organizational barriers to deploying big data analytics. The top reason most organizations do not plan to implement big data analytics or don't know when they will do so is concern about security risks (39%); this need not be an issue, but it is wise to make it part of your evaluation to ensure you learn how to avoid it. Half of those that have already deployed big data analytics said difficulty in identifying return on investment (ROI) and defining the business case holds back making improvements in it; undertake advance research on both of these issues. The other most troublesome barriers to improvement, for two in five organizations, are lack of skills and data quality and information management. Thus you also should address people and information issues prior to implementing big data analytics. Determining the expected ROI and

defining a strong business case up front can prevent a loss of confidence once the initiative is under way.

5. Exploit skills available now in the organization and develop others needed for big data analytics.

A significant percentage (39%) said lack of skills impedes improvement in big data analytics, but the research also shows more satisfaction when advanced tools are used. Deploying standardized commercial software can reduce the need for in-house expertise in analytics. We recommend a two-pronged approach: Exploit the skills you already have (as in using SQL and spreadsheets) while adding advanced skills designed for big data and analytics. Organizations that use specialized roles such as data scientist or data miner reported improvement in activities from using big data analytics more often (88%) than those that are led by IT (73%) or lines of business (78%). On the other hand new tools can help SQL and spreadsheet users apply their skills to analytics, and for advanced analytics other tools remove the complexity of custom coding. In any case, be sure to include training in your implementation plans. In the research the most important training needs are for applying analytics to business problems (in 54% of organizations), the use of big data analytics tools (53%), analytics concepts and techniques (46%) and visualizing big data (41%). Therefore concentrate on developing skills in visualization, mathematics and statistics.

6. Realize that big data analytics requires a variety of information sources.

Half or more of organizations use an array of different types of data for their analytics: transactional data from applications, external data sources, content such as documents and Web pages, and event-centric data. Consider which mix of these each of your different user groups need to access and how to integrate them for analysis. Significantly fewer (40%) organizations are satisfied with the integration of information than are satisfied with the process overall (56%). Keep in mind that external data sources coming from cloud-based applications, social media data and third-party sources are increasingly used for big data analytics and often require different integration approaches than internal data types. Having high-quality data in compatible formats is a necessary precursor to analysis. Don't

underestimate the importance of it, and find tools that can assist in the process.

7. Evaluate advanced tools such as Hadoop and in-memory systems for big data analytics.

Our analysis finds more significant improvement in organizational activities for those using in-memory systems (50%), Hadoop (42%) and real-time dashboards (39%) for big data analytics than other tools. Yet the tools most commonly used are not these but rather relational database management systems (RDBMSs) and flat files, which cannot exploit the power of big data. We strongly recommend evaluating tools designed for big data analytics. In-memory systems provide speed and flexibility and should be considered to enable functions such as real-time dashboards, visual analytics and data discovery, and analytic sandbox environments. Two-thirds of users of this technology said that real-time analytics is an important benefit of it. These functions can shorten analytic time to value, which many organizations consider an important metric for big data analytics. While in-memory technology is limited in the size of data it can handle, Hadoop provides a parallelized scalable architecture for big data. In addition, users can apply standard SQL approaches to implement analysis in Hadoop, which more than half do. Consider Hadoop for tasks such as ETL offloading, recommendation engines and fraud analytics, in which large-scale models must be frequently updated and scored.

8. Deploy analytic models that meet specific needs and update them regularly.

The research reveals that effective organizations manage their modeling processes by building and deploying them in a timely manner to meet specific needs. Most regularly update the analytic models: Seven in 10 do it weekly or more often, and half (54%) do it daily or more frequently. Those that update their models based on business need have more significant improvement in outcomes than those that do not (73% vs. 50%). Consider modern approaches to deploying big data analytics, including in-database approaches, that provide tight integration with the underlying architecture and enable timely deployments. The timing for model updates depends on changes in your business environment and the goals of the model, but we conclude from the research that organizations should update their

models at least once a month. Similarly, we recommend scoring models at least once a week (four out of five score more often than that) and considering modern methods for scoring such as PMML (which 28% currently use) and in-database approaches (62%). Also consider the advantages in time saved and accuracy of machine learning techniques such as closed-loop validation processes and continuous model updates over more manual modeling approaches.

9. Utilize cross-functional teams that include a data science resource.

The research shows that IT organizations (37%) are most often involved in designing and deploying big data analytics systems, followed by cross-functional teams (22%), data scientists or data miners (20%) and line-of-business analysts (15%). We believe that all these roles should participate in buying decisions for big data analytics, as should representatives from the departments that will deploy the analytics as well. More than three-fourths (78%) of participants said that collaboration is important to big data analytics, so it makes sense to institute it from the start. Remember also that communication and knowledge sharing is a highly ranked benefit of these initiatives. The research shows that when IT takes the lead, satisfaction is much less than when a specialized resource is involved. We suggest creating cross-functional teams, ideally including a data scientist, to manage the process; this is likely to produce the most significant impact on organizational activities. Furthermore, consider vendors that employ such specialized talent and can help lead the project from an analytical perspective rather than a technological one.

10. Take a broad view of big data analytics deployment.

We recommend considering multiple perspectives before deploying big data analytics. Include in your assessment the type of analytics to be deployed, the specific technologies to be included, the processes involved in customizing models for different purposes and the people and skills that must be involved to enable success. Consider a variety of technological approaches such as in-database analytics, in-memory, Hadoop, open source data-mining libraries and third-party analytics suppliers. These options can improve outcomes associated with big data analytics initiatives as well as supplement the expertise in your organization. An increasing number

(44%) of organizations plan to purchase a dedicated or packaged application when they deploy big data analytics. Fewer of them (36%) will choose custom builds than did those that have already implemented systems (54%). Consider evaluating packages that fit the deployment rather than creating a completely custom approach. Also understand how you want to deploy analytic models into the relevant lines of business. This may require integration with a variety of tools, including visualization, business intelligence, and business process and decision management.

About Ventana Research

Ventana Research is the most authoritative and respected benchmark business technology research and advisory services firm. We provide insight and expert guidance on mainstream and disruptive technologies through a unique set of research-based offerings including benchmark research and technology evaluation assessments, education workshops and our research and advisory services, Ventana On-Demand. Our unparalleled understanding of the role of technology in optimizing business processes and performance and our best practices guidance are rooted in our rigorous research-based benchmarking of people, processes, information and technology across business and IT functions in every industry. This benchmark research plus our market coverage and in-depth knowledge of hundreds of technology providers means we can deliver education and expertise to our clients to increase the value they derive from technology investments while reducing time, cost and risk.

Ventana Research provides the most comprehensive analyst and research coverage in the industry; business and IT professionals worldwide are members of our community and benefit from Ventana Research's insights, as do highly regarded media and association partners around the globe. Our views and analyses are distributed daily through blogs and social media channels including Twitter, Facebook, LinkedIn and Google+.

To learn how Ventana Research advances the maturity of organizations' use of information and technology through benchmark research, education and advisory services, visit www.ventanaresearch.com.

Appendix: About This Benchmark Research

Methodology

Ventana Research conducted this benchmark research on the Web from July through October 2013. We solicited survey participation via email, our website and social media invitations. Email invitations were also sent by our media partners and by vendor sponsors.

We presented this explanation of the topic to participants prior to their entry into the survey:

Applying analytics to big data can improve organizations' ability to address many information-dependent operations and management needs. But many in business and IT still aren't clear what big data analytics is or how to use it effectively. This benchmark research will for the first time assess the lessons learned by early adopters and identify, explore and quantify the ways in which organizations use big data analytics to gain business value.

The following promotion incented participants to complete the survey:

What's In It For You? Upon completion of the research, all qualified participants will receive a report on the findings of this benchmark research to support their organization's efforts, along with a $5 Amazon.com gift certificate. In addition, all qualified participants will be entered into a drawing to win one of 25 benchmark research reports and a 30-minute consultation, a package valued at US$1,495 or €1,232. Thank you for your participation!

Qualification

We designed the research to assess the use of and plans for spreadsheets across organizations and industries. Qualification to participate was presented to participants as follows:

The survey for this benchmark research is designed for business and IT managers connected with managing or using big data systems and business analytics. Solution providers, software vendors, consultants, media and systems integrators may participate in the survey, but they

are not eligible for incentives and their input will be used only if they meet the qualifications. Incentives are provided to qualified participants in the research and also are conditional on provision of accurate contact information including company name and company email address that can be used for fulfillment of incentives.

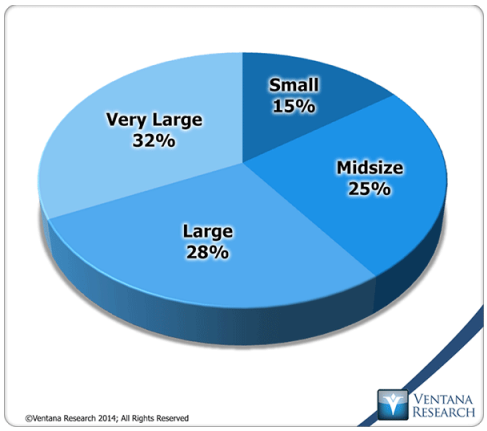
Further qualification evaluation of respondents was conducted as part of the research methodology and quality assurance processes. It entailed screening out responses from companies that are too small, questionnaires that were not materially complete, or those where the submission is from an inappropriate submitter or appears to be spurious.

Demographics

We designed the survey used for this research to be answered by executives and managers across a broad range of roles and titles working in organizations. We deemed 240 of those who clicked through to this survey to be qualified to have their answers analyzed in this research. In this report, the term “participants” refers to that group, and the charts in this section characterize various aspects of their demographics and qualifications.

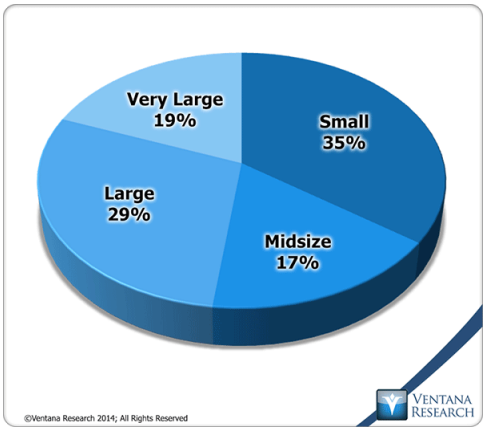
Company Size by Workforce

We require participants to indicate the size of their entire company. Our research repeatedly shows that size of organization, measured in this instance by employees, is a useful means of segmenting companies because it correlates with the complexity of processes, communications and organizational structure as well as the complexity of the IT infrastructure. In this research, participants represented a range of organization sizes, with three-fifths from the two largest size categories: 32 percent work in very large companies (having 10,000 or more employees), 28 percent work in large companies (with 1,000 to 9,999 employees), 25 percent work in midsize companies (with 100 to 999 employees), and 15 percent work in small companies (with fewer than 100 employees). This distribution is consistent with prior benchmark research and our research objectives and provides a suitably large sample from each size category.

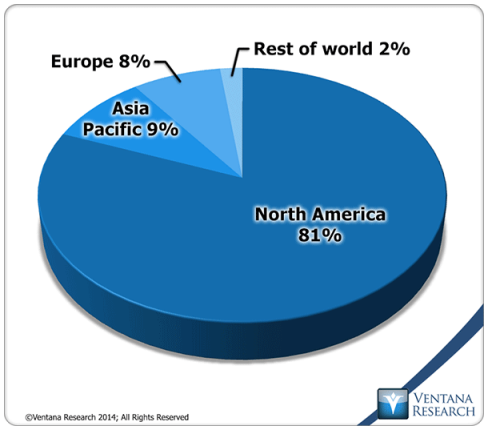


Company Size by Annual Revenue

When we measured size by annual revenue, the distribution of categories shifted downward; fewer companies fell into the largest category and more than twice as many are small. By this measure, 13 percent fewer are very large companies (having revenue of more than US$10 billion), 1 percent more are large companies (having revenue from US$500 million to US$10 billion), 8 percent fewer are midsize companies (having revenue from US$100 to US$500 million), and 20 percent more are small companies (with revenue of less than US$100 million). This sort of redistribution is typical, though usually less dramatic, in our research projects when we measure by revenue instead of headcount.

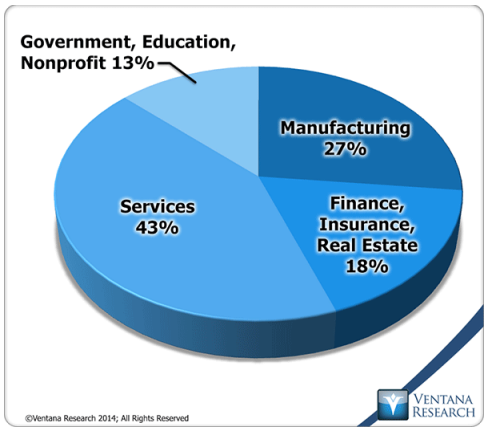


Geographic Distribution



A large majority (81%) of the participants were from companies located or headquartered in North America. Those based in Asia Pacific accounted for 9 percent, in Europe for another 8 percent and in the Middle East, Africa and Central and South America combined for 2 percent. This result was in keeping with our expectations at the start of this investigation, since organizations participating in our research most often are headquartered in North America. However, many of these are global organizations operating worldwide.

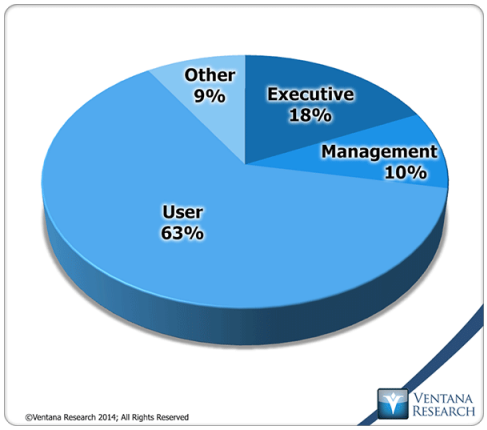
Industry



The companies of the participants in this benchmark research represented a broad range of industries, which we have categorized into four general categories as shown below. Companies that provide services accounted for 43 percent, those in manufacturing for 27 percent and those in finance, insurance and real estate for 18 percent. Government, education and nonprofits accounted for the balance.

Job Title

We asked participants to choose from among 15 titles the one that best describes theirs. We sorted these responses into four categories: executives, management, users and others. Slightly less than two-thirds identified themselves as having titles that we categorize as users, a grouping that includes director (18%), senior manager or manager (29%), analyst (14%) and staff (3%). Almost one in five are executives; most of them are CIOs (8%) or CEOs or presidents (6%). Another 10 percent are management, by which we mean vice presidents. Others, in this case consultants, teachers or students, accounted for the balance. We concluded after analysis that this response set provided a meaningfully broad distribution of job titles.



Role by Functional Area

We asked participants to identify their functional area of responsibility as well. This enabled us to identify differences between participants who have differing roles in the organization. In this rather technical area nearly two-fifths of the participants identified themselves as being in the IT function. The only line of business with more than 6 percent was finance or accounting, at 8 percent.