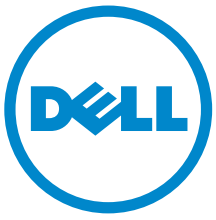
Insights Across the Hybrid Enterprise: Big Data 2015

An ENTERPRISE MANAGEMENT ASSOCIATES® (EMA™) and 9sight Consulting Research Report

March 2015

Premium Sponsors:



Gold Sponsor:

IT & DATA MANAGEMENT RESEARCH, INDUSTRY ANALYSIS & CONSULTING

Table of Contents

1. Executive Summary 1

1.1. Key Findings 2

2. Big Data—More Than Size Matters 3

2.1. A Maturing Market 3

2.2. On the Road to a Data-Driven Culture 5

2.3. The Internet of Things Has Already Arrived 7

2.4. Filling Up the Data Lake 11

2.5. Where Does Big Data Stand Today? 13

2.6. Future of Ethics in Big Data – Security and Privacy 14

3. Hybrid Data Ecosystem: Big Data Equals a Big Tent 15

3.1. Hybrid Means More than Multi-Structured 16

3.2. Expanding to the Cloud 17

3.3. Data Consumers and the Cloud 19

3.4. Updates to the Hybrid Data Ecosystem 19

4. 2014 EMA/9sight Survey Demographics 21

4.1. Company Size 21

4.2. Segmenting Industries 22

4.3. Geographic Distribution 23

5. EMA/9sight Survey Panelist Horizontal Strategies 25

5.1. Big Data 25

5.2. Cultural Change: Data-Driven Organizations 25

5.3. Connecting Everything: Internet of Things 27

5.4. Data Lakes 28

5.5. Use Cases 29

5.6. Managing Hurdles 31

6. Big Data Projects from EMA/9sight Survey Panelists 34

6.1. Project Maturity 35

Table of Contents (continued)

6.2. Build vs Buy 38
6.3. Project Business Challenge 39
6.4. Data Consumers 41
6.5. Building Big Data User Skills 44
6.6. Project Sponsors 46
6.7. Impact of Cloud on Implementation Avenues 48
6.8. Workload 49
6.9. Latency 51
7. What is Driving Big Data? Hybrid Data Ecosystem? 52
7.1. The Need for Speed 52
7.2. Economics of Big Data 55
7.3. Complex Workloads Go Real Time 58
7.4. Data Loads 61
7.5. Multi-Structured Data: Different Tune, Same Song 65
8. Methodology and Demographics 70
8.1. Research Methodology 70
8.2. 2014 Respondents 70
8.3. 2013 Respondents 70
8.4. 2012 Respondents 70
9. Authors 71
9.1. About Enterprise Management Associates 71
9.2. About 9sight 71

1. Executive Summary

Since its inception, the concept of "big data" has been widely associated with a single data management platform—Hadoop. This connection may be due to the popularity of the use of the Hadoop platform to store and process large amounts of multi-structured data. However, Hadoop is not the only answer to the question of "what is big data?"

As was explored in the inaugural EMA/9sight survey in 2012, big data is both a way to look at new sources of data and how organizations place that information under management. Big data has attracted a wide range of application innovators as well as many protesters against the dominance of relational databases and data warehouses. The EMA/9sight surveys use a deliberately broad definition of big data to inspire end users to think beyond the box of Hadoop. The survey explores the wide range of ways in which non-traditional data, often in combination with more traditional types, has enabled new or improved business processes. As was established in previous studies in 2012 and 2013 and again in 2014, big data offers a wide range of possibilities, but the name "big data" itself keeps media and industry eyes focused on size as the defining feature.

With the 2015 report, these previous observations continue to hold true. However, there is an evolving market in which size is not everything, and speed, in all its aspects, has grown in importance for respondents. Furthermore, respondents continue to include a wide range of data structures, from highly irregular to strongly modeled, within the scope of their projects. This refocusing of implementers' attention on speed and structure reduces overall growth in big data by some measures and requires EMA/9sight to explore how consideration of speed and structure are changing market dynamics. The 2014 survey also included investigations into the highly visible topics of data-driven cultures, the Internet of Things, and data lake architectures.

Relating to data-driven cultures, respondents with larger percentages of big data in their organizations were also more likely to have adopted data-driven strategies and vice versa. In addition, there will be significant growth in the area of workforce access to data-driven initiatives over the next year.

If any term eclipsed big data in attention and hype in 2014, it must have been Internet of Things (IoT) or one of its synonyms, such as Internet of Everything or Industrial Internet of Things. Suddenly, it seemed that IoT was the acronym on everybody's lips, as predictions of enormous growth and high economic impact have become commonplace. However, the devices themselves are only one side of the equation. On the other side are the data delivered to the enterprise and the analytics required to understand its significance—topics of key importance to survey respondents.

The concept of data lake architectures is also gaining traction in big data initiatives. Externally-sourced big data, including data from the Internet of Things, clearly demands a data management store, as information arrives at high speed and in significant volumes. This store must be large, flexible, and cost-effective. However, expanding the scope of the data lake to include business-critical data and legally-binding core business information could pose issues for organizations unless they master their data management practices first.

In 2012, EMA defined a next-generation data management architecture—the Hybrid Data Ecosystem. The HDE has been refined over the past three years. Each of the platforms within the Hybrid Data Ecosystem supports a particular combination of business requirements along with operational or analytical processing challenges. The HDE represents a unique approach compared with traditional best practices. Rather than advocating a single data store that supports all business and technical requirements at the center of its architecture, the Hybrid Data Ecosystem seeks to determine the best platforms for supporting a particular set of requirements and links those platforms together. This year the HDE expands to include the influence and impact of the cloud on big data environments and data consumers.

Copyright 2015, EMA Inc. and 9sight Consulting. All Rights Reserved.

1.1. Key Findings

The 2014 EMA/9sight Big Data research surveyed 351 business and technology stakeholders around the world. The survey instrument was designed to identify key trends surrounding the adoption, expectations, and challenges associated with strategies, technologies, and implementations of big data initiatives. The research identified the following highlights in the 2014 Big Data research with comparisons to past EMA/9sight studies in 2012 and 2013.

- **Growing Number of Projects –** In 2014, almost two thirds of organizations had three or more big data projects, and more than 20% reported five or more projects in progress.
- **IoT Importance** – Nearly 50% of respondents indicated that the Internet of Things was currently adopted and an important or essential part of business.
- **Data Lakes: An Acquired Taste** – Over 22% of respondents said a data lake is currently adopted as part of the strategy to replace existing operational and informational platforms.
- **Big Data Is a Maturing Strategy** – Over 55% said that a big data strategy was Adopted and Essential or Adopted and Important.
- **Data-Driven Is Driving Organizations** – Almost 63% of respondents included data-driven strategies in their organization at a significant level.
- **Speed Is Driving Competition** – Speed of processing response was the most frequently indicated use case by respondents at nearly 20%.
- **What's Stopping Big Data?** – Poor data management practices, like lack of data governance and not supporting SLAs, are the top obstacles for organizations implementing a big data initiative.
- **Time to Value with Applications** – Over 20% of respondents implemented big data projects using customizable applications from an external provider as their big data implementation strategy.
- **Revenue-Generating Big Data Projects** – The highest percentage of projects tackle market basket analysis and cross-sell/up-sell activities, targeting improvements to top-line revenue.
- **Continued Importance of Analytics in Business Processes** – Operational analytics leads all other categories, 42% stating they were executing projects around this workload to minimize cost.
- **Business Stakeholders Lead Consumption Again** – Line of business executives is the largest groups of data consumers. Marketing and financial analysts is the second largest group of users.
- **Looking Outside the Skills Box** – Over 20% of respondents utilize technical consulting services from external organizations, providing a significant opportunity for consulting groups.
- **Partly Cloudy Implementations** – Nearly 60% of big data projects have a primary implementation element as part of a cloud: private, public, hybrid, or as a managed service.
- **Low Latency is High Profile** – Big data projects are overwhelmingly near-real time, with over 32% described as real time/near-real time processing of data.
- **Two-time Use Case Champion** – For the second year in a row, the top use case for big data initiatives is speed of processing response at nearly 20% of mentions.
- **Financial Driver: Managing Maintenance –** The primary financial driver was that organizations wanted to reduce maintenance costs associated with their existing data management implementations.
- **What's in Big Data? –** Machine-generated data types dominate the data types for big data initiatives, with over 40% of respondent mentions.
- **How Is Big Data Moved?** – The top three integration strategies were real-time streaming data integration, data replication (e.g., standardized duplication of enterprise data sources), and change data capture.

Copyright 2015, EMA Inc. and 9sight Consulting. All Rights Reserved.

2. Big Data—More Than Size Matters

As was explored in the inaugural EMA/9sight survey in 2012[1], big data is both an ill-defined reformation of the data management market and a broad church. It has attracted a wide range of application innovators as well as many protesters against the hegemony of relational databases and data warehouses. The EMA/9sight surveys use a deliberately extensive definition of big data. As a result, respondents thought way beyond the box of Hadoop. They explored a wide range of ways in which non-traditional data, often in combination with more traditional types, has enabled new or improved business processes. Thus, in 2012 and 2013, there was an inchoate environment where big data offered a wide range of possibilities, but the name big data itself kept media and industry eyes focused on size as the defining feature.

With this third iteration of the survey, these previous observations continue to hold true. However, there is an evolving market where size is not everything. In particular, speed, in all its aspects, has grown in importance for respondents. Furthermore, respondents continue to include a wide range of data structures, from highly irregular to strongly modeled, within the remit of their projects. This refocusing of implementers' attention on speed and structure reduces overall growth in big data by some measures and requires us to explore how consideration of speed and structure are changing the market dynamics. Speed of data acquisition and processing appears most obviously in the concept of the Internet of Things. Considerations of data structure lead us to examine data lakes.

2.1. A Maturing Market

Between 2012 and 2013, there was significant growth in the percentage of projects[2] **In Operation** with corresponding decreases in percentages for **Serious Planning** and **Investigating**. In 2014, **In Operation** dropped just over one percentage point to just over 52%. Such a plateau is normally indicative of a maturing market, but might be expected to occur at a somewhat higher percentage. However, this statistically small change should be taken in the context of the strong 6-8% shift from **Investigating** to **Serious Planning**. This combination of trends suggests a pause in putting more "traditional" big data projects into production, while new and different projects are being planned. This is another aspect of a maturing market—one that needs to pause in implementation to take account of lessons learned and external market changes. These changes will be discussed further later in this paper.

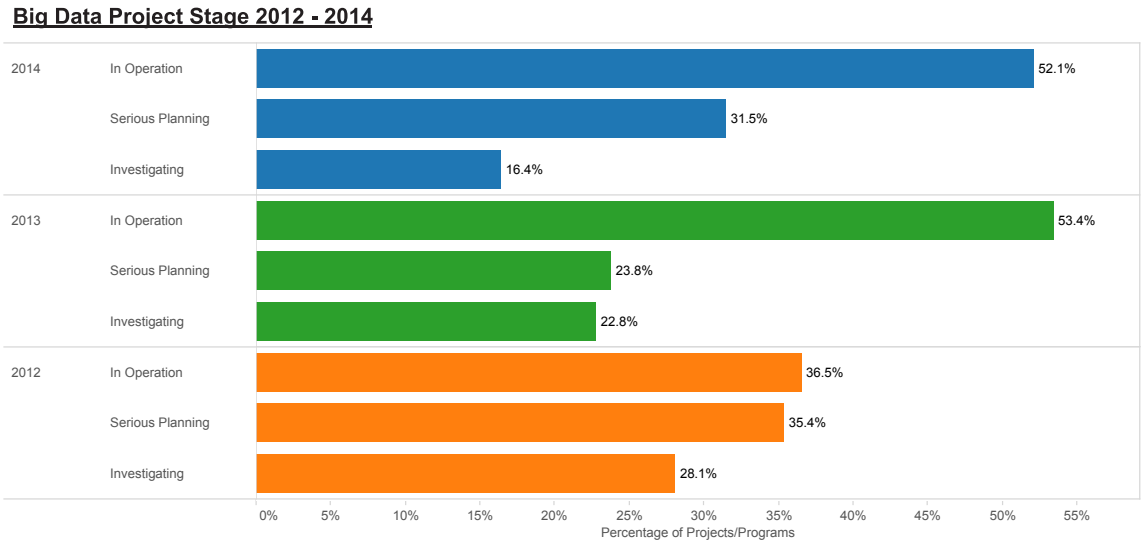


Figure 1

[1] http://www.enterprisemanagement.com/research/asset.php/2409/Big-Data-Comes-of-Age
[2] 2012 figures refer to the percentage of respondents rather than percentage of projects.

Copyright 2015, EMA Inc. and 9sight Consulting. All Rights Reserved.

Another sign of this growing maturity is the increasing number of big data projects found in each organization. Between 2013 and 2014, there was significant growth in the percentage of organizations with three or more big data projects, with a corresponding decrease in the percentage of organizations with one or two. Less than one third of respondents in 2013 had three or more projects. By 2014, that percentage had increased to almost two thirds of respondents, and more than 20% reported five or more projects in progress.

Big data is both an ill-defined reformation of the data management market and a broad church. It has attracted a wide range of application innovators as well as many protesters against the hegemony of relational databases and data warehouses.

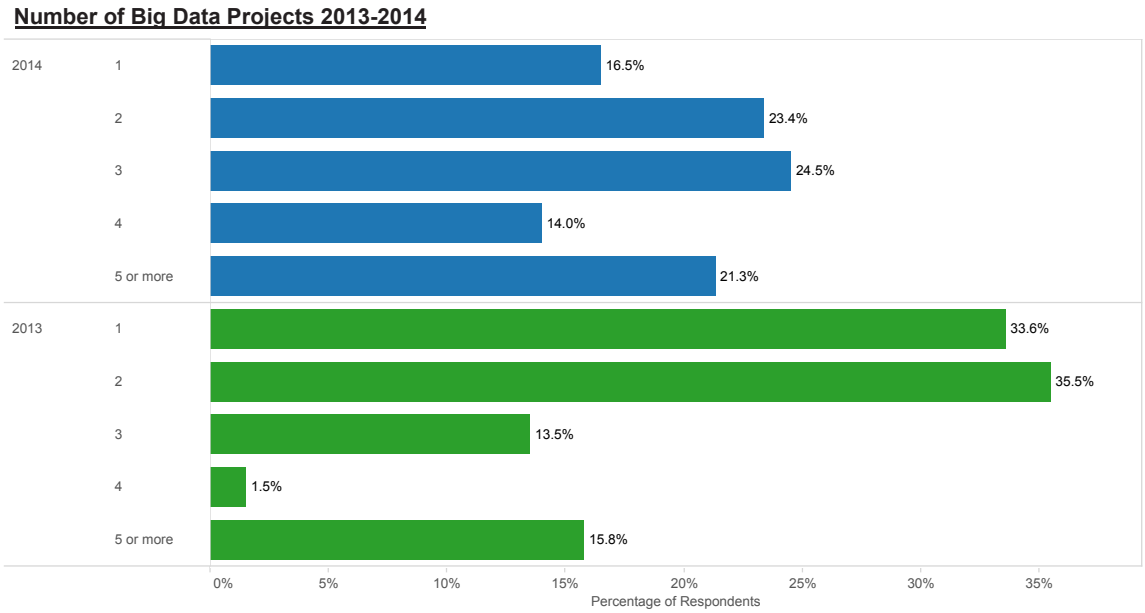


Figure 2

Another indication of this growing maturity of the big data market is a shift from **Stakeholder Issues** being the largest obstacle in 2013 to **Data Management** topping the list in 2014. With stakeholders more broadly on board, the focus has switched to more technical concerns. However, **Strategy Issues** still figure prominently on the obstacle list, indicating that more work is needed to align the value and use of big data with business goals.

Less than one third of respondents in 2013 had three or more projects. By 2014, that percentage had increased to almost two thirds of respondents, and more than 20% reported five or more projects in progress.

Copyright 2015, EMA Inc. and 9sight Consulting. All Rights Reserved.

2.2. On the Road to a Data-Driven Culture

The concept of being a data-driven organization or culture is closely associated with that of big data. And like big data, data-driven is a term that is widely used but loosely defined. The EMA/9sight survey defined data-driven organizations as:

> [c]ompanies that have either fully integrated data or started with data as the basis for their strategic and operational thinking. These organizations use data and information for finance, marketing, sales, and operations decisions and have built new business models in existing markets using data and information. Data-driven organizations have the following attributes: widespread access to and usage of data, integration of data analytics into strategic decision-making and pervasive usage of information in tactical decision-making.

With over 60% of respondents declaring that their organizations have already adopted a data-driven strategy, data-driven is clearly seen as important to future success. Indeed, nearly 20% of respondents described data-driven as already being a vital part of their business. The correlation between big data and being data-driven was also clearly seen.

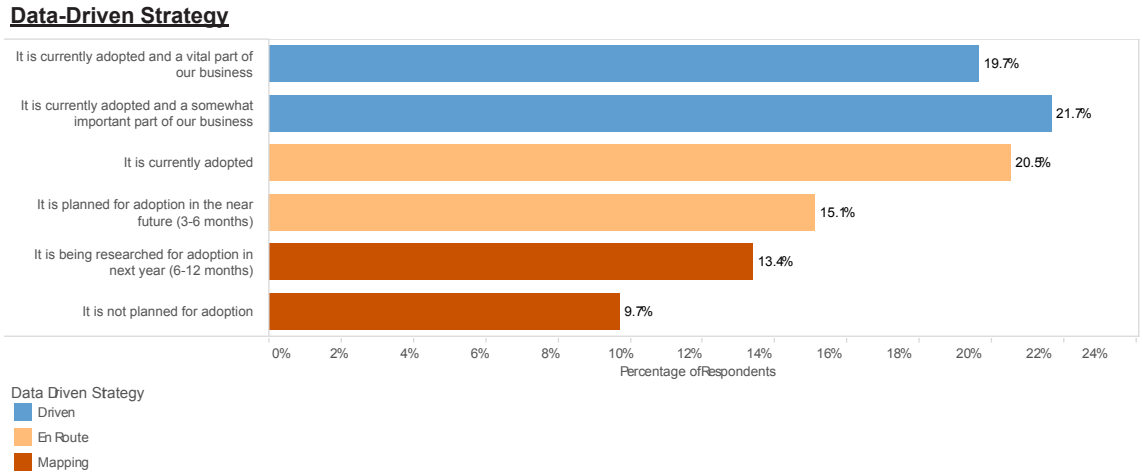


Figure 3

Respondents with larger percentages of big data in their organizations were also more likely to have adopted data-driven strategies and vice versa. In particular, 52% of respondents who were either not planning to adopt a data-driven strategy or are in the early stages of research recorded big data percentages of less than 10%, while 12% of respondents who declared data-driven a vital or important part of their business recorded big data percentages of more than 40%. When looking at current and expected workforce access to data-driven initiatives, there will be significant growth over the period 2014 to 2016. Currently, half the respondents show access rates of less than 15% of the workforce. By 2016, over 20% of respondents expect to have access rates to data-driven initiatives, reaching over 50% of the workforce.

Currently, half the respondents show access rates of less than 15% of the workforce. By 2016, over 20% of respondents expect to have access rates to data-driven initiatives, reaching over 50% of the workforce.

Copyright 2015, EMA Inc. and 9sight Consulting. All Rights Reserved.

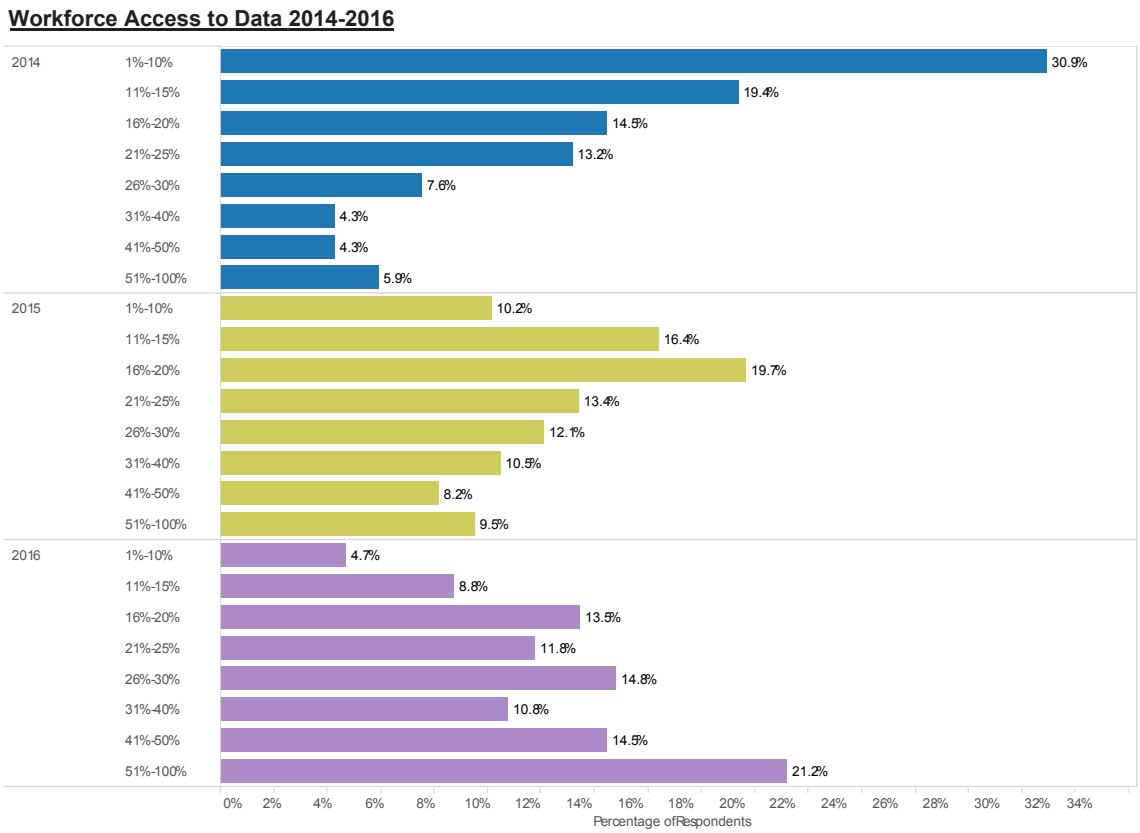


Figure 4

However, according to respondents, there is significant variation in uptake across industry segments. Comparing the **Retail** and **Industrial** industry segments with **Finance**, for example, data-driven strategies are strongly adopted by both industries. **Retail** and **Industrial** both have nearly a 68% self-description as **Driven** cultures. **Finance** is well over 50%.

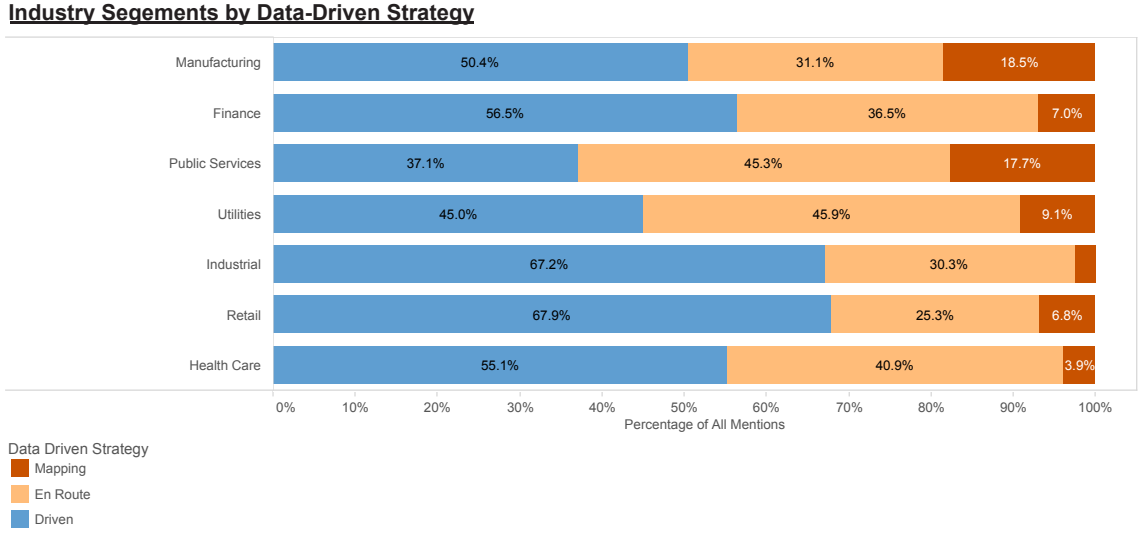


Figure 5

Copyright 2015, EMA Inc. and 9sight Consulting. All Rights Reserved.

However, according to respondents, there is significant variation in uptake across industry segments when you consider the general concept of adoption. EMA/9sight defines adoption as an organization indicating that a data-driven culture is **Currently Adopted**, **Adopted and Somewhat Important**, or **Adopted and Vital**. In this analysis, the finance industry segment has well over 90% adoption of data-driven concepts.

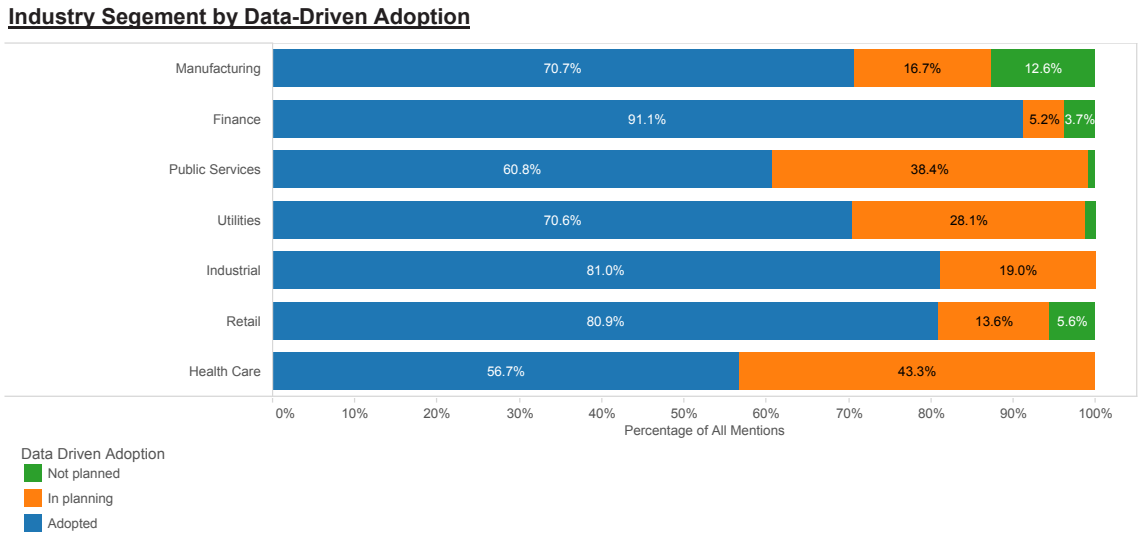


Figure 6

However, **Retail** and **Industrial** show a considerably lower overall adoption percentage than their peers in the finance industry segment. It can be speculated that this somewhat counterintuitive result arises more from individual industry segment perceptions of the meaning of data-driven than a real difference in the actual vital importance of the data between the industries.

2.3. The Internet of Things Has Already Arrived

If any term eclipsed big data in attention and hype in 2014, it must have been Internet of Things (IoT), or one of its synonyms, such as Internet of Everything or Industrial Internet of Things. Since 2014, the Consumer Electronics Show in Las Vegas was overrun with a plague of wearables of every conceivable shape and size. Also in January, Google acquired the home automation company, Nest, preceded and followed by a string of other automation and robotics acquisitions. Suddenly, it seemed IoT was the acronym on everybody's lips as predictions of enormous growth and high economic impact have become commonplace.

The devices themselves are only one side of the equation. On the other side reside the data delivered to the enterprise and the analytics required to understand its significance—topics of key importance to survey respondents. In between is the network where significant architectural and development work is required to carry the data safely, consistently, and securely to its intended destinations. The establishment of this infrastructure will occupy hardware and software vendors and communications providers for some years to come. Nonetheless, it is clear that enterprises who participated in the EMA/9sight survey are already considering data and analytics needs in this space.

The 2013 big data survey had already noted almost 15% growth in a class of data designated as "**Machine-generated**." The name is self-explanatory: the category covers data generated by sensors and similar devices, as well as that produced by computers as logs of significant events. It is thus a

Copyright 2015, EMA Inc. and 9sight Consulting. All Rights Reserved.

categorical superset of the data produced by the Internet of Things. **Machine-generated** data includes both internally- and externally-sourced data. 2014 saw small, approximately 3%, further growth in the **Machine-generated** category to just over 41%.

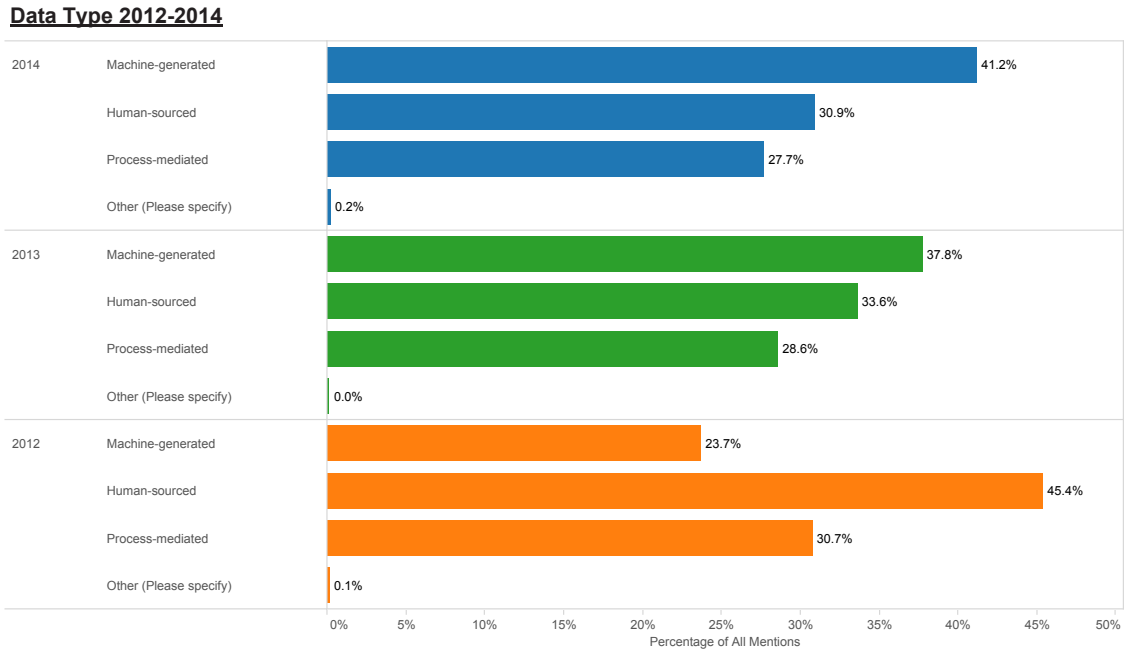


Figure 7

In the 2014 survey, respondents were asked how they were implementing the Internet of Things. The somewhat surprising response was that the Internet of Things was **currently adopted and an important or essential part of business** for nearly half (46%) of the respondents, while just over 10% were not even planning to research it.

The Internet of Things was currently adopted and an important or essential part of business for nearly half (46%) of the respondents.

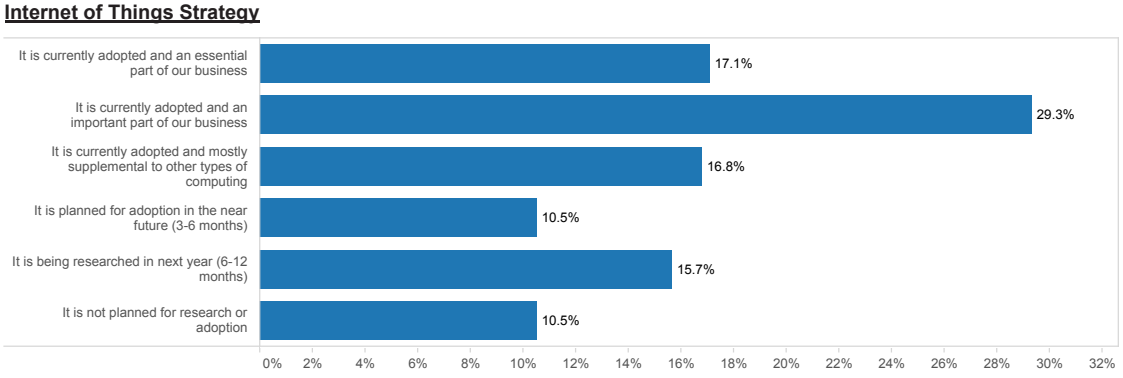


Figure 8

Copyright 2015, EMA Inc. and 9sight Consulting. All Rights Reserved.

The detailed responses to the Project Stage question (discussed in more detail in Section 6.1) provide additional insight into the IoT numbers.

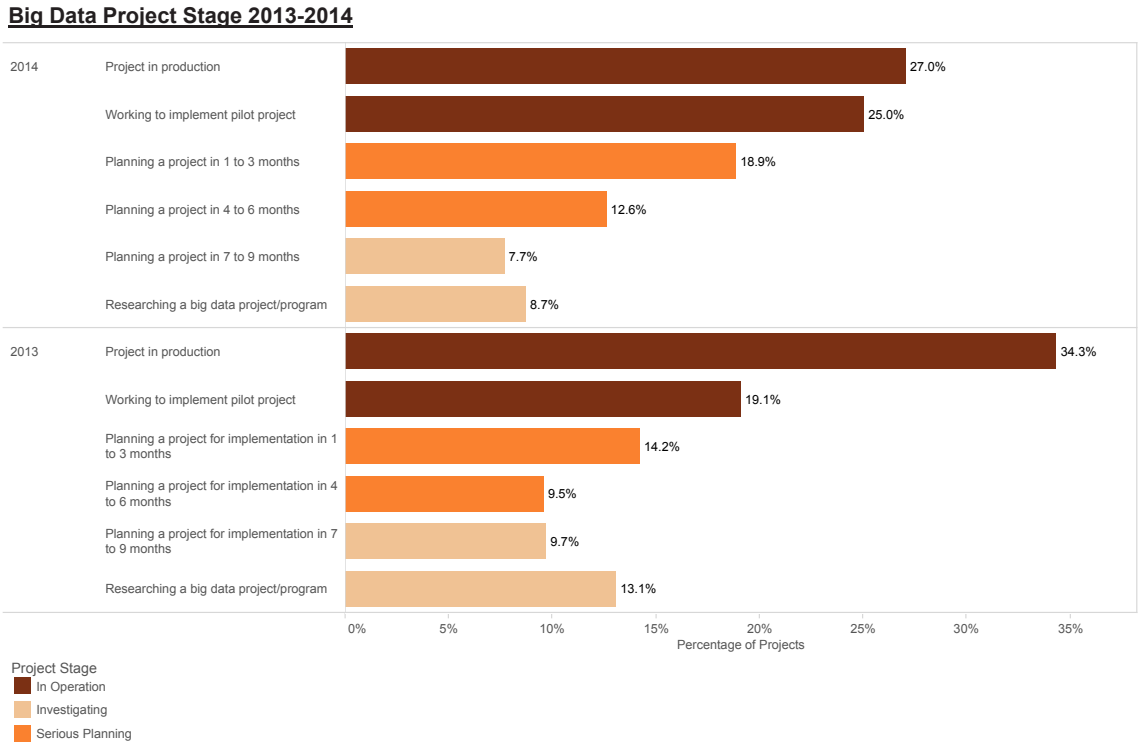


Figure 9

Between 2013 and 2014, there was a significant shift in the proportions of **Project in production** and **Working to implement a pilot project**. In 2013, **In Production** led **Pilot** by 15%; in 2014, the difference dropped dramatically to 2%. This shift may correlate to a significant uptick in the number of IoT projects being piloted in the past year.

Between 2013 and 2014, there was a significant shift in the proportions of Project in production and Working to implement a pilot project. In 2013, In Production led Pilot by 15%; in 2014, the difference dropped dramatically to 2%.

Given the rather small increase in **Machine-generated** data noted above and the relatively recent explosion of Internet-connected devices other than smartphones, these figures may represent the strategic aspirations of a large swathe of the industry and the initial rollout of pilot applications using existing data sources, both internal and external. Looking at the connected device types for IoT, geolocation (largely associated with smartphones) leads the pack with nearly 15%. Given the penetration of smartphones in the consumer space, the pathfinder role for smartphones in the IoT market is becoming clear.

Copyright 2015, EMA Inc. and 9sight Consulting. All Rights Reserved.

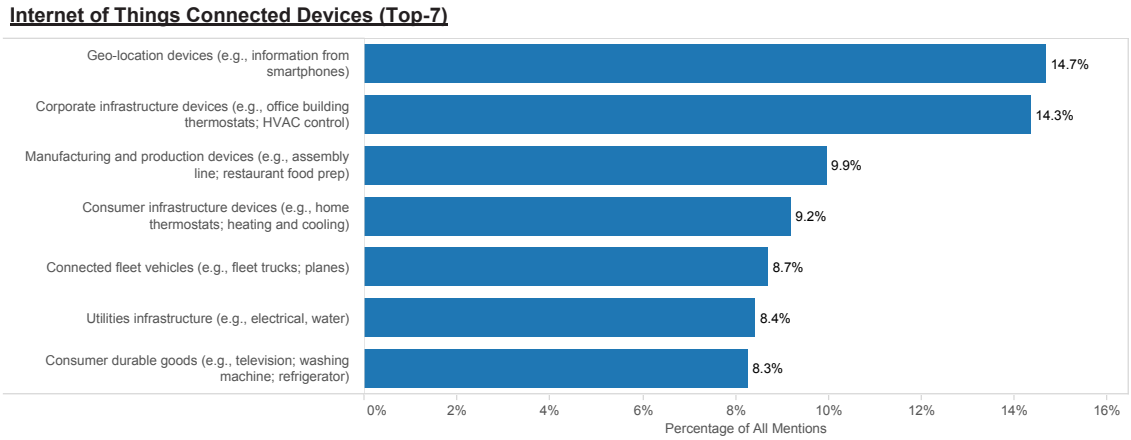


Figure 10

Recognizing that smartphones can be consolidation points for data collection from biomedical devices, consumer vehicles, and home automation, their central role in driving IoT is clear. A range of building and production monitoring devices, both corporate and consumer, take the next three places in the table. However, the span of percentages is rather small across the range of device types, indicating that IoT is expanding along a very wide front.

Recognizing that smartphones can be consolidation points for data collection from biomedical devices, consumer vehicles, and home automation, their central role in driving IoT is clear.

In terms of the **Primary Use Case** for the big data program, **Utilization of Streaming Data**, which corresponds strongly with IoT, has jumped from sixth place in 2013 to third in 2014, by almost 17%.

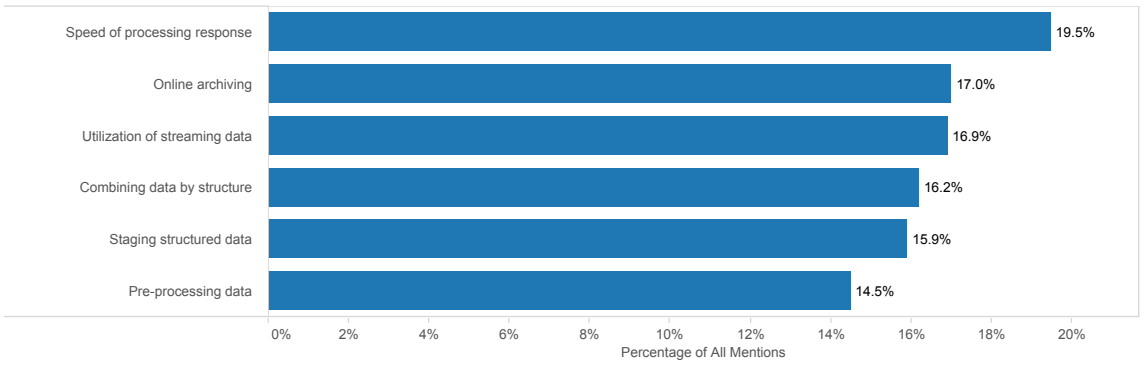


Figure 11

Copyright 2015, EMA Inc. and 9sight Consulting. All Rights Reserved.

Obstacles to IoT implementation fall into three broad areas: business and organization issues (37%), data collection and use (34%), and privacy/quality concerns (29%).

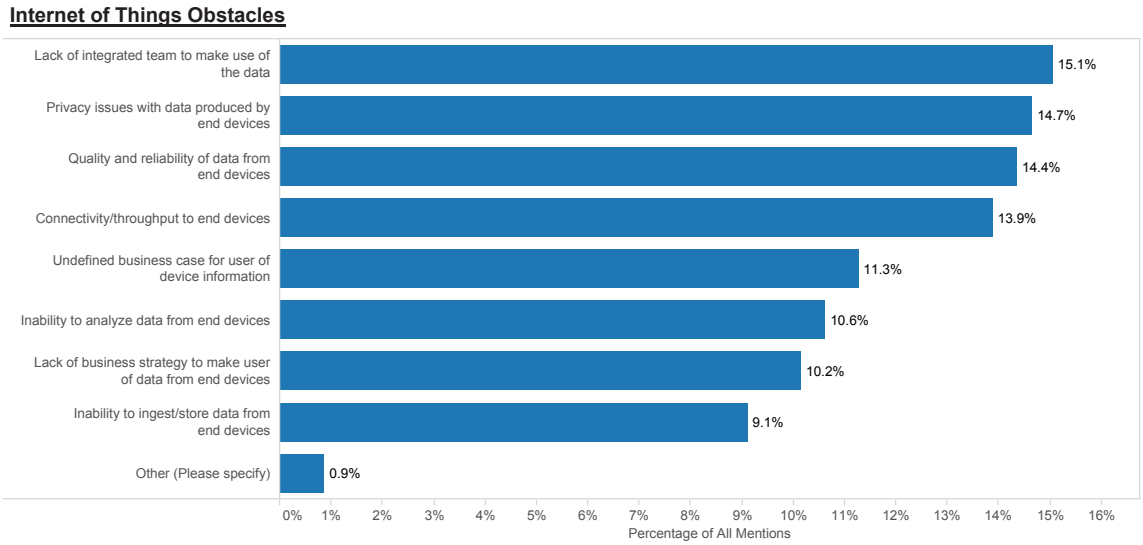


Figure 12

It is encouraging to note that approximately two thirds of respondents recognize the non-technological concerns around business, organization, privacy, and quality as important. It is only in addressing precisely these issues that acceptance of IoT in business and society at large can be achieved.

2.4. Filling Up the Data Lake

The concept of a data lake has become popular, if rather contentious, during the past 12 to 18 months. A data lake is, in its simplest usage, a Hadoop-based storage area for all the data—in whatever form or structure—that a business might ever gather. The metaphor of a lake offers, of course, the opportunity to talk about water and data flowing in freely, with users able to dip in with ease for whatever cupful they need. Playful images of recreational use suggest the freedom and fun that business users would have if only they did not have to worry about where the data comes from or how it is structured. It is suggested that data is like crystal clear lake water, waiting to be consumed. Deeper thinking, even at the level of the lake metaphor, reminds us that there is more to it. Lake water must undergo significant treatment before it is considered fit to drink. Many lakes are filled with effluent from the rivers that feed them. Even the pleasure seekers on the lake understand that there may be dangerous shallows or hidden rocks. Some analysts have begun to talk of the dangers of a data swamp, where misunderstood and unclassified data sinks and rots, waiting to suck down unwitting users.

Beyond the word games rests the serious question of what business requirements this data lake would satisfy. Externally-sourced big data, including that from the Internet of Things, clearly demands a data management store as information arrives at high speed and in significant volumes. This store must be large, flexible, and cost-effective. Hadoop has proven itself useful for these needs. The issues that arise lie in the challenges for data scientists in understanding the meaning of the data and assuring its cleanliness for analysis, and in the barrier presented to moving from exploration to production modes of operation. These issues present significant technical challenges to software vendors and serious data governance concerns for businesses. In the case of noncritical data, the risks can be mitigated.

Copyright 2015, EMA Inc. and 9sight Consulting. All Rights Reserved.

However, expanding the scope of the data lake to include business-critical data and legally binding core business information without addressing some of the perceived shortcoming associated with the data lake architecture seems particularly dangerous.

Uptake of the data lake among survey respondents is mixed. At one end of the scale, over 12% have **used it to replace their data warehouse environment**. Another 22% say that a data lake is **Currently adopted as part of the strategy to replace existing operational and informational platforms**. At the other end of the scale, nearly 20% of respondents have **no plans for research or adoption**. In light of the concerns voiced above, these numbers give rise to concerns that issues of data quality and governance may be taking second place in a drive for cheaper hardware and software solutions.

Uptake of the data lake among survey respondents is mixed. At one end of the scale, over 12% have used it to replace their data warehouse environment. Another 22% say that a data lake is Currently adopted.

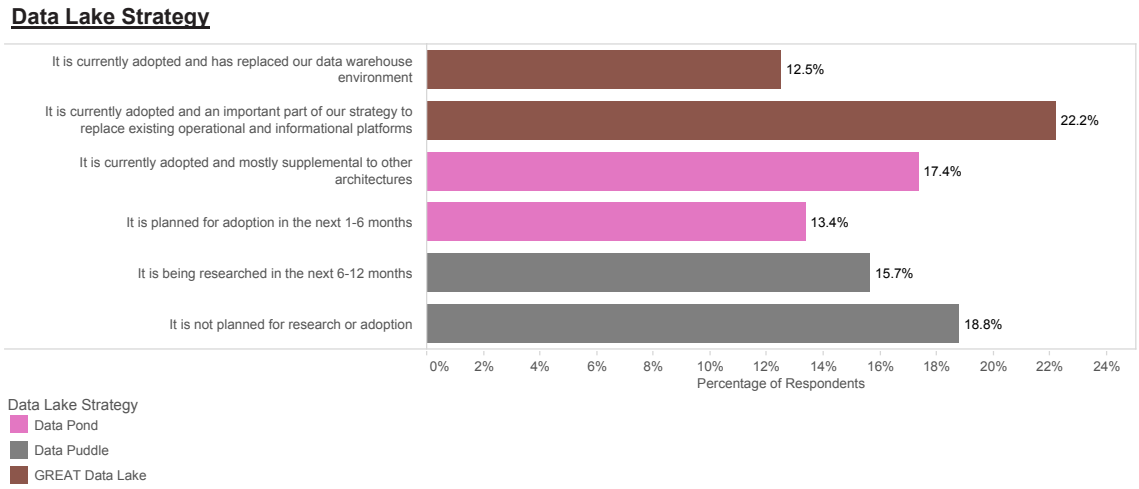


Figure 13

Answers from the panel respondents are segmented into the following analysis bands.

- **Great Data Lakes** are organizations that have truly jumped into the strategy of establishing a data lake. These organizations have indicated a **Data Lake Strategy** of **Adopted and Replaced Data Warehouse** and **Adopted and Important.**
- **Data Ponds** are organizations that are testing the concept of a data lake architecture. These organizations have indicated **Currently Adopted and Supplemental** and **Being Researched for Adoption in 1-6 months** as their **Data Lake Strategy**.
- **Data Puddles** are organizations that still have not embraced the concept of the data lake. These organizations are pushing the establishment of a data lake to late 2015 or beyond with **Being Researched for Adoption in 6-12 months** and **Not Planned for Adoption** indications for **Data Lake Strategy**.

Implementation levels vary significantly across industries. Retail shows the highest uptake by far, which may reflect the traditional cost consciousness of the industry together with a surfeit of data. Utilities and health care, where regulatory compliance and data governance have high priority, have the lowest adoption rates.

Copyright 2015, EMA Inc. and 9sight Consulting. All Rights Reserved.

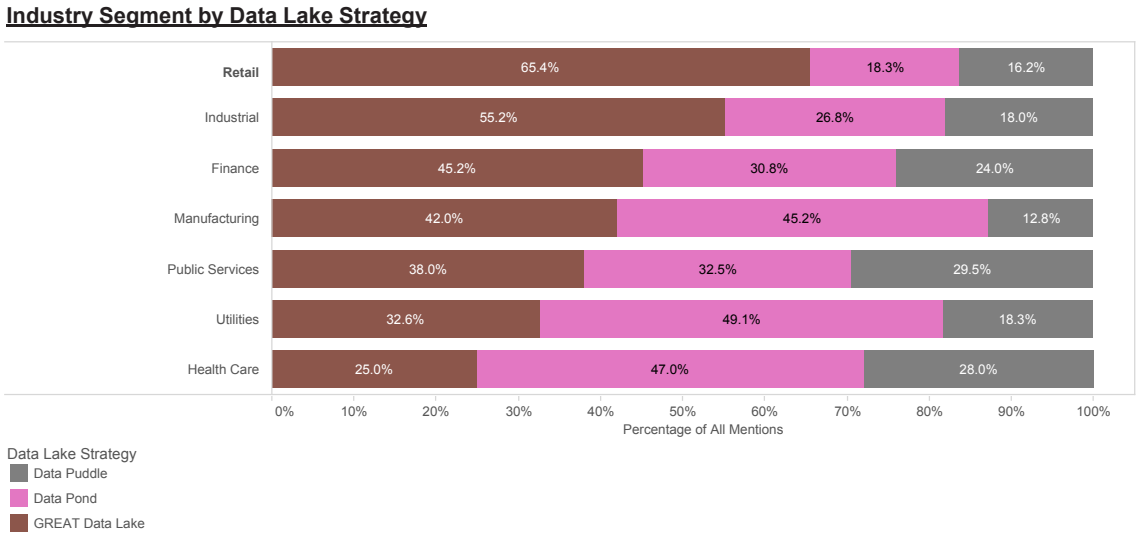


Figure 14

2.5. Where Does Big Data Stand Today?

In summation, 2014 saw the overall growth in big data adoption beginning to slow. Much of this slowdown may be due to the emergence of the IoT as a serious consideration in most industries. In effect, interest has moved, at least for the moment, from size as the primary driver of growth to initial investigations into the implications of structure and speed as defining characteristics of the next phase of adoption. With the infrastructure for the IoT still in the early stages of development, this movement among the respondents is most likely based on existing sources of **Machine-generated** data, both internal (from existing and upgraded production and monitoring equipment) and external (from the burgeoning growth of smartphones). Such early stage experimentation suggests high expectations for IoT across the industry. Encouragingly, the business and data obstacles are receiving higher levels of attention than the purely technical ones.

Such early stage experimentation suggests high expectations for IoT across the industry. Encouragingly, the business and data obstacles are receiving higher levels of attention than the purely technical ones.

From the point of view of business adoption, there is increasing interest and broadening adoption of big data both across industries and across functions within individual businesses.

Among the more significant trends on the business front were a significant (up to 5%) switch in big data use from sales to marketing and an 11% shift in funding of big data projects from sources other than the annual IT budget back to the IT budget.

As big data matures, many more pauses can certainly be expected, particularly as aspects of growth and shifts in emphasis in new technology, both hardware and software, emerge. As the market currently pauses and reorients, it seems an appropriate moment to consider what is next for big data in a more holistic sense. Will there be a continued evolution along current lines, or is something different in store? The former choice may be more appealing and easier to justify based on the trends in the survey, but the latter, more revolutionary answer is one explained here.

Copyright 2015, EMA Inc. and 9sight Consulting. All Rights Reserved.

2.6. Future of Ethics in Big Data – Security and Privacy

An increasing number of experts, academics, privacy advocates, and even members of the public have been speaking out and taking action at various levels about the ethics relating to use and implementation of big data initiatives[3]. In the 2014 EMA/9sight Big Data survey report, the authors expressed a hope that providers and implementers of big data solutions would actively consider the ethical implications of the tools and programs they were rolling out.

At a conceptual level, a number of statisticians and analytics experts have begun to question the widespread assumption that big data analytics can replace hypothesis creation/checking, eliminate bias, and automate most or all decision making. Lawyers and ethicists have begun to examine the societal implications of pervasive data collection. At a practical level, based on fears about privacy and in the light of a spate of major security breaches, the public is beginning to take actions like using ad and cookie blockers and encrypted communications, to limit the data collected by large corporations and the government. For now, these are small steps, but taken together, and with the near certainty of further major data privacy breaches, the playing field looks to be set to change.

The beginning stages of establishing ethical guidelines reside in the privacy policies and security practices associated with organizations that implement big data. The EMA/9sight panelists were asked about their privacy policies. The most popular indication was that respondents' organizations were implementing **Enterprise-wide data protection policies** (nearly 19%).

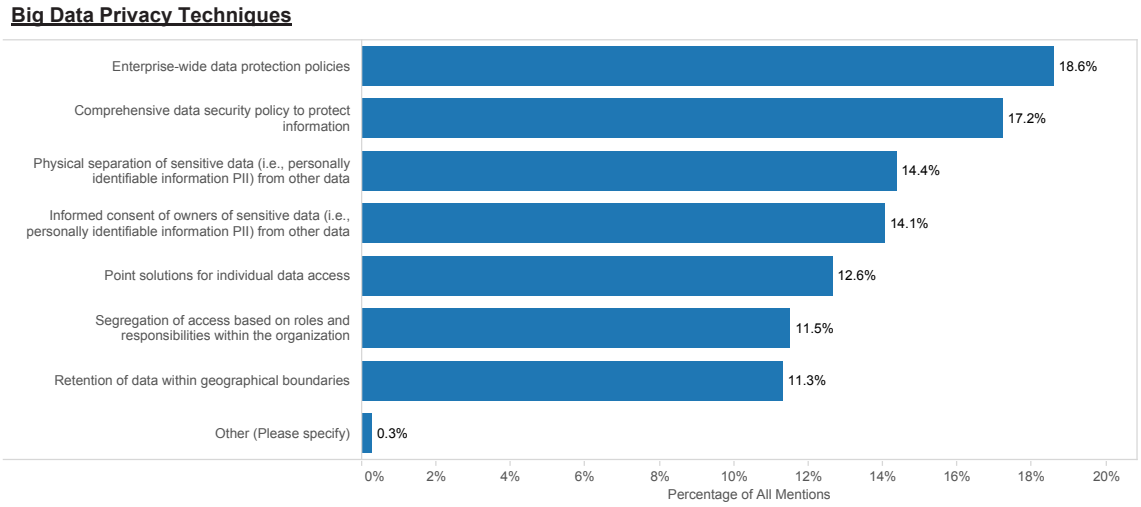


Figure 15

Respondents also indicated that **Comprehensive data security policies to protect information** and **Physical separation of sensitive information** was part of these privacy policies.

[3] http://www.informationweek.com/big-data/big-data-analytics/data-scientists-want-big-data-ethics-standards/d/d-id/1315798

Copyright 2015, EMA Inc. and 9sight Consulting. All Rights Reserved.

Linked to those privacy policies with a strong bent toward security; the EMA/9sight panel respondents indicated that they were taking steps to secure the information within their big data environments. The most selected security practices option was **Encryption of data at rest in a data store,** with over 18% of all mentions.

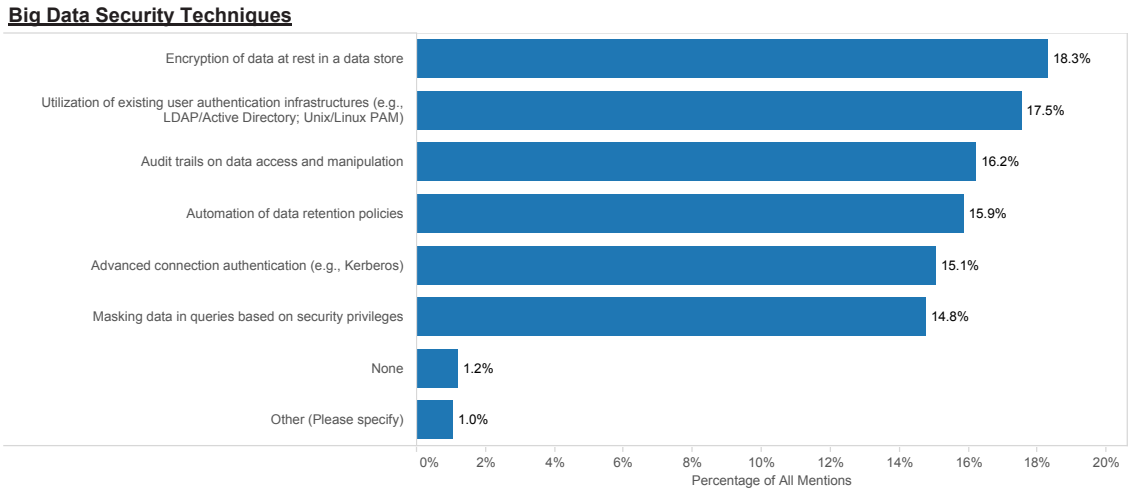


Figure 16

The next two security practices were **Utilization of existing user authentication infrastructure** (nearly 18%) and **Audit trails on data access and manipulation** (over 16%). In all of these cases, EMA/9sight panel respondents indicated that they are working to protect the information within their next generation data management environments to promote proper privacy policies.

The financing of much of the Internet is currently based on a personalized advertising revenue model, which depends on collecting and analyzing extensive personal and behavioral data. Given consumers' growing awareness of privacy/security challenges, vendors and implementers might be advised to consider the implications of this shift as part of their strategy.

3. Hybrid Data Ecosystem: Big Data Equals a Big Tent

Since its inception, the concept of "big data" has been widely associated with a single data management platform—Hadoop. This connection may be due to the popularity of the use of the Hadoop platform to **store** and **process** large amounts of **multi-structured** data. However, Hadoop is not the only platform that can meet these requirements.

Many data management platforms can **store** large amounts of information. Traditional relational databases have served the needs of organizations such as AT&T and Walmart for many years. These environments store the operational and analytical information associated with over 100 million telecommunications subscribers[4] and a retail organization with nearly $500[5] million in worldwide annual revenues.

Other traditional data management platforms can **process** massive amounts of information. Cray supercomputers[6] serve the processing requirements of government agencies such as the National Oceanic and Atmospheric Administration (NOAA) and National Center for Atmospheric Research

[4] http://www.att.com/gen/investor-relations?pid=5711
[5] http://news.walmart.com/walmart-facts/corporate-financial-fact-sheet
[6] http://en.wikipedia.org/wiki/Cray

Copyright 2015, EMA Inc. and 9sight Consulting. All Rights Reserved.

(NCAR), processing massive amounts of data for oceanographic and atmospheric research and prediction. In smaller environments, the columnar database[7] format became mainstream in the mid-2000s and provided commercial operations with the ability to process mass quantities of information in many disciplines.

Many different platforms can natively support **multi-structured** data. Some, such as Cassandra™[8] and Redis[9] support key-value formats. Others, such as CouchDB™[10] and MongoDB[11] base their approach on a document model. Still other platforms support a wide range of non-relational data models that include graph and object.

In each of these cases, a wide range of options can support many of the common concepts of big data initiatives. Big data was, and is, bigger than a single data management platform or paradigm. Big data initiatives encompass a wide range of business, technical, and financial requirements to move from our previous era of data—whether you call that traditional, relational, or enterprise—to a new era of data. This new era of big data is more than a "rip and replace" of previous times. It is an augmentation of traditional approaches that includes new capabilities and possibilities.

3.1. Hybrid Means More than Multi-Structured

The first EMA and 9sight Big Data survey in 2012, "Big Data Comes of Age," made very clear that big data implementers and consumers were taking a big tent view; their business needs were so broad that their solution could not be limited to a single platform. Their implementation choices were not being dictated by technical limitations of the underlying platforms. Rather, the emerging big data initiatives were being driven by a series of specific requirements, which EMA has since consolidated into the following five core requirements of big data initiatives.

- **Response** – New technology platforms such as big data tools and frameworks are at the core of this evolution, powering new solutions and improved speed of results.
- **Economics** – Big data platforms leverage commodity hardware, and the software is often free, substantially reducing financial barriers to adoption.
- **Workload** – Big data platforms play a role within the ecosystem to execute extremely complex analytic workloads; innovative companies are willing to invest early in these solutions to gain competitive advantage.
- **Load** – Data loads are growing more complex and the sources are more diverse. Driven by greater complexity and demand, big data adoption is driven by the need to provide flexibility.
- **Structure** – Data structure and schema flexibility are key to the foundation of big data utilization and adoption.

It is not just Hadoop and similar "new-age" platforms that meet the requirements of the big data era. Elements of the previous era of data management that include relational data from operational platforms such as enterprise resource planning (ERP), customer relationship management (CRM), and supply chain management (SCM), as well as analytical platforms such as enterprise data warehouses (EDW) and data marts are also part of the picture. This combined set of old and new data structures and data management platforms are woven into a hybrid environment that serves many different workloads in this new era of big data.

[7] http://en.wikipedia.org/wiki/Column-oriented_DBMS
[8] http://en.wikipedia.org/wiki/Apache_Cassandra
[9] http://redis.io/
[10] http://couchdb.apache.org/
[11] http://www.mongodb.org/

Copyright 2015, EMA Inc. and 9sight Consulting. All Rights Reserved.

EMA refers to the collection of big data management platforms and implementations as the **Hybrid Data Ecosystem** (HDE). These platforms include the following.

- **Operational Platforms –** Manage day-to-day transactions, resources, and billing management
- **Enterprise Data Warehouses –** Support "single version of the truth" corporate information
- **Data Marts –** Subject-specific and support individual analytical and reporting requirements
- **Analytical Database Platforms/Appliances –** Perform ad-hoc analysis across large data sets
- **Data Discovery Platforms –** Explore relational and multi-structured data sources
- **Hadoop Platforms –** Based on Apache Hadoop and its associated sub-projects
- **NoSQL Data Stores –** Handle multi-structured data sets and may not directly support data access standards such as the standard query language (SQL)
- **External Data Sources –** Provide business processes, contextual data, and other functionality

Each of the platforms within the **Hybrid Data Ecosystem** supports a particular combination of business requirements; and operational or analytical processing challenges. The **HDE** represents a unique approach when compared with traditional best practices. Rather than advocating a single data store that supports all business and technical requirements at the center of its architecture, the **Hybrid Data Ecosystem** seeks to determine the best platforms for supporting a particular set of requirements and link those platforms together.

3.2. Expanding to the Cloud

The platforms associated with the **HDE** are implemented in a wide range of avenues. It is not only on-premises or managed services data centers that house these data management platforms; private cloud environments located in multiple data center locations with all components situated behind the corporate firewall support these environments. Powerful and distributed public cloud resources power these environments. Even a combination of public and private cloud resources are being used to meet these requirements.

EMA/9sight panel respondents were asked about their choices to implement the platforms of the HDE from the following implementation avenues.

- **Bare Metal** installations are on-premises data center installations. They are often known as "bare metal" installations since the data management platform is installed on a server where you can touch the "bare metal."
- **Public Cloud** environments allow organizations to utilize a third party's infrastructure for the elastic provisioning and operation of data management platforms. Examples of public cloud infrastructures include Amazon Web Services (AWS) and OpenStack[12] environments from various vendors such as Rackspace and HP.
- **Private Cloud** includes the elastic provisioning and operation attributes of public cloud resources, but is implemented within the firewall of a data center.
- **Hybrid Cloud** allows organizations to securely store their data within the firewall of a private cloud environment and flexibly use additional resources in public cloud environments.
- **Managed Service** are those in which an organization has fully encapsulated the technical complexity of the operations of its data center to another organization. These environments can operate as a data center with bare metal installations or can operate in a manner similar to cloud environments.

[12] http://en.wikipedia.org/wiki/OpenStack

Copyright 2015, EMA Inc. and 9sight Consulting. All Rights Reserved.

Each of these implementation avenues has benefits and drawbacks for the implementation of a next generation data management environment such as the EMA Hybrid Data Ecosystem. On-premises **Bare Metal** installations provide the most security and hands-on administration. However, **Bare Metal** installations often have the highest administration costs as well as offer limited flexibility for implementation growth when compared with cloud implementations. **Private**, **Hybrid**, and **Public Cloud** implementations offer flexibility of provisioning and elastic operations, yet there can be issues with data transfer and data security in **Public** and **Hybrid Cloud** implementations. **Private Cloud** implementations can run into similar elasticity issues as other traditional data center implementations if internal teams do not plan sufficiently. **Managed Services** implementation provides the highest level of encapsulation from administrative levels of effort, but can also have issues in the transfer of data to and from those locations.

The graph below shows which implementation avenues participants in the 2014 EMA/9sight big data end-user survey use to implement their components of the Hybrid Data Ecosystem.

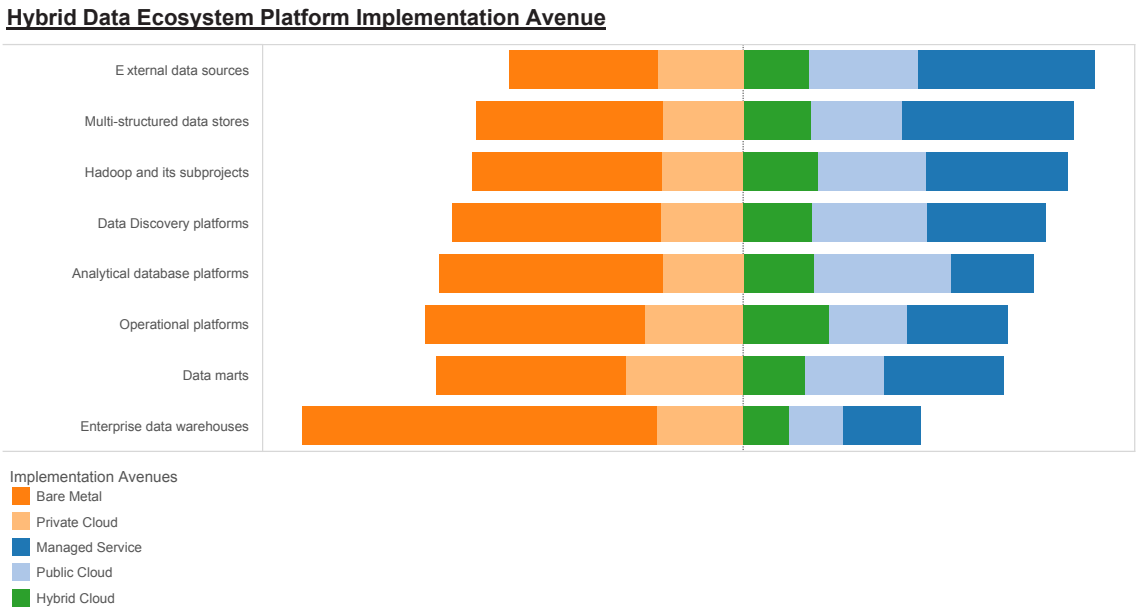


Figure 17

Bars on the right side of the graph represent platforms that are more likely to be implemented outside of the data center. Bars on the left side of the graph symbolize platforms that are more likely to be implemented in the data center as a **Bare Metal** installation.

Not surprisingly, **Enterprise data warehouses** and **Data marts** are the most likely platforms to be implemented via a **Bare Metal** installation, due to the long history of deploying these environments in the data center. Yet it is not unusual to see **Enterprise data warehouses** and **Data marts** implemented outside of the data center. The most popular option for such implementations is via **Managed Services.**

On the other hand, **External data source** platforms are often implemented outside the data center as a service from third party vendors. **External data sources** lend themselves to a **Managed Service** or **Cloud**-based implementation, with many organizations accessing a single demographic or firmographic set of data, or individual organizations accessing specific aspects of the external data that they are interested in.

Copyright 2015, EMA Inc. and 9sight Consulting. All Rights Reserved.

Between these two ends of the continuum, HDE platforms such as **NoSQL data stores**, **Hadoop platforms**, **Data Discovery platforms**, and **Analytical database platforms/appliances** have varying degrees of deployment, either in the data center or in the cloud.

3.3. Data Consumers and the Cloud

Just as data management platforms associated with the HDE are impacted by the continued influence of cloud implementation avenues, **Data Consumers** are also impacted. When asked about how **Implementation Avenues** such as **Bare Metal Installation** in data centers and various levels of cloud implementations were associated with the **Data Consumers** of their big data projects, EMA/9sight panelists indicated that some traditional line of business (LOB) data consumers were the most likely to be using projects based in the data center. **Line of Business executives** (nearly 54%), **Business Analysts** (almost 42%) and **IT Analysts** (over 40%) were the most likely to be associated with projects based in the data center.

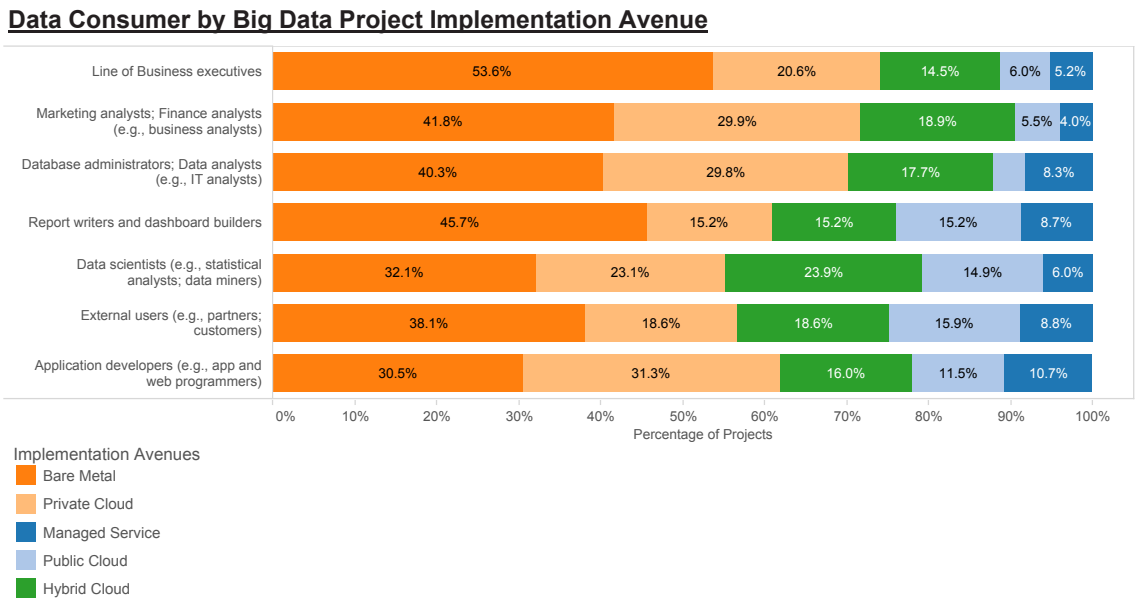


Figure 18

Conversely, **Application developers** such as application and web developers (over 10%) and **External users** including partners and customers (nearly 9%) were the most likely to be associated with **Externally Managed** projects.

In the middle, **BI Analysts** such as report writers and dashboard builders (over 60% between Bare Metal and Private Cloud) lean toward **On-premises projects** and **Data Scientists** (nearly 39% between Hybrid and Public cloud) lean toward projects outside the typical firewall.

3.4. Updates to the Hybrid Data Ecosystem

In 2013, EMA expanded the definition of the **Hybrid Data Ecosystem** to include **Information and Data Management** and a focus on **Information Consumers**. The EMA/9sight survey results in 2014 have also provided additional refinements.

Copyright 2015, EMA Inc. and 9sight Consulting. All Rights Reserved.

- **Impact of Cloud Implementations** – This year's version of the **HDE** includes the influence and impact from cloud implementations. Both the **Data Consumers** and **Data Management** platforms are influenced by the implementation choices afforded by the cloud. This year's HDE demonstrates those influences by grouping **Data Consumers** and **Data Management** platforms together.
- **Refined Workloads** – The concepts of **Speed of Response** and **Complex Workload** were established in 2012 as key components of the **Hybrid Data Ecosystem** requirements. The 2013 HDE identified the workloads that big data initiatives are tackling. They include **Operational, Analytical**, **Exploratory**, and **Operational Analytics** workloads. This year, EMA refined those workloads to include expectations for processing latency. Organizations expect **Operational Analytics** and **Operational** workloads to be completed in near-real time. **Analytical** workloads can be completed in longer time frames and **Exploratory** workloads have an even longer expectation.

The following details the 2015 EMA **Hybrid Data Ecosystem**.

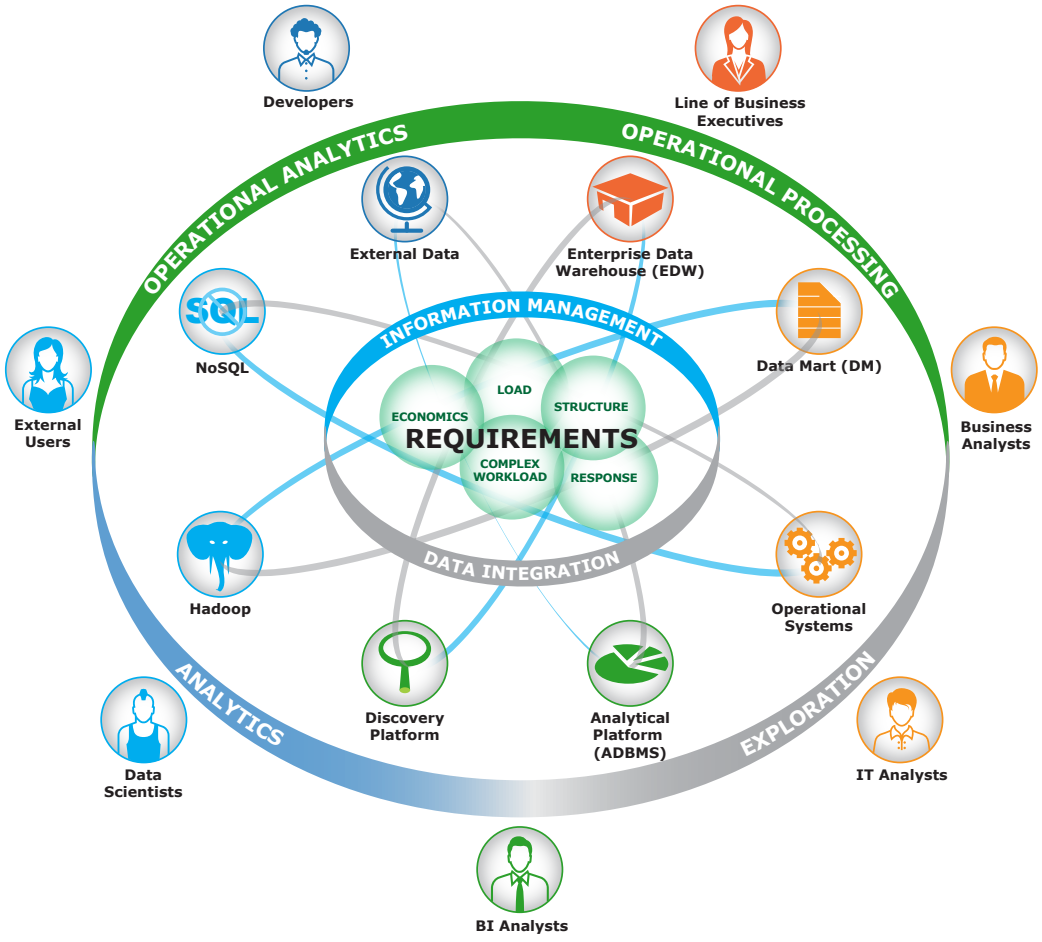


Figure 19

The **HDE** is supported by over three years of extensive research on big data initiatives both quantitatively via end-user surveys and qualitatively via interviews with thought leaders in technology, implementation specialists, and software vendor experts.

Copyright 2015, EMA Inc. and 9sight Consulting. All Rights Reserved.

4. 2014 EMA/9sight Survey Demographics

The 2014 EMA/9sight survey respondents represent a wide range of industries, company sizes, and geographic distribution. The survey process did not set particular company sizes, industries or geographic distributions as minimum requirements. The survey process did pull together a panel with insights into the strategy, budgets, and implementations of big data initiatives around the globe. This organic diversity of panel respondents provided a well-balanced view into the big data activities of data management technologists and business stakeholders worldwide.

4.1. Company Size

The EMA/9sight survey included organizations across a range of sizes. Corporate headcounts for organizations in the survey panel are distributed in the following manner.

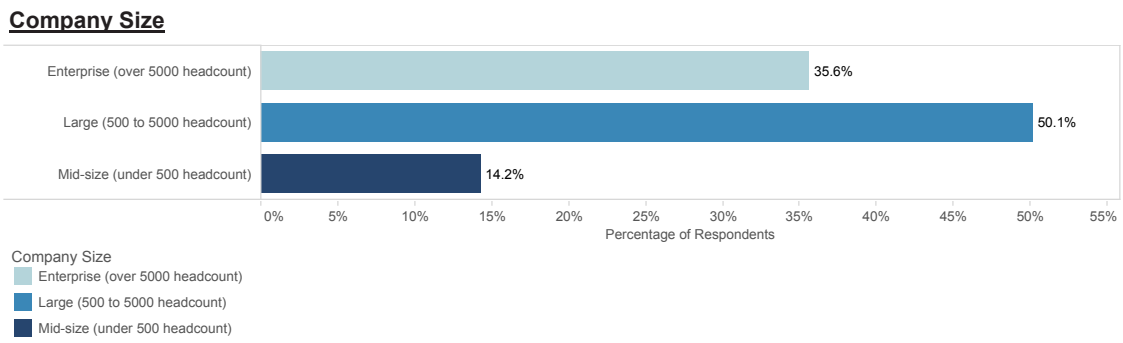


Figure 20

In comparison, this distribution is different from the distribution of company size from the 2012 and 2013 surveys. The trend of organization size is displayed here.

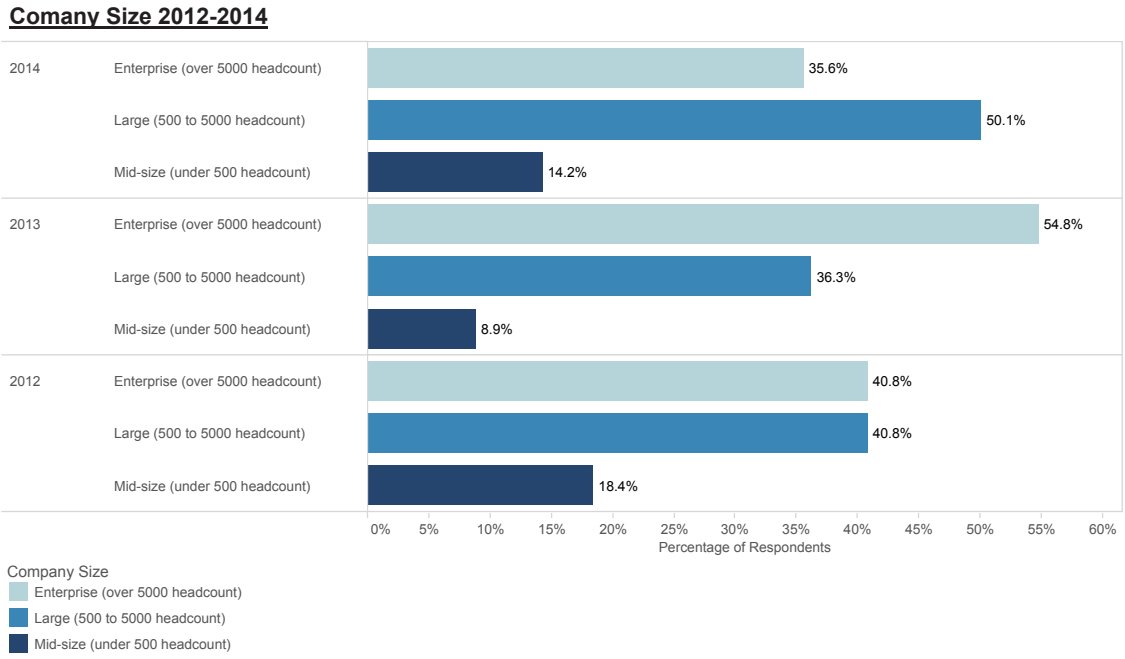


Figure 21

Copyright 2015, EMA Inc. and 9sight Consulting. All Rights Reserved.

Enterprise companies (500+ headcount) vary between 36% in 2014 and 55% in 2013 of the panel respondents. **Large** organizations fluctuate in a similar range of 36% in 2013 and 50% in 2014. **Mid-size** organizations, which have a lower number of technology resources required for big data initiatives, have been a lower percentage of the EMA/9sight survey panel over the years. The representation of **Mid-size** organizations has ranged from just under 9% in 2013 to nearly 19% in 2012.

4.2. Segmenting Industries

Any end-user analysis of big data initiatives should take into consideration the various industries and industry segments of the panel respondents. Some industries are on the cutting edge of developments while others are still gaining traction with their big data initiatives.

In this year's study, and in previous editions of the EMA/9sight Big Data research, individual industries were grouped into the following segmentation designations:

- **Finance** – Finance, banking, and insurance
- **Health care** – Medical device and supply and pharmaceutical production
- **Industrial** – Aerospace and defense manufacturing, oil and gas production and refining, chemical manufacturing, and transportation and logistics organizations like airlines, trucking, and rail
- **Leisure** – Hospitality, gaming and entertainment, as well as recreation and travel
- **Manufacturing** – All non-computer or networking-related manufacturing industries
- **Media and PR** – Marketing, advertising, public relations, and market research, and publishing and broadcasting
- **Public Services** – Government, education, non-profit/not for profit, and legal
- **Retail** – End consumer retail and wholesale and distribution
- **Utilities Infrastructure** – Telecommunications service providers, application, internet, and managed-network service providers, and energy production and distribution utilities

The breakdown by industry segment for the 2014 panel respondents is as follows.

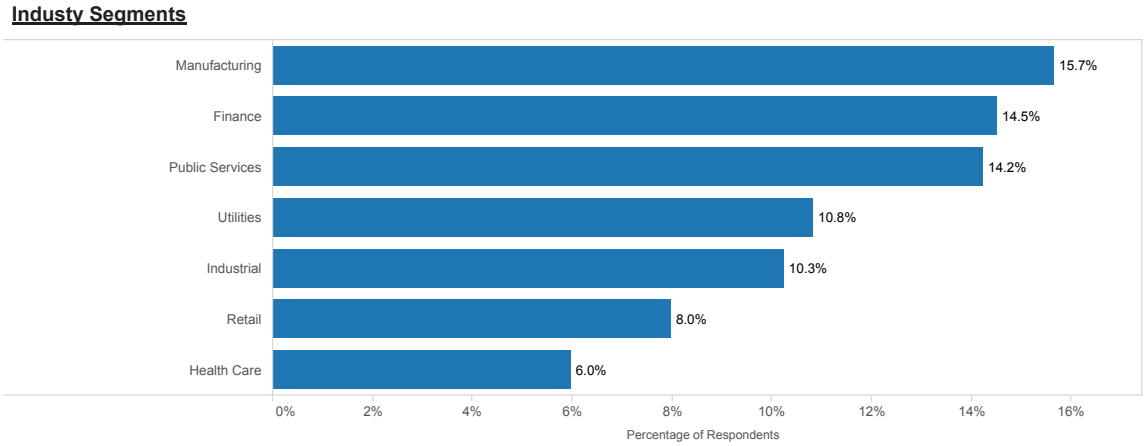


Figure 22

Copyright 2015, EMA Inc. and 9sight Consulting. All Rights Reserved.

In 2014, the number of the **Leisure** and **Media & PR** industry segments respondents was not statistically significant. Therefore, these industry segments are not included in the analysis for 2014. However, for overall comparison, those segments are represented in the trending analysis 2012 and 2013 result sets below.

In 2014, the top industry segment was **Manufacturing**. The **Finance**, **Public Services** and **Utilities** segments followed.

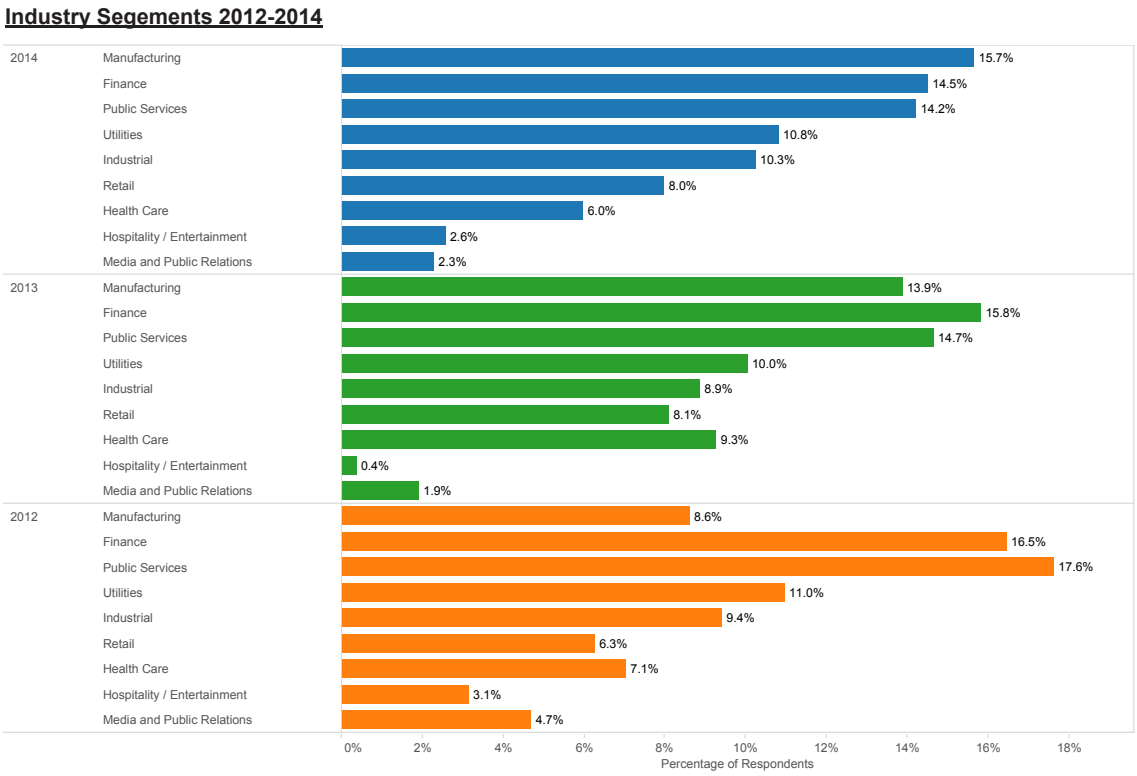


Figure 23

In 2014, the **Manufacturing** segment increased its representation from just under 14% in 2013 and 9% in 2012. The **Finance** segment has maintained a strong representation in the EMA/9sight survey with approximately 14% in 2014, 16% in 2013, and nearly 17% in 2012. With the data intensive nature of the financial industry, it is not surprising to see this type of continued significance within the EMA/9sight survey panel. The **Public Services** segment has maintained a significant representation within the panel respondents. With almost 18% of the survey panel in 2012, approximately 15% in 2013, just over 14% of respondents is consistent with the importance of big data to organization within the government and non-profit industries for "smart government" initiatives.

4.3. Geographic Distribution

The rate of big data adoption varies from country to country and region to region. This comes from the diverse attitudes regarding the usage of data around the globe. The United States, included in the North American region, has a generally lax set of cultural and regulatory constraints on the usage of information from sources such as social media and personally identifiable information (PII) data. However, Canada, European countries and those from Central and South America have much different mindsets regarding the usage of and the physical location of data sources.

Copyright 2015, EMA Inc. and 9sight Consulting. All Rights Reserved.

The 2014 EMA/9sight panel respondents are well distributed around the globe, providing insights from each of the following regions.

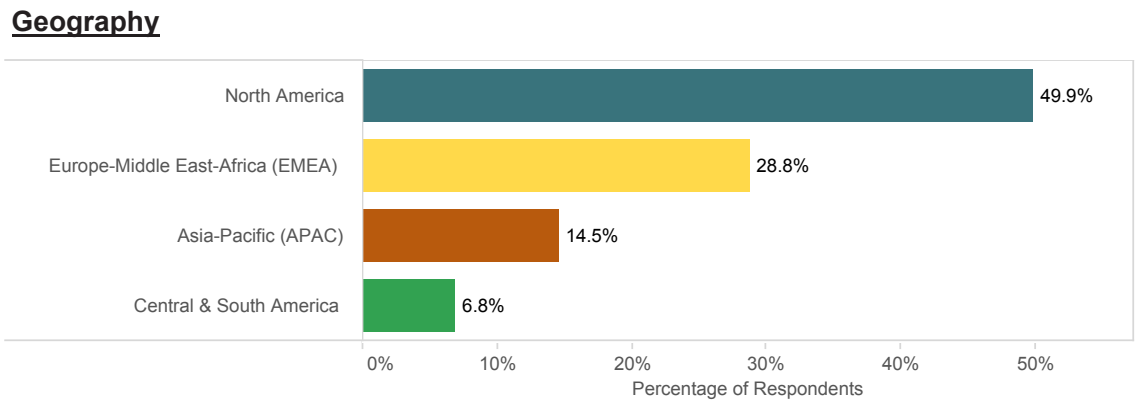


Figure 24

North America has represented a significant portion of the EMA/9sight survey panel respondents since 2012, consistently providing approximately half of the panel respondents, if not more, over the years.

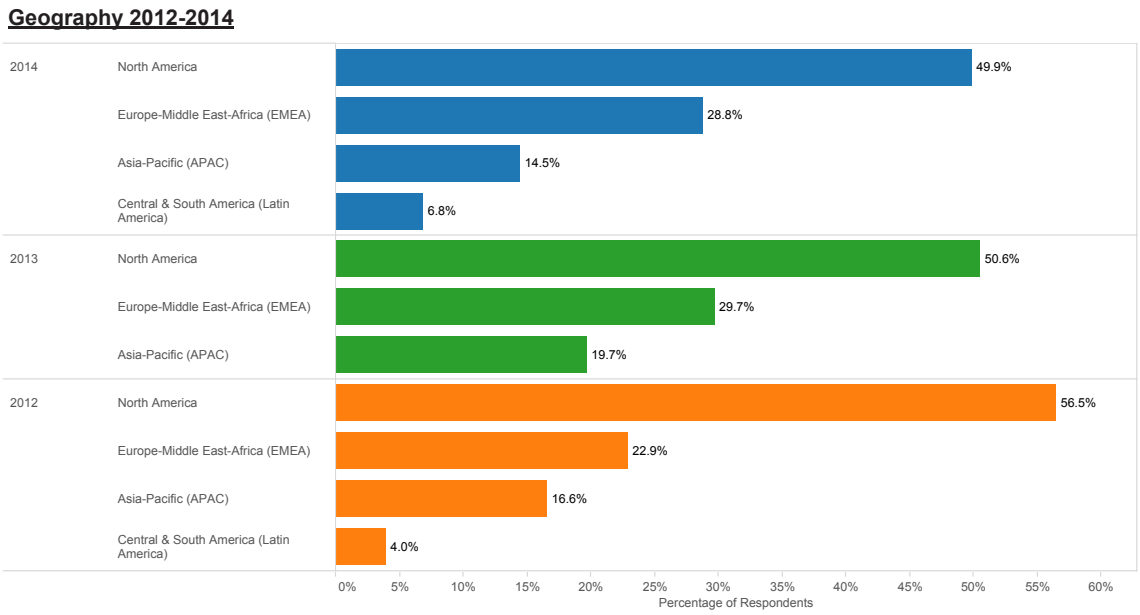


Figure 25

Respondents from the Europe-Middle East-Africa region have also provided a consistent level of input to the EMA/9sight panel with representation ranging from 23% in 2012 to just under 30% in 2013. Asia Pacific (APAC) respondents have also contributed a relatively consistent representation in the survey. APAC respondents provided between 20% in 2013 and 15% in 2014 of the respondent panel over the course of the survey's history.

Central and South American (CALA) respondents have been a relatively fluid set of contributors during the history of this survey. With a relatively small representation in 2012 and no representation in 2013, CALA respondents are back in the 2014 survey panel with 7% representation.

Copyright 2015, EMA Inc. and 9sight Consulting. All Rights Reserved.

5. EMA/9sight Survey Panelist Horizontal Strategies

As part of the EMA/9sight study, panel respondents were asked about their opinions on high-level strategies across their organizations. These horizontal strategies included the importance of big data to their organization, how pervasive the concept of Internet of Things is within their organization, how much they work with the concept of the data lake, and how deeply a data-driven culture is embedded within their organization.

5.1. Big Data

Naturally, as part of an end-user study on big data initiatives, it is important to understand how companies view big data as part of their organization. In previous surveys, the importance of big data was indicated with a "yes/no" as part of the survey qualification process. For 2014, the importance of big data was expanded to include many different aspects from essential to their business to supplemental to other computing paradigms to being researched.

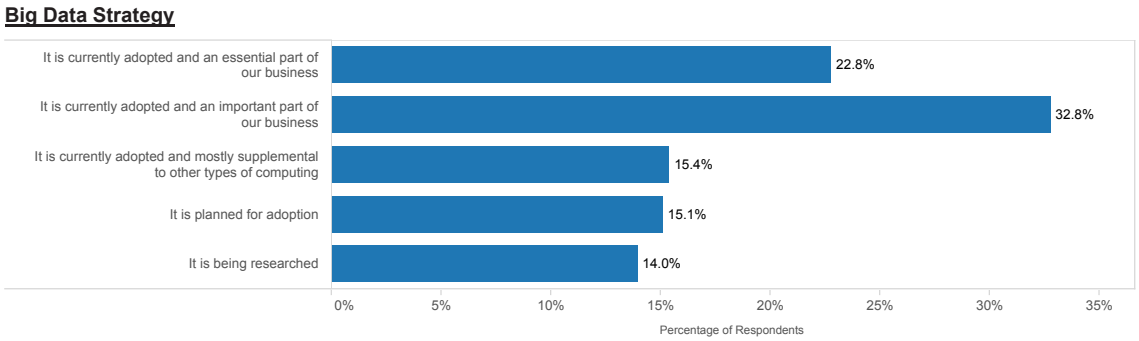


Figure 26

In the 2014 EMA/9sight survey, panel respondents indicated overwhelmingly that big data was part of their organizational strategies. Over 55% indicated that big data was **Adopted and Essential** (nearly 23%) or **Adopted and Important**. Another 15% indicated that big data was **Adopted and Supplemental**.

5.2. Cultural Change: Data-Driven Organizations

A key component of developing a big data strategy is how an organization's culture is focused on data-driven strategies. Without a strong culture surrounding the usage and application of data-driven concepts, including empowering employees with information and encouraging them to use that information in operational, tactical, and strategic decisions, the collection of data from big data sources such as sensor information, application log files, or social media simply becomes a technical exercise that is unlikely to be linked to higher-level business functions.

In the 2014 EMA/9sight survey, respondents were asked to provide information on how they assessed data-driven culture in their organization.

Without a strong culture surrounding the usage and application of data-driven concepts, the collection of data from big data sources such as sensor information, application log files, or social media simply becomes a technical exercise that is unlikely to be linked to higher-level business functions.

Copyright 2015, EMA Inc. and 9sight Consulting. All Rights Reserved.

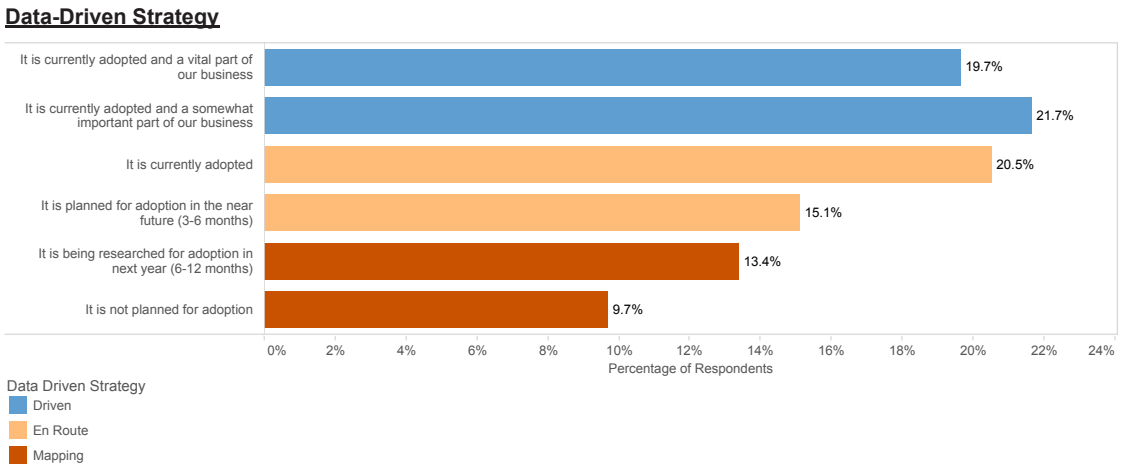


Figure 27

Almost 63% of respondents indicated that data-driven concepts were included in their organization at a significant level. **Adopted and Vital** (nearly 20%) and **Adopted and Important** (almost 22%) were the highest levels of data-driven acceptance. Another 21% indicated that a data-driven culture is **Currently Adopted** in their organization.

Almost 63% of respondents indicated that data-driven concepts were included in their organization at a significant level. Adopted and Vital (nearly 20%) and Adopted and Important (almost 22%) were the highest levels of data-driven acceptance.

Respondents' indications for data-driven culture have been segmented into the following bands.

- A **Driven** culture is one in which organizations have dedicated themselves to empowering employees and integrating information and data into their decision-making processes across many levels. These organizations indicate their **Data-Driven Strategy** as **Adopted and Essential** or **Adopted and Important**.
- Data-driven cultures that are described as **En Route** are those that moving in the direction of a deep integration of data into their collective corporate DNA, but those efforts are still in progress. These organizations described their **Data-Driven Strategy** as **Currently Adopted** and **Being Researched for Adoption in 3-6 months**.
- **Organizations with a Mapping** data-driven strategy as those farthest from that integration of data into their day-to-day operations. These organizations choose **Being Researched for Adoption in 6-12 months** or **Not Planned for Adoption** as their **Data-Driven Strategy**.

A data-driven culture is often a pre-cursor to a big data initiative. Whether originating from business stakeholders or technologists, a main driver for big initiatives is expanding and deepening the application of information within an organization. Data-driven cultures take this driver as one of their main tenets and seek to empower their employee base to utilize many different data sources to impact strategic, operational, and tactical decisions.

Copyright 2015, EMA Inc. and 9sight Consulting. All Rights Reserved.

5.3. Connecting Everything: Internet of Things

IoT is defined by many to be the "collection, processing, and analysis of information from sensors on a large number of disparate devices." These devices can include everything from smartphones to consumer durable goods to commercial-grade appliances to industrial manufacturing equipment. The Internet of Things is a concept that aligns well with big data initiatives because many of the core attributes of IoT such as data collection from multi-structured data sources (for example JSON-based sensor information), data management at large scale, and complex processing are the same as for a big data initiative.

The Internet of Things is a concept that aligns well with big data initiatives because many of the core attributes of IoT such as data collection from multi-structured data sources, data management at large scale, and complex processing are the same as for a big data initiative.

With these parallels, it was important to understand how panelists' adoption of Internet of Things concepts tracked with their adoption of big data.

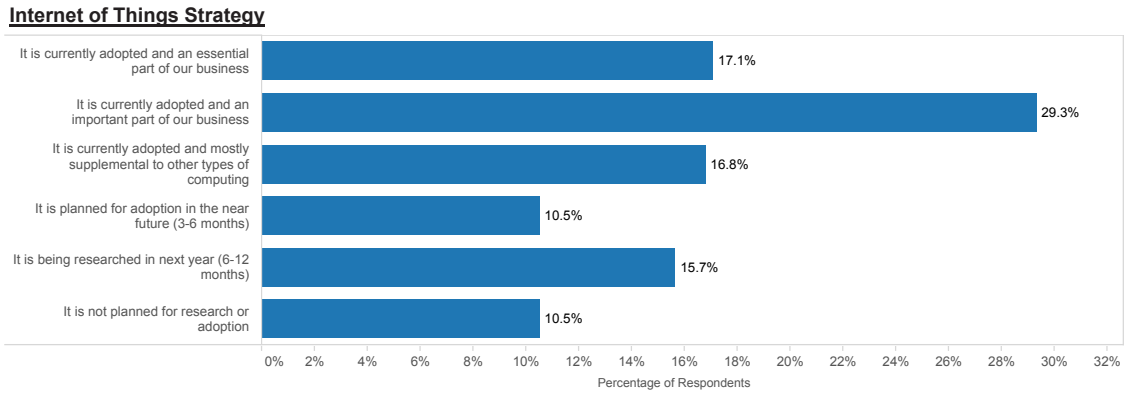


Figure 28

Nearly half of organizations in the EMA/9sight survey indicated the Internet of Things to be of significant importance to their organization. **Adopted and Essential** (17%) and **Adopted and Important** (nearly 30%) were the two highest responses from the panel. **Adopted and Supplemental** (almost 17%) was the third highest response.

Again, the relationship between the EMA/9sight panel responses for big data initiatives and IoT adoption brings to light the association between the higher-level concept of big data and the more concrete use case of Internet of Things.

Copyright 2015, EMA Inc. and 9sight Consulting. All Rights Reserved.

5.4. Data Lakes

With the importance of Hadoop as an economical storage option and a powerful processing platform, a new analysis architecture has grown in visibility over the past 12 to 18 months. This architecture is known as a "data lake." The data lake architecture uses the strengths of Hadoop to store large quantities of information and process that information for operational and/or analytical use cases. This approach is similar to that of an EDW in that information is accumulated, stored, and presented as a single set of facts for the organization.

With the importance of Hadoop as an economical storage option and a powerful processing platform, a new analysis architecture has grown in visibility over the past 12 to 18 months. This architecture is known as a "data lake."

Data lakes were included in the 2014 EMA/9sight survey to determine the extent of the adoption of the data lake architecture.

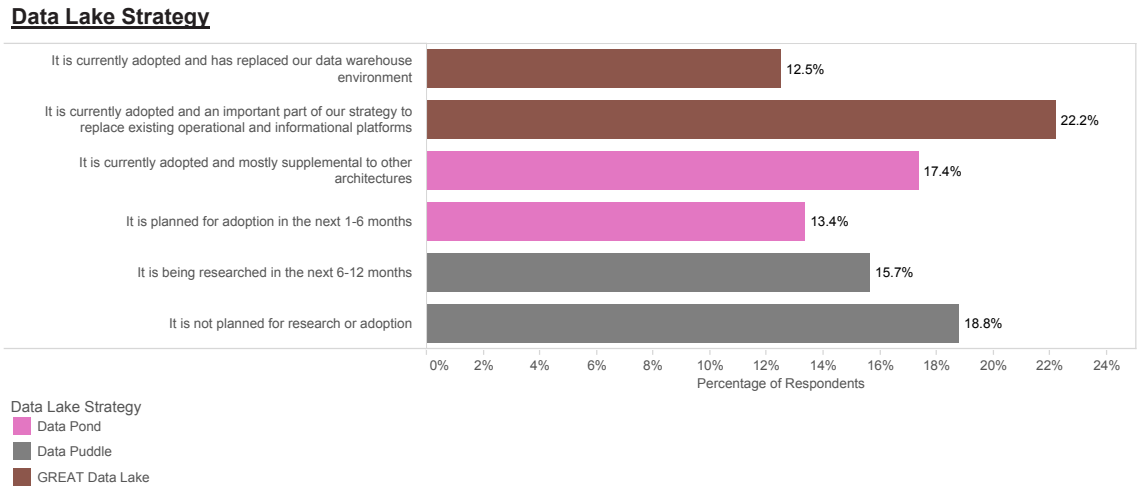


Figure 29

The responses from the survey panel showed an interesting pattern. Almost 35% of respondents indicated that they were using a data lake architecture within their organization. In addition, nearly the same number indicated either they had no plans for adoption of data lake architectures or their implementations were planned for the later parts of 2015, if not 2016.

Copyright 2015, EMA Inc. and 9sight Consulting. All Rights Reserved.

This divergent set of views shows that data lakes are popular with organizations with significant levels of Hadoop adoption. Below is a view of the **Data Lake Strategy** distribution for just those organizations with at least one Hadoop installation. Notably for respondents with Hadoop in their architectures, there is nearly a 50% **GREAT Data Lake** strategy as opposed to just over 35% for the wider EMA/9sight panel.

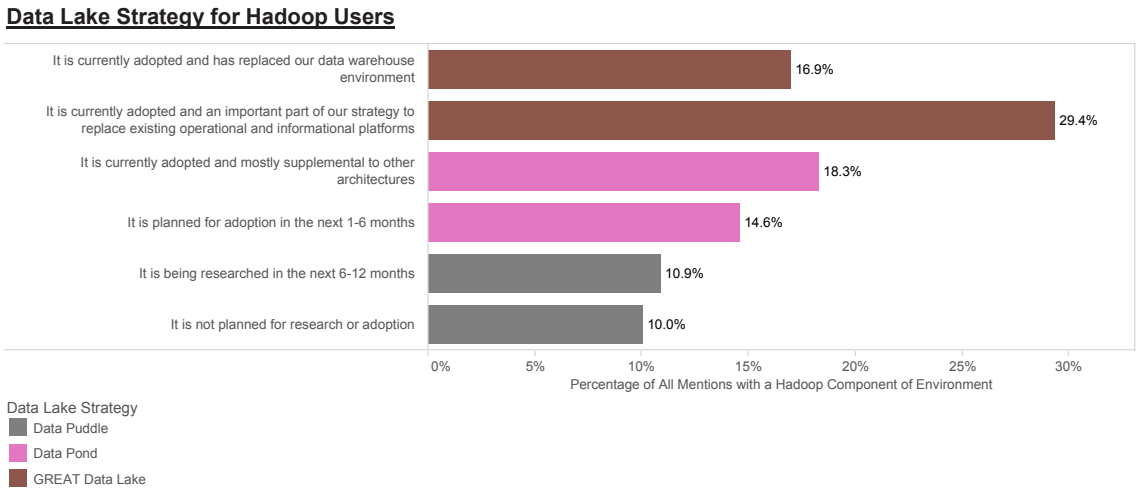


Figure 30

Also, there is a significant decline in the number of organizations that are putting off data lake architectures (nearly 19% for all respondents vs. just over 10% for respondents with Hadoop) and putting off data lakes until late 2015 or 2016 (almost 16% for all respondents and nearly 11% for respondents with Hadoop).

5.5. Use Cases

In order to determine how respondents are using, or planning to use, their big data implementations, EMA/9sight asked respondents which of a selection of **Use Cases** they are implementing. Respondents had the opportunity to select multiple **Use Cases** as appropriate to their situation.

Again, for 2014, **Speed of processing response** was the **Use Case** most frequently mentioned by respondents, although at nearly 20%, its importance has fallen by a little more than three percentage points since 2013.

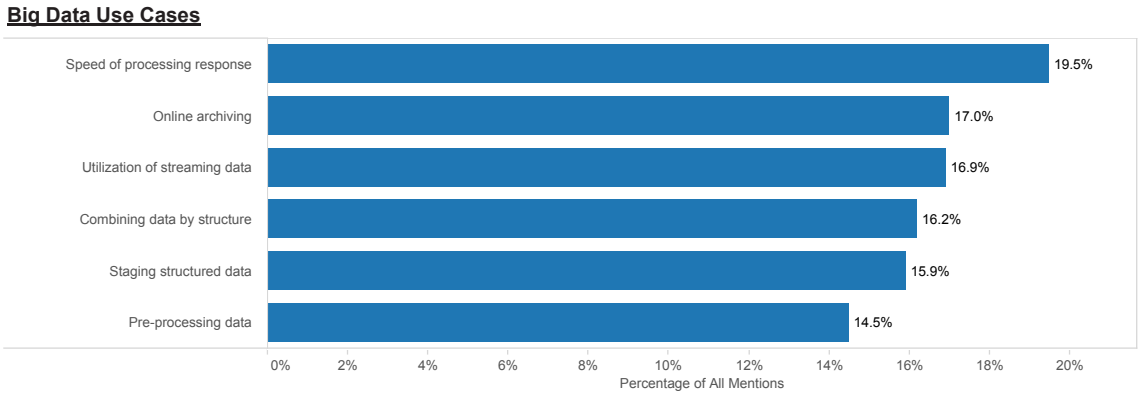


Figure 31

Copyright 2015, EMA Inc. and 9sight Consulting. All Rights Reserved.

The percentage for **Online archiving** increased by two points, while **Combining data by structure** lost 1.3 percentage points. **Utilization of streaming data** gained 2.5 points, thus moving it into a virtual tie for the second most commonly implemented **Use Case**. **Staging structured data** and **Pre-processing data** each fell in popularity by 1.2 points each.

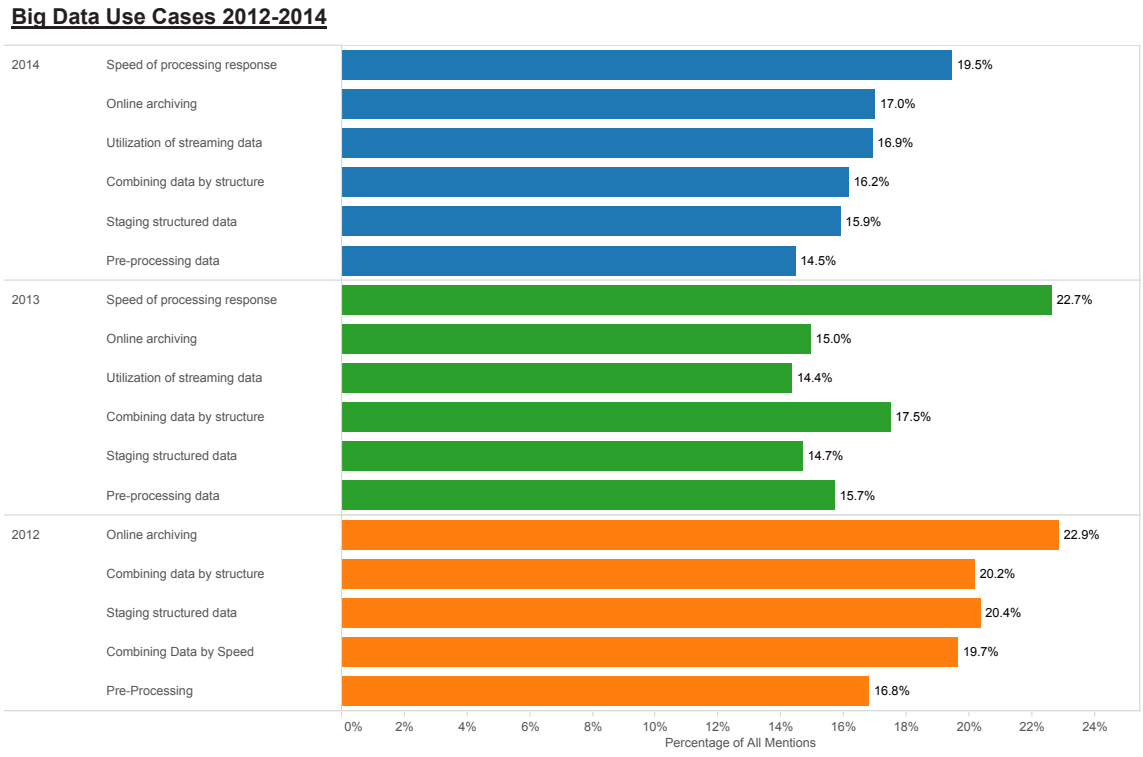


Figure 32

The changes in **Use Case** utilization from 2013 to 2014 demonstrates that **Speed of processing response** and the closely associated **Utilization of streaming data** is still important to maintain competitive advantage via workloads that stress fast reaction time and low latency.

Use Case utilization from 2013 to 2014 demonstrates that Speed of processing response and the closely associated Utilization of streaming data is still important to maintain competitive advantage via workloads that stress fast reaction time and low latency.

Copyright 2015, EMA Inc. and 9sight Consulting. All Rights Reserved.

5.5.1. Breaking Down Industry Cases

The results of the breakdown of **Use Cases** by industry segment show **Speed of processing** to be the highest percentage **Use Case** for the **Manufacturing** and **Finance** sectors. Timely processing of IoT data is critical for manufacturing enterprises, while high-speed processing of the streaming data that is the essence of automated trading is important for organizations in the world of finance.

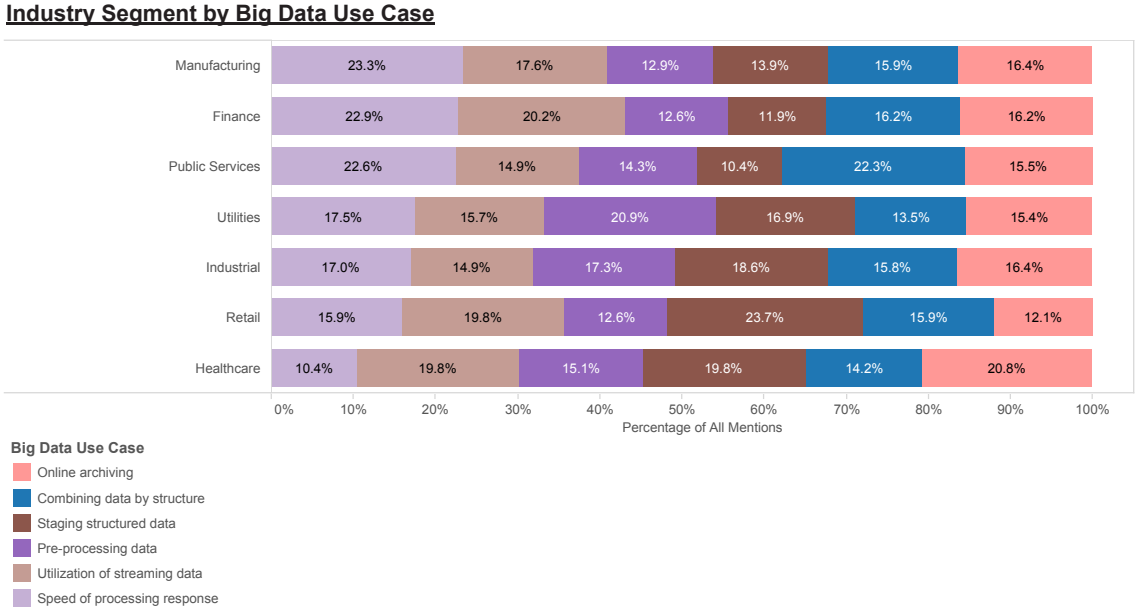


Figure 33

Speed of processing is least important to **Health care** companies, as they are currently focusing on **Online archiving** to store the avalanche of structured and unstructured Electronic Health Records (EHR) data that is now being generated by both humans and operational systems.

5.6. Managing Hurdles

Awareness of the obstacles that a project faces is critical to understanding how to make current and future efforts successful. EMA/9sight panel respondents were asked to identify the issues that they experience.

Answers from the panel respondents are segmented into the following analysis bands.

- **Poor Data Management** – A number of data management issues can negatively impact big data initiatives. Organizations with **Inadequacy of established data governance practices for multi-structured data;** and **Insufficient compliance with operational service level expectations for uptime and support** indicators fall into this category.
- **Strategy Issues** – Misalignments among strategic goals can impact a big data initiative. These indicators include **Absence of defined long-term big data strategy** and an **Unfocused implementation approach strategy**.
- **Stakeholder Issues** – Misalignments among elements of the corporate structure. These include **Lack of stakeholder buy-in for big data project** and **Need for additional communication/ alignment between business and IT stakeholders**.

Copyright 2015, EMA Inc. and 9sight Consulting. All Rights Reserved.

- **Lack of Skills for Multi-structured Data Platform Management –** New data management platforms such as Hadoop and MongoDB can have a steep learning curve for organizations familiar with traditional relational databases.
- **Complexity of Implementation** – Next generation data management platforms may have a lower level of deployment tools and functionality and be more complex to implement.
- **Shortage of Application Management Features** – Next generation data management platforms may need additional features and functions to be on a par with traditional data management platforms.

Respondents called out **Poor Data Management** as their most common project challenge. The prevalence of this challenge is likely because recent increases in the number and complexity of big data projects are stretching the limits of existing policies and developing best practices.

Respondents called out Poor Data Management as their most common project challenge. The prevalence of this challenge is likely because recent increases in the number and complexity of big data projects are stretching the limits of existing policies and developing best practices.

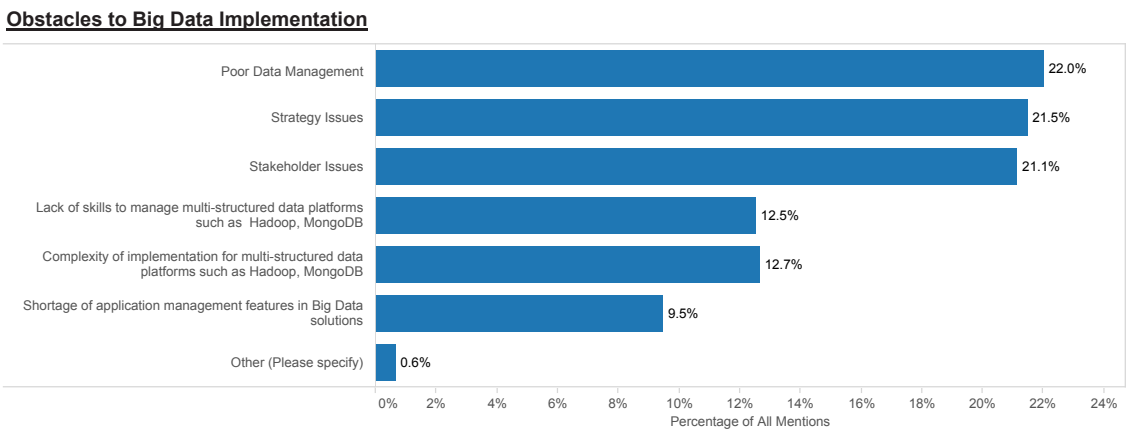


Figure 34

Strategy issues related to aligning on implementation strategy and **Stakeholder issues** of communicating with management and building the case for their approval closely trailed **Poor Data Management** by less than one percentage point. **Strategy** and **Stakeholder** challenges are not unique to big data initiatives, but are pervasive throughout business, especially in the area of IT.

Looking at changes in project hurdles between 2013 and 2014, we see that **Poor Data Management** increased significantly, while **Strategy issues** stayed constant and **Stakeholder issues** decreased. It appears that, with the proliferation of the adoption in big data projects, stakeholders are becoming more comfortable with them. The increase in **Poor Data Management** issues arises from the same root cause—the implementation of more projects and their associated growing pains.

Copyright 2015, EMA Inc. and 9sight Consulting. All Rights Reserved.

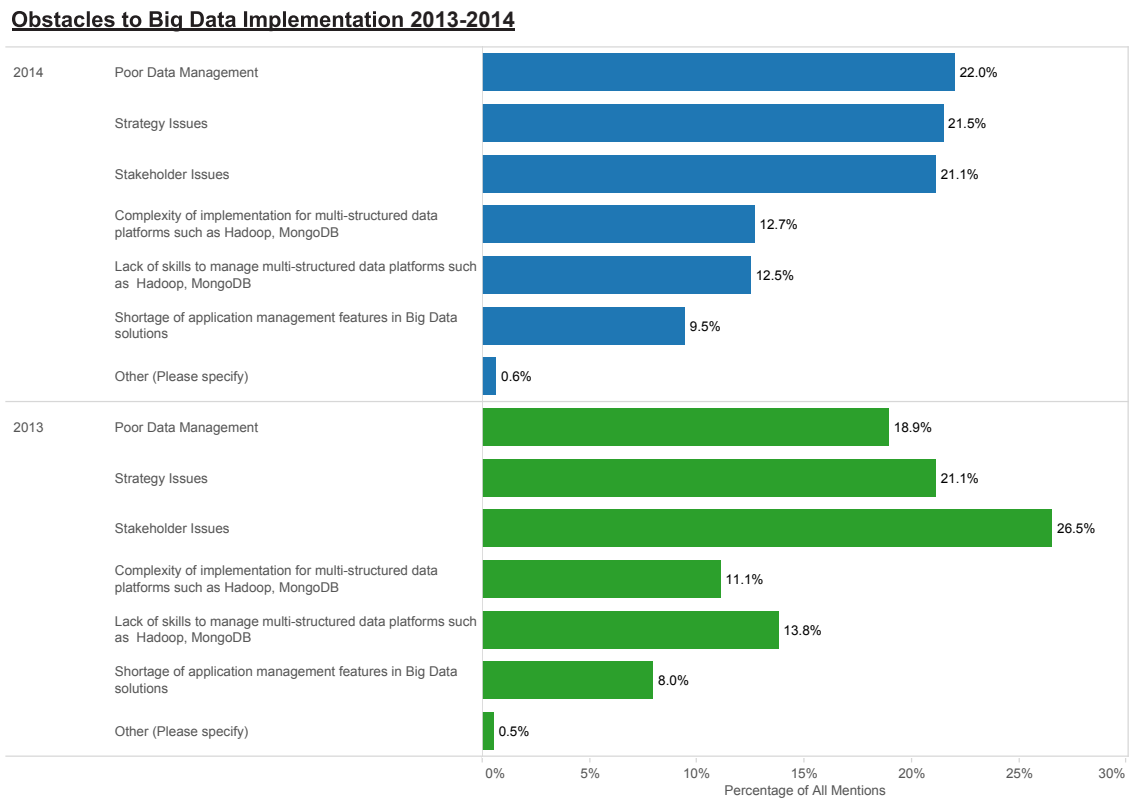


Figure 35

When challenges are broken down by industry segment, **Poor Data Management** issues are concentrated in the **Health Care** segment, followed by **Public Services**.

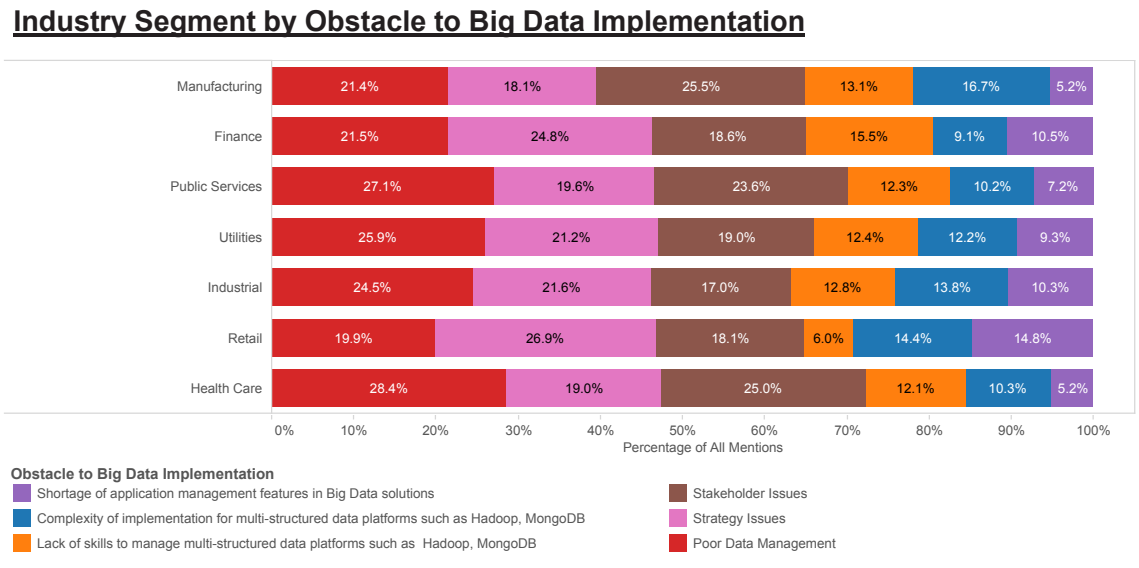


Figure 36

Strategy Issues are most acute in the **Retail** and **Finance** industries, while **Stakeholder Issues** are more critical to **Manufacturing** and **Health Care.**

Copyright 2015, EMA Inc. and 9sight Consulting. All Rights Reserved.

6. Big Data Projects from EMA/9sight Survey Panelists

EMA/9sight asked respondents asked survey respondents how many big data projects their organization is currently working on. In the 2014 EMA/9sight Big Data survey, there is information on 1054 big data projects from the panel respondents.

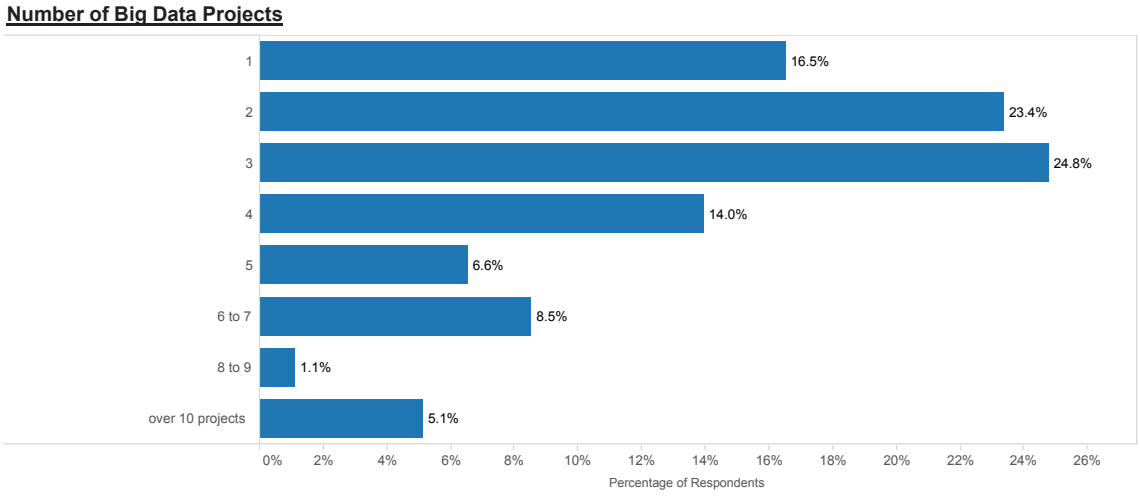


Figure 37

When 2014 and 2013 project numbers are compared, it is evident that the number of projects per organization has grown. Organizations in the **Robust** group have increased by almost 6%.

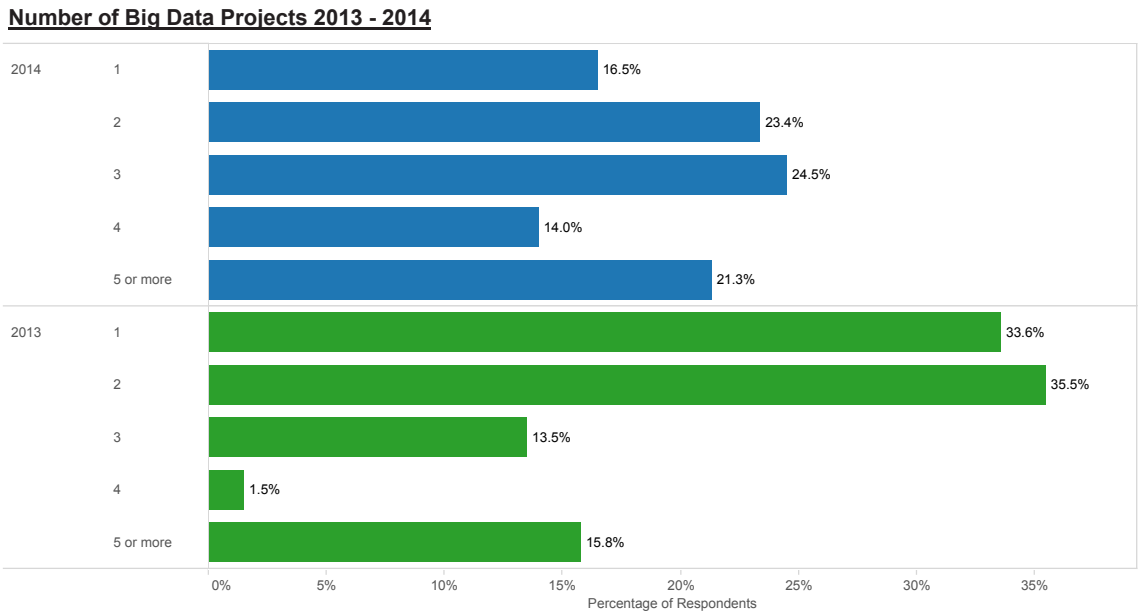


Figure 38

Organizations in the **Maturing** group have increased by almost 24%, while the **Beginning** band has shrunk by over 29%. These trends in number of projects per organizations are among the strongest indications of a rapidly maturing big data market.

Copyright 2015, EMA Inc. and 9sight Consulting. All Rights Reserved.

6.1. Project Maturity

Project stages are classified into three phases: **In Operation**, **Serious Planning,** and **Investigating**. Each phase identifies a different stage of implementation for an active big data project. Of the projects in this year's research, 52% are **In Operation**.

Project stages are classified into three phases: In Operation, Serious Planning, and Investigating. Of the projects in this year's research, 52% are In Operation.

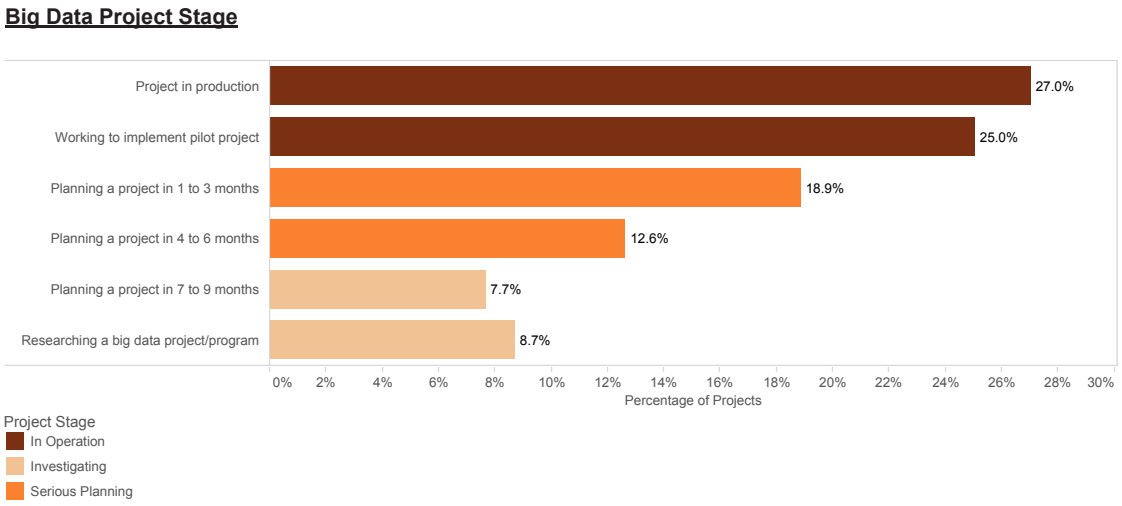


Figure 39

The following groupings are used to define an organization's status in relation to its big data initiatives.

- **In Operation** represents actual implementations of big data projects including **Project in production** and **Working to implement pilot project**. These respondents have hands-on experience with both big data business requirements and technologies that solve those requirements.
- **Serious Planning** is near-term to immediate big data projects. These include survey respondents who indicated **Planning a project in 1 to 6 months**. This group represents organizations that are close to or on the verge of signing contracts for hardware and software licenses associated with their big data implementation.
- **Investigating** signifies those organizations still looking at big data requirements and big data technologies. These respondents are **Planning a project in 7 to 9 months** or are **Researching a big data pilot program**.

Copyright 2015, EMA Inc. and 9sight Consulting. All Rights Reserved.

There was a significant shift in project stages from 2012-2013. It was clear that the 2013 EMA/9sight respondents were further along in their big data efforts than in 2012. Respondents indicated 17% more of their initiatives are **In Operation** over the previous year. In comparison, **Serious Planning** and **Investigating** are both lower in 2013 than 2012's indications. The increase in **In Operation** projects points to a growth in adoption and a maturing of the market.

In 2014, **In Operation** project stages fell slightly, but still within a single percentage point of 2013.

There was a significant shift in project stages from 2012-2013. Respondents indicated 17% more of their initiatives are In Operation over the previous year. In comparison, Serious Planning and Investigating are both lower in 2013 than 2012's indications. The increase in In Operation projects points to a growth in adoption and a maturing of the market.

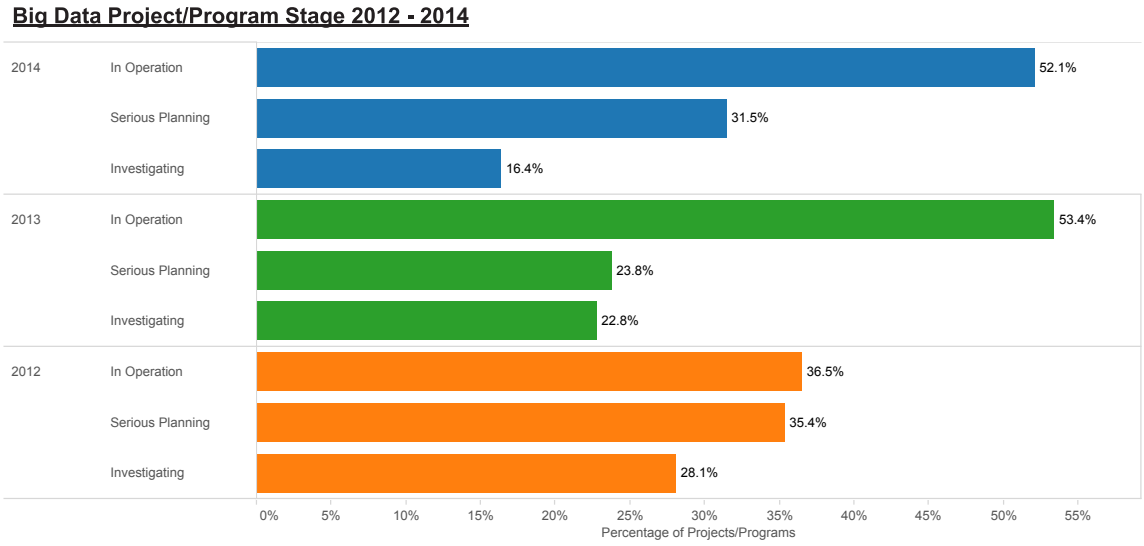


Figure 40

However, **Serious Planning** grew from nearly 24% in 2013 to nearly 32% in 2014 with a corresponding drop in the **Investigation** stage projects. This shows that big data project continue to evolve and mature.

Copyright 2015, EMA Inc. and 9sight Consulting. All Rights Reserved.

6.1.1. Project Implementation by Industries

Most industry segments have identified opportunities to deploy big data projects. In 2012, EMA/9sight survey respondents from **Industrial** and **Finance** industry segments were most likely to have some aspect of their big data initiative **In Operation**. In 2013, **Health care** led all industry segments for projects **In Operation** followed by **Retail**, **Industrial**, and **Manufacturing.** In 2014, **Finance** leads the pack, followed by **Industrial** and **Retail**. **Health care** has dropped back to fourth place, with over 50% of its projects **In Operation**.

In 2013, Health care led all industry segments for projects In Operation followed by Retail, Industrial, and Manufacturing. In 2014, Finance leads the pack, followed by Industrial and Retail. Health care has dropped back to fourth place, with over 50% of its projects In Operation.

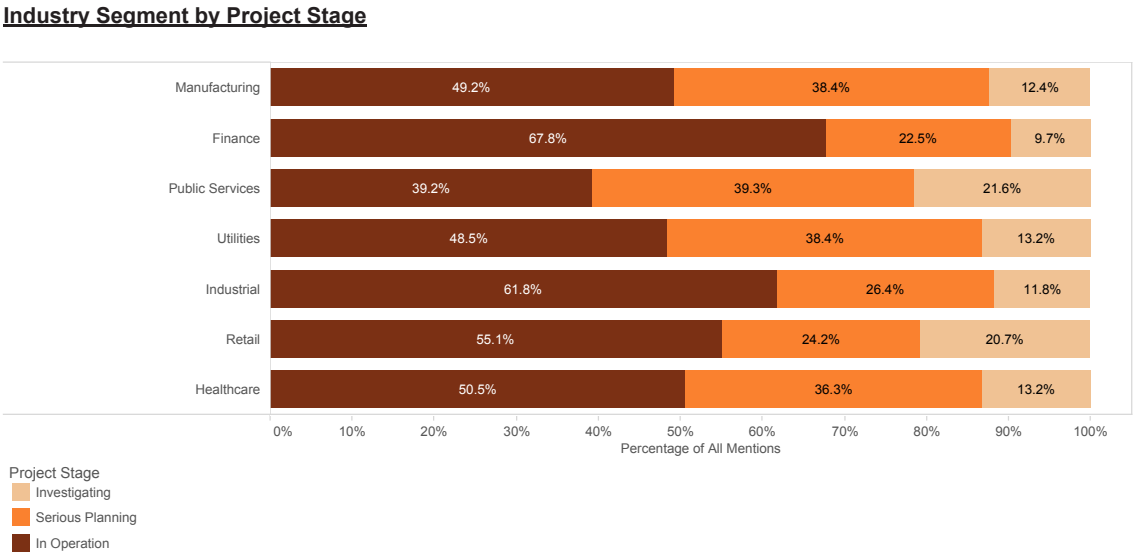


Figure 41

When the categories of **In Operation** and **Serious Planning** are combined together, percentages for **Manufacturing**, **Finance**, **Utilities**, **Industrial,** and **Health care** range between almost 87% to just over 90%, pulling all segments more closely together. The industries that appear most hesitant to implement big data projects are **Public Services** and **Retail**, each with 20-21% of projects in the investigation phase.

Copyright 2015, EMA Inc. and 9sight Consulting. All Rights Reserved.

6.2. Build vs Buy

The EMA/9sight survey asked how organizations were implementing their projects. Were they using templates and configurable applications from external providers such as software vendors or third party consultants? Were they "hand-rolling" their projects using toolsets and manual resources?

Over 20% of respondents indicated that they were implementing big data project using **Customizable applications from external provider** as their implementation strategy.

> Over 20% of respondents indicated that they were implementing big data project using Customizable applications from external provider as their implementation strategy.

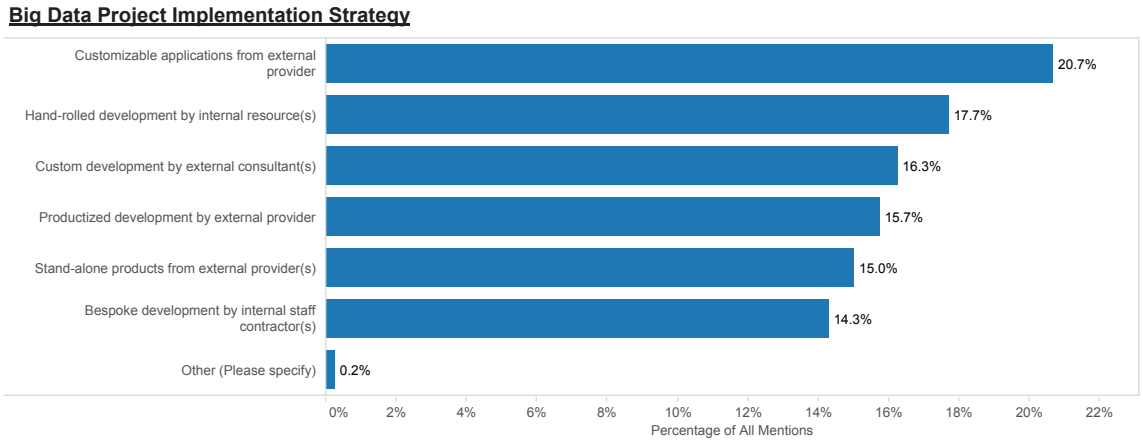


Figure 42

The next two big data project implementation strategies were to utilize some form of manual implementation: **Hand-rolled development by internal resources** and **Custom development by external consultants**.

Answers from the panel respondents are segmented into two analysis bands.

- **Configurable Products** are those project implementation strategies that organizations utilize to improve their time to implementation through the use of pre-packaged software bundles. Organizations that indicated **Customizable application from external provider**; **Productized development by external provider**; or **Stand-alone products from external providers(s)** are included in this analysis band.
- **Manual Development** is leveraging the expertise of developers to handcraft big data project applications. These organizations have self-described their big data project implementation strategy as **Hand-rolled development by internal resource(s)**; **Custom development by external consultant(s)**; or **Bespoke development by internal staff contractor(s)**.

Copyright 2015, EMA Inc. and 9sight Consulting. All Rights Reserved.

Over half of EMA/9sight panel respondents mentioned using **Configurable Products** for big data project implementations.

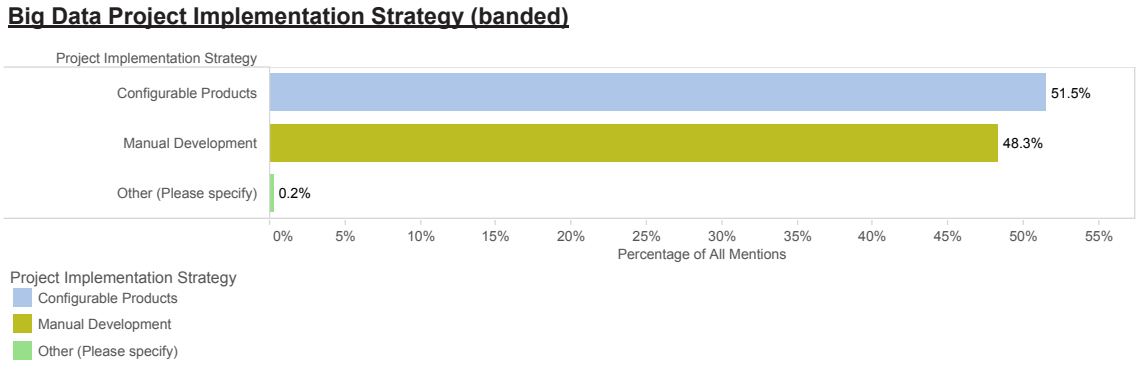


Figure 43

This metric shows significant interest in a faster time to implementation; the maturing field of big data initiatives is moving away from the "classic" data scientist and the manual effort of hand coding, and is instead moving toward a faster time to implementation and re-using components that will improve the productivity of their implementation teams. However, almost half of projects still utilize some form of **Manual Development** to implement their big data initiatives.

6.3. Project Business Challenge

Today's big data projects address a multitude of diverse workloads. Panel respondents divided their responses to the question of the workload that a project addresses among eleven choices, with percentages by workload ranging from almost 16% to approximately 5% of the total, clearly indicating a large variety of projects in the works.

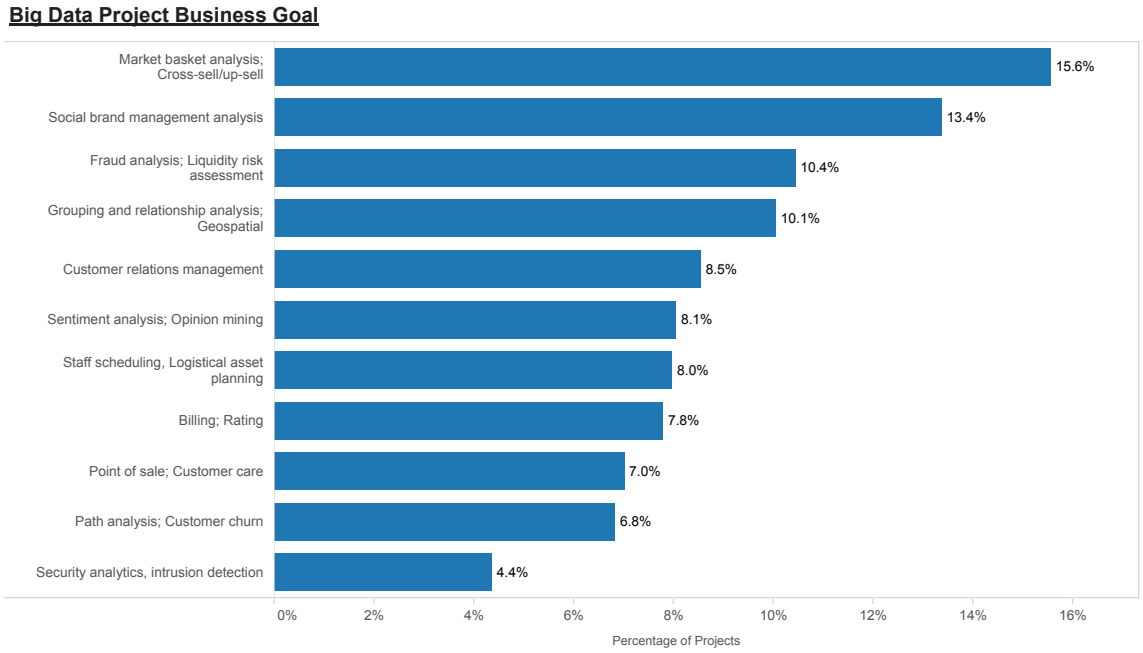


Figure 44

Copyright 2015, EMA Inc. and 9sight Consulting. All Rights Reserved.

The highest percentage of projects tackles **Market basket analysis and Cross-sell/up-sell** activities, targeting improvements to organizations' top-line revenue. Next in popularity is **Social brand management analysis**, which relates to the establishment and maintenance of customer relationships. Ranked third is **Fraud analysis and Liquidity risk assessment**, relating to reducing costs and improving margins. **Security** was a new option for 2014, as it is rapidly growing in importance to organizations that manage big data.

The highest percentage of projects tackles Market basket analysis and Cross-sell/ up-sell activities, targeting improvements to organizations' top-line revenue.

6.3.1. Putting Them Together

As the EMA/9sight survey examined the over 1000 active projects, groupings of business goals were identified.

- **Operational Analytics** includes the growing area in which operational processes are merged with analytical results to improve overall business performance with these workloads: **Campaign Optimization; Market Basket Analysis**, **Cross-sell/Up-sell Recommendation**; **Customer Relations Management**; **Staff Scheduling**, **Logistical Asset Planning**; and **Fraud Analysis and Liquidity Risk Assessment (e.g., risk management)**
- **Operational Processing** includes the beating heart of most organizations: **Billing, Rating (e.g., operational event and policy processing)**, **Point of Sale**, and **Customer Care (e.g., operational transaction processing)**
- **Relationship Analysis** is one of the key areas of big data projects is the ability to take new data sources and include them into the task workloads. **Grouping and Relationship Analysis**, **Geographic Optimization** and **Path Analysis**, and **Customer Churn** are examples of these business goals
- **Social Brand and Sentiment Analysis** represents one of the core functions of big data has been the ability to ingest and process information coming from social and sentiment data sources associated with social media and human-sourced information. **Sentiment Analysis**; **Opinion Mining**; and **Social Brand Management Analysis** are in this category
- **Security Analytics and Intrusion Detection** project goals are in this category

Operational Analytics leads all other categories with almost 43% of respondents stating they were executing projects around this workload. Organizations see the value of integrating analytics into operational processes to minimize total cost of ownership (TCO) while maximizing revenues.

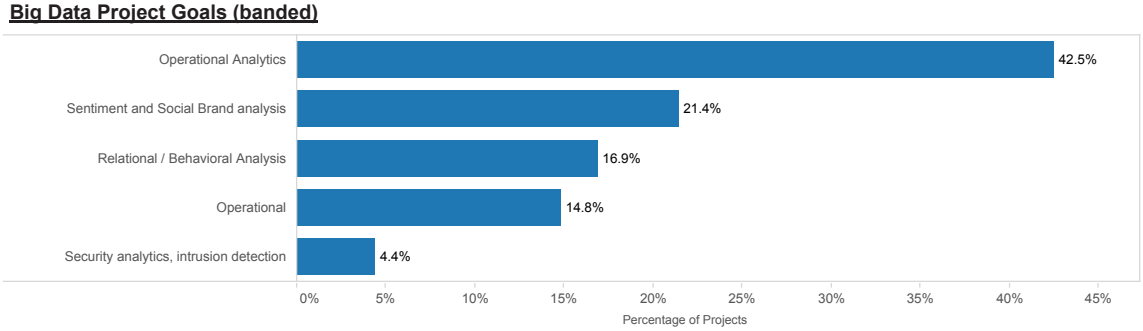


Figure 45

Copyright 2015, EMA Inc. and 9sight Consulting. All Rights Reserved.

When business goals are compared by industry segment, it is clear that the **Retail** sector leads all others in the area of **Operational Analytics**, demonstrating the importance that retailers place on **Fraud Analysis**, **Campaign Optimization**, **Market Basket Analysis,** and **Cross-sell/Up-sell Recommendation**. The **Manufacturing** segment comes in second, underscoring its attention to manufacturing quality.

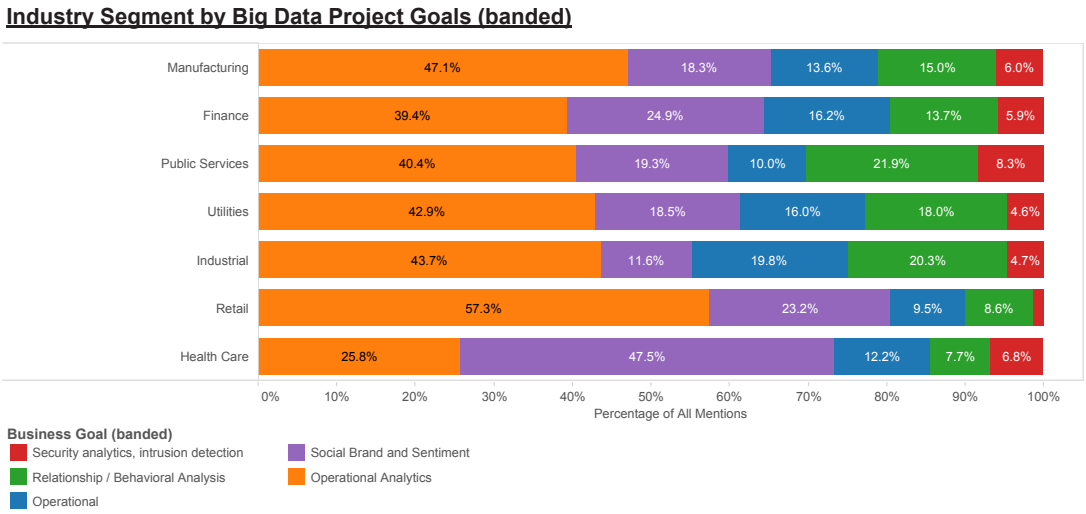


Figure 46

In the area of Security, **Public Services** and **Health care** lead the other segments, emphasizing needs to protect users' identities their applications due to regulatory and consumer demands.

6.4. Data Consumers

It is critical to understand who consumes big data projects. For the most part, users of these complex and expensive projects are inside the company with **Line of Business executives**—the group charged with keeping their finger on the pulse their companies—being the largest groups of consumers. **Marketing and Financial Analysts** are the second largest group of users—a group with the skills to dive deep into the outcomes of big data initiatives.

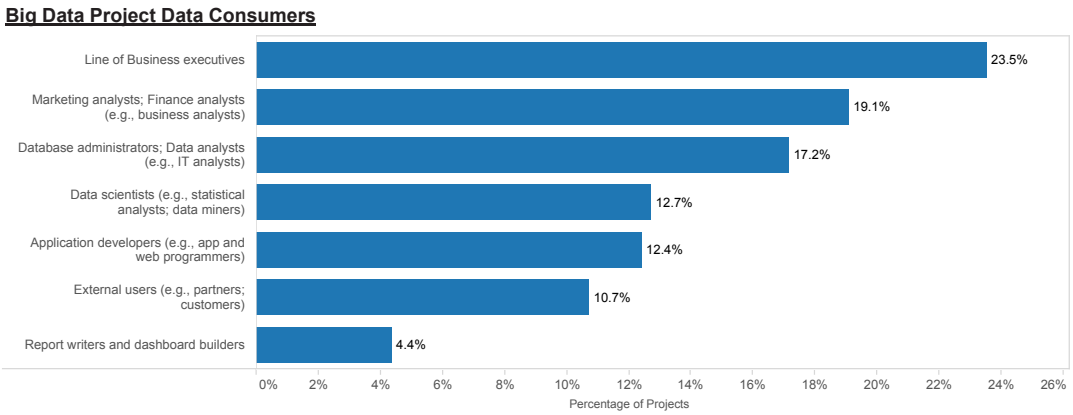


Figure 47

Database administrators and Data analysts come in third, using over 17% of big data projects, with their interests lying in the more technical aspects of projects.

Copyright 2015, EMA Inc. and 9sight Consulting. All Rights Reserved.

When 2014 big data project consumers are compared with those from the 2014 study, fewer projects were directed at **Line of Business** executives, **Marketing and Finance Analysts,** and **Database administrators and Data analysts**.

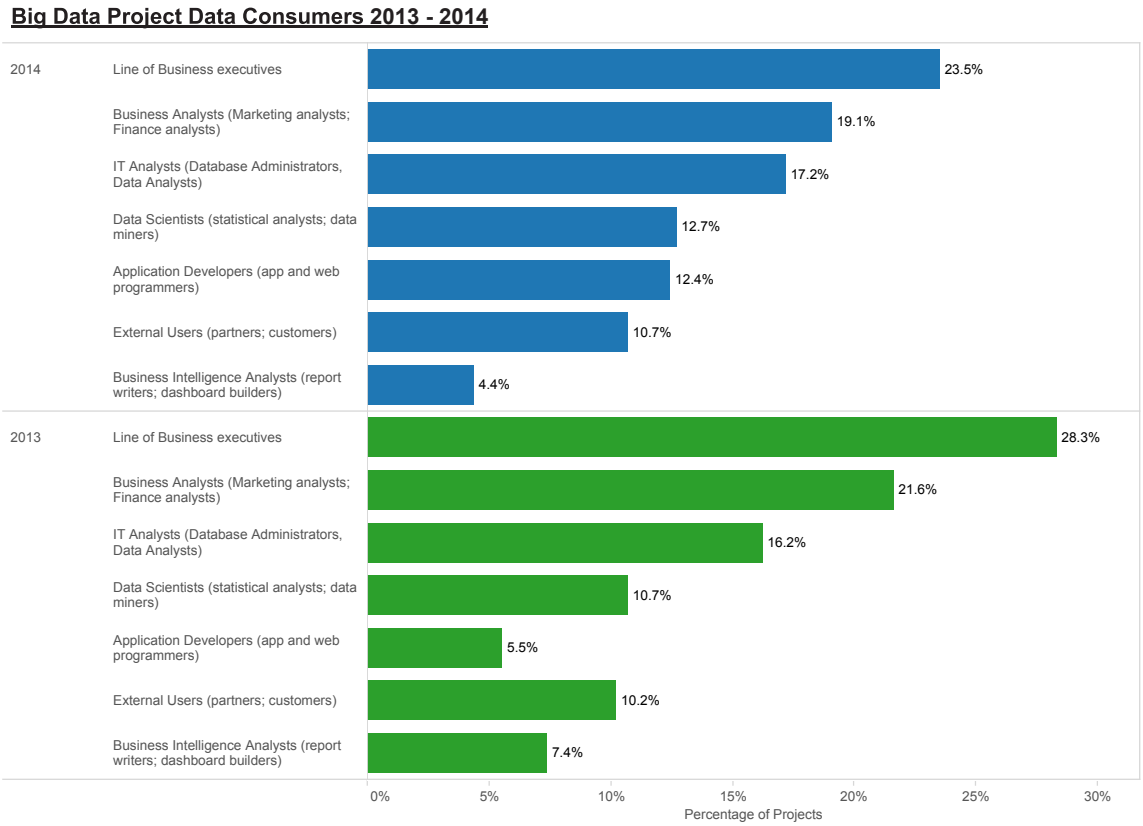


Figure 48

Projects targeted towards development groups, such as **Report Writers and Dashboard Developers** and **Application Developers**, increased.

Looking at various company sizes in association with their Data Consumers reveals that **Application developers** and **IT Analysts** are the top two consumers of projects for **Enterprise** sized organizations. For **Large** companies, **External Users** and BI Analysts such as **Report writers and dashboard builders** are the two largest users of big data projects.

Copyright 2015, EMA Inc. and 9sight Consulting. All Rights Reserved.

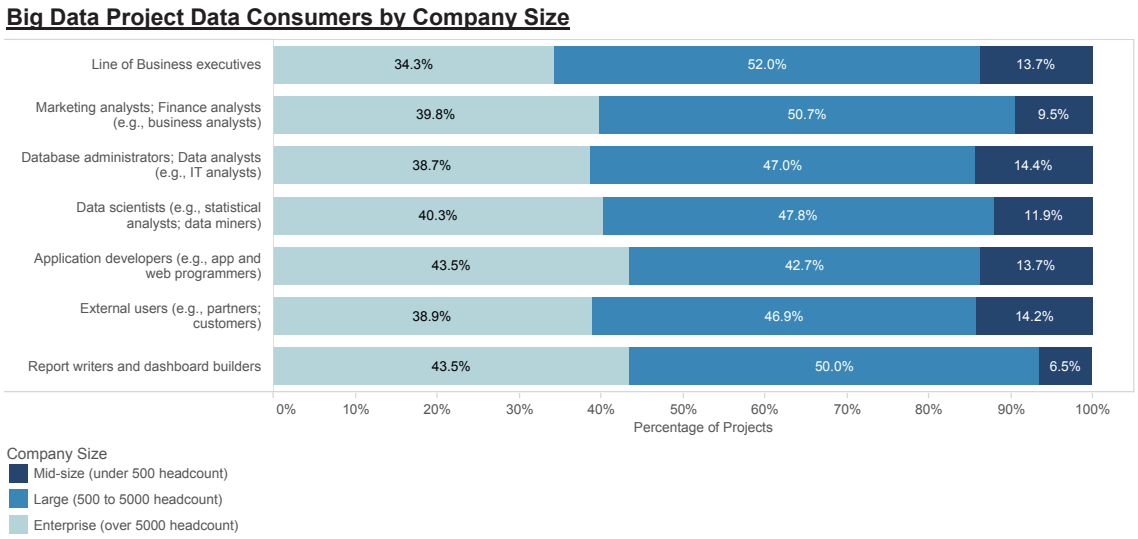


Figure 49

For **Mid-size** organizations, the top indications for **Data consumers** of big data projects are a disparate collection of business users and technologists—**Line of business executives**, IT Analysts such as **Data analysts**, and **Application developers** such as web developers—all between 11 and 12%.

6.4.1. Different Industries, Different Users

EMA/9sight analyzed information consumers of big data projects by industry segment. The largest percentage of **External users (partners and consumers)** was in the **Health care** sector, leading **Public Services**, the next closest sector, by a factor of two.

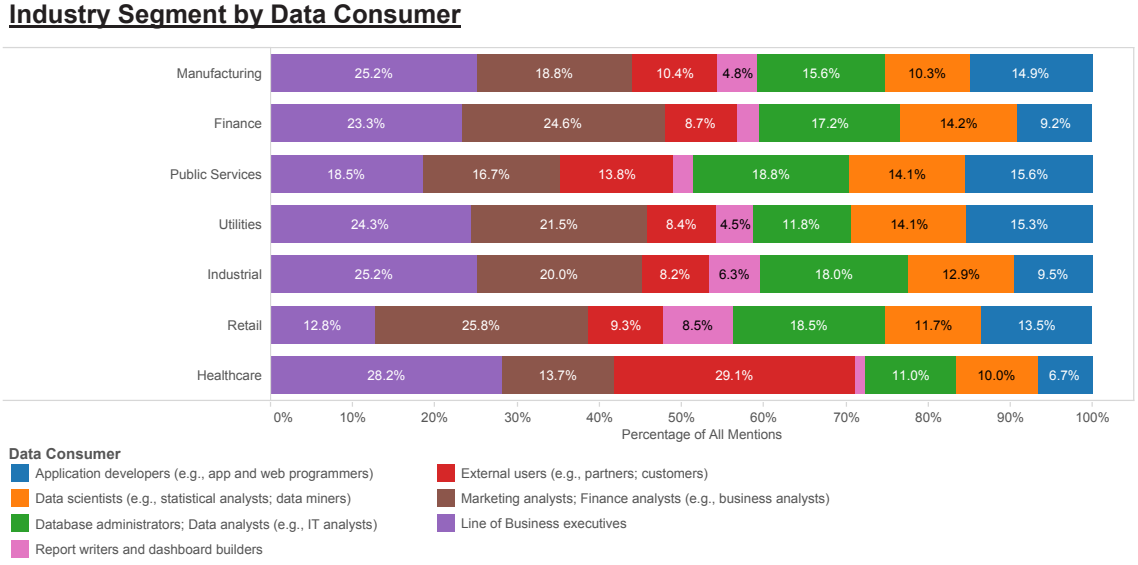


Figure 50

It is also notable that aside from **Retail** and **Public Services**, **Line of Business executives** are targeted in about one quarter of the projects of the industry segments.

Copyright 2015, EMA Inc. and 9sight Consulting. All Rights Reserved.

6.5. Building Big Data User Skills

Deciding how to accomplish the work required to deliver a successful big data project can be challenging for an organization. Big data projects often require new, multiple platforms that require skills that current staff may not have. In 2014, many organizations looked outside their companies to get their work done. Over 20% of respondents **Utilize technical consulting services from external organizations**, which provides a significant opportunity for consulting groups.

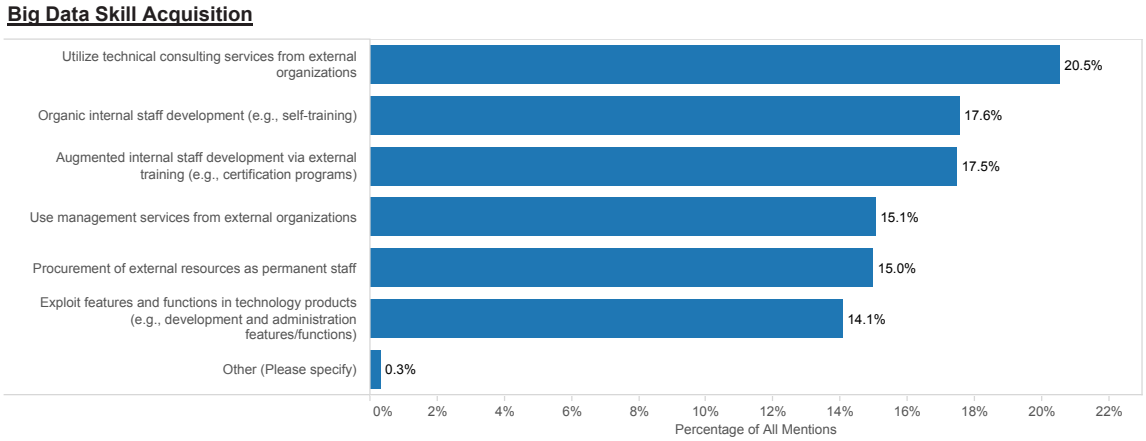


Figure 51

The next two most common responses looked to develop internal resources. **Organic internal staff development (self-training)** and **Augmented internal staff development via external training (certification programs)** each garnered over 17% of responses.

A comparison of the results of 2013 and 2014 demonstrated that the **Utilization of technical consulting services from external organizations** increased by over 2 percentage points between 2013 and 2014, while **self-training** decreased by over 4 percentage points.

Copyright 2015, EMA Inc. and 9sight Consulting. All Rights Reserved.

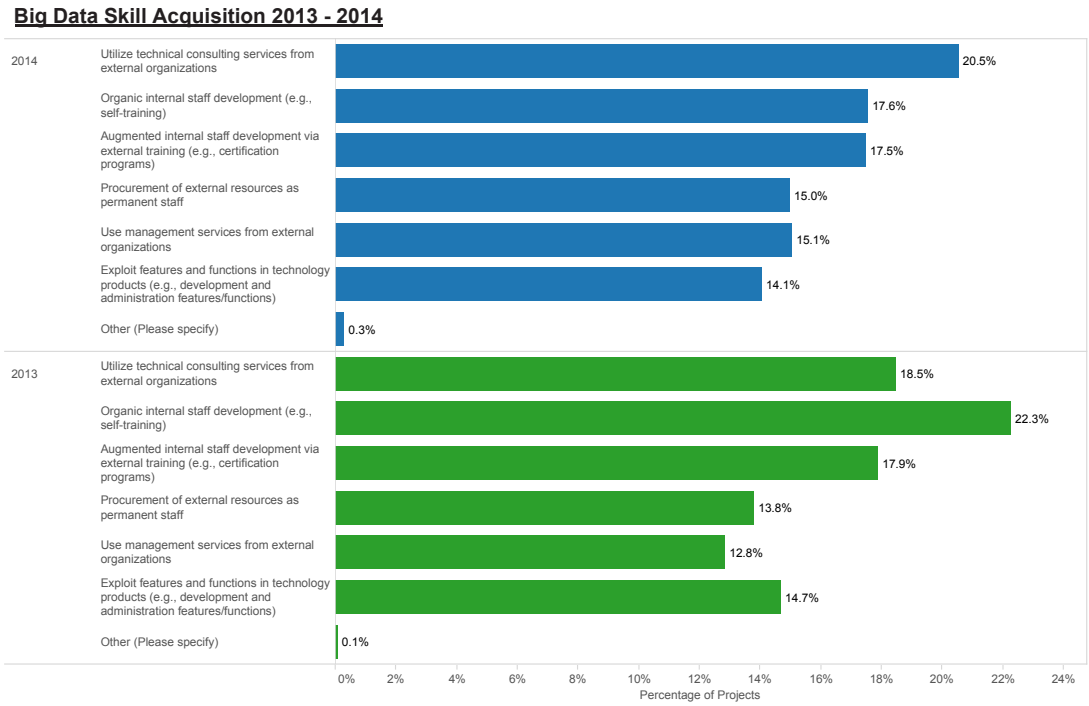


Figure 52

In another nod to external sourcing, the category **Use management services from external organizations** increased by over 2 percentage points in 2014.

When comparisons by industry segment are made, the largest user of **Technical consulting services from external organizations** is **Public Services**, with its critical need to deliver high-performing systems on a tight timeline.

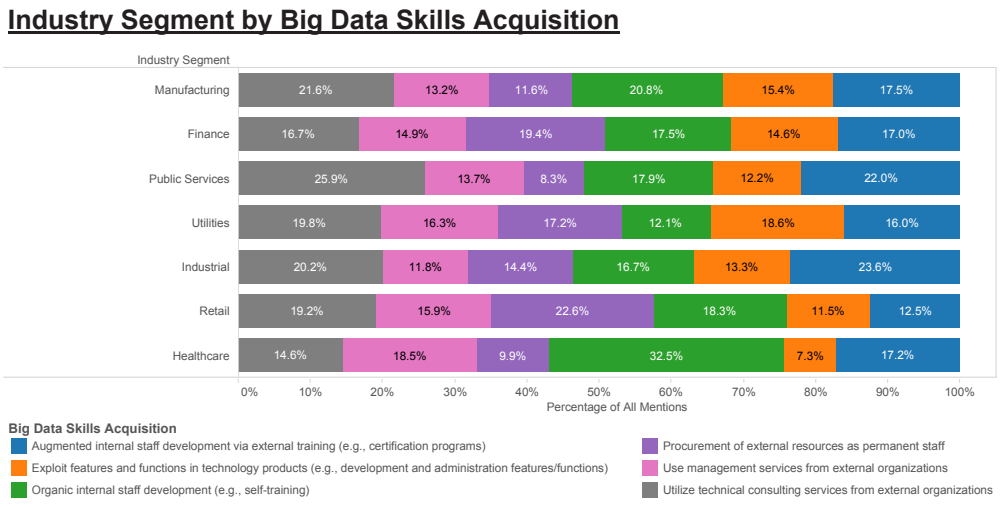


Figure 53

Health care is more inclined than other sectors to **Use management services from external organizations**, but leads the group in **Organic internal staff development (self-training).**

Copyright 2015, EMA Inc. and 9sight Consulting. All Rights Reserved.

6.6. Project Sponsors

All technology projects require sponsorship both in terms of IT support and budget funding, and this is no different in the area of big data. Big data projects are not trivial and many projects require the acquisition of new hardware and software infrastructure. Panel respondents rank **Information Technology/Data Center** as the sponsor for the largest percentage of projects.

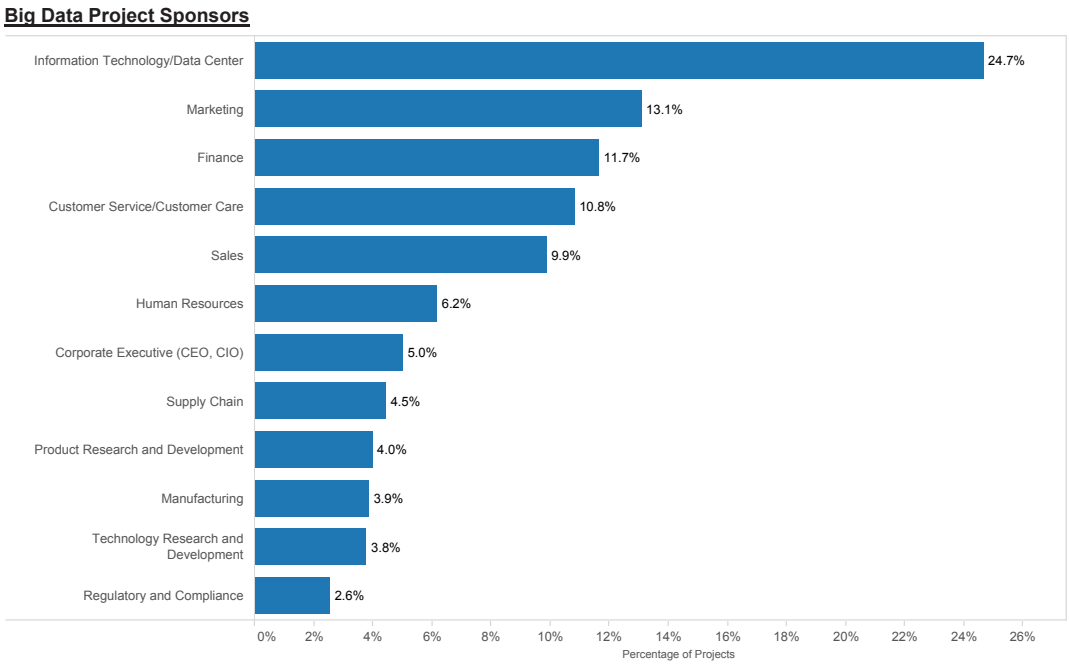


Figure 54

However, the next four most frequent sponsors, **Marketing**, **Finance**, **Customer Service/Customer Care,** and **Sales** are all similar, are **Line of Business** sponsors outside the office of the CEO/CIO that, taken together, account for over 45% of project sponsorships. This finding demonstrates the continuing maturation of the market, with diverse components of the organization embracing big data to solve critical business challenges.

Further demonstration of the maturation of the big data landscape comes from the comparison of data between 2013 and 2014, in which sponsorship from **Corporate Executives (CEO, CIO)** dropped by 3 percentage points.

Copyright 2015, EMA Inc. and 9sight Consulting. All Rights Reserved.

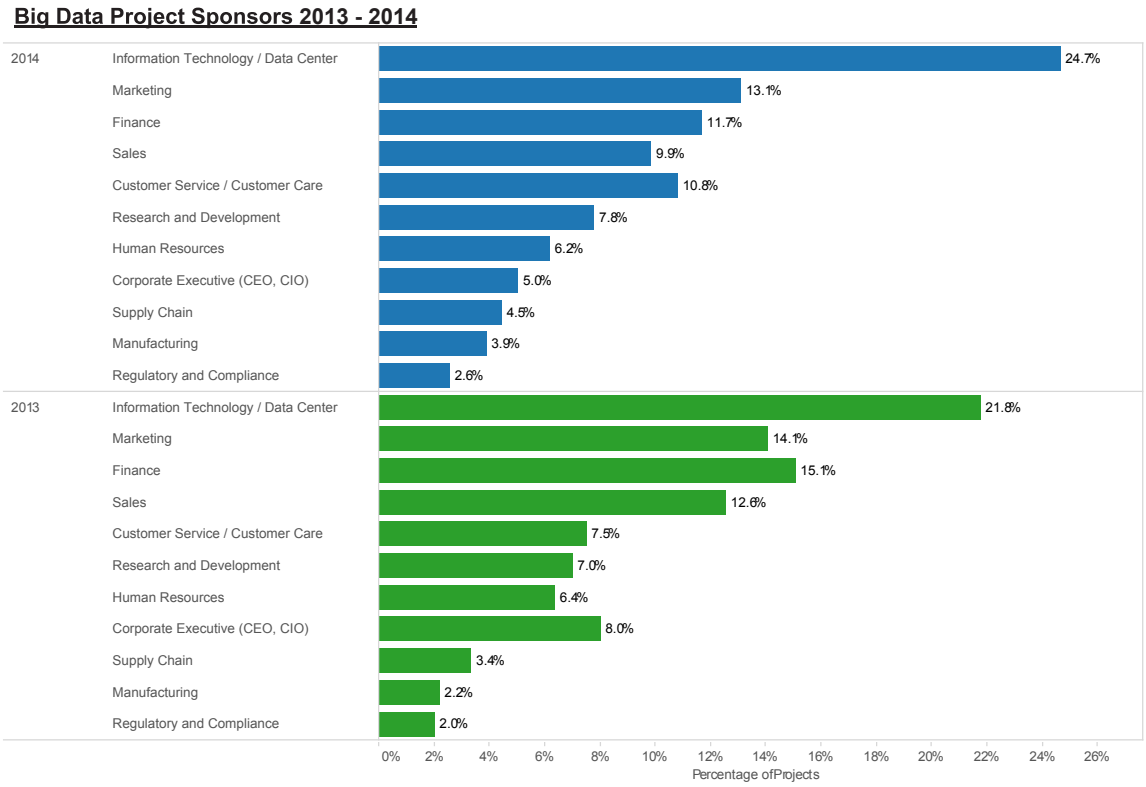


Figure 55

Sponsorship by **IT** and **Lines of Business** also dropped slightly, while **Customer Service/Customer Care** and **Manufacturing** sponsorship increased.

When project sponsorship is compared by the amount of external funding provided to big data initiatives, you can see that **Marketing** department sponsors contribute at both the top end of over 40% of a big data initiative and at lower ends between 1% and 15%. These numbers show a diverse set of maturity for **Marketing** project sponsorship. Some organizations are significantly funding their big data initiatives from the CMO's office, which shows the power of the CMO in driving some big data programs. Other **Marketing** departments are investing less in big data initiatives, but are still making those investments.

These numbers show a diverse set of maturity for Marketing project sponsorship. Some organizations are significantly funding their big data initiatives from the CMO's office, which shows the power of the CMO in driving some big data programs.

Copyright 2015, EMA Inc. and 9sight Consulting. All Rights Reserved.

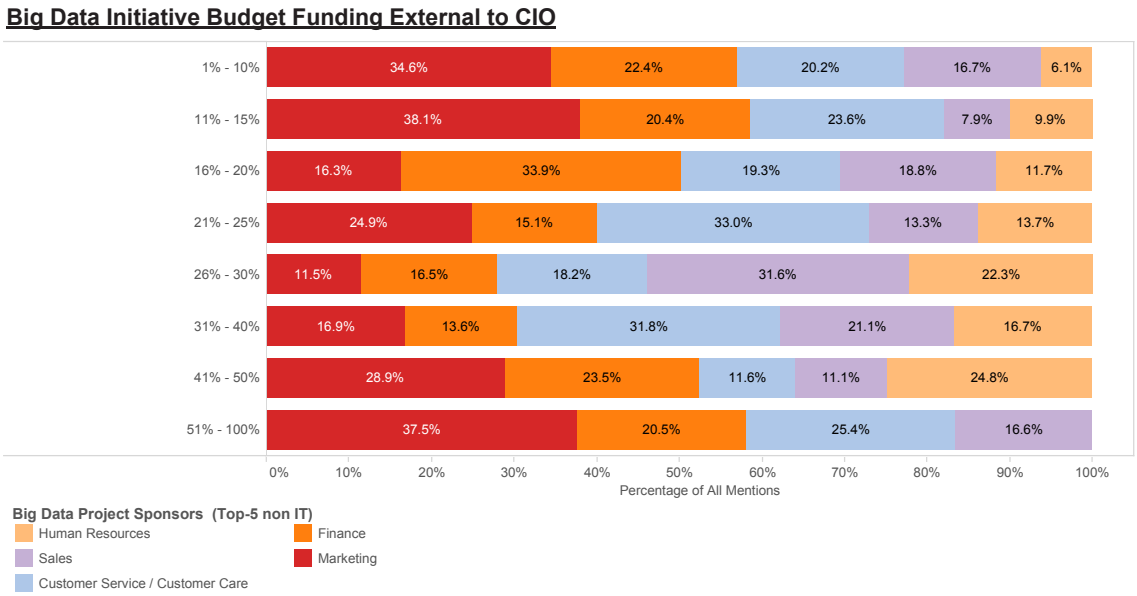


Figure 56

Finance's contributions to big data initiatives are focused in the 16-20% range of budget contribution. This finding shows that the overall big data budgets of these EMA/9sight panel respondents are still controlled and maintained by the CIO's office, but are receiving both support and guidance from the CFO on the direction of those projects.

6.7. Impact of Cloud on Implementation Avenues

When asked what the implementation strategy their projects used, the EMA/9sight panel respondents overwhelmingly favored on-premises implementations. In total, over 41% of projects used an **On-premises data center implementation**, while just over 25% used an **On-premises private cloud implementation**. When the data is split a different way, a total of almost 52% of implementations were in some flavor of cloud: **On-premises private cloud implementation**, **Hybrid on-premises private cloud with public cloud**, or **External public cloud**.

Copyright 2015, EMA Inc. and 9sight Consulting. All Rights Reserved.

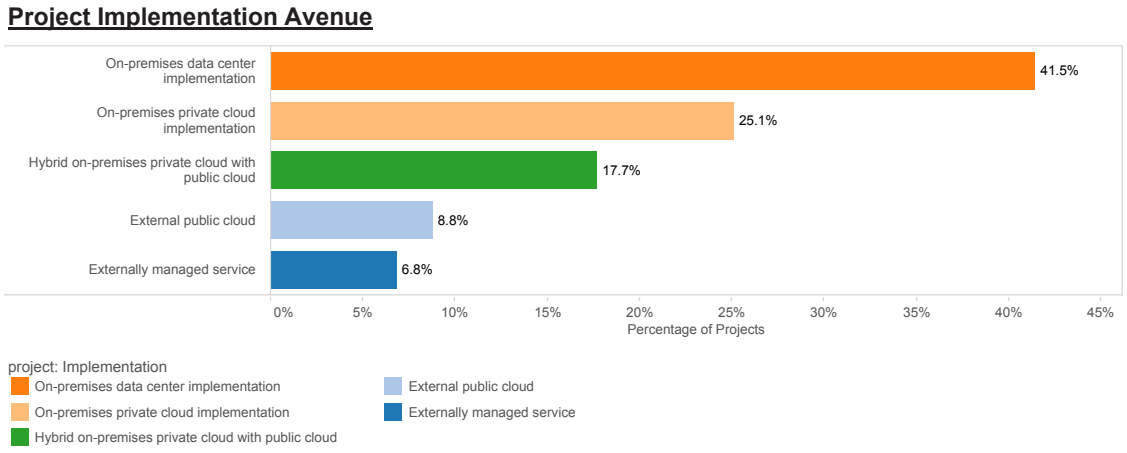


Figure 57

Public cloud-based infrastructures are becoming increasingly more popular, as they can meet modern organizations' needs for reporting and processing agility along with fast response time with minimal drain on a company's resources. These data management platforms can support the growth of data without placing an excessive capital expenditure burden such as license investment or data center capacity in the near term to accommodate potential (and possibly unattainable) future growth.

6.8. Workload

For the first time in 2014, survey respondents were asked to share the types of workloads addressed by their projects. The most common workload was **Operational processing (e.g., billing and rating)** with over 16% of projects reported with that workload. This is not unexpected, as this processing is the core of IT for many businesses.

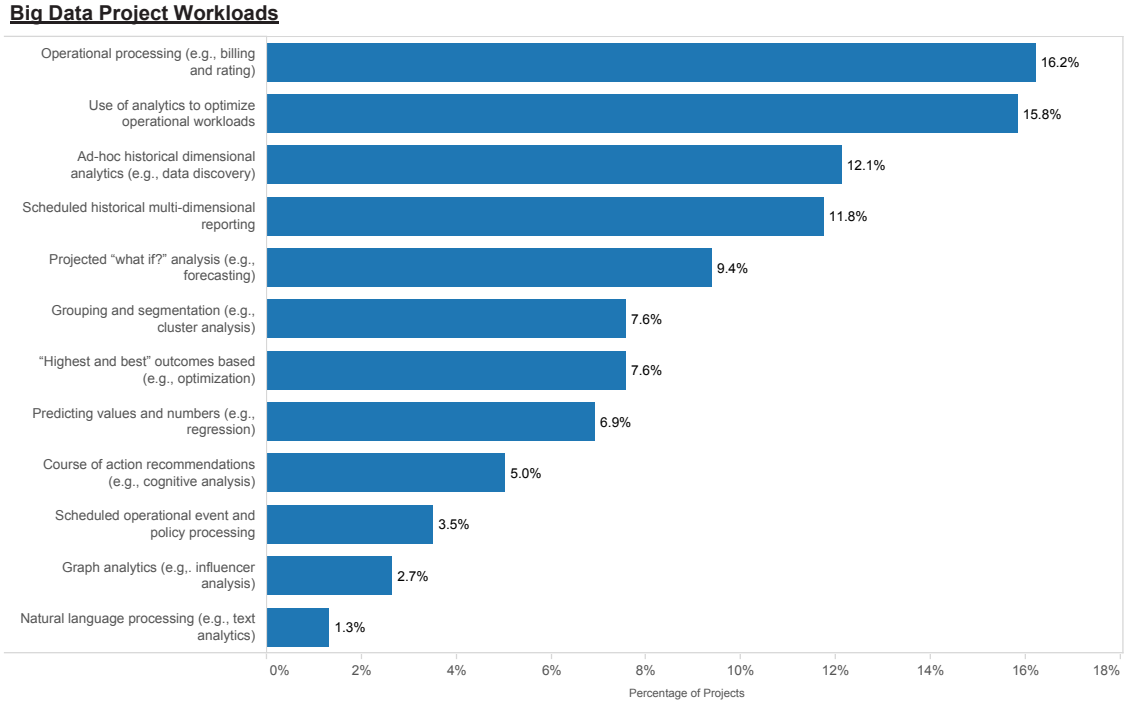


Figure 58

Copyright 2015, EMA Inc. and 9sight Consulting. All Rights Reserved.

However, **Use of analytics to optimize operational workloads** ranked in second place, coming in only .4 point behind **Operational processing**, again confirming that a large number of businesses find benefits in getting a better handle on their operational workloads. **Ad-hoc historical dimensional analysis (e.g., data discovery)** was the third most frequent workload, addressed by just over 12% of projects.

EMA/9sight also examined the relationships between companies with varying levels of data-driven strategy compared with the workloads that their projects tackle. The highest-ranking workload for **Data-driven** companies' projects is **Operational processing (e.g., billing and rating)**, which impacts revenue generation, **Ad-hoc historical dimensional analytics,** which helps discovery of new data, and **Natural language processing,** which provides advanced text analytics.

The highest-ranking workload for Data-driven companies' projects is Operational processing (e.g., billing and rating), which impacts revenue generation, Ad-hoc historical dimensional analytics, which helps discovery of new data, and Natural language processing, which provides advanced text analytics.

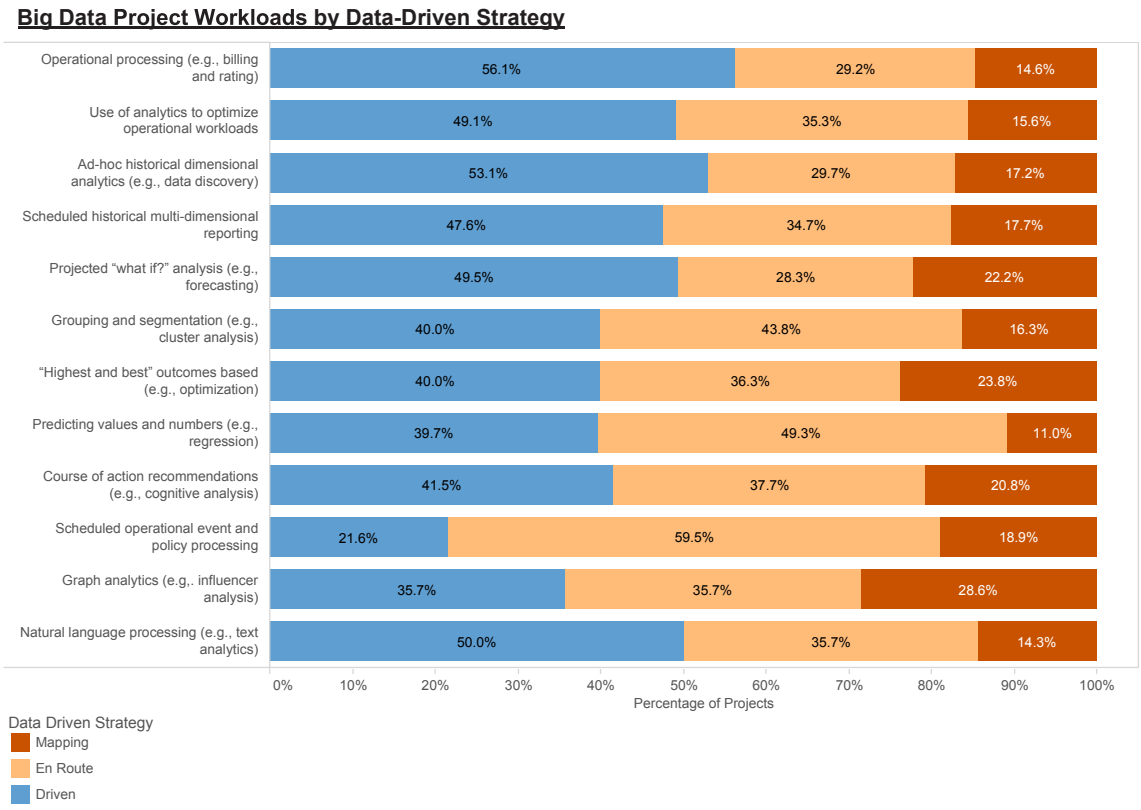


Figure 59

The three most frequent workloads for **En Route** companies are **Scheduled operational event and policy processing**, **Predicting values and numbers (e.g., regression)**, and **Grouping and segmentation (e.g., cluster analysis)**. These workloads demonstrate that **En Route** organizations are utilizing advanced analytical workloads as well as policy driven operational workloads. Both workloads are consistent with an organization that is still getting a handle on a **Data-driven Strategy** before embarking on more complex workloads.

Copyright 2015, EMA Inc. and 9sight Consulting. All Rights Reserved.

6.9. Latency

Panel respondents also disclosed the latency of each of their projects. Big data projects are overwhelmingly **Near-real time**, with nearly 33% described as **Real time/near-real time processing of data** and another 23% termed **Intra-hour processing of data**. Clearly, companies require fast response time from their big data projects to keep their organizations running smoothly.

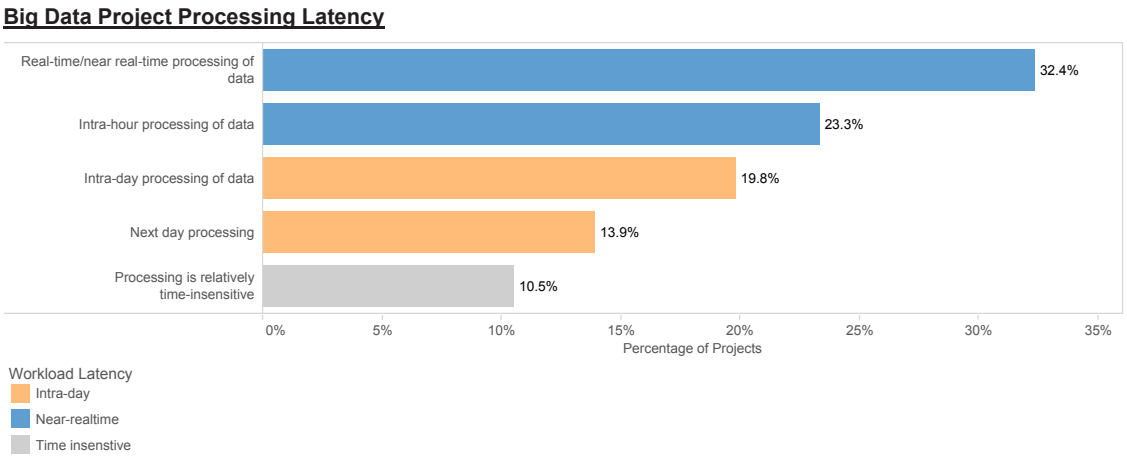


Figure 60

Just under 34% of projects are listed as **Intra-day**—either **intra-day** or **next day processing**. Only 11% of projects are **Time-insensitive**.

When a company's data-driven strategy is taken into account, it is evident that **Driven** organizations focus on lower latency for faster responses for their mission-critical projects.

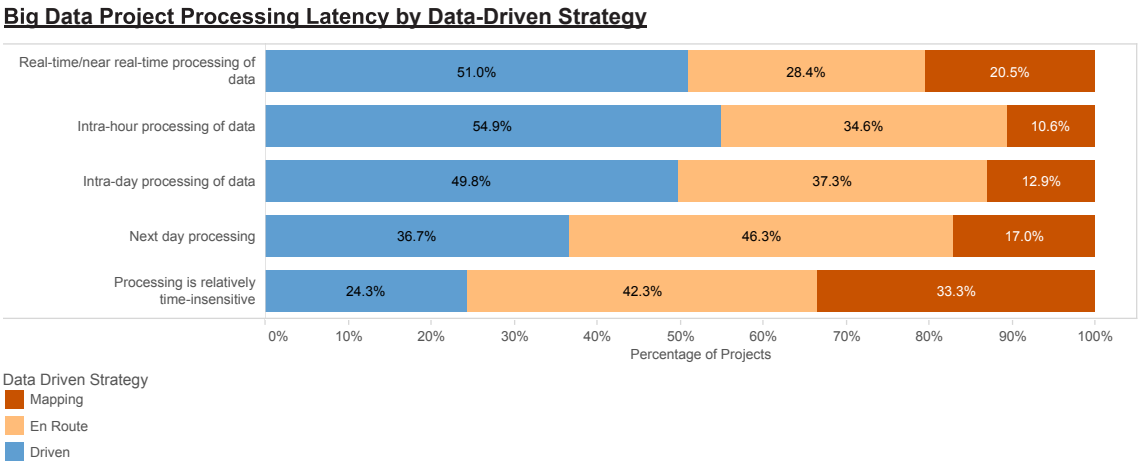
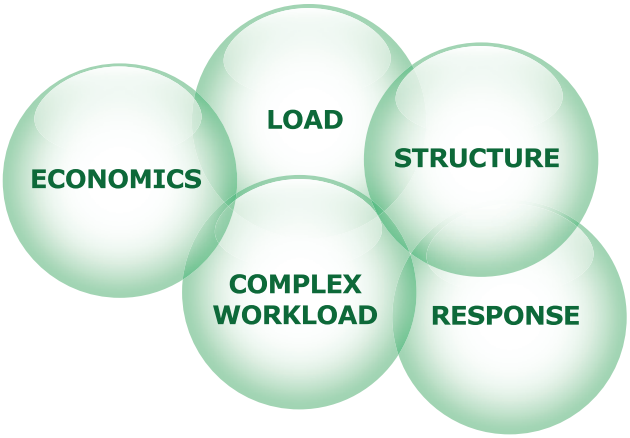


Figure 61

En Route groups have implemented projects in which response time is not as critical as those of data-driven companies have. In general, the projects of **Mapping** organizations are least time-sensitive.

Copyright 2015, EMA Inc. and 9sight Consulting. All Rights Reserved.

7. What is Driving Big Data? Hybrid Data Ecosystem?



With confirmation via the EMA/9sight Big Data research since 2012, EMA has identified five core requirements of big data initiatives and the implementation of an organization's Hybrid Data Ecosystem.

- **Response** – New technology platforms such as big data tools and frameworks are at the core of this evolution, powering new solutions and improved speed of results.
- **Economics** – Big data platforms leverage commodity hardware and the software is often free, substantially reducing the financial barriers to adoption.
- **Workload** – Big data platforms play a role within the ecosystem to execute extremely complex analytic workloads; innovative companies are willing to invest early in these solutions to gain competitive advantage.
- **Load** – Data loads are growing more complex and the sources are more diverse. Driven by greater complexity and demand, big data adoption is driven by the need to provide flexibility.
- **Structure** – Data structure and schema flexibility are key to the foundation of big data utilization and adoption.

7.1. The Need for Speed

The business requirement for faster processing response is core to the drivers of big data initiatives. It is not sufficient to provide processing in big data environments, or next generation data management environments such as an EMA **HDE**, in a batch processing time frame. This concept supports the premise that big data initiatives require data management platforms that go beyond the latency often associated with Hadoop's MapReduce and YARN frameworks. The result sets, whether they are operational or analytical in nature, need to show improvement over the speeds associated with existing or traditional data management platforms. The expectation is that, regardless of the processing being requested, the results will be faster than those in previous eras of computing.

7.1.1. Building the Use Case for Speed

As in prior years of the EMA/9sight survey, panel respondents were asked which use cases were applicable to their big data initiatives. Choices were defined as the following.

- **Online Archiving** – Active online storage of relational (e.g., customer care) or multi-structured (e.g., social media) data for capacity, data access, or economic reasons.
- **Pre-processing Data** – Utilizing a big data solution to search through and process high volumes of relational or multi-structured data to identify a subset of information that can be analyzed further.
- **Speed of Processing Response** – Utilizing a big data solution to process relational or multi-structured data with near-real time (low latency) service level agreements for response (e.g., online applications).

Copyright 2015, EMA Inc. and 9sight Consulting. All Rights Reserved.

- **Combining Data by Structure** – Utilizing a big data solution to combine highly structured relational data (e.g., customer relationship management) with multi-structured (e.g., social media) or unstructured (e.g., video; audio) data.
- **Staging Structured Data** – Utilizing a big data solution to analyze high volumes of traditional, structured relational data (e.g., point of sale; enterprise resource planning).
- **Utilization of Streaming Data** – Utilizing a big data solution to combine relational or multi-structured data at rest with real-time (low latency) event processing data.

In this year's study, when EMA/9sight examined the use cases of the respondents above, the **Speed of processing response** was the top choice among respondents.

In this year's study, the Speed of processing response was the top choice among respondents.

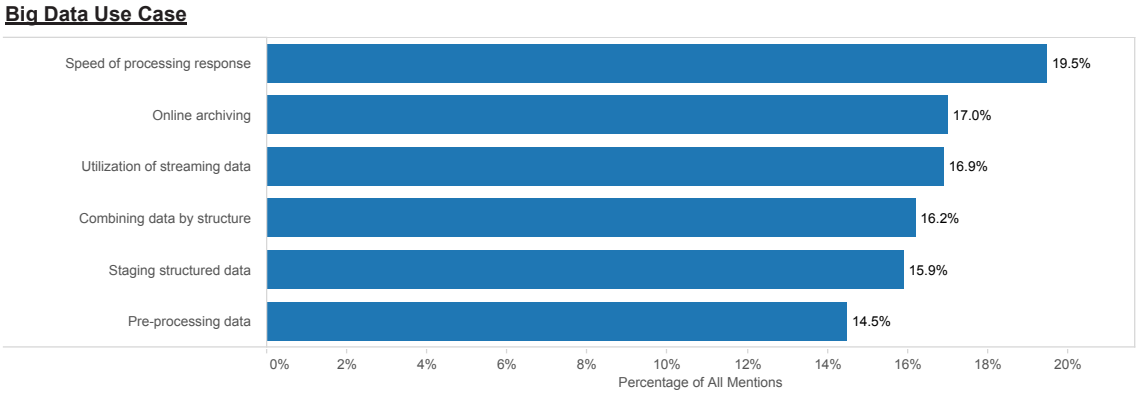


Figure 62

The second most common response was **Online archiving** for those organizations starting their big data initiatives and getting a handle on their new data strategies. The third most prevalent option was the **Utilization of streaming data**. With the importance of the Internet of Things in this year's study, it is not surprising to see organizations focusing on this use case.

The second most common response was Online archiving for those organizations starting their big data initiatives and getting a handle on their new data strategies.

Copyright 2015, EMA Inc. and 9sight Consulting. All Rights Reserved.

7.1.2. Trends in Use Cases

Over the past three editions of the EMA/9sight Big Data survey, panel respondents have been presented with similar use case options. As with last year's survey, the **Speed of processing response** is the top use case. However, it should be noted that the prevalence of indications for **Speed of processing response** declined from 2013 to 2014.

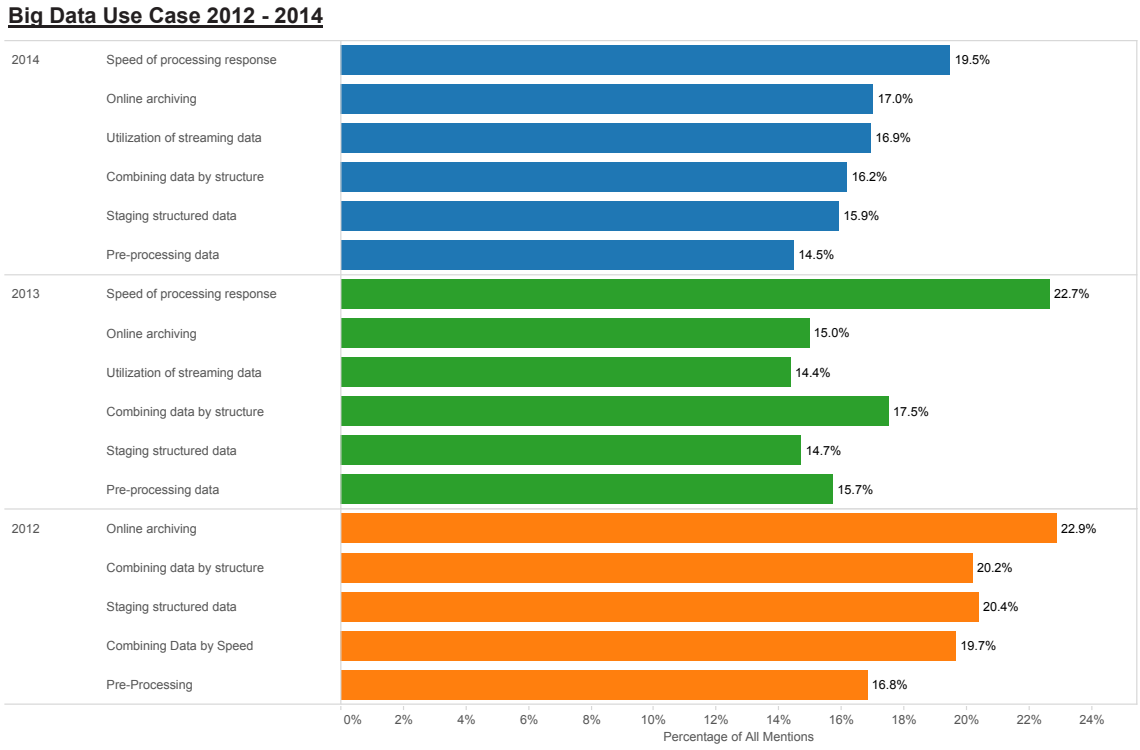


Figure 63

One of the main differences in the trends associated with the use case indications is how the use of **Combining data by structure**, which was the #2 response in 2013 and #3 in 2012 is now ranked at the fourth most popular indication in the 2014 survey. In addition, **Utilization of streaming data,** the sixth most popular use case last year, rose to the third most popular use case in 2014.

Copyright 2015, EMA Inc. and 9sight Consulting. All Rights Reserved.

7.1.3. Application of Speed of Response

In the last two years of EMA/9sight survey data, the importance of **Response** as a driver of big data initiatives has been clear. Its significance is reflected not only in the use cases of the EMA/9sight survey respondents, but also in how they are implementing their projects in terms of how business goals are accomplished.

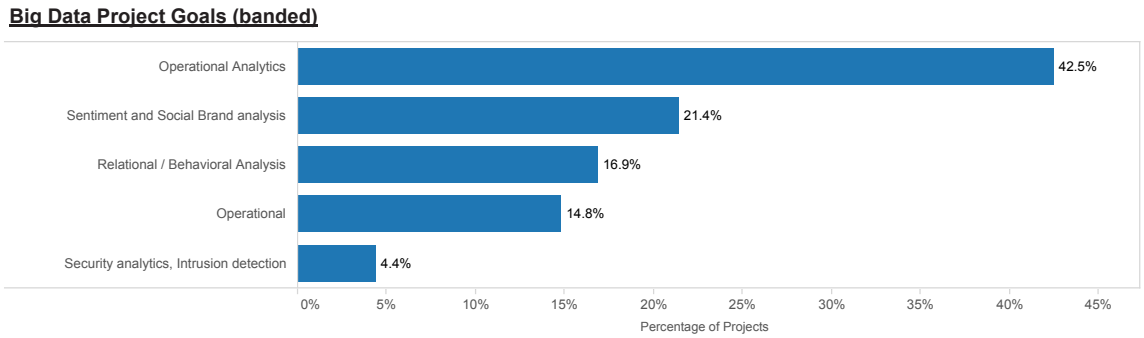


Figure 64

The key to many **Operational Analytics** projects is to provide a speed of **Response** consistent with the speed required by their operational workflows. For example, speed of processing is important in industries such as telecommunications and online consumer retail. In these industries, the businesses of intangible goods and services have low barriers to entry and exit. Fraud management and cross-sell/up-sell projects can provide key elements to competitive advantage as well as profit and loss. This comes not just as a technical initiative, but also as a clearly identifiable business requirement.

7.2. Economics of Big Data

Big data initiatives have impacted the economics of the data management landscape in a number of areas. The impact has been primarily associated with a downward price pressure on the overall costs associated with next-generation data management environments. Price pressure can range from overall budgets to the cost to implement a big data environment to how much an organization needs to invest in the "care and feeding" of a particular environment.

The following section addresses the financial drivers associated with big data environments and also looks at where costs are allocated between traditional sources and the line of business (LOB) when it comes to the budgetary "financing" of big data initiatives.

7.2.1. A Wider Approach to Costs

Economics, costs, and TCO are all ways that strategies have been impacted by big data initiatives. In the 2014 EMA/9sight study, panel respondents were given a new option: to provide feedback on the financial drivers associated with their big data initiatives.

Copyright 2015, EMA Inc. and 9sight Consulting. All Rights Reserved.

The primary driver reported by the EMA/9sight panel was that organizations wanted to **Reduce maintenance costs associated with their existing data management implementations**. The strategy that legacy traditional data management platforms have used to bundle (sometimes) excessive maintenance fees associated with license costs for relatively small returns for those maintenance fees is a major reason for this response.

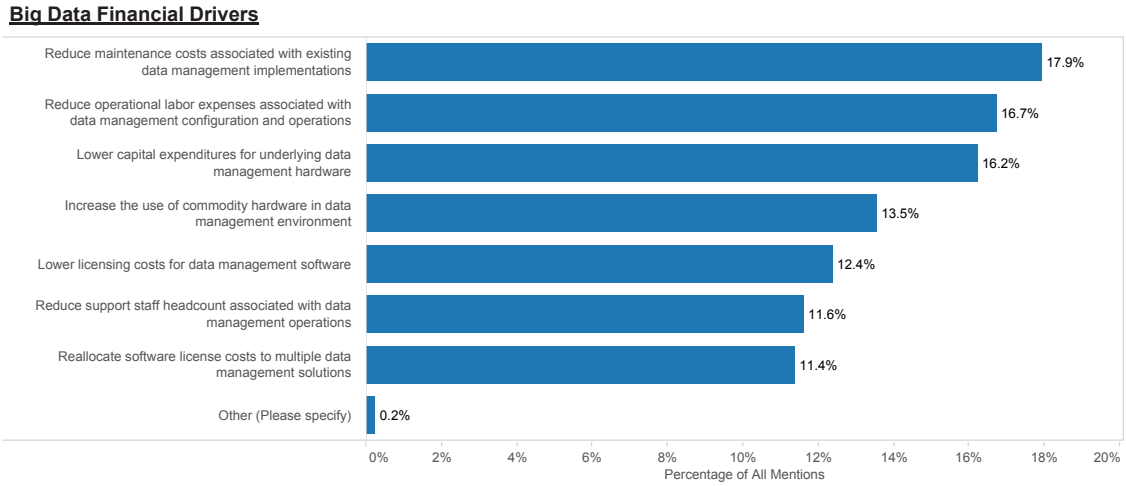


Figure 65

The next most frequent panelist indicator is the **Reduction of labor expenses associated with data management environments**. This cost driver includes the systems administration and data administration aspects of the operation of data management environments. Many next-generation data management platforms include decreased administration requirements through automated or simplified administration consoles due to a "re-thinking" of how data management platforms should be operated and maintained.

The next two most common financial drivers are closely linked. **Lowering the capital expenditures for data management hardware** and **Increasing the usage of commodity hardware** both lower the expense of a big data initiative by reducing the expense and complexity of a hardware environment. The reduction of custom designed or high end hardware allows for either a smaller investment over the same data management challenge, or it allows for organizations to expand to larger data capacities within the same budgetary footprint.

The need to reduce overall costs and the downward pressure on pricing derived from the perception that certain big data technologies are "free" that drives expectations that organizations should do more with less. In that vein, EMA/9sight panel respondents were asked to identify their overall IT and big data budgets.

7.2.2. Looking at Information Technology Budget

Budgets of IT departments are being adjusted to support the financial drivers outlined above. In the 2013 EMA/9sight study, panelists indicated that nearly 50% of big data initiatives received funding from external sources such as LOB organizations allocating direct funding to those initiatives. Not only

Copyright 2015, EMA Inc. and 9sight Consulting. All Rights Reserved.

were business stakeholders looking for big data initiatives to help with their business goals, they were willing to bring their "checkbooks" to the table to make those initiatives possible.

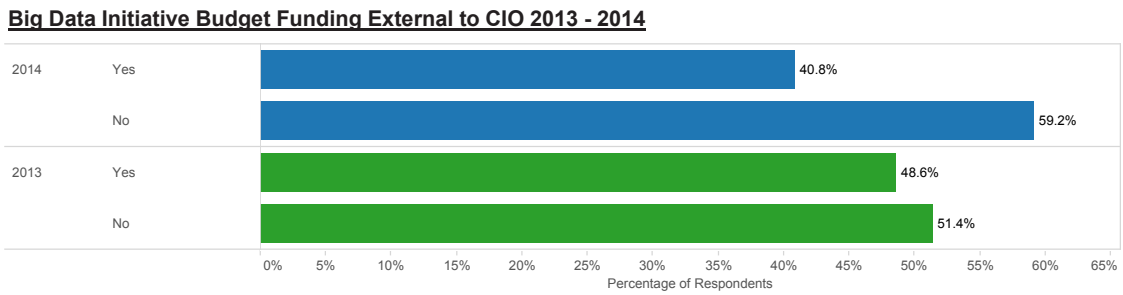


Figure 66

In 2014, there was a decrease in the number of organizations that were funding big data initiatives from line of business budgets. This reduction indicates that organizations have accepted big data as a mainstream initiative and that formalized budgets are being allocated to meet the full expenses of those projects without the additional resources from LOB in industries such as marketing and finance.

The percentage of IT budgets dedicated to big data projects confirms this observation. In 2013, EMA/9sight panel respondents were dedicating a relatively small amount of their budgets to big data initiatives. Nearly 60% of respondents indicated that between 1-15% of their IT budgets was dedicated to big data initiatives.

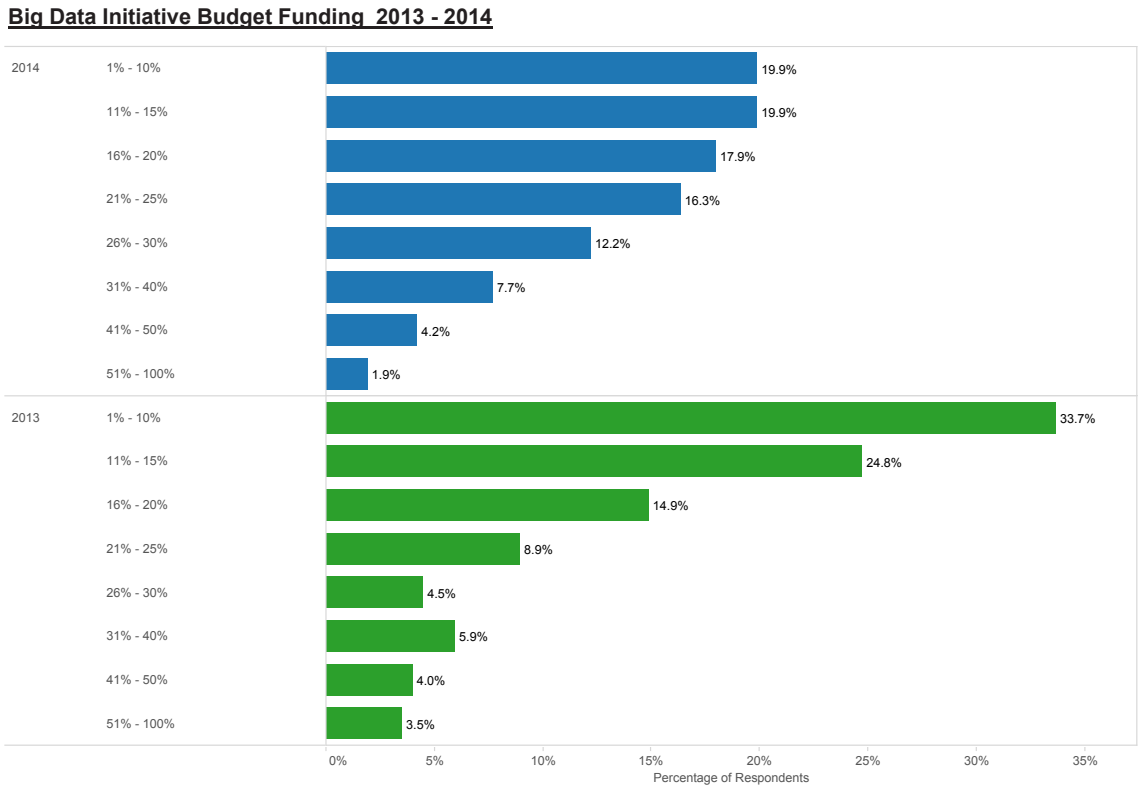


Figure 67

Copyright 2015, EMA Inc. and 9sight Consulting. All Rights Reserved.

In 2014, there is a shift towards the allocation of higher percentages of IT budgets to big data initiatives. Higher percentages of EMA/9sight respondents were dedicating higher allocations of IT budgets toward big data initiatives. Nearly 50% of respondents indicated that between 16-30% of IT budgets was allocated to big data.

When coupling the shift from additional LOB financing with increases in the percentage of IT funding, it is clear that organizations are allocating their budgets to align with the financial drivers detailed in the 2014 EMA/9sight study to manage the increasing size of environments and the workloads being managed by those environments.

7.3. Complex Workloads Go Real Time

Big data initiatives are often driven by the requirement to apply very complex workloads to data sets. This requirement can take the form of large processing sets for operational workloads such as billing and rating for use by online services. It can also appear as complex workloads associated with advanced analytical workloads. These workloads can run the gamut from descriptive analytics to cluster or determine the attributes of information, to predicting the performance of a product or when a particular part may fail. Analytics can expand to much more intensive and complex workloads that are associated with graph analytics to determine relationships across a customer network, or a series of transactions, or the ability of an organization to determine the best path of action associated with cognitive analytical models.

In each of these cases, the requirements of the business goals of an organization should drive the data management strategy of the initiative. The strategy must provide the appropriate level of processing capacity to meet those challenges in both context (operational vs analytical) and in complexity (traditional vs advanced analytics).

7.3.1. How Strategy Impacts Complexity

Looking again at that the business goals of projects, it is clear to see the workload types that the EMA/9sight panelists indicated were the most important. **Operational analytics**, with over 40% of the responses, is a key component of how organizations are looking for their next generation data management environments to support their business organizations.

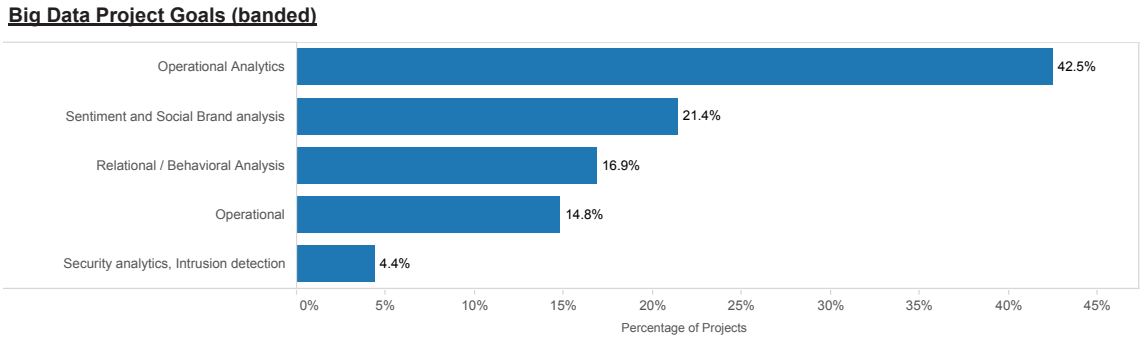


Figure 68

Sentiment and Social Brand analysis follows with over 20% of the project goals in which organizations are using big data projects to link with and understand the relationship between themselves and their customer base. These links can be in the form of social media or Facebook content or in how sentiment from blogs and other long-form review content provide a window into the activities and thoughts of the customer base.

Copyright 2015, EMA Inc. and 9sight Consulting. All Rights Reserved.

It should be noted that security analytics, a new area of indication in the 2014 survey, represents nearly 5% of big data projects. This percentage demonstrates the increasing importance of security and shows how big data can provide support to identify potential security breaches and attempts at intrusion.

7.3.2. How Latency Works with Complexity

Organizations are utilizing a wide array of workloads to support their engagements, and advanced analytical workloads are important members of the array. Over 30% of big data projects are implementing advanced analytical workloads in the form of predictive, descriptive, graph, and cognitive analytical workloads.

Over 30% of big data projects are implementing advanced analytical workloads in the form of predictive, descriptive, graph, and cognitive analytical workloads.

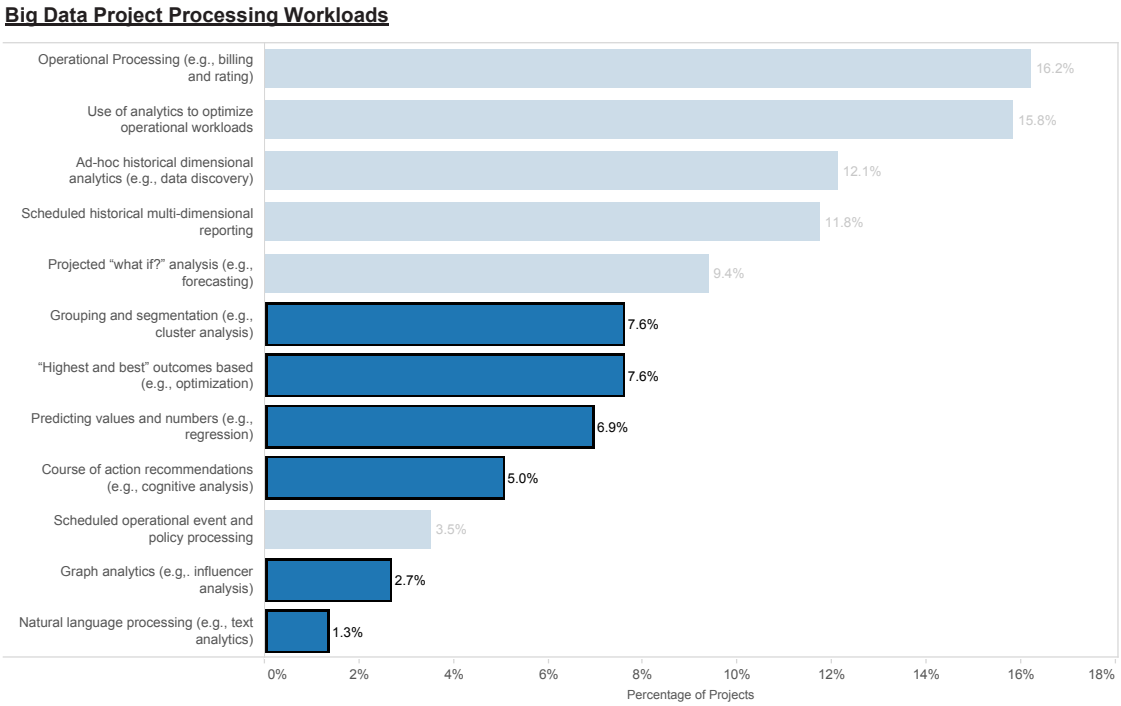


Figure 69

Examination of these complex workloads, reveals that they represent a significant portion the EMA/9sight panel respondents' big data projects. Another important component of big data projects is the speed at which the workloads need to be addressed. Over 56% of big data projects mentioned by EMA/9sight respondents require **Near-real time** latency.

Copyright 2015, EMA Inc. and 9sight Consulting. All Rights Reserved.

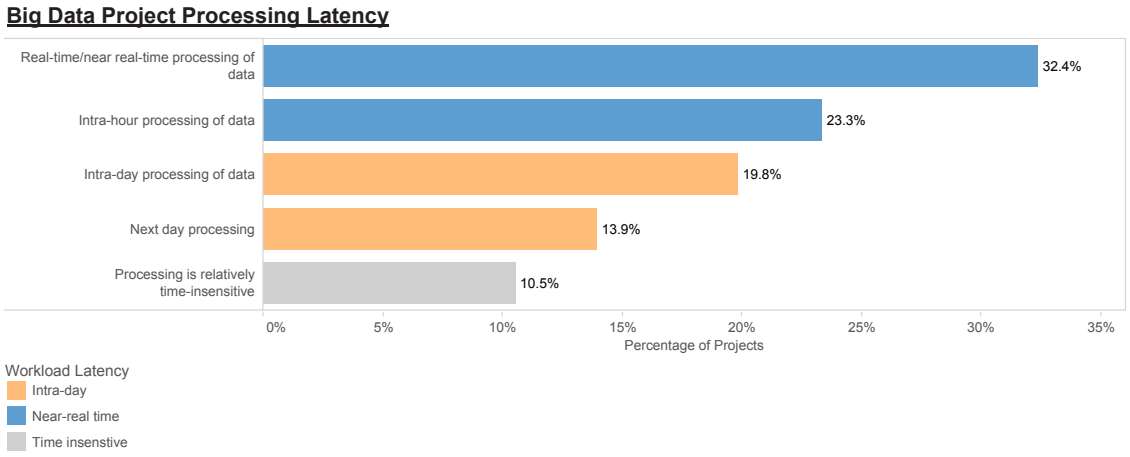


Figure 70

When workload latencies are applied to the aggregated **Project Goals** of various projects, you see that Security Analytics have the highest requirement for low latency processing. Almost 70% of security projects require a near-real-time processing load for their workloads.

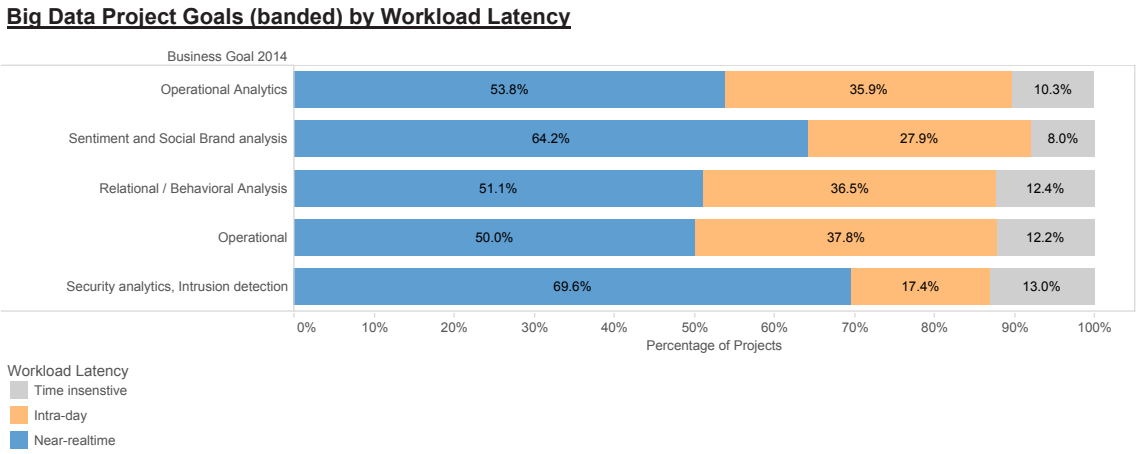
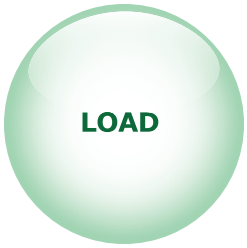


Figure 71

The next highest percentage of low latency for complex analytical workloads is **Sentiment and Social Brand Analysis**. These types of big data projects often involve complex analytical workloads associated with the predictive, descriptive, and graph workloads described earlier.

The next highest percentage of low latency for complex analytical workloads is Sentiment and Social Brand Analysis. These types of big data projects often involve complex analytical workloads associated with the predictive, descriptive, and graph workloads described earlier.

Copyright 2015, EMA Inc. and 9sight Consulting. All Rights Reserved.



7.4. Data Loads

Big data is called "big" for a reason. The data sets associated with big data initiatives generally are of a style and type that expand the overall data management environment for companies. In the past, many of these data sets were too large to be stored. Therefore, there was limited knowledge of the value in web server log files or geolocation information. Next generation data management environments enable storage and utilization of these data sets not just in the short term, but allow for a wider historical analysis that is beneficial in areas such as preventive maintenance associated with the IoT and customer path analysis associated with online and mobile application cross-sell/up-sell applications.

In 2012 and 2013, the EMA/9sight Big Data survey sought to quantify the size of the data stored within the next generation data management platforms associated with big data. In other words, it attempted to define the "big" of big data in terms of data under management. In 2014, EMA/9sight again aims to provide a guidance estimate on the big data portion of data management environments.

Next generation data management environments enable storage and utilization of these data sets not just in the short term.

7.4.1. Overall Environment Sizing

When EMA/9sight survey respondents were surveyed in regards to their overall data management environment, nearly 60% of the indicated that they were managing between 50 TB and 750 TB of data.

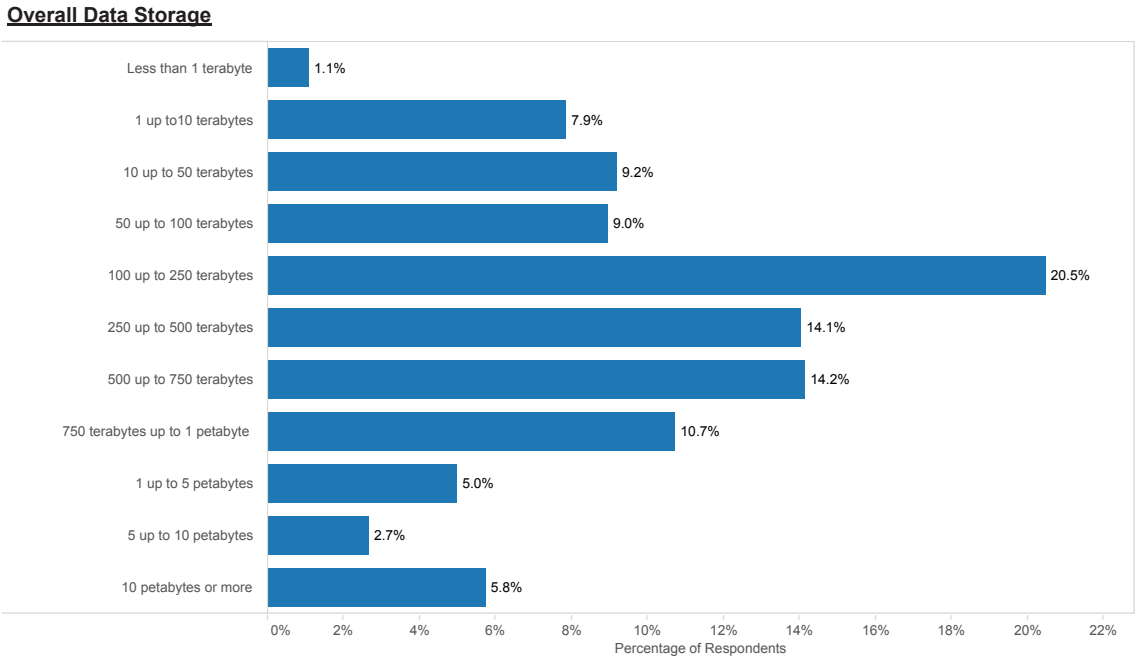


Figure 72

Copyright 2015, EMA Inc. and 9sight Consulting. All Rights Reserved.

The largest percentage of respondents indicated that they had a total percentage of information under data management (i.e., stored in some form or fashion on disk or in-memory) was nearly 21%, in the range of **100-250 TB**.

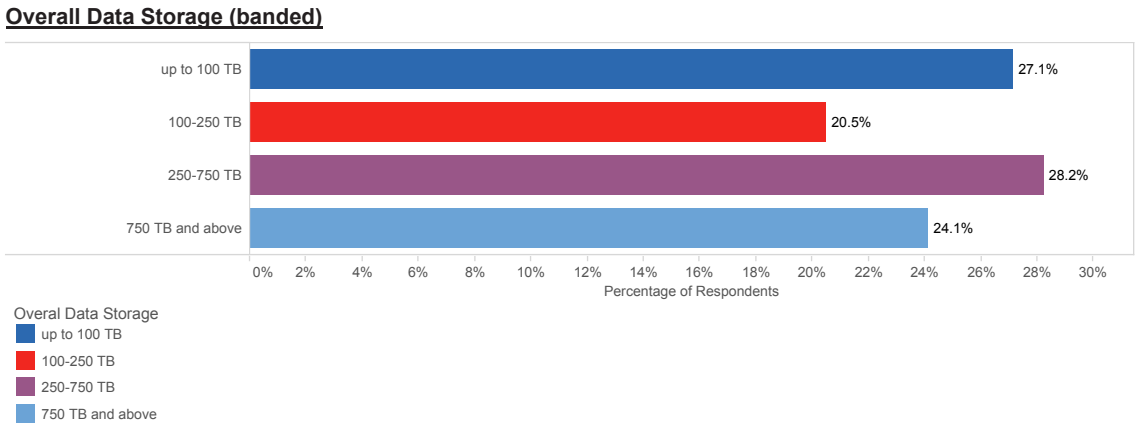


Figure 73

When these categories are banded into four approximate bands, you can see that over 50% of the EMA/9sight panel respondents had 250 TB or more of information stored in their overall environments. The largest grouping of these data storage bands is the **250-750 TB** range.

When you compare the 2014 EMA/9sight responses with those from 2013, it is clear that the lower end of the ranges remains approximately the same. The **750 TB and above** range has "reduced" considerably over the previous year.

When you compare the 2014 EMA/9sight responses with those from 2013 the 750 TB and above range has "reduced" considerably over the previous year.

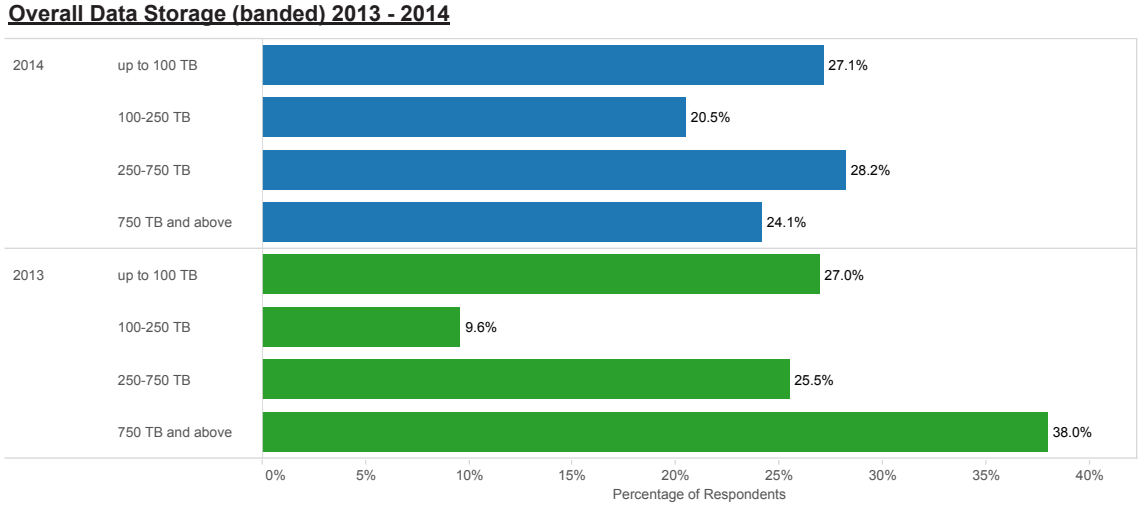


Figure 74

The largest gain from the 2013 survey to this year's is in the **100-250 TB** range.

Copyright 2015, EMA Inc. and 9sight Consulting. All Rights Reserved.

7.4.2. Impact of Company Size on Data Storage

Looking at how organizational headcount impacts the distribution of information under data management, you can see the year-over-year trends by company size. In both the 2013 and 2014 surveys, **Enterprise** organizations dominate the larger data storage bands. **Enterprise** companies represent nearly 80% of the organizations **750 TB and above** in 2013 and nearly over 65% in 2014.

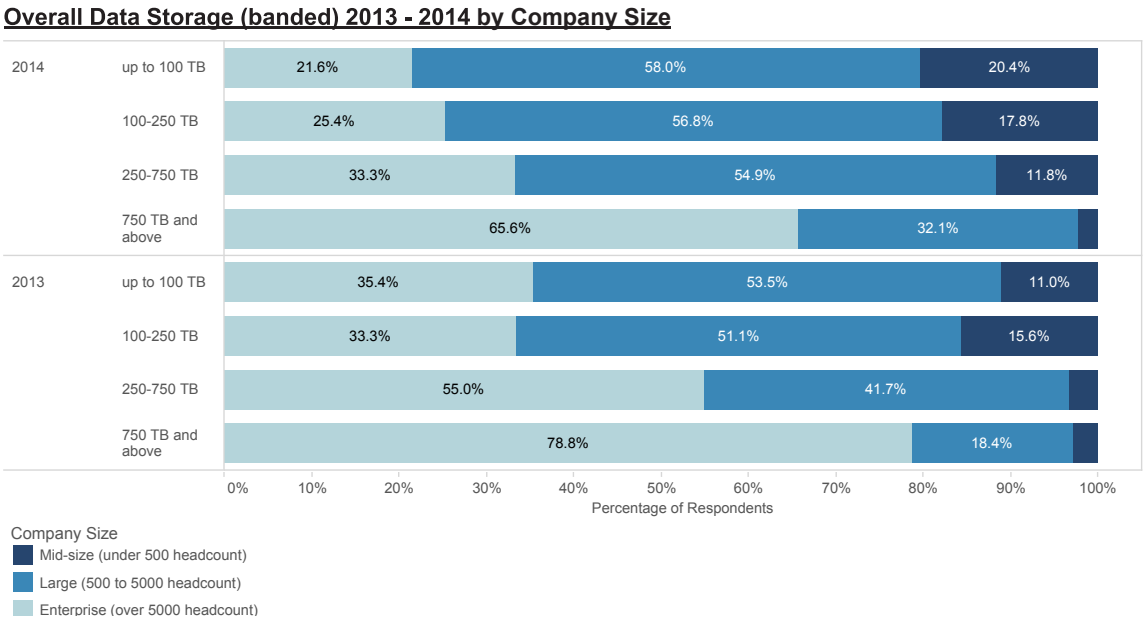


Figure 75

Participation by **Mid-size** companies increased in 2014 in both the overall survey respondent pool and in their representation within the data storage bands. This increase demonstrates how smaller companies are leveraging and growing their data management requirements. **Mid-size** organizations' representation in the **up to 100 TB** range nearly doubled between 2013 and 2014.

7.4.3. Sizing Big Data Environments in 2014

When asked what percentage of their overall environments was dedicated to big data-related data management in 2014, nearly 60% of respondents indicated that big data represented between 11-30% of their overall environment.

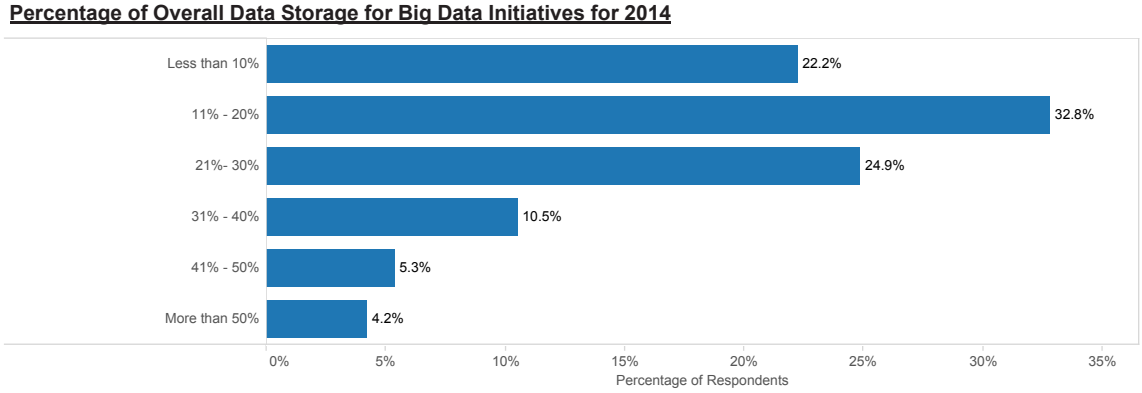


Figure 76

Copyright 2015, EMA Inc. and 9sight Consulting. All Rights Reserved.

The single largest indication among the panel respondents was in the **11-20%** range of data management environments associated with big data initiatives, with nearly one third of all indications.

Using the information from the EMA/9sight respondents on data management environment size and percentage dedicated to big data initiatives, the most "common" big data solution environment is sized to approximately 11 TB to 250 TB[13] as compared with a calculated "most common" range of 5.5 TB to 100 TB[14] in 2013.

7.4.4. Future Projections

Looking toward growth in the next 12 to 18 months, EMA/9sight respondents estimated how their big data environments would grow in 2016 as compared to 2014. Overall, there is a shift toward the higher percentages of big data as a percentage of their overall data storage component.

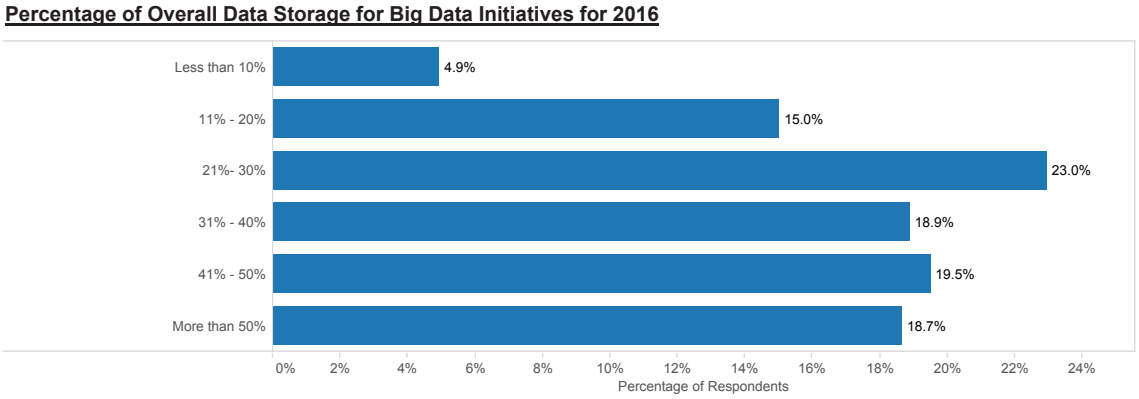


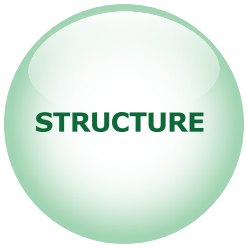
Figure 77

Over 70% of EMA/9sight panel respondents estimated that the big data segment of their data management environments would be between 21 and 50% in 2016. The single largest increase was in the 21-30% range with nearly a quarter of the responses.

Combining these growth projections with the 2014 data management numbers, a 2016 projection of big data environment sizes can be approximated to 21 TB to 375 TB.[15] This project continues the growth for big data initiatives and the environment size for next generation data management platforms.

[13] 100 TB-750 TB x 11%-30% = 11 TB – 250 TB size range
[14] 50 TB-500 TB x 11%-20% = 5.5 TB – 100 TB size range
[15] 100TB-750TB x 21%-50% = 21TB – 375TB size range

Copyright 2015, EMA Inc. and 9sight Consulting. All Rights Reserved.



7.5. Multi-Structured Data: Different Tune, Same Song

While some have simplified the description of big data to being "un-structured," big data environments are comprised of data that comes in many structures and formats. From existing relational structured data sets to images and audio files with vastly different types of information, all these data sources have their own specific data structures, if not data formats.

That being said, the big data movement has done an excellent marketing job of delineating the difference between "two eras" of data. "Before" is typified by traditional relational data structures. "After" is the new era that combines traditional with new, or newly utilized, multi-structured data sets. However, there is often confusion as to what exactly fits into the big data data-type. The main reason for this confusion is that "big data" fits much better on a bumper sticker or Twitter hashtag than "multi-structured event data with relational data to provide operational and analytical context."

7.5.1. Breaking Down Big Data Domains

EMA/9sight panel respondents were asked to detail which data types existed within their big data environments. The most common data type was **Operational application data** that comes from point of sale, customer care, or SCM applications. These data sources, which some may not consider "big data," are the key to many project applications. This data type is the curated information relating to customer, product, and other key aspects of applying a context to multi-structured data sources.

The next most common response was the unstructured data type of **Human-generated documents** with free form components. These "unstructured" data components include emails and application forms for jobs or loans. It should be noted that these types of data sources are not truly "unstructured"; many aspects of these free-form data sources have structure and metadata that can be utilized for operational and analytical uses. Emails have date/time stamps and addressee lists with cc and bcc information. Application forms have discrete areas such as date/time and name and address fields if they can be translated with optical recognition (OCR) technology.

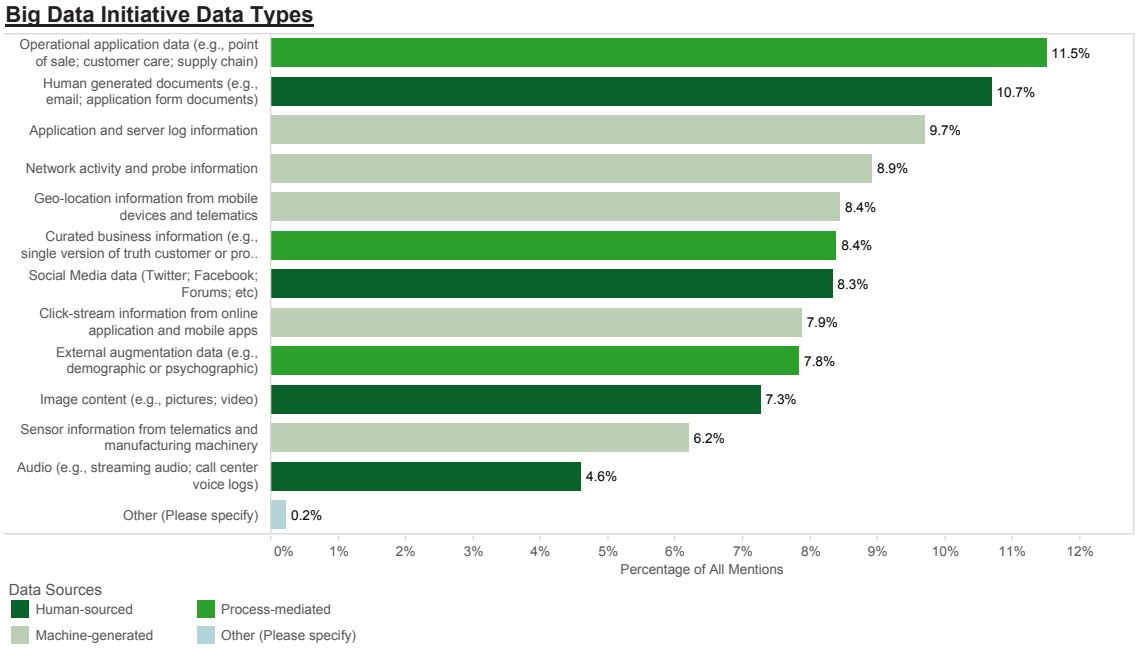


Figure 78

Copyright 2015, EMA Inc. and 9sight Consulting. All Rights Reserved.

The next three most common data types are the components that are most associated with big data environments: the multi-structured **Application and server log**, **Network activity,** and **Geolocation** data. In the "traditional" era of data, these data types were often not stored because of the perceived limited value of the information. For example, **Network activity information** was often viewed as important to the network engineering teams for operational purposes (i.e., the proper management of a corporate network), but not viewed as critical to a wider audience within the organization. However, in a retail setting, if the amount of time that customers spend in a brick-and-mortar store and the websites that those customers might use to "price compare" while inside the store could be identified, the information that comes from a network has relevance for both marketing and competitive awareness. No retailer wants to be the "showroom" for their online competitors, but they do want to better understand the needs of their customer base.

For clarification, these data types are segmented into the following areas.

- **Machine-generated** originates from sensor or automatically generated data sources such as **Geolocation information from mobile devices and telematics**; **Click-stream information from online application and mobile apps; Sensor information from telematics and manufacturing machinery; Application and server log information**; and **Network activity and probe information**. This information can take the form of multi-structured data.
- **Human-sourced** generated from sources where it is more difficult to determine and assign business and technical metadata for processing. These sources include **Audio (e.g., streaming audio; call center voice logs)**; **Social media data (Twitter; Facebook; forums; etc)**; **Human-generated documents (e.g., email; application form documents)**; and **Image content (e.g., pictures; video)**
- **Process-mediated** is typified by information from operational support systems (OSS). These platforms generally produce structured data with well-defined business and technical metadata. **Operational application data (e.g., point of sale; customer care; supply chain)**; **External augmentation data (e.g., demographic or psychographic)**; and **Curated business information (e.g., single version of truth customer or product information)** are examples.

As one would expect, with these segments applied, **Machine-generated** data types dominate the data types for big data initiatives with over 40% of respondent mentions.

As one would expect, Machine-generated data types dominate the data types for big data initiatives with over 40% of respondent mentions.

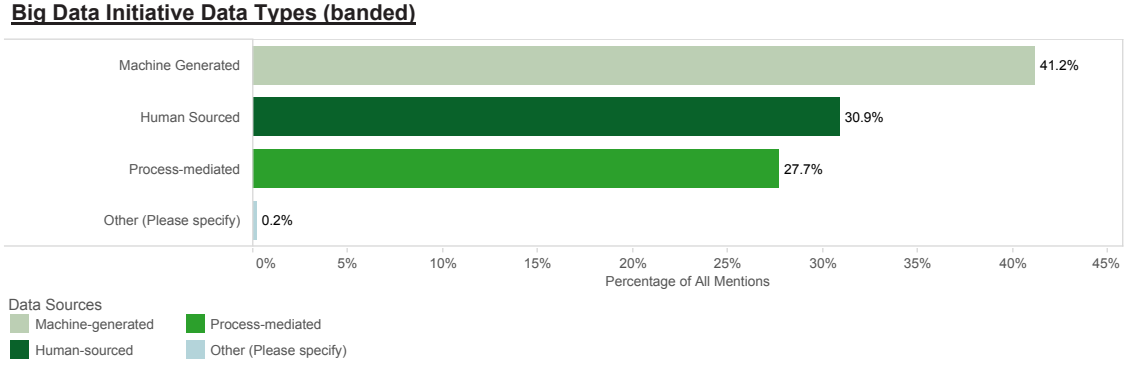


Figure 79

Copyright 2015, EMA Inc. and 9sight Consulting. All Rights Reserved.

Human-sourced data types associated with unstructured text and audio and video represent over 30% of the mentions. Traditional **Process-mediated** structured/relational data types are the remaining, but not insignificant, portion of the information associated with big data initiatives.

As mentioned earlier, when looking at the trends over the history of the EMA/9sight Big Data study, it is easy to see how **Machine-generated** sources have grown over time.

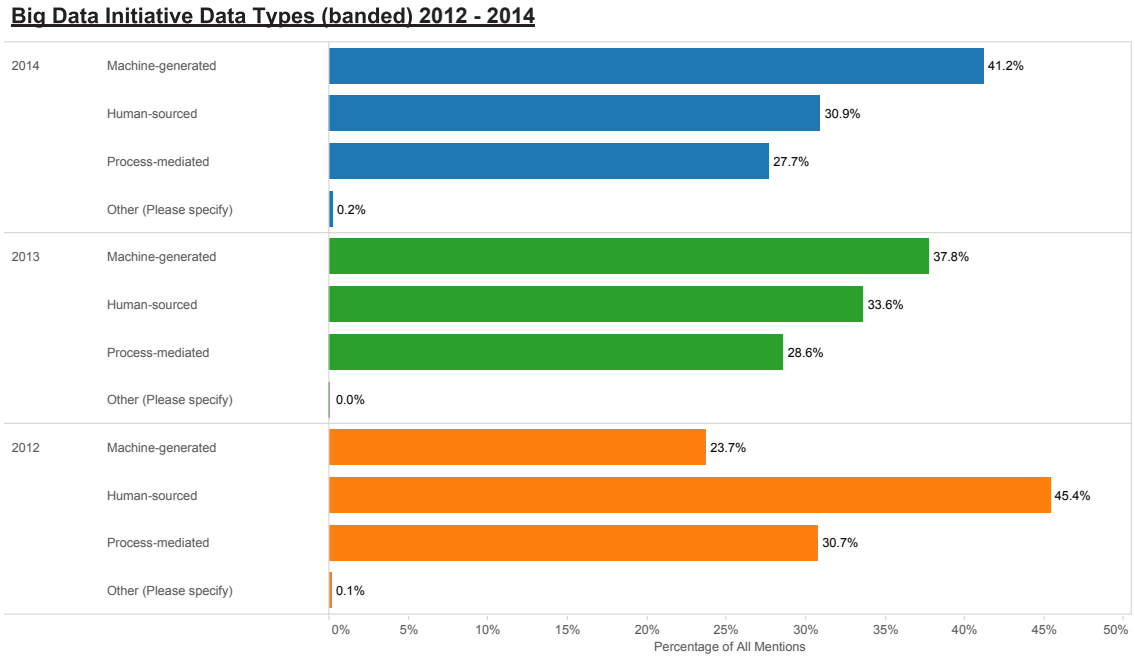


Figure 80

The role of **Process-mediated** data has maintained a fairly consistent level over the course of the EMA/9sight survey with approximately 28-30% of the data types within big data environments.

The role of Process-mediated data has maintained a fairly consistent level over the course of the EMA/9sight survey with approximately 28-30% of the data types within big data environments.

Copyright 2015, EMA Inc. and 9sight Consulting. All Rights Reserved.

7.5.2. Moving Big Data

Coordination of the wide variety of data sources into a big data environment to meet the workload requirements of various data consumers requires multiple data integration techniques. When operational and analytical systems were integrating data between relational-based data structures, the integration process was relatively simple. With the introduction of multi-structured data sets and image/audio formats, the integration process becomes more complex in two respects.

The first hurdle is the increase of the number of data formats. The introduction of multi-structured data sets in addition to relational formats causes an increase in the types of information to integrate. As stated previously, more than relational data sets must be integrated; new data types have many more complexities than previous relational structures. The second increase in complexity is the number of systems that organizations must now consider. Adding the relatively large number of platforms that exist outside the corporate firewall to the comparatively small number of platforms that are contained within the firewall also increases the complexity.

When asked to designate which data integration techniques are used in managing data within their big data environments, the 2014 EMA/9sight survey respondents indicated that their top three strategies were **Real-time streaming data integration**, **Data replication (e.g., standardized duplication of enterprise data sources)**, and **Change data capture**.

When asked to designate which data integration techniques are used in managing data within their big data environments, respondents indicated that their top three strategies were Real-time streaming data integration, Data replication (e.g., standardized duplication of enterprise data sources), and Change data capture.

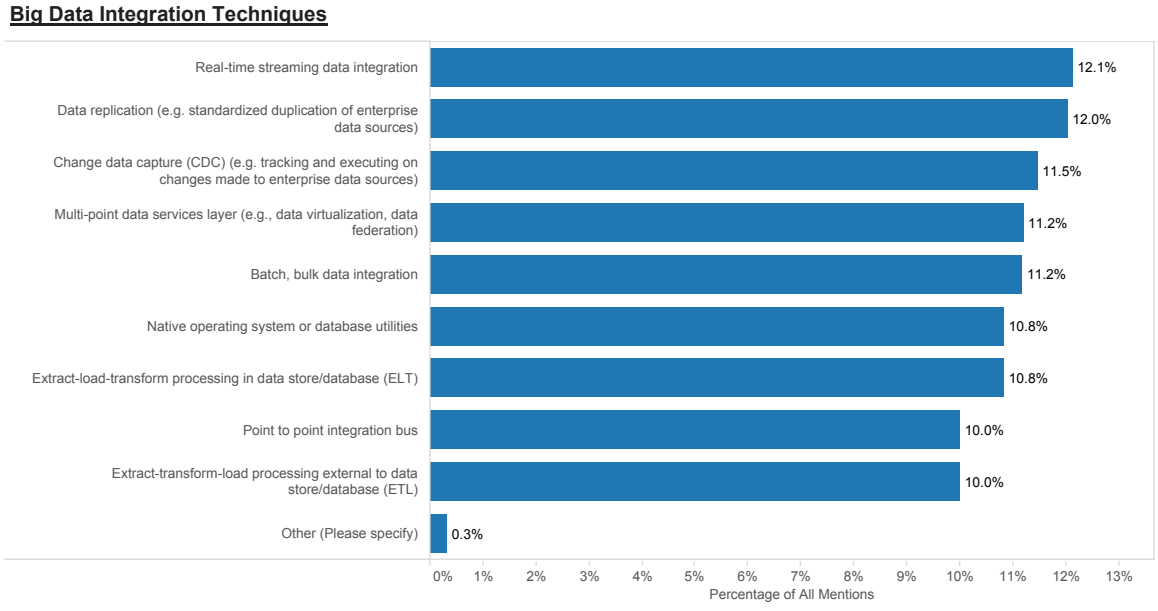


Figure 81

Copyright 2015, EMA Inc. and 9sight Consulting. All Rights Reserved.

Comparisons between the 2013 and 2014 surveys for data integration techniques in big data environments show that the concept of **Batch, bulk data integration** is losing traction.

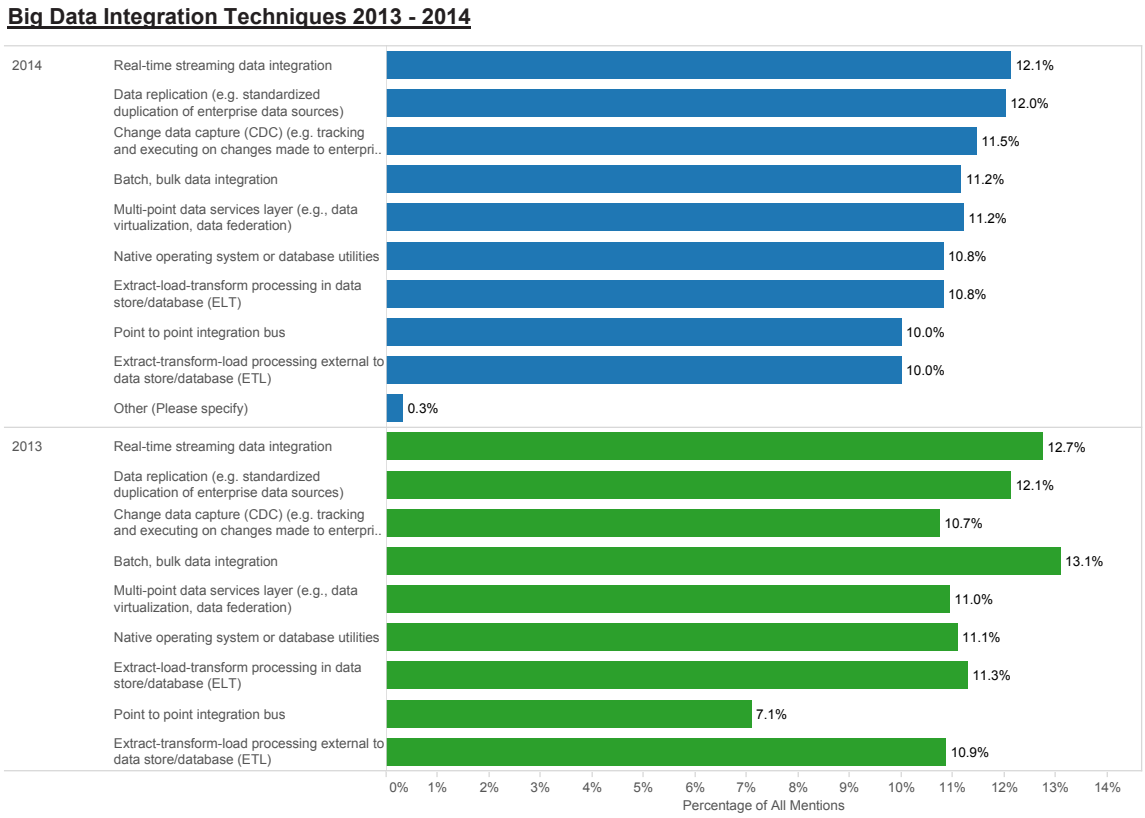


Figure 82

As organizations continue to refine the information management layer of their next generation data management environments, they are refining the amount of information moved around their environment. Targeted transfers of data are used rather than bulk transport from one platform to another. This change in tactics reduces the amount of uncoordinated replication of data within the wider environment as well as imposes a smaller impact on the network connections between disparately located data management platforms.

As organizations continue to refine the information management layer of their next generation data management environments, they are refining the amount of information moved around their environment.

Copyright 2015, EMA Inc. and 9sight Consulting. All Rights Reserved.

8. Methodology and Demographics

8.1. Research Methodology

EMA and 9sight Consulting crafted the big data user surveys that are the basis for this report. Before the survey was conducted, report sponsors were provided with a copy of the survey instrument. However, sponsors had no direct involvement in or influence on the survey creation, survey contents, survey execution, or any of the subsequent evaluation and analysis of the results for this report.

For this research, EMA and 9sight invited pre-qualified business intelligence (BI) and information technology (IT) professionals to complete an extensive web-based survey. These respondents were further qualified based on their responses to the following questions.

- What is your primary role in the usage and/or management of big data applications/technology within your organization?
- Which of the following best describes your company's primary industry?
- How would you describe the extent to which big data initiatives have been adopted within your business/organization?
- What is your relation to big data applications/products currently being used within your organization?
- At what phase of implementation are your business area /organization's big data initiative's project(s)?

Respondents who failed to qualify on these questions were rejected. As a result, all respondents (in addition to being independently pre-qualified through the initial invitation process) self-identified as being active participants with a working knowledge of current operational and analytical data management practices within their company that is presently researching, planning, or implementing big data strategies.

8.2. 2014 Respondents

In 2014, 351 business and technology professionals responded to an invitation to provide their insights on big data strategies and implementation practices. To offer a balanced enterprise view of the subject, the respondent pool was also restricted. Business stakeholders represented 57% of respondents. Technologists were 43%.

The 2014 survey instrument was executed in December 2014.

8.3. 2013 Respondents

In 2013, 259 business and technology professionals responded to an invitation to provide their insights on big data strategies and implementation practices. To offer a balanced enterprise view of the subject, the respondent pool was also restricted. Business stakeholders represented 51% of respondents. Technologists were 48%. Professional services consultants in IT represented less than 1% of the response panel.

The 2013 survey instrument was executed between July and August 2013.

8.4. 2012 Respondents

For 2012, 255 business and technology professionals responded to the survey invitation. To provide balance, EMA/9sight restricted the respondent pool to an approximate mix of 45% business stakeholders, 45% IT participants, and 10% IT consultants.

The 2012 survey instrument was conducted between July and August 2012.

Copyright 2015, EMA Inc. and 9sight Consulting. All Rights Reserved.

9. Authors

John Myers
EMA

Dr. Barry Devlin
9sight

9.1. About Enterprise Management Associates

Founded in 1996, Enterprise Management Associates (EMA) is a leading industry analyst firm that provides deep insight across the full spectrum of IT and data management technologies. EMA analysts leverage a unique combination of practical experience, insight into industry best practices, and in-depth knowledge of current and planned vendor solutions to help its clients achieve their goals. Learn more about EMA research, analysis, and consulting services for enterprise line of business users, IT professionals, and IT vendors at http://blogs.enterprisemanagement.com. You can also follow EMA on Twitter or Facebook.

9.2. About 9sight

Dr. Barry Devlin, founder and principal of 9sight Consulting (www.9sight.com), is among the foremost authorities on business intelligence (BI), big data and beyond. He is a founder of data warehousing, having defined its first architecture in 1985. A respected visionary and thought-leader in the evolving data industry, Barry has authored two ground-breaking books: the classic "Data Warehouse--from Architecture to Implementation" and "Business unIntelligence--Insight and Innovation Beyond Analytics and Big Data" (bit.ly/BunI_Book) in 2013.

With over 30 years of experience in IT, until 2007 with IBM as a consultant, manager and distinguished engineer, Barry provides strategic consulting and thought-leadership to buyers and vendors of BI and Big Data solutions. He is an associate editor of TDWI's Journal of Business Intelligence, and a regular keynote speaker, teacher and writer on all aspects of information creation and use. Barry operates worldwide from Cape Town, South Africa.

This report in whole or in part may not be duplicated, reproduced, stored in a retrieval system or retransmitted without prior written permission of Enterprise Management Associates, Inc. All opinions and estimates herein constitute our judgement as of this date and are subject to change without notice. Product names mentioned herein may be trademarks and/or registered trademarks of their respective companies. "EMA" and "Enterprise Management Associates" are trademarks of Enterprise Management Associates, Inc. in the United States and other countries.

©2015 Enterprise Management Associates, Inc. All Rights Reserved. EMA™, ENTERPRISE MANAGEMENT ASSOCIATES®, and the mobius symbol are registered trademarks or common-law trademarks of Enterprise Management Associates, Inc.

Corporate Headquarters:
1995 North 57th Court, Suite 120
Boulder, CO 80301
Phone: +1 303.543.9500
Fax: +1 303.543.7687
www.enterprisemanagement.com

3082.040115

Copyright 2015, EMA Inc. and 9sight Consulting. All Rights Reserved.