

Operationalizing the Buzz: Big Data 2013

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

April 2014

Prepared for: **Pivotal**™



IT & DATA MANAGEMENT RESEARCH,
INDUSTRY ANALYSIS & CONSULTING



Operationalizing the Buzz: Big Data 2013

Table of Contents

1. Executive Summary	1
1.1 Key Findings	2
2. Hybrid Data Ecosystem	2
2.1 Platform Trends	3
2.2 Ecosystem Diversity	4
2.3 Updates to the Ecosystem in 2013	5
Corporate Background	6
Product Description	6
Hybrid Data Ecosystem Product Positioning	7
EMA Perspective	7

Operationalizing the Buzz: Big Data 2013

1. Executive Summary

The 2013 EMA/9sight Big Data research makes a clear case for the maturation of Big Data as a critical approach for innovative companies. This year's survey went beyond simple questions of strategy, adoption and use to explore why and how companies are utilizing Big Data. This year's findings show an increased level of Big Data sophistication between 2012 and 2013 respondents. An improved understanding of the "domains of data" drives this increased sophistication and maturity. Highly developed use of **Process-mediated**, **Machine-generated** and **Human-sourced** information is prevalent throughout this year's study.

The 2013 study dives deep into the Big Data project initiatives of EMA/9sight respondents focusing on multiple characteristics within each. These 259 respondents, averaging between two and three projects in their Big Data programs, provided information on nearly 600 ongoing Big Data efforts. Over 50% of these projects have an implementation stage of **In Operation – In Production** or **Implemented as a Pilot**. Respondents indicated that the top three business challenges were associated with **Risk Management** activities, **Ad-Hoc Operational** queries, and **Asset Optimization** operations. These projects provide groundbreaking detail information into not just the strategy of Big Data implementations, but also the details on implementation choices: on-premises vs cloud; project sponsors throughout the organization, specifically outside the office of the CIO; and actual implementation stages.

Speed of Processing Response has replaced **Online Archiving** as the top Big Data use case in the 2013 study. This shows that organizational strategies are moving from discovering "*the things we don't know we don't know*" into managing Big Data initiatives toward achievable business objectives and "*the things we know we don't know*." That being said, many of the individual projects being implemented are still using an **Online Archiving** use case. **Speed of Processing Response** and **Online Archiving** are the two most popular uses cases in projects classified as **In Operation** indicating that these use cases are critical to early Big Data adopters.

Respondents in the 2013 survey indicated that the information consumers (users) of these Big Data projects are coming from the less technical ranks of their companies. Approximately 50% of users were from business backgrounds with **Line of Business Executives** and **Business Analysts** representing the top two responses. This shows that Big Data projects are moving beyond **Data Scientist** as the primary user of these projects. When examining the sponsors of Big Data projects, business is not only using the information results from these systems, but also "putting their money where their users are." Nearly 50% of all Big Data projects are sponsored by business organizations such as Finance, Marketing and Sales. Just over two of ten Big Data projects were sponsored directly by the CIO.

Integrating Big Data initiatives into the fabric of everyday business operations is growing in importance. The types of projects being implemented overwhelmingly favor **Operational Analytics**. **Operational Analytics** workloads are the integration of advanced analytics such as customer segmentation, predictive analytics and graph analysis into operational workflows to provide real-time enhancements to business processes. An excellent example of Operational Analytics can be found as organizations move toward the real-time provisioning of goods and services. It is critical to provide visibility into AND action regarding illicit activities among customers. In addition, risk assessments become more important as businesses use value-based decisions to determine courses of action to pursue new customers and/or to retain existing ones.

In summary, the world of Big Data is maturing at a dramatic pace and supporting many of the project activities, information users and financial sponsors that were once the domain of traditional structured data management projects. It is possible that within the next three to five years, Big Data will have fully absorbed those traditional approaches into a new world driven by a more open and dynamic set of data best practices.

Operationalizing the Buzz: Big Data 2013

1.1 Key Findings

The 2013 EMA/9sight Big Data research surveyed 259 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 2013 Big Data research and comparisons to the 2012 results:

- **The Internet of Things Is Coming...If Not Here: Machine-generated** data represents the fastest growing data source for Big Data projects. This includes machine-to-machine and application log file information that contributes to linking devices to the Internet.
- **Big Guys Are Getting Into Big Data:** Enterprise sized organizations made the largest jump in survey participation between 2012 and 2013. This indicates that Big Data programs are making their way into the most highly governed IT environment – the enterprise corporate data center.
- **Spreading Around The Globe:** Respondents in the Asia-Pacific (APAC) region showed the largest increase in response for the 2013 survey over 2012. Although the APAC region addresses Big Data with unique requirements, respondents provide insights into how Big Data is being utilized outside of North America.
- **Moving Faster Than Ever Before:** Of the Big Data Use Cases for our respondents, the top response was for **Speed of Processing Response** with over 50% of the total, illustrating that organizations are focusing less on exploring their data and more on how fast they can process information.
- **New Brand of Workload:** Operational Analytics – the integration of advanced analytics in real-time operational workflows – is the most prevalent type of project workload. From segmentation to asset optimization to risk management, Operational Analytics is pushing into critical business workflows.
- **Business Is Consuming Big Data Information:** Nearly 50% of Big Data project users detailed in the 2013 study were business stakeholders: Line of Business Executives and Business Analysts from marketing, finance and customer care departments.
- **Economics Are Important:** Big Data technologies are applying pressure to the costs associated with many processing platforms. Top business challenges for 2013 respondents are Improved Data Management, TCO and Improving Competitive Advantage.
- **Big Data Grows Beyond the Office of CIO:** Almost 50% of respondents indicated that funding for their Big Data initiatives originated from outside the overall IT budget. Finance, Marketing and Sales were the top non-CIO sponsors of Big Data projects.

Operational Analytics – the integration of advanced analytics in real-time operational workflows – is the most prevalent type of project workload. From segmentation to asset optimization to risk management, Operational Analytics is pushing into critical business workflows.

2. Hybrid Data Ecosystem

In the 2012 “Big Data Comes of Age” study, EMA and 9sight identified that Big Data implementers and consumers are relying on a variety of platforms (not just Hadoop) to meet their Big Data requirements. EMA has established there is a collection of platforms that support Big Data initiatives. These platforms include new data management technologies such as Hadoop, MongoDB and Cassandra. But the collection also includes traditional SQL-based data management technologies supporting data

Operationalizing the Buzz: Big Data 2013

warehouses and data marts; operational support systems such as customer relationship management (CRM) and enterprise resource planning (ERP); as well as cloud-based platforms leveraging freely available data sets from sources such as the Open Government Initiative (<http://www.data.gov/>) to software-as-a-service (SaaS) platforms such as Salesforce.com. EMA refers to this collection of platforms as the **Hybrid Data Ecosystem**. These platforms include:

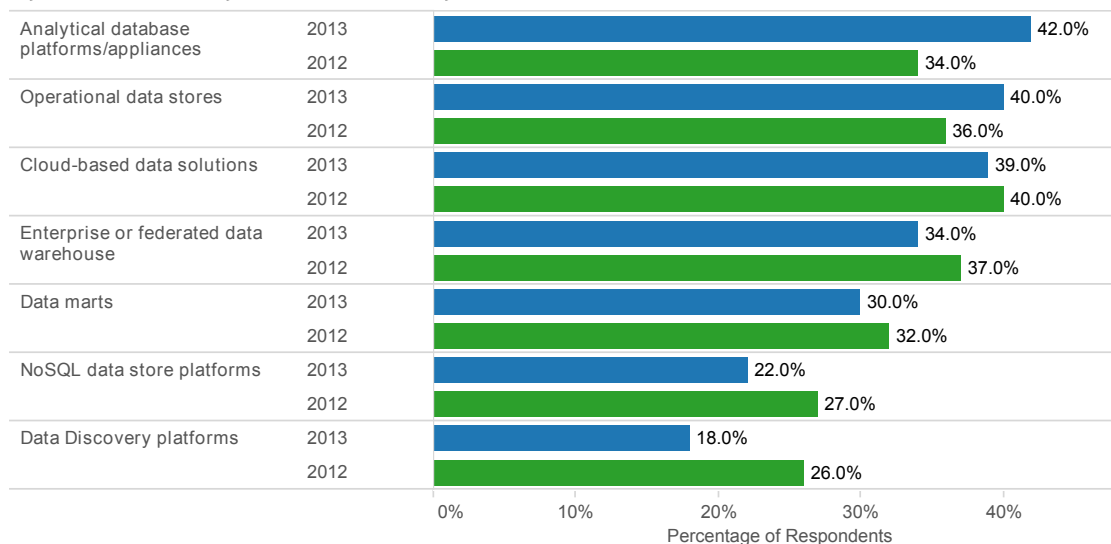
- Enterprise or federated data warehouse
- Data marts
- Operational data stores
- Analytical database platforms/appliances
- NoSQL data store platforms
- Data Discovery platforms
- Cloud-based data solutions
- Hadoop and its subprojects

Each of the platforms within the **Hybrid Data Ecosystem** supports a particular combination of business requirements and processing challenges. This is a relatively unique approach when compared to traditional best practices. Rather than maintaining a single data store that supports all business and technical requirements at the center of this architecture, the **Hybrid Data Ecosystem** seeks to find the best platform for a particular set of requirements and link those platforms together.

2.1 Platform Trends

There were changes in the choices of EMA/9sight panel respondents concerning technology platforms from 2012 to 2013. The most significant of these differences between the 2012 and 2013 surveys focus on two platform types in particular: **Analytical Data Platforms/Appliances** and **Operational Data Stores**.

Hybrid Data Ecosystem Platform by Year



Analytical Data Platforms/Appliances made the largest jump in utilization, from 34% to 42% of respondents. This change reflects how important **Speed of Processing Response** is in Big Data use

Operationalizing the Buzz: Big Data 2013

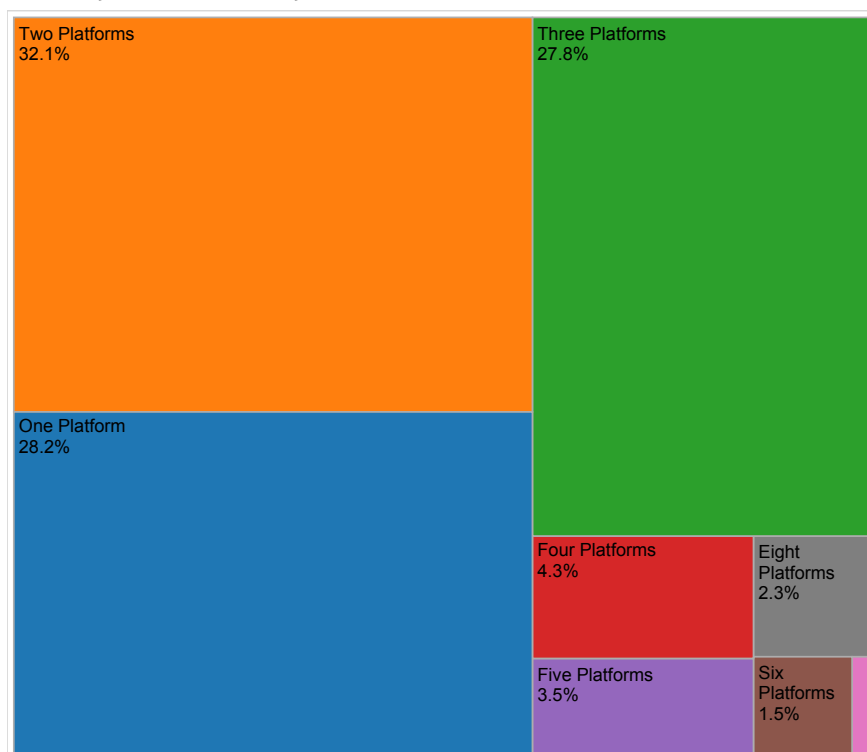
cases and the implementation of realtime **Operational Analytical** workloads. This also matches the workload types that **Analytical Data Platforms/Appliances** were designed to handle. The increase in responses for **Operational Data Stores** shows how Big Data initiatives are continuing to press into the everyday processes of organizations. From specific Big Data systems that handle order processing and point of sales to the inclusion of operational datasets into Exploratory and Analytical strategies, **Operational Data Stores** are some of the best sources of data to drive improvement in business processes, and by extension, competitive advantage.

Of the platforms that showed a decrease between 2012 and 2013, **NoSQL Data Stores** and **Data Discovery Platforms** fell to the last two places on the trend analysis. One of the main differences between the 2012 and 2013 surveys was the specific inclusion of Hadoop as a platform type separate from NoSQL Data Stores. This adjustment to the survey options also contributed to the drop in **Data Discovery Platforms**. Hadoop and Hadoop HDFS are considered components of many **Data Discovery Platforms** that bridge the gap between NoSQL and SQL access layers.

2.2 Ecosystem Diversity

When asked how many platforms were part of their Big Data initiatives, the EMA/9sight respondents indicated that a wide number of **Hybrid Data Ecosystem** platforms were important to their Big Data environments. The most common environment was **Two Platforms** with over 30% of responses.

2013 Hybrid Data Ecosystem Platform Distribution



Nearly 65% of respondents are using two to four platforms, which indicates that they are implementing fairly complex and diverse combinations of technology to power their **Hybrid Data Ecosystem** environments.

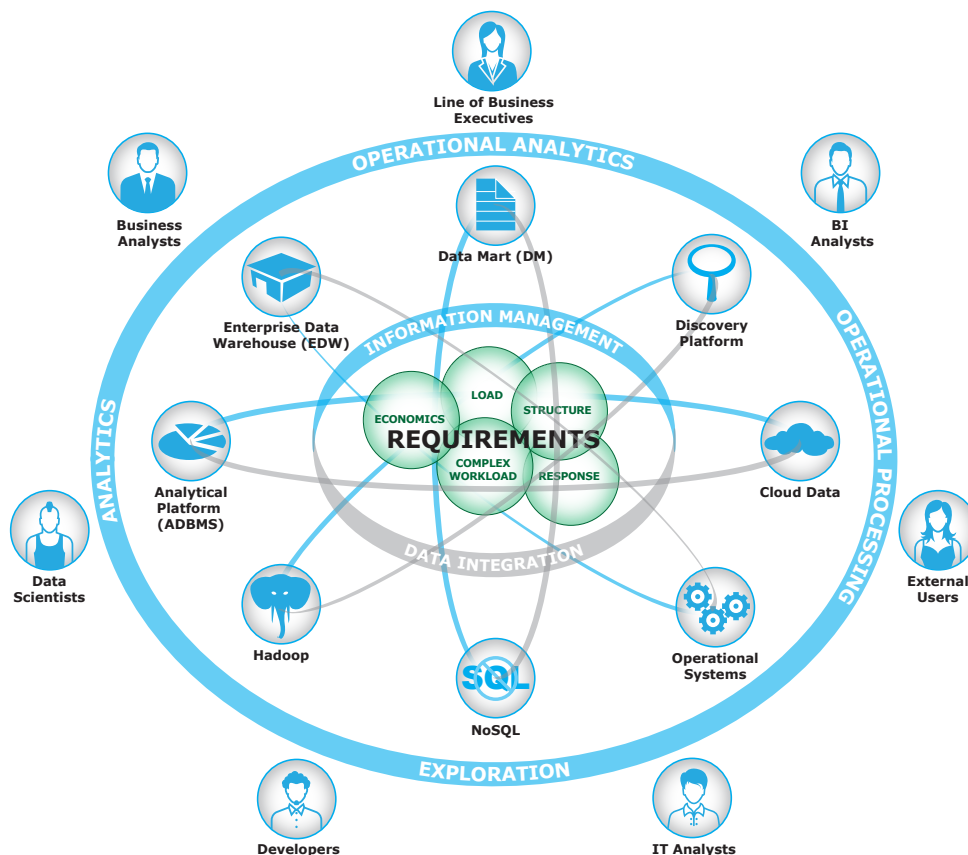
Operationalizing the Buzz: Big Data 2013

2.3 Updates to the Ecosystem in 2013

For 2013, EMA expanded the definition of the **Hybrid Data Ecosystem** to include **Information and Data Management** and a focus on **Information Consumers**. Our 2013 results have also provided deeper insights into the workloads of this environment.

- **Information and Data Management:** The 2012 research defined the number of platforms companies were using as well as how the platforms were related. In 2013, respondents provided deeper insights into how they choose to move information in a bi-directional manner between platforms and which technologies make that information management a reality.
- **Workloads:** The concepts of **Speed of Response** and **Complex Workload** were established in 2012 as key components of the **Hybrid Data Ecosystem** requirements. This year's research leveraged new project-based results to identify the workloads that Big Data initiatives are tackling. They included: Operational workloads associated with ordering, provisioning and billing for goods and services; Analytics workloads for summarizing, predicting and categorizing business operations; Operational Analytics workloads for the integration of analytical models into realtime business processes; and Exploration workloads designed to quickly and iteratively determine new uses for Big Data sources.
- **Information Consumers:** In 2013, the role of information consumer or user was added to the Hybrid Data Ecosystem framework. As important as the underlying technology and processing results are, the users are the most important aspect of a Big Data initiative. Users are the direct links to the top and bottom line of the balance sheet and the best way to gauge the success or failure of a Big Data initiative.

The following details the 2013 EMA Hybrid Data Ecosystem, supported by two years of extensive user research on Big Data initiatives.



Operationalizing the Buzz: Big Data 2013

Corporate Background

Created in April 2013, Pivotal includes assets from both EMC and VMware to create a 1,700 person independent company. Pivotal is owned in partnership by EMC, VMware and General Electric. The company's mission is to support customers in constructing a new class of applications, leveraging Big Data and fast implementation methodologies with the independence of cloud infrastructure. Pivotal serves customers in the following industries:

- Financial Services
- Healthcare
- Internet Services
- Media
- Travel

Headquartered in San Francisco, CA, Pivotal supports open source and open standards as part of its application and data infrastructure software, agile development services, and data science consulting. The following products and services are part of Pivotal and utilized with the Pivotal Big Data Suite: Greenplum DB, HAWQ, GemFire, SQLFire, GemFire XD, Pivotal HD.

HIGHLIGHTS

Vendor Name: Pivotal

Product Name: Pivotal Big Data Suite

Product function: Integrated Big Data storage, transformation and analytics platform

Vendor contact: info@gopivotal.com

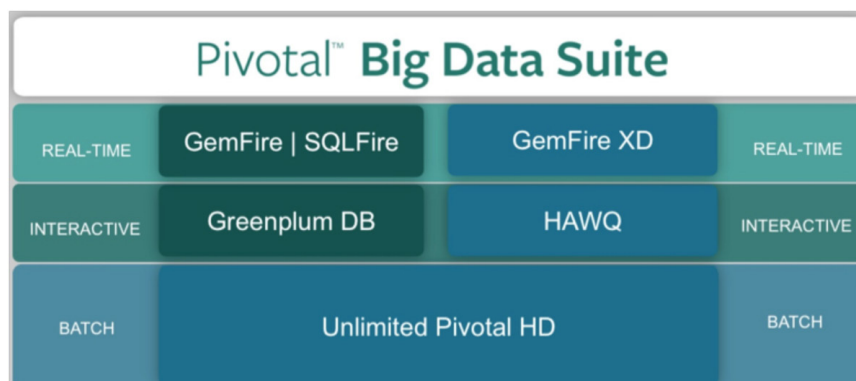
Availability: General Availability

Pivotal™

Product Description

Pivotal Big Data Suite is a unified set of Big Data technologies that offers a powerful, flexible and fast approach to building a Business Data Lake. This toolset enables companies to store all data, accelerate processing with flexible analytics and most importantly increase the amount of data being analyzed and operationalized within the business. Pivotal delivers these capabilities from long-term experience in the development and implementation of data management and analytical intelligence solutions. Pivotal Big Data Suite includes an unlimited usage of the Pivotal enterprise Hadoop distribution; Pivotal HD. Pivotal Big Data Suite integrates all the essentials for a Business Data Lake architecture: Storage, Analysis and Flexible Architecture.

The Pivotal Big Data Suite stores large amounts of information to create a rich data repository for business needs. Pivotal Big Data Suite enables organizations to store all their data in its native format using Pivotal HD. By storing larger volumes of data, Pivotal Big Data Suite delivers insights on long-term data patterns to help turn today's businesses into data-driven enterprises.



Operationalizing the Buzz: Big Data 2013

Using Pivotal Big Data Suite, organizations can analyze the information stored within the Pivotal HD platform with a wide set of analytical solutions to determine the “integration value” of multiple data sets and types. Today’s Big Data analytics require real-time, interactive and batch capabilities. Pivotal Big Data Suite provides these analytical engines and toolsets for a wide range of users such as data scientists and business analysts.

- **Batch:** All batch needs are delivered with Pivotal HD based on the Apache Hadoop Distribution.
- **Interactive:** Pivotal Greenplum Database uses a shared-nothing, massively parallel processing (MPP) database and flexible column and row orientated storage to deliver an advanced Analytical Data Warehouse (ADW). Simultaneously, HAWQ delivers a high performing SQL query engine over HDFS for interactive query analysis.
- **Real-time:** For real-time analytical and transactional needs, enterprises can extend their environment with in-memory data grid technology from Pivotal GemFire, Pivotal SQLFire and Pivotal GemFire XD.

Build the right thing with a flexible data infrastructure that is designed to deliver a transformative solution to meet an organization’s demanding business needs. Pivotal Big Data Suite, along with the flexible and modern Business Data Lake infrastructure, enables next generation, low-latency, data-intensive applications. Pivotal supports these powerful data management technologies with Spring - Java development framework; and Pivotal CF - platform-as-a-service technology - to accelerate the implementation applications, processing of data and speeding analytical cycles.

Hybrid Data Ecosystem Product Positioning

The Pivotal Big Data Suite is an integrated architecture for Big Data analytics. Pivotal Big Data Suite comprises multi-platforms within the EMA Hybrid Data Ecosystem. These include Hadoop (Pivotal HD), Analytical Platforms (Greenplum DB and GemFire) and Data Discovery (HAWQ). Pivotal also provides an integrated data management layer in the form of Pivotal Data Dispatch to enable data management services, metadata management and data lineage requirements associated with the HDE Information Management layer.

With these platforms working in concert, Pivotal Big Data Suite supports **Exploration, Analytics** and **Operational Analytics** workloads across multiple data latency levels. From real-time processing with the GemFire and GemFireXD products to batch processing with the MapReduce frameworks associated with the Pivotal HD Hadoop distribution, Pivotal allows data consumers from across the organization to manage their workloads at the speed of their business.

EMA Perspective

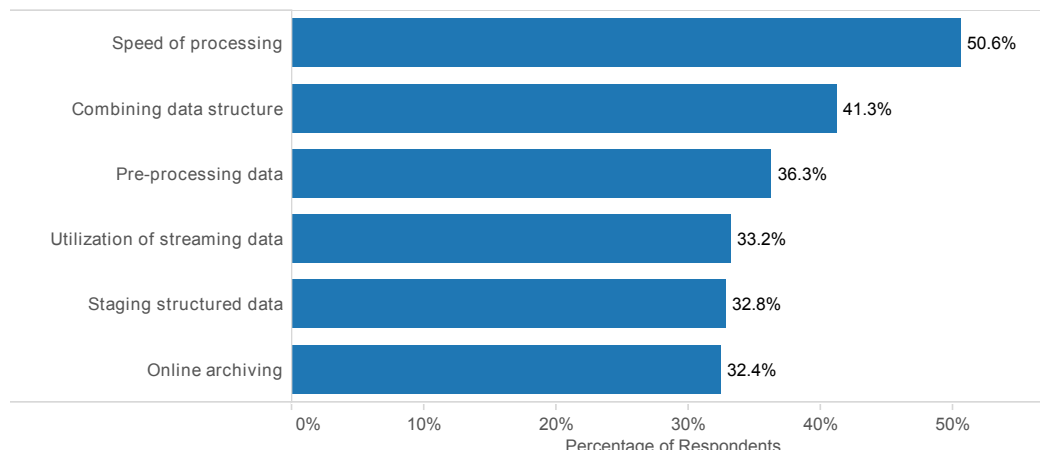
A new level of sophistication has emerged in Big Data over the past two years. Workloads have evolved beyond standard analytics to operational workloads that execute at the speed of the business. A new “art of the possible” is driving innovation and extending the demands on traditional data solutions creating a need for new data strategy. Speed of response is critical when supporting these processes and operational workloads and is creating new value for companies that embrace these new opportunities.

As discussed above companies are embracing Hybrid Data Ecosystem strategies to align data and workload to meet the demands of these new business opportunities and the value that speed can deliver.

Operationalizing the Buzz: Big Data 2013

The leading use case for the 2013 EMA Big Data survey respondents was the **Speed of Processing Response**:

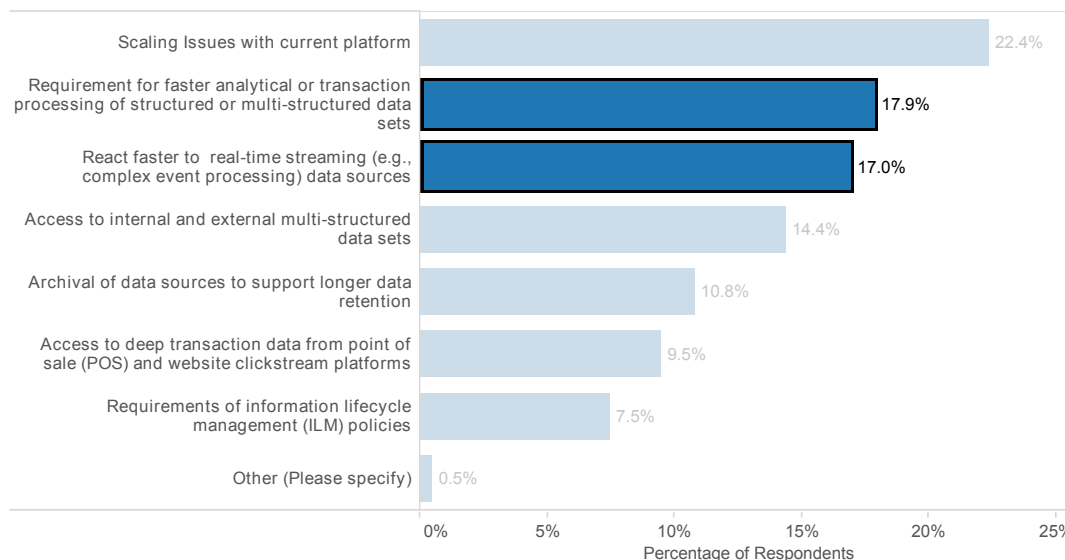
2013 Use Cases



This illustrates the importance of **Response** when considering the requirements associated with Big Data initiatives. It is reflected not only in the use cases of the EMA/9sight survey respondents, but also in how they are implementing their projects. As more initiatives are implemented, organizations are working to deliver critical and sophisticated projects to their internal and external stakeholders. Most of these workloads incorporate multiple data sources that include multi-structured as well as structured data.

There are times when IT departments will strive for technical solutions when the business requirements do not support the effort. With **Response**, this is not the case. While **Scaling Issues with Current Platforms** is the top response from the EMA/9sight panel respondents, the speed of **Response** technical drivers are the second and third highest responses further demonstrating the importance of speed in Big Data projects.

2013 Technical Drivers

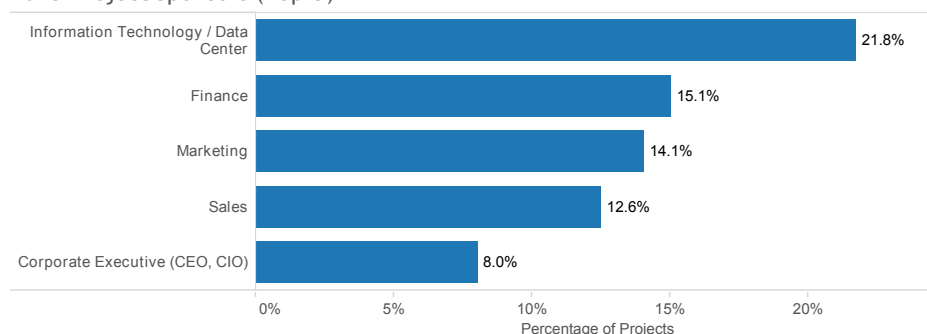


Operationalizing the Buzz: Big Data 2013

Whether it is for faster processing or faster access to streaming data sources, the technical drivers of the 2013 EMA/9sight panel respondents support the Big Data requirement for high rates of **Response** associated with their Big Data initiatives.

All technology projects require sponsorship both in terms of IT support and budget funding. This is no different in the area of Big Data. Big Data projects are not trivial and the majority of projects require the acquisition of new hardware and software infrastructure.

2013 Project Sponsors (Top-5)



In many instances, Big Data project sponsorship comes from departments outside of the office of CIO. This trend is another proof point in the maturity of the market. Diverse areas of the organization are embracing Big Data to solve critical business challenges. **Finance**, **Marketing** and **Sales** account for 41.8% of the sponsorship in the nearly 600 projects analyzed in this research.

The Pivotal Big Data Suite provides the ability for organizations to flexibly meet the challenges of speed of **Response** and data latency along with meeting the implementation and economic expectations of business stakeholders. Pivotal provides both Big Data processing and workloads with an flexible implementation architecture. With these attributes, Pivotal's architecture enables data consumers as a well project sponsors with a framework to implement Big Data workloads as the organization needs them without the constraint of fixed licensing costs and implementation paradigms. Pivotal also allows its customers to utilize a mix of solutions on its platforms without lock in enabling clients to shift priorities without additional costs or lose of original investment.

With these attributes, Pivotal's architecture enables data consumers as a well project sponsors with a framework to implement Big Data workloads as the organization needs them without the constraint of fixed licensing costs and implementation paradigms.

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.

©2014 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
2885.040714