INFORMATICA

WHITE PAPER

Balancing Opportunity and Risk in Big Data

A Survey of Enterprise Priorities and Strategies for Harnessing Big Data

This edition published May 2012

Table of Contents

Executive Summary

As the big data era unfolds, organizations are facing epic increases in the volume, variety, and velocity of data. From growing transactional information to vast reservoirs of social media data, from long-neglected unstructured content to real-time data streams from sensors and devices, big data presents both a business opportunity and an IT challenge for organizations in virtually every industry.

IT professionals and business managers seem to be up for the task. By a more than 2-to-1 margin, organizations view big data principally as a business opportunity rather than an IT challenge, according to a global survey conducted by Informatica in March and April 2012. Designed to assess the state of big data projects and understand big data priorities and strategies, our survey reveals an aggressive approach to the big data phenomenon, with nearly 70 percent or organizations considering, planning, testing, or running big data projects.

While the majority is focused on big transaction data and analytics, the survey shows strong interest in social media information, unstructured content, industry-specific data, and machine-generated information from sensors and devices. Similarly, while most respondents view big data projects as a way to improve operational efficiency and agility, many see opportunities to deliver new products and services and increase customer acquisition and retention by taking advantage of big data.

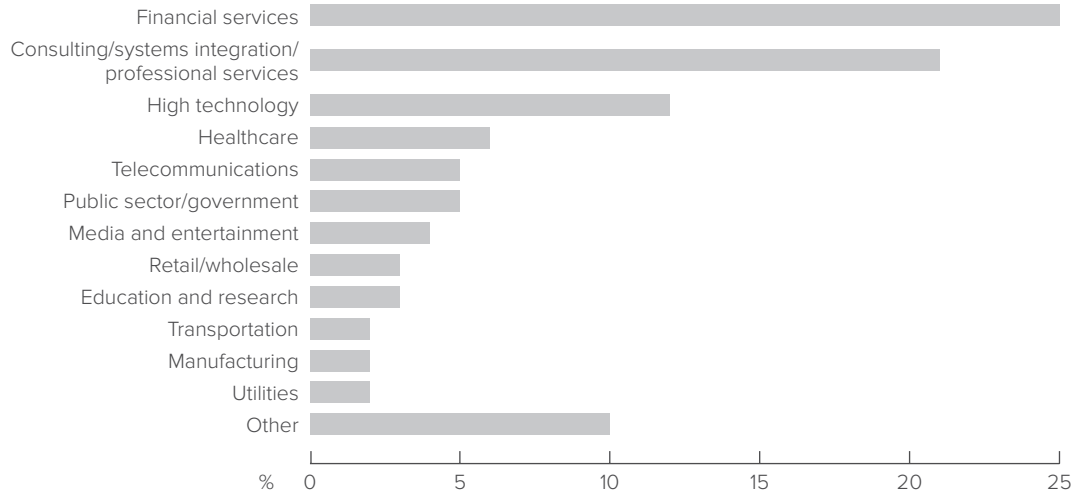
Yet organizations face significant obstacles—particularly a lack of maturity in big data tooling, difficult and time-consuming development, and a scarcity of skilled big data resources. Tellingly, however, just 17 percent of our respondents cited a lack of business justification and return on investment (ROI) as a challenge, making clear that an overwhelming majority of enterprises believe they can gain substantial business advantage by effectively harnessing big data. This report shares the results of our survey and is intended to help you benchmark your strategies, focus areas, challenges, and objectives for big data against the industry at large.

Methodology and Demographics

This report is based on a survey of 589 IT and business professionals in North America, Europe, and the Asia-Pacific region. Answers totaling over 100 percent are based on questions for which respondents were able to select multiple choices.

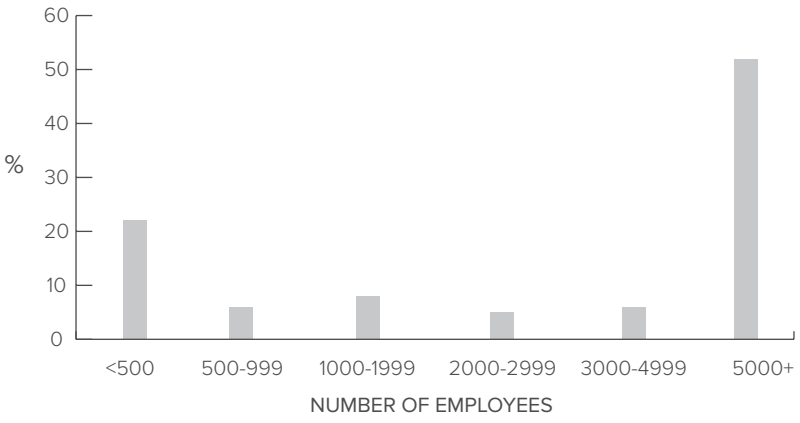
The majority (76 percent) of respondents work in IT, with the remainder in such functions as R&D, finance and procurement, marketing, sales and customer service. The principal responsibilities of our respondents are data warehousing and data integration (38 percent), business intelligence (BI) and reporting (19 percent), consulting (15 percent), applications (15 percent), data quality (4 percent), and C-level executive management (4 percent). Industries most strongly represented are financial services (25 percent), consulting, systems integration and professional services (21 percent) and high technology (12 percent).

Industry representation



While big data is often thought of as being of the greatest interest to larger organizations, our respondents are almost evenly split between organizations with more than 5,000 employees (52 percent) and organizations with fewer than 5,000 employees (48 percent). A sizable percentage—28 percent—work at organizations with fewer than 1,000 employees.

Company size (number of employees)



Most View Big Data as a Business Opportunity

Organizations are bullish on the business opportunities presented by big data. More than two-thirds (67 percent) view big data mostly as an opportunity, while the remaining 33 percent view it mostly as a challenge. This finding parallels optimism expressed by Informatica customers and industry experts at conferences, consultations and on social media. Across both our survey base and the industry at large, most IT and business professionals recognize that big data, still a relatively new phenomenon, introduces significant challenges in fundamental data management, interoperability with traditional data, and security and quality.

However, the majority appears to have eyes on the prize—the opportunities for increased efficiency, newfound insights, new market penetration, and customer engagement that leading organizations are already achieving by harnessing big transaction data (in traditional applications and data warehouses) and big interaction data (from social media, sensors and devices, web clickstreams and other sources).

Big Data: Challenge or Opportunity?

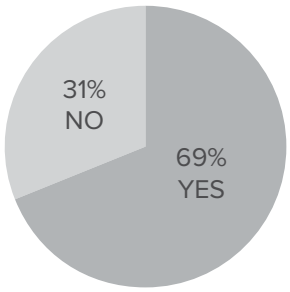
Do you see big data as mostly a challenge or an opportunity for your organization?

Big Data Projects are in Planning or Production at Most Organizations

Considering the relative nascence of big data, organizations are moving fairly aggressively to adapt to large-scale increases in data volumes, variety, and velocity. Nearly 70 percent of our respondents have big data projects in production or testing, or are planning or considering an implementation. On the other hand, 31 percent are on the big data sidelines for now. Some are deterred by the lack of maturity and standards in many big data technologies, while cost pressures have big data projects on the back burner for other enterprises. Based on the large number of organizations exploring big data technologies, we believe the commitment to big data projects will increase as newer big data technologies such as Hadoop, NoSQL databases, and MapReduce mature, and as companies take advantage of data management solutions geared toward integrating traditional and new data types while addressing concerns over data quality and security, and the availability of skilled resources.

State of Big Data Projects

Are you currently considering, planning or implementing big data projects?



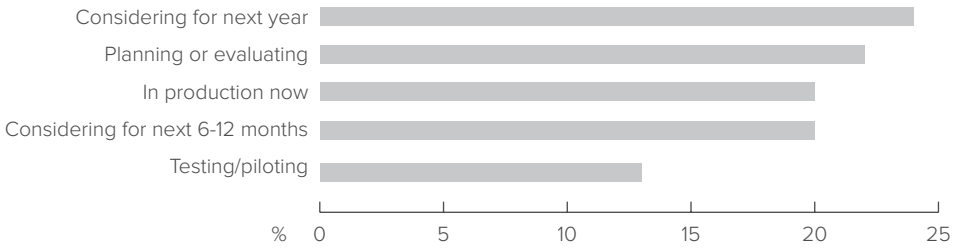
Many Big Data Projects are Already Operational

Among those organizations pursuing big data, 20 percent are already in production and 13 percent are in a test or pilot phase. This means that one-third of the organizations moving ahead with big data are already processing large and often unconventional data sets or are about to, reflecting both an enterprise approach to the big data opportunity and a recognition of the business benefits and competitive advantage that can be gained. Among these leading organizations are Informatica customers such as T-Mobile, which has reduced customer churn by integrating data across a federated architecture, including a Hadoop implementation that supports advanced churn analysis based on call detail record (CDRs), web logs, billing data, social media information, and more.

Adoption will increase as those planning or evaluating big data projects (22 percent) move into test or pilot phases. Another 20 percent are considering tackling big data in the next six to 12 months, with nearly one-quarter considering projects for the next year.

Tackling Big Data

Projects in production, testing, planning, or under consideration



Note: Among those answering "yes" to considering, planning or implementing big data projects

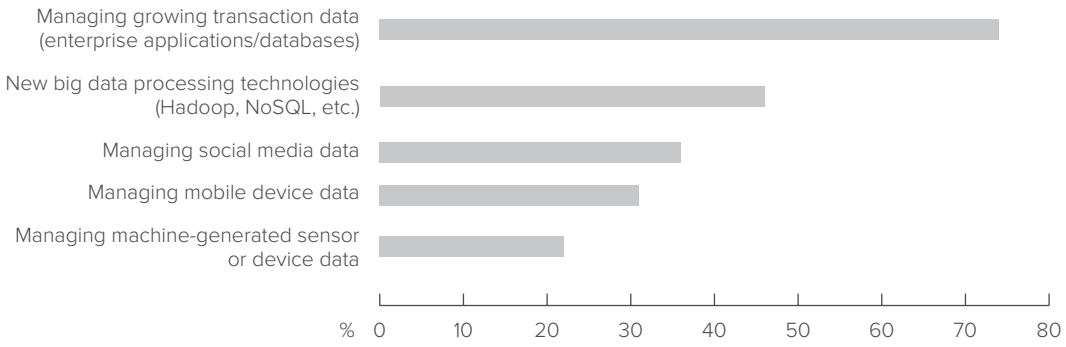
Big Transaction Data Ranks #1 in Relevance

Managing growing volumes of transactional data in enterprise applications and databases is the most relevant aspect of big data for 74 percent of our respondent organizations. This traditional relational data continues to grow steadily from a wide variety of sources as organizations build long-term customer profiles, maintain or archive information for compliance, apply analytics to large data sets for fraud prevention, and move towards hybrid on-premise and cloud application models. These trends are prompting many enterprises to invest in enterprise-grade information infrastructures, strategize cost-effective information lifecycle management approaches, rationalize complex application portfolios, and utilize new data appliances for greater capacity and performance.

Yet our survey reveals strong relevance for emerging big data technologies such as Hadoop and NoSQL databases, and big interaction data from social media, mobile devices, and machine-generated data. Nearly half (46 percent) ranked Hadoop as relevant to their organizations, no doubt far higher than it would have been just one year ago. Thirty-six percent regard social media data as relevant, due to the insights it offers for brand sentiment monitoring, crowd-sourcing, and personalized sales and marketing. Managing mobile device data (31 percent) and machine-generated sensor and device data (22 percent) also generated strong response, as these data types can streamline operations in telecommunications, supply chain, utilities, research and development, and other areas.

Key Big Data Focus Areas

Which of the following aspects of big data are relevant to your organization?



Most Initiatives Pursue Greater Efficiency

Improving operational efficiency is the #1 business driver for big data projects at 74 percent of respondent organizations, with a variety of strategies and focus areas. For instance, Informatica customer US Xpress, a trucking company, saves millions of dollars a year by collecting hundreds of data elements, including fuel, tire, and engine sensors and geospatial systems to optimize fleet usage, reduce idle time, and minimize fuel consumption. Other efficiencies are being gained by integrating disparate big data types in such industries as consumer retail, financial services, healthcare, the public sector, and ecommerce. With efficiency comes greater agility, ranked the #2 business driver at 51 percent.

Our respondents also highlighted introducing new products and services (50 percent) and attracting and retaining customers (49 percent) as key business drivers for big data. These strategies depend in large part on capturing social media and machine data to fine-tune product development and delivery, crowd-source product direction, identify development or manufacturing weaknesses, and tailor brand interactions to unique customer preferences revealed on Facebook, LinkedIn, Twitter, and other social media. Notably, reducing IT costs is a key business driver at 38 percent of organizations, with lower expenses for IT labor, software, and infrastructure achievable through development reusability and cost-effective Hadoop and cloud systems.

Key Business Drivers

What are the business drivers for your big data projects?

Top Challenges Include Lack of Tool Maturity, Real-Time Data, and Quality/Security

Lack of maturity in big data tooling is the top challenge (52 percent) that respondents face in big data projects. Current Hadoop technology, for instance, lacks support for reusability and metadata, making it difficult to extend projects and ensure consistency. Many organizations resort to manual scripting to achieve interoperability between Hadoop and relational systems. Compounded by lack of maturity in technologies such as MapReduce, Hive, and Pig, the fast-changing landscape of big data technologies frustrates efforts to find skilled big data developers (35 percent) and makes development time-consuming and difficult (34 percent).

Lack of support for real-time data was also viewed as a key challenge (39 percent), hindering objectives to utilize real-time steaming data in financial markets, supply chain, ecommerce, and other areas. Unsurprisingly, data quality, security, and privacy were top obstacles (38 percent) that could result in inaccurate data or expose enterprises to a security breach. Interestingly, a lack of business justification and ROI concerned just 17 percent of respondents, making it the least important big data challenge. This indicates that most organizations expect value from big data and have little concern over making a big data business case.

Key Big Data Challenges

What key challenges do you face or foresee in managing big data?

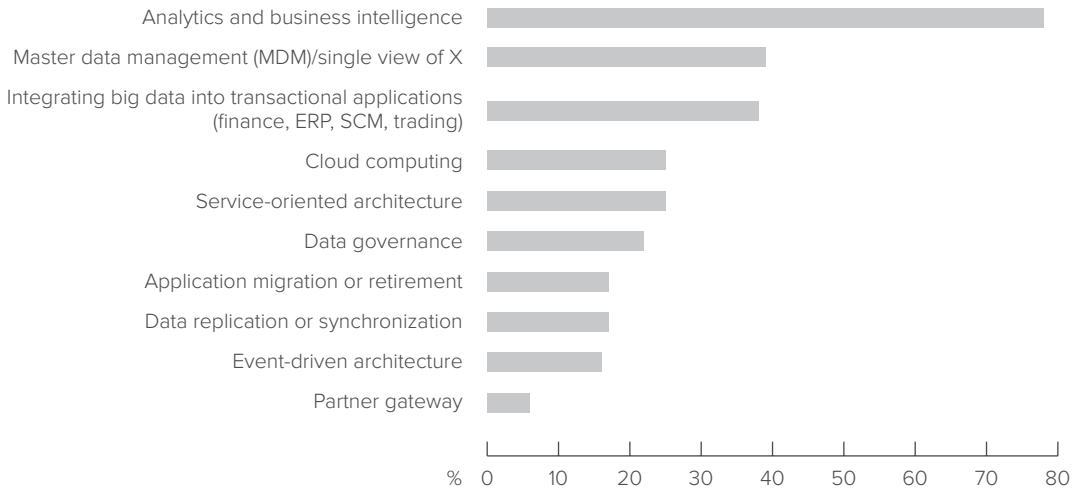
Analytics Leads Big Data IT Projects

By a large margin, analytics and BI are the top IT projects (78 percent) associated with big data. Of course, "analytics" is a broad term that in the big data realm can involve analysis of real-time data for fraud detection, analysis of customer profiles enriched with social media information, or the use of Hadoop as a petabyte-scale virtualized staging area for an enterprise data warehouse. It can also be used for embedded execution of predictive analytics, massive full data scans, and text (semantic) analysis. The top ranking of analytics underscores the value that enterprises put on the groundbreaking insights possible when disparate types of big data are integrated and analyzed.

Ranked second at 39 percent is the use of master data management (MDM) to achieve a single view of customers, products, or suppliers, by combining big transaction and big interaction data. Organizations are also aggressively pursuing the integration of big data into transactional applications for finance, ERP, supply chain, and trading (38 percent), and managing big data in cloud environments (25 percent) to prevent data disconnects between on-premise and cloud systems.

Top Big Data IT Projects

What are the top IT projects associated with your big data projects?



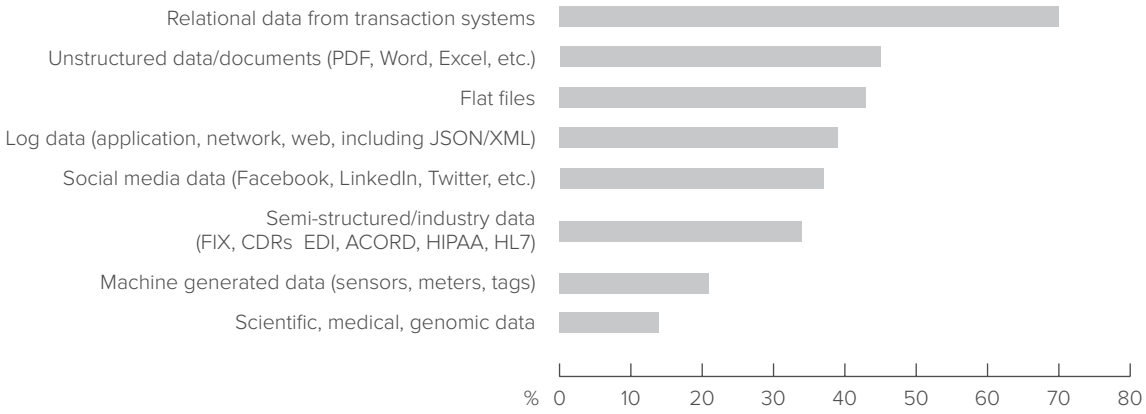
Interest is Strong in Non-Relational Big Data

Relational data from transaction systems is the leading data type for big data projects, cited by 70 percent of respondents. The survey also reveals strong interest in unstructured data (45 percent), flat files (43 percent), log data from application, network and web systems (including JSON/XL) (39 percent), social media data (37 percent), and semi-structured industry data including CDRs and FIX, EDI, ACORD, HIPAA and HL7 data (34 percent).

The findings underscore the broad scope of data that organizations intend to address with big data projects and the perception of value possible from harnessing non-relational data types. The high ranking of unstructured data, including Word, PDF, and Excel documents, suggests that organizations view big data technologies as a way to better manage and analyze non-relational content, estimated in several studies to constitute about 80 percent of enterprise information. Big data technologies also open opportunities for enterprises to leverage industry-specific information in such verticals as telecommunications, financial services, insurance, and healthcare.

Big Data Types

What data types are you interested in addressing in your big data projects?

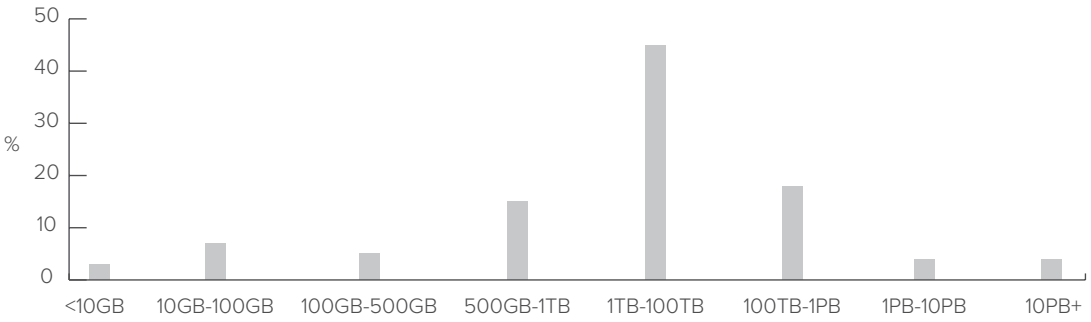


Big Data Volumes: From Terabytes to Petabytes

Big data is commonly thought of as being at least several dozen terabytes, up into petabyte scales. Among our respondents, 26 percent manage big data totaling at least 100TB, with 8 percent dealing with volumes of a petabyte or more. The single greatest share of respondents (45 percent) manage data between 1TB and 100TB, a percentage sure to diminish as data volumes continue to grow. IDC, for instance, predicts that volumes within enterprise data centers will grow 50 times over the next decade.

Big Data Volumes

How much data are you managing in your big data environment?



High Interest in Hadoop and Real-Time Data Streaming

Interest in Hadoop is high—at 42 percent, it's the #1 big data technology being evaluated or considered for deployment. Reflecting the early phase of the big data era, though, just 13 percent of organizations have Hadoop in production or testing (the same percentage as NoSQL databases). Our respondents also have real-time data streaming high on the priority list—at 56 percent, it's the top big data technology being evaluated or in planning/development.

Data warehouse appliances lead the list of big data technologies now in production or testing, at 40 percent of organizations. Other similarly established technologies in relatively broad deployment are data replication (31 percent), data compression or archiving (26 percent), and data parsing (25 percent).

Big Data Technologies

What big data technologies are you using or planning to use?

	NOT CONSIDERING	EVALUATING/ CONSIDERING	IN PLANNING/ DEVELOPMENT	TESTING PHASES	IN PRODUCTION	N/A
Data warehousing appliances	11%	30%	17%	6%	34%	2%
Data replication	17%	27%	17%	5%	26%	8%
Data compression or archiving	16%	34%	16%	6%	20%	8%
Data parsing	20%	30%	17%	8%	17%	8%
Event processing	23%	34%	12%	6%	15%	10%
Real-time data streaming	17%	40%	16%	8%	12%	8%
Data masking/data privacy	20%	35%	16%	7%	12%	9%
Semantic or text analysis/ entity extraction	23%	33%	16%	5%	10%	13%
Hadoop	23%	42%	10%	6%	7%	13%
NoSQL databases	25%	37%	11%	4%	9%	14%
Social media connectors (e.g. Twitter, LinkedIn, Facebook)	26%	36%	14%	6%	5%	13%

Note: Listed from high to low by those technologies currently highest in production or testing phases.

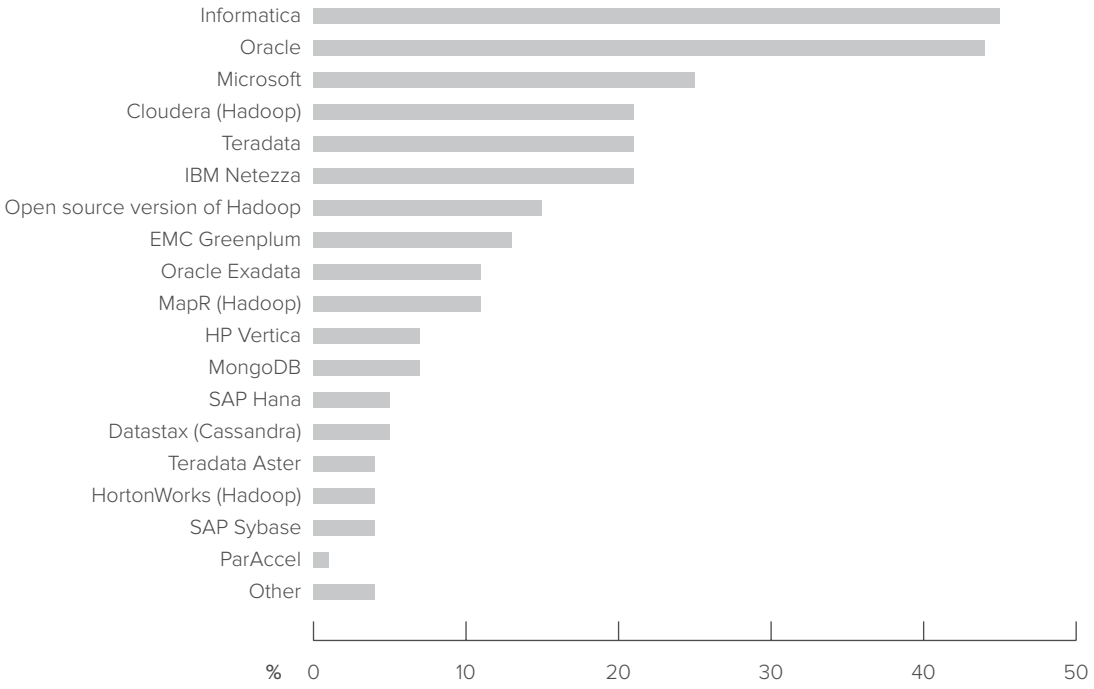
A Changing Landscape of Big Data Technology Vendors

The survey reveals a fairly high degree of fragmentation on the big data vendor landscape, making it likely that some consolidation will occur. Notably, no Hadoop or data warehouse appliance vendor has more than one-quarter market share. Informatica and Oracle lead in big data technology adoption among our respondents, which in part reflects the survey respondent base which included individuals from Informatica's customer database as well as individuals invited to participate via Twitter and on third party Big Data communities.

The key takeaway is that due to the rapidly changing vendor landscape, it is too early to call a winner in big data technologies. So it is prudent for companies to ensure that while they are adopting new technologies to drive innovation, they maintain a strategy for future-proofing themselves with a robust infrastructure that can adapt to the changes quickly.

Big Data Solutions

What technology vendors do you use for big data projects?



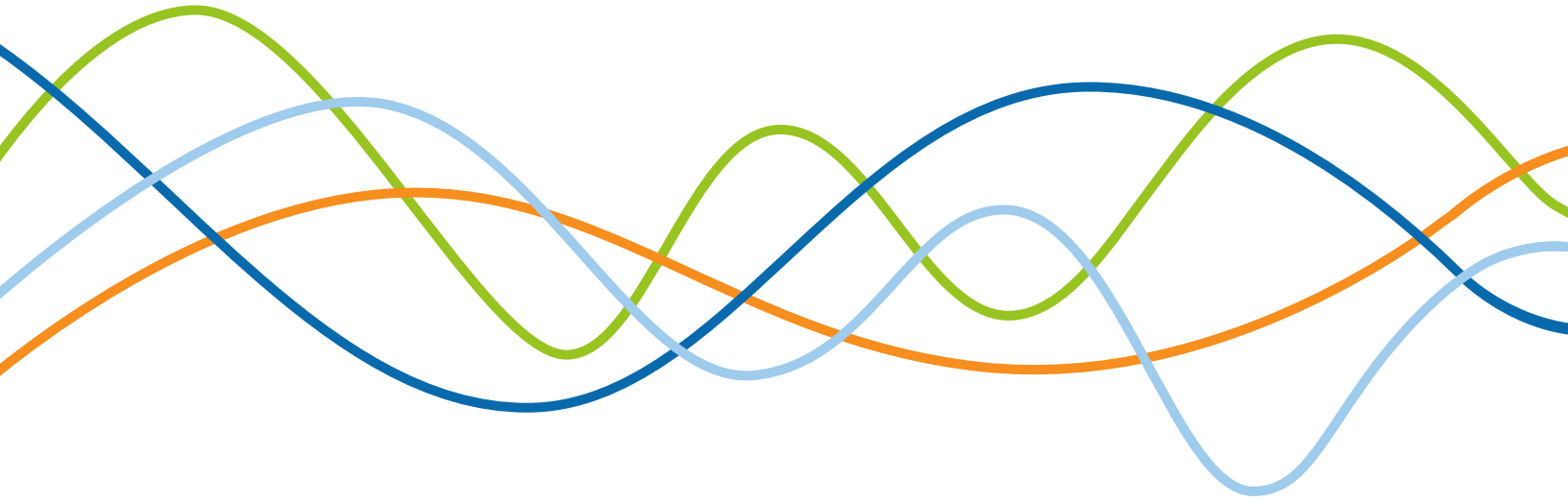
Conclusion

Informatica 9.5, the newest release of the Informatica Platform, is engineered expressly to help organizations maximize return on big data. Its capabilities address key big data challenges identified in this survey, providing a comprehensive, proven, and mature set of technologies that can make your data more relevant, trustworthy, timely, authoritative, actionable, accessible, holistic, and secure. Informatica 9.5 ensures interoperability between the Hadoop stack and the larger enterprise IT environment. Capabilities including real-time data streaming, data quality and privacy, social data integration, smart data partitioning, Hadoop data parsing, and any-to-any data transformations fills a void in emerging big data technologies, enabling you to avoid costly and time-consuming manual approaches. Finally, the widely used Informatica Platform positions you to leverage existing skillsets and assets for your big data projects, rather than hunting for high-priced talent with specialized skills.

Leading organizations are already putting Informatica technology to work to increase the business value of data, lower the cost of data, and realize the promise of big data. With Informatica, big data becomes less of a challenge and more of a business opportunity to gain a competitive advantage.

ABOUT INFORMATICA

Informatica Corporation (NASDAQ: INFA) is the world's number one independent provider of data integration software. Organizations around the world rely on Informatica for maximizing return on data to drive their top business imperatives. Worldwide, over 4,630 enterprises depend on Informatica to fully leverage their information assets residing on-premise, in the Cloud and across social networks.



INFORMATICA®

Worldwide Headquarters, 100 Cardinal Way, Redwood City, CA 94063, USA
phone: 650.385.5000 fax: 650.385.5500 toll-free in the US: 1.800.653.3871
informatica.com linkedin.com/company/informatica twitter.com/InformaticaCorp

IN09_0512_01943