GIGAOM RESEARCH

How enterprises will use the cloud for big data analytics

Lynn Langit

November 10, 2014

This report is underwritten by Cazena.

TABLE OF CONTENTS

Executive summary

Enterprise IT infrastructure largely predates the emergence of cloud computing as a viable choice for hosting data-driven applications. Large organizations are now showing real signs of adopting cloud computing for certain applications and a few forward-thinking enterprises are formulating the concept of Data as a Service (DaaS) based on performing big-data analytics in the cloud. However, exactly when big-data analytics will move to the cloud remains an open question.

Compatibility, security, and performance concerns have kept enterprise organizations from being completely comfortable with the idea of moving their complex core applications to the cloud. Without a seamless application migration blueprint, the project can seem more like a headache – and a risk – than it's worth. This report, which is based on a survey that Gigaom Research and Cazena sponsored in September 2014, will demonstrate the different considerations when moving some or all of the DaaS-type applications to the cloud. The report is will give guidance to CxOs, IT and business leaders, and decision-makers at Software as a Service (SaaS) companies and cloud service providers.

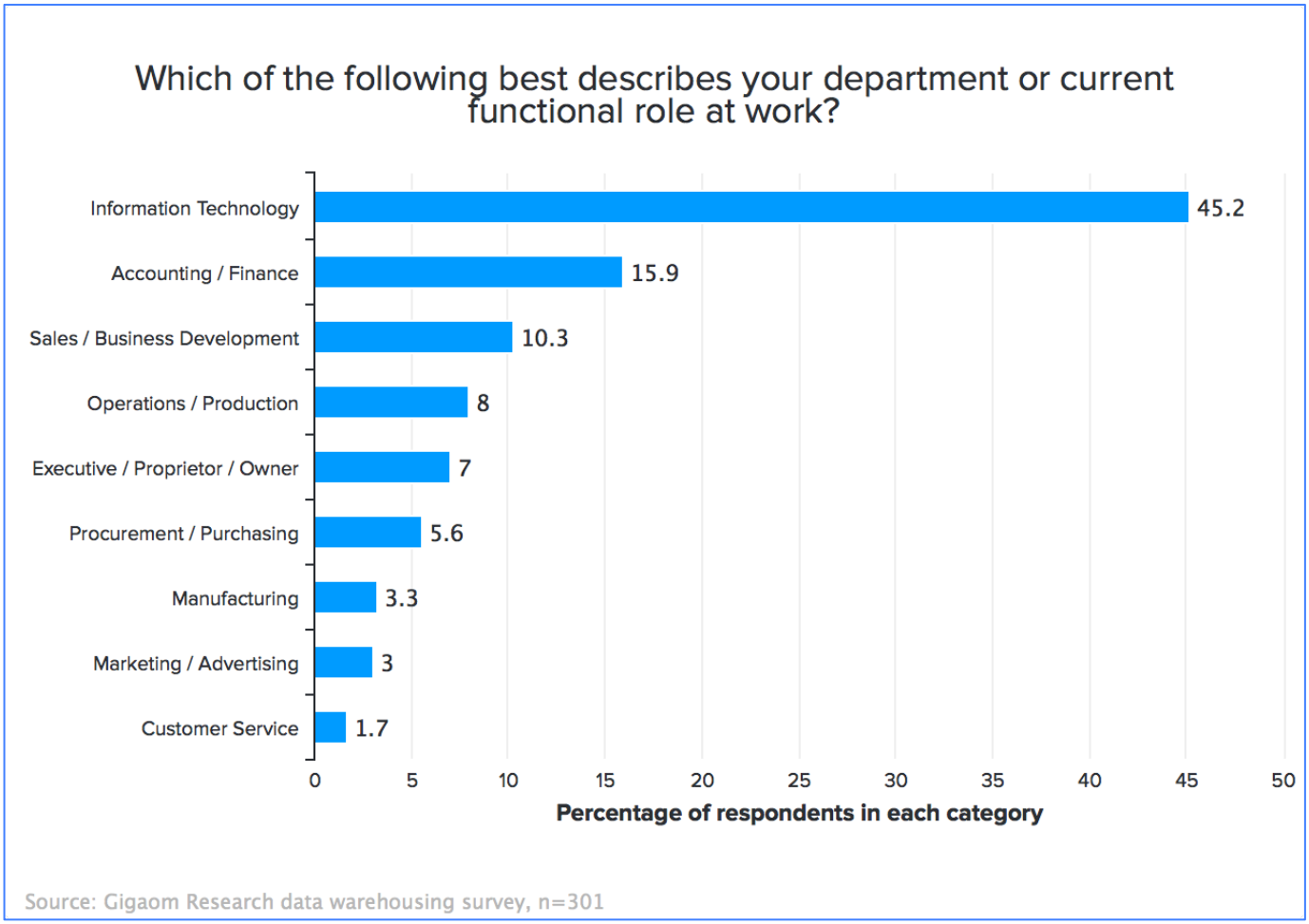
- Executives with an eye to costs are motivated to explore moving their data-intensive applications to the cloud. What are the primary drivers?
- What types of cloud solutions are most prevalent now – public, private, government, or hybrid clouds?
- What types of big-data workloads are best suited to the reality of what the vendors are providing currently?
- How significant is the maturity of cloud-tooling in general, and of tooling around data-analytics projects specifically to adoption?
- What are the primary blockers? Understanding security, privacy, and compliance needs are key components for ensuring that big-data analytics projects can be reliably migrated or newly hosted on the public cloud, but the complexities of vendor offers and of the current regulatory landscape are significant. Does the answer lie in education, legislation, or something else?

A majority of enterprises are interested in public cloud

Survey respondents for this study were primary decision-makers or influencers across a spectrum of industries, but most heavily representing the finance, technical, manufacturing, health care, and retail sectors. The majority is "somewhat" or "very" familiar with analytics, data warehousing, or big-data projects.

Only 45 percent of the respondents who are familiar with analytic concepts perform an IT function in their organizations. This statistic suggests that interacting with data at a detailed level has moved from being centralized in IT to a much wider community and throughout most organizations.

The majority of data interactions come from outside of IT



(Source: Gigaom Research)

A majority of the survey respondents (53 percent) reported that they "are already" or "planning to" leverage public cloud resources as part of their enterprise big-data analytic needs. Survey participants reported that their greatest motivation is either the potential for cost reduction (49 percent) or increased agility (30 percent). Only 13 percent of respondents had no intention of using public cloud resources; however, even within this group, 44 percent would reassess their positions if they had a deeper understanding of the cloud's security posture.

Of those verticals already leveraging public cloud resources, those most often in the cloud are: manufacturing (27 percent), technology, computing, consumer electronics (19 percent), or finance (14 percent). Those replying that they "plan on leveraging public cloud resources but haven't started yet" are most often from three verticals:

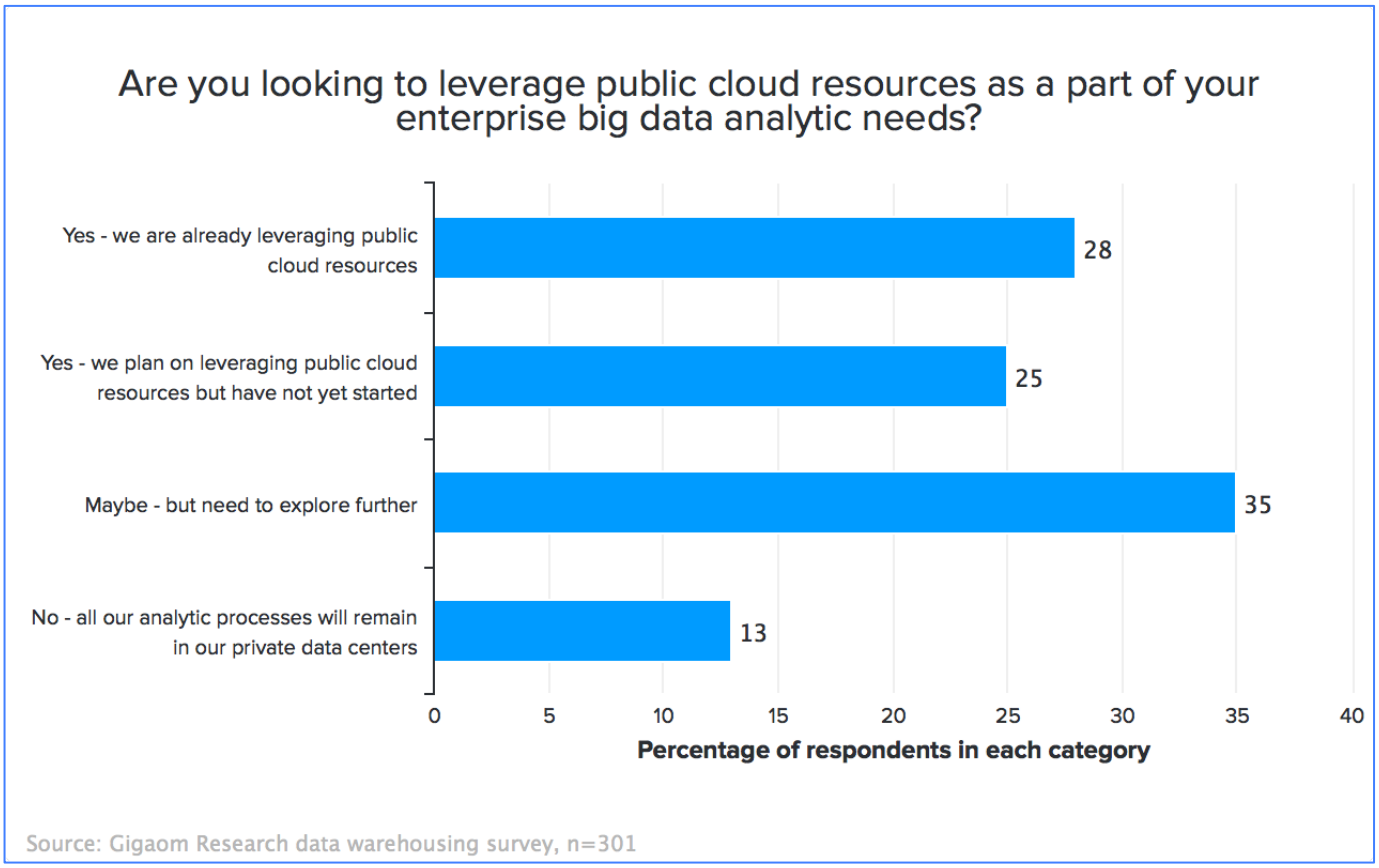
1. Technology, computing, consumer electronics (26 percent)
2. Retail, wholesale (15 percent)
3. Finance, insurance, real estate (15 percent)

The top three responding "maybe, but (we) need to explore further (before moving to the public cloud)":

1. Finance, insurance, real estate verticals (18 percent)
2. Manufacturing (15 percent)
3. Technology, computing, consumer electronics (12 percent)

These numbers strongly suggest significant interest in assessing how public cloud resources can be utilized as a part of enterprise analytic processes.

The enterprise is looking to leverage the public cloud



(Source: Gigaom Research)

Data workloads and tools

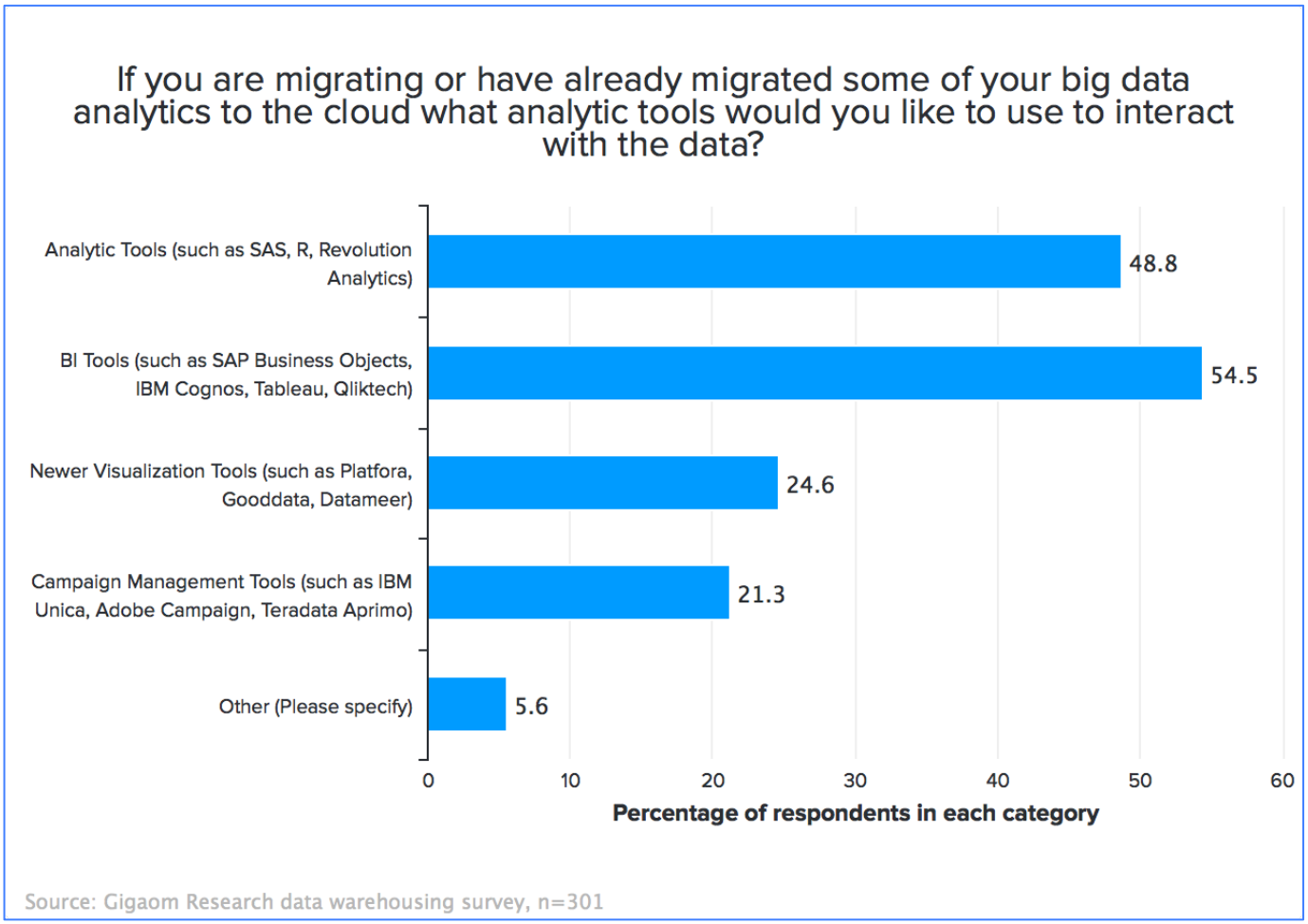
The survey participants looking to leverage the cloud for big-data analytics intend to move large volumes of data to the cloud, with 92 percent of survey participants wanting to move more than a terabyte and 20 percent wanting to move more than 100 terabytes. The preferred starting point for enterprise big-data projects migrating to the cloud are full or partial migrations of non-mission critical applications (65 percent).

The types of data-intensive workloads being considered for cloud-migrations most often are analytics-real-time operational workloads (43 percent) or processing, such as batch processing with SQL (43 percent). Sandboxes for data discovery are in third place (33 percent).

Standard BI tools (55 percent) and analytic tools (49 percent) are the preferred tooling types for analytics projects that have been migrated to the cloud. The preference for using familiar tools is seen in this

category. Consistent with the general tone of the results in this survey, enterprises value minimal disruption and prefer reusing skills, tools, and techniques whenever possible.

Familiar analytics tools are preferred



(Source: Gigaom Research)

A large group (53 percent) is analyzing CRM data (such as SalesForce.com), which falls into the category of external data for analytic workloads. Other significant external data sources are site analytics and log files (40 percent) and social media (26 percent) with CPG, services, and non-profit organizations preferring social media. Finance, insurance, real estate, government, and healthcare prefer site analytics.

Security, privacy, and complexity are top enterprise concerns

A key concern raised repeatedly in the survey is the perception of security practices that cloud vendors currently in the market practice. As shown below, 63 percent indicated "security considerations" as a key blocker for any potential cloud-based implementation. Among the issues raised were:

1) **A lack of understanding of security processes being used by cloud vendors in general**. A majority of the respondents (55 percent) replied that "better understanding of the security posture of the cloud" would cause them to reassess their current strategy of hesitating or not planning to implement any analytics processes to the cloud.

2) **A lack of trust of one or more cloud vendors.** Quotes from the survey results include "*need more security in cloud computing,*" "*the cloud is still too dangerous to put PII data on it,*" and "*our company's data is very sensitive and the arena of cloud computing needs a longer track record of secure functionality before we will consider using it.*" A total of 33 percent indicated that they "don't trust the providing vendor in general."

3) **A lack of understanding of compliance and regulatory implementations for various verticals** (i.e., healthcare and government) and via various standards (e.g., HIPAA and SOX). Another 35 percent replied that a lack of existing industry certifications is a significant blocker to cloud adoption. However, 39 percent see that cloud enabling regulatory changes are "imminent" or are "coming in the next 1-2 years."

   The industry certifications most often cited as useful are:

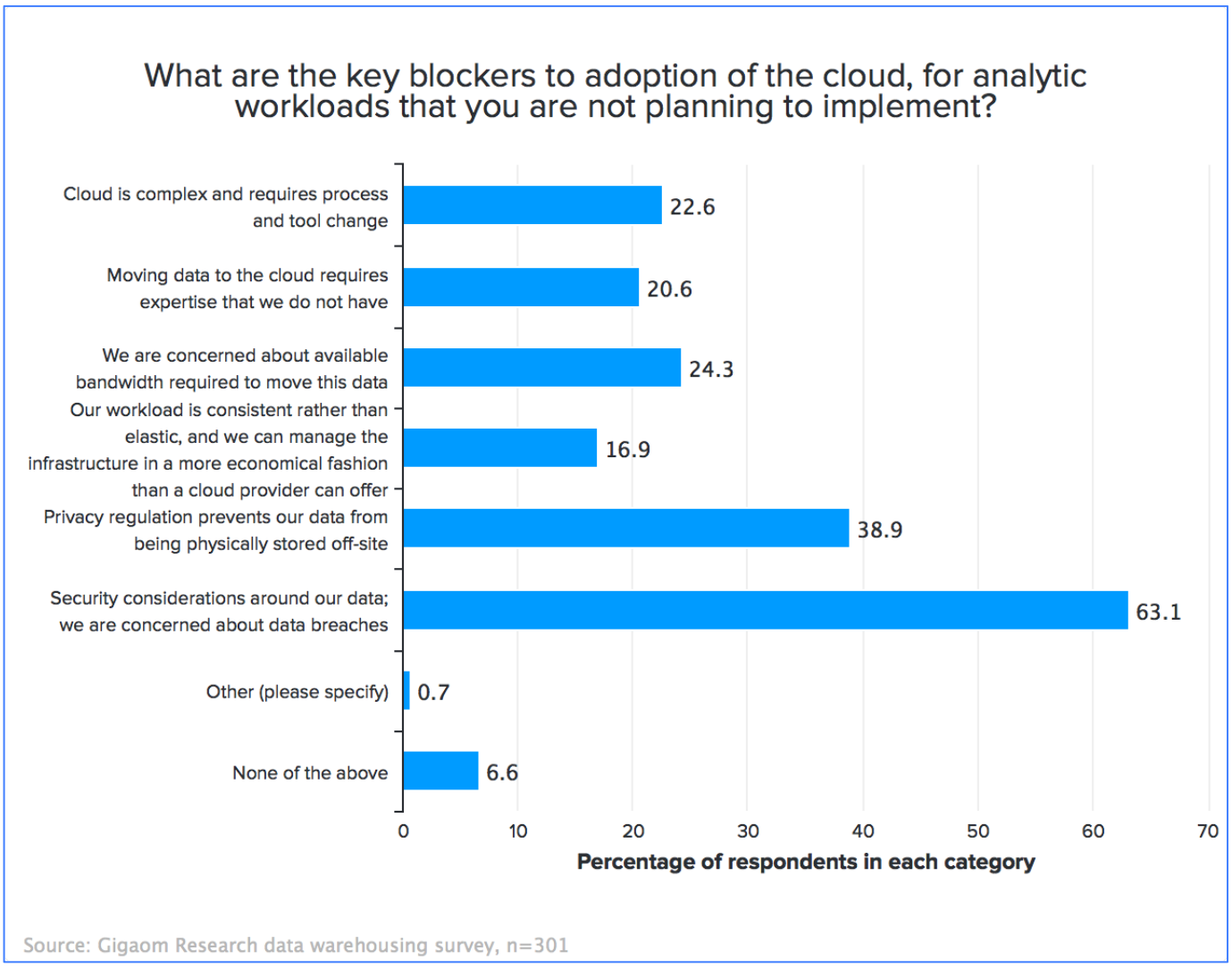
   a. SOX (35 percent)

   b. HIPAA (34 percent)

   c. ISO 2700 (26 percent)

The following figure illustrates the reported blockers to cloud adoption.

Key blockers to public cloud adoption



(Source: Gigaom Research)

Given the number of responses indicating strong hesitation to cloud migration because of a lack of understanding about the various vendors' approaches to security, a natural question is whether education could help mitigate this blocker. Survey results support this theory. More than half (55 percent) responded to the question "if you are hesitating or not planning to implement any analytic processes to the cloud what event would result in you reassessing your strategy" with "better understanding of the security posture of the cloud." A segment of this group (23 percent) responded that a service provider would be most valuable if it had "expertise in our specific area of concern, i.e. security, compliance."

This response points to the importance of any cloud service provider clearly explaining security practices and the value to potential customers of independent security certifications. Also, respondents appear to view vendors, who provide guidance about regulatory compliance for specific verticals, e.g., health-care and financial, as providing more value to potential customers. This fact can also be seen in the certifications being requested for the cloud. Even though 34 percent of respondents thought that HIPAA was crucial for their business, only 9 percent were from healthcare organizations. There is a need for HIPAA outside healthcare, in large HR departments, but no survey respondents identified them as part of an HR function.

Complexity is a key blocker to the cloud

The survey indicates that complexity is a key blocker to the cloud. Inherently the move to the cloud represents friction for the enterprise. Complexity is characterized in a number of ways:

- The movement of data to and from the cloud requires a different set of technologies and techniques for it to be efficient and reliable. The secondary set of concerns focused on this complexity with 24 percent of survey respondents having concerns about the available bandwidth to move their data onto a cloud platform. This concern is most likely related to the volume that organizations would like to move to the cloud, which is substantially more than 1 TB.

- Process and tool changes resulting from migrating big-data analytic workloads to the cloud concerned 23 percent of survey respondents.

- Of the respondents, 21 percent indicated that they currently do not have the expertise to move data to the cloud. Enterprises have significant investments in existing processes and tools, so any viable solution must fit into existing frameworks rather than cause significant disruption.

Key takeaways

- The majority of the survey respondents are either in or interested in sending some of their workloads to the public cloud. Only 13 percent responded "no public cloud." Nevertheless, the respondents have significant concerns about potential migrations. They have significant interest in specialty clouds (such as government, vertical-specific, etc.) as well as in hybrid clouds.

- The key concern is transparency about public-cloud vendor security and compliance practices. SOX and HIPAA are mentioned most often, but with an expressed need for more standards and understanding of how cloud vendors support these standards. Respondents appear to have a large amount of fear and confusion in this area and would welcome education about standards.

- Another concern is complexity. Cloud is a new architecture and impacts existing processes. Enterprises would like to work with solutions that have minimal impact to their existing processes that enable existing workloads and tools.

- The survey respondents distinctly conveyed a conservative response, particularly in concerns about security and trust, as well as a preference for existing tools that would need to function with this new infrastructure.

- Workloads being considered for the cloud center around analytics. Other considerations are maturity of tools, size of data (1 TB seems to be the magic starting point), and data sources. CRM data and log data appear at the top of the list for cloud projects that include an analytic focus. Respondents seem to have a general lack of imagination about potential data sources other than those listed.

About the author

Lynn Langit is an analyst for Gigaom Research and a big data and cloud architect, who has been working with database solutions for more than 15 years. Over the past three years, she has been working as an independent architect using these technologies, mostly in the biotech, education, manufacturing, and facilities verticals. She has done POCs and has helped teams build solutions on the AWS, Azure, Google, and Rackspace clouds. She worked with SQL Server, MySQL, AWS Redshift, AWS MapReduce, Cloudera Hadoop, MongoDB, Neo4j, and many other database systems. In addition to building solutions, Lynn partners with all major cloud vendors, providing early technical feedback into their big data and cloud offerings. She is a Google Developer Expert (Cloud), Microsoft MVP (SQL Server), and a MongoDB Master. Lynn is also a Cloudera-certified instructor for MapReduce Programming.

Prior to re-entering the consulting world three years ago, Lynn was a Microsoft Certified instructor for more than ten years, a Microsoft vendor, and a Microsoft employee for four years. She has published three books on SQL Server business intelligence and has most recently worked with the SQL Azure team at Microsoft. She continues to write and screencast and hosts a big-data channel on YouTube with more than 150 different technical videos on cloud and big-data topics. Lynn is also a committee member for several open source projects.

About Gigaom Research

Gigaom Research gives you insider access to expert industry insights on emerging markets. Focused on delivering highly relevant and timely research to the people who need it most, our analysis, reports, and original research come from the most respected voices in the industry. Whether you're beginning to learn about a new market or are an industry insider, Gigaom Research addresses the need for relevant, illuminating insights into the industry's most dynamic markets.

Visit us at: research.gigaom.com.