InformationWeek
:: reports

Reports.InformationWeek.com

July 2013 $99

Next

The NSA and Big Data: What IT Can Learn

Ever heard of Accumulo? CIOs can leverage this and other tools Big Brother uses to analyze online activities. Here's how.

By Kurt Marko

Report ID: S7250713

TABLE OF CONTENTS

ABOUT US

InformationWeek Reports' analysts arm business technology decision-makers with real-world perspective based on qualitative and quantitative research, business and technology assessment and planning tools, and adoption best practices gleaned from experience.

OUR STAFF

Lorna Garey, content director; lorna.garey@ubm.com
Heather Vallis, managing editor, research; heather.vallis@ubm.com
Elizabeth Chodak, copy chief; elizabeth.chodak@ubm.com
Tara DeFilippo, associate art director; tara.defilippo@ubm.com

Find all of our reports at reports.informationweek.com.

Kurt Marko
InformationWeek Reports

Kurt Marko is an *InformationWeek* and *Network Computing* contributor and IT industry veteran, pursuing his passion for communications after a varied career that has spanned virtually the entire high-tech food chain from chips to systems. Upon graduating from Stanford University with a BS and MS in electrical engineering, Kurt spent several years as a semiconductor device physicist, doing process design, modeling and testing. He then joined AT&T Bell Laboratories as a memory chip designer and CAD and simulation developer.

Moving to Hewlett-Packard, Kurt started in the laser printer R&D lab doing electrophotography development, for which he earned a patent, but his love of computers eventually led him to join HP's nascent technical IT group. He spent 15 years as an IT engineer and was a lead architect for several enterprise-wide infrastructure projects at HP, including the Windows domain infrastructure, remote access service, Exchange email infrastructure and managed Web services.

EXECUTIVE SUMMARY

Recent revelations about the NSA's cyber-spying capabilities are an IT version of sausage-making: Everyone knew something was going on, but it all seemed much more disturbing once the messy details were exposed. While there's plenty for privacy advocates and civil libertarians to grouse about, the technologies underpinning the NSA's data collection and analysis programs provide much for IT pros to cheer. The agency has built a scalable, extensible, secure big data system that rivals, and in some cases exceeds, anything cloud heavyweights like Google, Amazon or Apple have deployed.

We'll examine the open source and commercial technologies the agency uses in its big data collection and analysis infrastructure, diving into details about Accumulo, the BigTable- and Hadoop-based system it developed and later open sourced, and discuss promising applications for the enterprise. These cyber swords in the war on terror can indeed be bent into IT plowshares in the quest for business success.

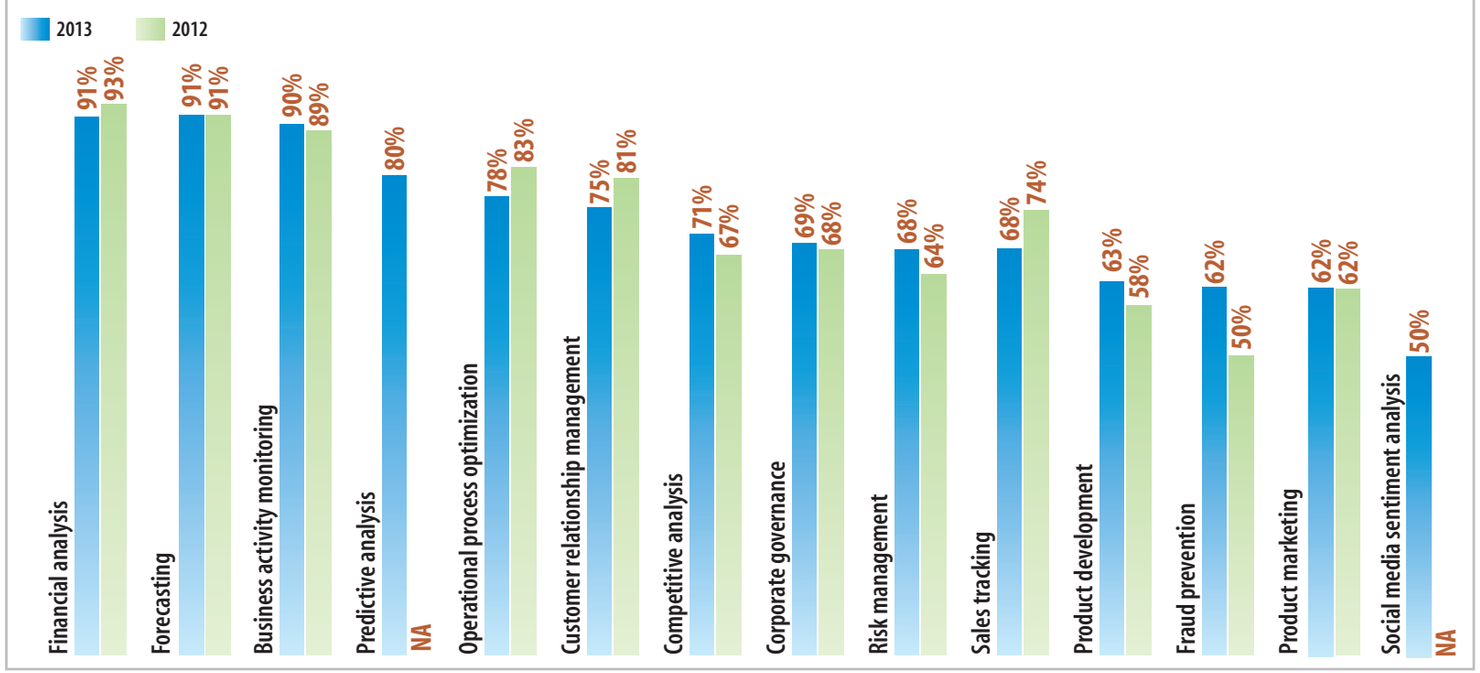
Technology vs. Privacy Smackdown

It's not the first time bleeding-edge technology has collided with constitutional law and public policy, and it won't be the last. The National Security Agency's electronic eavesdropping apparatus is one of those increasingly common situations where breakneck technological progress runs into enduring questions of personal rights and governmental powers. By now we're all familiar with the basics of the government data collection and analysis programs that were exposed by former contract employee Edward Snowden, whose subsequent odyssey to escape capture and prosecution by U.S. authorities is an ongoing saga. But the slides Snowden released describing the programs are anything but fiction. Two previously unknown systems, PRISM, which gathers online communications from the likes of Microsoft, Yahoo, Google and Facebook, and Blarney, a program to vacuum up metadata about every wireless call, text message or data

Figure 1

Utilization of Analytics and BI: 2013 vs. 2012

How do you currently utilize or plan to utilize analytics or business intelligence?



Note: Percentages reflect current or planned use

R6061112/6

Base: 417 respondents in October 2012 and 414 in October 2011 at organizations using or planning to deploy data analytics, BI or statistical analysis software

Data: *InformationWeek* Analytics, Business Intelligence and Information Management Survey of business technology professionals

connection from the major wireless carriers, have generated nonstop debate from Capitol Hill to the office watercooler.

What's gotten less attention, even among technologists, is what's underpinning these programs — technology that's largely based on open source software, commercially available hardware and private cloud infrastructure.

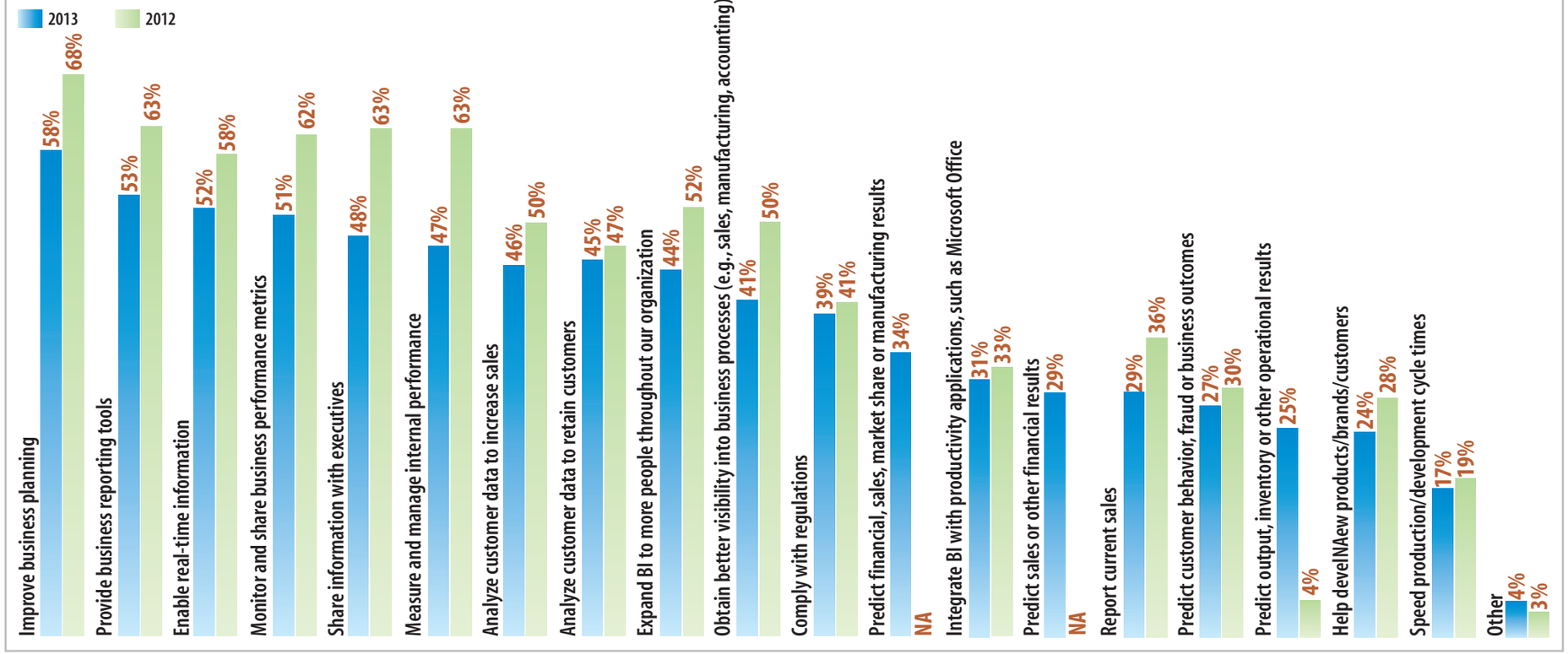
It turns out the NSA is using advanced but freely available tools to detect plots and connect terrorist networks. Taking a closer look at its portfolio reveals a number of promising opportunities for enterprises looking to connect a different set of dots, to better identify potential customers, target existing ones with more effective promotions, spot fraud or cybercrime in its early stages, or improve products and services by mixing lots of little data points into a useful mosaic.

PRISM, Blarney and the cloud data centers used to power them are part of a long-term, concerted strategy to redesign the NSA's IT infrastructure, which we've covered in some depth. As some of that archi-

Figure 2

Implementation Goals

What are your organization's current goals for implementing analytics or BI products?



Note: Multiple responses allowed
Base: 417 respondents in October 2012 and 414 in October 2011 at organizations using or planning to deploy data analytics, BI or statistical analysis software
Data: *InformationWeek* Analytics, Business Intelligence and Information Management Survey of business technology professionals

R6061112/8

tecture is made public, including key components that were developed by the NSA and subsequently put in the public domain, we're getting a much clearer picture of how the agency built its eavesdropping apparatus. Let's take a closer look at some of the IT building blocks and how they might be applied to the enterprise.

The Technology

As NSA CIO Lonny Anderson surveyed the technology landscape several years ago, he decided he wanted what Google had: a massive, distributed data collection and analysis system built on a new generation of big data software and commodity hardware, the IT equivalent of "e pluribus unum" — out of many, one. But unlike the cipher-cracking supercomputers and custom hardware the agency is known for, this new generation of cloud-scale software is mostly in the public domain running on commodity hardware. Harvard Business School IT professor Tom Davenport, talking about the NSA's data analysis prowess in a *Wall Street Journal* report, says, "They've substantially reduced the cost and greatly increased the [government's] ability to analyze this type of data," adding that the supporting technology has become "orders of magnitude" less expensive.

The pillars of the NSA's architecture are big data systems, particularly a new distributed data store called Accumulo, machine learning and natural language processing software, and scale-out cloud hardware. Let's look at each.

Step 1: Big Data Software

Like Google, the NSA needed to store, process and analyze incredible amounts of both structured and unstructured data, the former consisting of things such as call logs and email address headers, the latter being actual content. In fact, one estimate claims the government's intelligence organizations have the goal of eventually handling yottabytes of data. For the math-challenged, that's 1 billion petabytes, or the contents of more than 250 billion 4 TB drives, an impossible feat with SQL Server. That's where various big data systems using NoSQL data stores, like Cassandra, HBase, MongoDB and Google's own BigTable, come in. The NSA liked the BigTable architecture, which Google published in 2006, but saw that it lacked some necessary features, which the agency proceeded to add.

But BigTable provides only the data store; you still need a way to process, analyze and draw correlations across large, distributed data sets. Again, Google set the standard with its MapReduce programming model, but here the NSA had a readily available, free and open alternative, Hadoop. Hadoop adopts the MapReduce paradigm of splitting large data analysis problems into chunks that can be run across a multitude of servers using a distributed file system like BigTable or HDFS (Hadoop File System). Although there are a number of commercial Hadoop implementations, including Cloudera, Hortonworks and WANdisco, the NSA rolled the open source code into its own big data analysis system, called Accumulo.

The NSA's requirements, while extreme in scale, aren't unique in capability. Our *InformationWeek* 2013 Analytics, Business

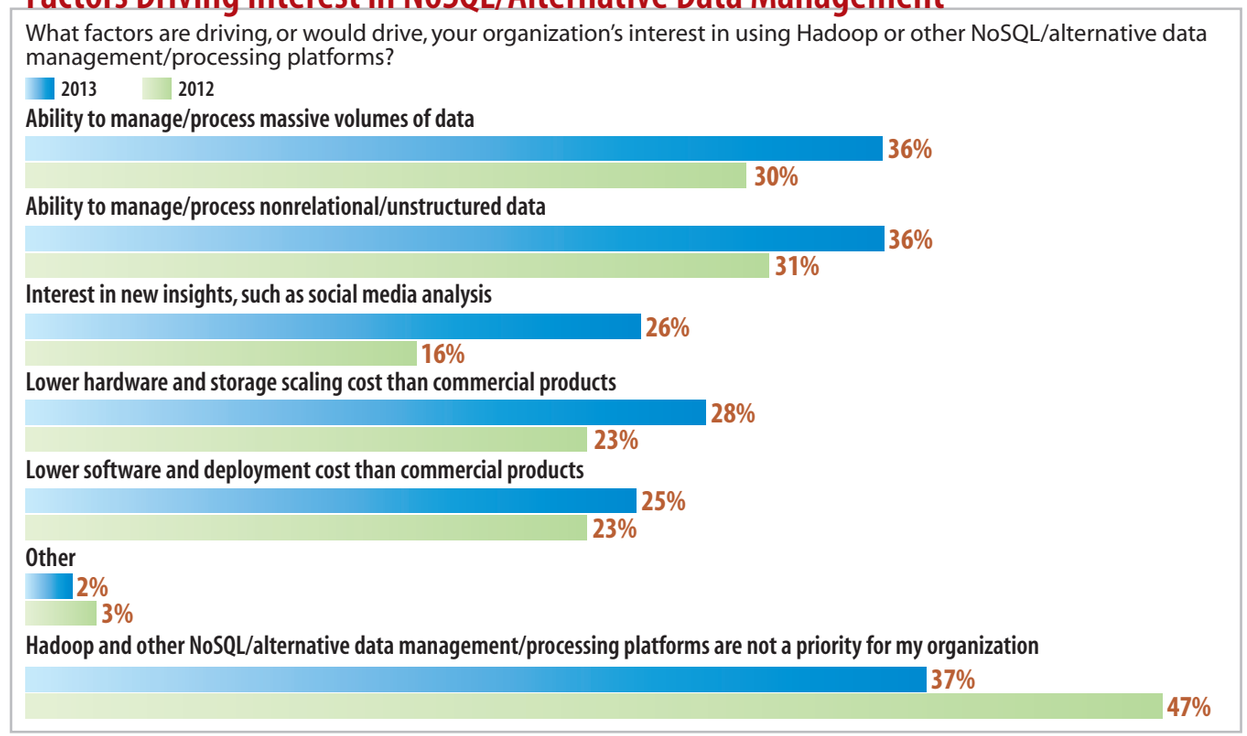
Intelligence, and Information Management Survey shows that the need to manage and process massive data volumes containing unstructured information, just the sort of problems the NSA was trying to solve, is most likely to spark interest in nonrelational databases like NoSQL and Hadoop.

Accumulo is the secret sauce behind the NSA's data collection and analysis systems. It's a reverse-engineered version of Google's BigTable distributed data store, with extensions, that's bundled with Hadoop, Apache ZooKeeper (a distributed application service and process management system) and Apache Thrift (a software library and set of code-generation tools used for back-end services that enables efficient and reliable communication across programming languages, more details here [PDF]). Accumulo is essentially a big data storage and application platform in a box, with some interesting new wrinkles.

Ely Kahn, co-founder and VP of business development of Sqrrl, a startup he and several former NSA developers launched with the goal of commercializing Accumulo, says the NSA had two primary requirements in a big data system that weren't being met by existing publicly available technology: granular data access control and massive scalability. So it's not surprising that the NSA's major exten-

Figure 3

Factors Driving Interest in NoSQL/Alternative Data Management



Note: Multiple responses allowed R6061112/21
Base: 517 respondents in October 2012 and 431 in October 2011 involved with information management technologies
Data: *InformationWeek* Analytics, Business Intelligence and Information Management Survey of business technology professionals

sion to the canonical, Google-like big data architecture is focused on security, which in Accumulo means new security attributes for individual data elements. According to Kahn, the system needed to tag every piece of data with a security label, which would dictate who can access particular data regardless of the application. This means the system does not return query results unless the user performing a search has access credentials that meet the aggregated union of each cell's "visibility" parameter.

This is a huge advantage over conventional big data architectures since it allows information with widely varying security requirements and access controls to be stored in a single system. Kahn says such "fine-grain tagging of data" enabled the NSA to aggregate information from multiple sources, expose it all to a wide population and not worry about individuals accessing something they shouldn't. "It solves the stovepiping problem," he says, describing the government's traditional means of controlling access by partitioning information into unique databases for different user groups and security clearances.

Accumulo is also eminently scalable. Kahn says people using HBase, a popular column-oriented NoSQL database that runs on top of HDFS, find it difficult to scale past a few hundred storage nodes. In contrast, he says, Accumulo instances with thousands of nodes have been in production for several years. Assuming each node is a typical 2U system with eight to 12 drive bays, thousands of nodes translates into tens if not hundreds of petabytes.

Accumulo also augments traditional NoSQL big data systems with a server-side programming facility that allows data to be modified as it flows through the system, thus enabling real-time data analysis, something not currently possible with Hadoop systems. Kahn likens these "iterators" (in Accumulo-speak) to user-defined functions in SQL or Excel. In big data parlance, "it's like a continuous MapReduce function," he says. It's a powerful feature for tasks like filtering, counting, categorizing and labeling data as it's being ingested into the system. Furthermore, Kahn says that since iterators operate in memory, they avoid performance-sapping disk reads and writes. While conventional Hadoop systems can perform such data analysis post hoc, iterators enable these sorts of MapReduce functions to happen continuously as new data enters the system.

From the Shadow Factory to the Public Square

Sqrrl owes its existence to the fact that the NSA didn't stop with just improving on BigTable and Hadoop — it open sourced it. As *InformationWeek* reported at the time, "in its submission to the Apache Foundation, the NSA said that it expects that verticals that have a keen interest in privacy, such as the government and healthcare providers, may find the most use for Accumulo. Other features include a storage format that improves compression. [The] NSA is releasing the project through Apache, rather than directly, because it is heavily reliant on other Apache technologies."

The strategy worked. Kahn says his motivation in founding Sqrrl was to take the expert-

ise acquired in building Accumulo and applications using the platform for the NSA and make its simpler to use, easier manage, more powerful and generally a better application development platform.

Our surveys indicate that the NSA and its Sqrrl progeny have identified an important need among big data practitioners. When we asked respondents to our *InformationWeek* 2013 Big Data Survey about the areas they were most concerned about, security came out on top, followed by data management and access speed. With security and access control being Accumulo's major selling points — indeed, its raison d'être — when paired with the system's massive scalability, it's clear the NSA built something with significant business value.

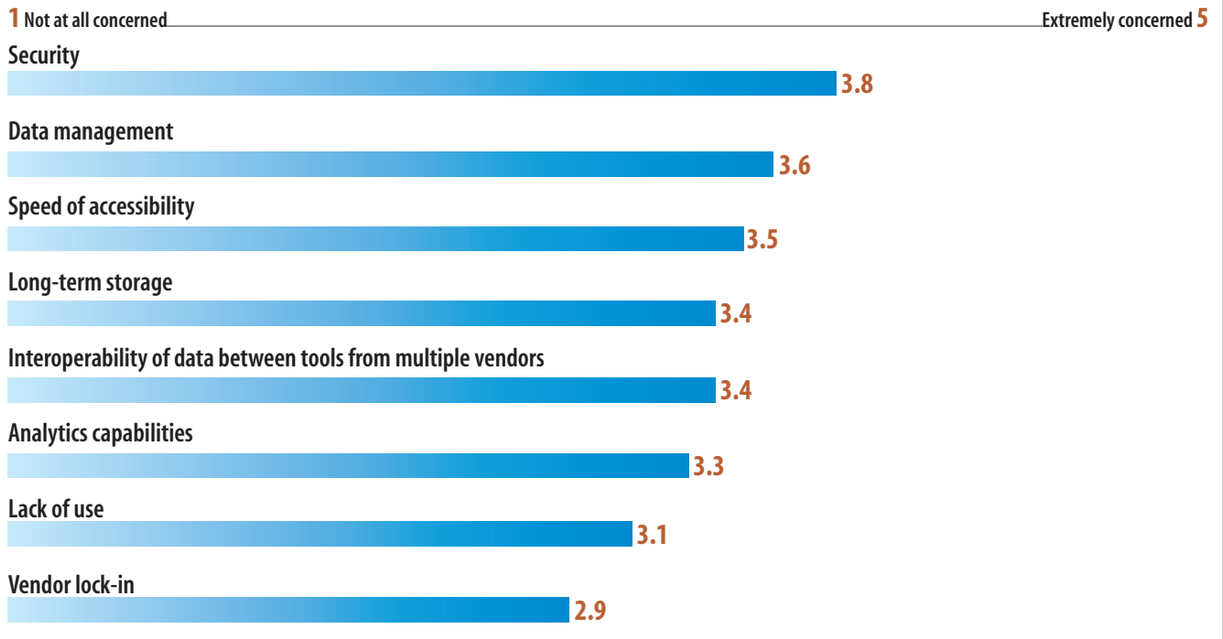
Step 2: Machine Learning and Natural Language Processing

Another important piece of the NSA's technology arsenal is machine learning, namely building adaptive, self-tuning systems that gradually, and automatically, evaluate incoming data to improve performance, update search queries, interpret ambiguous phrases or identify objects in digital images. Much like Google improves your search results by recording metadata about past searches, clicked links and Gmail document text, the NSA's systems can apply context to textual analysis to determine whether the phrase "this plot will destroy

Figure 4

Concern With Big Data-Related Issues

How concerned are you with the following issues as they relate to big data? Please use a scale of 1 to 5, where 1 is "not at all concerned" and 5 is "extremely concerned."



Note: Mean average ratings R5561012/12

Data: *InformationWeek* 2013 Big Data Survey of 257 business technology professionals at organizations with 50 or more employees, September 201

him" refers to an assassination scheme or a critic's opinion that a lousy novel will ruin the author's reputation for suspense.

For example, one public NSA project, KODA, can automatically create summaries of large sets of textual data using just the text itself. As the agency's public disclosure points out, current search engines and content management systems typically use dictionaries or standard text samples to search and categorize data; however, that approach breaks down when applied to large volumes of information. KODA (PDF) "works by measuring the similarity between passages of text and selects sentences it can use to summarize the text. KODA performs well on large data sets in many languages regardless of formatting. It can read thousands of documents and publish a summary of user-definable length for each document."

Other publicly disclosed projects demonstrating the agency's interest in adaptive systems include software (Renoir) that can analyze large data sets and create visualizations and associations that identify relevant and interesting relationships, a program that can extract text from color images, and another that can generate a topic description and categorization of an arbitrary text sample and automatically apply these machine-identified categories to search, sort and catalog new text by topic.

FAST FACT

26%

of respondents to our 2013 Big Data Survey say customer service is among the top two business areas most in need of improvement via data analysis.

Step 3: Hardware and Networks

Just as the NSA re-created the big data analysis features of a Google and Facebook in software, so too is it replicating their data storage and processing capabilities in hardware, notably in a new cloud-scale data center under construction in Utah. The fact that we know so much about this data center is likely attributable to the NSA's choice of contractor; because the U.S. Army Corps of Engineers is the builder, many design documents are in the public domain. They describe a massive Tier 3 facility comprising 100,000 square feet of raised floor space in four data halls, 900,000 square feet of technical support and administrative space, two substations supplying 65 megawatts to the data halls (which translates to roughly 20 kW per rack), all at a price of between $1.5 billion and $2 billion.

Impressive, particularly the cost, which is about an order of magnitude higher per square foot than other cloud facilities But it's hardly unprecedented given that mega-centers, like Facebook's Prineville, Ore., facility, routinely top out with more than 300,000 square feet of data hall space. Added security likely accounts for some of the price premium, as does the fact that the facility is on an old Army base and so could have required extra preparation and probably extra fiber pulls since it's not exactly on the info-superhighway interstate.

Fortunately for shops without such deep pockets, as we point out in *InformationWeek's* latest State of the Data Center report, such state-of-the-art facilities are readily available to any enterprise through a rich and competitive market for co-lo space. So whether your cloud spans 20 servers or 20,000, there's no trouble finding it a suitable home.

Enterprise Use Cases

Given what we know about the NSA's IT capabilities, CIOs have a couple of ways of look-

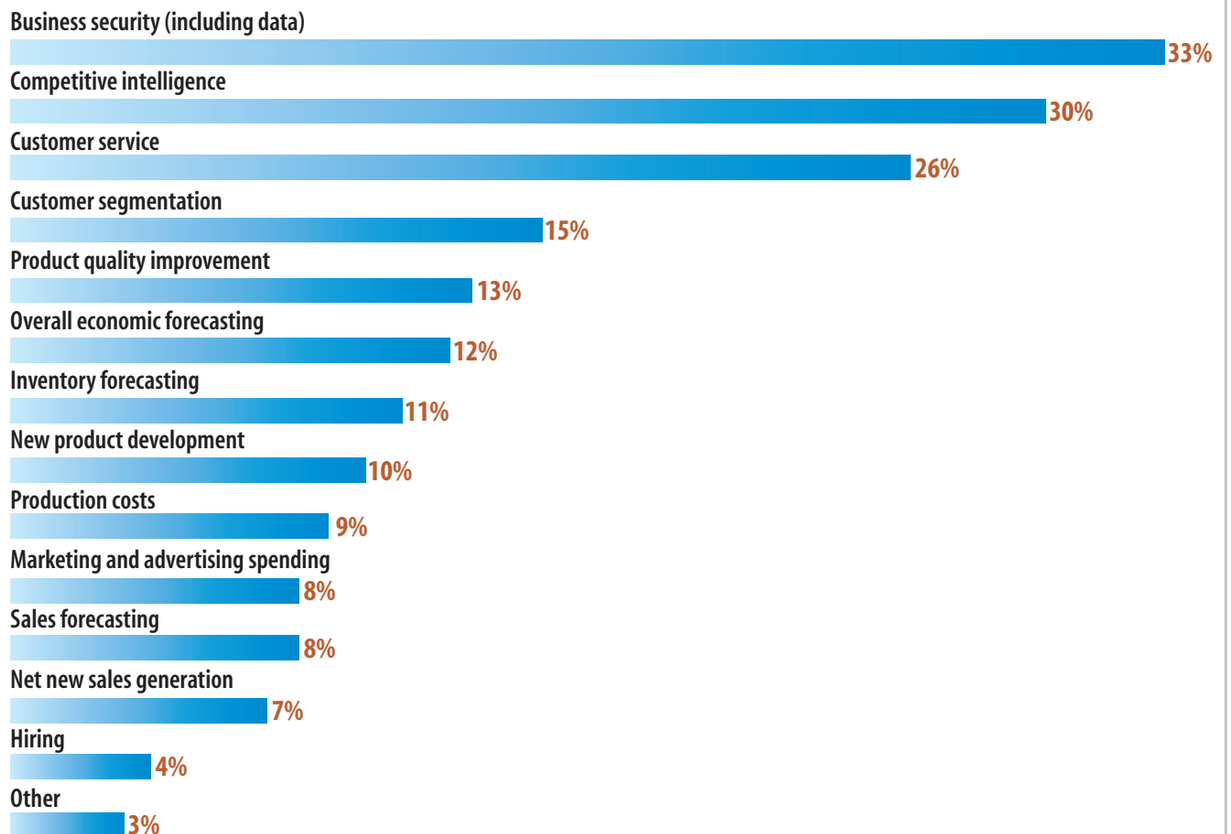
ing at the opportunity. Fundamentally, it's just another big data system capable of the same sorts of applications that have become regular features on the business pages. Whether it's a brick-and-mortar retailer tracking customer movements throughout its store to learn about shopping and buying habits, or Netflix using detailed viewer preferences to determine which new shows to license and thus be assured that *House of Cards* would be a hit before it ever aired, big data systems are the foundation for a growing number of innovative applications. On this front, many organizations already have a good idea of what they need to do to capitalize on today's proliferation of data and resulting storage, management and analysis technologies.

Our big data survey found that most data analysis projects target some combination of business security; competitive and financial benefit (through customer segmentation, better sales and supply forecasting, competitive analysis, and improved quality); improved business processes; and higher customer engagement, retention, satisfaction and loyalty — all

Figure 5

Business Areas Most in Need of Improvement

What are the top two business areas most ripe for improvement via better data analysis at your organization?



Note: Two responses allowed R5561012/4

Data: *InformationWeek* 2013 Big Data Survey of 257 business technology professionals at organizations with 50 or more employees, September 2012

areas that contribute to the bottom line.

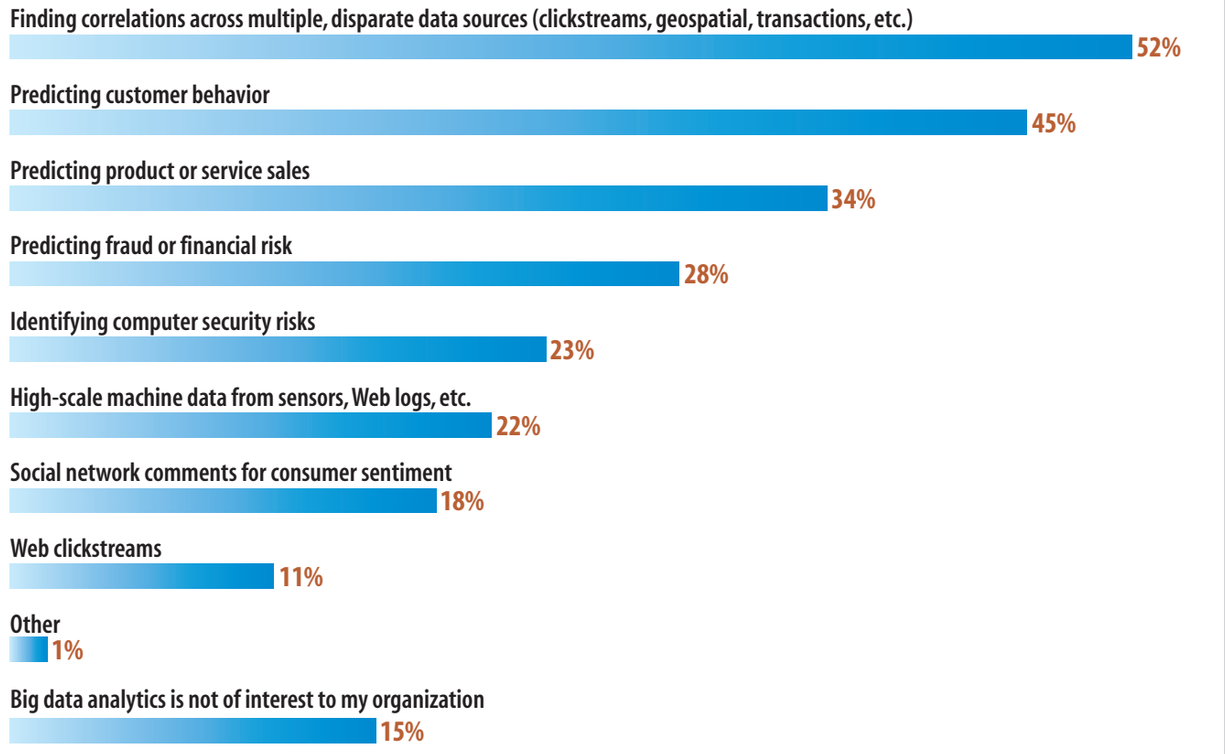
When it comes to big data applications and analysis of the sort done by the NSA, IT pros responding to our BI survey are most intrigued about the ability to find correlations across multiple data sets, using them to predict customer behavior, product sales, or the likelihood of fraudulent or financially risky activity.

So to the degree the NSA has implemented just another big data collection infrastructure and analysis platform of the type widely used in large enterprises, its technology is relevant to any of these applications. However, the NSA advances the state of the art in a few ways that are well suited to a several types of enterprise applications. Indeed, one respondent to our big data survey, an enterprise data architect for a large healthcare provider, hits the same set of challenges the NSA faced when developing Accumulo: security and data governance: "There is a lot of opportunity for us with big data, especially in getting analytics from unstructured data, but the governance, security and IT management aspects of it represent a culture change which we are not

Figure 6

Factors Driving Interest in Big Data Analysis

What data sources or challenges are driving, or would drive, your organization's interest in using big data analysis?



Note: Multiple responses allowed
Base: 417 respondents at organizations using or planning to deploy data analytics, BI or statistical analysis software
Data: *InformationWeek* 2013 Analytics, Business Intelligence and Information Management Survey of 541 business technology professionals, October 2012

R6061112/13

likely ready for."

In fact, of Sqrrl's three primary customer segments, Kahn says one is Hadoop users who want to add security and real-time analytics to their big data systems, and another is the healthcare business (cybersecurity being the third). We hope our commenter is watching.

The following are some application categories where NSA tech really shines:

>> **Real time versus batch analytics:** Accumulo's server-side programming features enable many data analysis tasks to operate continuously, thus providing real-time or interactive access to analytics features, something not possible with Hadoop-based systems. As a Sqrrl white paper puts it, Accumulo enables applications that "bridge the gap between Online Transaction Processing (OLTP) and Online Analytical Processing (OLAP) by being able to process hundreds of thousands of transactions in real time and conducting real-time analytics on those transactions using petabytes of historical data." For example, Kahn says that Sqrrl Enterprise, its Accumulo-based product, uses iterators to filter and count data in real time and in memory with virtually no performance hit.

>> **Data mining and graph analysis:** Perhaps the primary application of the PRISM program is exposing social connections and communication paths that aren't apparent when analyzing millions of separate actors. In fact, graph analysis is the application the NSA used earlier this year (PDF) to showcase Accumulo's speed and scalability, describing a system running a standard graph benchmark with 4.4 trillion nodes and 70 trillion edges. For comparison, Facebook's Graph Search currently scales to hundreds of billions of nodes and trillions of edges. We're impressed.

Aside from scalability, the advantages to doing data mining and graph analysis in a system like Accumulo are twofold. There's the speed factor of the server-side processing enabling IT to pre-cache certain common operations, and there's the security inherent in allowing only authorized individuals access to certain queries. For example, the leaked PRISM flowchart slides show that search queries pass through several layers of filtering and authorization prior to any results being returned.

>>**Predictive analytics:** Using vast quantities of data from multiple sources — sales transactions, customer loyalty cards, economic data, Twitter comments, even the weather — to model and predict events from demand for specific products to a customer's receptiveness to individual promotions is an analysis problem where the more data and the faster the processing the better. Again, the scalability and granular access control of the NSA's technology make it an ideal platform.

Some specific applications are garnering the interest of Accumulo early adopters, including:

>> **Cybersecurity:** PRISM-style log file, transaction and content analysis can identify and prevent malicious or fraudulent activity. Kahn says big data security analytics is an emerging trend, since traditional security information and event management and security forensics software can't handle the volume, diversity and complexity of data needed to identify new threats or predict incidents. For example, he says a financial institution might generate terabytes of security data from hundreds of

FAST FACT

80%

of respondents to our 2013 Analytics, Business Intelligence and Information Management Survey say they plan to use predictive analysis.

platforms per day and want to store years' worth in order to draw long-term correlations. From this petabyte-scale pool, they need to perform low-latency searches, getting results within a of couple seconds. But since the repository includes sensitive data mixed in with the mundane, Kahn says firms also need Accumulo's security features and API so that users get only what they're authorized to see.

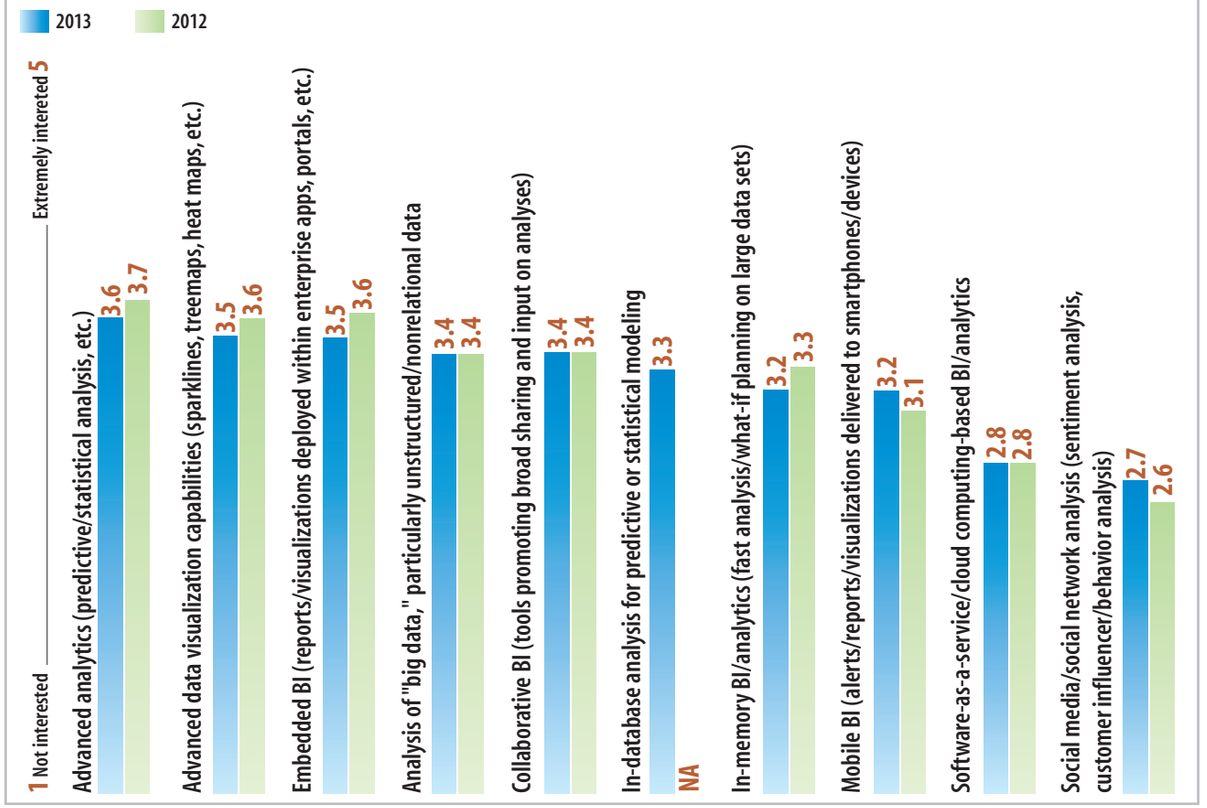
>> Healthcare: To get a consistent picture of an individual's health profile, organizations need to consolidate records from multiple systems: administrative, clinical, research/lab, pharmaceutical. Here, Accumulo's cell-level security can provide the necessary access controls for specific information based on user credentials, satisfying HIPAA. Furthermore, Kahn says that by consolidating data, it's much easier for clinicians to find subtle correlations across large patient groups to aid in diagnosis and treatment.

>> The Internet of things: Accumulo's scalability and real-time filtering/categorizing features make it an ideal data repository for telemetry from millions of physical objects, with applications allowing interactive data

Figure 7

Interest in BI Technologies

Please rate the level of interest within your organization in the following leading-edge analytics or BI technologies. Please use a scale of 1 to 5, where 1 is "not interested" and 5 is "extremely interested."



Note: Mean average ratings R6061112/11
Base: 417 respondents in October 2012 and 414 in October 2011 at organizations using or planning to deploy data analytics, BI or statistical analysis software
Data: *InformationWeek* Analytics, Business Intelligence and Information Management Survey of business technology professionals

mining across object types and customers. Although not yet a significant application category, we expect early IoT adopters will find Accumulo an attractive data repository and application platform.

>> Dynamic pricing and order management: Collecting and analyzing competitive pricing information and generating real-time sales and inventory trends is a big data application with big-dollar potential. Such information allows businesses to adjust prices and supply chains in response to changing conditions. Here too, Accumulo's scalability, speed and security make it a good choice. For example, a retail chain might want to consolidate all sales and inventory data into a single system but allow a local store manager to see trends only for that store, not the entire company. According to our surveys, predictive, statistical analytics and data visualization top the list of interesting capabilities of big data BI systems.

Conclusion and Recommendations

Unlike the NSA, we know enterprise IT doesn't have a multibillion-dollar big data and private cloud budget, but there are steps you can take to capitalize on its largess.

1. Develop a big data and BI strategy: It's long past time to build larger, more diverse data sets than your departmental Access or SQL Server databases and use tools more sophisticated than Excel to monitor and analyze what's going on in your organization. Yet, according to our BI survey, the most commonly used analytics tools are the most primitive: spreadsheets and formatted reports. Fewer than half of respondents use predictive analytics.

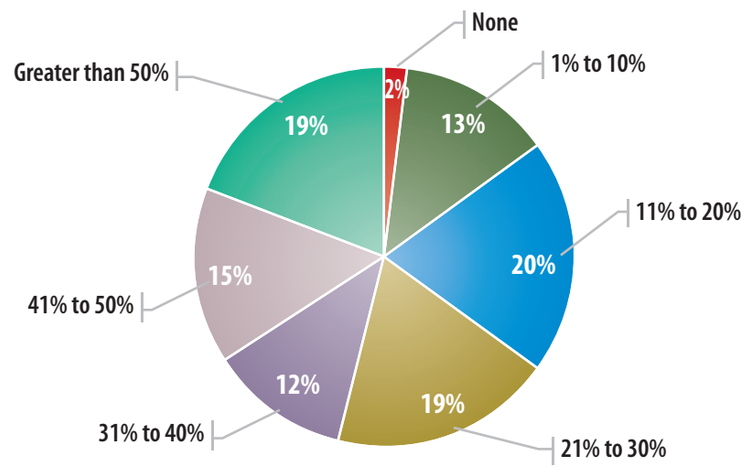
Memo to IT: It's 2013, not 1993. Get with the program.

2. Big data needs big security: Don't short-

Figure 8

Percentage of Data Analysis Done With Excel

What percentage of your data analysis is done with Excel?



R5561012/16

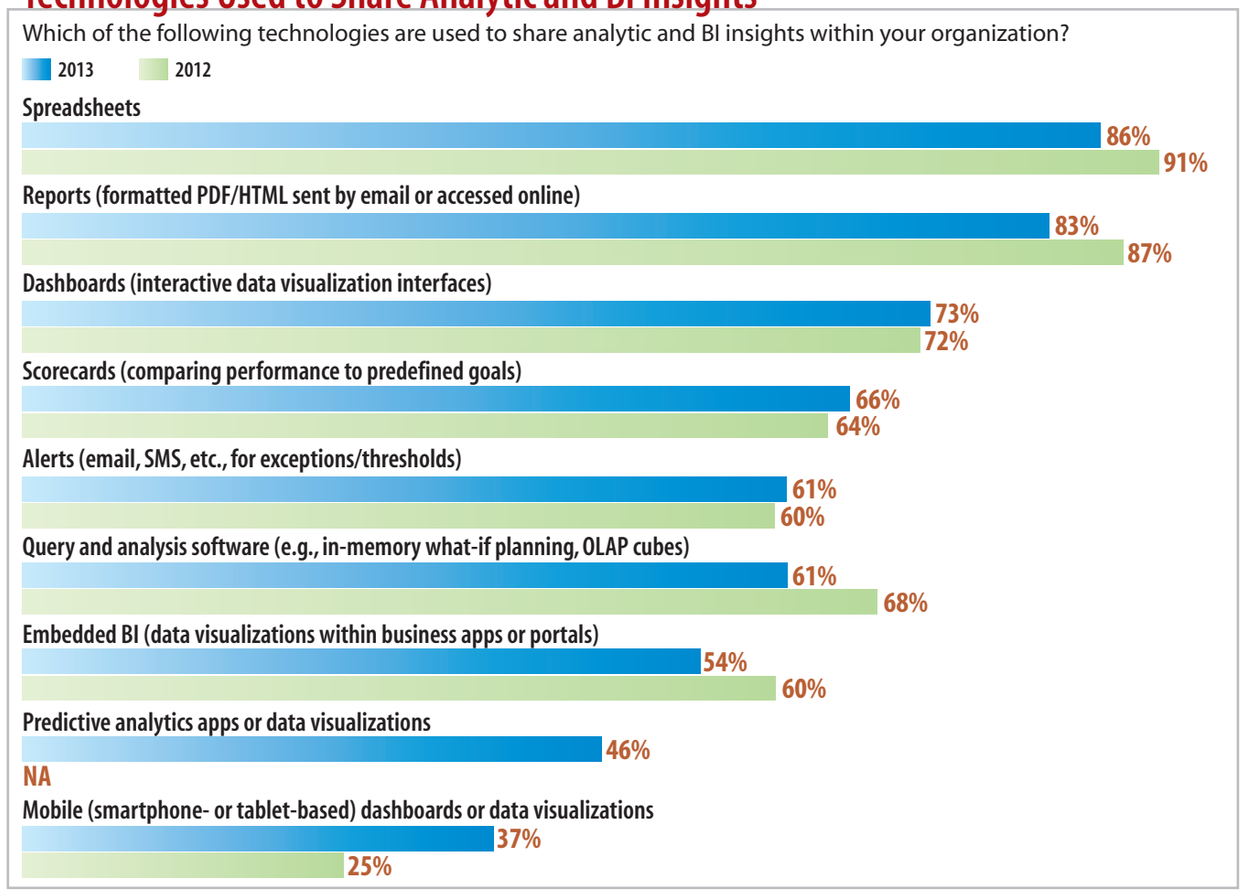
Data: *InformationWeek* 2013 Big Data Survey of 257 business technology professionals at organizations with 50 or more employees, September 2012

change security when planning and building your big data systems. The PRISM imbroglio is but the latest in a series of privacy flaps over data leaks at the likes of Facebook and LinkedIn. It's clear that big data is a target for hackers and industrial espionage and leads to big problems when lost. A key advantage of Accumulo is the ability to apply fine-grained access controls to information in a consolidated data store, where information sources and users might span the entire organization and even include business partners and contractors. But cell-level access controls should be coupled with well-known and widely available data encryption, both at the file system and network transport layers. Although open source Accumulo doesn't (yet) support encrypted file systems, Sqrrl's product does, and according to Khan, it adds a minimal 6% increase in latency. Seems like a good trade-off to us.

3. Identify one or two high-value big data analytics applications for a pilot: You don't want to boil the ocean with your first big data/BI projects. Instead, find a tightly focused application with substantial new business

Figure 9

Technologies Used to Share Analytic and BI Insights



Note: Percentages reflect extensive or limited use
R6061112/4
Base: 417 respondents in October 2012 and 414 in October 2011 at organizations using or planning to deploy data analytics, BI or statistical analysis software
Data: *InformationWeek* Analytics, Business Intelligence and Information Management Survey of business technology professionals

value. A good category to investigate is predictive analytics to identify your best customers, build loyalty, cross-sell new products and services, or better predict sales and buying patterns based on dynamic information such as the weather, social networking trends or current events.

Many of our BI survey respondents are already on board, as our survey finds the most common BI applications to be financial analysis, forecasting and business activity monitoring, although 80% plan on using predictive analysis. In fact, most respondents use the technology to improve business planning and report on current business activity and metrics. For them, software like Accumulo, MapReduce and other NoSQL distributed data stores can provide near-real-time information to predict customer behavior, sales trends, and inventory needs or identify fraudulent activity.

DRAWBACKS

Accumulo's Not-So-Silver Lining

While Accumulo is a good fit for a lot of existing big data applications and could easily displace conventional Hadoop/HDFS installations while improving scalability, real-time performance and security, there are downsides:

>> Accumulo is new and not widely used outside of government. Unless you use Sqrrl (itself a very new company), you have to manually install from the open source repository, which includes a few other Apache software components (ZooKeeper, Thrift), so there's more expertise, care and feeding required.

>> Unlike Hadoop/MapReduce and HDFS/HBase, no public cloud services offering Accumulo instances are yet available. Amazon has a tutorial on how to install it on top of Elastic MapReduce, but again, a lot of DIY is involved.

When applications are developed, new features like security labels and iterators come with a learning curve. This 451 Group survey of LinkedIn profiles finds increasing numbers of Accumulo-savvy developers, but it's still a very small base. Read: expensive.

>> The original Senate Armed Services Defense Authorization bill last year contained skeptical language directed at Accumulo. To wit, Section 929 "prohibits any DOD component from utilizing the cloud computing database developed by the National Security Agency (NSA) and known as 'Accumulo' after the end of FY2013, unless the DOD CIO certifies that: (1) there are no viable commercial open source databases that have such security features, or (2) Accumulo itself has become a successful open source database project." In end-of-the-year legislative haggling, the House Defense Authorization bill, which didn't contain references to Accumulo, was the basis for the final law. Although the Senate incorporated many sections of its bill into the final compromise, Section 929 wasn't one of them. The Senate's actions could have been the result of lobbying by companies like Oracle or Microsoft with skin in the game. But it does point out that, as with any new technology, long-term viability is not assured. Incumbents have an incentive to squash it.

MORE LIKE THIS

Want More Like This?

InformationWeek creates more than 150 reports like this each year, and they're all free to registered users. We'll help you sort through vendor claims, justify IT projects and implement new systems by providing analysis and advice from IT professionals. Right now on our site you'll find:

Research: Big Data, Smart Data: Our 2013 Big Data Survey shows we're not lacking facts, figures or the tools to wrangle them. So why do just 9% rate themselves as extremely effective users of data? And how do we expect to improve when just 31% have a wide array of business users accessing information and just 20% plan to grow their dedicated analytics teams?

Research: Data Center Decision Time: Is your glass house a sparkling hub of IT innovation or a financial albatross? For many, it's the latter. Worse, we often lack agility, 67% say application and hardware philosophies frequently or occasionally conflict with business demand, and just 27% say a private cloud is a high priority. Meanwhile, planned use of colocation facilities is up an anemic 5 points over last year.

Research: 2012 Big Data and Analytics Staffing Survey: It's a good time to be a big data and analytics expert: demand is high and supply is low. 18% of big data-focused companies in our survey want to increase staff in this area by more than 30% in the next two years, but 53% expect it will be hard to find people with the required skills. Find out how companies in diverse industries are coping with the shortage.

PLUS: Find signature reports, such as the *InformationWeek* Salary Survey, *InformationWeek 500* and the annual State of Security report; full issues; and much more.