TAKING THE GUESSWORK OUT OF BIG DATA

Understanding where the information comes from and what's required to process it will aid adoption

The proliferation of big data has government agencies scrambling to handle the new information forms even as they continue to multiply. The benefits of better, faster decision-making coupled with lower costs and improved citizen service make big data hugely attractive, but federal, state and local information technology managers are finding that new technology and new skills are required to tackle the explosion in information production, no easy feat that is made even more challenging by budget constraints. But before organizations can figure out how to manage and analyze big data, they must first understand what it is, said James McLeod-Warrick, managing partner at Beacon Technology Partners, during an Oct. 30, 2012, webcast titled "Big Data: From Hype to Reality."
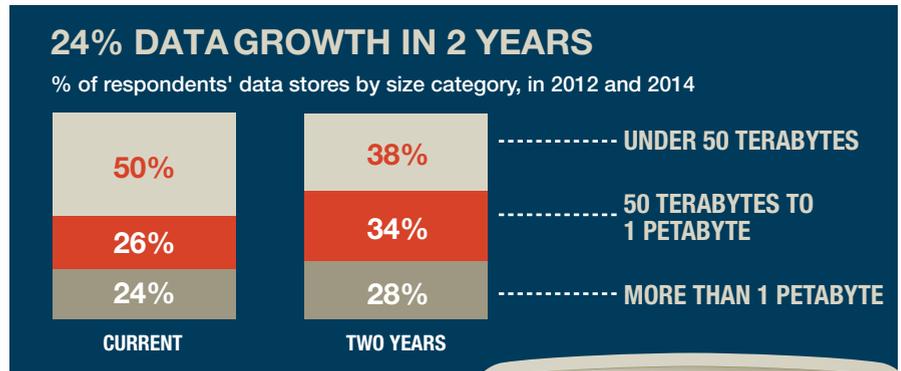
Big data describes a new generation of technologies and architectures designed to

Big Data Defined

Data sets whose size is beyond the abilities of the typical database software tools to capture, store, manage and analyze.

extract value from very large volumes of data flowing from a wide variety of sources, McLeod-Warrick said.

"There is no real consensus definition, but a common definition is data sets whose size is beyond the abilities of the typical database software tools to capture, store, manage and analyze," added Scott Pearson, director of Big Data Solutions at Brocade. "It's a horizontal IT solution crossing many

24% DATA GROWTH IN 2 YEARS
% of respondents' data stores by size category, in 2012 and 2014

CURRENT: 50%, 26%, 24%
TWO YEARS: 38%, 34%, 28%

UNDER 50 TERABYTES
50 TERABYTES TO 1 PETABYTE
MORE THAN 1 PETABYTE

verticals, such as Web 2.0, health care, finance and federal."

Big data isn't new but the advent of the iPhone in 2007 and the iPad in 2010 – "basically the proliferation of handheld devices," Pearson said – caused the explosion of data. Ninety percent of the world's compute data has been produced in the past two years, he added.

"Big data is growing seven times faster than the overall IT business," Pearson said. "It's here and it's real. It's not a marketing term."

Public and private organizations have found that big data can expedite fact-based decision-making, deliver services more efficiently, help detect security risks and enable more accurate planning and forecasting. Still, it's not without challenges.

"New skills sets are demanded to manage and analyze petabytes of data in real time," McLeod-Warrick said. "More sophisticated analytic tools are desired, and IT infrastructures need to be far more flexible and agile to handle the volume, variety and velocity that big data requires."

3 V's of Big Data

In August 2012, Federal Computer Week, along with Beacon Technology Partners,

> "Big data is growing seven times faster than the overall IT business. It's here and it's real."
>
> — *Scott Pearson, director of Big Data Solutions at Brocade*

conducted a survey of technical and business decision-makers at federal agencies, including defense and civilian, to get a sense of where the government stands with big data. The survey found that about half of the agencies have big data initiatives under way, with 21 percent having fully implemented one and 35 percent starting one. Almost half of the 193 respondents believe that agencies that are unable to implement and use big data will find it harder to meet their agency's mission.

The biggest draws of big data are improving the quality and speed of decision-making (76 percent), better planning and forecasting (68 percent) and improving the efficiency of internal processes (67 percent).

"Big data fundamentally involves the three V's: volume, variety and velocity of the data being received as well as analyzed," McLeod-Warrick said.

First, the average quantities of data used for a big data analytic project will grow 24

> "Big data has grown beyond the hype stage. It's seen as vital to many federal, state and local agencies."
>
> *— James McLeod-Warrick, managing partner at Beacon Technology Partners*

percent over the next two years with more than one-quarter of our foreseen projects relying on data sets in excess of 1 petabyte, he said, putting extreme stress on network infrastructures.

Second, IT managers worry about their agency's ability to handle the variety of data it receives. Traditional data warehouses and analytics were created to handle structured data, which is stored and indexed in traditional database formats. Big data comes from semi-structured data, such as Extensible Markup Language and RSS feeds, and unstructured data, such as voice, video and print or real-time feeds.

"Many observers feel that it's these real-time data flows which constitute the third V, velocity," McLeod-Warrick said. "This is where the real value in big data will reside."

Challenges of Big Data
It's also where agencies' concerns lie, Pearson said:
• Existing IT structures may hinder new types of data sourcing, storage and analysis.
• Existing IT architecture may prevent the integration of siloed information.
• Managing unstructured data often remains beyond traditional IT capabilities.
• Many existing systems were built to deliver data in batches, not to provide a continuous flow for real-time decision-making.

Agencies involved in or investigating big data are looking to hire people with experience in business in addition to statistics, computer science and machine learning. And they are focusing spending on building a network architecture that can handle high-speed data capture, including increased network bandwidth, server hardware, data warehousing and cloud storage.

Then there's the issue of sorting the data into meaningful patterns, McLeod-Warrick

IMPLEMENTATION STATUS

INVESTIGATING - 44%
JUST STARTING - 35%
FULLY IMPLEMENTED - 21%
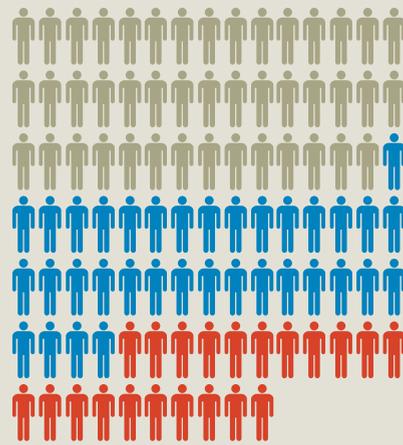
added. "Statistical tools are fine for starters, but the true promise of big data won't necessarily come from standardized toolsets."

Instead, he said, agencies are turning to advanced data visualization, which NASA has used to demonstrate the consequences of drought on groundwater levels. Text mining and sentiment analysis tools help make sense of social media and blog posts, while complex event processing would combine data from multiple sources to infer patterns to identify things like security threats. Lastly, Hadoop, open-source software developed to analyze structured and unstructured big data sets, is gaining momentum in the private sector at companies such as Google and Facebook, but also at government agencies.

Brocade's Big Data Solutions
"It's my view, and Brocade's view, that the network is at the core of big data, basi-

cally at the intersection," Pearson said. "Networks are crucial to process, transport, analyze and manage big data."

Moreover, the network must be able to scale incrementally, making Ethernet fabric the foundation of choice for handling big data, Pearson said. To that end, Brocade is working with partners such as IBM, Hadoop and Zettaset to provide several package solutions:
• An enterprise-grade platform for exploring big data, involving servers, fabric adapters, Ethernet fabric TOR switches and fabric aggregation chassis switches.
• A real-time platform for big data in motion, involving telemetry with a 10 gigabyte chassis switch. It takes Brocade's chassis MLX switches and loads IBM telemetry streams and uses a smart switch to grab data-in-motion, whether voice or Internet.
• The IBM intelligent cluster, where the networks are located inside the cluster as a management or a compute network.
• End-to-end big data analytics for data-in-motion and data-at-rest, involving Hadoop, telemetry InfoSphere streams and the IBM intelligent cluster.

Big data is here to stay, so finding ways to synthesize it is crucial and not impossible despite the challenges.

"Big data has grown beyond the hype stage. It's seen as vital to many federal, state and local agencies," McLeod-Warrick said. "Investment in more agile, flexible networking architectures as well as more robust analytic tools can help bring the promise of big data to fruition." •

SPONSORED BY:

BROCADE

For more from Brocade, go to
www.brocade.com