

BIG DATA WITHOUT BIG HEADACHES

THREE STRATEGIES TO HELP THE BOTTOM LINE

The history of business is replete with bad decisions made with incomplete information. One classic example is the introduction of New Coke in 1985. The Coca-Cola Company had done extensive market research and conducted numerous focus groups, and these efforts seemed to clearly indicate that a new, sweeter formula would be more popular than the beverage that had been the epitome of cola for generations. Of course, the old formula was reintroduced less than three months later—after a near-mutiny by consumers, shareholders, and bottlers. It turned out that consumers' emotional connection of the classic taste with the Coca-Cola brand was not a part of the research—and was very hard to break.

101001010110100100101

0010100101

1010010101101001001

0101011010010

0101011010010

To make an effective decision for any business, information is essential. These days, more information is available than ever before, but assembling and analyzing that information has become exponentially more complex. "Many new types of data are being collected in just the past few years, from GPS locations via mobile phones, to social media posts, to data from advanced medical devices," observes Ben Woo, managing director and founder of Neuralytix, a New York City-based analyst firm that helps organizations optimize their use of data. "The key is to be able to use the information to advance the business."

The practice of making such information actionable has come to be known as 'big data'. "It's really a convergence of internal data sets, such as transaction data, and external data sets, such as social media, to gain insight and ultimately innovation, to provide companies with economic benefit," Woo explains.

Early adopters of big data practices included e-commerce companies such as Amazon.com, which developed advanced algorithms to manage the customer experience precisely and individually.¹ Brick-and-mortar retail chains, among others, have also had success.² But the reality is that most companies are just beginning to dip their toes into big data. "A year ago, there was still sort of a science project mentality around big data," Woo relates. "As we stand here in 2013, many organizations are going beyond the science experiment to practical application."

For executives, big data presents great opportunity as well as potential management headaches if not approached strategically. In addition to Neuralytix's Woo, *CIO Digest*

spoke with three other IT leaders to get their take on where things stand today with big data, and where they foresee them heading. All agree that adopting big data is a larger process than deploying an application or opening a data warehouse. As *Harvard Business Review* recently argued, being successful requires "a new culture of decision making,"³ one that is based on data rather than intuition.

1. Find the right combination of data

One requirement of a data-driven decision making process is that the data be relevant. "Many use the term 'big data' because of the volume, but I really think that in a couple years we'll be talking about simply 'data', because the scalability problem has been solved," observes Scott Gnau, president of Teradata Labs, which develops



Paul Chang, CIO,
Information Technology Center,
Asia Pacific Telecom

... ➤ "We want to maximize the value we get from the investment and provide all the data that's needed to optimize our operations and increase revenue."

— Paul Chang, CIO, Information Technology Center, Asia Pacific Telecom

big data solutions built on Teradata Corporation's Unified Data Architecture. "Our customers don't need much help with storing and retrieving large amounts of data; they come to us because they want to make sense of that data and use it to build the business.

"The biggest thing right now is combining geospatial data and time

sequence data," Gnau continues.

"We've always done analysis involving who, what, where, and when, but the 'where' was very generalized. In today's world of sensors and GPS, the 'where' can be very specific and in very real time."

In Taiwan, Asia Pacific Telecom (APT) competes in a very crowded marketplace, with four major telco providers in a country of just 24 million people, and margins are thin. Data on the combination of 'where' and 'when' is crucial to the company's bottom line in a number of ways. "We're already using transaction and location data from users' devices to develop predictive models for network capacity," explains Paul Chang, CIO at the firm's Information Technology Center. "This helps us to provide the best service to customers at the lowest cost.

"We're also in the process of building a big data infrastructure to make customer behavior information available in real time," Chang adds. "This may enable us to work with advertisers to send offers to some consumers based on their interests."

Several years ago, Telefónica de España developed a big data infrastructure to support the company's internal needs, but now markets a

subscription-based big data service to other enterprises in Spain. Internally, the company uses the data for network resource allocation and location-based marketing. But Telefónica's customers have differing needs. "For example, we're working with a major beverage supplier," explains Sergio Díaz-Pérez, product manager for Storage and Backup. "People drink beverages in different amounts if it's sunny outside versus cloudy or rainy. And the amount goes up if there's an important football match being played. So we're integrating their transaction data with weather, sports, and calendar data to optimize their delivery strategies."

2. Build physical and communications infrastructures

Being strategic about big data involves more than throwing together a Hadoop cluster, but you don't always have to spend millions of dollars on an all-new infrastructure either. According to Neuraltix's Woo, the key is internal communication. "I've observed that there is too much emphasis, and often too much money spent, on the infrastructure elements, with little or no focus on talking with the business about what's available and what can be done with it," he asserts. "In other words, IT is not doing a good job selling information within the organization. As a result, a lot of line of business leaders are going outside of IT for solutions."

That's why APT's Chang is currently meeting with leaders from all of the company's business units as his team builds its new big data infrastructure. "We want to build the infrastructure that best meets our business needs, and the result is that we're going in a somewhat different direction than the other



Sergio Díaz-Pérez, Product Manager, Storage and Backup, Telefónica de España

...> "We have found that Symantec appliances deliver much better deduplication results than Data Domain for less cost."

— Sergio Díaz-Pérez, Product Manager, Storage and Backup, Telefónica de España

BIG DATA THREAT MONITORING AT SYMANTEC

telecommunications providers here,” he comments. “We want to maximize the value we get from the investment and provide all the data that’s needed to optimize our operations and increase revenue.”

Teradata’s Gnau emphasizes that a flexible infrastructure enables IT teams to build a big data practice according to the needs of the company. “The Teradata Unified Data Architecture supports various solution stacks depending on business need,” Gnau explains. “This gives customers the flexibility to adjust the infrastructure according to which people in the organization need data analytics at a given time without starting from scratch.”

The team at Telefónica also consulted extensively with the lines of business before deploying an EMC Greenplum system supporting a Hadoop cluster and Bayesian mathematical analysis software. “We expended considerable effort to make sure that the infrastructure was not only right for our company, but scalable to support services for our customers,” Diaz-Pérez explains.

3. Ensure data protection and availability

The information in a big data system is there because it’s important to the business, so it goes without saying that it should be backed up. But the vast volumes involved can potentially create an unwieldy system where recovery is slow and data is unavailable or slow to access. “One of the things we often forget is that data generates more data,” says Neuralytix’s Woo. “Over time, a data warehouse will have an incredible volume of duplicate data, and that affects the agility of the system.”

Among Teradata’s offerings is a Backup, Archive, Restore (BAR) solution, which includes Symantec

As the owner of the world’s largest threat database, Symantec has been involved in big data since long before the term was coined. The volume of threat data was massive even a decade ago, when most of it was in the form of virus signatures and firewall log data. Today, Symantec maintains around 600 terabytes of threat data, involving roughly 6 million daily incidents from about 200 million Symantec-protected endpoints around the world.

As threats have become exponentially more numerous and complex in recent years, Symantec engineers have increasingly leveraged big data techniques to stay ahead in a quickly evolving security landscape. Several engines were developed to manage advanced predictive modeling of security risks. A few examples include: (1) a behavior-based engine that identifies potential threats by their behavior, as opposed to relying on prewritten virus signatures, (2) a reputation-based engine that rates potential threats according to their prior impact on other Symantec-protected endpoints, and (3) another engine that rates applications for their impact on system stability. This advanced analytics infrastructure benefits Symantec customers in the enterprise, public sector, small and medium business, and consumer spaces.

In 2009, Symantec built the Symantec Data Analytics Platform (SDAP), a highly scalable big data platform that hosts all of the company’s security data, including the threat analytics infrastructure. The SDAP platform currently manages approximately 2 trillion database rows, growing at a rate of about 5 billion rows per day. Upon the deployment of SDAP, the team was able to merge several disparate threat databases that had been built at different times using different technologies. This allowed Symantec analysts and data scientists to correlate data from many different sources to mine new intelligence that was not available before. And the average time to develop new big data applications, from concept to deployment, was significantly reduced.

The SDAP data center is built to LEED Gold green building standards for minimal power consumption. As with all of Symantec’s data, SDAP data is protected by Symantec NetBackup, which deduplicates the massive data store at the source and replicates the data to the secondary data centers. Veritas Storage Foundation HA from Symantec provides high availability for the threat data. It uses N+1 clustering that enables automatic failover to one of two secondary data centers in the event of disaster.

COMPANY PROFILES

ASIA PACIFIC TELECOM

Founded: 2000

Headquarters: Taipei, Taiwan

Competitive Marketplace: APT is one of four major telcos in Taiwan

Website: www.aptg.com.tw

TELEFÓNICA DE ESPAÑA

Founded: 1924

Headquarters: Madrid, Spain

Global Reach: Parent company

Telefónica has a presence in 24 countries

Website: www.telefonica.es

TERADATA

Founded: 1979

Headquarters: Dayton, Ohio

Managing Big Data: More than 1,500 large enterprises rely on Teradata for data analytics

Website: www.teradata.com

NetBackup. “We’ve chosen to certify the various elements of our software stack so that we can guarantee an end-to-end process,” Gnau explains. “For us, it was an easy decision to certify NetBackup for our BAR offering, as it’s an industry standard. It’s deployed with the majority of our customers, and it works really well.”

NetBackup supports Teradata’s business model to provide a flexible solution. “It’s optimized for every operating system, every type of hardware, and physical, virtual, and cloud environments,” Gnau notes. “A company’s Teradata



Scott Gnau, President,
Teradata Labs,
Teradata Corp.

...➤ “For us, it was an easy decision to certify NetBackup for our BAR offering, as it’s an industry standard.”

– Scott Gnau, President, Teradata Labs, Teradata Corp.

services up and running for customers. Storage Foundation HA provides APT with automatic, policy-based failover in the event of a system or database failure, enabling the company to deliver on an SLA of 99.99 percent uptime for its key services.

As APT prepares to deploy a new big data infrastructure, Chang’s team is testing NetBackup’s deduplication capabilities and plans to begin deduplicating across the enterprise this summer. “This will be a key component of our big data initiative, as

integration with automatic image replication, which is useful to us in replicating between data centers.”

Helping the bottom line

The potential is there. At least one study has confirmed that data-driven decision-making helps improve companies’ performance in financial and operational results across a spectrum of industries.⁴ “The power of the technology, based on what we can be creative about as human beings, is absolutely tremendous,” asserts NeuraLytx’s Woo. “It’s really possible for companies to improve the efficiency of their operations, provide good value for their customers, and increase sales.”

By embracing big data strategically, IT leaders might just prevent a New Coke-style disaster from happening on their watch. ■

administrator can view backup status through the Teradata console, while the backup administrator can monitor the backups from the NetBackup console.”

APT’s Chang appreciates the flexibility of Symantec solutions as well. “We have a heterogeneous server environment, with Solaris, Linux, and Windows on physical and virtual servers,” he notes. “It’s really nice that Symantec’s backup and high availability solutions are optimized for all our diverse systems.” Chang’s team relies on NetBackup across the enterprise, and uses Veritas Storage Foundation High Availability (HA) from Symantec to keep its

duplicate data will slow the performance of the system,” he says. “And with Storage Foundation HA providing high availability, we will be able to set aggressive SLAs for our advertising partners.”

At Telefónica, Díaz-Pérez and his team have introduced Symantec NetBackup appliances to support the enterprise, including the firm’s big data infrastructure and a managed backup service offering. “We had been using Data Domain appliances for deduplication, but we have found that Symantec NetBackup appliances deliver much better deduplication results than Data Domain, and for less cost. NetBackup also provides

¹ Andrew McAfee and Erik Brynjolfsson, “Big Data: The Management Revolution,” *Harvard Business Review*, October 2012, p. 62.

² Kashmir Hill, “How Target Figured Out A Teen Girl Was Pregnant Before Her Father Did,” *Forbes.com*, 26 February 2012.

³ McAfee and Brynjolfsson, 65.

⁴ McAfee and Brynjolfsson, 63-64.

Mark Mullins is a managing editor of CIO Digest and senior manager of Symantec’s SMB and Channel Publishing Programs’ team.

PROTECTING BIG DATA WITH SYMANTEC

- > Symantec NetBackup
- > Symantec NetBackup Appliances
- > Veritas Storage Foundation HA
- > Veritas Cluster Server