VOLUME VELOCITY VARIETY VALUE

Delivering The Elusive Fourth V Out Of Big Data
WHAT IS BIG

THE EXPERTS WEIGH IN:

“What our research suggests is that scale and scope of changes has really brought us to an inflection point. There are sets of data that are really beyond the scope of our traditional ways of dealing with them.”
– Michael Chui, senior fellow
McKinsey Global Institute

“The problem is not ‘big’ data — it’s more data. The increased use of interactive applications and websites — as well as sensors, meters and other data-generating machines — has increased the volume, velocity and variety of data to store and process.”
– Matthew Aslett, analyst
451 Group

“IT leaders must educate their business counterparts on the challenges while ensuring some degree of control and coordination so that the big-data opportunity doesn’t become big-data chaos, which may raise compliance risks, increase costs and create yet more silos.”
– Mark Beyer, research VP
Gartner
Big data may be the topic du jour at IT conferences, on conference calls and for industry webinars. But much like ‘cloud computing’ eluded definition back in 2010, ‘big data’ still remains fuzzy in the minds of many IT pros today.

In fact, according to TIBCO LogLogic’s Big Data and Cloud Survey, just 27% of respondents said they clearly understood what big data is. And those respondents weren’t the average man on the street; they were IT professionals at the director level or above.

So what the heck is big data? And why should we care about it?

Let’s start with the second question first, because that may well be the most important and its answer is the most obvious. Consultants with McKinsey say that effective use of big data will become one of the biggest technology differentiators of the next decade. With the right analytics, data can be turned into actionable intelligence that can be used to make businesses maximize revenue, improve operations and mitigate risks and to generally unlock the value in the information we already collect today.

Companies that pin down this slippery beast and leverage the capabilities it offers stand to not only transform their operations, but perhaps even their entire business.
QUANTIFYING ‘BIG’

A lot of ink has been spilled by experts bickering over an exact numerical definition for big data, but that may actually be a wasted exercise. The fact is that big data is all relative.

It could be a petabyte or a few terabytes. It could come from networking equipment log data, from machines in a manufacturing plant or from endless reports spit out by an enormous ecommerce infrastructure. The truth is that if you’re having a hard time extracting worthwhile information from existing 20th century database architecture and information is sitting idly on servers without ever being collated or charted — much less interpreted — it probably counts as big data to you.

These days analysts have latched onto what they call the 3Vs of big data to act as a litmus test for whether something counts as big data and to explain some of the challenges of parsing it:

**VOLUME**: As data production grows, the volume of output makes it difficult to keep up.

**VARIETY**: Whether it is transactional data, log data, database data, document data, social media data, audio or video, the types of data, file formats and storage structures continue to expand, increasing the challenge of analyzing across all types.

**VELOCITY**: The speed at which data is created continues to rev up, but our ability to keep up with the velocity of analysis is straining.

DID YOU KNOW?

The volume of information around the world is growing by 59 Percent per year

Source: Gartner

Global data will reach 35 Zettabytes by 2020

Source: IDC

90 Percent of the data in the world today was created in the last two years

Source: IBM
WHERE DOES IT COME FROM?

IT didn’t just pull these large and complex data sets out of thin air. Rather, all of that data is piling up as a byproduct of cumulative technological progress in other areas.

Service-oriented architecture (SOA) and modern integration models expose more information from core business applications than ever before, but application data is only a fraction of the big data volume — automation and connectivity have intersected in what many in the industry call the “Internet of things,” or the network of devices, machines and apps that are networked to automate our businesses and our lives.

All of these “things” tend to spit out data, all of it created in different formats, many ‘contained’ in different file structures.

This development, combined with a continuing increase in complexity of IT architectures, creates data sources as fragmented as they are voluminous.

“We’re generating information through all these devices, through all these different industries, and all these different services. The challenge comes in gaining access to that information and creating an inclusive ecosystem where information can be exchanged,” says Dr. Abel Sanchez of MIT.

Organizations today are challenged with the complications of a fragmented ecosystem. When data is in different formats, different locations and different contexts, search becomes difficult and correlation of business or technological process data becomes nearly impossible.
With so much disparate data being produced every day, enterprises are simply failing to unlock the unnamed fourth V of big data: VALUE.

According to analysts with Forrester, most organizations today use less than 5 percent of the data that’s available to them. What’s more, analysts with the firm believe that the average enterprise only manages to capture between a quarter and a third of the data it produces each year. As our capability to collect data has increased, our ability to store, sort and analyze it has diminished.

It’s a problem for the enterprise, but even at a personal level most people have experienced the phenomenon. Case in point: the experience of the family shutterbug.

When we were all taking pictures using film cameras, the typical snapshot artist would take maybe a couple hundred photos per year. The photos and negatives would go into a shoebox and if the photographer was really organized, she’d file them according to date or event. But many of us craved a classification system more capable and convenient than that; we wished for a way to easily find photos by color scheme, location or people depicted.

With the advent of digital photography we thought that all would change. For some, it did. But for most,
it actually made sorting even more troublesome than before.

Now, the family photographer shoots with impunity. Without worrying about film and developing costs, she can take thousands of photos of special events, vacations, kids, dogs and scenery — often using more than one camera! While the pictures themselves tell a story and recall nice memories, they usually go unseen.

The truth is that even with tools out there to tag and classify photos, most people don’t use them. There are just so many photos from so many devices that these pictures usually sit unsorted on a hard drive somewhere, or worse yet, on a number of different SD cards, CDs, smartphones and laptops spread around the house.

And that’s just a microcosm of the big data conundrum. With enterprise data, the scale and scope of the problem is magnified. Files are spread across more PCs and servers. More devices are producing a larger volume of data in more file types than ever. And instead of enjoying the benefits of all that information, we’re essentially chucking it all into a proverbial shoebox.

5 WAYS TO GENERATE VALUE FROM BIG DATA

When it is parsed and available, big data can be indispensable to the business in so many ways. According to the McKinsey Global Institute, the following five ways are some of the most important:

- information transparency and access to information at a higher frequency
- detailed performance information about transactional data that can help the business conduct controlled experiments to make better management decisions
- enough information to further narrow customer segments and better tailor products or services
- the foundation for sophisticated analytics to improve decision-making
- more information to move forward the development of next-generation products and services
Like the shoemaker’s children going barefoot, IT departments today run the risk of being among the last within their organizations to benefit from the treasure trove of data being produced. While technologists scurry to help business stakeholders extract intelligence from business metrics like customer service data, IT departments fail to look inward at its own data to streamline operations and security practices.

Studies conducted by Forrester Research showed that in 2012, among those organizations that plan information analysis projects, 72 percent are doing it for improved business decisions while just 42 percent are doing it to improve IT initiatives such as security and compliance. The emphasis on better informing line-of-business leaders tracks well with business goals, but IT shouldn’t be lost in the shuffle, either.

According to a recent survey conducted by Echelon One, over half of IT decision-makers claim their organizations currently manage logs for more than 5,000 devices and applications. The sheer volume of data that these sources pump out boggles the mind — approximately 38 percent say that their log data totals more than a petabyte.

None of that data does IT departments any good sitting in the proverbial shoebox. According to researchers with the Institute for Applied Network Security (IANS), simply collecting logs is just the first step in the process of creating value from it:

“Making sense of [logs from web applications, middleware, custom backend applications, and databases] requires an ability to normalize the collected data, correlate it, report on it, and send actionable alerts for issues related to security and malicious activity like APTs, but also for availability, work flow issues, and help desk efficiency.”
While there has been much talk about organizations bringing in new tools like Hadoop and hiring ‘data scientists’ to mine big data for information to deliver better business results and improve IT operations, the problem is that these are long-term solutions to a problem that grows in immediacy every day.

The truth is that it will take the industry time to learn how to best take advantage of newer tools in the data analysis toolkit — perhaps even a whole generation of workers to do it right. In fact, according to McKinsey, in the U.S. alone the enterprise community could face a shortage of 140,000 to 190,000 people with the deep analytical skills to live up to the data scientist moniker.

As we brace for this transition, it is clear that organizations need a contingency plan.

According to Gartner vice president and distinguished analyst Yvonne Genovese, the lynchpin of managing big data effectively is to make decisions through “pattern-based” strategies:

“Pattern-based strategy, as an engine of change, utilizes all the dimensions in its pattern-seeking process. It then provides the basis of the modeling for new business solutions, which allows the business to adapt. The seek-model-and-adapt cycle can then be completed in various mediums, such as social computing analysis or context-aware computing engines.”

Fortunately for IT, those pattern-based strategies already exist within the technological framework we see today. Today’s log management solutions already have the pattern-based computational power behind them to help organizations manage the big data problems they face now and in the near future. It is only a question of how we use these tools.
A survey conducted by ESG showed the following challenges standing in the way of gaining value from Big Data:

- **47%** reported data integration complexity an obstacle
- **34%** said the lack of skills necessary to properly manage large data sets and derive value from them stood in the way
- **29%** said data set sizes limited their ability to perform analytics
- **28%** said difficulty in completing analytics within a reasonable period of time was an issue
BIG DATA TRANSFORMATION — UNDERSTAND, ANTICIPATE, ACT
Whether it is to tease out intelligence to help the business or to make business run more smoothly and securely, big data’s value is best extracted through a transformation of IT infrastructure and providing domain experts with interactive data analytics.

Whatever technology or practices you use to deliver value from big data, the process you establish will inevitably lead you to cycle through what we at TIBCO LogLogic have dubbed Understand — Anticipate — Act.

**UNDERSTAND:** using analytics and data discovery tools, we allow data domain experts and data scientists to visualize large data sets and interact with that data to understand the trends, outliers and emergent patterns. The data is rationalized across multiple types of logged machine data, merged with information from real-time applications and, if necessary, supplemented by historical data from databases, spreadsheets and files.

**ANTICIPATE:** once we have modeled a particular pattern through visual and/or computational analytics, the model can be used in real-time to detect re-occurrence of similar events. Using models in this way we can proactively anticipate opportunities or threats across the organization.

**ACT:** when events are detected, we must take action to derive the ultimate value — shutting down access, rerouting traffic, kicking off an investigation or maintenance process, even creating root cause applications on the fly complete with contextual information — when the likelihood of a critical event exceeds a certain probability, all of these are possible actions which can immediately be taken.
GETTING VALUE OUT OF DATA NOW

Getting value out of big data is no pie-in-the-sky proposition. In fact, many organizations are already using TIBCO LogLogic Analytics in two impactful areas:

Risk Mitigation • Operational Efficiency

Let’s examine a few case studies to see how it’s being done.

I WANTED BETA BLOCKERS, NOT VIAGRA!

A drug distribution company that runs production lines to pick, bottle, seal, label and pack millions of drug products once had the embarrassment of a disgruntled employee hijack the system in the middle of the night and change the process to serve up Viagra instead of Beta Blockers. The bottles were filled, labeled incorrectly, sealed, packed and distributed to the retail outlets. The problem was only discovered by an end customer with ‘undisclosed’ symptoms. This company had to recall all of the drugs from the entire lot and destroy them — along with their reputation.

Our customer, a competitor of the embarrassed firm, wanted a cost-effective way to keep tabs on employee actions to catch people should they choose to try such shenanigans. The company uses LogLogic to track every part of the process by barcoding and logging events like picking, bottling, labeling, sealing and shipping. Though they cannot prevent the action of a disgruntled employee tampering, they will know when they come into the building through logging of card access. They can then tie this time slice to which drugs were handled before and after that event, this massively limiting the cost.
PRESCRIPTION FOR LOWER PHARMA RISKS

One of our healthcare customers is developing a process to use our products to streamline and facilitate the monitoring of Schedule II and III controlled substances such as OxyContin, Vicodin and Percocet. Healthcare organizations are required to report suspected prescription abuses and frauds to the DEA. Before LogLogic, this process took upwards of a week to prepare a single alert for the DEA to take action. Now with LogLogic, this process requires no human involvement or effort at all and reduces the process time from a week down to less than an hour.

Here’s an example of how this alerting capability will work: One of the healthcare organization’s doctors prescribes five separate prescriptions of Schedule III narcotic drugs in less than an hour. The DEA will receive an automatic alert (based on alert parameters preset in LogLogic) and they can choose to take action as they see fit, determining if the prescriptions are legitimate, if the prescriptions are in fact being requested by the physician in question, and so on.

THREATENING EMPLOYEE NABBED IN MINUTES

A C-level executive had been receiving some seriously threatening emails to his company outlook account. Despite extensive efforts, not a single person in the company’s IT department could trace the origin of these threatening emails. That is, until someone thought about turning to the LogLogic appliance to review the logs. The tool was able to find out that it was actually an employee who was sending the executive threatening emails, which they were doing by tunneling in via VPN with a random Hotmail account. The process of locating the culprit using LogLogic took only 4 minutes!

KEEPING TABS, POST-911

An unnamed university uses LogLogic to provide reports to UK law enforcement agencies that track and investigate students suspected of involvement in terrorist activities. These law enforcement agencies use LogLogic to investigate what kind of information and which web sites these students are accessing on the internet, as well as who they are communicating with.

RISK MITIGATION
A distribution company found itself delivering up life-threatening situations all because of print server crashes. The company depended on very strategic driving routes to deliver drugs in its special refrigerated trucks. However, sporadic printer server crashes, were wiping out the routes and derailing their timing to the point that the drugs in the trucks were spoiled — a life-threatening situation for patients.

Using analytics, the firm is able to keep this situation from happening by flagging the few odd behaviors from the printer servers that cause the problems. By logging and alerting on three odd events that cause such a crash, and sending an emailing alerts to the printer server, they run a simple, automated script that keeps the server from ever crashing.
EFFICIENCY

AN IT CHRISTMAS LIGHT BULB CHECKER

A very common problem within our client base is the mystery of poorly running services or appliances. Trying to figure out why a service is not running and which router is the offending one is akin to hanging Christmas tree lights where one bulb is blown and you have to start at one end and work your way painfully through the line, one bulb at a time, to find the bad one. Wouldn’t it be great if you had something that could pinpoint the exact bulb in a split second? This is exactly what one of our largest telco and largest healthcare customers use LogLogic for, when they have a spurious router that’s gone off the reservation.

KEEPING INTERNET BANKING ONLINE

With a combined 7.8 million active internet banking customers purchasing 2.6 million products a year online, one of our customers needed a tool that could ensure its infrastructure stayed healthy at all times of the day. It brought in LogLogic to do just that, by constantly monitoring the functionality of this huge infrastructure. The retail online banking platform produces custom application and infrastructure logs which the bank uses to troubleshoot issues within the production. Loglogic was seen as the only way to monitor in production and have the scale to handle the job at hand.

TIME WASTED NO MORE

It’s a common scenario: a user has changed his password and some other machines or services still know his old password and try to authenticate, continuously locking the user out. Troubleshooting this kind of error takes on average six to eight hours. With Loglogic you can have your helpdesk agent receive a support ticket; they can then hit a button that shows them the users’ logs from the last hour, a simple, powerful use case saving hours of frustration and running through mazes.
To learn more about what kind of value TIBCO LogLogic Analytics can drive through your big data, visit www.tibco.com/loglogic-analytics