HOW TO OVERCOME 3 KEY BIG DATA CHALLENGES

LEARNING FROM COMMON MISTAKES TO TRANSFORM BIG DATA INTO INSIGHTS
OVERVIEW

Google’s Eric Schmidt has noted that, from the dawn of time through 2003, humans created five exabytes of data. That’s the equivalent of 250,000 years worth of DVD-quality video. ¹ This amount of data is now being produced every two days. The era of Big Data has arrived.

Organizations of all sizes understand the value of mining data to improve their businesses and are rushing to implement Big Data projects. Unfortunately, collecting, organizing, analyzing, and operationalizing the vast amounts of data required to yield actionable results is not easily done. Jim Kaskade, VP of Big Data & Analytics at CSC, commented on Big Data projects, “The

¹ TechTarget Search Storage.
rule of thumb is 100 percent of the Big Data projects go over schedule, 100 percent go over budget, and half of them fail.”

There are a number of causes for the high failure rate of Big Data projects. Three of the most common are failure to operationalize data insights, inefficient tools and processes, and a lack of tie in to business goals.

New data preparation and analytics tools have now come onto the scene and are enabling significant progress toward overcoming these obstacles. This paper will explore these three key challenges and show how businesses can surmount them to realize the full benefit of Big Data analytics.

---

THE DAWN OF BIG DATA

WHAT IS BIG DATA?
Big Data is the term adopted to describe the huge volume of structured and unstructured data that contains vast amounts of information but requires sophisticated analytics tools to produce insight. According to Webopedia, the term Big Data describes not only the mass of data but also the technology and processes involved in processing it.

Many industry analysts define Big Data using the four Vs – volume, velocity, variety, and veracity.³

VOLUME represents the actual amount of data available. Data is growing exponentially and businesses are struggling to manage it all.

VELOCITY is the speed at which data is being created and how fast it must be processed to meet business needs. Organizations are often burdened with historical information that may no longer be relevant, so it is essential to rapidly analyze the data and use only the most relevant data to drive decisions.

VARIETY refers to the multiple data sources and varied formats in which the data can be presented. Organizations must collect, organize, and analyze data from such disparate sources as websites, point-of-sale systems, CRM systems, ERP systems, Excel® files, social media, and legacy systems.

VERACITY denotes the uncertainty surrounding data caused by inconsistency and incompleteness, ambiguities, latency, inaccuracies, and trustworthiness. Organizations need to determine which data can be trusted when populating their analytics models.

WHY BIG DATA MATTERS:
Big Data represents a tremendous opportunity for organizations to enhance and transform their decision-making process. It can drive innovation, establish new pricing models, introduce more effective ways to engage with customers and partners, establish operational efficiencies, monitor compliance and risk, uncover new market opportunities, and help develop a better overall understanding of the business.

GROWING USE OF BIG DATA:
Businesses are increasingly buying into Big Data. A recent Gartner study revealed that three out of every four organizations surveyed have invested or plan to invest in Big Data in the next two years, up from 64% in 2013. The same study shows that more businesses are making the leap, with the number of respondents who had no plans for Big Data falling from 31% in 2013 to 24% in 2014.4

4 Gartner Survey reveals that 73 percent of organizations have invested or plan to invest in Big Data in the Next Two Years, Gartner Newsroom, September 17, 2014.
With the enormous resource commitment organizations are making, why do Big Data projects so often fail to deliver the desired business value? Here are three main reasons.

FAILURE TO OPERATIONALIZE RESULTS: Bridging the gap from data collection and organization to analysis and implementation is a leap many organizations are finding hard to make. Data is rushing in from a growing number of disparate sources, overwhelming the front end of the analysis pipeline. Complex, multiple, incompatible formats create a logjam, bogging down the analytic process right from the start.

Uncovering actionable results from a Big Data project requires that the data get into the hands of people who can do something with it. In too many cases, that person must be a data scientist who has to untangle the various data streams and complex patterns to present insights to those who can operationalize it for day-to-day use. The true value of data is unleashed only when it is put in a business context.

The problem is that the availability of data scientists is at an all-time low, with some estimating that between 140,000 and 180,000 data scientist positions will remain unfilled by 2018. Without this front-end expertise, many Big Data projects can derail before they even begin. Information remains unprocessed and unfocused, without yielding any relevant actionable insights that can help the business.

---

INEFFICIENT TOOLS AND PROCESSES:
Traditional tools and methodologies have proven to be only marginally effective for Big Data projects. Traditional enterprise data warehouses and other BI infrastructures aren’t able to perform well when faced with the volume, variety, and speed of the growing data streams. Integrating and blending data from multiple sources using these resources is simply too slow and expensive, and fails to produce acceptable results in a timely manner.

This has led many business users to create their own data and spread marts oriented to a specific business line or team, convoluting the data further, leading to more failed projects. The bottom line is that traditional extract, transform, and load (ETL) systems simply can’t effectively handle the complex new requirements of Big Data analysis.

LACK OF CLEAR BUSINESS GOALS:
Many Big Data projects fail because they do not tie directly to specific business goals. This problem often arises at the outset, when the project is focused as a technology project rather than a means to address a specific business issue. As IT leaders drive the process separately, without appropriate guidance from business leaders, the ultimate goal can be muddled and lost.

The project can also be hindered due to lack of proper scope. An open-ended project can drift and never reach a defined endpoint. Some organizations will set up a Big Data platform assuming that insights can be used in multiple use cases, without focusing on specific issues. VP of Big Data Jim Kaskade calls this “boiling the ocean.” The result is a meandering project that fails to present concrete, actionable conclusions to help resolve specific business problems.
In spite of the failure of many Big Data projects, businesses can’t afford to abandon the value to be gained from these initiatives. They need to face and overcome the challenges Big Data projects present. One important way many organizations are addressing the challenges is by adopting a self-service, agile data management and advanced analytics data prep solution.

ENABLE SELF-SERVICE ANALYSIS:
The first step to effective Big Data projects is to minimize the need for data guru’s/statisticians in the data prep process, and get information into the hands of business leaders. An agile data management solution empowers business analysts, by eliminating reliance on conventional, cumbersome tools like ETL, Excel, Access© and SAS©. Self-service allows analysts more time to devote to uncovering actionable insights, rather than being mired in the details of prepping the data. Agile data management and advanced analytics also eliminates the need to request scarce IT resources that can slow down projects.

The increased use of self-service analysis tools is another indicator of the growing democratization of data and consumerization of analytics. Businesses are broadening BI and analytics to a wider group of non-IT users throughout the organization from the front lines to the C-suite. This expands the utility of Big Data and makes analytics more available to other members of the organization, allowing them to incorporate relevant insights into every day business processes.
PROVIDE DATA AGILITY/FLEXIBILITY:
Advanced data prep tools enable business analysts to get to the data more quickly to begin discovering solutions to business problems. Because analysts can rapidly explore alternate solutions, they are able to “fail quickly and cheaply” and discard results that don’t work and move on to other alternatives.

Advanced analytics tools allow business analysts to easily blend new, diverse data with traditional data sources and apply business rules to handle data quality, conditions, and exceptions. They are able to create visual analytic models to automate various analytic processes like data cleansing and data quality. The end result is consistent and precise insights on how to improve every aspect of the business.

DEFINE SPECIFIC PROJECT GOALS:
An InfoChimps survey of 300 CIO’s revealed that 58% of CIOs attributed a Big Data failure to an inaccurate scope of the project. As the saying goes, if you don’t know where you’re going, any road will take you there. Each project needs to have a clearly defined ROI that justifies the expense. Research from The Wikibon Project determined that the average ROI for a Big Data project was $.55 for every dollar spent. They found that Big Data winners focused on operationalizing and automating projects.  

The key to a successful Big Data project is to ensure that business leaders define clear goals and objectives. Before collecting data, businesses need to have the end result in mind—what problem is being addressed and what is the value in solving it.

58% OF CIOS ATTRIBUTED A BIG DATA FAILURE TO AN INACCURATE SCOPE OF THE PROJECT.

---

6 Analysis from the Wikibon Project, by David Floyer, Wikibon.org, January 2015.
Organizations who hope to succeed in an increasingly competitive marketplace need to make Big Data analytics an integral part of their operation. To make sure these efforts meet targeted business goals, companies must provide easy access to data and flexible tools to data analysts to allow them to turn information into actionable insights.

Companies must build a collaborative partnership between IT and business managers to automate processes and develop a balance between demands for business speed and agility with IT governance and control. Rapid advances in tools and processes are enabling this shift and helping organizations capture Big Data insights through a more agile and analytics driven business culture.
Lavastorm is the agile data management and analytics company trusted by enterprises seeking an analytic advantage. The Lavastorm data discovery platform empowers business professionals and analysts with the fastest, most accurate way to discover and transform insights into business improvements, while providing IT with control over data governance.

The Lavastorm Analytics Engine’s visual, discovery-based environment enables organizations to reduce analytic development time by up to 90%. The core strength of the Lavastorm Analytics Engine is its ability to easily blend data from multiple sources, then apply business rules to ensure data quality and handle exceptions. Businesses can then continuously run the visual analytic models they create, allowing the automation of various analytic processes, such as data cleansing and data quality. The end result is an analytics infrastructure that supports larger amounts of data, responds with more speed, adapts to greater change, and empowers a growing number of decision makers.

For more information, please visit www.lavastorm.com.