



LAVASTORM  
analytics



# Why Most Big Data Projects Fail

Learning from Common Mistakes to Transform  
Big Data into Insights

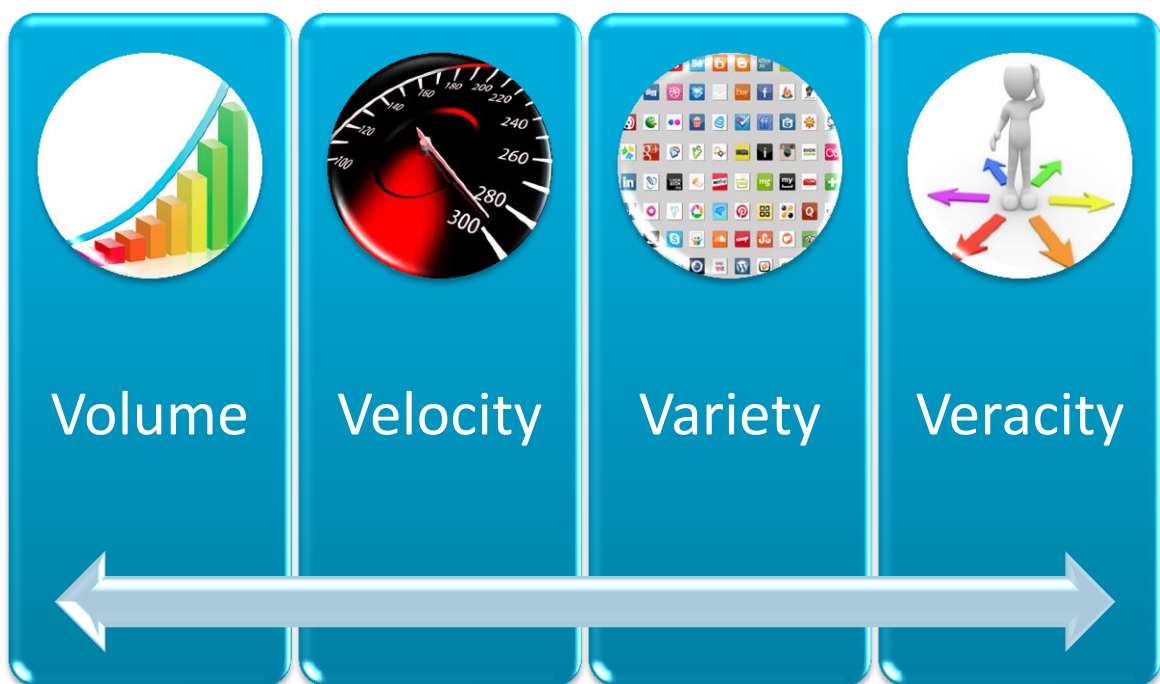


<b>What is Big Data? .....</b>	<b>2</b>
<b>Three Reasons Why Big Data Projects Fail .....</b>	<b>3</b>
<b>How Can Big Data Be Used? .....</b>	<b>5</b>
<b>The Lavastorm Approach to Big Data .....</b>	<b>5</b>
<b>Future Outlook.....</b>	<b>6</b>
<b>Getting Started With The Lavastorm Analytics Engine.....</b>	<b>6</b>

## What is Big Data?

Big data is often talked about as a “problem” because it cannot be easily processed with traditional systems based on relational databases. In reality it should be considered a tremendous opportunity to enhance and even transform how you run your business. It can lead to innovations, such as new pricing models, new ways to engage with your customers and partners, operational efficiencies, monitoring compliance and risk, new market opportunities or a better understanding of your business. But, how do we really define big data? Industry analysts<sup>1</sup> define big data using the four V’s:

Figure 1 – The Four V’s of Big Data



**Volume** represents the actual amount of data available. As this data is usually collected over a period of years, the sheer magnitude can be extremely high. According to TechCrunch<sup>2</sup>, 90% of the data in the world today has been created in the last two years alone.

**Velocity** refers to the speed of change in the industry, and taking action on relevant data today. 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.

<sup>1</sup>“3D Data Management: Controlling Data Volume, Velocity and Variety”, Douglas Laney, Gartner, 2001

<sup>2</sup>“If You Think Big Data’s Big Now, Just Wait”, Ron Miller, TechCrunch, 2014

**Variety** refers to the different types of data sources and forms. Organizations today collect data from a vast array of sources such as Websites, Point-Of-Sale systems, CRM systems, ERP systems, Excel® files, social media, legacy systems etc.

**Veracity**, the newest Big Data V, stands for the uncertainty about the data available. Due to the volume, velocity and variety of data, how can we be sure that the data we are using is accurate? Today, more business leaders do not trust the data they use to make decisions than ever before.

Using this definition of big data, let's learn from the mistakes of others and transform big data into insights.

## Three Reasons Why Big Data Projects Fail

Big data holds the promise for organizations to gain business advantages, yield new insights, make better decisions, and optimize business processes. Spending on big data initiatives is growing rapidly based on this promise, but the potential downside has been ignored. According to Gartner Research<sup>3</sup>, in 2013, investments in big data continued to rise, with 64% of organizations investing or planning to invest over the next year. However, according to InformationWeek<sup>4</sup>, in 2014, 92% of organizations are still stuck in neutral, either planning to get started or avoiding big data projects altogether due to the high rate of failure. So, what causes big data projects to fail?

1. **Big data is treated like a project with a known beginning and a known end rather than an agile exploration exercise.**



What are the patterns that will drive tomorrow's business success? To determine the answer takes exploration with a starting point, but not a deliverable in mind.

Gaining value from big data requires getting it into the right people's hands. Sometimes that is a data scientist looking for complex patterns, but if that's the case, he needs to pass it off to someone else to operationalize it – to use it day-to-day. The true value of data is unleashed only when it is put in a business context – otherwise it's just A LOT of data.

The availability of data scientists is at an all-time low – between 140K and 180K data science positions will remain unfilled by 2018<sup>5</sup>. Analysts empowered by agile solutions are the alternative but they are bound by conventional tools like ETL, Excel, and SAS, and are in high demand. So, as with most projects, there are limited resources and the team is a mix of full time employees (FTEs) and contractors with differing levels of experience.

<sup>3</sup>Gartner Survey, Gartner Symposium/ITxpo, Rob van der Meulen, Gartner, 2013

<sup>4</sup>"8 Reasons Big Data Projects Fail", Matt Asay, InformationWeek, 2014

<sup>5</sup>"Universities Can't Train Data Scientists Fast Enough for CIOs", Clint Boulton, The Wall Street Journal, 2014

To be successful, there needs to be flexibility to explore the data without restrictions. This includes flexibility in the analytics software to blend, integrate and visualize data from multiple sources, the business methodology to experiment with new theories and the working relationship between business and IT to make the analytics process seamless.

**2. The ROI hurdle is too high for big data initiatives. There needs to be flexibility and agility to fail cheaply and learn from mistakes.**



Traditionally, IT and BI organizations have spent almost all of their efforts caring for and managing but a small portion of the data available to business decision makers. Data from distribution partners, customer surveys, contracts, emails, government studies, real-time sensors, and customer social media data is estimated to represent a major portion of all enterprise data and is often ignored. This data is where the uncaptured insights for profit and revenue gains and operational efficiencies lie.

Traditional tools and methodologies have very limited application for big data. The enterprise data warehouses and other BI infrastructures didn't perform well when faced with the volume, variety or speed that characterized much of that data. Additionally, the turnaround time and cost for extending or making changes to existing data flows was unacceptable to the business. Further the ability to operate with partial requirements or recover from a failed initiative was non-existent.

Finally, integrating and blending big data from multiple sources also proved to be very costly and inefficient for organizations. This led a lot of business users to create their own data marts and spread marts oriented to a specific business line or team, in turn creating multiple versions of the truth and leading to further distrust of the data. Organizations invested heavily to include big data projects into a roadmap based on the fallacy that the EDW is achievable. Yet the big data promise is about high volumes of data of varying, even indeterminate value that could not possible go through the slow and expense ETL/EDW process, or even have heavy IT involvement.

**3. Big data projects are not tied to a business need. They are considered just for scientific purposes with no business goal or metrics.**

While big data needs to be an exploration exercise rather than a project, there is still a need to tie it to an overarching problem that matters to the organization. You need to be able to showcase results in order to justify ongoing big data investments, including the ability to continuously reap the benefits from both the vast scope of big data and the contextual information of enterprise data. In such cases, it is important to operationalize big data by applying it to operations, transactions and processes within your organization. The revolution today is around how the business needs, and must have, rapid and agile data access with a very low cost barrier for data-driven discovery and exploration.



Additionally, there is an ongoing debate about how big data doesn't stand alone. For example, a big data project to determine customer sentiment from call center records, Facebook posts, and tweets may be hard to act upon unless you can determine from which segment of your existing customer base it is originating and how profitable those products or customers are.

## How Can Big Data Be Used?

Big data can have a major impact on your business in three ways. It can help:

1. **Discover hidden insights** – For example, if you consider customer survey data when investigating a high service cancellation rate, you may detect a pattern or root cause not previously visible that can improve retention.
2. **Improve decisions, by enriching information for decision makers** – For example, if you consider a customer's social media profile, you can get a clearer picture of that customer and their place in the world. You can use that information to prioritize, improve customer satisfaction or create fraud alerts.
3. **Automate business processes** – For example, a bank can diagnose and improve existing intercompany reconciliation processes to provide ongoing governance of internal transaction processes and increase transparency in compliance reporting.

## The Lavastorm Approach to Big Data

Business users need to perform rapid data blending and integration, answer ad hoc questions, and respond to changes in the environment to fail cheaply when evaluating hypotheses. Now, with Lavastorm's sandbox approach to big data analytics, failing cheaply is a reality! The Lavastorm Analytics Engine has been proven successful across multiple industries for these types of use cases for years. With more than 15 years of use in the communications industry, the product has already generated billions of dollars in business value in the face of extremely high data volume, variety, velocity and veracity.

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 blend together new and diverse data alongside traditional data with tremendous ease, and then easily apply business rules to deal with data quality, conditions and exceptions. You can then continuously run the visual analytic models that you create, allowing the automation of various analytic processes, such as data cleansing and data quality.

*“Many companies are now relying on Hadoop and MongoDB to hold vast amounts of data, but they struggle to make it accessible to business users who know how to work with that data to improve their businesses. Lavastorm gives us the ability to ‘operationalize’ these data sources so that we can bring this data closer to business operations and use it to draw various insights on how to improve operations in sales, marketing, customer service, strategic planning, and other departments”*

*- VP and Global Head – Analytics, Mphasis*

Along with these capabilities, the Lavastorm Analytics Engine offers the following unique features to address big data project requirements:

Big Data Project Requirements	Lavastorm Analytics Engine Solution
Data Flow Approach for Agile Exploration	<ul style="list-style-type: none"> <li>Acquires virtually any source data – no schema required, just join and go</li> <li>“Fine” analytic control allows you to profile, inspect, and transform data any way necessary</li> <li>Visual models create an audit trail and transparency that builds trust in the results with managers/peers</li> </ul>
Flexibility and Agility to Fail Cheaply	<ul style="list-style-type: none"> <li>Reduces time of analytic development by 90% or more</li> <li>Desktop and server configurations provide a cost-effective solution entry points for any analytic application</li> <li>Designed for business analysts, pre-built functions and visual development empowers analysts to deliver results</li> </ul>
Operationalize Big Data	<ul style="list-style-type: none"> <li>Automates processes, including data blending/integration, data quality, exception handling, and analytics</li> <li>Easily extract data from Hadoop and MongoDB and blend with other data sources</li> <li>Small start-up costs mitigates risks and encourages greater use of analytics leading to a data-driven business culture</li> <li>IT friendly – offers IT governance capabilities while enabling business self-service</li> </ul>

## Future Outlook

A key aspect of big data like all data will remain exploratory. Organizations need to make big data analytics an integral part of their everyday workflows, which means continuously running analytics and automating processes. For this to work seamlessly, organizations need to tear down the walls between IT and business, remove hurdles, to strike a balance between demands for business speed and agility with IT governance and control. Tools such as the Lavastorm Analytics Engine help organizations capture big data-based insights while enabling a more agile and analytics-driven business culture.

## Getting Started With The Lavastorm Analytics Engine

To fully experience the benefits of the Lavastorm Analytics Engine in your organization, download a free 30-day trial today at: <http://www.lavastorm.com/resources/software-downloads-trials/>. Find out more about how Lavastorm can solve your big data challenges by visiting us at <http://www.lavastorm.com> and connect with our big data analytics experts.