

DATA DELUGE

Taming the World of Big Data

A SCALABLE WHITEPAPER

CATEGORY : DATA



SCALABLE
AI

TABLE OF CONTENT

INTRODUCTION.....3

WHY CARE ABOUT BIG DATA?.....4

THE THREE V’S OF BIG DATA.....5

HOW TO APPROACH BIG DATA.....7

THE BASICS OF WHAT IS AND WHY IT MATTERS.....8

VISUALIZE DATA AND COMMUNICATE INTELLIGENCE.....9

BEAUTIFUL VISUALIZATIONS10

POTENTIAL FOR BIG DATA.....11

BEAUTIFUL VISUALIZATIONS12

CONCLUSION.....13

ABOUT SCALABLE AI.....14

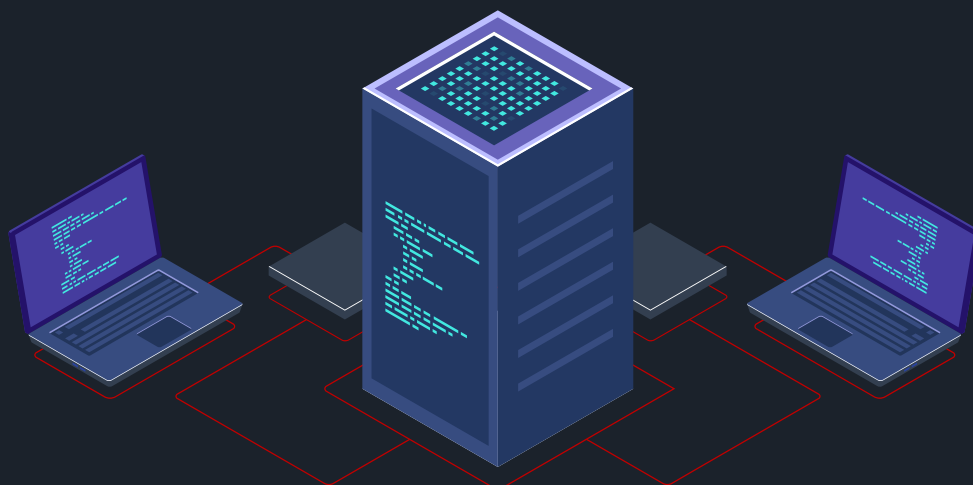
INTRODUCTION

Corporations from every industry worldwide are currently dealing with the rise of Big Data. Information stored from customer transactions and business operations is clogging data systems and puzzling decision-makers. Unstructured information from social media, reviews, and online shopping is also puzzling certain corporations while others use it to benefit. A report from the World Economic Forum titled “Big Data, Big Impact” proclaimed that data is becoming a lucrative economic asset viable with gold and currency. Big Data is a term that refers to this rapidly growing amalgam of stored information that is growing too large to be handled by conventional database tools, and it is currently redefining the future of entire industries, governments, and societies.

One main problem with this huge mass of data is that companies do not know how to analyze and utilize the information effectively, leaving them with a significant blind zone for data analysis. Well-applied data insights extracted from the mines of unstructured data can highlight trends in purchasing, patterns in customer service problems, assessment of social media campaigns, and even

monitor the social web chatter about your product. The hard part about data science involves asking insightful questions that increase success. Although there is high value and potential in combining Big Data analytics with traditional business intelligence, the truth is that conventional data mining and data warehousing techniques do not perform as needed. Because of the limits of the status quo database management systems, a Big Processing approach to Big Data is essential to utilizing massive datasets.

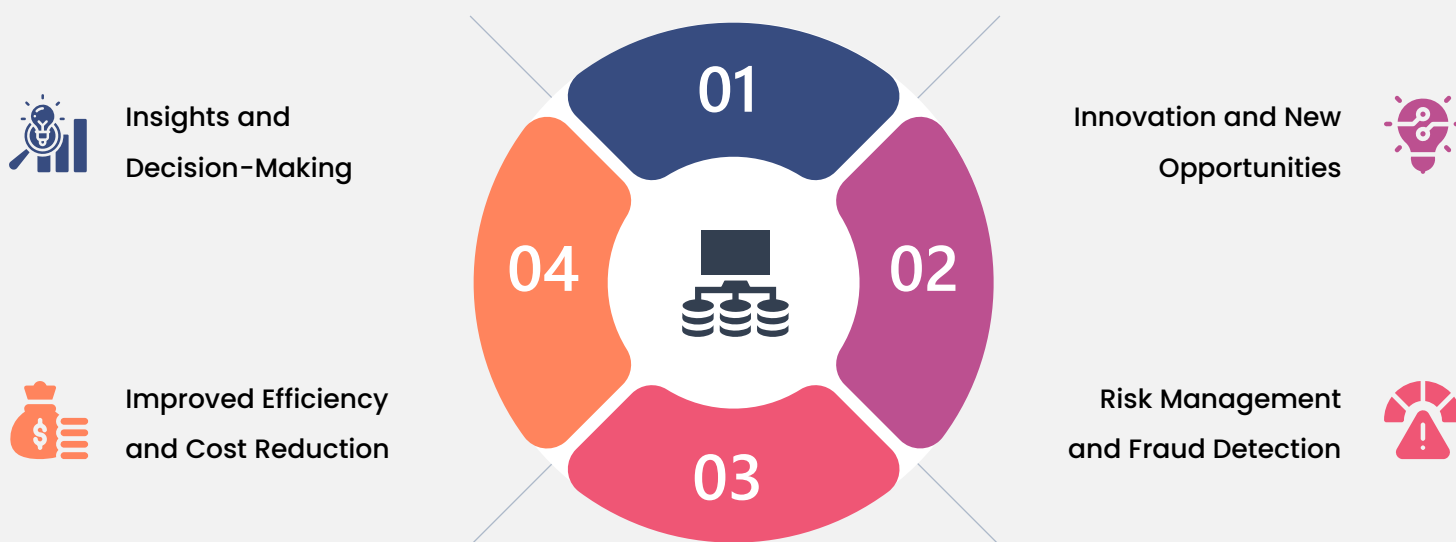
In a world where every individual connected to a smartphone can serve as a node in a massive network of potential customers, people can give instant feedback on a product, service, or government. Experts refer to the explosion of data accumulation as a primary contributing factor to the feedback economy. That feedback provides insights into consumer behavior trends and patterns that can prove invaluable to a corporation. We now live in an era where a company's stock performance and reputation can be stifled by a passionately negative YouTube video or Google review that has gone viral.



WHY CARE ABOUT BIG DATA?

Businesses across the globe are facing the same cumbersome problem: an ever-growing amount of data combined with limited IT architecture to interpret and analyze it into meaningful and insightful information. Due to the high volume, density, variety, and velocity of knowledge within data streams, companies need help finding efficient and productive ways to manage Big Data efficiently. Throughout any given day, a company may store terabytes of data ranging from customer transaction records, website referrals, traffic information, and more. On any given day, a 747 airliner amasses ten terabytes of data for each 30-minute interval of flying. The NYSE produces about one terabyte a day. And in the last two years, human beings have created more information than we have in the entire history of our species. [1] Dealing with this information overload will decide who wins in business.

Big Data is complex but can be processed with help from the existing IT architecture that most corporations already have. Options such as virtualizing servers through cloud processing technology are becoming commonplace in addition to onsite data warehousing efforts. Scalable-Systems focuses on the challenges of cleaning, storing, searching, and visualizing unstructured, structured, and raw data and then presenting it to provide incredibly valuable insights into a company's operations. As mentioned, traditional data warehousing systems are structured to deal with less than terabytes of data. While dealing with the sheer scale of the data, it is important to focus on analytical and efficient means for Big Data processing. Discovering value is the main goal of Big Data analytics, but the main trouble involves mining through massive data sets with no clear start or end point.



THE THREE V'S OF BIG DATA

Applying Big Data effectively to a corporation's decision-making involves a holistic approach. Like how the "cloud" encompasses various aspects of different technologies, Big Data also involves many components. It boggles the mind when contemplating all the possible applications and sources contributing to the zettabytes of information that comprise Big Data. Public transportation systems, financial transaction records, and social media chatter create terabytes of data every hour. Even forms of product views and traffic sensor data can serve as potential sources of Big Data. Value could be the fourth V,

but it cannot be extracted without attention to the data's volume, variety, and velocity. Before determining insights, the enrichment and data cleansing process must occur, or else the process will be like finding a needle in a haystack.

Due to the volume, variety, and velocity of all this information, it is important to break up data by these three characteristics. The three V's are vital to understanding this new field of information infrastructure and the software platforms used to harness their potential.



UNDERSTANDING THE THREE V'S OF BIG DATA

VARIETY

Raw structured and unstructured data could be more organized, manageable, and ready for analysis. The type of data for potential analysis ranges from social media discussions, YouTube videos, and customer complaints about the automated phone system. Due to the diversity and variety of data, analytics needs to produce valuable insight-hunting methods for specific organizations. Traditional data analytics involves cleaning the data by sifting through and extracting useful information while destroying the rest. With Big Data processing, it is possible to keep everything. Storage limits confine traditional datasets, while Big Data processing systems can save every bit of data exhaust an organization could desire. Certain data types require specific databases with advanced methods for aggregating data and converting it to meaningful insights.

VELOCITY

Speeds of data transmission and acquiring the fastest fiber-optic connection possible were once problems and concerns only associated with large-scale financial markets and mega-corporations such as Google and Wal-Mart. In today's hyper-connected feedback economy, businesses require simultaneous streaming of live information sets to capitalize on data's rapid inflow and outflow. Organizations that can rapidly and efficiently utilize and apply information to their goals will gain the upper hand. For example, through live data streams of real-time customer sentiment and discussion data, it is possible to watch a brand's online popularity rise and

hopefully not fall, similar to a hospital patient's lifeline.

VOLUME

Volume is certainly relative when considering the size of some corporations today, but a large amount of potentially relevant stored information is ready and waiting for analytics to extract insights. Although data analytic skills are integral to data science, possessing that large initial set of data to work with is the key to predicting demand and optimizing business. Big Data transforms the common prediction modeling process simply because of the many potential factors that can be considered. The volume of millions of bytes of data is the first and foremost obstacle to existing IT systems. Traditional querying techniques must be updated due to the need for scalable storage systems to handle this vast amount of data. Distributed querying methods dealt with through multiple servers via software platforms, such as the open-source Apache Hadoop, are the key to processing the massive amounts and types of data companies already possess.

Information infrastructure has moved mountains due to the desire for organizations to possess the fastest fiber optic wire connection possible to shave off a few milliseconds for less latency and greater competitive advantage. Mobile technologies have transformed how societies communicate, share information, and participate in the economy. It is possible to provide almost instant solutions to customer or logistic-related problems through sentiment analytics and additional methods.

HOW TO APPROACH BIG DATA...

Locality and latency are still issues for competitors in today's hyper-connected world. Updating data is often slow and painstaking. Datacenters are now more than ever an integral factor contributing to the success and failure of present-day and future competitors. Real estate companies can comb huge data sets for good prospects on loans and red lights for negligent borrowers. Restaurants can analyze food sales by specific items to maximize placement on the menu. Efficient means for organizing and cleaning large data sets are vital to an organization's Big Data planning success. Leading data scientists often highlight that they devote most of their efforts to tidying up massive amounts of data. Analysis of Big Data extracts information that was not explicitly designed for BI. The emergence of data scientists has truly revolutionized what business and social intelligence can be. Data science involves a hybrid melding of programming, scientific insight, and math skills. Although software and hardware are important, the personnel and human analysis behind all the hard work are essential to a successful Big Data strategy.

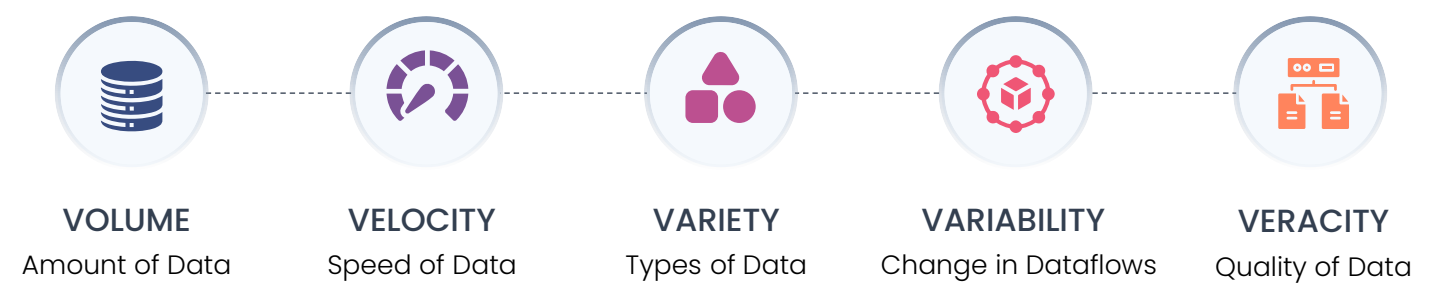
For example, Major League Baseball teams benefit highly from data science when asking

how many days of rest result in optimal performance for a pitcher. A great way to think about data scientists is to visualize a hybrid statistician possessing superior technical expertise, a curious mindset, and the ability to tell clever and insightful stories about a business, its customers, or competitors. Valuable insights are buried beneath mines of data, and it is the job of data analysts and scientists alike to spot trends and patterns and assist in visualizing the data in a way that best communicates the intelligence held within Big Data. Big Data is not an enigma. Even the "gray data" or information pulled from various sources without a specific need for traditional data intelligence can be analyzed and interpreted. With the right expertise and talent, it is possible to identify changes in customer tastes and anticipate stress on suppliers through analysis of weather and traffic conditions with minimal cost and high efficiency. Whether it is an inquiry into the effectiveness of an advertising campaign or input into customer satisfaction with an online customer service system, Big Data processing combined with effective visualization and data analytics can provide unheralded insights into your organization's past, present, and future.

THE BASICS OF WHAT IS AND WHY IT MATTERS

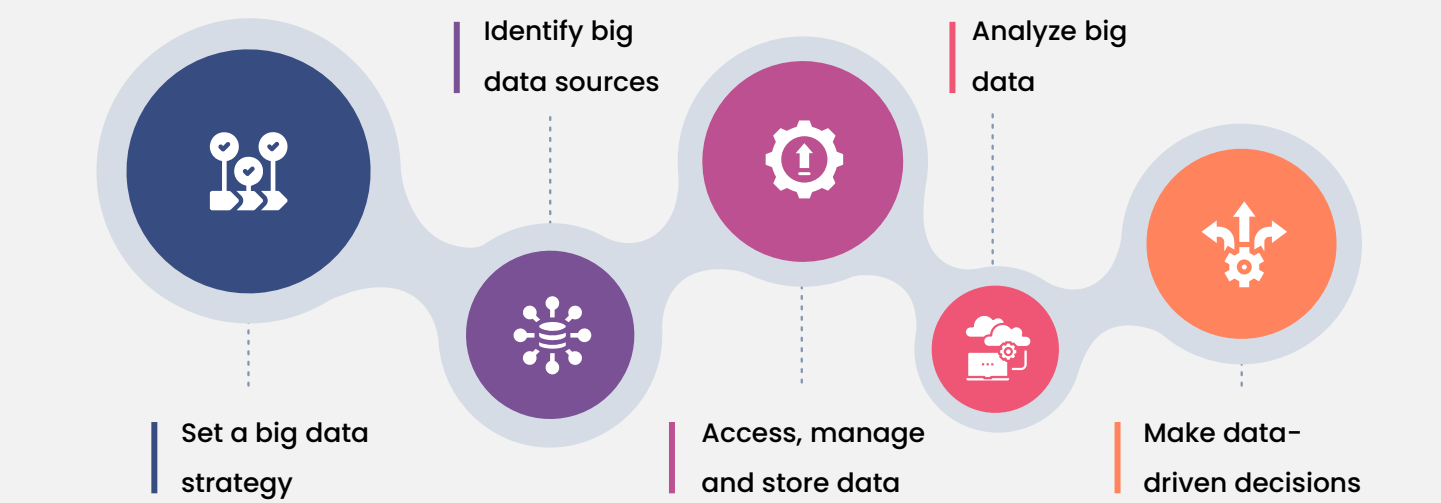
The 3 V's of big data: Plus 2

These are the defining properties or dimensions of big data.



How do organizations optimize the value of big data?

These are the defining properties or dimensions of big data.



Trends in big data

These are the defining properties or dimensions of big data.

| | |
|---|---|
| Mobile and real-time data dominate | By 2025, over a quarter of data will be real-time in nature and IoT real-time data will account for more than 95% of it |
| Artificial intelligence transforms the norm | Insights are generated via new technologies like machine learning and natural language processing |

VISUALIZE DATA AND COMMUNICATE INTELLIGENCE

Thousands of rows of numbers combined with countless subcategories are too difficult and time-consuming to read. Instead of a confusing spreadsheet, imagine a live updated visualization of customer service inquiries and feedback surveys as they occur. Or even a seamless visualization of customer transactions online? It is important to focus on imagery to communicate efficient yet insightful intelligence to members of different teams and departments. Visualizations of complex data sets are the most effective way to share highly complex statistical information. Incredibly important decisions are made based on the information provided by performance dashboards of visualized data. The information must be properly organized and displayed to comprehend the insights efficiently.

Making sense of complex data sets has always been challenging through data visualization. Professionals with a design and visually oriented mindset are vital to successful data visualization. Data science requires considerable skills spanning

technical prowess but may need more artistic vision. Graphic artists and media professionals understand the visual process.

Hiring designers and creative personnel who understand the visual process can turn data sets into powerful and insightful visualizations.

Explaining Big Data's insights requires much more than the standard bar graph can offer. A great deal of visualization does not provide the help and intelligence desired and needed by the person inspecting it. When imagery is the most efficient means of relaying such vast amounts of information, the personnel and expertise you put in charge of such tasks are essential investments. Visualization can provide a competitive advantage during high-risk decision-making. Due to its importance, it is important to know the differences between visualization to explore data and visualization to explain it in detail:

BEAUTIFUL VISUALIZATIONS

A perfect analogy would involve comparing the mining and analysis of data to the extraction and refinement process of a precious metal or gem. At first, the valuable material is in flux, but a practical and expensive result is created through various stages of production and division of labor efforts. In essence, effective visualization of Big Data is worth far more than its weight in gold. Insights that can alter the course of a quarter or a corporation's five-year plan can stem straight from insightful data visualization. Asking the right questions and seeking insightful answers are central to Big Data analysis. Understanding correlations and patterns

that encompass the actions that create data can provide real-time seamless insights into the real-world interactions and discussions that shape the vitality of a business. A time-tested example of a successful visualization would be the classic Periodic Table of Elements. With a few seconds of staring at the image, scientists and young students alike can easily comprehend the categorization and grouping of different elements. Data visualization can already accomplish so much with small sets of data. The true challenge comes with doing the same for petabytes and zeta data.



EXPLORING

Data visualizations to explore are highly effective when approaching a data set that needs to be well-known. Exploring occurs when data scientists attempt to discover what potentially valuable data lies within a group. Although it is a great start to Big Data visualization, exploring is often imprecise. Before extracting meaningful information and insights, the data must be cleaned and organized. The first step of exploring allows for the iteration of possible patterns and trends within data sets.



EXPLAINING

Interpreting the valuable information within a data set can only happen once data is cleaned and prioritized. Once data scientists have sifted through the mines of data, they can effectively communicate insightful patterns and trends. Once the professionals behind the scenes have contemplated the data and analyzed it for what's worth the valuable intelligence, the results can be visualized for decision-makers.

POTENTIAL FOR BIG DATA

In 2008, Google had the upper hand over the Center for Disease Control when they noticed an increase in flu activity via their search engine trends for specific geographic areas. Data is not only valuable; it can also be lifesaving. Even the logistics of a company website's automated services can be improved through synchronized updates of customer experience issues that automatically show up in a visualization dashboard to alert necessary departments. Consumer data from social intelligence is incredibly important; when combined with traditional market research, the possibilities for providing profitable products or services to customers are endless. Invaluable insights from social data analytics would make marketing directors of the past green with envy. Real-time streaming visualization dashboards can provide instant access to live customer sentiment and purchasing patterns. Energy companies can now monitor fluctuations in usage and prevent blackouts and power surges in real-time by tracking data visualizations. Free detail-rich data is now available for instant access within the mines of data accumulated by consumers' online interactions and conversations. Corporations can monitor conversations regarding a product or business, providing in-depth sentiment analysis, modeling metrics, and insights directly from the customer for a corporation's use. Text and social analytics also monitor customer discussions and conversations to

find insights regarding their sentiments toward a product, service, or event.

Web analytics, financial risk analysis, and active fraud protection are all examples of how Big Data analysis can augment the human analytical process. The potential applications for data science are limited only by the creativity and ingenuity of the personnel involved. There are many sources for Big Data analytics, but business and social intelligence held within large datasets are by far the most useful for a corporation or organization. Social networking sites alone create and store terabytes of data every single day. This mass of Big Data is immensely profitable for those willing to mine through it. Due to human communication embracing social networking, the data exhausts produced from such sites create astonishingly valuable insights into consumer behavior and business intelligence. Future victors in the business world will certainly be chosen by the organization's ability to become adept at analyzing and utilizing the massive amounts of social data at their disposal. By approaching Big Data with a customer-oriented approach, organizations can effectively track the live customer sentiment pulse through instant streaming of data analytics and the resulting insights.

BEAUTIFUL VISUALIZATIONS

Conversations and discussions regarding products are constantly occurring somewhere throughout the web. Velocity is one of the Three V's due to the importance of agile decision-making in today's hyper-connected world. Due to the time-sensitive nature of this valuable data, industry professionals are likely to witness major innovations in streaming data processing. Live streaming up-to-the-minute dashboard visualizations are incredibly powerful in the hands of the right decision-maker. Streaming analysis of large-scale data sets can potentially become more efficient with useless information being discarded and high-value data extracted and stored autonomously. Although distributed file systems have made major advances in terms of the storage of data, it is impractical to want to keep every piece of information generated. Streaming processing can automate the deletion of low-value

data and the extraction of high-value data. Data is valuable, and Big Data analytics are proving it.

Data marketplaces are also expected to rise in popularity due to the desire for organizations to combine their data with additional valuable datasets. Weather conditions are a perfect case. By comparing consumer behavior to weather conditions, it is possible to discover patterns between the weather and purchasing habits. Original methods for acquiring additional data sets were once a hassle. With the advent of data marketplaces, it is possible to streamline all datasets and integrate them into an organization's data analytics process. 2012 will likely see the ease and convenience of accessing Big Data.



Big Data will become more powerful as new programming languages gain a foothold of support in Hadoop-type processing engines. Shorter programs that work effectively with multiple markup languages will make Big Data processing more interactive and comprehensible.



One of the main weaknesses of Hadoop distributed management systems is the batch-oriented nature of its processing which lacks a certain level of interactivity that many data scientists and programmers' desire. Data scientists especially require efficient yet powerful tools to do their jobs.

CONCLUSION

Better information is our competitive advantage. The exponential growth in the data and the demand to cut down the response time have posed a crucial challenge to industries year after year. Big Data represents a class of solutions that enable organizations to capture, analyze, and mine very large data sets to gain insight into various problem domains ranging from customer preferences to complex forecasting.

Extracting derivative value from data enables organizations to enhance productivity and competitive advantage. Scalable Systems provides

Big Data consulting and implementation services that can Unlock business insights from all your structured and unstructured data, including large volumes of previously activated data, and fill the gap between data deluge & interpretive analysis.

Scalable Systems has a strong focus and is an established thought leader in Big Data and Analytics. We use a global delivery model to help the organization to evaluate and implement solutions tailored to their specific technical and business context.



About Scalable AI

We deliver actionable insights that organizations can use to identify opportunities, manage risks, achieve operational excellence, and to gain an innovative edge.

www.scalableai.com

About Scalable Systems

Scalable Systems is a Data, Analytics & AI Company focused on vertical-specific innovative solutions. By providing next-generation technology solutions and services, we help organizations to identify risks & opportunities, and achieve sales and operational excellence to gain an innovative edge.

www.scalable-systems.com