What’s the Big Deal about BIG DATA
# Table of Contents

- WHAT IS BIG DATA? .................................................. 03
- THE FOUR V’S OF BIG DATA ...................................... 05
- BIG DATA TECHNOLOGIES ........................................ 06
- 2015 STATE OF SELF-SERVICE BI FINDINGS .................. 10
- BIG DATA USE CASES .............................................. 11
- 5 STEPS FOR BUILDING A BIG DATA STRATEGY .......... 14
- LOGI ANALYTICS & BIG DATA .................................... 15
- ABOUT LOGI ANALYTICS .......................................... 16
WHAT IS BIG DATA?

Big data refers to the ever-growing volume of data, increasing velocity in the generation of that data, and increased variety of types of data. In 2015, adoption of big data skyrocketed across all sizes of businesses. And it’s not just IT teams that are looking to take advantage of all of that information. Business users want access to these data sets as well.

There really is no one concrete number that describes how big data is. The big data environment continues to grow more complex as the volume, variety and velocity of data increases. New technologies have emerged, improving how data is stored, managed, and retrieved.

LOOK BEYOND THE SURFACE

Think of your company’s data as an iceberg. The data your operations analysts and data scientists regularly access for the most part floats above the waterline. But the vast majority of data remains below that waterline. That’s big data. More than half of organizations using big data analytics have to break big data into smaller pieces just to work with it. This extra process requires significant investment in data preparation and resources, which slows down time to get immediate insights.

Big data has really accelerated the change in how people approach business intelligence and analytics today. Valuable business data no longer just comes from business applications or is solely managed by IT. You have applications in the cloud that you subscribe to one day then switch off the next. Other sources of data such as social media and video are pushing data volumes and velocity at a much greater scale, which plays a big part of the big data story. As a result, many new technologies have emerged to handle these data challenges. Unlike their peers in the traditional data warehousing space, these technologies are designed to scale, to be more open and less proprietary in nature, and ultimately be more flexible in how they handle data. Finally, big data applications aren’t just about reporting anymore, and they increasingly offer greater levels of interactivity with the data. These applications need to be developed with higher levels of agility, such that you shouldn’t be constrained by traditional and rigid processes that model the data and manually optimize queries for performance.
The challenge for organizations is to know how to appropriately use these new data repositories to improve business performance by analyzing data in as close to realtime as possible. What’s even more daunting is that mobile devices, social media, and other tools and technologies add to an already enormous stream of both structured and unstructured data.

Tapping into the data that lies below the waterline can take weeks or months — if it can be reached at all using the organization’s capabilities. However, with the right tools, your organization can analyze 100 percent of its data quickly and easily, regardless of how far the data sits below the surface. You can blend disparate data sets and empower users to explore that data on their own without technical assistance.

ENTER THE CHIEF DATA OFFICER

While this role varies by industry and is still evolving in today’s corporate world, the Chief Data Officer (CDO) is someone who helps transform a business into a data-driven company. This person is responsible for mapping the particulars of a company’s data needs to its overall business purpose in order to create and drive value. They may also work with departments across the organization to ensure everyone is working toward the same goal. Technical skills, marketing expertise, and business acumen help a CDO to see the big picture on big data and elevate the importance of that data to the top of the organization.

Because big data is really more of a concept that characterizes the changing nature of data, we’re going to break it down into what are commonly known as its four dimensions.

**STRUCTURED DATA**

refers to any data that lives in a fixed field within a record or file. This is typically data contained in relational databases and spreadsheets.

**UNSTRUCTURED DATA**

refers to information that either does not have a predefined data model or is not organized in a predefined manner. It can be text-heavy, such as emails and social posts, but also includes everything from images to video to audio.
THE 4 V’S OF BIG DATA

V IS FOR VOLUME
(the amount of data)

While one number cannot characterize big data, a few interesting ones are worth noting. Recent studies predict there will be 40 zettabytes, or a trillion gigabytes, of data generated in 2020 – which is 300 times that in the year 2005. Regardless, big data can be an issue specifically for those users who hoard every one of their email messages, take loads of pictures, and record video after video. What happens when these users run out of disk space? With this in mind, big data becomes a concept that applies on a more personal level, and scaling it then becomes difficult on many different levels.

V IS FOR VELOCITY
(the speed of data change)

Consider the billion pieces of content that are shared on Facebook every day. In London, an estimated more than 6 million closed-circuit camera TVs are capturing video on a daily basis. Each video is captured at 30 frames per second, which equates to roughly 100 million frames per second in total – that’s over 15 trillion frames per day! In Major League Baseball, a system in every stadium captures the movement of the players and the ball on the field using advanced video and radar. This system generates approximately seven terabytes of data per game. That’s a lot of data that must be turned around for real-time analysis during each and every event. The analytics challenges presented by this velocity of data demonstrate that data is not just coming from business applications anymore. It’s coming from everywhere!

V IS FOR VARIETY
(the different forms of data)

Data comes in many forms. Whether it’s text, images, audio, or video - the channels they feed into can be easily distractible and hard to decipher. Now some of this data is unstructured, which means it isn’t ready to be conventionally processed and analyzed. But even when the data is structured, the fact that it comes from different places ultimately means each piece of data may have a different structure. Within the realm of business applications, resolving such data inconsistencies across changing systems must be addressed, whether through sales analytics tools, marketing tools, finance, HR, or ERP systems.

V IS FOR VALUE
(the value of data)

Information about a transaction has become even more valuable than the transaction itself. For example, as a retailer, you want to know the sequence of events that leads to a transaction (what marketing campaign worked, the customer’s click path on the website, and so on). All of this information can help build value by driving more transactions and building stronger relationships with customers. But value is never a straightforward path; you often won’t know how some of the data you have today can help you answer a question tomorrow.
BIG DATA TECHNOLOGIES

In just a few short years, big data technologies have gone from nothing more than hype to being one of the core single disruptors in the digital age. Many classes of technology dominate this new landscape. These include data storage and collection systems that provide operational capabilities for real-time, interactive workloads, and systems that provide analytical capabilities for retrospective, complex analysis that may tap most or all of the data. It’s these big data challenges that have presented the opportunity for new technologies to emerge – technologies designed to handle data in much greater volumes and variety and at greater speeds. Let’s take a look at the technologies helping today’s users tackle big data.

COLUMNAR DATABASES

Compared to relational databases, which store data in rows and offer fast reads and writes for use with transactional applications, columnar databases store data in columns. They support fast read operations and analytical capabilities. Columnar databases also employ data compression to handle large data sets and enhance performance.

As an example, consider sensor data or logging of machine data for a device at rest. Data always needs to be recorded, but the values themselves don’t change much. By storing the data in columns, and with the help of compression algorithms, the database doesn’t need to store each of these repeated values, thus helping with data volumes. Columnar data stores are built for SQL querying, which make them friendly to interfacing with BI applications.

Examples of columnar data stores include HP Vertica, Amazon Redshift, and Infobright.
NoSQL
NoSQL addresses the issues of data variety by storing data with JSON documents or key-value pairs in a flexible structure, rather than solely in tables. With NoSQL, you don’t necessarily need to specify or adhere to a fixed data structure. A record or document can be saved with some set of attributes associated with it, while the next record can have a completely different set of attributes. The database will ultimately understand how to store and query data from a data store.

Examples of NoSQL technologies include MongoDB, Amazon DynamoDB, and Cassandra. (NoSQL contrasts directly with relational databases, which have a very well-defined table structure and only read and write rows that adhere to that structure).

Hadoop
Hadoop stores data simply as files. The Hadoop Distributed File System (HDFS) offers the highest scale, where a Hadoop cluster can have hundreds or even thousands of nodes. Hadoop is designed for large-scale processing, which is performed by distributing operations across the multiple nodes, where each node operates against smaller subsets of data.

Examples of Hadoop include Cloudera, Hortonworks, and MAPR.

Search and Processing Engines
It’s useful to consider the big data ecosystem as more than a set of big data repositories, but rather a set of different technologies that may be implemented for specific use cases. For example, columnar databases, Hadoop, and NoSQL usually have to make tradeoffs to balance the need for different ways of storing data and performing analytics. Really, there’s a need to augment these capabilities in very specific ways such as through search, which requires a different type of processing engine that sits on top of the databases and works alongside them.

Examples of search and processing engines include HP IDOL search for unstructured data as well as SOLR, Elasticsearch, or even Spark for large-scale data processing.

Now that we have a basic understanding of these technologies, let’s look at some common ways they are used in the evolving world of big data.
At the bottom of this graphic, there are a variety of data sources. At the top, there are business intelligence and analytics applications — not a single monolithic application, but rather distinct analytic applications tailored to specific use cases. Relational databases and data warehouses, however, are not going away any time soon, especially when it comes to business applications.

Here, we introduce Hadoop into the picture. In many cases, Hadoop serves as a central database, and it’s not uncommon for businesses to pursue big data because they are implementing Hadoop somewhere in their company. That’s all well and good; but for the purpose of this discussion, it’s important to realize that implementing interactive or self-service applications directly off of Hadoop can be difficult. It’s more suited to large-scale and batch processing than interactive analysis. Now, companies with existing data warehouses can structure and move data from their Hadoop stores to their data warehouses for reporting — though they may run into scaling and flexibility issues depending on the architecture of their data warehouse.

Because NoSQL stores are designed for operational applications, they can also act as centralized data stores. While both NoSQL and Hadoop are great for big data, they’re intended for different types of workloads. NoSQL is popular with application developers due to its flexibility in handling data. For analytic applications that connect directly to NoSQL sources, there can be moderate interactivity from both structured and semi-structured data.

Hadoop, on the other hand, is about large-scale processing of data. To process large volumes of data, you want to do the work in parallel, and typically across many servers.
Much like typical relational databases, columnar databases can also be used as centralized stores for structured data. What’s interesting about columnar stores is that they are starting to take on the role of big data warehouses. It’s the columnar store that offers the most interactive types of self-service analysis with structured data. This is where the high-performance scale and the use of SQL really make analytics shine.

One of the main benefits of a columnar database is that data can be highly compressed allowing columnar actions to be performed very quickly. Columnar databases can be self-indexing, thus optimizing performance for self-service analytics and reducing the maintenance overhead of a database administrator.

Search engines enable us to create more interactive applications with unstructured data. To extract meaning from unstructured data such as tweets and images, users can perform text searches or utilize search engines for specialized algorithms – for example, to uncover underlying sentiment from tweets.

At a high level, we’ve covered some of the ways these technologies are utilized. But this is just the beginning of the story, as the big data space continues to evolve and new innovations are introduced.
We’ve established that leveraging big data is incredibly important for businesses. But is anyone actually using it outside of the analyst community? According to our recent 2015 State of Self-Service BI Report, the answer is yes. We asked more than 400 IT professionals which data sources they’re providing to business users engaged in self-service BI.

As you can see, relational databases and data warehouses are still very relevant. What’s interesting is that adoption of big data stores has increased year over year. In turn, this has exacerbated some of the major challenges of using big data. For instance, the blue bar in the graphic on the right indicates the expectation that IT will implement such data stores for self-service analysis within the next one to two years. In looking at that bar, we see a rapidly evolving data landscape when it comes to analytics applications and underlying data sources.

Here’s an encouraging point for big data: In our 2014 survey, the percentage of IT professionals who said they had or plan to invest in big data in 1-2 years added up to 30 percent. In this year’s 2015 survey, they total almost 40 percent. From our point of view, this validates the increasing investment in big data technologies.

The value of big data really presents itself when business users can easily see and work with data to:

- Make their jobs easier
- Lower the cost of operations for the business
- Drive revenue and gain a competitive advantage

Let’s look at three examples of big data use cases.
BIG DATA USE CASES

Let’s take a look at some common use cases that work with big data.

1. INTERNET OF THINGS
   The growth of the Internet of Things (IoT) has been exploding - changing the way businesses and consumers interact with the physical world. With so many connected devices generating so much data, there’s often a need to derive insights and meaning from this data. Use cases might include a data center with thousands of machines generating machine logs, or a healthcare facility with medical devices or sensors monitoring activity.

CASE STUDY EXAMPLE: GLASSBEAM

- **WHO?** Glassbeam is a big data applications company specializing in multi-structured machine data analytics for IT and business users.

- **WHY LOGI?** Glassbeam needed an embeddable application that would provide their users with dashboards and reports, but that also allowed for control over how elements were placed and located. They needed to bring data into a data center in order to monitor device usage and performance. They also needed more developer control over options like placement of charts and filters within reports.

- **RESULTS:** With Logi, Glassbeam was able to build and customize dashboards they can frequently enhance and modify with new visualizations, interactivity, and data sources. They’re able to provide value through capacity planning – helping end-users to ensure devices have enough memory, disk space, and processing power to operate, and to proactively predict device failures and ensure uptime. Users can also utilize this data to perform audits and intrusions where unauthorized access can be detected in real time.

“Glassbeam collects a wide range of unstructured data from complex machines and converts that data into structured data. We needed an analytics solution that would help us analyze large amounts of data and provide customized insights to our customers.”

- Vivek Sundaram, Solutions Architect, Glassbeam

READ THE FULL CASE STUDY
2. MEASURING BRAND PERFORMANCE
Social intelligence is necessary for gaining insight on how consumers think and behave. As social technology matures, social intelligence can help companies overcome some of the limits of older intelligence-gathering approaches. These are often used with traditional reporting and business intelligence methods to help organizations make better data-driven decisions.

CASE STUDY EXAMPLE: SOCIAL MEDIA

• WHO? A local social media agency

• WHY LOGI? This business had brand managers and customer support agents who needed help tracking everything that was said about their company, and understand if the sentiments posted were positive or negative. They wanted to proactively monitor the health of their brands and engage with individuals coming through their numerous channels (Facebook, Twitter, blogs, forums, etc.).

• RESULTS: Logi’s big data technologies made it possible for the organization to get value from all the data collected at rates much faster than ever before, making complex problems much easier to digest and take action on.
3. BUSINESS PROCESS COMPLEXITY
Technology can wrangle the complexity in a business process to deliver results faster. Service warranties, as an example, are provided by many different agents and channels, and sorting through these relationships can be quite complex. What’s more, different warranties may have different terms, and with the business expanding, these documents are constantly evolving.

CASE STUDY EXAMPLE: WARRANTY SERVICES

• WHO? A mid-size global warranty services organization

• WHY LOGI? This organization faced many challenges when trying to bring service documents together, and structure them into a relational database. They needed us to help create those complex joins which before, had proved to be a long, time-consuming process for them. Ultimately, solving this problem required a NoSQL data store in order to efficiently store and query such documents.

• RESULTS: We were able to help them deliver much higher value to their business by helping them identify potential opportunity by policy renewals, up-sales, and cross-promotions of warranty products.

As consumers, we all take for granted the excellent user experiences offered by the Facebooks, LinkedIns, Amazons, and the Googles of the world. We don’t necessarily sit back and thank them for using big data. We simply enjoy an intuitive, seamless user experience. In turn, this heightens our expectation that business applications will provide as much utility as consumer applications provide for us. Ultimately, that is what makes big data relevant to those who are looking to implement big data projects.
Now that you have an understanding of big data, the next step is to build out a plan to deal with it. Get started on your big data strategy with these five easy steps:

1. UNDERSTAND YOUR BUSINESS GOALS
First, identify the business problem or case your organization is looking to address and map it to the right benchmarks, metrics, and KPIs. For example, is your goal to optimize operational levels? Increase sales forecast transparency? Or monitor the performance of equipment across regional locations? Insights into big data can help your business achieve all of these objectives and much more. Big data also gives IT and the line of business an unprecedented opportunity to work together to increase productivity, efficiency, and business processes. By increasing accountability and collaboration across the business — along with clearly outlining requirements and priorities — you will best position your company to uncover the hidden value in your data.

2. HAVE A CLEAR STRATEGY
It’s important to be strategic in your implementation of big data technology so you can make the most of your existing IT infrastructure and prevent the new technology from becoming a siloed part of your organization. For instance, if you decide to move to Hadoop, then you need to subsequently choose a distribution player so you can deploy it. And, you need to select a big data analytics platform that can transform the raw data you put into Hadoop into real-time insights for the organization. Logi Analytics’ end-to-end platform enables you to run analyses across your company’s data — transactions, customer interactions, and machine data.

3. SELECT THE RIGHT PLATFORM
When selecting a big data analytics platform, ask yourself if it has the following attributes:
- The ability to gain insights from multi-structured data
- Tools that show you all of your data, not just what’s at the top of the iceberg
- Freedom from IT — the ability to ask the questions you want, when you want
- Fast answers, regardless of how much data you have on hand
- Access to big data for everyone — not just users with “scientist” in their title
- Tools built natively so the business can make the most of the data

4. START SMALL AND MEASURE
Once you have the ability to access and analyze information, the temptation to go big and analyze all the data in sight is hard to resist. Instead, be strategic. Pick one business problem, perform an audit to understand what data you need, and then measure that particular set of data for insights. Focus on small wins first, as this will help all employees fully understand the data in their everyday work. This strategy will also enable you to build the momentum to change your organization into a data-driven enterprise.

5. BUILD A DATA-DRIVEN CULTURE
When users feel empowered to ask questions of big data, companies can build a data-driven culture fostered by collaboration and innovation. With self-service analytics...
- Users can examine data from every touch point — from transactions to social posts — and make informed decisions faster
- The power and flexibility to get answers to questions is much easier, and groups can easily share that information with others
- Data scientists can make their work more accessible to the organization, which makes what they do more meaningful to the business
- IT professionals can stop worrying about the volume, variety, and velocity of data; whether users have access to the data they need; and whether or not that data is secure
Logi Analytics offers a powerful platform that simplifies self-service analytics by eliminating concerns around data performance and preparation. Logi DataHub enables you to connect directly to multiple data sources, cache the data for high performance, and prepare the data for analysis in intuitive ways. This gives you the ability to deliver efficient reporting and analysis that doesn't affect your transactional systems, allowing for more insightful decision-making.

DIRECT CONNECTIVITY
Logi works with a variety of big data repositories to ensure a high level of connectivity with many of the top-tier technology providers. For providers that are not included in our out-of-the-box connectivity – such as search engines that don’t have a BI-friendly interface – we offer a plug-in model that interacts with such engines and other proprietary interface stores via code. This enables you to quickly view, understand, and act on critical information without a need for additional data engineering or architecting.

NEW STRATEGIC PARTNERSHIPS
Logi has strategic partnerships with the industry’s technology leaders for analytical data stores, including HP Vertica, Amazon Redshift, ParStream, Hortonworks, and Cloudera. Additionally, we optimize some of the querying to leverage the high performance these data sources offer.

HIGH-PERFORMANCE DATA REPOSITORY
Our solution fulfills your needs to cache data, blend data from multiple sources, and/or enrich that data for analysis.

INTEGRATION
Logi has built-in query optimization for self-service analytics. Many highly interactive and self-service capabilities can run directly through the underlying data sources of your choice.

SECURITY
Our platform is extremely flexible, and we offer many different ways to support your security needs by helping to detect suspicious patterns and prevent fraudulent behavior.
ABOUT LOGI ANALYTICS

Logi Analytics is the leader in self-service analytics, delivering tools designed to meet the needs of users and product managers. At Logi, we are re-imagining how software can empower individuals, and the organizations and products that serve them, with analytics that can be embedded directly into the business applications people use every day. From interactive dashboards to ad hoc queries and visual analysis, Logi enables users to explore and discover insights and make data-driven decisions.

More than 1,750 customers worldwide rely on Logi Analytics. The company is headquartered in McLean, Virginia, with offices in the UK and Europe. Logi Analytics is a privately held, venture-backed firm.