BIG DATA VS. DATA ANALYTICS VS. DATA SCIENCE: WHAT'S THE DIFFERENCE?



Data has become the most critical factor in business today. As a result, different technologies, methodologies, and systems have been invented to process, transform, analyze, and store data in this data-driven world.

However, there is still much confusion regarding the key areas of Big Data, Data Analytics, and Data Science. In this post, we will demystify these concepts to better understand each technology and how they relate to each other.

Data TL:DR

- **Big data** refers to any large and complex collection of data.
- **Data analytics** is the process of extracting meaningful information from data.
- **Data science** is a multidisciplinary field that aims to produce broader insights.

Each of these technologies complements one another yet can be used as separate entities. For instance, big data can be used to store large sets of data, and data analytics techniques can extract information from simpler datasets.

Read on for more detail.

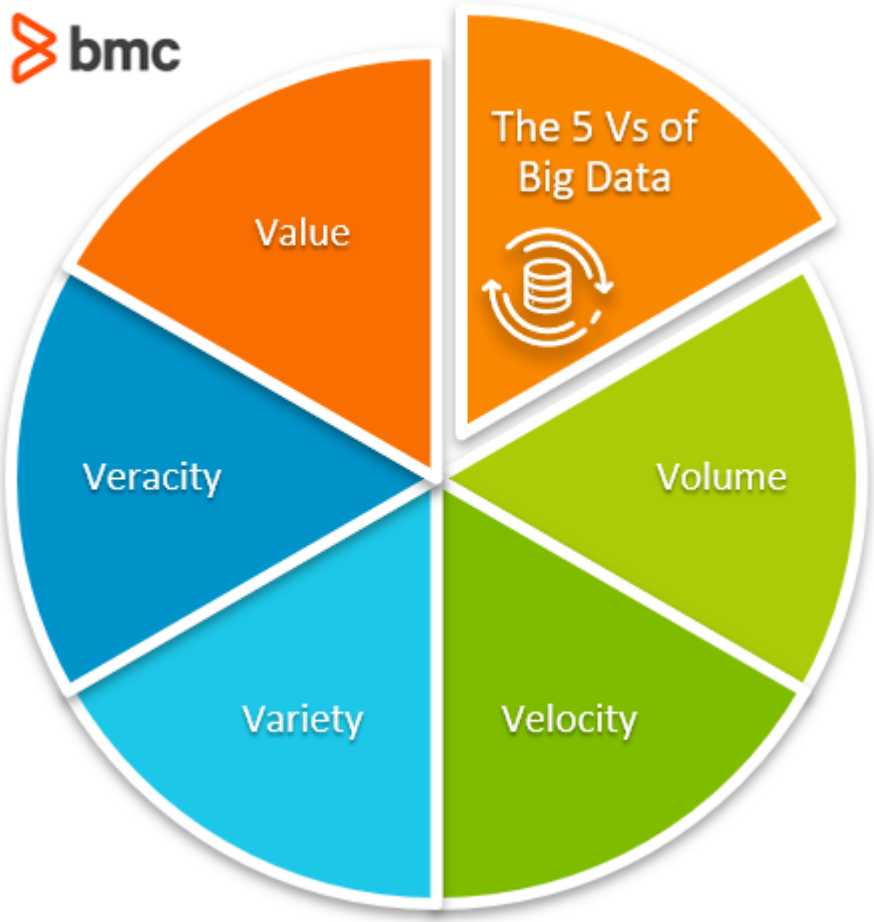
What is big data?

As the name suggests, big data simply refers to extremely large data sets. This size, combined with the complexity and evolving nature of these data sets, has enabled them to surpass the capabilities of traditional data management tools. This way, data warehouses and data lakes have emerged as the go-to solutions to handle big data, far surpassing the power of traditional databases.

Some data sets that we can consider truly big data include:

- Stock market data
- Social media
- Sporting events and games
- Scientific and research data

(Read our full primer on big data.)



Characteristics of big data

- **Volume.** Big data is enormous, far surpassing the capabilities of normal data storage and processing methods. The volume of data determines if it can be categorized as big data.
- **Variety.** Large data sets are not limited to a single kind of data—instead, they consist of various kinds of data. Big data consists of different kinds of data, from tabular databases to images and audio data regardless of data structure.
- **Velocity.** The speed at which data is generated. In Big Data, new data is constantly generated and added to the data sets frequently. This is highly prevalent when dealing with continuously

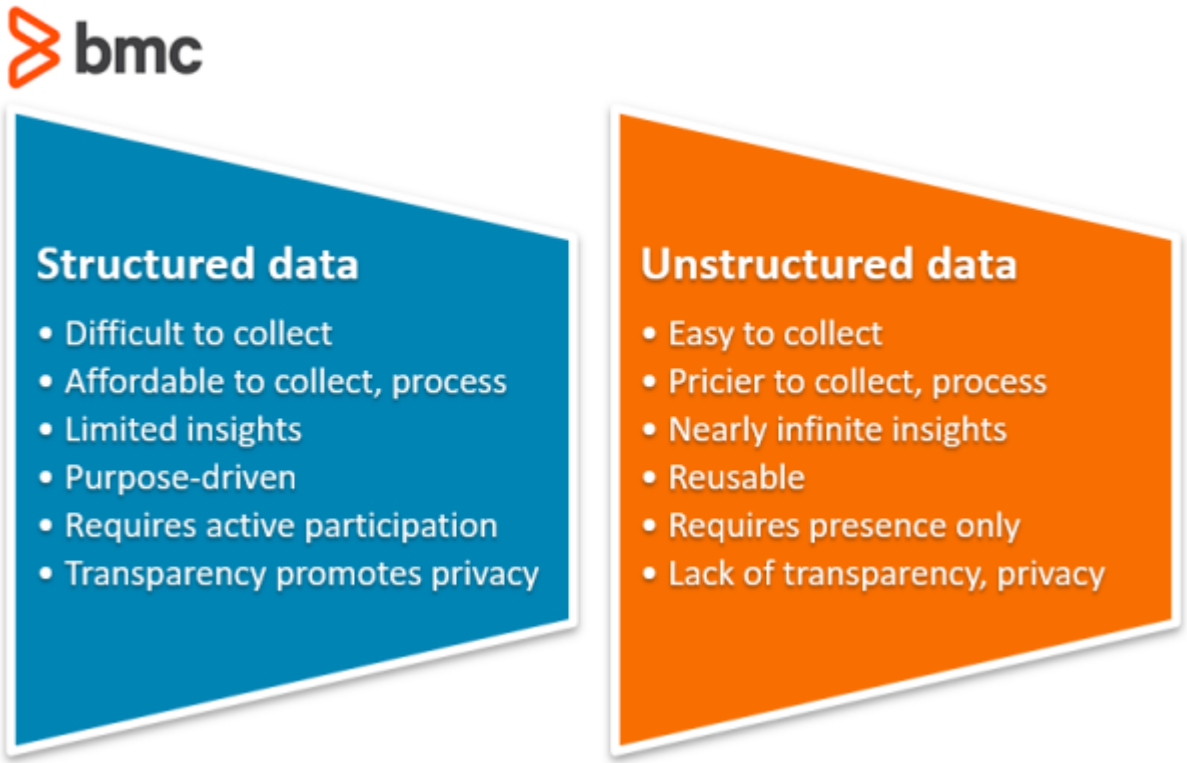
evolving data such as social media, IoT devices, and monitoring services.

- **Veracity or variability.** There will inevitably be some inconsistencies in the data sets due to the enormity and complexity of big data. Therefore, you must account for variability to properly manage and process big data.
- **Value.** The usefulness of Big Data assets. The worthiness of the output of big data analysis can be subjective and is evaluated based on unique business objectives.

Types of big data

- **Structured data.** Any data set that adheres to a specific structure can be called structured data. These structured data sets can be processed relatively easily compared to other data types as users can exactly identify the structure of the data. A good example for structured data will be a distributed RDBMS which contains data in organized table structures.
- **Semi-structured data.** This type of data does not adhere to a specific structure yet retains some kind of observable structure such as a grouping or an organized hierarchy. Some examples of semi-structured data will be markup languages (XML), web pages, emails, etc.
- **Unstructured data.** This type of data consists of data that does not adhere to a schema or a preset structure. It is the most common type of data when dealing with big data—things like text, pictures, video, and audio all come up under this type.

(Get a deeper understanding of structured and unstructured data types.)



Big data systems & tools

When it comes to managing big data, many solutions are available to store and process the data sets. Cloud providers like AWS, Azure, and GCP offer their own data warehousing and data lake implementations, such as:

- AWS Redshift

- GCP BigQuery
- Azure SQL Data Warehouse
- Azure Synapse Analytics
- Azure Data Lake

Apart from that, there are specialized providers such as Snowflake, Databriks, and even open-source solutions like Apache Hadoop, Apache Storm, Openrefine, etc., that provide robust Big Data solutions on any kind of hardware, including commodity hardware.

What is data analytics?

Data Analytics is the process of analyzing data in order to extract meaningful data from a given data set. These analytics techniques and methods are carried out on big data in most cases, though they certainly can be applied to any data set.

(Learn more about data analysis vs. data analytics)

The primary goal of data analytics is to help individuals or organizations to make informed decisions based on patterns, behaviors, trends, preferences, or any type of meaningful data extracted from a collection of data.

For example, businesses can use analytics to identify their customer preferences, purchase habits, and market trends and then create strategies to address them and handle evolving market conditions. In a scientific sense, a medical research organization can collect data from medical trials and evaluate the effectiveness of drugs or treatments accurately by analyzing those research data.

Combining these analytics with data visualization techniques will help you get a clearer picture of the underlying data and present them more flexibly and purposefully.

Types of analytics

While there are multiple analytics methods and techniques for data analytics, there are four types that apply to any data set.

- **Descriptive.** This refers to understanding what has happened in the data set. As the starting point in any analytics process, the descriptive analysis will help users understand what has happened in the past.
- **Diagnostic.** The next step of descriptive is diagnostic, which will consider the descriptive analysis and build on top of it to understand why something happened. It allows users to gain knowledge on the exact information of root causes of past events, patterns, etc.
- **Predictive.** As the name suggests, predictive analytics will predict what will happen in the future. This will combine data from descriptive and diagnostic analytics and use ML and AI techniques to predict future trends, patterns, problems, etc.
- **Prescriptive.** Prescriptive analytics takes predictions from predictive analytics and takes it a step further by exploring how the predictions will happen. This can be considered the most important type of analytics as it allows users to understand future events and tailor strategies to handle any predictions effectively.

Accuracy of data analytics

The most important thing to remember is that the accuracy of the analytics is based on the underlying data set. If there are inconsistencies or errors in the dataset, it will result in inefficiencies or outright incorrect analytics.

Any good analytical method will consider external factors like data purity, bias, and variance in the analytical methods. Data Normalization, purifying, and transforming raw data can significantly help in this aspect.

Data analytics tools & technologies

There are both open source and commercial products for data analytics. They will range from simple analytics tools such as Microsoft Excel's Analysis ToolPak that comes with Microsoft Office to SAP BusinessObjects suite and open source tools such as Apache Spark.

When considering cloud providers, Azure is known as the best platform for data analytics needs. It provides a complete toolset to cater to any need with its Azure Synapse Analytics suite, Apache Spark-based Databricks, HDInsights, Machine Learning, etc.

AWS and GCP also provide tools such as Amazon QuickSight, Amazon Kinesis, GCP Stream Analytics to cater to analytics needs.

Additionally, specialized BI tools provide powerful analytics functionality with relatively simple configurations. Examples here include Microsoft PowerBI, SAS Business Intelligence, and Periscope Data Even programming languages like Python or R can be used to create custom analytics scripts and visualizations for more targeted and advanced analytics needs.

Finally, ML algorithms like TensorFlow and scikit-learn can be considered part of the data analytics toolbox—they are popular tools to use in the analytics process.

Difference between data analytics and big data analytics

Size and scale are not the only things that distinguish big data analytics from ordinary data analytics. Understanding the differences will help you understand big data analytics concepts and the big data that powers it.

- **Scale**: Data analytics can handle small- to medium-sized datasets, whereas big data analytics is specifically designed for very large and complex datasets.
- **Tools and infrastructure**: Data analytics typically uses simpler tools and can often be performed on a single machine, while big data analytics requires distributed computing and more advanced, scalable tools.
- **Data types**: Data analytics often deals with structured data (like relational databases), while big data analytics often involves unstructured or semi-structured data (like text, images, or sensor data), in addition to structured data.
- **Processing**: Big data analytics often involves real-time or near-real-time processing due to the high velocity of data, while data analytics may not necessarily require such speed.
- **Functionality:** Big data analytics takes in more advanced capabilities, including machine learning and artificial intelligence, for in-depth analysis yielding valuable insights.
- **Security challenges:** Because of the size and scale of big data, third-party storage is the norm,

which requires much greater cybersecurity protections.

What is data science?

Now we have a clear understanding of big data and data analytics. So—what exactly is data science?

Unlike the first two, data science cannot be limited to a single function or field. Data science is a multidisciplinary approach that extracts information from data by combining:

- Scientific methods
- Maths and statistics
- Programming
- Advanced analytics
- ML and AI
- Deep learning

In data analytics, the primary focus is to gain meaningful insights from the underlying data. The scope of Data Science far exceeds this purpose—data science will deal with everything, from analyzing complex data, creating new analytics algorithms and tools for data processing and purification, and even building powerful, useful visualizations.

Data science tools & technologies

This includes programming languages like R, Python, Julia, which can be used to create new algorithms, ML models, AI processes for big data platforms like Apache Spark and Apache Hadoop.

Data processing and purification tools such as Winpure, Data Ladder, and data visualization tools such as Microsoft Power Platform, Google Data Studio, Tableau to visualization frameworks like matplotlib and ploty can also be considered as data science tools.

As data science covers everything related to data, any tool or technology that is used in Big Data and Data Analytics can somehow be utilized in the Data Science process.

Big data vs. data analytics

Big data is the term for large amounts of data that are growing in volume at a rapid rate. The data can be in different forms, such as structured, unstructured, and a blend of both. To process and store it, you need parallel computing capacity and advanced data management tools. Data scientists with big data expertise need to be well-versed in programming NoSQL databases, along with distributed systems and frameworks. Big data analytics is routinely used in financial services, retail, media, entertainment, and communications.

Data analytics is the process of drawing insights from the analysis of raw data, using techniques that range from descriptive to diagnostic, predictive, and prescriptive. While data scientists may be helpful, skilled data analysts with a background in programming, statistics, and mathematics are well able to handle the challenges. Data analytics techniques are used to identify and manage risks, in energy management, and for gaming, healthcare, travel, and science.

Big data vs. data science

- **Data management:** Big data is concerned with storing and processing large amounts of data, while data science uses data that is already cleaned and processed.
- **Data types**: Big data involves many types of data in structured, unstructured, and semi-structured forms. Data science constructs models and programs or algorithms that use various data types.
- **Speed of change**: Big data rapidly grows in volume and data science uses a variety of tools and techniques to handle the analysis of it.
- **Timing**: Big data supports real-time data processing. Data science applies statistical techniques and machine learning for real-time and near-real-time insights and trend tracking.
- **Distribution**: Big data is in repositories that can be queried by data scientists, typically in a networked computing environment.

Data analytics vs. data science

Both data analytics and data science are closely related fields that deal with gaining insights, tracking trends, and using past data to make decisions about the future.

Data science is the broader concept, of which data analytics is a part. It also includes data engineering, machine learning, statistics, programming predictive models, and the development and programming of algorithms.

Data analytics is focused on discovering answers and insights to specific questions in order to address the needs of an organization. The work involves preparing specific datasets, performing various types of analyses, finding insights, and presenting the information in ways that are useful for decision makers.

Data science vs. big data vs. data analytics

Aspect	Data Science	Big Data	Data Analytics
Definition	The interdisciplinary field that uses scientific methods, processes, algorithms, and systems to extract knowledge and insights from structured and unstructured data.	The collection, storage, and processing of large volumes of diverse data types that exceed traditional processing capabilities.	The process of examining datasets to draw conclusions, identify patterns, and support decision-making.
Objective	To derive actionable insights and build predictive models that inform decisions and drive innovation.	To store, manage, and process massive datasets efficiently and reliably, often in real-time or near-real-time.	To provide meaningful insights and support operational or strategic decisions through data examination.

Aspect	Data Science	Big Data	Data Analytics
Focus	Data exploration, model building, algorithm development, and interpretation of complex data.	Handling the 3Vs (volume, velocity, variety) of data and ensuring scalable and efficient data pipelines.	Analysis and reporting of data to identify trends, measure performance, and inform decisions.
Primary Tasks	Data collection, cleaning, exploration, modeling (machine learning), validation, visualization, and communication of insights.	Ingesting, storing, and processing data from diverse sources; ensuring data quality, scalability, and integration.	Descriptive statistics, trend analysis, visualization, and basic predictive insights (often via dashboards and reports).
Tools/Technologies	Python, R, Jupyter Notebook, TensorFlow, PyTorch, Scikit-Learn, SQL, Spark (for some ML tasks).	Hadoop, Spark, HDFS, NoSQL databases (e.g., MongoDB, Cassandra), real-time streaming platforms (e.g., Kafka).	SQL, Excel, Tableau, Power BI, Google Data Studio, some scripting in Python/R for lightweight analytics.
Data Types	Structured, semi-structured, and unstructured data that has been pre-processed and cleaned for analysis.	Structured, semi-structured, and unstructured data - often raw, massive, and in varied formats (text, images, logs, etc.).	Primarily structured and semi-structured data (often sourced from processed Big Data or data warehouses).

Data is the future

Ultimately, big data, data analytics, and data science all help individuals and organizations tackle enormous data sets and extract valuable information out of them. As the importance of data grows exponentially, they will become essential components in the technological landscape.

Related reading

- BMC Machine Learning & Big Data Blog
- DataOps Explained: Understand how DataOps leverages analytics to drive actionable business insights
- What Is a Data Pipeline?
- Database Administrator (DBA) Roles & Responsibilities in The Big Data Age
- Data Streaming Explained: Pros, Cons & How It Works
- Data Ethics for Companies