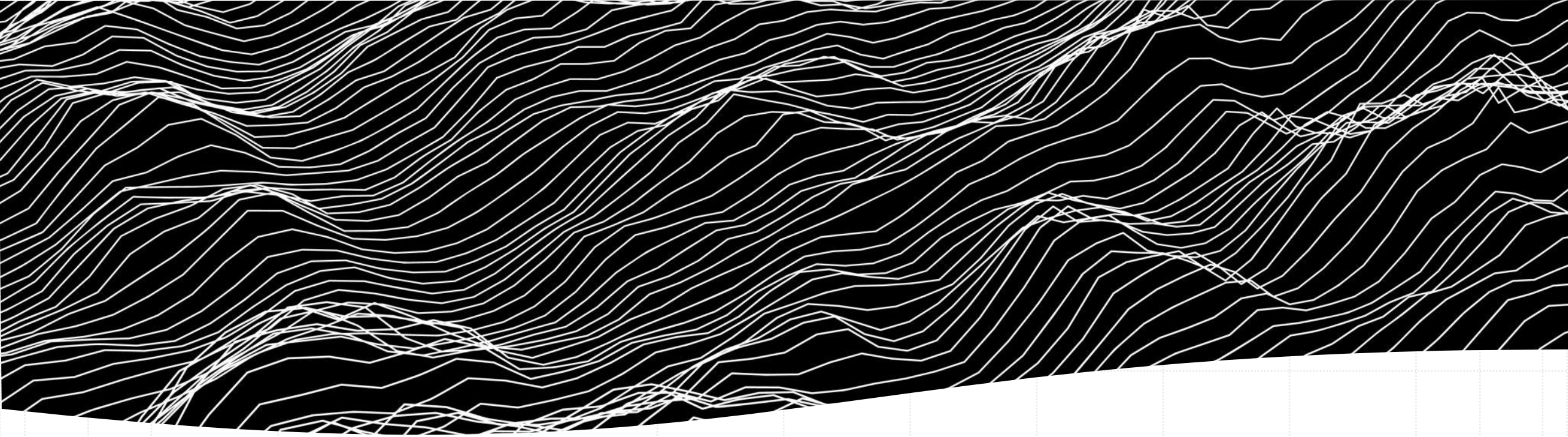


Big Data vs Data Science

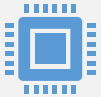
Unit 1

What is Data Science?

Data Science is the study of data. It is about finding patterns in data through an in-depth analysis.

The process of Data Science involves the extraction, data transformation, data analysis and prediction to gain insights about the data.

With Data Science, employees can assist in the **decision-making process** which will help the business to grow and enhance the quality of the product.



Data Science is the most sought after field today. Data is everywhere. It is being generated at an exponential rate and contains within insights that can shape the course of businesses.

There are several machine learning and **business intelligence tools** that help to find the likelihood of the outcome of the event. Data Science is like a sea of data operations. It stems from multiple disciplines like statistics, math and computer science.

What is Big Data?

Big Data is the extraction, analysis and management of processing a large volume of data. It revolves around the datatype – Big Data which is a collection of a colossal amount of data.

Such amount of data, which could not be processed earlier due to limitations in the computational techniques can now be performed with highly advanced tools and methodologies.

Some of the **tools for Big Data** are – Apache Hadoop, Spark, Flink etc. Big Data contains a pool of data that can be both structured and unstructured. By structured/unstructured data, we mean the data that mobile devices, services, & websites generate.

The unstructured data is the data that users generate themselves. **For example**, emails, chats, telephone conversations, reviews etc.

Difference Between Big Data & Data Science

Big Data deals with ***handling & managing huge amount of data***. Prior to Big Data, industries did not possess the required tools and resources to manage such a large volume of data. However, the emergence of MapReduce and Hadoop made it easier for them to handle this form of data. Data Science, on the other hand, is the ***scientific analysis of data***. It is more quantitative in nature and uses various statistical approaches to find insights within the data.

While Big Data is about **storing data**, Data Science is about **analyzing it**. However, it is to be kept in mind that Data Science is an ocean of data operations, one that also includes Big Data. A Data Scientist analyzes the data that is quite large and requires a big data platform. Therefore, an ideal data scientist must also possess the knowledge of big data tools.

Furthermore, Big Data is limited only to the **storage & management of data**. However, recently, more components like PIG & HIVE have been added to the Hadoop framework in order to facilitate the analysis of big data. Furthermore, newer frameworks like **Spark** have analytical features that are intrinsic to it.

The ROLES of Data Scientist & Big Data specialist also differ. A Data Scientist is required to **analyze, draw insights from the data**, visualize the data and communicate the results through robust storytelling. A Big Data Specialist, on the other hand, **develops, maintains & administers Big Data clusters** that hold the voluminous amount of data.

Similarities Between Big Data & Data Science

As mentioned above, Data Science is the ocean of data operations. These data operations also include Big Data. Data Science is like a bigger set that also contains Big Data as its sub-set along with other important data operations. Both of these fields deal with data.

Furthermore, a data scientist is required to handle big data which is frequently unstructured in nature.

In order to handle such type of data, a data scientist must possess the skills. If you are skilled at Hadoop or any other Big Data technology, it will add a great bonus to your profile. Furthermore, it will also increase your value in the market and give you a competitive edge over others.

Recently, the line between Big Data and Data Science has been becoming lesser. This is because recent Big Data platforms like Spark and Flink have data analytical engine as part of their framework.

Even the older platform like Hadoop has released Mahout, which is the data analytical engine comprising of machine learning algorithms. This makes the Big Data platform comprehensive and inclusive of all the data science tools.

Summary

- In the end, Big Data vs Data Science, we conclude that while Big Data and Data Science may share a common frontier of dealing with data, they are completely different.
- We also overviewed how Data Science is a bigger set that comprises of Big Data as its subpart.
- Furthermore, we learnt how newer Big Data platforms are utilizing analytical tools.