

Let's deeply look at the main characteristics of Big Data. They are generally known as Four V's of Big Data.

Volume, Velocity, Veracity and Variety.

First, Volume in Big Data refers to a very large amount of data.

We can see some examples in today's world such as social media platforms, GPS tracking apps,

Online shopping platforms where we generate petabytes and zettabytes of data

And this is nowhere compared to the traditional systems that we are used to in terms of size of the data.

Secondly, if you also look at the speed at which these data arrive, it is remarkably very high.

So the next important characteristic Velocity refers to the speed at which the data arrives.

Let's take the example of the data coming from the popular social media platform such as Twitter.

Think about how many people around the world are tweeting at the same time.

Next, when you collect a very large amount of data at higher speeds, the chances are that they may not be perfect at all times,

they may not be usable immediately, either they might be missing some information due to some connectivity issues or they could be noisy.

This is called Veracity and it refers to the uncertainty that comes with the data.

You need a mechanism to deal with these kind of issues.

We already looked at different types of data and it is called Variety of Big Data such as videos, audios, sensor data, and so on.

These are pretty much unstructured data versus the structured data we see in the typical traditional systems.

Next, let's take a look at some of the key platforms and technologies that were primarily developed as open-source to deal with the characteristics that we just discussed.

Hadoop, HDFS known as Hadoop Distributed File System & Spark are key examples of a MapReduce framework that relies heavily on distributed and functional programming models to ingest, process and analyze the large amount of data.

Mongo DB, Amazon DynamoDB, and Microsoft Cosmos Db are great examples of Document stores which are primarily used for unstructured data such as text documents, JSON files, images, audio and video files etc.

Redis is another great example of a key-value store which is used for in-memory data caching to speed up applications by minimizing reads and writes to slower disk-based systems.

Cassandra and HBase are good examples of a wide column store database that works well for storing enormous amounts of data that can be collected and analyzed that are not optimum for relational models.

Also Neo4j is a graph store where data is stored in vertices and edges such as in many social media platforms.

You will use these concepts throughout the remainder of this course.
Thanks for watching!