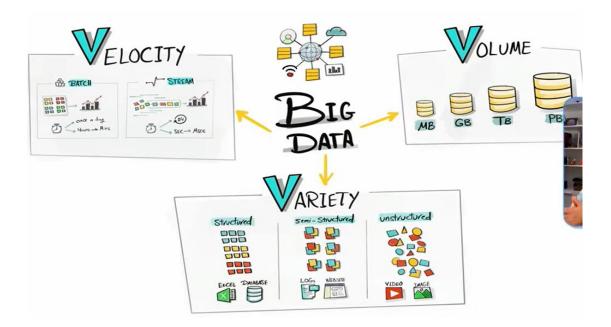Data is the new oil. We are living now in the data-driven age, and data is generated everywhere. We generate a massive amount of data. Each search, each link on the internet. Our smartphones are just constantly uploading data. Not just smartphones, but now we can connect any device to the internet. This is the concept of **IOT,** in which devices are connected to the internet in order to generate data. There are currently over 44zetabytes of data in the entire digital universe. So, we are now not dealing with traditional data; we are dealing with **Big Data**.

BIG DATA:

It is defined by the 3 V`s

1- Volume: The amount of data with mobile devices, IoT, and the Internet has grown dramatically.

2- Velocity: In normal data processing, we are used to processing slow data, or we call it patch data, once a day or something, and we store it on disk. But in BIG DATA, sources are generating streams of data at very high speed. We need to analyze it in real-time fashion, and then store it in memory instead of Disk.

3- Variety: In the big data world, Data often comes in semi-structured Format(Excel, CSV, JSON files, or server logs). Or comes in an Unstructured format like pics and videos.

So, summarizing it, the Big Data term means how we can efficiently store, Process, and analyze data, when it has huge volume, high speed, and different types to reveal significant values for the businesses. But we have a problem that all this generated data is raw Data. Raw data is just unprocessed rows and rows of numbers that are really hard to understand and have almost no value to the businesses. Almost 70% of the world`s data is unused.

Data Architecture:

It is the process of creating Blueprints on how we organize, process, and store our data into different layers for different purposes. So that architecture make it easier to manage, access,

and protect our data.

Data Engineering:
It is a complex process of designing and building data pipelines and data storage. In this, we usually build **ETL** processes to extract row data from multiple sources, then transform it, and then load it into the Target storage.

Data Modeling:
It is the process of connecting the dots-> We're going to put all the data into entities and objects, then we describe the relationship between those entities in order to help us and help the program to understand how the data is related to each other.

Data Mining:
It is the process of analyzing a massive amount of raw data to discover knowledge to discover business intelligence, like patterns and trends.

Machine Learning:
In this, we`re providing the computers with two things: First, the raw and historical data together with mathematical models and algorithms. So once the computer has those 2 things, it's going to start training and practicing in order to perform tasks like predictions. The more machines train, the better and more accurate the results are going to be.

Data Science:
It is the scientific study of data, and it combines three major powers. The power of programming languages, together with mathematics and statistics, and the knowledge of a specific domain In order to uncover valuable knowledge.

Data Visualization: It is the process of converting numbers and raw data into visuals and charts.