What is Big Data?

A massive amount of data which can not be stored,processed and analyzed using traditional tools is known as big data.

5Vs. Of Big Data

1. Value : the most important "V" from the perspective of the business, the value of big data usually comes from insight discovery and pattern recognition that lead to more effective operations, stronger customer relationships and other clear and quantifiable business benefits.

2. Volume : the size and amounts of big data that companies manage and analyze

3. Valocity : the speed at which companies receive, store and manage data – e.g., the specific number of social media posts or search queries received within a day, hour or other unit of time.

4. Variety : the diversity and range of different data types, including unstructured data, semi-structured data and raw data.

5. Veracity : the "truth" or accuracy of data and information assets, which often determines executive-level confidence.

What is Big Data Analytics?

Big Data analytics is a process used to extract meaningful insights, such as hidden patterns, unknown correlations, market trends, and customer preferences. Big Data analytics provides various advantages—it can be used for better decision making, preventing fraudulent activities, among other things.

Uses and Examples of Big Data Analytics

There are many different ways that Big Data analytics can be used in order to improve businesses and organizations. Here are some examples:

- Using analytics to understand customer behavior in order to optimize the customer experience
- Predicting future trends in order to make better business decisions
- Improving marketing campaigns by understanding what works and what doesn't
- Increasing operational efficiency by understanding where bottlenecks are and how to fix them
- Detecting fraud and other forms of misuse sooner

The Lifecycle Phases of Big Data Analytics

Now, let's review how Big Data analytics works:

- Stage 1 - **Business case evaluation** - The Big Data analytics lifecycle begins with a business case, which defines the reason and goal behind the analysis.
- Stage 2 - **Identification of data** - Here, a broad variety of data sources are identified.

- Stage 3 - **Data filtering** - All of the identified data from the previous stage is filtered here to remove corrupt data.
- Stage 4 - **Data extraction** - Data that is not compatible with the tool is extracted and then transformed into a compatible form.
- Stage 5 – **Data Validation and Cleansing –** Data that is invalid leads to inavlid result. In order to ensure only the appropiate data is analysed , the Data Validation and Cleansing stage of the BigData Lifecycle is required.
- Stage 6 - **Data aggregation** - In this stage, data with the same fields across different datasets are integrated.
- Stage 7 - **Data analysis** - Data is evaluated using analytical and statistical tools to discover useful information.
- Stage 8 - **Visualization of data** - With tools like Tableau, Power BI, and QlikView, Big Data analysts can produce graphic visualizations of the analysis.
- Stage 9 - **Final analysis result** - This is the last step of the Big Data analytics lifecycle, where the final results of the analysis are made available to business stakeholders who will take action.
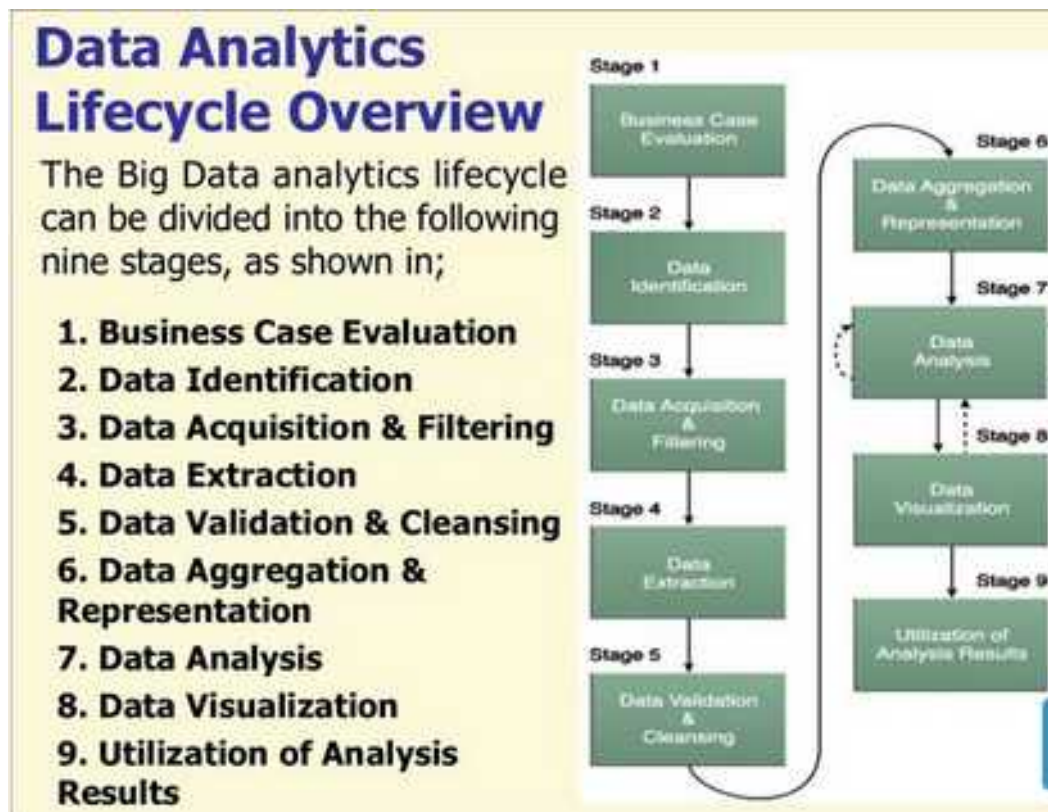
Big Data Analytics Tools

Here are some of the key big data analytics tools :

- **Hadoop** - helps in storing and analyzing data
- **MongoDB** - used on datasets that change frequently
- **Talend** - used for data integration and management
- **Cassandra** - a distributed database used to handle chunks of data
- **Spark** - used for real-time processing and analyzing large amounts of data

- **STORM** - an open-source real-time computational system
- **Kafka** - a distributed streaming platform that is used for fault-tolerant storage

Big Data Industry Applications

Here are some of the sectors where Big Data is actively used:

- **Ecommerce** - Predicting customer trends and optimizing prices are a few of the ways e-commerce uses Big Data analytics
- **Marketing** - Big Data analytics helps to drive high ROI marketing campaigns, which result in improved sales
- **Education** - Used to develop new and improve existing courses based on market requirements
- **Healthcare** - With the help of a patient's medical history, Big Data analytics is used to predict how likely they are to have health issues
- **Media and entertainment** - Used to understand the demand of shows, movies, songs, and more to deliver a personalized recommendation list to its users
- **Banking** - Customer income and spending patterns help to predict the likelihood of choosing various banking offers, like loans and credit cards
- **Telecommunications** - Used to forecast network capacity and improve customer experience
- **Government** - Big Data analytics helps governments in law enforcement, among other things

Difference between Data Mining and Data Analytics

Data Mining	Data Analytics
Data mining is a process of extracting useful information, patterns, and trends from raw data.	Data analytics is a process of evaluating data using analytical and logical concepts to examine a complete insight of all the employees, customers and business.
Data Mining is a step in the data analytics process.	Data Analytics is the umbrella that deals with every step involved in business purposes.
Data mining is performed on well-structured data.	Data Analytics can be performed on any data.
Data mining does not depend on any notions which are introduced before tacking the data.	It is mainly used for hypothesis testing.
It does not need any data visualizations such as bars, charts, etc.	It needs data visualization.