BIG DATA

Big Data:

Big data refers to extremely large and complex datasets that traditional data processing tools can't handle efficiently.

It is not just about size; it's about extracting meaningful insights from diverse. Fast-moving and messy data.

4Vs of Big Data:

The major four dimensions and challenges of Big Data are:

1. Volume
2. Velocity
3. Variety
4. Veracity

1.Volume:

- Refers to scale of data, generated by various sources.
- Big data environments typically involve massive amounts of data, often measured in terabytes, petabytes, or even exabytes.
- Example: A social media platform like Facebook generates enormous amounts of data
- daily, including posts, messages, photos, and videos from billions of users

2.Velocity:

- Refers to speed at which data is generated and processed.
- It encompasses the rate at which data is created and the speed at which it can be analysed.
- Example: Real-time data streams from financial transactions, sensor data from connected devices, or social media activity all generate data at high velocity.

3.Variety:

- Refers to different types of data that are generated from various sources.
- It encompasses structured, semi-structured, and unstructured data.
- Example: Big data can include structured data like databases, semi-structured data like XML files, and unstructured data like text documents, images, audio, and video.

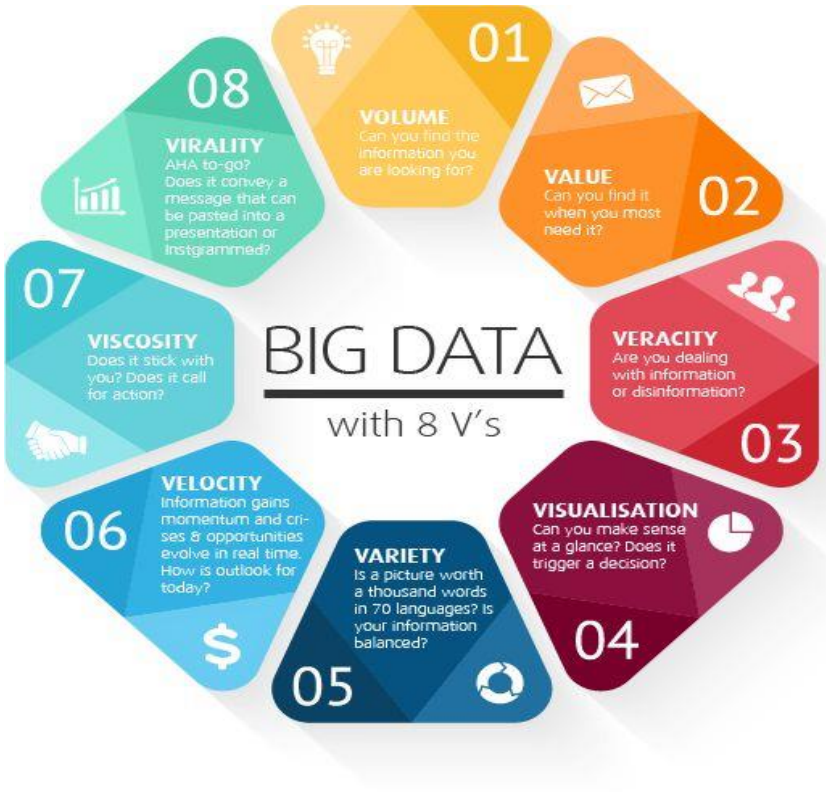
4.Veracity:

- Refers to the quality and trustworthiness of data.
- It encompasses the accuracy, consistency, and reliability of the data.

- Example: Inaccurate or incomplete data can lead to flawed analysis and poor decision-making.

V-Dimension	Description	Key Tools	Use Case Example
Volume	Size of data	Hadoop, SQL	Tracking billions of web clicks
Velocity	Speed of data	Kafka, Spark	Real-time health monitoring
Variety	Data formats	MongoDB, Pandas	Combining CSV, JSON, and images
Veracity	Data quality	Pandas, Data Validation	Filtering fake news in sentiment analysis

The 8 Vs of Big Data:



Beyond the classic 4 Vs:

There are also extra 4 Vs are there which extends the ideology of Big data and its functionalities

5.Value:

- Refers to the usefulness or business impact derived from data.
- Data is only "big" if it delivers insights, decisions, or innovation.
- Examples: E-commerce platforms use customer behavior data to improve product recommendations and boost sales.

6. Viscosity

- Describes the latency or friction in processing and transforming data.
- High viscosity means slow data movement or delayed insights.
- Examples: A legacy system that takes hours to process logs has high viscosity.
- In a MongoDB pipeline, if regex filters slow down query performance, that's viscosity in action.

7. Virality

- Measures the rate and reach of data propagation across networks.
- Viral data can influence trends, decisions, and public perception.
- Examples: A meme or tweet going viral can shape brand reputation in minutes.
  - In analytics: Tracking how fast a product review spreads across platforms helps gauge customer influence.

8.Visualization

- Refers to the presentation of data in graphical or interactive formats.
- Good visualization turns raw data into actionable insights.
- Examples: Dashboards showing sales trends, heatmaps of user clicks, or bar charts of inventory levels.
- In Pandas: Using .plot() or Seaborn to visualize cleaned data helps spot outliers or patterns.