

BIG DATA:

* Big Data refers to extremely large datasets that cannot be managed, processed, or analyzed using traditional data processing tools.
* It includes structured, semi-structured, and unstructured data.
* It is used in various fields to analyze trends, predict outcomes, and optimize operations.
* Big data analysis provides valuable insights for improved decision-making, process automation, and innovation.

Types of Big Data:

1. **Structured Data**
   * Data that resides in a fixed field within records or files (e.g., SQL databases).
2. **Semi-structured Data**
   * Data that does not reside in a relational database but has some structure (e.g., XML, JSON).
3. **Unstructured Data**
   * Data with no predefined format (e.g., images, videos, emails, audio).

Characteristics of Big Data – The 4 V’s:

1. **Volume**
   * Refers to the huge amount of data generated from different sources such as sensors, devices, social media platforms, transactions, etc.
   * Example: Facebook generates more than 4 petabytes of data every day.
2. **Velocity**
   * The speed at which data is generated, collected, and processed.
   * Real-time or near real-time processing is required in many applications like fraud detection or live traffic updates.
3. **Variety**
   * Big Data includes various formats such as structured data (databases), semi-structured data (XML, JSON), and unstructured data (videos, emails, social media).
4. **Veracity**
   * Refers to the trustworthiness and accuracy of the data. High veracity means more reliable insights.

