**Big Data – Short Notes**

**Definition:**  
Big Data refers to extremely large datasets that are too complex or voluminous to be handled by traditional data processing tools.

**Characteristics (The 5 V’s):**

1. **Volume** – Massive amounts of data.
2. **Velocity** – Speed at which data is generated and processed.
3. **Variety** – Different types (structured, unstructured, semi-structured).
4. **Veracity** – Trustworthiness of data.
5. **Value** – Usefulness of the data in decision-making.

**Technologies Used:**

* **Hadoop** – Open-source framework for distributed storage and processing.
* **Spark** – Fast data processing engine.
* **NoSQL** – Non-relational databases (e.g., MongoDB, Cassandra).
* **MapReduce** – Programming model for processing large datasets.

**Applications:**

* Predictive analytics
* Fraud detection
* Recommendation engines
* Social media analysis