

## **Introduction to Big Data - Detailed 10 Marks Answers**

### **1. Explain five effective characteristics of BIG DATA.**

Big Data refers to large and complex datasets that traditional data processing software cannot handle efficiently. The characteristics of Big Data are commonly known as the 'Five Vs' - Volume, Velocity, Variety, Veracity, and Value.

1. Volume: Refers to the vast amount of data generated every second from various sources such as social media, IoT devices, and sensors.
2. Velocity: Indicates the speed at which data is created, collected, and processed to provide timely insights.
3. Variety: Represents the different types and formats of data such as text, images, videos, and logs.
4. Veracity: Refers to the reliability and accuracy of data. Not all data collected is trustworthy.
5. Value: The ultimate goal of Big Data is to extract meaningful insights that add business or social value.

Conclusion: These five characteristics define the scope and complexity of Big Data, making it a vital concept in modern analytics.

### **2. Write down applications of BIG DATA.**

Big Data has numerous applications across various sectors, helping in data-driven decision making and predictive analysis.

Applications:

1. Healthcare: Used to predict disease outbreaks, manage patient records, and enhance diagnosis accuracy.
2. Banking and Finance: Detects fraud, manages risks, and improves customer targeting.
3. Retail: Helps understand consumer behavior, optimize pricing, and improve customer experience.
4. Transportation: Enhances traffic management and logistics planning using real-time analytics.

## **Introduction to Big Data - Detailed 10 Marks Answers**

5. Government: Used for policy making, crime prevention, and smart city management.
6. Education: Supports personalized learning and institutional decision-making.

Conclusion: Big Data applications have revolutionized industries by turning raw data into valuable insights for better performance.

### **3. How BIG DATA Vs. are classified? Explain in detail.**

Big Data is classified using the 'V' dimensions that represent its properties. Initially, there were three Vs - Volume, Velocity, and Variety - later expanded to five with Veracity and Value.

1. Volume: The quantity of data generated from sources like social media, IoT, and sensors. For example, Facebook generates terabytes of data daily.
2. Velocity: The speed at which data flows in from devices and applications. Real-time data analysis is essential for instant decisions.
3. Variety: The diversity in data formats such as structured (databases), semi-structured (XML, JSON), and unstructured (videos, text).
4. Veracity: Deals with the quality and accuracy of data. Data may be incomplete or inconsistent.
5. Value: Represents the usefulness of data in generating insights or profits.

Conclusion: The classification of Big Data based on these Vs helps organizations understand, store, and analyze large datasets efficiently.

### **4. Write down challenges of BIG DATA.**

Although Big Data offers great benefits, it also presents several challenges in handling, processing, and securing the data.

Major Challenges:

1. Data Storage: Managing and storing petabytes of data requires efficient infrastructure.

## **Introduction to Big Data - Detailed 10 Marks Answers**

2. Data Integration: Combining data from multiple sources in different formats is complex.
3. Data Quality: Ensuring accuracy, completeness, and reliability of data.
4. Data Security: Protecting sensitive data from unauthorized access and cyber-attacks.
5. Processing Speed: Analyzing large datasets in real time demands powerful computing systems.
6. Cost Management: Maintaining servers, tools, and skilled professionals is expensive.

Conclusion: Overcoming these challenges requires advanced technologies, skilled personnel, and robust data governance policies.

### **5. Explain the difference between Volume, Veracity and Velocity.**

Volume, Veracity, and Velocity are three essential characteristics of Big Data that describe its scale, accuracy, and speed.

1. Volume: Refers to the total amount of data generated and stored. The larger the data, the greater the need for advanced storage solutions.

Example: Social media posts, financial transactions.

2. Veracity: Indicates the trustworthiness and quality of the data. It ensures that decisions are based on accurate information.

Example: Filtering fake or duplicate data from datasets.

3. Velocity: Defines the speed at which new data is created and processed. Real-time analytics are essential for industries like banking or healthcare.

Example: Streaming analytics for fraud detection.

Conclusion: Together, these three Vs determine how efficiently Big Data can be processed and utilized for meaningful insights.

## **Introduction to Big Data - Detailed 10 Marks Answers**

### **6. Write down the tools used in BIG DATA.**

Big Data tools are technologies that help store, process, and analyze large datasets effectively.

Common Big Data Tools:

1. Hadoop: An open-source framework for distributed storage and processing of large datasets.
2. Spark: A powerful tool for real-time data analytics with faster processing than Hadoop.
3. Hive: A data warehouse system that allows SQL-like queries on Hadoop.
4. HBase: A NoSQL database for managing large, sparse datasets.
5. Flume: Used for collecting and transferring log data.
6. Pig: A high-level platform for creating data analysis programs.
7. Tableau: Used for data visualization and business intelligence.

Conclusion: These tools enable organizations to handle massive volumes of data efficiently and derive actionable insights.