

#1. What is the primary characteristic of Big Data?

- ☐  
Volume
- ☐  
Velocity
- ☐  
Variety
- ☐  
Veracity
- ☐  
Value

#2. Which technology is used for real-time Big Data processing?

- ☐  
Hadoop
- ☐  
Spark
- ☐  
Hive
- ☐  
Pig
- ☐  
Flink

#3. What type of data is easily stored, queried, and analyzed?

- ☐  
Structured Data
- ☐  
Semi-Structured Data
- ☐

Unstructured Data

☐

Raw Data

☐

Meta Data

#4. Which programming language is commonly used for data analysis in Big Data?

☐

Python

☐

Java

☐

C++

☐

R

☐

Scala

#5. Which component manages data storage in Hadoop?

☐

HDFS

☐

MapReduce

☐

YARN

☐

Hive

☐

Pig

#6. What is the term for analyzing large datasets to find patterns and correlations?

☐

Descriptive Analytics

☐

Diagnostic Analytics

☐

Predictive Analytics

☐

Prescriptive Analytics

☐

Data Mining

#7. Which technology allows distributed, fault-tolerant data storage and retrieval?

☐

NoSQL Databases

☐

SQL Databases

☐

Relational Databases

☐

In-Memory Databases

☐

Graph Databases

#8. What's the process of preparing raw data for analysis called?

☐

Data Wrangling

☐

Data Mining

☐

Data Aggregation

☐

Data Integration

☐

## Data Normalization

#9. Name a popular cloud-based platform for Big Data processing.

☐

AWS (Amazon Web Services)

☐

Azure (Microsoft)

☐

GCP (Google Cloud Platform)

☐

IBM Cloud

☐

Oracle Cloud

#10. What technique distributes data across multiple nodes for parallel processing?

☐

Sharding

☐

Replication

☐

Partitioning

☐

Indexing

☐

Joins

#11. Which term is used to describe data that is not in a structured format?

☐

Unstructured Data

☐

Semi-Structured Data

☐

Raw Data

☐

Meta Data

☐

Structured Data

#12. What technology is used for batch processing of Big Data?

☐

Hadoop

☐

Spark

☐

Hive

☐

Pig

☐

Flink

#13. Which language is commonly used for querying data in a relational database?

☐

SQL

☐

Python

☐

Java

☐

R

☐

C++

#14. In Hadoop, what is the processing engine responsible for distributed data

processing?

☐

MapReduce

☐

HDFS

☐

YARN

☐

Hive

☐

Pig

#15. Which type of analytics focuses on identifying the cause of a problem?

☐

Diagnostic Analytics

☐

Descriptive Analytics

☐

Predictive Analytics

☐

Prescriptive Analytics

☐

Data Mining

#16. Which database type is designed for high-performance, scalable, and fault-tolerant storage and retrieval?

☐

NoSQL Databases

☐

SQL Databases

☐

Relational Databases

☐

In-Memory Databases

☐

Graph Databases

#17. What is the term for the process of combining and restructuring data for easier analysis?

☐

Data Wrangling

☐

Data Mining

☐

Data Aggregation

☐

Data Integration

☐

Data Normalization

#18. Which cloud platform provides services like Amazon S3 and Amazon Redshift for Big Data processing?

☐

AWS (Amazon Web Services)

☐

Azure (Microsoft)

☐

GCP (Google Cloud Platform)

☐

IBM Cloud

☐

Oracle Cloud

#19. What technique is used to divide a dataset into smaller, manageable pieces

for parallel processing?

☐

Partitioning

☐

Sharding

☐

Replication

☐

Indexing

☐

Joins

#20. Which technology is known for its ability to handle real-time stream processing of data?

☐

Flink

☐

Spark

☐

Hadoop

☐

Hive

☐

Pig

Next

Results





