

#1. What is the term for data that falls in between structured and unstructured data?

☐

Semi-Structured Data

☐

Unstructured Data

☐

Raw Data

☐

Meta Data

☐

Structured Data

#2. Which technology is known for its in-memory processing capabilities for Big Data?

☐

Spark

☐

Hadoop

☐

Hive

☐

Pig

☐

Flink

#3. Which language is commonly used for developing applications on the Hadoop platform?

☐

Java

☐

Python

☐

C++

☐

Scala

☐

R

#4. In a Hadoop ecosystem, which component is responsible for resource management and job scheduling?

☐

YARN (Yet Another Resource Negotiator)

☐

HDFS (Hadoop Distributed File System)

☐

MapReduce

☐

Hive

☐

Pig

#5. What type of analytics focuses on predicting future outcomes based on historical data?

☐

Predictive Analytics

☐

Descriptive Analytics

☐

Diagnostic Analytics

☐

Prescriptive Analytics

☐

Data Mining

#6. Which database type is optimized for handling complex relationships and querying graph-like data structures?

☐

Graph Databases

☐

NoSQL Databases

☐

SQL Databases

☐

Relational Databases

☐

In-Memory Databases

#7. What is the term for combining and summarizing data to create useful information?

☐

Data Aggregation

☐

Data Wrangling

☐

Data Mining

☐

Data Integration

☐

Data Normalization

#8. Which cloud platform provides services like Azure Data Lake Storage and Azure HDInsight for Big Data processing?

☐

Azure (Microsoft)

☐

AWS (Amazon Web Services)

☐

GCP (Google Cloud Platform)

☐

IBM Cloud

☐

Oracle Cloud

#9. What technique is used to replicate data across multiple nodes for fault tolerance?

☐

Replication

☐

Sharding

☐

Partitioning

☐

Indexing

☐

Joins

#10. Which technology is commonly used for processing and analyzing large-scale data streams in real-time?

☐

Flink

☐

Spark

☐

Hadoop

☐

Hive

☐

Pig

#11. What is the term for the process of extracting meaningful insights from data?

☐

Data Mining

☐

Data Wrangling

☐

Data Aggregation

☐

Data Integration

☐

Data Normalization

#12. Which cloud platform provides services like Google BigQuery and Dataflow for Big Data processing?

☐

GCP (Google Cloud Platform)

☐

AWS (Amazon Web Services)

☐

Azure (Microsoft)

☐

IBM Cloud

☐

Oracle Cloud

#13. What technique is used to distribute data across multiple nodes based on a certain key or attribute?

☐

Sharding

☐

Partitioning

☐

Replication

☐

Indexing

☐

Joins

#14. Which technology is known for its ability to handle both batch and stream processing of data?

☐

Flink

☐

Spark

☐

Hadoop

☐

Hive

☐

Pig

#15. In a Hadoop ecosystem, which component is responsible for querying and analyzing data using SQL-like language?

☐

Hive

☐

YARN (Yet Another Resource Negotiator)

☐

MapReduce

☐

HDFS (Hadoop Distributed File System)

☐

Pig

#16. What type of analytics focuses on providing recommendations and actions based on data analysis?

☐

Prescriptive Analytics

☐

Descriptive Analytics

☐

Diagnostic Analytics

☐

Predictive Analytics

☐

Data Mining

#17. Which database type is designed for storing, querying, and managing interconnected data elements?

☐

Graph Databases

☐

NoSQL Databases

☐

SQL Databases

☐

Relational Databases

☐

In-Memory Databases

#18. What is the term for the process of combining multiple datasets into a unified view for analysis?

☐

Data Integration

☐

Data Wrangling

☐

Data Mining

☐

Data Aggregation

☐

Data Normalization

#19. Which cloud platform provides services like IBM Db2 and Db2 Warehouse for Big Data processing?

☐

IBM Cloud

☐

AWS (Amazon Web Services)

☐

Azure (Microsoft)

☐

GCP (Google Cloud Platform)

☐

Oracle Cloud

#20. Which technique is used to join related data from multiple tables in a database?

☐

Joins

☐

Replication

☐

Sharding

☐

Partitioning





Indexing

Next

Results

