

#1. What is the term for the process of identifying and handling missing or inaccurate data in a dataset?

☐

Data Cleansing

☐

Data Wrangling

☐

Data Aggregation

☐

Data Integration

☐

Data Normalization

#2. Which cloud platform provides services like Snowflake and Databricks for Big Data processing?

☐

AWS (Amazon Web Services)

☐

Azure (Microsoft)

☐

GCP (Google Cloud Platform)

☐

IBM Cloud

☐

Oracle Cloud

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

☐

Sharding

☐

Partitioning

☐

Replication

☐

Indexing

☐

Joins

#4. Which technology is known for its ability to process large volumes of data in real-time with low latency?

☐

Flink

☐

Spark

☐

Hadoop

☐

Hive

☐

Kafka

#5. In a Hadoop ecosystem, which component is responsible for managing and executing data processing tasks in a cluster?

☐

YARN (Yet Another Resource Negotiator)

☐

MapReduce

☐

HDFS (Hadoop Distributed File System)

☐

Hive

☐

Pig

#6. What type of analytics focuses on understanding why certain events or patterns occurred in the data?

☐

Diagnostic Analytics

☐

Descriptive Analytics

☐

Prescriptive Analytics

☐

Predictive Analytics

☐

Data Mining

#7. Which database type is designed for handling highly connected and interrelated data in a graph-like structure?

☐

Graph Databases

☐

NoSQL Databases

☐

SQL Databases

☐

Relational Databases

☐

In-Memory Databases

#8. What is the term for the process of combining and summarizing data to provide insights into business performance?

☐

Business Intelligence

☐

Data Wrangling

☐

Data Mining

☐

Data Integration

☐

Data Normalization

#9. Which cloud platform provides services like Alibaba Cloud AnalyticDB and MaxCompute for Big Data processing?

☐

Alibaba Cloud

☐

AWS (Amazon Web Services)

☐

Azure (Microsoft)

☐

GCP (Google Cloud Platform)

☐

Oracle Cloud

#10. Which technique is used to create copies of data on multiple nodes to ensure fault tolerance and availability?

☐

Replication

☐

Sharding

☐

Partitioning

☐

Indexing

☐

Joins

#11. What is the term for the process of organizing and storing data in a way that allows for efficient retrieval and analysis?

☐

Data Management

☐

Data Architecture

☐

Data Modeling

☐

Data Storage

☐

Data Warehousing

#12. Which technology is known for its ability to handle complex event processing and real-time analytics on data streams?

☐

Kafka

☐

Flink

☐

Spark

☐

Hadoop

☐

Hive

#13. In a Hadoop ecosystem, which component is responsible for resource allocation and scheduling of tasks on a cluster?

☐

YARN (Yet Another Resource Negotiator)

☐

HDFS (Hadoop Distributed File System)

☐

MapReduce

☐

Hive

☐

Pig

#14. What type of analytics focuses on suggesting actions to optimize future outcomes based on data analysis?

☐

Prescriptive Analytics

☐

Descriptive Analytics

☐

Diagnostic Analytics

☐

Predictive Analytics

☐

Data Mining

#15. Which database type is optimized for handling large volumes of unstructured data with high scalability?

☐

NoSQL Databases

☐

SQL Databases

☐

Relational Databases

☐

In-Memory Databases

☐

Graph Databases

#16. What is the term for the process of combining data from various sources to create a unified view for analysis?

☐

Data Integration

☐

Data Wrangling

☐

Data Mining

☐

Data Aggregation

☐

Data Normalization

#17. Which cloud platform provides services like Cloudera Data Platform and CDP Data Hub for Big Data processing?

☐

Cloudera

☐

AWS (Amazon Web Services)

☐

Azure (Microsoft)

☐

GCP (Google Cloud Platform)

☐

IBM Cloud

#18. What technique is used to divide a dataset into smaller, manageable parts for parallel processing?

☐

Partitioning

☐

Sharding

☐

Replication

☐

Indexing

☐

Joins

#19. Which technology is known for its ability to handle real-time processing of data streams and complex event processing?

☐

Flink

☐

Kafka

☐

Spark

☐

Hadoop

☐

Hive

#20. In a Hadoop ecosystem, which component is responsible for managing and allocating resources in a cluster?

☐

YARN (Yet Another Resource Negotiator)

☐

HDFS (Hadoop Distributed File System)

☐

MapReduce

☐

Hive





Pig

Finish

Results

