

#1. What is the term for the process of extracting insights and knowledge from data using advanced algorithms and machine learning techniques?

☐

Data Science

☐

Data Wrangling

☐

Data Aggregation

☐

Data Integration

☐

Data Normalization

#2. Which cloud platform provides services like Tencent Cloud TDSQL and Hadoop for Big Data processing?

☐

Tencent Cloud

☐

AWS (Amazon Web Services)

☐

Azure (Microsoft)

☐

GCP (Google Cloud Platform)

☐

Oracle Cloud

#3. What technique is used to distribute data across multiple nodes in a way that each node has a copy of the entire dataset?

☐

Replication

☐

Sharding

☐

Partitioning

☐

Indexing

☐

Joins

#4. Which technology is known for its ability to handle large-scale data processing and machine learning tasks?

☐

Spark

☐

Flink

☐

Hadoop

☐

Hive

☐

Pig

#5. In a Hadoop ecosystem, which component is responsible for processing and analyzing data using SQL-like queries?

☐

Hive

☐

MapReduce

☐

YARN (Yet Another Resource Negotiator)

☐

HDFS (Hadoop Distributed File System)

☐

Pig

#6. What type of analytics focuses on suggesting actions to optimize future outcomes?

☐

Prescriptive Analytics

☐

Descriptive Analytics

☐

Diagnostic Analytics

☐

Predictive Analytics

☐

Data Mining

#7. Which database type is designed for efficiently storing and retrieving highly connected data points 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 gain insights into trends and patterns?

☐

Data Analysis

☐

Data Wrangling

☐

Data Mining

☐

Data Integration

☐

Data Normalization

#9. Which cloud platform provides services like SAP HANA and SAP Data Intelligence for Big Data processing?

☐

AWS (Amazon Web Services)

☐

Azure (Microsoft)

☐

GCP (Google Cloud Platform)

☐

IBM Cloud

☐

SAP Cloud

#10. Which technique is used to optimize data retrieval by creating a smaller, indexed representation of the data in memory?

☐

In-Memory Indexing

☐

Joins

☐

Replication

☐

Sharding

☐

Partitioning

#11. What is the term for the process of storing data across multiple nodes to ensure availability and fault tolerance?

☐

Replication

☐

Sharding

☐

Partitioning

☐

Indexing

☐

Joins

#12. Which technology is known for its ability to handle complex graph algorithms and traversals on large-scale graph datasets?

☐

Graph Databases

☐

Spark

☐

Flink

☐

Hadoop

☐

Hive

#13. In a Hadoop ecosystem, which component is responsible for batch processing of large datasets?

☐

MapReduce

- ☐  
YARN (Yet Another Resource Negotiator)
  - ☐  
HDFS (Hadoop Distributed File System)
  - ☐  
Hive
  - ☐  
Pig
- #14. What type of analytics focuses on determining the root cause of a problem or finding solutions to improve outcomes?

- ☐  
Diagnostic Analytics
- ☐  
Descriptive Analytics
- ☐  
Prescriptive Analytics
- ☐  
Predictive Analytics
- ☐  
Data Mining

#15. Which database type is optimized for handling unstructured and semi-structured data?

- ☐  
NoSQL Databases
- ☐  
SQL Databases
- ☐  
Relational Databases
- ☐  
In-Memory Databases

☐

Graph Databases

#16. What is the term for the process of combining data from different 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 Redshift and EMR for Big Data processing?

☐

AWS (Amazon Web Services)

☐

Azure (Microsoft)

☐

GCP (Google Cloud Platform)

☐

IBM Cloud

☐

Oracle 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 stream processing and event-driven architectures?

☐

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

Next

Results

