

1. What is Hive primarily used for?

- a) Real-time data processing
- b) Analyzing structured data
- c) Web scraping
- d) Image recognition

Answer: b) Analyzing structured data

Explanation: Hive is primarily used for querying and analyzing structured data stored in Hadoop distributed file system (HDFS) using a SQL-like language called HiveQL.

2. Which component of Hadoop ecosystem does Hive rely on for distributed storage and processing?

- a) HBase
- b) MapReduce
- c) YARN
- d) ZooKeeper

Answer: b) MapReduce

Explanation: Hive relies on MapReduce for distributed storage and processing of data stored in HDFS.

3. Which of the following statements is true about Hive architecture?

- a) Hive directly stores data in a relational database.

- b) Hive translates SQL-like queries into MapReduce jobs for execution.
- c) Hive only supports unstructured data processing.
- d) Hive is independent of Hadoop ecosystem components.

Answer: b) Hive translates SQL-like queries into MapReduce jobs for execution.

Explanation: Hive architecture involves translating SQL-like queries into MapReduce jobs which are executed on the Hadoop cluster.

4. Which of the following is not a Hive data type?

- a) Array
- b) Map
- c) Tuple
- d) Struct

Answer: c) Tuple

Explanation: Hive supports complex data types such as Array, Map, Struct, but not Tuple.

5. In Hive, what is the purpose of the 'serde' in table creation?

- a) Serialization/Deserialization
- b) Sorting data
- c) Securing data
- d) Shuffling data

Answer: a) Serialization/Deserialization

Explanation: The 'serde' (serialization/deserialization) in Hive table creation specifies how data is serialized and deserialized when read from or written to the table.

6. Which query language is used in Hive for data manipulation and querying?

- a) HiveQL
- b) Pig Latin
- c) SQL
- d) Java

Answer: a) HiveQL

Explanation: Hive Query Language (HiveQL) is used in Hive for data manipulation and querying.

7. What is the primary function of Pig in the Hadoop ecosystem?

- a) Real-time data processing
- b) Analyzing structured data
- c) Batch processing of data
- d) Machine learning

Answer: c) Batch processing of data

Explanation: Pig is primarily used for batch processing of data in the Hadoop ecosystem.

8. Which programming language is used to write Pig scripts?

- a) Java
- b) Python
- c) Pig Latin
- d) Scala

Answer: c) Pig Latin

Explanation: Pig scripts are written in Pig Latin, a high-level data flow scripting language.

9. What does ETL stand for in the context of data processing?

- a) Extract, Transform, Load
- b) Examine, Test, Learn
- c) Execute, Transform, Log
- d) Edit, Transform, Load

Answer: a) Extract, Transform, Load

Explanation: ETL stands for Extract, Transform, Load, which refers to the process of extracting data from various sources, transforming it into a suitable format, and loading it into a target database or data warehouse.

10. Which of the following is not a data type supported by Pig?

- a) Integer
- b) Float
- c) Date
- d) Complex

Answer: c) Date

Explanation: Pig supports basic data types like Integer, Float, and also complex data types like Map, Tuple, and Bag, but it does not have a specific data type for handling dates.

11. How does Pig execute data processing tasks?

- a) Using MapReduce
- b) Using Spark
- c) Using Flink
- d) Using Storm

Answer: a) Using MapReduce

Explanation: Pig executes data processing tasks using MapReduce framework.

12. Which of the following Pig Latin operators is used for filtering data?

- a) GROUP
- b) JOIN
- c) FILTER
- d) FOREACH

Answer: c) FILTER

Explanation: FILTER operator in Pig Latin is used to filter rows of data based on a specified condition.

13. What is the purpose of Pig functions?

- a) To define complex data structures
- b) To perform data manipulation and transformation
- c) To optimize query execution
- d) To manage security access

Answer: b) To perform data manipulation and transformation

Explanation: Pig functions are used to perform various data manipulation and transformation tasks within Pig scripts.

14. Which of the following is a user-defined function (UDF) in Pig?

- a) CONCAT
- b) COUNT
- c) MAX
- d) SUM

Answer: a) CONCAT

Explanation: CONCAT is an example of a user-defined function (UDF) in Pig, used for concatenating strings.

15. What is the default execution mode of Pig?

- a) Local mode
- b) Distributed mode

- c) Standalone mode
- d) Parallel mode

Answer: b) Distributed mode

Explanation: The default execution mode of Pig is distributed mode, where tasks are executed across the Hadoop cluster.

16. Which of the following is a Pig Latin keyword used to define a schema for data?

- a) LOAD
- b) STORE
- c) SCHEMA
- d) DESCRIBE

Answer: c) SCHEMA

Explanation: SCHEMA keyword is used in Pig Latin to define a schema for data during loading or storing operations.

17. What is the purpose of the 'STORE' keyword in Pig Latin?

- a) To load data into a relation
- b) To filter data
- c) To store the output of a Pig script
- d) To group data

Answer: c) To store the output of a Pig script

Explanation: The 'STORE' keyword in Pig Latin is used to store the output of a Pig script into a specified location.

18. Which Pig Latin operator is used for joining two or more datasets?

- a) GROUP
- b) JOIN
- c) FILTER
- d) DISTINCT

Answer: b) JOIN

Explanation: JOIN operator in Pig Latin is used for joining two or more datasets based on a common field.

19. What does the 'GROUP' operator in Pig Latin do?

- a) Aggregates data based on a specified key
- b) Filters data based on a condition
- c) Joins two or more datasets
- d) Sorts data

Answer: a) Aggregates data based on a specified key

Explanation: The 'GROUP' operator in Pig Latin is used to aggregate data based on a specified key field.

20. Which data type in Pig represents an unordered collection of tuples?

- a) Bag
- b) Tuple
- c) Map
- d) Array

Answer: a) Bag

Explanation: In Pig, a Bag represents an unordered collection of tuples.

Related posts:

1. Introduction to Information Security
2. Introduction to Information Security MCQ
3. Introduction to Information Security MCQ
4. Symmetric Key Cryptography MCQ
5. Asymmetric Key Cryptography MCQ
6. Authentication & Integrity MCQ
7. E-mail, IP and Web Security MCQ