

CMT 428 module 4 quiz 1

Questions

1. Which tool is used for data warehousing on top of Hadoop? a) Pig b) Hive c) HBase d) Sqoop
2. What language is used to query data in Hive? a) Pig Latin b) HiveQL c) Java d) Python
3. Which tool is used for high-level data flow and parallel computation? a) Hive b) Pig c) HBase d) Oozie
4. What is the scripting language used in Pig? a) HiveQL b) Pig Latin c) Oozie d) Sqoop
5. Which tool is a NoSQL database for real-time read/write access to data? a) Hive b) Pig c) HBase d) Sqoop
6. What type of storage does HBase use? a) Row-oriented b) Column-oriented c) Document-oriented d) Graph-oriented
7. Which tool is used to transfer data between Hadoop and relational databases? a) Hive b) Pig c) HBase d) Sqoop
8. What are the two main functions of Sqoop? a) Import and Export b) Map and Reduce c) ETL and ELT d) Streaming and Batching
9. Which tool is used for collecting and aggregating log data? a) Hive b) Pig c) Flume d) Oozie
10. What are the main components of a Flume pipeline? a) Sources, Channels, Sinks b) Mappers, Reducers, Combiners c) Nodes, Edges, Graphs d) Tables, Partitions, Buckets
11. Which tool is used as a workflow scheduler for Hadoop jobs? a) Hive b) Pig c) Flume d) Oozie
12. What is the primary function of Oozie? a) Data warehousing b) Data analysis c) Data ingestion d) Workflow management
13. Which of the following is NOT a Hadoop ecosystem tool? a) Hive b) Pig c) MySQL d) Oozie
14. What is the purpose of Hive partitions? a) To improve query performance b) To store data in different formats c) To replicate data for fault tolerance d) To manage user permissions
15. What is the role of a SerDe in Hive? a) To serialize and deserialize data b) To schedule and manage workflows c) To transfer data between Hadoop and relational databases d) To collect and aggregate log data
16. What is the purpose of UDFs (User Defined Functions) in Pig? a) To extend Pig's functionality with custom functions b) To define data schemas for input and output data c) To connect to external data sources d) To manage and monitor Hadoop jobs
17. What is the purpose of row keys in HBase? a) To uniquely identify rows in a table b) To define the schema of a table c) To store data values d) To manage access control to data

18. What is the purpose of Flume interceptors? a) To modify or drop events in a Flume pipeline b) To define the source of data for Flume c) To specify the destination for Flume data d) To manage and monitor Flume agents
19. What is an Oozie coordinator used for? a) To schedule periodic workflows b) To define individual actions within a workflow c) To manage dependencies between workflow actions d) To monitor the execution of workflows
20. What is an Oozie bundle? a) A collection of workflows that can be managed and executed together b) A single action within a workflow c) A configuration file for an Oozie workflow d) A log file generated by an Oozie workflow

Answers

1. b
2. b
3. b
4. b
5. c
6. b
7. d
8. a
9. c
10. a
11. d
12. d
13. c
14. a
15. a
16. a
17. a
18. a
19. a
20. a