

# **Alibaba**

**ACA-BIGDATA1 Exam**

**ACA Big Data Certification**

**Questions & Answers  
Demo**

## Version: 4.0

---

### Question: 1

---

Scenario: Jack is the administrator of project prj1. The project involves a large volume of sensitive data such as bank account, medical record, etc. Jack wants to properly protect the data. Which of the follow statements is necessary?

- A. set ProjectACL=true;
- B. add accountprovider ram;
- C. set ProjectProtection=true;
- D. use prj1;

---

**Answer: C**

---

---

### Question: 2

---

Where is the meta data (e.g.,table schemas) in Hive?

- A. Stored as metadata on the NameNode
- B. Stored along with the data in HDFS
- C. Stored in the RDBMS like MySQL
- D. Stored in ZooKeeper

---

**Answer: C**

---

---

### Question: 3

---

MaxCompute tasks contain computational tasks and non-computational tasks. The computational tasks require actual operations on data stored in the table. MaxCompute parses the task to obtain its execution plan, and submits the task for execution. The noncomputational tasks require substantial reading of and modification to metadata information. Therefore, the task is not parsed, and no execution plan is provided. The task is directly submitted for execution. The latter one has a faster response speed than the former one. Which of the following operations on the table t\_test is a computational task?

- A. desc t\_test
- B. alter table t\_test add columns (comments string);
- C. select count(\*) from t\_test;
- D. truncate table t\_test;

---

**Answer: C**

---

---

**Question: 4**

---

When we use the MaxCompute tunnel command to upload the log.txt file to the t\_log table, the t\_log is a partition table and the partitioning column is (p1 string, p2 string). Which of the following commands is correct?

- A. tunnel upload log.txt t\_log/p1="b1", p2="b2"
- B. tunnel upload log.txt t\_log/(p1="b1", p2="b2")
- C. tunnel upload log.txt t\_log/p1="b1"/p2="b2"

---

**Answer: A**

---

---

**Question: 5**

---

A Log table named log in MaxCompute is a partition table, and the partition key is dt. A new partition is created daily to store the new data of that day. Now we have one month's data, starting from dt='20180101' to dt='20180131', and we may use \_\_\_\_\_ to delete the data on 20180101.

- A. delete from log where dt='20180101'
- B. truncate table where dt='20180101'
- C. drop partition log (dt='20180101')
- D. alter table log drop partition(dt='20180101')

---

**Answer: D**

---

---

**Question: 6**

---

DataV is a powerful yet accessible data visualization tool, which features geographic information systems allowing for rapid interpretation of data to understand relationships, patterns, and trends. When a DataV screen is ready, it can embed works to the existing portal of the enterprise through \_\_\_\_\_.

- A. URL after the release
- B. URL in the preview
- C. MD5 code obtained after the release
- D. Jar package imported after the release

---

**Answer: A**

---

---

**Question: 7**

---

By integrating live dashboards, DataV can present and monitor business data

simultaneously. This data-driven approach enables for well-organized data mining and analysis allowing the user to seize new opportunities that otherwise might remain hidden. It can support wide range of databases and data formats. Which of the following options DataV does not support?

- A. Alibaba Cloud' s AnalyticDB, ApsaraDB
- B. Static data in CSV and JSON formats
- C. Oracle Database
- D. MaxCompute Project

---

**Answer: D**

---

---

### **Question: 8**

---

You want to understand more about how users browse your public website. For example, you want to know which pages they visit prior to placing an order. You have a server farm of 100 web servers hosting your website. Which is the most efficient process to gather these web servers across logs into traditional Hadoop ecosystem.

- A. Just copy them into HDFS using curl
- B. Ingest the server web logs into HDFS using Apache Flume
- C. Channel these clickstreams into Hadoop using Hadoop Streaming
- D. Import all user clicks from your OLTP databases into Hadoop using Sqoop

---

**Answer: B**

---

---

### **Question: 9**

---

Your company stores user profile records in an OLTP databases. You want to join the serecords with web server logs you have already ingested into the Hadoop file system. What is the best way to obtain and ingest these user records?

- A. Ingest with Hadoop streaming
- B. Ingest using Hive
- C. Ingest with sqoop import
- D. Ingest with Pig's LOAD command

---

**Answer: C**

---