Question 1

These primary tool in use, and the data format is Optimized Row Columnar (ORC). All ORC files have been successfully copied to a Cloud Storage bucket. You need to replicate some data to the cluster's local Hadoop Distributed File System (HDFS) to maximize performance. What are two ways to start using Hive in Cloud Dataproc? (Choose two.)

Options:

A) Run the gsutil utility to transfer all ORC files from the Cloud Storage bucket to HDFS. Mount the Hive tables locally.

B) Run the gsutil utility to transfer all ORC files from the Cloud Storage bucket to any node of the Dataproc cluster. Mount the Hive tables locally.

C) Run the gsutil utility to transfer all ORC files from the Cloud Storage bucket to the master node of the Dataproc cluster. Then run the Hadoop utility to copy them do HDFS. Mount the Hive tables from HDFS.

D) Leverage Cloud Storage connector for Hadoop to mount the ORC files as external Hive tables. Replicate external Hive tables to the native ones.

E) Load the ORC files into BigQuery. Leverage BigQuery connector for Hadoop to mount the BigQuery tables as external Hive tables. Replicate external Hive tables to the native ones.

Answer:

B, C

Question 2

You have uploaded 5 years of log data to Cloud Storage A user reported that some data points in the log data are outside of their expected ranges, which indicates errors You need to address this issue and be able to run the process again in the future while keeping the original data for compliance reasons What should you do?

Options:

A) Import the data from Cloud Storage into BigQuery Create a new BigQuery table, and skip the rows with errors.

B) Create a Compute Engine instance and create a new copy of the data in Cloud Storage Skip the rows with errors

C) Create a Cloud Dataflow workflow that reads the data from Cloud Storage, checks for values outside the expected range, sets the value to an appropriate default, and writes the updated records to a new dataset in
Cloud Storage

D) Create a Cloud Dataflow workflow that reads the data from Cloud Storage, checks for values outside the expected range, sets the value to an appropriate default, and writes the updated records to the same dataset in Cloud Storage

Answer:

C

Question 3

MJTelco's Google Cloud Dataflow pipeline is now ready to start receiving data from the 50,000 installations. You want to allow Cloud Dataflow to scale its compute power up as required. Which Cloud Dataflow pipeline configuration setting should you update?

Options:

A) The zone

B) The number of workers

C) The disk size per worker

D) The maximum number of workers

Answer:

A

Question 4

You are designing the database schema for a machine learning-based food ordering service that will predict what users want to eat. Here is some of the information you need to store:

The user profile: What the user likes and doesn't like to eat

The user account information: Name, address, preferred meal times

The order information: When orders are made, from where, to whom

The database will be used to store all the transactional data of the product. You want to optimize the data schem

a. Which Google Cloud Platform product should you use?

Options:

A) BigQuery

B) Cloud SQL

C) Cloud Bigtable

D) Cloud Datastore

Answer:

A

Question 5

Question Type: **MultipleChoice**

Suppose you have a table that includes a nested column called "city" inside a column called "person", but when you try to submit the following query in BigQuery, it gives you an error.

SELECT person FROM `project1.example.table1` WHERE city = "London"

How would you correct the error?

Options:

A) Add ', UNNEST(person)' before the WHERE clause.

B) Change 'person' to 'person.city'.

C) Change 'person' to 'city.person'.

D) Add ', UNNEST(city)' before the WHERE clause.

Answer:

A

Question 6

Question Type: MultipleChoice

You are integrating one of your internal IT applications and Google BigQuery, so users can query BigQuery from the application's interface. You do not want individual users to authenticate to BigQuery and you do not want to give them access to the dataset. You need to securely access BigQuery from your IT application.

What should you do?

Options:

A) Create groups for your users and give those groups access to the dataset

B) Integrate with a single sign-on (SSO) platform, and pass each user's credentials along with the query request

C) Create a service account and grant dataset access to that account. Use the service account's private key to access the dataset

D) Create a dummy user and grant dataset access to that user. Store the username and password for that user in a file on the files system, and use those credentials to access the BigQuery dataset

Question 7

Question Type: MultipleChoice

Your company is using WHILECARD tables to query data across multiple tables with similar names. The SQL statement is currently failing with the following error:

Syntax error : Expected end of statement but got "-" at [4:11]

SELECT age

FROM

bigquery-public-data.noaa_gsod.gsod

WHERE

age != 99

AND_TABLE_SUFFIX = '1929'

ORDER BY

age DESC

Which table name will make the SQL statement work correctly?

Options:

A) 'bigquery-public-data.noaa_gsod.gsod'

B) bigquery-public-data.noaa_gsod.gsod*

C) 'bigquery-public-data.noaa_gsod.gsod'*

D) 'bigquery-public-data.noaa_gsod.gsod*`
References:

Answer:

B

Question 8

Question Type: **MultipleChoice**

Your company's on-premises Apache Hadoop servers are approaching end-of-life, and IT has decided to migrate the cluster to Google Cloud Dataproc. A like-for-like migration of the cluster would require 50 TB of Google Persistent Disk per node. The CIO is concerned about the cost of using that much block storage. You want to minimize the storage cost of the migration. What should you do?

Options:

A) Put the data into Google Cloud Storage.

B) Use preemptible virtual machines (VMs) for the Cloud Dataproc cluster.

C) Tune the Cloud Dataproc cluster so that there is just enough disk for all data.

D) Migrate some of the cold data into Google Cloud Storage, and keep only the hot data in Persistent Disk.
References:

Answer:

B