

Latest Version: 6.0

Question: 1

Your company is loading comma-separated values (CSV) files into BigQuery. The data is fully imported successfully; however, the imported data is not matching byte-to-byte to the source file.

What is the most likely cause of this problem?

Response:

- A. The CSV data loaded in BigQuery is not flagged as CSV.
- B. The CSV data had invalid rows that were skipped on import.
- C. The CSV data has not gone through an ETL phase before loading into BigQuery.
- D. The CSV data loaded in BigQuery is not using BigQuery's default encoding.

Answer: D

Question: 2

You are using Pub/Sub to stream inventory updates from many point-of-sale (POS) terminals into BigQuery. Each update event has the following information: product identifier "prodSku", change increment "quantityDelta", POS identification "termId", and "messageId" which is created for each push attempt from the terminal.

During a network outage, you discovered that duplicated messages were sent, causing the inventory system to over-count the changes. You determine that the terminal application has design problems and may send the same event more than once during push retries. You want to ensure that the inventory update is accurate.

What should you do?

Response:

- A. Inspect the "publishTime" of each message. Make sure that messages whose "publishTime" values match rows in the BigQuery table are discarded.
- B. Inspect the "messageId" of each message. Make sure that any messages whose "messageId" values match corresponding rows in the BigQuery table are discarded.
- C. Instead of specifying a change increment for "quantityDelta", always use the derived inventory value after the increment has been applied. Name the new attribute "adjustedQuantity".
- D. Add another attribute orderId to the message payload to mark the unique check-out order across all terminals. Make sure that messages whose "orderId" and "prodSku" values match corresponding rows in the BigQuery table are discarded.

Answer: D

Question: 3

You have 250,000 devices which produce a JSON device status event every 10 seconds. You want to capture this event data for outlier time series analysis. What should you do?

Response:

- A. Ship the data into BigQuery. Develop a custom application that uses the BigQuery API to query the dataset and displays device outlier data based on your business requirements.
- B. Ship the data into BigQuery. Use the BigQuery console to query the dataset and display device outlier data based on your business requirements.
- C. Ship the data into Cloud Bigtable. Use the Cloud Bigtable cbt tool to display device outlier data based on your business requirements.
- D. Ship the data into Cloud Bigtable. Install and use the HBase shell for Cloud Bigtable to query the table for device outlier data based on your business requirements.

Answer: C

Question: 4

You created a Cloud SQL database that uses replication to improve read performance. Occasionally, the read replica will be unavailable. You haven't noticed a pattern, but the disruptions occur once or twice a month. No DBA operations are occurring when the incidents occur. What might be the cause of this issue?

Response:

- A. The read replica is being promoted to a standalone Cloud SQL instance.
- B. Maintenance is occurring on the read replica.
- C. A backup is being performed on the read replica.
- D. The primary Cloud SQL instance is failing over to the read replica.

Answer: B

Question: 5

A team of machine learning engineers are creating a repository of data for training and testing machine learning models. All of the engineers work in the same city, and they all contribute datasets to the repository. The data files will be accessed frequently, usually at least once a week.

The data scientists want to minimize their storage costs. They plan to use Cloud Storage; what storage class would you recommend?

Response:

- A. Regional

- B. Multi-regional
- C. Nearline
- D. Coldline

Answer: A

Question: 6

The business owners of a data warehouse have determined that the current design of the data warehouse is not meeting their needs. In addition to having data about the state of systems at certain points in time, they need to know about all the times that data changed between those points in time. What kind of data warehousing pipeline should be used to meet this new requirement?

Response:

- A. ETL
- B. ELT
- C. Extraction and load
- D. Change data capture

Answer: D

Question: 7

You are working on a project with two compliance requirements. The first requirement states that your developers should be able to see the Google Cloud billing charges for only their own projects. The second requirement states that your finance team members can set budgets and view the current charges for all projects in the organization.

The finance team should not be able to view the project contents. You want to set permissions. What should you do?

Response:

- A. Add the finance team members to the Billing Administrator role for each of the billing accounts that they need to manage. Add the developers to the Viewer role for the Project.
- B. Add the finance team members to the default IAM Owner role. Add the developers to a custom role that allows them to see their own spend only.
- C. Add the developers and finance managers to the Viewer role for the Project.
- D. Add the finance team to the Viewer role for the Project. Add the developers to the Security Reviewer role for each of the billing accounts.

Answer: A

Question: 8

You are designing a relational data repository on Google Cloud to grow as needed. The data will be transactionally consistent and added from any location in the world. You want to monitor and adjust node count for input traffic, which can spike unpredictably.

What should you do?

Response:

- A. Use Cloud Spanner for storage. Monitor storage usage and increase node count if more than 70% utilized.
- B. Use Cloud Spanner for storage. Monitor CPU utilization and increase node count if more than 70% utilized for your time span.
- C. Use Cloud Bigtable for storage. Monitor data stored and increase node count if more than 70% utilized.
- D. Use Cloud Bigtable for storage. Monitor CPU utilization and increase node count if more than 70% utilized for your time span.

Answer: B

Question: 9

You need to stream time-series data in Avro format, and then write this to both BigQuery and Cloud Bigtable simultaneously using Dataflow. You want to achieve minimal end-to-end latency.

Your business requirements state this needs to be completed as quickly as possible. What should you do?

Response:

- A. Create a pipeline and use ParDo transform.
- B. Create a pipeline that groups the data into a PCollection and uses the Combine transform.
- C. Create a pipeline that groups data using a PCollection and then uses Bigtable and BigQueryIO transforms.
- D. Create a pipeline that groups data using a PCollection, and then use Avro I/O transform to write to Cloud Storage. After the data is written, load the data from Cloud Storage into BigQuery and Bigtable.

Answer: C

Question: 10

A team of researchers is analyzing buying patterns of customers of a national grocery store chain. They are especially interested in sets of products that customers frequently buy together. The researchers plan to use association rules for this frequent pattern mining. What machine learning option in GCP would you recommend?

Response:

- A. Cloud Dataflow

- B. Spark MLlib
- C. BigQuery ML
- D. AutoML Tables

Answer: B