

# Latest Version: 6.0

## Question: 1

The chief financial officer of your company would like to build a program to predict which customers will likely be late paying their bills. The company has an enterprise data warehouse in BigQuery containing all the data related to customers, billing, and payments.

The company does not have anyone with machine learning experience, but it does have analysts and data scientists experienced in SQL, Python, and Java.

The analysts and data scientists will generate and test a large number of models, so they prefer fast model building. What service would you recommend using to build the model?

Response:

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

**Answer: C**

## Question: 2

A database administrator (DBA) who is new to Google Cloud has asked for your help configuring network access to a Cloud SQL PostgreSQL database. The DBA wants to ensure that traffic is encrypted while minimizing administrative tasks, such as managing SQL certificates. What would you recommend?

Response:

- A. Use the TLS protocol
- B. Use Cloud SQL Proxy
- C. Use a private IP address
- D. Configure the database instance to use auto-encryption

**Answer: B**

## Question: 3

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**

### Question: 4

An analyst would like to build a machine learning model to classify rows of data in a dataset. There are two categories into which the rows can be grouped: Type A and Type B. The dataset has over 1 million rows, and each row has 32 attributes or features. The analyst does not know which features are important.

A labeled training set is available with a sufficient number of rows to train a model. The analyst would like the most accurate model possible with the least amount of effort on the analyst's part. What would you recommend?

Response:

- A. Kubeflow
- B. Spark MLlib
- C. AutoML Tables
- D. AutoML Natural Language

**Answer: C**

### Question: 5

Your department is experimenting with using Cloud Spanner for a globally accessible database. You are starting with a pilot project using a regional instance. You would like to follow Google's recommendations for the maximum sustained CPU utilization of a regional instance. What is the maximum CPU utilization that you would target?

Response:

- A. 50%
- B. 65%
- C. 75%
- D. 45%

**Answer: B**

### 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 want to publish system metrics to Google Cloud from a large number of on-prem hypervisors and VMs for analysis and creation of dashboards.

You have an existing custom monitoring agent deployed to all the hypervisors and your on-prem metrics system is unable to handle the load. You want to design a system that can collect and store metrics at scale. You don't want to manage your own time series database.

Metrics from all agents should be written to the same table but agents must not have permission to modify or read data written by other agents. What should you do?

Response:

- A. Modify the monitoring agent to publish protobuf messages to Pub/Sub. Use a Dataproc cluster or Dataflow job to consume messages from Pub/Sub and write to BigTable.
- B. Modify the monitoring agent to write protobuf messages directly to BigTable.
- C. Modify the monitoring agent to write protobuf messages to HBase deployed on Compute Engine VM Instances
- D. Modify the monitoring agent to write protobuf messages to Pub/Sub. Use a Dataproc cluster or Dataflow job to consume messages from Pub/Sub and write to Cassandra deployed on Compute Engine VM Instances.

**Answer: A**

### Question: 8

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: 9

Auditors have informed your company CFO that to comply with a new regulation, your company will need to ensure that financial reporting data is kept for at least three years. The CFO asks for your advice on how to comply with the regulation with the least administrative overhead. What would you recommend?

Response:

- A. Store the data on Coldline storage
- B. Store the data on multi-regional storage
- C. Define a data retention policy
- D. Define a lifecycle policy

**Answer: C**

## Question: 10

Your company is streaming real-time sensor data from their factory floor into Bigtable and they have noticed extremely poor performance. How should the row key be redesigned to improve Bigtable performance on queries that populate real-time dashboards?

Response:

- A. Use a row key of the form <timestamp>.
- B. Use a row key of the form <sensorid>.
- C. Use a row key of the form <timestamp>#<sensorid>.

D. Use a row key of the form <sensorid>#<timestamp>.

**Answer: D**