



Free Questions for Professional-Data-Engineer by dumpshq

Shared by Townsend on 20-10-2022

For More Free Questions and Preparation Resources

Check the Links on Last Page

Question 1

Question Type: MultipleChoice

You are building a data pipeline on Google Cloud. You need to prepare data using a casual method for a machine-learning process. You want to support a logistic regression model. You also need to monitor and adjust for null values, which must remain real-valued and cannot be removed. What should you do?

Options:

- A-** Use Cloud Dataprep to find null values in sample source data. Convert all nulls to 'none' using a Cloud Dataproc job.
- B-** Use Cloud Dataprep to find null values in sample source data. Convert all nulls to 0 using a Cloud Dataprep job.
- C-** Use Cloud Dataflow to find null values in sample source data. Convert all nulls to 'none' using a Cloud Dataprep job.
- D-** Use Cloud Dataflow to find null values in sample source data. Convert all nulls to using a custom script.

Answer:

C

Question 2

Question Type: MultipleChoice

You are migrating a table to BigQuery and are deciding on the data model. Your table stores information related to purchases made across several store locations and includes information like the time of the transaction, items purchased, the store ID and the city and state in which the store is located. You frequently query this table to see how many of each item were sold over the past 30 days and to look at purchasing trends by state, city, and individual store. You want to model this table to minimize query time and cost. What should you do?

Options:

- A- Partition by transaction time; cluster by state first, then city then store ID
- B- Partition by transaction time; cluster by store ID first, then city, then state
- C- Top-level cluster by state first, then city then store
- D- Top-level cluster by store ID first, then city then state.

Answer:

C

Question 3

Question Type: MultipleChoice

Government regulations in the banking industry mandate the protection of client's personally identifiable information (PII). Your company requires PII to be access controlled encrypted and compliant with major data protection standards. In addition to using Cloud Data Loss Prevention (Cloud DLP) you want to follow Google-recommended practices and use service accounts to control access to PII. What should you do?

Options:

- A-** Assign the required identity and Access Management (IAM) roles to every employee, and create a single service account to access protect resources
- B-** Use one service account to access a Cloud SQL database and use separate service accounts for each human user
- C-** Use Cloud Storage to comply with major data protection standards. Use one service account shared by all users
- D-** Use Cloud Storage to comply with major data protection standards. Use multiple service accounts attached to IAM groups to grant the appropriate access to each group

Answer:

D

Question 4

Question Type: MultipleChoice

You are updating the code for a subscriber to a Pub/Sub feed. You are concerned that upon deployment the subscriber may erroneously acknowledge messages, leading to message loss. Your subscriber is not set up to retain acknowledged messages. What should you do to ensure that you can recover from errors after deployment?

Options:

- A-** Use Cloud Build for your deployment. If an error occurs after deployment, use a Seek operation to locate a timestamp logged by Cloud Build at the start of the deployment.
- B-** Create a Pub/Sub snapshot before deploying new subscriber code. Use a Seek operation to re-deliver messages that became available after the snapshot was created.
- C-** Set up the Pub/Sub emulator on your local machine. Validate the behavior of your new subscriber code before deploying it to production.
- D-** Enable dead-lettering on the Pub/Sub topic to capture messages that aren't successfully acknowledged. If an error occurs after deployment, re-deliver any messages captured by the dead-letter queue.

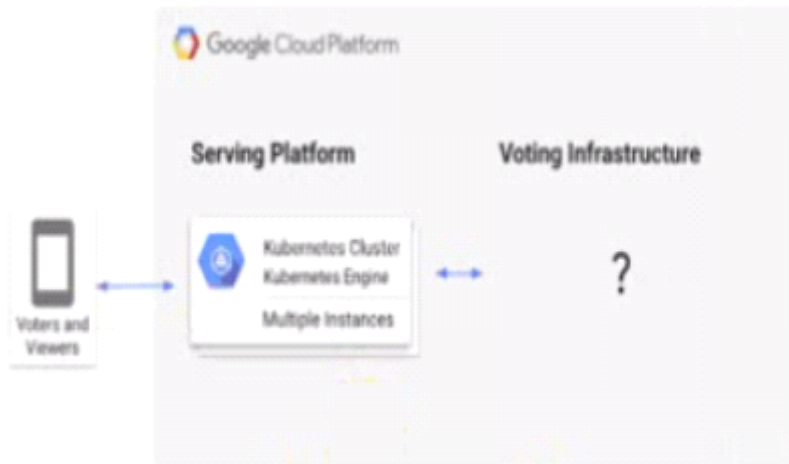
Answer:

B

Question 5

Question Type: MultipleChoice

A live TV show asks viewers to cast votes using their mobile phones. The event generates a large volume of data during a 3 minute period. You are in charge of the Voting restructure* and must ensure that the platform can handle the load and Hal all votes are processed. You must display partial results write voting is open. After voting doses you need to count the votes exactly once white optimizing cost. What should you do?



Options:

A- Create a Memorystore instance with a high availability (HA) configuration

B- Write votes to a Pub Sub topic and have Cloud Functions subscribe to it and write votes to BigQuery

C- Write votes to a Pub/Sub topic and load into both Bigtable and BigQuery via a Dataflow pipeline Query Bigtable for real-time results and BigQuery for later analysis Shutdown the Bigtable instance when voting concludes

D Create a Cloud SQL for PostgreSQL database with high availability (HA) configuration and multiple read replicas

Answer:

C

Question 6

Question Type: MultipleChoice

You want to optimize your queries for cost and performance. How should you structure your data?

Options:

A- Partition table data by create_date, location_id and device_version

B- Partition table data by create_date cluster table data by location_id and device_version

C- Cluster table data by create_date location_id and device_version

D- Cluster table data by create_date partition by location and device_version

Answer:

B

Question 7

Question Type: MultipleChoice

You have data pipelines running on BigQuery, Cloud Dataflow, and Cloud Dataproc. You need to perform health checks and monitor their behavior, and then notify the team managing the pipelines if they fail. You also need to be able to work across multiple projects. Your preference is to use managed products or features of the platform. What should you do?

Options:

- A-** Export the information to Cloud Stackdriver, and set up an Alerting policy
- B-** Run a Virtual Machine in Compute Engine with Airflow, and export the information to Stackdriver
- C-** Export the logs to BigQuery, and set up App Engine to read that information and send emails if you find a failure in the logs
- D-** Develop an App Engine application to consume logs using GCP API calls, and send emails if you find a failure in the logs

Answer:

B

Question 8

Question Type: MultipleChoice

You want to archive data in Cloud Storage. Because some data is very sensitive, you want to use the "Trust No One" (TNO) approach to encrypt your data to prevent the cloud provider staff from decrypting your data.

a. What should you do?

Options:

- A-** Use `gcloud kms keys create` to create a symmetric key. Then use `gcloud kms encrypt` to encrypt each archival file with the key and unique additional authenticated data (AAD). Use `gsutil cp` to upload each encrypted file to the Cloud Storage bucket, and keep the AAD outside of Google Cloud.
- B-** Use `gcloud kms keys create` to create a symmetric key. Then use `gcloud kms encrypt` to encrypt each archival file with the key. Use `gsutil cp` to upload each encrypted file to the Cloud Storage bucket. Manually destroy the key previously used for encryption, and rotate the key once and rotate the key once.
- C-** Specify customer-supplied encryption key (CSEK) in the `.boto` configuration file. Use `gsutil cp` to upload each archival file to the Cloud Storage bucket. Save the CSEK in Cloud Memorystore as permanent storage of the secret.

D- Specify customer-supplied encryption key (CSEK) in the .boto configuration file. Use gsutil cp to upload each archival file to the Cloud Storage bucket. Save the CSEK in a different project that only the security team can access.

Answer:

B

Question 9

Question Type: MultipleChoice

You need to choose a database to store time series CPU and memory usage for millions of computers. You need to store this data in one-second interval samples. Analysts will be performing real-time, ad hoc analytics against the database. You want to avoid being charged for every query executed and ensure that the schema design will allow for future growth of the dataset. Which database and data model should you choose?

Options:

A- Create a table in BigQuery, and append the new samples for CPU and memory to the table

B- Create a wide table in BigQuery, create a column for the sample value at each second, and update the row with the interval for each second

C- Create a narrow table in Cloud Bigtable with a row key that combines the Computer Engine computer identifier with the sample time at each second

D- Create a wide table in Cloud Bigtable with a row key that combines the computer identifier with the sample time at each minute, and combine the values for each second as column data.

Answer:

C

Explanation:

A tall and narrow table has a small number of events per row, which could be just one event, whereas a short and wide table has a large number of events per row. As explained in a moment, tall and narrow tables are best suited for time-series data. For time series, you should generally use tall and narrow tables. This is for two reasons: Storing one event per row makes it easier to run queries against your data. Storing many events per row makes it more likely that the total row size will exceed the recommended maximum (see Rows can be big but are not infinite). https://cloud.google.com/bigtable/docs/schema-design-time-series#patterns_for_row_key_design

Question 10

Question Type: MultipleChoice

You work for a mid-sized enterprise that needs to move its operational system transaction data from an on-premises database to GCP. The database is about 20 TB in size. Which database should you choose?

Options:

- A- Cloud SQL
- B- Cloud Bigtable
- C- Cloud Spanner
- D- Cloud Datastore

Answer:

A

To Get Premium Files for Professional-Data-Engineer Visit

<https://www.p2pexams.com/products/professional-data-engineer>

For More Free Questions Visit

<https://www.p2pexams.com/google/pdf/professional-data-engineer>

