

**Question #164Topic 1**

You are working on a linear regression model on BigQuery ML to predict a customer's likelihood of purchasing your company's products. Your model uses a city name variable as a key predictive component. In order to train and serve the model, your data must be organized in columns. You want to prepare your data using the least amount of coding while maintaining the predictable variables. What should you do?

- A. Create a new view with BigQuery that does not include a column with city information.
- B. Use SQL in BigQuery to transform the state column using a one-hot encoding method, and make each city a column with binary values.
- C. Use TensorFlow to create a categorical variable with a vocabulary list. Create the vocabulary file and upload that as part of your model to BigQuery ML.
- **D. Use Cloud Data Fusion to assign each city to a region that is labeled as 1, 2, 3, 4, or 5, and then use that number to represent the city in the model.**

**Question #165Topic 1**

You work for a large bank that operates in locations throughout North America. You are setting up a data storage system that will handle bank account transactions. You require ACID compliance and the ability to access data with SQL. Which solution is appropriate?

- A. Store transaction data in Cloud Spanner. Enable stale reads to reduce latency.
- B. Store transaction in Cloud Spanner. Use locking read-write transactions.
- **C. Store transaction data in BigQuery. Disabled the query cache to ensure consistency.**
- D. Store transaction data in Cloud SQL. Use a federated query BigQuery for analysis.

**Question #166Topic 1**

A shipping company has live package-tracking data that is sent to an Apache Kafka stream in real time. This is then loaded into BigQuery. Analysts in your company want to query the tracking data in BigQuery to analyze geospatial trends in the lifecycle of a package. The table was originally created with ingest-date partitioning. Over time, the query processing time has increased. You need to implement a change that would improve query performance in BigQuery. What should you do?

- A. Implement clustering in BigQuery on the ingest date column.
- **B. Implement clustering in BigQuery on the package-tracking ID column.**
- C. Tier older data onto Cloud Storage files and create a BigQuery table using Cloud Storage as an external data source.
- D. Re-create the table using data partitioning on the package delivery date.

**Question #167Topic 1**

Your company currently runs a large on-premises cluster using Spark, Hive, and HDFS in a colocation facility. The cluster is designed to accommodate peak usage on the system; however, many jobs are batch in nature, and usage of the cluster fluctuates quite dramatically. Your company is eager to move to the cloud to reduce the overhead associated with on-premises

infrastructure and maintenance and to benefit from the cost savings. They are also hoping to modernize their existing infrastructure to use more serverless offerings in order to take advantage of the cloud. Because of the timing of their contract renewal with the colocation facility, they have only 2 months for their initial migration. How would you recommend they approach their upcoming migration strategy so they can maximize their cost savings in the cloud while still executing the migration in time?

- A. Migrate the workloads to Dataproc plus HDFS; modernize later.
- B. Migrate the workloads to Dataproc plus Cloud Storage; modernize later.
- C. Migrate the Spark workload to Dataproc plus HDFS, and modernize the Hive workload for BigQuery.
- **D. Modernize the Spark workload for Dataflow and the Hive workload for BigQuery.**

#### **Question #168***Topic 1*

You work for a financial institution that lets customers register online. As new customers register, their user data is sent to Pub/Sub before being ingested into BigQuery. For security reasons, you decide to redact your customers' Government issued Identification Number while allowing customer service representatives to view the original values when necessary. What should you do?

- A. Use BigQuery's built-in AEAD encryption to encrypt the SSN column. Save the keys to a new table that is only viewable by permissioned users.
- B. Use BigQuery column-level security. Set the table permissions so that only members of the Customer Service user group can see the SSN column.
- C. Before loading the data into BigQuery, use Cloud Data Loss Prevention (DLP) to replace input values with a cryptographic hash.
- **D. Before loading the data into BigQuery, use Cloud Data Loss Prevention (DLP) to replace input values with a cryptographic format-preserving encryption token.**

#### **Question #169***Topic 1*

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. How would you model this table for the best query performance?

- **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 ID.
- D. Top-level cluster by store ID first, then city, then state.

#### **Question #170***Topic 1*

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?

- A. Set up the Pub/Sub emulator on your local machine. Validate the behavior of your new subscriber logic before deploying it to production.
- **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. 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.
- 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.

#### **Question #171Topic 1**

You work for a large real estate firm and are preparing 6 TB of home sales data to be used for machine learning. You will use SQL to transform the data and use BigQuery ML to create a machine learning model. You plan to use the model for predictions against a raw dataset that has not been transformed. How should you set up your workflow in order to prevent skew at prediction time?

- **A. When creating your model, use BigQuery's TRANSFORM clause to define preprocessing steps. At prediction time, use BigQuery's ML.EVALUATE clause without specifying any transformations on the raw input data.**
- B. When creating your model, use BigQuery's TRANSFORM clause to define preprocessing steps. Before requesting predictions, use a saved query to transform your raw input data, and then use ML.EVALUATE.
- C. Use a BigQuery view to define your preprocessing logic. When creating your model, use the view as your model training data. At prediction time, use BigQuery's ML.EVALUATE clause without specifying any transformations on the raw input data.
- D. Preprocess all data using Dataflow. At prediction time, use BigQuery's ML.EVALUATE clause without specifying any further transformations on the input data.

#### **Question #172Topic 1**

You are analyzing the price of a company's stock. Every 5 seconds, you need to compute a moving average of the past 30 seconds' worth of data. You are reading data from Pub/Sub and using DataFlow to conduct the analysis. How should you set up your windowed pipeline?

- A. Use a fixed window with a duration of 5 seconds. Emit results by setting the following trigger: `AfterProcessingTime.pastFirstElementInPane().plusDelayOf(Duration.standardSeconds(30))`

- B. Use a fixed window with a duration of 30 seconds. Emit results by setting the following trigger: `AfterWatermark.pastEndOfWindow().plusDelayOf (Duration.standardSeconds(5))`
- C. Use a sliding window with a duration of 5 seconds. Emit results by setting the following trigger: `AfterProcessingTime.pastFirstElementInPane().plusDelayOf (Duration.standardSeconds(30))`
- **D. Use a sliding window with a duration of 30 seconds and a period of 5 seconds. Emit results by setting the following trigger: `AfterWatermark.pastEndOfWindow ()`**

### Question #173Topic 1

You are designing a pipeline that publishes application events to a Pub/Sub topic. Although message ordering is not important, you need to be able to aggregate events across disjoint hourly intervals before loading the results to BigQuery for analysis. What technology should you use to process and load this data to BigQuery while ensuring that it will scale with large volumes of events?

- A. Create a Cloud Function to perform the necessary data processing that executes using the Pub/Sub trigger every time a new message is published to the topic.
- B. Schedule a Cloud Function to run hourly, pulling all available messages from the Pub/Sub topic and performing the necessary aggregations.
- C. Schedule a batch Dataflow job to run hourly, pulling all available messages from the Pub/Sub topic and performing the necessary aggregations.
- **D. Create a streaming Dataflow job that reads continually from the Pub/Sub topic and performs the necessary aggregations using tumbling windows.**

### Question #174Topic 1

You work for a large financial institution that is planning to use Dialogflow to create a chatbot for the company's mobile app. You have reviewed old chat logs and tagged each conversation for intent based on each customer's stated intention for contacting customer service. About 70% of customer requests are simple requests that are solved within 10 intents. The remaining 30% of inquiries require much longer, more complicated requests. Which intents should you automate first?

- **A. Automate the 10 intents that cover 70% of the requests so that live agents can handle more complicated requests.**
- B. Automate the more complicated requests first because those require more of the agents' time.
- C. Automate a blend of the shortest and longest intents to be representative of all intents.
- D. Automate intents in places where common words such as 'payment' appear only once so the software isn't confused.

### Question #175Topic 1

Your company is implementing a data warehouse using BigQuery, and you have been tasked with designing the data model. You move your on-premises sales data warehouse with a star data schema to BigQuery but notice performance issues when querying the data of the past 30 days.

Based on Google's recommended practices, what should you do to speed up the query without increasing storage costs?

- A. Denormalize the data.
- B. Shard the data by customer ID.
- C. Materialize the dimensional data in views.
- **D. Partition the data by transaction date.**

#### **Question #176Topic 1**

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?

- 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 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 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.

#### **Question #177Topic 1**

You want to rebuild your batch pipeline for structured data on Google Cloud. You are using PySpark to conduct data transformations at scale, but your pipelines are taking over twelve hours to run. To expedite development and pipeline run time, you want to use a serverless tool and SQL syntax. You have already moved your raw data into Cloud Storage. How should you build the pipeline on Google Cloud while meeting speed and processing requirements?

- A. Convert your PySpark commands into SparkSQL queries to transform the data, and then run your pipeline on Dataproc to write the data into BigQuery.
- B. Ingest your data into Cloud SQL, convert your PySpark commands into SparkSQL queries to transform the data, and then use federated queries from BigQuery for machine learning.
- **C. Ingest your data into BigQuery from Cloud Storage, convert your PySpark commands into BigQuery SQL queries to transform the data, and then write the transformations to a new table.**
- D. Use Apache Beam Python SDK to build the transformation pipelines, and write the data into BigQuery.

#### **Question #178Topic 1**

You are testing a Dataflow pipeline to ingest and transform text files. The files are compressed gzip, errors are written to a dead-letter queue, and you are using SideInputs to join data. You noticed that the pipeline is taking longer to complete than expected; what should you do to expedite the Dataflow job?

- A. Switch to compressed Avro files.
- **B. Reduce the batch size.**
- C. Retry records that throw an error.
- D. Use CoGroupByKey instead of the SideInput.

#### **Question #179Topic 1**

You are building a real-time prediction engine that streams files, which may contain PII (personal identifiable information) data, into Cloud Storage and eventually into BigQuery. You want to ensure that the sensitive data is masked but still maintains referential integrity, because names and emails are often used as join keys.

How should you use the Cloud Data Loss Prevention API (DLP API) to ensure that the PII data is not accessible by unauthorized individuals?

- **A. Create a pseudonym by replacing the PII data with cryptogenic tokens, and store the non-tokenized data in a locked-down bucket.**
- B. Redact all PII data, and store a version of the unredacted data in a locked-down bucket.
- C. Scan every table in BigQuery, and mask the data it finds that has PII.
- D. Create a pseudonym by replacing PII data with a cryptographic format-preserving token.

#### **Question #180Topic 1**

You are migrating an application that tracks library books and information about each book, such as author or year published, from an on-premises data warehouse to BigQuery. In your current relational database, the author information is kept in a separate table and joined to the book information on a common key. Based on Google's recommended practice for schema design, how would you structure the data to ensure optimal speed of queries about the author of each book that has been borrowed?

- A. Keep the schema the same, maintain the different tables for the book and each of the attributes, and query as you are doing today.
- B. Create a table that is wide and includes a column for each attribute, including the author's first name, last name, date of birth, etc.
- **C. Create a table that includes information about the books and authors, but nest the author fields inside the author column.**
- D. Keep the schema the same, create a view that joins all of the tables, and always query the view.

#### **Question #181Topic 1**

You need to give new website users a globally unique identifier (GUID) using a service that takes in data points and returns a GUID. This data is sourced from both internal and external systems via

HTTP calls that you will make via microservices within your pipeline. There will be tens of thousands of messages per second and that can be multi-threaded. and you worry about the backpressure on the system. How should you design your pipeline to minimize that backpressure?

- A. Call out to the service via HTTP.
- B. Create the pipeline statically in the class definition.
- C. Create a new object in the startBundle method of DoFn.
- **D. Batch the job into ten-second increments.**

### **Question #182Topic 1**

You are migrating your data warehouse to Google Cloud and decommissioning your on-premises data center. Because this is a priority for your company, you know that bandwidth will be made available for the initial data load to the cloud. The files being transferred are not large in number, but each file is 90 GB.

Additionally, you want your transactional systems to continually update the warehouse on Google Cloud in real time. What tools should you use to migrate the data and ensure that it continues to write to your warehouse?

- A. Storage Transfer Service for the migration; Pub/Sub and Cloud Data Fusion for the real-time updates
- B. BigQuery Data Transfer Service for the migration; Pub/Sub and Dataproc for the real-time updates
- **C. gsutil for the migration; Pub/Sub and Dataflow for the real-time updates**
- D. gsutil for both the migration and the real-time updates

### **Question #183Topic 1**

You are using Bigtable to persist and serve stock market data for each of the major indices. To serve the trading application, you need to access only the most recent stock prices that are streaming in. How should you design your row key and tables to ensure that you can access the data with the simplest query?

- A. Create one unique table for all of the indices, and then use the index and timestamp as the row key design.
- **B. Create one unique table for all of the indices, and then use a reverse timestamp as the row key design.**
- C. For each index, have a separate table and use a timestamp as the row key design.
- D. For each index, have a separate table and use a reverse timestamp as the row key design.

### **Question #184Topic 1**

You are building a report-only data warehouse where the data is streamed into BigQuery via the streaming API. Following Google's best practices, you have both a staging and a production table for the data. How should you design your data loading to ensure that there is only one master dataset without affecting performance on either the ingestion or reporting pieces?

- A. Have a staging table that is an append-only model, and then update the production table every three hours with the changes written to staging.
- B. Have a staging table that is an append-only model, and then update the production table every ninety minutes with the changes written to staging.
- C. Have a staging table that moves the staged data over to the production table and deletes the contents of the staging table every three hours.
- **D. Have a staging table that moves the staged data over to the production table and deletes the contents of the staging table every thirty minutes.**

#### **Question #185Topic 1**

You issue a new batch job to Dataflow. The job starts successfully, processes a few elements, and then suddenly fails and shuts down. You navigate to the Dataflow monitoring interface where you find errors related to a particular DoFn in your pipeline. What is the most likely cause of the errors?

- A. Job validation
- **B. Exceptions in worker code**
- C. Graph or pipeline construction
- D. Insufficient permissions

#### **Question #186Topic 1**

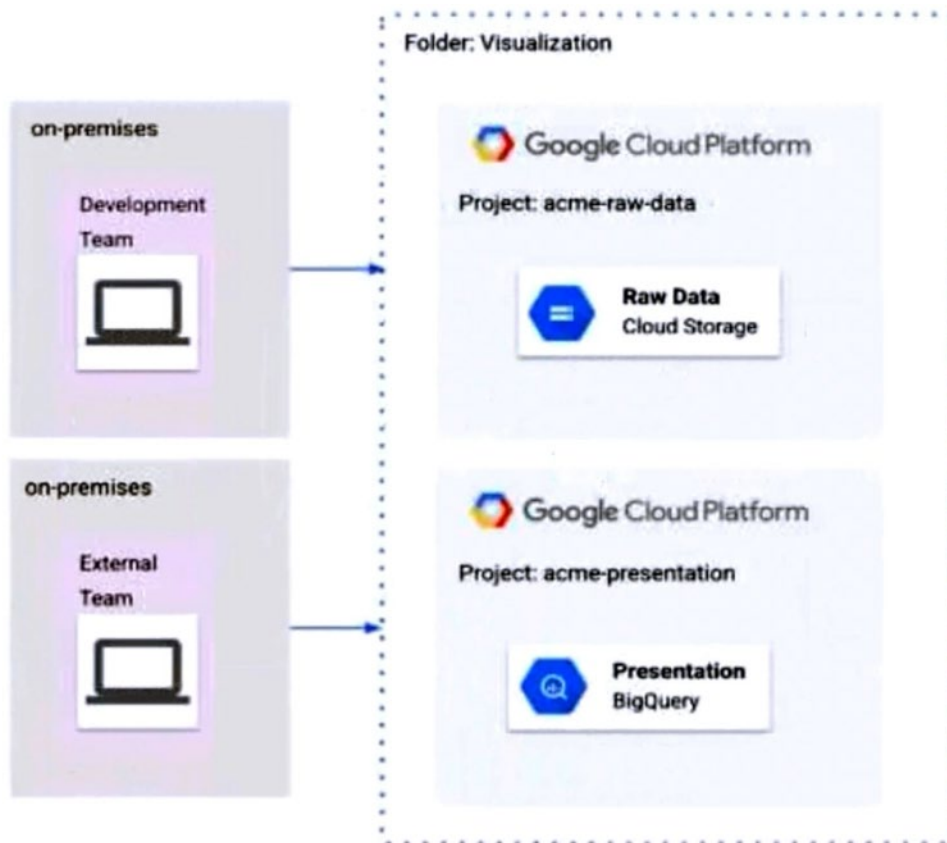
Your new customer has requested daily reports that show their net consumption of Google Cloud compute resources and who used the resources. You need to quickly and efficiently generate these daily reports. What should you do?

- A. Do daily exports of Cloud Logging data to BigQuery. Create views filtering by project, log type, resource, and user.
- B. Filter data in Cloud Logging by project, resource, and user; then export the data in CSV format.
- C. Filter data in Cloud Logging by project, log type, resource, and user, then import the data into BigQuery.
- **D. Export Cloud Logging data to Cloud Storage in CSV format. Cleanse the data using Dataprep, filtering by project, resource, and user.**

#### **Question #187Topic 1**

The Development and External teams have the project viewer Identity and Access Management (IAM) role in a folder named Visualization. You want the Development Team to be able to read data from both Cloud Storage and BigQuery, but the External Team should only be able to read data from BigQuery. What should you do?





- A. Remove Cloud Storage IAM permissions to the External Team on the acme-raw-data project.
- B. Create Virtual Private Cloud (VPC) firewall rules on the acme-raw-data project that deny all ingress traffic from the External Team CIDR range.
- **C. Create a VPC Service Controls perimeter containing both projects and BigQuery as a restricted API. Add the External Team users to the perimeter's Access Level.**
- D. Create a VPC Service Controls perimeter containing both projects and Cloud Storage as a restricted API. Add the Development Team users to the perimeter's Access Level.

### Question #188 Topic 1

Your startup has a web application that currently serves customers out of a single region in Asia. You are targeting funding that will allow your startup to serve customers globally. Your current goal is to optimize for cost, and your post-funding goal is to optimize for global presence and performance. You must use a native JDBC driver. What should you do?

- **A. Use Cloud Spanner to configure a single region instance initially, and then configure multi-region Cloud Spanner instances after securing funding.**
- B. Use a Cloud SQL for PostgreSQL highly available instance first, and Bigtable with US, Europe, and Asia replication after securing funding.

- C. Use a Cloud SQL for PostgreSQL zonal instance first, and Bigtable with US, Europe, and Asia after securing funding.
- D. Use a Cloud SQL for PostgreSQL zonal instance first, and Cloud SQL for PostgreSQL with highly available configuration after securing funding.

### Question #189Topic 1

You need to migrate 1 PB of data from an on-premises data center to Google Cloud. Data transfer time during the migration should take only a few hours. You want to follow Google-recommended practices to facilitate the large data transfer over a secure connection. What should you do?

- **A. Establish a Cloud Interconnect connection between the on-premises data center and Google Cloud, and then use the Storage Transfer Service.**
- B. Use a Transfer Appliance and have engineers manually encrypt, decrypt, and verify the data.
- C. Establish a Cloud VPN connection, start gcloud compute scp jobs in parallel, and run checksums to verify the data.
- D. Reduce the data into 3 TB batches, transfer the data using gsutil, and run checksums to verify the data.

### Question #190Topic 1

You are loading CSV files from Cloud Storage to BigQuery. The files have known data quality issues, including mismatched data types, such as STRINGS and INT64s in the same column, and inconsistent formatting of values such as phone numbers or addresses. You need to create the data pipeline to maintain data quality and perform the required cleansing and transformation. What should you do?

- **A. Use Data Fusion to transform the data before loading it into BigQuery.**
- B. Use Data Fusion to convert the CSV files to a self-describing data format, such as AVRO, before loading the data to BigQuery.
- C. Load the CSV files into a staging table with the desired schema, perform the transformations with SQL, and then write the results to the final destination table.
- D. Create a table with the desired schema, load the CSV files into the table, and perform the transformations in place using SQL.

### Question #191Topic 1

You are developing a new deep learning model that predicts a customer's likelihood to buy on your ecommerce site. After running an evaluation of the model against both the original training data and new test data, you find that your model is overfitting the data. You want to improve the accuracy of the model when predicting new data. What should you do?

- A. Increase the size of the training dataset, and increase the number of input features.
- **B. Increase the size of the training dataset, and decrease the number of input features.**
- C. Reduce the size of the training dataset, and increase the number of input features.
- D. Reduce the size of the training dataset, and decrease the number of input features.

**Question #192Topic 1**

You are implementing a chatbot to help an online retailer streamline their customer service. The chatbot must be able to respond to both text and voice inquiries.

You are looking for a low-code or no-code option, and you want to be able to easily train the chatbot to provide answers to keywords. What should you do?

- A. Use the Cloud Speech-to-Text API to build a Python application in App Engine.
- B. Use the Cloud Speech-to-Text API to build a Python application in a Compute Engine instance.
- C. Use Dialogflow for simple queries and the Cloud Speech-to-Text API for complex queries.
- **D. Use Dialogflow to implement the chatbot, defining the intents based on the most common queries collected.**

**Question #193Topic 1**

An aerospace company uses a proprietary data format to store its flight data. You need to connect this new data source to BigQuery and stream the data into

BigQuery. You want to efficiently import the data into BigQuery while consuming as few resources as possible. What should you do?

- A. Write a shell script that triggers a Cloud Function that performs periodic ETL batch jobs on the new data source.
- B. Use a standard Dataflow pipeline to store the raw data in BigQuery, and then transform the format later when the data is used.
- C. Use Apache Hive to write a Dataproc job that streams the data into BigQuery in CSV format.
- **D. Use an Apache Beam custom connector to write a Dataflow pipeline that streams the data into BigQuery in Avro format.**

**Question #194Topic 1**

An online brokerage company requires a high volume trade processing architecture. You need to create a secure queuing system that triggers jobs. The jobs will run in Google Cloud and call the company's Python API to execute trades. You need to efficiently implement a solution. What should you do?

- **A. Use a Pub/Sub push subscription to trigger a Cloud Function to pass the data to the Python API.**
- B. Write an application hosted on a Compute Engine instance that makes a push subscription to the Pub/Sub topic.
- C. Write an application that makes a queue in a NoSQL database.
- D. Use Cloud Composer to subscribe to a Pub/Sub topic and call the Python API.

**Question #195Topic 1**

Your company wants to be able to retrieve large result sets of medical information from your current system, which has over 10 TBs in the database, and store the data in new tables for further query. The database must have a low-maintenance architecture and be accessible via SQL. You need to implement a cost-effective solution that can support data analytics for large result sets. What should you do?

- A. Use Cloud SQL, but first organize the data into tables. Use JOIN in queries to retrieve data.
- **B. Use BigQuery as a data warehouse. Set output destinations for caching large queries.**
- C. Use a MySQL cluster installed on a Compute Engine managed instance group for scalability.
- D. Use Cloud Spanner to replicate the data across regions. Normalize the data in a series of tables.

#### **Question #196Topic 1**

You have 15 TB of data in your on-premises data center that you want to transfer to Google Cloud. Your data changes weekly and is stored in a POSIX-compliant source. The network operations team has granted you 500 Mbps bandwidth to the public internet. You want to follow Google-recommended practices to reliably transfer your data to Google Cloud on a weekly basis. What should you do?

- A. Use Cloud Scheduler to trigger the gsutil command. Use the -m parameter for optimal parallelism.
- B. Use Transfer Appliance to migrate your data into a Google Kubernetes Engine cluster, and then configure a weekly transfer job.
- **C. Install Storage Transfer Service for on-premises data in your data center, and then configure a weekly transfer job.**
- D. Install Storage Transfer Service for on-premises data on a Google Cloud virtual machine, and then configure a weekly transfer job.

#### **Question #197Topic 1**

You are designing a system that requires an ACID-compliant database. You must ensure that the system requires minimal human intervention in case of a failure. What should you do?

- A. Configure a Cloud SQL for MySQL instance with point-in-time recovery enabled.
- **B. Configure a Cloud SQL for PostgreSQL instance with high availability enabled.**
- C. Configure a Bigtable instance with more than one cluster.
- D. Configure a BigQuery table with a multi-region configuration.

#### **Question #198Topic 1**

You are implementing workflow pipeline scheduling using open source-based tools and Google Kubernetes Engine (GKE). You want to use a Google managed service to simplify and automate the task. You also want to accommodate Shared VPC networking considerations. What should you do?

- A. Use Dataflow for your workflow pipelines. Use Cloud Run triggers for scheduling.
- B. Use Dataflow for your workflow pipelines. Use shell scripts to schedule workflows.
- C. Use Cloud Composer in a Shared VPC configuration. Place the Cloud Composer resources in the host project.
- **D. Use Cloud Composer in a Shared VPC configuration. Place the Cloud Composer resources in the service project.**

#### **Question #199Topic 1**

You are using BigQuery and Data Studio to design a customer-facing dashboard that displays large quantities of aggregated data. You expect a high volume of concurrent users. You need to optimize the dashboard to provide quick visualizations with minimal latency. What should you do?

- **A. Use BigQuery BI Engine with materialized views.**
- B. Use BigQuery BI Engine with logical views.
- C. Use BigQuery BI Engine with streaming data.
- D. Use BigQuery BI Engine with authorized views.

#### **Question #200Topic 1**

Government regulations in the banking industry mandate the protection of clients' 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?

- A. Assign the required Identity and Access Management (IAM) roles to every employee, and create a single service account to access project 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.**

#### **Question #201Topic 1**

You need to migrate a Redis database from an on-premises data center to a Memorystore for Redis instance. You want to follow Google-recommended practices and perform the migration for minimal cost, time and effort. What should you do?

- **A. Make an RDB backup of the Redis database, use the gsutil utility to copy the RDB file into a Cloud Storage bucket, and then import the RDB file into the Memorystore for Redis instance.**
- B. Make a secondary instance of the Redis database on a Compute Engine instance and then perform a live cutover.

- C. Create a Dataflow job to read the Redis database from the on-premises data center and write the data to a Memorystore for Redis instance.
- D. Write a shell script to migrate the Redis data and create a new Memorystore for Redis instance.

**Question #202Topic 1**

Your platform on your on-premises environment generates 100 GB of data daily, composed of millions of structured JSON text files. Your on-premises environment cannot be accessed from the public internet. You want to use Google Cloud products to query and explore the platform data. What should you do?

- A. Use Cloud Scheduler to copy data daily from your on-premises environment to Cloud Storage. Use the BigQuery Data Transfer Service to import data into BigQuery.
- B. Use a Transfer Appliance to copy data from your on-premises environment to Cloud Storage. Use the BigQuery Data Transfer Service to import data into BigQuery.
- **C. Use Transfer Service for on-premises data to copy data from your on-premises environment to Cloud Storage. Use the BigQuery Data Transfer Service to import data into BigQuery.**
- D. Use the BigQuery Data Transfer Service dataset copy to transfer all data into BigQuery.