Number: Associate Data Practitioner
Passing Score: 800
Time Limit: 120
File Version: 3.0

Exam Code: Associate Data Practitioner

Exam Name: Google Cloud Associate Data Practitioner

Exam A

QUESTION 1
Your organization uses a BigQuery table that is partitioned by ingestion time. You need to remove data that is older than one year to reduce your organization's storage costs. You want to use the most efficient approach while minimizing cost. What should you do?

A. Create a scheduled query that periodically runs an update statement in SQL that sets the ''deleted' column to ''yes'' for data that is more than one year old. Create a view that filters out rows that have been marked deleted.
B. Create a view that filters out rows that are older than one year.
C. Require users to specify a partition filter using the alter table statement in SQL.
D. Set the table partition expiration period to one year using the ALTER TABLE statement in SQL.

Correct Answer: D
Section:
Explanation:
Setting the table partition expiration period to one year using the ALTER TABLE statement is the most efficient and cost-effective approach. This automatically deletes data in partitions older than one year, reducing storage costs without requiring manual intervention or additional queries. It minimizes administrative overhead and ensures compliance with your data retention policy while optimizing storage usage in BigQuery.

QUESTION 2
Your company is migrating their batch transformation pipelines to Google Cloud. You need to choose a solution that supports programmatic transformations using only SQL. You also want the technology to support Git integration for version control of your pipelines. What should you do?

A. Use Cloud Data Fusion pipelines.
B. Use Dataform workflows.
C. Use Dataflow pipelines.
D. Use Cloud Composer operators.

Correct Answer: B
Section:
Explanation:
Dataform workflows are the ideal solution for migrating batch transformation pipelines to Google Cloud when you want to perform programmatic transformations using only SQL. Dataform allows you to define SQL-based workflows for data transformations and supports Git integration for version control, enabling collaboration and version tracking of your pipelines. This approach is purpose-built for SQL-driven data pipeline management and aligns perfectly with your requirements.

QUESTION 3
You manage a BigQuery table that is used for critical end-of-month reports. The table is updated weekly with new sales data. You want to prevent data loss and reporting issues if the table is accidentally deleted. What should you do?

A. Configure the time travel duration on the table to be exactly seven days. On deletion, re-create the deleted table solely from the time travel data.
B. Schedule the creation of a new snapshot of the table once a week. On deletion, re-create the deleted table using the snapshot and time travel data.
C. Create a clone of the table. On deletion, re-create the deleted table by copying the content of the clone.
D. Create a view of the table. On deletion, re-create the deleted table from the view and time travel data.

Correct Answer: B
Section:

Explanation:
Scheduling the creation of a snapshot of the table weekly ensures that you have a point-in-time backup of the table. In case of accidental deletion, you can re-create the table from the snapshot. Additionally, BigQuery's time travel feature allows you to recover data from up to seven days prior to deletion. Combining snapshots with time travel provides a robust solution for preventing data loss and ensuring reporting continuity for critical tables. This approach minimizes risks while offering flexibility for recovery.

QUESTION 4
You manage a large amount of data in Cloud Storage, including raw data, processed data, and backups. Your organization is subject to strict compliance regulations that mandate data immutability for specific data types. You want to use an efficient process to reduce storage costs while ensuring that your storage strategy meets retention requirements. What should you do?

A. Configure lifecycle management rules to transition objects to appropriate storage classes based on access patterns. Set up Object Versioning for all objects to meet immutability requirements.
B. Move objects to different storage classes based on their age and access patterns. Use Cloud Key Management Service (Cloud KMS) to encrypt specific objects with customer-managed encryption keys (CMEK) to meet immutability requirements.
C. Create a Cloud Run function to periodically check object metadata, and move objects to the appropriate storage class based on age and access patterns. Use object holds to enforce immutability for specific objects.
D. Use object holds to enforce immutability for specific objects, and configure lifecycle management rules to transition objects to appropriate storage classes based on age and access patterns.

Correct Answer: D
Section:
Explanation:
Using object holds and lifecycle management rules is the most efficient and compliant strategy for this scenario because:
Immutability: Object holds (temporary or event-based) ensure that objects cannot be deleted or overwritten, meeting strict compliance regulations for data immutability.
Cost efficiency: Lifecycle management rules automatically transition objects to more cost-effective storage classes based on their age and access patterns.
Compliance and automation: This approach ensures compliance with retention requirements while reducing manual effort, leveraging built-in Cloud Storage features.

QUESTION 5
You work for an ecommerce company that has a BigQuery dataset that contains customer purchase history, demographics, and website interactions. You need to build a machine learning (ML) model to predict which customers are most likely to make a purchase in the next month. You have limited engineering resources and need to minimize the ML expertise required for the solution. What should you do?

A. Use BigQuery ML to create a logistic regression model for purchase prediction.
B. Use Vertex AI Workbench to develop a custom model for purchase prediction.
C. Use Colab Enterprise to develop a custom model for purchase prediction.
D. Export the data to Cloud Storage, and use AutoML Tables to build a classification model for purchase prediction.

Correct Answer: A
Section:
Explanation:
Using BigQuery ML is the best solution in this case because:
Ease of use: BigQuery ML allows users to build machine learning models using SQL, which requires minimal ML expertise.
Integrated platform: Since the data already exists in BigQuery, there's no need to move it to another service, saving time and engineering resources.
Logistic regression: This is an appropriate model for binary classification tasks like predicting the likelihood of a customer making a purchase in the next month.

QUESTION 6
You are designing a pipeline to process data files that arrive in Cloud Storage by 3:00 am each day. Data processing is performed in stages, where the output of one stage becomes the input of the next. Each stage takes a long time to run. Occasionally a stage fails, and you have to address the problem. You need to ensure that the final output is generated as quickly as possible. What should you do?

A. Design a Spark program that runs under Dataproc. Code the program to wait for user input when an error is detected. Rerun the last action after correcting any stage output data errors.
B. Design the pipeline as a set of PTransforms in Dataflow. Restart the pipeline after correcting any stage output data errors.
C. Design the workflow as a Cloud Workflow instance. Code the workflow to jump to a given stage based on an input parameter. Rerun the workflow after correcting any stage output data errors.
D. Design the processing as a directed acyclic graph (DAG) in Cloud Composer. Clear the state of the failed task after correcting any stage output data errors.

Correct Answer: D
Section:
Explanation:
Using Cloud Composer to design the processing pipeline as a Directed Acyclic Graph (DAG) is the most suitable approach because:
Fault tolerance: Cloud Composer (based on Apache Airflow) allows for handling failures at specific stages. You can clear the state of a failed task and rerun it without reprocessing the entire pipeline.
Stage-based processing: DAGs are ideal for workflows with interdependent stages where the output of one stage serves as input to the next.
Efficiency: This approach minimizes downtime and ensures that only failed stages are rerun, leading to faster final output generation.

QUESTION 7
Another team in your organization is requesting access to a BigQuery dataset. You need to share the dataset with the team while minimizing the risk of unauthorized copying of data. You also want to create a reusable framework in case you need to share this data with other teams in the future. What should you do?

A. Create authorized views in the team's Google Cloud project that is only accessible by the team.
B. Create a private exchange using Analytics Hub with data egress restriction, and grant access to the team members.
C. Enable domain restricted sharing on the project. Grant the team members the BigQuery Data Viewer IAM role on the dataset.
D. Export the dataset to a Cloud Storage bucket in the team's Google Cloud project that is only accessible by the team.

Correct Answer: B
Section:
Explanation:
Using Analytics Hub to create a private exchange with data egress restrictions ensures controlled sharing of the dataset while minimizing the risk of unauthorized copying. This approach allows you to provide secure, managed access to the dataset without giving direct access to the raw data. The egress restriction ensures that data cannot be exported or copied outside the designated boundaries. Additionally, this solution provides a reusable framework that simplifies future data sharing with other teams or projects while maintaining strict data governance.

QUESTION 8
Your company has developed a website that allows users to upload and share video files. These files are most frequently accessed and shared when they are initially uploaded. Over time, the files are accessed and shared less frequently, although some old video files may remain very popular.
You need to design a storage system that is simple and cost-effective. What should you do?

A. Create a single-region bucket with Autoclass enabled.
B. Create a single-region bucket. Configure a Cloud Scheduler job that runs every 24 hours and changes the storage class based on upload date.
C. Create a single-region bucket with custom Object Lifecycle Management policies based on upload date.
D. Create a single-region bucket with Archive as the default storage class.

Correct Answer: C
Section:
Explanation:
Creating a single-region bucket with custom Object Lifecycle Management policies based on upload date is the most appropriate solution. This approach allows you to automatically transition objects to less expensive storage classes as their access frequency decreases over time. For example, frequently accessed files can remain in the Standard storage class initially, then transition to Nearline, Coldline, or Archive storage as their popularity wanes. This strategy ensures a cost-effective and efficient storage system while maintaining simplicity by automating the lifecycle management of video files.

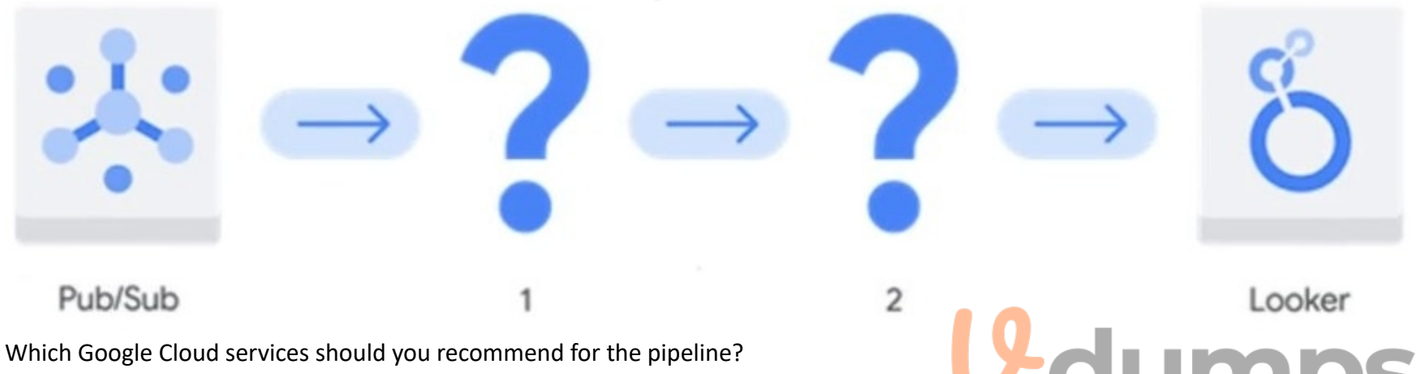
QUESTION 9
You recently inherited a task for managing Dataflow streaming pipelines in your organization and noticed that proper access had not been provisioned to you. You need to request a Google-provided IAM role so you can restart the pipelines. You need to follow the principle of least privilege. What should you do?

A. Request the Dataflow Developer role.
B. Request the Dataflow Viewer role.
C. Request the Dataflow Worker role.

D. Request the Dataflow Admin role.

Correct Answer: A
Section:
Explanation:
The Dataflow Developer role provides the necessary permissions to manage Dataflow streaming pipelines, including the ability to restart pipelines. This role adheres to the principle of least privilege, as it grants only the permissions required to manage and operate Dataflow jobs without unnecessary administrative access. Other roles, such as Dataflow Admin, would grant broader permissions, which are not needed in this scenario.

QUESTION 10
You need to create a new data pipeline. You want a serverless solution that meets the following requirements:
* Data is streamed from Pub/Sub and is processed in real-time.
* Data is transformed before being stored.
* Data is stored in a location that will allow it to be analyzed with SQL using Looker.



Which Google Cloud services should you recommend for the pipeline?

A. 1. Dataproc Serverless 2. Bigtable
B. 1. Cloud Composer 2. Cloud SQL for MySQL
C. 1. BigQuery 2. Analytics Hub
D. 1. Dataflow 2. BigQuery

Correct Answer: D
Section:
Explanation:
To build a serverless data pipeline that processes data in real-time from Pub/Sub, transforms it, and stores it for SQL-based analysis using Looker, the best solution is to use Dataflow and BigQuery. Dataflow is a fully managed service for real-time data processing and transformation, while BigQuery is a serverless data warehouse that supports SQL-based querying and integrates seamlessly with Looker for data analysis and visualization. This combination meets the requirements for real-time streaming, transformation, and efficient storage for analytical queries.

QUESTION 11
Your team wants to create a monthly report to analyze inventory data that is updated daily. You need to aggregate the inventory counts by using only the most recent month of data, and save the results to be used in a Looker Studio dashboard. What should you do?

A. Create a materialized view in BigQuery that uses the SUM( ) function and the DATE_SUB( ) function.
B. Create a saved query in the BigQuery console that uses the SUM( ) function and the DATE_SUB( ) function. Re-run the saved query every month, and save the results to a BigQuery table.
C. Create a BigQuery table that uses the SUM( ) function and the _PARTITIONDATE filter.
D. Create a BigQuery table that uses the SUM( ) function and the DATE_DIFF( ) function.

Correct Answer: A

Section:
Explanation:
Creating a materialized view in BigQuery with the SUM() function and the DATE_SUB() function is the best approach. Materialized views allow you to pre-aggregate and cache query results, making them efficient for repeated access, such as monthly reporting. By using the DATE_SUB() function, you can filter the inventory data to include only the most recent month. This approach ensures that the aggregation is up-to-date with minimal latency and provides efficient integration with Looker Studio for dashboarding.

QUESTION 12
You have a BigQuery dataset containing sales data. This data is actively queried for the first 6 months. After that, the data is not queried but needs to be retained for 3 years for compliance reasons. You need to implement a data management strategy that meets access and compliance requirements, while keeping cost and administrative overhead to a minimum. What should you do?

A. Use BigQuery long-term storage for the entire dataset. Set up a Cloud Run function to delete the data from BigQuery after 3 years.
B. Partition a BigQuery table by month. After 6 months, export the data to Coldline storage. Implement a lifecycle policy to delete the data from Cloud Storage after 3 years.
C. Set up a scheduled query to export the data to Cloud Storage after 6 months. Write a stored procedure to delete the data from BigQuery after 3 years.
D. Store all data in a single BigQuery table without partitioning or lifecycle policies.

Correct Answer: B
Section:
Explanation:
Partitioning the BigQuery table by month allows efficient querying of recent data for the first 6 months, reducing query costs. After 6 months, exporting the data to Coldline storage minimizes storage costs for data that is rarely accessed but needs to be retained for compliance. Implementing a lifecycle policy in Cloud Storage automates the deletion of the data after 3 years, ensuring compliance while reducing administrative overhead. This approach balances cost efficiency and compliance requirements effectively.

QUESTION 13
You have created a LookML model and dashboard that shows daily sales metrics for five regional managers to use. You want to ensure that the regional managers can only see sales metrics specific to their region. You need an easy-to-implement solution. What should you do?

A. Create a sales_region user attribute, and assign each manager's region as the value of their user attribute. Add an access_filter Explore filter on the region_name dimension by using the sales_region user attribute.
B. Create five different Explores with the sql_always_filter Explore filter applied on the region_name dimension. Set each region_name value to the corresponding region for each manager.
C. Create separate Looker dashboards for each regional manager. Set the default dashboard filter to the corresponding region for each manager.
D. Create separate Looker instances for each regional manager. Copy the LookML model and dashboard to each instance. Provision viewer access to the corresponding manager.

Correct Answer: A
Section:
Explanation:
Using a sales_region user attribute is the best solution because it allows you to dynamically filter data based on each manager's assigned region. By adding an access_filter Explore filter on the region_name dimension that references the sales_region user attribute, each manager sees only the sales metrics specific to their region. This approach is easy to implement, scalable, and avoids duplicating dashboards or Explores, making it both efficient and maintainable.

QUESTION 14
You need to design a data pipeline that ingests data from CSV, Avro, and Parquet files into Cloud Storage. The data includes raw user input. You need to remove all malicious SQL injections before storing the data in BigQuery. Which data manipulation methodology should you choose?

A. EL
B. ELT
C. ETL
D. ETLT

Correct Answer: C

Section:
Explanation:
The ETL (Extract, Transform, Load) methodology is the best approach for this scenario because it allows you to extract data from the files, transform it by applying the necessary data cleansing (including removing malicious SQL injections), and then load the sanitized data into BigQuery. By transforming the data before loading it into BigQuery, you ensure that only clean and safe data is stored, which is critical for security and data quality.

QUESTION 15
You are working with a large dataset of customer reviews stored in Cloud Storage. The dataset contains several inconsistencies, such as missing values, incorrect data types, and duplicate entries. You need to clean the data to ensure that it is accurate and consistent before using it for analysis. What should you do?

A. Use the PythonOperator in Cloud Composer to clean the data and load it into BigQuery. Use SQL for analysis.
B. Use BigQuery to batch load the data into BigQuery. Use SQL for cleaning and analysis.
C. Use Storage Transfer Service to move the data to a different Cloud Storage bucket. Use event triggers to invoke Cloud Run functions to load the data into BigQuery. Use SQL for analysis.
D. Use Cloud Run functions to clean the data and load it into BigQuery. Use SQL for analysis.

Correct Answer: B
Section:
Explanation:
Using BigQuery to batch load the data and perform cleaning and analysis with SQL is the best approach for this scenario. BigQuery provides powerful SQL capabilities to handle missing values, enforce correct data types, and remove duplicates efficiently. This method simplifies the pipeline by leveraging BigQuery's built-in processing power for both cleaning and analysis, reducing the need for additional tools or services and minimizing complexity.

QUESTION 16
Your retail organization stores sensitive application usage data in Cloud Storage. You need to encrypt the data without the operational overhead of managing encryption keys. What should you do?

A. Use Google-managed encryption keys (GMEK).
B. Use customer-managed encryption keys (CMEK).
C. Use customer-supplied encryption keys (CSEK).
D. Use customer-supplied encryption keys (CSEK) for the sensitive data and customer-managed encryption keys (CMEK) for the less sensitive data.

Correct Answer: A
Section:
Explanation:
Using Google-managed encryption keys (GMEK) is the best choice when you want to encrypt sensitive data in Cloud Storage without the operational overhead of managing encryption keys. GMEK is the default encryption mechanism in Google Cloud, and it ensures that data is automatically encrypted at rest with no additional setup or maintenance required. It provides strong security while eliminating the need for manual key management.

QUESTION 17
You work for a financial organization that stores transaction data in BigQuery. Your organization has a regulatory requirement to retain data for a minimum of seven years for auditing purposes. You need to ensure that the data is retained for seven years using an efficient and cost-optimized approach. What should you do?

A. Create a partition by transaction date, and set the partition expiration policy to seven years.
B. Set the table-level retention policy in BigQuery to seven years.
C. Set the dataset-level retention policy in BigQuery to seven years.
D. Export the BigQuery tables to Cloud Storage daily, and enforce a lifecycle management policy that has a seven-year retention rule.

Correct Answer: B
Section:
Explanation:
Setting a table-level retention policy in BigQuery to seven years is the most efficient and cost-optimized solution to meet the regulatory requirement. A table-level retention policy ensures that the data cannot be deleted or overwritten before the specified retention period expires, providing compliance with auditing requirements while keeping the data within BigQuery for easy access and analysis. This approach avoids the complexity and

additional costs of exporting data to Cloud Storage.

QUESTION 18
Your organization has a petabyte of application logs stored as Parquet files in Cloud Storage. You need to quickly perform a one-time SQL-based analysis of the files and join them to data that already resides in BigQuery. What should you do?

A. Create a Dataproc cluster, and write a PySpark job to join the data from BigQuery to the files in Cloud Storage.
B. Launch a Cloud Data Fusion environment, use plugins to connect to BigQuery and Cloud Storage, and use the SQL join operation to analyze the data.
C. Create external tables over the files in Cloud Storage, and perform SQL joins to tables in BigQuery to analyze the data.
D. Use the bq load command to load the Parquet files into BigQuery, and perform SQL joins to analyze the data.

Correct Answer: C
Section:
Explanation:
Creating external tables over the Parquet files in Cloud Storage allows you to perform SQL-based analysis and joins with data already in BigQuery without needing to load the files into BigQuery. This approach is efficient for a one-time analysis as it avoids the time and cost associated with loading large volumes of data into BigQuery. External tables provide seamless integration with Cloud Storage, enabling quick and cost-effective analysis of data stored in Parquet format.

QUESTION 19
Your organization uses Dataflow pipelines to process real-time financial transactions. You discover that one of your Dataflow jobs has failed. You need to troubleshoot the issue as quickly as possible. What should you do?

A. Set up a Cloud Monitoring dashboard to track key Dataflow metrics, such as data throughput, error rates, and resource utilization.
B. Create a custom script to periodically poll the Dataflow API for job status updates, and send email alerts if any errors are identified.
C. Navigate to the Dataflow Jobs page in the Google Cloud console. Use the job logs and worker logs to identify the error.
D. Use the gcloud CLI tool to retrieve job metrics and logs, and analyze them for errors and performance bottlenecks.

Correct Answer: C
Section:
Explanation:
To troubleshoot a failed Dataflow job as quickly as possible, you should navigate to the Dataflow Jobs page in the Google Cloud console. The console provides access to detailed job logs and worker logs, which can help you identify the cause of the failure. The graphical interface also allows you to visualize pipeline stages, monitor performance metrics, and pinpoint where the error occurred, making it the most efficient way to diagnose and resolve the issue promptly.

QUESTION 20
Your company uses Looker to generate and share reports with various stakeholders. You have a complex dashboard with several visualizations that needs to be delivered to specific stakeholders on a recurring basis, with customized filters applied for each recipient. You need an efficient and scalable solution to automate the delivery of this customized dashboard. You want to follow the Google-recommended approach. What should you do?

A. Create a separate LookML model for each stakeholder with predefined filters, and schedule the dashboards using the Looker Scheduler.
B. Create a script using the Looker Python SDK, and configure user attribute filter values. Generate a new scheduled plan for each stakeholder.
C. Embed the Looker dashboard in a custom web application, and use the application's scheduling features to send the report with personalized filters.
D. Use the Looker Scheduler with a user attribute filter on the dashboard, and send the dashboard with personalized filters to each stakeholder based on their attributes.

Correct Answer: D
Section:
Explanation:
Using the Looker Scheduler with user attribute filters is the Google-recommended approach to efficiently automate the delivery of a customized dashboard. User attribute filters allow you to dynamically customize the dashboard's content based on the recipient's attributes, ensuring each stakeholder sees data relevant to them. This approach is scalable, does not require creating separate models or custom scripts, and leverages Looker's built-in functionality to automate recurring deliveries effectively.

QUESTION 21
You are predicting customer churn for a subscription-based service. You have a 50 PB historical customer dataset in BigQuery that includes demographics, subscription information, and engagement metrics. You want to build a churn prediction model with minimal overhead. You want to follow the Google-recommended approach. What should you do?

A. Export the data from BigQuery to a local machine. Use scikit-learn in a Jupyter notebook to build the churn prediction model.
B. Use Dataproc to create a Spark cluster. Use the Spark MLlib within the cluster to build the churn prediction model.
C. Create a Looker dashboard that is connected to BigQuery. Use LookML to predict churn.
D. Use the BigQuery Python client library in a Jupyter notebook to query and preprocess the data in BigQuery. Use the CREATE MODEL statement in BigQueryML to train the churn prediction model.

Correct Answer: D
Section:
Explanation:
Using the BigQuery Python client library to query and preprocess data directly in BigQuery and then leveraging BigQueryML to train the churn prediction model is the Google-recommended approach for this scenario. BigQueryML allows you to build machine learning models directly within BigQuery using SQL, eliminating the need to export data or manage additional infrastructure. This minimizes overhead, scales effectively for a dataset as large as 50 PB, and simplifies the end-to-end process of building and training the churn prediction model.

QUESTION 22
You are a data analyst at your organization. You have been given a BigQuery dataset that includes customer information. The dataset contains inconsistencies and errors, such as missing values, duplicates, and formatting issues. You need to effectively and quickly clean the data. What should you do?

A. Develop a Dataflow pipeline to read the data from BigQuery, perform data quality rules and transformations, and write the cleaned data back to BigQuery.
B. Use Cloud Data Fusion to create a data pipeline to read the data from BigQuery, perform data quality transformations, and write the clean data back to BigQuery.
C. Export the data from BigQuery to CSV files. Resolve the errors using a spreadsheet editor, and re-import the cleaned data into BigQuery.
D. Use BigQuery's built-in functions to perform data quality transformations.

Correct Answer: D
Section:
Explanation:
Using BigQuery's built-in functions is the most effective and efficient way to clean the dataset directly within BigQuery. BigQuery provides powerful SQL capabilities to handle missing values, remove duplicates, and resolve formatting issues without needing to export data or create complex pipelines. This approach minimizes overhead and leverages the scalability of BigQuery for large datasets, making it an ideal solution for quickly addressing data quality issues.

QUESTION 23
Your organization has several datasets in their data warehouse in BigQuery. Several analyst teams in different departments use the datasets to run queries. Your organization is concerned about the variability of their monthly BigQuery costs. You need to identify a solution that creates a fixed budget for costs associated with the queries run by each department. What should you do?

A. Create a custom quota for each analyst in BigQuery.
B. Create a single reservation by using BigQuery editions. Assign all analysts to the reservation.
C. Assign each analyst to a separate project associated with their department. Create a single reservation by using BigQuery editions. Assign all projects to the reservation.
D. Assign each analyst to a separate project associated with their department. Create a single reservation for each department by using BigQuery editions. Create assignments for each project in the appropriate reservation.

Correct Answer: D
Section:
Explanation:
Assigning each analyst to a separate project associated with their department and creating a single reservation for each department using BigQuery editions allows for precise cost management. By assigning each project to its department's reservation, you can allocate fixed compute resources and budgets for each department, ensuring that their query costs are predictable and controlled. This approach aligns with your organization's goal of creating a fixed budget for query costs while maintaining departmental separation and accountability.

QUESTION 24
You manage a web application that stores data in a Cloud SQL database. You need to improve the read performance of the application by offloading read traffic from the primary database instance. You want to implement a solution that minimizes effort and cost. What should you do?

A. Use Cloud CDN to cache frequently accessed data.
B. Store frequently accessed data in a Memorystore instance.
C. Migrate the database to a larger Cloud SQL instance.
D. Enable automatic backups, and create a read replica of the Cloud SQL instance.

Correct Answer: D
Section:
Explanation:
Enabling automatic backups and creating a read replica of the Cloud SQL instance is the best solution to improve read performance. Read replicas allow you to offload read traffic from the primary database instance, reducing its load and improving overall performance. This approach is cost-effective and easy to implement within Cloud SQL. It ensures that the primary instance focuses on write operations while replicas handle read queries, providing a seamless performance boost with minimal effort.

QUESTION 25
Your organization plans to move their on-premises environment to Google Cloud. Your organization's network bandwidth is less than 1 Gbps. You need to move over 500 of data to Cloud Storage securely, and only have a few days to move the data. What should you do?

A. Request multiple Transfer Appliances, copy the data to the appliances, and ship the appliances back to Google Cloud to upload the data to Cloud Storage.
B. Connect to Google Cloud using VPN. Use Storage Transfer Service to move the data to Cloud Storage.
C. Connect to Google Cloud using VPN. Use the gcloud storage command to move the data to Cloud Storage.
D. Connect to Google Cloud using Dedicated Interconnect. Use the gcloud storage command to move the data to Cloud Storage.

Correct Answer: A
Section:
Explanation:
Using Transfer Appliances is the best solution for securely and efficiently moving over 500 TB of data to Cloud Storage within a limited timeframe, especially with network bandwidth below 1 Gbps. Transfer Appliances are physical devices provided by Google Cloud to securely transfer large amounts of data. After copying the data to the appliances, they are shipped back to Google, where the data is uploaded to Cloud Storage. This approach bypasses bandwidth limitations and ensures the data is migrated quickly and securely.

QUESTION 26
Your team is building several data pipelines that contain a collection of complex tasks and dependencies that you want to execute on a schedule, in a specific order. The tasks and dependencies consist of files in Cloud Storage, Apache Spark jobs, and data in BigQuery. You need to design a system that can schedule and automate these data processing tasks using a fully managed approach. What should you do?

A. Use Cloud Scheduler to schedule the jobs to run.
B. Use Cloud Tasks to schedule and run the jobs asynchronously.
C. Create directed acyclic graphs (DAGs) in Cloud Composer. Use the appropriate operators to connect to Cloud Storage, Spark, and BigQuery.
D. Create directed acyclic graphs (DAGs) in Apache Airflow deployed on Google Kubernetes Engine. Use the appropriate operators to connect to Cloud Storage, Spark, and BigQuery.

Correct Answer: C
Section:
Explanation:
Using Cloud Composer to create Directed Acyclic Graphs (DAGs) is the best solution because it is a fully managed, scalable workflow orchestration service based on Apache Airflow. Cloud Composer allows you to define complex task dependencies and schedules while integrating seamlessly with Google Cloud services such as Cloud Storage, BigQuery, and Dataproc for Apache Spark jobs. This approach minimizes operational overhead, supports scheduling and automation, and provides an efficient and fully managed way to orchestrate your data pipelines.

QUESTION 27
You are responsible for managing Cloud Storage buckets for a research company. Your company has well-defined data tiering and retention rules. You need to optimize storage costs while achieving your data retention needs. What should you do?

A. Configure the buckets to use the Archive storage class.
B. Configure a lifecycle management policy on each bucket to downgrade the storage class and remove objects based on age.
C. Configure the buckets to use the Standard storage class and enable Object Versioning.
D. Configure the buckets to use the Autoclass feature.

Correct Answer: B
Section:
Explanation:
Configuring a lifecycle management policy on each Cloud Storage bucket allows you to automatically transition objects to lower-cost storage classes (such as Nearline, Coldline, or Archive) based on their age or other criteria. Additionally, the policy can automate the removal of objects once they are no longer needed, ensuring compliance with retention rules and optimizing storage costs. This approach aligns well with well-defined data tiering and retention needs, providing cost efficiency and automation.

QUESTION 28
You are using your own data to demonstrate the capabilities of BigQuery to your organization's leadership team. You need to perform a one- time load of the files stored on your local machine into BigQuery using as little effort as possible. What should you do?

A. Write and execute a Python script using the BigQuery Storage Write API library.
B. Create a Dataproc cluster, copy the files to Cloud Storage, and write an Apache Spark job using the spark-bigquery-connector.
C. Execute the bq load command on your local machine.
D. Create a Dataflow job using the Apache Beam FileIO and BigQueryIO connectors with a local runner.

Correct Answer: C
Section:
Explanation:
Using the bq load command is the simplest and most efficient way to perform a one-time load of files from your local machine into BigQuery. This command-line tool is easy to use, requires minimal setup, and supports direct uploads from local files to BigQuery tables. It meets the requirement for minimal effort while allowing you to quickly demonstrate BigQuery's capabilities to your organization's leadership team.