Vendor: Google

Exam Code: Associate-Data-Practitioner

Exam Name: Google Cloud Associate Data Practitioner

Version: DEMO

QUESTION 1
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.

Answer: A
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 2
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

Answer: C
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 3
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.

Answer: B
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 4
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.

Answer: A
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.

Google Cloud encrypts all data at rest by default, and the simplest way to avoid key management overhead is to use Google-managed encryption keys (GMEK).

Option A: GMEK is fully managed by Google, requiring no user intervention, and meets the requirement of no operational overhead while ensuring encryption.

Option B: CMEK requires managing keys in Cloud KMS, adding operational overhead.

Option C: CSEK requires users to supply and manage keys externally, increasing complexity significantly.

Option D: Combining CSEK and CMEK adds unnecessary complexity and overhead for a uniform sensitive dataset.

QUESTION 5
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.

Answer: B
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 6
You need to create a weekly aggregated sales report based on a large volume of data. You want to use Python to design an efficient process for generating this report. What should you do?

A. Create a Cloud Run function that uses NumPy. Use Cloud Scheduler to schedule the function to run once a week.
B. Create a Colab Enterprise notebook and use the bigframes.pandas library. Schedule the notebook to execute once a week.
C. Create a Cloud Data Fusion and Wrangler flow. Schedule the flow to run once a week.
D. Create a Dataflow directed acyclic graph (DAG) coded in Python. Use Cloud Scheduler to schedule the code to run once a week.

Answer: D
Explanation:
Using Dataflow with a Python-coded Directed Acyclic Graph (DAG) is the most efficient solution for generating a weekly aggregated sales report based on a large volume of data. Dataflow is optimized for large-scale data processing and can handle aggregation efficiently. Python allows you to customize the pipeline logic, and Cloud Scheduler enables you to automate the process to run weekly. This approach ensures scalability, efficiency, and the ability to process large datasets in a cost-effective manner.

QUESTION 7
Your organization has decided to move their on-premises Apache Spark-based workload to Google Cloud. You want to be able to manage the code without needing to provision and manage your own cluster. What should you do?

A. Migrate the Spark jobs to Dataproc Serverless.
B. Configure a Google Kubernetes Engine cluster with Spark operators, and deploy the Spark jobs.
C. Migrate the Spark jobs to Dataproc on Google Kubernetes Engine.
D. Migrate the Spark jobs to Dataproc on Compute Engine.

Answer: A
Explanation:
Migrating the Spark jobs to Dataproc Serverless is the best approach because it allows you to run Spark workloads without the need to provision or manage clusters. Dataproc Serverless automatically scales resources based on workload requirements, simplifying operations and reducing administrative overhead. This solution is ideal for organizations that want to focus on managing their Spark code without worrying about the underlying infrastructure. It is cost-effective and fully managed, aligning well with the goal of minimizing cluster management.

QUESTION 8
You are developing a data ingestion pipeline to load small CSV files into BigQuery from Cloud Storage. You want to load these files upon arrival to minimize data latency. You want to accomplish this with minimal cost and maintenance. What should you do?

A. Use the bq command-line tool within a Cloud Shell instance to load the data into BigQuery.
B. Create a Cloud Composer pipeline to load new files from Cloud Storage to BigQuery and schedule it to run every 10 minutes.
C. Create a Cloud Run function to load the data into BigQuery that is triggered when data arrives in Cloud Storage.
D. Create a Dataproc cluster to pull CSV files from Cloud Storage, process them using Spark, and write the results to BigQuery.

Answer: C
Explanation:
Using a Cloud Run function triggered by Cloud Storage to load the data into BigQuery is the best solution because it minimizes both cost and maintenance while providing low-latency data ingestion. Cloud Run is a serverless platform that automatically scales based on the workload, ensuring efficient use of resources without requiring a dedicated instance or cluster. It integrates seamlessly with Cloud Storage event notifications, enabling real-time processing of incoming files and loading them into BigQuery. This approach is cost-effective, scalable, and easy to manage.

The goal is to load small CSV files into BigQuery upon arrival (event-driven) with minimal latency, cost, and maintenance. Google Cloud provides serverless, event-driven options that align with this requirement. Let's evaluate each option in detail:

Option A: Cloud Composer (managed Apache Airflow) can schedule a pipeline to check Cloud Storage every 10 minutes, but this polling approach introduces latency (up to 10 minutes) and incurs costs for running Composer even when no files arrive. Maintenance includes managing DAGs and the Composer environment, which adds overhead. This is better suited for scheduled batch jobs, not event-driven ingestion.

Option B: A Cloud Run function triggered by a Cloud Storage event (via Eventarc or Pub/Sub) loads files into BigQuery as soon as they arrive, minimizing latency. Cloud Run is serverless, scales to zero when idle (low cost), and requires minimal maintenance (deploy and forget). Using the BigQuery API in the function (e.g., Python client library) handles small CSV loads efficiently. This aligns with Google's serverless, event-driven best practices.

Option C: Dataproc with Spark is designed for large-scale, distributed processing, not small CSV ingestion. It requires cluster management, incurs higher costs (even with ephemeral clusters), and adds unnecessary complexity for a simple load task.

Option D: The bq command-line tool in Cloud Shell is manual and not automated, failing the "upon arrival" requirement. It's a one-off tool, not a pipeline solution, and Cloud Shell isn't designed for persistent automation.

Why B is Best: Cloud Run leverages Cloud Storage's object creation events, ensuring near-zero latency between file arrival and BigQuery ingestion. It's serverless, meaning no infrastructure to manage, and costs scale with usage (free when idle). For small CSVs, the BigQuery load job is lightweight, avoiding processing overhead.

QUESTION 9

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.

Answer: C
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 10
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.

Answer: C
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 11
Your company uses Looker to visualize and analyze sales data. You need to create a dashboard that displays sales metrics, such as sales by region, product category, and time period. Each metric relies on its own set of attributes distributed across several tables. You need to provide users the ability to filter the data by specific sales representatives and view individual transactions. You want to follow the Google-recommended approach. What should you do?

A. Create multiple Explores, each focusing on each sales metric. Link the Explores together in a

dashboard using drill-down functionality.

B. Use BigQuery to create multiple materialized views, each focusing on a specific sales metric. Build the dashboard using these views.
C. Create a single Explore with all sales metrics. Build the dashboard using this Explore.
D. Use Looker's custom visualization capabilities to create a single visualization that displays all the sales metrics with filtering and drill-down functionality.

Answer: C
Explanation:
Creating a single Explore with all the sales metrics is the Google-recommended approach. This Explore should be designed to include all relevant attributes and dimensions, enabling users to analyze sales data by region, product category, time period, and other filters like sales representatives. With a well-structured Explore, you can efficiently build a dashboard that supports filtering and drill-down functionality. This approach simplifies maintenance, provides a consistent data model, and ensures users have the flexibility to interact with and analyze the data seamlessly within a unified framework.

Looker's recommended approach for dashboards is a single, unified Explore for scalability and usability, supporting filters and drill-downs.

Option A: Materialized views in BigQuery optimize queries but bypass Looker's modeling layer, reducing flexibility.

Option B: Custom visualizations are for specific rendering, not multi-metric dashboards with filtering/drill-down.

Option C: Multiple Explores fragment the data model, complicating dashboard cohesion and maintenance.

Option D: A single Explore joins tables in LookML, enabling all metrics, filters (e.g., sales rep), and drill-downs to transactions, per Looker's best practices.

QUESTION 12
Your company's ecommerce website collects product reviews from customers. The reviews are loaded as CSV files daily to a Cloud Storage bucket. The reviews are in multiple languages and need to be translated to Spanish. You need to configure a pipeline that is serverless, efficient, and requires minimal maintenance. What should you do?

A. Load the data into BigQuery using Dataproc. Use Apache Spark to translate the reviews by invoking the Cloud Translation API. Set BigQuery as the sink.U
B. Use a Dataflow templates pipeline to translate the reviews using the Cloud Translation API. Set BigQuery as the sink.
C. Load the data into BigQuery using a Cloud Run function. Use the BigQuery ML create model statement to train a translation model. Use the model to translate the product reviews within BigQuery.
D. Load the data into BigQuery using a Cloud Run function. Create a BigQuery remote function that invokes the Cloud Translation API. Use a scheduled query to translate new reviews.

Answer: D
Explanation:
Loading the data into BigQuery using a Cloud Run function and creating a BigQuery remote function that invokes the Cloud Translation API is a serverless and efficient approach. With this setup, you can use a scheduled query in BigQuery to invoke the remote function and translate

new product reviews on a regular basis. This solution requires minimal maintenance, as BigQuery handles storage and querying, and the Cloud Translation API provides accurate translations without the need for custom ML model development.

QUESTION 13
You have a Dataproc cluster that performs batch processing on data stored in Cloud Storage. You need to schedule a daily Spark job to generate a report that will be emailed to stakeholders. You need a fully-managed solution that is easy to implement and minimizes complexity. What should you do?

A. Use Cloud Composer to orchestrate the Spark job and email the report.
B. Use Dataproc workflow templates to define and schedule the Spark job, and to email the report.
C. Use Cloud Run functions to trigger the Spark job and email the report.
D. Use Cloud Scheduler to trigger the Spark job. and use Cloud Run functions to email the report.

Answer: B
Explanation:
Using Dataproc workflow templates is a fully-managed and straightforward solution for defining and scheduling your Spark job on a Dataproc cluster. Workflow templates allow you to automate the execution of Spark jobs with predefined steps, including data processing and report generation. You can integrate email notifications by adding a step to the workflow that sends the report using tools like a Cloud Function or external email service. This approach minimizes complexity while leveraging Dataproc's managed capabilities for batch processing.

QUESTION 14
Your organization has highly sensitive data that gets updated once a day and is stored across multiple datasets in BigQuery. You need to provide a new data analyst access to query specific data in BigQuery while preventing access to sensitive data. What should you do?

A. Grant the data analyst the BigQuery Job User IAM role in the Google Cloud project.
B. Create a materialized view with the limited data in a new dataset. Grant the data analyst BigQuery Data Viewer IAM role in the dataset and the BigQuery Job User IAM role in the Google Cloud project.
C. Create a new Google Cloud project, and copy the limited data into a BigQuery table. Grant the data analyst the BigQuery Data Owner IAM role in the new Google Cloud project.
D. Grant the data analyst the BigQuery Data Viewer IAM role in the Google Cloud project.

Answer: B
Explanation:
Creating a materialized view with the limited data in a new dataset and granting the data analyst the BigQuery Data Viewer role on the dataset and the BigQuery Job User role in the project ensures that the analyst can query only the non-sensitive data without access to sensitive datasets. Materialized views allow you to predefine what subset of data is visible, providing a secure and efficient way to control access while maintaining compliance with data governance policies. This approach follows the principle of least privilege while meeting the requirements.

Thank You for Trying Our Product

Lead2pass Certification Exam Features:

★ More than **99,900** Satisfied Customers Worldwide.

★ Average **99.9%** Success Rate.

★ **Free Update** to match latest and real exam scenarios.

★ **Instant Download** Access! No Setup required.

★ Questions & Answers are downloadable in **PDF** format and **VCE** test engine format.

★ Multi-Platform capabilities - **Windows, Laptop, Mac, Android, iPhone, iPod, iPad**.

★ **100%** Guaranteed Success or **100%** Money Back Guarantee.

★ **Fast**, helpful support **24x7**.

View list of all certification exams: http://www.lead2pass.com/all-products.html

10% Discount Coupon Code: **ASTR14**