Google's Professional Cloud Architect Actual Exam Questions

## Question #: 1

Q: Your retail company wants to predict customer churn using historical purchase data stored in BigQuery. The dataset includes customer demographics, purchase history, and a label indicating whether the customer churned or not. You want to build a machine learning model to identify customers at risk of churning. You need to create and train a logistic regression model for predicting customer churn, using the customer\_data table with the churned column as the target label. Which BigQuery ML query should you use?

* A.  
  -------------------------
* **B.  
  -------------------------**
* C.  
  -------------------------
* D.  
  -------------------------

Answer: B

## Question #: 2

Q: Your company has several retail locations. Your company tracks the total number of sales made at each location each day. You want to use SQL to calculate the weekly moving average of sales by location to identify trends for each store. Which query should you use?

* A.  
  -------------------------
* B.  
  -------------------------
* **C.  
  -------------------------**
* D.  
  -------------------------

Answer: C

## Question #: 3

Q: Your company is building a near real-time streaming pipeline to process JSON telemetry data from small appliances. You need to process messages arriving at a Pub/Sub topic, capitalize letters in the serial number field, and write results to BigQuery. You want to use a managed service and write a minimal amount of code for underlying transformations. What should you do?

* A. Use a Pub/Sub to BigQuery subscription, write results directly to BigQuery, and schedule a transformation query to run every five minutes.
* B. Use a Pub/Sub to Cloud Storage subscription, write a Cloud Run service that is triggered when objects arrive in the bucket, performs the transformations, and writes the results to BigQuery.
* **C. Use the “Pub/Sub to BigQuery” Dataflow template with a UDF, and write the results to BigQuery.**
* D. Use a Pub/Sub push subscription, write a Cloud Run service that accepts the messages, performs the transformations, and writes the results to BigQuery.

Answer: C

## Question #: 4

Q: You want to process and load a daily sales CSV file stored in Cloud Storage into BigQuery for downstream reporting. You need to quickly build a scalable data pipeline that transforms the data while providing insights into data quality issues. What should you do?

* **A. Create a batch pipeline in Cloud Data Fusion by using a Cloud Storage source and a BigQuery sink.**
* B. Load the CSV file as a table in BigQuery, and use scheduled queries to run SQL transformation scripts.
* C. Load the CSV file as a table in BigQuery. Create a batch pipeline in Cloud Data Fusion by using a BigQuery source and sink.
* D. Create a batch pipeline in Dataflow by using the Cloud Storage CSV file to BigQuery batch template.

Answer: A

## Question #: 5

Q: You manage a Cloud Storage bucket that stores temporary files created during data processing. These temporary files are only needed for seven days, after which they are no longer needed. To reduce storage costs and keep your bucket organized, you want to automatically delete these files once they are older than seven days. What should you do?

* A. Set up a Cloud Scheduler job that invokes a weekly Cloud Run function to delete files older than seven days.
* **B. Configure a Cloud Storage lifecycle rule that automatically deletes objects older than seven days.**
* C. Develop a batch process using Dataflow that runs weekly and deletes files based on their age.
* D. Create a Cloud Run function that runs daily and deletes files older than seven days.

Answer: B

## Question #: 6

Q: You work for a healthcare company that has a large on-premises data system containing patient records with personally identifiable information (PII) such as names, addresses, and medical diagnoses. You need a standardized managed solution that de-identifies PII across all your data feeds prior to ingestion to Google Cloud. What should you do?

* A. Use Cloud Run functions to create a serverless data cleaning pipeline. Store the cleaned data in BigQuery.
* **B. Use Cloud Data Fusion to transform the data. Store the cleaned data in BigQuery.**
* C. Load the data into BigQuery, and inspect the data by using SQL queries. Use Dataflow to transform the data and remove any errors.
* D. Use Apache Beam to read the data and perform the necessary cleaning and transformation operations. Store the cleaned data in BigQuery.

Answer: B

## Question #: 7

Q: 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.**

Answer: D

## Question #: 8

Q: 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.

Answer: A

## Question #: 9

Q: 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.**

Answer: D

## Question #: 10

Q: 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.

Answer: A

## Question #: 11

Q: 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.

Answer: A

## Question #: 12

Q: 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.

Answer: A

## Question #: 13

Q: 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**

Answer: D

## Question #: 14

Q: 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.

Answer: A

## Question #: 15

Q: 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.

Answer: B

## Question #: 16

Q: 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

## Question #: 17

Q: 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

## Question #: 18

Q: 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

## Question #: 19

Q: 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

## Question #: 20

Q: 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

## Question #: 21

Q: 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

## Question #: 22

Q: 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

## Question #: 23

Q: 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

## Question #: 24

Q: 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

## Question #: 25

Q: 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

## Question #: 26

Q: 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.

Answer: B

## Question #: 27

Q: 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.

Answer: C

## Question #: 28

Q: 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.

Answer: C

## Question #: 29

Q: 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.**

Answer: D

## Question #: 30

Q: 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.**

Answer: D

## Question #: 31

Q: 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.**

Answer: D

## Question #: 32

Q: 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.**

Answer: D

## Question #: 33

Q: 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.**

Answer: D

## Question #: 34

Q: 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 ТВ 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.

Answer: A

## Question #: 35

Q: 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.**

Answer: D

## Question #: 36

Q: 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.

Answer: B

## Question #: 37

Q: 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.

Answer: B

## Question #: 38

Q: Your organization sends IoT event data to a Pub/Sub topic. Subscriber applications read and perform transformations on the messages before storing them in the data warehouse. During particularly busy times when more data is being written to the topic, you notice that the subscriber applications are not acknowledging messages within the deadline. You need to modify your pipeline to handle these activity spikes and continue to process the messages. What should you do?

* A. Retry messages until they are acknowledged.
* **B. Implement flow control on the subscribers.**
* C. Forward unacknowledged messages to a dead-letter topic.
* D. Seek back to the last acknowledged message.

Answer: B

## Question #: 39

Q: You have millions of customer feedback records stored in BigQuery. You want to summarize the data by using the large language model (LLM) Gemini. You need to plan and execute this analysis using the most efficient approach. What should you do?

* A. Query the BigQuery table from within a Python notebook, use the Gemini API to summarize the data within the notebook, and store the summaries in BigQuery.
* B. Use a BigQuery ML model to pre-process the text data, export the results to Cloud Storage, and use the Gemini API to summarize the pre- processed data.
* **C. Create a BigQuery Cloud resource connection to a remote model in Vertex Al, and use Gemini to summarize the data.**
* D. Export the raw BigQuery data to a CSV file, upload it to Cloud Storage, and use the Gemini API to summarize the data.

Answer: C

## Question #: 40

Q: You are working on a data pipeline that will validate and clean incoming data before loading it into BigQuery for real-time analysis. You want to ensure that the data validation and cleaning is performed efficiently and can handle high volumes of data. What should you do?

* A. Write custom scripts in Python to validate and clean the data outside of Google Cloud. Load the cleaned data into BigQuery.
* B. Use Cloud Run functions to trigger data validation and cleaning routines when new data arrives in Cloud Storage.
* **C. Use Dataflow to create a streaming pipeline that includes validation and transformation steps.**
* D. Load the raw data into BigQuery using Cloud Storage as a staging area, and use SQL queries in BigQuery to validate and clean the data.

Answer: C

## Question #: 41

Q: Your organization needs to implement near real-time analytics for thousands of events arriving each second in Pub/Sub. The incoming messages require transformations. You need to configure a pipeline that processes, transforms, and loads the data into BigQuery while minimizing development time. What should you do?

* **A. Use a Google-provided Dataflow template to process the Pub/Sub messages, perform transformations, and write the results to BigQuery.**
* B. Create a Cloud Data Fusion instance and configure Pub/Sub as a source. Use Data Fusion to process the Pub/Sub messages, perform transformations, and write the results to BigQuery.
* C. Load the data from Pub/Sub into Cloud Storage using a Cloud Storage subscription. Create a Dataproc cluster, use PySpark to perform transformations in Cloud Storage, and write the results to BigQuery.
* D. Use Cloud Run functions to process the Pub/Sub messages, perform transformations, and write the results to BigQuery.

Answer: A

## Question #: 42

Q: Your organization needs to store historical customer order data. The data will only be accessed once a month for analysis and must be readily available within a few seconds when it is accessed. You need to choose a storage class that minimizes storage costs while ensuring that the data can be retrieved quickly. What should you do?

* **A. Store the data in Cloud Storage using Nearline storage.**
* B. Store the data in Cloud Storage using Coldline storage.
* C. Store the data in Cloud Storage using Standard storage.
* D. Store the data in Cloud Storage using Archive storage.

Answer: A

## Question #: 43

Q: You have a Dataflow pipeline that processes website traffic logs stored in Cloud Storage and writes the processed data to BigQuery. You noticed that the pipeline is failing intermittently. You need to troubleshoot the issue. What should you do?

* A. Use Cloud Logging to identify error groups in the pipeline's logs. Use Cloud Monitoring to create a dashboard that tracks the number of errors in each group.
* B. Use Cloud Logging to create a chart displaying the pipeline’s error logs. Use Metrics Explorer to validate the findings from the chart.
* **C. Use Cloud Logging to view error messages in the pipeline's logs. Use Cloud Monitoring to analyze the pipeline's metrics, such as CPU utilization and memory usage.**
* D. Use the Dataflow job monitoring interface to check the pipeline's status every hour. Use Cloud Profiler to analyze the pipeline’s metrics, such as CPU utilization and memory usage.

Answer: C

## Question #: 44

Q: Your organization’s business analysts require near real-time access to streaming data. However, they are reporting that their dashboard queries are loading slowly. After investigating BigQuery query performance, you discover the slow dashboard queries perform several joins and aggregations.  
You need to improve the dashboard loading time and ensure that the dashboard data is as up-to-date as possible. What should you do?

* A. Disable BigQuery query result caching.
* B. Modify the schema to use parameterized data types.
* C. Create a scheduled query to calculate and store intermediate results.
* **D. Create materialized views.**

Answer: D

## Question #: 45

Q: You need to create a data pipeline that streams event information from applications in multiple Google Cloud regions into BigQuery for near real-time analysis. The data requires transformation before loading. You want to create the pipeline using a visual interface. What should you do?

* **A. Push event information to a Pub/Sub topic. Create a Dataflow job using the Dataflow job builder.**
* B. Push event information to a Pub/Sub topic. Create a Cloud Run function to subscribe to the Pub/Sub topic, apply transformations, and insert the data into BigQuery.
* C. Push event information to a Pub/Sub topic. Create a BigQuery subscription in Pub/Sub.
* D. Push event information to Cloud Storage, and create an external table in BigQuery. Create a BigQuery scheduled job that executes once each day to apply transformations.

Answer: A

## Question #: 46

Q: You work for an online retail company. Your company collects customer purchase data in CSV files and pushes them to Cloud Storage every 10 minutes. The data needs to be transformed and loaded into BigQuery for analysis. The transformation involves cleaning the data, removing duplicates, and enriching it with product information from a separate table in BigQuery. You need to implement a low-overhead solution that initiates data processing as soon as the files are loaded into Cloud Storage. What should you do?

* A. Use Cloud Composer sensors to detect files loading in Cloud Storage. Create a Dataproc cluster, and use a Composer task to execute a job on the cluster to process and load the data into BigQuery.
* B. Schedule a direct acyclic graph (DAG) in Cloud Composer to run hourly to batch load the data from Cloud Storage to BigQuery, and process the data in BigQuery using SQL.
* **C. Use Dataflow to implement a streaming pipeline using an OBJECT\_FINALIZE notification from Pub/Sub to read the data from Cloud Storage, perform the transformations, and write the data to BigQuery.**
* D. Create a Cloud Data Fusion job to process and load the data from Cloud Storage into BigQuery. Create an OBJECT\_FINALI ZE notification in Pub/Sub, and trigger a Cloud Run function to start the Cloud Data Fusion job as soon as new files are loaded.

Answer: C

## Question #: 47

Q: You work for a home insurance company. You are frequently asked to create and save risk reports with charts for specific areas using a publicly available storm event dataset. You want to be able to quickly create and re-run risk reports when new data becomes available. What should you do?

* A. Export the storm event dataset as a CSV file. Import the file to Google Sheets, and use cell data in the worksheets to create charts.
* **B. Copy the storm event dataset into your BigQuery project. Use BigQuery Studio to query and visualize the data in Looker Studio.**
* C. Reference and query the storm event dataset using SQL in BigQuery Studio. Export the results to Google Sheets, and use cell data in the worksheets to create charts.
* D. Reference and query the storm event dataset using SQL in a Colab Enterprise notebook. Display the table results and document with Markdown, and use Matplotlib to create charts.

Answer: B

## Question #: 48

Q: Your company currently uses an on-premises network file system (NFS) and is migrating data to Google Cloud. You want to be able to control how much bandwidth is used by the data migration while capturing detailed reporting on the migration status. What should you do?

* A. Use a Transfer Appliance.
* B. Use Cloud Storage FUSE.
* **C. Use Storage Transfer Service.**
* D. Use gcloud storage commands.

Answer: C

## Question #: 49

Q: You are a Looker analyst. You need to add a new field to your Looker report that generates SQL that will run against your company's database. You do not have the Develop permission. What should you do?

* A. Create a new field in the LookML layer, refresh your report, and select your new field from the field picker.
* B. Create a calculated field using the Add a field option in Looker Studio, and add it to your report.
* C. Create a table calculation from the field picker in Looker, and add it to your report.
* **D. Create a custom field from the field picker in Looker, and add it to your report.**

Answer: D

## Question #: 50

Q: Your organization’s ecommerce website collects user activity logs using a Pub/Sub topic. Your organization’s leadership team wants a dashboard that contains aggregated user engagement metrics. You need to create a solution that transforms the user activity logs into aggregated metrics, while ensuring that the raw data can be easily queried. What should you do?

* **A. Create a Dataflow subscription to the Pub/Sub topic, and transform the activity logs. Load the transformed data into a BigQuery table for reporting.**
* B. Create an event-driven Cloud Run function to trigger a data transformation pipeline to run. Load the transformed activity logs into a BigQuery table for reporting.
* C. Create a Cloud Storage subscription to the Pub/Sub topic. Load the activity logs into a bucket using the Avro file format. Use Dataflow to transform the data, and load it into a BigQuery table for reporting.
* D. Create a BigQuery subscription to the Pub/Sub topic, and load the activity logs into the table. Create a materialized view in BigQuery using SQL to transform the data for reporting

Answer: A

## Question #: 51

Q: You are constructing a data pipeline to process sensitive customer data stored in a Cloud Storage bucket. You need to ensure that this data remains accessible, even in the event of a single-zone outage. What should you do?

* A. Set up a Cloud CDN in front of the bucket.
* B. Enable Object Versioning on the bucket.
* **C. Store the data in a multi-region bucket.**
* D. Store the data in Nearline storage.

Answer: C

## Question #: 52

Q: Your retail company collects customer data from various sources:  
Online transactions: Stored in a MySQL database  
Customer feedback: Stored as text files on a company server  
Social media activity: Streamed in real-time from social media platforms  
You are designing a data pipeline to extract this data. Which Google Cloud storage system(s) should you select for further analysis and ML model training?

* A. 1. Online transactions: Cloud Storage  
  2. Customer feedback: Cloud Storage  
  3. Social media activity: Cloud Storage
* **B. 1. Online transactions: BigQuery  
  2. Customer feedback: Cloud Storage  
  3. Social media activity: BigQuery**
* C. 1. Online transactions: Bigtable  
  2. Customer feedback: Cloud Storage  
  3. Social media activity: CloudSQL for MySQL
* D. 1. Online transactions: Cloud SQL for MySQL  
  2. Customer feedback: BigQuery  
  3. Social media activity: Cloud Storage

Answer: B

## Question #: 53

Q: Your company uses Looker as its primary business intelligence platform. You want to use LookML to visualize the profit margin for each of your company’s products in your Looker Explores and dashboards. You need to implement a solution quickly and efficiently. What should you do?

* A. Create a derived table that pre-calculates the profit margin for each product, and include it in the Looker model.
* **B. Define a new measure that calculates the profit margin by using the existing revenue and cost fields.**
* C. Create a new dimension that categorizes products based on their profit margin ranges (e.g., high, medium, low).
* D. Apply a filter to only show products with a positive profit margin.

Answer: B

## Question #: 54

Q: You are a data analyst working with sensitive customer data in BigQuery. You need to ensure that only authorized personnel within your organization can query this data, while following the principle of least privilege. What should you do?

* **A. Enable access control by using IAM roles.**
* B. Encrypt the data by using customer-managed encryption keys (CMEK).
* C. Update dataset privileges by using the SQL GRANT statement.
* D. Export the data to Cloud Storage, and use signed URLs to authorize access.

Answer: A

## Question #: 55

Q: Your organization stores highly personal data in BigQuery and needs to comply with strict data privacy regulations. You need to ensure that sensitive data values are rendered unreadable whenever an employee leaves the organization. What should you do?

* **A. Use AEAD functions and delete keys when employees leave the organization.**
* B. Use dynamic data masking and revoke viewer permissions when employees leave the organization.
* C. Use customer-managed encryption keys (CMEK) and delete keys when employees leave the organization.
* D. Use column-level access controls with policy tags and revoke viewer permissions when employees leave the organization.

Answer: A

## Question #: 56

Q: You used BigQuery ML to build a customer purchase propensity model six months ago. You want to compare the current serving data with the historical serving data to determine whether you need to retrain the model. What should you do?

* A. Compare the two different models.
* B. Evaluate the data skewness.
* **C. Evaluate data drift.**
* D. Compare the confusion matrix.

Answer: C

## Question #: 57

Q: 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

## Question #: 58

Q: 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.
* **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: B

## Question #: 59

Q: 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: A

## Question #: 60

Q: 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

## Question #: 61

Q: You are a database administrator managing sales transaction data by region stored in a BigQuery table. You need to ensure that each sales representative can only see the transactions in their region. What should you do?

* A. Add a policy tag in BigQuery.
* **B. Create a row-level access policy.**
* C. Create a data masking rule.
* D. Grant the appropriate IAM permissions on the dataset.

Answer: B

## Question #: 62

Q: Your company’s customer support audio files are stored in a Cloud Storage bucket. You plan to analyze the audio files’ metadata and file content within BigQuery to create inference by using BigQuery ML. You need to create a corresponding table in BigQuery that represents the bucket containing the audio files. What should you do?

* A. Create an external table.
* B. Create a temporary table.
* C. Create a native table.
* **D. Create an object table.**

Answer: D

## Question #: 63

Q: You work for a financial services company that handles highly sensitive data. Due to regulatory requirements, your company is required to have complete and manual control of data encryption. Which type of keys should you recommend to use for data storage?

* **A. Use customer-supplied encryption keys (CSEK).**
* B. Use a dedicated third-party key management system (KMS) chosen by the company.
* C. Use Google-managed encryption keys (GMEK).
* D. Use customer-managed encryption keys (CMEK).

Answer: A

## Question #: 64

Q: Your team needs to analyze large datasets stored in BigQuery to identify trends in user behavior. The analysis will involve complex statistical calculations, Python packages, and visualizations. You need to recommend a managed collaborative environment to develop and share the analysis. What should you recommend?

* **A. Create a Colab Enterprise notebook and connect the notebook to BigQuery. Share the notebook with your team. Analyze the data and generate visualizations in Colab Enterprise.**
* B. Create a statistical model by using BigQuery ML. Share the query with your team. Analyze the data and generate visualizations in Looker Studio.
* C. Create a Looker Studio dashboard and connect the dashboard to BigQuery. Share the dashboard with your team. Analyze the data and generate visualizations in Looker Studio.
* D. Connect Google Sheets to BigQuery by using Connected Sheets. Share the Google Sheet with your team. Analyze the data and generate visualizations in Gooqle Sheets.

Answer: A

## Question #: 65

Q: Your organization has several datasets in BigQuery. The datasets need to be shared with your external partners so that they can run SQL queries without needing to copy the data to their own projects. You have organized each partner’s data in its own BigQuery dataset. Each partner should be able to access only their data. You want to share the data while following Google-recommended practices. What should you do?

* **A. Use Analytics Hub to create a listing on a private data exchange for each partner dataset. Allow each partner to subscribe to their respective listings.**
* B. Create a Dataflow job that reads from each BigQuery dataset and pushes the data into a dedicated Pub/Sub topic for each partner. Grant each partner the pubsub. subscriber IAM role.
* C. Export the BigQuery data to a Cloud Storage bucket. Grant the partners the storage.objectUser IAM role on the bucket.
* D. Grant the partners the bigquery.user IAM role on the BigQuery project.

Answer: A

## Question #: 66

Q: Your organization has decided to migrate their existing enterprise data warehouse to BigQuery. The existing data pipeline tools already support connectors to BigQuery. You need to identify a data migration approach that optimizes migration speed. What should you do?

* A. Create a temporary file system to facilitate data transfer from the existing environment to Cloud Storage. Use Storage Transfer Service to migrate the data into BigQuery.
* B. Use the Cloud Data Fusion web interface to build data pipelines. Create a directed acyclic graph (DAG) that facilitates pipeline orchestration.
* **C. Use the existing data pipeline tool’s BigQuery connector to reconfigure the data mapping.**
* D. Use the BigQuery Data Transfer Service to recreate the data pipeline and migrate the data into BigQuery.

Answer: C

## Question #: 67

Q: Your organization uses scheduled queries to perform transformations on data stored in BigQuery. You discover that one of your scheduled queries has failed. You need to troubleshoot the issue as quickly as possible. What should you do?

* A. Navigate to the Logs Explorer page in Cloud Logging. Use filters to find the failed job, and analyze the error details.
* B. Set up a log sink using the gcloud CLI to export BigQuery audit logs to BigQuery. Query those logs to identify the error associated with the failed job ID.
* C. Request access from your admin to the BigQuery information\_schema. Query the jobs view with the failed job ID, and analyze error details.
* **D. Navigate to the Scheduled queries page in the Google Cloud console. Select the failed job, and analyze the error details.**

Answer: D

## Question #: 68

Q: Your team uses the Google Ads platform to visualize metrics. You want to export the data to BigQuery to get more granular insights. You need to execute a one-time transfer of historical data and automatically update data daily. You want a solution that is low-code, serverless, and requires minimal maintenance. What should you do?

* A. Export the historical data to BigQuery by using BigQuery Data Transfer Service. Use Cloud Composer for daily automation.
* B. Export the historical data to Cloud Storage by using Storage Transfer Service. Use Pub/Sub to trigger a Dataflow template that loads data for daily automation.
* C. Export the historical data as a CSV file. Import the file into BigQuery for analysis. Use Cloud Composer for daily automation.
* **D. Export the historical data to BigQuery by using BigQuery Data Transfer Service. Use BigQuery Data Transfer Service for daily automation.**

Answer: D

## Question #: 69

Q: Your organization has a BigQuery dataset that contains sensitive employee information such as salaries and performance reviews. The payroll specialist in the HR department needs to have continuous access to aggregated performance data, but they do not need continuous access to other sensitive data. You need to grant the payroll specialist access to the performance data without granting them access to the entire dataset using the simplest and most secure approach. What should you do?

* **A. Use authorized views to share query results with the payroll specialist.**
* B. Create row-level and column-level permissions and policies on the table that contains performance data in the dataset. Provide the payroll specialist with the appropriate permission set.
* C. Create a table with the aggregated performance data. Use table-level permissions to grant access to the payroll specialist.
* D. Create a SQL query with the aggregated performance data. Export the results to an Avro file in a Cloud Storage bucket. Share the bucket with the payroll specialist.

Answer: A

## Question #: 70

Q: You work for a global financial services company that trades stocks 24/7. You have a Cloud SGL for PostgreSQL user database. You need to identify a solution that ensures that the database is continuously operational, minimizes downtime, and will not lose any data in the event of a zonal outage. What should you do?

* A. Continuously back up the Cloud SGL instance to Cloud Storage. Create a Compute Engine instance with PostgreSCL in a different region. Restore the backup in the Compute Engine instance if a failure occurs.
* B. Create a read replica in another region. Promote the replica to primary if a failure occurs.
* **C. Configure and create a high-availability Cloud SQL instance with the primary instance in zone A and a secondary instance in any zone other than zone A.**
* D. Create a read replica in the same region but in a different zone.

Answer: C

## Question #: 71

Q: You are migrating data from a legacy on-premises MySQL database to Google Cloud. The database contains various tables with different data types and sizes, including large tables with millions of rows and transactional data. You need to migrate this data while maintaining data integrity, and minimizing downtime and cost. What should you do?

* A. Set up a Cloud Composer environment to orchestrate a custom data pipeline. Use a Python script to extract data from the MySQL database and load it to MySQL on Compute Engine.
* B. Export the MySQL database to CSV files, transfer the files to Cloud Storage by using Storage Transfer Service, and load the files into a Cloud SQL for MySQL instance.
* **C. Use Database Migration Service to replicate the MySQL database to a Cloud SQL for MySQL instance.**
* D. Use Cloud Data Fusion to migrate the MySQL database to MySQL on Compute Engine.

Answer: C

## Question #: 72

Q: You created a customer support application that sends several forms of data to Google Cloud. Your application is sending:  
1. Audio files from phone interactions with support agents that will be accessed during trainings.  
2. CSV files of users’ personally identifiable information (Pll) that will be analyzed with SQL.  
3. A large volume of small document files that will power other applications.  
You need to select the appropriate tool for each data type given the required use case, while following Google-recommended practices. Which should you choose?

* A. 1. Cloud Storage  
  2. CloudSQL for PostgreSQL  
  3. Bigtable
* B. 1. Filestore  
  2. Cloud SQL for PostgreSQL  
  3. Datastore
* **C. 1. Cloud Storage  
  2. BigQuery  
  3. Firestore**
* D. 1. Filestore  
  2. Bigtable  
  3. BigQuery

Answer: C