Quiz: Ungraded Practice Exam

Your score: 60%   Passing score: 75%

Retake

Unfortunately, you need at least a 75% to pass this assessment. Not to worry though, review your answers and try again.

✕ 1. Low-cost one-way one-time migration of two 100-TB file servers to Google Cloud; data will be frequently accessed and only from Germany.

  ○ **Use Transfer Appliance. Transfer to a Cloud Storage Standard bucket.**

  ✕ ~~Use Storage Transfer Service. Transfer to a Cloud Storage Standard bucket.~~

  ○ **Use Storage Transfer Service. Transfer to a Cloud Storage Coldline bucket.**

  ○ **Use Transfer Appliance. Transfer to a Cloud Storage Nearline bucket.**

  This is not correct because you should only use Transfer Service for a one-time one-way transfer. Also, Storage Transfer Service does not work for data stored on-premises.

✓ 2. 250,000 devices produce a JSON device status every 10 seconds. How do you capture event data for outlier time series analysis?

  ○ **Capture data in BigQuery. Develop a BigQuery API custom application to query the dataset and display device outlier data.**

  ✓ **Capture data in Cloud Bigtable. Use the Cloud Bigtable cbt tool to display device outlier data.**

  ○ **Capture data in Cloud Bigtable. Install and use the HBase shell for Cloud Bigtable to query the table for device outlier data.**

  ○ **Capture data in BigQuery. Use the BigQuery console to query the dataset and display device outlier data.**

  This is correct because the data type, volume, and query pattern best fit Cloud Bigtable capabilities.

✓ 3. BigQuery data is stored in external CSV files in Cloud Storage; as the data has increased, the query performance has dropped.

  ✓ **Import the data into BigQuery for better performance.**

  ○ **Divide the data into partitions based on date.**

  ○ **Time to move to Cloud Bigtable; it is faster in all cases.**

  ○ **Request more slots for greater capacity to improve performance.**

  This is correct. The performance issue is because the data is stored in a non-optimal format in an external storage medium.

✓ 4. A client wants to store files from one location and retrieve them from another location. Security requirements are that **no one** should be able to access the contents of **the file** while it is hosted in the cloud. What is the best option?

  ○ Default encryption should be sufficient

  ○ Customer-Supplied Encryption Keys (CSEK)

  ○ Customer Managed Encryption Keys (CMEK)

  ✓ Client-side encryption

  This is correct. The requirement is that the file cannot be decrypted in the cloud, so encrypt it before it is uploaded and after it is downloaded adds a layer of encryption.

✕ 5. Cost-effective way to run non-critical Apache Spark jobs on Dataproc?

  ○ **Set up a cluster in high availability mode with high-memory machine types. Add 10 additional local SSDs.**

  ○ **Set up a cluster in standard mode with high-memory machine types. Add 10 additional preemptible worker nodes.**

  ✕ ~~Set up a cluster in standard mode with the default machine types. Add 10 additional local SSDs.~~

  ○ **Set up a cluster in high availability mode with default machine types. Add 10 additional preemptible worker nodes.**

  This is not correct because you should not add more local SSDs; instead, use preemptible nodes to meet your objective of delivering a cost-effective solution.

✓ 6. An application has the following data requirements.

  1. It requires strongly consistent transactions.
  2. Total data will be less than 500 GB.
  3. The data does not need to be streaming or real time.
  Which data technology would fit these requirements?

  ○ **BigQuery**

  ✓ **Cloud SQL**

  ○ **Memorystore**

  ○ **Cloud Bigtable**

  This is correct. Cloud SQL supports strongly consistent transactions. And the size requirements will fit with a Cloud SQL instance.

✓ 7. A company has migrated their Hadoop cluster to the cloud and is now using Dataproc with the same settings and methods as in the data center. What would you advise them to do to make better use of the cloud environment?

○ Upgrade to the latest version of HDFS. Change the settings in Hadoop components to optimize for the different kinds of work in the mix.

✓ Store persistent data off-cluster. Start a cluster for one kind of work then shut it down when it is not processing data.

○ Migrate from Dataproc to an open source Hadoop Cluster hosted on Compute Engine, because this is the only way to get all the Hadoop customizations needed for efficiency.

○ Find more jobs to run so the cluster utilizations will cost-justify the expense.

This is correct. Storing persistent data off the cluster allows the cluster to be shut down when not processing data. And it allows separate clusters to be started per job or per kind of work, so tuning is less important.

✓ 8. Source data is streamed in bursts and must be transformed before use.

○ **Use Cloud Bigtable for fast input and cbt for ETL.**

○ **Use Pub/Sub to buffer the data, and then use BigQuery for ETL.**

✓ **Use Pub/Sub to buffer the data, and then use Dataflow for ETL.**

○ **Ingest data to Cloud Storage. Use Dataproc for ETL.**

This is correct because the unpredictable data requires a buffer

✓ 9. Calculate a running average on streaming data that can arrive late and out of order.

✓ **Use Pub/Sub and Dataflow with Sliding Time Windows.**

○ **Pub/Sub can guarantee timely arrival and order.**

○ **Use Dataflow's built-in timestamps for ordering and filtering.**

○ **Use Pub/Sub and Google Data Studio.**

This is correct because together, Pub/Sub and Dataflow can provide a solution.

✗ 10. A company has a new IoT pipeline. Which services will make this design work?
Select the services that should be used to replace the icons with the number "1" and number "2" in the diagram.
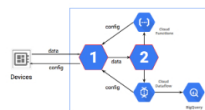
○ **IoT Core, Datastore**

○ **IoT Core, Pub/Sub**

✗ ~~**App Engine, IoT Core**~~

○ **Pub/Sub, Cloud Storage**

This is not correct. In theory, an App Engine application could duplicate the functions of IoT Core, but since IoT Core only publishes to Pub/Sub, in position 2, it would not communicate with either Cloud Functions or Dataflow.

✗ 11. Storage of JSON files with occasionally changing schema, for ANSI SQL queries.

✗ ~~**Store in BigQuery. Provide format files for data load and update them as needed.**~~

○ **Store in Cloud Storage. Link data as temporary tables in BigQuery and turn on the "Automatically detect" option in the Schema section of BigQuery.**

○ **Store in Cloud Storage. Link data as permanent tables in BigQuery and turn on the "Automatically detect" option in the Schema section of BigQuery.**

○ **Store in BigQuery. Select "Automatically detect" in the Schema section.**

This is not correct because you should not provide format files: you can simply turn on the 'Automatically detect' schema changes flag.

✗ 12. Cost-effective backup to Google Cloud of multi-TB databases from another cloud including monthly DR drills.

○ Use Transfer Appliance. Transfer to Cloud Storage Nearline bucket.

○ Use Storage Transfer Service. Transfer to Cloud Storage Nearline bucket.

○ Use Storage Transfer Service. Transfer to Cloud Storage Coldline bucket.

✗ ~~Use Transfer Appliance. Transfer to Cloud Storage Coldline bucket.~~

This is not correct because you should not use Coldline if you want to access the files monthly and you should not use Transfer Appliance for repeated data transfers.

✓ 13. Three Google Cloud services commonly used together in data engineering solutions. (Described in this course).

○ Dataproc, Cloud SQL, BigQuery

○ Pub/Sub, Google Kubernetes Engine, Cloud Spanner

○ Cloud Bigtable, Dataproc, Cloud Spanner

✓ Pub/Sub, Dataflow, BigQuery

Correct. Pub/Sub provides messaging, Dataflow is used for ETL and data transformation, and BigQuery is used for interactive queries.

✕ 14. Promote a Cloud Bigtable solution with a lot of data from development to production and optimize for performance.

○ Change your Cloud Bigtable instance type from Development to Production, and set the number of nodes to at least 3. Verify that the storage type is HDD.

○ Export the data from your current Cloud Bigtable instance to Cloud Storage. Create a new Cloud Bigtable Production instance type with at least 3 nodes. Select the HDD storage type. Import the data into the new instance from Cloud Storage.

✕ ~~Export the data from your current Cloud Bigtable instance to Cloud Storage. Create a new Cloud Bigtable Production instance type with at least 3 nodes. Select the SSD storage type. Import the data into the new instance from Cloud Storage.~~

○ Change your Cloud Bigtable instance type from Development to Production, and set the number of nodes to at least 3. Verify that the storage type is SSD.

This is not correct because creating a new Cloud Bigtable instance is extraneous and not needed to export; you can upgrade in place for nodes, but the storage type cannot be changed.

✓ 15. Event data in CSV format to be queried for individual values over time windows. Which storage and schema to minimize query costs?

✓ Use Cloud Bigtable. Design tall and narrow tables, and use a new row for each single event version.

○ Use Cloud Storage. Join the raw file data with a BigQuery log table.

○ Use Cloud Storage. Write a Dataprep job to split the data into partitioned tables.

○ Use Cloud Bigtable. Design short and wide tables, and use a new column for each single event version.

This is correct because it is a recommended best practice. Use Cloud Bigtable and this schema for this scenario. Cloud Storage would have cheaper STORAGE costs than Cloud Bigtable, but we want to minimize QUERY costs.

✓ 16. Customer wants to maintain investment in an existing Apache Spark code data pipeline.

○ BigQuery

✓ Dataproc

○ Dataprep

○ Dataflow

This is correct because Dataproc is a managed Hadoop service and runs Apache Spark applications.

✕ 17. A company wants to connect cloud applications to an Oracle database in its data center. Requirements are a maximum of 9 Gbps of data and a Service Level Agreement (SLA) of 99%.

✕ ~~Implement a high-throughput Cloud VPN connection~~

○ Dedicated Interconnect

○ Partner Interconnect

○ Cloud Router with VPN

This is not correct. Cloud VPN traverses the public internet. It is useful for low-volume connections. The SLA offered by Google covers the Cloud VPN service itself, and not the internet transport. So it would not meet the SLA requirement. Which is the best option?

✕ 18. A Data Analyst is concerned that a BigQuery query could be too expensive.

○ Use the LIMIT clause to limit the number of values in the results.

✕ ~~Set the Maximum Bytes Billed, which will limit the number of bytes processed but still run the query if the number of bytes requested goes over the limit.~~

○ Use GROUP BY so the results will be grouped into fewer output values.

○ Use the SELECT clause to limit the amount of data in the query. Partition data by date so the query can be more focused.

This is not correct because if the query contains too many bytes, the job will fail and not be run.

✓ 19. What is AVRO used for?

✓ Serialization and de-serialization of data so that it can be transmitted and stored while maintaining an object structure.

○ AVRO is a file type usually specified with *.avr and a common format for spreadsheets.

○ AVRO is a numerical type in SQL that stores a 38 digit value with 9 digit decimal representation. It avoids rounding errors in financial calculations.

○ AVRO is an encryption method. AVRO-256 is a 256-bit key standard.

This is correct. AVRO is a serialization / de-serialization standard.

✓ 20. Host a deep neural network machine learning model on Google Cloud. Run and monitor jobs that could occasionally fail.

○ **Use Vertex AI to host your model. Monitor the status of the Operation object for 'error' results.**

○ **Use a Google Kubernetes Engine cluster to host your model. Monitor the status of the Jobs object for 'failed' job states.**

○ **Use a Google Kubernetes Engine cluster to host your model. Monitor the status of the Operation object for 'error' results.**

✓ **Use Vertex AI to host your model. Monitor the status of the Jobs object for 'failed' job states.**

This is correct because of the requirement to host an ML DNN. Vertex AI for Tensorflow can handle DNNs. Google recommends monitoring Jobs, not Operations.

✓ 21. A client has been developing a pipeline based on PCollections using local programming techniques and is ready to scale up to production. What should they do?

✓ They should use the Dataflow Cloud Runner.

○ They should use the local version of runner.

○ Import the pipeline into BigQuery.

○ They should upload the pipeline to Dataproc.

This is correct. The PCollection indicates it is a Dataflow pipeline. And the Cloud Runner will enable the pipeline to scale to production levels.

✓ 22. You want to minimize costs to run Google Data Studio reports on BigQuery queries by using prefetch caching.

○ **Set up the report to use the Owner's credentials to access the underlying data in BigQuery, and direct the users to view the report only once per business day (24-hour period).**

○ **Set up the report to use the Viewer's credentials to access the underlying data in BigQuery, and also set it up to be a 'view-only' report.**

○ **Set up the report to use the Viewer's credentials to access the underlying data in BigQuery, and verify that the 'Enable cache' checkbox is not selected for the report.**

✓ **Set up the report to use the Owner's credentials to access the underlying data in BigQuery, and verify that the 'Enable cache' checkbox is selected for the report.**

This is correct because you must set Owner credentials to use the 'enable cache' option in BigQuery. It is also a Google best practice to use the 'enable cache' option when the business scenario calls for using prefetch caching. 1) Report must use Owner's Credentials. 2) You don't need to tell the users not to use the report, you need to tell the system to use Query and Pre-fetch caching to cut down on BigQuery jobs.

✗ 23. A client is using Cloud SQL database to serve infrequently changing lookup tables that host data used by applications. The applications will not modify the tables. As they expand into other geographic regions they want to ensure good performance. What do you recommend?

✗ ~~Migrate to Cloud Spanner~~

○ **Instance high availability configuration**

○ **Replicate from an external server**

○ **Read replicas**

This is not correct because there is no mention of a scale issue requiring a larger database or globally consistent transactions.