



# Quiz: Ungraded Practice Exam

Your score: 91%    Passing score: 75%

[Retake](#)

Congratulations! You passed this assessment.

- ✗ 1. 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%.

- ✗ **Dedicated Interconnect**
- ☐ **Partner Interconnect**
- ☐ **Cloud Router with VPN**
- ☐ **Implement a high-throughput Cloud VPN connection**

This is not correct. Direct Interconnect is useful for data from 10 Gbps to 80 Gbps. An ISP could offer a 99% SLA, but the max 9 Gbps requirement means this solution would not be optimal.

- ✓ 2. What is AVRO used for?
- ☐ 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.
- ✓ ☒ Serialization and de-serialization of data so that it can be transmitted and stored while maintaining an object structure.

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

- ✓ 3. Customer wants to maintain investment in an existing Apache Spark code data pipeline.
- ✓ **Dataproc**
- ☐ **Dataprep**
- ☐ **Dataflow**
- ☐ **BigQuery**

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

- ✓ 4. 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?

- ✓ ☒ 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.
- ☐ Upgrade to the latest version of HDFS. Change the settings in Hadoop components to optimize for the different kinds of work in the mix.

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.

- ✓ 5. 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?

- ☐ Customer-Supplied Encryption Keys (CSEK)
- ☐ Customer Managed Encryption Keys (CMEK)
- ✓ ☒ Client-side encryption
- ☐ Default encryption should be sufficient

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.

- ✓ 6. 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 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.
- ☐ 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).

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.

- ✓ 7. Promote a Cloud Bigtable solution with a lot of data from development to production and optimize for performance.
- ☐ 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.
  - ☐ 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.

This is correct because Cloud Bigtable allows you to 'scale in place,' which meets your requirements for this scenario.

- ✓ 8. 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?
- ☐ Instance high availability configuration
  - ☐ Replicate from an external server
  - ✓ Read replicas
  - ☐ Migrate to Cloud Spanner

This is correct. A read replica will increase the availability of the service and can be located closer to the users in the new geographies.

- ✓ 9. 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 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.
  - ☐ Capture data in BigQuery. Develop a BigQuery API custom application 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.

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

- ☐ 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.
- ✓ ☒ Use Pub/Sub and Dataflow with Sliding Time Windows.

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

- ✓ 11. 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 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.
- ✓ ☒ Use Transfer Appliance. Transfer to a Cloud Storage Standard bucket.

This is correct because you are performing a one-time (rather than an ongoing series) data transfer from on-premises to Google Cloud for users in a single region (Germany). Using a Standard storage bucket is best for data that is frequently accessed, will reduce cost and also conform to regulatory requirements.

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

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

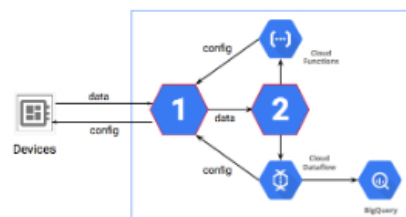
This is correct because of the requirement to host an ML DNN. Vertex AI for Tensorflow can handle DNNs. Google recommends monitoring Jobs, not Operations.

- ✓ 13. Cost-effective way to run non-critical Apache Spark jobs on Dataproc?

- ✓ ☒ 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.
- ☐ Set up a cluster in high availability mode with high-memory machine types. Add 10 additional local SSDs.

This is correct because Spark and high-memory machines only need the standard mode. Also, use preemptible nodes because you want to save money and this is not mission-critical.

- ✓ 14. 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, Pub/Sub
- ☐ App Engine, IoT Core
- ☐ Pub/Sub, Cloud Storage
- ☐ IoT Core, Datastore

This is correct because device data captured by IoT Core gets published to Pub/Sub.

- ✓ 15. 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 local version of runner.
  - ☐ Import the pipeline into BigQuery.
  - ☐ They should upload the pipeline to Dataproc.
  - ✓ ☒ They should use the Dataflow Cloud Runner.

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

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

- ☐ 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.
- ☐ Use Cloud Bigtable for fast input and cbt for ETL.

This is correct because the unpredictable data requires a buffer

- ✓ 17. Storage of JSON files with occasionally changing schema, for ANSI SQL queries.

- ☐ 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.
- ☐ Store in BigQuery. Provide format files for data load and update them as needed.

This is correct because of the requirement to support occasionally (schema) changing JSON files and aggregate ANSI SQL queries: you need to use BigQuery, and it is quickest to use 'Automatically detect' for schema changes.

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

- ☐ 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.
- ✓ ☒ Import the data into BigQuery for better performance.

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

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

- ☐ 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.
- ✗ ☒ Use Transfer Appliance. Transfer to Cloud Storage Nearline bucket.



bucket.

This is not correct because you should not use Transfer Appliance for repeated data transfers.

- ✓ 20. 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?

- ☒ **Cloud SQL**
- ☐ **Memorystore**
- ☐ **Cloud Bigtable**
- ☐ **BigQuery**

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

- ✓ 21. A Data Analyst is concerned that a BigQuery query could be too expensive.

- ☐ **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.**
- ☐ **Use the LIMIT clause to limit the number of values in the results.**

This is correct. SELECT limits the input data.

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

- ☐ **Pub/Sub, Google Kubernetes Engine, Cloud Spanner**
- ☐ **Cloud Bigtable, Dataproc, Cloud Spanner**
- ☒ **Pub/Sub, Dataflow, BigQuery**
- ☐ **Dataproc, Cloud SQL, BigQuery**

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

- ✓ 23. Event data in CSV format to be queried for individual values over time

windows. Which storage and schema to minimize query costs?

- ☐ 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.
- ☒ Use Cloud Bigtable. Design tall and narrow tables, and use a new row 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**.