Congratulations! You passed!

TO PASS 90% or higher

Keep Learning

GRADE
92%

Ungraded Practice Exam Quiz

TOTAL POINTS 25

1. Storage of JSON files with occasionally changing schema, for ANSI SQL queries. **1 / 1 point**

- ○ **Store in BigQuery. Provide format files for data load and update them as needed.**
- ◉ **Store in BigQuery. Select "Automatically detect" in the Schema section.**
- ○ **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.**

✓ **Correct**

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.

2. Low-cost one-way one-time migration of two 100-TB file servers to GCP; data will only be accessed from Germany. **1 / 1 point**

- ◉ **Use Transfer Appliance. Transfer to a Cloud Storage Regional storage bucket.**
- ○ **Use Transfer Appliance. Transfer to a Cloud Storage Multi-Regional bucket.**
- ○ **Use Storage Transfer Service. Transfer to a Cloud Storage Regional bucket.**
- ○ **Use Storage Transfer Service. Transfer to a Cloud Storage Multi-Regional bucket.**

✓ **Correct**

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

3. Cost-effective backup to GCP of multi-TB databases from another cloud including monthly DR drills. **1 / 1 point**

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

- [x] Use Storage Transfer Service. Transfer to Cloud Storage Nearline bucket.
- [ ] Use Storage Transfer Service. Transfer to Cloud Storage Coldline bucket.

Correct

This is correct because you will need to access your backup data monthly to test your disaster recovery process, so you should use a Nearline bucket; also because you will be performing ongoing, regular data transfers, so you should use Storage Transfer Service.

4. 250,000 devices produce a JSON device status every 10 seconds. How do you capture event data for outlier time series analysis? **1 / 1 point**

- [ ] **Capture data in BigQuery. Develop a BigQuery API custom application to query the dataset and display device outlier data.**
- [ ] **Capture data in BigQuery. Use the BigQuery console to query the dataset and display device outlier data.**
- [x] **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.**

Correct

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

5. Event data in CSV format to be queried for individual values over time windows. Which storage and schema to minimize query costs? **1 / 1 point**

- [ ] **Use Cloud Storage. Join the raw file data with a BigQuery log table.**
- [ ] **Use Cloud Storage. Write a Cloud 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.**
- [x] **Use Cloud Bigtable. Design tall and narrow tables, and use a new row for each single event version.**

Correct

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.***

6. Customer wants to maintain investment in existing Apache Spark code data pipeline. **1 / 1 point**

- [ ] **BigQuery**
- [ ] **Cloud Dataflow**
- [x] **Cloud Dataproc**

- Cloud Dataprep

✓ **Correct**

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

7. Host a deep neural network machine learning model on GCP. Run and monitor jobs that could occasionally fail. **1 / 1 point**

- **Use Cloud Machine Learning Engine to host your model. Monitor the status of the Operation object for 'error' results.**
- **Use Cloud Machine Learning Engine to host your model. Monitor the status of the Jobs object for 'failed' job states.**
- **Use a Kubernetes Engine cluster to host your model. Monitor the status of the Jobs object for 'failed' job states.**
- **Use a Kubernetes Engine cluster to host your model. Monitor the status of the Operation object for 'error' results.**

✓ **Correct**

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

8. Cost-effective way to run non-critical Apache Spark jobs on Cloud Dataproc? **1 / 1 point**

- **Set up a cluster in high availability mode with high-memory 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 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.**

✓ **Correct**

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.

9. Promote a Cloud Bigtable solution with a lot of data from development to production and optimize for performance. **1 / 1 point**

- **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.**
- **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.**
- **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.**

Correct

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

10. As part of your backup plan, you want to be able to restore snapshots of Compute Engine instances using the fewest steps. **1 / 1 point**

- **Export the snapshots to Cloud Storage. Create disks from the exported snapshot files. Create images from the new disks.**
- **Export the snapshots to Cloud Storage. Create images from the exported snapshot files.**
- **Use the snapshots to create replacement disks. Use the disks to create instances as needed.**
- **Use the snapshots to create replacement instances as needed.**

Correct

This is correct because the scenario asks how to recreate instances. You can create an instance directly from a snapshot without restoring to disk first.

11. You want to minimize costs to run Google Data Studio reports on BigQuery queries by using prefetch caching. **1 / 1 point**

- **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 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 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.**

Correct

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.

12. A Data Analyst is concerned that a BigQuery query could be too expensive. **1 / 1 point**

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

- (●) **Use the SELECT clause to limit the amount of data in the query. Partition data by date so the query can be more focused.**
- ( ) **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.**

✓ **Correct**

This is correct. SELECT limits the input data.

13. BigQuery data is stored in external CSV files in Cloud Storage; as the data has increased, the query performance has dropped. **1 / 1 point**

- (●) **Import the data into BigQuery for better performance.**
- ( ) **Request more slots for greater capacity to improve performance.**
- ( ) **Divide the data into partitions based on date.**
- ( ) **Time to move to Cloud Bigtable; it is faster in all cases.**

✓ **Correct**

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

14. Source data is streamed in bursts and must be transformed before use. **1 / 1 point**

- ( ) **Use Cloud Bigtable for fast input and cbt for ETL.**
- ( ) **Ingest data to Cloud Storage. Use Cloud Dataproc for ETL.**
- ( ) **Use Cloud Pub/Sub to buffer the data, and then use BigQuery for ETL.**
- (●) **Use Cloud Pub/Sub to buffer the data, and then use Cloud Dataflow for ETL.**

✓ **Correct**

This is correct because the unpredictable data requires a buffer

15. Calculate a running average on streaming data that can arrive late and out of order. **1 / 1 point**

- (●) **Use Cloud Pub/Sub and Cloud Dataflow with Sliding Time Windows.**
- ( ) **Use Cloud Pub/Sub and Google Data Studio.**
- ( ) **Cloud Pub/Sub can guarantee timely arrival and order.**
- ( ) **Use Cloud Dataflow's built-in timestamps for ordering and filtering.**

Correct

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

16. Testing a Machine Learning model with validation data returns 100% correct answers. **1 / 1 point**

- **The model is working extremely well, indicating the hyperparameters are set correctly.**
- **The model is overfit. There is a problem.**
- **The model is underfit. There is a problem.**
- **The model is perfectly fit. You do not need to continue training.**

Correct

This is correct. The 100% accuracy is an indicator that the validation data may have somehow gotten mixed in with the training data. You will need new validation data to generate recognizable error.

17. 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? **0 / 1 point**

- **Migrate to Cloud Spanner**
- **Read replicas**
- **Instance high availability configuration**
- **Replicate from an external server**

Incorrect

High availability will not improve performance or availability for additional geographic locations.

18. 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? **1 / 1 point**

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

Correct

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.

19. Three Google Cloud services commonly used together in data engineering solutions. *(Described in this course).* **1 / 1 point**

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

Correct

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

20. What is AVRO used for? **1 / 1 point**

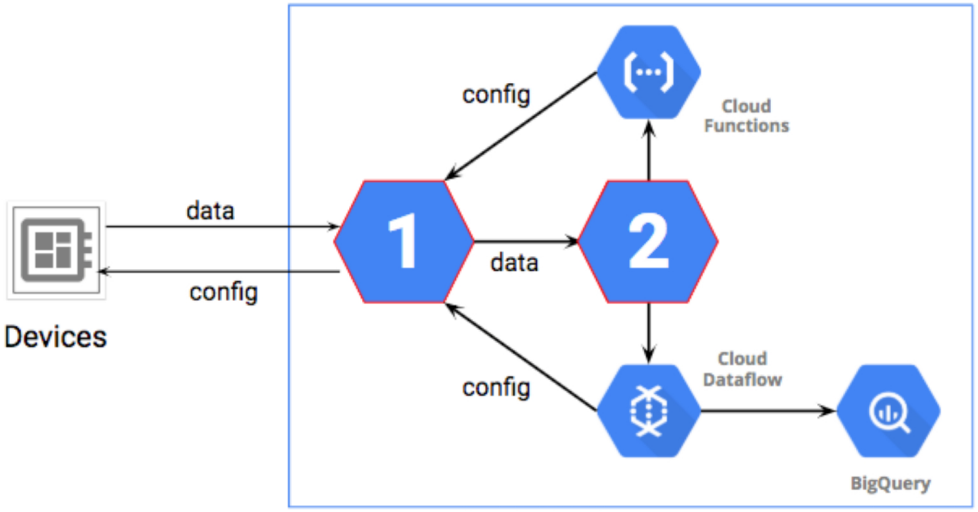
- **Serialization and de-serialization of data so that it can be transmitted and stored while maintaining an object structure.**
- AVRO is an encryption method. AVRO-256 is a 256-bit key standard.
- AVRO is a file type usually specified with *.avr and a common format for spreadsheets.
- AVRO is a numerical type in SQL that stores a 38 digit value with 9 digit decimal representation. It avoids rounding errors in financial calculations.

Correct

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

21. A company has a new IoT pipeline. Which services will make this design work? **1 / 1 point**

Select the services that should be used to replace the icons with the number "1" and number "2" in the diagram.



- Cloud IoT Core, Cloud Datastore
- Cloud Pub/Sub, Cloud Storage
- **Cloud IoT Core, Cloud Pub/Sub**
- App Engine, Cloud IoT Core

✓ **Correct**

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

22. 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%. **0 / 1 point**

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

! **Incorrect**

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?

23. 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? **1 / 1 point**

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

✓ **Correct**

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

24. A company has migrated their Hadoop cluster to the cloud and is now using Cloud Dataproc with the same settings and same methods as in the data center. What would you advise them to do to make better use of the cloud environment? **1 / 1 point**

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

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

Correct

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.

25. 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? **1 / 1 point**

- **BigQuery**
- **Cloud Bigtable**
- **Cloud SQL**
- **Cloud Memorystore**

Correct

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