✓ **Congratulations! You passed!**

Grade received 100%   **To pass** 80% or higher

Go to next item

Module Review

Latest Submission Grade 100%

---

1. Complete the following:

You should feed your machine learning model your _____ and not your _____. It will learn those for itself!

◉ data, rules

○ if/then statements, data

○ rules, data

1 / 1 point

> ✓ **Correct**
> Correct!

---

2. True or False: Cloud SQL is a big data analytics warehouse

○ True

◉ False

1 / 1 point

> ✓ **Correct**
> Correct - Cloud SQL is a transaction RDBMS or relational database management system. It is designed for many more WRITES than READS.
>
> Whereas BigQuery is a big data analytics warehouse which is optimized for reporting READS.

---

3. True or False: If you are migrating your Hadoop workload to the cloud, you must first rewrite all your Spark jobs to be compliant with the cloud.

○ True

◉ False

1 / 1 point

> ✓ **Correct**
> Correct - you can run your same Spark job code running on the same Hadoop software but running on cloud hardware with Cloud Dataproc.

---

4. You are thinking about migrating your Hadoop workloads to the cloud and you have a few workloads that are fault-tolerant (they can handle interruptions of individual VMs gracefully). What are some architecture considerations you should explore in the cloud? Choose all that apply

☑ Use PVMs or Preemptible Virtual Machines

1 / 1 point

> ✓ **Correct**
> Correct!

☑ Migrate your storage from on-cluster HDFS to off-cluster Google Cloud Storage (GCS)

> ✓ **Correct**
> Correct!

☑ Consider having multiple Cloud Dataproc instances for each priority workload and then turning them down when not in use

✓ **Correct**
Correct!

5. Google Cloud Storage is a good option for storing data that:

(Select the **2** correct options below).

☐ Is ingested in real-time from sensors and other devices and supports SQL-based queries

☑ May be required to be read at some later time (i.e. load a CSV file into BigQuery)

✓ **Correct**
Correct!

☐ Will be accessed frequently and updated constantly with new transactions from a front-end and needs to be stored in a relational database

☑ May be imported from a bucket into a Hadoop cluster for analysis

✓ **Correct**
Correct!

6. Relational databases are a good choice when you need:

◯ Aggregations on unstructured data

◯ Fast queries on terabytes of data

◯ Streaming, high-throughput writes

◉ Transactional updates on relatively small datasets

✓ **Correct**
Correct!

7. Cloud SQL and Cloud Dataproc offer familiar tools (MySQL and Hadoop/Pig/Hive/Spark). What is the value-add provided by Google Cloud Platform?

(Select the **2** correct options below)

☑ Running it on Google infrastructure offers reliability and cost savings

✓ **Correct**
Yes. You pay only for the resources you use. Cloud SQL can be shut down when it's not being used. Hadoop clusters can be of preemptible nodes, and so on.

☐ It's the same API, but Google implements it better

☑ Fully-managed versions of the software offer no-ops

✓ **Correct**
Yes. No-ops is the main value-add here.

☐ Google-proprietary extensions and bug fixes to MySQL, Hadoop, and so on

1 / 1 point

1 / 1 point

1 / 1 point