Module Review

TOTAL POINTS 7

1.Question 1
Complete the following:

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

1 point

if/then statements, data

data, rules

rules, data

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

1 point

True

False

3.Question 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.

1 point

True

False

4.Question 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

1 point

- [x] Use PVMs or Preemptible Virtual Machines
- [x] Migrate your storage from on-cluster HDFS to off-cluster Google Cloud Storage (GCS)
- [x] Consider having multiple Cloud Dataproc instances for each priority workload and then turning them down when not in use

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

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

1 point

- [ ] Is ingested in real-time from sensors and other devices and supports SQL-based queries
- [x] May be imported from a bucket into a Hadoop cluster for analysis
- [ ] May be required to be read at some later time (i.e. load a CSV file into BigQuery)
- [x] Will be accessed frequently and updated constantly with new transactions from a front-end and needs to be stored in a relational database

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

1 point

- Streaming, high-throughput writes
- Fast queries on terabytes of data

Transactional updates on relatively small datasets

Aggregations on unstructured data

7.Question 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)

1 point

Running it on Google infrastructure offers reliability and cost savings

Fully-managed versions of the software offer no-ops

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

It’s the same API, but Google implements it better