



Navigation


Read the [Linux Academy Oath](https://go.acloudguru.com/faq-la-customer-oath) (<https://go.acloudguru.com/faq-la-customer-oath>) from Sam Kroonenburg, our CEO.

[See What the LA Product Team Has Been Up To! \(/updates\)](#)

## Google Certified Professional Data Engineer - Managed Spark with Cloud Dataproc Quiz

 30  
minutes

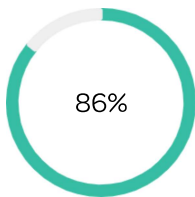
 7  
Questions

 4.29  
Minutes  
per  
Question

[Advanced \(/search?type=Practice Exam Challenge&difficulty=Advanced&categories=Cloud\)](#)

Go Back

Start Challenge



 Go Back

**Congratulations!**

You passed this challenge on this attempt.

### Expectations Report Card

Google Certified Professional Data Engineer - Managed Spark with Cloud Dataproc Quiz

85.71%

### Exam Breakdown

Google Certified Professional Data Engineer - Managed Spark with Cloud Dataproc Quiz 

INCORRECT

1. Which features are not compatible with Dataproc autoscaling?



**A** HDFS Storage

B Preemptible workers

C Spark Structured Streaming

D High-availability clusters

E MapReduce tasks

F Apache Pig jobs

**Your Answer: F**

**Why is this incorrect?**

**Correct Answer: A**

**Why is this correct?**

Dataproc autoscaling can automate the task of determining the correct number of workers required for a particular workload, but there are certain scenarios when its use would not be recommended.

**Correct Answer: C**

**Why is this correct?**

Dataproc autoscaling can automate the task of determining the correct number of workers required for a particular workload, but there are certain scenarios when its use would not be recommended.

**Correct Answer: D**

**Why is this correct?**

Dataproc autoscaling can automate the task of determining the correct number of workers required for a particular workload, but there are certain scenarios when its use would not be recommended.

---

2. Which native output connectors are supported by Dataproc?



A BigQuery

B Cloud Bigtable

C Cloud Firestore

D Cloud SQL

E Cloud Storage

**Correct Answer: A**

**Why is this correct?**

Cloud Dataproc has built-in integration with BigQuery, Cloud Storage, Cloud Bigtable, Stackdriver Logging, and Stackdriver Monitoring.

**Correct Answer: B**

**Why is this correct?**

Cloud Dataproc has built-in integration with BigQuery, Cloud Storage, Cloud Bigtable, Stackdriver Logging, and Stackdriver Monitoring.

**Correct Answer: E**

**Why is this correct?**

Cloud Dataproc has built-in integration with BigQuery, Cloud Storage, Cloud Bigtable, Stackdriver Logging, and Stackdriver Monitoring.

3. True or False: Preemptible workers in a Dataproc cluster cannot store HDFS data.



A False

**B True**

**Correct Answer: B**

**Why is this correct?**

Since preemptibles can be reclaimed at any time, preemptible workers do not store data.

4. A customer wants to run Spark jobs on a low-cost ephemeral Dataproc cluster, utilizing preemptible workers wherever possible, but needs to store the results of Dataproc jobs persistently. What would you recommend?



A Do not use preemptible workers at all, it will prevent you from choosing any persistent storage option.

B Use a secondary group of preemptible worker nodes, but ensure there is enough persistent storage on the primary (non-preemptible) worker nodes to store all of the data.

C Use a secondary group of preemptible worker nodes, but add custom code to a job that copies its results to Cloud Storage.

**D Use the Cloud Storage connector, and specify GCS locations for the input and output of jobs.**

**Correct Answer: D**

**Why is this correct?**

The Cloud Storage connector lets you run Apache Hadoop or Apache Spark jobs directly on data in Cloud Storage and offers a number of other benefits over HDFS.

**Question List** Show All Answers

5. Which primary Apache services does Dataproc run?



**A Hadoop**

**B Spark**

C Dataflow

D Cassandra

E Kafka

**Correct Answer: A**

**Why is this correct?**

Cloud Dataproc is a managed Spark and Hadoop service that lets you take advantage of open source data tools for batch processing, querying, streaming, and machine learning.

**Correct Answer: B**

**Why is this correct?**

Cloud Dataproc is a managed Spark and Hadoop service that lets you take advantage of open source data tools for batch processing, querying, streaming, and machine learning.

6. Your customer would like to use Dataproc, but the standard image does not contain some additional Spark components required to run their jobs. What would you recommend?



**A** Create custom Dataproc image that fulfils the customer requirements and use it to deploy a Dataproc cluster.

B Create an image that fulfils the customer requirements and use it to deploy a custom Spark cluster using Compute Engine.

C Split the customer workloads into 2 clusters. Where the extra components are not required, use Dataproc. Where extra components are required, build a custom image and use it to deploy a custom Spark cluster using Compute Engine.

D Use a Dataproc cluster, but specify an initialization action that installs all of the additional components.

**Correct Answer: A**

**Why is this correct?**

Cloud Dataproc clusters can be provisioned with a custom image that includes a user's pre-installed packages. You could alternatively use initialization actions to install the additional components, but this would be less efficient and incur more running time for ephemeral clusters.

7. Which GCP product implements the Apache Beam SDK and is sometimes recommended as an alternative to Dataproc particularly for streaming data?



A Cloud Data Fusion

**B** Cloud Dataflow

C Cloud Composer

D Cloud Datalab

**Correct Answer: B****Why is this correct?**

The Apache Beam SDK is an open source programming model that enables you to develop both batch and streaming pipelines. You create your pipelines with an Apache Beam program and then run them on the Dataflow service.

---