

✓ Congratulations! You passed!

Next item



1. Which of these accurately describes the relationship between Apache Beam and Cloud Dataflow?

1 / 1 point

☒ Apache Beam is the API for data pipeline building in java or python and Cloud Dataflow is the implementation and execution framework.

Correct

☐ Cloud Dataflow is the API for data pipeline building in java or python and Apache Beam is the implementation and execution framework.

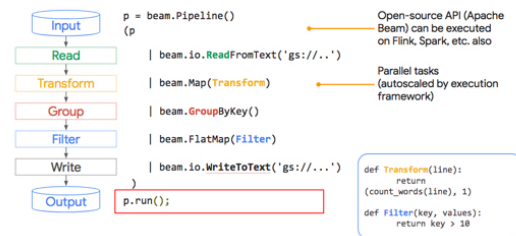
☐ Cloud Dataflow is the proprietary version of the Apache Beam API and the two are not compatible

☐ They are the same



2. TRUE or FALSE: The Filter method can be carried out in parallel and autoscaled by the execution framework:

1 / 1 point



☒ True

Correct

Anything in Map or FlatMap can be parallelized by the Beam execution framework

☐ False



3. What is the purpose of a Cloud Dataflow connector?

1 / 1 point

.apply(TextIO.write().to("gs://..."));

☐ Connectors allow you to authenticate your pipeline as specific users who may have greater access to datasets

☐ Connectors allow you to chain multiple data-processing steps together automatically so they process in parallel

☒ Connectors allow you to output the results of a pipeline to a specific data sink like Bigtable, Google Cloud Storage, flat file, BigQuery, and more...

Correct



4. Below you'll find a Cloud Dataflow preprocessing graph. Correctly identify the terms for A, B, and C.

1 / 1 point





- ☐ A is a data stream,
B are transformation steps, and
C is a data source
- ☒ A is a data source,
B are transformation steps, and
C is a data sink

Correct

- ☐ A is a data stream,
B are transformation steps, and
C is a data sink

✓ 5. To run a pipeline you need something called a _____

1 / 1
point

- ☐ executor
- ☒ runner

Correct

- ☐ pipeline
- ☐ Apache Beam

✓ 6. Your development team is about to execute this code block. What is your team about to do?

1 / 1
point

```

mvn compile -e exec:java \
  -Dexec.mainClass=$MAIN \
  -Dexec.args="--project=$PROJECT \
  --stagingLocation=gs://$BUCKET/staging/ \
  --tempLocation=gs://$BUCKET/staging/ \
  --runner=DataflowRunner"
  
```

- ☒ We are compiling our Cloud Dataflow pipeline written in Java and are submitting it to the cloud for execution

Correct

Notice that we are calling mvn compile and passing in --runner=DataflowRunner

- ☐ We are compiling our Cloud Dataflow pipeline written in Python and are loading the outputs of the executed pipeline inside of Google Cloud Storage (gs://)
- ☐ We are preparing a staging area in Google Cloud Storage for the output of our Cloud Dataflow pipeline and will be submitting our BigQuery job with a later command



7. TRUE or FALSE: A ParDo acts on all items at once (like a Map in MapReduce)

1 / 1
point



True



False

Correct

A ParDo acts on one item at a time (like a Map in MapReduce)

