

[Google Discussions](#)

## Exam Professional Data Engineer All Questions

View all questions & answers for the Professional Data Engineer exam

[Go to Exam](#)

### EXAM PROFESSIONAL DATA ENGINEER TOPIC 1 QUESTION 86 DISCUSSION

Actual exam question from Google's Professional Data Engineer

Question #: 86

Topic #: 1

[\[All Professional Data Engineer Questions\]](#)

You have an Apache Kafka cluster on-prem with topics containing web application logs. You need to replicate the data to Google Cloud for analysis in BigQuery and Cloud Storage. The preferred replication method is mirroring to avoid deployment of Kafka Connect plugins.

What should you do?

- A. Deploy a Kafka cluster on GCE VM Instances. Configure your on-prem cluster to mirror your topics to the cluster running in GCE. Use a Dataproc cluster or Dataflow job to read from Kafka and write to GCS.
- B. Deploy a Kafka cluster on GCE VM Instances with the Pub/Sub Kafka connector configured as a Sink connector. Use a Dataproc cluster or Dataflow job to read from Kafka and write to GCS.
- C. Deploy the Pub/Sub Kafka connector to your on-prem Kafka cluster and configure Pub/Sub as a Source connector. Use a Dataflow job to read from Pub/Sub and write to GCS.
- D. Deploy the Pub/Sub Kafka connector to your on-prem Kafka cluster and configure Pub/Sub as a Sink connector. Use a Dataflow job to read from Pub/Sub and write to GCS.

[Show Suggested Answer](#)

by [Rajokkiyam](#) at March 22, 2020, 4:20 a.m.

## Comments

Type your comment...

Submit

  **Ganshank** Highly Voted  3 years, 6 months ago

A.

<https://cwiki.apache.org/confluence/pages/viewpage.action?pagelId=27846330>

The solution specifically mentions mirroring and minimizing the use of Kafka Connect plugin.

D would be the more Google Cloud-native way of implementing the same, but the requirement is better met by A.

   upvoted 34 times

  **[Removed]** Highly Voted  3 years, 7 months ago

Answer: A

Description: Question says mirroring and avoid kafka connect plugins

   upvoted 11 times

  **Qix** Most Recent  4 months, 4 weeks ago

Pub/Sub Kafka connector requires Kafka Connect, as described here [https://cloud.google.com/pubsub/docs/connect\\_kafka](https://cloud.google.com/pubsub/docs/connect_kafka)  
Deployment of Kafka Connect is explicitly excluded by the requirements. So the only option available is A

   upvoted 4 times

  **samdhimal** 9 months ago

Option A: Deploy a Kafka cluster on GCE VM Instances. Configure your on-prem cluster to mirror your topics to the cluster running in GCE. Use a Dataproc cluster or Dataflow job to read from Kafka and write to GCS.

This option involves setting up a separate Kafka cluster in Google Cloud, and then configuring the on-prem cluster to mirror the topics to this cluster. The data from the Google Cloud Kafka cluster can then be read using either a Dataproc cluster or a Dataflow job and written to Cloud Storage for analysis in BigQuery.

   upvoted 3 times

  **samdhimal** 9 months ago

Option B: Deploy a Kafka cluster on GCE VM Instances with the Pub/Sub Kafka connector configured as a Sink connector. Use a Dataproc cluster or Dataflow job to read from Kafka and write to GCS.

This option is similar to Option A, but involves using the Pub/Sub Kafka connector as a sink connector instead of mirroring the topics from the on-prem cluster. This option would result in the same duplication of data and additional resources required as Option A, making it less desirable.

   upvoted 1 times

  **samdhimal** 9 months ago

Sorry. I messed up. The answer is probably A. My badd....



   upvoted 1 times

  **samdhimal** 9 months ago

Option D: Deploy the Pub/Sub Kafka connector to your on-prem Kafka cluster and configure Pub/Sub as a Sink connector. Use a Dataflow job to read from Pub/Sub and write to GCS.

This option involves deploying the Pub/Sub Kafka connector on the on-prem cluster, but configuring it as a sink connector. In this case, the data from the on-prem Kafka cluster would be sent directly to Pub/Sub, which would act as the final destination for the data. A Dataflow job would then be used to read the data from Pub/Sub and write it to Cloud Storage for analysis in BigQuery. This option would result in the data being stored in both the on-prem cluster and Pub/Sub, making it less desirable compared to option C, where the data is only stored in Pub/Sub as an intermediary between the on-prem cluster and Google Cloud.

   upvoted 1 times

  **musumusu** 8 months, 2 weeks ago

you use chatgpt replies, if you instruct chat gpt that you don't need to use plugins as per question say, it will answer A

   upvoted 1 times

  **samdhimal** 9 months ago

Option C: Deploy the Pub/Sub Kafka connector to your on-prem Kafka cluster and configure Pub/Sub as a Source connector. Use a Dataflow job to read from Pub/Sub and write to GCS.

This option involves deploying the Pub/Sub Kafka connector directly on the on-prem cluster, and configuring it as a source connector. The data from the on-prem Kafka cluster is then sent directly to Pub/Sub, which acts as an intermediary between the on-prem cluster and the data stored in Google Cloud. A Dataflow job is then used to read the data from Pub/Sub and write it to Cloud Storage for analysis in BigQuery. This option avoids the duplication of data and additional resources required by the other options, making it the preferred option.

   upvoted 2 times

  **zellick** 11 months ago

**Selected Answer: A**

A is the answer.

   upvoted 1 times

  **Afonya** 1 year ago

**Selected Answer: A**

"The preferred replication method is mirroring to avoid deployment of Kafka Connect plugins."

   upvoted 1 times

  **somnathmaddi** 1 year ago

D is the right answer

   upvoted 3 times

  **clauditis** 1 year, 1 month ago

D is the right answer

   upvoted 2 times

  **hendrixlives** 1 year, 10 months ago

**Selected Answer: A**

"A" is the answer which complies with the requirements (specifically, "The preferred replication method is mirroring to avoid deployment of Kafka Connect plugins"). Indeed, one of the uses of what is called "Geo-Replication" (or Cross-Cluster Data Mirroring) in Kafka is precisely cloud migrations: <https://kafka.apache.org/documentation/#georeplication>

However I agree with Ganshank, and the optimal "Google way" way would be "D", installing the Pub/Sub Kafka connector to move the data from on-prem to GCP.

   upvoted 6 times

  **gcp\_k** 2 years ago

Going with "D"

Refer: <https://stackoverflow.com/questions/55277188/kafka-to-google-pub-sub-using-sink-connector>

   upvoted 3 times

  **baubaumiaomiao** 1 year, 10 months ago

"avoid deployment of Kafka Connect plugins"

   upvoted 1 times

  **sumanshu** 2 years, 4 months ago

Vote for A

   upvoted 1 times

  **daghayeghi** 2 years, 7 months ago

Answer: A

Description: Question says mirroring to avoid kafka connect plugins

   upvoted 3 times

  **Allan222** 2 years, 8 months ago

Correct is D

   upvoted 1 times

  **sumanshu** 2 years, 3 months ago

As per question - "avoid deployment of Kafka Connect plugins."



   upvoted 1 times

  **vakati** 3 years, 1 month ago

A.

the best solution would be D but given the restriction here to use mirroring and avoid connectors, A would be the natural choice

   upvoted 3 times

  **Tanmoyk** 3 years, 1 month ago

D should be the correct answer. Configure pub/sub as sink

   upvoted 4 times

  **haroldbenites** 3 years, 2 months ago

C is correct.

<https://docs.confluent.io/current/connect/kafka-connect-gcp-pubsub/index.html>

   upvoted 2 times

  **haroldbenites** 3 years, 2 months ago

Correct Answer: D

Why is this correct?

You can connect Kafka to GCP by using a connector. The 'downstream' service (Pub/Sub) will use a sink connector.

   upvoted 1 times

  **sumanshu** 2 years, 4 months ago



Question says : avoid deployment of Kafka Connect plugins.

   upvoted 2 times

  **clauditis** 3 years, 2 months ago

its D, why would google prefer Kafka in their own cert questions! :)

   upvoted 3 times

  **Ral17** 2 years, 2 months ago

Because the questions mentions to avoid deployment of Kafka connect plugins

   upvoted 3 times

[Load full discussion...](#)



## Platform

> [Home](#)

> [Examtopics PRO](#)

> [All Exams](#)

> [Training Courses](#)



© 2024 ExamTopics