

🔗 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 268 DISCUSSION

Actual exam question from Google's Professional Data Engineer

Question #: 268

Topic #: 1

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

You created a new version of a Dataflow streaming data ingestion pipeline that reads from Pub/Sub and writes to BigQuery. The previous version of the pipeline that runs in production uses a 5-minute window for processing. You need to deploy the new version of the pipeline without losing any data, creating inconsistencies, or increasing the processing latency by more than 10 minutes. What should you do?

- A. Update the old pipeline with the new pipeline code.
- B. Snapshot the old pipeline, stop the old pipeline, and then start the new pipeline from the snapshot.
- C. Drain the old pipeline, then start the new pipeline.
- D. Cancel the old pipeline, then start the new pipeline.

[Show Suggested Answer](#)

by [scaenruy](#) at Jan. 3, 2024, 6:31 p.m.

Comments

Type your comment...

[Submit](#)

🗨️ [raaad](#) [Highly Voted](#) 1 year, 4 months ago

Selected Answer: C

- Graceful Data Transition: Draining the old pipeline ensures it processes all existing data in its buffers and watermarks before shutting down, preventing data loss or inconsistencies.

- Minimal Latency Increase: The latency increase will be limited to the amount of time it takes to drain the old pipeline, typically within the acceptable 10-minute threshold.

   upvoted 8 times

  **AlizCert** **Highly Voted**  1 year, 2 months ago

I don't think C is correct, as it will immediately fire the window:

"Draining can result in partially filled windows. In that case, if you restart the drained pipeline, the same window might fire a second time, which can cause issues with your data."

<https://cloud.google.com/dataflow/docs/guides/stopping-a-pipeline#effects>

Maybe "A" means launching a replacement job?


<https://cloud.google.com/dataflow/docs/guides/updating-a-pipeline#Launching>

   upvoted 6 times

  **SamuelTsch** 6 months, 1 week ago

we don't restart the drained pipeline.

   upvoted 1 times

  **d11379b** 1 year, 1 month ago

So why not B it is the better choice to save intermediate state and easy to use

   upvoted 2 times

  **petulda** **Most Recent**  5 months, 1 week ago

Why not B?

<https://cloud.google.com/dataflow/docs/guides/upgrade-guide#stop-and-replace>

   upvoted 1 times

  **STEVE_PEGLEG** 9 months ago

Selected Answer: C

There is requirement to avoid data loss.

<https://cloud.google.com/dataflow/docs/guides/upgrade-guide#stop-and-replace>

"To avoid data loss, in most cases, draining is the preferred action."

   upvoted 1 times

  **Ouss_123** 10 months, 3 weeks ago

Selected Answer: C

- Draining the old pipeline ensures that it finishes processing all in-flight data before stopping, which prevents data loss and inconsistencies.

- After draining, you can start the new pipeline, which will begin processing new data from where the old pipeline left off.

- This approach maintains a smooth transition between the old and new versions, minimizing latency increases and avoiding data gaps or overlaps.

==> Other options, such as updating, snapshotting, or canceling, might not provide the same level of consistency and could lead to data loss or increased latency beyond the acceptable 10-minute window. Draining is the safest method to ensure a seamless transition.

   upvoted 2 times

  **d11379b** 1 year, 1 month ago

Selected Answer: B

I would choose B as mentioned by Alizcert, a simple drain may cause problem

Dataflow snapshots save the state of a streaming pipeline, which lets you start a new version of your Dataflow job without losing state. Snapshots are useful for backup and recovery, testing and rolling back updates to streaming pipelines, and other similar scenarios.

   upvoted 2 times

  **hanoverquay** 1 year, 1 month ago

Selected Answer: C


C option



   upvoted 1 times

  **Matt_108** 1 year, 3 months ago

Selected Answer: C

Option C, draining the old pipeline solves all requests

   upvoted 1 times

  **scaenrui** 1 year, 4 months ago

Selected Answer: C

C. Drain the old pipeline, then start the new pipeline.

👍 ↩ 🚩 upvoted 2 times



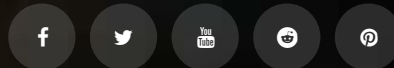
Platform

> [Home](#)

> [Examtopics PRO](#)

> [All Exams](#)

> [Training Courses](#)



© 2024 ExamTopics