

[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 280 DISCUSSION

Actual exam question from Google's Professional Data Engineer

Question #: 280

Topic #: 1

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

You are running a streaming pipeline with Dataflow and are using hopping windows to group the data as the data arrives. You noticed that some data is arriving late but is not being marked as late data, which is resulting in inaccurate aggregations downstream. You need to find a solution that allows you to capture the late data in the appropriate window. What should you do?

- A. Use watermarks to define the expected data arrival window. Allow late data as it arrives.
- B. Change your windowing function to tumbling windows to avoid overlapping window periods.
- C. Change your windowing function to session windows to define your windows based on certain activity.
- D. Expand your hopping window so that the late data has more time to arrive within the grouping.

[Show Suggested Answer](#)

by [scaenruy](#) at Jan. 4, 2024, 5:32 a.m.

Comments

Type your comment...

[Submit](#)

[raaad](#) [Highly Voted](#) 10 months ago

Selected Answer: A

- Watermarks: Watermarks in a streaming pipeline are used to specify the point in time when Dataflow expects all data up to that point to have arrived.
- Allow Late Data: configure the pipeline to accept and correctly process data that arrives after the watermark, ensuring it's captured in the appropriate window.

👍 ↩ 🚩 upvoted 7 times

📄 👤 **Pime13** Most Recent 3 months, 4 weeks ago

Selected Answer: A

<https://cloud.google.com/dataflow/docs/concepts/streaming-pipelines#watermarks>

A watermark is a threshold that indicates when Dataflow expects all of the data in a window to have arrived. If the watermark has progressed past the end of the window and new data arrives with a timestamp within the window, the data is considered late data. For more information, see Watermarks and late data in the Apache Beam documentation.

Dataflow tracks watermarks because of the following reasons:

Data is not guaranteed to arrive in time order or at predictable intervals.

Data events are not guaranteed to appear in pipelines in the same order that they were generated.

👍 ↩ 🚩 upvoted 1 times

📄 👤 **m_a_p_s** 4 months, 3 weeks ago

Selected Answer: A

A - <https://cloud.google.com/dataflow/docs/concepts/streaming-pipelines#watermarks>

👍 ↩ 🚩 upvoted 1 times

📄 👤 **JyoGCP** 8 months, 2 weeks ago

Selected Answer: A

Option A

👍 ↩ 🚩 upvoted 1 times

📄 👤 **Matt_108** 9 months, 3 weeks ago

Selected Answer: A

Option A - <https://cloud.google.com/dataflow/docs/concepts/streaming-pipelines#watermarks>

👍 ↩ 🚩 upvoted 3 times

📄 👤 **Sofia98** 10 months ago

Selected Answer: A

<https://cloud.google.com/dataflow/docs/concepts/streaming-pipelines#watermarks>

👍 ↩ 🚩 upvoted 2 times

📄 👤 **scaenruy** 10 months ago

Selected Answer: A

A. Use watermarks to define the expected data arrival window. Allow late data as it arrives.

👍 ↩ 🚩 upvoted 1 times



Platform

> Home

> Examtopics PRO

> All Exams

> Training Courses



