

[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 172 DISCUSSION

Actual exam question from Google's Professional Data Engineer

Question #: 172

Topic #: 1

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

You are analyzing the price of a company's stock. Every 5 seconds, you need to compute a moving average of the past 30 seconds' worth of data. You are reading data from Pub/Sub and using DataFlow to conduct the analysis. How should you set up your windowed pipeline?




- A. Use a fixed window with a duration of 5 seconds. Emit results by setting the following trigger:  
`AfterProcessingTime.pastFirstElementInPane().plusDelayOf (Duration.standardSeconds(30))`
- B. Use a fixed window with a duration of 30 seconds. Emit results by setting the following trigger:  
`AfterWatermark.pastEndOfWindow().plusDelayOf (Duration.standardSeconds(5))`
- C. Use a sliding window with a duration of 5 seconds. Emit results by setting the following trigger:  
`AfterProcessingTime.pastFirstElementInPane().plusDelayOf (Duration.standardSeconds(30))`
- D. Use a sliding window with a duration of 30 seconds and a period of 5 seconds. Emit results by setting the following trigger:  
`AfterWatermark.pastEndOfWindow ()`

[Show Suggested Answer](#)

by [AWSandeep](#) at *Sept. 2, 2022, 7:48 p.m.*

### Comments

Type your comment...

  **vamgcp** Highly Voted  1 year, 9 months ago

**Selected Answer: D**

Option D: Sliding Window: Since you need to compute a moving average of the past 30 seconds' worth of data every 5 seconds, a sliding window is appropriate. A sliding window allows overlapping intervals and is well-suited for computing rolling aggregates.

Window Duration: The window duration should be set to 30 seconds to cover the required 30 seconds' worth of data for the moving average calculation.

Window Period: The window period or sliding interval should be set to 5 seconds to move the window every 5 seconds and recalculate the moving average with the latest data.

Trigger: The trigger should be set to `AfterWatermark.pastEndOfWindow()` to emit the computed moving average results when the watermark advances past the end of the window. This ensures that all data within the window is considered before emitting the result.

   upvoted 9 times




  **AWSandeep** Highly Voted  2 years, 8 months ago

**Selected Answer: D**

D. Use a sliding window with a duration of 30 seconds and a period of 5 seconds. Emit results by setting the following trigger: `AfterWatermark.pastEndOfWindow()`

Reveal Solution

   upvoted 7 times

  **Anudeep58** Most Recent  10 months, 3 weeks ago

**Selected Answer: D**

Option D is the correct configuration because it uses a sliding window of 30 seconds with a period of 5 seconds, ensuring that the moving average is computed every 5 seconds based on the past 30 seconds of data. The trigger `AfterWatermark.pastEndOfWindow()` ensures timely and accurate results are emitted as the watermark progresses.

   upvoted 1 times

  **Kimich** 1 year, 5 months ago

`AfterWatermark` is an essential triggering condition in Dataflow that allows computations to be triggered based on event time rather than processing time. Then eliminate A&C. Comparing B&D, B will generate outcome every 30 seconds which is not what we want

D. Using a sliding window with a duration of 30 seconds and a period of 5 seconds, and setting the trigger as `AfterWatermark.pastEndOfWindow()`, is a sliding window that generates results every 5 seconds, and each result includes data from the past 30 seconds. In other words, every 5 seconds, you get the average value of the most recent 30 seconds' data, and there is a 5-second overlap between these windows. This is what we want.

   upvoted 2 times

  **zellck** 2 years, 5 months ago

**Selected Answer: D**

D is the answer.

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

You set the following windows with the Apache Beam SDK or Dataflow SQL streaming extensions:

Hopping windows (called sliding windows in Apache Beam)

A hopping window represents a consistent time interval in the data stream. Hopping windows can overlap, whereas tumbling windows are disjoint.

For example, a hopping window can start every thirty seconds and capture one minute of data. The frequency with which hopping windows begin is called the period. This example has a one-minute window and thirty-second period.

   upvoted 2 times

  **pluiedust** 2 years, 7 months ago

**Selected Answer: D**

Moving average —> sliding window

   upvoted 4 times



## Platform

> [Home](#)

> [Examtopics PRO](#)

> [All Exams](#)

> [Training Courses](#)



© 2024 ExamTopics