○ **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 153 DISCUSSION

Actual exam question from Google's Professional Data Engineer

Question #: 153

Topic #: 1

[All Professional Data Engineer Questions]

You operate an IoT pipeline built around Apache Kafka that normally receives around 5000 messages per second. You want to use Google Cloud Platform to create an alert as soon as the moving average over 1 hour drops below 4000 messages per second. What should you do?

A. Consume the stream of data in Dataflow using Kafka IO. Set a sliding time window of 1 hour every 5 minutes. Compute the average when the window closes, and send an alert if the average is less than 4000 messages.

B. Consume the stream of data in Dataflow using Kafka IO. Set a fixed time window of 1 hour. Compute the average when the window closes, and send an alert if the average is less than 4000 messages.

C. Use Kafka Connect to link your Kafka message queue to Pub/Sub. Use a Dataflow template to write your messages from Pub/Sub to Bigtable. Use Cloud Scheduler to run a script every hour that counts the number of rows created in Bigtable in the last hour. If that number falls below 4000, send an alert.

D. Use Kafka Connect to link your Kafka message queue to Pub/Sub. Use a Dataflow template to write your messages from Pub/Sub to BigQuery. Use Cloud Scheduler to run a script every five minutes that counts the number of rows created in BigQuery in the last hour. If that number falls below 4000, send an alert.

Show Suggested Answer

by [deleted] at *March 22, 2020, 8:12 a.m.*

Comments

Type your comment...

Submit

⊟ 👤 **[Removed]** `Highly Voted 👍` **4 years, 7 months ago**
Should be A
👍 ↩ 🚩 upvoted 27 times

⊟ 👤 **[Removed]** `Highly Voted 👍` **4 years, 7 months ago**
Correct: A

Dataflow can connect with Kafka and sliding window is used for taking averages
👍 ↩ 🚩 upvoted 17 times

⊟ 👤 **mothkuri** `Most Recent ⊘` **8 months ago**
`Selected Answer: A`
Option A is correct answer.
Option B is not correct. There could be a chance middle of 1st window to middle of 2nd window less messages(i.e > 4000).
Option C & D out of scope.
👍 ↩ 🚩 upvoted 2 times

⊟ 👤 **barnac1es** **1 year, 1 month ago**
`Selected Answer: A`
Dataflow with Sliding Time Windows: Dataflow allows you to work with event-time windows, making it suitable for time-series data like incoming IoT messages. Using sliding windows every 5 minutes allows you to compute moving averages efficiently.

Sliding Time Window: The sliding time window of 1 hour every 5 minutes enables you to calculate the moving average over the specified time frame.

Computing Averages: You can efficiently compute the average when each sliding window closes. This approach ensures that you have real-time visibility into the message rate and can detect deviations from the expected rate.

Alerting: When the calculated average drops below 4000 messages per second, you can trigger an alert from within the Dataflow pipeline, sending it to your desired alerting mechanism, such as Cloud Monitoring, Pub/Sub, or another notification service.

Scalability: Dataflow can scale automatically based on the incoming data volume, ensuring that you can handle the expected rate of 5000 messages per second.
👍 ↩ 🚩 upvoted 2 times

⊟ 👤 **vamgcp** **1 year, 3 months ago**
`Selected Answer: A`
Option A

Pros:

This option is relatively simple to implement.
It can be used to compute the moving average over any time window.
Cons:

This option can be computationally expensive, especially if the data stream is large.
It can be difficult to troubleshoot if the alert does not fire when it is supposed to.
👍 ↩ 🚩 upvoted 2 times

⊟ 👤 **vaga1** **1 year, 5 months ago**
`Selected Answer: A`
the correct answer is between A and B since it doesn't make sense to use Pub/Sub combined with Kafka. To have a Moving Average then we should go for A, updating the average estimation every 5 minutes using the new data that came in and eliminating the "most far" 5 minutes.
👍 ↩ 🚩 upvoted 2 times

⊟ 👤 **zellck** **1 year, 11 months ago**
`Selected Answer: A`
A is the answer.

https://cloud.google.com/dataflow/docs/concepts/streaming-pipelines#windows
Windowing functions divide unbounded collections into logical components, or windows. Windowing functions group unbounded collections by the timestamps of the individual elements. Each window contains a finite number of elements.

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 4 times

☐ 👤 **medeis_jar** 2 years, 10 months ago

Selected Answer: A

as explained by Alasmindas

👍 ↩ 🏳 upvoted 2 times

☐ 👤 **AACHB** 2 years, 10 months ago

Selected Answer: A

Correct Answer: A

👍 ↩ 🏳 upvoted 2 times

☐ 👤 **JG123** 2 years, 11 months ago

Correct: A

👍 ↩ 🏳 upvoted 1 times

☐ 👤 **Chelseajcole** 3 years, 1 month ago

A is enough

👍 ↩ 🏳 upvoted 1 times

☐ 👤 **daghayeghi** 3 years, 8 months ago

A:
the correct answer is between A and B, But because used "Moving Average" then we should go for A.

👍 ↩ 🏳 upvoted 2 times

☐ 👤 **apnu** 3 years, 10 months ago

yes , using KafkaIO , we can connect to Kafka cluster.

👍 ↩ 🏳 upvoted 2 times

☐ 👤 **ashuchip** 3 years, 10 months ago

yes A is correct , because sliding window can only help here.

👍 ↩ 🏳 upvoted 3 times

☐ 👤 **Alasmindas** 3 years, 11 months ago

Option A is the correct answer. Reasons:-
a) Kafka IO and Dataflow is a valid option for interconnect (needless where Kafka is located - On Prem/Google Cloud/Other cloud)
b) Sliding Window will help to calculate average.

Option C and D are overkill and complex, considering the scenario in the question,
https://cloud.google.com/solutions/processing-messages-from-kafka-hosted-outside-gcp

👍 ↩ 🏳 upvoted 7 times

☐ 👤 **Alasmindas** 3 years, 11 months ago

Option A is the correct answer. Reasons:-
a) Kafka IO and Dataflow is a valid option for interconnect (needless where Kafka is located - On Prem/Google Cloud/Other cloud)
b) Sliding Window will help to calculate average.

Option C and D are overkill and complex, considering the scenario in the question,

👍 ↩ 🏳 upvoted 6 times

☐ 👤 **atnafu2020** 4 years, 2 months ago

A
To take running averages of data, use hopping windows. You can use one-minute hopping windows with a thirty-second period to compute a one-minute running average every thirty seconds.

👍 ↩ 🏳 upvoted 2 times

Load full discussion…

EXAMTOPICS

Platform

> Home

> Examtopics PRO

> All Exams

> Training Courses