EXAM PROFESSIONAL DATA ENGINEER TOPIC 1 QUESTION 11 DISCUSSION

Actual exam question from Google's Professional Data Engineer
Question #: 11
Topic #: 1
[All Professional Data Engineer Questions]

You are designing a basket abandonment system for an ecommerce company. The system will send a message to a user based on these rules:
✑ No interaction by the user on the site for 1 hour
Has added more than $30 worth of products to the basket
▪
✑ Has not completed a transaction
You use Google Cloud Dataflow to process the data and decide if a message should be sent. How should you design the pipeline?

A. Use a fixed-time window with a duration of 60 minutes.

B. Use a sliding time window with a duration of 60 minutes.

C. Use a session window with a gap time duration of 60 minutes.

D. Use a global window with a time based trigger with a delay of 60 minutes.

Show Suggested Answer

by AWSandeep at *Sept. 3, 2022, 6:50 a.m.*

Comments

Type your comment...

Submit

vetaal Highly Voted 2 years, 6 months ago

There are 3 windowing concepts in dataflow and each can be used for below use case
1) Fixed window
2) Sliding window and
3) Session window.

Fixed window = any aggregation use cases, any batch analysis of data, relatively simple use cases.

Sliding window = Moving averages of data
Session window = user session data, click data and real time gaming analysis.

The question here is about user session data and hence session window.

Reference:
https://cloud.google.com/dataflow/docs/concepts/streaming-pipelines

upvoted 33 times

cqrm3n Most Recent 3 months, 2 weeks ago

Selected Answer: C

The answer is C because session window is specifically designed to handle use cases where activity is grouped by gaps.

A. Fixed-time window divides data into non-overlapping, equally-size intervals but do not track gaps in user activity.
B. Sliding-time window process overlapping intervals and are better suited for periodic aggregation.
D. Global windows process all data over the pipeline's lifetime and rely on custom triggers to handle time-based logic. It is technically possible but unnecessarily complex so no.

upvoted 3 times

rtcpost 7 months, 1 week ago

Selected Answer: C

C. Use a session window with a gap time duration of 60 minutes.

A session window with a gap time duration of 60 minutes is appropriate for capturing user sessions where there has been no interaction on the site for 1 hour. It allows you to group user activity within a session, and when the session becomes inactive for the defined gap time, you can evaluate whether the user added more than $30 worth of products to the basket and has not completed a transaction.

Options A and B (fixed-time window and sliding time window) might not capture the specific session-based criteria of inactivity and user interaction effectively.

Option D (global window with a time-based trigger) is not suitable for capturing user sessions and checking inactivity based on a specific time duration. It's more appropriate for cases where you need a single global view of the data.

upvoted 3 times

RT_G 1 year, 5 months ago

Selected Answer: C

Session window since the question specifically talks about a specific user for a fixed duration.

upvoted 1 times

rocky48 1 year, 6 months ago

Selected Answer: C

Session window = user session data, click data and real time gaming analysis.

upvoted 1 times

imran79 1 year, 7 months ago

The basket abandonment system needs to determine if a user hasn't interacted with the site for 1 hour, has added products worth more than $30, and hasn't completed a transaction. Therefore, the pipeline should account for periods of user activity and inactivity. A session-based windowing approach is appropriate here.

The right choice is:

C. Use a session window with a gap time duration of 60 minutes.

Session windows group data based on periods of activity and inactivity. If there's no interaction for the duration of the gap time (in this case, 60 minutes), a new window is started. This would help identify users who haven't interacted with the site for the specified duration, fulfilling the requirement for the basket abandonment system.

upvoted 2 times

MikkelRev 1 year, 7 months ago

Selected Answer: C

session windows can divide a data stream representing user activity
https://cloud.google.com/dataflow/docs/concepts/streaming-pipelines#session-windows

upvoted 1 times

Chesternut999 2 years, 1 month ago

Selected Answer: C

C - The best option for this use case.

upvoted 2 times

bha11111 2 years, 1 month ago

Selected Answer: C

Session window is used for these type of scenario

upvoted 2 times

samdhimal 2 years, 2 months ago

C. Use a session window with a gap time duration of 60 minutes.

A session window would be the most appropriate option to use in this case, as it would allow you to group events into sessions based on time gaps. In this case, the gap time of 60 minutes could be used to define a session, and if there is no interaction from the user for 60 minutes, a new session would be created. By using a session window, you can track the behavior of the user during each session, including the products added to the basket, and determine if the conditions for sending a message have been met (i.e., the user has added more than $30 worth of products to the basket and has not completed a transaction).

upvoted 2 times

kennyloo 2 years, 6 months ago

Only C is feasible for this question

upvoted 1 times

AWSandeep 2 years, 8 months ago

Selected Answer: C

C. Use a session window with a gap time duration of 60 minutes.

upvoted 1 times