← **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 20 DISCUSSION**

Actual exam question from Google's Professional Data Engineer

Question #: 20

Topic #: 1

[All Professional Data Engineer Questions]

You work for a car manufacturer and have set up a data pipeline using Google Cloud Pub/Sub to capture anomalous sensor events. You are using a push subscription in Cloud Pub/Sub that calls a custom HTTPS endpoint that you have created to take action of these anomalous events as they occur. Your custom HTTPS endpoint keeps getting an inordinate amount of duplicate messages. What is the most likely cause of these duplicate messages?

- A. The message body for the sensor event is too large.

- B. Your custom endpoint has an out-of-date SSL certificate.

- C. The Cloud Pub/Sub topic has too many messages published to it.

- D. Your custom endpoint is not acknowledging messages within the acknowledgement deadline.

Show Suggested Answer

by 👤 **jvg637** at *March 11, 2020, 6:37 p.m.*

Comments

Type your comment...

Submit

☐ 👤 **jvg637** `Highly Voted 👍` 5 years, 1 month ago

The Answer should be D. The custom endpoint is not acknowledging the message, that is the reason for Pub/Sub to send the message again and again. Not B.

👍 ↩ 🏳 upvoted 88 times

☐ 👤 **MauryaSushil** `Highly Voted 👍` 5 years, 1 month ago

D : Doubt should be only between B & D. But B is not possible because if SSL is expired then endpoint URL will not receive any messages forget about duplicates. So It should be D for Duplicates.

👍 ↩ 🏳 upvoted 10 times

☐ 👤 **atnafu2020** `Most Recent ⊙` 7 months, 1 week ago

D
Why are there too many duplicate messages?

Pub/Sub guarantees at-least-once message delivery, which means that occasional duplicates are to be expected. However, a high rate of duplicates may indicate that the client is not acknowledging messages within the configured ack_deadline_seconds, and Pub/Sub is retrying the message delivery. This can be observed in the monitoring metrics pubsub.googleapis.com/subscription/pull_ack_message_operation_count for pull subscriptions, and pubsub.googleapis.com/subscription/push_request_count for push subscriptions. Look for elevated expired or webhook_timeout values in the /response_code. This is particularly likely if there are many small messages, since Pub/Sub may batch messages internally and a partially acknowledged batch will be fully redelivered.

👍 ↩ 🏳 upvoted 5 times

☐ 👤 **ganesh2121** 7 months, 1 week ago

D is correct
As per google docs- When you do not acknowledge a message before its acknowledgement deadline has expired, Pub/Sub resends the message. As a result, Pub/Sub can send duplicate messages. Use Google Cloud's operations suite to monitor acknowledge operations with the expired response code to detect this condition

👍 ↩ 🏳 upvoted 3 times

☐ 👤 **Radhika7983** 7 months, 1 week ago

The correct answer is D. Look for the link
https://cloud.google.com/pubsub/docs/faq

Why are there too many duplicate messages?
Pub/Sub guarantees at-least-once message delivery, which means that occasional duplicates are to be expected. However, a high rate of duplicates may indicate that the client is not acknowledging messages within the configured ack_deadline_seconds, and Pub/Sub is retrying the message delivery. This can be observed in the monitoring metrics pubsub.googleapis.com/subscription/pull_ack_message_operation_count for pull subscriptions, and pubsub.googleapis.com/subscription/push_request_count for push subscriptions. Look for elevated expired or webhook_timeout values in the /response_code. This is particularly likely if there are many small messages, since Pub/Sub may batch messages internally and a partially acknowledged batch will be fully redelivered.

Another possibility is that the subscriber is not acknowledging some messages because the code path processing those specific messages fails, and the Acknowledge call is never made; or the push endpoint never responds or responds with an error.

👍 ↩ 🏳 upvoted 6 times

☐ 👤 **MaxNRG** 7 months, 1 week ago

D as the Cloud Pub/Sub will deliver duplicate messages only if there has been no acknowledgement from the subscriber.
Refer GCP documentation - Cloud Pub/Sub FAQs - Duplicates:
https://cloud.google.com/pubsub/docs/faq#duplicates
Why are there too many duplicate messages?
Cloud Pub/Sub guarantees at-least-once message delivery, which means that occasional duplicates are to be expected. However, a high rate of duplicates may indicate that the client is not acknowledging messages within the configured ack_deadline_seconds, and Cloud Pub/Sub is retrying the message delivery.

👍 ↩ 🏳 upvoted 2 times

☐ 👤 **crismo04** 7 months, 1 week ago

`Selected Answer: D`

If the answer had been B the messages had not reached the HTTP page since the push subscription requires an SSL certificate (https://cloud.google.com/pubsub/docs/push#:~:text=Endpoint%20URL%20(required)), so the answer must be D.

I think it is due to a bad response due to overcharge in the HTTPS page, which is sending a non correct status code (https://cloud.google.com/pubsub/docs/push#:~:text=VPC%20Service%20Controls.-,Receive%20messages,-When%20Pub/Sub)

👍 ↩ 🏳 upvoted 1 times

☐ 👤 **rtcpost** 7 months, 1 week ago

`Selected Answer: D`

In Google Cloud Pub/Sub, when you use a push subscription, messages are delivered to the specified endpoint (in this case, your custom HTTPS endpoint). The acknowledgment deadline is the time given to the endpoint to acknowledge that it has

your custom HTTPS endpoint). The acknowledgment deadline is the time given to the endpoint to acknowledge that it has received and processed the message. If the acknowledgment is not received within the deadline, Pub/Sub may consider the message as unacknowledged and may attempt redelivery, which can lead to duplicate messages.

You should ensure that your custom HTTPS endpoint acknowledges messages within the acknowledgment deadline to prevent duplicate messages from being sent. Additionally, it's essential to handle messages in an idempotent way, so even if duplicates do occur, the action taken by your endpoint doesn't have unintended consequences.

👍 ↩ ⚑ upvoted 2 times

⊟ 👤 **petergjohnson** 7 months, 1 week ago

Selected Answer: B

After re-reading the question it seem to me that it is asking for a root cause. It is possible that the most common cause of this symptom is and expired certificate. Once expired duplicates would be received for every message.

👍 ↩ ⚑ upvoted 1 times

⊟ 👤 **VictorBa** 11 months, 2 weeks ago

Selected Answer: D

Agree with previous explanations regarding validity of D

👍 ↩ ⚑ upvoted 1 times

⊟ 👤 **searching4alicense** 1 year, 2 months ago

D - If a message has not been acknowledged within its acknowledgement deadline, Dataflow attempts to maintain the lease on the message by repeatedly extending the acknowledgement deadline to prevent redelivery from Pub/Sub. However this is best effort and there is a possibility that messages may be redelivered. This can be monitored using metrics listed here. https://cloud.google.com/blog/products/data-analytics/handling-duplicate-data-in-streaming-pipeline-using-pubsub-dataflow

👍 ↩ ⚑ upvoted 1 times

⊟ 👤 **philli1011** 1 year, 3 months ago

D should be the answer. If acknowlegement is not received back to pub/sub , pub/sub may resend meassages.

👍 ↩ ⚑ upvoted 1 times

⊟ 👤 **imran79** 1 year, 7 months ago

The correct answer is:
D. Your custom endpoint is not acknowledging messages within the acknowledgement deadline.

👍 ↩ ⚑ upvoted 2 times

⊟ 👤 **emmylou** 1 year, 7 months ago

if there were an out of date certificate then nothing would get through. D

👍 ↩ ⚑ upvoted 1 times

⊟ 👤 **FP77** 1 year, 9 months ago

Selected Answer: D

It should be D
https://cloud.google.com/pubsub/docs/troubleshooting#dupes

👍 ↩ ⚑ upvoted 3 times

⊟ 👤 **dgteixeira** 1 year, 11 months ago

Selected Answer: D

The correct answer is D, because it's how Pub/Sub works.
Documentation here: https://cloud.google.com/pubsub/docs/troubleshooting#dupes

👍 ↩ ⚑ upvoted 3 times

⊟ 👤 **boca_2022** 2 years ago

Selected Answer: D

D for sure

👍 ↩ ⚑ upvoted 2 times

Load full discussion...