

🔗 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 112 DISCUSSION

Actual exam question from Google's Professional Data Engineer

Question #: 112

Topic #: 1

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

You operate a logistics company, and you want to improve event delivery reliability for vehicle-based sensors. You operate small data centers around the world to capture these events, but leased lines that provide connectivity from your event collection infrastructure to your event processing infrastructure are unreliable, with unpredictable latency. You want to address this issue in the most cost-effective way. What should you do?

- A. Deploy small Kafka clusters in your data centers to buffer events.
- B. Have the data acquisition devices publish data to Cloud Pub/Sub.
- C. Establish a Cloud Interconnect between all remote data centers and Google.
- D. Write a Cloud Dataflow pipeline that aggregates all data in session windows.

Show Suggested Answer

by [deleted] at March 22, 2020, 1:58 p.m.

Comments




Type your comment...

Submit

🗨️ 👤 [Removed] Highly Voted 5 years, 1 month ago

Should be B

   upvoted 31 times

  **Ganshank** Highly Voted  5 years ago

C.

This is a tricky one. The issue here is the unreliable connection between data collection and data processing infrastructure, and to resolve it in a cost-effective manner. However, it also mentions that the company is using leased lines. I think replacing the leased lines with Cloud InterConnect would solve the problem, and hopefully not be an added expense. <https://cloud.google.com/interconnect/docs/concepts/overview>

   upvoted 22 times

  **serg3d** 4 years, 11 months ago


Yea, this would definitely solve the issue, but it's not "the most cost-effective way". I think PubSub is the correct answer.

   upvoted 7 times

  **snamburi3** 4 years, 5 months ago

the question also talks about a cost effective way...

   upvoted 3 times

  **sh2020** 4 years, 10 months ago

I agree, C is the only choice that addresses the problem. The problem is caused by leased line. How come pub/sub service can resolve it? Pub/sub will still use the leased line




   upvoted 5 times

  **awssp12345** 3 years, 10 months ago

DEFINITELY NOT COST EFFECT. C IS THE WORST CHOICE.

   upvoted 7 times

[Load full discussion...](#)

  **grshankar9** Most Recent  3 months, 2 weeks ago

Selected Answer: B

The data acquisition devices are spread across the globe. How many cloud interconnects would this require? Option C is definitely not cost effective. Option B seems to make sense as it suggests publishing from the acquisition device to Pub/Sub and not from the collection infrastructure. The problem clearly stated the connection from collection infrastructure to the processing infrastructure was unreliable.

   upvoted 1 times

  **Anudeep58** 10 months, 3 weeks ago

Selected Answer: B

Option B: Have the data acquisition devices publish data to Cloud Pub/Sub.

Rationale:


Managed Service: Cloud Pub/Sub is a fully managed service, reducing the operational overhead compared to managing Kafka clusters.

Reliability and Scalability: Cloud Pub/Sub can handle high volumes of data with low latency and provides built-in mechanisms for reliable message delivery, even in the face of intermittent connectivity.

Cost-Effective: Cloud Pub/Sub offers a pay-as-you-go pricing model, which can be more cost-effective than setting up and maintaining dedicated network infrastructure like Cloud Interconnect.

Global Availability: Cloud Pub/Sub is available globally and can handle data from multiple regions efficiently.

   upvoted 1 times



  **Nandababy** 1 year, 4 months ago

Even with Cloud Pub/Sub, unpredictable latency or delays could still occur due to the unreliable leased lines connecting your event collection infrastructure and event processing infrastructure. While Cloud Pub/Sub offers reliable message delivery within its own network, the handoff to your processing infrastructure is still dependent on the leased lines.

Replacing leased lines with Cloud Interconnect could potentially resolve the overall issue of unpredictable latency in event processing pipeline but it could be unnecessary expense provided data centers distributed world wide.

Cloud Pub/Sub along with other optimization techniques like Cloud VPN or edge computing might be sufficient.

   upvoted 1 times

  **FP77** 1 year, 8 months ago




Selected Answer: C

I don't know why B is the most voted. The issue here is unreliable connectivity and C is the perfect use-case for that

   upvoted 1 times

  **NeoNitin** 1 year, 9 months ago

its says with unpredictable latency and here no need to worry about connection
So B is the right one

   upvoted 1 times

🗨️ **ZZHZZH** 1 year, 10 months ago

Selected Answer: C

The question is misleading. But should be C since it addresses the unpredictability and latency directly.

👍 🔄 🚩 upvoted 1 times

🗨️ **musumusu** 2 years, 2 months ago

Best answer is A, By using Kafka, you can buffer the events in the data centers until a reliable connection is established with the event processing infrastructure.

But go with B, its google asking :P

👍 🔄 🚩 upvoted 2 times

🗨️ **musumusu** 2 years, 2 months ago

I read this question again, I wanna answer C. Buying Data acquisition devices and set them up with sensor, i dont think its practical approach. Imagine, Aduino is cheapest IOT available in market for 15 dollars, but who will open the sensor box and install it .. omg,, its a big job. This question depends if IOT devices that are attached to sensor needs to be programmed. Big Headache right. Use google cloud connect to deal with current situation. Or reprogramme IOT if they have connected with sensors.

👍 🔄 🚩 upvoted 1 times

🗨️ **ayush_1995** 2 years, 3 months ago

Selected Answer: B

B. Have the data acquisition devices publish data to Cloud Pub/Sub. This would provide a reliable messaging service for your event data, allowing you to ingest and process your data in a timely manner, regardless of the reliability of the leased lines. Cloud Pub/Sub also offers automatic retries and fault-tolerance, which would further improve the reliability of your event delivery. Additionally, using Cloud Pub/Sub would allow you to easily scale up or down your event processing infrastructure as needed, which would help to minimize costs.

👍 🔄 🚩 upvoted 10 times

🗨️ **desertlotus1211** 2 years, 3 months ago

Are they talking about GCP in this question?

Where is the event processing infrastructure?

Answer A, might be correct!

👍 🔄 🚩 upvoted 2 times

🗨️ **PrashantGupta1616** 2 years, 4 months ago

Selected Answer: B

pub/sub is region is a global service

It's important to note that the term "global" in this context refers to the geographical scope of the service

👍 🔄 🚩 upvoted 1 times

🗨️ **NicolasN** 2 years, 4 months ago

Selected Answer: A

As usual the answer is hidden somewhere in the Google Cloud Blog:

"In the case of our automotive company, the data is already stored and processed in local data centers in different regions. This happens by streaming all sensor data from the cars via MQTT to local Kafka Clusters that leverage Confluent's MQTT Proxy."

"This integration from devices to a local Kafka cluster typically is its own standalone project, because you need to handle IoT-specific challenges like constrained devices and unreliable networks."

<https://cloud.google.com/blog/products/ai-machine-learning/enabling-connected-transformation-with-apache-kafka-and-tensorflow-on-google-cloud-platform>

👍 🔄 🚩 upvoted 3 times

🗨️ **desertlotus1211** 2 years, 3 months ago

The question is asking from the on-premise infrastructure, which already has the data, to the event processing infrastructure, which is in the GCP, is unreliable....

it not asking from the sensors to the on-premise...

👍 🔄 🚩 upvoted 2 times

🗨️ **desertlotus1211** 2 years, 3 months ago

I might have to retract my answer... Are they talking about GCP in this question?

where is the event processing infrastructure?

👍 🔄 🚩 upvoted 1 times

🗨️ **zelck** 2 years, 5 months ago

Selected Answer: B

B is the answer.

👍 ↩ 🚩 upvoted 2 times

🗨️ 👤 **AzureDP900** 2 years, 4 months ago

yes it is B. Have the data acquisition devices publish data to Cloud Pub/Sub.

👍 ↩ 🚩 upvoted 1 times

🗨️ 👤 **piotrpiskorski** 2 years, 5 months ago

yeah, changing whole architecture around the world for the use of pub/sub is so much more cost efficient than Cloud Interconnect (which is like 3k\$)..

It's C.

👍 ↩ 🚩 upvoted 1 times

🗨️ 👤 **odacir** 2 years, 4 months ago

It's not a Cloud Interconnect, it's a lot of interconnect ones per data center, PUB/SUB addresses all the requirements. Its B

👍 ↩ 🚩 upvoted 1 times

🗨️ 👤 **odacir** 2 years, 4 months ago

ALSO, the problem it's not your connection, its the connectivity BT your event collection infrastructure to your event processing infrastructure, so PUSUB it's perfect for this

👍 ↩ 🚩 upvoted 1 times

🗨️ 👤 **jkhong** 2 years, 4 months ago

Wouldn't using cloud interconnect also result in amendments to each of the data center around the world? I don't see why there would be a huge architecture change when using PubSub, the publishers would just need to push messages directly to pubsub, instead of pushing to their own cost center.

Also, if the script for pushing messages can be standardised, the data centers can share it around to

👍 ↩ 🚩 upvoted 1 times

🗨️ 👤 **TNT87** 2 years, 7 months ago

Selected Answer: B

Cloud Pub/Sub, it supports batch & streaming , push and pull capabilities
Answer B

👍 ↩ 🚩 upvoted 1 times

🗨️ 👤 **t11** 2 years, 8 months ago

It has to be B.

👍 ↩ 🚩 upvoted 1 times

[Load full discussion...](#)



Platform

> Home

> All Exams

> Examtopics PRO

> Training Courses



