EXAMTOPICS

- Expert Verified, Online, **Free.**

☰ MENU

🔍

⬅ Google Discussions

Exam Professional Machine Learning Engineer All Questions

View all questions & answers for the Professional Machine Learning Engineer exam

Go to Exam

📄 **EXAM PROFESSIONAL MACHINE LEARNING ENGINEER TOPIC 1 QUESTION 113 DISCUSSI...**

Actual exam question from Google's Professional Machine Learning Engineer

Question #: 113

Topic #: 1

[All Professional Machine Learning Engineer Questions]

You need to analyze user activity data from your company's mobile applications. Your team will use BigQuery for data analysis, transformation, and experimentation with ML algorithms. You need to ensure real-time ingestion of the user activity data into BigQuery. What should you do?

A. Configure Pub/Sub to stream the data into BigQuery.

B. Run an Apache Spark streaming job on Dataproc to ingest the data into BigQuery.

C. Run a Dataflow streaming job to ingest the data into BigQuery.

D. Configure Pub/Sub and a Dataflow streaming job to ingest the data into BigQuery,

Show Suggested Answer

by 👤 **mymy9418** at *Dec. 18, 2022, 2:27 a.m.*

Comments

Type your comment...

Submit

➖ 👤 **pshemol** `Highly Voted 👍` 1 year, 10 months ago

Previously Google pattern was Pub/Sub -> Dataflow -> BQ
but now it looks as there is new Pub/Sub -> BQ
https://cloud.google.com/blog/products/data-analytics/pub-sub-launches-direct-path-to-bigquery-for-streaming-analytics

👍 ↩ 🚩 upvoted 19 times

☐ 👤 **TNT87** 1 year, 7 months ago

New pub sub??? heheheh

👍 ↩ 🚩 upvoted 1 times

☐ 👤 **TNT87** 1 year, 7 months ago

https://cloud.google.com/blog/products/data-analytics/pub-sub-launches-direct-path-to-bigquery-for-streaming-analytics
You should have said pub sub has been upgrade to directly stream to bigquery templates...not new pub sub

👍 ↩ 🚩 upvoted 2 times

☐ 👤 **baimus** `Most Recent ⊘` 1 month, 1 week ago

The question specifies that transformation occurs in Bigquery. This means the new direct pub/sub to bigquery streaming path is correct.

👍 ↩ 🚩 upvoted 1 times

☐ 👤 **Prakzz** 3 months, 3 weeks ago

Need PubSub and Dataflow both for this

👍 ↩ 🚩 upvoted 1 times

☐ 👤 **ludovikush** 7 months, 3 weeks ago

Werner123 i agree

👍 ↩ 🚩 upvoted 1 times

☐ 👤 **Werner123** 7 months, 3 weeks ago

User data would most likely include PII, for that case it is still recommended to use Dataflow since you need to remove/anonymise sensitive data.

👍 ↩ 🚩 upvoted 2 times

☐ 👤 **pico** 11 months, 1 week ago

I would have added "with / without data transformation" to the question to choose the right answer between A or D

👍 ↩ 🚩 upvoted 1 times

☐ 👤 **andresvelasco** 1 year, 1 month ago

I had my doubts between A and D.
But since the transformation will occur in bigquery I think Pubsub suffices.

👍 ↩ 🚩 upvoted 3 times

☐ 👤 **M25** 1 year, 5 months ago

Agree with TNT87. From the same link: "For Pub/Sub messages where advanced preload transformations or data processing before landing data in BigQuery (such as masking PII) is necessary, we still recommend going through Dataflow." It's "analyze user activity data", not merely streaming IoT into BigQuery so that concerns like privacy are per se n/a. One can deal with PII after landing in BigQuery as well, but apparently that's not what they recommend.

👍 ↩ 🚩 upvoted 3 times

☐ 👤 **PHD_CHENG** 1 year, 6 months ago

Pub/Sub -> DataFlow -> BigQuery

👍 ↩ 🚩 upvoted 2 times

☐ 👤 **TNT87** 1 year, 7 months ago

D. Configure Pub/Sub and a Dataflow streaming job to ingest the data into BigQuery.

This solution involves using Google Cloud Pub/Sub as the messaging service to receive the data from the mobile application, and then using Google Cloud Dataflow to transform and load the data into BigQuery in real time. Pub/Sub is a scalable and reliable messaging service that can handle high-volume real-time data streaming, while Dataflow provides a unified programming model to develop and run data processing pipelines. This solution is suitable for handling large volumes of user activity data from mobile applications and ingesting it into BigQuery in real time for analysis and ML experimentation

activity data from mobile applications and ingesting it into BigQuery in real-time for analysis and ML experimentation.

👍 ↩ ⚑ upvoted 2 times

⊟ 👤 **TNT87** 1 year, 7 months ago

Starting today, you no longer have to write or run your own pipelines for data ingestion from Pub/Sub into BigQuery. We are introducing a new type of Pub/Sub subscription called a "BigQuery subscription" that writes directly from Cloud Pub/Sub to BigQuery. This new extract, load, and transform (ELT) path will be able to simplify your event-driven architecture. For Pub/Sub messages where advanced preload transformations or data processing before landing data in BigQuery (such as masking PII) is necessary, we still recommend going through Dataflow

👍 ↩ ⚑ upvoted 1 times

⊟ 👤 **hiromi** 1 year, 10 months ago

Selected Answer: A

A
agree with pshemol

👍 ↩ ⚑ upvoted 3 times

⊟ 👤 **mymy9418** 1 year, 10 months ago

Selected Answer: D

need dataflow

👍 ↩ ⚑ upvoted 2 times

⊟ 👤 **mil_spyro** 1 year, 10 months ago

transformation will be handled in BQ hence I think A

👍 ↩ ⚑ upvoted 6 times

⊟ 👤 **mymy9418** 1 year, 9 months ago

agree.

👍 ↩ ⚑ upvoted 1 times

We are the biggest and most updated IT certification exam material website.

Using our own resources, we strive to strengthen the IT professionals community for free.