Google Discussions

EXAM PROFESSIONAL DATA ENGINEER TOPIC 1 QUESTION 134 DISCUSSION

Actual exam question from Google's Professional Data Engineer
Question #: 134
Topic #: 1
[All Professional Data Engineer Questions]

You are building an application to share financial market data with consumers, who will receive data feeds. Data is collected from the markets in real time.
Consumers will receive the data in the following ways:
- Real-time event stream
- ANSI SQL access to real-time stream and historical data
- Batch historical exports

Which solution should you use?

A. Cloud Dataflow, Cloud SQL, Cloud Spanner

B. Cloud Pub/Sub, Cloud Storage, BigQuery

C. Cloud Dataproc, Cloud Dataflow, BigQuery

D. Cloud Pub/Sub, Cloud Dataproc, Cloud SQL

Show Suggested Answer

by [deleted] at *March 22, 2020, 10:43 a.m.*

Comments

Type your comment...

Submit

[Removed] Highly Voted 5 years, 1 month ago

should be B

upvoted 24 times

itche_scratche Highly Voted 4 years, 7 months ago

D, not ideal but only option that work. You need pubsub, then a processing layer (dataflow or dataproc), then storage (some sql database).

upvoted 12 times

seiyassa 4 years, 4 months ago

I think pubsub doesn't have good connection to dataproc, so D is not the answer

upvoted 3 times

jkhong 2 years, 4 months ago

As of Dec 2022,there is the PubSub Lite connector to Dataproc

upvoted 2 times

jkhong 2 years, 4 months ago

We can have our pubsub topics to have BigQuery subscriptions, where data is automatically streamed into our BQ tables. Autoscaling is already handled automatically so this renders Dataflow and Dataproc pretty irrelevant for our usecase

upvoted 1 times

cetanx 1 year, 10 months ago

Here is the reference:
https://cloud.google.com/blog/products/data-analytics/pub-sub-launches-direct-path-to-bigquery-for-streaming-analytics

upvoted 2 times

SamuelTsch Most Recent 6 months, 1 week ago

Selected Answer: D

why B? The main goal of the question is data storage. Thus BigQuery is not neccessary for this situation. Option D from my point of view cover the whole requirements. Pub/Sub for streaming data, dataproc for data processing, SQL for storage.

upvoted 1 times

barnac1es 1 year, 7 months ago

B. Cloud Pub/Sub, Cloud Storage, BigQuery.

Here's how this solution aligns with your requirements:
Real-time Event Stream: Cloud Pub/Sub is a managed messaging service that can handle real-time event streams efficiently. You can use Pub/Sub to ingest and publish real-time market data to consumers.
ANSI SQL Access: BigQuery supports ANSI SQL queries, making it suitable for both real-time and historical data analysis. You can stream data into BigQuery tables from Pub/Sub and provide ANSI SQL access to consumers.
Batch Historical Exports: Cloud Storage can be used for batch historical exports. You can export data from BigQuery to Cloud Storage in batch, making it available for consumers to download.

upvoted 4 times

vaga1 1 year, 11 months ago

Selected Answer: B

I was in doubt as I did not know that BQ handles real-time access to data without dataflow underneath.

https://cloud.google.com/bigquery/docs/write-api#:~:text=You%20can%20use%20the%20Storage,in%20a%20single%20atomic%20operation.

upvoted 2 times

midgoo 2 years, 1 month ago

Selected Answer: B

Event Stream -> PubSub
PubSub has direct Write to BigQuery
Historical Exports to GCS

upvoted 1 times

AzureDP900 2 years, 4 months ago

B. Cloud Pub/Sub, Cloud Storage, BigQuery

upvoted 3 times

AzureDP900 2 years, 4 months ago

https://cloud.google.com/solutions/stream-analytics/

upvoted 1 times

zellck 2 years, 5 months ago

Selected Answer: B

B is the answer.

upvoted 3 times

John_Pongthorn 2 years, 7 months ago

Selected Answer: B

B: https://cloud.google.com/solutions/stream-analytics/
Real-time made real easy
Adopt simple ingestion for complex events
Ingest and analyze hundreds of millions of events per second from applications or devices virtually anywhere on the globe with Pub/Sub. Or directly stream millions of events per second into your data warehouse for SQL-based analysis with BigQuery's streaming API.

upvoted 3 times

John_Pongthorn 2 years, 7 months ago

Selected Answer: B

No matter what the last of it must end up with bigquery and the first service is pubsub I think intimidate service it should be dataflow

upvoted 1 times

Motivated_Gamer 3 years, 1 month ago

Selected Answer: A

Dataflow: Streaming data
CLoud SQL: for ansi sql support
Spanner: for batch historical data export

upvoted 1 times

tavva_prudhvi 3 years ago

You gonna use batch historical export for Spanner? It's B!

upvoted 2 times

Prasanna_kumar 3 years, 2 months ago

Answer is B

upvoted 1 times

MaxNRG 3 years, 3 months ago

Selected Answer: B

Cloud Pub/Sub, Cloud Dataflow, BigQuery
https://cloud.google.com/solutions/stream-analytics/

upvoted 4 times

MaxNRG 1 year, 4 months ago

B. Cloud Pub/Sub, Cloud Storage, BigQuery
The key requirements here are:
1. Real-time event stream (Pub/Sub)
2. ANSI SQL access to real-time and historical data (BigQuery)
3. Batch historical exports (Cloud Storage)
So Cloud Pub/Sub provides the real-time stream, BigQuery provides ANSI SQL access to stream and historical data, and Cloud Storage enables batch historical exports.
Option A is incorrect because Cloud Spanner does not offer batch exports and Dataflow is overkill for just SQL access.
Option C is incorrect as Dataproc is for spark workloads, not serving consumer data.
Option D is incorrect as Cloud SQL does not provide batch export capabilities.
Therefore, option B with Pub/Sub, Storage, and BigQuery is the best solution given the stated requirements. Dataflow
https://cloud.google.com/solutions/stream-analytics/

upvoted 1 times

medeis_jar 3 years, 3 months ago

Selected Answer: B

➪ Real-time event stream -> Pub/Sub
➪ ANSI SQL access to real-time stream and historical data -> BigQuery
➪ Batch historical exports -> Cloud Storage

upvoted 10 times

JG123 3 years, 5 months ago

Correct: B

upvoted 1 times

AdrianMonter26 3 years, 5 months ago

I think it must be D because you need Pub/Sub for streaming data, Dataflow or DataProc to get the data from Pub/Sub and store it in a database and finally the Cloud SQL database to store the data.
A and C cannot be because it is missing something for streaming data
B It can't be because you need something to pass the data from Pub/Sub to Cloud storage

upvoted 3 times

sumanshu 3 years, 10 months ago

Vote for B

upvoted 3 times

Load full discussion...