Google Discussions

EXAM PROFESSIONAL DATA ENGINEER TOPIC 1 QUESTION 220 DISCUSSION

Actual exam question from Google's Professional Data Engineer
Question #: 220
Topic #: 1
[All Professional Data Engineer Questions]

You are migrating your on-premises data warehouse to BigQuery. One of the upstream data sources resides on a MySQL. database that runs in your on-premises data center with no public IP addresses. You want to ensure that the data ingestion into BigQuery is done securely and does not go through the public internet. What should you do?

A. Update your existing on-premises ETL tool to write to BigQuery by using the BigQuery Open Database Connectivity (ODBC) driver. Set up the proxy parameter in the simba.googlebigqueryodbc.ini file to point to your data center's NAT gateway.

B. Use Datastream to replicate data from your on-premises MySQL database to BigQuery. Set up Cloud Interconnect between your on-premises data center and Google Cloud. Use Private connectivity as the connectivity method and allocate an IP address range within your VPC network to the Datastream connectivity configuration. Use Server-only as the encryption type when setting up the connection profile in Datastream.

C. Use Datastream to replicate data from your on-premises MySQL database to BigQuery. Use Forward-SSH tunnel as the connectivity method to establish a secure tunnel between Datastream and your on-premises MySQL database through a tunnel server in your on-premises data center. Use None as the encryption type when setting up the connection profile in Datastream.

D. Use Datastream to replicate data from your on-premises MySQL database to BigQuery. Gather Datastream public IP addresses of the Google Cloud region that will be used to set up the stream. Add those IP addresses to the firewall allowlist of your on-premises data center. Use IP Allowlisting as the connectivity method and Server-only as the encryption type when setting up the connection profile in Datastream.

Show Suggested Answer

by [redacted] at Dec. 30, 2023, 9:46 a.m.

by e70ea9e at Dec. 30, 2023, 9:46 a.m.

Comments

Type your comment...

Submit

raaad Highly Voted 10 months ago

Selected Answer: B

- Datastream is a serverless change data capture and replication service, which can be used to replicate data changes from MySQL to BigQuery.
- Using Cloud Interconnect provides a private, secure connection between your on-premises environment and Google Cloud ==> This method ensures that data doesn't go through the public internet and is a recommended approach for secure, large-scale data migrations.
- Setting up private connectivity with Datastream allows for secure and direct data transfer.

upvoted 8 times

josech Most Recent 5 months, 2 weeks ago

Selected Answer: B

https://cloud.google.com/datastream/docs/network-connectivity-options

upvoted 2 times

datapassionate 9 months, 3 weeks ago

Selected Answer: B

Datastream is a seamless replication from relational databases directly to BigQuery. The source database can be hosted on-premises, on Google Cloud services such as Cloud SQL or Bare Metal Solution for Oracle, or anywhere else on any cloud. https://cloud.google.com/datastream-for-bigquery#benefits

upvoted 1 times

datapassionate 9 months, 3 weeks ago

It is required that the data ingestion into BigQuery is done securely and does not go through the public internet. It can be done by Interconnect.

upvoted 1 times

e70ea9e 10 months, 1 week ago

Selected Answer: B

Secure Private Connection:

Cloud Interconnect establishes a direct, private connection between your on-premises network and Google Cloud, bypassing the public internet and ensuring data confidentiality.
Datastream Integration:

Datastream seamlessly replicates data from your MySQL database to BigQuery, handling the complexities of data transfer and synchronization.

upvoted 2 times