EXAM PROFESSIONAL DATA ENGINEER TOPIC 1 QUESTION 305 DISCUSSION

Actual exam question from Google's Professional Data Engineer
Question #: 305
Topic #: 1
[All Professional Data Engineer Questions]

Your organization uses a multi-cloud data storage strategy, storing data in Cloud Storage, and data in Amazon Web Services' (AWS) S3 storage buckets. All data resides in US regions. You want to query up-to-date data by using BigQuery, regardless of which cloud the data is stored in. You need to allow users to query the tables from BigQuery without giving direct access to the data in the storage buckets. What should you do?

A. Setup a BigQuery Omni connection to the AWS S3 bucket data. Create BigLake tables over the Cloud Storage and S3 data and query the data using BigQuery directly.

B. Set up a BigQuery Omni connection to the AWS S3 bucket data. Create external tables over the Cloud Storage and S3 data and query the data using BigQuery directly.

C. Use the Storage Transfer Service to copy data from the AWS S3 buckets to Cloud Storage buckets. Create BigLake tables over the Cloud Storage data and query the data using BigQuery directly.

D. Use the Storage Transfer Service to copy data from the AWS S3 buckets to Cloud Storage buckets. Create external tables over the Cloud Storage data and query the data using BigQuery directly.

Show Suggested Answer

by rahulvin at *Dec. 30, 2023, 10:03 p.m.*

Comments

Type your comment...

Submit

raaad Highly Voted 1 year, 4 months ago

Selected Answer: A

- BigQuery Omni: This is an extension of BigQuery that allows you to analyze data across Google Cloud, AWS, and Azure without having to manage the infrastructure or move data across clouds. It's suitable for querying data stored in AWS S3 buckets directly.
- BigLake: Allows you to create a logical abstraction (table) over data stored in Cloud Storage and S3, so you can query data using BigQuery without moving it.
- Unified Querying: By setting up BigQuery Omni to connect to AWS S3 and creating BigLake tables over both Cloud Storage and S3 data, you can query all data using BigQuery directly.

upvoted 13 times

ML6 1 year, 2 months ago

I wonder, why BigLake tables (A) over external tables (B)?

upvoted 4 times

aoifneofi_ef 8 months, 2 weeks ago

external tables can be created only on data residing in Cloud Storage, BigTable or Google Drive: https://cloud.google.com/bigquery/docs/external-tables. Hence creating external tables WITHOUT BQ Omni is not an option

upvoted 1 times

AllenChen123 1 year, 3 months ago

Agree. https://cloud.google.com/bigquery/docs/omni-introduction
"To run BigQuery analytics on your external data, you first need to connect to Amazon S3 or Blob Storage. If you want to query external data, you would need to create a BigLake table that references Amazon S3 or Blob Storage data."

upvoted 6 times

JyoGCP Most Recent 1 year, 2 months ago

Selected Answer: A

Option A

upvoted 2 times

Matt_108 1 year, 3 months ago

Selected Answer: A

Option A - clearly explained in comments

upvoted 2 times

rahulvin 1 year, 4 months ago

Selected Answer: A

A - BigLake tables work for S3 and GCS

upvoted 3 times

rahulvin 1 year, 4 months ago

https://cloud.google.com/bigquery/docs/external-data-sources#external_data_source_feature_comparison

upvoted 2 times