Google Discussions

EXAM PROFESSIONAL DATA ENGINEER TOPIC 1 QUESTION 195 DISCUSSION

Actual exam question from Google's Professional Data Engineer
Question #: 195
Topic #: 1
[All Professional Data Engineer Questions]

Your company wants to be able to retrieve large result sets of medical information from your current system, which has over 10 TBs in the database, and store the data in new tables for further query. The database must have a low-maintenance architecture and be accessible via SQL. You need to implement a cost-effective solution that can support data analytics for large result sets. What should you do?

A. Use Cloud SQL, but first organize the data into tables. Use JOIN in queries to retrieve data.

B. Use BigQuery as a data warehouse. Set output destinations for caching large queries.

C. Use a MySQL cluster installed on a Compute Engine managed instance group for scalability.

D. Use Cloud Spanner to replicate the data across regions. Normalize the data in a series of tables.

Show Suggested Answer

by ducc at *Sept. 3, 2022, 3:56 a.m.*

Comments

Type your comment...

Submit

AWSandeep Highly Voted 2 years, 2 months ago

Selected Answer: B

B. Use BigQuery as a data warehouse. Set output destinations for caching large queries.

upvoted 8 times

MaxNRG Most Recent 10 months, 2 weeks ago

Selected Answer: B

Option B is the best approach - use BigQuery as a data warehouse, and set output destinations for caching large queries.

The key reasons why BigQuery fits the requirements:

It is a fully managed data warehouse built to scale to handle massive datasets and perform fast SQL analytics
It has a low maintenance architecture with no infrastructure to manage
SQL capabilities allow easy querying of the medical data
Output destinations allow configurable caching for fast retrieval of large result sets
It provides a very cost-effective solution for these large scale analytics use cases
In contrast, Cloud Spanner and Cloud SQL would not scale as cost effectively for 10TB+ data volumes. Self-managed MySQL on Compute Engine also requires more maintenance. Hence, leveraging BigQuery as a fully managed data warehouse is the optimal solution here.

upvoted 3 times

AzureDP900 1 year, 10 months ago

B. Use BigQuery as a data warehouse. Set output destinations for caching large queries. Most Voted

upvoted 2 times

zellck 1 year, 11 months ago

Selected Answer: B

B is the answer.

upvoted 3 times

TNT87 2 years, 1 month ago

Answer B.
https://cloud.google.com/bigquery/docs/query-overview

upvoted 4 times

ducc 2 years, 2 months ago

Selected Answer: B

B is correct

upvoted 2 times