Google Discussions

EXAM PROFESSIONAL DATA ENGINEER TOPIC 1 QUESTION 73 DISCUSSION

Actual exam question from Google's Professional Data Engineer
Question #: 73
Topic #: 1
[All Professional Data Engineer Questions]

You are designing storage for two relational tables that are part of a 10-TB database on Google Cloud. You want to support transactions that scale horizontally.
You also want to optimize data for range queries on non-key columns. What should you do?

A. Use Cloud SQL for storage. Add secondary indexes to support query patterns.

B. Use Cloud SQL for storage. Use Cloud Dataflow to transform data to support query patterns.

C. Use Cloud Spanner for storage. Add secondary indexes to support query patterns.

D. Use Cloud Spanner for storage. Use Cloud Dataflow to transform data to support query patterns.

Show Suggested Answer

by [deleted] at *March 21, 2020, 4:52 p.m.*

Comments

Type your comment...

Submit

[Removed] Highly Voted 4 years, 1 month ago
Answer: C
Description: Spanner allows transaction tables to scale horizontally and secondary indexes for range queries

Description: Spanner allows transaction tables to scale horizontally and secondary indexes for range queries

upvoted 31 times

[Removed] Highly Voted 4 years, 1 month ago

Correct: C

upvoted 9 times

nhanhoangle Most Recent 1 year ago

Selected Answer: C

Correct: C

upvoted 1 times

PolyMoe 1 year, 3 months ago

Selected Answer: C

Cloud Spanner is a fully-managed, horizontally scalable relational database service that supports transactions and allows you to optimize data for range queries on non-key columns. By using Cloud Spanner for storage, you can ensure that your database can scale horizontally to meet the needs of your application.
To optimize data for range queries on non-key columns, you can add secondary indexes, this will allow you to perform range scans on non-key columns, which can improve the performance of queries that filter on non-key columns.

upvoted 2 times

samdhimal 1 year, 3 months ago

C. Use Cloud Spanner for storage. Add secondary indexes to support query patterns.

Cloud Spanner is a fully-managed, horizontally scalable relational database service that supports transactions and allows you to optimize data for range queries on non-key columns. By using Cloud Spanner for storage, you can ensure that your database can scale horizontally to meet the needs of your application.
To optimize data for range queries on non-key columns, you can add secondary indexes, this will allow you to perform range scans on non-key columns, which can improve the performance of queries that filter on non-key columns.

upvoted 3 times

samdhimal 1 year, 3 months ago

- Option A, Using Cloud SQL for storage and adding secondary indexes to support query patterns, may not be the best option as Cloud SQL is a relational database service that does not support horizontal scaling and may not be able to handle the large amount of data and the number of queries required by your application.

upvoted 2 times

samdhimal 1 year, 3 months ago

- Option B, Using Cloud SQL for storage and using Cloud Dataflow to transform data to support query patterns, may not be the best option as Cloud SQL is a relational database service that does not support horizontal scaling and may not be able to handle the large amount of data and the number of queries required by your application. Additionally, Cloud Dataflow is a data processing service and not a storage service, so it may not be the best fit for this use case.

- Option D, Using Cloud Spanner for storage and using Cloud Dataflow to transform data to support query patterns, is not necessary as Cloud Spanner provides the ability to optimize data for range queries on non-key columns by adding secondary indexes. Cloud Spanner also supports transactional consistency, which is a feature that allows you to perform multiple operations that must be performed together in a single transaction. Additionally, Cloud Dataflow is a data processing service and not a storage service, so it may not be the best fit for this use case.

upvoted 2 times

Mathew106 9 months, 2 weeks ago

Cloud SQL does support replicas to increase availability. Why is that not considered horizontal scaling?

upvoted 2 times

zellck 1 year, 5 months ago

Selected Answer: C

C is the answer.

https://cloud.google.com/architecture/autoscaling-cloud-spanner
When you create a Cloud Spanner instance, you choose the number of compute capacity nodes or processing units to serve your data. However, if the workload of an instance changes, Cloud Spanner doesn't automatically adjust the size of the instance. This document introduces the Autoscaler tool for Cloud Spanner (Autoscaler), an open source tool that you can use as a companion tool to Cloud Spanner. This tool lets you automatically increase or reduce the number of nodes or processing units in one or more Spanner instances based on how their capacity is being used.

https://cloud.google.com/spanner/docs/secondary-indexes
You can also create secondary indexes for other columns. Adding a secondary index on a column makes it more efficient to look up data in that column.

upvoted 1 times

sedado77 1 year, 7 months ago

Selected Answer: C

As sumanshu said

upvoted 1 times

tsoetan001 2 years, 6 months ago

Answer: C

upvoted 1 times

sumanshu 2 years, 10 months ago

Vote for C

upvoted 4 times

sumanshu 2 years, 10 months ago

A is not correct because Cloud SQL does not natively scale horizontally.
B is not correct because Cloud SQL does not natively scale horizontally.
C is correct because Cloud Spanner scales horizontally, and you can create secondary indexes for the range queries that are required.
D is not correct because Dataflow is a data pipelining tool to move and transform data, but the use case is centered around querying.

upvoted 8 times

timolo 3 years, 1 month ago

Answer: C
https://cloud.google.com/spanner/docs/secondary-indexes

upvoted 2 times

Nileshk611 3 years, 4 months ago

Correct: C

upvoted 3 times

arghya13 3 years, 5 months ago

Correct answers is C

upvoted 2 times