Google Discussions

EXAM PROFESSIONAL DATA ENGINEER TOPIC 1 QUESTION 47 DISCUSSION

Actual exam question from Google's Professional Data Engineer
Question #: 47
Topic #: 1
[All Professional Data Engineer Questions]

You are designing the database schema for a machine learning-based food ordering service that will predict what users want to eat. Here is some of the information you need to store:
- The user profile: What the user likes and doesn't like to eat
- The user account information: Name, address, preferred meal times
- The order information: When orders are made, from where, to whom

The database will be used to store all the transactional data of the product. You want to optimize the data schema. Which Google Cloud Platform product should you use?

A. BigQuery

B. Cloud SQL

C. Cloud Bigtable

D. Cloud Datastore

Show Suggested Answer

by rickywck at *March 17, 2020, 4:50 a.m.*

Comments

Type your comment...

Submit

jvg637 Highly Voted 5 years, 1 month ago

You want to optimize the data schema + Machine Learning --> Bigquery. So A

upvoted 61 times

yoshik 3 years, 7 months ago

BigQuery is a datawarehouse, not a transactional db. You need to store transactional data as a requirement.

upvoted 28 times

alexmirmao 3 years, 7 months ago

In my opinion transactional data doesnt mean transactions they could be grouped so there is no need to write register by register.

upvoted 9 times

yoshik 3 years, 6 months ago

In other questions they talk about 'transactional log data' when referring to past transactions, but you could be right, agree. In that case ok A BigQuery. Nevertheless, the question is formulated ambiguously.

upvoted 5 times

alecuba16 2 years, 9 months ago

Biquery Supports transactions:
https://cloud.google.com/bigquery/docs/reference/standard-sql/transactions
, but indeed is not a good DB for OLTP.

But I would said or CloudSQL or BigQuery

upvoted 4 times

[Removed] Highly Voted 5 years, 1 month ago

Answer: Should be D - Datastore

upvoted 26 times

GeeBeeEl 4 years, 5 months ago

There is SQLML with BigQuery, you know that?
You cannot optimize a schema in datastore, it is a NoSQL document database built for automatic scaling, high performance, and ease of application development. It does not work based on schemas!

upvoted 20 times

BigQuery 3 years, 5 months ago

BQML is there. But, In question do they want to do ML on BQ?? Its saying just ML Based Company.

upvoted 7 times

cetanx 1 year, 11 months ago

It was also a difficult one for Chat GPT, it did give different answers each time I inquiry more about the question. After a few iterations, we also agreed on "D" :) - because;

In the context of a food ordering service, storing data about what a user likes and doesn't like to eat can potentially involve a varied and dynamic set of data. Some users might have a long list of food preferences, while others might have only a few. Some users might update their likes and dislikes frequently, while others rarely or never. This kind of data is a good match for a NoSQL database like Datastore, which can easily accommodate such variations.

upvoted 5 times

abhaya2608 Most Recent 1 month, 1 week ago

Selected Answer: A

Cloud SQL doesn't support ML so the right answer will be BigQuery

upvoted 1 times

dcruzado 2 months, 1 week ago

Selected Answer: B

Transactional db -> CloudSQL

upvoted 1 times

dcruzado 2 months, 1 week ago

Selected Answer: B

Since they need transactional data i would say B
However thinking on machine learning is better A

upvoted 1 times

cqrm3n 3 months ago

Selected Answer: B

The answer should be Cloud SQL because it is a relational database suitable for transactional data.

BigQuery is for analytics and querying - not suitable for transactional workload.
Bigtable is for unstructured and time series data.
Datastore is a nosql document database for semi structured data.

upvoted 2 times

Yad_datatonic 3 months, 1 week ago

Selected Answer: A

For a machine learning-based food ordering service that requires optimised storage of transactional data, Google Cloud BigQuery is a suitable choice

upvoted 1 times

grshankar9 3 months, 2 weeks ago

Selected Answer: A

ithin Google Cloud, the database that most readily allows for data schema optimization is BigQuery; it provides features like schema auto-detection, columnar storage, and the ability to manually define your schema to tailor it for efficient querying and analysis of large datasets.

upvoted 1 times

manikolbe 3 months, 4 weeks ago

Selected Answer: B

Cloud SQL is the best choice for your application as it provides relational database management and is optimized for storing transactional data with SQL querying capabilities. It is well-suited for managing user profiles, account information, and orders, ensuring data integrity, and supporting complex queries necessary for the food ordering service.

upvoted 3 times

Ronn27 4 months ago

Selected Answer: B

Use BigQuery for analyzing aggregated data (e.g., predicting food trends or training ML models).
Use Cloud Bigtable for large-scale real-time recommendation engines if needed in the future.
Use Firestore for dynamic, semi-structured data with real-time updates if you need flexibility over transactional consistency.
Cloud SQL strikes the right balance for this use case due to its support for structured data, transactions, and easy integration with other GCP services.
So B. CloudSQL is the right answer

upvoted 2 times

sravi1200 4 months, 2 weeks ago

Selected Answer: B

Cloud SQL can store transactional data not Big Query. Big Query is an analytical service.

upvoted 1 times

DGames 4 months, 2 weeks ago

Selected Answer: A

Easy implement data schema + Machine Learning model in Big Query

upvoted 1 times

julydev82 4 months, 3 weeks ago

Selected Answer: B

database will be used to storage all transactional data.... I think that you need a relational database for that, then federated tables to bigquery to analysis.

upvoted 1 times

decipher9 5 months, 2 weeks ago

For a machine learning-based food ordering service that needs to store transactional data, Cloud SQL is the most suitable option. Here's why:

Cloud SQL is a fully-managed relational database service that supports transactional workloads, making it ideal for storing user profiles, account information, and order details.
It provides strong consistency and supports complex queries, which are essential for managing and retrieving transactional data efficiently.
While BigQuery is excellent for large-scale data analysis, it is not optimized for transactional data storage.
Cloud Bigtable is designed for high-throughput and low-latency workloads but lacks the transactional capabilities needed for this use case.
Cloud Datastore is a NoSQL database that supports transactions but is generally less powerful than a relational database for complex transactional schemas12.
So, the best choice for your needs is B. Cloud SQL.

upvoted 1 times

SamuelTsch 6 months, 2 weeks ago

Selected Answer: B

No idea why so many people go to A. But as transactional data, I think B is correct.

upvoted 1 times

baimus 7 months, 2 weeks ago

Selected Answer: D

The details of the information definitely look suited to noSql to me, so that means C or D. Datastore is designed for this sort of thing - transactional nosql for an App. I took the question to mean "the machine learning app already exists" so the fact bigquery allows ML isn't relevant. It would be a leap to assume that the ML is done in Bigquery (I have a current Google ML pro cert, and this wouldn't say bigquery to me from that cert)

upvoted 1 times

Nittin 8 months, 1 week ago

Selected Answer: B

Cloud SQL is a fully-managed relational database service that supports MySQL, PostgreSQL, and SQL Server. It is well-suited for transactional workloads, allowing you to store structured data with relationships between different entities, such as users, orders, and profiles.

upvoted 1 times

Load full discussion...