Google Discussions

EXAM PROFESSIONAL DATA ENGINEER TOPIC 1 QUESTION 152 DISCUSSION

Actual exam question from Google's Professional Data Engineer
Question #: 152
Topic #: 1
[All Professional Data Engineer Questions]

You work for a global shipping company. You want to train a model on 40 TB of data to predict which ships in each geographic region are likely to cause delivery delays on any given day. The model will be based on multiple attributes collected from multiple sources. Telemetry data, including location in GeoJSON format, will be pulled from each ship and loaded every hour. You want to have a dashboard that shows how many and which ships are likely to cause delays within a region. You want to use a storage solution that has native functionality for prediction and geospatial processing. Which storage solution should you use?

A. BigQuery

B. Cloud Bigtable

C. Cloud Datastore

D. Cloud SQL for PostgreSQL

Show Suggested Answer

by [deleted] at *March 22, 2020, 8:05 a.m.*

Comments

Type your comment...

Submit

[Removed] Highly Voted 4 years, 7 months ago

Answer: A
Description: Geospatial and ML functionality is with bigquery

upvoted 22 times

[Removed] Highly Voted 4 years, 7 months ago

Answer : A

upvoted 15 times

moumou Most Recent 8 months ago

Selected Answer: A

[Removed] Highly Voted 3 years, 11 months ago
Answer: A

upvoted 1 times

mothkuri 8 months ago

Answer : A
Statement "You want to have a dashboard that shows how many and which ships are likely to cause delays within a region" means we run analytical queries using ML. So BigQuery is Correct answer and it can able to store large volume of data

upvoted 1 times

barnac1es 1 year, 1 month ago

Selected Answer: A

Here's why BigQuery is a good choice:

Scalable Data Storage: BigQuery is a fully managed, highly scalable data warehouse that can handle large volumes of data, including your 40 TB dataset. It allows you to store and manage your data efficiently.

SQL for Predictive Analytics: BigQuery supports standard SQL and has built-in machine learning capabilities through BigQuery ML. You can easily build predictive models using SQL queries, which aligns with your goal of predicting ship delays.

Geospatial Processing: BigQuery has robust support for geospatial data processing. It provides functions for working with GeoJSON and geospatial data types, making it suitable for your ship telemetry data and geospatial analysis.

Integration with Dashboards: BigQuery can be easily integrated with visualization tools like Google Data Studio or other BI tools. You can create interactive dashboards to monitor ship delays based on your model's predictions.

upvoted 3 times

musumusu 1 year, 8 months ago

Answer B: BigTable,
Catchup words: Telemetry (sensor- semi structured data) as data is bigger than 500GB, datastore is not a good option.
GEOJSON , bigquery has geospatical capabilites but still not quick enough for semi structure geojson data.
Prediction for delay of ships <<likely to>> For me its time crucial and almost real time requirement. BigQuery is not suitable for it.
Best solution for this case is: Use BigTable for storage, create a datflow pipeline / google cloud AI platform for time senstive prediction.

upvoted 1 times

musumusu 1 year, 8 months ago

answer A: You are just looking for a storage solution not a workflow

upvoted 4 times

zellck 1 year, 11 months ago

Selected Answer: A

A is the answer.

https://cloud.google.com/bigquery/docs/geospatial-intro
In a data warehouse like BigQuery, location information is very common. Many critical business decisions revolve around location data. For example, you may record the latitude and longitude of your delivery vehicles or packages over time. You may also record customer transactions and join the data to another table with store location data.

You can use this type of location data to determine when a package is likely to arrive or to determine which customers should receive a mailer for a particular store location. Geospatial analytics let you analyze and visualize geospatial data in BigQuery by using geography data types and Google Standard SQL geography functions.

upvoted 2 times

Atnafu 1 year, 11 months ago

A
Geospatial analytics let you analyze and visualize geospatial data in BigQuery by using geography data types and Google Standard SQL geography functions. + BigqueryML

upvoted 1 times

JG123 2 years, 11 months ago

Answer: C

upvoted 1 times

Chihhanyu 2 years, 11 months ago

GeoJson + Native functionality for prediction -> BigQuery

upvoted 3 times

singh_payal_1404 2 years, 11 months ago

Answer : A

upvoted 1 times

PM17 3 years ago

This is more of a question that an answer but: How much data can Bigquery handle?

40TB seems to be a lot and bigtable can handle that, but of course Bigquery is better when it comes to ML and GIS.

upvoted 3 times

haroldbenites 4 years, 2 months ago

A is correct

upvoted 3 times

FARR 4 years, 2 months ago

A
https://cloud.google.com/bigquery/docs/gis-intro

upvoted 3 times

Rajokkiyam 4 years, 7 months ago

Answer A

upvoted 2 times