EXAM PROFESSIONAL DATA ENGINEER TOPIC 1 QUESTION 309 DISCUSSION

Actual exam question from Google's Professional Data Engineer
Question #: 309
Topic #: 1
[All Professional Data Engineer Questions]

You work for an airline and you need to store weather data in a BigQuery table. Weather data will be used as input to a machine learning model. The model only uses the last 30 days of weather data. You want to avoid storing unnecessary data and minimize costs. What should you do?

A. Create a BigQuery table where each record has an ingestion timestamp. Run a scheduled query to delete all the rows with an ingestion timestamp older than 30 days.

B. Create a BigQuery table partitioned by datetime value of the weather date. Set up partition expiration to 30 days.

C. Create a BigQuery table partitioned by ingestion time. Set up partition expiration to 30 days.

D. Create a BigQuery table with a datetime column for the day the weather data refers to. Run a scheduled query to delete rows with a datetime value older than 30 days.

Show Suggested Answer

by AllenChen123 at *Jan. 26, 2024, 12:30 a.m.*

Comments

Type your comment...

Submit

AllenChen123 Highly Voted 1 year, 3 months ago

Selected Answer: B

Partitioned based on weather date, with partition expiration set

upvoted 6 times

iooj Highly Voted 9 months ago

Selected Answer: B

got this one on the exam, aug 2024, passed

upvoted 6 times

desertlotus1211 Most Recent 1 month, 1 week ago

Selected Answer: C

Partitioning by ingestion time is simpler and sufficient if data retention is based on load time, not the data's internal timestamp

upvoted 1 times

juliorevk 3 months, 1 week ago

Selected Answer: B

B

BQ partitioning with partition expiration of 30 days allows you to only filter for the last 30 days and delete days that are beyond 30 days.

upvoted 1 times

d11379b 1 year, 1 month ago

https://cloud.google.com/bigquery/docs/partitioned-tables
Here it mentions " For TIMESTAMP and DATETIME columns, the partitions can have either hourly, daily, monthly, or yearly granularity.l
So you should not calculate the amount of partitions on second granularity

upvoted 3 times

chambg 1 year, 1 month ago

Selected Answer: D

Skeptical about Option B as maximum partitions in a BQ table is 4000.Since Datetime value is a timestamp it will have more than 4000 values in a duration for 30 days (30*24*60*60 = 259,200). So Option D is right imo

upvoted 1 times

ce9e395 1 year, 1 month ago

This is a good point

upvoted 1 times

joao_01 1 year ago

It's not a good point. The granularity goes to DAYs, not SECONDs. So, the right answer is B.

upvoted 6 times

JyoGCP 1 year, 2 months ago

Selected Answer: B

Option B

upvoted 1 times

Sofiia98 1 year, 3 months ago

Selected Answer: B

We need the last 30 days, we don't care about ingestion time

upvoted 4 times