

🔗 Google Discussions



Exam Professional Data Engineer All Questions

View all questions & answers for the Professional Data Engineer exam

[Go to Exam](#)



EXAM PROFESSIONAL DATA ENGINEER TOPIC 1 QUESTION 169 DISCUSSION

Actual exam question from Google's Professional Data Engineer

Question #: 169

Topic #: 1

[\[All Professional Data Engineer Questions\]](#)

You are migrating a table to BigQuery and are deciding on the data model. Your table stores information related to purchases made across several store locations and includes information like the time of the transaction, items purchased, the store ID, and the city and state in which the store is located. You frequently query this table to see how many of each item were sold over the past 30 days and to look at purchasing trends by state, city, and individual store. How would you model this table for the best query performance?

- A. Partition by transaction time; cluster by state first, then city, then store ID.
- B. Partition by transaction time; cluster by store ID first, then city, then state.
- C. Top-level cluster by state first, then city, then store ID.
- D. Top-level cluster by store ID first, then city, then state.

[Show Suggested Answer](#)

by **ducc** at *Sept. 3, 2022, 7:10 a.m.*

Comments

Type your comment...

[Submit](#)

🗄️ 👤 **AWSandeep** Highly Voted 2 years, 8 months ago

Selected Answer: A

A. Partition by transaction time; cluster by state first, then city, then store ID.

👍 🔄 🚩 upvoted 9 times

🗄️ 👤 **Atnafu** Highly Voted 2 years, 4 months ago

A

Partitioning is obvious

Clustering is already mentioned in the question
past 30 days and to look at purchasing trends by
state,
city, and
individual store

👍 🔄 🚩 upvoted 7 times

🗄️ 👤 **SamuelTsch** Most Recent 6 months, 1 week ago

Selected Answer: A

go to A.

👍 🔄 🚩 upvoted 1 times

🗄️ 👤 **MaxNRG** 1 year, 4 months ago

Selected Answer: B

over the past 30 days -> partitioning
by state, city, and individual store -> cluster order

👍 🔄 🚩 upvoted 3 times

🗄️ 👤 **MaxNRG** 1 year, 4 months ago

For optimal query performance in BigQuery, especially for the described use cases of analyzing sales data by time and geographical hierarchies, the data should be organized to minimize the amount of data scanned during queries. Given the frequent queries over the past 30 days and analysis by location, the best approach is:

Option A: Partition by transaction time; cluster by state first, then city, then store ID.

👍 🔄 🚩 upvoted 1 times

🗄️ 👤 **MaxNRG** 1 year, 4 months ago

Partitioning the table by transaction time allows for efficient querying over specific time ranges, such as the past 30 days, which reduces costs and improves performance because it limits the amount of data scanned.

Clustering by state, then city, and then store ID aligns with the hierarchy of geographical data and the types of queries that are run against the dataset. It organizes the data within each partition so that queries filtering by state, city, or store ID—or any combination of these—are optimized, as BigQuery can limit the scan to just the relevant clusters within the partitions.

👍 🔄 🚩 upvoted 2 times

🗄️ 👤 **tibuenoc** 1 year, 5 months ago

Selected Answer: B

Partition by ingest time

Partition by specified data column (Id, State and City)

👍 🔄 🚩 upvoted 1 times

🗄️ 👤 **ffgrrre** 1 year, 6 months ago

Selected Answer: C

Partition by transaction time would lead to too many partitions - if it was a date, it would have made sense.

👍 🔄 🚩 upvoted 1 times

🗄️ 👤 **sylva1212** 9 months, 1 week ago

Even though its a timestamp, the partitioning can be configured on a daily granularity, so A is correct (https://cloud.google.com/bigquery/docs/partitioned-tables#date_timestamp_partitioned_tables)

👍 🔄 🚩 upvoted 1 times

🗄️ 👤 **aureole** 1 year, 7 months ago

Selected Answer: C

It should be C. not A

👍 🔄 🚩 upvoted 1 times

🗄️ 👤 **aureole** 1 year, 7 months ago

I think it should be C.

The fact that we partition the table with the time of the transaction will result many transactions in each day, so it will affect negatively the query performance.

negatively the query performance.

i.e : by the end of the day I will have many partitions if I use the transaction time. A would be correct if the partition was by date and not by time.

Response: C.

👍 ↩ 🚩 upvoted 1 times

📄 👤 **vaga1** 1 year, 11 months ago

Selected Answer: A

Partitioning for time is obvious to improve performance and costs of querying only the last 30 days of the table.

So, the answer is A or B.

<https://cloud.google.com/bigquery/docs/querying-clustered-tables>

"... To get the benefits of clustering, include all of the clustered columns or a subset of the columns in left-to-right sort order, starting with the first column."

This means that it is a better choice to sort the table rows by region-province-city (region-state-city in the US case).

So, the answer is A.

👍 ↩ 🚩 upvoted 4 times

📄 👤 **Prakzz** 2 years, 4 months ago

Selected Answer: B

Should be B

The clustering should be according to the filtering needs

👍 ↩ 🚩 upvoted 2 times

📄 👤 **zellick** 2 years, 5 months ago

Selected Answer: A

A is the answer.

<https://cloud.google.com/bigquery/docs/partitioned-tables>

This page provides an overview of partitioned tables in BigQuery. A partitioned table is a special table that is divided into segments, called partitions, that make it easier to manage and query your data. By dividing a large table into smaller partitions, you can improve query performance, and you can control costs by reducing the number of bytes read by a query. You can partition BigQuery tables by:

- Time-unit column: Tables are partitioned based on a TIMESTAMP, DATE, or DATETIME column in the table.

<https://cloud.google.com/bigquery/docs/clustered-tables>

Clustered tables in BigQuery are tables that have a user-defined column sort order using clustered columns. Clustered tables can improve query performance and reduce query costs.

👍 ↩ 🚩 upvoted 4 times

📄 👤 **TNT87** 2 years, 7 months ago

<https://cloud.google.com/bigquery/docs/querying-clustered-tables>

👍 ↩ 🚩 upvoted 2 times

📄 👤 **ducc** 2 years, 8 months ago

Selected Answer: A

A

The question mention that the query is 30 days recently

👍 ↩ 🚩 upvoted 3 times



Platform

> Home

> Examtopics PRO

> All Exams

> Training Courses

