

[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 304 DISCUSSION

Actual exam question from Google's Professional Data Engineer

Question #: 304

Topic #: 1

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

You have a table that contains millions of rows of sales data, partitioned by date. Various applications and users query this data many times a minute. The query requires aggregating values by using AVG, MAX, and SUM, and does not require joining to other tables. The required aggregations are only computed over the past year of data, though you need to retain full historical data in the base tables. You want to ensure that the query results always include the latest data from the tables, while also reducing computation cost, maintenance overhead, and duration. What should you do?

- A. Create a materialized view to aggregate the base table data. Include a filter clause to specify the last one year of partitions.
- B. Create a materialized view to aggregate the base table data. Configure a partition expiration on the base table to retain only the last one year of partitions.
- C. Create a view to aggregate the base table data. Include a filter clause to specify the last year of partitions.
- D. Create a new table that aggregates the base table data. Include a filter clause to specify the last year of partitions. Set up a scheduled query to recreate the new table every hour.

[Show Suggested Answer](#)

by [scaenruy](#) at Jan. 4, 2024, 1:41 p.m.

Comments

Submit

 **raaad** Highly Voted 9 months, 4 weeks ago

Selected Answer: A

- Materialized View: Materialized views in BigQuery are precomputed views that periodically cache the result of a query for increased performance and efficiency. They are especially beneficial for heavy and repetitive aggregation queries.
- Filter for Recent Data: Including a clause to focus on the last year of partitions ensures that the materialized view is only storing and updating the relevant data, optimizing storage and refresh time.
- Always Up-to-date: Materialized views are maintained by BigQuery and automatically updated at regular intervals, ensuring they include the latest data up to a certain freshness point.

   upvoted 11 times

 **MBNR** Most Recent 1 month, 3 weeks ago

Selected Answer: A

Answer : A


Question has below three requirements , it did NOT talk about STORAGE cost

Reducing computation cost: Using Materialized views in BigQuery, query costs can be lower due to faster performance

maintenance overhead : Bigquery takes care of data updates

and duration: Since the results are precomputed and stored , it takes very less time for the query output

   upvoted 1 times


 **Pime13** 3 months, 4 weeks ago

Selected Answer: A

Option B, creating a materialized view and configuring a partition expiration on the base table to retain only the last one year of partitions, would not meet the requirement of retaining full historical data in the base tables. Partition expiration would delete older data, which is not desirable if you need to keep the full historical data.

Option A, on the other hand, allows you to create a materialized view that aggregates the data for the past year without deleting any historical data from the base table. This ensures that you always have access to the latest data while retaining the full history.


   upvoted 1 times

 **JyoGCP** 8 months, 2 weeks ago

Selected Answer: A

Option A

   upvoted 2 times

 **et2137** 8 months, 2 weeks ago

Selected Answer: C

materialized view requires refreshing so it might not fulfill the requirement: "results always include the latest data from the tables". Opt. C will give you the newest data every time you execute the query but it will have to be computed every time

   upvoted 2 times

 **d11379b** 7 months, 1 week ago

Agree, these questions always play with words, making many of options seem plausible

   upvoted 1 times

 **d11379b** 7 months, 1 week ago

But materialized view always returns fresh data

Fresh data. Materialized views return fresh data. If changes to base tables might invalidate the materialized view, then data is read directly from the base tables. If the changes to the base tables don't invalidate the materialized view, then rest of the data is read from the materialized view and only the changes are read from the base tables.

<https://cloud.google.com/bigquery/docs/materialized-views-intro>

   upvoted 2 times

 **casadocc** 9 months ago




A We can do aggregations, but if not specified table will not be partitioned on the view.

B partition expiration is not possible, as expiration is the same as base table

C It might be the right one, although not specific savings vs the original query, but here we would guarantee accessing only last year data.

D not a good one in any sense

A and C might be equally good solutions depending on some understandings. Would probably opt for A

   upvoted 1 times

 **Matt_108** 9 months, 3 weeks ago

Selected Answer: A

. Create a materialized view to aggregate the base table data. Include a filter clause to specify the last one year of partitions.

👍 ↩ 🚩 upvoted 2 times

🗂 👤 **Sofiia98** 9 months, 4 weeks ago

Selected Answer: A

To preserve the historical data

👍 ↩ 🚩 upvoted 2 times

🗂 👤 **scaenruy** 10 months ago

Selected Answer: B

B. Create a materialized view to aggregate the base table data. Configure a partition expiration on the base table to retain only the last one year of partitions.

👍 ↩ 🚩 upvoted 1 times

🗂 👤 **Sofiia98** 9 months, 4 weeks ago

Don't agree, it is said thad that we need to store the historical data, so answer A is correct

👍 ↩ 🚩 upvoted 4 times

🗂 👤 **raaad** 9 months, 4 weeks ago

Why not B

- Configuring partition expiration on the BASE TABLE is a way to manage storage and costs by automatically dropping old data. However, the question specifies the need to retain full historical data, making this approach not suitable since it doesnt keep all historical records.

👍 ↩ 🚩 upvoted 4 times



Platform

> [Home](#)

> [Examtopics PRO](#)

> [All Exams](#)

> [Training Courses](#)

