

🔗 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 229 DISCUSSION

Actual exam question from Google's Professional Data Engineer

Question #: 229

Topic #: 1

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

You currently use a SQL-based tool to visualize your data stored in BigQuery. The data visualizations require the use of outer joins and analytic functions. Visualizations must be based on data that is no less than 4 hours old. Business users are complaining that the visualizations are too slow to generate. You want to improve the performance of the visualization queries while minimizing the maintenance overhead of the data preparation pipeline. What should you do?

- A. Create materialized views with the `allow_non_incremental_definition` option set to true for the visualization queries. Specify the `max_staleness` parameter to 4 hours and the `enable_refresh` parameter to true. Reference the materialized views in the data visualization tool.
- B. Create views for the visualization queries. Reference the views in the data visualization tool.
- C. Create a Cloud Function instance to export the visualization query results as parquet files to a Cloud Storage bucket. Use Cloud Scheduler to trigger the Cloud Function every 4 hours. Reference the parquet files in the data visualization tool.
- D. Create materialized views for the visualization queries. Use the incremental updates capability of BigQuery materialized views to handle changed data automatically. Reference the materialized views in the data visualization tool.

[Show Suggested Answer](#)

by [e70ea9e](#) at Dec. 30, 2023, 9:55 a.m.

Comments

Type your comment...

Submit

  **baimus** 7 months ago

Selected Answer: A

Just a note, the question saying "data no less than 4 hours old" presumably means "no more than 4 hours old"

   upvoted 1 times

  **JamesKarianis** 8 months, 3 weeks ago

Selected Answer: B

Unfortunately the correct answer is B due to the limitations of materialized views, doesn't support any other join than inner and no analytical function is supported

   upvoted 2 times

  **AWSandeep** 4 months, 2 weeks ago

Yes, they do if they are non-incremental.

   upvoted 1 times

  **ricardovazz** 1 year, 1 month ago

Selected Answer: A

A

<https://cloud.google.com/bigquery/docs/materialized-views-create#non-incremental>

In scenarios where data staleness is acceptable, for example for batch data processing or reporting, non-incremental materialized views can improve query performance and reduce cost.

allow_non_incremental_definition option. This option must be accompanied by the max_staleness option. To ensure a periodic refresh of the materialized view, you should also configure a refresh policy.

   upvoted 3 times

  **Matt_108** 1 year, 3 months ago

Selected Answer: A

Option A is better than D, since it accounts for data staleness and is better suited for heavy querying, thanks to the allow_non_incremental_definition

   upvoted 2 times

  **Jordan18** 1 year, 4 months ago

A seems right but whats wrong with option D, can anybody please explain?

   upvoted 4 times

  **datapassionate** 1 year, 3 months ago


Seems like materialized views can use incremental updates only if data was not deleted or updated in original table.

Here the data changes so I think that's the reason why it's not correct answer

https://cloud.google.com/bigquery/docs/materialized-views-use#incremental_updates

"BigQuery combines the cached view's data with new data to provide consistent query results while still using the materialized view. For single-table materialized views, this is possible if the base table is unchanged since the last refresh, or if only new data was added. For multi-table views, no more than one table can have appended data. If more than one of a multi-table view's base tables has changed, then the view cannot be incrementally updated."

   upvoted 2 times

  **raaad** 1 year, 4 months ago

Selected Answer: A

- Materialized views in BigQuery precompute and store the result of a base query, which can speed up data retrieval for complex queries used in visualizations.
- The max_staleness parameter allows us to specify how old the data can be, ensuring that the visualizations are based on data no less than 4 hours old.
- The enable_refresh parameter ensures that the materialized view is periodically refreshed.
- The allow_non_incremental_definition is used for enabling the creation of non-incrementally refreshable materialized views.

   upvoted 3 times

  **e70ea9e** 1 year, 4 months ago

Selected Answer: A

Precomputed Results: Materialized views store precomputed results of complex queries, significantly accelerating subsequent query performance, addressing the slow visualization issue.

Allow Non-Incremental Views: Using allow_non_incremental_definition circumvents the limitation of incremental updates for outer joins and analytic functions, ensuring views can be created for the specified queries.

Near-Real-Time Data: Setting max_staleness to 4 hours guarantees data freshness within the acceptable latency for

near-real-time data. Setting max_staleness to 4 hours guarantees data freshness within the acceptable latency for visualizations.

Automatic Refresh: Enabling refresh with enable_refresh maintains view consistency with minimal maintenance overhead.

Minimal Overhead: Materialized views automatically update as underlying data changes, reducing maintenance compared to manual exports or view definitions.

👍 ↩ 🚩 upvoted 3 times



Platform

> [Home](#)

> [Examtopics PRO](#)

> [All Exams](#)

> [Training Courses](#)



© 2024 ExamTopics