

[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 129 DISCUSSION

Actual exam question from Google's Professional Data Engineer

Question #: 129

Topic #: 1

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

You use BigQuery as your centralized analytics platform. New data is loaded every day, and an ETL pipeline modifies the original data and prepares it for the final users. This ETL pipeline is regularly modified and can generate errors, but sometimes the errors are detected only after 2 weeks. You need to provide a method to recover from these errors, and your backups should be optimized for storage costs. How should you organize your data in BigQuery and store your backups?

- A. Organize your data in a single table, export, and compress and store the BigQuery data in Cloud Storage.
- B. Organize your data in separate tables for each month, and export, compress, and store the data in Cloud Storage.
- C. Organize your data in separate tables for each month, and duplicate your data on a separate dataset in BigQuery.
- D. Organize your data in separate tables for each month, and use snapshot decorators to restore the table to a time prior to the corruption.

[Show Suggested Answer](#)

by [deleted] at March 22, 2020, 11:14 a.m.

Comments

[Submit](#)

🗄️ 👤 [Removed] Highly Voted 👍 5 years, 1 month ago

Should be B

👍 ↩️ 🚩 upvoted 22 times

🗄️ 👤 Ganshank Highly Voted 👍 5 years ago

B

The questions is specifically about organizing the data in BigQuery and storing backups.

👍 ↩️ 🚩 upvoted 12 times

🗄️ 👤 desertlotus1211 Most Recent 🕒 1 month, 2 weeks ago

Selected Answer: B

Snapshot decorators in BigQuery allow you to query a table at a past point in time, but they are limited by BigQuery's time travel window (which is typically 7 days). Since errors are sometimes only detected after 2 weeks, snapshot decorators won't be effective for recovering data beyond their retention period.

👍 ↩️ 🚩 upvoted 2 times

🗄️ 👤 grshankar9 3 months, 2 weeks ago

Selected Answer: D

With Storage Decorators, BigQuery only stores the differences between a snapshot and its base table, minimizing storage costs.

👍 ↩️ 🚩 upvoted 1 times

🗄️ 👤 desertlotus1211 1 month, 2 weeks ago

Does this answer address how should you organize your data in BigQuery and store your backups?

👍 ↩️ 🚩 upvoted 1 times

🗄️ 👤 cloud_rider 5 months ago

Selected Answer: D

D is the most cost optimized solution to keep the backup. please read the link - https://cloud.google.com/bigquery/docs/table-snapshots-intro#table_snapshots

👍 ↩️ 🚩 upvoted 2 times

🗄️ 👤 SamuelTsch 6 months, 1 week ago

Selected Answer: D

I think D is better.

👍 ↩️ 🚩 upvoted 2 times

🗄️ 👤 Lenifia 10 months ago

Selected Answer: D

The best option is D. Organize your data in separate tables for each month, and use snapshot decorators to restore the table to a time prior to the corruption.

👍 ↩️ 🚩 upvoted 2 times

🗄️ 👤 zevexWM 1 year ago

Selected Answer: D

Answer is D:

Snapshots are different from time travel. They can hold data as long as we want.

Furthermore "BigQuery only stores bytes that are different between a snapshot and its base table" so pretty cost effective as well.

https://cloud.google.com/bigquery/docs/table-snapshots-intro#table_snapshots

👍 ↩️ 🚩 upvoted 2 times

🗄️ 👤 Farah_007 1 year ago

Selected Answer: B

From : <https://cloud.google.com/architecture/dr-scenarios-for-data#BigQuery>

It can't be D

If the corruption is caught within 7 days, query the table to a point in time in the past to recover the table prior to the corruption using snapshot decorators.

Store the original data on Cloud Storage. This allows you to create a new table and reload the uncorrupted data. From there, you can adjust your applications to point to the new table. => D

👍 ↩️ 🚩 upvoted 2 times

🗄️ 👤 Nirca 1 year, 6 months ago

Selected Answer: D

D - this solution is integrated. No core is needed

👍 ↩️ 🚩 upvoted 5 times

🗄️ 👤 Rahuhali1088 1 year, 7 months ago

— **Barabari1900** 1 year, 7 months ago

90% of questions are having multiple answers and its very hard to get into every discussion where the conclusion is not there

   upvoted 7 times

🗨️ **ckanaar** 1 year, 7 months ago

Selected Answer: B

The answer is B:

Why not D? Because snapshot costs can become high if a lot of small changes are made to the base table:

[https://cloud.google.com/bigquery/docs/table-snapshots-](https://cloud.google.com/bigquery/docs/table-snapshots-intro#:~:text=Because%20BigQuery%20storage%20is%20column%2Dbased%2C%20small%20changes%20to%20the%20data%20in%20a%20base%20table%20can%20result%20in%20large%20increases%20in%20storage%20cost%20for%20its%20table%20snapshot.)

intro#:~:text=Because%20BigQuery%20storage%20is%20column%2Dbased%2C%20small%20changes%20to%20the%20data%20in%20a%20base%20table%20can%20result%20in%20large%20increases%20in%20storage%20cost%20for%20its%20table%20snapshot.

Since the question specifically states that the ETL pipeline is regularly modified, this means that lots of small changes are present. In combination with the requirement to optimize for storage costs, this means that option B is the way to go.

   upvoted 7 times

🗨️ **arien_chen** 1 year, 8 months ago

Selected Answer: D

keyword: detected after 2 weeks.

only snapshot could resolve the problem.

   upvoted 2 times

🗨️ **Lanro** 1 year, 9 months ago

Selected Answer: D

From BigQuery documentation - Benefits of using table snapshots include the following:

- Keep a record for longer than seven days. With BigQuery time travel, you can only access a table's data from seven days ago or more recently. With table snapshots, you can preserve a table's data from a specified point in time for as long as you want.

- Minimize storage cost. BigQuery only stores bytes that are different between a snapshot and its base table, so a table snapshot typically uses less storage than a full copy of the table.

So storing data in GCS will make copies of data for each table. Table snapshots are more optimal in this scenario.

   upvoted 8 times

🗨️ **vamgcp** 1 year, 9 months ago

Selected Answer: B

Organizing your data in separate tables for each month will make it easier to identify the affected data and restore it.

Exporting and compressing the data will reduce storage costs, as you will only need to store the compressed data in Cloud Storage.

Storing your backups in Cloud Storage will make it easier to restore the data, as you can restore the data from Cloud Storage directly

   upvoted 2 times

🗨️ **phidelics** 1 year, 10 months ago

Selected Answer: B

Organize in separate tables and store in GCS

   upvoted 3 times

🗨️ **cetanx** 1 year, 10 months ago

Just an additional info!

Here is an example for an export job;

```
$ bq extract --destination_format CSV --compression GZIP 'your_project:your_dataset.your_new_table' 'gs://your_bucket/your_object.csv.gz'
```

   upvoted 1 times

🗨️ **cetanx** 1 year, 10 months ago

I will update my answer to D.

Think of a scenario that you are in the last week of June and an error occurred 3 weeks ago (so still in June) however you do not have an export of the June table yet therefore you cannot recover the data simply because you don't have an export just yet.

So snapshots are way to go!

   upvoted 3 times

🗨️ **sdi_studiers** 1 year, 11 months ago

Selected Answer: D

D

"With BigQuery time travel, you can only access a table's data from seven days ago or more recently. With table snapshots, you can preserve a table's data from a specified point in time for as long as you want." [source: <https://cloud.google.com/bigquery/docs/table-snapshots-intro>]

   upvoted 3 times

[Load full discussion...](#)



Platform

> Home

> All Exams

> Examtopics PRO

> Training Courses



© 2024 ExamTopics