Google Discussions

EXAM PROFESSIONAL DATA ENGINEER TOPIC 1 QUESTION 211 DISCUSSION

Actual exam question from Google's Professional Data Engineer
Question #: 211
Topic #: 1
[All Professional Data Engineer Questions]

You are using BigQuery with a multi-region dataset that includes a table with the daily sales volumes. This table is updated multiple times per day. You need to protect your sales table in case of regional failures with a recovery point objective (RPO) of less than 24 hours, while keeping costs to a minimum. What should you do?

A. Schedule a daily export of the table to a Cloud Storage dual or multi-region bucket.

B. Schedule a daily copy of the dataset to a backup region.

C. Schedule a daily BigQuery snapshot of the table.

D. Modify ETL job to load the data into both the current and another backup region.

Show Suggested Answer

by e70ea9e at *Dec. 30, 2023, 9:34 a.m.*

Comments

Type your comment...

Submit

MaxNRG Highly Voted 1 year, 3 months ago

Selected Answer: A

Why not C:

A table snapshot must be in the same region, and under the same organization, as its base table.
https://cloud.google.com/bigquery/docs/table-snapshots-intro#limitations

upvoted 15 times

MaxNRG 1 year, 3 months ago

Based on the information provided and the need to avoid data loss in the case of a hard regional failure in BigQuery, which could result in the destruction of all data in that region, the focus should be on creating backups in a geographically distinct region. Considering this scenario, the most suitable option would be Option A

Here's why this option is the most appropriate:

upvoted 5 times

MaxNRG 1 year, 3 months ago

• Cross-Region Backup: Exporting the data to a Google Cloud Storage bucket that is either dual or multi-regional ensures that your backups are stored in a different geographic location. This is critical for protecting against hard regional failures.
• Data Durability: Cloud Storage provides high durability for stored data, making it a reliable option for backups in the case of regional disasters.
• Cost-Effectiveness: While there are costs associated with storage and data transfer, this method can be more cost-effective compared to maintaining active replicas of the data in multiple regions, especially if the data is large.

upvoted 3 times

MaxNRG 1 year, 3 months ago

• Flexibility and Automation: The export process can be automated and scheduled to occur daily, aligning with your RPO of less than 24 hours. This ensures that the most recent data is always backed up.
• Recovery Process: In the event of a hard regional failure, the data can be restored from the Cloud Storage backup to another operational BigQuery region, ensuring continuity of operations.

upvoted 3 times

MaxNRG 1 year, 3 months ago

The other options, while viable in certain scenarios, do not provide the same level of protection against a hard regional failure:
• Option B (Copy to Backup Region) and Option D (Modify ETL to Load into Backup Region) do not address the possibility of a hard regional failure adequately, as they do not necessarily imply storing data in a geographically distinct region.
• Option C (BigQuery Snapshot) is useful for point-in-time recovery but does not inherently protect against hard regional failures since the snapshots are within the same BigQuery service.
Focusing on a robust disaster recovery strategy is crucial. Option A provides a balance between ensuring data availability in the event of a regional disaster and managing costs, aligning with best practices for data management in the cloud.

upvoted 4 times

raaad Highly Voted 1 year, 4 months ago

Selected Answer: C

Option C provides cost-effective way.
- BigQuery table snapshots are a feature that allows you to capture the state of a table at a particular point in time.
- Snapshots are incremental, so they only store the data that has changed, making them more cost-effective than full table copies.
- In the event of a regional failure, you can quickly restore the table from a snapshot.

upvoted 7 times

ToiToi Most Recent 6 months ago

Selected Answer: A

Why other options are not as suitable:

B (Copy of the dataset): Copying the entire dataset daily is more expensive and less efficient than exporting just the table data.
C (BigQuery snapshot): snapshots are within the same region and won't protect against a regional outage.
D (Modify ETL job): This adds complexity to your ETL process and might not be the most efficient or cost-effective way to achieve your RPO.

upvoted 2 times

desertlotus1211 1 month, 2 weeks ago

What about the cost inplications?

BigQuery exports are full-table exports, so you pay for every row scanned and written.
Storage in multi-region buckets is a bit more expensive than single-region.

Answer C is better suited

upvoted 1 times

Priyal19 8 months, 1 week ago

A : BQ snapshot should be in same region, and if so the region fails so does the snapshot.

upvoted 1 times

viciousjpjp 8 months, 2 weeks ago

Selected Answer: C

Why Option A is not suitable: Restoring data from Option A would require reloading it back into BigQuery, which is time-consuming. This process cannot guarantee a recovery point objective (RPO) of less than 24 hours.

upvoted 2 times

meh_33 8 months, 3 weeks ago

Believe me all questions were from Exam topic all were there yesterday in exam. But yes dont go with starting questions mainly focus questions after 200 and latest questions are at last page.

upvoted 2 times

meh_33 8 months, 4 weeks ago

Selected Answer: C

C seems correct and Raaad also saying same.

upvoted 1 times

ostora 1 year ago

Selected Answer: C

it is c

upvoted 1 times

pandeyspecial 1 year, 2 months ago

Selected Answer: C

C. Schedule a daily BigQuery snapshot of the table.

Here's why:

Cost-effective: BigQuery snapshots are significantly cheaper than daily exports to Cloud Storage or copying the entire dataset to a backup region. They offer point-in-time backups with minimal storage costs.
Fast recovery: Snapshots can be restored quickly, meeting your RPO requirement of less than 24 hours.
Multi-regional: By default, BigQuery snapshots are automatically stored in a different region from the source data, ensuring redundancy and disaster recovery.

upvoted 2 times

Sergei_B 1 year, 1 month ago

At the beginning I also thought that "C" is a correct answer, but futher I found documentation https://cloud.google.com/bigquery/docs/locations. According this documentation
"Selecting a multi-region location does not provide cross-region replication or regional redundancy, so there is no increase in dataset availability in the event of a regional outage. Data is stored in a single region within the geographic location.

Data located in the EU multi-region is only stored in the europe-west1 (Belgium) or europe-west4 (Netherlands) data centers"
So, multi-region dataset just means locating data inside one of US regions, hence snapshot also will be stored in the same region what means that answer C is not correct

upvoted 2 times

JyoGCP 1 year, 2 months ago

Selected Answer: A

A. Schedule a daily export of the table to a Cloud Storage dual or multi-region bucket.

upvoted 1 times

GCP001 1 year, 3 months ago

Selected Answer: A

Option A. Check the ref for regional loss - https://cloud.google.com/bigquery/docs/reliability-intro#scenario_loss_of_region

upvoted 2 times

datapassionate 1 year, 3 months ago

Selected Answer: A

A. Schedule a daily export of the table to a Cloud Storage dual or multi-region bucket.

upvoted 1 times

Matt_108 1 year, 3 months ago

Selected Answer: A

A: MaxNRG and Helinia cleared the reasons very well

upvoted 2 times

Helinia 1 year, 4 months ago

Selected Answer: A

"BigQuery does not offer durability or availability in the extraordinarily unlikely and unprecedented event of physical region loss. This is true for both "regions and multi-region" configurations. Hence maintaining durability and availability under such a scenario requires customer planning."

"To avoid data loss in the face of destructive regional loss, you need to back up data to another geographic location. For example, you could periodically export a snapshot of your data to Google Cloud Storage in another geographically distinct region."

Ref: https://cloud.google.com/bigquery/docs/reliability-intro#scenario_loss_of_region

upvoted 5 times

MaxNRG 1 year, 3 months ago

Option A (Export to Cloud Storage): While exporting to Cloud Storage is a viable backup strategy, it can be more expensive and less efficient than using snapshots, especially if the table is large and updated frequently.

upvoted 1 times

MaxNRG 1 year, 3 months ago

I agree, It's A:
A table snapshot must be in the same region, and under the same organization, as its base table.
https://cloud.google.com/bigquery/docs/table-snapshots-intro#limitations

upvoted 1 times

Helinia 1 year, 3 months ago

Why not C:
"BigQuery also supports the ability to snapshot tables. With this feature you can explicitly backup data within the same region for longer than the 7 day time travel window. A snapshot is purely a metadata operation and results in no additional storage bytes. While this can add protection against accidental deletion, it does not increase the durability of the data."

https://cloud.google.com/bigquery/docs/reliability-intro#scenario_accidental_deletion_or_data_corruption

upvoted 3 times

qq589539483084gfrgrgfr 1 year, 4 months ago

Option A

upvoted 3 times

qq589539483084gfrgrgfr 1 year, 4 months ago

https://cloud.google.com/bigquery/docs/reliability-intro

upvoted 3 times

e70ea9e 1 year, 4 months ago

Selected Answer: D

Automatically replicates data to a backup region upon each update, ensuring an RPO of less than 24 hours, even with multiple daily updates.

upvoted 2 times

raaad 1 year, 4 months ago

Option D:
Doubles the write load and storage costs since you are maintaining two live datasets.

upvoted 4 times