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 318 DISCUSSION

Actual exam question from Google's Professional Data Engineer
Question #: 318
Topic #: 1
[All Professional Data Engineer Questions]

You are using BigQuery with a regional 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 HectorLeon2099 at *Dec. 4, 2024, 5:57 p.m.*

Comments

Type your comment...

Submit

desertlotus1211 3 weeks, 3 days ago

Selected Answer: A

almost the same as 211, 211 says multi-region vs regional...

upvoted 1 times

Parandhaman_Margan 1 month, 3 weeks ago

Selected Answer: C

Meets the RPO requirement (< 24 hours)
Cost-effective solution
Quick recovery from regional failures

upvoted 1 times

gabbferreira 1 week, 5 days ago

snapshots are stored in the same region, so it dont protect from regional failure

upvoted 1 times

MarcoPellegrino 2 months, 2 weeks ago

Selected Answer: D

https://cloud.google.com/blog/topics/developers-practitioners/backup-disaster-recovery-strategies-bigquery

Google presents both A and D
Why A:
- Cost: Lower. GCS storage is significantly cheaper than BigQuery storage. You pay for storage in GCS and minimal egress charges when exporting.
- Complexity: Simpler. You schedule a daily export job. Restoring involves importing from GCS to BigQuery in another region.
- Consistency: Easier to manage. The export process creates a consistent snapshot of the data at the time of export. You might have some latency (up to 24 hours in this scenario), but the data within the export is consistent.
- RPO: Meets the requirement. A daily export ensures an RPO of less than 24 hours.
- RTO: Depends on the restore process from GCS to BigQuery. You can pre-provision slots in the backup region to minimize restore time.

upvoted 1 times

Pime13 3 months, 4 weeks ago

Selected Answer: A

This approach ensures that your data is stored in multiple regions, providing redundancy and protection against regional failures

c is not possible it has some limitations compared to exporting the table to a Cloud Storage dual or multi-region bucket:

Regional Limitation: BigQuery table snapshots are limited to the same region as the base table. This means that if the entire region experiences a failure, the snapshot may not be accessible.

Storage Costs: While snapshots can be cost-effective, they still incur storage costs for the data that is different from the base table. Exporting to Cloud Storage can be more cost-effective, especially when using dual or multi-region buckets.

RPO Considerations: Both options can meet an RPO of less than 24 hours, but exporting to Cloud Storage provides additional redundancy by storing data in multiple regions, enhancing data availability and durability

upvoted 1 times

FireAtMe 4 months, 1 week ago

Selected Answer: A

Both A and B works. But it is cheaper to save data in GCS.

upvoted 1 times

joelcaro 4 months, 2 weeks ago

Selected Answer: D

Opción D: Modify ETL job to load the data into both the current and another backup region
Evaluación:
Ajustar el ETL para escribir en dos tablas (una en la región principal y otra en una región de respaldo) asegura que los datos estén disponibles en ambas ubicaciones casi en tiempo real.
Esto garantiza un RPO de menos de 24 horas, ya que las actualizaciones intradía se reflejan en ambas regiones.
Aunque podría aumentar los costos de almacenamiento por duplicar los datos, es la solución más efectiva y directa para proteger contra fallos regionales.

upvoted 2 times

mdell 4 months, 2 weeks ago

Selected Answer: B

In most cases, it is cheaper to copy a BigQuery dataset to a new region directly rather than exporting it to a Cloud Storage bucket and then loading it into a new BigQuery dataset in the desired region, as you only pay for data transfer costs when copying within BigQuery, while exporting to a bucket incurs additional storage charges for the exported data in Cloud Storage, even if it's only temporary.

Key points to consider:

No extra storage cost for copying:
When copying a BigQuery dataset to a new region, you only pay for the data transfer cost, not the storage of the data in a separate location.

Storage cost for exporting:
Exporting data to a Cloud Storage bucket means you are charged for the storage of that data in the bucket until you delete it, even if you are just temporarily storing it for transfer.

upvoted 1 times

HectorLeon2099 5 months ago

Selected Answer: A

Option A is the most cost efficient: https://cloud.google.com/blog/topics/developers-practitioners/backup-disaster-recovery-strategies-bigquery

upvoted 4 times

mdell 4 months, 2 weeks ago

Additionally it only mentions backing up the sales table and not the entire dataset

upvoted 1 times