← 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 249 DISCUSSION

Actual exam question from Google's Professional Data Engineer

Question #: 249

Topic #: 1

[All Professional Data Engineer Questions]

Your team is building a data lake platform on Google Cloud. As a part of the data foundation design, you are planning to store all the raw data in Cloud Storage. You are expecting to ingest approximately 25 GB of data a day and your billing department is worried about the increasing cost of storing old data. The current business requirements are:

• The old data can be deleted anytime.
• There is no predefined access pattern of the old data.
• The old data should be available instantly when accessed.
• There should not be any charges for data retrieval.

What should you do to optimize for cost?

> A. Create the bucket with the Autoclass storage class feature.
>
> B. Create an Object Lifecycle Management policy to modify the storage class for data older than 30 days to nearline, 90 days to coldline, and 365 days to archive storage class. Delete old data as needed.
>
> C. Create an Object Lifecycle Management policy to modify the storage class for data older than 30 days to coldline, 90 days to nearline, and 365 days to archive storage class. Delete old data as needed.
>
> D. Create an Object Lifecycle Management policy to modify the storage class for data older than 30 days to nearline, 45 days to coldline, and 60 days to archive storage class. Delete old data as needed.

Show Suggested Answer

by ☺ rasad at Jan, 5, 2024, 12:26 p.m.

Comments

Type your comment...

Submit

👤 **Smakyel79** Highly Voted 👍 1 year, 3 months ago

Selected Answer: A

https://cloud.google.com/storage/docs/autoclass

👍 ↩ 🚩 upvoted 10 times

👤 **raaad** Highly Voted 👍 1 year, 4 months ago

Selected Answer: A

- Autoclass automatically moves objects between storage classes without impacting performance or availability, nor incurring retrieval costs.
- It continuously optimizes storage costs based on access patterns without the need to set specific lifecycle management policies.

👍 ↩ 🚩 upvoted 8 times

👤 **SamuelTsch** Most Recent ⊙ 6 months, 1 week ago

Selected Answer: A

From the documentation https://cloud.google.com/storage/docs/autoclass

👍 ↩ 🚩 upvoted 1 times

👤 **hussain.sain** 9 months, 3 weeks ago

Selected Answer: B

the question clearly specifies there should not be any retrieval charges. so enabling autoclass is not recommended because we have to pay one time fees while retrieving the data. and usually soft delete is enable.

👍 ↩ 🚩 upvoted 1 times

  👤 **nadavw** 8 months, 1 week ago

  A one-time pay isn't considered a retrieval charge. A is correct

  👍 ↩ 🚩 upvoted 2 times

👤 **CGS22** 1 year ago

Selected Answer: B

Why B is the best choice:

Cost Optimization: This option leverages Cloud Storage's different storage classes to significantly reduce costs for storing older data. Nearline, coldline, and archive storage classes are progressively cheaper than the standard storage class, with trade-offs in availability and retrieval times.
Meets Requirements:
Old data deletion: You can manually delete old data whenever needed, fulfilling the first requirement.
No predefined access pattern: The policy automatically transitions data to cheaper storage classes based on age, regardless of access patterns.
Instant availability: Nearline storage provides immediate access to data, meeting the third requirement.
No retrieval charges: While there are retrieval charges for coldline and archive storage, nearline storage has no retrieval fees, satisfying the fourth requirement.

👍 ↩ 🚩 upvoted 3 times

👤 **Sofiia98** 1 year, 3 months ago

Selected Answer: A

For sure A, read the documentation

👍 ↩ 🚩 upvoted 4 times

👤 **GCP001** 1 year, 3 months ago

Selected Answer: A

autoclass is the correct way to handle all business cases

👍 ↩ 🚩 upvoted 2 times

👤 **therealsohail** 1 year, 4 months ago

Selected Answer: B

Create an Object Lifecycle Management policy to modify the storage class for data older than 30 days to nearline, 90 days to

coldline, and 365 days to archive storage class. Delete old data as needed.

👍 ↩ 🏴 upvoted 3 times