

[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 76 DISCUSSION

Actual exam question from Google's Professional Data Engineer

Question #: 76

Topic #: 1

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

Government regulations in your industry mandate that you have to maintain an auditable record of access to certain types of data. Assuming that all expiring logs will be archived correctly, where should you store data that is subject to that mandate?

- A. Encrypted on Cloud Storage with user-supplied encryption keys. A separate decryption key will be given to each authorized user.
- B. In a BigQuery dataset that is viewable only by authorized personnel, with the Data Access log used to provide the auditability.
- C. In Cloud SQL, with separate database user names to each user. The Cloud SQL Admin activity logs will be used to provide the auditability.
- D. In a bucket on Cloud Storage that is accessible only by an AppEngine service that collects user information and logs the access before providing a link to the bucket.

[Show Suggested Answer](#)

by [deleted] at *March 21, 2020, 6:01 p.m.*

Comments

[Submit](#)

  **Mitra123** Highly Voted  3 years, 7 months ago

Keywords here are

1. "Archived": Immutable and hence, BQ and Cloud SQL are ruled out
 2. "Auditable": Means track any changes done.
- Only D can provide the audibility piece!
I will go with D

   upvoted 50 times

  **Jarek7** 1 year, 6 months ago

I have no idea why so many upvotes on this answer:

- 1) archived doesn't mean immutable and cloud storage is not immutable too.
- 2) auditable means viewable for authorized personel - and in this case not changes need to be monitored but any access.
- 3) with option D it is easy to go around logging - you can add another access to the bucket read the data remove the access and no one will ever know that you accessed the data.
- 4) option D is much more difficult - you need to application on AppEngine to log the data and provide access for users.
- 5) option D doesn't explain where and how it stores the audit data - it could be accessed and modified from some side app/service.

   upvoted 11 times

  **[Removed]** Highly Voted  4 years, 7 months ago

Answer: B

Description: Bigquery is used to analyse access logs, data access logs capture the details of the user that accessed the data

   upvoted 23 times

  **sraakesh95** 2 years, 9 months ago

There is no option for archiving with BQ

   upvoted 1 times

  **tavva_prudhvi** 2 years, 6 months ago



You dont need to archive the expiring logs, you have to archive the un-archived data here! See the question, it says "Assuming that all expiring logs will be archived correctly", which means they are already stored somewhere like in GCS!!! Hence, better to store the remaining un-archived data in BQ.

   upvoted 5 times

  **vartiklis** 2 years, 3 months ago

The question is about where to store the `_data_` for which the logs will be generated.

The bit you quoted is about the `_logs_` that will be generated when accessing data. The "archived correctly" implies that proper retention policies will be set up if you choose GCS.

   upvoted 3 times

  **awssp12345** 3 years, 4 months ago

The question has no mention of ANALYZE.. BQ is not correct. I would go with D.

   upvoted 12 times

  **philli1011** Most Recent  9 months ago

In recent GCP, we have cloud audit.

   upvoted 1 times

  **Nandababy** 11 months, 1 week ago

Option B is valid only when analytics to be performed over logs, which is not mentioned anywhere

   upvoted 1 times

  **rocky48** 11 months, 1 week ago

Selected Answer: B

For maintaining an auditable record of access to certain types of data, especially when government regulations are in place, the most suitable option would be:

B. In a BigQuery dataset that is viewable only by authorized personnel, with the Data Access log used to provide the auditability.

Storing the data in a BigQuery dataset with restricted access ensures control over who can view the data, and utilizing Data Access logs provides a comprehensive audit trail for compliance purposes. This option aligns well with the need for maintaining an auditable record as mandated by government regulations.

   upvoted 3 times

  **Jarek7** 1 year, 6 months ago

Selected Answer: B

There are notes for option D, which says - limited to option BQ. The only REAL difference is that for option D you need to

if you are going for option D, why do you eliminate option B? The only REAL difference is that for option D you need to develop an app for storing log data and providing bucket link and in option B you have it all done BETTER by GCP. You might also pay a bit more for BQ storage, but the question never mentions about cost optimization. BTW in the D option the bucket is accessible only by AppEngine service, so what will the user do with the provided link? he has no access anyway... And if he even has the access to this link what stops him from using the same link many times? How the AppEngine get and store the information what specific data he accessed and how?

   upvoted 9 times

  **Kiroo** 1 year, 5 months ago

That was my thought, either B or D could work but D it's a little bit odd create an app to do something that could be achieved natively gcp

   upvoted 3 times

  **phidelics** 1 year, 5 months ago

I was about to say the same thing. Why go through that stress?


   upvoted 2 times

  **Rodrigo4N** 1 year, 6 months ago

Selected Answer: D

D amongus

   upvoted 2 times

  **juliojobs** 1 year, 7 months ago

Selected Answer: B

They want to know where you can store ****data**** in a way that every access is logged in an auditable way.

Both BQ and GCS have audit logs, except that in alternative D you're circumventing it by creating your own logs. I doubt Google would recommend that.

By types of data you can understand "financial type", "marketing type", etc.

   upvoted 3 times

  **midgoo** 1 year, 8 months ago

Selected Answer: D

I was thinking it should be A. However, 'data' in this question is too vague. It does not say anywhere that the data could fit in BigQuery tables. It could be unstructure data such as videos or images

Option D seems to involve more setup but it is the only viable option for this scenario. Note that GCS do have Cloud Audit logs. That should be the best option. Maybe this question was asked when Cloud Audit log is not yet available for GCS.

   upvoted 4 times

  **aleixfc96** 1 year, 9 months ago

It is so clear that is B lol

   upvoted 1 times

  **NamitSehgal** 1 year, 9 months ago

B bigquery for a record set store

   upvoted 1 times

  **PolyMoe** 1 year, 9 months ago

Selected Answer: B

B. In a BigQuery dataset that is viewable only by authorized personnel, with the Data Access log used to provide the auditability.


BigQuery provides built-in logging of all data access, including the user's identity, the specific query run and the time of the query. This log can be used to provide an auditable record of access to the data. Additionally, BigQuery allows you to control access to the dataset using Identity and Access Management (IAM) roles, so you can ensure that only authorized personnel can view the dataset.



   upvoted 2 times

  **samdhimal** 1 year, 9 months ago

B. In a BigQuery dataset that is viewable only by authorized personnel, with the Data Access log used to provide the auditability.

BigQuery provides built-in logging of all data access, including the user's identity, the specific query run and the time of the query. This log can be used to provide an auditable record of access to the data. Additionally, BigQuery allows you to control access to the dataset using Identity and Access Management (IAM) roles, so you can ensure that only authorized personnel can view the dataset.

   upvoted 3 times

  **Oleksandr0501** 1 year, 6 months ago

note: You are asked that option A does not provide an auditable record of access to the data, so it's not the correct data

gpt: You are correct that option A does not provide an auditable record of access to the data, as it only addresses data security through encryption. Option C provides auditability through Cloud SQL Admin activity logs, but it may not be the best option as it requires additional setup and management.

Option D is a feasible solution, but as you mentioned, it requires additional setup and maintenance of the AppEngine service. It also may not provide a comprehensive audit log of all data access.

Option B, storing the data in a BigQuery dataset that is viewable only by authorized personnel and using the Data Access log to provide auditability, is the most appropriate option as it provides built-in logging of all data access and allows you to control access to the dataset using IAM roles. Therefore, it provides both data security and auditable access to the data.

/// ok let it be B

👍 ↩ 🚩 upvoted 2 times

🗨️ 👤 **Oleksandr0501** 1 year, 6 months ago

OR MAYBE D....

👍 ↩ 🚩 upvoted 1 times

🗨️ 👤 **Oleksandr0501** 1 year, 5 months ago

!!! confused. Give 69% confidence to B, as user Jarek7 explained

👍 ↩ 🚩 upvoted 1 times

🗨️ 👤 **samdhimal** 1 year, 9 months ago

A. Encrypted on Cloud Storage with user-supplied encryption keys. A separate decryption key will be given to each authorized user. is a good option for data security but it does not provide an auditable record of access to the data.

C. In Cloud SQL, with separate database user names to each user. The Cloud SQL Admin activity logs will be used to provide the auditability. is also a good option for data security but it does not provide an auditable record of access to the data.

D. In a bucket on Cloud Storage that is accessible only by an AppEngine service that collects user information and logs the access before providing a link to the bucket. is also a good option but it requires additional setup and maintenance of the AppEngine service, and it may not provide an auditable record of access to the data.

👍 ↩ 🚩 upvoted 2 times

🗨️ 👤 **GCPpro** 1 year, 9 months ago

D is the correct answer

👍 ↩ 🚩 upvoted 1 times

🗨️ 👤 **RoshanAshraf** 1 year, 9 months ago

Selected Answer: D

Keys

TYPES of data --> Cloud Storage not BQ

Archival --> Cloud Storage

Access --> No decryption keys to all users

👍 ↩ 🚩 upvoted 1 times

🗨️ 👤 **PrashantGupta1616** 1 year, 10 months ago

Selected Answer: D

I will go with D

👍 ↩ 🚩 upvoted 1 times

🗨️ 👤 **DGames** 1 year, 10 months ago

Selected Answer: D

Keyword, Archiver, certain type of data, auditable, GCS is better option. Durability 11 time 9 to store log immutable for long time.

👍 ↩ 🚩 upvoted 1 times

[Load full discussion...](#)

> Home

> All Exams

> Examtopics PRO

> Training Courses



© 2024 ExamTopics