EXAM PROFESSIONAL DATA ENGINEER TOPIC 1 QUESTION 293 DISCUSSION

Actual exam question from Google's Professional Data Engineer
Question #: 293
Topic #: 1
[All Professional Data Engineer Questions]

Your organization is modernizing their IT services and migrating to Google Cloud. You need to organize the data that will be stored in Cloud Storage and BigQuery. You need to enable a data mesh approach to share the data between sales, product design, and marketing departments. What should you do?

A. 1. Create a project for storage of the data for each of your departments.
2. Enable each department to create Cloud Storage buckets and BigQuery datasets.
3. Create user groups for authorized readers for each bucket and dataset.
4. Enable the IT team to administer the user groups to add or remove users as the departments' request.

B. 1. Create multiple projects for storage of the data for each of your departments' applications.
2. Enable each department to create Cloud Storage buckets and BigQuery datasets.
3. Publish the data that each department shared in Analytics Hub.
4. Enable all departments to discover and subscribe to the data they need in Analytics Hub.

C. 1. Create a project for storage of the data for your organization.
2. Create a central Cloud Storage bucket with three folders to store the files for each department.
3. Create a central BigQuery dataset with tables prefixed with the department name.
4. Give viewer rights for the storage project for the users of your departments.

D. 1. Create multiple projects for storage of the data for each of your departments' applications.
2. Enable each department to create Cloud Storage buckets and BigQuery datasets.
3. In Dataplex, map each department to a data lake and the Cloud Storage buckets, and map the BigQuery datasets to zones.
4. Enable each department to own and share the data of their data lakes.

Show Suggested Answer

Show Suggested Answer

by scaenruy at *Jan. 4, 2024, 11:37 a.m.*

Comments

Type your comment...

Submit

raaad Highly Voted 1 year, 3 months ago

Selected Answer: D

- Decentralized ownership: Each department controls its data lake, aligning with the core principle of data ownership in a data mesh.
- Self-service data access: Departments can create and manage their own Cloud Storage buckets and BigQuery datasets within their data lakes, enabling self-service data access.
- Interdepartmental sharing: Dataplex facilitates data sharing by enabling departments to publish their data products from their data lakes, making it easily discoverable and usable by other departments.

upvoted 11 times

daed09 Most Recent 1 month, 4 weeks ago

It says we want to work with GCS data assets, thus Dataplex is a better option, so it's option D!

For those who says B as correct answer, I'd say that question states data assets in both GCS and BQ, however Analytics Hub is focused primarily for BQ assets.

upvoted 1 times

plum21 2 months, 4 weeks ago

Selected Answer: D

D because it looks like B is impossible due to the lack of GCS support in Analytics Hub

upvoted 1 times

marlon.andrei 3 months, 4 weeks ago

Selected Answer: B

In "a data mesh approach to share the data between sales, product design, and marketing departments", Analytics Hub is the solution.

upvoted 2 times

Nandababy 9 months ago

B is better option as organization is migrating to google cloud, that means teams doesnt have much hands on, analytical hub is more ease to use and solved the purpose as compared to dataplex were setup itself if very complex.

upvoted 2 times

987af6b 9 months, 2 weeks ago

Selected Answer: B

For a straightforward data mesh approach where the focus is on decentralizing data management while enabling easy data sharing and discovery, Analytics Hub is often the more appropriate choice due to its simplicity and directness. It facilitates the core objectives of a data mesh—decentralized data ownership and accessible data sharing—without the added complexity of managing data lakes and advanced governance features.

upvoted 3 times

joao_01 1 year ago

I think its B. I know since we are talking about Datamesh we want to go to the Dataplex service suddenly. However, in Dataplex a Lake can only have assets (bq tables etc) that are in the same project as the Dataplex service.

Example: There is bq table in project A and B. I want to to create a Lake in Dataplex in Project A that contains tables of project B. I can´t do that, i can only host tables of the Project A, since the Lake is in project A.

With this said, I think the best option is B, because the datamesh approach is related to "to share the data between sales, product design, and marketing departments". So the question is focusing only in the sharing part of the datamesh. Option B fits just fine.

upvoted 2 times

joao_01 1 year ago

I was wrong in my explanation guys. Look at this link:
https://cloud.google.com/dataplex/docs/add-zone

"A lake can include one or more zones. While a zone can only be part of one lake, it may contain assets that point to resources that are part of projects outside of its parent project."

So, option D seems good.

upvoted 2 times

JyoGCP 1 year, 2 months ago

Selected Answer: D

Option D

upvoted 1 times

Matt_108 1 year, 3 months ago

Selected Answer: D

that's pure data mesh, which is what dataplex has been built for

upvoted 2 times

Sofiia98 1 year, 3 months ago

Selected Answer: D

For me, Dataplex looks more logical

upvoted 1 times

GCP001 1 year, 3 months ago

D. Dataplex looks more suitable for data mesh approach, Check the ref - https://cloud.google.com/dataplex/docs/introduction

upvoted 1 times