

[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 210 DISCUSSION

Actual exam question from Google's Professional Data Engineer

Question #: 210

Topic #: 1

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

You are designing a data mesh on Google Cloud with multiple distinct data engineering teams building data products. The typical data curation design pattern consists of landing files in Cloud Storage, transforming raw data in Cloud Storage and BigQuery datasets, and storing the final curated data product in BigQuery datasets. You need to configure Dataplex to ensure that each team can access only the assets needed to build their data products. You also need to ensure that teams can easily share the curated data product. What should you do?

- A. 1. Create a single Dataplex virtual lake and create a single zone to contain landing, raw, and curated data.
2. Provide each data engineering team access to the virtual lake.
- B. 1. Create a single Dataplex virtual lake and create a single zone to contain landing, raw, and curated data.
2. Build separate assets for each data product within the zone.
3. Assign permissions to the data engineering teams at the zone level.
- C. 1. Create a Dataplex virtual lake for each data product, and create a single zone to contain landing, raw, and curated data.
2. Provide the data engineering teams with full access to the virtual lake assigned to their data product.
- D. 1. Create a Dataplex virtual lake for each data product, and create multiple zones for landing, raw, and curated data.
2. Provide the data engineering teams with full access to the virtual lake assigned to their data product.

[Show Suggested Answer](#)

by [e70ea9e](#) at Dec. 30, 2023, 9:33 a.m.

Comments

Type your comment...

Submit

  **f74ca0c** 4 months, 1 week ago

Selected Answer: D


Create a Dataplex lake that acts as the domain for your data mesh.

Add zones to your lake that represents individual teams within each domain and provide managed data contracts.

Attach assets that map to data stored in Cloud Storage.

https://cloud.google.com/transfer-appliance/docs/4.0/overview?_gl=1*8pbq1*_up*MQ..&gclid=CjwKCAiAg8S7BhATEiwAO2-R6gOYtlc2FJa7zE4lh3-2f00x9F3hwgul9lYjfJs2cAprxOleXq_NhoCw-8QAvD_BwE&gclsrc=aw.ds

   upvoted 1 times

  **SamuelTsch** 6 months, 1 week ago

Selected Answer: D

just like MaxNRG said

   upvoted 1 times

  **JyoGCP** 1 year, 2 months ago

Selected Answer: D

Answer D

   upvoted 1 times

  **datapassionate** 1 year, 3 months ago

Selected Answer: D

D. 1. Create a Dataplex virtual lake for each data product, and create multiple zones for landing, raw, and curated data.
2. Provide the data engineering teams with full access to the virtual lake assigned to their data product.

Lake: A logical construct representing a data domain or business unit. For example, to organize data based on group usage, you can set up a lake for each department (for example, Retail, Sales, Finance).

Zone: A subdomain within a lake, which is useful to categorize data by the following:

Stage: For example, landing, raw, curated data analytics, and curated data science.

   upvoted 1 times

  **datapassionate** 1 year, 3 months ago

<https://cloud.google.com/dataplex/docs/introduction>

   upvoted 1 times

  **Matt_108** 1 year, 3 months ago

Selected Answer: D

D: 1 virtual lake per Data Product (which stands for domain basically), zones to split data by "status". Each Data Eng team can access their own data exclusively and in a data mesh compliant way

   upvoted 1 times



  **MaxNRG** 1 year, 3 months ago

Selected Answer: D

The best approach is to create a Dataplex virtual lake for each data product, with multiple zones for landing, raw, and curated data. Then provide the data engineering teams with access only to the zones they need within the virtual lake assigned to their product.

To enable teams to easily share curated data products, you should use cross-lake sharing in Dataplex. This allows curated zones to be shared across virtual lakes while maintaining data isolation for other zones.

   upvoted 4 times

  **MaxNRG** 1 year, 3 months ago

So the steps would be:

1. Create a Dataplex virtual lake for each data product.

2. Within each lake, create separate zones for landing, raw, and curated data.

3. Provide each data engineering team with access only to the zones they need within their assigned virtual lake.

4. Configure cross-lake sharing on the curated data zones to share curated data products between teams.

This provides isolation and access control between teams for raw data while enabling easy sharing of curated data products.

https://cloud.google.com/dataplex/docs/introduction#a_domain-centric_data_mesh

   upvoted 3 times

  **Smakye179** 1 year, 3 months ago

I believe the answer is B, but there is a misspelling in the answer, should say "create multiple zones"

👍 ↩ 🚩 upvoted 2 times

🗄️ 👤 **Helinia** 1 year, 4 months ago

Selected Answer: D

Each lake should be created per data product since data product sounds like a domain in this question.

Since we have landing, raw, curated data, we should create different zones.

"Zones are of two types: raw and curated.

Raw zone: Contains data that is in its raw format and not subject to strict type-checking.

Curated zone: Contains data that is cleaned, formatted, and ready for analytics. The data is columnar, Hive-partitioned, and stored in Parquet, Avro, Orc files, or BigQuery tables. Data undergoes type-checking- for example, to prohibit the use of CSV files because they don't perform as well for SQL access."

Ref: <https://cloud.google.com/dataplex/docs/introduction#terminology>

👍 ↩ 🚩 upvoted 1 times

🗄️ 👤 **Jordan18** 1 year, 4 months ago

why not B?

👍 ↩ 🚩 upvoted 4 times

🗄️ 👤 **Sofia98** 1 year, 4 months ago

Why not B?

👍 ↩ 🚩 upvoted 3 times

🗄️ 👤 **tibuenoc** 1 year, 3 months ago

Because it's the best practice is separated zones. One zone for landing, raw and curated.

The answer B - has this part that excluded it "create a single zone to contain landing"

The correct answer is D

👍 ↩ 🚩 upvoted 2 times

🗄️ 👤 **Ed_Kim** 1 year, 4 months ago

Selected Answer: D

The answer is D

👍 ↩ 🚩 upvoted 2 times

🗄️ 👤 **e70ea9e** 1 year, 4 months ago

Selected Answer: C

Virtual Lake per Data Product: Each virtual lake acts as a self-contained domain for a specific data product, aligning with the data mesh principle of decentralized ownership and responsibility.

Team Autonomy: Teams have full control over their virtual lake, enabling independent development, management, and sharing of their data products.

👍 ↩ 🚩 upvoted 2 times



Platform

> Home

> Examtopics PRO

> All Exams

> Training Courses

