☰ MENU 🔍

← 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 269 DISCUSSION**

Actual exam question from Google's Professional Data Engineer

Question #: 269

Topic #: 1

[All Professional Data Engineer Questions]

Your organization's data assets are stored in BigQuery, Pub/Sub, and a PostgreSQL instance running on Compute Engine. Because there are multiple domains and diverse teams using the data, teams in your organization are unable to discover existing data assets. You need to design a solution to improve data discoverability while keeping development and configuration efforts to a minimum. What should you do?

A. Use Data Catalog to automatically catalog BigQuery datasets. Use Data Catalog APIs to manually catalog Pub/Sub topics and PostgreSQL tables.

B. Use Data Catalog to automatically catalog BigQuery datasets and Pub/Sub topics. Use Data Catalog APIs to manually catalog PostgreSQL tables.

C. Use Data Catalog to automatically catalog BigQuery datasets and Pub/Sub topics. Use custom connectors to manually catalog PostgreSQL tables.

D. Use customer connectors to manually catalog BigQuery datasets, Pub/Sub topics, and PostgreSQL tables.

Show Suggested Answer

by 👤 raaad at *Jan. 5, 2024, 6:08 p.m.*

Comments

Type your comment...

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

`Selected Answer: B`

- It utilizes Data Catalog's native support for both BigQuery datasets and Pub/Sub topics.
- For PostgreSQL tables running on a Compute Engine instance, you'd use Data Catalog APIs to create custom entries, as Data Catalog does not automatically discover external databases like PostgreSQL.

👍 ↩ 🏳 upvoted 15 times

⊟ 👤 **AllenChen123** 1 year, 3 months ago

Agree. https://cloud.google.com/data-catalog/docs/concepts/overview#catalog-non-google-cloud-assets

👍 ↩ 🏳 upvoted 5 times

⊟ 👤 **datapassionate** `Highly Voted 👍` 1 year, 3 months ago

`Selected Answer: C`

Data Catalog is the best choice. But for catalogging PostgreSQL it is better to use a connector when available, instead of using API.
https://cloud.google.com/data-catalog/docs/integrate-data-sources#integrate_unsupported_data_sources

👍 ↩ 🏳 upvoted 12 times

⊟ 👤 **7787de3** 7 months, 3 weeks ago

I agree. On the linked page:
If you can't find a connector for your data source, you can still manually integrate it by creating entry groups and custom entries.
As we can find a connector there, it should be used.

👍 ↩ 🏳 upvoted 1 times

⊟ 👤 **tibuenoc** 1 year, 3 months ago

Agree. If it doesn't have a connector, it must be manually built on the Data Catalog API.
As PostgreSQL already has a connector it's the best option is C

👍 ↩ 🏳 upvoted 4 times

⊟ 👤 **aaaaaaaasdasdasfs** `Most Recent ⊘` 3 weeks ago

`Selected Answer: B`

This is the correct approach. Data Catalog provides automatic discovery and metadata extraction for both BigQuery datasets and Pub/Sub topics as native Google Cloud services. You'll only need to use the Data Catalog APIs to manually catalog the PostgreSQL tables running on Compute Engine, as this is a non-native data source. This maximizes automation while minimizing development effort.

👍 ↩ 🏳 upvoted 1 times

⊟ 👤 **Abizi** 1 month, 1 week ago

`Selected Answer: C`

Why C is Correct?
BigQuery datasets →   Automatically cataloged in Data Catalog

Pub/Sub topics →   Automatically cataloged in Data Catalog

PostgreSQL on Compute Engine →   Not automatically cataloged

Requires a custom connector to extract metadata and push it to Data Catalog.

Option B (using Data Catalog APIs manually) is not enough because PostgreSQL metadata isn't natively supported.

Why Not B?
Option B suggests using Data Catalog APIs manually for PostgreSQL.

However, Data Catalog does not natively support PostgreSQL metadata extraction.

You need a custom connector to first extract PostgreSQL schema information, then push it to Data Catalog.

👍 ↩ 🏳 upvoted 1 times

⊟ 👤 **Pime13** 3 months, 4 weeks ago

`Selected Answer: A`

https://cloud.google.com/data-catalog/docs/concepts/overview#automatic_cataloging_of_assets
https://cloud.google.com/data-catalog/docs/concepts/overview#catalog-non-google-cloud-assets

👍 ↩ 🏳 upvoted 1 times

⊟ 👤 **AWSandeep** 4 months, 2 weeks ago

This section explains it clearly: https://cloud.google.com/data-catalog/docs/integrate-data-sources#integrate_unsupported_data_sources.

👍 ↩ ⚑ upvoted 2 times

☐ 👤 **baimus** 6 months, 4 weeks ago

This is C. To clarify some issues below with B, the links provided by supporters of B actually do say that it's preferable to use a community connector where available, and to only use the API when the case is genuinely not supported by community connectors.
In this case it's Postgresql, so it's supported, see here for full list: https://cloud.google.com/data-catalog/docs/integrate-data-sources#integrate_on-premises_data_sources

So this would be B if it was something like Q+ or some genuinely unsupported database, but postgres is supported for community connector.

👍 ↩ ⚑ upvoted 3 times

☐ 👤 **shanks_t** 8 months, 2 weeks ago

Data Catalog automatically catalogs metadata from Google Cloud sources such as BigQuery, Vertex AI, Pub/Sub, Spanner, Bigtable, and more.

To catalog metadata from non-Google Cloud systems in your organization, you can use the following:

Community-contributed connectors to multiple popular on-premises data sources
Manually build on the Data Catalog APIs for custom entries

👍 ↩ ⚑ upvoted 3 times

   ☐ 👤 **shanks_t** 8 months, 2 weeks ago

   C. While similar to B, using custom connectors for PostgreSQL might involve more development effort than using the Data Catalog APIs directly.

   👍 ↩ ⚑ upvoted 1 times

☐ 👤 **meh_33** 8 months, 4 weeks ago

raaad mostly correct and we can check his description supporting his answer so we can go with it .Cheers mate

👍 ↩ ⚑ upvoted 1 times

☐ 👤 **987af6b** 9 months, 1 week ago

I'm voting for C because the documentation states that Postgres is a custom connector developed by the community.

👍 ↩ ⚑ upvoted 2 times

   ☐ 👤 **987af6b** 9 months, 1 week ago

   Changed my mind. B.
   -This is not on premise, so the custom connector should not be applicable
   -Question says keep manual dev and config to a minimum

   👍 ↩ ⚑ upvoted 1 times

☐ 👤 **fitri001** 10 months, 3 weeks ago

BigQuery Datasets and Pub/Sub Topics: Google Data Catalog can automatically catalog metadata from BigQuery and Pub/Sub, making it easy to discover and manage these data assets without additional development effort.

PostgreSQL Tables: While Data Catalog does not have built-in connectors for PostgreSQL, you can use the Data Catalog APIs to manually catalog the PostgreSQL tables. This requires some custom development but is manageable compared to creating custom connectors for everything.

👍 ↩ ⚑ upvoted 4 times

☐ 👤 **virat_kohli** 11 months, 2 weeks ago

B. Use Data Catalog to automatically catalog BigQuery datasets and Pub/Sub topics. Use Data Catalog APIs to manually catalog PostgreSQL tables.

👍 ↩ ⚑ upvoted 2 times

☐ 👤 **Cassim** 11 months, 3 weeks ago

Option B leverages Data Catalog to automatically catalog BigQuery datasets and Pub/Sub topics, which streamlines the process and reduces manual effort. Using Data Catalog APIs to manually catalog PostgreSQL tables ensures consistency across all data assets while minimizing development and configuration efforts.

👍 ↩ 🏳 upvoted 2 times

⊟ 👤 **LaxmanTiwari** 1 year ago

Selected Answer: C

I vote for c as per Integrate on-premises data sources
To integrate on-premises data sources, you can use the corresponding Python connectors contributed by the community:

under the link

https://cloud.google.com/data-catalog/docs/integrate-data-sources

👍 ↩ 🏳 upvoted 3 times

⊟ 👤 **LaxmanTiwari** 1 year ago

data catalog api will come into effect if custom connectors are not available via community repos.

👍 ↩ 🏳 upvoted 1 times

⊟ 👤 **joao_01** 1 year ago

In the opction C, the expression "Use custom connectors to manually catalog PostgreSQL tables." is refering to the use case of Google when you want to use "Community-contributed connectors to multiple popular on-premises data sources". As you can see, this connectors are for ON-PREMISSES data sources ONLY. In this case the Postgres is in a VM in the cloud. Thus, the option correct is B.

👍 ↩ 🏳 upvoted 3 times

⊟ 👤 **joao_01** 1 year ago

Link: https://cloud.google.com/data-catalog/docs/concepts/overview#catalog-non-google-cloud-assets

👍 ↩ 🏳 upvoted 1 times

⊟ 👤 **hanoverquay** 1 year, 1 month ago

Selected Answer: B

option B, there's no need to build a custom connector now, postgreSQL is now supported
https://github.com/GoogleCloudPlatform/datacatalog-connectors-rdbms/tree/master/google-datacatalog-postgresql-connector

👍 ↩ 🏳 upvoted 2 times

⊟ 👤 **d11379b** 1 year, 1 month ago

I think "custom connector" here may just infer that this is not official tools? as the doc mentioned " connectors contributed by the community"
And should not be B as "manually catalog by API " this is a way even more basic than using connector

👍 ↩ 🏳 upvoted 1 times

⊟ 👤 **Y___ash** 1 year, 1 month ago

Selected Answer: B

Use Data Catalog to automatically catalog BigQuery datasets and Pub/Sub topics. Use Data Catalog APIs to manually catalog PostgreSQL tables.

👍 ↩ 🏳 upvoted 2 times

Load full discussion…