🔍

⊙ **Google Discussions**

📄 **EXAM PROFESSIONAL DATA ENGINEER TOPIC 1 QUESTION 303 DISCUSSION**

Actual exam question from Google's Professional Data Engineer

Question #: 303

Topic #: 1

[All Professional Data Engineer Questions]

You are managing a Dataplex environment with raw and curated zones. A data engineering team is uploading JSON and CSV files to a bucket asset in the curated zone but the files are not being automatically discovered by Dataplex. What should you do to ensure that the files are discovered by Dataplex?

A. Move the JSON and CSV files to the raw zone.

B. Enable auto-discovery of files for the curated zone.

C. Use the bg command-line tool to load the JSON and CSV files into BigQuery tables.

D. Grant object level access to the CSV and JSON files in Cloud Storage.

Show Suggested Answer

by 👤 **scaenruy** at *Jan. 4, 2024, 1:45 p.m.*

Comments

Type your comment...

Submit

⊟ 👤 **GCP001** `Highly Voted 👍` 1 year, 3 months ago

`Selected Answer: A`

Should be A. Curated zone need Parquet, Avro, ORC format not CSV or JSON. Check the ref - https://cloud.google.com/dataplex/docs/add-zone#curated-zones

👍 ↩ 🚩 upvoted 26 times

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

`Selected Answer: B`

- Auto-Discovery Feature: Dataplex has an auto-discovery feature that, when enabled, automatically discovers and catalogs data assets within a zone.
- Appropriate for Both Raw and Curated Zones: This feature is applicable to both raw and curated zones, and it should be tailored to the specific data governance and cataloging needs of the organization.

👍 ↩ 🚩 upvoted 10 times

☐ 👤 **cloud_rider** 5 months, 1 week ago

A is correct, Auto-Discovery features works on both curated and raw zones, but to keep JSON and CSV in curated zone, they must be kept along with the specification. Whereas in RAW zone, the discovery of these files happens even without specification file. refer to this link->> https://cloud.google.com/dataplex/docs/discover-data#discovery-configuration

👍 ↩ 🚩 upvoted 1 times

☐ 👤 **rajshiv** `Most Recent ⊘` 3 weeks, 2 days ago

`Selected Answer: B`

We can store JSON and CSV in the curated zone if those files represent curated data. Usually we store JSON/CSV files in the raw zone if they are straight from source. But nowhere in the question is any of that detail mentioned. So I think the correct answer is :
B - Dataplex automatically discovers and catalogs data in the zones only if auto-discovery is enabled for the zone or asset. In this scenario - JSON and CSV files are being uploaded to a curated zone, which is fine. But if files are not being discovered, it's likely because auto-discovery is not enabled for that zone.

👍 ↩ 🚩 upvoted 2 times

☐ 👤 **MBNR** 1 month, 3 weeks ago

`Selected Answer: A`

Answer is A
Data Format supported: Data in curated zones is typically columnar, Hive-partitioned, and stored in formats like Parquet, Avro, or ORC
Restrictions: Dataplex does NOT allow users to create CSV files within a "curated zone

👍 ↩ 🚩 upvoted 1 times

☐ 👤 **desertlotus1211** 1 month, 1 week ago

Auto-Discovery is the better option

👍 ↩ 🚩 upvoted 1 times

☐ 👤 **juliorevk** 3 months, 1 week ago

`Selected Answer: A`

- Raw zones store structured data, semi-structured data such as CSV files and JSON files, and unstructured data in any format from external sources. Raw zones are useful for staging raw data before performing any transformations. Data can be stored in Cloud Storage buckets or BigQuery datasets.
- Curated Zones do not support JSON / CSV

👍 ↩ 🚩 upvoted 1 times

☐ 👤 **Pime13** 3 months, 4 weeks ago

`Selected Answer: B`

Auto-discovery needs to be enabled for the curated zone to ensure that Dataplex can scan and register the files. You can configure this setting at the zone or asset level.
Option A, moving the JSON and CSV files to the raw zone, would not solve the issue of automatic discovery in the curated zone. The problem lies in the configuration of the curated zone, not the location of the files.

👍 ↩ 🚩 upvoted 3 times

☐ 👤 **SamuelTsch** 6 months ago

`Selected Answer: A`

Raw zones store structured data, semi-structured data such as CSV files and JSON files, and unstructured data in any format from external sources. Curated zones store structured data. Data can be stored in Cloud Storage buckets or BigQuery datasets. Supported formats for Cloud Storage buckets include Parquet, Avro, and ORC.

👍 ↩ 🚩 upvoted 1 times

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

`Selected Answer: B`

Discovery configuration
Discovery is enabled by default when you create a new zone or asset. You can disable Discovery at the zone or asset level.

For each Dataplex asset with Discovery enabled, Dataplex does the following:

Scans the data associated with the asset.
Groups structured and semi-structured files into tables.
Collects technical metadata, such as table name, schema, and partition definition.
For unstructured data, such as images and videos, Dataplex Discovery automatically detects and registers groups of files sharing media type as filesets. For example, if gs://images/group1 contains GIF images, and gs://images/group2 contains JPEG images, Dataplex Discovery detects and registers two filesets. For structured data, such as Avro, Discovery detects files only if they are located in folders that contain the same data format and schema.

Reference : https://cloud.google.com/dataplex/docs/discover-data#exclude-files-from-Discovery

👍 ↩ 🚩 **upvoted 3 times**

⊟ 👤 **hussain.sain** 10 months ago

Selected Answer: B

While JSON and CSV can technically be stored in curated zones, it is not a common practice due to the reasons mentioned above. no where in the mention link its mention that there is a restriction.

👍 ↩ 🚩 **upvoted 3 times**

⊟ 👤 **Anudeep58** 10 months, 3 weeks ago

Selected Answer: A

While none of the original options (A, B, C, or D) directly address the issue, the closest solution is:

Move the JSON and CSV files to a raw zone. (This was previously marked as the most voted option, but it's not ideal due to data organization disruption)
Here's why this approach might be necessary (but not ideal):

Dataplex curated zones currently don't support native processing of JSON and CSV formats. They are designed for structured data formats like Parquet, Avro, or ORC.

👍 ↩ 🚩 **upvoted 4 times**

⊟ 👤 **chrissamharris** 1 year ago

Selected Answer: A

Option A
https://cloud.google.com/dataplex/docs/add-zone#raw-zones

Raw zones are the only zones that support CSV & JSON

👍 ↩ 🚩 **upvoted 1 times**

⊟ 👤 **joao_01** 1 year ago

Its B guys, i encounter this in my job, and I had to do B to make it work

👍 ↩ 🚩 **upvoted 1 times**

⊟ 👤 **joao_01** 1 year ago

Actually I did this in a Raw zone, not Curated.

👍 ↩ 🚩 **upvoted 1 times**

⊟ 👤 **joao_01** 1 year ago

Its A :)

👍 ↩ 🚩 **upvoted 4 times**

⊟ 👤 **demoro86** 1 year, 2 months ago

Selected Answer: A

GCP001 agree with him

👍 ↩ 🚩 **upvoted 2 times**

⊟ 👤 **Moss2011** 1 year, 2 months ago

Selected Answer: A

The answer can be found reading a common config of Dataplex in this URL: https://medium.com/google-cloud/google-cloud-dataplex-part-1-lakes-zones-assets-and-discovery-5f288486cb2f

👍 ↩ 🚩 **upvoted 2 times**

⊟ 👤 **kck6ra4214wm** 1 year, 2 months ago

Selected Answer: A

Dataplex does not allow users to create CSV files within a "curated zone"

👍 ↩ 🚩 **upvoted 1 times**

⊟ 👤 **daidai75** 1 year, 2 months ago

Selected Answer: B

According to this URL : https://cloud.google.com/dataplex/docs/discover-data, the auto-discovery can support CSV and Json

According to this URL: https://cloud.google.com/dataplex/docs/discover-data, the auto-discovery can support CSV and Json in both Raw-Zone and Curated-Zone. I also open a console the verify it, both Raw and Curated zone can set up csv&json auto-discovery.

👍 ↩ 🏳 upvoted 2 times

⊟ 👤 **dungct** 1 year, 2 months ago

Selected Answer: B

Discovery raises the following administrator actions whenever data-related issues are detected during scans : Inconsistent data format in a table. For example, files of different formats exist with the same table prefix. Inconsistent data format in a table. For example, files of different formats exist with the same table prefix.

👍 ↩ 🏳 upvoted 4 times

⊟ 👤 **dungct** 1 year, 2 months ago

https://cloud.google.com/dataplex/docs/discover-data#invalid_data_format

👍 ↩ 🏳 upvoted 3 times

Load full discussion…