

Databricks Databricks-Certified-Data-Engineer-Associate Deutsch Prüfung, Databricks-Certified-Data-Engineer- Associate Echte Fragen



P.S. Kostenlose und neue Databricks-Certified-Data-Engineer-Associate Prüfungsfragen sind auf Google Drive freigegeben von Pass4Test verfügbar: https://drive.google.com/open?id=1fLZ0GOOwWvtSubGUgOWbg0K65TF8j_1

Jede Version der Databricks Databricks-Certified-Data-Engineer-Associate Prüfungsunterlagen von uns hat ihre eigene Überlegenheit. PDF Version hat keine Beschränkung für Anlage, deshalb können Sie irgendwo die Unterlagen lesen. Wenn Sie Internet benutzen können, die Online Test Engine der Databricks Databricks-Certified-Data-Engineer-Associate können Sie sowohl mit Windows, Mac als auch Android, iOS benutzen. Mit Simulations-Software können Sie die Prüfungsumwelt der Databricks Databricks-Certified-Data-Engineer-Associate erfahren und bessere Kenntnisse darüber erwerben. Übrigens, Sie dürfen die Prüfungssoftware irgendwie viele Male installieren.

Um sich auf die Databricks Certified Data Engineer Associate Prüfung vorzubereiten, können Kandidaten eine Reihe von Ressourcen nutzen, einschließlich Studienführern, Prüfungsübungen und Schulungskursen. Die Prüfung wird online durchgeführt und kann von überall auf der Welt aus abgelegt werden. Erfolgreiche Kandidaten erhalten ein digitales Abzeichen und ein Zertifikat, das sie verwenden können, um ihre Fähigkeiten und Kenntnisse potenziellen Arbeitgebern zu präsentieren.

Das GAQM Databricks-zertifizierte Data-Engineer-Associate (Databricks Certified Data Engineer Associate) ist eine herausfordernde und belohnende Zertifizierungsprüfung, mit der Dateningenieure ihre Fähigkeiten und Kenntnisse bei der Arbeit mit Datenbanken validieren können. Die Zertifizierung wird von führenden Organisationen anerkannt und kann Dateningenieuren helfen, ihre Karriere voranzutreiben. Wenn Sie ein Dateningenieur sind, der daran interessiert ist, mit Datenbanken zu arbeiten, ist diese

Zertifizierungsprüfung auf jeden Fall eine Überlegung wert.

>> **Databricks Databricks-Certified-Data-Engineer-Associate Deutsch Prüfung** <<

Databricks-Certified-Data-Engineer-Associate Echte Fragen - Databricks-Certified-Data-Engineer-Associate Kostenlos Downloaden

Egal wie anziehend die Werbung ist, ist nicht so überzeugend wie Ihre eigene Erfahrung. Auf unserer Webseite können Sie die Demo der Databricks Databricks-Certified-Data-Engineer-Associate Prüfungssoftware kostenlos herunterladen. Wir glauben, solange Sie diese Software, die vielen Leuten bei der Databricks Databricks-Certified-Data-Engineer-Associate geholfen hat, probiert haben, werden Sie diese Software sofort mögen. Benutzen Sie unsere Produkte! Sie können auch ein IT-Spezialist mit Databricks Databricks-Certified-Data-Engineer-Associate Prüfungszeugnis werden!

Databricks Certified Data Engineer Associate Exam Databricks-Certified-Data-Engineer-Associate Prüfungsfragen mit Lösungen (Q57-Q62):

57. Frage

A Data Engineer is building a simple data pipeline using Lakeflow Declarative Pipelines (LDP) in Databricks to ingest customer data. The raw customer data is stored in a cloud storage location in JSON format. The task is to create Lakeflow Declarative Pipelines that read the raw JSON data and write it into a Delta table for further processing.

Which code snippet will correctly ingest the raw JSON data and create a Delta table using LDP?

- A. `import dlt`
`@dlt.view`
`def raw_customers():`
`return spark.format.json("s3://my-bucket/raw-customers/")`
- B. `import dlt`
`@dlt.table`
`def raw_customers():`
`return spark.read.format("csv").load("s3://my-bucket/raw-customers/")`
- C. `import dlt`
`@dlt.table`
`def raw_customers():`
`return spark.read.json("s3://my-bucket/raw-customers/")`
- D. `import dlt`
`@dlt.table`
`def raw_customers():`
`return spark.read.format("parquet").load("s3://my-bucket/raw-customers/")`

Antwort: C

Begründung:

The correct method to define a table using Lakeflow Declarative Pipelines (LDP) is with the `@dlt.table` decorator, which persists the output as a managed Delta table. When ingesting raw JSON data, `spark.read`.

`json()` or `spark.read.format("json").load()` is the standard approach. This reads JSON-formatted files from the source and stores them in Delta format automatically managed by Databricks.

Reference Source: Databricks Lakeflow Declarative Pipelines Developer Guide - "Create tables from raw JSON and Delta sources."

58. Frage

A data engineer needs to apply custom logic to string column city in table stores for a specific use case. In order to apply this custom logic at scale, the data engineer wants to create a SQL user-defined function (UDF).

Which of the following code blocks creates this SQL UDF?

• A.

```
CREATE UDF combine_nyc(city STRING)
RETURN CASE
  WHEN city = "brooklyn" THEN "new york"
  ELSE city
END;
```

• B.

```
CREATE FUNCTION combine_nyc(city STRING)
RETURN CASE
  WHEN city = "brooklyn" THEN "new york"
  ELSE city
END;
```

• C.

```
CREATE UDF combine_nyc(city STRING)
RETURNS STRING
RETURN CASE
  WHEN city = "brooklyn" THEN "new york"
  ELSE city
END;
```

• D.

```
CREATE FUNCTION combine_nyc(city STRING)
RETURNS STRING
RETURN CASE
  WHEN city = "brooklyn" THEN "new york"
  ELSE city
END;
```

• E.

```
CREATE UDF combine_nyc(city STRING)
RETURNS STRING
CASE
  WHEN city = "brooklyn" THEN "new york"
  ELSE city
END;
```

Antwort: D

Begründung:

<https://www.databricks.com/blog/2021/10/20/introducing-sql-user-defined-functions.html>

59. Frage

A data engineer wants to schedule their Databricks SQL dashboard to refresh every hour, but they only want the associated SQL endpoint to be running when it is necessary. The dashboard has multiple queries on multiple datasets associated with it. The data that feeds the dashboard is automatically processed using a Databricks Job.

Which of the following approaches can the data engineer use to minimize the total running time of the SQL endpoint used in the refresh schedule of their dashboard?

- A. They can turn on the Auto Stop feature for the SQL endpoint.
- B. They can ensure the dashboard's SQL endpoint matches each of the queries' SQL endpoints.
- C. They can set up the dashboard's SQL endpoint to be serverless.
- D. They can reduce the cluster size of the SQL endpoint.
- E. They can ensure the dashboard's SQL endpoint is not one of the included query's SQL endpoint.

Antwort: A

Begründung:

The Auto Stop feature allows the SQL endpoint to automatically stop after a specified period of inactivity.

This can help reduce the cost and resource consumption of the SQL endpoint, as it will only run when it is needed to refresh the dashboard or execute queries. The data engineer can configure the Auto Stop setting for the SQL endpoint from the SQL Endpoints UI, by selecting the desired idle time from the Auto Stop dropdown menu. The default idle time is 120 minutes, but it can be set to as low as 15 minutes or as high as

240 minutes. Alternatively, the data engineer can also use the SQL Endpoints REST API to set the Auto Stop setting programmatically. References: SQL Endpoints UI, SQL Endpoints REST API, Refreshing SQL Dashboard

60. Frage

A dataset has been defined using Delta Live Tables and includes an expectations clause:

CONSTRAINT valid_timestamp EXPECT (timestamp > '2020-01-01') ON VIOLATION FAIL UPDATE What is the expected behavior when a batch of data containing data that violates these constraints is processed?

- A. Records that violate the expectation cause the job to fail.
- B. Records that violate the expectation are dropped from the target dataset and loaded into a quarantine table.
- C. Records that violate the expectation are dropped from the target dataset and recorded as invalid in the event log.
- D. Records that violate the expectation are added to the target dataset and recorded as invalid in the event log.
- E. Records that violate the expectation are added to the target dataset and flagged as invalid in a field added to the target dataset.

Antwort: A

Begründung:

Explanation

<https://docs.databricks.com/en/delta-live-tables/expectations.html>

Action

Result

warn (default)

Invalid records are written to the target; failure is reported as a metric for the dataset.

drop

Invalid records are dropped before data is written to the target; failure is reported as a metrics for the dataset.

fail

Invalid records prevent the update from succeeding. Manual intervention is required before re-processing.

61. Frage

A data engineer has configured a Structured Streaming job to read from a table, manipulate the data, and then perform a streaming write into a new table.

The code block used by the data engineer is below:

```
(spark.readStream
  .table("sales")
  .withColumn("avg_price", col("sales") / col("units"))
  .writeStream
  .option("checkpointLocation", checkpointPath)
  .outputMode("complete")
  ._____
  .table("new_sales")
)
```

If the data engineer only wants the query to process all of the available data in as many batches as required, which of the following lines of code should the data engineer use to fill in the blank?

- A. trigger(processingTime="once")
- B. processingTime(1)
- C. trigger(continuous="once")
- D. trigger(parallelBatch=True)
- E. trigger(availableNow=True)

Antwort: E

Begründung:

<https://spark.apache.org/docs/latest/api/python/reference/pyspark.ss/api/pyspark.sql.streaming.DataStreamWriter>

• • • • •

Databricks-Certified-Data-Engineer-Associate Echte Fragen: <https://www.pass4test.de/Databricks-Certified-Data-Engineer-Associate.html>

- [illegible]

myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt,
myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, liberationmeditation.org, Disposable vapes

Außerdem sind jetzt einige Teile dieser Pass4Test Databricks-Certified-Data-Engineer-Associate Prüfungsfragen kostenlos
erhältlich: https://drive.google.com/open?id=1fLZ0GOOwWvtSubGUgOWbg0K65TF8i_1