

Databricks-Certified-Professional-Data-Engineer Prüfung, Databricks-Certified-Professional-Data-Engineer Prüfungen



BONUS!!! Laden Sie die vollständige Version der Zertpruefung Databricks-Certified-Professional-Data-Engineer Prüfungsfragen kostenlos herunter: <https://drive.google.com/open?id=1rIJ9-8d1ZyIHyetrlf8Akj9vLOHFB8Od>

Wollen Sie Databricks Databricks-Certified-Professional-Data-Engineer Zertifizierungsprüfung ablegen? Wollen Sie die Databricks Databricks-Certified-Professional-Data-Engineer Zertifizierung bekommen? Wie können Sie ohne sehr gute Vorbereitung diese Prüfung ablegen? Tatsächlich gibt es eine Weise für Sie, in sehr beschränkter Zeit die Databricks Databricks-Certified-Professional-Data-Engineer Prüfung leicht zu bestehen. Was können Sie machen? Es ist erreichbar, dass Sie die Databricks Databricks-Certified-Professional-Data-Engineer Dumps von Zertpruefung benutzen.

Viele IT-Fachleute wollen Databricks Databricks-Certified-Professional-Data-Engineer Zertifikate erhalten. Die IT-Zertifikate werden Ihnen helfen, in der IT-Branche befördert zu werden. Das Databricks Databricks-Certified-Professional-Data-Engineer Zertifikat ist ein beliebtes unter den vielen Zertifikaten. Obwohl es nicht so leicht ist, die Databricks Databricks-Certified-Professional-Data-Engineer Zertifizierungsprüfung zu bestehen, gibt es doch Methoden. Sie können viel Zeit und Energie für die Prüfung benutzen, um Ihr Know-How zu konsolidieren, oder an den effizienten Kursen teilnehmen. Die speziellen Simulationsprüfungen von Zertpruefung, die Ihnen viel Zeit und Energie ersparen und Ihr Ziel erreichen können, ist sehr effizient. Zertpruefung ist eine gute Wahl für Sie.

>> **Databricks-Certified-Professional-Data-Engineer Prüfung** <<

Databricks Databricks-Certified-Professional-Data-Engineer Prüfungen, Databricks-Certified-Professional-Data-Engineer Prüfungsunterlagen

Ea ist Traum der Angestellten, sich in der IT-Branche engagieren zu können, die Databricks Databricks-Certified-Professional-Data-Engineer Zertifizierungsprüfung zu bestehen. Wenn Sie Ihren Traum verwirklichen wollen, brauchen Sie nur fachliche Ausbildung zu wählen. Zertpruefung ist eine fachliche Website, die Schulungsunterlagen zur Databricks Databricks-Certified-Professional-Data-Engineer Zertifizierung bietet. Wählen Sie Zertpruefung. Und wir versprechen, dass Sie den Erfolg erlangen und Ihren Traum verwirklichen, egal welches hohes Ziel Sie anstreben, können.

Databricks Certified Professional Data Engineer Exam Databricks-Certified-Professional-Data-Engineer Prüfungsfragen mit Lösungen (Q12-Q17):

12. Frage

A production workload incrementally applies updates from an external Change Data Capture feed to a Delta Lake table as an always-on Structured Stream job. When data was initially migrated for this table, OPTIMIZE was executed and most data files were resized to 1 GB. Auto Optimize and Auto Compaction were both turned on for the streaming production job. Recent review of data files shows that most data files are under 64 MB, although each partition in the table contains at least 1 GB of data and the total table size is over 10 TB.

Which of the following likely explains these smaller file sizes?

- A. Databricks has autotuned to a smaller target file size based on the overall size of data in the table
- B. Databricks has autotuned to a smaller target file size based on the amount of data in each partition
- **C. Databricks has autotuned to a smaller target file size to reduce duration of MERGE operations**
- D. Z-order indices calculated on the table are preventing file compaction C Bloom filter indices calculated on the table are preventing file compaction

Antwort: C

Begründung:

This is the correct answer because Databricks has a feature called Auto Optimize, which automatically optimizes the layout of Delta Lake tables by coalescing small files into larger ones and sorting data within each file by a specified column. However, Auto Optimize also considers the trade-off between file size and merge performance, and may choose a smaller target file size to reduce the duration of merge operations, especially for streaming workloads that frequently update existing records. Therefore, it is possible that Auto Optimize has autotuned to a smaller target file size based on the characteristics of the streaming production job. Verified References: [Databricks Certified Data Engineer Professional], under "Delta Lake" section; Databricks Documentation, under "Auto Optimize" section.

<https://docs.databricks.com/en/delta/tune-file-size.html#autotune-table> 'Autotune file size based on workload'

13. Frage

A junior data engineer has configured a workload that posts the following JSON to the Databricks REST API endpoint `2.0/jobs/create`.

```
{
  "name": "Ingest new data",
  "existing_cluster_id": "6015-954420-peace720",
  "notebook_task": {
    "notebook_path": "/Prod/ingest.py"
  }
}
```

Assuming that all configurations and referenced resources are available, which statement describes the result of executing this workload three times?

- A. Three new jobs named "Ingest new data" will be defined in the workspace, but no jobs will be executed.
- B. The logic defined in the referenced notebook will be executed three times on new clusters with the configurations of the provided cluster ID.
- **C. The logic defined in the referenced notebook will be executed three times on the referenced existing all purpose cluster.**
- D. Three new jobs named "Ingest new data" will be defined in the workspace, and they will each run once daily.
- E. One new job named "Ingest new data" will be defined in the workspace, but it will not be executed.

Antwort: C

Begründung:

This is the correct answer because the JSON posted to the Databricks REST API endpoint `2.0/jobs/create` defines a new job with a name, an existing cluster id, and a notebook task. However, it does not specify any schedule or trigger for the job execution. Therefore, three new jobs with the same name and configuration will be created in the workspace, but none of them will be executed until they are manually triggered or scheduled.

Verified References: [Databricks Certified Data Engineer Professional], under "Monitoring & Logging" section; [Databricks Documentation], under "Jobs API - Create" section.

14. Frage

The data engineering team maintains a table of aggregate statistics through batch nightly updates. This includes total sales for the previous day alongside totals and averages for a variety of time periods including the 7 previous days, year-to-date, and quarter-to-date. This table is named `store_sales_summary` and the schema is as follows:

```
store_id INT, total_sales_qtd FLOAT, avg_daily_sales_qtd FLOAT, total_sales_ytd FLOAT,
avg_daily_sales_ytd FLOAT, previous_day_sales FLOAT, total_sales_7d FLOAT, avg_daily_sales_7d
FLOAT, updated TIMESTAMP
```

The table `daily_store_sales` contains all the information needed to update `store_sales_summary`. The schema for this table is:

`store_id` INT, `sales_date` DATE, `total_sales` FLOAT

If `daily_store_sales` is implemented as a Type 1 table and the `total_sales` column might be adjusted after manual data auditing, which approach is the safest to generate accurate reports in the `store_sales_summary` table?

- A. Implement the appropriate aggregate logic as a batch read against the `daily_store_sales` table and append new rows nightly to the `store_sales_summary` table.
- **B. Use Structured Streaming to subscribe to the change data feed for `daily_store_sales` and apply changes to the aggregates in the `store_sales_summary` table with each update.**
- C. Implement the appropriate aggregate logic as a Structured Streaming read against the `daily_store_sales` table and use upsert logic to update results in the `store_sales_summary` table.
- D. Implement the appropriate aggregate logic as a batch read against the `daily_store_sales` table and use upsert logic to update results in the `store_sales_summary` table.
- E. Implement the appropriate aggregate logic as a batch read against the `daily_store_sales` table and overwrite the `store_sales_summary` table with each Update.

Antwort: B

Begründung:

The `daily_store_sales` table contains all the information needed to update `store_sales_summary`. The schema of the table is:

`store_id` INT, `sales_date` DATE, `total_sales` FLOAT

The `daily_store_sales` table is implemented as a Type 1 table, which means that old values are overwritten by new values and no history is maintained. The `total_sales` column might be adjusted after manual data auditing, which means that the data in the table may change over time.

The safest approach to generate accurate reports in the `store_sales_summary` table is to use Structured Streaming to subscribe to the change data feed for `daily_store_sales` and apply changes to the aggregates in the `store_sales_summary` table with each update.

Structured Streaming is a scalable and fault-tolerant stream processing engine built on Spark SQL. Structured Streaming allows processing data streams as if they were tables or DataFrames, using familiar operations such as select, filter, groupBy, or join. Structured Streaming also supports output modes that specify how to write the results of a streaming query to a sink, such as append, update, or complete. Structured Streaming can handle both streaming and batch data sources in a unified manner.

The change data feed is a feature of Delta Lake that provides structured streaming sources that can subscribe to changes made to a Delta Lake table. The change data feed captures both data changes and schema changes as ordered events that can be processed by downstream applications or services. The change data feed can be configured with different options, such as starting from a specific version or timestamp, filtering by operation type or partition values, or excluding no-op changes.

By using Structured Streaming to subscribe to the change data feed for `daily_store_sales`, one can capture and process any changes made to the `total_sales` column due to manual data auditing. By applying these changes to the aggregates in the `store_sales_summary` table with each update, one can ensure that the reports are always consistent and accurate with the latest data. Verified References:

[Databricks Certified Data Engineer Professional], under "Spark Core" section; Databricks Documentation, under "Structured Streaming" section; Databricks Documentation, under "Delta Change Data Feed" section.

15. Frage

Spill occurs as a result of executing various wide transformations. However, diagnosing spill requires one to proactively look for key indicators.

Where in the Spark UI are two of the primary indicators that a partition is spilling to disk?

- A. Stage's detail screen and Executor's files
- B. Driver's and Executor's log files
- **C. Stage's detail screen and Query's detail screen**
- D. Executor's detail screen and Executor's log files

Antwort: C

Begründung:

In Apache Spark's UI, indicators of data spilling to disk during the execution of wide transformations can be found in the Stage's detail screen and the Query's detail screen. These screens provide detailed metrics about each stage of a Spark job, including information about memory usage and spill data. If a task is spilling data to disk, it indicates that the data being processed exceeds the available memory, causing Spark to spill data to disk to free up memory. This is an important performance metric as excessive spill can significantly slow down the processing.

Reference:

Apache Spark Monitoring and Instrumentation: Spark Monitoring Guide

16. Frage

A data engineer is using Lakeflow Declarative Pipelines Expectations feature to track the data quality of their incoming sensor data. Periodically, sensors send bad readings that are out of range, and they are currently flagging those rows with a warning and writing them to the silver table along with the good data. They've been given a new requirement - the bad rows need to be quarantined in a separate quarantine table and no longer included in the silver table.

This is the existing code for their silver table:

```
@dlt.table
@dlt.expect("valid_sensor_reading", "reading < 120")
def silver_sensor_readings():
    return spark.readStream.table("bronze_sensor_readings")
```

What code will satisfy the requirements?

- A. `@dlt.table`
`@dlt.expect_or_drop("valid_sensor_reading", "reading < 120")`
`def silver_sensor_readings():`
 `return spark.readStream.table("bronze_sensor_readings")`
`@dlt.table`
`@dlt.expect_or_drop("invalid_sensor_reading", "reading >= 120")`
`def quarantine_sensor_readings():`
 `return spark.readStream.table("bronze_sensor_readings")`
- B. `@dlt.table`
`@dlt.expect("valid_sensor_reading", "reading < 120")`
`def silver_sensor_readings():`
 `return spark.readStream.table("bronze_sensor_readings")`
`@dlt.table`
`@dlt.expect("invalid_sensor_reading", "reading >= 120")`
`def quarantine_sensor_readings():`
 `return spark.readStream.table("bronze_sensor_readings")`
- C. `@dlt.table`
`@dlt.expect_or_drop("valid_sensor_reading", "reading < 120")`
`def silver_sensor_readings():`
 `return spark.readStream.table("bronze_sensor_readings")`
`@dlt.table`
`@dlt.expect("invalid_sensor_reading", "reading >= 120")`
`def quarantine_sensor_readings():`
 `return spark.readStream.table("bronze_sensor_readings")`
- D. `@dlt.table`
`@dlt.expect_or_drop("valid_sensor_reading", "reading < 120")`
`def silver_sensor_readings():`
 `return spark.readStream.table("bronze_sensor_readings")`
`@dlt.table`
`@dlt.expect("invalid_sensor_reading", "reading < 120")`
`def quarantine_sensor_readings():`
 `return spark.readStream.table("bronze_sensor_readings")`

Antwort: B

Begründung:

Comprehensive and Detailed Explanation from Databricks Documentation:

Lakeflow Declarative Pipelines (DLT) supports data quality enforcement using `@dlt.expect`, `@dlt.expect_or_drop`, and `@dlt.expect_all`.

`@dlt.expect` applies a rule and records whether rows pass or fail the condition but does not drop failing rows. Instead, failing rows can be written to a quarantine table.

`@dlt.expect_or_drop` enforces that only rows passing the condition flow downstream, dropping bad records automatically.

In this case, the requirement is:

Good rows (reading < 120) go to the silver table.

Bad rows (reading >= 120) go to a quarantine table.

Bad rows should not be included in silver.

The correct implementation is Option A, where:

The silver table uses `@dlt.expect` to validate reading < 120 . These rows flow normally.

The quarantine table applies an expectation for reading ≥ 120 , ensuring bad records are captured separately.

Other options are incorrect:

Option B/D: These either use `expect_or_drop` incorrectly or apply wrong conditions, leading to dropped rows without quarantining properly.

Option C: Uses `expect_or_drop` for both tables, which would discard bad rows instead of persisting them into a quarantine table.

Thus, Option A meets the business requirement to split good and bad data streams while ensuring both are captured for auditing and processing.

17. Frage

.....

Unser Zertpruefung ist eine fachliche Website, die Prüfungsmaterialien für zahlreiche Databricks Databricks-Certified-Professional-Data-Engineer Zertifizierungsprüfung bieten. Unser Zertpruefung wird vielen IT-Fachleuten zum Berufsaufstieg verhelfen. Die Kraft unserer Eliteteams ist unglaublich. Sie können die Databricks Databricks-Certified-Professional-Data-Engineer Examensübungen- und antworten teilweise als Probe umsonst herunterladen, so dass Sie die Glaubwürdigkeit vom Zertpruefung testen können.

Databricks-Certified-Professional-Data-Engineer Prüfungen: https://www.zertpruefung.de/Databricks-Certified-Professional-Data-Engineer_exam.html

Zertpruefung verspricht, dass Sie die Databricks Databricks-Certified-Professional-Data-Engineer Zertifizierungsprüfung 100% zum ersten Mal bestehen können, Die Databricks Databricks-Certified-Professional-Data-Engineer Zertifizierungsprüfung ist heutzutage sehr beliebt, Zertpruefung Databricks-Certified-Professional-Data-Engineer Prüfungen ist Ihnen die beste Wahl, Databricks Databricks-Certified-Professional-Data-Engineer Prüfung Erfahrenes Team von zertifizierten Fachleuten, Als weltweiter Führer im Bereich der Databricks-Certified-Professional-Data-Engineer Übungswerkstätte, sind wir determiniert, unseren Kunden einen verständlichen Service anzubieten und einen integrierten Service aufzubauen.

Edward nickte nur und schaute mir eindringlich in die Augen, Der Anwalt zeigte offen seine Ironie, Zertpruefung verspricht, dass Sie die Databricks Databricks-Certified-Professional-Data-Engineer Zertifizierungsprüfung 100% zum ersten Mal bestehen können.

Kostenlos Databricks-Certified-Professional-Data-Engineer dumps torrent & Databricks Databricks-Certified-Professional-Data-Engineer Prüfung prep & Databricks-Certified-Professional-Data-Engineer examcollection braindumps

Die Databricks Databricks-Certified-Professional-Data-Engineer Zertifizierungsprüfung ist heutzutage sehr beliebt, Zertpruefung ist Ihnen die beste Wahl, Erfahrenes Team von zertifizierten Fachleuten, Als weltweiter Führer im Bereich der Databricks-Certified-Professional-Data-Engineer Übungswerkstätte, sind wir determiniert, unseren Kunden einen verständlichen Service anzubieten und einen integrierten Service aufzubauen.

- Databricks-Certified-Professional-Data-Engineer Studienmaterialien: Databricks Certified Professional Data Engineer Exam - Databricks-Certified-Professional-Data-Engineer Torrent Prüfung - Databricks-Certified-Professional-Data-Engineer wirkliche Prüfung ☐ Öffnen Sie die Webseite **【 www.it-pruefung.com 】** und suchen Sie nach kostenloser Download von [Databricks-Certified-Professional-Data-Engineer] ☐ Databricks-Certified-Professional-Data-Engineer Unterlage
- Databricks-Certified-Professional-Data-Engineer Zertifizierungsantworten ☐ Databricks-Certified-Professional-Data-Engineer Fragen&Antworten ☐ Databricks-Certified-Professional-Data-Engineer Deutsche Prüfungsfragen ☐ Sie müssen nur zu ➡ www.itcert.com ☐ gehen um nach kostenloser Download von (Databricks-Certified-Professional-Data-Engineer) zu suchen ☐ Databricks-Certified-Professional-Data-Engineer Deutsch
- Databricks-Certified-Professional-Data-Engineer Prüfungen ☐ Databricks-Certified-Professional-Data-Engineer Testking ☐ Databricks-Certified-Professional-Data-Engineer Online Tests ☐ Suchen Sie jetzt auf ☐ www.zertpruefung.ch ☐ nach (Databricks-Certified-Professional-Data-Engineer) und laden Sie es kostenlos herunter ☐ Databricks-Certified-Professional-Data-Engineer Prüfungsmaterialien
- Databricks-Certified-Professional-Data-Engineer Studienmaterialien: Databricks Certified Professional Data Engineer Exam - Databricks-Certified-Professional-Data-Engineer Torrent Prüfung - Databricks-Certified-Professional-Data-Engineer wirkliche Prüfung ☐ Öffnen Sie die Website [www.itcert.com] Suchen Sie ➤ Databricks-Certified-Professional-Data-Engineer ☐ Kostenloser Download ☐ Databricks-Certified-Professional-Data-Engineer Deutsch Prüfung
- Neuester und gültiger Databricks-Certified-Professional-Data-Engineer Test VCE Motoren-Dumps und Databricks-Certified-Professional-Data-Engineer neueste Testfragen für die IT-Prüfungen ☐ Suchen Sie auf der Webseite ➡ www.deutschpruefung.com ☐ nach ➡ Databricks-Certified-Professional-Data-Engineer ☐ und laden Sie es kostenlos

- [illegible]

Außerdem sind jetzt einige Teile dieser Zertprüfung Databricks-Certified-Professional-Data-Engineer Prüfungsfragen kostenlos erhältlich: <https://drive.google.com/open?id=1rIJ-8d1ZyIHvyr1f8Akj9vLOHFB8Od>