

Databricks-Certified-Professional-Data-Engineer Testfagen & Databricks-Certified-Professional-Data-Engineer Echte Fragen



2025 Die neuesten Zertpruefung Databricks-Certified-Professional-Data-Engineer PDF- Versionen Prüfungsfragen und Databricks-Certified-Professional-Data-Engineer Fragen und Antworten sind kostenlos verfügbar: https://drive.google.com/open?id=1uX__75M2dLnGjTfjZarUQGrYeSFux3

Sich für IT-Branche interessierend Sie bereiten sich jetzt auf die wichtige Databricks Databricks-Certified-Professional-Data-Engineer Prüfung? Lassen wir Zertpruefung Ihnen helfen! Was wir Ihnen garantieren ist, dass Sie nicht nur die Databricks Databricks-Certified-Professional-Data-Engineer Prüfung bestehen können, sondern auch Sie der leichte Vorbereitungsprozess und guter Kundendienst genießen.

Die Datenbank für Datenbanken misst die Fähigkeit eines Kandidaten, Datenpipelines mithilfe von Datenbäumen zu entwerfen, zu erstellen und zu verwalten. Es deckt eine breite Palette von Themen ab, einschließlich Datenaufnahme, Transformation, Speicherung und Analyse. Die Kandidaten müssen ihre Kenntnisse bei der Verwendung von Datenbank-Tools und -Techniken demonstrieren, um reale Daten technische Probleme zu lösen. Diese Zertifizierungsprüfung ist ideal für Dateningenieure, die ihre Fähigkeiten und Fachkenntnisse bei der Verwendung von Datenbanken zum Erstellen und Verwalten von Datenpipelines validieren möchten.

Databricks ist eine Cloud-basierte Plattform, die Datenentwicklung, maschinelles Lernen und Analysedienste anbietet. Die DCPDE - Prüfung (Databricks Certified Professional Data Engineer) wurde entwickelt, um die Fähigkeiten und Kenntnisse von Dateningenieuren zu testen, die für das Erstellen von Datenpipelines, die Verwaltung von Dateninfrastrukturen und die Optimierung von Datenverarbeitungsworkflows auf der Datenbankplattform verantwortlich sind.

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

Databricks-Certified-Professional-Data-Engineer Pass Dumps & PassGuide Databricks-Certified-Professional-Data-Engineer Prüfung & Databricks-Certified-Professional-Data-Engineer Guide

Die Databricks Databricks-Certified-Professional-Data-Engineer Zertifizierung ist den IT-Fachleute eine unentbehrliche Prüfung, weil sie ihres Schicksal bestimmt. Die Fragenkataloge zur Databricks Databricks-Certified-Professional-Data-Engineer Prüfung brauchen alle Kandidaten. Mit ihr kann der Kandidat sich gut auf die Databricks-Certified-Professional-Data-Engineer Prüfung vorbereiten und nicht so sehr unter Druck stehen. Und die Fragenkataloge in Zertpruefung sind einzigartig. Mit ihr können Sie die Databricks Databricks-Certified-Professional-Data-Engineer Prüfung ganz mühlos bestehen.

Die Prüfung ist ideal für Daten-Ingenieure, die ihre Expertise im Entwurf, Aufbau und der Wartung von Big-Data-Verarbeitungspipelines mit Databricks demonstrieren möchten. Die Zertifizierungsprüfung ist auf die spezifischen Rollen und Verantwortlichkeiten eines Daten-Ingenieurs zugeschnitten und umfasst eine Reihe von Themen, einschließlich Daten-Erfassung, Daten-Transformationen, Daten-Speicherung und Daten-Analyse. Durch das Bestehen der Prüfung können Kandidaten ihre Kompetenz im Umgang mit Databricks bei der Lösung komplexer Big-Data-Herausforderungen demonstrieren.

Databricks Certified Professional Data Engineer Exam Databricks-Certified-

Professional-Data-Engineer Prüfungsfragen mit Lösungen (Q96-Q101):

96. Frage

Which of the following statements are true about a lakehouse?

- A. Lakehouse does not support ACID
- B. Lakehouse only supports end-to-end streaming workloads and Data warehouses support Batch workloads
- C. Lakehouse do not support SQL
- D. Lakehouse only supports Machine learning workloads and Data warehouses support BI workloads
- E. Lakehouse supports Transactions

Antwort: E

Begründung:

Explanation

What Is a Lakehouse? - The Databricks Blog

Text Description automatically generated

A lakehouse has the following key features:

- **Transaction support:** In an enterprise lakehouse many data pipelines will often be reading and writing data concurrently. Support for ACID transactions ensures consistency as multiple parties concurrently read or write data, typically using SQL.
- **Schema enforcement and governance:** The Lakehouse should have a way to support schema enforcement and evolution, supporting DW schema architectures such as star/snowflake-schemas. The system should be able to reason about data integrity, and it should have robust governance and auditing mechanisms.
- **BI support:** Lakehouses enable using BI tools directly on the source data. This reduces staleness and improves recency, reduces latency, and lowers the cost of having to operationalize two copies of the data in both a data lake and a warehouse.
- **Storage is decoupled from compute:** In practice this means storage and compute use separate clusters, thus these systems are able to scale to many more concurrent users and larger data sizes. Some modern data warehouses also have this property.
- **Openness:** The storage formats they use are open and standardized, such as Parquet, and they provide an API so a variety of tools and engines, including machine learning and Python/R libraries, can efficiently access the data directly.
- **Support for diverse data types ranging from unstructured to structured data:** The lakehouse can be used to store, refine, analyze, and access data types needed for many new data applications, including images, video, audio, semi-structured data, and text.
- **Support for diverse workloads:** including data science, machine learning, and SQL and analytics. Multiple tools might be needed to support all these workloads but they all rely on the same data repository.
- **End-to-end streaming:** Real-time reports are the norm in many enterprises. Support for streaming eliminates the need for separate systems dedicated to serving real-time data applications.

97. Frage

A data engineer is using Auto Loader to read incoming JSON data as it arrives. They have configured Auto Loader to quarantine invalid JSON records but notice that over time, some records are being quarantined even though they are well-formed JSON.

The code snippet is:

```
df= (spark.readStream
```

```
.format("cloudFiles")
.option("cloudFiles.format", "json")
.option("badRecordsPath", "/tmp/somewhere/badRecordsPath")
.schema("a int, b int")
.load("/Volumes/catalog/schema/raw_data/"))
```

What is the cause of the missing data?

- A. The source data is valid JSON but does not conform to the defined schema in some way.
- B. The badRecordsPath location is accumulating many small files.
- C. At some point, the upstream data provider switched everything to multi-line JSON.
- D. The engineer forgot to set the option "cloudFiles.quarantineMode" = "rescue".

Antwort: A

Begründung:

Comprehensive and Detailed Explanation From Exact Extract of Databricks Data Engineer Documents:

Databricks Auto Loader quarantines records that fail schema validation, even if they are syntactically valid JSON. The documentation explains that "records that do not match the defined schema are considered corrupt and written to the bad records path." In this case, the data likely contains valid JSON fields whose structure or data types differ from the declared schema (a int, b int). For example, if a field expected as an integer arrives as a string, Auto Loader will quarantine it. The presence of badRecordsPath confirms this behavior. The cloudFiles.quarantineMode option controls quarantine behavior but is unrelated to valid data that fails schema validation. Thus, the correct cause is C.

98. Frage

The business intelligence team has a dashboard configured to track various summary metrics for retail stores.

This includes total sales for the previous day alongside totals and averages for a variety of time periods. The fields required to populate this dashboard have the following schema:

```
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
```

For Demand forecasting, the Lakehouse contains a validated table of all itemized sales updated incrementally in near real-time. This table named products_per_order, includes the following fields:

```
store_id INT, order_id INT, product_id INT, quantity INT, price FLOAT,
order_timestamp TIMESTAMP
```

Because reporting on long-term sales trends is less volatile, analysts using the new dashboard only require data to be refreshed once daily. Because the dashboard will be queried interactively by many users throughout a normal business day, it should return results quickly and reduce total compute associated with each materialization.

Which solution meets the expectations of the end users while controlling and limiting possible costs?

- A. Use Structure Streaming to configure a live dashboard against the products_per_order table within a Databricks notebook.
- B. Use the Delta Cache to persist the products_per_order table in memory to quickly refresh the dashboard with each query.
- C. Define a view against the products_per_order table and define the dashboard against this view.
- D. Populate the dashboard by configuring a nightly batch job to save the required data to quickly update the dashboard with each query.

Antwort: D

99. Frage

The business reporting team requires that data for their dashboards be updated every hour. The total processing time for the pipeline that extracts, transforms and loads the data for their pipeline runs in 10 minutes.

Assuming normal operating conditions, which configuration will meet their service-level agreement requirements with the lowest cost?

- A. Schedule a job to execute the pipeline once an hour on a new job cluster.
- B. Configure a job that executes every time new data lands in a given directory.
- C. Schedule a Structured Streaming job with a trigger interval of 60 minutes.
- D. Schedule a job to execute the pipeline once an hour on a dedicated interactive cluster.

Antwort: A

Begründung:

Scheduling a job to execute the data processing pipeline once an hour on a new job cluster is the most cost-effective solution given the scenario. Job clusters are ephemeral in nature; they are spun up just before the job execution and terminated upon completion, which means you only incur costs for the time the cluster is active. Since the total processing time is only 10 minutes, a new job cluster created for each hourly execution minimizes the running time and thus the cost, while also fulfilling the requirement for hourly data updates for the business reporting team's dashboards.

:

Databricks documentation on jobs and job clusters: <https://docs.databricks.com/jobs.html>

100. Frage

A team member is leaving the team and he/she is currently the owner of the few tables, instead of transferring the ownership to a user you have decided to transfer the ownership to a group so in the future anyone in the group can manage the permissions rather than a single individual, which of the following commands help you accomplish this?

- A. GRANT OWNER table_name to 'group'*
- B. TRANSFER OWNER table_name to 'group'
- C. ALTER OWNER ON table_name to 'group'
- D. GRANT OWNER On table_name to 'group'
- E. ALTER TABLE table_name OWNER to 'group'

Antwort: E

Begründung:

Explanation

The answer is ALTER TABLE table_name OWNER to 'group'

Assign owner to object

101. Frage

.....

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

- Databricks-Certified-Professional-Data-Engineer Lernressourcen ☐ Databricks-Certified-Professional-Data-Engineer Prüfungsfragen ☐ Databricks-Certified-Professional-Data-Engineer Prüfung ☐ Suchen Sie einfach auf ➡ www.zertpruefung.ch ☐ nach kostenloser Download von « Databricks-Certified-Professional-Data-Engineer » ☐ ☐ Databricks-Certified-Professional-Data-Engineer Fragenkatalog
- Neuester und gültiger Databricks-Certified-Professional-Data-Engineer Test VCE Motoren-Dumps und Databricks-Certified-Professional-Data-Engineer neueste Testfragen für die IT-Prüfungen ☐ Öffnen Sie (www.itzert.com) geben Sie ➡ Databricks-Certified-Professional-Data-Engineer ☐ ein und erhalten Sie den kostenlosen Download ☐ ☐ Databricks-Certified-Professional-Data-Engineer Exam
- Databricks-Certified-Professional-Data-Engineer Lernressourcen ☐ Databricks-Certified-Professional-Data-Engineer Online Prüfungen ☐ Databricks-Certified-Professional-Data-Engineer Prüfungs-Guide ☐ URL kopieren ☼ de.fast2test.com ☐ ☼ ☐ Öffnen und suchen Sie ➤ Databricks-Certified-Professional-Data-Engineer ☐ Kostenloser Download ☼ Databricks-Certified-Professional-Data-Engineer PDF
- Databricks-Certified-Professional-Data-Engineer Schulungsmaterialien - Databricks-Certified-Professional-Data-Engineer Dumps Prüfung - Databricks-Certified-Professional-Data-Engineer Studienguide ☐ Suchen Sie auf « www.itzert.com » nach “Databricks-Certified-Professional-Data-Engineer” und erhalten Sie den kostenlosen Download mühelos ☐ ☐ Databricks-Certified-Professional-Data-Engineer Zertifizierungsfragen
- Databricks-Certified-Professional-Data-Engineer Schulungsmaterialien - Databricks-Certified-Professional-Data-Engineer Dumps Prüfung - Databricks-Certified-Professional-Data-Engineer Studienguide ☐ Öffnen Sie die Webseite « www.deutschpruefung.com » und suchen Sie nach kostenloser Download von ➤ Databricks-Certified-Professional-Data-Engineer ☐ ☐ Databricks-Certified-Professional-Data-Engineer Demotesten
- Databricks-Certified-Professional-Data-Engineer Buch ☐ Databricks-Certified-Professional-Data-Engineer Lernressourcen ☐ Databricks-Certified-Professional-Data-Engineer Tests ☐ Suchen Sie jetzt auf ➤ www.itzert.com ☐ nach ☼ Databricks-Certified-Professional-Data-Engineer ☐ ☼ ☐ und laden Sie es kostenlos herunter ☐ Databricks-Certified-Professional-Data-Engineer Prüfungsvorbereitung

