Databricks-Certified-Professional-Data-Engineer퍼펙트 공부자료 & Databricks-Certified-Professional-Data-Engineer최신버전시험덤프문제

그리고 Itcertkr Databricks-Certified-Professional-Data-Engineer 시험 문제집의 전체 버전을 클라우드 저장소에서 다운로드할 수 있습니다: https://drive.google.com/open?id=1Q7Y2pJkvVRmmULbwjTqBF328DqEpLlVR

Itcertkr의Databricks인증 Databricks-Certified-Professional-Data-Engineer덤프는 고객님의 IT인증자격증을 취득하는 소원을들어줍니다. IT업계에 금방 종사한 분은 자격증을 많이 취득하여 자신만의 가치를 업그레이드할수 있습니다. Itcertkr의Databricks인증 Databricks-Certified-Professional-Data-Engineer덤프는 실제 시험문제에 대비하여 연구제작된 퍼펙트한 시험전 공부자료로서 시험이 더는 어렵지 않게 느끼도록 편하게 도와드립니다.

Databricks 인증 전문 데이터 엔지니어 자격증은 기술 산업의 주요 회사들에 의해 데이터 엔지니어링 분야에서 우수성의 기준으로 인정됩니다. 이는 이 분야에서 경력을 발전시키고 Databricks 기술에 대한 전문 지식을 입증하려는 개인들에게 유용한 자격증입니다. 게다가, Databricks 인증 전문 데이터 엔지니어를 고용하는 기업들은 복잡한 데이터 파이프라인 및 분석 시스템을 구축 및 유지 관리할 필요한 기술을 보유하고 있음을 확신할 수 있습니다.

Databricks Certified Professional Data Engineer 시험은 데이터 엔지니어링에 대한 광범위한 지식과 경험이 필요한 엄격한 인증 시험입니다. 응시자는 데이터 모델링, 데이터웨어 하우징, ETL, 데이터 거버넌스 및 데이터 보안과 같은 데이터 엔지니어링 개념을 깊이 이해해야합니다. 또한 Apache Spark, Delta Lake 및 MLFlow와 같은 Databricks 도구 및 기술을 사용한 경험이 있어야합니다. 이 시험을 통과하면 응시자는 데이터 사업 플랫폼에서 데이터 파이프 라인을 구축하고 최적화하는 데 필요한 기술과 지식이 있음을 보여줍니다.

>> Databricks-Certified-Professional-Data-Engineer퍼펙트 공부자료 <<

Databricks-Certified-Professional-Data-Engineer최신버전 시험덤프문제, Databricks-Certified-Professional-Data-Engineer퍼펙트 덤프 최신문제

Databricks Databricks-Certified-Professional-Data-Engineer인증시험이 이토록 인기가 많으니 우리Itcertkr에서는 모든 힘을 다하여 여러분이 응시에 도움을 드리겠으며 또 일년무료 업뎃서비스를 제공하며, Itcertkr 선택으로 여러분은 자신의 꿈과 더 가까워질 수 있습니다. 희망찬 내일을 위하여 Itcertkr선택은 정답입니다. Itcertkr선택함으로 당신이 바로 진정한IT인사입니다.

Databricks Certified Professional Data Engineer 시험은 Databricks를 사용하여 대규모 데이터를 처리하는 데 전문성을 증명하려는 데이터 엔지니어들에게 탁월한 선택입니다. 이 자격증 시험은 전 세계적으로 인정되며 Databricks를 사용하는 기관에서 높은 가치를 두고 있습니다. 시험에 통과함으로써, 데이터 엔지니어는 자신의 지식과 기술을 검증하고 경력 발전의 기회를 높일 수 있습니다.

최신 Databricks Certification Databricks-Certified-Professional-Data-Engineer 무료샘플문제 (Q90-Q95):

질문 # 90

A data pipeline uses Structured Streaming to ingest data from kafka to Delta Lake. Data is being stored in a bronze table, and includes the Kafka_generated timesamp, key, and value. Three months after the pipeline is deployed the data engineering team has noticed some latency issued during certain times of the day.
A senior data engineer updates the Delta Table's schema and ingestion logic to include the current timestamp (as recoded by Apache Spark) as well the Kafka topic and partition. The team plans to use the additional metadata fields to diagnose the transient processing delays:
Which limitation will the team face while diagnosing this problem?

- A. Spark cannot capture the topic partition fields from the kafka source.
- B. New fields not be computed for historic records.
- C. Updating the table schema will invalidate the Delta transaction log metadata.
- D. Updating the table schema requires a default value provided for each file added.

정답： B

설명:
When adding new fields to a Delta table's schema, these fields will not be retrospectively applied to historical records that were ingested before the schema change. Consequently, while the team can use the new metadata fields to investigate transient processing delays moving forward, they will be unable to apply this diagnostic approach to past data that lacks these fields.
Reference:
Databricks documentation on Delta Lake schema management: https://docs.databricks.com/delta/delta-batch.html#schema-management

질문 # 91
The marketing team is looking to share data in an aggregate table with the sales organization, but the field names used by the teams do not match, and a number of marketing specific fields have not been approval for the sales org.
Which of the following solutions addresses the situation while emphasizing simplicity?

- A. Create a new table with the required schema and use Delta Lake's DEEP CLONE functionality to sync up changes committed to one table to the corresponding table.
- B. Use a CTAS statement to create a derivative table from the marketing table configure a production jon to propagation changes.
- C. Create a view on the marketing table selecting only these fields approved for the sales team alias the names of any fields that should be standardized to the sales naming conventions.
- D. Add a parallel table write to the current production pipeline, updating a new sales table that varies as required from marketing table.

정답： C

설명:
Creating a view is a straightforward solution that can address the need for field name standardization and selective field sharing between departments. A view allows for presenting a transformed version of the underlying data without duplicating it. In this scenario, the view would only include the approved fields for the sales team and rename any fields as per their naming conventions.
:
Databricks documentation on using SQL views in Delta Lake: https://docs.databricks.com/delta/quick-start.
html#sql-views

질문 # 92
How VACCUM and OPTIMIZE commands can be used to manage the DELTA lake?

- A. VACCUM command can be used to compact small parquet files, and the OP-TIMZE command can be used to delete parquet files that are marked for dele-tion/unused.
- B. VACCUM command can be used to compress the parquet files to reduce the size of the table, OPTIMIZE command can be used to cache frequently delta tables for better performance.
- C. VACCUM command can be used to delete empty/blank parquet files in a delta table, OPTIMIZE command can be used to cache frequently delta tables for better perfor-mance.
- D. OPTIMIZE command can be used to compact small parquet files, and the VAC-CUM command can be used to delete parquet files that are marked for deletion/unused.
(Correct)

- E. VACCUM command can be used to delete empty/blank parquet files in a delta table. OPTIMIZE command can be used to update stale statistics on a delta table.

정답： D

설명：

Explanation

VACCUM:

You can remove files no longer referenced by a Delta table and are older than the retention thresh-old by running the vacuum command on the table. vacuum is not triggered automatically. The de-fault retention threshold for the files is 7 days. To change this behavior, see Configure data reten-tion for time travel.

OPTIMIZE:

Using OPTIMIZE you can compact data files on Delta Lake, this can improve the speed of read queries on the table. Too many small files can significantly degrade the performance of the query.

질문 # 93

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_saies_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 Structured Streaming read against the daily_store_sales table and use upsert logic to update results in the store_sales_summary table.
- B. 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.
- C. 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.
- D. 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.
- 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.

정답： C

설명：

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 Reference: [Databricks Certified Data Engineer Professional], under "Spark Core" section; Databricks Documentation, under "Structured

Streaming" section; Databricks Documentation, under "Delta Change Data Feed" section.

질문 # 94

A data engineer is configuring a pipeline that will potentially see late-arriving, duplicate records.
In addition to de-duplicating records within the batch, which of the following approaches allows the data engineer to deduplicate data against previously processed records as it is inserted into a Delta table?

- A. VACUUM the Delta table after each batch completes.
- B. Perform an insert-only merge with a matching condition on a unique key.
- C. Perform a full outer join on a unique key and overwrite existing data.
- D. Rely on Delta Lake schema enforcement to prevent duplicate records.
- E. Set the configuration delta.deduplicate = true.

정답： B

설명：

To deduplicate data against previously processed records as it is inserted into a Delta table, you can use the merge operation with an insert-only clause. This allows you to insert new records that do not match any existing records based on a unique key, while ignoring duplicate records that match existing records. For example, you can use the following syntax:
MERGE INTO target_table USING source_table ON target_table.unique_key = source_table.unique_key WHEN NOT MATCHED THEN INSERT * This will insert only the records from the source table that have a unique key that is not present in the target table, and skip the records that have a matching key. This way, you can avoid inserting duplicate records into the Delta table.
:
https://docs.databricks.com/delta/delta-update.html#upsert-into-a-table-using-merge
https://docs.databricks.com/delta/delta-update.html#insert-only-merge

질문 # 95

......

Databricks-Certified-Professional-Data-Engineer최신버전 시험덤프문제: https://www.itcertkr.com/Databricks-Certified-Professional-Data-Engineer_exam.html

- Databricks-Certified-Professional-Data-Engineer최신버전 시험공부자료 □ Databricks-Certified-Professional-Data-Engineer퍼펙트 덤프 최신 데모 □ Databricks-Certified-Professional-Data-Engineer시험대비 덤프문제 □ 지금「kr.fast2test.com」에서□ Databricks-Certified-Professional-Data-Engineer □를 검색하고 무료로 다운로드하세요 Databricks-Certified-Professional-Data-Engineer Vce
- Databricks-Certified-Professional-Data-Engineer높은 통과율 공부문제 □ Databricks-Certified-Professional-Data-Engineer최신버전 시험공부자료 □ Databricks-Certified-Professional-Data-Engineer높은 통과율 덤프공부 □ □ www.itdumpskr.com □에서 검색만 하면➡ Databricks-Certified-Professional-Data-Engineer □를 무료로 다운로드할 수 있습니다Databricks-Certified-Professional-Data-Engineer최고합격덤프
- 100% 합격보장 가능한 Databricks-Certified-Professional-Data-Engineer퍼펙트 공부자료 공부문제 ☣ 오픈 웹 사이트□ kr.fast2test.com □검색▷ Databricks-Certified-Professional-Data-Engineer ◁무료 다운로드Databricks-Certified-Professional-Data-Engineer Vce
- Databricks-Certified-Professional-Data-Engineer퍼펙트 공부자료 최신 덤프문제모음집 □ □ www.itdumpskr.com □에서▶ Databricks-Certified-Professional-Data-Engineer ◀를 검색하고 무료 다운로드 받기Databricks-Certified-Professional-Data-Engineer시험응시
- Databricks-Certified-Professional-Data-Engineer퍼펙트 공부자료 인기 인증 시험덤프자료 □ ☀ Databricks-Certified-Professional-Data-Engineer □☀□를 무료로 다운로드하려면⇛ www.dumptop.com ⇚웹사이트를 입력하세요Databricks-Certified-Professional-Data-Engineer최신버전 시험공부자료
- Databricks-Certified-Professional-Data-Engineer퍼펙트 공부자료 최신버전 시험대비자료 □ 무료로 쉽게 다운로드하려면□ www.itdumpskr.com □에서☀ Databricks-Certified-Professional-Data-Engineer □☀□를 검색하세요 Databricks-Certified-Professional-Data-Engineer Vce
- Databricks-Certified-Professional-Data-Engineer시험응시 □ Databricks-Certified-Professional-Data-Engineer시험대비 덤프문제 □ Databricks-Certified-Professional-Data-Engineer완벽한 시험기출자료 □ ▷ www.koreadumps.com ◁에서▶ Databricks-Certified-Professional-Data-Engineer ◀를 검색하고 무료로 다운로드하세요Databricks-Certified-Professional-Data-Engineer인증시험
- Databricks-Certified-Professional-Data-Engineer높은 통과율 덤프공부 ↗ Databricks-Certified-Professional-Data-Engineer최신 업데이트 인증공부자료 □ Databricks-Certified-Professional-Data-Engineer최신 기출문제 □《

www.itdumpskr.com 》에서□ Databricks-Certified-Professional-Data-Engineer □를 검색하고 무료로 다운로드하세요Databricks-Certified-Professional-Data-Engineer최신버전 시험공부자료

- Databricks-Certified-Professional-Data-Engineer시험응시 □ Databricks-Certified-Professional-Data-Engineer Vce □ Databricks-Certified-Professional-Data-Engineer최고품질 인증시험공부자료 □ ➤ www.pass4test.net □의 무료 다운로드➤ Databricks-Certified-Professional-Data-Engineer □페이지가 지금 열립니다Databricks-Certified-Professional-Data-Engineer최신 업데이트 인증공부자료
- Databricks-Certified-Professional-Data-Engineer시험패스보장덤프 □ Databricks-Certified-Professional-Data-Engineer시험패스보장덤프 □ Databricks-Certified-Professional-Data-Engineer최신 기출문제 □ 【 www.itdumpskr.com 】을 통해 쉽게□ Databricks-Certified-Professional-Data-Engineer □무료 다운로드 받기Databricks-Certified-Professional-Data-Engineer시험대비 덤프공부
- Databricks-Certified-Professional-Data-Engineer최신 기출문제 □ Databricks-Certified-Professional-Data-Engineer완벽한 시험기출자료 □ Databricks-Certified-Professional-Data-Engineer최고합격덤프 □ ▷ Databricks-Certified-Professional-Data-Engineer ◁를 무료로 다운로드하려면□ www.dumptop.com □웹사이트를 입력하세요Databricks-Certified-Professional-Data-Engineer최신 업데이트 시험덤프문제
- myportal.utt.edu.tt, www.stes.tyc.edu.tw, 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, yogesganesan.com, aavitechveda.com, 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, www.stes.tyc.edu.tw, learn.psmsurat.com, Disposable vapes

그 외, Itcertkr Databricks-Certified-Professional-Data-Engineer 시험 문제집 일부가 지금은 무료입니다: https://drive.google.com/open?id=1Q7Y2pJkvVRmmULbwjTqBF328DqEpLIVR