試験の準備方法-100％合格率のDatabricks-Certified-Professional-Data-Engineer的中問題集試験-高品質なDatabricks-Certified-Professional-Data-Engineer最新関連参考書

ちなみに、Pass4Test Databricks-Certified-Professional-Data-Engineerの一部をクラウドストレージからダウンロードできます：https://drive.google.com/open?id=1ioqWY1q8h6-jycCX2uT-M74B82AKZU_N

Pass4TestはDatabricksのDatabricks-Certified-Professional-Data-Engineer認定試験に受かりたい各受験生に明確かつ顕著なソリューションを提供しました。当社はDatabricksのDatabricks-Certified-Professional-Data-Engineer認定試験の詳しい問題と解答を提供します。当社のIT専門家が最も経験と資格があるプロな人々で、我々が提供したテストの問題と解答は実際の認定試験と殆ど同じです。これは本当に素晴らしいことです。それにもっと大切なのは、Pass4Testのサイトは世界的でDatabricks-Certified-Professional-Data-Engineer試験トレーニングによっての試験合格率が一番高いです。

試験は、Databricksを使用して大規模なデータ処理パイプラインを設計、構築、および維持することに熟練したデータエンジニアが自分の専門知識を証明したい場合に最適です。認定試験はデータエンジニアの特定の役割と責任に合わせて調整されており、データインジェスション、データ変換、データストレージ、およびデータ分析などのトピックをカバーしています。試験に合格することで、候補者はDatabricksを使用して複雑なビッグデータ課題を解決する能力を証明できます。

Databricks Certified Professional Data Engineer（Databricks-Certified-Professional-Data-Engineer）試験は、Databricksを使用してビッグデータパイプラインを開発および管理するデータエンジニアのスキルと専門知識を検証するために設計された認定プログラムです。この試験は、Databricksで作業するデータエンジニア、ETL開発者、およびデータアーキテクトに最適であり、彼らのスキルと熟練度をアピールすることができます。

>> Databricks-Certified-Professional-Data-Engineer的中問題集 <<

Pass4Testは Databricks Databricks-Certified-Professional-Data-Engineer試験の実践訓練を提供する

最近Databricks試験に参加する人が多くなっています。どのように試験を準備すべきですか？受験生たちはまず試験センターでDatabricks-Certified-Professional-Data-Engineer認証試験に関する情報を了解してください。順調にDatabricks-Certified-Professional-Data-Engineer試験に合格するために、我々の問題集で復習することができます。我々の問題集は的中率が高いですから、あなたのDatabricks-Certified-Professional-Data-Engineer試験への復習に役立つことができます。

Databricks Certified Professional Data Engineer Exam 認定 Databricks-Certified-Professional-Data-Engineer 試験問題 (Q143-Q148):

質問 # 143
A data team's Structured Streaming job is configured to calculate running aggregates for item sales to update a downstream marketing dashboard. The marketing team has introduced a new field to track the number of times this promotion code is used for each item. A junior data engineer suggests updating the existing query as follows: Note that proposed changes are in bold.

Which step must also be completed to put the proposed query into production?

- A. Run REFRESH TABLE delta, /item_agg'
- B. Remove .option (mergeSchema', true') from the streaming write
- C. Increase the shuffle partitions to account for additional aggregates
- D. Specify a new checkpointlocation

正解：D

解説：
When introducing a new aggregation or a change in the logic of a Structured Streaming query, it is generally necessary to specify a new checkpoint location. This is because the checkpoint directory contains metadata about the offsets and the state of the aggregations of a streaming query. If the logic of the query changes, such as including a new aggregation field, the state information saved in the current checkpoint would not be compatible with the new logic, potentially leading to incorrect results or failures. Therefore, to accommodate the new field and ensure the streaming job has the correct starting point and state information for aggregations, a new checkpoint location should be specified.
References:
* Databricks documentation on Structured Streaming:
https://docs.databricks.com/spark/latest/structured-streaming/index.html
* Databricks documentation on streaming checkpoints:
https://docs.databricks.com/spark/latest/structured-streaming/production.html#checkpointing

質問 # 144
The data engineering team has configured a Databricks SQL query and alert to monitor the values in a Delta Lake table. The recent_sensor_recordings table contains an identifying sensor_id alongside the timestamp and temperature for the most recent 5 minutes of recordings.
The below query is used to create the alert:

```
SELECT MEAN(temperature), MAX(temperature), MIN(temperature)
FROM recent_sensor_recordings
GROUP BY sensor_id
```

The query is set to refresh each minute and always completes in less than 10 seconds. The alert is set to trigger when mean (temperature) > 120. Notifications are triggered to be sent at most every 1 minute.
If this alert raises notifications for 3 consecutive minutes and then stops, which statement must be true?

- A. The average temperature recordings for at least one sensor exceeded 120 on three consecutive executions of the query
- B. The total average temperature across all sensors exceeded 120 on three consecutive executions of the query
- C. The recent_sensor_recordingstable was unresponsive for three consecutive runs of the query
- D. The maximum temperature recording for at least one sensor exceeded 120 on three consecutive executions of the query
- E. The source query failed to update properly for three consecutive minutes and then restarted

正解：A

解説：

This is the correct answer because the query is using a GROUP BY clause on the sensor_id column, which means it will calculate the mean temperature for each sensor separately. The alert will trigger when the mean temperature for any sensor is greater than 120, which means at least one sensor had an average temperature above 120 for three consecutive minutes. The alert will stop when the mean temperature for all sensors drops below 120. Verified Reference: [Databricks Certified Data Engineer Professional], under "SQL Analytics" section; Databricks Documentation, under "Alerts" section.

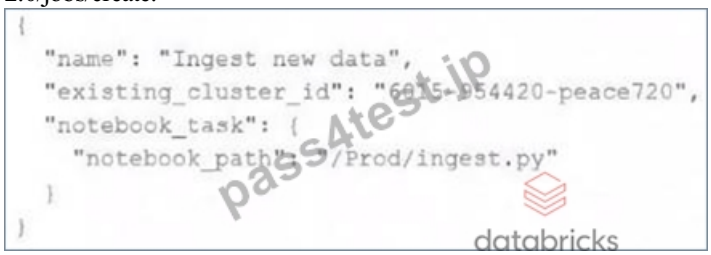
質問 # 145
A Delta Lake table representing metadata about content from user has the following schema:
user_id LONG, post_text STRING, post_id STRING, longitude FLOAT, latitude FLOAT, post_time TIMESTAMP, date DATE
Based on the above schema, which column is a good candidate for partitioning the Delta Table?

- A. Post_id
- B. Post_time
- C. User_id
- D. Date

正解：D

解説：
Partitioning a Delta Lake table improves query performance by organizing data into partitions based on the values of a column. In the given schema, the date column is a good candidate for partitioning for several reasons:
* Time-Based Queries: If queries frequently filter or group by date, partitioning by the date column can significantly improve performance by limiting the amount of data scanned.
* Granularity: The date column likely has a granularity that leads to a reasonable number of partitions (not too many and not too few). This balance is important for optimizing both read and write performance.
* Data Skew: Other columns like post_id or user_id might lead to uneven partition sizes (data skew), which can negatively impact performance.
Partitioning by post_time could also be considered, but typically date is preferred due to its more manageable granularity.
:
Delta Lake Documentation on Table Partitioning: Optimizing Layout with Partitioning

質問 # 146
A junior data engineer has configured a workload that posts the following JSON to the Databricks REST API endpoint 2.0/jobs/create.



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

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

正解：E

解説：
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 Reference: [Databricks Certified Data Engineer Professional], under

"Monitoring & Logging" section; [Databricks Documentation], under "Jobs API - Create" section.

質問 # 147
A data engineer is configuring Delta Sharing for a Databricks-to-Databricks scenario to optimize read performance. The recipient needs to perform time travel queries and streaming reads on shared sales data.
Which configuration will provide the optimal performance while enabling these capabilities?

- A. Share the entire schema WITHOUT HISTORY and rely on recipient-side caching for performance.
- B. Use the open sharing protocol instead of Databricks-to-Databricks sharing for better performance.
- C. Share tables WITHOUT HISTORY and enable partitioning for better query performance.
- D. Share tables WITH HISTORY, ensure tables don't have partitioning enabled, and enable CDF before sharing.

正解： D

解説：
Comprehensive and Detailed Explanation From Exact Extract of Databricks Data Engineer Documents:
The official Delta Sharing guidance specifies that in order for recipients to use time travel queries and streaming reads, providers must share Delta tables WITH HISTORY. Sharing history ensures the Delta log is included, which enables efficient access to table snapshots and incremental data streams. Additionally, Change Data Feed (CDF) must be enabled prior to sharing if downstream consumers require streaming CDC queries. Without history, recipients cannot perform time travel or streaming queries. Open sharing supports static Delta tables but lacks streaming support. Therefore, sharing tables WITH HISTORY and enabling CDF is the required configuration for both performance and functionality.

質問 # 148
.....

Databricks-Certified-Professional-Data-Engineer試験ガイドは、ビジネスマンであろうと学生であろうと、すべての人に適しています。試験に参加するには、20〜30時間で練習できます。あなたが素晴らしい成績をとれることは間違いありません。私たちの学習ペースに従えば、予想外の驚きがあります。当社のDatabricks-Certified-Professional-Data-Engineerガイドトレントを選択した場合にのみ、この重要な試験に合格し、Databricks-Certified-Professional-Data-Engineer試験の準備に関するまったく新しい経験を得ることが容易になります。

ンロードしてくださいDatabricks-Certified-Professional-Data-Engineer全真模擬試験

- Databricks-Certified-Professional-Data-Engineer全真模擬試験 ✍ Databricks-Certified-Professional-Data-Engineer日本語 □ Databricks-Certified-Professional-Data-Engineer模試エンジン □ ➡ www.xhs1991.com □□□は、➽ Databricks-Certified-Professional-Data-Engineer □を無料でダウンロードするのに最適なサイトですDatabricks-Certified-Professional-Data-Engineer模擬試験最新版
- 権威のあるDatabricks-Certified-Professional-Data-Engineer的中問題集 - 合格スムーズDatabricks-Certified-Professional-Data-Engineer最新関連参考書 | 信頼的なDatabricks-Certified-Professional-Data-Engineer再テスト □ ⇛ Databricks-Certified-Professional-Data-Engineer ⇚を無料でダウンロード{ www.goshiken.com }で検索するだけDatabricks-Certified-Professional-Data-Engineer模試エンジン
- Databricks-Certified-Professional-Data-Engineer日本語対策問題集 □ Databricks-Certified-Professional-Data-Engineer日本語版問題集 ◴ Databricks-Certified-Professional-Data-Engineer試験合格攻略 □ 今すぐ▹ www.xhs1991.com ◃で➽ Databricks-Certified-Professional-Data-Engineer □を検索して、無料でダウンロードしてくださいDatabricks-Certified-Professional-Data-Engineer日本語
- Databricks Databricks-Certified-Professional-Data-Engineer試験の認証資格は一層重要になった □ 時間限定無料で使える“Databricks-Certified-Professional-Data-Engineer”の試験問題は「 www.goshiken.com 」サイトで検索Databricks-Certified-Professional-Data-Engineerテスト問題集
- Databricks-Certified-Professional-Data-Engineer試験復習 □ Databricks-Certified-Professional-Data-Engineer日本語 □ □ Databricks-Certified-Professional-Data-Engineer模擬試験最新版 □ ✔ www.goshiken.com □✔□は、➤ Databricks-Certified-Professional-Data-Engineer □を無料でダウンロードするのに最適なサイトですDatabricks-Certified-Professional-Data-Engineer関連試験
- tomascuirolo.com, www.stes.tyc.edu.tw, www.stes.tyc.edu.tw, www.stes.tyc.edu.tw, www.stes.tyc.edu.tw, cou.alnoor.edu.iq, www.stes.tyc.edu.tw, www.stes.tyc.edu.tw, mzansiempowerment.com, summerschool.entrehubs.com, Disposable vapes

BONUS！！！ Pass4Test Databricks-Certified-Professional-Data-Engineerダンプの一部を無料でダウンロード：https://drive.google.com/open?id=1ioqWY1q8h6-jycCX2uT-M74B82AKZU_N