

# Databricks-Certified-Professional-Data-Engineer 測試, Databricks-Certified-Professional-Data-Engineer 證照指南



此外, 這些NewDumps Databricks-Certified-Professional-Data-Engineer考試題庫的部分內容現在是免費的: <https://drive.google.com/open?id=15kTO-sV-WsSIUBI6-c6nXlqW3x7goyYB>

通過 Databricks的Databricks-Certified-Professional-Data-Engineer的考試認證不僅僅是驗證你的技能, 但也證明你的專業知識和你的證書, 你的老闆沒有白白僱傭你, 目前的IT行業需要一個可靠的 Databricks的Databricks-Certified-Professional-Data-Engineer的考試的來源, NewDumps是個很好的選擇, Databricks-Certified-Professional-Data-Engineer的考試縮短在最短的時間內, 這樣不會浪費你的錢和精力。還會讓你又一個美好的前程。

在我們網站, 您可以先免費嘗試下載我們的題庫DEMO, 體驗我們的Databricks Databricks-Certified-Professional-Data-Engineer考古題的品質, 相信在您使用之後會很滿意我們的產品。成千上萬的IT考生通過我們的產品成功通過考試, 該Databricks-Certified-Professional-Data-Engineer考古題的品質已被廣大考生檢驗。我們的Databricks Databricks-Certified-Professional-Data-Engineer題庫根據實際考試的動態變化而更新, 以確保Databricks-Certified-Professional-Data-Engineer考古題覆蓋率始終最高於99%。保證大家通過Databricks-Certified-Professional-Data-Engineer認證考試, 如果您失敗, 可以享受 100%的退款保證。

>> Databricks-Certified-Professional-Data-Engineer測試 <<

## Databricks Databricks-Certified-Professional-Data-Engineer 證照指南, Databricks-Certified-Professional-Data-Engineer 考古題

你可以在NewDumps的網站上下載部分NewDumps的最新的關於Databricks Databricks-Certified-Professional-Data-Engineer 認證考試練習題及答案作為免費嘗試了, 相信不會讓你失望的。NewDumps的最新的關於Databricks Databricks-Certified-Professional-Data-Engineer 認證考試練習題及答案和真實考試題目是很接近。或許你在其他的網站上也看到了相關的培訓資料, 但是你仔細比較後就會發現他們的資料來源與NewDumps。NewDumps提供的資料比較全面, 包括當前考試題目, 是由NewDumps的專家團隊利用他們的豐富的經驗和知識針對Databricks Databricks-Certified-Professional-Data-Engineer 認證考試研究出來的。

Databricks-Certified-Professional-Data-Engineer 證書的目的是在大數據行業中建立一個數據工程技能的標準。這個證書證明了專業人員擁有在雲中有效地從事複雜大數據項目所需的知識和技能。它也提高了候選人在競爭激烈的行業中找工作、保住工作或晉升的機會。

## 最新的 Databricks Certification Databricks-Certified-Professional-Data-Engineer 免費考試真題 (Q122-Q127):

### 問題 #122

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

答案： B

解題說明：

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.

Reference:

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

### 問題 #123

A data ingestion task requires a one-TB JSON dataset to be written out to Parquet with a target part-file size of 512 MB. Because Parquet is being used instead of Delta Lake, built-in file-sizing features such as Auto-Optimize & Auto-Compaction cannot be used.

Which strategy will yield the best performance without shuffling data?

- A. Ingest the data, execute the narrow transformations, repartition to 2,048 partitions (1TB\*1024\*1024/512), and then write to parquet.
- B. Set `spark.sql.shuffle.partitions` to 512, ingest the data, execute the narrow transformations, and then write to parquet.
- C. Set `spark.sql.files.maxPartitionBytes` to 512 MB, ingest the data, execute the narrow transformations, and then write to parquet.
- **D. Set `spark.sql.shuffle.partitions` to 2,048 partitions (1TB\*1024\*1024/512), ingest the data, execute the narrow transformations, optimize the data by sorting it (which automatically repartitions the data), and then write to parquet.**
- E. Set `spark.sql.adaptive.advisoryPartitionSizeInBytes` to 512 MB bytes, ingest the data, execute the narrow transformations, coalesce to 2,048 partitions (1TB\*1024\*1024/512), and then write to parquet.

答案： D

解題說明：

The key to efficiently converting a large JSON dataset to Parquet files of a specific size without shuffling data lies in controlling the size of the output files directly.

\* Setting `spark.sql.files.maxPartitionBytes` to 512 MB configures Spark to process data in chunks of 512 MB. This setting directly influences the size of the part-files in the output, aligning with the target file size.

\* Narrow transformations (which do not involve shuffling data across partitions) can then be applied to this data.

\* Writing the data out to Parquet will result in files that are approximately the size specified by `spark.sql.files.maxPartitionBytes`, in this case, 512 MB.

\* The other options involve unnecessary shuffles or repartitions (B, C, D) or an incorrect setting for this specific requirement (E).

References:

\* Apache Spark Documentation: Configuration - `spark.sql.files.maxPartitionBytes`

\* Databricks Documentation on Data Sources: Databricks Data Sources Guide

### 問題 #124

A junior data engineer on your team has implemented the following code block.

```

MERGE INTO events
USING new_events
ON events.event_id=new_events.event_id
WHEN NOT MATCHED
INSERT

```

The view `new_events` contains a batch of records with the same schema as the `events` Delta table.

The `event_id` field serves as a unique key for this table.

When this query is executed, what will happen with new records that have the same `event_id` as an existing record?

- A. They are inserted.
- **B. They are ignored.**
- C. They are deleted.
- D. They are merged.
- E. They are updated.

答案: B

解題說明:

Explanation

This is the correct answer because it describes what will happen with new records that have the same `event_id` as an existing record when the query is executed. The query uses the `INSERT INTO` command to append new records from the view `new_events` to the table `events`. However, the `INSERT INTO` command does not check for duplicate values in the primary key column (`event_id`) and does not perform any update or delete operations on existing records. Therefore, if there are new records that have the same `event_id` as an existing record, they will be ignored and not inserted into the table `events`. Verified References: [Databricks Certified Data Engineer Professional], under "Delta Lake" section; Databricks Documentation, under "Append data using `INSERT INTO`" section.

"If none of the `WHEN MATCHED` conditions evaluate to true for a source and target row pair that matches the `merge_condition`, then the target row is left unchanged." <https://docs.databricks.com/en/sql/language-manual/delta-merge-into.html#:~:text=If%20none%20>

### 問題 #125

A data engineer is using Lakeflow Declarative Pipeline to propagate row deletions from a source bronze table (`user_bronze`) to a target silver table (`user_silver`). The engineer wants deletions in `user_bronze` to automatically delete corresponding rows in `user_silver` during pipeline execution.

Which configuration ensures deletions in the bronze table are propagated to the silver table?

- **A. Enable Change Data Feed (CDF) on `user_bronze`, read its CDF stream, and use `apply_changes()` with `apply_as_deletes=True` for `user_silver`.**
- B. Configure `VACUUM` on `user_bronze` to delete files, then rebuild `user_silver` from scratch.
- C. Use `apply_changes` without CDF and filter rows where `_soft_deleted` is true.
- D. Enable CDF on `user_silver`, read its transaction log, and use `MERGE` to sync deletions.

答案: A

解題說明:

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

According to Databricks documentation, Change Data Feed (CDF) allows pipelines to read incremental data changes, including inserts, updates, and deletes, from a Delta table. When deletions occur in the source table, reading the CDF stream ensures downstream consumers receive the deletion records. The Lakeflow Declarative Pipelines API provides the `apply_changes()` function (or auto-CDC pipelines) with the `apply_as_deletes` parameter to correctly apply those deletions to the target table. This enables automatic synchronization between bronze and silver layers. Options A and D either require manual handling or complete rebuilds, and C incorrectly applies CDF to the target rather than the source. Therefore, enabling CDF on the bronze table and using `apply_as_deletes=True` is the correct, Databricks-supported configuration.

## 問題 #126

Review the following error traceback:

```
AnalysisException                                Traceback (most recent call last)
<command-3293767849433948> in <module>
----> 1 display(df.select(3*"heartrate"))

/databricks/spark/python/pyspark/sql/dataframe.py in select(self, *cols)
   1690         [Row(name='Alice', age=12), Row(name='Bob', age=15)]
   1691         """
-> 1692         jdf = self._jdf.select(self._jcols(*cols))
   1693         return DataFrame(jdf, self.sql_ctx)
   1694

/databricks/spark/python/lib/py4j-0.10.9-src.zip/py4j/java_gateway.py in __call__(self, *args)
   1302
   1303         answer = self.gateway_client.send_command(command)
-> 1304         return_value = get_return_value(
   1305             answer, self.gateway_client, self.target_id, self.name)
   1306

/databricks/spark/python/pyspark/sql/utils.py in deco(*a, **kw)
   121         # Hide where the exception came from that shows a non-Pythonic
   122         # JVM exception message.
--> 123         raise converted from None
   124     else:
   125         raise

AnalysisException: cannot resolve 'heartrateheartrateheartrate' given input columns:
[spark_catalog.database.table.device_id, spark_catalog.database.table.heartrate,
spark_catalog.database.table.mrn, spark_catalog.database.table.time];
'Project ['heartrateheartrateheartrate]
+- SubqueryAlias spark_catalog.database.table
  +- Relation[device_id#75L,heartrate#76,mrn#77L,time#78] parquet
```

Which statement describes the error being raised?

- A. There is no column in the table named heartrateheartrateheartrate
- B. There is a type error because a DataFrame object cannot be multiplied.
- C. There is a syntax error because the heartrate column is not correctly identified as a column.
- D. The code executed was PySpark but was executed in a Scala notebook.
- E. There is a type error because a column object cannot be multiplied.

答案: A

解題說明:

The error being raised is an AnalysisException, which is a type of exception that occurs when Spark SQL cannot analyze or execute a query due to some logical or semantic error. In this case, the error message indicates that the query cannot resolve the column name 'heartrateheartrateheartrate' given the input columns 'heartrate' and 'age'. This means that there is no column in the table named 'heartrateheartrateheartrate', and the query is invalid. A possible cause of this error is a typo or a copy-paste mistake in the query. To fix this error, the query should use a valid column name that exists in the table, such as 'heartrate'. Reference: AnalysisException

## 問題 #127

.....

NewDumps為Databricks Databricks-Certified-Professional-Data-Engineer 認證考試準備的培訓包括Databricks Databricks-Certified-Professional-Data-Engineer認證考試的模擬測試題和當前考試題。在互聯網上你也可以看到幾個也提供相關的培訓的網站，但是你比較之後，你就會發現NewDumps的關於Databricks Databricks-Certified-Professional-Data-Engineer 認證考試的培訓比較有針對性，不僅品質是最高的，而且內容是最全面的。

**Databricks-Certified-Professional-Data-Engineer 證照指南:** <https://www.newdumps.com/Databricks-Certified-Professional-Data-Engineer-exam-new-dumps.html>

所有購買我們“Databricks-Certified-Professional-Data-Engineer 題庫學習資料”的客戶，都將獲得三個月免費更新新題學

習資料的售後服務(三個月內參加且通過考試的客戶將不提供更新), 確保您能壹直擁有最新的認證考試題庫學習資料, Databricks Certified Professional Data Engineer Exam故障排除考試Databricks-Certified-Professional-Data-Engineer考生需要花費很多美元費用在全球的VUE考試機構預約考試時間, 主要測試Mac OS X實用知識, 重點強調Databricks Certified Professional Data Engineer Exam故障排除, 並適用於Mac OS系統, Databricks-Certified-Professional-Data-Engineer 題庫資料能有效地幫考生通過考試, 所有購買 NewDumps Databricks-Certified-Professional-Data-Engineer證照指南NewDumps Databricks-Certified-Professional-Data-Engineer證照指南 Databricks-Certified-Professional-Data-Engineer證照指南認證題庫學習資料的客戶都將得到半年的免費升級服務, 確保您的題庫學習資料始終保持最新狀態, 利用 Databricks-Certified-Professional-Data-Engineer 考試資料, 你肯定可以得到你想要的成功。

壹個豪華的話的語氣低低的, 已經固執的拉住蕭峰的胳膊不讓去開門, 慶戰半個時辰, 兩敗俱傷, 所有購買我們“Databricks-Certified-Professional-Data-Engineer題庫學習資料”的客戶, 都將獲得三個月免費更新新題學習資料的售後服務(三個月內參加且通過考試的客戶將不提供更新), 確保您能壹直擁有最新的認證考試題庫學習資料。

## 值得信賴的Databricks-Certified-Professional-Data-Engineer測試 |第一次嘗試輕鬆學習並通過考試和最佳的Databricks-Certified-Professional-Data-Engineer: Databricks Certified Professional Data Engineer Exam

Databricks Certified Professional Data Engineer Exam故障排除考試Databricks-Certified-Professional-Data-Engineer考生需要花費很多美元費用在全球的VUE考試機構預約考試時間, 主要測試Mac OS X實用知識, 重點強調Databricks Certified Professional Data Engineer Exam故障排除, 並適用於Mac OS系統, Databricks-Certified-Professional-Data-Engineer 題庫資料能有效地幫考生通過考試。

所有購買 NewDumpsNewDumps Databricks Certification認證題庫學習資料的客戶都將得到半年的免費升級服務, 確保您的題庫學習資料始終保持最新狀態, 利用 Databricks-Certified-Professional-Data-Engineer 考試資料, 你肯定可以得到你想要的成功。

- Databricks-Certified-Professional-Data-Engineer 考試題庫 – 專業的 Databricks-Certified-Professional-Data-Engineer 認證題庫學習資料  ⇒ [www.vcesoft.com](http://www.vcesoft.com) ⇒最新▶ Databricks-Certified-Professional-Data-Engineer ◀問題集合 Databricks-Certified-Professional-Data-Engineer證照信息
- Databricks-Certified-Professional-Data-Engineer考試資料  Databricks-Certified-Professional-Data-Engineer最新考證  最新Databricks-Certified-Professional-Data-Engineer考證  到✓ [www.newdumpsdf.com](http://www.newdumpsdf.com) ✓搜尋➡ Databricks-Certified-Professional-Data-Engineer 以獲取免費下載考試資料新版Databricks-Certified-Professional-Data-Engineer考古題
- Databricks-Certified-Professional-Data-Engineer最新考證  最新Databricks-Certified-Professional-Data-Engineer考證  新版Databricks-Certified-Professional-Data-Engineer考古題↑在( [www.testpdf.net](http://www.testpdf.net) ) 網站上查找 ➡ Databricks-Certified-Professional-Data-Engineer 的最新題庫最新Databricks-Certified-Professional-Data-Engineer題庫資源
- 選擇Databricks-Certified-Professional-Data-Engineer測試表示您已通過Databricks Certified Professional Data Engineer Exam無憂  透過▶ [www.newdumpsdf.com](http://www.newdumpsdf.com) ◀搜索▶ Databricks-Certified-Professional-Data-Engineer ◀免費下載考試資料Databricks-Certified-Professional-Data-Engineer權威認證
- Databricks-Certified-Professional-Data-Engineer最新考證 ♥ Databricks-Certified-Professional-Data-Engineer考題套裝  Databricks-Certified-Professional-Data-Engineer考題套裝  { [www.testpdf.net](http://www.testpdf.net) } 上的免費下載 ( Databricks-Certified-Professional-Data-Engineer ) 頁面立即打開最新Databricks-Certified-Professional-Data-Engineer題庫資源
- 真實的Databricks-Certified-Professional-Data-Engineer測試&保證Databricks Databricks-Certified-Professional-Data-Engineer考試成功與頂級的Databricks-Certified-Professional-Data-Engineer證照指南  立即在【 [www.newdumpsdf.com](http://www.newdumpsdf.com) 】上搜尋 ➡ Databricks-Certified-Professional-Data-Engineer 並免費下載Databricks-Certified-Professional-Data-Engineer考題套裝
- 專業的Databricks Databricks-Certified-Professional-Data-Engineer測試是行業領先材料&授權的Databricks-Certified-Professional-Data-Engineer證照指南  ▶ [www.testpdf.net](http://www.testpdf.net) ◀是獲取【 Databricks-Certified-Professional-Data-Engineer 】免費下載的最佳網站Databricks-Certified-Professional-Data-Engineer最新考證
- 有效的Databricks-Certified-Professional-Data-Engineer測試: Databricks Certified Professional Data Engineer Exam - Databricks Databricks-Certified-Professional-Data-Engineer證照指南確定通過  請在 [www.newdumpsdf.com](http://www.newdumpsdf.com)  網站上免費下載 ➡ Databricks-Certified-Professional-Data-Engineer 題庫Databricks-Certified-Professional-Data-Engineer證照信息
- 新版Databricks-Certified-Professional-Data-Engineer考古題  最新Databricks-Certified-Professional-Data-Engineer題庫資源  Databricks-Certified-Professional-Data-Engineer證照資訊  到➡ [www.testpdf.net](http://www.testpdf.net) 搜索➡ Databricks-Certified-Professional-Data-Engineer 輕鬆取得免費下載Databricks-Certified-Professional-Data-Engineer考古題介紹
- 選擇Databricks-Certified-Professional-Data-Engineer測試表示您已通過Databricks Certified Professional Data Engineer Exam無憂  複製網址 ➡ [www.newdumpsdf.com](http://www.newdumpsdf.com) 打開並搜索 ➡ Databricks-Certified-Professional-Data-Engineer 免費下載Databricks-Certified-Professional-Data-Engineer考試資料

- 100%權威的Databricks-Certified-Professional-Data-Engineer測試，最好的考試指南幫助妳快速通過Databricks-Certified-Professional-Data-Engineer考試 □ 請在《[www.pdfexamdumps.com](http://www.pdfexamdumps.com)》網站上免費下載{ Databricks-Certified-Professional-Data-Engineer }題庫Databricks-Certified-Professional-Data-Engineer信息資訊
- [gerardkhdu872982.blog2freedom.com](http://gerardkhdu872982.blog2freedom.com), [guidemysocial.com](http://guidemysocial.com), [rajanccpy063469.topbloghub.com](http://rajanccpy063469.topbloghub.com), [bookmarkcolumn.com](http://bookmarkcolumn.com), [cheapbookmarking.com](http://cheapbookmarking.com), [elodievouq947046.wikihearsay.com](http://elodievouq947046.wikihearsay.com), [elijahujpv210157.wikigop.com](http://elijahujpv210157.wikigop.com), [chiaraonex828024.liberty-blog.com](http://chiaraonex828024.liberty-blog.com), [susanyqjg443786.wiki-racconti.com](http://susanyqjg443786.wiki-racconti.com), [theoqnvml75988.aboutyoublog.com](http://theoqnvml75988.aboutyoublog.com), Disposable vapes

P.S. NewDumps在Google Drive上分享了免費的、最新的Databricks-Certified-Professional-Data-Engineer考試題庫：<https://drive.google.com/open?id=15kTO-sV-WsSIUBI6-c6nxIqW3x7goyYB>