

Valid Databricks-Certified-Data-Engineer-Professional Exam Guide & Databricks-Certified-Data-Engineer-Professional Valid Learning Materials



Purpose of this Exam Guide

The purpose of this exam guide is to give you an overview of the exam and what is covered on the exam to help you determine your exam readiness. This document will get updated anytime there are any changes to an exam (and when those changes will take effect on an exam) so that you can be prepared. This version covers the currently live version as of August 1, 2023

Audience Description

The Databricks Certified Data Engineer Professional certification exam assesses an individual's ability to use Databricks to perform advanced data engineering tasks. This includes an understanding of the Databricks platform and developer tools like Apache Spark, Delta Lake, MLflow, and the Databricks CLI and REST API. It also assesses the ability to build optimized and cleaned ETL pipelines. Additionally, modeling data into a Lakehouse using knowledge of general data modeling concepts will also be assessed. Finally, ensuring that data pipelines are secure, reliable, monitored, and tested before deployment will also be included in this exam. Individuals who pass this certification exam can be expected to complete advanced data engineering tasks using Databricks and its associated tools.

About the Exam

- Number of items: 60 multiple-choice questions
- Time limit: 120 minutes
- Registration fee: USD 200, plus applicable taxes as required per local law
- Delivery method: Online Proctored
- Test aides: none allowed.
- Prerequisite: None required; course attendance and 1 year of hands-on experience in Databricks is highly recommended
- Validity: 2 years
- Unscored Content: Exams may include unscored items to gather statistical information for future use. These items are not identified on the form and do not impact your score, and additional time is factored into account for this content.

2026 Latest PDF4Test Databricks-Certified-Data-Engineer-Professional PDF Dumps and Databricks-Certified-Data-Engineer-Professional Exam Engine Free Share: <https://drive.google.com/open?id=1YW8YRrWQiqGdgiefSiGZWOlmfPZB0yJ>

The Internet is increasingly becoming a platform for us to work and learn, while many products are unreasonable in web design, and too much information is not properly classified. Our Databricks-Certified-Data-Engineer-Professional exam materials draw lessons from the experience of failure, will all kinds of Databricks-Certified-Data-Engineer-Professional qualification examination has carried on the classification of clear layout, at the same time the user when they entered the Databricks-Certified-Data-Engineer-Professional Study Guide materials page in the test module classification of clear, convenient to use a very short time to find what they want to study for the Databricks-Certified-Data-Engineer-Professional exam

Unlike many other learning materials, our Databricks-Certified-Data-Engineer-Professional study materials are specially designed to help people pass the exam in a more productive and time-saving way, and such an efficient feature makes it a wonderful assistant in personal achievement as people have less spare time nowadays. On the other hand, Databricks-Certified-Data-Engineer-Professional Study Materials are aimed to help users make best use of their sporadic time by adopting flexible and safe study access.

>> Valid Databricks-Certified-Data-Engineer-Professional Exam Guide <<

Databricks-Certified-Data-Engineer-Professional Valid Learning Materials - New Databricks-Certified-Data-Engineer-Professional Exam Guide

It is seen as a challenging task to pass the Databricks-Certified-Data-Engineer-Professional exam. Tests like these demand profound knowledge. The Databricks Databricks-Certified-Data-Engineer-Professional certification is absolute proof of your talent and ticket to high-paying jobs in a renowned firm. Databricks Databricks-Certified-Data-Engineer-Professional test every year to shortlist applicants who are eligible for the Databricks-Certified-Data-Engineer-Professional exam certificate.

Databricks Certified Data Engineer Professional Exam Sample Questions (Q215-Q220):

NEW QUESTION # 215

The data engineering team has configured a job to process customer requests to be forgotten (have their data deleted). All user data that needs to be deleted is stored in Delta Lake tables using default table settings.

The team has decided to process all deletions from the previous week as a batch job at 1am each Sunday. The total duration of this job is less than one hour. Every Monday at 3am, a batch job executes a series of VACUUM commands on all Delta Lake tables throughout the organization.

The compliance officer has recently learned about Delta Lake's time travel functionality. They are concerned that this might allow continued access to deleted data.

Assuming all delete logic is correctly implemented, which statement correctly addresses this concern?

- A. Because Delta Lake's delete statements have ACID guarantees, deleted records will be permanently purged from all storage systems as soon as a delete job completes.
- B. Because Delta Lake time travel provides full access to the entire history of a table, deleted records can always be recreated by users with full admin privileges.
- C. Because the default data retention threshold is 24 hours, data files containing deleted records will be retained until the vacuum job is run the following day.
- **D. Because the default data retention threshold is 7 days, data files containing deleted records will be retained until the vacuum job is run 8 days later.**
- E. Because the vacuum command permanently deletes all files containing deleted records, deleted records may be accessible with time travel for around 24 hours.

Answer: D

Explanation:

<https://learn.microsoft.com/en-us/azure/databricks/delta/vacuum>

NEW QUESTION # 216

A member of the data engineering team has submitted a short notebook that they wish to schedule as part of a larger data pipeline.

Assume that the commands provided below produce the logically correct results when run as presented.

Get Latest & Actual Certified-Data-Engineer-Professional Exam's Question and Answers from

Cmd 1

```
rawDF = spark.table("raw_data")
```

Cmd 2

```
rawDF.printSchema()
```

Cmd 3

```
flattenedDF = rawDF.select("*", "values.*")
```

Cmd 4

```
finalDF = flattenedDF.drop("values")
```

Cmd 5

```
finalDF.explain()
```

Cmd 6

```
display(finalDF)
```

Cmd 7

```
finalDF.write.mode("append").saveAsTable("flat_data")
```

Which command should be removed from the notebook before scheduling it as a job?

- A. Cmd 2
- B. Cmd 4
- C. Cmd 5
- D. Cmd 3
- **E. Cmd 6**

Answer: E

Explanation:

When scheduling a Databricks notebook as a job, it's generally recommended to remove or modify commands that involve displaying output, such as using the `display()` function. Displaying data using `display()` is an interactive feature designed for exploration and visualization within the notebook interface and may not work well in a production job context.

The `finalDF.explain()` command, which provides the execution plan of the DataFrame transformations and actions, is often useful for debugging and optimizing queries. While it doesn't Get Latest & Actual Certified-Data-Engineer-Professional Exam's Question and Answers from display interactive visualizations like `display()`, it can still be informative for understanding how Spark is executing the operations on your DataFrame.

An organization processes customer data from web and mobile applications. Data includes names, emails, phone numbers, and location history. Data arrives both as batch files (from SFTP daily) and streaming JSON events (from Kafka in real-time).

To comply with data privacy policies, the following requirements must be met:

- Personally Identifiable Information (PII) such as email, phone number, and IP address must be masked or anonymized before storage.
- Both batch and streaming pipelines must apply consistent PII handling.
- Masking logic must be auditable and reproducible.
- The masked data must remain usable for downstream analytics.

How should the data engineer design a compliant data pipeline on Databricks that supports both batch and streaming modes, applies data masking to PII, and maintains traceability for audits?

- A. Ingest both batch and streaming data using Lakeflow Declarative Pipelines, and apply masking via Unity Catalog column masks at read time to avoid modifying the data during ingestion.
- B. Load batch data with notebooks and ingest streaming data with SQL Warehouses; use Unity Catalog column masks on Silver tables to redact fields after storage.
- C. Use Lakeflow Declarative Pipelines for batch and streaming ingestion, define a PII masking function, and apply it during Bronze ingestion before writing to Delta Lake.
- D. Allow PII to be stored unmasked in Bronze for lineage tracking, then apply masking logic in Gold tables used for reporting.

Answer: C

Explanation:

Databricks recommends applying data masking or anonymization before persisting PII to ensure compliance with privacy regulations such as GDPR and HIPAA. In a Lakeflow Declarative Pipeline, developers can define custom Python or SQL-based masking functions to standardize PII handling across both batch and streaming inputs.

This approach ensures that data entering the Delta Lake is already anonymized, guaranteeing consistent and auditable behavior. By applying masking during ingestion (in the Bronze layer), audit trails are preserved through pipeline event logs.

While Unity Catalog column masks (option C) can enforce dynamic masking at query time, they do not prevent PII storage. Thus, option D aligns with the best practice of securing PII before storage, while still supporting reproducibility and analytics usability.

NEW QUESTION # 218

The data science team has created and logged a production model using MLflow. The following code correctly imports and applies the production model to output the predictions as a new DataFrame named preds with the schema "customer_id LONG, predictions DOUBLE, date DATE".

```
from pyspark.sql.functions import current_date

model = mlflow.pyfunc.spark_udf(spark, model_uri="models:/churn/pred")
df = spark.table("customers")
columns = ["account_age", "time_since_last_seen", "app_rating"]
preds = (df.select(
    "customer_id",
    model(*columns).alias("predictions"),
    current_date().alias("date")
))
```

The data science team would like predictions saved to a Delta Lake table with the ability to compare all predictions across time. Churn predictions will be made at most once per day.

Which code block accomplishes this task while minimizing potential compute costs?

```
(preds.write
 .format("delta")
 .mode("overwrite")
 .saveAsTable("churn_preds"))
```

- A. `preds.write.format("delta").mode("overwrite").saveAsTable("churn_preds")`
- B. `preds.write.mode("append").saveAsTable("churn_preds")`

```
(preds.writeStream
  databricks
  .outputMode("overwrite")
  .option("checkpointPath", "_checkpoints/churn_preds")
  .start("/preds/churn_preds")
)
```

- C.


```
(preds.writeStream
        .outputMode("append")
        .option("checkpointPath", "_checkpoints/churn_preds")
        .table("churn_preds")
        databricks
      )
```
- D.)
- E. preds.write.format("delta").save("/preds/churn_preds")

Answer: B

NEW QUESTION # 219

A junior data engineer has manually configured a series of jobs using the Databricks Jobs UI. Upon reviewing their work, the engineer realizes that they are listed as the "Owner" for each job. They attempt to transfer "Owner" privileges to the "DevOps" group, but cannot successfully accomplish this task. Which statement explains what is preventing this privilege transfer?

- A. Other than the default "admins" group, only individual users can be granted privileges on jobs.
- B. A user can only transfer job ownership to a group if they are also a member of that group.
- C. Databricks jobs must have exactly one owner; "Owner" privileges cannot be assigned to a group.
- D. The creator of a Databricks job will always have "Owner" privileges; this configuration cannot be changed.
- E. Only workspace administrators can grant "Owner" privileges to a group.

Answer: C

Explanation:

A job cannot have more than one owner. A job cannot have a group as an owner.

NEW QUESTION # 220

.....

You want to get the most practical and useful certificate which can reflect your ability in some area. If you choose to attend the test Databricks-Certified-Data-Engineer-Professional certification buying our Databricks-Certified-Data-Engineer-Professional study materials can help you pass the test and get the valuable certificate. Our company has invested a lot of personnel, technology and capitals on our products and is always committed to provide the top-ranking Databricks-Certified-Data-Engineer-Professional Study Materials to the clients and serve for the client wholeheartedly.

Databricks-Certified-Data-Engineer-Professional Valid Learning Materials: <https://www.pdf4test.com/Databricks-Certified-Data-Engineer-Professional-dump-torrent.html>

Databricks Valid Databricks-Certified-Data-Engineer-Professional Exam Guide Make sure you are filling the correct username and password, you can apply for getting password through email, Databricks Valid Databricks-Certified-Data-Engineer-Professional Exam Guide We are not afraid of your disturbing; please choose our products as your top priority, Databricks Valid Databricks-Certified-Data-Engineer-Professional Exam Guide It is a simulation test system and you can do elevation for your knowledge, thus you can improve yourself with effective method, You may know from your friends, colleagues or classmates that some Databricks-Certified-Data-Engineer-Professional actual test dumps pdf is very useful to help them pass exams easily.

Inside, he'll give you the latest updates on new Databricks-Certified-Data-Engineer-Professional Detailed Study Dumps players, services, and sites, the truth about copyright issues, and sources for legally freemusic, Many small businesses have already been Databricks-Certified-Data-Engineer-Professional Valid Learning Materials established by placing a cell phone and a microloan in the hands of industrious entrepreneurs.

Free PDF Databricks - Databricks-Certified-Data-Engineer-Professional - Accurate Valid Databricks Certified Data Engineer Professional Exam Exam

<https://drive.google.com/open?id=1YW8YRrWQiQGdjefSiGZW0lmmfPZB0yJ>