

# Study Databricks-Certified-Professional-Data-Engineer Group - Databricks-Certified-Professional-Data-Engineer Reliable Guide Files



BONUS!!! Download part of Prep4pass Databricks-Certified-Professional-Data-Engineer dumps for free:  
[https://drive.google.com/open?id=1RgoPXbpqFsl6Wyt1z9IOMzeGpuKMMG\\_](https://drive.google.com/open?id=1RgoPXbpqFsl6Wyt1z9IOMzeGpuKMMG_)

With the qualification certificate, you are qualified to do this professional job. Therefore, getting the test Databricks-Certified-Professional-Data-Engineer certification is of vital importance to our future employment. Our Databricks-Certified-Professional-Data-Engineer practice materials are updating according to the precise of the real exam. Our test prep can help you to conquer all difficulties you may encounter. In other words, we will be your best helper. Pass the Databricks-Certified-Professional-Data-Engineer Exam, for most people, is an ability to live the life they want, and the realization of these goals needs to be established on a good basis of having a good job. A good job requires a certain amount of competence, and the most intuitive way to measure competence is whether you get a series of the test Databricks-Certified-Professional-Data-Engineer certification and obtain enough qualifications.

Databricks-Certified-Professional-Data-Engineer exam consists of multiple-choice questions and hands-on, real-world scenarios that test the candidate's ability to design, build, and deploy data pipelines on Databricks. Databricks-Certified-Professional-Data-Engineer Exam covers various topics, including data engineering concepts, Databricks architecture, data processing using Spark, and data integration with other systems. Databricks Certified Professional Data Engineer Exam certification program provides a comprehensive learning experience that prepares candidates to become skilled data engineers and provides them with a competitive edge in the job market.

>> Study Databricks-Certified-Professional-Data-Engineer Group <<

## Databricks Databricks-Certified-Professional-Data-Engineer Reliable Guide Files | Databricks-Certified-Professional-Data-Engineer Test Centres

Success in the Databricks-Certified-Professional-Data-Engineer test of the Databricks Databricks-Certified-Professional-Data-Engineer credential is essential in today's industry to verify the skills and get well-paying jobs in reputed firms around the whole globe. Earning the Databricks Certified Professional Data Engineer Exam Databricks-Certified-Professional-Data-Engineer Certification sharpens your skills and helps you to accelerate your career in today's cut throat competition in the Databricks industry. It is not easy to clear the Databricks-Certified-Professional-Data-Engineer exam on the maiden attempt.

## Databricks Certified Professional Data Engineer Exam Sample Questions (Q105-Q110):

### NEW QUESTION # 105

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

**Answer: A**

Explanation:

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>

### NEW QUESTION # 106

Given the following error traceback (from `display(df.select(3*"heartrate"))`) which shows `AnalysisException: cannot resolve 'heartrateheartrateheartrate'`, 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 type error because a column object cannot be multiplied.
- D. There is a syntax error because the heartrate column is not correctly identified as a column.

**Answer: A**

Explanation:

Comprehensive and Detailed Explanation From Exact Extract:

Exact extract: "select() expects column names or Column expressions."

Exact extract: "Use col('name') (or df['name']) to reference a column; Python string operations act on strings, not columns." In

Python, "heartrate"\*3 is string multiplication that yields "heartrateheartrateheartrate". select() then looks for a column literally named that string. Because no such column exists, Spark raises an analysis error that it "cannot resolve" that column. The correct expression to multiply the column is `df.select((F.col("heartrate")*3).alias("heartrate_x3"))`.

Reference:

### NEW QUESTION # 107

The data governance team has instituted a requirement that all tables containing Personal Identifiable Information (PH) must be clearly annotated. This includes adding column comments, table comments, and setting the custom table property "contains\_pii" = true.

The following SQL DDL statement is executed to create a new table:

Which command allows manual confirmation that these three requirements have been met?

- A. DESCRIBE DETAIL dev.pii test
- B. DESCRIBE HISTORY dev.pii test
- **C. DESCRIBE EXTENDED dev.pii test**
- D. SHOW TBLPROPERTIES dev.pii test
- E. SHOW TABLES dev

**Answer: C**

Explanation:

Explanation

This is the correct answer because it allows manual confirmation that these three requirements have been met.

The requirements are that all tables containing Personal Identifiable Information (PII) must be clearly annotated, which includes adding column comments, table comments, and setting the custom table property "contains\_pii" = true. The DESCRIBE EXTENDED command is used to display detailed information about a table, such as its schema, location, properties, and comments. By using this command on the dev.pii\_test table, one can verify that the table has been created with the correct column comments, table comment, and custom table property as specified in the SQL DDL statement. Verified References: [Databricks Certified Data Engineer Professional], under "Lakehouse" section; Databricks Documentation, under "DESCRIBE EXTENDED" section.

### NEW QUESTION # 108

A junior data engineer has been asked to develop a streaming data pipeline with a grouped aggregation using DataFrame df. The pipeline needs to calculate the average humidity and average temperature for each non-overlapping five-minute interval. Incremental state information should be maintained for 10 minutes for late-arriving data.

Streaming DataFrame df has the following schema:

```
"device_id INT, event_time TIMESTAMP, temp FLOAT, humidity FLOAT"
```

Code block:

Choose the response that correctly fills in the blank within the code block to complete this task.

- A. slidingWindow("event\_time", "10 minutes")
- **B. withWatermark("event\_time", "10 minutes")**
- C. delayWrite("event\_time", "10 minutes")
- D. awaitArrival("event\_time", "10 minutes")
- E. await("event\_time + '10 minutes'")

**Answer: B**

Explanation:

Explanation

The correct answer is A. withWatermark("event\_time", "10 minutes"). This is because the question asks for incremental state information to be maintained for 10 minutes for late-arriving data. The withWatermark method is used to define the watermark for late data. The watermark is a timestamp column and a threshold that tells the system how long to wait for late data. In this case, the watermark is set to 10 minutes. The other options are incorrect because they are not valid methods or syntax for watermarking in Structured Streaming. References:

Watermarking: <https://docs.databricks.com/spark/latest/structured-streaming/watermarks.html> Windowed aggregations: <https://docs.databricks.com/spark/latest/structured-streaming/window-operations.html>

### NEW QUESTION # 109

What is the purpose of the bronze layer in a Multi-hop architecture?

- A. Can be used to eliminate duplicate records
- B. Used as a data source for Machine learning applications.
- C. Perform data quality checks, corrupt data quarantined
- D. Contains aggregated data that is to be consumed into Silver
- **E. Provides efficient storage and querying of full unprocessed history of data**

**Answer: E**

Explanation:

Explanation

The answer is Provides efficient storage and querying of full unprocessed history of data Medallion Architecture - Databricks Bronze Layer:

- 1.Raw copy of ingested data
- 2.Replaces traditional data lake
- 3.Provides efficient storage and querying of full, unprocessed history of data
- 4.No schema is applied at this layer

Exam focus: Please review the below image and understand the role of each layer(bronze, silver, gold) in medallion architecture, you will see varying questions targeting each layer and its purpose.

Sorry I had to add the watermark some people in Udemy are copying my content.



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, Disposable vapes

BTW, DOWNLOAD part of Prep4pass Databricks-Certified-Professional-Data-Engineer dumps from Cloud Storage:  
[https://drive.google.com/open?id=1RgoPXbpqFsl6WyT1z9IOMzeGpuKMMG\\_](https://drive.google.com/open?id=1RgoPXbpqFsl6WyT1z9IOMzeGpuKMMG_)