

Databricks-Certified-Professional-Data-Engineer valid cram guide & Databricks-Certified-Professional-Data-Engineer training prep & Databricks-Certified-Professional-Data-Engineer sure pass



Test4Cram almost aimed to meet the needs of all candidates who want to pass the Databricks-Certified-Professional-Data-Engineer exam. If someone who don't have enough time to prepare for their exam, our website provide they with test answers which only need 20-30 hours to grasp; If someone who worry about failed the Databricks-Certified-Professional-Data-Engineer Exam, our website can guarantee that they can get full refund. In summary, the easiest way to prepare for Databricks-Certified-Professional-Data-Engineer certification exam is to complete Databricks-Certified-Professional-Data-Engineer study material.

Databricks Databricks-Certified-Professional-Data-Engineer certification is a valuable credential for professionals working with big data and data engineering. Databricks Certified Professional Data Engineer Exam certification validates the candidates' technical skills in working with big data projects implemented on the Databricks platform. It aims to create a standard for big data engineering skills and provides a valuable addition to a candidate's resume. Earning this certification opens up doors for career advancement and can improve a professional's ability to secure a high-paying job in the big data industry.

The Databricks Databricks-Certified-Professional-Data-Engineer Exam is designed to test the candidate's ability to work with Databricks in a real-world setting. Candidates are required to demonstrate their ability to design and implement data pipelines that are scalable, efficient, and reliable. They must also be able to troubleshoot issues that arise during the data engineering process and optimize performance to ensure that pipelines run smoothly.

>> **Databricks-Certified-Professional-Data-Engineer Test Valid** <<

Databricks-Certified-Professional-Data-Engineer Reliable Braindumps Pdf | Databricks-Certified-Professional-Data-Engineer Download Pdf

I am proud to tell you that our company is definitely one of the most authoritative companies in the international market for Databricks-Certified-Professional-Data-Engineer exam. What's more, we will provide the most considerate after sale service for our customers in twenty four hours a day seven days a week, therefore, our company is really the best choice for you to buy the Databricks-Certified-Professional-Data-Engineer Training Materials. You can just feel rest assured that our after sale service staffs are always here waiting for offering you our services on our Databricks-Certified-Professional-Data-Engineer exam questions. Please feel free to contact us. You will be surprised by our good Databricks-Certified-Professional-Data-Engineer study guide.

Databricks Certified Professional Data Engineer Exam Sample Questions (Q156-Q161):

NEW QUESTION # 156

Below sample input data contains two columns, one cartId also known as session id, and the second column is called items, every time a customer makes a change to the cart this is stored as an array in the table, the Marketing team asked you to create a unique list of items that were ever added to the cart by each customer, fill in blanks by choosing the appropriate array function so the query produces below expected result as shown below.

Schema: cartId INT, items Array<INT>

Sample Data

1. SELECT cartId, ____ (____(items)) as items
2. FROM carts GROUP BY cartId

Expected result:

cartId items

1 [1,100,200,300,250]

- A. ARRAY_UNION, FLATTEN
- **B. ARRAY_UNION, COLLECT_SET**
- C. FLATTEN, COLLECT_UNION
- D. ARRAY_UNION, ARRAY_DISTINCT
- E. ARRAY_DISTINCT, ARRAY_UNION

Answer: B

Explanation:

Explanation

COLLECT SET is a kind of aggregate function that combines a column value from all rows into a unique list ARRAY_UNION combines and removes any duplicates, Graphical user interface, application Description automatically generated with medium confidence

NEW QUESTION # 157

A data engineering team has a time-consuming data ingestion job with three data sources. Each notebook takes about one hour to load new data. One day, the job fails because a notebook update introduced a new required configuration parameter. The team must quickly fix the issue and load the latest data from the failing source.

Which action should the team take?

- A. Repair the run with the new parameter.
- B. Share the analysis with the failing notebook owner so that they can fix it quickly.
- C. Update the task by adding the missing task parameter, and manually run the job.
- **D. Repair the run with the new parameter, and update the task by adding the missing task parameter.**

Answer: D

Explanation:

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

The repair run capability in Databricks Jobs allows re-execution of failed tasks without re-running successful ones. When a parameterized job fails due to missing or incorrect task configuration, engineers can perform a repair run to fix inputs or parameters and resume from the failed state.

This approach saves time, reduces cost, and ensures workflow continuity by avoiding unnecessary recomputation. Additionally, updating the task definition with the missing parameter prevents future runs from failing.

Running the job manually (B) loses run context; (C) alone does not prevent recurrence; (D) delays resolution. Thus, A follows the correct operational and recovery practice.

NEW QUESTION # 158

Which of the following commands results in the successful creation of a view on top of the delta stream(stream on delta table)?

- **A. Spark.readStream.format("delta").table("sales").createOrReplaceTempView("streaming_vw")**
- B. Spark.read.format("delta").stream("sales").createOrReplaceTempView("streaming_vw")
- C. Spark.read.format("delta").table("sales").trigger("stream").createOrReplaceTempView("streaming_vw")
- D. You can not create a view on streaming data source.
- E. Spark.read.format("delta").table("sales").createOrReplaceTempView("streaming_vw")
- F. Spark.read.format("delta").table("sales").mode("stream").createOrReplaceTempView("streaming_vw")

Answer: A

Explanation:

Explanation

The answer is

`Spark.readStream.table("sales").createOrReplaceTempView("streaming_vw")` When you load a Delta table as a stream source and use it in a streaming query, the query processes all of the data present in the table as well as any new data that arrives after the stream is started.

You can load both paths and tables as a stream, you also have the ability to ignore deletes and changes (updates, Merge, overwrites) on the delta table.

Here is more information,

<https://docs.databricks.com/delta/delta-streaming.html#delta-table-as-a-source>

NEW QUESTION # 159

You are working on a process to load external CSV files into a delta table by leveraging the COPY INTO command, but after running the command for the second time no data was loaded into the table name, why is that?

1. COPY INTO table_name
2. FROM 'dbfs:/mnt/raw/*.csv'
3. FILEFORMAT = CSV

- A. COPY INTO only works one time data load
- B. Run REFRESH TABLE sales before running COPY INTO
- C. COPY INTO does not support incremental load, use AUTO LOADER
- D. Use incremental = TRUE option to load new files
- E. COPY INTO did not detect new files after the last load

Answer: E

Explanation:

Explanation

The answer is COPY INTO did not detect new files after the last load,

COPY INTO keeps track of files that were successfully loaded into the table, the next time when the COPY INTO runs it skips them

FYI, you can change this behavior by using COPY_OPTIONS 'force' = 'true', when this option is enabled all files in the path/pattern are loaded.

1. COPY INTO table_identifier
2. FROM [file_location | (SELECT identifier_list FROM file_location)]
3. FILEFORMAT = data_source
4. [FILES = [file_name, ... | PATTERN = 'regex_pattern']
5. [FORMAT_OPTIONS ('data_source_reader_option' = 'value', ...)]
6. [COPY_OPTIONS 'force' = ('false'|'true')]

NEW QUESTION # 160

Which is a key benefit of an end-to-end test?

- A. It pinpoint errors in the building blocks of your application.
- B. It closely simulates real world usage of your application.
- C. It provides testing coverage for all code paths and branches.
- D. It makes it easier to automate your test suite

Answer: B

Explanation:

End-to-end testing is a methodology used to test whether the flow of an application, from start to finish, behaves as expected. The key benefit of an end-to-end test is that it closely simulates real-world, user behavior, ensuring that the system as a whole operates correctly.

Reference:

Software Testing: End-to-End Testing

NEW QUESTION # 161

.....

