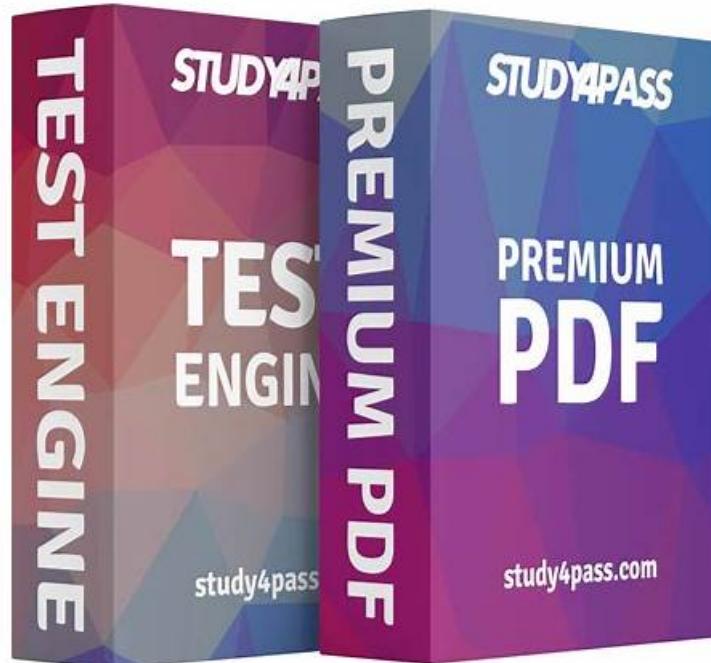


Valid Associate-Developer-Apache-Spark-3.5 Exam Prep & Accurate Associate-Developer-Apache-Spark-3.5 Answers



What's more, part of that DumpsTests Associate-Developer-Apache-Spark-3.5 dumps now are free:
<https://drive.google.com/open?id=1nZfWv9fb4aYQrsueQBWffpAKTEt1B-Y1>

DumpsTests Databricks Associate-Developer-Apache-Spark-3.5 Exam Questions are made in accordance with the latest syllabus and the actual Databricks Associate-Developer-Apache-Spark-3.5 certification exam. We constantly upgrade our training materials, all the products you get with one year of free updates. You can always extend the to update subscription time, so that you will get more time to fully prepare for the exam. If you still confused to use the training materials of DumpsTests, then you can download part of the examination questions and answers in DumpsTests website. It is free to try, and if it is suitable for you, then go to buy it, to ensure that you will never regret.

There is always a fear of losing the Associate-Developer-Apache-Spark-3.5 exam and this causes you may loss your money and waste the time. There is no such issue if you study our Associate-Developer-Apache-Spark-3.5 exam questions. Your money and exam attempt is bound to award you a sure and definite success if you study with our Associate-Developer-Apache-Spark-3.5 Study Guide to prapare for the exam. According to our data, our pass rate of the Associate-Developer-Apache-Spark-3.5 practice engine is high as 98% to 100%. So if you choose our Associate-Developer-Apache-Spark-3.5 learning quiz, you will pass for sure.

>> Valid Associate-Developer-Apache-Spark-3.5 Exam Prep <<

Valid Associate-Developer-Apache-Spark-3.5 Exam Prep - The Best Databricks Accurate Associate-Developer-Apache-Spark-3.5 Answers: Databricks Certified Associate Developer for Apache Spark 3.5 - Python

If you feel that you just don't have enough competitiveness to find a desirable job. Then it is time to strengthen your skills. Our Associate-Developer-Apache-Spark-3.5 exam simulating will help you master the most popular skills in the job market. Then you will have a greater chance to find a desirable job. Also, it doesn't matter whether have basic knowledge about the Associate-Developer-Apache-Spark-3.5 training quiz for the content of our Associate-Developer-Apache-Spark-3.5 study guide contains all the exam keypoints which you need to cope with the real exam

Databricks Certified Associate Developer for Apache Spark 3.5 - Python Sample Questions (Q68-Q73):

NEW QUESTION # 68

15 of 55.

A data engineer is working on a Streaming DataFrame (`streaming_df`) with the following streaming data:

```
id
name
count
timestamp
1
Delhi
20
2024-09-19T10:11
1
Delhi
50
2024-09-19T10:12
2
London
50
2024-09-19T10:15
3
Paris
30
2024-09-19T10:18
3
Paris
20
2024-09-19T10:20
4
Washington
10
2024-09-19T10:22
```

Which operation is supported with `streaming_df`?

- A. `streaming_df.select(countDistinct('name'))`
- B. `streaming_df.count()`
- C. `streaming_df.filter("count < 30")`
- D. `streaming_df.show()`

Answer: C

Explanation:

In Structured Streaming, only transformation operations are allowed on streaming DataFrames. These include `select()`, `filter()`, `where()`, `groupBy()`, `withColumn()`, etc.

Example of supported transformation:

```
filtered_df = streaming_df.filter("count < 30")
```

However, actions such as `count()`, `show()`, and `collect()` are not supported directly on streaming DataFrames because streaming queries are unbounded and never finish until stopped.

To perform aggregations, the query must be executed through `writeStream` and an output sink.

Why the other options are incorrect:

A: `count()` is an action, not allowed directly on streaming DataFrames.

C: `countDistinct()` is a stateful aggregation, not supported outside of a proper streaming query.

D: `show()` is also an action, unsupported on streaming queries.

Reference:

PySpark Structured Streaming Programming Guide - supported transformations and actions.

Databricks Exam Guide (June 2025): Section "Structured Streaming" - performing operations on streaming DataFrames and understanding supported transformations.

NEW QUESTION # 69

An engineer has a large ORC file located at /file/test_data.orc and wants to read only specific columns to reduce memory usage. Which code fragment will select the columns, i.e., col1, col2, during the reading process?

- A. `spark.read.format("orc").load("/file/test_data.orc").select("col1", "col2")`
- B. `spark.read.orc("/file/test_data.orc").selected("col1", "col2")`
- C. `spark.read.orc("/file/test_data.orc").filter("col1 = 'value' ").select("col2")`
- D. `spark.read.format("orc").select("col1", "col2").load("/file/test_data.orc")`

Answer: A

Explanation:

The correct way to load specific columns from an ORC file is to first load the file using `.load()` and then apply `.select()` on the resulting DataFrame. This is valid with `.read.format("orc")` or the shortcut `.read.orc()`.

`df = spark.read.format("orc").load("/file/test_data.orc").select("col1", "col2")` Why others are incorrect:

A performs selection after filtering, but doesn't match the intention to minimize memory at load.

B incorrectly tries to use `.select()` before `.load()`, which is invalid.

C uses a non-existent `.selected()` method.

D correctly loads and then selects.

NEW QUESTION # 70

What is the behavior for function `date_sub(start, days)` if a negative value is passed into the days parameter?

- A. The number of days specified will be added to the start date
- B. The same start date will be returned
- C. The number of days specified will be removed from the start date
- D. An error message of an invalid parameter will be returned

Answer: A

Explanation:

The function `date_sub(start, days)` subtracts the number of days from the start date. If a negative number is passed, the behavior becomes a date addition.

Example:

`SELECT date_sub('2024-05-01', -5)`

-- Returns: 2024-05-06

So, a negative value effectively adds the absolute number of days to the date.

NEW QUESTION # 71

A developer wants to refactor some older Spark code to leverage built-in functions introduced in Spark 3.5.0. The existing code performs array manipulations manually. Which of the following code snippets utilizes new built-in functions in Spark 3.5.0 for array operations?

```
import pyspark.sql.functions as F
min_price = 110.50

result_df = prices_df \
    .filter(F.col("spot price") >= F.lit(min_price)) \
    .agg(F.count("*"))
```



```
result_df = prices_df \
    .agg(F.count("spot_price").alias("spot_price")) \
    .filter(F.col("spot_price") > F.lit("min_price"))
```

- A.

```
result_df = prices_df \
    .agg(F.count("spot_price").alias("spot_price")) \
    .filter(F.col("spot_price") > F.lit("min_price"))
```

- B.

```
result_df = prices_df \
    .agg(F.count_if(F.col("spot_price") >= F.lit(min_price)))
```

- C.

```
result_df = prices_df \
    .withColumn("valid_price", F.when(F.col("spot_price") > F.lit(min_price), 1).otherwise(0))
```

```
result_df = prices_df \
    .withColumn("valid_price", F.when(F.col("spot_price") > F.lit(min_price), 1).otherwise(0))
```

- D.

```
result_df = prices_df \
    .agg(F.min("spot_price"), F.max("spot_price"))
```

Answer: B

Explanation:

count_if(condition) counts the number of rows that meet the specified boolean condition.

In this example, it directly counts how many times spot_price >= min_price evaluates to true, replacing the older verbose combination of when/otherwise and filtering or summing.

Official Spark 3.5.0 documentation notes the addition of count_if to simplify this kind of logic:

"Added count_if aggregate function to count only the rows where a boolean condition holds (SPARK-43773)." Why other options are incorrect or outdated:

A uses a legacy-style method of adding a flag column (when().otherwise()), which is verbose compared to count_if.

C performs a simple min/max aggregation-useful but unrelated to conditional array operations or the updated functionality.

D incorrectly applies .filter() after .agg() which will cause an error, and misuses string "min_price" rather than the variable. Therefore, B is the only option leveraging new functionality from Spark 3.5.0 correctly and efficiently.

Explanation:

The correct answer is B because it uses the new function count_if, introduced in Spark 3.5.0, which simplifies conditional counting within aggregations.

NEW QUESTION # 72

An MLOps engineer is building a Pandas UDF that applies a language model that translates English strings into Spanish. The initial code is loading the model on every call to the UDF, which is hurting the performance of the data pipeline.

The initial code is:

```
def in_spanish_inner(df: pd.Series) -> pd.Series:
    model = get_translation_model(target_lang='es')
    return df.apply(model)
```

```
in_spanish = sf.pandas_udf(in_spanish_inner, StringType())
def in_spanish_inner(df: pd.Series) -> pd.Series:
    model = get_translation_model(target_lang='es')
    return df.apply(model)
in_spanish = sf.pandas_udf(in_spanish_inner, StringType())
```

How can the MLOps engineer change this code to reduce how many times the language model is loaded?

- A. Convert the Pandas UDF from a Series # Series UDF to an Iterator[Series] # Iterator[Series] UDF
- B. Run the `in Spanish_inner()` function in `amapInPandas()` function call
- C. Convert the Pandas UDF to a PySpark UDF
- D. Convert the Pandas UDF from a Series # Series UDF to a Series # Scalar UDF

Answer: A

Explanation:

Comprehensive and Detailed Explanation From Exact Extract:

The provided code defines a Pandas UDF of type Series-to-Series, where a new instance of the language model is created on each call, which happens per batch. This is inefficient and results in significant overhead due to repeated model initialization.

To reduce the frequency of model loading, the engineer should convert the UDF to an iterator-based Pandas UDF (`Iterator[pd.Series] -> Iterator[pd.Series]`). This allows the model to be loaded once per executor and reused across multiple batches, rather than once per call.

From the official Databricks documentation:

"Iterator of Series to Iterator of Series UDFs are useful when the UDF initialization is expensive... For example, loading a ML model once per executor rather than once per row/batch."

- Databricks Official Docs: Pandas UDFs

Correct implementation looks like:

```
python
CopyEdit
@pandas_udf("string")
def translate_udf(batch_iter: Iterator[pd.Series]) -> Iterator[pd.Series]:
    model = get_translation_model(target_lang='es')
    for batch in batch_iter:
        yield batch.apply(model)
```

This refactor ensures the `get_translation_model()` is invoked once per executor process, not per batch, significantly improving pipeline performance.

NEW QUESTION # 73

.....

The high quality and high efficiency of our Associate-Developer-Apache-Spark-3.5 exam materials has helped many people pass exams quickly. And we can proudly claim that if you study with our Associate-Developer-Apache-Spark-3.5 study questions for 20 to 30 hours, then you can confidently pass the exam for sure. After our worthy customers get a Associate-Developer-Apache-Spark-3.5 certificate, they now have more job opportunities. The current situation is very serious. Selecting Associate-Developer-Apache-Spark-3.5 training guide is your best decision.

Accurate Associate-Developer-Apache-Spark-3.5 Answers: <https://www.dumpstests.com/Associate-Developer-Apache-Spark-3.5-latest-test-dumps.html>

Therefore, for your convenience, more choices are provided for you, we are pleased to suggest you to choose our Accurate Associate-Developer-Apache-Spark-3.5 Answers - Databricks Certified Associate Developer for Apache Spark 3.5 - Python guide torrent for your exam, So you will have real Databricks Certified Associate Developer for Apache Spark 3.5 - Python (Associate-Developer-Apache-Spark-3.5) questions with accurate answers at your disposal in a Associate-Developer-Apache-Spark-3.5 dumps pdf document, Databricks Valid Associate-Developer-Apache-Spark-3.5 Exam Prep It has also been made sure that you get exposure to the exam format and practice the attempt before the appearing in the final exam.

This means looking at your network from all angles, finding Reliable Associate-Developer-Apache-Spark-3.5 Study Notes openings, and tracing out how an attacker could combine multiple exploits to achieve a malicious goal.

Why Hadoop for Data Munging, Therefore, for your convenience, Associate-Developer-Apache-Spark-3.5 more choices are provided for you, we are pleased to suggest you to choose our Databricks Certified Associate Developer for Apache Spark 3.5 - Python guide torrent for your exam.

Quiz 2026 Associate-Developer-Apache-Spark-3.5: Marvelous Valid Databricks Certified Associate Developer for Apache Spark 3.5 - Python Exam Prep

So you will have real Databricks Certified Associate Developer for Apache Spark 3.5 - Python (Associate-Developer-Apache-Spark-3.5) questions with accurate answers at your disposal in a Associate-Developer-Apache-Spark-3.5 dumps pdf document, It has also been made sure that you get exposure Accurate Associate-Developer-Apache-Spark-3.5 Answers to the exam format and practice the attempt before the appearing in the final exam

The after-sales service of our company completely Valid Associate-Developer-Apache-Spark-3.5 Torrent gives you a satisfying experience, which is unique in the world, Considerate after-sale services.

DOWNLOAD the newest DumpsTests Associate-Developer-Apache-Spark-3.5 PDF dumps from Cloud Storage for free:
<https://drive.google.com/open?id=1nZfWv9fb4aYQrsueQBWfpAKTEt1B-Y1>