

Valid Associate-Developer-Apache-Spark Exam Experience, New Associate-Developer-Apache-Spark Test Fee

Databricks Certification Details		
Databricks Certified Associate Developer for Apache Spark		
 Prior Certification Not Required	 Exam Validity 2 Years	 Exam Fee \$200 USD
 Exam Duration 120 Minutes	 No. of Questions 60 Questions	 Passing Marks 70%
 Recommended Experience Basic coding knowledge in Python or Scala		 Exam Format Multiple choice
 Languages English		

Our Associate-Developer-Apache-Spark Research materials design three different versions for all customers. These three different versions include PDF version, software version and online version, they can help customers solve any problems in use, meet all their needs. Although the three major versions of our Associate-Developer-Apache-Spark Learning Materials provide a demo of the same content for all customers, they will meet different unique requirements from a variety of users based on specific functionality.

Databricks Certified Associate Developer for Apache Spark 3.0 certification exam is a professional-level exam that is designed to test the skills and knowledge of software developers who work with Apache Spark. Databricks Certified Associate Developer for Apache Spark 3.0 Exam certification is offered by Databricks, which is a leading provider of cloud-based data analytics and machine learning platforms. Databricks Certified Associate Developer for Apache Spark 3.0 Exam certification is an excellent way for software developers to demonstrate their proficiency in using Apache Spark to build data-intensive applications and analytics systems.

Databricks Certified Associate Developer for Apache Spark 3.0 is a certification exam designed to validate the skills and knowledge of developers who work with Apache Spark. Associate-Developer-Apache-Spark Exam is offered by Databricks, a company that provides a cloud-based platform for data engineering, data science, and machine learning. Associate-Developer-Apache-Spark exam is intended for developers who have experience with Apache Spark and are seeking to demonstrate their proficiency in using the technology to build data-driven applications.

>> Valid Associate-Developer-Apache-Spark Exam Experience <<

2025 Valid Associate-Developer-Apache-Spark Exam Experience | Reliable Associate-Developer-Apache-Spark: Databricks Certified Associate Developer for Apache Spark 3.0 Exam 100% Pass

CertkingdomPDF have the latest Databricks certification Associate-Developer-Apache-Spark exam training materials. The industrious CertkingdomPDF's IT experts through their own expertise and experience continuously produce the latest Databricks Associate-Developer-Apache-Spark training materials to facilitate IT professionals to pass the Databricks Certification Associate-Developer-Apache-Spark Exam. The certification of Databricks Associate-Developer-Apache-Spark more and more valuable in the IT area and a lot people use the products of CertkingdomPDF to pass Databricks certification Associate-Developer-Apache-Spark exam. Through so many feedbacks of these products, our CertkingdomPDF products prove to be trusted.

For more info read the reference:

Databricks Associate Developer Apache Spark Exam

Databricks Certified Associate Developer for Apache Spark 3.0 Exam Sample Questions (Q10-Q15):

NEW QUESTION # 10

Which of the following code blocks returns approximately 1000 rows, some of them potentially being duplicates, from the 2000-row DataFrame `transactionsDf` that only has unique rows?

- A. `transactionsDf.sample(True, 0.5, force=True)`
- B. `transactionsDf.take(1000)`
- C. `transactionsDf.sample(False, 0.5)`
- D. `transactionsDf.take(1000).distinct()`
- E. `transactionsDf.sample(True, 0.5)`

Answer: E

Explanation:

Explanation

To solve this question, you need to know that `DataFrame.sample()` is not guaranteed to return the exact fraction of the number of rows specified as an argument. Furthermore, since duplicates may be returned, you should understand that the operator's `withReplacement` argument should be set to `True`. A `force=` argument for the operator does not exist.

While the `take` argument returns an exact number of rows, it will just take the first specified number of rows (1000 in this question) from the DataFrame. Since the DataFrame does not include duplicate rows, there is no potential of any of those returned rows being duplicates when using `take()`, so the correct answer cannot involve `take()`.

More info: `pyspark.sql.DataFrame.sample` - PySpark 3.1.2 documentation

Static notebook | Dynamic notebook: See test 2

NEW QUESTION # 11

Which of the following code blocks shows the structure of a DataFrame in a tree-like way, containing both column names and types?

- A. `itemsDf.printSchema()`
- B. `itemsDf.print.schema()`
- C. `spark.schema(itemsDf)`
- D. `1.print(itemsDf.columns)`
`2.print(itemsDf.types)`
- E. `itemsDf.rdd.printSchema()`

Answer: A

Explanation:

Explanation

`itemsDf.printSchema()`

Correct! Here is an example of what `itemsDf.printSchema()` shows, you can see the tree-like structure containing both column names and types:

root

-- itemId: integer (nullable = true)

-- attributes: array (nullable = true)

| |-- element: string (containsNull = true)

-- supplier: string (nullable = true)

`itemsDf.rdd.printSchema()`

No, the DataFrame's underlying RDD does not have a `printSchema()` method.

`spark.schema(itemsDf)`

Incorrect, there is no `spark.schema` command.

`print(itemsDf.columns)`

`print(itemsDf.dtypes)`

Wrong. While the output of this code blocks contains both column names and column types, the information is not arranged in a tree-like way.

`itemsDf.print.schema()`

No, DataFrame does not have a `print` method.

Static notebook | Dynamic notebook: See test 3

NEW QUESTION # 12

The code block shown below should return the number of columns in the CSV file stored at location filePath.

From the CSV file, only lines should be read that do not start with a # character. Choose the answer that correctly fills the blanks in the code block to accomplish this.

Code block:

```
__1__ ( __2__ . __3__ . csv(filePath, __4__ ). __5__ )
```

- A. 1. DataFrame
2. spark
3. read()
4. escape='#'
5. shape[0]
- B. 1. size
2. spark
3. read()
4. escape='#'
5. columns
- C. 1. len
2. spark
3. read
4. comment='#'
5. columns
- D. 1. len
2. pyspark
3. DataFrameReader
4. comment='#'
5. columns
- E. 1. size
2. pyspark
3. DataFrameReader
4. comment='#'
5. columns

Answer: C

Explanation:

Explanation

Correct code block:

```
len(spark.read.csv(filePath, comment='#').columns)
```

This is a challenging question with difficulties in an unusual context: The boundary between DataFrame and the DataFrameReader. It is unlikely that a question of this difficulty level appears in the exam. However, solving it helps you get more comfortable with the DataFrameReader, a subject you will likely have to deal with in the exam.

Before dealing with the inner parentheses, it is easier to figure out the outer parentheses, gaps 1 and 5. Given the code block, the object in gap 5 would have to be evaluated by the object in gap 1, returning the number of columns in the read-in CSV. One answer option includes DataFrame in gap 1 and shape[0] in gap 2. Since DataFrame cannot be used to evaluate shape[0], we can discard this answer option.

Other answer options include size in gap 1. size() is not a built-in Python command, so if we use it, it would have to come from somewhere else. pyspark.sql.functions includes a size() method, but this method only returns the length of an array or map stored within a column (documentation linked below).

So, using a size() method is not an option here. This leaves us with two potentially valid answers.

We have to pick between gaps 2 and 3 being spark.read or pyspark.DataFrameReader. Looking at the documentation (linked below), the DataFrameReader is actually a child class of pyspark.sql, which means that we cannot import it using pyspark.DataFrameReader. Moreover, spark.read makes sense because on Databricks, spark references current Spark session (pyspark.sql.SparkSession) and spark.read therefore returns a DataFrameReader (also see documentation below). Finally, there is only one correct answer option remaining.

More info:

- pyspark.sql.functions.size - PySpark 3.1.2 documentation
- pyspark.sql.DataFrameReader.csv - PySpark 3.1.2 documentation
- pyspark.sql.SparkSession.read - PySpark 3.1.2 documentation

Static notebook | Dynamic notebook: See test 3

NEW QUESTION # 13

Which of the following code blocks reads in the JSON file stored at filePath as a DataFrame?

- A. `spark.read().path(filePath)`
- B. `spark.read().json(filePath)`
- C. `spark.read.path(filePath)`
- D. `spark.read.json(filePath)`
- E. `spark.read.path(filePath, source="json")`

Answer: D

Explanation:

Explanation

`spark.read.json(filePath)`

Correct. `spark.read` accesses Spark's `DataFrameReader`. Then, Spark identifies the file type to be read as JSON type by passing `filePath` into the `DataFrameReader.json()` method.

`spark.read.path(filePath)`

Incorrect. Spark's `DataFrameReader` does not have a `path` method. A universal way to read in files is provided by the `DataFrameReader.load()` method (link below).

`spark.read.path(filePath, source="json")`

Wrong. A `DataFrameReader.path()` method does not exist (see above).

`spark.read().json(filePath)`

Incorrect. `spark.read` is a way to access Spark's `DataFrameReader`. However, the `DataFrameReader` is not callable, so calling it via `spark.read()` will fail.

`spark.read().path(filePath)`

No, Spark's `DataFrameReader` is not callable (see above).

More info: `pyspark.sql.DataFrameReader.json` - PySpark 3.1.2 documentation, `pyspark.sql.DataFrameReader.load` - PySpark 3.1.2 documentation Static notebook | Dynamic notebook: See test 3

NEW QUESTION # 14

Which of the following code blocks writes DataFrame itemsDf to disk at storage location filePath, making sure to substitute any existing data at that location?

- A. `itemsDf.write().parquet(filePath, mode="overwrite")`
- B. `itemsDf.write(filePath, mode="overwrite")`
- C. `itemsDf.write.option("parquet").mode("overwrite").path(filePath)`
- D. `itemsDf.write.mode("overwrite").path(filePath)`
- E. `itemsDf.write.mode("overwrite").parquet(filePath)`

Answer: E

Explanation:

Explanation

`itemsDf.write.mode("overwrite").parquet(filePath)`

Correct! `itemsDf.write` returns a `pyspark.sql.DataFrameWriter` instance whose overwriting behavior can be modified via the `mode` setting or by passing `mode="overwrite"` to the `parquet()` command.

Although the `parquet` format is not prescribed for solving this question, `parquet()` is a valid operator to initiate Spark to write the data to disk.

`itemsDf.write.mode("overwrite").path(filePath)`

No. A `pyspark.sql.DataFrameWriter` instance does not have a `path()` method.

`itemsDf.write.option("parquet").mode("overwrite").path(filePath)`

Incorrect, see above. In addition, a file format cannot be passed via the `option()` method.

`itemsDf.write(filePath, mode="overwrite")`

Wrong. Unfortunately, this is too simple. You need to obtain access to a `DataFrameWriter` for the `DataFrame` through calling `itemsDf.write` upon which you can apply further methods to control how Spark data should be written to disk. You cannot, however, pass arguments to `itemsDf.write` directly.

`itemsDf.write().parquet(filePath, mode="overwrite")`

False. See above.

More info: `pyspark.sql.DataFrameWriter.parquet` - PySpark 3.1.2 documentation Static notebook | Dynamic notebook: See test 3

• • • • •

[illegible]