



VCE & PDF

PassApply.com

<https://www.passapply.com/databricks-certified-associate-developer-for-apache-spark>
2024 Latest passapply DATABRICKS-CERTIFIED-ASSOCIATE-DEVELOPER-FOR-APACHE-SPARK PDF and VCE dumps Download

DATABRICKS-CERTIFIED-ASSOCIATE-DEVELOPER-FOR-APACHE-SPARK

Q&As

Databricks Certified Associate Developer for Apache Spark 3.0

Pass Databricks DATABRICKS-CERTIFIED-ASSOCIATE-DEVELOPER-FOR-APACHE-SPARK Exam with 100% Guarantee

Free Download Real Questions & Answers **PDF** and **VCE** file from:

<https://www.passapply.com/databricks-certified-associate-developer-for-apache-spark.html>

100% Passing Guarantee
100% Money Back Assurance

Following Questions and Answers are all new published by Databricks
Official Exam Center



VCE & PDF

PassApply.com

<https://www.passapply.com/databricks-certified-associate-developer-for-apache-spark>
2024 Latest passapply DATABRICKS-CERTIFIED-ASSOCIATE-DEVELOPER-FOR-APACHE-SPARK PDF and VCE dumps Download

- ⚙️ **Instant Download** After Purchase
- ⚙️ **100% Money Back** Guarantee
- ⚙️ **365 Days** Free Update
- ⚙️ **800,000+** Satisfied Customers





QUESTION 1

The code block shown below should show information about the data type that column storeId of DataFrame transactionsDf contains. Choose the answer that correctly fills the blanks in the code block to accomplish this.

Code block:

```
transactionsDf.__1__(__2__).__3__
```

A. 1. select

2.

"storeId"

3.

print_schema()

B. 1. limit

2.

1

3.

columns

C. 1. select

2.

"storeId"

3.

printSchema()

D. 1. limit

2.

"storeId"

3.

printSchema()

E. 1. select

2.

storeId



3.

dtypes

Correct Answer: B

Correct code block: `transactionsDf.select("storeId").printSchema()` The difficulty of this is that it is hard to solve with the stepwise first-to-last-gap approach that has worked well for similar questions, since the answer options are so different from one another. Instead, you might want to eliminate answers by looking for patterns of frequently wrong answers. A first pattern that you may recognize by now is that column names are not expressed in quotes. For this reason, the answer that includes `storeId` should be eliminated. By now, you may have understood that the `DataFrame.limit()` is useful for returning a specified amount of rows. It has nothing to do with specific columns. For this reason, the answer that resolves to

`limit("storeId")` can be eliminated. Given that we are interested in information about the data type, you should

QUESTION 2

Which of the following code blocks performs an inner join of DataFrames `transactionsDf` and `itemsDf` on columns `productId` and `itemId`, respectively, excluding columns `value` and `storeId` from DataFrame `transactionsDf` and column `attributes` from DataFrame `itemsDf`?

A. `transactionsDf.drop("\\value\\", "\\storeId\\").join(itemsDf.select("\\attributes\\"), transactionsDf.productId==itemsDf.itemId)`

B. `1.transactionsDf.createOrReplaceTempView("\\transactionsDf\\") 2.itemsDf.createOrReplaceTempView("\\itemsDf\\") 3.spark.sql("SELECT -value, -storeId FROM transactionsDf INNER JOIN itemsDf ON productId==itemId").drop("attributes")`

C. `transactionsDf.drop("value", "storeId").join(itemsDf.drop("attributes"), "transactionsDf.productId==itemsDf.itemId")`

D. `1.transactionsDf \`

2.

`.drop(col("\\value\\"), col("\\storeId\\")) \`

3.

`.join(itemsDf.drop(col("\\attributes\\")), col("\\productId\\")==col("\\itemId\\"))`

E. `1.transactionsDf.createOrReplaceTempView("\\transactionsDf\\") 2.itemsDf.createOrReplaceTempView("\\itemsDf\\") 3.statement = "" 4.SELECT * FROM transactionsDf 5.INNER JOIN itemsDf 6.ON transactionsDf.productId==itemsDf.itemId`

7. `""`

8. `spark.sql(statement).drop("value", "storeId", "attributes")`

Correct Answer: E



QUESTION 3

Which of the following code blocks returns about 150 randomly selected rows from the 1000-row DataFrame transactionsDf, assuming that any row can appear more than once in the returned DataFrame?

- A. transactionsDf.resample(0.15, False, 3142)
- B. transactionsDf.sample(0.15, False, 3142)
- C. transactionsDf.sample(0.15)
- D. transactionsDf.sample(0.85, 8429)
- E. transactionsDf.sample(True, 0.15, 8261)

Correct Answer: E

Answering this correctly depends on whether you understand the arguments to the DataFrame.sample() method (link to the documentation below). The arguments are as follows: DataFrame.sample(withReplacement=None, fraction=None, seed=None). The first argument withReplacement specified whether a row can be drawn from the DataFrame multiple times. By default, this option is disabled in Spark. But we have to enable it here, since the question asks for a row being able to appear more than once. So, we need to pass True for this argument.

About replacement: "Replacement" is easiest explained with the example of removing random items from a box. When you remove those "with replacement" it means that after you have taken an item out of the box, you put it back inside. So, essentially, if you would randomly take 10 items out of a box with 100 items, there is a chance you take the same item twice or more times. "Without replacement" means that you would not put the item back into the box after removing it. So, every time you remove an item from the box, there is one less item in the box and you can never take the same item twice. The second argument to the withReplacement method is fraction. This refers to the fraction of items that should be returned. In the we are asked for 150 out of 1000 items ?a fraction of 0.15. The last argument is a random seed. A random seed makes a randomized processed repeatable. This means that if you would re-run the same sample() operation with the same random seed, you would get the same rows returned from the sample() command. There is no behavior around the random seed specified in the question. The varying random seeds are only there to confuse you!

More info: `pyspark.sql.DataFrame.sample` -- PySpark 3.1.1 documentation Static notebook | Dynamic notebook: See test 1, 49 (Databricks import instructions)

QUESTION 4

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



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

Correct Answer: A

QUESTION 5

The code block shown below should return an exact copy of DataFrame `transactionsDf` that does not include rows in which values in column `storeId` have the value 25. Choose the answer that correctly fills the blanks in the code block to accomplish this.

A. `transactionsDf.remove(transactionsDf.storeId==25)`

B. `transactionsDf.where(transactionsDf.storeId!=25)`

C. `transactionsDf.filter(transactionsDf.storeId==25)`

D. `transactionsDf.drop(transactionsDf.storeId==25)`

E. `transactionsDf.select(transactionsDf.storeId!=25)`

Correct Answer: B

`transactionsDf.where(transactionsDf.storeId!=25)`

Correct. `DataFrame.where()` is an alias for the `DataFrame.filter()` method. Using this method, it is straightforward to filter out rows that do not have value 25 in column `storeId`.

`transactionsDf.select(transactionsDf.storeId!=25)`

Wrong. The select operator allows you to build DataFrames column-wise, but when using it as shown, it does not filter out rows.

`transactionsDf.filter(transactionsDf.storeId==25)`

Incorrect. Although the filter expression works for filtering rows, the `==` in the filtering condition is inappropriate. It should be `!=` instead.

`transactionsDf.drop(transactionsDf.storeId==25)`

No. `DataFrame.drop()` is used to remove specific columns, but not rows, from the DataFrame.

`transactionsDf.remove(transactionsDf.storeId==25)`

False. There is no `DataFrame.remove()` operator in PySpark. More info: `pyspark.sql.DataFrame.where` --

PySpark 3.1.2 documentation Static notebook | Dynamic notebook: See test 3, 48 (Databricks import instructions)



VCE & PDF

PassApply.com

<https://www.passapply.com/databricks-certified-associate-developer-for-apache-spark>

2024 Latest passapply DATABRICKS-CERTIFIED-ASSOCIATE-DEVELOPER-FOR-APACHE-SPARK PDF and VCE dumps Download

[DATABRICKS-CERTIFIED-ASSOCIATE-DEVELOPER-FOR-APACHE-SPARK PDF Dumps](#)

[DATABRICKS-CERTIFIED-ASSOCIATE-DEVELOPER-FOR-APACHE-SPARK Practice Test](#)

[DATABRICKS-CERTIFIED-ASSOCIATE-DEVELOPER-FOR-APACHE-SPARK Braindumps](#)