



DATABRICKS-CERTIFIED-ASSOCIAT

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.pass4itsure.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

Pass4itSure.com

<https://www.pass4itsure.com/databricks-certified-associate-developer-for-apache-spark>
2024 Latest pass4itsure 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 displayed below contains an error. The code block should return a new DataFrame that only contains rows from DataFrame transactionsDf in which the value in column predError is at least 5.

Find the error.

Code block:

```
transactionsDf.where("col(predError) >= 5")
```

- A. The argument to the where method should be "predError >= 5".
- B. Instead of where(), filter() should be used.
- C. The expression returns the original DataFrame transactionsDf and not a new DataFrame. To avoid this, the code block should be transactionsDf.toNewDataFrame().where("col(predError) >= 5").
- D. The argument to the where method cannot be a string.
- E. Instead of >=, the SQL operator GEQ should be used.

Correct Answer: A

QUESTION 2

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 3

Which of the following code blocks displays the 10 rows with the smallest values of column value in DataFrame `transactionsDf` in a nicely formatted way?

- A. `transactionsDf.sort(asc(value)).show(10)`
- B. `transactionsDf.sort(col("value")).show(10)`
- C. `transactionsDf.sort(col("value").desc()).head()`
- D. `transactionsDf.sort(col("value").asc()).print(10)`
- E. `transactionsDf.orderBy("value").asc().show(10)`

Correct Answer: B

QUESTION 4

The code block displayed below contains an error. The code block should save DataFrame `transactionsDf` at path `path` as a parquet file, appending to any existing parquet file. Find the error.

Code block:

- A. `transactionsDf.format("parquet").option("mode", "append").save(path)`
- B. The code block is missing a reference to the `DataFrameWriter`.
- C. `save()` is evaluated lazily and needs to be followed by an action.
- D. The mode option should be omitted so that the command uses the default mode.
- E. The code block is missing a `bucketBy` command that takes care of partitions.
- F. Given that the DataFrame should be saved as parquet file, `path` is being passed to the wrong method.

Correct Answer: B



Correct code block:

```
transactionsDf.write.format("parquet").option("mode", "append").save(path)
```

QUESTION 5

Which of the following statements about stages is correct?

- A. Different stages in a job may be executed in parallel.
- B. Stages consist of one or more jobs.
- C. Stages ephemerally store transactions, before they are committed through actions.
- D. Tasks in a stage may be executed by multiple machines at the same time.
- E. Stages may contain multiple actions, narrow, and wide transformations.

Correct Answer: D

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

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

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