



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.geekcert.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

GeekCert.com

<https://www.geekcert.com/databricks-certified-associate-developer-for-apache-spark-2024-latest-geekcert-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

Which of the following code blocks concatenates rows of DataFrames transactionsDf and transactionsNewDf, omitting any duplicates?

- A. transactionsDf.concat(transactionsNewDf).unique()
- B. transactionsDf.union(transactionsNewDf).distinct()
- C. spark.union(transactionsDf, transactionsNewDf).distinct()
- D. transactionsDf.join(transactionsNewDf, how="union").distinct()
- E. transactionsDf.union(transactionsNewDf).unique()

Correct Answer: B

DataFrame.unique() and DataFrame.concat() do not exist and union() is not a method of the SparkSession. In addition, there is no union option for the join method in the DataFrame.join() statement.

More info: [pyspark.sql.DataFrame.union -- PySpark 3.1.2 documentation](#)

Static notebook | Dynamic notebook: See test 2, 43 (Databricks import instructions)

QUESTION 2

The code block shown below should return a copy of DataFrame transactionsDf with an added column cos. This column should have the values in column value converted to degrees and having the cosine of those converted values taken, rounded to two decimals. Choose the answer that correctly fills the blanks in the code block to accomplish this.

Code block:

```
transactionsDf.__1__(__2__, round(__3__(__4__(__5__)),2))
```

A. 1. withColumn

2.

```
col("cos")
```

3.

```
cos
```

4.

```
degrees
```



5.

transactionsDf.value

B. 1. withColumnRenamed

2.

"cos"

3.

cos

4.

degrees

5.

"transactionsDf.value"

C. 1. withColumn

2.

"cos"

3.

cos

4.

degrees

5.

transactionsDf.value

D. 1. withColumn

2.

col("cos")

3.

cos

4.

degrees

5.

col("value")



E. 1. withColumn

2.

"cos"

3.

degrees

4.

cos

5.

col("value")

Correct Answer: C

QUESTION 3

The code block shown below should write DataFrame transactionsDf as a parquet file to path storeDir, using brotli compression and replacing any previously existing file. Choose the answer that correctly fills the blanks in the code block to accomplish this.

```
transactionsDf.__1__.format("parquet").__2__(__3__).option(__4__, "brotli").__5__(storeDir)
```

A. 1. save

2.

mode

3.

"ignore"

4.

"compression"

5.

path

B. 1. store

2.

with

3.



"replacement"

4.

"compression"

5.

path

C. 1. write

2.

mode

3.

"overwrite"

4.

"compression"

5.

save

(Correct)

D. 1. save

2.

mode

3.

"replace"

4.

"compression"

5.

path

E. 1. write

2.

mode

3.

"overwrite"



4.

compression

5.

parquet

Correct Answer: C

QUESTION 4

Which of the following code blocks creates a new DataFrame with two columns season and wind_speed_ms where column season is of data type string and column wind_speed_ms is of data type double?

- A. `spark.DataFrame({"season": ["winter","summer"], "wind_speed_ms": [4.5, 7.5]})`
- B. `spark.createDataFrame([(("summer", 4.5), ("winter", 7.5)), ("season", "wind_speed_ms")])`
- C. 1. `from pyspark.sql import types as T`
2. `spark.createDataFrame(((("summer", 4.5), ("winter", 7.5)),`
- D. `StructType([T.StructField("season", T.CharType()), T.StructField("season",`
- E. `DoubleType()))))`
- F. `spark.newDataFrame([(("summer", 4.5), ("winter", 7.5)), ("season", "wind_speed_ms")])`
- G. `spark.createDataFrame({"season": ["winter","summer"], "wind_speed_ms": [4.5, 7.5]})`

Correct Answer: B

QUESTION 5

Which of the following is a viable way to improve Spark's performance when dealing with large amounts of data, given that there is only a single application running on the cluster?

- A. Increase values for the properties `spark.default.parallelism` and `spark.sql.shuffle.partitions`
- B. Decrease values for the properties `spark.default.parallelism` and `spark.sql.partitions`
- C. Increase values for the properties `spark.sql.parallelism` and `spark.sql.partitions`
- D. Increase values for the properties `spark.sql.parallelism` and `spark.sql.shuffle.partitions`
- E. Increase values for the properties `spark.dynamicAllocation.maxExecutors`, `spark.default.parallelism`, and `spark.sql.shuffle.partitions`



Correct Answer: A

[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 Exam Questions](#)