



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 returns the number of unique values in column storeId of DataFrame transactionsDf?

- A. `transactionsDf.select("storeId").dropDuplicates().count()`
- B. `transactionsDf.select(count("storeId")).dropDuplicates()`
- C. `transactionsDf.select(distinct("storeId")).count()`
- D. `transactionsDf.dropDuplicates().agg(count("storeId"))`
- E. `transactionsDf.distinct().select("storeId").count()`

Correct Answer: A

`transactionsDf.select("storeId").dropDuplicates().count()` Correct! After dropping all duplicates from column storeId, the remaining rows get counted, representing the number of unique values in the column.
`transactionsDf.select(count("storeId")).dropDuplicates()` No. `transactionsDf.select(count("storeId"))` just returns a single-row DataFrame showing the number of non-null rows. `dropDuplicates()` does not have any effect in this context.

`transactionsDf.dropDuplicates().agg(count("storeId"))` Incorrect. While `transactionsDf.dropDuplicates()` removes duplicate rows from `transactionsDf`, it does not do so taking only column storeId into consideration, but eliminates full row duplicates instead.
`transactionsDf.distinct().select("storeId").count()` Wrong. `transactionsDf.distinct()` identifies unique rows across all columns, but not only unique rows with respect to column storeId. This may leave duplicate values in the column, making the count not represent the number of unique values in that column.
`transactionsDf.select(distinct("storeId")).count()` False. There is no `distinct` method in `pyspark.sql.functions`.

QUESTION 2

The code block displayed below contains an error. The code block is intended to return all columns of DataFrame transactionsDf except for columns predError, productId, and value.

Find the error.

Excerpt of DataFrame transactionsDf:

```
transactionsDf.select(~col("predError"), ~col("productId"), ~col("value"))
```

- A. The select operator should be replaced by the drop operator and the arguments to the drop operator should be column names predError, productId and value wrapped in the col operator so they should be expressed like `drop(col(predError), col(productId), col(value))`.
- B. The select operator should be replaced with the deselect operator.
- C. The column names in the select operator should not be strings and wrapped in the col operator, so they should be expressed like `select(~col(predError), ~col(productId), ~col(value))`.
- D. The select operator should be replaced by the drop operator.



E. The select operator should be replaced by the drop operator and the arguments to the drop operator should be column names predError, productId and value as strings.

Correct Answer: E

Correct code block:

```
transactionsDf.drop("predError", "productId", "value")
```

 Static notebook | Dynamic notebook: See test 1, 37

(Databricks import instructions)

QUESTION 3

Which of the following code blocks reads in the JSON file stored at filePath, enforcing the schema expressed in JSON format in variable json_schema, shown in the code block below?

Code block: 1.json_schema = "" 2.{"type": "struct",

3.

```
"fields": [
```

4.

```
{
```

5.

```
"name": "itemId",
```

6.

```
"type": "integer",
```

7.

```
"nullable": true,
```

8.

```
"metadata": {}
```

9.

```
},
```

10.

```
{
```

11.

```
"name": "supplier",
```



12.

```
"type": "string",
```

13.

```
"nullable": true,
```

14.

```
"metadata": {}
```

15.

```
}
```

16.

```
] 17.} 18. ""
```

A. `spark.read.json(filePath, schema=json_schema)`

B. `spark.read.schema(json_schema).json(filePath)` 1. `schema = StructType.fromJson(json.loads(json_schema))`
2. `spark.read.json(filePath, schema=schema)`

C. `spark.read.json(filePath, schema=schema_of_json(json_schema))`

D. `spark.read.json(filePath, schema=spark.read.json(json_schema))`

Correct Answer: C

QUESTION 4

Which of the following describes the difference between client and cluster execution modes?

A. In cluster mode, the driver runs on the worker nodes, while the client mode runs the driver on the client machine.

B. In cluster mode, the driver runs on the edge node, while the client mode runs the driver in a worker node.

C. In cluster mode, each node will launch its own executor, while in client mode, executors will exclusively run on the client machine.

D. In client mode, the cluster manager runs on the same host as the driver, while in cluster mode, the cluster manager runs on a separate node.

E. In cluster mode, the driver runs on the master node, while in client mode, the driver runs on a virtual machine in the cloud.

Correct Answer: A



QUESTION 5

Which of the following code blocks returns a copy of DataFrame transactionsDf where the column storeId has been converted to string type?

- A. `transactionsDf.withColumn("storeId", convert("storeId", "string"))`
- B. `transactionsDf.withColumn("storeId", col("storeId", "string"))`
- C. `transactionsDf.withColumn("storeId", col("storeId").convert("string"))`
- D. `transactionsDf.withColumn("storeId", col("storeId").cast("string"))`
- E. `transactionsDf.withColumn("storeId", convert("storeId").as("string"))`

Correct Answer: D

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

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

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