



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

Which of the following code blocks returns all unique values across all values in columns value and productId in DataFrame transactionsDf in a one-column DataFrame?

- A. `transactionsDf.select(\value\).join(transactionsDf.select(\productId\), col(\value\)==col(\productId\), \outer\)`
- B. `transactionsDf.select(col(\value\), col(\productId\)).agg({**: \count\})`
- C. `transactionsDf.select(\value\, \productId\).distinct()`
- D. `transactionsDf.select(\value\).union(transactionsDf.select(\productId\)).distinct()`
- E. `transactionsDf.agg({\value*: \collect_set\, \productId*: \collect_set\})`

Correct Answer: D

QUESTION 2

Which of the following code blocks reads the parquet file stored at filePath into DataFrame itemsDf, using a valid schema for the sample of itemsDf shown below?

Sample of itemsDf:

1.	+-----+-----+-----+
2.	itemId attributes supplier
3.	+-----+-----+-----+
4.	1 [blue, winter, cozy] Sports Company Inc.
5.	2 [red, summer, fresh, cooling] YetiX
6.	3 [green, summer, travel] Sports Company Inc.
7.	+-----+-----+-----+

- A. `1.itemsDfSchema = StructType([`
- 2.
- 3.
- 4.
- `StructField("supplier", StringType())])`



5.

```
6.itemsDf = spark.read.schema(itemsDfSchema).parquet(filePath)
```

```
B. 1.itemsDfSchema = StructType([
```

2.

```
  StructField("itemId", IntegerType),
```

3.

```
  StructField("attributes", ArrayType(StringType)),
```

4.

```
  StructField("supplier", StringType)])
```

5.

```
6.itemsDf = spark.read.schema(itemsDfSchema).parquet(filePath)
```

```
C. 1.itemsDf = spark.read.schema("\\itemId integer, attributes , supplier string\\").parquet(filePath)
```

```
D. 1.itemsDfSchema = StructType([
```

2.

```
  StructField("itemId", IntegerType()),
```

3.

```
  StructField("attributes", ArrayType(StringType())),
```

4.

```
  StructField("supplier", StringType())])
```

5.

```
6.itemsDf = spark.read.schema(itemsDfSchema).parquet(filePath)
```

```
E. 1.itemsDfSchema = StructType([
```

2.

```
  StructField("itemId", IntegerType()),
```

3.

```
  StructField("attributes", ArrayType([StringType()])),
```

4.

```
  StructField("supplier", StringType())])
```

5.



```
6.itemsDf = spark.read(schema=itemsDfSchema).parquet(filePath)
```

Correct Answer: D

The challenge in this comes from there being an array variable in the schema. In addition, you should know how to pass a schema to the DataFrameReader that is invoked by spark.read. The correct way to define an array of strings in a schema is through `ArrayType(StringType())`. A schema can be passed to the DataFrameReader by simply appending `schema(structType)` to the `read()` operator. Alternatively, you can also define a schema as a string. For example, for the schema of `itemsDf`, the following string would make sense: `itemId integer, attributes array, supplier string`. A thing to keep in mind is that in schema definitions, you always need to instantiate the types, like so: `StringType()`. Just using `StringType` does not work in pySpark and will fail. Another concern with schemas is whether columns should be nullable, so allowed to have null values. In the case at hand, this is not a concern however, since the just asks for a "valid" schema. Both non-nullable and nullable column schemas would be valid here, since no null value appears in the DataFrame sample. More info: Learning Spark, 2nd Edition, Chapter 3 Static notebook | Dynamic notebook: See test 3, 19 (Databricks import instructions)

QUESTION 3

Which of the following code blocks efficiently converts DataFrame `transactionsDf` from 12 into 24 partitions?

- A. `transactionsDf.repartition(24, boost=True)`
- B. `transactionsDf.repartition()`
- C. `transactionsDf.repartition("itemId", 24)`
- D. `transactionsDf.coalesce(24)`
- E. `transactionsDf.repartition(24)`

Correct Answer: E

QUESTION 4

The code block displayed below contains an error. The code block should return all rows of DataFrame `transactionsDf`, but including only columns `storeId` and `predError`. Find the error.

Code block:

```
spark.collect(transactionsDf.select("storeId", "predError"))
```

- A. Instead of `select`, DataFrame `transactionsDf` needs to be filtered using the filter operator.
- B. Columns `storeId` and `predError` need to be represented as a Python list, so they need to be wrapped in brackets (`[]`).
- C. The `take` method should be used instead of the `collect` method.
- D. Instead of `collect`, `collectAsRows` needs to be called.



E. The collect method is not a method of the SparkSession object.

Correct Answer: E

Correct code block: `transactionsDf.select("storeId", "predError").collect()` `collect()` is a method of the DataFrame object.
More info: `pyspark.sql.DataFrame.collect` -- PySpark 3.1.2 documentation Static notebook | Dynamic notebook: See test 2, 24 (Databricks import instructions)

QUESTION 5

Which of the following code blocks stores a part of the data in DataFrame itemsDf on executors?

- A. `itemsDf.cache().count()`
- B. `itemsDf.cache(eager=True)`
- C. `cache(itemsDf)`
- D. `itemsDf.cache().filter()`
- E. `itemsDf.rdd.storeCopy()`

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 Braindumps](#)