



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 store DataFrame transactionsDf on two different executors, utilizing the executors' memory as much as possible, but not writing anything to disk. Choose the answer that correctly fills the blanks in the code block to accomplish this.

1.from pyspark import StorageLevel 2.transactionsDf.__1__(StorageLevel.__2__).__3__

A. 1. cache

2.

MEMORY_ONLY_2

3.

count()

B. 1. persist

2.

DISK_ONLY_2

3.

count()

C. 1. persist

2.

MEMORY_ONLY_2

3.

select()

D. 1. cache

2.

DISK_ONLY_2

3.

count()

E. 1. persist

2.

MEMORY_ONLY_2

3.



count()

Correct Answer: E

QUESTION 2

The code block displayed below contains an error. The code block should read the csv file located at path data/transactions.csv into DataFrame transactionsDf, using the first row as column header and casting the columns in the most appropriate type. Find the error. First 3 rows of transactions.csv: 1.transactionId;storeId;productId;name 2.1;23;12;green grass 3.2;35;31;yellow sun 4.3;23;12;green grass Code block: transactionsDf = spark.read.load("data/transactions.csv", sep=";", format="csv", header=True)

- A. The DataFrameReader is not accessed correctly.
- B. The transaction is evaluated lazily, so no file will be read.
- C. Spark is unable to understand the file type.
- D. The code block is unable to capture all columns.
- E. The resulting DataFrame will not have the appropriate schema.

Correct Answer: E

Correct code block:

```
transactionsDf = spark.read.load("data/transactions.csv", sep=";", format="csv", header=True, inferSchema=True)
```

By default, Spark does not infer the schema of the CSV (since this usually takes some time). So, you need to add the inferSchema=True option to the code block.

More info: [pyspark.sql.DataFrameReader.csv -- PySpark 3.1.2 documentation](#)

QUESTION 3

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 4

The code block shown below should return a DataFrame with all columns of DataFrame `transactionsDf`, but only maximum 2 rows in which column `productId` has at least the value 2. Choose the answer that correctly fills the blanks in the code block to accomplish this.

`transactionsDf.__1__(__2__).__3__`

A. 1. `where`

2.

`"productId" > 2`

3.

`max(2)`

B. 1. `where`

2.

`transactionsDf[productId] >= 2`

3.

`limit(2)`

C. 1. `filter`

2.

`productId > 2`

3.

`max(2)`

D. 1. `filter`

2.

`col("productId") >= 2`

3.

`limit(2)`

E. 1. `where`



2.

```
productId >= 2
```

3.

```
limit(2)
```

Correct Answer: D

Correct code block:

`transactionsDf.filter(col("productId") >= 2).limit(2)` The filter and where operators in gap 1 are just aliases of one another, so you cannot use them to pick the right answer.

The column definition in gap 2 is more helpful. The `DataFrame.filter()` method takes an argument of type `Column` or `str`. From all possible answers, only the one including `col("productId") >= 2` fits this profile, since it returns a `Column` type. The answer option using `"productId" > 2` is invalid, since Spark does not understand that `"productId"` refers to column `productId`. The answer option using `transactionsDf[productId] >= 2` is wrong because you cannot refer to a column using square bracket notation in Spark (if you are coming from Python using Pandas, this is something to watch out for). In all other options, `productId` is being referred to as a Python variable, so they are relatively easy to eliminate. Also note that the asks for the value in column `productId` being at least 2. This translates to a "greater or equal" sign (`>= 2`), but not a "greater" sign (`> 2`). Another thing worth noting is that there is no `DataFrame.max()` method. If you picked any option including this, you may be confusing it with the `pyspark.sql.functions.max` method. The correct method to limit the amount of rows is the `DataFrame.limit()` method.

More info:

-`pyspark.sql.DataFrame.filter` -- PySpark 3.1.2 documentation

-`pyspark.sql.DataFrame.limit` -- PySpark 3.1.2 documentation

Static notebook | Dynamic notebook: See test 3, 54 (Databricks import instructions)

QUESTION 5

Which of the following code blocks returns all unique values of column `storeId` in `DataFrame transactionsDf`?

- A. `transactionsDf["storeId"].distinct()`
- B. `transactionsDf.select("storeId").distinct()`
- C. `transactionsDf.filter("storeId").distinct()`
- D. `transactionsDf.select(col("storeId").distinct())`
- E. `transactionsDf.distinct("storeId")`

Correct Answer: B



distinct() is a method of a DataFrame. Knowing this, or recognizing this from the documentation, is the key to solving this question. More info: `pyspark.sql.DataFrame.distinct` -- PySpark 3.1.2 documentation Static notebook | Dynamic notebook: See test 2, 19 (Databricks import instructions)

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

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

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