

# Databricks

*Developer-for-Apache-Spark-Python*  
*Certified Associate Developer for Apache Spark - Python*

**For More Information – Visit link below:**

**<https://www.examsempire.com/>**

**Product Version**

- 1. Up to Date products, reliable and verified.**
- 2. Questions and Answers in PDF Format.**



**<https://examsempire.com/>**

---

# Latest Version: 6.0

## Question: 1

The code blown down below intends to join df1 with df2 with inner join but it contains an error. Identify the error.

```
d1.join(d2, "inner", d1.col("id") === df2.col("id"))
```

Response:

- A. The join type is not in right order. The correct query should be `d2.join(d1, d1.col("id") === df2.col("id"), "inner")`
- B. There should be two `==` instead of `===`. So the correct query is `d1.join(d2, "inner", d1.col("id") == df2.col("id"))`
- C. Syntax is not correct `d1.join(d2, d1.col("id") == df2.col("id"), "inner")`
- D. We cannot do inner join in spark 3.0, but it is in the roadmap.

**Answer: C**

## Question: 2

If spark is running in client mode, which of the following statement about is correct?

Response:

- A. Spark driver is randomly attributed to a machine in the cluster
- B. Spark driver is attributed to the machine that has the most resources
- C. Spark driver remains on the client machine that submitted the application
- D. The entire spark application is run on a single machine.

**Answer: C**

## Question: 3

Which of the following code blocks adds a column `predErrorSqrt` to DataFrame `transactionsDf` that is the square root of column `predError`?

Response:

- A. `transactionsDf.withColumn("predErrorSqrt", sqrt(predError))`
- B. `transactionsDf.select(sqrt(predError))`
- C. `transactionsDf.withColumn("predErrorSqrt", col("predError").sqrt())`
- D. `transactionsDf.withColumn("predErrorSqrt", sqrt(col("predError")))`
- E. `transactionsDf.select(sqrt("predError"))`

---

**Answer: D**

### Question: 4

Which of the following are valid execution modes?

Response:

- A. Kubernetes, Local, Client
- B. Client, Cluster, Local
- C. Server, Standalone, Client
- D. Cluster, Server, Local
- E. Standalone, Client, Cluster

**Answer: B**

### Question: 5

Which of the following DataFrame methods is classified as a transformation?

Response:

- A. DataFrame.count()
- B. DataFrame.show()
- C. DataFrame.select()
- D. DataFrame.foreach()
- E. DataFrame.first()

**Answer: C**

### Question: 6

Which of the following 3 DataFrame operations are classified as an action?

Choose 3 answers:

Response:

- A. PrintSchema()
- B. Show()
- C. First()
- D. limit()
- E. foreach()
- F. cache

**Answer: B,C,E**

### Question: 7

If we want to create a constant integer 1 as a new column 'new\_column' in a dataframe df, which code block we should select?

Response:

- A. df.withColumnRenamed('new\_column', lit(1))
- B. df.withColumn(new\_column, lit(1))
- C. df.withColumn("new\_column", lit("1"))
- D. df.withColumn("new\_column", 1)
- E. df.withColumn("new\_column", lit(1))

**Answer: E**

### Question: 8

What command we can use to get the number of partition of a dataframe named df?

Response:

- A. df.rdd.getPartitionSize()
- B. df.getPartitionSize()
- C. df.getNumPartitions()
- D. df.rdd.getNumPartitions()

**Answer: D**

### Question: 9

Which of the following statements is NOT true for broadcast variables?

Response:

- A. It provides a mutable variable that a Spark cluster can safely update on a per-row basis.
- B. It is a way of updating a value inside of a variety of transformations and propagating that value to the driver node in an efficient and fault-tolerant way.
- C. You can define your own custom broadcast class by extending org.apache.spark.util.BroadcastV2 in Java or Scala or pyspark.AccumulatorParams in Python.
- D. Broadcast variables are shared, immutable variables that are cached on every machine in the cluster instead of serialized with every single task.
- E. The canonical use case is to pass around a small large table that does fit in memory on the executors.

---

**Answer: A,B,C**

**Question: 10**

The code block displayed below contains an error. The code block is intended to join DataFrame itemsDf with the larger DataFrame transactionsDf on column itemId. Find the error.

Code block: transactionsDf.join(itemsDf, "itemId", how="broadcast")

Response:

- A. The syntax is wrong, how= should be removed from the code block.
- B. The join method should be replaced by the broadcast method.
- C. Spark will only perform the broadcast operation if this behavior has been enabled on the Spark cluster.
- D. The larger DataFrame transactionsDf is being broadcasted, rather than the smaller DataFrame itemsDf
- E. broadcast is not a valid join type.

**Answer: E**

## Thank You for Trying Our Product

Special 16 USD Discount Coupon: **NSZUBG3X**

**Email:** [support@examsempire.com](mailto:support@examsempire.com)

**Check our Customer Testimonials and ratings  
available on every product page.**

**Visit our website.**

**<https://examsempire.com/>**