



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 DataFrame methods is classified as a transformation?

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

Correct Answer: C

`DataFrame.select()`

Correct, `DataFrame.select()` is a transformation. When the command is executed, it is evaluated lazily and returns an RDD when it is triggered by an action.

`DataFrame.foreach()`

Incorrect, `DataFrame.foreach()` is not a transformation, but an action. The intention of `foreach()` is to apply code to each element of a DataFrame to update accumulator variables or write the

elements to external storage. The process does not return an RDD - it is an action! `DataFrame.first()`

Wrong. As an action, `DataFrame.first()` executed immediately and returns the first row of a DataFrame.

`DataFrame.count()`

Incorrect. `DataFrame.count()` is an action and returns the number of rows in a DataFrame.

`DataFrame.show()`

No, `DataFrame.show()` is an action and displays the DataFrame upon execution of the command.

QUESTION 2

Which of the following describes Spark's way of managing memory?

- A. Spark uses a subset of the reserved system memory.
- B. Storage memory is used for caching partitions derived from DataFrames.
- C. As a general rule for garbage collection, Spark performs better on many small objects than few big objects.
- D. Disabling serialization potentially greatly reduces the memory footprint of a Spark application.



E. Spark's memory usage can be divided into three categories: Execution, transaction, and storage.

Correct Answer: B

Spark's memory usage can be divided into three categories: Execution, transaction, and storage.

No, it is either execution or storage.

As a general rule for garbage collection, Spark performs better on many small objects than few big objects.

No, Spark's garbage collection runs faster on fewer big objects than many small objects. Disabling serialization potentially greatly reduces the memory footprint of a Spark application.

The opposite is true ?serialization reduces the memory footprint, but may impact performance in a negative way.

Spark uses a subset of the reserved system memory. No, the reserved system memory is separate from Spark memory. Reserved memory stores Spark's internal objects.

More info: Tuning - Spark 3.1.2 Documentation, Spark Memory Management | Distributed Systems

Architecture, Learning Spark, 2nd Edition, Chapter 7

QUESTION 3

Which of the following statements about stages is correct?

- A. Different stages in a job may be executed in parallel.
- B. Stages consist of one or more jobs.
- C. Stages ephemerally store transactions, before they are committed through actions.
- D. Tasks in a stage may be executed by multiple machines at the same time.
- E. Stages may contain multiple actions, narrow, and wide transformations.

Correct Answer: D

QUESTION 4

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 5

Which of the following code blocks can be used to save DataFrame transactionsDf to memory only, recalculating



partitions that do not fit in memory when they are needed?

- A. from pyspark import StorageLevel transactionsDf.cache(StorageLevel.MEMORY_ONLY)
- B. transactionsDf.cache()
- C. transactionsDf.storage_level(\\MEMORY_ONLY\\)
- D. transactionsDf.persist()
- E. transactionsDf.clear_persist()
- F. from pyspark import StorageLevel transactionsDf.persist(StorageLevel.MEMORY_ONLY)

Correct Answer: F

[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 Braindumps](#)