DATABRICKS-CERTIFIED-ASSOCIAT

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

Latest DATABRICKS-CERTIFIED-ASSOCIATE-DEVELOPER-FOR-APACHE-SPARK Dumps | DATABRICKS-CERTIFIED-ASSOCIATE-DEVELOPER-FOR-APACHE-SPARK Exam Questions | DATABRICKS-CERTIFIED-ASSOCIATE-DEVELOPER-FOR-APACHE-SPARK Braindumps

1 / 6

QUESTION 1

Which of the following describes Spark actions?

A. Writing data to disk is the primary purpose of actions.

B. Actions are Spark\\'s way of exchanging data between executors.

C. The driver receives data upon request by actions.

D. Stage boundaries are commonly established by actions.

E. Actions are Spark\\'s way of modifying RDDs.

Correct Answer: C

The driver receives data upon request by actions.

Correct! Actions trigger the distributed execution of tasks on executors which, upon task completion,

transfer result data back to the driver. Actions are Spark\\'s way of exchanging data between executors. No.

In Spark, data is exchanged between executors via shuffles. Writing data to disk is the primary purpose of

actions. No. The primary purpose of actions is to access data that is stored in Spark\\'s RDDs and return the

data, often in aggregated form, back to the driver.

Actions are Spark\\'s way of modifying RDDs.

Incorrect. Firstly, RDDs are immutable ?they cannot be modified. Secondly, Spark generates new RDDs

via transformations and not actions. Stage boundaries are commonly established by actions. Wrong. A

stage boundary is commonly established by a shuffle, for example caused by a wide transformation.

QUESTION 2

Which of the following code blocks returns a single-row DataFrame that only has a column corr which shows the
Pearson correlation coefficient between columns predError and value in DataFrame transactionsDf?

A. transactionsDf.select(corr(["predError", "value"]).alias("corr")).first()

B. transactionsDf.select(corr(col("predError"), col("value")).alias("corr")).first()

C. transactionsDf.select(corr(predError, value).alias("corr"))

D. transactionsDf.select(corr(col("predError"), col("value")).alias("corr"))

E. transactionsDf.select(corr("predError", "value"))

Latest DATABRICKS-CERTIFIED-ASSOCIATE-DEVELOPER-FOR-APACHE-SPARK Dumps | DATABRICKS-
CERTIFIED-ASSOCIATE-DEVELOPER-FOR-APACHE-SPARK Exam Questions | DATABRICKS-CERTIFIED-
ASSOCIATE-DEVELOPER-FOR-APACHE-SPARK Braindumps

3 / 6

Correct Answer: D

In difficulty, this is above what you can expect from the exam. What this wants to teach you, however, is

to pay attention to the useful details included in the

documentation.

pyspark.sql.corr is not a very common method, but it deals with Spark\'s data structure in an interesting

way. The command takes two columns over multiple rows and returns a single row - similar to

an aggregation function. When examining the documentation (linked below), you will find this code

example:

a = range(20)

b = [2 * x for x in range(20)]

df = spark.createDataFrame(zip(a, b), ["a", "b"])

df.agg(corr("a", "b").alias(\\'c\\')).collect()

[Row(c=1.0)]

See how corr just returns a single row? Once you understand this, you should be suspicious about

answers that include first(), since there is no need to just select a single row. A reason to eliminate

those answers is that DataFrame.first() returns an object of type Row, but not DataFrame, as requested in

the question.

transactionsDf.select(corr(col("predError"), col("value")).alias("corr")) Correct! After calculating the Pearson

correlation coefficient, the resulting column is correctly renamed to corr.

transactionsDf.select(corr(predError, value).alias("corr")) No. In this answer, Python will interpret column

names predError and value as variable names.

transactionsDf.select(corr(col("predError"), col("value")).alias("corr")).first() Incorrect. first() returns a row,

not a DataFrame (see above and linked documentation below).

transactionsDf.select(corr("predError", "value"))

Wrong. Whie this statement returns a DataFrame in the desired shape, the column will have the name corr

(predError, value) and not corr.

transactionsDf.select(corr(["predError", "value"]).alias("corr")).first() False. In addition to first() returning a

row, this code block also uses the wrong call structure for command corr which takes two arguments (the

two columns to correlate).

More info:

-pyspark.sql.functions.corr -- PySpark 3.1.2 documentation

-pyspark.sql.DataFrame.first -- PySpark 3.1.2 documentation

Static notebook | Dynamic notebook: See test 3, 53 (Databricks import instructions)

QUESTION 3

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 4

Which of the following statements about executors is correct?

A. Executors are launched by the driver.

B. Executors stop upon application completion by default.

C. Each node hosts a single executor.

D. Executors store data in memory only.

E. An executor can serve multiple applications.

Correct Answer: B

QUESTION 5

The code block shown below should return a DataFrame with columns transactionsId, predError, value, and f from DataFrame transactionsDf. Choose the answer that correctly fills the blanks in the code block to accomplish this.

transactionsDf.__1__(__2__)

A. 1. filter

2. "transactionId", "predError", "value", "f"

B. 1. select

2. "transactionId, predError, value, f"

C. 1. select

2. ["transactionId", "predError", "value", "f"]

D. 1. where

2. col("transactionId"), col("predError"), col("value"), col("f") E. 1. select

2. col(["transactionId", "predError", "value", "f"])

Correct Answer: C