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.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

Instant Download After Purchase
100% Money Back Guarantee
365 Days Free Update
800,000+ Satisfied Customers

QUESTION 1

Which of the following code blocks generally causes a great amount of network traffic?

A. DataFrame.select()

B. DataFrame.coalesce()

C. DataFrame.collect()

D. DataFrame.rdd.map()

E. DataFrame.count()

Correct Answer: C

QUESTION 2

The code block displayed below contains an error. The code block should display the schema of DataFrame transactionsDf. Find the error.

Code block:

transactionsDf.rdd.printSchema

A. There is no way to print a schema directly in Spark, since the schema can be printed easily through using print(transactionsDf.columns), so that should be used instead.

B. The code block should be wrapped into a print() operation.

C. PrintSchema is only accessible through the spark session, so the code block should be rewritten as spark.printSchema(transactionsDf).

D. PrintSchema is a method and should be written as printSchema(). It is also not callable through transactionsDf.rdd, but should be called directly from transactionsDf.

E. PrintSchema is a not a method of transactionsDf.rdd. Instead, the schema should be printed via transactionsDf.print_schema().

Correct Answer: D

Correct code block: transactionsDf.printSchema() This is more of a knowledge that you should just memorize or look up in the provided documentation during the exam. You can get more info about DataFrame.printSchema() in the documentation (link below). However - it is a plain simple method without any arguments. One answer points to an alternative of printing the schema: You could also use print(transactionsDf.schema). This will give you readable, but not nicely formatted, description of the schema. More info: pyspark.sql.DataFrame.printSchema -- PySpark 3.1.1 documentation Static notebook | Dynamic notebook: See test 1, 54 (Databricks import instructions)

QUESTION 3

Which of the following is one of the big performance advantages that Spark has over Hadoop?

A. Spark achieves great performance by storing data in the DAG format, whereas Hadoop can only use parquet files.

B. Spark achieves higher resiliency for queries since, different from Hadoop, it can be deployed on Kubernetes.

C. Spark achieves great performance by storing data and performing computation in memory, whereas large jobs in Hadoop require a large amount of relatively slow disk I/O operations.

D. Spark achieves great performance by storing data in the HDFS format, whereas Hadoop can only use parquet files.

E. Spark achieves performance gains for developers by extending Hadoop\\'s DataFrames with a user-friendly API.

Correct Answer: C

QUESTION 4

The code block shown below should set the number of partitions that Spark uses when shuffling data for joins or aggregations to 100. Choose the answer that correctly fills the blanks in the code block to accomplish this.

spark.sql.shuffle.partitions

__1__.__2__.__3__(__4__, 100)

A. 1. spark

2.

conf

3.

set

4.

"spark.sql.shuffle.partitions"

B. 1. pyspark

2.

config

3.

set

4.

spark.shuffle.partitions

C. 1. spark

2.

conf

3.

get

4.

"spark.sql.shuffle.partitions"

D. 1. pyspark

2.

config

3.

set

4.

"spark.sql.shuffle.partitions"

E. 1. spark

2.

conf

3.

set

4.

"spark.sql.aggregate.partitions"

Correct Answer: A

Correct code block:

spark.conf.set("spark.sql.shuffle.partitions", 100) The conf interface is part of the SparkSession, so you need to call it through spark and not pyspark. To configure spark, you need to use the set method, not the get method. get reads a property, but does not write it. The correct property to achieve what is outlined in the

QUESTION 5

The code block displayed below contains an error. The code block should save DataFrame transactionsDf at path path

as a parquet file, appending to any existing parquet file. Find the error.

Code block:

A. transactionsDf.format("parquet").option("mode", "append").save(path)

B. The code block is missing a reference to the DataFrameWriter.

C. save() is evaluated lazily and needs to be followed by an action.

D. The mode option should be omitted so that the command uses the default mode.

E. The code block is missing a bucketBy command that takes care of partitions.

F. Given that the DataFrame should be saved as parquet file, path is being passed to the wrong method.

Correct Answer: B

Correct code block:

transactionsDf.write.format("parquet").option("mode", "append").save(path)

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