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

QUESTION 1

Which of the following describes the characteristics of accumulators?

A. Accumulators are used to pass around lookup tables across the cluster.

B. All accumulators used in a Spark application are listed in the Spark UI.

C. Accumulators can be instantiated directly via the accumulator(n) method of the pyspark.RDD module.

D. Accumulators are immutable.

E. If an action including an accumulator fails during execution and Spark manages to restart the action and complete it successfully, only the successful attempt will be counted in the accumulator.

Correct Answer: E

QUESTION 2

Which of the following describes a narrow transformation?

A. narrow transformation is an operation in which data is exchanged across partitions.

B. A narrow transformation is a process in which data from multiple RDDs is used.

C. A narrow transformation is a process in which 32-bit float variables are cast to smaller float variables, like 16-bit or 8-bit float variables.

D. A narrow transformation is an operation in which data is exchanged across the cluster.

E. A narrow transformation is an operation in which no data is exchanged across the cluster.

Correct Answer: E

A narrow transformation is an operation in which no data is exchanged across the cluster. Correct! In narrow transformations, no data is exchanged across the cluster, since these transformations do not require any data from outside of the partition they are applied on. Typical narrow transformations include filter, drop, and coalesce. A narrow transformation is an operation in which data is exchanged across partitions. No, that would be one definition of a wide transformation, but not of a narrow transformation. Wide transformations typically cause a shuffle, in which data is exchanged across partitions, executors, and the cluster. A narrow transformation is an operation in which data is exchanged across the cluster. No, see explanation just above this one. A narrow transformation is a process in which 32-bit float variables are cast to smaller float variables, like 16-bit or 8-bit float variables. No, type conversion has nothing to do with narrow transformations in Spark. A narrow transformation is a process in which data from multiple RDDs is used. No. A resilient distributed dataset (RDD) can be described as a collection of partitions. In a narrow transformation, no data is exchanged between partitions. Thus, no data is exchanged between RDDs. One could say though that a narrow transformation and, in fact, any transformation results in a new RDD being created. This is because a transformation results in a change to an existing RDD (RDDs are the foundation of other Spark data structures, like DataFrames). But, since RDDs are immutable, a new RDD needs to be created to reflect the change caused by the transformation. More info: Spark Transformation and Action: A Deep Dive | by Misbah Uddin | CodeX |

Medium

QUESTION 3

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")

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.

Correct Answer: E

QUESTION 4

Which of the following describes the difference between client and cluster execution modes?

A. In cluster mode, the driver runs on the worker nodes, while the client mode runs the driver on the client machine.

B. In cluster mode, the driver runs on the edge node, while the client mode runs the driver in a worker node.

C. In cluster mode, each node will launch its own executor, while in client mode, executors will exclusively run on the client machine.

D. In client mode, the cluster manager runs on the same host as the driver, while in cluster mode, the cluster manager runs on a separate node.

E. In cluster mode, the driver runs on the master node, while in client mode, the driver runs on a virtual machine in the cloud.

Correct Answer: A

QUESTION 5

The code block displayed below contains an error. The code block should write DataFrame transactionsDf

as a parquet file to location filePath after partitioning it on column storeId.

Find the error.

Code block:

transactionsDf.write.partitionOn("storeId").parquet(filePath)

A. The partitioning column as well as the file path should be passed to the write() method of DataFrame transactionsDf directly and not as appended commands as in the code block.

B. The partitionOn method should be called before the write method.

C. The operator should use the mode() option to configure the DataFrameWriter so that it replaces any existing files at location filePath.

D. Column storeId should be wrapped in a col() operator.

E. No method partitionOn() exists for the DataFrame class, partitionBy() should be used instead.

Correct Answer: E