

Databricks

Developer-for-Apache-Spark-Scala
Certified Associate Developer for Apache Spark - Scala

Questions And Answers PDF Format:

For More Information – Visit link below:
<https://www.certsgrade.com/>

Version = Product



Latest Version: 6.0

Question: 1

How to make sure that dataframe df has 12 partitions given that df has 4 partitions?

Response:

- A. df.setPartition(12)
- B. df.repartition(12)
- C. df.repartition()
- D. df.setPartition()

Answer: B

Question: 2

Your application on production is crashing lately and your application gets stuck at the same level every time you restart the spark job . You know that it is the toLocalIterator function is causing the problem.

What are the possible solutions to this problem?

Response:

- A. There is nothing to worry, application crashes are expected and will not affect your application at all.
- B. Use collect function instead of to localIterator
- C. Reduce the memory of the driver
- D. Reduce the size of your partitions if possible.

Answer: D

Question: 3

Spark dynamically handles skew in sort-merge join by splitting (and replicating if needed) skewed partitions. Which property needs to be enabled to achieve this?

Response:

- A. spark.sql.adaptive.skewJoin.enabled
- B. spark.sql.skewJoin.enabled
- C. spark.sql.adaptive.skewJoin.enable
- D. spark.sql.adaptive.optimeze.skewJoin

Answer: A

Question: 4

If spark is running in cluster mode, which of the following statements about nodes is incorrect?

Response:

- A. Each executor is running in a JVM inside of a worker node
- B. There might be more executors than total number of nodes
- C. There is at least one worker node in the cluster
- D. There is one single worker node that contains the Spark driver and the executors
- E. The spark driver runs in its own non-worker node without any executors

Answer: E

Question: 5

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

Response:

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

Answer: B

Question: 6

Which of the following three operations are classified as a wide transformation?

Choose 3 answers:

Response:

- A. drop()
- B. repartition()
- C. filter()
- D. flatMap()
- E. orderBy()
- F. selectDistinct()

Answer: B,E,F

Question: 7

Which of the following statements about the Spark driver is incorrect?

Response:

- A. The Spark driver is the node in which the Spark application's main method runs to coordinate the Spark application.
- B. The Spark driver is horizontally scaled to increase overall processing throughput.
- C. The Spark driver contains the SparkContext object.
- D. The Spark driver is responsible for scheduling the execution of data by various worker nodes in cluster mode.
- E. The Spark driver should be as close as possible to worker nodes for optimal performance.

Answer: B

Question: 8

Which of the followings is NOT a useful use cases of spark?

Response:

- A. Performing ad hoc or interactive queries to explore and visualize data sets
- B. Building, training, and evaluating machine learning models using MLlib
- C. Analyzing graph data sets and social networks
- D. Processing in parallel small data sets distributed across a cluster

Answer: D

Question: 9

At which stage do the first set of optimizations take place?

Response:

- A. Physical Planning
- B. Code Generation
- C. Logical Optimization
- D. Analysis

Answer: C

Question: 10

At which stage Catalyst optimizer generates one or more physical plans?

Response:

- A. Code Generation
- B. Logical Optimization
- C. Physical Planning
- D. Analysis

Answer: C

For More Information – **Visit link below:**
<https://www.certsgrade.com/>

PRODUCT FEATURES

-  **100% Money Back Guarantee**
-  **90 Days Free updates**
-  **Special Discounts on Bulk Orders**
-  **Guaranteed Success**
-  **50,000 Satisfied Customers**
-  **100% Secure Shopping**
-  **Privacy Policy**
-  **Refund Policy**

16 USD Discount Coupon Code: **NB4XKTMZ**

