CCA175 Q&As

CCA Spark and Hadoop Developer Exam

Pass Cloudera CCA175 Exam with 100% Guarantee

Free Download Real Questions & Answers **PDF** and **VCE** file from:

https://www.geekcert.com/cca175.html

100% Passing Guarantee
100% Money Back Assurance

Following Questions and Answers are all new published by Cloudera Official Exam Center

QUESTION 1

Problem Scenario 55 : You have been given below code snippet.

val pairRDDI = sc.parallelize(List(("cat",2), ("cat", 5), ("book", 4),("cat", 12))) val

pairRDD2 = sc.parallelize(List(("cat",2), ("cup", 5), ("mouse", 4),("cat", 12)))

operation1

Write a correct code snippet for operationl which will produce desired output, shown below.

Array[(String, (Option[lnt], Option[lnt]))] = Array((book,(Some(4},None)),

(mouse,(None,Some(4))), (cup,(None,Some(5))), (cat,(Some(2),Some(2)),

(cat,(Some(2),Some(12))), (cat,(Some(5),Some(2))), (cat,(Some(5),Some(12))),

(cat,(Some(12),Some(2))), (cat,(Some(12),Some(12)))J

Correct Answer: See the explanation for Step by Step Solution and configuration.

Solution : pairRDD1.fullOuterJoin(pairRDD2).collect

fullOuterJoin [Pair]

Performs the full outer join between two paired RDDs.

Listing Variants

def fullOuterJoin[W](other: RDD[(K, W)], numPartitions: Int): RDD[(K, (Option[V],

OptionfW]))]

def fullOuterJoin[W](other: RDD[(K, W}]}: RDD[(K, (Option[V], OptionfW]))]

def fullOuterJoin[W](other: RDD[(K, W)], partitioner: Partitioner): RDD[(K, (Option[V],

Option[W]))]

QUESTION 2

Problem Scenario 72 : You have been given a table named "employee2" with following detail. first_name string
last_name string Write a spark script in python which read this table and print all the rows and individual column values.

Correct Answer: See the explanation for Step by Step Solution and configuration.

Solution :

Step 1 : Import statements for HiveContext from pyspark.sql import HiveContext

Step 2 : Create sqlContext sqlContext = HiveContext(sc)

Step 3 : Query hive

employee2 = sqlContext.sql("select\\' from employee2")

Step 4 : Now prints the data for row in employee2.collect(): print(row)

Step 5 : Print specific column for row in employee2.collect(): print(row.fi rst_name)

QUESTION 3

Problem Scenario 9 : You have been given following mysql database details as well as other info. user=retail_dba password=cloudera database=retail_db jdbc URL = jdbc:mysql://quickstart:3306/retail_db Please accomplish following.

1.

Import departments table in a directory.

2.

Again import departments table same directory (However, directory already exist hence it should not overrride and append the results)

3.

Also make sure your results fields are terminated by \\'|\\' and lines terminated by \\'\n\

Correct Answer: See the explanation for Step by Step Solution and configuration.

Solutions :

Step 1 : Clean the hdfs file system, if they exists clean out.

hadoop fs -rm -R departments

hadoop fs -rm -R categories

hadoop fs -rm -R products

hadoop fs -rm -R orders

hadoop fs -rm -R order_items

hadoop fs -rm -R customers

Step 2 : Now import the department table as per requirement.

sqoop import \

-connect jdbc:mysql://quickstart:330G/retaiI_db \

--username=retail_dba \

-password=cloudera \

-table departments \

-target-dir=departments \

-fields-terminated-by \\'|\\' \

-lines-terminated-by \\'\n\\' \

-ml

Step 3 : Check imported data.

hdfs dfs -Is departments

hdfs dfs -cat departments/part-m-00000

Step 4 : Now again import data and needs to appended.

sqoop import \

-connect jdbc:mysql://quickstart:3306/retail_db \

--username=retail_dba \

-password=cloudera \

-table departments \

-target-dir departments \

-append \

-tields-terminated-by \\'|\\' \

-lines-termtnated-by \\'\n\\' \

-ml

Step 5 : Again Check the results

hdfs dfs -Is departments

hdfs dfs -cat departments/part-m-00001

QUESTION 4

Problem Scenario 59 : You have been given below code snippet.

val x = sc.parallelize(1 to 20)

val y = sc.parallelize(10 to 30) operationl

z.collect

Write a correct code snippet for operationl which will produce desired output, shown below.

Array[Int] = Array(16,12, 20,13,17,14,18,10,19,15,11)

Correct Answer: See the explanation for Step by Step Solution and configuration.

Solution :

val z = x.intersection(y)

intersection : Returns the elements in the two RDDs which are the same.

QUESTION 5

Problem Scenario 96 : Your spark application required extra Java options as below. XX:+PrintGCDetails-XX:+PrintGCTimeStamps Please replace the XXX values correctly ./bin/spark-submit --name "My app" --master local[4] --conf spark.eventLog.enabled=talse -conf XXX hadoopexam.jar

Correct Answer: See the explanation for Step by Step Solution and configuration.

Solution

XXX: Mspark.executoi\extraJavaOptions=-XX:+PrintGCDetails -XX:+PrintGCTimeStamps"

Notes: ./bin/spark-submit \

--class

--master \

--deploy-mode \

-conf = \

other options

\

[application-arguments]

Here, conf is used to pass the Spark related contigs which are required for the application

to run like any specific property(executor memory) or if you want to override the default

property which is set in Spark-default.conf.