



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.pass4itsure.com/cca175.html>

100% Passing Guarantee
100% Money Back Assurance

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

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



**QUESTION 1**

Problem Scenario 70 : Write down a Spark Application using Python, In which it read a file "Content.txt" (On hdfs) with following content. Do the word count and save the results in a directory called "problem85" (On hdfs)

Content.txt

Hello this is ABCTECH.com

This is XYZTECH.com

Apache Spark Training

This is Spark Learning Session Spark is faster than MapReduce

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

Solution :

Step 1 : Create an application with following code and store it in problem84.py

```
# Import SparkContext and SparkConf
from pyspark import SparkContext, SparkConf

# Create configuration object and set App name
conf = SparkConf().setAppName("CCA 175 Problem 85") sc = sparkContext(conf=conf)

#load data from hdfs
contentRDD = sc.textFile(MContent.txt")

#filter out non-empty lines
nonemptyjines = contentRDD.filter(lambda x: len(x) > 0)

#Split line based on space
words = nonempty_lines.ffatMap(lambda x: x.split("\\\\"))

#Do the word count
wordcounts = words.map(lambda x: (x, 1)) \
reduceByKey(lambda x, y: x+y) \
map(lambda x: (x[1], x[0])).sortByKey(False)

for word in wordcounts.collect(): print(word)

#Save final data " wordcounts.saveAsTextFile("problem85")
```



step 2 : Submit this application

```
spark-submit -master yarn problem85.py
```

QUESTION 2

Problem Scenario 16 : 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 below assignment.

1.

Create a table in hive as below.

```
create table departments_hive(department_id int, department_name string);
```

2.

Now import data from mysql table departments to this hive table. Please make sure that

data should be visible using below hive command, select" from departments_hive

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

Solution :

Step 1 : Create hive table as said.

hive

```
show tables;
```

```
create table departments_hive(department_id int, department_name string);
```

Step 2 : The important here is, when we create a table without delimiter fields. Then default

delimiter for hive is ^A (\001). Hence, while importing data we have to provide proper

delimiter.

```
sqoop import \
```

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

```
~username=retail_dba \
```

```
-password=cloudera \
```

```
--table departments \
```

```
--hive-home /user/hive/warehouse \
```

```
-hive-import \
```

```
-hive-overwrite \
```



```
--hive-table departments_hive \
```

```
--fields-terminated-by '\\001\\'
```

Step 3 : Check-the data in directory.

```
hdfs dfs -ls /user/hive/warehouse/departments_hive
```

```
hdfs dfs -cat/user/hive/warehouse/departmentshive/part\\
```

Check data in hive table.

```
Select * from departments_hive;
```

QUESTION 3

Problem Scenario 8 : You have been given following mysql database details as well as

other info.

Please accomplish following.

1.
Import joined result of orders and order_items table join on orders.order_id = order_items.order_item_order_id.

2.
Also make sure each tables file is partitioned in 2 files e.g. part-00000, part-00002

3.
Also make sure you use orderid columns for sqoop to use for boundary conditions.

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:3306/retail_db \  
-username=retail_dba \  
-password=cloudera \  
-query="select\` from orders join order_items on orders.orderid =  
order_items.order_item_order_id where \SCONDITIONS" \  
-target-dir /user/cloudera/order_join \  
-split-by order_id \  
--num-mappers 2
```

Step 3 : Check imported data.

```
hdfs dfs -ls order_join
```

```
hdfs dfs -cat order_join/part-m-00000
```

```
hdfs dfs -cat order_join/part-m-00001
```

QUESTION 4

Problem Scenario 38 : You have been given an RDD as below,

```
val rdd: RDD[Array[Byte]]
```

Now you have to save this RDD as a SequenceFile. And below is the code snippet.

```
import org.apache.hadoop.io.compress.GzipCodec
```

```
rdd.map(byteArray => (A.get(), new B(byteArray))).saveAsSequenceFile(\\7output/path",classOf[GzipCodec])
```

 What would be the correct replacement for A and B in above snippet.

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

Solution :

A. NullWritable

B. BytesWritable

QUESTION 5

Problem Scenario 94 : You have to run your Spark application on yarn with each executor

20GB and number of executors should be 50. Please replace XXX, YYY, ZZZ

```
export HADOOP_CONF_DIR=XXX
```



```
./bin/spark-submit \  
-class com.hadoopexam.MyTask \  
xxx\  
-deploy-mode cluster \ # can be client for client mode  
YYY\  
222 \  
/path/to/hadoopexam.jar \  
1000
```

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

Solution

XXX: -master yarn YYY : -executor-memory 20G ZZZ: -num-executors 50

[Latest CCA175 Dumps](#)

[CCA175 Study Guide](#)

[CCA175 Brindumps](#)