sqoop import --connect jdbc:mysql://localhost:3306/retail\_db --username root --password cloudera --table departments

specify the target-dir and partitions:

sqoop import --connect jdbc:mysql://localhost:3306/retail\_db --username root --password cloudera --table departments --target-dir spark11/departments --m 2

specify the delete target-dir:

sqoop import --connect jdbc:mysql://localhost:3306/retail\_db --username root --password cloudera --table departments --delete-target-dir --target-dir spark11/departments --m 2

###############################SQOOP###################################

Sqoop connect with mysql:

sqoop import --driver com.mysql.jdbc.Driver --connect jdbc:mysql://localhost/users --username root --password cloudera --table user

Importing the all tables:

sqoop import --driver com.mysql.jdbc.Driver --connect jdbc:mysql://localhost/users --username root --password cloudera --table user

UDEMY URL:

https://www.udemy.com/apache-sqoop-for-certifications-cca-and-hdpcd/learn/v4/t/lecture/8631410?start=765

Sqoop User Guide:

https://sqoop.apache.org/docs/1.4.6/SqoopUserGuide.html#\_connecting\_to\_a\_database\_server

Import statement:

sqoop import --connect jdbc:mysql://10.0.0.41:3306/users --username root --password cloudera --table user --target-dir = /user/root/data/

Still Running and no output

In ruuning state still accepted only:

The application might not be running yet or there is no Node Manager or Container available. This page will be automatically refreshed.

List the yarn cluster:

yarn node -list

Start the yarn cluster:

sudo service hadoop-yarn-resourcemanager start

sudo service hadoop-yarn-nodemanager start

sudo service hadoop-mapreduce-historyserver start

Delete the existing directory:

sqoop import --connect jdbc:mysql://localhost:3306/users --username root --P --table student --delete-target-dir --target-dir /user/root --m 2

M-using for giving the no of reducers in jobs.

Applying the where clause in Sqoop command:

sqoop import --connect jdbc:mysql://localhost:3306/users --username root --P --table student --where "s\_id = 10" --delete-target-dir --target-dir /user/root --m 2

Append the values in particular existing folder:

–incremental <mode>

–check-column <column name>

–last value <last check column value>

sqoop import --connect jdbc:mysql://localhost:3306/users --username root --P --table student --target-dir /user/root --incremental append --check-column s\_id --last-value 39 --m 2

Import all tables:

sqoop import-all-tables --connect jdbc:mysql://localhost:3306/users --username root --P

List the tables and databases:

sqoop list-tables --connect jdbc:mysql://localhost:3306/users --username root --P

sqoop list-databases --connect jdbc:mysql://localhost:3306/users --username root --P

Importing the selected column only:

sqoop import --connect jdbc:mysql://localhost:3306/users --username root --P --table student --columns "s\_id,s\_name" --where "s\_id > 10" --delete-target-dir --target-dir /user/root --m 2

AVRO file format:--as-sequencefile

sqoop import --connect jdbc:mysql://localhost:3306/users --username root --P --table student --columns "s\_id,s\_name" --where "s\_id > 10" --delete-target-dir --target-dir /user/root --as -avrodatafile

Parquet file format:

sqoop import --connect jdbc:mysql://localhost:3306/users --username root --P --table student --columns "s\_id,s\_name" --where "s\_id > 10" --delete-target-dir --target-dir /user/root --as-parquetfile

Delimeter example:

sqoop import --connect jdbc:mysql://localhost:3306/users --username root --P --table student --columns "s\_id,s\_name" --where "s\_id > 10" --delete-target-dir --target-dir /user/root --fields-terminated-by '\t' --lines-terminated-by '\n' -m1

=====================================================================================================================================

Sqoop Job:

You have been given MySQL DB with following details.

user=retail\_dba

password=cloudera

database=retail\_db

table=retail\_db.categories

jdbc URL = jdbc:mysql://quickstart:3306/retail\_db

Please accomplish following activities.

1. Write a Sqoop Job which will import "retaildb.categories" table to hdfs, in a directory

name "categories\_targetJob".

sqoop job --create myJob -- import --connect jdbc:mysql://localhost:3306/retail\_db --username root --P --table categories --target-dir /user/root/spark16/categories

sqoop job --list

sqoop job --show myJob

sqoop job --exec myJob

====================================================================================================================================

Sqoop Export:

Problem Scenario 13 : You have been given following mysql database details as well as

other info.

user=retail\_dba

password=cloudera

Question No : 50 CORRECT TEXT

Cloudera CCA175 : Practice Exam

Leaders in it certification 75

Visit us athttps://www.examsboost.com/test/cca175/

database=retail\_db

jdbc URL = jdbc:mysql://quickstart:3306/retail\_db

Please accomplish following.

1. Create a table in retailedb with following definition.

CREATE table departments\_export (department\_id int(11), department\_name varchar(45),

created\_date T1MESTAMP DEFAULT NOWQ);

2. Now import the data from following directory into departments\_export table,

/user/cloudera/departments new

step 1:

table creation with default time stamp:

CREATE table departments\_export (department\_id int(11), department\_name varchar(45), created\_date TIMESTAMP NOT NULL DEFAULT CURRENT\_TIMESTAMP );

INSERT INTO departments\_export SELECT \* from departments\_new;

sqoop export --connect jdbc:mysql://localhost:3306/retail\_db --username root -P --table departments\_export --export-dir departments\_new

=====================================================================================================================================

Examples:

1.

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\

Step 1:

sqoop import --driver com.mysql.jdbc.Driver --connect jdbc:mysql://localhost:3306/retail\_db --username root -P --table departments --target-dir spark16/departments

Step2:

sqoop import --driver com.mysql.jdbc.Driver --connect jdbc:mysql://localhost:3306/retail\_db --username root -P --table departments --append -fields-terminated-by '|' -lines-terminated-by '\n' --target-dir spark16/departments

2.

Problem Scenario 1 : You have been given MySQL DB with following details.

user=retail\_dba

password=cloudera

database=retail\_db

table=retail\_db.products

jdbc URL = jdbc:mysql://quickstart:3306/retail\_db

Columns of products table : (product\_id | product\_category\_id | product\_name |

product\_description | product\_price | product\_image )

Please accomplish following activities.

1. Copy "retaildb.products" table to hdfs in a directory p93\_products

2. Now sort the products data sorted by product price per category, use productcategoryid

colunm to group by category

sqoop import --driver com.mysql.jdbc.Driver --connect jdbc:mysql://localhost/retail\_db --username root --password cloudera --table products

sudo service hadoop-yarn-resourcemanager start

sudo service hadoop-yarn-nodemanager start

sudo service hadoop-mapreduce-historyserver start

Solutions:

Step1:

sqoop import --driver com.mysql.jdbc.Driver --connect jdbc:mysql://localhost/retail\_db --username root --password cloudera --table products --target-dir /user/root/p93\_products

Read the table in parquet file format:

sqoop import --driver com.mysql.jdbc.Driver --connect jdbc:mysql://localhost/retail\_db --username root --password cloudera --table products --target-dir /user/root/p92\_products --as-parquetfile --m 1

Read the file df:

var df = sqlContext.read.parquet("hdfs:///user/root/p92\_products/8355ac71-615f-41ed-92dd-6adb2dffed16.parquet")

Step2:

var df1 = df.sort("product\_price")

var df2 = df1.groupBy("product\_category\_id")

var df2 = df1.groupBy("product\_category\_id").agg(sum("product\_id") as "count\_per\_role")

3.

Problem Scenario 12 : 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. Create a table in retailedb with following definition.

CREATE table departments\_new (department\_id int(11), department\_name varchar(45), created\_date T1MESTAMP DEFAULT NOW());

2. Now insert records from departments table to departments\_new

3. Now import data from departments\_new table to hdfs.

4. Insert following 5 records in departmentsnew table. insert into departments\_new values(110, "Civil" , null); Insert into departments\_new values(111, "Mechanical" , null); Insert into departments\_new values(112, "Automobile" , null); Insert into departments\_new values(113, "Pharma" , null); Insert into departments\_new values(114, "Social Engineering" , null);

5. Now do the incremental import based on created\_date column.

dt DATETIME DEFAULT CURRENT\_TIMESTAMP

CREATE table departments\_new (department\_id int(11), department\_name varchar(45), created\_date TIMESTAMP DEFAULT CURRENT\_TIMESTAMP);

sqoop import --driver com.mysql.jdbc.Driver --connect jdbc:mysql://localhost/retail\_db --username root -P --table departments

sqoop export –connect jdbc:mysql://localhost/retail\_db –username root –P –table departments\_new –export-dir /user/root/departments

sqoop export --driver com.mysql.jdbc.Driver --connect jdbc:mysql://localhost/retail\_db --username root -P --table departments\_new --export-dir /user/root/departments --fields-terminated-by ',' --columns "department\_id,department\_name"

ALTER TABLE departments\_new ADD PRIMARY KEY (department\_id);

sqoop import --driver com.mysql.jdbc.Driver --connect jdbc:mysql://localhost:3306/retail\_db --username root -P --table departments\_new

insert into departments\_new(department\_id,department\_name,created\_date) values(110, "Civil" , null);

insert into departments\_new(department\_id,department\_name,created\_date) values(114, "Social Engineering" , null);

--incremental lastmodified -check-column created\_date --target-dir sqin -m 1

sqoop import --driver com.mysql.jdbc.Driver --connect jdbc:mysql://localhost:3306/retail\_db --username root -P --table departments\_new --incremental lastmodified -check-column created\_date --target-dir sqin -m 1