Lab: Exporting HDFS Data to an RDBMS

About this Lab

Objective: Export data from HDFS into a MySQL table using Sqoop.

File locations: /root/devph/labs/Lab3.2

Successful outcome: The data in salarydata.txt in HDFS will appear in a table in MySQL named salary2.

Before you begin: Your HDP 2.3 cluster should be up and running within your VM.

Related lesson: ***Inputting Data into HDFS***

Lab Steps

1) Put the Data into HDFS

a. If not already done, open a Terminal in your VM and type "ssh sandbox".
b. Change directories to `/root/devph/labs/Lab3.2`:

```
# cd ~/devph/labs/Lab3.2
View the contents of salarydata.txt:
# tail salarydata.txt
M,49,29000,95103
M,44,34000,95102
M,99,25000,94041
F,93,96000,95105
F,75,9000,94040
F,14,0,95102
M,68,1000,94040
F,45,78000,94041
M,40,6000,95103
F,82,5000,95050

Notice the records in this file contain four values separated by
commas, and the values represent a gender, age, salary, and zip
code, respectively.
```

c. Create a new directory in HDFS named `salarydata`.

```
# hdfs dfs -mkdir salarydata
```

d. Put salarydata.txt into the `salarydata` directory in HDFS.

```
# hdfs dfs –put salarydata.txt salarydata
```

2) Create a Table in the Database

a. There is a script in the Exporting HDFS Data to an RDBMS lab folder that creates a table in MySQL that matches the records in salarydata.txt. View the SQL script:

```
# more salaries2.sql
```

b. Run this script using the following command:

```
# mysql test < salaries2.sql
```

c. Verify that the table was created successfully in MySQL:

```
# mysql
mysql> use test;
mysql> describe salaries2;
+---------+------------+------+-----+---------+-------+
| Field   | Type       | Null | Key | Default | Extra |
+---------+------------+------+-----+---------+-------+
| gender  | varchar(1) | YES  |     | NULL    |       |
| age     | int(11)    | YES  |     | NULL    |       |
| salary  | double     | YES  |     | NULL    |       |
| zipcode | int(11)    | YES  |     | NULL    |       |
+---------+------------+------+-----+---------+-------+
```

d. Exit the mysql prompt:

```
mysql> exit
```

3) Export the Data

a. Run a Sqoop command that exports the `salarydata` folder in HDFS into the salaries2 table in MySQL. At the end of the MapReduce output, you should see a log event stating that 10,000 records were exported.

```
# sqoop export \
--connect jdbc:mysql://sandbox/test?user=root \
--table salaries2 \
--export-dir salarydata \
--input-fields-terminated-by ","
```

b. Verify it worked by viewing the table's contents from the mysql prompt. The output should look like the following:

```
# mysql
mysql> use test;
mysql> select * from salaries2 limit 10;
+--------+------+--------+---------+
| gender | age  | salary | zipcode |
+--------+------+--------+---------+
| M      |   57 |  39000 |   95050 |
| F      |   63 |  41000 |   95102 |
| M      |   55 |  99000 |   94040 |
| M      |   51 |  58000 |   95102 |
| M      |   75 |  43000 |   95101 |
| M      |   94 |  11000 |   95051 |
| M      |   28 |   6000 |   94041 |
| M      |   14 |      0 |   95102 |
| M      |    3 |      0 |   95101 |
| M      |   25 |  26000 |   94040 |
+--------+------+--------+---------+
```

c. Exit the mysql prompt.

Result

You have now used Sqoop to export data from HDFS into a database table in MySQL.