Session 11

Assignment 1

|  |  |
| --- | --- |
| **Prepared For:** | AcadGild |
|  |  |
| **Document Approval:** | **AcadGild** |
|  |  |
|  |  |
|  |  |
|  |  |
| **Project Title:** | Session 11 – Assignment 1 |
|  |  |
| **Prepared By:** | Duncan Burgess |
|  |  |
|  | dburgess@duncb.com |
|  |  |
| **Primary Engineer:** | Duncan Burgess |
|  |  |
| **Document Reference:** | **Session 10 – Assignment 1** |
|  |  |
| **Start Date:** | 06/10/2017 |
|  |  |
|  |  |



# 

# Contents

[Contents 2](#_Toc495040013)

[Change History 3](#_Toc495040014)

[1. Problem Statement 4](#_Toc495040015)

[2. Dataset 5](#_Toc495040016)

[3. Solution 5](#_Toc495040017)

[3.1. Transfer data between MySQL and HDFS (Import and Export) using Sqoop 5](#_Toc495040018)

[3.1.1. Preparing Mysql 5](#_Toc495040019)

[3.1.2. IMPORT FROM MySQL to HDFS: 6](#_Toc495040020)

[3.1.3. IMPORT FROM MySQL to HDFS 8](#_Toc495040021)

[3.1.4. Export FROM HDFS to MYSQL 9](#_Toc495040022)

[3.1.5. Transfer data between MySQL and Hive (Import and Export only selected columns) using Sqoop 10](#_Toc495040023)

[3.1.6. IMPORT FROM HIVE to MySQL: 12](#_Toc495040024)

# Change History

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Document Revision** | **Date** | **Authored By** | **Authorised By** | **Sections Affected** | **Reason for Change** |
| Rev 01 | 06/10/2017 | Duncan Burgess |  | All | Initial release. |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |

# Problem Statement

Perform and explain the code flow and the associated result for the below tasks. Candidates should create and use their own employee dataset for the same. Share the screenshot of the commands used and its associated result.

* Transfer data between MySQL and HDFS (Import and Export) using Sqoop.
* Transfer data between MySQL and Hive (Import and Export only selected columns) using Sqoop.

# Dataset

The following Dataset was used for this assignment

**optuminf.txt**

1,Optumuk,Duncan,Burgess,52,9848022337,Mawsley,89,windows

2,Optumuk,Mike,Tartaglia,32,9848022338,Burton Latimer,78,windows

3,Optumuk,Ben,Ludford,22,9848022339,Swindon,90,windows

4,Optumuk,Tracey,Baker,21,9848022330,Birmingham,93,management

5,Optumuk,Chris,Ginn,23,9848022336,Cardif,14,SQL

6,Optumuk,Chris,Reid,23,9848022335,Leek,87,linux

7,Optumuk,Stuart,Ford,24,9848022334,Edingborough,83,linux

8,Optumuk,Shaun,Turner,38,9848022388,Leicester,67,SQL

10,Optumuk,Rosario,Marino,24,9848022333,Oxford,72,windows

# Solution

## Transfer data between MySQL and HDFS (Import and Export) using Sqoop

Tasks required for analysing Data

* Starting MySQL as root user/mode
* Creating the Database sqoop\_db
* Creating table optumstaff

### Preparing Mysql

**Create database squoop\_db**

*mysql> create database sqoop\_db;*

*Query OK, 1 row affected (0.00 sec)*

*mysql> use sqoop\_db;*

*Database changed*

**Create Table optumstaff**

*mysql> create table optumstaff*

*-> (*

*-> emp\_ID int,*

*-> company varchar(20),*

*-> firstname varchar(30),*

*-> lastname varchar(30),*

*-> age int,*

*-> city varchar (30),*

*-> jobs int,*

*-> skill varchar(30)*

*-> );*

*Query OK, 0 rows affected (0.09 sec)*

**Check Table**

*mysql> describe optumstaff;*

*+-----------+-------------+------+-----------+---------+-----------------------+*

*| Field | Type | Null | Key | Default | Extra |*

*+-----------+-------------+------+-----+---------+----------------------------+*

*| emp\_ID | int(11) | YES | | NULL | |*

*| company | varchar(20) | YES | | NULL | |*

*| firstname | varchar(30) | YES | | NULL | |*

*| lastname | varchar(30) | YES | | NULL | |*

*| age | int(11) | YES | | NULL | |*

*| city | varchar(30) | YES | | NULL | |*

*| jobs | int(11) | YES | | NULL | |*

*| skill | varchar(30) | YES | | NULL | |*

*+-----------+-------------+------+------------+---------+---------------------+*

*8 rows in set (0.00 sec)*

**Inserting records/values into the table optumstaff**

*mysql> insert into optumstaff values (1,'Optumuk','Duncan','Burgess',52,'Mawsley',89,'windows');*

*Query OK, 1 row affected (0.01 sec)*

*mysql> insert into optumstaff values (2,'Optumuk','Mike','Tartaglia',32,'Burton Latimer',78,'windows');*

*Query OK, 1 row affected (0.00 sec)*

*mysql> insert into optumstaff values (3,'Optumuk','Ben','Ludford',22,'Swindon',90,'windows');*

*Query OK, 1 row affected (0.00 sec)*

*mysql> insert into optumstaff values (4,'Optumuk','Tracey','Baker',21,'Birmingham',93,'management');*

*Query OK, 1 row affected (0.00 sec)*

*mysql> insert into optumstaff values (5,'Optumuk','Chris','Ginn',23,'Cardif',14,'SQL');*

*Query OK, 1 row affected (0.01 sec)*

*mysql> insert into optumstaff values (6,'Optumuk','Chris','Reid',23,'Leek',87,'linux');*

*Query OK, 1 row affected (0.01 sec)*

*mysql>*

**Check records have been created**

*mysql> select \* from optumstaff;*

*+--------+---------+-----------+-----------+------+----------------+------+------------+*

*| emp\_ID | company | firstname | lastname | age | city | jobs | skill |*

*+--------+---------+-----------+-----------+------+----------------+------+------------+*

*| 1 | Optumuk | Duncan | Burgess | 52 | Mawsley | 89 | windows |*

*| 2 | Optumuk | Mike | Tartaglia | 32 | Burton Latimer | 78 | windows |*

*| 3 | Optumuk | Ben | Ludford | 22 | Swindon | 90 | windows |*

*| 4 | Optumuk | Tracey | Baker | 21 | Birmingham | 93 | management |*

*| 5 | Optumuk | Chris | Ginn | 23 | Cardif | 14 | SQL |*

*| 6 | Optumuk | Chris | Reid | 23 | Leek | 87 | linux |*

*+--------+---------+-----------+-----------+------+----------------+------+------------+*

*6 rows in set (0.00 sec)*

*mysql>*

### IMPORT FROM MySQL to HDFS:

To transfer the data from the from the MySQL table to a HDFS Location, we use the Sqoop **import** command with the following parameters:

Connecting to the database **squoop\_db** in the MySQL server using the **JDBC Driver**

Entering the system username (**root**) & –P to manually enter password.)

Specifying the table in MySQL **‘optumstaff**, where the data is to come from

Specifying the number of mapper jobs to execute 1. The default is 4.

**Sqoop command run**

*sqoop import --connect jdbc:mysql://localhost/sqoop\_db --username 'root' -P --table 'optumstaff' --target-dir '/user/cloudera/sqoop\_import1' -m 1;*

**Results**

*17/10/04 05:54:05 INFO mapreduce.Job: Job job\_1507104061255\_0017 completed successfully*

*17/10/04 05:54:05 INFO mapreduce.Job: Counters: 30*

*File System Counters*

*FILE: Number of bytes read=0*

*FILE: Number of bytes written=151480*

*FILE: Number of read operations=0*

*FILE: Number of large read operations=0*

*FILE: Number of write operations=0*

*HDFS: Number of bytes read=87*

*HDFS: Number of bytes written=272*

*HDFS: Number of read operations=4*

*HDFS: Number of large read operations=0*

*HDFS: Number of write operations=2*

*Job Counters*

*Launched map tasks=1*

*Other local map tasks=1*

*Total time spent by all maps in occupied slots (ms)=23910*

*Total time spent by all reduces in occupied slots (ms)=0*

*Total time spent by all map tasks (ms)=23910*

*Total vcore-milliseconds taken by all map tasks=23910*

*Total megabyte-milliseconds taken by all map tasks=24483840*

*Map-Reduce Framework*

*Map input records=6*

*Map output records=6*

*Input split bytes=87*

*Spilled Records=0*

*Failed Shuffles=0*

*Merged Map outputs=0*

*GC time elapsed (ms)=119*

*CPU time spent (ms)=1310*

*Physical memory (bytes) snapshot=122650624*

*Virtual memory (bytes) snapshot=1510170624*

*Total committed heap usage (bytes)=60882944*

*File Input Format Counters*

*Bytes Read=0*

*File Output Format Counters*

*Bytes Written=272*

*17/10/04 05:54:05 INFO mapreduce.ImportJobBase: Transferred 272 bytes in 49.8991 seconds (5.451 bytes/sec)*

*17/10/04 05:54:05 INFO mapreduce.ImportJobBase: Retrieved 6 records.*

**Check Results**

Using the **ls** command to check the content of the HDFS.

The file has been successfully transferred to the HDFS. Here because only one mapper is given, only one output file is found. (No split)

*[cloudera@quickstart ~]$ hadoop fs -ls sqoop\_import1*

*Found 2 items*

*-rw-r--r-- 1 cloudera cloudera 0 2017-10-04 05:54 sqoop\_import1/\_SUCCESS*

*-rw-r--r-- 1 cloudera cloudera 272 2017-10-04 05:54 sqoop\_import1/part-m-00000*

*[cloudera@quickstart ~]$ hadoop fs -cat sqoop\_import1/part-m-00000*

*1,Optumuk,Duncan,Burgess,52,Mawsley,89,windows*

*2,Optumuk,Mike,Tartaglia,32,Burton Latimer,78,windows*

*3,Optumuk,Ben,Ludford,22,Swindon,90,windows*

*4,Optumuk,Tracey,Baker,21,Birmingham,93,management*

*5,Optumuk,Chris,Ginn,23,Cardif,14,SQL*

*6,Optumuk,Chris,Reid,23,Leek,87,linux*

*[cloudera@quickstart ~]$*

**Note** This could have also been performed using multiple mappers using the –split option see below with 2 mappers.

*sqoop import --connect jdbc:mysql://localhost/sqoop\_db --username 'root' -P --table 'optumstaff' --target-dir '/user/cloudera/sqoop\_import2' --split-by emp\_ID -m 2;*

*Job Counters*

*Launched map tasks=2*

*Other local map tasks=2*

*hadoop fs -ls sqoop\_import2*

*Found 3 items*

*-rw-r--r-- 1 cloudera cloudera 0 2017-10-04 06:07 sqoop\_import2/\_SUCCESS*

*-rw-r--r-- 1 cloudera cloudera 145 2017-10-04 06:07 sqoop\_import2/part-m-00000*

*-rw-r--r-- 1 cloudera cloudera 127 2017-10-04 06:07 sqoop\_import2/part-m-00001*

### IMPORT FROM MySQL to HDFS

**Creating table optum\_staging with all columns**

*mysql> use sqoop\_db;*

*Database changed*

*mysql> create table optum\_staging*

*-> (*

*-> emp\_ID int,*

*-> company varchar(20),*

*-> firstname varchar(30),*

*-> lastname varchar(30),*

*-> age int,*

*-> city varchar (30),*

*-> jobs int,*

*-> skill varchar(30)*

*-> );*

*Query OK, 0 rows affected (0.05 sec)*

1. To transfer the data from the HDFS to a MySQL table, we use **optum\_staging.** Here we are using the sqoop **export** command with the following parameters:
2. Connecting to the database **sqoop\_db** in the MySQL server using the **JDBC Driver**
3. Entering the system username (**root**) & -P
4. Specifying the MySQL table name ‘ optum\_staging**’** from where the data is to come from
5. Specifying the directory in the HDFS to store the exported value from the MySQL table, **Here it is sqoop\_import1**
6. We are also specifying that the data in the HDFS directory is a comma- separated file
7. Specifying the number of mapper jobs to execute. Here 2. Default 4
8. Also, specifying all the columns

*sqoop export --connect jdbc:mysql://localhost:3306/sqoop\_db --username 'root' -P --table 'optum\_staging' --export-dir 'sqoop\_import1' --input-fields-terminated-by ',' -m 2 --columns emp\_ID,company,firstname,lastname,age,city,jobs,skill*

*17/10/04 06:23:17 INFO mapreduce.Job: Job job\_1507104061255\_0020 completed successfully*

*17/10/04 06:23:17 INFO mapreduce.Job: Counters: 30*

*File System Counters*

*FILE: Number of bytes read=0*

*FILE: Number of bytes written=302916*

*FILE: Number of read operations=0*

*FILE: Number of large read operations=0*

*FILE: Number of write operations=0*

*HDFS: Number of bytes read=720*

*HDFS: Number of bytes written=0*

*HDFS: Number of read operations=8*

*HDFS: Number of large read operations=0*

*HDFS: Number of write operations=0*

*Job Counters*

*Launched map tasks=2*

*Data-local map tasks=2*

*Total time spent by all maps in occupied slots (ms)=42302*

*Total time spent by all reduces in occupied slots (ms)=0*

*Total time spent by all map tasks (ms)=42302*

*Total vcore-milliseconds taken by all map tasks=42302*

*Total megabyte-milliseconds taken by all map tasks=43317248*

*Map-Reduce Framework*

*Map input records=6*

*Map output records=6*

*Input split bytes=306*

*Spilled Records=0*

*Failed Shuffles=0*

*Merged Map outputs=0*

*GC time elapsed (ms)=488*

*CPU time spent (ms)=2290*

*Physical memory (bytes) snapshot=215842816*

*Virtual memory (bytes) snapshot=3016105984*

*Total committed heap usage (bytes)=121765888*

*File Input Format Counters*

*Bytes Read=0*

*File Output Format Counters*

*Bytes Written=0*

*17/10/04 06:23:17 INFO mapreduce.ExportJobBase: Transferred 720 bytes in 42.4462 seconds (16.9626 bytes/sec)*

*17/10/04 06:23:17 INFO mapreduce.ExportJobBase: Exported 6 records.*

Selecting/Listing all the records in the created table to check if all the records have indeed been created successfully

*mysql> select \* from optum\_staging;*

*+--------+---------+-----------+-----------+------+----------------+------+------------+*

*| emp\_ID | company | firstname | lastname | age | city | jobs | skill |*

*+--------+---------+-----------+-----------+------+----------------+------+------------+*

*| 4 | Optumuk | Tracey | Baker | 21 | Birmingham | 93 | management |*

*| 1 | Optumuk | Duncan | Burgess | 52 | Mawsley | 89 | windows |*

*| 2 | Optumuk | Mike | Tartaglia | 32 | Burton Latimer | 78 | windows |*

*| 3 | Optumuk | Ben | Ludford | 22 | Swindon | 90 | windows |*

*| 5 | Optumuk | Chris | Ginn | 23 | Cardif | 14 | SQL |*

*| 6 | Optumuk | Chris | Reid | 23 | Leek | 87 | linux |*

*+--------+---------+-----------+-----------+------+----------------+------+------------+*

*6 rows in set (0.00 sec)*

### Export FROM HDFS to MYSQL

To transfer the data from the HDFS to a MySQL table, we use **SQOOP.** Here we are using the Sqoop **export** command with the following parameters:

1. Connecting to the database **sqoop\_db** in the MySQL server using the **JDBC Driver**
2. Entering the system username (**root**) & -P
3. Specifying the MySQL table name ‘ optum\_staging**’** from where the data is to come from
4. Specifying the directory in the HDFS to store the exported value from the MySQL table, **sqoop\_import1**
5. We are also specifying that the data in the HDFS directory is a comma separated file
6. Specifying the number of mappers =2.
7. Also, specifying all the columns

sqoop export --connect jdbc:mysql://localhost:3306/sqoop\_db --username 'root' -P --table 'optum\_staging' --export-dir 'sqoop\_import1' --input-fields-terminated-by ',' -m 2 --columns emp\_ID,company,firstname,lastname,age,city,jobs,skill

*17/10/04 06:23:17 INFO mapreduce.Job: Job job\_1507104061255\_0020 completed successfully*

*17/10/04 06:23:17 INFO mapreduce.Job: Counters: 30*

*File System Counters*

*FILE: Number of bytes read=0*

*FILE: Number of bytes written=302916*

*FILE: Number of read operations=0*

*FILE: Number of large read operations=0*

*FILE: Number of write operations=0*

*HDFS: Number of bytes read=720*

*HDFS: Number of bytes written=0*

*HDFS: Number of read operations=8*

*HDFS: Number of large read operations=0*

*HDFS: Number of write operations=0*

*Job Counters*

*Launched map tasks=2*

*Data-local map tasks=2*

*Total time spent by all maps in occupied slots (ms)=42302*

*Total time spent by all reduces in occupied slots (ms)=0*

*Total time spent by all map tasks (ms)=42302*

*Total vcore-milliseconds taken by all map tasks=42302*

*Total megabyte-milliseconds taken by all map tasks=43317248*

*Map-Reduce Framework*

*Map input records=6*

*Map output records=6*

*Input split bytes=306*

*Spilled Records=0*

*Failed Shuffles=0*

*Merged Map outputs=0*

*GC time elapsed (ms)=488*

*CPU time spent (ms)=2290*

*Physical memory (bytes) snapshot=215842816*

*Virtual memory (bytes) snapshot=3016105984*

*Total committed heap usage (bytes)=121765888*

*File Input Format Counters*

*Bytes Read=0*

*File Output Format Counters*

*Bytes Written=0*

*17/10/04 06:23:17 INFO mapreduce.ExportJobBase: Transferred 720 bytes in 42.4462 seconds (16.9626 bytes/sec)*

*17/10/04 06:23:17 INFO mapreduce.ExportJobBase: Exported 6 records.*

**Checking Results**

*mysql> select \* from optum\_staging;*

*+--------+---------+-----------+-----------+------+----------------+------+------------+*

*| emp\_ID | company | firstname | lastname | age | city | jobs | skill |*

*+--------+---------+-----------+-----------+------+----------------+------+------------+*

*| 4 | Optumuk | Tracey | Baker | 21 | Birmingham | 93 | management |*

*| 1 | Optumuk | Duncan | Burgess | 52 | Mawsley | 89 | windows |*

*| 2 | Optumuk | Mike | Tartaglia | 32 | Burton Latimer | 78 | windows |*

*| 3 | Optumuk | Ben | Ludford | 22 | Swindon | 90 | windows |*

*| 5 | Optumuk | Chris | Ginn | 23 | Cardif | 14 | SQL |*

*| 6 | Optumuk | Chris | Reid | 23 | Leek | 87 | linux |*

*+--------+---------+-----------+-----------+------+----------------+------+------------+*

*6 rows in set (0.00 sec)*

### Transfer data between MySQL and Hive (Import and Export only selected columns) using Sqoop

To transfer the data from the from the MySQL table to a Hive table, we use the Sqoop **import** command with the following parameters:

1. Connecting to the database sqoop\_db in the MySQL server using the JDBC Driver
2. Entering the system username (root) & -P
3. For the purpose of parallel execution (which increases the speed of execution), we use the split-by emp\_ID
4. 4 columns selected 3 columns from the MySQL table: ID, firstname, lastname and jobs.
5. Specifying the MySQL table name to get the data from. **‘optum\_staff’**
6. Specifying the target directory to store the data and specifying that the data is ‘,’ delimted
7. Using hive import command, we create the hive table to load the data in in the query itself, giving it the name optum\_hiveand storing it in the **default** database.

*sqoop import --connect jdbc:mysql://localhost/sqoop\_db --username 'root' -P --split-by emp\_ID --columns emp\_ID,firstname,latname,jobs -input-fields-terminated-by ',' --table 'optum\_staff' --target-dir '/sqoop\_hive' --hive-import --fields-terminated-by ',' --create-hive-table --hive-table default.optum\_hive -m 1;*

*17/10/04 06:29:39 INFO mapreduce.Job: Job job\_1507104061255\_0021 completed successfully*

*17/10/04 06:29:39 INFO mapreduce.Job: Counters: 30*

*File System Counters*

*FILE: Number of bytes read=0*

*FILE: Number of bytes written=152008*

*FILE: Number of read operations=0*

*FILE: Number of large read operations=0*

*FILE: Number of write operations=0*

*HDFS: Number of bytes read=87*

*HDFS: Number of bytes written=107*

*HDFS: Number of read operations=4*

*HDFS: Number of large read operations=0*

*HDFS: Number of write operations=2*

*Job Counters*

*Launched map tasks=1*

*Other local map tasks=1*

*Total time spent by all maps in occupied slots (ms)=13603*

*Total time spent by all reduces in occupied slots (ms)=0*

*Total time spent by all map tasks (ms)=13603*

*Total vcore-milliseconds taken by all map tasks=13603*

*Total megabyte-milliseconds taken by all map tasks=13929472*

*Map-Reduce Framework*

*Map input records=6*

*Map output records=6*

*Input split bytes=87*

*Spilled Records=0*

*Failed Shuffles=0*

*Merged Map outputs=0*

*GC time elapsed (ms)=108*

*CPU time spent (ms)=1130*

*Physical memory (bytes) snapshot=129265664*

*Virtual memory (bytes) snapshot=1510166528*

*Total committed heap usage (bytes)=60882944*

*File Input Format Counters*

*Bytes Read=0*

*File Output Format Counters*

*Bytes Written=107*

*17/10/04 06:29:39 INFO mapreduce.ImportJobBase: Transferred 107 bytes in 34.8201 seconds (3.0729 bytes/sec)*

*17/10/04 06:29:39 INFO mapreduce.ImportJobBase: Retrieved 6 records.*

*17/10/04 06:29:39 INFO manager.SqlManager: Executing SQL statement: SELECT t.\* FROM `optum\_staff` AS t LIMIT 1*

*17/10/04 06:29:39 INFO hive.HiveImport: Loading uploaded data into Hive*

*Logging initialized using configuration in jar:file:/usr/lib/hive/lib/hive-common-1.1.0-cdh5.12.0.jar!/hive-log4j.properties*

*OK*

*Time taken: 3.977 seconds*

*Loading data to table default.optum\_hive*

*Table default.optum\_hive stats: [numFiles=1, totalSize=107]*

*OK*

*Time taken: 0.68 seconds*

Using default database and table **optum\_hive** table to check all the records have been created successfully.

*hive> select \* from optum\_hive;*

*OK*

*1 Duncan Burgess 89*

*2 Mike Tartaglia 78*

*3 Ben Ludford 90*

*4 Tracey Baker 93*

*5 Chris Ginn 14*

*6 Chris Reid 87*

*Time taken: 0.719 seconds, Fetched: 6 row(s)*

*hive>*

### IMPORT FROM HIVE to MySQL:

Creating an empty table in sqoop\_db databases - optum\_hive\_mysql to contain the data imported from the Hive table optum\_hive

*mysql> create table optum\_mysql*

*-> (*

*-> emp\_id int,*

*-> firstname varchar(30)*

*-> );*

*Query OK, 0 rows affected (0.12 sec)*

To transfer the data from the from the Hive table to a MySQL table, we use the Sqoop **export** command with the following parameters:

Connecting to the database **Sqoop** in the MySQL server using the **JDBC Driver**

Entering the system username (**root**) & -P

Using the **--direct** mode for fat imports

Only selected 2 columns from the Hive table: ID and Name to be transferred to the MySQL Table to be created

Specifying the MySQL table name to store the data in **‘optum\_hive\_mysql’.** It has just 2 columns :emp\_ID and Name.

Specifying the export target directory i.e. the hive warehouse to get the data and specifying that the data is ‘\t’ delimited

Specifying 1 map job. So you get only 1 output file

**Run command**

*The data in the Hive table optum\_hive on the HDFS. As it is run by 2 map jobs while storing it has two files of data*

*sqoop export --connect jdbc:mysql://localhost/sqoop\_db --username 'root' -P --direct --columns emp\_ID,firstname --table optum\_hive\_mysql --export-dir /user/hive/warehouse/optum\_hive --input-fields-terminated-by '/t' -m 1;*

**Results**

*17/10/04 07:06:21 INFO mapreduce.Job: Job job\_1507104061255\_0023 completed successfully*

*17/10/04 07:06:21 INFO mapreduce.Job: Counters: 30*

*File System Counters*

*FILE: Number of bytes read=0*

*FILE: Number of bytes written=152044*

*FILE: Number of read operations=0*

*FILE: Number of large read operations=0*

*FILE: Number of write operations=0*

*HDFS: Number of bytes read=266*

*HDFS: Number of bytes written=0*

*HDFS: Number of read operations=4*

*HDFS: Number of large read operations=0*

*HDFS: Number of write operations=0*

*Job Counters*

*Launched map tasks=1*

*Data-local map tasks=1*

*Total time spent by all maps in occupied slots (ms)=10107*

*Total time spent by all reduces in occupied slots (ms)=0*

*Total time spent by all map tasks (ms)=10107*

*Total vcore-milliseconds taken by all map tasks=10107*

*Total megabyte-milliseconds taken by all map tasks=10349568*

*Map-Reduce Framework*

*Map input records=6*

*Map output records=0*

*Input split bytes=156*

*Spilled Records=0*

*Failed Shuffles=0*

*Merged Map outputs=0*

*GC time elapsed (ms)=60*

*CPU time spent (ms)=570*

*Physical memory (bytes) snapshot=117161984*

*Virtual memory (bytes) snapshot=1507880960*

*Total committed heap usage (bytes)=60817408*

*File Input Format Counters*

*Bytes Read=0*

*File Output Format Counters*

*Bytes Written=0*

*17/10/04 07:06:21 INFO mapreduce.ExportJobBase: Transferred 266 bytes in 32.1941 seconds (8.2624 bytes/sec)*

*17/10/04 07:06:21 INFO mapreduce.ExportJobBase: Exported 6 records.*

Finally check the data has been inserted into the MySQL table successfully.

mysql> select \* from optum\_hive\_mysql;

*+--------+-----------+*

*| emp\_id | firstname |*

*+--------+-----------+*

*| 1 | Duncan |*

*| 2 | Mike |*

*| 3 | Ben |*

*| 4 | Tracey |*

*| 5 | Chris |*

*| 6 | Chris |*

*+--------+-----------+*

*6 rows in set (0.01 sec)*