

## Assignment 6.1

### DATA INGESTION TOOL SQOOP

#### TASK 1-

We will verify table called Person in simplidb database. Person table has person\_id as primary key.

```
mysql> show tables;
+-----+
| Tables_in_simplidb |
+-----+
| Person              |
| backup_person       |
| employee             |
+-----+
3 rows in set (0.00 sec)

mysql> select * from Person;
+-----+-----+-----+-----+-----+
| person_id | lname | fname | area      | city      |
+-----+-----+-----+-----+-----+
| 1         | Rizvi | Syed   | Whitefield | Bangalore |
| 2         | Rizvi | Syed   | Whitefield | Bangalore |
| 3         | Rizvi | Syed   | Whitefield | Bangalore |
| 4         | Rizvi | Syed   | Whitefield | Bangalore |
| 5         | Rizvi | Syed   | Whitefield | Bangalore |
| 6         | Rizvi | Syed   | Whitefield | Bangalore |
| 7         | Rizvi | Syed   | Whitefield | Bangalore |
| 8         | Rizvi | Syed   | Whitefield | Bangalore |
| 9         | Rizvi | Syed   | Whitefield | Bangalore |
| 10        | Rizvi | Syed   | Whitefield | Bangalore |
| 11        | Rizvi | Syed   | Whitefield | Bangalore |
| 12        | Rizvi | Syed   | Whitefield | Bangalore |
+-----+-----+-----+-----+-----+
12 rows in set (0.00 sec)
```

Below is the data present in HDFS at 'sqoopout' folder

```
[acdgild@localhost ~]$ hadoop fs -cat /sqoopout/*
18/07/29 13:38:21 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
1,Rizvi,Syed,Whitefield,Bangalore
2,Rizvi,Syed,Whitefield,Bangalore
3,Rizvi,Syed,Whitefield,Bangalore
4,Rizvi,Syed,Whitefield,Bangalore
5,Rizvi,Syed,Whitefield,Bangalore
6,Rizvi,Syed,Whitefield,Bangalore
7,Rizvi,Syed,Whitefield,Bangalore
8,Rizvi,Syed,Whitefield,Bangalore
9,Rizvi,Syed,Whitefield,Bangalore
10,Rizvi,Syed,Whitefield,Bangalore
11,Rizvi,Syed,Whitefield,Bangalore
12,Rizvi,Syed,Whitefield,Bangalore
```

Now we will delete the data from Person table :

```
mysql> truncate table Person;
Query OK, 0 rows affected (0.18 sec)

mysql> select * from Person;
Empty set (0.00 sec)
```

## Assignment 6.1

### DATA INGESTION TOOL SQOOP

Below we run the sqoop export command to export contents from /sqoopout folder in HDFS to table Person

```
sqoop export --connect jdbc:mysql://localhost/simplidb --table Person --username root -P --export-dir /sqoopout --input-fields-terminated-by ',' -m 1 --columns person_id,lname,fname,area,city;
```

```
[acacgild@localhost ~]$ sqoop export --connect jdbc:mysql://localhost/simplidb --table Person --username root -P --export-dir /sqoopout --input-fields-terminated-by ',' -m 1 --columns person_id,lname,fname,area,city;
Warning: /home/acacgild/install/sqoop/sqoop-1.4.6.bin_hadoop-2.0.4-alpha/./hcatalog does not exist! HCatalog jobs will fail.
Please set $HCAT_HOME to the root of your HCatalog installation.
Warning: /home/acacgild/install/sqoop/sqoop-1.4.6.bin_hadoop-2.0.4-alpha/./accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
18/07/29 13:58:06 INFO Sqoop.Sqoop: Running Sqoop version: 1.4.6
Enter password:
18/07/29 13:58:17 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
18/07/29 13:58:17 INFO tool.CodeGenTool: Beginning code generation
Sun Jul 29 13:58:18 IST 2018 WARN: Establishing SSL connection without server's identity verification is not recommended. According to MySQL 5.5.45+, 5.6.26+ and 5.7.6+ requirements SSL connection must be established by default if explicit option isn't set. For compliance with existing applications not using SSL the verifyServerCertificate property is set to 'false'. You need either to explicitly disable SSL by setting useSSL=false, or set useSSL=true and provide truststore for server certificate verification.
18/07/29 13:58:19 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM 'Person' AS t LIMIT 1
18/07/29 13:58:19 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM 'Person' AS t LIMIT 1
18/07/29 13:58:19 INFO orm.CompilationManager: HADOOP_MAPRED_HOME is /home/acacgild/install/hadoop/hadoop-2.6.5
Note: /tmp/sqoop-acacgild/compile/80b9b78172f83a79a7decc9d67ba5f06/Person.java uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
18/07/29 13:58:24 INFO orm.CompilationManager: Writing jar file: /tmp/sqoop-acacgild/compile/80b9b78172f83a79a7decc9d67ba5f06/Person.jar
18/07/29 13:58:24 INFO mapreduce.ExportJobBase: Beginning export of Person
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/acacgild/install/hadoop/hadoop-2.6.5/share/hadoop/common/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/home/acacgild/install/hbase/hbase-1.2.6/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
18/07/29 13:58:24 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/07/29 13:58:24 INFO Configuration.deprecation: mapred.jar is deprecated. Instead, use mapreduce.job.jar
18/07/29 13:58:27 INFO Configuration.deprecation: mapred.reduce.tasks.speculative.execution is deprecated. Instead, use mapreduce.reduce.speculative
18/07/29 13:58:27 INFO Configuration.deprecation: mapred.map.tasks.speculative.execution is deprecated. Instead, use mapreduce.map.speculative
18/07/29 13:58:27 INFO Configuration.deprecation: mapred.map.tasks is deprecated. Instead, use mapreduce.job.maps
18/07/29 13:58:27 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/07/29 13:58:32 INFO input.FileInputFormat: Total input paths to process : 4
18/07/29 13:58:32 INFO input.FileInputFormat: Total input paths to process : 4
18/07/29 13:58:32 INFO mapreduce.JobSubmitter: number of splits:1
18/07/29 13:58:32 INFO Configuration.deprecation: mapred.map.tasks.speculative.execution is deprecated. Instead, use mapreduce.map.speculative
18/07/29 13:58:32 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1532845056857_0009
18/07/29 13:58:33 INFO impl.YarnClientImpl: Submitted application application_1532845056857_0009
18/07/29 13:58:33 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1532845056857_0009/
18/07/29 13:58:33 INFO mapreduce.Job: Running job: job_1532845056857_0009
18/07/29 13:58:49 INFO mapreduce.Job: Job job_1532845056857_0009 running in uber mode : false
18/07/29 13:58:49 INFO mapreduce.Job: map 0% reduce 0%
18/07/29 13:59:01 INFO mapreduce.Job: map 100% reduce 0%
18/07/29 13:59:01 INFO mapreduce.Job: Job job_1532845056857_0009 completed successfully
18/07/29 13:59:02 INFO mapreduce.Job: Counters: 30
File System Counters
  FILE: Number of bytes read=0
  FILE: Number of bytes written=127541
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=727
  HDFS: Number of bytes written=0
  HDFS: Number of read operations=13
  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)=9266
  Total time spent by all reduces in occupied slots (ms)=0
  Total time spent by all map tasks (ms)=9266
  Total vcore-milliseconds taken by all map tasks=9266
  Total megabyte-milliseconds taken by all map tasks=9488384
Map-Reduce Framework
  Map input records=12
  Map output records=12
```

## Assignment 6.1

### DATA INGESTION TOOL SQOOP

Below screenshot image shows that the data from sqoopout folder is loaded in the table Person.

```
mysql> select * from Person;
+-----+-----+-----+-----+-----+
| person_id | lname | fname | area | city |
+-----+-----+-----+-----+-----+
| 1 | Rizvi | Syed | Whitefield | Bangalore |
| 2 | Rizvi | Syed | Whitefield | Bangalore |
| 3 | Rizvi | Syed | Whitefield | Bangalore |
| 4 | Rizvi | Syed | Whitefield | Bangalore |
| 5 | Rizvi | Syed | Whitefield | Bangalore |
| 6 | Rizvi | Syed | Whitefield | Bangalore |
| 7 | Rizvi | Syed | Whitefield | Bangalore |
| 8 | Rizvi | Syed | Whitefield | Bangalore |
| 9 | Rizvi | Syed | Whitefield | Bangalore |
| 10 | Rizvi | Syed | Whitefield | Bangalore |
| 11 | Rizvi | Syed | Whitefield | Bangalore |
| 12 | Rizvi | Syed | Whitefield | Bangalore |
+-----+-----+-----+-----+-----+
12 rows in set (0.00 sec)
```

#### TASK 2:

Below we have executed sqoop import command to place the data from the table Person having Person\_id =3 to HDFS location /sqoopout2 as shown in below screenshot :

```
sqoop import --connect jdbc:mysql://localhost/simplidb --table Person --username root --password Root@123 --target-dir /sqoopout2 --split-by person_id -m 1 --where "person_id=3"
```

```
[acadgild@localhost ~]$ sqoop import --connect jdbc:mysql://localhost/simplidb --table Person --username root --password Root@123 --target-dir /sqoopout2 --split-by Person_id -m 1 --where "person_id=3" --columns person_id,lname,fname,area,city;
Warning: /home/acadgild/install/sqoop/sqoop-1.4.6/bin_hadoop-2.0.4-alpha/./hcatalog does not exist! HCatalog jobs will fail.
Please set $HCAT_HOME to the root of your HCatalog installation.
Warning: /home/acadgild/install/sqoop/sqoop-1.4.6/bin_hadoop-2.0.4-alpha/./accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
18/07/29 15:02:20 INFO sqoop.Sqoop: Running Sqoop version: 1.4.6
18/07/29 15:02:20 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.
18/07/29 15:02:21 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
18/07/29 15:02:21 INFO tool.CodeGenTool: Beginning code generation
Sun Jul 29 15:02:21 IST 2018 WARN: Establishing SSL connection without server's identity verification is not recommended. According to MySQL 5.5.45+, 5.6.26+ and 5.7.6+ requirements SSL connection must be established by default if explicit option isn't set. For compliance with existing applications not using SSL the verifyServerCertificate property is set to 'false'. You need either to explicitly disable SSL by setting useSSL=false, or set useSSL=true and provide truststore for server certificate verification.
18/07/29 15:02:23 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM 'Person' AS t LIMIT 1
18/07/29 15:02:23 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM 'Person' AS t LIMIT 1
18/07/29 15:02:23 INFO orm.CompilationManager: HADOOP_MAPRED_HOME is /home/acadgild/install/hadoop/hadoop-2.6.5
Note: /tmp/sqoop-acadgild/compile/ded622aba0a1bcbd6ae555b03580c5f4/Person.java uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
18/07/29 15:02:27 INFO orm.CompilationManager: Writing jar file: /tmp/sqoop-acadgild/compile/ded622aba0a1bcbd6ae555b03580c5f4/Person.jar
18/07/29 15:02:27 WARN manager.MySQLManager: It looks like you are importing from mysql.
18/07/29 15:02:27 WARN manager.MySQLManager: This transfer can be faster! Use the --direct
18/07/29 15:02:27 WARN manager.MySQLManager: option to exercise a MySQL-specific fast path.
18/07/29 15:02:27 INFO manager.MySQLManager: Setting zero DATETIME behavior to convertToNull (mysql)
18/07/29 15:02:27 INFO mapreduce.ImportJobBase: Beginning import of Person
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/common/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/home/acadgild/install/hbase/hbase-1.2.6/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
18/07/29 15:02:28 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/07/29 15:02:28 INFO Configuration.deprecation: mapred.jar is deprecated. Instead, use mapreduce.job.jar
```

## Assignment 6.1

### DATA INGESTION TOOL SQOOP

Below image shows that data present at /sqoopout2 folder with Person\_id=3

```
[acadgild@localhost ~]$ hadoop fs -ls /sqoopout2
18/07/29 15:03:26 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 2 items
-rw-r--r-- 1 acadgild supergroup          0 2018-07-29 15:03 /sqoopout2/_SUCCESS
-rw-r--r-- 1 acadgild supergroup        34 2018-07-29 15:03 /sqoopout2/part-m-00000
[acadgild@localhost ~]$ hadoop fs -cat /sqoopout2/part-m-00000
18/07/29 15:03:52 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
3,Rizvi,Syed,Whitefield,Bangalore
```