Name : Manav Pahilwani	Class/Roll No. : D16AD/ 37	Grade :

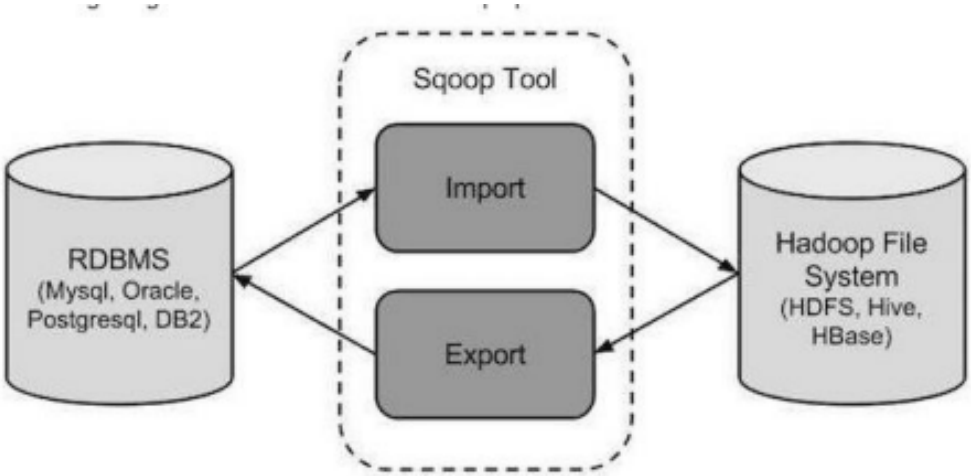
Title of Experiment : Use Sqoop to load data from RDBMS (weblog/ transactions data) and analyze it using HIVE/PIG.

Theory:

Sqoop – "SQL to Hadoop and Hadoop to SQL"

Sqoop is a tool designed to transfer data between Hadoop and relational database servers. It is used to import data from relational databases such as MySQL, Oracle to Hadoop HDFS, and export from Hadoop file system to relational databases. It is provided by the Apache Software Foundation.

How Sqoop works:



Enter MySql cmd prompt:

```
[cloudera@quickstart ~]$ mysql -u root -p
Enter password:
Welcome to the MySQL monitor.  Commands end with ; or \g.
Your MySQL connection id is 756
Server version: 5.1.73 Source distribution

Copyright (c) 2000, 2013, Oracle and/or its affiliates. All rights reserved.

Oracle is a registered trademark of Oracle Corporation and/or its
affiliates. Other names may be trademarks of their respective
owners.

Type 'help;' or '\h' for help. Type '\c' to clear the current input statement.
```

Create a database sales and change to that database.

```
mysql> create database sales;
Query OK, 1 row affected (0.00 sec)

mysql> use sales;
Database changed
```

Create a table sales in the sales database:

Command -

Create table sales(month_number int(10) not null primary key, facecream int(5), facewash int(5), toothpaste int(5), bathingsoap int(5), shampoo int(5), moisturizer int(5), total_units int(10), total_profit int(10));

```
mysql> Create table sales(month_number int(10) not null primary key, facecream i
nt(5), facewash int(5), toothpaste int(5), bathingsoap int(5), shampoo int(5), m
oisturizer int(5), total_units int(10), total_profit int(10));
Query OK, 0 rows affected (0.02 sec)
```

Import values in the table:

We import values from a CSV file into the table using the following command:

LOAD Data Local infile '/home/cloudera/Desktop/sales.csv' into table sales fields terminated by ',' lines terminated by '\n';

```
mysql> LOAD Data Local infile '/home/cloudera/Desktop/sales.csv' into table sales fields
terminated by ',' lines terminated by '\n';
Query OK, 13 rows affected, 9 warnings (0.00 sec)
Records: 13  Deleted: 0  Skipped: 0  Warnings: 0
```

Check if values are inserted

```
mysql> select * from sales limit 5;
+--------------+-----------+----------+------------+-------------+---------+-------------+-------------+--------------+
| month_number | facecream | facewash | toothpaste | bathingsoap | shampoo | moisturizer | total_units | total_profit |
+--------------+-----------+----------+------------+-------------+---------+-------------+-------------+--------------+
|            0 |         0 |        0 |          0 |           0 |       0 |           0 |           0 |            0 |
|            1 |      2500 |     1500 |       5200 |        9200 |    1200 |        1500 |       21100 |       211000 |
|            2 |      2630 |     1200 |       5100 |        6100 |    2100 |        1200 |       18330 |       183300 |
|            3 |      2140 |     1340 |       4550 |        9550 |    3550 |        1340 |       22470 |       224700 |
|            4 |      3400 |     1130 |       5870 |        8870 |    1870 |        1130 |       22270 |       222700 |
+--------------+-----------+----------+------------+-------------+---------+-------------+-------------+--------------+
5 rows in set (0.00 sec)
```

List all the tables present in mysql database:

```
[cloudera@quickstart ~]$ sqoop list-tables --connect jdbc:mysql://localhost/sales --username root --password "cloudera"
Warning: /usr/lib/sqoop/../accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
23/10/15 08:38:39 INFO sqoop.Sqoop: Running Sqoop version: 1.4.6-cdh5.13.0
23/10/15 08:38:39 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.
23/10/15 08:38:39 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
sales
```

Import tables from RDMS to HDFS using Sqoop:

```
[cloudera@quickstart ~]$ sqoop import --connect jdbc:mysql://localhost/sales --username=root --password="cloudera" --table=sales --target-dir
=/sales/sales -incremental append --check-column month_number --fields-terminated-by='\t';
Warning: /usr/lib/sqoop/../accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
23/10/15 08:41:10 INFO sqoop.Sqoop: Running Sqoop version: 1.4.6-cdh5.13.0
23/10/15 08:41:10 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.
23/10/15 08:41:10 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
23/10/15 08:41:10 INFO tool.CodeGenTool: Beginning code generation
23/10/15 08:41:11 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `sales` AS t LIMIT 1
23/10/15 08:41:11 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `sales` AS t LIMIT 1
23/10/15 08:41:11 INFO orm.CompilationManager: HADOOP_MAPRED_HOME is /usr/lib/hadoop-mapreduce
Note: /tmp/sqoop-cloudera/compile/1bb201fb9b1ca12490e5cb0c303f9d7a/sales.java uses or overrides a deprecated API.
```

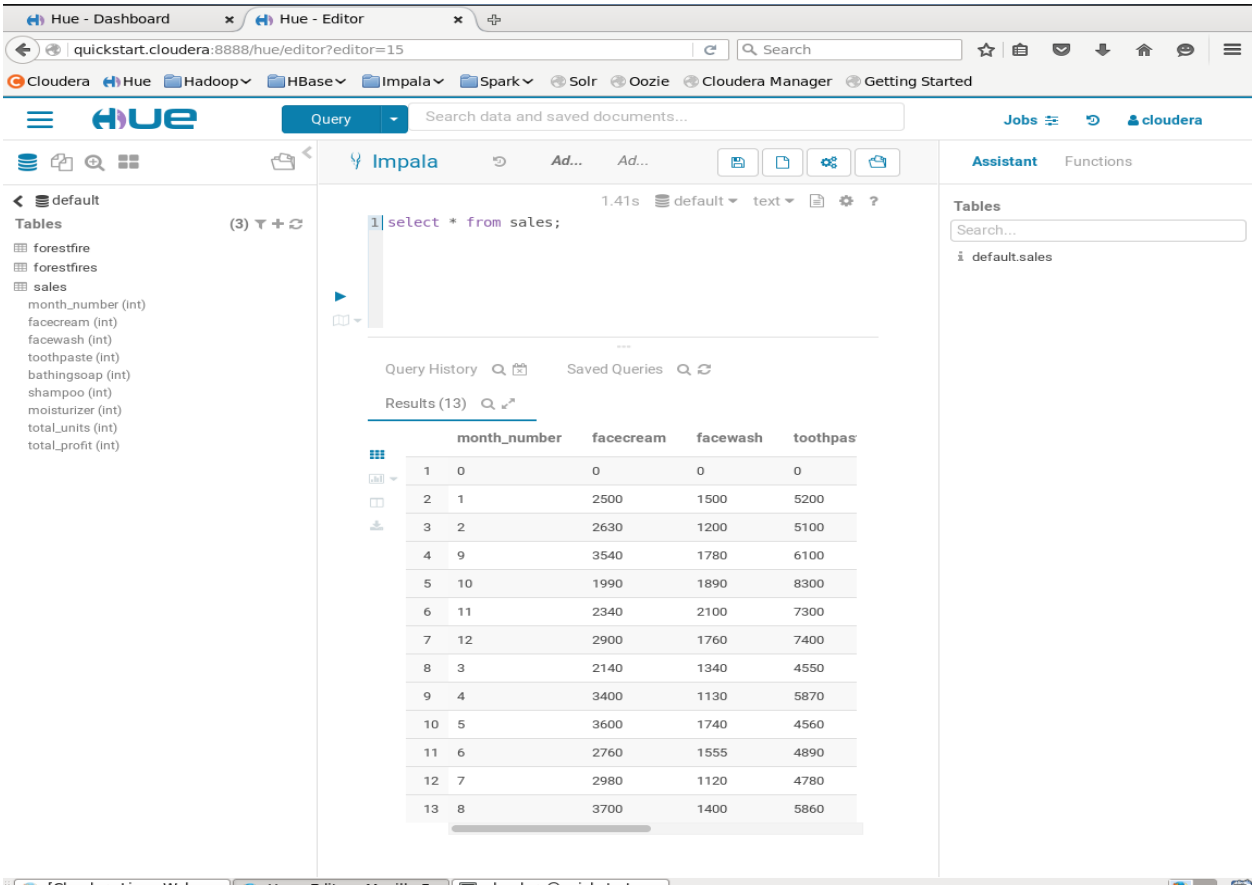
```
23/10/15 08:41:48 INFO mapreduce.ImportJobBase: Transferred 564 bytes in 33.5221 seconds (16.8247 bytes/sec)
23/10/15 08:41:48 INFO mapreduce.ImportJobBase: Retrieved 13 records.
23/10/15 08:41:48 INFO util.AppendUtils: Creating missing output directory - sales
23/10/15 08:41:48 INFO tool.ImportTool: Incremental import complete! To run another incremental import of all data following this import, sup
ply the following arguments:
23/10/15 08:41:48 INFO tool.ImportTool:  --incremental append
23/10/15 08:41:48 INFO tool.ImportTool:   --check-column month_number
23/10/15 08:41:48 INFO tool.ImportTool:   --last-value 12
23/10/15 08:41:48 INFO tool.ImportTool: (Consider saving this with 'sqoop job --create')
```

Import tables from HDFS to Hive

```
[cloudera@quickstart ~]$ sqoop import-all-tables --connect jdbc:mysql://localhost/sales --username root --password "cloudera" --compression-c
odec=snappy --as-parquetfile --warehouse-dir=/user/hive/warehouse --hive-import
Warning: /usr/lib/sqoop/../accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
23/10/15 08:46:22 INFO sqoop.Sqoop: Running Sqoop version: 1.4.6-cdh5.13.0
23/10/15 08:46:22 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.
23/10/15 08:46:22 INFO tool.BaseSqoopTool: Using Hive-specific delimiters for output. You can override
        Map-Reduce Framework
                Map input records=13
                Map output records=13
                Input split bytes=474
                Spilled Records=0
                Failed Shuffles=0
                Merged Map outputs=0
                GC time elapsed (ms)=3572
                CPU time spent (ms)=9620
                Physical memory (bytes) snapshot=596922368
                Virtual memory (bytes) snapshot=6105870336
                Total committed heap usage (bytes)=243531776
        File Input Format Counters
                Bytes Read=0
        File Output Format Counters
                Bytes Written=0
23/10/15 08:47:31 INFO mapreduce.ImportJobBase: Transferred 20.7383 KB in 61.3796 seconds (345.9781 bytes/sec)
23/10/15 08:47:31 INFO mapreduce.ImportJobBase: Retrieved 13 records.
```

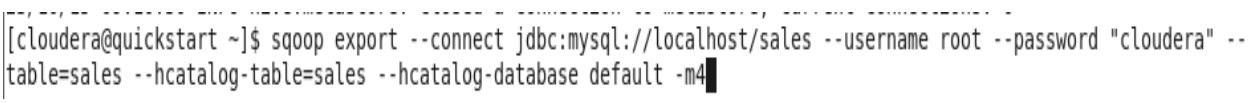
```
[cloudera@quickstart ~]$ hive

Logging initialized using configuration in file:/etc/hive/conf.dist/hive-log4j.properties
WARNING: Hive CLI is deprecated and migration to Beeline is recommended.
hive> Show Tables;
OK
forestfire
forestfires
sales
Time taken: 0.452 seconds, Fetched: 3 row(s)
```



	month_number	facecream	facewash	toothpas
1	0	0	0	0
2	1	2500	1500	5200
3	2	2630	1200	5100
4	9	3540	1780	6100
5	10	1990	1890	8300
6	11	2340	2100	7300
7	12	2900	1760	7400
8	3	2140	1340	4550
9	4	3400	1130	5870
10	5	3600	1740	4560
11	6	2760	1555	4890
12	7	2980	1120	4780
13	8	3700	1400	5860

Export data from Hive to MySQL



```
23/10/15 09:10:00 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1697375854682_0010
23/10/15 09:10:01 INFO impl.YarnClientImpl: Submitted application application_1697375854682_0010
23/10/15 09:10:01 INFO mapreduce.Job: The url to track the job: http://quickstart.cloudera:8088/proxy/application_169
7375854682_0010/
23/10/15 09:10:01 INFO mapreduce.Job: Running job: job_1697375854682_0010
23/10/15 09:10:11 INFO mapreduce.Job: Job job_1697375854682_0010 running in uber mode : false
23/10/15 09:10:11 INFO mapreduce.Job:  map 0% reduce 0%
23/10/15 09:10:29 INFO mapreduce.Job:  map 100% reduce 0%
```

Check in RDBMS:

```
mysql> select * from sales where month_number=13;
+--------------+-----------+----------+------------+-------------+---------+-------------+-------------+-------------
-+
| month_number | facecream | facewash | toothpaste | bathingsoap | shampoo | moisturizer | total_units | total_profit
 |
+--------------+-----------+----------+------------+-------------+---------+-------------+-------------+-------------
-+
|           13 |      1000 |     1000 |       1000 |        1000 |    1000 |        1000 |        6000 |       120000
 |
+--------------+-----------+----------+------------+-------------+---------+-------------+-------------+-------------
-+
1 row in set (0.00 sec)
```

Results and Discussions :

Successfully, created table in MySql and imported it to Hdfs and then to Hive, made changes and exported the data from Hive to MySql.