SQOOP COMMAND

sqoop job --create Patients_Contact_Info -- import \
--connect jdbc:mysql://upgraddetest.cyaielc9bmnf.us-east-1.rds.amazonaws.com/testdatabase \
--username student \
--password STUDENT123 \
--table patients_information \
--hive-import \
--create-hive-table \
--hive-table Capstone_Project.Patients_Contact_Info

<u>Sqoop Job List</u>
sqoop job --list

<u>Sqoop Execute Command</u>
sqoop job --exec Patients_Contact_Info

SNAPSHOT OF DATA COLLECTED

root@ip-10-0-0-135:~

```
(base) [root@ip-10-0-0-135 ~]# sqoop job --create Patients_Contact_Info -- import \
> --connect jdbc:mysql://upgradtest1.cyaielc9bmnf.us-east-1.rds.amazonaws.com/testdatabase \
> --username student \
> --password STUDENT123 \
> --table patients_information \
> --hive-import \
> --create-hive-table \
> --hive-table Capstone_Project.Patients_Contact_Info
21/04/04 13:50:18 INFO sqoop.Sqoop: Running Sqoop version: 1.4.6-cdh5.15.1
21/04/04 13:50:18 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.
21/04/04 13:50:18 INFO tool.BaseSqoopTool: Using Hive-specific delimiters for output. You can override
21/04/04 13:50:18 INFO tool.BaseSqoopTool: delimiters with --fields-terminated-by, etc.
(base) [root@ip-10-0-0-135 ~]# sqoop job --list
21/04/04 13:51:00 INFO sqoop.Sqoop: Running Sqoop version: 1.4.6-cdh5.15.1
Available jobs:
  Patients_Contact_Info
```

```
21/04/05 03:07:58 INFO mapreduce.Job:  map 80% reduce 0%
21/04/05 03:07:59 INFO mapreduce.Job:  map 100% reduce 0%
21/04/05 03:08:00 INFO mapreduce.Job: Job job_1617591216787_0001 completed successfully
21/04/05 03:08:00 INFO mapreduce.Job: Counters: 30
        File System Counters
                FILE: Number of bytes read=0
                FILE: Number of bytes written=884930
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=561
                HDFS: Number of bytes written=230
                HDFS: Number of read operations=20
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=10
        Job Counters
                Launched map tasks=5
                Other local map tasks=5
                Total time spent by all maps in occupied slots (ms)=44431
                Total time spent by all reduces in occupied slots (ms)=0
                Total time spent by all map tasks (ms)=44431
                Total vcore-milliseconds taken by all map tasks=44431
                Total megabyte-milliseconds taken by all map tasks=45497344
        Map-Reduce Framework
                Map input records=5
                Map output records=5
                Input split bytes=561
                Spilled Records=0
                Failed Shuffles=0
                Merged Map outputs=0
                GC time elapsed (ms)=1066
                CPU time spent (ms)=15180
                Physical memory (bytes) snapshot=1381687296
                Virtual memory (bytes) snapshot=14023221248
                Total committed heap usage (bytes)=1503133696
        File Input Format Counters
                Bytes Read=0
        File Output Format Counters
                Bytes Written=230
21/04/05 03:08:00 INFO mapreduce.ImportJobBase: Transferred 230 bytes in 44.8588 seconds (5.1272 bytes/sec)
21/04/05 03:08:00 INFO mapreduce.ImportJobBase: Retrieved 5 records.
21/04/05 03:08:00 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `patients_information` AS t LIMIT 1
21/04/05 03:08:00 INFO hive.HiveImport: Loading uploaded data into Hive

Logging initialized using configuration in jar:file:/opt/cloudera/parcels/CDH-5.15.1-1.cdh5.15.1.p0.4/jars/hive-common-1.1.0-cdh5.15.1.jar!/hive-log4j.properties
OK
Time taken: 3.455 seconds
Loading data to table capstone_project.patients_contact_info
Table capstone_project.patients_contact_info stats: [numFiles=5, totalSize=230]
OK
```

Check the data in Hive Table

```
hive> show databases;
OK
amz_review
capstone_project
credit_card
default
demo
demo2
demo3
events
Time taken: 2.014 seconds, Fetched: 8 row(s)
hive> use capstone_project;
OK
Time taken: 0.024 seconds
hive> select * from Patients_Contact_Info;
OK
1       Alex S  XDC test Address        8982739282      1       23      null
2       Sammy A New Building Address    2382739282      2       45      null
3       Karan C Aws Address     8923739282      3       56      null
4       Dara M  India Address   2182739282      4       67      null
5       Pam     ABC test Address        4982739282      5       72      null
Time taken: 0.894 seconds, Fetched: 5 row(s)
hive>
```