

## 11.2. Perform incremental load in Hive

Read from MySQL Table and load it in Hive table.

Create hive table if it does not exist.

If it exists, perform the incremental load.

=====

### INPUT COMMANDS TO IMPORT DATA FROM MYSQL INTO HIVE:

```
[cloudera@quickstart ~]$ sudo service mysqld start
[cloudera@quickstart ~]$ mysql -u root -pcloudera
mysql> create database mysqlpdb1 ;
mysql> use mysqlpdb1 ;
mysql> CREATE TABLE source_mysql_table
  ( id int not null primary key,
    name varchar(20),
    designation varchar(20),
    salary int
  );
mysql> SELECT * FROM source_mysql_table;
mysql> exit;
mysql> insert into source_mysql_table values(12300,'Mark','Architect',50000);
mysql> insert into source_mysql_table values(12301,'Nick','PM',40000);
mysql> insert into source_mysql_table values(1203,'Markus','Architect',55000);
mysql> insert into source_mysql_table values(1204,'Nicholas','PM',45000);
mysql> insert into source_mysql_table values(1205,'Arnica','Developer',36000);
mysql> select * from source_mysql_table;
mysql> exit;
[cloudera@quickstart ~]$ sudo hive
hive> create database hivedb1;
hive> exit;
```

=====

### CREATION OF NEW HIVE TABLE FOR THE FIRST IMPORT

```
[cloudera@quickstart ~]$ sqoop import --connect jdbc:mysql://localhost/mysqlpdb1 \
--username 'root' -P \
--table 'source_mysql_table' \
--fields-terminated-by ',' \
--target-dir '/sqoopoutputnew1' \
--hive-import \
--create-hive-table \
--hive-table hivedb1.destination_hive_table ;
```

**Input MYSQL Table:**

```
mysql> select * from source_mysql_table;
```

id	name	designation	salary
12300	Mark	Architect	50000
12301	Nick	PM	40000
12302	Arthur	Developer	30000
12303	Markus	Architect	55000
12304	Nicholas	PM	45000
12305	Arnica	Developer	36000

```
6 rows in set (0.00 sec)
```

```
mysql> █
```

```
[cloudera@quickstart ~]$ sqoop import --connect jdbc:mysql://localhost/mysql --username 'root' -P --table 'source_mysql_table' \
> --fields-terminated-by ',' \
> --target-dir '/sqoopoutputnew1' --hive-import --create-hive-table --hive-table hivedb1.destination_hive_table ;
```

### Output:

```
17/10/19 11:28:37 INFO mapreduce.ImportJobBase: Transferred 158 bytes in 76.5115 seconds (2.065 bytes/sec)
17/10/19 11:28:37 INFO mapreduce.ImportJobBase: Retrieved 6 records.
17/10/19 11:28:37 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `source_mysql_table` AS t LIMIT 1
17/10/19 11:28:37 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: 4.946 seconds
Loading data to table hivedb1.destination_hive_table
Table hivedb1.destination_hive_table stats: [numFiles=4, totalSize=158]
OK
Time taken: 1.049 seconds
[cloudera@quickstart ~]$ hadoop fs -ls /user/hive/warehouse/hivedb1.destination_hive_table
ls: '/user/hive/warehouse/hivedb1.destination_hive_table': No such file or directory
[cloudera@quickstart ~]$ hadoop fs -ls /user/hive/warehouse/hivedb1.db/destination_hive_table
Found 4 items
-rwxrwxrwx 1 cloudera supergroup 47 2017-10-19 11:28 /user/hive/warehouse/hivedb1.db/destination_hive_table/part-m-00000
-rwxrwxrwx 1 cloudera supergroup 29 2017-10-19 11:28 /user/hive/warehouse/hivedb1.db/destination_hive_table/part-m-00001
-rwxrwxrwx 1 cloudera supergroup 29 2017-10-19 11:28 /user/hive/warehouse/hivedb1.db/destination_hive_table/part-m-00002
-rwxrwxrwx 1 cloudera supergroup 53 2017-10-19 11:28 /user/hive/warehouse/hivedb1.db/destination_hive_table/part-m-00003
[cloudera@quickstart ~]$ hadoop fs -cat /user/hive/warehouse/hivedb1.db/destination_hive_table/part-m-00000
12300,Mark,Architect,50000
12301,Nick,PM,40000
[cloudera@quickstart ~]$ hadoop fs -cat /user/hive/warehouse/hivedb1.db/destination_hive_table/part-m-00001
12302,Arthur,Developer,30000
[cloudera@quickstart ~]$ hadoop fs -cat /user/hive/warehouse/hivedb1.db/destination_hive_table/part-m-00002
12303,Markus,Architect,55000
[cloudera@quickstart ~]$ hadoop fs -cat /user/hive/warehouse/hivedb1.db/destination_hive_table/part-m-00003
12304,Nicholas,PM,45000
12305,Arnica,Developer,36000
[cloudera@quickstart ~]$ █
```

```
hive> select * from destination_hive_table;
OK
12300   Mark      Architect      50000
12301   Nick      PM          40000
12302   Arthur    Developer     30000
12303   Markus    Architect     55000
12304   Nicholas   PM          45000
12305   Arnica    Developer     36000
Time taken: 0.121 seconds, Fetched: 6 row(s)
hive> █
```

```
mysql> insert into source_mysql_table values(12306,'Monica','Architect',50100);
```

### Incremental Load

```
[cloudera@quickstart ~]$ sqoop import --connect jdbc:mysql://localhost/mysqlpdb1 \
--username 'root' -P \
--table 'source_mysql_table' \
--fields-terminated-by "," \
--m 1 \
--incremental append \
--check-column id \
--last-value 12305 \
--target-dir '/user/hive/warehouse/hivedb1.db/destination_hive_table';
```

### Input MYSQL Table:

```
mysql> insert into source_mysql_table values(12306,'Monica','Architect',50100);
Query OK, 1 row affected (0.01 sec)
```

```
mysql> select * from source_mysql_table;
+-----+-----+-----+-----+
| id    | name    | designation | salary |
+-----+-----+-----+-----+
| 12300 | Mark    | Architect   | 50000  |
| 12301 | Nick    | PM          | 40000  |
| 12302 | Arthur  | Developer   | 30000  |
| 12303 | Markus  | Architect   | 55000  |
| 12304 | Nicholas| PM          | 45000  |
| 12305 | Arnica  | Developer   | 36000  |
| 12306 | Monica  | Architect   | 50100  |
+-----+-----+-----+-----+
7 rows in set (0.00 sec)
```

```
mysql> █
```

```
[cloudera@quickstart ~]$ sqoop import --connect jdbc:mysql://localhost/mysqlpdb1 --username 'root' -P --table 'source_mysql_table' \
> --fields-terminated-by "," --m 1 --incremental append --check-column id --last-value 12305 \
> --target-dir '/user/hive/warehouse/hivedb1.db/destination_hive_table';
```

## Output:

```
17/10/19 11:37:36 INFO mapreduce.ImportJobBase: Transferred 29 bytes in 29.7917 seconds (0.9734 bytes/sec)
17/10/19 11:37:36 INFO mapreduce.ImportJobBase: Retrieved 1 records.
17/10/19 11:37:36 INFO util.AppendUtils: Appending to directory destination_hive_table
17/10/19 11:37:36 INFO util.AppendUtils: Using found partition 4
17/10/19 11:37:36 INFO tool.ImportTool: Incremental import complete! To run another incremental import of all data following this im
17/10/19 11:37:36 INFO tool.ImportTool: --incremental append
17/10/19 11:37:36 INFO tool.ImportTool: --check-column id
17/10/19 11:37:36 INFO tool.ImportTool: --last-value 12306
17/10/19 11:37:36 INFO tool.ImportTool: (Consider saving this with 'sqoop job --create')
[cloudera@quickstart ~]$ hadoop fs -ls /user/hive/warehouse/hivedb1.db/destination_hive_table
Found 5 items
-rwxrwxrwx 1 cloudera supergroup 47 2017-10-19 11:28 /user/hive/warehouse/hivedb1.db/destination_hive_table/part-m-00000
-rwxrwxrwx 1 cloudera supergroup 29 2017-10-19 11:28 /user/hive/warehouse/hivedb1.db/destination_hive_table/part-m-00001
-rwxrwxrwx 1 cloudera supergroup 29 2017-10-19 11:28 /user/hive/warehouse/hivedb1.db/destination_hive_table/part-m-00002
-rwxrwxrwx 1 cloudera supergroup 53 2017-10-19 11:28 /user/hive/warehouse/hivedb1.db/destination_hive_table/part-m-00003
-rw-r--r-- 1 cloudera cloudera 29 2017-10-19 11:37 /user/hive/warehouse/hivedb1.db/destination_hive_table/part-m-00004
[cloudera@quickstart ~]$ hadoop fs -cat /user/hive/warehouse/hivedb1.db/destination_hive_table/part-m-00000
12300,Mark,Architect,50000
12301,Nick,PM,40000
[cloudera@quickstart ~]$ hadoop fs -cat /user/hive/warehouse/hivedb1.db/destination_hive_table/part-m-00001
12302,Arthur,Developer,30000
[cloudera@quickstart ~]$ hadoop fs -cat /user/hive/warehouse/hivedb1.db/destination_hive_table/part-m-00002
12303,Markus,Architect,55000
[cloudera@quickstart ~]$ hadoop fs -cat /user/hive/warehouse/hivedb1.db/destination_hive_table/part-m-00003
12304,Nicholas,PM,45000
12305,Arnica,Developer,36000
[cloudera@quickstart ~]$ hadoop fs -cat /user/hive/warehouse/hivedb1.db/destination_hive_table/part-m-00004
12306,Monica,Architect,50100
[cloudera@quickstart ~]$ █
```

```
hive> select * from destination_hive_table;
```

OK

```
12300    Mark    Architect    50000
12301    Nick    PM        40000
12302    Arthur  Developer  30000
12303    Markus  Architect  55000
12304    Nicholas      PM        45000
12305    Arnica  Developer  36000
12306    Monica  Architect  50100
```

Time taken: 0.155 seconds, Fetched: 7 row(s)

```
hive> █
```

```
=====
```

```
mysql> insert into source_mysql_table values(12307,'Nancy','PM',90000);
```

```
mysql> insert into source_mysql_table values(12308,'Angeline','Developer',60000);
```

## Incremental Load

```
[cloudera@quickstart ~]$ sqoop import --connect jdbc:mysql://localhost/mysqlpdb1 \
--username 'root' -P \
--table 'source_mysql_table' \
--fields-terminated-by "," \
--m 1 \
```

```
--incremental append \
--check-column id \
--last-value 12306 \
--target-dir '/user/hive/warehouse/hivedb1.db/destination_hive_table';
```

### Input MYSQL Table:

```
mysql> insert into source_mysql_table values(12307,'Nancy','PM',90000);
Query OK, 1 row affected (0.00 sec)
```

```
mysql> insert into source_mysql_table values(12308,'Angeline','Developer',60000);
Query OK, 1 row affected (0.00 sec)
```

```
mysql> select * from source_mysql_table;
```

id	name	designation	salary
12300	Mark	Architect	50000
12301	Nick	PM	40000
12302	Arthur	Developer	30000
12303	Markus	Architect	55000
12304	Nicholas	PM	45000
12305	Arnica	Developer	36000
12306	Monica	Architect	50100
12307	Nancy	PM	90000
12308	Angeline	Developer	60000

```
9 rows in set (0.00 sec)
```

```
mysql> █
```

```
[cloudera@quickstart ~]$ sqoop import --connect jdbc:mysql://localhost/mysqlpdb1 --username 'root' -P --table 'source_mysql_table' \
> --fields-terminated-by "," --m 1 --incremental append --check-column id --last-value 12306 \
> --target-dir '/user/hive/warehouse/hivedb1.db/destination_hive_table';
```

### Output:

```
17/10/19 11:45:24 INFO mapreduce.ImportJobBase: Transferred 52 bytes in 26.623 seconds (1.9532 bytes/sec)
```

```
17/10/19 11:45:24 INFO mapreduce.ImportJobBase: Retrieved 2 records.
```

```
17/10/19 11:45:24 INFO util.AppendUtils: Appending to directory destination_hive_table
```

```
17/10/19 11:45:24 INFO util.AppendUtils: Using found partition 5
```

```
17/10/19 11:45:24 INFO tool.ImportTool: Incremental import complete! To run another incremental import of all data following this import, supply the follow
```

```
17/10/19 11:45:24 INFO tool.ImportTool: --incremental append
```

```
17/10/19 11:45:24 INFO tool.ImportTool: --check-column id
```

```
17/10/19 11:45:24 INFO tool.ImportTool: --last-value 12308
```

```
17/10/19 11:45:24 INFO tool.ImportTool: (Consider saving this with 'sqoop job --create')
```

```
[cloudera@quickstart ~]$ hadoop fs -cat /user/hive/warehouse/hivedb1.db/destination_hive_table/part-m-00005
```

```
12307,Nancy,PM,90000
```

```
12308,Angeline,Developer,60000
```

```
[cloudera@quickstart ~]$ █
```

```
hive> select * from destination_hive_table;
OK
12300  Mark    Architect    50000
12301  Nick     PM          40000
12302  Arthur   Developer    30000
12303  Markus   Architect    55000
12304  Nicholas      PM          45000
12305  Arnica   Developer    36000
12306  Monica   Architect    50100
12307  Nancy    PM           90000
12308  Angeline      Developer    60000
Time taken: 0.072 seconds, Fetched: 9 row(s)
hive> █
```

=====