

Problem Statement

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.

Solution:

--Creating table 'employee' in MYSQL

```
mysql> create table employee(e_id int,e_name varchar(20),e_unit varchar(20));
Query OK, 0 rows affected (0.01 sec)
```

--Inserting and show data in the table 'employee' created in MYSQL

```
mysql> insert into employee values (1,'Nitisha','CA');
Query OK, 1 row affected (0.01 sec)
```

```
mysql> insert into employee values (2,'Shika','Test Engineer');
Query OK, 1 row affected (0.00 sec)
```

```
mysql> insert into employee values (3,'Pranshu','Senior Engineer');
Query OK, 1 row affected (0.00 sec)
```

```
mysql> select * from employee;
```

e_id	e_name	e_unit
1	Nitisha	CA
2	Shika	Test Engineer
3	Pranshu	Senior Engineer

3 rows in set (0.00 sec)

--Importing data from table 'employee' in MYSQL to HIVE table 'employee_incre'

```
[cloudera@quickstart ~]$ sqoop import --connect jdbc:mysql://localhost/sqoop --username 'root' -P --table employee --split-by e_id --columns e_id,e_name --fields-terminated-by "," --target-dir /user/cloudera/employee --hive-import --create-hive-table --hive-table sqoop.employee_incre --m 1;
Warning: /usr/lib/sqoop/./accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
17/11/20 09:25:53 INFO sqoop.Sqoop: Running Sqoop version: 1.4.6-cdh5.12.0
Enter password:
17/11/20 09:26:00 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
17/11/20 09:26:00 INFO tool.CodeGenTool: Beginning code generation
17/11/20 09:26:04 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `employee` AS t LIMIT 1
17/11/20 09:26:04 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `employee` AS t LIMIT 1
17/11/20 09:26:04 INFO orm.CompilationManager: HADOOP MAPRED HOME is /usr/lib/hadoop-mapreduce
Physical memory (bytes) snapshot=136118272
Virtual memory (bytes) snapshot=1510133760
Total committed heap usage (bytes)=60751872
File Input Format Counters
  Bytes Read=0
File Output Format Counters
  Bytes Written=28
17/11/20 09:27:36 INFO mapreduce.ImportJobBase: Transferred 28 bytes in 71.8467 seconds (0.3897 bytes/sec)
17/11/20 09:27:36 INFO mapreduce.ImportJobBase: Retrieved 3 records.
17/11/20 09:27:37 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `employee` AS t LIMIT 1
17/11/20 09:27:37 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: 13.145 seconds
Loading data to table sqoop.employee_incre
Table sqoop.employee_incre stats: [numFiles=1, totalSize=28]
OK
Time taken: 2.898 seconds
[cloudera@quickstart ~]$
[cloudera@quickstart ~]$
```

--Checking data imported in table 'employee_incre' or not in hive

```
hive> select * from employee_incre;
OK
1      Nitisha
2      Shika
3      Pranshu
Time taken: 0.25 seconds, Fetched: 3 row(s)
```

--Inserting more data in 'employee' table created in MYSQL

```
mysql> insert into employe values (4,'Sonal','Engineer');
Query OK, 1 row affected (0.01 sec)
```

```
mysql> insert into employe values (5,'Rohit','CS');
Query OK, 1 row affected (0.00 sec)
```

```
mysql> insert into employe values (6,'Priti','Software Engineer');
Query OK, 1 row affected (0.00 sec)
```

```
mysql> select * from employe;
+-----+-----+-----+
| e_id | e_name | e_unit |
+-----+-----+-----+
| 1    | Nitisha | CA     |
| 2    | Shika  | Test Engineer |
| 3    | Pranshu | Senior Engineer |
| 4    | Sonal  | Engineer |
| 5    | Rohit  | CS     |
| 6    | Priti  | Software Engineer |
+-----+-----+-----+
6 rows in set (0.00 sec)
```

mysql> █

--Performing Incremental Load in hive using Sqoop

```
[cloudera@quickstart ~]$ sqoop import --connect jdbc:mysql://localhost/sqoop --username 'root' -P --table employe --hive-database sqoop --hive-table employee_incre --check-column e_id --last-value 3 --incremental append --fields-terminated-by "," --target-dir hdfs://quickstart.cloudera:8020/user/hive/warehouse/sqoop.db/employee_incre --m 1;
```

Warning: /usr/lib/sqoop/./accumulo does not exist! Accumulo imports will fail.

Please set \$ACCUMULO_HOME to the root of your Accumulo installation.

17/11/20 09:42:55 INFO sqoop.Sqoop: Running Sqoop version: 1.4.6-cdh5.12.0

Enter password:

17/11/20 09:43:00 WARN tool.BaseSqoopTool: It seems that you've specified at least one of following:

17/11/20 09:43:00 WARN tool.BaseSqoopTool: --hive-home

17/11/20 09:43:00 WARN tool.BaseSqoopTool: --hive-overwrite
Total committed heap usage (bytes)=60751872

File Input Format Counters

Bytes Read=0

File Output Format Counters

Bytes Written=54

17/11/20 09:44:24 INFO mapreduce.ImportJobBase: Transferred 54 bytes in 66.5599 seconds (0.8113 bytes/sec)

17/11/20 09:44:24 INFO mapreduce.ImportJobBase: Retrieved 3 records.

17/11/20 09:44:24 INFO util.AppendUtils: Appending to directory employee_incre

17/11/20 09:44:24 INFO util.AppendUtils: Using found partition 1

17/11/20 09:44:24 INFO tool.ImportTool: Incremental import complete! To run another incremental import of all data following this import, supply the following arguments:

17/11/20 09:44:24 INFO tool.ImportTool: --incremental append

17/11/20 09:44:24 INFO tool.ImportTool: --check-column e_id

17/11/20 09:44:24 INFO tool.ImportTool: --last-value 6

17/11/20 09:44:24 INFO tool.ImportTool: (Consider saving this with 'sqoop job --create')

[cloudera@quickstart ~]\$ █

--Checking whether the new data added in table 'employee_incre' or not in hive

```
hive> select * from employee_incre;
```

OK

1 Nitisha

2 Shika

3 Pranshu

4 Sonal

5 Rohit

6 Priti

Time taken: 0.416 seconds, Fetched: 6 row(s)

```
hive> █
```