● Explain and perform Sqoop Incremental Import operation to load data from Mysql To HDFS using Sqoop

**Solution:**

Sqoop supports two types of incremental imports: append and lastmodified. We can use the –incremental argument to specify the type of incremental import to perform.

We should specify the append mode when importing a table, where new rows are continually added with increasing row id values.

We must specify the column containing the row’s id with –check-column.

Sqoop imports rows where the check column has a value greater than the one specified with –last-value.

An alternate table update strategy supported by Sqoop is called lastmodified mode. This should be used when rows of the source table is updated, and each such update will set the value of a last-modified column to the current timestamp.

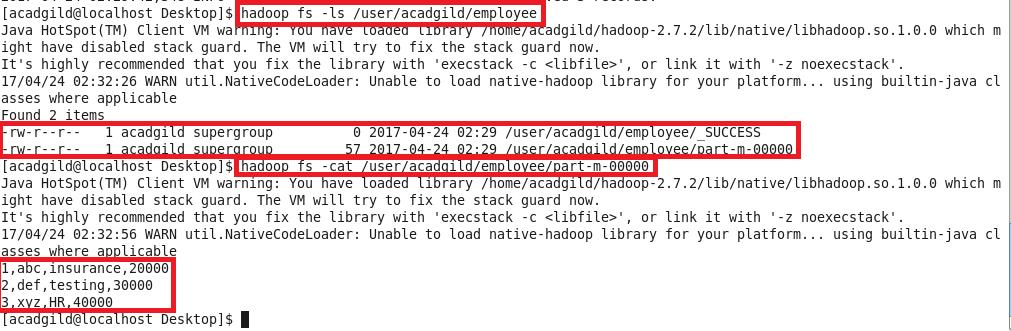
Rows where the check column holds a timestamp more recent than the timestamp specified with –last-value are imported.

At the end of an incremental import, the value which should be specified as –last-value for a subsequent import is printed to the screen.

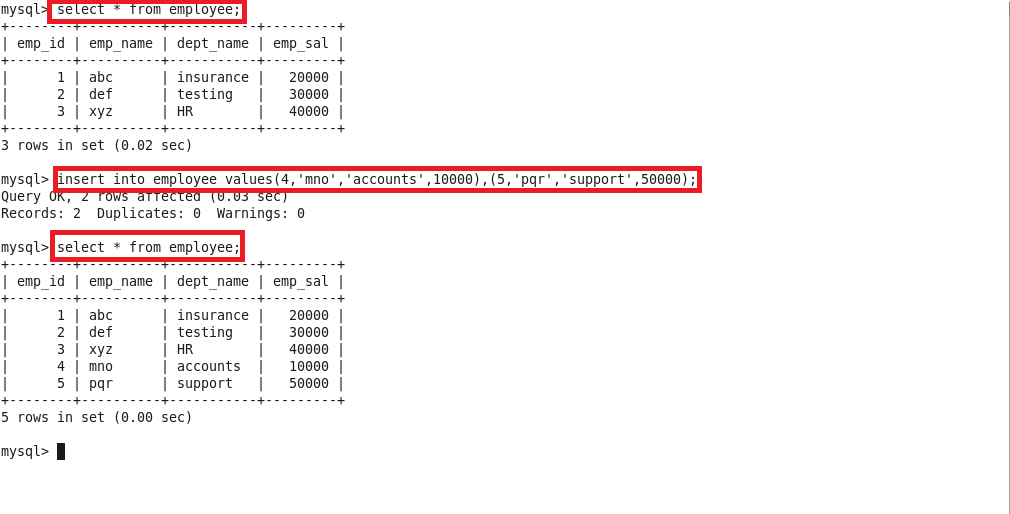
When running a subsequent import, we should specify –last-value in this way to ensure we import only the new or updated data. This is handled automatically by creating an incremental import as a saved job, which is the preferred mechanism for performing a recurring incremental import.

● Explain the procedures performed, Share the screenshots of commands and results for the same.

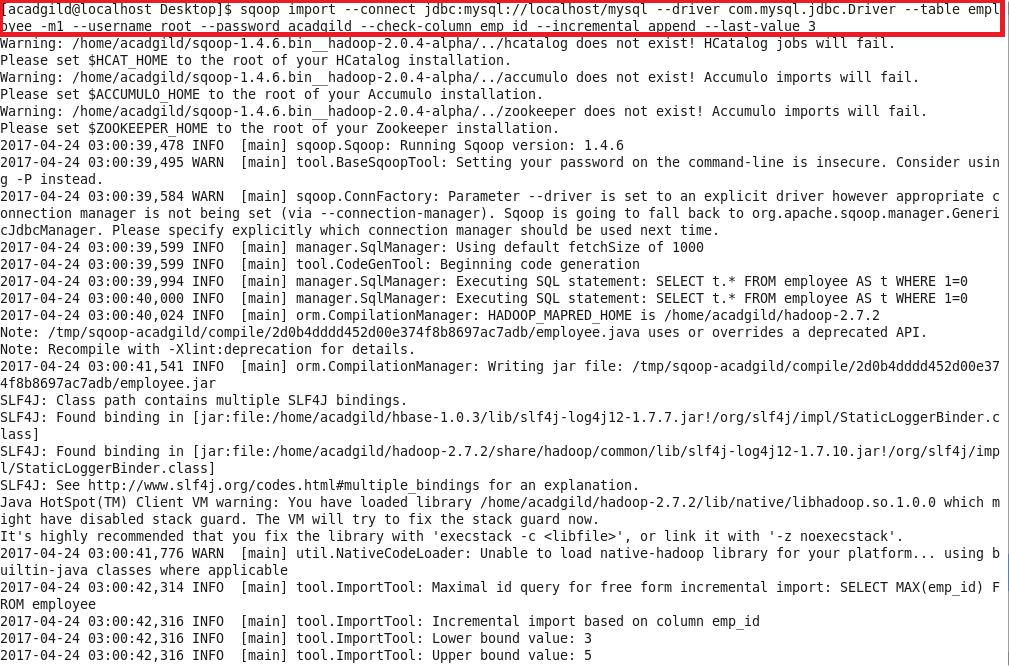
Displaying the contents of the file already present in hdfs

****

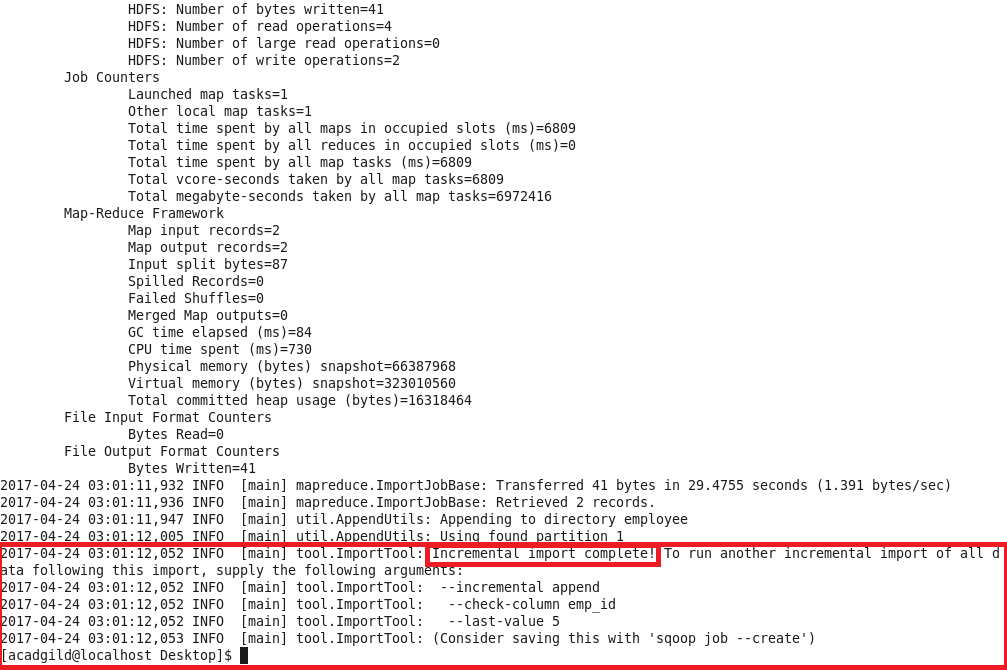
Displaying the contents of table , adding new records and displaying new entries of table



Sqoop incremental import command



Incremental import complete



Displaying the contents of new file created

