Problem Statement:

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

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

**Incremental Import**

Suppose we assume that we have already a imported a table from Mysql to HDFS and suppose if the data is again updated (either the data is changed or additional rows are added in mysql),In this case we can use Incremental Import so that only additional data or updated data is alone added instead of adding the full data again

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

**Append method**

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

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.

**Lastmodified** 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.

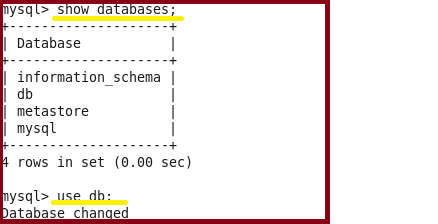
At the end of an incremental import, the value which should be specified as –last-value for a subsequent import is printed to the screen.

When running a subsequent import, you should specify –last-value in this way to ensure you import only the new or updated data.

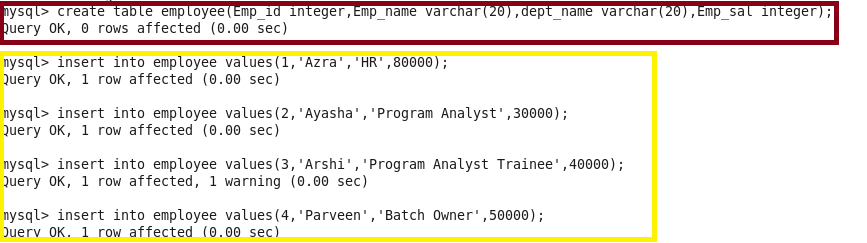
This is handled automatically by creating an incremental import as a saved job, which is the preferred mechanism for performing a recurring incremental import.

**To perform Sqoop Incremental Import operation to load data from Mysql To HDFS using Sqoop we need to create a table.**

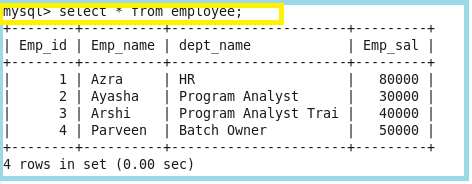
Using show databases we can check the databases available and we can use one of them using use command.



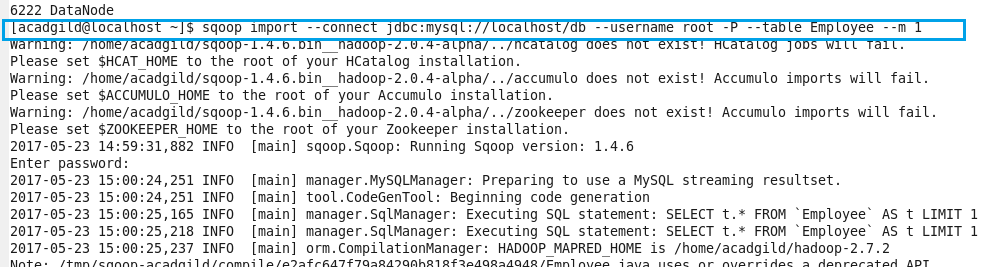
**Creating a table named Employee and inserting the values using insert command.**

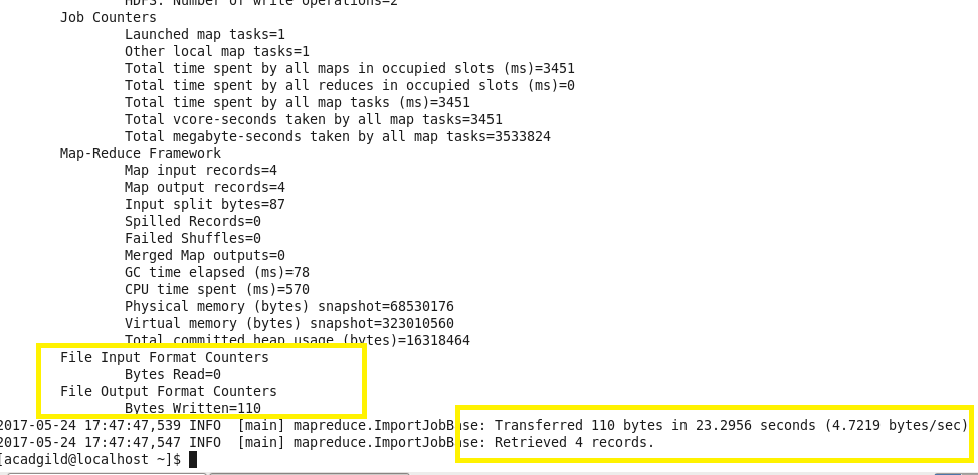


Performing basic select operation to display the contents of the table

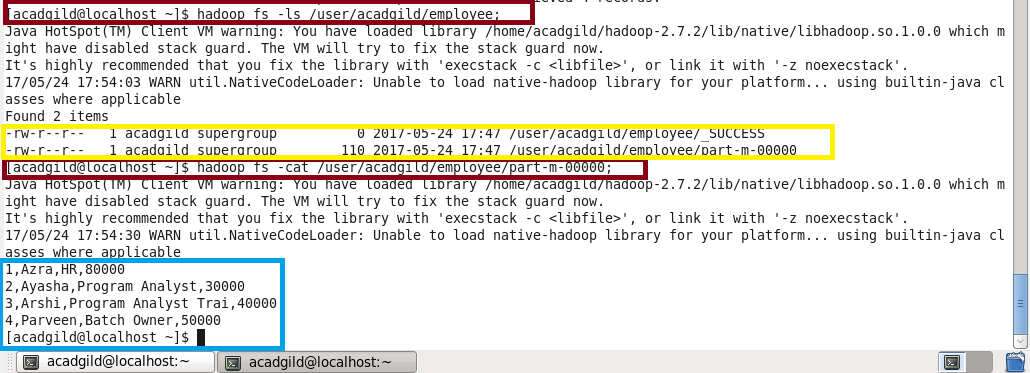


**Importing the data from mysql to hdfs using the below command**

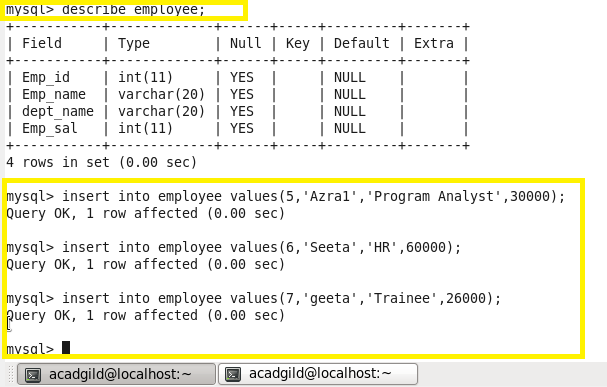




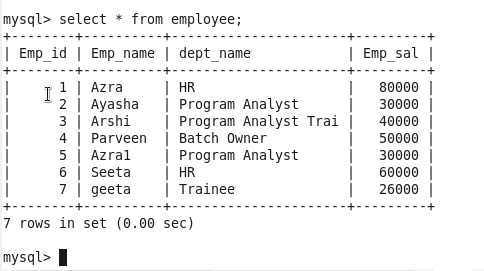
**Checking whether the files are imported and displaying the contents of the Employee**



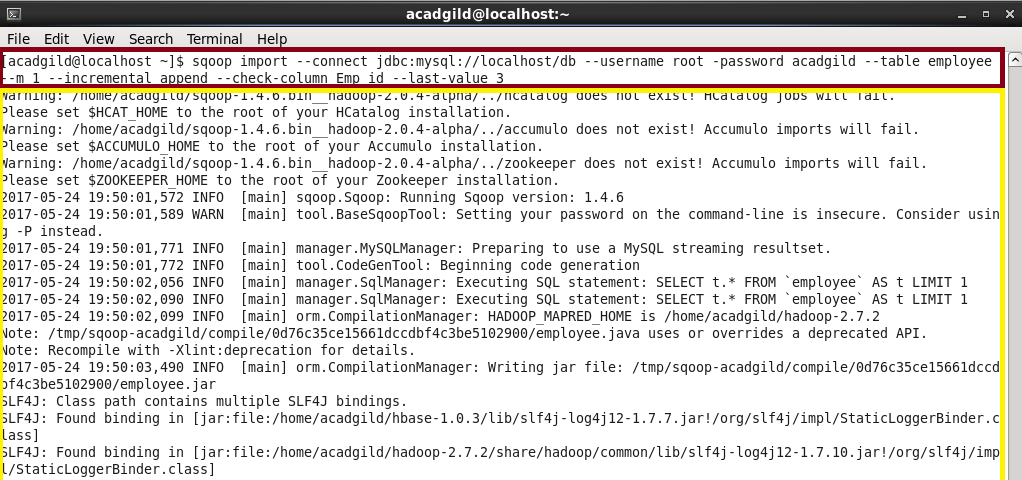
**Inserting 3 more records in the existing table**

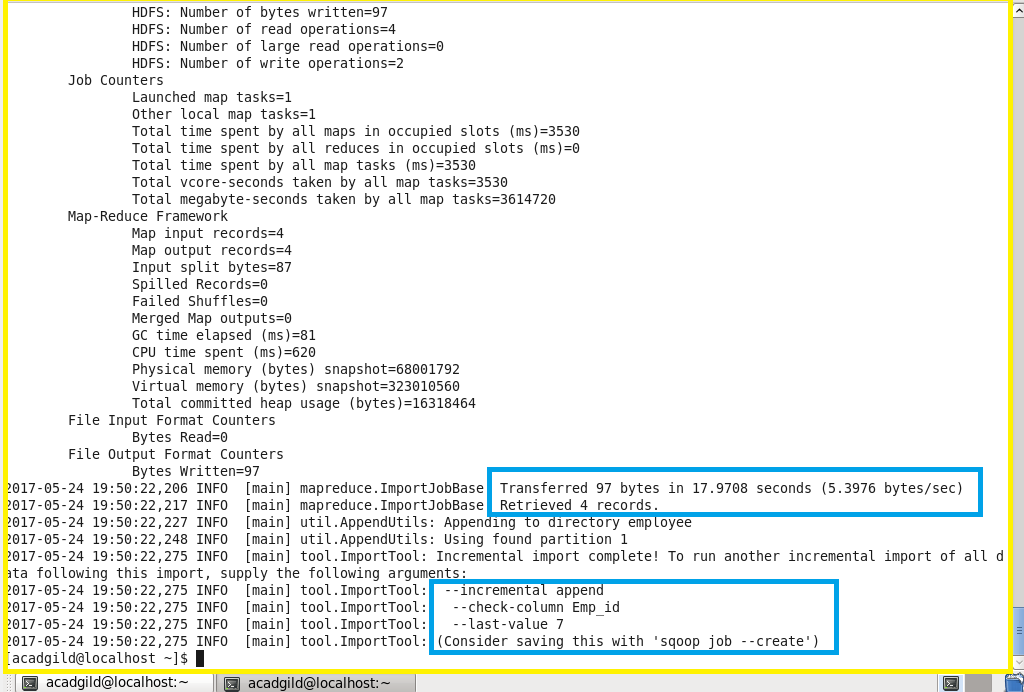


**Displaying all the contents using select command**



**Performing incremental append using below command**

****



**Sqoop import - import is used when we are importing data from rdbms to Hadoop ecosystem**

**connect - It is used to give the JDBC Url of database**

**Username - user name of database**

**password - password of database**

**table - give the table you want to copy from MySql**

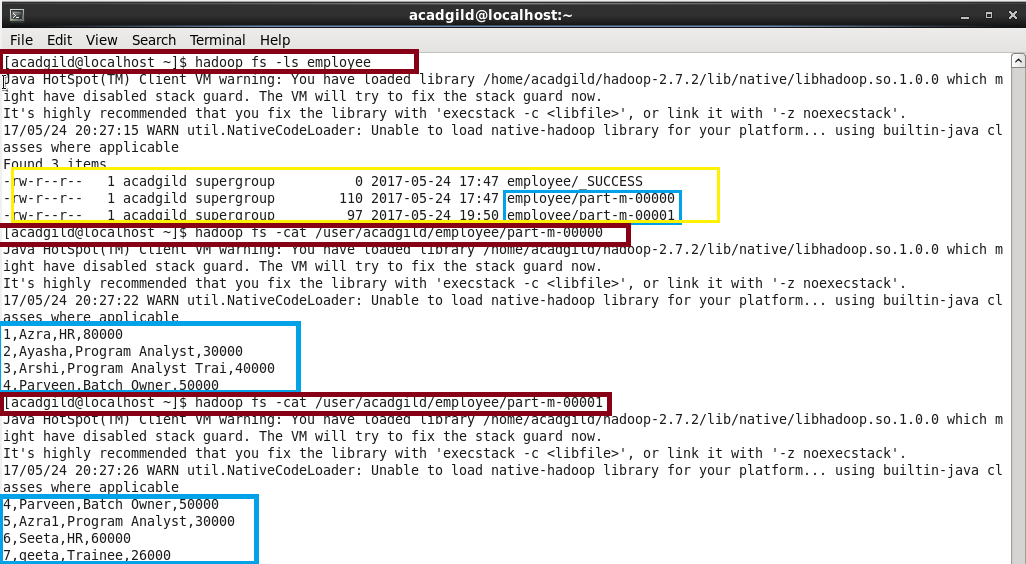
**target-dir - temporary target directory used by Sqoop Under which the data will be stored in HDFS**

**incremental append - tells sqoop the import is of increamental Import Type**

**check-column - column based on which data will be checked while importing**

**last-value - specifies the value after which data will be imported**

**Imported data is displayed from hdfs**

****

**Output**

