Creating Hive table using SEQUENCEFILE format and importing data

We will see how to create a table in Hive using SEQUENCEFILE format and how to import data into the table.

Create table

CREATE TABLE Employee(

ID BIGINT,

NAME STRING,

AGE INT,

SALARY BIGINT

)

COMMENT 'This is Employee table stored as sequencefile'

ROW FORMAT DELIMITED

FIELDS TERMINATED BY ','

STORED AS SEQUENCEFILE;

Load data into the table from Sequencefiles

To load data in the table directly from files, we need to have the input files as SEQUENCEFILEs.

**Input File :-**

On our HDFS, we have a sequencefile ‘/home/hadoop/data/employeeseq‘ containing employee data. We will load this data in our Employee table.

**Load Data from file:-**

#To overwrite the data in the table use -

LOAD DATA INPATH '/home/hadoop/data/employeeseq' OVERWRITE INTO TABLE Employee;

#To append the data in the table use -

LOAD DATA INPATH '/home/hadoop/data/employeeseq' INTO TABLE Employee;

Load data into the table using a select query

To load data into the table using a select query use the following:-

#Overwrite data from result of a select query into the table

INSERT OVERWRITE TABLE Employee SELECT id, name, age, salary from Employee\_text;

#Append data from result of a select query into the table

INSERT INTO TABLE Employee SELECT id, name, age, salary from Employee\_text;

## Load Data from Text or CSV files

If you have data in text or csv format, then first load that data in a temporary table, which stores data in TEXTFILE format, and then import data from the temporary table into this seq\_table using a select query.