Creating Hive table using ORC format and importing data

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

ORC format

ORC (Optimized Row Columnar) file format provides a highly efficient way to store Hive data. Using ORC format  improves performance when reading, writing, and processing data in Hive. We can specify compression to further compress data files. It could result in a small performance loss while writing, but there will be huge performance gain in reading. Compression available are SNAPPY, ZLIB, NONE.

Create table

We will see creation of Employee table in ORC format using compression as ZLIB.

CREATE TABLE Employee(

ID BIGINT,

NAME STRING,

AGE INT,

SALARY BIGINT

)

COMMENT 'This is Employee table in ORC file format'

ROW FORMAT DELIMITED

FIELDS TERMINATED BY ','

STORED AS ORC tblproperties ("orc.compress"="ZLIB");

Load data into the table from ORC files

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

**Input File :-**

On our HDFS, we have an ORC file  ‘/home/hadoop/data/employeeorc‘ 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/employeeorc' OVERWRITE INTO TABLE Employee;

#To append the data in the table use -

LOAD DATA INPATH '/home/hadoop/data/employeeorc' 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 table using a select query.