Hive Join operations

Create a table named CUSTOMERS(ID | NAME | AGE | ADDRESS | SALARY)

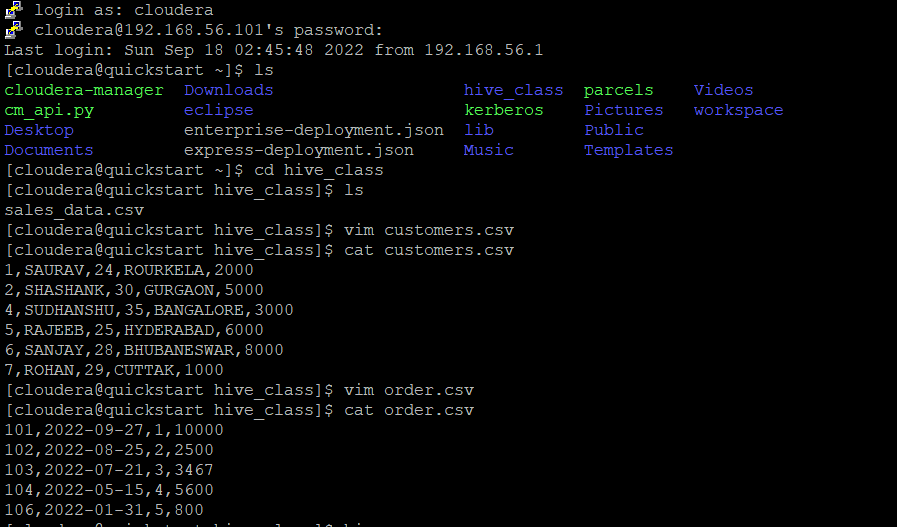
Create a Second table ORDER(OID | DATE | CUSTOMER\_ID | AMOUNT

)

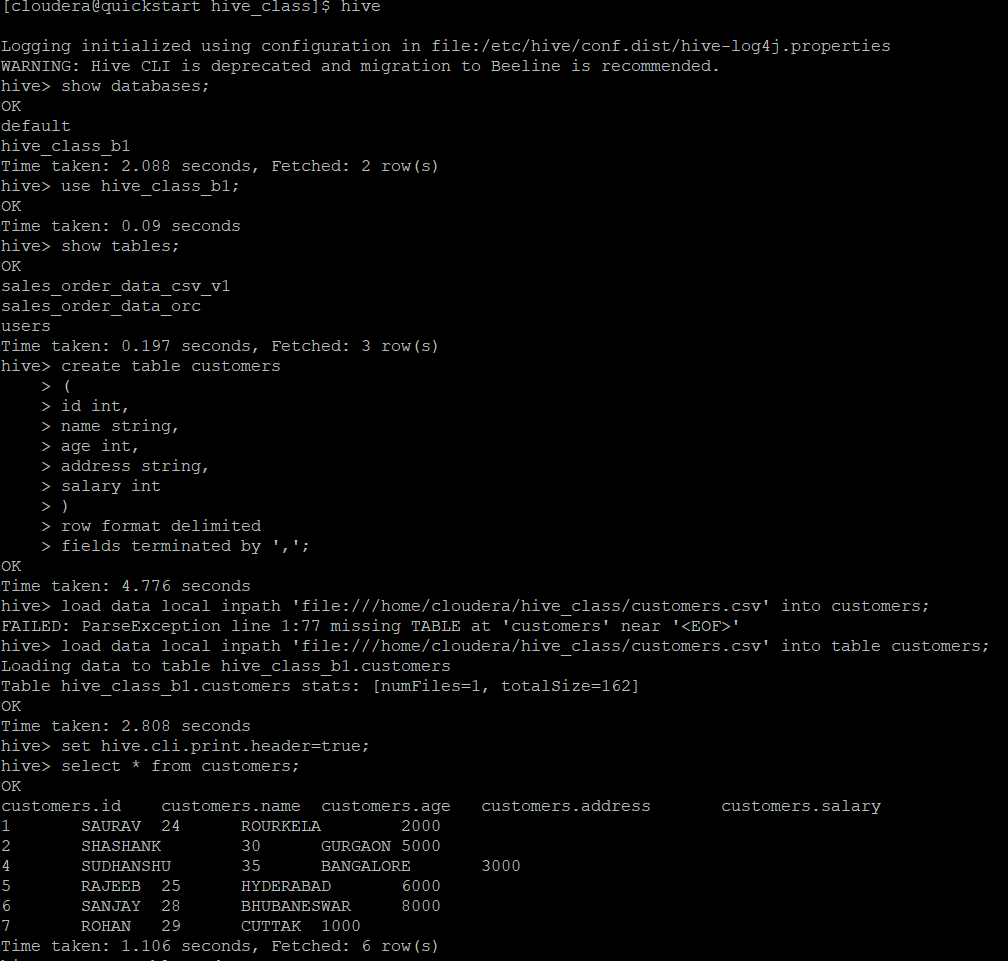
Now perform different joins operations on top of these tables

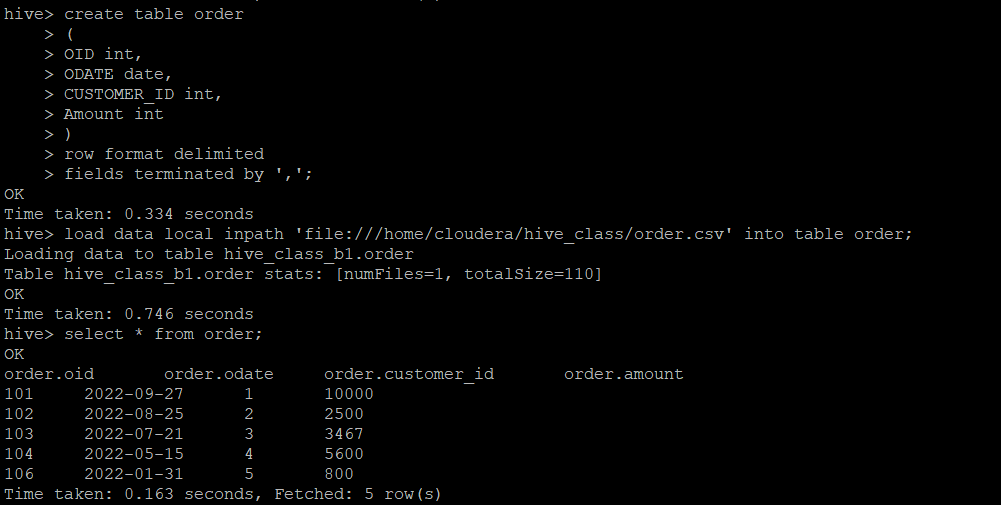
(Inner JOIN, LEFT OUTER JOIN ,RIGHT OUTER JOIN ,FULL OUTER JOIN)

1)Create csv files in local using vim

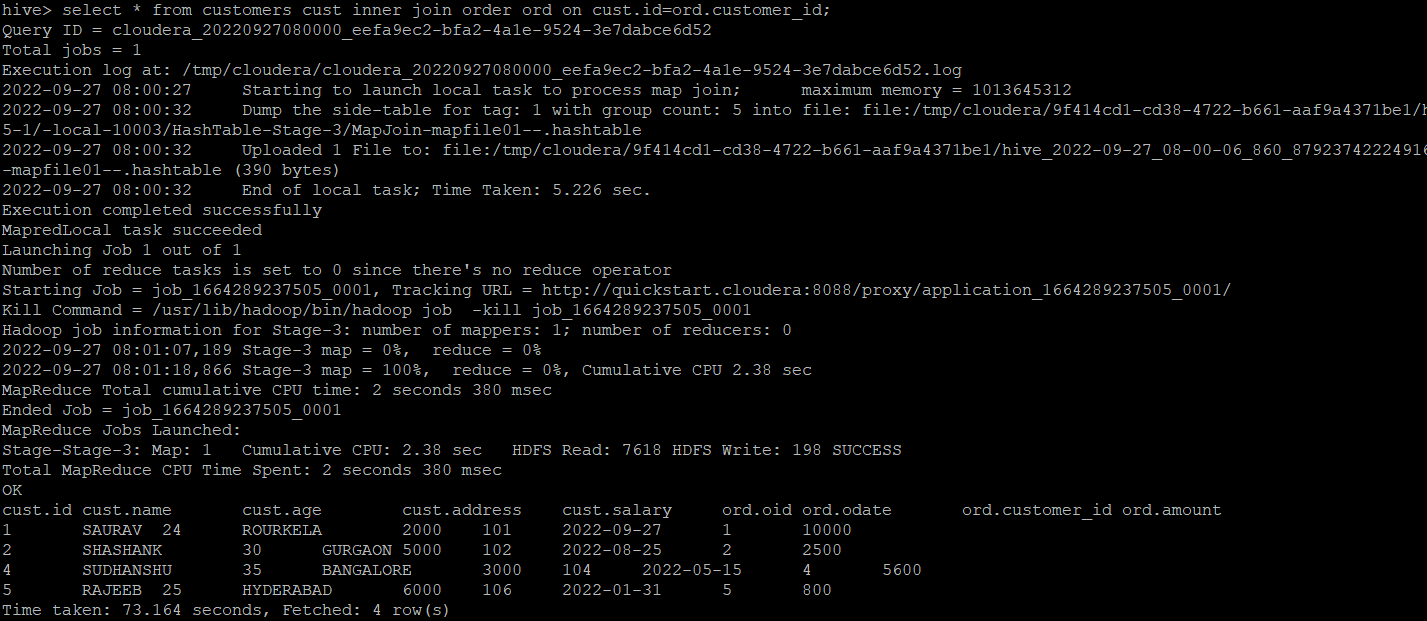


2)Create tables and load data from local

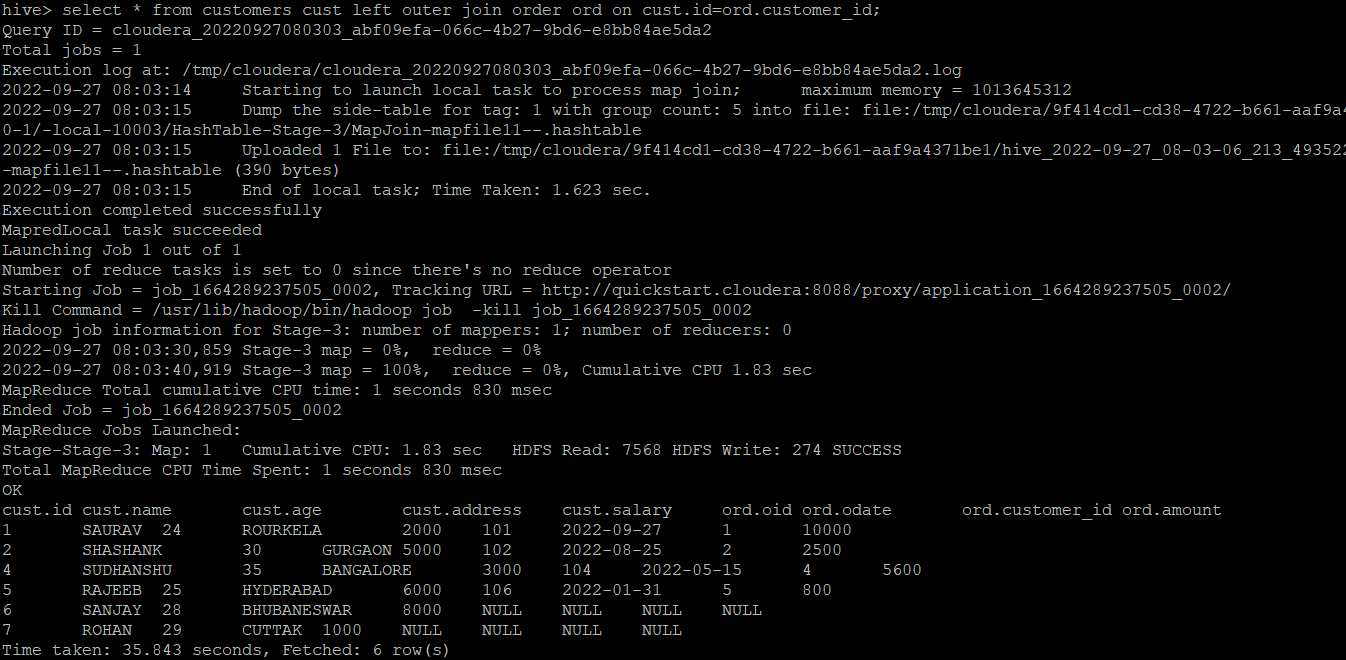




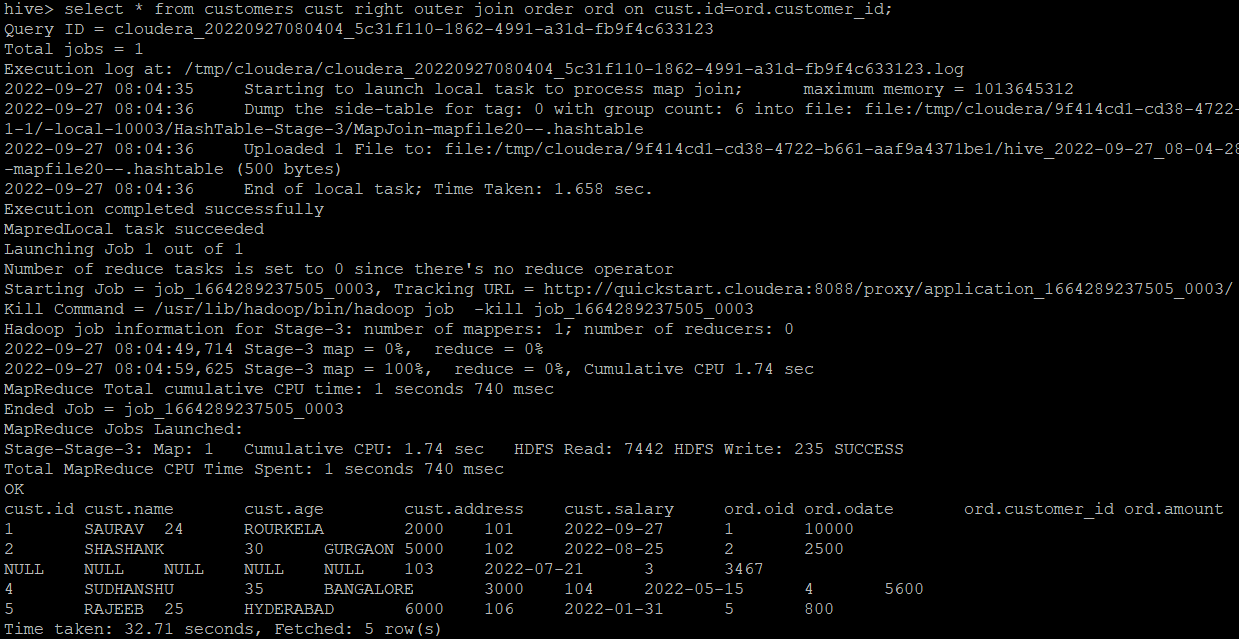
3)Inner join



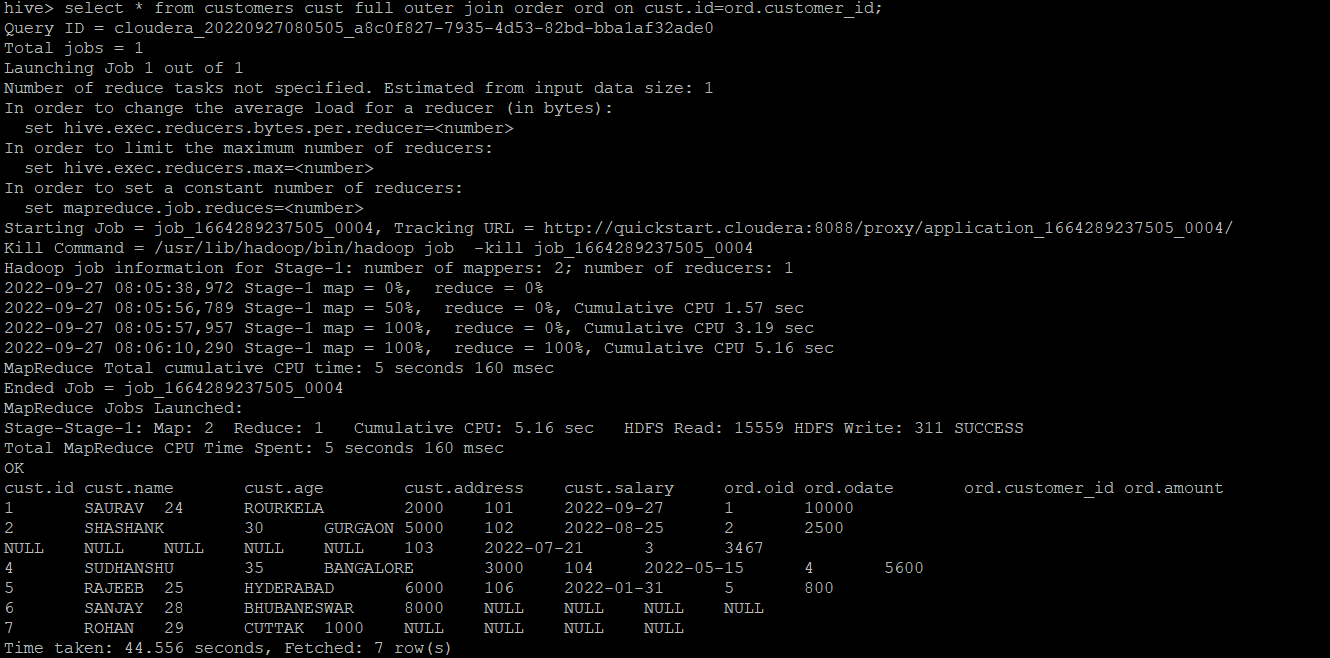
4)left outer join



5)right outer join



6)full outer join



BUILD A DATA PIPELINE WITH HIVE

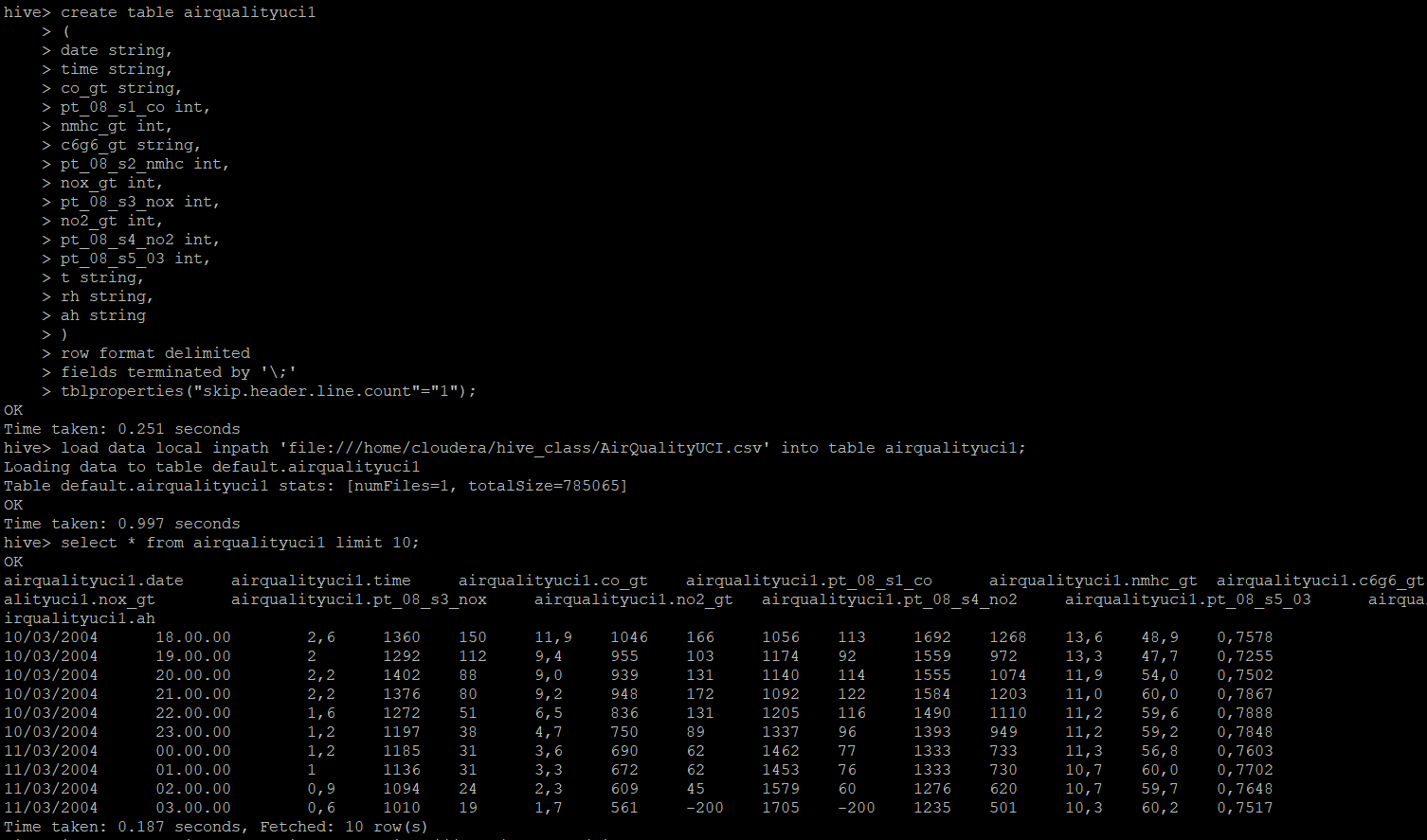
Download a data from the given location -

<https://archive.ics.uci.edu/ml/machine-learning-databases/00360/>

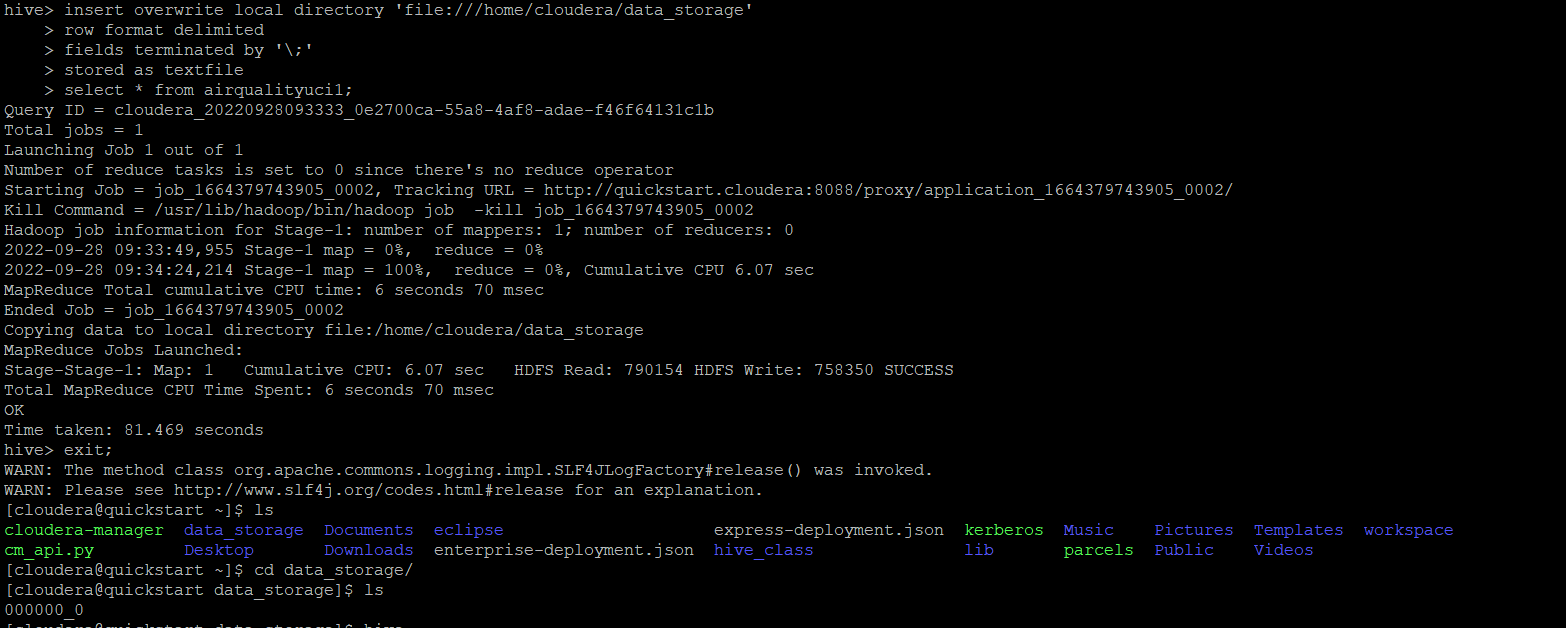
1. Create a hive table as per given schema in your dataset

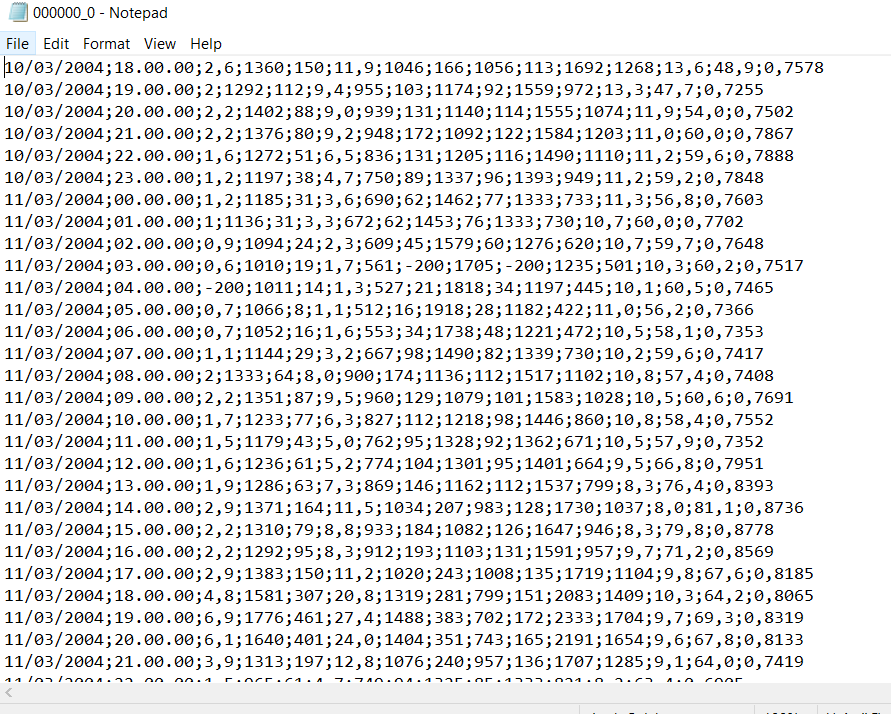
2. try to place a data into table location

3. Perform a select operation .

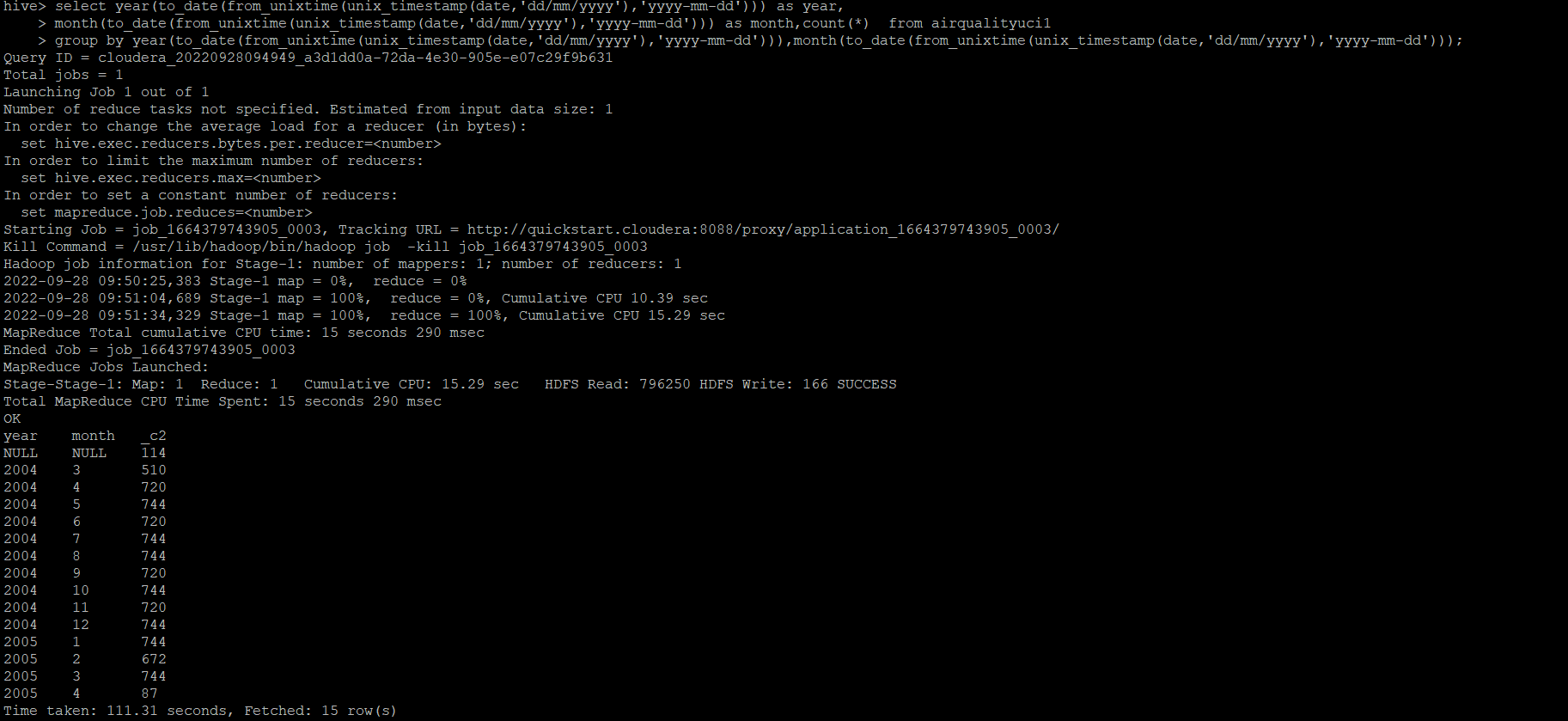


4. Fetch the result of the select operation in your local as a csv file .



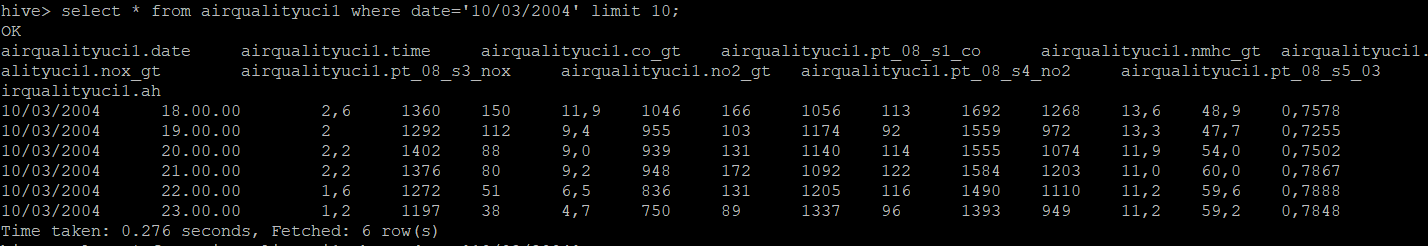


5. Perform group by operation .

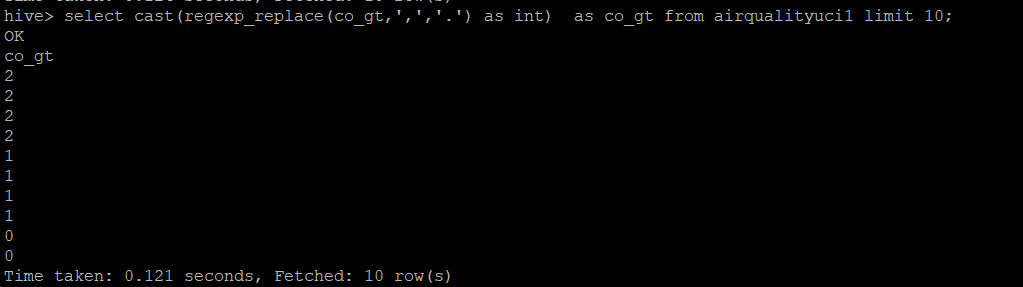


7. Perform filter operation at least 5 kinds of filter examples .

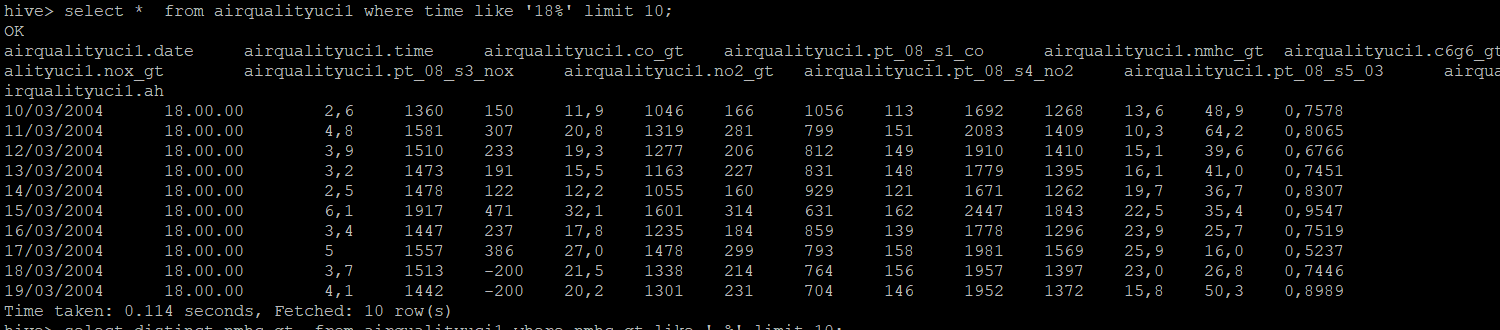
hive> select \* from airqualityuci1 where date='10/03/2004' limit 10;



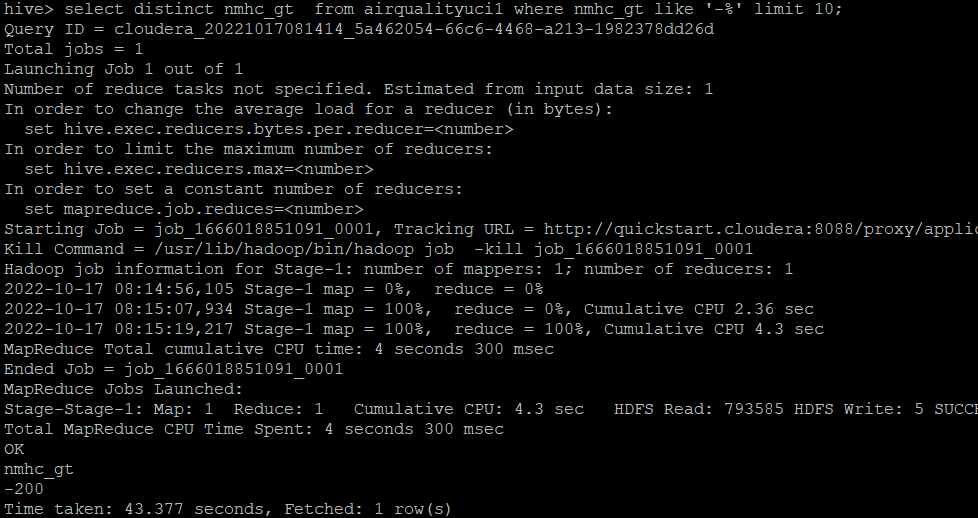
hive> select cast(regexp\_replace(co\_gt,',','.') as int) as co\_gt from airqualityuci1 limit 10;



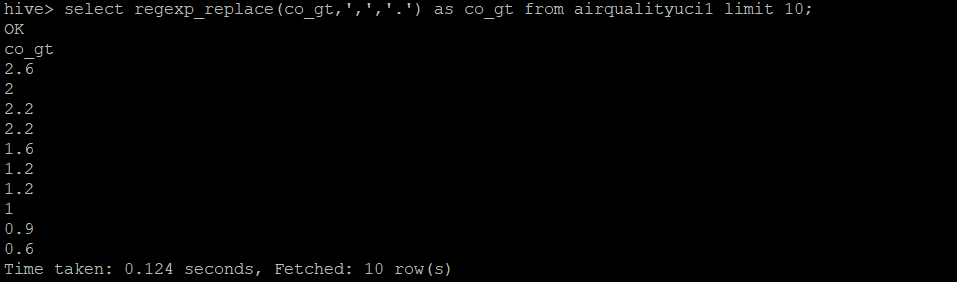
hive> select \* from airqualityuci1 where time like '18%' limit 10;



hive> select distinct nmhc\_gt from airqualityuci1 where nmhc\_gt like '-%' limit 10;

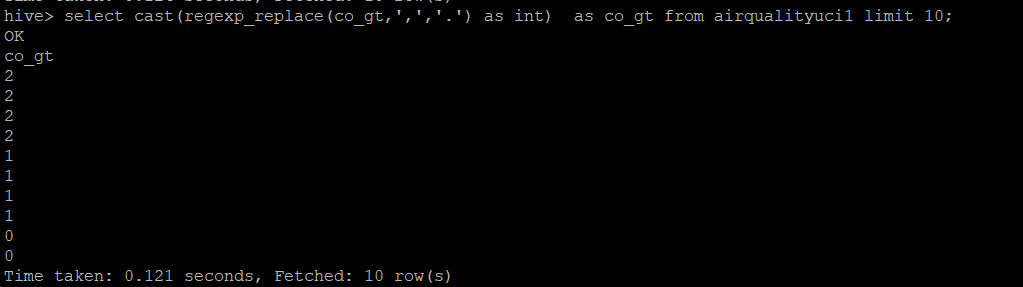


hive> select regexp\_replace(co\_gt,',','.') as co\_gt from airqualityuci1 limit 10;

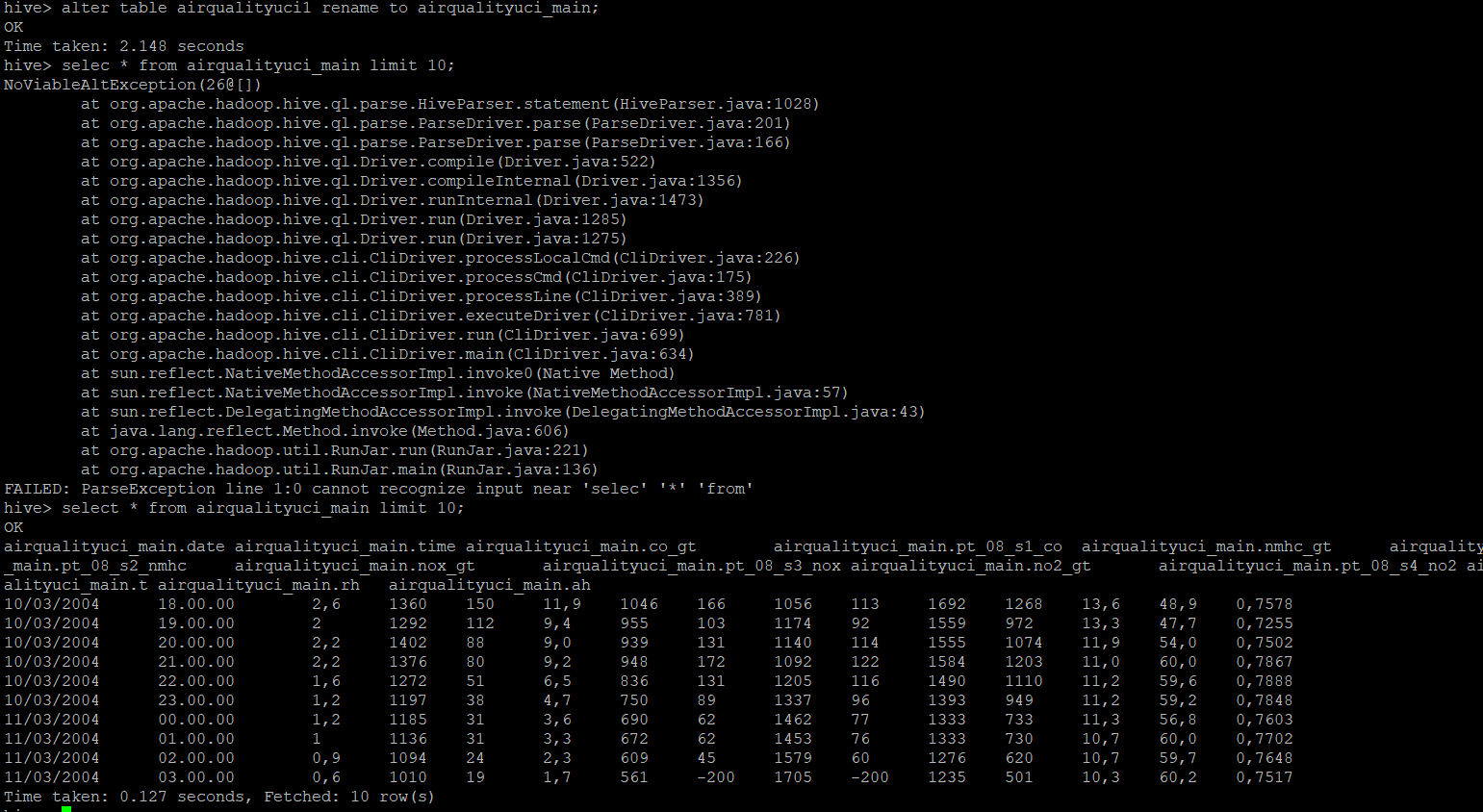


8. show and example of regex operation

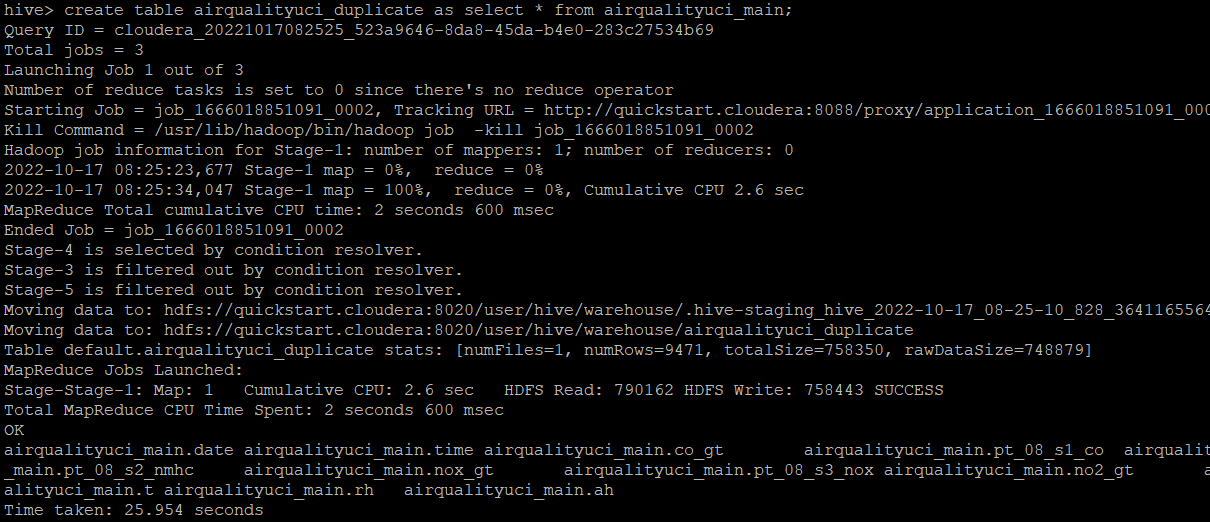
hive> select cast(regexp\_replace(co\_gt,',','.') as int) as co\_gt from airqualityuci1 limit 10;

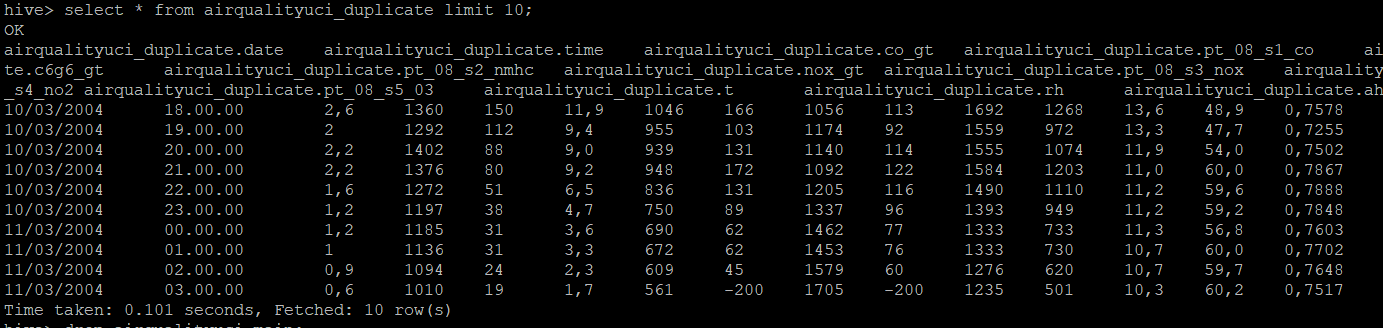


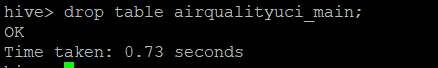
9. alter table operation



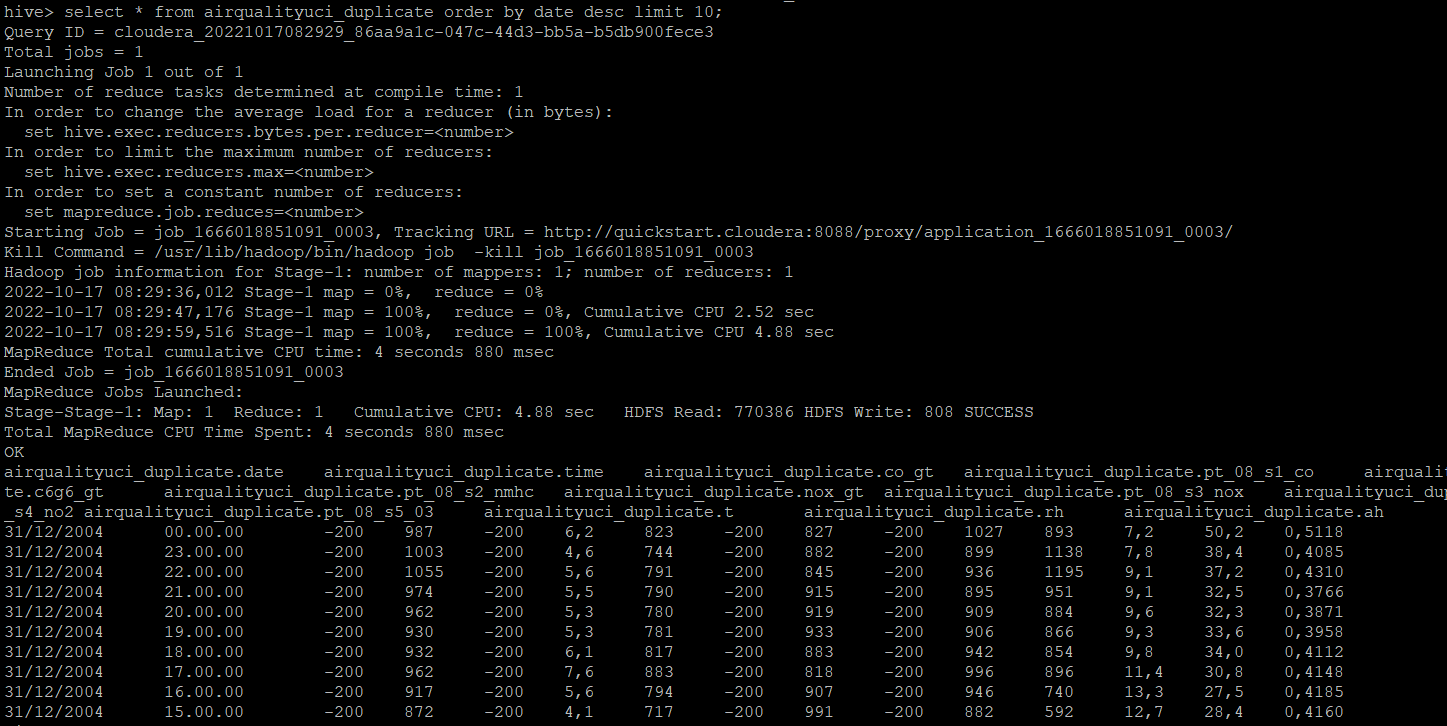
10 . drop table operation





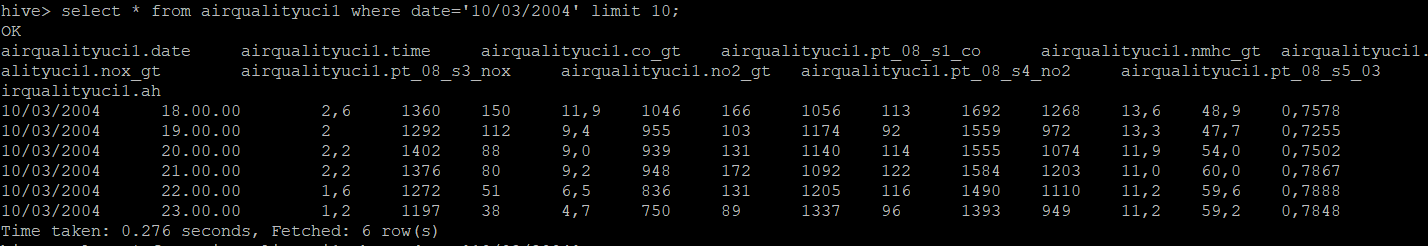


12 . order by operation .

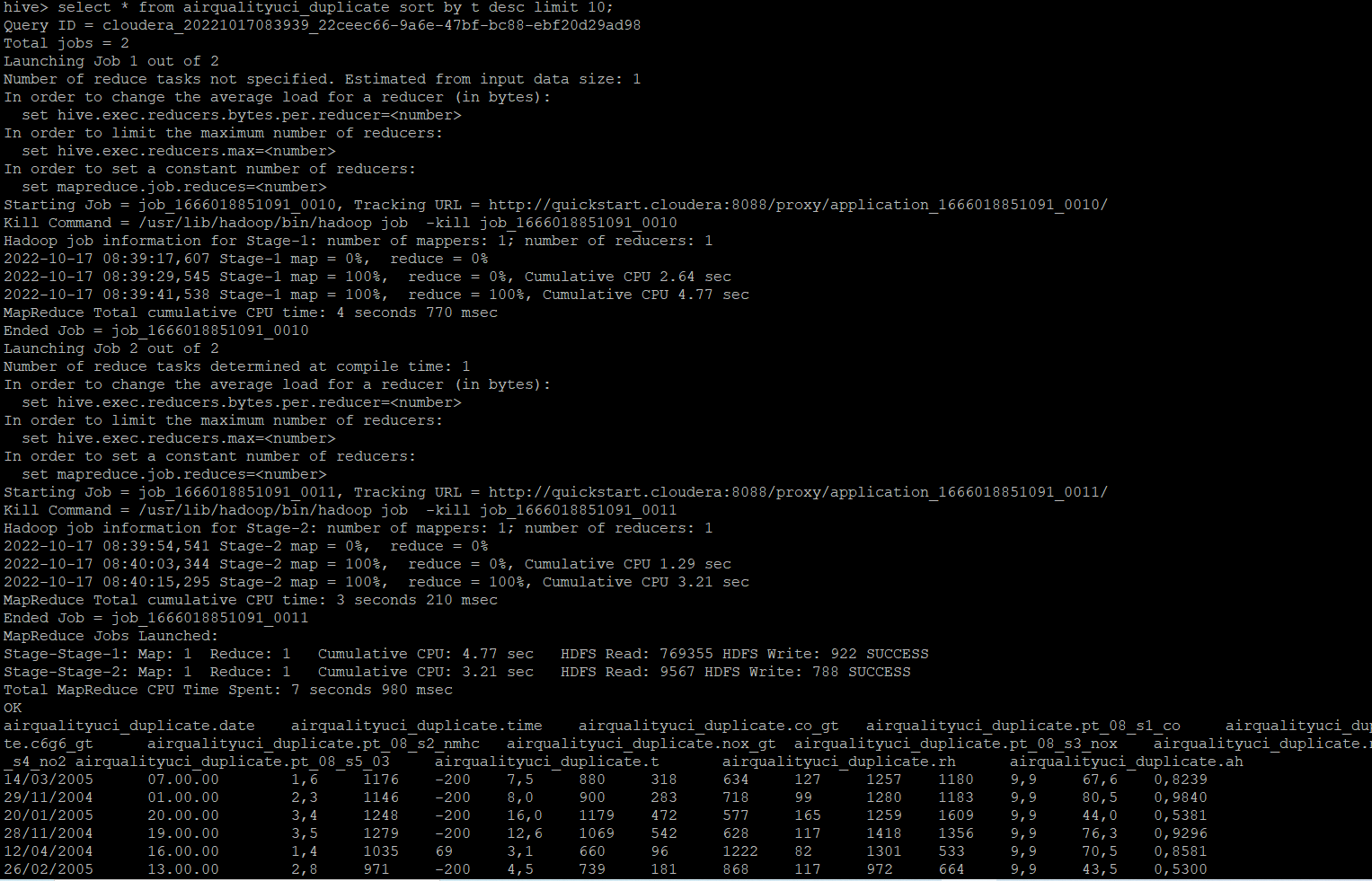


13 . where clause operations you have to perform .

select \* from airqualityuci1 where date='10/03/2004' limit 10;

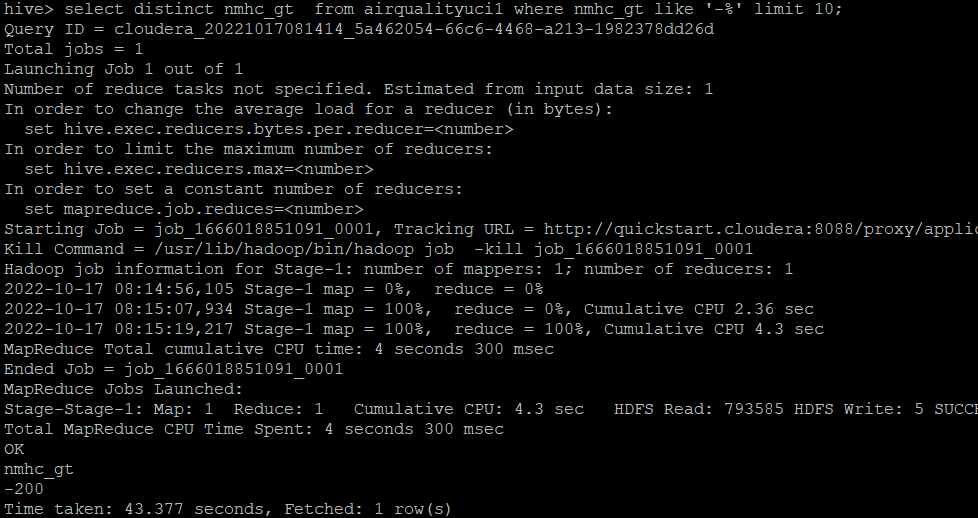


14 . sorting operation you have to perform .



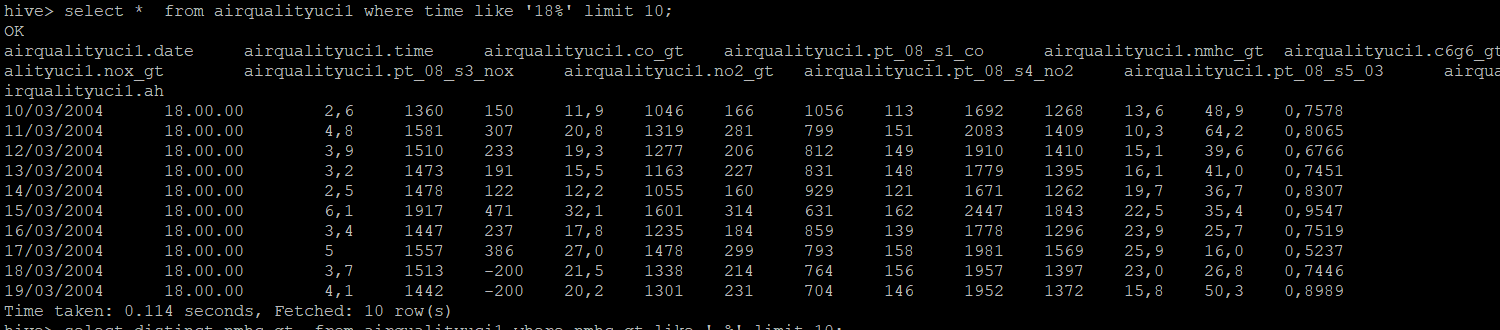
15 . distinct operation you have to perform .

hive> select distinct nmhc\_gt from airqualityuci1 where nmhc\_gt like '-%' limit 10;

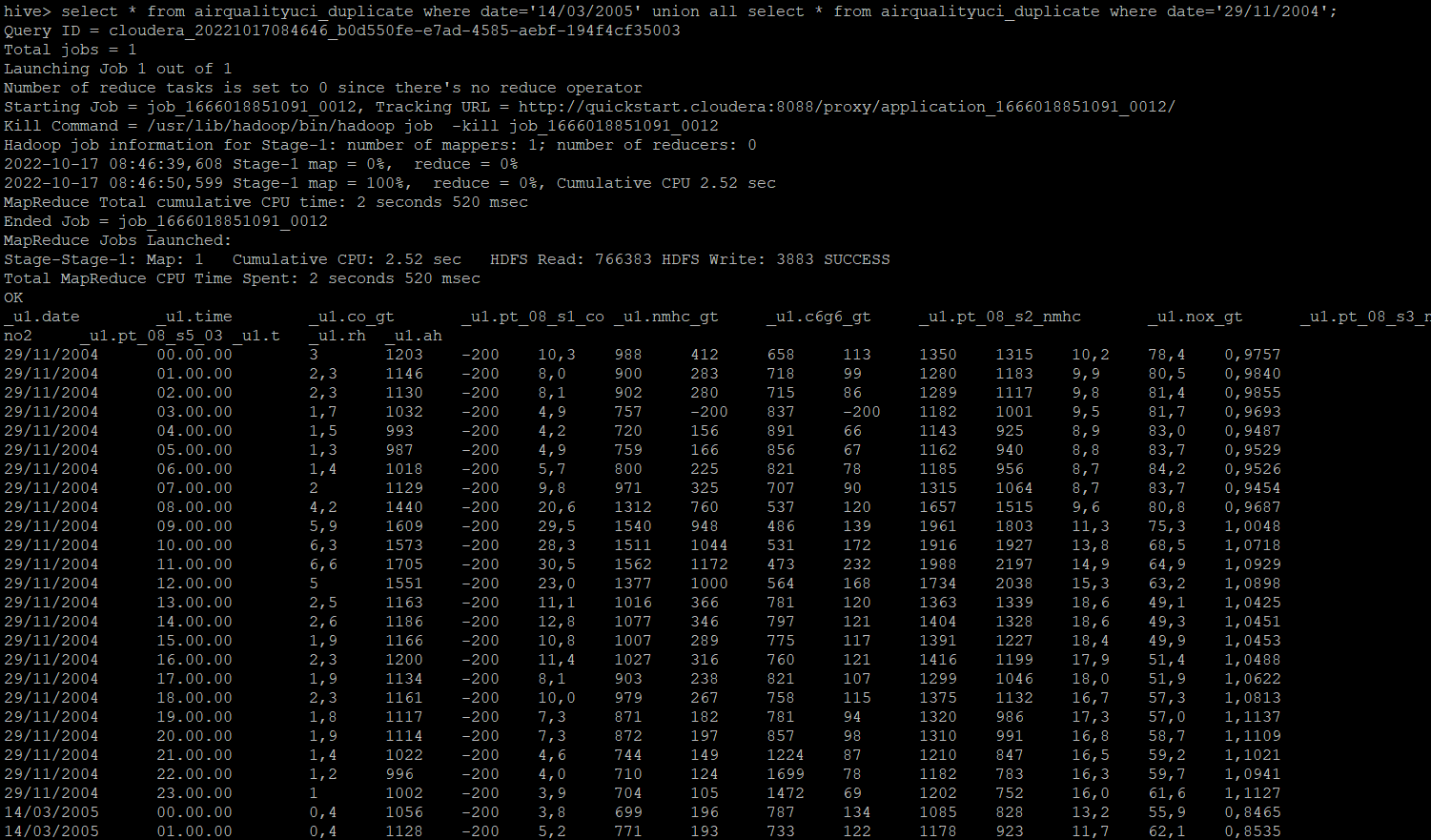


16 . like an operation you have to perform .

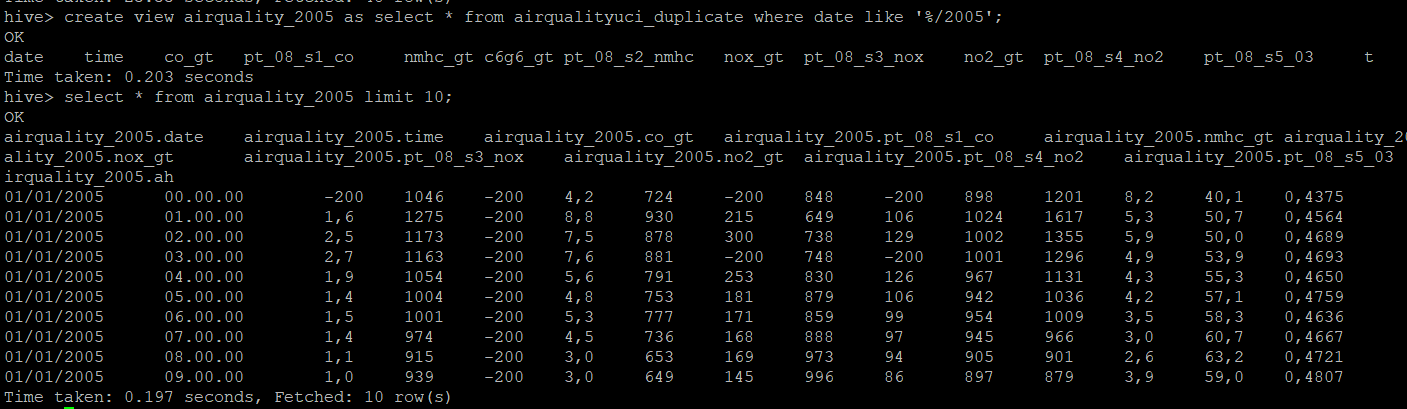
hive> select \* from airqualityuci1 where time like '18%' limit 10;



17 . union operation you have to perform .



18 . table view operation you have to perform .



hive operation with python

Create a python application that connects to the Hive database for extracting data,

creating sub tables for data processing,

drops temporary tables.

fetch rows to python itself into a list of tuples and mimic the join or filter operations

Note:see ipynb file in repo