set hive.support.concurrency=true;

set hive.enforce.bucketing=true;

set hive.exec.dynamic.partition.mode=nonstrict;

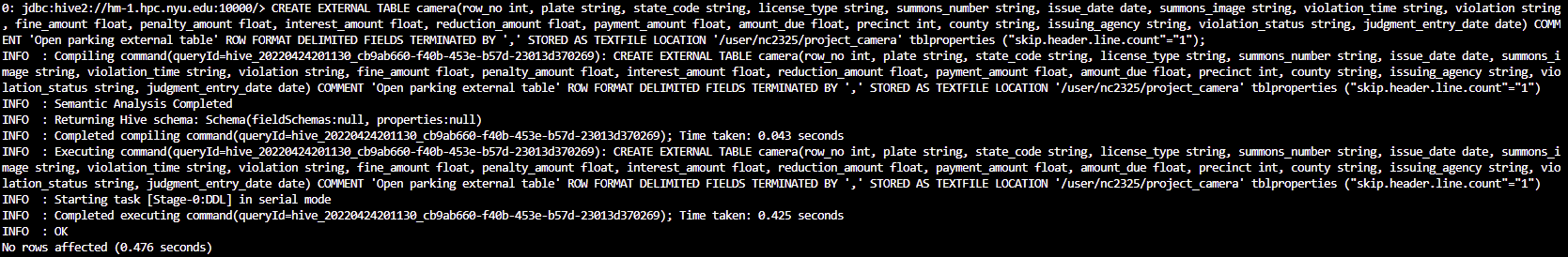
set hive.txn.manager=org.apache.hadoop.hive.ql.lockmgr.DbTxnManager;

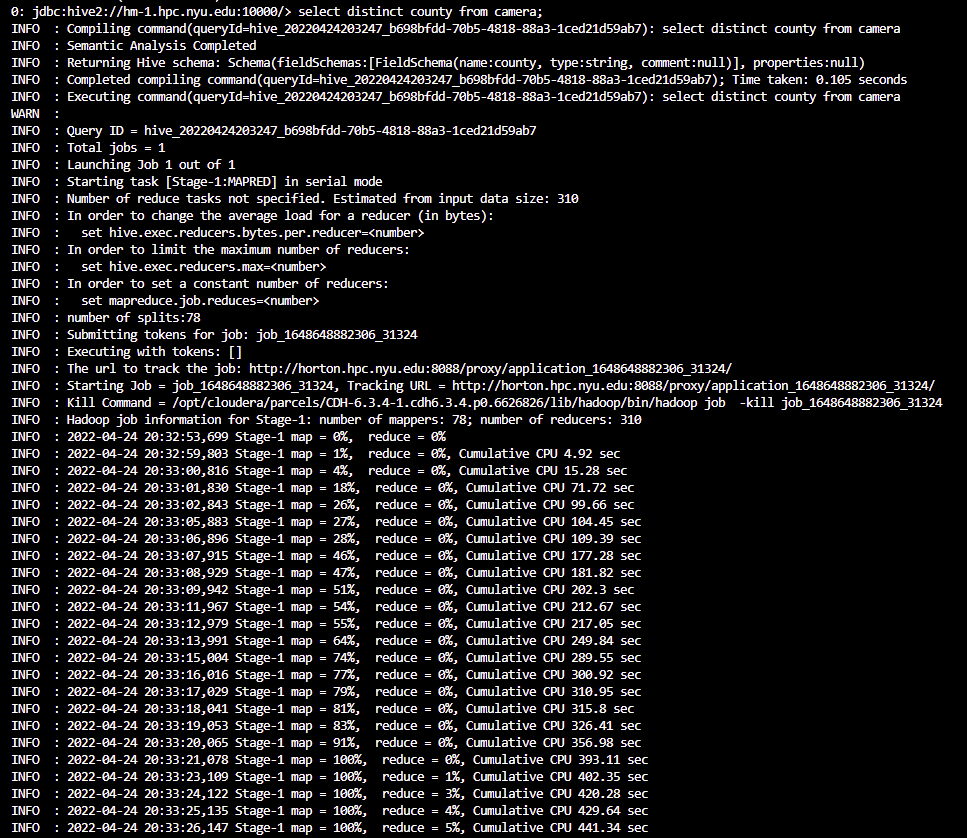
set hive.compactor.initiator.on=true;

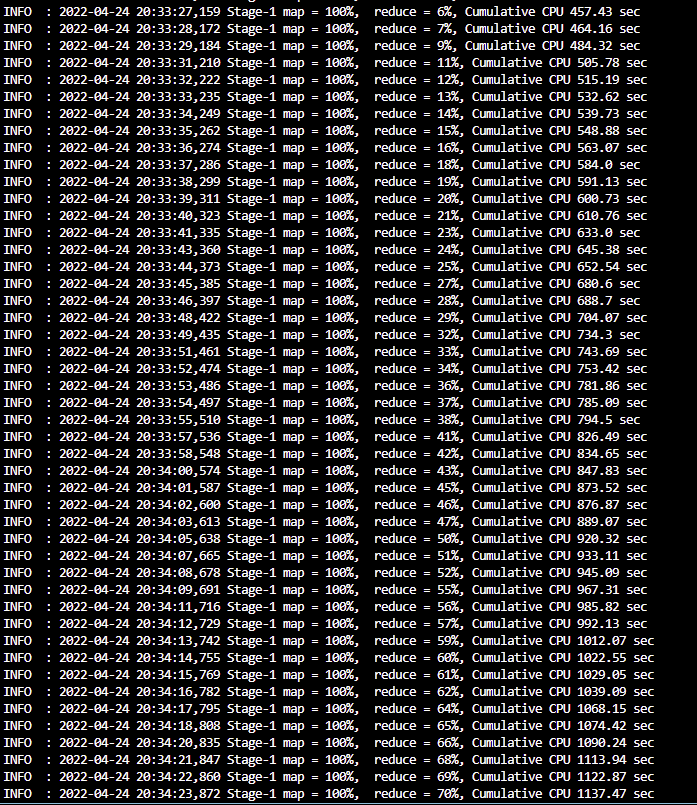
set hive.compactor.worker.threads=2;

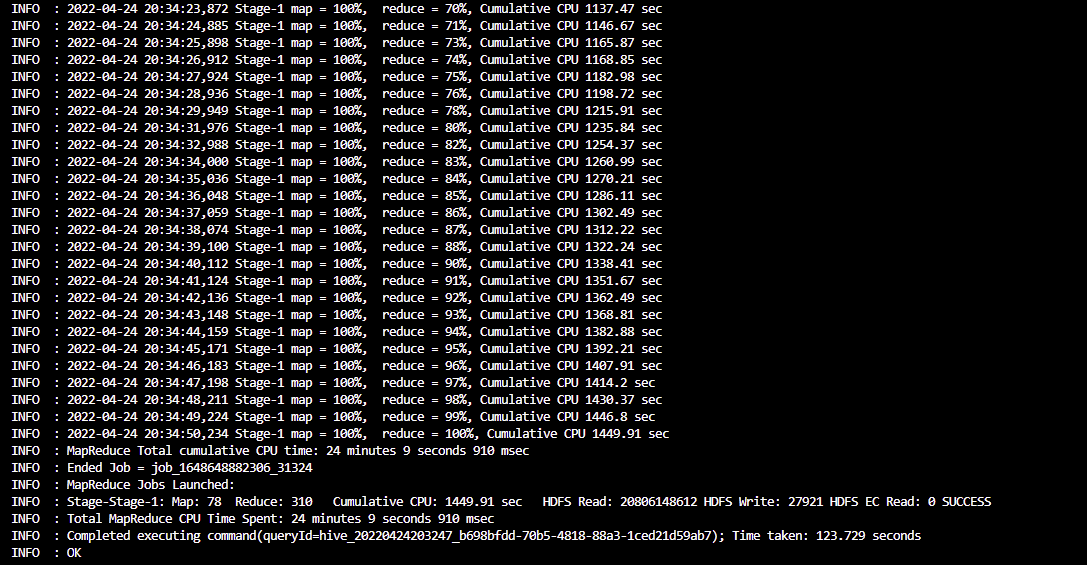
Evidence for create and import data into camera table

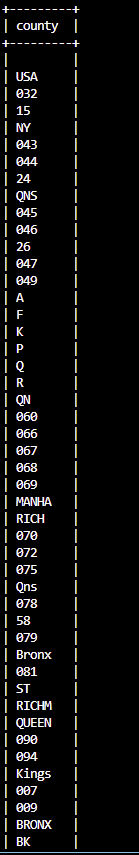
CREATE EXTERNAL TABLE camera(plate string, state\_code string, license\_type string, summons\_number string, issue\_date date, summons\_image string, violation\_time string, violation string, fine\_amount float, penalty\_amount float, interest\_amount float, reduction\_amount float, payment\_amount float, amount\_due float, precinct int, county string, issuing\_agency string, violation\_status string, judgment\_entry\_date date) COMMENT 'Open parking external table' ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' STORED AS TEXTFILE LOCATION '/user/nc2325/project\_camera' tblproperties ("skip.header.line.count"="1");

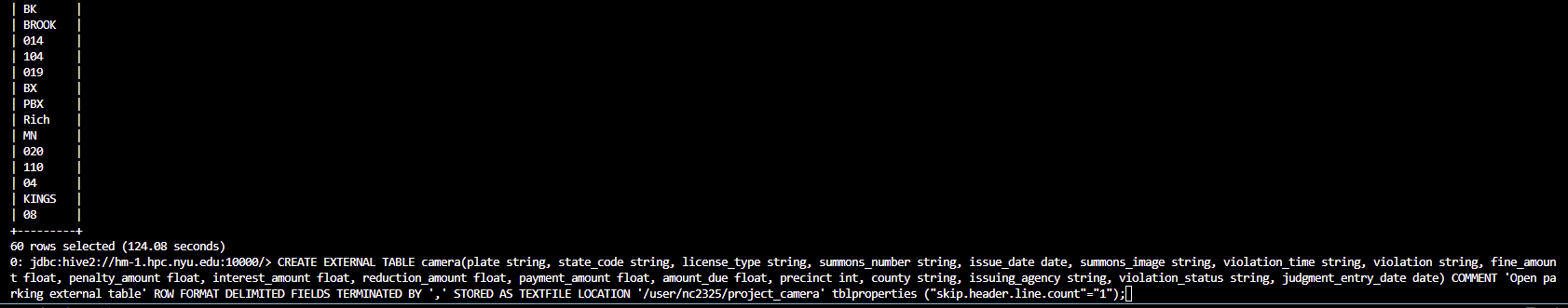
See distinct type in county column





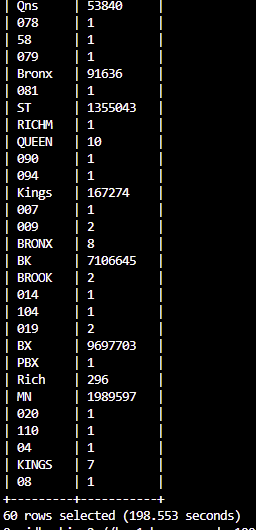
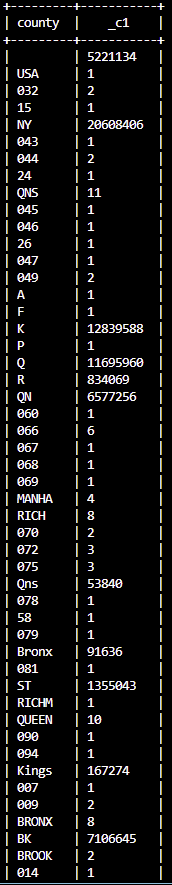
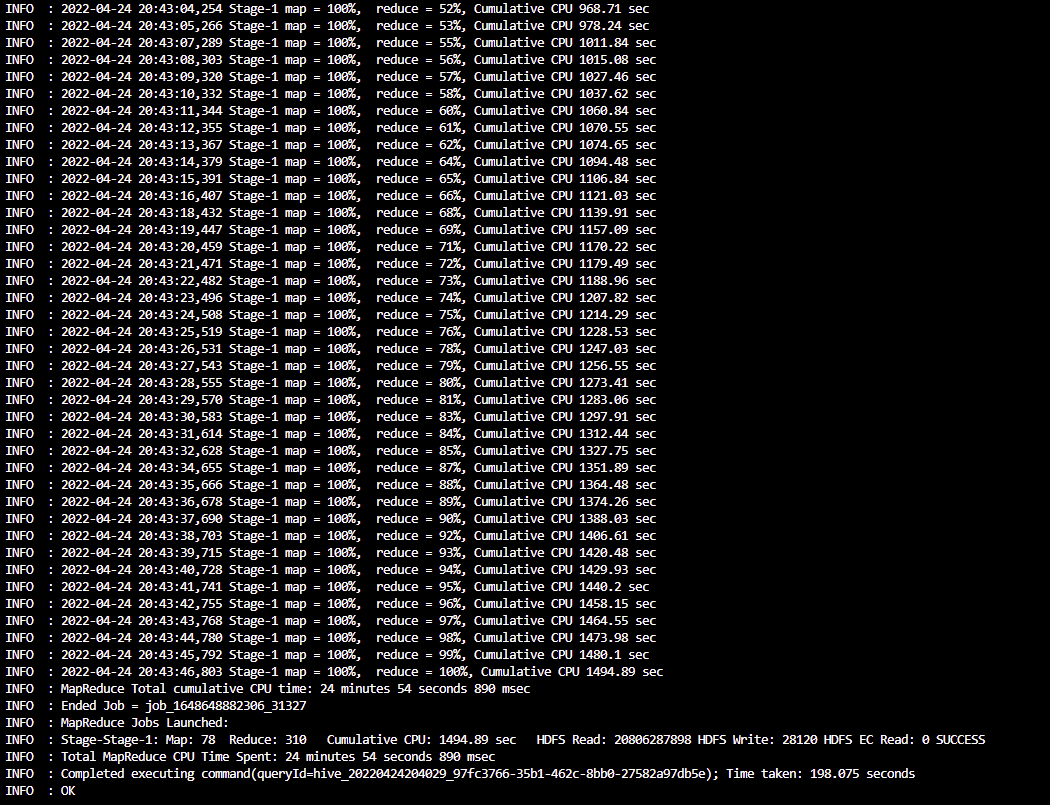
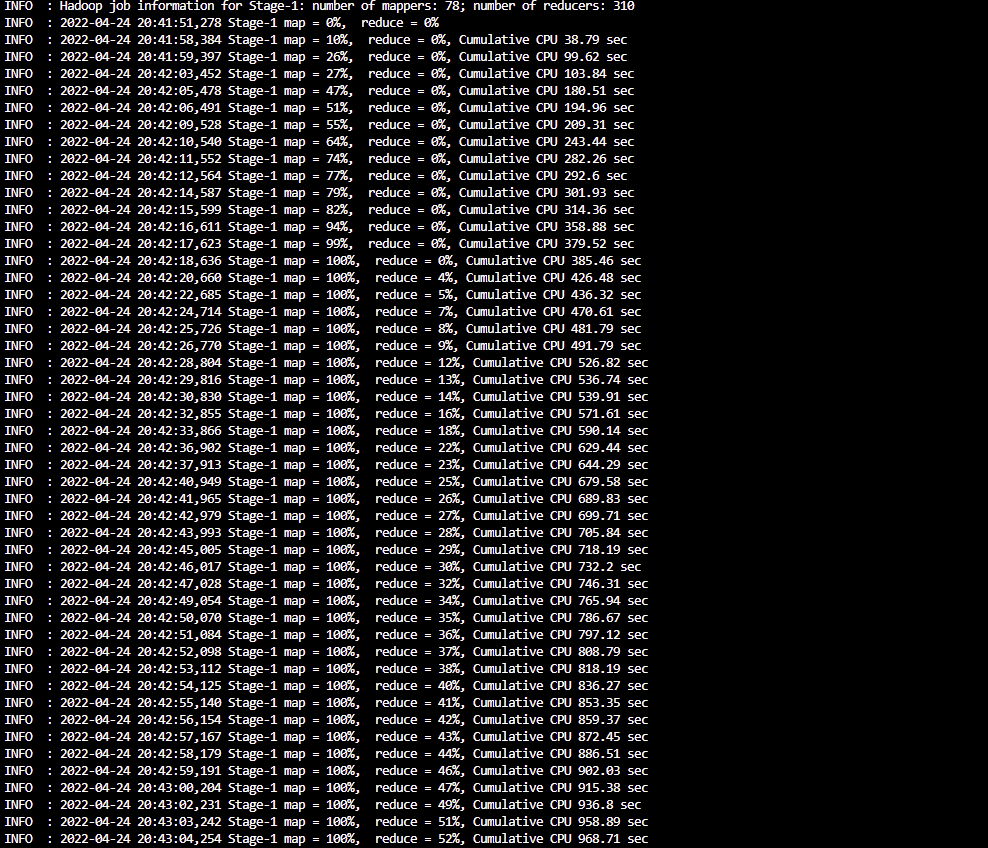
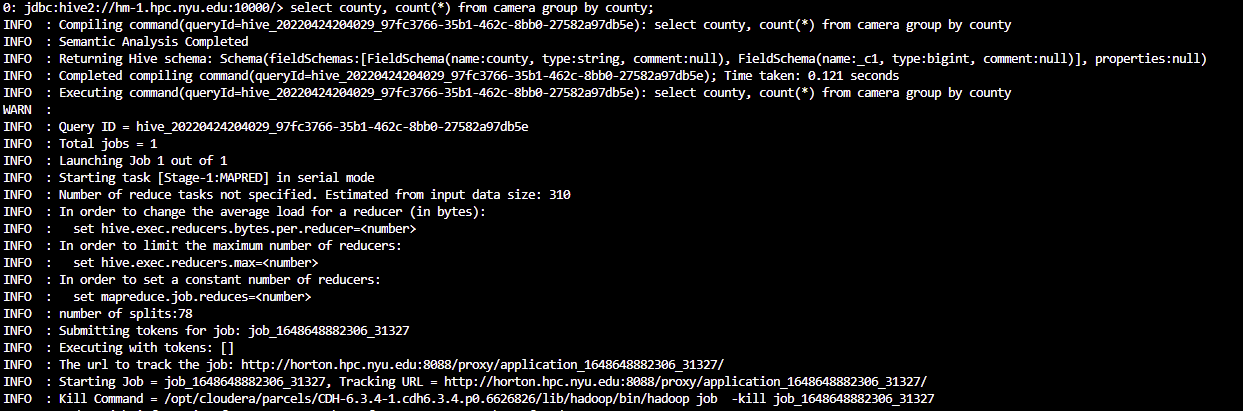






Count of each distinct value in county with our decision on the table

0: jdbc:hive2://hm-1.hpc.nyu.edu:10000/> select county, count(\*) from camera group by county;



beeline -u jdbc:hive2://hm-1.hpc.nyu.edu:10000/nc2325 -n nc2325 -p PASSWORD --outputformat=csv2 -e "SELECT

county, count(\*)

FROM

camera

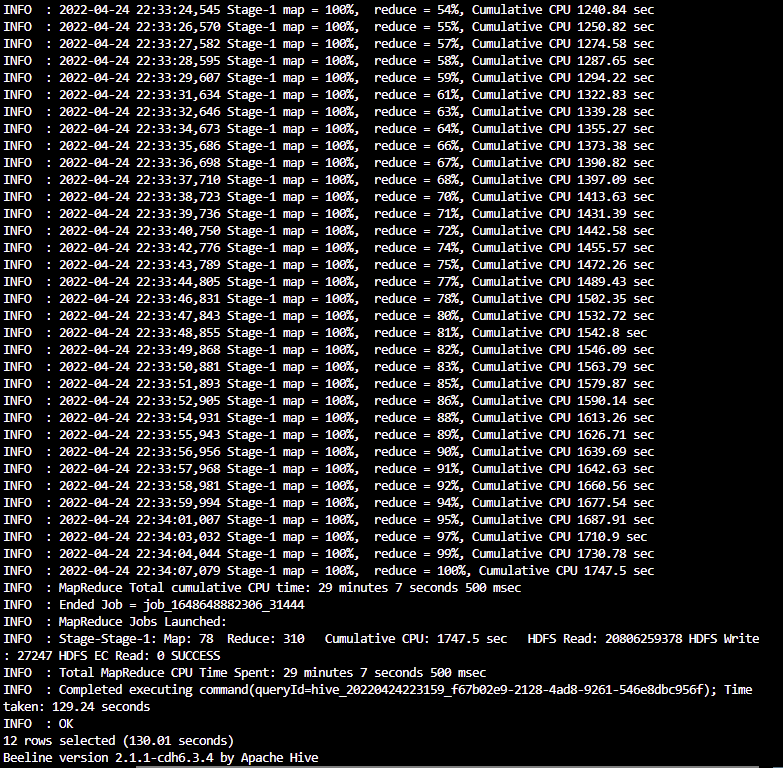
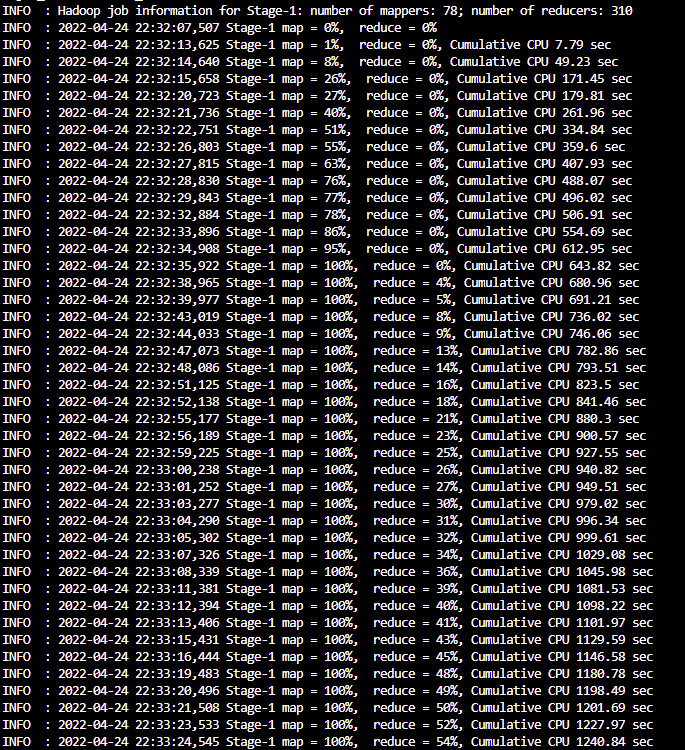
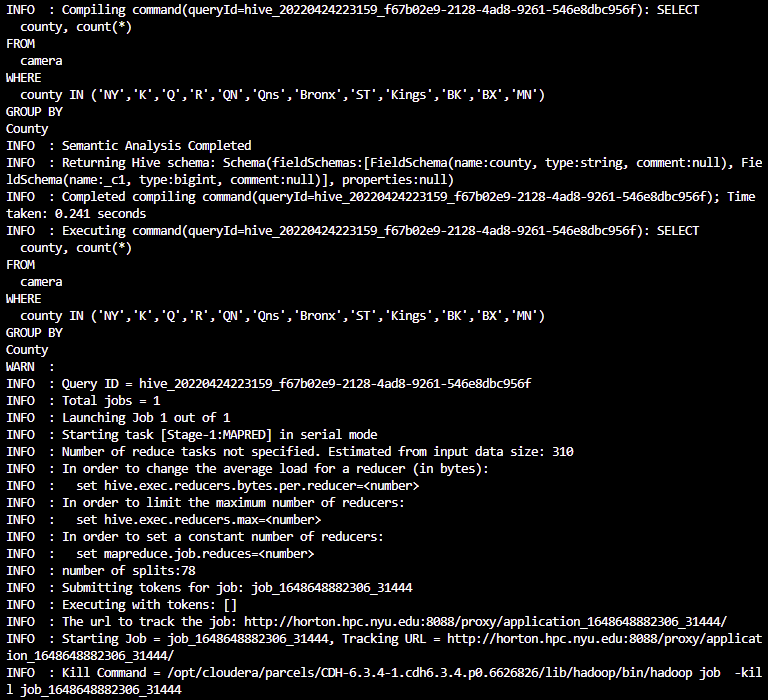
WHERE

county IN ('NY','K','Q','R','QN','Qns','Bronx','ST','Kings','BK','BX','MN')

GROUP BY

County;" > county\_camera.csv





COMPLAINT TABLE

create table crimes(ID string, COMPLAINT\_DATE date, COMPLAINT\_TIME string, COMPLAINT\_PROCESSED\_DATE date, COMPLAINT\_PROCESSED\_TIME string, PRECINCT string, REPORTED\_DATE date, OFFENSE\_CODE string, OFFENSE\_DESC string, INTERNAL\_CODE string, PD\_DESCRIPTION string, CRIME\_ATTEMPT string, LEVEL\_OF\_OFFENSE string, BOROUGH string, LOCATION string, PREMISE string, JURISDICTION\_DESC string, JURISDICTION\_CODE string, PARKS string, HADEVELOPT string, HOUSING\_PSA string, X\_COORD string, Y\_COORD string, AGE\_GROUP string, SUSP\_RACE string, SUSP\_SEX string, TRANSIT\_DISTRICT string, LATITUDE string, LONGITUDE string, LAT\_LONG string, PATROL\_BOROUGH string, STATION\_NAME string, VIC\_AGE string, VIC\_RACE string, VIC\_SEX string)

COMMENT 'CRIMES'

ROW FORMAT SERDE 'org.apache.hadoop.hive.serde2.OpenCSVSerde'

WITH SERDEPROPERTIES (

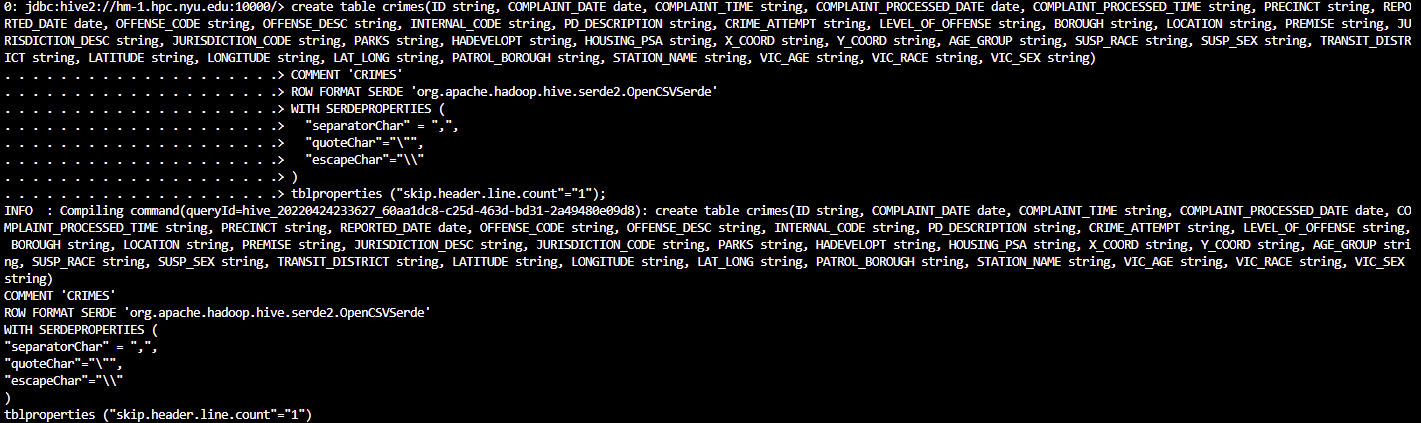
"separatorChar" = ",",

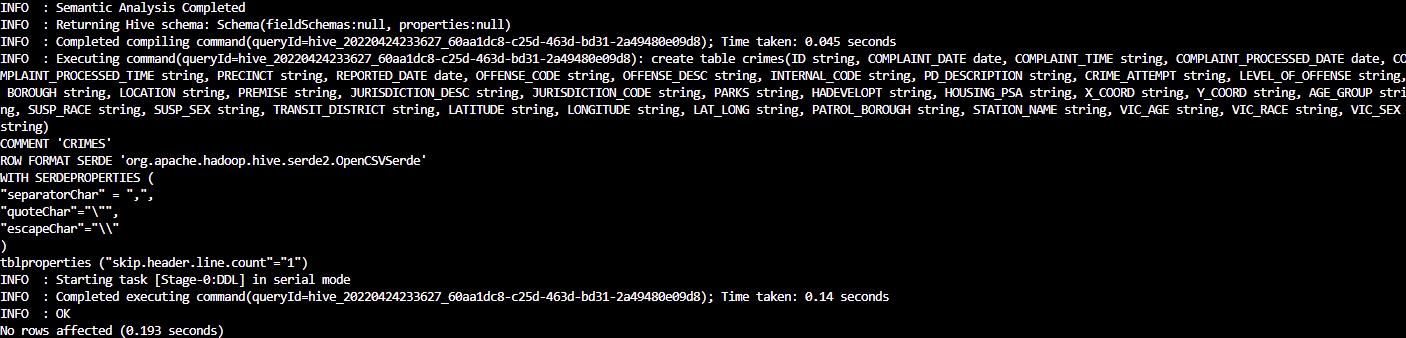
"quoteChar"="\"",

"escapeChar"="\\"

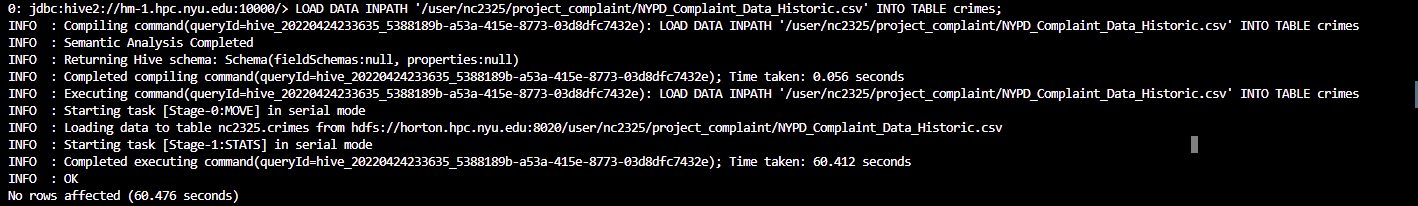
)

tblproperties ("skip.header.line.count"="1");





LOAD DATA INPATH '/user/nc2325/project\_complaint/NYPD\_Complaint\_Data\_Historic.csv' INTO TABLE crimes;



beeline -u jdbc:hive2://hm-1.hpc.nyu.edu:10000/nc2325 -n nc2325 -p PASSWORD --outputformat=csv2 -e "SELECT

BOROUGH, count(\*)

FROM

crimes

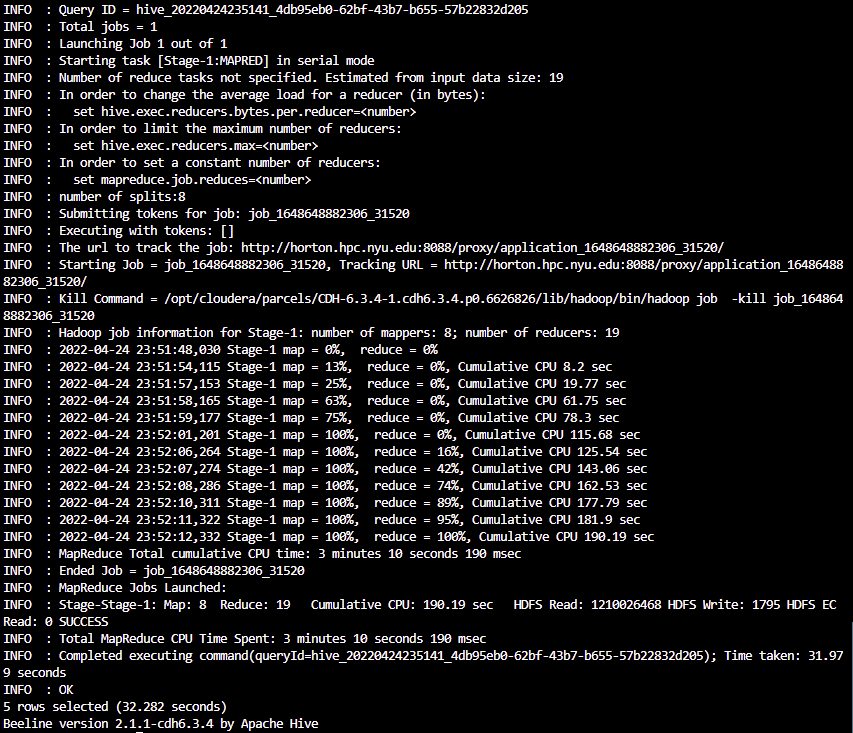
WHERE

BOROUGH IN ('MANHATTAN','QUEENS', 'STATEN ISLAND', 'BRONX', 'BROOKLYN')

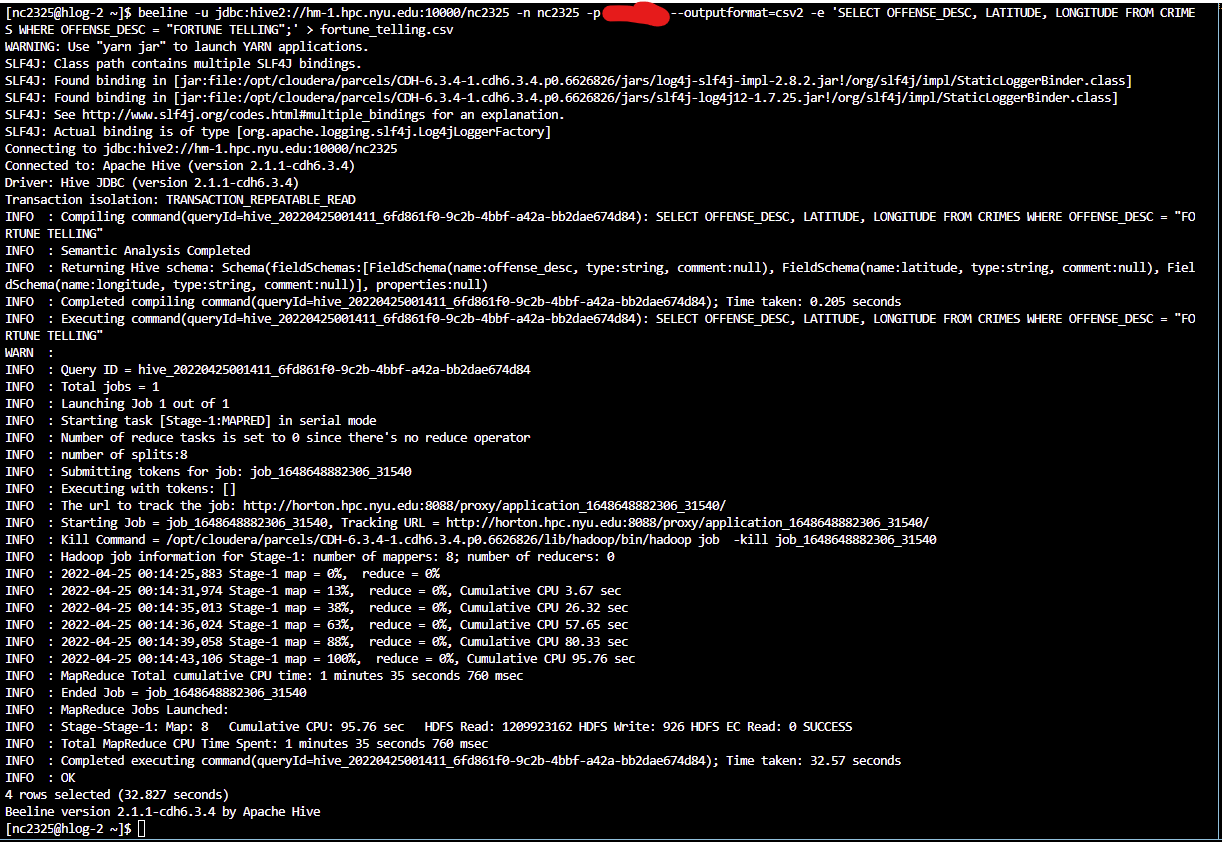
GROUP BY

BOROUGH;" > crimes.csv



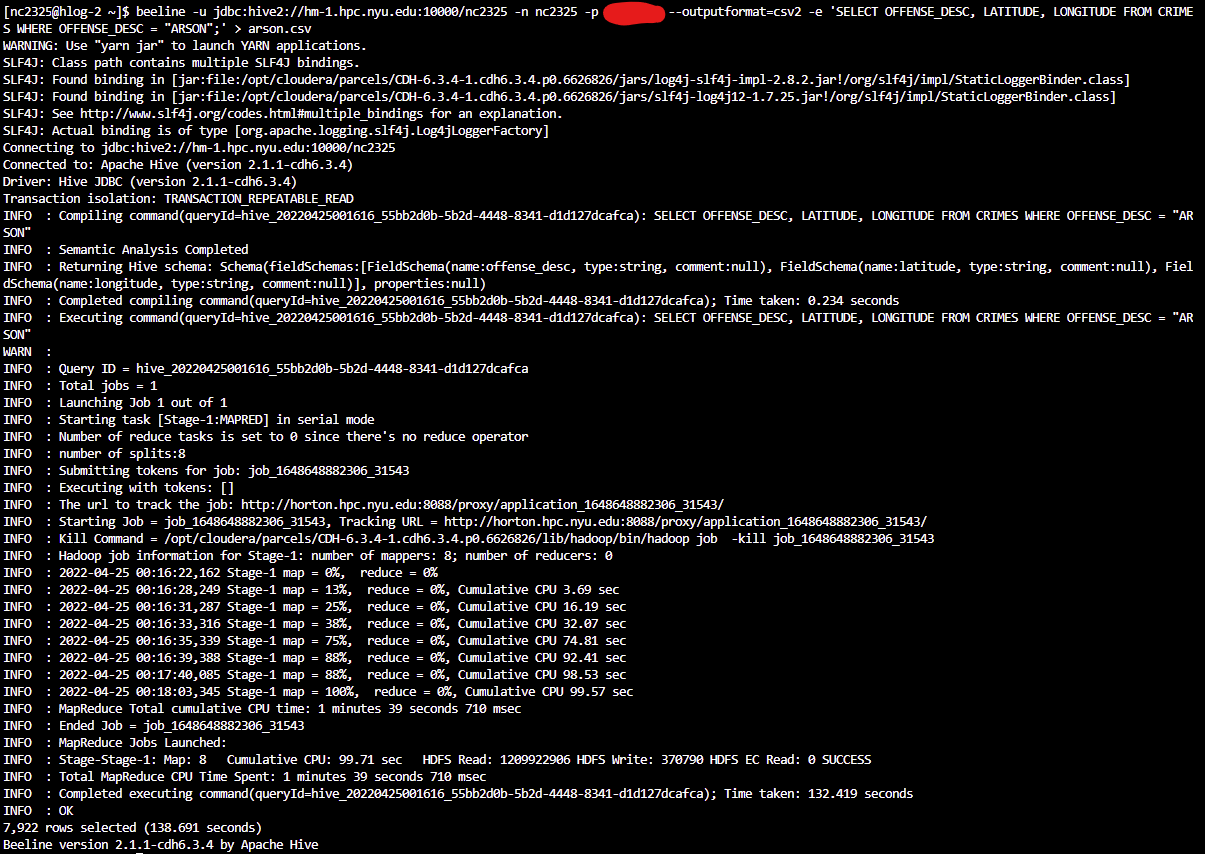


beeline -u jdbc:hive2://hm-1.hpc.nyu.edu:10000/nc2325 -n nc2325 -p PASSWORD --outputformat=csv2 -e ‘SELECT OFFENSE\_DESC, LATITUDE, LONGITUDE FROM CRIMES WHERE OFFENSE\_DESC = "FORTUNE TELLING";’ > fortune\_telling.csv

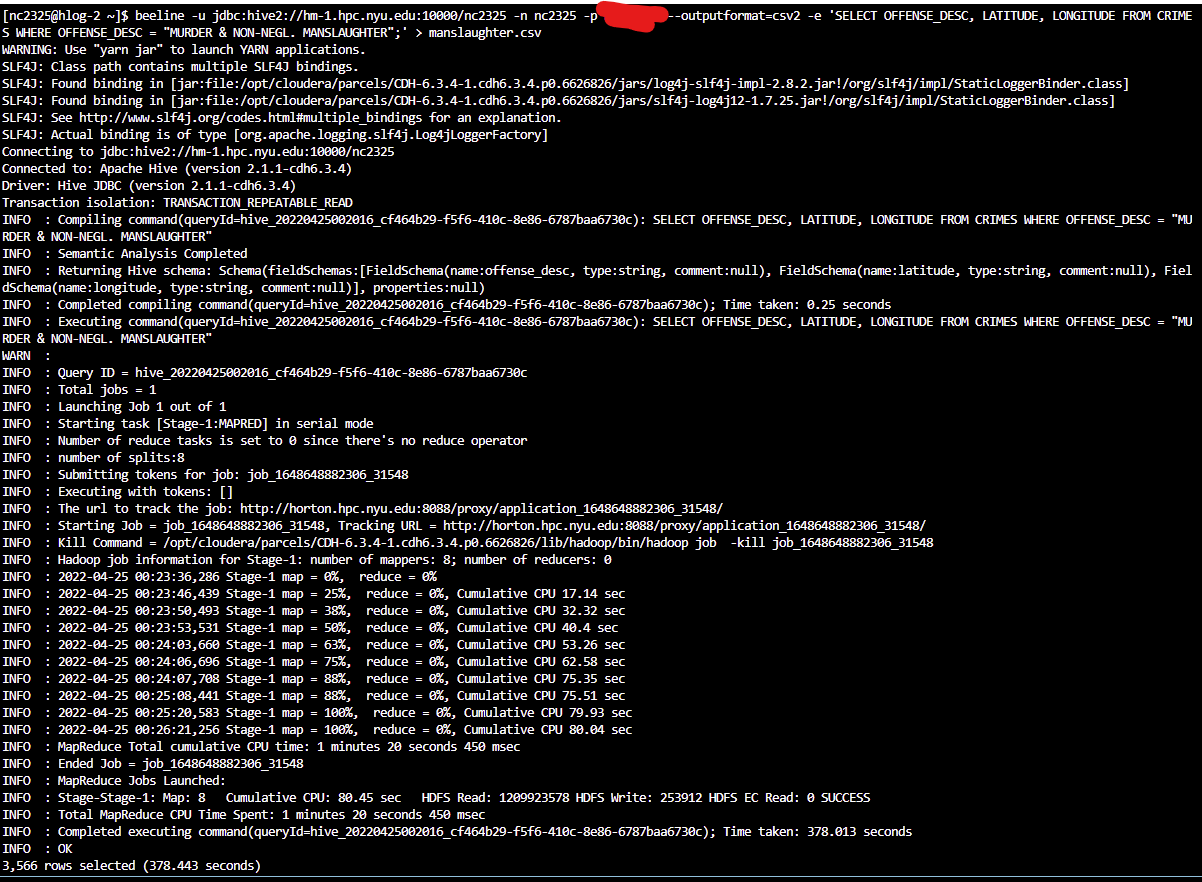


beeline -u jdbc:hive2://hm-1.hpc.nyu.edu:10000/nc2325 -n nc2325 -p PASSWORD --outputformat=csv2 -e ‘SELECT OFFENSE\_DESC, LATITUDE, LONGITUDE FROM CRIMES WHERE OFFENSE\_DESC = "ARSON";

’ > arson.csv



beeline -u jdbc:hive2://hm-1.hpc.nyu.edu:10000/nc2325 -n nc2325 -p PASSWORD --outputformat=csv2 -e ‘SELECT OFFENSE\_DESC, LATITUDE, LONGITUDE FROM CRIMES WHERE OFFENSE\_DESC = "MURDER & NON-NEGL. MANSLAUGHTER";’ > manslaughter.csv



beeline -u jdbc:hive2://hm-1.hpc.nyu.edu:10000/nc2325 -n nc2325 -p PASSWORD --outputformat=csv2 -e ‘SELECT OFFENSE\_DESC, LATITUDE, LONGITUDE, LAT\_LONG FROM CRIMES WHERE OFFENSE\_DESC = "KIDNAPPING & RELATED OFFENSES";’ > kidnapping.csv

