wget -O Newyork\_min\_data.csv <https://data.cityofnewyork.us/api/views/qgea-i56i/rows.csv?accessType=DOWNLOAD&bom=true&format=true>/Newyork\_min\_data.csv

hdfs dfs -mkdir NYPD

hdfs dfs -put Newyork\_min\_data.csv NYPD

hdfs dfs -ls

wget -O chicago\_min\_data.csv

https://data.cityofchicago.org/api/views/ijzp-q8t2/rows.csv?accessType=DOWNLOAD&bom=true&format=true/chicago\_min\_data.csv

hdfs dfs -mkdir Chicago

hdfs dfs -put chicago\_min\_data.csv Chicago

hdfs dfs -ls

## for Chicago dataset

CREATE EXTERNAL TABLE IF NOT EXISTS chicago (id BIGINT,

casenumber STRING, rgdate STRING, block STRING, IUCR STRING, primarytype STRING, lawcat STRING, locationdescription STRING, year STRING) ROW FORMAT DELIMITED FIELDS TERMINATED BY ','

STORED AS TEXTFILE LOCATION '/tmp/chicago'

TBLPROPERTIES ('skip.header.line.count'='1');

##Query To catergories crimes into Felony,Violation and Misdemeanor

create table chicago1 row format delimited fields terminitated by ',' stores as textfile location '/tmp/chicago/' as select \*,

if Primary Type == 'Arson' | 'Carry license violation' | 'Criminal trespass' | 'Gambling' | 'Public peace violation' | 'Liquor law violation' | 'intimidation'

| 'Public indecency', 'VIOLATION',

if Primary Type == 'non criminal Obscenity' | 'Other offense' | 'Ritualism' | 'Robbery' | 'Theft' | 'Burglary' | 'Weapons violation' | 'Other narcotic violation' | 'Deceptive practice' |'Interference with public' | 'Non criminal', 'MISDEMENOR', 'FELONY')) AS lawcat from chicago

##query to find out where crimes take place more in chicago top 3 locations

select locationdescription, count(iucr), rank() over (order by count(iucr)desc) AS rank from chicago1

group by locationdescription limit 3;

##query to find the top 5 cities with higest crimes recorded

SELECT block, count (iucr), rank () over (ORDER BY count (iucr) desc) AS rank from chicago

GROUP BY block limit 5;

##query to catergories the crims in the top city

SELECT lawcat, count (iucr) from chicago1 where block = "100XX W OHARE ST" GROUP BY lawcat;

##table to display top crime areas

CREATE TABLE area row format delimited fields terminated by ',' STORED AS RCFile

AS select block, count (iucr) AS total\_crimes, rank () over (ORDER BY count (iucr) desc) AS rank from chicago1

GROUP BY address;

##query for displaying the content of the above table

Select \* from RCFile limit 3;

## for NYPD dataset

DROP TABLE IF EXISTS nypd;

CREATE EXTERNAL TABLE IF NOT EXISTS nypd (cmplnt\_num BIGINT,

cmplnt\_fr\_dt STRING, rpt\_dt STRING, ky\_cd BIGINT, ofns\_desc STRING, law\_cat\_cd STRING, boro\_nm STRING, loc\_of\_occur\_desc STRING, prem\_typ\_desc STRING) ROW FORMAT DELIMITED FIELDS TERMINATED BY ','

STORED AS TEXTFILE LOCATION '/tmp/nypd'

TBLPROPERTIES ('skip.header.line.count'='1');

##query to find out where crimes take place more in New York top 3 locations

Select PREM\_TYP\_DESC, count (KY\_CD), rank() over (order by count(KY\_CD)desc) AS rank from nypd

group by PREM\_TYP\_DESC limit 3;

##query to find the top 5 cities with highest crimes recorded

SELECT BORO\_NM, count (KY\_CD), rank () over (ORDER BY count (KY\_CD) desc) AS rank from nypd

GROUP BY BORO\_NM limit 5;

##query to catergories the crimes in the top city

SELECT LAW\_CAT\_CD, count (KY\_CD) from nypd where BORO\_NM = "MANHATTAN" GROUP BY LAW\_CAT\_CD;