LA Hive Queries

For Los Angeles Area:

Download enriched\_data for LA from google drive to system hosting hadoop

1. wget -O lacrimedata.zip "<https://drive.google.com/uc?export=download&id=1pyMmySMP5MvaEBcuLQ3pA8CSyP_JTqcT>"
2. unzip lacrimedata.zip
3. Open hive terminal using: hive
4. Run Following commands on hive shell do:
   1. create database if not exists crime\_data;
   2. show databases;
   3. use crime\_data;
   4. CREATE TABLE IF NOT EXISTS la\_crime\_data(

dr\_no STRING,

date\_rptd DATE,

date\_occ DATE,

time\_occ STRING,

area\_id STRING,

area\_name STRING,

rpt\_dist\_no STRING,

crm\_cd STRING,

crm\_cd\_desc STRING,

mocodes STRING,

vict\_age STRING,

vict\_sex STRING,

vict\_descent STRING,

premis\_cd STRING,

premis\_desc STRING,

weapon\_used\_cd STRING,

weapon\_desc STRING,

status STRING,

status\_desc STRING,

crm\_cd\_1 STRING,

crm\_cd\_2 STRING,

crm\_cd\_3 STRING,

crm\_cd\_4 STRING,

location STRING,

cross\_street STRING,

location\_1 STRING,

zipcode STRING,

city STRING,

state STRING)ROW FORMAT SERDE'org.apache.hadoop.hive.serde2.OpenCSVSerde'WITH SERDEPROPERTIES ("separatorChar" = ",","quoteChar" = "\"")TBLPROPERTIES ("skip.header.line.count"="1");

* 1. load data local inpath '/home/ngupta8/tmp/finaloutput.csv' into table la\_crime\_data;
  2. select \* from la\_crime\_data limit 10;

1. Run Following commands from the bash shell and your home directory for instance /home/ngupta8 (NOT from Hive shell).

**Query -1: Top 10 crimes:**

hive -e 'set hive.cli.print.header=true;use crime\_data;Select count(\*) as crime\_count,crm\_cd\_desc from la\_crime\_data group by crm\_cd\_desc order by crime\_count DESC limit 10;'| perl -lpe 's/"/\\"/g; s/^|$/"/g; s/\t/","/g' > Top\_10\_Crime\_In\_LA.csv;

**Query -2 : Top 10 crimes per year :**

hive -e 'set hive.cli.print.header=true ; SET hive.groupby.orderby.position.alias=true ; use crime\_data; select count(\*) as crime\_count,from\_unixtime(unix\_timestamp(date\_occ,"mm/dd/yyyy"),"Y") as year from la\_crime\_data where crm\_cd\_desc is not NULL and crm\_cd\_desc != "" AND crm\_cd\_desc IN(

"BATTERY - SIMPLE ASSAULT",

"BURGLARY FROM VEHICLE",

"VEHICLE - STOLEN",

"BURGLARY",

"THEFT PLAIN - PETTY ($950 & UNDER)",

"THEFT OF IDENTITY",

"INTIMATE PARTNER - SIMPLE ASSAULT",

"VANDALISM - FELONY ($400 & OVER, ALL CHURCH VANDALISMS) 0114",

"VANDALISM - MISDEAMEANOR ($399 OR UNDER)",

"ASSAULT WITH DEADLY WEAPON, AGGRAVATED ASSAULT") group by 2'| perl -lpe 's/"/\\"/g; s/^|$/"/g; s/\t/","/g' >Top\_10\_Crime\_Per\_Year\_LA.csv ;

**Query -3 : Top 10 crimes by day :**

hive -e 'set hive.cli.print.header=true ; SET hive.groupby.orderby.position.alias=true ; use crime\_data ;with query as (Select count(\*) as crime\_count,crm\_cd\_desc,from\_unixtime(unix\_timestamp(date\_occ,"mm/dd/yyyy"),"E") as day,zipcode,city,state from la\_crime\_data where crm\_cd\_desc is not NULL and crm\_cd\_desc != "" AND crm\_cd\_desc IN(

"BATTERY - SIMPLE ASSAULT",

"BURGLARY FROM VEHICLE",

"VEHICLE - STOLEN",

"BURGLARY",

"THEFT PLAIN - PETTY ($950 & UNDER)",

"THEFT OF IDENTITY",

"INTIMATE PARTNER - SIMPLE ASSAULT",

"VANDALISM - FELONY ($400 & OVER, ALL CHURCH VANDALISMS) 0114",

"VANDALISM - MISDEAMEANOR ($399 OR UNDER)",

"ASSAULT WITH DEADLY WEAPON, AGGRAVATED ASSAULT") AND zipcode != "NONE" group by crm\_cd\_desc,3, zipcode,city,state) select \* from query where day != "NULL";'| perl -lpe 's/"/\\"/g; s/^|$/"/g; s/\t/","/g' >Top\_10\_Crime\_Per\_Day\_LA.csv;

**Query -4 : Top 10 crimes by hour :**

hive -e 'set hive.cli.print.header=true ; SET hive.groupby.orderby.position.alias=true ; use crime\_data ; with query as (Select count(\*) as crime\_count,crm\_cd\_desc,CAST(round((time\_occ/100),0) as INT) as hour,zipcode,city,state from la\_crime\_data where crm\_cd\_desc is not NULL and crm\_cd\_desc != "" AND crm\_cd\_desc IN(

"BATTERY - SIMPLE ASSAULT",

"BURGLARY FROM VEHICLE",

"VEHICLE - STOLEN",

"BURGLARY",

"THEFT PLAIN - PETTY ($950 & UNDER)",

"THEFT OF IDENTITY",

"INTIMATE PARTNER - SIMPLE ASSAULT",

"VANDALISM - FELONY ($400 & OVER, ALL CHURCH VANDALISMS) 0114",

"VANDALISM - MISDEAMEANOR ($399 OR UNDER)",

"ASSAULT WITH DEADLY WEAPON, AGGRAVATED ASSAULT") AND zipcode != "NONE" group by crm\_cd\_desc,3,zipcode,city,state) select \* from query where hour is not NULL;'| perl -lpe 's/"/\\"/g; s/^|$/"/g; s/\t/","/g' >Top\_10\_Crime\_Per\_Hour\_LA.csv;

**Query -5: Top 10 zipcodes :**

hive -e 'set hive.cli.print.header=true;use crime\_data ;select count(\*) as crime\_count ,zipcode from la\_crime\_data group by zipcode order by crime\_count DESC limit 10;'| perl -lpe 's/"/\\"/g; s/^|$/"/g; s/\t/","/g' > Top\_10\_Crime\_Prone\_zipcode\_LA.csv;

**Download processed data on local:**

Scp ngupta8@bi-hadoop-prod-4126.bi.services.us-south.bluemix.net:\*\_LA.csv .