Chicago Hive Queries:

1. curl -c /tmp/cookies "https://drive.google.com/uc?export=download&id=1IKFo8GFWu4gshICahcaQlwT7UcXVCCg2"> /tmp/enriched\_ny.html
2. curl -L -b /tmp/cookies "https://drive.google.com$(cat /tmp/enriched\_ny.html | grep -Po 'uc-download-link" [^>]\* href="\K[^"]\*' | sed 's/\&amp;/\&/g')" > chicagocrimedata.zip
3. unzip chicagocrimedata.zip
4. Open hive terminal using: hive
5. 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 chicago\_crime\_data(ID INT,  
      Case\_Number STRING,  
      Crime\_date STRING,  
      Block STRING,  
      IUCR INT,  
      Primary\_Type STRING,  
      Description STRING,  
      Location\_Description STRING,  
      Arrest BOOLEAN,  
      Domestic BOOLEAN,  
      Beat INT,  
      District INT,  
      Ward INT,  
      Community\_Area INT,  
      FBI\_Code INT,  
      X\_Coordinate INT,  
      Y\_Coordinate INT,  
      Year INT,  
      Updated\_On STRING,  
      Latitude DOUBLE,  
      Longitude DOUBLE,  
      Location DOUBLE,  
      Zipcode INT,  
      City STRING,  
      State STRING)  
      ROW FORMAT SERDE'org.apache.hadoop.hive.serde2.OpenCSVSerde'WITH SERDEPROPERTIES ("separatorChar" = ",","quoteChar" = "\"")TBLPROPERTIES ("skip.header.line.count"="1");
   5. load data local inpath '/home/ngupta8/chicagooutput.csv' into table chicago\_crime\_data;
   6. select \* from chicago\_crime\_data limit 10;

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

hive -e "set hive.cli.print.header=true ;use crime\_data ;Select count(\*) as crime\_count,Primary\_Type from chicago\_crime\_data group by Primary\_Type order by crime\_count DESC limit 10;"| perl -lpe 's/"/\\"/g; s/^|$/"/g; s/\t/","/g' > Top\_10\_Crime\_Chicago.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(Crime\_date,"mm/dd/yyyy"),"Y") as year from chicago\_crime\_data where Primary\_Type is not NULL and Primary\_Type != "" AND Primary\_Type IN(

"THEFT",

"BATTERY",

"CRIMINAL DAMAGE",

"NARCOTICS",

"OTHER OFFENSE",

"ASSAULT",

"BURGLARY",

"MOTOR VEHICLE THEFT",

"ROBBERY",

"DECEPTIVE PRACTICE") group by 2'| perl -lpe 's/"/\\"/g; s/^|$/"/g; s/\t/","/g' >Top\_10\_Crime\_Per\_Year\_Chicago.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,Primary\_Type,from\_unixtime(unix\_timestamp(Crime\_date,"mm/dd/yyyy"),"E") as day ,zipcode,city,state from chicago\_crime\_data where Primary\_Type is not NULL and Primary\_Type != "" AND Primary\_Type IN(

"THEFT",

"BATTERY",

"CRIMINAL DAMAGE",

"NARCOTICS",

"OTHER OFFENSE",

"ASSAULT",

"BURGLARY",

"MOTOR VEHICLE THEFT",

"ROBBERY",

"DECEPTIVE PRACTICE") AND zipcode != "NONE" group by Primary\_Type,3, zipcode,city,state) select \* from query where day != "NULL";'| perl -lpe 's/"/\\"/g; s/^|$/"/g; s/\t/","/g' >Top\_10\_Crime\_Per\_Day\_Chicago.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,Primary\_Type,from\_unixtime(unix\_timestamp(Crime\_date,"mm/dd/yyyy hh:mm:ss a"),"H") as hour ,zipcode,city,state from chicago\_crime\_data where Primary\_Type is not NULL and Primary\_Type != "" AND Primary\_Type IN(

"THEFT",

"BATTERY",

"CRIMINAL DAMAGE",

"NARCOTICS",

"OTHER OFFENSE",

"ASSAULT",

"BURGLARY",

"MOTOR VEHICLE THEFT",

"ROBBERY",

"DECEPTIVE PRACTICE") AND zipcode != "NONE" group by Primary\_Type,3, zipcode,city,state) select \* from query order by hour;'| perl -lpe 's/"/\\"/g; s/^|$/"/g; s/\t/","/g' >Top\_10\_Crime\_Per\_Hour\_Chicago.csv;

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

hive -e 'set hive.cli.print.header=true;use crime\_data ; select count(\*) as crime\_count,zipcode from chicago\_crime\_data where zipcode is not NULL and zipcode !="NONE" and zipcode != "" group by zipcode order by crime\_count DESC limit 10;'| perl -lpe 's/"/\\"/g; s/^|$/"/g; s/\t/","/g' > Top\_10\_Crime\_Prone\_zipcode\_Chicago.csv;

**Download processed data on local:**

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