Crime Dataset Analysis

The dataset contains Crimes data. **crime statistics** are used in many ways and serve many purposes. They provide law enforcement with data for use in budget formulation, planning, resource allocation, assessment of police operations, etc., to help address the crime problem at various levels.

Understanding Data

The data format is comma separated values. There are 200 observations.

Schema Description

Id,case_no,date,block,IUCR,Primary_type,description,Loc_des,arrest,domestic,beat,district,ward,community,fbicode,XCor,YCor,year,Updated_on

Problem Statements

1) Load the crimes file from HDFS and make it an RDD & Convert the RDD into dataframe.

$spark-shell

```
val sqlContext = new org.apache.spark.sql.SQLContext(sc)

import sqlContext.implicits._

scala> case class
CrimeRecords(id:String,case_no:String,date:String,block:String,IUCR:String,Primary_type:String,description:String,Loc_des:String,arrest:String,domestic:String,beat:String,district:String,ward:String,community:String,fbicode:String,XCor:String,YCor:String,year:String,Updated_on:String)

scala> val crime_fram =
sc.textFile("file:///home/cloudera/khasimbabu/Spark_Excercise/crimes.txt").map(_.split(',')).map(i
=>
CrimeRecords(i(0),i(1),i(2),i(3),i(4),i(5),i(6),i(7),i(8),i(9),i(10),i(11),i(12),i(13),i(14).trim,i(15),i(16),i(17),i(18))).toDF()
```

2) Register the temp table for the dataframe.

```
scala> crime_fram.printSchema()

scala> crime_fram.registerTempTable("CrimeRecordsTable")

scala> val allrecords = sqlContext.sql("select * from CrimeRecordsTable")

scala> allrecords.show()
```

4) Find number of crimes that happened under each FBI code

```
val allrecords = sqlContext.sql("select fbicode,count(fbicode) as count from CrimeRecordsTable
group by fbicode")

allrecords.show()
```

5) Find number of 'NARCOTICS' cases filed in the year 2015

```
val allrecords = sqlContext.sql("select count(*) as count from CrimeRecordsTable where Primary_type ='NARCOTICS' and year = '2015' ")

allrecords.show()
```

6) Find the number of theft related arrests that happened in each district.

```
val allrecords = sqlContext.sql("select district ,count(*) as count from CrimeRecordsTable where Primary_type ='THEFT' and arrest = 'true' group by district")

allrecords.show()
```