

BIG DATA ANALYTICS

ITE-2013

J-Component

***“Crime Rate Analysis using Hadoop
and Python”***

17BIT0336-Shubhang

17BIT0036-Amandeep Singh

17BIT0238-Jay Paliwal

Problem Statement

Crimes have been severely increased in past few years, the Problem Statement includes analysis of crimes with different perspectives including utmost attributes possible and predicting via the study of nature of crimes committed. The problem statement is described to initially predict the crime-type based on location and time. We worked on data about historical crimes in California.

We had close to 13,000 records of crimes with data on the date and time of the crime, its location, and its type. Common types of crime include theft, criminal damage, criminal trespass, and assault. This project took on the task of predicting the type of crime that was committed given a police report in two ways one according to time that is when crime took place and another is location that is where crime took place. From a small number of overly detailed features, in time it will give the detail that at which time slot which crime is maximum and in location it will tell at which place which type of crime is maximum. They then trained various diagram based models (Graphs and Pie charts) to classify crimes by type using the generated features.

Finally, they tested the performance of their models on testing data. They conclude that predicting the type of crimes committed by time and location alone is quite difficult, but that the feature engineering greatly increases predictive power. Predictions will be made to provide local authorities with an upper hand on crime and help them plan a better strategy to tackle the same.

Introduction

At a glance, here's what we did:

- *Data Set Collection*
- *Implementation of our code*
- *Displaying results according to different objectives*

Data was collected from <https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-present/ijzp-q8t2> (2017 Dataset Crime). Data was pre-processed and cleaned to remove missing values and garbage values at various positions. Then, it was given as input to our code in mapper code. The output from mapper was sent to reducer code and the corresponding results were printed. The results were saved in a text file as the code was run in Hadoop. The part of code was then selected and pasted in excel file. Corresponding graphs were created to display the results more beautifully.

Output of the code:

- ***Based on Time of Crime*** – Analysis and predictions will be based on time of crimes i.e. which time have the maximum crime rates and needs to be inspected more efficiently.
- ***Based on Location of Crime*** – The most Crime prone locations will provide local authorities to target specific area clusters to counter crime.
- ***Based on District of Crime***– The most Crime prone Districts will provide local authorities to target specific area clusters to counter crime.

Software/Hardware Requirements:

Tools used for this project:

- *VMware Virtual Workstation*
- *Windows and Cloudera OS*
- *Rapid Miner*
- *Hadoop HDFS*
- *Python*
- *Microsoft Excel*

DATA PREPARATION PHASE

2.a)

DATA SOURCE:

<https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-present/ijzp-q8t2>

DATA SET DESCRIPTION:

Dataset contains the following attributes:

- | | | |
|------------------------------------|---------------------------|-------------|
| <i>1. ID</i> | <i>11. District</i> | <i>Ward</i> |
| <i>2. Case Number</i> | <i>12. Community Area</i> | |
| <i>3. Date</i> | <i>13. FBI Code</i> | |
| <i>4. Block</i> | <i>14. X Coordinate</i> | |
| <i>5. IUCR</i> | <i>15. Y Coordinate</i> | |
| <i>6. Primary Type Description</i> | <i>16. Year</i> | |
| <i>7. Location Description</i> | <i>17. Updated On</i> | |
| <i>8. Arrest</i> | <i>18. Latitude</i> | |
| <i>9. Domestic</i> | <i>19. Longitude</i> | |
| <i>10. Beat</i> | <i>20. Location</i> | |

Column Name	Description	Type	
<i>ID</i>	<i>Unique identifier for the record.</i>	<i>Number</i>	
<i>Case Number</i>	<i>The Chicago Police Department RD Number (Records Division Number), which is unique to the incident.</i>	<i>Plain Text</i>	
<i>Date</i>	<i>Date when the incident occurred. this is sometimes a best estimate.</i>	<i>Date & Time</i>	
<i>Block</i>	<i>The partially redacted address where the incident occurred, placing it on the same block as the actual address.</i>	<i>Plain Text</i>	
<i>IUCR</i>	<i>The Illinois Uniform Crime Reporting code. This is directly linked to the Primary Type and Description. See the list of IUCR codes at https://data.cityofchicago.org/d/c7ck-438e.</i>	<i>Plain Text</i>	
<i>Primary Type</i>	<i>The primary description of the IUCR code.</i>	<i>Plain Text</i>	
<i>Description</i>	<i>The secondary description of the IUCR code, a subcategory of the primary description.</i>	<i>Plain Text</i>	
<i>Location Description</i>	<i>Description of the location where the incident occurred.</i>	<i>Plain Text</i>	
<i>Arrest</i>	<i>Indicates whether an arrest was made.</i>	<i>Checkbox</i>	
<i>Domestic</i>	<i>Indicates whether the incident was domestic-related as defined by the Illinois Domestic Violence Act.</i>	<i>Checkbox</i>	
<i>Beat</i>	<i>Indicates the beat where the incident occurred. A beat is the smallest police geographic area – each beat has a dedicated police beat car. Three to five beats make up a police sector, and three sectors make up a police district. The Chicago Police Department has 22 police districts. See the beats at https://data.cityofchicago.org/d/aerh-rz74.</i>	<i>Plain Text</i>	
<i>District</i>	<i>Indicates the police district where the incident occurred. See the districts at https://data.cityofchicago.org/d/fthy-xz3r.</i>	<i>Plain Text</i>	

<i>Ward</i>	<i>The ward (City Council district) where the incident occurred. See the wards at https://data.cityofchicago.org/d/sp34-6z76.</i>	<i>Number</i>	
<i>Community Area</i>	<i>Indicates the community area where the incident occurred. Chicago has 77 community areas. See the community areas at https://data.cityofchicago.org/d/cauq-8yn6.</i>	<i>Plain Text</i>	
<i>FBI Code</i>	<i>Indicates the crime classification as outlined in the FBI's National Incident-Based Reporting System (NIBRS). See the Chicago Police Department listing of these classifications at http://gis.chicagopolice.org/clearmap_crime_sums/crime_types.html.</i>	<i>Plain Text</i>	
<i>X Coordinate</i>	<i>The x coordinate of the location where the incident occurred in State Plane Illinois East NAD 1983 projection. This location is shifted from the actual location for partial redaction but falls on the same block.</i>	<i>Number</i>	
<i>Y Coordinate</i>	<i>The y coordinate of the location where the incident occurred in State Plane Illinois East NAD 1983 projection. This location is shifted from the actual location for partial redaction but falls on the same block.</i>	<i>Number</i>	
<i>Year</i>	<i>Year the incident occurred.</i>	<i>Number</i>	
<i>Updated On</i>	<i>Date and time the record was last updated.</i>	<i>Date & Time</i>	
<i>Latitude</i>	<i>The latitude of the location where the incident occurred. This location is shifted from the actual location for partial redaction but falls on the same block.</i>	<i>Number</i>	
<i>Longitude</i>	<i>The longitude of the location where the incident occurred. This location is shifted from the actual location for partial redaction but falls on the same block.</i>	<i>Number</i>	
<i>Location</i>	<i>The location where the incident occurred in a format that allows for creation of maps and other geographic operations on this data portal. This location is shifted from the actual location for partial redaction but falls on the same block.</i>	<i>Location</i>	

DATASET SCREENSHOTS:

ID	CASE_NUM	DATE	BLOCK	IUCR	PRIMARY	DESCRIPTION	LOCATION	ARREST	DOMESTIC	BEAT	DISTRICT	WARD	COMMUN	FBKCODE	X_COORD	Y_COORD	YEAR	UPDATED
1	3512276	HK567712	06/28/200	0470X S KI	890	THEFT	FROM BUI SMALL RET	FALSE	FALSE	911	9	14	58	6	1155838	1879050	2004	02/28/201
2	3406013	HK456306	06/26/200	0090X N C	820	THEFT	\$500 AND OTHER	FALSE	FALSE	1112	11	27	23	6	1152206	1906127	2004	02/28/201
3	8002131	HT233595	04/30/201	0430X S VA	820	THEFT	\$500 AND NURSING	FALSE	FALSE	221	2	3	38	6	1177436	1876313	2011	02/28/201
4	7903289	HT133522	12/30/201	0830X S KI	840	THEFT	FINANCIA RESIDENC	FALSE	FALSE	423	4	7	46	6	1154622	1850125	2010	02/28/201
5	10402076	HZ138551	05/13/200	0330X W E	820	THEFT	\$500 AND ALLEY	FALSE	FALSE	831	8	15	66	6	1155240	1860961	2010	02/28/201
6	7732712	H5540108	09/29/201	0060X W C	810	THEFT	OVER \$50K PARKING	FALSE	FALSE	1523	12	27	24	6	1171668	1905607	2010	02/28/201
7	10789475	HZ534775	11/30/201	0120X N K	810	THEFT	OVER \$50K STREET	FALSE	FALSE	1713	17	33	14	6	1154133	1933114	2010	02/28/201
8	4494340	HL793243	12/16/200	0050X S PI	860	THEFT	RETAIL TH GROCERY	TRUE	FALSE	213	2	3	38	6	1180448	1879234	2005	02/28/201
9	3778925	HL349610	01/28/200	1000X S VA	810	THEFT	OVER \$50K STREET	FALSE	FALSE	2211	22	19	72	6	1160129	1838040	2005	02/28/201
10	3124217	HK316151	05/13/200	0330X W E	820	THEFT	\$500 AND SMALL RET	FALSE	FALSE	1733	17	35	21	6	1155390	1921064	2004	02/28/201
11	9902691	HK552797	12/23/201	0760X S CI	860	THEFT	RETAIL TH DEPARTM	TRUE	FALSE	833	8	13	65	6	1145727	1853720	2014	02/28/201
12	7941984	HT180228	02/26/201	0120X S KI	820	THEFT	\$500 AND BAR OR TA	FALSE	FALSE	2232	12	25	28	6	1171230	1894421	2011	02/28/201
13	8568365	HV243612	04/15/201	1080X S EF	820	THEFT	\$500 AND RESIDENT	FALSE	FALSE	2233	22	34	49	6	1173239	1832994	2012	02/28/201
14	5342947	HN200365	02/26/200	0120X S CI	842	THEFT	AGG: FINA OTHER	FALSE	FALSE	131	1	2	28	6	1173307	1895342	2007	02/28/201
15	3121541	HK108680	03/28/200	0380X W F	890	THEFT	FROM BUI RESIDENC	FALSE	FALSE	1133	11	24	26	6	1150808	1896753	2003	02/28/201
16	10209850	HY96729	08/24/201	0260X N E	820	THEFT	\$500 AND PARKING	FALSE	FALSE	1432	14	1	22	6	1160867	1917657	2015	02/28/201
17	6866056	HR271603	04/16/200	0890X S ST	860	THEFT	RETAIL TH OTHER	TRUE	FALSE	413	4	8	48	6	1188475	1846086	2009	02/28/201
18	9681782	HK332132	07/03/201	0700X S ST	820	THEFT	\$500 AND CTA PLATI	FALSE	FALSE	623	6	6	44	6	1177660	1852569	2014	02/28/201
19	5166453	HW303959	05/03/200	0530X W C	810	THEFT	OVER \$50K VACANT L	FALSE	FALSE	1522	15	29	25	6	1139693	1897797	2013	02/28/201
20	7762423	H5509246	10/17/201	0580X W I	880	THEFT	PURSE-SN SIDEWALK	FALSE	FALSE	1513	15	29	25	6	1137635	1895420	2010	02/28/201
21	10046030	HY24744	04/24/201	0080X N N	860	THEFT	RETAIL TH DEPARTM	TRUE	FALSE	1833	18	42	8	6	1177338	1906213	2015	02/28/201
22	6733710	HP107451	09/03/200	0900X S ST	860	THEFT	RETAIL TH DEPARTM	FALSE	FALSE	171	1	42	32	6	1176390	1900234	2008	02/28/201

ID	CASE_NUM	DATE	BLOCK	IUCR	PRIMARY	DESCRIPTION	LOCATION	ARREST	DOMESTIC	BEAT	DISTRICT	WARD	COMMUN	FBKCODE	X_COORD	Y_COORD	YEAR	UPDATED
26	0050X S PI	860	THEFT	RETAIL TH GROCERY	TRUE	FALSE	213	2	3	38	6	1180448	1879234	2005	02/28/201	41.82388	-87.0135	[41.823879880]
27	1000X S VA	810	THEFT	OVER \$50K STREET	FALSE	FALSE	2211	22	19	72	6	1160129	1838040	2005	02/28/201	41.71128	-87.6892	[41.711280519]
28	0120X W F	820	THEFT	\$500 AND SMALL RET	FALSE	FALSE	1733	17	35	21	6	1155390	1921064	2004	02/28/201	41.9393	-87.7109	[41.939295821]
29	0760X S CI	860	THEFT	RETAIL TH DEPARTM	TRUE	FALSE	833	8	13	65	6	1145727	1853720	2014	02/28/201	41.75459	-87.7415	[41.754592961]
30	0120X S KI	820	THEFT	\$500 AND BAR OR TA	FALSE	FALSE	2232	12	25	28	6	1171230	1894421	2011	02/28/201	41.86376	-87.6469	[41.863761261]
31	1080X S EF	820	THEFT	\$500 AND RESIDENT	FALSE	FALSE	2233	22	34	49	6	1173239	1832994	2012	02/28/201	41.69715	-87.6413	[41.697154372]
32	0110X S CI	842	THEFT	AGG: FINA OTHER	FALSE	FALSE	131	1	2	28	6	1173307	1895342	2007	02/28/201	41.86824	-87.6392	[41.868242722]
33	0380X W F	890	THEFT	FROM BUI RESIDENC	FALSE	FALSE	1133	11	24	26	6	1150808	1896753	2003	02/28/201	41.87258	-87.7318	[41.872584171]
34	0260X N E	820	THEFT	\$500 AND PARKING	FALSE	FALSE	1432	14	1	22	6	1160867	1917657	2015	02/28/201	41.92974	-87.6843	[41.929743816]
35	0890X S ST	860	THEFT	RETAIL TH OTHER	TRUE	FALSE	413	4	8	48	6	1188475	1846086	2009	02/28/201	41.73273	-87.5851	[41.732730802]
36	0790X S ST	820	THEFT	\$500 AND CTA PLATI	FALSE	FALSE	623	6	6	44	6	1177660	1852569	2014	02/28/201	41.75077	-87.6245	[41.750772111]
37	0550X W C	810	THEFT	OVER \$50K VACANT L	FALSE	FALSE	1522	15	29	25	6	1139693	1897797	2013	02/28/201	41.87566	-87.7626	[41.875659106]
38	0580X W I	880	THEFT	PURSE-SN SIDEWALK	FALSE	FALSE	1513	15	29	25	6	1137635	1895420	2010	02/28/201	41.88917	-87.7702	[41.889173601]
39	0080X N N	860	THEFT	RETAIL TH DEPARTM	TRUE	FALSE	1833	18	42	8	6	1177338	1906213	2015	02/28/201	41.89796	-87.6241	[41.897962931]
40	0000X S ST	860	THEFT	RETAIL TH DEPARTM	FALSE	FALSE	171	1	42	32	6	1176390	1900234	2008	02/28/201	41.8816	-87.6278	[41.881597699]
41	0000X E BE	820	THEFT	\$500 AND STREET	FALSE	FALSE	322	3	20	69	6	1178177	1859290	2007	02/28/201	41.7892	-87.6234	[41.789203083]
42	0360X S AI	860	THEFT	RETAIL TH GROCERY	TRUE	FALSE	912	9	11	59	6	1161654	1880357	2018	02/28/201	41.82737	-87.6826	[41.827375751]
43	0800X S KI	810	THEFT	OVER \$50K STREET	FALSE	FALSE	712	7	16	68	6	1171999	1865041	2009	02/28/201	41.78512	-87.6449	[41.785122891]
44	0440X N E	820	THEFT	\$500 AND ALTHETIC	FALSE	FALSE	1722	17	39	14	6	1148062	1929262	2009	02/28/201	41.96185	-87.731	[41.961845244]
45	0000X W F	870	THEFT	POCKET-P AIRPORT	FALSE	FALSE	2651	18	41	76	6	1106658	1934241	2002	02/28/201	41.97629	-87.9652	[41.976290414]
46	0600X S VA	820	THEFT	\$500 AND STREET	FALSE	FALSE	832	8	13	66	6	1161515	1860616	2013	02/28/201	41.7732	-87.6835	[41.773203999]
47	1040X S KI	810	THEFT	OVER \$50K PARKING	FALSE	FALSE	2233	22	34	73	6	1172824	1835651	2003	02/28/201	41.70445	-87.6428	[41.704454734]
48	0790X S PI	810	THEFT	OVER \$50K STREET	FALSE	FALSE	671	6	17	48	6	1176098	1852235	2008	02/28/201	41.7501	-87.6125	[41.750101935]

2.b)

```
In [1]: import pandas as pd

In [2]: df=pd.read_csv('Desktop/crimedata.csv')
df.head()
```

Out[2]:

	ID	CASE_NUMBER	DATE	BLOCK	IUCR	PRIMARY_TYPE	DESCRIPTION	LOCATION_DESCRIPTION	ARREST	DOMESTIC	...	WARD	COMM
0	3512275	HK587712	08/28/2004 05:50:56 PM	047XX S KEDZIE AVE	890	THEFT	FROM BUILDING	SMALL RETAIL STORE	False	False	...	14.0	
1	3406613	HK456306	06/28/2004 12:40:00 PM	060XX N CENTRAL PARK AVE	820	THEFT	\$500 AND UNDER	OTHER	False	False	...	27.0	
2	8002131	HT233595	04-04-2011 5:45	043XX S WABASH AVE	820	THEFT	\$500 AND UNDER	NURSING HOME/RETIREMENT HOME	False	False	...	3.0	
3	7900209	HT130522	12/30/2010 04:30:00 PM	083XX S KINGSTON AVE	840	THEFT	FINANCIAL ID THEFT) OVER \$300	RESIDENCE	False	False	...	7.0	
4	10400075	HZ138551	02-02- 2016 19:30	033XX W 86TH ST	820	THEFT	\$500 AND UNDER	ALLEY	False	False	...	15.0	

5 rows x 22 columns

Structural analysis:

```
In [3]: df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 533 entries, 0 to 532
Data columns (total 22 columns):
ID                533 non-null int64
CASE_NUMBER       533 non-null object
DATE              533 non-null object
BLOCK             533 non-null object
IUCR              533 non-null object
PRIMARY_TYPE      533 non-null object
DESCRIPTION        533 non-null object
LOCATION_DESCRIPTION 533 non-null object
ARREST            533 non-null bool
DOMESTIC          533 non-null bool
BEAT              533 non-null int64
DISTRICT          533 non-null int64
WARD              490 non-null float64
COMMUNITY_AREA_NUMBER 490 non-null float64
FBI CODE          533 non-null object
X COORDINATE      529 non-null float64
Y COORDINATE      529 non-null float64
YEAR              533 non-null int64
UPDATED ON        533 non-null object
LATITUDE          529 non-null float64
LONGITUDE         529 non-null float64
LOCATION            529 non-null object
dtypes: bool(2), float64(6), int64(4), object(10)
memory usage: 84.4+ KB
```

Statistical analysis:

```
[4]: handle.describe()
```

	ID	BEAT	DISTRICT	WARD	COMMUNITY_AREA_NUMBER	X_COORDINATE	Y_COORDINATE	YEAR	LATITUDE
count	5.330000e+02	533.000000	533.000000	490.000000	490.000000	5.290000e+02	5.290000e+02	533.00000	529.000000
mean	6.226301e+06	1187.037523	11.195122	22.608163	37.542857	1.162876e+06	1.886223e+06	2008.34334	41.843429
std	3.027454e+06	660.592263	6.496631	13.080119	21.409514	1.650197e+04	3.009980e+04	4.94940	0.082798
min	2.114900e+04	111.000000	1.000000	1.000000	1.000000	1.100658e+06	1.814512e+06	2001.00000	41.645796
25%	3.420516e+06	711.000000	6.000000	12.000000	23.000000	1.151715e+06	1.860454e+06	2004.00000	41.772556
50%	6.231712e+06	1111.000000	10.000000	24.000000	30.000000	1.163235e+06	1.891447e+06	2008.00000	41.857737
75%	8.875038e+06	1634.000000	16.000000	32.000000	58.000000	1.174618e+06	1.907711e+06	2012.00000	41.902298
max	1.127717e+07	2533.000000	25.000000	50.000000	77.000000	1.204126e+06	1.951001e+06	2016.00000	42.021178

BEFORE CLEANING:

```
In [4]: file1 = open("Desktop/crimedata.txt","r")
        for i in range(0,534):
            print(file1.readline().split(",")[11])
```

```
12
3
11
25
FALSE
10
10
25
25
4
11
19
19
25
11
3
8
7
2
7
```

```
In [7]: file2=open("untitled1.txt","r")
        print(file2.read())
```

```
ID,CASE_NUMBER,DATE,BLOCK,IUCR,PRIMARY_TYPE,DESCRIPTION,LOCATION_DESCRIPTION,ARREST,DOMESTIC,BEAT,DISTRICT,WARD,COMMUNITY_ARE
A NUMBER,FBICODE,X_COORDINATE,Y_COORDINATE,YEAR,UPDATEDON,LATITUDE,LONGITUDE,LOCATION
3512276,HK587712,08/28/2004 05:58:56 PM,047XX S KEDZIE AVE,890,THEFT,FROM BUILDING,SMALL RETAIL STORE,FALSE,FALSE,911,9,14,5
8,6,1155838,1873050,2004,02-10-2018 15:50,41.8874405,-87.70395585,(41.8874405, -87.703955849)
3406613,HK456306,06/26/2004 12:40:00 PM,009XX N CENTRAL PARK AVE,820,THEFT,$500 AND UNDER,OTHER,FALSE,FALSE,1112,11,27,23,6,1
152206,1906127,2004,02/28/2018 03:56:25 PM,41.89827996,-87.71640551,(41.89827996, -87.716405505)
8002131,HT23595,04-04-2011 5:45,043XX S WABASH AVE,820,THEFT,$500 AND UNDER,MURSING HOME[RETIREMENT HOME,FALSE,FALSE,221,2,
3,38,6,1177436,1876313,2011,02-10-2018 15:50,41.81593313,-87.62464213,(41.81593313, -87.624642127)
7903289,HT133522,12/30/2010 04:30:00 PM,083XX S KINGSTON AVE,840,THEFT,FINAKIAL ID THEFT[ OVER $300,RESIDENCE,FALSE,FALSE,42
3,4,7,46,6,1194622,1858125,2010,02-10-2018 15:50,41.74366532,-87.56246276,(41.74366532, -87.562462756)
10402076,H2138551,02-02-2016 19:30,033XX W 66TH ST,820,THEFT,$500 AND UNDER,ALLEY,FALSE,FALSE,831,8,15,66,6,1155240,1868661,2
016,02-10-2018 15:50,41.7734553,-87.70648847,(41.773455295, -87.706488471)
7732712,H5540106,09/29/2010 07:59:00 AM,006XX W CHICAGO AVE,810,THEFT,OVER $500,PARKING LOT[GARAGE(NON RESID ),FALSE,FALSE,13
23,12,27,24,6,1171668,1905607,2010,02-10-2018 15:50,41.89644677,-87.64493868,(41.89644677, -87.644938678)
10769475,H2534771,11/30/2016 01:15:00 AM,050XX N KEDZIE AVE,810,THEFT,OVER $500,STREET,FALSE,FALSE,1713,17,33,14,6,1154133,19
33314,2016,02-10-2018 15:50,41.97284491,-87.70860008,(41.972844913, -87.708600079)
4494340,HL793243,12/16/2005 04:45:00 PM,005XX E PERSHING RD,860,THEFT,RETAIL THEFT,GROCERY FOOD STORE,TRUE,FALSE,213,2,3,38,
6,1188448,1879234,2005,02/28/2018 03:56:25 PM,41.82387989,-87.61350386,(41.823879885, -87.613503857)
3778925,HL140610,01/28/2005 05:00:00 PM,100XX S WASHTENAW AVE,810,THEFT,OVER $500,STREET,FALSE,FALSE,2211,22,19,72,6,1168129,
1838848,2005,03/20/2018 03:56:35 PM,41.71128851,-87.6801701,(41.711288513, -87.680170067)
```

DATE FORMAT

C
DATE
08/28/2004 05:50:56 PM
06/26/2004 12:40:00 PM
04-04-2011 5:45
12/30/2010 04:30:00 PM
02-02-2016 19:30
09/29/2010 07:59:00 AM
11/30/2016 01:15:00 AM
12/16/2005 04:45:00 PM
01/28/2005 05:00:00 PM
05/13/2004 02:15:00 PM
12/23/2014 02:15:00 PM
02/26/2011 04:00:00 PM
04/15/2012 07:30:00 PM
02/26/2007 04:00:00 PM
01-05-2003 18:30
08/24/2015 07:30:00 PM
04/16/2009 05:30:00 PM
07-04-2014 21:45
06-04-2013 9:00
10/17/2010 06:00:00 PM
04/24/2015 07:55:00 PM
05-06-2008 15:20
07/22/2007 07:25:00 AM
02/14/2018 11:15:00 AM
10-11-2009 19:00
09/13/2009 10:30:00 PM
07-03-2002 20:00
09-07-2013 2:30
08/13/2003 08:00:00 PM

LOCATION DESCRIPTION

H
LOCATION_DESCRIPTION
SMALL RETAIL STORE
OTHER
NURSING HOME, RETIREMENT
RESIDENCE
ALLEY
PARKING LOT, GARAGE (N
STREET
GROCERY FOOD STORE
STREET
SMALL RETAIL STORE
DEPARTMENT STORE
BAR OR TAVERN
RESIDENTIAL YARD (FRO
OTHER
RESIDENCE PORCH, HALL'
PARKING LOT, GARAGE (N
OTHER
CTA PLATFORM
VACANT LOT, LAND
SIDEWALK
DEPARTMENT STORE
DEPARTMENT STORE
STREET
GROCERY FOOD STORE
STREET
ATHLETIC CLUB
AIRPORT, AIRCRAFT
STREET
PARKING LOT, GARAGE (N

AFTER CLEANING

H
LOCATION_DESCRIPTION
SMALL RETAIL STORE
OTHER
NURSING HOME RETIRE
RESIDENCE
ALLEY
PARKING LOT GARAGE(N
STREET
GROCERY FOOD STORE
STREET
SMALL RETAIL STORE
DEPARTMENT STORE
BAR OR TAVERN
RESIDENTIAL YARD (FROI
OTHER
RESIDENCE PORCH HALL
PARKING LOT GARAGE(N
OTHER
CTA PLATFORM
VACANT LOT LAND
SIDEWALK
DEPARTMENT STORE
DEPARTMENT STORE
STREET
GROCERY FOOD STORE
STREET
ATHLETIC CLUB
AIRPORT AIRCRAFT
STREET
PARKING LOT GARAGE(N

TIME CLEANING CODE

```
In [9]: file1 = open("untitled.txt", "r")
arr=[]
for i in range(0,534):
    k=file1.readline().split(",")[2]
    time=k.split(" ")
    if (time[len(time)-1]=="AM" or time[len(time)-1]=="PM"):
        hhmmss=time[1].split(":")
        hour=hhmmss[0]
        if (time[len(time)-1]=="PM"):
            if(hour=="12"):
                hour=hour
            else:
                hour=int(hour)+12
            hhmmss[0]=str(hour)
            time[1]=":".join(hhmmss)
            time.pop()
            time="|".join(time)

        if (time[len(time)-1]=="AM"):
            if(hour=="12"):
                hour="00"
            time.pop()
            time="|".join(time)
        else:
            date=time[0].split('-')
            time[0]="/".join(date)
            time="|".join(time)
        arr.append(time)
        print(time)
```

NEW FORMAT OF DATE:

```
DATE
08/28/2004|17:50:56
06/26/2004|12:40:00
04/04/2011|5:45
12/30/2010|16:30:00
02/02/2016|19:30
09/29/2010|07:59:00
11/30/2016|01:15:00
12/16/2005|16:45:00
01/28/2005|17:00:00
05/13/2004|14:15:00
12/23/2014|14:15:00
02/26/2011|16:00:00
04/15/2012|19:30:00
02/26/2007|16:00:00
01/05/2003|18:30
08/24/2015|19:30:00
04/16/2009|17:30:00
07/04/2014|21:45
06/04/2011|0:00
```

COPYING EVERYTHING TO NEW TEXT FILE:

```
In [10]: file1 = open("untitled.txt", "r")
file2 = open("Desktop/crimedata_afterprep.txt", "w")
for i in range(0,534):
    k=file1.readline().split(",")
    k[2]=arr[i]
    k=', '.join(k)
    file2.write(k)
    print(k)
```

```
ID,CASE_NUMBER,DATE,BLOCK,IUCR,PRIMARY_TYPE,DESCRIPTION,LOCATION_DESCRIPTION,ARREST,DOMESTIC,BEAT,DISTRICT,WARD,COMMUNITY_AREA
A_NUMBER,FBICODE,X_COORDINATE,Y_COORDINATE,YEAR,UPDATEDON,LATITUDE,LONGITUDE,LOCATION
```

```
3512276,HK587712,08/28/2004|17:50:56,047XX S KEDZIE AVE,890,THEFT,FROM BUILDING,SMALL RETAIL STORE,FALSE,FALSE,911,9,14,58,
6,1155838,1873850,2004,02-10-2018 15:50,41.8074405,-87.70395585,(41.8074405,-87.703955849)
```

```
3406613,HK456306,06/26/2004|12:40:00,089XX N CENTRAL PARK AVE,820,THEFT,$500 AND UNDER,OTHER,FALSE,FALSE,1112,11,27,23,6,115
2206,1906127,2804,02/28/2018 03:56:25 PM,41.89827996,-87.71640551,(41.89827996,-87.716405505)
```

```
8002131,HT233595,04/04/2011|5:45,043XX S WABASH AVE,820,THEFT,$500 AND UNDER,NURSING HOME[RETIREMENT HOME,FALSE,FALSE,221,2,
3,38,6,1177436,1876213,2011,02-10-2018 15:50,41.81593313,-87.62464213,(41.81593313,-87.624642127)
```

```
7903289,HT133522,12/30/2010|16:30:00,083XX S KINGSTON AVE,840,THEFT,FINANCIAL ID THEFT| OVER $300,RESIDENCE,FALSE,FALSE,423,
4,7,46,6,1194622,1850125,2010,02-10-2018 15:50,41.74366532,-87.56246276,(41.74366532,-87.562462756)
```

```
10402076,HZ130551,02/02/2016|19:30,033XX W 66TH ST,820,THEFT,$500 AND UNDER,ALLEY,FALSE,FALSE,831,8,15,66,6,1155240,1800601,
2016,02-10-2018 15:50,41.7734553,-87.70648047,(41.773455295,-87.706480471)
```

```
7732712,H5540106,09/29/2010|07:59:00,006XX W CHICAGO AVE,810,THEFT,OVER $500,PARKING LOT[GARAGE(NON RESID)],FALSE,FALSE,132,
3,13,37,38,6,1175660,14006607,2010,02-10-2018 15:50,41.80666672,-87.64403066,(41.80666672,-87.644030670)
```

CONVERTING TEXT FILE TO CSV FOR FUTURE REFERENCE:

```
In [104]: import csv
csv_file = 'Desktop/crimedata.csv'
txt_file = 'untitled.txt'
with open(txt_file, "u") as my_output_file:
    with open(csv_file, "r") as my_input_file:
        [ my_output_file.write(",".join(row)+"\n") for row in csv.reader(my_input_file)]
    my_output_file.close()
```

Model Planning:

The given analysis consists of three modules of data analysis namely location, time and district.

Each module consists of an individual mapper and reducer functions and data visualisation outputs. The functions of the given modules are as follows:

- **LOCATION BASED**

The mapper functions used in this module uses the x-coordinate and y-coordinates of the place where the crime is registered.

The function links each line to its coordinates and passes on as input to the reducer function which in turn counts the occurrence of each crime in each area and displays the results area wise with numbers associated with each crime.

- **TIME BASED**

The mapper functions used in this module uses the time at which the crime is registered. The hour component of the time is extracted and divided into 8 time slots and forwarded to the reducer where further analysis takes place.

The function links each line to its hour component and the reducer function counts the occurrence of each crime in every

time slot displays the results time slot wise with numbers associated with each crime.

- **DISTRICT BASED**

The mapper functions used in this module uses the district of the place where the crime is registered.

The function links each line to its district and passes on as input to the reducer function which in turn counts the occurrence of each crime in each area and displays the results district wise with numbers associated with each crime.

DIAGRAM:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S		
1	ID	CASE_NUM	DATE	BLOCK	IUCR	PRIMARY	DESCRIPT	LOCATION	ARREST	DOMESTIC	BEAT	DISTRICT	WARD	COMMUN	FB	CODE	X_COORD	Y_COORD	YEAR	UPDATED	LAT
2	3512276	HKS87712	08/28/200	047XX S KI	890	THEFT	FROM BUI	SMALL RE	FALSE	FALSE	911	9	14	58	6	1155838	1873050	2004	#####	41	
3	3406613	HK456306	06/26/200	009XX N C	820	THEFT	\$500 AND	OTHER	FALSE	FALSE	1112	11	27	23	6	1152206	1906127	2004	02/28/201	41	
4	8002131	HT233595	#####	043XX S W	820	THEFT	\$500 AND	NURSING	FALSE	FALSE	221	2	3	38	6	1177436	1876313	2011	#####	42	

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S		
1	ID	CASE_NUM	DATE	BLOCK	IUCR	PRIMARY	DESCRIPT	LOCATION	ARREST	DOMESTIC	BEAT	DISTRICT	WARD	COMMUN	FB	CODE	X_COORD	Y_COORD	YEAR	UPDATED	LAT
2	3512276	HKS87712	08/28/200	047XX S KI	890	THEFT	FROM BUI	SMALL RE	FALSE	FALSE	911	9	14	58	6	1155838	1873050	2004	#####	41	

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S		
1	ID	CASE_NUM	DATE	BLOCK	IUCR	PRIMARY	DESCRIPT	LOCATION	ARREST	DOMESTIC	BEAT	DISTRICT	WARD	COMMUN	FB	CODE	X_COORD	Y_COORD	YEAR	UPDATED	LAT
2	3512276	HKS87712	08/28/200	047XX S KI	890	THEFT	FROM BUI	SMALL RE	FALSE	FALSE	911	9	14	58	6	1155838	1873050	2004	#####	41	

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S		
1	ID	CASE_NUM	DATE	BLOCK	IUCR	PRIMARY	DESCRIPT	LOCATION	ARREST	DOMESTIC	BEAT	DISTRICT	WARD	COMMUN	FB	CODE	X_COORD	Y_COORD	YEAR	UPDATED	LAT
2	3512276	HKS87712	08/28/200	047XX S KI	890	THEFT	FROM BUS	SMALL RE	FALSE	FALSE	911	9	14	58	6	1155838	1873050	2004	#####	41	

P	Q	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	
X_COORD	Y_COORD	ID	CASE_NUM	DATE	BLOCK	IUCR	PRIMARY	DESCRIPT	LOCATION	ARREST	DOMESTIC	BEAT	DISTRICT	WARD	COMMUN	FB	CODE
1155838	1873050	2	3512276	HK587712	08/28/200	047XX S KI	890	THEFT	FROM BUI	SMALL RE	FALSE	FALSE	911	9	14	58	6
1152206	1906127	3	3406613	HK456306	06/26/200	009XX N C	820	THEFT	\$500 AND	OTHER	FALSE	FALSE	1112	11	27	23	6
1177436	1876313	4	8002131	HT233595	#####	043XX S W	820	THEFT	\$500 AND	NURSING	FALSE	FALSE	221	2	3	38	6
1194622	1850125	5	7903289	HT133522	12/30/201	083XX S KI	840	THEFT	FINANCIA	RESIDENC	FALSE	FALSE	423	4	7	46	6
		6	10402076	HZ138551	#####	033XX W E	820	THEFT	\$500 AND	ALLEY	FALSE	FALSE	831	8	15	66	6

THEFT=n1

RAPE=n2

BURGLARY=n3

NARCOTICS=n4

ASSAULT=n5

OUTPUT

Pseudocodes:

Location mapper:

- Start
- Input the .csv file
- using a for loop for taking each line for processing
- Split each line using split() function
- Store x and y coordinate from each array which is result of split.
- Map (x,y) coordinate to the line.

Location reducer:

- Start
- Input the from mapper.py
- using a for loop for taking each line for processing
- Split each line using split() function
- Obtain x,y coordinate and tuple in variables.
- Divide according to range of x,y coordinates in 5 areas.
- For each area distinguish crime type and their count.
- In each area, get total crime and type of crime most committed.
- Display results.

District mapper:

- *Start*
- *Input the .csv file*
- *using a for loop for taking each line for processing*
- *Split each line using split() function*
- *Store the district from each array which is result of split.*
- *Map district to the line.*

District reducer:

- *Start*
- *Input the from mapper.py*
- *using a for loop for taking each line for processing*
- *Split each line using split() function*
- *Obtain district and tuple in variables.*
- *Divide according to range of x,y coordinates in 5 areas.*
- *For each district distinguish crime type and their count.*
- *In each district, get total crime and type of crime most committed.*

Display results.

Time mapper:

- *Start*
- *Input the .csv file*
- *using a for loop for taking each line for processing*
- *Split each line using split() function*
- *Take time attribute in a variable and split it to obtain hour of crime.*
- *Store hour from each array which is result of split.*
- *Map hour component to the line.*

Time reducer:

- *Start*
- *Input the from mapper.py*
- *using a for loop for taking each line for processing*
- *Split each line using split() function*
- *Obtain hour component and tuple in variables.*
- *Divide according to range of hour component into 8 time slots.*
- *For each time slot distinguish crime type and count.*
- *In each time slot, get total crime and type of crime most committed.*

Display results.

MODEL BUILDING:

We have built mapper and reducer code identifying various sectors of dataset which analytics would be helpful for the society and Chicago government.

TIME BASED ANALYSIS:

TIME_MAP

#!/usr/bin/python

import sys

flag=0

for input_line in sys.stdin:

line = input_line.strip().split(",")

if (flag==0):

columns=line

flag=1

else:

time = line[2].split(" | ")[1].split(":")

hour = float(time[0])

print "{0}\t{1}".format(hour,str(input_line.strip()))

TIME_REDUCE:

#!/usr/bin/python

coding: utf-8

import sys

quad = []

tab = []

index=-1

count=0

theft="THEFT"

for i in range(8):

quad.append([])

tab.append([])

for input_line in sys.stdin:

line = input_line.strip().split("\t") *# Time \t Tuple*

time = str(line[0]) *#time*

count=count+1

#print 'count'+str(count)

hour = 0 + float(time)

if hour>=0 and hour<3:

quad[0].append(line[1])

elif hour>=3 and hour<6:

quad[1].append(line[1])

elif hour>=6 and hour<9:

quad[2].append(line[1])

elif hour>=9 and hour<12:

quad[3].append(line[1])

```
elif hour>=12 and hour<15:  
    quad[4].append(line[1])  
elif hour>=15 and hour<18:  
    quad[5].append(line[1])  
elif hour>=18 and hour<21:  
    quad[6].append(line[1])  
elif hour>=21 and hour<24:  
    quad[7].append(line[1])
```

```
for i in range(0,8):  
    theft_c = 0  
    criminal_c=0  
    burglary_c=0  
    assault_c=0  
    narcotics_c=0
```

```
if i==0:  
    time_slot="00:00 – 02:59"  
if i==1:  
    time_slot="03:00 – 05:59"  
if i==2:  
    time_slot="06:00 – 08:59"  
if i==3:  
    time_slot="09:00 – 11:59"  
if i==4:  
    time_slot="12:00 – 14:59"  
if i==5:  
    time_slot="15:00 – 17:59"  
if i==6:  
    time_slot="18:00 – 20:59"  
if i==7:  
    time_slot="21:00 – 23:59"
```

```
for j in range(len(quad[i])):  
    crime_array = quad[i][j].strip().split(",")
```

```

crime_type = str(crime_array[5])

if crime_type == "THEFT":
    theft_c = theft_c + 1
elif crime_type=="CRIMINAL DAMAGE":
    criminal_c = criminal_c + 1
elif crime_type=="BURGLARY":
    burglary_c = burglary_c + 1
elif crime_type=="ASSAULT":
    assault_c = assault_c + 1
elif crime_type=="NARCOTICS":
    narcotics_c = narcotics_c + 1

total_crimes = theft_c + criminal_c + burglary_c + assault_c +
narcotics_c

print '\n\nFor Time slot { ' + time_slot + ' } Occurrence of Crime
types are: '

print '-----'

print '\t\tTHEFT\t\t\t: '+str(theft_c)
print '\t\tCRIMINAL DAMAGE\t\t: '+str(criminal_c)
print '\t\tBURGLARY\t\t\t: '+str(burglary_c)
print '\t\tASSAULT\t\t\t: '+str(assault_c)
print '\t\tNARCOTICS\t\t: '+str(narcotics_c)

most    =    max(int(theft_c),int(criminal_c),    int(burglary_c),
int(assault_c), int(narcotics_c))

if most==theft_c:
    most_crime="THEFT"

elif most==criminal_c:
    most_crime="CRIMINAL DAMAGE"

```



```

elif most==burglary_c:
    most_crime="BURGLARY"
elif most==assault_c:
    most_crime="ASSAULT"
elif most==narcotics_c:
    most_crime="NARCOTICS"

print '\n\t\t=> Total Crimes ccurred and reported in above time
slot is : ' + str(total_crimes) + ' reports.'

print '\n\t\t=> Also the most ccurred crime in above time slot is :
' + most_crime + str(most) + ' times.'

tab[i].append(theft_c)
tab[i].append(criminal_c)
tab[i].append(burglary_c)
tab[i].append(assault_c)
tab[i].append(narcotics_c)

print
'\n\n*****
*****'

print
'\n\t\t\tTHEFT\t\t\tCRIMINAL\t\t\tBURGLARY\t\t\tASSAULT\t\t\tNARCOTICS'

for i in range(0,8):

    #for j in range(0,5):

        print
        "\nSlot
"+str(i+1)+"\t\t\t"+str(tab[i][0])+"\t\t\t"+str(tab[i][1])+"\t\t\t"+str(tab[i][
2])+"\t\t\t"+str(tab[i][3])+"\t\t\t"+str(tab[i][4])

print '\n\t\t* Total number of Tuples analysed are : ' + str(count)

```

print

```
'\n*****  
*****\n\n'
```

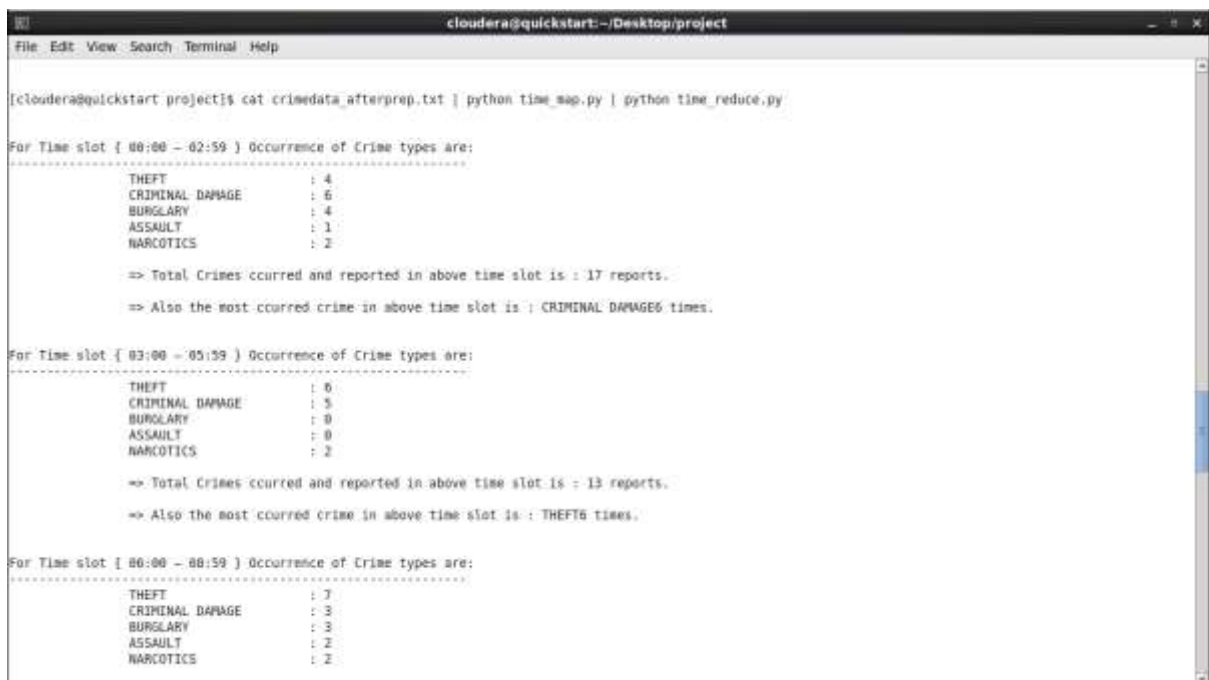
OUTPUT:

LOCAL EXECUTION:

COMMANDS USED:

cd Desktop/project

*cat crimedata_afterprep.txt | python time_map.py | python
time_reduce.py*



```
cloudera@quickstart:~/Desktop/project
File Edit View Search Terminal Help

[cloudera@quickstart project]$ cat crimedata_afterprep.txt | python time_map.py | python time_reduce.py

For Time slot [ 00:00 - 02:59 ] Occurrence of Crime types are:
-----
      THEFT           : 4
    CRIMINAL DAMAGE   : 6
      BURGLARY        : 4
      ASSAULT         : 1
      NARCOTICS       : 2

=> Total Crimes curred and reported in above time slot is : 17 reports.
=> Also the most curred crime in above time slot is : CRIMINAL DAMAGE 6 times.

For Time slot [ 03:00 - 05:59 ] Occurrence of Crime types are:
-----
      THEFT           : 6
    CRIMINAL DAMAGE   : 5
      BURGLARY        : 0
      ASSAULT         : 0
      NARCOTICS       : 2

=> Total Crimes curred and reported in above time slot is : 13 reports.
=> Also the most curred crime in above time slot is : THEFT 6 times.

For Time slot [ 06:00 - 08:59 ] Occurrence of Crime types are:
-----
      THEFT           : 3
    CRIMINAL DAMAGE   : 3
      BURGLARY        : 3
      ASSAULT         : 2
      NARCOTICS       : 2
```

```
Applications Places System cloudera@quickstart:~/Desktop/project
File Edit View Search Terminal Help

For Time slot { 09:00 - 11:59 } Occurrence of Crime types are:
-----
THEFT : 13
CRIMINAL DAMAGE : 6
BURGLARY : 5
ASSAULT : 4
NARCOTICS : 8

=> Total Crimes ccurred and reported in above time slot is : 36 reports.
=> Also the most ccurred crime in above time slot is : THEFT13 times.

For Time slot { 12:00 - 14:59 } Occurrence of Crime types are:
-----
THEFT : 18
CRIMINAL DAMAGE : 7
BURGLARY : 4
ASSAULT : 10
NARCOTICS : 10

=> Total Crimes ccurred and reported in above time slot is : 49 reports.
=> Also the most ccurred crime in above time slot is : THEFT18 times.

For Time slot { 15:00 - 17:59 } Occurrence of Crime types are:
-----
THEFT : 20
CRIMINAL DAMAGE : 7
BURGLARY : 5
ASSAULT : 5
NARCOTICS : 8

=> Total Crimes ccurred and reported in above time slot is : 45 reports.
=> Also the most ccurred crime in above time slot is : THEFT20 times.

cloudera@quickstart:~/Desktop/project [job work (~/.Desktop) ...]
```

```
Applications Places System cloudera@quickstart:~/Desktop/project
File Edit View Search Terminal Help

For Time slot { 18:00 - 20:59 } Occurrence of Crime types are:
-----
THEFT : 23
CRIMINAL DAMAGE : 8
BURGLARY : 7
ASSAULT : 6
NARCOTICS : 11

=> Total Crimes ccurred and reported in above time slot is : 55 reports.
=> Also the most ccurred crime in above time slot is : THEFT23 times.

For Time slot { 21:00 - 23:59 } Occurrence of Crime types are:
-----
THEFT : 15
CRIMINAL DAMAGE : 16
BURGLARY : 2
ASSAULT : 4
NARCOTICS : 11

=> Total Crimes ccurred and reported in above time slot is : 48 reports.
=> Also the most ccurred crime in above time slot is : CRIMINAL DAMAGE16 times.
```

CONSLIDATED OUTPUT:

	THEFT	CRIPINAL	BURGLARY	ASSAULT	NARCOTICS
Slot 1	4	6	4	1	2
Slot 2	6	5	0	0	2
Slot 3	7	3	3	2	2
Slot 4	13	6	5	4	8
Slot 5	18	7	4	10	10
Slot 6	20	7	5	5	8
Slot 7	23	8	7	6	11
Slot 8	15	10	2	4	11
* Total number of Tuples analysed are : 533					

[cloudera@quickstart projec job work (~/.Desktop) - gedit
cloudera@quickstart: job work (~/.Desktop) ...

HADOOP EXECUTION:

```
hadoop jar /usr/lib/hadoop-mapreduce/hadoop-streaming-2.6.0-  
cdh5.13.0.jar \  
-Dmapred.reduce.tasks=1 \  
-input /user/big12/crimedata_afterprep.txt \  
-file Desktop/project/district_map.py Desktop/project/district_reduce.py \  
-mapper "python time_map.py" \  
-reducer "python time_reduce.py" \  
-output /user/time_output/output
```



The screenshot shows a terminal window titled "cloudera@quickstart:~" with a menu bar (File, Edit, View, Search, Terminal, Help). The terminal displays the execution of the Hadoop streaming command. The command is: `hadoop jar /usr/lib/hadoop-mapreduce/hadoop-streaming-2.6.0-cdh5.13.0.jar -Dmapred.reduce.tasks=1 -input /user/big12/crimedata_afterprep.txt -file Desktop/project/time_map.py Desktop/project/time_reduce.py -mapper "python time_map.py" -reducer "python time_reduce.py" -output /user/n4/output`. The output shows a warning about the deprecated `-file` option, followed by log messages from the Hadoop streaming job, including package information, client connections to the Resource Manager, and a `java.lang.InterruptedOperationException` caught by the DFS client. The terminal also shows the status bar at the bottom with "[job work (~/.Desktop) ...]" and "cloudera@quickstart:~".

```
cloudera@quickstart:~  
File Edit View Search Terminal Help  
[cloudera@quickstart ~]$ hadoop jar /usr/lib/hadoop-mapreduce/hadoop-streaming-2  
.6.0-cdh5.13.0.jar \  
> -Dmapred.reduce.tasks=1 \  
> -input /user/big12/crimedata_afterprep.txt \  
> -file Desktop/project/time_map.py Desktop/project/time_reduce.py \  
> -mapper "python time_map.py" \  
> -reducer "python time_reduce.py" \  
> -output /user/n4/output  
19/04/07 13:18:30 WARN streaming.StreamJob: -file option is deprecated, please u  
se generic option -files instead.  
packageJobJar: [Desktop/project/time_map.py, Desktop/project/time_reduce.py] [/u  
sr/lib/hadoop-mapreduce/hadoop-streaming-2.6.0-cdh5.13.0.jar] /tmp/streamjob7622  
956244152526227.jar tmpDir=null  
19/04/07 13:18:30 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0  
:8032  
19/04/07 13:18:30 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0  
:8032  
19/04/07 13:18:30 WARN hdfs.DFSClient: Caught exception  
java.lang.InterruptedOperationException  
    at java.lang.Object.wait(Native Method)  
    at java.lang.Thread.join(Thread.java:1281)  
    at java.lang.Thread.join(Thread.java:1355)  
    at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.closeResponder(DF  
SOutputStream.java:967)  
    at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.endBlock(DFSOutpu  
tStream.java:705)  
    at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.run(DFSOutputStre  
am.java:894)  
19/04/07 13:18:30 INFO mapred.FileInputFormat: Total input paths to process : 1  
19/04/07 13:18:30 WARN hdfs.DFSClient: Caught exception  
java.lang.InterruptedOperationException  
    at java.lang.Object.wait(Native Method)  
    at java.lang.Thread.join(Thread.java:1281)  
    at java.lang.Thread.join(Thread.java:1355)  
    at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.closeResponder(DF  
SOutputStream.java:967)  
    at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.endBlock(DFSOutpu  
tStream.java:705)  
    at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.run(DFSOutputStre
```

```
Applications Places System cloudera@quickstart:~
File Edit View Search Terminal Help
am.java:804)
19/04/07 13:18:38 INFO mapred.FileInputFormat: Total input paths to process : 1
19/04/07 13:18:38 WARN hdfs.DFSClient: Caught exception
java.lang.InterruptedException
    at java.lang.Object.wait(Native Method)
    at java.lang.Thread.join(Thread.java:1281)
    at java.lang.Thread.join(Thread.java:1355)
    at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.closeResponder(DFSOutputStream.java:967)
    at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.endBlock(DFSOutputStream.java:785)
    at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.run(DFSOutputStream.java:804)
19/04/07 13:18:38 INFO mapreduce.JobSubmitter: number of splits:2
19/04/07 13:18:38 INFO Configuration.deprecation: mapred.reduce.tasks is deprecated. Instead, use mapreduce.job.reduces
19/04/07 13:18:39 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_154406383082_0010
19/04/07 13:18:40 INFO impl.YarnClientImpl: Submitted application application_154406383082_0010
19/04/07 13:18:40 INFO mapreduce.Job: The url to track the job: http://quickstart.cloudera:8088/proxy/application_154406383082_0010/
19/04/07 13:18:40 INFO mapreduce.Job: Running job: job_154406383082_0010
19/04/07 13:19:01 INFO mapreduce.Job: Job job_154406383082_0010 running in uber mode : false
19/04/07 13:19:01 INFO mapreduce.Job:  map 0% reduce 0%
19/04/07 13:19:32 INFO mapreduce.Job:  map 100% reduce 0%
19/04/07 13:19:48 INFO mapreduce.Job:  map 100% reduce 100%
19/04/07 13:19:48 INFO mapreduce.Job: Job job_154406383082_0010 completed successfully
19/04/07 13:19:58 INFO mapreduce.Job: Counters: 49
    File System Counters
        FILE: Number of bytes read=123335
        FILE: Number of bytes written=688473
        FILE: Number of read operations=0
        FILE: Number of large read operations=0
        FILE: Number of write operations=0
        HDFS: Number of bytes read=123975
        HDFS: Number of bytes written=3519

[job work (-~/Desktop)... cloudera@quickstart:~
```

hdfs dfs -cat/user/n4/part-00000

```
Applications Places System cloudera@quickstart:~
File Edit View Search Terminal Help
19/04/07 13:19:58 INFO streaming.StreamJob: Output directory: /user/n4/output
[cloudera@quickstart ~]$ hadoop dfs -cat /user/n4/output/part-00000
DEPRECATED: Use of this script to execute hdfs command is deprecated.
Instead use the hdfs command for it.

For Time slot ( 00:00 - 02:59 ) Occurrence of Crime types are:
-----
THEFT                : 4
CRIMINAL DAMAGE      : 6
BURGLARY             : 4
ASSAULT              : 1
MARCOTICS            : 2

=> Total Crimes occurred and reported in above time slot is : 17 reports.
=> Also the most occurred crime in above time slot is : CRIMINAL DAMAGES times.

For Time slot ( 03:00 - 05:59 ) Occurrence of Crime types are:
-----
THEFT                : 6
CRIMINAL DAMAGE      : 5
BURGLARY             : 0
ASSAULT              : 0
MARCOTICS            : 2

=> Total Crimes occurred and reported in above time slot is : 13 reports.
=> Also the most occurred crime in above time slot is : THEFT times.

For Time slot ( 06:00 - 08:59 ) Occurrence of Crime types are:
-----
THEFT                : 7
CRIMINAL DAMAGE      : 3
BURGLARY             : 3
ASSAULT              : 2

[job work (-~/Desktop)... cloudera@quickstart:~
```

HADOOP EXECUTION STATUS :

MapReduce Job job_1554699191533_0005 - Mozilla Firefox

Cloudera Live: Welcom... x MapReduce Job job_155... x

quickstart.cloudera:19000/jobhistory/job/job_1554699191533_0005

Cloudera Hue Hadoop HBase Impala Spark Solr Oozie Cloudera Manager Getting Started

hadoop

MapReduce Job job_1554699191533_0005

Logged in as: drovho

Application

Job

Overview

Counters

Configuration

Map tasks

Reduce tasks

Tools

Job Overview

Job Name: streamjob4102852038249471605.jar

User Name: cloudera

Queue: root.cloudera

State: SUCCEEDED

Uberized: false

Submitted: Sun Apr 07 23:40:45 PDT 2019

Started: Sun Apr 07 23:40:54 PDT 2019

Finished: Sun Apr 07 23:41:16 PDT 2019

Elapsed: 21sec

Diagnostics:

Average Map Time: 10sec

Average Shuffle Time: 5sec

Average Merge Time: 0sec

Average Reduce Time: 0sec

ApplicationMaster

Attempt Number	Start Time	Node	Logs
1	Sun Apr 07 23:40:48 PDT 2019	quickstart.cloudera:8042	logs

Task Type	Total	Complete
Map	2	2
Reduce	1	1

*job work (~/.Desktop)... [cloudera@quickstart:~] MapReduce job job_15... cloudera@quickstart:~

Elapsed: 21sec				
Diagnostics:				
Average Map Time 10sec				
Average Shuffle Time 5sec				
Average Merge Time 0sec				
Average Reduce Time 0sec				
ApplicationMaster				
Attempt Number	Start Time		Node	Logs
1	Sun Apr 07 23:40:48 PDT 2019		quickstart.cloudera:8042	logs
Task Type	Total		Complete	
Map	2		2	
Reduce	1		1	
Attempt Type	Failed	Killed	Successful	
Maps	0	0	2	
Reduces	0	0	1	

DISTRICT_MAP:

#!/usr/bin/python

import sys

flag=0

for input_line in sys.stdin:

line = input_line.strip().split(",")

if (flag ==0):

columns=line

flag=1

else:

disno = float(line[11])

print "{0}\t{1}".format(disno,str(input_line.strip()))

DISTRICT_REDUCE:

#!/usr/bin/python

coding: utf-8

import sys

tab = []

quad = []

count=0

for i in range(5):

quad.append([])

tab.append([])

for input_line in sys.stdin:

line = input_line.strip().split("\t")

disno = float(line[0])

count=count+1

if disno>=1 and disno<5:

quad[0].append(line[1])

elif disno>=5 and disno<10:

quad[1].append(line[1])

elif disno>=10 and disno<15:

quad[2].append(line[1])

elif disno>=15 and disno<20:

quad[3].append(line[1])

```
elif disno>=20 and disno<25:  
    quad[4].append(line[1])
```

```
for i in range(0,5):
```

```
    theft_c = 0
```

```
    criminal_c=0
```

```
    burglary_c=0
```

```
    assault_c=0
```

```
    narcotics_c=0
```

```
for j in range(len(quad[i])):
```

```
    crime_array = quad[i][j].strip().split(',')
```

```
    crime_type = str(crime_array[5])
```

```
    if crime_type == 'THEFT':
```

```
        theft_c = theft_c + 1
```

```
    elif crime_type=='CRIMINAL DAMAGE':
```

```
        criminal_c = criminal_c + 1
```

```
    elif crime_type=='BURGLARY':
```

```
        burglary_c = burglary_c + 1
```

```
    elif crime_type=='ASSAULT':
```

```
        assault_c = assault_c + 1
```

```
    elif crime_type=='NARCOTICS':
```

```
        narcotics_c = narcotics_c + 1
```

```
total_crimes = theft_c + criminal_c + burglary_c + assault_c +  
narcotics_c
```

```
print '\n\nFor sector ' + str(i+1) + ' – Occurrence of Crime types are:  
,
```

```
print '-----'\n  
print '\t\tTHEFT\t\t\t: '+str(theft_c)\nprint '\t\tCRIMINAL DAMAGE\t\t: '+str(criminal_c)\nprint '\t\tBURGLARY\t\t: '+str(burglary_c)\nprint '\t\tASSAULT\t\t\t: '+str(assault_c)\nprint '\t\tNARCOTICS\t\t: '+str(narcotics_c)\n  
most = max(int(theft_c),int(criminal_c), int(burglary_c),  
int(assault_c), int(narcotics_c))
```

```
if most==theft_c:  
    most_crime='THEFT'\n  
elif most==criminal_c:  
    most_crime='CRIMINAL DAMAGE'\n  
elif most==burglary_c:  
    most_crime='BURGLARY'\n  
elif most==assault_c:  
    most_crime='ASSAULT'\n  
elif most==narcotics_c:
```

```
most_crime='NARCOTICS'
```

```
print '\n\t\t=> Total Crimes ccurred and reported in above sector is  
: ' + str(total_crimes) + ' reports.'
```

```
print '\n\t\t=> Also the most ccurred crime in above sector is : ' +  
most_crime + str(most) + ' times.'
```

```
tab[i].append(theft_c)
```

```
tab[i].append(criminal_c)
```

```
tab[i].append(burglary_c)
```

```
tab[i].append(assault_c)
```

```
tab[i].append(narcotics_c)
```

```
print
```

```
'\n\n*****  
*****'
```

```
print
```

```
'\n\t\t\tTHEFT\t\t\tCRIMINAL\t\t\tBURGLARY\t\t\tASSAULT\t\t\tNARCOTICS'
```

```
for i in range(0,5):
```

```
    print
```

```
        "\nSector
```

```
"+str(i+1)+"\t\t\t"+str(tab[i][0])+"\t\t\t"+str(tab[i][1])+"\t\t\t"+str(tab[i][  
2])+"\t\t\t"+str(tab[i][3])+"\t\t\t"+str(tab[i][4])
```

```
print '\n\t\t* Total number of Tuples analysed are : ' + str(count)
```

[illegible]

COMMANDS USED:

```
cloudera@quickstart:~/Desktop/project
File Edit View Search Terminal Help

[cloudera@quickstart project]$ cat crimedata_afterprep.txt | python district_map.py | python district_reduce.py
\

For sector 1 - Occurrence of Crime types are:
-----
THEFT                : 18
CRIMINAL DAMAGE      : 8
BURGLARY             : 7
ASSAULT              : 8
NARCOTICS            : 8

=> Total Crimes occurred and reported in above sector is : 47 reports.

=> Also the most occurred crime in above sector is : THEFT 18 times.

For sector 2 - Occurrence of Crime types are:
-----
THEFT                : 28
CRIMINAL DAMAGE      : 22
BURGLARY             : 9
ASSAULT              : 11
NARCOTICS            : 14

=> Total Crimes occurred and reported in above sector is : 84 reports.

=> Also the most occurred crime in above sector is : THEFT 28 times.

For sector 3 - Occurrence of Crime types are:
-----
THEFT                : 24
CRIMINAL DAMAGE      : 12
BURGLARY             : 4
ASSAULT              : 6
NARCOTICS            : 18

=> Total Crimes occurred and reported in above sector is : 64 reports.
```

```
Applications Places System cloudera@quickstart:~/Desktop/project
File Edit View Search Terminal Help

For sector 1 - Occurrence of Crime types are:
-----
THEFT : 18
CRIMINAL DAMAGE : 6
BURGLARY : 7
ASSAULT : 8
NARCOTICS : 8

=> Total Crimes curred and reported in above sector is : 47 reports.
=> Also the most curred crime in above sector is : THEFT18 times.

For sector 2 - Occurrence of Crime types are:
-----
THEFT : 28
CRIMINAL DAMAGE : 22
BURGLARY : 9
ASSAULT : 11
NARCOTICS : 14

=> Total Crimes curred and reported in above sector is : 84 reports.
=> Also the most curred crime in above sector is : THEFT28 times.

For sector 3 - Occurrence of Crime types are:
-----
THEFT : 24
CRIMINAL DAMAGE : 12
BURGLARY : 4
ASSAULT : 6
NARCOTICS : 18

=> Total Crimes curred and reported in above sector is : 64 reports.
=> Also the most curred crime in above sector is : THEFT24 times.

cloudera@quickstart:~/Desktop/project
```

```
Applications Places System cloudera@quickstart:~/Desktop/project
File Edit View Search Terminal Help

For sector 4 - Occurrence of Crime types are:
-----
THEFT : 22
CRIMINAL DAMAGE : 13
BURGLARY : 4
ASSAULT : 5
NARCOTICS : 10

=> Total Crimes curred and reported in above sector is : 54 reports.
=> Also the most curred crime in above sector is : THEFT22 times.

For sector 5 - Occurrence of Crime types are:
-----
THEFT : 10
CRIMINAL DAMAGE : 2
BURGLARY : 2
ASSAULT : 1
NARCOTICS : 2

=> Total Crimes curred and reported in above sector is : 17 reports.
=> Also the most curred crime in above sector is : THEFT10 times.

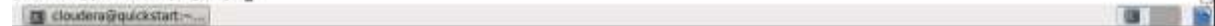
cloudera@quickstart:~/Desktop/project
```

CONSOLIDATED OUTPUT:

```
*****
      THEFT      CRIMINAL      BURGLARY      ASSAULT      NARCOTICS
Sector 1         18          6          7          8          8
Sector 2         28         22          9         11         14
Sector 3         24         12          4          6         18
Sector 4         22         13          4          5         10
Sector 5         10          2          2          1          2

* Total number of tuples analysed are : 533
*****

[cloudera@quickstart project]$
```

A terminal window with a title bar showing 'cloudera@quickstart'. The window contains the command prompt and the output of a script, which is a table of crime statistics for five sectors. The output is enclosed in a block of asterisks.

HADOOP EXECUTION:

`hadoop jar /usr/lib/hadoop-mapreduce/hadoop-streaming-2.6.0-cdh5.13.0.jar \`

`-Dmapred.reduce.tasks=1 \`

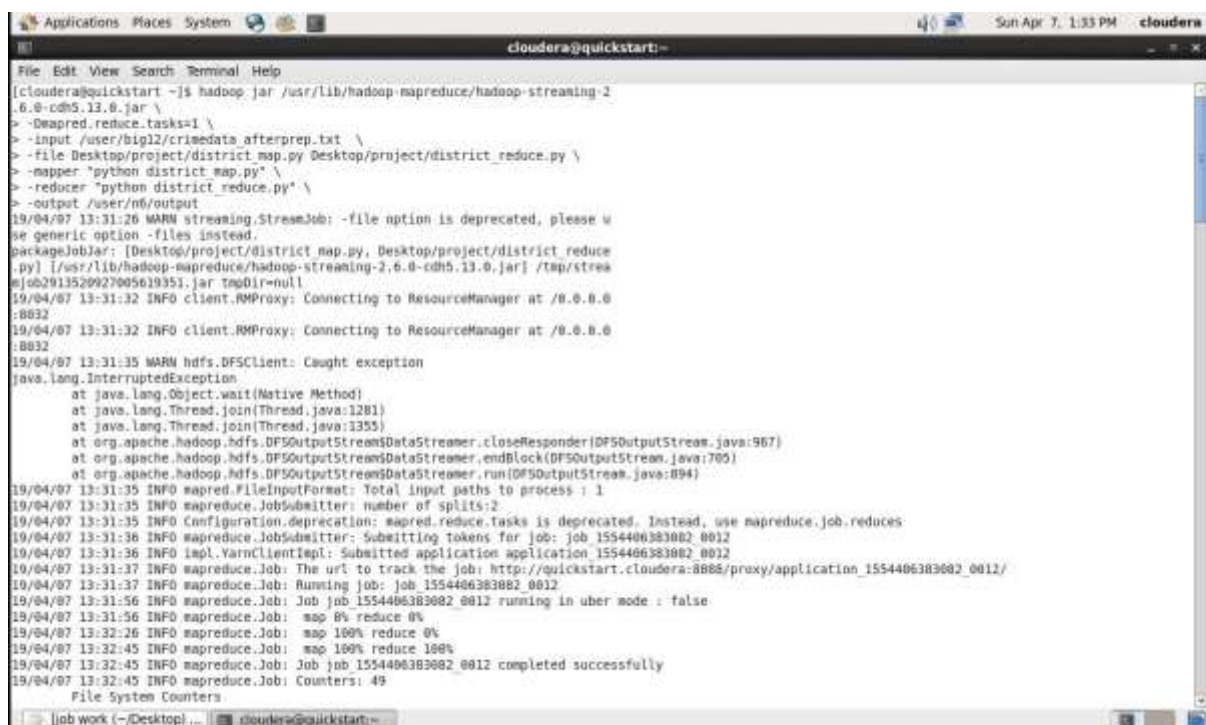
`-input /user/big12/crimedata_afterprep.txt \`

`-file Desktop/project/district_map.py
Desktop/project/district_reduce.py \`

`-mapper "python district_map.py" \`

`-reducer "python district_reduce.py" \`

`-output /user/n6/output`



```
cloudera@quickstart:~$ hadoop jar /usr/lib/hadoop-mapreduce/hadoop-streaming-2.6.0-cdh5.13.0.jar \
-Dmapred.reduce.tasks=1 \
-input /user/big12/crimedata_afterprep.txt \
-file Desktop/project/district_map.py Desktop/project/district_reduce.py \
-mapper "python district_map.py" \
-reducer "python district_reduce.py" \
-output /user/n6/output
19/04/07 13:31:26 WARN streaming.StreamJob: -file option is deprecated, please use generic option -files instead.
packageJobJar: [Desktop/project/district_map.py, Desktop/project/district_reduce.py] [/usr/lib/hadoop-mapreduce/hadoop-streaming-2.6.0-cdh5.13.0.jar] /tmp/streamjob2913520927005619351.jar tmpDir=null
19/04/07 13:31:32 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
19/04/07 13:31:32 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
19/04/07 13:31:35 WARN hdfs.DFSClient: Caught exception
java.lang.InterruptedException
    at java.lang.Object.wait(Native Method)
    at java.lang.Thread.join(Thread.java:1281)
    at java.lang.Thread.join(Thread.java:1355)
    at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.closeResponder(DFSOutputStream.java:967)
    at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.endBlock(DFSOutputStream.java:705)
    at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.run(DFSOutputStream.java:894)
19/04/07 13:31:35 INFO mapred.FileInputFormat: Total input paths to process : 1
19/04/07 13:31:35 INFO mapreduce.JobSubmitter: number of splits:2
19/04/07 13:31:35 INFO Configuration.deprecation: mapred.reduce.tasks is deprecated. Instead, use mapreduce.job.reduces
19/04/07 13:31:36 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1554406383082_0012
19/04/07 13:31:36 INFO impl.YarnClientImpl: Submitted application application_1554406383082_0012
19/04/07 13:31:37 INFO mapreduce.Job: The url to track the job: http://quickstart.cloudera:8888/proxy/application_1554406383082_0012/
19/04/07 13:31:37 INFO mapreduce.Job: Running job: job_1554406383082_0012
19/04/07 13:31:56 INFO mapreduce.Job: Job job_1554406383082_0012 running in uber mode : false
19/04/07 13:31:56 INFO mapreduce.Job:  map 0% reduce 0%
19/04/07 13:32:26 INFO mapreduce.Job:  map 100% reduce 0%
19/04/07 13:32:45 INFO mapreduce.Job:  map 100% reduce 100%
19/04/07 13:32:45 INFO mapreduce.Job: Job job_1554406383082_0012 completed successfully
19/04/07 13:32:45 INFO mapreduce.Job: Counters: 49
File System Counters
```



```
cloudera@quickstart:~$  
File Edit View Search Terminal Help  
19/04/07 22:27:03 INFO mapreduce.Job: map 100% reduce 0%  
19/04/07 22:27:12 INFO mapreduce.Job: map 100% reduce 100%  
19/04/07 22:27:12 INFO mapreduce.Job: Job job_1554699191533_0003 completed successfully  
19/04/07 22:27:12 INFO mapreduce.Job: Counters: 50  
File System Counters  
FILE: Number of bytes read=123217  
FILE: Number of bytes written=688357  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=123975  
HDFS: Number of bytes written=2210  
HDFS: Number of read operations=9  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
Job Counters  
Killed map tasks=1  
Launched map tasks=2  
Launched reduce tasks=1  
Data-local map tasks=2  
Total time spent by all maps in occupied slots (ms)=22762  
Total time spent by all reduces in occupied slots (ms)=6962  
Total time spent by all map tasks (ms)=22762  
Total time spent by all reduce tasks (ms)=6962  
Total vcore-milliseconds taken by all map tasks=22762  
Total vcore-milliseconds taken by all reduce tasks=6962  
Total megabyte-milliseconds taken by all map tasks=23306288  
Total megabyte-milliseconds taken by all reduce tasks=7129088  
Map-Reduce Framework  
Map input records=534  
Map output records=532  
Map output bytes=121606  
Map output materialized bytes=123223  
Input split bytes=236  
Combine input records=0  
Combine output records=0  
Reduce input groups=22  
Reduce shuffle bytes=123223  
Reduce input records=532
```

```
Applications Places System cloudera@quickstart:~$  
Sun Apr 7, 1:35 PM cloudera  
File Edit View Search Terminal Help  
CPU time spent (ms)=4910  
Physical memory (bytes) snapshot=580120576  
Virtual memory (bytes) snapshot=4522307504  
Total committed heap usage (bytes)=392372224  
Shuffle Errors  
BAD_ID=0  
CONNECTION=0  
IO_ERROR=0  
WRONG_LENGTH=0  
WRONG_MAP=0  
WRONG_REDUCE=0  
File Input Format Counters  
Bytes Read=123739  
File Output Format Counters  
Bytes Written=2210  
19/04/07 13:32:45 INFO streaming.StreamJob: Output directory: /user/n6/output  
[cloudera@quickstart ~]$ hadoop dfs -cat /user/n6/output/part-00000  
DEPRECATED: Use of this script to execute hdfs command is deprecated.  
Instead use the hdfs command for it.  
  
For sector 1 - Occurrence of Crime types are:  
-----  
THEFT : 18  
CRIMINAL DAMAGE : 6  
BURGLARY : 7  
ASSAULT : 8  
NARCOTICS : 8  
  
=> Total Crimes occurred and reported in above sector is : 47 reports.  
=> Also the most occurred crime in above sector is : THEFT 18 times.  
  
For sector 2 - Occurrence of Crime types are:  
-----  
THEFT : 28  
CRIMINAL DAMAGE : 22  
[job work (~/Desktop) ... cloudera@quickstart:~$
```

LOCATION_MAP:

#!/usr/bin/python

import sys

cnt=0

for input_line in sys.stdin:

line = input_line.strip().split(",")

if (flag ==0):

columns=line

flag=1

else:

x = float(line[15])

y = float(line[16])

print "{0}\t{1}\t{2}".format(x,y,str(input_line.strip()))

LOCATION_REDUCE:

#!/usr/bin/python

import sys

tab = []

quad = []

count=0

for i in range(5):

quad.append([])

tab.append([])

for input_line in sys.stdin:

line = input_line.strip().split("\t") # X-coordinate \t Y-

coordinate \t Tuple

x = float(line[0]) # x-coordinate

y = float(line[1]) # y-coordinate

count=count+1

if x>=1095000 and x<1117000 and y>=1836000 and y<1856600:

quad[0].append(line[2])

*elif x>=1117000 and x<1139000 and y>=1856600 and
y<1877200:*

quad[1].append(line[2])

*elif x>=1139000 and x<1161000 and y>=1877200 and
y<1897800:*

```

        quad[2].append(line[2])

    elif x>=1161000 and x<1183000 and y>=1897800 and
y<1918400:

        quad[3].append(line[2])

    elif x>=1183000 and x<1205000 y>=1918400 and y<1939000:

        quad[4].append(line[2])

for i in range(0,5):

    theft_c = 0
    criminal_c=0
    burglary_c=0
    assault_c=0
    narcotics_c=0
    for j in range(len(quad[i])):
        crime_array = quad[i][j].strip().split(',')
        crime_type = str(crime_array[5])
        if crime_type == 'THEFT':
            theft_c = theft_c + 1
        elif crime_type=='CRIMINAL DAMAGE':
            criminal_c = criminal_c + 1
        elif crime_type=='BURGLARY':
            burglary_c = burglary_c + 1
        elif crime_type=='ASSAULT':
            assault_c = assault_c + 1

```

```
elif crime_type=='NARCOTICS':
```

```
    narcotics_c = narcotics_c + 1
```

```
total_crimes = theft_c + criminal_c + burglary_c + assault_c +  
narcotics_c
```

```
print '\n\nFor Area ' + str(i+1) + ' – Occurrence of Crime types  
are: '
```

```
print '-----'
```

```
print '\t\tTHEFT\t\t\t: '+str(theft_c)
```

```
print '\t\tCRIMINAL DAMAGE\t\t: '+str(criminal_c)
```

```
print '\t\tBURGLARY\t\t: '+str(burglary_c)
```

```
print '\t\tASSAULT\t\t\t: '+str(assault_c)
```

```
print '\t\tNARCOTICS\t\t: '+str(narcotics_c)
```

```
most = max(int(theft_c),int(criminal_c), int(burglary_c),  
int(assault_c), int(narcotics_c))
```

```
if most==theft_c:
```

```
    most_crime='THEFT'
```

```
elif most==criminal_c:
```

```
    most_crime='CRIMINAL DAMAGE'
```

```
elif most==burglary_c:
```

```
    most_crime='BURGLARY'
```

```
elif most==assault_c:
```

```
    most_crime='ASSAULT'
```

```

elif most==narcotics_c:

    most_crime='NARCOTICS'

    print '\n\t\t=> Total Crimes ccurred and reported in above Area
is : ' + str(total_crimes) + ' reports.'

    Print '\n\t\t=> Also the most ccurred crime in above Area is : ' +
most_crime + ' – ' +str(most) + ' times.'

    tab[i].append(theft_c)
    tab[i].append(criminal_c)
    tab[i].append(burglary_c)
    tab[i].append(assault_c)
    tab[i].append(narcotics_c)

print

'\n\n*****

*****'

print

'\n\t\t\tTHEFT\t\tCRIMINAL\tBURGLARY\tASSAULT\t\tNARCOTICS'

for i in range(0,5):

    print                                                         "\nArea

"+str(i+1)+"\t\t\t"+str(tab[i][0])+"\t\t"+str(tab[i][1])+"\t\t"+str(tab[i][
2])+"\t\t"+str(tab[i][3])+"\t\t"+str(tab[i][4])

print

'\n\n*****

*****'

print '\n\t* Total number of Tuples analysed are : ' + str(count)

```

print

```
'\n*****
```

```
*****\n\n'
```

```
#print "{0}\t{1}".format(output, 1)
```

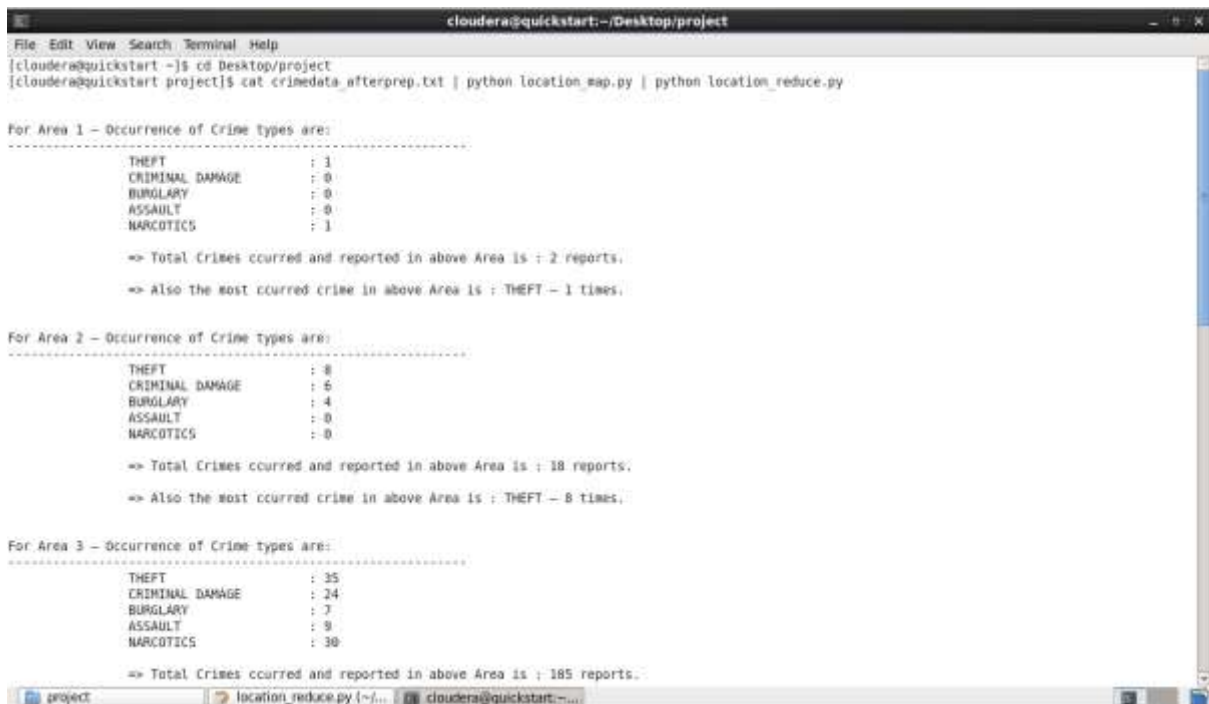
```
#print "{0}\t{1}".format(output, 1)
```

LOCAL EXECUTION:

COMMANDS USED:

cd Desktop/project

cat crimedata_afterprep.txt | python location_map.py | python location_reduce.py



```
cloudera@quickstart:~/Desktop/project
File Edit View Search Terminal Help
[cloudera@quickstart ~]$ cd Desktop/project
[cloudera@quickstart project]$ cat crimedata_afterprep.txt | python location_map.py | python location_reduce.py

For Area 1 - Occurrence of Crime Types are:
-----
THEFT          : 1
CRIMINAL DAMAGE : 0
BURGLARY       : 0
ASSAULT        : 0
NARCOTICS      : 1

=> Total Crimes occurred and reported in above Area is : 2 reports.
=> Also the most occurred crime in above Area is : THEFT - 1 times.

For Area 2 - Occurrence of Crime Types are:
-----
THEFT          : 8
CRIMINAL DAMAGE : 6
BURGLARY       : 4
ASSAULT        : 0
NARCOTICS      : 0

=> Total Crimes occurred and reported in above Area is : 18 reports.
=> Also the most occurred crime in above Area is : THEFT - 8 times.

For Area 3 - Occurrence of Crime Types are:
-----
THEFT          : 35
CRIMINAL DAMAGE : 24
BURGLARY       : 7
ASSAULT        : 9
NARCOTICS      : 30

=> Total Crimes occurred and reported in above Area is : 185 reports.
```

```
cloudera@quickstart:~/Desktop/project
File Edit View Search Terminal Help
ASSAULT : 9
NARCOTICS : 30
=> Total Crimes occurred and reported in above Area is : 185 reports.
=> Also the most occurred crime in above Area is : THEFT - 35 times.

For Area 4 - Occurrence of Crime types are:
-----
THEFT : 56
CRIMINAL DAMAGE : 21
BURGLARY : 13
ASSAULT : 18
NARCOTICS : 17
=> Total Crimes occurred and reported in above Area is : 125 reports.
=> Also the most occurred crime in above Area is : THEFT - 56 times.

For Area 5 - Occurrence of Crime types are:
-----
THEFT : 6
CRIMINAL DAMAGE : 7
BURGLARY : 6
ASSAULT : 5
NARCOTICS : 5
=> Total Crimes occurred and reported in above Area is : 29 reports.
=> Also the most occurred crime in above Area is : CRIMINAL DAMAGE - 7 times.
```

Consolidated output:

	THEFT	CRIMINAL	BURGLARY	ASSAULT	NARCOTICS
Area 1	1	0	0	0	1
Area 2	8	6	4	0	0
Area 3	35	24	7	9	30
Area 4	56	21	13	18	17
Area 5	6	7	6	5	5

* Total number of Tuples analysed are : 529