**Info 7250 – Engineering Big-Data System**

**Analysis Of Crimes In Chicago Using Hadoop Environment Tools**

** **

**A close up of a logo

Description generated with very high confidence A close up of a logo

Description generated with very high confidence**

**Submitted By:**

**Shantanu Deosthale**

**001851612**

**Project Description**

This project deals with the analysis of the crimes in Chicago between years 2006-2016. The project also deals with the number of Big-Data technologies like Hadoop, Pig and Hive. To visualize the results to obtain a clear understanding of crime patterns according to Month, Week, Crime Type, and location prone to crime.

The Dataset can be downloaded from the below link:

<https://www.kaggle.com/currie32/crimes-in-chicago>

The following Dataset contains following:

The dataset contains more than 6,000,000 records/rows of data. Wrote custom python program to merge Datasets from year 2006-2016.

**Key Technologies Used:**

* **Hadoop Map Reduce**
* **Pig**
* **Hive**
* **Amazon EMR (Elastic Map Reduce)**
* **PowerBI for visualization**

**Hadoop Map-Reduce to find number crime per week**

The Date column contains “04/02/2006 01:00:00 PM”. Wrote MapReduce program to parse Week from the Date column.

**Reducer**

A screenshot of a cell phone

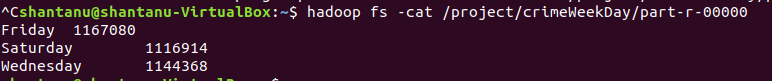
Description generated with very high confidence

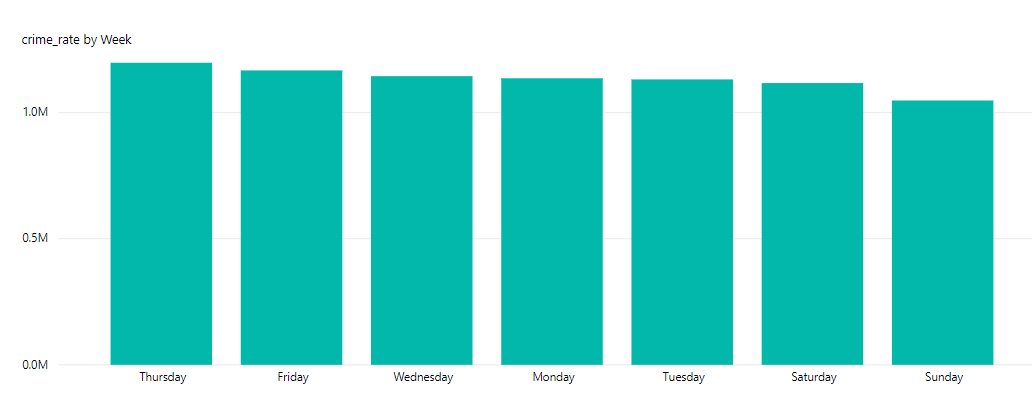
**Mapper**

A screenshot of a cell phone

Description generated with very high confidence

The figure shows the mapper and the reducer function.





**Hadoop MapReduce to find crime based upon Primary Type**

The primary type consists of crime type happened between year 2006-2017. We can find out the maximum number of crimes that took place was from **THEFT.**

**Mapper**

A screenshot of a cell phone

Description generated with very high confidence

**Reducer**

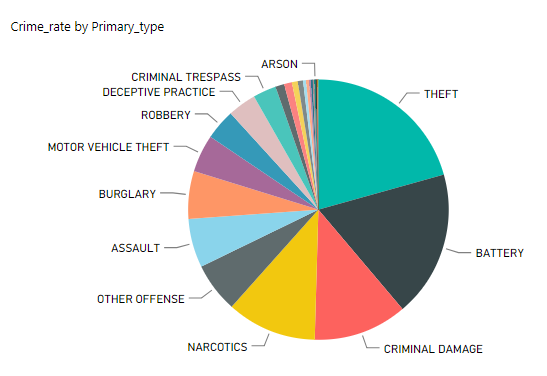
A screenshot of a cell phone

Description generated with high confidence

The mapper and reducer of the primary type

A screenshot of a cell phone

Description generated with high confidence



**Hadoop MapReduce to find count of arrest made or not in each district.**

**Performed Secondary Sorting** to find every type of arrest per district. Performed Job Chaining and used the file obtained from 1st job to identify the number of false and true arrest per district.

A screenshot of a cell phone

Description generated with very high confidence

**Output-**

A screenshot of a cell phone

Description generated with very high confidence

A screenshot of a cell phone

Description generated with very high confidence

**Hadoop MapReduce Binning Pattern to find Crimes Between specific years**

* **Binning pattern to identify crimes between 2001-2017 with a four years gap**
* **Used MultipleOutputFormat to save files.**

**Mapper**

**A screenshot of a social media post

Description generated with very high confidence**

**Hadoop MapReduce Top10 Crime Location**

Used top10 summarization to find the top10 crime location in Chicago. Used Job Chaining to find the count of crime per location and the used the output file to find the top10 unsafe location to live in Chicago.

A screenshot of a cell phone

Description generated with very high confidence

**Output**

A screenshot of a cell phone

Description generated with high confidenceA screenshot of a cell phone

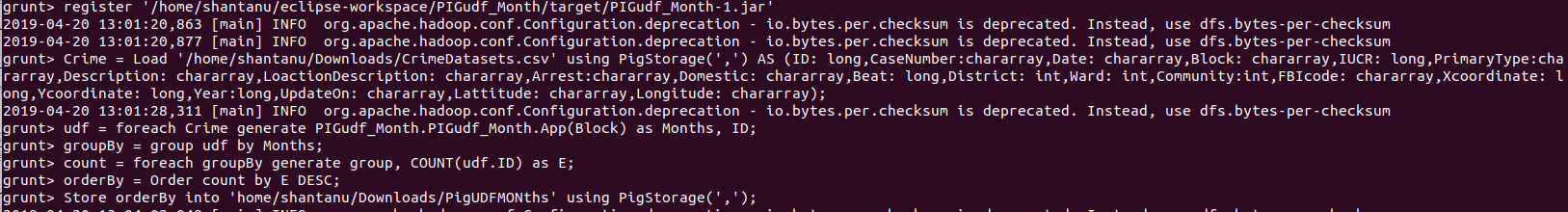
Description generated with high confidence

**PIG to find the count of crimes recorded per month using PIG UDF.**

**A screenshot of a cell phone

Description generated with very high confidence**

**PIG SCRIPT**

****

**Output**

**A screen shot of a social media post

Description generated with very high confidence**

**A picture containing sky

Description generated with very high confidence**

**PIG to find top 10 hours with most crime using PIG UDF**

**A screenshot of a social media post

Description generated with very high confidence**

**A screenshot of a computer

Description generated with high confidence**

**Output**

**A picture containing wall, bottle

Description generated with high confidence**

**A close up of a map

Description generated with high confidence**

**HIVE**

**Top 10 Xcoordinate and Ycoordinate of location where Crime Type was THEFT**

**A screenshot of a cell phone

Description generated with very high confidence**

**Hive Query for number of false and true arrest crime for year 2011**

**A screenshot of a cell phone

Description generated with very high confidence**

**HIVE Query for distinct Crime Type**

**A screenshot of text

Description generated with very high confidence**

**Elastic Map Reduce (EMR)**

**Performed two jobs on AWS EMR using custom jar and saved the output in S3.**

**Secondary Sorting Job**

**A screenshot of a social media post

Description generated with very high confidence**

**Output**

**A screenshot of a cell phone

Description generated with very high confidence**

**Top 10 Summarization Job**

**A screenshot of a computer

Description generated with very high confidence**

**Output**

**A screenshot of a computer

Description generated with very high confidence**