



BIG DATA
DEVELOPMENT

Project 1

ACADGILD

Project 1 - USA Crime Analysis

Table of Contents

1. Introduction3
2. Objective3
3. Prerequisites3
4. Associated Data Files3
5. Problem Statement3
6. Approximate Time to Complete Task3

1. Introduction

This dataset contains attributes related to crimes taking place in various areas like type of crime, FBI code related to that criminal case, arrest frequency, location of crime etc.

2. Objective

3. Prerequisites

You should have Hadoop cluster installed in your system.

4. Associated Data Files

<https://drive.google.com/file/d/0B1QaXx7tpw3SaUJHOHBZcIBXWG8/view?usp=sharing>

Dataset Description:

ID,Case Number,Date,Block,IUCR,Primary Type,Description,Location
Description,Arrest,Domestic,Beat,District,Ward,Community Area,FBI Code,X Coordinate,Y
Coordinate,Year,Updated On,Latitude,Longitude,Location

5. Problem Statement

1. Write a mapreduce/pig program to calculate the number of cases investigated under each FBI code
2. Write a mapreduce/pig program to calculate the number of cases investigated under FBI code 32.
3. Write a mapreduce/pig program to calculate the number of arrests in theft district wise.
4. Write a mapreduce/pig program to calculate the number of arrests done between october 2014 and october 2015.

6. Approximate Time to Complete Task