**Project 1 - USA Crime Analysis**

**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.

**Dataset Description:**

|  |  |
| --- | --- |
| Column 1 | ID |
| Column 2 | Case Number |
| Column 3 | Date |
| Column 4 | Block |
| Column 5 | IUCR |
| Column 6 | Primary Type |
| Column 7 | Description |
| Column 8 | Location Description |
| Column 9 | Arrest |
| Column 10 | Domestic |
| Column 11 | Beat |
| Column 12 | District |
| Column 13 | Ward |
| Column 14 | Community Area |
| Column 15 | FBICode |
| Column 16 | X Coordinate |
| Column 17 | Y Coordinate |
| Column 18 | Year |
| Column 19 | Updated On |
| Column 20 | Latitude |
| Column 21 | Longitude |
| Column 22 | Location |

**Problem Statement:**

1. Write a mapreduce and pig program to calculate the number of cases investigated under each FBI code

2. Write a mapreduce and pig program to calculate the number of cases investigated under FBI code 32. 3. Write a mapreduce and pig program to calculate the number of arrests in theft district wise.

4. Write a mapreduce and pig program to calculate the number of arrests done between October 2014 and October 2015.

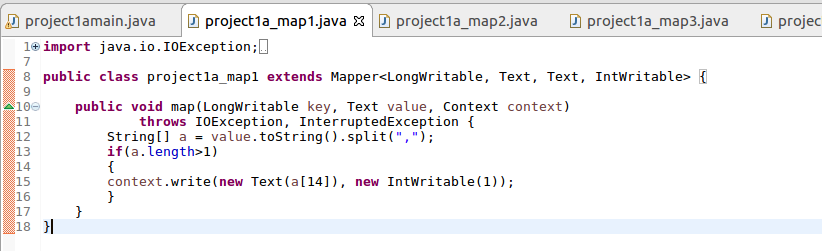
**MAP REDUCE APPROACH**

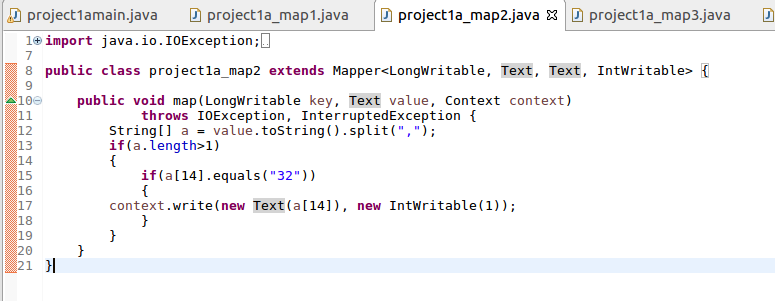
**DRIVER CLASS:**



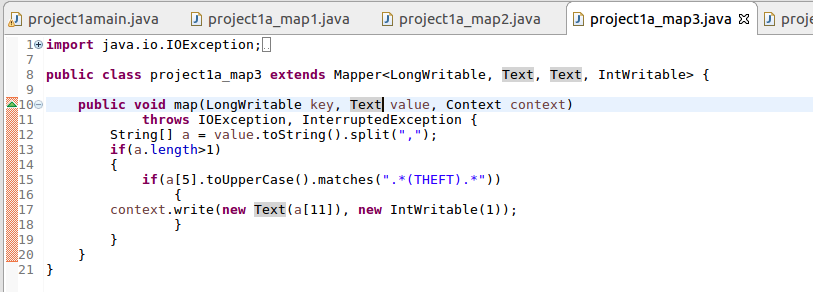


**MAPPER 1:**

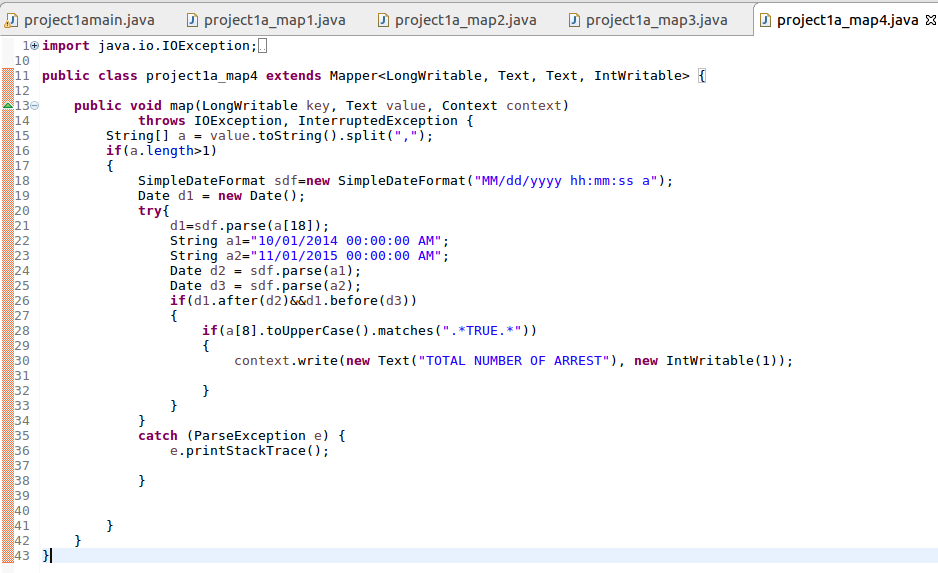
**MAPPER 2:**



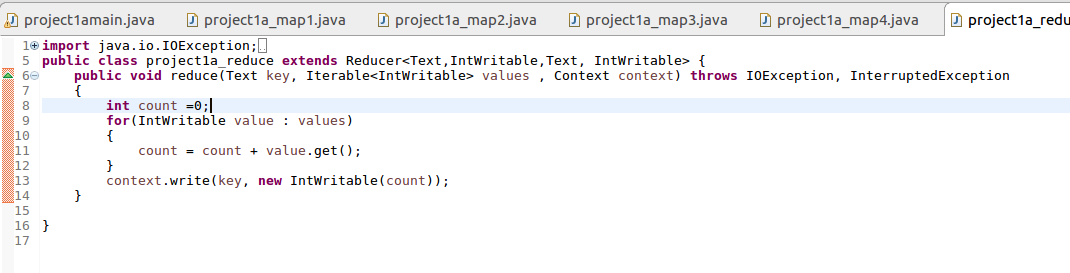
**MAPPER 3:**



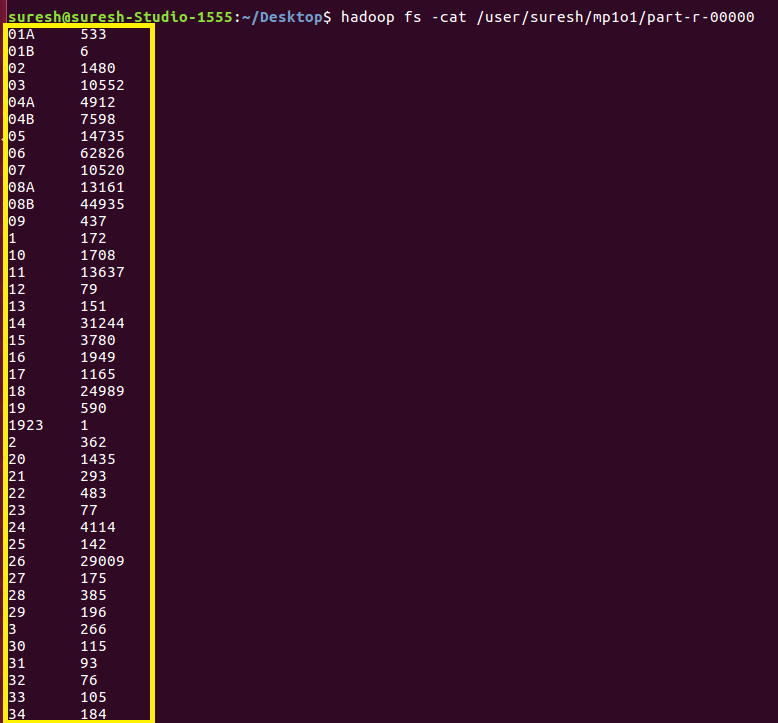
**MAPPER 4:**

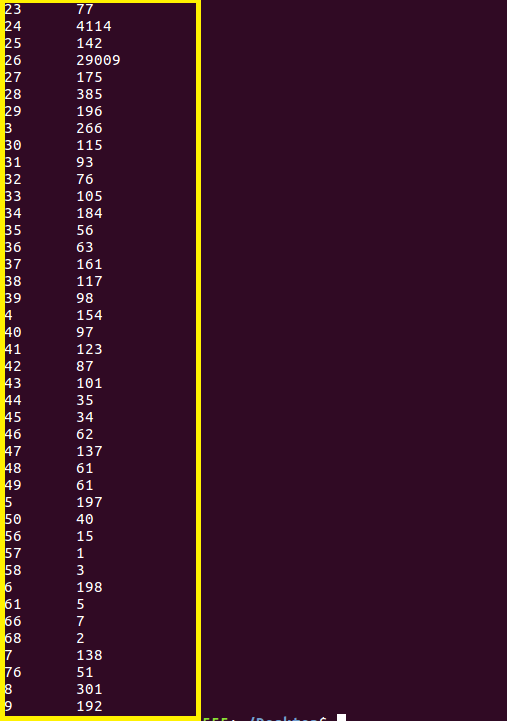


**REDUCER CLASS:**

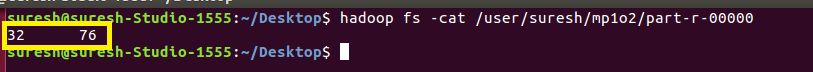


**OUTPUT 1:**





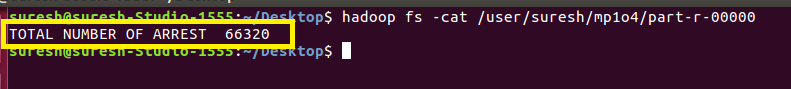
**OUTPUT 2:**



**OUTPUT 3:**

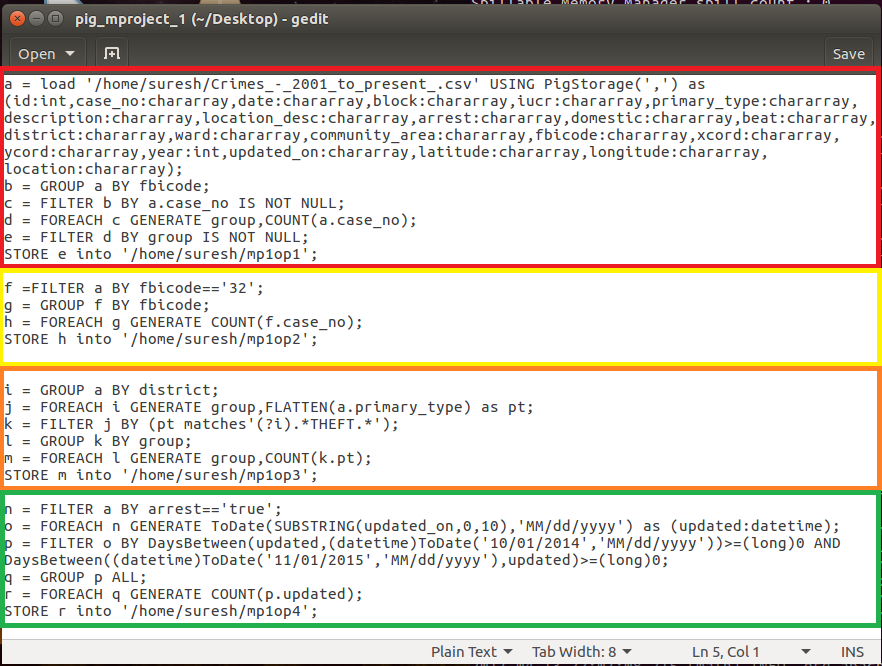


**OUTPUT 4:**

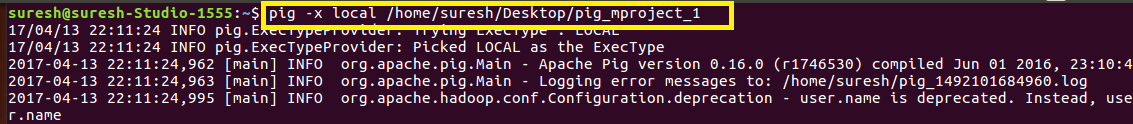


**PIG APPROACH:**

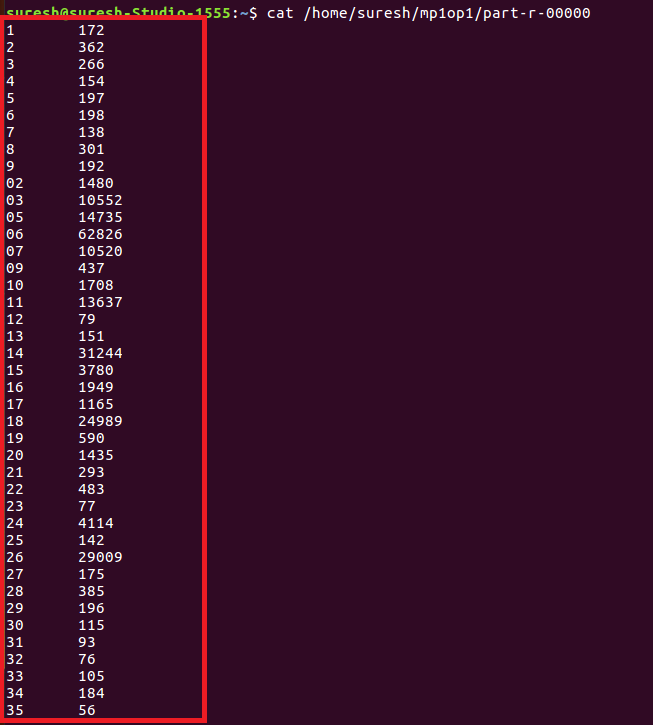
**PIG SCRIPT:**

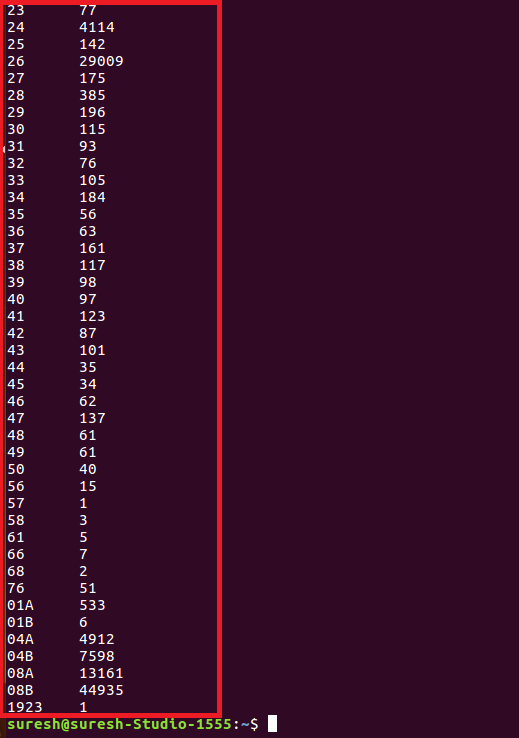


**RUNNING PIG SCRIPT:**



**OUTPUT 1:**

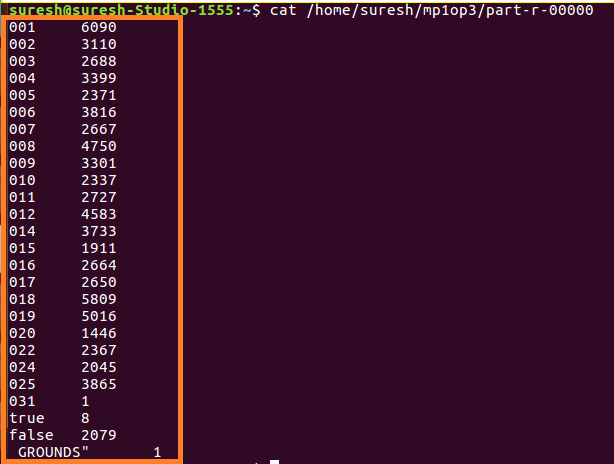




**OUTPUT 2:**



**OUTPUT 3:**



**OUTPUT 4:**

