**BigData And Hadoop**

**Project1.1 - Task 3**

**Dataset Description:**

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

**Problem Statement:**

Write a MapReduce/Pig program to calculate the number of arrests in theft district wise.

**Solution**

**Code files are as follows:**

Mapper class: Task3Mapper.java

Reducer class: Task3Reducer.java

Combiner class: Task3Reducer.java

Driver class: Task3Driver.java

**Solution logic:**

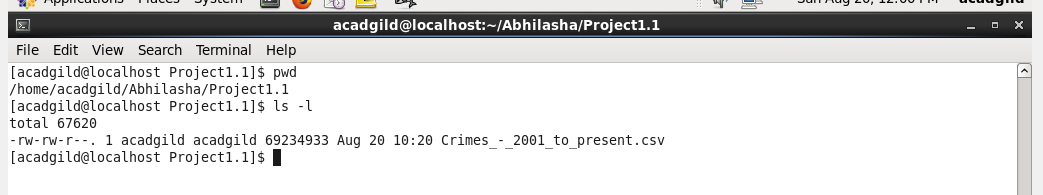
In mapper, we apply a check on PrimaryType and Arrest attributes as we need records with primary type = THEFT and Arrest = TRUE. Mapper emits the District as the key and ‘1’ as the value as we need to group the records by District.

In reducer, we find the total number of cases grouped by District, i.e. by summing it up per District.

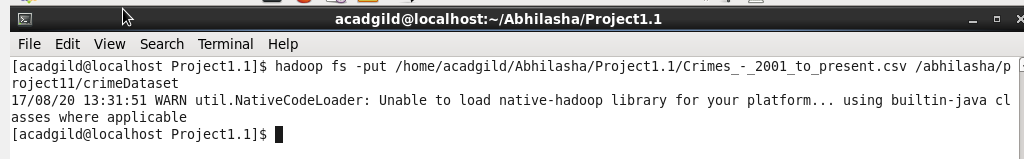
We have used combiner as well, to improve performance, which is same as reducer.

**Snapshots of the output are as follows:**

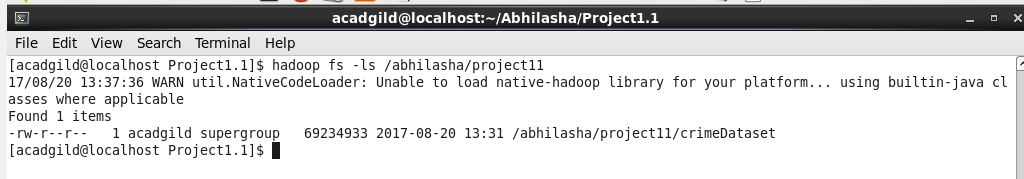
1. Dataset to be used is present locally at /home/acadgild/Abhilasha/Project1.1 as follows:



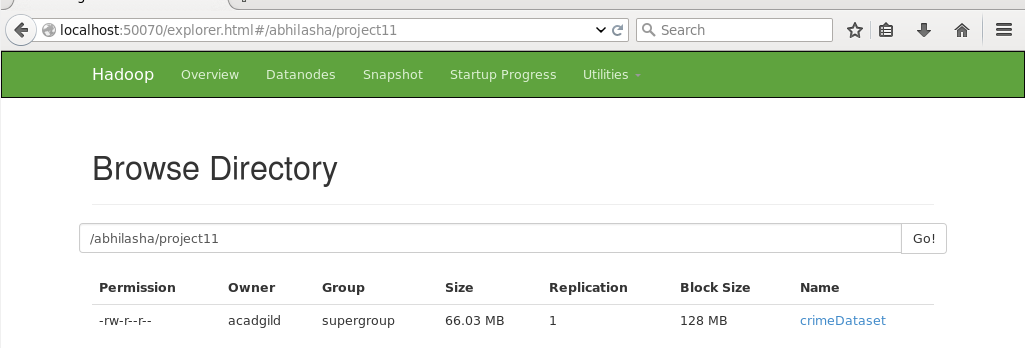
1. We now put the dataset into hdfs using put command at /abhilasha/project11 and name it as crimeDataset.



1. Verify the presence of crimeDataset into hdfs using ls command



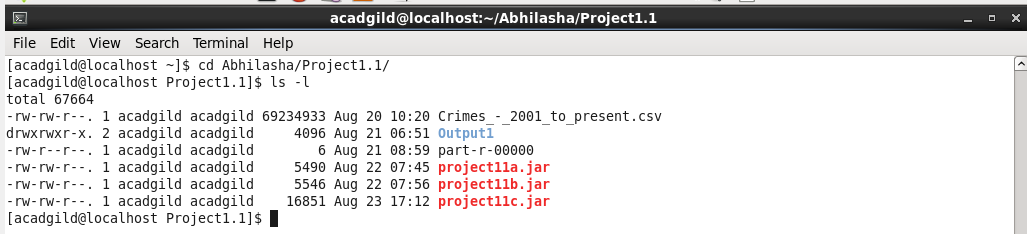
Verifying it from hdfs UI as well



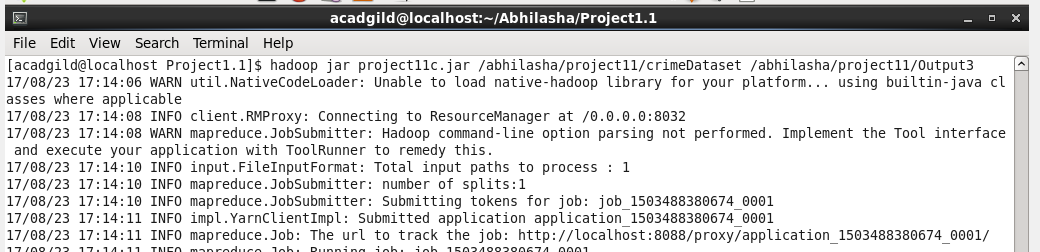
1. Viewing a few records from the dataset using cat command and then using head command to limit the records to display



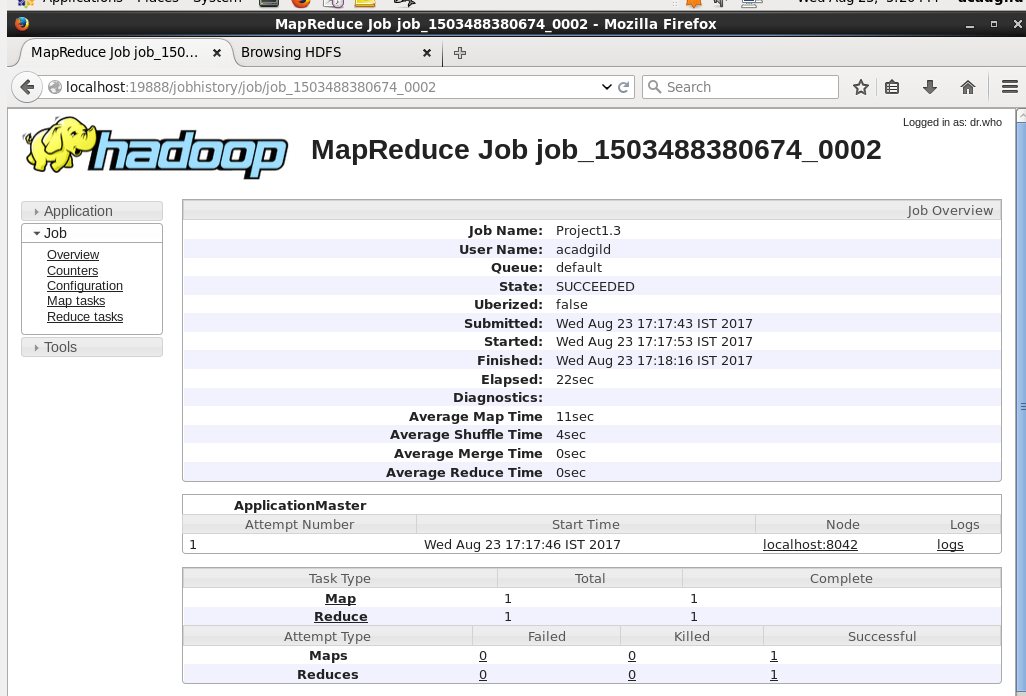
1. The Jar being used is project11c.jar and is placed at /home/acadgild/Abhilasha/Project1.1



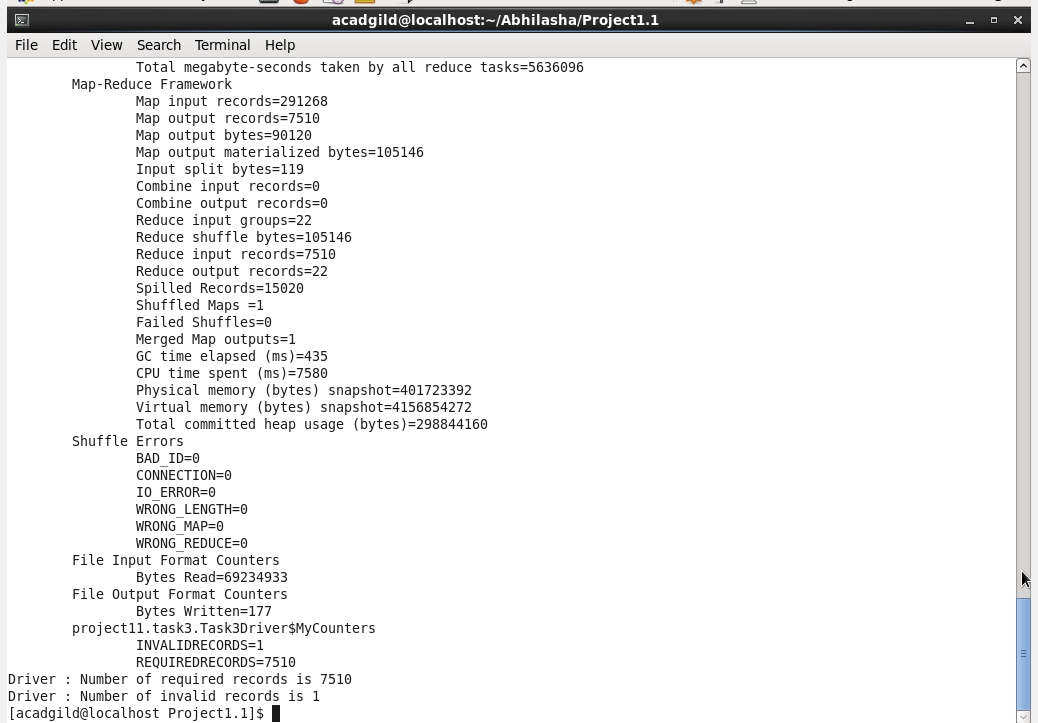
1. Executing the jar containing map-reduce code as follows. Output directory mentioned is /abhilasha/project11/Output3



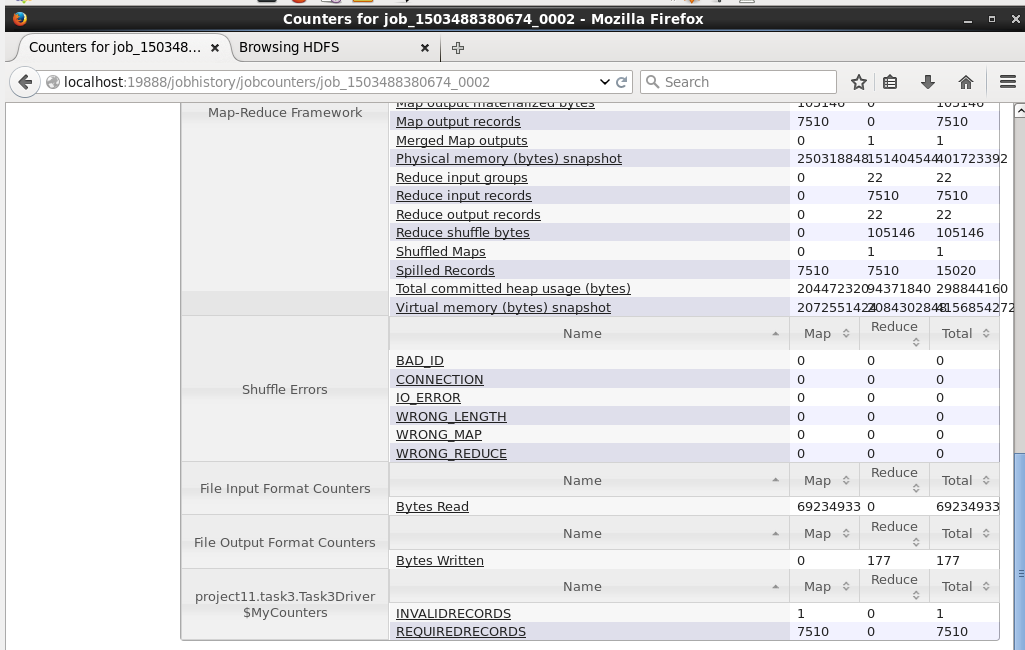
1. Job’s successful completion is shown on the job history server as follows:



1. We used custom counters in the code, their values are as follows:

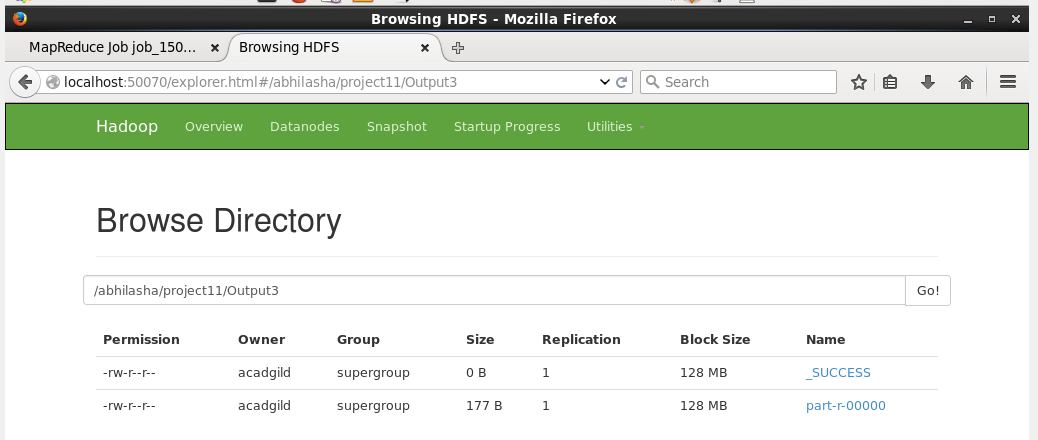


These custom counters are also visible on history server in Counters section as follows:

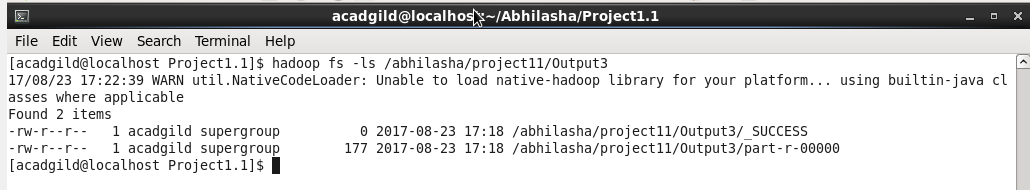


1. Listing the directory Output3, which contains output of map-reduce as follows:

From hdfs UI



Also listing it using ls command



1. Displaying the output on console using cat command. Output displays the FBI code i.e., 32 and number of records belonging to it.

