**BigData And Hadoop**

**Project1.1 - Task 1**

**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 cases investigated under each FBI code.

**Solution**

**Code files are as follows:**

Mapper class: Mapper.java

Reducer class: Reducer.java

Combiner class: Reducer.java

Driver class: Driver.java

**Solution logic:**

We need to find number of cases under each FBI code.

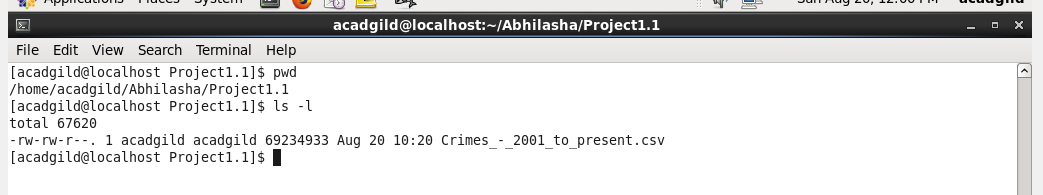
In mapper, we emit the FBI Code as the key and ‘1’ as the value. Using FBI Code as key as we need number of cases per FBI code.

In reducer, we find the total number of cases under each FBI code by summing up the count. As key of mapper is FBI code, reducer will get records grouped by FBI code and hence can calculate the records per FBI code.

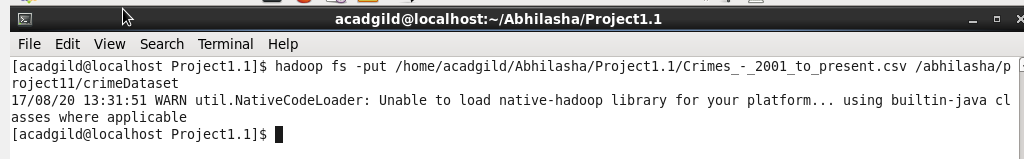
We have used combiner as well, to improve performance, which is same as reducer.

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

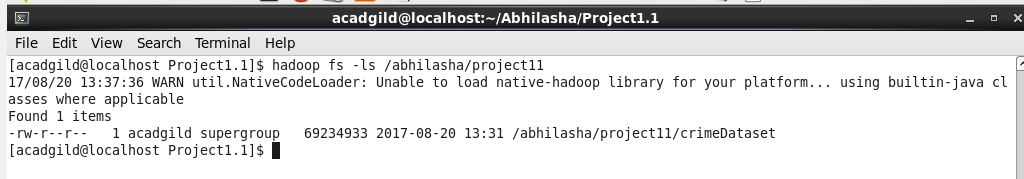
1. Dataset to be used is present locally at /home/acadgild/Abhilasha/Project1.1 as follows:



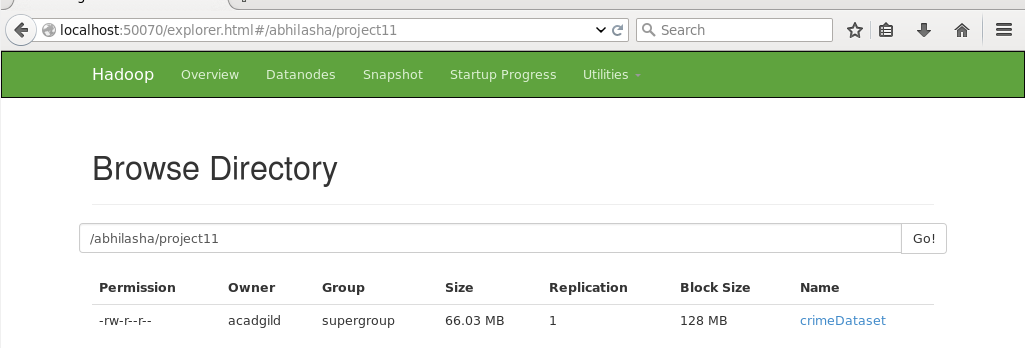
1. We now put the dataset into hdfs using put command at /abhilasha/project11 and name it as crimeDataset.



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



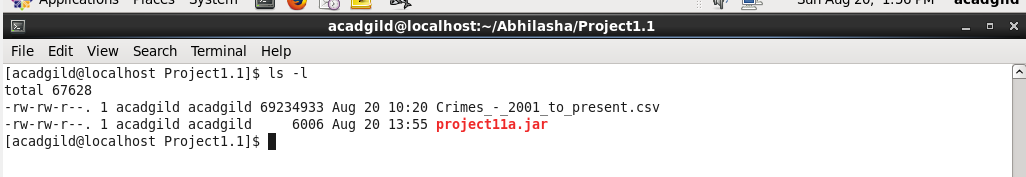
Verifying it from hdfs UI as well



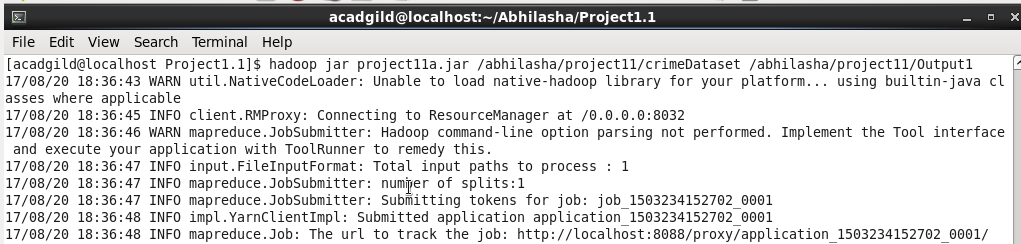
1. Viewing a few records from the dataset using cat command and then using head command to limit the records to display



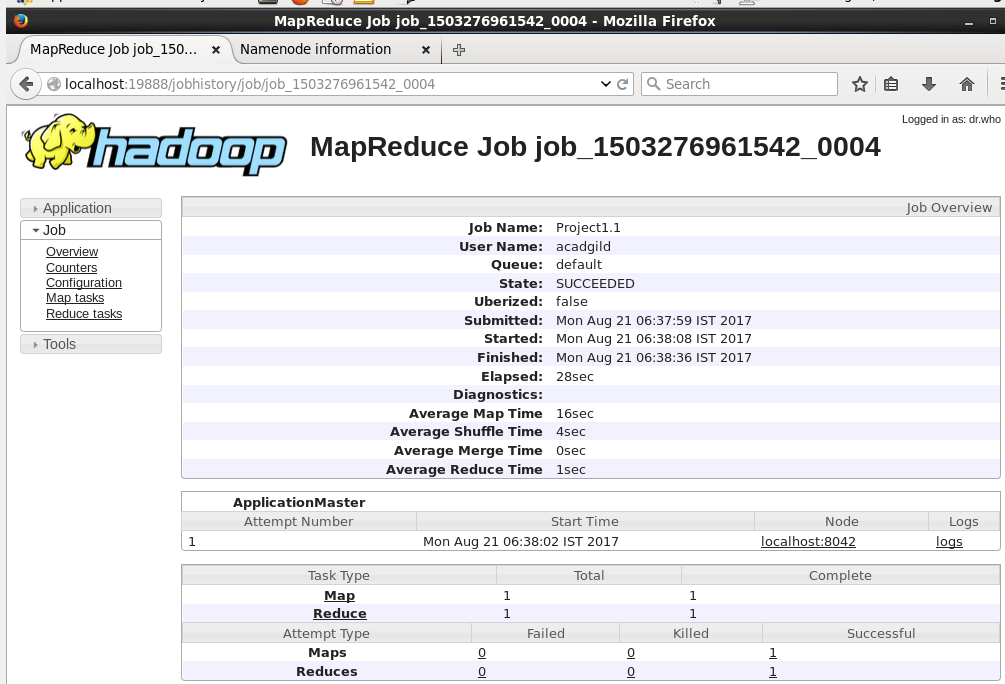
1. The Jar being used is project11a.jar and is placed at /home/acadgild/Abhilasha/Project1.1



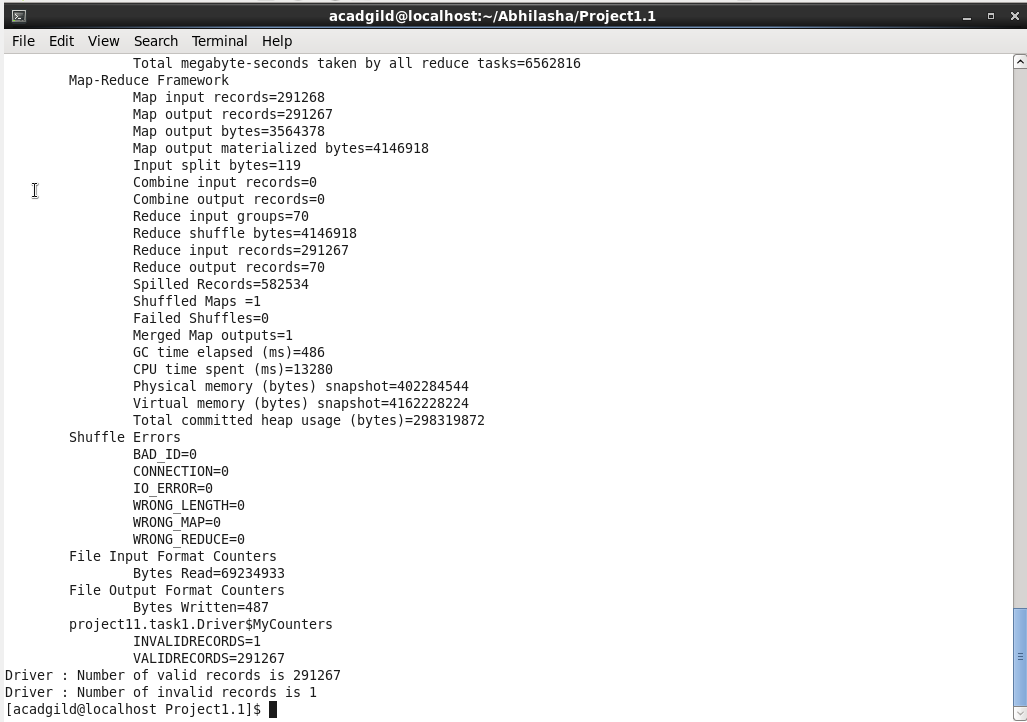
1. Executing the jar containing map-reduce code as follows. Output directory mentioned is /abhilasha/project11/Output1



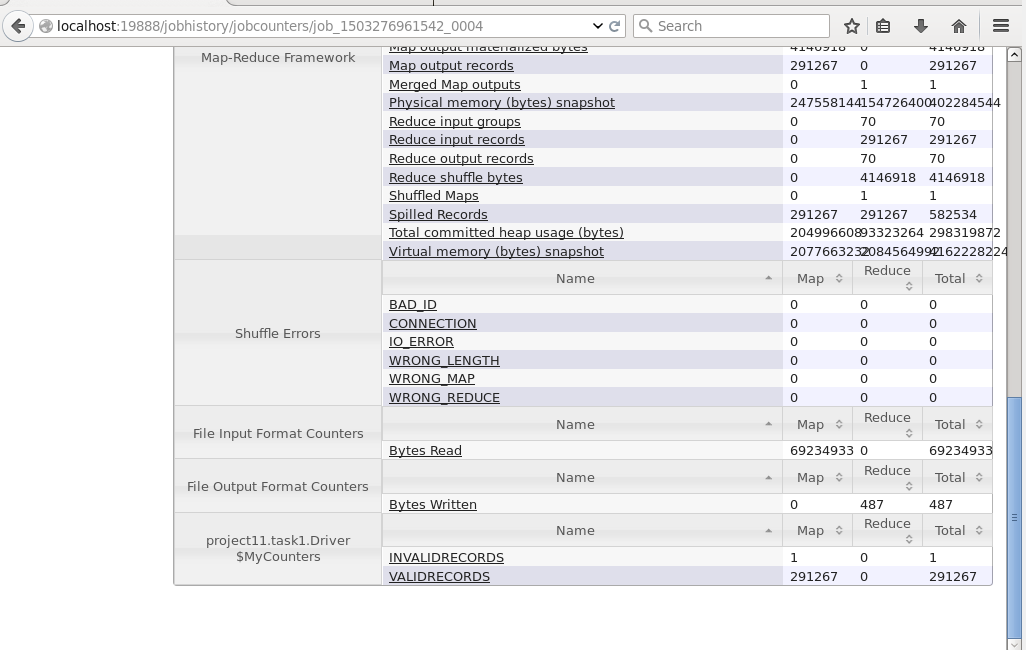
1. Job’s successful completion is shown on the job history server as follows:



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

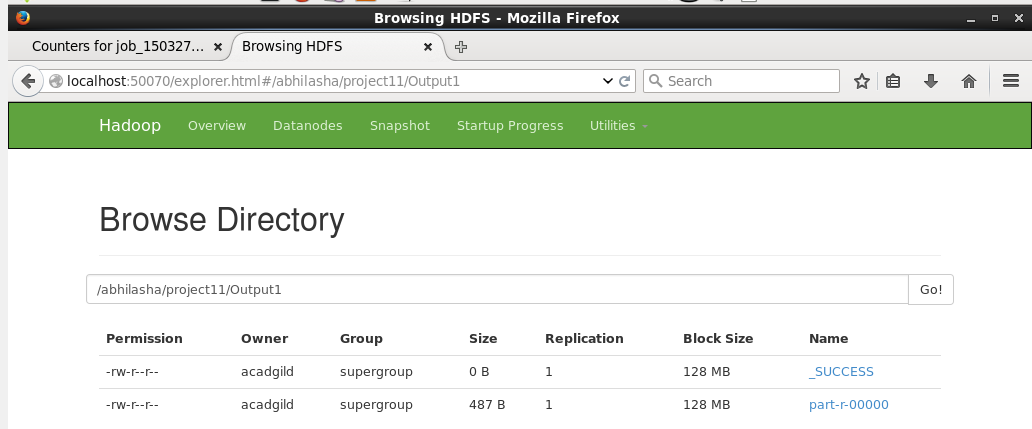


These custom counters are also visible on history server in Counters section as follows:

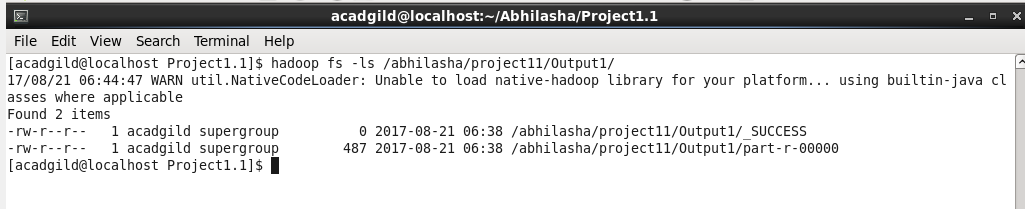


1. Listing the directory Output1, which contains output of map-reduce as follows:

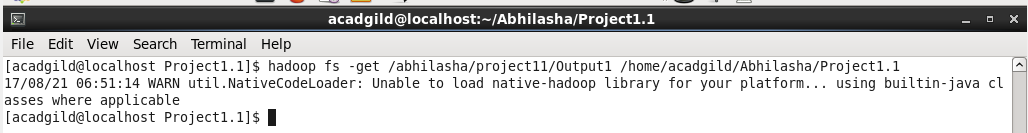
From hdfs UI



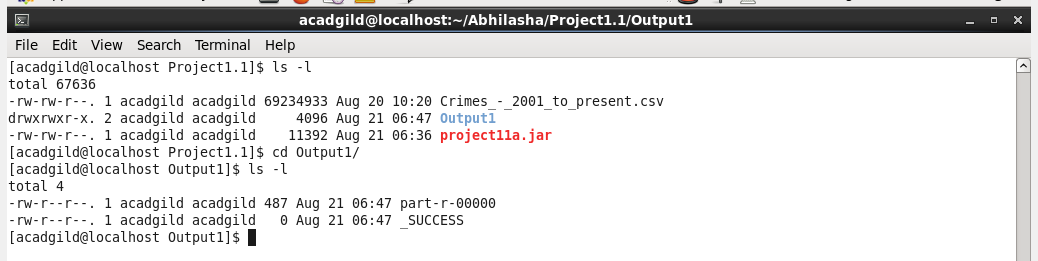
Also listing it using ls command



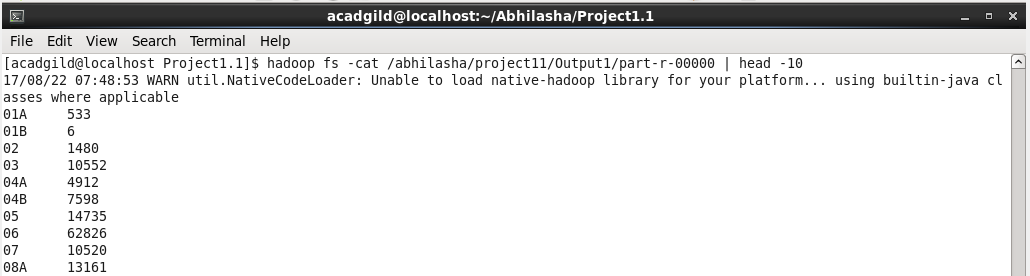
1. Placing the output on local file system from hdfs using get command and placing it at /home/acadgild/Abhilasha/Project1.1 as follows:



1. Verifying presence of output folder on local file system as follows:



1. Displaying a few records from output file on console using cat command. Used head command to limit the number of records to display



Complete output is placed in the file part-r-00000.