Department of Computing

Syed Muhammad Dawoud Sheraz Ali-111417

BESE-5B

**Lab 13: Hadoop**

**Date: 31/05/2017**

**SE312: Software Construction**

**Instructor: Fahad Ahmed Satti**

**Lab Engineer: Ms. Maryam Sajjad**

**GitHub:** <https://github.com/SyedDawoud/University/tree/master/Software%20Construction/Labs>

# Introduction

This lab is intended to give a deep knowledge of how to use Hadoop for the analysis purpose. We have been provided with the data in csv file. The data is the call records of the different users. We have to analyze and make the relationship between the caller and callee. The relationship can be either the number of calls made, the duration of the call, call made on the day etc. The purpose is to see how different jobs can be done using the Hadoop.

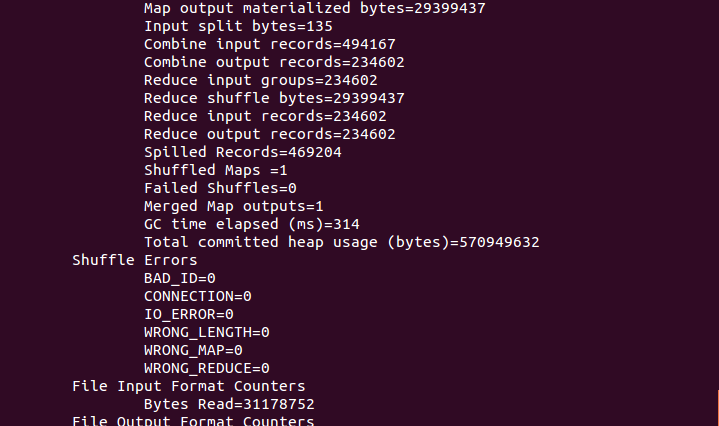
# Approach & Analysis

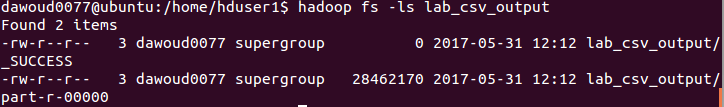
1. For loading the file, following command is used:



1. Next is the running of a job to see csv file has been loaded:







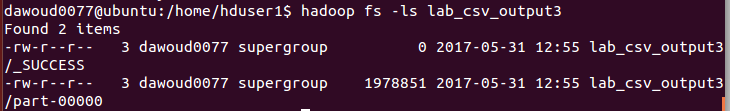
The output of Hadoop word count is as:



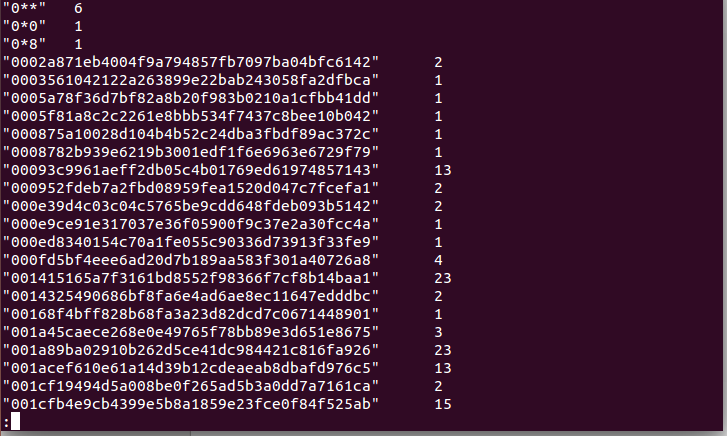
1. Next part is the running of custom jar on the csv file.











# Running the Code

# For running the code, the Hadoop must be setup.

# Start the service using “start-all.sh” command.

1. Copy the file from local directory to hdfs using “Hadoop fs –copyFromLocal local\_path Hadoop\_path”.
2. Run the job from jar using “Hadoop jar jar\_file.jar ClassName Hadoop\_path output\_path”.
3. See whether output is generated from the command “Hadoop fs –ls output\_path”
4. View the output using “Hadoop dfs –cat output\_path/file\_name | less”.