Name of the student:		**Roll No.**	
Practical Number:2		**Date of Practical:**	
Relevant CO's: ITC802.2	**At the end of the course students will be able to use tools like hadoop and NoSQL to solve big data related problems.**		
Sign here to indicate that you have read all the relevant material provided before attempting this practical			**Sign:**

Practical grading using Rubrics

Indicator	**Very Poor**	**Poor**	**Average**	**Good**	**Excellent**
Timeline (2)	More than a session late (0)	NA	NA	NA	Early or on time (2)
Code design (2)	N/A	Very poor code design with no comments and indentation(0.5)	Poor code design with very comments and indentation (1)	Design with good coding standards (1.5)	Accurate design with better coding satndards (2)
Performance (4)	Unable to perform the experiment (0)	Able to partially perform the experiment (1)	Able to perform the experiment for certain use cases (2)	Able to perform the experiment considering most of the use cases (3)	Able to perform the experiment considering all use cases (4)
Postlab (2)	No Execution(0)	N/A	Partially Executed (1)	N/A	Fully Executed (2)

Total Marks (10)	**Sign of instructor**

Practical

Course title: Big Data Analytics
Course term: 2019-2020
Instructor name: Saurabh Kulkarni

Problem Statement: Counting number of words in given text file using map reduce.

Theory:Explain the working of word count using map reduce with small example and diagrams

Code:

code for mapper:

```
import java.io.IOException;
import java.util.StringTokenizer;

import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.LongWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Mapper;
import org.apache.hadoop.mapreduce.Mapper.Context;

public class WCMapper extends Mapper<LongWritable, Text, Text,
    ↪ IntWritable>
 {
        // Create object of type Text to hold strings created per word
            ↪ of given document
```

Code for Reducer:

Code for Driver Class:

PostLab:Find inverted index

In this assignment you have to implement a simple map reduce job that builds an inverted index on the set of input documents. An inverted index maps each word to a list of documents that contain the word, and additionally records the position of each occurrence of the word within the document. For the purpose of this assignment, the position will be based on counting words, not characters.

Ex: Assume below are the input Documents.
file1="data is good."
file2="data is not good?"

Output:
data (file1,1)(file2,1)
good (file1,3)(file2,4)
is (file1,2)(file2,2)
not (file2,3)
For more details on inverted indices, you can check out the Wikipedia page on inverted indices.
Now in this assignment you need to implement above map-reduce job.
Input: A set of documents

Output:
Map: word1 (filename, position)
word2 (filename, position)
word1 (filename, position)
and so on for each occurrence of each word.
Reduce: word1 (filename, position)(filename,position)
word2 (filename, position)
and so on for each word.
Code for getting file name in Hadoop, which can be used in the Map function:

```
String filename=null;
filename = ((FileSplit) context.getInputSplit()).
    ↪ getPath().getName();
```

Code for postlab question