

# Distributed Systems

## Assignment 4 – MapReduce

Assignment given on: 31<sup>th</sup> Oct 2020

Due Date: 13th Nov 2020

---

### Assignment 4 – MapReduce with Hadoop.

The objective of this assignment is to gain experience on Hadoop setup and understand the MapReduce. You will implement a Hadoop program that counts the number of occurrences of each word in a text file.

#### Setup:

Setup a Pseudo Distributed Mode mode Hadoop environment on your local machine or VM.

#### Solution spec:

Write a scala program that counts the number of occurrences of each word in a text file. Place the input text file into HDFS. Implement MapReduce algorithm to read the input file, count the number of occurrences of each word and write the output (that is word and count) into a output file.

Create a jar file and name it as Assignment4.jar and run the jar file

```
$hadoop jar /home/user/WordCount_Assignment4.jar WordCount  
/test/input_data.txt WCOuput
```

Where,

input\_data.tx – is the input file

WCOuput – is the output file to store the word count, each line contains word and word count.

#### What to Submit

A single zip file containing:

- WordCount.scala
- input\_data.txt
- WCOuput
- README File
- WordCount\_Assignment4.jar
- Output Screenshot

## Grading – 20 Marks

- WordCount.scala– 15 Marks
- Demo – 5 Marks

## What to Demo

- Pseudo Distributed Mode setup (configuration files).
- Upload the input\_data.txt file HDFS using the command prompt.
- Create a jar file of the program from command prompt.
- Run the jar file from the command prompt.
- Show the result.

## Note:

- Penalty of 10% per day will be issued if the deadline is not met
- If found copied, you will fetch 0 score.

## Submission Details:

1. Please read the questions carefully and complete it.
2. Make a directory with name <Your\_Roll\_Number> and copy your all program (source code) and output file to that folder.
3. Implement the solution using Scala.
4. Please take screen shot of output, name it with its question number and put it in a same folder.
5. Test well before submission. Follow some coding style uniformly. Provide proper comments in your code.
6. Submit only through moodle and well in advance. Any hiccups in the moodle/Internet at the last minute is never acceptable as an excuse for late submission. Submissions through email will be ignored.
7. Please attach a readme.txt file
8. Please zip your folder and submit to moodle on or before 13-11-2020 (23:55 PM)