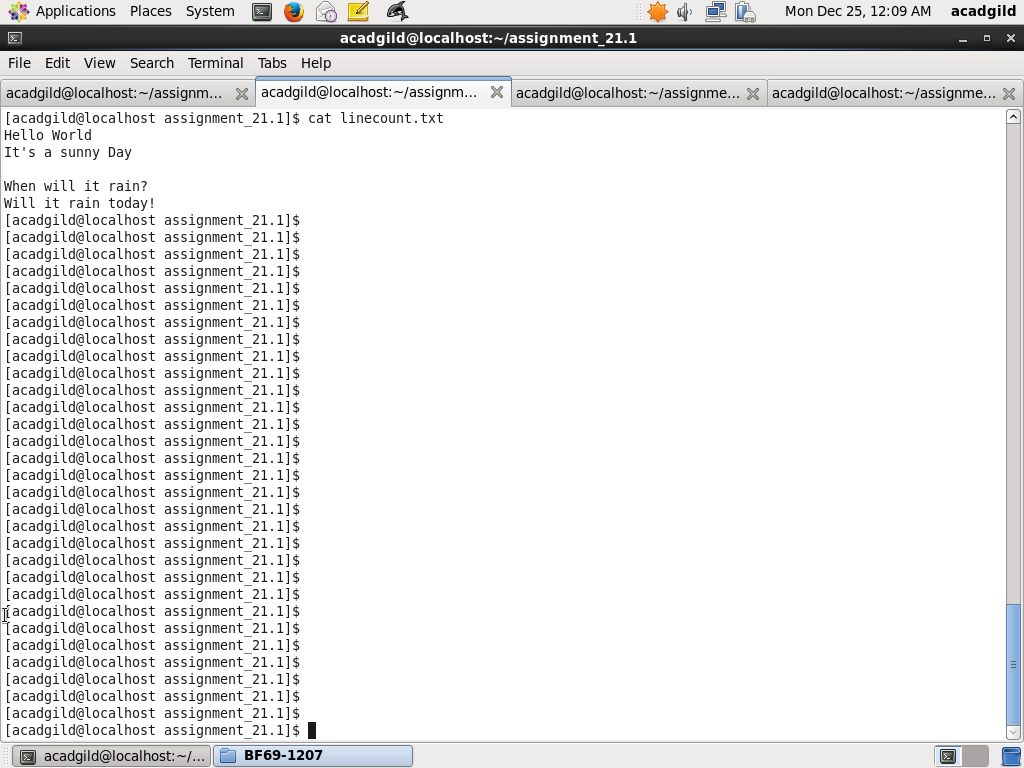
Assignment 21.1

Step1: Copy the dataset linecount.txt

The Screenshot for linecount.txt is as below



Step2: define a accumulator to keep track of blank line count

Define accumulator blankLinesCount and initialize to 0

val blankLinesCount = sc.accumulator(0)

Step3: Define a function to calculate blank line count

Define a function doBlankLineCount to calculate blank line count. It will trim line and checks if it is a blank line then increment accumulator blankLinesCount by 1

Code is as below:

def doBlankLineCount(line:String):Array[String] = {

if (line.trim().equals("")) {

blankLinesCount += 1

Array[String]()

} else {

line.split(" ")

}

}

Step4: Read file linecount.txt and process blank line count

Read file linecount.txt and put to fileRR. Next use flatMap on fileRDD with method doBlankLIneCount and call the action collect

val fileRDD = sc.textFile("/home/acadgild/assignment\_21.1/linecount.txt")

fileRDD.flatMap(doBlankLineCount).collect

Step5: Display blank line count

Use println function to display value of blankLinesCount accumulator

println("No of blank lines = " + blankLinesCount.value.toInt)

Screenshot steps 2 to 5 is as below:

