**HW 8/15 Spark Practice**

Not using Hadoop file system

Create input.txt in hadoopuser

And put the below content

people are not as beautiful as they look,

as they walk or as they talk.

they are only as beautiful as they love,

as they care as they share.

People 1

Are 1

Not 1

As 1

Are 1

-----

Are 2

scala> val inputfile = sc.textFile("file:///home/hadoopuser/input.txt")

inputfile: org.apache.spark.rdd.RDD[String] = file:///home/hadoopuser/input.txt MapPartitionsRDD[1] at textFile at <console>:24

scala> val counts = inputfile.flatMap(line => line.split(" ")).map(word => (word,

1)).reduceByKey(\_+\_);

counts: org.apache.spark.rdd.RDD[(String, Int)] = ShuffledRDD[4] at reduceByKey at <console>:25

scala> counts.saveAsTextFile("file:///home/hadoopuser/output")

scala>

------------------------------------------------------------------------------------------------

cd

nano input.txt

people are not as beautiful as they look,

as they walk or as they talk.

they are only as beautiful as they love,

as they care as they share.

val inputfile = sc.textFile("file:///home/hadoopuser/input.txt")

.

counts.saveAsTextFile("file:///home/hadoopuser/output")

-----

cat input.txt

--------------8/24/18------------------------------------------

Practice Scala online:

<https://www.tutorialspoint.com/scala/scala_basic_syntax.htm>

sudo apt-get install scala

# this didnt work

<https://stackoverflow.com/questions/50876388/scala-script-wont-run-on-ubuntu>

Version check:

<https://docs.scala-lang.org/overviews/jdk-compatibility/overview.html>

sudo apt remove scala-library scala

wget www.scala-lang.org/files/archive/scala-2.12.6.deb

sudo dpkg -i scala-2.12.6.deb

-----------------------