from pyspark.sql import SparkSession

import getpass

username = getpass.getuser()

spark = SparkSession. \

builder. \

config('spark.ui.port','0'). \

config("spark.sql.warehouse.dir", f"/user/{username}/warehouse"). \

enableHiveSupport(). \

master('yarn'). \

getOrCreate()

rdd1 = spark.sparkContext.textFile("/user/itv017244/data/input/inputfile.txt")

rdd2 = rdd1.flatMap(lambda line : line.split(" "))

rdd3 = rdd2.map(lambda word : (word,1))

rdd4 = rdd3.reduceByKey(lambda x,y : x+y)

rdd4.collect()

rdd4.saveAsTextFile("/user/itv017244/data/newputput")