Q3- Spark code of word count

Solution:-

from pyspark.sql import SparkSession

from pyspark.sql.functions import split, explode

spark = SparkSession.builder.appName("WordCount").getOrCreate()

lines = spark.read.text("<input\_file>").rdd.map(lambda r: r[0])

words = lines.flatMap(lambda line: line.split(" "))

word\_counts = words.groupBy(lambda word: word.lower()).count()

word\_counts.show()

spark.stop()