University of Wolverhampton

School of Mathematics and Computer Science

5CS022 Distributed and Cloud Systems Programming

Workshop 5: The Apache Spark Framework

Question 1

Word count

```
SLF4J: Failed to load class "org.slf4j.impl.StaticLoggerBinder".
SLF4J: Defaulting to no-operation (NOP) logger implementation
SLF4J: See http://www.slf4j.org/codes.html#StaticLoggerBinder for further details.
SLF4J: Failed to load class "org.slf4j.impl.StaticMDCBinder".
SLF4J: Defaulting to no-operation MDCAdapter implementation.
SLF4J: See http://www.slf4j.org/codes.html#no_static_mdc_binder for further details.
Word Counts:
became: 1
save: 1
play: 1
4: 1
Cross-platform: 1
it: 2
is: 2
2020,: 1
X/S: 1
December: 1
action: 1
2021: 1
Support: 1
multiplayer: 1
Series: 1
developed: 1
later: 1
January: 1
September: 1
groups: 1
A: 1
in: 12
2023,[1]: 1
Switch: 1
development: 1
online: 1
Android: 1
2022.: 1
free-to-play: 1
2024.: 2
```

```
shooter: 1
ported: 1
Nintendo: 1
role-playing: 1
fully: 1
April: 1
began: 1
different: 1
iOS: 1
March: 1
players: 2
computers: 1
5: 1
was: 3
a: 1
One: 1
2018,: 1
2014,: 1
First: 1
all: 1
to: 3
available: 1
November: 3
Xbox: 2
Extremes.: 1
third-person: 1
of: 1
released: 2
Warframe: 1
by: 1
waves: 1
published: 1
PlayStation: 2
Windows: 1
and: 3
February: 1
```

Question 2

Create a Spark program to count letters instead of words.

```
package uk.ac.wlv;
import java.io.IOException;
import java.util.Arrays;
import org.apache.hadoop.fs.FileSystem;
import org.apache.hadoop.fs.Path;
import org.apache.spark.SparkConf;
import org.apache.spark.api.java.JavaPairRDD;
import org.apache.spark.api.java.JavaRDD;
import org.apache.spark.api.java.JavaSparkContext;
import scala.Tuple2;

public class LetterCount {

    public static void main(String[] args) {

        SparkConf sparkConf = new SparkConf();
        sparkConf.setAppName("Spark LetterCount example using Java");

        sparkConf.setMaster("local");
        JavaSparkContext sparkContext = new
JavaSparkContext(sparkConf);
        JavaRDD<String> textFile = sparkContext.textFile("input.txt");

        JavaRDD<String> letters = textFile.flatMap(l ->
Arrays.asList(l.split("")).iterator());
        /*Generate Pair of Letter with count */

        JavaPairRDD<String, Integer> pairs = letters.mapToPair(w ->
new Tuple2<String, Integer>(w, 1));

        /* Aggregate Pairs of Same Letters with count */

        JavaPairRDD<String, Integer> counts = pairs.reduceByKey((x, y)
-> x + y);

        /* Deleting output directory if it already exists and saving
the result file */

        String outputPath = "output1"; // Change this to your desired
output directory

        try {


FileSystem.get(sparkContext.hadoopConfiguration()).delete(new
Path(outputPath), true);
        } catch (IOException e) {
            e.printStackTrace();
        }

        /* Saving the result file */
        try {
            counts.saveAsTextFile(outputPath);
        } catch (Exception e) {
```

```
            e.printStackTrace();
        }

        /* System.out.println(counts.collect()); */
        System.out.println("Letter Counts:");

        for (Tuple2<String, Integer> tuple : counts.collect()) {
            System.out.println(tuple._1() + ": " + tuple._2());
        }
        sparkContext.stop();

        sparkContext.close();

    }

}
```

```
<terminated> LetterCount [Java Application] C:\Program Files\Java\jdk-17\bin\javaw.exe  (Mar 30, 2024, 2:38:47 PM – 2:39:05 PM) [pid: 11152]
SLF4J: Failed to load class "org.slf4j.impl.StaticMDCBinder".
SLF4J: Defaulting to no-operation MDCAdapter implementation.
SLF4J: See http://www.slf4j.org/codes.html#no_static_mdc_binder for further details.
Letter Counts:
w: 6
4: 4
e: 56
P: 2
O: 2
b: 13
y: 12
A: 3
h: 5
2: 17
o: 38
i: 36
n: 32
3: 3
-: 6
J: 1
S: 8
F: 2
1: 6
g: 5
l: 30
N: 4
m: 14
c: 7
d: 17
s: 25
8: 1
/: 1
p: 21
x: 3
M: 1
5: 1
a: 40
```

```
<terminated> LetterCount [Java Application] C:\Program Files\Java\jdk-17\bin\javaw.exe  (Mar 30, 2024, 2:38:47 PM
3: 3
-: 6
J: 1
S: 8
F: 2
1: 6
g: 5
l: 30
N: 4
m: 14
c: 7
d: 17
s: 25
8: 1
/: 1
p: 21
x: 3
M: 1
5: 1
a: 40
t: 30
.: 4
0: 11
u: 9
 : 103
]: 1
f: 11
v: 8
C: 1
E: 1
r: 42
,: 6
[: 1
X: 3
W: 2
D: 2
```