Project 5

Name: Muyu Tong

Andrew ID: muyut

Email: muyut@andrew.cmu.edu

Part 2 Program

```
1   /**
2    * Name: muyu Tong
3    * Andrew ID: muyut*/
4   import org.apache.spark.SparkConf;
5   import org.apache.spark.api.java.JavaPairRDD;
6   import org.apache.spark.api.java.JavaRDD;
7   import org.apache.spark.api.java.JavaSparkContext;
8   import scala.Tuple2;
9
10  import java.util.Arrays;
11  import java.util.Scanner;
12  //https://github.com/CMU-Heinz-95702/lab9-MapReduceAndSpark
13  public class TempestAnalytics {
14
15      //task 0
16      private static void countLine(String fileName) {
17
18          SparkConf sparkConf = new
    SparkConf().setMaster("local").setAppName("TempestAnalytics");
19
20          JavaSparkContext sparkContext = new JavaSparkContext(sparkConf);
21
22          JavaRDD<String> inputFile = sparkContext.textFile(fileName);
23
24          //        Split by lines
25
26          JavaRDD<String> linesFromFile = inputFile.flatMap(content ->
    Arrays.asList(content.split("\n")));
27
28          double count = linesFromFile.count();
29          System.out.println("the total number of lines is "+ count);
30      }
31
32      //task1
33      private static void wordCount(String fileName) {
34
35          SparkConf sparkConf = new
    SparkConf().setMaster("local").setAppName("TempestAnalytics");
```

```java
36
37          JavaSparkContext sparkContext = new JavaSparkContext(sparkConf);
38
39          JavaRDD<String> inputFile = sparkContext.textFile(fileName);
40
41          JavaRDD<String> wordsFromFile = inputFile.flatMap(content ->
    Arrays.asList(content.split("[^a-zA-Z]+")));
42
43          double count = wordsFromFile.count();
44
45          System.out.println("the total number of words is "+count);
46      }
47
48      //task2
49      private static void distinctWordCount(String fileName) {
50
51          SparkConf sparkConf = new
    SparkConf().setMaster("local").setAppName("TempestAnalytics");
52
53          JavaSparkContext sparkContext = new JavaSparkContext(sparkConf);
54
55          JavaRDD<String> inputFile = sparkContext.textFile(fileName);
56
57          JavaRDD<String> wordsFromFile = inputFile.flatMap(content ->
    Arrays.asList(content.split("[^a-zA-Z]+")));
58
59          double count = wordsFromFile.distinct().count();
60
61          System.out.println("the total number of distinct word is " +count);
62      }
63
64      //task3
65      private static void symbolCount(String fileName) {
66
67          SparkConf sparkConf = new
    SparkConf().setMaster("local").setAppName("TempestAnalytics");
68
69          JavaSparkContext sparkContext = new JavaSparkContext(sparkConf);
70
71          JavaRDD<String> inputFile = sparkContext.textFile(fileName);
72
73          JavaRDD<String> symbols = inputFile.flatMap(content ->
    Arrays.asList(content.split("")));
74
75          double count = symbols.count();
76
77          System.out.println("the total number of symbols is "+count);
78      }
79
80      //task4
81      private static void distinctSymbolCount(String fileName) {
```

```java
82
83          SparkConf sparkConf = new
     SparkConf().setMaster("local").setAppName("TempestAnalytics");
84
85          JavaSparkContext sparkContext = new JavaSparkContext(sparkConf);
86
87          JavaRDD<String> inputFile = sparkContext.textFile(fileName);
88
89          JavaRDD<String> symbols = inputFile.flatMap(content ->
     Arrays.asList(content.split("")));
90
91          double count = symbols.distinct().count();
92
93          System.out.println("the total number of distinct symbols is "+ count);
94      }
95
96      //task5
97      private static void distinctLetterCount(String fileName) {
98
99          SparkConf sparkConf = new
     SparkConf().setMaster("local").setAppName("TempestAnalytics");
100
101         JavaSparkContext sparkContext = new JavaSparkContext(sparkConf);
102
103         JavaRDD<String> inputFile = sparkContext.textFile(fileName);
104
105         JavaRDD<String> lettersFromFile = inputFile.flatMap(content ->
     Arrays.asList(content.split("")));
106         lettersFromFile = lettersFromFile.filter(k -> (k.matches("[a-zA-Z]")));
107         double count = lettersFromFile.distinct().count();
108
109         System.out.println("the total number of distinct letter is " + count);
110     }
111
112     //task6
113     private static void show(String fileName) {
114         System.out.println("Please enter the search word: ");
115         Scanner scanner = new Scanner(System.in);
116         String word = scanner.nextLine();
117
118         SparkConf sparkConf = new
     SparkConf().setMaster("local").setAppName("TempestAnalytics");
119
120         JavaSparkContext sparkContext = new JavaSparkContext(sparkConf);
121
122         JavaRDD<String> inputFile = sparkContext.textFile(fileName);
123
124         JavaRDD<String> lines = inputFile.flatMap(content ->
     Arrays.asList(content.split("\n")));
125         lines = lines.filter(k -> (k.contains(word)));
126         for (String line: lines.collect()) {
```

```
127              System.out.println(line);
128          }
129      }
130
131      public static void main(String[] args) {
132
133          if (args.length == 0) {
134              System.out.println("No files provided.");
135              System.exit(0);
136          }
137          show(args[0]);
138      }
139
```

Excution of Part 2 Task 6

```
Please enter the search word:
love
22/04/24 22:16:56 INFO spark.SecurityManager: Changing view acls to: tongmuyu
22/04/24 22:16:56 INFO spark.SecurityManager: SecurityManager: authentication disabled; ui acls disabled; users with vi
22/04/24 22:16:57 INFO slf4j.Slf4jLogger: Slf4jLogger started
22/04/24 22:16:57 INFO Remoting: Starting remoting
22/04/24 22:16:57 INFO Remoting: Remoting started; listening on addresses :[akka.tcp://spark@192.168.0.5:62112]
22/04/24 22:16:57 INFO Remoting: Remoting now listens on addresses: [akka.tcp://spark@192.168.0.5:62112]
22/04/24 22:16:57 INFO spark.SparkEnv: Registering MapOutputTracker
22/04/24 22:16:57 INFO spark.SparkEnv: Registering BlockManagerMaster
22/04/24 22:16:57 INFO storage.DiskBlockManager: Created local directory at /var/folders/81/zk47q6td0glgyn5x9wjx8g_m000
22/04/24 22:16:57 INFO storage.MemoryStore: MemoryStore started with capacity 2.1 GB.
22/04/24 22:16:57 INFO network.ConnectionManager: Bound socket to port 62113 with id = ConnectionManagerId(192.168.0.5,
22/04/24 22:16:57 INFO storage.BlockManagerMaster: Trying to register BlockManager
22/04/24 22:16:57 INFO storage.BlockManagerInfo: Registering block manager 192.168.0.5:62113 with 2.1 GB RAM
22/04/24 22:16:57 INFO storage.BlockManagerMaster: Registered BlockManager
22/04/24 22:16:57 INFO spark.HttpServer: Starting HTTP Server
22/04/24 22:16:57 INFO server.Server: jetty-8.1.14.v20131031
22/04/24 22:16:57 INFO server.AbstractConnector: Started SocketConnector@0.0.0.0:62114
22/04/24 22:16:57 INFO broadcast.HttpBroadcast: Broadcast server started at http://192.168.0.5:62114
22/04/24 22:16:57 INFO spark.HttpFileServer: HTTP File server directory is /var/folders/81/zk47q6td0glgyn5x9wjx8g_m000
22/04/24 22:16:57 INFO spark.HttpServer: Starting HTTP Server
22/04/24 22:16:57 INFO server.Server: jetty-8.1.14.v20131031
22/04/24 22:16:57 INFO server.AbstractConnector: Started SocketConnector@0.0.0.0:62115
22/04/24 22:16:57 INFO server.Server: jetty-8.1.14.v20131031
22/04/24 22:16:57 INFO server.AbstractConnector: Started SelectChannelConnector@0.0.0.0:4040
22/04/24 22:16:57 INFO ui.SparkUI: Started SparkUI at http://192.168.0.5:4040
22/04/24 22:16:58 INFO storage.MemoryStore: ensureFreeSpace(32816) called with curMem=0, maxMem=2290719129
22/04/24 22:16:58 INFO storage.MemoryStore: Block broadcast_0 stored as values to memory (estimated size 32.0 KB, free
22/04/24 22:16:58 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-j
22/04/24 22:16:58 WARN snappy.LoadSnappy: Snappy native library not loaded
22/04/24 22:16:58 INFO mapred.FileInputFormat: Total input paths to process : 1
```

22/04/24 22:16:58 INFO spark.SparkContext: Starting job: collect at TempestAnalytics.java:141
22/04/24 22:16:58 INFO scheduler.DAGScheduler: Got job 0 (collect at TempestAnalytics.java:141) with 1 output partitions
22/04/24 22:16:58 INFO scheduler.DAGScheduler: Final stage: Stage 0(collect at TempestAnalytics.java:141)
22/04/24 22:16:58 INFO scheduler.DAGScheduler: Parents of final stage: List()
22/04/24 22:16:58 INFO scheduler.DAGScheduler: Missing parents: List()
22/04/24 22:16:58 INFO scheduler.DAGScheduler: Submitting Stage 0 (FilteredRDD[3] at filter at TempestAnalytics.java:140
22/04/24 22:16:58 INFO scheduler.DAGScheduler: Submitting 1 missing tasks from Stage 0 (FilteredRDD[3] at filter at Temp
22/04/24 22:16:58 INFO scheduler.TaskSchedulerImpl: Adding task set 0.0 with 1 tasks
22/04/24 22:16:58 INFO scheduler.TaskSetManager: Starting task 0.0:0 as TID 0 on executor localhost: localhost (PROCESS_
22/04/24 22:16:58 INFO scheduler.TaskSetManager: Serialized task 0.0:0 as 2370 bytes in 3 ms
22/04/24 22:16:58 INFO executor.Executor: Running task ID 0
22/04/24 22:16:58 INFO storage.BlockManager: Found block broadcast_0 locally
22/04/24 22:16:58 INFO rdd.HadoopRDD: Input split: file:/Users/tongmuyu/Documents/courses/distributed/project5/Project5P
22/04/24 22:16:58 INFO executor.Executor: Serialized size of result for 0 is 1460
22/04/24 22:16:58 INFO executor.Executor: Sending result for 0 directly to driver
22/04/24 22:16:58 INFO executor.Executor: Finished task ID 0
22/04/24 22:16:58 INFO scheduler.TaskSetManager: Finished TID 0 in 131 ms on localhost (progress: 1/1)
22/04/24 22:16:58 INFO scheduler.DAGScheduler: Completed ResultTask(0, 0)
22/04/24 22:16:58 INFO scheduler.TaskSchedulerImpl: Removed TaskSet 0.0, whose tasks have all completed, from pool
22/04/24 22:16:58 INFO scheduler.DAGScheduler: Stage 0 (collect at TempestAnalytics.java:141) finished in 0.159 s
22/04/24 22:16:58 INFO spark.SparkContext: Job finished: collect at TempestAnalytics.java:141, took 0.248694833 s
BOATSWAIN None that I more love than myself. You are
Of all the world I loved, and to him put
So dear the love my people bore me, nor set
Knowing I loved my books, he furnished me
Into a cloven pine, within which rift
I do not love to look on.
That burn by day and night. And then I loved thee,
And I the King shall love thee.

```
And I the King shall love thee.
All wound with adders, who with cloven tongues
    She loved not the savor of tar nor of pitch,
MIRANDA  Do you love me?
Do love, prize, honor you.
And his and mine loved darling. [He exits, above.]
Were but my trials of thy love, and thou
With such love as 'tis now, the murkiest den,
    Do you love me, master? No?
Whose shadow the dismissed bachelor loves,
A contract of true love to celebrate,
On the blest lovers.
A contract of true love. Be not too late.
FERDINAND  No, my dearest love,
Of these our dear-beloved solemnized,


Process finished with exit code 0
```

ID and Password

199

Lkdmw26@