

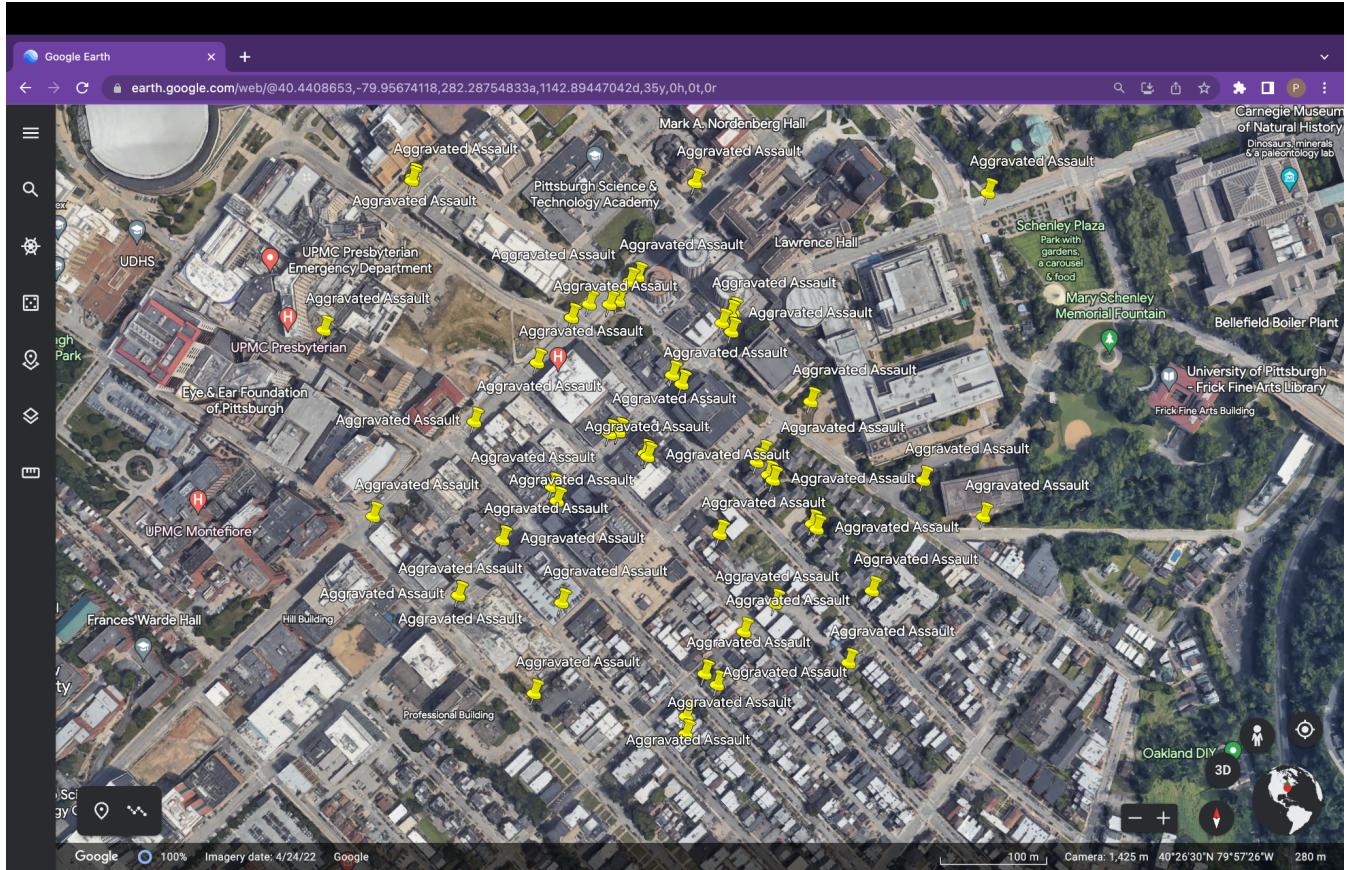
Praveen Ramesh  
pramesh2@andrew.cmu.edu

## Project 5

Cluster ID: student119  
Password: Pramesh@2

### Task7

screenshot of Google Earth from Part 1, Task 7



## Part 2 Code

```
/*
@Author : Praveen Ramesh
@Andrew ID: pramesh2@andrew.cmu.edu
This Java application utilizes Spark to perform text analysis on a given input
text file. The class provides
various functionalities such as counting lines, words, distinct words,
symbols, distinct symbols, and distinct letters.
Additionally, it can return lines containing a specific word. The
implementation relies on Spark's built-in methods,
such as flatMap, count, and distinct, to carry out these operations.
*/
//import call the required packages
import org.apache.spark.SparkConf;
import org.apache.spark.api.java.JavaRDD;
import org.apache.spark.api.java.JavaSparkContext;
import java.util.Arrays;
import java.util.List;
import java.util.Scanner;
import org.apache.spark.api.java.function.Function;

/*
The public class ShakespeareAnalytics contains all the necessary methods and
functionalities to perform text
analysis operations on a given input text file.
*/
public class ShakespeareAnalytics {
    /*lineCount method counts the number of lines in the textfile using count
method of the JavaRDD class
     * @param fileName
     */
    private static void lineCount(String fileName) {
        //Set the configuration for spark
        SparkConf sparkConf = new SparkConf().setMaster("local").setAppName("JD
Line Counter");
        //Create a JavaSparkContext object
        JavaSparkContext sparkContext = new JavaSparkContext(sparkConf);
        //Read the inputfile through textFile method
        JavaRDD<String> inputFile = sparkContext.textFile(fileName);
        //Get the number of lines through Count() method of JavaRDD
        long lineCount = inputFile.count();
        //Print the number of lines
        System.out.println("Number of lines: " + lineCount);
        //Stop the spark context
        sparkContext.stop();
    }
}
```

```

/**
 * wordCount method of ShakespeareAnalytics counts the number of words in
the textfile using flatMap
 * count() methods of JavaRDD class
 * @param fileName
 */
private static void wordCount(String fileName) {
    //Set the configuration for spark
    SparkConf sparkConf = new SparkConf().setMaster("local").setAppName("JD
Word Counter");
    //Create a JavaSparkContext object
    JavaSparkContext sparkContext = new JavaSparkContext(sparkConf);
    //Read the inputfile through textFile method
    JavaRDD<String> inputFile = sparkContext.textFile(fileName);
    //Filter to check the empty words
    Function<String, Boolean> filter = k -> (!k.isEmpty());
    //Apply flatmap to RDD to split each line into words through the
specified regex and filter out empty words
    JavaRDD<String> words = inputFile.flatMap(content ->
Arrays.asList(content.split("[^a-zA-Z]+")))
        .filter(filter);
    //Get the number of words through Count() method of JavaRDD
    long wordCount = words.count();
    //Print the number of words
    System.out.println("Number of words: " + wordCount);
    //Stop the spark context
    sparkContext.stop();
}

/**
 * distinctWordCount method of ShakespeareAnalytics counts the distinct
number of words in the textfile using flatMap,
 * count and distinct() methods
 * @param fileName
 */
private static void distinctWordCount(String fileName) {
    //Set the configuration for spark
    SparkConf sparkConf = new SparkConf().setMaster("local").setAppName("JD
Distinct Word Counter");
    //Create a JavaSparkContext object
    JavaSparkContext sparkContext = new JavaSparkContext(sparkConf);
    //Read the inputfile through textFile method
    JavaRDD<String> inputFile = sparkContext.textFile(fileName);
    //Filter to check the empty words
    Function<String, Boolean> filter = k -> (!k.isEmpty());
    //Apply flatmap to RDD to split each line into words through the
specified regex and filter out empty words
    JavaRDD<String> words = inputFile.flatMap(content ->
Arrays.asList(content.split("[^a-zA-Z]+")))

```

```

        .filter(filter);
    //Get the distinct words through distinct() method
    JavaRDD<String> distinctWords = words.distinct();
    //Get the distinct number of words through Count() method of JavaRDD
    long distinctWordsCount = distinctWords.count();
    //Print the distinct number of words
    System.out.println("Number of distinct words: " + distinctWordsCount);
    //Stop the spark context
    sparkContext.stop();
}

/**
 * countSymbols method of ShakespeareAnalytics counts the number of symbols
in the textfile using flatMap,
 * count methods
 * @param fileName
 */
private static void countSymbols(String fileName) {
    //Set the configuration for spark
    SparkConf sparkConf = new SparkConf().setMaster("local").setAppName("JD
Symbol Counter");
    //Create a JavaSparkContext object
    JavaSparkContext sparkContext = new JavaSparkContext(sparkConf);
    //Read the inputfile through textFile method
    JavaRDD<String> inputFile = sparkContext.textFile(fileName);
    //Apply flatmap to RDD to split each line into individual characters
and special symbol
    JavaRDD<String> symbols = inputFile.flatMap(content ->
Arrays.asList(content.split(""))));
    //Get the number of symbols through Count() method of JavaRDD
    long symbolsCount = symbols.count();
    //Print the number of symbols
    System.out.println("Number of symbols: " + symbolsCount);
    //Stop the spark context
    sparkContext.stop();
}

/**
 * distinctSymbolCount method of ShakespeareAnalytics counts the distinct
number of symbols in the textfile using flatMap,
 * count and distinct methods
 * @param fileName
 */
private static void distinctSymbolCount(String fileName) {
    //Set the configuration for spark
    SparkConf sparkConf = new SparkConf().setMaster("local").setAppName("JD
Symbol Counter");
    //Create a JavaSparkContext object
    JavaSparkContext sparkContext = new JavaSparkContext(sparkConf);

```

```

    //Read the inputfile through textFile method
    JavaRDD<String> inputFile = sparkContext.textFile(fileName);
    //Apply flatmap to RDD to split each line into individual characters
    //and special symbol
    JavaRDD<String> symbols = inputFile.flatMap(content ->
Arrays.asList(content.split(""))));
    //Get the distinct symbols through distinct() method
    JavaRDD<String> distinctSymbols = symbols.distinct();
    //Get the distinct number of symbols through Count() method of JavaRDD
    long distinctSymbolsCount = distinctSymbols.count();
    //Print the distinct number of symbols
    System.out.println("Number of distinct symbols: " +
distinctSymbolsCount);
    //Stop the spark context
    sparkContext.stop();
}

/**
 * distinctLetterCount method of ShakespeareAnalytics counts the distinct
 * number of letters in the textfile using flatMap,
 * count and distinct methods
 * @param fileName
 */
private static void distinctLetterCount(String fileName) {
    //Set the configuration for spark
    SparkConf sparkConf = new SparkConf().setMaster("local").setAppName("JD
Distinct Letter Counter");
    //Create a JavaSparkContext object
    JavaSparkContext sparkContext = new JavaSparkContext(sparkConf);
    //Read the inputfile through textFile method
    JavaRDD<String> inputFile = sparkContext.textFile(fileName);
    //Apply flatmap to RDD to split each line into individual symbols and
    //apply filter to get the letters alone
    JavaRDD<String> letters = inputFile.flatMap(content ->
Arrays.asList(content.split(""))))
        .filter(letter -> letter.matches("[a-zA-Z]"));
    //Get the distinct letters through distinct() method
    JavaRDD<String> distinctLetters = letters.distinct();
    //Get the distinct number of letters through Count() method of JavaRDD
    long distinctLetterCount = distinctLetters.count();
    //Print the distinct number of letters
    System.out.println("Number of distinct letters: " +
distinctLetterCount);
    //Stop the spark context
    sparkContext.stop();
}

/**

```

```

    * lineWordMatch method of ShakespeareAnalytics displays all the lines in
the textfile which contains the searchword
    * It performs a case-sensitive match through contains() method
    * @param fileName
    * @param searchWord
    */
private static void lineWordMatch(String fileName, String searchWord) {
    //Set the configuration for spark
    SparkConf sparkConf = new SparkConf().setMaster("local").setAppName("JD
Lines with Word");
    //Create a JavaSparkContext object
    JavaSparkContext sparkContext = new JavaSparkContext(sparkConf);
    //Read the inputfile through textFile method
    JavaRDD<String> inputFile = sparkContext.textFile(fileName);
    //Apply filter to get the lines which has the searchWord in it through
contains() which does a case-sensitive match
    JavaRDD<String> lines = inputFile.filter(line ->
line.contains(searchWord));
    //Get the lines in a list of string through collect() method
    List<String> filteredResult = lines.collect();
    //Stop the spark context
    sparkContext.stop();
    //Print the lines
    System.out.println("Lines containing the word " + searchWord + ":" );
    for (String line : filteredResult) {
        System.out.println(line);
    }
}

/**
 * Main method calls each of the functionalities with the filename give in
the command line arguments
 * @param args
 */
public static void main(String[] args) {
    // Check if any input files were provided
    if (args.length == 0) {
        System.out.println("No files provided.");
        System.exit(0);
    }
    //Initialize a scanner object
    Scanner scanner = new Scanner(System.in);
    //call various methods to display the respective counts
    lineCount(args[0]);
    wordCount(args[0]);
    distinctWordCount(args[0]);
    countSymbols(args[0]);
    distinctSymbolCount(args[0]);
    distinctLetterCount(args[0]);
}

```

```

//      Get the word to find the match
System.out.println("Enter a word :");
String searchWord = scanner.nextLine();
//call the lineWordMatch with the input file and searchWord
lineWordMatch(args[0],searchWord);
}
}

```

## Part 2

### Task 0

#### Output

23/04/21 00:23:31 INFO spark.SecurityManager: Changing view acls to: praveenramesh  
 23/04/21 00:23:31 INFO spark.SecurityManager: SecurityManager: authentication disabled; ui  
 acls disabled; users with view permissions: Set(praveenramesh)  
 23/04/21 00:23:31 INFO slf4j.Slf4jLogger: Slf4jLogger started  
 23/04/21 00:23:31 INFO Remoting: Starting remoting  
 23/04/21 00:23:32 INFO Remoting: Remoting started; listening on addresses  
 :[akka.tcp://spark@10.0.0.18:54515]  
 23/04/21 00:23:32 INFO Remoting: Remoting now listens on addresses:  
 [akka.tcp://spark@10.0.0.18:54515]  
 23/04/21 00:23:32 INFO spark.SparkEnv: Registering MapOutputTracker  
 23/04/21 00:23:32 INFO spark.SparkEnv: Registering BlockManagerMaster  
 23/04/21 00:23:32 INFO storage.DiskBlockManager: Created local directory at  
 /var/folders/1x/z4spc31954n2gycqnlph1rf40000gn/T/spark-local-20230421002332-5bf3  
 23/04/21 00:23:32 INFO storage.MemoryStore: MemoryStore started with capacity 2.1 GB.  
 23/04/21 00:23:32 INFO network.ConnectionManager: Bound socket to port 54516 with id =  
 ConnectionManagerId(10.0.0.18,54516)  
 23/04/21 00:23:32 INFO storage.BlockManagerMaster: Trying to register BlockManager  
 23/04/21 00:23:32 INFO storage.BlockManagerInfo: Registering block manager  
 10.0.0.18:54516 with 2.1 GB RAM  
 23/04/21 00:23:32 INFO storage.BlockManagerMaster: Registered BlockManager  
 23/04/21 00:23:32 INFO spark.HttpServer: Starting HTTP Server  
 23/04/21 00:23:32 INFO server.Server: jetty-8.1.14.v20131031  
 23/04/21 00:23:32 INFO server.AbstractConnector: Started SocketConnector@0.0.0.0:54517  
 23/04/21 00:23:32 INFO broadcast.HttpBroadcast: Broadcast server started at  
 http://10.0.0.18:54517  
 23/04/21 00:23:32 INFO spark.HttpFileServer: HTTP File server directory is  
 /var/folders/1x/z4spc31954n2gycqnlph1rf40000gn/T/spark-f059823a-3373-4cb8-b1b0-b336c22  
 446b2  
 23/04/21 00:23:32 INFO spark.HttpServer: Starting HTTP Server  
 23/04/21 00:23:32 INFO server.Server: jetty-8.1.14.v20131031  
 23/04/21 00:23:32 INFO server.AbstractConnector: Started SocketConnector@0.0.0.0:54518

23/04/21 00:23:32 INFO server.Server: jetty-8.1.14.v20131031  
23/04/21 00:23:32 INFO server.AbstractConnector: Started  
SelectChannelConnector@0.0.0.0:4040  
23/04/21 00:23:32 INFO ui.SparkUI: Started SparkUI at http://10.0.0.18:4040  
23/04/21 00:23:32 INFO storage.MemoryStore: ensureFreeSpace(32816) called with  
curMem=0, maxMem=2290719129  
23/04/21 00:23:32 INFO storage.MemoryStore: Block broadcast\_0 stored as values to memory  
(estimated size 32.0 KB, free 2.1 GB)  
23/04/21 00:23:32 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your  
platform... using builtin-java classes where applicable  
23/04/21 00:23:32 WARN snappy.LoadSnappy: Snappy native library not loaded  
23/04/21 00:23:32 INFO mapred.FileInputFormat: Total input paths to process : 1  
23/04/21 00:23:32 INFO spark.SparkContext: Starting job: count at  
ShakespeareAnalytics.java:34  
23/04/21 00:23:32 INFO scheduler.DAGScheduler: Got job 0 (count at  
ShakespeareAnalytics.java:34) with 1 output partitions (allowLocal=false)  
23/04/21 00:23:32 INFO scheduler.DAGScheduler: Final stage: Stage 0(count at  
ShakespeareAnalytics.java:34)  
23/04/21 00:23:32 INFO scheduler.DAGScheduler: Parents of final stage: List()  
23/04/21 00:23:32 INFO scheduler.DAGScheduler: Missing parents: List()  
23/04/21 00:23:32 INFO scheduler.DAGScheduler: Submitting Stage 0 (MappedRDD[1] at  
textFile at ShakespeareAnalytics.java:32), which has no missing parents  
23/04/21 00:23:32 INFO scheduler.DAGScheduler: Submitting 1 missing tasks from Stage 0  
(MappedRDD[1] at textFile at ShakespeareAnalytics.java:32)  
23/04/21 00:23:32 INFO scheduler.TaskSchedulerImpl: Adding task set 0.0 with 1 tasks  
23/04/21 00:23:32 INFO scheduler.TaskSetManager: Starting task 0.0:0 as TID 0 on executor  
localhost: localhost (PROCESS\_LOCAL)  
23/04/21 00:23:32 INFO scheduler.TaskSetManager: Serialized task 0.0:0 as 1750 bytes in 1  
ms  
23/04/21 00:23:32 INFO executor.Executor: Running task ID 0  
23/04/21 00:23:32 INFO storage.BlockManager: Found block broadcast\_0 locally  
23/04/21 00:23:32 INFO rdd.HadoopRDD: Input split: file:/Users/praveenramesh/Distributed  
Systems/Project5Part2/AllsWellThatEndsWell:0+133886  
23/04/21 00:23:32 INFO executor.Executor: Serialized size of result for 0 is 597  
23/04/21 00:23:32 INFO executor.Executor: Sending result for 0 directly to driver  
23/04/21 00:23:32 INFO executor.Executor: Finished task ID 0  
23/04/21 00:23:32 INFO scheduler.TaskSetManager: Finished TID 0 in 35 ms on localhost  
(progress: 1/1)  
23/04/21 00:23:32 INFO scheduler.TaskSchedulerImpl: Removed TaskSet 0.0, whose tasks  
have all completed, from pool  
23/04/21 00:23:32 INFO scheduler.DAGScheduler: Completed ResultTask(0, 0)  
23/04/21 00:23:32 INFO scheduler.DAGScheduler: Stage 0 (count at  
ShakespeareAnalytics.java:34) finished in 0.039 s

23/04/21 00:23:32 INFO spark.SparkContext: Job finished: count at ShakespeareAnalytics.java:34, took 0.074483791 s

**Number of lines: 4570**

23/04/21 00:23:32 INFO handler.ContextHandler: stopped  
o.e.j.s.ServletContextHandler{/metrics/json,null}  
23/04/21 00:23:32 INFO handler.ContextHandler: stopped  
o.e.j.s.ServletContextHandler{/stages/stage/kill,null}  
23/04/21 00:23:32 INFO handler.ContextHandler: stopped o.e.j.s.ServletContextHandler{/,null}  
23/04/21 00:23:32 INFO handler.ContextHandler: stopped  
o.e.j.s.ServletContextHandler{/static,null}  
23/04/21 00:23:32 INFO handler.ContextHandler: stopped  
o.e.j.s.ServletContextHandler{/executors/json,null}  
23/04/21 00:23:32 INFO handler.ContextHandler: stopped  
o.e.j.s.ServletContextHandler{/executors,null}  
23/04/21 00:23:32 INFO handler.ContextHandler: stopped  
o.e.j.s.ServletContextHandler{/environment/json,null}  
23/04/21 00:23:32 INFO handler.ContextHandler: stopped  
o.e.j.s.ServletContextHandler{/environment,null}  
23/04/21 00:23:32 INFO handler.ContextHandler: stopped  
o.e.j.s.ServletContextHandler{/storage/rdd/json,null}  
23/04/21 00:23:32 INFO handler.ContextHandler: stopped  
o.e.j.s.ServletContextHandler{/storage/rdd,null}  
23/04/21 00:23:32 INFO handler.ContextHandler: stopped  
o.e.j.s.ServletContextHandler{/storage/json,null}  
23/04/21 00:23:32 INFO handler.ContextHandler: stopped  
o.e.j.s.ServletContextHandler{/storage,null}  
23/04/21 00:23:32 INFO handler.ContextHandler: stopped  
o.e.j.s.ServletContextHandler{/stages/pool/json,null}  
23/04/21 00:23:32 INFO handler.ContextHandler: stopped  
o.e.j.s.ServletContextHandler{/stages/pool,null}  
23/04/21 00:23:32 INFO handler.ContextHandler: stopped  
o.e.j.s.ServletContextHandler{/stages/stage/json,null}  
23/04/21 00:23:32 INFO handler.ContextHandler: stopped  
o.e.j.s.ServletContextHandler{/stages/stage,null}  
23/04/21 00:23:32 INFO handler.ContextHandler: stopped  
o.e.j.s.ServletContextHandler{/stages/json,null}  
23/04/21 00:23:32 INFO handler.ContextHandler: stopped  
o.e.j.s.ServletContextHandler{/stages,null}  
23/04/21 00:23:32 INFO ui.SparkUI: Stopped Spark web UI at http://10.0.0.18:4040  
23/04/21 00:23:32 INFO scheduler.DAGScheduler: Stopping DAGScheduler  
23/04/21 00:23:33 INFO spark.MapOutputTrackerMasterActor: MapOutputTrackerActor stopped!  
23/04/21 00:23:33 INFO network.ConnectionManager: Selector thread was interrupted!  
23/04/21 00:23:33 INFO network.ConnectionManager: ConnectionManager stopped

```
23/04/21 00:23:33 INFO storage.MemoryStore: MemoryStore cleared
23/04/21 00:23:33 INFO storage.BlockManager: BlockManager stopped
23/04/21 00:23:33 INFO storage.BlockManagerMasterActor: Stopping BlockManagerMaster
23/04/21 00:23:33 INFO storage.BlockManagerMaster: BlockManagerMaster stopped
23/04/21 00:23:33 INFO spark.SparkContext: Successfully stopped SparkContext
```

Process finished with exit code 0

## Task 1

### Output

```
23/04/21 00:30:45 WARN util.Utils: Your hostname, Praveens-MacBook-Pro-177.local resolves
to a loopback address: 127.0.0.1; using 172.31.78.32 instead (on interface utun5)
23/04/21 00:30:45 WARN util.Utils: Set SPARK_LOCAL_IP if you need to bind to another
address
23/04/21 00:30:45 INFO spark.SecurityManager: Changing view acls to: praveenramesh
23/04/21 00:30:45 INFO spark.SecurityManager: SecurityManager: authentication disabled; ui
acls disabled; users with view permissions: Set(praveenramesh)
23/04/21 00:30:45 INFO slf4j.Slf4jLogger: Slf4jLogger started
23/04/21 00:30:45 INFO Remoting: Starting remoting
23/04/21 00:30:45 INFO Remoting: Remoting started; listening on addresses
:[akka.tcp://spark@172.31.78.32:54557]
23/04/21 00:30:45 INFO Remoting: Remoting now listens on addresses:
[akka.tcp://spark@172.31.78.32:54557]
23/04/21 00:30:45 INFO spark.SparkEnv: Registering MapOutputTracker
23/04/21 00:30:45 INFO spark.SparkEnv: Registering BlockManagerMaster
23/04/21 00:30:45 INFO storage.DiskBlockManager: Created local directory at
/var/folders/1x/z4spc31954n2gycqnlph1rf40000gn/T/spark-local-20230421003045-0771
23/04/21 00:30:45 INFO storage.MemoryStore: MemoryStore started with capacity 2.1 GB.
23/04/21 00:30:45 INFO network.ConnectionManager: Bound socket to port 54558 with id =
ConnectionManagerId(172.31.78.32,54558)
23/04/21 00:30:45 INFO storage.BlockManagerMaster: Trying to register BlockManager
23/04/21 00:30:45 INFO storage.BlockManagerInfo: Registering block manager
172.31.78.32:54558 with 2.1 GB RAM
23/04/21 00:30:45 INFO storage.BlockManagerMaster: Registered BlockManager
23/04/21 00:30:45 INFO spark.HttpServer: Starting HTTP Server
23/04/21 00:30:45 INFO server.Server: jetty-8.1.14.v20131031
23/04/21 00:30:45 INFO server.AbstractConnector: Started SocketConnector@0.0.0.0:54559
23/04/21 00:30:45 INFO broadcast.HttpBroadcast: Broadcast server started at
http://172.31.78.32:54559
```

23/04/21 00:30:45 INFO spark.HttpFileServer: HTTP File server directory is /var/folders/1x/z4spc31954n2gycqnlph1rf40000gn/T/spark-ea08071c-475b-4381-a729-d0c21a8a0617

23/04/21 00:30:45 INFO spark.HttpServer: Starting HTTP Server

23/04/21 00:30:45 INFO server.Server: jetty-8.1.14.v20131031

23/04/21 00:30:45 INFO server.AbstractConnector: Started SocketConnector@0.0.0.0:54560

23/04/21 00:30:45 INFO server.Server: jetty-8.1.14.v20131031

23/04/21 00:30:45 INFO server.AbstractConnector: Started SelectChannelConnector@0.0.0.0:4040

23/04/21 00:30:45 INFO ui.SparkUI: Started SparkUI at http://172.31.78.32:4040

23/04/21 00:30:45 INFO storage.MemoryStore: ensureFreeSpace(32816) called with curMem=0, maxMem=2290719129

23/04/21 00:30:45 INFO storage.MemoryStore: Block broadcast\_0 stored as values to memory (estimated size 32.0 KB, free 2.1 GB)

23/04/21 00:30:45 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable

23/04/21 00:30:45 WARN snappy.LoadSnappy: Snappy native library not loaded

23/04/21 00:30:45 INFO mapred.FileInputFormat: Total input paths to process : 1

23/04/21 00:30:45 INFO spark.SparkContext: Starting job: count at ShakespeareAnalytics.java:60

23/04/21 00:30:45 INFO scheduler.DAGScheduler: Got job 0 (count at ShakespeareAnalytics.java:60) with 1 output partitions (allowLocal=false)

23/04/21 00:30:45 INFO scheduler.DAGScheduler: Final stage: Stage 0(count at ShakespeareAnalytics.java:60)

23/04/21 00:30:45 INFO scheduler.DAGScheduler: Parents of final stage: List()

23/04/21 00:30:45 INFO scheduler.DAGScheduler: Missing parents: List()

23/04/21 00:30:45 INFO scheduler.DAGScheduler: Submitting Stage 0 (FilteredRDD[3] at filter at ShakespeareAnalytics.java:58), which has no missing parents

23/04/21 00:30:45 INFO scheduler.DAGScheduler: Submitting 1 missing tasks from Stage 0 (FilteredRDD[3] at filter at ShakespeareAnalytics.java:58)

23/04/21 00:30:45 INFO scheduler.TaskSchedulerImpl: Adding task set 0.0 with 1 tasks

23/04/21 00:30:45 INFO scheduler.TaskSetManager: Starting task 0.0:0 as TID 0 on executor localhost: localhost (PROCESS\_LOCAL)

23/04/21 00:30:45 INFO scheduler.TaskSetManager: Serialized task 0.0:0 as 2344 bytes in 1 ms

23/04/21 00:30:45 INFO executor.Executor: Running task ID 0

23/04/21 00:30:45 INFO storage.BlockManager: Found block broadcast\_0 locally

23/04/21 00:30:45 INFO rdd.HadoopRDD: Input split: file:/Users/praveenramesh/Distributed Systems/Project5Part2/AllsWellThatEndsWell:0+133886

23/04/21 00:30:45 INFO executor.Executor: Serialized size of result for 0 is 597

23/04/21 00:30:45 INFO executor.Executor: Sending result for 0 directly to driver

23/04/21 00:30:45 INFO executor.Executor: Finished task ID 0

23/04/21 00:30:45 INFO scheduler.TaskSetManager: Finished TID 0 in 69 ms on localhost (progress: 1/1)

23/04/21 00:30:45 INFO scheduler.TaskSchedulerImpl: Removed TaskSet 0.0, whose tasks have all completed, from pool

23/04/21 00:30:45 INFO scheduler.DAGScheduler: Completed ResultTask(0, 0)

23/04/21 00:30:45 INFO scheduler.DAGScheduler: Stage 0 (count at ShakespeareAnalytics.java:60) finished in 0.074 s

23/04/21 00:30:45 INFO spark.SparkContext: Job finished: count at ShakespeareAnalytics.java:60, took 0.122313208 s

### **Number of words: 25085**

23/04/21 00:30:45 INFO handler.ContextHandler: stopped

o.e.j.s.ServletContextHandler{/metrics/json,null}

23/04/21 00:30:45 INFO handler.ContextHandler: stopped

o.e.j.s.ServletContextHandler{/stages/stage/kill,null}

23/04/21 00:30:45 INFO handler.ContextHandler: stopped o.e.j.s.ServletContextHandler{/null}

23/04/21 00:30:45 INFO handler.ContextHandler: stopped

o.e.j.s.ServletContextHandler{/static,null}

23/04/21 00:30:45 INFO handler.ContextHandler: stopped

o.e.j.s.ServletContextHandler{/executors/json,null}

23/04/21 00:30:45 INFO handler.ContextHandler: stopped

o.e.j.s.ServletContextHandler{/executors,null}

23/04/21 00:30:45 INFO handler.ContextHandler: stopped

o.e.j.s.ServletContextHandler{/environment/json,null}

23/04/21 00:30:45 INFO handler.ContextHandler: stopped

o.e.j.s.ServletContextHandler{/environment,null}

23/04/21 00:30:45 INFO handler.ContextHandler: stopped

o.e.j.s.ServletContextHandler{/storage/rdd/json,null}

23/04/21 00:30:45 INFO handler.ContextHandler: stopped

o.e.j.s.ServletContextHandler{/storage/rdd,null}

23/04/21 00:30:45 INFO handler.ContextHandler: stopped

o.e.j.s.ServletContextHandler{/storage/json,null}

23/04/21 00:30:45 INFO handler.ContextHandler: stopped

o.e.j.s.ServletContextHandler{/storage,null}

23/04/21 00:30:45 INFO handler.ContextHandler: stopped

o.e.j.s.ServletContextHandler{/stages/pool/json,null}

23/04/21 00:30:45 INFO handler.ContextHandler: stopped

o.e.j.s.ServletContextHandler{/stages/pool,null}

23/04/21 00:30:45 INFO handler.ContextHandler: stopped

o.e.j.s.ServletContextHandler{/stages/stage/json,null}

23/04/21 00:30:45 INFO handler.ContextHandler: stopped

o.e.j.s.ServletContextHandler{/stages/stage,null}

23/04/21 00:30:45 INFO handler.ContextHandler: stopped

o.e.j.s.ServletContextHandler{/stages/json,null}

23/04/21 00:30:45 INFO handler.ContextHandler: stopped

o.e.j.s.ServletContextHandler{/stages,null}

23/04/21 00:30:46 INFO ui.SparkUI: Stopped Spark web UI at http://172.31.78.32:4040

```
23/04/21 00:30:46 INFO scheduler.DAGScheduler: Stopping DAGScheduler
23/04/21 00:30:47 INFO spark.MapOutputTrackerMasterActor: MapOutputTrackerActor
stopped!
23/04/21 00:30:47 INFO network.ConnectionManager: Selector thread was interrupted!
23/04/21 00:30:47 INFO network.ConnectionManager: ConnectionManager stopped
23/04/21 00:30:47 INFO storage.MemoryStore: MemoryStore cleared
23/04/21 00:30:47 INFO storage.BlockManager: BlockManager stopped
23/04/21 00:30:47 INFO storage.BlockManagerMasterActor: Stopping BlockManagerMaster
23/04/21 00:30:47 INFO storage.BlockManagerMaster: BlockManagerMaster stopped
23/04/21 00:30:47 INFO spark.SparkContext: Successfully stopped SparkContext
```

Process finished with exit code 0

## Task 2

### Output

```
23/04/21 00:33:53 INFO spark.SecurityManager: Changing view acls to: praveenramesh
23/04/21 00:33:53 INFO spark.SecurityManager: SecurityManager: authentication disabled; ui
acls disabled; users with view permissions: Set(praveenramesh)
23/04/21 00:33:53 INFO slf4j.Slf4jLogger: Slf4jLogger started
23/04/21 00:33:53 INFO Remoting: Starting remoting
23/04/21 00:33:53 INFO Remoting: Remoting started; listening on addresses
:[akka.tcp://spark@10.0.0.18:54571]
23/04/21 00:33:53 INFO Remoting: Remoting now listens on addresses:
[akka.tcp://spark@10.0.0.18:54571]
23/04/21 00:33:53 INFO spark.SparkEnv: Registering MapOutputTracker
23/04/21 00:33:53 INFO spark.SparkEnv: Registering BlockManagerMaster
23/04/21 00:33:53 INFO storage.DiskBlockManager: Created local directory at
/var/folders/1x/z4spc31954n2gycqnlph1rf40000gn/T/spark-local-20230421003353-5e45
23/04/21 00:33:53 INFO storage.MemoryStore: MemoryStore started with capacity 2.1 GB.
23/04/21 00:33:53 INFO network.ConnectionManager: Bound socket to port 54572 with id =
ConnectionManagerId(10.0.0.18,54572)
23/04/21 00:33:53 INFO storage.BlockManagerMaster: Trying to register BlockManager
23/04/21 00:33:53 INFO storage.BlockManagerInfo: Registering block manager
10.0.0.18:54572 with 2.1 GB RAM
23/04/21 00:33:53 INFO storage.BlockManagerMaster: Registered BlockManager
23/04/21 00:33:53 INFO spark.HttpServer: Starting HTTP Server
23/04/21 00:33:53 INFO server.Server: jetty-8.1.14.v20131031
23/04/21 00:33:53 INFO server.AbstractConnector: Started SocketConnector@0.0.0.0:54573
23/04/21 00:33:53 INFO broadcast.HttpBroadcast: Broadcast server started at
http://10.0.0.18:54573
```

23/04/21 00:33:53 INFO spark.HttpFileServer: HTTP File server directory is /var/folders/1x/z4spc31954n2gycqnlph1rf40000gn/T/spark-fdd9b28f-86ce-4fd4-9a5f-92128c90bc5b

23/04/21 00:33:53 INFO spark.HttpServer: Starting HTTP Server

23/04/21 00:33:53 INFO server.Server: jetty-8.1.14.v20131031

23/04/21 00:33:53 INFO server.AbstractConnector: Started SocketConnector@0.0.0.0:54574

23/04/21 00:33:54 INFO server.Server: jetty-8.1.14.v20131031

23/04/21 00:33:54 INFO server.AbstractConnector: Started SelectChannelConnector@0.0.0.0:4040

23/04/21 00:33:54 INFO ui.SparkUI: Started SparkUI at http://10.0.0.18:4040

23/04/21 00:33:54 INFO storage.MemoryStore: ensureFreeSpace(32816) called with curMem=0, maxMem=2290719129

23/04/21 00:33:54 INFO storage.MemoryStore: Block broadcast\_0 stored as values to memory (estimated size 32.0 KB, free 2.1 GB)

23/04/21 00:33:54 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable

23/04/21 00:33:54 WARN snappy.LoadSnappy: Snappy native library not loaded

23/04/21 00:33:54 INFO mapred.FileInputFormat: Total input paths to process : 1

23/04/21 00:33:54 INFO spark.SparkContext: Starting job: count at ShakespeareAnalytics.java:87

23/04/21 00:33:54 INFO scheduler.DAGScheduler: Registering RDD 5 (distinct at ShakespeareAnalytics.java:85)

23/04/21 00:33:54 INFO scheduler.DAGScheduler: Got job 0 (count at ShakespeareAnalytics.java:87) with 1 output partitions (allowLocal=false)

23/04/21 00:33:54 INFO scheduler.DAGScheduler: Final stage: Stage 0(count at ShakespeareAnalytics.java:87)

23/04/21 00:33:54 INFO scheduler.DAGScheduler: Parents of final stage: List(Stage 1)

23/04/21 00:33:54 INFO scheduler.DAGScheduler: Missing parents: List(Stage 1)

23/04/21 00:33:54 INFO scheduler.DAGScheduler: Submitting Stage 1 (MapPartitionsRDD[5] at distinct at ShakespeareAnalytics.java:85), which has no missing parents

23/04/21 00:33:54 INFO scheduler.DAGScheduler: Submitting 1 missing tasks from Stage 1 (MapPartitionsRDD[5] at distinct at ShakespeareAnalytics.java:85)

23/04/21 00:33:54 INFO scheduler.TaskSchedulerImpl: Adding task set 1.0 with 1 tasks

23/04/21 00:33:54 INFO scheduler.TaskSetManager: Starting task 1.0:0 as TID 0 on executor localhost: localhost (PROCESS\_LOCAL)

23/04/21 00:33:54 INFO scheduler.TaskSetManager: Serialized task 1.0:0 as 2648 bytes in 1 ms

23/04/21 00:33:54 INFO executor.Executor: Running task ID 0

23/04/21 00:33:54 INFO storage.BlockManager: Found block broadcast\_0 locally

23/04/21 00:33:54 INFO rdd.HadoopRDD: Input split: file:/Users/praveenramesh/Distributed Systems/Project5Part2/AllsWellThatEndsWell:0+133886

23/04/21 00:33:54 INFO executor.Executor: Serialized size of result for 0 is 779

23/04/21 00:33:54 INFO executor.Executor: Sending result for 0 directly to driver

23/04/21 00:33:54 INFO executor.Executor: Finished task ID 0

23/04/21 00:33:54 INFO scheduler.TaskSetManager: Finished TID 0 in 143 ms on localhost  
(progress: 1/1)

23/04/21 00:33:54 INFO scheduler.DAGScheduler: Completed ShuffleMapTask(1, 0)

23/04/21 00:33:54 INFO scheduler.TaskSchedulerImpl: Removed TaskSet 1.0, whose tasks have all completed, from pool

23/04/21 00:33:54 INFO scheduler.DAGScheduler: Stage 1 (distinct at ShakespeareAnalytics.java:85) finished in 0.147 s

23/04/21 00:33:54 INFO scheduler.DAGScheduler: looking for newly runnable stages

23/04/21 00:33:54 INFO scheduler.DAGScheduler: running: Set()

23/04/21 00:33:54 INFO scheduler.DAGScheduler: waiting: Set(Stage 0)

23/04/21 00:33:54 INFO scheduler.DAGScheduler: failed: Set()

23/04/21 00:33:54 INFO scheduler.DAGScheduler: Missing parents for Stage 0: List()

23/04/21 00:33:54 INFO scheduler.DAGScheduler: Submitting Stage 0 (MappedRDD[8] at distinct at ShakespeareAnalytics.java:85), which is now runnable

23/04/21 00:33:54 INFO scheduler.DAGScheduler: Submitting 1 missing tasks from Stage 0 (MappedRDD[8] at distinct at ShakespeareAnalytics.java:85)

23/04/21 00:33:54 INFO scheduler.TaskSchedulerImpl: Adding task set 0.0 with 1 tasks

23/04/21 00:33:54 INFO scheduler.TaskSetManager: Starting task 0.0:0 as TID 1 on executor localhost: localhost (PROCESS\_LOCAL)

23/04/21 00:33:54 INFO scheduler.TaskSetManager: Serialized task 0.0:0 as 2489 bytes in 0 ms

23/04/21 00:33:54 INFO executor.Executor: Running task ID 1

23/04/21 00:33:54 INFO storage.BlockManager: Found block broadcast\_0 locally

23/04/21 00:33:54 INFO storage.BlockFetcherIterator\$BasicBlockFetcherIterator: maxBytesInFlight: 50331648, targetRequestSize: 10066329

23/04/21 00:33:54 INFO storage.BlockFetcherIterator\$BasicBlockFetcherIterator: Getting 1 non-empty blocks out of 1 blocks

23/04/21 00:33:54 INFO storage.BlockFetcherIterator\$BasicBlockFetcherIterator: Started 0 remote fetches in 2 ms

23/04/21 00:33:54 INFO executor.Executor: Serialized size of result for 1 is 863

23/04/21 00:33:54 INFO executor.Executor: Sending result for 1 directly to driver

23/04/21 00:33:54 INFO executor.Executor: Finished task ID 1

23/04/21 00:33:54 INFO scheduler.TaskSetManager: Finished TID 1 in 39 ms on localhost  
(progress: 1/1)

23/04/21 00:33:54 INFO scheduler.TaskSchedulerImpl: Removed TaskSet 0.0, whose tasks have all completed, from pool

23/04/21 00:33:54 INFO scheduler.DAGScheduler: Completed ResultTask(0, 0)

23/04/21 00:33:54 INFO scheduler.DAGScheduler: Stage 0 (count at ShakespeareAnalytics.java:87) finished in 0.039 s

23/04/21 00:33:54 INFO spark.SparkContext: Job finished: count at ShakespeareAnalytics.java:87, took 0.241093625 s

**Number of distinct words: 3866**

23/04/21 00:33:54 INFO handler.ContextHandler: stopped  
o.e.j.s.ServletContextHandler{/metrics/json,null}

23/04/21 00:33:54 INFO handler.ContextHandler: stopped  
o.e.j.s.ServletContextHandler{/stages/stage/kill,null}  
23/04/21 00:33:54 INFO handler.ContextHandler: stopped o.e.j.s.ServletContextHandler{/,null}  
23/04/21 00:33:54 INFO handler.ContextHandler: stopped  
o.e.j.s.ServletContextHandler{/static,null}  
23/04/21 00:33:54 INFO handler.ContextHandler: stopped  
o.e.j.s.ServletContextHandler{/executors/json,null}  
23/04/21 00:33:54 INFO handler.ContextHandler: stopped  
o.e.j.s.ServletContextHandler{/executors,null}  
23/04/21 00:33:54 INFO handler.ContextHandler: stopped  
o.e.j.s.ServletContextHandler{/environment/json,null}  
23/04/21 00:33:54 INFO handler.ContextHandler: stopped  
o.e.j.s.ServletContextHandler{/environment,null}  
23/04/21 00:33:54 INFO handler.ContextHandler: stopped  
o.e.j.s.ServletContextHandler{/storage/rdd/json,null}  
23/04/21 00:33:54 INFO handler.ContextHandler: stopped  
o.e.j.s.ServletContextHandler{/storage/rdd,null}  
23/04/21 00:33:54 INFO handler.ContextHandler: stopped  
o.e.j.s.ServletContextHandler{/storage/json,null}  
23/04/21 00:33:54 INFO handler.ContextHandler: stopped  
o.e.j.s.ServletContextHandler{/storage,null}  
23/04/21 00:33:54 INFO handler.ContextHandler: stopped  
o.e.j.s.ServletContextHandler{/stages/pool/json,null}  
23/04/21 00:33:54 INFO handler.ContextHandler: stopped  
o.e.j.s.ServletContextHandler{/stages/pool,null}  
23/04/21 00:33:54 INFO handler.ContextHandler: stopped  
o.e.j.s.ServletContextHandler{/stages/stage/json,null}  
23/04/21 00:33:54 INFO handler.ContextHandler: stopped  
o.e.j.s.ServletContextHandler{/stages/stage,null}  
23/04/21 00:33:54 INFO handler.ContextHandler: stopped  
o.e.j.s.ServletContextHandler{/stages/json,null}  
23/04/21 00:33:54 INFO handler.ContextHandler: stopped  
o.e.j.s.ServletContextHandler{/stages,null}  
23/04/21 00:33:54 INFO ui.SparkUI: Stopped Spark web UI at http://10.0.0.18:4040  
23/04/21 00:33:54 INFO scheduler.DAGScheduler: Stopping DAGScheduler  
23/04/21 00:33:55 INFO spark.MapOutputTrackerMasterActor: MapOutputTrackerActor  
stopped!  
23/04/21 00:33:55 INFO network.ConnectionManager: Selector thread was interrupted!  
23/04/21 00:33:55 INFO network.ConnectionManager: ConnectionManager stopped  
23/04/21 00:33:55 INFO storage.MemoryStore: MemoryStore cleared  
23/04/21 00:33:55 INFO storage.BlockManager: BlockManager stopped  
23/04/21 00:33:55 INFO storage.BlockManagerMasterActor: Stopping BlockManagerMaster  
23/04/21 00:33:55 INFO storage.BlockManagerMaster: BlockManagerMaster stopped  
23/04/21 00:33:55 INFO spark.SparkContext: Successfully stopped SparkContext

Process finished with exit code 0

## Task 3 Output

```
23/04/21 00:38:17 INFO spark.SecurityManager: Changing view acls to: praveenramesh
23/04/21 00:38:17 INFO spark.SecurityManager: SecurityManager: authentication disabled; ui
acls disabled; users with view permissions: Set(praveenramesh)
23/04/21 00:38:17 INFO slf4j.Slf4jLogger: Slf4jLogger started
23/04/21 00:38:17 INFO Remoting: Starting remoting
23/04/21 00:38:17 INFO Remoting: Remoting started; listening on addresses
:[akka.tcp://spark@10.0.0.18:54583]
23/04/21 00:38:17 INFO Remoting: Remoting now listens on addresses:
[akka.tcp://spark@10.0.0.18:54583]
23/04/21 00:38:17 INFO spark.SparkEnv: Registering MapOutputTracker
23/04/21 00:38:17 INFO spark.SparkEnv: Registering BlockManagerMaster
23/04/21 00:38:17 INFO storage.DiskBlockManager: Created local directory at
/var/folders/1x/z4spc31954n2gycqnlph1rf40000gn/T/spark-local-20230421003817-30f1
23/04/21 00:38:17 INFO storage.MemoryStore: MemoryStore started with capacity 2.1 GB.
23/04/21 00:38:17 INFO network.ConnectionManager: Bound socket to port 54584 with id =
ConnectionManagerId(10.0.0.18,54584)
23/04/21 00:38:17 INFO storage.BlockManagerMaster: Trying to register BlockManager
23/04/21 00:38:17 INFO storage.BlockManagerInfo: Registering block manager
10.0.0.18:54584 with 2.1 GB RAM
23/04/21 00:38:17 INFO storage.BlockManagerMaster: Registered BlockManager
23/04/21 00:38:17 INFO spark.HttpServer: Starting HTTP Server
23/04/21 00:38:17 INFO server.Server: jetty-8.1.14.v20131031
23/04/21 00:38:17 INFO server.AbstractConnector: Started SocketConnector@0.0.0.0:54585
23/04/21 00:38:17 INFO broadcast.HttpBroadcast: Broadcast server started at
http://10.0.0.18:54585
23/04/21 00:38:17 INFO spark.HttpFileServer: HTTP File server directory is
/var/folders/1x/z4spc31954n2gycqnlph1rf40000gn/T/spark-a5774bac-3a4e-47b9-b6a6-19466a5
d94d2
23/04/21 00:38:17 INFO spark.HttpServer: Starting HTTP Server
23/04/21 00:38:17 INFO server.Server: jetty-8.1.14.v20131031
23/04/21 00:38:17 INFO server.AbstractConnector: Started SocketConnector@0.0.0.0:54586
23/04/21 00:38:17 INFO server.Server: jetty-8.1.14.v20131031
```

23/04/21 00:38:17 INFO server.AbstractConnector: Started  
SelectChannelConnector@0.0.0.0:4040  
23/04/21 00:38:17 INFO ui.SparkUI: Started SparkUI at http://10.0.0.18:4040  
23/04/21 00:38:17 INFO storage.MemoryStore: ensureFreeSpace(32816) called with  
curMem=0, maxMem=2290719129  
23/04/21 00:38:17 INFO storage.MemoryStore: Block broadcast\_0 stored as values to memory  
(estimated size 32.0 KB, free 2.1 GB)  
23/04/21 00:38:17 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your  
platform... using builtin-java classes where applicable  
23/04/21 00:38:17 WARN snappy.LoadSnappy: Snappy native library not loaded  
23/04/21 00:38:17 INFO mapred.FileInputFormat: Total input paths to process : 1  
23/04/21 00:38:17 INFO spark.SparkContext: Starting job: count at  
ShakespeareAnalytics.java:109  
23/04/21 00:38:17 INFO scheduler.DAGScheduler: Got job 0 (count at  
ShakespeareAnalytics.java:109) with 1 output partitions (allowLocal=false)  
23/04/21 00:38:17 INFO scheduler.DAGScheduler: Final stage: Stage 0(count at  
ShakespeareAnalytics.java:109)  
23/04/21 00:38:17 INFO scheduler.DAGScheduler: Parents of final stage: List()  
23/04/21 00:38:17 INFO scheduler.DAGScheduler: Missing parents: List()  
23/04/21 00:38:17 INFO scheduler.DAGScheduler: Submitting Stage 0 (FlatMappedRDD[2] at  
flatMap at ShakespeareAnalytics.java:107), which has no missing parents  
23/04/21 00:38:17 INFO scheduler.DAGScheduler: Submitting 1 missing tasks from Stage 0  
(FlatMappedRDD[2] at flatMap at ShakespeareAnalytics.java:107)  
23/04/21 00:38:17 INFO scheduler.TaskSchedulerImpl: Adding task set 0.0 with 1 tasks  
23/04/21 00:38:17 INFO scheduler.TaskSetManager: Starting task 0.0:0 as TID 0 on executor  
localhost: localhost (PROCESS\_LOCAL)  
23/04/21 00:38:17 INFO scheduler.TaskSetManager: Serialized task 0.0:0 as 2228 bytes in 0  
ms  
23/04/21 00:38:17 INFO executor.Executor: Running task ID 0  
23/04/21 00:38:17 INFO storage.BlockManager: Found block broadcast\_0 locally  
23/04/21 00:38:17 INFO rdd.HadoopRDD: Input split: file:/Users/praveenramesh/Distributed  
Systems/Project5Part2/AllsWellThatEndsWell:0+133886  
23/04/21 00:38:17 INFO executor.Executor: Serialized size of result for 0 is 597  
23/04/21 00:38:17 INFO executor.Executor: Sending result for 0 directly to driver  
23/04/21 00:38:17 INFO executor.Executor: Finished task ID 0  
23/04/21 00:38:17 INFO scheduler.TaskSetManager: Finished TID 0 in 59 ms on localhost  
(progress: 1/1)  
23/04/21 00:38:17 INFO scheduler.DAGScheduler: Completed ResultTask(0, 0)  
23/04/21 00:38:17 INFO scheduler.TaskSchedulerImpl: Removed TaskSet 0.0, whose tasks  
have all completed, from pool  
23/04/21 00:38:17 INFO scheduler.DAGScheduler: Stage 0 (count at  
ShakespeareAnalytics.java:109) finished in 0.065 s  
23/04/21 00:38:17 INFO spark.SparkContext: Job finished: count at  
ShakespeareAnalytics.java:109, took 0.105579875 s

### **Number of symbols: 130373**

```
23/04/21 00:38:18 INFO handler.ContextHandler: stopped
o.e.j.s.ServletContextHandler{/metrics/json,null}
23/04/21 00:38:18 INFO handler.ContextHandler: stopped
o.e.j.s.ServletContextHandler{/stages/stage/kill,null}
23/04/21 00:38:18 INFO handler.ContextHandler: stopped o.e.j.s.ServletContextHandler{/,null}
23/04/21 00:38:18 INFO handler.ContextHandler: stopped
o.e.j.s.ServletContextHandler{/static,null}
23/04/21 00:38:18 INFO handler.ContextHandler: stopped
o.e.j.s.ServletContextHandler{/executors/json,null}
23/04/21 00:38:18 INFO handler.ContextHandler: stopped
o.e.j.s.ServletContextHandler{/executors,null}
23/04/21 00:38:18 INFO handler.ContextHandler: stopped
o.e.j.s.ServletContextHandler{/environment/json,null}
23/04/21 00:38:18 INFO handler.ContextHandler: stopped
o.e.j.s.ServletContextHandler{/environment,null}
23/04/21 00:38:18 INFO handler.ContextHandler: stopped
o.e.j.s.ServletContextHandler{/storage/rdd/json,null}
23/04/21 00:38:18 INFO handler.ContextHandler: stopped
o.e.j.s.ServletContextHandler{/storage/rdd,null}
23/04/21 00:38:18 INFO handler.ContextHandler: stopped
o.e.j.s.ServletContextHandler{/storage/json,null}
23/04/21 00:38:18 INFO handler.ContextHandler: stopped
o.e.j.s.ServletContextHandler{/storage,null}
23/04/21 00:38:18 INFO handler.ContextHandler: stopped
o.e.j.s.ServletContextHandler{/stages/pool/json,null}
23/04/21 00:38:18 INFO handler.ContextHandler: stopped
o.e.j.s.ServletContextHandler{/stages/pool,null}
23/04/21 00:38:18 INFO handler.ContextHandler: stopped
o.e.j.s.ServletContextHandler{/stages/stage/json,null}
23/04/21 00:38:18 INFO handler.ContextHandler: stopped
o.e.j.s.ServletContextHandler{/stages/stage,null}
23/04/21 00:38:18 INFO handler.ContextHandler: stopped
o.e.j.s.ServletContextHandler{/stages/json,null}
23/04/21 00:38:18 INFO handler.ContextHandler: stopped
o.e.j.s.ServletContextHandler{/stages,null}
23/04/21 00:38:18 INFO ui.SparkUI: Stopped Spark web UI at http://10.0.0.18:4040
23/04/21 00:38:18 INFO scheduler.DAGScheduler: Stopping DAGScheduler
23/04/21 00:38:19 INFO spark.MapOutputTrackerMasterActor: MapOutputTrackerActor
stopped!
23/04/21 00:38:19 INFO network.ConnectionManager: Selector thread was interrupted!
23/04/21 00:38:19 INFO network.ConnectionManager: ConnectionManager stopped
23/04/21 00:38:19 INFO storage.MemoryStore: MemoryStore cleared
23/04/21 00:38:19 INFO storage.BlockManager: BlockManager stopped
```

```
23/04/21 00:38:19 INFO storage.BlockManagerMasterActor: Stopping BlockManagerMaster
23/04/21 00:38:19 INFO storage.BlockManagerMaster: BlockManagerMaster stopped
23/04/21 00:38:19 INFO spark.SparkContext: Successfully stopped SparkContext
```

Process finished with exit code 0

## Task 4

### Output

```
23/04/21 00:47:09 INFO spark.SecurityManager: Changing view acls to: praveenramesh
23/04/21 00:47:09 INFO spark.SecurityManager: SecurityManager: authentication disabled; ui
acls disabled; users with view permissions: Set(praveenramesh)
23/04/21 00:47:10 INFO slf4j.Slf4jLogger: Slf4jLogger started
23/04/21 00:47:10 INFO Remoting: Starting remoting
23/04/21 00:47:10 INFO Remoting: Remoting started; listening on addresses
:[akka.tcp://spark@10.0.0.18:54643]
23/04/21 00:47:10 INFO Remoting: Remoting now listens on addresses:
[akka.tcp://spark@10.0.0.18:54643]
23/04/21 00:47:10 INFO spark.SparkEnv: Registering MapOutputTracker
23/04/21 00:47:10 INFO spark.SparkEnv: Registering BlockManagerMaster
23/04/21 00:47:10 INFO storage.DiskBlockManager: Created local directory at
/var/folders/1x/z4spc31954n2gycqnlph1rf40000gn/T/spark-local-20230421004710-40ca
23/04/21 00:47:10 INFO storage.MemoryStore: MemoryStore started with capacity 2.1 GB.
23/04/21 00:47:10 INFO network.ConnectionManager: Bound socket to port 54644 with id =
ConnectionManagerId(10.0.0.18,54644)
23/04/21 00:47:10 INFO storage.BlockManagerMaster: Trying to register BlockManager
23/04/21 00:47:10 INFO storage.BlockManagerInfo: Registering block manager
10.0.0.18:54644 with 2.1 GB RAM
23/04/21 00:47:10 INFO storage.BlockManagerMaster: Registered BlockManager
23/04/21 00:47:10 INFO spark.HttpServer: Starting HTTP Server
23/04/21 00:47:10 INFO server.Server: jetty-8.1.14.v20131031
23/04/21 00:47:10 INFO server.AbstractConnector: Started SocketConnector@0.0.0.0:54645
23/04/21 00:47:10 INFO broadcast.HttpBroadcast: Broadcast server started at
http://10.0.0.18:54645
23/04/21 00:47:10 INFO spark.HttpFileServer: HTTP File server directory is
/var/folders/1x/z4spc31954n2gycqnlph1rf40000gn/T/spark-e52a5b1e-4109-4b1a-b4b7-27f06f27
1b6d
23/04/21 00:47:10 INFO spark.HttpServer: Starting HTTP Server
23/04/21 00:47:10 INFO server.Server: jetty-8.1.14.v20131031
23/04/21 00:47:10 INFO server.AbstractConnector: Started SocketConnector@0.0.0.0:54646
23/04/21 00:47:10 INFO server.Server: jetty-8.1.14.v20131031
```

23/04/21 00:47:10 INFO server.AbstractConnector: Started  
SelectChannelConnector@0.0.0.0:4040  
23/04/21 00:47:10 INFO ui.SparkUI: Started SparkUI at http://10.0.0.18:4040  
23/04/21 00:47:10 INFO storage.MemoryStore: ensureFreeSpace(32816) called with  
curMem=0, maxMem=2290719129  
23/04/21 00:47:10 INFO storage.MemoryStore: Block broadcast\_0 stored as values to memory  
(estimated size 32.0 KB, free 2.1 GB)  
23/04/21 00:47:10 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your  
platform... using builtin-java classes where applicable  
23/04/21 00:47:10 WARN snappy.LoadSnappy: Snappy native library not loaded  
23/04/21 00:47:10 INFO mapred.FileInputFormat: Total input paths to process : 1  
23/04/21 00:47:10 INFO spark.SparkContext: Starting job: count at  
ShakespeareAnalytics.java:133  
23/04/21 00:47:10 INFO scheduler.DAGScheduler: Registering RDD 4 (distinct at  
ShakespeareAnalytics.java:131)  
23/04/21 00:47:10 INFO scheduler.DAGScheduler: Got job 0 (count at  
ShakespeareAnalytics.java:133) with 1 output partitions (allowLocal=false)  
23/04/21 00:47:10 INFO scheduler.DAGScheduler: Final stage: Stage 0(count at  
ShakespeareAnalytics.java:133)  
23/04/21 00:47:10 INFO scheduler.DAGScheduler: Parents of final stage: List(Stage 1)  
23/04/21 00:47:10 INFO scheduler.DAGScheduler: Missing parents: List(Stage 1)  
23/04/21 00:47:10 INFO scheduler.DAGScheduler: Submitting Stage 1 (MapPartitionsRDD[4] at  
distinct at ShakespeareAnalytics.java:131), which has no missing parents  
23/04/21 00:47:10 INFO scheduler.DAGScheduler: Submitting 1 missing tasks from Stage 1  
(MapPartitionsRDD[4] at distinct at ShakespeareAnalytics.java:131)  
23/04/21 00:47:10 INFO scheduler.TaskSchedulerImpl: Adding task set 1.0 with 1 tasks  
23/04/21 00:47:10 INFO scheduler.TaskSetManager: Starting task 1.0:0 as TID 0 on executor  
localhost: localhost (PROCESS\_LOCAL)  
23/04/21 00:47:10 INFO scheduler.TaskSetManager: Serialized task 1.0:0 as 2531 bytes in 1  
ms  
23/04/21 00:47:10 INFO executor.Executor: Running task ID 0  
23/04/21 00:47:10 INFO storage.BlockManager: Found block broadcast\_0 locally  
23/04/21 00:47:10 INFO rdd.HadoopRDD: Input split: file:/Users/praveenramesh/Distributed  
Systems/Project5Part2/AllsWellThatEndsWell:0+133886  
23/04/21 00:47:10 INFO executor.Executor: Serialized size of result for 0 is 779  
23/04/21 00:47:10 INFO executor.Executor: Sending result for 0 directly to driver  
23/04/21 00:47:10 INFO executor.Executor: Finished task ID 0  
23/04/21 00:47:10 INFO scheduler.TaskSetManager: Finished TID 0 in 119 ms on localhost  
(progress: 1/1)  
23/04/21 00:47:10 INFO scheduler.TaskSchedulerImpl: Removed TaskSet 1.0, whose tasks  
have all completed, from pool  
23/04/21 00:47:10 INFO scheduler.DAGScheduler: Completed ShuffleMapTask(1, 0)  
23/04/21 00:47:10 INFO scheduler.DAGScheduler: Stage 1 (distinct at  
ShakespeareAnalytics.java:131) finished in 0.123 s

23/04/21 00:47:10 INFO scheduler.DAGScheduler: looking for newly runnable stages  
23/04/21 00:47:10 INFO scheduler.DAGScheduler: running: Set()  
23/04/21 00:47:10 INFO scheduler.DAGScheduler: waiting: Set(Stage 0)  
23/04/21 00:47:10 INFO scheduler.DAGScheduler: failed: Set()  
23/04/21 00:47:10 INFO scheduler.DAGScheduler: Missing parents for Stage 0: List()  
23/04/21 00:47:10 INFO scheduler.DAGScheduler: Submitting Stage 0 (MappedRDD[7] at distinct at ShakespeareAnalytics.java:131), which is now runnable  
23/04/21 00:47:10 INFO scheduler.DAGScheduler: Submitting 1 missing tasks from Stage 0 (MappedRDD[7] at distinct at ShakespeareAnalytics.java:131)  
23/04/21 00:47:10 INFO scheduler.TaskSchedulerImpl: Adding task set 0.0 with 1 tasks  
23/04/21 00:47:10 INFO scheduler.TaskSetManager: Starting task 0.0:0 as TID 1 on executor localhost: localhost (PROCESS\_LOCAL)  
23/04/21 00:47:10 INFO scheduler.TaskSetManager: Serialized task 0.0:0 as 2383 bytes in 0 ms  
23/04/21 00:47:10 INFO executor.Executor: Running task ID 1  
23/04/21 00:47:10 INFO storage.BlockManager: Found block broadcast\_0 locally  
23/04/21 00:47:10 INFO storage.BlockFetcherIterator\$BasicBlockFetcherIterator: maxBytesInFlight: 50331648, targetRequestSize: 10066329  
23/04/21 00:47:10 INFO storage.BlockFetcherIterator\$BasicBlockFetcherIterator: Getting 1 non-empty blocks out of 1 blocks  
23/04/21 00:47:10 INFO storage.BlockFetcherIterator\$BasicBlockFetcherIterator: Started 0 remote fetches in 2 ms  
23/04/21 00:47:10 INFO executor.Executor: Serialized size of result for 1 is 863  
23/04/21 00:47:10 INFO executor.Executor: Sending result for 1 directly to driver  
23/04/21 00:47:10 INFO executor.Executor: Finished task ID 1  
23/04/21 00:47:10 INFO scheduler.TaskSetManager: Finished TID 1 in 14 ms on localhost (progress: 1/1)  
23/04/21 00:47:10 INFO scheduler.DAGScheduler: Completed ResultTask(0, 0)  
23/04/21 00:47:10 INFO scheduler.TaskSchedulerImpl: Removed TaskSet 0.0, whose tasks have all completed, from pool  
23/04/21 00:47:10 INFO scheduler.DAGScheduler: Stage 0 (count at ShakespeareAnalytics.java:133) finished in 0.014 s  
23/04/21 00:47:10 INFO spark.SparkContext: Job finished: count at ShakespeareAnalytics.java:133, took 0.194534042 s

### **Number of distinct symbols: 78**

23/04/21 00:47:10 INFO handler.ContextHandler: stopped o.e.j.s.ServletContextHandler{/metrics/json,null}  
23/04/21 00:47:10 INFO handler.ContextHandler: stopped o.e.j.s.ServletContextHandler{/stages/stage/kill,null}  
23/04/21 00:47:10 INFO handler.ContextHandler: stopped o.e.j.s.ServletContextHandler{/,null}  
23/04/21 00:47:10 INFO handler.ContextHandler: stopped o.e.j.s.ServletContextHandler{/static,null}  
23/04/21 00:47:10 INFO handler.ContextHandler: stopped o.e.j.s.ServletContextHandler{/executors/json,null}

```
23/04/21 00:47:10 INFO handler.ContextHandler: stopped
o.e.j.s.ServletContextHandler{/executors,null}
23/04/21 00:47:10 INFO handler.ContextHandler: stopped
o.e.j.s.ServletContextHandler{/environment/json,null}
23/04/21 00:47:10 INFO handler.ContextHandler: stopped
o.e.j.s.ServletContextHandler{/environment,null}
23/04/21 00:47:10 INFO handler.ContextHandler: stopped
o.e.j.s.ServletContextHandler{/storage/rdd/json,null}
23/04/21 00:47:10 INFO handler.ContextHandler: stopped
o.e.j.s.ServletContextHandler{/storage/rdd,null}
23/04/21 00:47:10 INFO handler.ContextHandler: stopped
o.e.j.s.ServletContextHandler{/storage/json,null}
23/04/21 00:47:10 INFO handler.ContextHandler: stopped
o.e.j.s.ServletContextHandler{/storage,null}
23/04/21 00:47:10 INFO handler.ContextHandler: stopped
o.e.j.s.ServletContextHandler{/stages/pool/json,null}
23/04/21 00:47:10 INFO handler.ContextHandler: stopped
o.e.j.s.ServletContextHandler{/stages/pool,null}
23/04/21 00:47:10 INFO handler.ContextHandler: stopped
o.e.j.s.ServletContextHandler{/stages/stage/json,null}
23/04/21 00:47:10 INFO handler.ContextHandler: stopped
o.e.j.s.ServletContextHandler{/stages/stage,null}
23/04/21 00:47:10 INFO handler.ContextHandler: stopped
o.e.j.s.ServletContextHandler{/stages/json,null}
23/04/21 00:47:10 INFO handler.ContextHandler: stopped
o.e.j.s.ServletContextHandler{/stages,null}
23/04/21 00:47:10 INFO ui.SparkUI: Stopped Spark web UI at http://10.0.0.18:4040
23/04/21 00:47:10 INFO scheduler.DAGScheduler: Stopping DAGScheduler
23/04/21 00:47:11 INFO spark.MapOutputTrackerMasterActor: MapOutputTrackerActor
stopped!
23/04/21 00:47:12 INFO network.ConnectionManager: Selector thread was interrupted!
23/04/21 00:47:12 INFO network.ConnectionManager: ConnectionManager stopped
23/04/21 00:47:12 INFO storage.MemoryStore: MemoryStore cleared
23/04/21 00:47:12 INFO storage.BlockManager: BlockManager stopped
23/04/21 00:47:12 INFO storage.BlockManagerMasterActor: Stopping BlockManagerMaster
23/04/21 00:47:12 INFO storage.BlockManagerMaster: BlockManagerMaster stopped
23/04/21 00:47:12 INFO spark.SparkContext: Successfully stopped SparkContext
```

Process finished with exit code 0

## Task 5

### Output

```
23/04/21 00:49:49 WARN util.Utils: Your hostname, Praveens-MacBook-Pro-177.local resolves  
to a loopback address: 127.0.0.1; using 172.31.78.32 instead (on interface utun5)  
23/04/21 00:49:49 WARN util.Utils: Set SPARK_LOCAL_IP if you need to bind to another  
address  
23/04/21 00:49:49 INFO spark.SecurityManager: Changing view acls to: praveenramesh  
23/04/21 00:49:49 INFO spark.SecurityManager: SecurityManager: authentication disabled; ui  
acls disabled; users with view permissions: Set(praveenramesh)  
23/04/21 00:49:49 INFO slf4j.Slf4jLogger: Slf4jLogger started  
23/04/21 00:49:49 INFO Remoting: Starting remoting  
23/04/21 00:49:49 INFO Remoting: Remoting started; listening on addresses  
:[akka.tcp://spark@172.31.78.32:54653]  
23/04/21 00:49:49 INFO Remoting: Remoting now listens on addresses:  
[akka.tcp://spark@172.31.78.32:54653]  
23/04/21 00:49:49 INFO spark.SparkEnv: Registering MapOutputTracker  
23/04/21 00:49:49 INFO spark.SparkEnv: Registering BlockManagerMaster  
23/04/21 00:49:49 INFO storage.DiskBlockManager: Created local directory at  
/var/folders/1x/z4spc31954n2gycqnlph1rf40000gn/T/spark-local-20230421004949-6b54  
23/04/21 00:49:49 INFO storage.MemoryStore: MemoryStore started with capacity 2.1 GB.  
23/04/21 00:49:49 INFO network.ConnectionManager: Bound socket to port 54654 with id =  
ConnectionManagerId(172.31.78.32,54654)  
23/04/21 00:49:49 INFO storage.BlockManagerMaster: Trying to register BlockManager  
23/04/21 00:49:49 INFO storage.BlockManagerInfo: Registering block manager  
172.31.78.32:54654 with 2.1 GB RAM  
23/04/21 00:49:49 INFO storage.BlockManagerMaster: Registered BlockManager  
23/04/21 00:49:49 INFO spark.HttpServer: Starting HTTP Server  
23/04/21 00:49:49 INFO server.Server: jetty-8.1.14.v20131031  
23/04/21 00:49:49 INFO server.AbstractConnector: Started SocketConnector@0.0.0.0:54655  
23/04/21 00:49:49 INFO broadcast.HttpBroadcast: Broadcast server started at  
http://172.31.78.32:54655  
23/04/21 00:49:49 INFO spark.HttpFileServer: HTTP File server directory is  
/var/folders/1x/z4spc31954n2gycqnlph1rf40000gn/T/spark-bdbc3b3f-1805-4a2c-a3dd-fad70a1e  
4da2  
23/04/21 00:49:49 INFO spark.HttpServer: Starting HTTP Server
```

23/04/21 00:49:49 INFO server.Server: jetty-8.1.14.v20131031  
23/04/21 00:49:49 INFO server.AbstractConnector: Started SocketConnector@0.0.0.0:54656  
23/04/21 00:49:50 INFO server.Server: jetty-8.1.14.v20131031  
23/04/21 00:49:50 INFO server.AbstractConnector: Started  
SelectChannelConnector@0.0.0.0:4040  
23/04/21 00:49:50 INFO ui.SparkUI: Started SparkUI at http://172.31.78.32:4040  
23/04/21 00:49:50 INFO storage.MemoryStore: ensureFreeSpace(32816) called with  
curMem=0, maxMem=2290719129  
23/04/21 00:49:50 INFO storage.MemoryStore: Block broadcast\_0 stored as values to memory  
(estimated size 32.0 KB, free 2.1 GB)  
23/04/21 00:49:50 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your  
platform... using builtin-java classes where applicable  
23/04/21 00:49:50 WARN snappy.LoadSnappy: Snappy native library not loaded  
23/04/21 00:49:50 INFO mapred.FileInputFormat: Total input paths to process : 1  
23/04/21 00:49:50 INFO spark.SparkContext: Starting job: count at  
ShakespeareAnalytics.java:158  
23/04/21 00:49:50 INFO scheduler.DAGScheduler: Registering RDD 5 (distinct at  
ShakespeareAnalytics.java:156)  
23/04/21 00:49:50 INFO scheduler.DAGScheduler: Got job 0 (count at  
ShakespeareAnalytics.java:158) with 1 output partitions (allowLocal=false)  
23/04/21 00:49:50 INFO scheduler.DAGScheduler: Final stage: Stage 0(count at  
ShakespeareAnalytics.java:158)  
23/04/21 00:49:50 INFO scheduler.DAGScheduler: Parents of final stage: List(Stage 1)  
23/04/21 00:49:50 INFO scheduler.DAGScheduler: Missing parents: List(Stage 1)  
23/04/21 00:49:50 INFO scheduler.DAGScheduler: Submitting Stage 1 (MapPartitionsRDD[5] at  
distinct at ShakespeareAnalytics.java:156), which has no missing parents  
23/04/21 00:49:50 INFO scheduler.DAGScheduler: Submitting 1 missing tasks from Stage 1  
(MapPartitionsRDD[5] at distinct at ShakespeareAnalytics.java:156)  
23/04/21 00:49:50 INFO scheduler.TaskSchedulerImpl: Adding task set 1.0 with 1 tasks  
23/04/21 00:49:50 INFO scheduler.TaskSetManager: Starting task 1.0:0 as TID 0 on executor  
localhost: localhost (PROCESS\_LOCAL)  
23/04/21 00:49:50 INFO scheduler.TaskSetManager: Serialized task 1.0:0 as 2647 bytes in 0  
ms  
23/04/21 00:49:50 INFO executor.Executor: Running task ID 0  
23/04/21 00:49:50 INFO storage.BlockManager: Found block broadcast\_0 locally  
23/04/21 00:49:50 INFO rdd.HadoopRDD: Input split: file:/Users/praveenramesh/Distributed  
Systems/Project5Part2/AllsWellThatEndsWell:0+133886  
23/04/21 00:49:50 INFO executor.Executor: Serialized size of result for 0 is 782  
23/04/21 00:49:50 INFO executor.Executor: Sending result for 0 directly to driver  
23/04/21 00:49:50 INFO executor.Executor: Finished task ID 0  
23/04/21 00:49:50 INFO scheduler.TaskSetManager: Finished TID 0 in 152 ms on localhost  
(progress: 1/1)  
23/04/21 00:49:50 INFO scheduler.TaskSchedulerImpl: Removed TaskSet 1.0, whose tasks  
have all completed, from pool

23/04/21 00:49:50 INFO scheduler.DAGScheduler: Completed ShuffleMapTask(1, 0)  
23/04/21 00:49:50 INFO scheduler.DAGScheduler: Stage 1 (distinct at  
ShakespeareAnalytics.java:156) finished in 0.158 s  
23/04/21 00:49:50 INFO scheduler.DAGScheduler: looking for newly runnable stages  
23/04/21 00:49:50 INFO scheduler.DAGScheduler: running: Set()  
23/04/21 00:49:50 INFO scheduler.DAGScheduler: waiting: Set(Stage 0)  
23/04/21 00:49:50 INFO scheduler.DAGScheduler: failed: Set()  
23/04/21 00:49:50 INFO scheduler.DAGScheduler: Missing parents for Stage 0: List()  
23/04/21 00:49:50 INFO scheduler.DAGScheduler: Submitting Stage 0 (MappedRDD[8] at  
distinct at ShakespeareAnalytics.java:156), which is now runnable  
23/04/21 00:49:50 INFO scheduler.DAGScheduler: Submitting 1 missing tasks from Stage 0  
(MappedRDD[8] at distinct at ShakespeareAnalytics.java:156)  
23/04/21 00:49:50 INFO scheduler.TaskSchedulerImpl: Adding task set 0.0 with 1 tasks  
23/04/21 00:49:50 INFO scheduler.TaskSetManager: Starting task 0.0:0 as TID 1 on executor  
localhost: localhost (PROCESS\_LOCAL)  
23/04/21 00:49:50 INFO scheduler.TaskSetManager: Serialized task 0.0:0 as 2489 bytes in 0  
ms  
23/04/21 00:49:50 INFO executor.Executor: Running task ID 1  
23/04/21 00:49:50 INFO storage.BlockManager: Found block broadcast\_0 locally  
23/04/21 00:49:50 INFO storage.BlockFetcherIterator\$BasicBlockFetcherIterator:  
maxBytesInFlight: 50331648, targetRequestSize: 10066329  
23/04/21 00:49:50 INFO storage.BlockFetcherIterator\$BasicBlockFetcherIterator: Getting 1  
non-empty blocks out of 1 blocks  
23/04/21 00:49:50 INFO storage.BlockFetcherIterator\$BasicBlockFetcherIterator: Started 0  
remote fetches in 2 ms  
23/04/21 00:49:50 INFO executor.Executor: Serialized size of result for 1 is 863  
23/04/21 00:49:50 INFO executor.Executor: Sending result for 1 directly to driver  
23/04/21 00:49:50 INFO executor.Executor: Finished task ID 1  
23/04/21 00:49:50 INFO scheduler.DAGScheduler: Completed ResultTask(0, 0)  
23/04/21 00:49:50 INFO scheduler.DAGScheduler: Stage 0 (count at  
ShakespeareAnalytics.java:158) finished in 0.016 s  
23/04/21 00:49:50 INFO scheduler.TaskSetManager: Finished TID 1 in 14 ms on localhost  
(progress: 1/1)  
23/04/21 00:49:50 INFO scheduler.TaskSchedulerImpl: Removed TaskSet 0.0, whose tasks  
have all completed, from pool  
23/04/21 00:49:50 INFO spark.SparkContext: Job finished: count at  
ShakespeareAnalytics.java:158, took 0.227180541 s

### **Number of distinct letters: 50**

23/04/21 00:49:50 INFO handler.ContextHandler: stopped  
o.e.j.s.ServletContextHandler{/metrics/json,null}  
23/04/21 00:49:50 INFO handler.ContextHandler: stopped  
o.e.j.s.ServletContextHandler{/stages/stage/kill,null}  
23/04/21 00:49:50 INFO handler.ContextHandler: stopped o.e.j.s.ServletContextHandler{/,null}

```
23/04/21 00:49:50 INFO handler.ContextHandler: stopped
o.e.j.s.ServletContextHandler{/static,null}
23/04/21 00:49:50 INFO handler.ContextHandler: stopped
o.e.j.s.ServletContextHandler{/executors/json,null}
23/04/21 00:49:50 INFO handler.ContextHandler: stopped
o.e.j.s.ServletContextHandler{/executors,null}
23/04/21 00:49:50 INFO handler.ContextHandler: stopped
o.e.j.s.ServletContextHandler{/environment/json,null}
23/04/21 00:49:50 INFO handler.ContextHandler: stopped
o.e.j.s.ServletContextHandler{/environment,null}
23/04/21 00:49:50 INFO handler.ContextHandler: stopped
o.e.j.s.ServletContextHandler{/storage/rdd/json,null}
23/04/21 00:49:50 INFO handler.ContextHandler: stopped
o.e.j.s.ServletContextHandler{/storage/rdd,null}
23/04/21 00:49:50 INFO handler.ContextHandler: stopped
o.e.j.s.ServletContextHandler{/storage/json,null}
23/04/21 00:49:50 INFO handler.ContextHandler: stopped
o.e.j.s.ServletContextHandler{/storage,null}
23/04/21 00:49:50 INFO handler.ContextHandler: stopped
o.e.j.s.ServletContextHandler{/stages/pool/json,null}
23/04/21 00:49:50 INFO handler.ContextHandler: stopped
o.e.j.s.ServletContextHandler{/stages/pool,null}
23/04/21 00:49:50 INFO handler.ContextHandler: stopped
o.e.j.s.ServletContextHandler{/stages/stage/json,null}
23/04/21 00:49:50 INFO handler.ContextHandler: stopped
o.e.j.s.ServletContextHandler{/stages/stage,null}
23/04/21 00:49:50 INFO handler.ContextHandler: stopped
o.e.j.s.ServletContextHandler{/stages/json,null}
23/04/21 00:49:50 INFO handler.ContextHandler: stopped
o.e.j.s.ServletContextHandler{/stages,null}
23/04/21 00:49:50 INFO ui.SparkUI: Stopped Spark web UI at http://172.31.78.32:4040
23/04/21 00:49:50 INFO scheduler.DAGScheduler: Stopping DAGScheduler
23/04/21 00:49:51 INFO spark.MapOutputTrackerMasterActor: MapOutputTrackerActor
stopped!
23/04/21 00:49:51 INFO network.ConnectionManager: Selector thread was interrupted!
23/04/21 00:49:51 INFO network.ConnectionManager: ConnectionManager stopped
23/04/21 00:49:51 INFO storage.MemoryStore: MemoryStore cleared
23/04/21 00:49:51 INFO storage.BlockManager: BlockManager stopped
23/04/21 00:49:51 INFO storage.BlockManagerMasterActor: Stopping BlockManagerMaster
23/04/21 00:49:51 INFO storage.BlockManagerMaster: BlockManagerMaster stopped
23/04/21 00:49:51 INFO spark.SparkContext: Successfully stopped SparkContext
```

Process finished with exit code 0

## Task 6

### Output

Enter a word :

Love

```
23/04/21 00:51:59 INFO spark.SecurityManager: Changing view acls to: praveenramesh
23/04/21 00:51:59 INFO spark.SecurityManager: SecurityManager: authentication disabled; ui
acls disabled; users with view permissions: Set(praveenramesh)
23/04/21 00:51:59 INFO slf4j.Slf4jLogger: Slf4jLogger started
23/04/21 00:51:59 INFO Remoting: Starting remoting
23/04/21 00:51:59 INFO Remoting: Remoting started; listening on addresses
:[akka.tcp://spark@10.0.0.18:54668]
23/04/21 00:51:59 INFO Remoting: Remoting now listens on addresses:
[akka.tcp://spark@10.0.0.18:54668]
23/04/21 00:51:59 INFO spark.SparkEnv: Registering MapOutputTracker
23/04/21 00:51:59 INFO spark.SparkEnv: Registering BlockManagerMaster
23/04/21 00:51:59 INFO storage.DiskBlockManager: Created local directory at
/var/folders/1x/z4spc31954n2gycqnlph1rf40000gn/T/spark-local-20230421005159-dfe4
23/04/21 00:51:59 INFO storage.MemoryStore: MemoryStore started with capacity 2.1 GB.
23/04/21 00:51:59 INFO network.ConnectionManager: Bound socket to port 54669 with id =
ConnectionManagerId(10.0.0.18,54669)
23/04/21 00:51:59 INFO storage.BlockManagerMaster: Trying to register BlockManager
23/04/21 00:51:59 INFO storage.BlockManagerInfo: Registering block manager
10.0.0.18:54669 with 2.1 GB RAM
23/04/21 00:51:59 INFO storage.BlockManagerMaster: Registered BlockManager
23/04/21 00:51:59 INFO spark.HttpServer: Starting HTTP Server
23/04/21 00:51:59 INFO server.Server: jetty-8.1.14.v20131031
23/04/21 00:51:59 INFO server.AbstractConnector: Started SocketConnector@0.0.0.0:54670
23/04/21 00:51:59 INFO broadcast.HttpBroadcast: Broadcast server started at
http://10.0.0.18:54670
23/04/21 00:51:59 INFO spark.HttpFileServer: HTTP File server directory is
/var/folders/1x/z4spc31954n2gycqnlph1rf40000gn/T/spark-7fac7406-540e-4c9f-9d6f-6c4264c28
43a
23/04/21 00:51:59 INFO spark.HttpServer: Starting HTTP Server
23/04/21 00:51:59 INFO server.Server: jetty-8.1.14.v20131031
23/04/21 00:51:59 INFO server.AbstractConnector: Started SocketConnector@0.0.0.0:54671
```

23/04/21 00:51:59 INFO server.Server: jetty-8.1.14.v20131031  
23/04/21 00:51:59 INFO server.AbstractConnector: Started  
SelectChannelConnector@0.0.0.0:4040  
23/04/21 00:51:59 INFO ui.SparkUI: Started SparkUI at http://10.0.0.18:4040  
23/04/21 00:52:00 INFO storage.MemoryStore: ensureFreeSpace(32816) called with  
curMem=0, maxMem=2290719129  
23/04/21 00:52:00 INFO storage.MemoryStore: Block broadcast\_0 stored as values to memory  
(estimated size 32.0 KB, free 2.1 GB)  
23/04/21 00:52:00 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your  
platform... using builtin-java classes where applicable  
23/04/21 00:52:00 WARN snappy.LoadSnappy: Snappy native library not loaded  
23/04/21 00:52:00 INFO mapred.FileInputFormat: Total input paths to process : 1  
23/04/21 00:52:00 INFO spark.SparkContext: Starting job: collect at  
ShakespeareAnalytics.java:181  
23/04/21 00:52:00 INFO scheduler.DAGScheduler: Got job 0 (collect at  
ShakespeareAnalytics.java:181) with 1 output partitions (allowLocal=false)  
23/04/21 00:52:00 INFO scheduler.DAGScheduler: Final stage: Stage 0(collect at  
ShakespeareAnalytics.java:181)  
23/04/21 00:52:00 INFO scheduler.DAGScheduler: Parents of final stage: List()  
23/04/21 00:52:00 INFO scheduler.DAGScheduler: Missing parents: List()  
23/04/21 00:52:00 INFO scheduler.DAGScheduler: Submitting Stage 0 (FilteredRDD[2] at filter  
at ShakespeareAnalytics.java:179), which has no missing parents  
23/04/21 00:52:00 INFO scheduler.DAGScheduler: Submitting 1 missing tasks from Stage 0  
(FilteredRDD[2] at filter at ShakespeareAnalytics.java:179)  
23/04/21 00:52:00 INFO scheduler.TaskSchedulerImpl: Adding task set 0.0 with 1 tasks  
23/04/21 00:52:00 INFO scheduler.TaskSetManager: Starting task 0.0:0 as TID 0 on executor  
localhost: localhost (PROCESS\_LOCAL)  
23/04/21 00:52:00 INFO scheduler.TaskSetManager: Serialized task 0.0:0 as 2122 bytes in 0  
ms  
23/04/21 00:52:00 INFO executor.Executor: Running task ID 0  
23/04/21 00:52:00 INFO storage.BlockManager: Found block broadcast\_0 locally  
23/04/21 00:52:00 INFO rdd.HadoopRDD: Input split: file:/Users/praveenramesh/Distributed  
Systems/Project5Part2/AllsWellThatEndsWell:0+133886  
23/04/21 00:52:00 INFO executor.Executor: Serialized size of result for 0 is 957  
23/04/21 00:52:00 INFO executor.Executor: Sending result for 0 directly to driver  
23/04/21 00:52:00 INFO executor.Executor: Finished task ID 0  
23/04/21 00:52:00 INFO scheduler.TaskSetManager: Finished TID 0 in 46 ms on localhost  
(progress: 1/1)  
23/04/21 00:52:00 INFO scheduler.DAGScheduler: Completed ResultTask(0, 0)  
23/04/21 00:52:00 INFO scheduler.TaskSchedulerImpl: Removed TaskSet 0.0, whose tasks  
have all completed, from pool  
23/04/21 00:52:00 INFO scheduler.DAGScheduler: Stage 0 (collect at  
ShakespeareAnalytics.java:181) finished in 0.051 s

23/04/21 00:52:00 INFO spark.SparkContext: Job finished: collect at ShakespeareAnalytics.java:181, took 0.088110125 s  
23/04/21 00:52:00 INFO handler.ContextHandler: stopped o.e.j.s.ServletContextHandler{/metrics/json,null}  
23/04/21 00:52:00 INFO handler.ContextHandler: stopped o.e.j.s.ServletContextHandler{/stages/stage/kill,null}  
23/04/21 00:52:00 INFO handler.ContextHandler: stopped o.e.j.s.ServletContextHandler{/,null}  
23/04/21 00:52:00 INFO handler.ContextHandler: stopped o.e.j.s.ServletContextHandler{/static,null}  
23/04/21 00:52:00 INFO handler.ContextHandler: stopped o.e.j.s.ServletContextHandler{/executors/json,null}  
23/04/21 00:52:00 INFO handler.ContextHandler: stopped o.e.j.s.ServletContextHandler{/executors,null}  
23/04/21 00:52:00 INFO handler.ContextHandler: stopped o.e.j.s.ServletContextHandler{/environment/json,null}  
23/04/21 00:52:00 INFO handler.ContextHandler: stopped o.e.j.s.ServletContextHandler{/environment,null}  
23/04/21 00:52:00 INFO handler.ContextHandler: stopped o.e.j.s.ServletContextHandler{/storage/rdd/json,null}  
23/04/21 00:52:00 INFO handler.ContextHandler: stopped o.e.j.s.ServletContextHandler{/storage/rdd,null}  
23/04/21 00:52:00 INFO handler.ContextHandler: stopped o.e.j.s.ServletContextHandler{/storage/json,null}  
23/04/21 00:52:00 INFO handler.ContextHandler: stopped o.e.j.s.ServletContextHandler{/storage,null}  
23/04/21 00:52:00 INFO handler.ContextHandler: stopped o.e.j.s.ServletContextHandler{/stages/pool/json,null}  
23/04/21 00:52:00 INFO handler.ContextHandler: stopped o.e.j.s.ServletContextHandler{/stages/pool,null}  
23/04/21 00:52:00 INFO handler.ContextHandler: stopped o.e.j.s.ServletContextHandler{/stages/stage/json,null}  
23/04/21 00:52:00 INFO handler.ContextHandler: stopped o.e.j.s.ServletContextHandler{/stages/stage,null}  
23/04/21 00:52:00 INFO handler.ContextHandler: stopped o.e.j.s.ServletContextHandler{/stages/json,null}  
23/04/21 00:52:00 INFO handler.ContextHandler: stopped o.e.j.s.ServletContextHandler{/stages,null}  
23/04/21 00:52:00 INFO ui.SparkUI: Stopped Spark web UI at http://10.0.0.18:4040  
23/04/21 00:52:00 INFO scheduler.DAGScheduler: Stopping DAGScheduler  
23/04/21 00:52:01 INFO spark.MapOutputTrackerMasterActor: MapOutputTrackerActor stopped!  
23/04/21 00:52:01 INFO network.ConnectionManager: Selector thread was interrupted!  
23/04/21 00:52:01 INFO network.ConnectionManager: ConnectionManager stopped  
23/04/21 00:52:01 INFO storage.MemoryStore: MemoryStore cleared

23/04/21 00:52:01 INFO storage.BlockManager: BlockManager stopped  
23/04/21 00:52:01 INFO storage.BlockManagerMasterActor: Stopping BlockManagerMaster  
23/04/21 00:52:01 INFO storage.BlockManagerMaster: BlockManagerMaster stopped  
23/04/21 00:52:01 INFO spark.SparkContext: Successfully stopped SparkContext

**Lines containing the word Love:**

Share with thy birthright. Love all, trust a few,  
Love no god, that would not extend his might only  
Love you my son?  
Was both herself and Love, O then give pity  
Fall when Love please! Marry, to each but one.  
And to imperial Love, that god most high,  
Love make your fortunes twenty times above  
Which great Love grant, and so I take my leave.  
Be not so holy-cruel. Love is holy,

Process finished with exit code 0