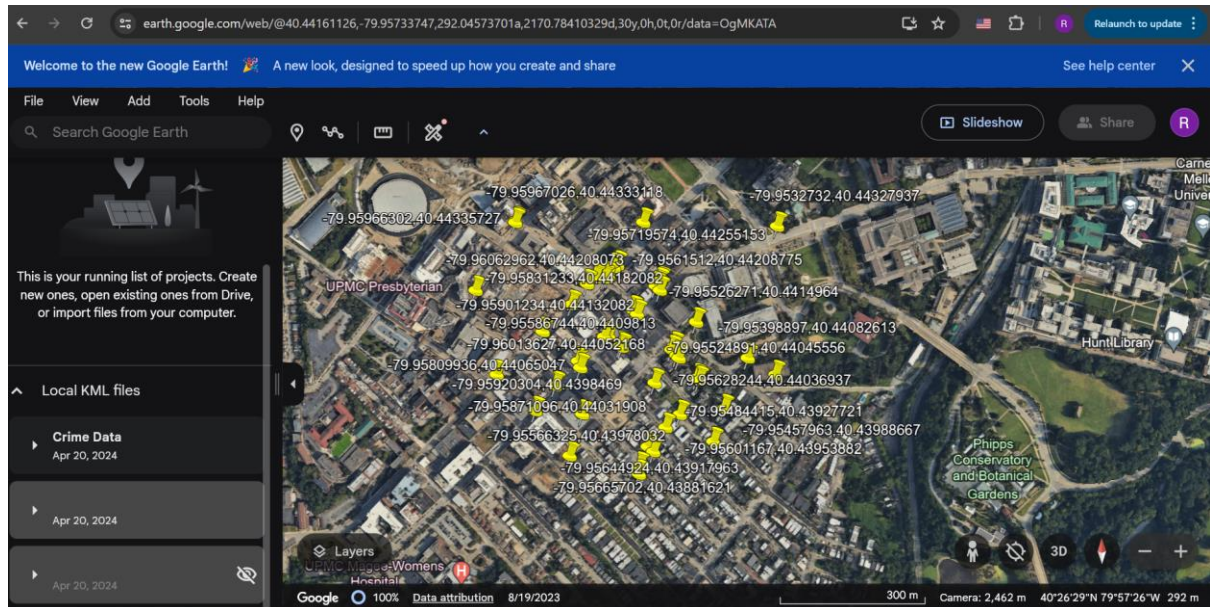


## Part 1

ClusterID: student137

Password: DistributedCMU1\$

Task 7 screenshot



## Part 2

Code:

```
package org.myorg;

import org.apache.spark.SparkConf;
import org.apache.spark.api.java.JavaRDD;
import org.apache.spark.api.java.JavaSparkContext;
import org.apache.spark.api.java.function.Function;

import java.io.BufferedReader;
import java.io.InputStreamReader;
import java.util.Arrays;
import java.util.List;

/**
 * ShakespeareAnalytics.java
 * Author: Rheann Sequeira
 * @author: rsequeir
 */
public class ShakespeareAnalytics {

    public static void main(String[] args) {
        // Check if input file is provided
        if (args.length == 0) {
            System.out.println("NO INPUT FILE");
        }
    }
}
```

```
        System.exit(1);
    }

    // Create a SparkConf object with local master and app name
    SparkConf conf = new
SparkConf().setMaster("local").setAppName("ShakespeareAnalytics");
    JavaSparkContext sc = new JavaSparkContext(conf);

    // Load the input file into a JavaRDD
    JavaRDD<String> lines = sc.textFile(args[0]);

    // Count the number of lines in the input file
    long noOfLines = lines.count();

    // Task 1: Count the number of words
    Function<String, Boolean> filter = k -> (!k.isEmpty());
    JavaRDD<String> words = lines.flatMap(line ->
Arrays.asList(line.split("[^a-zA-Z]+")));
    long noOfWords = words.filter(word -> !word.isEmpty()).count();

    // Task 2: Count the number of distinct words
    long noOfDistinctWords =
words.filter(filter).distinct().count();

    // Task 3: Count the number of symbols
    JavaRDD<String> symbols = lines.flatMap(line ->
Arrays.asList(line.split("")));
    long noOfSymbols = symbols.count();

    // Task 4: Count the number of distinct symbols
    long noOfDistinctSymbols = symbols.distinct().count();

    // Task 5: Count the number of distinct letters
    JavaRDD<String> letters = symbols.filter(s -> s.matches("[a-zA-
Z]"));
    long noOfDistinctLetters = letters.distinct().count();

    // Display the results
    System.out.println("Number of lines: " + noOfLines);
    System.out.println("Number of words: " + noOfWords);
    System.out.println("Number of distinct words in 'All's Well
That Ends Well': " + noOfDistinctWords);
    System.out.println("Number of symbols: " + noOfSymbols);
    System.out.println("Number of distinct symbols: " +
noOfDistinctSymbols);
    System.out.println("Number of distinct letters: " +
noOfDistinctLetters);

    // Task 6: Search for a word in the input file
    try (BufferedReader reader = new BufferedReader(new
InputStreamReader(System.in))) {
        System.out.print("Enter search word (This search is case-
sensitive): ");
        String input = reader.readLine();

        JavaRDD<String> linesFiltered = lines.filter(line ->
line.contains(input));
```

```

        List<String> result = linesFiltered.collect();
        if (result.isEmpty()) {
            System.out.println("No lines found containing the word
\"\" + input + "\".");
        } else {
            System.out.println("Lines containing the word \"\" +
input + "\".");
            result.forEach(System.out::println);
        }
    } catch (Exception e) {
        System.out.println("Error processing input: " +
e.getMessage());
        e.printStackTrace();
    }

    // Stop the Spark context
    sc.stop();
}
}

```

## Output:

```

Run  ShakespeareAnalytics x
C:\Program Files\Java\jdk1.8.0_291\bin\java.exe ...
24/04/20 22:37:06 INFO spark.SecurityManager: Changing view acls to: Rheann
24/04/20 22:37:06 INFO spark.SecurityManager: SecurityManager: authentication disabled; ui acls disabled; users with view permissions: Set(Rheann)
24/04/20 22:37:07 INFO slf4j.Slf4jLogger: Slf4jLogger started
24/04/20 22:37:07 INFO Remoting: Starting remoting
24/04/20 22:37:07 INFO Remoting: Remoting started; listening on addresses :[akka.tcp://spark@172.31.64.167:51241]
24/04/20 22:37:07 INFO Remoting: Remoting now listens on addresses: [akka.tcp://spark@172.31.64.167:51241]
24/04/20 22:37:07 INFO spark.SparkEnv: Registering MapOutputTracker
24/04/20 22:37:07 INFO spark.SparkEnv: Registering BlockManagerMaster
24/04/20 22:37:07 INFO storage.DiskBlockManager: Created local directory at C:\Users\Rheann\AppData\Local\Temp\spark-local-20240420223707-c36c
24/04/20 22:37:07 INFO storage.MemoryStore: MemoryStore started with capacity 1063.5 MB.
24/04/20 22:37:07 INFO network.ConnectionManager: Bound socket to port 51244 with id = ConnectionManagerId(172.31.64.167,51244)
24/04/20 22:37:07 INFO storage.BlockManagerMaster: Trying to register BlockManager
24/04/20 22:37:07 INFO storage.BlockManagerInfo: Registering block manager 172.31.64.167:51244 with 1063.5 MB RAM
24/04/20 22:37:07 INFO storage.BlockManagerMaster: Registered BlockManager
24/04/20 22:37:07 INFO spark.HttpServer: Starting HTTP Server
24/04/20 22:37:08 INFO server.Server: jetty-8.1.14.v20131031
24/04/20 22:37:08 INFO server.AbstractConnector: Started SocketConnector@0.0.0.0:51245
24/04/20 22:37:08 INFO broadcast.HttpBroadcast: Broadcast server started at http://172.31.64.167:51245
24/04/20 22:37:08 INFO spark.HttpFileServer: HTTP File server directory is C:\Users\Rheann\AppData\Local\Temp\spark-9858cc16-d37c-4c37-92bb-72b4ab634280
24/04/20 22:37:08 INFO spark.HttpServer: Starting HTTP Server
24/04/20 22:37:08 INFO server.Server: jetty-8.1.14.v20131031
24/04/20 22:37:08 INFO server.AbstractConnector: Started SocketConnector@0.0.0.0:51246
24/04/20 22:37:08 INFO server.Server: jetty-8.1.14.v20131031
24/04/20 22:37:08 INFO server.AbstractConnector: Started SelectChannelConnector@0.0.0.0:4040
24/04/20 22:37:08 INFO ui.SparkUI: Started SparkUI at http://172.31.64.167:4040
24/04/20 22:37:09 INFO storage.MemoryStore: ensureFreeSpace(57476) called with curMem=0, maxMem=1115160576
24/04/20 22:37:09 INFO storage.MemoryStore: Block broadcast_0 stored as values to memory (estimated size 56.1 KB, free 1063.4 MB)
24/04/20 22:37:09 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
24/04/20 22:37:09 WARN snappy.LoadSnappy: Snappy native library not loaded
24/04/20 22:37:09 INFO mapred.FileInputFormat: Total input paths to process : 1
24/04/20 22:37:09 INFO spark.SparkContext: Starting job: count at ShakespeareAnalytics.java:35
24/04/20 22:37:09 INFO scheduler.DAGScheduler: Got job 0 (count at ShakespeareAnalytics.java:35) with 1 output partitions (allowLocal=false)
24/04/20 22:37:09 INFO scheduler.DAGScheduler: Final stage: Stage 0(count at ShakespeareAnalytics.java:35)
24/04/20 22:37:09 INFO scheduler.DAGScheduler: Parents of final stage: List()
24/04/20 22:37:09 INFO scheduler.DAGScheduler: Missing parents: List()
24/04/20 22:37:09 INFO scheduler.DAGScheduler: Submitting Stage 0 (MappedRDD[1] at textFile at ShakespeareAnalytics.java:32), which has no missing parents
24/04/20 22:37:09 INFO scheduler.DAGScheduler: Submitting 1 missing tasks from Stage 0 (MappedRDD[1] at textFile at ShakespeareAnalytics.java:32)
24/04/20 22:37:09 INFO scheduler.TaskSchedulerImpl: Adding task set 0.0 with 1 tasks
24/04/20 22:37:09 INFO scheduler.TaskSetManager: Starting task 0.0:0 as TID 0 on executor localhost: localhost (PROCESS_LOCAL)
24/04/20 22:37:09 INFO scheduler.TaskSetManager: Serialized task 0.0:0 as 1771 bytes in 3 ms
24/04/20 22:37:09 INFO executor.Executor: Running task ID 0
24/04/20 22:37:09 INFO storage.BlockManager: Found block broadcast_0 locally

```

```
Run ShakespeareAnalytics x
24/04/20 22:37:09 INFO scheduler.TaskSetManager: Starting task 0.0:0 as TID 0 on executor localhost: localhost (PROCESS_LOCAL)
24/04/20 22:37:09 INFO scheduler.TaskSetManager: Serialized task 0.0:0 as 1771 bytes in 3 ms
24/04/20 22:37:09 INFO executor.Executor: Running task ID 0
24/04/20 22:37:09 INFO storage.BlockManager: Found block broadcast_0 locally
24/04/20 22:37:09 INFO rdd.HadoopRDD: Input split: file:/C:/Users/Rheann/Documents/CMU/Spring_24/Distributed_Systems/Project5/AllsWellThatEndsWell.txt:0+133886
24/04/20 22:37:09 INFO executor.Executor: Serialized size of result for 0 is 597
24/04/20 22:37:09 INFO executor.Executor: Sending result for 0 directly to driver
24/04/20 22:37:09 INFO executor.Executor: Finished task ID 0
24/04/20 22:37:09 INFO scheduler.TaskSetManager: Finished TID 0 in 165 ms on localhost (progress: 1/1)
24/04/20 22:37:09 INFO scheduler.TaskSchedulerImpl: Removed TaskSet 0.0, whose tasks have all completed, from pool
24/04/20 22:37:09 INFO scheduler.DAGScheduler: Completed ResultTask(0, 0)
24/04/20 22:37:09 INFO scheduler.DAGScheduler: Stage 0 (count at ShakespeareAnalytics.java:35) finished in 0.184 s
24/04/20 22:37:09 INFO spark.SparkContext: Job finished: count at ShakespeareAnalytics.java:35, took 0.2793563 s
24/04/20 22:37:09 INFO spark.SparkContext: Starting job: count at ShakespeareAnalytics.java:40
24/04/20 22:37:09 INFO scheduler.DAGScheduler: Got job 1 (count at ShakespeareAnalytics.java:40) with 1 output partitions (allowLocal=false)
24/04/20 22:37:09 INFO scheduler.DAGScheduler: Final stage: Stage 1(count at ShakespeareAnalytics.java:40)
24/04/20 22:37:09 INFO scheduler.DAGScheduler: Parents of final stage: List()
24/04/20 22:37:09 INFO scheduler.DAGScheduler: Missing parents: List()
24/04/20 22:37:09 INFO scheduler.DAGScheduler: Submitting Stage 1 (FilteredRDD[3] at filter at ShakespeareAnalytics.java:40), which has no missing parents
24/04/20 22:37:09 INFO scheduler.DAGScheduler: Submitting 1 missing tasks from Stage 1 (FilteredRDD[3] at filter at ShakespeareAnalytics.java:40)
24/04/20 22:37:09 INFO scheduler.TaskSchedulerImpl: Adding task set 1.0 with 1 tasks
24/04/20 22:37:09 INFO scheduler.TaskSetManager: Starting task 1.0:0 as TID 1 on executor localhost: localhost (PROCESS_LOCAL)
24/04/20 22:37:09 INFO scheduler.TaskSetManager: Serialized task 1.0:0 as 2370 bytes in 1 ms
24/04/20 22:37:09 INFO executor.Executor: Running task ID 1
24/04/20 22:37:09 INFO storage.BlockManager: Found block broadcast_0 locally
24/04/20 22:37:09 INFO rdd.HadoopRDD: Input split: file:/C:/Users/Rheann/Documents/CMU/Spring_24/Distributed_Systems/Project5/AllsWellThatEndsWell.txt:0+133886
24/04/20 22:37:09 INFO executor.Executor: Serialized size of result for 1 is 597
24/04/20 22:37:09 INFO executor.Executor: Sending result for 1 directly to driver
24/04/20 22:37:09 INFO executor.Executor: Finished task ID 1
24/04/20 22:37:09 INFO scheduler.DAGScheduler: Completed ResultTask(1, 0)
24/04/20 22:37:09 INFO scheduler.TaskSetManager: Finished TID 1 in 129 ms on localhost (progress: 1/1)
24/04/20 22:37:09 INFO scheduler.TaskSchedulerImpl: Removed TaskSet 1.0, whose tasks have all completed, from pool
24/04/20 22:37:09 INFO scheduler.DAGScheduler: Stage 1 (count at ShakespeareAnalytics.java:40) finished in 0.130 s
24/04/20 22:37:09 INFO spark.SparkContext: Job finished: count at ShakespeareAnalytics.java:40, took 0.1467592 s
24/04/20 22:37:09 INFO spark.SparkContext: Starting job: count at ShakespeareAnalytics.java:43
24/04/20 22:37:09 INFO scheduler.DAGScheduler: Registering RDD 6 (distinct at ShakespeareAnalytics.java:43)
24/04/20 22:37:09 INFO scheduler.DAGScheduler: Got job 2 (count at ShakespeareAnalytics.java:43) with 1 output partitions (allowLocal=false)
24/04/20 22:37:09 INFO scheduler.DAGScheduler: Final stage: Stage 2(count at ShakespeareAnalytics.java:43)
24/04/20 22:37:09 INFO scheduler.DAGScheduler: Parents of final stage: List(Stage 3)
24/04/20 22:37:09 INFO scheduler.DAGScheduler: Missing parents: List(Stage 3)
24/04/20 22:37:09 INFO scheduler.DAGScheduler: Submitting Stage 3 (MapPartitionsRDD[6] at distinct at ShakespeareAnalytics.java:43), which has no missing parents
24/04/20 22:37:09 INFO scheduler.DAGScheduler: Submitting 1 missing tasks from Stage 3 (MapPartitionsRDD[6] at distinct at ShakespeareAnalytics.java:43)
24/04/20 22:37:09 INFO scheduler.TaskSchedulerImpl: Adding task set 3.0 with 1 tasks
```

Run ShakespeareAnalytics x

More tool windows

```
24/04/20 22:37:09 INFO scheduler.DAGScheduler: Missing parents: List(Stage 3)
24/04/20 22:37:09 INFO scheduler.DAGScheduler: Submitting Stage 3 (MapPartitionsRDD[6] at distinct at ShakespeareAnalytics.java:43), which has no missing parents
24/04/20 22:37:09 INFO scheduler.DAGScheduler: Submitting 1 missing tasks from Stage 3 (MapPartitionsRDD[6] at distinct at ShakespeareAnalytics.java:43)
24/04/20 22:37:09 INFO scheduler.TaskSchedulerImpl: Adding task set 3.0 with 1 tasks
24/04/20 22:37:09 INFO scheduler.TaskSetManager: Starting task 3.0:0 as TID 2 on executor localhost: localhost (PROCESS_LOCAL)
24/04/20 22:37:09 INFO scheduler.TaskSetManager: Serialized task 3.0:0 as 2672 bytes in 0 ms
24/04/20 22:37:09 INFO executor.Executor: Running task ID 2
24/04/20 22:37:09 INFO storage.BlockManager: Found block broadcast_0 locally
24/04/20 22:37:09 INFO rdd.HadoopRDD: Input split: file:/C:/Users/Rheann/Documents/CMU/Spring_24/Distributed_Systems/Project5/AllsWellThatEndsWell.txt:0+133886
24/04/20 22:37:09 INFO executor.Executor: Serialized size of result for 2 is 783
24/04/20 22:37:09 INFO executor.Executor: Sending result for 2 directly to driver
24/04/20 22:37:09 INFO executor.Executor: Finished task ID 2
24/04/20 22:37:09 INFO scheduler.DAGScheduler: Completed ShuffleMapTask(3, 0)
24/04/20 22:37:09 INFO scheduler.TaskSetManager: Finished TID 2 in 299 ms on localhost (progress: 1/1)
24/04/20 22:37:09 INFO scheduler.DAGScheduler: Stage 3 (distinct at ShakespeareAnalytics.java:43) finished in 0.301 s
24/04/20 22:37:09 INFO scheduler.TaskSchedulerImpl: Removed TaskSet 3.0, whose tasks have all completed, from pool
24/04/20 22:37:09 INFO scheduler.DAGScheduler: looking for newly runnable stages
24/04/20 22:37:09 INFO scheduler.DAGScheduler: running: Set()
24/04/20 22:37:09 INFO scheduler.DAGScheduler: waiting: Set(Stage 2)
24/04/20 22:37:09 INFO scheduler.DAGScheduler: failed: Set()
24/04/20 22:37:09 INFO scheduler.DAGScheduler: Missing parents for Stage 2: List()
24/04/20 22:37:09 INFO scheduler.DAGScheduler: Submitting Stage 2 (MappedRDD[9] at distinct at ShakespeareAnalytics.java:43), which is now runnable
24/04/20 22:37:09 INFO scheduler.DAGScheduler: Submitting 1 missing tasks from Stage 2 (MappedRDD[9] at distinct at ShakespeareAnalytics.java:43)
24/04/20 22:37:09 INFO scheduler.TaskSchedulerImpl: Adding task set 2.0 with 1 tasks
24/04/20 22:37:09 INFO scheduler.TaskSetManager: Starting task 2.0:0 as TID 3 on executor localhost: localhost (PROCESS_LOCAL)
24/04/20 22:37:09 INFO scheduler.TaskSetManager: Serialized task 2.0:0 as 2491 bytes in 1 ms
24/04/20 22:37:09 INFO executor.Executor: Running task ID 3
24/04/20 22:37:10 INFO storage.BlockManager: Found block broadcast_0 locally
24/04/20 22:37:10 INFO storage.BlockFetcherIterator$BasicBlockFetcherIterator: maxBytesInFlight: 50331648, targetRequestSize: 10066329
24/04/20 22:37:10 INFO storage.BlockFetcherIterator$BasicBlockFetcherIterator: Getting 1 non-empty blocks out of 1 blocks
24/04/20 22:37:10 INFO storage.BlockFetcherIterator$BasicBlockFetcherIterator: Started 0 remote fetches in 7 ms
24/04/20 22:37:10 INFO executor.Executor: Serialized size of result for 3 is 863
24/04/20 22:37:10 INFO executor.Executor: Sending result for 3 directly to driver
24/04/20 22:37:10 INFO executor.Executor: Finished task ID 3
24/04/20 22:37:10 INFO scheduler.DAGScheduler: Completed ResultTask(2, 0)
24/04/20 22:37:10 INFO scheduler.TaskSetManager: Finished TID 3 in 91 ms on localhost (progress: 1/1)
24/04/20 22:37:10 INFO scheduler.DAGScheduler: Stage 2 (count at ShakespeareAnalytics.java:43) finished in 0.092 s
24/04/20 22:37:10 INFO scheduler.TaskSchedulerImpl: Removed TaskSet 2.0, whose tasks have all completed, from pool
24/04/20 22:37:10 INFO spark.SparkContext: Job finished: count at ShakespeareAnalytics.java:43, took 0.4411493 s
24/04/20 22:37:10 INFO spark.SparkContext: Starting job: count at ShakespeareAnalytics.java:47
24/04/20 22:37:10 INFO scheduler.DAGScheduler: Got job 3 (count at ShakespeareAnalytics.java:47) with 1 output partitions (allowLocal=false)
24/04/20 22:37:10 INFO scheduler.DAGScheduler: Final stage: Stage 4(count at ShakespeareAnalytics.java:47)
```



Run ShakespeareAnalytics x

↑

↓

≡

↶

↷

🖨

🗑

```
24/04/20 22:37:10 INFO spark.SparkContext: Starting job: count at ShakespeareAnalytics.java:47
24/04/20 22:37:10 INFO scheduler.DAGScheduler: Got job 3 (count at ShakespeareAnalytics.java:47) with 1 output partitions (allowLocal=false)
24/04/20 22:37:10 INFO scheduler.DAGScheduler: Final stage: Stage 4(count at ShakespeareAnalytics.java:47)
24/04/20 22:37:10 INFO scheduler.DAGScheduler: Parents of final stage: List()
24/04/20 22:37:10 INFO scheduler.DAGScheduler: Missing parents: List()
24/04/20 22:37:10 INFO scheduler.DAGScheduler: Submitting Stage 4 (FlatMappedRDD[10] at flatMap at ShakespeareAnalytics.java:46), which has no missing parents
24/04/20 22:37:10 INFO scheduler.DAGScheduler: Submitting 1 missing tasks from Stage 4 (FlatMappedRDD[10] at flatMap at ShakespeareAnalytics.java:46)
24/04/20 22:37:10 INFO scheduler.TaskSchedulerImpl: Adding task set 4.0 with 1 tasks
24/04/20 22:37:10 INFO scheduler.TaskSetManager: Starting task 4.0:0 as TID 4 on executor localhost: localhost (PROCESS_LOCAL)
24/04/20 22:37:10 INFO scheduler.TaskSetManager: Serialized task 4.0:0 as 2251 bytes in 1 ms
24/04/20 22:37:10 INFO executor.Executor: Running task ID 4
24/04/20 22:37:10 INFO storage.BlockManager: Found block broadcast_0 locally
24/04/20 22:37:10 INFO rdd.HadoopRDD: Input split: file:/C:/Users/Rheann/Documents/CMU/Spring 24/Distributed Systems/Project5/AllsWellThatEndsWell.txt:0+133886
24/04/20 22:37:10 INFO executor.Executor: Serialized size of result for 4 is 597
24/04/20 22:37:10 INFO executor.Executor: Sending result for 4 directly to driver
24/04/20 22:37:10 INFO executor.Executor: Finished task ID 4
24/04/20 22:37:10 INFO scheduler.DAGScheduler: Completed ResultTask(4, 0)
24/04/20 22:37:10 INFO scheduler.TaskSetManager: Finished TID 4 in 50 ms on localhost (progress: 1/1)
24/04/20 22:37:10 INFO scheduler.TaskSchedulerImpl: Removed TaskSet 4.0, whose tasks have all completed, from pool
24/04/20 22:37:10 INFO scheduler.DAGScheduler: Stage 4 (count at ShakespeareAnalytics.java:47) finished in 0.050 s
24/04/20 22:37:10 INFO spark.SparkContext: Job finished: count at ShakespeareAnalytics.java:47, took 0.057092599 s
24/04/20 22:37:10 INFO spark.SparkContext: Starting job: count at ShakespeareAnalytics.java:50
24/04/20 22:37:10 INFO scheduler.DAGScheduler: Registering RDD 12 (distinct at ShakespeareAnalytics.java:50)
24/04/20 22:37:10 INFO scheduler.DAGScheduler: Got job 4 (count at ShakespeareAnalytics.java:50) with 1 output partitions (allowLocal=false)
24/04/20 22:37:10 INFO scheduler.DAGScheduler: Final stage: Stage 5(count at ShakespeareAnalytics.java:50)
24/04/20 22:37:10 INFO scheduler.DAGScheduler: Parents of final stage: List(Stage 6)
24/04/20 22:37:10 INFO scheduler.DAGScheduler: Missing parents: List(Stage 6)
24/04/20 22:37:10 INFO scheduler.DAGScheduler: Submitting Stage 6 (MapPartitionsRDD[12] at distinct at ShakespeareAnalytics.java:50), which has no missing parents
24/04/20 22:37:10 INFO scheduler.DAGScheduler: Submitting 1 missing tasks from Stage 6 (MapPartitionsRDD[12] at distinct at ShakespeareAnalytics.java:50)
24/04/20 22:37:10 INFO scheduler.TaskSchedulerImpl: Adding task set 6.0 with 1 tasks
24/04/20 22:37:10 INFO scheduler.TaskSetManager: Starting task 6.0:0 as TID 5 on executor localhost: localhost (PROCESS_LOCAL)
24/04/20 22:37:10 INFO scheduler.TaskSetManager: Serialized task 6.0:0 as 2555 bytes in 0 ms
24/04/20 22:37:10 INFO executor.Executor: Running task ID 5
24/04/20 22:37:10 INFO storage.BlockManager: Found block broadcast_0 locally
24/04/20 22:37:10 INFO rdd.HadoopRDD: Input split: file:/C:/Users/Rheann/Documents/CMU/Spring 24/Distributed Systems/Project5/AllsWellThatEndsWell.txt:0+133886
24/04/20 22:37:10 INFO storage.ShuffleBlockManager: Deleted all files for shuffle 0
24/04/20 22:37:10 INFO spark.ContextCleaner: Cleaned shuffle 0
24/04/20 22:37:10 INFO executor.Executor: Serialized size of result for 5 is 783
24/04/20 22:37:10 INFO executor.Executor: Sending result for 5 directly to driver
24/04/20 22:37:10 INFO executor.Executor: Finished task ID 5
24/04/20 22:37:10 INFO scheduler.TaskSetManager: Finished TID 5 in 149 ms on localhost (progress: 1/1)
24/04/20 22:37:10 INFO scheduler.DAGScheduler: Completed ShuffleMapTask(6, 0)
24/04/20 22:37:10 INFO scheduler.TaskSchedulerImpl: Removed TaskSet 6.0, whose tasks have all completed, from pool
```

ject5 > ShakespeareAnalytics






```
Run ShakespeareAnalytics x
24/04/20 22:37:10 INFO scheduler.TaskSchedulerImpl: Removed TaskSet 6.0, whose tasks have all completed, from pool
24/04/20 22:37:10 INFO scheduler.DAGScheduler: Stage 6 (distinct at ShakespeareAnalytics.java:50) finished in 0.151 s
24/04/20 22:37:10 INFO scheduler.DAGScheduler: Looking for newly runnable stages
24/04/20 22:37:10 INFO scheduler.DAGScheduler: running: Set()
24/04/20 22:37:10 INFO scheduler.DAGScheduler: waiting: Set(Stage 5)
24/04/20 22:37:10 INFO scheduler.DAGScheduler: failed: Set()
24/04/20 22:37:10 INFO scheduler.DAGScheduler: Missing parents for Stage 5: List()
24/04/20 22:37:10 INFO scheduler.DAGScheduler: Submitting Stage 5 (MappedRDD[15] at distinct at ShakespeareAnalytics.java:50), which is now runnable
24/04/20 22:37:10 INFO scheduler.DAGScheduler: Submitting 1 missing tasks from Stage 5 (MappedRDD[15] at distinct at ShakespeareAnalytics.java:50)
24/04/20 22:37:10 INFO scheduler.TaskSchedulerImpl: Adding task set 5.0 with 1 tasks
24/04/20 22:37:10 INFO scheduler.TaskSetManager: Starting task 5.0:0 as TID 6 on executor localhost: localhost (PROCESS_LOCAL)
24/04/20 22:37:10 INFO scheduler.TaskSetManager: Serialized task 5.0:0 as 2385 bytes in 0 ms
24/04/20 22:37:10 INFO executor.Executor: Running task ID 6
24/04/20 22:37:10 INFO storage.BlockManager: Found block broadcast_0 locally
24/04/20 22:37:10 INFO storage.BlockFetcherIterator$BasicBlockFetcherIterator: maxBytesInFlight: 50331648, targetRequestSize: 10066329
24/04/20 22:37:10 INFO storage.BlockFetcherIterator$BasicBlockFetcherIterator: Getting 1 non-empty blocks out of 1 blocks
24/04/20 22:37:10 INFO storage.BlockFetcherIterator$BasicBlockFetcherIterator: Started 0 remote fetches in 0 ms
24/04/20 22:37:10 INFO executor.Executor: Serialized size of result for 6 is 863
24/04/20 22:37:10 INFO executor.Executor: Sending result for 6 directly to driver
24/04/20 22:37:10 INFO executor.Executor: Finished task ID 6
24/04/20 22:37:10 INFO scheduler.DAGScheduler: Completed ResultTask(5, 0)
24/04/20 22:37:10 INFO scheduler.TaskSetManager: Finished TID 6 in 12 ms on localhost (progress: 1/1)
24/04/20 22:37:10 INFO scheduler.TaskSchedulerImpl: Removed TaskSet 5.0, whose tasks have all completed, from pool
24/04/20 22:37:10 INFO scheduler.DAGScheduler: Stage 5 (count at ShakespeareAnalytics.java:50) finished in 0.013 s
24/04/20 22:37:10 INFO spark.SparkContext: Job finished: count at ShakespeareAnalytics.java:50, took 0.1759263 s
24/04/20 22:37:10 INFO spark.SparkContext: Starting job: count at ShakespeareAnalytics.java:54
24/04/20 22:37:10 INFO scheduler.DAGScheduler: Registering RDD 18 (distinct at ShakespeareAnalytics.java:54)
24/04/20 22:37:10 INFO scheduler.DAGScheduler: Got job 5 (count at ShakespeareAnalytics.java:54) with 1 output partitions (allowLocal=false)
24/04/20 22:37:10 INFO scheduler.DAGScheduler: Final stage: Stage 7(count at ShakespeareAnalytics.java:54)
24/04/20 22:37:10 INFO scheduler.DAGScheduler: Parents of final stage: List(Stage 8)
24/04/20 22:37:10 INFO scheduler.DAGScheduler: Missing parents: List(Stage 8)
24/04/20 22:37:10 INFO scheduler.DAGScheduler: Submitting Stage 8 (MapPartitionsRDD[18] at distinct at ShakespeareAnalytics.java:54), which has no missing parents
24/04/20 22:37:10 INFO scheduler.DAGScheduler: Submitting 1 missing tasks from Stage 8 (MapPartitionsRDD[18] at distinct at ShakespeareAnalytics.java:54)
24/04/20 22:37:10 INFO scheduler.TaskSchedulerImpl: Adding task set 8.0 with 1 tasks
24/04/20 22:37:10 INFO scheduler.TaskSetManager: Starting task 8.0:0 as TID 7 on executor localhost: localhost (PROCESS_LOCAL)
24/04/20 22:37:10 INFO scheduler.TaskSetManager: Serialized task 8.0:0 as 2673 bytes in 0 ms
24/04/20 22:37:10 INFO executor.Executor: Running task ID 7
24/04/20 22:37:10 INFO storage.BlockManager: Found block broadcast_0 locally
24/04/20 22:37:10 INFO rdd.HadoopRDD: Input split: file:/C:/Users/Rheann/Documents/CMU/Spring_24/Distributed_Systems/Project5/AllsWellThatEndsWell.txt:0+133886
24/04/20 22:37:10 INFO spark.ContextCleaner: Cleaned shuffle 1
24/04/20 22:37:10 INFO storage.ShuffleBlockManager: Deleted all files for shuffle 1
24/04/20 22:37:10 INFO executor.Executor: Serialized size of result for 7 is 783
24/04/20 22:37:10 INFO executor.Executor: Serializing result for 7 directly to driver...
```

```
24/04/20 22:37:10 INFO spark.ContextCleaner: Cleaned shuffle 1
24/04/20 22:37:10 INFO storage.ShuffleBlockManager: Deleted all files for shuffle 1
24/04/20 22:37:10 INFO executor.Executor: Serialized size of result for 7 is 783
24/04/20 22:37:10 INFO executor.Executor: Sending result for 7 directly to driver
24/04/20 22:37:10 INFO executor.Executor: Finished task ID 7
24/04/20 22:37:10 INFO scheduler.TaskSetManager: Finished TID 7 in 129 ms on localhost (progress: 1/1)
24/04/20 22:37:10 INFO scheduler.DAGScheduler: Completed ShuffleMapTask(8, 0)
24/04/20 22:37:10 INFO scheduler.TaskSchedulerImpl: Removed TaskSet 8.0, whose tasks have all completed, from pool
24/04/20 22:37:10 INFO scheduler.DAGScheduler: Stage 8 (distinct at ShakespeareAnalytics.java:54) finished in 0.129 s
24/04/20 22:37:10 INFO scheduler.DAGScheduler: looking for newly runnable stages
24/04/20 22:37:10 INFO scheduler.DAGScheduler: running: Set()
24/04/20 22:37:10 INFO scheduler.DAGScheduler: waiting: Set(Stage 7)
24/04/20 22:37:10 INFO scheduler.DAGScheduler: failed: Set()
24/04/20 22:37:10 INFO scheduler.DAGScheduler: Missing parents for Stage 7: List()
24/04/20 22:37:10 INFO scheduler.DAGScheduler: Submitting Stage 7 (MappedRDD[21] at distinct at ShakespeareAnalytics.java:54), which is now runnable
24/04/20 22:37:10 INFO scheduler.DAGScheduler: Submitting 1 missing tasks from Stage 7 (MappedRDD[21] at distinct at ShakespeareAnalytics.java:54)
24/04/20 22:37:10 INFO scheduler.TaskSchedulerImpl: Adding task set 7.0 with 1 tasks
24/04/20 22:37:10 INFO scheduler.TaskSetManager: Starting task 7.0:0 as TID 8 on executor localhost: localhost (PROCESS_LOCAL)
24/04/20 22:37:10 INFO scheduler.TaskSetManager: Serialized task 7.0:0 as 2495 bytes in 0 ms
24/04/20 22:37:10 INFO executor.Executor: Running task ID 8
24/04/20 22:37:10 INFO storage.BlockManager: Found block broadcast_0 locally
24/04/20 22:37:10 INFO storage.BlockFetcherIterator$BasicBlockFetcherIterator: maxBytesInFlight: 50331648, targetRequestSize: 10066329
24/04/20 22:37:10 INFO storage.BlockFetcherIterator$BasicBlockFetcherIterator: Getting 1 non-empty blocks out of 1 blocks
24/04/20 22:37:10 INFO storage.BlockFetcherIterator$BasicBlockFetcherIterator: Started 0 remote fetches in 0 ms
24/04/20 22:37:10 INFO executor.Executor: Serialized size of result for 8 is 863
24/04/20 22:37:10 INFO executor.Executor: Sending result for 8 directly to driver
24/04/20 22:37:10 INFO executor.Executor: Finished task ID 8
24/04/20 22:37:10 INFO scheduler.TaskSetManager: Finished TID 8 in 6 ms on localhost (progress: 1/1)
24/04/20 22:37:10 INFO scheduler.DAGScheduler: Completed ResultTask(7, 0)
24/04/20 22:37:10 INFO scheduler.TaskSchedulerImpl: Removed TaskSet 7.0, whose tasks have all completed, from pool
24/04/20 22:37:10 INFO scheduler.DAGScheduler: Stage 7 (count at ShakespeareAnalytics.java:54) finished in 0.006 s
24/04/20 22:37:10 INFO spark.SparkContext: Job finished: count at ShakespeareAnalytics.java:54, took 0.147567799 s
Number of lines: 4570
Number of words: 25085
Number of distinct words in 'All's Well That Ends Well': 3866
Number of symbols: 130373
Number of distinct symbols: 78
Number of distinct letters: 50
Enter search word (This search is case-sensitive): love
Lines containing the word "love":
24/04/20 22:37:12 INFO spark.SparkContext: Starting job: collect at ShakespeareAnalytics.java:71
24/04/20 22:37:12 INFO scheduler.DAGScheduler: Got job 6 (collect at ShakespeareAnalytics.java:71) with 1 output partitions (allowLocal=false)
24/04/20 22:37:12 INFO scheduler.DAGScheduler: Final stage: Stage 9[collect_at.ShakespeareAnalytics.java:71]
```



```
Run ShakespeareAnalytics x
Lines containing the word "love":
24/04/20 22:37:12 INFO spark.SparkContext: Starting job: collect at ShakespeareAnalytics.java:71
24/04/20 22:37:12 INFO scheduler.DAGScheduler: Got job 6 (collect at ShakespeareAnalytics.java:71) with 1 output partitions (allowLocal=false)
24/04/20 22:37:12 INFO scheduler.DAGScheduler: Final stage: Stage 9(collect at ShakespeareAnalytics.java:71)
24/04/20 22:37:12 INFO scheduler.DAGScheduler: Parents of final stage: List()
24/04/20 22:37:12 INFO scheduler.DAGScheduler: Missing parents: List()
24/04/20 22:37:12 INFO scheduler.DAGScheduler: Submitting Stage 9 (FilteredRDD[22] at filter at ShakespeareAnalytics.java:69), which has no missing parents
24/04/20 22:37:12 INFO scheduler.DAGScheduler: Submitting 1 missing tasks from Stage 9 (FilteredRDD[22] at filter at ShakespeareAnalytics.java:69)
24/04/20 22:37:12 INFO scheduler.TaskSchedulerImpl: Adding task set 9.0 with 1 tasks
24/04/20 22:37:12 INFO scheduler.TaskSetManager: Starting task 9.0:0 as TID 9 on executor localhost: localhost (PROCESS_LOCAL)
24/04/20 22:37:12 INFO scheduler.TaskSetManager: Serialized task 9.0:0 as 2145 bytes in 0 ms
24/04/20 22:37:12 INFO executor.Executor: Running task ID 9
24/04/20 22:37:12 INFO storage.BlockManager: Found block broadcast_0 locally
24/04/20 22:37:12 INFO rdd.HadoopRDD: Input split: file:/C:/Users/Rheann/Documents/CMU/Spring_24/Distributed_Systems/Project5/AllsWellThatEndsWell.txt:0+133886
24/04/20 22:37:12 INFO executor.Executor: Serialized size of result for 9 is 3801
24/04/20 22:37:12 INFO executor.Executor: Sending result for 9 directly to driver
24/04/20 22:37:12 INFO executor.Executor: Finished task ID 9
24/04/20 22:37:12 INFO scheduler.DAGScheduler: Completed ResultTask(9, 0)
24/04/20 22:37:12 INFO scheduler.TaskSetManager: Finished TID 9 in 12 ms on localhost (progress: 1/1)
24/04/20 22:37:12 INFO scheduler.DAGScheduler: Stage 9 (collect at ShakespeareAnalytics.java:71) finished in 0.012 s
24/04/20 22:37:12 INFO scheduler.TaskSchedulerImpl: Removed TaskSet 9.0, whose tasks have all completed, from pool
24/04/20 22:37:12 INFO spark.SparkContext: Job finished: collect at ShakespeareAnalytics.java:71, took 0.017815601 s
Attend his love.
That I should love a bright particular star
Th' ambition in my love thus plagues itself:
Must die for love. 'Twas pretty, though a plague,
One that goes with him. I love him for his sake,
self-love, which is the most inhibited sin in the
There shall your master have a thousand loves,
What power is it which mounts my love so high,
To show her merit that did miss her love?
FIRST LORD His love and wisdom,
SECOND LORD You're loved, sir.
and blood loves my flesh and blood; he that loves
STEWARD I know, madam, you love your gentlewoman
lawfully make title to as much love as she finds.
not any stranger sense. Her matter was she loved
Where love's strong passion is impressed in youth.
You love my son. Invention is ashamed
Do you love my son?
HELEN Do not you love him, madam?
Go not about. My love hath in 't a bond
I love your son.
```

Run ShakespeareAnalytics x



↑

↓

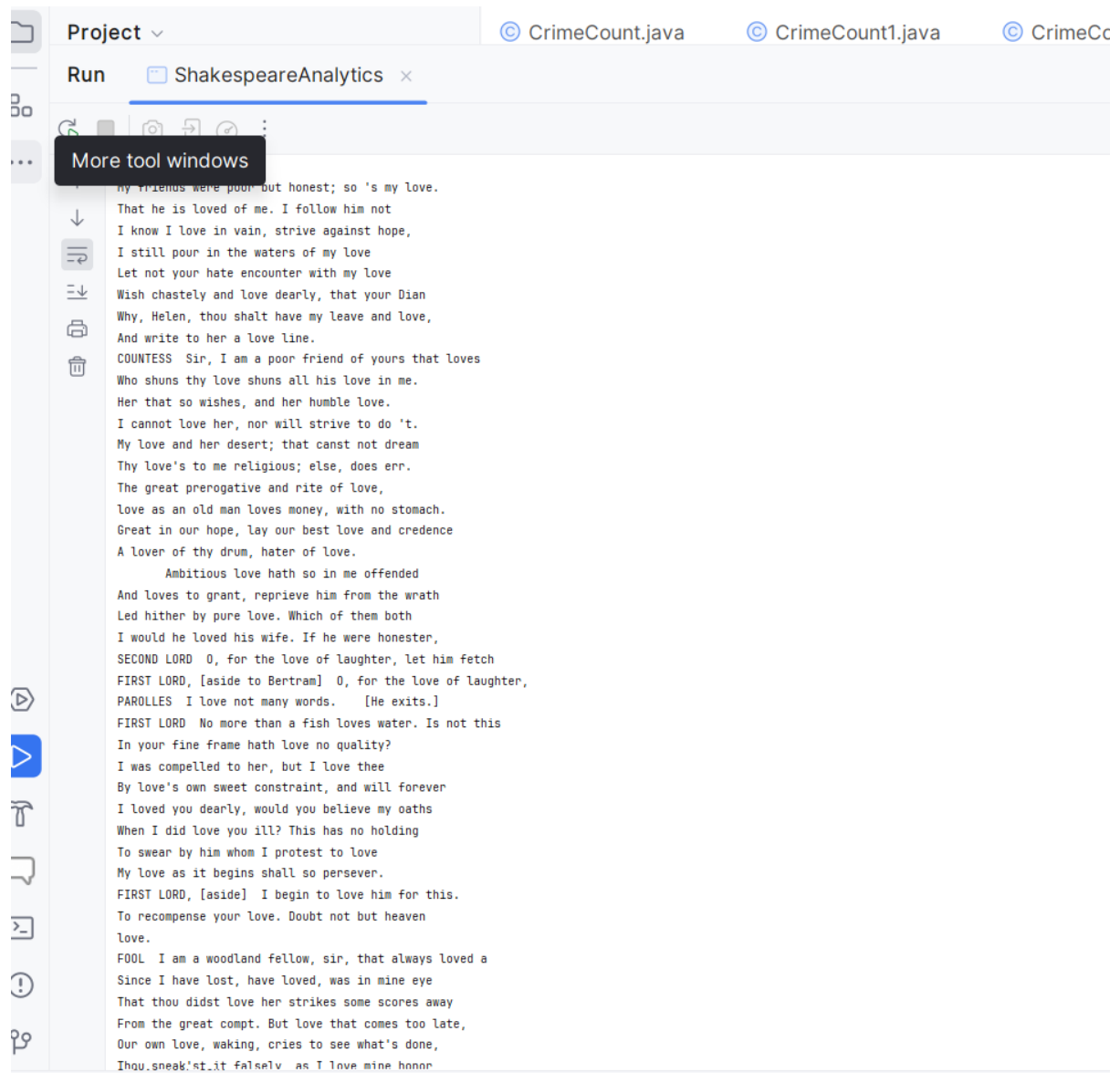
↶

↷

🖨

🗑

Go not about. My love hath in 't a bond  
I love your son.  
My friends were poor but honest; so 's my love.  
That he is loved of me. I follow him not  
I know I love in vain, strive against hope,  
I still pour in the waters of my love  
Let not your hate encounter with my love  
Wish chastely and love dearly, that your Dian  
Why, Helen, thou shalt have my leave and love,  
And write to her a love line.  
COUNTESS Sir, I am a poor friend of yours that loves  
Who shuns thy love shuns all his love in me.  
Her that so wishes, and her humble love.  
I cannot love her, nor will strive to do 't.  
My love and her desert; that canst not dream  
Thy love's to me religious; else, does err.  
The great prerogative and rite of love,  
Love as an old man loves money, with no stomach.  
Great in our hope, lay our best love and credence  
A lover of thy drum, hater of love.  
Ambitious love hath so in me offended  
And loves to grant, reprieve him from the wrath  
Led hither by pure love. Which of them both  
I would he loved his wife. If he were honester,  
SECOND LORD O, for the love of laughter, let him fetch  
FIRST LORD, [aside to Bertram] O, for the love of laughter,  
PAROLLES I love not many words. [He exits.]  
FIRST LORD No more than a fish loves water. Is not this  
In your fine frame hath love no quality?  
I was compelled to her, but I love thee  
By love's own sweet constraint, and will forever  
I loved you dearly, would you believe my oaths  
When I did love you ill? This has no holding  
To swear by him whom I protest to love  
My love as it begins shall so persever.  
FIRST LORD, [aside] I begin to love him for this.  
To recompense your love. Doubt not but heaven  
love.  
FOOL I am a woodland fellow, sir, that always loved a  
Since I have lost, have loved, was in mine eye  
That thou didst love her strikes some scores away  
From the great compt. But love that comes too late,  
Our own love, waking, cries to see what's done,



```
Run ShakespeareAnalytics x
Let not your hate encounter with my love
Wish chastely and love dearly, that your Dian
Why, Helen, thou shalt have my leave and love,
And write to her a love line.
COUNTESS Sir, I am a poor friend of yours that loves
Who shuns thy love shuns all his love in me.
Her that so wishes, and her humble love.
I cannot love her, nor will strive to do 't.
My love and her desert; that canst not dream
Thy love's to me religious; else, does err.
The great prerogative and rite of love,
Love as an old man loves money, with no stomach.
Great in our hope, lay our best love and credence
A lover of thy drum, hater of love.
    Ambitious love hath so in me offended
And loves to grant, reprieve him from the wrath
Led hither by pure love. Which of them both
I would he loved his wife. If he were honest,
SECOND LORD O, for the love of laughter, let him fetch
FIRST LORD, [aside to Bertram] O, for the love of laughter,
PAROLLES I love not many words. [He exits.]
FIRST LORD No more than a fish loves water. Is not this
In your fine frame hath love no quality?
I was compelled to her, but I love thee
By love's own sweet constraint, and will forever
I loved you dearly, would you believe my oaths
When I did love you ill? This has no holding
To swear by him whom I protest to love
My love as it begins shall so persever.
FIRST LORD, [aside] I begin to love him for this.
To recompense your love. Doubt not but heaven
love.
FOOL I am a woodland fellow, sir, that always loved a
Since I have lost, have loved, was in mine eye
That thou didst love her strikes some scores away
From the great compt. But love that comes too late,
Our own love, waking, cries to see what's done,
Thou speak'st it falsely, as I love mine honor,
KING Come, come, to th' purpose. Did he love this
PAROLLES Faith, sir, he did love her, but how?
PAROLLES He did love her, sir, as a gentleman loves a
PAROLLES He loved her, sir, and loved her not.
loved her. for indeed he was mad for her. and
```