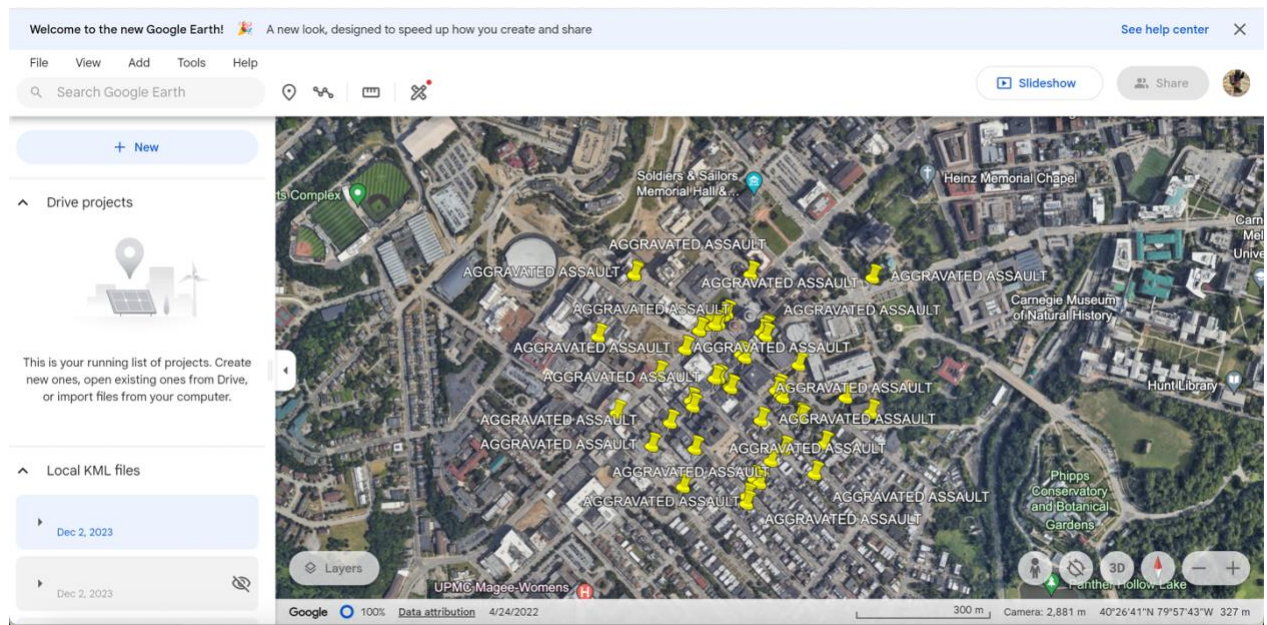


Title: Project5
Name: Aditi Gupta
Email: argupta@andrew.cmu.edu

Project5Part1

Task7 Output

Screen shot of Google Earth



Project5Part2

Complete Spark program

```
// Authored by: Aditi Gupta - argupta
// Taken reference from https://github.com/CMU-Heinz-95702/Project5
documentation
// Taken code from https://github.com/CMU-Heinz-95702/lab9-MapReduceAndSpark
Lab9
// Taken help from ChatGPT for generating Task1

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.util.Arrays;
import java.util.Scanner;

public class ShakespeareAnalytics {
    public static void main(String[] args) {
```

```

// Configure Spark
SparkConf sparkConf = new
SparkConf().setAppName("ShakespeareAnalytics").setMaster("local");
JavaSparkContext sparkContext = new JavaSparkContext(sparkConf);

// Load the text file
JavaRDD<String> lines =
sparkContext.textFile("AllsWellThatEndsWell.txt");

// Task 0: Count the number of lines
long numberOfLines = lines.count();
System.out.println("Task 0: Number of lines in 'All's Well That Ends
Well': " + numberOfLines);

// Task 1: Count the number of words
JavaRDD<String> words = lines.flatMap(line ->
Arrays.asList(line.split("[^a-zA-Z]+")))
    .filter((Function<String, Boolean>) k -> (!k.isEmpty()));
long numberOfWords = words.count();
System.out.println("Task 1: Number of words in 'All's Well That Ends
Well': " + numberOfWords);

//Task2: Count the number of distinct words
long numberOfDistinctWords = words.distinct().count();
System.out.println("Task 2: Number of distinct words in 'All's Well
That Ends Well': " + numberOfDistinctWords);

// Task 3: Count the number of symbols
long numberOfSymbols = lines.flatMap(line ->
Arrays.asList(line.split(""))).count();
System.out.println("Task 3: Number of symbols in 'All's Well That
Ends Well': " + numberOfSymbols);

// Task 4: Count the number of distinct symbols
long numberOfDistinctSymbols = lines.flatMap(line ->
Arrays.asList(line.split(""))).distinct().count();
System.out.println("Task 4: Number of distinct symbols in 'All's Well
That Ends Well': " + numberOfDistinctSymbols);

// Task 5: Count the number of distinct letters
long numberOfDistinctLetters = words.flatMap(word ->
Arrays.asList(word.split(""))).filter(letter -> letter.matches("[a-zA-
Z]")).distinct().count();
System.out.println("Task 5: Number of distinct letters in 'All's Well
That Ends Well': " + numberOfDistinctLetters);

// Task 6: Interactive search for lines containing a specific word
Scanner scanner = new Scanner(System.in);
System.out.print("Task 6: Enter a word to search for in 'All's Well
That Ends Well': ");
String searchWord = scanner.nextLine();

JavaRDD<String> linesContainingWord = lines.filter(line ->
line.contains(searchWord));
System.out.println("Lines containing '" + searchWord + "':");
for (String line : linesContainingWord.collect()) {

```

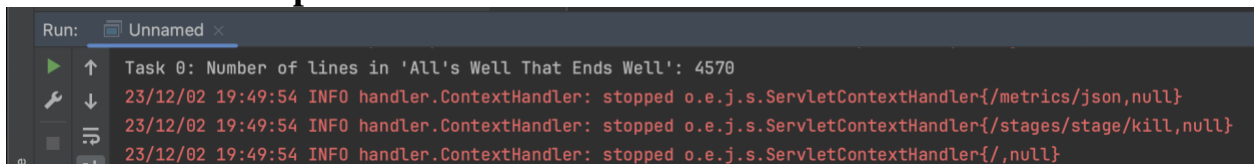
```

        System.out.println(line);
    }

    // Stop the SparkContext
    sparkContext.stop();
}
}

```

Screenshot of Output of Task0



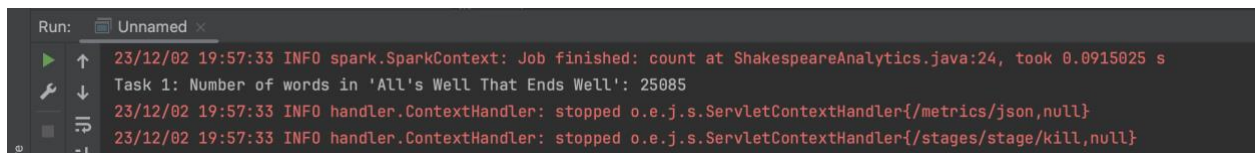
Run: Unnamed x

```

Task 0: Number of lines in 'All's Well That Ends Well': 4570
23/12/02 19:49:54 INFO handler.ContextHandler: stopped o.e.j.s.ServletContextHandler{/metrics/json,null}
23/12/02 19:49:54 INFO handler.ContextHandler: stopped o.e.j.s.ServletContextHandler{/stages/stage/kill,null}
23/12/02 19:49:54 INFO handler.ContextHandler: stopped o.e.j.s.ServletContextHandler{/,null}

```

Screenshot of Output of Task1



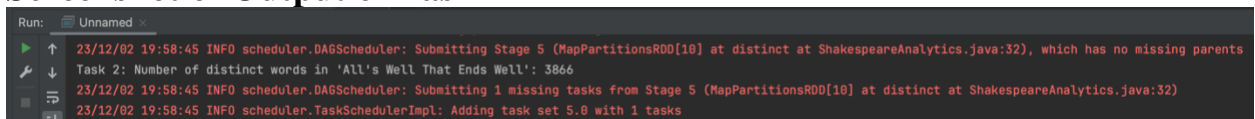
Run: Unnamed x

```

23/12/02 19:57:33 INFO spark.SparkContext: Job finished: count at ShakespeareAnalytics.java:24, took 0.0915025 s
Task 1: Number of words in 'All's Well That Ends Well': 25085
23/12/02 19:57:33 INFO handler.ContextHandler: stopped o.e.j.s.ServletContextHandler{/metrics/json,null}
23/12/02 19:57:33 INFO handler.ContextHandler: stopped o.e.j.s.ServletContextHandler{/stages/stage/kill,null}

```

Screenshot of Output of Task2



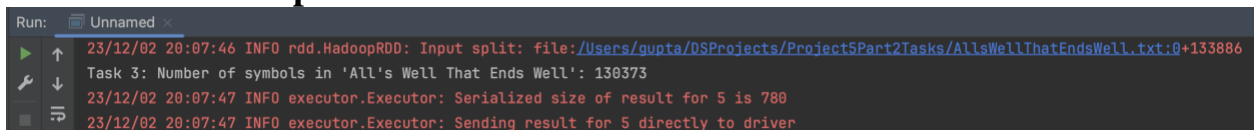
Run: Unnamed x

```

23/12/02 19:58:45 INFO scheduler.DAGScheduler: Submitting Stage 5 (MapPartitionsRDD[10] at distinct at ShakespeareAnalytics.java:32), which has no missing parents
Task 2: Number of distinct words in 'All's Well That Ends Well': 3866
23/12/02 19:58:45 INFO scheduler.DAGScheduler: Submitting 1 missing tasks from Stage 5 (MapPartitionsRDD[10] at distinct at ShakespeareAnalytics.java:32)
23/12/02 19:58:45 INFO scheduler.TaskSchedulerImpl: Adding task set 5.0 with 1 tasks

```

Screenshot of Output of Task3



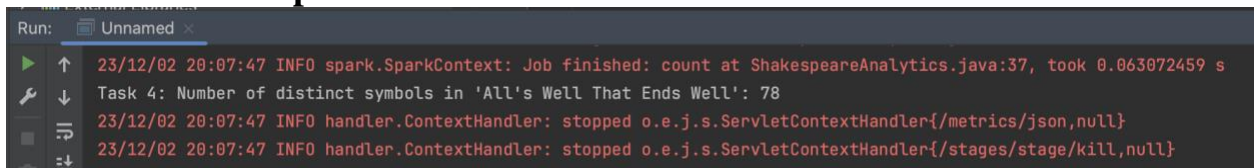
Run: Unnamed x

```

23/12/02 20:07:46 INFO rdd.HadoopRDD: Input split: file:/Users/gupta/DSProjects/Project5Part2Tasks/AllsWellThatEndsWell.txt:0+133886
Task 3: Number of symbols in 'All's Well That Ends Well': 130373
23/12/02 20:07:47 INFO executor.Executor: Serialized size of result for 5 is 780
23/12/02 20:07:47 INFO executor.Executor: Sending result for 5 directly to driver

```

Screenshot of Output of Task4




Run: Unnamed x

```

23/12/02 20:07:47 INFO spark.SparkContext: Job finished: count at ShakespeareAnalytics.java:37, took 0.063072459 s
Task 4: Number of distinct symbols in 'All's Well That Ends Well': 78
23/12/02 20:07:47 INFO handler.ContextHandler: stopped o.e.j.s.ServletContextHandler{/metrics/json,null}
23/12/02 20:07:47 INFO handler.ContextHandler: stopped o.e.j.s.ServletContextHandler{/stages/stage/kill,null}

```

Screenshot of Output of Task5



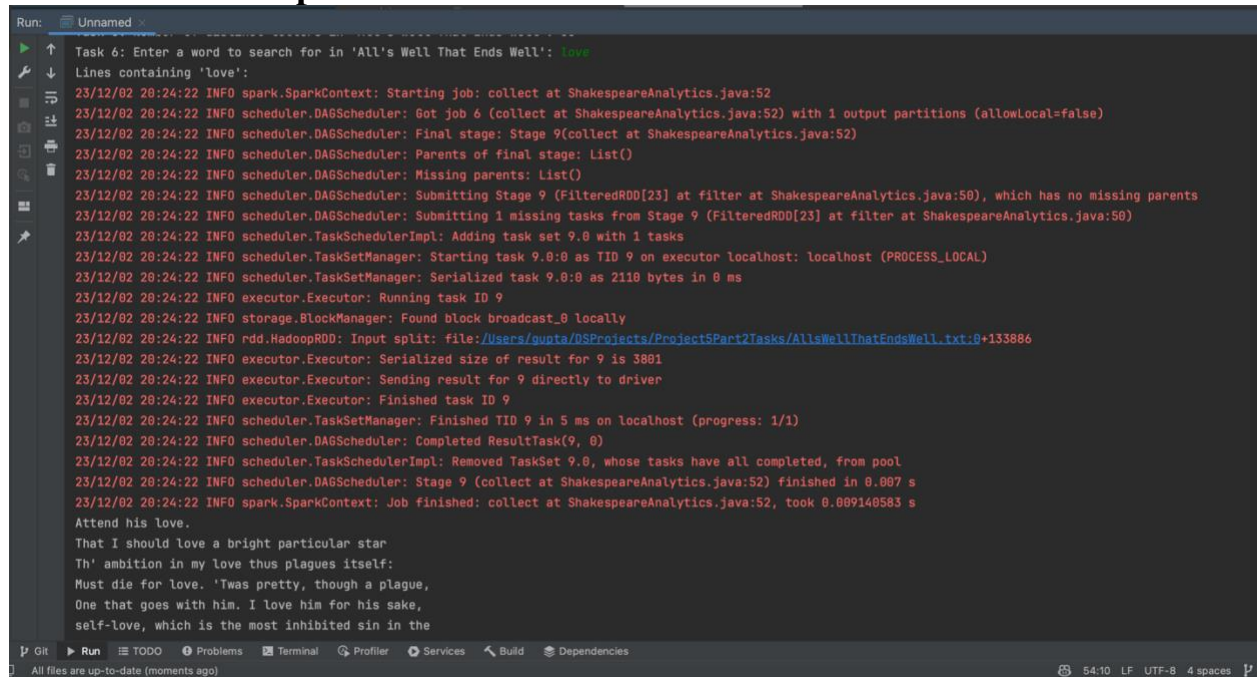
Run: Unnamed x

```

23/12/02 20:09:45 INFO spark.SparkContext: Job finished: count at ShakespeareAnalytics.java:41, took 0.13371825 s
Task 5: Number of distinct letters in 'All's Well That Ends Well': 50
23/12/02 20:09:45 INFO handler.ContextHandler: stopped o.e.j.s.ServletContextHandler{/metrics/json,null}
23/12/02 20:09:45 INFO handler.ContextHandler: stopped o.e.j.s.ServletContextHandler{/stages/stage/kill,null}

```

Screenshot of Output of Task6



```
Run: Unnamed
Task 6: Enter a word to search for in 'All's Well That Ends Well': love
Lines containing 'love':
23/12/02 20:24:22 INFO spark.SparkContext: Starting job: collect at ShakespeareAnalytics.java:52
23/12/02 20:24:22 INFO scheduler.DAGScheduler: Got job 6 (collect at ShakespeareAnalytics.java:52) with 1 output partitions (allowLocal=false)
23/12/02 20:24:22 INFO scheduler.DAGScheduler: Final stage: Stage 9(collect at ShakespeareAnalytics.java:52)
23/12/02 20:24:22 INFO scheduler.DAGScheduler: Parents of final stage: List()
23/12/02 20:24:22 INFO scheduler.DAGScheduler: Missing parents: List()
23/12/02 20:24:22 INFO scheduler.DAGScheduler: Submitting 1 missing tasks from Stage 9 (FilteredRDD[23] at filter at ShakespeareAnalytics.java:50), which has no missing parents
23/12/02 20:24:22 INFO scheduler.TaskSchedulerImpl: Adding task set 9.0 with 1 tasks
23/12/02 20:24:22 INFO scheduler.TaskSetManager: Starting task 9.0:0 as TID 9 on executor localhost (PROCESS_LOCAL)
23/12/02 20:24:22 INFO scheduler.TaskSetManager: Serialized task 9.0:0 as 2118 bytes in 8 ms
23/12/02 20:24:22 INFO executor.Executor: Running task ID 9
23/12/02 20:24:22 INFO storage.BlockManager: Found block broadcast_0 locally
23/12/02 20:24:22 INFO rdd.HadoopRDD: Input split: file:/Users/gupta/DSProjects/Project5Part2Tasks/AllsWellThatEndsWell.txt:0+133886
23/12/02 20:24:22 INFO executor.Executor: Serialized size of result for 9 is 3801
23/12/02 20:24:22 INFO executor.Executor: Sending result for 9 directly to driver
23/12/02 20:24:22 INFO executor.Executor: Finished task ID 9
23/12/02 20:24:22 INFO scheduler.TaskSetManager: Finished TID 9 in 5 ms on localhost (progress: 1/1)
23/12/02 20:24:22 INFO scheduler.DAGScheduler: Completed ResultTask(9, 0)
23/12/02 20:24:22 INFO scheduler.TaskSchedulerImpl: Removed TaskSet 9.0, whose tasks have all completed, from pool
23/12/02 20:24:22 INFO scheduler.DAGScheduler: Stage 9 (collect at ShakespeareAnalytics.java:52) finished in 0.007 s
23/12/02 20:24:22 INFO spark.SparkContext: Job finished: collect at ShakespeareAnalytics.java:52, took 0.009140583 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
```

Total Output:

```
/Users/gupta/Library/Java/JavaVirtualMachines/corretto-1.8.0_392/Contents/Home/bin/java -
javaagent:/Applications/IntelliJ IDEA.app/Contents/lib/idea_rt.jar=58240:/Applications/IntelliJ
IDEA.app/Contents/bin -Dfile.encoding=UTF-8 -classpath
/Users/gupta/Library/Java/JavaVirtualMachines/corretto-
1.8.0_392/Contents/Home/jre/lib/charsets.jar:/Users/gupta/Library/Java/JavaVirtualMachines/corretto
-
1.8.0_392/Contents/Home/jre/lib/ext/cldrdata.jar:/Users/gupta/Library/Java/JavaVirtualMachines/corr
etto-
1.8.0_392/Contents/Home/jre/lib/ext/dnsns.jar:/Users/gupta/Library/Java/JavaVirtualMachines/corrett
o-
1.8.0_392/Contents/Home/jre/lib/ext/jaccess.jar:/Users/gupta/Library/Java/JavaVirtualMachines/corre
tto-
1.8.0_392/Contents/Home/jre/lib/ext/localedata.jar:/Users/gupta/Library/Java/JavaVirtualMachines/co
rretto-
1.8.0_392/Contents/Home/jre/lib/ext/nashorn.jar:/Users/gupta/Library/Java/JavaVirtualMachines/corr
etto-
1.8.0_392/Contents/Home/jre/lib/ext/sunec.jar:/Users/gupta/Library/Java/JavaVirtualMachines/corrett
o-
1.8.0_392/Contents/Home/jre/lib/ext/sunjce_provider.jar:/Users/gupta/Library/Java/JavaVirtualMachi
nes/corretto-
1.8.0_392/Contents/Home/jre/lib/ext/sunpkcs11.jar:/Users/gupta/Library/Java/JavaVirtualMachines/c
orretto-
1.8.0_392/Contents/Home/jre/lib/ext/zipfs.jar:/Users/gupta/Library/Java/JavaVirtualMachines/corretto
-1.8.0_392/Contents/Home/jre/lib/jce.jar:/Users/gupta/Library/Java/JavaVirtualMachines/corretto-
1.8.0_392/Contents/Home/jre/lib/jfr.jar:/Users/gupta/Library/Java/JavaVirtualMachines/corretto-
1.8.0_392/Contents/Home/jre/lib/jsse.jar:/Users/gupta/Library/Java/JavaVirtualMachines/corretto-
1.8.0_392/Contents/Home/jre/lib/management-
```

agent.jar:/Users/gupta/Library/Java/JavaVirtualMachines/corretto-
1.8.0_392/Contents/Home/jre/lib/resources.jar:/Users/gupta/Library/Java/JavaVirtualMachines/corrett
o-
1.8.0_392/Contents/Home/jre/lib/rt.jar:/Users/gupta/DSProjects/Project5Part2Tasks/target/classes:/Us
ers/gupta/.m2/repository/org/apache/spark/spark-core_2.10/1.0.0/spark-core_2.10-
1.0.0.jar:/Users/gupta/.m2/repository/org/apache/hadoop/hadoop-client/1.0.4/hadoop-client-
1.0.4.jar:/Users/gupta/.m2/repository/org/apache/hadoop/hadoop-core/1.0.4/hadoop-core-
1.0.4.jar:/Users/gupta/.m2/repository/xmlenc/xmlenc/0.52/xmlenc-
0.52.jar:/Users/gupta/.m2/repository/org/apache/commons/commons-math/2.1/commons-math-
2.1.jar:/Users/gupta/.m2/repository/commons-configuration/commons-configuration/1.6/commons-
configuration-1.6.jar:/Users/gupta/.m2/repository/commons-collections/commons-
collections/3.2.1/commons-collections-3.2.1.jar:/Users/gupta/.m2/repository/commons-
lang/commons-lang/2.4/commons-lang-2.4.jar:/Users/gupta/.m2/repository/commons-
digester/commons-digester/1.8/commons-digester-1.8.jar:/Users/gupta/.m2/repository/commons-
beanutils/commons-beanutils/1.7.0/commons-beanutils-
1.7.0.jar:/Users/gupta/.m2/repository/commons-beanutils/commons-beanutils-core/1.8.0/commons-
beanutils-core-1.8.0.jar:/Users/gupta/.m2/repository/commons-el/commons-el/1.0/commons-el-
1.0.jar:/Users/gupta/.m2/repository/hsqldb/hsqldb/1.8.0.10/hsqldb-
1.8.0.10.jar:/Users/gupta/.m2/repository/oro/oro/2.0.8/oro-
2.0.8.jar:/Users/gupta/.m2/repository/org/codehaus/jackson/jackson-mapper-asl/1.0.1/jackson-mapper-
asl-1.0.1.jar:/Users/gupta/.m2/repository/org/codehaus/jackson/jackson-core-asl/1.0.1/jackson-core-
asl-1.0.1.jar:/Users/gupta/.m2/repository/net/java/dev/jets3t/jets3t/0.7.1/jets3t-
0.7.1.jar:/Users/gupta/.m2/repository/commons-codec/commons-codec/1.3/commons-codec-
1.3.jar:/Users/gupta/.m2/repository/commons-httpclient/commons-httpclient/3.1/commons-httpclient-
3.1.jar:/Users/gupta/.m2/repository/org/apache/curator/curator-recipes/2.4.0/curator-recipes-
2.4.0.jar:/Users/gupta/.m2/repository/org/apache/curator/curator-framework/2.4.0/curator-framework-
2.4.0.jar:/Users/gupta/.m2/repository/org/apache/curator/curator-client/2.4.0/curator-client-
2.4.0.jar:/Users/gupta/.m2/repository/org/apache/zookeeper/zookeeper/3.4.5/zookeeper-
3.4.5.jar:/Users/gupta/.m2/repository/jline/jline/0.9.94/jline-
0.9.94.jar:/Users/gupta/.m2/repository/org/eclipse/jetty/jetty-plus/8.1.14.v20131031/jetty-plus-
8.1.14.v20131031.jar:/Users/gupta/.m2/repository/org/eclipse/jetty/orbit/javax.transaction/1.1.1.v2011
05210645/javax.transaction-
1.1.1.v201105210645.jar:/Users/gupta/.m2/repository/org/eclipse/jetty/jetty-
webapp/8.1.14.v20131031/jetty-webapp-
8.1.14.v20131031.jar:/Users/gupta/.m2/repository/org/eclipse/jetty/jetty-xml/8.1.14.v20131031/jetty-
xml-8.1.14.v20131031.jar:/Users/gupta/.m2/repository/org/eclipse/jetty/jetty-
servlet/8.1.14.v20131031/jetty-servlet-
8.1.14.v20131031.jar:/Users/gupta/.m2/repository/org/eclipse/jetty/jetty-jndi/8.1.14.v20131031/jetty-
jndi-
8.1.14.v20131031.jar:/Users/gupta/.m2/repository/org/eclipse/jetty/orbit/javax.mail.glassfish/1.4.1.v2
01005082020/javax.mail.glassfish-
1.4.1.v201005082020.jar:/Users/gupta/.m2/repository/org/eclipse/jetty/orbit/javax.activation/1.1.0.v20
1105071233/javax.activation-
1.1.0.v201105071233.jar:/Users/gupta/.m2/repository/org/eclipse/jetty/jetty-
security/8.1.14.v20131031/jetty-security-
8.1.14.v20131031.jar:/Users/gupta/.m2/repository/org/eclipse/jetty/jetty-util/8.1.14.v20131031/jetty-
util-8.1.14.v20131031.jar:/Users/gupta/.m2/repository/org/eclipse/jetty/jetty-
server/8.1.14.v20131031/jetty-server-
8.1.14.v20131031.jar:/Users/gupta/.m2/repository/org/eclipse/jetty/orbit/javax.servlet/3.0.0.v2011120
11016/javax.servlet-3.0.0.v201112011016.jar:/Users/gupta/.m2/repository/org/eclipse/jetty/jetty-
continuation/8.1.14.v20131031/jetty-continuation-
8.1.14.v20131031.jar:/Users/gupta/.m2/repository/org/eclipse/jetty/jetty-http/8.1.14.v20131031/jetty-
http-8.1.14.v20131031.jar:/Users/gupta/.m2/repository/org/eclipse/jetty/jetty-

io/8.1.14.v20131031/jetty-io-
8.1.14.v20131031.jar:/Users/gupta/.m2/repository/com/google/guava/guava/14.0.1/guava-
14.0.1.jar:/Users/gupta/.m2/repository/org/apache/commons/commons-lang3/3.3.2/commons-lang3-
3.3.2.jar:/Users/gupta/.m2/repository/com/google/code/findbugs/jsr305/1.3.9/jsr305-
1.3.9.jar:/Users/gupta/.m2/repository/org/slf4j/slf4j-api/1.7.5/slf4j-api-
1.7.5.jar:/Users/gupta/.m2/repository/org/slf4j/jul-to-slf4j/1.7.5/jul-to-slf4j-
1.7.5.jar:/Users/gupta/.m2/repository/org/slf4j/jcl-over-slf4j/1.7.5/jcl-over-slf4j-
1.7.5.jar:/Users/gupta/.m2/repository/log4j/log4j/1.2.17/log4j-
1.2.17.jar:/Users/gupta/.m2/repository/org/slf4j/slf4j-log4j12/1.7.5/slf4j-log4j12-
1.7.5.jar:/Users/gupta/.m2/repository/com/ning/compress-lzf/1.0.0/compress-lzf-
1.0.0.jar:/Users/gupta/.m2/repository/org/xerial/snappy/snappy-java/1.0.5/snappy-java-
1.0.5.jar:/Users/gupta/.m2/repository/com/twitter/chill_2.10/0.3.6/chill_2.10-
0.3.6.jar:/Users/gupta/.m2/repository/com/esotericsoftware/kryo/kryo/2.21/kryo-
2.21.jar:/Users/gupta/.m2/repository/com/esotericsoftware/reflectasm/reflectasm/1.07/reflectasm-1.07-
shaded.jar:/Users/gupta/.m2/repository/com/esotericsoftware/minlog/minlog/1.2/minlog-
1.2.jar:/Users/gupta/.m2/repository/org/objenesis/objenesis/1.2/objenesis-
1.2.jar:/Users/gupta/.m2/repository/com/twitter/chill-java/0.3.6/chill-java-
0.3.6.jar:/Users/gupta/.m2/repository/commons-net/commons-net/2.2/commons-net-
2.2.jar:/Users/gupta/.m2/repository/org/spark-project/akka/akka-remote_2.10/2.2.3-shaded-
protobuf/akka-remote_2.10-2.2.3-shaded-protobuf.jar:/Users/gupta/.m2/repository/org/spark-
project/akka/akka-actor_2.10/2.2.3-shaded-protobuf/akka-actor_2.10-2.2.3-shaded-
protobuf.jar:/Users/gupta/.m2/repository/com/typesafe/config/1.0.2/config-
1.0.2.jar:/Users/gupta/.m2/repository/io/netty/netty/3.6.6.Final/netty-
3.6.6.Final.jar:/Users/gupta/.m2/repository/org/spark-project/protobuf/protobuf-java/2.4.1-
shaded/protobuf-java-2.4.1-
shaded.jar:/Users/gupta/.m2/repository/org/uncommons/maths/uncommons-maths/1.2.2a/uncommons-
maths-1.2.2a.jar:/Users/gupta/.m2/repository/org/spark-project/akka/akka-slf4j_2.10/2.2.3-shaded-
protobuf/akka-slf4j_2.10-2.2.3-shaded-protobuf.jar:/Users/gupta/.m2/repository/org/scala-lang/scala-
library/2.10.4/scala-library-2.10.4.jar:/Users/gupta/.m2/repository/org/json4s/json4s-
jackson_2.10/3.2.6/json4s-jackson_2.10-3.2.6.jar:/Users/gupta/.m2/repository/org/json4s/json4s-
core_2.10/3.2.6/json4s-core_2.10-3.2.6.jar:/Users/gupta/.m2/repository/org/json4s/json4s-
ast_2.10/3.2.6/json4s-ast_2.10-
3.2.6.jar:/Users/gupta/.m2/repository/com/thoughtworks/paranamer/paranamer/2.6/paranamer-
2.6.jar:/Users/gupta/.m2/repository/org/scala-lang/scalap/2.10.0/scalap-
2.10.0.jar:/Users/gupta/.m2/repository/org/scala-lang/scala-compiler/2.10.0/scala-compiler-
2.10.0.jar:/Users/gupta/.m2/repository/org/scala-lang/scala-reflect/2.10.0/scala-reflect-
2.10.0.jar:/Users/gupta/.m2/repository/com/fasterxml/jackson/core/jackson-databind/2.3.0/jackson-
databind-2.3.0.jar:/Users/gupta/.m2/repository/com/fasterxml/jackson/core/jackson-
annotations/2.3.0/jackson-annotations-
2.3.0.jar:/Users/gupta/.m2/repository/com/fasterxml/jackson/core/jackson-core/2.3.0/jackson-core-
2.3.0.jar:/Users/gupta/.m2/repository/org/colt/colt/1.2.0/colt-
1.2.0.jar:/Users/gupta/.m2/repository/concurrent/concurrent/1.3.4/concurrent-
1.3.4.jar:/Users/gupta/.m2/repository/org/apache/mesos/mesos/0.18.1/mesos-0.18.1-shaded-
protobuf.jar:/Users/gupta/.m2/repository/io/netty/netty-all/4.0.17.Final/netty-all-
4.0.17.Final.jar:/Users/gupta/.m2/repository/com/clearspring/analytics/stream/2.5.1/stream-
2.5.1.jar:/Users/gupta/.m2/repository/com/codahale/metrics/metrics-core/3.0.0/metrics-core-
3.0.0.jar:/Users/gupta/.m2/repository/com/codahale/metrics/metrics-jvm/3.0.0/metrics-jvm-
3.0.0.jar:/Users/gupta/.m2/repository/com/codahale/metrics/metrics-json/3.0.0/metrics-json-
3.0.0.jar:/Users/gupta/.m2/repository/com/codahale/metrics/metrics-graphite/3.0.0/metrics-graphite-
3.0.0.jar:/Users/gupta/.m2/repository/org/tachyonproject/tachyon/0.4.1-thrift/tachyon-0.4.1-
thrift.jar:/Users/gupta/.m2/repository/org/apache/ant/ant/1.9.0/ant-
1.9.0.jar:/Users/gupta/.m2/repository/org/apache/ant/ant-launcher/1.9.0/ant-launcher-
1.9.0.jar:/Users/gupta/.m2/repository/commons-io/commons-io/2.4/commons-io-

2.4.jar:/Users/gupta/.m2/repository/org/spark-project/pyrolite/2.0.1/pyrolite-
 2.0.1.jar:/Users/gupta/.m2/repository/net/sf/py4j/py4j/0.8.1/py4j-0.8.1.jar ShakespeareAnalytics
 AllsWellThatEndsWell.txt
 23/12/02 20:24:15 INFO spark.SecurityManager: Changing view acls to: gupta
 23/12/02 20:24:15 INFO spark.SecurityManager: SecurityManager: authentication disabled; ui acls
 disabled; users with view permissions: Set(gupta)
 23/12/02 20:24:15 INFO slf4j.Slf4jLogger: Slf4jLogger started
 23/12/02 20:24:15 INFO Remoting: Starting remoting
 23/12/02 20:24:15 INFO Remoting: Remoting started; listening on
 addresses :[akka.tcp://spark@10.0.0.187:58243]
 23/12/02 20:24:15 INFO Remoting: Remoting now listens on addresses:
 [akka.tcp://spark@10.0.0.187:58243]
 23/12/02 20:24:15 INFO spark.SparkEnv: Registering MapOutputTracker
 23/12/02 20:24:15 INFO spark.SparkEnv: Registering BlockManagerMaster
 23/12/02 20:24:15 INFO storage.DiskBlockManager: Created local directory at
 /var/folders/5t/_27frd4j1xd25sj5h3t3gv280000gn/T/spark-local-20231202202415-b8aa
 23/12/02 20:24:15 INFO storage.MemoryStore: MemoryStore started with capacity 1092.3 MB.
 23/12/02 20:24:15 INFO network.ConnectionManager: Bound socket to port 58244 with id =
 ConnectionManagerId(10.0.0.187,58244)
 23/12/02 20:24:15 INFO storage.BlockManagerMaster: Trying to register BlockManager
 23/12/02 20:24:15 INFO storage.BlockManagerInfo: Registering block manager 10.0.0.187:58244 with
 1092.3 MB RAM
 23/12/02 20:24:15 INFO storage.BlockManagerMaster: Registered BlockManager
 23/12/02 20:24:15 INFO spark.HttpServer: Starting HTTP Server
 23/12/02 20:24:15 INFO server.Server: jetty-8.1.14.v20131031
 23/12/02 20:24:15 INFO server.AbstractConnector: Started SocketConnector@0.0.0.0:58245
 23/12/02 20:24:15 INFO broadcast.HttpBroadcast: Broadcast server started at http://10.0.0.187:58245
 23/12/02 20:24:15 INFO spark.HttpFileServer: HTTP File server directory is
 /var/folders/5t/_27frd4j1xd25sj5h3t3gv280000gn/T/spark-7fd2a00b-2b53-442c-8d8f-df6e3ae3b090
 23/12/02 20:24:15 INFO spark.HttpServer: Starting HTTP Server
 23/12/02 20:24:15 INFO server.Server: jetty-8.1.14.v20131031
 23/12/02 20:24:15 INFO server.AbstractConnector: Started SocketConnector@0.0.0.0:58246
 23/12/02 20:24:16 INFO server.Server: jetty-8.1.14.v20131031
 23/12/02 20:24:16 INFO server.AbstractConnector: Started SelectChannelConnector@0.0.0.0:4040
 23/12/02 20:24:16 INFO ui.SparkUI: Started SparkUI at http://10.0.0.187:4040
 23/12/02 20:24:16 INFO storage.MemoryStore: ensureFreeSpace(32816) called with curMem=0,
 maxMem=1145359564
 23/12/02 20:24:16 INFO storage.MemoryStore: Block broadcast_0 stored as values to memory (estimated
 size 32.0 KB, free 1092.3 MB)
 23/12/02 20:24:16 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your
 platform... using builtin-java classes where applicable
 23/12/02 20:24:16 WARN snappy.LoadSnappy: Snappy native library not loaded
 23/12/02 20:24:16 INFO mapred.FileInputFormat: Total input paths to process : 1
 23/12/02 20:24:16 INFO spark.SparkContext: Starting job: count at ShakespeareAnalytics.java:19
 23/12/02 20:24:16 INFO scheduler.DAGScheduler: Got job 0 (count at ShakespeareAnalytics.java:19)
 with 1 output partitions (allowLocal=false)
 23/12/02 20:24:16 INFO scheduler.DAGScheduler: Final stage: Stage 0(count at
 ShakespeareAnalytics.java:19)
 23/12/02 20:24:16 INFO scheduler.DAGScheduler: Parents of final stage: List()
 23/12/02 20:24:16 INFO scheduler.DAGScheduler: Missing parents: List()
 23/12/02 20:24:16 INFO scheduler.DAGScheduler: Submitting Stage 0 (MappedRDD[1] at textFile at
 ShakespeareAnalytics.java:16), which has no missing parents

23/12/02 20:24:16 INFO scheduler.DAGScheduler: Submitting 1 missing tasks from Stage 0 (MappedRDD[1] at textFile at ShakespeareAnalytics.java:16)

23/12/02 20:24:16 INFO scheduler.TaskSchedulerImpl: Adding task set 0.0 with 1 tasks

23/12/02 20:24:16 INFO scheduler.TaskSetManager: Starting task 0.0:0 as TID 0 on executor localhost: localhost (PROCESS_LOCAL)

23/12/02 20:24:16 INFO scheduler.TaskSetManager: Serialized task 0.0:0 as 1746 bytes in 1 ms

23/12/02 20:24:16 INFO executor.Executor: Running task ID 0

23/12/02 20:24:16 INFO storage.BlockManager: Found block broadcast_0 locally

23/12/02 20:24:16 INFO rdd.HadoopRDD: Input split: file:/Users/gupta/DSProjects/Project5Part2Tasks/AllsWellThatEndsWell.txt:0+133886

23/12/02 20:24:16 INFO executor.Executor: Serialized size of result for 0 is 597

23/12/02 20:24:16 INFO executor.Executor: Sending result for 0 directly to driver

23/12/02 20:24:16 INFO executor.Executor: Finished task ID 0

23/12/02 20:24:16 INFO scheduler.TaskSetManager: Finished TID 0 in 58 ms on localhost (progress: 1/1)

23/12/02 20:24:16 INFO scheduler.DAGScheduler: Completed ResultTask(0, 0)

23/12/02 20:24:16 INFO scheduler.TaskSchedulerImpl: Removed TaskSet 0.0, whose tasks have all completed, from pool

23/12/02 20:24:16 INFO scheduler.DAGScheduler: Stage 0 (count at ShakespeareAnalytics.java:19) finished in 0.065 s

23/12/02 20:24:16 INFO spark.SparkContext: Job finished: count at ShakespeareAnalytics.java:19, took 0.123057208 s

23/12/02 20:24:16 INFO spark.SparkContext: Starting job: count at ShakespeareAnalytics.java:25

23/12/02 20:24:16 INFO scheduler.DAGScheduler: Got job 1 (count at ShakespeareAnalytics.java:25) with 1 output partitions (allowLocal=false)

23/12/02 20:24:16 INFO scheduler.DAGScheduler: Final stage: Stage 1(count at ShakespeareAnalytics.java:25)

23/12/02 20:24:16 INFO scheduler.DAGScheduler: Parents of final stage: List()

23/12/02 20:24:16 INFO scheduler.DAGScheduler: Missing parents: List()

23/12/02 20:24:16 INFO scheduler.DAGScheduler: Submitting Stage 1 (FilteredRDD[3] at filter at ShakespeareAnalytics.java:24), which has no missing parents

Task 0: Number of lines in 'All's Well That Ends Well': 4570

23/12/02 20:24:16 INFO scheduler.DAGScheduler: Submitting 1 missing tasks from Stage 1 (FilteredRDD[3] at filter at ShakespeareAnalytics.java:24)

23/12/02 20:24:16 INFO scheduler.TaskSchedulerImpl: Adding task set 1.0 with 1 tasks

23/12/02 20:24:16 INFO scheduler.TaskSetManager: Starting task 1.0:0 as TID 1 on executor localhost: localhost (PROCESS_LOCAL)

23/12/02 20:24:16 INFO scheduler.TaskSetManager: Serialized task 1.0:0 as 2338 bytes in 0 ms

23/12/02 20:24:16 INFO executor.Executor: Running task ID 1

23/12/02 20:24:16 INFO storage.BlockManager: Found block broadcast_0 locally

23/12/02 20:24:16 INFO rdd.HadoopRDD: Input split: file:/Users/gupta/DSProjects/Project5Part2Tasks/AllsWellThatEndsWell.txt:0+133886

23/12/02 20:24:16 INFO executor.Executor: Serialized size of result for 1 is 597

23/12/02 20:24:16 INFO executor.Executor: Sending result for 1 directly to driver

23/12/02 20:24:16 INFO executor.Executor: Finished task ID 1

23/12/02 20:24:16 INFO scheduler.DAGScheduler: Completed ResultTask(1, 0)

23/12/02 20:24:16 INFO scheduler.TaskSetManager: Finished TID 1 in 52 ms on localhost (progress: 1/1)

23/12/02 20:24:16 INFO scheduler.TaskSchedulerImpl: Removed TaskSet 1.0, whose tasks have all completed, from pool

23/12/02 20:24:16 INFO scheduler.DAGScheduler: Stage 1 (count at ShakespeareAnalytics.java:25) finished in 0.053 s

23/12/02 20:24:16 INFO spark.SparkContext: Job finished: count at ShakespeareAnalytics.java:25, took 0.061630292 s

23/12/02 20:24:16 INFO spark.SparkContext: Starting job: count at ShakespeareAnalytics.java:29

23/12/02 20:24:16 INFO scheduler.DAGScheduler: Registering RDD 5 (distinct at ShakespeareAnalytics.java:29)
23/12/02 20:24:16 INFO scheduler.DAGScheduler: Got job 2 (count at ShakespeareAnalytics.java:29) with 1 output partitions (allowLocal=false)
23/12/02 20:24:16 INFO scheduler.DAGScheduler: Final stage: Stage 2(count at ShakespeareAnalytics.java:29)
23/12/02 20:24:16 INFO scheduler.DAGScheduler: Parents of final stage: List(Stage 3)
23/12/02 20:24:16 INFO scheduler.DAGScheduler: Missing parents: List(Stage 3)
23/12/02 20:24:16 INFO scheduler.DAGScheduler: Submitting Stage 3 (MapPartitionsRDD[5] at distinct at ShakespeareAnalytics.java:29), which has no missing parents
Task 1: Number of words in 'All's Well That Ends Well': 25085
23/12/02 20:24:16 INFO scheduler.DAGScheduler: Submitting 1 missing tasks from Stage 3 (MapPartitionsRDD[5] at distinct at ShakespeareAnalytics.java:29)
23/12/02 20:24:16 INFO scheduler.TaskSchedulerImpl: Adding task set 3.0 with 1 tasks
23/12/02 20:24:16 INFO scheduler.TaskSetManager: Starting task 3.0:0 as TID 2 on executor localhost: localhost (PROCESS_LOCAL)
23/12/02 20:24:16 INFO scheduler.TaskSetManager: Serialized task 3.0:0 as 2637 bytes in 0 ms
23/12/02 20:24:16 INFO executor.Executor: Running task ID 2
23/12/02 20:24:16 INFO storage.BlockManager: Found block broadcast_0 locally
23/12/02 20:24:16 INFO rdd.HadoopRDD: Input split:
file:/Users/gupta/DSProjects/Project5Part2Tasks/AllsWellThatEndsWell.txt:0+133886
23/12/02 20:24:16 INFO executor.Executor: Serialized size of result for 2 is 780
23/12/02 20:24:16 INFO executor.Executor: Sending result for 2 directly to driver
23/12/02 20:24:16 INFO executor.Executor: Finished task ID 2
23/12/02 20:24:16 INFO scheduler.TaskSetManager: Finished TID 2 in 146 ms on localhost (progress: 1/1)
23/12/02 20:24:16 INFO scheduler.TaskSchedulerImpl: Removed TaskSet 3.0, whose tasks have all completed, from pool
23/12/02 20:24:16 INFO scheduler.DAGScheduler: Completed ShuffleMapTask(3, 0)
23/12/02 20:24:16 INFO scheduler.DAGScheduler: Stage 3 (distinct at ShakespeareAnalytics.java:29) finished in 0.147 s
23/12/02 20:24:16 INFO scheduler.DAGScheduler: looking for newly runnable stages
23/12/02 20:24:16 INFO scheduler.DAGScheduler: running: Set()
23/12/02 20:24:16 INFO scheduler.DAGScheduler: waiting: Set(Stage 2)
23/12/02 20:24:16 INFO scheduler.DAGScheduler: failed: Set()
23/12/02 20:24:16 INFO scheduler.DAGScheduler: Missing parents for Stage 2: List()
23/12/02 20:24:16 INFO scheduler.DAGScheduler: Submitting Stage 2 (MappedRDD[8] at distinct at ShakespeareAnalytics.java:29), which is now runnable
23/12/02 20:24:16 INFO scheduler.DAGScheduler: Submitting 1 missing tasks from Stage 2 (MappedRDD[8] at distinct at ShakespeareAnalytics.java:29)
23/12/02 20:24:16 INFO scheduler.TaskSchedulerImpl: Adding task set 2.0 with 1 tasks
23/12/02 20:24:16 INFO scheduler.TaskSetManager: Starting task 2.0:0 as TID 3 on executor localhost: localhost (PROCESS_LOCAL)
23/12/02 20:24:16 INFO scheduler.TaskSetManager: Serialized task 2.0:0 as 2489 bytes in 0 ms
23/12/02 20:24:16 INFO executor.Executor: Running task ID 3
23/12/02 20:24:16 INFO storage.BlockManager: Found block broadcast_0 locally
23/12/02 20:24:16 INFO storage.BlockFetcherIterator\$BasicBlockFetcherIterator: maxBytesInFlight: 50331648, targetRequestSize: 10066329
23/12/02 20:24:16 INFO storage.BlockFetcherIterator\$BasicBlockFetcherIterator: Getting 1 non-empty blocks out of 1 blocks
23/12/02 20:24:16 INFO storage.BlockFetcherIterator\$BasicBlockFetcherIterator: Started 0 remote fetches in 2 ms
23/12/02 20:24:16 INFO executor.Executor: Serialized size of result for 3 is 863

23/12/02 20:24:16 INFO executor.Executor: Sending result for 3 directly to driver
23/12/02 20:24:16 INFO executor.Executor: Finished task ID 3
23/12/02 20:24:16 INFO scheduler.TaskSetManager: Finished TID 3 in 83 ms on localhost (progress: 1/1)
23/12/02 20:24:16 INFO scheduler.TaskSchedulerImpl: Removed TaskSet 2.0, whose tasks have all completed, from pool
23/12/02 20:24:16 INFO scheduler.DAGScheduler: Completed ResultTask(2, 0)
23/12/02 20:24:16 INFO scheduler.DAGScheduler: Stage 2 (count at ShakespeareAnalytics.java:29) finished in 0.084 s
23/12/02 20:24:16 INFO spark.SparkContext: Job finished: count at ShakespeareAnalytics.java:29, took 0.255784917 s
23/12/02 20:24:16 INFO spark.SparkContext: Starting job: count at ShakespeareAnalytics.java:33
23/12/02 20:24:16 INFO scheduler.DAGScheduler: Got job 3 (count at ShakespeareAnalytics.java:33) with 1 output partitions (allowLocal=false)
23/12/02 20:24:16 INFO scheduler.DAGScheduler: Final stage: Stage 4(count at ShakespeareAnalytics.java:33)
23/12/02 20:24:16 INFO scheduler.DAGScheduler: Parents of final stage: List()
23/12/02 20:24:16 INFO scheduler.DAGScheduler: Missing parents: List()
23/12/02 20:24:16 INFO scheduler.DAGScheduler: Submitting Stage 4 (FlatMappedRDD[9] at flatMap at ShakespeareAnalytics.java:33), which has no missing parents
23/12/02 20:24:16 INFO scheduler.DAGScheduler: Submitting 1 missing tasks from Stage 4 (FlatMappedRDD[9] at flatMap at ShakespeareAnalytics.java:33)
23/12/02 20:24:16 INFO scheduler.TaskSchedulerImpl: Adding task set 4.0 with 1 tasks
23/12/02 20:24:16 INFO scheduler.TaskSetManager: Starting task 4.0:0 as TID 4 on executor localhost: localhost (PROCESS_LOCAL)
23/12/02 20:24:16 INFO scheduler.TaskSetManager: Serialized task 4.0:0 as 2222 bytes in 0 ms

Task 2: Number of distinct words in 'All's Well That Ends Well': 3866

23/12/02 20:24:16 INFO executor.Executor: Running task ID 4
23/12/02 20:24:16 INFO storage.BlockManager: Found block broadcast_0 locally
23/12/02 20:24:16 INFO rdd.HadoopRDD: Input split:
file:/Users/gupta/DSProjects/Project5Part2Tasks/AllsWellThatEndsWell.txt:0+133886
23/12/02 20:24:16 INFO storage.ShuffleBlockManager: Deleted all files for shuffle 0
23/12/02 20:24:16 INFO executor.Executor: Serialized size of result for 4 is 597
23/12/02 20:24:16 INFO executor.Executor: Sending result for 4 directly to driver
23/12/02 20:24:16 INFO executor.Executor: Finished task ID 4
23/12/02 20:24:16 INFO scheduler.TaskSetManager: Finished TID 4 in 48 ms on localhost (progress: 1/1)
23/12/02 20:24:16 INFO scheduler.TaskSchedulerImpl: Removed TaskSet 4.0, whose tasks have all completed, from pool
23/12/02 20:24:16 INFO scheduler.DAGScheduler: Completed ResultTask(4, 0)
23/12/02 20:24:16 INFO spark.ContextCleaner: Cleaned shuffle 0
23/12/02 20:24:16 INFO scheduler.DAGScheduler: Stage 4 (count at ShakespeareAnalytics.java:33) finished in 0.049 s
23/12/02 20:24:16 INFO spark.SparkContext: Job finished: count at ShakespeareAnalytics.java:33, took 0.0568595 s
23/12/02 20:24:16 INFO spark.SparkContext: Starting job: count at ShakespeareAnalytics.java:38

Task 3: Number of symbols in 'All's Well That Ends Well': 130373

23/12/02 20:24:16 INFO scheduler.DAGScheduler: Registering RDD 12 (distinct at ShakespeareAnalytics.java:38)
23/12/02 20:24:16 INFO scheduler.DAGScheduler: Got job 4 (count at ShakespeareAnalytics.java:38) with 1 output partitions (allowLocal=false)
23/12/02 20:24:16 INFO scheduler.DAGScheduler: Final stage: Stage 5(count at ShakespeareAnalytics.java:38)
23/12/02 20:24:16 INFO scheduler.DAGScheduler: Parents of final stage: List(Stage 6)
23/12/02 20:24:16 INFO scheduler.DAGScheduler: Missing parents: List(Stage 6)

23/12/02 20:24:16 INFO scheduler.DAGScheduler: Submitting Stage 6 (MapPartitionsRDD[12] at distinct at ShakespeareAnalytics.java:38), which has no missing parents
23/12/02 20:24:16 INFO scheduler.DAGScheduler: Submitting 1 missing tasks from Stage 6 (MapPartitionsRDD[12] at distinct at ShakespeareAnalytics.java:38)
23/12/02 20:24:16 INFO scheduler.TaskSchedulerImpl: Adding task set 6.0 with 1 tasks
23/12/02 20:24:16 INFO scheduler.TaskSetManager: Starting task 6.0:0 as TID 5 on executor localhost: localhost (PROCESS_LOCAL)
23/12/02 20:24:16 INFO scheduler.TaskSetManager: Serialized task 6.0:0 as 2520 bytes in 0 ms
23/12/02 20:24:16 INFO executor.Executor: Running task ID 5
23/12/02 20:24:16 INFO storage.BlockManager: Found block broadcast_0 locally
23/12/02 20:24:16 INFO rdd.HadoopRDD: Input split:
file:/Users/gupta/DSProjects/Project5Part2Tasks/AllsWellThatEndsWell.txt:0+133886
23/12/02 20:24:17 INFO executor.Executor: Serialized size of result for 5 is 780
23/12/02 20:24:17 INFO executor.Executor: Sending result for 5 directly to driver
23/12/02 20:24:17 INFO executor.Executor: Finished task ID 5
23/12/02 20:24:17 INFO scheduler.TaskSetManager: Finished TID 5 in 57 ms on localhost (progress: 1/1)
23/12/02 20:24:17 INFO scheduler.DAGScheduler: Completed ShuffleMapTask(6, 0)
23/12/02 20:24:17 INFO scheduler.TaskSchedulerImpl: Removed TaskSet 6.0, whose tasks have all completed, from pool
23/12/02 20:24:17 INFO scheduler.DAGScheduler: Stage 6 (distinct at ShakespeareAnalytics.java:38) finished in 0.057 s
23/12/02 20:24:17 INFO scheduler.DAGScheduler: looking for newly runnable stages
23/12/02 20:24:17 INFO scheduler.DAGScheduler: running: Set()
23/12/02 20:24:17 INFO scheduler.DAGScheduler: waiting: Set(Stage 5)
23/12/02 20:24:17 INFO scheduler.DAGScheduler: failed: Set()
23/12/02 20:24:17 INFO scheduler.DAGScheduler: Missing parents for Stage 5: List()
23/12/02 20:24:17 INFO scheduler.DAGScheduler: Submitting Stage 5 (MappedRDD[15] at distinct at ShakespeareAnalytics.java:38), which is now runnable
23/12/02 20:24:17 INFO scheduler.DAGScheduler: Submitting 1 missing tasks from Stage 5 (MappedRDD[15] at distinct at ShakespeareAnalytics.java:38)
23/12/02 20:24:17 INFO scheduler.TaskSchedulerImpl: Adding task set 5.0 with 1 tasks
23/12/02 20:24:17 INFO scheduler.TaskSetManager: Starting task 5.0:0 as TID 6 on executor localhost: localhost (PROCESS_LOCAL)
23/12/02 20:24:17 INFO scheduler.TaskSetManager: Serialized task 5.0:0 as 2378 bytes in 1 ms
23/12/02 20:24:17 INFO executor.Executor: Running task ID 6
23/12/02 20:24:17 INFO storage.BlockManager: Found block broadcast_0 locally
23/12/02 20:24:17 INFO storage.BlockFetcherIterator\$BasicBlockFetcherIterator: maxBytesInFlight: 50331648, targetRequestSize: 10066329
23/12/02 20:24:17 INFO storage.BlockFetcherIterator\$BasicBlockFetcherIterator: Getting 1 non-empty blocks out of 1 blocks
23/12/02 20:24:17 INFO storage.BlockFetcherIterator\$BasicBlockFetcherIterator: Started 0 remote fetches in 0 ms
23/12/02 20:24:17 INFO executor.Executor: Serialized size of result for 6 is 863
23/12/02 20:24:17 INFO executor.Executor: Sending result for 6 directly to driver
23/12/02 20:24:17 INFO executor.Executor: Finished task ID 6
23/12/02 20:24:17 INFO scheduler.DAGScheduler: Completed ResultTask(5, 0)
23/12/02 20:24:17 INFO scheduler.TaskSetManager: Finished TID 6 in 9 ms on localhost (progress: 1/1)
23/12/02 20:24:17 INFO scheduler.DAGScheduler: Stage 5 (count at ShakespeareAnalytics.java:38) finished in 0.009 s
23/12/02 20:24:17 INFO scheduler.TaskSchedulerImpl: Removed TaskSet 5.0, whose tasks have all completed, from pool
23/12/02 20:24:17 INFO spark.SparkContext: Job finished: count at ShakespeareAnalytics.java:38, took 0.083360042 s

23/12/02 20:24:17 INFO spark.SparkContext: Starting job: count at ShakespeareAnalytics.java:42
 23/12/02 20:24:17 INFO scheduler.DAGScheduler: Registering RDD 19 (distinct at ShakespeareAnalytics.java:42)
 23/12/02 20:24:17 INFO scheduler.DAGScheduler: Got job 5 (count at ShakespeareAnalytics.java:42) with 1 output partitions (allowLocal=false)
 23/12/02 20:24:17 INFO scheduler.DAGScheduler: Final stage: Stage 7(count at ShakespeareAnalytics.java:42)
 23/12/02 20:24:17 INFO scheduler.DAGScheduler: Parents of final stage: List(Stage 8)
 23/12/02 20:24:17 INFO scheduler.DAGScheduler: Missing parents: List(Stage 8)
 23/12/02 20:24:17 INFO scheduler.DAGScheduler: Submitting Stage 8 (MapPartitionsRDD[19] at distinct at ShakespeareAnalytics.java:42), which has no missing parents
 23/12/02 20:24:17 INFO scheduler.DAGScheduler: Submitting 1 missing tasks from Stage 8 (MapPartitionsRDD[19] at distinct at ShakespeareAnalytics.java:42)
 23/12/02 20:24:17 INFO scheduler.TaskSchedulerImpl: Adding task set 8.0 with 1 tasks
 23/12/02 20:24:17 INFO scheduler.TaskSetManager: Starting task 8.0:0 as TID 7 on executor localhost: localhost (PROCESS_LOCAL)
 23/12/02 20:24:17 INFO scheduler.TaskSetManager: Serialized task 8.0:0 as 2703 bytes in 0 ms
 23/12/02 20:24:17 INFO executor.Executor: Running task ID 7
 23/12/02 20:24:17 INFO storage.BlockManager: Found block broadcast_0 locally
 23/12/02 20:24:17 INFO rdd.HadoopRDD: Input split: file:/Users/gupta/DSProjects/Project5Part2Tasks/AllsWellThatEndsWell.txt:0+133886
Task 4: Number of distinct symbols in 'All's Well That Ends Well': 78
 23/12/02 20:24:17 INFO storage.ShuffleBlockManager: Deleted all files for shuffle 1
 23/12/02 20:24:17 INFO spark.ContextCleaner: Cleaned shuffle 1
 23/12/02 20:24:17 INFO executor.Executor: Serialized size of result for 7 is 780
 23/12/02 20:24:17 INFO executor.Executor: Sending result for 7 directly to driver
 23/12/02 20:24:17 INFO executor.Executor: Finished task ID 7
 23/12/02 20:24:17 INFO scheduler.DAGScheduler: Completed ShuffleMapTask(8, 0)
 23/12/02 20:24:17 INFO scheduler.TaskSetManager: Finished TID 7 in 92 ms on localhost (progress: 1/1)
 23/12/02 20:24:17 INFO scheduler.DAGScheduler: Stage 8 (distinct at ShakespeareAnalytics.java:42) finished in 0.093 s
 23/12/02 20:24:17 INFO scheduler.DAGScheduler: looking for newly runnable stages
 23/12/02 20:24:17 INFO scheduler.DAGScheduler: running: Set()
 23/12/02 20:24:17 INFO scheduler.DAGScheduler: waiting: Set(Stage 7)
 23/12/02 20:24:17 INFO scheduler.DAGScheduler: failed: Set()
 23/12/02 20:24:17 INFO scheduler.TaskSchedulerImpl: Removed TaskSet 8.0, whose tasks have all completed, from pool
 23/12/02 20:24:17 INFO scheduler.DAGScheduler: Missing parents for Stage 7: List()
 23/12/02 20:24:17 INFO scheduler.DAGScheduler: Submitting Stage 7 (MappedRDD[22] at distinct at ShakespeareAnalytics.java:42), which is now runnable
 23/12/02 20:24:17 INFO scheduler.DAGScheduler: Submitting 1 missing tasks from Stage 7 (MappedRDD[22] at distinct at ShakespeareAnalytics.java:42)
 23/12/02 20:24:17 INFO scheduler.TaskSchedulerImpl: Adding task set 7.0 with 1 tasks
 23/12/02 20:24:17 INFO scheduler.TaskSetManager: Starting task 7.0:0 as TID 8 on executor localhost: localhost (PROCESS_LOCAL)
 23/12/02 20:24:17 INFO scheduler.TaskSetManager: Serialized task 7.0:0 as 2548 bytes in 0 ms
 23/12/02 20:24:17 INFO executor.Executor: Running task ID 8
 23/12/02 20:24:17 INFO storage.BlockManager: Found block broadcast_0 locally
 23/12/02 20:24:17 INFO storage.BlockFetcherIterator\$BasicBlockFetcherIterator: maxBytesInFlight: 50331648, targetRequestSize: 10066329
 23/12/02 20:24:17 INFO storage.BlockFetcherIterator\$BasicBlockFetcherIterator: Getting 1 non-empty blocks out of 1 blocks

23/12/02 20:24:17 INFO storage.BlockFetcherIterator\$BasicBlockFetcherIterator: Started 0 remote fetches in 1 ms
23/12/02 20:24:17 INFO executor.Executor: Serialized size of result for 8 is 863
23/12/02 20:24:17 INFO executor.Executor: Sending result for 8 directly to driver
23/12/02 20:24:17 INFO executor.Executor: Finished task ID 8
23/12/02 20:24:17 INFO scheduler.TaskSetManager: Finished TID 8 in 11 ms on localhost (progress: 1/1)
23/12/02 20:24:17 INFO scheduler.DAGScheduler: Completed ResultTask(7, 0)
23/12/02 20:24:17 INFO scheduler.TaskSchedulerImpl: Removed TaskSet 7.0, whose tasks have all completed, from pool
23/12/02 20:24:17 INFO scheduler.DAGScheduler: Stage 7 (count at ShakespeareAnalytics.java:42) finished in 0.011 s
23/12/02 20:24:17 INFO spark.SparkContext: Job finished: count at ShakespeareAnalytics.java:42, took 0.11544725 s

Task 5: Number of distinct letters in 'All's Well That Ends Well': 50

Task 6: Enter a word to search for in 'All's Well That Ends Well': love

Lines containing 'love':

23/12/02 20:24:22 INFO spark.SparkContext: Starting job: collect at ShakespeareAnalytics.java:52
23/12/02 20:24:22 INFO scheduler.DAGScheduler: Got job 6 (collect at ShakespeareAnalytics.java:52) with 1 output partitions (allowLocal=false)
23/12/02 20:24:22 INFO scheduler.DAGScheduler: Final stage: Stage 9(collect at ShakespeareAnalytics.java:52)
23/12/02 20:24:22 INFO scheduler.DAGScheduler: Parents of final stage: List()
23/12/02 20:24:22 INFO scheduler.DAGScheduler: Missing parents: List()
23/12/02 20:24:22 INFO scheduler.DAGScheduler: Submitting Stage 9 (FilteredRDD[23] at filter at ShakespeareAnalytics.java:50), which has no missing parents
23/12/02 20:24:22 INFO scheduler.DAGScheduler: Submitting 1 missing tasks from Stage 9 (FilteredRDD[23] at filter at ShakespeareAnalytics.java:50)
23/12/02 20:24:22 INFO scheduler.TaskSchedulerImpl: Adding task set 9.0 with 1 tasks
23/12/02 20:24:22 INFO scheduler.TaskSetManager: Starting task 9.0:0 as TID 9 on executor localhost: localhost (PROCESS_LOCAL)
23/12/02 20:24:22 INFO scheduler.TaskSetManager: Serialized task 9.0:0 as 2110 bytes in 0 ms
23/12/02 20:24:22 INFO executor.Executor: Running task ID 9
23/12/02 20:24:22 INFO storage.BlockManager: Found block broadcast_0 locally
23/12/02 20:24:22 INFO rdd.HadoopRDD: Input split: file:/Users/gupta/DSProjects/Project5Part2Tasks/AllsWellThatEndsWell.txt:0+133886
23/12/02 20:24:22 INFO executor.Executor: Serialized size of result for 9 is 3801
23/12/02 20:24:22 INFO executor.Executor: Sending result for 9 directly to driver
23/12/02 20:24:22 INFO executor.Executor: Finished task ID 9
23/12/02 20:24:22 INFO scheduler.TaskSetManager: Finished TID 9 in 5 ms on localhost (progress: 1/1)
23/12/02 20:24:22 INFO scheduler.DAGScheduler: Completed ResultTask(9, 0)
23/12/02 20:24:22 INFO scheduler.TaskSchedulerImpl: Removed TaskSet 9.0, whose tasks have all completed, from pool
23/12/02 20:24:22 INFO scheduler.DAGScheduler: Stage 9 (collect at ShakespeareAnalytics.java:52) finished in 0.007 s
23/12/02 20:24:22 INFO spark.SparkContext: Job finished: collect at ShakespeareAnalytics.java:52, took 0.009140583 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.
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 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
LAFEW This woman's an easy glove, my lord; she goes
I'll love her dearly, ever, ever dearly.

```
23/12/02 20:24:22 INFO handler.ContextHandler: stopped  
    o.e.j.s.ServletContextHandler{/metrics/json,null}  
23/12/02 20:24:22 INFO handler.ContextHandler: stopped  
    o.e.j.s.ServletContextHandler{/stages/stage/kill,null}  
23/12/02 20:24:22 INFO handler.ContextHandler: stopped o.e.j.s.ServletContextHandler{/,null}  
23/12/02 20:24:22 INFO handler.ContextHandler: stopped o.e.j.s.ServletContextHandler{/static,null}  
23/12/02 20:24:22 INFO handler.ContextHandler: stopped  
    o.e.j.s.ServletContextHandler{/executors/json,null}  
23/12/02 20:24:22 INFO handler.ContextHandler: stopped o.e.j.s.ServletContextHandler{/executors,null}  
23/12/02 20:24:22 INFO handler.ContextHandler: stopped  
    o.e.j.s.ServletContextHandler{/environment/json,null}  
23/12/02 20:24:22 INFO handler.ContextHandler: stopped  
    o.e.j.s.ServletContextHandler{/environment,null}  
23/12/02 20:24:22 INFO handler.ContextHandler: stopped  
    o.e.j.s.ServletContextHandler{/storage/rdd/json,null}  
23/12/02 20:24:22 INFO handler.ContextHandler: stopped  
    o.e.j.s.ServletContextHandler{/storage/rdd,null}  
23/12/02 20:24:22 INFO handler.ContextHandler: stopped  
    o.e.j.s.ServletContextHandler{/storage/json,null}  
23/12/02 20:24:22 INFO handler.ContextHandler: stopped o.e.j.s.ServletContextHandler{/storage,null}  
23/12/02 20:24:22 INFO handler.ContextHandler: stopped  
    o.e.j.s.ServletContextHandler{/stages/pool/json,null}  
23/12/02 20:24:22 INFO handler.ContextHandler: stopped  
    o.e.j.s.ServletContextHandler{/stages/pool,null}  
23/12/02 20:24:22 INFO handler.ContextHandler: stopped  
    o.e.j.s.ServletContextHandler{/stages/stage/json,null}  
23/12/02 20:24:22 INFO handler.ContextHandler: stopped  
    o.e.j.s.ServletContextHandler{/stages/stage,null}  
23/12/02 20:24:22 INFO handler.ContextHandler: stopped  
    o.e.j.s.ServletContextHandler{/stages/json,null}  
23/12/02 20:24:22 INFO handler.ContextHandler: stopped o.e.j.s.ServletContextHandler{/stages,null}  
23/12/02 20:24:22 INFO ui.SparkUI: Stopped Spark web UI at http://10.0.0.187:4040  
23/12/02 20:24:22 INFO scheduler.DAGScheduler: Stopping DAGScheduler  
23/12/02 20:24:23 INFO spark.MapOutputTrackerMasterActor: MapOutputTrackerActor stopped!  
23/12/02 20:24:23 INFO network.ConnectionManager: Selector thread was interrupted!  
23/12/02 20:24:23 INFO network.ConnectionManager: ConnectionManager stopped  
23/12/02 20:24:23 INFO storage.MemoryStore: MemoryStore cleared  
23/12/02 20:24:23 INFO storage.BlockManager: BlockManager stopped  
23/12/02 20:24:23 INFO storage.BlockManagerMasterActor: Stopping BlockManagerMaster  
23/12/02 20:24:23 INFO storage.BlockManagerMaster: BlockManagerMaster stopped  
23/12/02 20:24:23 INFO spark.SparkContext: Successfully stopped SparkContext
```

Process finished with exit code 0

Project5Part1

- Heinz Cluster user id - student034
- Heinz Cluster password – 2June1997@