

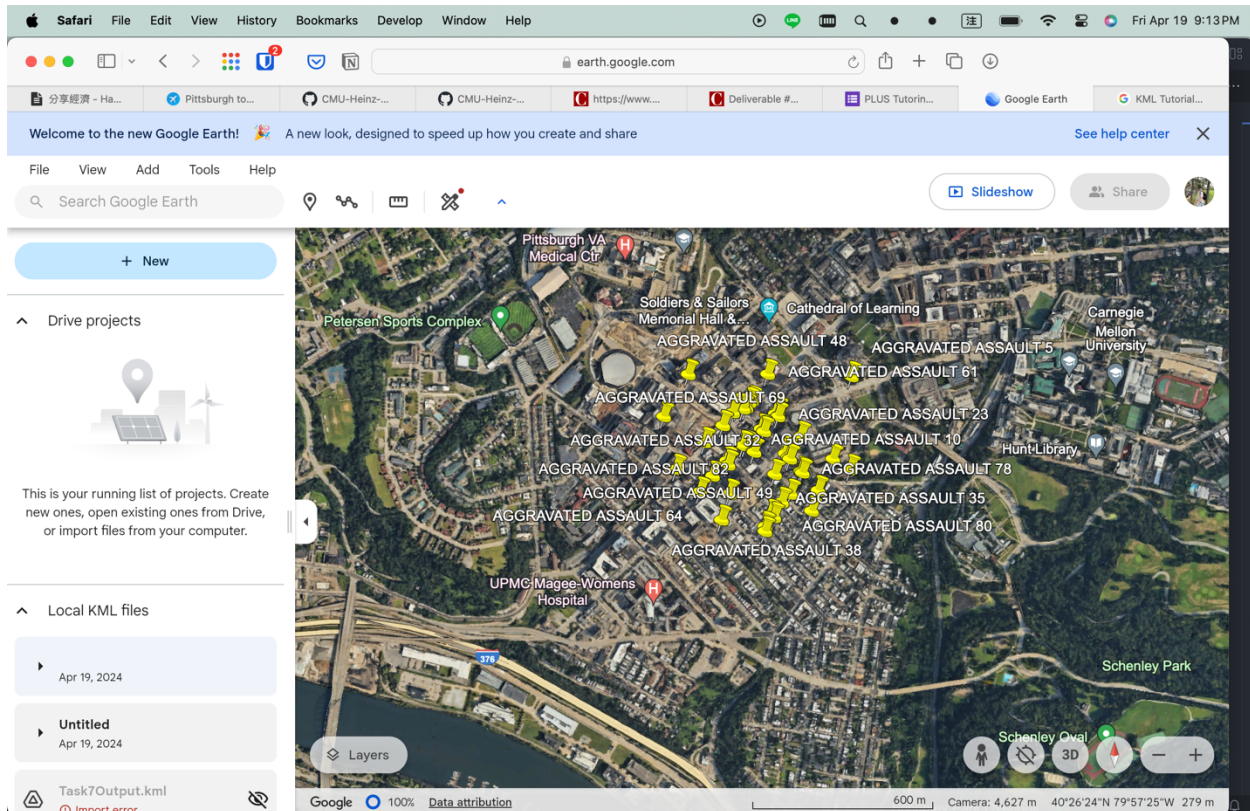
# Project 5

Louis Chang (hungyic)

Cluster User ID: student017, Password: Louis1999!

## Part 1 Map Reduce

### Task 7



## Part 2

### Task 0 – Task 5

```
import org.apache.spark.SparkConf;
import org.apache.spark.api.java.JavaPairRDD;
import org.apache.spark.api.java.JavaRDD;
import org.apache.spark.api.java.JavaSparkContext;
import scala.Tuple2;

import java.util.Arrays;

public class WordCounter {
```

```

private static void wordCount(String fileName) {

    SparkConf sparkConf = new
SparkConf().setMaster("local").setAppName("Extended Word Counter");
    JavaSparkContext sparkContext = new JavaSparkContext(sparkConf);

    // Read the file into an RDD
    JavaRDD<String> inputFile = sparkContext.textFile(fileName);

    // Count the number of lines
    long numberOfLines = inputFile.count();

    // Flatten the lines into words
    JavaRDD<String> wordsFromFile = inputFile.flatMap(content ->
Arrays.asList(content.split("[^a-zA-Z]+")));
    JavaRDD<String> filteredWords = wordsFromFile.filter(s -
> !s.isEmpty());

    // Count the number of words
    long numberOfWords = filteredWords.count();

    // Create a PairRDD with each word and count each word
    JavaPairRDD<String, Integer> wordCounts =
filteredWords.mapToPair(word -> new Tuple2<>(word, 1))
                .reduceByKey(Integer::sum);

    // Count the number of distinct words
    long numberOfDistinctWords = wordCounts.keys().distinct().count();

    // Flatten the lines into characters
    JavaRDD<String> symbolsFromFile = inputFile.flatMap(content ->
Arrays.asList(content.split("")));

    // Count the number of symbols
    long numberOfSymbols = symbolsFromFile.count();

    // Count distinct symbols
    long numberOfDistinctSymbols = symbolsFromFile.distinct().count();

    // Filter only letters and count distinct letters
    JavaRDD<String> lettersFromFile = symbolsFromFile.filter(s ->
s.matches("[a-zA-Z]"));
    long numberOfDistinctLetters = lettersFromFile.distinct().count();

    // Output the results to the console or write them back to a file
system
    System.out.println("Number of lines: " + numberOfLines);
    System.out.println("Number of words: " + numberOfWords);
    System.out.println("Number of distinct words: " +
numberOfDistinctWords);
    System.out.println("Number of symbols: " + numberOfSymbols);
    System.out.println("Number of distinct symbols: " +
numberOfDistinctSymbols);
}

```

```

        System.out.println("Number of distinct letters: " +
numberOfDistinctLetters);

    }

    public static void main(String[] args) {

        if (args.length == 0) {
            System.out.println("No files provided.");
            System.exit(0);
        }

        wordCount(args[0]);

    }
}

```

/Library/Java/JavaVirtualMachines/jdk1.8.0\_151.jdk/Contents/Home/bin/java -  
javaagent:/Users/louischang/Applications/IntelliJ IDEA  
Ultimate.app/Contents/lib/idea\_rt.jar=54445:/Users/louischang/Applications/IntelliJ IDEA  
Ultimate.app/Contents/bin -Dfile.encoding=UTF-8 -classpath  
/Library/Java/JavaVirtualMachines/jdk1.8.0\_151.jdk/Contents/Home/jre/lib/charsets.jar:/Li  
brary/Java/JavaVirtualMachines/jdk1.8.0\_151.jdk/Contents/Home/jre/lib/deploy.jar:/Librar  
y/Java/JavaVirtualMachines/jdk1.8.0\_151.jdk/Contents/Home/jre/lib/ext/cldrdata.jar:/Librar  
y/Java/JavaVirtualMachines/jdk1.8.0\_151.jdk/Contents/Home/jre/lib/ext/dnsns.jar:/Librar  
y/Java/JavaVirtualMachines/jdk1.8.0\_151.jdk/Contents/Home/jre/lib/ext/jaccess.jar:/Librar  
y/Java/JavaVirtualMachines/jdk1.8.0\_151.jdk/Contents/Home/jre/lib/ext/jfxrt.jar:/Library/Ja  
va/JavaVirtualMachines/jdk1.8.0\_151.jdk/Contents/Home/jre/lib/ext/localedata.jar:/Librar  
y/Java/JavaVirtualMachines/jdk1.8.0\_151.jdk/Contents/Home/jre/lib/ext/nashorn.jar:/Librar  
y/Java/JavaVirtualMachines/jdk1.8.0\_151.jdk/Contents/Home/jre/lib/ext/sunec.jar:/Librar  
y/Java/JavaVirtualMachines/jdk1.8.0\_151.jdk/Contents/Home/jre/lib/ext/sunjce\_provider.ja  
r:/Library/Java/JavaVirtualMachines/jdk1.8.0\_151.jdk/Contents/Home/jre/lib/ext/sunpkcs1  
1.jar:/Library/Java/JavaVirtualMachines/jdk1.8.0\_151.jdk/Contents/Home/jre/lib/ext/zipfs.j  
ar:/Library/Java/JavaVirtualMachines/jdk1.8.0\_151.jdk/Contents/Home/jre/lib/javaws.jar:/  
Library/Java/JavaVirtualMachines/jdk1.8.0\_151.jdk/Contents/Home/jre/lib/jce.jar:/Library/J  
ava/JavaVirtualMachines/jdk1.8.0\_151.jdk/Contents/Home/jre/lib/jfr.jar:/Library/Java/Java  
VirtualMachines/jdk1.8.0\_151.jdk/Contents/Home/jre/lib/jfxswt.jar:/Library/Java/JavaVirtu  
alMachines/jdk1.8.0\_151.jdk/Contents/Home/jre/lib/jsse.jar:/Library/Java/JavaVirtualMach  
ines/jdk1.8.0\_151.jdk/Contents/Home/jre/lib/management-  
agent.jar:/Library/Java/JavaVirtualMachines/jdk1.8.0\_151.jdk/Contents/Home/jre/lib/plugi  
n.jar:/Library/Java/JavaVirtualMachines/jdk1.8.0\_151.jdk/Contents/Home/jre/lib/resources  
.jar:/Library/Java/JavaVirtualMachines/jdk1.8.0\_151.jdk/Contents/Home/jre/lib/rt.jar:/Libra  
ry/Java/JavaVirtualMachines/jdk1.8.0\_151.jdk/Contents/Home/lib/ant-

javafx.jar:/Library/Java/JavaVirtualMachines/jdk1.8.0\_151.jdk/Contents/Home/lib/dt.jar:/Library/Java/JavaVirtualMachines/jdk1.8.0\_151.jdk/Contents/Home/lib/javafx-mx.jar:/Library/Java/JavaVirtualMachines/jdk1.8.0\_151.jdk/Contents/Home/lib/jconsole.jar:/Library/Java/JavaVirtualMachines/jdk1.8.0\_151.jdk/Contents/Home/lib/packager.jar:/Library/Java/JavaVirtualMachines/jdk1.8.0\_151.jdk/Contents/Home/lib/sa-jdi.jar:/Library/Java/JavaVirtualMachines/jdk1.8.0\_151.jdk/Contents/Home/lib/tools.jar:/Users/louischang/Library/CloudStorage/OneDrive-andrew.cmu.edu/DS/Project5/Project5Part2/target/classes:/Users/louischang/.m2/repository/org/apache/spark/spark-core\_2.10/1.0.0/spark-core\_2.10-1.0.0.jar:/Users/louischang/.m2/repository/org/apache/hadoop/hadoop-client/1.0.4/hadoop-client-1.0.4.jar:/Users/louischang/.m2/repository/org/apache/hadoop/hadoop-core/1.0.4/hadoop-core-1.0.4.jar:/Users/louischang/.m2/repository/xmlenc/xmlenc/0.52/xmlenc-0.52.jar:/Users/louischang/.m2/repository/org/apache/commons/commons-math/2.1/commons-math-2.1.jar:/Users/louischang/.m2/repository/commons-configuration/commons-configuration/1.6/commons-configuration-1.6.jar:/Users/louischang/.m2/repository/commons-collections/commons-collections/3.2.1/commons-collections-3.2.1.jar:/Users/louischang/.m2/repository/commons-lang/commons-lang/2.4/commons-lang-2.4.jar:/Users/louischang/.m2/repository/commons-digester/commons-digester/1.8/commons-digester-1.8.jar:/Users/louischang/.m2/repository/commons-beanutils/commons-beanutils/1.7.0/commons-beanutils-1.7.0.jar:/Users/louischang/.m2/repository/commons-beanutils/commons-beanutils-core/1.8.0/commons-beanutils-core-1.8.0.jar:/Users/louischang/.m2/repository/commons-el/commons-el/1.0/commons-el-1.0.jar:/Users/louischang/.m2/repository/hsqldb/hsqldb/1.8.0.10/hsqldb-1.8.0.10.jar:/Users/louischang/.m2/repository/oro/oro/2.0.8/oro-2.0.8.jar:/Users/louischang/.m2/repository/org/codehaus/jackson/jackson-mapper-asl/1.0.1/jackson-mapper-asl-1.0.1.jar:/Users/louischang/.m2/repository/org/codehaus/jackson/jackson-core-asl/1.0.1/jackson-core-asl-1.0.1.jar:/Users/louischang/.m2/repository/net/java/dev/jets3t/jets3t/0.7.1/jets3t-0.7.1.jar:/Users/louischang/.m2/repository/commons-codec/commons-codec/1.3/commons-codec-1.3.jar:/Users/louischang/.m2/repository/commons-httpclient/commons-httpclient/3.1/commons-httpclient-3.1.jar:/Users/louischang/.m2/repository/org/apache/curator/curator-recipes/2.4.0/curator-recipes-

2.4.0.jar:/Users/louischang/.m2/repository/org/apache/curator/curator-framework/2.4.0/curator-framework-  
2.4.0.jar:/Users/louischang/.m2/repository/org/apache/curator/curator-client/2.4.0/curator-client-  
2.4.0.jar:/Users/louischang/.m2/repository/org/apache/zookeeper/zookeeper/3.4.5/zookeeper-3.4.5.jar:/Users/louischang/.m2/repository/jline/jline/0.9.94/jline-  
0.9.94.jar:/Users/louischang/.m2/repository/org/eclipse/jetty/jetty-plus/8.1.14.v20131031/jetty-plus-  
8.1.14.v20131031.jar:/Users/louischang/.m2/repository/org/eclipse/jetty/orbit/javax.transaction/1.1.1.v201105210645/javax.transaction-  
1.1.1.v201105210645.jar:/Users/louischang/.m2/repository/org/eclipse/jetty/jetty-webapp/8.1.14.v20131031/jetty-webapp-  
8.1.14.v20131031.jar:/Users/louischang/.m2/repository/org/eclipse/jetty/jetty-xml/8.1.14.v20131031/jetty-xml-  
8.1.14.v20131031.jar:/Users/louischang/.m2/repository/org/eclipse/jetty/jetty-servlet/8.1.14.v20131031/jetty-servlet-  
8.1.14.v20131031.jar:/Users/louischang/.m2/repository/org/eclipse/jetty/jetty-jndi/8.1.14.v20131031/jetty-jndi-  
8.1.14.v20131031.jar:/Users/louischang/.m2/repository/org/eclipse/jetty/orbit/javax.mail.glassfish/1.4.1.v201005082020/javax.mail.glassfish-  
1.4.1.v201005082020.jar:/Users/louischang/.m2/repository/org/eclipse/jetty/orbit/javax.activation/1.1.0.v201105071233/javax.activation-  
1.1.0.v201105071233.jar:/Users/louischang/.m2/repository/org/eclipse/jetty/jetty-security/8.1.14.v20131031/jetty-security-  
8.1.14.v20131031.jar:/Users/louischang/.m2/repository/org/eclipse/jetty/jetty-util/8.1.14.v20131031/jetty-util-  
8.1.14.v20131031.jar:/Users/louischang/.m2/repository/org/eclipse/jetty/jetty-server/8.1.14.v20131031/jetty-server-  
8.1.14.v20131031.jar:/Users/louischang/.m2/repository/org/eclipse/jetty/orbit/javax.servlet/3.0.0.v201112011016/javax.servlet-  
3.0.0.v201112011016.jar:/Users/louischang/.m2/repository/org/eclipse/jetty/jetty-continuation/8.1.14.v20131031/jetty-continuation-  
8.1.14.v20131031.jar:/Users/louischang/.m2/repository/org/eclipse/jetty/jetty-http/8.1.14.v20131031/jetty-http-  
8.1.14.v20131031.jar:/Users/louischang/.m2/repository/org/eclipse/jetty/jetty-io/8.1.14.v20131031/jetty-io-  
8.1.14.v20131031.jar:/Users/louischang/.m2/repository/com/google/guava/guava/14.0.1/guava-14.0.1.jar:/Users/louischang/.m2/repository/org/apache/commons/commons-

lang3/3.3.2/commons-lang3-  
3.3.2.jar:/Users/louischang/.m2/repository/com/google/code/findbugs/jsr305/1.3.9/jsr305-  
1.3.9.jar:/Users/louischang/.m2/repository/org/slf4j/slf4j-api/1.7.5/slf4j-api-  
1.7.5.jar:/Users/louischang/.m2/repository/org/slf4j/jul-to-slf4j/1.7.5/jul-to-slf4j-  
1.7.5.jar:/Users/louischang/.m2/repository/org/slf4j/jcl-over-slf4j/1.7.5/jcl-over-slf4j-  
1.7.5.jar:/Users/louischang/.m2/repository/log4j/log4j/1.2.17/log4j-  
1.2.17.jar:/Users/louischang/.m2/repository/org/slf4j/slf4j-log4j12/1.7.5/slf4j-log4j12-  
1.7.5.jar:/Users/louischang/.m2/repository/com/ning/compress-lzf/1.0.0/compress-lzf-  
1.0.0.jar:/Users/louischang/.m2/repository/org/xerial/snappy/snappy-java/1.0.5/snappy-  
java-1.0.5.jar:/Users/louischang/.m2/repository/com/twitter/chill\_2.10/0.3.6/chill\_2.10-  
0.3.6.jar:/Users/louischang/.m2/repository/com/esotericsoftware/kryo/kryo/2.21/kryo-  
2.21.jar:/Users/louischang/.m2/repository/com/esotericsoftware/reflectasm/reflectasm/1  
.07/reflectasm-1.07-  
shaded.jar:/Users/louischang/.m2/repository/com/esotericsoftware/minlog/minlog/1.2/mi  
nlog-1.2.jar:/Users/louischang/.m2/repository/org/objenesis/objenesis/1.2/objenesis-  
1.2.jar:/Users/louischang/.m2/repository/com/twitter/chill-java/0.3.6/chill-java-  
0.3.6.jar:/Users/louischang/.m2/repository/commons-net/commons-net/2.2/commons-  
net-2.2.jar:/Users/louischang/.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/louischang/.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/louischang/.m2/repository/com/typesafe/config/1.0.2/config-  
1.0.2.jar:/Users/louischang/.m2/repository/io/netty/netty/3.6.6.Final/netty-  
3.6.6.Final.jar:/Users/louischang/.m2/repository/org/spark-project/protobuf/protobuf-  
java/2.4.1-shaded/protobuf-java-2.4.1-  
shaded.jar:/Users/louischang/.m2/repository/org/uncommons/maths/uncommons-  
maths/1.2.2a/uncommons-maths-1.2.2a.jar:/Users/louischang/.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/louischang/.m2/repository/org/scala-lang/scala-library/2.10.4/scala-  
library-2.10.4.jar:/Users/louischang/.m2/repository/org/json4s/json4s-  
jackson\_2.10/3.2.6/json4s-jackson\_2.10-  
3.2.6.jar:/Users/louischang/.m2/repository/org/json4s/json4s-core\_2.10/3.2.6/json4s-  
core\_2.10-3.2.6.jar:/Users/louischang/.m2/repository/org/json4s/json4s-  
ast\_2.10/3.2.6/json4s-ast\_2.10-  
3.2.6.jar:/Users/louischang/.m2/repository/com/thoughtworks/paranamer/paranamer/2.6/  
paranamer-2.6.jar:/Users/louischang/.m2/repository/org/scala-lang/scalap/2.10.0/scalap-  
2.10.0.jar:/Users/louischang/.m2/repository/org/scala-lang/scala-compiler/2.10.0/scala-  
compiler-2.10.0.jar:/Users/louischang/.m2/repository/org/scala-lang/scala-

reflect/2.10.0/scala-reflect-  
2.10.0.jar:/Users/louischang/.m2/repository/com/fasterxml/jackson/core/jackson-  
databind/2.3.0/jackson-databind-  
2.3.0.jar:/Users/louischang/.m2/repository/com/fasterxml/jackson/core/jackson-  
annotations/2.3.0/jackson-annotations-  
2.3.0.jar:/Users/louischang/.m2/repository/com/fasterxml/jackson/core/jackson-  
core/2.3.0/jackson-core-2.3.0.jar:/Users/louischang/.m2/repository/colt/colt/1.2.0/colt-  
1.2.0.jar:/Users/louischang/.m2/repository/concurrent/concurrent/1.3.4/concurrent-  
1.3.4.jar:/Users/louischang/.m2/repository/org/apache/mesos/mesos/0.18.1/mesos-  
0.18.1-shaded-protobuf.jar:/Users/louischang/.m2/repository/io/netty/netty-  
all/4.0.17.Final/netty-all-  
4.0.17.Final.jar:/Users/louischang/.m2/repository/com/clearspring/analytics/stream/2.5.1  
/stream-2.5.1.jar:/Users/louischang/.m2/repository/com/codahale/metrics/metrics-  
core/3.0.0/metrics-core-  
3.0.0.jar:/Users/louischang/.m2/repository/com/codahale/metrics/metrics-  
jvm/3.0.0/metrics-jvm-  
3.0.0.jar:/Users/louischang/.m2/repository/com/codahale/metrics/metrics-  
json/3.0.0/metrics-json-  
3.0.0.jar:/Users/louischang/.m2/repository/com/codahale/metrics/metrics-  
graphite/3.0.0/metrics-graphite-  
3.0.0.jar:/Users/louischang/.m2/repository/org/tachyonproject/tachyon/0.4.1-  
thrift/tachyon-0.4.1-  
thrift.jar:/Users/louischang/.m2/repository/org/apache/ant/ant/1.9.0/ant-  
1.9.0.jar:/Users/louischang/.m2/repository/org/apache/ant/ant-launcher/1.9.0/ant-  
launcher-1.9.0.jar:/Users/louischang/.m2/repository/commons-io/commons-  
io/2.4/commons-io-2.4.jar:/Users/louischang/.m2/repository/org/spark-  
project/pyrolite/2.0.1/pyrolite-  
2.0.1.jar:/Users/louischang/.m2/repository/net/sf/py4j/py4j/0.8.1/py4j-0.8.1.jar  
WordCounter All'sWellThatEndsWell.txt

objc[98873]: Class JavaLaunchHelper is implemented in both  
/Library/Java/JavaVirtualMachines/jdk1.8.0\_151.jdk/Contents/Home/bin/java  
(0x102ebb4c0) and  
/Library/Java/JavaVirtualMachines/jdk1.8.0\_151.jdk/Contents/Home/jre/lib/libinstrument.  
dylib (0x10ba234e0). One of the two will be used. Which one is undefined.

24/04/20 22:30:02 WARN util.Utils: Your hostname, Louis-MacBook-Pro.local resolves to a  
loopback address: 127.0.0.1; using 172.31.67.229 instead (on interface utun6)

24/04/20 22:30:02 WARN util.Utils: Set SPARK\_LOCAL\_IP if you need to bind to another address

24/04/20 22:30:02 INFO spark.SecurityManager: Changing view acls to: louischang

24/04/20 22:30:02 INFO spark.SecurityManager: SecurityManager: authentication disabled; ui acls disabled; users with view permissions: Set(louischang)

24/04/20 22:30:03 INFO slf4j.Slf4jLogger: Slf4jLogger started

24/04/20 22:30:03 INFO Remoting: Starting remoting

24/04/20 22:30:03 INFO Remoting: Remoting started; listening on addresses :[akka.tcp://spark@vpn-172-31-67-229.vpn.local.cmu.edu:54451]

24/04/20 22:30:03 INFO Remoting: Remoting now listens on addresses: [akka.tcp://spark@vpn-172-31-67-229.vpn.local.cmu.edu:54451]

24/04/20 22:30:03 INFO spark.SparkEnv: Registering MapOutputTracker

24/04/20 22:30:03 INFO spark.SparkEnv: Registering BlockManagerMaster

24/04/20 22:30:03 INFO storage.DiskBlockManager: Created local directory at /var/folders/rq/ytt2\_vp15\_3cmwy08gvnql0c0000gn/T/spark-local-20240420223003-6d1e

24/04/20 22:30:03 INFO storage.MemoryStore: MemoryStore started with capacity 4.8 GB.

24/04/20 22:30:03 INFO network.ConnectionManager: Bound socket to port 54452 with id = ConnectionManagerId(vpn-172-31-67-229.vpn.local.cmu.edu,54452)

24/04/20 22:30:03 INFO storage.BlockManagerMaster: Trying to register BlockManager

24/04/20 22:30:03 INFO storage.BlockManagerInfo: Registering block manager vpn-172-31-67-229.vpn.local.cmu.edu:54452 with 4.8 GB RAM

24/04/20 22:30:03 INFO storage.BlockManagerMaster: Registered BlockManager

24/04/20 22:30:03 INFO spark.HttpServer: Starting HTTP Server

24/04/20 22:30:03 INFO server.Server: jetty-8.1.14.v20131031

24/04/20 22:30:03 INFO server.AbstractConnector: Started SocketConnector@0.0.0.0:54453

24/04/20 22:30:03 INFO broadcast.HttpBroadcast: Broadcast server started at http://172.31.67.229:54453



24/04/20 22:30:03 INFO spark.HttpFileServer: HTTP File server directory is  
/var/folders/rq/ytt2\_vp15\_3cmwy08gvnql0c0000gn/T/spark-70aba6f1-6be9-4c7f-ad73-  
4ce7b3bb8674

24/04/20 22:30:03 INFO spark.HttpServer: Starting HTTP Server

24/04/20 22:30:03 INFO server.Server: jetty-8.1.14.v20131031

24/04/20 22:30:03 INFO server.AbstractConnector: Started  
SocketConnector@0.0.0.0:54454

24/04/20 22:30:03 INFO server.Server: jetty-8.1.14.v20131031

24/04/20 22:30:03 INFO server.AbstractConnector: Started  
SelectChannelConnector@0.0.0.0:4040

24/04/20 22:30:03 INFO ui.SparkUI: Started SparkUI at http://vpn-172-31-67-  
229.vpn.local.cmu.edu:4040

24/04/20 22:30:04 INFO storage.MemoryStore: ensureFreeSpace(32816) called with  
curMem=0, maxMem=5153960755

24/04/20 22:30:04 INFO storage.MemoryStore: Block broadcast\_0 stored as values to  
memory (estimated size 32.0 KB, free 4.8 GB)

24/04/20 22:30:04 WARN util.NativeCodeLoader: Unable to load native-hadoop library for  
your platform... using builtin-java classes where applicable

24/04/20 22:30:04 WARN snappy.LoadSnappy: Snappy native library not loaded

24/04/20 22:30:04 INFO mapred.FileInputFormat: Total input paths to process : 1

24/04/20 22:30:04 INFO spark.SparkContext: Starting job: count at WordCounter.java:20

24/04/20 22:30:04 INFO scheduler.DAGScheduler: Got job 0 (count at  
WordCounter.java:20) with 1 output partitions (allowLocal=false)

24/04/20 22:30:04 INFO scheduler.DAGScheduler: Final stage: Stage 0(count at  
WordCounter.java:20)

24/04/20 22:30:04 INFO scheduler.DAGScheduler: Parents of final stage: List()

24/04/20 22:30:04 INFO scheduler.DAGScheduler: Missing parents: List()

24/04/20 22:30:04 INFO scheduler.DAGScheduler: Submitting Stage 0 (MappedRDD[1] at  
textFile at WordCounter.java:17), which has no missing parents

24/04/20 22:30:04 INFO scheduler.DAGScheduler: Submitting 1 missing tasks from Stage 0 (MappedRDD[1] at textFile at WordCounter.java:17)

24/04/20 22:30:04 INFO scheduler.TaskSchedulerImpl: Adding task set 0.0 with 1 tasks

24/04/20 22:30:04 INFO scheduler.TaskSetManager: Starting task 0.0:0 as TID 0 on executor localhost: localhost (PROCESS\_LOCAL)

24/04/20 22:30:04 INFO scheduler.TaskSetManager: Serialized task 0.0:0 as 1792 bytes in 2 ms

24/04/20 22:30:04 INFO executor.Executor: Running task ID 0

24/04/20 22:30:04 INFO storage.BlockManager: Found block broadcast\_0 locally

24/04/20 22:30:04 INFO rdd.HadoopRDD: Input split:  
file:/Users/louischang/Library/CloudStorage/OneDrive-  
andrew.cmu.edu/DS/Project5/Project5Part2/All'sWellThatEndsWell.txt:0+133885

24/04/20 22:30:04 INFO executor.Executor: Serialized size of result for 0 is 597

24/04/20 22:30:04 INFO executor.Executor: Sending result for 0 directly to driver

24/04/20 22:30:04 INFO executor.Executor: Finished task ID 0

24/04/20 22:30:04 INFO scheduler.TaskSetManager: Finished TID 0 in 74 ms on localhost (progress: 1/1)

24/04/20 22:30:04 INFO scheduler.DAGScheduler: Completed ResultTask(0, 0)

24/04/20 22:30:04 INFO scheduler.TaskSchedulerImpl: Removed TaskSet 0.0, whose tasks have all completed, from pool

24/04/20 22:30:04 INFO scheduler.DAGScheduler: Stage 0 (count at WordCounter.java:20) finished in 0.083 s

24/04/20 22:30:04 INFO spark.SparkContext: Job finished: count at WordCounter.java:20, took 0.148292041 s

24/04/20 22:30:04 INFO spark.SparkContext: Starting job: count at WordCounter.java:28

24/04/20 22:30:04 INFO scheduler.DAGScheduler: Got job 1 (count at WordCounter.java:28) with 1 output partitions (allowLocal=false)

24/04/20 22:30:04 INFO scheduler.DAGScheduler: Final stage: Stage 1(count at WordCounter.java:28)

24/04/20 22:30:04 INFO scheduler.DAGScheduler: Parents of final stage: List()

24/04/20 22:30:04 INFO scheduler.DAGScheduler: Missing parents: List()

24/04/20 22:30:04 INFO scheduler.DAGScheduler: Submitting Stage 1 (FilteredRDD[3] at filter at WordCounter.java:25), which has no missing parents

24/04/20 22:30:04 INFO scheduler.DAGScheduler: Submitting 1 missing tasks from Stage 1 (FilteredRDD[3] at filter at WordCounter.java:25)

24/04/20 22:30:04 INFO scheduler.TaskSchedulerImpl: Adding task set 1.0 with 1 tasks

24/04/20 22:30:04 INFO scheduler.TaskSetManager: Starting task 1.0:0 as TID 1 on executor localhost: localhost (PROCESS\_LOCAL)

24/04/20 22:30:04 INFO scheduler.TaskSetManager: Serialized task 1.0:0 as 2379 bytes in 1 ms

24/04/20 22:30:04 INFO executor.Executor: Running task ID 1

24/04/20 22:30:04 INFO storage.BlockManager: Found block broadcast\_0 locally

24/04/20 22:30:04 INFO rdd.HadoopRDD: Input split:  
file:/Users/louischang/Library/CloudStorage/OneDrive-andrew.cmu.edu/DS/Project5/Project5Part2/All'sWellThatEndsWell.txt:0+133885

24/04/20 22:30:04 INFO executor.Executor: Serialized size of result for 1 is 597

24/04/20 22:30:04 INFO executor.Executor: Sending result for 1 directly to driver

24/04/20 22:30:04 INFO executor.Executor: Finished task ID 1

24/04/20 22:30:04 INFO scheduler.TaskSetManager: Finished TID 1 in 99 ms on localhost (progress: 1/1)

24/04/20 22:30:04 INFO scheduler.TaskSchedulerImpl: Removed TaskSet 1.0, whose tasks have all completed, from pool

24/04/20 22:30:04 INFO scheduler.DAGScheduler: Completed ResultTask(1, 0)

24/04/20 22:30:04 INFO scheduler.DAGScheduler: Stage 1 (count at WordCounter.java:28) finished in 0.102 s

24/04/20 22:30:04 INFO spark.SparkContext: Job finished: count at WordCounter.java:28, took 0.120541583 s

24/04/20 22:30:04 INFO spark.SparkContext: Starting job: count at WordCounter.java:35

24/04/20 22:30:04 INFO scheduler.DAGScheduler: Registering RDD 5 (reduceByKey at WordCounter.java:32)

24/04/20 22:30:04 INFO scheduler.DAGScheduler: Registering RDD 10 (distinct at WordCounter.java:35)

24/04/20 22:30:04 INFO scheduler.DAGScheduler: Got job 2 (count at WordCounter.java:35) with 1 output partitions (allowLocal=false)

24/04/20 22:30:04 INFO scheduler.DAGScheduler: Final stage: Stage 2(count at WordCounter.java:35)

24/04/20 22:30:04 INFO scheduler.DAGScheduler: Parents of final stage: List(Stage 3)

24/04/20 22:30:04 INFO scheduler.DAGScheduler: Missing parents: List(Stage 3)

24/04/20 22:30:04 INFO scheduler.DAGScheduler: Submitting Stage 4 (MapPartitionsRDD[5] at reduceByKey at WordCounter.java:32), which has no missing parents

24/04/20 22:30:04 INFO scheduler.DAGScheduler: Submitting 1 missing tasks from Stage 4 (MapPartitionsRDD[5] at reduceByKey at WordCounter.java:32)

24/04/20 22:30:04 INFO scheduler.TaskSchedulerImpl: Adding task set 4.0 with 1 tasks

24/04/20 22:30:04 INFO scheduler.TaskSetManager: Starting task 4.0:0 as TID 2 on executor localhost: localhost (PROCESS\_LOCAL)

24/04/20 22:30:04 INFO scheduler.TaskSetManager: Serialized task 4.0:0 as 2806 bytes in 0 ms

24/04/20 22:30:04 INFO executor.Executor: Running task ID 2

24/04/20 22:30:04 INFO storage.BlockManager: Found block broadcast\_0 locally

24/04/20 22:30:04 INFO rdd.HadoopRDD: Input split:  
file:/Users/louischang/Library/CloudStorage/OneDrive-andrew.cmu.edu/DS/Project5/Project5Part2/All'sWellThatEndsWell.txt:0+133885

24/04/20 22:30:04 INFO executor.Executor: Serialized size of result for 2 is 805

24/04/20 22:30:04 INFO executor.Executor: Sending result for 2 directly to driver

24/04/20 22:30:04 INFO executor.Executor: Finished task ID 2

24/04/20 22:30:04 INFO scheduler.TaskSetManager: Finished TID 2 in 288 ms on localhost (progress: 1/1)

24/04/20 22:30:04 INFO scheduler.TaskSchedulerImpl: Removed TaskSet 4.0, whose tasks have all completed, from pool

24/04/20 22:30:04 INFO scheduler.DAGScheduler: Completed ShuffleMapTask(4, 0)

24/04/20 22:30:04 INFO scheduler.DAGScheduler: Stage 4 (reduceByKey at WordCounter.java:32) finished in 0.294 s

24/04/20 22:30:04 INFO scheduler.DAGScheduler: looking for newly runnable stages

24/04/20 22:30:04 INFO scheduler.DAGScheduler: running: Set()

24/04/20 22:30:04 INFO scheduler.DAGScheduler: waiting: Set(Stage 2, Stage 3)

24/04/20 22:30:04 INFO scheduler.DAGScheduler: failed: Set()

24/04/20 22:30:04 INFO scheduler.DAGScheduler: Missing parents for Stage 2: List(Stage 3)

24/04/20 22:30:04 INFO scheduler.DAGScheduler: Missing parents for Stage 3: List()

24/04/20 22:30:04 INFO scheduler.DAGScheduler: Submitting Stage 3 (MapPartitionsRDD[10] at distinct at WordCounter.java:35), which is now runnable

24/04/20 22:30:04 INFO scheduler.DAGScheduler: Submitting 1 missing tasks from Stage 3 (MapPartitionsRDD[10] at distinct at WordCounter.java:35)

24/04/20 22:30:04 INFO scheduler.TaskSchedulerImpl: Adding task set 3.0 with 1 tasks

24/04/20 22:30:04 INFO scheduler.TaskSetManager: Starting task 3.0:0 as TID 3 on executor localhost: localhost (PROCESS\_LOCAL)

24/04/20 22:30:04 INFO scheduler.TaskSetManager: Serialized task 3.0:0 as 2647 bytes in 0 ms

24/04/20 22:30:04 INFO executor.Executor: Running task ID 3

24/04/20 22:30:04 INFO storage.BlockManager: Found block broadcast\_0 locally

24/04/20 22:30:04 INFO storage.BlockFetcherIterator\$BasicBlockFetcherIterator: maxBytesInFlight: 50331648, targetRequestSize: 10066329

24/04/20 22:30:04 INFO storage.BlockFetcherIterator\$BasicBlockFetcherIterator: Getting 1 non-empty blocks out of 1 blocks

24/04/20 22:30:04 INFO storage.BlockFetcherIterator\$BasicBlockFetcherIterator: Started 0 remote fetches in 2 ms

24/04/20 22:30:05 INFO executor.Executor: Serialized size of result for 3 is 1021

24/04/20 22:30:05 INFO executor.Executor: Sending result for 3 directly to driver

24/04/20 22:30:05 INFO executor.Executor: Finished task ID 3

24/04/20 22:30:05 INFO scheduler.DAGScheduler: Completed ShuffleMapTask(3, 0)

24/04/20 22:30:05 INFO scheduler.TaskSetManager: Finished TID 3 in 187 ms on localhost (progress: 1/1)

24/04/20 22:30:05 INFO scheduler.TaskSchedulerImpl: Removed TaskSet 3.0, whose tasks have all completed, from pool

24/04/20 22:30:05 INFO scheduler.DAGScheduler: Stage 3 (distinct at WordCounter.java:35) finished in 0.192 s

24/04/20 22:30:05 INFO scheduler.DAGScheduler: looking for newly runnable stages

24/04/20 22:30:05 INFO scheduler.DAGScheduler: running: Set()

24/04/20 22:30:05 INFO scheduler.DAGScheduler: waiting: Set(Stage 2)

24/04/20 22:30:05 INFO scheduler.DAGScheduler: failed: Set()

24/04/20 22:30:05 INFO scheduler.DAGScheduler: Missing parents for Stage 2: List()

24/04/20 22:30:05 INFO scheduler.DAGScheduler: Submitting Stage 2 (MappedRDD[13] at distinct at WordCounter.java:35), which is now runnable

24/04/20 22:30:05 INFO scheduler.DAGScheduler: Submitting 1 missing tasks from Stage 2 (MappedRDD[13] at distinct at WordCounter.java:35)

24/04/20 22:30:05 INFO scheduler.TaskSchedulerImpl: Adding task set 2.0 with 1 tasks

24/04/20 22:30:05 INFO scheduler.TaskSetManager: Starting task 2.0:0 as TID 4 on executor localhost: localhost (PROCESS\_LOCAL)

24/04/20 22:30:05 INFO scheduler.TaskSetManager: Serialized task 2.0:0 as 2710 bytes in 2 ms

24/04/20 22:30:05 INFO executor.Executor: Running task ID 4

24/04/20 22:30:05 INFO storage.BlockManager: Found block broadcast\_0 locally

24/04/20 22:30:05 INFO storage.BlockFetcherIterator\$BasicBlockFetcherIterator: maxBytesInFlight: 50331648, targetRequestSize: 10066329

24/04/20 22:30:05 INFO storage.BlockFetcherIterator\$BasicBlockFetcherIterator: Getting 1 non-empty blocks out of 1 blocks

24/04/20 22:30:05 INFO storage.BlockFetcherIterator\$BasicBlockFetcherIterator: Started 0 remote fetches in 0 ms

24/04/20 22:30:05 INFO executor.Executor: Serialized size of result for 4 is 863

24/04/20 22:30:05 INFO executor.Executor: Sending result for 4 directly to driver

24/04/20 22:30:05 INFO executor.Executor: Finished task ID 4

24/04/20 22:30:05 INFO scheduler.TaskSetManager: Finished TID 4 in 58 ms on localhost (progress: 1/1)

24/04/20 22:30:05 INFO scheduler.DAGScheduler: Completed ResultTask(2, 0)

24/04/20 22:30:05 INFO scheduler.TaskSchedulerImpl: Removed TaskSet 2.0, whose tasks have all completed, from pool

24/04/20 22:30:05 INFO scheduler.DAGScheduler: Stage 2 (count at WordCounter.java:35) finished in 0.059 s

24/04/20 22:30:05 INFO spark.SparkContext: Job finished: count at WordCounter.java:35, took 0.634305333 s

24/04/20 22:30:05 INFO spark.SparkContext: Starting job: count at WordCounter.java:41

24/04/20 22:30:05 INFO scheduler.DAGScheduler: Got job 3 (count at WordCounter.java:41) with 1 output partitions (allowLocal=false)

24/04/20 22:30:05 INFO scheduler.DAGScheduler: Final stage: Stage 5(count at WordCounter.java:41)

24/04/20 22:30:05 INFO scheduler.DAGScheduler: Parents of final stage: List()

24/04/20 22:30:05 INFO scheduler.DAGScheduler: Missing parents: List()

24/04/20 22:30:05 INFO scheduler.DAGScheduler: Submitting Stage 5 (FlatMappedRDD[14] at flatMap at WordCounter.java:38), which has no missing parents

24/04/20 22:30:05 INFO scheduler.DAGScheduler: Submitting 1 missing tasks from Stage 5 (FlatMappedRDD[14] at flatMap at WordCounter.java:38)

24/04/20 22:30:05 INFO scheduler.TaskSchedulerImpl: Adding task set 5.0 with 1 tasks

24/04/20 22:30:05 INFO scheduler.TaskSetManager: Starting task 5.0:0 as TID 5 on executor localhost: localhost (PROCESS\_LOCAL)

24/04/20 22:30:05 INFO scheduler.TaskSetManager: Serialized task 5.0:0 as 2257 bytes in 0 ms

24/04/20 22:30:05 INFO executor.Executor: Running task ID 5

24/04/20 22:30:05 INFO storage.BlockManager: Found block broadcast\_0 locally

24/04/20 22:30:05 INFO rdd.HadoopRDD: Input split:  
file:/Users/louischang/Library/CloudStorage/OneDrive-andrew.cmu.edu/DS/Project5/Project5Part2/All'sWellThatEndsWell.txt:0+133885

24/04/20 22:30:05 INFO executor.Executor: Serialized size of result for 5 is 597

24/04/20 22:30:05 INFO executor.Executor: Sending result for 5 directly to driver

24/04/20 22:30:05 INFO executor.Executor: Finished task ID 5

24/04/20 22:30:05 INFO scheduler.TaskSetManager: Finished TID 5 in 77 ms on localhost (progress: 1/1)

24/04/20 22:30:05 INFO scheduler.TaskSchedulerImpl: Removed TaskSet 5.0, whose tasks have all completed, from pool

24/04/20 22:30:05 INFO scheduler.DAGScheduler: Completed ResultTask(5, 0)

24/04/20 22:30:05 INFO scheduler.DAGScheduler: Stage 5 (count at WordCounter.java:41) finished in 0.080 s

24/04/20 22:30:05 INFO spark.SparkContext: Job finished: count at WordCounter.java:41, took 0.093885416 s

24/04/20 22:30:05 INFO spark.SparkContext: Starting job: count at WordCounter.java:44

24/04/20 22:30:05 INFO scheduler.DAGScheduler: Registering RDD 16 (distinct at WordCounter.java:44)

24/04/20 22:30:05 INFO scheduler.DAGScheduler: Got job 4 (count at WordCounter.java:44) with 1 output partitions (allowLocal=false)

24/04/20 22:30:05 INFO scheduler.DAGScheduler: Final stage: Stage 6(count at WordCounter.java:44)

24/04/20 22:30:05 INFO scheduler.DAGScheduler: Parents of final stage: List(Stage 7)



24/04/20 22:30:05 INFO scheduler.DAGScheduler: Missing parents: List(Stage 7)

24/04/20 22:30:05 INFO scheduler.DAGScheduler: Submitting Stage 7  
(MapPartitionsRDD[16] at distinct at WordCounter.java:44), which has no missing parents

24/04/20 22:30:05 INFO scheduler.DAGScheduler: Submitting 1 missing tasks from Stage 7  
(MapPartitionsRDD[16] at distinct at WordCounter.java:44)

24/04/20 22:30:05 INFO scheduler.TaskSchedulerImpl: Adding task set 7.0 with 1 tasks

24/04/20 22:30:05 INFO scheduler.TaskSetManager: Starting task 7.0:0 as TID 6 on  
executor localhost: localhost (PROCESS\_LOCAL)

24/04/20 22:30:05 INFO scheduler.TaskSetManager: Serialized task 7.0:0 as 2562 bytes in 1  
ms

24/04/20 22:30:05 INFO executor.Executor: Running task ID 6

24/04/20 22:30:05 INFO storage.BlockManager: Found block broadcast\_0 locally

24/04/20 22:30:05 INFO rdd.HadoopRDD: Input split:  
file:/Users/louischang/Library/CloudStorage/OneDrive-  
andrew.cmu.edu/DS/Project5/Project5Part2/All'sWellThatEndsWell.txt:0+133885

24/04/20 22:30:05 INFO executor.Executor: Serialized size of result for 6 is 805

24/04/20 22:30:05 INFO executor.Executor: Sending result for 6 directly to driver

24/04/20 22:30:05 INFO executor.Executor: Finished task ID 6

24/04/20 22:30:05 INFO scheduler.TaskSetManager: Finished TID 6 in 73 ms on localhost  
(progress: 1/1)

24/04/20 22:30:05 INFO scheduler.TaskSchedulerImpl: Removed TaskSet 7.0, whose tasks  
have all completed, from pool

24/04/20 22:30:05 INFO scheduler.DAGScheduler: Completed ShuffleMapTask(7, 0)

24/04/20 22:30:05 INFO scheduler.DAGScheduler: Stage 7 (distinct at  
WordCounter.java:44) finished in 0.074 s

24/04/20 22:30:05 INFO scheduler.DAGScheduler: looking for newly runnable stages

24/04/20 22:30:05 INFO scheduler.DAGScheduler: running: Set()

24/04/20 22:30:05 INFO scheduler.DAGScheduler: waiting: Set(Stage 6)

24/04/20 22:30:05 INFO scheduler.DAGScheduler: failed: Set()

24/04/20 22:30:05 INFO scheduler.DAGScheduler: Missing parents for Stage 6: List()  
24/04/20 22:30:05 INFO scheduler.DAGScheduler: Submitting Stage 6 (MappedRDD[19] at distinct at WordCounter.java:44), which is now runnable  
24/04/20 22:30:05 INFO scheduler.DAGScheduler: Submitting 1 missing tasks from Stage 6 (MappedRDD[19] at distinct at WordCounter.java:44)  
24/04/20 22:30:05 INFO scheduler.TaskSchedulerImpl: Adding task set 6.0 with 1 tasks  
24/04/20 22:30:05 INFO scheduler.TaskSetManager: Starting task 6.0:0 as TID 7 on executor localhost: localhost (PROCESS\_LOCAL)  
24/04/20 22:30:05 INFO scheduler.TaskSetManager: Serialized task 6.0:0 as 2373 bytes in 0 ms  
24/04/20 22:30:05 INFO executor.Executor: Running task ID 7  
24/04/20 22:30:05 INFO storage.BlockManager: Found block broadcast\_0 locally  
24/04/20 22:30:05 INFO storage.BlockFetcherIterator\$BasicBlockFetcherIterator: maxBytesInFlight: 50331648, targetRequestSize: 10066329  
24/04/20 22:30:05 INFO storage.BlockFetcherIterator\$BasicBlockFetcherIterator: Getting 1 non-empty blocks out of 1 blocks  
24/04/20 22:30:05 INFO storage.BlockFetcherIterator\$BasicBlockFetcherIterator: Started 0 remote fetches in 0 ms  
24/04/20 22:30:05 INFO executor.Executor: Serialized size of result for 7 is 863  
24/04/20 22:30:05 INFO executor.Executor: Sending result for 7 directly to driver  
24/04/20 22:30:05 INFO executor.Executor: Finished task ID 7  
24/04/20 22:30:05 INFO scheduler.DAGScheduler: Completed ResultTask(6, 0)  
24/04/20 22:30:05 INFO scheduler.TaskSetManager: Finished TID 7 in 10 ms on localhost (progress: 1/1)  
24/04/20 22:30:05 INFO scheduler.TaskSchedulerImpl: Removed TaskSet 6.0, whose tasks have all completed, from pool  
24/04/20 22:30:05 INFO scheduler.DAGScheduler: Stage 6 (count at WordCounter.java:44) finished in 0.011 s

24/04/20 22:30:05 INFO spark.SparkContext: Job finished: count at WordCounter.java:44, took 0.099526958 s

24/04/20 22:30:05 INFO spark.SparkContext: Starting job: count at WordCounter.java:48

24/04/20 22:30:05 INFO scheduler.DAGScheduler: Registering RDD 22 (distinct at WordCounter.java:48)

24/04/20 22:30:05 INFO scheduler.DAGScheduler: Got job 5 (count at WordCounter.java:48) with 1 output partitions (allowLocal=false)

24/04/20 22:30:05 INFO scheduler.DAGScheduler: Final stage: Stage 8(count at WordCounter.java:48)

24/04/20 22:30:05 INFO scheduler.DAGScheduler: Parents of final stage: List(Stage 9)

24/04/20 22:30:05 INFO scheduler.DAGScheduler: Missing parents: List(Stage 9)

24/04/20 22:30:05 INFO scheduler.DAGScheduler: Submitting Stage 9 (MapPartitionsRDD[22] at distinct at WordCounter.java:48), which has no missing parents

24/04/20 22:30:05 INFO scheduler.DAGScheduler: Submitting 1 missing tasks from Stage 9 (MapPartitionsRDD[22] at distinct at WordCounter.java:48)

24/04/20 22:30:05 INFO scheduler.TaskSchedulerImpl: Adding task set 9.0 with 1 tasks

24/04/20 22:30:05 INFO scheduler.TaskSetManager: Starting task 9.0:0 as TID 8 on executor localhost: localhost (PROCESS\_LOCAL)

24/04/20 22:30:05 INFO scheduler.TaskSetManager: Serialized task 9.0:0 as 2681 bytes in 0 ms

24/04/20 22:30:05 INFO executor.Executor: Running task ID 8

24/04/20 22:30:05 INFO storage.BlockManager: Found block broadcast\_0 locally

24/04/20 22:30:05 INFO rdd.HadoopRDD: Input split:  
file:/Users/louischang/Library/CloudStorage/OneDrive-andrew.cmu.edu/DS/Project5/Project5Part2/All'sWellThatEndsWell.txt:0+133885

24/04/20 22:30:05 INFO storage.ShuffleBlockManager: Deleted all files for shuffle 2

24/04/20 22:30:05 INFO spark.ContextCleaner: Cleaned shuffle 2

24/04/20 22:30:05 INFO storage.ShuffleBlockManager: Deleted all files for shuffle 0

24/04/20 22:30:05 INFO spark.ContextCleaner: Cleaned shuffle 0

24/04/20 22:30:05 INFO executor.Executor: Serialized size of result for 8 is 805

24/04/20 22:30:05 INFO executor.Executor: Sending result for 8 directly to driver

24/04/20 22:30:05 INFO executor.Executor: Finished task ID 8

24/04/20 22:30:05 INFO scheduler.DAGScheduler: Completed ShuffleMapTask(9, 0)

24/04/20 22:30:05 INFO scheduler.TaskSetManager: Finished TID 8 in 153 ms on localhost (progress: 1/1)

24/04/20 22:30:05 INFO scheduler.TaskSchedulerImpl: Removed TaskSet 9.0, whose tasks have all completed, from pool

24/04/20 22:30:05 INFO scheduler.DAGScheduler: Stage 9 (distinct at WordCounter.java:48) finished in 0.156 s

24/04/20 22:30:05 INFO scheduler.DAGScheduler: looking for newly runnable stages

24/04/20 22:30:05 INFO scheduler.DAGScheduler: running: Set()

24/04/20 22:30:05 INFO scheduler.DAGScheduler: waiting: Set(Stage 8)

24/04/20 22:30:05 INFO scheduler.DAGScheduler: failed: Set()

24/04/20 22:30:05 INFO scheduler.DAGScheduler: Missing parents for Stage 8: List()

24/04/20 22:30:05 INFO scheduler.DAGScheduler: Submitting Stage 8 (MappedRDD[25] at distinct at WordCounter.java:48), which is now runnable

24/04/20 22:30:05 INFO scheduler.DAGScheduler: Submitting 1 missing tasks from Stage 8 (MappedRDD[25] at distinct at WordCounter.java:48)

24/04/20 22:30:05 INFO scheduler.TaskSchedulerImpl: Adding task set 8.0 with 1 tasks

24/04/20 22:30:05 INFO scheduler.TaskSetManager: Starting task 8.0:0 as TID 9 on executor localhost: localhost (PROCESS\_LOCAL)

24/04/20 22:30:05 INFO scheduler.TaskSetManager: Serialized task 8.0:0 as 2482 bytes in 0 ms

24/04/20 22:30:05 INFO executor.Executor: Running task ID 9

24/04/20 22:30:05 INFO storage.BlockManager: Found block broadcast\_0 locally

24/04/20 22:30:05 INFO storage.BlockFetcherIterator\$BasicBlockFetcherIterator: maxBytesInFlight: 50331648, targetRequestSize: 10066329

24/04/20 22:30:05 INFO storage.BlockFetcherIterator\$BasicBlockFetcherIterator: Getting 1 non-empty blocks out of 1 blocks

24/04/20 22:30:05 INFO storage.BlockFetcherIterator\$BasicBlockFetcherIterator: Started 0 remote fetches in 1 ms

24/04/20 22:30:05 INFO executor.Executor: Serialized size of result for 9 is 863

24/04/20 22:30:05 INFO executor.Executor: Sending result for 9 directly to driver

24/04/20 22:30:05 INFO executor.Executor: Finished task ID 9

24/04/20 22:30:05 INFO scheduler.TaskSetManager: Finished TID 9 in 8 ms on localhost (progress: 1/1)

24/04/20 22:30:05 INFO scheduler.DAGScheduler: Completed ResultTask(8, 0)

24/04/20 22:30:05 INFO scheduler.TaskSchedulerImpl: Removed TaskSet 8.0, whose tasks have all completed, from pool

24/04/20 22:30:05 INFO scheduler.DAGScheduler: Stage 8 (count at WordCounter.java:48) finished in 0.011 s

24/04/20 22:30:05 INFO spark.SparkContext: Job finished: count at WordCounter.java:48, took 0.1799065 s

Number of lines: 4570

Number of words: 25085

Number of distinct words: 3866

Number of symbols: 130373

Number of distinct symbols: 78

Number of distinct letters: 50

Process finished with exit code 0

## Task 6

```
import org.apache.spark.SparkConf;
import org.apache.spark.api.java.JavaRDD;
import org.apache.spark.api.java.JavaSparkContext;

import java.util.List;
import java.util.Scanner;
```

```

public class WordFinder {

    private static void findLinesWithWord(String fileName) {
        SparkConf sparkConf = new
SparkConf().setMaster("local").setAppName("Line Finder");
        JavaSparkContext sparkContext = new JavaSparkContext(sparkConf);

        // Read the file into an RDD
        JavaRDD<String> inputFile = sparkContext.textFile(fileName);

        Scanner scanner = new Scanner(System.in);
        System.out.println("Please enter a word:");
        String word = scanner.nextLine();
        scanner.close();

        // Filter lines that contain the specified word
        JavaRDD<String> linesWithWord = inputFile.filter(line ->
line.contains(word));

        // Collect and print lines
        List<String> resultLines = linesWithWord.collect();
        if (resultLines.isEmpty()) {
            System.out.println("No lines found containing the word: " +
word);
        } else {
            System.out.println("Lines containing the word '" + word +
"':");
            resultLines.forEach(System.out::println);
        }
    }

    public static void main(String[] args) {
        if (args.length == 0) {
            System.out.println("No file provided.");
            System.exit(0);
        }

        findLinesWithWord(args[0]);
    }
}

```

/Library/Java/JavaVirtualMachines/jdk1.8.0\_151.jdk/Contents/Home/bin/java -  
javaagent:/Users/louischang/Applications/IntelliJ IDEA  
Ultimate.app/Contents/lib/idea\_rt.jar=54918:/Users/louischang/Applications/IntelliJ IDEA  
Ultimate.app/Contents/bin -Dfile.encoding=UTF-8 -classpath  
/Library/Java/JavaVirtualMachines/jdk1.8.0\_151.jdk/Contents/Home/jre/lib/charsets.jar:/Li  
brary/Java/JavaVirtualMachines/jdk1.8.0\_151.jdk/Contents/Home/jre/lib/deploy.jar:/Librar  
y/Java/JavaVirtualMachines/jdk1.8.0\_151.jdk/Contents/Home/jre/lib/ext/cldrdata.jar:/Libra

ry/Java/JavaVirtualMachines/jdk1.8.0\_151.jdk/Contents/Home/jre/lib/ext/dnsns.jar:/Library/Java/JavaVirtualMachines/jdk1.8.0\_151.jdk/Contents/Home/jre/lib/ext/jaccess.jar:/Library/Java/JavaVirtualMachines/jdk1.8.0\_151.jdk/Contents/Home/jre/lib/ext/jfxrt.jar:/Library/Java/JavaVirtualMachines/jdk1.8.0\_151.jdk/Contents/Home/jre/lib/ext/localedata.jar:/Library/Java/JavaVirtualMachines/jdk1.8.0\_151.jdk/Contents/Home/jre/lib/ext/nashorn.jar:/Library/Java/JavaVirtualMachines/jdk1.8.0\_151.jdk/Contents/Home/jre/lib/ext/sunec.jar:/Library/Java/JavaVirtualMachines/jdk1.8.0\_151.jdk/Contents/Home/jre/lib/ext/sunjce\_provider.jar:/Library/Java/JavaVirtualMachines/jdk1.8.0\_151.jdk/Contents/Home/jre/lib/ext/sunpkcs11.jar:/Library/Java/JavaVirtualMachines/jdk1.8.0\_151.jdk/Contents/Home/jre/lib/ext/zipfs.jar:/Library/Java/JavaVirtualMachines/jdk1.8.0\_151.jdk/Contents/Home/jre/lib/javaws.jar:/Library/Java/JavaVirtualMachines/jdk1.8.0\_151.jdk/Contents/Home/jre/lib/jce.jar:/Library/Java/JavaVirtualMachines/jdk1.8.0\_151.jdk/Contents/Home/jre/lib/jfr.jar:/Library/Java/JavaVirtualMachines/jdk1.8.0\_151.jdk/Contents/Home/jre/lib/jfxswt.jar:/Library/Java/JavaVirtualMachines/jdk1.8.0\_151.jdk/Contents/Home/jre/lib/jsse.jar:/Library/Java/JavaVirtualMachines/jdk1.8.0\_151.jdk/Contents/Home/jre/lib/management-agent.jar:/Library/Java/JavaVirtualMachines/jdk1.8.0\_151.jdk/Contents/Home/jre/lib/plugin.jar:/Library/Java/JavaVirtualMachines/jdk1.8.0\_151.jdk/Contents/Home/jre/lib/resources.jar:/Library/Java/JavaVirtualMachines/jdk1.8.0\_151.jdk/Contents/Home/jre/lib/rt.jar:/Library/Java/JavaVirtualMachines/jdk1.8.0\_151.jdk/Contents/Home/lib/ant-javafx.jar:/Library/Java/JavaVirtualMachines/jdk1.8.0\_151.jdk/Contents/Home/lib/dt.jar:/Library/Java/JavaVirtualMachines/jdk1.8.0\_151.jdk/Contents/Home/lib/javafx-mx.jar:/Library/Java/JavaVirtualMachines/jdk1.8.0\_151.jdk/Contents/Home/lib/jconsole.jar:/Library/Java/JavaVirtualMachines/jdk1.8.0\_151.jdk/Contents/Home/lib/packager.jar:/Library/Java/JavaVirtualMachines/jdk1.8.0\_151.jdk/Contents/Home/lib/sa-jdi.jar:/Library/Java/JavaVirtualMachines/jdk1.8.0\_151.jdk/Contents/Home/lib/tools.jar:/Users/louischang/Library/CloudStorage/OneDrive-andrew.cmu.edu/DS/Project5/Project5Part2/target/classes:/Users/louischang/.m2/repository/org/apache/spark/spark-core\_2.10/1.0.0/spark-core\_2.10-1.0.0.jar:/Users/louischang/.m2/repository/org/apache/hadoop/hadoop-client/1.0.4/hadoop-client-1.0.4.jar:/Users/louischang/.m2/repository/org/apache/hadoop/hadoop-core/1.0.4/hadoop-core-1.0.4.jar:/Users/louischang/.m2/repository/xmlenc/xmlenc/0.52/xmlenc-0.52.jar:/Users/louischang/.m2/repository/org/apache/commons/commons-math/2.1/commons-math-2.1.jar:/Users/louischang/.m2/repository/commons-configuration/commons-configuration/1.6/commons-configuration-1.6.jar:/Users/louischang/.m2/repository/commons-collections/commons-collections/3.2.1/commons-collections-

3.2.1.jar:/Users/louischang/.m2/repository/commons-lang/commons-lang/2.4/commons-lang-2.4.jar:/Users/louischang/.m2/repository/commons-digester/commons-digester/1.8/commons-digester-1.8.jar:/Users/louischang/.m2/repository/commons-beanutils/commons-beanutils/1.7.0/commons-beanutils-1.7.0.jar:/Users/louischang/.m2/repository/commons-beanutils/commons-beanutils-core/1.8.0/commons-beanutils-core-1.8.0.jar:/Users/louischang/.m2/repository/commons-el/commons-el/1.0/commons-el-1.0.jar:/Users/louischang/.m2/repository/hsqldb/hsqldb/1.8.0.10/hsqldb-1.8.0.10.jar:/Users/louischang/.m2/repository/oro/oro/2.0.8/oro-2.0.8.jar:/Users/louischang/.m2/repository/org/codehaus/jackson/jackson-mapper-asl/1.0.1/jackson-mapper-asl-1.0.1.jar:/Users/louischang/.m2/repository/org/codehaus/jackson/jackson-core-asl/1.0.1/jackson-core-asl-1.0.1.jar:/Users/louischang/.m2/repository/net/java/dev/jets3t/jets3t/0.7.1/jets3t-0.7.1.jar:/Users/louischang/.m2/repository/commons-codec/commons-codec/1.3/commons-codec-1.3.jar:/Users/louischang/.m2/repository/commons-httpclient/commons-httpclient/3.1/commons-httpclient-3.1.jar:/Users/louischang/.m2/repository/org/apache/curator/curator-recipes/2.4.0/curator-recipes-2.4.0.jar:/Users/louischang/.m2/repository/org/apache/curator/curator-framework/2.4.0/curator-framework-2.4.0.jar:/Users/louischang/.m2/repository/org/apache/curator/curator-client/2.4.0/curator-client-2.4.0.jar:/Users/louischang/.m2/repository/org/apache/zookeeper/zookeeper/3.4.5/zookeeper-3.4.5.jar:/Users/louischang/.m2/repository/jline/jline/0.9.94/jline-0.9.94.jar:/Users/louischang/.m2/repository/org/eclipse/jetty/jetty-plus/8.1.14.v20131031/jetty-plus-8.1.14.v20131031.jar:/Users/louischang/.m2/repository/org/eclipse/jetty/orbit/javax.transaction/1.1.1.v201105210645/javax.transaction-1.1.1.v201105210645.jar:/Users/louischang/.m2/repository/org/eclipse/jetty/jetty-webapp/8.1.14.v20131031/jetty-webapp-8.1.14.v20131031.jar:/Users/louischang/.m2/repository/org/eclipse/jetty/jetty-xml/8.1.14.v20131031/jetty-xml-8.1.14.v20131031.jar:/Users/louischang/.m2/repository/org/eclipse/jetty/jetty-servlet/8.1.14.v20131031/jetty-servlet-8.1.14.v20131031.jar:/Users/louischang/.m2/repository/org/eclipse/jetty/jetty-jndi/8.1.14.v20131031/jetty-jndi-8.1.14.v20131031.jar:/Users/louischang/.m2/repository/org/eclipse/jetty/orbit/javax.mail.g



lassfish/1.4.1.v201005082020/javax.mail.glassfish-  
1.4.1.v201005082020.jar:/Users/louischang/.m2/repository/org/eclipse/jetty/orbit/javax.ac  
tivation/1.1.0.v201105071233/javax.activation-  
1.1.0.v201105071233.jar:/Users/louischang/.m2/repository/org/eclipse/jetty/jetty-  
security/8.1.14.v20131031/jetty-security-  
8.1.14.v20131031.jar:/Users/louischang/.m2/repository/org/eclipse/jetty/jetty-  
util/8.1.14.v20131031/jetty-util-  
8.1.14.v20131031.jar:/Users/louischang/.m2/repository/org/eclipse/jetty/jetty-  
server/8.1.14.v20131031/jetty-server-  
8.1.14.v20131031.jar:/Users/louischang/.m2/repository/org/eclipse/jetty/orbit/javax.servle  
t/3.0.0.v201112011016/javax.servlet-  
3.0.0.v201112011016.jar:/Users/louischang/.m2/repository/org/eclipse/jetty/jetty-  
continuation/8.1.14.v20131031/jetty-continuation-  
8.1.14.v20131031.jar:/Users/louischang/.m2/repository/org/eclipse/jetty/jetty-  
http/8.1.14.v20131031/jetty-http-  
8.1.14.v20131031.jar:/Users/louischang/.m2/repository/org/eclipse/jetty/jetty-  
io/8.1.14.v20131031/jetty-io-  
8.1.14.v20131031.jar:/Users/louischang/.m2/repository/com/google/guava/guava/14.0.1/g  
uava-14.0.1.jar:/Users/louischang/.m2/repository/org/apache/commons/commons-  
lang3/3.3.2/commons-lang3-  
3.3.2.jar:/Users/louischang/.m2/repository/com/google/code/findbugs/jsr305/1.3.9/jsr305  
-1.3.9.jar:/Users/louischang/.m2/repository/org/slf4j/slf4j-api/1.7.5/slf4j-api-  
1.7.5.jar:/Users/louischang/.m2/repository/org/slf4j/jul-to-slf4j/1.7.5/jul-to-slf4j-  
1.7.5.jar:/Users/louischang/.m2/repository/org/slf4j/jcl-over-slf4j/1.7.5/jcl-over-slf4j-  
1.7.5.jar:/Users/louischang/.m2/repository/log4j/log4j/1.2.17/log4j-  
1.2.17.jar:/Users/louischang/.m2/repository/org/slf4j/slf4j-log4j12/1.7.5/slf4j-log4j12-  
1.7.5.jar:/Users/louischang/.m2/repository/com/ning/compress-lzf/1.0.0/compress-lzf-  
1.0.0.jar:/Users/louischang/.m2/repository/org/xerial/snappy/snappy-java/1.0.5/snappy-  
java-1.0.5.jar:/Users/louischang/.m2/repository/com/twitter/chill\_2.10/0.3.6/chill\_2.10-  
0.3.6.jar:/Users/louischang/.m2/repository/com/esotericsoftware/kryo/kryo/2.21/kryo-  
2.21.jar:/Users/louischang/.m2/repository/com/esotericsoftware/reflectasm/reflectasm/1  
.07/reflectasm-1.07-  
shaded.jar:/Users/louischang/.m2/repository/com/esotericsoftware/minlog/minlog/1.2/mi  
nlog-1.2.jar:/Users/louischang/.m2/repository/org/objenesis/objenesis/1.2/objenesis-  
1.2.jar:/Users/louischang/.m2/repository/com/twitter/chill-java/0.3.6/chill-java-  
0.3.6.jar:/Users/louischang/.m2/repository/commons-net/commons-net/2.2/commons-  
net-2.2.jar:/Users/louischang/.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/louischang/.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/louischang/.m2/repository/com/typesafe/config/1.0.2/config-1.0.2.jar:/Users/louischang/.m2/repository/io/netty/netty/3.6.6.Final/netty-3.6.6.Final.jar:/Users/louischang/.m2/repository/org/spark-project/protobuf/protobuf-java/2.4.1-shaded/protobuf-java-2.4.1-shaded.jar:/Users/louischang/.m2/repository/org/uncommons/maths/uncommons-maths/1.2.2a/uncommons-maths-1.2.2a.jar:/Users/louischang/.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/louischang/.m2/repository/org/scala-lang/scala-library/2.10.4/scala-library-2.10.4.jar:/Users/louischang/.m2/repository/org/json4s/json4s-jackson\_2.10/3.2.6/json4s-jackson\_2.10-3.2.6.jar:/Users/louischang/.m2/repository/org/json4s/json4s-core\_2.10/3.2.6/json4s-core\_2.10-3.2.6.jar:/Users/louischang/.m2/repository/org/json4s/json4s-ast\_2.10/3.2.6/json4s-ast\_2.10-3.2.6.jar:/Users/louischang/.m2/repository/com/thoughtworks/paranamer/paranamer/2.6/paranamer-2.6.jar:/Users/louischang/.m2/repository/org/scala-lang/scalap/2.10.0/scalap-2.10.0.jar:/Users/louischang/.m2/repository/org/scala-lang/scala-compiler/2.10.0/scala-compiler-2.10.0.jar:/Users/louischang/.m2/repository/org/scala-lang/scala-reflect/2.10.0/scala-reflect-2.10.0.jar:/Users/louischang/.m2/repository/com/fasterxml/jackson/core/jackson-databind/2.3.0/jackson-databind-2.3.0.jar:/Users/louischang/.m2/repository/com/fasterxml/jackson/core/jackson-annotations/2.3.0/jackson-annotations-2.3.0.jar:/Users/louischang/.m2/repository/com/fasterxml/jackson/core/jackson-core/2.3.0/jackson-core-2.3.0.jar:/Users/louischang/.m2/repository/colt/colt/1.2.0/colt-1.2.0.jar:/Users/louischang/.m2/repository/concurrent/concurrent/1.3.4/concurrent-1.3.4.jar:/Users/louischang/.m2/repository/org/apache/mesos/mesos/0.18.1/mesos-0.18.1-shaded-protobuf.jar:/Users/louischang/.m2/repository/io/netty/netty-all/4.0.17.Final/netty-all-4.0.17.Final.jar:/Users/louischang/.m2/repository/com/clearspring/analytics/stream/2.5.1/stream-2.5.1.jar:/Users/louischang/.m2/repository/com/codahale/metrics/metrics-core/3.0.0/metrics-core-3.0.0.jar:/Users/louischang/.m2/repository/com/codahale/metrics/metrics-jvm/3.0.0/metrics-jvm-3.0.0.jar:/Users/louischang/.m2/repository/com/codahale/metrics/metrics-json/3.0.0/metrics-json-3.0.0.jar:/Users/louischang/.m2/repository/com/codahale/metrics/metrics-

graphite/3.0.0/metrics-graphite-  
3.0.0.jar:/Users/louischang/.m2/repository/org/tachyonproject/tachyon/0.4.1-  
thrift/tachyon-0.4.1-  
thrift.jar:/Users/louischang/.m2/repository/org/apache/ant/ant/1.9.0/ant-  
1.9.0.jar:/Users/louischang/.m2/repository/org/apache/ant/ant-launcher/1.9.0/ant-  
launcher-1.9.0.jar:/Users/louischang/.m2/repository/commons-io/commons-  
io/2.4/commons-io-2.4.jar:/Users/louischang/.m2/repository/org/spark-  
project/pyrolite/2.0.1/pyrolite-  
2.0.1.jar:/Users/louischang/.m2/repository/net/sf/py4j/py4j/0.8.1/py4j-0.8.1.jar  
WordFinder All'sWellThatEndsWell.txt

objc[21509]: Class JavaLaunchHelper is implemented in both  
/Library/Java/JavaVirtualMachines/jdk1.8.0\_151.jdk/Contents/Home/bin/java  
(0x1008264c0) and  
/Library/Java/JavaVirtualMachines/jdk1.8.0\_151.jdk/Contents/Home/jre/lib/libinstrument.  
dylib (0x1093754e0). One of the two will be used. Which one is undefined.

24/04/19 21:44:18 WARN util.Utils: Your hostname, Louis-MacBook-Pro.local resolves to a  
loopback address: 127.0.0.1; using 172.31.67.122 instead (on interface utun6)

24/04/19 21:44:18 WARN util.Utils: Set SPARK\_LOCAL\_IP if you need to bind to another  
address

24/04/19 21:44:18 INFO spark.SecurityManager: Changing view acls to: louischang

24/04/19 21:44:18 INFO spark.SecurityManager: SecurityManager: authentication  
disabled; ui acls disabled; users with view permissions: Set(louischang)

24/04/19 21:44:18 INFO slf4j.Slf4jLogger: Slf4jLogger started

24/04/19 21:44:18 INFO Remoting: Starting remoting

24/04/19 21:44:18 INFO Remoting: Remoting started; listening on  
addresses :[akka.tcp://spark@vpn-172-31-67-122.vpn.local.cmu.edu:54925]

24/04/19 21:44:18 INFO Remoting: Remoting now listens on addresses:  
[akka.tcp://spark@vpn-172-31-67-122.vpn.local.cmu.edu:54925]

24/04/19 21:44:18 INFO spark.SparkEnv: Registering MapOutputTracker

24/04/19 21:44:18 INFO spark.SparkEnv: Registering BlockManagerMaster

24/04/19 21:44:18 INFO storage.DiskBlockManager: Created local directory at  
/var/folders/rq/ytt2\_vp15\_3cmwy08gvnql0c0000gn/T/spark-local-20240419214418-0cf2

24/04/19 21:44:18 INFO storage.MemoryStore: MemoryStore started with capacity 4.8 GB.

24/04/19 21:44:18 INFO network.ConnectionManager: Bound socket to port 54926 with id = ConnectionManagerId(vpn-172-31-67-122.vpn.local.cmu.edu,54926)

24/04/19 21:44:18 INFO storage.BlockManagerMaster: Trying to register BlockManager

24/04/19 21:44:18 INFO storage.BlockManagerInfo: Registering block manager vpn-172-31-67-122.vpn.local.cmu.edu:54926 with 4.8 GB RAM

24/04/19 21:44:18 INFO storage.BlockManagerMaster: Registered BlockManager

24/04/19 21:44:18 INFO spark.HttpServer: Starting HTTP Server

24/04/19 21:44:19 INFO server.Server: jetty-8.1.14.v20131031

24/04/19 21:44:19 INFO server.AbstractConnector: Started SocketConnector@0.0.0.0:54927

24/04/19 21:44:19 INFO broadcast.HttpBroadcast: Broadcast server started at http://172.31.67.122:54927

24/04/19 21:44:19 INFO spark.HttpFileServer: HTTP File server directory is /var/folders/rq/ytt2\_vp15\_3cmwy08gvnql0c0000gn/T/spark-16f92819-3029-4bef-9582-fc5a8cbc9ff3

24/04/19 21:44:19 INFO spark.HttpServer: Starting HTTP Server

24/04/19 21:44:19 INFO server.Server: jetty-8.1.14.v20131031

24/04/19 21:44:19 INFO server.AbstractConnector: Started SocketConnector@0.0.0.0:54928

24/04/19 21:44:19 INFO server.Server: jetty-8.1.14.v20131031

24/04/19 21:44:19 INFO server.AbstractConnector: Started SelectChannelConnector@0.0.0.0:4040

24/04/19 21:44:19 INFO ui.SparkUI: Started SparkUI at http://vpn-172-31-67-122.vpn.local.cmu.edu:4040

24/04/19 21:44:19 INFO storage.MemoryStore: ensureFreeSpace(32816) called with curMem=0, maxMem=5153960755

24/04/19 21:44:19 INFO storage.MemoryStore: Block broadcast\_0 stored as values to memory (estimated size 32.0 KB, free 4.8 GB)

Please enter a word:

love

24/04/19 21:44:32 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable

24/04/19 21:44:32 WARN snappy.LoadSnappy: Snappy native library not loaded

24/04/19 21:44:32 INFO mapred.FileInputFormat: Total input paths to process : 1

24/04/19 21:44:32 INFO spark.SparkContext: Starting job: collect at WordFinder.java:26

24/04/19 21:44:32 INFO scheduler.DAGScheduler: Got job 0 (collect at WordFinder.java:26) with 1 output partitions (allowLocal=false)

24/04/19 21:44:32 INFO scheduler.DAGScheduler: Final stage: Stage 0(collect at WordFinder.java:26)

24/04/19 21:44:32 INFO scheduler.DAGScheduler: Parents of final stage: List()

24/04/19 21:44:32 INFO scheduler.DAGScheduler: Missing parents: List()

24/04/19 21:44:32 INFO scheduler.DAGScheduler: Submitting Stage 0 (FilteredRDD[2] at filter at WordFinder.java:23), which has no missing parents

24/04/19 21:44:32 INFO scheduler.DAGScheduler: Submitting 1 missing tasks from Stage 0 (FilteredRDD[2] at filter at WordFinder.java:23)

24/04/19 21:44:32 INFO scheduler.TaskSchedulerImpl: Adding task set 0.0 with 1 tasks

24/04/19 21:44:32 INFO scheduler.TaskSetManager: Starting task 0.0:0 as TID 0 on executor localhost: localhost (PROCESS\_LOCAL)

24/04/19 21:44:32 INFO scheduler.TaskSetManager: Serialized task 0.0:0 as 2158 bytes in 1 ms

24/04/19 21:44:32 INFO executor.Executor: Running task ID 0

24/04/19 21:44:32 INFO storage.BlockManager: Found block broadcast\_0 locally

24/04/19 21:44:32 INFO rdd.HadoopRDD: Input split:  
file:/Users/louischang/Library/CloudStorage/OneDrive-andrew.cmu.edu/DS/Project5/Project5Part2/All'sWellThatEndsWell.txt:0+133885

24/04/19 21:44:32 INFO executor.Executor: Serialized size of result for 0 is 3801

24/04/19 21:44:32 INFO executor.Executor: Sending result for 0 directly to driver

24/04/19 21:44:32 INFO executor.Executor: Finished task ID 0

24/04/19 21:44:32 INFO scheduler.TaskSetManager: Finished TID 0 in 87 ms on localhost (progress: 1/1)

24/04/19 21:44:32 INFO scheduler.DAGScheduler: Completed ResultTask(0, 0)

24/04/19 21:44:32 INFO scheduler.TaskSchedulerImpl: Removed TaskSet 0.0, whose tasks have all completed, from pool

24/04/19 21:44:32 INFO scheduler.DAGScheduler: Stage 0 (collect at WordFinder.java:26) finished in 0.094 s

24/04/19 21:44:32 INFO spark.SparkContext: Job finished: collect at WordFinder.java:26, took 0.159223167 s

Lines containing the word 'love':

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 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 persevere.

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.

Process finished with exit code 0