

Project5
Ruidi Chang, ruidic
ruidic@andrew.cmu.edu

Table of Contents

Project5Part1	2
Heinz Cluster.....	2
Screenshot for Task7	2
Project5Part2	3
Code	3
The TempestAnalytics.java	3
Main.java	4
Output Screenshot.....	5
Heinz Cluster	5
All “info” lines:	5

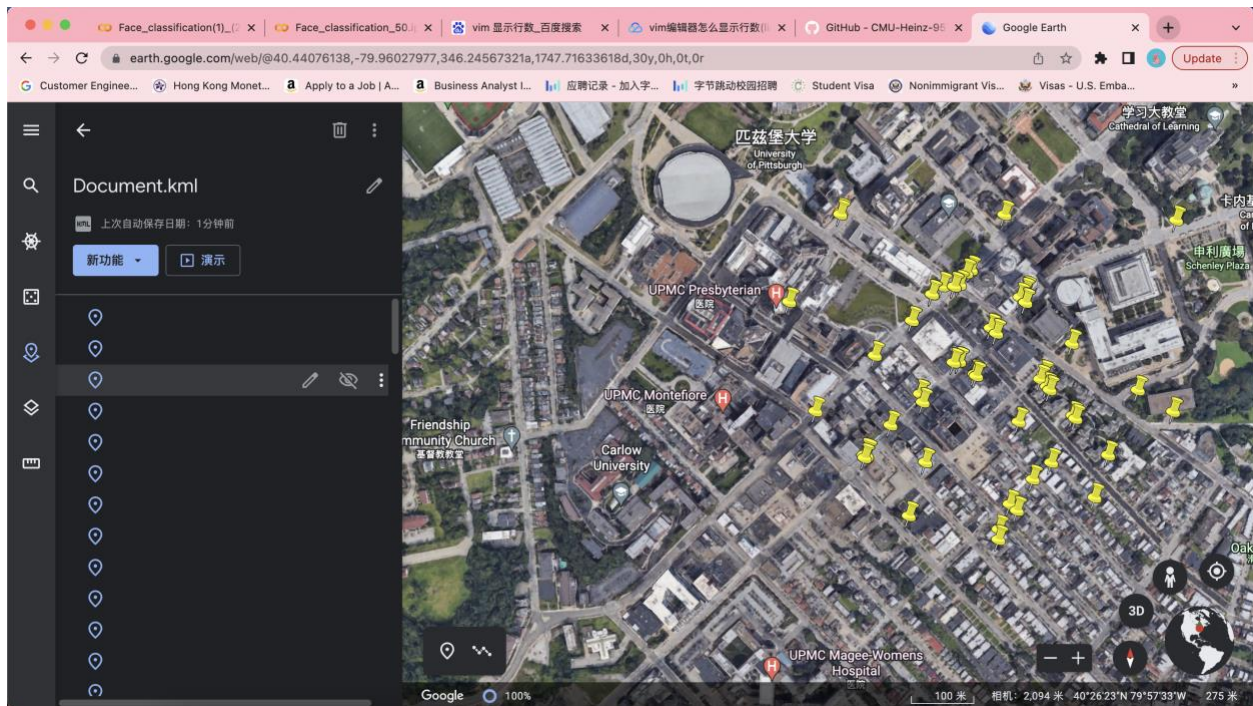
Project5Part1

Heinz Cluster

userID: student148

password: @Mia611

Screenshot for Task7



Project5Part2

Code

The TempestAnalytics.java

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

import java.util.Arrays;
import java.util.List;

/**
 * This program that uses Spark to read The Tempest and perform various
 * calculations, including:
 * Number of lines
 * Number of words
 * Number of distinct words
 * Number of symbols
 * Number of distinct symbols
 * Number of distinct letters
 * It also provides user interaction to let user search for word.
 *
 * Author: Ruidi Chang
 * andrew id: ruidic
 */
public class TempestAnalytics {
    public static void tempest(String fileName, String SearchTerm) {

        // Create a SparkConf object
        SparkConf sparkConf = new
SparkConf().setMaster("local").setAppName("Project5Part2");
        // JavaSparkContext
        JavaSparkContext sparkContext = new JavaSparkContext(sparkConf);
        // RDD
        JavaRDD<String> input = sparkContext.textFile(fileName);

        // Task 0
        // Map: number of lines
        JavaRDD<String> Lines = input.flatMap(content ->
Arrays.asList(content.split("\n")));
        // Reduce: count
        long count0 = Lines.count();

        // Task 1
        // Map: number of words
        JavaRDD<String> Words = input.flatMap(content ->
Arrays.asList(content.split("[^a-zA-Z]+"))).filter(word ->
(!word.isEmpty()));
        // Reduce: count
        long count1 = Words.count();

        // Task 2
```

```

        // Map: Number of distinct words
        JavaRDD<String> DistinctWords = Words.distinct();
        // Reduce: count
        long count2 = DistinctWords.count();

        // Task 3
        // Map: number of symbols
        JavaRDD<String> Symbol = input.flatMap(content ->
Arrays.asList(content.split("")));
        // Reduce: count
        long count3 = Symbol.count();

        // Task 4
        // Map: Number of distinct symbol
        JavaRDD<String> DistinctSymbol = Symbol.distinct();
        // Reduce: count
        long count4 = DistinctSymbol.count();

        //Task5
        // Map: number of distinct letter
        JavaRDD<String> DistinctLetters = Symbol.filter(l -> l.matches("[a-
zA-Z]+")).distinct();
        // Reduce: count
        long count5 = DistinctLetters.count();

        // Task6
        // Map: search word
        JavaRDD<String> Result = Lines.filter(l -> l.contains(SearchTerm));
        // Reduce: collect
        List<String> result = Result.collect();

        System.out.println("Number of lines: " + count0);
        System.out.println("Number of words: " + count1);
        System.out.println("Number of distinct words: " + count2);
        System.out.println("Number of symbols: " + count3);
        System.out.println("Number of distinct symbols: " + count4);
        System.out.println("Number of distinct letters: " + count5);
        System.out.println("lines that contains " + SearchTerm);
        for (String l : result) {
            System.out.println(l);
        }
    }
}

```

Main.java

```

import java.util.Scanner;
/**
 * Main.java
 *
 * Author: Ruidi Chang
 * andrew id: ruidic
 */
public class Main {
    public static void main(String[] args) {
        if (args.length == 0) {
            System.out.println("No files provided.");
        }
    }
}

```

```

        System.exit(0);
    }
    // Interact with user
    System.out.println("Please enter a word: ");
    Scanner scan = new Scanner(System.in);
    String searchWord = scan.nextLine().trim();
    TempestAnalytics t = new TempestAnalytics();
    t.tempest(args[0], searchWord);
}

```

Output Screenshot

```

21/12/20 20:42:05 INFO spark.driver.Context: add finished: collect at TempestAnalytics.java:72, took 0.0073000s
Number of lines: 3466
Number of words: 18167
Number of distinct words: 3595
Number of symbols: 97132
Number of distinct symbols: 77
Number of distinct letters: 51

Lines that contains love
BOATSWAIN None that I more love than myself. You are
Of all the world I loved, and to him put
So dear the love my people bore me, nor set
Knowing I loved my books, he furnished me
Into a cloven pine, within which rift
I do not love to look on.
That burn by day and night. And then I loved thee,
And I the King shall love thee.
All wound with adders, who with cloven tongues
    She loved not the savor of tar nor of pitch,
MIRANDA Do you love me?
Do love, prize, honor you.
And his and mine loved darling. [He exits, above.]
Were but my trials of thy love, and thou
With such love as 'tis now, the murkiest den,
    Do you love me, master? No?
Whose shadow the dismissed bachelor loves,
A contract of true love to celebrate,
On the blest lovers.
A contract of true love. Be not too late.
FERDINAND No, my dearest love,
Of these our dear-beloved solemnized,

Process finished with exit code 0

```

Heinz Cluster

userID: student148

password: @Mia611

All “info” lines:

```

/Library/Java/JavaVirtualMachines/jdk1.8.0_301.jdk/Contents/Home/bin/java -
javaagent:/Applications/IntelliJ IDEA.app/Contents/lib/idea_rt.jar=52669:/Applications/IntelliJ
IDEA.app/Contents/bin -Dfile.encoding=UTF-8 -classpath
/Library/Java/JavaVirtualMachines/jdk1.8.0_301.jdk/Contents/Home/jre/lib/charsets.jar:/Libra
ry/Java/JavaVirtualMachines/jdk1.8.0_301.jdk/Contents/Home/jre/lib/deploy.jar:/Library/Java/
JavaVirtualMachines/jdk1.8.0_301.jdk/Contents/Home/jre/lib/ext/cldrdata.jar:/Library/Java/Ja
vaVirtualMachines/jdk1.8.0_301.jdk/Contents/Home/jre/lib/ext/dnsns.jar:/Library/Java/JavaVir
tualMachines/jdk1.8.0_301.jdk/Contents/Home/jre/lib/ext/jaccess.jar:/Library/Java/JavaVirtual
Machines/jdk1.8.0_301.jdk/Contents/Home/jre/lib/ext/jfxrt.jar:/Library/Java/JavaVirtualMachi

```

nes/jdk1.8.0_301.jdk/Contents/Home/jre/lib/ext/localedata.jar:/Library/Java/JavaVirtualMachines/jdk1.8.0_301.jdk/Contents/Home/jre/lib/ext/nashorn.jar:/Library/Java/JavaVirtualMachines/jdk1.8.0_301.jdk/Contents/Home/jre/lib/ext/sunec.jar:/Library/Java/JavaVirtualMachines/jdk1.8.0_301.jdk/Contents/Home/jre/lib/ext/sunjce_provider.jar:/Library/Java/JavaVirtualMachines/jdk1.8.0_301.jdk/Contents/Home/jre/lib/ext/sunpkcs11.jar:/Library/Java/JavaVirtualMachines/jdk1.8.0_301.jdk/Contents/Home/jre/lib/ext/zipfs.jar:/Library/Java/JavaVirtualMachines/jdk1.8.0_301.jdk/Contents/Home/jre/lib/javaws.jar:/Library/Java/JavaVirtualMachines/jdk1.8.0_301.jdk/Contents/Home/jre/lib/jce.jar:/Library/Java/JavaVirtualMachines/jdk1.8.0_301.jdk/Contents/Home/jre/lib/jfr.jar:/Library/Java/JavaVirtualMachines/jdk1.8.0_301.jdk/Contents/Home/jre/lib/jfxswt.jar:/Library/Java/JavaVirtualMachines/jdk1.8.0_301.jdk/Contents/Home/jre/lib/jssse.jar:/Library/Java/JavaVirtualMachines/jdk1.8.0_301.jdk/Contents/Home/jre/lib/management-

agent.jar:/Library/Java/JavaVirtualMachines/jdk1.8.0_301.jdk/Contents/Home/jre/lib/plugin.jar:/Library/Java/JavaVirtualMachines/jdk1.8.0_301.jdk/Contents/Home/jre/lib/resources.jar:/Library/Java/JavaVirtualMachines/jdk1.8.0_301.jdk/Contents/Home/jre/lib/rt.jar:/Users/changruidi/CMUS3/95-702 Distributed

Systems/Week15/Project5/Project5Part2/target/classes:/Users/changruidi/.m2/repository/org/apache/spark/spark-core_2.10/1.0.0/spark-core_2.10-

1.0.0.jar:/Users/changruidi/.m2/repository/org/apache/hadoop/hadoop-client/1.0.4/hadoop-client-1.0.4.jar:/Users/changruidi/.m2/repository/org/apache/hadoop/hadoop-core/1.0.4/hadoop-core-

1.0.4.jar:/Users/changruidi/.m2/repository/xmlenc/xmlenc/0.52/xmlenc-

0.52.jar:/Users/changruidi/.m2/repository/org/apache/commons/commons-

math/2.1/commons-math-2.1.jar:/Users/changruidi/.m2/repository/commons-configuration/commons-configuration/1.6/commons-configuration-

1.6.jar:/Users/changruidi/.m2/repository/commons-collections/commons-

collections/3.2.1/commons-collections-3.2.1.jar:/Users/changruidi/.m2/repository/commons-lang/commons-lang/2.4/commons-lang-2.4.jar:/Users/changruidi/.m2/repository/commons-digester/commons-digester/1.8/commons-digester-

1.8.jar:/Users/changruidi/.m2/repository/commons-beanutils/commons-

beanutils/1.7.0/commons-beanutils-1.7.0.jar:/Users/changruidi/.m2/repository/commons-beanutils/commons-beanutils-core/1.8.0/commons-beanutils-core-

1.8.0.jar:/Users/changruidi/.m2/repository/commons-el/commons-el/1.0/commons-el-

1.0.jar:/Users/changruidi/.m2/repository/hsqldb/hsqldb/1.8.0.10/hsqldb-

1.8.0.10.jar:/Users/changruidi/.m2/repository/oro/oro/2.0.8/oro-

2.0.8.jar:/Users/changruidi/.m2/repository/org/codehaus/jackson/jackson-mapper-asl/1.0.1/jackson-mapper-asl-

1.0.1.jar:/Users/changruidi/.m2/repository/org/codehaus/jackson/jackson-core-asl/1.0.1/jackson-core-asl-

1.0.1.jar:/Users/changruidi/.m2/repository/net/java/dev/jets3t/jets3t/0.7.1/jets3t-

0.7.1.jar:/Users/changruidi/.m2/repository/commons-codec/commons-codec/1.3/commons-codec-1.3.jar:/Users/changruidi/.m2/repository/commons-httpclient/commons-

httpclient/3.1/commons-httpclient-

3.1.jar:/Users/changruidi/.m2/repository/org/apache/curator/curator-recipes/2.4.0/curator-

recipes-2.4.0.jar:/Users/changruidi/.m2/repository/org/apache/curator/curator-framework/2.4.0/curator-framework-2.4.0.jar:/Users/changruidi/.m2/repository/org/apache/curator/curator-client-2.4.0.jar:/Users/changruidi/.m2/repository/org/apache/curator/curator-client-2.4.0.jar:/Users/changruidi/.m2/repository/org/apache/zookeeper/zookeeper/3.4.5/zookeeper-3.4.5.jar:/Users/changruidi/.m2/repository/jline/jline/0.9.94/jline-0.9.94.jar:/Users/changruidi/.m2/repository/org/eclipse/jetty/jetty-plus/8.1.14.v20131031/jetty-plus-8.1.14.v20131031.jar:/Users/changruidi/.m2/repository/org/eclipse/jetty/orbit/javax.transaction/1.1.1.v201105210645/javax.transaction-1.1.1.v201105210645.jar:/Users/changruidi/.m2/repository/org/eclipse/jetty/jetty-webapp/8.1.14.v20131031/jetty-webapp-8.1.14.v20131031.jar:/Users/changruidi/.m2/repository/org/eclipse/jetty/jetty-xml/8.1.14.v20131031/jetty-xml-8.1.14.v20131031.jar:/Users/changruidi/.m2/repository/org/eclipse/jetty/jetty-servlet/8.1.14.v20131031/jetty-servlet-8.1.14.v20131031.jar:/Users/changruidi/.m2/repository/org/eclipse/jetty/jetty-jndi/8.1.14.v20131031/jetty-jndi-8.1.14.v20131031.jar:/Users/changruidi/.m2/repository/org/eclipse/jetty/orbit/javax.mail.glassfish/1.4.1.v201005082020/javax.mail.glassfish-1.4.1.v201005082020.jar:/Users/changruidi/.m2/repository/org/eclipse/jetty/orbit/javax.activation/1.1.0.v201105071233/javax.activation-1.1.0.v201105071233.jar:/Users/changruidi/.m2/repository/org/eclipse/jetty/jetty-security/8.1.14.v20131031/jetty-security-8.1.14.v20131031.jar:/Users/changruidi/.m2/repository/org/eclipse/jetty/jetty-util/8.1.14.v20131031/jetty-util-8.1.14.v20131031.jar:/Users/changruidi/.m2/repository/org/eclipse/jetty/jetty-server/8.1.14.v20131031/jetty-server-8.1.14.v20131031.jar:/Users/changruidi/.m2/repository/org/eclipse/jetty/orbit/javax.servlet/3.0.0.v201112011016/javax.servlet-3.0.0.v201112011016.jar:/Users/changruidi/.m2/repository/org/eclipse/jetty/jetty-continuation/8.1.14.v20131031/jetty-continuation-8.1.14.v20131031.jar:/Users/changruidi/.m2/repository/org/eclipse/jetty/jetty-http/8.1.14.v20131031/jetty-http-8.1.14.v20131031.jar:/Users/changruidi/.m2/repository/org/eclipse/jetty/jetty-io/8.1.14.v20131031/jetty-io-8.1.14.v20131031.jar:/Users/changruidi/.m2/repository/com/google/guava/guava/14.0.1/guava-14.0.1.jar:/Users/changruidi/.m2/repository/org/apache/commons/commons-lang3/3.3.2/commons-lang3-3.3.2.jar:/Users/changruidi/.m2/repository/com/google/code/findbugs/jsr305/1.3.9/jsr305-1.3.9.jar:/Users/changruidi/.m2/repository/org/slf4j/slf4j-api/1.7.5/slf4j-api-1.7.5.jar:/Users/changruidi/.m2/repository/org/slf4j/jul-to-slf4j/1.7.5/jul-to-slf4j-1.7.5.jar:/Users/changruidi/.m2/repository/org/slf4j/jcl-over-slf4j/1.7.5/jcl-over-slf4j-1.7.5.jar:/Users/changruidi/.m2/repository/log4j/log4j/1.2.17/log4j-

1.2.17.jar:/Users/changruidi/.m2/repository/org/slf4j/slf4j-log4j12/1.7.5/slf4j-log4j12-
1.7.5.jar:/Users/changruidi/.m2/repository/com/ning/compress-lzf/1.0.0/compress-lzf-
1.0.0.jar:/Users/changruidi/.m2/repository/org/xerial/snappy/snappy-java/1.0.5/snappy-java-
1.0.5.jar:/Users/changruidi/.m2/repository/com/twitter/chill_2.10/0.3.6/chill_2.10-
0.3.6.jar:/Users/changruidi/.m2/repository/com/esotericsoftware/kryo/kryo/2.21/kryo-
2.21.jar:/Users/changruidi/.m2/repository/com/esotericsoftware/reflectasm/reflectasm/1.07/r
eflectasm-1.07-
shaded.jar:/Users/changruidi/.m2/repository/com/esotericsoftware/minlog/minlog/1.2/minlog
-1.2.jar:/Users/changruidi/.m2/repository/org/objenesis/objenesis/1.2/objenesis-
1.2.jar:/Users/changruidi/.m2/repository/com/twitter/chill-java/0.3.6/chill-java-
0.3.6.jar:/Users/changruidi/.m2/repository/commons-net/commons-net/2.2/commons-net-
2.2.jar:/Users/changruidi/.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/changruidi/.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/changruidi/.m2/repository/com/typesafe/config/1.0.2/config-
1.0.2.jar:/Users/changruidi/.m2/repository/io/netty/netty/3.6.6.Final/netty-
3.6.6.Final.jar:/Users/changruidi/.m2/repository/org/spark-project/protobuf/protobuf-
java/2.4.1-shaded/protobuf-java-2.4.1-
shaded.jar:/Users/changruidi/.m2/repository/org/uncommons/maths/uncommons-
maths/1.2.2a/uncommons-maths-1.2.2a.jar:/Users/changruidi/.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/changruidi/.m2/repository/org/scala-lang/scala-library/2.10.4/scala-
library-2.10.4.jar:/Users/changruidi/.m2/repository/org/json4s/json4s-
jackson_2.10/3.2.6/json4s-jackson_2.10-
3.2.6.jar:/Users/changruidi/.m2/repository/org/json4s/json4s-core_2.10/3.2.6/json4s-
core_2.10-3.2.6.jar:/Users/changruidi/.m2/repository/org/json4s/json4s-ast_2.10/3.2.6/json4s-
ast_2.10-
3.2.6.jar:/Users/changruidi/.m2/repository/com/thoughtworks/paranamer/paranamer/2.6/par
anamer-2.6.jar:/Users/changruidi/.m2/repository/org/scala-lang/scalap/2.10.0/scalap-
2.10.0.jar:/Users/changruidi/.m2/repository/org/scala-lang/scala-compiler/2.10.0/scala-
compiler-2.10.0.jar:/Users/changruidi/.m2/repository/org/scala-lang/scala-reflect/2.10.0/scala-
reflect-2.10.0.jar:/Users/changruidi/.m2/repository/com/fasterxml/jackson/core/jackson-
databind/2.3.0/jackson-databind-
2.3.0.jar:/Users/changruidi/.m2/repository/com/fasterxml/jackson/core/jackson-
annotations/2.3.0/jackson-annotations-
2.3.0.jar:/Users/changruidi/.m2/repository/com/fasterxml/jackson/core/jackson-
core/2.3.0/jackson-core-2.3.0.jar:/Users/changruidi/.m2/repository/org/colt/colt/1.2.0/colt-
1.2.0.jar:/Users/changruidi/.m2/repository/concurrent/concurrent/1.3.4/concurrent-
1.3.4.jar:/Users/changruidi/.m2/repository/org/apache/mesos/mesos/0.18.1/mesos-0.18.1-
shaded-protobuf.jar:/Users/changruidi/.m2/repository/io/netty/netty-all/4.0.17.Final/netty-all-
4.0.17.Final.jar:/Users/changruidi/.m2/repository/com/clearspring/analytics/stream/2.5.1/stre
am-2.5.1.jar:/Users/changruidi/.m2/repository/com/codahale/metrics/metrics-
core/3.0.0/metrics-core-

3.0.0.jar:/Users/changruidi/.m2/repository/com/codahale/metrics/metrics-jvm/3.0.0/metrics-jvm-3.0.0.jar:/Users/changruidi/.m2/repository/com/codahale/metrics/metrics-json/3.0.0/metrics-json-3.0.0.jar:/Users/changruidi/.m2/repository/com/codahale/metrics/metrics-graphite/3.0.0/metrics-graphite-3.0.0.jar:/Users/changruidi/.m2/repository/org/tachyonproject/tachyon/0.4.1-thrift/tachyon-0.4.1-thrift.jar:/Users/changruidi/.m2/repository/org/apache/ant/ant/1.9.0/ant-1.9.0.jar:/Users/changruidi/.m2/repository/org/apache/ant/ant-launcher/1.9.0/ant-launcher-1.9.0.jar:/Users/changruidi/.m2/repository/commons-io/commons-io/2.4/commons-io-2.4.jar:/Users/changruidi/.m2/repository/org/spark-project/pyrolite/2.0.1/pyrolite-2.0.1.jar:/Users/changruidi/.m2/repository/net/sf/py4j/py4j/0.8.1/py4j-0.8.1.jar Main TheTempest.txt

Please enter a word:

love

22/12/09 16:41:58 INFO spark.SecurityManager: Changing view acls to: changruidi
22/12/09 16:41:58 INFO spark.SecurityManager: SecurityManager: authentication disabled; ui acls disabled; users with view permissions: Set(changruidi)
22/12/09 16:41:58 INFO slf4j.Slf4jLogger: Slf4jLogger started
22/12/09 16:41:58 INFO Remoting: Starting remoting
22/12/09 16:41:58 INFO Remoting: Remoting started; listening on addresses :[akka.tcp://spark@192.168.1.14:52691]
22/12/09 16:41:58 INFO Remoting: Remoting now listens on addresses: [akka.tcp://spark@192.168.1.14:52691]
22/12/09 16:41:58 INFO spark.SparkEnv: Registering MapOutputTracker
22/12/09 16:41:58 INFO spark.SparkEnv: Registering BlockManagerMaster
22/12/09 16:41:58 INFO storage.DiskBlockManager: Created local directory at /var/folders/07/h1tjpsdj51qcs7bl_48v_10w0000gn/T/spark-local-20221209164158-fb5c
22/12/09 16:41:58 INFO storage.MemoryStore: MemoryStore started with capacity 2.1 GB.
22/12/09 16:41:58 INFO network.ConnectionManager: Bound socket to port 52692 with id = ConnectionManagerId(192.168.1.14,52692)
22/12/09 16:41:58 INFO storage.BlockManagerMaster: Trying to register BlockManager
22/12/09 16:41:58 INFO storage.BlockManagerInfo: Registering block manager 192.168.1.14:52692 with 2.1 GB RAM
22/12/09 16:41:58 INFO storage.BlockManagerMaster: Registered BlockManager
22/12/09 16:41:58 INFO spark.HttpServer: Starting HTTP Server
22/12/09 16:41:58 INFO server.Server: jetty-8.1.14.v20131031
22/12/09 16:41:58 INFO server.AbstractConnector: Started SocketConnector@0.0.0.0:52693
22/12/09 16:41:58 INFO broadcast.HttpBroadcast: Broadcast server started at http://192.168.1.14:52693
22/12/09 16:41:58 INFO spark.HttpFileServer: HTTP File server directory is /var/folders/07/h1tjpsdj51qcs7bl_48v_10w0000gn/T/spark-c9ecc894-2038-4313-8574-f099efe02703
22/12/09 16:41:58 INFO spark.HttpServer: Starting HTTP Server
22/12/09 16:41:58 INFO server.Server: jetty-8.1.14.v20131031

22/12/09 16:41:58 INFO server.AbstractConnector: Started SocketConnector@0.0.0.0:52694
22/12/09 16:41:59 INFO server.Server: jetty-8.1.14.v20131031
22/12/09 16:41:59 INFO server.AbstractConnector: Started
SelectChannelConnector@0.0.0.0:4040
22/12/09 16:41:59 INFO ui.SparkUI: Started SparkUI at http://192.168.1.14:4040
22/12/09 16:41:59 INFO storage.MemoryStore: ensureFreeSpace(32816) called with
curMem=0, maxMem=2290719129
22/12/09 16:41:59 INFO storage.MemoryStore: Block broadcast_0 stored as values to memory
(estimated size 32.0 KB, free 2.1 GB)
22/12/09 16:41:59 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your
platform... using builtin-java classes where applicable
22/12/09 16:41:59 WARN snappy.LoadSnappy: Snappy native library not loaded
22/12/09 16:41:59 INFO mapred.FileInputFormat: Total input paths to process : 1
22/12/09 16:41:59 INFO spark.SparkContext: Starting job: count at TempestAnalytics.java:35
22/12/09 16:41:59 INFO scheduler.DAGScheduler: Got job 0 (count at
TempestAnalytics.java:35) with 1 output partitions (allowLocal=false)
22/12/09 16:41:59 INFO scheduler.DAGScheduler: Final stage: Stage 0(count at
TempestAnalytics.java:35)
22/12/09 16:41:59 INFO scheduler.DAGScheduler: Parents of final stage: List()
22/12/09 16:41:59 INFO scheduler.DAGScheduler: Missing parents: List()
22/12/09 16:41:59 INFO scheduler.DAGScheduler: Submitting Stage 0 (FlatMappedRDD[2] at
flatMap at TempestAnalytics.java:33), which has no missing parents
22/12/09 16:41:59 INFO scheduler.DAGScheduler: Submitting 1 missing tasks from Stage 0
(FlatMappedRDD[2] at flatMap at TempestAnalytics.java:33)
22/12/09 16:41:59 INFO scheduler.TaskSchedulerImpl: Adding task set 0.0 with 1 tasks
22/12/09 16:41:59 INFO scheduler.TaskSetManager: Starting task 0.0:0 as TID 0 on executor
localhost: localhost (PROCESS_LOCAL)
22/12/09 16:41:59 INFO scheduler.TaskSetManager: Serialized task 0.0:0 as 2234 bytes in 2 ms
22/12/09 16:41:59 INFO executor.Executor: Running task ID 0
22/12/09 16:41:59 INFO storage.BlockManager: Found block broadcast_0 locally
22/12/09 16:41:59 INFO rdd.HadoopRDD: Input split: file:/Users/changruidi/CMUS3/95-702
Distributed Systems/Week15/Project5/Project5Part2/TheTempest.txt:0+103313
22/12/09 16:41:59 INFO executor.Executor: Serialized size of result for 0 is 597
22/12/09 16:41:59 INFO executor.Executor: Sending result for 0 directly to driver
22/12/09 16:41:59 INFO executor.Executor: Finished task ID 0
22/12/09 16:41:59 INFO scheduler.TaskSetManager: Finished TID 0 in 206 ms on localhost
(progress: 1/1)
22/12/09 16:41:59 INFO scheduler.DAGScheduler: Completed ResultTask(0, 0)
22/12/09 16:41:59 INFO scheduler.TaskSchedulerImpl: Removed TaskSet 0.0, whose tasks have
all completed, from pool
22/12/09 16:41:59 INFO scheduler.DAGScheduler: Stage 0 (count at TempestAnalytics.java:35)
finished in 0.219 s
22/12/09 16:41:59 INFO spark.SparkContext: Job finished: count at TempestAnalytics.java:35,
took 0.322116334 s

22/12/09 16:41:59 INFO spark.SparkContext: Starting job: count at TempestAnalytics.java:41
22/12/09 16:41:59 INFO scheduler.DAGScheduler: Got job 1 (count at TempestAnalytics.java:41) with 1 output partitions (allowLocal=false)
22/12/09 16:41:59 INFO scheduler.DAGScheduler: Final stage: Stage 1(count at TempestAnalytics.java:41)
22/12/09 16:41:59 INFO scheduler.DAGScheduler: Parents of final stage: List()
22/12/09 16:41:59 INFO scheduler.DAGScheduler: Missing parents: List()
22/12/09 16:41:59 INFO scheduler.DAGScheduler: Submitting Stage 1 (FilteredRDD[4] at filter at TempestAnalytics.java:39), which has no missing parents
22/12/09 16:41:59 INFO scheduler.DAGScheduler: Submitting 1 missing tasks from Stage 1 (FilteredRDD[4] at filter at TempestAnalytics.java:39)
22/12/09 16:41:59 INFO scheduler.TaskSchedulerImpl: Adding task set 1.0 with 1 tasks
22/12/09 16:41:59 INFO scheduler.TaskSetManager: Starting task 1.0:0 as TID 1 on executor localhost: localhost (PROCESS_LOCAL)
22/12/09 16:41:59 INFO scheduler.TaskSetManager: Serialized task 1.0:0 as 2351 bytes in 1 ms
22/12/09 16:41:59 INFO executor.Executor: Running task ID 1
22/12/09 16:41:59 INFO storage.BlockManager: Found block broadcast_0 locally
22/12/09 16:41:59 INFO rdd.HadoopRDD: Input split: file:/Users/changruidi/CMUS3/95-702 Distributed Systems/Week15/Project5/Project5Part2/TheTempest.txt:0+103313
22/12/09 16:42:00 INFO executor.Executor: Serialized size of result for 1 is 597
22/12/09 16:42:00 INFO executor.Executor: Sending result for 1 directly to driver
22/12/09 16:42:00 INFO executor.Executor: Finished task ID 1
22/12/09 16:42:00 INFO scheduler.TaskSetManager: Finished TID 1 in 116 ms on localhost (progress: 1/1)
22/12/09 16:42:00 INFO scheduler.DAGScheduler: Completed ResultTask(1, 0)
22/12/09 16:42:00 INFO scheduler.TaskSchedulerImpl: Removed TaskSet 1.0, whose tasks have all completed, from pool
22/12/09 16:42:00 INFO scheduler.DAGScheduler: Stage 1 (count at TempestAnalytics.java:41) finished in 0.123 s
22/12/09 16:42:00 INFO spark.SparkContext: Job finished: count at TempestAnalytics.java:41, took 0.142987208 s
22/12/09 16:42:00 INFO spark.SparkContext: Starting job: count at TempestAnalytics.java:47
22/12/09 16:42:00 INFO scheduler.DAGScheduler: Registering RDD 6 (distinct at TempestAnalytics.java:45)
22/12/09 16:42:00 INFO scheduler.DAGScheduler: Got job 2 (count at TempestAnalytics.java:47) with 1 output partitions (allowLocal=false)
22/12/09 16:42:00 INFO scheduler.DAGScheduler: Final stage: Stage 2(count at TempestAnalytics.java:47)
22/12/09 16:42:00 INFO scheduler.DAGScheduler: Parents of final stage: List(Stage 3)
22/12/09 16:42:00 INFO scheduler.DAGScheduler: Missing parents: List(Stage 3)
22/12/09 16:42:00 INFO scheduler.DAGScheduler: Submitting Stage 3 (MapPartitionsRDD[6] at distinct at TempestAnalytics.java:45), which has no missing parents
22/12/09 16:42:00 INFO scheduler.DAGScheduler: Submitting 1 missing tasks from Stage 3 (MapPartitionsRDD[6] at distinct at TempestAnalytics.java:45)

22/12/09 16:42:00 INFO scheduler.TaskSchedulerImpl: Adding task set 3.0 with 1 tasks
22/12/09 16:42:00 INFO scheduler.TaskSetManager: Starting task 3.0:0 as TID 2 on executor localhost: localhost (PROCESS_LOCAL)
22/12/09 16:42:00 INFO scheduler.TaskSetManager: Serialized task 3.0:0 as 2652 bytes in 1 ms
22/12/09 16:42:00 INFO executor.Executor: Running task ID 2
22/12/09 16:42:00 INFO storage.BlockManager: Found block broadcast_0 locally
22/12/09 16:42:00 INFO rdd.HadoopRDD: Input split: file:/Users/changruidi/CMUS3/95-702 Distributed Systems/Week15/Project5/Project5Part2/TheTempest.txt:0+103313
22/12/09 16:42:00 INFO executor.Executor: Serialized size of result for 2 is 782
22/12/09 16:42:00 INFO executor.Executor: Sending result for 2 directly to driver
22/12/09 16:42:00 INFO executor.Executor: Finished task ID 2
22/12/09 16:42:00 INFO scheduler.TaskSetManager: Finished TID 2 in 293 ms on localhost (progress: 1/1)
22/12/09 16:42:00 INFO scheduler.DAGScheduler: Completed ShuffleMapTask(3, 0)
22/12/09 16:42:00 INFO scheduler.TaskSchedulerImpl: Removed TaskSet 3.0, whose tasks have all completed, from pool
22/12/09 16:42:00 INFO scheduler.DAGScheduler: Stage 3 (distinct at TempestAnalytics.java:45) finished in 0.298 s
22/12/09 16:42:00 INFO scheduler.DAGScheduler: looking for newly runnable stages
22/12/09 16:42:00 INFO scheduler.DAGScheduler: running: Set()
22/12/09 16:42:00 INFO scheduler.DAGScheduler: waiting: Set(Stage 2)
22/12/09 16:42:00 INFO scheduler.DAGScheduler: failed: Set()
22/12/09 16:42:00 INFO scheduler.DAGScheduler: Missing parents for Stage 2: List()
22/12/09 16:42:00 INFO scheduler.DAGScheduler: Submitting Stage 2 (MappedRDD[9] at distinct at TempestAnalytics.java:45), which is now runnable
22/12/09 16:42:00 INFO scheduler.DAGScheduler: Submitting 1 missing tasks from Stage 2 (MappedRDD[9] at distinct at TempestAnalytics.java:45)
22/12/09 16:42:00 INFO scheduler.TaskSchedulerImpl: Adding task set 2.0 with 1 tasks
22/12/09 16:42:00 INFO scheduler.TaskSetManager: Starting task 2.0:0 as TID 3 on executor localhost: localhost (PROCESS_LOCAL)
22/12/09 16:42:00 INFO scheduler.TaskSetManager: Serialized task 2.0:0 as 2474 bytes in 0 ms
22/12/09 16:42:00 INFO executor.Executor: Running task ID 3
22/12/09 16:42:00 INFO storage.BlockManager: Found block broadcast_0 locally
22/12/09 16:42:00 INFO storage.BlockFetcherIterator\$BasicBlockFetcherIterator: maxBytesInFlight: 50331648, targetRequestSize: 10066329
22/12/09 16:42:00 INFO storage.BlockFetcherIterator\$BasicBlockFetcherIterator: Getting 1 non-empty blocks out of 1 blocks
22/12/09 16:42:00 INFO storage.BlockFetcherIterator\$BasicBlockFetcherIterator: Started 0 remote fetches in 3 ms
22/12/09 16:42:00 INFO executor.Executor: Serialized size of result for 3 is 863
22/12/09 16:42:00 INFO executor.Executor: Sending result for 3 directly to driver
22/12/09 16:42:00 INFO executor.Executor: Finished task ID 3
22/12/09 16:42:00 INFO scheduler.DAGScheduler: Completed ResultTask(2, 0)

22/12/09 16:42:00 INFO scheduler.TaskSetManager: Finished TID 3 in 190 ms on localhost (progress: 1/1)

22/12/09 16:42:00 INFO scheduler.TaskSchedulerImpl: Removed TaskSet 2.0, whose tasks have all completed, from pool

22/12/09 16:42:00 INFO scheduler.DAGScheduler: Stage 2 (count at TempestAnalytics.java:47) finished in 0.193 s

22/12/09 16:42:00 INFO spark.SparkContext: Job finished: count at TempestAnalytics.java:47, took 0.555167625 s

22/12/09 16:42:00 INFO spark.SparkContext: Starting job: count at TempestAnalytics.java:53

22/12/09 16:42:00 INFO scheduler.DAGScheduler: Got job 3 (count at TempestAnalytics.java:53) with 1 output partitions (allowLocal=false)

22/12/09 16:42:00 INFO scheduler.DAGScheduler: Final stage: Stage 4(count at TempestAnalytics.java:53)

22/12/09 16:42:00 INFO scheduler.DAGScheduler: Parents of final stage: List()

22/12/09 16:42:00 INFO scheduler.DAGScheduler: Missing parents: List()

22/12/09 16:42:00 INFO scheduler.DAGScheduler: Submitting Stage 4 (FlatMappedRDD[10] at flatMap at TempestAnalytics.java:51), which has no missing parents

22/12/09 16:42:00 INFO scheduler.DAGScheduler: Submitting 1 missing tasks from Stage 4 (FlatMappedRDD[10] at flatMap at TempestAnalytics.java:51)

22/12/09 16:42:00 INFO scheduler.TaskSchedulerImpl: Adding task set 4.0 with 1 tasks

22/12/09 16:42:00 INFO scheduler.TaskSetManager: Starting task 4.0:0 as TID 4 on executor localhost: localhost (PROCESS_LOCAL)

22/12/09 16:42:00 INFO scheduler.TaskSetManager: Serialized task 4.0:0 as 2233 bytes in 1 ms

22/12/09 16:42:00 INFO executor.Executor: Running task ID 4

22/12/09 16:42:00 INFO storage.BlockManager: Found block broadcast_0 locally

22/12/09 16:42:00 INFO rdd.HadoopRDD: Input split: file:/Users/changruidi/CMUS3/95-702 Distributed Systems/Week15/Project5/Project5Part2/TheTempest.txt:0+103313

22/12/09 16:42:00 INFO executor.Executor: Serialized size of result for 4 is 597

22/12/09 16:42:00 INFO executor.Executor: Sending result for 4 directly to driver

22/12/09 16:42:00 INFO executor.Executor: Finished task ID 4

22/12/09 16:42:00 INFO scheduler.TaskSetManager: Finished TID 4 in 68 ms on localhost (progress: 1/1)

22/12/09 16:42:00 INFO scheduler.DAGScheduler: Completed ResultTask(4, 0)

22/12/09 16:42:00 INFO scheduler.TaskSchedulerImpl: Removed TaskSet 4.0, whose tasks have all completed, from pool

22/12/09 16:42:00 INFO scheduler.DAGScheduler: Stage 4 (count at TempestAnalytics.java:53) finished in 0.070 s

22/12/09 16:42:00 INFO spark.SparkContext: Job finished: count at TempestAnalytics.java:53, took 0.082984334 s

22/12/09 16:42:00 INFO spark.SparkContext: Starting job: count at TempestAnalytics.java:59

22/12/09 16:42:00 INFO scheduler.DAGScheduler: Registering RDD 12 (distinct at TempestAnalytics.java:57)

22/12/09 16:42:00 INFO scheduler.DAGScheduler: Got job 4 (count at TempestAnalytics.java:59) with 1 output partitions (allowLocal=false)

22/12/09 16:42:00 INFO scheduler.DAGScheduler: Final stage: Stage 5(count at TempestAnalytics.java:59)

22/12/09 16:42:00 INFO scheduler.DAGScheduler: Parents of final stage: List(Stage 6)

22/12/09 16:42:00 INFO scheduler.DAGScheduler: Missing parents: List(Stage 6)

22/12/09 16:42:00 INFO scheduler.DAGScheduler: Submitting Stage 6 (MapPartitionsRDD[12] at distinct at TempestAnalytics.java:57), which has no missing parents

22/12/09 16:42:00 INFO scheduler.DAGScheduler: Submitting 1 missing tasks from Stage 6 (MapPartitionsRDD[12] at distinct at TempestAnalytics.java:57)

22/12/09 16:42:00 INFO scheduler.TaskSchedulerImpl: Adding task set 6.0 with 1 tasks

22/12/09 16:42:00 INFO scheduler.TaskSetManager: Starting task 6.0:0 as TID 5 on executor localhost: localhost (PROCESS_LOCAL)

22/12/09 16:42:00 INFO scheduler.TaskSetManager: Serialized task 6.0:0 as 2532 bytes in 2 ms

22/12/09 16:42:00 INFO executor.Executor: Running task ID 5

22/12/09 16:42:00 INFO storage.BlockManager: Found block broadcast_0 locally

22/12/09 16:42:00 INFO rdd.HadoopRDD: Input split: file:/Users/changruidi/CMUS3/95-702 Distributed Systems/Week15/Project5/Project5Part2/TheTempest.txt:0+103313

22/12/09 16:42:00 INFO executor.Executor: Serialized size of result for 5 is 782

22/12/09 16:42:00 INFO executor.Executor: Sending result for 5 directly to driver

22/12/09 16:42:00 INFO executor.Executor: Finished task ID 5

22/12/09 16:42:00 INFO scheduler.TaskSetManager: Finished TID 5 in 54 ms on localhost (progress: 1/1)

22/12/09 16:42:00 INFO scheduler.DAGScheduler: Completed ShuffleMapTask(6, 0)

22/12/09 16:42:00 INFO scheduler.TaskSchedulerImpl: Removed TaskSet 6.0, whose tasks have all completed, from pool

22/12/09 16:42:00 INFO scheduler.DAGScheduler: Stage 6 (distinct at TempestAnalytics.java:57) finished in 0.055 s

22/12/09 16:42:00 INFO scheduler.DAGScheduler: looking for newly runnable stages

22/12/09 16:42:00 INFO scheduler.DAGScheduler: running: Set()

22/12/09 16:42:00 INFO scheduler.DAGScheduler: waiting: Set(Stage 5)

22/12/09 16:42:00 INFO scheduler.DAGScheduler: failed: Set()

22/12/09 16:42:00 INFO scheduler.DAGScheduler: Missing parents for Stage 5: List()

22/12/09 16:42:00 INFO scheduler.DAGScheduler: Submitting Stage 5 (MappedRDD[15] at distinct at TempestAnalytics.java:57), which is now runnable

22/12/09 16:42:00 INFO scheduler.DAGScheduler: Submitting 1 missing tasks from Stage 5 (MappedRDD[15] at distinct at TempestAnalytics.java:57)

22/12/09 16:42:00 INFO scheduler.TaskSchedulerImpl: Adding task set 5.0 with 1 tasks

22/12/09 16:42:00 INFO scheduler.TaskSetManager: Starting task 5.0:0 as TID 6 on executor localhost: localhost (PROCESS_LOCAL)

22/12/09 16:42:00 INFO scheduler.TaskSetManager: Serialized task 5.0:0 as 2365 bytes in 1 ms

22/12/09 16:42:00 INFO executor.Executor: Running task ID 6

22/12/09 16:42:00 INFO storage.BlockManager: Found block broadcast_0 locally

22/12/09 16:42:00 INFO storage.BlockFetcherIterator\$BasicBlockFetcherIterator: maxBytesInFlight: 50331648, targetRequestSize: 10066329

22/12/09 16:42:00 INFO storage.BlockFetcherIterator\$BasicBlockFetcherIterator: Getting 1 non-empty blocks out of 1 blocks
22/12/09 16:42:00 INFO storage.BlockFetcherIterator\$BasicBlockFetcherIterator: Started 0 remote fetches in 1 ms
22/12/09 16:42:00 INFO executor.Executor: Serialized size of result for 6 is 863
22/12/09 16:42:00 INFO executor.Executor: Sending result for 6 directly to driver
22/12/09 16:42:00 INFO executor.Executor: Finished task ID 6
22/12/09 16:42:00 INFO scheduler.DAGScheduler: Completed ResultTask(5, 0)
22/12/09 16:42:00 INFO scheduler.TaskSetManager: Finished TID 6 in 24 ms on localhost (progress: 1/1)
22/12/09 16:42:00 INFO scheduler.TaskSchedulerImpl: Removed TaskSet 5.0, whose tasks have all completed, from pool
22/12/09 16:42:00 INFO scheduler.DAGScheduler: Stage 5 (count at TempestAnalytics.java:59) finished in 0.025 s
22/12/09 16:42:00 INFO spark.SparkContext: Job finished: count at TempestAnalytics.java:59, took 0.097473417 s
22/12/09 16:42:00 INFO spark.SparkContext: Starting job: count at TempestAnalytics.java:65
22/12/09 16:42:00 INFO scheduler.DAGScheduler: Registering RDD 18 (distinct at TempestAnalytics.java:63)
22/12/09 16:42:00 INFO scheduler.DAGScheduler: Got job 5 (count at TempestAnalytics.java:65) with 1 output partitions (allowLocal=false)
22/12/09 16:42:00 INFO scheduler.DAGScheduler: Final stage: Stage 7(count at TempestAnalytics.java:65)
22/12/09 16:42:00 INFO scheduler.DAGScheduler: Parents of final stage: List(Stage 8)
22/12/09 16:42:00 INFO scheduler.DAGScheduler: Missing parents: List(Stage 8)
22/12/09 16:42:00 INFO scheduler.DAGScheduler: Submitting Stage 8 (MapPartitionsRDD[18] at distinct at TempestAnalytics.java:63), which has no missing parents
22/12/09 16:42:00 INFO scheduler.DAGScheduler: Submitting 1 missing tasks from Stage 8 (MapPartitionsRDD[18] at distinct at TempestAnalytics.java:63)
22/12/09 16:42:00 INFO scheduler.TaskSchedulerImpl: Adding task set 8.0 with 1 tasks
22/12/09 16:42:00 INFO scheduler.TaskSetManager: Starting task 8.0:0 as TID 7 on executor localhost: localhost (PROCESS_LOCAL)
22/12/09 16:42:00 INFO scheduler.TaskSetManager: Serialized task 8.0:0 as 2651 bytes in 1 ms
22/12/09 16:42:00 INFO executor.Executor: Running task ID 7
22/12/09 16:42:00 INFO storage.BlockManager: Found block broadcast_0 locally
22/12/09 16:42:00 INFO rdd.HadoopRDD: Input split: file:/Users/changruidi/CMUS3/95-702 Distributed Systems/Week15/Project5/Project5Part2/TheTempest.txt:0+103313
22/12/09 16:42:01 INFO executor.Executor: Serialized size of result for 7 is 782
22/12/09 16:42:01 INFO executor.Executor: Sending result for 7 directly to driver
22/12/09 16:42:01 INFO executor.Executor: Finished task ID 7
22/12/09 16:42:01 INFO scheduler.DAGScheduler: Completed ShuffleMapTask(8, 0)
22/12/09 16:42:01 INFO scheduler.TaskSetManager: Finished TID 7 in 149 ms on localhost (progress: 1/1)

22/12/09 16:42:01 INFO scheduler.TaskSchedulerImpl: Removed TaskSet 8.0, whose tasks have all completed, from pool

22/12/09 16:42:01 INFO scheduler.DAGScheduler: Stage 8 (distinct at TempestAnalytics.java:63) finished in 0.155 s

22/12/09 16:42:01 INFO scheduler.DAGScheduler: looking for newly runnable stages

22/12/09 16:42:01 INFO scheduler.DAGScheduler: running: Set()

22/12/09 16:42:01 INFO scheduler.DAGScheduler: waiting: Set(Stage 7)

22/12/09 16:42:01 INFO scheduler.DAGScheduler: failed: Set()

22/12/09 16:42:01 INFO scheduler.DAGScheduler: Missing parents for Stage 7: List()

22/12/09 16:42:01 INFO scheduler.DAGScheduler: Submitting Stage 7 (MappedRDD[21] at distinct at TempestAnalytics.java:63), which is now runnable

22/12/09 16:42:01 INFO scheduler.DAGScheduler: Submitting 1 missing tasks from Stage 7 (MappedRDD[21] at distinct at TempestAnalytics.java:63)

22/12/09 16:42:01 INFO scheduler.TaskSchedulerImpl: Adding task set 7.0 with 1 tasks

22/12/09 16:42:01 INFO scheduler.TaskSetManager: Starting task 7.0:0 as TID 8 on executor localhost: localhost (PROCESS_LOCAL)

22/12/09 16:42:01 INFO scheduler.TaskSetManager: Serialized task 7.0:0 as 2473 bytes in 0 ms

22/12/09 16:42:01 INFO executor.Executor: Running task ID 8

22/12/09 16:42:01 INFO storage.BlockManager: Found block broadcast_0 locally

22/12/09 16:42:01 INFO storage.BlockFetcherIterator\$BasicBlockFetcherIterator: maxBytesInFlight: 50331648, targetRequestSize: 10066329

22/12/09 16:42:01 INFO storage.BlockFetcherIterator\$BasicBlockFetcherIterator: Getting 1 non-empty blocks out of 1 blocks

22/12/09 16:42:01 INFO storage.BlockFetcherIterator\$BasicBlockFetcherIterator: Started 0 remote fetches in 0 ms

22/12/09 16:42:01 INFO executor.Executor: Serialized size of result for 8 is 863

22/12/09 16:42:01 INFO executor.Executor: Sending result for 8 directly to driver

22/12/09 16:42:01 INFO executor.Executor: Finished task ID 8

22/12/09 16:42:01 INFO scheduler.TaskSetManager: Finished TID 8 in 11 ms on localhost (progress: 1/1)

22/12/09 16:42:01 INFO scheduler.DAGScheduler: Completed ResultTask(7, 0)

22/12/09 16:42:01 INFO scheduler.TaskSchedulerImpl: Removed TaskSet 7.0, whose tasks have all completed, from pool

22/12/09 16:42:01 INFO scheduler.DAGScheduler: Stage 7 (count at TempestAnalytics.java:65) finished in 0.012 s

22/12/09 16:42:01 INFO spark.SparkContext: Job finished: count at TempestAnalytics.java:65, took 0.203830125 s

22/12/09 16:42:01 INFO spark.SparkContext: Starting job: collect at TempestAnalytics.java:71

22/12/09 16:42:01 INFO scheduler.DAGScheduler: Got job 6 (collect at TempestAnalytics.java:71) with 1 output partitions (allowLocal=false)

22/12/09 16:42:01 INFO scheduler.DAGScheduler: Final stage: Stage 9(collect at TempestAnalytics.java:71)

22/12/09 16:42:01 INFO scheduler.DAGScheduler: Parents of final stage: List()

22/12/09 16:42:01 INFO scheduler.DAGScheduler: Missing parents: List()

22/12/09 16:42:01 INFO scheduler.DAGScheduler: Submitting Stage 9 (FilteredRDD[22] at filter at TempestAnalytics.java:69), which has no missing parents
22/12/09 16:42:01 INFO scheduler.DAGScheduler: Submitting 1 missing tasks from Stage 9 (FilteredRDD[22] at filter at TempestAnalytics.java:69)
22/12/09 16:42:01 INFO scheduler.TaskSchedulerImpl: Adding task set 9.0 with 1 tasks
22/12/09 16:42:01 INFO scheduler.TaskSetManager: Starting task 9.0:0 as TID 9 on executor localhost: localhost (PROCESS_LOCAL)
22/12/09 16:42:01 INFO scheduler.TaskSetManager: Serialized task 9.0:0 as 2372 bytes in 2 ms
22/12/09 16:42:01 INFO executor.Executor: Running task ID 9
22/12/09 16:42:01 INFO storage.BlockManager: Found block broadcast_0 locally
22/12/09 16:42:01 INFO rdd.HadoopRDD: Input split: file:/Users/changruidi/CMUS3/95-702 Distributed Systems/Week15/Project5/Project5Part2/TheTempest.txt:0+103313
22/12/09 16:42:01 INFO executor.Executor: Serialized size of result for 9 is 1460
22/12/09 16:42:01 INFO executor.Executor: Sending result for 9 directly to driver
22/12/09 16:42:01 INFO executor.Executor: Finished task ID 9
22/12/09 16:42:01 INFO scheduler.TaskSetManager: Finished TID 9 in 25 ms on localhost (progress: 1/1)
22/12/09 16:42:01 INFO scheduler.TaskSchedulerImpl: Removed TaskSet 9.0, whose tasks have all completed, from pool
22/12/09 16:42:01 INFO scheduler.DAGScheduler: Completed ResultTask(9, 0)
22/12/09 16:42:01 INFO scheduler.DAGScheduler: Stage 9 (collect at TempestAnalytics.java:71) finished in 0.026 s
22/12/09 16:42:01 INFO spark.SparkContext: Job finished: collect at TempestAnalytics.java:71, took 0.039506833 s
Number of lines: 3466
Number of words: 18167
Number of distinct words: 3595
Number of symbols: 97132
Number of distinct symbols: 77
Number of distinct letters: 51
lines that contains love
BOATSWAIN None that I more love than myself. You are
Of all the world I loved, and to him put
So dear the love my people bore me, nor set
Knowing I loved my books, he furnished me
Into a cloven pine, within which rift
I do not love to look on.
That burn by day and night. And then I loved thee,
And I the King shall love thee.
All wound with adders, who with cloven tongues
 She loved not the savor of tar nor of pitch,
MIRANDA Do you love me?
Do love, prize, honor you.
And his and mine loved darling. [He exits, above.]

Were but my trials of thy love, and thou
With such love as 'tis now, the murkiest den,
 Do you love me, master? No?
Whose shadow the dismissed bachelor loves,
A contract of true love to celebrate,
On the blest lovers.
A contract of true love. Be not too late.
FERDINAND No, my dearest love,
Of these our dear-beloved solemnized,

Process finished with exit code 0