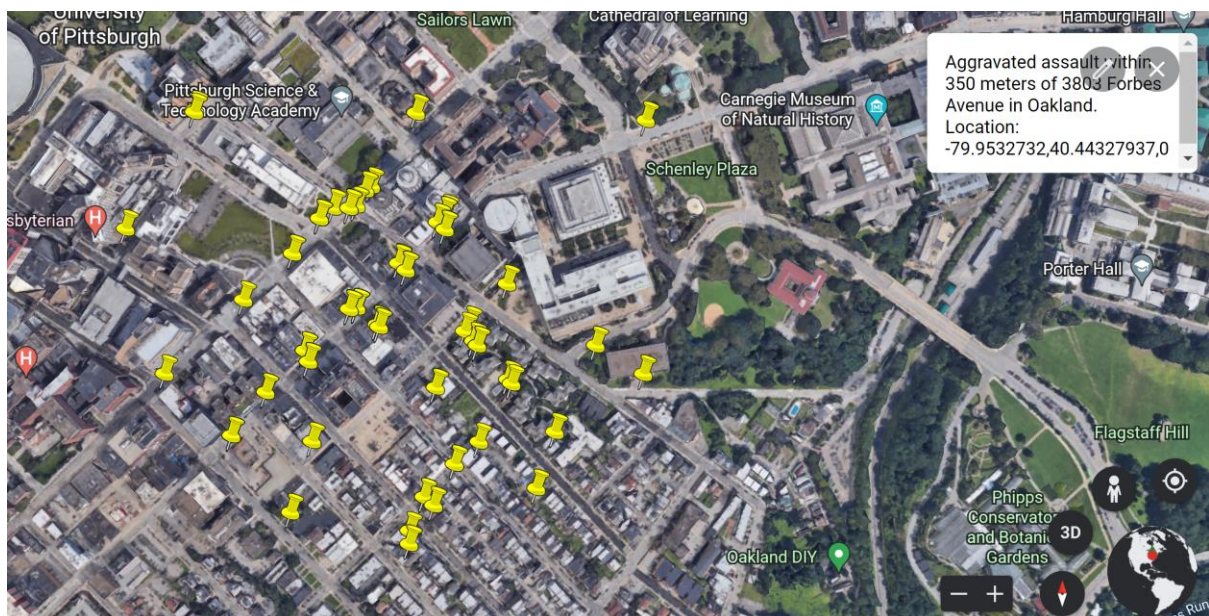```
/**
 * Author: Shivam Patel
 * Andrew IDs: shpatel
 * Emails: shpatel@cmu.edu
 * Last Modified: December 10, 2022
 * File: shpatel_Project5_PDF.pdf
 * Part Of: Project5
 *
**/
```

# Part 1 – Task 7

Screenshot from Google Earth containing locations of aggravated assault crimes that occurred within 350 meters of 3803 Forbes Avenue in Oakland.

# Part 2 – Source Code – Tasks 0 to 6

```java
/**
 * Author: Shivam Patel
 * Andrew IDs: shpatel
 * Emails: shpatel@cmu.edu
 * Last Modified: December 10, 2022
 * File: TempestAnalytics.java
 * Part Of: Project5Part2
 *
 * This Java file uses Spark to read TheTempest.txt file and perform
various
 * calculations on its contents. The location of the file is given as
arguments
 * to the main() function. The calculations performed on TheTempest.txt
file are:
 *
 * - Calculate total number of lines in the file
 * - Calculate total number of words in the file
 * - Calculate total number of distinct words in the file
 * - Calculate total number of symbols in the file
 * - Calculate total number of distinct symbols in the file
 * - Calculate total number of distinct letters in the file
 * - Request the user for an input word and print all the lines
 *   that contain the word entered by the user
 */

// Imports necessary for executing Spark, use Arrays and Scanner classes
import org.apache.spark.SparkConf;
import org.apache.spark.api.java.JavaRDD;
import org.apache.spark.api.java.JavaSparkContext;
import org.apache.spark.api.java.function.VoidFunction;
import java.util.Arrays;
import java.util.Scanner;

public class TempestAnalytics {

    public static void main(String[] args) {

        // Create a Spark configuration object, setting the master to local
and the app name to Project 5 Part 2
        SparkConf sparkConf = new
SparkConf().setMaster("local").setAppName("Project 5 Part 2");

        // Create a JavaSparkContext object with the required spark
configurations
        JavaSparkContext sparkContext = new JavaSparkContext(sparkConf);

        // Create an RDD that represents the data file
        JavaRDD<String> inputFile = sparkContext.textFile(args[0]);

        // Create an RDD built up from each line split into words
        JavaRDD<String> wordsFromFile = inputFile.flatMap(content ->
Arrays.asList(content.split("[^a-zA-Z]+")));

        // ##### Project 5 - Part 2 - Task 0 #####
        // Counts the number of lines in the input file
        long numberOfLines = inputFile.count();

        // Displays number of lines in the input file to the user
```

```java
        System.out.println("Number of lines: " + numberOfLines);

        // ##### Project 5 - Part 2 - Task 1 #####
        // Counts the number of words in the input file
        long numberOfWords = wordsFromFile.filter(k -> (
!k.isEmpty())).count();

        // Displays number of words in the input file to the user
        System.out.println("Number of words: " + numberOfWords);

        // ##### Project 5 - Part 2 - Task 2 #####
        // Counts the number of distinct words in the input file
        long numberOfDistinctWords = wordsFromFile.filter(k -> (
!k.isEmpty())).distinct().count();

        // Displays number of distinct words in the input file to the user
        System.out.println("Number of distinct words: " +
numberOfDistinctWords);

        // ##### Project 5 - Part 2 - Task 3 #####
        // Create an RDD built up from each line split into symbols
        JavaRDD<String> symbolsFromFile = inputFile.flatMap(content ->
Arrays.asList(content.split("")));

        // Counts the number of symbols in the input file
        long numberOfSymbols = symbolsFromFile.count();

        // Displays number of symbols in the input file to the user
        System.out.println("Number of symbols: " + numberOfSymbols);

        // ##### Project 5 - Part 2 - Task 4 #####
        // Counts the number of distinct symbols in the input file
        long numberOfDistinctSymbols = symbolsFromFile.distinct().count();

        // Displays number of distinct symbols in the input file to the
user
        System.out.println("Number of distinct symbols: " +
numberOfDistinctSymbols);

        // ##### Project 5 - Part 2 - Task 5 #####
        // Counts the number of distinct letters in the input file
        long numberOfDistinctLetters = symbolsFromFile.filter(k ->
(k.matches("[a-zA-Z]"))).distinct().count();

        // Displays number of distinct letters in the input file to the
user
        System.out.println("Number of distinct letters: " +
numberOfDistinctLetters);

        // ##### Project 5 - Part 2 - Task 6 #####
        // Create an object of Scanner class to get input from the user
        Scanner s = new Scanner(System.in);

        // Function to print line containing the word entered by user to
the output screen
        VoidFunction<String> printLineContainingUserWord = line ->
System.out.println(line);

        while (true) {
            // Prompt the user for input
            System.out.println("\nEnter a word to search in TheTempest.txt
```

```java
(Enter \"#$@Done@$#\" to stop): ");

            // Get input from user
            String userWord = s.next();

            // If the user enters "#$@Done@$#", stop the execution of while
loop
            // Kept the word to be "#$@Done@$#", as that would not be
present in the input file
            if (userWord.equals("#$@Done@$#")) {
                break;
            }

            // Create an RDD built up from each line in the input file
split by a new line character
            JavaRDD<String> linesFromFile = inputFile.flatMap(content ->
Arrays.asList(content.split("\n")));

            // Update RDD to contain only the lines that contain the word
entered by the user
            linesFromFile = linesFromFile.filter(k ->
(k.contains(userWord)));

            // Loop over the RDD and print lines containing the input word
to the console
            linesFromFile.foreach(printLineContainingUserWord);
        }
        // Exit message
        System.out.println("\nQuitting! Thank you :)");
    }
}
```

Part 2 – Execution Output

C:\Users\shiva\.jdks\temurin-1.8.0_352\bin\java.exe "-javaagent:C:\Program Files\JetBrains\IntelliJ IDEA 2022.2.1\lib\idea_rt.jar=53102:C:\Program Files\JetBrains\IntelliJ IDEA 2022.2.1\bin" -Dfile.encoding=UTF-8 -classpath "C:\Users\shiva\.jdks\temurin-1.8.0_352\jre\lib\charsets.jar;C:\Users\shiva\.jdks\temurin-1.8.0_352\jre\lib\ext\access-bridge-64.jar;C:\Users\shiva\.jdks\temurin-1.8.0_352\jre\lib\ext\cldrdata.jar;C:\Users\shiva\.jdks\temurin-1.8.0_352\jre\lib\ext\dnsns.jar;C:\Users\shiva\.jdks\temurin-1.8.0_352\jre\lib\ext\jaccess.jar;C:\Users\shiva\.jdks\temurin-1.8.0_352\jre\lib\ext\localedata.jar;C:\Users\shiva\.jdks\temurin-1.8.0_352\jre\lib\ext\nashorn.jar;C:\Users\shiva\.jdks\temurin-1.8.0_352\jre\lib\ext\sunec.jar;C:\Users\shiva\.jdks\temurin-1.8.0_352\jre\lib\ext\sunjce_provider.jar;C:\Users\shiva\.jdks\temurin-1.8.0_352\jre\lib\ext\sunmscapi.jar;C:\Users\shiva\.jdks\temurin-1.8.0_352\jre\lib\ext\sunpkcs11.jar;C:\Users\shiva\.jdks\temurin-1.8.0_352\jre\lib\ext\zipfs.jar;C:\Users\shiva\.jdks\temurin-1.8.0_352\jre\lib\jce.jar;C:\Users\shiva\.jdks\temurin-1.8.0_352\jre\lib\jfr.jar;C:\Users\shiva\.jdks\temurin-1.8.0_352\jre\lib\jsse.jar;C:\Users\shiva\.jdks\temurin-1.8.0_352\jre\lib\management-agent.jar;C:\Users\shiva\.jdks\temurin-1.8.0_352\jre\lib\resources.jar;C:\Users\shiva\.jdks\temurin-1.8.0_352\jre\lib\rt.jar;F:\Shivam\CMU Material\A - CMU Study Material\Semester 3\S1 - Distributed Systems for ISM\Projects\Project5\Project5Part2\target\classes;F:\Shivam\CMU Material\A - CMU Study Material\Semester 3\S1 - Distributed Systems for ISM\Projects\Project5\Project5Part2\lib\spark-core_2.10-1.0.0.jar;F:\Shivam\CMU Material\A - CMU Study Material\Semester 3\S1 - Distributed Systems for ISM\Projects\Project5\Project5Part2\lib\hadoop-client-1.0.4.jar;F:\Shivam\CMU Material\A - CMU Study Material\Semester 3\S1 - Distributed Systems for ISM\Projects\Project5\Project5Part2\lib\hadoop-core-1.0.4.jar;F:\Shivam\CMU Material\A - CMU Study Material\Semester 3\S1 - Distributed Systems for ISM\Projects\Project5\Project5Part2\lib\xmlenc-0.52.jar;F:\Shivam\CMU Material\A - CMU Study Material\Semester 3\S1 - Distributed Systems for ISM\Projects\Project5\Project5Part2\lib\commons-math-2.1.jar;F:\Shivam\CMU Material\A - CMU Study Material\Semester 3\S1 - Distributed Systems for ISM\Projects\Project5\Project5Part2\lib\commons-configuration-1.6.jar;F:\Shivam\CMU Material\A - CMU Study Material\Semester 3\S1 - Distributed Systems for ISM\Projects\Project5\Project5Part2\lib\commons-collections-3.2.1.jar;F:\Shivam\CMU Material\A - CMU Study Material\Semester 3\S1 - Distributed Systems for ISM\Projects\Project5\Project5Part2\lib\commons-lang-2.4.jar;F:\Shivam\CMU Material\A - CMU Study Material\Semester 3\S1 - Distributed Systems for ISM\Projects\Project5\Project5Part2\lib\commons-digester-1.8.jar;F:\Shivam\CMU Material\A - CMU Study Material\Semester 3\S1 - Distributed Systems for ISM\Projects\Project5\Project5Part2\lib\commons-beanutils-1.7.0.jar;F:\Shivam\CMU Material\A - CMU Study Material\Semester 3\S1 - Distributed Systems for ISM\Projects\Project5\Project5Part2\lib\commons-beanutils-core-1.8.0.jar;F:\Shivam\CMU Material\A - CMU Study Material\Semester 3\S1 - Distributed Systems for

ISM\Projects\Project5\Project5Part2\lib\commons-el-1.0.jar;F:\Shivam\CMU Material\A - CMU Study Material\Semester 3\S1 - Distributed Systems for ISM\Projects\Project5\Project5Part2\lib\hsqldb-1.8.0.10.jar;F:\Shivam\CMU Material\A - CMU Study Material\Semester 3\S1 - Distributed Systems for ISM\Projects\Project5\Project5Part2\lib\oro-2.0.8.jar;F:\Shivam\CMU Material\A - CMU Study Material\Semester 3\S1 - Distributed Systems for ISM\Projects\Project5\Project5Part2\lib\jackson-mapper-asl-1.0.1.jar;F:\Shivam\CMU Material\A - CMU Study Material\Semester 3\S1 - Distributed Systems for ISM\Projects\Project5\Project5Part2\lib\jackson-core-asl-1.0.1.jar;F:\Shivam\CMU Material\A - CMU Study Material\Semester 3\S1 - Distributed Systems for ISM\Projects\Project5\Project5Part2\lib\jets3t-0.7.1.jar;F:\Shivam\CMU Material\A - CMU Study Material\Semester 3\S1 - Distributed Systems for ISM\Projects\Project5\Project5Part2\lib\commons-codec-1.3.jar;F:\Shivam\CMU Material\A - CMU Study Material\Semester 3\S1 - Distributed Systems for ISM\Projects\Project5\Project5Part2\lib\commons-httpclient-3.1.jar;F:\Shivam\CMU Material\A - CMU Study Material\Semester 3\S1 - Distributed Systems for ISM\Projects\Project5\Project5Part2\lib\curator-recipes-2.4.0.jar;F:\Shivam\CMU Material\A - CMU Study Material\Semester 3\S1 - Distributed Systems for ISM\Projects\Project5\Project5Part2\lib\curator-framework-2.4.0.jar;F:\Shivam\CMU Material\A - CMU Study Material\Semester 3\S1 - Distributed Systems for ISM\Projects\Project5\Project5Part2\lib\curator-client-2.4.0.jar;F:\Shivam\CMU Material\A - CMU Study Material\Semester 3\S1 - Distributed Systems for ISM\Projects\Project5\Project5Part2\lib\zookeeper-3.4.5.jar;F:\Shivam\CMU Material\A - CMU Study Material\Semester 3\S1 - Distributed Systems for ISM\Projects\Project5\Project5Part2\lib\jline-0.9.94.jar;F:\Shivam\CMU Material\A - CMU Study Material\Semester 3\S1 - Distributed Systems for ISM\Projects\Project5\Project5Part2\lib\jetty-plus-8.1.14.v20131031.jar;F:\Shivam\CMU Material\A - CMU Study Material\Semester 3\S1 - Distributed Systems for ISM\Projects\Project5\Project5Part2\lib\javax.transaction-1.1.1.v201105210645.jar;F:\Shivam\CMU Material\A - CMU Study Material\Semester 3\S1 - Distributed Systems for ISM\Projects\Project5\Project5Part2\lib\jetty-webapp-8.1.14.v20131031.jar;F:\Shivam\CMU Material\A - CMU Study Material\Semester 3\S1 - Distributed Systems for ISM\Projects\Project5\Project5Part2\lib\jetty-xml-8.1.14.v20131031.jar;F:\Shivam\CMU Material\A - CMU Study Material\Semester 3\S1 - Distributed Systems for ISM\Projects\Project5\Project5Part2\lib\jetty-servlet-8.1.14.v20131031.jar;F:\Shivam\CMU Material\A - CMU Study Material\Semester 3\S1 - Distributed Systems for ISM\Projects\Project5\Project5Part2\lib\jetty-jndi-8.1.14.v20131031.jar;F:\Shivam\CMU Material\A - CMU Study Material\Semester 3\S1 - Distributed Systems for ISM\Projects\Project5\Project5Part2\lib\javax.mail.glassfish-1.4.1.v201005082020.jar;F:\Shivam\CMU Material\A - CMU Study Material\Semester 3\S1 - Distributed Systems for ISM\Projects\Project5\Project5Part2\lib\javax.activation-1.1.0.v201105071233.jar;F:\Shivam\CMU Material\A - CMU Study Material\Semester 3\S1 - Distributed Systems for ISM\Projects\Project5\Project5Part2\lib\jetty-security-8.1.14.v20131031.jar;F:\Shivam\CMU Material\A - CMU Study Material\Semester 3\S1 - Distributed Systems for ISM\Projects\Project5\Project5Part2\lib\jetty-util-8.1.14.v20131031.jar;F:\Shivam\CMU Material\A - CMU Study Material\Semester 3\S1 - Distributed Systems for ISM\Projects\Project5\Project5Part2\lib\jetty-server-8.1.14.v20131031.jar;F:\Shivam\CMU Material\A - CMU Study Material\Semester 3\S1 - Distributed Systems for ISM\Projects\Project5\Project5Part2\lib\javax.servlet-3.0.0.v201112011016.jar;F:\Shivam\CMU Material\A - CMU Study Material\Semester 3\S1 - Distributed Systems for

ISM\Projects\Project5\Project5Part2\lib\jetty-continuation-8.1.14.v20131031.jar;F:\Shivam\CMU Material\A - CMU Study Material\Semester 3\S1 - Distributed Systems for ISM\Projects\Project5\Project5Part2\lib\jetty-http-8.1.14.v20131031.jar;F:\Shivam\CMU Material\A - CMU Study Material\Semester 3\S1 - Distributed Systems for ISM\Projects\Project5\Project5Part2\lib\jetty-io-8.1.14.v20131031.jar;F:\Shivam\CMU Material\A - CMU Study Material\Semester 3\S1 - Distributed Systems for ISM\Projects\Project5\Project5Part2\lib\guava-14.0.1.jar;F:\Shivam\CMU Material\A - CMU Study Material\Semester 3\S1 - Distributed Systems for ISM\Projects\Project5\Project5Part2\lib\commons-lang3-3.3.2.jar;F:\Shivam\CMU Material\A - CMU Study Material\Semester 3\S1 - Distributed Systems for ISM\Projects\Project5\Project5Part2\lib\jsr305-1.3.9.jar;F:\Shivam\CMU Material\A - CMU Study Material\Semester 3\S1 - Distributed Systems for ISM\Projects\Project5\Project5Part2\lib\slf4j-api-1.7.5.jar;F:\Shivam\CMU Material\A - CMU Study Material\Semester 3\S1 - Distributed Systems for ISM\Projects\Project5\Project5Part2\lib\jul-to-slf4j-1.7.5.jar;F:\Shivam\CMU Material\A - CMU Study Material\Semester 3\S1 - Distributed Systems for ISM\Projects\Project5\Project5Part2\lib\jcl-over-slf4j-1.7.5.jar;F:\Shivam\CMU Material\A - CMU Study Material\Semester 3\S1 - Distributed Systems for ISM\Projects\Project5\Project5Part2\lib\log4j-1.2.17.jar;F:\Shivam\CMU Material\A - CMU Study Material\Semester 3\S1 - Distributed Systems for ISM\Projects\Project5\Project5Part2\lib\slf4j-log4j12-1.7.5.jar;F:\Shivam\CMU Material\A - CMU Study Material\Semester 3\S1 - Distributed Systems for ISM\Projects\Project5\Project5Part2\lib\compress-lzf-1.0.0.jar;F:\Shivam\CMU Material\A - CMU Study Material\Semester 3\S1 - Distributed Systems for ISM\Projects\Project5\Project5Part2\lib\snappy-java-1.0.5.jar;F:\Shivam\CMU Material\A - CMU Study Material\Semester 3\S1 - Distributed Systems for ISM\Projects\Project5\Project5Part2\lib\chill_2.10-0.3.6.jar;F:\Shivam\CMU Material\A - CMU Study Material\Semester 3\S1 - Distributed Systems for ISM\Projects\Project5\Project5Part2\lib\kryo-2.21.jar;F:\Shivam\CMU Material\A - CMU Study Material\Semester 3\S1 - Distributed Systems for ISM\Projects\Project5\Project5Part2\lib\reflectasm-1.07-shaded.jar;F:\Shivam\CMU Material\A - CMU Study Material\Semester 3\S1 - Distributed Systems for ISM\Projects\Project5\Project5Part2\lib\minlog-1.2.jar;F:\Shivam\CMU Material\A - CMU Study Material\Semester 3\S1 - Distributed Systems for ISM\Projects\Project5\Project5Part2\lib\objenesis-1.2.jar;F:\Shivam\CMU Material\A - CMU Study Material\Semester 3\S1 - Distributed Systems for ISM\Projects\Project5\Project5Part2\lib\chill-java-0.3.6.jar;F:\Shivam\CMU Material\A - CMU Study Material\Semester 3\S1 - Distributed Systems for ISM\Projects\Project5\Project5Part2\lib\commons-net-2.2.jar;F:\Shivam\CMU Material\A - CMU Study Material\Semester 3\S1 - Distributed Systems for ISM\Projects\Project5\Project5Part2\lib\akka-remote_2.10-2.2.3-shaded-protobuf.jar;F:\Shivam\CMU Material\A - CMU Study Material\Semester 3\S1 - Distributed Systems for ISM\Projects\Project5\Project5Part2\lib\akka-actor_2.10-2.2.3-shaded-protobuf.jar;F:\Shivam\CMU Material\A - CMU Study Material\Semester 3\S1 - Distributed Systems for ISM\Projects\Project5\Project5Part2\lib\config-1.0.2.jar;F:\Shivam\CMU Material\A - CMU Study Material\Semester 3\S1 - Distributed Systems for ISM\Projects\Project5\Project5Part2\lib\netty-3.6.6.Final.jar;F:\Shivam\CMU Material\A - CMU Study Material\Semester 3\S1 - Distributed Systems for ISM\Projects\Project5\Project5Part2\lib\protobuf-java-2.4.1-shaded.jar;F:\Shivam\CMU Material\A - CMU Study Material\Semester 3\S1 - Distributed Systems for ISM\Projects\Project5\Project5Part2\lib\uncommons-maths-1.2.2a.jar;F:\Shivam\CMU Material\A - CMU Study Material\Semester 3\S1 - Distributed Systems for ISM\Projects\Project5\Project5Part2\lib\akka-slf4j_2.10-2.2.3-shaded-protobuf.jar;F:\Shivam\CMU

Material\A - CMU Study Material\Semester 3\S1 - Distributed Systems for ISM\Projects\Project5\Project5Part2\lib\scala-library-2.10.4.jar;F:\Shivam\CMU Material\A - CMU Study Material\Semester 3\S1 - Distributed Systems for ISM\Projects\Project5\Project5Part2\lib\json4s-jackson_2.10-3.2.6.jar;F:\Shivam\CMU Material\A - CMU Study Material\Semester 3\S1 - Distributed Systems for ISM\Projects\Project5\Project5Part2\lib\json4s-core_2.10-3.2.6.jar;F:\Shivam\CMU Material\A - CMU Study Material\Semester 3\S1 - Distributed Systems for ISM\Projects\Project5\Project5Part2\lib\json4s-ast_2.10-3.2.6.jar;F:\Shivam\CMU Material\A - CMU Study Material\Semester 3\S1 - Distributed Systems for ISM\Projects\Project5\Project5Part2\lib\paranamer-2.6.jar;F:\Shivam\CMU Material\A - CMU Study Material\Semester 3\S1 - Distributed Systems for ISM\Projects\Project5\Project5Part2\lib\scalap-2.10.0.jar;F:\Shivam\CMU Material\A - CMU Study Material\Semester 3\S1 - Distributed Systems for ISM\Projects\Project5\Project5Part2\lib\scala-compiler-2.10.0.jar;F:\Shivam\CMU Material\A - CMU Study Material\Semester 3\S1 - Distributed Systems for ISM\Projects\Project5\Project5Part2\lib\scala-reflect-2.10.0.jar;F:\Shivam\CMU Material\A - CMU Study Material\Semester 3\S1 - Distributed Systems for ISM\Projects\Project5\Project5Part2\lib\jackson-databind-2.3.0.jar;F:\Shivam\CMU Material\A - CMU Study Material\Semester 3\S1 - Distributed Systems for ISM\Projects\Project5\Project5Part2\lib\jackson-annotations-2.3.0.jar;F:\Shivam\CMU Material\A - CMU Study Material\Semester 3\S1 - Distributed Systems for ISM\Projects\Project5\Project5Part2\lib\jackson-core-2.3.0.jar;F:\Shivam\CMU Material\A - CMU Study Material\Semester 3\S1 - Distributed Systems for ISM\Projects\Project5\Project5Part2\lib\colt-1.2.0.jar;F:\Shivam\CMU Material\A - CMU Study Material\Semester 3\S1 - Distributed Systems for ISM\Projects\Project5\Project5Part2\lib\concurrent-1.3.4.jar;F:\Shivam\CMU Material\A - CMU Study Material\Semester 3\S1 - Distributed Systems for ISM\Projects\Project5\Project5Part2\lib\mesos-0.18.1-shaded-protobuf.jar;F:\Shivam\CMU Material\A - CMU Study Material\Semester 3\S1 - Distributed Systems for ISM\Projects\Project5\Project5Part2\lib\netty-all-4.0.17.Final.jar;F:\Shivam\CMU Material\A - CMU Study Material\Semester 3\S1 - Distributed Systems for ISM\Projects\Project5\Project5Part2\lib\stream-2.5.1.jar;F:\Shivam\CMU Material\A - CMU Study Material\Semester 3\S1 - Distributed Systems for ISM\Projects\Project5\Project5Part2\lib\metrics-core-3.0.0.jar;F:\Shivam\CMU Material\A - CMU Study Material\Semester 3\S1 - Distributed Systems for ISM\Projects\Project5\Project5Part2\lib\metrics-jvm-3.0.0.jar;F:\Shivam\CMU Material\A - CMU Study Material\Semester 3\S1 - Distributed Systems for ISM\Projects\Project5\Project5Part2\lib\metrics-json-3.0.0.jar;F:\Shivam\CMU Material\A - CMU Study Material\Semester 3\S1 - Distributed Systems for ISM\Projects\Project5\Project5Part2\lib\metrics-graphite-3.0.0.jar;F:\Shivam\CMU Material\A - CMU Study Material\Semester 3\S1 - Distributed Systems for ISM\Projects\Project5\Project5Part2\lib\tachyon-0.4.1-thrift.jar;F:\Shivam\CMU Material\A - CMU Study Material\Semester 3\S1 - Distributed Systems for ISM\Projects\Project5\Project5Part2\lib\ant-1.9.0.jar;F:\Shivam\CMU Material\A - CMU Study Material\Semester 3\S1 - Distributed Systems for ISM\Projects\Project5\Project5Part2\lib\ant-launcher-1.9.0.jar;F:\Shivam\CMU Material\A - CMU Study Material\Semester 3\S1 - Distributed Systems for ISM\Projects\Project5\Project5Part2\lib\commons-io-2.4.jar;F:\Shivam\CMU Material\A - CMU Study Material\Semester 3\S1 - Distributed Systems for ISM\Projects\Project5\Project5Part2\lib\pyrolite-2.0.1.jar;F:\Shivam\CMU Material\A - CMU Study

Material\Semester 3\S1 - Distributed Systems for ISM\Projects\Project5\Project5Part2\lib\py4j-0.8.1.jar" TempestAnalytics TheTempest.txt

22/12/10 21:08:52 INFO spark.SecurityManager: Changing view acls to: shiva

22/12/10 21:08:52 INFO spark.SecurityManager: SecurityManager: authentication disabled; ui acls disabled; users with view permissions: Set(shiva)

22/12/10 21:08:52 INFO slf4j.Slf4jLogger: Slf4jLogger started

22/12/10 21:08:52 INFO Remoting: Starting remoting

22/12/10 21:08:53 INFO Remoting: Remoting started; listening on addresses :[akka.tcp://spark@172.31.66.14:53119]

22/12/10 21:08:53 INFO Remoting: Remoting now listens on addresses: [akka.tcp://spark@172.31.66.14:53119]

22/12/10 21:08:53 INFO spark.SparkEnv: Registering MapOutputTracker

22/12/10 21:08:53 INFO spark.SparkEnv: Registering BlockManagerMaster

22/12/10 21:08:53 INFO storage.DiskBlockManager: Created local directory at C:\Users\shiva\AppData\Local\Temp\spark-local-20221210210853-f784

22/12/10 21:08:53 INFO storage.MemoryStore: MemoryStore started with capacity 2.1 GB.

22/12/10 21:08:53 INFO network.ConnectionManager: Bound socket to port 53122 with id = ConnectionManagerId(172.31.66.14,53122)

22/12/10 21:08:53 INFO storage.BlockManagerMaster: Trying to register BlockManager

22/12/10 21:08:53 INFO storage.BlockManagerInfo: Registering block manager 172.31.66.14:53122 with 2.1 GB RAM

22/12/10 21:08:53 INFO storage.BlockManagerMaster: Registered BlockManager

22/12/10 21:08:53 INFO spark.HttpServer: Starting HTTP Server

22/12/10 21:08:53 INFO server.Server: jetty-8.1.14.v20131031

22/12/10 21:08:53 INFO server.AbstractConnector: Started SocketConnector@0.0.0.0:53123

22/12/10 21:08:53 INFO broadcast.HttpBroadcast: Broadcast server started at http://172.31.66.14:53123

22/12/10 21:08:53 INFO spark.HttpFileServer: HTTP File server directory is C:\Users\shiva\AppData\Local\Temp\spark-bfc92403-de16-4534-8a2f-d8695068f044

22/12/10 21:08:53 INFO spark.HttpServer: Starting HTTP Server

22/12/10 21:08:53 INFO server.Server: jetty-8.1.14.v20131031

22/12/10 21:08:53 INFO server.AbstractConnector: Started SocketConnector@0.0.0.0:53124

22/12/10 21:08:53 INFO server.Server: jetty-8.1.14.v20131031

22/12/10 21:08:53 INFO server.AbstractConnector: Started SelectChannelConnector@0.0.0.0:4040

22/12/10 21:08:53 INFO ui.SparkUI: Started SparkUI at http://172.31.66.14:4040

22/12/10 21:08:54 INFO storage.MemoryStore: ensureFreeSpace(32816) called with curMem=0, maxMem=2274990489

22/12/10 21:08:54 INFO storage.MemoryStore: Block broadcast_0 stored as values to memory (estimated size 32.0 KB, free 2.1 GB)

22/12/10 21:08:54 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable

22/12/10 21:08:54 WARN snappy.LoadSnappy: Snappy native library not loaded

22/12/10 21:08:54 INFO mapred.FileInputFormat: Total input paths to process : 1

22/12/10 21:08:54 INFO spark.SparkContext: Starting job: count at TempestAnalytics.java:49

22/12/10 21:08:54 INFO scheduler.DAGScheduler: Got job 0 (count at TempestAnalytics.java:49) with 1 output partitions (allowLocal=false)

22/12/10 21:08:54 INFO scheduler.DAGScheduler: Final stage: Stage 0(count at TempestAnalytics.java:49)

22/12/10 21:08:54 INFO scheduler.DAGScheduler: Parents of final stage: List()

22/12/10 21:08:54 INFO scheduler.DAGScheduler: Missing parents: List()

22/12/10 21:08:54 INFO scheduler.DAGScheduler: Submitting Stage 0 (MappedRDD[1] at textFile at TempestAnalytics.java:42), which has no missing parents

22/12/10 21:08:54 INFO scheduler.DAGScheduler: Submitting 1 missing tasks from Stage 0 (MappedRDD[1] at textFile at TempestAnalytics.java:42)

22/12/10 21:08:54 INFO scheduler.TaskSchedulerImpl: Adding task set 0.0 with 1 tasks

22/12/10 21:08:54 INFO scheduler.TaskSetManager: Starting task 0.0:0 as TID 0 on executor localhost: localhost (PROCESS_LOCAL)

22/12/10 21:08:54 INFO scheduler.TaskSetManager: Serialized task 0.0:0 as 1809 bytes in 4 ms

22/12/10 21:08:54 INFO executor.Executor: Running task ID 0

22/12/10 21:08:54 INFO storage.BlockManager: Found block broadcast_0 locally

22/12/10 21:08:54 INFO rdd.HadoopRDD: Input split: file:/F:/Shivam/CMU Material/A - CMU Study Material/Semester 3/S1 - Distributed Systems for ISM/Projects/Project5/Project5Part2/TheTempest.txt:0+103311

22/12/10 21:08:54 INFO executor.Executor: Serialized size of result for 0 is 597

22/12/10 21:08:54 INFO executor.Executor: Sending result for 0 directly to driver

22/12/10 21:08:54 INFO executor.Executor: Finished task ID 0

22/12/10 21:08:54 INFO scheduler.TaskSetManager: Finished TID 0 in 192 ms on localhost (progress: 1/1)

22/12/10 21:08:54 INFO scheduler.DAGScheduler: Completed ResultTask(0, 0)

22/12/10 21:08:54 INFO scheduler.DAGScheduler: Stage 0 (count at TempestAnalytics.java:49) finished in 0.213 s

22/12/10 21:08:54 INFO scheduler.TaskSchedulerImpl: Removed TaskSet 0.0, whose tasks have all completed, from pool

22/12/10 21:08:54 INFO spark.SparkContext: Job finished: count at TempestAnalytics.java:49, took 0.3284624 s

Number of lines: 3466

22/12/10 21:08:54 INFO spark.SparkContext: Starting job: count at TempestAnalytics.java:56

22/12/10 21:08:54 INFO scheduler.DAGScheduler: Got job 1 (count at TempestAnalytics.java:56) with 1 output partitions (allowLocal=false)

22/12/10 21:08:54 INFO scheduler.DAGScheduler: Final stage: Stage 1(count at TempestAnalytics.java:56)

22/12/10 21:08:54 INFO scheduler.DAGScheduler: Parents of final stage: List()

22/12/10 21:08:54 INFO scheduler.DAGScheduler: Missing parents: List()

22/12/10 21:08:54 INFO scheduler.DAGScheduler: Submitting Stage 1 (FilteredRDD[3] at filter at TempestAnalytics.java:56), which has no missing parents

22/12/10 21:08:54 INFO scheduler.DAGScheduler: Submitting 1 missing tasks from Stage 1 (FilteredRDD[3] at filter at TempestAnalytics.java:56)

22/12/10 21:08:54 INFO scheduler.TaskSchedulerImpl: Adding task set 1.0 with 1 tasks

22/12/10 21:08:54 INFO scheduler.TaskSetManager: Starting task 1.0:0 as TID 1 on executor localhost: localhost (PROCESS_LOCAL)

22/12/10 21:08:54 INFO scheduler.TaskSetManager: Serialized task 1.0:0 as 2398 bytes in 1 ms

22/12/10 21:08:54 INFO executor.Executor: Running task ID 1

22/12/10 21:08:54 INFO storage.BlockManager: Found block broadcast_0 locally

22/12/10 21:08:54 INFO rdd.HadoopRDD: Input split: file:/F:/Shivam/CMU Material/A - CMU Study Material/Semester 3/S1 - Distributed Systems for ISM/Projects/Project5/Project5Part2/TheTempest.txt:0+103311

22/12/10 21:08:55 INFO executor.Executor: Serialized size of result for 1 is 597

22/12/10 21:08:55 INFO executor.Executor: Sending result for 1 directly to driver

22/12/10 21:08:55 INFO executor.Executor: Finished task ID 1

22/12/10 21:08:55 INFO scheduler.DAGScheduler: Completed ResultTask(1, 0)

22/12/10 21:08:55 INFO scheduler.DAGScheduler: Stage 1 (count at TempestAnalytics.java:56) finished in 0.109 s

22/12/10 21:08:55 INFO spark.SparkContext: Job finished: count at TempestAnalytics.java:56, took 0.1583267 s

22/12/10 21:08:55 INFO scheduler.TaskSetManager: Finished TID 1 in 107 ms on localhost (progress: 1/1)

22/12/10 21:08:55 INFO scheduler.TaskSchedulerImpl: Removed TaskSet 1.0, whose tasks have all completed, from pool

Number of words: 18167

22/12/10 21:08:55 INFO spark.SparkContext: Starting job: count at TempestAnalytics.java:63

22/12/10 21:08:55 INFO scheduler.DAGScheduler: Registering RDD 6 (distinct at TempestAnalytics.java:63)

22/12/10 21:08:55 INFO scheduler.DAGScheduler: Got job 2 (count at TempestAnalytics.java:63) with 1 output partitions (allowLocal=false)

22/12/10 21:08:55 INFO scheduler.DAGScheduler: Final stage: Stage 2(count at TempestAnalytics.java:63)

22/12/10 21:08:55 INFO scheduler.DAGScheduler: Parents of final stage: List(Stage 3)

22/12/10 21:08:55 INFO scheduler.DAGScheduler: Missing parents: List(Stage 3)

22/12/10 21:08:55 INFO scheduler.DAGScheduler: Submitting Stage 3 (MapPartitionsRDD[6] at distinct at TempestAnalytics.java:63), which has no missing parents

22/12/10 21:08:55 INFO scheduler.DAGScheduler: Submitting 1 missing tasks from Stage 3 (MapPartitionsRDD[6] at distinct at TempestAnalytics.java:63)

22/12/10 21:08:55 INFO scheduler.TaskSchedulerImpl: Adding task set 3.0 with 1 tasks

22/12/10 21:08:55 INFO scheduler.TaskSetManager: Starting task 3.0:0 as TID 2 on executor localhost: localhost (PROCESS_LOCAL)

22/12/10 21:08:55 INFO scheduler.TaskSetManager: Serialized task 3.0:0 as 2696 bytes in 1 ms

22/12/10 21:08:55 INFO executor.Executor: Running task ID 2

22/12/10 21:08:55 INFO storage.BlockManager: Found block broadcast_0 locally

22/12/10 21:08:55 INFO rdd.HadoopRDD: Input split: file:/F:/Shivam/CMU Material/A - CMU Study Material/Semester 3/S1 - Distributed Systems for ISM/Projects/Project5/Project5Part2/TheTempest.txt:0+103311

22/12/10 21:08:55 INFO executor.Executor: Serialized size of result for 2 is 782

22/12/10 21:08:55 INFO executor.Executor: Sending result for 2 directly to driver

22/12/10 21:08:55 INFO executor.Executor: Finished task ID 2

22/12/10 21:08:55 INFO scheduler.DAGScheduler: Completed ShuffleMapTask(3, 0)

22/12/10 21:08:55 INFO scheduler.TaskSetManager: Finished TID 2 in 332 ms on localhost (progress: 1/1)

22/12/10 21:08:55 INFO scheduler.DAGScheduler: Stage 3 (distinct at TempestAnalytics.java:63) finished in 0.315 s

22/12/10 21:08:55 INFO scheduler.TaskSchedulerImpl: Removed TaskSet 3.0, whose tasks have all completed, from pool

22/12/10 21:08:55 INFO scheduler.DAGScheduler: looking for newly runnable stages

22/12/10 21:08:55 INFO scheduler.DAGScheduler: running: Set()

22/12/10 21:08:55 INFO scheduler.DAGScheduler: waiting: Set(Stage 2)

22/12/10 21:08:55 INFO scheduler.DAGScheduler: failed: Set()

22/12/10 21:08:55 INFO scheduler.DAGScheduler: Missing parents for Stage 2: List()

22/12/10 21:08:55 INFO scheduler.DAGScheduler: Submitting Stage 2 (MappedRDD[9] at distinct at TempestAnalytics.java:63), which is now runnable

22/12/10 21:08:55 INFO scheduler.DAGScheduler: Submitting 1 missing tasks from Stage 2 (MappedRDD[9] at distinct at TempestAnalytics.java:63)

22/12/10 21:08:55 INFO scheduler.TaskSchedulerImpl: Adding task set 2.0 with 1 tasks

22/12/10 21:08:55 INFO scheduler.TaskSetManager: Starting task 2.0:0 as TID 3 on executor localhost: localhost (PROCESS_LOCAL)

22/12/10 21:08:55 INFO scheduler.TaskSetManager: Serialized task 2.0:0 as 2476 bytes in 0 ms

22/12/10 21:08:55 INFO executor.Executor: Running task ID 3

22/12/10 21:08:55 INFO storage.BlockManager: Found block broadcast_0 locally

22/12/10 21:08:55 INFO storage.BlockFetcherIterator$BasicBlockFetcherIterator: maxBytesInFlight: 50331648, targetRequestSize: 10066329

22/12/10 21:08:55 INFO storage.BlockFetcherIterator$BasicBlockFetcherIterator: Getting 1 non-empty blocks out of 1 blocks

22/12/10 21:08:55 INFO storage.BlockFetcherIterator$BasicBlockFetcherIterator: Started 0 remote fetches in 6 ms

22/12/10 21:08:55 INFO executor.Executor: Serialized size of result for 3 is 863

22/12/10 21:08:55 INFO executor.Executor: Sending result for 3 directly to driver

22/12/10 21:08:55 INFO executor.Executor: Finished task ID 3

22/12/10 21:08:55 INFO scheduler.DAGScheduler: Completed ResultTask(2, 0)

22/12/10 21:08:55 INFO scheduler.DAGScheduler: Stage 2 (count at TempestAnalytics.java:63) finished in 0.103 s

22/12/10 21:08:55 INFO spark.SparkContext: Job finished: count at TempestAnalytics.java:63, took 0.5451095 s

22/12/10 21:08:55 INFO scheduler.TaskSetManager: Finished TID 3 in 102 ms on localhost (progress: 1/1)

22/12/10 21:08:55 INFO scheduler.TaskSchedulerImpl: Removed TaskSet 2.0, whose tasks have all completed, from pool

22/12/10 21:08:55 INFO spark.SparkContext: Starting job: count at TempestAnalytics.java:73

22/12/10 21:08:55 INFO scheduler.DAGScheduler: Got job 3 (count at TempestAnalytics.java:73) with 1 output partitions (allowLocal=false)

22/12/10 21:08:55 INFO scheduler.DAGScheduler: Final stage: Stage 4(count at TempestAnalytics.java:73)

22/12/10 21:08:55 INFO scheduler.DAGScheduler: Parents of final stage: List()

22/12/10 21:08:55 INFO scheduler.DAGScheduler: Missing parents: List()

22/12/10 21:08:55 INFO scheduler.DAGScheduler: Submitting Stage 4 (FlatMappedRDD[10] at flatMap at TempestAnalytics.java:70), which has no missing parents

22/12/10 21:08:55 INFO scheduler.DAGScheduler: Submitting 1 missing tasks from Stage 4 (FlatMappedRDD[10] at flatMap at TempestAnalytics.java:70)

22/12/10 21:08:55 INFO scheduler.TaskSchedulerImpl: Adding task set 4.0 with 1 tasks

22/12/10 21:08:55 INFO scheduler.TaskSetManager: Starting task 4.0:0 as TID 4 on executor localhost: localhost (PROCESS_LOCAL)

22/12/10 21:08:55 INFO scheduler.TaskSetManager: Serialized task 4.0:0 as 2276 bytes in 0 ms

22/12/10 21:08:55 INFO executor.Executor: Running task ID 4

22/12/10 21:08:55 INFO storage.BlockManager: Found block broadcast_0 locally

22/12/10 21:08:55 INFO rdd.HadoopRDD: Input split: file:/F:/Shivam/CMU Material/A - CMU Study Material/Semester 3/S1 - Distributed Systems for ISM/Projects/Project5/Project5Part2/TheTempest.txt:0+103311

Number of distinct words: 3595

22/12/10 21:08:55 INFO executor.Executor: Serialized size of result for 4 is 597

22/12/10 21:08:55 INFO executor.Executor: Sending result for 4 directly to driver

22/12/10 21:08:55 INFO executor.Executor: Finished task ID 4

22/12/10 21:08:55 INFO scheduler.DAGScheduler: Completed ResultTask(4, 0)

22/12/10 21:08:55 INFO scheduler.DAGScheduler: Stage 4 (count at TempestAnalytics.java:73) finished in 0.057 s

22/12/10 21:08:55 INFO scheduler.TaskSetManager: Finished TID 4 in 57 ms on localhost (progress: 1/1)

22/12/10 21:08:55 INFO scheduler.TaskSchedulerImpl: Removed TaskSet 4.0, whose tasks have all completed, from pool

22/12/10 21:08:55 INFO spark.SparkContext: Job finished: count at TempestAnalytics.java:73, took 0.0655384 s

22/12/10 21:08:55 INFO spark.SparkContext: Starting job: count at TempestAnalytics.java:80

Number of symbols: 97132

22/12/10 21:08:55 INFO scheduler.DAGScheduler: Registering RDD 12 (distinct at TempestAnalytics.java:80)

22/12/10 21:08:55 INFO scheduler.DAGScheduler: Got job 4 (count at TempestAnalytics.java:80) with 1 output partitions (allowLocal=false)

22/12/10 21:08:55 INFO scheduler.DAGScheduler: Final stage: Stage 5(count at TempestAnalytics.java:80)

22/12/10 21:08:55 INFO scheduler.DAGScheduler: Parents of final stage: List(Stage 6)

22/12/10 21:08:55 INFO scheduler.DAGScheduler: Missing parents: List(Stage 6)

22/12/10 21:08:55 INFO scheduler.DAGScheduler: Submitting Stage 6 (MapPartitionsRDD[12] at distinct at TempestAnalytics.java:80), which has no missing parents

22/12/10 21:08:55 INFO scheduler.DAGScheduler: Submitting 1 missing tasks from Stage 6 (MapPartitionsRDD[12] at distinct at TempestAnalytics.java:80)

22/12/10 21:08:55 INFO scheduler.TaskSchedulerImpl: Adding task set 6.0 with 1 tasks

22/12/10 21:08:55 INFO scheduler.TaskSetManager: Starting task 6.0:0 as TID 5 on executor localhost: localhost (PROCESS_LOCAL)

22/12/10 21:08:55 INFO scheduler.TaskSetManager: Serialized task 6.0:0 as 2578 bytes in 0 ms

22/12/10 21:08:55 INFO executor.Executor: Running task ID 5

22/12/10 21:08:55 INFO storage.BlockManager: Found block broadcast_0 locally

22/12/10 21:08:55 INFO rdd.HadoopRDD: Input split: file:/F:/Shivam/CMU Material/A - CMU Study Material/Semester 3/S1 - Distributed Systems for ISM/Projects/Project5/Project5Part2/TheTempest.txt:0+103311

22/12/10 21:08:55 INFO executor.Executor: Serialized size of result for 5 is 782

22/12/10 21:08:55 INFO executor.Executor: Sending result for 5 directly to driver

22/12/10 21:08:55 INFO executor.Executor: Finished task ID 5

22/12/10 21:08:55 INFO scheduler.TaskSetManager: Finished TID 5 in 167 ms on localhost (progress: 1/1)

22/12/10 21:08:55 INFO scheduler.TaskSchedulerImpl: Removed TaskSet 6.0, whose tasks have all completed, from pool

22/12/10 21:08:55 INFO scheduler.DAGScheduler: Completed ShuffleMapTask(6, 0)

22/12/10 21:08:55 INFO scheduler.DAGScheduler: Stage 6 (distinct at TempestAnalytics.java:80) finished in 0.172 s

22/12/10 21:08:55 INFO scheduler.DAGScheduler: looking for newly runnable stages

22/12/10 21:08:55 INFO scheduler.DAGScheduler: running: Set()

22/12/10 21:08:55 INFO scheduler.DAGScheduler: waiting: Set(Stage 5)

22/12/10 21:08:55 INFO scheduler.DAGScheduler: failed: Set()

22/12/10 21:08:55 INFO scheduler.DAGScheduler: Missing parents for Stage 5: List()

22/12/10 21:08:55 INFO scheduler.DAGScheduler: Submitting Stage 5 (MappedRDD[15] at distinct at TempestAnalytics.java:80), which is now runnable

22/12/10 21:08:55 INFO scheduler.DAGScheduler: Submitting 1 missing tasks from Stage 5 (MappedRDD[15] at distinct at TempestAnalytics.java:80)

22/12/10 21:08:55 INFO scheduler.TaskSchedulerImpl: Adding task set 5.0 with 1 tasks

22/12/10 21:08:55 INFO scheduler.TaskSetManager: Starting task 5.0:0 as TID 6 on executor localhost: localhost (PROCESS_LOCAL)

22/12/10 21:08:55 INFO scheduler.TaskSetManager: Serialized task 5.0:0 as 2368 bytes in 1 ms

22/12/10 21:08:55 INFO executor.Executor: Running task ID 6

22/12/10 21:08:55 INFO storage.BlockManager: Found block broadcast_0 locally

22/12/10 21:08:55 INFO storage.BlockFetcherIterator$BasicBlockFetcherIterator: maxBytesInFlight: 50331648, targetRequestSize: 10066329

22/12/10 21:08:55 INFO storage.BlockFetcherIterator$BasicBlockFetcherIterator: Getting 1 non-empty blocks out of 1 blocks

22/12/10 21:08:55 INFO storage.BlockFetcherIterator$BasicBlockFetcherIterator: Started 0 remote fetches in 1 ms

22/12/10 21:08:55 INFO executor.Executor: Serialized size of result for 6 is 863

22/12/10 21:08:55 INFO executor.Executor: Sending result for 6 directly to driver

22/12/10 21:08:55 INFO scheduler.TaskSetManager: Finished TID 6 in 19 ms on localhost (progress: 1/1)

22/12/10 21:08:55 INFO scheduler.TaskSchedulerImpl: Removed TaskSet 5.0, whose tasks have all completed, from pool

22/12/10 21:08:55 INFO scheduler.DAGScheduler: Completed ResultTask(5, 0)

22/12/10 21:08:55 INFO scheduler.DAGScheduler: Stage 5 (count at TempestAnalytics.java:80) finished in 0.022 s

22/12/10 21:08:55 INFO spark.SparkContext: Job finished: count at TempestAnalytics.java:80, took 0.2191118 s

22/12/10 21:08:55 INFO executor.Executor: Finished task ID 6

22/12/10 21:08:55 INFO spark.SparkContext: Starting job: count at TempestAnalytics.java:87

22/12/10 21:08:55 INFO scheduler.DAGScheduler: Registering RDD 18 (distinct at TempestAnalytics.java:87)

22/12/10 21:08:55 INFO scheduler.DAGScheduler: Got job 5 (count at TempestAnalytics.java:87) with 1 output partitions (allowLocal=false)

22/12/10 21:08:55 INFO scheduler.DAGScheduler: Final stage: Stage 7(count at TempestAnalytics.java:87)

22/12/10 21:08:55 INFO scheduler.DAGScheduler: Parents of final stage: List(Stage 8)

Number of distinct symbols: 77

22/12/10 21:08:55 INFO scheduler.DAGScheduler: Missing parents: List(Stage 8)

22/12/10 21:08:55 INFO scheduler.DAGScheduler: Submitting Stage 8 (MapPartitionsRDD[18] at distinct at TempestAnalytics.java:87), which has no missing parents

22/12/10 21:08:55 INFO scheduler.DAGScheduler: Submitting 1 missing tasks from Stage 8 (MapPartitionsRDD[18] at distinct at TempestAnalytics.java:87)

22/12/10 21:08:55 INFO scheduler.TaskSchedulerImpl: Adding task set 8.0 with 1 tasks

22/12/10 21:08:55 INFO scheduler.TaskSetManager: Starting task 8.0:0 as TID 7 on executor localhost: localhost (PROCESS_LOCAL)

22/12/10 21:08:55 INFO scheduler.TaskSetManager: Serialized task 8.0:0 as 2696 bytes in 0 ms

22/12/10 21:08:55 INFO executor.Executor: Running task ID 7

22/12/10 21:08:55 INFO storage.BlockManager: Found block broadcast_0 locally

22/12/10 21:08:55 INFO rdd.HadoopRDD: Input split: file:/F:/Shivam/CMU Material/A - CMU Study Material/Semester 3/S1 - Distributed Systems for ISM/Projects/Project5/Project5Part2/TheTempest.txt:0+103311

22/12/10 21:08:56 INFO spark.ContextCleaner: Cleaned shuffle 1

22/12/10 21:08:56 INFO spark.ContextCleaner: Cleaned shuffle 0

22/12/10 21:08:56 INFO storage.ShuffleBlockManager: Deleted all files for shuffle 1

22/12/10 21:08:56 INFO storage.ShuffleBlockManager: Deleted all files for shuffle 0

22/12/10 21:08:56 INFO executor.Executor: Serialized size of result for 7 is 782

22/12/10 21:08:56 INFO executor.Executor: Sending result for 7 directly to driver

22/12/10 21:08:56 INFO scheduler.TaskSetManager: Finished TID 7 in 237 ms on localhost (progress: 1/1)

22/12/10 21:08:56 INFO scheduler.TaskSchedulerImpl: Removed TaskSet 8.0, whose tasks have all completed, from pool

22/12/10 21:08:56 INFO scheduler.DAGScheduler: Completed ShuffleMapTask(8, 0)

22/12/10 21:08:56 INFO scheduler.DAGScheduler: Stage 8 (distinct at TempestAnalytics.java:87) finished in 0.240 s

22/12/10 21:08:56 INFO scheduler.DAGScheduler: looking for newly runnable stages

22/12/10 21:08:56 INFO scheduler.DAGScheduler: running: Set()

22/12/10 21:08:56 INFO scheduler.DAGScheduler: waiting: Set(Stage 7)

22/12/10 21:08:56 INFO scheduler.DAGScheduler: failed: Set()

22/12/10 21:08:56 INFO executor.Executor: Finished task ID 7

22/12/10 21:08:56 INFO scheduler.DAGScheduler: Missing parents for Stage 7: List()

22/12/10 21:08:56 INFO scheduler.DAGScheduler: Submitting Stage 7 (MappedRDD[21] at distinct at TempestAnalytics.java:87), which is now runnable

22/12/10 21:08:56 INFO scheduler.DAGScheduler: Submitting 1 missing tasks from Stage 7 (MappedRDD[21] at distinct at TempestAnalytics.java:87)

22/12/10 21:08:56 INFO scheduler.TaskSchedulerImpl: Adding task set 7.0 with 1 tasks

22/12/10 21:08:56 INFO scheduler.TaskSetManager: Starting task 7.0:0 as TID 8 on executor localhost: localhost (PROCESS_LOCAL)

22/12/10 21:08:56 INFO scheduler.TaskSetManager: Serialized task 7.0:0 as 2476 bytes in 0 ms

22/12/10 21:08:56 INFO executor.Executor: Running task ID 8

22/12/10 21:08:56 INFO storage.BlockManager: Found block broadcast_0 locally

22/12/10 21:08:56 INFO storage.BlockFetcherIterator$BasicBlockFetcherIterator: maxBytesInFlight: 50331648, targetRequestSize: 10066329

22/12/10 21:08:56 INFO storage.BlockFetcherIterator$BasicBlockFetcherIterator: Getting 1 non-empty blocks out of 1 blocks

22/12/10 21:08:56 INFO storage.BlockFetcherIterator$BasicBlockFetcherIterator: Started 0 remote fetches in 1 ms

22/12/10 21:08:56 INFO executor.Executor: Serialized size of result for 8 is 863

22/12/10 21:08:56 INFO executor.Executor: Sending result for 8 directly to driver

22/12/10 21:08:56 INFO executor.Executor: Finished task ID 8

22/12/10 21:08:56 INFO scheduler.TaskSetManager: Finished TID 8 in 27 ms on localhost (progress: 1/1)

22/12/10 21:08:56 INFO scheduler.TaskSchedulerImpl: Removed TaskSet 7.0, whose tasks have all completed, from pool

22/12/10 21:08:56 INFO scheduler.DAGScheduler: Completed ResultTask(7, 0)

22/12/10 21:08:56 INFO scheduler.DAGScheduler: Stage 7 (count at TempestAnalytics.java:87) finished in 0.029 s

22/12/10 21:08:56 INFO spark.SparkContext: Job finished: count at TempestAnalytics.java:87, took 0.312917 s

Number of distinct letters: 51

Enter a word to search in TheTempest.txt (Enter "#$@Done@$#" to stop):

love

22/12/10 21:08:59 INFO spark.SparkContext: Starting job: foreach at TempestAnalytics.java:119

22/12/10 21:08:59 INFO scheduler.DAGScheduler: Got job 6 (foreach at TempestAnalytics.java:119) with 1 output partitions (allowLocal=false)

22/12/10 21:08:59 INFO scheduler.DAGScheduler: Final stage: Stage 9(foreach at TempestAnalytics.java:119)

22/12/10 21:08:59 INFO scheduler.DAGScheduler: Parents of final stage: List()

22/12/10 21:08:59 INFO scheduler.DAGScheduler: Missing parents: List()

22/12/10 21:08:59 INFO scheduler.DAGScheduler: Submitting Stage 9 (FilteredRDD[23] at filter at TempestAnalytics.java:116), which has no missing parents

22/12/10 21:08:59 INFO scheduler.DAGScheduler: Submitting 1 missing tasks from Stage 9 (FilteredRDD[23] at filter at TempestAnalytics.java:116)

22/12/10 21:08:59 INFO scheduler.TaskSchedulerImpl: Adding task set 9.0 with 1 tasks

22/12/10 21:08:59 INFO scheduler.TaskSetManager: Starting task 9.0:0 as TID 9 on executor localhost: localhost (PROCESS_LOCAL)

22/12/10 21:08:59 INFO scheduler.TaskSetManager: Serialized task 9.0:0 as 2466 bytes in 1 ms

22/12/10 21:08:59 INFO executor.Executor: Running task ID 9

22/12/10 21:08:59 INFO storage.BlockManager: Found block broadcast_0 locally

22/12/10 21:08:59 INFO rdd.HadoopRDD: Input split: file:/F:/Shivam/CMU Material/A - CMU Study Material/Semester 3/S1 - Distributed Systems for ISM/Projects/Project5/Project5Part2/TheTempest.txt:0+103311

22/12/10 21:08:59 INFO executor.Executor: Serialized size of result for 9 is 559

22/12/10 21:08:59 INFO executor.Executor: Sending result for 9 directly to driver

22/12/10 21:08:59 INFO executor.Executor: Finished task ID 9

22/12/10 21:08:59 INFO scheduler.DAGScheduler: Completed ResultTask(9, 0)

22/12/10 21:08:59 INFO scheduler.TaskSetManager: Finished TID 9 in 20 ms on localhost (progress: 1/1)

22/12/10 21:08:59 INFO scheduler.DAGScheduler: Stage 9 (foreach at TempestAnalytics.java:119) finished in 0.020 s

22/12/10 21:08:59 INFO scheduler.TaskSchedulerImpl: Removed TaskSet 9.0, whose tasks have all completed, from pool

22/12/10 21:08:59 INFO spark.SparkContext: Job finished: foreach at TempestAnalytics.java:119, took 0.0288018 s

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,

Enter a word to search in TheTempest.txt (Enter "#$@Done@$#" to stop):

hate

22/12/10 21:09:01 INFO spark.SparkContext: Starting job: foreach at TempestAnalytics.java:119

22/12/10 21:09:01 INFO scheduler.DAGScheduler: Got job 7 (foreach at TempestAnalytics.java:119) with 1 output partitions (allowLocal=false)

22/12/10 21:09:01 INFO scheduler.DAGScheduler: Final stage: Stage 10(foreach at TempestAnalytics.java:119)

22/12/10 21:09:01 INFO scheduler.DAGScheduler: Parents of final stage: List()

22/12/10 21:09:01 INFO scheduler.DAGScheduler: Missing parents: List()

22/12/10 21:09:01 INFO scheduler.DAGScheduler: Submitting Stage 10 (FilteredRDD[25] at filter at TempestAnalytics.java:116), which has no missing parents

22/12/10 21:09:01 INFO scheduler.DAGScheduler: Submitting 1 missing tasks from Stage 10 (FilteredRDD[25] at filter at TempestAnalytics.java:116)

22/12/10 21:09:01 INFO scheduler.TaskSchedulerImpl: Adding task set 10.0 with 1 tasks

22/12/10 21:09:01 INFO scheduler.TaskSetManager: Starting task 10.0:0 as TID 10 on executor localhost: localhost (PROCESS_LOCAL)

22/12/10 21:09:01 INFO scheduler.TaskSetManager: Serialized task 10.0:0 as 2466 bytes in 1 ms

22/12/10 21:09:01 INFO executor.Executor: Running task ID 10

22/12/10 21:09:01 INFO storage.BlockManager: Found block broadcast_0 locally

22/12/10 21:09:01 INFO rdd.HadoopRDD: Input split: file:/F:/Shivam/CMU Material/A - CMU Study Material/Semester 3/S1 - Distributed Systems for ISM/Projects/Project5/Project5Part2/TheTempest.txt:0+103311

22/12/10 21:09:01 INFO executor.Executor: Serialized size of result for 10 is 559

22/12/10 21:09:01 INFO executor.Executor: Sending result for 10 directly to driver

22/12/10 21:09:01 INFO executor.Executor: Finished task ID 10

22/12/10 21:09:01 INFO scheduler.DAGScheduler: Completed ResultTask(10, 0)

22/12/10 21:09:01 INFO scheduler.DAGScheduler: Stage 10 (foreach at TempestAnalytics.java:119) finished in 0.011 s

22/12/10 21:09:01 INFO spark.SparkContext: Job finished: foreach at TempestAnalytics.java:119, took 0.017549 s

22/12/10 21:09:01 INFO scheduler.TaskSetManager: Finished TID 10 in 10 ms on localhost (progress: 1/1)

22/12/10 21:09:01 INFO scheduler.TaskSchedulerImpl: Removed TaskSet 10.0, whose tasks have all completed, from pool

Shall make me chide thee, if not hate thee. What,

One spirit to command. They all do hate him

To make this contract grow; but barren hate,

That you shall hate it both. Therefore take heed,

Enter a word to search in TheTempest.txt (Enter "#$@Done@$#" to stop):

#$@Done@$#

Quitting! Thank you :)

Process finished with exit code 0

Heinz Cluster ID – student106

Heinz Cluster Password – CMUDS@student106