

Lab – 5

CE004

20CEUOD004

```
MutualFriends.java
import java.io.IOException;
import org.apache.hadoop.conf.Configuration;
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.io.*;
import org.apache.hadoop.mapreduce.Job;
import org.apache.hadoop.mapreduce.Mapper;
import org.apache.hadoop.mapreduce.Reducer;
import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;
import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;
import org.apache.hadoop.util.GenericOptionsParser;
import java.util.*;

public class MutualFriends {
    public static class Map
    extends Mapper < LongWritable, Text, Text, Text > {
        Text user = new Text();
        Text friends = new Text();

        public void map(LongWritable key, Text value, Context context)
        throws IOException,
        InterruptedException {
            /*Splitting each line into user and corresponding friend
list*/

            String[] split = value.toString().split("\\t");
            /*split[0] is user and split[1] is friend list */
            String userId = split[0];
            if (split.length == 1) {
                return;
            }
            String[] friendIds = split[1].split(",");
            for (String friend: friendIds) {
                if (userId.equals(friend)) {
                    continue;
                }
                String userKey = (Integer.parseInt(userId) <
                    Integer.parseInt(friend)) ? userId + "," + friend : friend + "," + user
Id;

                String regex = "((\\b" + friend + "[^\\w]+)\\b,?" + friend +
                    "$)";
                friends.set(split[1].replaceAll(regex, ""));
                user.set(userKey);
                context.write(user, friends);
            }
        }
    }
}
```

```

}

public static class Reduce
extends Reducer < Text, Text, Text, Text > {

    private String matchingFriends(String firstList, String secondList) {

        if (firstList == null || secondList == null) {
            return null;
        }

        String[] list1 = firstList.split(",");
        String[] list2 = secondList.split(",");

        LinkedHashSet < String > firstSet = new LinkedHashSet();
        for (String user: list1) {
            firstSet.add(user);
        }
        /*Retaining sort order*/
        LinkedHashSet < String > secondSet = new LinkedHashSet();
        for (String user: list2) {
            secondSet.add(user);
        }
        firstSet.retainAll(secondSet);
        /*Keeping only the matched friends*/
        return firstSet.toString().replaceAll("\\[|\\]", "");
    }

    public void reduce(Text key, Iterable < Text > values, Context context)
    throws IOException,
    InterruptedException {

        String[] friendsList = new String[2];
        int index = 0;

        for (Text value: values) {
            friendsList[index++] = value.toString();
        }
        String mutualFriends =
            matchingFriends(friendsList[0], friendsList[1]);
        if (mutualFriends != null && mutualFriends.length() != 0) {
            context.write(key, new Text(mutualFriends));
        }
    }
}

// Driver program
public static void main(String[] args) throws Exception {
    Configuration conf = new Configuration();
    String[] otherArgs = new GenericOptionsParser(conf,
        args).getRemainingArgs();
    // get all args

```

```

if (otherArgs.length != 2) {
    System.err.println("Usage: MutualFriends <input file name> <
        output file name > ");
    System.exit(2);
}
// create a job with name "mutual friends"
@SuppressWarnings("deprecation")
Job job = new Job(conf, "mutualfriends");
job.setJarByClass(MutualFriends.class);
job.setMapperClass(Map.class);
job.setReducerClass(Reduce.class);
// uncomment the following line to add the Combiner
job.setCombinerClass(Reduce.class);
// set output key type
job.setOutputKeyClass(Text.class);
// set output value type
job.setOutputValueClass(Text.class);
//set the HDFS path of the input data
FileInputFormat.addInputPath(job, new Path(otherArgs[0]));
// set the HDFS path for the output
FileOutputFormat.setOutputPath(job, new Path(otherArgs[1]));
//Wait till job completion
System.exit(job.waitForCompletion(true) ? 0 : 1);
}

```

Sample.txt

```

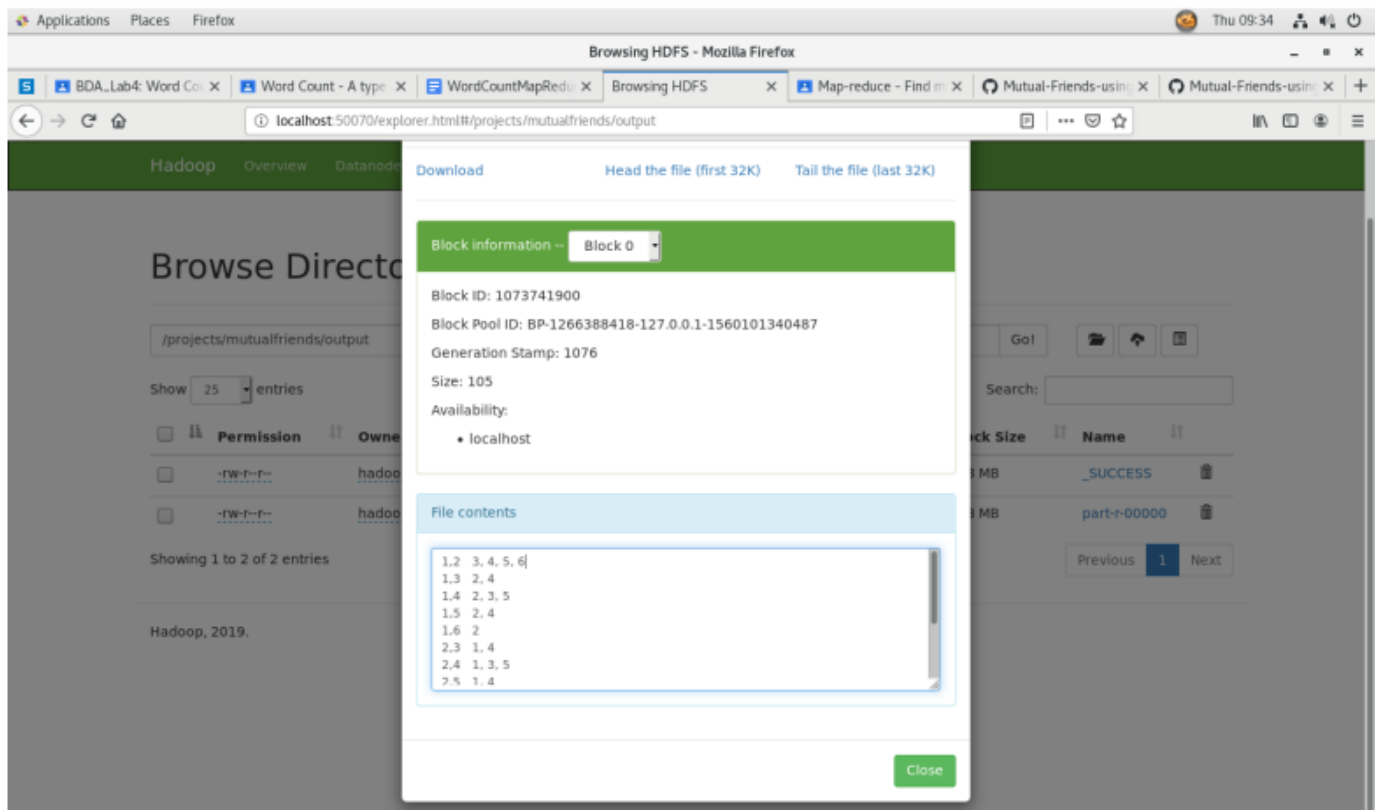
1 2, 3, 4, 5, 6
2 1, 3, 4, 5, 6
3 1, 2, 4
4 1, 2, 3, 5
5 1, 2, 4
6 1, 2

```

```
File Edit View Search Terminal Tabs Help
hadoop@hadoop-clone:~/Desktop
hadoop@hadoop-clone:/opt/hadoop

[hadoop@hadoop-clone Desktop]$ cd /opt/
Adobe/          hadoop/          lampp/          rh/             sublime_text/
eclipse/        hbase/          ns-allinone-2.35/ sqldeveloper/   TIB_js-studiocomm_6.8.0/
google/         hive/           pig/            sts/

[hadoop@hadoop-clone Desktop]$ cd /opt/hadoop/
[hadoop@hadoop-clone hadoop]$ df -mkdir -p
df      dfctest  dfu-tool
[hadoop@hadoop-clone hadoop]$ df -mkdir -p
df      dfctest  dfu-tool
[hadoop@hadoop-clone hadoop]$ hdfs -mkdir -p
hdfs     hdfs.cmd
[hadoop@hadoop-clone hadoop]$ hdfs dfs -mkdir -p /projects/mutualfriends/input
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
2022-08-25 09:16:02,685 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
[hadoop@hadoop-clone hadoop]$ hdfs dfs -put *.txt /projects/mutualfriends/input
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
2022-08-25 09:19:54,555 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
[hadoop@hadoop-clone hadoop]$ hdfs dfs -chmod -R 775 /projects
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
2022-08-25 09:20:30,528 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
[hadoop@hadoop-clone hadoop]$
```



← → ↺ 🏠

localhost:50070/explorer.html#/projects/mutualfriends/output

🔍 🗑️ ⌂

Hadoop Overview Datanode

Browse Directory

/projects/mutualfriends/output

Show 25 entries

| Permission | Owner |
|------------|--------|
| -rwxr-xr-x | hadoop |
| -rwxr-xr-x | hadoop |

Showing 1 to 2 of 2 entries

Hadoop, 2019.

Download Head the file (first 32K) Tail the file (last 32K)

Block information -- Block 0

Block ID: 1073741900

Block Pool ID: BP-1266388418-127.0.0.1-1560101340487

Generation Stamp: 1076

Size: 105

Availability:

- localhost

File contents

```
1.0 4
2.3 1, 4
2.4 1, 3, 5
2.5 1, 4
2.6 1
3.4 1, 2
4.5 1, 2
```

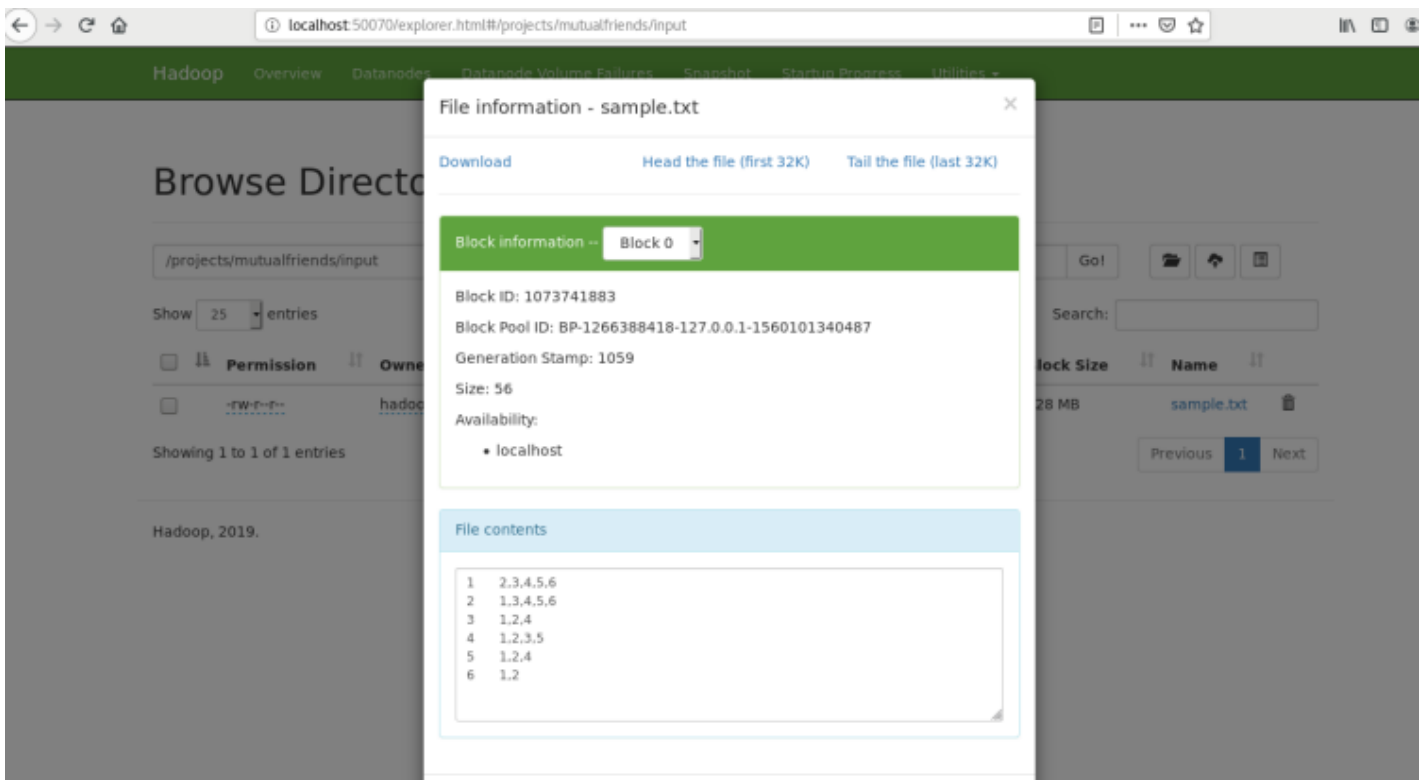
Close

File Edit View Search Terminal Tabs Help

hadoop@hadoop-clone:~/Desktop

hadoop@hadoop-clone:~/Desktop

```
[hadoop@hadoop-clone Desktop]$ cd /opt/
Adobe/          hadoop/
eclipse/        hbase/
google/         hive/
[hadoop@hadoop-clone Desktop]$ cd /opt/hadoop/
[hadoop@hadoop-clone hadoop]$ df -mkdir -p
df      dfest  dfu-tool
[hadoop@hadoop-clone hadoop]$ df -mkdir -p
df      dfest  dfu-tool
[hadoop@hadoop-clone hadoop]$ hdfs -mkdir -p
hdfs     hdfs.cmd
[hadoop@hadoop-clone hadoop]$ hdfs dfs -mkdir -p /projects/mutualfriends/input
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
2022-08-25 09:16:02,685 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
[hadoop@hadoop-clone hadoop]$ hdfs dfs -put *.txt /projects/mutualfriends/input
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
2022-08-25 09:19:54,555 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
[hadoop@hadoop-clone hadoop]$ hdfs dfs -chmod -R 775 /projects
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
2022-08-25 09:20:30,528 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
[hadoop@hadoop-clone hadoop]$ hadoop jar mf.jar MutualFriends /projects/mutualfriends/input /projects/mutualfriends/output
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
JAR does not exist or is not a normal file: /opt/hadoop/mf.jar
[hadoop@hadoop-clone hadoop]$ cd /home/hadoop/Desktop/
[hadoop@hadoop-clone Desktop]$ hadoop jar mf.jar MutualFriends /projects/mutualfriends/input /projects/mutualfriends/output
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
2022-08-25 09:23:04,502 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
2022-08-25 09:23:04,895 INFO client.RMProxy: Connecting to ResourceManager at hadoop-clone/127.0.0.1:8032
2022-08-25 09:23:05,253 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/hadoop/.staging/job_1661397719090_0001
2022-08-25 09:23:05,883 INFO input.FileInputFormat: Total input files to process : 3
2022-08-25 09:23:06,469 INFO mapreduce.JobSubmitter: number of splits:3
2022-08-25 09:23:06,572 INFO Configuration.deprecation: yarn.resourcemanager.system-metrics-publisher.enabled is deprecated. Instead, use yarn.system-metrics-publisher.enabled
2022-08-25 09:23:06,644 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1661397719090_0001
2022-08-25 09:23:06,645 INFO mapreduce.JobSubmitter: Executing with tokens: []
2022-08-25 09:23:06,752 INFO conf.Configuration: resource-types.xml not found
2022-08-25 09:23:06,752 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2022-08-25 09:23:06,902 INFO impl.YarnClientImpl: Submitted application application_1661397719090_0001
2022-08-25 09:23:06,959 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1661397719090_0001/
2022-08-25 09:23:06,960 INFO mapreduce.Job: Running job: job_1661397719090_0001
```



(2)

```
[hadoop@hadoop-clone ~]$ cd Desktop/
[hadoop@hadoop-clone Desktop]$ ls
BookCount.java~  MutualFriends.class  MutualFriends$Map.class    sample.txt
mf.jar           MutualFriends.java    MutualFriends$Reduce.class
[hadoop@hadoop-clone Desktop]$ hadoop com.sun.tools.javac.Main MutualFriends.java
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
Note: MutualFriends.java uses unchecked or unsafe operations.
Note: Recompile with -Xlint:unchecked for details.
[hadoop@hadoop-clone Desktop]$ jar cf wc.jar MutualFriends*.class
```

```
[hadoop@hadoop-clone ~]$ cd Desktop/
[hadoop@hadoop-clone Desktop]$ ls
BookCount.java~ MutualFriends.class MutualFriends$Map.class sample.txt
mf.jar MutualFriends.java MutualFriends$Reduce.class
[hadoop@hadoop-clone Desktop]$ hadoop com.sun.tools.javac.Main MutualFriends.java
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
Note: MutualFriends.java uses unchecked or unsafe operations.
Note: Recompile with -Xlint:unchecked for details.
[hadoop@hadoop-clone Desktop]$ jar cf wc.jar MutualFriends*.class
[hadoop@hadoop-clone Desktop]$ hdfs dfs -put *.txt /projects/mutualfriends/input
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
2022-08-25 10:10:00,753 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
[hadoop@hadoop-clone Desktop]$ hadoop jar mf.jar MutualFriends /projects/mutualfriends/input /projects/mutualfriends/output
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
2022-08-25 10:11:14,850 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
2022-08-25 10:11:15,255 INFO client.RMProxy: Connecting to ResourceManager at hadoop-clone/127.0.0.1:8032
2022-08-25 10:11:15,609 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/hadoop/.staging/job_1661401953007_0001/
2022-08-25 10:11:15,954 INFO input.FileInputFormat: Total input files to process : 1
2022-08-25 10:11:16,117 INFO mapreduce.JobSubmitter: number of splits:1
2022-08-25 10:11:16,181 INFO Configuration.deprecation: yarn.resourcemanager.system-metrics-publisher.enabled is deprecated. Instead, use yarn.system-metrics-publisher.enabled
2022-08-25 10:11:16,259 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1661401953007_0001
2022-08-25 10:11:16,260 INFO mapreduce.JobSubmitter: Executing with tokens: []
2022-08-25 10:11:16,373 INFO conf.Configuration: resource-types.xml not found
2022-08-25 10:11:16,373 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2022-08-25 10:11:16,548 INFO impl.YarnClientImpl: Submitted application application_1661401953007_0001
2022-08-25 10:11:16,616 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1661401953007_0001/
2022-08-25 10:11:16,617 INFO mapreduce.Job: Running job: job_1661401953007_0001
2022-08-25 10:11:22,677 INFO mapreduce.Job: Job job_1661401953007_0001 running in uber mode : false
2022-08-25 10:11:22,679 INFO mapreduce.Job: map 0% reduce 0%
2022-08-25 10:11:25,729 INFO mapreduce.Job: map 100% reduce 0%
2022-08-25 10:11:29,760 INFO mapreduce.Job: map 100% reduce 100%
2022-08-25 10:11:30,780 INFO mapreduce.Job: Job job_1661401953007_0001 completed successfully
2022-08-25 10:11:30,840 INFO mapreduce.Job: Counters: 54
```

The screenshot shows the Hadoop web interface at localhost:50070. A modal window titled "File information - part-r-00000" is open, displaying details for a specific file block. The background shows the "Browse Directories" page for the path /projects/mutualfriends/output, listing two entries with permissions -rw-r--r-- and owner hadoop.

File information - part-r-00000

Download Head the file (first 32K) Tail the file (last 32K)

Block information -- Block 0

Block ID: 1073741911
 Block Pool ID: BP-1266388418-127.0.0.1-1560101340487
 Generation Stamp: 1087
 Size: 33
 Availability:
 • localhost

File contents

```
1.2 5
1.4 5
1.5 2, 4
2.5 1
4.5 1
```

← → ↺ 🏠

localhost:50070/explorer.html#/projects/mutualfriends/input

📄 ⋮ 🗑️ ☆

🖨️ 📄 📄

Hadoop

Overview

Datanodes

Datanode Volume Failures

Searches

Startup Progress

Utilities

Browse Directory

/projects/mutualfriends/input

Show 25 entries

☐

Permission

☐

Ownership

Showing 1 to 1 of 1 entries

Hadoop, 2019.

Go!

📄 📄 📄

Search:

Block Size

Name

28 MB

sample.txt

Previous 1 Next

File information - sample.txt

Download

Head the file (first 32K)

Tail the file (last 32K)

Block information -- Block 0

Block ID: 1073741904

Block Pool ID: BP-1266388418-127.0.0.1-1560101340487

Generation Stamp: 1080

Size: 34

Availability:

• localhost

File contents

1 2.4.5

2 1.3.5

3 2

4 1.5

5 1.2.4