

# CS570 BigData Processing And Analysis

## Homework3

```
import java.io.*;
import java.util.*;
import java.lang.Object;
import java.net.URI;
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.conf.*;
import org.apache.hadoop.io.*;
import org.apache.hadoop.mapreduce.*;
import org.apache.hadoop.mapreduce.Mapper.Context;
import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;
import org.apache.hadoop.mapreduce.lib.input.TextInputFormat;
import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;
import org.apache.hadoop.mapreduce.lib.output.TextOutputFormat;
import org.apache.hadoop.fs.*;

public class MapReducepi {

    public static class Map extends Mapper<LongWritable, Text, Text, IntWritable> {

        private final static IntWritable one = new IntWritable (1);
        private Text word = new Text();

        public void map(LongWritablekey,Textvalue,Contextcontext)throws IOException{

            int totals = 0;

            for (IntWritable val : values) {

                totals += val.get();

            }

        }

        context.write(key, new IntWritable(totals));

    }

}

public static void main(String[] args) throws IOException InterruptedException {
```

```

Configuration conf = new Configuration();
Job job = new Job(conf, "MapReducepi");
job.setJarByClass(MapReducepi.class);
job.setOutputKeyClass(Text.class);
job.setOutputValueClass(IntWritable.class);
job.setMapperClass(Map.class);
job.setCombinerClass(Reduce.class);
job.setReducerClass(Reduce.class);
job.setInputFormatClass(TextInputFormat.class);
job.setOutputFormatClass(TextOutputFormat.class);
job.setNumReduceTasks(1);
FileInputFormat.addInputPath(job, new Path(args[0]));
FileOutputFormat.setOutputPath(job, new Path(args[1]));
circle1 = circle1.replace("outsidevalue", "").trim();
System.out.println("insidevalue:"+circle2+", outsidevalue:"+circle1)
if (circle2 != null && circle1 != null) {
    double valueinside = Double.valueOf(circle2);
    double valueoutside = Double.valueOf(circle1);
    double pi_value = 4*(valueinside/(valueinside+valueoutside));
    System.out.println("Pi value:"+pi_value);
}
fs.close();
}
}

```

Now setup for Hadoop Installation;

Sudo addgroup hadoop

Sudo adduser -ingroup Hadoop hduser sudo su

hduser

For this I used Hadoop version hadoop-2.10.2.tar.gz Sudo tar xzf hadoop-2.10.2.tar.gz -C /usr/local Cd /usr/local

Sudo mv hadoop-2.10.2 hadoop  
Sudo chown -R hduser:Hadoop Hadoop

```
hduser@cs570bigdata: ~  
hduser@cs570bigdata:~$ sudo nano ~/.bashrc  
[sudo] password for hduser: █
```

```
hduser@cs570bigdata: ~  
GNU nano 4.8 /home/hduser/.bashrc  
~/.bashrc: executed by bash(1) for non-login shells.  
# see /usr/share/doc/bash/examples/startup-files (in the package bash) for examples  
  
# If not running interactively, don't do anything  
case $- in  
  *i*) ;;  
  *) return;;  
esac  
  
# don't put duplicate lines or lines starting with space in the history  
# See bash(1) for more options  
HISTCONTROL=ignoreboth  
  
# append to the history file, don't overwrite it  
shopt -s histappend  
  
# for setting history length see HISTSIZE and HISTFILESIZE in bash(1)
```

```
hduser@cs570bigdata: ~  
GNU nano 4.8 /home/hduser/.bashrc  
# sources /etc/bash.bashrc).  
if ! shopt -oq posix; then  
  if [ -f /usr/share/bash-completion/bash_completion ]; then  
    . /usr/share/bash-completion/bash_completion  
  elif [ -f /etc/bash_completion ]; then  
    . /etc/bash_completion  
  fi  
fi  
export JAVA_HOME=/usr/lib/jvm/jdk/  
export HADOOP_INSTALL=/usr/local/hadoop  
export PATH=$PATH:$HADOOP_INSTALL/bin  
export PATH=$PATH:$HADOOP_INSTALL/sbin  
export HADOOP_MAPRED_HOME=$HADOOP_INSTALL  
export HADOOP_COMMON_HOME=$HADOOP_INSTALL  
export HADOOP_HDFS_HOME=$HADOOP_INSTALL  
export YARN_HOME=$HADOOP_INSTALL  
export HADOOP_COMMON_LIB_NATIVE_DIR=$HADOOP_HOME/lib/native  
export HADOOP_OPTS="-Djava.library.path=$HADOOP_HOME/lib"
```

To save use source

~/ .bashrc systemctl reboot -i export JAVA\_HOME=/usr/lib/jvm/jdk

sudo nano /usr/local/hadoop/etc/hadoop/yarn-site.xml

```
<?xml version="1.0"?>  
<!--  
Licensed under the Apache License, Version 2.0 (the "License");  
you may not use this file except in compliance with the License.  
You may obtain a copy of the License at  
  
http://www.apache.org/licenses/LICENSE-2.0  
  
Unless required by applicable law or agreed to in writing, software  
distributed under the License is distributed on an "AS IS" BASIS,  
WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.  
See the License for the specific language governing permissions and  
limitations under the License. See accompanying LICENSE file.  
-->  
<configuration>  
  
<!-- Site specific YARN configuration properties -->  
<property>  
  <name>yarn.nodemanager.aux-services</name>  
  <value>mapreduce_shuffle</value>  
</property>  
<property>  
  <name>yarn.nodemanager.aux-services.mapreduce.shuffle.class</name>  
  <value>org.apache.hadoop.mapred.ShuffleHandler</value>  
</property>  
</configuration>
```

```

<?xml version="1.0" encoding="UTF-8"?>
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
<!--
  Licensed under the Apache License, Version 2.0 (the "License");
  you may not use this file except in compliance with the License.
  You may obtain a copy of the License at

    http://www.apache.org/licenses/LICENSE-2.0

  Unless required by applicable law or agreed to in writing, software
  distributed under the License is distributed on an "AS IS" BASIS,
  WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
  See the License for the specific language governing permissions and
  limitations under the License. See accompanying LICENSE file.
-->

<!-- Put site-specific property overrides in this file. -->

<configuration>
<property>
<name>dfs.replication</name>
<value>1 </value>
</property>
<property>
<name>dfs.namenode.name.dir</name>
<value>file:/home/hduser/mydata/hdfs/namenode</value>
</property>
<property>
<name>dfs.datanode.data.dir</name>
<value>file:/home/hduser/mydata/hdfs/datanode</value>
</property>
</configuration>

```

```

hduser@cs570bigdata:/usr/local/hadoop$ cd hadoop
-bash: cd: hadoop: No such file or directory
hduser@cs570bigdata:/usr/local/hadoop$ ls
bin  etc  include  lib  libexec  LICENSE.txt  logs  NOTICE.txt  README.txt  sbin  share
hduser@cs570bigdata:/usr/local/hadoop$ cd hadoop
-bash: cd: hadoop: No such file or directory
hduser@cs570bigdata:/usr/local/hadoop$ cd sbin
hduser@cs570bigdata:/usr/local/hadoop/sbin$ hdfs dfs -mkdir /user

```

```

hduser@cs570bigdata:/usr/local/hadoop/sbin$ start-dfs.sh
WARNING: An illegal reflective access operation has occurred
WARNING: Illegal reflective access by org.apache.hadoop.security.authentication.util.KerberosUtil (file:/usr/local/hadoop/share/hadoop/common/lib/hadoop-
ity.krb5.Config.getInstance()
WARNING: Please consider reporting this to the maintainers of org.apache.hadoop.security.authentication.util.KerberosUtil
WARNING: Use --illegal-access=warn to enable warnings of further illegal reflective access operations
WARNING: All illegal access operations will be denied in a future release
Starting namenodes on [cs570bigdata]
cs570bigdata: starting namenode, logging to /usr/local/hadoop/logs/hadoop-hduser-namenode-cs570bigdata.out
localhost: starting datanode, logging to /usr/local/hadoop/logs/hadoop-hduser-datanode-cs570bigdata.out
Starting secondary namenodes [0.0.0.0]
0.0.0.0: starting secondarynamenode, logging to /usr/local/hadoop/logs/hadoop-hduser-secondarynamenode-cs570bigdata.out
WARNING: An illegal reflective access operation has occurred
WARNING: Illegal reflective access by org.apache.hadoop.security.authentication.util.KerberosUtil (file:/usr/local/hadoop/share/hadoop/common/lib/hadoop-
ity.krb5.Config.getInstance()
WARNING: Please consider reporting this to the maintainers of org.apache.hadoop.security.authentication.util.KerberosUtil
WARNING: Use --illegal-access=warn to enable warnings of further illegal reflective access operations
WARNING: All illegal access operations will be denied in a future release
hduser@cs570bigdata:/usr/local/hadoop/sbin$

```

```

hduser@cs570bigdata:/usr/local/hadoop/sbin$ start-yarn.sh
starting yarn daemons
starting resourcemanager, logging to /usr/local/hadoop/logs/yarn-hduser-resourcemanager-cs570bigdata.out
localhost: starting nodemanager, logging to /usr/local/hadoop/logs/yarn-hduser-nodemanager-cs570bigdata.out
hduser@cs570bigdata:/usr/local/hadoop/sbin$ jps
3347 NodeManager
2998 SecondaryNameNode
3208 ResourceManager
3660 Jps
2638 NameNode
2702 DataNode
hduser@cs570bigdata:/usr/local/hadoop/sbin$

```

## Hadoop running Pi Jar file command:

```

hduser@cs570bigdata:/usr/local/hadoop$ bin/hadoop jar
/home/hduser/GPi/pivalueproj.jar PiValuepro /user/hduser/inputdata outputwc

```

**Previous slide output continued :** hduser@cs570bigdata:/usr/local/hadoop\$  
bin/hadoop jar

```

/home/hduser/GPi/pivalueproj.jar PiValuepro /user/hduser/inputdata outputwc

```

```

hduser@cs570bigdata:/usr/local/hadoop$ bin/hadoop jar /home/hduser/GPi/pivalueproj.jar PiValuepro /user/hduser/inputdata outputwc
WARNING: An illegal reflective access operation has occurred
WARNING: Illegal reflective access by org.apache.hadoop.security.authentication.util.KerberosUtil (file:/usr/local/hadoop/share/hadoop/common/lib/hadoop-auth-2.10.2.jar) to method sun.security.krb5.Config.getInstance()
WARNING: Please consider reporting this to the maintainers of org.apache.hadoop.security.authentication.util.KerberosUtil
WARNING: Use --illegal-access=warn to enable warnings of further illegal reflective access operations
WARNING: All illegal access operations will be denied in a future release
22/10/10 00:52:04 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
22/10/10 00:52:13 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and
execute your application with ToolRunner to remedy this.
22/10/10 00:52:17 INFO input.FileInputFormat: Total input files to process : 1
22/10/10 00:52:18 INFO mapreduce.JobSubmitter: number of splits:1
22/10/10 00:52:27 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1665330928624_0001
22/10/10 00:52:33 INFO conf.Configuration: resource-types.xml not found
22/10/10 00:52:33 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
22/10/10 00:52:33 INFO resource.ResourceUtils: Adding resource type - name = memory-mb, units = Mi, type = COUNTABLE
22/10/10 00:52:33 INFO resource.ResourceUtils: Adding resource type - name = vcores, units = , type = COUNTABLE
22/10/10 00:52:40 INFO impl.YarnClientImpl: Submitted application application_1665330928624_0001
22/10/10 00:52:42 INFO mapreduce.Job: The url to track the job: http://cs570bigdata:8088/proxy/application_1665330928624_0001/
22/10/10 00:52:42 INFO mapreduce.Job: Running job: job_1665330928624_0001
22/10/10 00:54:14 INFO mapreduce.Job: Job job_1665330928624_0001 running in uber mode : false
22/10/10 00:54:14 INFO mapreduce.Job: map 0% reduce 0%
22/10/10 00:55:06 INFO mapreduce.Job: map 100% reduce 0%
22/10/10 00:55:34 INFO mapreduce.Job: map 100% reduce 100%
22/10/10 00:55:36 INFO mapreduce.Job: Job job_1665330928624_0001 completed successfully
22/10/10 00:55:37 INFO mapreduce.Job: Counters: 49
  File System Counters
    FILE: Number of bytes read=54
    FILE: Number of bytes written=421529
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=150
    HDFS: Number of bytes written=28
    HDFS: Number of read operations=6
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=2
  Job Counters
    Launched map tasks=1
    Launched reduce tasks=1
    Data-local map tasks=1
    Total time spent by all maps in occupied slots (ms)=44660
    Total time spent by all reduces in occupied slots (ms)=24845
    Total time spent by all map tasks (ms)=44660
    Total time spent by all reduce tasks (ms)=24845
    Total vcore-milliseconds taken by all map tasks=44660
    Total vcore-milliseconds taken by all reduce tasks=24845
    Total megabyte-milliseconds taken by all map tasks=45731840
    Total megabyte-milliseconds taken by all reduce tasks=25441280
  Map-Reduce Framework

```



hduser@cs570bigdata: /usr/local/hadoop

```
HDFS: Number of read operations=6
HDFS: Number of large read operations=0
HDFS: Number of write operations=2
Job Counters
  Launched map tasks=1
  Launched reduce tasks=1
  Data-local map tasks=1
  Total time spent by all maps in occupied slots (ms)=44660
  Total time spent by all reduces in occupied slots (ms)=24845
  Total time spent by all map tasks (ms)=44660
  Total time spent by all reduce tasks (ms)=24845
  Total vcore-milliseconds taken by all map tasks=44660
  Total vcore-milliseconds taken by all reduce tasks=24845
  Total megabyte-milliseconds taken by all map tasks=45731840
  Total megabyte-milliseconds taken by all reduce tasks=25441280
Map-Reduce Framework
  Map input records=8
  Map output records=8
  Map output bytes=62
  Map output materialized bytes=54
  Input split bytes=120
  Combine input records=8
  Combine output records=5
  Reduce input groups=5
  Reduce shuffle bytes=54
  Reduce input records=5
  Reduce output records=5
  Spilled Records=10
  Shuffled Maps =1
  Failed Shuffles=0
  Merged Map outputs=1
  GC time elapsed (ms)=767
  CPU time spent (ms)=12160
  Physical memory (bytes) snapshot=432816128
  Virtual memory (bytes) snapshot=4139466752
  Total committed heap usage (bytes)=211812352
Shuffle Errors
  BAD_ID=0
  CONNECTION=0
  IO_ERROR=0
  WRONG_LENGTH=0
  WRONG_MAP=0
  WRONG_REDUCE=0
File Input Format Counters
  Bytes Read=30
File Output Format Counters
  Bytes Written=28
1,1      2
2        1
Inside:1,1    2, Outside:2    1
```