Project : MapReduce Pi

Presented by MANICKAM RAVISEKAR, Master of Science in Computer Science, 19599,

Spring 2022

Guidance from Dr., Professor Henry Chang
TA Liang

SAN FRANCISCO BAY UNIVERSITY 47671 WestingHouse Dr., Fremont, CA 94539

ACKNOWLEDGEMENT

One of our master's degree Project for Hadoop MapReduce for PI computation, Is Interesting, which made me to learn new things, it is useful in designing and applying on Google Cloud

Platform (Hadoop MapReduce) on Google Cloud Platform and Oracle Virtual Machine /

For deploying this project, I would like to thank Dr. Henry Chang an TA Liang for providing all the required input.

Also, for all I would like to always pray to Almighty for giving us wisdom and power to understand things.

Index

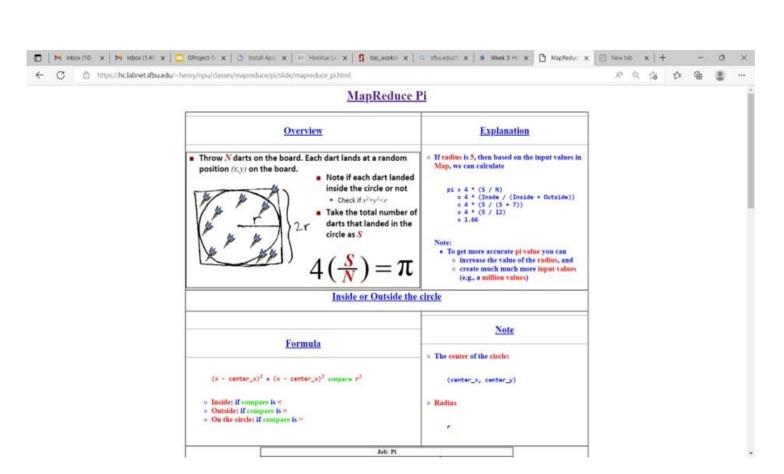
- 1. Abstract
- 2 About MapReduce Pi
- 3.Installation of Oracle VM, Java and Apache Hadoop
- 4. Steps to be followed to test MapReduce Pi
- 5. Results of Output
- 6. Conclusion
- 7. Reference

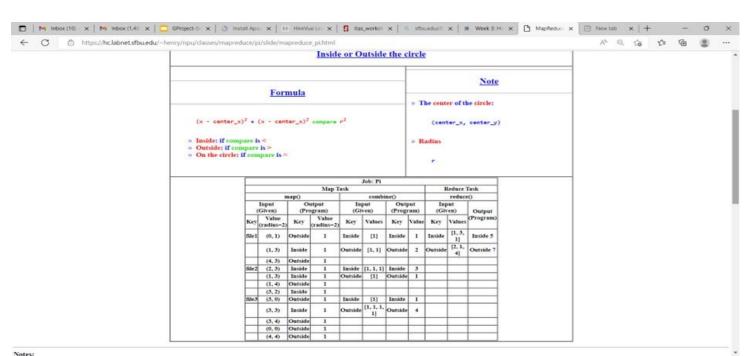
ABSTRACT

The main lesson about this assignment is to study details, about deploying applications on both Google Cloud Platform and Oracle Virtual Machine to learn

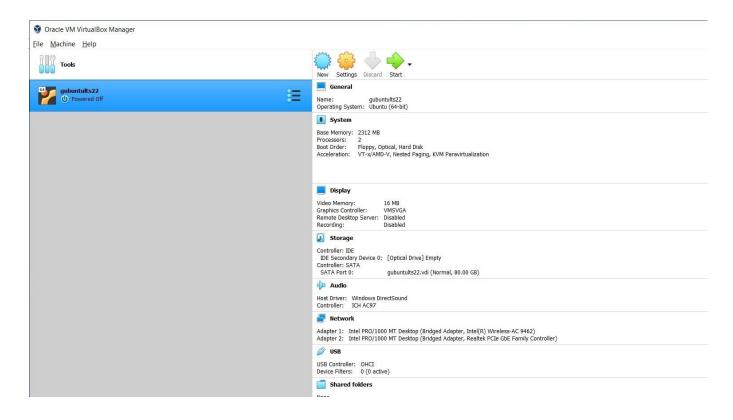
about Hadoop MapReduce to apply for Pi calculations with

radius and random points for a circle to compute points to find whether points are inside (darts) or outside circle.





MapReduce on Ubuntu on Oracle Virtual Machine local:





Shared Folders

User Interface

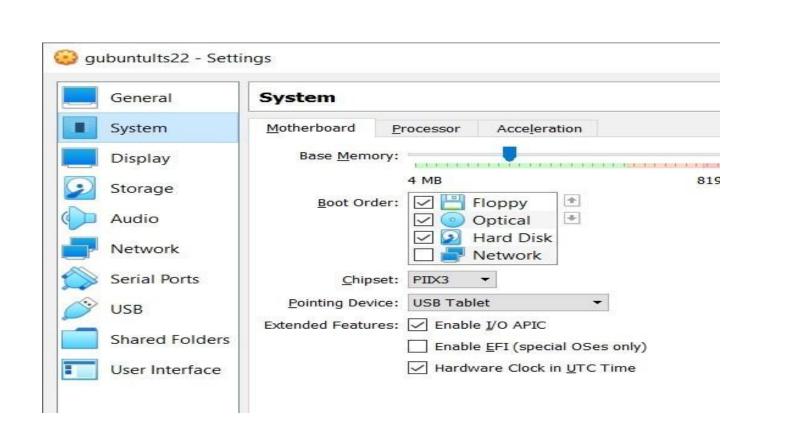
General

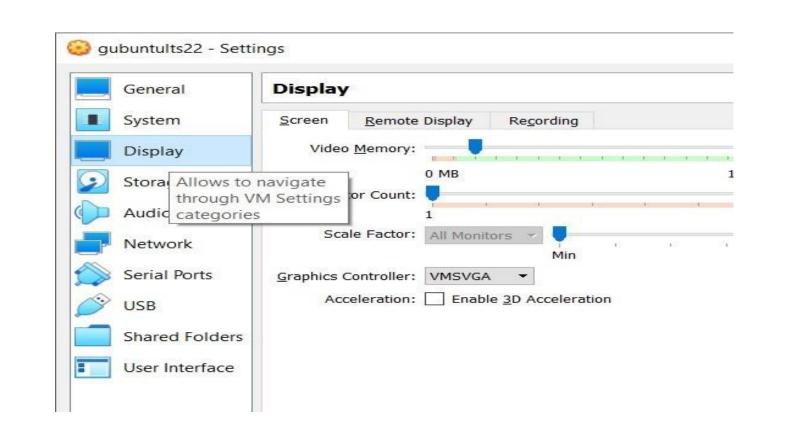
Advanced

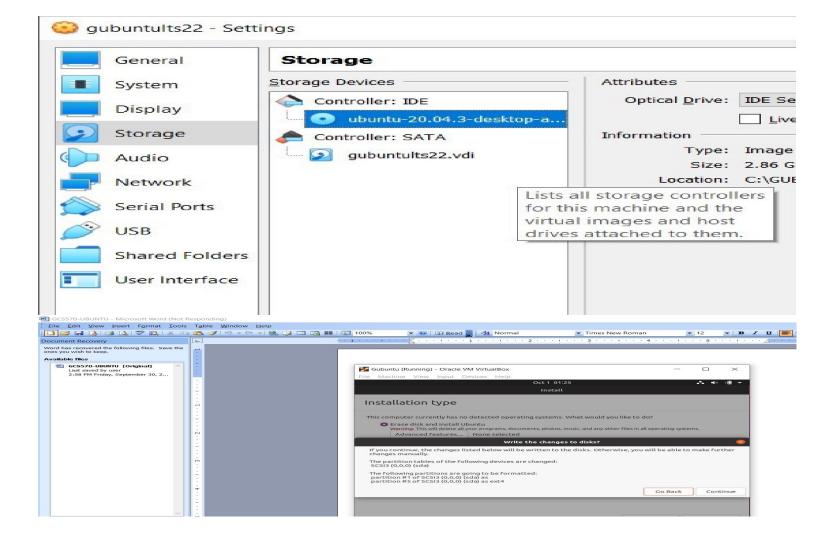
Description

Disk Encryption

Version: Ubuntu (64-bit)







Select start-up disk

Please select a virtual optical disk file or a physica drive containing a disk to start your new virtual m from.

The disk should be suitable for starting a compute should contain the operating system you wish to it the virtual machine if you want to do that now. The ejected from the virtual drive automatically new switch the virtual machine off, but you can also do yourself if needed using the Devices menu.

After completion of installation Oracle VM, install Java, Hadoop and the script files Download hadoop Download Java Extract hadoop Java installation steps

sudo apt-get install openjdk-11-jre sudo apt-get install openjdk-11-jdk java - version

hduser@cs570bigdata:~
hduser@cs570bigdata:~\$ java -version
openjdk version "11.0.16" 2022-07-19
OpenJDK Runtime Environment (build 11.0.16+8-post-Ubuntu-Oubuntu12
OpenJDK 64-Bit Server VM (build 11.0.16+8-post-Ubuntu-Oubuntu120.0
sharing)
hduser@cs570bigdata:~\$

Setup hadoop user 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

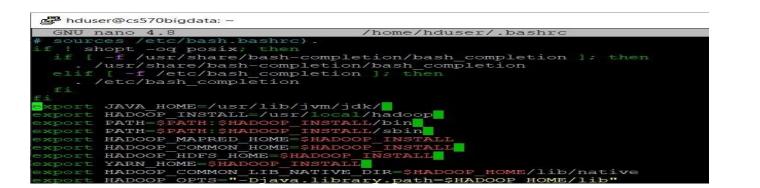
xzf hadoop-2.10.2.tar.gz –C /usr/local Cd / usr/local

Sudo mv hadoop-2.10.2 hadoop

Sudo chown –R hduser:hadoop hadoop

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

Move to the end and add following lines for hadoop



To save use source ~/.bashrc systemetl reboot –i

Now do the following setting for hduser

export JAVA_HOME=/usr/lib/jvm/jdk

Ssh generation and creation of authorized keys from public ssh keys

```
ravisekar@ravisekar-VirtualBox: ~/ssh
rav1sekar@ravisekar-VirtualBox: ~$ ssh-keygen -t rsa -P ""

rav1sekar@ravisekar-VirtualBox: ~$ ssh-keygen -t rsa -P ""

Inter file in which to save the key palr.

Inter file in which to save the key (home/ravisekar/.ssh/id_rsa):

Created directory '/home/ravisekar/.ssh'.

Your identification has been saved in /home/ravisekar/.ssh/id_rsa

Inter file in which has been saved in /home/ravisekar/.ssh/id_rsa

Inter key fingerprint is:

Inter key fingerprint is:

Inter key randomart image is:

Inter key's randomart image is:

Inter key's randomart image is:

Inter key is randomart image is randomart image is:

Inter key is randomart image is:

Inter key is randomart image is:

Inter key is randomart image is:

Int
```

Now we need following files to be set for hadoop:

```
ORV Industry (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (19
```

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

 $3. \ sudo \ nano \ /usr/local/hadoop/etc/hadoop/mapred-site.xml$

```
GRIP And 4.0

GRIP AND 4.0

GRIP AND 4.0

GRIP AND 4.0

A
```

4 For doing hdfs.xml

First complete this task's mkdir -p

```
Thusersety topiquate year the type-tiesty and the file an
```

Create following test data file for hdfs

Phuwewce.S/Chigdata.-/Desktop/inputdata
rduros/esis/Objetata./Objetata/p/inputdatas cat test.txt
heilo world world hello hello
hduse.ge.s570bigdata:-/Desktop/inputdatas

Create directory as mentioned



Now copy to file created for input /usr/local/hadoop/bin/hdfs dfs -put '/home/hduser/Desktop/input data' /user Next final step

for hadoop

hdfs namenode -format

now everything is set we can start hadoop and run the jar for the word count and get the output as shown in below screen

now start hadoop as shown below by running start-dfs.sh start- yarn.sh and jps to check the status as shown in below screens

Advances: All All Anti-All Ant

where cs70bigdata:/ws/local/hadoop/sbin | https://docs.cs/hadoop/sbin | start-yarn.sh | https://docs.cs/hadoop/sbin | start-yarn.sh | https://docs.cs/hadoop/sbin | start-yarn.sh | start-yarn

```
Sample program to generate input file for MapReduce Pi:
import java.io.*; import java.util.*;
import java.lang.Object; import
java.net.URI;
         org.apache.hadoop.fs.Path;
                                           import
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 PiValue {
  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(LongWritable key, Text value, Context context) throws

IOEvention InterruptedEvention

```
int sum = 0;
              for (IntWritable val : values) { sum +=
                      val.get();
              context.write(key, new IntWritable(sum));
public static void main(String[] args) throws Exception { Configuration
 conf = new Configuration();
 Job job = new Job(conf, "PiValue");
 job.setJarByClass(PiValue.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]));
```

ioh waitForCompletion(true):

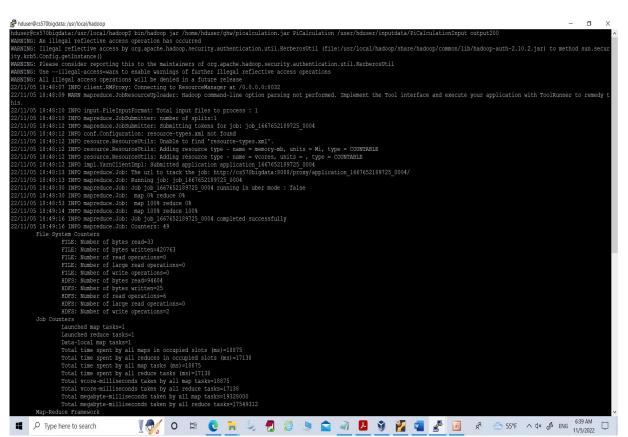
```
line2 = line2.replace("outside","").trim();
System.out.println("Inside:"+line1+", Outside:"+line2); if (line1 !=
null && line2 != null) {
  double invalue = Double.valueOf(line1); double
  outvalue = Double.valueOf(line2); double pi =4*(
```

invalue /(invalue+outvalue));
System.out.println("PI:"+pi);

fs.close();

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

Random number: 10000 (Ten Thousand)



Random number: 10000 (Ten Thousand)

```
A hduser@cs570bigdata; /usr/local/hadoop
                                                                                                                                                                               - 0 X
               HDFS: Number of write operations=2
                Launched map tasks=1
                Launched reduce tasks=1
               Data-local map tasks=1
               Total time spent by all maps in occupied slots (ms)=18875
               Total time spent by all reduces in occupied slots (ms)=17138
               Total time spent by all map tasks (ms)=18875
               Total time spent by all reduce tasks (ms)=17138
               Total vcore-milliseconds taken by all map tasks=18875
               Total vcore-milliseconds taken by all reduce tasks=17138
               Total megabyte-milliseconds taken by all map tasks=19328000
                Total megabyte-milliseconds taken by all reduce tasks=17549312
       Map-Reduce Framework
               Map input records=1
               Map output records=10000
               Map output bytes=112145
               Map output materialized bytes=33
                Input split bytes=130
               Combine input records=10000
               Combine output records=2
               Reduce input groups=2
               Reduce shuffle bytes=33
               Reduce input records=2
               Reduce output records=2
               Spilled Records=4
               Shuffled Maps =1
               Failed Shuffles=0
               Merged Map outputs=1
               GC time elapsed (ms)=2308
               CPU time spent (ms)=4730
               Physical memory (bytes) snapshot=448245760
               Virtual memory (bytes) snapshot=4147322880
               Total committed heap usage (bytes)=258998272
       Shuffle Errors
               BAD ID=0
                CONNECTION=0
                IO ERROR=0
               WRONG MAP=0
               WRONG REDUCE=0
       File Input Format Counters
               Bytes Read=94474
       File Output Format Counters
               Bytes Written=25
inside 7855
outside 2145
Inside:7855, Outside:2145
  duser@cs570bigdata:/usr/local/hadoop$
```

Random number: 50000 (Fifty Thousand)

```
Aduser@cs570bigdata: /usr/local/hadoop
                                                                                                                                                                                     П
hduser@cs570bigdata:/usr/local/hadoop$ bin/hadoop jar /home/hduser/ghw/picalculation.jar PiCalculation /user/hduser/inputdata/PiCalculationInput 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.secur
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
22/11/05 19:22:42 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
22/11/05 19:22:45 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy
22/11/05 19:22:46 INFO input.FileInputFormat: Total input files to process: 1
22/11/05 19:22:47 INFO mapreduce.JobSubmitter: number of splits:1
22/11/05 19:22:49 INFO mapreduce.JobSubmitter: Submitting tokens for job: job 1667652189725 0005
22/11/05 19:22:50 INFO conf.Configuration: resource-types.xml not found
22/11/05 19:22:50 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
22/11/05 19:22:50 INFO resource.ResourceUtils: Adding resource type - name = memory-mb, units = Mi, type = COUNTABLE
22/11/05 19:22:50 INFO resource.ResourceUtils: Adding resource type - name = vcores, units = , type = COUNTABLE
22/11/05 19:22:53 INFO impl.YarnClientImpl: Submitted application application 1667652189725 0005
22/11/05 19:22:53 INFO mapreduce.Job: The url to track the job: http://cs570bigdata:8088/proxy/application 1667652189725 0005/
22/11/05 19:22:53 INFO mapreduce.Job: Running job: job 1667652189725 0005
22/11/05 19:23:38 INFO mapreduce.Job: Job job 1667652189725 0005 running in uber mode : false
22/11/05 19:23:38 INFO mapreduce.Job: map 0% reduce 0%
22/11/05 19:24:08 INFO mapreduce.Job: map 100% reduce 0%
22/11/05 19:24:45 INFO mapreduce.Job: map 100% reduce 100%
22/11/05 19:24:52 INFO mapreduce.Job: Job job 1667652189725 0005 completed successfully
22/11/05 19:24:53 INFO mapreduce.Job: Counters: 49
       File System Counters
               FILE: Number of bytes read=33
               FILE: Number of bytes written=420761
               FILE: Number of read operations=0
               FILE: Number of large read operations=0
               FILE: Number of write operations=0
               HDFS: Number of bytes read=472721
               HDFS: Number of bytes written=27
               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)=23590
               Total time spent by all reduces in occupied slots (ms)=29968
               Total time spent by all map tasks (ms)=23590
               Total time spent by all reduce tasks (ms)=29968
               Total vcore-milliseconds taken by all map tasks=23590
               Total vcore-milliseconds taken by all reduce tasks=29968
               Total megabyte-milliseconds taken by all map tasks=24156160
                Total megabyte-milliseconds taken by all reduce tasks=30687232
       Map-Reduce Framework
```

Random number: 50000 (Fifty Thousand)

```
hduser@cs570bigdata: /usr/local/hadoop
                                                                                                                                                                                   0 X
               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)=23590
               Total time spent by all reduces in occupied slots (ms)=29968
               Total time spent by all map tasks (ms)=23590
               Total time spent by all reduce tasks (ms)=29968
               Total vcore-milliseconds taken by all map tasks=23590
               Total vcore-milliseconds taken by all reduce tasks=29968
               Total megabyte-milliseconds taken by all map tasks=24156160
               Total megabyte-milliseconds taken by all reduce tasks=30687232
       Map-Reduce Framework
               Map input records=1
               Map output records=50000
               Map output bytes=560756
               Map output materialized bytes=33
               Input split bytes=130
               Combine input records=50000
               Combine output records=2
               Reduce input groups=2
               Reduce shuffle bytes=33
               Reduce input records=2
               Reduce output records=2
               Spilled Records=4
               Shuffled Maps =1
               Failed Shuffles=0
               Merged Map outputs=1
               GC time elapsed (ms)=948
               CPU time spent (ms)=4970
               Physical memory (bytes) snapshot=416432128
               Virtual memory (bytes) snapshot=4134461440
               Total committed heap usage (bytes)=205520896
       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=472591
       File Output Format Counters
               Bytes Written=27
outside 10756
Inside: 39244, Outside: 10756
nduser@cs570bigdata:/usr/local/hadoop$
```

Random number: 1000000 (One Million)

```
hduser@cs570bigdata: /usr/local/hadoop
                                                                                                                                                                                      П
hduser@cs570bigdata:/usr/local/hadoop$ bin/hadoop jar /home/hduser/ghw/picalculation.jar PiCalculation /user/hduser/inputdata/PiCalculationInput outputwo
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.secur
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
22/11/06 05:01:51 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
22/11/06 05:01:53 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy t
22/11/06 05:01:55 INFO input.FileInputFormat: Total input files to process: 1
22/11/06 05:01:55 INFO mapreduce.JobSubmitter: number of splits:1
22/11/06 05:01:57 INFO mapreduce.JobSubmitter: Submitting tokens for job: job 1667690462450 0001
22/11/06 05:01:58 INFO conf.Configuration: resource-types.xml not found
22/11/06 05:01:58 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
22/11/06 05:01:58 INFO resource.ResourceUtils: Adding resource type - name = memory-mb, units = Mi, type = COUNTABLE
22/11/06 05:01:58 INFO resource.ResourceUtils: Adding resource type - name = vcores, units = , type = COUNTABLE
22/11/06 05:01:59 INFO impl.YarnClientImpl: Submitted application application 1667690462450 0001
22/11/06 05:02:00 INFO mapreduce.Job: The url to track the job: http://cs570bigdata:8088/proxy/application 1667690462450 0001/
22/11/06 05:02:00 INFO mapreduce. Job: Running job: job 1667690462450 0001
22/11/06 05:02:25 INFO mapreduce.Job: Job job 1667690462450 0001 running in uber mode: false
22/11/06 05:02:25 INFO mapreduce.Job: map 0% reduce 0%
22/11/06 05:02:46 INFO mapreduce.Job: map 100% reduce 0%
22/11/06 05:03:03 INFO mapreduce.Job: map 100% reduce 100%
22/11/06 05:03:07 INFO mapreduce.Job: Job job 1667690462450 0001 completed successfully
22/11/06 05:03:08 INFO mapreduce.Job: Counters: 49
       File System Counters
                FILE: Number of bytes read=33
               FILE: Number of bytes written=420761
               FILE: Number of read operations=0
               FILE: Number of large read operations=0
               FILE: Number of write operations=0
                HDFS: Number of bytes read=945592
               HDFS: Number of bytes written=27
               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)=16802
               Total time spent by all reduces in occupied slots (ms)=16083
               Total time spent by all map tasks (ms)=16802
               Total time spent by all reduce tasks (ms)=16083
               Total vcore-milliseconds taken by all map tasks=16802
               Total vcore-milliseconds taken by all reduce tasks=16083
               Total megabyte-milliseconds taken by all map tasks=17205248
                Total megabyte-milliseconds taken by all reduce tasks=16468992
       Map-Reduce Framework
```

Random number: 1000000 (One Million)

```
O
hduser@cs570bigdata: /usr/local/hadoop
                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)=16802
                Total time spent by all reduces in occupied slots (ms)=16083
                Total time spent by all map tasks (ms)=16802
                Total time spent by all reduce tasks (ms)=16083
                Total vcore-milliseconds taken by all map tasks=16802
                Total vcore-milliseconds taken by all reduce tasks=16083
                Total megabyte-milliseconds taken by all map tasks=17205248
               Total megabyte-milliseconds taken by all reduce tasks=16468992
       Map-Reduce Framework
                Map input records=1
                Map output records=100000
                Map output bytes=1121405
                Map output materialized bytes=33
                Input split bytes=130
                Combine input records=100000
               Combine output records=2
               Reduce input groups=2
               Reduce shuffle bytes=33
                Reduce output records=2
                Spilled Records=4
                Shuffled Maps =1
               Failed Shuffles=0
               Merged Map outputs=1
               GC time elapsed (ms)=498
               CPU time spent (ms)=4690
               Physical memory (bytes) snapshot=471113728
                Virtual memory (bytes) snapshot=4132884480
                Total committed heap usage (bytes) = 252706816
       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=945462
       File Output Format Counters
                Bytes Written=27
inside 78595
outside 21405
Inside: 78595, Outside: 21405
 duser@cs570bigdata:/usr/local/hadoop$
```

Conclusion

Application usage of MapReduce - PI, this lessons learned

Can be applied in many application design for example in study of

Astronomy to find the details of the location of planets, meteors.

Reference:
SFBU course materials