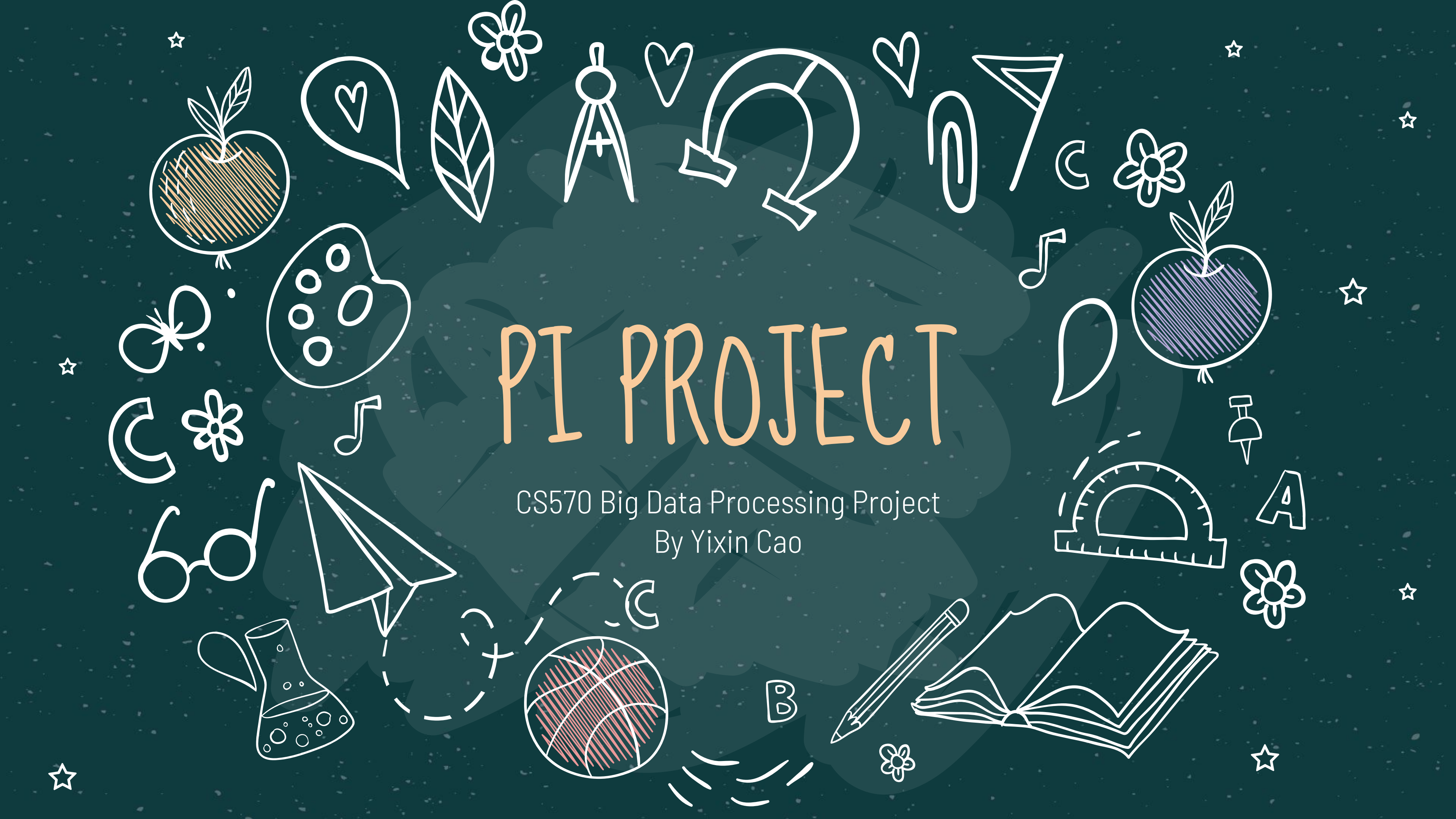


PI PROJECT

CS570 Big Data Processing Project

By Yixin Cao

TABLE OF CONTENT

1. Introduction
2. Design
3. Implementation
4. Test
5. Enhancement
6. Conclusion
7. Reference

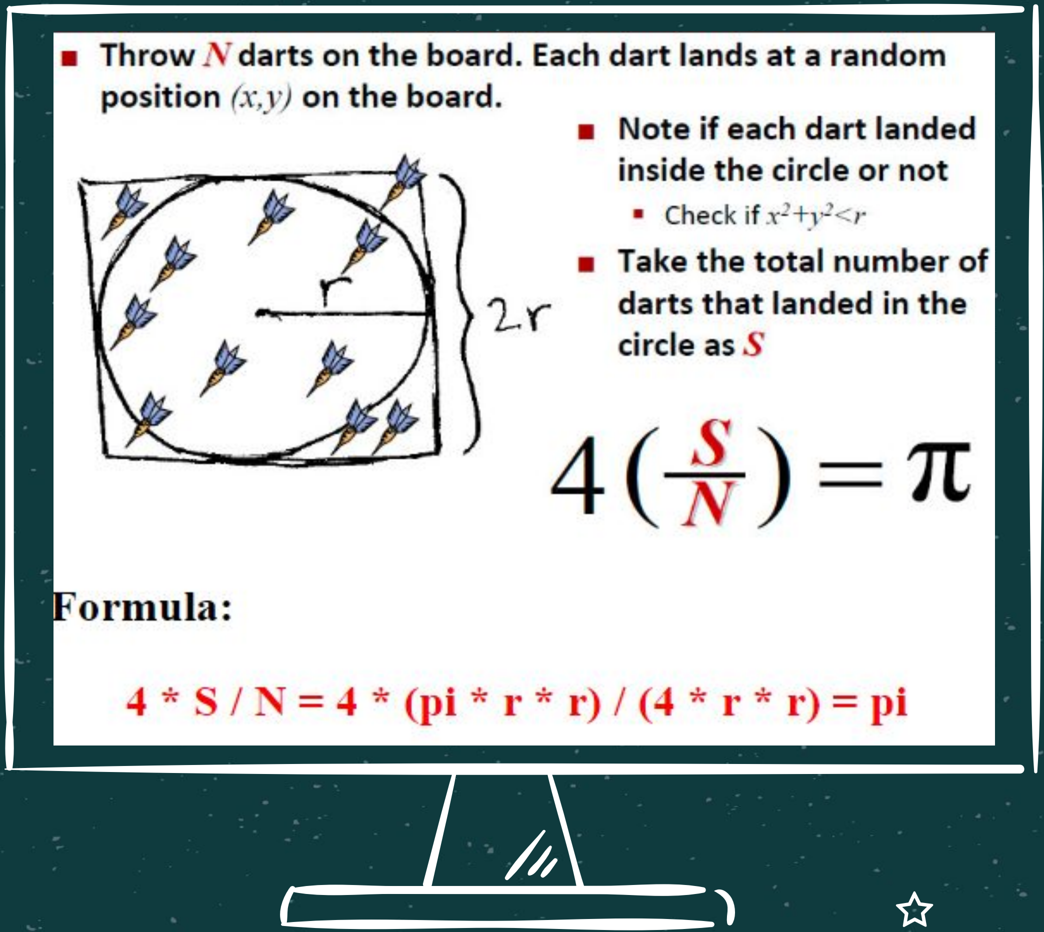
INTRODUCTION

This Pi Project is to use Google Cloud Platform to implement Hadoop with MapReduce to calculate pi value.

STUDY
HARD!

THEORY OF
Pi Calculation

As the illustrated on the right, the value of pi can be calculated by counting the number of random darts that falls in the circle and outside the circle.



- Throw N darts on the board. Each dart lands at a random position (x,y) on the board.
- Note if each dart landed inside the circle or not
  - Check if $x^2+y^2<r$
- Take the total number of darts that landed in the circle as S

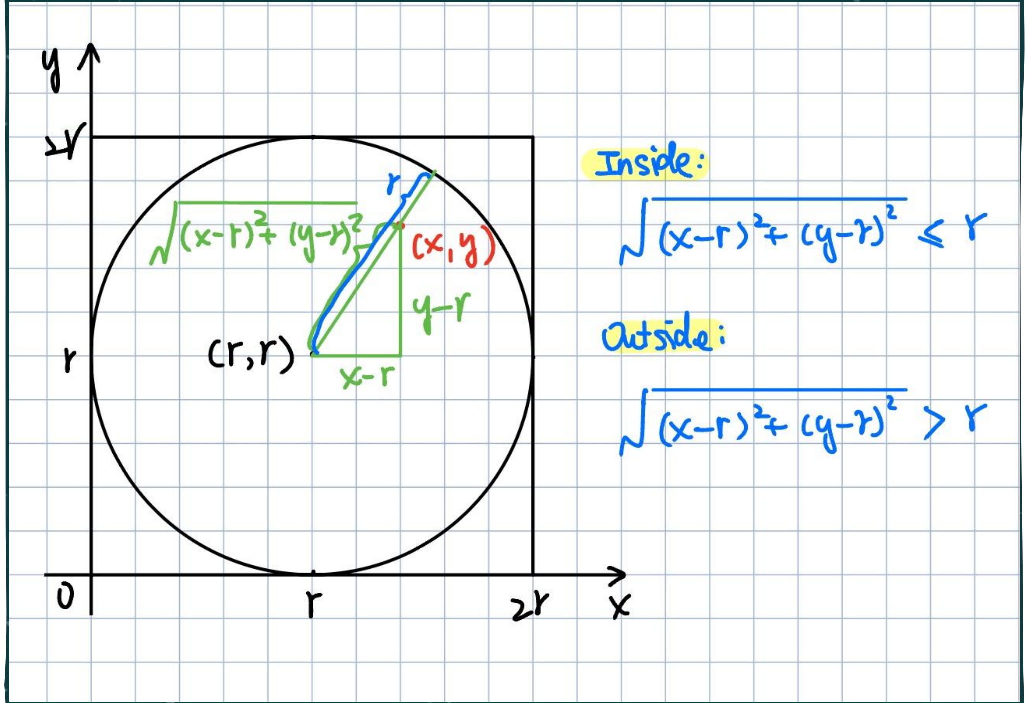
$$4\left(\frac{S}{N}\right)=\pi$$

Formula:

4 * S / N = 4 * (pi * r * r) / (4 * r * r) = pi

THEORY OF Pi Calculation

To determine whether the dot is inside or outside, us the formula for distance for reference

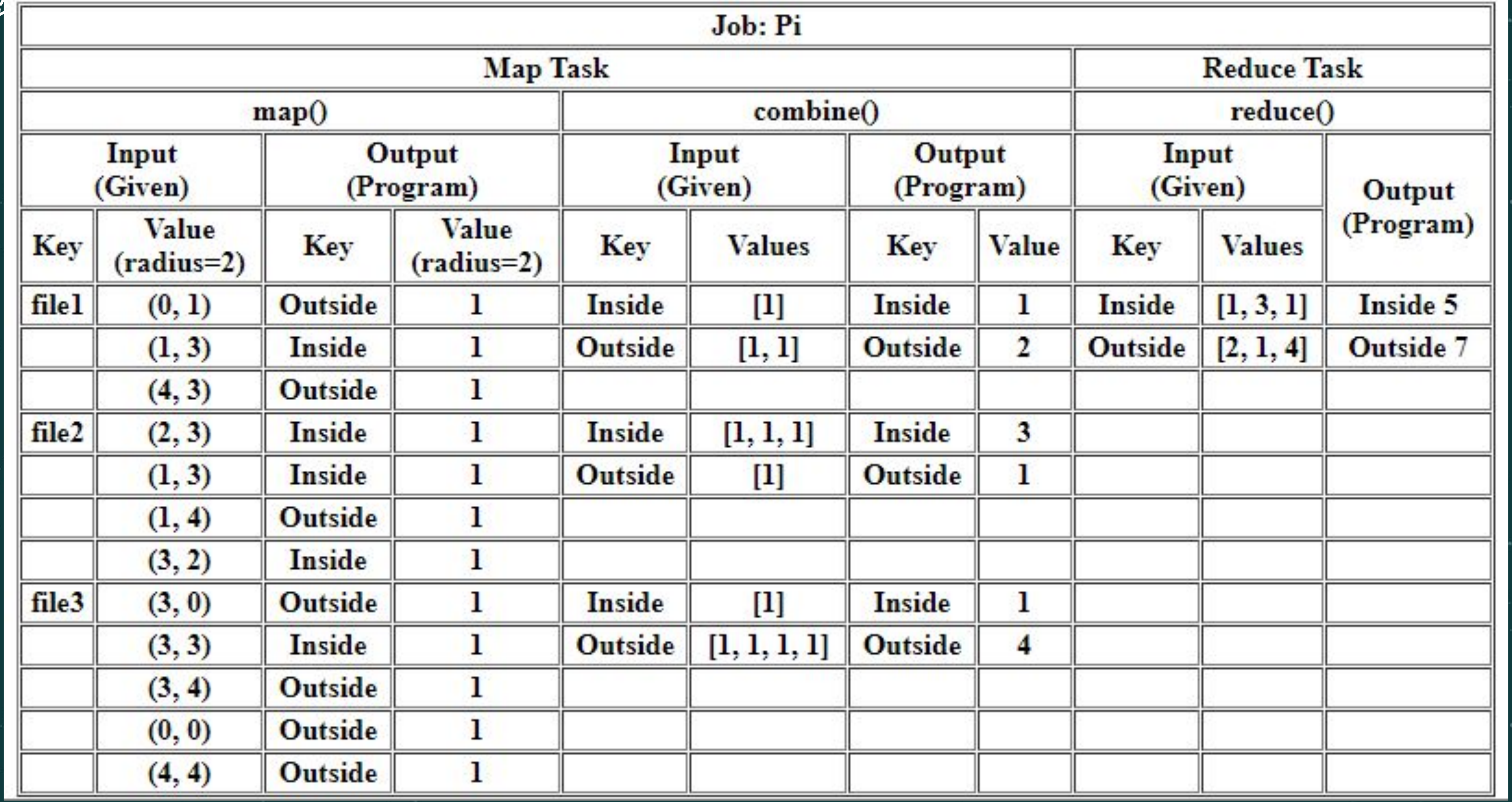


DESIGN

This section will discuss about the process and methods designed to solve pi calculation.

TECHNOLOGY USED

- Using GCP Ubuntu as project environment.
- Using Hadoop framework to implement MapReduce model.
- Program in Java Language.



Job: Pi										
Map Task								Reduce Task		
map()				combine()				reduce()		
Input (Given)		Output (Program)		Input (Given)		Output (Program)		Input (Given)		Output (Program)
Key	Value (radius=2)	Key	Value (radius=2)	Key	Values	Key	Value	Key	Values	
file1	(0, 1)	Outside	1	Inside	[1]	Inside	1	Inside	[1, 3, 1]	Inside 5
	(1, 3)	Inside	1	Outside	[1, 1]	Outside	2	Outside	[2, 1, 4]	Outside 7
	(4, 3)	Outside	1							
file2	(2, 3)	Inside	1	Inside	[1, 1, 1]	Inside	3			
	(1, 3)	Inside	1	Outside	[1]	Outside	1			
	(1, 4)	Outside	1							
	(3, 2)	Inside	1							
file3	(3, 0)	Outside	1	Inside	[1]	Inside	1			
	(3, 3)	Inside	1	Outside	[1, 1, 1, 1]	Outside	4			
	(3, 4)	Outside	1							
	(0, 0)	Outside	1							
	(4, 4)	Outside	1							

PROCESS

1. Prepare Input File

- Write a Java program to generate numbers of random pairs of point(x, y) with given radius
- Save the result in file to use as MapReduce input file

2. Code for MapReduce

- Write MapReduce program in Java Language to count number of points inside and outside of the circle with given radius.

3. Run Mapreduce on GCP

- Using the input file generated in step 1 to run MapReduce program in Step 2
- Output should be like:
  Inside xxx
  Outside xxx

4. Calculate Pi

- Write a Java Program to calculate pi value
- Using the output from Step 3 get pi value

$\sqrt{123}$

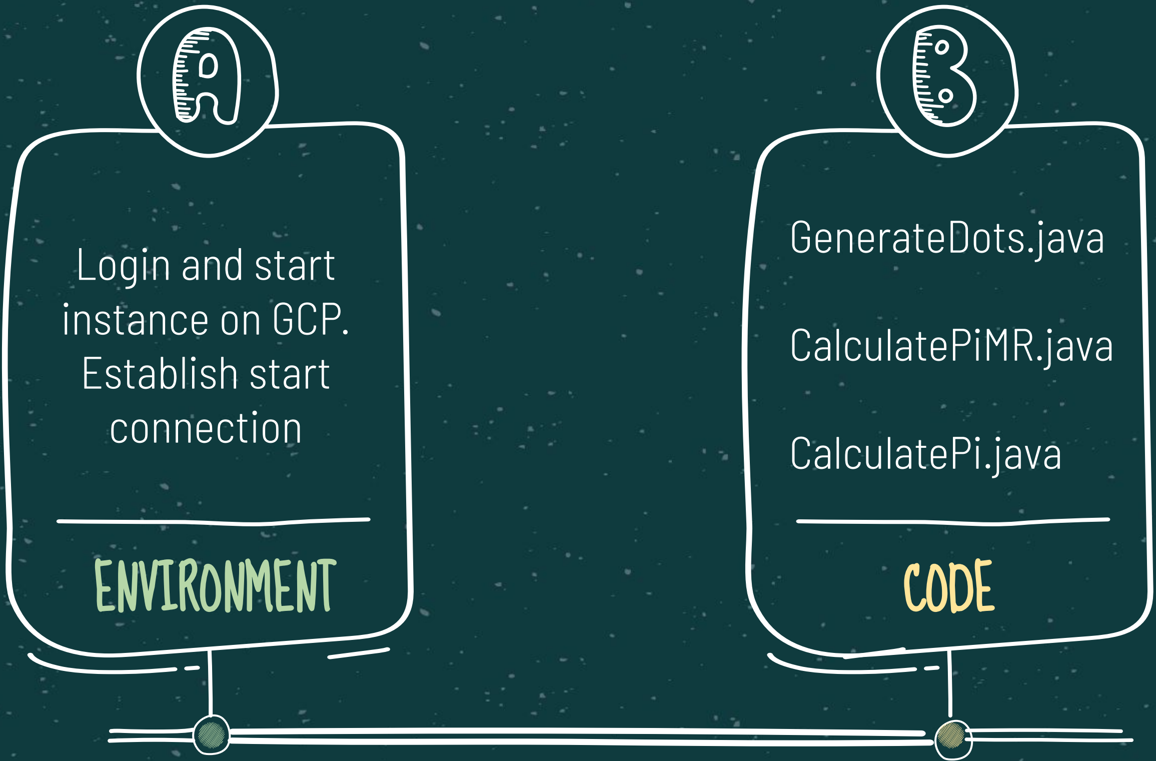
IMPLEMENTATION

Getting ready to test

STUDY HARD!

+×÷

PROJECT IMPLEMENTATION



Login and start instance on GCP. Establish start connection

ENVIRONMENT

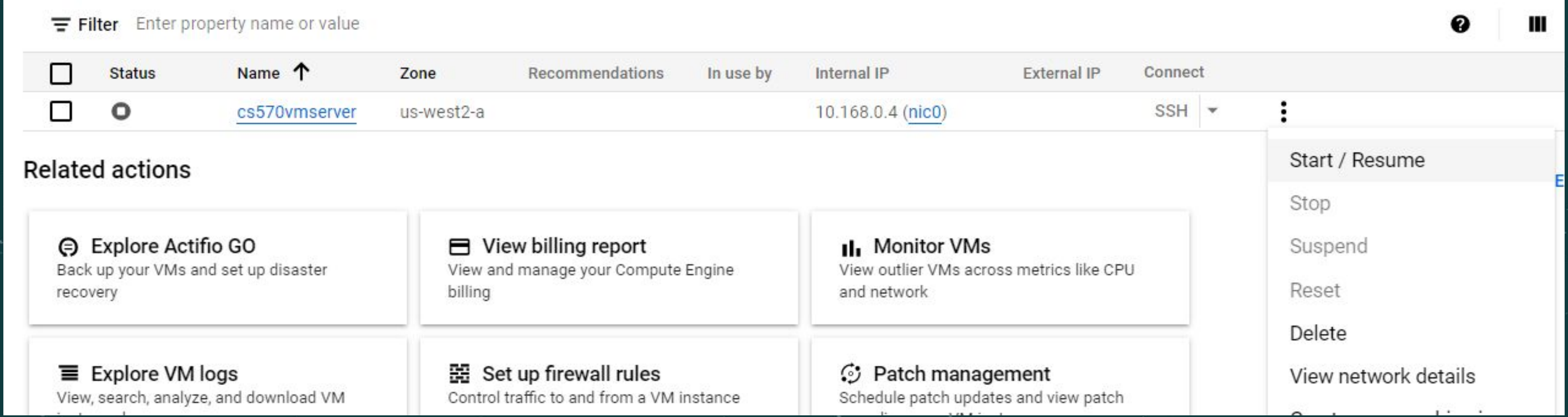
GenerateDots.java

CalculatePiMR.java

CalculatePi.java

CODE

ENVIRONMENT--GCP



Filter Enter property name or value

☐	Status	Name ↑	Zone	Recommendations	In use by	Internal IP	External IP	Connect	
☐	○	cs570vmserver	us-west2-a			10.168.0.4 (nic0)		SSH ▾	⋮

Related actions

Explore Actifio GO
Back up your VMs and set up disaster recovery

View billing report
View and manage your Compute Engine billing

Monitor VMs
View outlier VMs across metrics like CPU and network

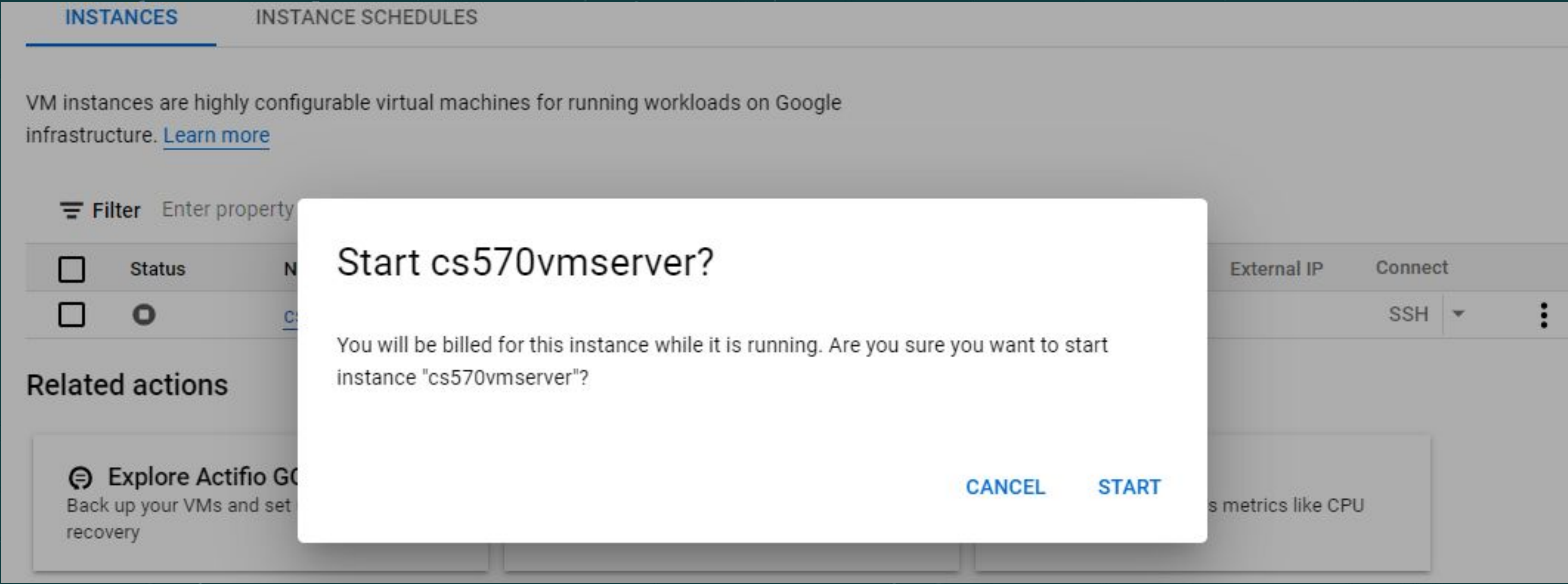
Explore VM logs
View, search, analyze, and download VM

Set up firewall rules
Control traffic to and from a VM instance

Patch management
Schedule patch updates and view patch

VM instance is stopped while not on GCP

ENVIRONMENT--GCP



VM instances are highly configurable virtual machines for running workloads on Google infrastructure. Learn more

Related actions

Explore Actifio GC

Back up your VMs and set recovery

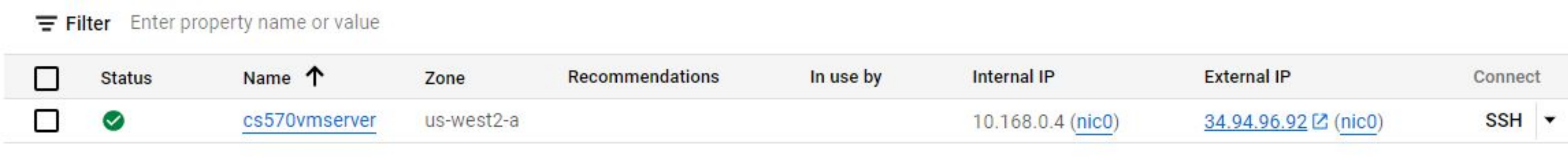
s metrics like CPU

Start cs570vmserver?

You will be billed for this instance while it is running. Are you sure you want to start instance "cs570vmserver"?

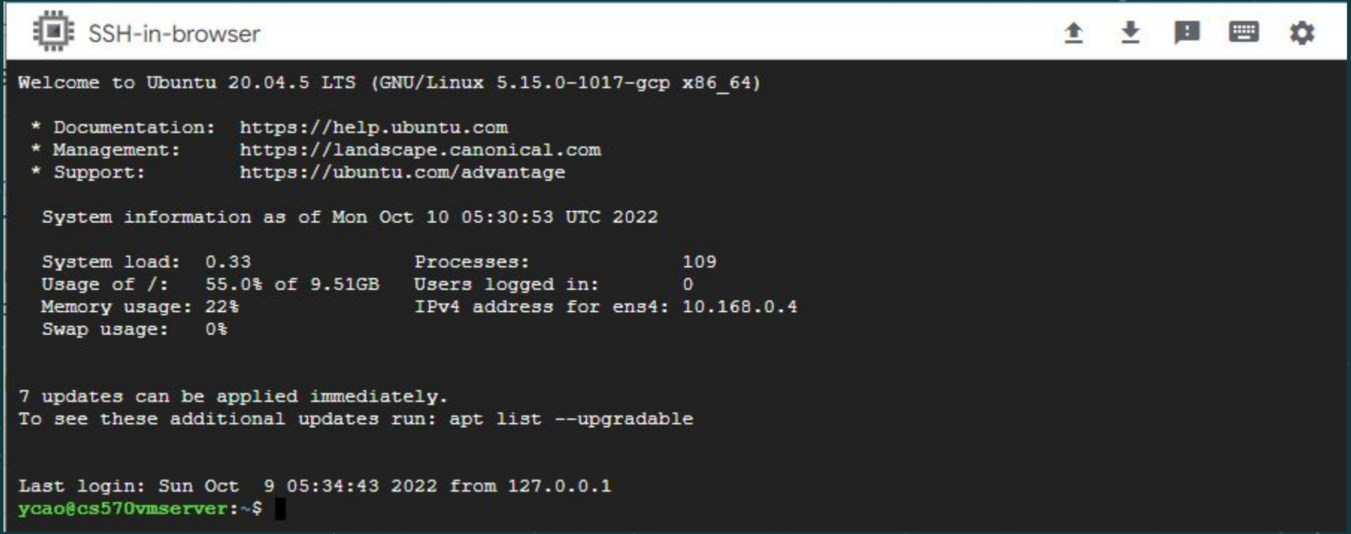
Start VM instance on GCP

ENVIRONMENT--GCP



Filter Enter property name or value

□	Status	Name ↑	Zone	Recommendations	In use by	Internal IP	External IP	Connect
□	✓	cs570vmserver	us-west2-a			10.168.0.4 (nic0)	34.94.96.92 (nic0)	SSH ▾

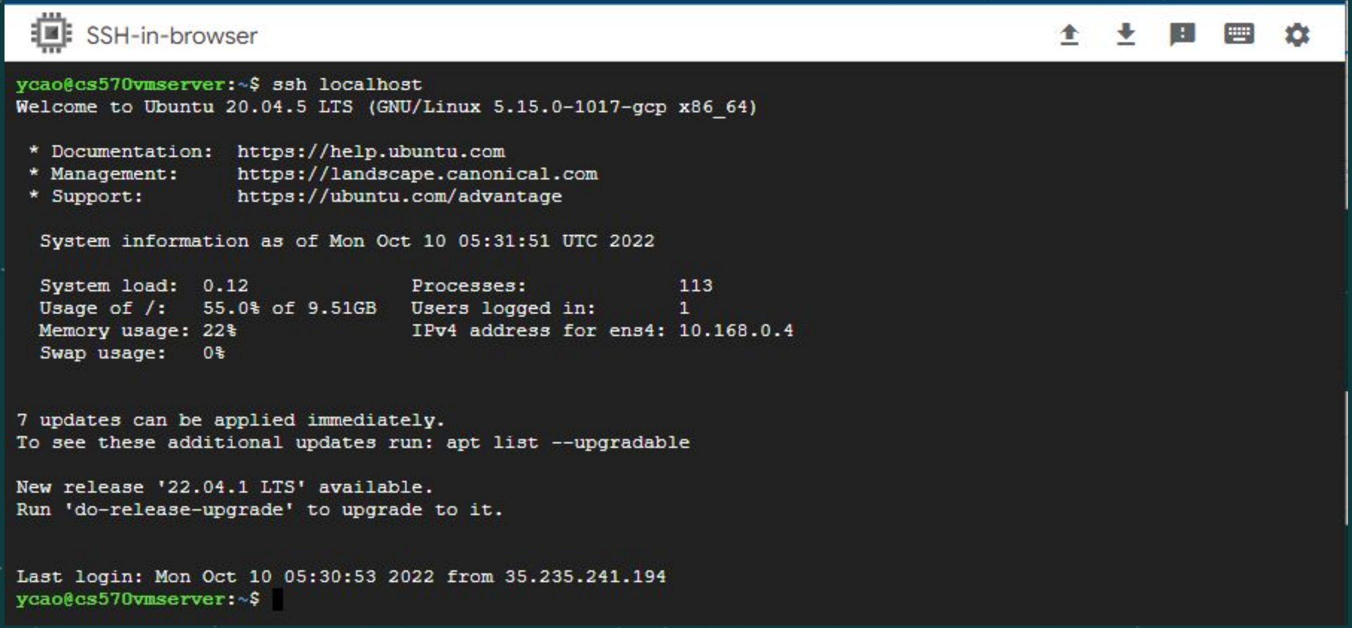


SSH-in-browser

```
Welcome to Ubuntu 20.04.5 LTS (GNU/Linux 5.15.0-1017-gcp x86_64)

 * Documentation:  https://help.ubuntu.com
 * Management:     https://landscape.canonical.com
 * Support:        https://ubuntu.com/advantage

  System information as of Mon Oct 10 05:30:53 UTC 2022

  System load:  0.33              Processes:             109
  Usage of /:   55.0% of 9.51GB   Users logged in:       0
  Memory usage: 22%               IPv4 address for ens4: 10.168.0.4
  Swap usage:   0%


7 updates can be applied immediately.
To see these additional updates run: apt list --upgradable


Last login: Sun Oct  9 05:34:43 2022 from 127.0.0.1
ycao@cs570vmserver:~$
```

Instance Started
and
Connect VM through SSH

ENVIRONMENT--Connection

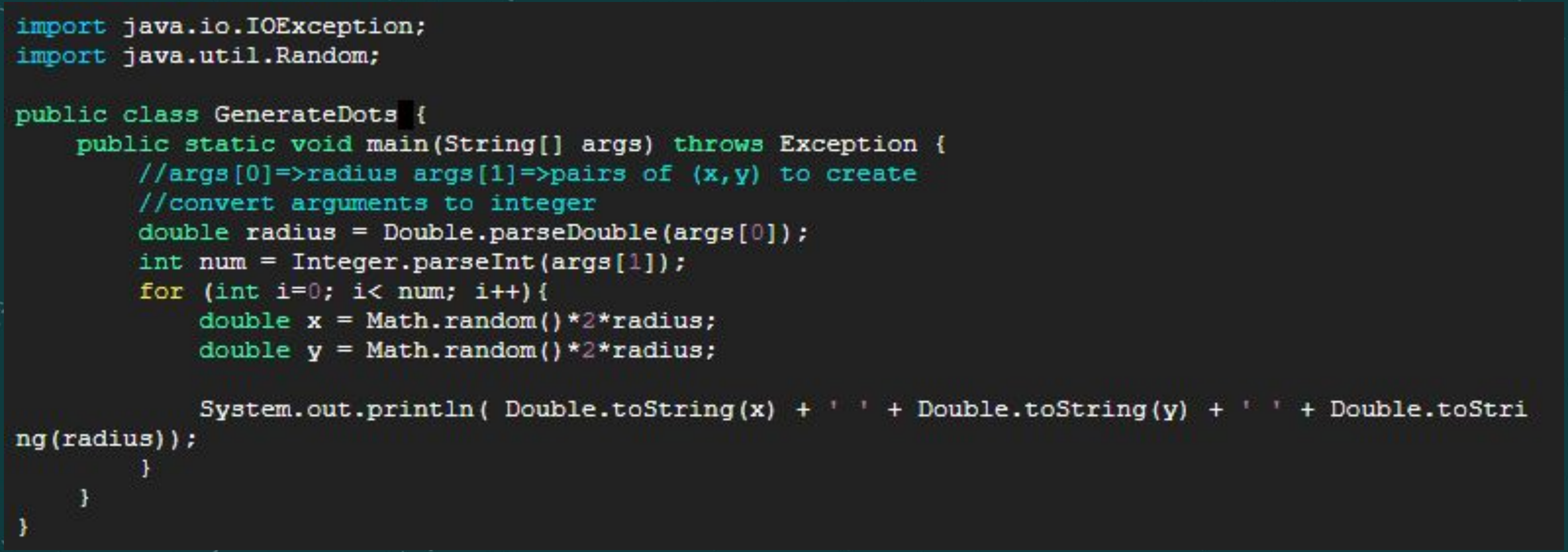


```
ycao@cs570vmserver:~$ ssh localhost
Welcome to Ubuntu 20.04.5 LTS (GNU/Linux 5.15.0-1017-gcp x86_64)

 * Documentation:  https://help.ubuntu.com
 * Management:     https://landscape.canonical.com
 * Support:        https://ubuntu.com/advantage

  System information as of Mon Oct 10 05:31:51 UTC 2022

  System load:  0.12               Processes:             113
  Usage of /:   55.0% of 9.51GB    Users logged in:       1
  Memory usage: 22%                IPv4 address for ens4: 10.168.0.4
  Swap usage:   0%


7 updates can be applied immediately.
To see these additional updates run: apt list --upgradable

New release '22.04.1 LTS' available.
Run 'do-release-upgrade' to upgrade to it.


Last login: Mon Oct 10 05:30:53 2022 from 35.235.241.194
ycao@cs570vmserver:~$
```

Connect with localhost

3

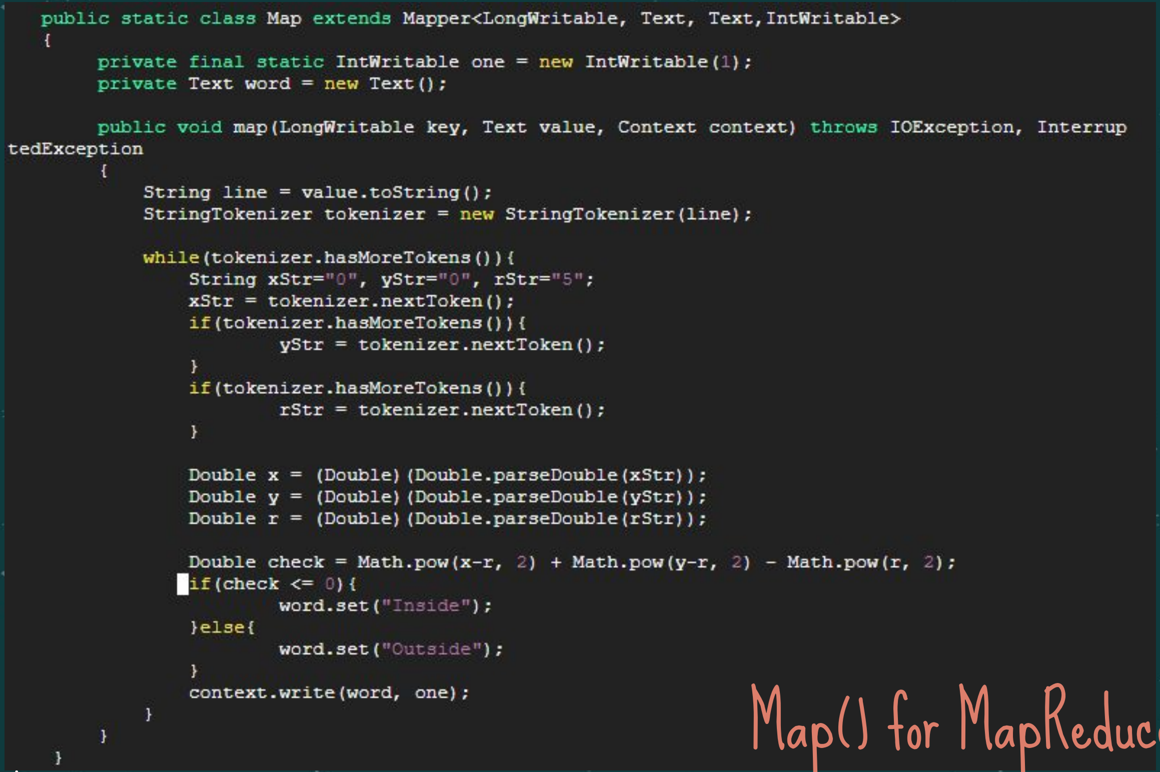
CODE--GenerateDots.java



```
import java.io.IOException;
import java.util.Random;

public class GenerateDots {
    public static void main(String[] args) throws Exception {
        //args[0]=>radius args[1]=>pairs of (x,y) to create
        //convert arguments to integer
        double radius = Double.parseDouble(args[0]);
        int num = Integer.parseInt(args[1]);
        for (int i=0; i< num; i++){
            double x = Math.random()*2*radius;
            double y = Math.random()*2*radius;

            System.out.println( Double.toString(x) + ' ' + Double.toString(y) + ' ' + Double.toStri
ng(radius));
        }
    }
}
```

Java Program to generate random dot pairs with command line arguments taken in as radius and number of pairs. Output format: x y radius

CODE--CalculatePiMR.java

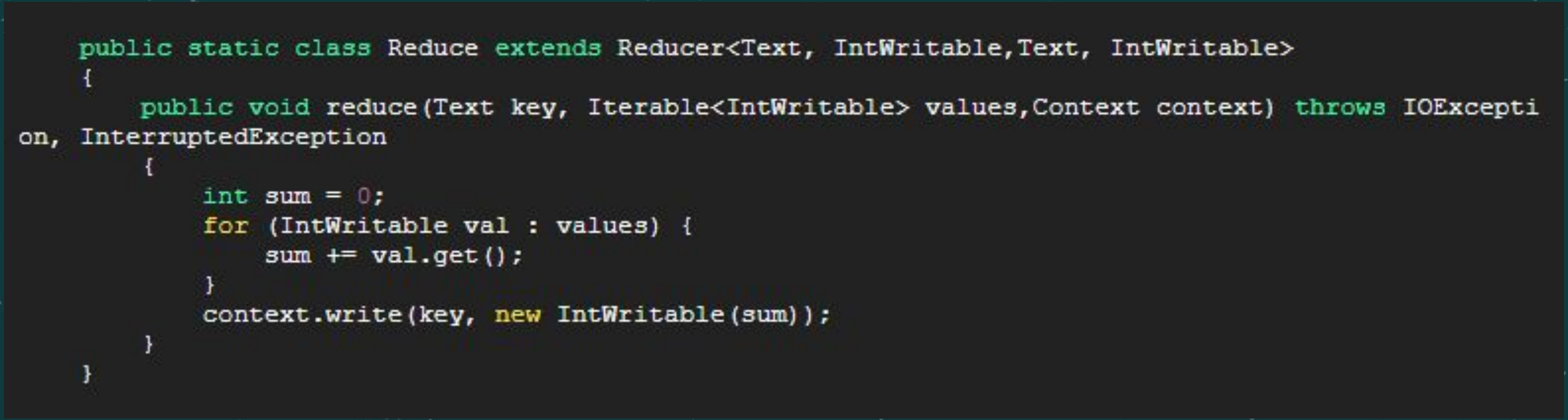


```
   public static class Map extends Mapper<LongWritable, Text, Text,IntWritable>
   {
        private final static IntWritable one = new IntWritable(1);
        private Text word = new Text();

        public void map(LongWritable key, Text value, Context context) throws IOException, Interrup
tedException
        {
            String line = value.toString();
            StringTokenizer tokenizer = new StringTokenizer(line);

            while(tokenizer.hasMoreTokens()){
                String xStr="0", yStr="0", rStr="5";
                xStr = tokenizer.nextToken();
                if(tokenizer.hasMoreTokens()){
                        yStr = tokenizer.nextToken();
                }
                if(tokenizer.hasMoreTokens()){
                        rStr = tokenizer.nextToken();
                }

                Double x = (Double)(Double.parseDouble(xStr));
                Double y = (Double)(Double.parseDouble(yStr));
                Double r = (Double)(Double.parseDouble(rStr));

                Double check = Math.pow(x-r, 2) + Math.pow(y-r, 2) - Math.pow(r, 2);
                if(check <= 0){
                        word.set("Inside");
                }else{
                        word.set("Outside");
                }
                context.write(word, one);
            }
        }
    }
```

Map() for MapReduce

3

CODE--CalculatePiMR.java

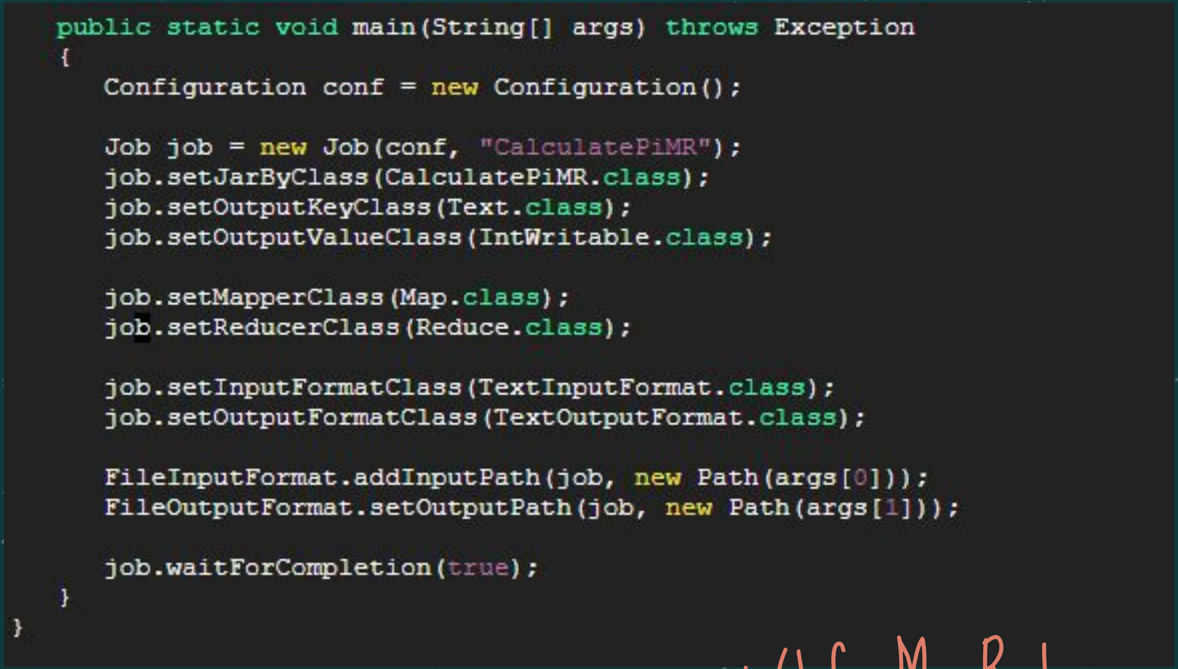


```
    public static class Reduce extends Reducer<Text, IntWritable,Text, IntWritable>
    {
        public void reduce(Text key, Iterable<IntWritable> values,Context context) throws IOExcepti
on, InterruptedException
        {
            int sum = 0;
            for (IntWritable val : values) {
                sum += val.get();
            }
            context.write(key, new IntWritable(sum));
        }
    }
```

Reduce() for MapReduce

3

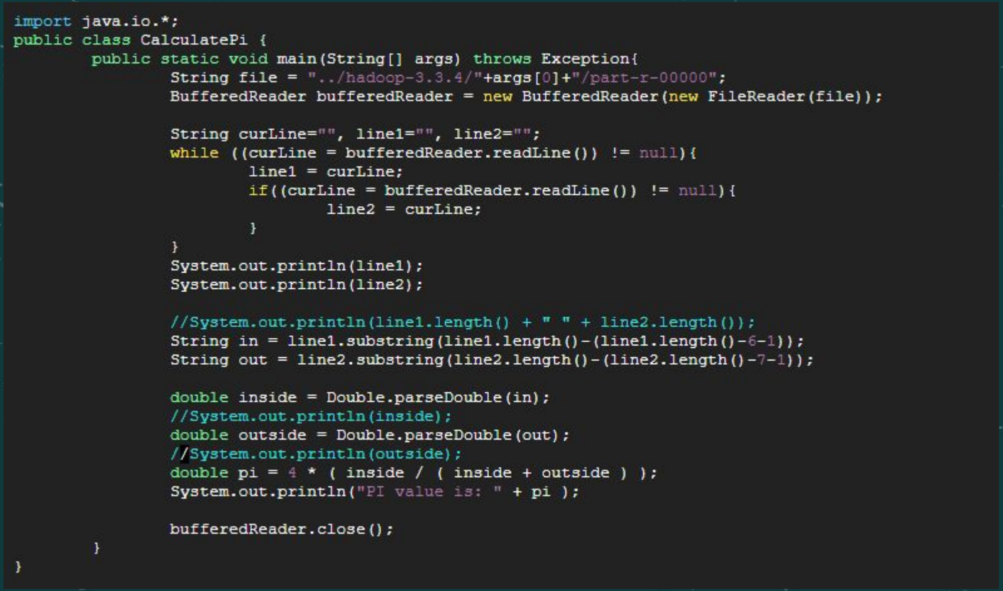
CODE--CalculatePiMR.java



```java
   public static void main(String[] args) throws Exception
   {
      Configuration conf = new Configuration();

      Job job = new Job(conf, "CalculatePiMR");
      job.setJarByClass(CalculatePiMR.class);
      job.setOutputKeyClass(Text.class);
      job.setOutputValueClass(IntWritable.class);

      job.setMapperClass(Map.class);
      job.setReducerClass(Reduce.class);

      job.setInputFormatClass(TextInputFormat.class);
      job.setOutputFormatClass(TextOutputFormat.class);

      FileInputFormat.addInputPath(job, new Path(args[0]));
      FileOutputFormat.setOutputPath(job, new Path(args[1]));

      job.waitForCompletion(true);
   }
}
```

main() for MapReduce

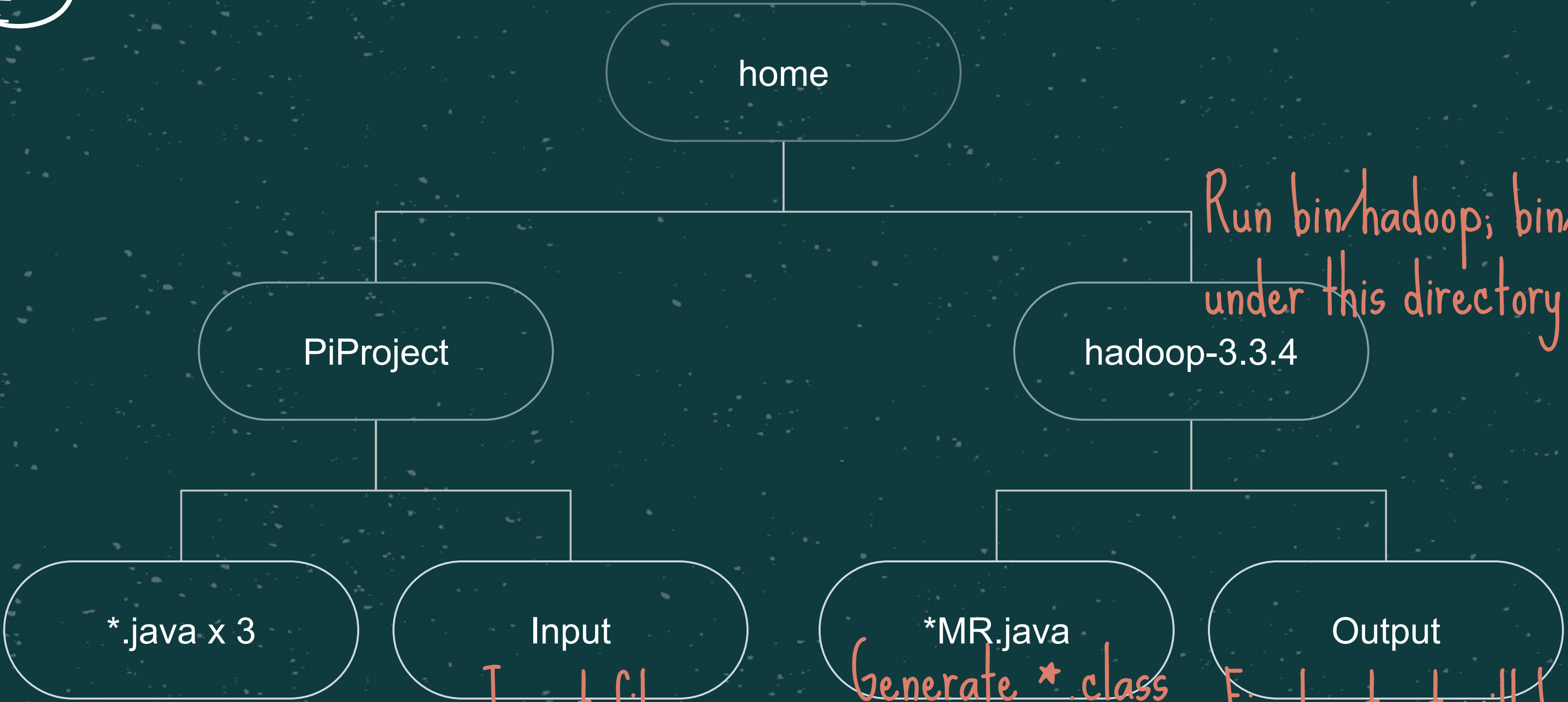
CODE--CalculatePi.java



```
import java.io.*;
public class CalculatePi {
        public static void main(String[] args) throws Exception{
                String file = "../hadoop-3.3.4/"+args[0]+"/part-r-00000";
                BufferedReader bufferedReader = new BufferedReader(new FileReader(file));

                String curLine="", line1="", line2="";
                while ((curLine = bufferedReader.readLine()) != null){
                        line1 = curLine;
                        if((curLine = bufferedReader.readLine()) != null){
                                line2 = curLine;
                        }
                }
                System.out.println(line1);
                System.out.println(line2);

                //System.out.println(line1.length() + " " + line2.length());
                String in = line1.substring(line1.length()-(line1.length()-6-1));
                String out = line2.substring(line2.length()-(line2.length()-7-1));

                double inside = Double.parseDouble(in);
                //System.out.println(inside);
                double outside = Double.parseDouble(out);
                //System.out.println(outside);
                double pi = 4 * ( inside / ( inside + outside ) );
                System.out.println("PI value is: " + pi );

                bufferedReader.close();
        }
}
```

Java Program to calculate pi value with MapReduce result taken in by reading the file.

CODE--Structure



home

PiProject

hadoop-3.3.4

*.java x 3

Input

*MR.java

Output

Run bin/hadoop; bin/hdfs under this directory

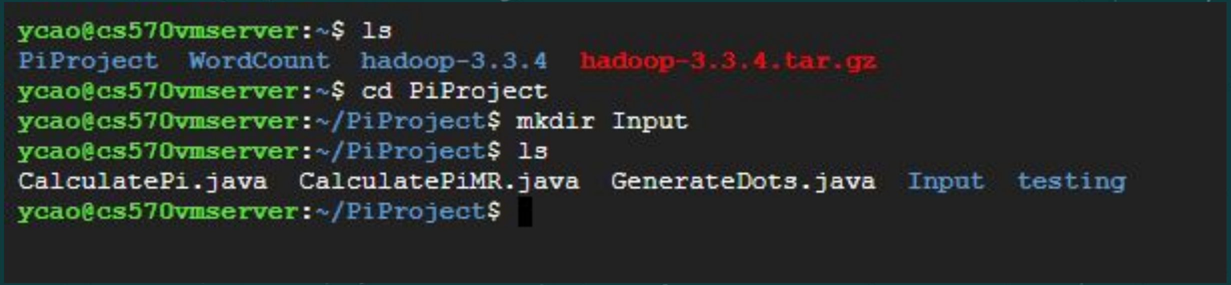
Input files

Generate *.class files here

Final output will be saved locally in this folder

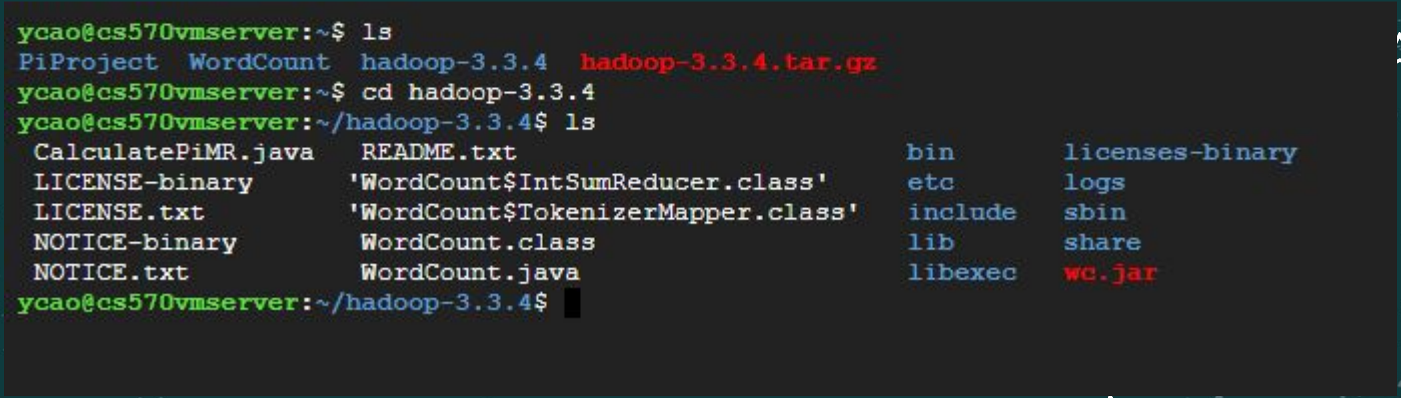
3

CODE--Structure



```
ycao@cs570vmserver:~$ ls
PiProject  WordCount  hadoop-3.3.4  hadoop-3.3.4.tar.gz
ycao@cs570vmserver:~$ cd PiProject
ycao@cs570vmserver:~/PiProject$ mkdir Input
ycao@cs570vmserver:~/PiProject$ ls
CalculatePi.java  CalculatePiMR.java  GenerateDots.java  Input  testing
ycao@cs570vmserver:~/PiProject$
```

PiProject directory



```
ycao@cs570vmserver:~$ ls
PiProject  WordCount  hadoop-3.3.4  hadoop-3.3.4.tar.gz
ycao@cs570vmserver:~$ cd hadoop-3.3.4
ycao@cs570vmserver:~/hadoop-3.3.4$ ls
 CalculatePiMR.java   README.txt                               bin        licenses-binary
 LICENSE-binary      'WordCount$IntSumReducer.class'           etc        logs
 LICENSE.txt         'WordCount$TokenizerMapper.class'         include    sbin
 NOTICE-binary        WordCount.class                          lib        share
 NOTICE.txt           WordCount.java                           libexec    wc.jar
ycao@cs570vmserver:~/hadoop-3.3.4$
```

hadoop-3.3.4 directory

TEST

Process to test the project

GCP-HADOOP-MAPREDUCE
STEPS & RESULT

1

STEPS

Detailed steps of running the project and outputs.

2

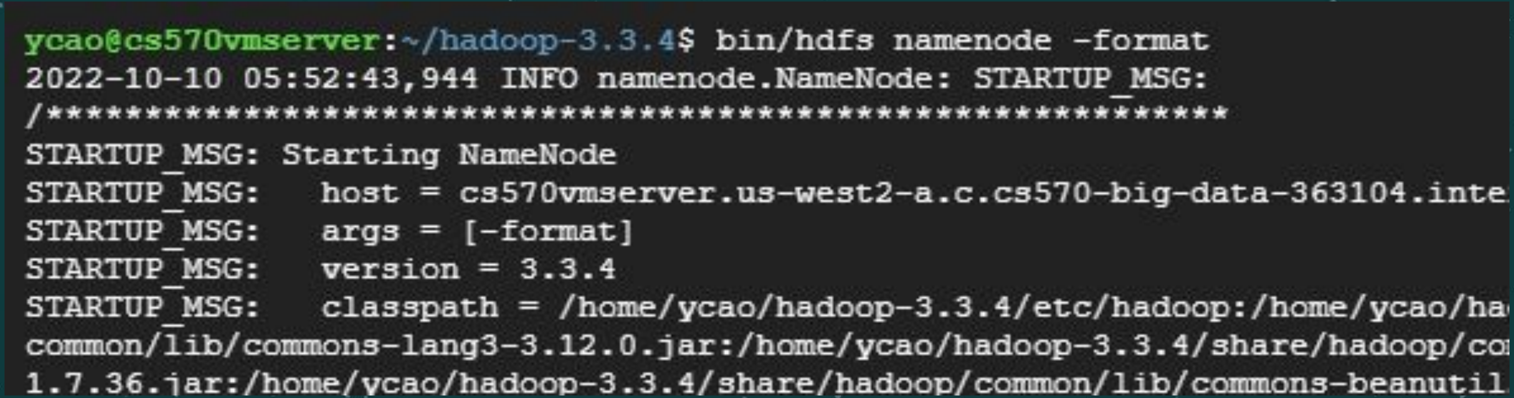
RESULT

Final Result for pi value calculated.

1

STEPS

$ bin/hdfs namenode -format



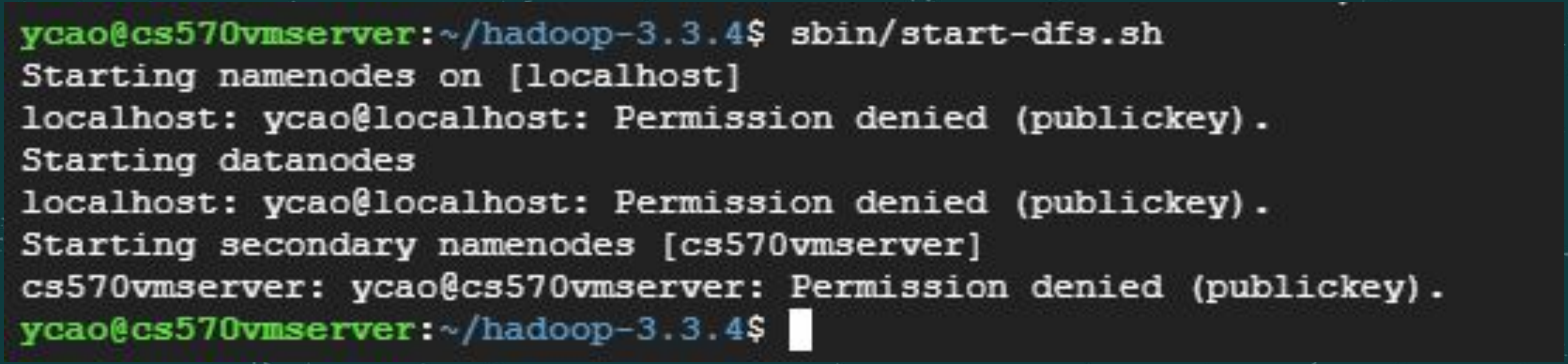
```
ycao@cs570vmserver:~/hadoop-3.3.4$ bin/hdfs namenode -format
2022-10-10 05:52:43,944 INFO namenode.NameNode: STARTUP_MSG:
/************************************************************
STARTUP_MSG: Starting NameNode
STARTUP_MSG:   host = cs570vmserver.us-west2-a.c.cs570-big-data-363104.inte
STARTUP_MSG:   args = [-format]
STARTUP_MSG:   version = 3.3.4
STARTUP_MSG:   classpath = /home/ycao/hadoop-3.3.4/etc/hadoop:/home/ycao/ha
common/lib/commons-lang3-3.12.0.jar:/home/ycao/hadoop-3.3.4/share/hadoop/co
1.7.36.jar:/home/ycao/hadoop-3.3.4/share/hadoop/common/lib/commons-beanutil
```

Format the file system

1

STEPS

$ sbin/start-dfs.sh

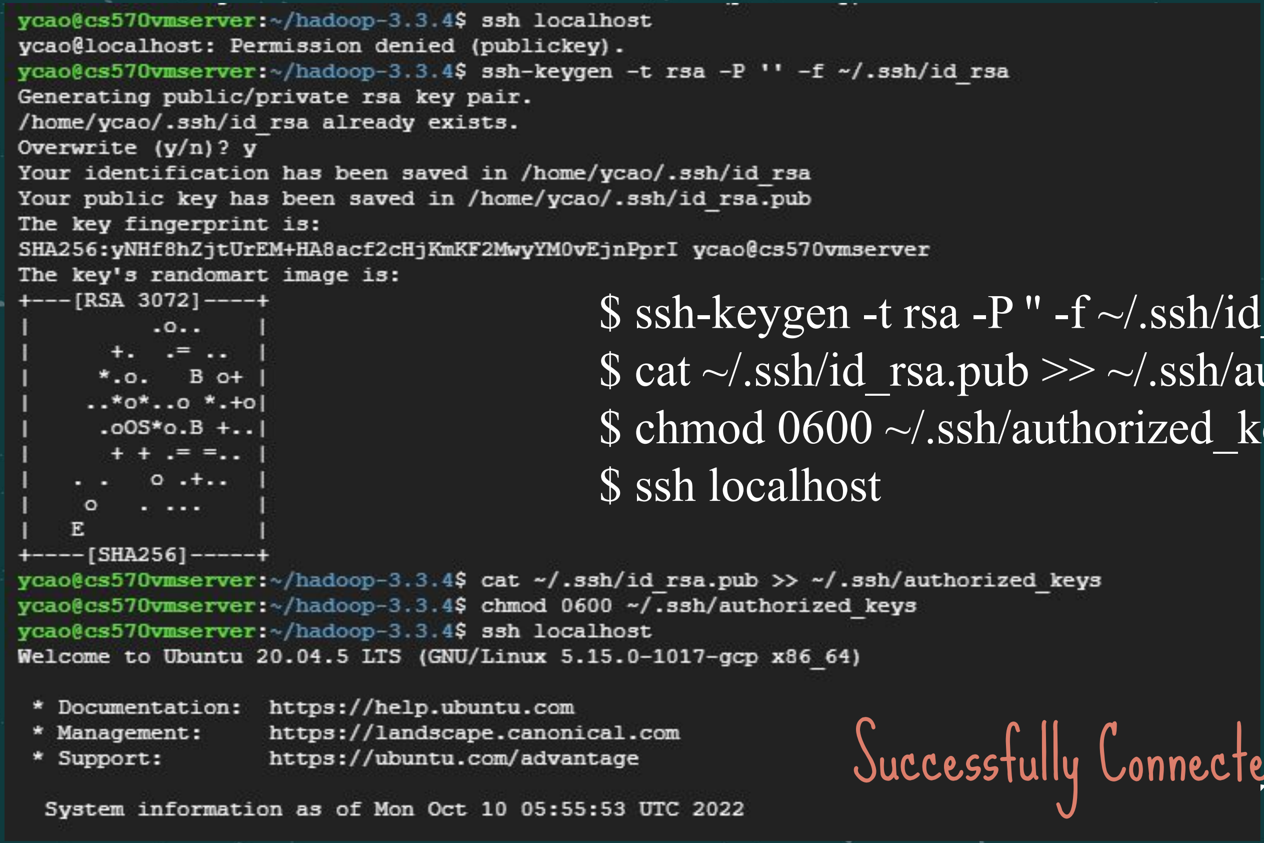


```
ycao@cs570vmserver:~/hadoop-3.3.4$ sbin/start-dfs.sh
Starting namenodes on [localhost]
localhost: ycao@localhost: Permission denied (publickey).
Starting datanodes
localhost: ycao@localhost: Permission denied (publickey).
Starting secondary namenodes [cs570vmserver]
cs570vmserver: ycao@cs570vmserver: Permission denied (publickey).
ycao@cs570vmserver:~/hadoop-3.3.4$
```

Start NameNode daemon and DataNode daemon
Permission Denied, need to connect ssh again.

1

STEPS



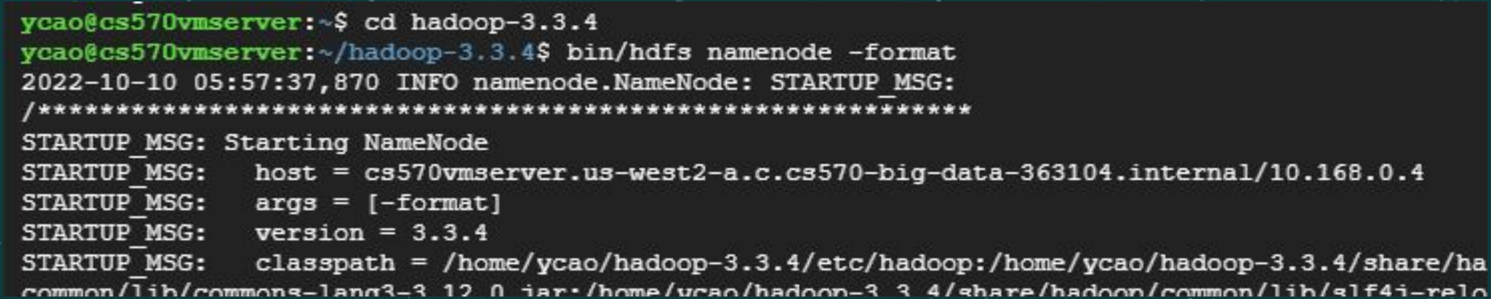
```
ycao@cs570vmserver:~/hadoop-3.3.4$ ssh localhost
ycao@localhost: Permission denied (publickey).
ycao@cs570vmserver:~/hadoop-3.3.4$ ssh-keygen -t rsa -P '' -f ~/.ssh/id_rsa
Generating public/private rsa key pair.
/home/ycao/.ssh/id_rsa already exists.
Overwrite (y/n)? y
Your identification has been saved in /home/ycao/.ssh/id_rsa
Your public key has been saved in /home/ycao/.ssh/id_rsa.pub
The key fingerprint is:
SHA256:yNHf8hZjtUrEM+HA8acf2cHjKmKF2MwyYM0vEjnPprI ycao@cs570vmserver
The key's randomart image is:
+---[RSA 3072]----+
|        .o..     |
|      +.  .= ..  |
|     *.o.   B o+ |
|    ..*o*..o *.+o|
|     .oOS*o.B +..|
|      + + .= =.. |
|   . .   o .+..  |
|    o   . ...    |
|   E             |
+----[SHA256]-----+
ycao@cs570vmserver:~/hadoop-3.3.4$ cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys
ycao@cs570vmserver:~/hadoop-3.3.4$ chmod 0600 ~/.ssh/authorized_keys
ycao@cs570vmserver:~/hadoop-3.3.4$ ssh localhost
Welcome to Ubuntu 20.04.5 LTS (GNU/Linux 5.15.0-1017-gcp x86_64)

 * Documentation:  https://help.ubuntu.com
 * Management:     https://landscape.canonical.com
 * Support:        https://ubuntu.com/advantage

  System information as of Mon Oct 10 05:55:53 UTC 2022
```

```
$ ssh-keygen -t rsa -P '' -f ~/.ssh/id_rsa
$ cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys
$ chmod 0600 ~/.ssh/authorized_keys
$ ssh localhost
```

Successfully Connected!

1

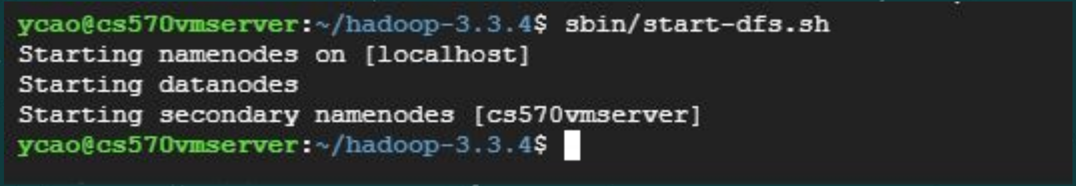
STEPS



```
ycao@cs570vmserver:~$ cd hadoop-3.3.4
ycao@cs570vmserver:~/hadoop-3.3.4$ bin/hdfs namenode -format
2022-10-10 05:57:37,870 INFO namenode.NameNode: STARTUP_MSG:
/************************************************************
STARTUP_MSG: Starting NameNode
STARTUP_MSG:   host = cs570vmserver.us-west2-a.c.cs570-big-data-363104.internal/10.168.0.4
STARTUP_MSG:   args = [-format]
STARTUP_MSG:   version = 3.3.4
STARTUP_MSG:   classpath = /home/ycao/hadoop-3.3.4/etc/hadoop:/home/ycao/hadoop-3.3.4/share/ha
common/lib/commons-lang3-3.12.0.jar:/home/ycao/hadoop-3.3.4/share/hadoop/common/lib/slf4j-relo
```

Format again!

Successful started!

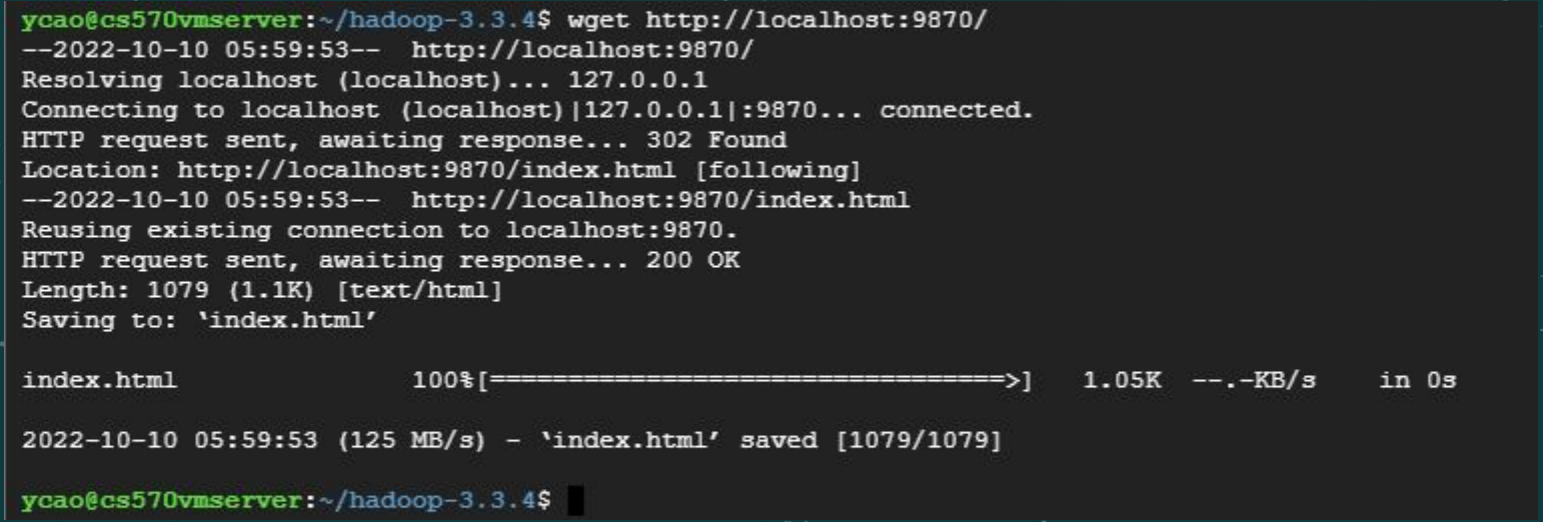


```
ycao@cs570vmserver:~/hadoop-3.3.4$ sbin/start-dfs.sh
Starting namenodes on [localhost]
Starting datanodes
Starting secondary namenodes [cs570vmserver]
ycao@cs570vmserver:~/hadoop-3.3.4$ █
```

1

STEPS

$ wget http://localhost:9870/



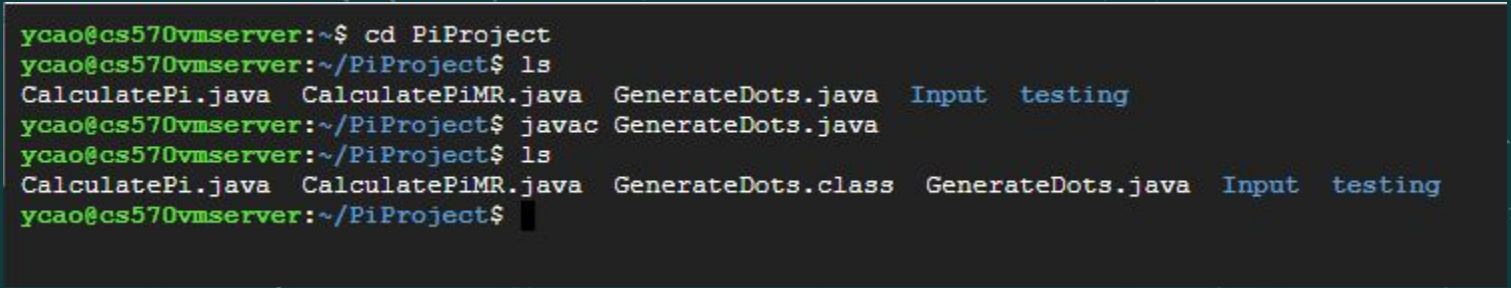
```
ycao@cs570vmserver:~/hadoop-3.3.4$ wget http://localhost:9870/
--2022-10-10 05:59:53--  http://localhost:9870/
Resolving localhost (localhost)... 127.0.0.1
Connecting to localhost (localhost)|127.0.0.1|:9870... connected.
HTTP request sent, awaiting response... 302 Found
Location: http://localhost:9870/index.html [following]
--2022-10-10 05:59:53--  http://localhost:9870/index.html
Reusing existing connection to localhost:9870.
HTTP request sent, awaiting response... 200 OK
Length: 1079 (1.1K) [text/html]
Saving to: 'index.html'

index.html              100%[==================================>]   1.05K  --.-KB/s    in 0s

2022-10-10 05:59:53 (125 MB/s) - 'index.html' saved [1079/1079]

ycao@cs570vmserver:~/hadoop-3.3.4$
```

Test Connection with localhost

1

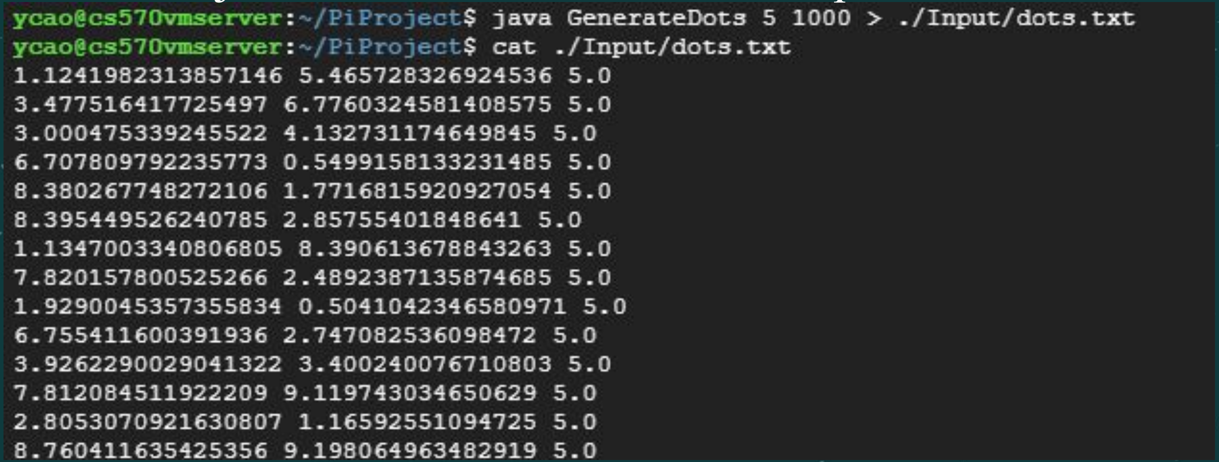
STEPS

$ javac GenerateDots.java



```
ycao@cs570vmserver:~$ cd PiProject
ycao@cs570vmserver:~/PiProject$ ls
CalculatePi.java  CalculatePiMR.java  GenerateDots.java  Input  testing
ycao@cs570vmserver:~/PiProject$ javac GenerateDots.java
ycao@cs570vmserver:~/PiProject$ ls
CalculatePi.java  CalculatePiMR.java  GenerateDots.class  GenerateDots.java  Input  testing
ycao@cs570vmserver:~/PiProject$ 
```

$ java GenerateDots 5 1000 > ./Input/dots.txt



```
ycao@cs570vmserver:~/PiProject$ java GenerateDots 5 1000 > ./Input/dots.txt
ycao@cs570vmserver:~/PiProject$ cat ./Input/dots.txt
1.1241982313857146 5.465728326924536 5.0
3.477516417725497 6.776032458140857 5 5.0
3.000475339245522 4.132731174649845 5.0
6.707809792235773 0.5499158133231485 5.0
8.380267748272106 1.7716815920927054 5.0
8.395449526240785 2.85755401848641 5.0
1.1347003340806805 8.390613678843263 5.0
7.820157800525266 2.4892387135874685 5.0
1.9290045357355834 0.5041042346580971 5.0
6.755411600391936 2.747082536098472 5.0
3.9262290029041322 3.400240076710803 5.0
7.812084511922209 9.119743034650629 5.0
2.8053070921630807 1.16592551094725 5.0
8.760411635425356 9.198064963482919 5.0
```

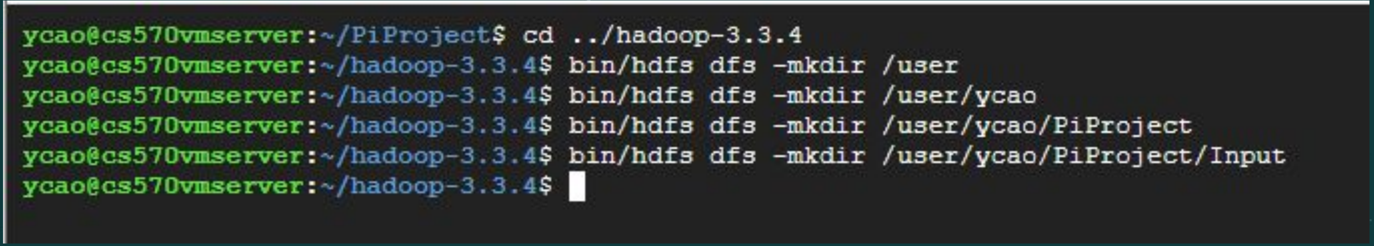
Compile and run java program to generate dots with radius=5, number = 1000

Output save in ./Input/dots.txt

1

STEPS

$ bin/hdfs dfs -mkdir /user/ycao/PiProject/Input

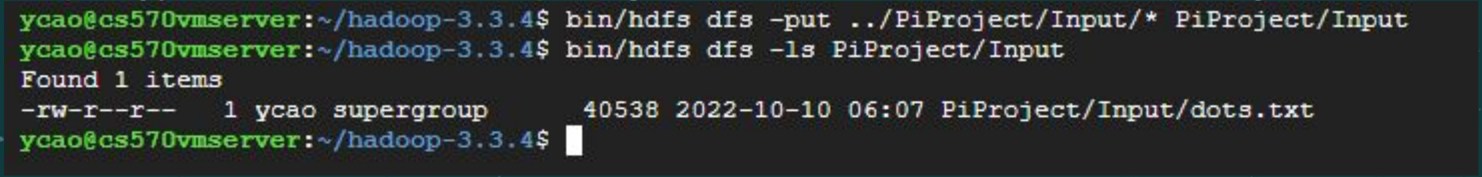


```
ycao@cs570vmserver:~/PiProject$ cd ../hadoop-3.3.4
ycao@cs570vmserver:~/hadoop-3.3.4$ bin/hdfs dfs -mkdir /user
ycao@cs570vmserver:~/hadoop-3.3.4$ bin/hdfs dfs -mkdir /user/ycao
ycao@cs570vmserver:~/hadoop-3.3.4$ bin/hdfs dfs -mkdir /user/ycao/PiProject
ycao@cs570vmserver:~/hadoop-3.3.4$ bin/hdfs dfs -mkdir /user/ycao/PiProject/Input
ycao@cs570vmserver:~/hadoop-3.3.4$ 
```

$ bin/hdfs dfs -put ../PiProject/Input/* PiProject/Input

$ bin/hdfs dfs -ls PiProject/Input

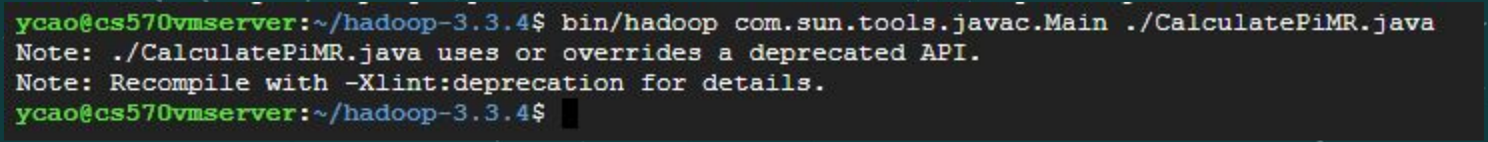
Copy file from local to hadoop and check



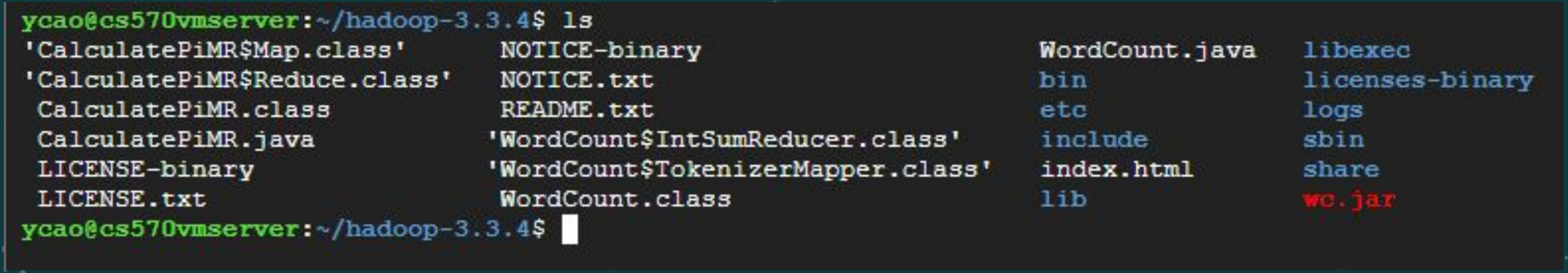
```
ycao@cs570vmserver:~/hadoop-3.3.4$ bin/hdfs dfs -put ../PiProject/Input/* PiProject/Input
ycao@cs570vmserver:~/hadoop-3.3.4$ bin/hdfs dfs -ls PiProject/Input
Found 1 items
-rw-r--r--   1 ycao supergroup      40538 2022-10-10 06:07 PiProject/Input/dots.txt
ycao@cs570vmserver:~/hadoop-3.3.4$ 
```

1

STEPS



```
ycao@cs570vmserver:~/hadoop-3.3.4$ bin/hadoop com.sun.tools.javac.Main ./CalculatePiMR.java
Note: ./CalculatePiMR.java uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
ycao@cs570vmserver:~/hadoop-3.3.4$
```



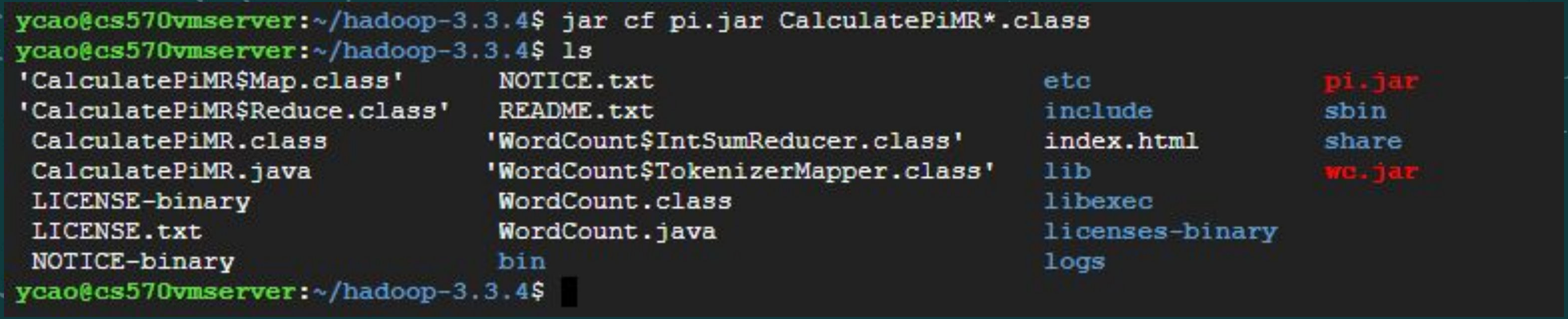
```
ycao@cs570vmserver:~/hadoop-3.3.4$ ls
'CalculatePiMR$Map.class'       NOTICE-binary                         WordCount.java   libexec
'CalculatePiMR$Reduce.class'    NOTICE.txt                            bin              licenses-binary
 CalculatePiMR.class            README.txt                            etc              logs
 CalculatePiMR.java            'WordCount$IntSumReducer.class'        include          sbin
 LICENSE-binary                'WordCount$TokenizerMapper.class'      index.html       share
 LICENSE.txt                    WordCount.class                       lib              wc.jar
ycao@cs570vmserver:~/hadoop-3.3.4$
```

$ bin/hadoop com.sun.tools.javac.Main ./CalculatePiMR.java

Compile Mapreduce program in Hadoop with *.class files created

1

STEPS



```
ycao@cs570vmserver:~/hadoop-3.3.4$ jar cf pi.jar CalculatePiMR*.class
ycao@cs570vmserver:~/hadoop-3.3.4$ ls
'CalculatePiMR$Map.class'      NOTICE.txt                            etc               pi.jar
'CalculatePiMR$Reduce.class'   README.txt                            include           sbin
 CalculatePiMR.class          'WordCount$IntSumReducer.class'        index.html        share
 CalculatePiMR.java           'WordCount$TokenizerMapper.class'      lib               wc.jar
 LICENSE-binary                WordCount.class                       libexec
 LICENSE.txt                   WordCount.java                        licenses-binary
 NOTICE-binary                 bin                                   logs
ycao@cs570vmserver:~/hadoop-3.3.4$ 
```

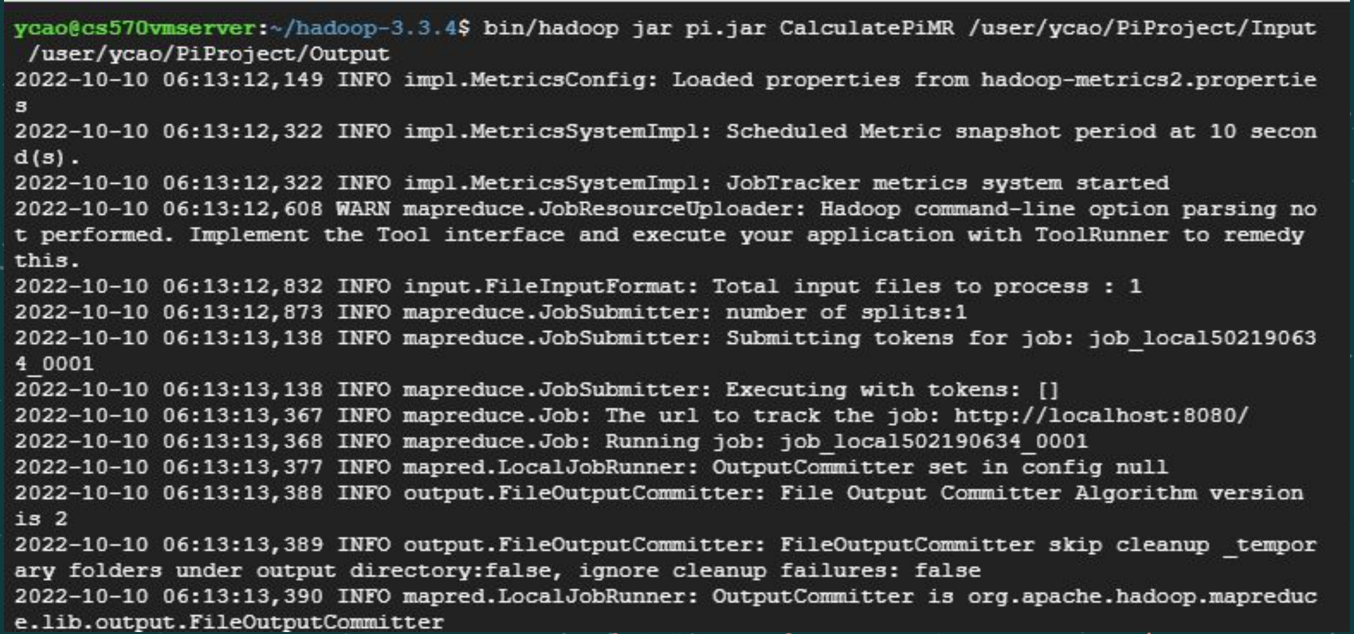
$ jar cf pi.jar CalculatePiMR*.class

Create .jar file with *.class files

STEPS

1

$ bin/hadoop jar pi.jar CalculatePiMR /user/ycao/PiProject/Input /user/ycao/PiProject/Output

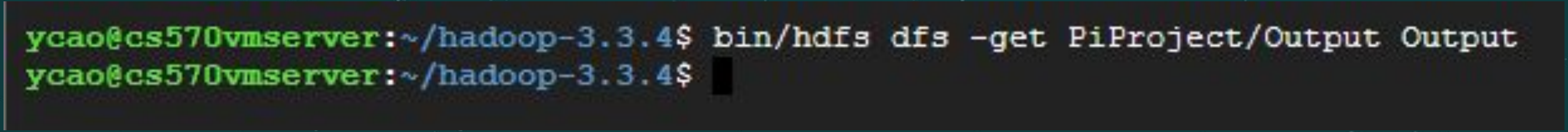


```
ycao@cs570vmserver:~/hadoop-3.3.4$ bin/hadoop jar pi.jar CalculatePiMR /user/ycao/PiProject/Input
 /user/ycao/PiProject/Output
2022-10-10 06:13:12,149 INFO impl.MetricsConfig: Loaded properties from hadoop-metrics2.propertie
s
2022-10-10 06:13:12,322 INFO impl.MetricsSystemImpl: Scheduled Metric snapshot period at 10 secon
d(s).
2022-10-10 06:13:12,322 INFO impl.MetricsSystemImpl: JobTracker metrics system started
2022-10-10 06:13:12,608 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing no
t performed. Implement the Tool interface and execute your application with ToolRunner to remedy
this.
2022-10-10 06:13:12,832 INFO input.FileInputFormat: Total input files to process : 1
2022-10-10 06:13:12,873 INFO mapreduce.JobSubmitter: number of splits:1
2022-10-10 06:13:13,138 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_local50219063
4_0001
2022-10-10 06:13:13,138 INFO mapreduce.JobSubmitter: Executing with tokens: []
2022-10-10 06:13:13,367 INFO mapreduce.Job: The url to track the job: http://localhost:8080/
2022-10-10 06:13:13,368 INFO mapreduce.Job: Running job: job_local502190634_0001
2022-10-10 06:13:13,377 INFO mapred.LocalJobRunner: OutputCommitter set in config null
2022-10-10 06:13:13,388 INFO output.FileOutputCommitter: File Output Committer Algorithm version
is 2
2022-10-10 06:13:13,389 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _tempor
ary folders under output directory:false, ignore cleanup failures: false
2022-10-10 06:13:13,390 INFO mapred.LocalJobRunner: OutputCommitter is org.apache.hadoop.mapreduc
e.lib.output.FileOutputCommitter
```

Run MapReduce Program with input file and save result in Output

2

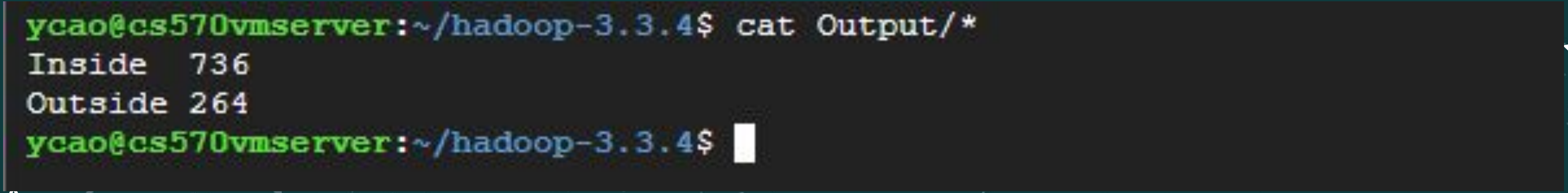
RESULT



```
ycao@cs570vmserver:~/hadoop-3.3.4$ bin/hdfs dfs -get PiProject/Output Output
ycao@cs570vmserver:~/hadoop-3.3.4$ 
```

$ bin/hdfs dfs -get PiProject/Output Output

Get output and save to local



```
ycao@cs570vmserver:~/hadoop-3.3.4$ cat Output/*
Inside  736
Outside 264
ycao@cs570vmserver:~/hadoop-3.3.4$ 
```

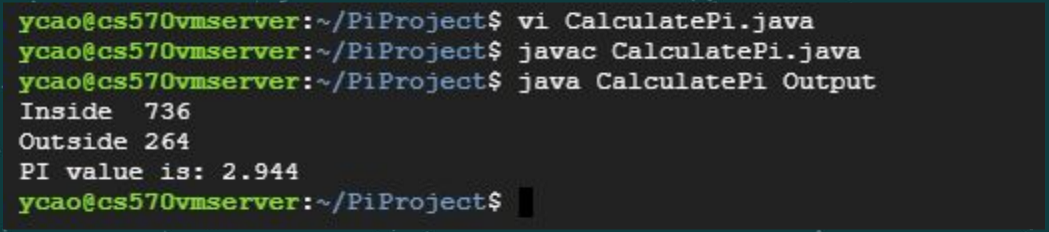
$ cat Output/*

Display Output

2

RESULT

$ jvac CalculatePi.java
$ java CalculatePi Output



```
ycao@cs570vmserver:~/PiProject$ vi CalculatePi.java
ycao@cs570vmserver:~/PiProject$ javac CalculatePi.java
ycao@cs570vmserver:~/PiProject$ java CalculatePi Output
Inside  736
Outside 264
PI value is: 2.944
ycao@cs570vmserver:~/PiProject$
```

Using the output (local output folder as command line arguments) from MapReduce Program to compile and run java program to get pi value

The pi value calculated is 2.944, and it is quite off from 3.1415926

$\sqrt{123}$

ENHANCEMENT

Can we get better result?

STUDY
HARD!

+×÷

HOW TO PROVE TEST RESULTS?

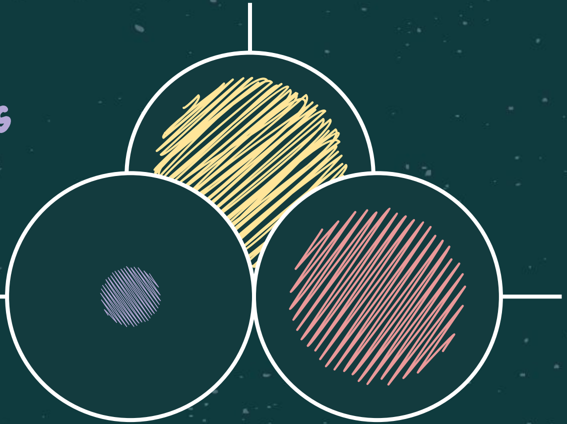
2.944

radius = 5
number = 1000

Base Case

To make the dots cover more area, we can decrease the radius

Or to increase the number of dots



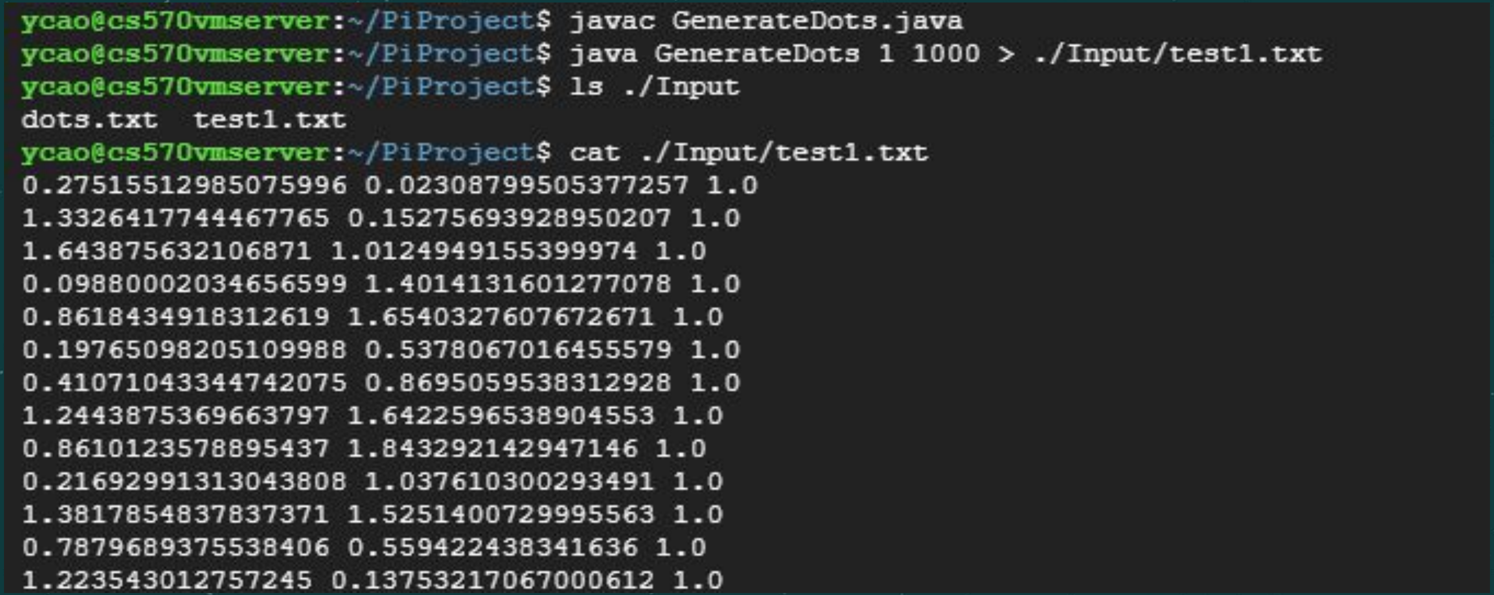
?

radius = 1
number = 1000

?

radius = 5
number = 100000

ENHANCED RESULT -- Decrease Radius

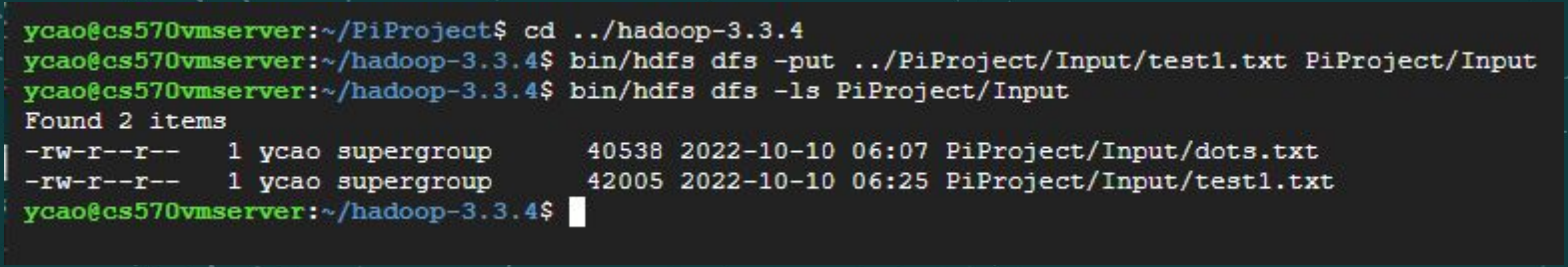


```
ycao@cs570vmserver:~/PiProject$ javac GenerateDots.java
ycao@cs570vmserver:~/PiProject$ java GenerateDots 1 1000 > ./Input/test1.txt
ycao@cs570vmserver:~/PiProject$ ls ./Input
dots.txt  test1.txt
ycao@cs570vmserver:~/PiProject$ cat ./Input/test1.txt
0.2751551298507599‌6 0.02308799505377257 1.0
1.3326417744467765 0.15275693928950207 1.0
1.643875632106871 1.0124949155399974 1.0
0.0988000203465659‌9 1.401413160127707‌8 1.0
0.8618434918312619 1.6540327607672671 1.0
0.1976509820510998‌8 0.5378067016455579 1.0
0.4107104334474207‌5 0.8695059538312928 1.0
1.2443875369663797 1.6422596538904553 1.0
0.8610123578895437 1.843292142947146 1.0
0.2169299131304380‌8 1.037610300293491 1.0
1.3817854837837371 1.5251400729995563 1.0
0.7879689375538406 0.559422438341636 1.0
1.22354301275724‌5 0.13753217067000612 1.0
```

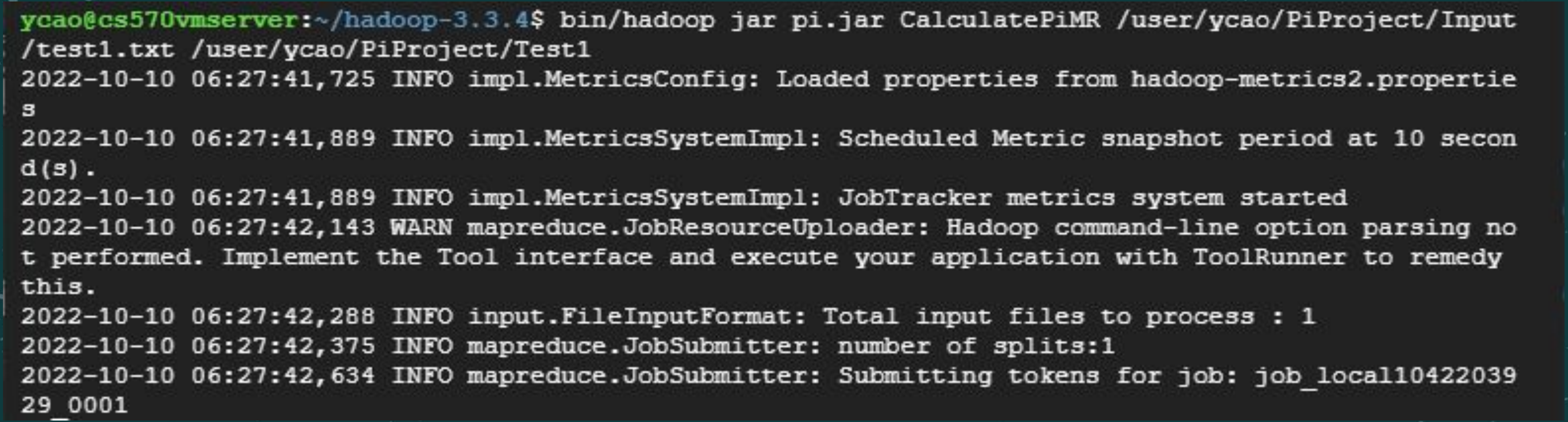
radius = 1

number = 1000

ENHANCED RESULT-- Decrease Radius

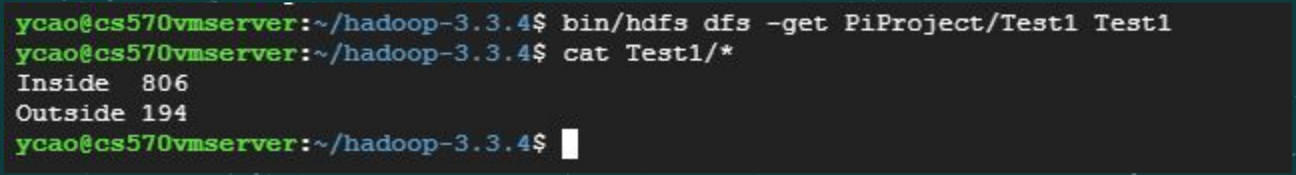


```
ycao@cs570vmserver:~/PiProject$ cd ../hadoop-3.3.4
ycao@cs570vmserver:~/hadoop-3.3.4$ bin/hdfs dfs -put ../PiProject/Input/test1.txt PiProject/Input
ycao@cs570vmserver:~/hadoop-3.3.4$ bin/hdfs dfs -ls PiProject/Input
Found 2 items
-rw-r--r--   1 ycao supergroup      40538 2022-10-10 06:07 PiProject/Input/dots.txt
-rw-r--r--   1 ycao supergroup      42005 2022-10-10 06:25 PiProject/Input/test1.txt
ycao@cs570vmserver:~/hadoop-3.3.4$ 
```

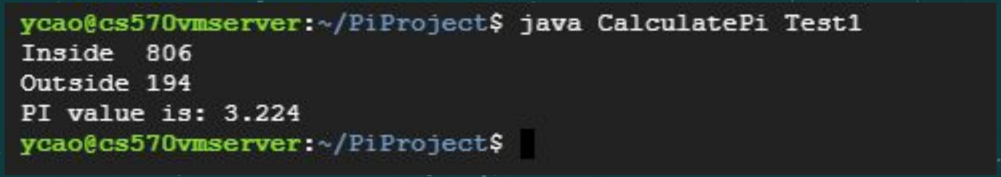


```
ycao@cs570vmserver:~/hadoop-3.3.4$ bin/hadoop jar pi.jar CalculatePiMR /user/ycao/PiProject/Input
/test1.txt /user/ycao/PiProject/Test1
2022-10-10 06:27:41,725 INFO impl.MetricsConfig: Loaded properties from hadoop-metrics2.propertie
s
2022-10-10 06:27:41,889 INFO impl.MetricsSystemImpl: Scheduled Metric snapshot period at 10 secon
d(s).
2022-10-10 06:27:41,889 INFO impl.MetricsSystemImpl: JobTracker metrics system started
2022-10-10 06:27:42,143 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing no
t performed. Implement the Tool interface and execute your application with ToolRunner to remedy
this.
2022-10-10 06:27:42,288 INFO input.FileInputFormat: Total input files to process : 1
2022-10-10 06:27:42,375 INFO mapreduce.JobSubmitter: number of splits:1
2022-10-10 06:27:42,634 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_local10422039
29_0001
```

ENHANCED RESULT-- Decrease Radius



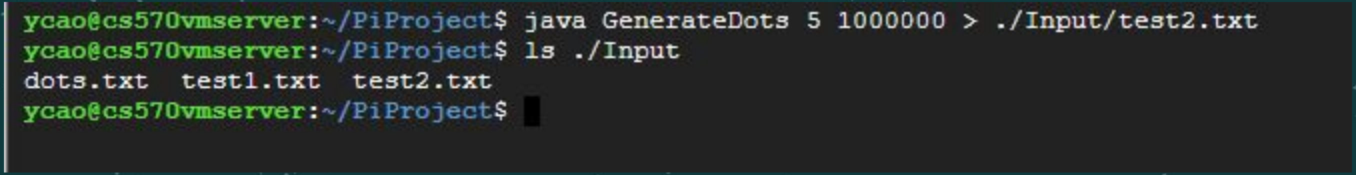
```
ycao@cs570vmserver:~/hadoop-3.3.4$ bin/hdfs dfs -get PiProject/Test1 Test1
ycao@cs570vmserver:~/hadoop-3.3.4$ cat Test1/*
Inside  806
Outside 194
ycao@cs570vmserver:~/hadoop-3.3.4$ 
```



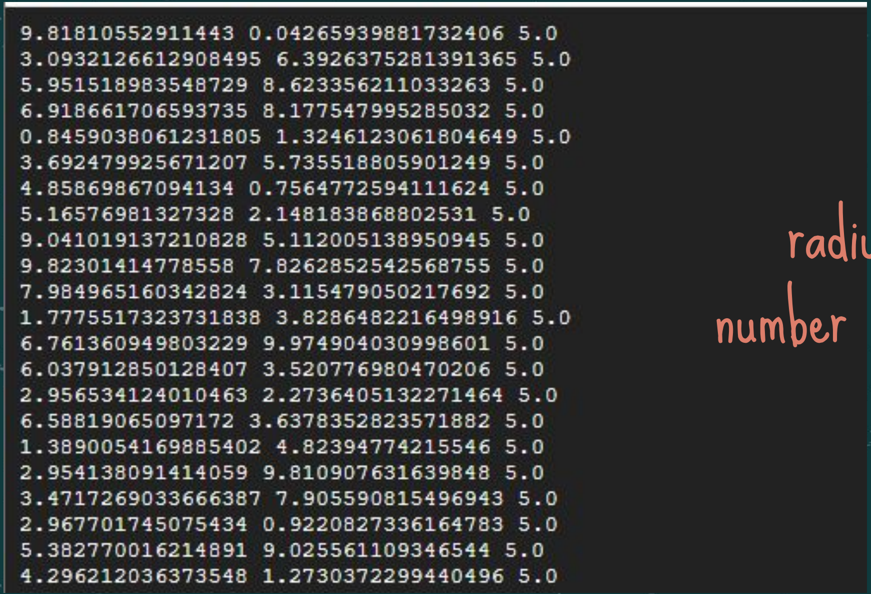
```
ycao@cs570vmserver:~/PiProject$ java CalculatePi Test1
Inside  806
Outside 194
PI value is: 3.224
ycao@cs570vmserver:~/PiProject$ 
```

Pi value calculate is 3.224 which is a better value to the real pi value then the base case value

ENHANCED RESULT -- Increase Number



```
ycao@cs570vmserver:~/PiProject$ java GenerateDots 5 1000000 > ./Input/test2.txt
ycao@cs570vmserver:~/PiProject$ ls ./Input
dots.txt  test1.txt  test2.txt
ycao@cs570vmserver:~/PiProject$ 
```

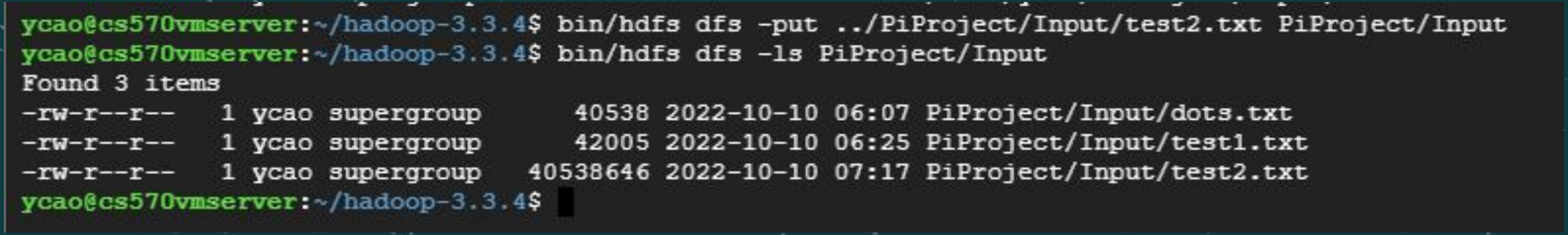


```
9.81810552911443 0.04265939881732406 5.0
3.0932126612908495 6.3926375281391365 5.0
5.951518983548729 8.623356211033263 5.0
6.918661706593735 8.177547995285032 5.0
0.8459038061231805 1.3246123061804649 5.0
3.692479925671207 5.735518805901249 5.0
4.85869867094134 0.7564772594111624 5.0
5.16576981327328 2.148183868802531 5.0
9.041019137210828 5.112005138950945 5.0
9.82301414778558 7.8262852542568755 5.0
7.984965160342824 3.115479050217692 5.0
1.7775517323731838 3.8286482216498916 5.0
6.761360949803229 9.974904030998601 5.0
6.037912850128407 3.520776980470206 5.0
2.956534124010463 2.2736405132271464 5.0
6.58819065097172 3.6378352823571882 5.0
1.3890054169885402 4.82394774215546 5.0
2.954138091414059 9.810907631639848 5.0
3.4717269033666387 7.905590815496943 5.0
2.967701745075434 0.9220827336164783 5.0
5.382770016214891 9.025561109346544 5.0
4.296212036373548 1.2730372299440496 5.0
```

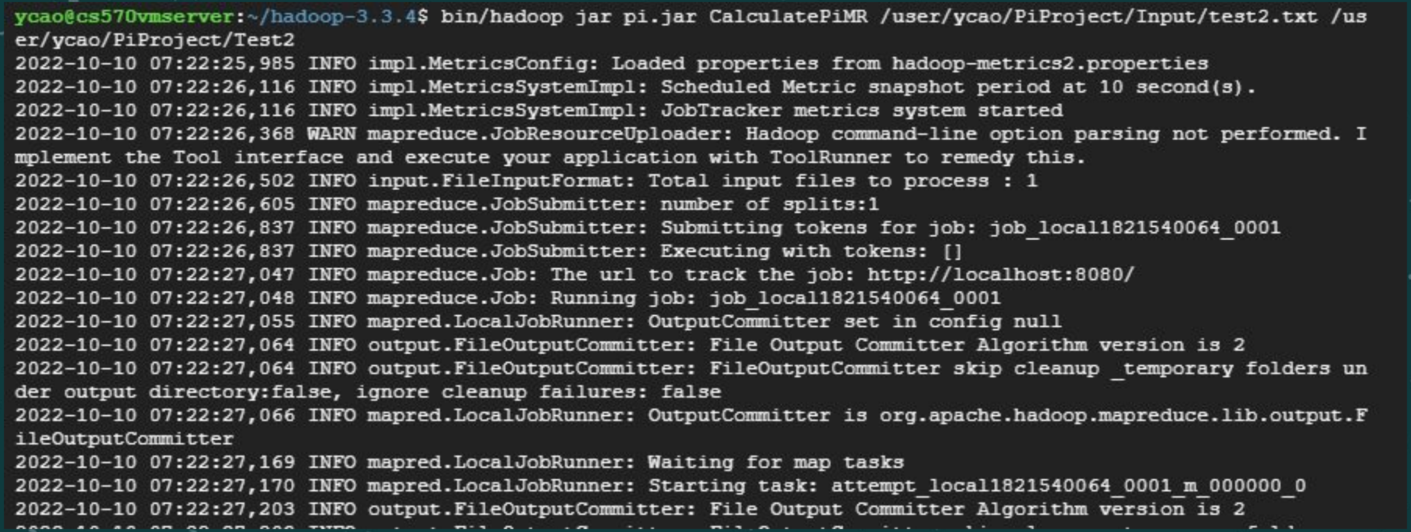
radius = 5

number = 100000

ENHANCED RESULT-- Increase Number

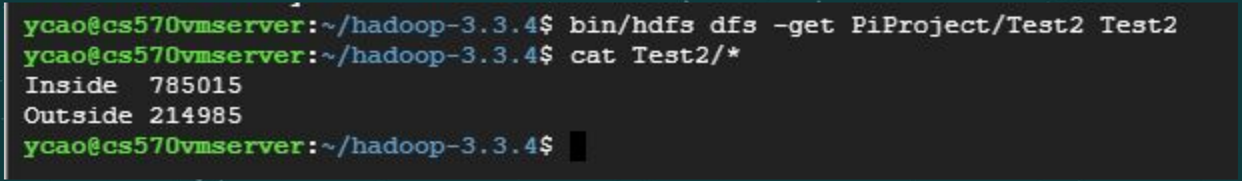


```
ycao@cs570vmserver:~/hadoop-3.3.4$ bin/hdfs dfs -put ../PiProject/Input/test2.txt PiProject/Input
ycao@cs570vmserver:~/hadoop-3.3.4$ bin/hdfs dfs -ls PiProject/Input
Found 3 items
-rw-r--r--   1 ycao supergroup      40538 2022-10-10 06:07 PiProject/Input/dots.txt
-rw-r--r--   1 ycao supergroup      42005 2022-10-10 06:25 PiProject/Input/test1.txt
-rw-r--r--   1 ycao supergroup   40538646 2022-10-10 07:17 PiProject/Input/test2.txt
ycao@cs570vmserver:~/hadoop-3.3.4$ 
```

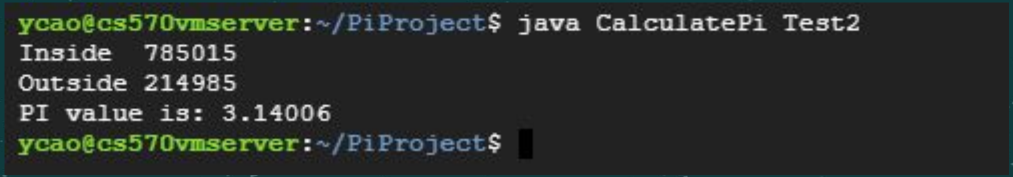


```
ycao@cs570vmserver:~/hadoop-3.3.4$ bin/hadoop jar pi.jar CalculatePiMR /user/ycao/PiProject/Input/test2.txt /us
er/ycao/PiProject/Test2
2022-10-10 07:22:25,985 INFO impl.MetricsConfig: Loaded properties from hadoop-metrics2.properties
2022-10-10 07:22:26,116 INFO impl.MetricsSystemImpl: Scheduled Metric snapshot period at 10 second(s).
2022-10-10 07:22:26,116 INFO impl.MetricsSystemImpl: JobTracker metrics system started
2022-10-10 07:22:26,368 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. I
mplement the Tool interface and execute your application with ToolRunner to remedy this.
2022-10-10 07:22:26,502 INFO input.FileInputFormat: Total input files to process : 1
2022-10-10 07:22:26,605 INFO mapreduce.JobSubmitter: number of splits:1
2022-10-10 07:22:26,837 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_local1821540064_0001
2022-10-10 07:22:26,837 INFO mapreduce.JobSubmitter: Executing with tokens: []
2022-10-10 07:22:27,047 INFO mapreduce.Job: The url to track the job: http://localhost:8080/
2022-10-10 07:22:27,048 INFO mapreduce.Job: Running job: job_local1821540064_0001
2022-10-10 07:22:27,055 INFO mapred.LocalJobRunner: OutputCommitter set in config null
2022-10-10 07:22:27,064 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2022-10-10 07:22:27,064 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders un
der output directory:false, ignore cleanup failures: false
2022-10-10 07:22:27,066 INFO mapred.LocalJobRunner: OutputCommitter is org.apache.hadoop.mapreduce.lib.output.F
ileOutputCommitter
2022-10-10 07:22:27,169 INFO mapred.LocalJobRunner: Waiting for map tasks
2022-10-10 07:22:27,170 INFO mapred.LocalJobRunner: Starting task: attempt_local1821540064_0001_m_000000_0
2022-10-10 07:22:27,203 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
```

ENHANCED RESULT-- Increase Number



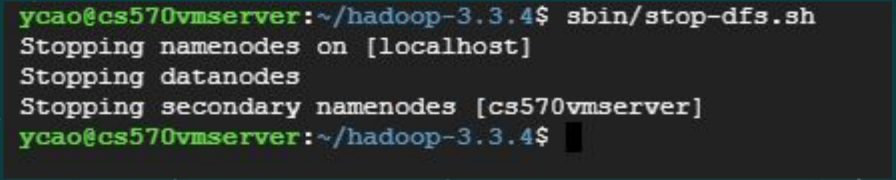
```
ycao@cs570vmserver:~/hadoop-3.3.4$ bin/hdfs dfs -get PiProject/Test2 Test2
ycao@cs570vmserver:~/hadoop-3.3.4$ cat Test2/*
Inside  785015
Outside 214985
ycao@cs570vmserver:~/hadoop-3.3.4$
```



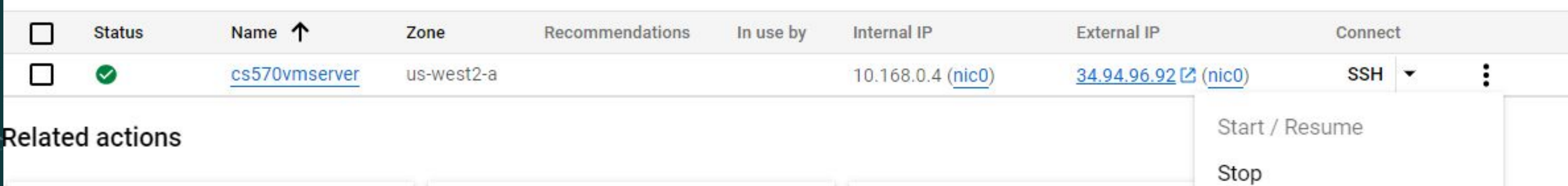
```
ycao@cs570vmserver:~/PiProject$ java CalculatePi Test2
Inside  785015
Outside 214985
PI value is: 3.14006
ycao@cs570vmserver:~/PiProject$
```

Pi value calculate is 3.14006 which is very close to the real pi value

STOP INSTANCE ON GCP



```
ycao@cs570vmserver:~/hadoop-3.3.4$ sbin/stop-dfs.sh
Stopping namenodes on [localhost]
Stopping datanodes
Stopping secondary namenodes [cs570vmserver]
ycao@cs570vmserver:~/hadoop-3.3.4$ 
```



□	Status	Name ↑	Zone	Recommendations	In use by	Internal IP	External IP	Connect
□	✓	cs570vmserver	us-west2-a			10.168.0.4 (nic0)	34.94.96.92 (nic0)	SSH

After done with project, stop namenode and stop the instance on GCP.

CONCLUSION

Summarize for Pi Project

The more random dots generated to cover the area, the more accurate pi value we will get. This is determined by radius and number of dots generated.

MapReduce is good for dealing with large data set using minimal amount of memory and get result fast.

REFERENCES

Chang, H. (2022, 10 09). *Overview of Pi Calculation*. Overview of Pi Calculation. https://hc.labnet.sfbu.edu/~henry/npu/classes//mapreduce/pi/slide/overview.html

Strengths and Weaknesses of MapReduce. (2016, September 11). LinkedIn. Retrieved October 10, 2022, from https://www.linkedin.com/pulse/strengths-weaknesses-mapreduce-muazzam-ali

Taylor, D. (2022, September 17). What is MapReduce in Hadoop? Big Data Architecture. Guru99. Retrieved October 10, 2022, from https://www.guru99.com/introduction-to-mapreduce.html

Value of Pi in Maths - Definition, Forms & Solved Examples. (n.d.). Byju's. Retrieved October 10, 2022, from https://byjus.com/maths/value-of-pi/

THANKS!

Do you have any questions?

CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon**, and infographics & images by **Freepik**