



Pi Project

CS 570 Big Data Processing



Belsabel Woldemichael

Table of Contents

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



Introduction

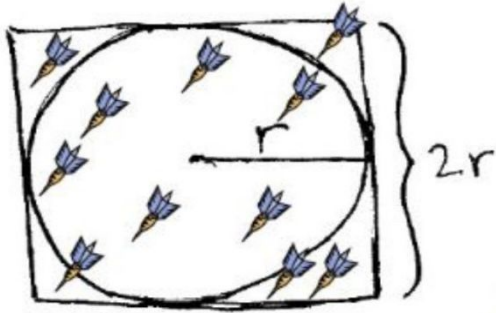
- This project utilizes the Google Cloud Platform (GCP) to set up a distributed computing environment.
- Hadoop, an open-source framework, is employed to handle large data sets and perform parallel processing.
- MapReduce, a programming model within Hadoop, is used to efficiently process and generate large data sets.
- The primary goal is to calculate the value of pi using these technologies.



Theory of Pi Calculation



- 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 $x^2 + y^2 < r^2$

- Take the total number of darts that landed in the circle as S

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

Formula:

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

The value of pi can be determined by counting the number of random darts that land inside the circle compared to those that land outside the circle.



Design



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							

Processes

1. 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. Write MapReduce program in Java Language to count number of points inside and outside of the circle with given radius.
3. Using the input file generated in step 1 to run MapReduce program in Step 2 Output should be like: Inside xxx Outside xxx .
4. Write a Java Program to calculate pi value - Using the output from Step 3 get pi value.



Implementation

Step 1 – Setting up an Ubuntu VM on Google Cloud



VM instances

Filter

Enter property name or value

?

⋮

<input type="checkbox"/>	Status	Name ↑	Zone	Recommendations	In use by	Internal IP	External IP	Connect
<input type="checkbox"/>	✓	cs-570	us-west2-a			10.168.0.3 (nic0)	34.94.210.89 ↗ (nic0)	SSH ▾ ⋮



Step 2. Hadoop: Setting up a Single Node Cluster

```
belsabelteklemariam@cs-570:~/hadoop-3.4.0$ bin/hdfs namenode -format
WARNING: /home/belsabelteklemariam/hadoop-3.4.0/logs does not exist. Creating.
2024-06-05 10:42:06,241 INFO namenode.NameNode: STARTUP MSG:
/*****
STARTUP MSG: Starting NameNode
STARTUP MSG:   host = cs-570/10.168.0.3
STARTUP MSG:   args = [-format]
STARTUP MSG:   version = 3.4.0
STARTUP MSG:   classpath = /home/belsabelteklemariam/hadoop-3.4.0/etc/hadoop:/home/belsabelteklemariam/hadoop-3.4.0/share/hadoop/common/lib/kerb-client-2.0.3.jar:/home/belsabelteklemariam/hadoop-3.4.0/share/hadoop/common/lib/curator-client-5.2.0.jar:/home/belsabelteklemariam/hadoop-3.4.0/share/hadoop/common/lib/netty-resolver-dns-native-macos-4.1.100.Final-osx-aarch_64.jar:/home/belsabelteklemariam/hadoop-3.4.0/share/hadoop/common/lib/netty-codec-http-4.1.100.Final.jar:/home/belsabelteklemariam/hadoop-3.4.0/share/hadoop/common/lib/netty-codec-4.1.100.Final.jar:/home/belsabelteklemariam/hadoop-3.4.0/share/hadoop/common/lib/kerb-util-2.0.3.jar:/home/belsabelteklemariam/hadoop-3.4.0/share/hadoop/common/lib/jetty-security-9.4.53.v20231009.jar:/home/belsabelteklemariam/hadoop-3.4.0/share/hadoop/common/lib/jakarta.activation-api-1.2.1.jar:/home/belsabelteklemariam/hadoop-3.4.0/share/hadoop/common/lib/netty-codec-stomp-4.1.100.Final.jar:/home/belsabelteklemariam/hadoop-3.4.0/share/hadoop/common/lib/jetty-server-9.4.53.v20231009.jar:/home/belsabelteklemariam/hadoop-3.4.0/share/hadoop/common/lib/kerby-xdr-2.0.3.jar:/home/belsabelteklemariam/hadoop-3.4.0/share/hadoop/common/lib/netty-transport-native-unix-common-4.1.100.Final.jar:/home/belsabelteklemariam/hadoop-3.4.0/share/hadoop/common/lib/netty-transport-classes-epoll-4.1.100.Final.jar:/home/belsabelteklemariam/hadoop-3.4.0/share/hadoop/common/lib/netty-handler-proxy-4.1.100.Final.jar:/home/belsabelteklemariam/hadoop-3.4.0/share/hadoop/common/lib/netty-handler-ssl-ocsp-4.1.100.Final.jar:/home/belsabelteklemariam/hadoop-3.4.0/share/hadoop/common/lib/netty-handler-ssl-ocsp-4.1.100.Final.jar:/home/belsabelteklemariam/hadoop-3.4.0/share/hadoop/common/lib/netty-codec-socks-4.1.100.Final.jar:/home/belsabelteklemariam/hadoop-3.4.0/share/hadoop/common/lib/jetty-http-9.4.53.v20231009.jar:/home/belsabelteklemariam/hadoop-3.4.0/share/hadoop/common/lib/netty-resolver-4.1.100.Final.jar:/home/belsabelteklemariam/hadoop-3.4.0/share/hadoop/common/lib/hadoop-shaded-guava-1.2.0.jar:/home/belsabelteklemariam/hadoop-3.4.0/share/hadoop/common/lib/netty-transport-native-epoll-4.1.100.Final.jar:/home/belsabelteklemariam/hadoop-3.4.0/share/hadoop/common/lib/jetty-util-9.4.53.v20231009.jar:/home/belsabelteklemariam/hadoop-3.4.0/share/hadoop/common/lib/jackson-annotations-2.12.7.jar:/home/belsabelteklemariam/hadoop-3.4.0/share/hadoop/common/lib/nimbus-jose-jwt-9.31.jar:/home/belsabelteklemariam/hadoop-3.4.0/share/hadoop/common/lib/kerby-pkix-2.0.3.jar:/home/belsabelteklemariam/hadoop-3.4.0/share/hadoop/common/lib/gson-2.9.0.jar:/home/belsabelteklemariam/hadoop-3.4.0/share/hadoop/common/lib/commons-lang3-3.12.0.jar:/home/belsabelteklemariam/hadoop-3.4.0/share/hadoop/common/lib/slf4j-api-1.7.36.jar:/home/belsabelteklemariam/hadoop-3.4.0/share/hadoop/common/lib/commons-io-2.14.0.jar:/home/belsabelteklemariam/hadoop-3.4.0/share/hadoop/common/lib/jetty-util-ajax-9.4.53.v20231009.jar:/home/belsabelteklemariam/hadoop-3.4.0/share/hadoop/common/lib/commons-compress-1.24.0.jar:/home/belsabelteklemariam/hadoop-3.4.0/share/hadoop/common/lib/animal-sniffer-annotations-1.17.jar:/home/belsabelteklemariam/hadoop-3.4.0/share/hadoop/common/lib/jersey-server-1.19.4.jar:/home/belsabelteklemariam/hadoop-3.4.0/share/hadoop/common/lib/re2j-1.1.jar:/home/belsabelteklemariam/hadoop-3.4.0/share/hadoop/common/lib/jackson-core-2.12.7.jar:/home/belsabelteklemariam/hadoop-3.4.0/share/hadoop/common/lib/jsr305-3.0.2.jar:/home/belsabelteklemariam/hadoop-3.4.0/share/hadoop/common/lib/netty-codec-mqtt-4.1.100.Final.jar:/home/belsabelteklemariam/hadoop-3.4.0/share/hadoop/common/lib/guava-27.0-jre.jar:/home/belsabelteklemariam/hadoop-3.4.0/share/hadoop/common/lib/kerb-core-2.0.3.jar:/home/belsabelteklemariam/hadoop-3.4.0/share/hadoop/common/lib/checker-qual-2.5.2.jar:/home/belsabelteklemariam/hadoop-3.4.0/share/hadoop/common/lib/avro-1.9.2.jar:/home/belsabelteklemariam/hadoop-3.4.0/share/hadoop/common/lib/commons-codec-1.15.jar:/home/belsabelteklemariam/hadoop-3.4.0/share/hadoop/common/lib/j2objc-annotations-1.1.jar:/home/belsabelteklemariam/hadoop-3.4.0/share/hadoop/common/lib/commons-configuration2-2.8.0.jar:/home/belsabelteklemariam/hadoop-3.4.0/share/hadoop/common/lib/netty-transport-socks-4.1.100.Final.jar:/home/belsabelteklemariam/hadoop-3.4.0/share/hadoop/common/lib/bcprov-jdk15on-1.70.jar:/home/belsabelteklemariam/hadoop-3.4.0/share/hadoop/common/lib/netty-transport-udt-4.1.100.Final.jar:/home/belsabelteklemariam/hadoop-3.4.0/share/hadoop/common/lib/jsch-0.1.55.jar:/home/belsabelteklemariam/hadoop-3.4.0/share/hadoop/common/lib/httpclient-4.5.13.jar:/home/belsabelteklemariam/hadoop-3.4.0/share/hadoop/common/lib/jcip-annotations-1.0-1.jar:/home/belsabelteklemariam/hadoop-3.4.0/share/hadoop/common/lib/jsr311-api-1.1.1.jar:/home/belsabelteklemariam/hadoop-3.4.0/share/hadoop/common/lib/jul-to-slf4j-1.7.36.jar:/home/belsabelteklemariam/hadoop-3.4.0/share/hadoop/common/lib/kerby-config-2.0.3.jar:/home/belsabelteklemariam/hadoop-3.4.0/share/hadoop/common/lib/netty-tcnative-2.0.40.Final.jar:/home/belsabelteklemariam/hadoop-3.4.0/share/hadoop/common/lib/netty-tcnative-2.0.40.Final.jar
```


Start NameNode daemon and DataNode

```
belsabelteklemaria@cs-570:~/hadoop-3.4.0$ nano etc/hadoop/hadoop-env.sh
belsabelteklemaria@cs-570:~/hadoop-3.4.0$ sbin/start-dfs.sh
Starting namenodes on [localhost]
Starting datanodes
Starting secondary namenodes [cs-570]
belsabelteklemaria@cs-570:~/hadoop-3.4.0$ wget http://localhost:9870/
--2024-06-05 10:51:47-- 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]
--2024-06-05 10:51:47-- 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

2024-06-05 10:51:47 (113 MB/s) - 'index.html' saved [1079/1079]
```



Step 3: PI_on_Mapreduce



```
belabelteklemariam@cs570:~$ ls
PiProject  hadoop-3.4.0  hadoop-3.4.0.tar.gz
belabelteklemariam@cs570:~$ cd PiProject
belabelteklemariam@cs570:~/PiProject$ mkdir input
belabelteklemariam@cs570:~/PiProject$ ls
CalculatePi.java  CalculatePiMR.java  GenerateDots.java  input
```



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.toString(radius));
        }
    }
}
```



CODE--CalculatePiMR.java



```
import java.io.IOException; import java.util.*;
import java.lang.Object;
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.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;

public class CalculatePiMR {
    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, InterruptedException
        {
            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);
            }
        }
    }
}
```

CODE--CalculatePi.java

```
import java.io.*;
public class CalculatePi {
    public static void main(String[] args) throws Exception{
        String file = "../hadoop-3.4.0/"+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();
        }
    }
}
```



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



Output save in ./Input/dots.txt

\$ java GenerateDots 5 1000 > ./Input/dots.tx

```
belsabelteklemariam@cs-570:~/PiProject$ java GenerateDots 5 1000 > ./input/dots.txt
belsabelteklemariam@cs-570:~/PiProject$ cat ./input/dots.txt
1.4406666166691517 3.008770756986985 5.0
5.408335796652937 2.9815032546113898 5.0
3.997345429830168 9.946000528287355 5.0
9.295169620361465 9.144400679853685 5.0
0.29292010391042345 7.536568894697078 5.0
5.636436915214842 9.657083087647129 5.0
3.412193157841461 3.4387923740145663 5.0
1.5352840507035648 7.774427946620649 5.0
4.73645982262977 8.725268713969056 5.0
5.635535862527447 3.1061206941878816 5.0
8.378991002057122 7.821445977817709 5.0
8.780879833043716 7.722201445373078 5.0
6.84027626184616 9.101890509383297 5.0
```



Copy file from local to hadoop and check

```
$ bin/hdfs dfs -mkdir /user/belsabelteklemariam/PiProject/Input
```

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

```
$ bin/hdfs dfs -ls PiProject/Input
```

```
belsabelteklemariam@cs-570:~/hadoop-3.4.0$ bin/hdfs dfs -mkdir -p /user/belsabelteklemariam/PiProject/input
belsabelteklemariam@cs-570:~/hadoop-3.4.0$ bin/hdfs dfs -put ../PiProject/input/* PiProject/input
belsabelteklemariam@cs-570:~/hadoop-3.4.0$ bin/hdfs dfs -ls PiProject/input
Found 1 items
-rw-r--r--    1 belsabelteklemariam supergroup    40557 2024-06-05 11:34 PiProject/input/dots.txt
```



Compile Mapreduce program in Hadoop with *.class files created



```
$bin/hadoop jar  
~/hadoop-3.4.0/share/hadoop/mapreduce/hadoop-mapreduce-client-core-3.4.0.jar  
com.sun.tools.javac.Main ~/PiProject/CalculatePiMR.java
```

```
belsabelteklemarian@cs-570:~/hadoop-3.4.0$ bin/hadoop jar ~/hadoop-3.4.0/share/hadoop/mapreduce/hadoop-mapreduce-client-core-3.4.0.jar com.sun.tools.javac.Main ~/PiProject/CalculatePiMR.java  
Note: /home/belsabelteklemarian/PiProject/CalculatePiMR.java uses or overrides a deprecated API.  
Note: Recompile with -Xlint:deprecation for details.
```

Create .jar file with *.class files


```
$Jar cf pi.jar CalculatePiMR*.class
```

```
belsabelteklemarian@cs-570:~/PiProject$ ls  
CalculatePi.java 'CalculatePiMR$Map.class' 'CalculatePiMR$Reduce.class' CalculatePiMR.class CalculatePiMR.java GenerateDots.class GenerateDots.java input testing  
belsabelteklemarian@cs-570:~/PiProject$ jar cf pi.jar CalculatePiMR*.class  
belsabelteklemarian@cs-570:~/PiProject$ cd ..
```



Run MapReduce Program with input file and save result in Output

\$ bin/hadoop jar ~/PiProject/pi.jar CalculatePiMR /user/belsabelteklemariam/PiProject/input
/user/belsabelteklemariam/PiProject/Output



```
belsabelteklemariam@cs-570:~/hadoop-3.4.0$ bin/hadoop jar ~/PiProject/pi.jar CalculatePiMR /user/belsabelteklemariam/PiProject/input /user/belsabelteklemariam/PiProject/Output
2024-06-05 11:55:15,705 INFO impl.MetricsConfig: Loaded properties from hadoop-metrics2.properties
2024-06-05 11:55:15,880 INFO impl.MetricsSystemImpl: Scheduled Metric snapshot period at 10 second(s).
2024-06-05 11:55:15,881 INFO impl.MetricsSystemImpl: JobTracker metrics system started
2024-06-05 11:55:16,249 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
2024-06-05 11:55:16,621 INFO input.FileInputFormat: Total input files to process : 1
2024-06-05 11:55:16,666 INFO mapreduce.JobSubmitter: number of splits:1
2024-06-05 11:55:17,036 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_local1814531961_0001
2024-06-05 11:55:17,037 INFO mapreduce.JobSubmitter: Executing with tokens: []
2024-06-05 11:55:17,351 INFO mapreduce.Job: The url to track the job: http://localhost:8080/
2024-06-05 11:55:17,351 INFO mapreduce.Job: Running job: job_local1814531961_0001
2024-06-05 11:55:17,360 INFO mapred.LocalJobRunner: OutputCommitter set in config null
2024-06-05 11:55:17,370 INFO output.PathOutputCommitterFactory: No output committer factory defined, defaulting to FileOutputCommitterFactory
2024-06-05 11:55:17,375 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2024-06-05 11:55:17,375 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup_temporary folders under output directory:false, ignore cleanup failures: false
```



Get output and save to local, the show output

```
$bin/hdfs dfs -get PiProject/Output Output
```

```
Cat Output/*
```

```
belsabelteklemariam@cs-570:~$ cd hadoop-3.4.0
```

```
belsabelteklemariam@cs-570:~/hadoop-3.4.0$ cat Output/*
```

```
Inside 781
```

```
Outside 219
```



Result

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



```
belsabelteklemariam@cs-570:~/PiProject$ java CalculatePi Output
Inside 781
Outside 219
PI value is: 3.124
```

The pi value calculated is 3.124, and it is quite off from 3.1415926




Enhancement

Decrease Radius to get better result

Radius = 1 and number = 1000

```
belsabelteklemariam@cs-570:~/PiProject$ javac GenerateDots.java
belsabelteklemariam@cs-570:~/PiProject$ java GenerateDots 1 1000 > ./input/test1.txt
belsabelteklemariam@cs-570:~/PiProject$ ls ./input
dots.txt  test1.txt
belsabelteklemariam@cs-570:~/PiProject$ cat ./input/test1.txt
1.8058669369195384 0.3490079194216753 1.0
1.7502526423492648 1.7374129161575977 1.0
1.3217944153431407 0.5399359750318693 1.0
1.708214442237073 0.17051646569541457 1.0
0.7039970640417672 0.9806576932362423 1.0
0.8444387559538782 1.4838562777081679 1.0
0.46033178798283814 1.7547635472229284 1.0
0.579020331068921 1.0995519871214199 1.0
1.7754944121527043 1.2643792687603244 1.0
0.2405854132503551 0.9064907521706083 1.0
1.9685267563849158 0.8618356169445511 1.0
1.8407248667789486 1.4997932610301083 1.0
```





```
belsabelteklemariam@cs-570:~$ cd hadoop-3.4.0/
belsabelteklemariam@cs-570:~/hadoop-3.4.0$ bin/hdfs dfs -put ../PiProject/input/test1.txt PiProject/input
belsabelteklemariam@cs-570:~/hadoop-3.4.0$ bin/hdfs dfs -ls PiProject/input
```

Found 2 items

```
-rw-r--r-- 1 belsabelteklemariam supergroup 40557 2024-06-05 11:34 PiProject/input/dots.txt
-rw-r--r-- 1 belsabelteklemariam supergroup 42052 2024-06-05 12:21 PiProject/input/test1.txt
```

```
belsabelteklemariam@cs-570:~/hadoop-3.4.0$ bin/hadoop jar ~/PiProject/pi.jar CalculatePIMR /user/belsabelteklemariam/PiProject/input/test1.txt /user/belsabelteklemariam/PiProject/est1
2024-06-05 12:29:01,685 INFO impl.MetricsConfig: Loaded properties from hadoop-metrics2.properties
2024-06-05 12:29:01,850 INFO impl.MetricsSystemImpl: Scheduled Metric snapshot period at 10 second(s).
2024-06-05 12:29:01,850 INFO impl.MetricsSystemImpl: JobTracker metrics system started
2024-06-05 12:29:02,143 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
2024-06-05 12:29:02,335 INFO input.FileInputFormat: Total input files to process : 1
2024-06-05 12:29:02,450 INFO mapreduce.JobSubmitter: number of splits:1
2024-06-05 12:29:02,723 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_local1719683798_0001
2024-06-05 12:29:02,723 INFO mapreduce.JobSubmitter: Executing with tokens: []
2024-06-05 12:29:02,969 INFO mapreduce.Job: The url to track the job: http://localhost:8080/
2024-06-05 12:29:02,970 INFO mapreduce.Job: Running job: job_local1719683798_0001
2024-06-05 12:29:02,977 INFO mapred.LocalJobRunner: OutputCommitter set in config null
2024-06-05 12:29:02,988 INFO output.PathOutputCommitterFactory: No output committer factory defined, defaulting to FileOutputCommitterFactory
2024-06-05 12:29:02,990 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2024-06-05 12:29:02,990 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup_temporary folders under output directory:false, ignore cleanup failures: false
2024-06-05 12:29:02,992 INFO mapred.LocalJobRunner: OutputCommitter is org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter
2024-06-05 12:29:03,076 INFO mapred.LocalJobRunner: Waiting for map tasks
2024-06-05 12:29:03,077 INFO mapred.LocalJobRunner: Starting task: attempt_local1719683798_0001_m_000000_0
2024-06-05 12:29:03,117 INFO output.PathOutputCommitterFactory: No output committer factory defined, defaulting to FileOutputCommitterFactory
2024-06-05 12:29:03,122 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2024-06-05 12:29:03,122 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup_temporary folders under output directory:false, ignore cleanup failures: false
```

```
belsabelteklemariam@cs-570:~/hadoop-3.4.0$ bin/hdfs dfs -get PiProject/Test1 Test1
belsabelteklemariam@cs-570:~/hadoop-3.4.0$ cat Test1/*
Inside 775
Outside 225
```

```
belsabelteklemariam@cs-570:~/PiProject$ java CalculatePi Test1
Inside 775
Outside 225
PI value is: 3.1
belsabelteklemariam@cs-570:~/PiProject$
```

Pi value calculated is 3.1 which is a better value to the real pi value

Increase number to get better result


Radius = 5 and number = 1000,000

```


Pi value is: 3.14159
aghebre423@mapreduce:~/PiProject$ java GenerateDots 5 1000000 > ./input/test2.txt
aghebre423@mapreduce:~/PiProject$ ls ./input/test2.txt
./input/test2.txt
aghebre423@mapreduce:~/PiProject$ ls ./input
dots.txt  test1.txt  test2.txt
aghebre423@mapreduce:~/PiProject$ cat ./input/test2.txt
2.4409513178371336 1.9695968916104478 5.0
6.039596943158905 1.946277459843908 5.0
7.34317341682304 9.64860808775004 5.0
2.6616950654632565 2.589232923294439 5.0
3.495537161083142 8.291024720380582 5.0
6.371800950987319 0.4486674244486122 5.0
9.300473331723488 7.773773117188401 5.0
8.291425720800357 0.9219277488584798 5.0
5.642389490486829 0.0012242655171057493 5.0
6.335145390203549 6.418908354091643 5.0
8.934590875828349 5.823586718904402 5.0
4.601557809759406 0.1391976117294913 5.0
9.090069650570543 4.063996612243868 5.0
2.4441202544231686 6.7298038781988 5.0
4.235313396113073 0.6966038934684193 5.0
5.94434163930047 5.214390014240774 5.0
8.996055945545837 7.243340433171827 5.0
7.1758141315547705 4.516588019987867 5.0
6.00614873233384 8.794951079325372 5.0
2.1260939728793837 2.1441056545946022 5.0
0.0255070025641506 2.7122558459730803 5.0

```





```
belsabelteklemariam@cs-570:~/PiProject$ ls ./input/test2.txt
```



```
belsabelteklemariam@cs-570:~/PiProject$ cat ./input/test2.txt
```

```
belsabelteklemariam@cs-570:~$ cd hadoop-3.4.0
belsabelteklemariam@cs-570:~/hadoop-3.4.0$ bin/hdfs dfs -put ../PiProject/input/test2.txt PiProject/input
belsabelteklemariam@cs-570:~/hadoop-3.4.0$ bin/hdfs dfs -ls PiProject/input
Found 3 items
-rw-r--r-- 1 belsabelteklemariam supergroup 40557 2024-06-05 11:34 PiProject/input/dots.txt
-rw-r--r-- 1 belsabelteklemariam supergroup 42052 2024-06-05 12:21 PiProject/input/test1.txt
-rw-r--r-- 1 belsabelteklemariam supergroup 40539287 2024-06-05 13:11 PiProject/input/test2.txt
```

```
belsabeltekleklemariam@cs-570:~/hadoop-3.4.0$ bin/hadoop jar ~/PiProject/pi.jar CalculatePiMR /user/belsabeltekleklemariam/PiProject/input/test2.txt /user/belsabeltekleklemariam/PiProject/est2
2024-06-05 13:14:15,623 INFO impl.MetricsConfig: Loaded properties from hadoop-metrics2.properties
2024-06-05 13:14:15,844 INFO impl.MetricsSystemImpl: Scheduled Metric snapshot period at 10 second(s).
2024-06-05 13:14:15,845 INFO impl.MetricsSystemImpl: JobTracker metrics system started
2024-06-05 13:14:16,065 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
2024-06-05 13:14:16,302 INFO input.FileInputFormat: Total input files to process : 1
2024-06-05 13:14:16,417 INFO mapreduce.JobSubmitter: number of splits:1
2024-06-05 13:14:16,696 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_local284321300_0001
2024-06-05 13:14:16,697 INFO mapreduce.JobSubmitter: Executing with tokens: []
2024-06-05 13:14:16,961 INFO mapreduce.Job: The url to track the job: http://localhost:8080/
2024-06-05 13:14:16,962 INFO mapreduce.Job: Running job: job_local284321300_0001
2024-06-05 13:14:16,970 INFO mapred.LocalJobRunner: OutputCommitter set in config null
2024-06-05 13:14:16,981 INFO output.PathOutputCommitterFactory: No output committer factory defined, defaulting to FileOutputCommitterFactory
2024-06-05 13:14:16,983 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2024-06-05 13:14:16,983 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
2024-06-05 13:14:16,985 INFO mapred.LocalJobRunner: OutputCommitter is org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter
2024-06-05 13:14:17,070 INFO mapred.LocalJobRunner: Waiting for map tasks
2024-06-05 13:14:17,071 INFO mapred.LocalJobRunner: Starting task: attempt_local284321300_0001_m_000000_0
2024-06-05 13:14:17,121 INFO output.PathOutputCommitterFactory: No output committer factory defined, defaulting to FileOutputCommitterFactory
2024-06-05 13:14:17,126 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2024-06-05 13:14:17,126 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
2024-06-05 13:14:17,166 INFO mapred.Task: Using ResourceCalculatorProcessTree : [ ]
```

Bytes Written=29

```
belsabelteklemariam@cs-570:~/hadoop-3.4.0$ bin/hdfs dfs -get PiProject/Test2 Test2
```

```
belsabelteklemariam@cs-570:~/hadoop-3.4.0$ cat Test2/*
```

```
Inside 785611
```

```
Outside 214389
```

```
belsabelteklemariam@cs-570:~/hadoop-3.4.0$ cd
```

```
belsabelteklemariam@cs-570:~$ cd PiProject
```

```
belsabelteklemariam@cs-570:~/PiProject$ java CalculatePi Test2
```

```
Inside 785611
```


```
Outside 214389
```

```
PI value is: 3.142444
```

```
belsabelteklemariam@cs-570:~/PiProject$
```

Pi value calculate is 3.142444 which is very close to the real pi value

Stop Instance on GCP

 Filter Enter property name or value

☐

Status

☐

☒

Name

↑

Zone

Recommendations

In use by

Internal IP

External IP

Connect

<input type="checkbox"/>	<input checked="" type="checkbox"/>	cs-570	us-west2-a		10.168.0.3 (nic0)	35.235.87.130 ↗ (nic0)	SSH ▾	⋮
--------------------------	-------------------------------------	------------------------	------------	--	-----------------------------------	--	-------	---

Related actions

Start / Resume

Stop

Suspend

IDE



Conclusion

- Increasing the quantity of random dots significantly enhances the precision of the pi estimation. This improvement is directly influenced by both the radius of the circle and the total number of dots sampled.
- The MapReduce framework is exceptionally proficient at handling extensive datasets with speed and efficiency, all while requiring minimal memory resources. This makes it particularly well-suited for large-scale data processing tasks.



Reference

Overview of Pi calculation using Map Reduce

How to calculate Pi

Value of Pi



Github link

https://github.com/BelsabelTekle/Cloud_Computing/tree/main/MapReduce

