



Calculating Pi using MapReduce

By
Shruti Kavishwar
San Francisco Bay University

Guided By: Prof. Henry Chang



Agenda

- Introduction to Pi
- Design
- Implementation
- Test
- Enhancement
- Conclusion
- References

Process

01

Prepare Input File

Write a Java program to generate numbers of random pairs of point(x,y) with given radius

02

Code for MapReduce

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

03

Execute Mapreduce on GCP

Using the input file to run MapReduce program

04

Calculate Pi

Write Java Program to calculate pi value using the output from step 3

Setup

- Create a Ubuntu VM instance on Google Cloud Platform

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"/>	mapreduce-week2-hw1	us-central1-c			10.128.0.2 (nic0)		SSH ▾ ⋮

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"/>	mapre				10.128.0.2 (nic0)		SSH ▾ ⋮

Related actions

You'll be charged for running this VM according to its configuration.

Start mapreduce-week2-hw1?

[CANCEL](#) [START](#)

Monitor VMs
View outlier VMs across metrics like CPU and network

[HIDE](#)

Setup

- Connect VM through SSH
- Connect to the localhost after the instance is up and running..

```
skavishw276@mapreduce-week2-hw1:~$ ssh localhost
Welcome to Ubuntu 20.04.6 LTS (GNU/Linux 5.15.0-1060-gcp x86_64)

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

System information as of Tue Jun  4 22:26:27 UTC 2024

System load:  0.06               Processes:           111
Usage of /:   54.3% of 9.51GB    Users logged in:    1
Memory usage: 22%               IPv4 address for ens4: 10.128.0.2
Swap usage:   0%

 * Strictly confined Kubernetes makes edge and IoT secure. Learn how MicroK8s
   just raised the bar for easy, resilient and secure K8s cluster deployment.

   https://ubuntu.com/engage/secure-kubernetes-at-the-edge

Expanded Security Maintenance for Applications is not enabled.

3 updates can be applied immediately.
3 of these updates are standard security updates.
To see these additional updates run: apt list --upgradable

Enable ESM Apps to receive additional future security updates.
See https://ubuntu.com/esm or run: sudo pro status

New release '22.04.3 LTS' available.
Run 'do-release-upgrade' to upgrade to it.

Last login: Tue Jun  4 22:24:42 2024 from 35.235.244.34
skavishw276@mapreduce-week2-hw1:~$
```

Setup

- Code to generate random dot pairs with command line argument taken in as radius and number of pairs. Output will be x y radius

```
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));
        }
    }
}
```

Setup

- Map() for MapReduce

```
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);
        }
    }
}
```

Setup

- Reduce() for MapReduce

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


Setup

- main() for MapReduce

```
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);
}
```

Setup

- Java code to calculate pi value with MapReduce result taken in by reading the file

```
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();
        }
    }
}
```

Code Structure

Pi Directory and content of the Pi directory with the input file and code files created

```
skavishw276@mapreduce-week2-hw1:~$ ls
Pi WordCount hadoop-3.4.0 hadoop-3.4.0.tar.gz
skavishw276@mapreduce-week2-hw1:~$
```

```
skavishw276@mapreduce-week2-hw1:~$ cd Pi
skavishw276@mapreduce-week2-hw1:~/Pi$ ls
CalculatePi.java CalculatePiMR.java GenerateDots.java input
skavishw276@mapreduce-week2-hw1:~/Pi$
```

Test

Start the Cluster and start the namenode and datanode services

```
skavishw276@mapreduce-week2-hw1:~/hadoop-3.4.0$ bin/hdfs namenode -format
2024-06-04 22:53:06,261 INFO namenode.NameNode: STARTUP_MSG:
/*****
STARTUP_MSG: Starting NameNode
STARTUP_MSG:   host = mapreduce-week2-hw1.us-central1-c.c.mapreduce-week2-hw1
-cs570.internal/10.128.0.2
STARTUP_MSG:   args = [-format]
STARTUP_MSG:   version = 3.4.0
STARTUP_MSG:   classpath = /home/skavishw276/hadoop-3.4.0/etc/hadoop:/home/sk
avishw276/hadoop-3.4.0/share/hadoop/common/lib/kerb-client-2.0.3.jar:/home/sk
```

```
skavishw276@mapreduce-week2-hw1:~/hadoop-3.4.0$ sbin/start-dfs.sh
Starting namenodes on [localhost]
localhost: skavishw276@localhost: Permission denied (publickey).
Starting datanodes
localhost: skavishw276@localhost: Permission denied (publickey).
Starting secondary namenodes [mapreduce-week2-hw1]
mapreduce-week2-hw1: skavishw276@mapreduce-week2-hw1: Permission denied (publ
ickey).
skavishw276@mapreduce-week2-hw1:~/hadoop-3.4.0$
```

Test

Permission denied error ssh to localhost again

```
skavishw276@mapreduce-week2-hw1:~/hadoop-3.4.0$ ssh localhost
skavishw276@localhost: Permission denied (publickey).
skavishw276@mapreduce-week2-hw1:~/hadoop-3.4.0$
```

```
skavishw276@mapreduce-week2-hw1:~/hadoop-3.4.0$ ssh-keygen -t rsa -P '' -f ~/.ssh/id_rsa
Generating public/private rsa key pair.
/home/skavishw276/.ssh/id_rsa already exists.
Overwrite (y/n)? y
Your identification has been saved in /home/skavishw276/.ssh/id_rsa
Your public key has been saved in /home/skavishw276/.ssh/id_rsa.pub
The key fingerprint is:
SHA256: iTU7rnYOCr9cNT0C+B6GrGP9Kes38ad60rtj2XKJL8s skavishw276@mapreduce-week2-hw1
The key's randomart image is:
+---[RSA 3072]-----+
|
|             o
|      .    o +
| . .    . S
| .o . . + o
|..* . O = .
|o+ B B+#. =
|o.ooX*BE&.
+----[SHA256]-----+
skavishw276@mapreduce-week2-hw1:~/hadoop-3.4.0$
```

```
skavishw276@mapreduce-week2-hw1:~/hadoop-3.4.0$ cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys
skavishw276@mapreduce-week2-hw1:~/hadoop-3.4.0$ chmod 0600 ~/.ssh/authorized_keys
```

Test

- Connect to the localhost it should work now

```
skavishw276@mapreduce-week2-hw1:~/hadoop-3.4.0$ ssh localhost
Welcome to Ubuntu 20.04.6 LTS (GNU/Linux 5.15.0-1060-gcp x86_64)

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

System information as of Tue Jun  4 22:58:04 UTC 2024

System load:  0.01               Processes:    110
Usage of /:   54.3% of 9.51GB    Users logged in: 1
Memory usage: 24%               IPv4 address for ens4: 10.128.0.2
Swap usage:   0%

 * Strictly confined Kubernetes makes edge and IoT secure. Learn how MicroK8s
   just raised the bar for easy, resilient and secure K8s cluster deployment.

   https://ubuntu.com/engage/secure-kubernetes-at-the-edge

Expanded Security Maintenance for Applications is not enabled.

3 updates can be applied immediately.
3 of these updates are standard security updates.
To see these additional updates run: apt list --upgradable

Enable ESM Apps to receive additional future security updates.
See https://ubuntu.com/esm or run: sudo pro status

New release '22.04.3 LTS' available.
Run 'do-release-upgrade' to upgrade to it.

Last login: Tue Jun  4 22:26:27 2024 from 127.0.0.1
skavishw276@mapreduce-week2-hw1:~$
```


Test

- Continue to start the cluster and start the services. Test connection with localhost.

```
skavishw276@mapreduce-week2-hw1:~/hadoop-3.4.0$ wget http://localhost:9870/
--2024-06-04 23:01:38-- 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]
--2024-06-04 23:01:38-- 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.14'

index.html.14      100%[=====>]    1.05K  --.-KB/s    in 0s

2024-06-04 23:01:38 (117 MB/s) - 'index.html.14' saved [1079/1079]

skavishw276@mapreduce-week2-hw1:~/hadoop-3.4.0$
```

Test

- Compile the GenerateDots.java
- Run the java code with radius 5 and 1000 random numbers

```
skavishw276@mapreduce-week2-hw1:~/Pi$ javac GenerateDots.java
skavishw276@mapreduce-week2-hw1:~/Pi$ ls
CalculatePi.java    GenerateDots.class  input
CalculatePiMR.java  GenerateDots.java
```

```
skavishw276@mapreduce-week2-hw1:~/Pi$ java GenerateDots 5 1000 > ./input/dots
.txt
skavishw276@mapreduce-week2-hw1:~/Pi$ head 10 ./input/dots.txt
head: cannot open '10' for reading: No such file or directory
==> ./input/dots.txt <==
1.7692907060846363 2.738563128317506 5.0
3.5356536405175163 6.570534980209852 5.0
6.6715899037452715 7.052608326683471 5.0
7.040217823977011 7.812642393491405 5.0
4.1282046757072575 4.785400174092062 5.0
7.504404011955912 6.355162272665623 5.0
3.3483131820619283 0.5025313515966423 5.0
5.584599565550805 3.0607094377238364 5.0
7.206123334587603 9.258170140579068 5.0
8.268198086780538 0.04199631225596412 5.0
skavishw276@mapreduce-week2-hw1:~/Pi$
```


Test

- Create following directories
- Copy file from local machine to hadoop
- Compile in hadoop

```
skavishw276@mapreduce-week2-hw1:~/Pi$  
skavishw276@mapreduce-week2-hw1:~/Pi$ cd ../hadoop-3.4.0/  
skavishw276@mapreduce-week2-hw1:~/hadoop-3.4.0$ bin/hdfs dfs -mkdir /user  
skavishw276@mapreduce-week2-hw1:~/hadoop-3.4.0$ bin/hdfs dfs -mkdir /user/skavishw276  
skavishw276@mapreduce-week2-hw1:~/hadoop-3.4.0$ bin/hdfs dfs -mkdir /user/skavishw276/Pi  
skavishw276@mapreduce-week2-hw1:~/hadoop-3.4.0$ bin/hdfs dfs -mkdir /user/skavishw276/Pi/input  
skavishw276@mapreduce-week2-hw1:~/hadoop-3.4.0$
```

```
skavishw276@mapreduce-week2-hw1:~/hadoop-3.4.0$ bin/hdfs dfs -put ../Pi/input/* /user/skavishw276/Pi/input  
skavishw276@mapreduce-week2-hw1:~/hadoop-3.4.0$ bin/hdfs dfs -ls Pi/input  
Found 1 items  
-rw-r--r-- 1 skavishw276 supergroup 40569 2024-06-04 23:20 Pi/input/dots.txt  
skavishw276@mapreduce-week2-hw1:~/hadoop-3.4.0$
```

```
skavishw276@mapreduce-week2-hw1:~/hadoop-3.4.0$ bin/hdfs dfs -ls /user/skavishw276/Pi/input  
Found 1 items  
-rw-r--r-- 1 skavishw276 supergroup 40569 2024-06-04 23:20 /user/skavishw276/Pi/input/dots.txt
```

```
skavishw276@mapreduce-week2-hw1:~/hadoop-3.4.0$ 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.
```

Test

- Mapper and reduce files are created after compiling

```
skavishw276@mapreduce-week2-hw1:~/hadoop-3.4.0$ ls -lrt | grep -i cal*
-rw-rw-r-- 1 skavishw276 skavishw276 1330 Jun  4 23:37 CalculatePi.java
-rw-rw-r-- 1 skavishw276 skavishw276 2877 Jun  4 23:38 CalculatePiMR.java
-rw-rw-r-- 1 skavishw276 skavishw276 2404 Jun  4 23:39 CalculatePiMR$Map.class
-rw-rw-r-- 1 skavishw276 skavishw276 1639 Jun  4 23:39 CalculatePiMR$Reduce.class
-rw-rw-r-- 1 skavishw276 skavishw276 1483 Jun  4 23:39 CalculatePiMR.class
skavishw276@mapreduce-week2-hw1:~/hadoop-3.4.0$
```

```
skavishw276@mapreduce-week2-hw1:~/hadoop-3.4.0$ jar cf pi.jar CalculatePiMR*.class
```

```
skavishw276@mapreduce-week2-hw1:~/hadoop-3.4.0$ ls -lrt | grep -i jar
-rw-rw-r-- 1 skavishw276 skavishw276 3069 May 30 06:10 wc.jar
-rw-rw-r-- 1 skavishw276 skavishw276 3272 Jun  4 23:46 pi.jar
```

Test

- Run MapReduce program with input file and save the output file
- Get command on hdfs filesystem to get the output and save the file to local machine

```
skavishw276@mapreduce-week2-hw1:~/hadoop-3.4.0$ bin/hadoop jar pi.jar CalculatePiMR /user/skavishw276/Pi/input /user/skavishw276/Pi/Output
2024-06-04 23:50:02,265 INFO impl.MetricsConfig: Loaded properties from hadoop-metrics2.properties
2024-06-04 23:50:02,511 INFO impl.MetricsSystemImpl: Scheduled Metric snapshot period at 10 second(s).
2024-06-04 23:50:02,512 INFO impl.MetricsSystemImpl: JobTracker metrics system started
2024-06-04 23:50:02,884 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface
ication with ToolRunner to remedy this.
2024-06-04 23:50:03,235 INFO input.FileInputFormat: Total input files to process : 1
2024-06-04 23:50:03,313 INFO mapreduce.JobSubmitter: number of splits:1
2024-06-04 23:50:03,743 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_local436295171_0001
2024-06-04 23:50:03,744 INFO mapreduce.JobSubmitter: Executing with tokens: []
2024-06-04 23:50:04,030 INFO mapreduce.Job: The url to track the job: http://localhost:8080/
```

```
skavishw276@mapreduce-week2-hw1:~/hadoop-3.4.0$ bin/hdfs dfs -get Pi/Output
skavishw276@mapreduce-week2-hw1:~/hadoop-3.4.0$
```

```
skavishw276@mapreduce-week2-hw1:~/hadoop-3.4.0$ ls -lrt | grep Output
drwxr-xr-x 2 skavishw276 skavishw276 4096 Jun  4 23:52 Output
skavishw276@mapreduce-week2-hw1:~/hadoop-3.4.0$
```

Test

- Number of inside and outside points

```
skavishw276@mapreduce-week2-hw1:~/hadoop-3.4.0$ cat Output/*  
Inside 775  
Outside 225  
skavishw276@mapreduce-week2-hw1:~/hadoop-3.4.0$
```

Result

- Number of inside and outside points
- The value of $\pi = 3.1$ which is pretty close to the actual value of π

```
skavishw276@mapreduce-week2-hw1:~/hadoop-3.4.0$ vi CalculatePi.java
skavishw276@mapreduce-week2-hw1:~/hadoop-3.4.0$ javac CalculatePi.java
skavishw276@mapreduce-week2-hw1:~/hadoop-3.4.0$ java CalculatePi Output1
Inside 775
Outside 225
PI value is: 3.1
skavishw276@mapreduce-week2-hw1:~/hadoop-3.4.0$
```

Enhanced Result

- Increase the number of dots to 1000000. As we increase the number of dots the accuracy tends to increase.

```
skavishw276@mapreduce-week2-hw1:~/Pi$ java GenerateDots 5 1000000 > ./input/p
oints.txt
skavishw276@mapreduce-week2-hw1:~/Pi$ ls ./input/
dots.txt  points.txt
skavishw276@mapreduce-week2-hw1:~/Pi$
```

```
skavishw276@mapreduce-week2-hw1:~/Pi$ head -10 ./input/points.txt
6.26519912004941 5.5207256711663755 5.0
6.4935353124386666 6.732341661204758 5.0
5.826616089580955 2.4617657489413625 5.0
8.594162799345526 4.79803177870831 5.0
7.2259203273970085 6.1482829980085025 5.0
5.423297623469873 0.7784022493094422 5.0
1.6526242988991124 4.401180908414524 5.0
7.767727960121387 9.341840933240071 5.0
4.7167917290821295 0.24515002867913305 5.0
0.03201684699011054 6.898210369139509 5.0
```

Enhanced Result

- Copy points input file to hdfs

```
skavishw276@mapreduce-week2-hw1:~/hadoop-3.4.0$ bin/hdfs dfs -put ../Pi/input/points.txt Pi/input
skavishw276@mapreduce-week2-hw1:~/hadoop-3.4.0$ bin/hdfs dfs -ls Pi/input
Found 2 items
-rw-r--r--   1 skavishw276 supergroup      40569 2024-06-04 23:20 Pi/input/dots.txt
-rw-r--r--   1 skavishw276 supergroup 40538882 2024-06-05 00:34 Pi/input/points.txt
skavishw276@mapreduce-week2-hw1:~/hadoop-3.4.0$
```

```
skavishw276@mapreduce-week2-hw1:~/hadoop-3.4.0$ bin/hadoop jar pi.jar CalculatePiMR /user/skavishw276/Pi/input/points.txt /user/skavishw276/Pi/Points
2024-06-05 00:40:37,526 INFO Impl.MetricsConfig: Loaded properties from hadoop-metrics2.properties
2024-06-05 00:40:37,761 INFO Impl.MetricsSystemImpl: Scheduled Metric snapshot period at 10 second(s).
2024-06-05 00:40:37,761 INFO Impl.MetricsSystemImpl: JobTracker metrics system started
2024-06-05 00:40:39,112 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 00:40:38,324 INFO input.FileInputFormat: Total input files to process : 1
2024-06-05 00:40:38,457 INFO mapreduce.JobSubmitter: number of splits:1
2024-06-05 00:40:38,821 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_local192535477_0001
```

```
skavishw276@mapreduce-week2-hw1:~/hadoop-3.4.0$ bin/hdfs dfs -get Pi/Points Points
skavishw276@mapreduce-week2-hw1:~/hadoop-3.4.0$ cat Points/*
Inside 784833
Outside 215167
skavishw276@mapreduce-week2-hw1:~/hadoop-3.4.0$
```

Final Result

- The value of $\pi = 3.139332$ when the generated dots were 1M.

```
skavishw276@mapreduce-week2-hw1:~/hadoop-3.4.0$ java CalculatePi Points  
Inside 784833  
Outside 215167  
PI value is: 3.139332  
skavishw276@mapreduce-week2-hw1:~/hadoop-3.4.0$
```

Conclusion

- The accuracy of value of π increases as the number of generated points increases.