

Step 1: Generate an input file to the Pi MapReduce program

Step 1.1: Create a regular Java program which accepts two command line arguments.

R: The radius

N: The number of (x, y) pairs to create The Java program then randomly generates N pairs of (x, y) and displays them on the standard output.

Step 1.2: Run the program created in Step 1.1 and save the result in a file. The file is the input to Step 2's Pi MapReduce program.

Step 2: Create a MapReduce program to calculate the numbers of inside darts and outside darts.

Step 3: Use the file generated in Step 1.2 as the input to execute the MapReduce program created in Step 2

Step 4: Calculate Pi in the driver program based on the numbers of inside darts and outside darts.

GenerateRandomNumbers.java

```
import java.io.File;
import java.io.FileWriter;
import java.io.IOException;
import java.util.Scanner;

public class GenerateRandomNumbers {
    public static void main(String[] args) {
        System.out.println("How many random numbers to generate:"); // we use 1000000 to test
        Scanner input = new Scanner(System.in);
        int RandomNumCount = input.nextInt();

        System.out.println("What's the radius?"); //we use 200 to test
        int radius = input.nextInt();
        int diameter = radius * 2;
```

```

input.close();

try {

    // it creates file input4
    File file = new File("./PiCalculationInput");
    file.createNewFile();

    // Prepare input data
    FileWriter writer = new FileWriter(file);
    //writer.write(radius + "\r\n");
    //writer.write(System.getProperty("line.separator"));

    for (int i = 0; i < RandomNumCount; i++) {
        int xvalue = (int) (Math.random() * diameter);
        int yvalue = (int) (Math.random() * diameter);
        writer.write("(" + xvalue + "," + yvalue + ") ");
    }

    // send the data into the file
    writer.flush();

    // closing the write after pushing the data inside the .txt file
    writer.close();
} catch (IOException e) {
    e.printStackTrace();
}
}
}

```

Picalculation.java

```

import java.io.*;
import java.util.*;
import java.lang.Object;
import java.net.URI;

import org.apache.hadoop.fs.Path;
import org.apache.hadoop.conf.*;
import org.apache.hadoop.io.*;
import org.apache.hadoop.mapreduce.*;
import org.apache.hadoop.mapreduce.Mapper.Context;

```

```

import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;
import org.apache.hadoop.mapreduce.lib.input.TextInputFormat;
import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;
import org.apache.hadoop.mapreduce.lib.output.TextOutputFormat;
import org.apache.hadoop.fs.*;

public class PiCalculation {

    public static class TokenizerMapper
        extends Mapper<Object, Text, Text, IntWritable> {

        private final static IntWritable one = new IntWritable(1);
        private Text word = new Text();
        private int totalLines = 0;

        public void map(Object key, Text value, Context context) throws IOException,
        InterruptedException {

            totalLines += 1;
            String line = value.toString();
            line = line.replace("(", "");
            line = line.replace(")", "");
            line = line.replace(", ", " ");

            StringTokenizer itr = new StringTokenizer(line);
            int radius = 200; // Same as the one you give in PiDataGenerator stage
            while (itr.hasMoreTokens()) {
                String x, y;
                x = itr.nextToken();
                if (itr.hasMoreTokens()) {
                    y = itr.nextToken();
                } else {
                    y = "0";
                }
                int xvalue = (int) (Integer.parseInt(x));
                int yvalue = (int) (Integer.parseInt(y));
                double check = Math.sqrt(Math.pow((radius - xvalue), 2) + Math.pow((radius - yvalue),
2));

                if (check < radius) {
                    word.set("inside");
                } else {
                    word.set("outside");
                }
            }
        }
    }
}

```

```

        context.write(word, one);
    }
}

```

```

public static class IntSumReducer
    extends Reducer<Text, IntWritable, Text, IntWritable> {
    private IntWritable result = new IntWritable();

    public void reduce(Text key, Iterable<IntWritable> values,
        Context context) throws IOException, InterruptedException {
        int sum = 0;
        for (IntWritable val : values) {
            sum += val.get();
        }
        result.set(sum);
        context.write(key, result);
    }
}

```

```

public static void main(String[] args) throws Exception {
    Configuration conf = new Configuration();
    Job job = Job.getInstance(conf, "pi calculation");
    job.setJarByClass(PiCalculation.class);
    job.setMapperClass(TokenizerMapper.class);
    job.setCombinerClass(IntSumReducer.class);
    job.setReducerClass(IntSumReducer.class);
    job.setOutputKeyClass(Text.class);
    job.setOutputValueClass(IntWritable.class);
    FileInputFormat.addInputPath(job, new Path(args[0]));
    FileOutputFormat.setOutputPath(job, new Path(args[1]));
    // System.exit(job.waitForCompletion(true) ? 0 : 1);
    job.waitForCompletion(true);
    String filePath = args[1] + "/" + "part-r-00000";
    Path path = new Path(filePath);
    FileSystem fs = FileSystem.get(path.toUri(), conf);

    BufferedReader br = new BufferedReader(new InputStreamReader(fs.open(path)));

    String z, inside = null, outside = null;

    String line1, line2;

    line1 = br.readLine();

```

```
System.out.println(line1);
line2 = br.readLine();
System.out.println(line2);

line1 = line1.replace("inside", "").trim();
line2 = line2.replace("outside", "").trim();

System.out.println("Inside:" + line1 + ", Outside:" + line2);

if (line1 != null && line2 != null) {
    double inval = Double.valueOf(line1);
    double outval = Double.valueOf(line2);
    double pi = 4 * (inval / (inval + outval));
    System.out.println("PI:" + pi);
}

fs.close();
}
}
```

GCP Environment

Google Cloud CS570BigData Search (/) for resources, docs, products, and more

Compute Engine VM instances CREATE INSTANCE IMPORT VM REFRESH HELP ASSISTANT LEARN

Virtual machines VM instances Instance templates Sole-tenant nodes Machine images TPUs Committed use discounts Reservations Migrate to Virtual Machin...

Storage Disks Snapshots Images Marketplace Release Notes

VM instances

Filter Enter property name or value

Status	Name	Zone	Recommendations	In use by	Internal IP	External IP	Connect
	bigdata-week2	us-central1-a			10.128.0.3 (nic0)		SSH

Related actions

- Explore Backup and DR NEW 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 and network
- Explore VM logs View, search, analyze, and download VM instance logs
- Set up firewall rules Control traffic to and from a VM instance
- Patch management Schedule patch updates and view patch compliance on VM instances
- Load balance between VMs Set up Load Balancing for your applications as your traffic and users grow

Start / Resume Stop Suspend Reset Delete View network details Create new machine image View logs View monitoring

Hadoop Environment

SSH-in-browser UPLOAD FILE DOWNLOAD FILE

```

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

System information as of Tue Jun 6 18:51:45 UTC 2023

System load: 0.1 Processes: 117
Usage of /: 57.5% of 9.51GB Users logged in: 1
Memory usage: 6% IPv4 address for ens4: 10.128.0.3
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

* Introducing Expanded Security Maintenance for Applications.
Receive updates to over 25,000 software packages with your
Ubuntu Pro subscription. Free for personal use.

https://ubuntu.com/gcp/pro

Expanded Security Maintenance for Applications is not enabled.

25 updates can be applied immediately.
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

The list of available updates is more than a week old.
To check for new updates run: sudo apt update
New release '22.04.2 LTS' available.
Run 'do-release-upgrade' to upgrade to it.

Last login: Tue Jun 6 18:50:43 2023 from 35.235.244.32
alope111777@bigdata-week2:~$ ls
WordCount hadoop-3.3.4 hadoop-3.3.4.tar.gz
alope111777@bigdata-week2:~$

```

Prepare input data

Commands:

```
$ mkdir PiCalculation
$ cd PiCalculation
$ vi GenerateRandomNumbers.java
$ javac GenerateRandomNumbers.java
$ java -cp . GenerateRandomNumbers
```

Setup passphraseless ssh

check that you can ssh to the localhost without a passphrase:

```
$ cd hadoop-3.3.4/
$ ssh localhost
```

If you cannot ssh to localhost without a passphrase, execute the following commands:

```
$ ssh-keygen -t rsa -P "" -f ~/.ssh/id_rsa
$ cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys
$ chmod 0600 ~/.ssh/authorized_keys
$ ssh localhost
```

```

alopelli777@bigdata-week2:~$ cd hadoop-3.3.4/
alopelli777@bigdata-week2:~/hadoop-3.3.4$ ssh localhost
alopelli777@localhost: Permission denied (publickey).
alopelli777@bigdata-week2:~/hadoop-3.3.4$ ssh-keygen -t rsa -P '' -f ~/.ssh/id_rsa
Generating public/private rsa key pair.
/home/alopelli777/.ssh/id_rsa already exists.
Overwrite (y/n)? y
Your identification has been saved in /home/alopelli777/.ssh/id_rsa
Your public key has been saved in /home/alopelli777/.ssh/id_rsa.pub
The key fingerprint is:
SHA256:u9dAdqemyMw12KP08H4/KjbR14qLi2icuX/kk2zEc0U alopelli777@bigdata-week2
The key's randomart image is:
+---[RSA 3072]-----+
|
|             E |
|             . |
|          o . o |
|        S* o + . |
|       +.% = . . |
|      . B.% % o . |
|     =.*o/.+ o |
|    .oo+**+*o+.. |
+-----[SHA256]-----+
alopelli777@bigdata-week2:~/hadoop-3.3.4$ cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys
alopelli777@bigdata-week2:~/hadoop-3.3.4$ chmod 0600 ~/.ssh/authorized_keys
alopelli777@bigdata-week2:~/hadoop-3.3.4$ ssh localhost
Welcome to Ubuntu 20.04.5 LTS (GNU/Linux 5.15.0-1034-gcp x86_64)

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

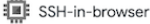
```

Make the HDFS directories required to execute MapReduce jobs(Copy input data to HDFS)

```

$ cd ..
$ cd hadoop-3.3.4/
$ bin/hdfs namenode -format
$ sbin/start-dfs.sh
$ wget http://localhost:9870/
$ bin/hdfs dfs -mkdir /user
$ bin/hdfs dfs -mkdir /user/alopelli777
$ bin/hdfs dfs -mkdir /user/alopelli777/picalculate
$ bin/hdfs dfs -mkdir /user/alopelli777/picalculate/input
$ bin/hdfs dfs -put ../PiCalculation/PiCalculationInput /user/alopelli777/picalculate/input

```

 UPLOAD FILE
 DOWNLOAD FILE



 UPLOAD FILE
  DOWNLOAD FILE
 



```

alopelli777@bigdata-week2:~/hadoop-3.3.4$ wget http://localhost:9870/
--2023-06-06 19:24:24-- 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]
--2023-06-06 19:24:24-- 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.5'

index.html.5          100%[=====>] 1.05K  --.-
KB/s   in 0s

2023-06-06 19:24:24 (317 MB/s) - 'index.html.5' saved [1079/1079]

alopelli777@bigdata-week2:~/hadoop-3.3.4$

```

```

alopelli777@bigdata-week2:~/hadoop-3.3.4$ bin/hdfs dfs -mkdir /user
alopelli777@bigdata-week2:~/hadoop-3.3.4$ bin/hdfs dfs -mkdir /user/alopelli777
alopelli777@bigdata-week2:~/hadoop-3.3.4$ bin/hdfs dfs -mkdir /user/alopelli777/picalculate
alopelli777@bigdata-week2:~/hadoop-3.3.4$ bin/hdfs dfs -mkdir /user/alopelli777/picalculate/input
alopelli777@bigdata-week2:~/hadoop-3.3.4$ bin/hdfs dfs -put ../PiCalculation/PiCalculationInput /user

```

Code preparation

Build PiCalculation java file

```

$ cd /hadoop-3.3.4
$ vi PiCalculation.java

```

Compile PiCalculation.java and create a jar

```

$ bin/hadoop com.sun.tools.javac.Main PiCalculation.java
$ jar cf wc.jar PiCalculation*.class

```

```

alopelli777@bigdata-week2:~$ cd /hadoop-3.3.4/
alopelli777@bigdata-week2:~/hadoop-3.3.4$ vi PiCalculation.java
alopelli777@bigdata-week2:~/hadoop-3.3.4$ ls
LICENSE-binary      PiCalculation.java      etc                  index.html.5        logs
LICENSE.txt         README.txt              include             index.html.6        output
NOTICE-binary       'WordCount$IntSumReducer.class'  index.html          index.html.7        sbin
NOTICE.txt          'WordCount$TokenizerMapper.class'  index.html.1        input              share
'PiCalculation$IntSumReducer.class'  WordCount.class         index.html.2        lib                wc.jar
'PiCalculation$TokenizerMapper.class' WordCount.java           index.html.3        libexec
PiCalculation.class bin                      index.html.4        licenses-binary
alopelli777@bigdata-week2:~/hadoop-3.3.4$

```

```

alopelli777@bigdata-week2:~/hadoop-3.3.4$ bin/hadoop com.sun.tools.javac.Main PiCalculation.java
alopelli777@bigdata-week2:~/hadoop-3.3.4$ jar cf wc.jar PiCalculation*.class

```

Run

Execute

```
$ bin/hadoop jar wc.jar PiCalculation /user/alopelli777/picalculate/input
/user/alopelli777/picalculate/output7
```

Output

```
$ bin/hdfs dfs -ls /user/alopelli777/picalculate/output7
$ bin/hdfs dfs -cat /user/alopelli777/picalculate/output7/part-r-00000
```

Stop

```
$ sbin/stop-dfs.sh
```

```
alopelli777@bigdata-week2:~/hadoop-3.3.4$ bin/hadoop jar wc.jar PiCalculation /user/alopelli777/picalculate/input2 /use
r/alopelli777/picalculate/output7
2023-06-06 22:26:55,117 INFO impl.MetricsConfig: Loaded properties from hadoop-metrics2.properties
2023-06-06 22:26:55,256 INFO impl.MetricsSystemImpl: Scheduled Metric snapshot period at 10 second(s).
2023-06-06 22:26:55,256 INFO impl.MetricsSystemImpl: JobTracker metrics system started
2023-06-06 22:26:55,531 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement
the Tool interface and execute your application with ToolRunner to remedy this.
2023-06-06 22:26:55,752 INFO input.FileInputFormat: Total input files to process : 1
2023-06-06 22:26:55,782 INFO mapreduce.JobSubmitter: number of splits:1
2023-06-06 22:26:55,965 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_local1544285763_0001
2023-06-06 22:26:55,965 INFO mapreduce.JobSubmitter: Executing with tokens: []
2023-06-06 22:26:56,205 INFO mapreduce.Job: The url to track the job: http://localhost:8080/
2023-06-06 22:26:56,207 INFO mapreduce.Job: Running job: job_local1544285763_0001
2023-06-06 22:26:56,216 INFO mapred.LocalJobRunner: OutputCommitter set in config null
2023-06-06 22:26:56,232 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2023-06-06 22:26:56,232 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under outp
ut directory:false, ignore cleanup failures: false
2023-06-06 22:26:56,234 INFO mapred.LocalJobRunner: OutputCommitter is org.apache.hadoop.mapreduce.lib.output.FileOutpu
tCommitter
2023-06-06 22:26:56,331 INFO mapred.LocalJobRunner: Waiting for map tasks
2023-06-06 22:26:56,332 INFO mapred.LocalJobRunner: Starting task: attempt_local1544285763_0001_m_000000_0
2023-06-06 22:26:56,376 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2023-06-06 22:26:56,376 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under outp
ut directory:false, ignore cleanup failures: false
2023-06-06 22:26:56,405 INFO mapred.Task: Using ResourceCalculatorProcessTree : [ ]
```



```

Map output materialized bytes=33
Input split bytes=141
Combine input records=1000000
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)=356
Total committed heap usage (bytes)=1109393408

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=9450165
File Output Format Counters
  Bytes Written=29

inside  785619
outside 214381
Inside:785619, Outside:214381
PI:3.142476
alopelli777@bigdata-week2:~/hadoop-3.3.4$ █

```

```

alopelli777@bigdata-week2:~/hadoop-3.3.4$ bin/hdfs dfs -ls /user/alopelli777/picalculate/output7
Found 2 items
-rw-r--r--  1 alopelli777 supergroup          0 2023-06-06 22:26 /user/alopelli777/picalculate/output7/_SUCCESS
-rw-r--r--  1 alopelli777 supergroup        29 2023-06-06 22:26 /user/alopelli777/picalculate/output7/part-r-00000
alopelli777@bigdata-week2:~/hadoop-3.3.4$ bin/hdfs dfs -cat /user/alopelli777/picalculate/output7/part-r-00000
inside  785619
outside 214381

```

```

alopelli777@bigdata-week2:~/hadoop-3.3.4$ sbin/stop-dfs.sh

Stopping namenodes on [localhost]
Stopping datanodes
Stopping secondary namenodes [bigdata-week2]
alopelli777@bigdata-week2:~/hadoop-3.3.4$
alopelli777@bigdata-week2:~/hadoop-3.3.4$ █

```