MapReduce in Calculating Pi

Linh Bien
Professor: Henry Chang

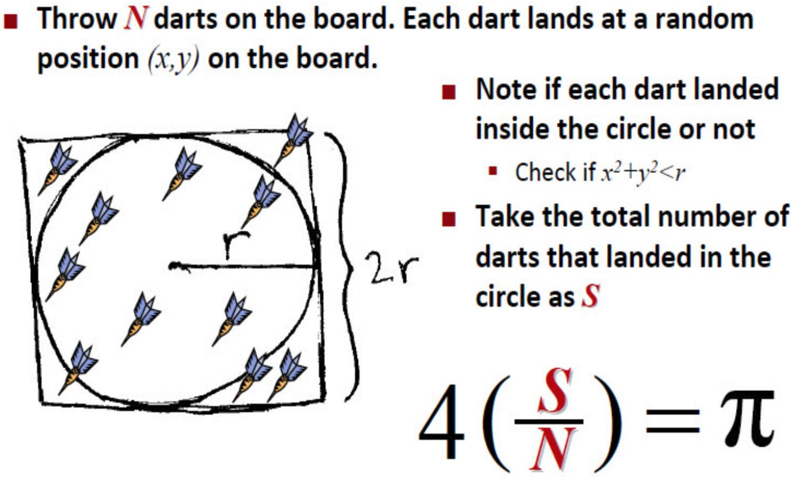
Table of content

- Introduction
- Design
- Implementation
- Test
- Enhancement ideas
- Conclusion
- Reference

Introduction

In this project, the Hadoop environment is used to calculate Pi

Design



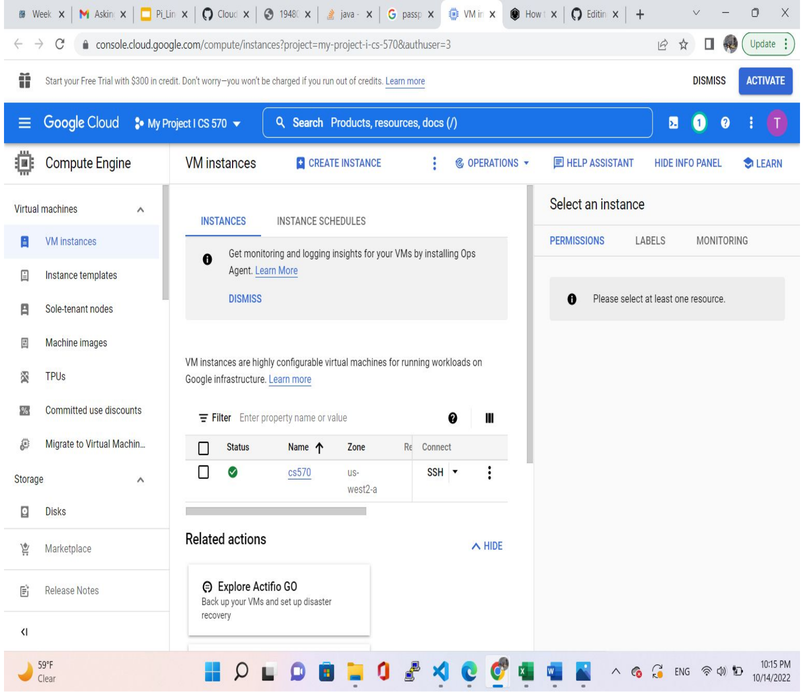
- Throw N darts on the board. Each dart lands at a random position (x,y) on the board.
- Note if each dart landed inside the circle or not
  - Check if $x^2+y^2<r$
- Take the total number of darts that landed in the circle as S

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

Formula:

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

Implementation



Start your Free Trial with $300 in credit. Don't worry—you won't be charged if you run out of credits. Learn more

Get monitoring and logging insights for your VMs by installing Ops Agent. Learn More

DISMISS

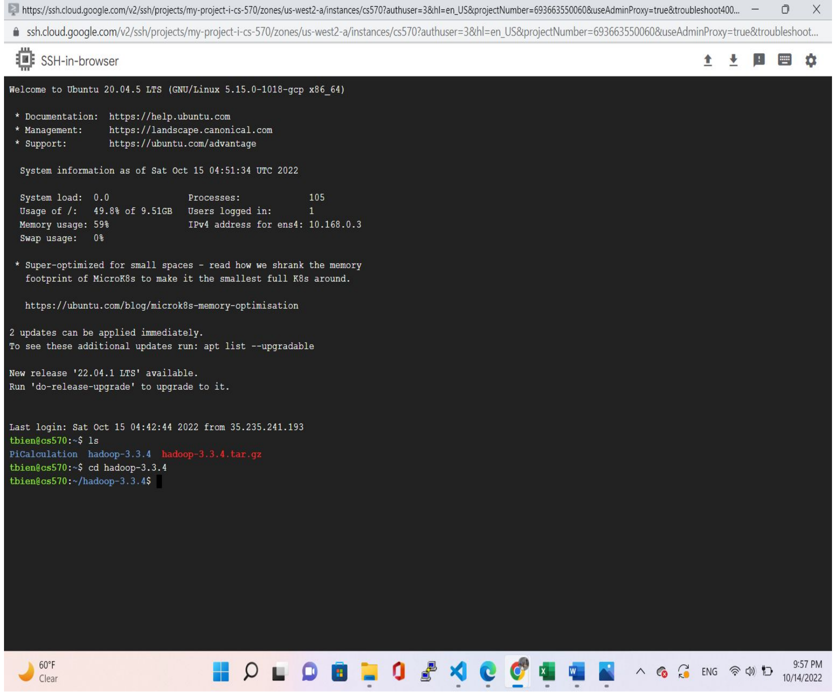
Please select at least one resource.

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

□	Status	Name ↑	Zone	Re	Connect
□	✓	cs570	us-west2-a		SSH ▾

Explore Actifio GO

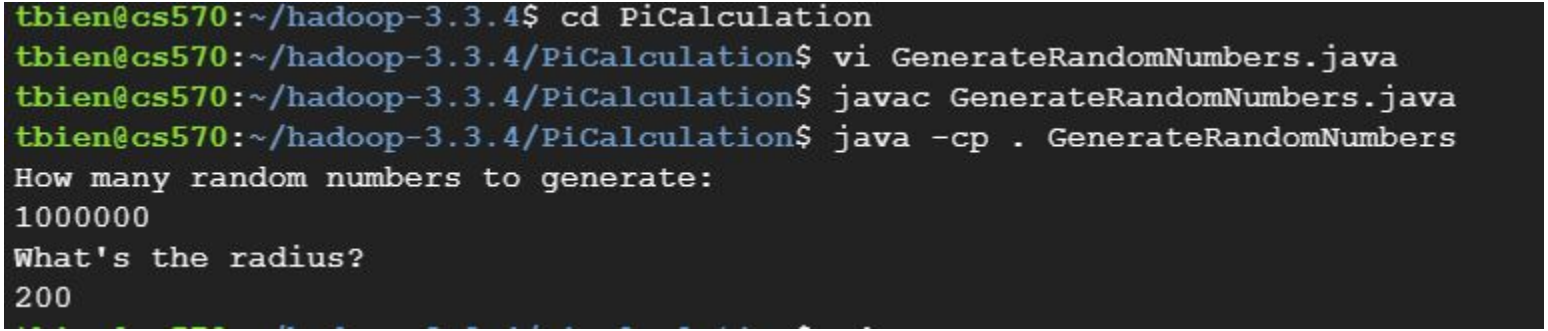
Back up your VMs and set up disaster recovery


https://ssh.cloud.google.com/v2/ssh/projects/my-project-i-cs-570/zones/us-west2-a/instances/cs570?authuser=3&hl=en_US&projectNumber=693663550060&useAdminProxy=true&troubleshoot400...
SSH-in-browser
60°F
Clear
ENG
9:57 PM
10/14/2022

```
Welcome to Ubuntu 20.04.5 LTS (GNU/Linux 5.15.0-1018-gcp x86_64)

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

  System information as of Sat Oct 15 04:51:34 UTC 2022

  System load:  0.0               Processes:             105
  Usage of /:   49.8% of 9.51GB   Users logged in:       1
  Memory usage: 59%               IPv4 address for ens4: 10.168.0.3
  Swap usage:   0%

 * Super-optimized for small spaces - read how we shrank the memory
   footprint of MicroK8s to make it the smallest full K8s around.

   https://ubuntu.com/blog/microk8s-memory-optimisation

2 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: Sat Oct 15 04:42:44 2022 from 35.235.241.193
tbien@cs570:~$ ls
PiCalculation  hadoop-3.3.4  hadoop-3.3.4.tar.gz
tbien@cs570:~$ cd hadoop-3.3.4
tbien@cs570:~/hadoop-3.3.4$
```

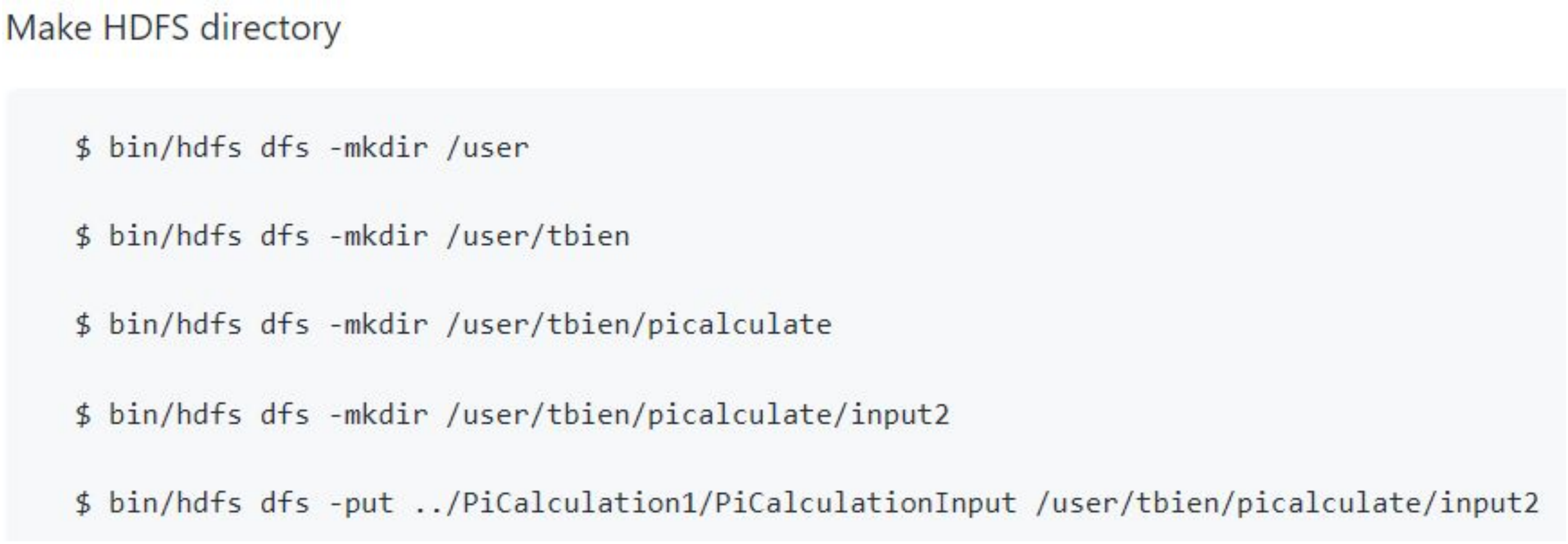
Implementation

```
$ mkdir PiCalculation1
```



```
tbien@cs570:~/hadoop-3.3.4$ cd PiCalculation
tbien@cs570:~/hadoop-3.3.4/PiCalculation$ vi GenerateRandomNumbers.java
tbien@cs570:~/hadoop-3.3.4/PiCalculation$ javac GenerateRandomNumbers.java
tbien@cs570:~/hadoop-3.3.4/PiCalculation$ java -cp . GenerateRandomNumbers
How many random numbers to generate:
1000000
What's the radius?
200
```

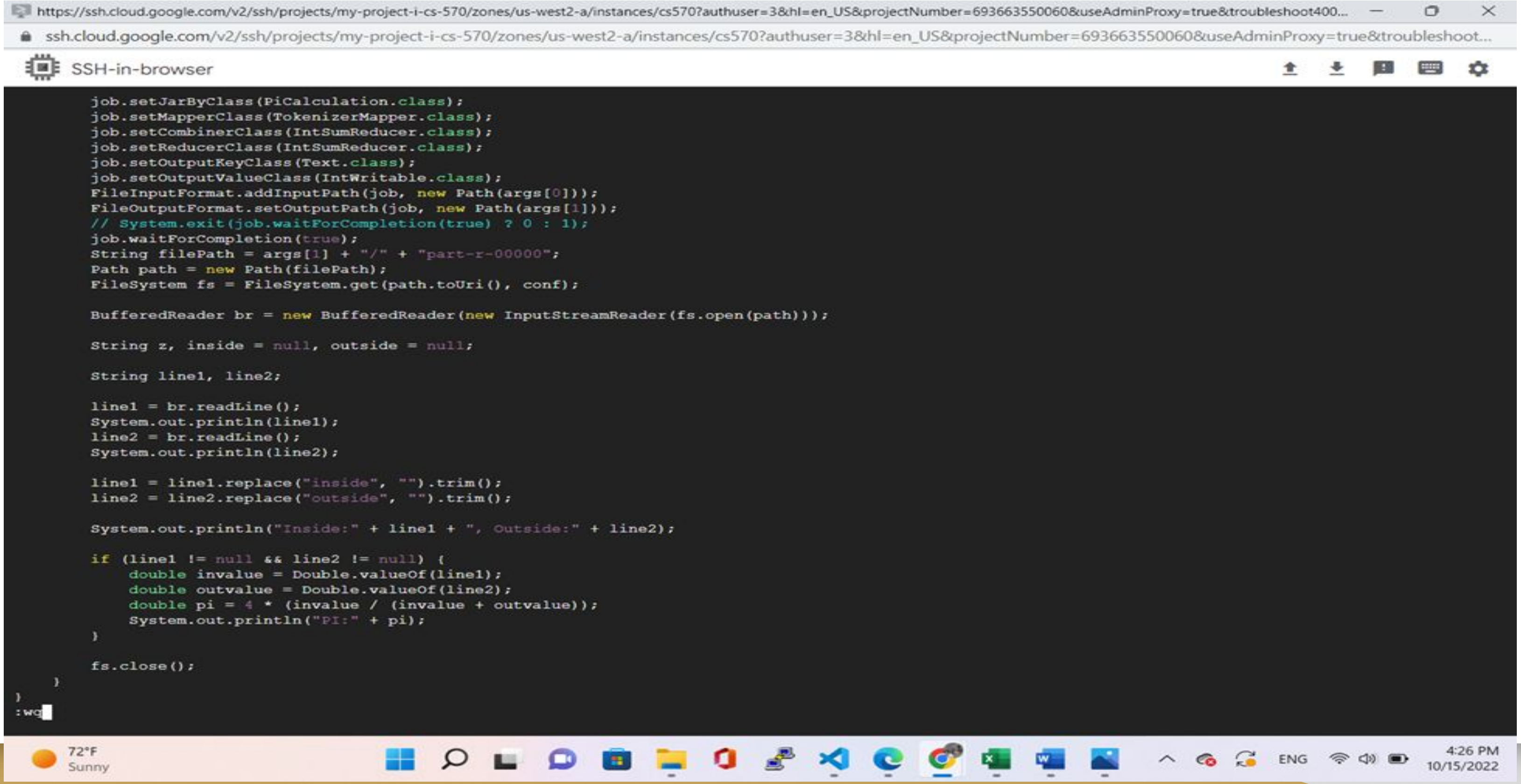
Implementation



Make HDFS directory

```
$ bin/hdfs dfs -mkdir /user

$ bin/hdfs dfs -mkdir /user/tbien

$ bin/hdfs dfs -mkdir /user/tbien/picalculate

$ bin/hdfs dfs -mkdir /user/tbien/picalculate/input2

$ bin/hdfs dfs -put ../PiCalculation1/PiCalculationInput /user/tbien/picalculate/input2
```

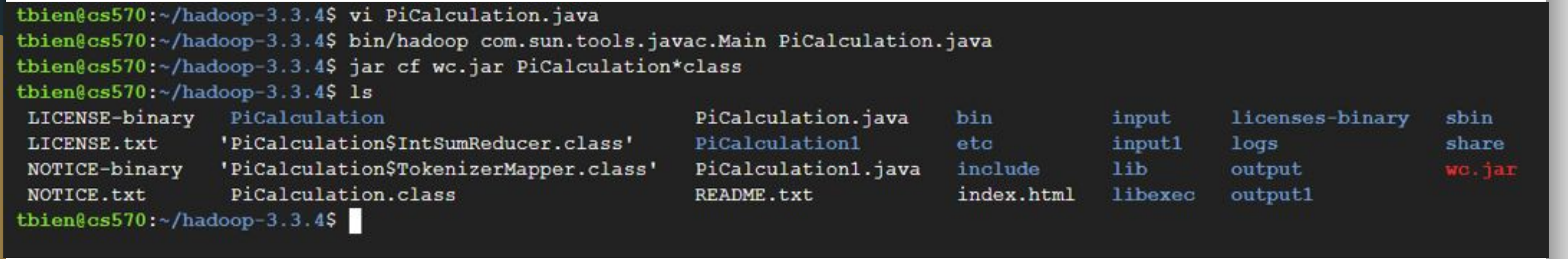
Implementation


https://ssh.cloud.google.com/v2/ssh/projects/my-project-i-cs-570/zones/us-west2-a/instances/cs570?authuser=3&hl=en_US&projectNumber=693663550060&useAdminProxy=true&troubleshoot400...
ssh.cloud.google.com/v2/ssh/projects/my-project-i-cs-570/zones/us-west2-a/instances/cs570?authuser=3&hl=en_US&projectNumber=693663550060&useAdminProxy=true&troubleshoot...
SSH-in-browser
72°F
Sunny
ENG
4:26 PM
10/15/2022

```
        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 invalue = Double.valueOf(line1);
            double outvalue = Double.valueOf(line2);
            double pi = 4 * (invalue / (invalue + outvalue));
            System.out.println("PI:" + pi);
        }

        fs.close();
    }
}
:wq
```

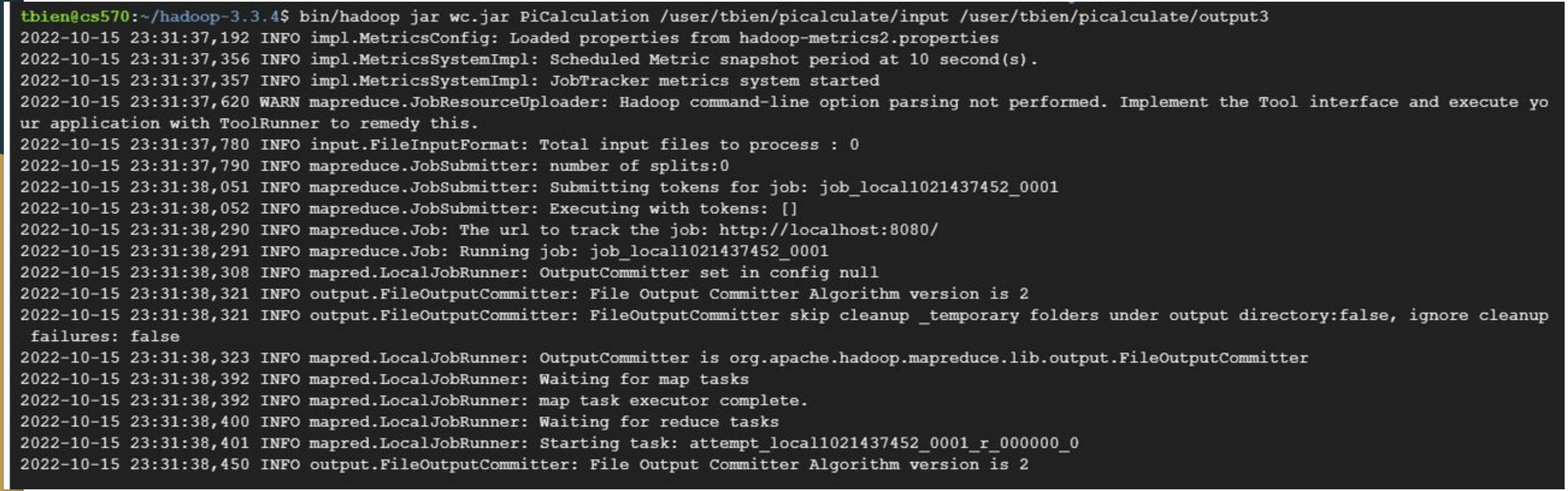
Implementation - MapReduce program

Create Mapreduce to calculate number of inside and outside darts



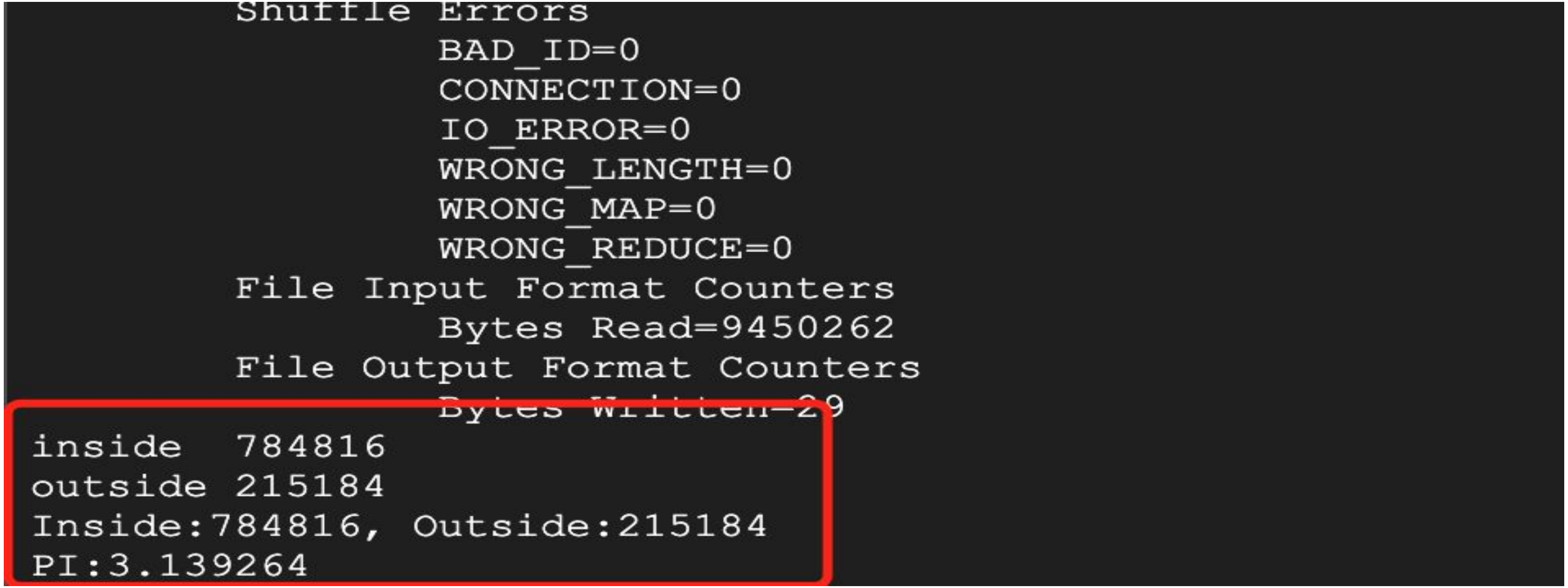
```
tbien@cs570:~/hadoop-3.3.4$ vi PiCalculation.java
tbien@cs570:~/hadoop-3.3.4$ bin/hadoop com.sun.tools.javac.Main PiCalculation.java
tbien@cs570:~/hadoop-3.3.4$ jar cf wc.jar PiCalculation*class
tbien@cs570:~/hadoop-3.3.4$ ls
 LICENSE-binary   PiCalculation                              PiCalculation.java    bin        input     licenses-binary   sbin
 LICENSE.txt     'PiCalculation$IntSumReducer.class'         PiCalculation1        etc        input1    logs              share
 NOTICE-binary   'PiCalculation$TokenizerMapper.class'       PiCalculation1.java   include    lib       output            wc.jar
 NOTICE.txt       PiCalculation.class                        README.txt            index.html libexec   output1
tbien@cs570:~/hadoop-3.3.4$
```

Implementation



```
tbien@cs570:~/hadoop-3.3.4$ bin/hadoop jar wc.jar PiCalculation /user/tbien/picalculate/input /user/tbien/picalculate/output3
2022-10-15 23:31:37,192 INFO impl.MetricsConfig: Loaded properties from hadoop-metrics2.properties
2022-10-15 23:31:37,356 INFO impl.MetricsSystemImpl: Scheduled Metric snapshot period at 10 second(s).
2022-10-15 23:31:37,357 INFO impl.MetricsSystemImpl: JobTracker metrics system started
2022-10-15 23:31:37,620 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute yo
ur application with ToolRunner to remedy this.
2022-10-15 23:31:37,780 INFO input.FileInputFormat: Total input files to process : 0
2022-10-15 23:31:37,790 INFO mapreduce.JobSubmitter: number of splits:0
2022-10-15 23:31:38,051 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_local1021437452_0001
2022-10-15 23:31:38,052 INFO mapreduce.JobSubmitter: Executing with tokens: []
2022-10-15 23:31:38,290 INFO mapreduce.Job: The url to track the job: http://localhost:8080/
2022-10-15 23:31:38,291 INFO mapreduce.Job: Running job: job_local1021437452_0001
2022-10-15 23:31:38,308 INFO mapred.LocalJobRunner: OutputCommitter set in config null
2022-10-15 23:31:38,321 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2022-10-15 23:31:38,321 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup
 failures: false
2022-10-15 23:31:38,323 INFO mapred.LocalJobRunner: OutputCommitter is org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter
2022-10-15 23:31:38,392 INFO mapred.LocalJobRunner: Waiting for map tasks
2022-10-15 23:31:38,392 INFO mapred.LocalJobRunner: map task executor complete.
2022-10-15 23:31:38,400 INFO mapred.LocalJobRunner: Waiting for reduce tasks
2022-10-15 23:31:38,401 INFO mapred.LocalJobRunner: Starting task: attempt_local1021437452_0001_r_000000_0
2022-10-15 23:31:38,450 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
```

Test



```
        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=9450262
        File Output Format Counters
                Bytes Written=29
inside  784816
outside 215184
Inside:784816, Outside:215184
PI:3.139264
```

Enhancement idea

Test more on larger numbers then we will have more accurate results

Conclusion

It is important to write right codes and run the correct results

Test on larger numbers to have more accurate results

References

Research Gate. https://www.researchgate.net/figure/MapReduce-calculation-process_fig2_359948761

Overview of Pi Calculation using MapReduce