Project : MapReduce Pi

Presented by MANICKAM RAVISEKAR ,
Master of Science in Computer Science, 19599 ,

Spring 2022

Guidance from Dr., Professor Henry Chang

SAN FRANCISCO BAY UNIVERSITY
47671 WestingHouse Dr.,
Fremont, CA 94539

ACKNOWLEDGEMENT

One of our master's degree Project for Hadoop MapReduce for PI computation,
Is Interesting, which made me to learn new things, it is useful in designing and applying on Google Cloud

 Platform (Hadoop MapReduce) on Google Cloud Platform and Oracle Virtual Machine /

For deploying this project , I would like to thank Dr. Henry Chang for providing all the required input .

Also, for all I would like to always pray to Almighty for giving us wisdom and power to understand things.
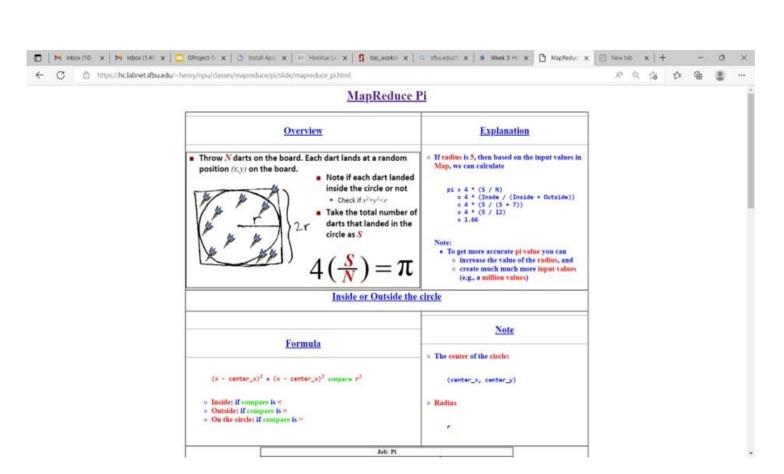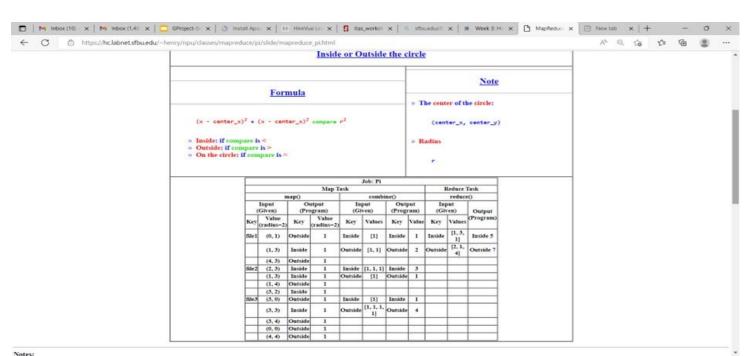
Index

ABSTRACT

The main lesson about this assignment is to study details,
about deploying applications on both Google Cloud Platform and Oracle Virtual Machine to learn
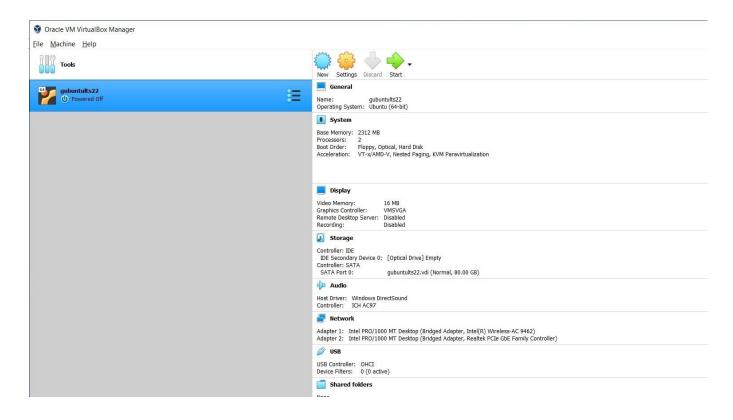
about Hadoop MapReduce to apply for Pi calculations with

radius and random points for a circle to compute points to find whether points are inside (darts) or

outside circle.

MapReduce Pi

Overview

- 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

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

Explanation

- If radius is 5, then based on the input values in Map, we can calculate

```
pi = 4 * (S / N)
   = 4 * (Inside / (Inside + Outside))
   = 4 * (5 / (5 + 7))
   = 4 * (5 / 12)
   = 1.66
```

Note:
- To get more accurate pi value you can
  - increase the value of the radius, and
  - create much much more input values (e.g., a million values)

Inside or Outside the circle

Formula

$(x - center\_x)^2 + (x - center\_x)^2$ compare r^2

- Inside: if compare is <
- Outside: if compare is >
- On the circle: if compare is =

Note

- The center of the circle:

  (center_x, center_y)

- Radius

  r

Job: Pi

M Inbox (10) × | M Inbox (1,40 × | GProject-S × | Install Apa × | HireVue L × | itas_worksh × | sfbu.edu:sl × | Week 3: H: × | MapRedu:: × | New tab × | +

https://hc.labnet.sfbu.edu/~henry/npu/classes/mapreduce/pi/slide/mapreduce_pi.html

Inside or Outside the circle

Formula

$$(x - center\_x)^2 + (x - center\_x)^2 \text{ compare } r^2$$

- Inside: if compare is <
- Outside: if compare is >
- On the circle: if compare is =

Note

- The center of the circle:

  (center_x, center_y)

- Radius

  r

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			
	(3, 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							

Notes:

MapReduce on Ubuntu on Oracle Virtual Machine local:

gubuntults22 - Settings

General
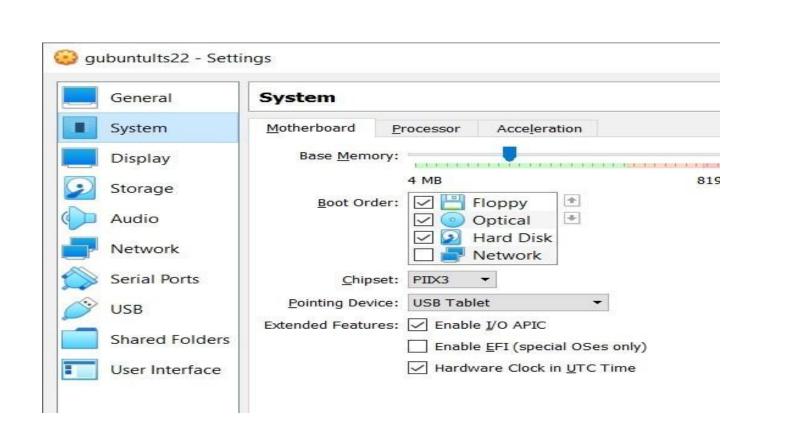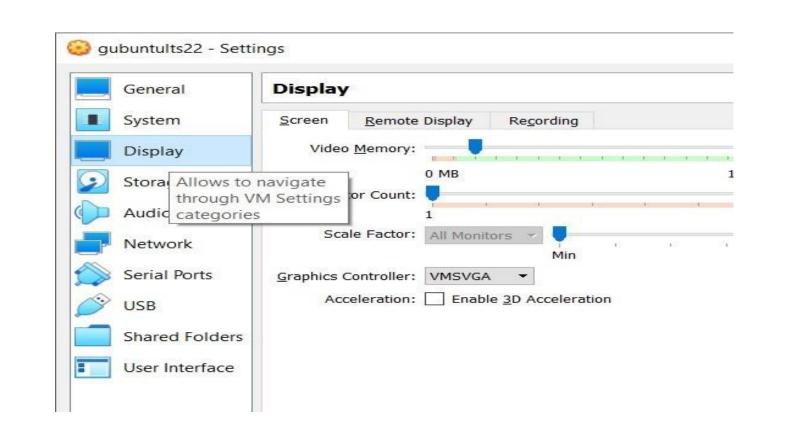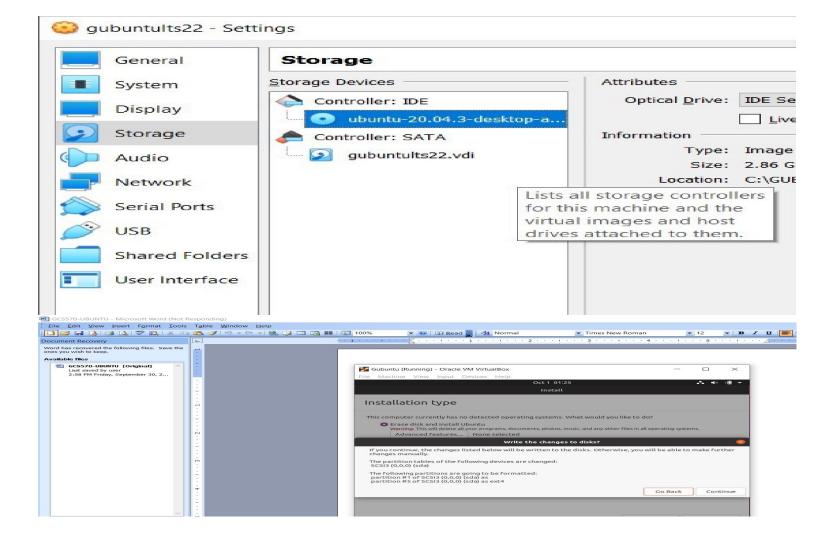
Basic	Advanced	Description	Disk Encryption

Name: gubuntults22

Type: Linux

Version: Ubuntu (64-bit)

gubuntults22 - Settings

Storage

Storage Devices

- Controller: IDE
  - ubuntu-20.04.3-desktop-a...
- Controller: SATA
  - gubuntults22.vdi

Attributes

Optical Drive: IDE Se...

☐ Live

Information

Type: Image
Size: 2.86 G
Location: C:\GUE

Lists all storage controllers for this machine and the virtual images and host drives attached to them.

GCS570-UBUNTU - Microsoft Word (Not Responding)

File Edit View Insert Format Tools Table Window Help

100% Read Normal Times New Roman 12 B I U

Document Recovery

Word has recovered the following files. Save the ones you wish to keep.

Available files

GCS570-UBUNTU [Original]
Last saved by user
2:58 PM Friday, September 30, 2...

Gubuntu [Running] - Oracle VM VirtualBox

File Machine View Input Devices Help

Oct 1 01:25

Install

Installation type

This computer currently has no detected operating systems. What would you like to do?

○ Erase disk and install Ubuntu
Warning: This will delete all your programs, documents, photos, music, and any other files in all operating systems.

Advanced features... None selected

Write the changes to disks?

If you continue, the changes listed below will be written to the disks. Otherwise, you will be able to make further changes manually.

The partition tables of the following devices are changed:
SCSI3 (0,0,0) (sda)

The following partitions are going to be formatted:
partition #1 of SCSI3 (0,0,0) (sda) as
partition #5 of SCSI3 (0,0,0) (sda) as ext4

Go Back   Continue

After completion of installation Oracle VM , install Java , Hadoop and the script files Download hadoop
Download Java Extract hadoop
Java installation steps

sudo apt-get install openjdk-11-jre sudo apt-get install
openjdk-11-jdk java - version

```
hduser@cs570bigdata: ~
hduser@cs570bigdata:~$ java -version
openjdk version "11.0.16" 2022-07-19
OpenJDK Runtime Environment (build 11.0.16+8-post-Ubuntu-0ubuntu12
OpenJDK 64-Bit Server VM (build 11.0.16+8-post-Ubuntu-0ubuntu120.0
 sharing)
hduser@cs570bigdata:~$
```

Setup hadoop user for Hadoop Installation

sudo addgroup hadoop
sudo adduser --ingroup hadoop hduser sudo su

hduser

For this I used hadoop version hadoop-2.10.2.tar.gz Sudo tar
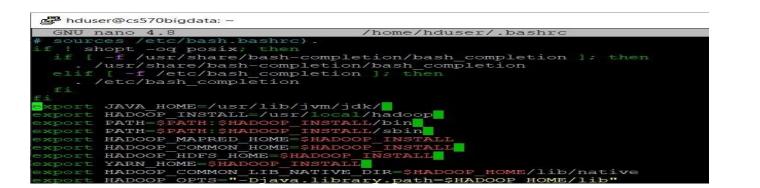
xzf hadoop-2.10.2.tar.gz –C /usr/local Cd / usr/local

Sudo mv hadoop-2.10.2 hadoop

Sudo chown –R hduser:hadoop hadoop

```
hduser@cs570bigdata:~$ sudo nano ~/.bashrc
[sudo] password for hduser:
```

```
  GNU nano 4.8                    /home/hduser/.bashrc
# ~/.bashrc: executed by bash(1) for non-login shells.
# see /usr/share/doc/bash/examples/startup-files (in the package b
# for examples

# If not running interactively, don't do anything
case $- in
    *i*) ;;
      *) return;;
esac

# don't put duplicate lines or lines starting with space in the hi
# See bash(1) for more options
HISTCONTROL=ignoreboth

# append to the history file, don't overwrite it
shopt -s histappend
# for setting history length see HISTSIZE and HISTFILESIZE in bash
```

Move to the end and add following lines for hadoop

```
 GNU nano 4.8                          /home/hduser/.bashrc
# sources /etc/bash.bashrc).
if ! shopt -oq posix; then
  if [ -f /usr/share/bash-completion/bash_completion ]; then
    . /usr/share/bash-completion/bash_completion
  elif [ -f /etc/bash_completion ]; then
    . /etc/bash_completion
  fi
fi
export JAVA_HOME=/usr/lib/jvm/jdk/
export HADOOP_INSTALL=/usr/local/hadoop
export PATH=$PATH:$HADOOP_INSTALL/bin
export PATH=$PATH:$HADOOP_INSTALL/sbin
export HADOOP_MAPRED_HOME=$HADOOP_INSTALL
export HADOOP_COMMON_HOME=$HADOOP_INSTALL
export HADOOP_HDFS_HOME=$HADOOP_INSTALL
export YARN_HOME=$HADOOP_INSTALL
export HADOOP_COMMON_LIB_NATIVE_DIR=$HADOOP_HOME/lib/native
export HADOOP_OPTS="-Djava.library.path=$HADOOP_HOME/lib"
```

To save use source ~/.bashrc
systemctl reboot –i

Now do the following setting for hduser


```
hduser@cs570bigdata:~$ sudo nano ~/.bashrc
[sudo] password for hduser:
hduser@cs570bigdata:~$ sudo nano /usr/local/hadoop/etc/hadoop/hadoop-env.sh
```

export JAVA_HOME=/usr/lib/jvm/jdk

Ssh generation and creation of authorized keys from public ssh keys

Now we need following files to be set for hadoop :

2. sudo nano /usr/local/hadoop/etc/hadoop/yarn-site.xml

3. sudo nano /usr/local/hadoop/etc/hadoop/mapred-site.xml

4 For doing hdfs.xml

First complete this task's mkdir –p

Create following test data file for hdfs

```
hduser@cs570bigdata: ~/Desktop/inputdata
hduser@cs570bigdata:~/Desktop/inputdata$ cat test.txt
hello world world hello hello
hduser@cs570bigdata:~/Desktop/inputdata$ ▉
```

Create directory as mentioned

```
hduser@cs570bigdata:/usr/local/hadoop/sbin
hduser@cs570bigdata:/usr/local/hadoop$ cd hadoop
-bash: cd: hadoop: No such file or directory
hduser@cs570bigdata:/usr/local/hadoop$ ls
bin   etc   include   lib   libexec   LICENSE.txt   logs   NOTICE.txt   README.txt   sbin   share
hduser@cs570bigdata:/usr/local/hadoop$ cd hadoop
-bash: cd: hadoop: No such file or directory
hduser@cs570bigdata:/usr/local/hadoop$ cd sbin
hduser@cs570bigdata:/usr/local/hadoop/sbin$ hdfs dfs -mkdir /user
```

Now copy to file created for input
/usr/local/hadoop/bin/hdfs dfs -put '/home/hduser/Desktop/input data' /user Next final step

for hadoop

hdfs namenode –format

now everything is set we can start hadoop and run the jar for the word count and get the output as shown
in below screen

now start hadoop as shown below by running start-dfs.sh start- yarn.sh and jps
to check the status as shown in below screens

```
hduser@cs570bigdata:/usr/local/hadoop/sbin$ start-dfs.sh
WARNING: An illegal reflective access operation has occurred
WARNING: Illegal reflective access by org.apache.hadoop.security.authentication.util.KerberosUtil (file:/usr/local/hadoop/share/hadoop/common/lib/hadoop-
ity.krb5.Config.getInstance()
WARNING: Please consider reporting this to the maintainers of org.apache.hadoop.security.authentication.util.KerberosUtil
WARNING: Use --illegal-access=warn to enable warnings of further illegal reflective access operations
WARNING: All illegal access operations will be denied in a future release
Starting namenodes on [cs570bigdata]
cs570bigdata: starting namenode, logging to /usr/local/hadoop/logs/hadoop-hduser-namenode-cs570bigdata.out
localhost: starting datanode, logging to /usr/local/hadoop/logs/hadoop-hduser-datanode-cs570bigdata.out
Starting secondary namenodes [0.0.0.0]
0.0.0.0: starting secondarynamenode, logging to /usr/local/hadoop/logs/hadoop-hduser-secondarynamenode-cs570bigdata.out
WARNING: An illegal reflective access operation has occurred
WARNING: Illegal reflective access by org.apache.hadoop.security.authentication.util.KerberosUtil (file:/usr/local/hadoop/share/hadoop/common/lib/hadoop-
ity.krb5.Config.getInstance()
WARNING: Please consider reporting this to the maintainers of org.apache.hadoop.security.authentication.util.KerberosUtil
WARNING: Use --illegal-access=warn to enable warnings of further illegal reflective access operations
WARNING: All illegal access operations will be denied in a future release
hduser@cs570bigdata:/usr/local/hadoop/sbin$
```

```
hduser@cs570bigdata:/usr/local/hadoop/sbin$ start-yarn.sh
starting yarn daemons
starting resourcemanager, logging to /usr/local/hadoop/logs/yarn-hduser-resourcemanager-cs570bigdata.out
localhost: starting nodemanager, logging to /usr/local/hadoop/logs/yarn-hduser-nodemanager-cs570bigdata.out
hduser@cs570bigdata:/usr/local/hadoop/sbin$ jps
3347 NodeManager
2998 SecondaryNameNode
3280 ResourceManager
3660 Jps
2638 NameNode
2782 DataNode
hduser@cs570bigdata:/usr/local/hadoop/sbin$
```

Sample program to generate input file for MapReduce Pi:

```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 PiValue {
  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
```

```java
        int sum = 0;
                for (IntWritable val : values) {  sum +=
                        val.get();
                }
                context.write(key, new IntWritable(sum));

    }
}

public static void main(String[] args) throws Exception {  Configuration
  conf = new Configuration();
  Job job = new Job(conf, "PiValue");
  job.setJarByClass(PiValue.class);  job.setOutputKeyClass(Text.class);
  job.setOutputValueClass(IntWritable.class);
  job.setMapperClass(Map.class);
  job.setCombinerClass(Reduce.class);
  job.setReducerClass(Reduce.class);
  job.setInputFormatClass(TextInputFormat.class);
  job.setOutputFormatClass(TextOutputFormat.class);
  job.setNumReduceTasks(1);

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

  job.waitForCompletion(true);
```

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

hduser@cs570bigdata:/usr/local/hadoop$ bin/hadoop jar /home/hduser/GPi/pivalueproj.jar PiValuepro /user/hduser/inputdata outputwc

Hadoop running Pi Jar file command:

hduser@cs570bigdata:/usr/local/hadoop$ bin/hadoop jar /home/hduser/GPi/pivalueproj.jar
PiValuepro /user/hduser/inputdata outputwc


```
hduser@cs570bigdata: /usr/local/hadoop                                                                              —    □    ×
hduser@cs570bigdata:/usr/local/hadoop$ bin/hadoop jar /home/hduser/GPi/pivalueproj.jar PiValuepro /user/hduser/inputdata outputwc
WARNING: An illegal reflective access operation has occurred
WARNING: Illegal reflective access by org.apache.hadoop.security.authentication.util.KerberosUtil (file:/usr/local/hadoop/share/hadoop/c
ommon/lib/hadoop-auth-2.10.2.jar) to method sun.security.krb5.Config.getInstance()
WARNING: Please consider reporting this to the maintainers of org.apache.hadoop.security.authentication.util.KerberosUtil
WARNING: Use --illegal-access=warn to enable warnings of further illegal reflective access operations
WARNING: All illegal access operations will be denied in a future release
22/10/10 00:52:06 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
22/10/10 00:52:13 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and
 execute your application with ToolRunner to remedy this.
22/10/10 00:52:17 INFO input.FileInputFormat: Total input files to process : 1
22/10/10 00:52:18 INFO mapreduce.JobSubmitter: number of splits:1
22/10/10 00:52:27 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1665330928624_0001
22/10/10 00:52:33 INFO conf.Configuration: resource-types.xml not found
22/10/10 00:52:33 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
22/10/10 00:52:33 INFO resource.ResourceUtils: Adding resource type - name = memory-mb, units = Mi, type = COUNTABLE
22/10/10 00:52:33 INFO resource.ResourceUtils: Adding resource type - name = vcores, units = , type = COUNTABLE
22/10/10 00:52:40 INFO impl.YarnClientImpl: Submitted application application_1665330928624_0001
22/10/10 00:52:42 INFO mapreduce.Job: The url to track the job: http://cs570bigdata:8088/proxy/application_1665330928624_0001/
22/10/10 00:52:42 INFO mapreduce.Job: Running job: job_1665330928624_0001
22/10/10 00:54:14 INFO mapreduce.Job: Job job_1665330928624_0001 running in uber mode : false
22/10/10 00:54:14 INFO mapreduce.Job:  map 0% reduce 0%
22/10/10 00:55:06 INFO mapreduce.Job:  map 100% reduce 0%
22/10/10 00:55:34 INFO mapreduce.Job:  map 100% reduce 100%
22/10/10 00:55:36 INFO mapreduce.Job: Job job_1665330928624_0001 completed successfully
22/10/10 00:55:37 INFO mapreduce.Job: Counters: 49
        File System Counters
                FILE: Number of bytes read=54
                FILE: Number of bytes written=421529
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=150
                HDFS: Number of bytes written=28
                HDFS: Number of read operations=6
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=2
        Job Counters
                Launched map tasks=1
                Launched reduce tasks=1
                Data-local map tasks=1
                Total time spent by all maps in occupied slots (ms)=44660
                Total time spent by all reduces in occupied slots (ms)=24845
                Total time spent by all map tasks (ms)=44660
                Total time spent by all reduce tasks (ms)=24845
                Total vcore-milliseconds taken by all map tasks=44660
                Total vcore-milliseconds taken by all reduce tasks=24845
                Total megabyte-milliseconds taken by all map tasks=45731840
                Total megabyte-milliseconds taken by all reduce tasks=25441280
        Map-Reduce Framework
```

Previous slide output continued :

hduser@cs570bigdata:/usr/local/hadoop$ bin/hadoop jar /home/hduser/GPi/pivalueproj.jar
PiValuepro /user/hduser/inputdata outputwc

hduser@cs570bigdata:/usr/local/hadoop$ bin/hadoop jar /home/hduser/GPi/pivalueproj.jar PiValuepro /user/hduser/inputdata outputwc

```
                HDFS: Number of read operations=6
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=2
        Job Counters
                Launched map tasks=1
                Launched reduce tasks=1
                Data-local map tasks=1
                Total time spent by all maps in occupied slots (ms)=44660
                Total time spent by all reduces in occupied slots (ms)=24845
                Total time spent by all map tasks (ms)=44660
                Total time spent by all reduce tasks (ms)=24845
                Total vcore-milliseconds taken by all map tasks=44660
                Total vcore-milliseconds taken by all reduce tasks=24845
                Total megabyte-milliseconds taken by all map tasks=45731840
                Total megabyte-milliseconds taken by all reduce tasks=25441280
        Map-Reduce Framework
                Map input records=8
                Map output records=8
                Map output bytes=62
                Map output materialized bytes=54
                Input split bytes=120
                Combine input records=8
                Combine output records=5
                Reduce input groups=5
                Reduce shuffle bytes=54
                Reduce input records=5
                Reduce output records=5
                Spilled Records=10
                Shuffled Maps =1
                Failed Shuffles=0
                Merged Map outputs=1
                GC time elapsed (ms)=767
                CPU time spent (ms)=12160
                Physical memory (bytes) snapshot=432816128
                Virtual memory (bytes) snapshot=4139466752
                Total committed heap usage (bytes)=211812352
        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=30
        File Output Format Counters
                Bytes Written=28
1,1     2
2       1
Inside:1,1      2, Outside:2    1
```

Conclusion

Application usage of MapReduce - PI , this lessons learned

Can be applied in many application design for example in study of

Astronomy to find the details of the location of planets , meteors.

Reference :

SFBU course materials