

Week 2

Felipe Rodriguez

Bellevue University

DSC 650 Big Data

Professor Nasheb Ismaily

December 10, 2023

HDFS Report

```
Terminal Shell Edit View Window Help
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/program/hadoop/share/hadoop/common/lib/slf4j-log4j12-1.7.25.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/program/tez/lib/slf4j-log4j12-1.7.10.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/program/hive/lib/log4j-slf4j-impl-2.10.0.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
2023-12-08 22:59:25.450 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Configured Capacity: 103670202368 (96.55 GB)
Present Capacity: 79750917200 (74.27 GB)
DFS Remaining: 79370712969 (73.92 GB)
DFS Used: 380204231 (362.59 MB)
DFS Used%: 0.48%
Replicated Blocks:
  Under replicated blocks: 0
  Blocks with corrupt replicas: 0
  Missing blocks: 0
  Missing blocks (with replication factor 1): 0
  Low redundancy blocks with highest priority to recover: 0
  Pending deletion blocks: 0
Erasure Coded Block Groups:
  Low redundancy block groups: 0
  Block groups with corrupt internal blocks: 0
  Missing block groups: 0
  Low redundancy blocks with highest priority to recover: 0
  Pending deletion blocks: 0

-----
Live datanodes (2):

Name: 172.28.1.2:9866 (worker1)
Hostname: worker1
Decommission Status : Normal
Configured Capacity: 51835101184 (48.28 GB)
DFS Used: 191703152 (182.82 MB)
Non DFS Used: 11672844176 (10.87 GB)
DFS Remaining: 39685343555 (36.96 GB)
DFS Used%: 0.37%
DFS Remaining%: 76.56%
Configured Cache Capacity: 0 (0 B)
Cache Used: 0 (0 B)
Cache Remaining: 0 (0 B)
Cache Used%: 100.00%
Cache Remaining%: 0.00%
Xceivers: 5
Last contact: Fri Dec 08 22:59:26 GMT 2023
Last Block Report: Fri Dec 08 22:53:17 GMT 2023
Num of Blocks: 142

Name: 172.28.1.3:9866 (worker2)
Hostname: worker2
Decommission Status : Normal
Configured Capacity: 51835101184 (48.28 GB)
DFS Used: 188501079 (179.77 MB)
Non DFS Used: 11676046249 (10.87 GB)
DFS Remaining: 39685369414 (36.96 GB)
DFS Used%: 0.36%
DFS Remaining%: 76.56%
Configured Cache Capacity: 0 (0 B)
Cache Used: 0 (0 B)
Cache Remaining: 0 (0 B)
Cache Used%: 100.00%
Cache Remaining%: 0.00%
Xceivers: 5
Last contact: Fri Dec 08 22:59:26 GMT 2023
Last Block Report: Fri Dec 08 22:53:17 GMT 2023
Num of Blocks: 134
```

Loaded Data

```
bash-5.0# hdfs dfs -put /data/grades.csv /
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/program/hadoop/share/hadoop/common/lib/slf4j-log4j12-1.7.25.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/program/tez/lib/slf4j-log4j12-1.7.10.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/program/hive/lib/log4j-slf4j-impl-2.10.0.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
2023-12-08 23:03:20,522 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
bash-5.0# hdfs dfs -ls /
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/program/hadoop/share/hadoop/common/lib/slf4j-log4j12-1.7.25.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/program/tez/lib/slf4j-log4j12-1.7.10.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/program/hive/lib/log4j-slf4j-impl-2.10.0.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
2023-12-08 23:04:04,295 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 7 items
-rw-r--r-- 1 root supergroup 747 2023-12-08 23:03 /grades.csv
drwxr-xr-x - root supergroup 0 2023-12-08 22:55 /hbase
drwxr-xr-x - root supergroup 0 2023-12-08 22:54 /log
drwxr-xr-x - root supergroup 0 2023-12-08 22:55 /spark-jars
drwxr-xr-x - root supergroup 0 2023-12-08 22:55 /tez
drwxr-xr-x - root supergroup 0 2023-12-08 22:53 /tmp
drwxrwx--- - root supergroup 0 2023-12-08 22:55 /user
bash-5.0#
```

Loaded Data Worker 1

```
[feliperodriguez@bigdata-week1:~/dsc650-infra/bellevue-bigdata/hadoop-hive-spark-hbase]$ docker-compose exec worker1 bash
bash-5.0# hdfs dfs -ls /
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/program/hadoop/share/hadoop/common/lib/slf4j-log4j12-1.7.25.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/program/tez/lib/slf4j-log4j12-1.7.10.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/program/hive/lib/log4j-slf4j-impl-2.10.0.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
2023-12-08 23:10:29,107 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 7 items
-rw-r--r-- 1 root supergroup 747 2023-12-08 23:03 /grades.csv
drwxr-xr-x - root supergroup 0 2023-12-08 22:55 /hbase
drwxr-xr-x - root supergroup 0 2023-12-08 22:54 /log
drwxr-xr-x - root supergroup 0 2023-12-08 22:55 /spark-jars
drwxr-xr-x - root supergroup 0 2023-12-08 22:55 /tez
drwxr-xr-x - root supergroup 0 2023-12-08 22:53 /tmp
drwxrwx--- - root supergroup 0 2023-12-08 22:55 /user
```

Loaded Data Worker 2

```
[feliperodriguez@bigdata-week1:~/dsc650-infra/bellevue-bigdata/hadoop-hive-spark-hbase$ docker-compose exec worker2 bash
bash-5.0# hdfs dfs -ls /
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/program/hadoop/share/hadoop/common/lib/slf4j-log4j12-1.7.25.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/program/tez/lib/slf4j-log4j12-1.7.10.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/program/hive/lib/log4j-slf4j-impl-2.10.0.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
2023-12-08 23:13:04,254 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 7 items
-rw-r--r-- 1 root supergroup 747 2023-12-08 23:03 /grades.csv
drwxr-xr-x - root supergroup 0 2023-12-08 22:55 /hbase
drwxr-xr-x - root supergroup 0 2023-12-08 22:54 /log
drwxr-xr-x - root supergroup 0 2023-12-08 22:55 /spark-jars
drwxr-xr-x - root supergroup 0 2023-12-08 22:55 /tez
drwxr-xr-x - root supergroup 0 2023-12-08 22:53 /tmp
drwxrwx--- - root supergroup 0 2023-12-08 22:55 /user
```

HDFS Command 1 – Count

```
bash-5.0# hdfs dfs -count /
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/program/hadoop/share/hadoop/common/lib/slf4j-log4j12-1.7.25.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/program/tez/lib/slf4j-log4j12-1.7.10.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/program/hive/lib/log4j-slf4j-impl-2.10.0.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
2023-12-08 23:19:15,463 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
72 289 377268168 /
bash-5.0#
```

HDFS Command 2 – Mkdir

```
bash-5.0# hdfs dfs -mkdir /test
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/program/hadoop/share/hadoop/common/lib/slf4j-log4j12-1.7.25.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/program/tez/lib/slf4j-log4j12-1.7.10.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/program/hive/lib/log4j-slf4j-impl-2.10.0.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
2023-12-09 00:01:03,772 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
bash-5.0# hdfs dfs -ls /
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/program/hadoop/share/hadoop/common/lib/slf4j-log4j12-1.7.25.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/program/tez/lib/slf4j-log4j12-1.7.10.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/program/hive/lib/log4j-slf4j-impl-2.10.0.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
2023-12-09 00:01:15,330 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 8 items
-rw-r--r--  1 root supergroup      747 2023-12-08 23:03 /grades.csv
drwxr-xr-x  - root supergroup        0 2023-12-08 22:55 /hbase
drwxr-xr-x  - root supergroup        0 2023-12-08 22:54 /log
drwxr-xr-x  - root supergroup        0 2023-12-08 22:55 /spark-jars
drwxr-xr-x  - root supergroup        0 2023-12-09 00:01 /test
drwxr-xr-x  - root supergroup        0 2023-12-08 22:55 /tez
drwxr-xr-x  - root supergroup        0 2023-12-08 22:53 /tmp
drwxrwx---  - root supergroup        0 2023-12-08 22:55 /user
bash-5.0#
```

HDFS Command 3 – Stat

```
bash-5.0# hdfs dfs -stat "%b %F" /grades.csv
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/program/hadoop/share/hadoop/common/lib/slf4j-log4j12-1.7.25.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/program/tez/lib/slf4j-log4j12-1.7.10.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/program/hive/lib/log4j-slf4j-impl-2.10.0.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
2023-12-10 22:02:19,143 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
747 regular file
bash-5.0#
```

Yarn Node –List

```
bash-5.0# yarn node -list
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/program/hadoop/share/hadoop/common/lib/slf4j-log4j12-1.7.25.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/program/tez/lib/slf4j-log4j12-1.7.10.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/program/hive/lib/log4j-slf4j-impl-2.10.0.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
2023-12-09 01:26:23,797 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
2023-12-09 01:26:23,897 INFO client.RMProxy: Connecting to ResourceManager at master/172.28.1.1:8032
Total Nodes:2
  Node-Id          Node-State Node-Http-Address      Number-of-Running-Containers
  worker1:45583    RUNNING   worker1:8042           0
  worker2:33977    RUNNING   worker2:8042           0
bash-5.0#
```

Maximum Capacity Increased

Safari File Edit View History Bookmarks Develop Window Help

localhost

Apple Google The Weather Channel Rent Blackboard Learn MCC MAYA Playground - OpenAI API Mushroom Cla...ion | Kaggle Home | Fstopshop

https://learn-us-e... Namemode infor... All Applications Spark Master at s... History Server VM instances - C... Visualize This: Th... Casino | Hard Ro...

Logged in as: dr.who

hadoop All Applications

Cluster

About

Nodes

Node Labels

Applications

NEW

NEW SAVING

SUBMITTED

ACCEPTED

RUNNING

FINISHED

FAILED

KILLED

Scheduler

Tools

Cluster Metrics

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Used Resources	Total Resources	Reserved Resources	Physical Mem Used %	Physical VCores Used %
0	0	0	0	0	<memory:0 B, vCores:0>	<memory:0 B, vCores:0>	<memory:0 B, vCores:0>	0	0

Cluster Nodes Metrics

Active Nodes	Decommissioning Nodes	Decommissioned Nodes	Lost Nodes	Unhealthy Nodes	Rebooted Nodes	Shutdown Nodes
0	0	0	0	0	0	0

User Metrics for dr.who

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Containers Pending	Containers Reserved	Memory Used	Memory Pending	Memory Reserved	VCores Used	VCores Pending	VCores Reserved
0	0	0	0	0	0	0	0 B	0 B	0 B	0	0	0

Scheduler Metrics

Scheduler Type	Scheduling Resource Type	Minimum Allocation	Maximum Allocation	Maximum Cluster Application Priority	Scheduler Busy %
Fair Scheduler	[memory-mb (unit=M), vcores]	<memory:512, vCores:1>	<memory:2048, vCores:4>	0	0

Show 20 entries

ID

User

Name

Application Type

Queue

Application Priority

StartTime

LaunchTime

FinishTime

State

FinalStatus

Running Containers

Allocated CPU VCores

Allocated Memory MB

Allocated GPUs

Reserved CPU VCores

Reserved Memory MB

Reserved GPUs

% of Queue

% of Cluster

Progress

Tracking UI

Blacklisted Nodes

No data available in table

Showing 0 to 0 of 0 entries

First Previous Next Last

MapReduce Results

```
SLF4J: Class path contains multiple SLF4J bindings.  
SLF4J: Found binding in [jarfile:/usr/program/hadoop/share/hadoop/common/lib/slf4j-log4j-1.7.25.jar/org/slf4j/impl/StaticLoggerBinder.class]  
SLF4J: Found binding in [jarfile:/usr/program/tez/lib/slf4j-log4j-1.7.8.jar/org/slf4j/impl/StaticLoggerBinder.class]  
SLF4J: Found binding in [jarfile:/usr/program/hive/lib/log4j-slf4j-impl-2.10.6.jar/org/slf4j/impl/StaticLoggerBinder.class]  
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.  
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]  
2023-12-10 19:44:17,154 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable  
Number of Maps = 2  
Samples per Map = 10  
Wrote input for Map #0  
Wrote input for Map #1  
Starting Job  
2023-12-10 18:44:18,635 INFO client.RMProxy: Connecting to ResourceManager at master/172.28.1.1:8032  
2023-12-10 18:44:19,870 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/root/.staging/job_1702233404743_0001  
2023-12-10 18:44:24,093 INFO Input.FileInputFormat: Total input files to process : 2  
2023-12-10 18:44:24,156 INFO mapreduce.JobSubmitter: number of splits=2  
2023-12-10 18:44:24,353 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1702233404743_0001  
2023-12-10 18:44:24,353 INFO mapreduce.JobSubmitter: Executing with tokens: []  
2023-12-10 18:44:24,623 INFO conf.Configuration: resource-types.xml not found  
2023-12-10 18:44:24,624 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.  
2023-12-10 18:44:24,636 WARN mapped.YARNRunner: Configuration yarn.app.mapreduce.am.resource.memory-mb=512Mi is overriding the yarn.app.mapreduce.am.resource.mb=1536 configuration  
2023-12-10 18:44:25,185 INFO impl.YarnClientImpl: Submitted application application_1702233404743_0001  
2023-12-10 18:44:25,269 INFO mapreduce.Job: The url to track the job: http://master:8080/proxy/application_1702233404743_0001/  
2023-12-10 18:44:25,270 INFO mapreduce.Job: Running job: job_1702233404743_0001  
2023-12-10 18:44:36,414 INFO mapreduce.Job: Job job_1702233404743_0001 running in uber mode : false  
2023-12-10 18:44:36,416 INFO mapreduce.Job: map 0% reduce 0%  
2023-12-10 18:44:44,510 INFO mapreduce.Job: map 100% reduce 0%  
2023-12-10 18:44:49,540 INFO mapreduce.Job: map 100% reduce 100%  
2023-12-10 18:44:50,554 INFO mapreduce.Job: Job job_1702233404743_0001 completed successfully  
2023-12-10 18:44:50,716 INFO mapreduce.Job: Counters: 54  
  
File System Counters  
FILE: Number of bytes read=58  
FILE: Number of bytes written=738268  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=522  
HDFS: Number of bytes written=215  
HDFS: Number of read operations=3  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=3  
HDFS: Number of bytes read erasure-coded=0  
  
Job Counters  
Launched map tasks=2  
Launched reduce tasks=1  
Data-local map tasks=2  
Total time spent by all maps in occupied slots (ms)=18758  
Total time spent by all reduces in occupied slots (ms)=12868  
Total time spent by all map tasks (ms)=9379  
Total time spent by all reduce tasks (ms)=3217  
Total vcore-milliseconds taken by all map tasks=9379  
Total vcore-milliseconds taken by all reduce tasks=3217  
Total megabyte-milliseconds taken by all map tasks=664006  
Total megabyte-milliseconds taken by all reduce tasks=6588416  
  
Map-Reduce Framework  
Map input records=2  
Map output records=4  
Map output bytes=36  
Map output materialized bytes=56  
Input split bytes=286  
Combine input records=0  
Combine output records=0  
Reduce input groups=2  
Reduce shuffle bytes=56  
Reduce input records=4  
Reduce output records=0  
Spilled Records=0  
Shuffled Maps=2  
Failed Shuffles=0  
Merged Map outputs=2  
GC time elapsed (ms)=257  
CPU time spent (ms)=1890  
Physical memory (bytes) snapshot=758051328  
Virtual memory (bytes) snapshot=768979712  
Total committed heap usage (bytes)=784118784  
Peak Map Physical memory (bytes)=285916064  
Peak Map Virtual memory (bytes)=2355364496  
Peak Reduce Physical memory (bytes)=180933088  
Peak Reduce Virtual memory (bytes)=2355864832  
  
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=236  
File Output Format Counters  
Bytes Written=97  
  
Job Finished in 32.221 seconds  
Estimated value of Pi is 3.800000000000000000000000  
bash-5.0#
```


Summary of Results

The method used for calculating π in this assignment is the Monte Carlo Method. This method takes the number of samples, in this case 10, randomly scatter points inside a square and counts the number of points that fall inside a circle inscribed in the square. This number counted is the estimate of π . π value is 3.1415~, however in the test ran, the π value is 3.8. This is likely because of the small sample size. Because the Monte Carlo method is looking at the amount of scatter points within the inscribed circle, the more samples that are observed, the closer the value will be to π .