

Monica Santana – Week 2 Exercise

HDFS Report

```
santanamonica@dscbigdata: ~/dsc650-infra/bellevue-bigdata/hadoop-hive-spark-hbase
DFS Used: 0.47%
Replicated Blocks:
  Under replicated blocks: 0
  Blocks with corrupt replican: 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: 164003840 (156.41 MB)
Non DFS Used: 10963152096 (10.21 GB)
DFS Remaining: 40422732964 (37.65 GB)
DFS Used: 0.32%
DFS Remaining: 77.98%
Configured Cache Capacity: 0 (0 B)
Cache Used: 0 (0 B)
Cache Remaining: 0 (0 B)
Cache Used: 100.00%
Cache Remaining: 0.00%
Xceiver: 5
Last contact: Sat Jun 15 05:49:37 GMT 2024
Last Block Report: Sat Jun 15 04:51:55 GMT 2024
Num of Blocks: 146

Name: 172.28.1.3:9866 (worker2)
Hostname: worker2
Decommission Status : Normal
Configured Capacity: 51835101184 (48.28 GB)
DFS Used: 217731072 (207.64 MB)
Non DFS Used: 10909425664 (10.16 GB)
DFS Remaining: 40422734085 (37.65 GB)
DFS Used: 0.42%
DFS Remaining: 77.98%
Configured Cache Capacity: 0 (0 B)
Cache Used: 0 (0 B)
Cache Remaining: 0 (0 B)
Cache Used: 100.00%
Cache Remaining: 0.00%
Xceiver: 5
Last contact: Sat Jun 15 05:49:37 GMT 2024
Last Block Report: Sat Jun 15 04:51:55 GMT 2024
Num of Blocks: 138

bash-5.0#
```

Load grades.csv into HDFS

```
santanamonica@dscbigdata: ~/dsc650-infra/bellevue-bigdata/hadoop-hive-spark-hbase
Xceiver: 5
Last contact: Sat Jun 15 05:49:37 GMT 2024
Last Block Report: Sat Jun 15 04:51:55 GMT 2024
Num of Blocks: 146

Name: 172.28.1.3:9866 (worker2)
Hostname: worker2
Decommission Status : Normal
Configured Capacity: 51835101184 (48.28 GB)
DFS Used: 217731072 (207.64 MB)
Non DFS Used: 10909425664 (10.16 GB)
DFS Remaining: 40422734085 (37.65 GB)
DFS Used: 0.42%
DFS Remaining: 77.98%
Configured Cache Capacity: 0 (0 B)
Cache Used: 0 (0 B)
Cache Remaining: 0 (0 B)
Cache Used: 100.00%
Cache Remaining: 0.00%
Xceiver: 5
Last contact: Sat Jun 15 05:49:37 GMT 2024
Last Block Report: Sat Jun 15 04:51:55 GMT 2024
Num of Blocks: 138

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]
2024-06-15 05:51:47,889 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]
2024-06-15 05:52:16,743 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 2024-06-15 05:51 /grades.csv
drwxr-xr-x - root supergroup 0 2024-06-15 04:55 /hbase
drwxr-xr-x - root supergroup 0 2024-06-15 04:53 /log
drwxr-xr-x - root supergroup 0 2024-06-15 04:54 /spark-jars
drwxr-xr-x - root supergroup 0 2024-06-15 04:54 /tez
drwxr-xr-x - root supergroup 0 2024-06-15 04:52 /tmp
drwxrwx--- - root supergroup 0 2024-06-15 04:53 /user

bash-5.0#
```

SSH into worker1 and worker2 nodes

```
santanamonica@dscbigdata: ~/dsc650-infra/bellevue-bigdata/hadoop-hive-spark-hbase
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
2024-06-16 05:35:25,446 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
put: /grades.csv: File exists
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]
2024-06-16 05:36:14,850 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 2024-06-15 05:51 /grades.csv
drwxr-xr-x - root supergroup          0 2024-06-16 05:34 /hbase
drwxr-xr-x - root supergroup          0 2024-06-15 04:53 /log
drwxr-xr-x - root supergroup          0 2024-06-15 04:54 /spark-jars
drwxr-xr-x - root supergroup          0 2024-06-15 04:54 /tez
drwxr-xr-x - root supergroup          0 2024-06-16 05:35 /tmp
drwxrwx-- - root supergroup          0 2024-06-15 04:53 /user
bash-5.0# exit
exit
santanamonica@dscbigdata:~/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]
2024-06-16 05:37:43,170 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 2024-06-15 05:51 /grades.csv
drwxr-xr-x - root supergroup          0 2024-06-16 05:34 /hbase
drwxr-xr-x - root supergroup          0 2024-06-15 04:53 /log
drwxr-xr-x - root supergroup          0 2024-06-15 04:54 /spark-jars
drwxr-xr-x - root supergroup          0 2024-06-15 04:54 /tez
drwxr-xr-x - root supergroup          0 2024-06-16 05:35 /tmp
drwxrwx-- - root supergroup          0 2024-06-15 04:53 /user
bash-5.0# exit
exit
santanamonica@dscbigdata:~/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]
2024-06-16 05:38:15,423 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 2024-06-15 05:51 /grades.csv
drwxr-xr-x - root supergroup          0 2024-06-16 05:34 /hbase
drwxr-xr-x - root supergroup          0 2024-06-15 04:53 /log
drwxr-xr-x - root supergroup          0 2024-06-15 04:54 /spark-jars
drwxr-xr-x - root supergroup          0 2024-06-15 04:54 /tez
drwxr-xr-x - root supergroup          0 2024-06-16 05:35 /tmp
drwxrwx-- - root supergroup          0 2024-06-15 04:53 /user
bash-5.0# exit
exit
```

3 chosen HDFS commands – 1st one is the -cat command to print file contents

```
santanamonica@dscbigdata: ~/dsc650-infra/bellevue-bigdata/hadoop-hive-spark-hbase
bash-5.0# hdfs dfs -cat /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]
2024-06-16 06:06:04,567 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Last name,First name,SN,Test1,Test2,Test3,Test4,Final,Grade
Alfalfa,Xiyouhua,123-45-6789,40,90,100,83,49,B+
Alfred,University,123-12-1234,41,97,96,97,48,B+
Certy,Gramma,567-89-0123,41,80,60,40,44,C
Android,Electric,087-65-4321,42,23,36,45,47,B-
Bumkin,Fred,456-78-9012,43,78,86,77,45,A-
Bubble,Betty,234-56-7890,44,90,80,90,46,C-
Noshov,Cecil,345-67-8901,45,11,-1,4,43,F
Buff,Bif,632-79-9939,46,20,30,40,50,B+
Airpump,Andrew,223-45-6789,49,1,90,100,83,A
Backus,Jim,143-12-1234,48,1,97,96,97,A+
Carnivore,Art,565-89-0123,44,1,80,60,40,D+
Dandy,Tim,007-75-4321,47,1,23,36,46,C+
Elephant,Ira,456-71-9012,45,1,78,88,77,B-
Franklin,Benny,234-56-2890,50,1,90,80,90,B-
George,Boy,345-67-3901,40,1,11,-1,4,B
Hefclump,Harvey,632-79-9439,30,1,20,30,40,C
bash-5.0#
```

2nd command is `-stat` using different format sequences: `%a` = permissions in octal, `%A` = permissions in symbolic style, `%b` = file size in bytes

```
santanamonica@dscbigdata: ~/dsc650-infra/bellevue-bigdata/hadoop-hive-spark-hbase
necessary.

-usage [cmd ...] :
  Displays the usage for given command or all commands if none is specified.

Generic options supported are:
--conf <configuration file>      specify an application configuration file
--D <property value>            define a value for a given property
--fs <file:///hdfs://namenode:port> specify default filesystem URL; to use, override
  the 'fs.defaultFS' property from configurations.
--jt <local|resource-manager:port> specify a Resource Manager
--files <file1,...>              specify a comma-separated list of files to be
  copied to the map reduce cluster
--libjars <jar1,...>            specify a comma-separated list of jar files to
  be included in the classpath
--archives <archive1,...>       specify a comma-separated list of archives to
  be unarchived on the compute machines

The general command line syntax is:
command [genericOptions] [commandOptions]

bash-5.0# hdfs dfs -stat %a /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]
2024-06-16 19:53:32,663 WARN util.NativeCodeLoader: Unable to load native-hadoop
library for your platform... using builtin-java classes where applicable
644
bash-5.0# hdfs dfs -stat %A /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]
2024-06-16 19:53:33,195 WARN util.NativeCodeLoader: Unable to load native-hadoop
library for your platform... using builtin-java classes where applicable
rw-r--r--
bash-5.0# hdfs dfs -stat %b /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]
2024-06-16 19:56:32,012 WARN util.NativeCodeLoader: Unable to load native-hadoop
library for your platform... using builtin-java classes where applicable
747
bash-5.0#
```

3rd command is `-touchz` to create an empty file and `-rm` to delete it

```
santanamonica@dscbigdata: ~/dsc650-infra/bellevue-bigdata/hadoop-hive-spark-hbase
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]
2024-06-16 20:20:34,801 WARN util.NativeCodeLoader: Unable to load native-hadoop
library for your platform... using builtin-java classes where applicable
Found 9 items
-rw-r--r-- 1 root supergroup 747 2024-06-16 20:00 /grades.csv
drwxr-xr-x - root supergroup 0 2024-06-16 19:50 /hbase
drwxr-xr-x - root supergroup 0 2024-06-15 04:53 /log
drwxr-xr-x - root supergroup 0 2024-06-15 04:54 /spark-jars
drwxr-xr-x - root supergroup 0 2024-06-16 05:42 /test
drwxr-xr-x - root supergroup 0 2024-06-15 04:54 /tez
drwxr-xr-x - root supergroup 0 2024-06-16 08:35 /tmp
drwxrwxr-x - root supergroup 0 2024-06-15 04:53 /user
drwxr-xr-x - root supergroup 0 2024-06-16 20:18 /week2
bash-5.0# hdfs dfs -touchz /test/emptyfile.txt
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]
2024-06-16 20:21:03,373 WARN util.NativeCodeLoader: Unable to load native-hadoop
library for your platform... using builtin-java classes where applicable
bash-5.0# hdfs dfs -ls /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]
2024-06-16 20:21:19,529 WARN util.NativeCodeLoader: Unable to load native-hadoop
library for your platform... using builtin-java classes where applicable
Found 1 item
-rw-r--r-- 1 root supergroup 0 2024-06-16 20:21 /test/emptyfile.txt
bash-5.0# hdfs dfs -rm /test/emptyfile.txt
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]
2024-06-16 20:23:54,547 WARN util.NativeCodeLoader: Unable to load native-hadoop
library for your platform... using builtin-java classes where applicable
Deleted /test/emptyfile.txt
bash-5.0#
```

Inspecting YARN nodes

```

santanamonica@dscbigdata: ~/dsc650-infra/bellevue-bigdata/hadoop-hive-spark-hbase
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/sl
f4j-log4j12-1.7.25.jar!/org.slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/program/tez/lib/sl4j-log4j12-1.7.10.jar!
/org/it4j/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]
2024-06-16 20:25:48,317 WARN util.NativeCodeLoader: Unable to load native-hadoop
 library for your platform... using builtin-java classes where applicable
2024-06-16 20:29:48,995 INFO client.RMProxy: Connecting to ResourceManager at
ater/172.28.1.1:8032
Total Nodes: 2
      Node-Id      Node-State Node-Http-Address  Number-of-Runnin
g-Containers
worker2:35407      RUNNING    worker2:8042
      0
worker1:44807      RUNNING    worker1:8042
      0
bash-5.0#

```

YARN UI with updated maximum memory of 2048 mb

All Applications

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:8 GB, vCores:16>	<memory:0 B, vCores:0>	65	0

Cluster Nodes Metrics

Active Nodes	Decommissioning Nodes	Decommissioned Nodes	Lost Nodes	Unhealthy Nodes	Rebooted Nodes	Shutdown Nodes
2	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

Running the MapReduce Pi

```

santanamonica@dscbigdata:~/dsc650/infra/belloo-hive-spark-hbase$
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=520
HDFS: Number of bytes written=215
HDFS: Number of read operations=13
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)=23438
  Total time spent by all reduces in occupied slots (ms)=15352
  Total time spent by all map tasks (ms)=11719
  Total time spent by all reduce tasks (ms)=3838
  Total vcore-milliseconds taken by all map tasks=11719
  Total vcore-milliseconds taken by all reduce tasks=3838
  Total megabyte-milliseconds taken by all map tasks=12000256
  Total megabyte-milliseconds taken by all reduce tasks=7600224
Map-Reduce Framework
  Map input records=2
  Map output records=4
  Map output bytes=36
  Map output materialized bytes=56
  Input split bytes=284
  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=8
  Shuffled Maps=2
  Failed Shuffles=0
  Merged Map outputs=2
  GC time elapsed (ms)=266
  CPU time spent (ms)=1500
  Physical memory (bytes) snapshot=740081664
  Virtual memory (bytes) snapshot=7065341952
  Total committed heap usage (bytes)=620756992
  Peak Map Physical memory (bytes)=277319680
  Peak Map Virtual memory (bytes)=2354552832
  Peak Reduce Physical memory (bytes)=185462496
  Peak Reduce Virtual memory (bytes)=2359795712
Shuffle Errors
  BAD_ID=0
  CONNECTION=0
  IO_ERROR=0
  WRONG_LENGTH=0
  WRONG_MAP=0
  WRONG_REFS=0
File Input Format Counters
  Bytes Read=236
File Output Format Counters
  Bytes Written=97
Job Finished in 32.269 seconds
Estimated value of EI is 3.800000000000000000000000
bash-5.0#

```

The program calculated the pi to be 3.80 in 32.27 seconds. It uses the quasi-monte carlo method using the number of maps which is 2, and the number of sample of maps which is 10.

MapReduce uses the mapper to generate points in a unit square to count the points inside and outside the square. The reducer then accumulates the points inside and outside of the mapper.

For pi, it was calculated by this formula of 1 (side of the square) and the area of the circle/square would be pi/4 (sides of the square).

The area of the circle is divided by the area of the square which is equivalent to the number of points inside the circle divided by the number of points of both inside and outside of the circle.

Number of points inside the circle / number of points of both inside and outside of the circle = $\pi/4$ which is value of $\pi = 4(\text{number of points inside} / \text{total number inside \& outside})$.

I assume larger sample sizes would calculate the pi size to be more accurate to 3.14, this sample was small with only 10.