

Week 2 Assignment

Guruprasad Velikadu Krishnamoorthy

College of Science and Technology, Bellevue University

DSC650-T301: Big Data

Professor. Nasheb Ismaily

December 8, 2023

1. Screenshot of dfsadmin report command:

```
bash-5.0# hdfs dfsadmin -report
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/tee/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]
2023-12-03 16:02:14,503 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Configured Capacity: 10387020288 (96.65 GB)
Present Capacity: 7842317696 (74.03 GB)
DFS Remaining: 7808041536 (73.65 GB)
DFS Used: 381702144 (364.02 MB)
DFS Used%: 0.45%
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: 172.28.1.219866 (worker1)
Hostnames: worker1
Decommission Status: Normal
Configured Capacity: 8135101184 (40.28 GB)
DFS Used: 150925200 (143.48 MB)
Non DFS Used: 11559731008 (11.04 GB)
DFS Remaining: 39540323685 (36.82 GB)
DFS Used%: 0.29%
DFS Remaining%: 76.28%
Configured Cache Capacity: 0 (0 B)
Cache Used: 0 (0 B)
Cache Remaining: 0 (0 B)
Cache Used%: 100.00%
Cache Remaining%: 0.00%
Xorivers: 5
Last contact: Sun Dec 03 16:02:12 GMT 2023
Last Block Report: Sun Dec 03 15:34:10 GMT 2023
Num of Blocks: 130
Name: 172.28.1.319866 (worker2)
Hostnames: worker2
Decommission Status: Normal
Configured Capacity: 8135101184 (40.28 GB)
DFS Used: 230848944 (220.17 MB)
Non DFS Used: 11778780384 (10.97 GB)
DFS Remaining: 39540291845 (36.82 GB)
DFS Used%: 0.45%
DFS Remaining%: 76.28%
Configured Cache Capacity: 0 (0 B)
Cache Used: 0 (0 B)
Cache Remaining: 0 (0 B)
Cache Used%: 100.00%
```

2.1. Screenshot showing the grades.csv file is copied in Hadoop master node / directory:

```
bash-5.0# hdfs dfs -ls /
Found 7 items
drwxr-xr-x - root supergroup 0 2023-12-03 16:20 /grades.csv
drwxr-xr-x - root supergroup 0 2023-12-03 15:36 /logs
drwxr-xr-x - root supergroup 0 2023-12-03 15:35 /log
drwxr-xr-x - root supergroup 0 2023-12-03 15:35 /spark-jars
drwxr-xr-x - root supergroup 0 2023-12-03 15:36 /tmp
drwxr-xr-x - root supergroup 0 2023-12-03 15:34 /tmp
drwxr-xr-x - root supergroup 0 2023-12-03 15:36 /user
```

2.2. Screenshot showing the grades.csv file is present in worker1 node / directory:

```
PRISUN@PRISUNbigdata-ds650:/data50-infra/hellovue-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 [jarfile:/usr/program/hadoop/share/hadoop/common/lib/slf4j-log4j12-1.7.25.jar/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jarfile:/usr/program/tez/lib/slf4j-log4j12-1.7.10.jar/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jarfile:/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]
2023-12-03 16:31:14,514 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 2023-12-03 16:20 /grades.csv
drwxr-xr-x - root supergroup 0 2023-12-03 15:35 /log
drwxr-xr-x - root supergroup 0 2023-12-03 15:35 /spark-jars
drwxr-xr-x - root supergroup 0 2023-12-03 15:36 /tmp
drwxr-xr-x - root supergroup 0 2023-12-03 15:36 /user
bash-5.0#
```

2.3. Screenshot showing the grades.csv file is present in worker2 node / directory:

```
PRISUN@PRISUNbigdata-ds650:/data50-infra/hellovue-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 [jarfile:/usr/program/hadoop/share/hadoop/common/lib/slf4j-log4j12-1.7.25.jar/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jarfile:/usr/program/tez/lib/slf4j-log4j12-1.7.10.jar/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jarfile:/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]
2023-12-03 16:32:47,967 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 2023-12-03 16:20 /grades.csv
drwxr-xr-x - root supergroup 0 2023-12-03 15:35 /log
drwxr-xr-x - root supergroup 0 2023-12-03 15:35 /spark-jars
drwxr-xr-x - root supergroup 0 2023-12-03 15:36 /tmp
drwxr-xr-x - root supergroup 0 2023-12-03 15:36 /user
bash-5.0#
```

3.1. Screenshot of HDFS command 1 → touch & rm commands.

A new file named testfile1.txt was created in hdfs /tmp directory using touch command and removed using rm command.

```
bash-5.0#  
bash-5.0# pwd  
/usr/bin  
bash-5.0# hdfs dfs -touch /tmp/testfile1.txt  
touch: /usr/bin: cannot create directory: not a directory.  
SLF4J: Found binding in [jarfile:/usr/program/hadoop/share/hadoop/common/lib/slf4j-log4j12-1.7.25.jar/org/slf4j/impl/StaticLoggerBinder.class]  
SLF4J: Found binding in [jarfile:/usr/program/tee/lib/slf4j-log4j12-1.7.10.jar/org/slf4j/impl/StaticLoggerBinder.class]  
SLF4J: Found binding in [jarfile:/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]  
2023-12-03 16:43:30.543 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable  
bash-5.0# hdfs dfs -ls /tmp  
SLF4J: Class path contains multiple SLF4J bindings.  
SLF4J: Found binding in [jarfile:/usr/program/hadoop/share/hadoop/common/lib/slf4j-log4j12-1.7.25.jar/org/slf4j/impl/StaticLoggerBinder.class]  
SLF4J: Found binding in [jarfile:/usr/program/tee/lib/slf4j-log4j12-1.7.10.jar/org/slf4j/impl/StaticLoggerBinder.class]  
SLF4J: Found binding in [jarfile:/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]  
2023-12-03 16:43:33.543 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable  
Found 2 items  
drwxr-xr-x  2 root supergroup      0 2023-12-03 16:35 /tmp/Chive  
-r-r--r--  1 root supergroup      0 2023-12-03 16:43 /tmp/testfile1.txt  
bash-5.0# hdfs dfs -rm /tmp/testfile1.txt  
SLF4J: Class path contains multiple SLF4J bindings.  
SLF4J: Found binding in [jarfile:/usr/program/hadoop/share/hadoop/common/lib/slf4j-log4j12-1.7.25.jar/org/slf4j/impl/StaticLoggerBinder.class]  
SLF4J: Found binding in [jarfile:/usr/program/tee/lib/slf4j-log4j12-1.7.10.jar/org/slf4j/impl/StaticLoggerBinder.class]  
SLF4J: Found binding in [jarfile:/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]  
2023-12-03 16:43:34.543 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable  
Deleted /tmp/testfile1.txt  
bash-5.0#
```

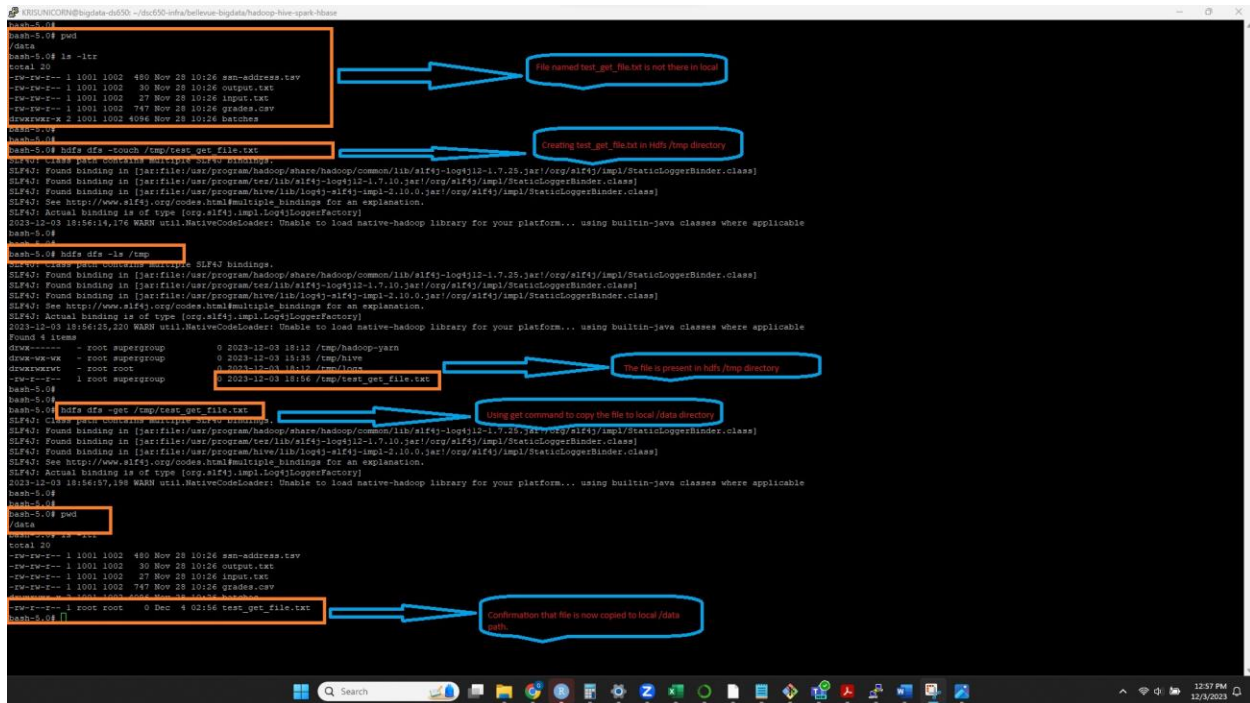
3.2. Screenshot of HDFS command 2 → copyToLocal command.

A new file testfile3.txt was created in hdfs /tmp directory and was copied to local /data directory using copyToLocal command

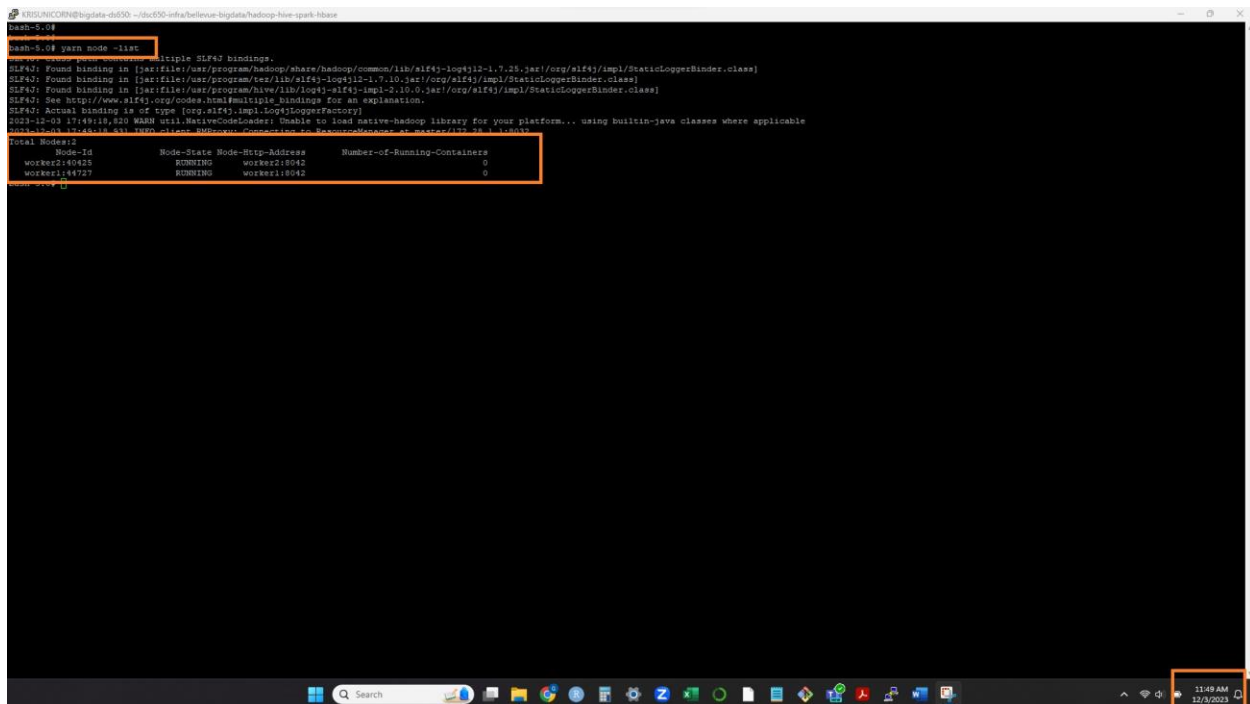
```
bash-5.0# cd /data
bash-5.0# pwd
/data
hdfs-5.0# ls -ls
total 28
drwxrwxr-x 3 1001 1002 4096 Nov 28 10:16 .
drwxr-xr-x 1 root root 4096 Dec 3 22:18 ..
drwxrwxr-x 2 1001 1002 4096 Nov 28 10:16 hachana
-rw-rw-r-- 1 1001 1002 747 Nov 28 10:16 grades.csv
-rw-rw-r-- 1 1001 1002 27 Nov 28 10:16 input.txt
-rw-rw-r-- 1 1001 1002 30 Nov 28 10:16 output.txt
-rw-rw-r-- 1 1001 1002 485 Nov 28 10:16 sm-addresses.tsv
hdfs-5.0# hdfs dfs -cp /tmp/testfile3.txt /data
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jarfile:/usr/program/hadoop/share/hadoop/common/lib/slf4j-log4j12-1.7.25.jar/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jarfile:/usr/program/tez/lib/slf4j-log4j12-1.7.10.jar/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jarfile:/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]
2023-12-03 17:16:09,859 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
hdfs-5.0# hdfs dfs -ls /tmp
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jarfile:/usr/program/hadoop/share/hadoop/common/lib/slf4j-log4j12-1.7.25.jar/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jarfile:/usr/program/tez/lib/slf4j-log4j12-1.7.10.jar/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jarfile:/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]
2023-12-03 17:16:09,859 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 2 items
drwxrwxr-x 3 1001 1002 4096 Nov 28 10:16 .
drwxr-xr-x 1 root root 4096 Dec 3 22:18 ..
-rw-rw-r-- 1 root supergroup 0 2023-12-03 15:08 /tmp/_all
-rw-rw-r-- 1 root supergroup 0 2023-12-03 17:14 /tmp/testfile3.txt
hdfs-5.0# hdfs dfs -copyToLocal /tmp/testfile3.txt /data
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jarfile:/usr/program/hadoop/share/hadoop/common/lib/slf4j-log4j12-1.7.25.jar/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jarfile:/usr/program/tez/lib/slf4j-log4j12-1.7.10.jar/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jarfile:/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]
hdfs-5.0# hdfs dfs -ls /tmp
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jarfile:/usr/program/hadoop/share/hadoop/common/lib/slf4j-log4j12-1.7.25.jar/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jarfile:/usr/program/tez/lib/slf4j-log4j12-1.7.10.jar/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jarfile:/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]
2023-12-03 17:16:09,859 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 2 items
drwxrwxr-x 3 1001 1002 4096 Nov 28 10:16 .
drwxr-xr-x 1 root root 4096 Dec 3 22:18 ..
-rw-rw-r-- 1 1001 1002 480 Nov 28 10:16 sm-addresses.tsv
-rw-rw-r-- 1 1001 1002 30 Nov 28 10:16 output.txt
-rw-rw-r-- 1 1001 1002 27 Nov 28 10:16 input.txt
-rw-rw-r-- 1 1001 1002 747 Nov 28 10:16 grades.csv
drwxrwxr-x 2 1001 1002 4096 Nov 28 10:16 hachana
-rw-rw-r-- 1 1001 1002 485 Nov 28 10:16 testfile3.txt
hdfs-5.0#
```

3.3. Screenshot of HDFS command 3 → get command.

A new file test_get_file.txt was created in hdfs /tmp directory and was copied to local /data directory using get command



4. Screenshot of yarn list command:



5. Screenshots of Yarn yarn.scheduler.maximum-allocation-mb :

5.1. Before making the change:

```
<?xml version="1.0" encoding="UTF-8"?>
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>

<!-- http://hadoop.apache.org/docs/r3.1.2/hadoop-project-dist/hadoop-common/ClusterSetup.html -->
<configuration>

  <!-- Single hostname that can be set in place of setting all
  yarn.resourcemanager.address resources. Results in default ports
  for ResourceManager components: scheduler - 8030, resource-tracker - 8031,
  resource-manager - 8032, admin - 8033, webapp - 8088. -->
  <property>
    <name>yarn.resourcemanager.hostname</name>
    <value>master</value>
  </property>

  <!-- Shuffle service that needs to be set for Map Reduce applications. -->
  <property>
    <name>yarn.nodemanager.aux-services</name>
    <value>mapreduce_shuffle</value>
  </property>
  <property>
    <name>yarn.nodemanager.aux-services.mapreduce_shuffle.class</name>
    <value>org.apache.hadoop.mapred.ShuffleHandler</value>
  </property>

  <!-- Memory and CPU constraints -->

  <!-- Amount of physical memory, in MB, that can be allocated for containers.
  It means the amount of memory YARN can utilize on this node and therefore
  this property should be lower than the total memory of that machine. -->
  <property>
    <name>yarn.nodemanager.resource.memory-mb</name>
    <value>4096</value>
  </property>

  <!-- Resource manager can only allocate memory to containers in increments
  of "yarn.scheduler.minimum-allocation-mb" (1024 by default) and not exceed
  "yarn.scheduler.maximum-allocation-mb" (8192 by default).
  It should not be more than total memory of the Node.
  YARN processes each map or reduce task in a container so this
  param divided by param above is effectively how many jobs you can
  run at once. -->
  <property>
    <name>yarn.scheduler.maximum-allocation-mb</name>
    <value>1536</value>
  </property>
  <property>
    <name>yarn.scheduler.minimum-allocation-mb</name>
    <value>512</value>
  </property>

  <!-- And it can only allocate CPU vcores to containers
  in increments of "yarn.scheduler.minimum-allocation-vcores"
  and not exceed "yarn.scheduler.maximum-allocation-vcores". -->
  <property>
    <name>yarn.scheduler.maximum-allocation-vcores</name>
    <value>4</value>
  </property>
  <property>
    <name>yarn.scheduler.minimum-allocation-vcores</name>
    <value>1</value>
  </property>
  <!--More-->
</configuration>
```

Google Account | VM Instances - Compute Engine | All Applications

http://localhost:8088/cluster

All Applications

Cluster Metrics

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Used Resources	Total Resources	Reserved Resources	Physical Mem Used %	Physical VCore Used %
0	0	0	0	0	<memory 0 B, vCores 0>	<memory 8 GB, vCores 16>	<memory 0 B, vCores 0>	75	0

Cluster Nodes Metrics

Active Nodes	Decommissioning Nodes	Decommissioned Nodes	Lost Nodes	Unhealthy Nodes	Relocated 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	VCore Used	VCore Pending	VCore Reserved
0	0	0	0	0	0	0	0 B	0 B	0 B	0	0	0

Scheduler Metrics

Fair Scheduler	Scheduler Type	Scheduling Resource Type	Minimum Allocation	Maximum Allocation	Minimum Cluster Application Priority	Scheduler Busy %
1	memory-mb (unit-MB, vCores)		<memory 512, vCores 1>	<memory 1536, vCores 4>	0	0

Showing 0 to 0 of 0 entries

ID	User	Name	Application Type	Queue	Application Priority	Start Time	Launch Time	Finish Time	State	Final Status	Running Containers	Allocated CPU VCore	Allocated Memory MB	Allocated GPU	Reserved CPU VCore	Reserved Memory MB	Reserved GPU	% of Queue	% of Cluster	Progress	Tracking UI	Blacklisted Nodes
No data available in table																						

Showing 0 to 0 of 0 entries

5.2. After making the changes:

```
CR00NC0RN@bigdata-d000: ~/bigdata-hadoop$ cat hdfs-site.xml
<?xml version="1.0" encoding="UTF-8"?>
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>

<!-- http://hadoop.apache.org/docs/r3.1.2/hadoop-project-dist/hadoop-common/ClusterSetup.html -->
configuration)

<!-- Single hostname that can be set in place of setting all
yarn.resourcemanager.address resources. Results in default ports
for ResourceManager components: scheduler - 8030, resource-tracker - 8031,
resource-manager - 8032, admin - 8033, webapp - 8088. -->
<property>
  <name>yarn.resourcemanager.hostname</name>
  <value>master</value>
</property>

<!-- Shuffle service that needs to be set for Map Reduce applications. -->
<property>
  <name>yarn.nodemanager.aux-services</name>
  <value>mapreduce_shuffle</value>
</property>
<property>
  <name>yarn.nodemanager.aux-services.mapreduce_shuffle.class</name>
  <value>org.apache.hadoop.mapred.ShuffleHandler</value>
</property>

<!-- Memory and CPU constraints -->

<!-- Amount of physical memory, in MB, that can be allocated for containers.
It means the amount of memory YARN can utilize on this node and therefore
this property should be lower than the total memory of that machine. -->
<property>
  <name>yarn.nodemanager.resource.memory-mb</name>
  <value>4096</value>
</property>

<!-- Resource manager can only allocate memory to containers in increments
of "yarn.scheduler.minimum-allocation-mb" (1024 by default) and not exceed
"yarn.scheduler.maximum-allocation-mb" (8192 by default).
It should not be more than total memory of the Node.
YARN processes each map or reduce task in a container so this
param divided by param above is effectively how many jobs you can
run simultaneously. -->
<property>
  <name>yarn.scheduler.maximum-allocation-mb</name>
  <value>2048</value>
</property>
<property>
  <name>yarn.scheduler.minimum-allocation-mb</name>
  <value>512</value>
</property>

<!-- And it can only allocate CPU vcores to containers
in increments of "yarn.scheduler.minimum-allocation-vcores"
and not exceed "yarn.scheduler.maximum-allocation-vcores". -->
<property>
  <name>yarn.scheduler.maximum-allocation-vcores</name>
  <value>4</value>
</property>
<property>
  <name>yarn.scheduler.minimum-allocation-vcores</name>
  <value>1</value>
</property>
--More--
```

hadoop

Cluster

Nodes

Node Labels

Associations

NEW

NEW SAVING

SUBMITTED

ACCEPTED

RUNNING

FINISHING

FAILED

SHUTDOWN

Scheduler

Tools

All Applications

Cluster Metrics

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Used Resources	Total Resources	Reserved Resources	Physical Mem Used %	Physical VCore Used %
0	0	0	0	0	<memory 0 B, vCores 0>	<memory 8 GB, vCores 16>	<memory 0 B, vCores 0>	75	0

Cluster Nodes Metrics

Active Nodes	Decommissioning Nodes	Decommissioned Nodes	Last Nodes	Unhealthy Nodes	Relocated 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	VCore Used	VCore Pending	VCore 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	Minimum Cluster Application Priority	Scheduler Busy %
Fair Scheduler	[memory-mb (unit-MB), vCores]	<memory 512, vCores 1>	<memory 2048, vCores 4>	0	0

Show 29 entries

ID	User	Name	Application Type	Queue	Application Priority	StartTime	LaunchTime	FinishTime	State	FinalStatus	Running Containers	Allocated CPU V-Cores	Allocated Memory MB	Allocated GPU	Reserved CPU V-Cores	Reserved Memory MB	Reserved GPU	% of Queue	% of Cluster	Progress	Tracking UI	Backlisted Nodes
No data available in table																						

Showing 0 to 0 of 0 entries

12:08 PM 12/3/2023

[illegible]

Google Account

VM instances - Compute Engine

All Applications

← → ↻

http://localhost:8088/cluster

🔍 ⚙️ ⌵ 🗖 📄

Cluster

About Nodes Node Labels Associations NEW NEW ADDED SUBMITTED ACCEPTED RUNNING FAILED SUCCEEDED Scheduler

Tools

Cluster Metrics

Cluster Nodes Metrics

User Metrics for dr who

Scheduler Metrics

Scheduler Type

All Applications

Logged in as: dr who

Cluster Metrics

Cluster Nodes Metrics

User Metrics for dr who

Scheduler Metrics

Scheduler Type

1 Apps Submitted 1 Apps Pending 0 Apps Running 0 Apps Completed 1 Containers Running <memory 1 GB, vCores 1> Total Resources <memory 8 GB, vCores 16> Reserved Resources <memory 9 B, vCores 0> Physical Mem Used % 86 Physical VCoers Used % 37

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

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

0 Scheduler Metrics

0 Scheduler Type

100% 100%

100% 100%

100% 100%

100% 100%

100% 100%

100% 100%

100% 100%

100% 100%

100% 100%

100% 100%

100% 100%

100% 100%

100% 100%

100% 100%

100% 100%

100% 100%

100% 100%

100% 100%

100% 100%

100% 100%

100% 100%

100% 100%

100% 100%

100% 100%

100% 100%

100% 100%

100% 100%

100% 100%

100% 100%

100% 100%

100% 100%

100% 100%

100% 100%

100% 100%

100% 100%

100% 100%

100% 100%

100% 100%

100% 100%

100% 100%

100% 100%

100% 100%

100% 100%

100% 100%

100% 100%

100% 100%

100% 100%

100% 100%

100% 100%

100% 100%

100% 100%

100% 100%

100% 100%

100% 100%

100% 100%

100% 100%

100% 100%

100% 100%

100% 100%

100% 100%

100% 100%

100% 100%

100% 100%

100% 100%

100% 100%

100% 100%

100% 100%

100% 100%

100% 100%

100% 100%

100% 100%

100% 100%

100% 100%

100% 100%

100% 100%

100% 100%

100% 100%

100% 100%

100% 100%

100% 100%

100% 100%

100% 100%

100% 100%

100% 100%

100% 100%

100% 100%

100% 100%

100% 100%

100% 100%

100% 100% 100% 100% 100% 100% 100% 100% 100% 1

The `hadoop-mapreduce-examples-3.2.3.jar` program is used to calculate the pi value using quasi-monte Carlo method. It takes 2 arguments, the first one is the number of maps and the second one is the number of samples per map. The

numerical approximation of arbitrary examples method is used in this program. The Mapper class generates points in the Unit square and then counts points inside or outside of the inscribed circle of the square. The Reducer combines the points inside or outside results from the mappers. Once the results are calculated it is sent to the output file.

Formula Calculation:

In this method, the value of pi is calculated using the formula $\pi = 4 \times \frac{\text{numInside}}{\text{numTotal}}$. So the value can be calculated once the value of $\frac{\text{numInside}}{\text{numTotal}}$ is calculated.

The fraction $\frac{\text{numInside}}{\text{numTotal}}$ is approximately equivalent to $\frac{(\text{Area of the circle})}{(\text{Area of the square})}$ which is the same as $\frac{\pi}{4}$.

$$\frac{\text{numInside}}{\text{numTotal}} = \frac{(\text{Area of the circle})}{(\text{Area of the square})} = \frac{\pi}{4}$$

The Area of the unit square is 1.

The Area of the inscribed circle is $\frac{\pi}{4}$

So, the equation can be broken down as follows:

$$\frac{\text{numInside}}{\text{numTotal}} = \frac{\pi}{4} \Rightarrow \pi = 4 \times \frac{\text{numInside}}{\text{numTotal}}$$

$$\text{so, } \pi = 4 \left(\frac{\text{numInside}}{\text{numTotal}} \right)$$

Observations:

In the first screenshot, we used 2 maps and 10 samples per map and the value was calculated as 3.8 and the program completed in 25.1 seconds. In the second test run, 2 maps with 100 samples per map were used (the screenshot below) and the value pi was calculated as 3.12 in 23.7 seconds.

```
CRDUNCORP@bigdata-d000: ~/bigdata-hadoop-hive-spark-flow
File System Counters
  FILE: Number of bytes read=0
  FILE: Number of bytes written=738252
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=820
  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)=16338
  Total time spent by all reduces in occupied slots (ms)=13536
  Total time spent by all map tasks (ms)=8149
  Total time spent by all reduce tasks (ms)=3484
  Total vcore-milliseconds taken by all map tasks=8149
  Total vcore-milliseconds taken by all reduce tasks=3484
  Total megabyte-milliseconds taken by all map tasks=8365056
  Total megabyte-milliseconds taken by all reduce tasks=7135232
Map-Reduce Framework
  Map input records=2
  Map output records=4
  Map output bytes=14
  Map output materialized bytes=56
  Input split bytes=294
  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 =3
  Failed Shuffles=0
  Merged Map outputs=2
  GC time elapsed (ms)=205
  CPU time spent (ms)=1830
  Physical memory (bytes) snapshot=747536354
  Virtual memory (bytes) snapshot=7938452480
  Total committed heap usage (bytes)=7035184486
  Peak Map Physical memory (bytes)=254050368
  Peak Map Virtual memory (bytes)=1353277840
  Peak Reduce Physical memory (bytes)=1504500544
  Peak Reduce Virtual memory (bytes)=2354784304
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=0
Job Finished in 23.704 seconds
Estimated value of Pi is 3.120200000000000000000000
bash-5.0$
```