ASSIGNMENT 02

Distributed Hadoop Clusters

Sabaragamuwa University of Sri Lanka

Faculty of Computing

Department of Software Engineering

SE6103 – Parallel and Distributed Systems

Name   : W.H.M.L.S. Bandara

Reg. No   : 19APSE4298

Academic Period : 3rd Year 2nd Semester

- docker pull bde2020/hadoop-datanode:latest
- docker pull bde2020/hadoop-historyserver

```
PS E:\Sabra 3.2\Docker\Labsheet2> docker pull bde2020/hadoop-datanode:latest
latest: Pulling from bde2020/hadoop-datanode
Digest: sha256:35f899bcbe9f983825a8a3bdc135ed0e8e0eaf3b58f9b08bf257b5e86bae3b47
Status: Image is up to date for bde2020/hadoop-datanode:latest
docker.io/bde2020/hadoop-datanode:latest

What's Next?
  View a summary of image vulnerabilities and recommendations → docker scout quickview bde2020/hadoop-datanode:latest
PS E:\Sabra 3.2\Docker\Labsheet2> docker pull bde2020/hadoop-historyserver
Using default tag: latest
latest: Pulling from bde2020/hadoop-historyserver
Digest: sha256:216100a96a73717006031ff0c8b72effdc7acffca0a6c647f8820cb7eabc81fd
Status: Image is up to date for bde2020/hadoop-historyserver:latest
docker.io/bde2020/hadoop-historyserver:latest
```

- docker volume create namenode
- docker volume create datanode

```
PS E:\Sabra 3.2\Docker\Labsheet2> docker volume create namenode
namenode
PS E:\Sabra 3.2\Docker\Labsheet2> docker volume create datanode
datanode
```

- docker-compose up -d

```
PS E:\Sabra 3.2\Docker\Labsheet2> docker-compose up -d
[+] Running 13/13
 ✓ resourcemanager 12 layers [██████████████]      0B/0B        Pulled
   ✓ 3192219afd04 Already exists
   ✓ 7127a1d8cced Already exists
   ✓ 883a89599900 Already exists
   ✓ 77920a3e82af Already exists
   ✓ 92329e81aec4 Already exists
   ✓ f373218fec59 Already exists
   ✓ aa53513fe997 Already exists
   ✓ 8b1800105b98 Already exists
   ✓ c3a84a3e49c8 Already exists
   ✓ a65640a64a76 Already exists
   ✓ 80be1b2ed3e1 Pull complete
   ✓ 26a86cba0190 Pull complete
[+] Running 5/5
 ✓ Network labsheet2_default   Created
 ✓ Container resourcemanager   Started
 ✓ Container namenode          Started
 ✓ Container datanode          Started
 ✓ Container historyserver     Started
```
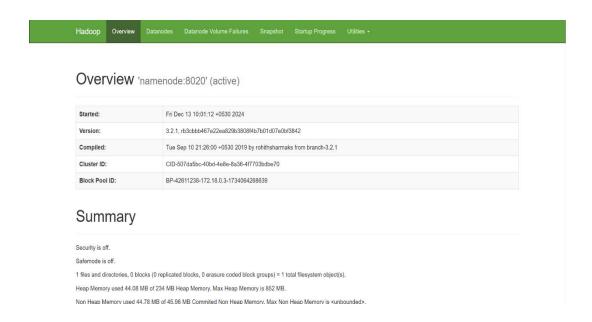
- Start-Process http://localhost:9870

```
PS E:\Sabra 3.2\Docker\Labsheet2> Start-Process http://localhost:9870
PS E:\Sabra 3.2\Docker\Labsheet2>
```

| Hadoop | Overview | Datanodes | Datanode Volume Failures | Snapshot | Startup Progress | Utilities ▾ |

Overview 'namenode:8020' (active)

Started:	Fri Dec 13 10:01:12 +0530 2024
Version:	3.2.1, rb3cbbb467e22ea829b3808f4b7b01d07e0bf3842
Compiled:	Tue Sep 10 21:26:00 +0530 2019 by rohithsharmaks from branch-3.2.1
Cluster ID:	CID-507da5bc-40bd-4e8e-8a36-4f7703bdbe70
Block Pool ID:	BP-42611238-172.18.0.3-1734064268639

Summary

Security is off.

Safemode is off.

1 files and directories, 0 blocks (0 replicated blocks, 0 erasure coded block groups) = 1 total filesystem object(s).

Heap Memory used 44.08 MB of 234 MB Heap Memory. Max Heap Memory is 852 MB.

Non Heap Memory used 44.78 MB of 45.96 MB Commited Non Heap Memory. Max Non Heap Memory is <unbounded>.

- docker exec -it namenode hdfs dfs -mkdir -p /input
- docker exec -it namenode hdfs dfs -put /path/to/sample.txt /input

```
PS E:\Sabra 3.2\Docker\Labsheet2> docker exec -it namenode hdfs dfs -mkdir -p /input
PS E:\Sabra 3.2\Docker\Labsheet2> docker exec -it namenode hdfs dfs -mkdir -p /input
PS E:\Sabra 3.2\Docker\Labsheet2> docker exec -it namenode hdfs dfs -put /path/to/sample.txt /input
put: `/path/to/sample.txt': No such file or directory
PS E:\Sabra 3.2\Docker\Labsheet2> docker exec -it namenode hdfs dfs -put /path/to/sample.txt /input
put: `/path/to/sample.txt': No such file or directory
PS E:\Sabra 3.2\Docker\Labsheet2> docker exec -it namenode hdfs dfs -mkdir -p /input
PS E:\Sabra 3.2\Docker\Labsheet2> docker exec -it namenode hdfs dfs -put /path/to/sample.txt /input
put: `/path/to/sample.txt': No such file or directory
```

- docker cp "E:\Sabra 3.2\Docker\Labsheet2\sample.txt" namenode:/sample.txt
- docker cp namenode:/sample.txt "E:\Sabra 3.2\Docker\Labsheet2\sample.txt"

```
PS E:\Sabra 3.2\Docker\Labsheet2> docker cp "E:\Sabra 3.2\Docker\Labsheet2\sample.txt" namenode:/sample.txt
Successfully copied 2.56kB to namenode:/sample.txt
PS E:\Sabra 3.2\Docker\Labsheet2> docker cp namenode:/sample.txt "E:\Sabra 3.2\Docker\Labsheet2\sample.txt"
Successfully copied 2.56kB to E:\Sabra 3.2\Docker\Labsheet2\sample.txt
```

- docker exec -it namenode ls /

```
PS E:\Sabra 3.2\Docker\Labsheet2> docker ps
CONTAINER ID   IMAGE                                COMMAND                CREATED        STATUS                  PORTS
    NAMES
af546a663540   bde2020/hadoop-historyserver:latest  "/entrypoint.sh /run…"  34 minutes ago  Up 34 minutes (healthy)  8188/tcp, 0.0.0.0:19888->19888/tcp
    historyserver
a43cd1f39db1   bde2020/hadoop-datanode:latest       "/entrypoint.sh /run…"  34 minutes ago  Up 34 minutes (healthy)  9864/tcp
    datanode
4819c368ee15   bde2020/hadoop-namenode:latest       "/entrypoint.sh /run…"  34 minutes ago  Up 34 minutes (healthy)  0.0.0.0:9000->9000/tcp, 0.0.0.0:9870->9870
/tcp    namenode
d64f529d36f1   bde2020/hadoop-resourcemanager:latest "/entrypoint.sh /run…"  34 minutes ago  Up 34 minutes (healthy)  0.0.0.0:8088->8088/tcp
    resourcemanager
PS E:\Sabra 3.2\Docker\Labsheet2> docker start namenode
namenode
PS E:\Sabra 3.2\Docker\Labsheet2> docker exec -it namenode ls /
KEYS  bin  boot  dev  entrypoint.sh  etc  hadoop  hadoop-data  home  lib  lib64  media  mnt  opt  proc  root  run  run.sh  sample.txt  sbin  srv  sys  tmp  usr  var
PS E:\Sabra 3.2\Docker\Labsheet2>
```

- Download the .jar file

https://repo1.maven.org/maven2/org/apache/hadoop/hadoop-mapreduce-examples/2.7.1/

- docker cp "E:/Sabra 3.2/Docker/hadoop-mapreduce-examples-2.7.1-javadoc.jar" namenode:/root/
- docker exec -it namenode hadoop jar /root/hadoop-mapreduce-examples-2.7.1-javadoc.jar org.apache.hadoop.examples.WordCount /input /output

```
PS E:\Sabra 3.2\Docker\Labsheet2> docker exec -it namenode hadoop jar /root/hadoop-mapreduce-examples-2.7.1-javadoc.jar org.apache.hadoop.examples.W
ordCount /input /output
2024-12-13 05:59:31,238 INFO impl.MetricsConfig: Loaded properties from hadoop-metrics2.properties
2024-12-13 05:59:31,279 INFO impl.MetricsSystemImpl: Scheduled Metric snapshot period at 10 second(s).
2024-12-13 05:59:31,279 INFO impl.MetricsSystemImpl: JobTracker metrics system started
2024-12-13 05:59:31,547 INFO input.FileInputFormat: Total input files to process : 0
2024-12-13 05:59:31,554 INFO mapreduce.JobSubmitter: number of splits:0
2024-12-13 05:59:31,672 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_local618763851_0001
2024-12-13 05:59:31,672 INFO mapreduce.JobSubmitter: Executing with tokens: []
2024-12-13 05:59:31,748 INFO mapreduce.Job: The url to track the job: http://localhost:8080/
2024-12-13 05:59:31,749 INFO mapreduce.Job: Running job: job_local618763851_0001
2024-12-13 05:59:31,750 INFO mapred.LocalJobRunner: OutputCommitter set in config null
2024-12-13 05:59:31,756 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2024-12-13 05:59:31,756 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cl
eanup failures: false
2024-12-13 05:59:31,757 INFO mapred.LocalJobRunner: OutputCommitter is org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter
2024-12-13 05:59:31,799 INFO mapred.LocalJobRunner: Waiting for map tasks
2024-12-13 05:59:31,799 INFO mapred.LocalJobRunner: map task executor complete.
2024-12-13 05:59:31,804 INFO mapred.LocalJobRunner: Waiting for reduce tasks
2024-12-13 05:59:31,804 INFO mapred.LocalJobRunner: Starting task: attempt_local618763851_0001_r_000000_0
2024-12-13 05:59:31,825 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2024-12-13 05:59:31,825 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cl
eanup failures: false
2024-12-13 05:59:31,842 INFO mapred.Task:  Using ResourceCalculatorProcessTree : [ ]
2024-12-13 05:59:31,845 INFO mapred.ReduceTask: Using ShuffleConsumerPlugin: org.apache.hadoop.mapreduce.task.reduce.Shuffle@505b28b
2024-12-13 05:59:31,846 WARN impl.MetricsSystemImpl: JobTracker metrics system already initialized!
2024-12-13 05:59:31,859 INFO reduce.MergeManagerImpl: MergerManager: memoryLimit=625370688, maxSingleShuffleLimit=156342672, mergeThreshold=41274467
2, ioSortFactor=10, memToMemMergeOutputsThreshold=10
2024-12-13 05:59:31,865 INFO reduce.EventFetcher: attempt_local618763851_0001_r_000000_0 Thread started: EventFetcher for fetching Map Completion Ev
ents
2024-12-13 05:59:31,870 INFO reduce.EventFetcher: EventFetcher is interrupted.. Returning
2024-12-13 05:59:31,873 INFO mapred.LocalJobRunner:
2024-12-13 05:59:31,873 INFO reduce.MergeManagerImpl: finalMerge called with 0 in-memory map-outputs and 0 on-disk map-outputs
2024-12-13 05:59:31,875 INFO reduce.MergeManagerImpl: Merging 0 files, 0 bytes from disk
```
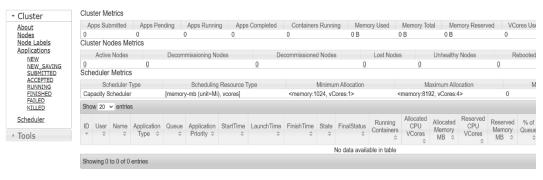
run you should add the

- docker-compose.yaml update

```
resourcemanager:
  image: bde2020/hadoop-resourcemanager:latest
  container_name: resourcemanager
  environment:
    - CORE_CONF_fs_defaultFS=hdfs://namenode:8020
  ports:
    - "8088:8088"
```

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	Memory Used	Memory Total	Memory Reserved	VCores Use
0	0	0	0	0	0 B	0 B	0 B	0

Cluster Nodes Metrics

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

Scheduler Metrics

Scheduler Type	Scheduling Resource Type	Minimum Allocation	Maximum Allocation	M
Capacity Scheduler	[memory-mb (unit=Mi), vcores]	<memory:1024, vCores:1>	<memory:8192, vCores:4>	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	Reserved CPU VCores	Reserved Memory MB	% of Queue
								No data available in table								

Showing 0 to 0 of 0 entries

- docker exec -it namenode hdfs dfs -ls /output

```
PS E:\Sabra 3.2\Docker\Labsheet2> docker exec -it namenode hdfs dfs -ls /output
Found 2 items
-rw-r--r--   3 root supergroup          0 2024-12-13 05:59 /output/_SUCCESS
-rw-r--r--   3 root supergroup          0 2024-12-13 05:59 /output/part-r-00000
```

- docker exec -it namenode hdfs dfs -ls /output
- docker exec -it namenode hdfs dfs -cat /output/part-r-00000

```
Found 2 items
-rw-r--r--   3 root supergroup          0 2024-11-25 08:58 /output2/_SUCCESS
-rw-r--r--   3 root supergroup         38 2024-11-25 08:58 /output2/part-r-00000
PS F:\SE\VI Semester\SE6103 Parallel and Distributed Systems\repo> docker exec -it namenode hdfs dfs -cat /output/part-r-00000
2024-11-25 09:03:28,371 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
This    1
for     1
hadoop  1
sample  1
text    1
```

- docker-compose down

```
PS E:\Sabra 3.2\Docker\Labsheet2> docker-compose down
[+] Running 5/5
 ✓ Container historyserver     Removed
 ✓ Container resourcemanager   Removed
 ✓ Container datanode          Removed
 ✓ Container namenode          Removed
 ✓ Network labsheet2_default   Removed
```