

ASSIGNMENT 02

Sabaragamuwa University of Sri Lanka

Faculty of Computing

Department of Software Engineering

SE6103 – Parallel & Distributed Systems

Name	: V Partheepan
Reg. No	: 19APSE4284
Academic Period	: 03 rd Year 02 nd Semester
Degree Program	: Software Engineering
Due Date	: 25/11/2024

Compose up -d

```
PS D:\SOFTWARE ENGINEERING\SEMESTER 6\Parallel and Distributed Systems\Hadoop> docker compose up -d
empty compose file
PS D:\SOFTWARE ENGINEERING\SEMESTER 6\Parallel and Distributed Systems\Hadoop> docker compose up -d
[+] Running 6/6
 ✓ Network hadoop_default          Created
 ✓ Volume "hadoop_namenode-data"   Created
 ✓ Volume "hadoop_datanode-data"   Created
 ✓ Container namenode              Started
 ✓ Container datanode              Started
 ✓ Container historyserver         Started
PS D:\SOFTWARE ENGINEERING\SEMESTER 6\Parallel and Distributed Systems\Hadoop>
```

Display all running containers

```
PS D:\SOFTWARE ENGINEERING\SEMESTER 6\Parallel and Distributed Systems\Hadoop> docker ps
CONTAINER ID   IMAGE                                COMMAND                  CREATED        STATUS        PORTS        NAMES
PS D:\SOFTWARE ENGINEERING\SEMESTER 6\Parallel and Distributed Systems\Hadoop> docker ps -a
CONTAINER ID   IMAGE                                COMMAND                  CREATED        STATUS        PORTS        NAMES
ec5dfcb3a8d1   bde2020/hadoop-namenode:latest      "/entrypoint.sh /bin..." 3 weeks ago    Exited (255) 3 weeks ago    0.0.0.0:8088->8088/tcp, hadoop-cluster-2
75eafa9aeafd   bde2020/hadoop-namenode:latest      "/entrypoint.sh /run..." 3 weeks ago    Exited (2) 3 weeks ago    goofy_robinson
b5f80af5a861   bde2020/hadoop-namenode:latest      "/entrypoint.sh /run..." 3 weeks ago    Exited (2) 3 weeks ago    tender_carson
4b099dfbda32   bde2020/hadoop-namenode:latest      "/entrypoint.sh /bin..." 5 weeks ago    Exited (130) 5 weeks ago    hadoop-cluster1
8903a01628af   bde2020/hadoop-namenode:latest      "/entrypoint.sh /run..." 5 weeks ago    Exited (2) 5 weeks ago    romantic_mclean
```

http://Localhost:9870

localhost:9870/dfshealth.html#tab-overview

Hadoop

Overview

Datanodes

Datanode Volume Failures

Snapshot

Startup Progress

Utilities

Overview

'namenode:8020' (active)

Started:	Thu Dec 12 10:32:07 +0530 2024
Version:	3.2.1, rb3cbb467e22ea829b3808f4b7b01d07e0bf3842
Compiled:	Tue Sep 10 21:26:00 +0530 2019 by rohitsharmaks from branch-3.2.1
Cluster ID:	CID-9ba5646a-a7f1-4fac-8f81-3330f12ab27e
Block Pool ID:	BP-326806746-172.18.0.2-1733979724461

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 42.62 MB of 258 MB Heap Memory. Max Heap Memory is 850 MB.

Non Heap Memory used 44.44 MB of 45.9 MB Committed Non Heap Memory. Max Non Heap Memory is <unbounded>.

Copy sample text file to namenode:sample.txt

```
PS D:\SOFTWARE ENGINEERING\SEMESTER 6\Parallel and Distributed Systems\Hadoop> docker cp "D:\SOFTWARE ENGINEERING\SEMESTER 6\Parallel and Distributed Systems\Hadoop\sample.txt" namenode:/sample.txt
Successfully copied 2.05kB to namenode:/sample.txt
PS D:\SOFTWARE ENGINEERING\SEMESTER 6\Parallel and Distributed Systems\Hadoop>
```

docker exec -it namenode ls /

```
PS D:\SOFTWARE ENGINEERING\SEMESTER 6\Parallel and Distributed Systems\Hadoop> docker exec -it namenode ls
KEYS dev          hadoop    lib      mnt      root     sample.txt sys var
bin  entrypoint.sh    hadoop-data lib64    opt      run      sbin     tmp
boot etc           home     media    proc     run.sh   srv      usr

What's next:
  Try Docker Debug for seamless, persistent debugging tools in any container or image → docker debug namenode
  Learn more at https://docs.docker.com/go/debug-cli/
PS D:\SOFTWARE ENGINEERING\SEMESTER 6\Parallel and Distributed Systems\Hadoop>
```

docker exec -it namenode hdfs dfs -mkdir -p /input

docker exec -it namenode hdfs dfs -put /sample.txt /input

docker exec -it namenode hdfs dfs -ls /input

```
PS D:\SOFTWARE ENGINEERING\SEMESTER 6\Parallel and Distributed Systems\Hadoop> docker exec -it namenode hdfs dfs -mkdir -p /input

What's next:
  Try Docker Debug for seamless, persistent debugging tools in any container or image → docker debug namenode
  Learn more at https://docs.docker.com/go/debug-cli/
PS D:\SOFTWARE ENGINEERING\SEMESTER 6\Parallel and Distributed Systems\Hadoop> docker exec -it namenode hdfs dfs -ls /input
Found 1 items
-rw-r--r--   3 root supergroup      85 2024-12-12 05:20 /input/sample.txt

What's next:
  Try Docker Debug for seamless, persistent debugging tools in any container or image → docker debug namenode
  Learn more at https://docs.docker.com/go/debug-cli/
PS D:\SOFTWARE ENGINEERING\SEMESTER 6\Parallel and Distributed Systems\Hadoop>
```

`docker exec -it namenode hadoop jar /opt/hadoop-3.2.1/share/hadoop/mapreduce/hadoop-mapreduce-examples-3.2.1.jar wordcount /input /output`

```
PS D:\SOFTWARE ENGINEERING\SEMESTER 6\Parallel and Distributed Systems\Hadoop> docker exec -it namenode hadoop jar /opt/hadoop-3.2.1/share/hadoop/mapreduce/hadoop-mapreduce-examples-3.2.1.jar wordcount /input /output
>>
2024-12-12 07:44:51,538 INFO impl.MetricsConfig: Loaded properties from hadoop-metrics2.properties
2024-12-12 07:44:51,699 INFO impl.MetricsSystemImpl: Scheduled Metric snapshot period at 10 second(s).
2024-12-12 07:44:51,699 INFO impl.MetricsSystemImpl: JobTracker metrics system started
2024-12-12 07:44:52,085 INFO input.FileInputFormat: Total input files to process : 1
2024-12-12 07:44:52,037 INFO mapreduce.JobSubmitter: number of splits:1
2024-12-12 07:44:52,277 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_local1928355047_0001
2024-12-12 07:44:52,277 INFO mapreduce.JobSubmitter: Executing with tokens: []
2024-12-12 07:44:52,395 INFO mapreduce.Job: The url to track the job: http://localhost:8080/
2024-12-12 07:44:52,396 INFO mapreduce.Job: Running job: job_local1928355047_0001
2024-12-12 07:44:52,398 INFO mapred.LocalJobRunner: OutputCommitter set in config null
2024-12-12 07:44:52,396 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2024-12-12 07:44:52,414 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
2024-12-12 07:44:52,415 INFO mapred.LocalJobRunner: OutputCommitter is org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter
2024-12-12 07:44:52,462 INFO mapred.LocalJobRunner: Waiting for map tasks
2024-12-12 07:44:52,463 INFO mapred.LocalJobRunner: Starting task: attempt_local1928355047_0001_m_000000_0
2024-12-12 07:44:52,491 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2024-12-12 07:44:52,491 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
2024-12-12 07:44:52,517 INFO mapred.Task: Using ResourceCalculatorProcessTree : [ ]
2024-12-12 07:44:52,521 INFO mapped.MapTask: Processing split: hdfs://namenode:8020/input/sample.txt:0-85
```

```
PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS POSTMAN CONSOLE COMMENTS
Combine input records=16
Combine output records=12
Reduce input groups=12
Reduce shuffle bytes=146
Reduce input records=12
Reduce output records=12
Spilled Records=24
Shuffled Maps =1
Failed Shuffles=0
Merged Map outputs=1
GC time elapsed (ms)=12
Total committed heap usage (bytes)=526385152
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=85
File Output Format Counters
Bytes Written=92
What's next:
Try Docker Debug for seamless, persistent debugging tools in any container or image -> docker debug namenode
Learn more at https://docs.docker.com/go/debug-cli/
PS D:\SOFTWARE ENGINEERING\SEMESTER 6\Parallel and Distributed Systems\Hadoop>
```

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

```
PS D:\SOFTWARE ENGINEERING\SEMESTER 6\Parallel and Distributed Systems\Hadoop> docker exec -it namenode hdfs dfs -ls /output
>>
Found 2 items
-rw-r--r-- 3 root supergroup 0 2024-12-12 07:44 /output/_SUCCESS
-rw-r--r-- 3 root supergroup 92 2024-12-12 07:44 /output/part-r-00000

What's next:
Try Docker Debug for seamless, persistent debugging tools in any container or image -> docker debug namenode
Learn more at https://docs.docker.com/go/debug-cli/
PS D:\SOFTWARE ENGINEERING\SEMESTER 6\Parallel and Distributed Systems\Hadoop>
```

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

```
PS D:\SOFTWARE ENGINEERING\SEMESTER 6\Parallel and Distributed Systems\Hadoop> docker exec -it namenode hdfs dfs -cat /output/part-r-00000
>>
2024-12-12 07:47:51,103 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
Big      1
Hadoop   1
This     1
a        2
data     1
demo     1
file.    2
hadoop   1
hadoop.  1
is       3
processing 1
test     1

What's next:
  Try Docker Debug for seamless, persistent debugging tools in any container or image → docker debug namenode
  Learn more at https://docs.docker.com/go/debug-cli/
PS D:\SOFTWARE ENGINEERING\SEMESTER 6\Parallel and Distributed Systems\Hadoop>
```

`docker-compose down`

```
PS D:\SOFTWARE ENGINEERING\SEMESTER 6\Parallel and Distributed Systems\Hadoop> docker-compose down
>>
[+] Running 4/4
✓ Container historyserver Removed
✓ Container datanode      Removed
✓ Container namenode      Removed
✓ Network hadoop_default  Removed
PS D:\SOFTWARE ENGINEERING\SEMESTER 6\Parallel and Distributed Systems\Hadoop>
```