

Rapport des TPs - Big Data

1) Launching the Cluster

```
PS C:\Users\NITRO 5\Desktop\S4\Big Data\tp> docker compose up -d
```

```
time="2026-02-23T15:12:01Z" level=warning msg='C:\\Users\\NITRO 5\\Desktop\\S4\\Big Data\\tp\\tp3\\docker-compose.yaml: the attribute `version` is obsolete, it will be ignored, please remove it to avoid potential confusion'
```

```
time="2026-02-23T15:12:01Z" level=warning msg='The \\HADOOP_HOME\\ variable is not set. Defaulting to a blank string.'
```

```
time="2026-02-23T15:12:01Z" level=warning msg='The \\HADOOP_HOME\\ variable is not set. Defaulting to a blank string.'
```

```
time="2026-02-23T15:12:01Z" level=warning msg='The \\HADOOP_HOME\\ variable is not set. Defaulting to a blank string.'
```

```
time="2026-02-23T15:12:01Z" level=warning msg='The \\HADOOP_HOME\\ variable is not set. Defaulting to a blank string.'
```

```
time="2026-02-23T15:12:01Z" level=warning msg='The \\HADOOP_HOME\\ variable is not set. Defaulting to a blank string.'
```

```
time="2026-02-23T15:12:01Z" level=warning msg='The \\HADOOP_HOME\\ variable is not set. Defaulting to a blank string.'
```

```
time="2026-02-23T15:12:01Z" level=warning msg='The \\HADOOP_HOME\\ variable is not set. Defaulting to a blank string.'
```

```
time="2026-02-23T15:12:01Z" level=warning msg='The \\HADOOP_HOME\\ variable is not set. Defaulting to a blank string.'
```

```
time="2026-02-23T15:12:01Z" level=warning msg='The \\HADOOP_HOME\\ variable is not set. Defaulting to a blank string.'
```

```
(+) Running 7/7
```

```
✓Network tp3_hadoopnet Created 0.8s
```

```
✓Container tp3-resourcemanager-1 Started 0.5s
```

```
✓Container tp3-nodemanager-1 Started 0.5s
```

```
✓Container grafana Started 0.6s
```

```
✓Container namenode Started 0.7s
```

```
✓Container prometheus Started 0.6s
```

```
✓Container datanode Started 0.6s
```

```
PS C:\Users\NITRO 5\Desktop\S4\Big Data\tp>
```

```
PS C:\Users\NITRO 5\Desktop\S4\Big Data\tp> docker compose ps
```

```
time="2026-02-23T15:12:35Z" level=warning msg='C:\\Users\\NITRO 5\\Desktop\\S4\\Big Data\\tp\\tp3\\docker-compose.yaml: the attribute `version` is obsolete, it will be ignored, please remove it to avoid potential confusion'
```

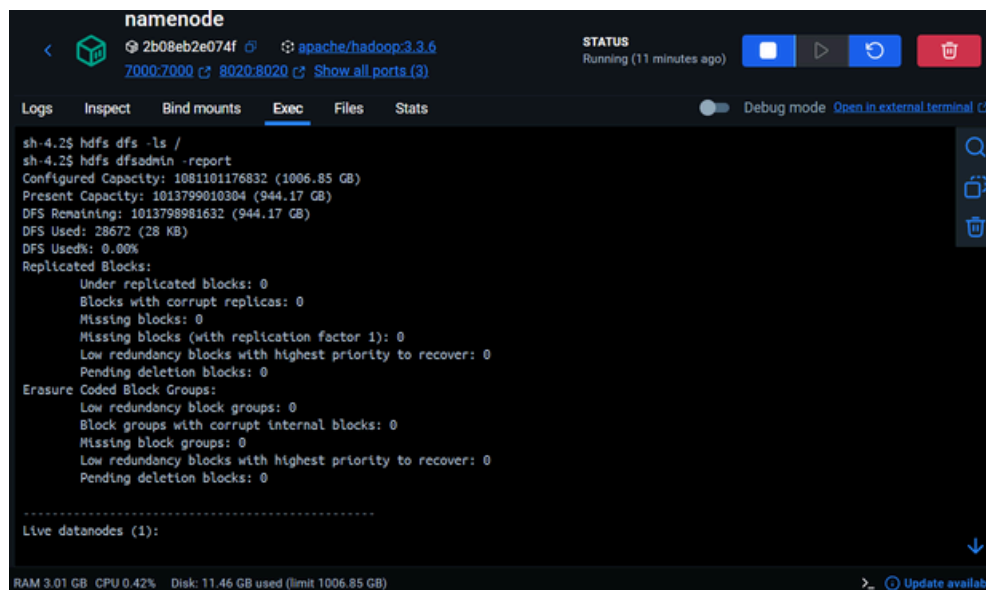
	NAME	IMAGE	CMD	SERVICE	CREATED	STATUS
PORTS						
	datanode	apache/hadoop:3.3.6	"usr/local/bin/dumb-"	datanode	33 seconds ago	Up 33 second
	ds_0.0.0.0:7001->7001/tcp,	[::]:7001->7001/tcp,	0.0.0.0:9864->9864/tcp,	:::9864->9864/tcp		
	grafana	grafana/grafana-latest	".run.sh"	grafana	33 seconds ago	Up 33 second
	ds_0.0.0.0:3000->3000/tcp,	[::]:3000->3000/tcp				
	namenode	apache/hadoop:3.3.6	"usr/local/bin/dumb-"	namenode	33 seconds ago	Up 33 second
	ds_0.0.0.0:7000->7000/tcp,	[::]:7000->7000/tcp,	0.0.0.0:8020->8020/tcp,	:::8020->8020/tcp,	0.0.0.0:9870->9870/tcp,	[::]:9870->9870/tcp
	prometheus	prom/prometheus:latest	"bin/prometheus --c-	prometheus	33 seconds ago	Up 33 second
	ds_0.0.0.0:9090->9090/tcp,	[::]:9090->9090/tcp				
	tp3-nodemanager-1	apache/hadoop:3.3.6	"usr/local/bin/dumb_"	nodemanager	33 seconds ago	Up 33 second
	ds					
	tp3-resourcemanager-1	apache/hadoop:3.3.6	"usr/local/bin/dumb_"	resourcemanager	33 seconds ago	Up 33 second
	ds_0.0.0.0:8088->8088/tcp,	[::]:8088->8088/tcp				

```
PS C:\Users\NITRO 5\Desktop\S4\Big Data\tp>
```

2) Quick Checks via Web UI

The screenshot displays the Hadoop cluster management interface. The top navigation bar includes links for 'Cluster', 'Nodes', 'Node Labels', 'Applications', 'Scheduler', and 'Tools'. The main content area is titled 'All Applications'. It features a 'Cluster Metrics' section with a table showing various metrics like 'Apps Submitted', 'Apps Pending', 'Apps Running', 'Apps Completed', 'Containers Running', 'Used Resources', and 'Lost Nodes'. Below this is a 'Cluster Nodes Metrics' section with a table showing 'Active Nodes', 'Decommissioning Nodes', 'Decommissioned Nodes', and 'Lost Nodes'. The 'Applications' table is currently empty, displaying 'No data available in table' and 'Showing 0 to 0 of 0 entries'.

3) Admin Shell in the NameNode



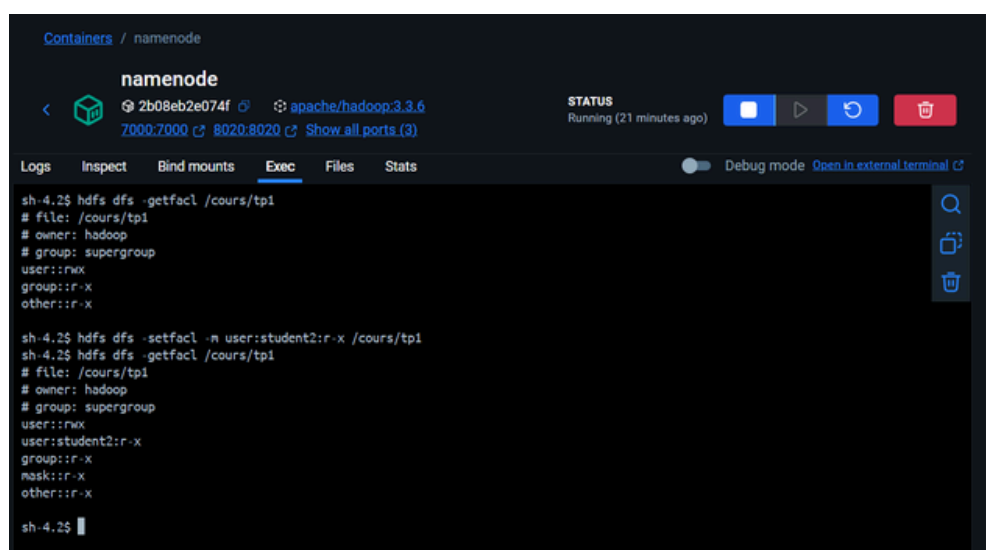
```
namenode
2b08eb2e074f apache/hadoop:3.3.6
7000:7000 8020:8020 Show all ports (3)
STATUS
Running (11 minutes ago)
Logs Inspect Bind mounts Exec Files Stats
Debug mode Open in external terminal
sh-4.2$ hdfs dfs -ls /
sh-4.2$ hdfs dfsadmin -report
Configured Capacity: 1081101176832 (1006.85 GB)
Present Capacity: 1013799010304 (944.17 GB)
DFS Remaining: 1013798981632 (944.17 GB)
DFS Used: 28672 (28 KB)
DFS Used%: 0.00%
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 (1):
RAM 3.01 GB CPU 0.42% Disk: 11.46 GB used (limit 1006.85 GB)
Update available
```

4) User Tree and Permissions



```
namenode
2b08eb2e074f apache/hadoop:3.3.6
7000:7000 8020:8020 Show all ports (3)
STATUS
Running (21 minutes ago)
Logs Inspect Bind mounts Exec Files Stats
Debug mode Open in external terminal
sh-4.2$ hdfs dfs -mkdir -p /courses/tp1
sh-4.2$ hdfs dfs -mkdir -p /user/student1 /user/student2
sh-4.2$ hdfs dfs -chmod 755 /user /courses
sh-4.2$ hdfs dfs -chmod 700 /user/student1 /user/student2
sh-4.2$ cat > /tmp/sales.csv << 'EOF'
> id,client,date,total
> 101,2025-09-01,120.50
> 102,2025-09-01,45.00
> 101,2025-09-02,33.20
> 103,2025-09-02,250.00
> EOF
sh-4.2$ hdfs dfs -put -f /tmp/sales.csv /courses/tp1/
sh-4.2$ hdfs dfs -ls -h /courses/tp1
Found 1 items
-rw-r--r-- 3 hadoop supergroup 107 2026-02-23 15:34 /courses/tp1/sales.csv
sh-4.2$
```

5) ACLs (Fine-Grained Access Control) on HDFS



```
Containers / namenode
namenode
2b08eb2e074f apache/hadoop:3.3.6
7000:7000 8020:8020 Show all ports (3)
STATUS
Running (21 minutes ago)
Logs Inspect Bind mounts Exec Files Stats
Debug mode Open in external terminal
sh-4.2$ hdfs dfs -getfacl /courses/tp1
# file: /courses/tp1
# owner: hadoop
# group: supergroup
user::rwx
group::r-x
other::r-x
sh-4.2$ hdfs dfs -setfacl -m user:student2:r-x /courses/tp1
sh-4.2$ hdfs dfs -getfacl /courses/tp1
# file: /courses/tp1
# owner: hadoop
# group: supergroup
user::rwx
user:student2:r-x
group::r-x
mask::r-x
other::r-x
sh-4.2$
```

6) Snapshots (Logical Directory Backup)

```
namenode
2b08eb2e074f apache/hadoop:3.3.6
7000:7000 8020:8020 Show all ports (3)
STATUS
Running (21 minutes ago)
Logs Inspect Bind mounts Exec Files Stats
Debug mode Open in external terminal

sh-4.2$ hdfs dfsadmin -allowSnapshot /cours
Allowing snapshot on /cours succeeded
sh-4.2$ SNAP="pre_modif_$(date +%Y%m%d_%H%M%S)"
sh-4.2$ hdfs dfs -createSnapshot /cours "$SNAP"
Created snapshot /cours/.snapshot/pre_modif_20260223_153854
sh-4.2$ hdfs dfs -ls /cours/.snapshot
Found 1 items
drwxr-xr-x 1 hadoop supergroup 0 2026-02-23 15:39 /cours/.snapshot/pre_modif_20260223_153854
sh-4.2$ hdfs dfs -rm /cours/tp1/sales.csv
Deleted /cours/tp1/sales.csv
sh-4.2$ hdfs dfs -ls /cours/tp1
sh-4.2$ hdfs dfs -cp /cours/.snapshot/$SNAP/tp1/sales.csv /cours/tp1/
sh-4.2$ hdfs dfs -ls /cours/tp1
Found 1 items
-rw-r--r-- 3 hadoop supergroup 107 2026-02-23 15:40 /cours/tp1/sales.csv
sh-4.2$
```

7) Quotas (Space & File Limits)

```
sh-4.2$ hdfs dfsadmin -setSpaceQuota 1m /user/student1
sh-4.2$ hdfs dfsadmin -setQuota 100 /user/student1
sh-4.2$ hdfs dfs -count -q -h /user/student1
100 99 1 M 1 0 0 /user/student1
sh-4.2$ dd if=/dev/zero of=/tmp/bigfile.bin bs=64K count=40 # ~2.5 MB
40+0 records in
40+0 records out
2621440 bytes (2.6 MB) copied, 0.00118465 s, 2.2 GB/s
sh-4.2$ hdfs dfs -put /tmp/bigfile.bin /user/student1/ || echo "Expected failure"
put: The DiskSpace quota of /user/student1 is exceeded: quota = 1048576 B = 1 MB but disk space consumed = 402653184 B = 384 MB
Expected failure
(quota exceeded)
sh-4.2$ hdfs dfsadmin -clrSpaceQuota /user/student1
sh-4.2$ hdfs dfsadmin -clrQuota /user/student1
sh-4.2$
```

8) Replication Factor & Data Integrity

```
namenode
2b08eb2e074f apache/hadoop:3.3.6
7000:7000 8020:8020 Show all ports (3)
STATUS
Running (21 minutes ago)
Logs Inspect Bind mounts Exec Files Stats
Debug mode Open in external terminal

sh-4.2$ hdfs getconf -confKey dfs.replication || true
3
sh-4.2$ hdfs dfs -stat %r /cours/tp1/sales.csv
3
sh-4.2$ hdfs dfs -setrep 2 /cours/tp1/sales.csv
Replication 2 set: /cours/tp1/sales.csv
sh-4.2$ hdfs dfs -stat %r /cours/tp1/sales.csv
2
sh-4.2$ hdfs balancer -threshold 10
2026-02-23 15:50:00 INFO Balancer:938 - Using a threshold of 10.0
2026-02-23 15:50:00 INFO Balancer:738 - namenodes = [hdfs://namenode:8020]

sh-4.2$ hdfs fsck / -files -blocks -locations | head -n 50
Connecting to namenode via http://namenode:9870/fsck?ugi=hadoop&files=1&blocks=1&locations=1&path=%2F
FSCK started by hadoop (auth:SIMPLE) from /172.18.0.7 for path / at Mon Feb 23 15:51:16 UTC 2026

/ <dir>
/cours <dir>
/cours/tp1 <dir>
/cours/tp1/sales.csv 107 bytes, replicated: replication=2, 1 block(s): Under replicated BP-466312492-172.18.0.7-1771859526502:b1k_1073741826_1002. Target Replicas is 2 but found 1 live replica(s), 0 decommissioned replica(s), 0 decommissioning replica(s).
0. BP-466312492-172.18.0.7-1771859526502:b1k_1073741826_1002 len=107 Live_repl=1 [DataNodeInfoWithStorage[172.18.0.6:9866,DS-7a842416-5e72-4267-bea9-9e4d8ca07af2,DISK]]

/system <dir>
/user <dir>
/user/student1 <dir>
/user/student2 <dir>

Status: HEALTHY
Number of data-nodes: 1
Number of racks: 1
Total dirs: 7
Total symlinks: 0

Replicated Blocks:
Total size: 107 B
Total files: 1
Total blocks (validated): 1 (avg. block size 107 B)
Minimally replicated blocks: 1 (100.0 %)
Over-replicated blocks: 0 (0.0 %)
Under-replicated blocks: 1 (100.0 %)
MIs-replicated blocks: 0 (0.0 %)
Default replication factor: 3
Average block replication: 1.0
Missing blocks: 0
Corrupt blocks: 0
```

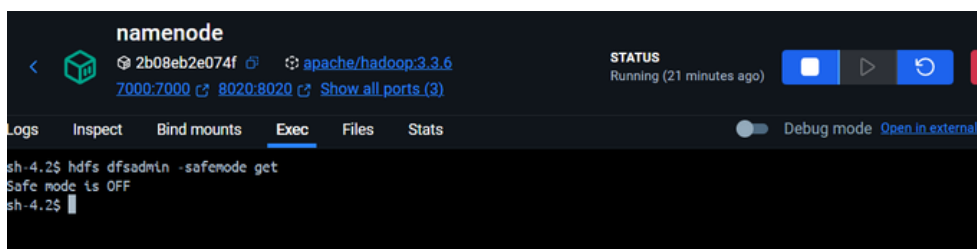
```
Missing replicas: 1 (50.0 %)
Blocks queued for replication: 0

Erasure Coded Block Groups:
Total size: 0 B
Total files: 0
Total block groups (validated): 0
Minimally erasure-coded block groups: 0
Over-erasure-coded block groups: 0
Under-erasure-coded block groups: 0

Unsatisfactory placement block groups: 0
Average block group size: 0.0
Missing block groups: 0
Corrupt block groups: 0
Missing internal blocks: 0
Blocks queued for replication: 0
FSCK ended at Mon Feb 23 15:51:16 UTC 2026 in 4 milliseconds

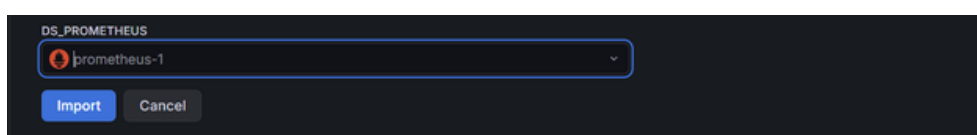
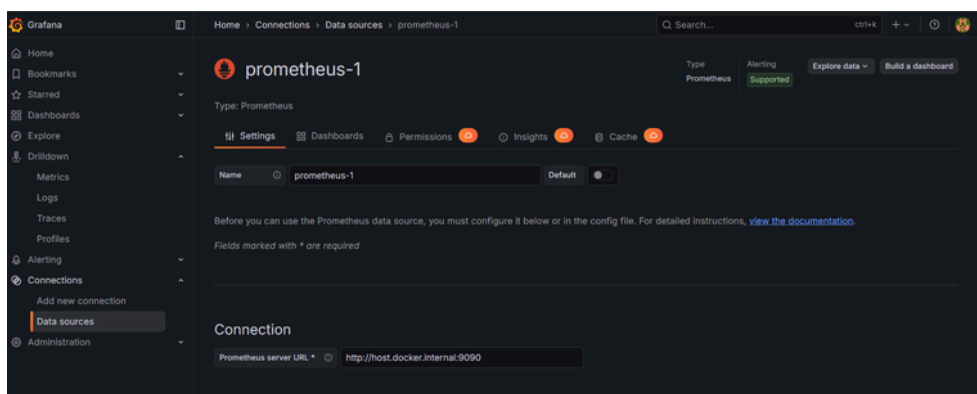
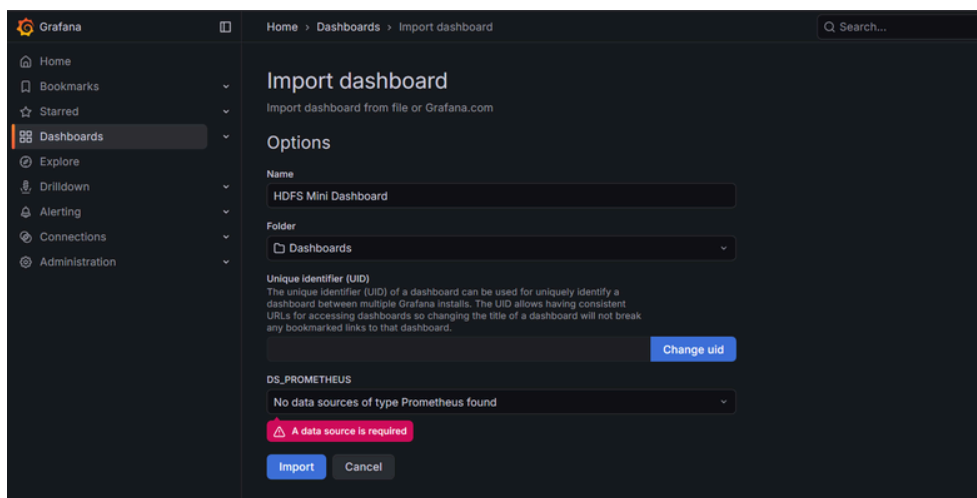
sh-4.2$
```

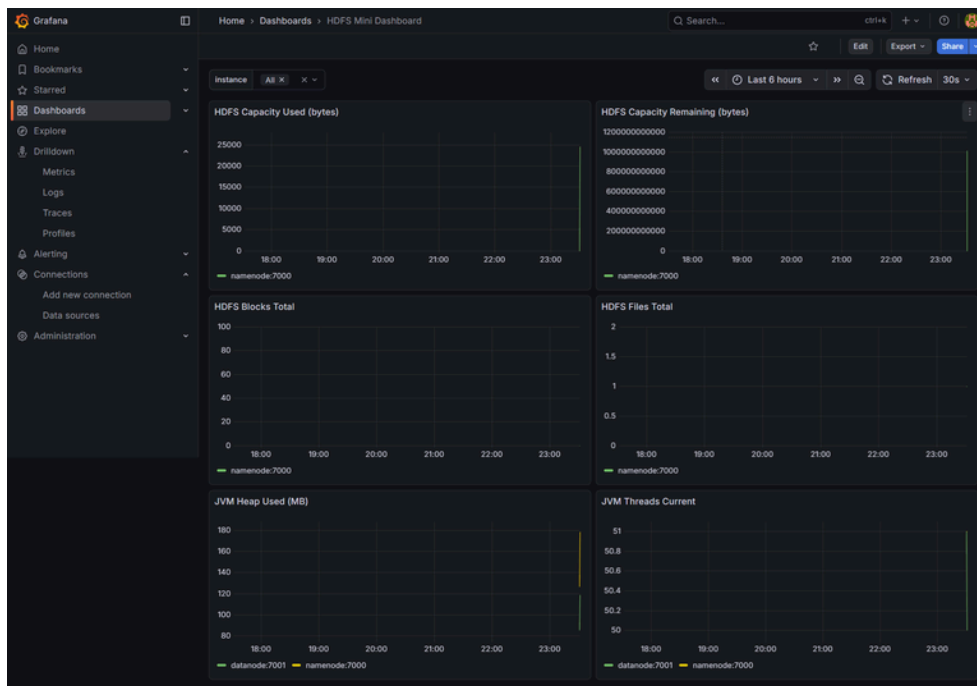
9) Safemode and Admin Operations



The screenshot shows the namenode web interface. The 'Exec' tab is active, displaying a terminal session where the command `hdfs dfsadmin -safemode get` was executed, resulting in the output `Safe mode is OFF`. The interface also shows the namenode's status as 'Running (21 minutes ago)' and various tabs like 'Logs', 'Inspect', 'Bind mounts', 'Exec', 'Files', and 'Stats'.

10) Cluster Monitoring with Prometheus and Grafana





namenode

4ed0fb8d657 [apache/hadoop:3.3.6](#)

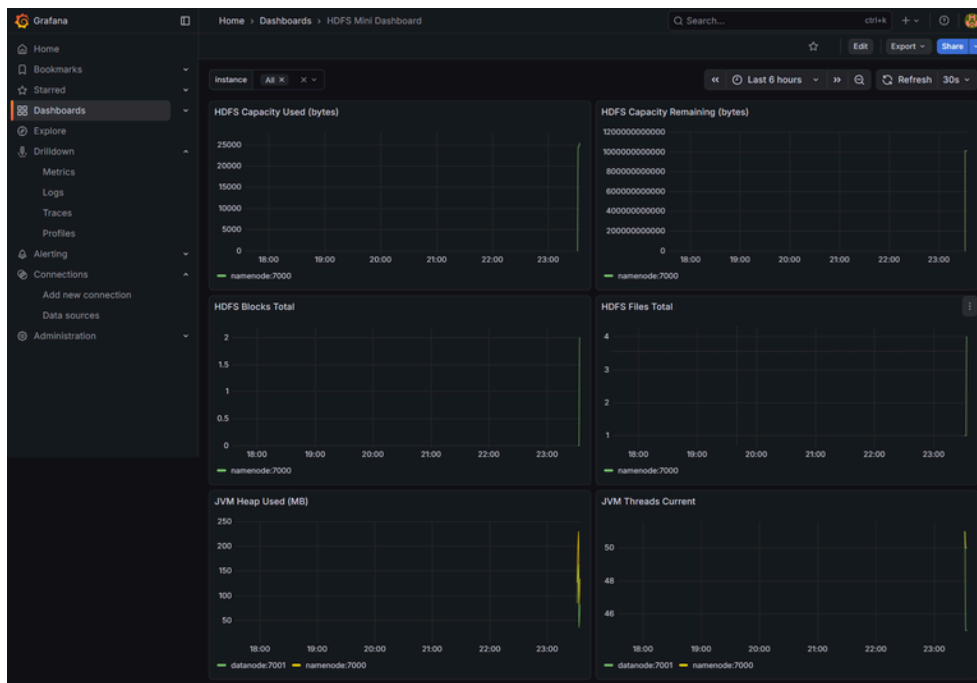
7000:7000 [8020:8020](#) [Show all ports \(3\)](#)

STATUS
Running (1 minute ago)

Logs Inspect Bind mounts **Exec** Files Stats

Debug mode [Open in external terminal](#)

```
sh-4.2$ hdfs dfs -mkdir /test
sh-4.2$ hdfs dfs -put /etc/passwd /test/
sh-4.2$ hdfs dfs -put /etc/hosts /test/
sh-4.2$
```

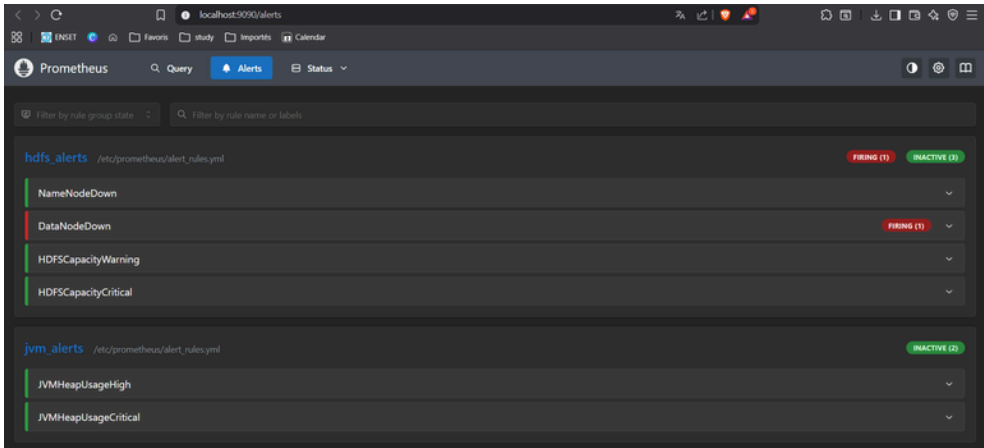


11) Prometheus Alert System Integration

```
[+] Running 8/8
✔ Network tp3_hadoopnet          Created          0.0s
✔ Container tp3-resourcemanager-1 Started         0.5s
✔ Container namenode              Started         0.7s
✔ Container datanode              Started         0.7s
✔ Container alertmanager          Started         0.6s
✔ Container tp3-nodemanager-1     Started         0.6s
✔ Container prometheus            Started         0.7s
✔ Container grafana                Started         0.8s

C:\Users\NITRO 5\Desktop\S4\Big Data\tp\tp3>
```

```
C:\Users\NITRO 5\Desktop\S4\Big Data\tp\tp3>docker stop datanode
datanode
```



```
C:\Users\NITRO 5\Desktop\S4\Big Data\tp\tp3>docker start datanode
datanode
```

