

TP 3: HDFS File System Administration

Rapport des TPs - Big Data

1) Launching the Cluster

2) Quick Checks via Web UI

localhost:8088/cluster

ENSET Favorites study Imports Calendar

All Applications

hadoop

Cluster Metrics							
	Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Used Resources	Memory
About	0	0	0	0	0	<memory:0 B, vCores:0>	<memory:8
Nodes							
Node Labels							
Applications							
NEW							
NEW_SAVING							
SUBMITTED							
ACCEPTED							
RUNNING							
FINISHED							
FAILED							
KILLED							
Scheduler							
Tools							

Cluster Nodes Metrics

Active Nodes		Decommissioning Nodes	Decommissioned Nodes	Lost Nodes
1	0	0	0	0

Scheduler Metrics

Scheduler Type		Scheduling Resource Type	Minimum Allocation	Maximum Allocation
Capacity Scheduler	[memory-mb (unit=M), vcores]	<memory:1024, vCores:1>	<memory:8192, vCores:4>	<memory:8192, vCores:4>

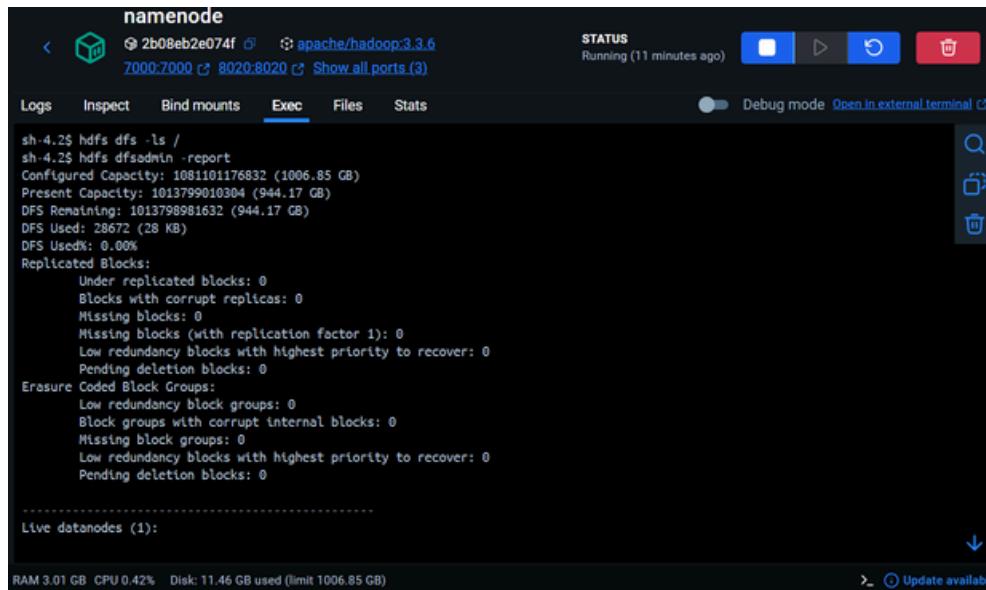
Show 20 entries

ID	User	Name	Application Type	Application Tags	Queue	Application Priority	StartTime	LaunchTime	FinishTime	State	FinalStatus	Running Containers	Allocated Capacity
No data available in table													

Showing 0 to 0 of 0 entries

Overview 'namenode:8020' (✓active)	
Started:	Mon Feb 23 15:12:00 +0000 2026
Version:	3.3.6, r1be78238728da9266a4f888195058f08d012bf8c
Compiled:	Sun Jun 18 09:22:00 +0100 2023 by ubuntu from (HEAD detached at release-3.3.6-RC1)
Cluster ID:	CID-4ef24449-9857-4fd5-9bac-34034dec0195
Block Pool ID:	BP-466312492-172.18.0.7-1771859526502

3) Admin Shell in the NameNode

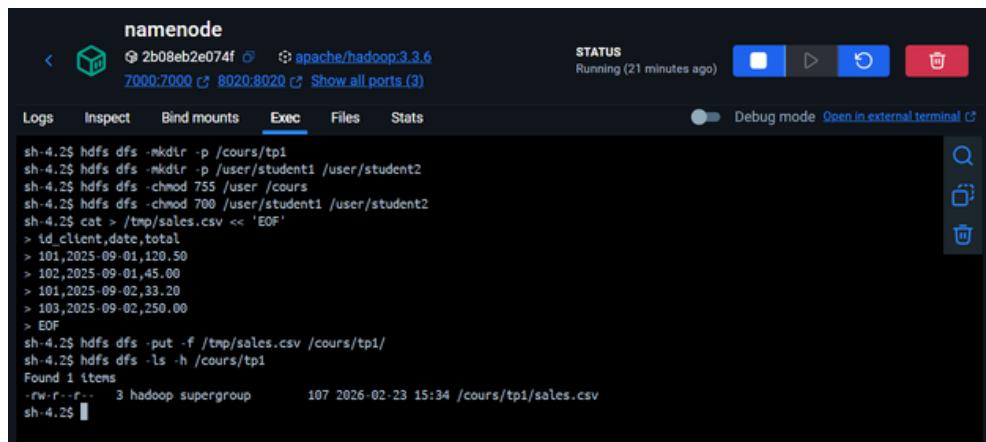


```
namenode
2b08eb2e074f apache/hadoop:3.3.6
7000:7000 8020:8020 Show all ports (3)

Logs Inspect Bind mounts Exec Files Stats STATUS Running (11 minutes ago)
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

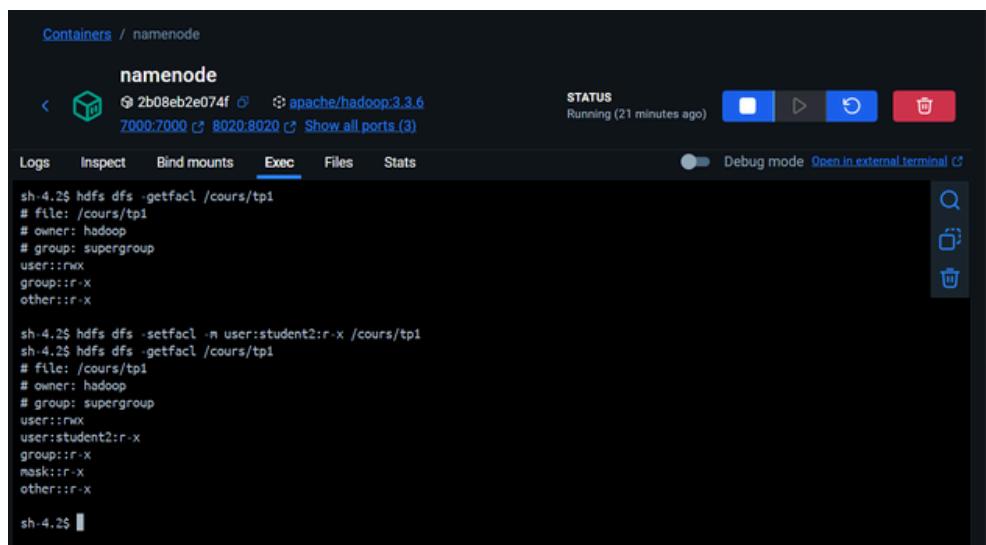


```
namenode
2b08eb2e074f apache/hadoop:3.3.6
7000:7000 8020:8020 Show all ports (3)

Logs Inspect Bind mounts Exec Files Stats STATUS Running (21 minutes ago)
Debug mode Open in external terminal

sh-4.2$ hdfs dfs -mkdirr -p /cours/tp1
sh-4.2$ hdfs dfs -mkdirr -p /user/student1 /user/student2
sh-4.2$ hdfs dfs -chmod 755 /user /cours
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 /cours/tp1/
sh-4.2$ hdfs dfs -ls -h /cours/tp1
Found 1 items
-rw-r--r-- 3 hadoop supergroup 107 2026-02-23 15:34 /cours/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)

Logs Inspect Bind mounts Exec Files Stats STATUS Running (21 minutes ago)
Debug mode Open in external terminal

sh-4.2$ hdfs dfs -getfacl /cours/tp1
# file: /cours/tp1
# owner: hadoop
# group: supergroup
user::rwx
group::r-x
other::r-X

sh-4.2$ hdfs dfs -setfacl -m user:student2:r-x /cours/tp1
sh-4.2$ hdfs dfs -getfacl /cours/tp1
# file: /cours/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)

The screenshot shows a terminal window titled "namenode" with the following command history:

```
sh-4.2$ hdfs dfsadmin -allowSnapshot /cours
Allowing snapshot on /cours succeeded
sh-4.2$ SNAP="`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 - hadoop supergroup          0 2026-02-23 15:39 /cours/.snapshot/pre_modif_20260223_153854
sh-4.2$ hdfs dfs -rm /cours/tpl/sales.csv
Deleted /cours/tpl/sales.csv
sh-4.2$ hdfs dfs -ls /cours/tpl
sh-4.2$ hdfs dfs -cp /cours/.snapshot/$SNAP/tpl/sales.csv /cours/tpl/
sh-4.2$ hdfs dfs -ls /cours/tpl
Found 1 items
-rw-r--r-- 3 hadoop supergroup      107 2026-02-23 15:40 /cours/tpl/sales.csv
sh-4.2$
```

7) Quotas (Space & File Limits)

The screenshot shows a terminal window with the following command history:

```
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 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
> (quota exceeded)"
put: The DiskSpace quota of /user/student1 is exceeded: quota = 1048576 B = 1 MB but diskspace 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

The screenshot shows a terminal window titled "namenode" with the following command history:

```
sh-4.2$ hdfs getconf -confKey dfs.replication || true
3
sh-4.2$ hdfs dfs -stat Xr /cours/tpl/sales.csv
3
sh-4.2$ hdfs dfs -setrep 2 /cours/tpl/sales.csv
Replication 2 set: /cours/tpl/sales.csv
sh-4.2$ hdfs dfs -stat Xr /cours/tpl/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/tpl <dir>
/cours/tpl/sales.csv 107 bytes, replicated: replication=2, 1 block(s): Under replicated BP-466312492-172.18.0.7-1771859526502:b
lk_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:blk_1073741826_1002 len=107 Live_repl=1 [DatanodeInfoWithStorage[172.18.0.6:9866,DS-7a
842416-5e72-4267-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$ 

```

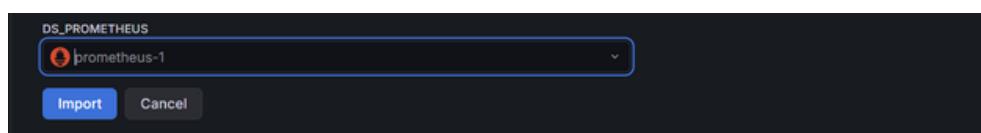
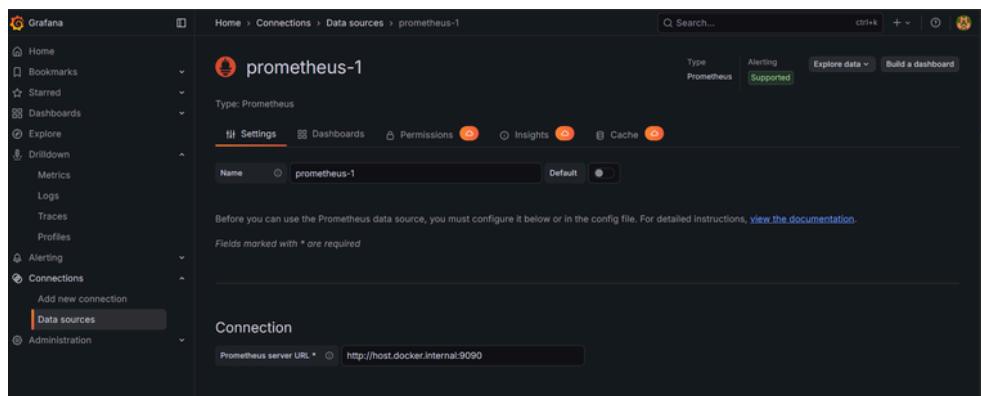
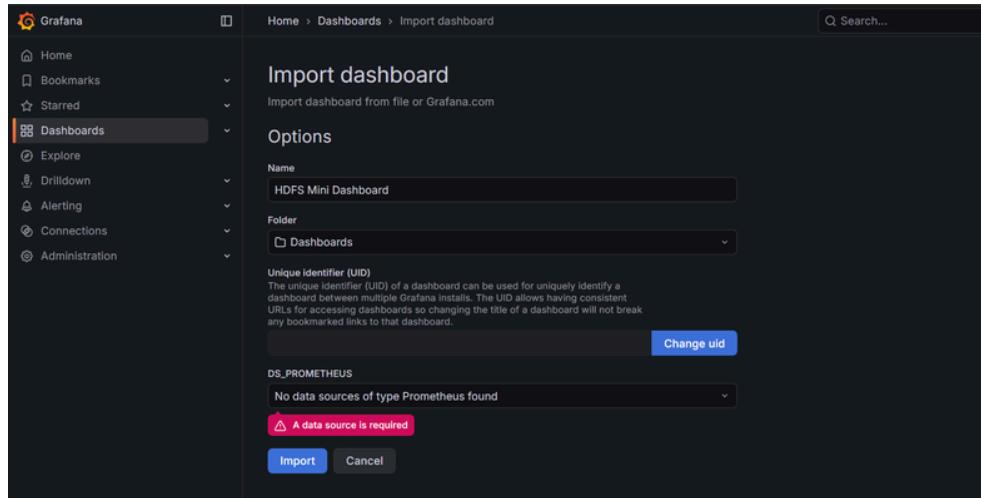
9) Safemode and Admin Operations

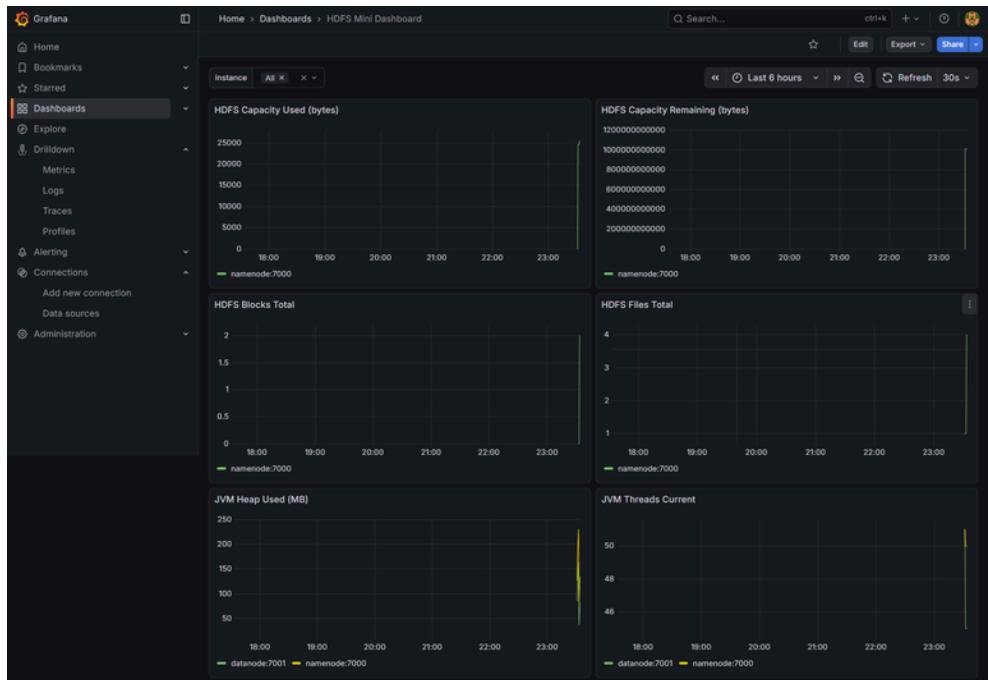
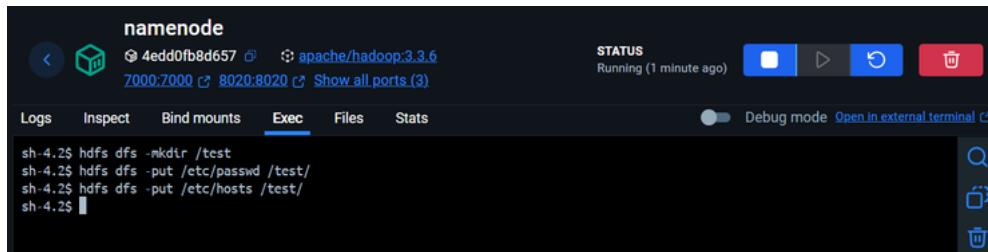
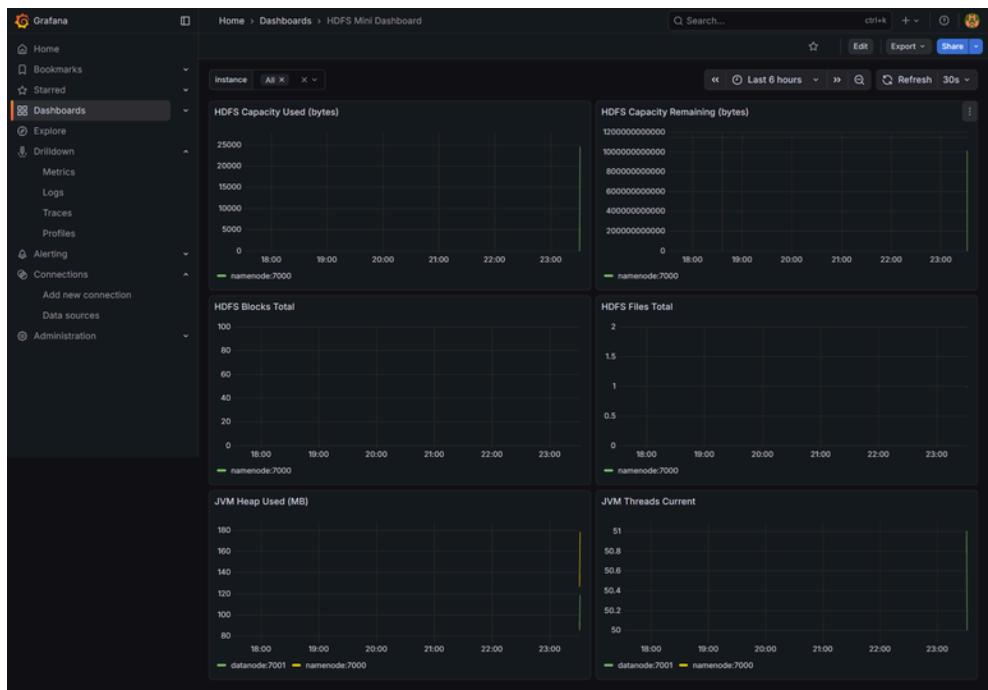
namenode
↳ 2b08eb2e074f ⚡ apache/hadoop:3.3.6
7000:7000 ⚡ 8020:8020 ⚡ Show all ports (3)

Logs Inspect Bind mounts Exec Files Stats

sh-4.2\$ hdfs dfsadmin -safemode get
Safe Mode is OFF
sh-4.2\$

10) Cluster Monitoring with Prometheus and Grafana





11) Prometheus Alert System Integration

```
[+] * Running 8/8
✓ Network tp3.hadoopnet      Created          0.0s
✓ Container tp3-resourcemanager-1 Started        0.5s
✓ Container namenode         Started        0.75s
✓ Container datanode         Started        0.76s
✓ Container alertmanager     Started        0.65s
✓ Container tp3-nodemanager-1 Started        0.65s
✓ Container prometheus       Started        0.75s
✓ 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
```

The screenshot shows the Prometheus web interface at `localhost:9090/alerts`. The 'Alerts' tab is selected. There are two main sections: 'hdfs_alerts' and 'jvm_alerts'. In the 'hdfs_alerts' section, there are two active alerts: 'NameNodeDown' and 'DataNodeDown', each with a red warning icon and a count of 1 under 'FIRING'. In the 'jvm_alerts' section, there are two inactive alerts: 'JVMHeapUsageHigh' and 'JVMHeapUsageCritical', each with a green info icon and a count of 0 under 'INACTIVE'.

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

This screenshot shows the same Prometheus interface after the 'datanode' container was restarted. The 'Alerts' tab is still selected. The 'hdfs_alerts' and 'jvm_alerts' sections now both show 'INACTIVE (0)' for all items, with no active alerts displayed.