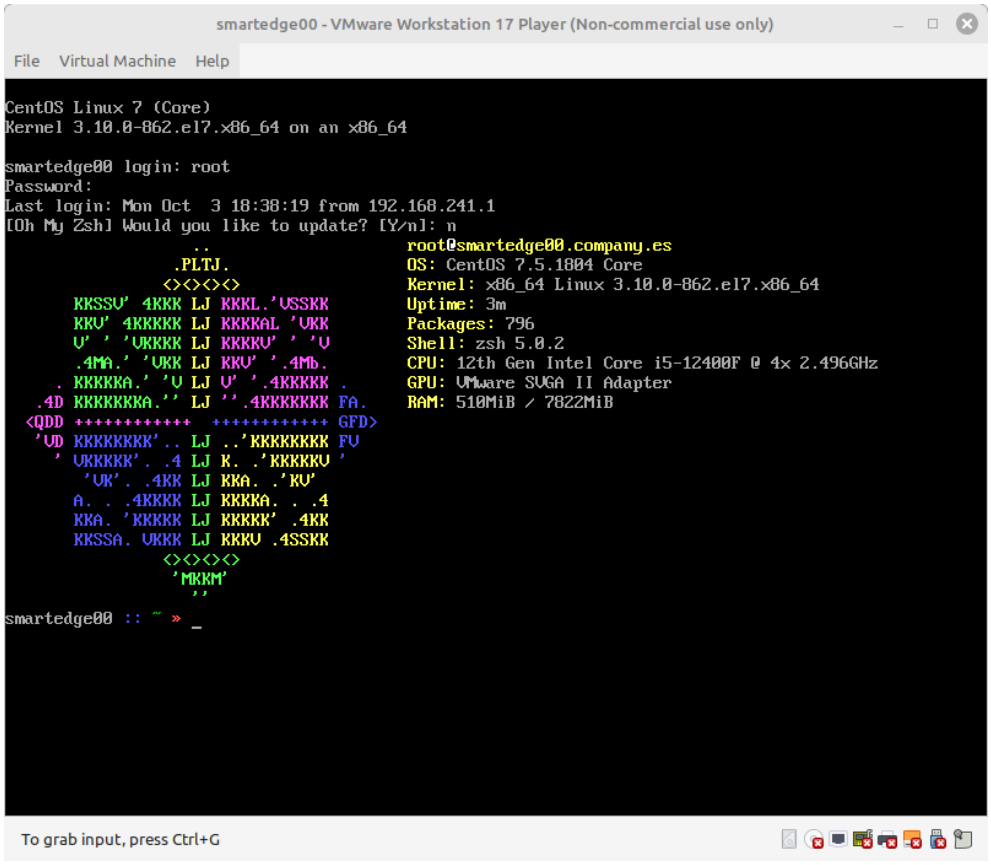
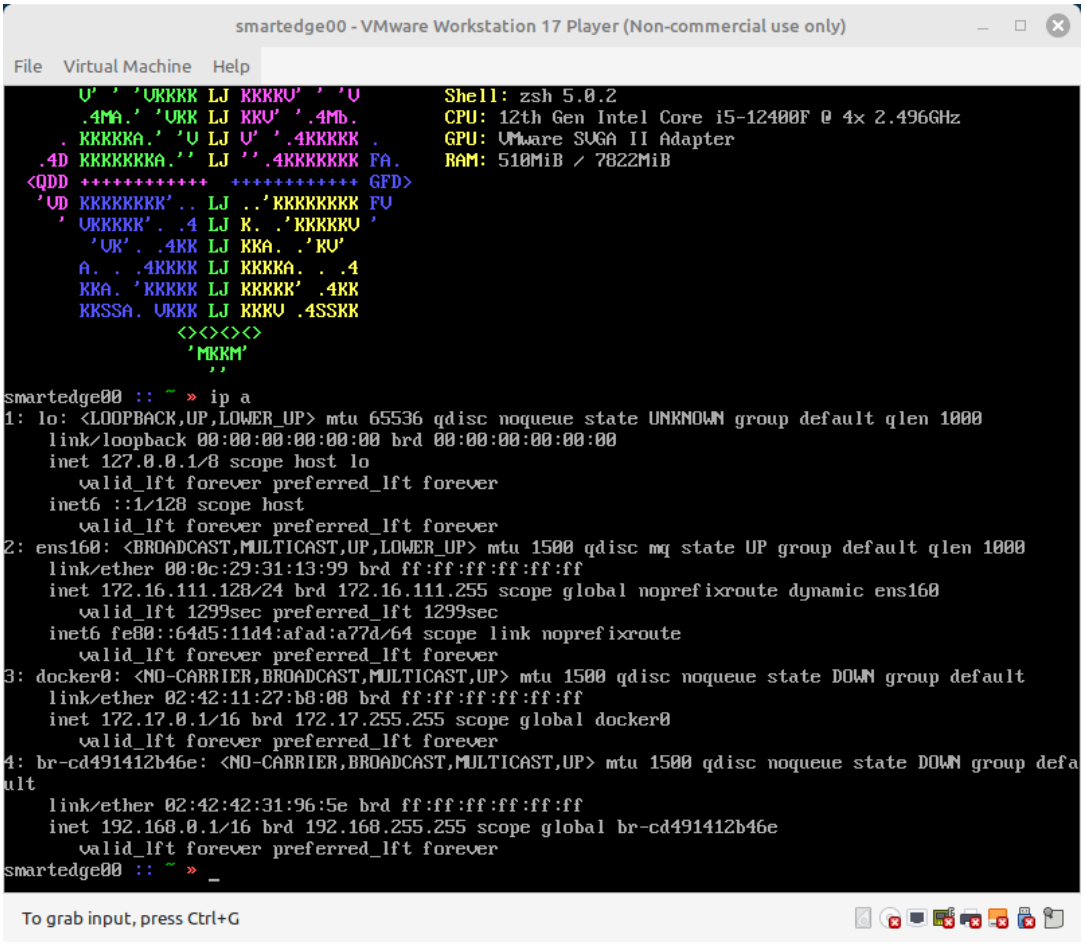
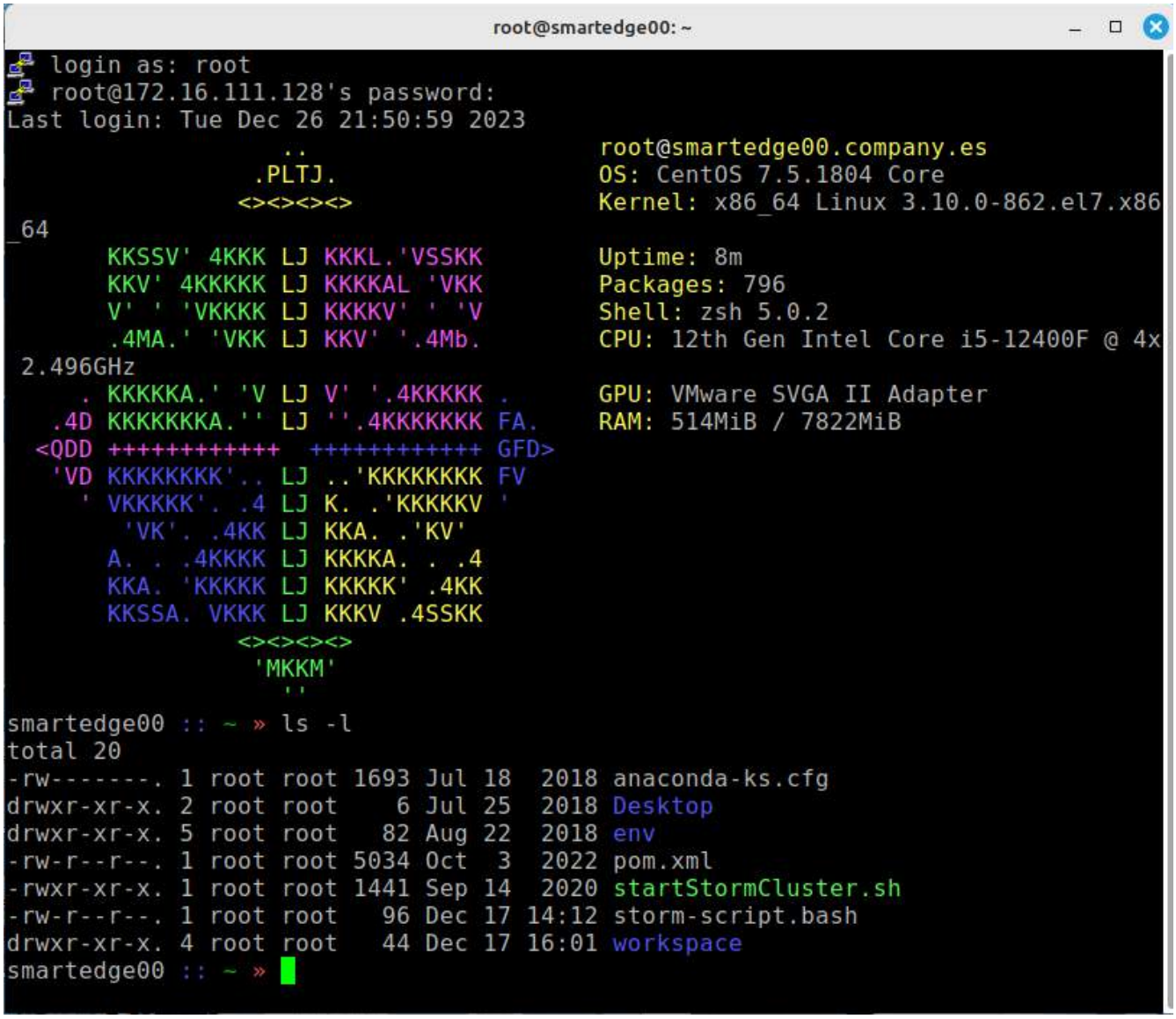
LAB 1

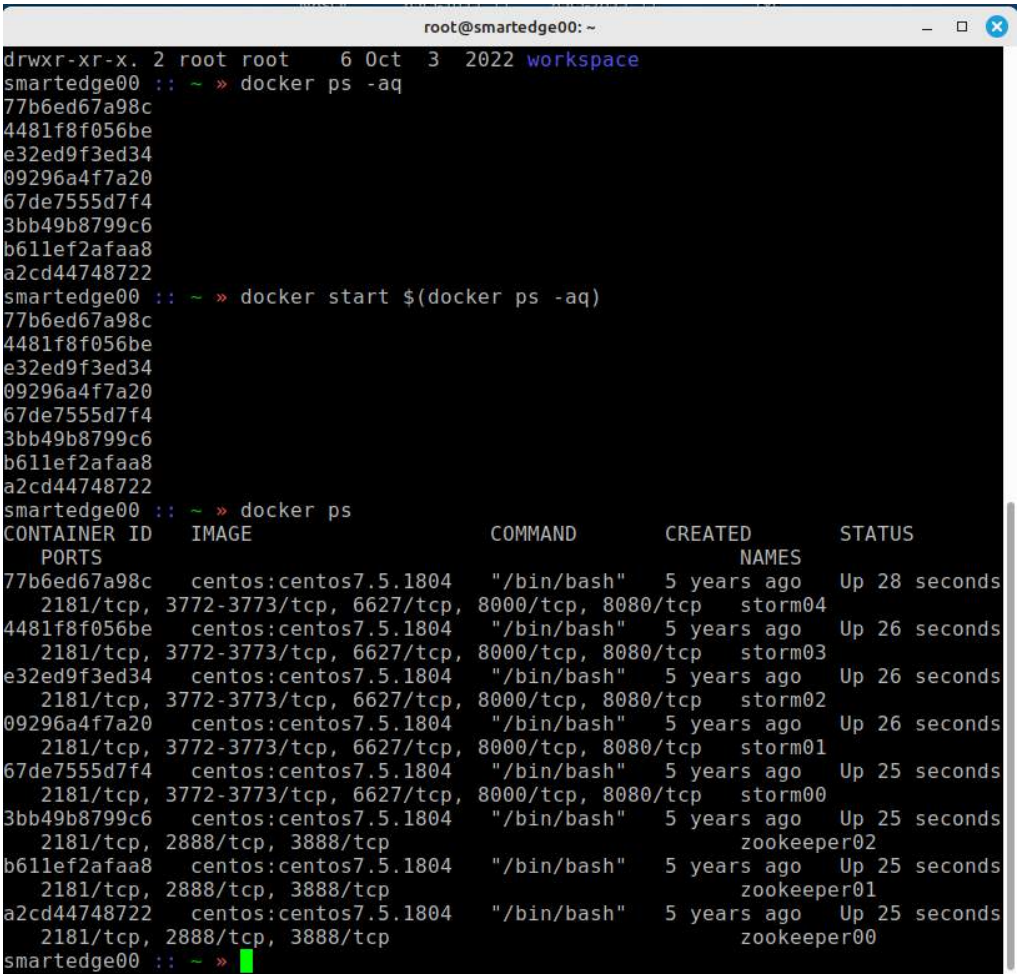

smartedge00 - VMware Workstation 17 Player (Non-commercial use only)
File Virtual Machine Help
CentOS Linux 7 (Core)
Kernel 3.10.0-862.el7.x86_64 on an x86_64
smartedge00 login: root
Password:
Last login: Mon Oct 3 18:38:19 from 192.168.241.1
[Oh My Zsh] Would you like to update? [Y/n]: n
root@smartedge00.company.es
OS: CentOS 7.5.1804 Core
Kernel: x86_64 Linux 3.10.0-862.el7.x86_64
Uptime: 3m
Packages: 796
Shell: zsh 5.0.2
CPU: 12th Gen Intel Core i5-12400F @ 4x 2.496GHz
GPU: VMware SVGA II Adapter
RAM: 510MiB / 7822MiB
smartedge00 :: ~ » _
To grab input, press Ctrl+G


smartedge00 - VMware Workstation 17 Player (Non-commercial use only)
File Virtual Machine Help
Shell: zsh 5.0.2
CPU: 12th Gen Intel Core i5-12400F @ 4x 2.496GHz
GPU: VMware SVGA II Adapter
RAM: 510MiB / 7822MiB
smartedge00 :: ~ » ip a
1: lo: <LOOPBACK,UP,LOWER_UP> mtu 65536 qdisc noqueue state UNKNOWN group default qlen 1000
link/loopback 00:00:00:00:00:00 brd 00:00:00:00:00:00
inet 127.0.0.1/8 scope host lo
valid_lft forever preferred_lft forever
inet6 ::1/128 scope host
valid_lft forever preferred_lft forever
2: ens160: <BROADCAST,MULTICAST,UP,LOWER_UP> mtu 1500 qdisc mq state UP group default qlen 1000
link/ether 00:0c:29:31:13:99 brd ff:ff:ff:ff:ff:ff
inet 172.16.111.128/24 brd 172.16.111.255 scope global noprefixroute dynamic ens160
valid_lft 1299sec preferred_lft 1299sec
inet6 fe80::64d5:11d4:afad:a77d/64 scope link noprefixroute
valid_lft forever preferred_lft forever
3: docker0: <NO-CARRIER,BROADCAST,MULTICAST,UP> mtu 1500 qdisc noqueue state DOWN group default
link/ether 02:42:11:27:b8:08 brd ff:ff:ff:ff:ff:ff
inet 172.17.0.1/16 brd 172.17.255.255 scope global docker0
valid_lft forever preferred_lft forever
4: br-cd491412b46e: <NO-CARRIER,BROADCAST,MULTICAST,UP> mtu 1500 qdisc noqueue state DOWN group default
link/ether 02:42:42:31:96:5e brd ff:ff:ff:ff:ff:ff
inet 192.168.0.1/16 brd 192.168.255.255 scope global br-cd491412b46e
valid_lft forever preferred_lft forever
smartedge00 :: ~ » _
To grab input, press Ctrl+G

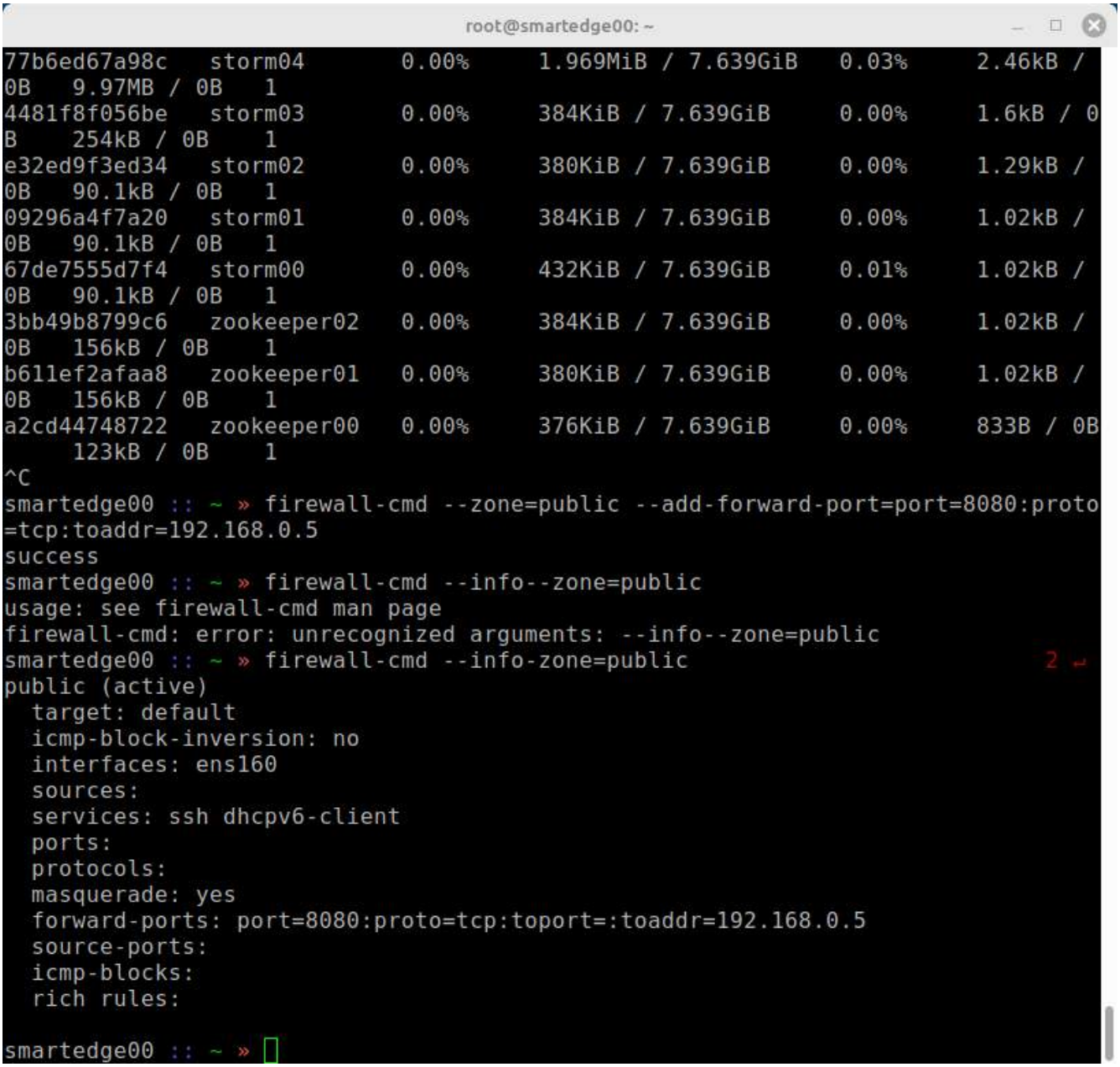
En PUTTY

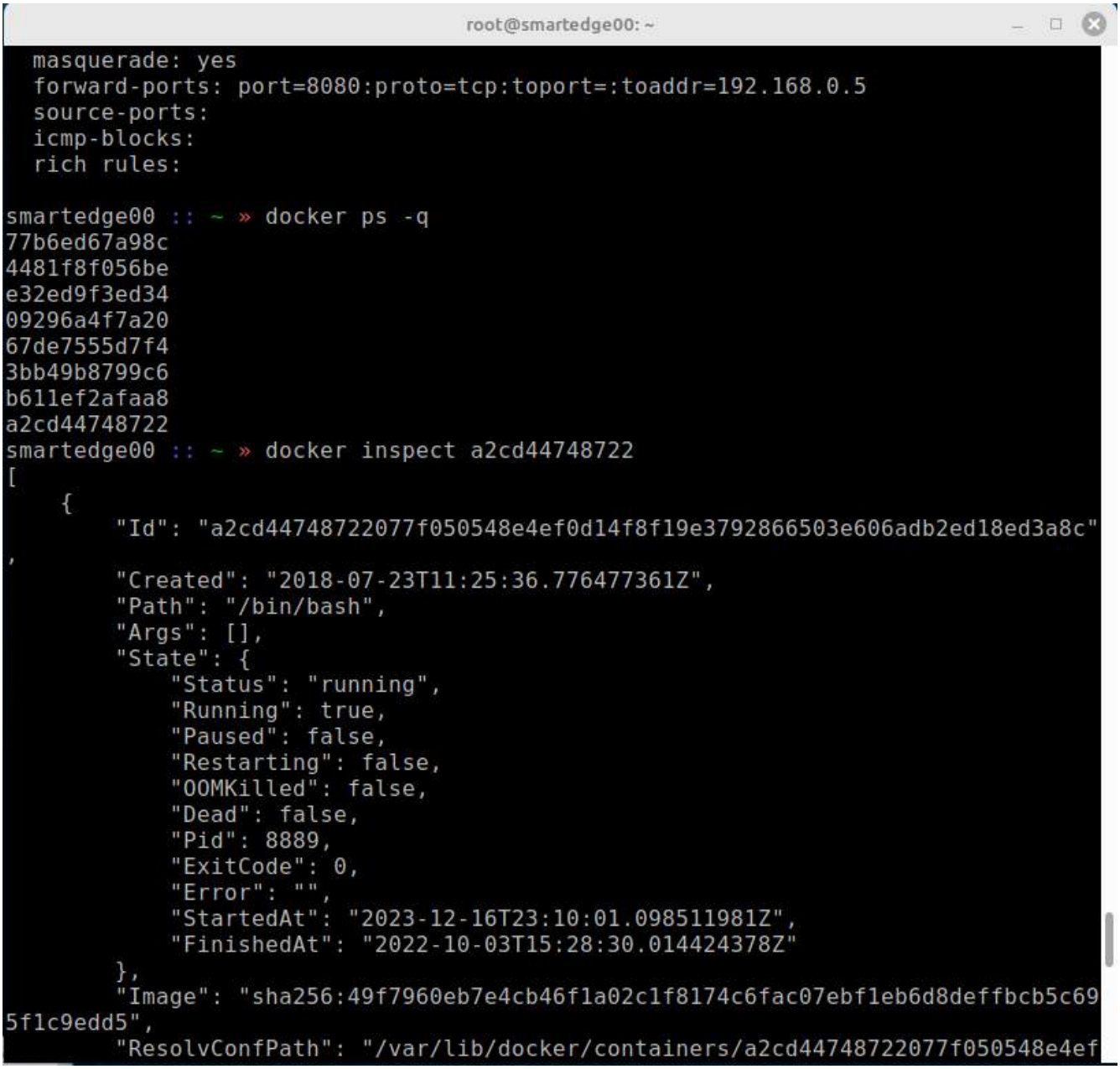

root@smartedge00: ~
login as: root
root@172.16.111.128's password:
Last login: Tue Dec 26 21:50:59 2023
root@smartedge00.company.es
OS: CentOS 7.5.1804 Core
Kernel: x86_64 Linux 3.10.0-862.el7.x86_64
Uptime: 8m
Packages: 796
Shell: zsh 5.0.2
CPU: 12th Gen Intel Core i5-12400F @ 4x 2.496GHz
GPU: VMware SVGA II Adapter
RAM: 514MiB / 7822MiB
smartedge00 :: ~ » ls -l
total 20
-rw-------. 1 root root 1693 Jul 18 2018 anaconda-ks.cfg
drwxr-xr-x. 2 root root 6 Jul 25 2018 Desktop
drwxr-xr-x. 5 root root 82 Aug 22 2018 env
-rw-r--r--. 1 root root 5034 Oct 3 2022 pom.xml
-rwxr-xr-x. 1 root root 1441 Sep 14 2020 startStormCluster.sh
-rw-r--r--. 1 root root 96 Dec 17 14:12 storm-script.bash
drwxr-xr-x. 4 root root 44 Dec 17 16:01 workspace
smartedge00 :: ~ »

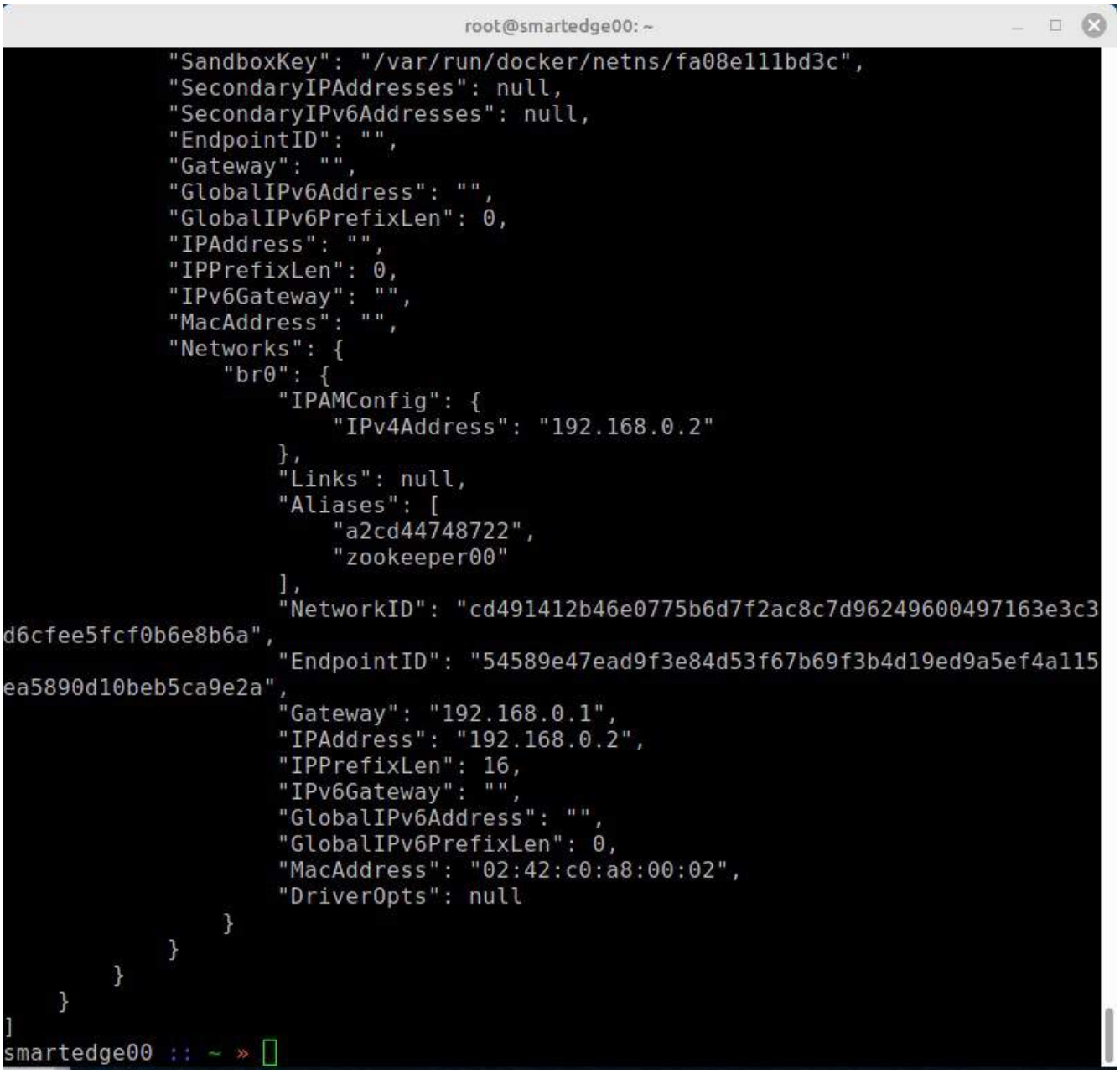
Arrancamos dockers

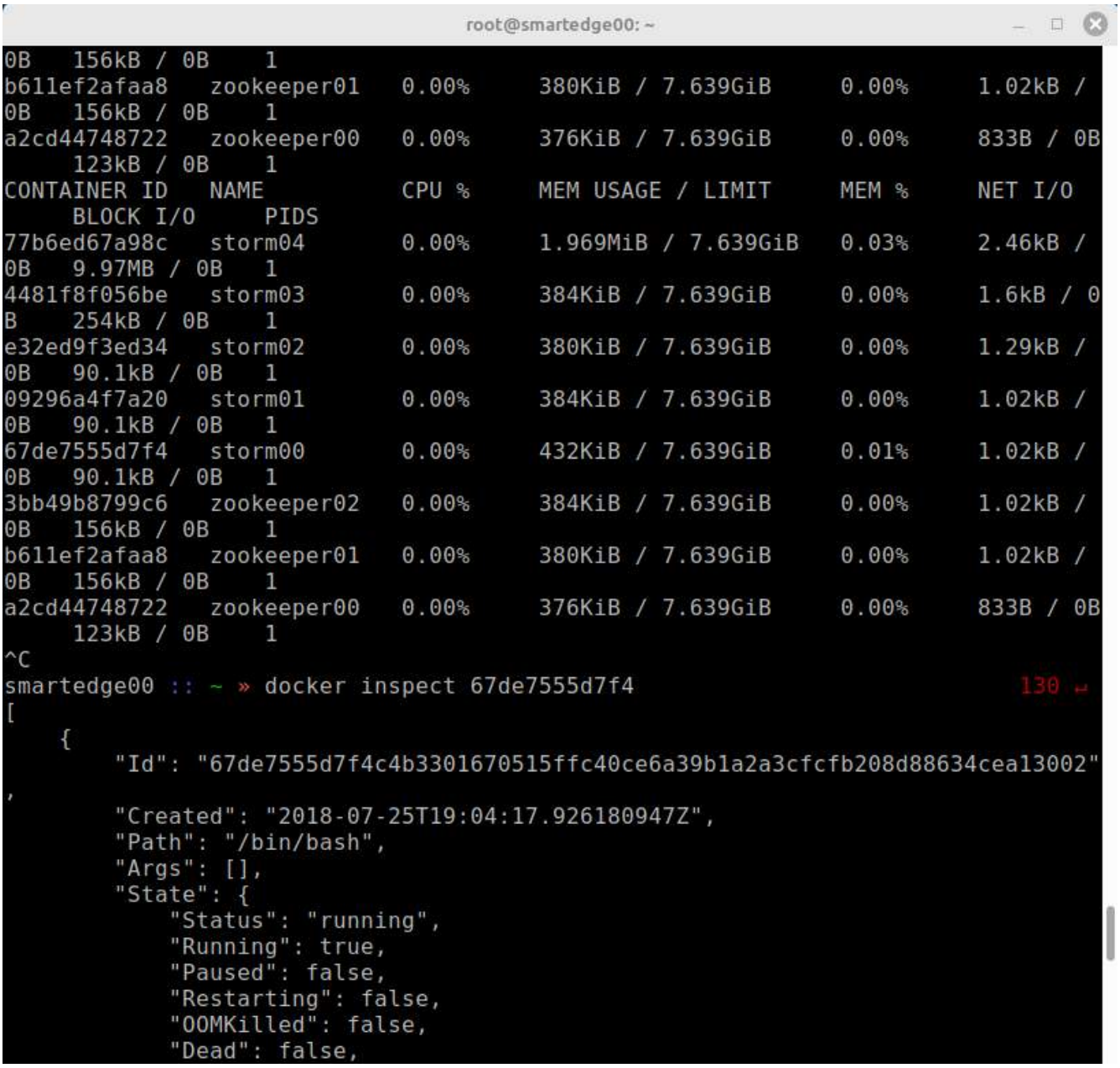


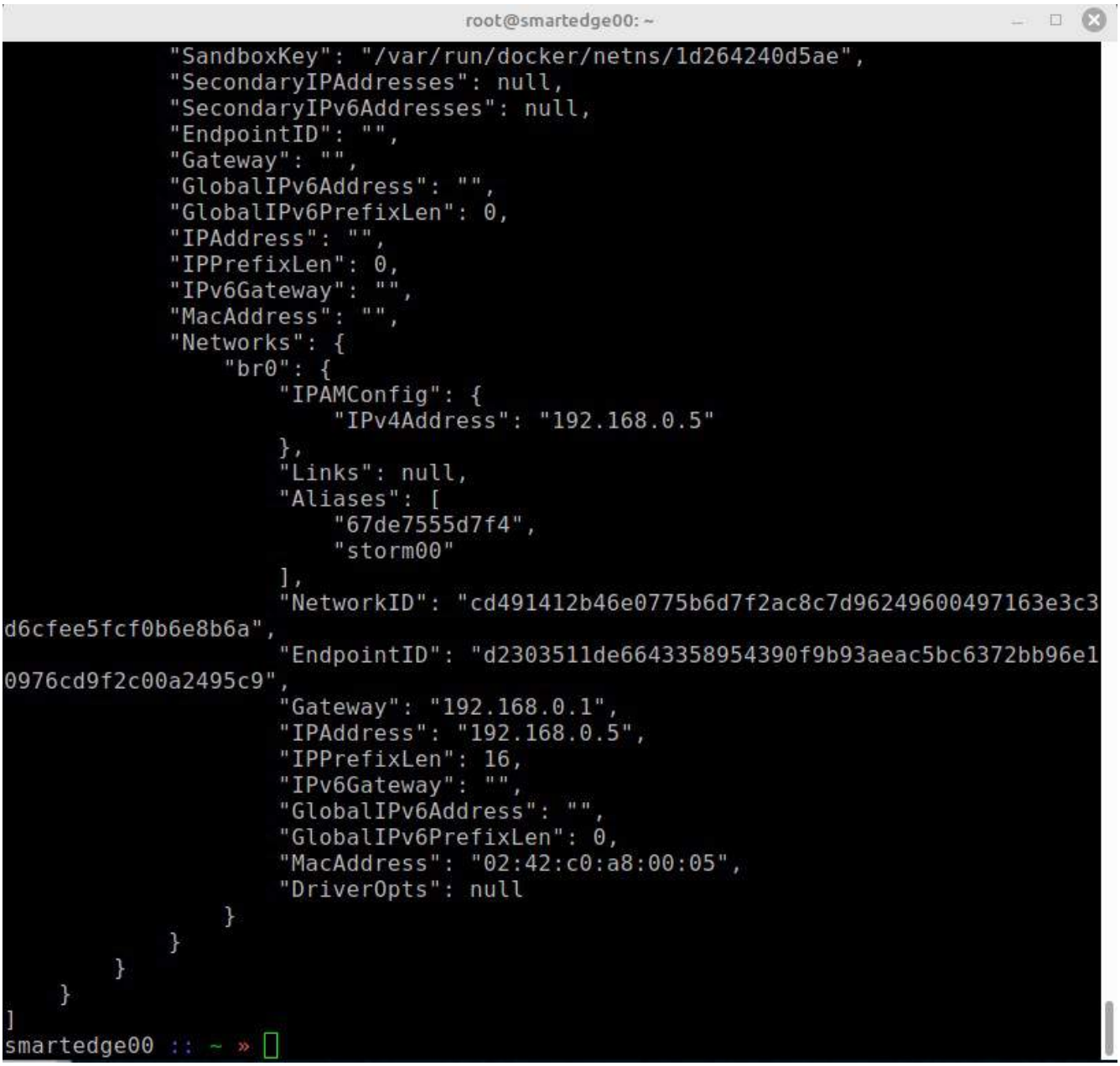
Establecemos port-forwarding



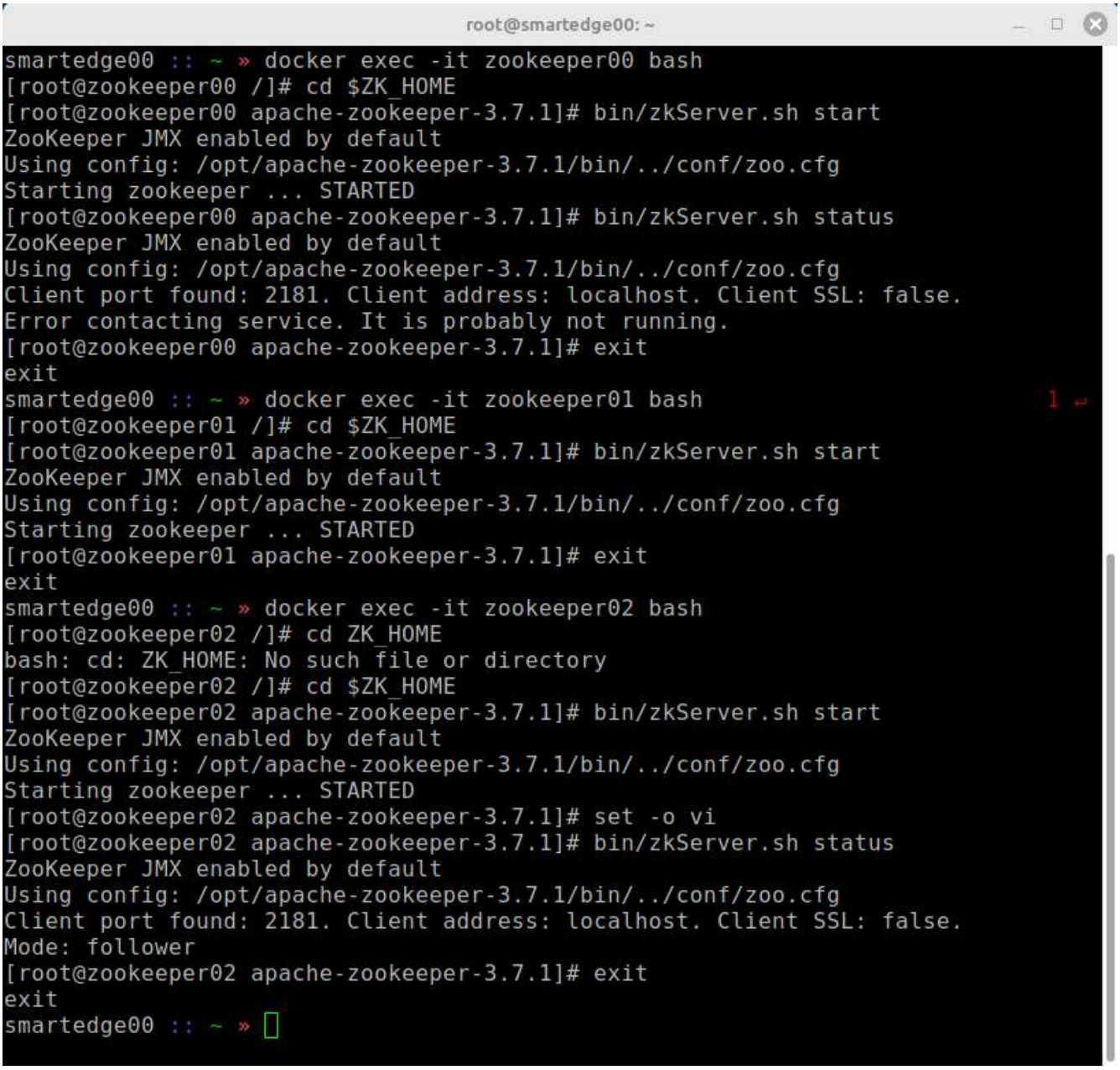

root@smartedge00: ~
masquerade: yes
forward-ports: port=8080:proto=tcp:toport=:toaddr=192.168.0.5
source-ports:
icmp-blocks:
rich rules:
smartedge00 :: ~ » docker ps -q
77b6ed67a98c
4481f8f056be
e32ed9f3ed34
09296a4f7a20
67de7555d7f4
3bb49b8799c6
b611ef2afaa8
a2cd44748722
smartedge00 :: ~ » docker inspect a2cd44748722
[
{
"Id": "a2cd44748722077f050548e4ef0d14f8f19e3792866503e606adb2ed18ed3a8c"
,
"Created": "2018-07-23T11:25:36.776477361Z",
"Path": "/bin/bash",
"Args": [],
"State": {
"Status": "running",
"Running": true,
"Paused": false,
"Restarting": false,
"OOMKilled": false,
"Dead": false,
"Pid": 8889,
"ExitCode": 0,
"Error": "",
"StartedAt": "2023-12-16T23:10:01.098511981Z",
"FinishedAt": "2022-10-03T15:28:30.014424378Z"
},
"Image": "sha256:49f7960eb7e4cb46f1a02c1f8174c6fac07ebf1eb6d8deffbcb5c69
5f1c9edd5",
"ResolvConfPath": "/var/lib/docker/containers/a2cd44748722077f050548e4ef


root@smartedge00: ~
"SandboxKey": "/var/run/docker/netns/fa08e111bd3c",
"SecondaryIPAddresses": null,
"SecondaryIPv6Addresses": null,
"EndpointID": "",
"Gateway": "",
"GlobalIPv6Address": "",
"GlobalIPv6PrefixLen": 0,
"IPAddress": "",
"IPPrefixLen": 0,
"IPv6Gateway": "",
"MacAddress": "",
"Networks": {
"br0": {
"IPAMConfig": {
"IPv4Address": "192.168.0.2"
},
"Links": null,
"Aliases": [
"a2cd44748722",
"zookeeper00"
],
"NetworkID": "cd491412b46e0775b6d7f2ac8c7d96249600497163e3c3
d6cfee5fcf0b6e8b6a",
"EndpointID": "54589e47ead9f3e84d53f67b69f3b4d19ed9a5ef4a115
ea5890d10beb5ca9e2a",
"Gateway": "192.168.0.1",
"IPAddress": "192.168.0.2",
"IPPrefixLen": 16,
"IPv6Gateway": "",
"GlobalIPv6Address": "",
"GlobalIPv6PrefixLen": 0,
"MacAddress": "02:42:c0:a8:00:02",
"DriverOpts": null
}
}
}
}
]
smartedge00 :: ~ »


root@smartedge00: ~
0B 156kB / 0B 1
b611ef2afaa8 zookeeper01 0.00% 380KiB / 7.639GiB 0.00% 1.02kB /
0B 156kB / 0B 1
a2cd44748722 zookeeper00 0.00% 376KiB / 7.639GiB 0.00% 833B / 0B
123kB / 0B 1
CONTAINER ID NAME CPU % MEM USAGE / LIMIT MEM % NET I/O
BLOCK I/O PIDS
77b6ed67a98c storm04 0.00% 1.969MiB / 7.639GiB 0.03% 2.46kB /
0B 9.97MB / 0B 1
4481f8f056be storm03 0.00% 384KiB / 7.639GiB 0.00% 1.6kB / 0
B 254kB / 0B 1
e32ed9f3ed34 storm02 0.00% 380KiB / 7.639GiB 0.00% 1.29kB /
0B 90.1kB / 0B 1
09296a4f7a20 storm01 0.00% 384KiB / 7.639GiB 0.00% 1.02kB /
0B 90.1kB / 0B 1
67de7555d7f4 storm00 0.00% 432KiB / 7.639GiB 0.01% 1.02kB /
0B 90.1kB / 0B 1
3bb49b8799c6 zookeeper02 0.00% 384KiB / 7.639GiB 0.00% 1.02kB /
0B 156kB / 0B 1
b611ef2afaa8 zookeeper01 0.00% 380KiB / 7.639GiB 0.00% 1.02kB /
0B 156kB / 0B 1
a2cd44748722 zookeeper00 0.00% 376KiB / 7.639GiB 0.00% 833B / 0B
123kB / 0B 1
^C
smartedge00 :: ~ » docker inspect 67de7555d7f4 130 ↵
[
{
"Id": "67de7555d7f4c4b3301670515ffc40ce6a39b1a2a3cfcfb208d88634cea13002"
,
"Created": "2018-07-25T19:04:17.926180947Z",
"Path": "/bin/bash",
"Args": [],
"State": {
"Status": "running",
"Running": true,
"Paused": false,
"Restarting": false,
"OOMKilled": false,
"Dead": false,

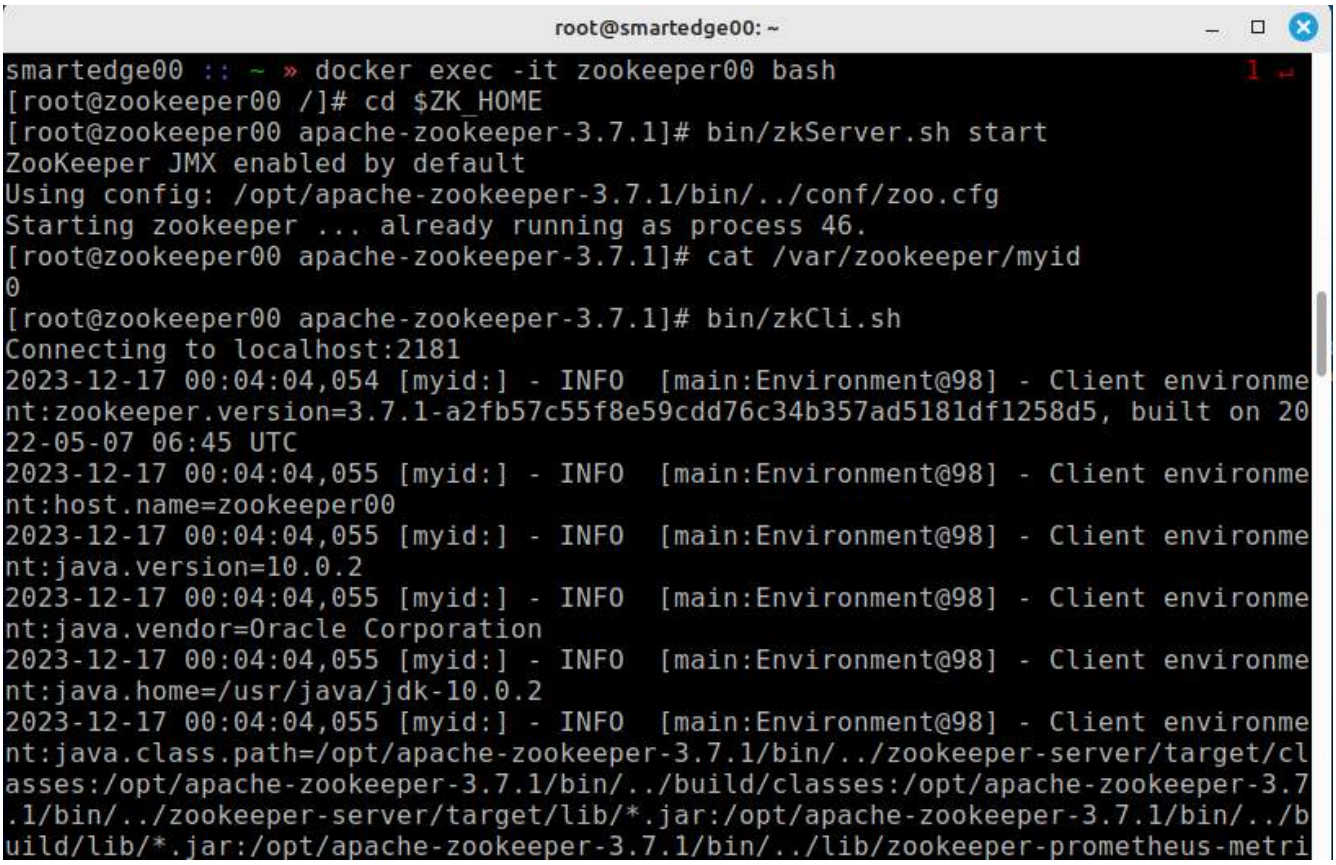

root@smartedge00: ~
"SandboxKey": "/var/run/docker/netns/1d264240d5ae",
"SecondaryIPAddresses": null,
"SecondaryIPv6Addresses": null,
"EndpointID": "",
"Gateway": "",
"GlobalIPv6Address": "",
"GlobalIPv6PrefixLen": 0,
"IPAddress": "",
"IPPrefixLen": 0,
"IPv6Gateway": "",
"MacAddress": "",
"Networks": {
"br0": {
"IPAMConfig": {
"IPv4Address": "192.168.0.5"
},
"Links": null,
"Aliases": [
"67de7555d7f4",
"storm00"
],
"NetworkID": "cd491412b46e0775b6d7f2ac8c7d96249600497163e3c3
d6cfee5fcf0b6e8b6a",
"EndpointID": "d2303511de6643358954390f9b93aeac5bc6372bb96e1
0976cd9f2c00a2495c9",
"Gateway": "192.168.0.1",
"IPAddress": "192.168.0.5",
"IPPrefixLen": 16,
"IPv6Gateway": "",
"GlobalIPv6Address": "",
"GlobalIPv6PrefixLen": 0,
"MacAddress": "02:42:c0:a8:00:05",
"DriverOpts": null
}
}
}
}
]
smartedge00 :: ~ »

Arrancamos Zookeeper
Accedemos a los contenedores zookeeper00, zookeeper01, zookeeper02

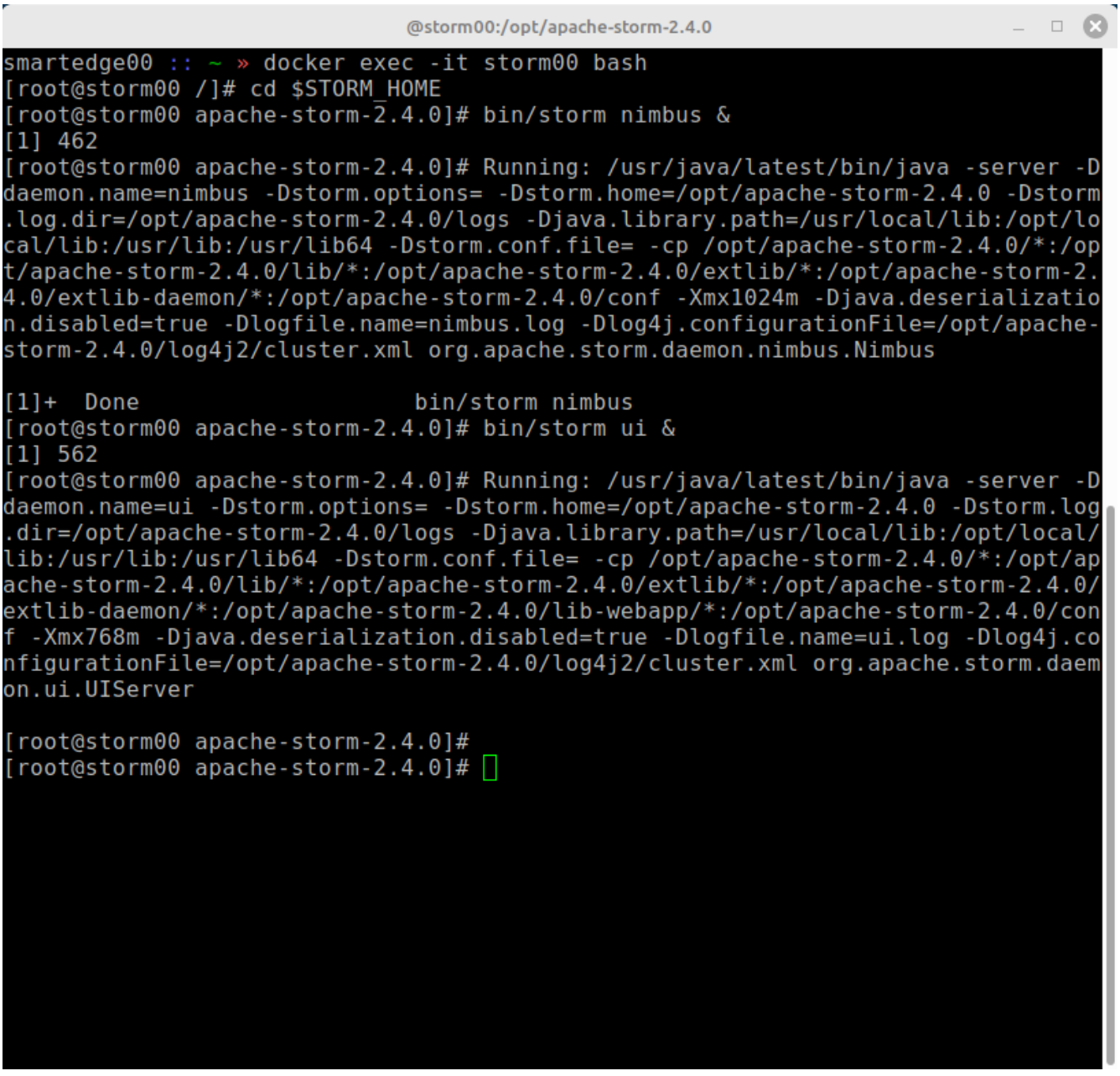


```
smartedge00 :: ~ » docker exec -it zookeeper00 bash
[root@zookeeper00 /]# cd $ZK_HOME
[root@zookeeper00 apache-zookeeper-3.7.1]# bin/zkServer.sh status
ZooKeeper JMX enabled by default
Using config: /opt/apache-zookeeper-3.7.1/bin/../conf/zoo.cfg
Client port found: 2181. Client address: localhost. Client SSL: false.
Mode: follower
[root@zookeeper00 apache-zookeeper-3.7.1]# exit
exit
smartedge00 :: ~ » docker exec -it zookeeper01 bash
[root@zookeeper01 /]# cd $ZK_HOME
[root@zookeeper01 apache-zookeeper-3.7.1]# bin/zkServer.sh status
ZooKeeper JMX enabled by default
Using config: /opt/apache-zookeeper-3.7.1/bin/../conf/zoo.cfg
Client port found: 2181. Client address: localhost. Client SSL: false.
Mode: leader
[root@zookeeper01 apache-zookeeper-3.7.1]#
```

```
@zookeeper00:/opt/apache-zookeeper-3.7.1
 connection, connectString=localhost:2181 sessionTimeout=30000 watcher=org.apach
e.zookeeper.ZooKeeperMain$MyWatcher@2ed0fbae
2023-12-17 00:05:58,211 [myid:] - INFO  [main:X509Util@77] - Setting -D jdk.tls.
rejectClientInitiatedRenegotiation=true to disable client-initiated TLS renegoti
ation
2023-12-17 00:05:58,214 [myid:] - INFO  [main:ClientCnxnSocket@239] - jute.maxbu
ffer value is 1048575 Bytes
2023-12-17 00:05:58,219 [myid:] - INFO  [main:ClientCnxn@1735] - zookeeper.reque
st.timeout value is 0. feature enabled=false
Welcome to ZooKeeper!
2023-12-17 00:05:58,228 [myid:localhost:2181] - INFO  [main-SendThread(localhost
:2181):ClientCnxn$SendThread@1171] - Opening socket connection to server localho
st/127.0.0.1:2181.
2023-12-17 00:05:58,228 [myid:localhost:2181] - INFO  [main-SendThread(localhost
:2181):ClientCnxn$SendThread@1173] - SASL config status: Will not attempt to aut
henticate using SASL (unknown error)
2023-12-17 00:05:58,232 [myid:localhost:2181] - INFO  [main-SendThread(localhost
:2181):ClientCnxn$SendThread@1005] - Socket connection established, initiating s
ession, client: /127.0.0.1:52290, server: localhost/127.0.0.1:2181
2023-12-17 00:05:58,240 [myid:localhost:2181] - INFO  [main-SendThread(localhost
:2181):ClientCnxn$SendThread@1446] - Session establishment complete on server lo
calhost/127.0.0.1:2181, session id = 0x90c8380001, negotiated timeout = 30000

WATCHER::

WatchedEvent state:SyncConnected type:None path:null
JLine support is enabled
[zk: localhost:2181(CONNECTED) 0] ls /storm
[assignments, backpressure, blobstore, blobstoremaxkeysequencenumber, credential
s, errors, leader-lock, logconfigs, nimbuses, storm, storms, supervisors, worker
beats]
[zk: localhost:2181(CONNECTED) 1] ls /storm/storms
[]
```


root@smartedge00: ~
smartedge00 :: ~ » docker exec -it zookeeper00 bash
[root@zookeeper00 /]# cd $ZK_HOME
[root@zookeeper00 apache-zookeeper-3.7.1]# bin/zkServer.sh start
ZooKeeper JMX enabled by default
Using config: /opt/apache-zookeeper-3.7.1/bin/../conf/zoo.cfg
Starting zookeeper ... already running as process 46.
[root@zookeeper00 apache-zookeeper-3.7.1]# cat /var/zookeeper/myid
0
[root@zookeeper00 apache-zookeeper-3.7.1]# bin/zkCli.sh
Connecting to localhost:2181
2023-12-17 00:04:04,054 [myid:] - INFO [main:Environment@98] - Client environment:zookeeper.version=3.7.1-a2fb57c55f8e59cdd76c34b357ad5181df1258d5, built on 2022-05-07 06:45 UTC
2023-12-17 00:04:04,055 [myid:] - INFO [main:Environment@98] - Client environment:host.name=zookeeper00
2023-12-17 00:04:04,055 [myid:] - INFO [main:Environment@98] - Client environment:java.version=10.0.2
2023-12-17 00:04:04,055 [myid:] - INFO [main:Environment@98] - Client environment:java.vendor=Oracle Corporation
2023-12-17 00:04:04,055 [myid:] - INFO [main:Environment@98] - Client environment:java.home=/usr/java/jdk-10.0.2
2023-12-17 00:04:04,055 [myid:] - INFO [main:Environment@98] - Client environment:java.class.path=/opt/apache-zookeeper-3.7.1/bin/../zookeeper-server/target/classes:/opt/apache-zookeeper-3.7.1/bin/../build/classes:/opt/apache-zookeeper-3.7.1/bin/../zookeeper-server/target/lib/*.jar:/opt/apache-zookeeper-3.7.1/bin/../build/lib/*.jar:/opt/apache-zookeeper-3.7.1/bin/../lib/zookeeper-prometheus-metri

LAB 2
Parte 1
Arrancamos Storm

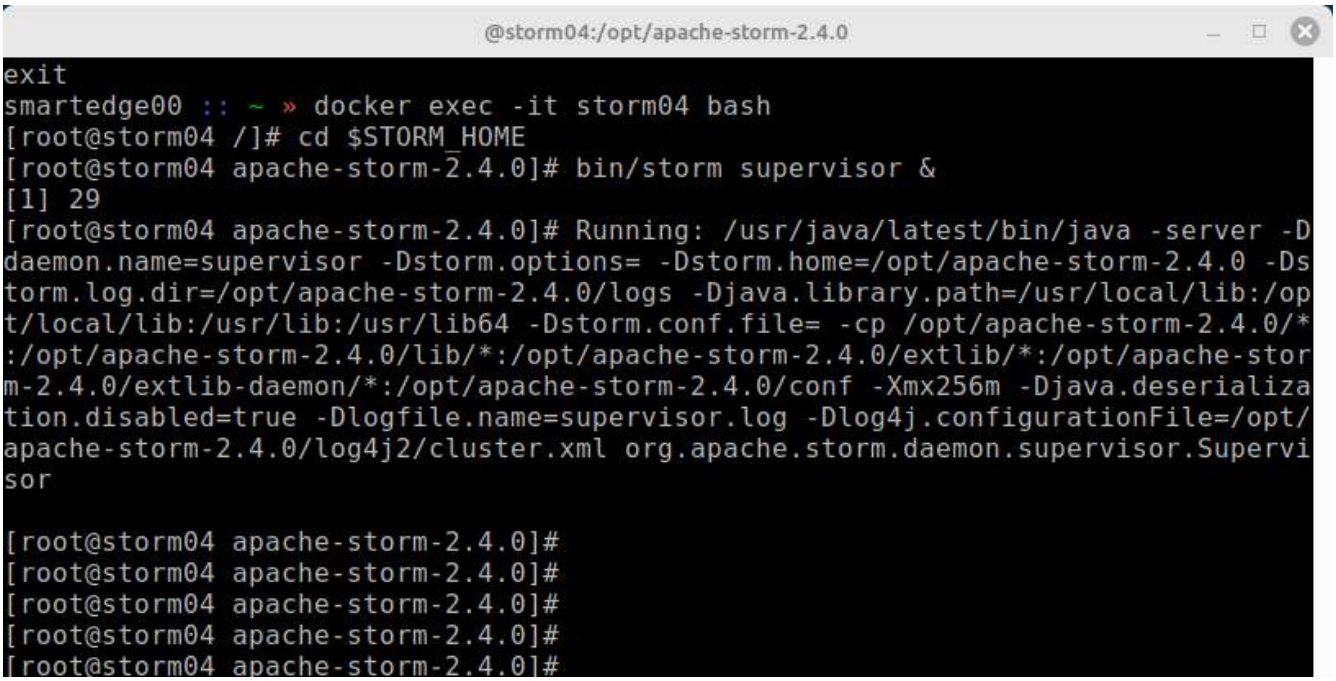


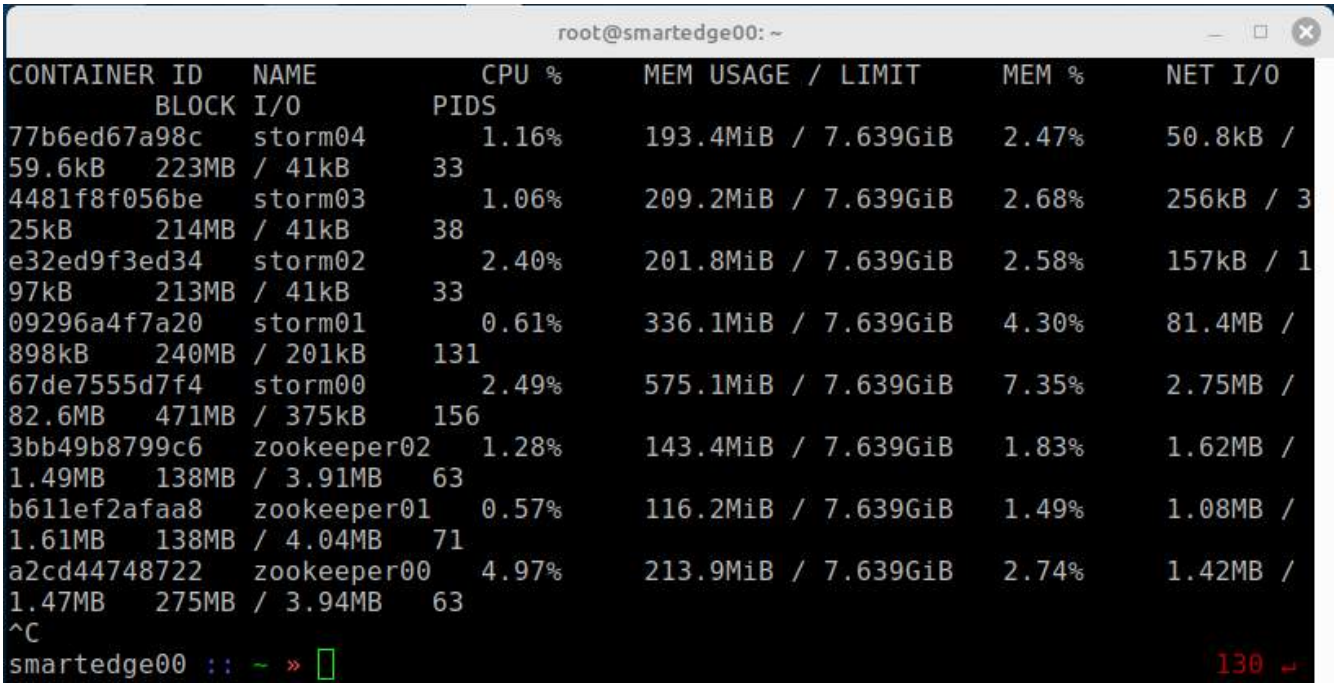
@storm01:/opt/apache-storm-2.4.0

```
smartedge00 :: ~ » docker exec -it storm01 bash
[root@storm01 /]# cd $STORM_HOME
[root@storm01 apache-storm-2.4.0]# bin/storm nimbus &
[1] 29
[root@storm01 apache-storm-2.4.0]# Running: /usr/java/latest/bin/java -server -D
daemon.name=nimbus -Dstorm.options= -Dstorm.home=/opt/apache-storm-2.4.0 -Dstorm
.log.dir=/opt/apache-storm-2.4.0/logs -Djava.library.path=/usr/local/lib:/opt/lo
cal/lib:/usr/lib:/usr/lib64 -Dstorm.conf.file= -cp /opt/apache-storm-2.4.0/*:/op
t/apache-storm-2.4.0/lib/*:/opt/apache-storm-2.4.0/extlib/*:/opt/apache-storm-2.
4.0/extlib-daemon/*:/opt/apache-storm-2.4.0/conf -Xmx1024m -Djava.deserializatio
n.disabled=true -Dlogfile.name=nimbus.log -Dlog4j.configurationFile=/opt/apache-
storm-2.4.0/log4j2/cluster.xml org.apache.storm.daemon.nimbus.Nimbus
```

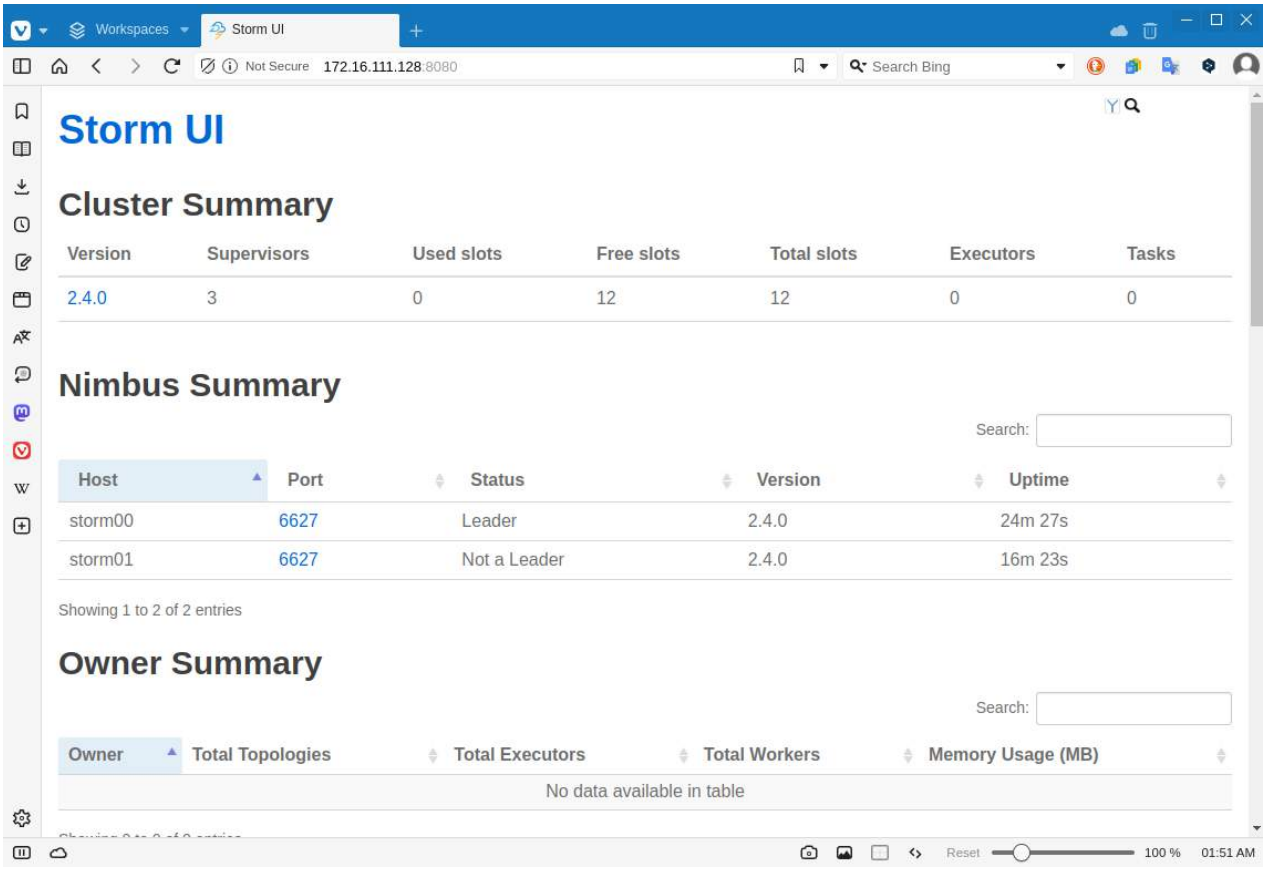
@storm03:/opt/apache-storm-2.4.0

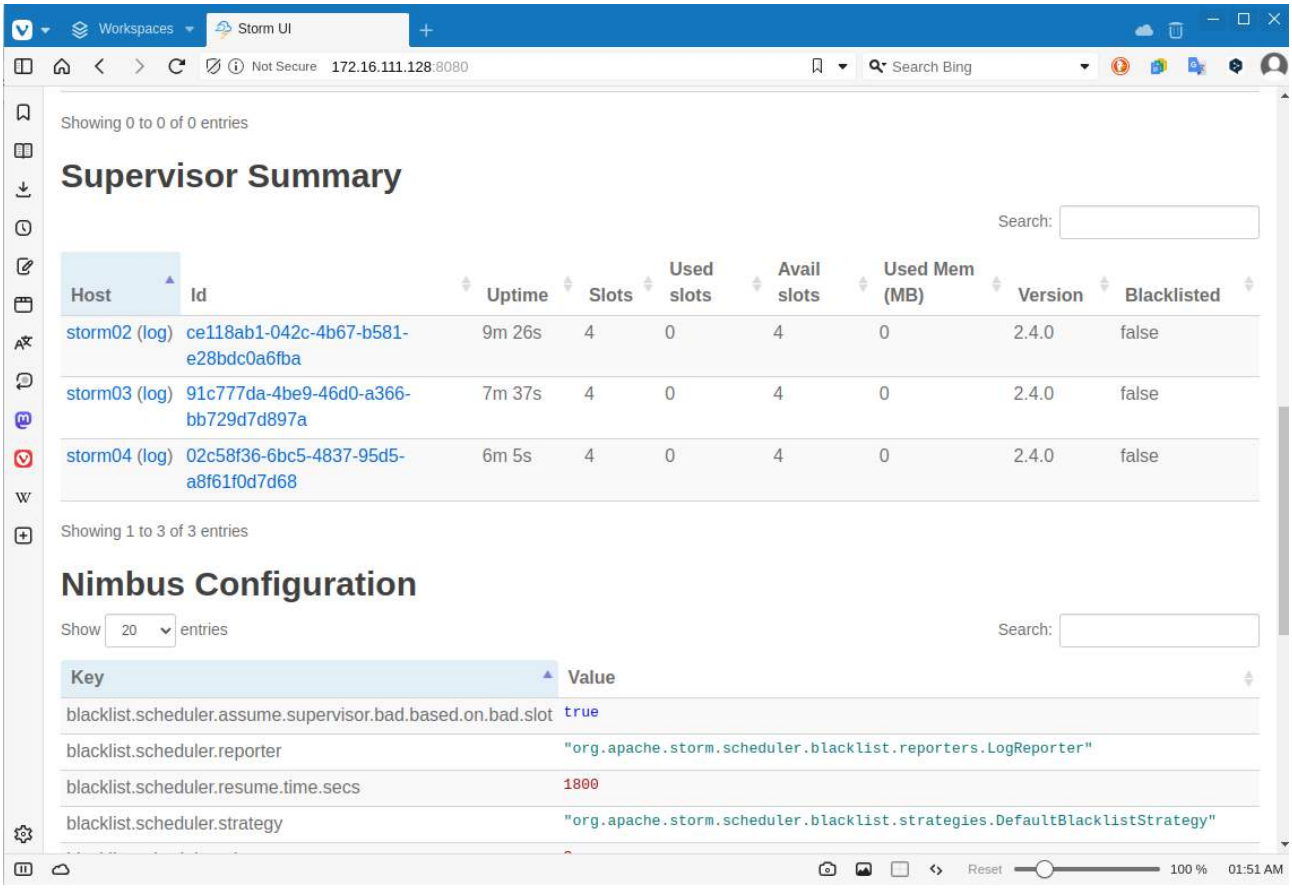
```
exit
smartedge00 :: ~ » docker exec -it storm02 bash
[root@storm02 /]# cd $STORM_HOME
[root@storm02 apache-storm-2.4.0]# bin/storm supervisor &
[1] 30
[root@storm02 apache-storm-2.4.0]#
[root@storm02 apache-storm-2.4.0]#
[root@storm02 apache-storm-2.4.0]# Running: /usr/java/latest/bin/java -server -D
daemon.name=supervisor -Dstorm.options= -Dstorm.home=/opt/apache-storm-2.4.0 -Ds
torm.log.dir=/opt/apache-storm-2.4.0/logs -Djava.library.path=/usr/local/lib:/op
t/local/lib:/usr/lib:/usr/lib64 -Dstorm.conf.file= -cp /opt/apache-storm-2.4.0/*
:/opt/apache-storm-2.4.0/lib/*:/opt/apache-storm-2.4.0/extlib/*:/opt/apache-stor
m-2.4.0/extlib-daemon/*:/opt/apache-storm-2.4.0/conf -Xmx256m -Djava.deserializa
tion.disabled=true -Dlogfile.name=supervisor.log -Dlog4j.configurationFile=/opt/
apache-storm-2.4.0/log4j2/cluster.xml org.apache.storm.daemon.supervisor.Supervi
sor

[root@storm02 apache-storm-2.4.0]# exit
exit
smartedge00 :: ~ » docker exec -it storm03 bash
[root@storm03 /]# cd $STORM_HOME
[root@storm03 apache-storm-2.4.0]# bin/storm supervisor &
[1] 29
[root@storm03 apache-storm-2.4.0]# Running: /usr/java/latest/bin/java -server -D
daemon.name=supervisor -Dstorm.options= -Dstorm.home=/opt/apache-storm-2.4.0 -Ds
torm.log.dir=/opt/apache-storm-2.4.0/logs -Djava.library.path=/usr/local/lib:/op
t/local/lib:/usr/lib:/usr/lib64 -Dstorm.conf.file= -cp /opt/apache-storm-2.4.0/*
:/opt/apache-storm-2.4.0/lib/*:/opt/apache-storm-2.4.0/extlib/*:/opt/apache-stor
m-2.4.0/extlib-daemon/*:/opt/apache-storm-2.4.0/conf -Xmx256m -Djava.deserializa
tion.disabled=true -Dlogfile.name=supervisor.log -Dlog4j.configurationFile=/opt/
apache-storm-2.4.0/log4j2/cluster.xml org.apache.storm.daemon.supervisor.Supervi
sor

[root@storm03 apache-storm-2.4.0]#
```

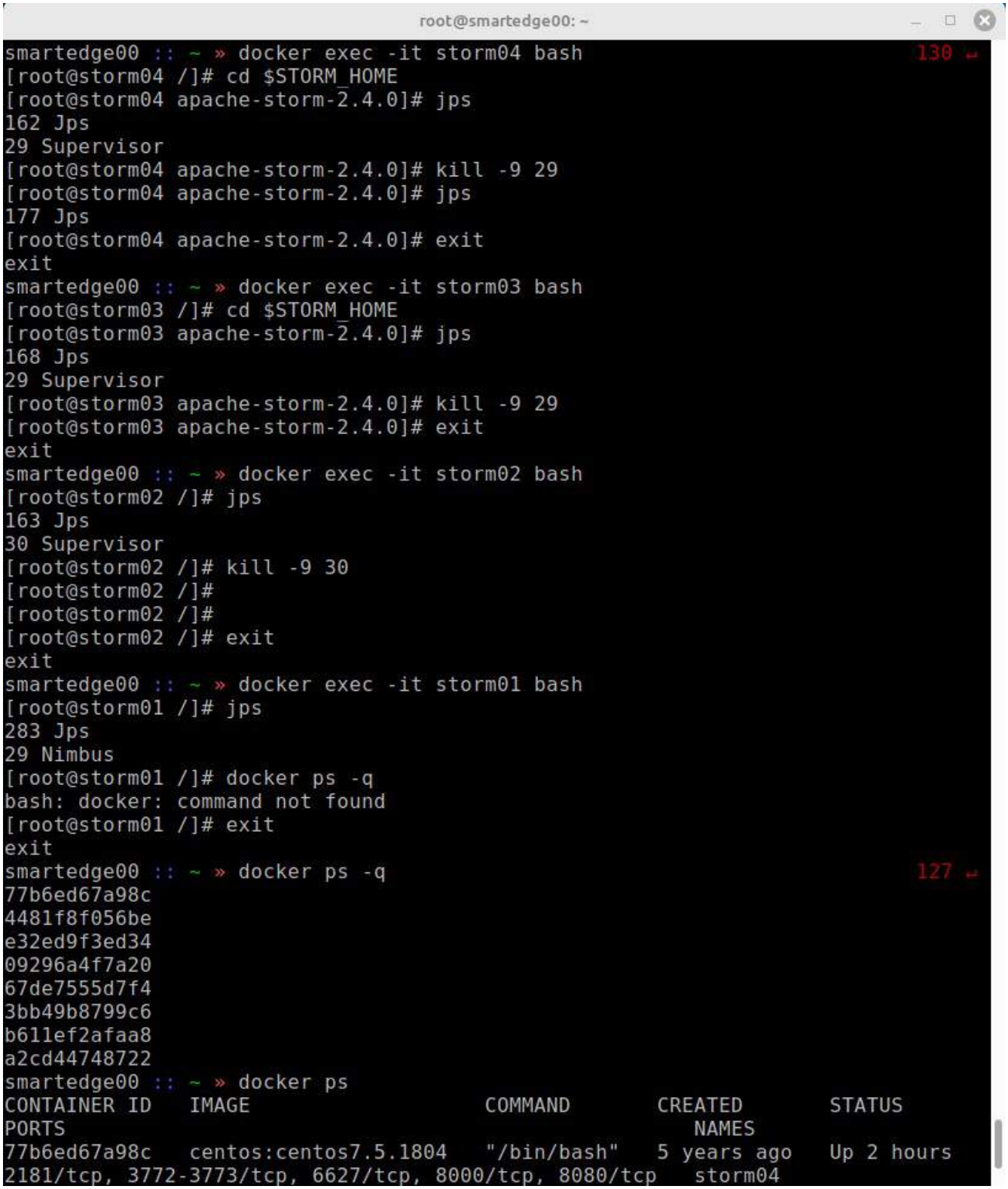

@storm04:/opt/apache-storm-2.4.0
exit
smartedge00 :: ~ » docker exec -it storm04 bash
[root@storm04 /]# cd $STORM_HOME
[root@storm04 apache-storm-2.4.0]# bin/storm supervisor &
[1] 29
[root@storm04 apache-storm-2.4.0]# Running: /usr/java/latest/bin/java -server -D
daemon.name=supervisor -Dstorm.options= -Dstorm.home=/opt/apache-storm-2.4.0 -Ds
torm.log.dir=/opt/apache-storm-2.4.0/logs -Djava.library.path=/usr/local/lib:/op
t/local/lib:/usr/lib:/usr/lib64 -Dstorm.conf.file= -cp /opt/apache-storm-2.4.0/*
:/opt/apache-storm-2.4.0/lib/*:/opt/apache-storm-2.4.0/extlib/*:/opt/apache-stor
m-2.4.0/extlib-daemon/*:/opt/apache-storm-2.4.0/conf -Xmx256m -Djava.deserializa
tion.disabled=true -Dlogfile.name=supervisor.log -Dlog4j.configurationFile=/opt/
apache-storm-2.4.0/log4j2/cluster.xml org.apache.storm.daemon.supervisor.Supervi
sor
[root@storm04 apache-storm-2.4.0]#
[root@storm04 apache-storm-2.4.0]#
[root@storm04 apache-storm-2.4.0]#
[root@storm04 apache-storm-2.4.0]#
[root@storm04 apache-storm-2.4.0]#

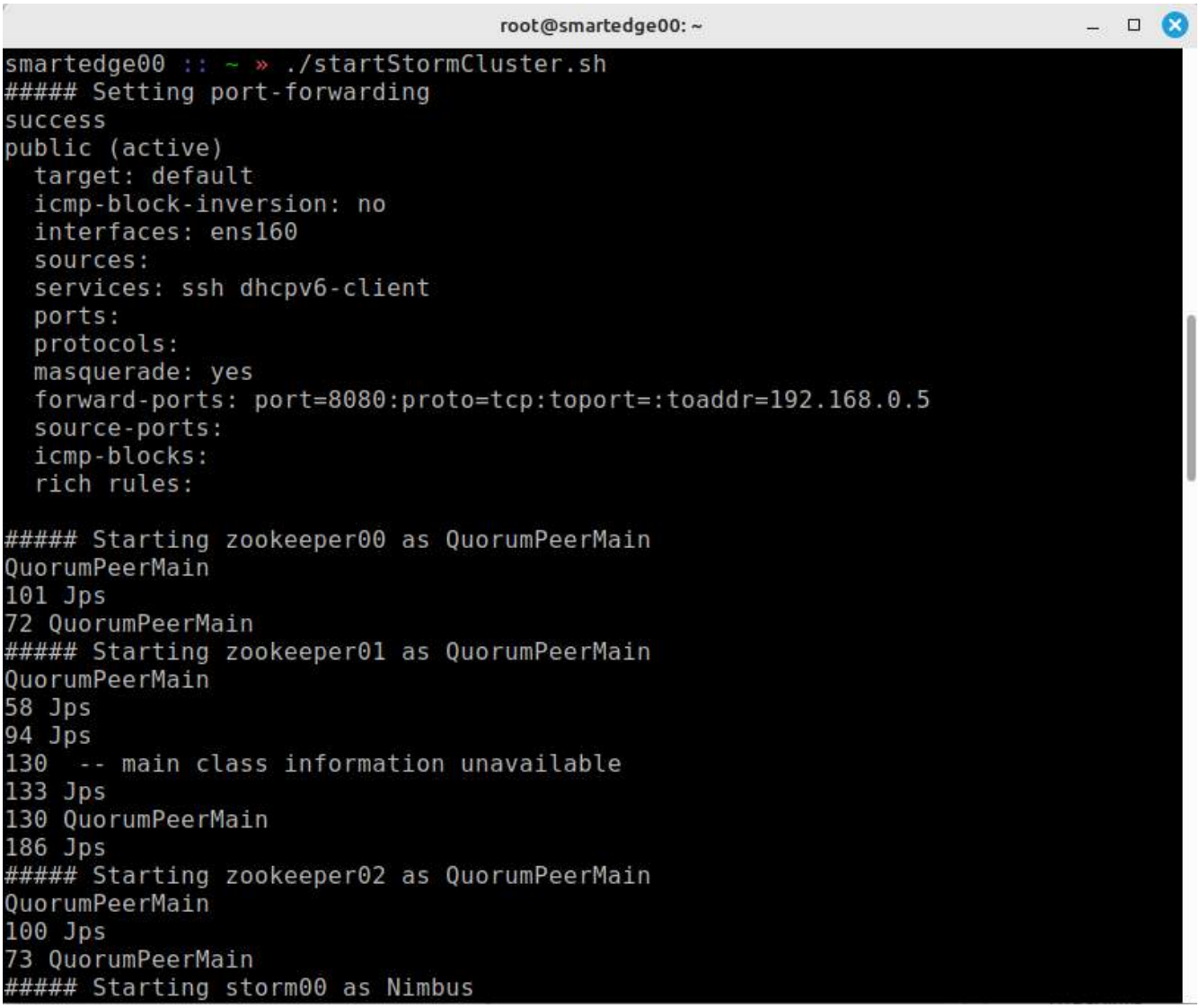

root@smartedge00: ~
CONTAINER ID NAME CPU % MEM USAGE / LIMIT MEM % NET I/O
BLOCK I/O PIDS
77b6ed67a98c storm04 1.16% 193.4MiB / 7.639GiB 2.47% 50.8kB /
59.6kB 223MB / 41kB 33
4481f8f056be storm03 1.06% 209.2MiB / 7.639GiB 2.68% 256kB / 3
25kB 214MB / 41kB 38
e32ed9f3ed34 storm02 2.40% 201.8MiB / 7.639GiB 2.58% 157kB / 1
97kB 213MB / 41kB 33
09296a4f7a20 storm01 0.61% 336.1MiB / 7.639GiB 4.30% 81.4MB /
898kB 240MB / 201kB 131
67de7555d7f4 storm00 2.49% 575.1MiB / 7.639GiB 7.35% 2.75MB /
82.6MB 471MB / 375kB 156
3bb49b8799c6 zookeeper02 1.28% 143.4MiB / 7.639GiB 1.83% 1.62MB /
1.49MB 138MB / 3.91MB 63
b611ef2afaa8 zookeeper01 0.57% 116.2MiB / 7.639GiB 1.49% 1.08MB /
1.61MB 138MB / 4.04MB 71
a2cd44748722 zookeeper00 4.97% 213.9MiB / 7.639GiB 2.74% 1.42MB /
1.47MB 275MB / 3.94MB 63
^C
smartedge00 :: ~ »
130

Podemos ver la consola gráfica desde el navegador en el sistema host
http://172.16.111.128:8080



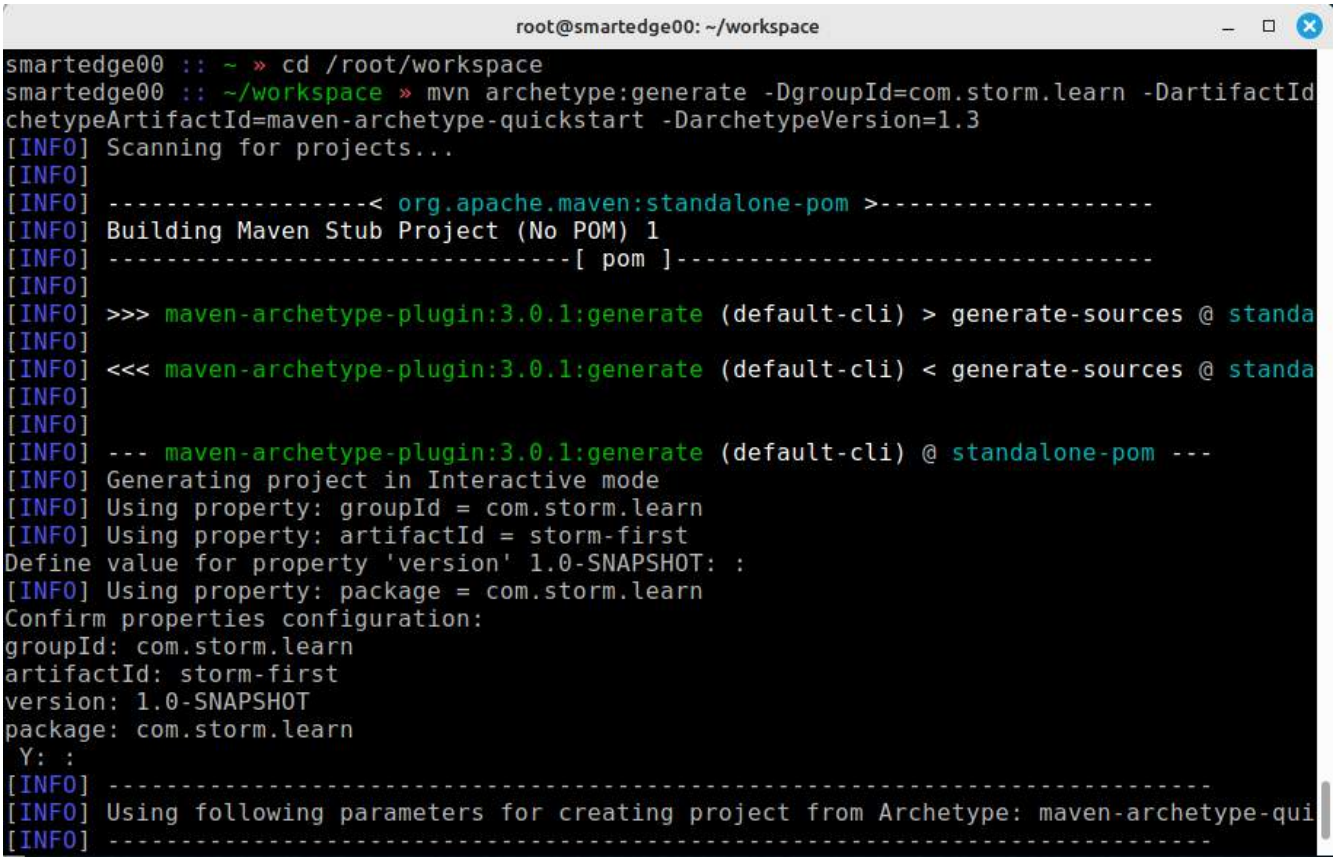

Workspaces
Storm UI
Not Secure
172.16.111.128:8080
Search Bing
Showing 0 to 0 of 0 entries
Supervisor Summary
Search:
Host
Id
Uptime
Slots
Used slots
Avail slots
Used Mem (MB)
Version
Blacklisted
storm02 (log)
ce118ab1-042c-4b67-b581-e28bdc0a6fba
9m 26s
4
0
4
0
2.4.0
false
storm03 (log)
91c777da-4be9-46d0-a366-bb729d7d897a
7m 37s
4
0
4
0
2.4.0
false
storm04 (log)
02c58f36-6bc5-4837-95d5-a8f61f0d7d68
6m 5s
4
0
4
0
2.4.0
false
Showing 1 to 3 of 3 entries
Nimbus Configuration
Show 20 entries
Search:
Key
Value
blacklist.scheduler.assume.supervisor.bad.based.on.bad.slot
true
blacklist.scheduler.reporter
"org.apache.storm.scheduler.blacklist.reporters.LogReporter"
blacklist.scheduler.resume.time.secs
1800
blacklist.scheduler.strategy
"org.apache.storm.scheduler.blacklist.strategies.DefaultBlacklistStrategy"
Reset
100 %
01:51 AM


root@smartedge00: ~
smartedge00 :: ~ » docker exec -it storm04 bash 130
[root@storm04 /]# cd $STORM_HOME
[root@storm04 apache-storm-2.4.0]# jps
162 Jps
29 Supervisor
[root@storm04 apache-storm-2.4.0]# kill -9 29
[root@storm04 apache-storm-2.4.0]# jps
177 Jps
[root@storm04 apache-storm-2.4.0]# exit
exit
smartedge00 :: ~ » docker exec -it storm03 bash
[root@storm03 /]# cd $STORM_HOME
[root@storm03 apache-storm-2.4.0]# jps
168 Jps
29 Supervisor
[root@storm03 apache-storm-2.4.0]# kill -9 29
[root@storm03 apache-storm-2.4.0]# exit
exit
smartedge00 :: ~ » docker exec -it storm02 bash
[root@storm02 /]# jps
163 Jps
30 Supervisor
[root@storm02 /]# kill -9 30
[root@storm02 /]#
[root@storm02 /]#
[root@storm02 /]# exit
exit
smartedge00 :: ~ » docker exec -it storm01 bash
[root@storm01 /]# jps
283 Jps
29 Nimbus
[root@storm01 /]# docker ps -q
bash: docker: command not found
[root@storm01 /]# exit
exit
smartedge00 :: ~ » docker ps -q 127
77b6ed67a98c
4481f8f056be
e32ed9f3ed34
09296a4f7a20
67de7555d7f4
3bb49b8799c6
b611ef2afaa8
a2cd44748722
smartedge00 :: ~ » docker ps
CONTAINER ID IMAGE COMMAND CREATED STATUS
PORTS NAMES
77b6ed67a98c centos:centos7.5.1804 "/bin/bash" 5 years ago Up 2 hours
2181/tcp, 3772-3773/tcp, 6627/tcp, 8000/tcp, 8080/tcp storm04

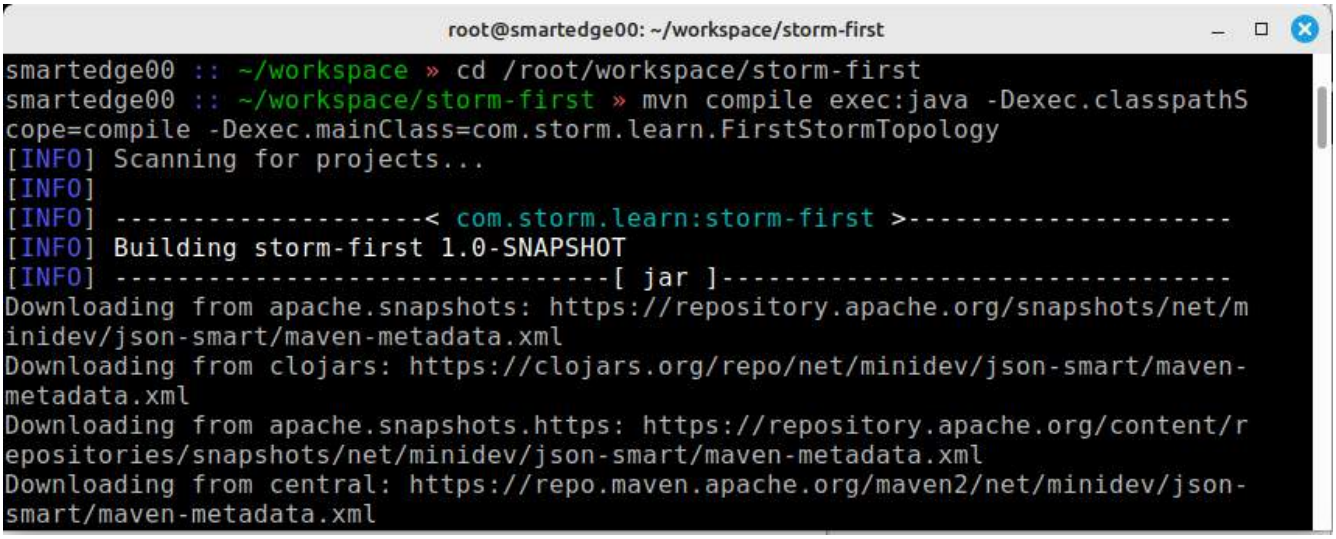

root@smartedge00: ~
smartedge00 :: ~ » ./startStormCluster.sh
Setting port-forwarding
success
public (active)
target: default
icmp-block-inversion: no
interfaces: ens160
sources:
services: ssh dhcpv6-client
ports:
protocols:
masquerade: yes
forward-ports: port=8080:proto=tcp:toport=:toaddr=192.168.0.5
source-ports:
icmp-blocks:
rich rules:
Starting zookeeper00 as QuorumPeerMain
QuorumPeerMain
101 Jps
72 QuorumPeerMain
Starting zookeeper01 as QuorumPeerMain
QuorumPeerMain
58 Jps
94 Jps
130 -- main class information unavailable
133 Jps
130 QuorumPeerMain
186 Jps
Starting zookeeper02 as QuorumPeerMain
QuorumPeerMain
100 Jps
73 QuorumPeerMain
Starting storm00 as Nimbus

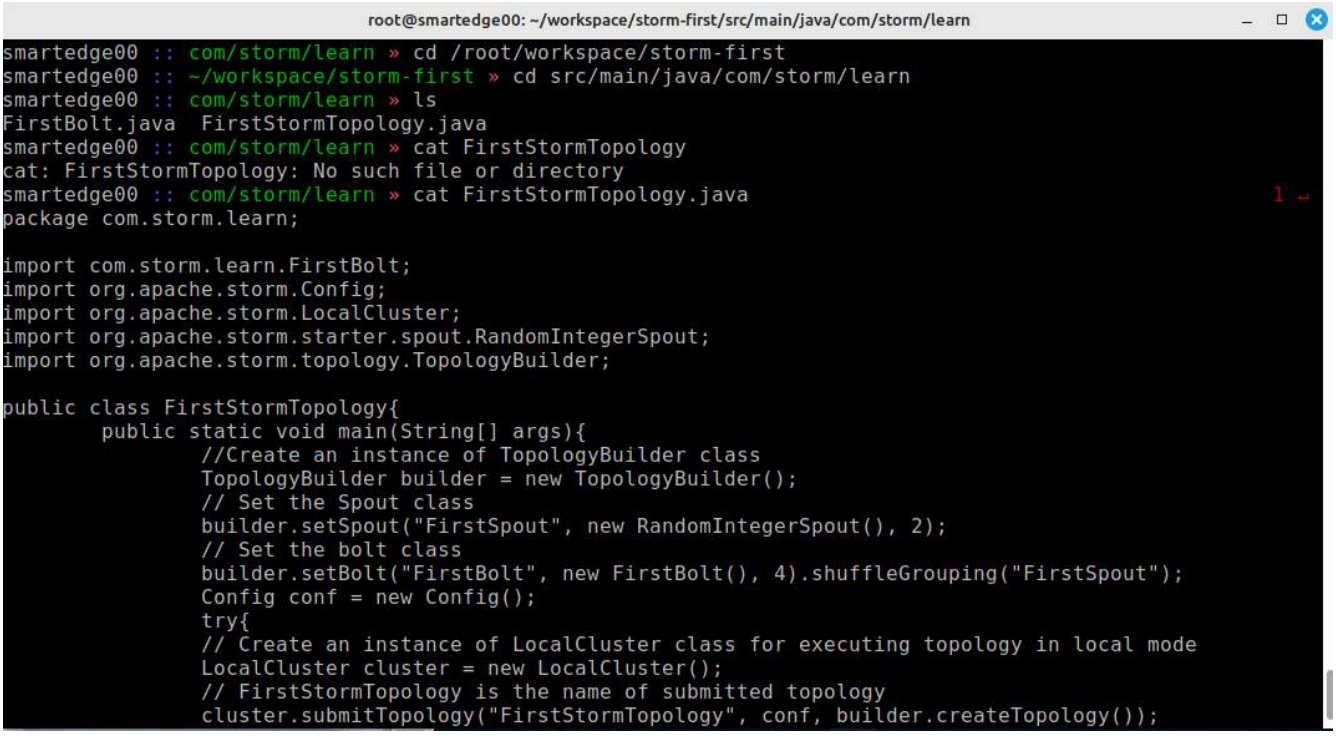

vim storm-script.bash
smartedge00 :: ~ » vim storm-script.bash
smartedge00 :: ~ » vim storm-script.bash
#!/bin/bash
for i in 0 1 2 3 4
do
echo "Comprobando storm0$i"
docker exec -it storm0$i jps
done

Creamos el proyecto storm-first con maven



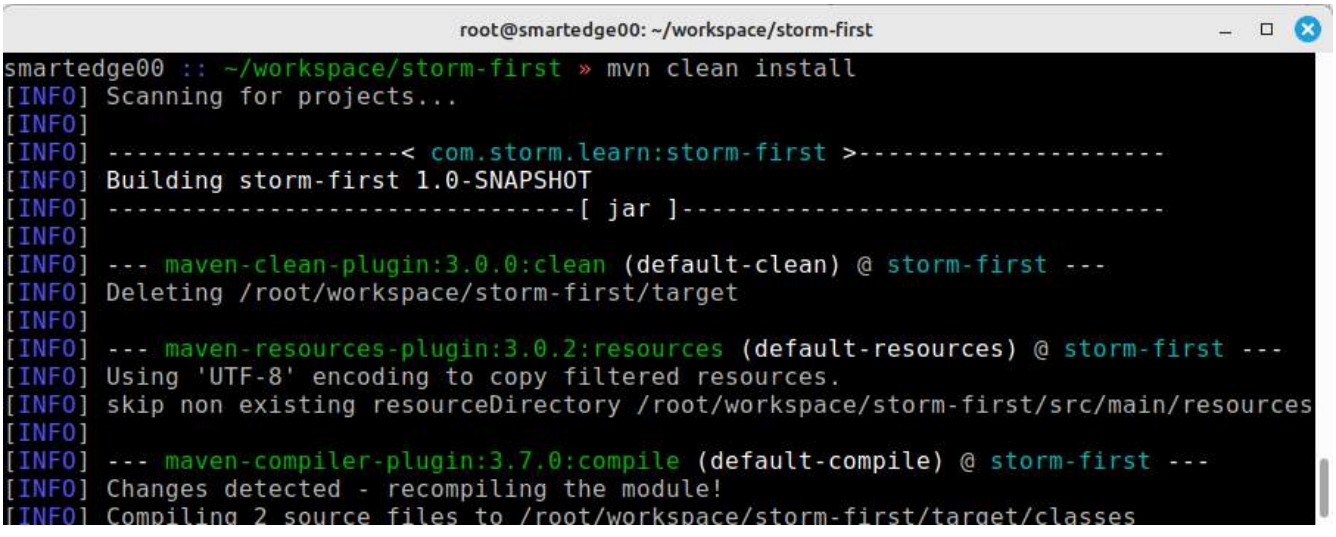
Removemos el fichero ejemplo “App.java”



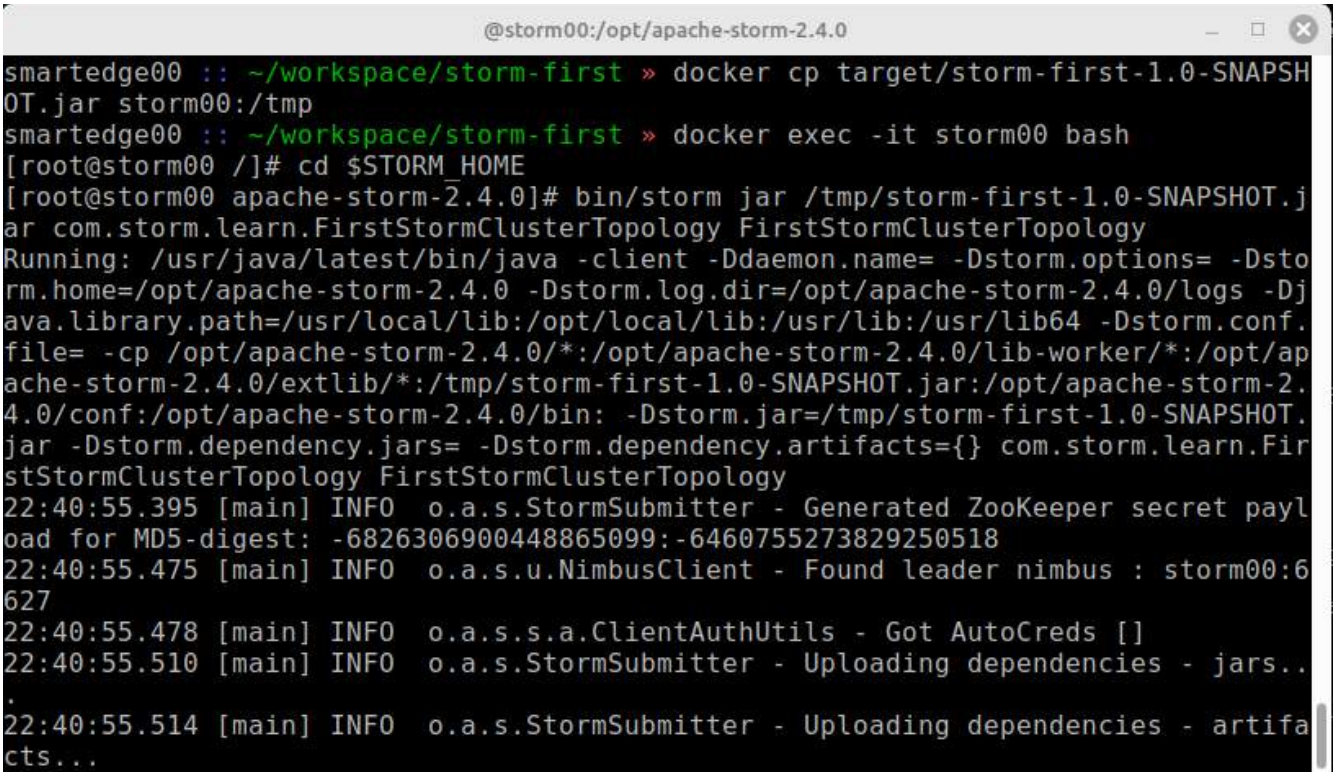

root@smartedge00: ~/workspace/storm-first/src/main/java/com/storm/learn
smartedge00 :: com/storm/learn » cd /root/workspace/storm-first
smartedge00 :: ~/workspace/storm-first » cd src/main/java/com/storm/learn
smartedge00 :: com/storm/learn » ls
FirstBolt.java FirstStormTopology.java
smartedge00 :: com/storm/learn » cat FirstStormTopology
cat: FirstStormTopology: No such file or directory
smartedge00 :: com/storm/learn » cat FirstStormTopology.java
1
package com.storm.learn;
import com.storm.learn.FirstBolt;
import org.apache.storm.Config;
import org.apache.storm.LocalCluster;
import org.apache.storm.starter.spout.RandomIntegerSpout;
import org.apache.storm.topology.TopologyBuilder;
public class FirstStormTopology{
public static void main(String[] args){
//Create an instance of TopologyBuilder class
TopologyBuilder builder = new TopologyBuilder();
// Set the Spout class
builder.setSpout("FirstSpout", new RandomIntegerSpout(), 2);
// Set the bolt class
builder.setBolt("FirstBolt", new FirstBolt(), 4).shuffleGrouping("FirstSpout");
Config conf = new Config();
try{
// Create an instance of LocalCluster class for executing topology in local mode
LocalCluster cluster = new LocalCluster();
// FirstStormTopology is the name of submitted topology
cluster.submitTopology("FirstStormTopology", conf, builder.createTopology());

Parte 2

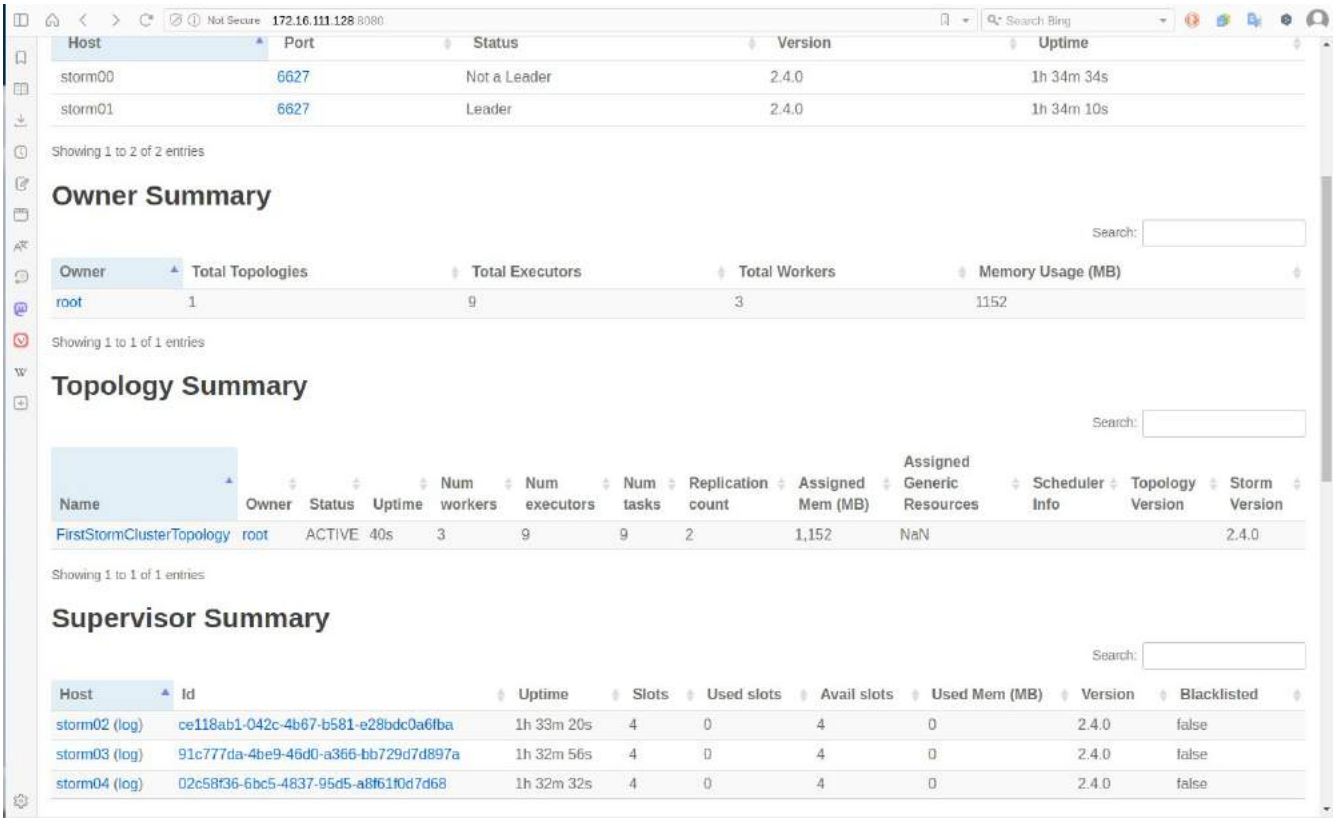

root@smartedge00: ~/workspace/storm-first/src/main/java/com/storm/learn
smartedge00 :: com/storm/learn » cp ~/workspace/storm_first/FirstStormClusterTopology.java ~/workspace/sto
rm-first/src/main/java/com/storm/learn
smartedge00 :: com/storm/learn » ls
FirstBolt.java FirstStormClusterTopology.java
smartedge00 :: com/storm/learn »

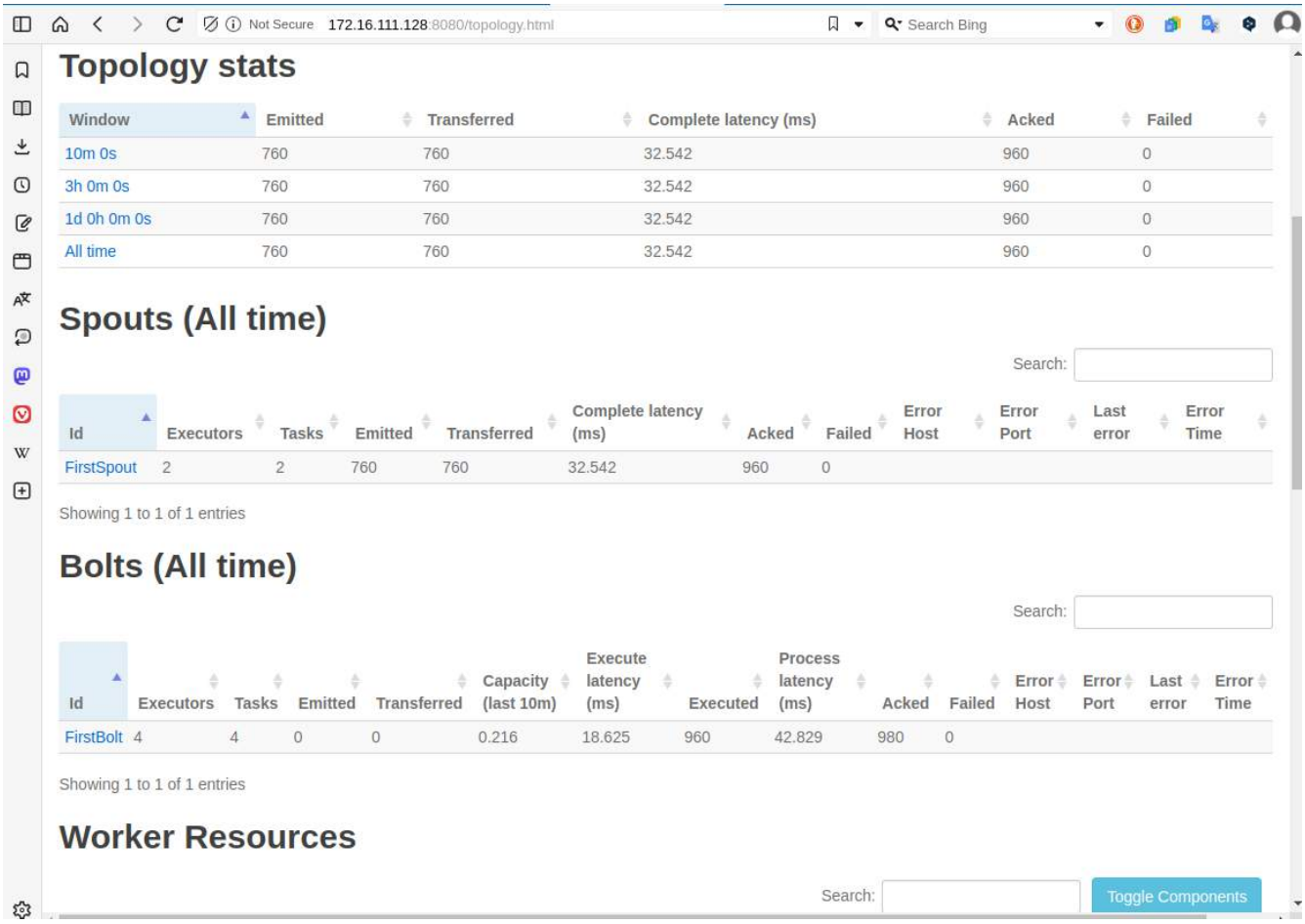

root@smartedge00: ~/workspace/storm-first
smartedge00 :: ~/workspace/storm-first » mvn clean install
[INFO] Scanning for projects...
[INFO]
[INFO] --------------------< com.storm.learn:storm-first >--------------------
[INFO] Building storm-first 1.0-SNAPSHOT
[INFO] --------------------------------[jar]---------------------------------
[INFO]
[INFO] --- maven-clean-plugin:3.0.0:clean (default-clean) @ storm-first ---
[INFO] Deleting /root/workspace/storm-first/target
[INFO]
[INFO] --- maven-resources-plugin:3.0.2:resources (default-resources) @ storm-first ---
[INFO] Using 'UTF-8' encoding to copy filtered resources.
[INFO] skip non existing resourceDirectory /root/workspace/storm-first/src/main/resources
[INFO]
[INFO] --- maven-compiler-plugin:3.7.0:compile (default-compile) @ storm-first ---
[INFO] Changes detected - recompiling the module!
[INFO] Compiling 2 source files to /root/workspace/storm-first/target/classes

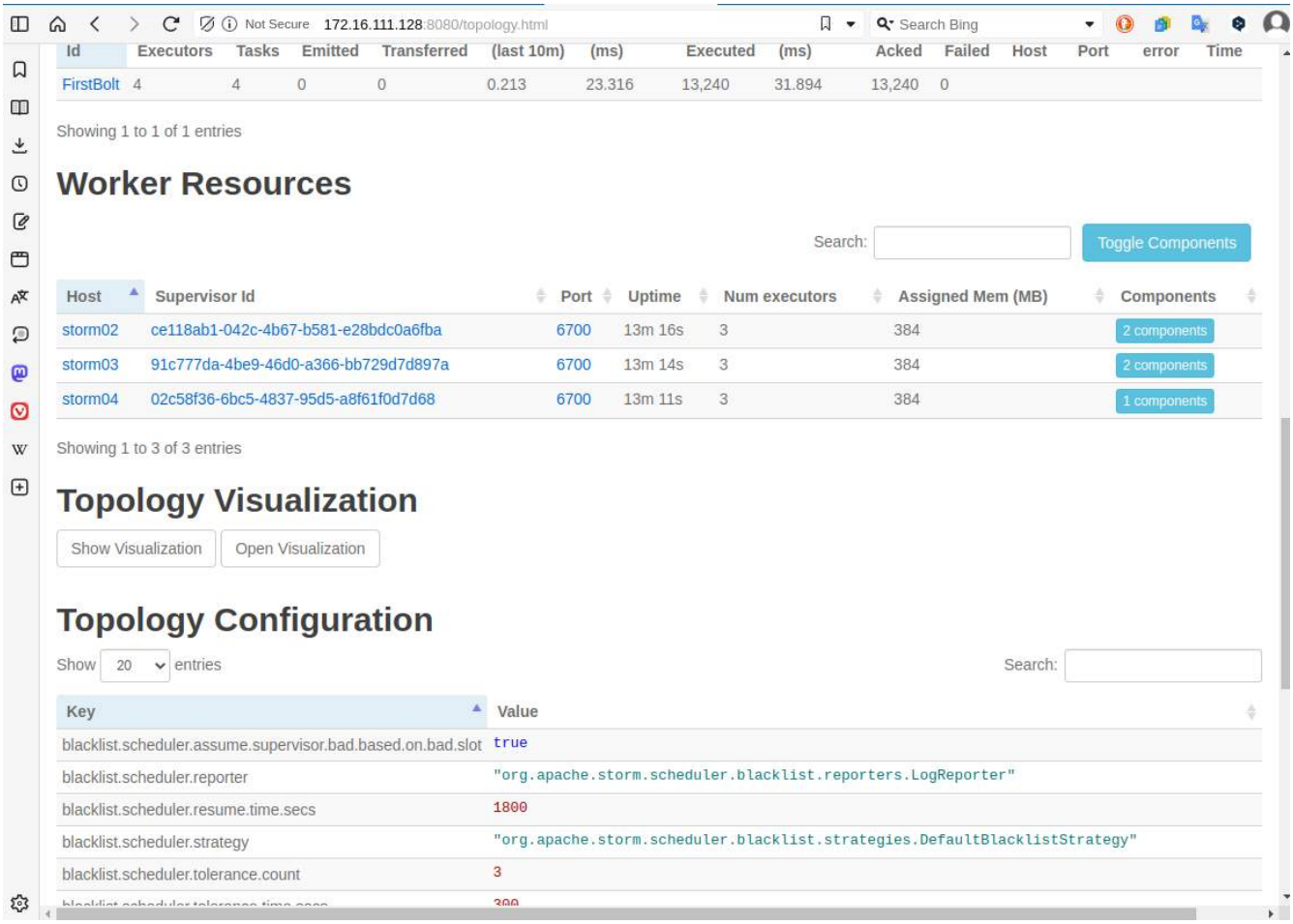
Copiamos el fichero de la topología 1.0-SNAPSHOT.jar en storm00, donde se encuentra el máster (Nimbus) y ejecutamos la nueva topología



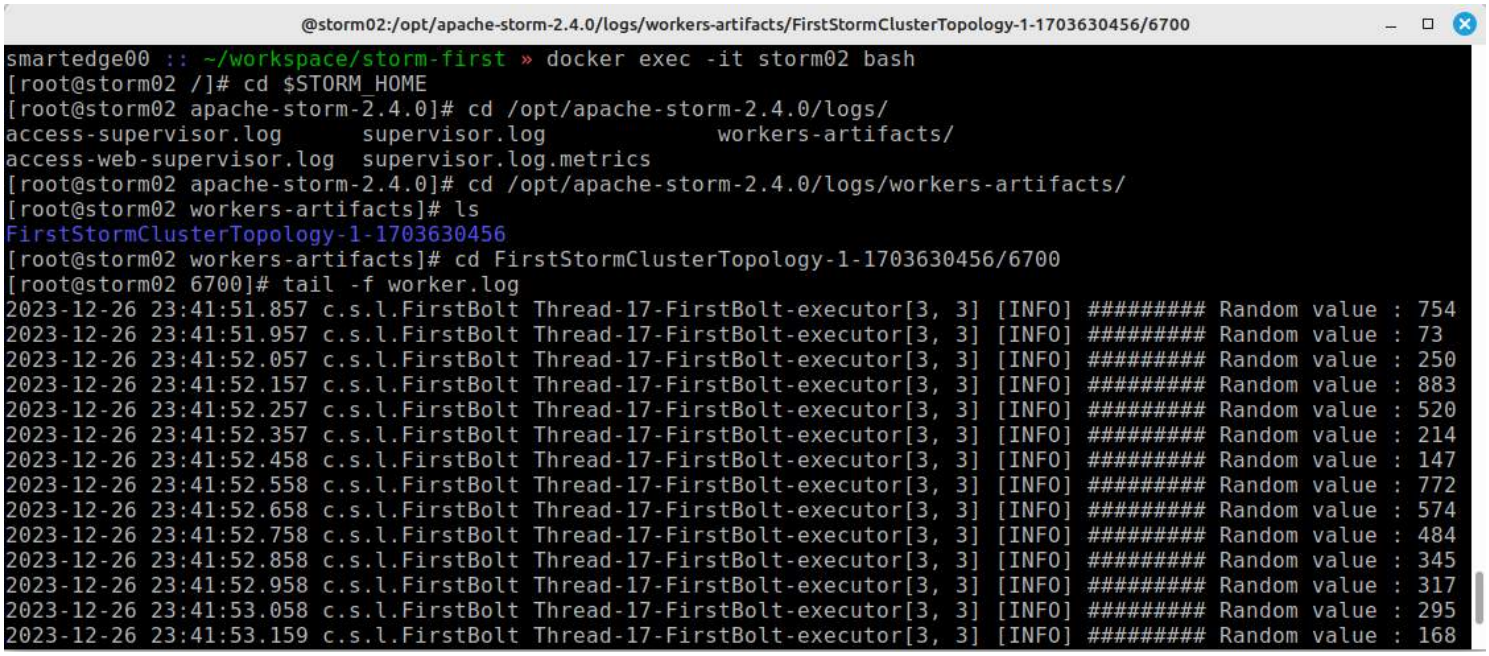
Una imagen que represente el clúster de Storm utilizado en los laboratorios: (3 puntos)
Comprobamos que se ha desplegado correctamente la topología accediendo a la UI de Storm.

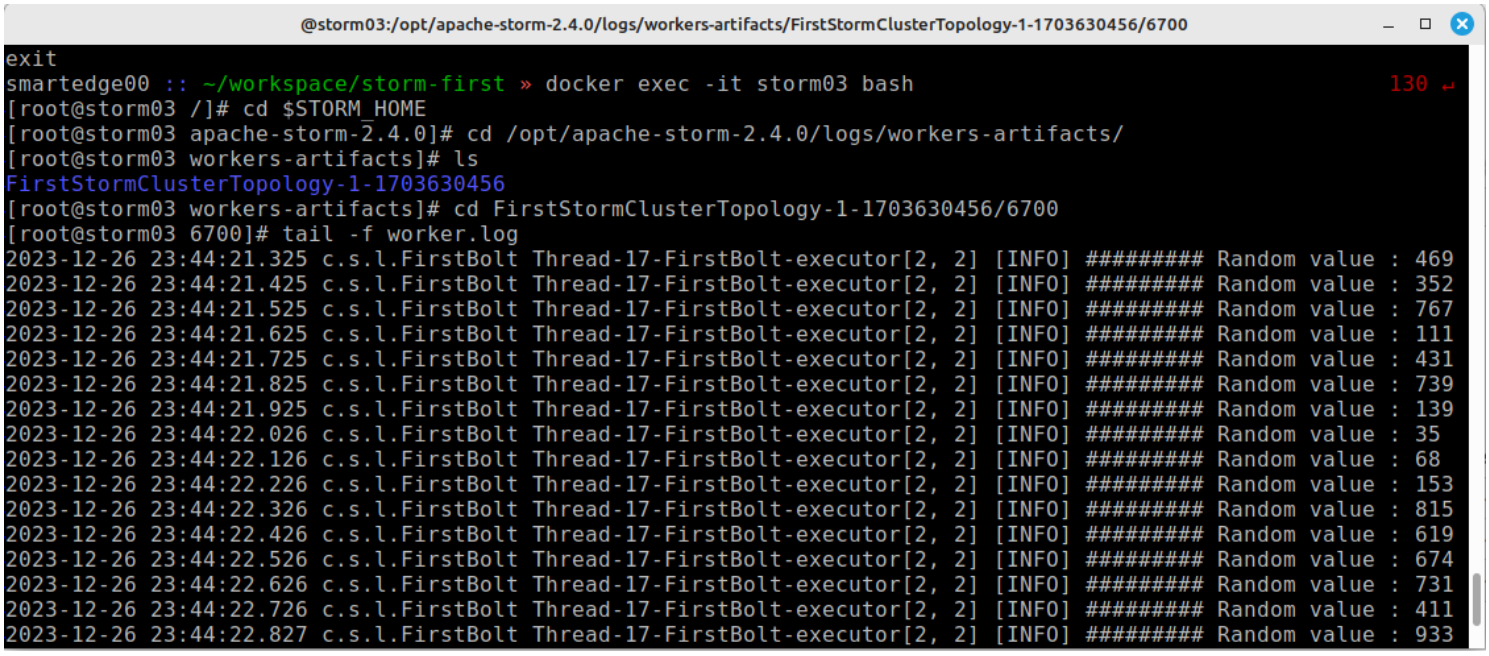


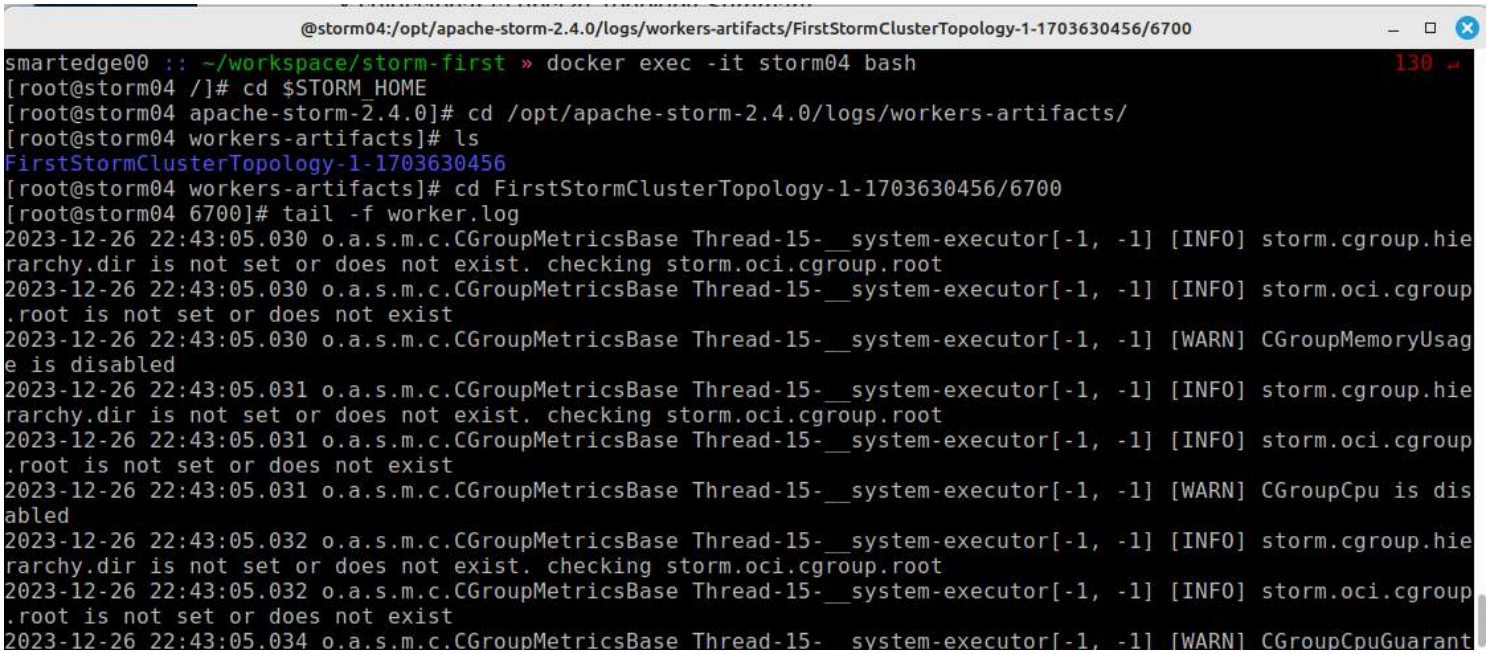

Not Secure 172.16.111.128:8080/topology.html
Search Bing
Topology stats
Window Emitted Transferred Complete latency (ms) Acked Failed
10m 0s 760 760 32.542 960 0
3h 0m 0s 760 760 32.542 960 0
1d 0h 0m 0s 760 760 32.542 960 0
All time 760 760 32.542 960 0
Spouts (All time)
Search:
Id Executors Tasks Emitted Transferred Complete latency (ms) Acked Failed Error Host Error Port Last error Error Time
FirstSpout 2 2 760 760 32.542 960 0
Showing 1 to 1 of 1 entries
Bolts (All time)
Search:
Id Executors Tasks Emitted Transferred Capacity (last 10m) Execute latency (ms) Executed Process latency (ms) Acked Failed Error Host Error Port Last error Error Time
FirstBolt 4 4 0 0 0.216 18.625 960 42.829 980 0
Showing 1 to 1 of 1 entries
Worker Resources
Search:
Toggle Components


Not Secure 172.16.111.128:8080/topology.html
Search Bing
Id Executors Tasks Emitted Transferred (last 10m) (ms) Executed (ms) Acked Failed Host Port error Time
FirstBolt 4 4 0 0 0.213 23.316 13,240 31.894 13,240 0
Showing 1 to 1 of 1 entries
Worker Resources
Search:
Toggle Components
Host Supervisor Id Port Uptime Num executors Assigned Mem (MB) Components
storm02 ce118ab1-042c-4b67-b581-e28bdc0a6fba 6700 13m 16s 3 384 2 components
storm03 91c777da-4be9-46d0-a366-bb729d7d897a 6700 13m 14s 3 384 2 components
storm04 02c58f36-6bc5-4837-95d5-a8f61f0d7d68 6700 13m 11s 3 384 1 components
Showing 1 to 3 of 3 entries
Topology Visualization
Show Visualization
Open Visualization
Topology Configuration
Show 20 entries
Search:
Key Value
blacklist.scheduler.assume.supervisor.bad.based.on.bad.slot true
blacklist.scheduler.reporter "org.apache.storm.scheduler.blacklist.reporters.LogReporter"
blacklist.scheduler.resume.time.secs 1800
blacklist.scheduler.strategy "org.apache.storm.scheduler.blacklist.strategies.DefaultBlacklistStrategy"
blacklist.scheduler.tolerance.count 3
300

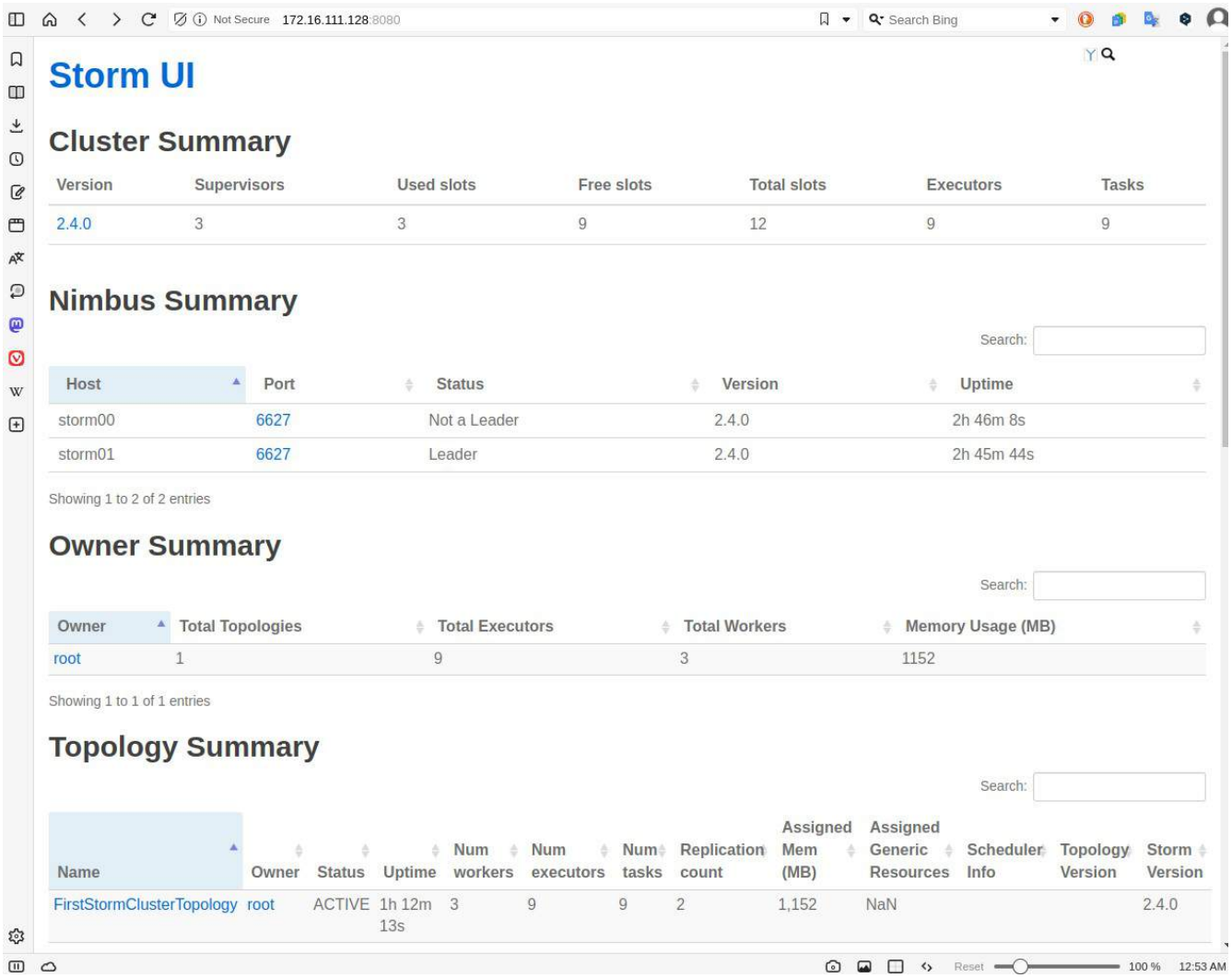
ENTREGA PARA LAB2

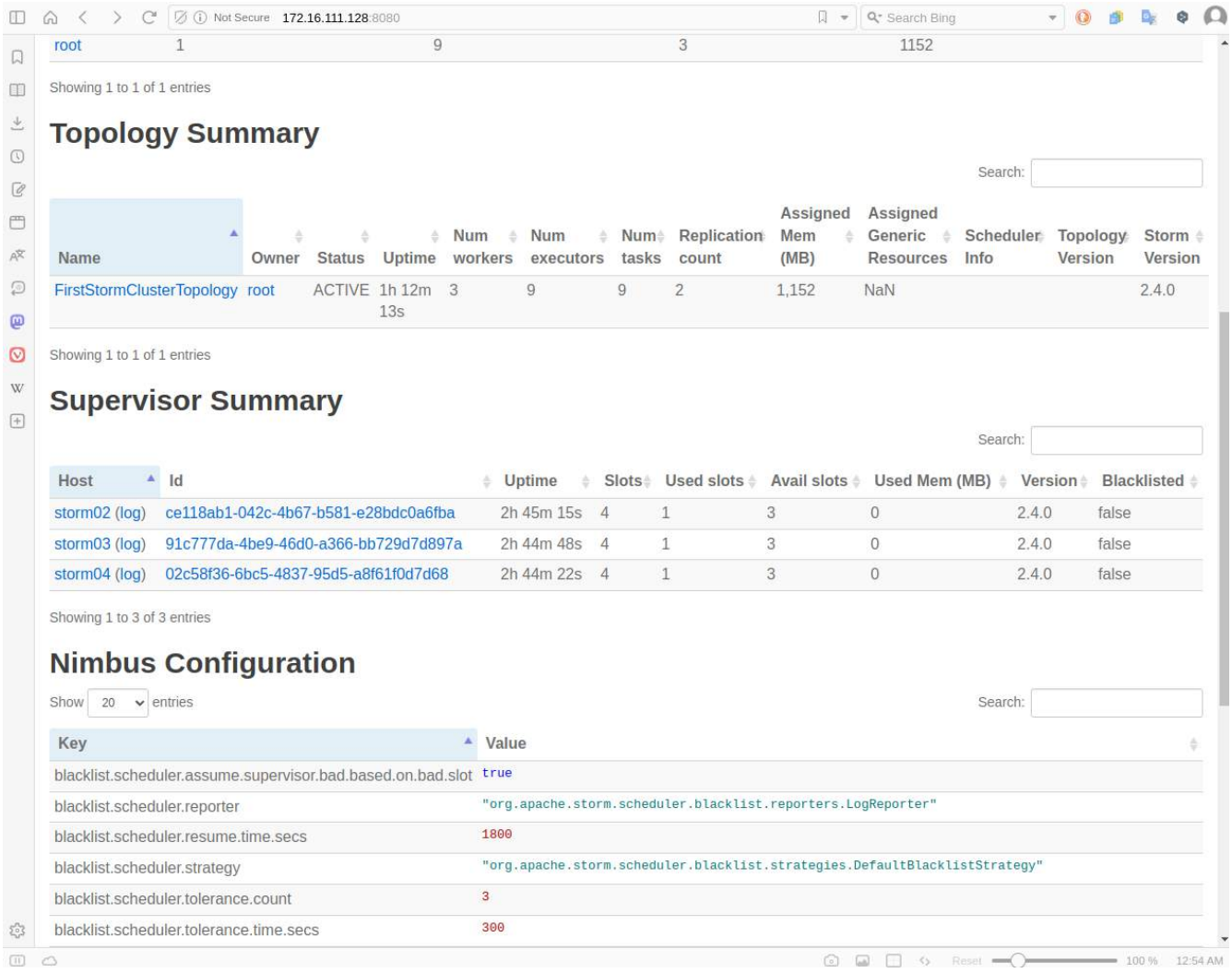


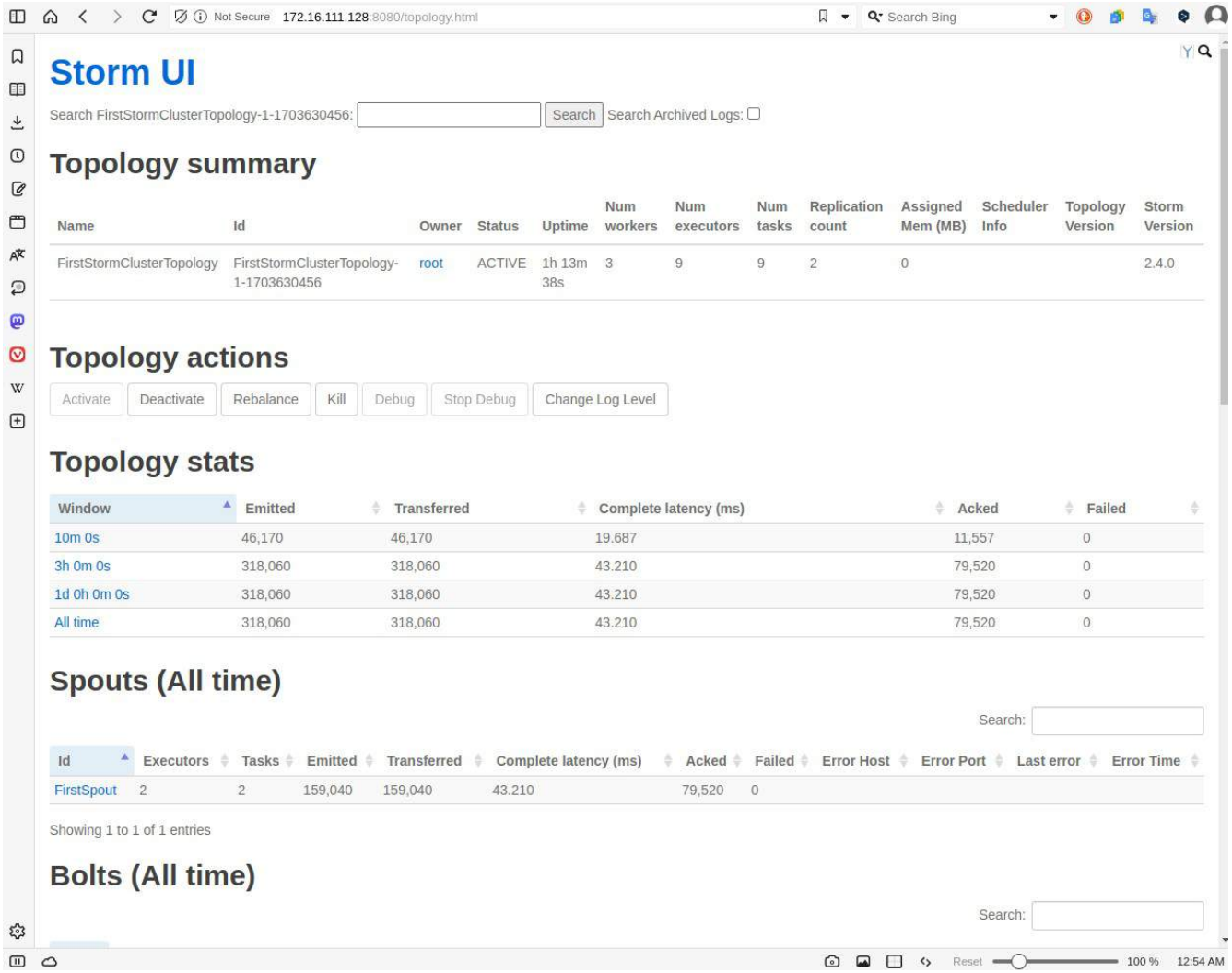


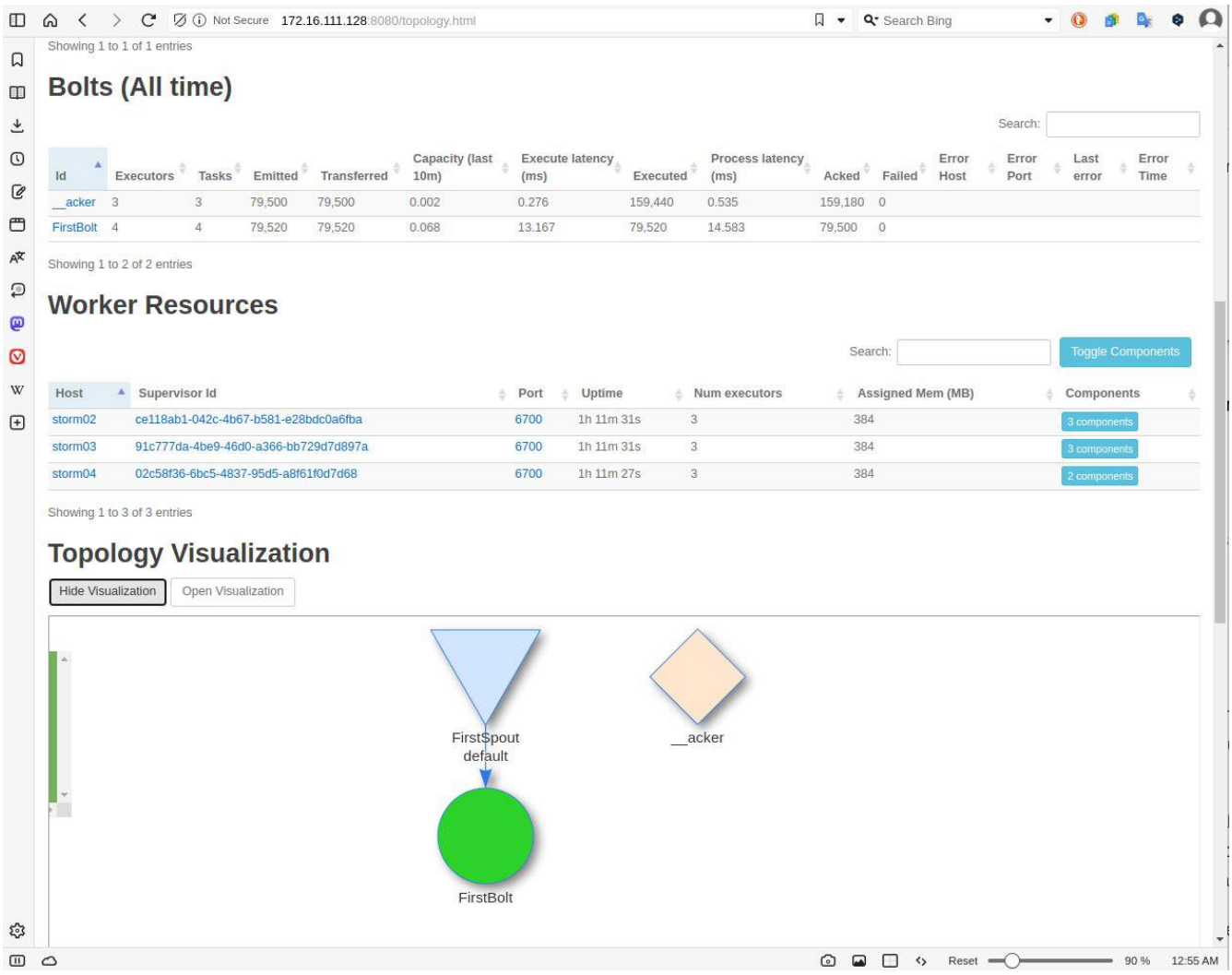

@storm04:/opt/apache-storm-2.4.0/logs/workers-artifacts/FirstStormClusterTopology-1-1703630456/6700
smartedge00 :: ~/workspace/storm-first » docker exec -it storm04 bash
130
[root@storm04 /]# cd $STORM_HOME
[root@storm04 apache-storm-2.4.0]# cd /opt/apache-storm-2.4.0/logs/workers-artifacts/
[root@storm04 workers-artifacts]# ls
FirstStormClusterTopology-1-1703630456
[root@storm04 workers-artifacts]# cd FirstStormClusterTopology-1-1703630456/6700
[root@storm04 6700]# tail -f worker.log
2023-12-26 22:43:05.030 o.a.s.m.c.CGroupMetricsBase Thread-15-__system-executor[-1, -1] [INFO] storm.cgroup.hierarchy.dir is not set or does not exist. checking storm.oci.cgroup.root
2023-12-26 22:43:05.030 o.a.s.m.c.CGroupMetricsBase Thread-15-__system-executor[-1, -1] [INFO] storm.oci.cgroup.root is not set or does not exist
2023-12-26 22:43:05.030 o.a.s.m.c.CGroupMetricsBase Thread-15-__system-executor[-1, -1] [WARN] CGroupMemoryUsage is disabled
2023-12-26 22:43:05.031 o.a.s.m.c.CGroupMetricsBase Thread-15-__system-executor[-1, -1] [INFO] storm.cgroup.hierarchy.dir is not set or does not exist. checking storm.oci.cgroup.root
2023-12-26 22:43:05.031 o.a.s.m.c.CGroupMetricsBase Thread-15-__system-executor[-1, -1] [INFO] storm.oci.cgroup.root is not set or does not exist
2023-12-26 22:43:05.031 o.a.s.m.c.CGroupMetricsBase Thread-15-__system-executor[-1, -1] [WARN] CGroupCpu is disabled
2023-12-26 22:43:05.032 o.a.s.m.c.CGroupMetricsBase Thread-15-__system-executor[-1, -1] [INFO] storm.cgroup.hierarchy.dir is not set or does not exist. checking storm.oci.cgroup.root
2023-12-26 22:43:05.032 o.a.s.m.c.CGroupMetricsBase Thread-15-__system-executor[-1, -1] [INFO] storm.oci.cgroup.root is not set or does not exist
2023-12-26 22:43:05.034 o.a.s.m.c.CGroupMetricsBase Thread-15-__system-executor[-1, -1] [WARN] CGroupCpuGuarant

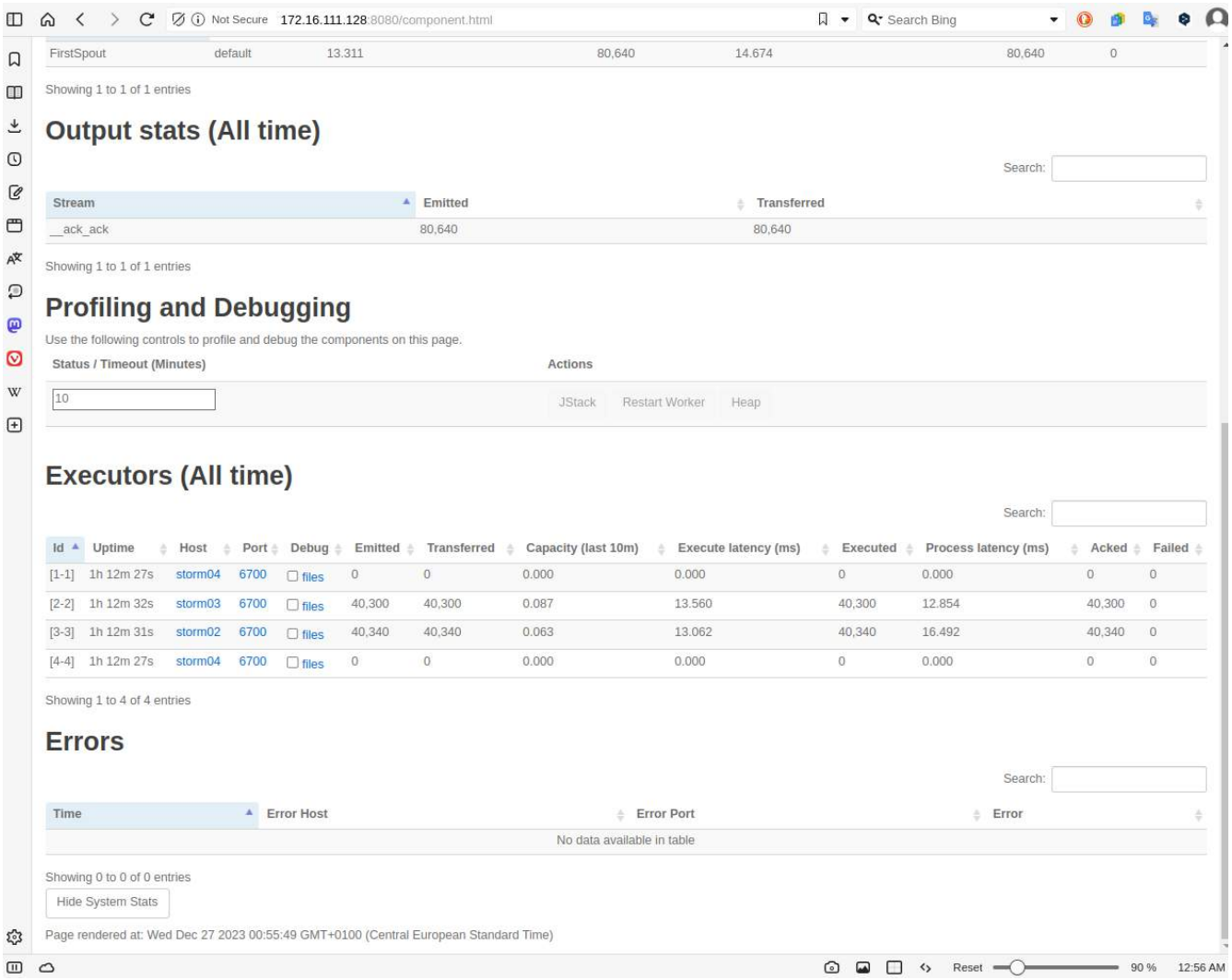
Capturas de pantalla de las topologías desplegadas en el laboratorio 2 (2 puntos):



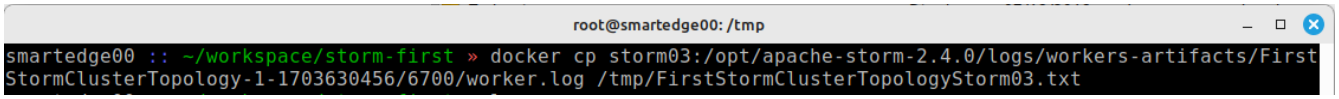

Not Secure 172.16.111.128:8080
Search Bing
root 1 9 3 1152
Showing 1 to 1 of 1 entries
Topology Summary
Search:
Name Owner Status Uptime Num workers Num executors Num tasks Replication count Assigned Mem (MB) Assigned Generic Resources Scheduler Info Topology Version Storm Version
FirstStormClusterTopology root ACTIVE 1h 12m 13s 3 9 9 2 1,152 NaN 2.4.0
Showing 1 to 1 of 1 entries
Supervisor Summary
Search:
Host Id Uptime Slots Used slots Avail slots Used Mem (MB) Version Blacklisted
storm02 (log) ce118ab1-042c-4b67-b581-e28bdc0a6fba 2h 45m 15s 4 1 3 0 2.4.0 false
storm03 (log) 91c777da-4be9-46d0-a366-bb729d7d897a 2h 44m 48s 4 1 3 0 2.4.0 false
storm04 (log) 02c58f36-6bc5-4837-95d5-a8f61f0d7d68 2h 44m 22s 4 1 3 0 2.4.0 false
Showing 1 to 3 of 3 entries
Nimbus Configuration
Show 20 entries
Search:
Key Value
blacklist.scheduler.assume.supervisor.bad.based.on.bad.slot true
blacklist.scheduler.reporter "org.apache.storm.scheduler.blacklist.reporters.LogReporter"
blacklist.scheduler.resume.time.secs 1800
blacklist.scheduler.strategy "org.apache.storm.scheduler.blacklist.strategies.DefaultBlacklistStrategy"
blacklist.scheduler.tolerance.count 3
blacklist.scheduler.tolerance.time.secs 300
Reset 100 % 12:54 AM


Not Secure 172.16.111.128:8080/topology.html
Search Bing
Storm UI
Search FirstStormClusterTopology-1-1703630456: Search Search Archived Logs:
Topology summary
Name Id Owner Status Uptime Num workers Num executors Num tasks Replication count Assigned Mem (MB) Scheduler Info Topology Version Storm Version
FirstStormClusterTopology FirstStormClusterTopology-1-1703630456 root ACTIVE 1h 13m 38s 3 9 9 2 0 2.4.0
Topology actions
Activate Deactivate Rebalance Kill Debug Stop Debug Change Log Level
Topology stats
Window Emitted Transferred Complete latency (ms) Acked Failed
10m 0s 46,170 46,170 19.687 11,557 0
3h 0m 0s 318,060 318,060 43.210 79,520 0
1d 0h 0m 0s 318,060 318,060 43.210 79,520 0
All time 318,060 318,060 43.210 79,520 0
Spouts (All time)
Search:
Id Executors Tasks Emitted Transferred Complete latency (ms) Acked Failed Error Host Error Port Last error Error Time
FirstSpout 2 2 159,040 159,040 43.210 79,520 0
Showing 1 to 1 of 1 entries
Bolts (All time)
Search:
Reset 100 % 12:54 AM

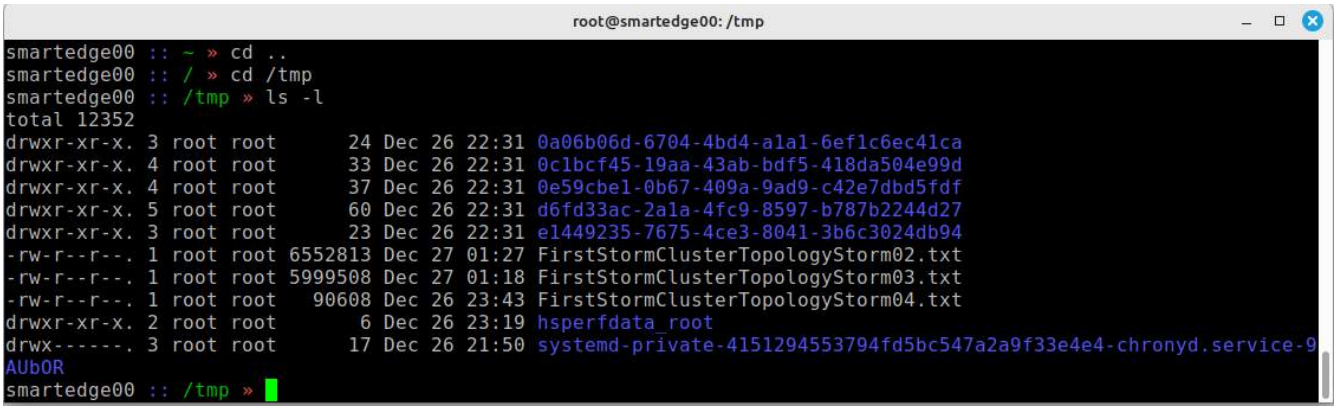

Not Secure 172.16.111.128:8080/topology.html
Search Bing
Showing 1 to 1 of 1 entries
Bolts (All time)
Search:
Id Executors Tasks Emitted Transferred Capacity (last 10m) Execute latency (ms) Executed Process latency (ms) Acked Failed Error Host Error Port Last error Error Time
__acker 3 3 79,500 79,500 0.002 0.276 159,440 0.535 159,180 0
FirstBolt 4 4 79,520 79,520 0.068 13.167 79,520 14.583 79,500 0
Showing 1 to 2 of 2 entries
Worker Resources
Search:
Toggle Components
Host Supervisor Id Port Uptime Num executors Assigned Mem (MB) Components
storm02 ce118ab1-042c-4b67-b581-e28bdc0a6fba 6700 1h 11m 31s 3 384 3 components
storm03 91c777da-4be9-46d0-a366-bb729d7d897a 6700 1h 11m 31s 3 384 3 components
storm04 02c58f36-6bc5-4837-95d5-a8f61f0d7d68 6700 1h 11m 27s 3 384 2 components
Showing 1 to 3 of 3 entries
Topology Visualization
Hide Visualization
Open Visualization
First$pout
default
__acker
FirstBolt
Reset
90 %
12:55 AM



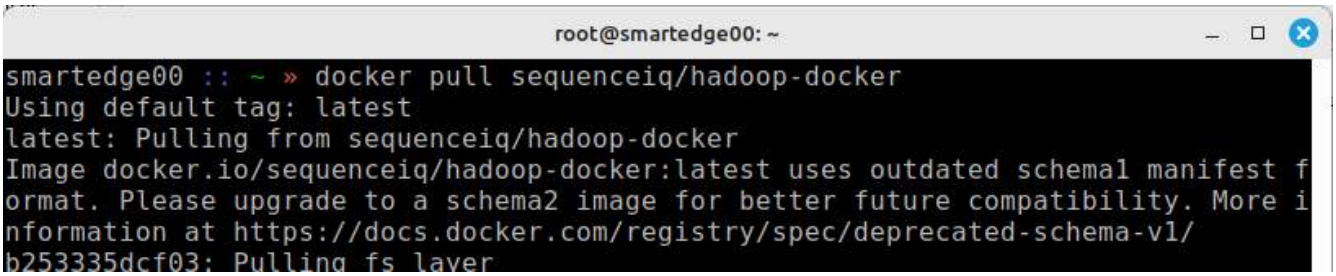
SalidaLab2.[zip/tgz] (0.5 puntos)
Debe contener el fichero worker.log de la topología desplegada en el laboratorio 2.

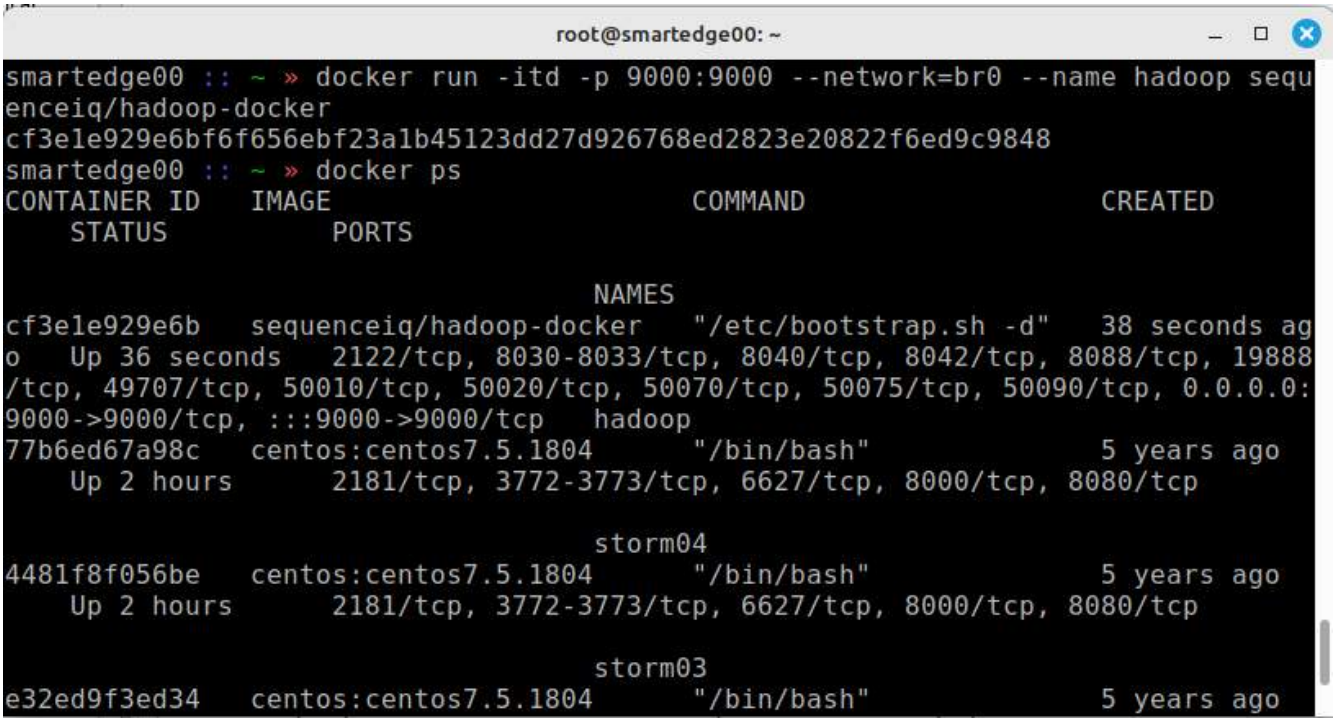


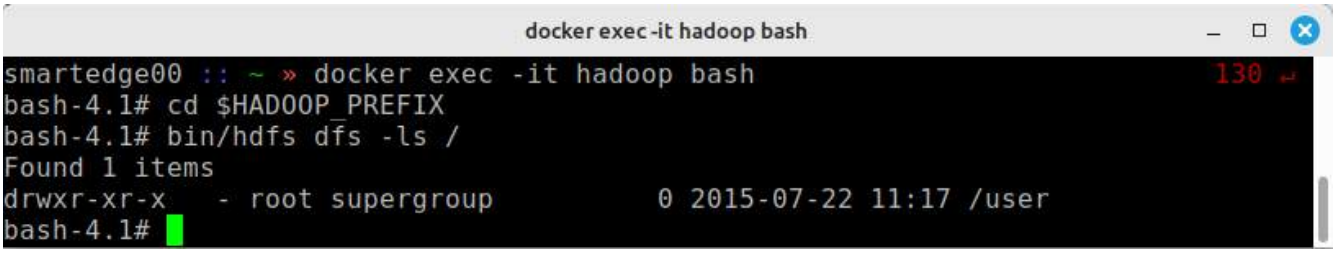
Adjunto a esta práctica el ficheros worker.log correspondiente al que he llamado “FirstStormClusterTopologyStorm02.txt” “FirstStormClusterTopologyStorm03.txt” y “FirstStormClusterTopologyStorm04.txt”

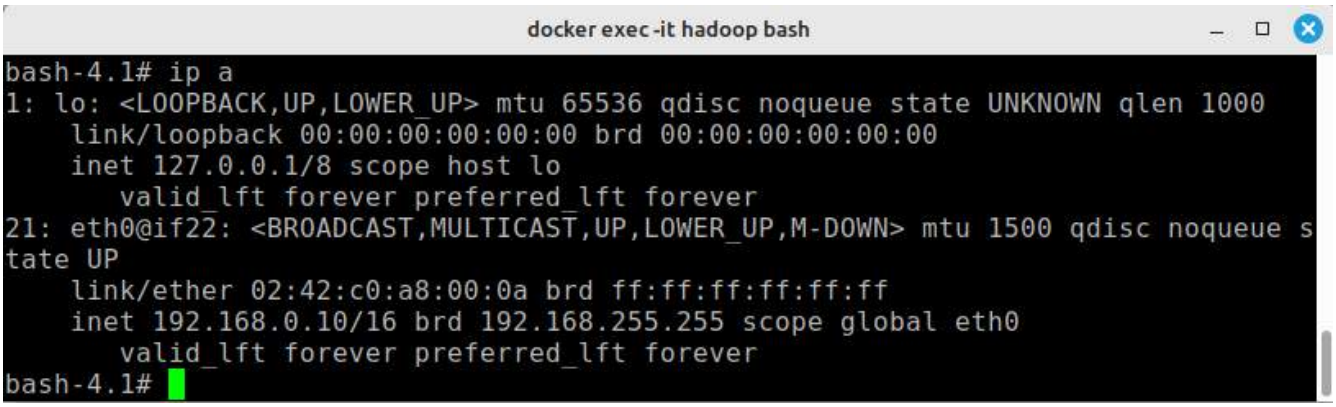


LAB 3

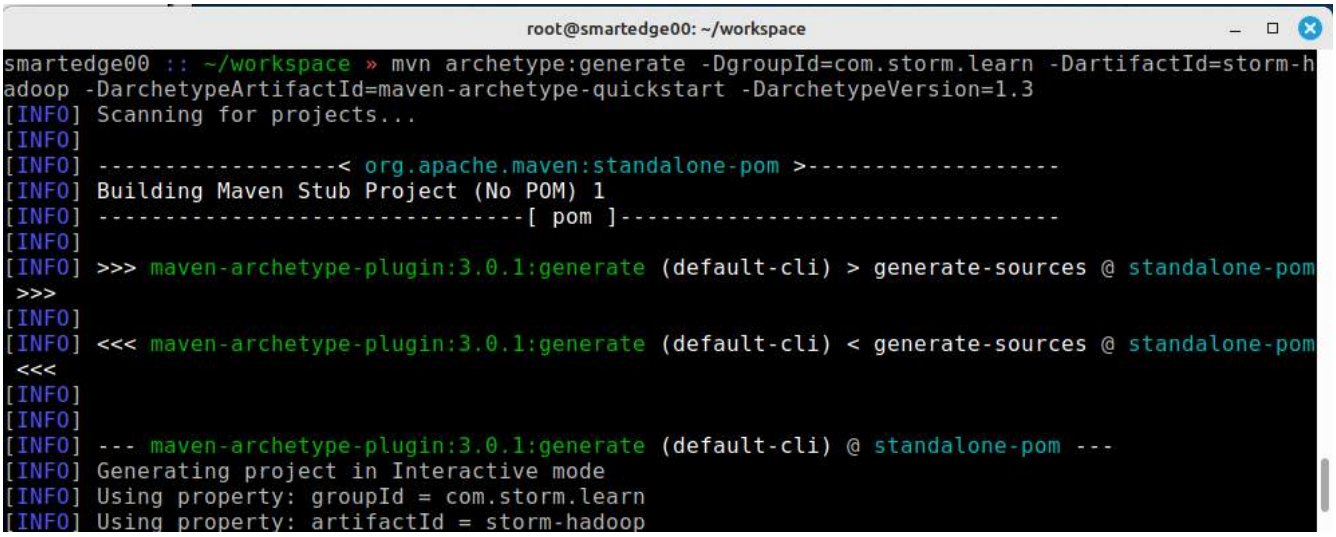

root@smartedge00: ~
smartedge00 :: ~ » docker pull sequenceiq/hadoop-docker
Using default tag: latest
latest: Pulling from sequenceiq/hadoop-docker
Image docker.io/sequenceiq/hadoop-docker:latest uses outdated schema1 manifest f
ormat. Please upgrade to a schema2 image for better future compatibility. More i
nformation at https://docs.docker.com/registry/spec/deprecated-schema-v1/
b253335dcf03: Pulling fs layer

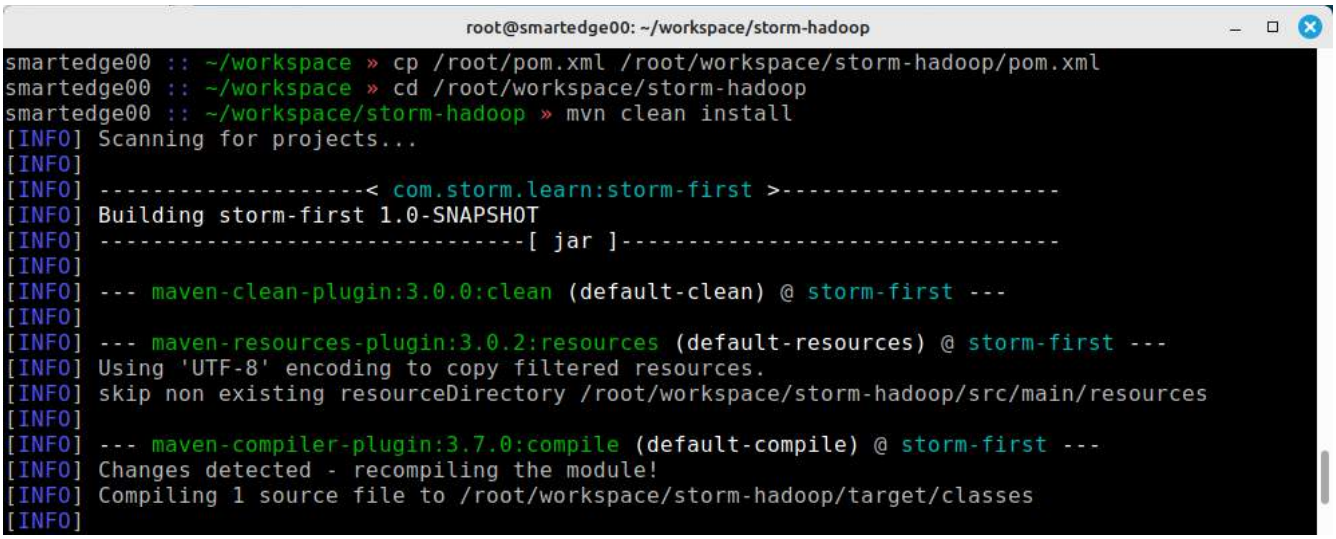

root@smartedge00: ~
smartedge00 :: ~ » docker run -itd -p 9000:9000 --network=br0 --name hadoop sequ
enceiq/hadoop-docker
cf3e1e929e6bf6f656ebf23a1b45123dd27d926768ed2823e20822f6ed9c9848
smartedge00 :: ~ » docker ps
CONTAINER ID IMAGE COMMAND CREATED
STATUS PORTS
NAMES
cf3e1e929e6b sequenceiq/hadoop-docker "/etc/bootstrap.sh -d" 38 seconds ag
o Up 36 seconds 2122/tcp, 8030-8033/tcp, 8040/tcp, 8042/tcp, 8088/tcp, 19888
/tcp, 49707/tcp, 50010/tcp, 50020/tcp, 50070/tcp, 50075/tcp, 50090/tcp, 0.0.0.0:
9000->9000/tcp, :::9000->9000/tcp hadoop
77b6ed67a98c centos:centos7.5.1804 "/bin/bash" 5 years ago
Up 2 hours 2181/tcp, 3772-3773/tcp, 6627/tcp, 8000/tcp, 8080/tcp
storm04
4481f8f056be centos:centos7.5.1804 "/bin/bash" 5 years ago
Up 2 hours 2181/tcp, 3772-3773/tcp, 6627/tcp, 8000/tcp, 8080/tcp
storm03
e32ed9f3ed34 centos:centos7.5.1804 "/bin/bash" 5 years ago


docker exec -it hadoop bash
smartedge00 :: ~ » docker exec -it hadoop bash
130
bash-4.1# cd $HADOOP_PREFIX
bash-4.1# bin/hdfs dfs -ls /
Found 1 items
drwxr-xr-x - root supergroup 0 2015-07-22 11:17 /user
bash-4.1#


docker exec -it hadoop bash
bash-4.1# ip a
1: lo: <LOOPBACK,UP,LOWER_UP> mtu 65536 qdisc noqueue state UNKNOWN qlen 1000
link/loopback 00:00:00:00:00:00 brd 00:00:00:00:00:00
inet 127.0.0.1/8 scope host lo
valid_lft forever preferred_lft forever
21: eth0@if22: <BROADCAST,MULTICAST,UP,LOWER_UP,M-DOWN> mtu 1500 qdisc noqueue s
tate UP
link/ether 02:42:c0:a8:00:0a brd ff:ff:ff:ff:ff:ff
inet 192.168.0.10/16 brd 192.168.255.255 scope global eth0
valid_lft forever preferred_lft forever
bash-4.1#

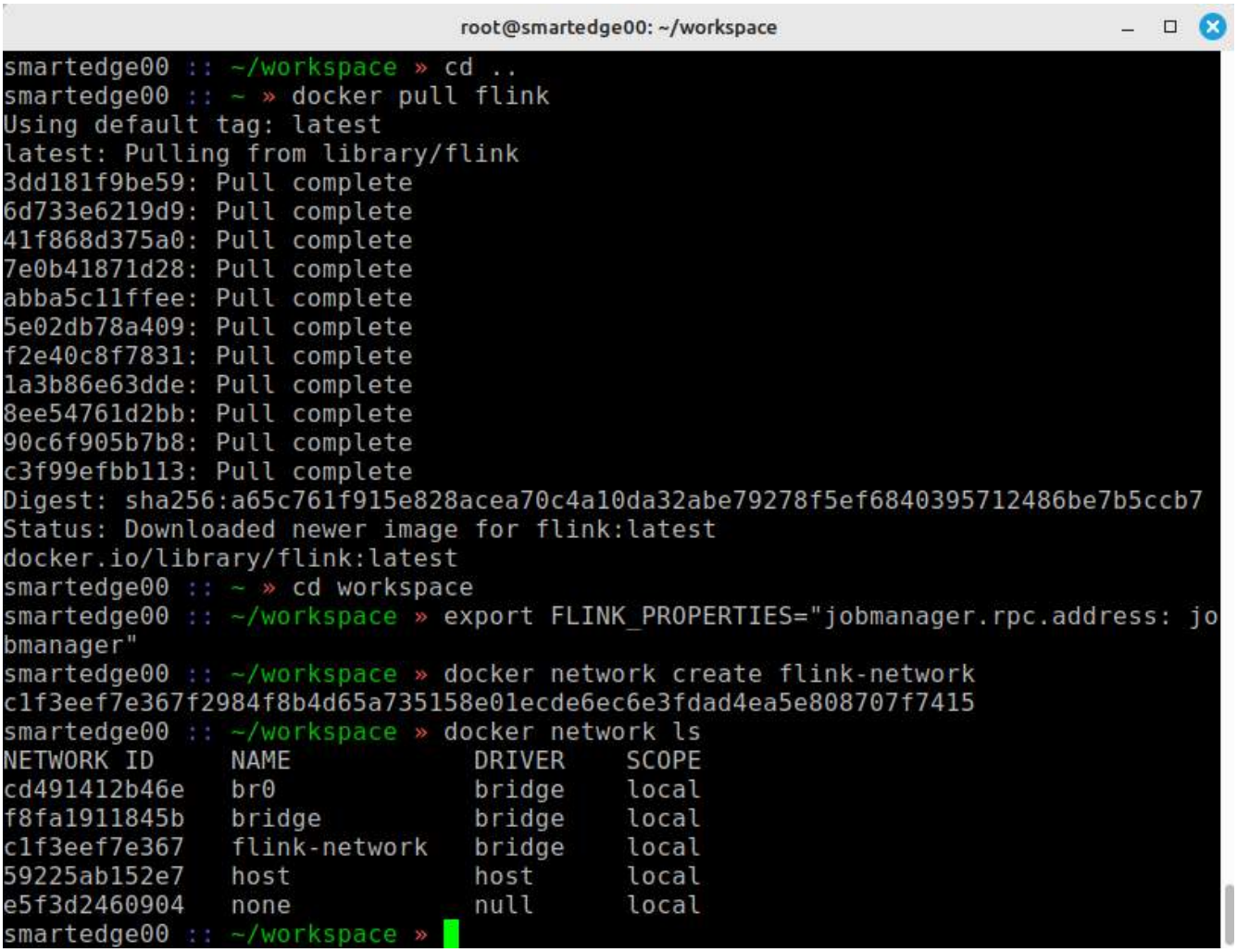

root@smartedge00: ~
smartedge00 :: ~ » docker inspect hadoop | grep IPAddress
"SecondaryIPAddresses": null,
"IPAddress": "",
"IPAddress": "192.168.0.10",
smartedge00 :: ~ »


root@smartedge00: ~/workspace
smartedge00 :: ~/workspace » mvn archetype:generate -DgroupId=com.storm.learn -DartifactId=storm-h
adoop -DarchetypeArtifactId=maven-archetype-quickstart -DarchetypeVersion=1.3
[INFO] Scanning for projects...
[INFO]
[INFO] ------------------< org.apache.maven:standalone-pom >------------------
[INFO] Building Maven Stub Project (No POM) 1
[INFO] --------------------------------[pom]---------------------------------
[INFO]
[INFO] >>> maven-archetype-plugin:3.0.1:generate (default-cli) > generate-sources @ standalone-pom
>>>
[INFO]
[INFO] <<< maven-archetype-plugin:3.0.1:generate (default-cli) < generate-sources @ standalone-pom
<<<
[INFO]
[INFO]
[INFO] --- maven-archetype-plugin:3.0.1:generate (default-cli) @ standalone-pom ---
[INFO] Generating project in Interactive mode
[INFO] Using property: groupId = com.storm.learn
[INFO] Using property: artifactId = storm-hadoop

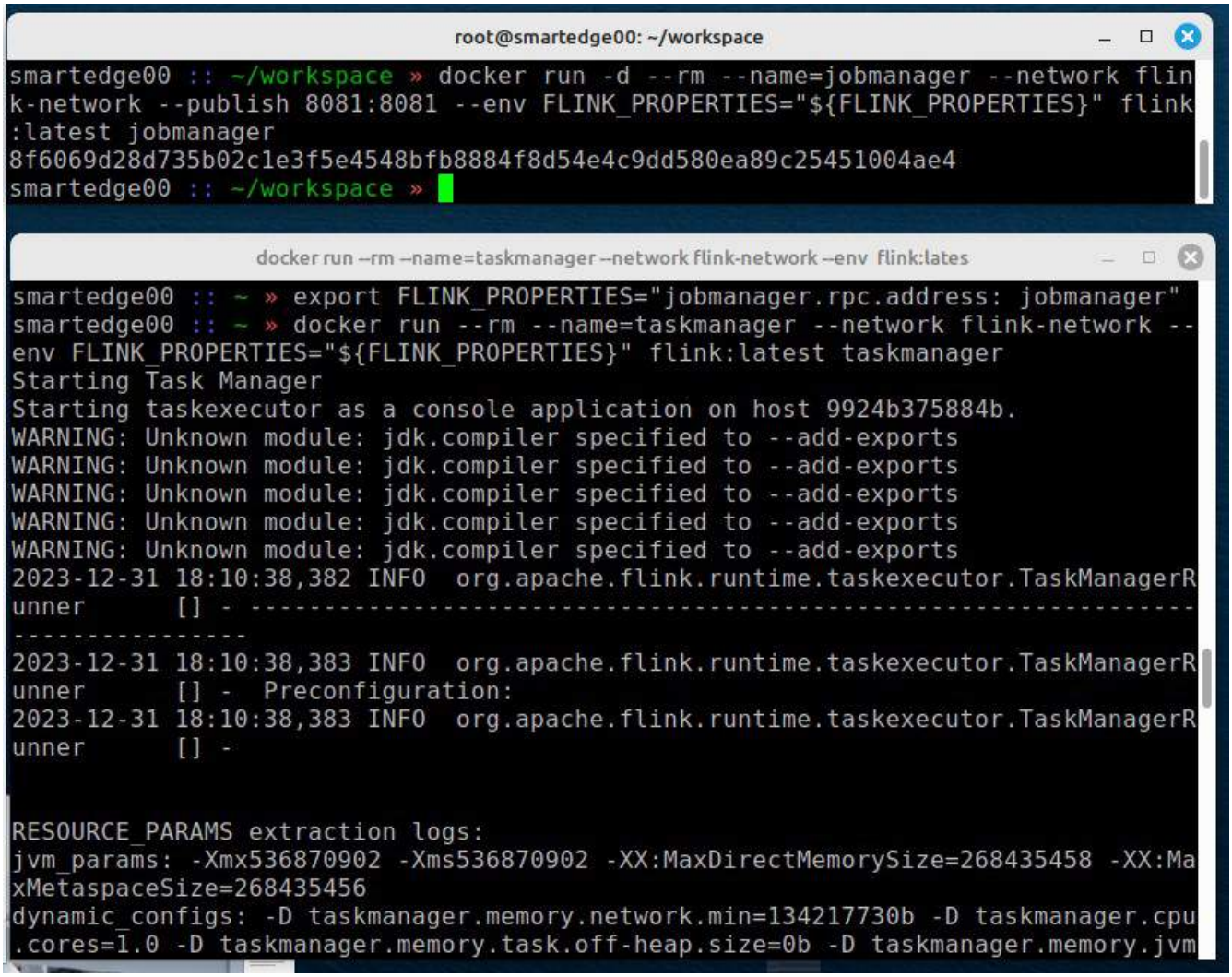

root@smartedge00: ~/workspace/storm-hadoop
smartedge00 :: ~/workspace » cp /root/pom.xml /root/workspace/storm-hadoop/pom.xml
smartedge00 :: ~/workspace » cd /root/workspace/storm-hadoop
smartedge00 :: ~/workspace/storm-hadoop » mvn clean install
[INFO] Scanning for projects...
[INFO]
[INFO] --------------------< com.storm.learn:storm-first >--------------------
[INFO] Building storm-first 1.0-SNAPSHOT
[INFO] --------------------------------[jar]---------------------------------
[INFO]
[INFO] --- maven-clean-plugin:3.0.0:clean (default-clean) @ storm-first ---
[INFO]
[INFO] --- maven-resources-plugin:3.0.2:resources (default-resources) @ storm-first ---
[INFO] Using 'UTF-8' encoding to copy filtered resources.
[INFO] skip non existing resourceDirectory /root/workspace/storm-hadoop/src/main/resources
[INFO]
[INFO] --- maven-compiler-plugin:3.7.0:compile (default-compile) @ storm-first ---
[INFO] Changes detected - recompiling the module!
[INFO] Compiling 1 source file to /root/workspace/storm-hadoop/target/classes
[INFO]

LAB 4

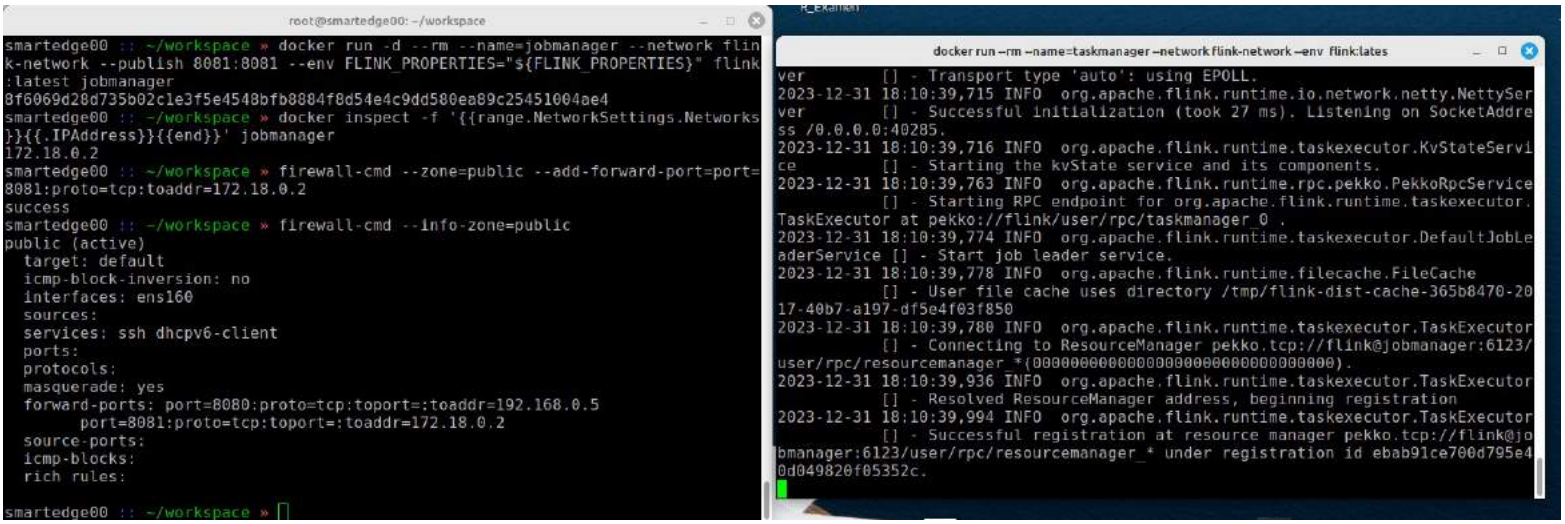
Aprovisionamos Apache Flink



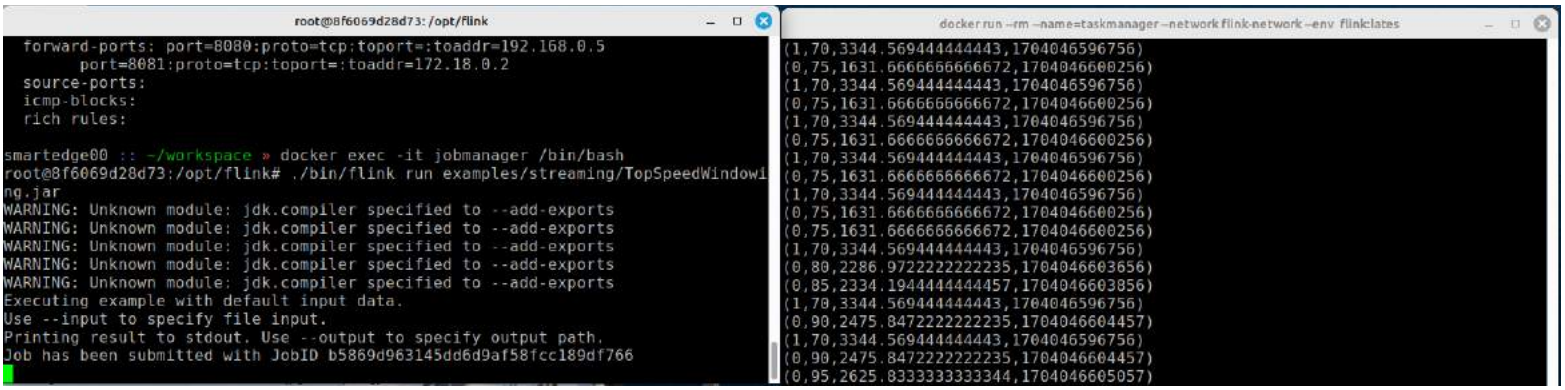
Ejecutar en dos terminales o ventanas divididas



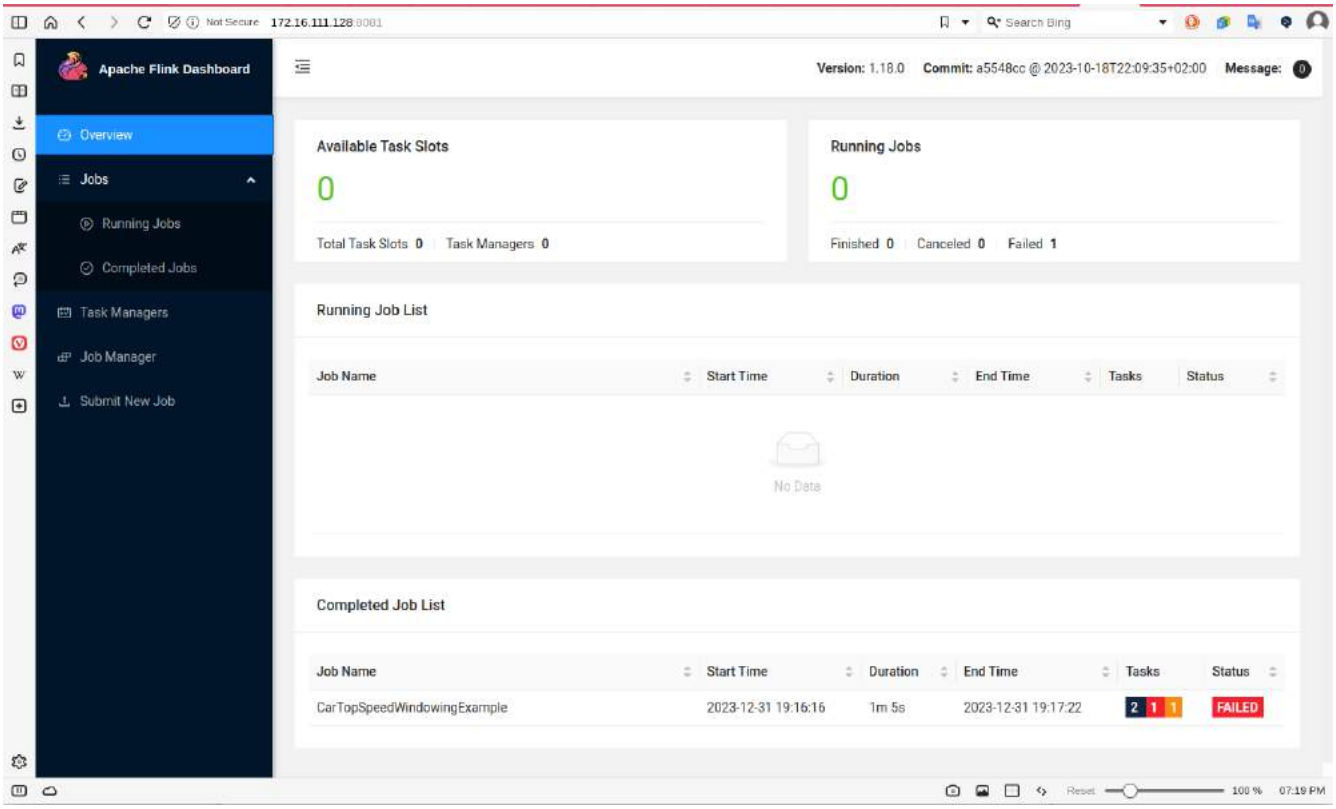
Configuramos firewall para poder acceder desde la máquina anfitriona

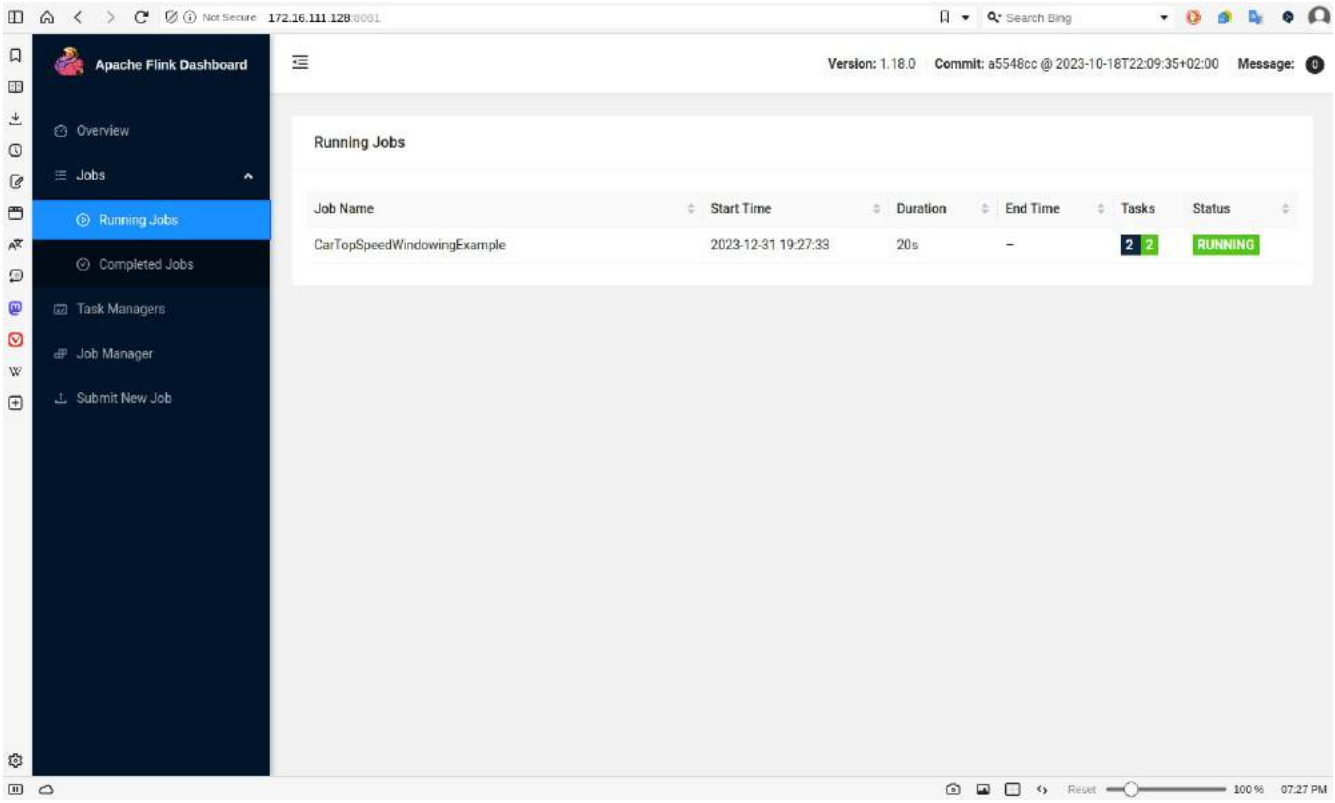


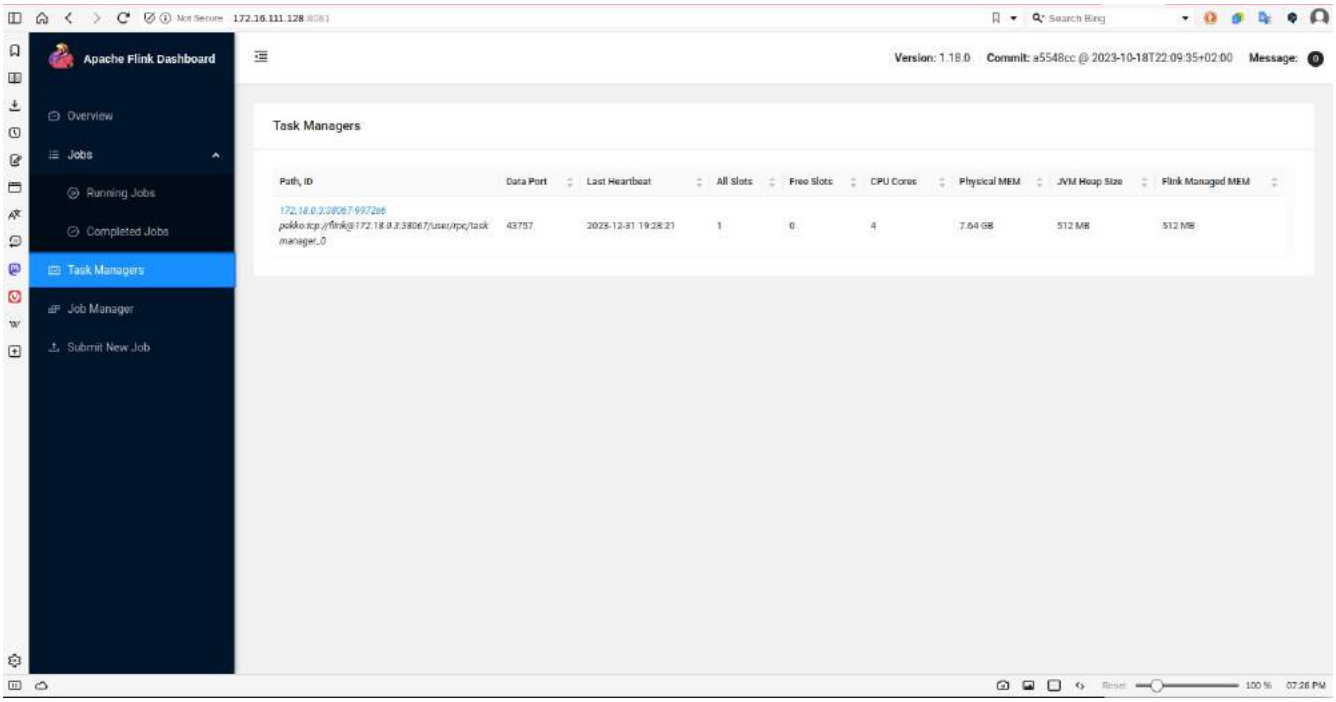
Ejecutamos ejemplos

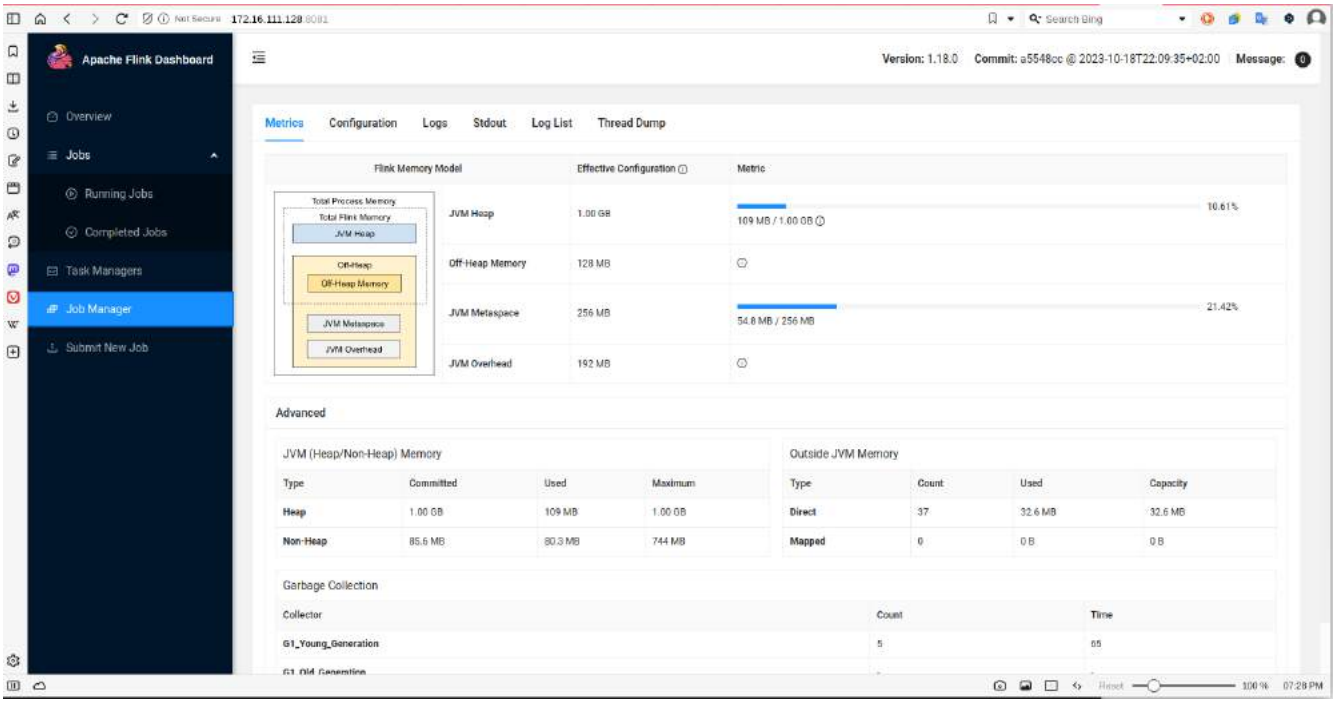


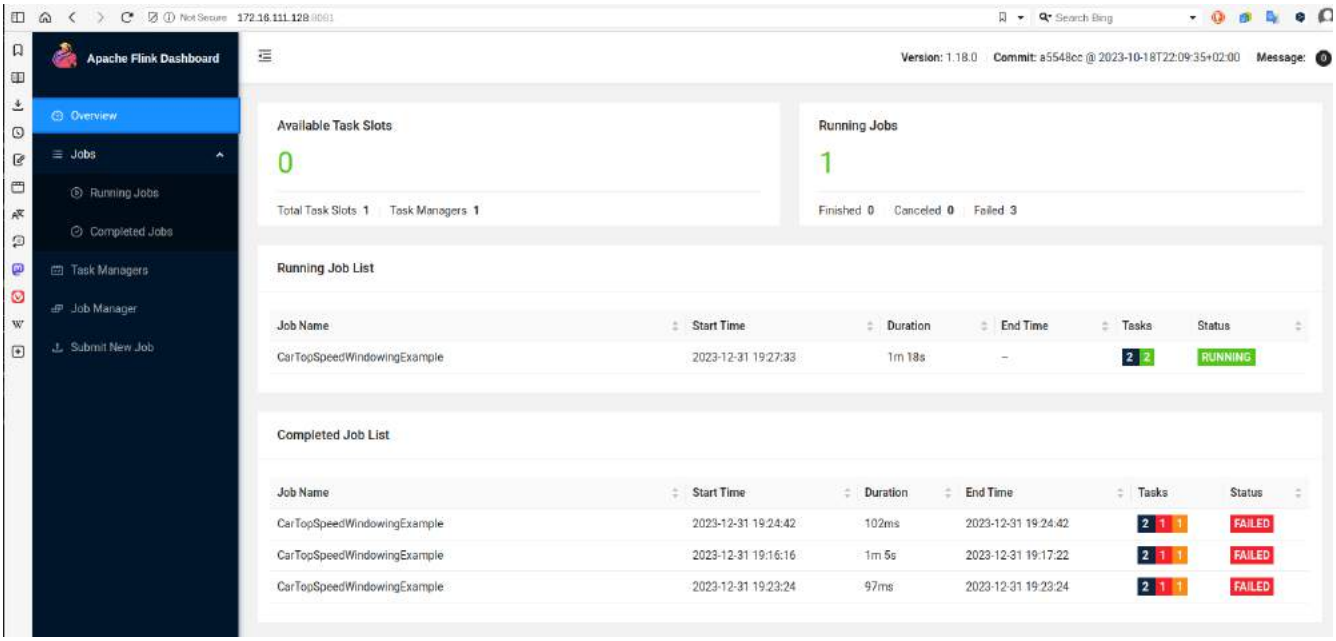
Podemos ver la consola gráfica desde el navegador en el sistema host
http://172.16.111.128:8081

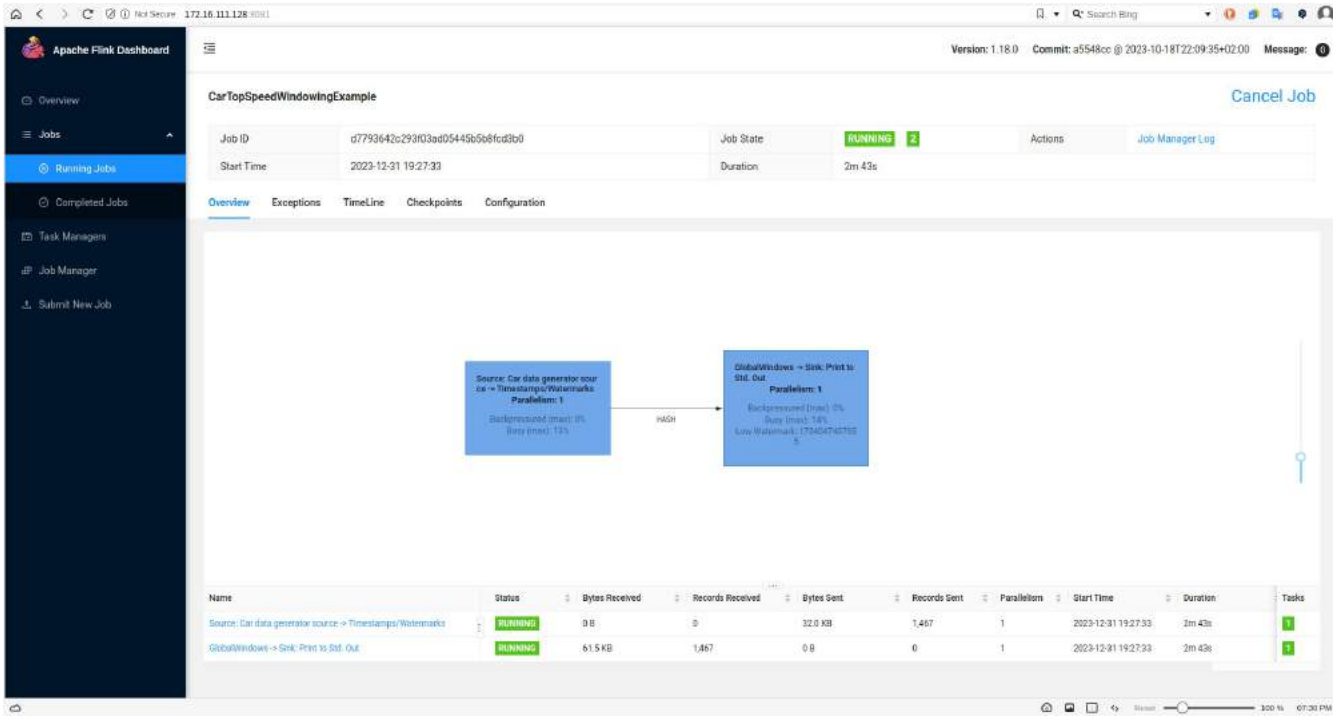




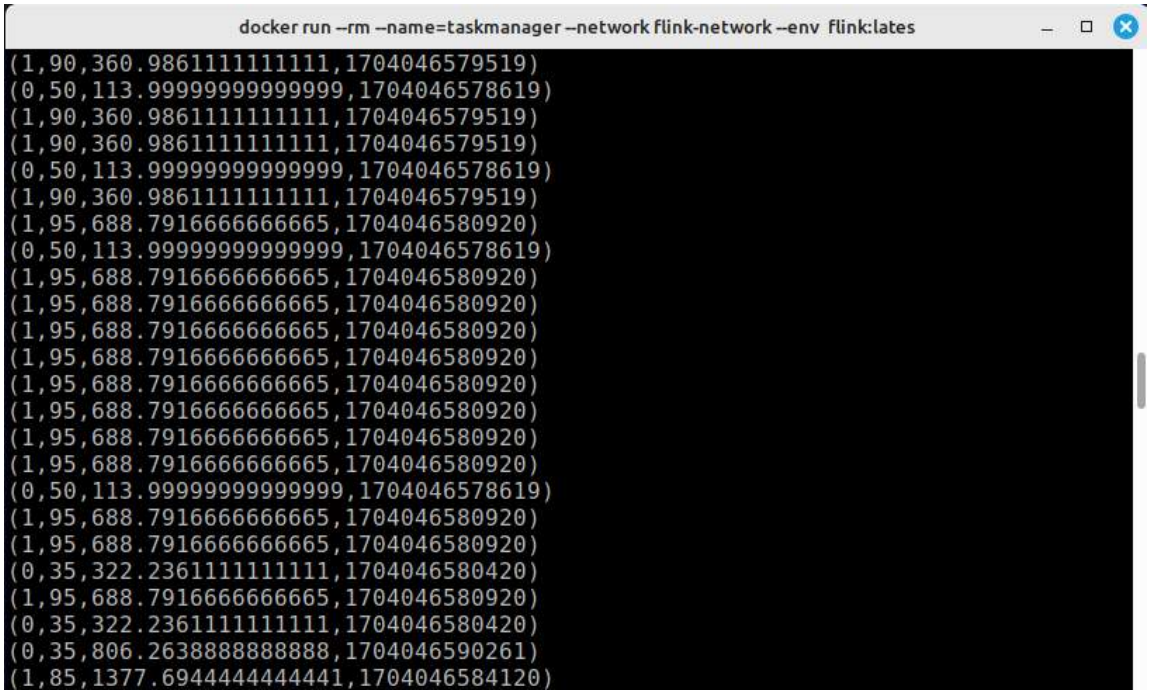

Apache Flink Dashboard
Overview
Jobs
Running Jobs
Completed Jobs
Task Managers
Job Manager
Submit New Job
Version: 1.18.0
Commit: a5548cc @ 2023-10-18T22:09:35+02:00
Message:
Task Managers
Path, ID
Data Port
Last Heartbeat
All Slots
Free Slots
CPU Cores
Physical MEM
JVM Heap Size
Flink Managed MEM
43757
2023-12-31 19:28:21
1
0
4
7.64 GB
512 MB
512 MB


Apache Flink Dashboard
Overview
Jobs
Running Jobs
Completed Jobs
Task Managers
Job Manager
Submit New Job
Version: 1.18.0
Commit: a5548cc @ 2023-10-18T22:09:35+02:00
Message:
Metrics
Configuration
Logs
Stdout
Log List
Thread Dump
Flink Memory Model
Effective Configuration
Metric
JVM Heap
1.00 GB
109 MB / 1.00 GB
10.61%
Off-Heap Memory
128 MB
JVM Metaspace
256 MB
54.8 MB / 256 MB
21.42%
JVM Overhead
192 MB
Advanced
JVM (Heap/Non-Heap) Memory
Type
Committed
Used
Maximum
Heap
1.00 GB
109 MB
1.00 GB
Non-Heap
85.6 MB
80.3 MB
744 MB
Outside JVM Memory
Type
Count
Used
Capacity
Direct
37
32.6 MB
32.6 MB
Mapped
0
0 B
0 B
Garbage Collection
Collector
Count
Time
G1_Young_Generation
5
65

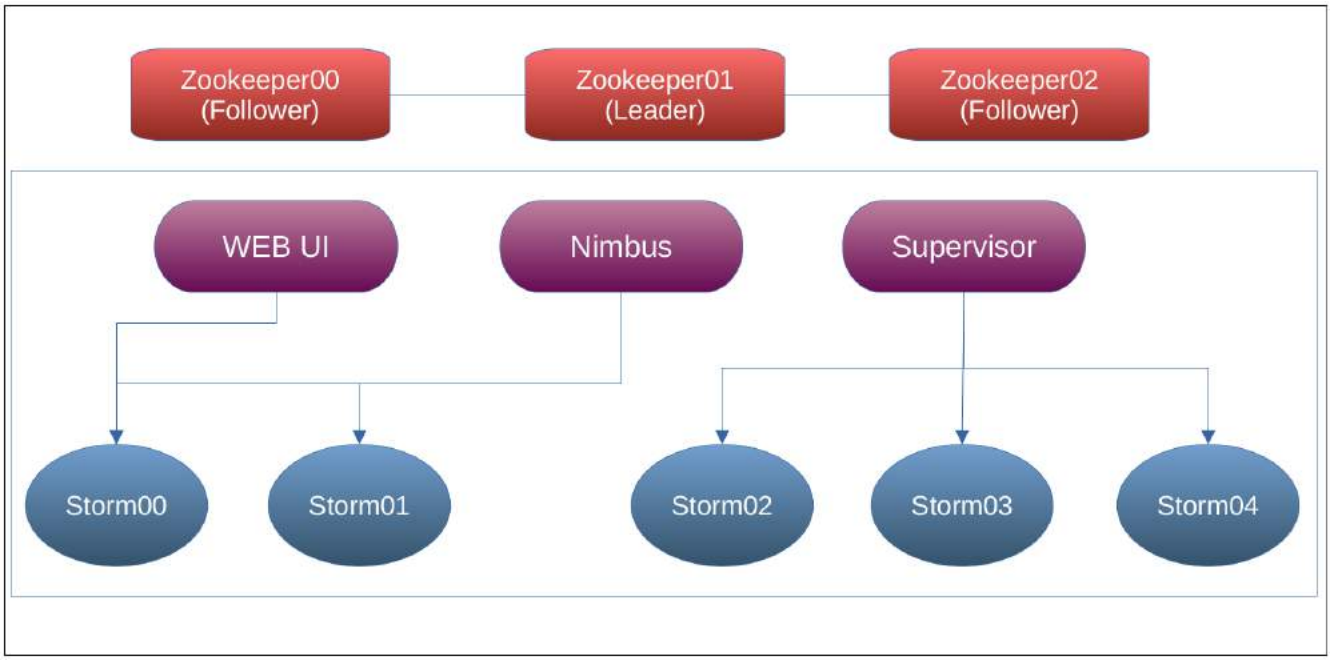

Apache Flink Dashboard
Version: 1.18.0 Commit: a5548cc @ 2023-10-18T22:09:35+02:00 Message:
Overview
Jobs
Running Jobs
Completed Jobs
Task Managers
Job Manager
Submit New Job
Available Task Slots
0
Total Task Slots 1 Task Managers 1
Running Jobs
1
Finished 0 Canceled 0 Failed 3
Running Job List
Job Name Start Time Duration End Time Tasks Status
CarTopSpeedWindowingExample 2023-12-31 19:27:33 1m 18s – 2 2 RUNNING
Completed Job List
Job Name Start Time Duration End Time Tasks Status
CarTopSpeedWindowingExample 2023-12-31 19:24:42 102ms 2023-12-31 19:24:42 2 1 1 FAILED
CarTopSpeedWindowingExample 2023-12-31 19:16:16 1m 5s 2023-12-31 19:17:22 2 1 1 FAILED
CarTopSpeedWindowingExample 2023-12-31 19:23:24 97ms 2023-12-31 19:23:24 2 1 1 FAILED


Apache Flink Dashboard
Version: 1.18.0 Commit: a5548cc @ 2023-10-18T22:09:35+02:00 Message:
Overview
Jobs
Running Jobs
Completed Jobs
Task Managers
Job Manager
Submit New Job
CarTopSpeedWindowingExample
Cancel Job
Job ID d7793642c293f03ad05445b5b8fcd3b0
Job State RUNNING 2
Actions Job Manager Log
Start Time 2023-12-31 19:27:33
Duration 2m 43s
Overview Exceptions TimeLine Checkpoints Configuration
HASH
Name Status Bytes Received Records Received Bytes Sent Records Sent Parallelism Start Time Duration Tasks
Source: Car data generator source -> Timestamps/Watermarks RUNNING 0 B 0 32.0 KB 1,467 1 2023-12-31 19:27:33 2m 43s 1
GlobalWindows -> Sink: Print to Std. Out RUNNING 61.5 KB 1,467 0 B 0 1 2023-12-31 19:27:33 2m 43s 1

Consola de taskmanager



Una imagen que represente el clúster de Storm utilizado en los laboratorios



Entregables

1. 1. ¿Qué es una topología, un Spout y un Bolt en Apache Storm?

En Apache Storm, una topología es un flujo de procesos formado por una serie de Bolts y Spouts. Las topologías se utilizan para manejar flujos de datos instantáneos y en constante cambio.

Un Bolt es la unidad de procesamiento de una topología. Los bolts se utilizan para recibir y procesar datos y generar otros nuevos. Los Bolts pueden recibir datos de uno o varios Spouts y enviarlos a otros Bolts o a fuentes externas.

Un Spout es la fuente de datos en una topología. Los Spouts reciben datos del mundo exterior y los envían a los Bolts. Los Spouts pueden recibir diferentes tipos de datos, como datos de sensores, datos de redes sociales, datos de tráfico web, etc.

Para crear una topología en Apache Storm, primero necesitamos extender las clases Bolt y Spout. Después, necesitamos definir las funciones de Bolts y Spouts.

2. ¿Qué tipos de procesamiento puedo ejecutar con Apache Storm en relación con la entrega de mensajes? ¿Puedes proporcionar una breve descripción y un caso de uso para cada uno de ellos?

Apache Storm es un sistema de procesamiento de datos en tiempo real que permite garantizar el procesamiento de diversos eventos en tiempo real, como analizar flujos de datos en tiempo real, realizar tareas de aprendizaje automático y organizar cálculos continuos. En términos de entrega de mensajes, Apache Storm puede procesar nuevos flujos de datos o actualizaciones de bases de datos en tiempo real, realizar peticiones continuas y procesar flujos continuos, transfiriendo los resultados del procesamiento al cliente en tiempo real. Además, una tormenta puede usarse para proporcionar concurrencia en la ejecución de consultas intensivas en recursos. Algunos tipos de procesamiento que pueden ejecutarse con Apache Storm en relación con la entrega de mensajes:

Best effort processing: Este es el modo de procesamiento por defecto en Apache Storm. No ofrece garantías de que un mensaje vaya a ser procesado, pero es el modo de procesamiento más rápido y eficiente.
Este modo es útil cuando la velocidad de procesamiento es más importante que la precisión del mensaje, como cuando se procesan datos no críticos.

At least once processing: Este modo garantiza que cada mensaje se procesará al menos una vez, pero puede procesarse varias veces.
Este modo es útil cuando es aceptable que los mensajes se procesen dos veces, como cuando se procesan datos que no son críticos pero que deben procesarse con precisión.

Exactly once processing: Este modo garantiza que cada mensaje se procesará exactamente una vez, asegurando que no haya duplicados.
Este modo es útil cuando es crítico evitar el procesamiento duplicado de mensajes, como cuando se procesan transacciones financieras u otros datos críticos.

3. ¿Una tarea ejecuta executors o un executor ejecuta tareas?

Los executors son los responsables de ejecutar las tareas.

En Apache Storm, un proceso worker ejecuta un subconjunto de una topología. Un worker process pertenece a una topología específica y puede ejecutar uno o más executors para uno o más componentes (spouts o bolts) de esta topología. Un executors es un hilo generado por un worker process. Puede ejecutar una o más tareas para el mismo componente (spouts o bolts). Una tarea realiza el procesamiento real de los datos: cada spouts o bolts que implementamos en código ejecuta tantas tareas en todo el clúster. Por lo tanto, un executors ejecuta tareas en Apache Storm.

4. ¿Qué comando debo ejecutar para cambiar el paralelismo de una topología en funcionamiento?

Para cambiar el paralelismo de una topología en ejecución en Apache Storm, podemos utilizar el comando rebalance del cliente de línea de comandos storm.

$ storm rebalance <topology-name> -n <new-number-of-workers> -e <component-id>:<new-number-of-executors>

$ storm rebalance mytopology -n 5 -e blue-spout=3 -e yellow-bolt=10

mytopology : Es el nombre de la topología que deseamos rebalance.

-n 5: Esta opción establece el número de workers en 5.

-e blue-spout=3: Esta opción establece el número de executors para el componente blue-spout en 3.

-e yellow-bolt=10: Esta opción establece el número de executors para el componente yellow-bolt en 10.

Así, el comando que hemos proporcionado cambiaría el paralelismo de la topología mytopology para utilizar 5 workers, 3 executors para el componente blue-spout y 10 executors para el componente yellow-bolt.

5. Si uno de los nodos que forman mi clúster de Apache Storm es de alto rendimiento (ej GPU), ¿qué planificador debería utilizar para extraer todo su valor y optimizar los recursos?

Para utilizar eficazmente un nodo de alto rendimiento (por ejemplo, una GPU) en nuestro clúster Apache Storm, empleamos el ResourceAwareScheduler.
Este algoritmo de planificación prioriza las tareas específicas de los recursos, asignándolas al nodo con las capacidades más adecuadas. Garantiza que las tareas se asignen al nodo equipado para gestionar sus requisitos de recursos, maximizando la utilización del nodo de alto rendimiento y evitando al mismo tiempo que otros nodos se sobrecarguen.

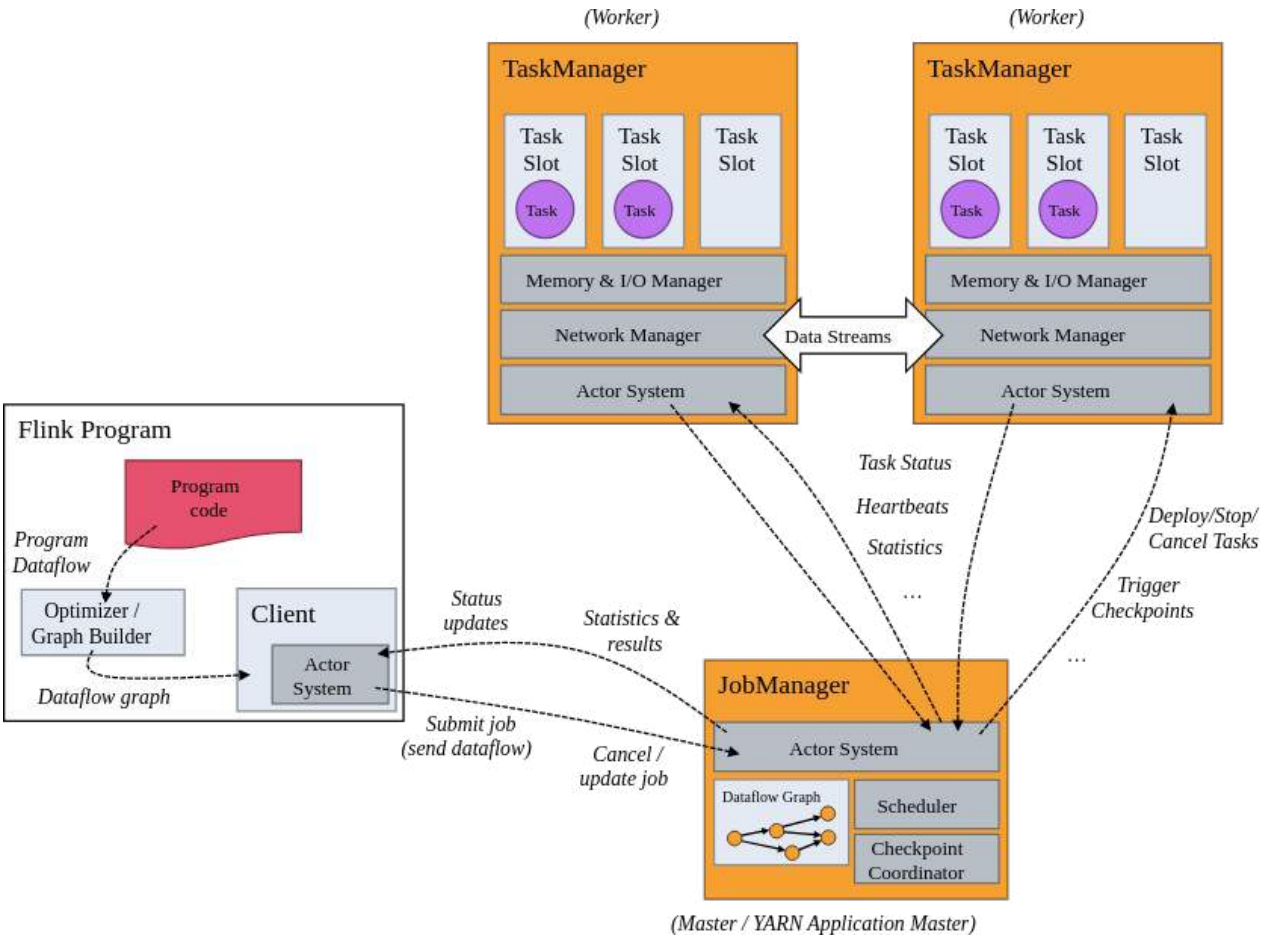
Maximiza la utilización de los recursos: El ResourceAwareScheduler garantiza que las tareas se asignen a nodos que dispongan de los recursos necesarios para ejecutarse de forma eficiente. Esto ayuda a maximizar la utilización del nodo de alto rendimiento con capacidades de GPU, ya que las tareas que pueden beneficiarse de sus capacidades se asignan a este nodo.

Evita la sobrecarga: El ResourceAwareScheduler también evita la sobrecarga de un solo nodo. Esto ayuda a garantizar que el clúster funcione de forma fluida y eficiente, y que ningún nodo se vea obligado a trabajar por encima de su capacidad.

Simplifica la configuración: ResourceAwareScheduler es relativamente fácil de configurar. Basta con especificar los recursos necesarios para cada tarea, como el uso mínimo y máximo de memoria y GPU. El programador asignará las tareas a los nodos en función de estos requisitos.

En resumen, ResourceAwareScheduler es la mejor opción para optimizar la utilización de recursos y extraer todo el valor de un nodo de alto rendimiento con capacidades de GPU en nuestro cluster de Apache Storm. Es una forma sencilla y eficaz de garantizar que nuestro cluster funcione de forma fluida y eficiente, y que todos nuestros recursos se utilicen al máximo de su potencial.

6. Describe la arquitectura de ejecución de trabajos en Apache Flink.



La arquitectura de ejecución de trabajos de Apache Flink está diseñada para el procesamiento de datos distribuidos con requisitos de alto rendimiento, baja latencia y para ser altamente escalable y tolerante a fallos.

Puede gestionar trabajos de procesamiento de datos a gran escala y recuperarse de fallos sin perder datos.

La arquitectura de ejecución de trabajos consta de varios componentes:

JobManager: El JobManager es el nodo master de un cluster Flink. Es responsable de coordinar la ejecución de los trabajos Flink. Sólo hay un JobManager en un cluster Flink. El JobManager realiza la coordinación de los trabajos, la planificación y la recuperación de tareas en caso de fallos. También gestiona el despliegue y la coordinación de los TaskManagers.

Task Managers: Los Task Managers son los nodos de worker de un clúster Flink. Se encargan de ejecutar las tareas y de mantener el estado del flujo de datos. Los Task Managers se ejecutan en los nodos workers del clúster y se comunican con el Job Manager para recibir instrucciones e informar de su progreso.

Job Client: Job Client es la aplicación que envía trabajos al clúster Flink. Se encarga de comunicarse con el gestor de trabajos y de monitorizar el progreso del trabajo.

La arquitectura de ejecución de trabajos de Flink funciona de la siguiente manera:

1. La aplicación envía un trabajo a Job Client.
2. Job Client envía la tarea a Job Manager.
3. Job Manager analiza el trabajo y determina las tareas que deben ejecutarse.
4. Job Manager asigna las tareas a Task Managers del clúster.
5. Task Managers ejecutan las tareas y envían los resultados a Job Manager.
6. Job Manager comprueba los resultados y decide si el trabajo se ha completado correctamente.

Si falla un Task Manager, el Job Manager reiniciará la tarea que ha fallado en otro Task Manager del clúster. Si falla el Job Manager, el Job Client puede volver a enviar la tarea al clúster.

Los mecanismos de tolerancia a fallos de Flink garantizan que los trabajos puedan seguir ejecutándose aunque fallen algunos de los Task Managers.

Los mecanismos de escalabilidad de Flink permiten ampliar o reducir fácilmente el clúster para gestionar cargas de trabajo cambiantes.

Estas características hacen de Flink una herramienta potente y fiable para aplicaciones de procesamiento de flujos.