

MONITORIZACIÓN DE CLÚSTERES

ALFREDO TALAVERA RAMAJO

Vamos a crear una carpeta en hdfs llamada trabajo con el comando: hdfs dfs -mkdir /trabajo

Y con el comando:

hadoop jar ContarPalabras.jar ContarPalabras /prueba/access_log /trabajo/salida subimos a hdfs el archivo ContarPalabras

Encendemos dfs, yarn y el history en el maestro

The screenshot shows a Linux desktop environment with three terminal windows and a Firefox browser window.

- Firefox Window:** Shows the Hadoop cluster metrics and nodes list. Cluster Metrics table:

	Submitted	Pending	Running	Completed	Containers Running	Used Resources
0	0	0	0	0	<memory:0, vCores:0>	

 Cluster Nodes Metrics table:

	Active Nodes	Decommissioning Nodes	Decommissioned Nodes	Lost Nodes	Unhealthy Nodes
2	0	0	0	0	0

 Scheduler Metrics table:

Scheduler Type	Scheduling Resource Type	Minimum Allocation	Maximum Allocation	Maximum Cluster Application Priority
Capacity Scheduler	<name=memory-mb default-unit=Mb type=COUNTABLE>, <name=vcores default-unit= type=COUNTABLE>	<memory:1024, vCores:1>	<memory:1024, vCores:4>	
- Top Terminal Window (agora@nodo3):**

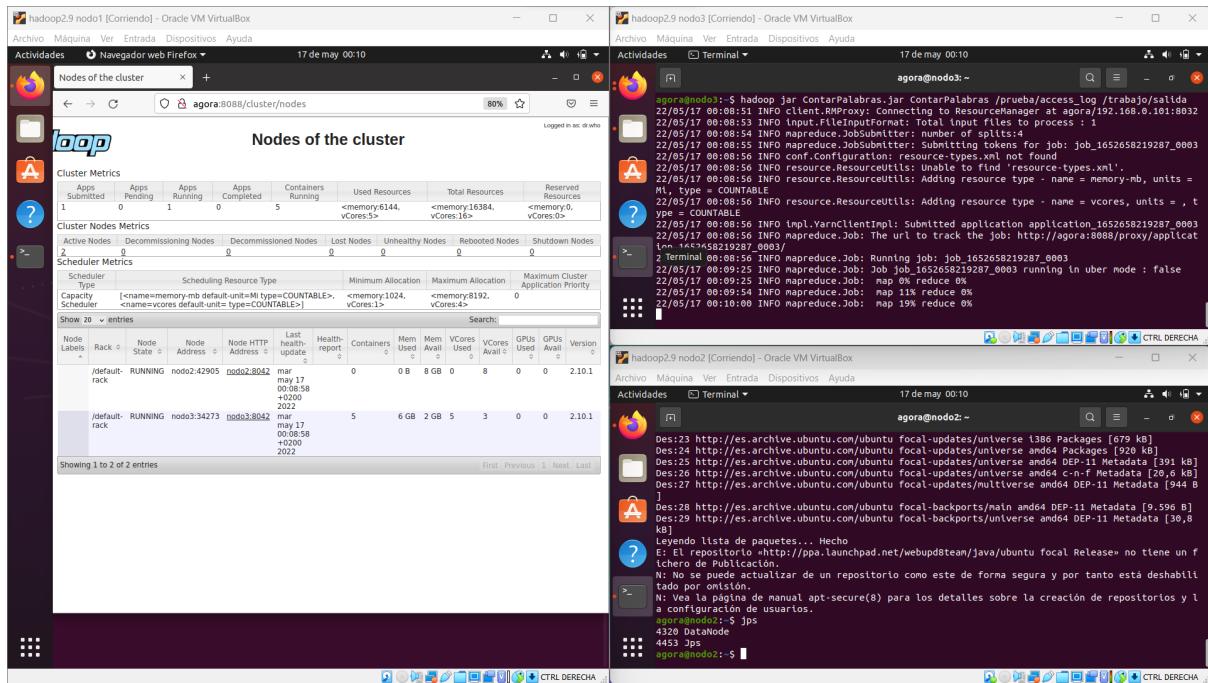
```
deslizamientos_retocado.csv order_reviews dataset.csv
agora@nodo3:~$ javac ContarPalabras.jar
error: class names 'ContarPalabras', are only accepted if annotation processing is explicitly requested
1 error
agora@nodo3:~$ jps
11371 Jps
11274 DataNode
agora@nodo3:~$ nano ContarPalabras.jar
agora@nodo3:~$ hdfs dfs -mkdir /trabajo
agora@nodo3:~$ hdfs dfs -ls
ls: /: No such file or directory
agora@nodo3:~$ hdfs dfs -ls /
Found 3 items:
drwxr-xr-x  2 agora supergroup 0 2022-05-11 18:40 /prueba
drwxr-xr-x  2 agora supergroup 0 2022-04-21 18:47 /tmp
drwxr-xr-x  2 agora supergroup 0 2022-05-17 00:06 /trabajo
agora@nodo3:~$
```
- Bottom Terminal Window (agora@nodo2):**

```
Des:23 http://es.archive.ubuntu.com/ubuntu focal-updates/universe 1386 Packages [679 kB]
Des:24 http://es.archive.ubuntu.com/ubuntu focal-updates/universe andd4 Packages [926 kB]
Des:25 http://es.archive.ubuntu.com/ubuntu focal-updates/universe andd4 DEP-11 Metadata [391 kB]
Des:26 http://es.archive.ubuntu.com/ubuntu focal-updates/universe andd4 c-n-f Metadata [20,6 kB]
Des:27 http://es.archive.ubuntu.com/ubuntu focal-updates/multiverse andd4 DEP-11 Metadata [944 kB]
Des:28 http://es.archive.ubuntu.com/ubuntu focal-backports/main andd4 DEP-11 Metadata [9,596 kB]
Des:29 http://es.archive.ubuntu.com/ubuntu focal-backports/universe andd4 DEP-11 Metadata [30,8 kB]
Leyendo lista de paquetes... Hecho
E: El repositorio «http://ppa.launchpad.net/webupdteam/java/ubuntu focal Release» no tiene un fichero de Publicación.
N: No se puede actualizar de un repositorio como este de forma segura y por tanto está deshabilitado por omisión
N: Vea la página de manual apt::secure(8) para los detalles sobre la creación de repositorios y la configuración de usuarios.
agora@nodo2:~$ jps
4320 DataNode
4 Mostrar aplicaciones
```

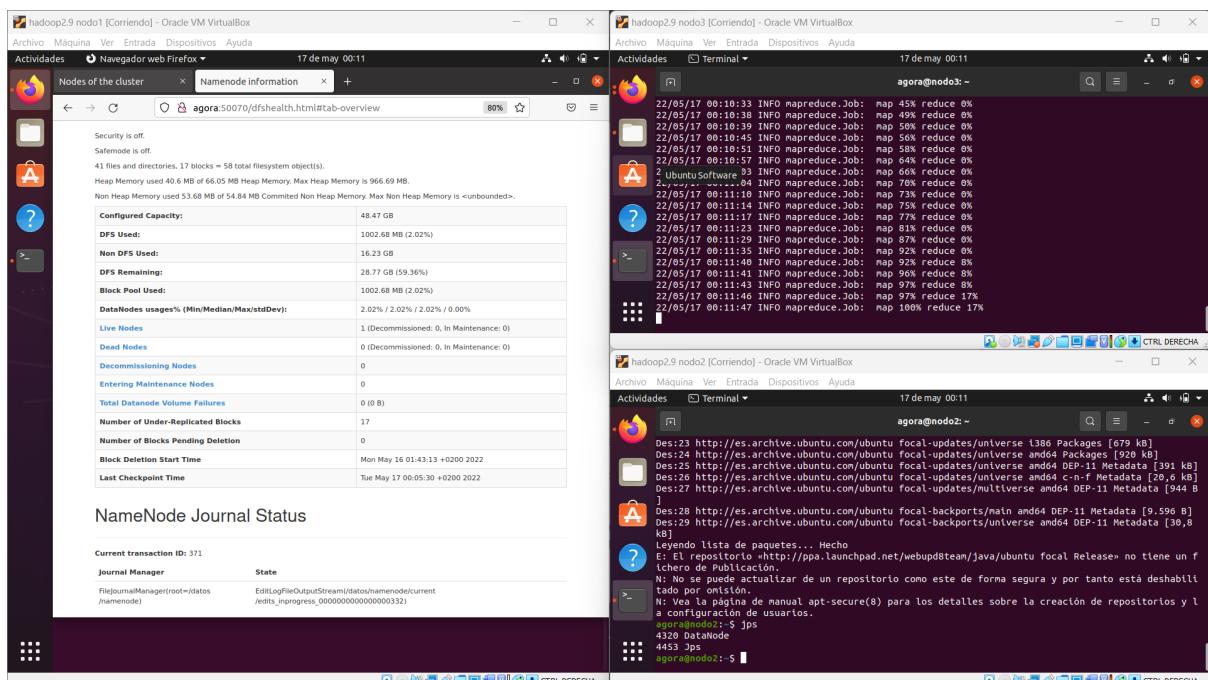
Mientras se va subiendo el archivo ContarPalabras, entramos al navegador en la dirección:

agora:8088

Y accederemos a una página que nos da la información que en el nodo3 cuando va por un 19% nos indica que está gastando 6GB de memoria y hay 2GB de memoria disponible, mientras que en el nodo2 que no estamos haciendo nada no está consumiendo nada



Si accedemos a la dirección agora:50070, accederemos a otra página que nos dirá que hay un 2% de la memoria DFS utilizada



Vemos que nos brinda informacion de los datanodes la cual nos dice que la capacidad total del nodo2 la capacidad es de 48GB y hay un 2% utilizandose

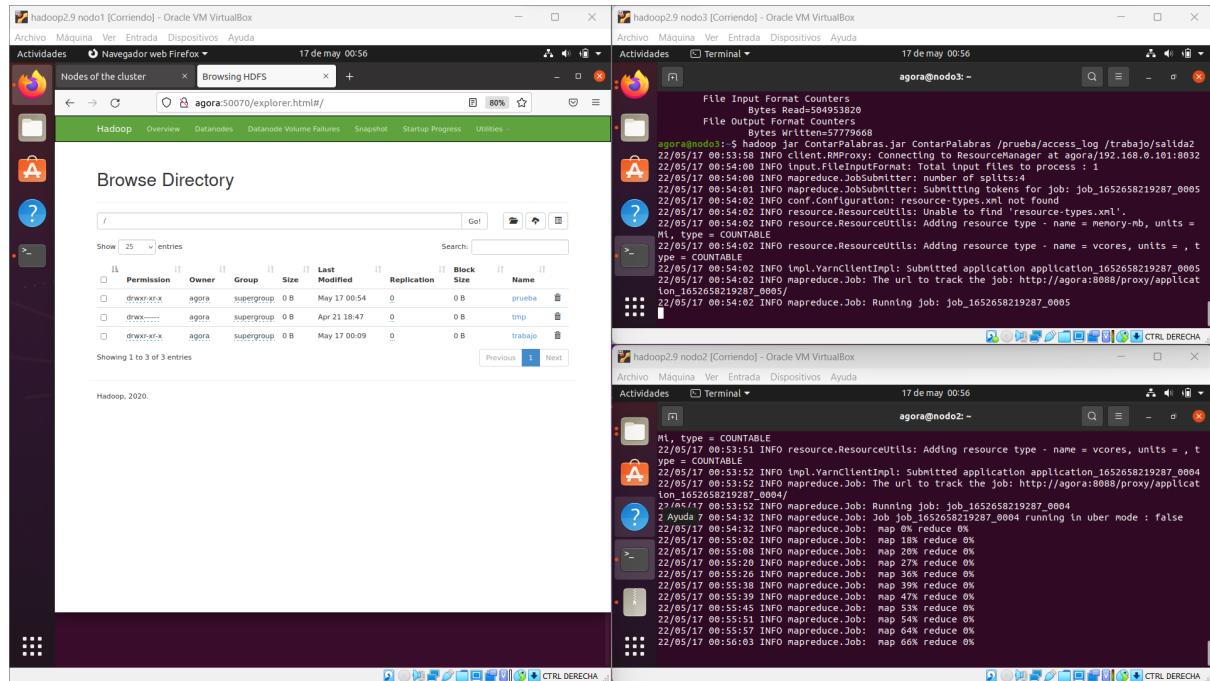
Vemos que cuando ha llegado a un 100% el nodo3 vuelve a que hay disponible 8GB de memoria

Ahora vamos a volver a la dirección de agora:8088, pero esta vez vamos a utilizar los dos esclavos y vamos a subir el mismo archivo anterior a dfs

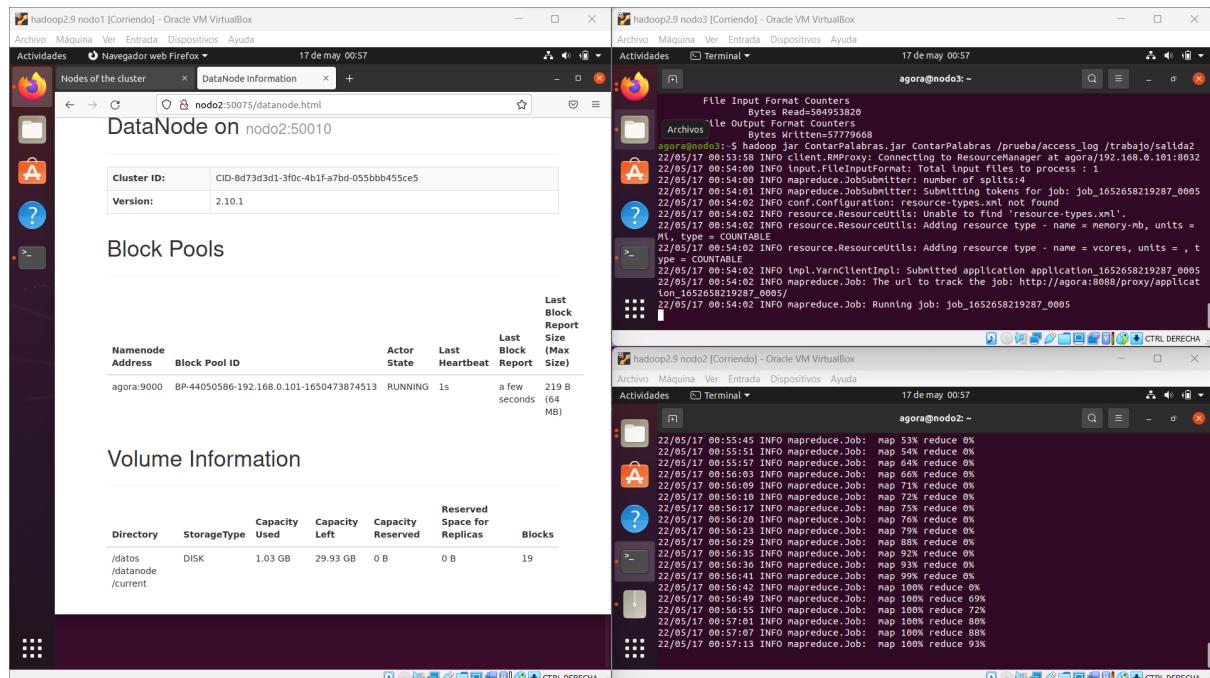
Vemos que primero se sube uno y luego el otro y que los dos gastan memoria mientras se esta subiendo simultáneamente

Node Labels	Rack	State	Node Address	Node HTTP Address	Last health-report	Health-report	Containers	Mem Used	Mem Avail	Vcores Used	Vcores Avail	GPUs Used	GPUs Avail	Version
/default-/rack	RUNNING	nodo2:42905	nodo2:8042	mar 17 00:52:58 +0200	2022		1	2 GB	6 GB	1	7	0	0	2.10.1
/default-/rack	RUNNING	nodo3:34273	nodo3:8042	mar 17 00:52:58 +0200	2022		4	4 GB	4 GB	4	0	0	0	2.10.1

Podemos entrar a ver los directorios que se han creado prueba, tmp y trabajo, desde agora:50070

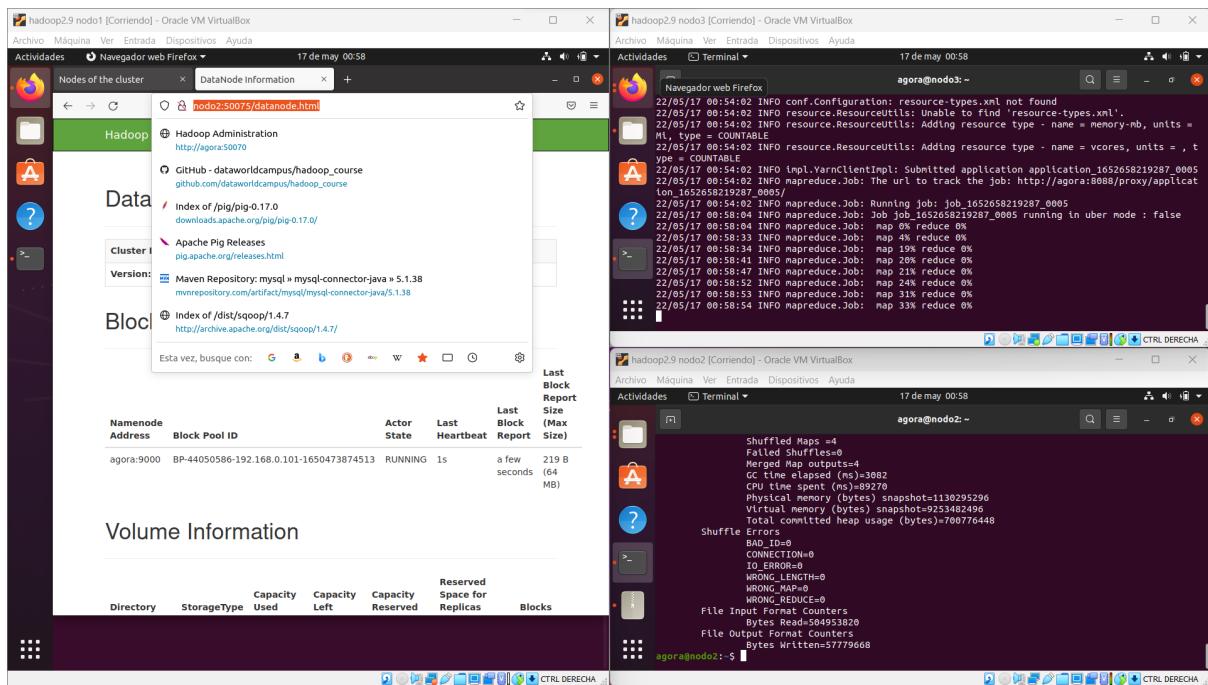


También podemos ver información del nodo2, tanto de la ID, la dirección con el puerto, los directorios creados, el volumen que ocupa...

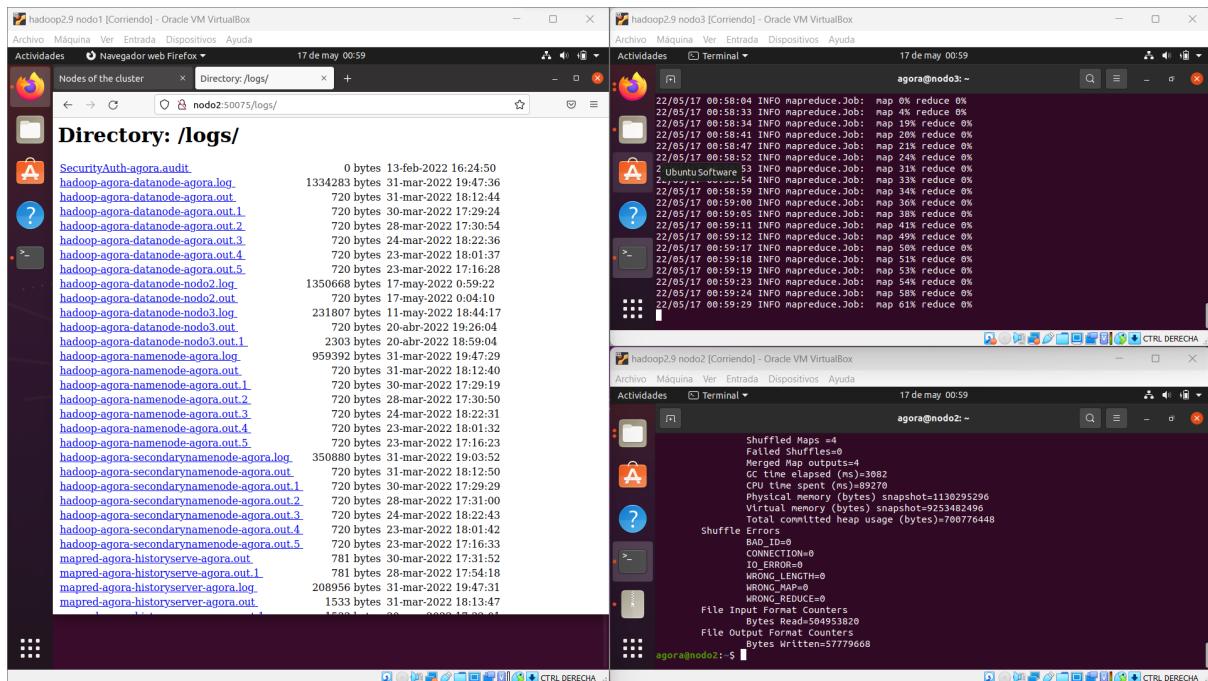


Podemos ver tambien la misma informaci髇 anterior pero del otro nodo

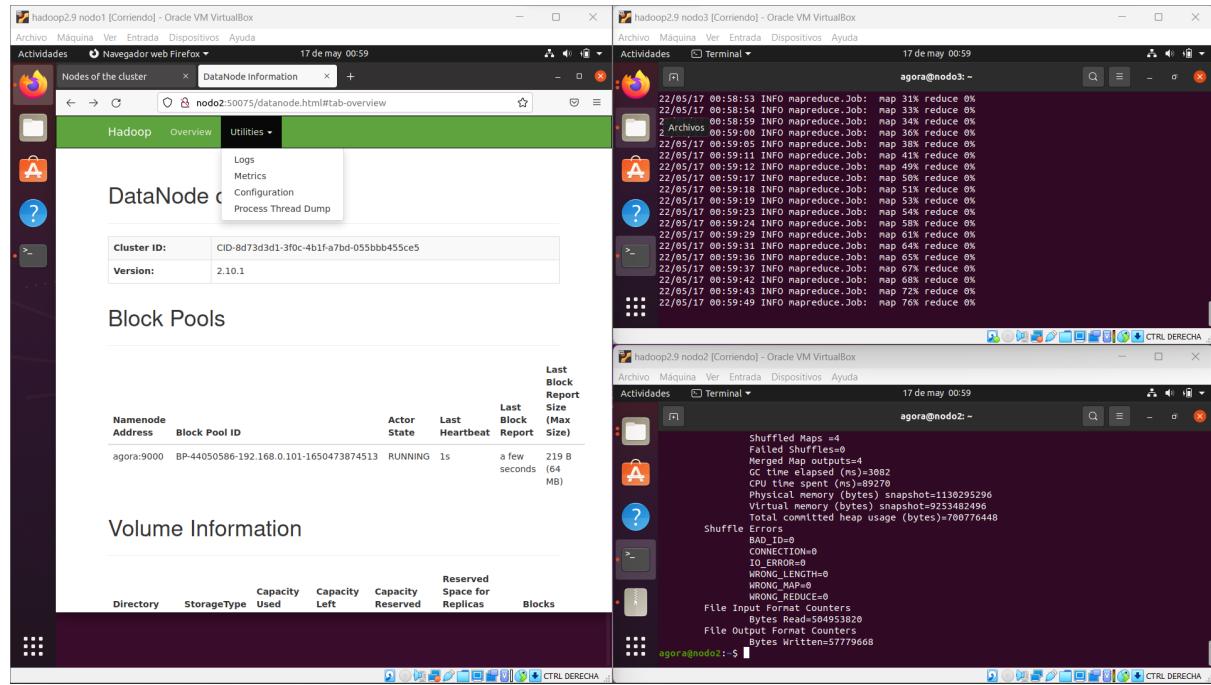
Para ver la informaci髇 de un nodo en concreto, solo nos basta con escribir el nombre del nodo y :50070



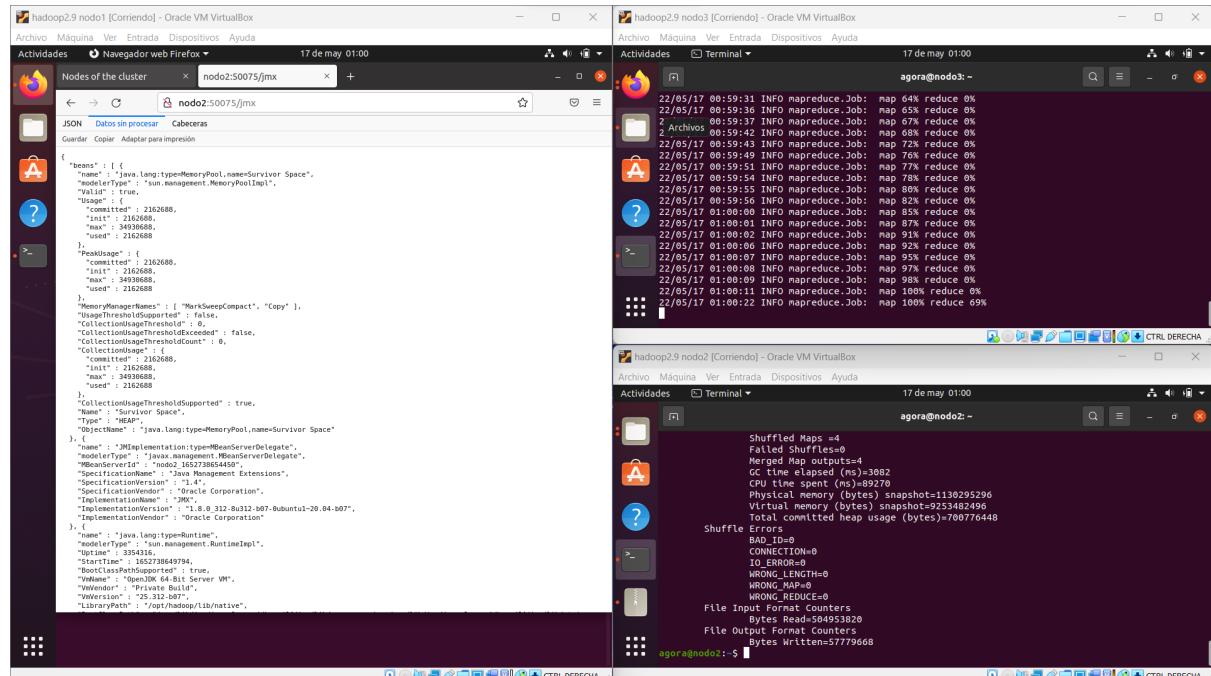
Nos dejará ver los log, que nos permite ver qué ha pasado, en los archivos ya que queda todo registrado ahí y podemos ver las incidencias con fecha y hora además



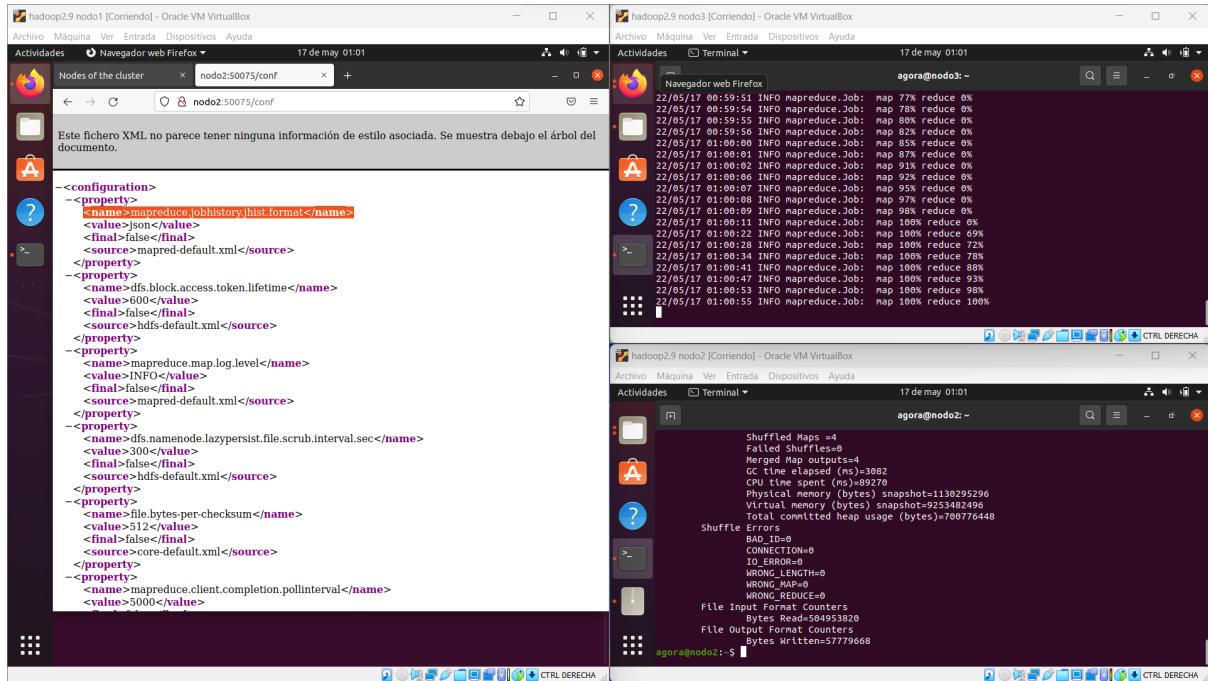
Podemos ver las utilidades que tenemos utilidades donde podemos acceder a los Log, las métricas, acceder a la configuración...



También nos permite ver en formato JSON datos sin procesar y datos ya procesados



O en fichero XML donde podemos interpretar diferentes datos



Si le damos a beans vemos cabeceras, bloques, nombres... Que se podrían interpretar

