

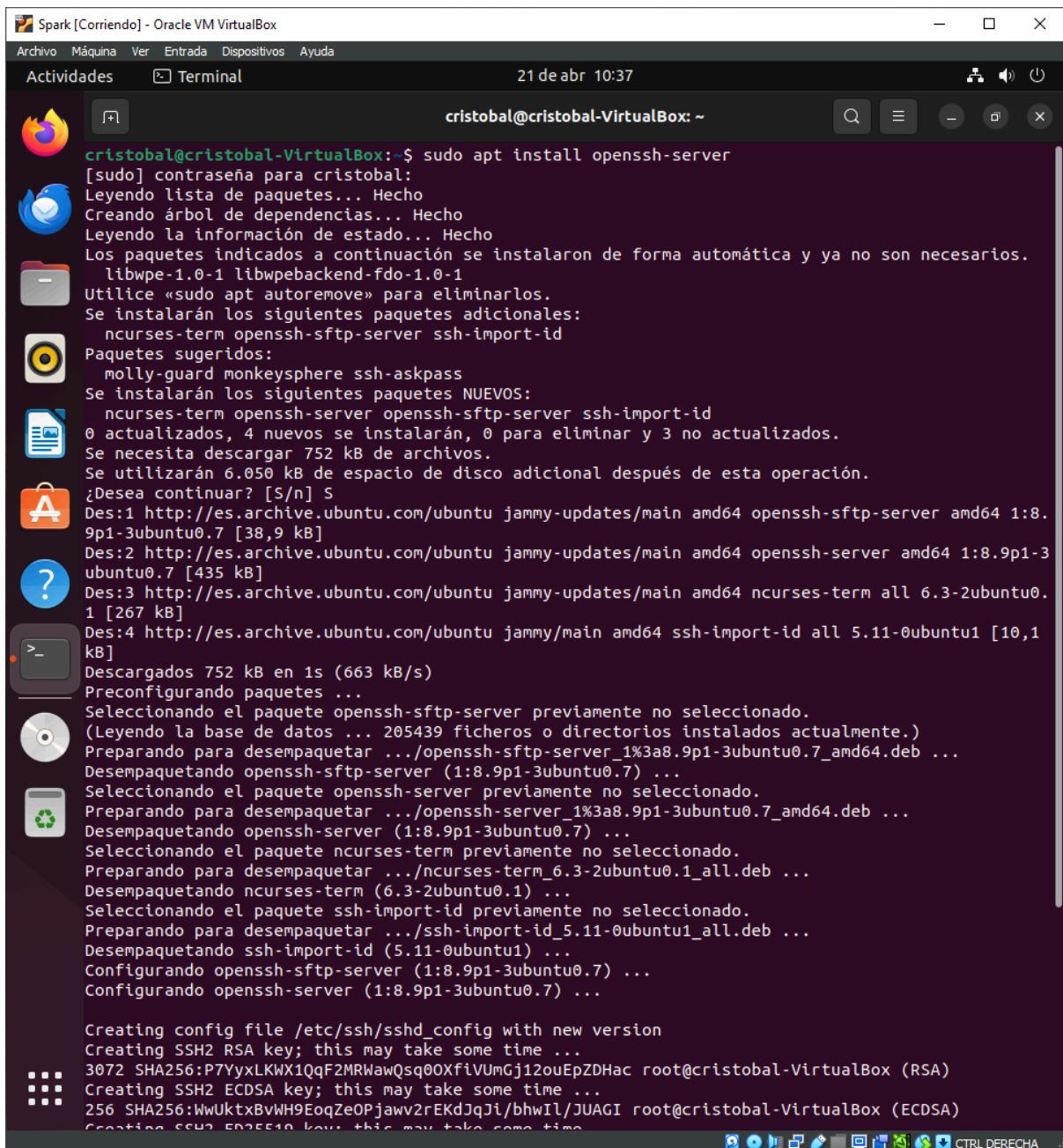
Apartat 1

Seguint les instruccions que hem vist en l'apartat 6 dels apunts, crea un clúster sobre Docker amb 1 node mestre, 3 nodes worker i un node per a JupyterLab. Pots emprar les mateixes imatges que hem fet servir als apunts. El worker 3 ha d'exposar el port 8083 per a la seva interfície web.

Fes un document amb Word, Write o Docs on has d'explicar les passes que has seguit i incloure captures de pantalla de la interfície web de cada un dels nodes del clúster.

Ejecutar primero de todo 'sudo apt update' y 'sudo apt upgrade'

Instalar ssh por comodidad (opcional)



```
Spark [Corriendo] - Oracle VM VirtualBox
Archivo Máquina Ver Entrada Dispositivos Ayuda
Actividades Terminal 21 de abr 10:37
cristobal@cristobal-VirtualBox: ~
cristobal@cristobal-VirtualBox:~$ sudo apt install openssh-server
[sudo] contraseña para cristobal:
Leyendo lista de paquetes... Hecho
Creando árbol de dependencias... Hecho
Leyendo la información de estado... Hecho
Los paquetes indicados a continuación se instalaron de forma automática y ya no son necesarios.
libwpe-1.0-1 libwpebackend-fdo-1.0-1
Utilice «sudo apt autoremove» para eliminarlos.
Se instalarán los siguientes paquetes adicionales:
ncurses-term openssh-sftp-server ssh-import-id
Paquetes sugeridos:
molly-guard monkeysphere ssh-askpass
Se instalarán los siguientes paquetes NUEVOS:
ncurses-term openssh-server openssh-sftp-server ssh-import-id
0 actualizados, 4 nuevos se instalarán, 0 para eliminar y 3 no actualizados.
Se necesita descargar 752 kB de archivos.
Se utilizarán 6.050 kB de espacio de disco adicional después de esta operación.
¿Desea continuar? [S/n] S
Des:1 http://es.archive.ubuntu.com/ubuntu jammy-updates/main amd64 openssh-sftp-server amd64 1:8.9p1-3ubuntu0.7 [38,9 kB]
Des:2 http://es.archive.ubuntu.com/ubuntu jammy-updates/main amd64 openssh-server amd64 1:8.9p1-3ubuntu0.7 [435 kB]
Des:3 http://es.archive.ubuntu.com/ubuntu jammy-updates/main amd64 ncurses-term all 6.3-2ubuntu0.1 [267 kB]
Des:4 http://es.archive.ubuntu.com/ubuntu jammy/main amd64 ssh-import-id all 5.11-0ubuntu1 [10,1 kB]
Descargados 752 kB en 1s (663 kB/s)
Preconfigurando paquetes ...
Seleccionando el paquete openssh-sftp-server previamente no seleccionado.
(Leyendo la base de datos ... 205439 ficheros o directorios instalados actualmente.)
Preparando para desempaquetar .../openssh-sftp-server_1%3a8.9p1-3ubuntu0.7_amd64.deb ...
Desempaquetando openssh-sftp-server (1:8.9p1-3ubuntu0.7) ...
Seleccionando el paquete openssh-server previamente no seleccionado.
Preparando para desempaquetar .../openssh-server_1%3a8.9p1-3ubuntu0.7_amd64.deb ...
Desempaquetando openssh-server (1:8.9p1-3ubuntu0.7) ...
Seleccionando el paquete ncurses-term previamente no seleccionado.
Preparando para desempaquetar .../ncurses-term_6.3-2ubuntu0.1_all.deb ...
Desempaquetando ncurses-term (6.3-2ubuntu0.1) ...
Seleccionando el paquete ssh-import-id previamente no seleccionado.
Preparando para desempaquetar .../ssh-import-id_5.11-0ubuntu1_all.deb ...
Desempaquetando ssh-import-id (5.11-0ubuntu1) ...
Configurando openssh-sftp-server (1:8.9p1-3ubuntu0.7) ...
Configurando openssh-server (1:8.9p1-3ubuntu0.7) ...

Creating config file /etc/ssh/sshd_config with new version
Creating SSH2 RSA key; this may take some time ...
3072 SHA256:P7YyxLKWX1QqF2MRWawQsq00XfivUMGj12ouEpZDHac root@cristobal-VirtualBox (RSA)
Creating SSH2 ECDSA key; this may take some time ...
256 SHA256:WwUktxBvWH9EqZeOPjwv2rEKdJqJi/bhwil/JUAGI root@cristobal-VirtualBox (ECDSA)
Creating SSH2 ED25519 key; this may take some time ...
```

Tras conectar por ssh instalar algunas dependencias necesarias:

```
cristobal@cristobal-VirtualBox: ~  
cristobal@cristobal-VirtualBox:~$ sudo apt install apt-transport-https ca-certificates curl software-properties-common -y  
[sudo] contraseña para cristobal:  
Leyendo lista de paquetes... Hecho  
Creando árbol de dependencias... Hecho  
Leyendo la información de estado... Hecho  
ca-certificates ya está en su versión más reciente (20230311ubuntu0.22.04.1).  
fijado ca-certificates como instalado manualmente.  
software-properties-common ya está en su versión más reciente (0.99.22.9).  
fijado software-properties-common como instalado manualmente.  
Los paquetes indicados a continuación se instalaron de forma automática y ya no son necesarios.  
  libwpe-1.0-1 libwpebackend-fdo-1.0-1  
Utilice «sudo apt autoremove» para eliminarlos.  
Se instalarán los siguientes paquetes NUEVOS:  
  apt-transport-https curl  
0 actualizados, 2 nuevos se instalarán, 0 para eliminar y 3 no actualizados.  
Se necesita descargar 196 kB de archivos.  
Se utilizarán 624 kB de espacio de disco adicional después de esta operación.  
Des:1 http://es.archive.ubuntu.com/ubuntu jammy-updates/universe amd64 apt-transport-https all 2.4.12 [1,510 B]  
Des:2 http://es.archive.ubuntu.com/ubuntu jammy-updates/main amd64 curl amd64 7.81.0-1ubuntu1.16 [194 kB]  
Descargados 196 kB en 1s (236 kB/s)  
Seleccionando el paquete apt-transport-https previamente no seleccionado.  
(Leyendo la base de datos ... 208364 ficheros o directorios instalados actualmente.)  
Preparando para desempaquetar .../apt-transport-https_2.4.12_all.deb ...  
Desempaquetando apt-transport-https (2.4.12) ...  
Seleccionando el paquete curl previamente no seleccionado.  
Preparando para desempaquetar .../curl_7.81.0-1ubuntu1.16_amd64.deb ...  
Desempaquetando curl (7.81.0-1ubuntu1.16) ...  
Configurando apt-transport-https (2.4.12) ...  
Configurando curl (7.81.0-1ubuntu1.16) ...  
Procesando disparadores para man-db (2.10.2-1) ...  
cristobal@cristobal-VirtualBox:~$
```

Añadir la GPG key para poder descargar docker desde el repositorio local

```
root@cristobal-VirtualBox: /home/cristobal  
cristobal@cristobal-VirtualBox:~$ sudo su  
root@cristobal-VirtualBox:/home/cristobal# curl -fsSL https://download.docker.com/linux/ubuntu/gpg | apt-key add -  
Warning: apt-key is deprecated. Manage keyring files in trusted.gpg.d instead (see apt-key(8)).  
OK  
root@cristobal-VirtualBox:/home/cristobal#
```

Añadir el repositorio de Docker en la lista de repositorios

```
root@cristobal-VirtualBox: /home/cristobal  
root@cristobal-VirtualBox:/home/cristobal# add-apt-repository "deb [arch=amd64] https://download.docker.com/linux/ubuntu focal stable"  
Repositorio: «deb [arch=amd64] https://download.docker.com/linux/ubuntu focal stable»  
Descripción:  
Archive for codename: focal components: stable  
Más información: https://download.docker.com/linux/ubuntu  
Añadiendo repositorio.  
Oprima [INTRO] para continuar o Ctrl+c para cancelar.  
Adding deb entry to /etc/apt/sources.list.d/archive_uri-https_download_docker_com_linux_ubuntu-jammy.list  
Adding disabled deb-src entry to /etc/apt/sources.list.d/archive_uri-https_download_docker_com_linux_ubuntu-jammy.list  
Des:1 https://download.docker.com/linux/ubuntu focal InRelease [57,7 kB]  
Obj:2 http://es.archive.ubuntu.com/ubuntu jammy InRelease  
Des:3 http://security.ubuntu.com/ubuntu jammy-security InRelease [110 kB]  
Des:4 http://es.archive.ubuntu.com/ubuntu jammy-updates InRelease [119 kB]  
Des:5 https://download.docker.com/linux/ubuntu focal/stable amd64 Packages [41,5 kB]  
Obj:6 http://es.archive.ubuntu.com/ubuntu jammy-backports InRelease  
Des:7 http://es.archive.ubuntu.com/ubuntu jammy-updates/main i386 Packages [616 kB]  
Des:8 http://es.archive.ubuntu.com/ubuntu jammy-updates/main amd64 Packages [1.597 kB]  
Des:9 http://es.archive.ubuntu.com/ubuntu jammy-updates/universe i386 Packages [699 kB]  
Des:10 http://es.archive.ubuntu.com/ubuntu jammy-updates/universe amd64 Packages [1.070 kB]  
Descargados 4.311 kB en 3s (1.680 kB/s)  
Leyendo lista de paquetes... Hecho  
W: https://download.docker.com/linux/ubuntu/dists/focal/InRelease: Key is stored in legacy trusted.gpg keyring (/etc/apt/trusted.gpg), see the DEPRECATION section in apt-key(8) for details.  
root@cristobal-VirtualBox:/home/cristobal#
```

Actualizar la lista de repositorios

```
cristobal@cristobal-VirtualBox: ~
root@cristobal-VirtualBox:/home/cristobal# root@cristobal-VirtualBox:/home/cristobal# exit
exit
cristobal@cristobal-VirtualBox:~$ sudo apt update
Obj:1 https://download.docker.com/linux/ubuntu focal InRelease
Obj:2 http://es.archive.ubuntu.com/ubuntu jammy InRelease
Obj:3 http://security.ubuntu.com/ubuntu jammy-security InRelease
Obj:4 http://es.archive.ubuntu.com/ubuntu jammy-updates InRelease
Obj:5 http://es.archive.ubuntu.com/ubuntu jammy-backports InRelease
Leyendo lista de paquetes... Hecho
Creando árbol de dependencias... Hecho
Leyendo la información de estado... Hecho
Se pueden actualizar 3 paquetes. Ejecute «apt list --upgradable» para verlos.
W: https://download.docker.com/linux/ubuntu/dists/focal/InRelease: Key is stored in legacy trusted.gpg keyring (/etc/apt/t
trusted.gpg), see the DEPRECATION section in apt-key(8) for details.
cristobal@cristobal-VirtualBox:~$
```

Algunas veces hay errores visuales que hacen que se muestre el texto del command prompt por duplicado, sólo es un error estético y no afecta al funcionamiento de la máquina.

Instalar Docker y Docker compose

```
cristobal@cristobal-VirtualBox: ~
cristobal@cristobal-VirtualBox:~$ sudo apt install docker-ce docker-compose -y
Leyendo lista de paquetes... Hecho
Creando árbol de dependencias... Hecho
Leyendo la información de estado... Hecho
Los paquetes indicados a continuación se instalaron de forma automática y ya no son necesarios.
  libwpe-1.0-1 libwpebackend-fdo-1.0-1
Utilice «sudo apt autoremove» para eliminarlos.
Se instalarán los siguientes paquetes adicionales:
  containerd.io docker-buildx-plugin docker-ce-cli docker-ce-rootless-extras docker-compose-plugin git git-man
  liberror-perl libslirp0 pigz python3-attr python3-distutils python3-docker python3-dockerpty python3-dockerpt
  python3-dotenv python3-jjsonschema python3-pyrsistent python3-setuptools python3-texttable python3-websocket
  slirp4netns
Paquetes sugeridos:
  aufs-tools cgroupfs-mount | cgroup-lite git-daemon-run | git-daemon-sysvinit git-doc git-email git-gui gitk gitweb
  git-cvs git-mediawiki git-svn python-attr-doc python-jjsonschema-doc python-setuptools-doc
Paquetes recomendados:
  docker.io
Se instalarán los siguientes paquetes NUEVOS:
  containerd.io docker-buildx-plugin docker-ce docker-ce-cli docker-ce-rootless-extras docker-compose
  docker-compose-plugin git git-man liberror-perl libslirp0 pigz python3-attr python3-distutils python3-docker
  python3-dockerpty python3-dockerpt python3-dotenv python3-jjsonschema python3-pyrsistent python3-setuptools
  python3-texttable python3-websocket slirp4netns
0 actualizados, 24 nuevos se instalarán, 0 para eliminar y 3 no actualizados.
Se necesita descargar 125 MB de archivos.
Se utilizarán 456 MB de espacio de disco adicional después de esta operación.
Des:1 https://download.docker.com/linux/ubuntu focal/stable amd64 containerd.io amd64 1.6.31-1 [29,8 MB]
Des:2 http://es.archive.ubuntu.com/ubuntu jammy/universe amd64 pigz amd64 2.6-1 [63,6 kB]
Des:3 http://es.archive.ubuntu.com/ubuntu jammy-updates/main amd64 python3-distutils all 3.10.8-1~22.04 [139 kB]
Des:4 http://es.archive.ubuntu.com/ubuntu jammy/universe amd64 python3-websocket all 1.2.3-1 [34,7 kB]
Des:5 http://es.archive.ubuntu.com/ubuntu jammy/universe amd64 python3-docker all 5.0.3-1 [89,3 kB]
Des:6 http://es.archive.ubuntu.com/ubuntu jammy/universe amd64 python3-dockerpty all 0.4.1-2 [11,1 kB]
Des:7 http://es.archive.ubuntu.com/ubuntu jammy/universe amd64 python3-dockerpt all 0.6.2-4 [26,9 kB]
Des:8 http://es.archive.ubuntu.com/ubuntu jammy/universe amd64 python3-dotenv all 0.19.2-1 [20,5 kB]
Des:9 http://es.archive.ubuntu.com/ubuntu jammy/main amd64 python3-attr all 21.2.0-1 [44,0 kB]
Des:10 http://es.archive.ubuntu.com/ubuntu jammy-updates/main amd64 python3-setuptools all 59.6.0-1.2ubuntu0.22.04.1 [339
kB]
Des:11 http://es.archive.ubuntu.com/ubuntu jammy/main amd64 python3-pyrsistent amd64 0.18.1-1build1 [55,5 kB]
Des:12 http://es.archive.ubuntu.com/ubuntu jammy/main amd64 python3-jjsonschema all 3.2.0-0ubuntu2 [43,1 kB]
Des:13 http://es.archive.ubuntu.com/ubuntu jammy/universe amd64 python3-texttable all 1.6.4-1 [11,4 kB]
Des:14 http://es.archive.ubuntu.com/ubuntu jammy/universe amd64 docker-compose all 1.29.2-1 [95,8 kB]
Des:15 http://es.archive.ubuntu.com/ubuntu jammy/main amd64 liberror-perl all 0.17029-1 [26,5 kB]
Des:16 http://es.archive.ubuntu.com/ubuntu jammy-updates/main amd64 git-man all 1:2.34.1-1ubuntu1.10 [954 kB]
Des:17 https://download.docker.com/linux/ubuntu focal/stable amd64 docker-buildx-plugin amd64 0.14.0-1~ubuntu.20.04~focal
[29,7 MB]
Des:18 http://es.archive.ubuntu.com/ubuntu jammy-updates/main amd64 git amd64 1:2.34.1-1ubuntu1.10 [3.166 kB]
Des:19 http://es.archive.ubuntu.com/ubuntu jammy/main amd64 libslirp0 amd64 4.6.1-1build1 [61,5 kB]
Des:20 http://es.archive.ubuntu.com/ubuntu jammy/universe amd64 slirp4netns amd64 1.0.1-2 [28,2 kB]
Des:21 https://download.docker.com/linux/ubuntu focal/stable amd64 docker-ce-cli amd64 5:26.0.2-1~ubuntu.20.04~focal
[13,8 MB]
Des:22 https://download.docker.com/linux/ubuntu focal/stable amd64 docker-ce amd64 5:26.0.2-1~ubuntu.20.04~focal [25,1 MB]
Des:23 https://download.docker.com/linux/ubuntu focal/stable amd64 docker-ce-rootless-extras amd64 5:26.0.2-1~ubuntu.20.04
~focal [9.329 kB]
Des:24 https://download.docker.com/linux/ubuntu focal/stable amd64 docker-compose-plugin amd64 2.26.1-1~ubuntu.20.04~focal
[12,4 MB]
Descargados 125 MB en 4s (29,0 MB/s)
Seleccionando el paquete pigz previamente no seleccionado.
(Leyendo la base de datos ... 208375 ficheros o directorios instalados actualmente.)
Preparando para desempaquetar .../00-pigz_2.6-1_amd64.deb ...
Desempaquetando pigz (2.6-1) ...
Seleccionando el paquete containerd.io previamente no seleccionado.
Preparando para desempaquetar .../01-containerd.io_1.6.31-1_amd64.deb ...
Desempaquetando containerd.io (1.6.31-1) ...
Seleccionando el paquete docker-buildx-plugin previamente no seleccionado.
```

Activar el servicio

```
cristobal@cristobal-VirtualBox: ~
cristobal@cristobal-VirtualBox:~$ systemctl start docker
==== AUTHENTICATING FOR org.freedesktop.systemd1.manage-units ====
Necesita autenticarse para iniciar «docker.service».
Authenticating as: cristobal,, (cristobal)
Password:
==== AUTHENTICATION COMPLETE ====
cristobal@cristobal-VirtualBox:~$
```

Comprobar el estado del servicio

```
cristobal@cristobal-VirtualBox: ~
cristobal@cristobal-VirtualBox:~$ systemctl status docker
● docker.service - Docker Application Container Engine
   Loaded: loaded (/lib/systemd/system/docker.service; enabled; vendor preset: enabled)
   Active: active (running) since Sun 2024-04-21 11:00:36 CEST; 11min ago
 TriggeredBy: ● docker.socket
    Docs: https://docs.docker.com
   Main PID: 5344 (dockerd)
     Tasks: 10
    Memory: 29.1M
       CPU: 1.482s
    CGroup: /system.slice/docker.service
           └─5344 /usr/bin/dockerd -H fd:// --containerd=/run/containerd/containerd.sock

abr 21 11:00:32 cristobal-VirtualBox systemd[1]: Starting Docker Application Container Engine...
abr 21 11:00:32 cristobal-VirtualBox dockerd[5344]: time="2024-04-21T11:00:32.809181689+02:00" level=info msg="Starting up"
abr 21 11:00:32 cristobal-VirtualBox dockerd[5344]: time="2024-04-21T11:00:32.827538696+02:00" level=info msg="detected 1"
abr 21 11:00:35 cristobal-VirtualBox dockerd[5344]: time="2024-04-21T11:00:35.360869700+02:00" level=info msg="Loading co"
abr 21 11:00:36 cristobal-VirtualBox dockerd[5344]: time="2024-04-21T11:00:36.170106388+02:00" level=info msg="Loading co"
abr 21 11:00:36 cristobal-VirtualBox dockerd[5344]: time="2024-04-21T11:00:36.356753136+02:00" level=info msg="Docker dae"
abr 21 11:00:36 cristobal-VirtualBox dockerd[5344]: time="2024-04-21T11:00:36.357059112+02:00" level=info msg="Daemon has"
abr 21 11:00:36 cristobal-VirtualBox dockerd[5344]: time="2024-04-21T11:00:36.530248767+02:00" level=info msg="API listen"
abr 21 11:00:36 cristobal-VirtualBox systemd[1]: Started Docker Application Container Engine.
lines 1-21/21 (END)
^C
cristobal@cristobal-VirtualBox:~$
```

Ejecutar el “hola mundo” para comprobar que funciona

```
cristobal@cristobal-VirtualBox: ~
cristobal@cristobal-VirtualBox:~$ sudo docker run hello-world
[sudo] contraseña para cristobal:
Unable to find image 'hello-world:latest' locally
latest: Pulling from library/hello-world
c1ec31eb5944: Pull complete
Digest: sha256:a26bff933ddc26d5cdf7faa98b4ae1e3ec20c4985e6f87ac0973052224d24302
Status: Downloaded newer image for hello-world:latest

Hello from Docker!
This message shows that your installation appears to be working correctly.

To generate this message, Docker took the following steps:
 1. The Docker client contacted the Docker daemon.
 2. The Docker daemon pulled the "hello-world" image from the Docker Hub.
    (amd64)
 3. The Docker daemon created a new container from that image which runs the
    executable that produces the output you are currently reading.
 4. The Docker daemon streamed that output to the Docker client, which sent it
    to your terminal.

To try something more ambitious, you can run an Ubuntu container with:
$ docker run -it ubuntu bash

Share images, automate workflows, and more with a free Docker ID:
https://hub.docker.com/

For more examples and ideas, visit:
https://docs.docker.com/get-started/
cristobal@cristobal-VirtualBox:~$
```

Crear la carpeta cluster y descargar dentro las 3 imágenes

spark-master.Dockerfile:

<https://raw.githubusercontent.com/tnavarrete-iedib/bigdata-23-24/main/docker/spark-master.Dockerfile>

```
cristobal@cristobal-VirtualBox: ~/cluster
cristobal@cristobal-VirtualBox:~/cluster$ wget https://raw.githubusercontent.com/tnavarrete-iedib/bigdata-23-24/main/docker/spark-master.Dockerfile
--2024-04-21 11:22:40-- https://raw.githubusercontent.com/tnavarrete-iedib/bigdata-23-24/main/docker/spark-master.Dockerfile
Resolviendo raw.githubusercontent.com (raw.githubusercontent.com)... 185.199.108.133, 185.199.109.133, 185.199.111.133, ..
Conectando con raw.githubusercontent.com (raw.githubusercontent.com)[185.199.108.133]:443... conectado.
Petición HTTP enviada, esperando respuesta... 200 OK
Longitud: 1957 (1,9K) [text/plain]
Guardando como: 'spark-master.Dockerfile'

spark-master.Dockerfile      100%[=====>] 1,91K  --.-KB/s  en 0s
2024-04-21 11:22:40 (34,7 MB/s) - 'spark-master.Dockerfile' guardado [1957/1957]
```

spark-worker.Dockerfile:

<https://raw.githubusercontent.com/tnavarrete-iedib/bigdata-23-24/main/docker/spark-worker.Dockerfile>

```
cristobal@cristobal-VirtualBox: ~/cluster
cristobal@cristobal-VirtualBox:~/cluster$ wget https://raw.githubusercontent.com/tnavarrete-iedib/bigdata-23-24/main/docker/spark-worker.Dockerfile
--2024-04-21 11:25:35-- https://raw.githubusercontent.com/tnavarrete-iedib/bigdata-23-24/main/docker/spark-worker.Dockerfile
Resolviendo raw.githubusercontent.com (raw.githubusercontent.com)... 185.199.108.133, 185.199.109.133, 185.199.110.133, ..
Conectando con raw.githubusercontent.com (raw.githubusercontent.com)[185.199.108.133]:443... conectado.
Petición HTTP enviada, esperando respuesta... 200 OK
Longitud: 1901 (1,9K) [text/plain]
Guardando como: 'spark-worker.Dockerfile'

spark-worker.Dockerfile      100%[=====>] 1,86K  --.-KB/s  en 0s
2024-04-21 11:25:35 (22,0 MB/s) - 'spark-worker.Dockerfile' guardado [1901/1901]
```

jupyterlab.Dockerfile:

<https://raw.githubusercontent.com/tnavarrete-iedib/bigdata-23-24/main/docker/jupyterlab.Dockerfile>

```
cristobal@cristobal-VirtualBox: ~/cluster
cristobal@cristobal-VirtualBox:~/cluster$ wget https://raw.githubusercontent.com/tnavarrete-iedib/bigdata-23-24/main/docker/jupyterlab.Dockerfile
--2024-04-21 11:27:22-- https://raw.githubusercontent.com/tnavarrete-iedib/bigdata-23-24/main/docker/jupyterlab.Dockerfile
Resolviendo raw.githubusercontent.com (raw.githubusercontent.com)... 185.199.108.133, 185.199.110.133, 185.199.109.133, ..
Conectando con raw.githubusercontent.com (raw.githubusercontent.com)[185.199.108.133]:443... conectado.
Petición HTTP enviada, esperando respuesta... 200 OK
Longitud: 1550 (1,5K) [text/plain]
Guardando como: 'jupyterlab.Dockerfile'

jupyterlab.Dockerfile        100%[=====>] 1,51K  --.-KB/s  en 0s
2024-04-21 11:27:23 (33,3 MB/s) - 'jupyterlab.Dockerfile' guardado [1550/1550]
```


Descargar en la carpeta cluster el script bash que montará las imágenes:

<https://raw.githubusercontent.com/tnavarrete-iedib/bigdata-23-24/main/docker/build.sh>

```
cristobal@cristobal-VirtualBox: ~/cluster
cristobal@cristobal-VirtualBox:~/cluster$ wget https://raw.githubusercontent.com/tnavarrete-iedib/bigdata-23-24/main/docker/build.sh
--2024-04-21 18:35:00-- https://raw.githubusercontent.com/tnavarrete-iedib/bigdata-23-24/main/docker/build.sh
Resolviendo raw.githubusercontent.com (raw.githubusercontent.com)... 185.199.108.133, 185.199.109.133, 185.199.110.133, ...
Conectando con raw.githubusercontent.com (raw.githubusercontent.com)[185.199.108.133]:443... conectado.
Petición HTTP enviada, esperando respuesta... 200 OK
Longitud: 626 [text/plain]
Guardando como: 'build.sh'

build.sh          100%[=====>]          626 --.-KB/s   en 0s

2024-04-21 18:35:00 (40,2 MB/s) - 'build.sh' guardado [626/626]

cristobal@cristobal-VirtualBox:~/cluster$
```

Ejecutar el script bash descargado

```
cristobal@cristobal-VirtualBox: ~/cluster
cristobal@cristobal-VirtualBox:~/cluster$ sudo bash build.sh
[sudo] contraseña para cristobal:
[+] Building 239.1s (9/9) FINISHED                                docker:default
=> [internal] load build definition from spark-master.Dockerfile                                0.2s
=> => transferring dockerfile: 2.01kB                                                            0.0s
=> [internal] load metadata for docker.io/library/openjdk:8-jre-slim                          2.3s
=> [internal] load .dockerignore                                                                0.1s
=> => transferring context: 2B                                                                    0.0s
=> [1/5] FROM docker.io/library/openjdk:8-jre-slim@sha256:53186129237fbb8bc0a12dd36da6761f4c7a2a20233c20d4eb 17.1s
=> => resolve docker.io/library/openjdk:8-jre-slim@sha256:53186129237fbb8bc0a12dd36da6761f4c7a2a20233c20d4eb0 0.1s
=> => sha256:53186129237fbb8bc0a12dd36da6761f4c7a2a20233c20d4eb0d497e4045a4f5 549B / 549B          0.0s
=> => sha256:285c61a1e5e6b7b3709729b69558670148c5fddc6eb7104fae7dd370042c51430 1.16kB / 1.16kB 0.0s
=> => sha256:85b121affeddccffc7bc6618140bb0285ad1257bd318676ddc67816863c0686c0 7.47kB / 7.47kB 0.0s
=> => sha256:1efc276f4ff952c055dea726cfc96ec6a4fdb8b62d9eed816bd2b788f2860ad7 31.37MB / 31.37MB 3.4s
=> => sha256:a2f2f93da48276873890ac821b3c991d53a7e864791aaf82c39b7863c908b93b 1.58MB / 1.58MB 1.1s
=> => sha256:1a2de4cc94315f2ba5015e6781672aa8e0b1456a4d488694bb5f016d8f59fa70 210B / 210B   0.7s
=> => sha256:d2421c7a4bbfc037d7fb487893cc5fe145e329dfb39b5ee6557016bf6c34072d 41.70MB / 41.70MB 6.7s
=> => extracting sha256:1efc276f4ff952c055dea726cfc96ec6a4fdb8b62d9eed816bd2b788f2860ad7 8.5s
=> => extracting sha256:a2f2f93da48276873890ac821b3c991d53a7e864791aaf82c39b7863c908b93b 0.4s
=> => extracting sha256:1a2de4cc94315f2ba5015e6781672aa8e0b1456a4d488694bb5f016d8f59fa70 0.0s
=> => extracting sha256:d2421c7a4bbfc037d7fb487893cc5fe145e329dfb39b5ee6557016bf6c34072d 3.6s
=> [2/5] RUN mkdir -p /opt/workspace && apt-get update -y && apt-get install -y python3 && ln -s 13.8s
=> [3/5] RUN apt-get update -y && apt-get install -y python3-pip && pip3 install gdown numpy matplo 107.3s
=> [4/5] RUN apt-get update -y && apt-get install -y curl && curl https://archive.apache.org/dist/sp 75.0s
=> [5/5] WORKDIR /usr/bin/spark-3.5.1-bin-hadoop3                                              0.4s
=> exporting to image                                                                           22.3s
=> => exporting layers                                                                            22.1s
=> => writing image sha256:f54305cac10d8586ec628590b4456314809b93fa7da4c4fbc5e44041dc0504580 0.0s
=> => naming to docker.io/library/spark-master                                                  0.1s
[+] Building 223.0s (9/9) FINISHED                                docker:default
=> [internal] load build definition from spark-worker.Dockerfile                                0.0s
=> => transferring dockerfile: 1.95kB                                                            0.0s
=> [internal] load metadata for docker.io/library/openjdk:8-jre-slim                          0.7s
=> [internal] load .dockerignore                                                                0.2s
=> => transferring context: 2B                                                                    0.0s
=> CACHED [1/5] FROM docker.io/library/openjdk:8-jre-slim@sha256:53186129237fbb8bc0a12dd36da6761f4c7a2a20233c 0.0s
=> [2/5] RUN mkdir -p /opt/workspace && apt-get update -y && apt-get install -y python3 && ln -s 11.7s
=> [3/5] RUN apt-get update -y && apt-get install -y python3-pip && pip3 install gdown numpy matplo 104.6s
=> [4/5] RUN apt-get update -y && apt-get install -y curl && curl https://archive.apache.org/dist/sp 73.3s
=> [5/5] WORKDIR /usr/bin/spark-3.5.1-bin-hadoop3                                              0.4s
=> exporting to image                                                                           31.4s
=> => exporting layers                                                                            31.2s
=> => writing image sha256:e6b44602629542b2cc31f5641074a7c5870896dc9e4bc35666a1d82b849f24f9 0.0s
=> => naming to docker.io/library/spark-worker                                                  0.0s
[+] Building 244.6s (9/9) FINISHED                                docker:default
=> [internal] load build definition from jupyterlab.Dockerfile                                0.1s
=> => transferring dockerfile: 1.60kB                                                            0.0s
=> [internal] load metadata for docker.io/library/openjdk:8-jre-slim                          1.2s
=> [internal] load .dockerignore                                                                0.2s
=> => transferring context: 2B                                                                    0.0s
=> CACHED [1/5] FROM docker.io/library/openjdk:8-jre-slim@sha256:53186129237fbb8bc0a12dd36da6761f4c7a2a20233c 0.0s
=> [2/5] RUN mkdir -p /opt/workspace && apt-get update -y && apt-get install -y python3 && ln -s 14.9s
=> [3/5] RUN apt-get update -y && apt-get install -y python3-pip && pip3 install gdown numpy matplo 99.2s
=> [4/5] RUN pip3 install wget pyspark==3.5.1 jupyterlab==4.1.5                               95.0s
=> [5/5] WORKDIR /opt/workspace                                                                0.2s
=> exporting to image                                                                           33.4s
=> => exporting layers                                                                            33.2s
=> => writing image sha256:af204e2c16c75648f4cfaae63c75bcb7fb422b3a16ef69d319cb43dab73089c2 0.0s
=> => naming to docker.io/library/jupyterlab                                                  0.0s
cristobal@cristobal-VirtualBox:~/cluster$
```

Comprobar que se han montado correctamente

```
cristobal@cristobal-VirtualBox: ~/cluster
cristobal@cristobal-VirtualBox:~/cluster$ sudo docker images
REPOSITORY          TAG             IMAGE ID        CREATED         SIZE
jupyterlab           latest          af204e2c16c7    2 minutes ago   2.18GB
spark-worker         latest          e6b446026295    6 minutes ago   1.46GB
spark-master         latest          f54305cac10d    10 minutes ago  1.46GB
hello-world          latest          d2c94e258dcb    11 months ago   13.3kB
cristobal@cristobal-VirtualBox:~/cluster$
```

Descargar el archivo de configuración docker-compose.yml:

<https://raw.githubusercontent.com/tnavarrete-iedib/bigdata-23-24/main/docker/docker-compose.yml>

```
cristobal@cristobal-VirtualBox: ~/cluster
cristobal@cristobal-VirtualBox:~/cluster$ wget https://raw.githubusercontent.com/tnavarrete-iedib/bigdata-23-24/main/docker/docker-compose.yml
--2024-04-21 18:54:50-- https://raw.githubusercontent.com/tnavarrete-iedib/bigdata-23-24/main/docker/docker-compose.yml
Resolviendo raw.githubusercontent.com (raw.githubusercontent.com)... 185.199.108.133, 185.199.110.133, 185.199.111.133, ...
Conectando con raw.githubusercontent.com (raw.githubusercontent.com)[185.199.108.133]:443... conectado.
Petición HTTP enviada, esperando respuesta... 200 OK
Longitud: 819 [text/plain]
Guardando como: 'docker-compose.yml'

docker-compose.yml          100%[=====>]          819  --.-KB/s    en 0s

2024-04-21 18:54:50 (40,9 MB/s) - 'docker-compose.yml' guardado [819/819]
cristobal@cristobal-VirtualBox:~/cluster$
```

Añadir spark-worker-3 exponiendo el puerto 8083 para su interfaz web

```
cristobal@cristobal-VirtualBox: ~/cluster
GNU nano 6.2 docker-compose.yml
version: "3.8"

volumes:
  hdfs-simulado:
    name: "hadoop-distributed-file-system"
    driver: local

services:
  spark-master:
    image: spark-master
    container_name: spark-master
    ports:
      - 7077:7077
      - 8080:8080
    volumes:
      - hdfs-simulado:/opt/workspace

  spark-worker-1:
    image: spark-worker
    container_name: spark-worker-1
    ports:
      - 8081:8081
    volumes:
      - hdfs-simulado:/opt/workspace
    depends_on:
      - spark-master

  spark-worker-2:
    image: spark-worker
    container_name: spark-worker-2
    ports:
      - 8082:8081
    volumes:
      - hdfs-simulado:/opt/workspace
    depends_on:
      - spark-master

  spark-worker-3:
    image: spark-worker
    container_name: spark-worker-3
    ports:
      - 8083:8081
    volumes:
      - hdfs-simulado:/opt/workspace
    depends_on:
      - spark-master

  jupyterlab:
    image: jupyterlab
    container_name: jupyterlab
    ports:
      - 8888:8888
    volumes:
      - hdfs-simulado:/opt/workspace
```

[54 líneas escritas]

^G Ayuda	^O Guardar	^W Buscar	^K Cortar	^T Ejecutar	^C Ubicación	M-U Deshacer
^X Salir	^R Leer fich.	^_ Reemplazar	^U Pegar	^J Justificar	^_ In a línea	M-E Rehacer

Ejecutar los contenedores desde la carpeta cluster

```
crisobal@crisobal-VirtualBox: ~/cluster
crisobal@crisobal-VirtualBox:~/cluster$ sudo docker-compose up
Creating network "cluster_default" with the default driver
Creating volume "hadoop-distributed-file-system" with local driver
Creating jupyterlab ... done
Creating spark-master ... done
Creating spark-worker-3 ... done
Creating spark-worker-1 ... done
Creating spark-worker-2 ... done
Attaching to jupyterlab, spark-master, spark-worker-2, spark-worker-3, spark-worker-1
jupyterlab | [I 2024-04-21 18:10:22.853 ServerApp] jupyter_lsp | extension was successfully linked.
jupyterlab | [I 2024-04-21 18:10:22.884 ServerApp] jupyter_server_terminals | extension was successfully linked.
jupyterlab | [W 2024-04-21 18:10:22.887 LabApp] 'token' has moved from NotebookApp to ServerApp. This config will be passed to ServerApp. Be sure to update your config before our next release.
jupyterlab | [W 2024-04-21 18:10:22.893 ServerApp] ServerApp.token config is deprecated in 2.0. Use IdentityProvider.token.
jupyterlab | [I 2024-04-21 18:10:22.893 ServerApp] jupyterlab | extension was successfully linked.
jupyterlab | [I 2024-04-21 18:10:22.896 ServerApp] Writing Jupyter server cookie secret to /root/.local/share/jupyter/runtime/jupyter_cookie_secret
jupyterlab | [I 2024-04-21 18:10:23.307 ServerApp] notebook_shim | extension was successfully linked.
jupyterlab | [W 2024-04-21 18:10:23.338 ServerApp] All authentication is disabled. Anyone who can connect to this server will be able to run code.
jupyterlab | [I 2024-04-21 18:10:23.340 ServerApp] notebook_shim | extension was successfully loaded.
jupyterlab | [I 2024-04-21 18:10:23.351 ServerApp] jupyter_lsp | extension was successfully loaded.
jupyterlab | [I 2024-04-21 18:10:23.357 ServerApp] jupyter_server_terminals | extension was successfully loaded.
jupyterlab | [I 2024-04-21 18:10:23.363 LabApp] JupyterLab extension loaded from /usr/local/lib/python3.9/dist-packages/jupyterlab
jupyterlab | [I 2024-04-21 18:10:23.363 LabApp] JupyterLab application directory is /usr/local/share/jupyter/lab
jupyterlab | [I 2024-04-21 18:10:23.365 LabApp] Extension Manager is 'pypi'.
jupyterlab | [I 2024-04-21 18:10:23.399 ServerApp] jupyterlab | extension was successfully loaded.
jupyterlab | [I 2024-04-21 18:10:23.400 ServerApp] Serving notebooks from local directory: /opt/workspace
jupyterlab | [I 2024-04-21 18:10:23.400 ServerApp] Jupyter Server 2.14.0 is running at:
jupyterlab | [I 2024-04-21 18:10:23.400 ServerApp] http://2a2df8d4ae53:8888/lab
jupyterlab | [I 2024-04-21 18:10:23.400 ServerApp] http://127.0.0.1:8888/lab
jupyterlab | [I 2024-04-21 18:10:23.400 ServerApp] Use Control-C to stop this server and shut down all kernels (twice to skip confirmation).
jupyterlab | [I 2024-04-21 18:10:23.439 ServerApp] Skipped non-installed server(s): bash-language-server, dock
erfile-language-server-nodejs, javascript-typescript-langserver, jedi-language-server, julia-language-server, pyright
, python-language-server, python-lsp-server, r-languageserver, sql-language-server, texlab, typescript-language-server
, unified-language-server, vscode-css-languageserver-bin, vscode-html-languageserver-bin, vscode-json-languageserver
-bin, yaml-language-server
spark-master | /usr/bin/spark-3.5.1-bin-hadoop3/bin/load-spark-env.sh: line 68: ps: command not found
spark-worker-1 | /usr/bin/spark-3.5.1-bin-hadoop3/bin/load-spark-env.sh: line 68: ps: command not found
spark-worker-2 | /usr/bin/spark-3.5.1-bin-hadoop3/bin/load-spark-env.sh: line 68: ps: command not found
spark-worker-3 | /usr/bin/spark-3.5.1-bin-hadoop3/bin/load-spark-env.sh: line 68: ps: command not found
spark-master | Using Spark's default log4j profile: org/apache/spark/log4j2-defaults.properties
spark-master | 24/04/21 18:10:25 INFO Master: Started daemon with process name: 7@dec5581c8dcf
spark-master | 24/04/21 18:10:25 INFO SignalUtils: Registering signal handler for TERM
spark-master | 24/04/21 18:10:25 INFO SignalUtils: Registering signal handler for HUP
spark-master | 24/04/21 18:10:25 INFO SignalUtils: Registering signal handler for INT
spark-master | 24/04/21 18:10:26 WARN NativeCodeLoader: Unable to load native-hadoop library for your platform..
. using builtin-java classes where applicable
spark-master | 24/04/21 18:10:26 INFO SecurityManager: Changing view acls to: root
spark-master | 24/04/21 18:10:26 INFO SecurityManager: Changing modify acls to: root
spark-master | 24/04/21 18:10:26 INFO SecurityManager: Changing view acls groups to:
spark-master | 24/04/21 18:10:26 INFO SecurityManager: Changing modify acls groups to:
spark-master | 24/04/21 18:10:26 INFO SecurityManager: SecurityManager: authentication disabled; ui acls disabled;
users with view permissions: root; groups with view permissions: EMPTY; users with modify permissions: root; group
s with modify permissions: EMPTY
spark-worker-2 | Using Spark's default log4j profile: org/apache/spark/log4j2-defaults.properties
spark-worker-2 | 24/04/21 18:10:26 INFO Worker: Started daemon with process name: 7@f9566155106f
```

Interfaz web del nodo maestro

Spark [Corriendo] - Oracle VM VirtualBox

Archivo Máquina Ver Entrada Dispositivos Ayuda

Actividades

Navegador web Firefox

21 de abr 20:20

Spark Master at spark://s ×

localhost:8080

Spark Master at spark://spark-master:7077

URL: spark://spark-master:7077

Alive Workers: 3

Cores in use: 12 Total, 0 Used

Memory in use: 8.4 GiB Total, 0.0 B Used

Resources in use:

Applications: 0 Running, 0 Completed

Drivers: 0 Running, 0 Completed

Status: ALIVE

Workers (3)

Worker Id	Address	State	Cores	Memory	Resources
worker-20240421181029-172.18.0.4-33945	172.18.0.4:33945	ALIVE	4 (0 Used)	2.8 GiB (0.0 B Used)	
worker-20240421181029-172.18.0.5-46179	172.18.0.5:46179	ALIVE	4 (0 Used)	2.8 GiB (0.0 B Used)	
worker-20240421181030-172.18.0.6-36609	172.18.0.6:36609	ALIVE	4 (0 Used)	2.8 GiB (0.0 B Used)	

Running Applications (0)

Application ID	Name	Cores	Memory per Executor	Resources Per Executor	Submitted Time	User	State	Duration
----------------	------	-------	---------------------	------------------------	----------------	------	-------	----------

Completed Applications (0)

Application ID	Name	Cores	Memory per Executor	Resources Per Executor	Submitted Time	User	State	Duration
----------------	------	-------	---------------------	------------------------	----------------	------	-------	----------

CTRL DERECHA

Interfaz web de worker 1

Spark [Corriendo] - Oracle VM VirtualBox

ArchivoMáquinaVerEntradaDispositivosAyuda

ActividadesNavegador web Firefox21 de abr 20:21

Spark Worker at 172.18.0.6

localhost:8081

APACHE
spark3.5.1

Spark Worker at 172.18.0.6:36609

ID: worker-20240421181030-172.18.0.6-36609

Master URL: spark://spark-master:7077

Cores: 4 (0 Used)

Memory: 2.8 GiB (0.0 B Used)

Resources:

[Back to Master](#)

▼ Running Executors (0)

ExecutorID	State	Cores	Memory	Resources	Job Details	Logs
------------	-------	-------	--------	-----------	-------------	------

CTRL DERECHA

Interfaz web del worker 2

Spark [Corriendo] - Oracle VM VirtualBox

ArchivoMáquinaVerEntradaDispositivosAyuda

ActividadesNavegador web Firefox21 de abr 20:23

Spark Worker at 172.18.0.4

localhost:8082

APACHE
spark3.5.1

Spark Worker at 172.18.0.4:33945

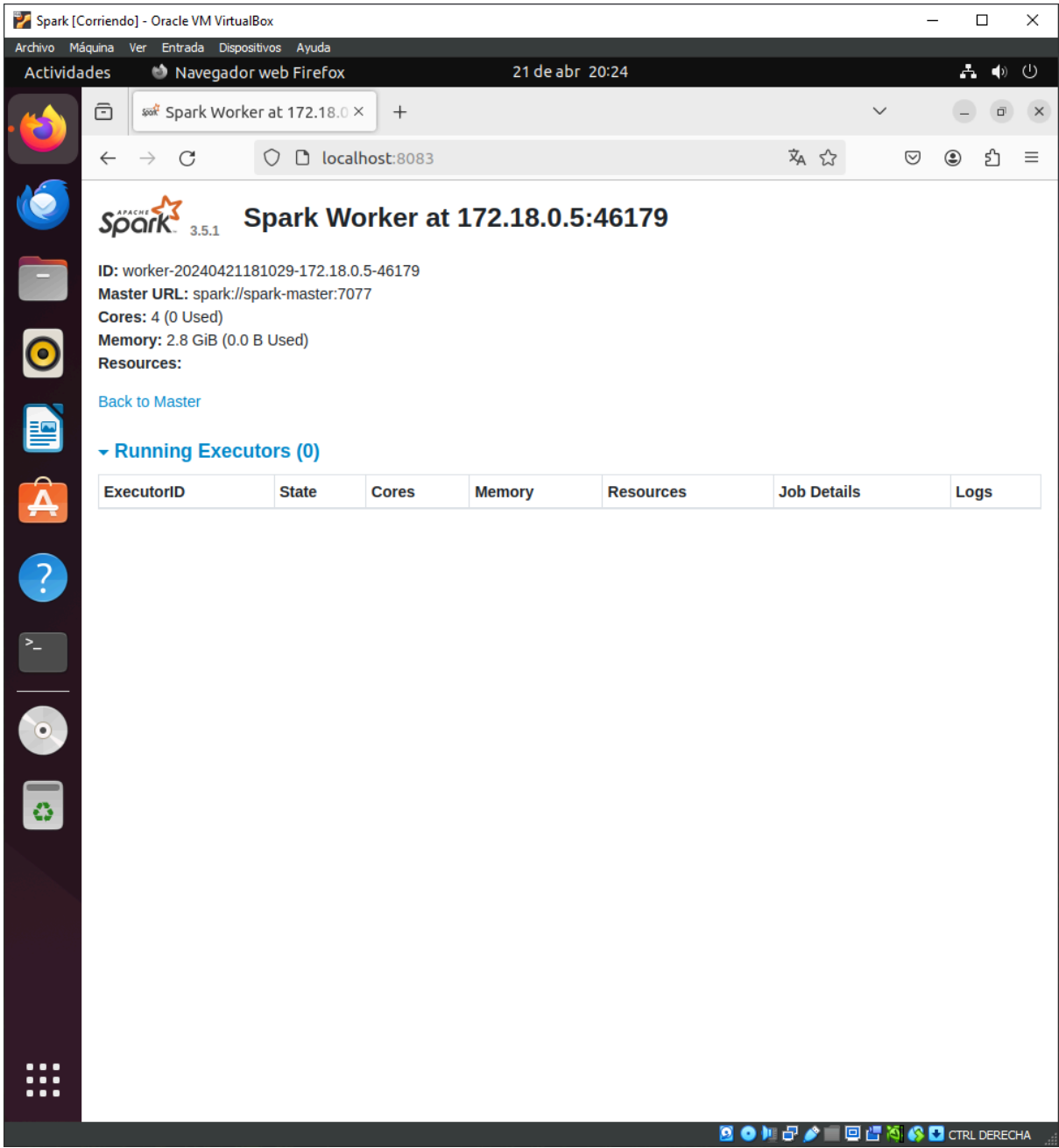
ID: worker-20240421181029-172.18.0.4-33945
Master URL: spark://spark-master:7077
Cores: 4 (0 Used)
Memory: 2.8 GiB (0.0 B Used)
Resources:
[Back to Master](#)

▼ Running Executors (0)

ExecutorID	State	Cores	Memory	Resources	Job Details	Logs
------------	-------	-------	--------	-----------	-------------	------

CTRL DERECHA

Interfaz web del worker 3 (desde el puerto 8083)



Interfaz web de JupyterLab

