



BIG DATA

2025/26

Ciclo	Especialización IA & BIG DATA
Nombre	Rodrigo Medina
Correo	YMQ06518@educastur.es
Nº Unidad Didáctica	02

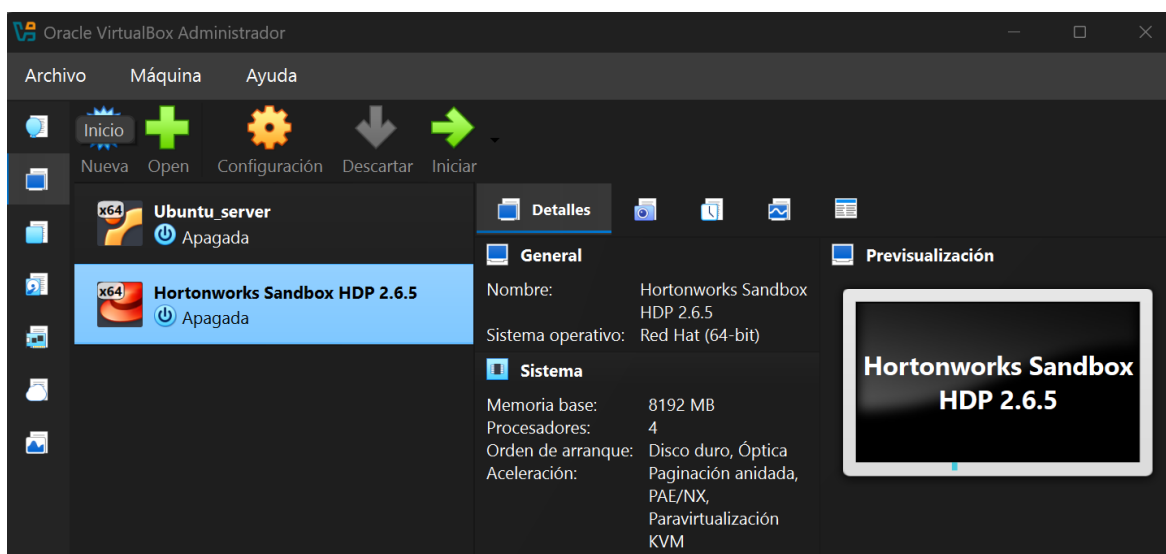
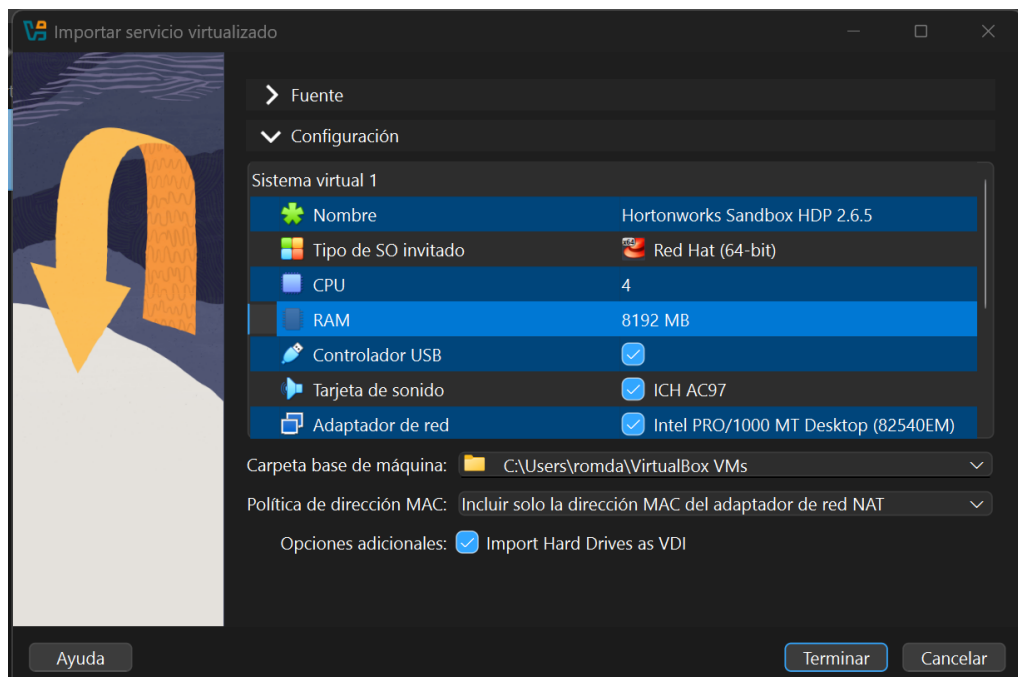
PR_02.2

1.	HORTONWORKS HDP SANDBOX.....	2
----	------------------------------	---

1. HORTONWORKS HDP SANDBOX

1. IMPORTACION .OVA EN VIRTUAL BOX

Tras descargar la imagen de la máquina virtual (.ova), ejecutamos el archivo para instalar la maquina virtual. De esta forma, nos abrirá Virtual Box con la configuración de la máquina ya preinstalada para iniciar la instalación.



BIG DATA

Tras ello, y al ejecutarla, comenzará la instalación que tardará unos minutos. Al completarse, nos aparecerá en pantalla la información sobre las IPs de localhost y sus puertos para el acceso por navegador o por ssh desde el navegador.

```
Hortonworks HDP Sandbox
https://hortonworks.com/products/sandbox

To quickly get started with the Hortonworks Sandbox, follow this tutorial:
https://hortonworks.com/tutorial/hadoop-tutorial-getting-started-with-hdp/

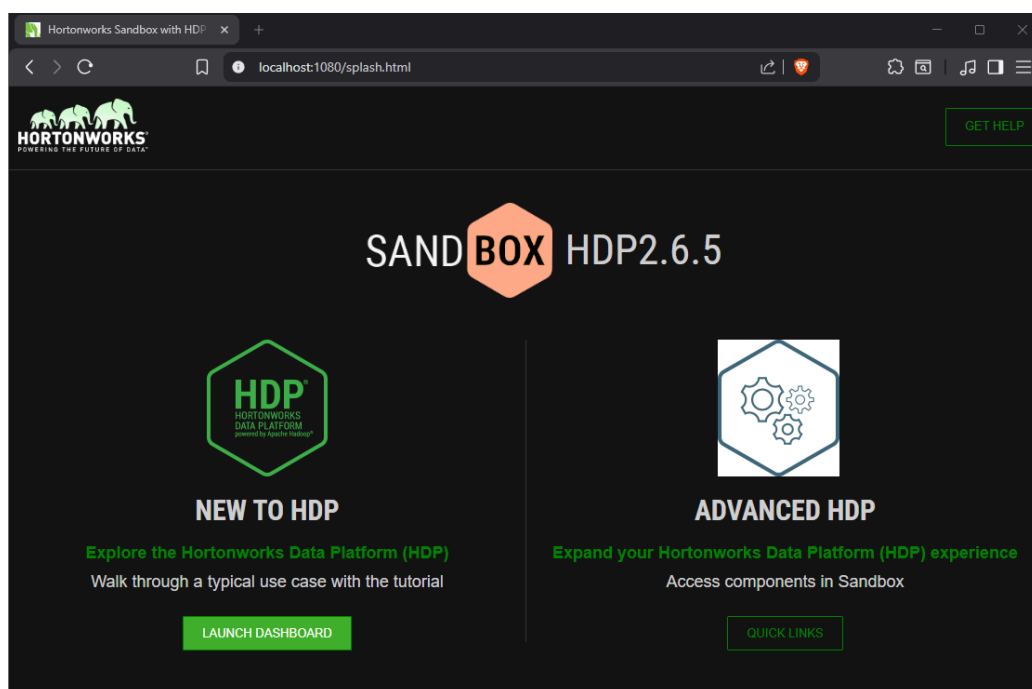
To initiate your Hortonworks Sandbox session, open a browser to this address:

For VirtualBox:
Welcome screen: http://localhost:1080
SSH: http://localhost:4200

For VMWare:
Welcome screen: http://10.0.2.15:1080
SSH: http://10.0.2.15:4200
```

En nuestro navegador, introducimos la IP correspondiente:

http://localhost:1080



2. ACCESO POR TERMINAL SSH (OPCIONAL)

En cambio, si queremos acceder por terminal, existen una serie de usuarios por defecto ya establecidos.

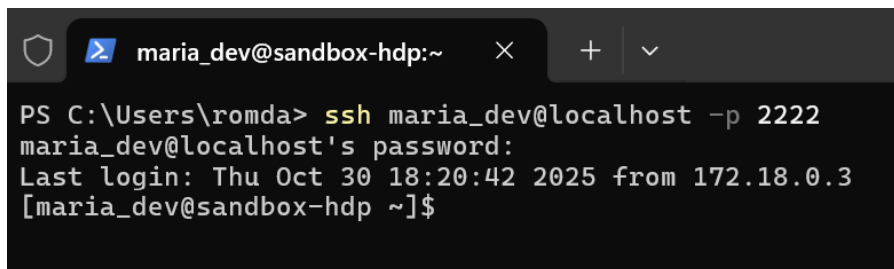
Para acceder como root, deberemos configurar su contraseña previamente.

Seleccionamos el primer usuario, maria_dev, el hostname correspondiente (localhost) y el puerto (-p) 2222. Tras ello introducimos su contraseña correspondiente.

Appendix A: Reference Sheet
Login Credentials:

User	Password
admin	refer to Admin Password Reset
maria_dev	maria_dev
raj_ops	raj_ops
holger_gov	holger_gov
amy_ds	amy_ds

```
ssh maria_dev@localhost -p 2222
```



```
PS C:\Users\romda> ssh maria_dev@localhost -p 2222
maria_dev@localhost's password:
Last login: Thu Oct 30 18:20:42 2025 from 172.18.0.3
[maria_dev@sandbox-hdp ~]$
```

3. MODIFICACIÓN HOSTS (OPCIONAL)

Para poder acceder al servicio por navegador sin tener que introducir la dirección IP, debemos modificar el archivo hosts. (*Ejecutamos bloc de notas como administrador*)

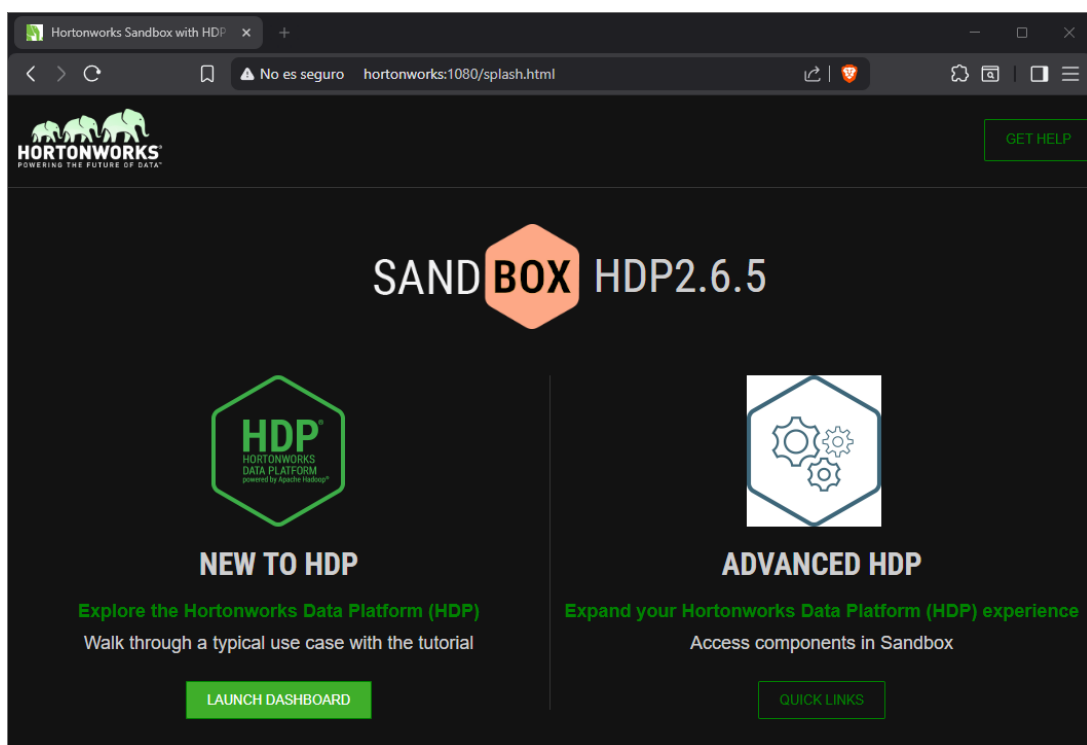
```
C:\Windows\System32\drivers\etc\hosts
```

Y añadimos al final del archivo las url que queramos a nuestro gusto para asociar a la IP.

```
127.0.0.1 localhost sandbox-hdp.hortonworks.com sandbox-  
hdf.hortonworks.com hortonworks
```

Ahora, si desde el navegador introducimos una de las anteriores url con su puerto correspondiente (:1080) podremos acceder a la máquina.

```
http://hortonworks:1080/
```



4. CAMBIO CONTRASEÑA ADMIN (AMBARI)

La primera vez que nos logueamos como root desde la terminal, nos solicita un cambio de contraseña. Por defecto las credenciales son:

Usuario: root

Contraseña: hadoop

A través de ssh sería:

```
ssh root@localhost -p 2222
```

```
PS C:\Users\romda> ssh root@localhost -p 2222
root@localhost's password:
Permission denied, please try again.
root@localhost's password:
You are required to change your password immediately (root enforced)
Last failed login: Fri Oct 31 17:51:32 UTC 2025 from 172.18.0.3 on ssh:notty
There was 1 failed login attempt since the last successful login.
Last login: Fri Oct 31 17:50:36 2025 from 172.18.0.3
Changing password for root.
(current) UNIX password:
New password:
Retype new password:
[root@sandbox-hdp ~]#
```

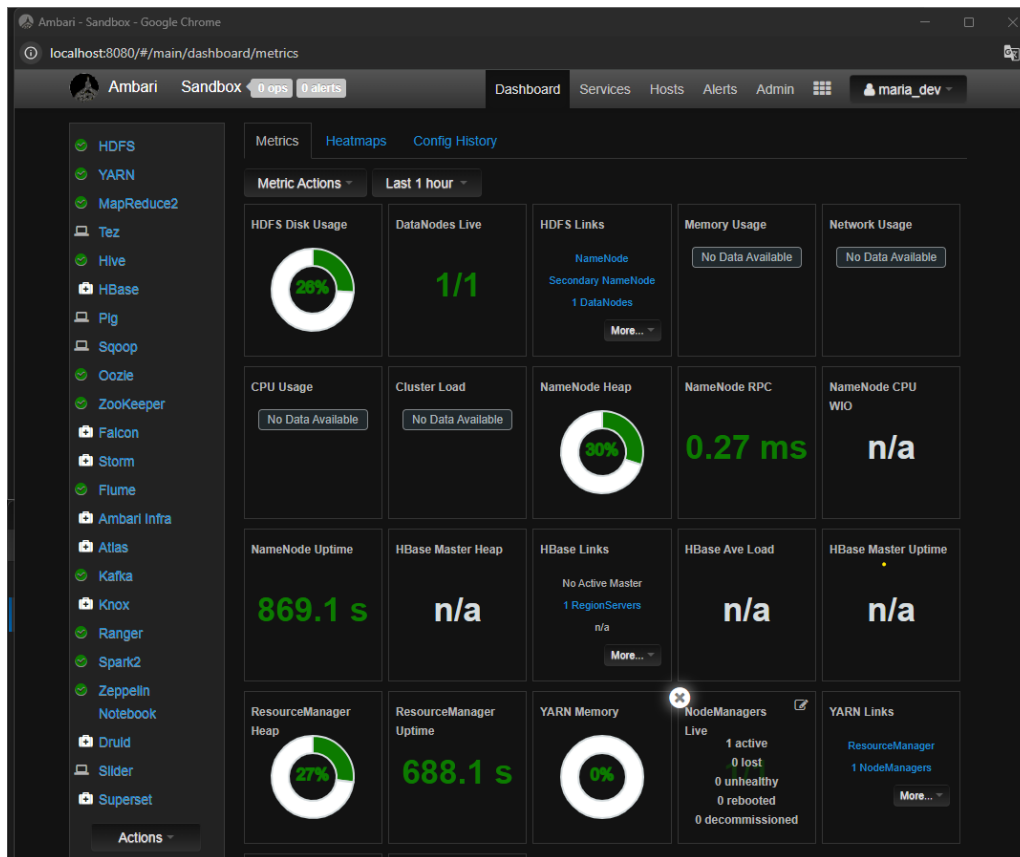
Por otro lado, para cambiar la contraseña del admin del servicio de Ambari, sería a través del siguiente comando:

```
ambari-admin-password-reset
```

***Importante → root** corresponde al usuario de Linux y **admin** del usuario del servicio de Ambari.

5. INTERFAZ AMBARI

→ Iniciar servicio



→ Panel de ejecución de operaciones

Operations	Start Time	Duration	Show:	All (10)
✓ Start All Services	Today 11:00	475.96 secs	<div></div>	100% ▶
✓ Stop required services	Mon Jun 18 2018 18:17	22.30 secs	<div></div>	100% ▶
✓ Stop required services	Mon Jun 18 2018 18:17	8.83 secs	<div></div>	100% ▶
✓ Stop required services	Mon Jun 18 2018 18:17	7.35 secs	<div></div>	100% ▶
✓ Stop required services	Mon Jun 18 2018 18:17	3.30 secs	<div></div>	100% ▶
✓ Stop required services	Mon Jun 18 2018 18:16	33.38 secs	<div></div>	100% ▶

BIG DATA

→ Información del servicio

The screenshot shows the Ambari interface with the 'Services' tab selected. The left sidebar lists various services, with HDFS at the top. The main content area displays the 'Summary' for the HDFS service, which is highlighted with a yellow border. The summary includes status indicators for NameNode, SNameNode, and DataNodes, along with disk usage, block errors, and other metrics. A 'No alerts' badge is visible in the top right corner of the summary box.

Component	Status	Alerts
NameNode	Started	No alerts
SNameNode	Started	No alerts
DataNodes	1/1 Started	

Additional metrics shown:

- Disk Remaining: 78.4 GB / 106.0 GB (73.97%)
- Blocks (total): 1121
- Block Errors: 0 corrupt replica / 0 missing / 0 under replicated
- Total Files + Directories: 1352
- Upgrade Status: No pending upgrade
- Safe Mode Status: Not in safe mode
- NameNode Uptime: 16.99 mins
- NameNode Heap: 24.1 MB / 240.0 MB (10.0% used)
- Disk Usage (DFS Used): 2.1 GB / 106.0 GB (1.95%)
- Disk Usage (Non DFS Used): 25.5 GB / 106.0 GB (24.08%)

→ Dashboard

Información de todos los servicios actuales

The screenshot shows the Ambari Dashboard with the 'Dashboard' tab selected. The left sidebar lists various services. The main content area displays a grid of metrics for different services, highlighted with a yellow border. The metrics include HDFS Disk Usage, DataNodes Live, HDFS Links, Memory Usage, Network Usage, CPU Usage, Cluster Load, NameNode Heap, NameNode RPC, NameNode CPU WIO, NameNode Uptime, HBase Master Heap, HBase Links, HBase Ave Load, HBase Master Uptime, ResourceManager Heap, ResourceManager Uptime, YARN Memory, NodeManagers Live, and YARN Links. Each metric is represented by a gauge or a text display showing the current value and status.

Metric	Value
HDFS Disk Usage	28%
DataNodes Live	1/1
HDFS Links	NameNode, Secondary NameNode, 1 DataNodes
Memory Usage	No Data Available
Network Usage	No Data Available
CPU Usage	No Data Available
Cluster Load	No Data Available
NameNode Heap	17%
NameNode RPC	0.55 ms
NameNode CPU WIO	n/a
NameNode Uptime	18.4 min
HBase Master Heap	n/a
HBase Links	No Active Master, 1 RegionServers, n/a
HBase Ave Load	n/a
HBase Master Uptime	n/a
ResourceManager Heap	25%
ResourceManager Uptime	922.2 s
YARN Memory	0%
NodeManagers Live	1/1
YARN Links	ResourceManager, 1 NodeManagers

BIG DATA

→ HDFS → Quick Links → NameNode UI

The screenshot shows the Hadoop NameNode UI Overview page for 'sandbox-hdp.hortonworks.com:8020' (active). The page has a green navigation bar with tabs: Hadoop, Overview, Datanodes, Datanode Volume Failures, Snapshot, Startup Progress, and Utilities. The Overview tab is selected. Below the navigation bar, the title 'Overview 'sandbox-hdp.hortonworks.com:8020' (active)' is displayed. A table provides key information:

Started:	Wed Nov 05 10:02:17 UTC 2025
Version:	2.7.3.2.6.5.0-292, r3091053c59a62c82d82c9f778c48bde5ef0a89a1
Compiled:	2018-05-11T07:53Z by jenkins from (detached from 3091053)
Cluster ID:	CID-a915d782-8c97-4ece-9351-cd595a0c9b61
Block Pool ID:	BP-243674277-172.17.0.2-1529333510191

Below the table, a 'Summary' section is visible. It states: 'Security is off.', 'Safemode is off.', '1352 files and directories, 1121 blocks = 2473 total filesystem object(s).', 'Heap Memory used 49.85 MB of 240 MB Heap Memory. Max Heap Memory is 240 MB.', and 'Non Heap Memory used 81.83 MB of 83.52 MB Committed Non Heap Memory. Max Non Heap Memory is <unbonded>.'

6. Remote SSH (Visual Studio Code)

Desde Visual Studio Code, buscamos e instalamos la extension *Remote SSH*

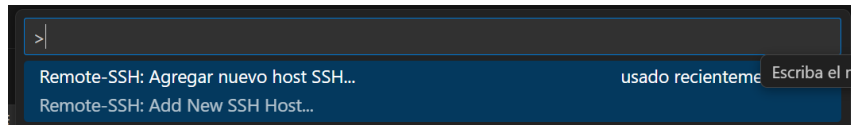
The screenshot shows the Visual Studio Code Extensions Marketplace. The search bar contains 'remote ssh'. The results list several extensions, with 'Remote - SSH' by Microsoft highlighted by a yellow box. The details for 'Remote - SSH' are:

- Icon: A blue circle with a white terminal icon.
- Name: **Remote - SSH**
- Description: Open any folder on a remote machine usin...
- Publisher: Microsoft
- Version: 133ms
- Settings icon: A gear icon.

Below 'Remote - SSH', other extensions are visible, including 'Remote - SSH: Editing Configura...' and 'Remote Development'.

BIG DATA

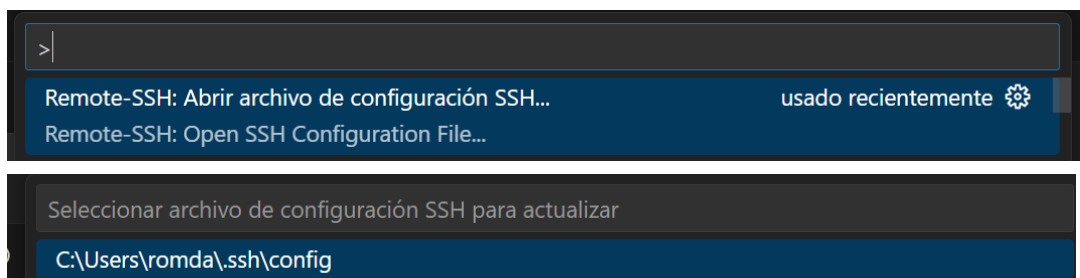
Una vez instalado, accedemos a través de Ctrl+Shift+P, al menú para añadir un nuevo host.



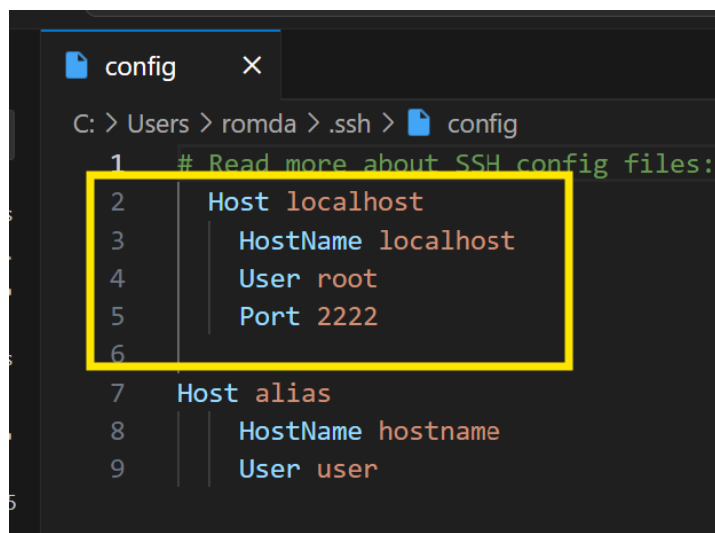
Y añadimos la línea para el archivo de configuración:

```
ssh root@localhost -p 2222
```

O abrimos el archivo de configuración y añadimos las siguientes líneas:

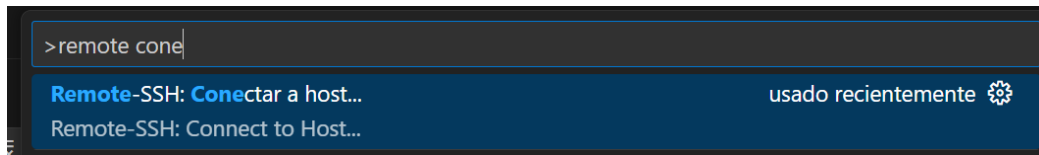


```
Host localhost
  HostName localhost
  User root
  Port 2222
```



BIG DATA

Una vez realizada la configuración, nos conectamos al host a través del nombre que le hayamos asignado:



Al llegar a este punto e intentar conectarse, daría error por incompatibilidad entre la versión de CentOS en la que está montada la máquina, y la actual versión de Visual Studio Code.