

Explicación previa

Vamos a crear una máquina virtual Ubuntu Server con Spark y sus dependencias para funcionar.

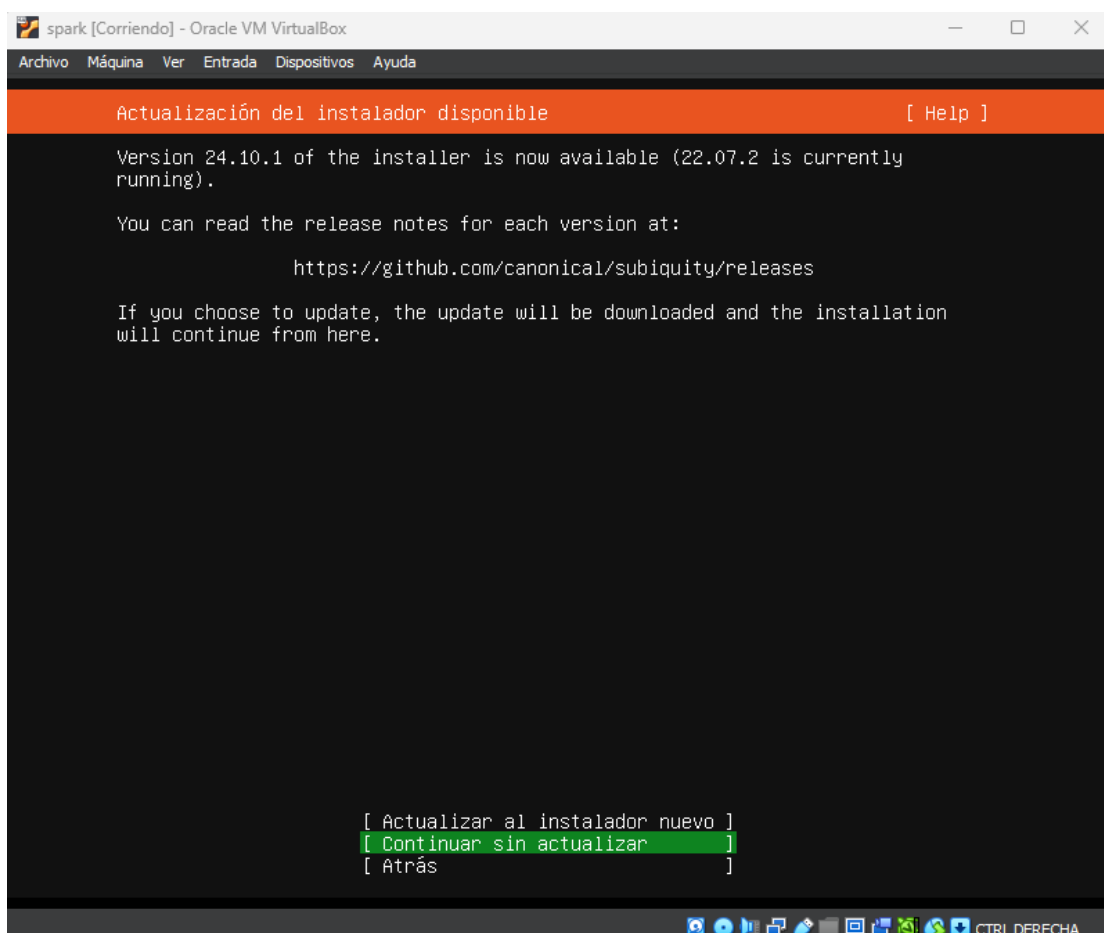
Requisitos:

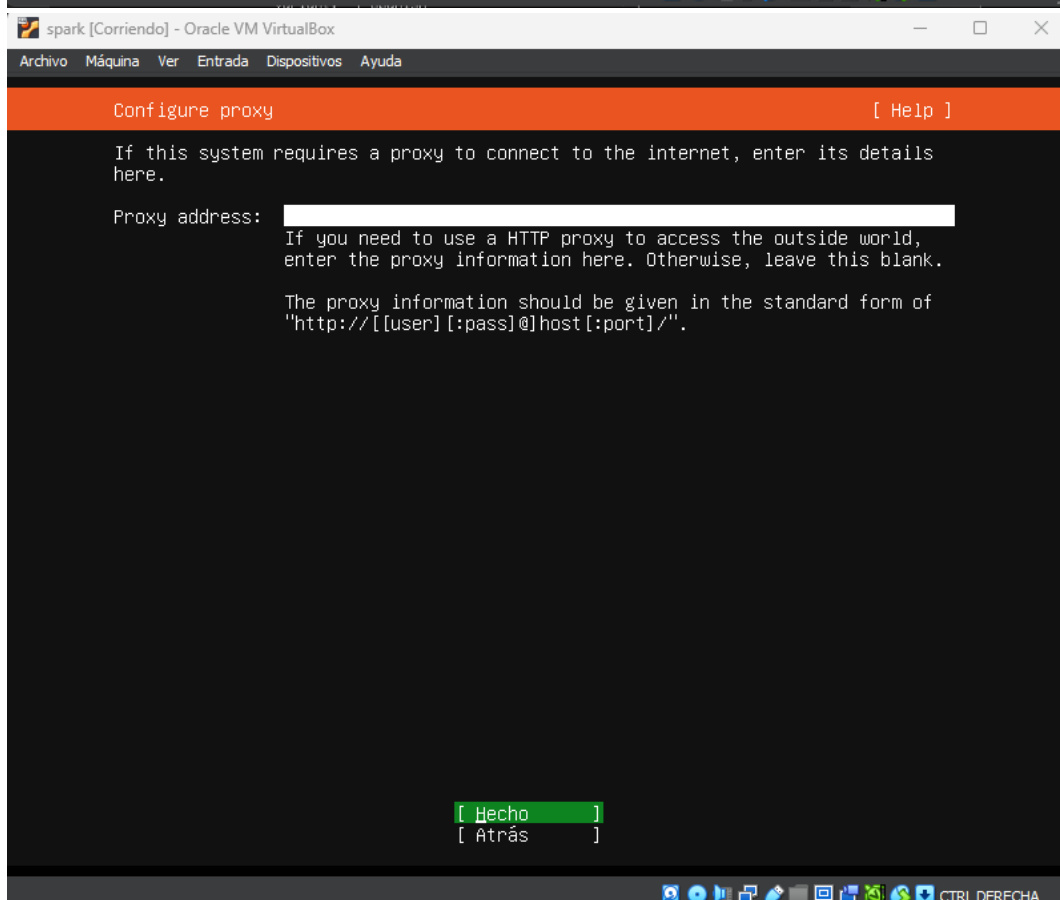
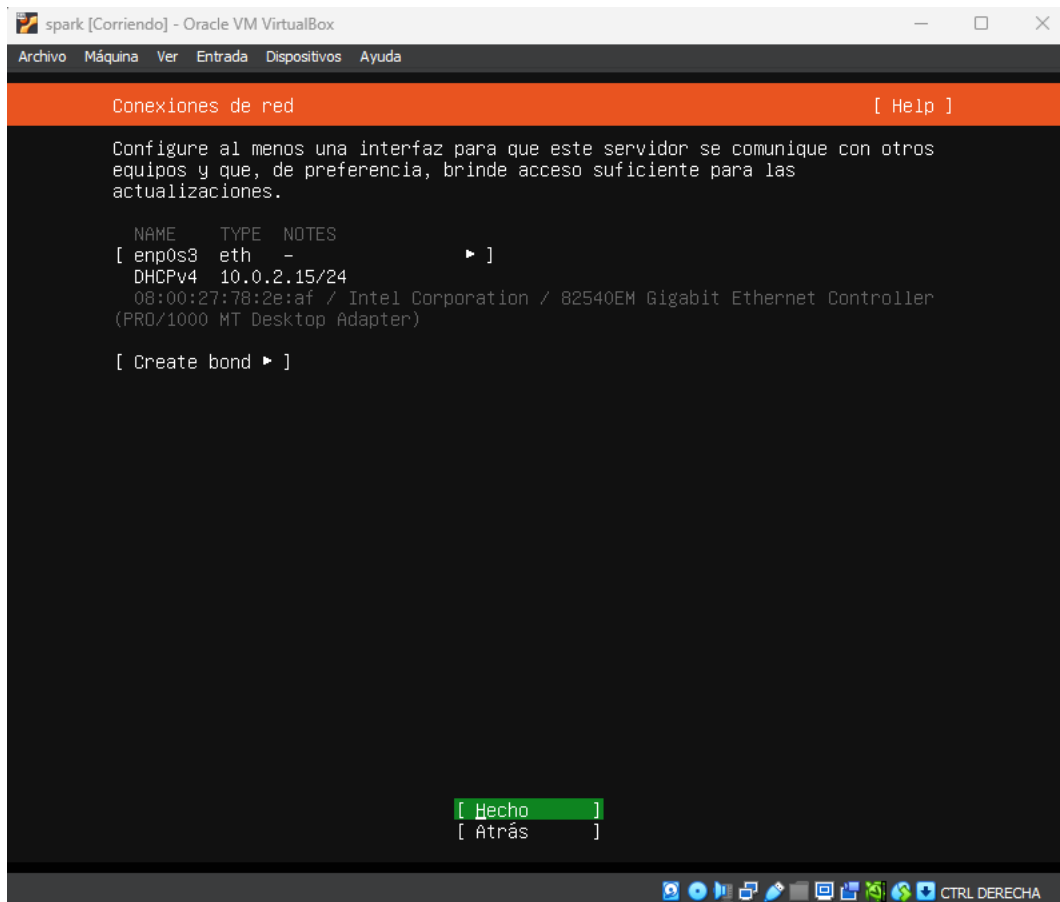
- En host:
 - Ubuntu desktop [20.04]
 - Virtual Box
- En la máquina virtual:
 - Spark
 - Java JDK
 - Python

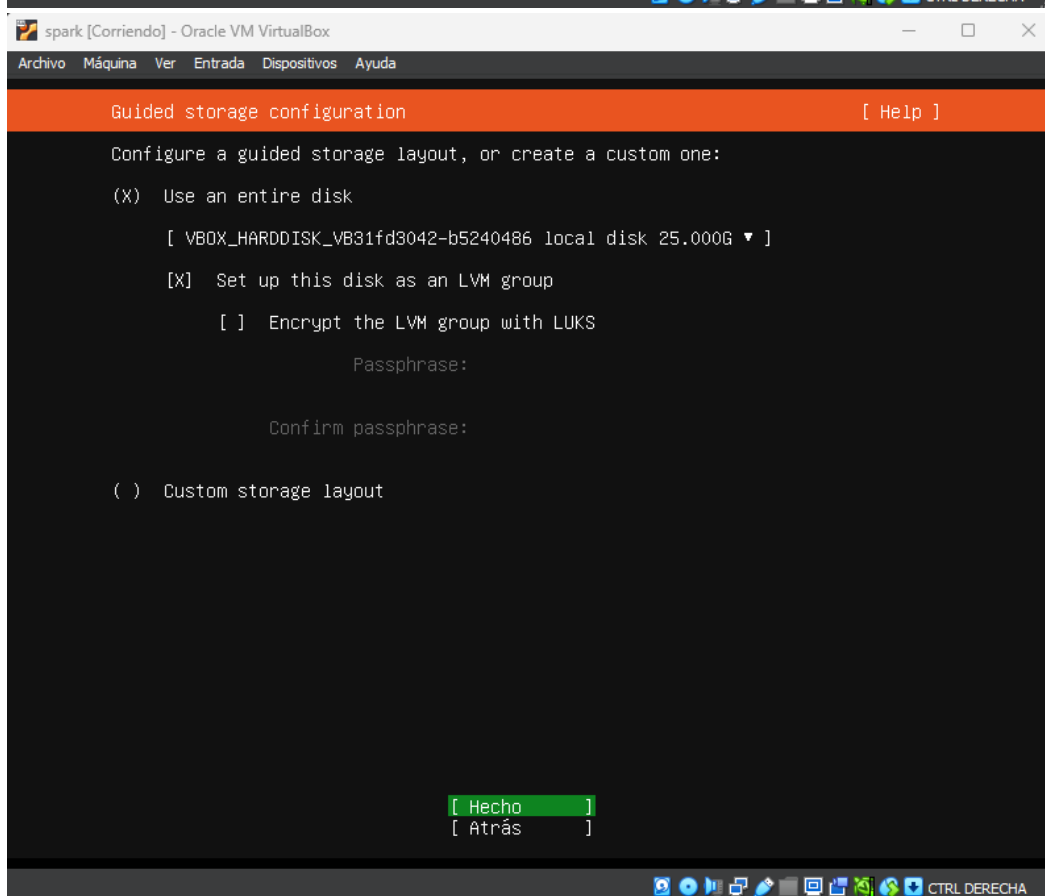
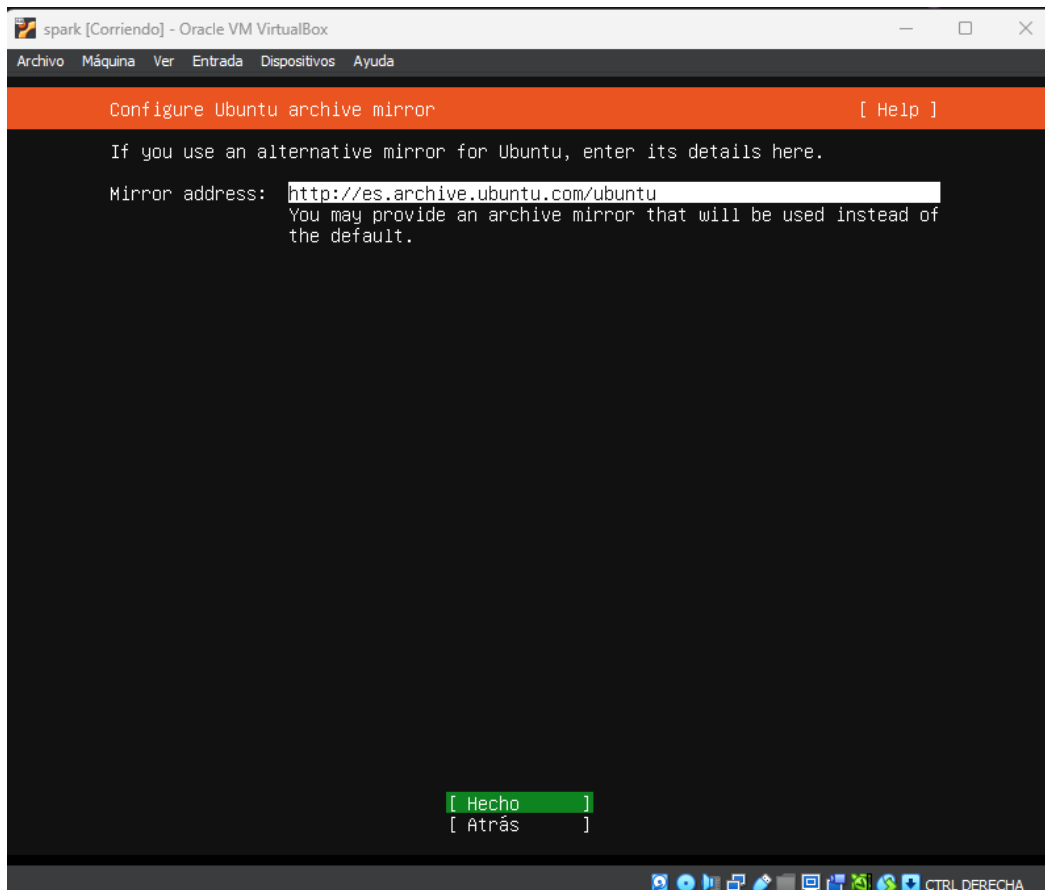
Creación de la máquina virtual en Virtual Box

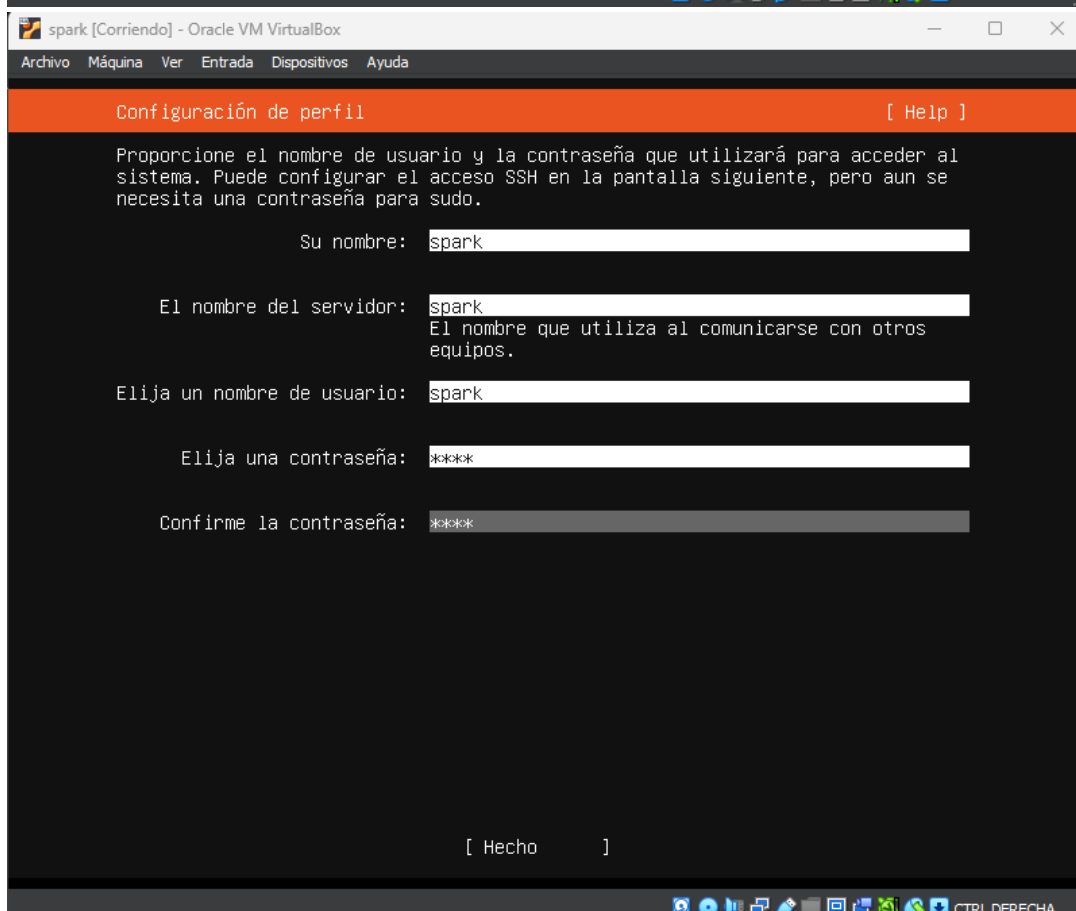
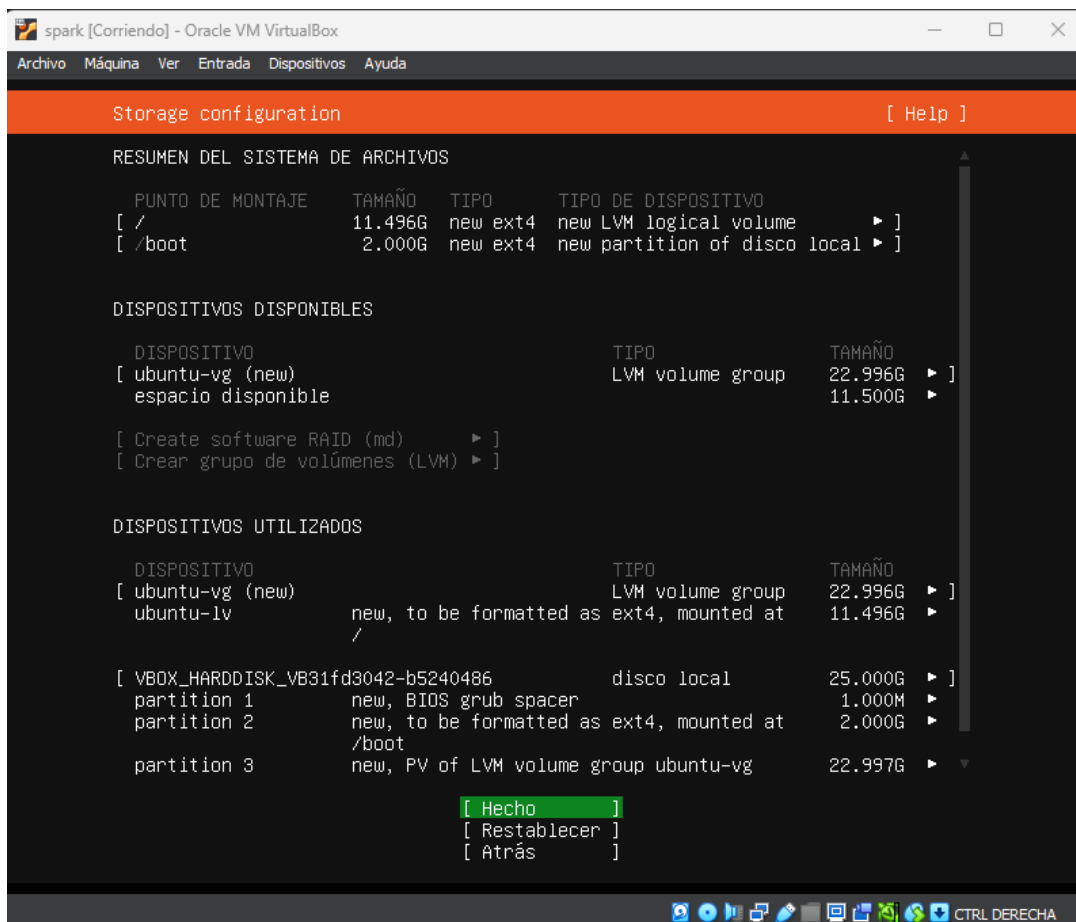
Configuración:

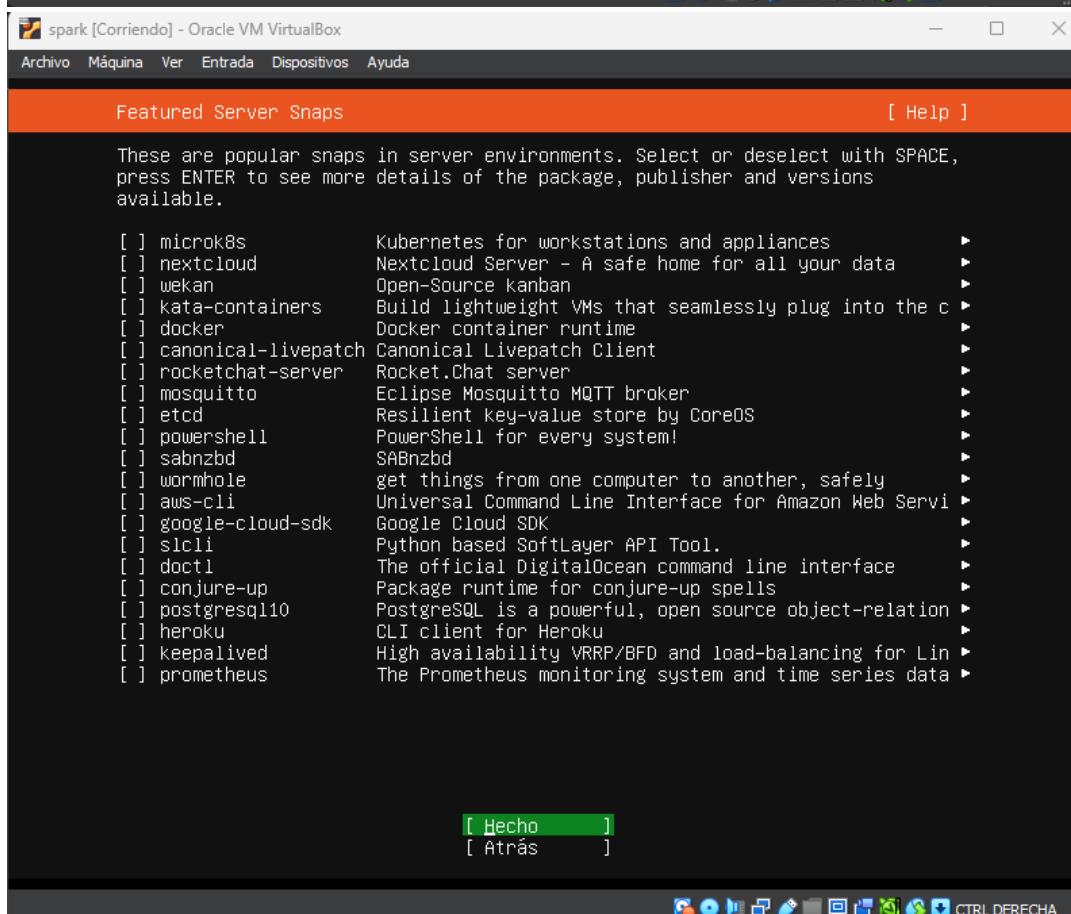
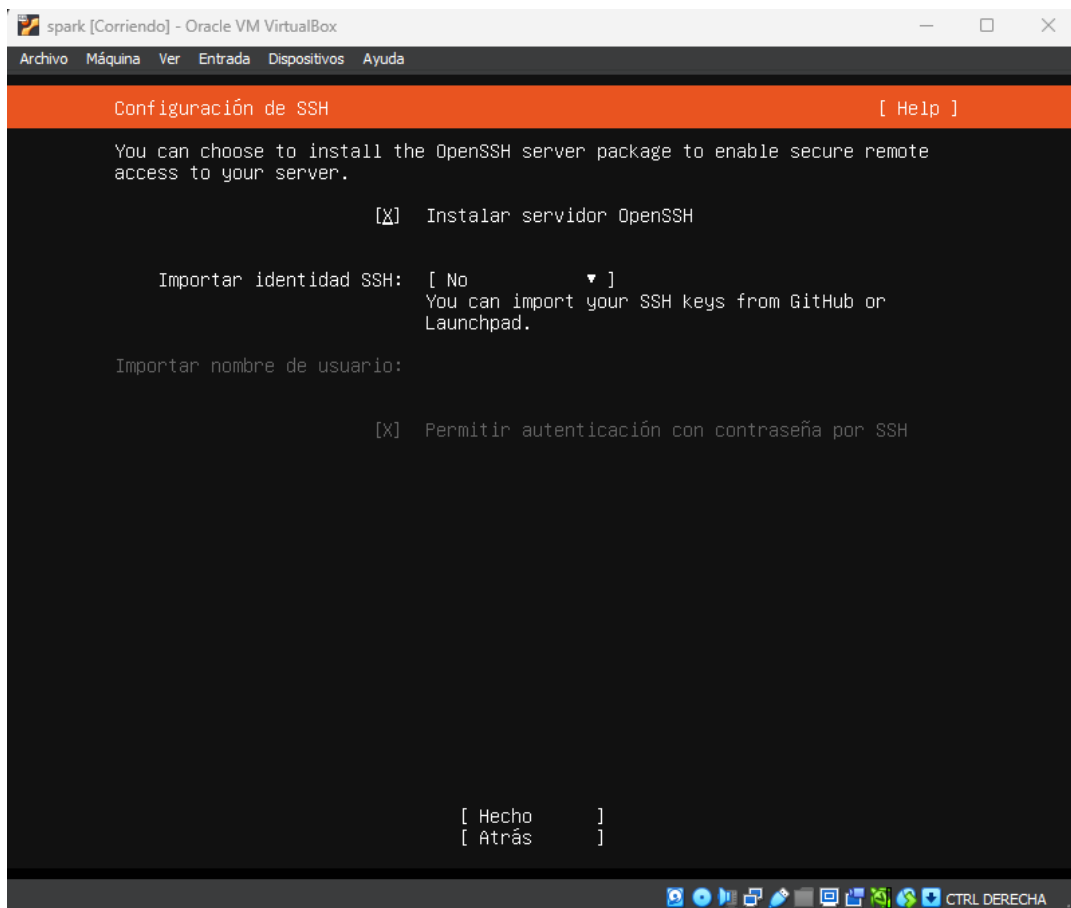
- Memoria 4096, 2 CPUs, 25 Gb de disco duro











Pasos para instalación de Spark en el servidor

```
spark@spark:~$ sudo apt update  
[sudo] password for spark: _
```

```
spark@spark:~$ sudo apt upgrade_
```

Instalar java

```
spark@spark:~$ sudo apt install openjdk-8-jdk_
```

```
spark@spark:~$ java -version  
openjdk version "1.8.0_432"  
OpenJDK Runtime Environment (build 1.8.0_432-8u432-ga~us1-0ubuntu2~20.04-ga)  
OpenJDK 64-Bit Server VM (build 25.432-bga, mixed mode)
```

Instalar Python3

```
spark@spark:~$ sudo apt install python3  
Leyendo lista de paquetes... Hecho  
Creando árbol de dependencias  
Leyendo la información de estado... Hecho  
python3 ya está en su versión más reciente (3.8.2-0ubuntu2).  
fijado python3 como instalado manualmente.  
0 actualizados, 0 nuevos se instalarán, 0 para eliminar y 0 no actualizados.
```

```
spark@spark:~$ python3 --version  
Python 3.8.10
```

Descargar version de spark

```
spark@spark:~$ wget https://downloads.apache.org/spark/spark-3.5.3/spark-3.5.3-bin-hadoop3.tgz_
```

Descomprimir tar y comprobar archivos

```
spark@spark:~$ tar -xvf spark-3.5.3-bin-hadoop3.tgz_
```

```
spark@spark:~$ ls  
spark-3.5.3-bin-hadoop3  spark-3.5.3-bin-hadoop3.tgz  
spark@spark:~$ ls spark-3.5.3-bin-hadoop3  
bin  data  jars  LICENSE  NOTICE  R  RELEASE  yarn  
conf  examples  kubernetes  licenses  python  README.md  sbin  
spark@spark:~$
```

Mover el contenido de tar a /opt/spark

```
spark@spark:~$ sudo mv spark-3.5.3-bin-hadoop3 /opt/spark  
spark@spark:~$ find / .bashrc_
```

Añadir variables de entorno al archivo .bashrc

```
spark@spark:~$ sudo nano .bashrc
```

