

# Guía de instalación del contenedor para el seguimiento de la sesión

Eduardo Cano García

Alejandro Fernández Camello

## Índice

Introducción .....	1
1. Docker Desktop .....	1
2. Descarga de la imagen.....	2
3. Creación y ejecución del contenedor .....	2

## Introducción

Mediante este breve documento se pretende proporcionar todo el proceso a seguir para instalar y configurar el entorno de trabajo.

Una vez seguidos todos los pasos, se podrá utilizar *Apache Spark* en *Python* mediante *Jupyter notebook* dentro de un contenedor *Docker*.

El motivo de usar *Docker* recae en que consideramos que es la manera más simple de utilizar instalar *Apache Spark*.

Se probó a instalar de una forma más convencional en varios sistemas operativos, pero era bastante complejo y propenso a errores, además de consumir mucho más espacio en disco que un contenedor y requerir variables de entorno.

Como tuvimos las sesiones sobre *Docker*, se cayó en la cuenta del poder del mismo, por lo que se aplicaron los conocimientos aprendidos, resultando en este documento.

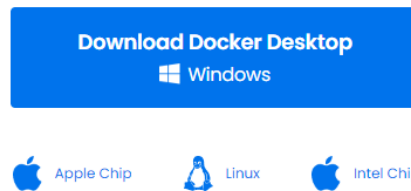
## 1. Docker Desktop

Lo primero que hay que hacer es descargar **Docker Desktop** desde su página oficial:

<https://www.docker.com/products/docker-desktop/>

# Docker Desktop

Install Docker Desktop – the fastest way to containerize applications.



**Figura 1. Descarga de Docker Desktop.**

Simplemente se selecciona el sistema operativo y se descarga.

Tras ello se siguen las instrucciones del instalador.

**Nota:** en caso de pedir instalar **WSL** (Windows Subsystem for Linux), instalarlo también.

El propio *Docker Desktop* indicará como hacerlo, aunque se puede hacer desde la tienda de *Microsoft*:

<https://apps.microsoft.com/store/detail/ubuntu/9PDXGNCFSZV?hl=es-es&gl=es>

## 2. Descarga de la imagen

Una vez se tiene *Docker Desktop* instalado, el primer paso a realizar es descargar la imagen mediante el siguiente comando:

```
docker pull jupyter/pyspark-notebook:latest
```

Con eso se descarga la imagen correspondiente a la última versión de *Jupyter notebook* con *pyspark*.

## 3. Creación y ejecución del contenedor

En este punto se debe abrir una terminal.

Ahora con el comando que se muestra a continuación (aunque salga en 2 líneas va todo en la misma, pues es un único comando) se crea el contenedor con la imagen descargada anteriormente:

```
docker run -p 10000:8888 -p 4050:4050 --name container jupyter/pyspark-notebook
```

**IMPORTANTE:** no cerrar la terminal.

Tras ello, el contenedor estaría ejecutándose y se puede acceder al entorno de trabajo.

Para acceder, simplemente se accede a la dirección “**localhost:10000**”, ya que es en dicho puerto en que se ha configurado el entorno con el comando anterior.

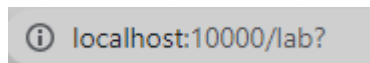


Figura 2. Dirección web para el entorno de trabajo.

Al navegar a dicha web, se solicitará un **token de acceso**.

Dicho **token** se encuentra en la terminal en que se ha ejecutado el comando anteriormente:

```
Símbolo del sistema - docker run -p 10000:8888 -p 4050:4050 --name container jupyter/pyspark-notebook

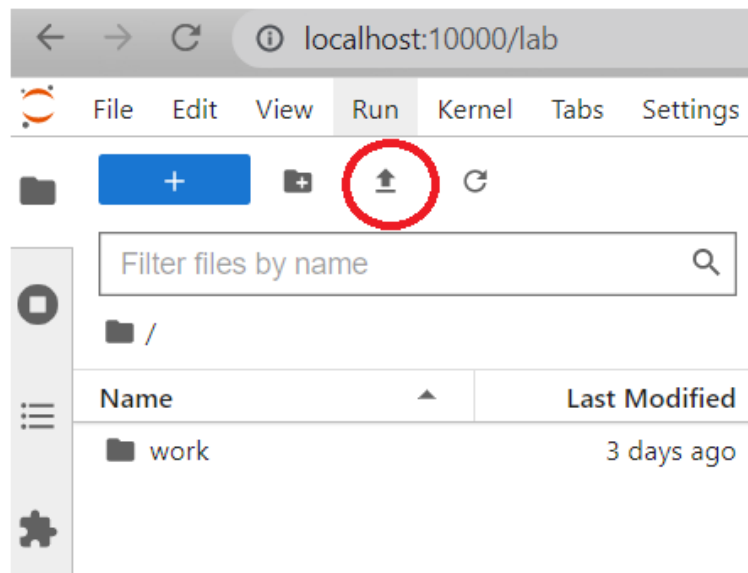
C:\Users\Usuario>docker run -p 10000:8888 -p 4050:4050 --name container jupyter/pyspark-notebook
Entered start.sh with args: jupyter lab
/usr/local/bin/start.sh: running hooks in /usr/local/bin/before-notebook.d as uid / gid: 1000 / 100
/usr/local/bin/start.sh: running script /usr/local/bin/before-notebook.d/spark-config.sh
/usr/local/bin/start.sh: done running hooks in /usr/local/bin/before-notebook.d
Executing the command: jupyter lab
[I 2022-11-30 19:21:29.314 ServerApp] jupyterlab | extension was successfully linked.
[W 2022-11-30 19:21:29.320 NotebookApp] 'ip' has moved from NotebookApp to ServerApp. This config will be passed to ServerApp. Be
sure to update your config before our next release.
[W 2022-11-30 19:21:29.320 NotebookApp] 'port' has moved from NotebookApp to ServerApp. This config will be passed to ServerApp.
Be sure to update your config before our next release.
[W 2022-11-30 19:21:29.320 NotebookApp] 'port' has moved from NotebookApp to ServerApp. This config will be passed to ServerApp.
Be sure to update your config before our next release.
[I 2022-11-30 19:21:29.328 ServerApp] nbclassic | extension was successfully linked.
[I 2022-11-30 19:21:29.330 ServerApp] Writing Jupyter server cookie secret to /home/jovyan/.local/share/jupyter/runtime/jupyter_c
ookie_secret
[I 2022-11-30 19:21:29.554 ServerApp] notebook_shim | extension was successfully linked.
[I 2022-11-30 19:21:29.570 ServerApp] notebook_shim | extension was successfully loaded.
[I 2022-11-30 19:21:29.571 LabApp] JupyterLab extension loaded from /opt/conda/lib/python3.10/site-packages/jupyterlab
[I 2022-11-30 19:21:29.571 LabApp] JupyterLab application directory is /opt/conda/share/jupyter/lab
[I 2022-11-30 19:21:29.575 ServerApp] jupyterlab | extension was successfully loaded.
[I 2022-11-30 19:21:29.578 ServerApp] nbclassic | extension was successfully loaded.
[I 2022-11-30 19:21:29.578 ServerApp] Serving notebooks from local directory: /home/jovyan
[I 2022-11-30 19:21:29.578 ServerApp] Jupyter Server 1.23.3 is running at:
[I 2022-11-30 19:21:29.578 ServerApp] http://e4aa52c59012:8888/lab?token=cbe2f5629cb26c6a65a7bb7ec474dfdac105cd0dbe194bff
[I 2022-11-30 19:21:29.578 ServerApp] or http://127.0.0.1:8888/lab?token=cbe2f5629cb26c6a65a7bb7ec474dfdac105cd0dbe194bff
[I 2022-11-30 19:21:29.578 ServerApp] Use Control-C to stop this server and shut down all kernels (twice to skip confirmation).
[C 2022-11-30 19:21:29.583 ServerApp]

To access the server, open this file in a browser:
file:///home/jovyan/.local/share/jupyter/runtime/jpserver-7-open.html
Or copy and paste one of these URLs:
http://e4aa52c59012:8888/lab?token=cbe2f5629cb26c6a65a7bb7ec474dfdac105cd0dbe194bff
or http://127.0.0.1:8888/lab?token=cbe2f5629cb26c6a65a7bb7ec474dfdac105cd0dbe194bff
```

Figura 3. Localización del token para abrir el notebook.

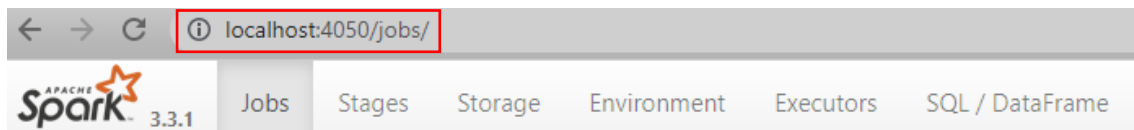
Una vez se ha accedido, se tiene preparado el entorno de trabajo.

Solo falta subir los archivos que se indiquen durante la exposición. Para ello, se deben subir desde la carpeta en que se hayan descargado, haciendo clic en el icono resaltado en la Figura que se muestra a continuación:



**Figura 4. Subida de archivos al notebook.**

**Nota:** para visualizar la consola de trabajos, se debe navegar a la dirección localhost:4050.



**Figura 5. Consola de trabajos.**