



## MANUAL DE INSTALACIÓN



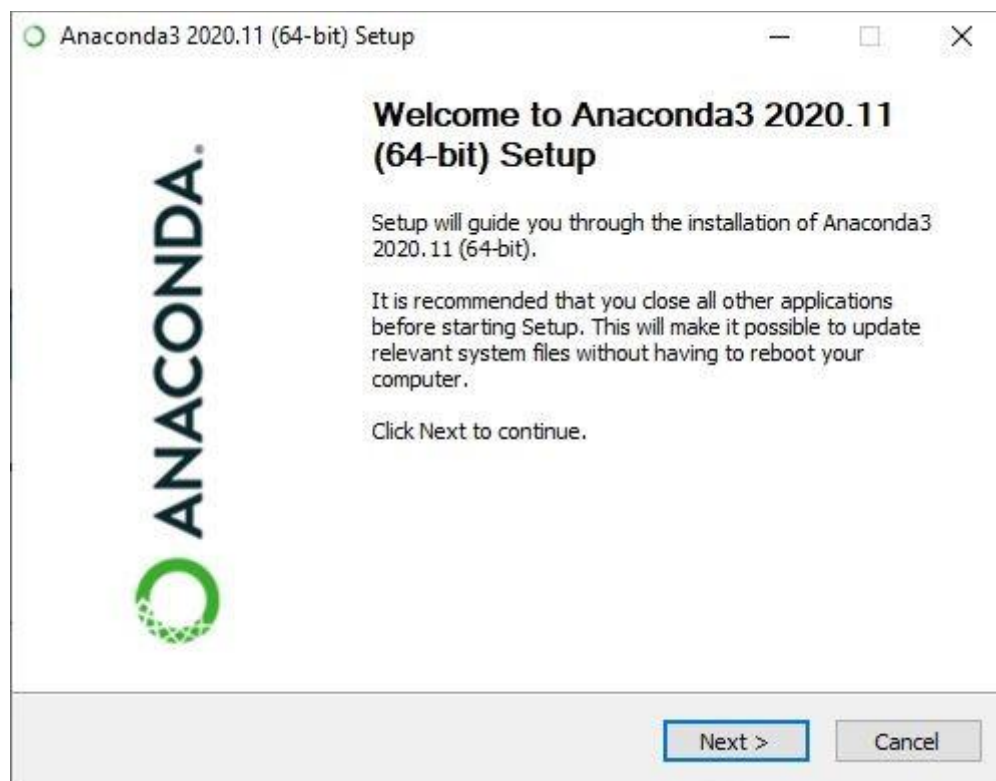
## Spark - Jupyter - Scala

Para poder utilizar estos componentes, una de las formas de descarga es por medio de Anaconda. Para ello se realizan los siguientes pasos.

1. Ir directamente a descargar el instalador de Anaconda, dependiendo del sistema operativo: <https://www.anaconda.com/products/individual#Downloads>

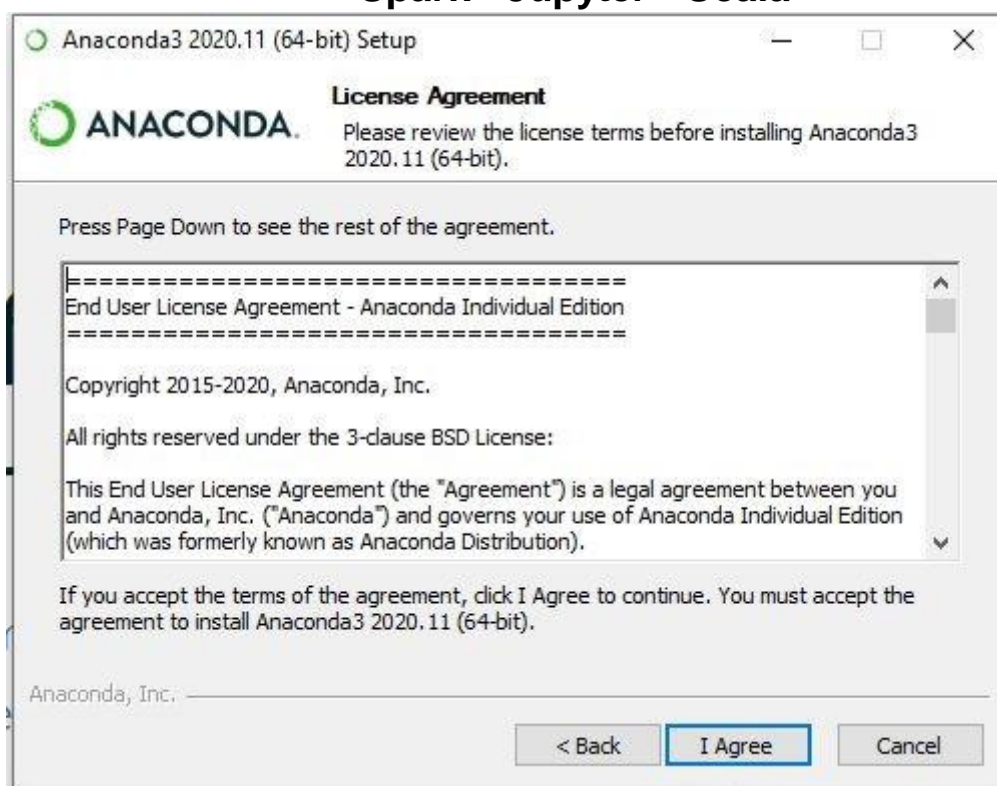


2. Continuamos con el proceso de instalación.

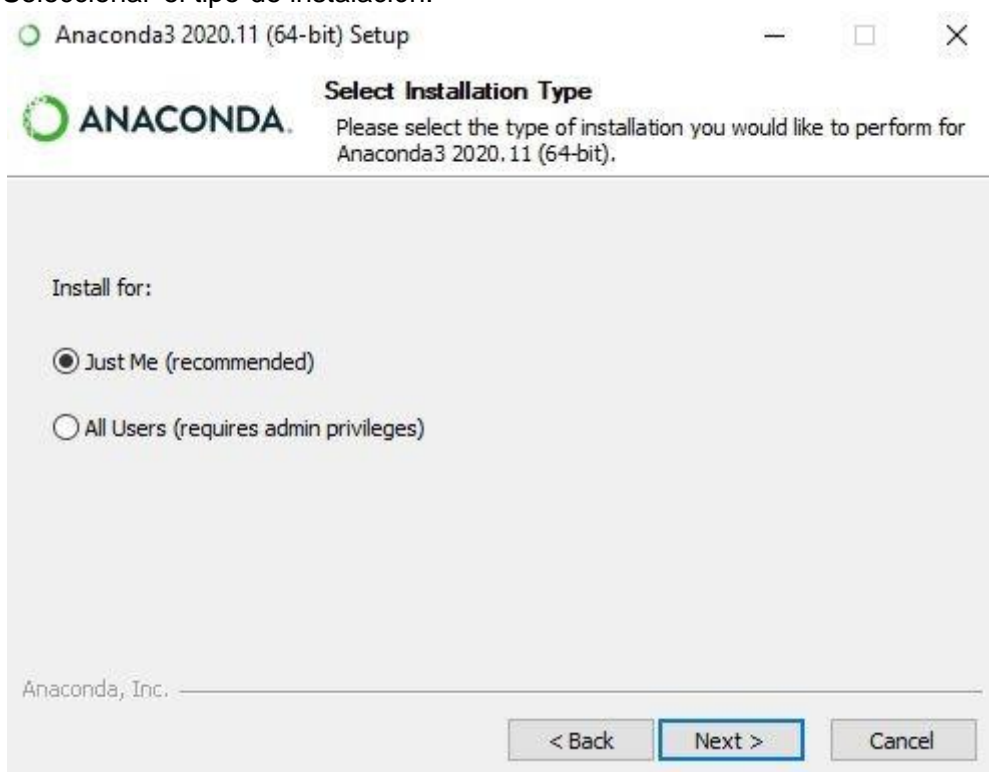


3. Leer y aceptar **License Agreement**

## Spark - Jupyter - Scala

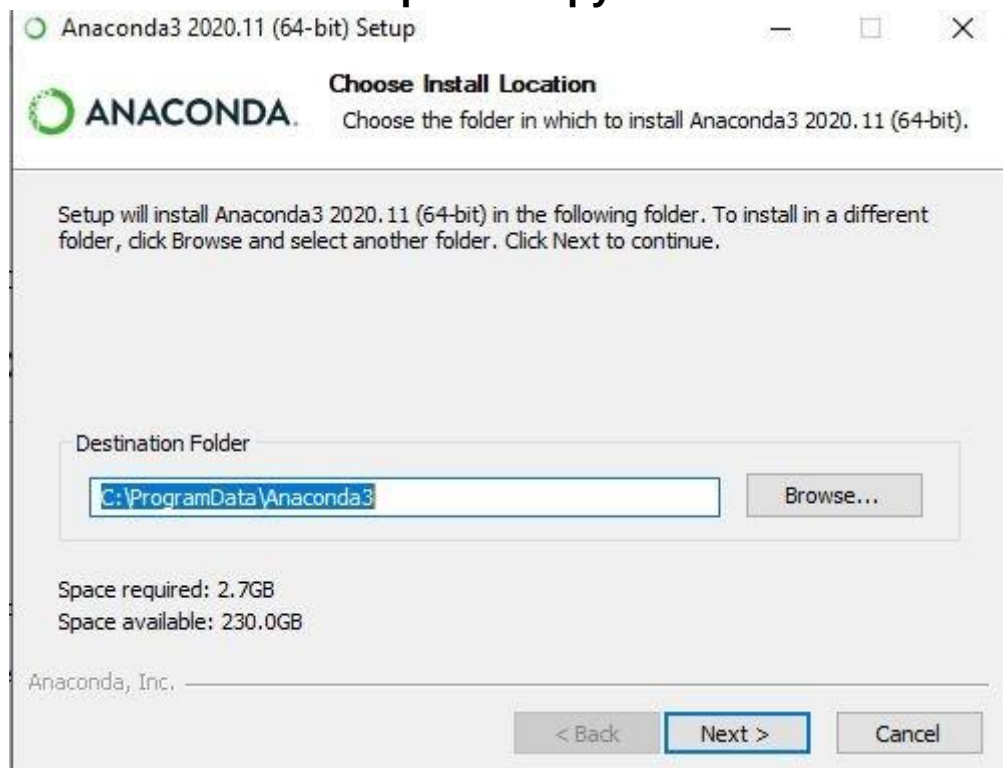


4. Seleccionar el tipo de instalación.

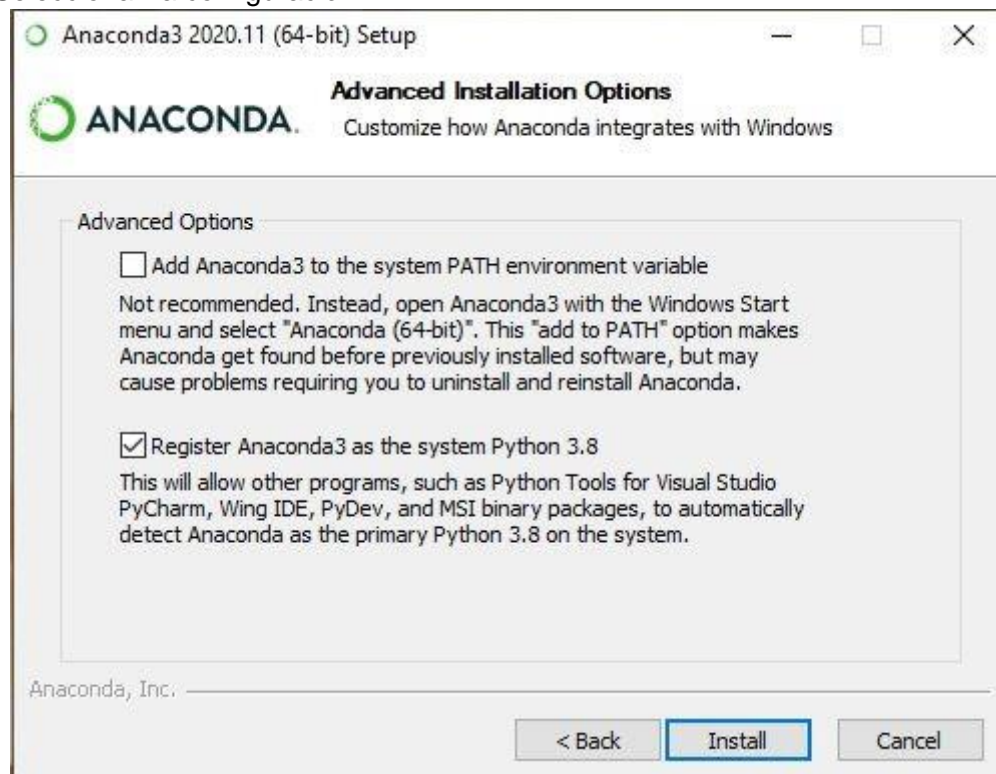


5. Seleccionar la ubicación de la instalación

## Spark - Jupyter - Scala

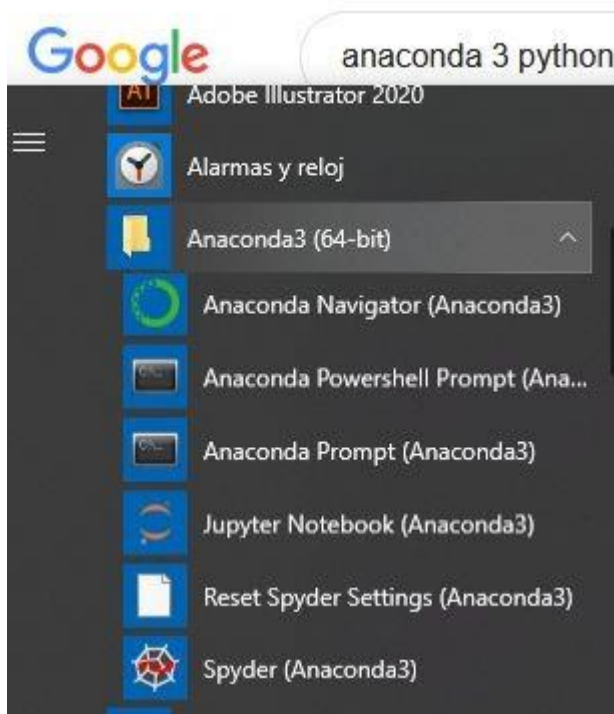


### 6. Seleccionar la configuración



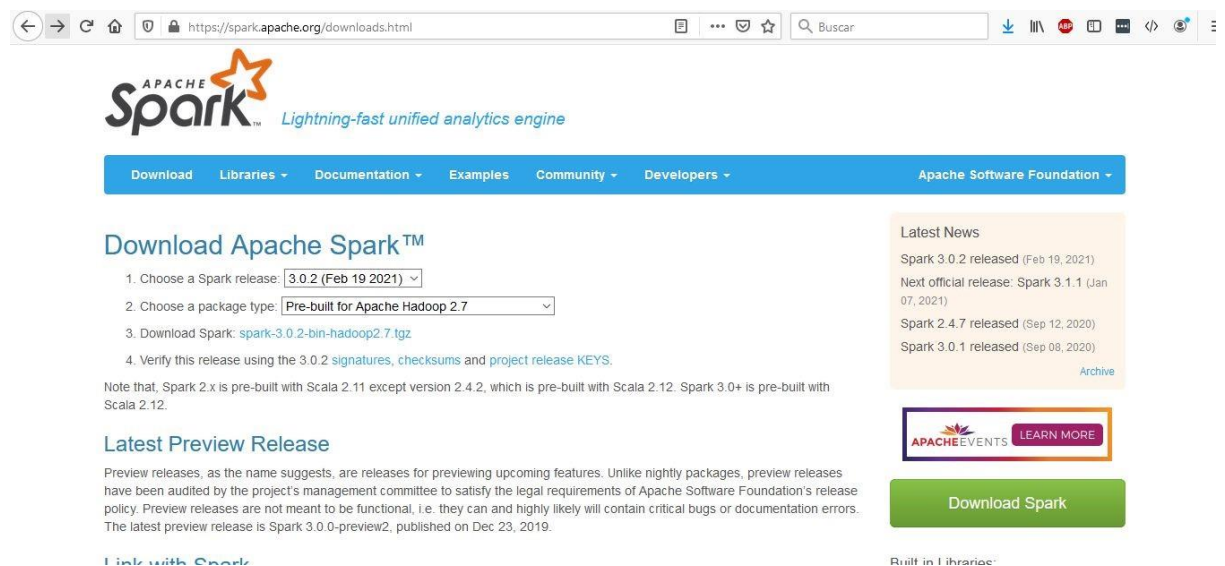
### 7. Al terminar la instalación ya deberá tener los accesos a las aplicaciones

## Spark - Jupyter - Scala



8. Descargar Spark, para ello nos dirigimos a la siguiente página:

<https://spark.apache.org/downloads.html>.




9. Procedemos con la descarga seleccionada



## Spark - Jupyter - Scala


← → ↻ 🏠 🔒 <https://www.apache.org/dyn/doser.lua/spark/spark-3.0.2/spark-3.0.2-bin-hadoop2.7.tgz> 🔍 Buscar ⬇️ 📄 📄 📄 📄 📄

News About ▾ Make a Donation ▾ The Apache Way ▾ Join Us ▾ Downloads ▾ 🔍



COMMUNITY-LED DEVELOPMENT "THE APACHE WAY"

Projects ▾ People ▾ Community ▾ License ▾ Sponsors ▾



We suggest the following mirror site for your download:  
<https://downloads.apache.org/spark/spark-3.0.2/spark-3.0.2-bin-hadoop2.7.tgz>  
 Other mirror sites are suggested below.  
 It is essential that you verify the integrity of the downloaded file using the PGP signature (.asc file) or a hash (.md5 or .sha\* file).  
 Please only use the backup mirrors to download KEYS, PGP signatures and hashes (SHA\* etc) -- or if no other mirrors are working.

### HTTP

10. Creamos una carpeta llamada Spark:

📁 > Este equipo > Windows (C:)

	Nombre	Fecha de modifica...	Tipo	Tamaño
📁	Spark	24/02/2021 9:43 p....	Carpeta de archivos	

11. Allí pegamos todo lo que se ha descargado en el punto 9.

📁 > Este equipo > Windows (C:) > Spark

	Nombre	Fecha de modificación	Tipo	Tamaño
📁	bin	24/02/2021 9:44 p.m.	Carpeta de archivos	
📁	conf	24/02/2021 9:44 p.m.	Carpeta de archivos	
📁	data	24/02/2021 9:39 p.m.	Carpeta de archivos	
📁	examples	24/02/2021 9:39 p.m.	Carpeta de archivos	
📁	jars	24/02/2021 9:45 p.m.	Carpeta de archivos	
📁	kubernetes	24/02/2021 9:39 p.m.	Carpeta de archivos	
📁	licenses	24/02/2021 9:45 p.m.	Carpeta de archivos	
📁	python	24/02/2021 9:45 p.m.	Carpeta de archivos	
📁	R	24/02/2021 9:39 p.m.	Carpeta de archivos	
📁	sbin	24/02/2021 9:45 p.m.	Carpeta de archivos	
📁	yarn	24/02/2021 9:45 p.m.	Carpeta de archivos	
📄	LICENSE	16/02/2021 12:21 a.m.	Archivo	23 KB
📄	NOTICE	16/02/2021 12:21 a.m.	Archivo	57 KB
📄	README.md	16/02/2021 12:21 a.m.	Archivo MD	5 KB
📄	RELEASE	16/02/2021 12:21 a.m.	Archivo	1 KB

12. Procedemos a descargar el WinUtils. <https://github.com/stevcloughran/winutils>

## Spark - Jupyter - Scala

master 1 branch 3 tags Go to file Code

steveloughran link to cdarlint/winutils

Folder	Description	Time
hadoop-2.6.0/bin	Add Hadoop-2.6.0/HDP-2.2 windows binaries	
hadoop-2.6.3/bin	add gpg2 signatures	
hadoop-2.6.4	add 2.6.4 and 2.7.1 windows binaries	
hadoop-2.7.1	add 2.6.4 and 2.7.1 windows binaries	
hadoop-2.8.0-RC3/bin	sign Hadoop artifacts	
hadoop-2.8.1	sign Hadoop artifacts	4 years ago
hadoop-2.8.3/bin	Windows binaries for hadoop-2.8.3	3 years ago
hadoop-3.0.0/bin	Hadoop 3.0.0 windows binaries; off the release 3.0 tag, patched with ...	3 years ago

Clone  
HTTPS GitHub CLI  
<https://github.com/steveloughran/winutils>  
Use Git or checkout with SVN using the web URL.

Open with GitHub Desktop

Download ZIP

13. Dentro de la carpeta de Spark creada en el punto 10, creamos una llamada hadoop.

Este equipo > Windows (C:) > Spark

Nombre	Fecha de modificación	Tipo	Tamaño
LICENSE	16/02/2021 12:21 a.m.	Archivo	23 KB
NOTICE	16/02/2021 12:21 a.m.	Archivo	57 KB
README.md	16/02/2021 12:21 a.m.	Archivo MD	5 KB
RELEASE	16/02/2021 12:21 a.m.	Archivo	1 KB
<b>hadoop</b>	24/02/2021 9:48 p.m.	Carpeta de archivos	
sbin	24/02/2021 9:45 p.m.	Carpeta de archivos	
yarn	24/02/2021 9:45 p.m.	Carpeta de archivos	
python	24/02/2021 9:45 p.m.	Carpeta de archivos	
R	24/02/2021 9:45 p.m.	Carpeta de archivos	
licenses	24/02/2021 9:45 p.m.	Carpeta de archivos	
jars	24/02/2021 9:45 p.m.	Carpeta de archivos	
kubernetes	24/02/2021 9:45 p.m.	Carpeta de archivos	

14. Dentro de esta pegamos lo descargado en el punto 12. (En mi caso solo descargue hadoop-3.0.0/bin)

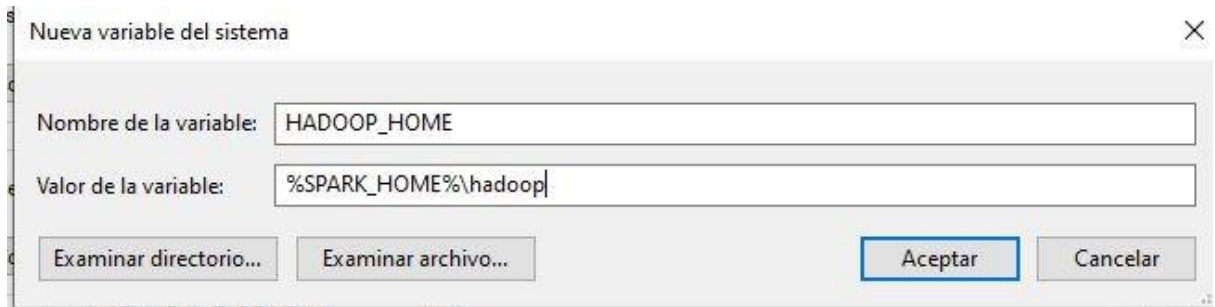
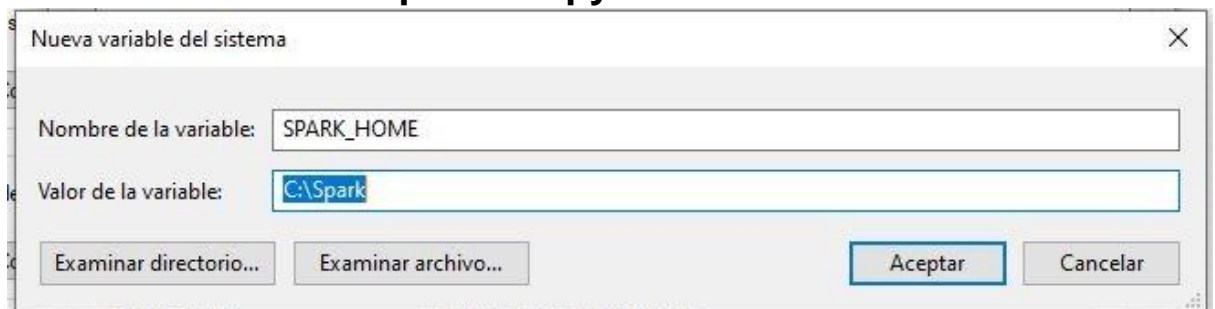
Este equipo > Windows (C:) > Spark > hadoop

Nombre	Fecha de modificación	Tipo	Tamaño
<b>bin</b>	24/02/2021 9:49 p.m.	Carpeta de archivos	

15. Creamos las variables de entorno correspondientes:

JAVA\_HOME C:\Program Files\Java\jdk1.8.0\_144

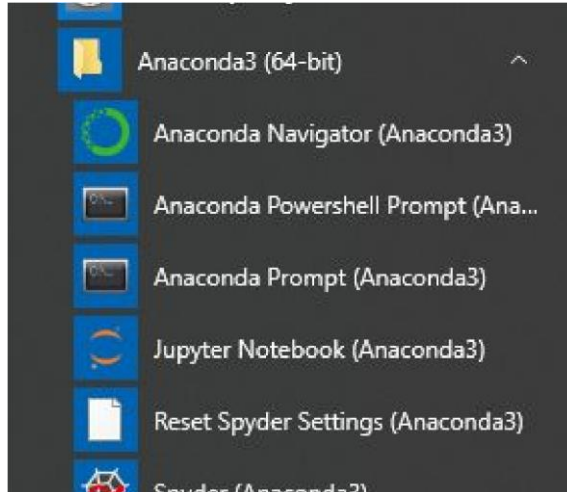
## Spark - Jupyter - Scala



### Variable Path



16. Abrimos la consola de Anaconda como administradores (Anaconda Prompt)



17. Ejecutamos los siguientes comandos:

- a. ***pip install spylon-kernel***



## Spark - Jupyter - Scala

```
Administrator: Command Prompt
[I 11:47:43.691 NotebookApp] Kernel shutdown: 6520a8ec-cde2-427a-ac90-984e0975cf4c
[I 11:47:43.692 NotebookApp] Kernel shutdown: 3e6eb459-8e16-45c2-9ee5-0451635f3cb5

C:\Windows\system32>pip install spylon-kernel
Collecting spylon-kernel
  Downloading https://files.pythonhosted.org/packages/3b/26/0c1c289ab535489b0e290461b0f2c45a00d2033a50a58a45f6d00c5fb205/spylon-kernel-0.4.1.tar.gz
Requirement already satisfied: ipykernel in c:\program files\python37\lib\site-packages (from spylon-kernel) (5.1.1)
Requirement already satisfied: jedi>=0.10 in c:\program files\python37\lib\site-packages (from spylon-kernel) (0.14.1)
Collecting metakernel (from spylon-kernel)
  Downloading https://files.pythonhosted.org/packages/bb/bd/658103e652c2c12a791c83539496541e8f8d755c1cc3293861983f0a6742/metakernel-0.24.2-py2.py3-none-any.whl (208kB)
  | 215kB 1.7MB/s
Collecting spylon[spark] (from spylon-kernel)
  Downloading https://files.pythonhosted.org/packages/a1/1b/236051a959a6cc76f5962e32aa743a31e77e3acd27a742e625d49eef4c7a/spylon-0.3.0.tar.gz (140kB)
  | 143kB 3.3MB/s
Requirement already satisfied: tornado in c:\program files\python37\lib\site-packages (from spylon-kernel) (6.0.3)
Requirement already satisfied: traitlets>=4.1.0 in c:\program files\python37\lib\site-packages (from ipykernel->spylon-kernel) (4.3.2)
Requirement already satisfied: jupyter-client in c:\program files\python37\lib\site-packages (from ipykernel->spylon-kernel) (5.3.1)
Requirement already satisfied: ipython>=5.0.0 in c:\program files\python37\lib\site-packages (from ipykernel->spylon-kernel) (7.6.1)
Requirement already satisfied: parso>=0.5.0 in c:\program files\python37\lib\site-packages (from jedi>=0.10->spylon-kernel) (0.5.1)
Collecting ipyparallel (from metakernel->spylon-kernel)
  Downloading https://files.pythonhosted.org/packages/3f/82/aaa7a357845a98d4028f27c799f0d3bb2fe55fc1247c73dc712b4ae2344c/ipyparallel-6.2.4-py2.py3-none-any.whl (198kB)
  | 204kB 939kB/s
```

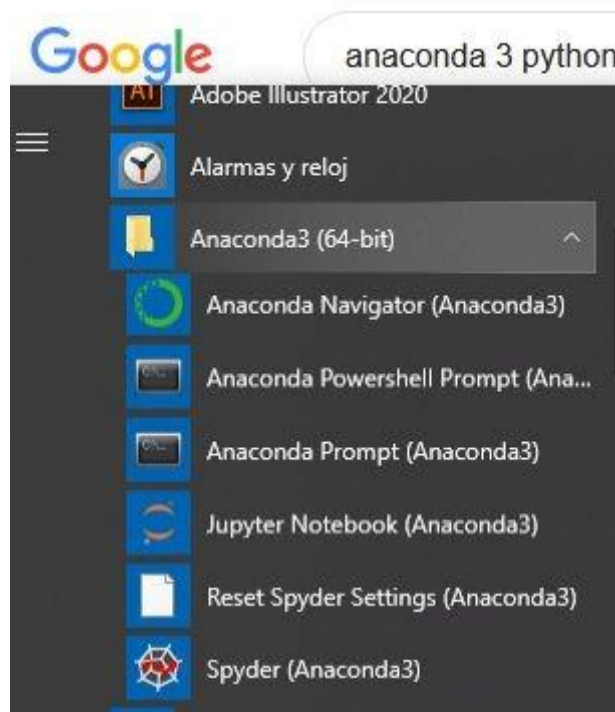
- b. Creamos el kernel spec, el cual nos permitirá seleccionar el kernel de scala a la hora de usar Jupyter notebook. Comando: ***python -m spylon\_kernel install***

```
Administrator: Command Prompt
C:\Windows\system32>
C:\Windows\system32>python -m spylon_kernel install
[InstallKernelSpec] Installed kernelspec spylon-kernel in C:\ProgramData\jupyter\kernels\spylon-kernel
C:\Windows\system32>
```

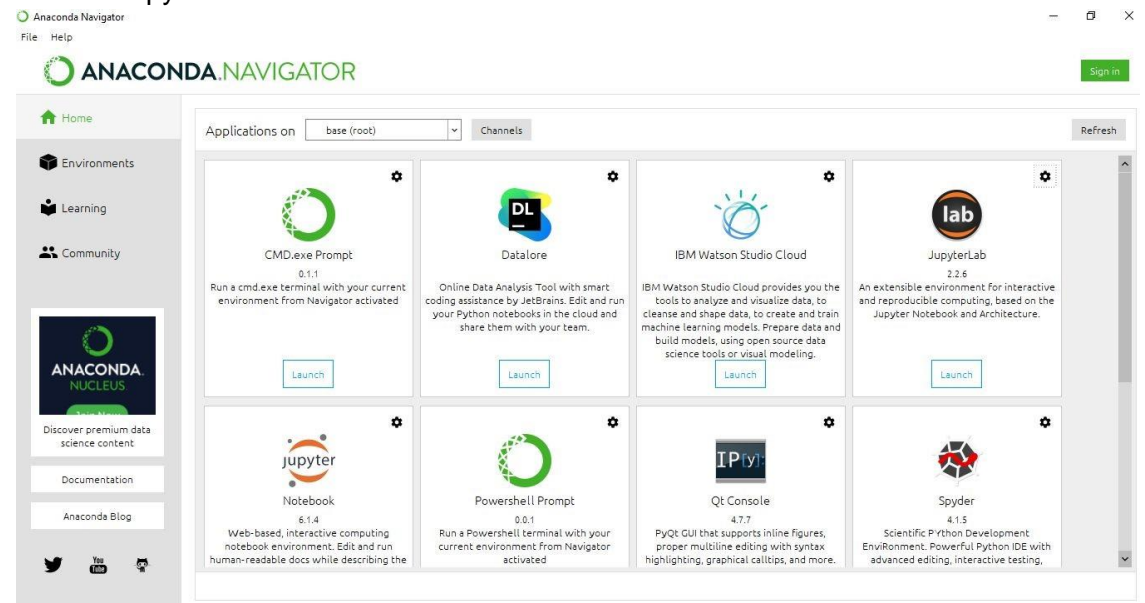
Con los pasos realizados anteriormente, completamos la instalación. Ahora procedemos a abrir el Jupyter.

1. Abrir Anaconda Navigator

## Spark - Jupyter - Scala

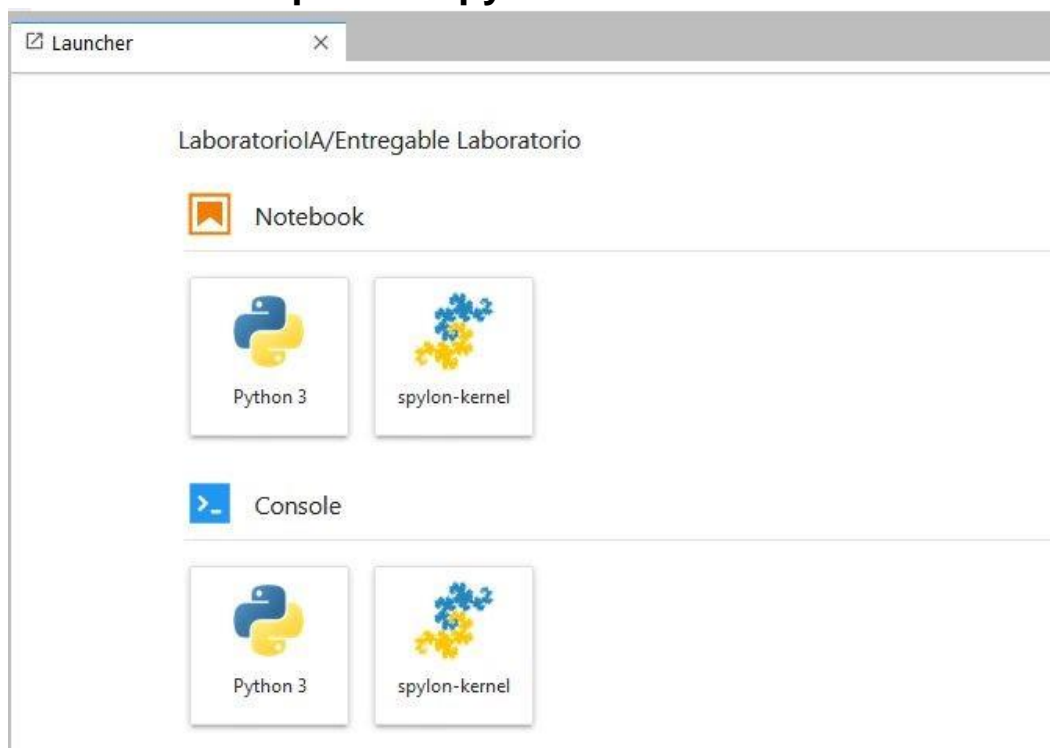


### 2. Abrimos JupyterLab o Notebook



### 3. Utilizamos el kernel instalado (Splyon-kernel)

## Spark - Jupyter - Scala



### 4. Validamos que funciona correctamente

```
[1]: spark

Intitializing Scala interpreter ...
Spark Web UI available at http://JuanArevaloM.mshome.net:4041
SparkContext available as 'sc' (version = 3.0.2, master = local[*], app id = local-1614230379827)
SparkSession available as 'spark'

[1]: res0: org.apache.spark.sql.SparkSession = org.apache.spark.sql.SparkSession@643c926e

[2]: val xValue = "X"

[2]: xValue: String = X

[ ]:
```