

Instituto Tecnológico de Costa Rica

Programa de Ciencia de los Datos

Curso: Big Data

Tarea #1 – Manual de Instalación

Ambiente de Trabajo para curso Big Data

Estudiante:

Natalia Rojas Canales

Profesor:

Felipe Meza

Noviembre 2019

Índice

Instalar Miniconda3 para Windows	1
Instalar JDK 8 en Windows 10.....	4
Instalar Spark	9
Crear ambiente de trabajo - BigData	13

Instalar Miniconda3 para Windows

Pasos de instalación:

1. Bajar el instalador desde: <https://docs.conda.io/en/latest/miniconda.html>
Debes seleccionar el instalador de acuerdo al sistema operativo de tu equipo, en mi caso es 64bits:

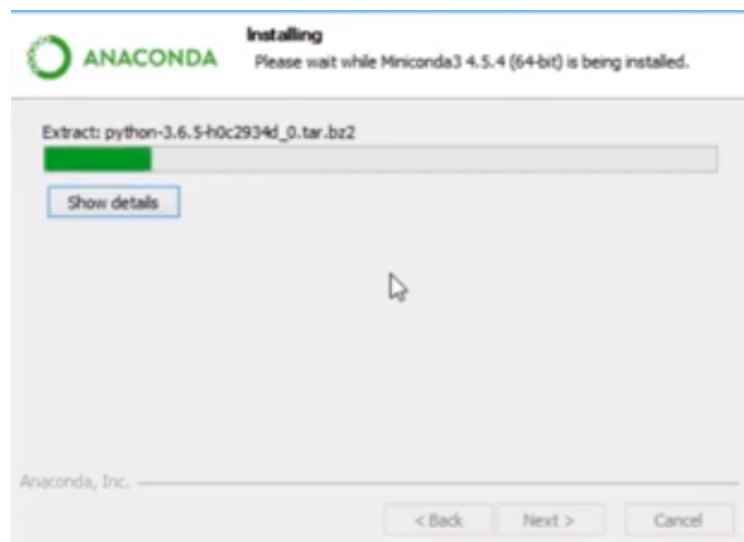
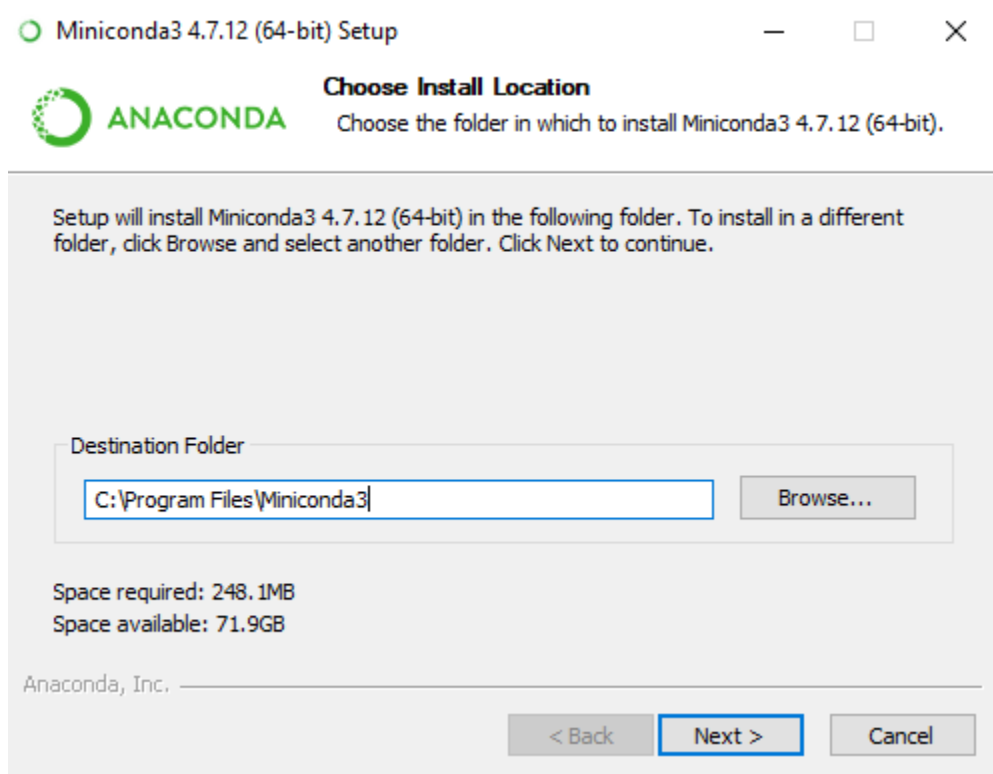
Windows installers

Windows			
Python version	Name	Size	SHA256 hash
Python 3.7	Miniconda3 Windows 64-bit	51.5 MiB	f18060cc0bb50ae75e4d602b7ce35197c8e31e81288d069b758594f1bb46ab45
	Miniconda3 Windows 32-bit	54.0 MiB	7c30778941d2bba03531ba269a78a108b01fa366530290376e7c3b467f3c66ba
Python 2.7	Miniconda2 Windows 64-bit	50.9 MiB	8647c54058f11842c37854edeff4d20bc1fbdad8b88d9d34d76fda1630e64846
	Miniconda2 Windows 32-bit	48.7 MiB	0d106228d6a4610b599df965dd6d9bb659329a17e3d693e3274b20291a7c6f94

2. Una vez que se bajo el instalador, se ejecuta, aparecerán las ventanas del wizard de instalación:



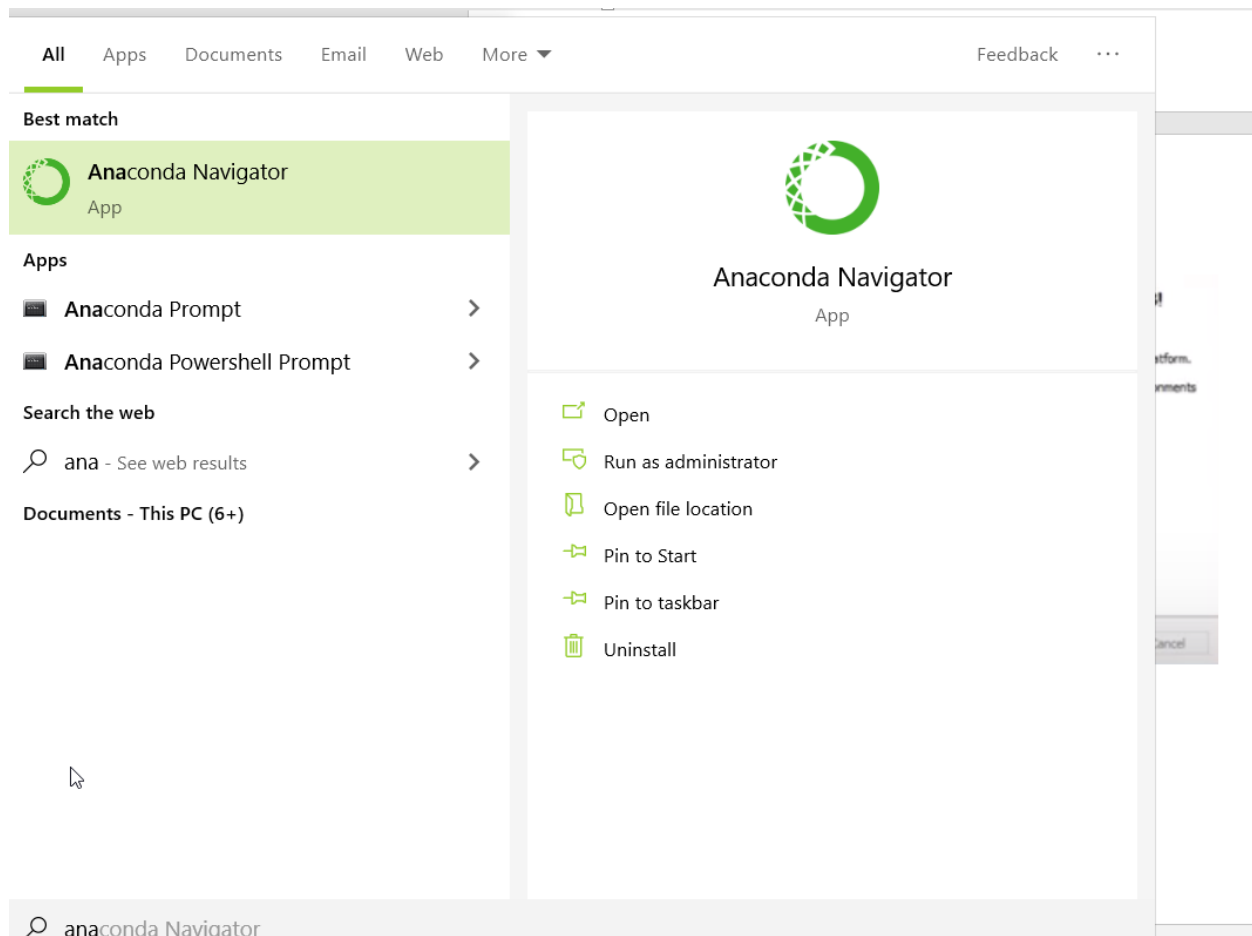
Van a dar Next hasta llegar a seleccionar el directorio de instalación, seleccionen la dirección con el botón Browse o bien, pueden dejar la que trae por defecto, en mi caso, seleccione: C:\Program Files\Miniconda3



3. Finaliza la instalación, se da click sobre el botón Finish:



Si vamos al menú de inicio y digitamos anaconda, nos mostrará lo que tenemos disponible:



Instalar JDK 8 en Windows 10

JDK es el kit de desarrollo Java, el mismo nos permite utilizar el lenguaje para programar sin necesidad de instalar un entorno de desarrollo, de modo que se puede programar desde consola. Para poder ejecutar esta instalación se ejecutan estos pasos:

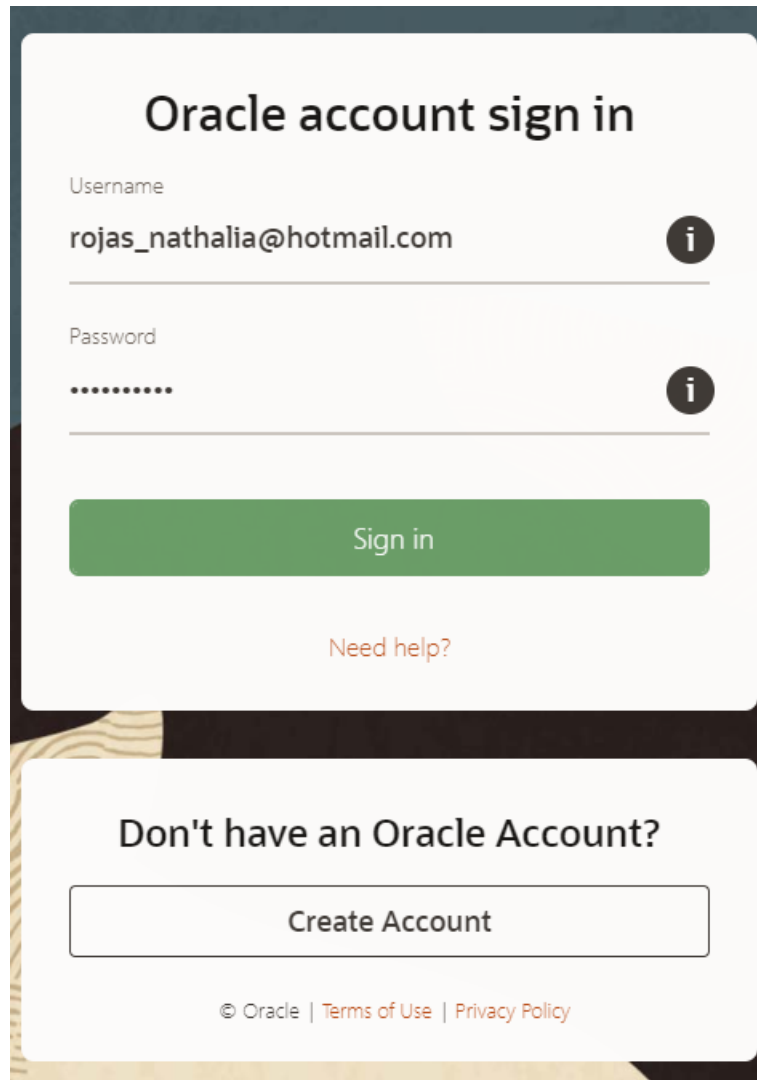
1. Ingresar a la dirección: <https://www.oracle.com/technetwork/java/javase/downloads/jdk8-downloads-2133151.html>

Voy a usar JDK 8 dado que la versión 13 le ha dado errores a algunos compañeros.

Java SE Development Kit 8u231		
You must accept the Oracle Technology Network License Agreement for Oracle Java SE to download this software.		
<input type="radio"/> Accept License Agreement <input checked="" type="radio"/> Decline License Agreement		
Product / File Description	File Size	Download
Linux ARM 32 Hard Float ABI	72.9 MB	jdk-8u231-linux-arm32-vfp-hflt.tar.gz
Linux ARM 64 Hard Float ABI	69.8 MB	jdk-8u231-linux-arm64-vfp-hflt.tar.gz
Linux x86	170.93 MB	jdk-8u231-linux-i586.rpm
Linux x86	185.75 MB	jdk-8u231-linux-i586.tar.gz
Linux x64	170.32 MB	jdk-8u231-linux-x64.rpm
Linux x64	185.16 MB	jdk-8u231-linux-x64.tar.gz
Mac OS X x64	253.4 MB	jdk-8u231-macosx-x64.dmg
Solaris SPARC 64-bit (SVR4 package)	132.98 MB	jdk-8u231-solaris-sparcv9.tar.Z
Solaris SPARC 64-bit	94.16 MB	jdk-8u231-solaris-sparcv9.tar.gz
Solaris x64 (SVR4 package)	133.73 MB	jdk-8u231-solaris-x64.tar.Z
Solaris x64	91.96 MB	jdk-8u231-solaris-x64.tar.gz
Windows x86	200.22 MB	jdk-8u231-windows-i586.exe
Windows x64	210.18 MB	jdk-8u231-windows-x64.exe

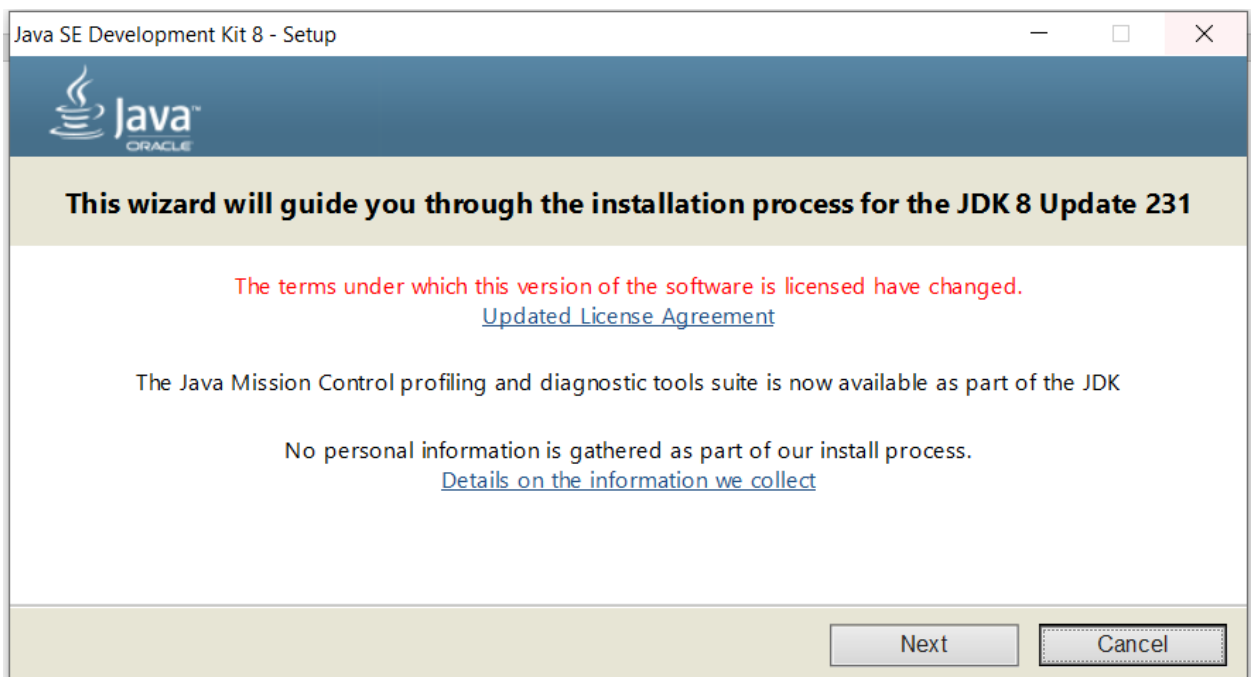
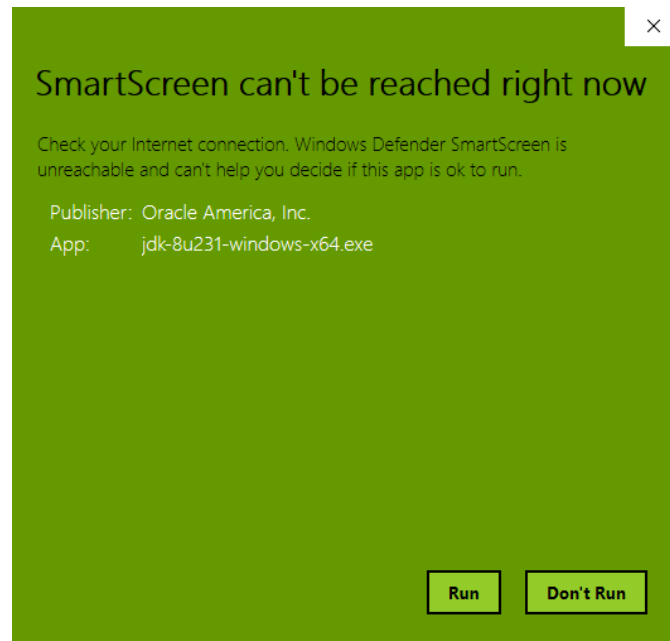
Antes de dar click para bajar la versión según el SO, debe aceptar la licencia.

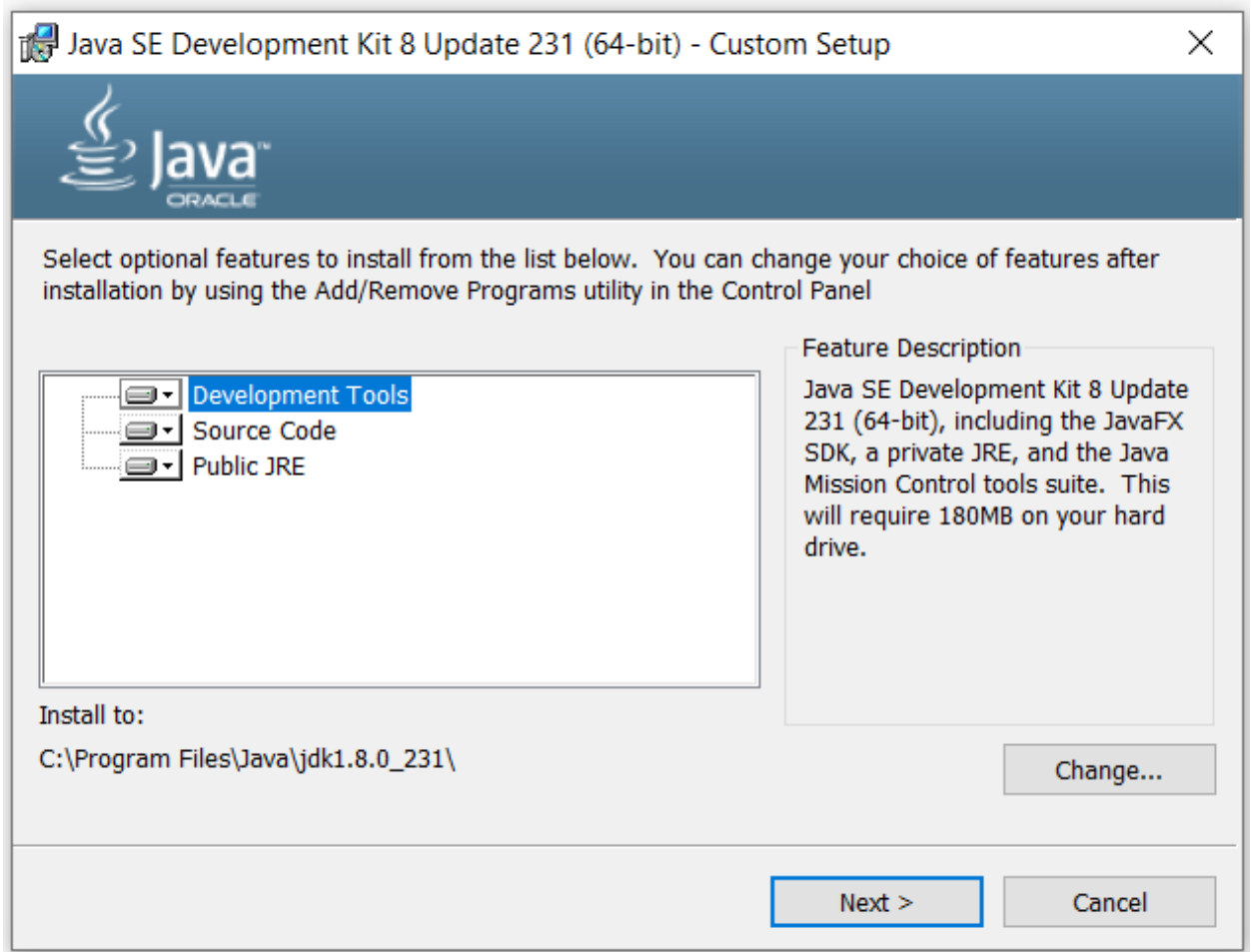
2. Le pedirá crear una cuenta o bien crear una:



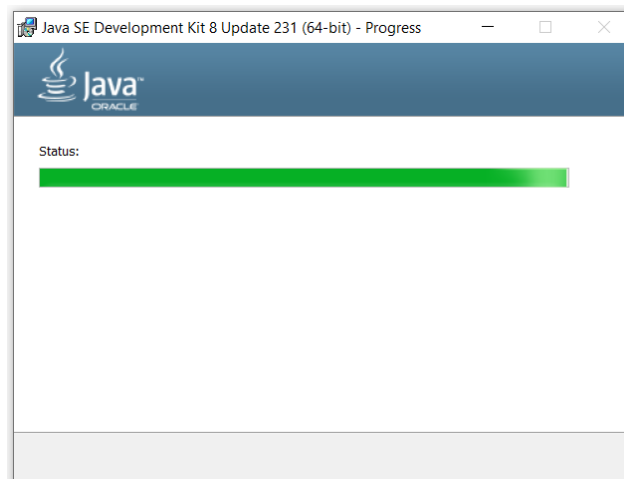
The screenshot displays the Oracle account sign in and creation interface. The top section, titled "Oracle account sign in", contains two input fields: "Username" with the value "rojas_nathalia@hotmail.com" and "Password" with masked characters ".....". Both fields have an information icon (i) to their right. Below these fields is a green "Sign in" button. Underneath the button is a link that says "Need help?". The bottom section, titled "Don't have an Oracle Account?", features a "Create Account" button. At the very bottom, there is a footer with the text "© Oracle | Terms of Use | Privacy Policy".

- Después de descargar el instalador, se procede a ejecutarlo:





Iniciará la instalación y se mostrará el avance:



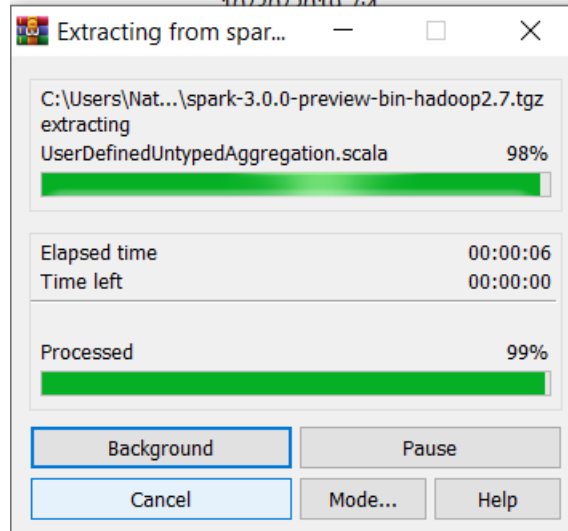
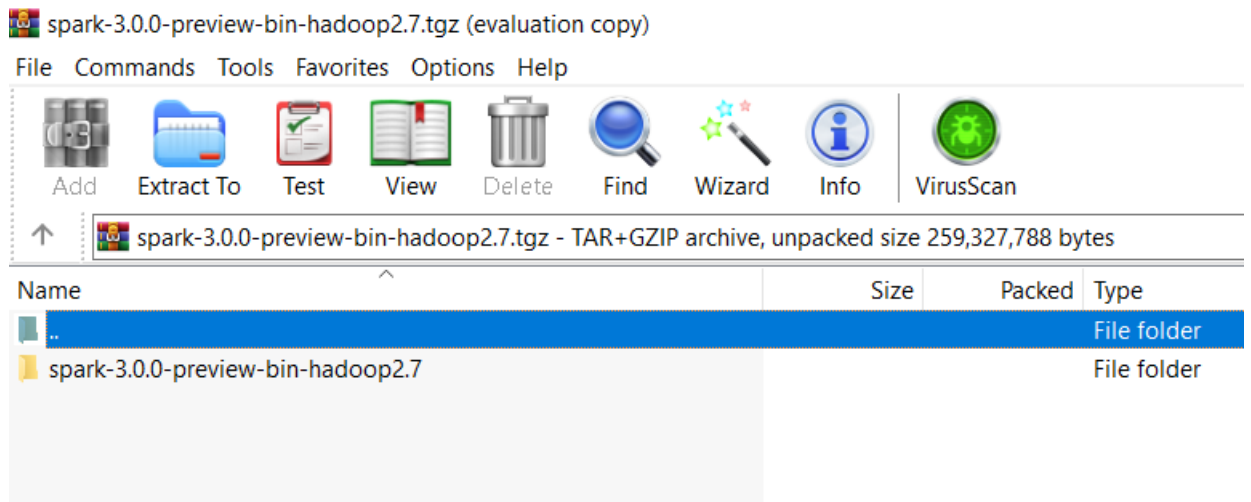


Instalar Spark

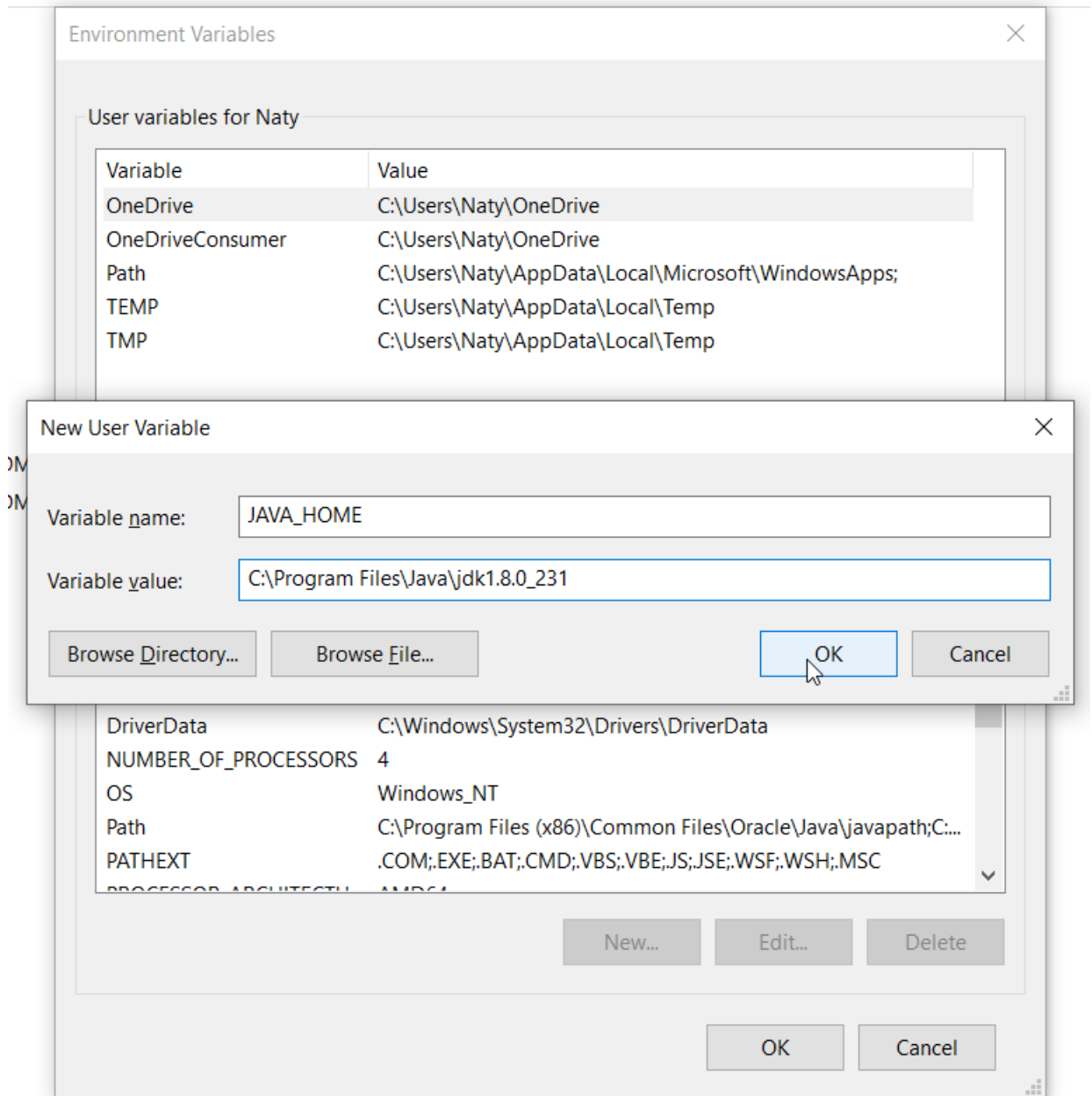
Para instalar Spark los pasos son muy sencillos:

Paso 1: Descargar Spark para Windows desde <https://spark.apache.org/downloads.html>

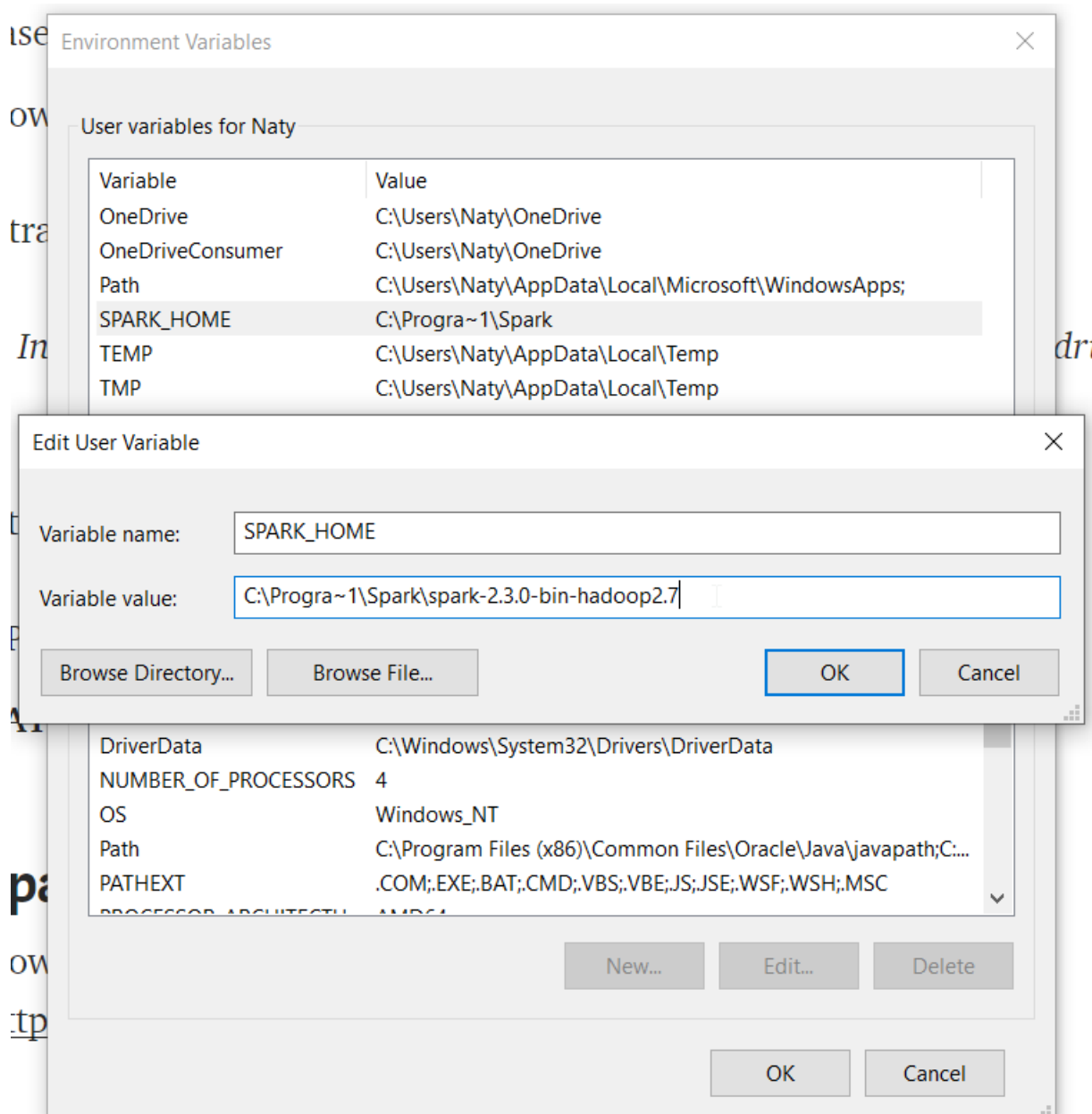
Paso 2: Descomprima el instalador.

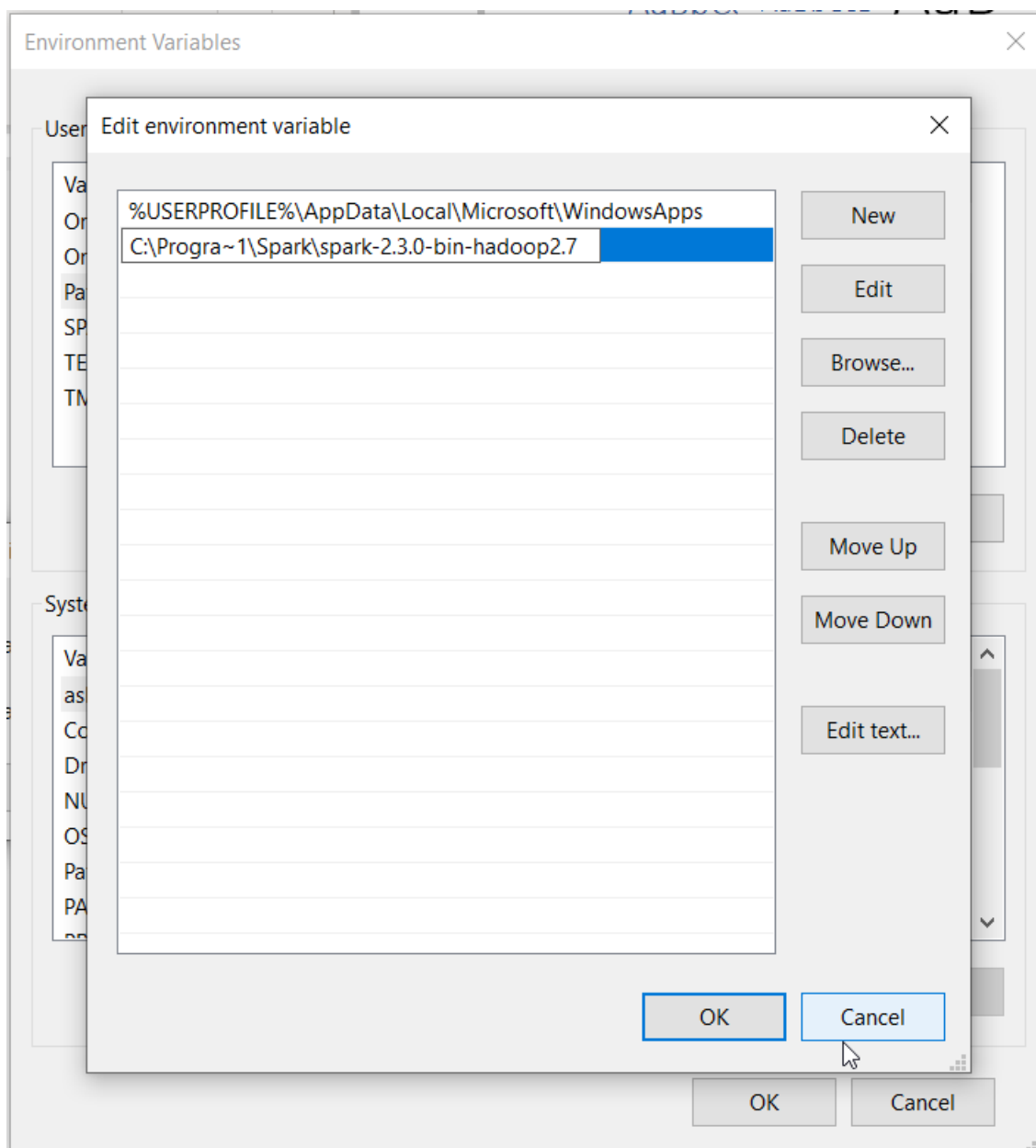


Paso 3: Setear las variables de entorno



Además debe setear las variables de SPARK





Crear ambiente de trabajo - BigData

Paso 1: Abrir el Anaconda Power Shell

Paso 2: ejecutar el comando `conda create --name BigData python=3`

Aparecerá la consulta que si realiza el proceso, se debe contestar **y** para continuar.

```
Administrator: Anaconda Powershell Prompt

setuptools-42.0.1 | py38_0 | 660 KB
sqlite-3.30.1 | he774522_0 | 962 KB
vc-14.1 | h0510ff6_4 | 6 KB
vs2015_runtime-14.16.27012 | hf0eaf9b_0 | 1.1 MB
wheel-0.33.6 | py38_0 | 53 KB
wincertstore-0.2 | py38_0 | 15 KB
-----
Total: 24.4 MB

The following NEW packages will be INSTALLED:

ca-certificates | pkgs/main/win-64::ca-certificates-2019.10.16-0
certifi | pkgs/main/win-64::certifi-2019.9.11-py38_0
openssl | pkgs/main/win-64::openssl-1.1.1d-he774522_3
pip | pkgs/main/win-64::pip-19.3.1-py38_0
python | pkgs/main/win-64::python-3.8.0-hff0d562_2
setuptools | pkgs/main/win-64::setuptools-42.0.1-py38_0
sqlite | pkgs/main/win-64::sqlite-3.30.1-he774522_0
vc | pkgs/main/win-64::vc-14.1-h0510ff6_4
vs2015_runtime | pkgs/main/win-64::vs2015_runtime-14.16.27012-hf0eaf9b_0
wheel | pkgs/main/win-64::wheel-0.33.6-py38_0
wincertstore | pkgs/main/win-64::wincertstore-0.2-py38_0

Proceed ([y]/n)? y

Downloading and Extracting Packages
setuptools-42.0.1 | 660 KB | ##### | 100%
python-3.8.0 | 19.6 MB | #####3 | 48%
```

Proceso finalizado

```
Administrator: Anaconda Powershell Prompt

wheel          pkgs/main/win-64::wheel-0.33.6-py38_0
wincertstore    pkgs/main/win-64::wincertstore-0.2-py38_0

Proceed ([y]/n)? y

Downloading and Extracting Packages
setuptools-42.0.1 | 660 KB | ##### | 100%
python-3.8.0      | 19.6 MB | ##### | 100%
sqlite-3.30.1     | 962 KB | ##### | 100%
certifi-2019.9.11 | 155 KB | ##### | 100%
pip-19.3.1        | 1.9 MB | ##### | 100%
vs2015_runtime-14.16 | 1.1 MB | ##### | 100%
wincertstore-0.2  | 15 KB | ##### | 100%
vc-14.1           | 6 KB | ##### | 100%
wheel-0.33.6      | 53 KB | ##### | 100%
Preparing transaction: done
Verifying transaction: done
Executing transaction: done
#
# To activate this environment, use
#
#     $ conda activate BigData
#
# To deactivate an active environment, use
#
#     $ conda deactivate
```

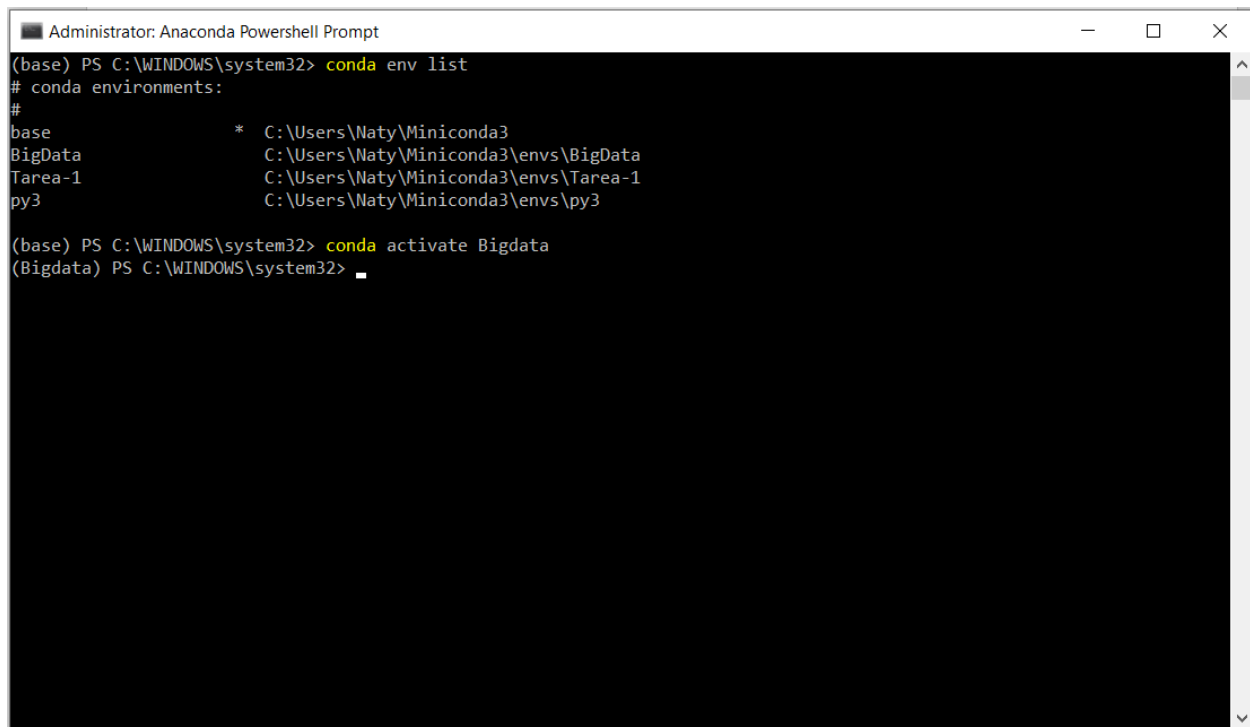
Paso 3: Para corroborar que fue creado se utiliza el comando `conda env list`

```
Administrator: Anaconda Powershell Prompt

(base) PS C:\WINDOWS\system32> conda env list
# conda environments:
#
base                * C:\Users\Naty\Miniconda3
BigData             C:\Users\Naty\Miniconda3\envs\BigData
Tarea-1             C:\Users\Naty\Miniconda3\envs\Tarea-1
py3                 C:\Users\Naty\Miniconda3\envs\py3

(base) PS C:\WINDOWS\system32>
```


Paso 4: Activar el ambiente creado: conda activate Bigdata

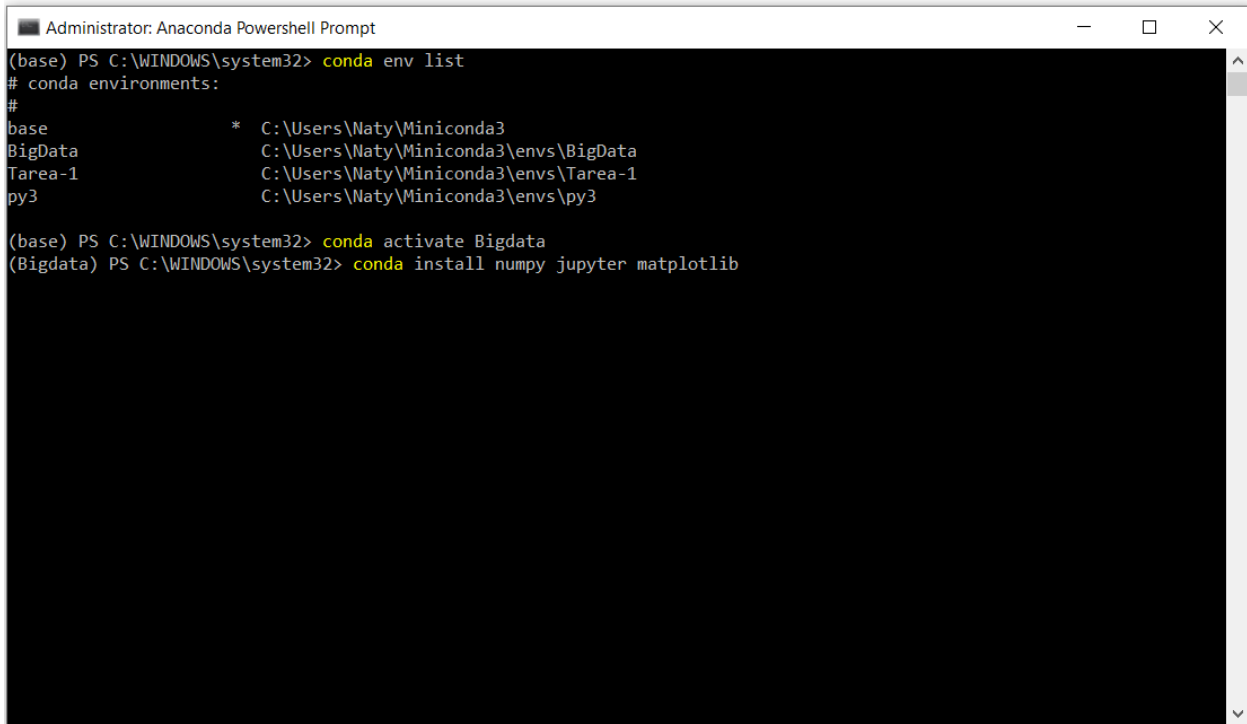


```
Administrator: Anaconda PowerShell Prompt
(base) PS C:\WINDOWS\system32> conda env list
# conda environments:
#
base                * C:\Users\Naty\Miniconda3
BigData             C:\Users\Naty\Miniconda3\envs\BigData
Tarea-1             C:\Users\Naty\Miniconda3\envs\Tarea-1
py3                 C:\Users\Naty\Miniconda3\envs\py3

(base) PS C:\WINDOWS\system32> conda activate Bigdata
(Bigdata) PS C:\WINDOWS\system32> _
```

Paso 5: Luego se realiza la instalación de las librerías necesarias:

`conda install numpy jupyter matplotlib`

A screenshot of an Anaconda PowerShell Prompt window. The window title is "Administrator: Anaconda PowerShell Prompt". The terminal shows the following commands and output:

```
(base) PS C:\WINDOWS\system32> conda env list
# conda environments:
#
base                * C:\Users\Naty\Miniconda3
BigData              C:\Users\Naty\Miniconda3\envs\BigData
Tarea-1              C:\Users\Naty\Miniconda3\envs\Tarea-1
py3                  C:\Users\Naty\Miniconda3\envs\py3

(base) PS C:\WINDOWS\system32> conda activate Bigdata
(Bigdata) PS C:\WINDOWS\system32> conda install numpy jupyter matplotlib
```

Preguntara si desea proceder: y para iniciar.

```
Administrator: Anaconda Powershell Prompt
wincertstore 0.2-py38_0 --> 0.2-py37_0

Proceed ([y]/n)? y

Downloading and Extracting Packages
testpath-0.4.4 | 88 KB | ##### | 100%
jupyter_console-6.0. | 52 KB | ##### | 100%
jupyter-1.0.0 | 6 KB | ##### | 100%
python-dateutil-2.8. | 224 KB | ##### | 100%
pywinpty-0.5.5 | 44 KB | ##### | 100%
qtconsole-4.6.0 | 95 KB | ##### | 100%
jpeg-9b | 245 KB | ##### | 100%
icu-58.2 | 9.4 MB | ##### | 100%
mistune-0.8.4 | 55 KB | ##### | 100%
pyzmq-18.1.0 | 400 KB | ##### | 100%
ipython_genutils-0.2 | 39 KB | ##### | 100%
mkl_random-1.1.0 | 232 KB | ##### | 100%
setuptools-42.0.1 | 692 KB | ##### | 100%
kiwisolver-1.1.0 | 53 KB | ##### | 100%
importlib_metadata-0 | 44 KB | ##### | 100%
zipp-0.6.0 | 9 KB | ##### | 100%
pyrsistent-0.15.6 | 95 KB | ##### | 100%
winpty-0.4.3 | 678 KB | ##### | 100%
intel-openmp-2019.4 | 1.4 MB | ##### | 100%
pandoc-2.2.3.2 | 14.3 MB | ##### | 100%
parso-0.5.1 | 68 KB | ##### | 100%
attrs-19.3.0 | 39 KB | ##### | 100%
notebook-6.0.2 | 7.7 MB | #5 | 2%
```

Se finaliza la instalación:

```
Administrator: Anaconda Powershell Prompt

entrypoints-0.3 | 12 KB | ##### | 100%
libsodium-1.0.16 | 416 KB | ##### | 100%
ipython-7.9.0 | 1.1 MB | ##### | 100%
prompt_toolkit-2.0.1 | 227 KB | ##### | 100%
jedi-0.15.1 | 704 KB | ##### | 100%
backcall-0.1.0 | 20 KB | ##### | 100%
numpy-base-1.17.3 | 4.8 MB | ##### | 100%
pyqt-5.9.2 | 3.3 MB | ##### | 100%
pickleshare-0.7.5 | 13 KB | ##### | 100%
libc_rt-2019.0.0 | 6.0 MB | ##### | 100%
pywin32-223 | 5.4 MB | ##### | 100%
bleach-3.1.0 | 218 KB | ##### | 100%
tornado-6.0.3 | 594 KB | ##### | 100%
jsonschema-3.2.0 | 112 KB | ##### | 100%
pyparsing-2.4.5 | 62 KB | ##### | 100%
jupyter_client-5.3.4 | 160 KB | ##### | 100%
blas-1.0 | 6 KB | ##### | 100%
matplotlib-3.1.1 | 4.9 MB | ##### | 100%
inbconvert-5.6.1 | 512 KB | ##### | 100%
six-1.13.0 | 27 KB | ##### | 100%
zeromq-4.3.1 | 10.6 MB | ##### | 100%
Preparing transaction: done
Verifying transaction: done
Executing transaction: / DEBUG menuinst_win32: __init__(199): Menu: name: 'Anaconda${PY_VER} ${PLATFORM}', prefix: 'C:\Users\Naty\Miniconda3\envs\Bigdata', env_name: 'Bigdata', mode: 'user', used_mode: 'user'
DEBUG menuinst_win32:create(323): Shortcut cmd is C:\Users\Naty\Miniconda3\python.exe, args are ['C:\Users\Naty\Miniconda3\cwp.py', 'C:\Users\Naty\Miniconda3\envs\Bigdata', 'C:\Users\Naty\Miniconda3\envs\Bigdata\python.exe', 'C:\Users\Naty\Miniconda3\envs\Bigdata\Scripts\jupyter-notebook-script.py', '%USERPROFILE%\']
done
(Bigdata) PS C:\WINDOWS\system32>
```

conda install pyspark

Igual que en los casos anteriores preguntara si procesa:

```
Administrator: Anaconda Powershell Prompt

environment location: C:\Users\Naty\Miniconda3\envs\Bigdata

added / updated specs:
- pyspark

The following packages will be downloaded:

package | build | size
-----|-----|-----
pandas-0.25.3 | py37ha925a31_0 | 9.8 MB
py4j-0.10.7 | py37_0 | 240 KB
pyspark-2.4.4 | py_0 | 205.0 MB
-----|-----|-----
Total: | 215.1 MB

The following NEW packages will be INSTALLED:

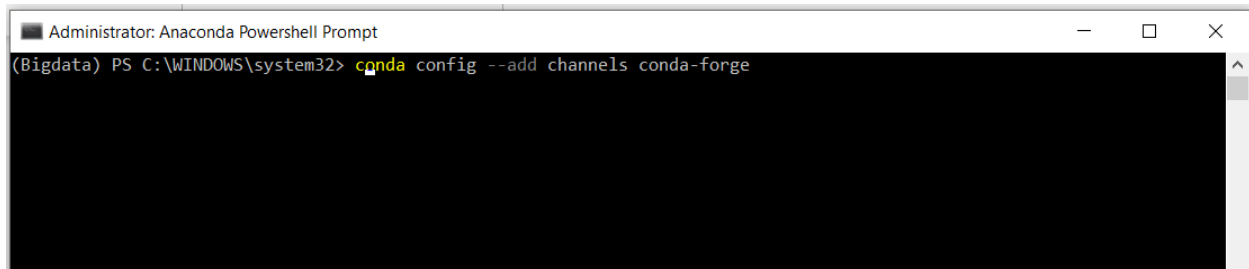
pandas pkgs/main/win-64::pandas-0.25.3-py37ha925a31_0
py4j pkgs/main/win-64::py4j-0.10.7-py37_0
pyspark pkgs/main/noarch::pyspark-2.4.4-py_0

Proceed ([y]/n)? y

Downloading and Extracting Packages
py4j-0.10.7 | 240 KB | ##### | 100%
pyspark-2.4.4 | 205.0 MB | #####8 | 14%
```

conda install findspark

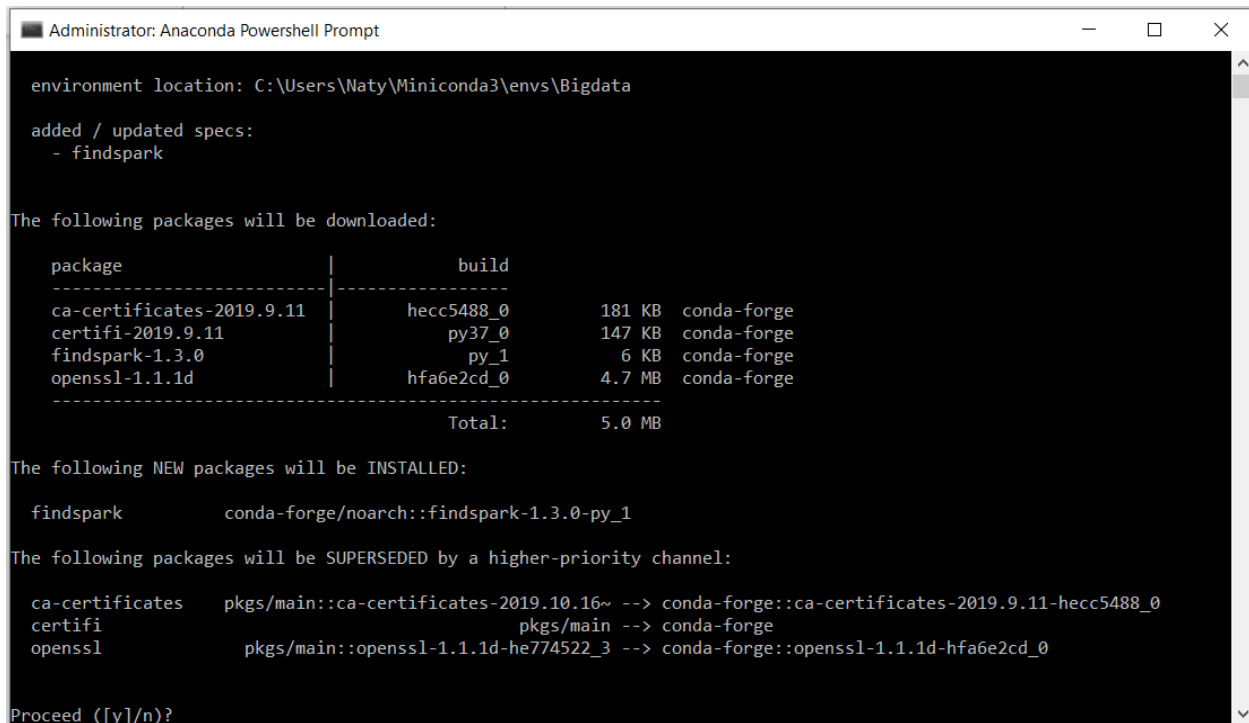
Primero se debe instalar un canal : **conda config --add channels conda-forge**



```
Administrator: Anaconda PowerShell Prompt
(Bigdata) PS C:\WINDOWS\system32> conda config --add channels conda-forge
```

Luego se ejecuta el comando de instalación: **conda install findspark**

Y volverá a preguntar si desea continuar:



```
Administrator: Anaconda PowerShell Prompt

environment location: C:\Users\Naty\Miniconda3\envs\Bigdata

added / updated specs:
- findspark

The following packages will be downloaded:

package | build | size | channel
-----|-----|-----|-----
ca-certificates-2019.9.11 | hecc5488_0 | 181 KB | conda-forge
certifi-2019.9.11 | py37_0 | 147 KB | conda-forge
findspark-1.3.0 | py_1 | 6 KB | conda-forge
openssl-1.1.1d | hfa6e2cd_0 | 4.7 MB | conda-forge
-----|-----|-----|-----
Total: | 5.0 MB |

The following NEW packages will be INSTALLED:

findspark conda-forge/noarch::findspark-1.3.0-py_1

The following packages will be SUPERSEDED by a higher-priority channel:

ca-certificates pkgs/main::ca-certificates-2019.10.16~ --> conda-forge::ca-certificates-2019.9.11-hecc5488_0
certifi pkgs/main --> conda-forge
openssl pkgs/main::openssl-1.1.1d-he774522_3 --> conda-forge::openssl-1.1.1d-hfa6e2cd_0

Proceed ([y]/n)?
```

Hasta finalizar:

```
Administrator: Anaconda Powershell Prompt

ca-certificates-2019.9.11 | hecc5488_0 | 181 KB | conda-forge
certifi-2019.9.11 | py37_0 | 147 KB | conda-forge
findspark-1.3.0 | py_1 | 6 KB | conda-forge
openssl-1.1.1d | hfa6e2cd_0 | 4.7 MB | conda-forge
-----
Total: 5.0 MB

The following NEW packages will be INSTALLED:

findspark          conda-forge/noarch::findspark-1.3.0-py_1

The following packages will be SUPERSEDED by a higher-priority channel:

ca-certificates    pkgs/main::ca-certificates-2019.10.16~ --> conda-forge::ca-certificates-2019.9.11-hecc5488_0
certifi            pkgs/main --> conda-forge
openssl            pkgs/main::openssl-1.1.1d-he774522_3 --> conda-forge::openssl-1.1.1d-hfa6e2cd_0

Proceed ([y]/n)? y

Downloading and Extracting Packages
certifi-2019.9.11 | 147 KB | ##### | 100%
openssl-1.1.1d | 4.7 MB | ##### | 100%
ca-certificates-2019 | 181 KB | ##### | 100%
findspark-1.3.0 | 6 KB | ##### | 100%
Preparing transaction: done
Verifying transaction: done
Executing transaction: done
(Bigdata) PS C:\WINDOWS\system32>
```

Y ESTAMOS LISTOS
PARA TRABAJAR