



Tarea 1

Manual de Instalación de Spark

Profesor: Felipe Meza

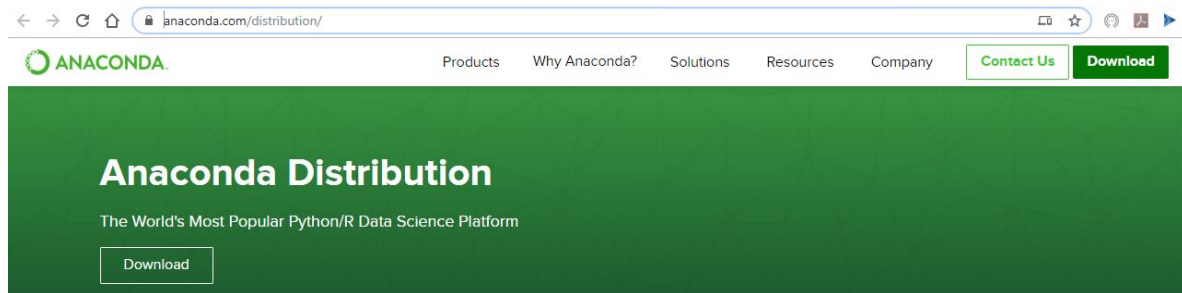
Alumno:

Lester Salazar Viales

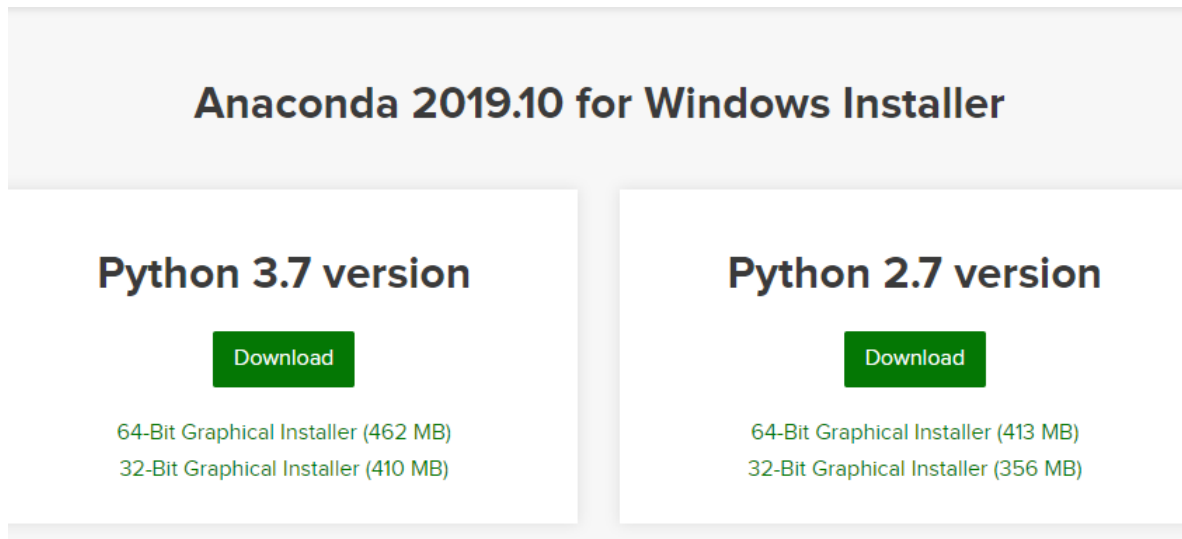
Instalación Conda:

Ingresa al sitio


<https://www.anaconda.com/distribution/>



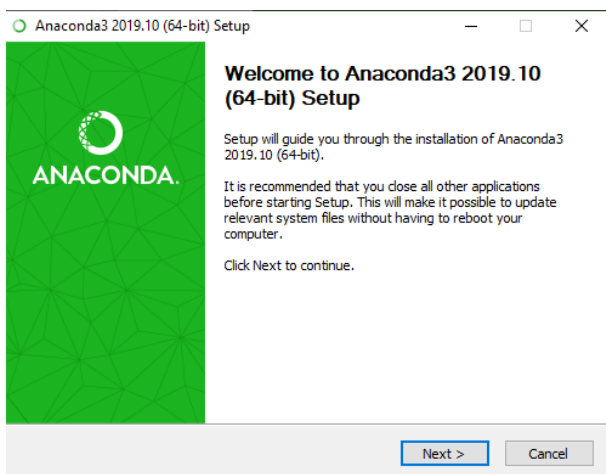
Presionar download que los lleva y muestra la siguiente parte de la pagina



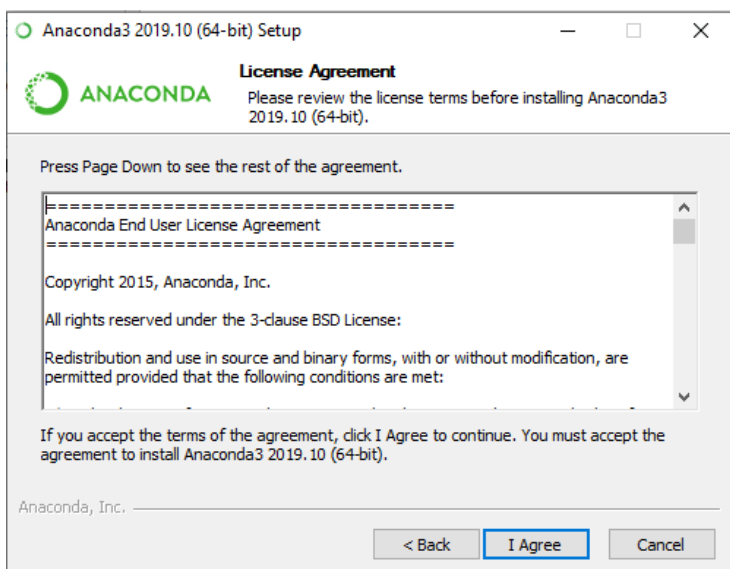
Descargar la versión para Python 3.7 y una vez descargada dar doble click sobre el archivo de instalación para que comience la misma

 Anaconda3-2019.10-Windows-x86_64

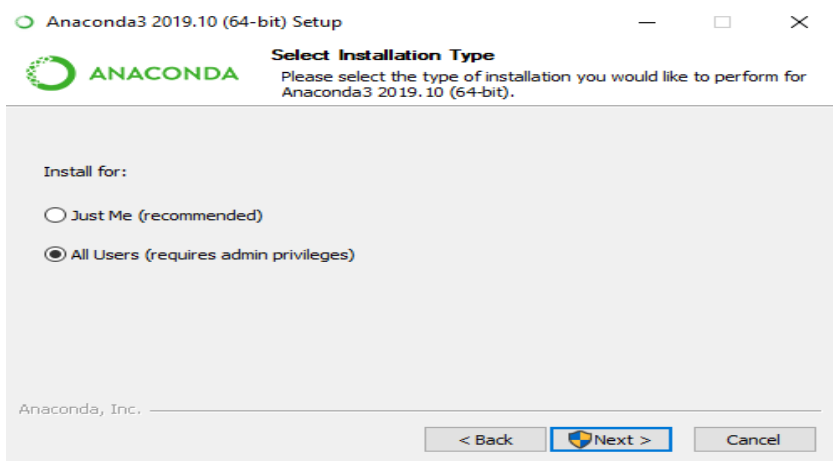
Y presionar Next en la pantalla de inicio de instalación



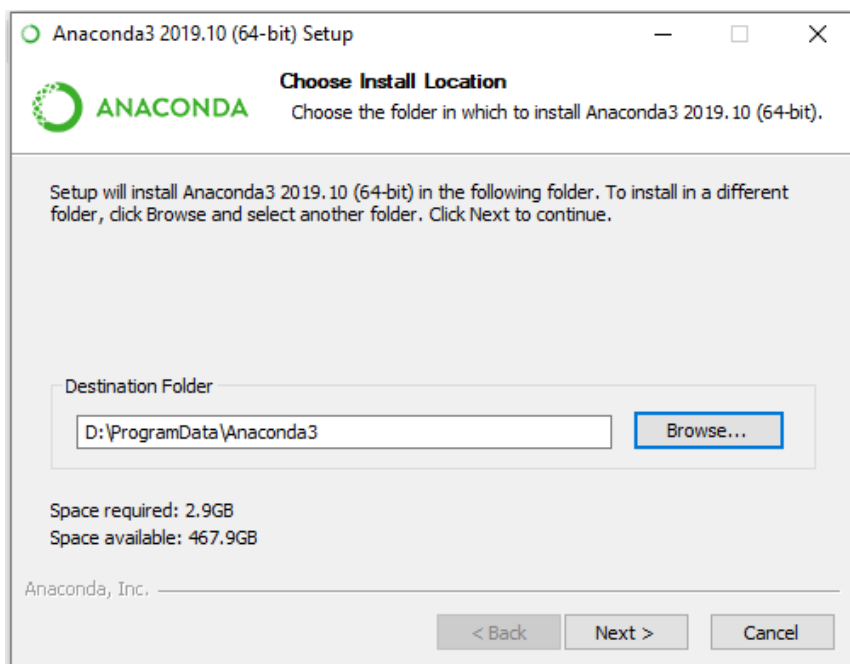
Se debe de aceptar los términos de la licencia presionando el botón que dice I agree



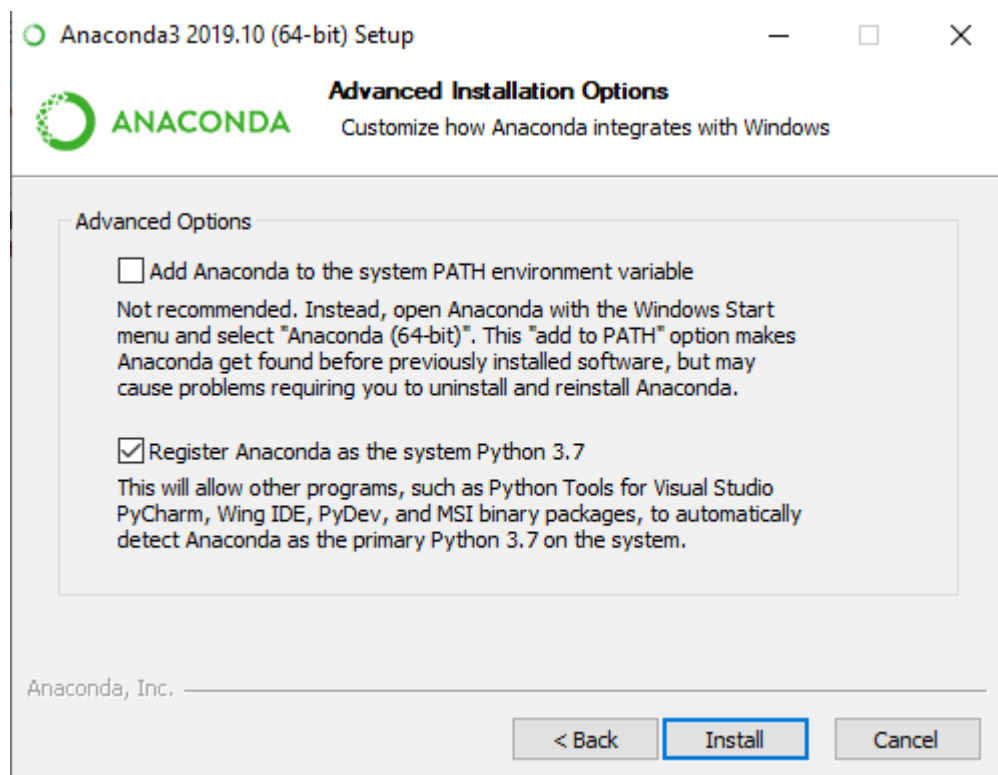
Y luego seleccionar que la instalación se realice para todos los usuarios y presionar Next

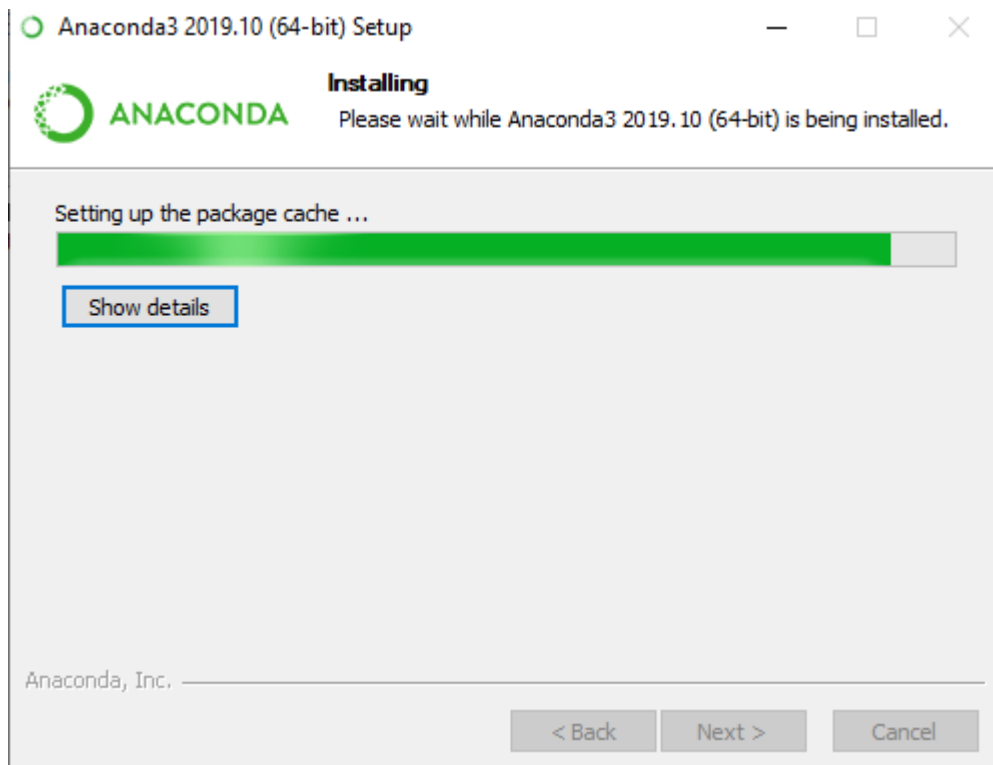


Presionar Next en la pantalla similar a la siguiente

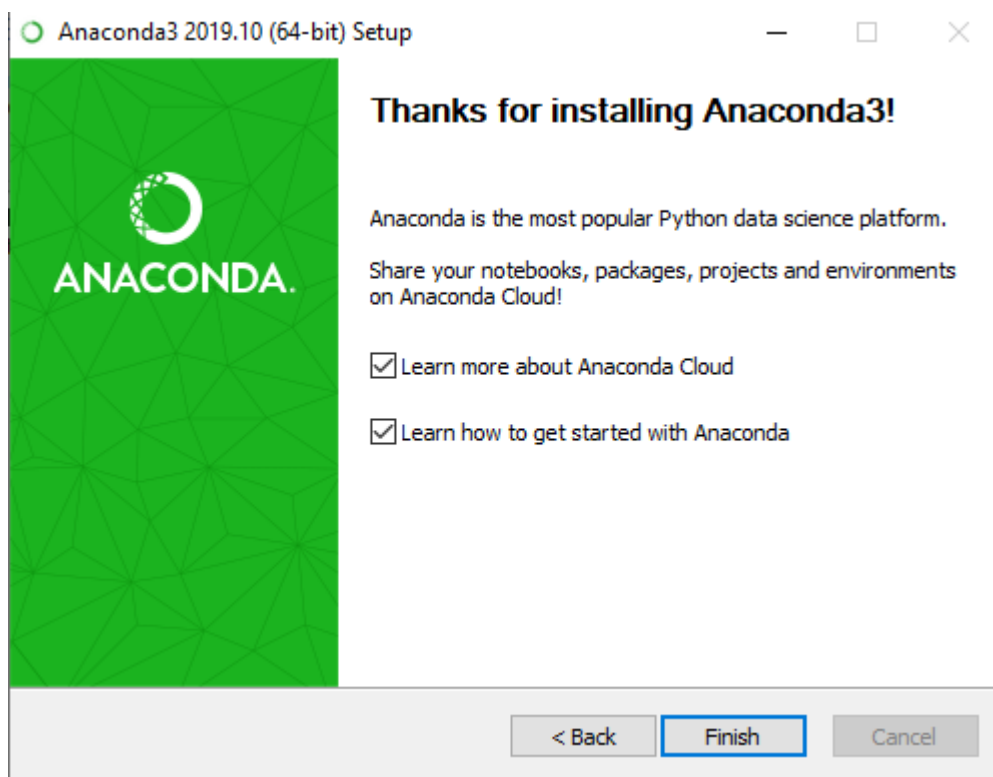


Una vez realizado el paso anterior se muestra la siguiente pantalla donde se debe de dejar con los valores que trae por defecto y se presiona Install para que comience la instalación.





Una vez terminada la instalación se presiona el botón que dice Finish



Instalación de Java

<https://www.java.com/es/download/>

Descarga gratuita de Java

Descargue Java para su computadora de escritorio ahora

Version 8 Update 231

Fecha de versión: 15 de octubre de 2019



Actualización importante de la licencia de Oracle Java

La licencia de Oracle Java ha cambiado para las versiones publicadas a partir del 16 de abril de 2019.

El nuevo [acuerdo de licencia de Oracle Technology Network para Oracle Java SE](#) es sustancialmente diferente a las licencias de Oracle Java anteriores. La nueva licencia permite ciertos usos, como el uso personal y de desarrollo, sin coste alguno (aunque podría haber otros usos autorizados en licencias de Oracle Java anteriores que ya no estén disponibles). Revise las condiciones con atención antes de descargar y utilizar este producto. Puede consultar las preguntas frecuentes [aquí](#).

La licencia comercial y el soporte están disponibles con una [suscripción de Java SE](#) de bajo coste.

Oracle también ofrece la última versión de OpenJDK con la [licencia pública general](#) de código abierto en jdk.java.net.

Descarga gratuita de Java

Descargar Java para Windows

Recomendado Version 8 Update 231 (Tamaño de archivo: 1.97 MB)

Fecha de versión: 15 de octubre de 2019



Actualización importante de la licencia de Oracle Java

La licencia de Oracle Java ha cambiado para las versiones publicadas a partir del 16 de abril de 2019.

El nuevo [acuerdo de licencia de Oracle Technology Network para Oracle Java SE](#) es sustancialmente diferente a las licencias de Oracle Java anteriores. La nueva licencia permite ciertos usos, como el uso personal y de desarrollo, sin coste alguno (aunque podría haber otros usos autorizados en licencias de Oracle Java anteriores que ya no estén disponibles). Revise las condiciones con atención antes de descargar y utilizar este producto. Puede consultar las preguntas frecuentes [aquí](#).

La licencia comercial y el soporte están disponibles con una [suscripción de Java SE](#) de bajo coste.

Oracle también ofrece la última versión de OpenJDK con la [licencia pública general](#) de código abierto en jdk.java.net.

Aceptar e iniciar descarga gratuita

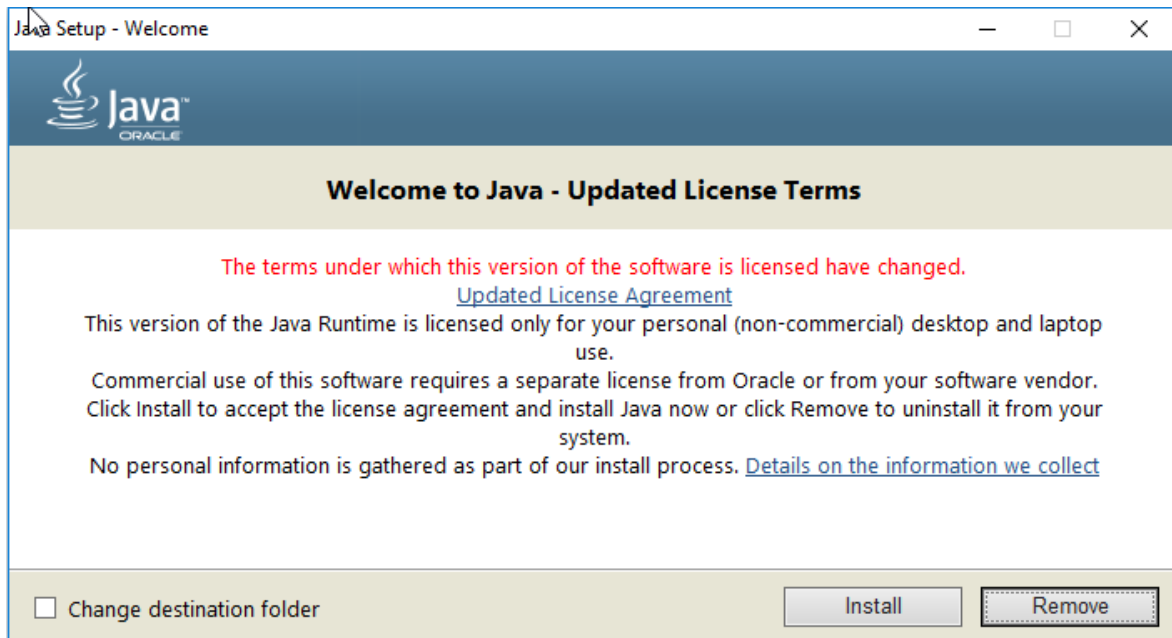
ejecutar



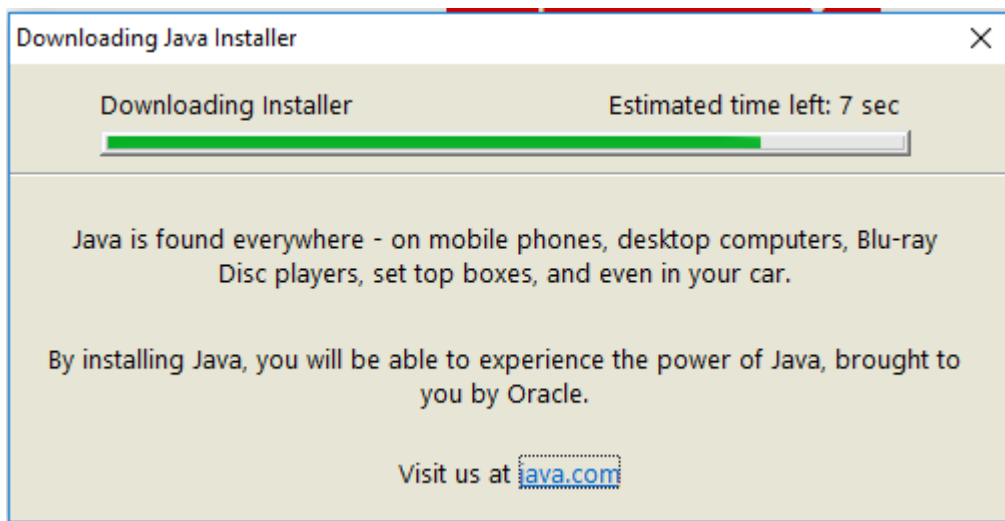
JavaSetup8u231.exe

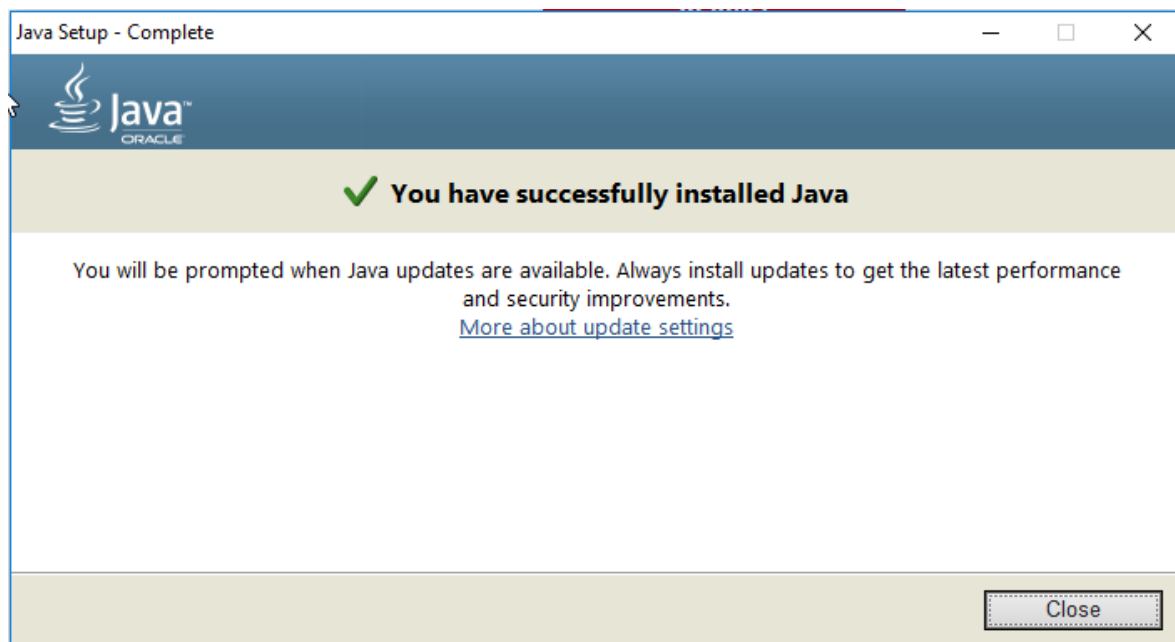


Presionar Yes



Presionar Install





Instalación SDK de Java

Ingresa al siguiente URL

<https://www.oracle.com/technetwork/es/java/javase/downloads/index.html>

se muestra la siguiente pantalla y se debe seleccionar SDK

Descargas de Java SE



Java Platform (JDK) 8u111 / 8u112



NetBeans con JDK 8

Java Platform, Standard Edition	
<p>Java SE 8u111 / 8u112</p> <p>Java SE 8u111 incluye importantes soluciones de seguridad. Oracle recomienda que todos los usuarios de Java SE 8 actualicen esta versión. Java SE 8u112 es una actualización con un conjunto de parches, que incluye todas las características adicionales de 8u111 (descritas en las notas de la versión).</p> <p>Lea más aquí (en inglés) ▶</p>	
<div><p>Importante cambio planificado para MD5-signed JARs</p><p>A partir de las versiones de la revisión crítica de abril, previstas para el 18 de abril de 2017, todas las versiones de JRE tratarán a los JARs firmados con MD5 como no firmado. Obtenga más información y vea las instrucciones de prueba. (en inglés)</p><p>Para obtener más información sobre el soporte del algoritmo criptográfico, por favor chequee este documento: JRE and JDK Crypto Roadmap. (en inglés)</p></div>	
<ul style="list-style-type: none">Instrucciones de instalación (en inglés)Notas de la versión (en inglés)Licencia de Oracle (en inglés)Productos Java SE (en inglés)Licencias de terceros (en inglés)Configuraciones del sistema certificadas (en inglés) <p>Archivos LeaMe</p> <ul style="list-style-type: none">Archivo LeaMe JDK (en inglés)Archivo LeaMe JRE (en inglés)	<div><p>JDK</p><p>DOWNLOAD ↓</p></div> <div><p>Servidor JRE</p><p>DOWNLOAD ↓</p></div> <div><p>JRE</p><p>DOWNLOAD ↓</p></div>

Nos lleva a la siguiente pantalla

Java SE Development Kit 8u231

You must accept the [Oracle Technology Network License Agreement for Oracle Java SE](#) to download this software.

Thank you for accepting the Oracle Technology Network License Agreement for Oracle Java SE; you may now download this software.

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

Ingresar la cuenta de Oracle

Oracle account sign in

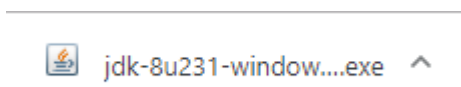
Username

Password

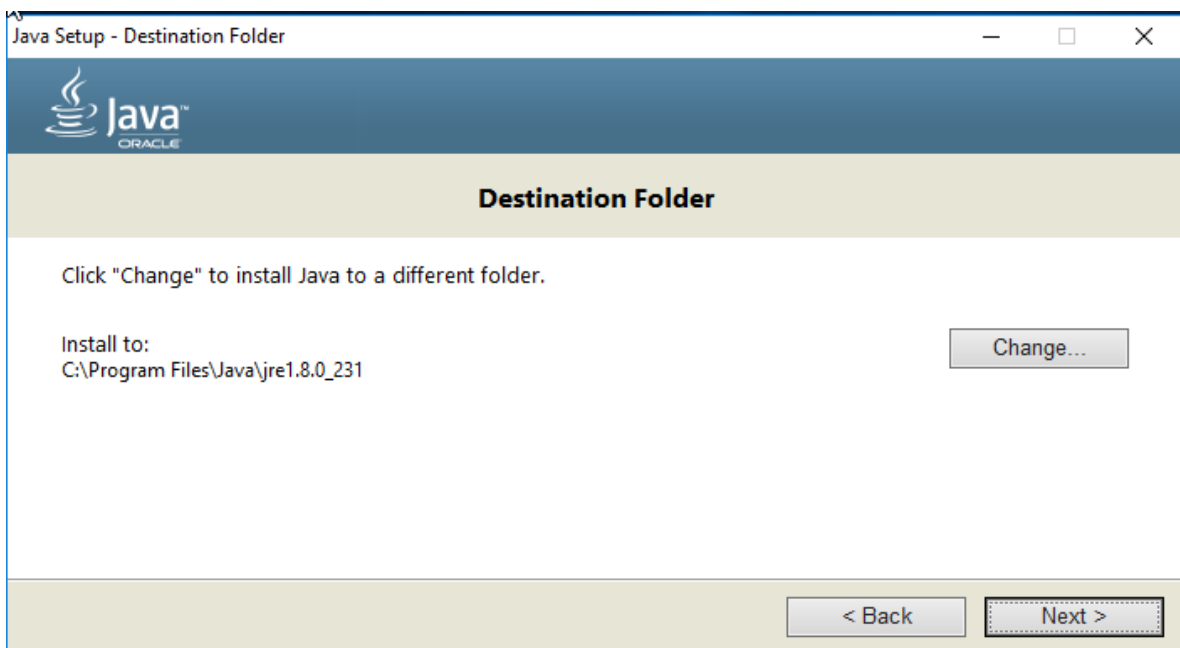
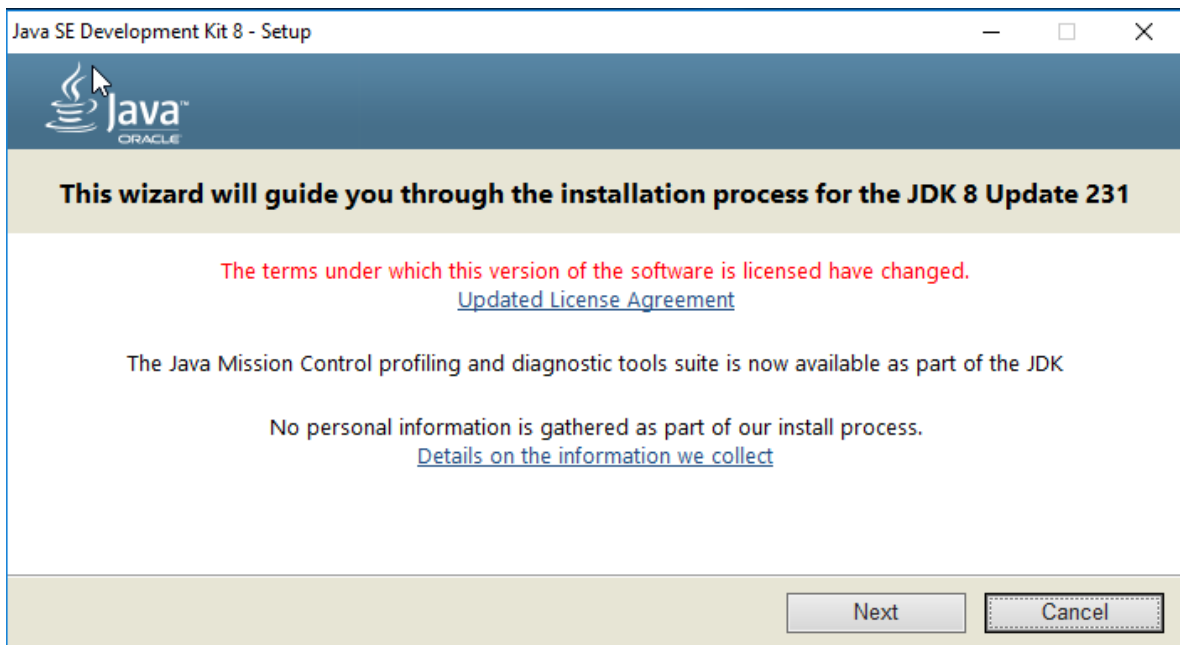
Invalid username and/or password.

Sign in

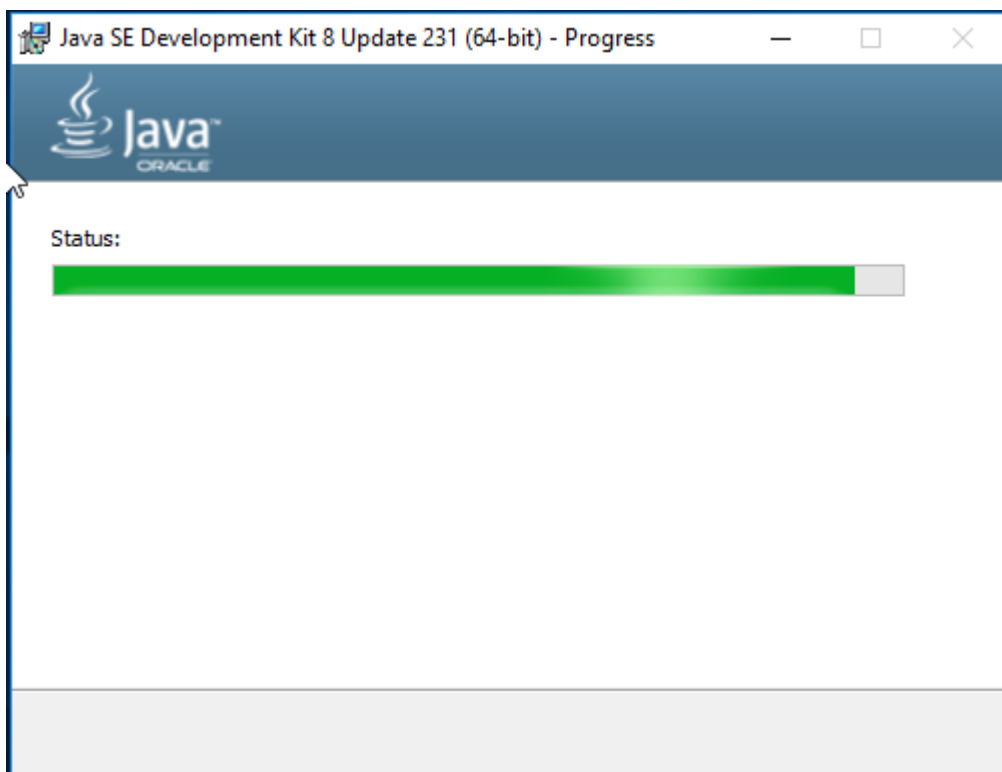
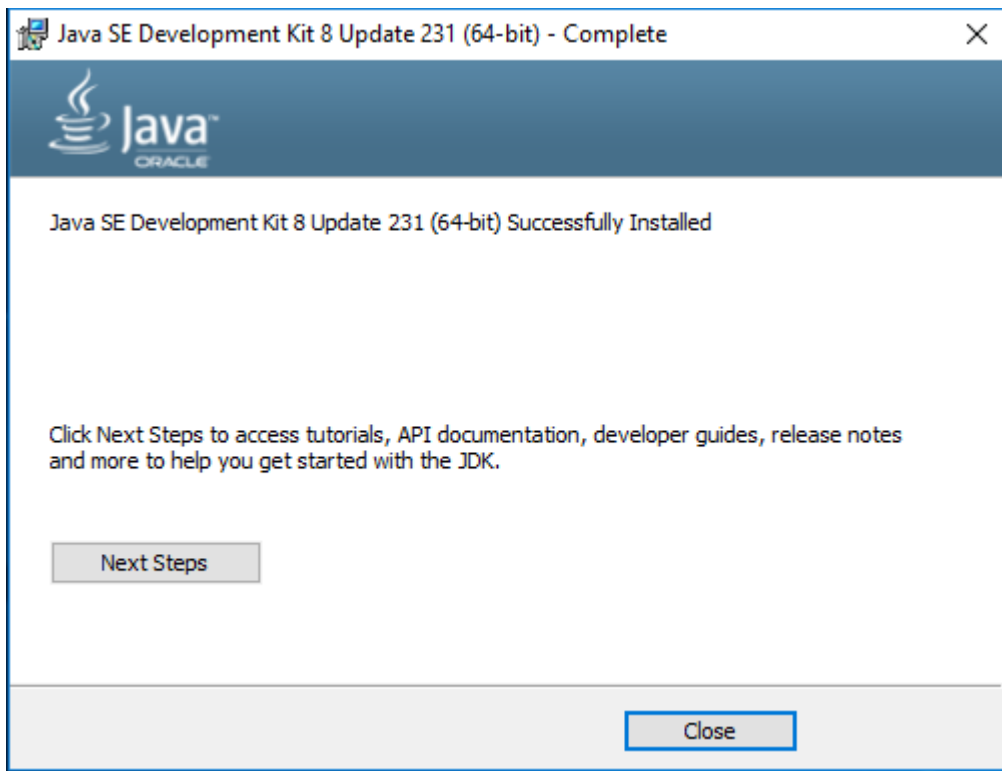
Ejecutar el siguiente instalador



Presionar next en la pantalla siguiente.



En la siguiente pantalla se presiona close



Creacion de ambiente de trabajo.

Se ingresa a la consola de ejecución de Anaconda y se ingresa el siguiente comando

```
Anaconda Prompt (Anaconda3)
```

```
(base) C:\Users\lmsv0>conda create -n BigData python=3.74
```

Y se presiona enter, lo que nos muestra la siguiente pantalla donde debemos de presionar y para que se realice la creación del ambiente y la instalación de aplicaciones básicas en dicho ambiente

```
Anaconda Prompt (Anaconda3) - conda create -n BigData python=3.7
```

The following packages will be downloaded:

package	build	
ca-certificates-2019.10.16	0	163 KB
pip-19.3.1	py37_0	1.9 MB
python-3.7.5	h8c8aaf0_0	14.7 MB
setuptools-42.0.1	py37_0	692 KB
sqlite-3.30.1	he774522_0	962 KB
Total:		18.4 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-py37_0
openssl	pkgs/main/win-64::openssl-1.1.1d-he774522_3
pip	pkgs/main/win-64::pip-19.3.1-py37_0
python	pkgs/main/win-64::python-3.7.5-h8c8aaf0_0
setuptools	pkgs/main/win-64::setuptools-42.0.1-py37_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-py37_0
wincertstore	pkgs/main/win-64::wincertstore-0.2-py37_0

Proceed ([y]/n)?

```
Anaconda Prompt (Anaconda3)

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-py37_0
wincertstore pkgs/main/win-64::wincertstore-0.2-py37_0

Proceed ([y]/n)? y

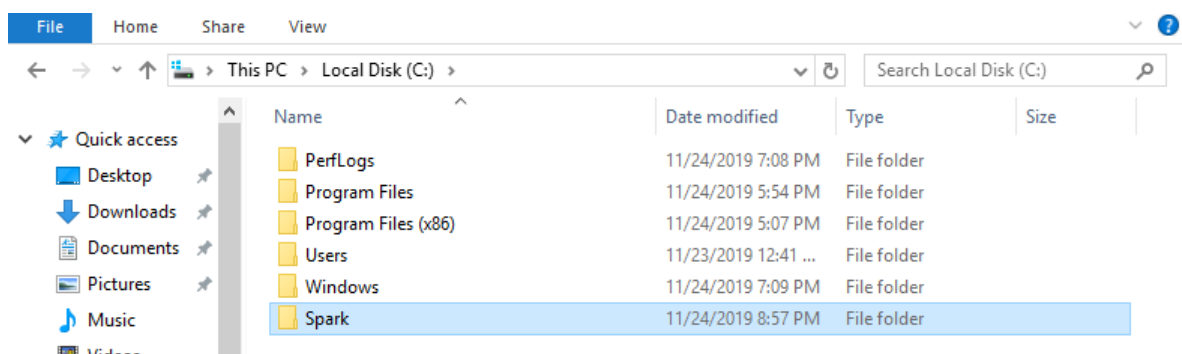
Downloading and Extracting Packages
sqlite-3.30.1      | 962 KB | ##### | 100%
python-3.7.5      | 14.7 MB | ##### | 100%
ca-certificates-2019 | 163 KB | ##### | 100%
pip-19.3.1       | 1.9 MB | ##### | 100%
setuptools-42.0.1 | 692 KB | ##### | 100%
Preparing transaction: done
Verifying transaction: done
Executing transaction: done
#
# To activate this environment, use
#
#   $ conda activate BigData
#
# To deactivate an active environment, use
#
#   $ conda deactivate
#
(base) C:\Users\lmsv0>
```

Se procede a ingresar al ambiente de trabajo utilizando el siguiente comando y presionando enter

```
(base) C:\Users\lmsv0>conda activate BigData
(BigData) C:\Users\lmsv0>
```

Instalación de SPARK

En la unidad C, creamos una carpeta que se llame SPARK



Ir al siguiente url:

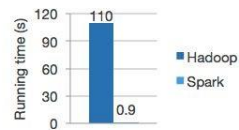
<https://spark.apache.org/>

Apache Spark™ is a unified analytics engine for large-scale data processing.

Speed

Run workloads 100x faster.

Apache Spark achieves high performance for both batch and streaming data, using a state-of-the-art DAG scheduler, a query optimizer, and a physical execution engine.



Logistic regression in Hadoop and Spark

Presionar para descargar

```
df = spark.read.json("logs.json")
df.where("age > 21")
  .select("name.first").show()
```

Spark's Python DataFrame API
Read JSON files with automatic schema inference

Latest News

Preview release of Spark 3.0 (Nov 06, 2019)
Spark 2.3.4 released (Sep 09, 2019)
Spark 2.4.4 released (Sep 01, 2019)
Plan for dropping Python 2 support (Jun 03, 2019)

[Archive](#)



Download Spark

Ease of Use

Write applications quickly in Java, Scala, Python, R, and SQL.

Spark offers over 80 high-level operators that make it easy to build parallel apps. And you can use it *interactively* from the Scala, Python, R, and SQL shells.

Built-in Libraries:

[SQL and DataFrames](#)
[Spark Streaming](#)
[MLlib \(machine learning\)](#)
[GraphX \(graph\)](#)
[Third-Party Projects](#)

Al presionar el botón de download nos lleva a la siguiente pantalla donde procederemos a descargar el framework de Spark



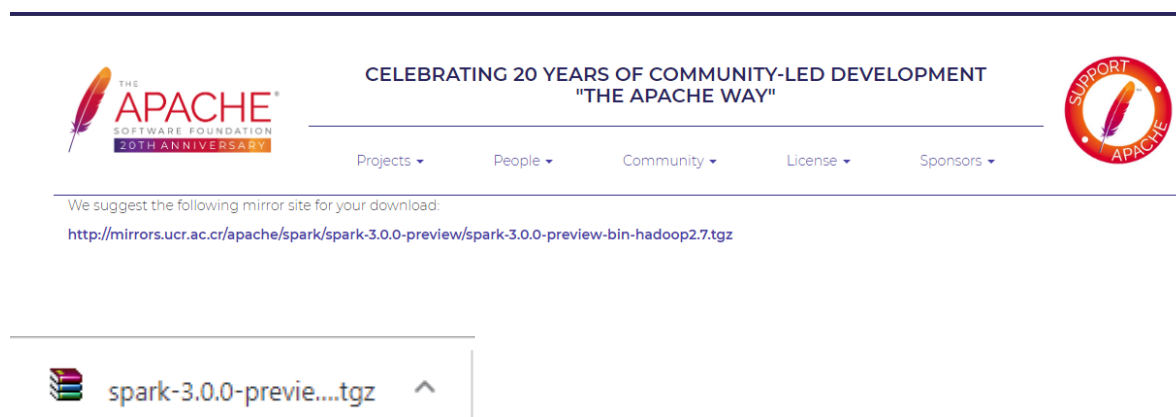
Download Apache Spark™

1. Choose a Spark release:
2. Choose a package type:
3. Download Spark: [spark-3.0.0-preview-bin-hadoop2.7.tgz](#)
4. Verify this release using the 3.0.0-preview [signatures](#), [checksums](#) and [project release KEYS](#).

Presionar este link para descargar Spark

Note that, Spark is pre-built with Scala 2.11 except version 2.4.2, which is pre-built with Scala 2.12.

El link anterior nos lleva a la siguiente donde se selecciona el link recomendado para que comience la descarga

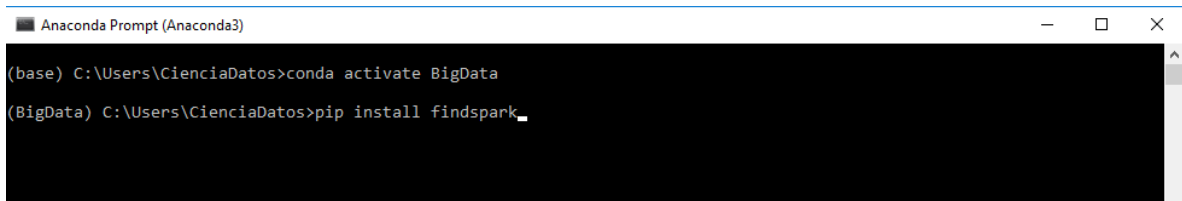


Se copia y se descomprime el archivo en la carpeta Spark y se crea la variable de entorno

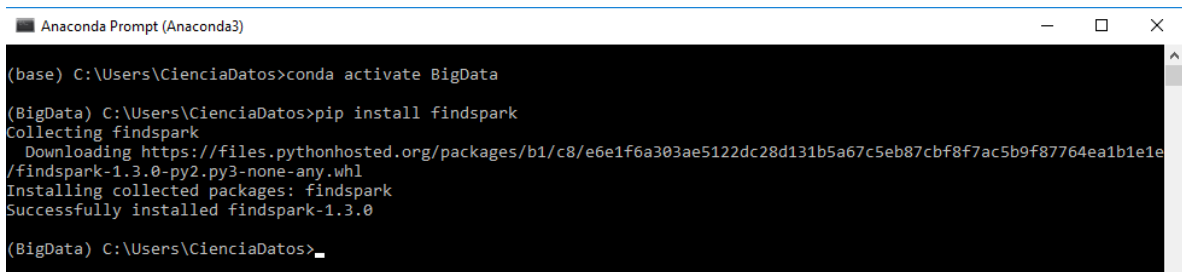
Instalación de findSpark

Para instalar el findSpark primero debemos de levantar la consola de comandos de Conda, ingresamos al ambiente de trabajo que vamos a utilizar y luego digitar el siguiente comando:

```
pip install findspark
```



```
Anaconda Prompt (Anaconda3)
(base) C:\Users\CienciaDatos>conda activate BigData
(BigData) C:\Users\CienciaDatos>pip install findspark_
```



```
Anaconda Prompt (Anaconda3)
(base) C:\Users\CienciaDatos>conda activate BigData
(BigData) C:\Users\CienciaDatos>pip install findspark
Collecting findspark
  Downloading https://files.pythonhosted.org/packages/b1/c8/e6e1f6a303ae5122dc28d131b5a67c5eb87cbf8f7ac5b9f87764ea1b1e1e/findspark-1.3.0-py2.py3-none-any.whl
Installing collected packages: findspark
Successfully installed findspark-1.3.0
(BigData) C:\Users\CienciaDatos>
```