SEP
SECRETARÍA DE EDUCACIÓN PÚBLICA

TECNOLÓGICO NACIONAL DE MÉXICO

TECNOLÓGICO NACIONAL DE MÉXICO
INSTITUTO TECNOLÓGICO DE TIJUANA

SUBDIRECCIÓN ACADÉMICA

DEPARTAMENTO DE SISTEMAS Y COMPUTACIÓN

PERIODO:

AGOSTO-DICIEMBRE 2020

ING. SISTEMAS COMPUTACIONALES

Materia:
Datos masivos

Bloque 1

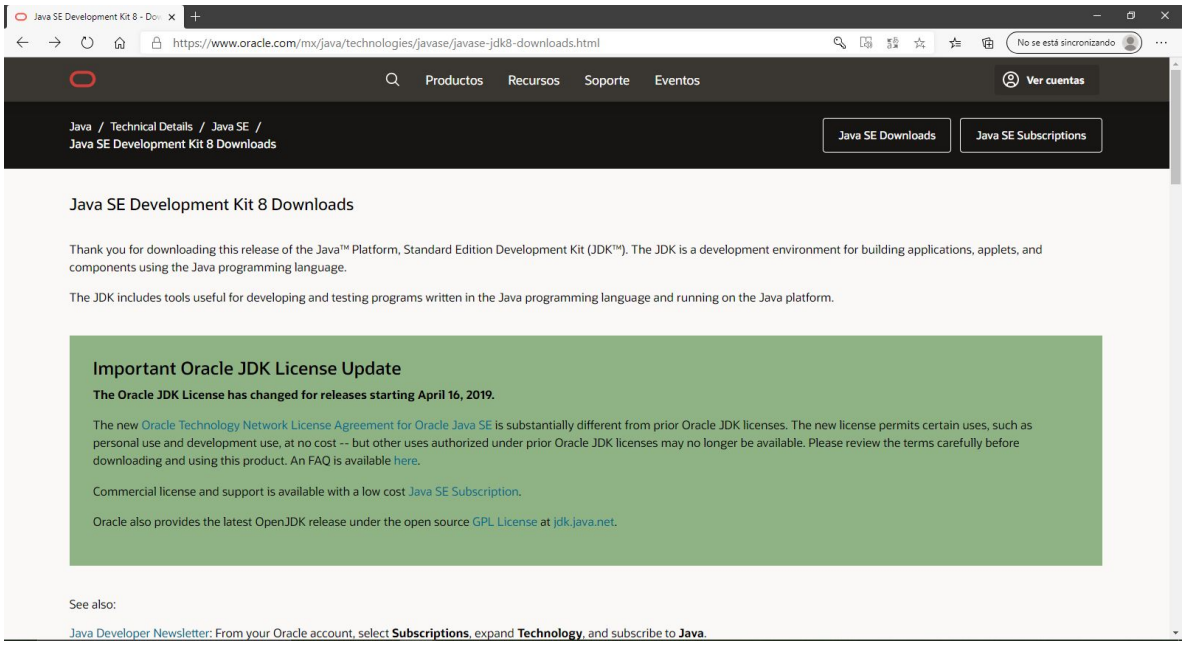
Documento de instalación de Scala y spark.

Docente:
José Christian Romero Hernández

Alumno:
Jonathan Hernández Iglesias

Index

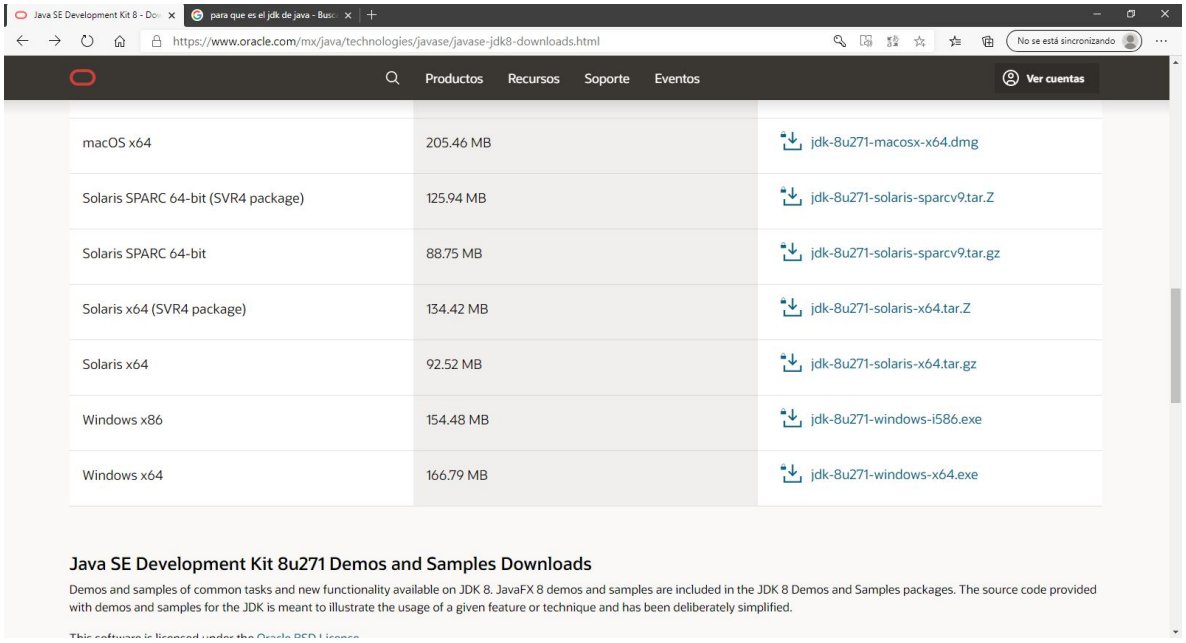
Installing the java JDK.



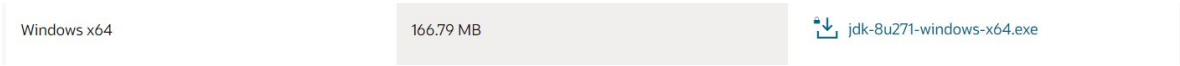
To begin with the installation process we must install the latest version of the Java JDK (Java development kit), its installation is necessary because it is required for the use of spark, this is necessary because it provides us with tools to install and work with various languages different programming and framework.

To find the jdk we enter the following link:

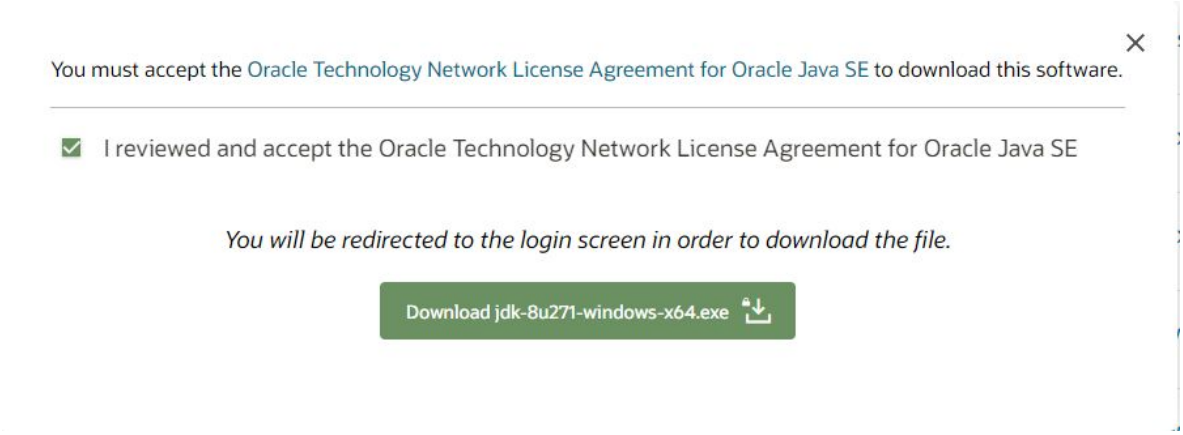
https://www.oracle.com/mx/java/technologies/javase/javase-jdk8-downloads.html



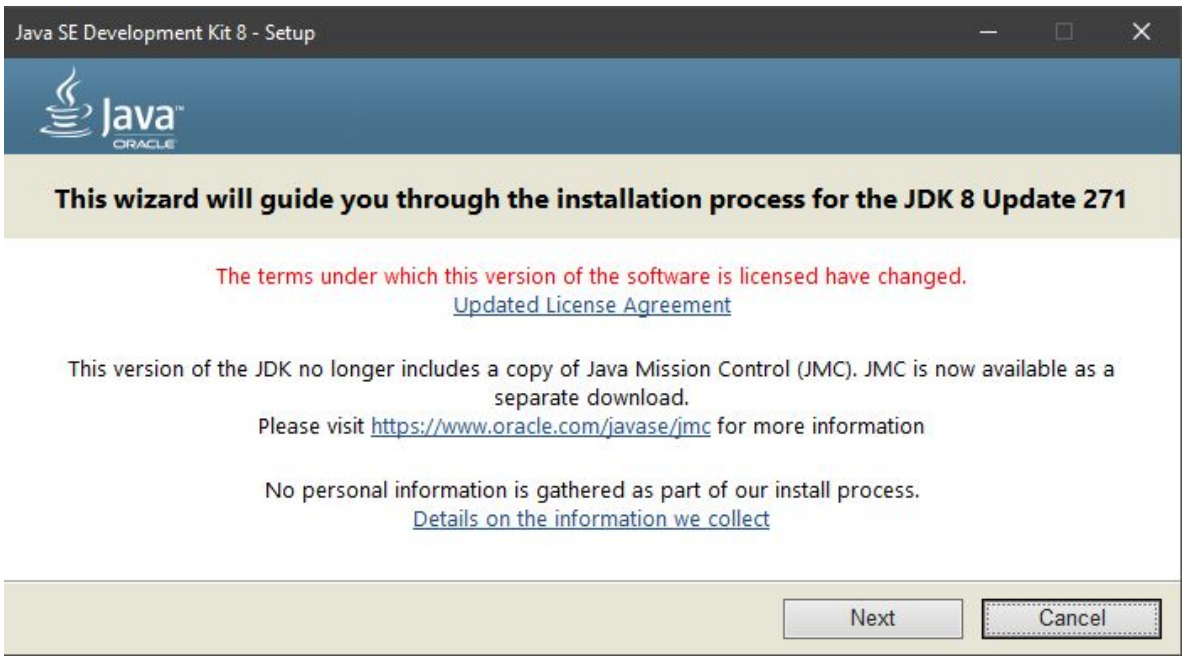
On the page we must scroll down where we will have to select the option according to our operating system, on the page you can find installation options for Linux, mac os, solaris and Windows; in this case my installation will proceed with the option for Windows in 64 bits.



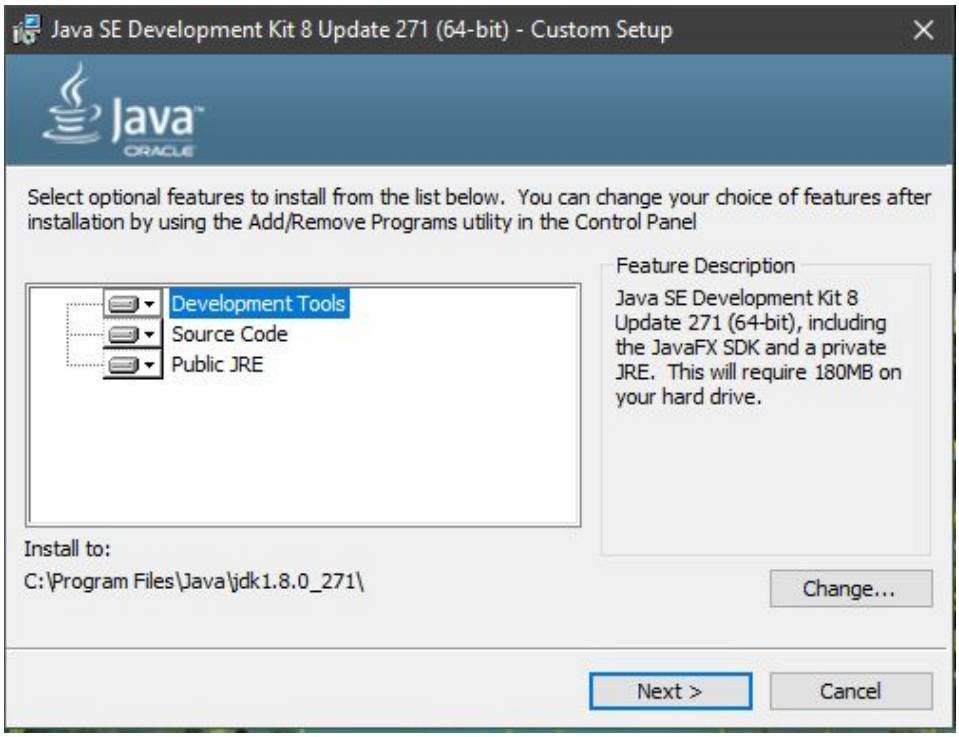
When selecting the option to download, the system will show us a window about the Oracle license terms, they must be accepted to start the download.



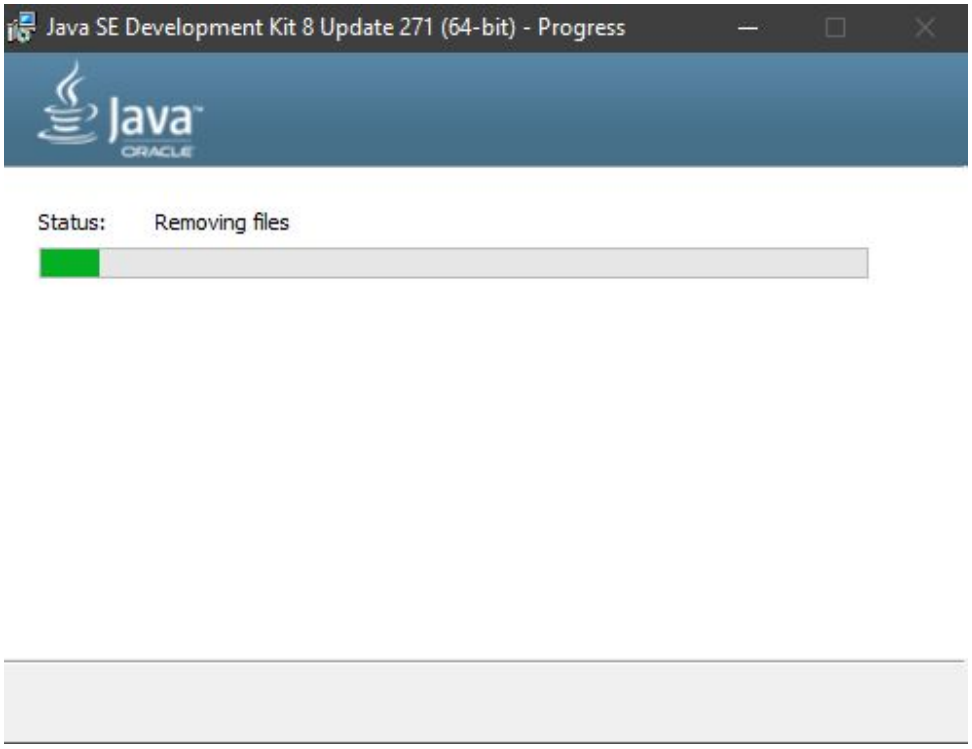
Once the executable is downloaded, we will open it; After this, a window with the installer will be displayed where we will begin the process for its installation.



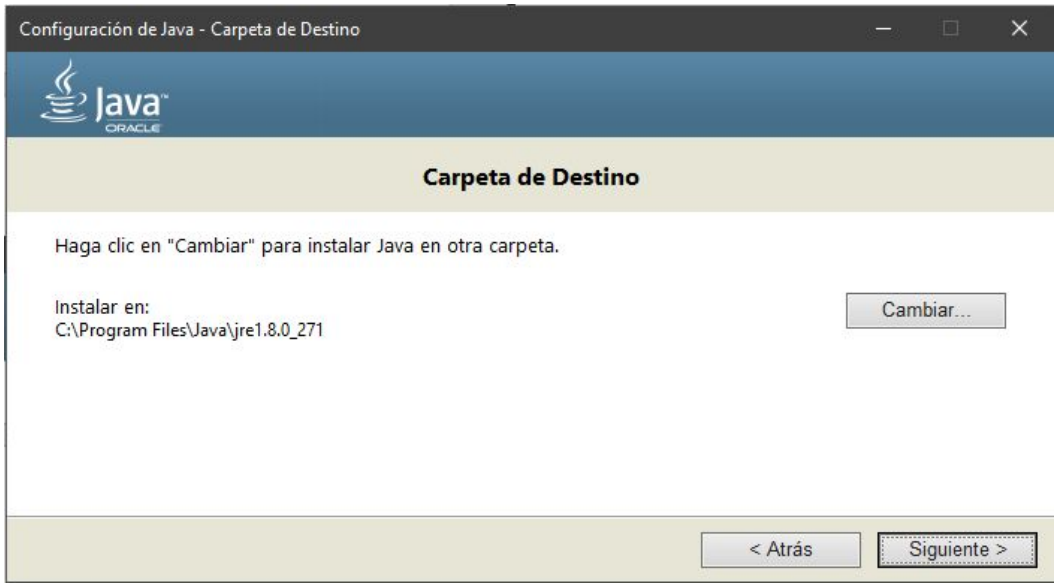
We will proceed to select the “Next” button, a new window will be displayed where we can select the features to install, in this case we will leave the default options by pressing “Next”.



The necessary files will begin to install.

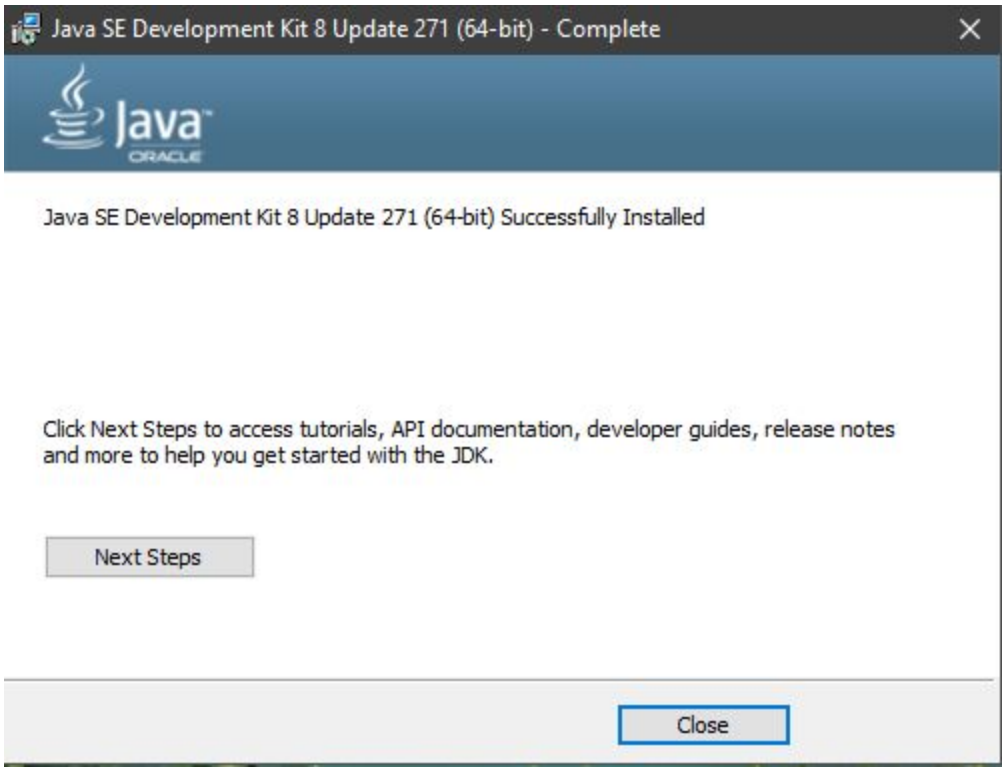


Once the installer is loaded we will begin to configure java, in this section we will select the destination folder, in this case we will leave the folder created by default and select "Next"

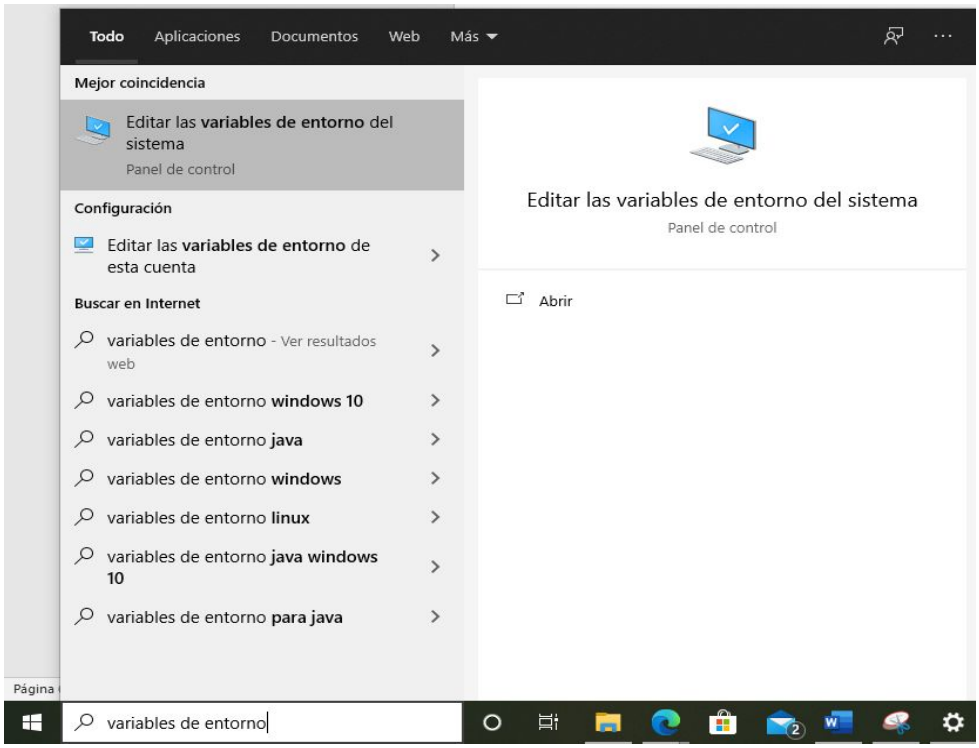


Once this is done, Java will begin to install.

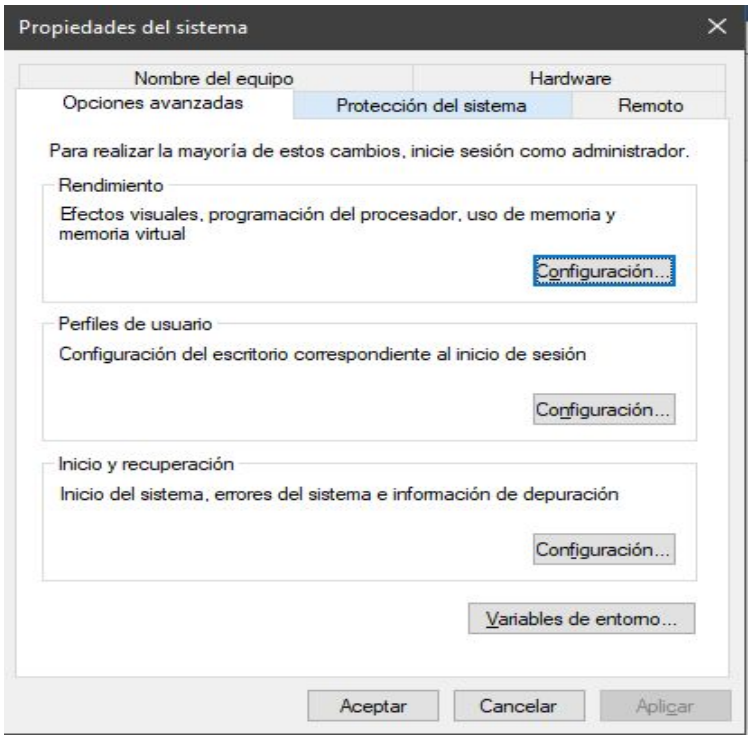
When we finish loading, we will have installed Java on our computer and we will select "Close" to finish the JDK installation process.



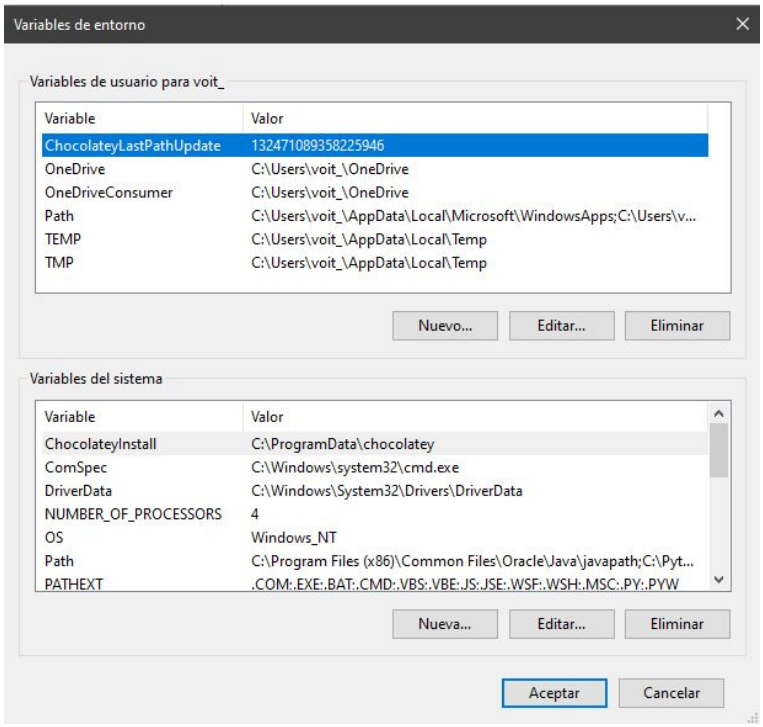
Now we must insert Java in the environment variables for the process, to do this, we must do the following.



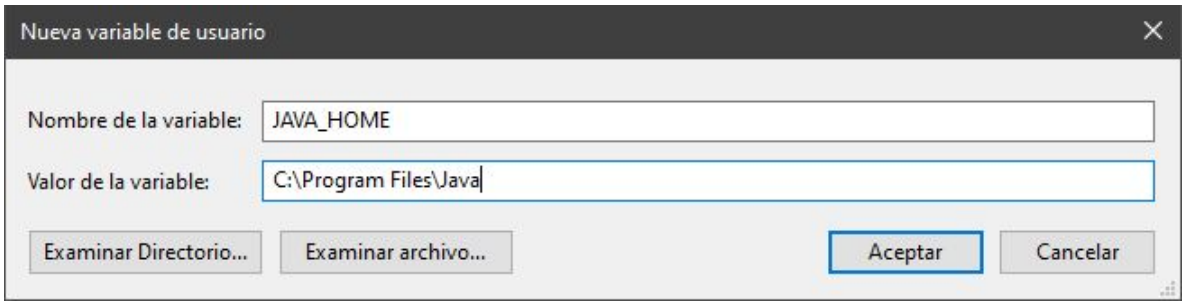
In the search bar, we write "Environment variables". And we select the option that says "Edit the system environment variables".



A window like the following one will appear, in this we must select the option of "environment variables"



The following window will open where we must select the option that says "NEW", after this the following window will open.



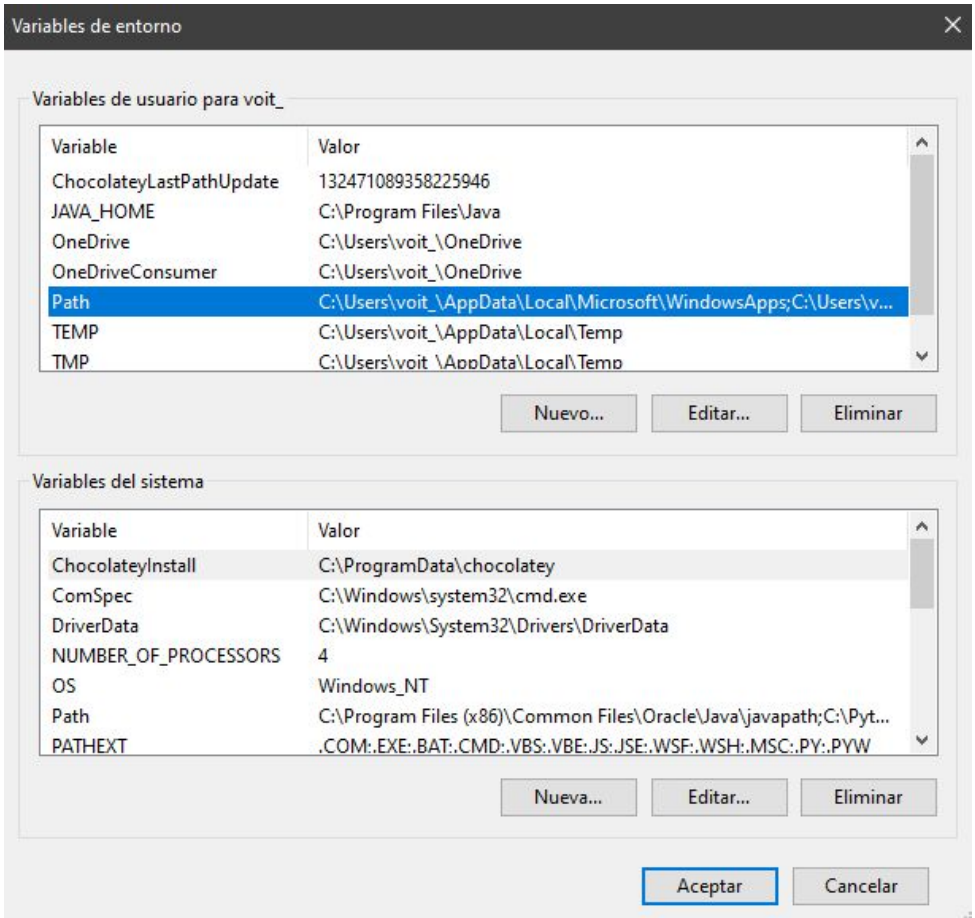
We are going to add the new environment variable to be able to use it.

In the variable name section, we write the following:

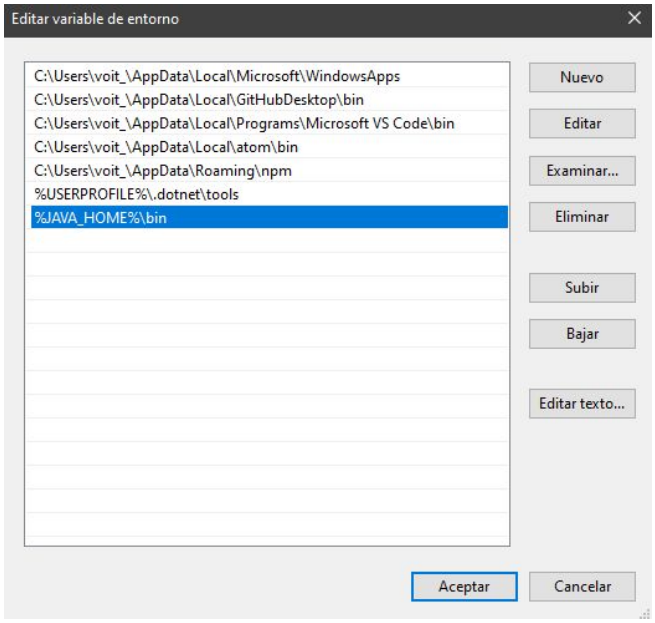
JAVA_HOME

In variable value we must select the path where the java file was installed.

And we click on "Accept"



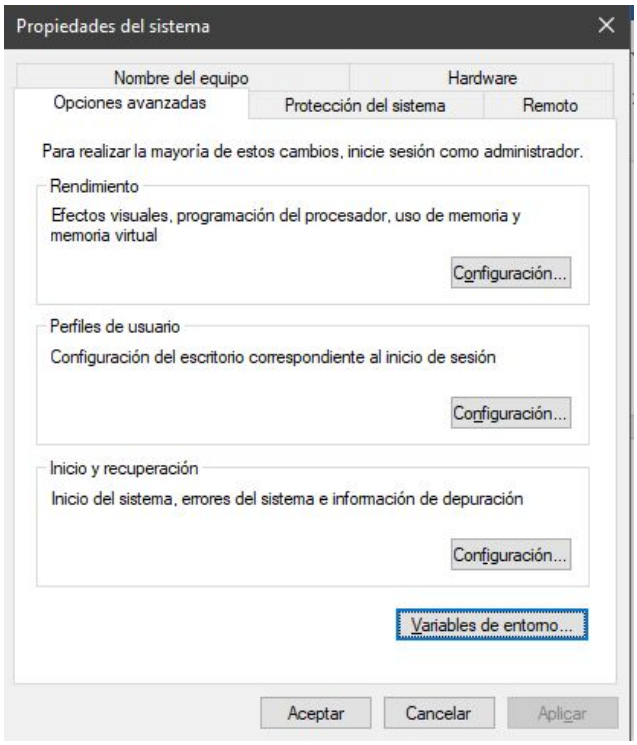
After this we must add the variable in the option "Path" for which we will select it and click on edit.



The following window will open and we must click on "New" and with we will write the following to add the variable to the path

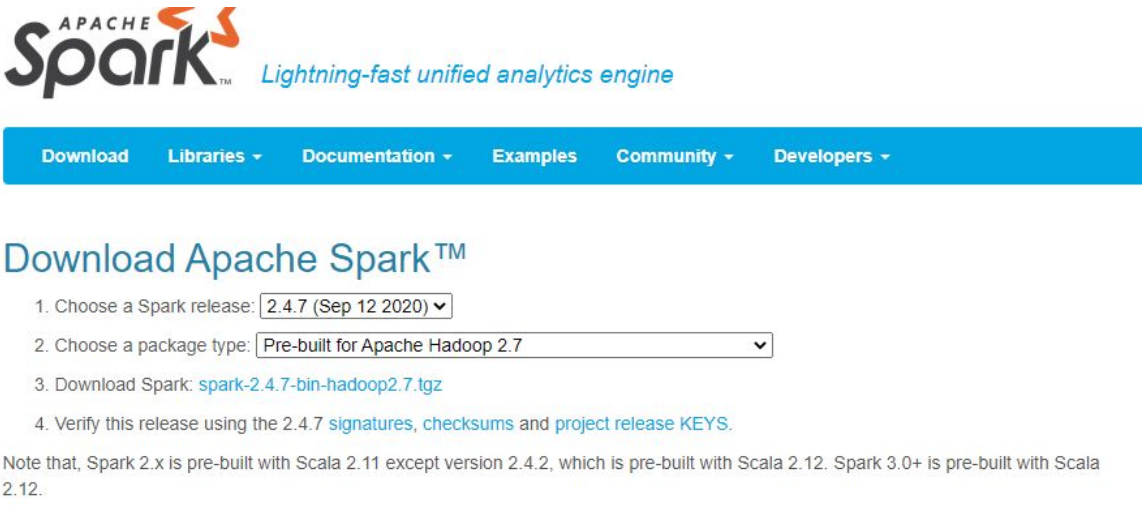
% JAVA_HOME% \ bin

And we will click on accept.



We will click on accept to finish the environment variables process.

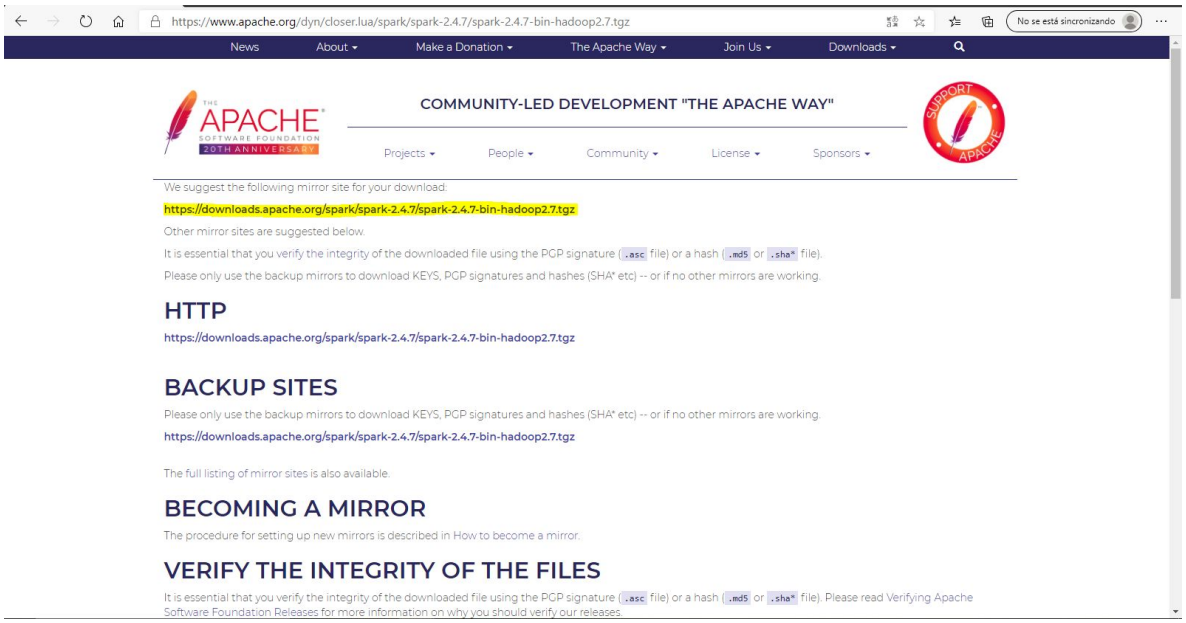
Spark installation



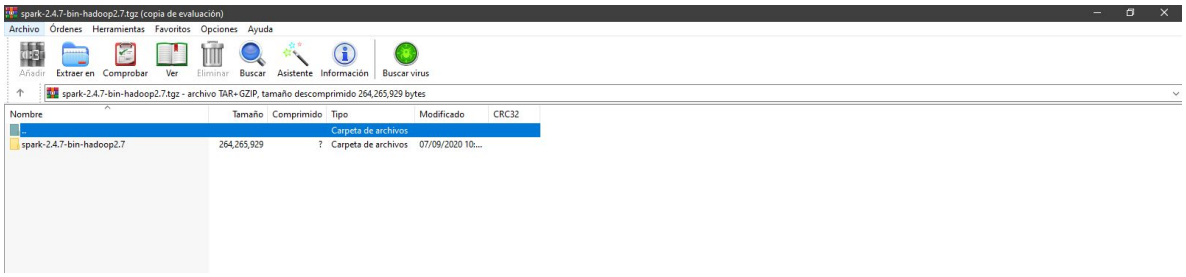
To install spark we must enter the following link:

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

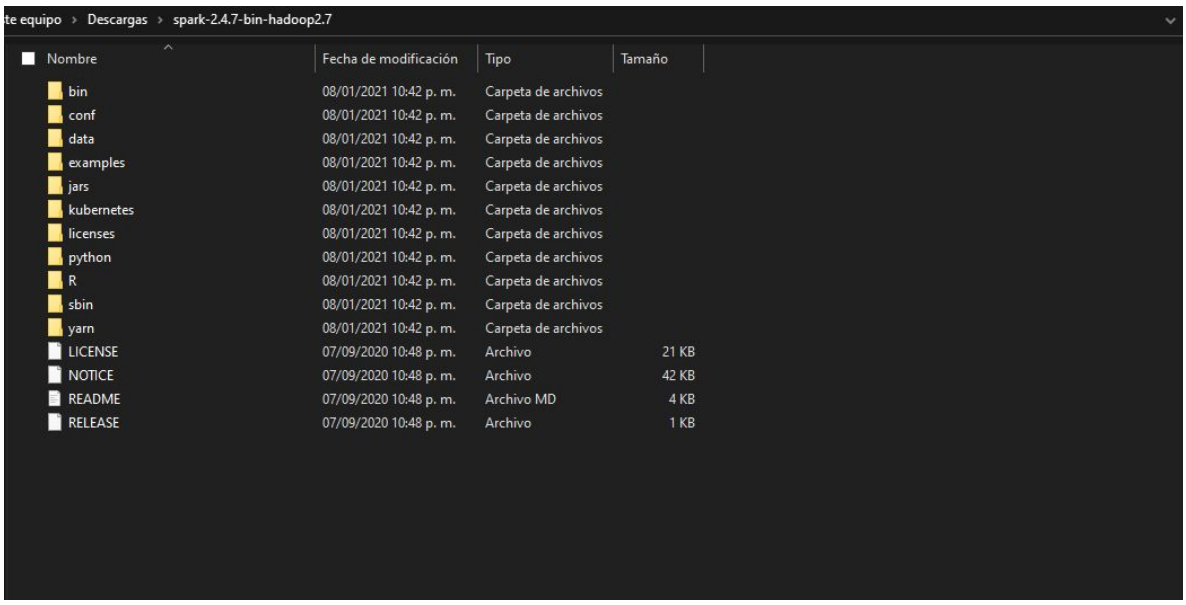
In this section we must choose the following values and click on option 3 that says "download spark"



The following window will open in which we select the first option to download the file.

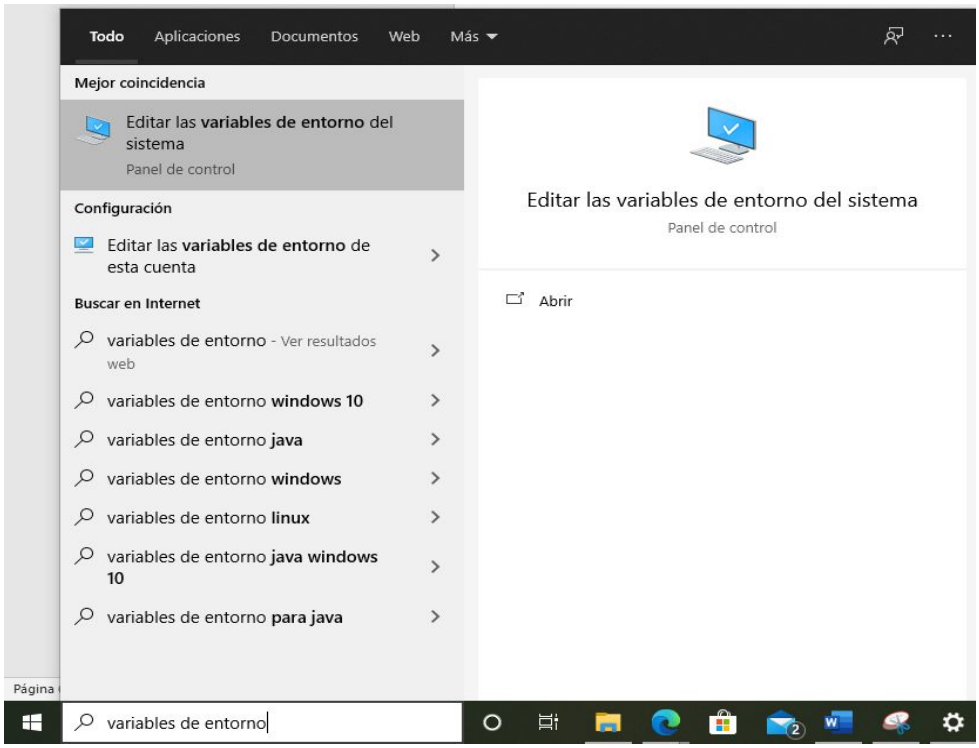


Once the file is downloaded, we must unzip the file.

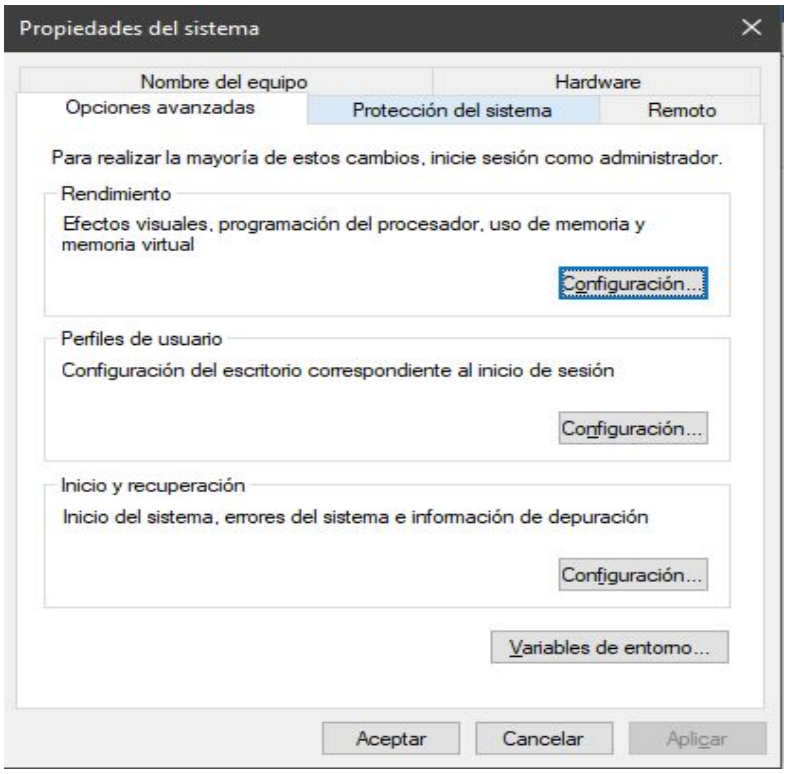


Once unzipped, we create a folder where we will save all the files in order to have a path for the environment variable.

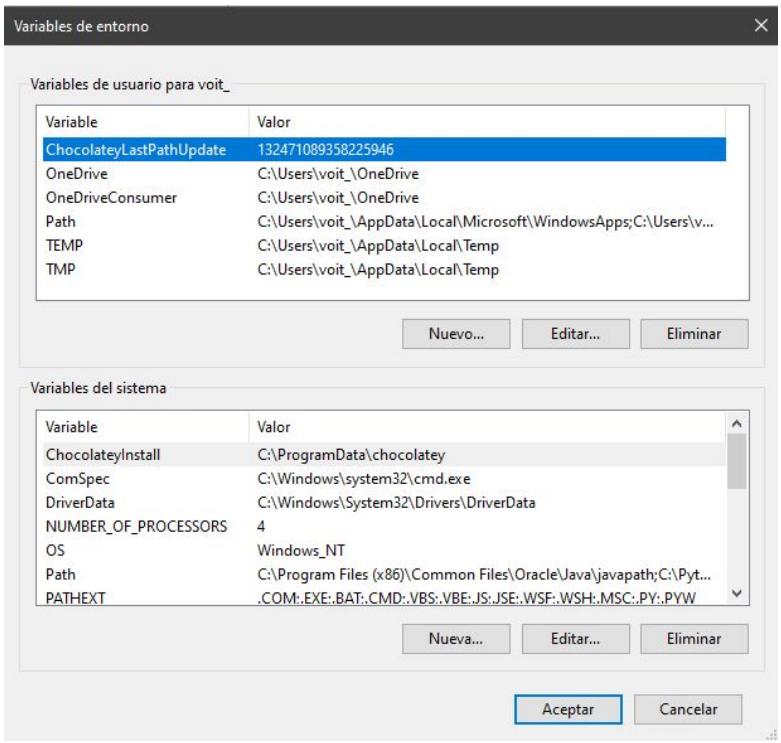
Now we must insert spark in the environment variables for the process, to do this, we must do the following.



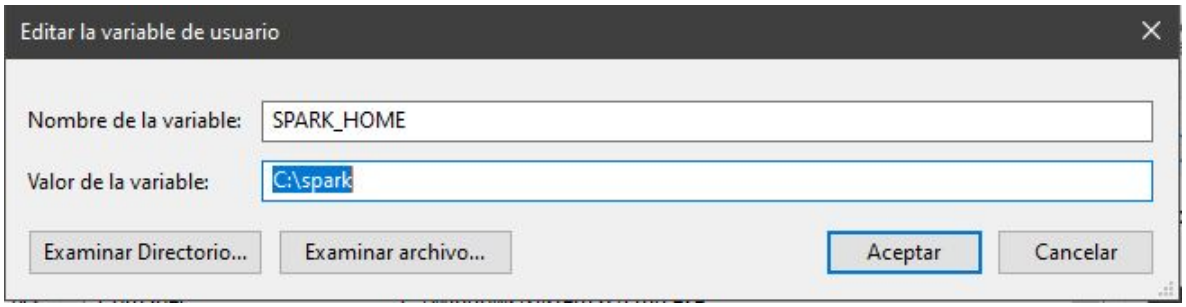
In the search bar, we write "Environment variables". And we select the option that says "Edit the system environment variables".



A window like the following one will appear, in this we must select the option of "environment variables"



The following window will open where we must select the option that says "NEW", after this the following window will open.



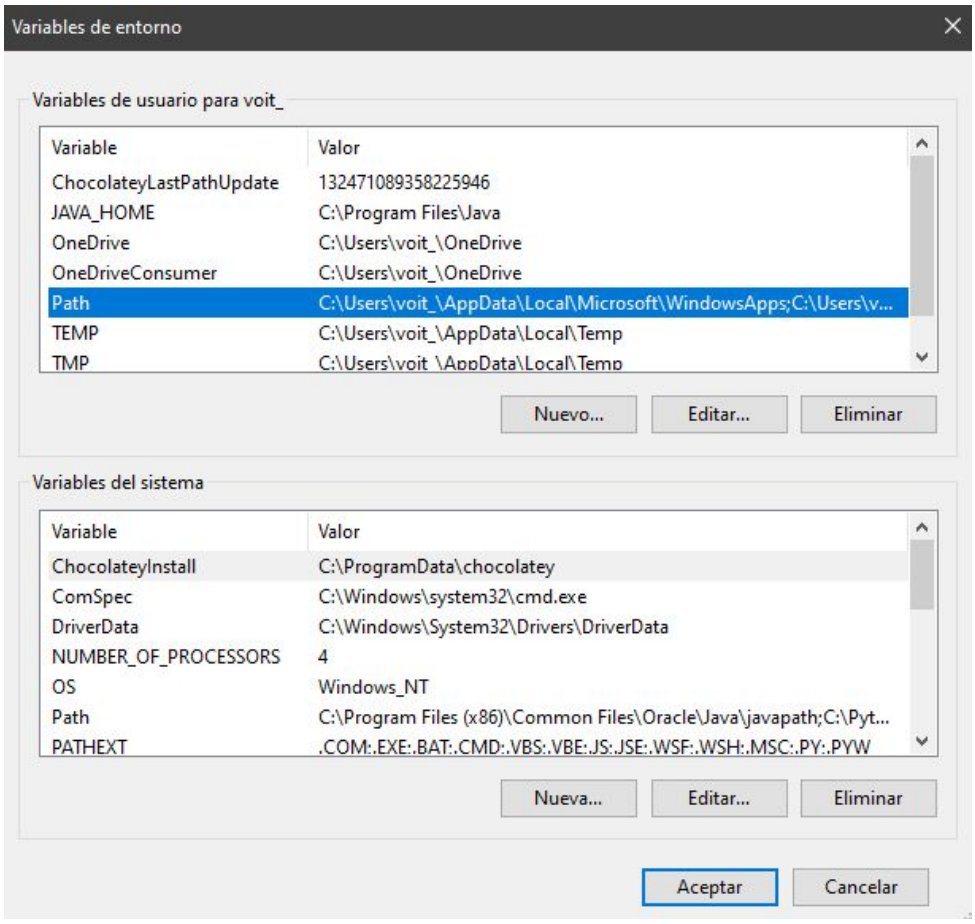
We are going to add the new environment variable to be able to use spark in any direction.

In the variable name section, we write the following:

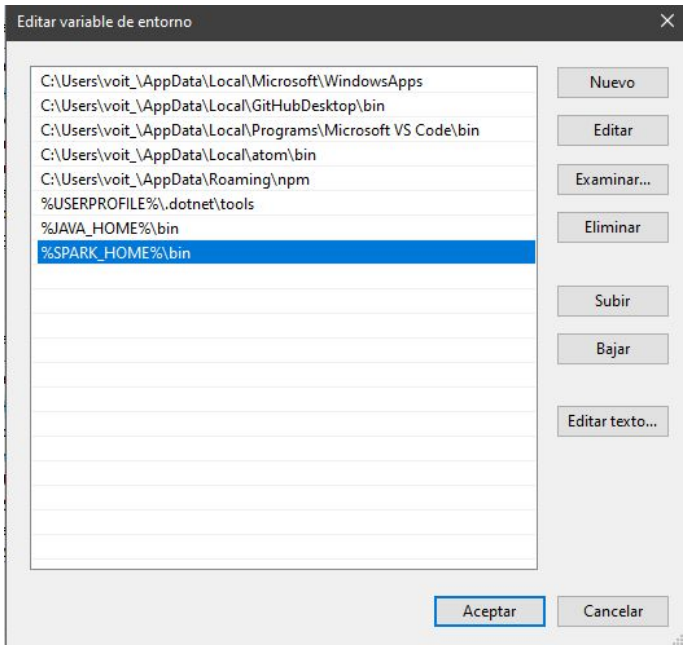
SPARK_HOME

In variable value we must select the path where the spark file was installed.

And we click on "Accept"



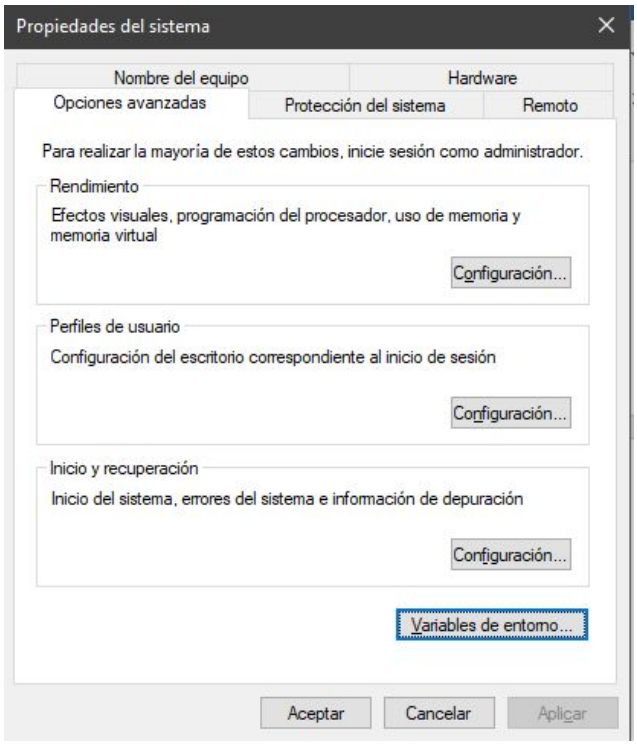
After this we must add the variable in the option "Path" for which we will select it and click on edit.



The following window will open and we must click on "New" and with we will write the following to add the variable to the path

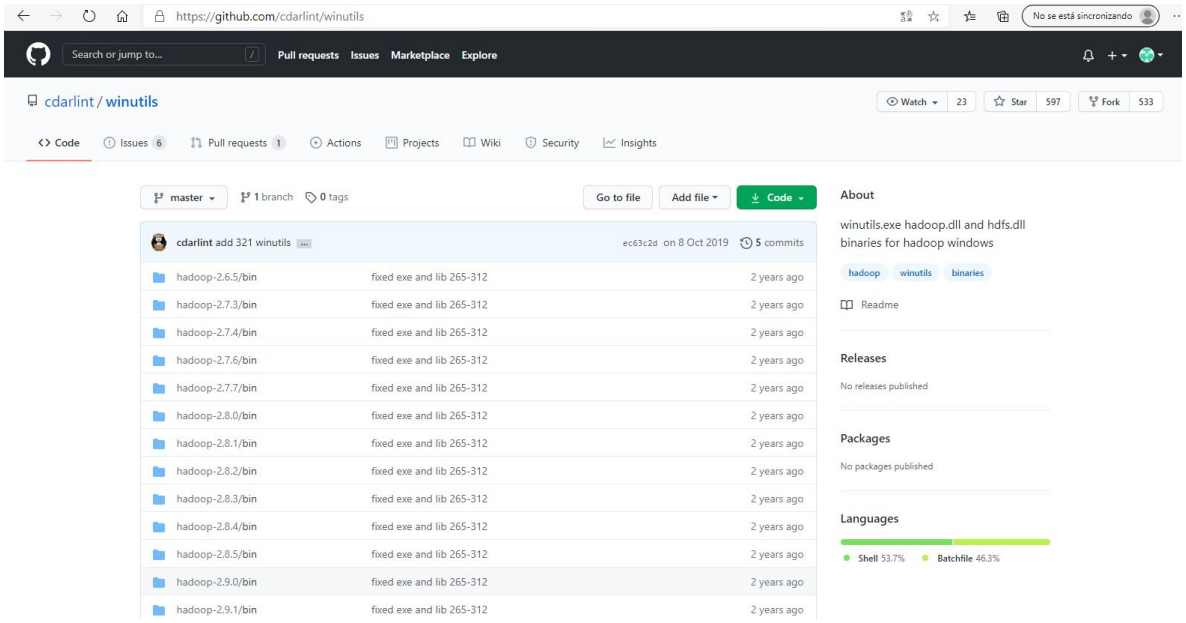
% SPARK_HOME% \ bin

And we will click on accept.



We will click on accept to finish the environment variables process.

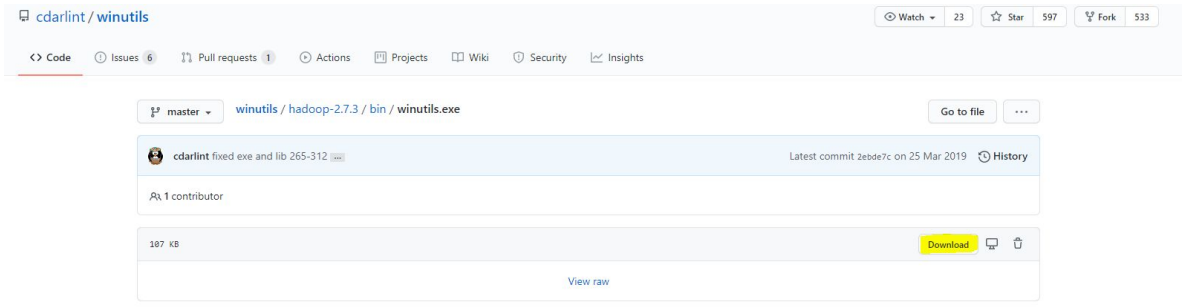
Install winutils.exe



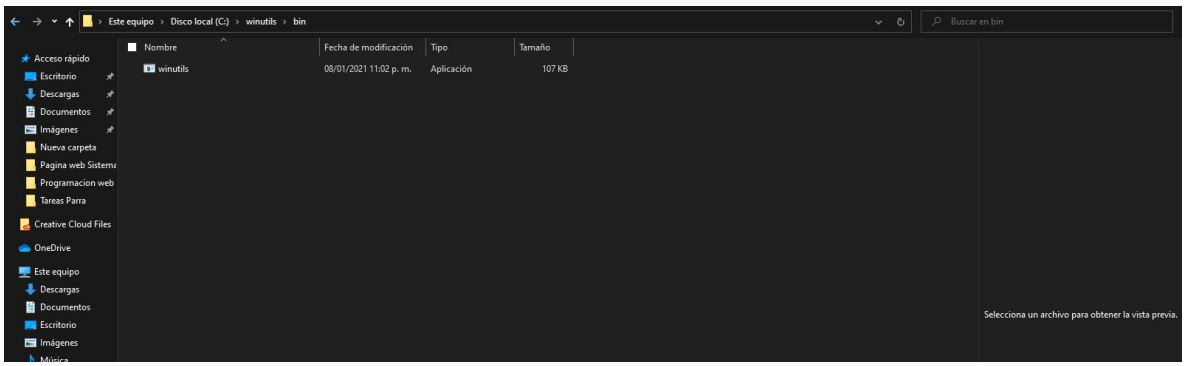
To install winutil.exe we must enter the following github link

https://github.com/cdarlint/winutils

In this github we must look for the file according to the version of spark that we installed.

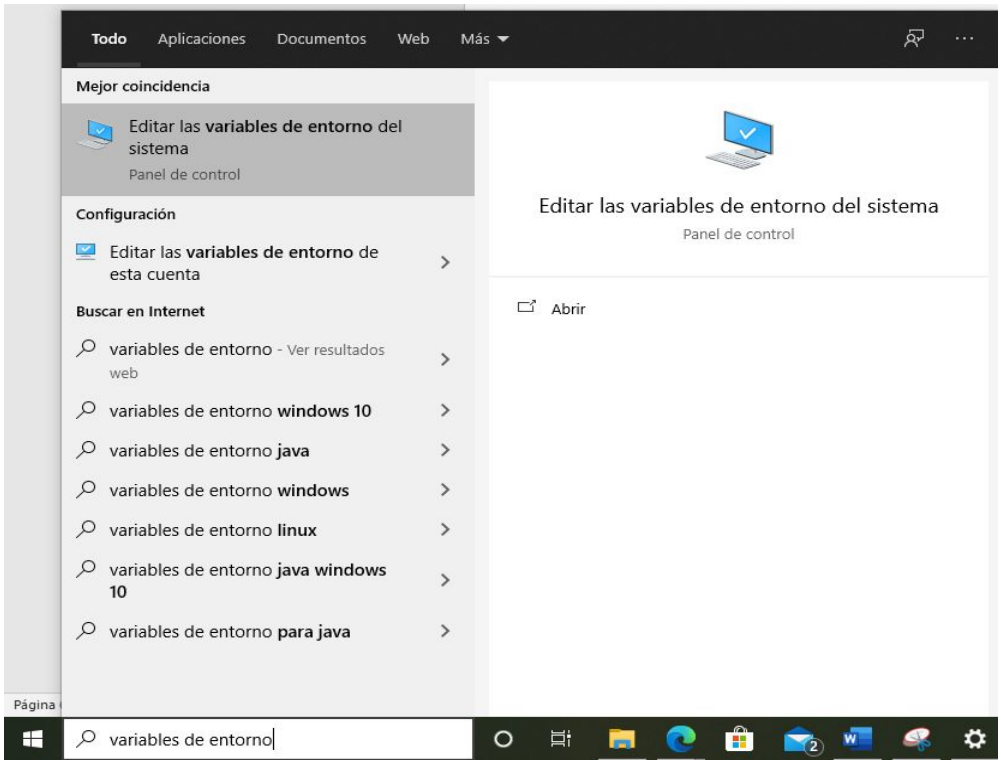


We will enter and look for the winutil.exe file and click, after this another tab will open, in which the "download" button will appear to download the file.

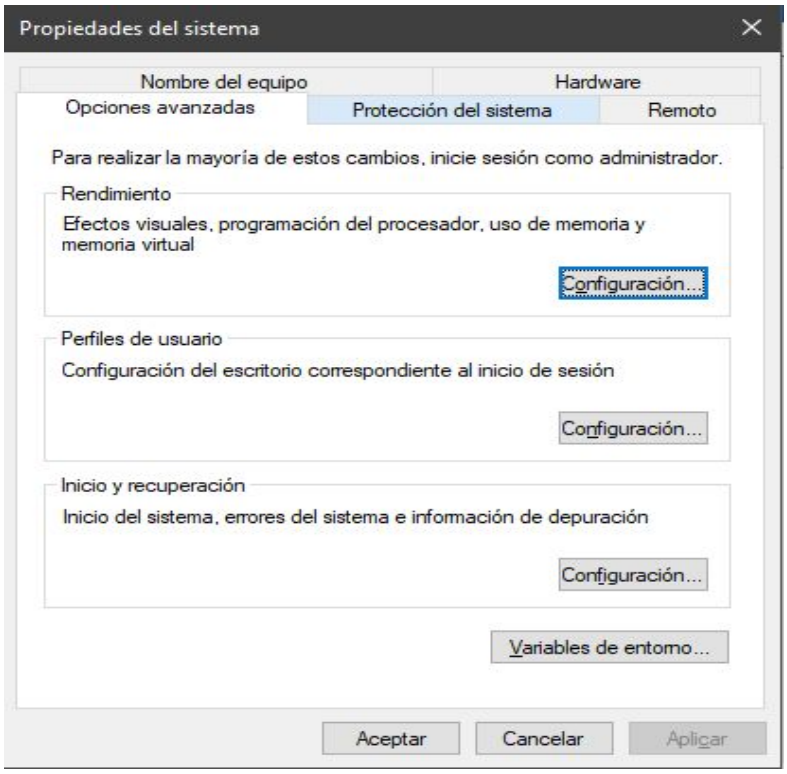


Once downloaded we must create a folder in the root (c :) with the name of winutils, within this folder we must create another folder called bin and we must leave the downloaded file.

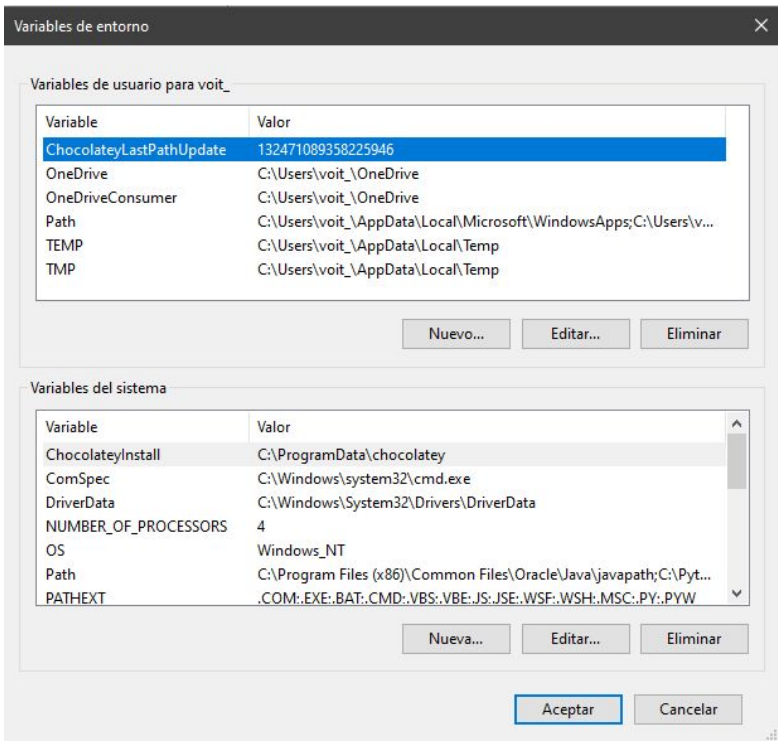
Now we must insert HADOOP in the environment variables for the process, to do this, we must do the following.



In the search bar, we write "Environment variables". And we select the option that says "Edit the system environment variables".



A window like the following one will appear, in this we must select the option of "environment variables"



The following window will open where we must select the option that says "NEW", after this the following window will open.

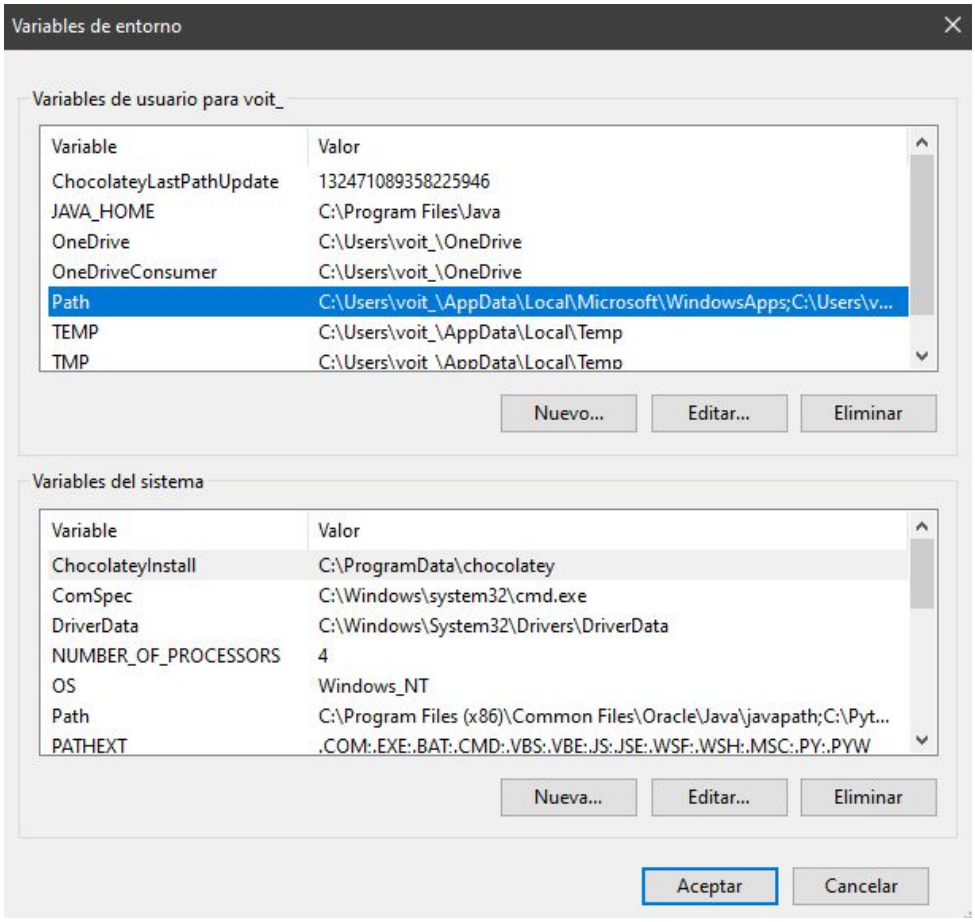
We are going to add the new environment variable to be able to use HADOOP in any direction.

In the variable name section, we write the following:

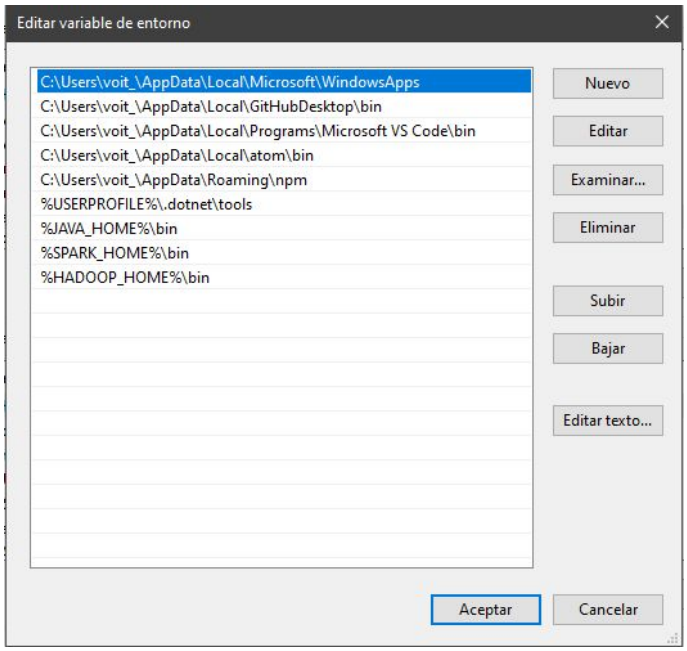
HADOOP_HOME

In variable value we must select the path where the HADOOP file was installed.

And we click on "Accept"



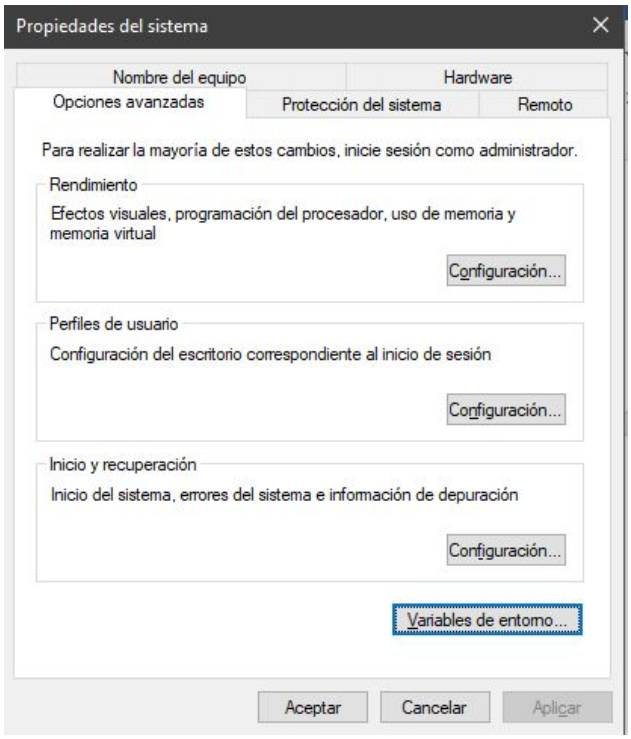
After this we must add the variable in the option "Path" for which we will select it and click on edit.



The following window will open and we must click on "New" and with we will write the following to add the variable to the path

% HADOOP_HOME% \ bin

And we will click on accept.



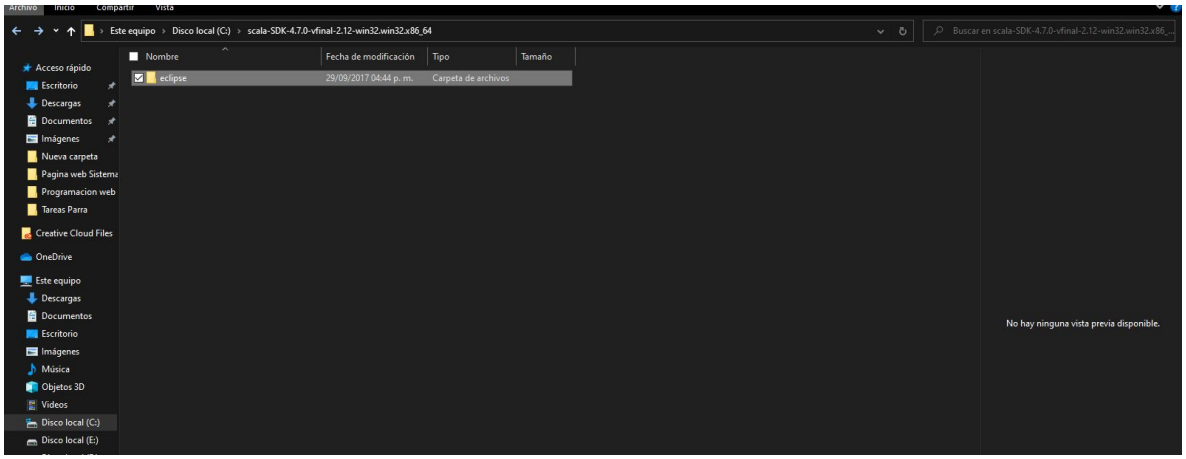
We will click on accept to finish the environment variables process.

Install scala

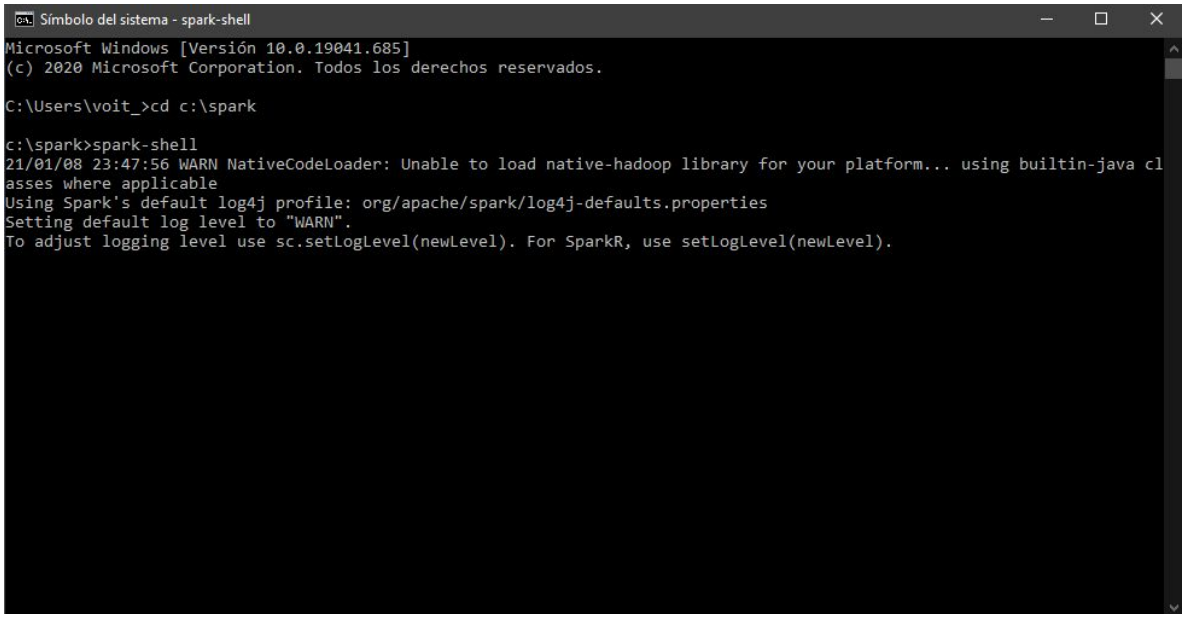
To install scala we must go to the following link:

http://scala-ide.org/download/sdk.html

Here we must click on the button "Download IDE"

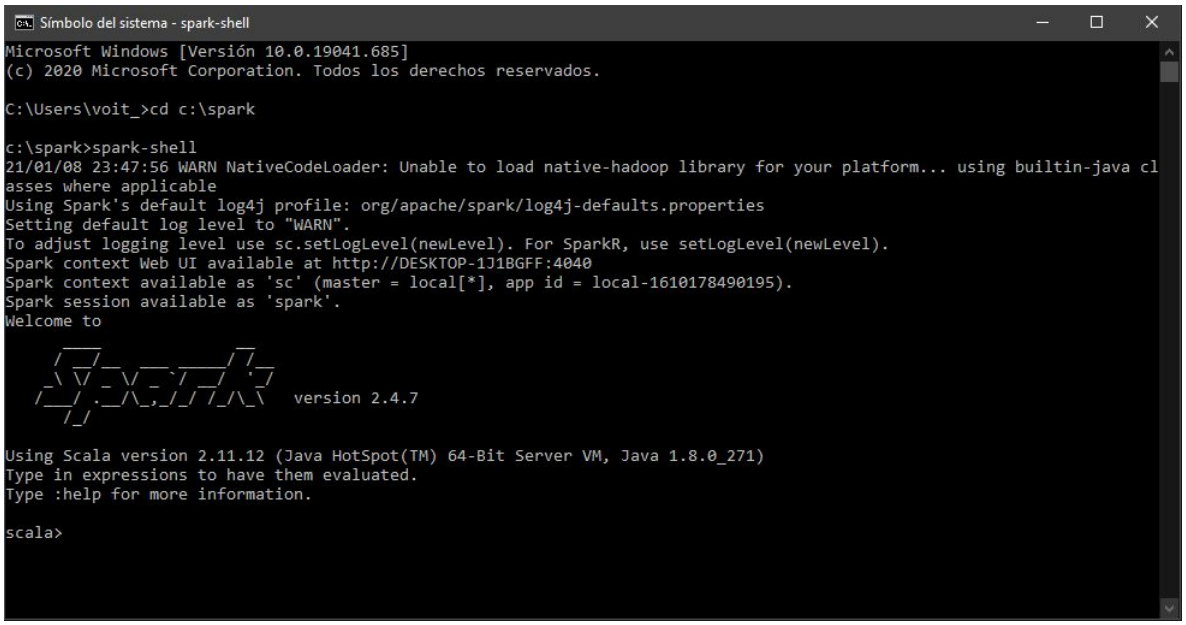


Once downloaded we must unzip it in a desired path.



Once the above is ready, we open a terminal, typing cmd, we proceed to change directory to enter the folder where we install spark, in my case it is in c: \ spark, once this is done we are going to

write the spark-shell command to enter the spark environment.



If everything was done correctly, we will have scala and spark installed on Windows.