



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.	2
Instalación de spark	11
Instalar winutils.exe	18
Instalar scala	23

Installing the java JDK.

The screenshot shows the Oracle website's 'Java SE Development Kit 8 Downloads' page. The page has a dark header with navigation links: 'Productos', 'Recursos', 'Soporte', and 'Eventos'. Below the header, there's a breadcrumb trail: 'Java / Technical Details / Java SE / Java SE Development Kit 8 Downloads'. Two buttons are visible: 'Java SE Downloads' and 'Java SE Subscriptions'. The main content area has a title 'Java SE Development Kit 8 Downloads' followed by a thank-you message and a link to the JDK. A green box highlights an 'Important Oracle JDK License Update' with details about the new license starting April 16, 2019. At the bottom, there's a 'See also:' section with a link to the 'Java Developer Newsletter'.

Java SE Development Kit 8 Downloads

Thank you for downloading this release of the Java™ Platform, Standard Edition Development Kit (JDK™). The JDK is a development environment for building applications, applets, and components using the Java programming language.

The JDK includes tools useful for developing and testing programs written in the Java programming language and running on the Java platform.

Important Oracle JDK License Update
The Oracle JDK License has changed for releases starting April 16, 2019.

The new Oracle Technology Network License Agreement for Oracle Java SE is substantially different from prior Oracle JDK licenses. The new license permits certain uses, such as personal use and development use, at no cost -- but other uses authorized under prior Oracle JDK licenses may no longer be available. Please review the terms carefully before downloading and using this product. An FAQ is available [here](#).

Commercial license and support is available with a low cost [Java SE Subscription](#).

Oracle also provides the latest OpenJDK release under the open source [GPL License](#) at [jdk.java.net](#).

See also:
[Java Developer Newsletter](#): From your Oracle account, select **Subscriptions**, expand **Technology**, and subscribe to **Java**.

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>

The screenshot shows the Oracle website's 'Java SE Development Kit 8u271 Downloads' page. The page has a dark header with navigation links: 'Productos', 'Recursos', 'Soporte', and 'Eventos'. Below the header, there's a breadcrumb trail: 'Java / Technical Details / Java SE / Java SE Development Kit 8 Downloads'. Two buttons are visible: 'Java SE Downloads' and 'Java SE Subscriptions'. The main content area has a title 'Java SE Development Kit 8u271 Downloads' followed by a table of download links for various operating systems. Below the table, there's a section for 'Java SE Development Kit 8u271 Demos and Samples Downloads' with a brief description and a link to the source code.

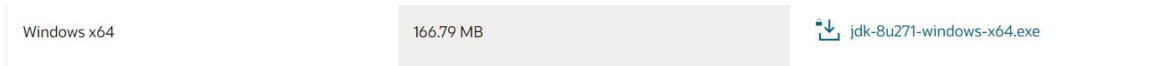
Operating System	Size	Download Link
macOS x64	205.46 MB	jdk-8u271-macosx-x64.dmg
Solaris SPARC 64-bit (SVR4 package)	125.94 MB	jdk-8u271-solaris-sparcv9.tar.Z
Solaris SPARC 64-bit	88.75 MB	jdk-8u271-solaris-sparcv9.tar.gz
Solaris x64 (SVR4 package)	134.42 MB	jdk-8u271-solaris-x64.tar.Z
Solaris x64	92.52 MB	jdk-8u271-solaris-x64.tar.gz
Windows x86	154.48 MB	jdk-8u271-windows-i586.exe
Windows x64	166.79 MB	jdk-8u271-windows-x64.exe

Java SE Development Kit 8u271 Demos and Samples Downloads

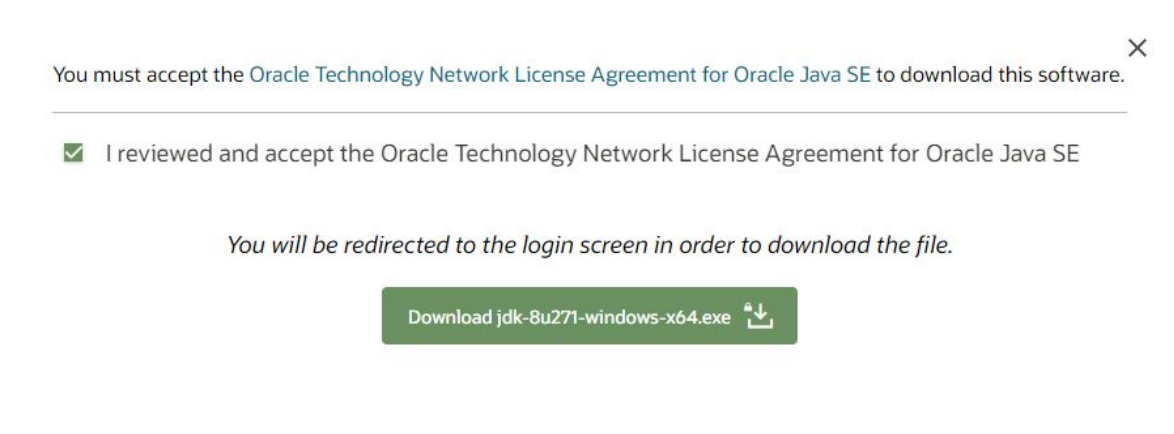
Demos and samples of common tasks and new functionality available on JDK 8. JavaFX 8 demos and samples are included in the JDK 8 Demos and Samples packages. The source code provided with demos and samples for the JDK is meant to illustrate the usage of a given feature or technique and has been deliberately simplified.

[This software is licensed under the Oracle BCL license.](#)

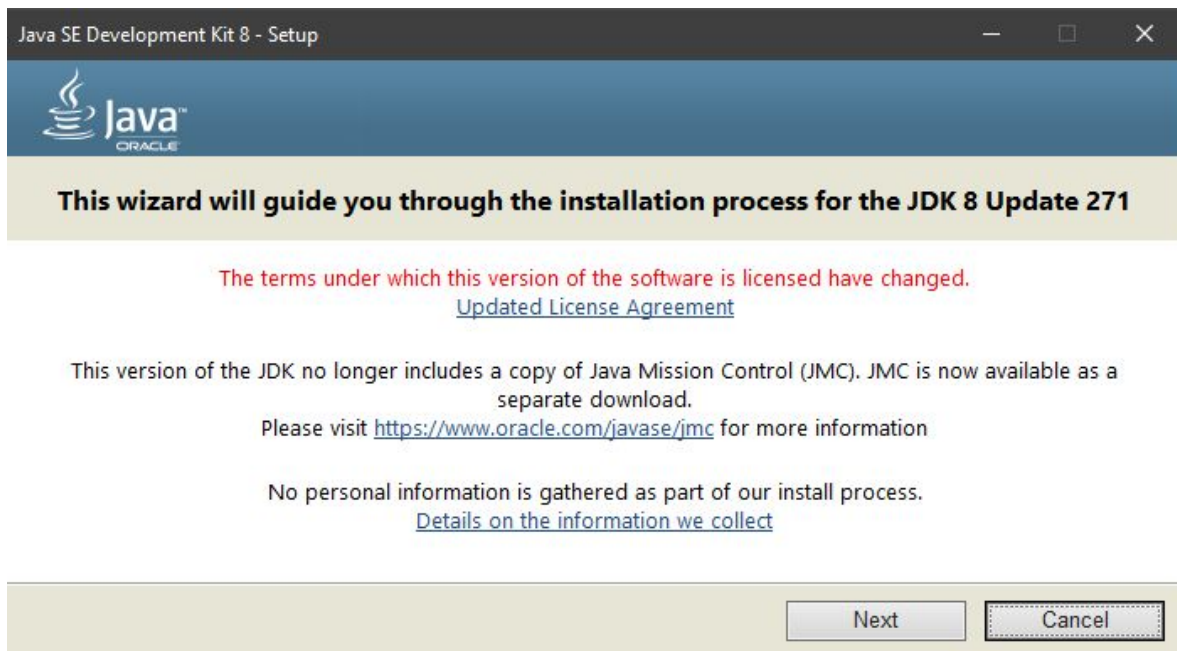
On the page we must scroll down where we will have to select the option according to our operating system, on the page you can find installation options for Linux, mac os, solaris and Windows; in this case my installation will proceed with the option for Windows in 64 bits.



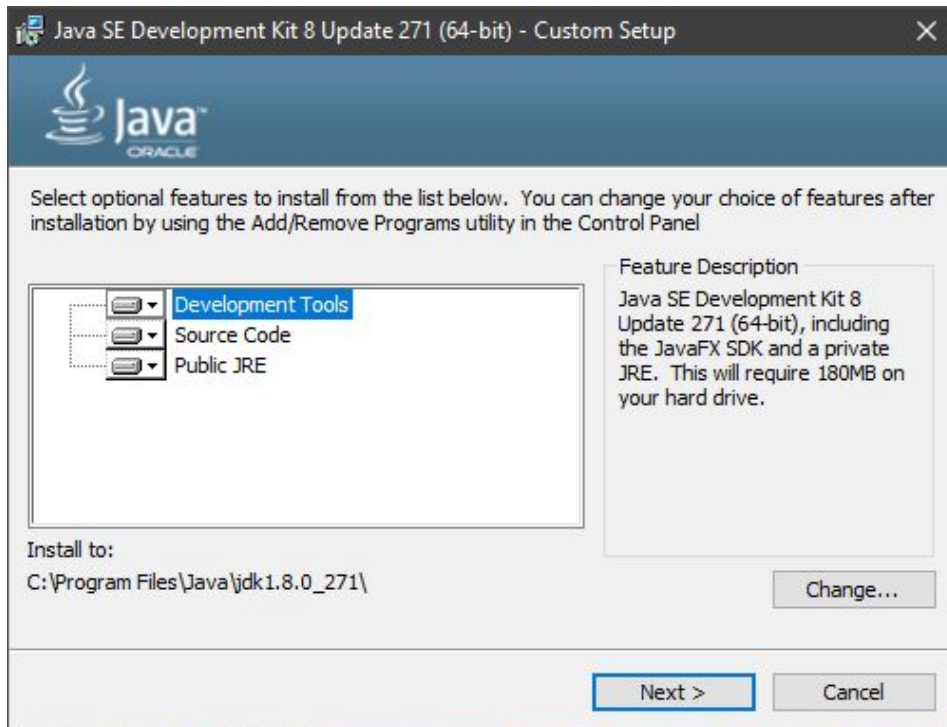
When selecting the option to download, the system will show us a window about the Oracle license terms, they must be accepted to start the download.



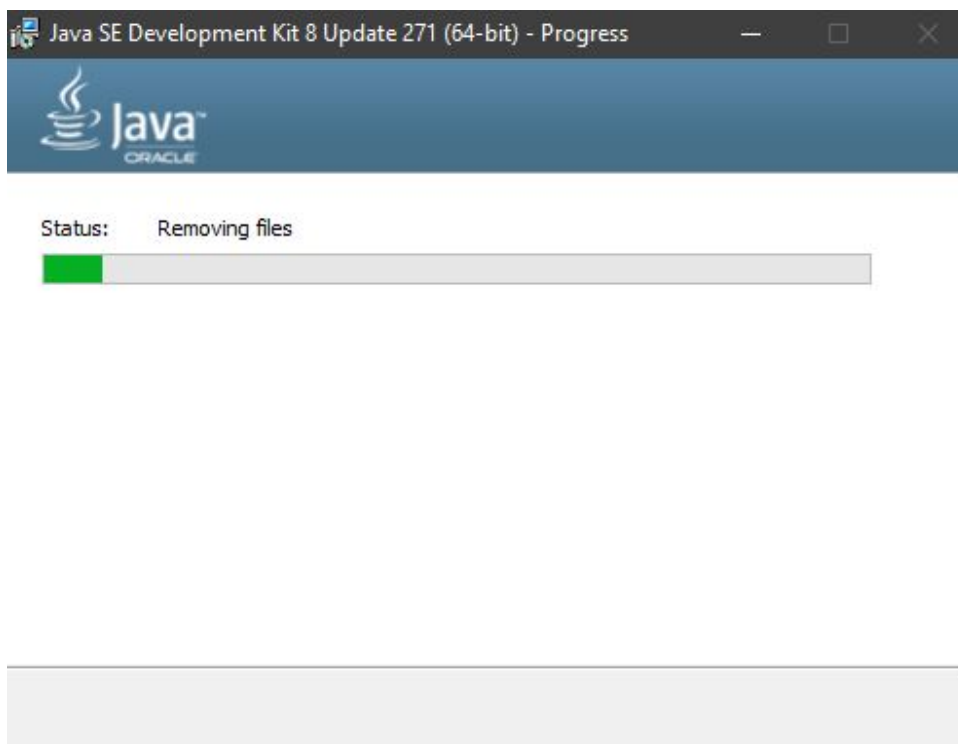
Once the executable is downloaded, we will open it; After this, a window with the installer will be displayed where we will begin the process for its installation.



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.



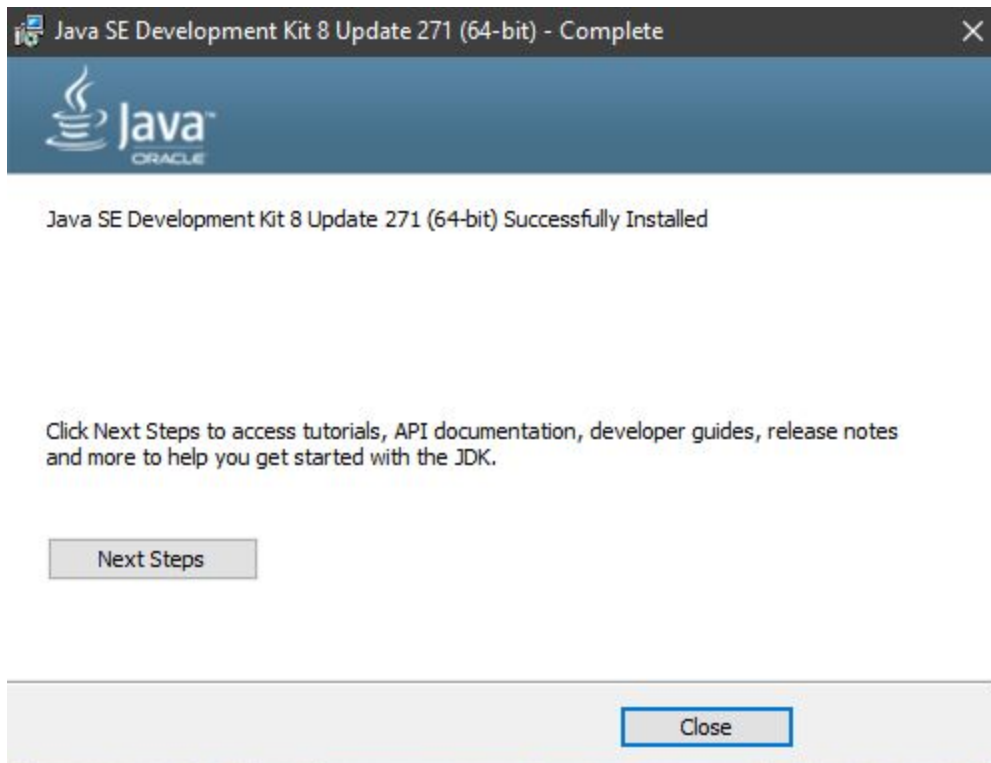
Once the installer is loaded we will begin to configure java, in this section we will select the destination folder, in this case we will leave the folder created by default and select "Next"



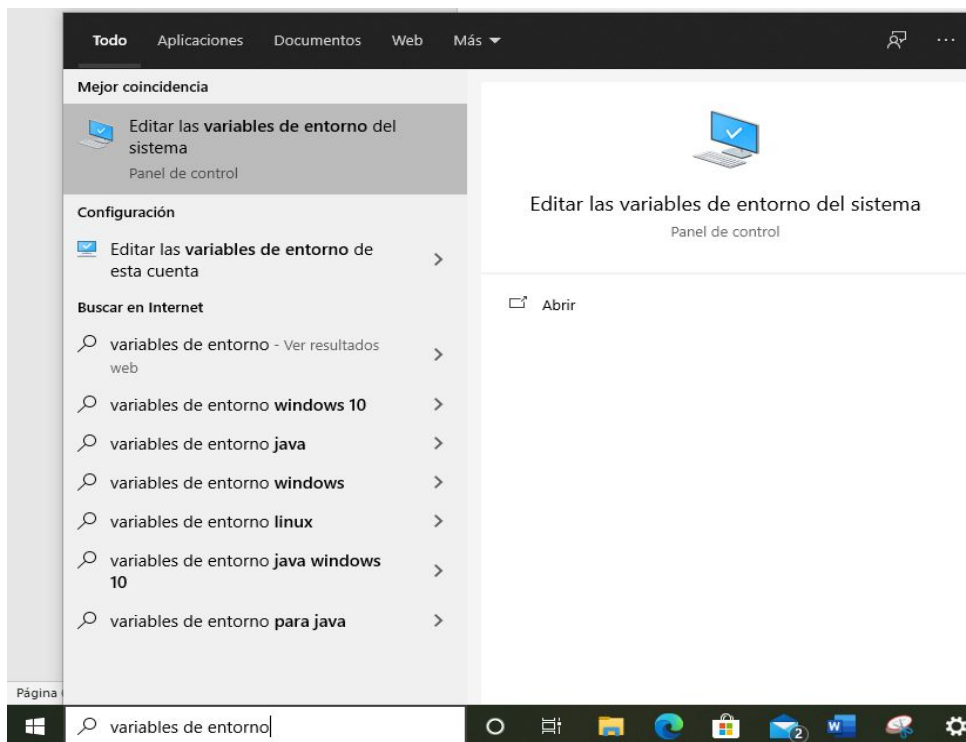
Once this is done, Java will begin to install.



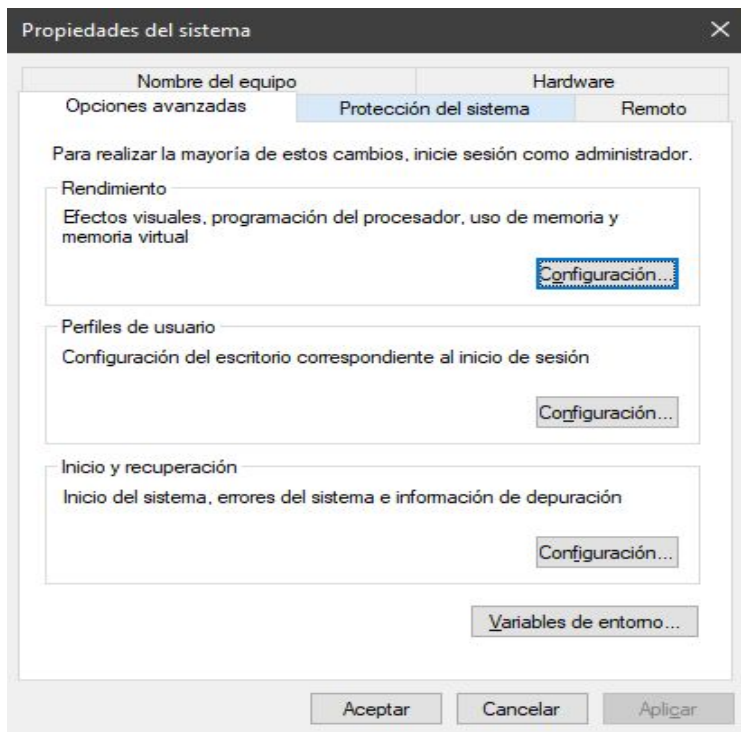
When we finish loading, we will have installed Java on our computer and we will select "Close" to finish the JDK installation process.



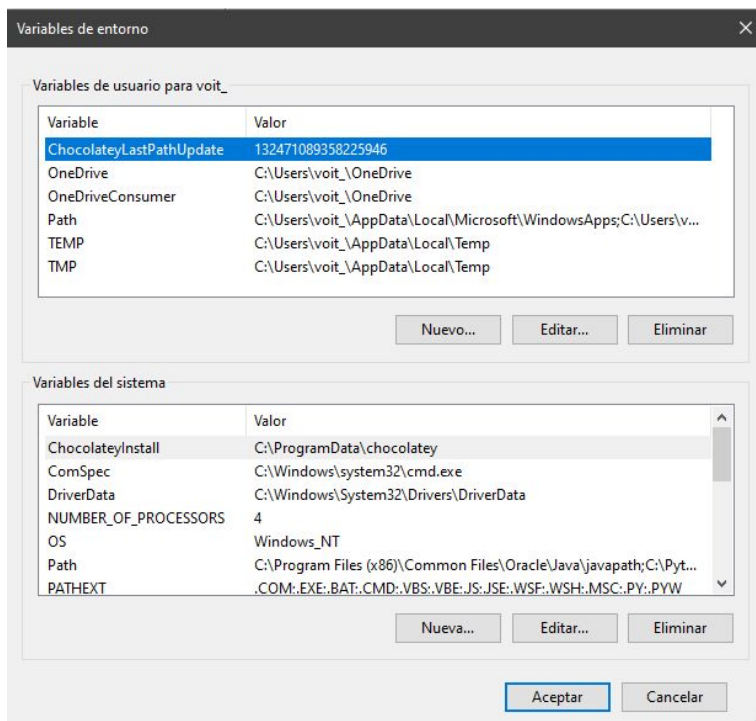
Now we must insert Java in the environment variables for the process, to do this, we must do the following.



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

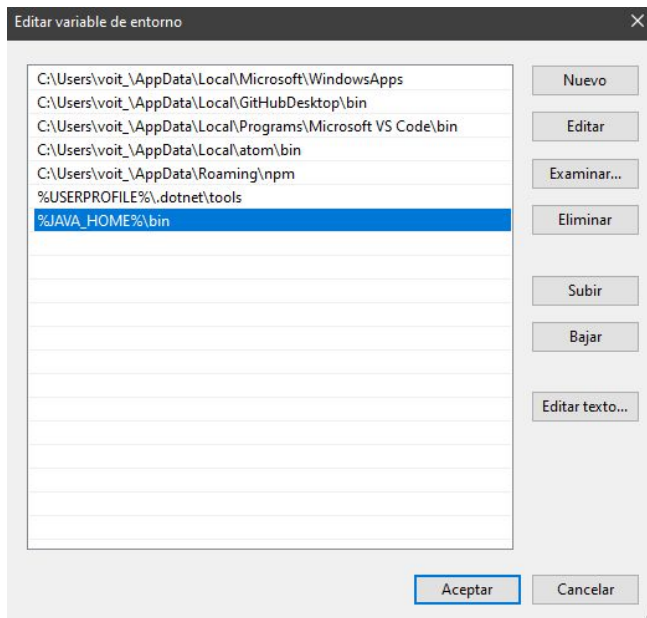
In variable value we must select the path where the java file was installed.

And we click on "Accept"

Variable	Valor
ChocolateyLastPathUpdate	132471089358225946
JAVA_HOME	C:\Program Files\Java
OneDrive	C:\Users\voit_OneDrive
OneDriveConsumer	C:\Users\voit_OneDrive
Path	C:\Users\voit\AppData\Local\Microsoft\WindowsApps;C:\Users\v...
TEMP	C:\Users\voit\AppData\Local\Temp
TMP	C:\Users\voit\AppData\Local\Temp

Variable	Valor
ChocolateyInstall	C:\ProgramData\chocolatey
ComSpec	C:\Windows\system32\cmd.exe
DriverData	C:\Windows\System32\Drivers\DriverData
NUMBER_OF_PROCESSORS	4
OS	Windows_NT
Path	C:\Program Files (x86)\Common Files\Oracle\Java\javapath;C:\Pyt...
PATHEXT	.COM;.EXE;.BAT;.CMD;.VBS;.VBE;.JS;.JSE;.WSF;.WSH;.MSC;.PY;.PYW

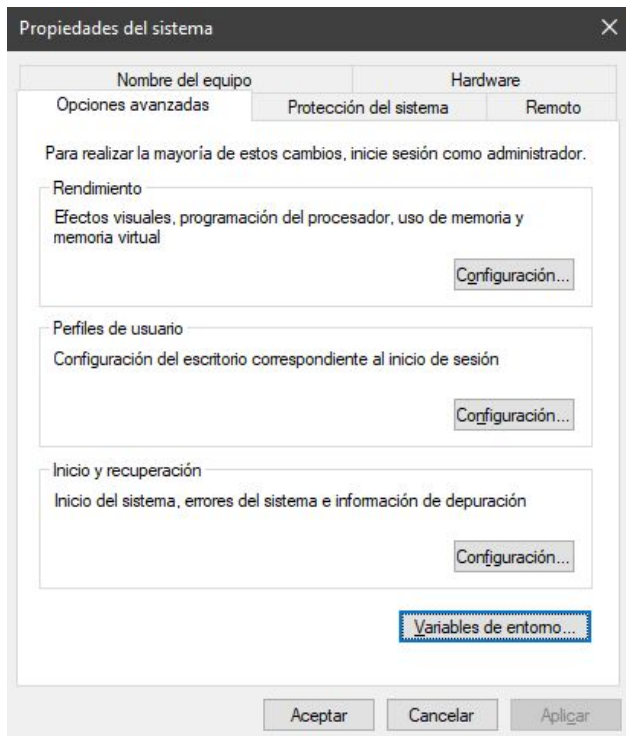
After this we must add the variable in the option "Path" for which we will select it and click on edit.



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



Download Apache Spark™

1. Choose a Spark release:
2. Choose a package type:
3. Download Spark: [spark-2.4.7-bin-hadoop2.7.tgz](#)
4. Verify this release using the 2.4.7 [signatures](#), [checksums](#) and [project release KEYS](#).

Note that, Spark 2.x is pre-built with Scala 2.11 except version 2.4.2, which is pre-built with Scala 2.12. Spark 3.0+ is pre-built with Scala 2.12.

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"

← → ↻ 🏠 🔒 https://www.apache.org/dyn/closer.lua/spark/spark-2.4.7/spark-2.4.7-bin-hadoop2.7.tgz

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-2.4.7/spark-2.4.7-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

<https://downloads.apache.org/spark/spark-2.4.7/spark-2.4.7-bin-hadoop2.7.tgz>

BACKUP SITES

Please only use the backup mirrors to download KEYS, PGP signatures and hashes (SHA* etc) -- or if no other mirrors are working.

<https://downloads.apache.org/spark/spark-2.4.7/spark-2.4.7-bin-hadoop2.7.tgz>

The full listing of mirror sites is also available.

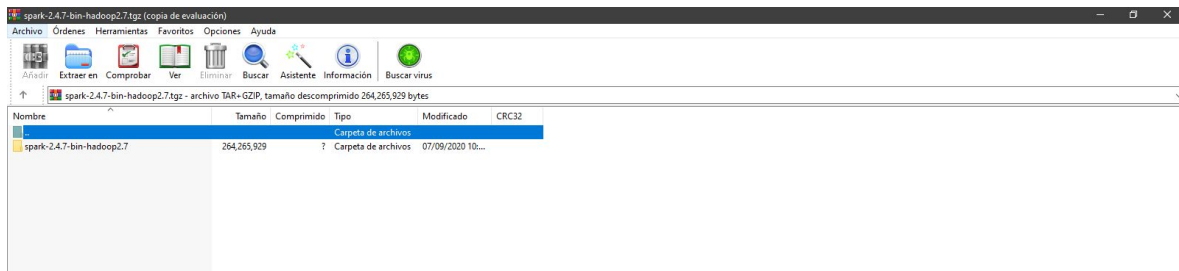
BECOMING A MIRROR

The procedure for setting up new mirrors is described in [How to become a mirror](#).

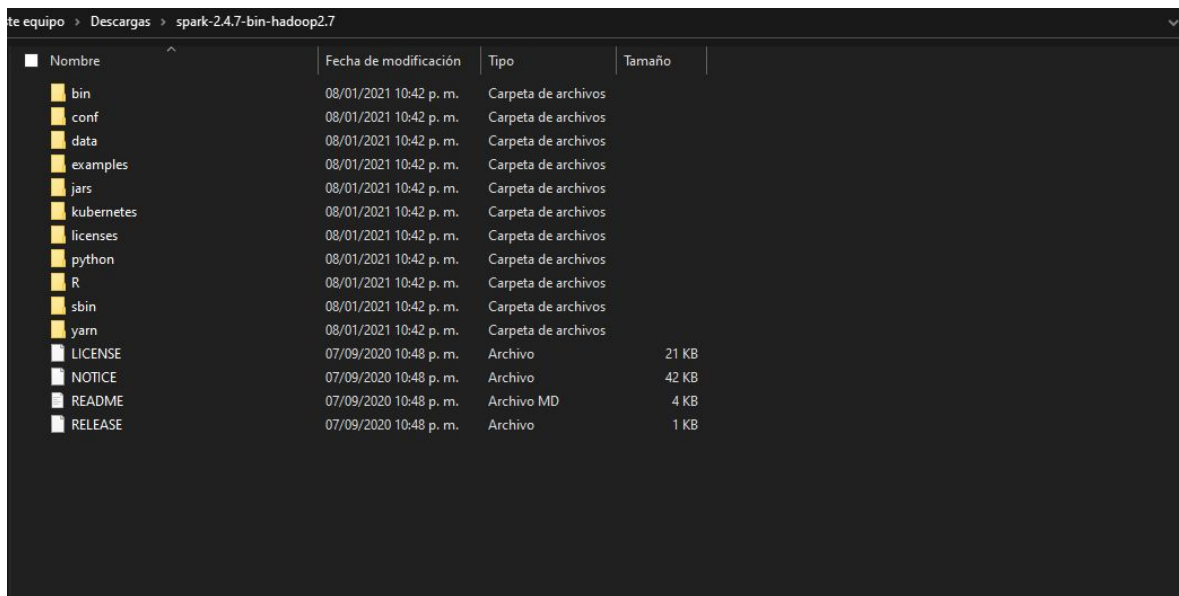
VERIFY THE INTEGRITY OF THE FILES

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 read [Verifying Apache Software Foundation Releases](#) for more information on why you should verify our releases.

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

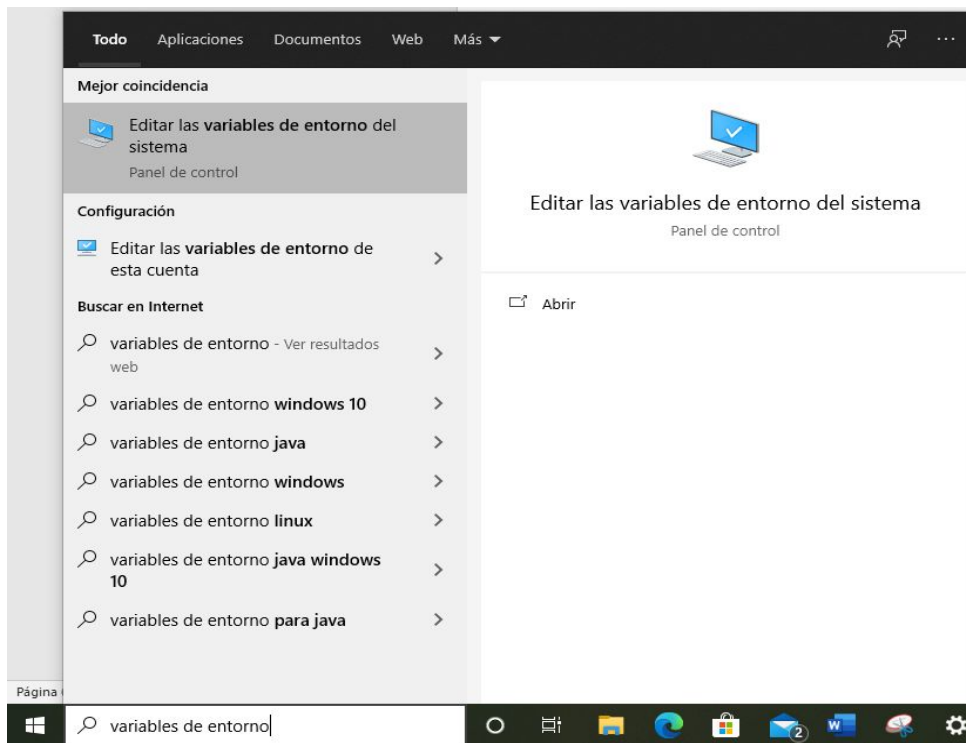


Once the file is downloaded, we must unzip the file.

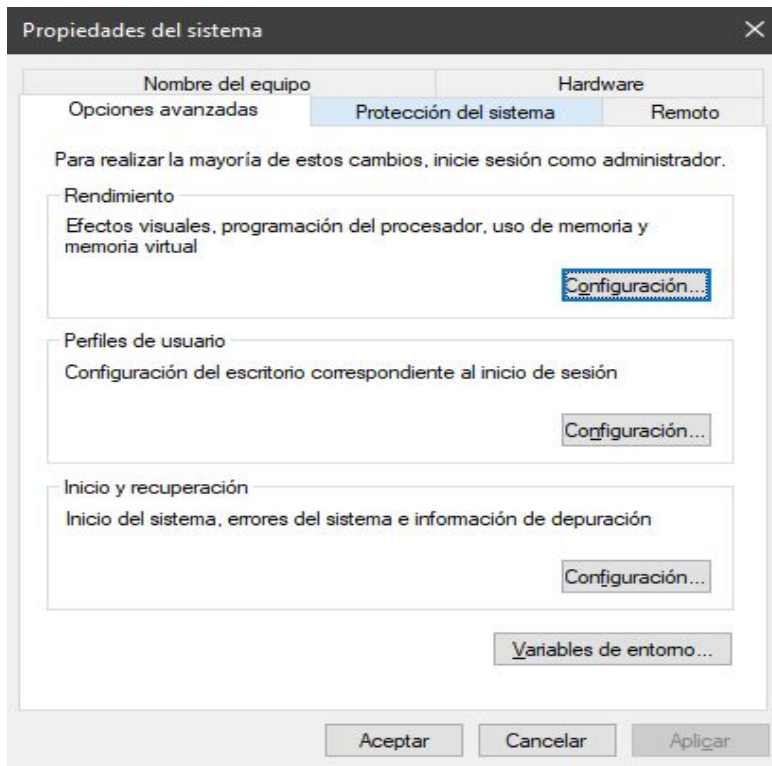


Once unzipped, we create a folder where we will save all the files in order to have a path for the environment variable.

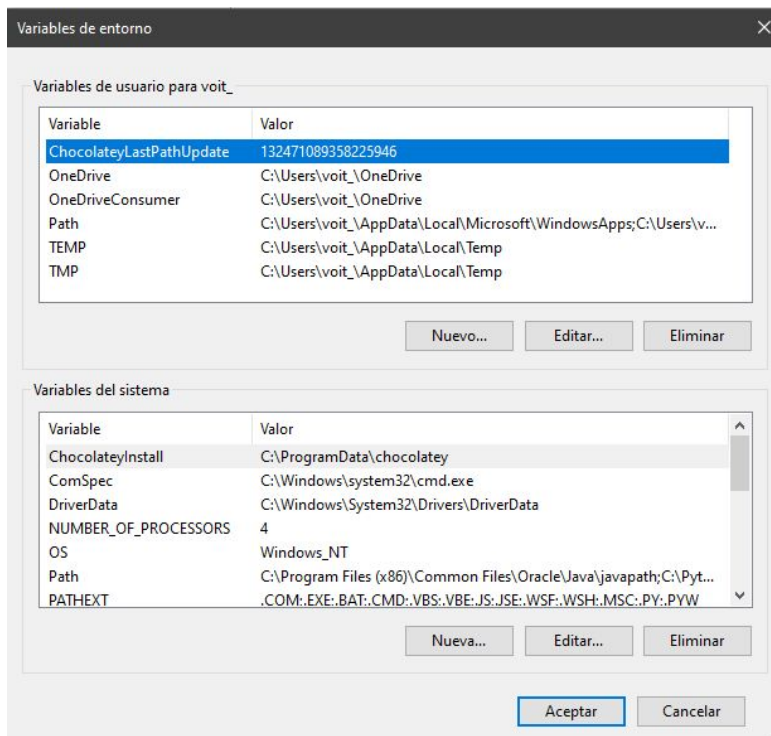
Now we must insert spark in the environment variables for the process, to do this, we must do the following.



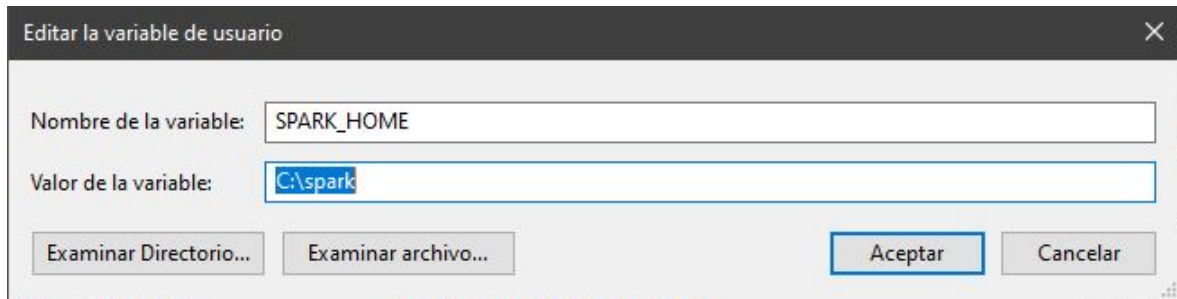
In the search bar, we write "Environment variables". And we select the option that says "Edit the system environment variables".



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.



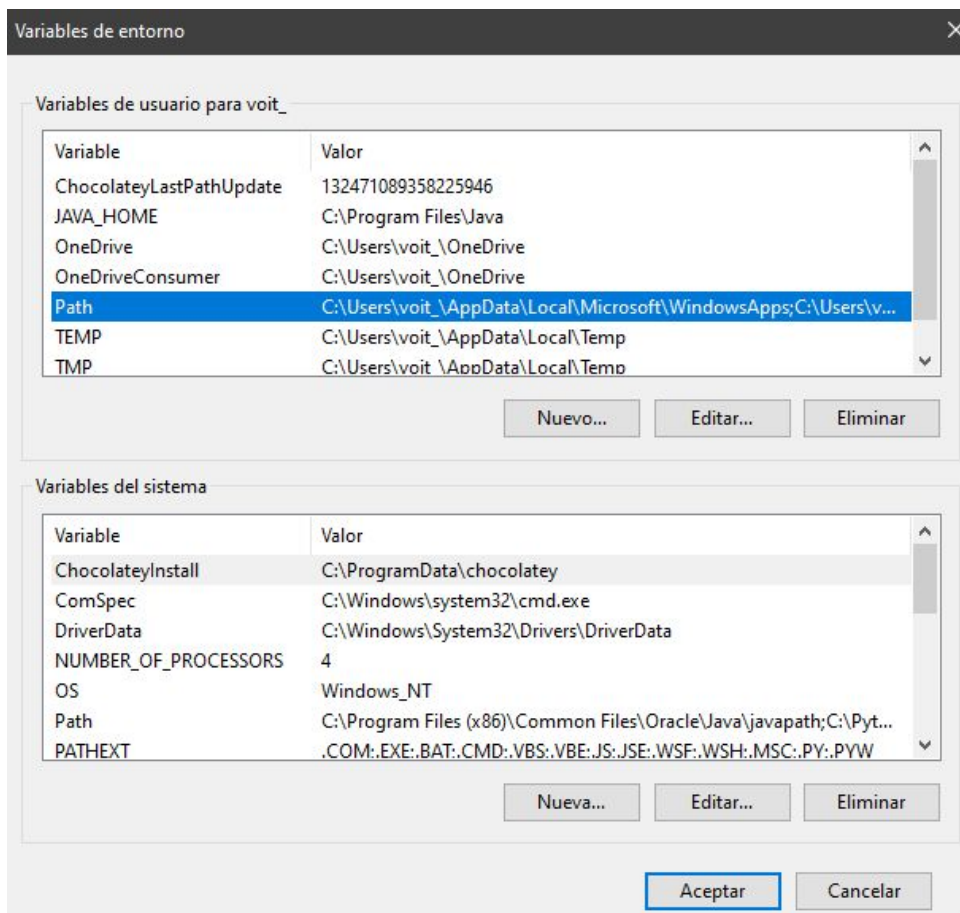
We are going to add the new environment variable to be able to use spark in any direction.

In the variable name section, we write the following:

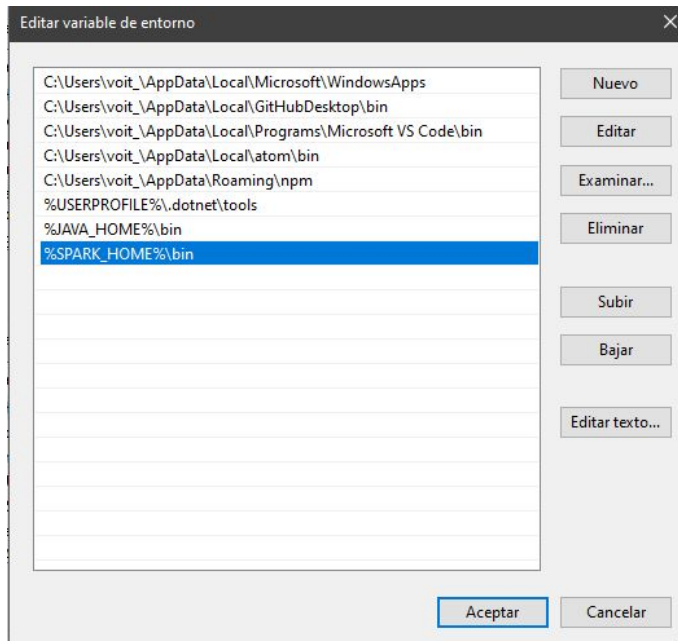
SPARK_HOME

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

And we click on "Accept"



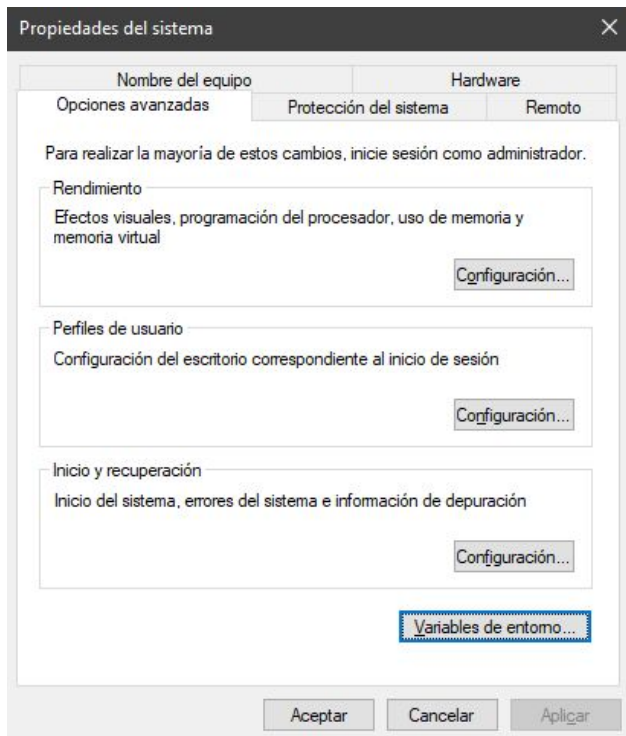
After this we must add the variable in the option "Path" for which we will select it and click on edit.



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

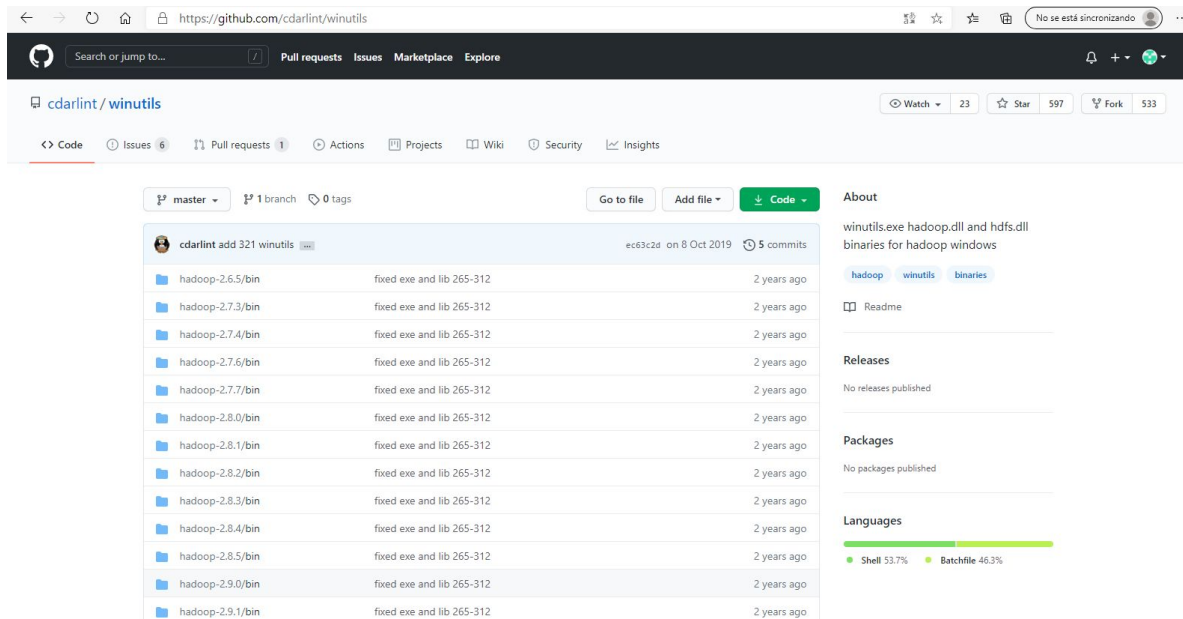
% SPARK_HOME% \ bin

And we will click on accept.



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

Install winutils.exe



cdarlint / winutils

master 1 branch 0 tags

Go to file Add file Code

File	Description	Commit
hadoop-2.6.5/bin	fixed exe and lib 265-312	2 years ago
hadoop-2.7.3/bin	fixed exe and lib 265-312	2 years ago
hadoop-2.7.4/bin	fixed exe and lib 265-312	2 years ago
hadoop-2.7.6/bin	fixed exe and lib 265-312	2 years ago
hadoop-2.7.7/bin	fixed exe and lib 265-312	2 years ago
hadoop-2.8.0/bin	fixed exe and lib 265-312	2 years ago
hadoop-2.8.1/bin	fixed exe and lib 265-312	2 years ago
hadoop-2.8.2/bin	fixed exe and lib 265-312	2 years ago
hadoop-2.8.3/bin	fixed exe and lib 265-312	2 years ago
hadoop-2.8.4/bin	fixed exe and lib 265-312	2 years ago
hadoop-2.8.5/bin	fixed exe and lib 265-312	2 years ago
hadoop-2.9.0/bin	fixed exe and lib 265-312	2 years ago
hadoop-2.9.1/bin	fixed exe and lib 265-312	2 years ago

About

winutils.exe hadoop.dll and hdfs.dll binaries for hadoop windows

hadoop winutils binaries

Readme

Releases

No releases published

Packages

No packages published

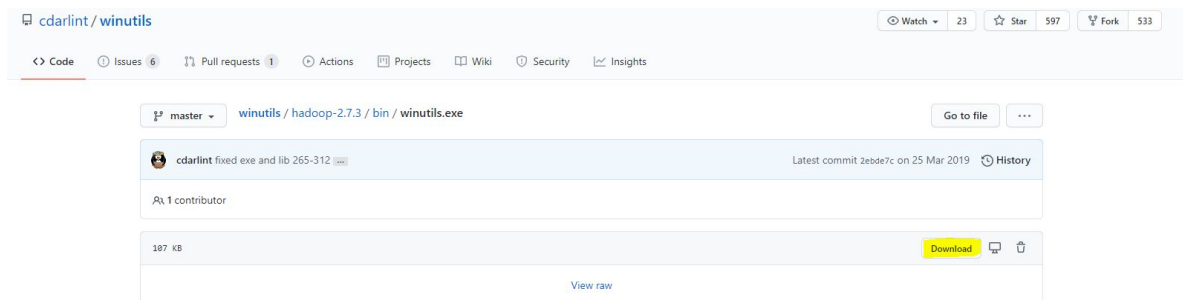
Languages

Shell 53.7% Batchfile 46.3%

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.



cdarlint / winutils

Watch 23 Star 597 Fork 533

Code Issues 6 Pull requests 1 Actions Projects Wiki Security Insights

master winutils / hadoop-2.7.3 / bin / winutils.exe

Go to file

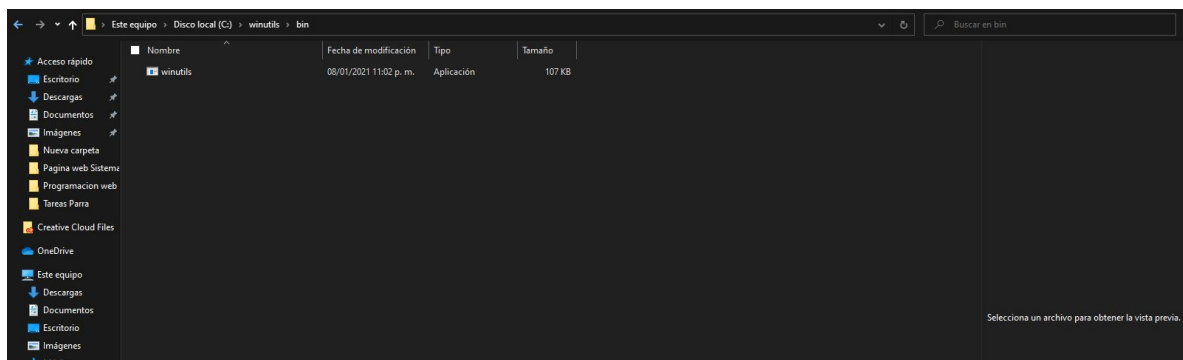
cdarlint fixed exe and lib 265-312 Latest commit zeb07c on 25 Mar 2019 History

1 contributor

107 KB Download

View raw

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.



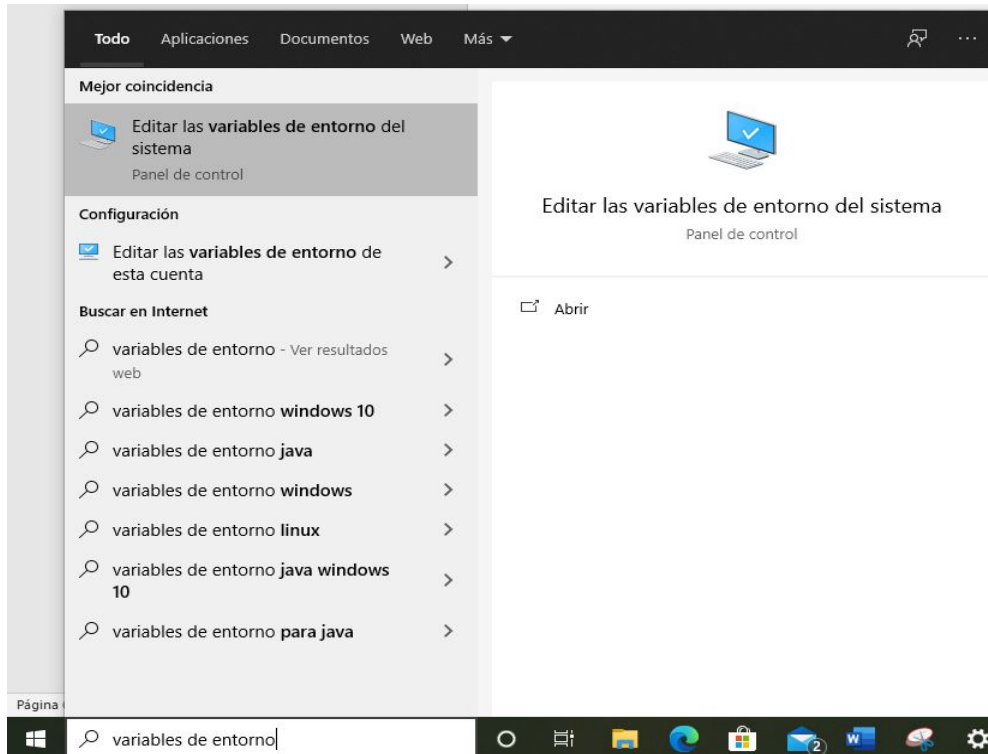
Este equipo > Disco local (C:) > winutils > bin

Nombre	Fecha de modificación	Tipo	Tamaño
winutils	08/01/2021 11:02 p. m.	Aplicación	107 KB

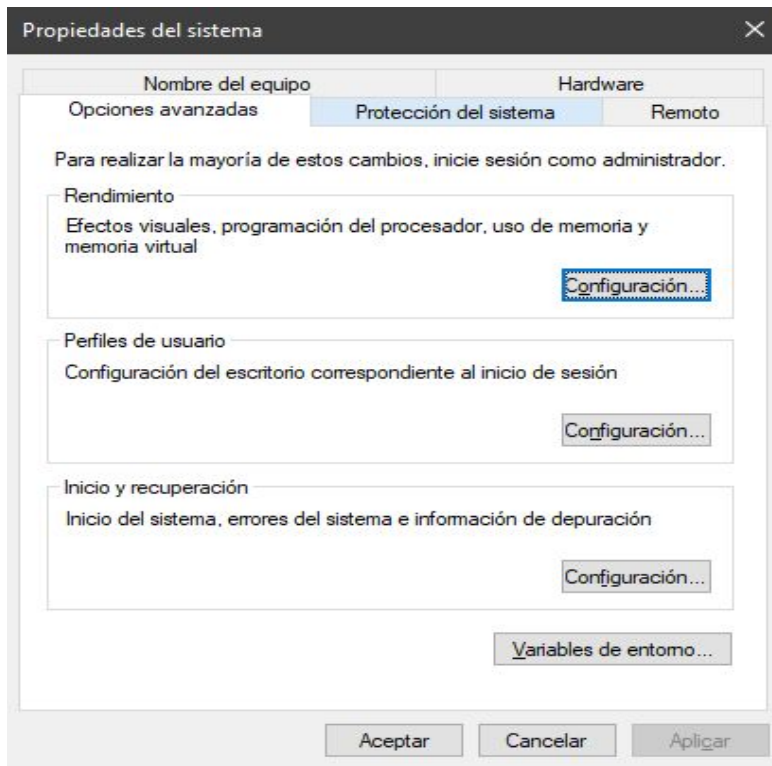
Selecciona un archivo para obtener la vista previa.

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.

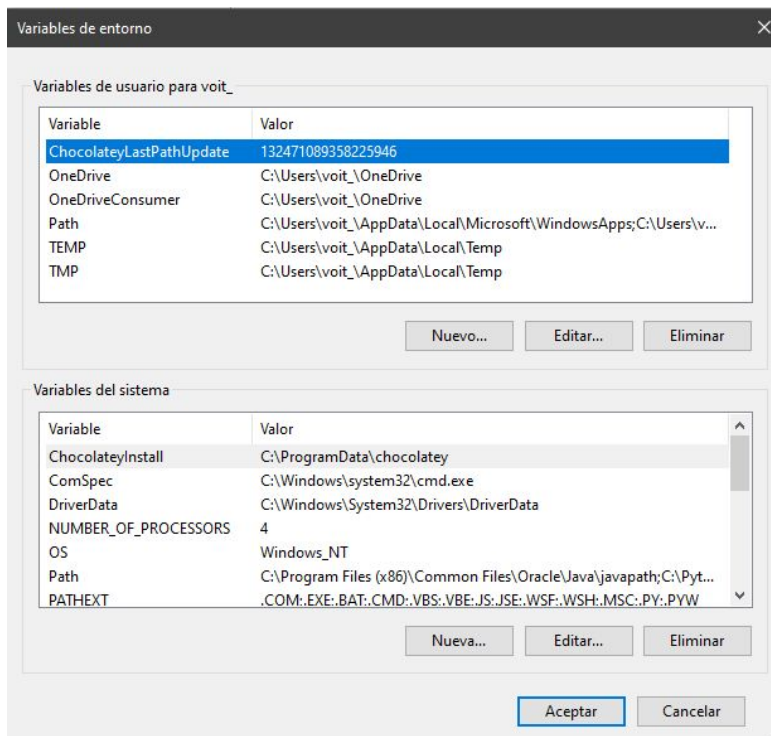
Now we must insert HADOOP in the environment variables for the process, to do this, we must do the following.



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.



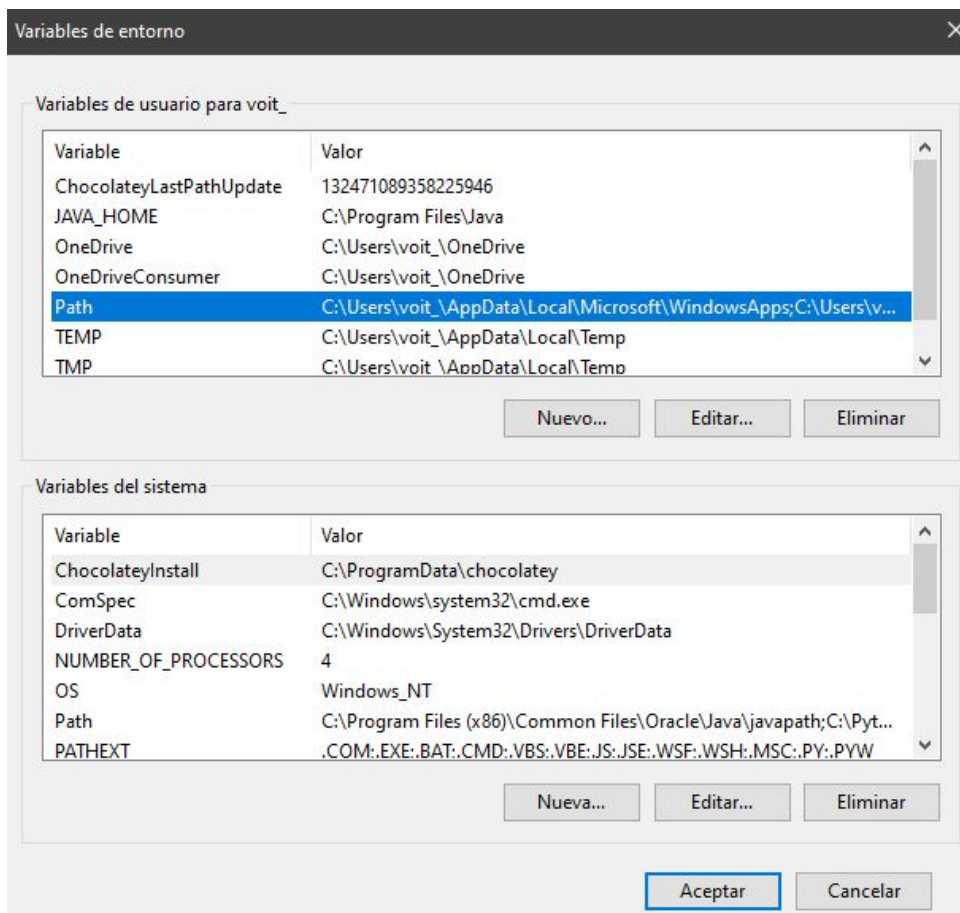
We are going to add the new environment variable to be able to use HADOOP in any direction.

In the variable name section, we write the following:

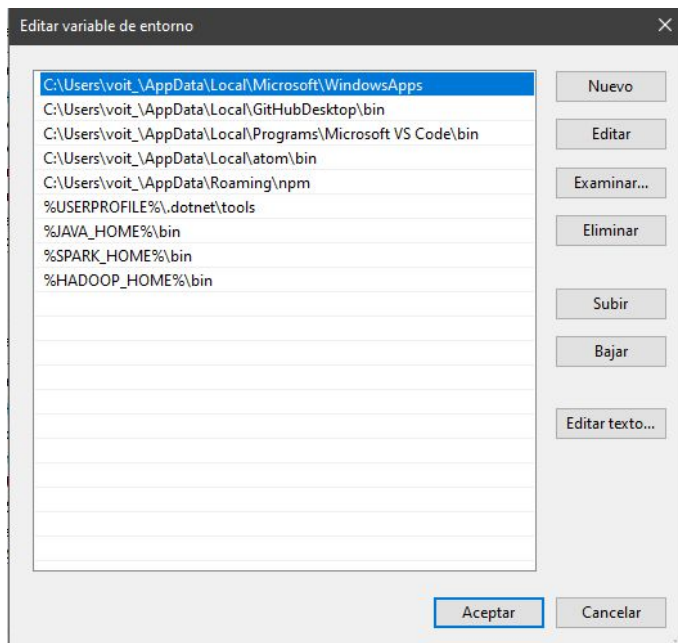
HADOOP_HOME

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

And we click on "Accept"



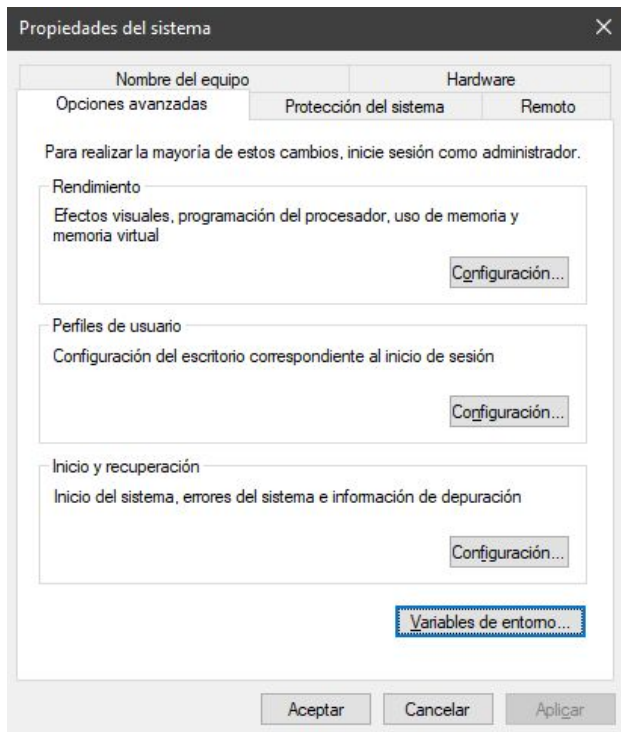
After this we must add the variable in the option "Path" for which we will select it and click on edit.



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.

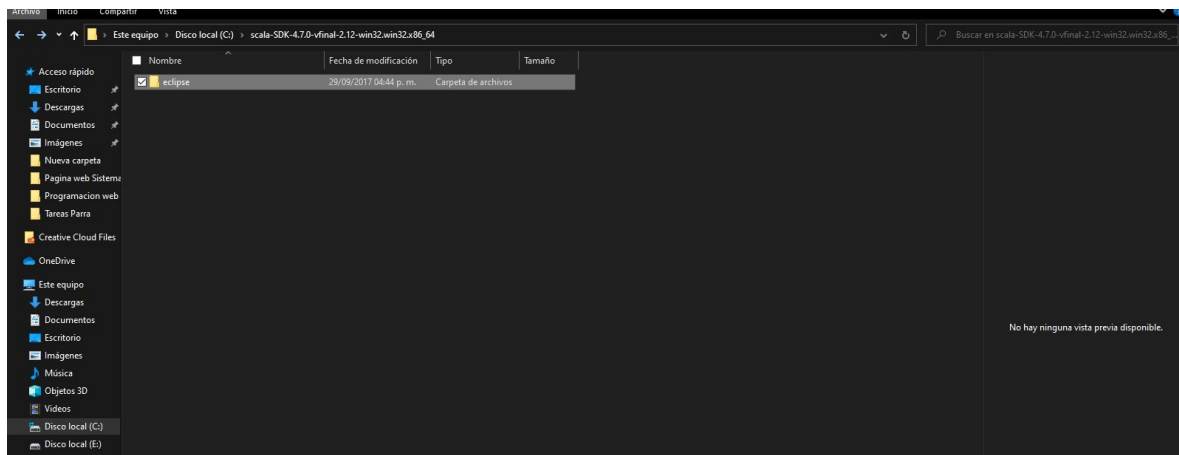
Install scala



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.

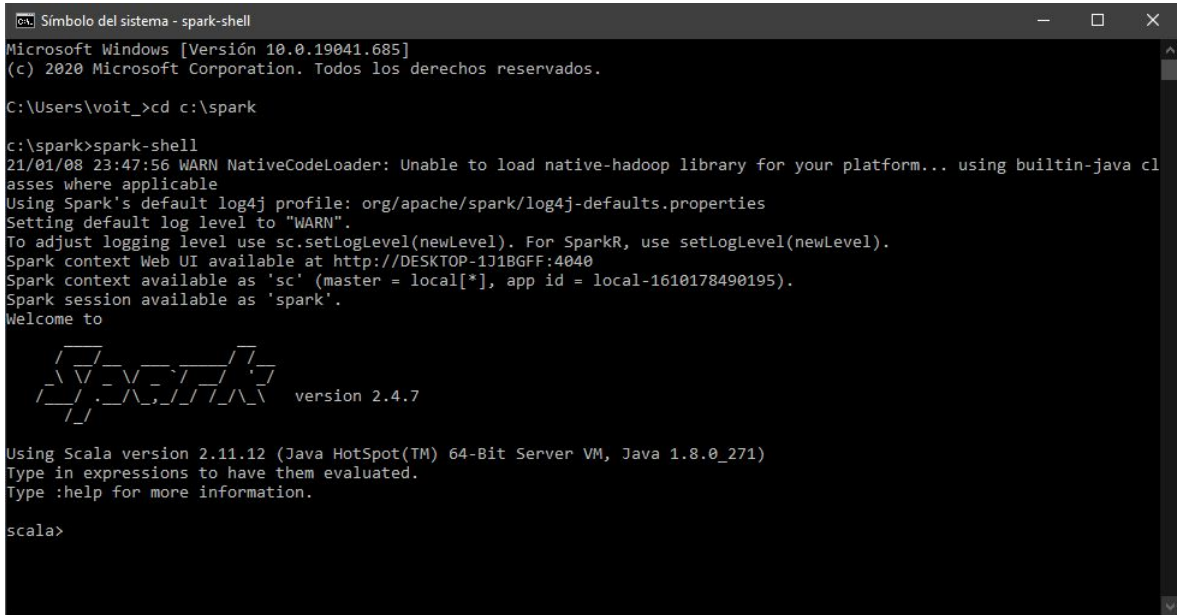
```
Símbolo del sistema - spark-shell
Microsoft Windows [Versión 10.0.19041.685]
(c) 2020 Microsoft Corporation. Todos los derechos reservados.

C:\Users\voit_>cd c:\spark

c:\spark>spark-shell
21/01/08 23:47:56 WARN NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Using Spark's default log4j profile: org/apache/spark/log4j-defaults.properties
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLevel).
```

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.



```
Símbolo del sistema - spark-shell
Microsoft Windows [Versión 10.0.19041.685]
(c) 2020 Microsoft Corporation. Todos los derechos reservados.

C:\Users\voit_>cd c:\spark

c:\spark>spark-shell
21/01/08 23:47:56 WARN NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
Using Spark's default log4j profile: org/apache/spark/log4j-defaults.properties
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLevel).
Spark context Web UI available at http://DESKTOP-1J1BGFF:4040
Spark context available as 'sc' (master = local[*], app id = local-1610178490195).
Spark session available as 'spark'.
Welcome to

  ____      __
 / ___ |    /  \
| |  \| |  /    \
| |__| | /  /\
|  __  |/  /  \
| |  | | /  /\
|_|  |_|/____\

version 2.4.7

Using Scala version 2.11.12 (Java HotSpot(TM) 64-Bit Server VM, Java 1.8.0_271)
Type in expressions to have them evaluated.
Type :help for more information.

scala>
```

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