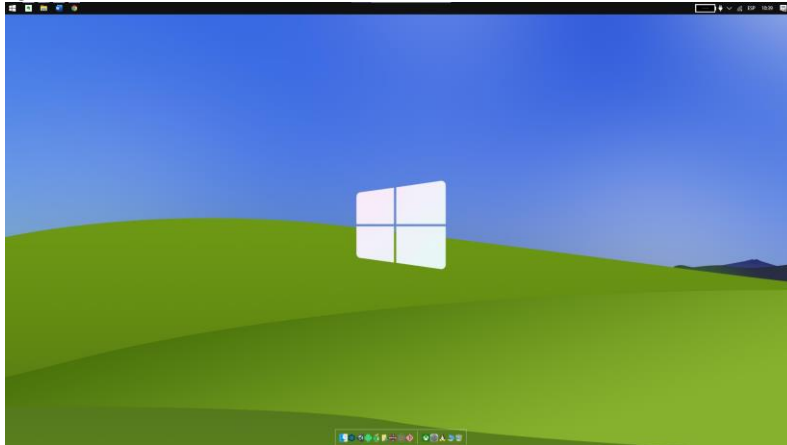


HADOOP 3.3.4:

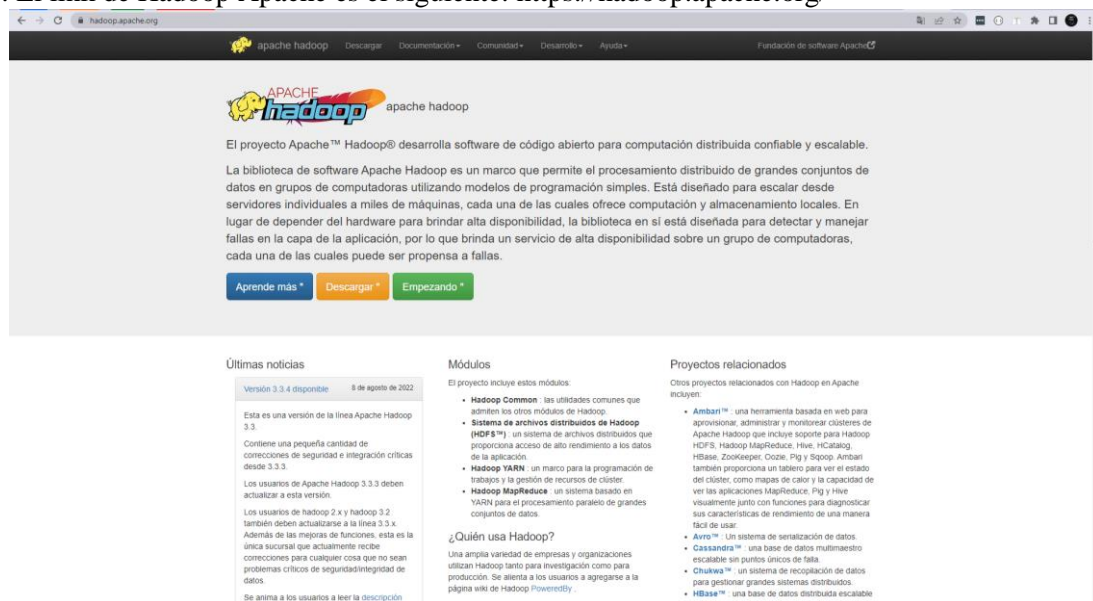
Guía Práctica de Instalación en Windows.

Brayan David Ahumada Castañeda. E-mail: brayan.ahumada@usantoto.edu.co. Cod: 2309109. Facultad de Ingeniería de Sistemas.
Juana Valentina Mendoza Santamaría. E-Mail: mailto:juana.mendoza@usantoto.edu.co. Cod: 2274474. Facultad de Ingeniería de Sistemas
Ing. Luis Felipe Narvaez Gomez. E-Mail: luis.narvaez@usantoto.edu.co. Cod: 2312660. Facultad de Ingeniería de Sistemas.

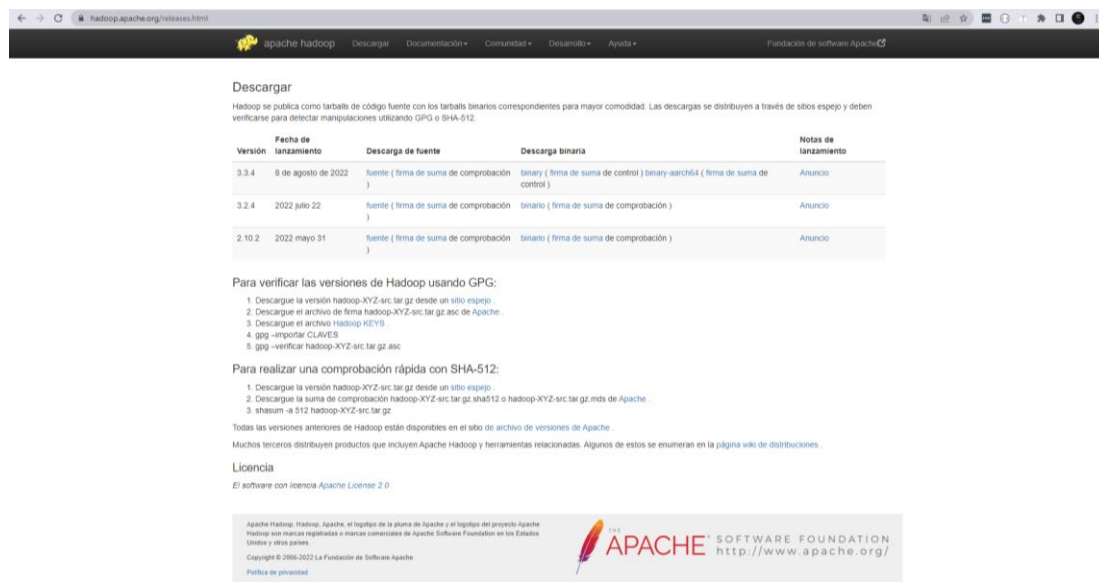
Para esta práctica instalaremos Hadoop en una Maquina Windows, en concreto la maquina en donde la instalaremos es un Windows 10 Home Single Basic.



Para instalar Hadoop, nos dirigiremos a la página principal de la distribución de este software de Apache, allí encontraremos documentación relacionado y claro, los links de descarga y diferentes versiones del programa que necesitamos. El link de Hadoop Apache es el siguiente: <https://hadoop.apache.org/>



Aquí nos dirigiremos a la parte de descarga, donde nos llevara a una nueva página en donde podremos descargar Hadoop para nuestra maquina en tres versiones disponibles, bajo dos formas de descarga.

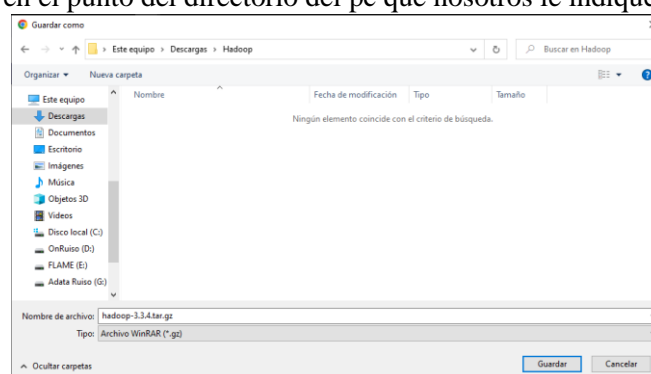


Como se especifica en la imagen anterior, hay dos formas de descargar e instalar Hadoop, por binarios o por fuente, a cada una se agrega una comprobación recomendada de los archivos con su guía. En mi caso selecciono la opción de Binarios para la versión (3.3.4)

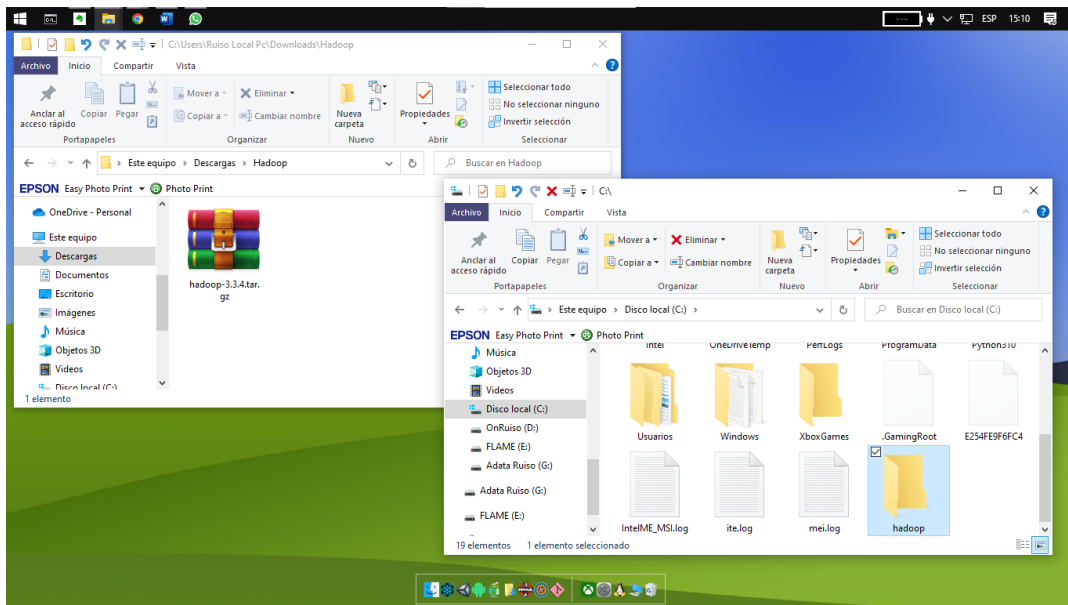
Para instalar Hadoop se recomienda tener preparado un sitio lo más cercano de la raíz de directorios de nuestro equipo, sin embargo, no hay ningún problema en principio con que lo instalemos en otra parte de nuestro equipo. Sin embargo, algo que si se le sugiere a todo usuario que quiera instalar hadoop en su Equipo es que la dirección de rutas, los nombres de las carpetas, no contengan espacios, evitando así errores al llamar algún archivo o alguna configuración. En mi caso la ruta de instalación será la siguiente: C:\hadoop



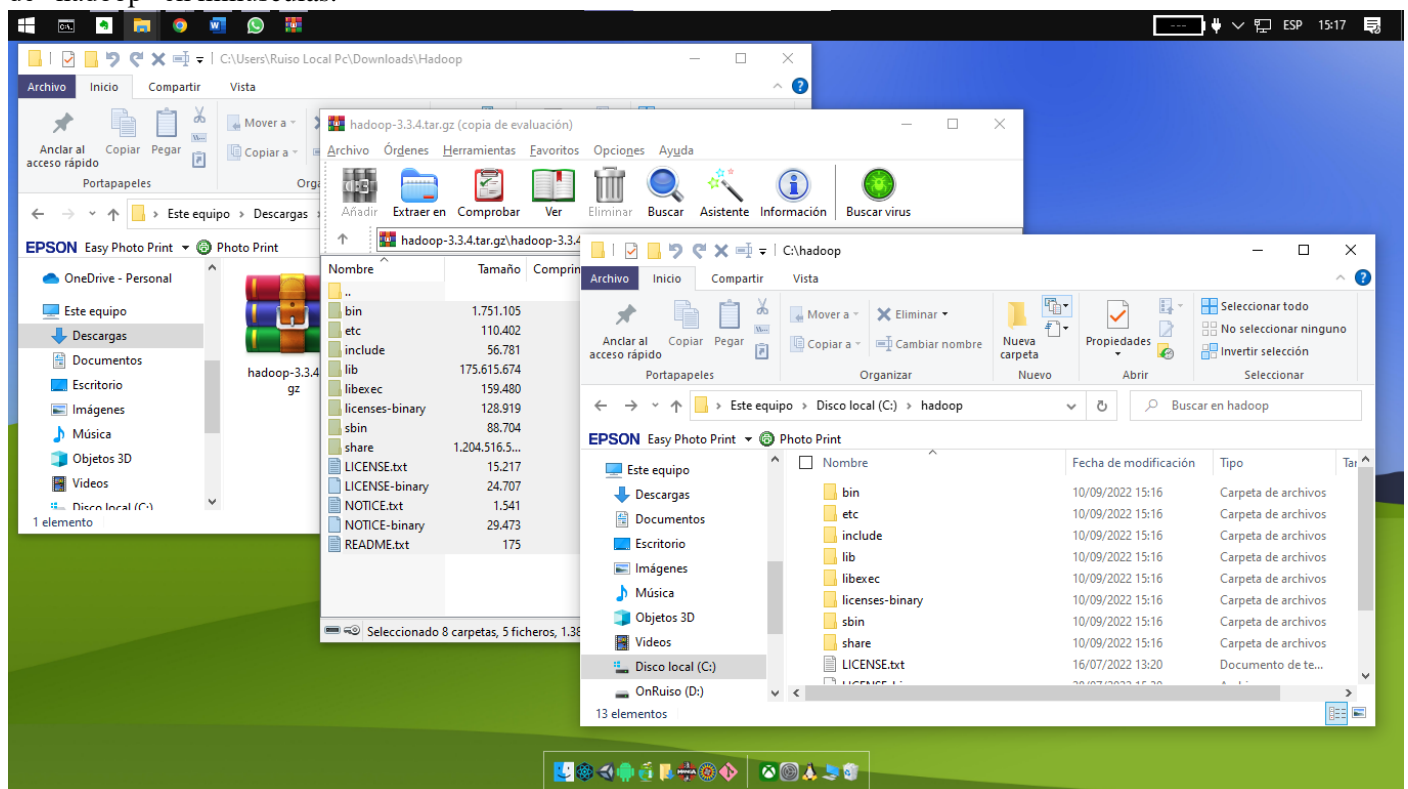
En el link de descarga seleccionaremos el enlace sugerido que termina con extensión "hadoop-3.3.4.tar.gz". Al darle clic se nos iniciara la descarga en el punto del directorio del pc que nosotros le indiquemos.



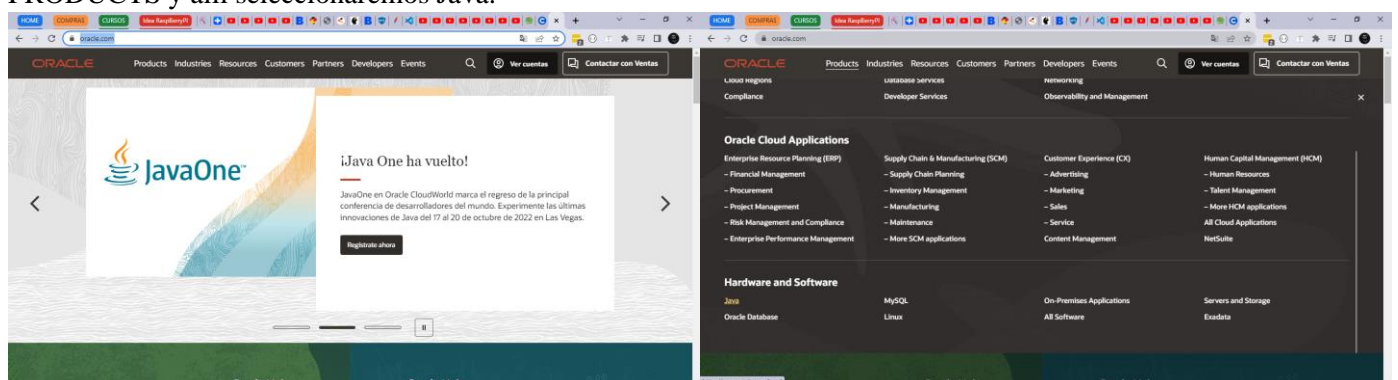
En esta instalación mantendremos nuestra carpeta de descarga del sistema y el sitio donde queremos instalar Hadoop abiertos. Es recomendable manejar los archivos de Hadoop de forma descomprimida.



La extracción de los archivos lo copiaremos en nuestra carpeta destino, la misma debe tener obligatoriamente el nombre de “hadoop” en minúsculas.



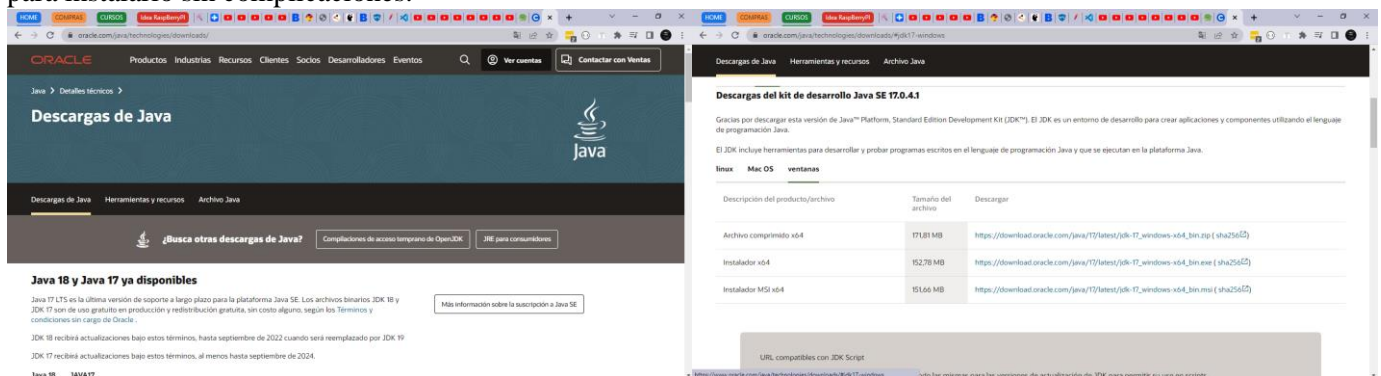
Antes de continuar con la instalación propia de Hadoop debemos tener en nuestro sistema el JDK de Java. Para esto iremos al enlace: <https://www.oracle.com/> . Aquí seleccionaremos en la barra de menus de la pagina la parte de PRODUCTS y allí seleccionaremos Java.



Se abrirá una nueva carpeta donde ahora daremos clic en el botón que dice Downloads o Descarga.

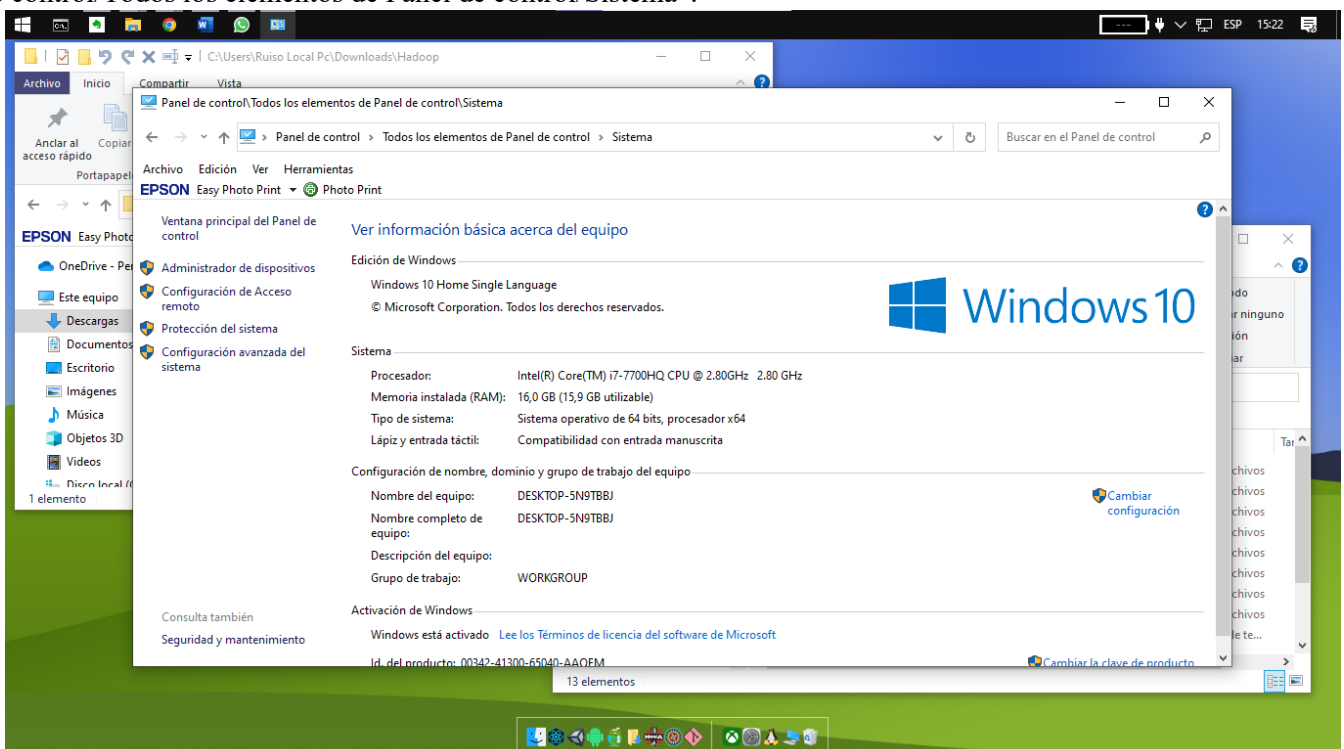


Ahora nos encontraremos en una pagina donde podremos seleccionar la versión que mas se nos ajuste de JDK de Java para nuestros propósitos, aquí recomendamos instalar la penúltima, la cual presenta siempre mas estabilidad que otras mas recientes. En mi caso es el JDK 17 de Java, para Windows de x64 de Arquitectura y el enlace de descarga tipo MSI para instalarlo sin complicaciones.

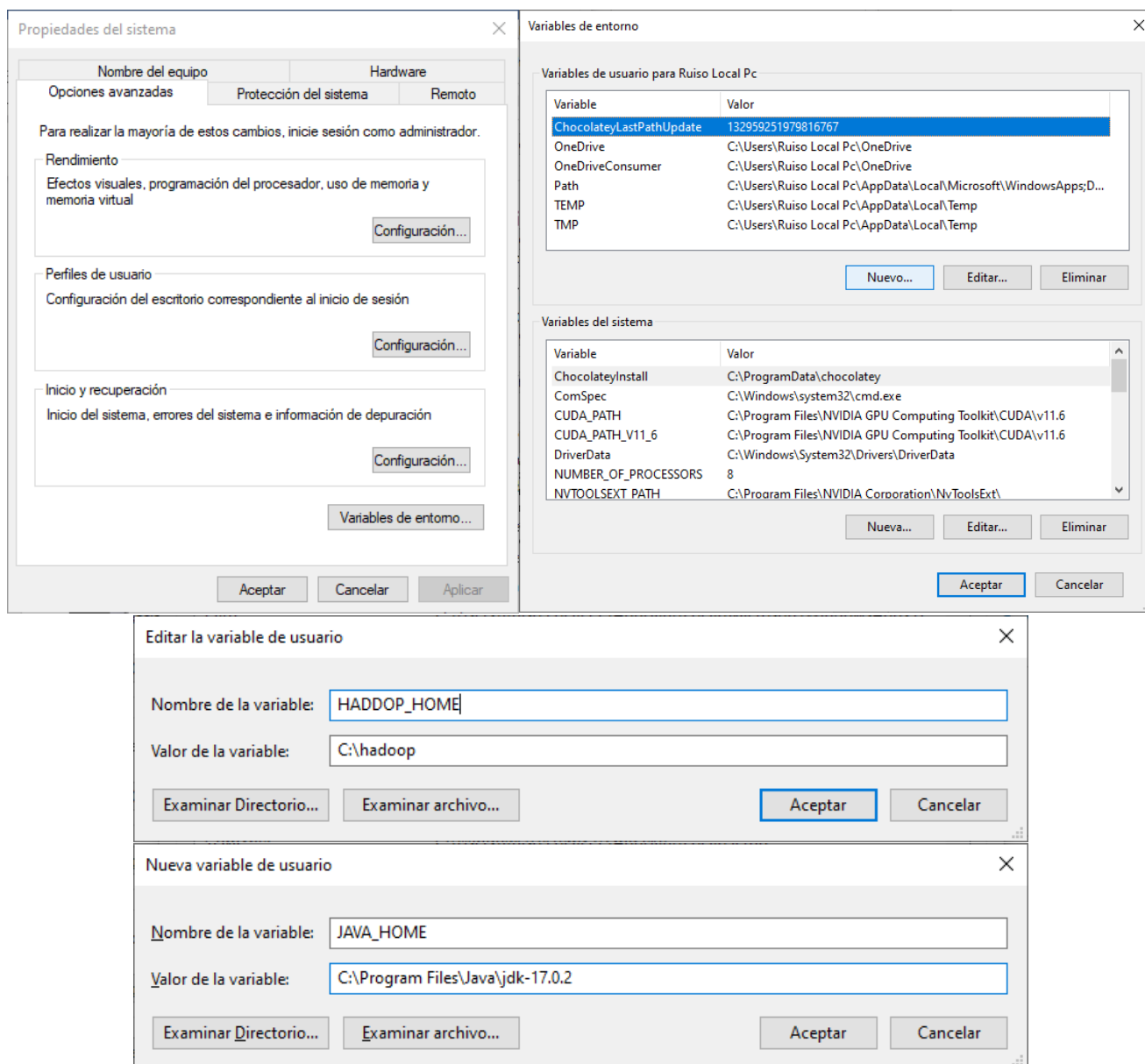


La instalación de Java es sencilla, es dar clic en siguiente de las ventanas de la interfaz de instalación y se recomienda siempre dejarlo en la ruta preestablecida. Sigamos con Hadoop.

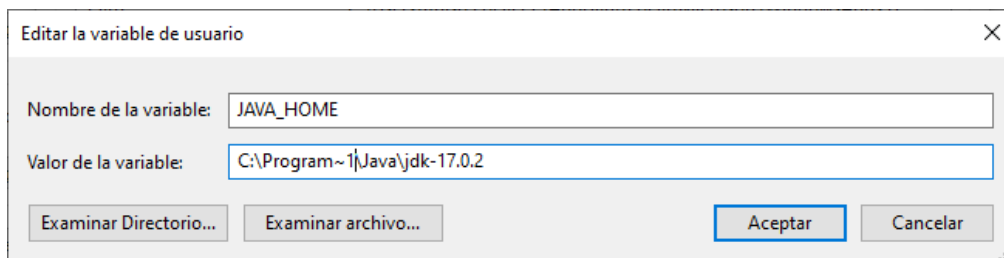
Ahora abriremos las propiedades del sistema, para esto basta con estar en un explorador de archivos de Windows, dar clic derecho sobre el Disco Local C o donde tengamos instalado Windows (el cual es el disco Raíz del sistema operativo) y seleccionar en el menú de herramientas desplegado la opción de “Propiedades”. En caso de no tener esta unidad asociada, aun siendo la raíz, se puede dar el doble clic sobre “Este Equipo” o ir a la dirección de Panel de control “Panel de control/Todos los elementos de Panel de control/Sistema”.



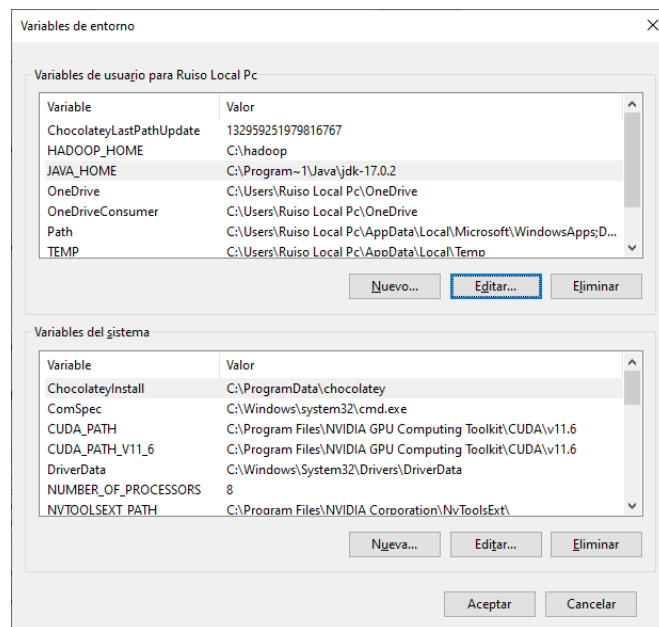
Allí seleccionaremos en la barra lateral la opción de “Configuración Avanzada del Sistema”. Seleccionamos “variables de entorno”. Creamos dos nuevas variables de entorno seleccionando la opción de “Nuevo...”. Las variables son:



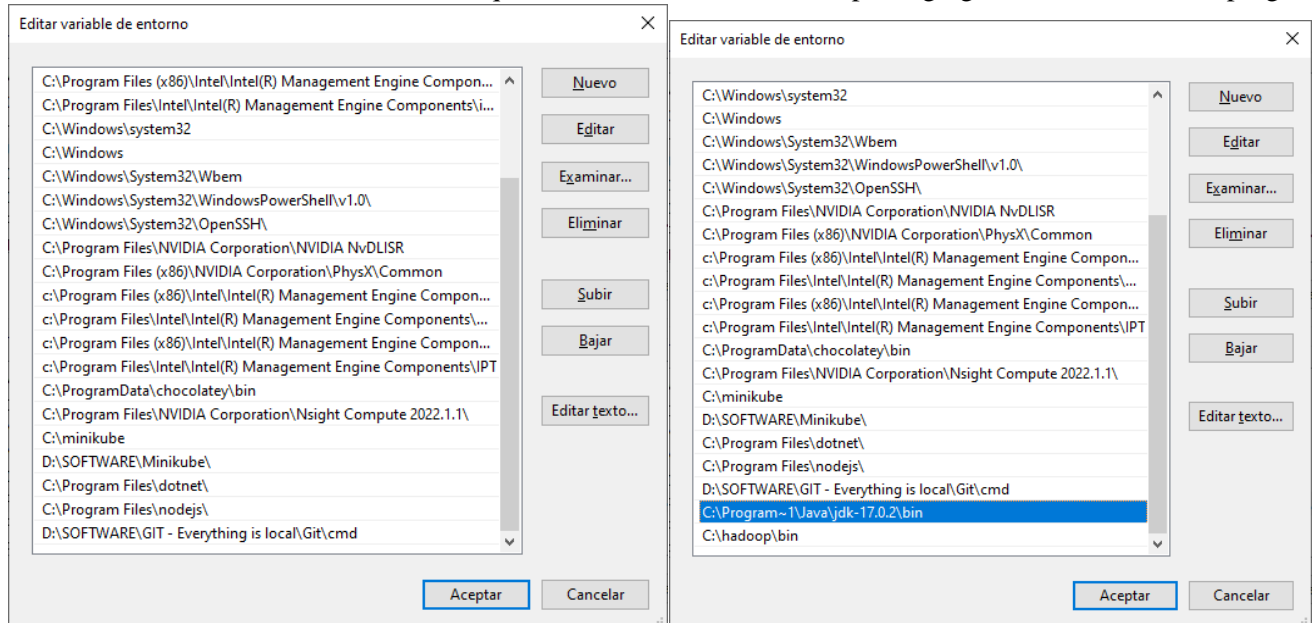
De la anterior dirección debemos cambiar el espacio que hay en el nombre de la carpeta “Program Files” esto lo arreglamos de la siguiente manera:



De esta manera obtenemos un resultado similar a lo siguiente:

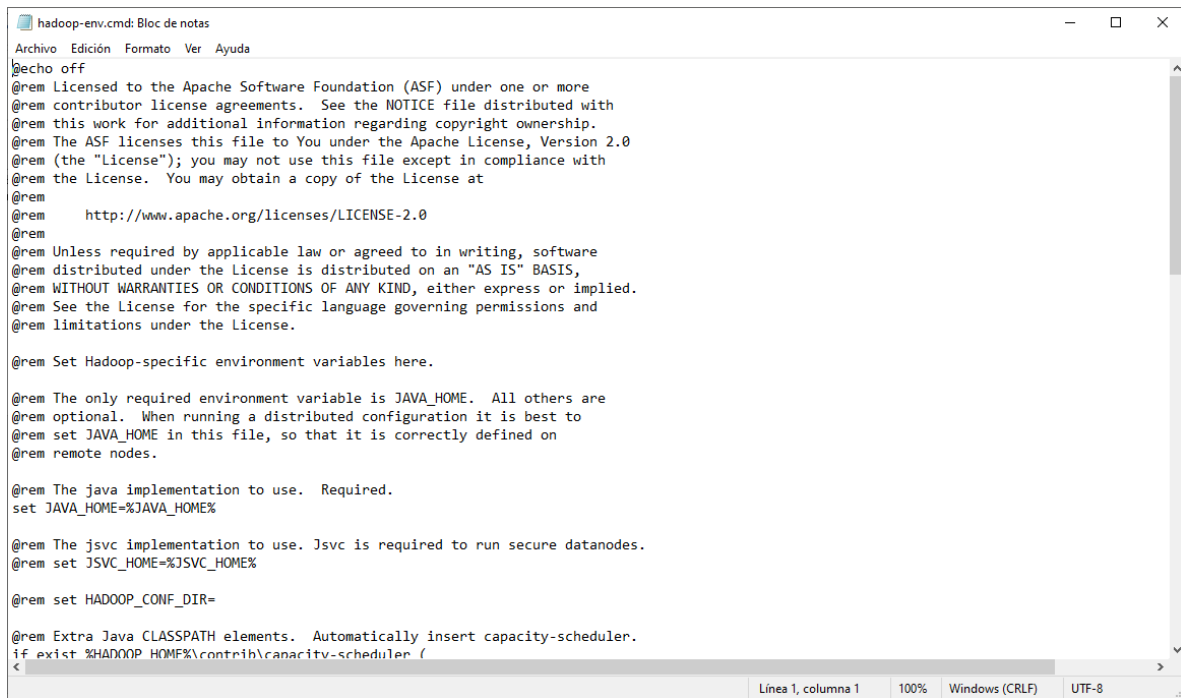


Ahora vamos a la parte de “Variables del Sistema” y generamos dos nuevas variantes, para esto buscamos la que se llame “Path”, la abrimos con un doble clic izquierdo. Seleccionamos nuevo para agregar las rutas de ambos programas.



Solo queda aceptar todos los cambios que hemos realizado en las variables de entorno y del sistema. Ahora nos dirigiremos a nuestra carpeta de Hadoop y haremos lo siguiente.

1. Crear la carpeta: C:\hadoop\data
2. Ir a la dirección: C:\hadoop\etc\hadoop y abrir con bloc de notas o clic derecho y opción editar “hadoop-env.cmd”



```
hadoop-env.cmd: Bloc de notas
Archivo Edición Formato Ver Ayuda
@echo off
@rem Licensed to the Apache Software Foundation (ASF) under one or more
@rem contributor license agreements. See the NOTICE file distributed with
@rem this work for additional information regarding copyright ownership.
@rem The ASF licenses this file to You under the Apache License, Version 2.0
@rem (the "License"); you may not use this file except in compliance with
@rem the License. You may obtain a copy of the License at
@rem
@rem http://www.apache.org/licenses/LICENSE-2.0
@rem
@rem Unless required by applicable law or agreed to in writing, software
@rem distributed under the License is distributed on an "AS IS" BASIS,
@rem WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
@rem See the License for the specific language governing permissions and
@rem limitations under the License.
@rem Set Hadoop-specific environment variables here.
@rem The only required environment variable is JAVA_HOME. All others are
@rem optional. When running a distributed configuration it is best to
@rem set JAVA_HOME in this file, so that it is correctly defined on
@rem remote nodes.
@rem The java implementation to use. Required.
set JAVA_HOME=%JAVA_HOME%
@rem The jsvc implementation to use. Jsvc is required to run secure datanodes.
@rem set JSVC_HOME=%JSVC_HOME%
@rem set HADOOP_CONF_DIR=
@rem Extra Java CLASSPATH elements. Automatically insert capacity-scheduler.
if exist %HADOOP_HOME%\contrib\capacity-scheduler (
<
```

Aquí nos dirigimos a una parte del archivo iniciada con SET JAVA.

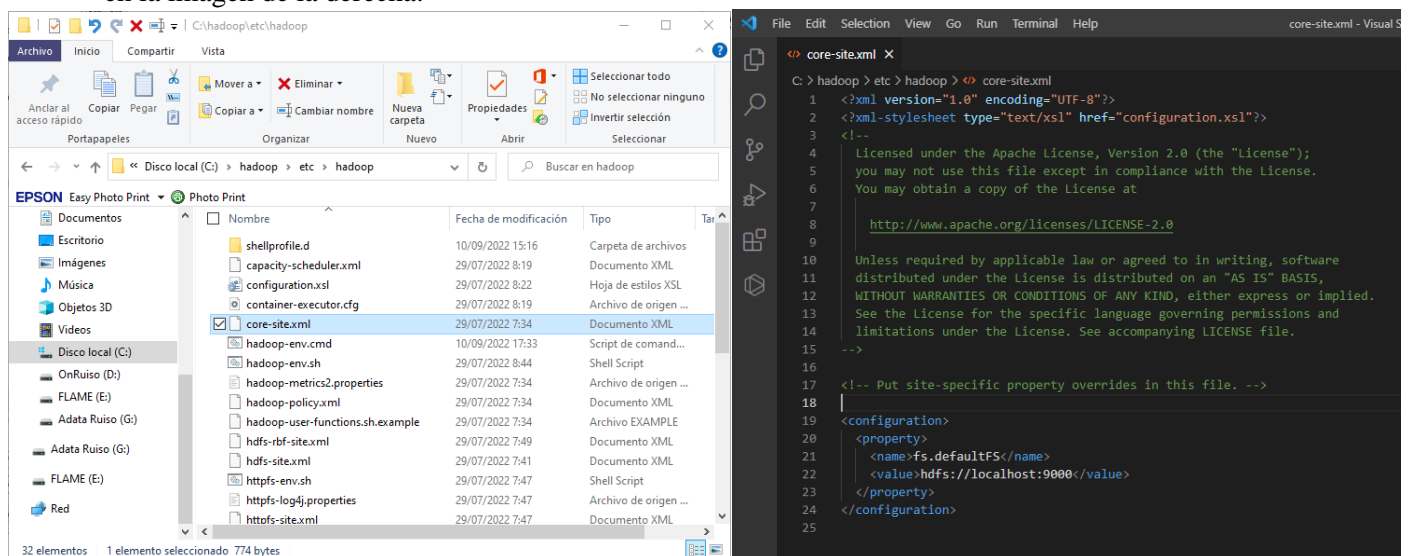
```
@rem The java implementation to use. Required.
set JAVA_HOME=%JAVA_HOME%
```

Comentaríamos esta línea y cambiamos por:

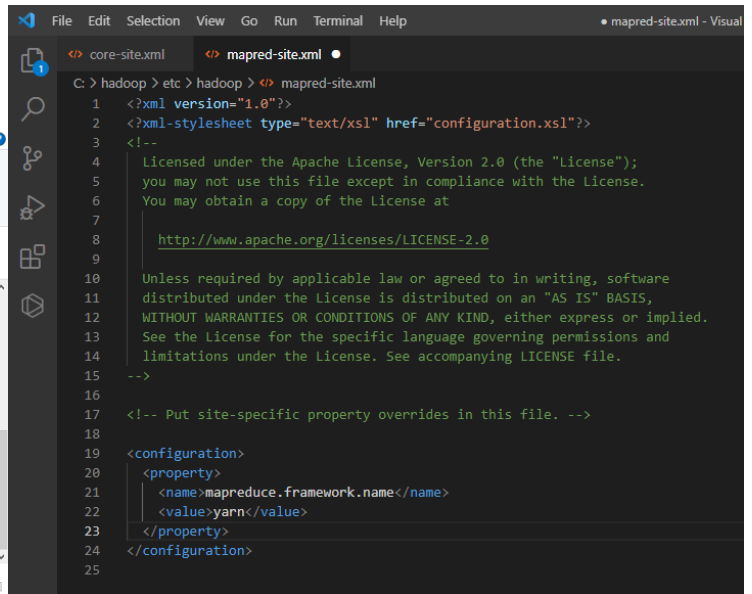
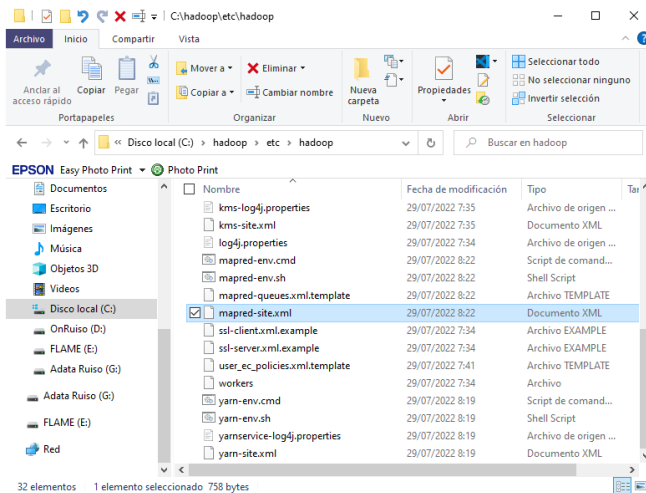
```
@rem The java implementation to use. Required.
@rem set JAVA_HOME=%JAVA_HOME%
set JAVA_HOME="C:\Program~1\Java\jdk-17.0.2"
```

Guardamos [CTRL]+[G] y salimos del archivo. Ahora transformaremos unos archivos de Hadoop que corresponden a su arranque en las paginas web que veremos al iniciar Hadoop, para ir a estos archivos deberas entrar a la carpeta de instalación de Hadoop, entrar al directorio ETC y luego al directorio HADOOP es decir una ruta similar a /HADOOP/ETC/HADOOP/, estos archivos son:

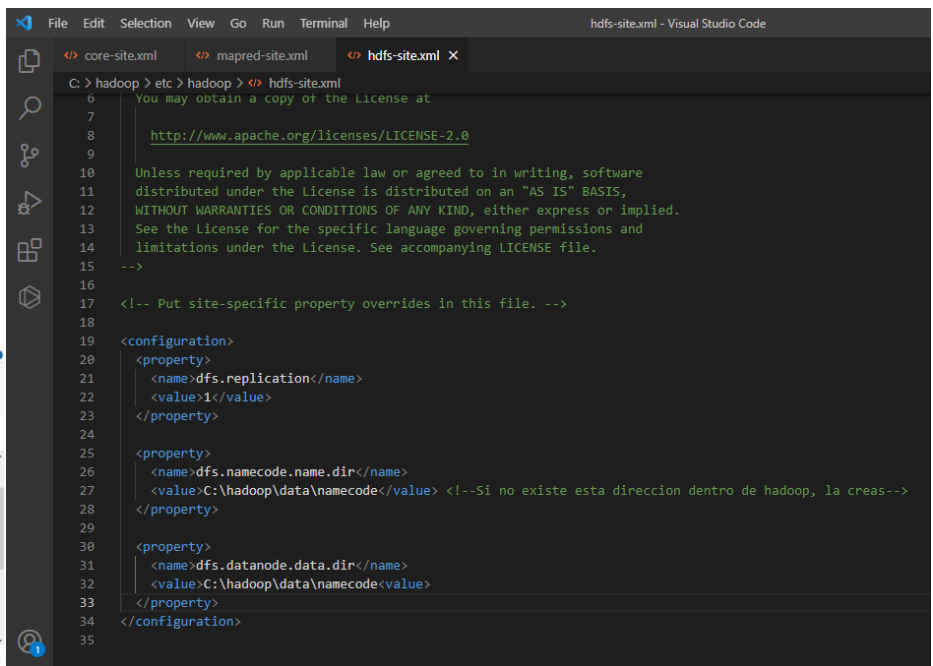
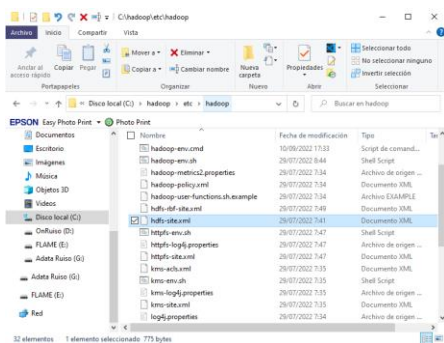
1. Core-site.xml, el código que deberas cambiar esta en la etiqueta de Configuración y debe lucir igual al mostrado en la imagen de la derecha.



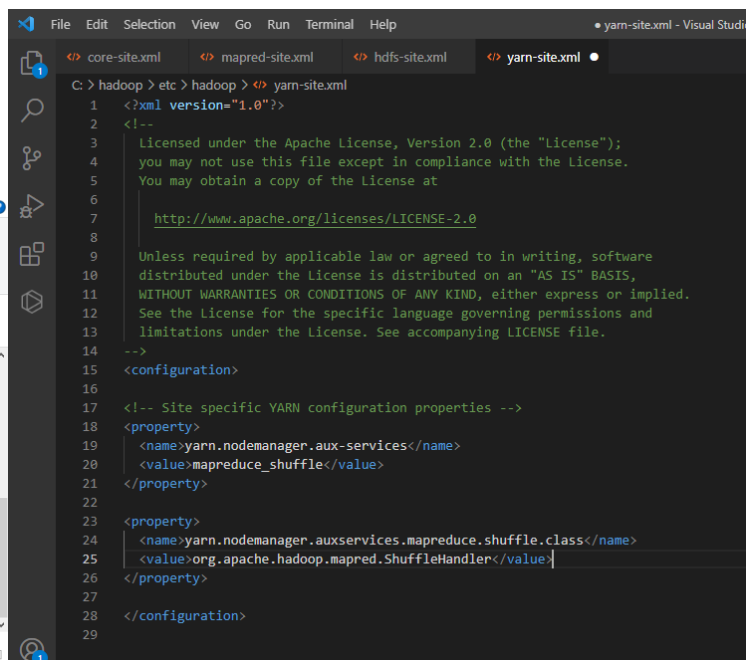
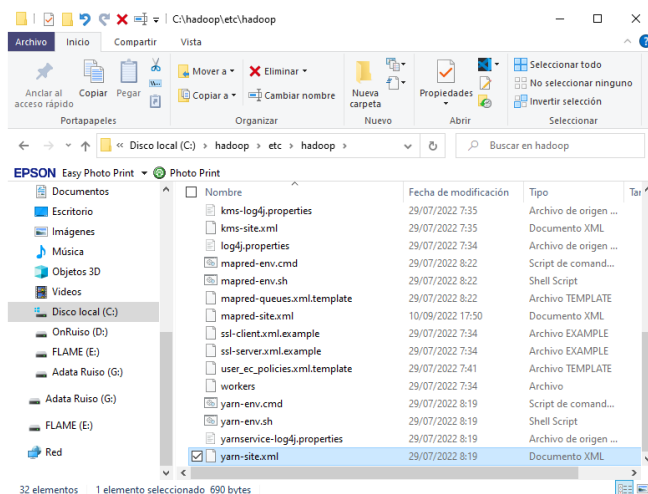
2. Mapred-site.xml, el código que se debe cambiar es el que esta encerrado en las etiquetas de configuracion, tal y como se muestra en la figura de la derecha.



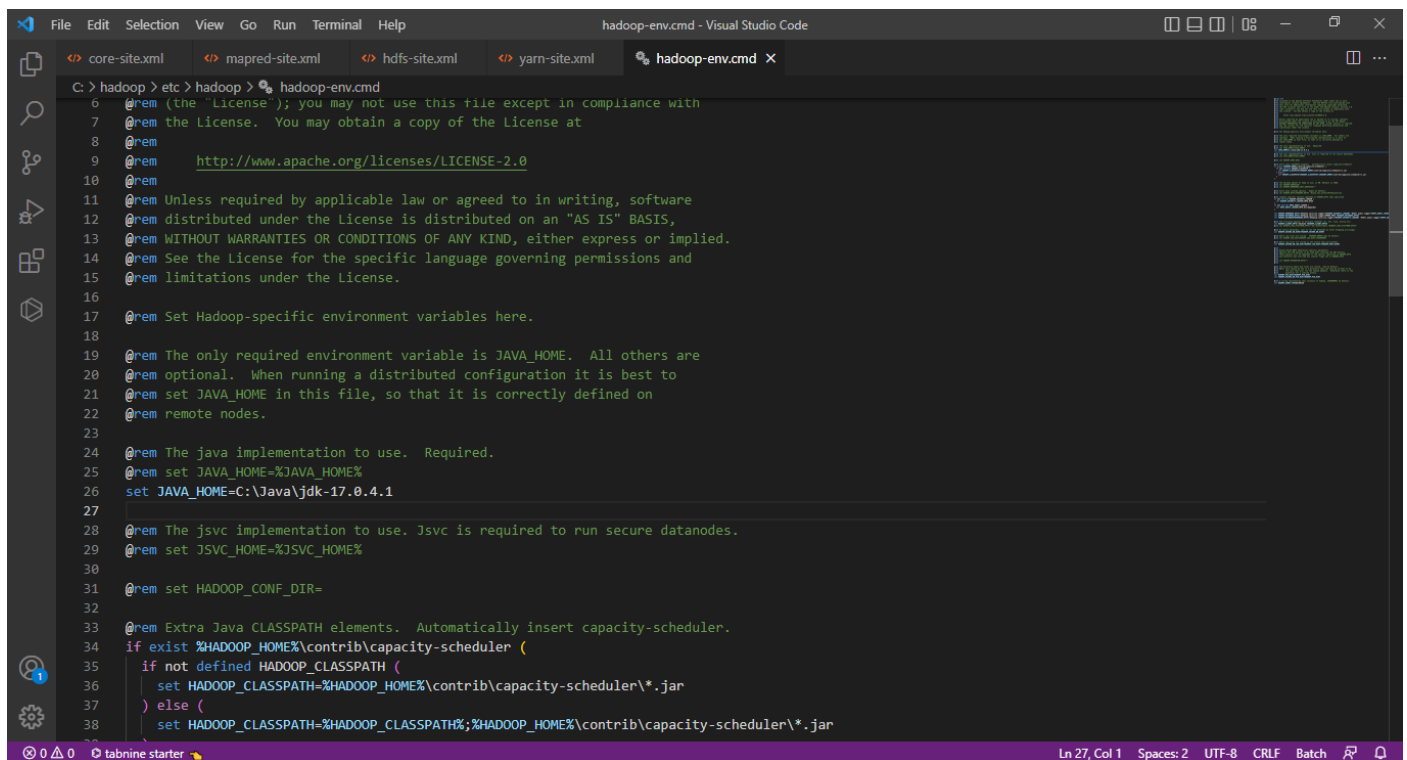
3. Hdfs-site.xml, El código que se debe cambiar es el que esta encerrado en las etiquetas de configuración, las tres propiedades teniendo especial cuidado en que las direcciones que se muestren ahí coincidan con las que hay en su computadora. En caso de no ser así, se deberá obtener la ruta real correspondiente y de no existir los directorios, los mismos se deberán crear.



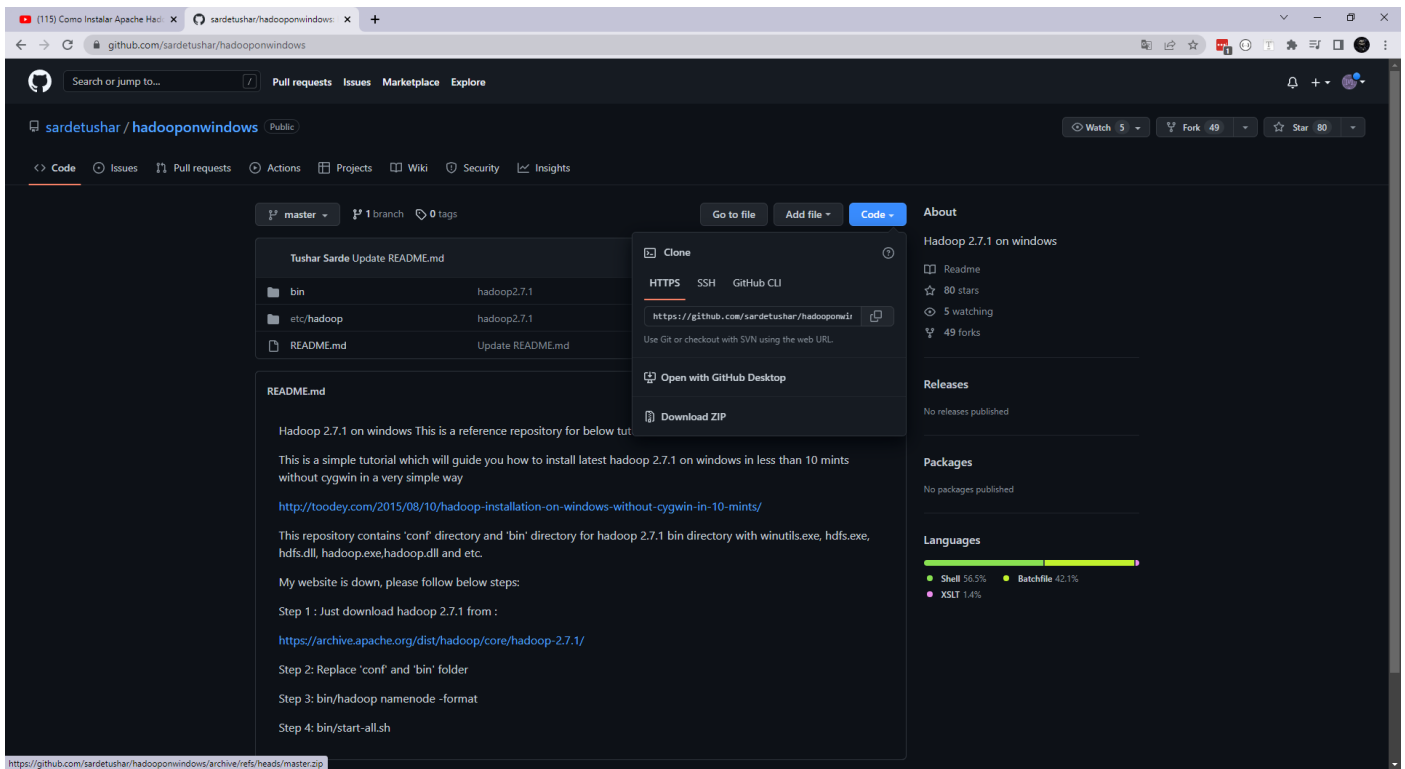
4. Yarn-site.xml, Se debera cambiar las propiedades internas a las etiquetas de configuracion, tal y como se ve en la imagen de la derecha.



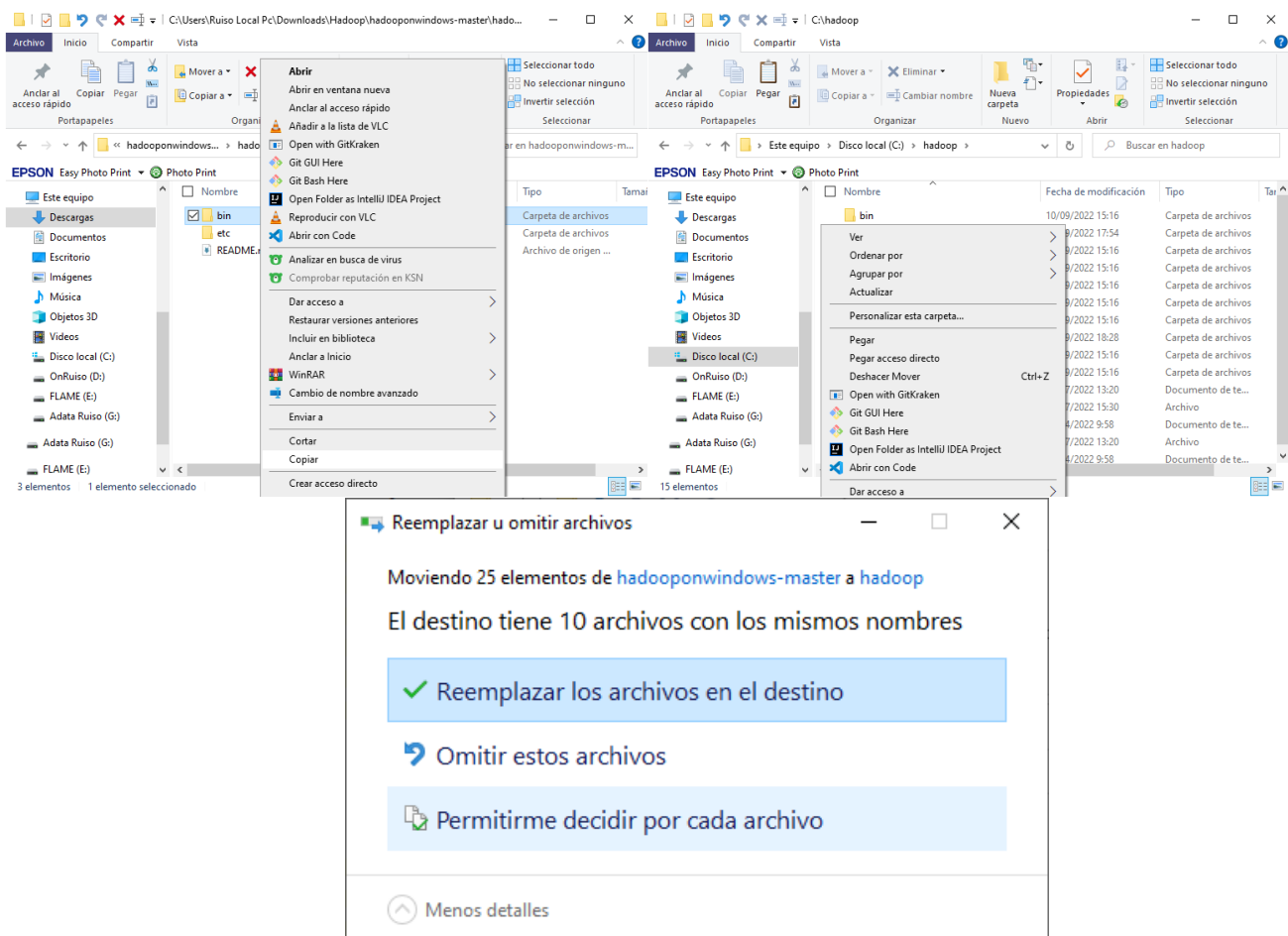
5. Una última revisada al hadoop-env.cmd, rectificando la variable de entorno JAVA_HOME.



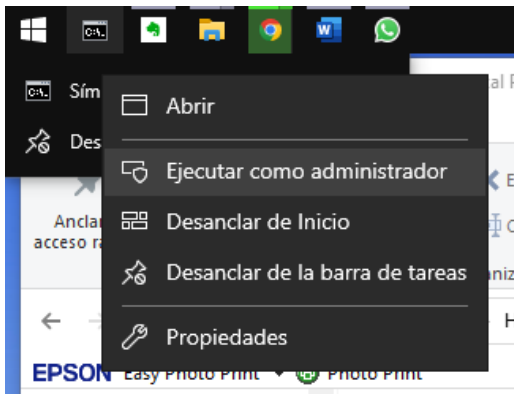
Ahora bien, hay algunas configuraciones propias de archivos BIN y ETC que son complicadas de toquetear como lo hicimos en los anteriores archivos, para nuestra fortuna, alguien ya lo hizo por nosotros y subió estos archivos a un repositorio de GITHUB siendo este el siguiente enlace, ahí encontraremos los archivos que podremos descargar luego en formato de ZIP o si es mas cómodo para usted, clonar el repositorio. Enlace: <https://github.com/sardetushar/hadooponwindows>



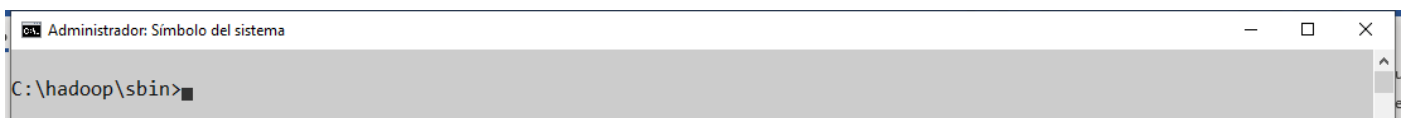
Una vez tenemos nuestros archivos descargados nos remitiremos a /HADOOP/ETC/HADOOP/ y moveremos nuestros archivos previamente modificados y guardados a otra ruta del pc con el fin de no comprometerlos, como por ejemplo escritorio. Una vez hecho esto podemos remitirnos al sitio de descarga descomprimir la carpeta y mover su contenido al sitio donde instalamos Hadoop en nuestra máquina, reemplazando todos los archivos. Una vez hecho esto podemos volver a mover nuestro archivo que habíamos movido a otra instancia como el escritorio y devolverlos a su ruta original, reemplazando alguno ya existente si es el caso.



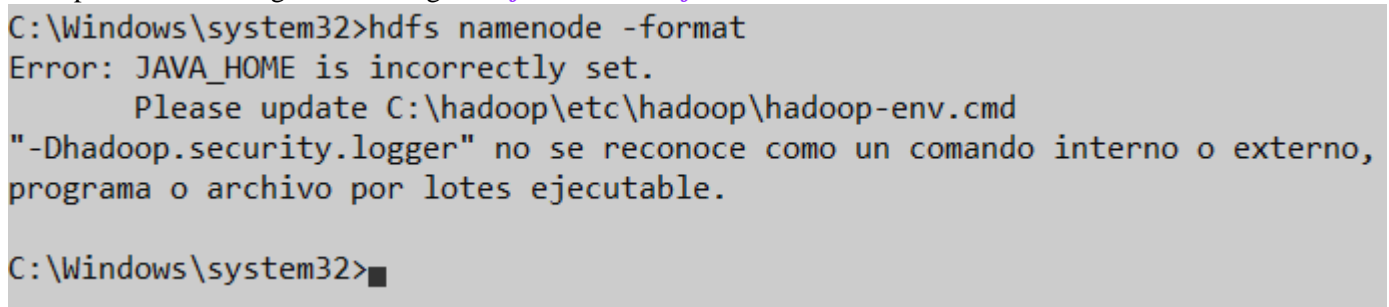
Ahora vamos a ejecutar la consola de comandos de Windows, podemos buscarla en el menú de aplicaciones de Windows como CMD. Lo importante en este paso es abrirla como Administradores del sistema, lo que se conoce como SuperUsuario o Root en las Maquinas Linux.



Dentro de la consola de comandos navegaremos hasta llegar a la dirección:



Allí aplicaremos los siguientes códigos: *hdfs namenode -format*

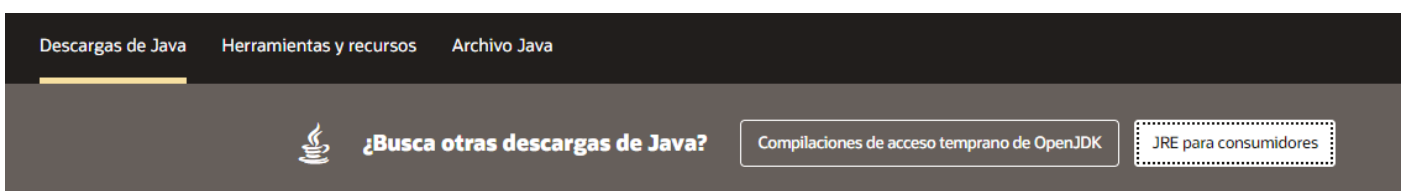


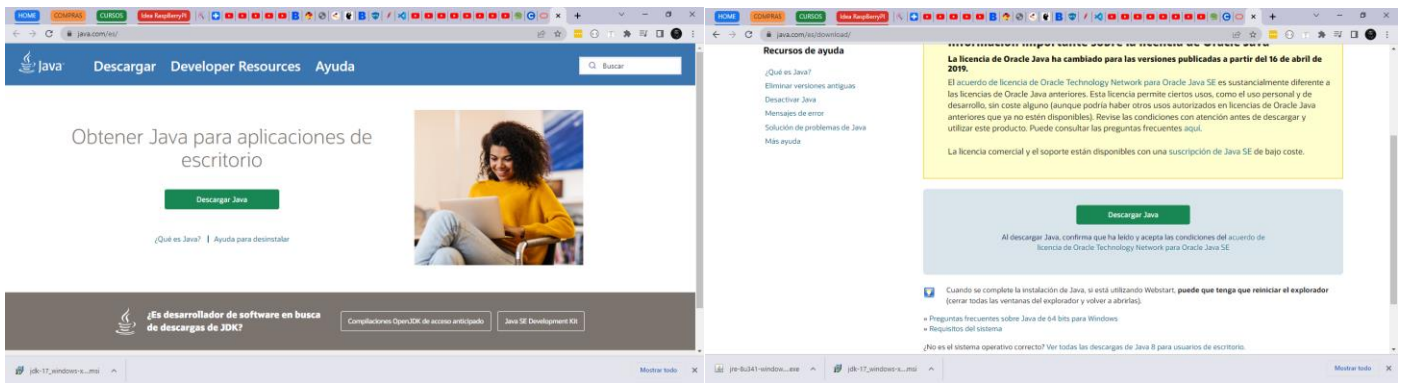
Podemos llegar a tener el error arrojado en la anterior pantalla. Existen varias formas de poder solucionarlo, lo primero es remitirnos nuevamente al documento “hadoop-env.cmd” y cambiar el renglón ultimo se se encarga de acceder al nombre del equipo y plantearlo como usuario de hadoop. Muchas veces pasa que los nombres de las maquinas contiene espacios, esto es leído como error y puede ser reemplazado por una palabra sin espacios cualquiera como se presenta a continuación.

```
@rem A string representing this instance of hadoop. %USERNAME% by default.
set HADOOP_IDENT_STRING=%USERNAME%

@rem A string representing this instance of hadoop. %USERNAME% by default.
set HADOOP_IDENT_STRING=RUIISO
```

En caso de que aun no funcione puede ser que el JDK no sea el competente para funcionar, en ese caso se debe instalar de la misma pagina de Oracle el JRE. Una vez descargado he instalado debemos alterar todas las rutas de configuración que utilizaban este JAMA_HOME.





Editar la variable de usuario

Nombre de la variable:

Valor de la variable:

Editar variable de entorno

```
@rem The java implementation to use. Required.  
@rem set JAVA_HOME=%JAVA_HOME%  
set JAVA_HOME=C:\Program~1\Java\jre1.8.0_341
```

En caso de que aun no funcione, puede deberse a que, aunque cambiamos las direcciones de JDK y JRE para que pudiesen estar dentro de carpetas con nombres que contienen espacios, las mismas rutas no pueden leerse dentro de la instalación de Hadoop. Por tanto una solución sería desinstalar las antiguas versiones de Java y reinstalar dentro de una carpeta próxima a la raíz del sistema.

Aplicaciones y características

Características opcionales

Alias de ejecución de aplicaciones

Busca, ordena y filtra por unidad. Si quieres desinstalar o mover una aplicación, selecciónala de la lista.

Ordenar por: Nombre ▾

Filtrar por: Todas las unidades ▾

Se encontraron 2 aplicaciones



Java 8 Update 341 (64-bit)

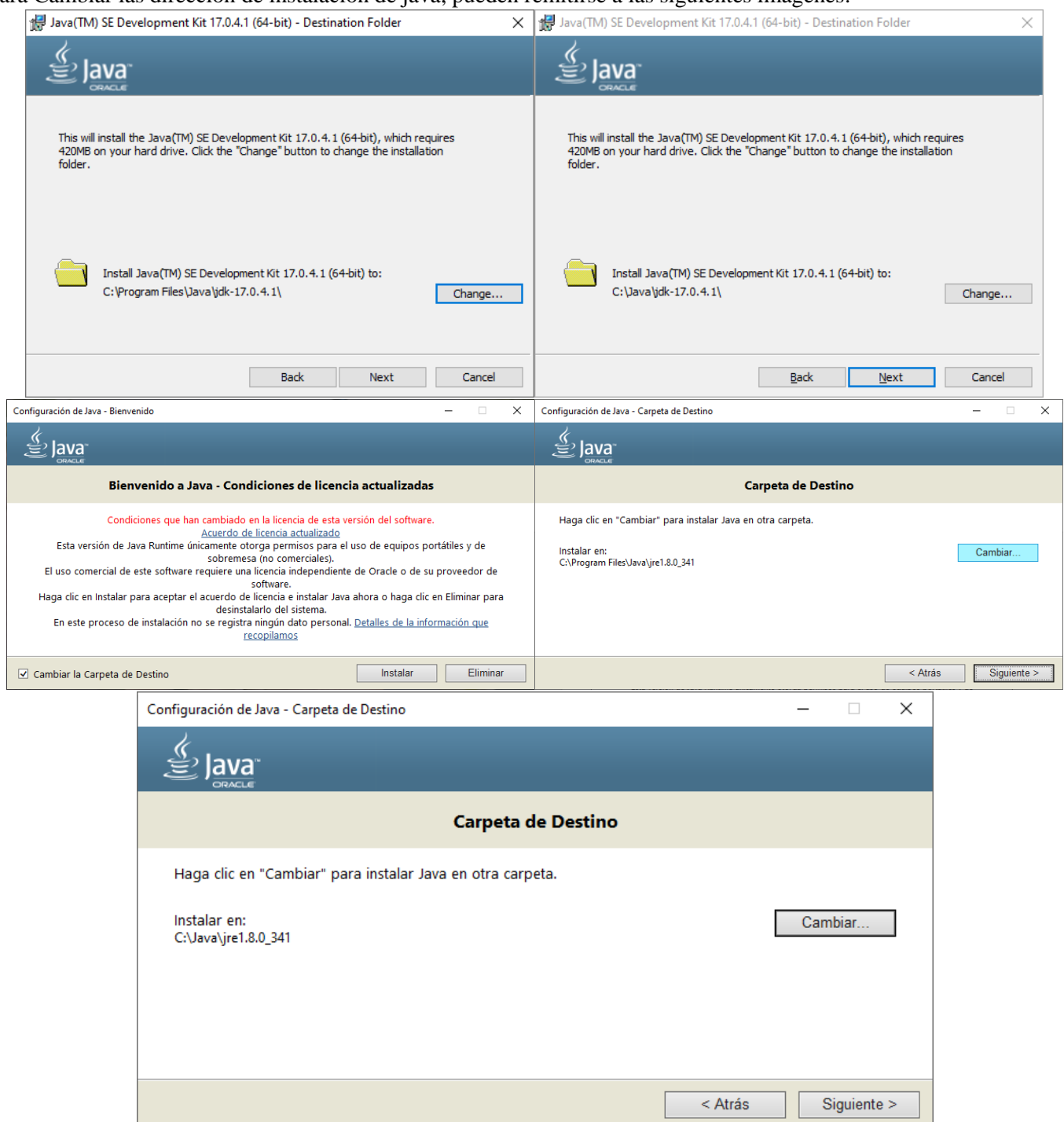
129 MB
10/09/2022



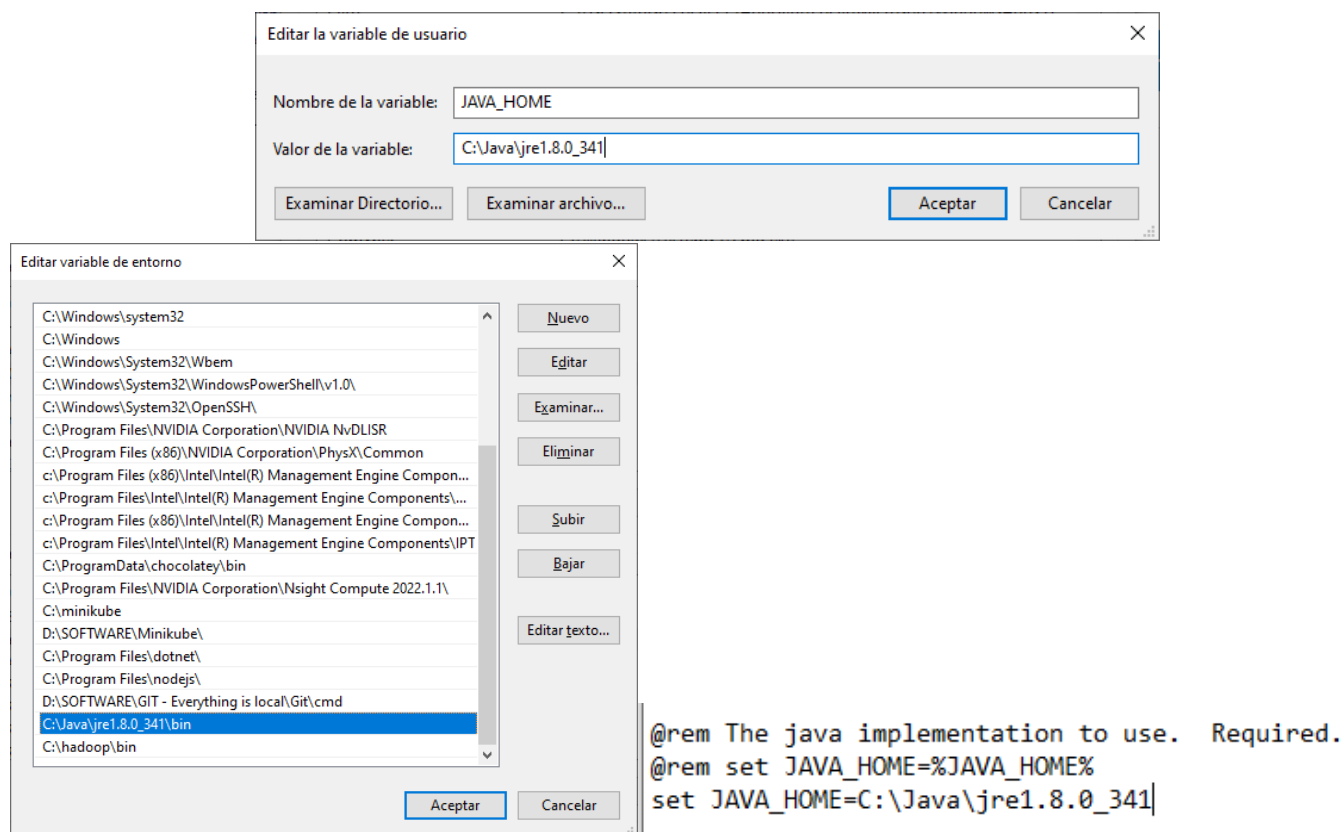
Java(TM) SE Development Kit 17.0.2 (64-bit)

290 MB
12/03/2022

Para Cambiar las dirección de instalación de java, pueden remitirse a las siguientes imágenes:



Una vez tenemos los nuevos JDK o JRE podemos probar con cada una de las versiones a ver cual nos permite continuar con la instalación, cambiando las direcciones en las variables de entorno y archivos de hadoop. Tanto si es para JRE como si es para JDK.



Editar la variable de usuario

Nombre de la variable: JAVA_HOME

Valor de la variable: C:\Java\jre1.8.0_341

Examinar Directorio... Examinar archivo... Aceptar Cancelar

Editar variable de entorno

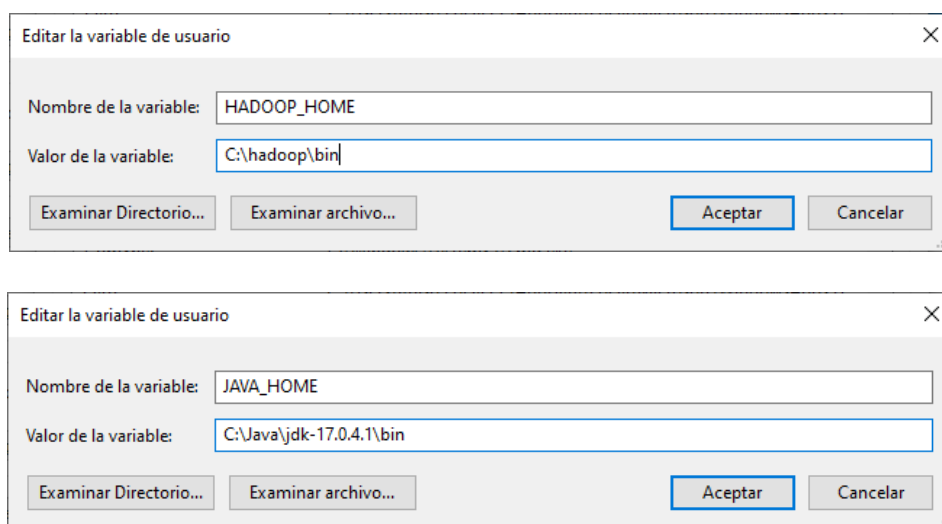
C:\Windows\system32
C:\Windows
C:\Windows\System32\Wbem
C:\Windows\System32\WindowsPowerShell\v1.0\
C:\Windows\System32\OpenSSH\
C:\Program Files\NVIDIA Corporation\NVIDIA NvDLISR
C:\Program Files (x86)\NVIDIA Corporation\PhysX\Common
c:\Program Files (x86)\Intel\Intel(R) Management Engine Compon...
c:\Program Files\Intel\Intel(R) Management Engine Components\...
c:\Program Files (x86)\Intel\Intel(R) Management Engine Compon...
c:\Program Files\Intel\Intel(R) Management Engine Components\IPT
C:\ProgramData\chocolatey\bin
C:\Program Files\NVIDIA Corporation\Nsight Compute 2022.1.1\
C:\minikube
D:\SOFTWARE\Minikube\
C:\Program Files\dotnet\
C:\Program Files\nodejs\
D:\SOFTWARE\GIT - Everything is local\Git\cmd
C:\Java\jre1.8.0_341\bin
C:\hadoop\bin

Nuevo
Editar
Examinar...
Eliminar
Subir
Bajar
Editar texto...

Aceptar Cancelar

```
@rem The java implementation to use. Required.  
@rem set JAVA_HOME=%JAVA_HOME%  
set JAVA_HOME=C:\Java\jre1.8.0_341
```

También se puede probar cambiando las direcciones de las variables de entorno sumándole el \bin como si se tratasen de las direcciones del Path.



Editar la variable de usuario

Nombre de la variable: HADOOP_HOME

Valor de la variable: C:\hadoop\bin

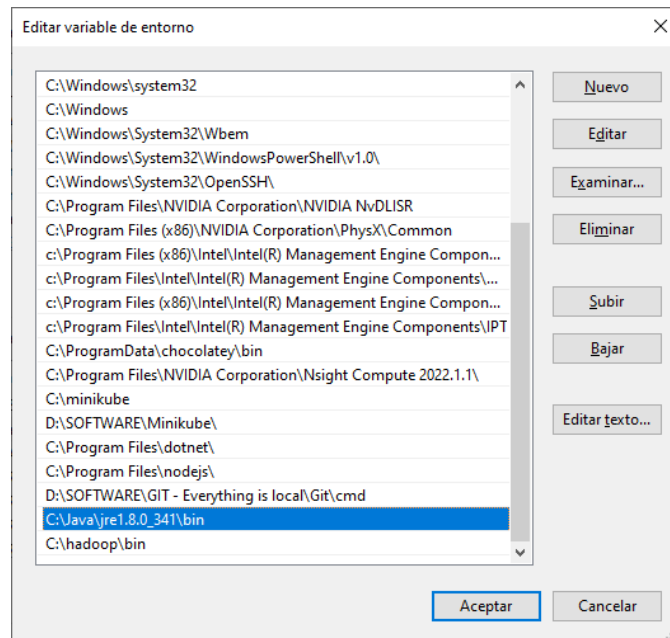
Examinar Directorio... Examinar archivo... Aceptar Cancelar

Editar la variable de usuario

Nombre de la variable: JAVA_HOME

Valor de la variable: C:\Java\jdk-17.0.4.1\bin

Examinar Directorio... Examinar archivo... Aceptar Cancelar



Ahora volvemos a ejecutar el anterior comando, el cual no debe ya arrojar error.

```
Administrador: Símbolo del sistema
2022-09-10 18:24:48,509 INFO util.GSet: Computing capacity for map NameNodeRetryCache
2022-09-10 18:24:48,509 INFO util.GSet: VM type = 64-bit
2022-09-10 18:24:48,510 INFO util.GSet: 0.029999999329447746% max memory 1000 MB = 307.2 KB
2022-09-10 18:24:48,510 INFO util.GSet: capacity = 2^15 = 32768 entries
2022-09-10 18:24:53,072 INFO namenode.FSImage: Allocated new BlockPoolId: BP-738757005-192.168.56.1-1662852293065
2022-09-10 18:24:53,075 WARN namenode.NameNode: Encountered exception during format
java.io.IOException: No image directories available!
    at org.apache.hadoop.hdfs.server.namenode.FSImage.saveFSImageInAllDirs(FSImage.java:1219)
    at org.apache.hadoop.hdfs.server.namenode.FSImage.saveFSImageInAllDirs(FSImage.java:1200)
    at org.apache.hadoop.hdfs.server.namenode.FSImage.format(FSImage.java:191)
    at org.apache.hadoop.hdfs.server.namenode.NameNode.format(NameNode.java:1278)
    at org.apache.hadoop.hdfs.server.namenode.NameNode.createNameNode(NameNode.java:1726)
    at org.apache.hadoop.hdfs.server.namenode.NameNode.main(NameNode.java:1834)
2022-09-10 18:24:53,094 INFO namenode.FSNamesystem: Stopping services started for active state
2022-09-10 18:24:53,097 INFO namenode.FSNamesystem: Stopping services started for standby state
2022-09-10 18:24:53,097 ERROR namenode.NameNode: Failed to start namenode.
java.io.IOException: No image directories available!
    at org.apache.hadoop.hdfs.server.namenode.FSImage.saveFSImageInAllDirs(FSImage.java:1219)
    at org.apache.hadoop.hdfs.server.namenode.FSImage.saveFSImageInAllDirs(FSImage.java:1200)
    at org.apache.hadoop.hdfs.server.namenode.FSImage.format(FSImage.java:191)
    at org.apache.hadoop.hdfs.server.namenode.NameNode.format(NameNode.java:1278)
    at org.apache.hadoop.hdfs.server.namenode.NameNode.createNameNode(NameNode.java:1726)
    at org.apache.hadoop.hdfs.server.namenode.NameNode.main(NameNode.java:1834)
2022-09-10 18:24:53,099 INFO util.ExitUtil: Exiting with status 1: java.io.IOException: No image directories available!
2022-09-10 18:24:53,102 INFO namenode.NameNode: SHUTDOWN_MSG:
/*****
SHUTDOWN_MSG: Shutting down NameNode at DESKTOP-5W9TBBJ/192.168.56.1
*****/
C:\hadoop\sbin>
```

Rectificamos nuestro JAVA en la maquina

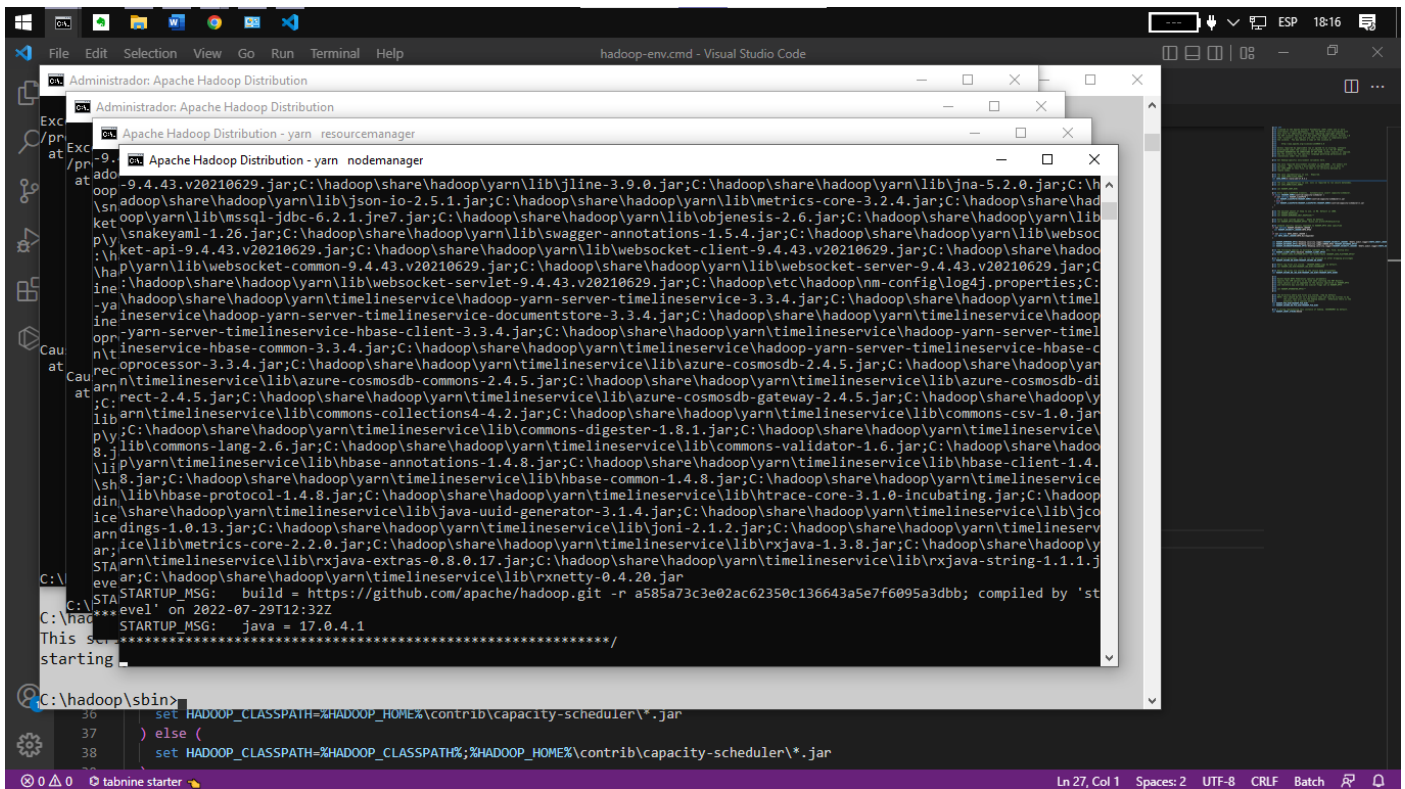
```
C:\Windows\system32>java -version
java version "17.0.4.1" 2022-08-18 LTS
Java(TM) SE Runtime Environment (build 17.0.4.1+1-LTS-2)
Java HotSpot(TM) 64-Bit Server VM (build 17.0.4.1+1-LTS-2, mixed mode, sharing)

C:\Windows\system32>
```

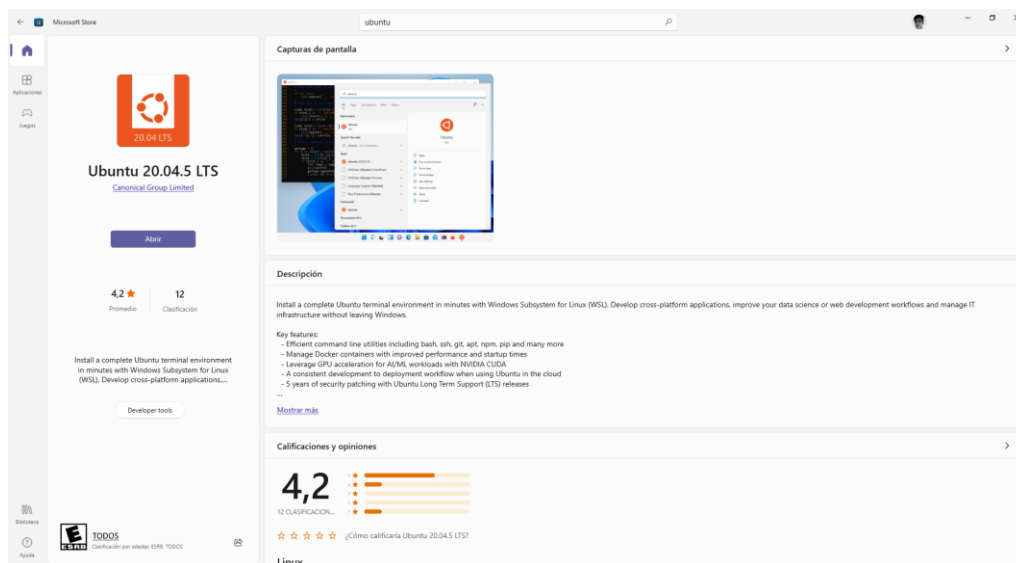
Ya solo queda iniciarlo con el siguiente comando:

```
C:\hadoop\sbin>start-all.cmd
This script is Deprecated. Instead use start-dfs.cmd and start-yarn.cmd
starting yarn daemons
C:\hadoop\sbin>
```

Al hacerlo se ejecutaran 4 procesos en 4 consolas independientes que nos rectificaran que Hadoop ya esta iniciado en nuestra computadora.



Ahora bien, en el caso de que aun existan complicaciones con manejar Hadoop en Windows o el mismo no funcione como tiene que funcionar, siempre se puede utilizar Ubuntu para Windows 10, el mismo se puede conseguir en la tienda de Microsoft Store como Ubuntu LTS, la versión que trabajaremos para esta guía es la 20.04.5.

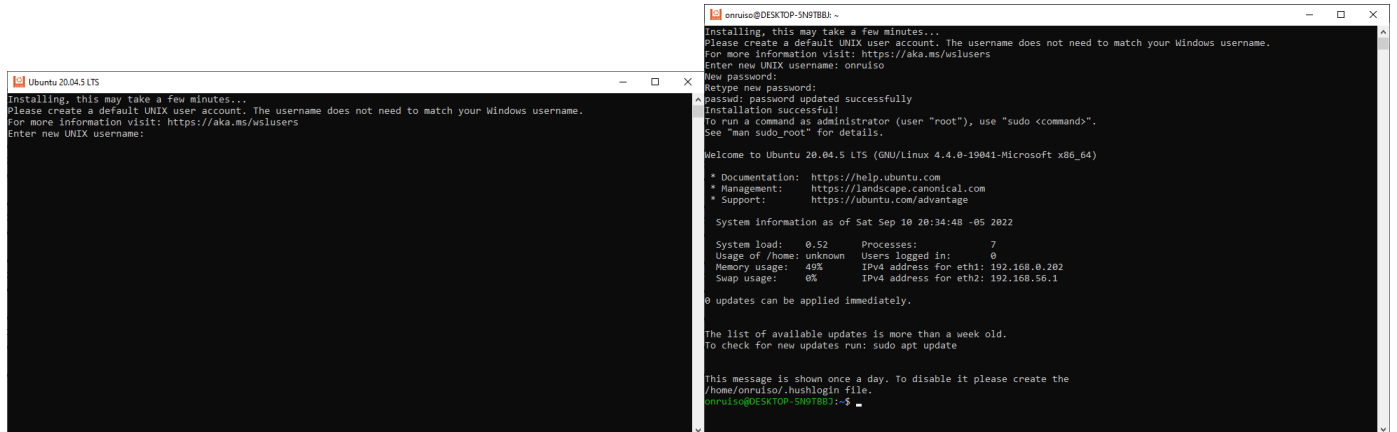


Este Ubuntu es en toda regla una maquina virtual o un subsistema de Linux corriendo en toda regla bajo Windows sin necesidad de algún intermediario como lo es Virtual Box, sin embargo es el SO Crudo, manejado por consola de comandos, nada de interfaces graficas para usuario u otras cosas añadidas a los SO's modernos.

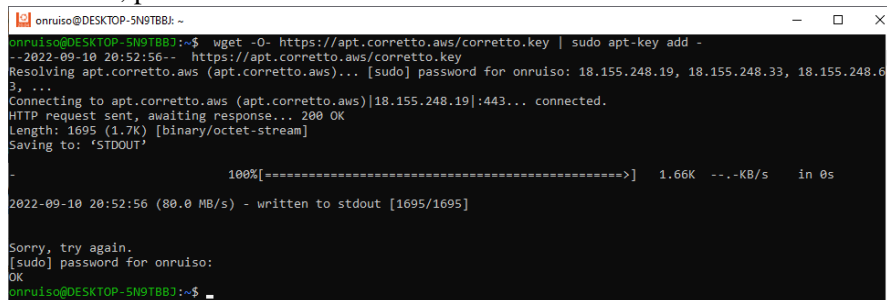
El proceso de instalación de Hadoop en este Ubuntu puede seguirse en las siguientes imágenes, sin embargo es recomendable dirigirse a los siguientes enlaces para tener una guía mas concreta de la instalación en la distro de Linux. Los enlaces son:

1. <https://docs.aws.amazon.com/corretto/latest/corretto-11-ug/generic-linux-install.html>
2. <https://docs.aws.amazon.com/corretto/latest/corretto-11-ug/downloads-list.html>
3. <https://hadoop.apache.org/releases.html>
4. <https://hadoop.apache.org/docs/stable/hadoop-project-dist/hadoop-common/SingleCluster.html>
- 5.

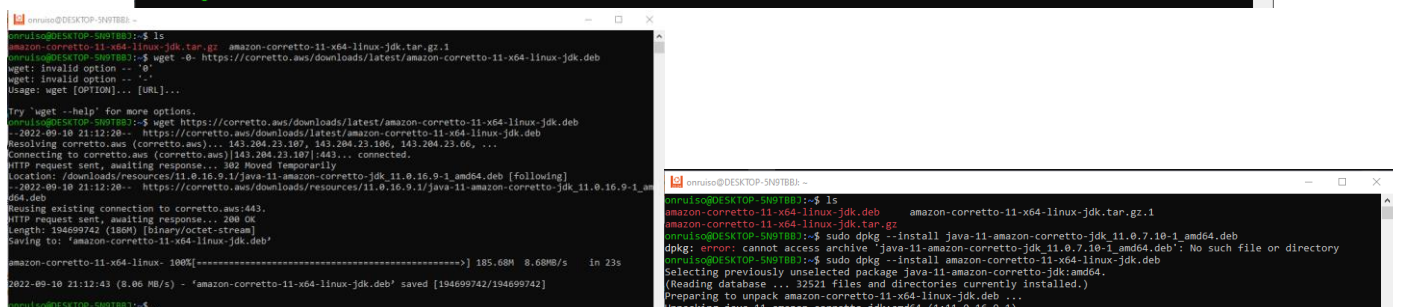
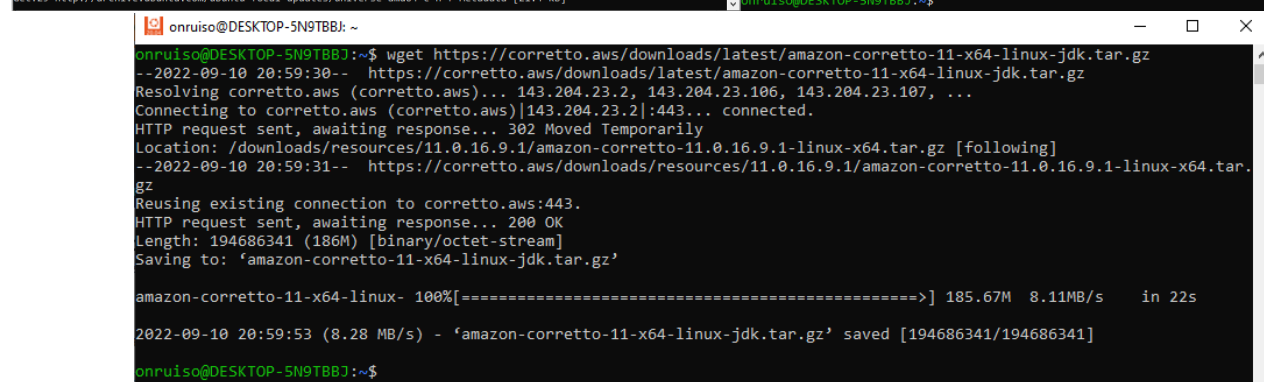
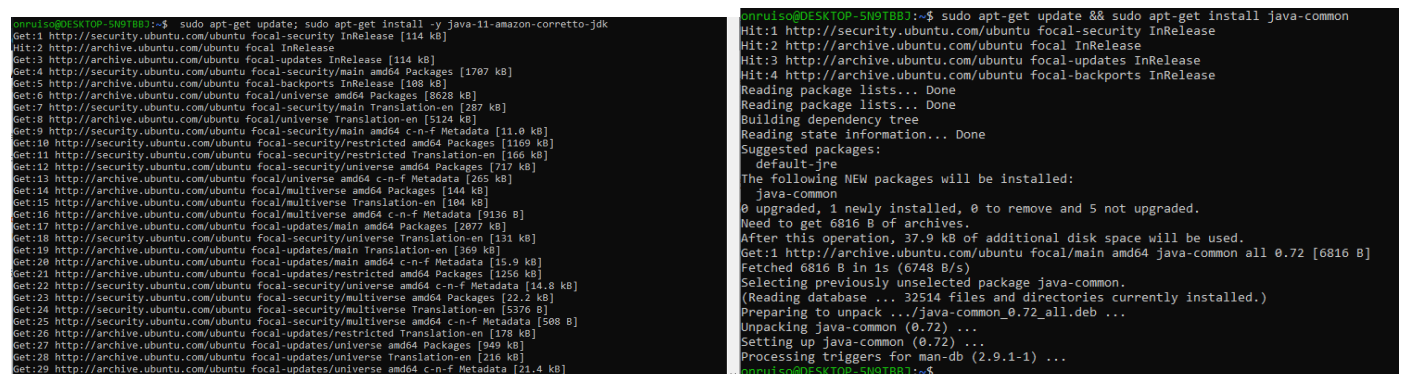
Cabe mencionar que para que no surja algún percance o error con la instalación de los archivos en nuestro Ubuntu/Windows tendremos que desactivar el Antivirus de nuestra computadora. Iniciamos primero nuestro Ubuntu/Windows.



Descargamos Correto de AWS, para mas información del mismo es recomendable ver los enlaces sugeridos.



Instalamos nuestro JAVA JDK e instalamos nuestro CORRETO AWS.



```
onruiso@DESKTOP-5N9TBBJ: ~  
onruiso@DESKTOP-5N9TBBJ:~$ ls  
amazon-corretto-11-x64-linux-jdk.deb      amazon-corretto-11-x64-linux-jdk.tar.gz.1  
amazon-corretto-11-x64-linux-jdk.tar.gz  
onruiso@DESKTOP-5N9TBBJ:~$ java -version  
openjdk version "11.0.16.1" 2022-08-12 LTS  
OpenJDK Runtime Environment Corretto-11.0.16.9.1 (build 11.0.16.1+9-LTS)  
OpenJDK 64-Bit Server VM Corretto-11.0.16.9.1 (build 11.0.16.1+9-LTS, mixed mode)  
onruiso@DESKTOP-5N9TBBJ:~$
```

Tendremos que Volver a ingresar a la pagina oficial de Hadoop para obtener el enlace de descarga de los binarios.

Descargar

Hadoop se publica como tarballs de código fuente con los tarballs binarios correspondientes para mayor comodidad. Las descargas se distribuyen a través de sitios espejo y deben verificarse para detectar manipulaciones utilizando GPG o SHA-512.

| Versión | Fecha de lanzamiento | Descarga de fuente | Descarga binaria | Notas de lanzamiento |
|---------|----------------------|--|---|----------------------|
| 3.3.4 | 8 de agosto de 2022 | fuelle (firma de suma de comprobación) | binary (firma de suma de control) binary-aarch64 (firma de suma de control) | Anuncio |
| 3.2.4 | 2022 julio 22 | fuelle (firma de suma de comprobación) | binario (firma de suma de comprobación) | Anuncio |
| 2.10.2 | 2022 mayo 31 | fuelle (firma de suma de comprobación) | binario (firma de suma de comprobación) | Anuncio |



DESARROLLO LIDERADO POR LA COMUNIDAD "THE APACHE WAY"

Proyectos Gente Comunidad Licencia Patrocinadores



Sugerimos el siguiente sitio para su descarga:

<https://dlcdn.apache.org/hadoop/common/hadoop-3.3.4/hadoop-3.3.4.tar.gz>

A continuación se sugieren ubicaciones de descarga alternativas.

Es esencial que verifique la integridad del archivo descargado mediante la firma PGP (.asc archivo) o un hash (.md5 o .sha* archivo).

Con el enlace ya podremos descargar el comprimido con la función WGET.

```
onruiso@DESKTOP-5N9TBBJ:~$ wget https://dlcdn.apache.org/hadoop/common/hadoop-3.3.4/hadoop-3.3.4.tar.gz  
--2022-09-10 21:24:26-- https://dlcdn.apache.org/hadoop/common/hadoop-3.3.4/hadoop-3.3.4.tar.gz  
Resolving dlcdn.apache.org (dlcdn.apache.org)... 151.101.2.132, 2a04:4e42::644  
Connecting to dlcdn.apache.org (dlcdn.apache.org)|151.101.2.132|:443... connected.  
HTTP request sent, awaiting response... 200 OK  
Length: 695457782 (663M) [application/x-gzip]  
Saving to: 'hadoop-3.3.4.tar.gz'  
hadoop-3.3.4.tar.gz 100%[=====] 663.24M 182KB/s in 44m 29s  
2022-09-10 22:08:57 (254 KB/s) - 'hadoop-3.3.4.tar.gz' saved [695457782/695457782]  
onruiso@DESKTOP-5N9TBBJ:~$
```

Descomprimir el archive de Hadoop ya descargado.

```
onruiso@DESKTOP-5N9TBBJ:~$ tar -xzf hadoop-3.3.4.tar.gz
```

Comprobar el lugar de instalación de java y exportar la variable de ambiente JAVA_HOME

```
onruiso@DESKTOP-5N9TBBJ:~$ ls  
amazon-corretto-11-x64-linux-jdk.deb      amazon-corretto-11-x64-linux-jdk.tar.gz.1  hadoop-3.3.4.tar.gz  
amazon-corretto-11-x64-linux-jdk.tar.gz  hadoop-3.3.4  
onruiso@DESKTOP-5N9TBBJ:~$ which java | xargs ls -lta | awk '{print $11}' | xargs ls -lta | rev | cut -c10- | rev;  
lrwxrwxrwx 1 root root 45 Sep 10 21:15 /etc/alternatives/java -> /usr/lib/jvm/java-11-amazon-corretto  
onruiso@DESKTOP-5N9TBBJ:~$ export JAVA_HOME=/usr/lib/jvm/java-11-amazon-corretto  
onruiso@DESKTOP-5N9TBBJ:~$
```

Ahora verificamos el directorio donde se descomprimió hadoop y corroboramos las versiones que tenemos del mismo programa, puede que nativamente ya este uno instalado.

```
onruiso@DESKTOP-5N9TBBJ:~$ ls  
amazon-corretto-11-x64-linux-jdk.deb      amazon-corretto-11-x64-linux-jdk.tar.gz.1  hadoop-3.3.4.tar.gz  
amazon-corretto-11-x64-linux-jdk.tar.gz  hadoop-3.3.4  
onruiso@DESKTOP-5N9TBBJ:~$ hadoop version  
Hadoop 2.10.2  
Subversion Unknown -r 965fd38006fa78b2315668fbc7eb432e1d8200f  
Compiled by ubuntu on 2022-05-24T22:35Z  
Compiled with protoc 2.5.0  
From source with checksum d3ab737f7788f05d467784f0a86573fe  
This command was run using /mnt/c/hadoop/share/hadoop/common/hadoop-common-2.10.2.jar  
onruiso@DESKTOP-5N9TBBJ:~$ cd hadoop-3.3.4/  
onruiso@DESKTOP-5N9TBBJ:~/hadoop-3.3.4$ ls  
LICENSE-binary NOTICE-binary README.txt etc lib licenses-binary share  
LICENSE.txt NOTICE.txt bin include libexec sbin  
onruiso@DESKTOP-5N9TBBJ:~/hadoop-3.3.4$ bin/hadoop version  
Hadoop 3.3.4  
Source code repository https://github.com/apache/hadoop.git -r a585a73c3e02ac62350c136643a5e7f6095a3ddb  
Compiled by stevel on 2022-07-29T12:32Z  
Compiled with protoc 3.7.1  
From source with checksum fb9dd8918a7b8a5b430d61af858f6ec  
This command was run using /home/onruiso/hadoop-3.3.4/share/hadoop/common/hadoop-common-3.3.4.jar  
onruiso@DESKTOP-5N9TBBJ:~/hadoop-3.3.4$
```

Instalar SSH (security Shell) para ejecutar de forma segura el ejecutable de Hadoop.


```
onruiso@DESKTOP-5N9TBBJ:~$ ls
amazon-corretto-11-x64-linux-jdk.deb  amazon-corretto-11-x64-linux-jdk.tar.gz.1  hadoop-3.3.4.tar.gz
onruiso@DESKTOP-5N9TBBJ:~$ sudo apt-get install ssh
[sudo] password for onruiso:
Reading package lists... Done
Building dependency tree
Reading state information... Done
The following NEW packages will be installed:
  ssh
0 upgraded, 1 newly installed, 0 to remove and 5 not upgraded.
Need to get 5084 B of archives.
After this operation, 120 kB of additional disk space will be used.
Get:1 http://archive.ubuntu.com/ubuntu focal-updates/main amd64 ssh all 1:8.2p1-4ubuntu0.5 [5084 B]
Fetched 5084 B in 0s (11.9 kB/s)
Selecting previously unselected package ssh.
(Reading database ... 33079 files and directories currently installed.)
Preparing to unpack .../ssh_1:8.2p1-4ubuntu0.5_all.deb ...
Unpacking ssh (1:8.2p1-4ubuntu0.5) ...
Setting up ssh (1:8.2p1-4ubuntu0.5) ...
onruiso@DESKTOP-5N9TBBJ:~$
```

```
Setting up ssh (1:8.2p1-4ubuntu0.5) ...
onruiso@DESKTOP-5N9TBBJ:~$ sudo apt-get install pdsh
Reading package lists... Done
Building dependency tree
Reading state information... Done
The following additional packages will be installed:
  genders libgenders0
Suggested packages:
  rdist
The following NEW packages will be installed:
  genders libgenders0 pdsh
0 upgraded, 3 newly installed, 0 to remove and 5 not upgraded.
Need to get 167 kB of archives.
After this operation, 519 kB of additional disk space will be used.
Do you want to continue? [Y/n] Y
Get:1 http://archive.ubuntu.com/ubuntu focal/universe amd64 libgenders0 amd64 1.22-1build2 [29.2 kB]
Get:2 http://archive.ubuntu.com/ubuntu focal/universe amd64 genders amd64 1.22-1build2 [29.7 kB]
Get:3 http://archive.ubuntu.com/ubuntu focal/universe amd64 pdsh amd64 2.31-3build2 [108 kB]
Preconfiguring packages ...
Selecting previously unselected package libgenders0:amd64.
(Reading database ... 33082 files and directories currently installed.)
Preparing to unpack .../libgenders0_1.22-1build2_amd64.deb ...
Unpacking libgenders0:amd64 (1.22-1build2) ...
Selecting previously unselected package genders.
Preparing to unpack .../genders_1.22-1build2_amd64.deb ...
Unpacking genders (1.22-1build2) ...
Selecting previously unselected package pdsh.
Preparing to unpack .../pdsh_2.31-3build2_amd64.deb ...
Unpacking pdsh (2.31-3build2) ...
Setting up libgenders0:amd64 (1.22-1build2) ...
Setting up genders (1.22-1build2) ...
Setting up pdsh (2.31-3build2) ...
Processing triggers for libc-bin (2.31-0ubuntu9.9) ...
Processing triggers for man-db (2.9.1-1) ...
onruiso@DESKTOP-5N9TBBJ:~$
```

De forma predeterminada, Hadoop está configurado para ejecutarse en un modo no distribuido, como un único proceso de Java. Vamos a transformar los mismos archivos que utilizamos la anterior vez en Windows de extensión XML.

```
onruiso@DESKTOP-5N9TBBJ:~/hadoop-3.3.4/etc/hadoop$ nano hadoop-env.sh
onruiso@DESKTOP-5N9TBBJ:~/hadoop-3.3.4/etc/hadoop$

GNU nano 4.8 hadoop-env.sh

# Licensed to the Apache Software Foundation (ASF) under one
# or more contributor license agreements. See the NOTICE file
# distributed with this work for additional information
# regarding copyright ownership. The ASF licenses this file
# to you under the Apache License, Version 2.0 (the
# "License"); you may not use this file except in compliance
# with the License. You may obtain a copy of the License at
#
# http://www.apache.org/licenses/LICENSE-2.0
#
# Unless required by applicable law or agreed to in writing, software
# distributed under the License is distributed on an "AS IS" BASIS,
# WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
# See the License for the specific language governing permissions and
# limitations under the License.

# Set Hadoop-specific environment variables here.

##
## THIS FILE ACTS AS THE MASTER FILE FOR ALL HADOOP PROJECTS.
## SETTINGS HERE WILL BE READ BY ALL HADOOP COMMANDS. THEREFORE,
## ONE CAN USE THIS FILE TO SET YARN, HDFS, AND MAPREDUCE
## CONFIGURATION OPTIONS INSTEAD OF xxx-env.sh.
##
## Precedence rules:
##
## {yarn-env.sh|hdfs-env.sh} > hadoop-env.sh > hard-coded defaults
##
## {YARN_xyz|HDFS_xyz} > HADOOP_xyz > hard-coded defaults
##

[ Read 430 lines ]
^G Get Help  ^O Write Out  ^W Where Is  ^K Cut Text  ^J Justify   ^C Cur Pos   M-U Undo     M-A Mark Text
^X Exit      ^R Read File  ^_ Replace   ^U Paste Text ^T To Spell  ^_ Go To Line M-E Redo     M-6 Copy Text
```

Cambiar de izquierda a derecha:

```
# The java implementation to use. By default, this environment
# variable is REQUIRED on ALL platforms except OS X!
# export JAVA_HOME=

# The java implementation to use. By default, this environment
# variable is REQUIRED on ALL platforms except OS X!
export JAVA_HOME=/usr/lib/jvm/java-11-amazon-corretto_
```

```
onruiso@DESKTOP-5N9TBBJ:~/hadoop-3.3.4/etc/hadoop$ nano core-site.xml
```

```
onruiso@DESKTOP-5N9TBBJ: ~/hadoop-3.3.4/etc/hadoop
GNU nano 4.8 core-site.xml
<?xml version="1.0" encoding="UTF-8"?>
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
<!--
Licensed under the Apache License, Version 2.0 (the "License");
you may not use this file except in compliance with the License.
You may obtain a copy of the License at

http://www.apache.org/licenses/LICENSE-2.0

Unless required by applicable law or agreed to in writing, software
distributed under the License is distributed on an "AS IS" BASIS,
WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
See the license for the specific language governing permissions and
limitations under the License. See accompanying LICENSE file.
-->

<!-- Put site-specific property overrides in this file. -->
<configuration>
  <property>
    <name>fs.defaultFS</name>
    <value>hdfs://localhost:9000</value>
  </property>
</configuration>

```

```
onruiso@DESKTOP-5N9TBBJ:~/hadoop-3.3.4/etc/hadoop$ nano hdfs-site.xml
onruiso@DESKTOP-5N9TBBJ: ~/hadoop-3.3.4/etc/hadoop
GNU nano 4.8 hdfs-site.xml Modified
<!--
Licensed under the Apache License, Version 2.0 (the "License");
you may not use this file except in compliance with the License.
You may obtain a copy of the License at

http://www.apache.org/licenses/LICENSE-2.0

Unless required by applicable law or agreed to in writing, software
distributed under the License is distributed on an "AS IS" BASIS,
WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
See the license for the specific language governing permissions and
limitations under the License. See accompanying LICENSE file.
-->

<!-- Put site-specific property overrides in this file. -->
<configuration>
  <property>
    <name>dfs.replication</name>
    <value>1</value>
  </property>

  <property>
    <name>dfs.name.dir</name>
    <value>file:///home/parallels/Downloads/hadoop-3.3.3/namenode </value>
  </property>

  <property>
    <name>dfs.data.dir</name>
    <value>file:///home/parallels/Downloads/hadoop-3.3.3/datanode </value>
  </property>
</configuration>

```

Creamos los dos nodos principales de uso de Hadoop, los mismos tipados en el archivo de hds-site.xml.

```
onruiso@DESKTOP-5N9TBBJ:~/hadoop-3.3.4/etc/hadoop$ nano hdfs-site.xml
onruiso@DESKTOP-5N9TBBJ:~/hadoop-3.3.4/etc/hadoop$ pwd
/home/onruiso/hadoop-3.3.4/etc/hadoop
onruiso@DESKTOP-5N9TBBJ:~/hadoop-3.3.4/etc/hadoop$ cd ..
onruiso@DESKTOP-5N9TBBJ:~/hadoop-3.3.4/etc$ cd ..
onruiso@DESKTOP-5N9TBBJ:~/hadoop-3.3.4$ ls
LICENSE-binary  NOTICE-binary  README.txt  etc  lib  licenses-binary  share
LICENSE.txt  NOTICE.txt  bin  include  libexec  sbin
onruiso@DESKTOP-5N9TBBJ:~/hadoop-3.3.4$ mkdir namenode
onruiso@DESKTOP-5N9TBBJ:~/hadoop-3.3.4$ ls
LICENSE-binary  NOTICE-binary  README.txt  etc  lib  licenses-binary  sbin
LICENSE.txt  NOTICE.txt  bin  include  libexec  namenode  share
onruiso@DESKTOP-5N9TBBJ:~/hadoop-3.3.4$ mkdir datanode
onruiso@DESKTOP-5N9TBBJ:~/hadoop-3.3.4$ ls
LICENSE-binary  NOTICE-binary  README.txt  datanode  include  libexec  namenode  share
LICENSE.txt  NOTICE.txt  bin  etc  lib  licenses-binary  sbin
onruiso@DESKTOP-5N9TBBJ:~/hadoop-3.3.4$ cd namenode/
onruiso@DESKTOP-5N9TBBJ:~/hadoop-3.3.4/namenode$ pwd
/home/onruiso/hadoop-3.3.4/namenode
onruiso@DESKTOP-5N9TBBJ:~/hadoop-3.3.4/namenode$ cd ..
onruiso@DESKTOP-5N9TBBJ:~/hadoop-3.3.4$ cd datanode/
onruiso@DESKTOP-5N9TBBJ:~/hadoop-3.3.4/datanode$ pwd
/home/onruiso/hadoop-3.3.4/datanode
onruiso@DESKTOP-5N9TBBJ:~/hadoop-3.3.4/datanode$ cd ..
onruiso@DESKTOP-5N9TBBJ:~/hadoop-3.3.4$
```

```
onruiso@DESKTOP-5N9TBBJ: ~/hadoop-3.3.4/etc/hadoop
GNU nano 4.8 hdfs-site.xml
<?xml version="1.0" encoding="UTF-8"?>
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
<!--
  Licensed under the Apache License, Version 2.0 (the "License");
  you may not use this file except in compliance with the License.
  You may obtain a copy of the License at

    http://www.apache.org/licenses/LICENSE-2.0

  Unless required by applicable law or agreed to in writing, software
  distributed under the License is distributed on an "AS IS" BASIS,
  WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
  See the License for the specific language governing permissions and
  limitations under the License. See accompanying LICENSE file.
-->

<!-- Put site-specific property overrides in this file. -->
<configuration>
  <property>
    <name>dfs.replication</name>
    <value>1</value>
  </property>

  <property>
    <name>dfs.name.dir</name>
    <value>file:///home/onruiso/hadoop-3.3.4/namenode </value>
  </property>

  <property>
    <name>dfs.data.dir</name>
    <value>file:///home/onruiso/hadoop-3.3.4/datanode </value>
  </property>
</configuration>
I wrote 34 lines
^G Get Help ^O Write Out ^W Where Is ^K Cut Text ^J Justify ^C Cur Pos ^U Undo ^M Mark Text
^X Exit ^R Read File ^A Replace ^N Paste Text ^T To Spell ^_ Go To Line ^E Redo ^-M-6 Copy Text
```

Ahora si podemos correr Hadoop.

```
onruiso@DESKTOP-5N9TBBJ: ~/hadoop-3.3.4
onruiso@DESKTOP-5N9TBBJ:~/hadoop-3.3.4/etc/hadoop$ cd ..
onruiso@DESKTOP-5N9TBBJ:~/hadoop-3.3.4/etc$ cd ..
onruiso@DESKTOP-5N9TBBJ:~/hadoop-3.3.4$ ls
LICENSE-binary NOTICE-binary README.txt datanode include libexec namenode share
LICENSE.txt NOTICE.txt bin etc lib licenses-binary sbin
onruiso@DESKTOP-5N9TBBJ:~/hadoop-3.3.4$ ./sbin/start-all.sh
WARNING: Attempting to start all Apache Hadoop daemons as onruiso in 10 seconds.
WARNING: This is not a recommended production deployment configuration.
WARNING: Use CTRL-C to abort.
Starting namenodes on [localhost]
pdsh@DESKTOP-5N9TBBJ: localhost: rcmd: socket: Permission denied
Starting datanodes
pdsh@DESKTOP-5N9TBBJ: localhost: rcmd: socket: Permission denied
Starting secondary namenodes [DESKTOP-5N9TBBJ]
pdsh@DESKTOP-5N9TBBJ: DESKTOP-5N9TBBJ: rcmd: socket: Permission denied
Starting resource manager
WARNING: /home/onruiso/hadoop-3.3.4/logs does not exist. Creating.
Starting nodemanagers
pdsh@DESKTOP-5N9TBBJ: localhost: rcmd: socket: Permission denied
onruiso@DESKTOP-5N9TBBJ:~/hadoop-3.3.4$
```

```
onruiso@DESKTOP-5N9TBBJ:~/hadoop-3.3.4$ ./sbin/stop-all.sh
WARNING: Stopping all Apache Hadoop daemons as onruiso in 10 seconds.
WARNING: Use CTRL-C to abort.
Stopping namenodes on [localhost]
pdsh@DESKTOP-5N9TBBJ: localhost: rcmd: socket: Permission denied
Stopping datanodes
pdsh@DESKTOP-5N9TBBJ: localhost: rcmd: socket: Permission denied
Stopping secondary namenodes [DESKTOP-5N9TBBJ]
pdsh@DESKTOP-5N9TBBJ: DESKTOP-5N9TBBJ: rcmd: socket: Permission denied
Stopping nodemanagers
pdsh@DESKTOP-5N9TBBJ: localhost: rcmd: socket: Permission denied
Stopping resource manager
onruiso@DESKTOP-5N9TBBJ:~/hadoop-3.3.4$
```

```
onruiso@DESKTOP-5N9TBBJ: ~/hadoop-3.3.4/bin
onruiso@DESKTOP-5N9TBBJ:~/hadoop-3.3.4$ ls
LICENSE-binary NOTICE-binary README.txt datanode include libexec logs sbin
LICENSE.txt NOTICE.txt bin etc lib licenses-binary namenode share
onruiso@DESKTOP-5N9TBBJ:~/hadoop-3.3.4$ cd bin/
onruiso@DESKTOP-5N9TBBJ:~/hadoop-3.3.4/bin$ hdfs namenode -format
22/09/10 23:55:02 INFO namenode.NameNode: STARTUP_MSG:
/*****
STARTUP_MSG: Starting NameNode
STARTUP_MSG: host = DESKTOP-5N9TBBJ/127.0.1.1
STARTUP_MSG: args = [-format]
STARTUP_MSG: version = 2.10.2
STARTUP_MSG: classpath = /mnt/c/hadoop/etc/hadoop:/mnt/c/hadoop/share/hadoop/common/lib/activation-1.1.jar:/mnt/c/hadoop/share/hadoop/common/lib/activation.jar:/mnt/c/hadoop/share/hadoop/common/lib/activation-1.1.jar:/mnt/c/hadoop/share/hadoop/common/lib/activation.jar:
*****/
```

```
onruiso@DESKTOP-5N9TBBJ: ~/hadoop-3.3.4
onruiso@DESKTOP-5N9TBBJ:~/hadoop-3.3.4/bin$ cd ..
onruiso@DESKTOP-5N9TBBJ:~/hadoop-3.3.4$ ./sbin/start-all.sh
WARNING: Attempting to start all Apache Hadoop daemons as onruiso in 10 seconds.
WARNING: This is not a recommended production deployment configuration.
WARNING: Use CTRL-C to abort.
Starting namenodes on [localhost]
pdsh@DESKTOP-5N9TBBJ: localhost: rcmd: socket: Permission denied
Starting datanodes
pdsh@DESKTOP-5N9TBBJ: localhost: rcmd: socket: Permission denied
Starting secondary namenodes [DESKTOP-5N9TBBJ]
pdsh@DESKTOP-5N9TBBJ: DESKTOP-5N9TBBJ: rcmd: socket: Permission denied
Starting resource manager
Starting nodemanagers
pdsh@DESKTOP-5N9TBBJ: localhost: rcmd: socket: Permission denied
onruiso@DESKTOP-5N9TBBJ:~/hadoop-3.3.4$
```

