

# Tema 3 Tarea 2 NiFi

Juan Manuel García Moyano  
IABD  
Informática y comunicaciones

## Índice

1. Instalación de NiFi.....	3
2. Trabajando con NiFi.....	6
1. Básico: Un CSV o JSON cualquiera. Previamente descargamos desde Kaggle.....	11

# 1. Instalación de NiFi

## 1. Descargamos Apache NiFi.

```
hadoop@dhcp2:~/Descargas$ wget https://dlcdn.apache.org/nifi/2.2.0/nifi-2.2.0-bin.zip
--2025-02-12 17:22:21-- https://dlcdn.apache.org/nifi/2.2.0/nifi-2.2.0-bin.zip
Resolviendo dlcdn.apache.org (dlcdn.apache.org)... 151.101.2.132, 2a04:4e42::644
Conectando con dlcdn.apache.org (dlcdn.apache.org)[151.101.2.132]:443... conectado.
Petición HTTP enviada, esperando respuesta... 200 OK
Longitud: 729771289 (696M) [application/zip]
Guardando como: "nifi-2.2.0-bin.zip"

nifi-2.2.0-bin.zip 100%[=====] 695,96M 2,11MB/s en 5m 6s

2025-02-12 17:27:27 (2,27 MB/s) - "nifi-2.2.0-bin.zip" guardado [729771289/729771289]

hadoop@dhcp2:~/Descargas$
```

## 2. Desempaquetamos nifi.

```
hadoop@dhcp2:~/Descargas$ unzip nifi-2.2.0-bin.zip
Archive: nifi-2.2.0-bin.zip
  creating: nifi-2.2.0/
  creating: nifi-2.2.0/extensions/
  creating: nifi-2.2.0/lib/
  creating: nifi-2.2.0/lib/bootstrap/
  creating: nifi-2.2.0/lib/aspectj/
  creating: nifi-2.2.0/conf/
  creating: nifi-2.2.0/bin/
  creating: nifi-2.2.0/docs/
  creating: nifi-2.2.0/docs/html/
  creating: nifi-2.2.0/docs/html/images/
  creating: nifi-2.2.0/docs/html/misc/
  creating: nifi-2.2.0/python/
  creating: nifi-2.2.0/python/framework/
  creating: nifi-2.2.0/python/framework/py4j/
  creating: nifi-2.2.0/python/api/
  creating: nifi-2.2.0/python/api/nifiapi/
  inflating: nifi-2.2.0/conf/authorizers.xml
  inflating: nifi-2.2.0/conf/bootstrap.conf
  inflating: nifi-2.2.0/conf/logback.xml
  inflating: nifi-2.2.0/conf/login-identity-providers.xml
```

3. Nos movemos al directorio bin de nifi. Ejecutamos el script tal y como se ve en la captura y nos debería de dar un error si no tenemos instalado el jdk 21 de Java.

```
hadoop@dhcp2:~$ cd Descargas/nifi-2.2.0/bin/
hadoop@dhcp2:~/Descargas/nifi-2.2.0/bin$ ./nifi.sh start

JAVA_HOME=/usr/lib/jvm/java-11-openjdk-amd64
NIFI_HOME=/home/hadoop/Descargas/nifi-2.2.0

Error: Se ha producido un error de enlace al cargar la clase principal org.apache.nifi.bootstrap.BootstrapProcess
    java.lang.UnsupportedClassVersionError: org/apache/nifi/bootstrap/BootstrapProcess has been compiled by a more recent version of the Java Runtime (class file version 65.0), this version of the Java Runtime only recognizes class file versions up to 55.0

hadoop@dhcp2:~/Descargas/nifi-2.2.0/bin$
```

4. Actualizamos la lista de paquetes.

```
hadoop@dhcp2:~$ sudo apt update
[sudo] contraseña para hadoop:
Des:1 http://security.ubuntu.com/ubuntu focal-security InRelease [128 kB]
Des:2 https://download.docker.com/linux/ubuntu focal InRelease [57,7 kB]
Obj:3 http://es.archive.ubuntu.com/ubuntu focal InRelease
Des:4 https://download.docker.com/linux/ubuntu focal/stable amd64 Packages [54,9
```

5. Descargamos la lista de paquetes instalados.

```
hadoop@dhcp2:~$ sudo apt upgrade -y
Leyendo lista de paquetes... Hecho
Creando árbol de dependencias
Leyendo la información de estado... Hecho
Calculando la actualización... Hecho
Los paquetes indicados a continuación se instalaron de forma automática y ya no son necesarios.
  genders gir1.2-goa-1.0 libgenders0 python3-ply
Utilice «sudo apt autoremove» para eliminarlos.
Get more security updates through Ubuntu Pro with 'esm-apps' enabled:
```

6. Comprobamos la versión de java y confirmamos que no tiene la 21.

```
hadoop@dhcp2:~$ java -version
openjdk version "11.0.26" 2025-01-21
OpenJDK Runtime Environment (build 11.0.26+4-post-Ubuntu-1ubuntu120.04)
OpenJDK 64-Bit Server VM (build 11.0.26+4-post-Ubuntu-1ubuntu120.04, mixed mode, sharing)
hadoop@dhcp2:~$
```

7. Instalamos la versión 21 de Java.

```
hadoop@dhcp2:~$ sudo apt install openjdk-21-jdk -y
Leyendo lista de paquetes... Hecho
Creando árbol de dependencias
Leyendo la información de estado... Hecho
Los paquetes indicados a continuación se instalaron de forma automática y
son necesarios
```

8. Comprobamos que la versión de nuevo y ya tenemos la versión 21.

```
hadoop@dhcp2:~$ java -version
openjdk version "21.0.6" 2025-01-21
OpenJDK Runtime Environment (build 21.0.6+7-Ubuntu-120.04.1)
OpenJDK 64-Bit Server VM (build 21.0.6+7-Ubuntu-120.04.1, mixed mode, sharing)
hadoop@dhcp2:~$
```

9. En caso de que no nos aparezca todavía puede ser que tengamos varias versiones instaladas de Java. Para ello ejecutamos el comando “sudo update-alternatives –config java” y seleccionamos la nueva.

10. Nos metemos en el siguiente fichero.

```
hadoop@dhcp2:~$ nano ~/.bashrc
```

11. Exportamos la variable de entorno JAVA\_HOME con el siguiente valor, /usr/lib/jvm/java-1.21.0-openjdk.amd64.

```
export JAVA_HOME=/usr/lib/jvm/java-1.21.0-openjdk-amd64
```

12. Guardamos y reiniciamos el bashrc. Si queremos comprobar que se ha realizado correctamente podemos hacerlo con “echo \$JAVA\_HOME”.

```
hadoop@dhcp2:~$ source ~/.bashrc
hadoop@dhcp2:~$
```

## 2. Trabajando con NiFi

1. Ahora si podemos iniciar nifi.

```
hadoop@dhcp2:~/Descargas/nifi-2.2.0/bin$ ./nifi.sh start
JAVA_HOME=/usr/lib/jvm/java-1.21.0-openjdk-amd64
NIFI_HOME=/home/hadoop/Descargas/nifi-2.2.0
```

2. Ejecutamos el siguiente comando: `./nifi.sh set-single-user-credentials <usuario> <contraseña>`. Nuestra función es cambiar donde pone usuario y contraseña por los valores que queramos.

```
hadoop@dhcp2:~/Descargas/nifi-2.2.0/bin$ ./nifi.sh set-single-user-credentials u
suario usuario12345
JAVA_HOME=/usr/lib/jvm/java-1.21.0-openjdk-amd64
NIFI_HOME=/home/hadoop/Descargas/nifi-2.2.0
Login Identity Providers Processed [/home/hadoop/Descargas/nifi-2.2.0/./conf/log
in-identity-providers.xml]
hadoop@dhcp2:~/Descargas/nifi-2.2.0/bin$
```

3. Reiniciamos nifi para que pille de nuevo los cambios del nuevo usuario.

```
hadoop@dhcp2:~/Descargas/nifi-2.2.0/bin$ ./nifi.sh restart

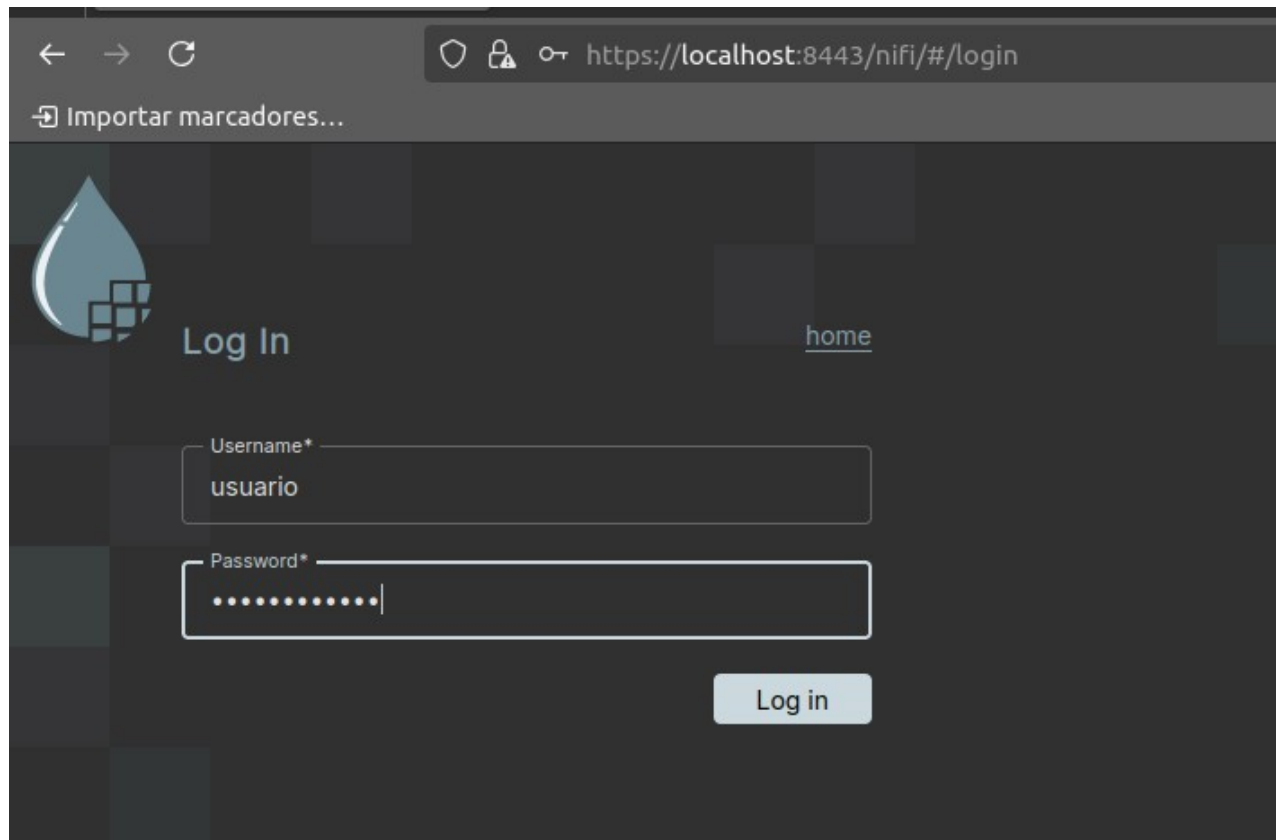
JAVA_HOME=/usr/lib/jvm/java-1.21.0-openjdk-amd64
NIFI_HOME=/home/hadoop/Descargas/nifi-2.2.0

2025-02-12 18:21:42,845 INFO [main] org.apache.nifi.bootstrap.Command Bootstrap
Process [28235] termination requested
2025-02-12 18:21:43,490 INFO [main] org.apache.nifi.bootstrap.Command Bootstrap
Process [28235] termination completed
2025-02-12 18:21:43,494 INFO [main] org.apache.nifi.bootstrap.Command Applicatio
n Process [28260] termination requested
2025-02-12 18:21:44,134 INFO [main] org.apache.nifi.bootstrap.Command Applicatio
n Process [28260] termination completed

JAVA_HOME=/usr/lib/jvm/java-1.21.0-openjdk-amd64
NIFI_HOME=/home/hadoop/Descargas/nifi-2.2.0

hadoop@dhcp2:~/Descargas/nifi-2.2.0/bin$
```

4. Para acceder a nifi, en un navegador ponemos la siguiente URL, <https://localhost:8443/nifi>. Nos mostrará una interfaz para iniciar sesión y las credenciales son las que hemos creado anteriormente.



**NOTA:** Me descargo XAMPP que ya trae MariaDB y MYSQL es el [enlace](#).

Ejecutamos los siguientes comando en phpmyadmin.

```
CREATE DATABASE nifi_db;
```

```
USE nifi_db;

CREATE TABLE steam_games (
  steam_appid VARCHAR(50),
  name VARCHAR(255),
  genres TEXT,
  release_date DATE,
  is_free BOOLEAN,
  price_usd DECIMAL(10,2),
  price_category VARCHAR(20)
);
```



```
CREATE USER 'nifi_user'@'localhost' IDENTIFIED BY 'nifi_password';  
GRANT ALL PRIVILEGES ON nifi_db.* TO 'nifi_user'@'localhost';  
FLUSH PRIVILEGES;
```

Descargar el .jar de los drivers [aquí](#). Nos descargamos el .zip:


**General Availability (GA) Releases** Archives ⓘ

## Connector/J 9.2.0

Select Operating System:

Platform Independent ▼

<b>Platform Independent (Architecture Independent), Compressed TAR Archive</b> (mysql-connector-j-9.2.0.tar.gz)	9.2.0	4.3M	<a href="#">Download</a>
	MD5: 7b5193cf541c99111309241886784aa7   <a href="#">Signature</a>		
<b>Platform Independent (Architecture Independent), ZIP Archive</b> (mysql-connector-j-9.2.0.zip)	9.2.0	5.1M	<a href="#">Download</a>
	MD5: 1535354d37b937bf792d618f5aafb6ab   <a href="#">Signature</a>		

 We suggest that you use the [MD5 checksums](#) and [GnuPG signatures](#) to verify the integrity of the packages you download.

Extraemos y buscamos el archivo mysql-connector-java-8.x.x.jar. Copiamos y pegamos en nifi/lib. Después reiniciamos NiFi.

```
hadoop@kaju-VMware-Virtual-Platform:~/nifi-2.2.0/bin$ ./nifi.sh restart
```

**NOTA:** Instalar los controladores de Hadoop en Nifi.

```
hadoop@kaju-VMware-Virtual-Platform:~/hadoop/etc/hadoop$ cp /home/hadoop/hadoop/  
etc/hadoop/core-site.xml /home/hadoop/nifi-2.2.0  
hadoop@kaju-VMware-Virtual-Platform:~/hadoop/etc/hadoop$ cp /home/hadoop/hadoop/  
etc/hadoop/hdfs-site.xml /home/hadoop/nifi-2.2.0  
hadoop@kaju-VMware-Virtual-Platform:~/hadoop/etc/hadoop$ S
```

```
hadoop@kaju-VMware-Virtual-Platform:~/nifi-2.2.0/lib$ cp /home/hadoop/hadoop/sha  
re/hadoop/hdfs/*.jar /home/hadoop/nifi-2.2.0/lib/
```

```
hadoop@kaju-VMware-Virtual-Platform:~/nifi-2.2.0/lib$ cp /home/hadoop/hadoop/sha  
re/hadoop/common/*.jar /home/hadoop/nifi-2.2.0/lib/
```


Descargar los .nar:

<https://mvnrepository.com/artifact/org.apache.nifi/nifi-hadoop-nar/2.2.0>

<https://mvnrepository.com/artifact/org.apache.nifi/nifi-hadoop-libraries-nar/2.2.0>


```
hadoop@kaju-VMware-Virtual-Platform:~/nifi-2.2.0/bin$ ./nifi.sh restart
```






## 1. Básico: Un CSV o JSON cualquiera. Previamente descargamos desde Kaggle.




<b>ObtenerFichero</b> GetFile 2.2.0 org.apache.nifi - nifi-standard-nar		
In	0 (0 bytes)	5 min
Read/Write	19,45 MB / 19,45 MB	5 min
Out	1 (19,45 MB)	5 min
Tasks/Time	1 / 00:00:00.028	5 min

Edit Processor   GetFile 2.2.0	
Settings	Scheduling
Properties	
Required field +	
Property	Value
Input Directory	/home/hadoop/datasets
File Filter	[^\.]*
Path Filter	No value set
Batch Size	10
Keep Source File	true
Recurse Subdirectories	true
Polling Interval	0 sec
Ignore Hidden Files	true

	<b>ConvertirCsvJson</b> ConvertRecord 2.2.0 org.apache.nifi - nifi-standard-nar	
In	1 (19,45 MB)	5 min
Read/Write	19,45 MB / 41,5 MB	5 min
Out	1 (41,5 MB)	5 min
Tasks/Time	1 / 00:00:00.925	5 min

Processor Details   ConvertRecord 2.2.0	
Settings	Scheduling
Properties	
Required field	
Property	Value
Record Reader	 CSVReader v1-1 
Record Writer	 JsonRecordSetWriter v1-1 
Include Zero Record FlowFiles	 true

	<b>TransformerExtensionCsVJson</b> UpdateAttribute 2.2.0 org.apache.nifi - nifi-update-attribute-nar	
In	1 (41,5 MB)	5 min
Read/Write	0 bytes / 0 bytes	5 min
Out	1 (41,5 MB)	5 min
Tasks/Time	1 / 00:00:00.001	5 min

### Processor Details | UpdateAttribute 2.2.0






Settings      Scheduling      **Properties**

**Required field**

Property	Value
Delete Attributes Expression	No value set
Store State	Do not store state
Stateful Variables Initial Value	No value set
Cache Value Lookup Cache Size	100
filename	\${filename:replaceAll('.csv', '.json')}

**FiltrarAtributos**  
JoltTransformJSON 2.2.0  
org.apache.nifi - nifi-jolt-nar

In	0 (0 bytes)	5 min
Read/Write	0 bytes / 0 bytes	5 min
Out	0 (0 bytes)	5 min
Tasks/Time	0 / 00:00:00.000	5 min

Processor Details   JoltTransformJSON 2.2.0	
Settings	Scheduling
Properties	
Required field	
Property	Value
Jolt Transform	 Chain
Jolt Specification	 [ { "operation": "shift", "spec": { "**": { ...
Transform Cache Size	 1
Pretty Print	 false
Max String Length	 20 MB

```
[
  {
    "operation": "shift",
    "spec": {
      "**": {
        "steam_appid": "[&1].steam_appid",
        "name": "[&1].name",
        "genres": "[&1].genres",
        "release_date": "[&1].release_date",
        "is_free": "[&1].is_free",
        "price_initial": "[&1].price_usd"
      }
    }
  }
]
```

```
}  
]
```



The image shows a NiFi processor card for 'PasarIdString'. It includes a play button icon, the processor name, version (JoltTransformJSON 2.2.0), and organization (org.apache.nifi - nifi-jolt-nar). Below this is a table with performance metrics.

In	0 (0 bytes)	5 min
Read/Write	0 bytes / 0 bytes	5 min
Out	0 (0 bytes)	5 min
Tasks/Time	0 / 00:00:00.000	5 min


Processor Details   JoltTransformJSON 2.2.0	
Settings	Scheduling
Properties	
Required field	
Property	Value
Jolt Transform	<i>i</i> Chain
Jolt Specification	<i>i</i> [ { "operation": "modify-overwrite-b... <i>i</i>
Transform Cache Size	<i>i</i> 1
Pretty Print	<i>i</i> false
Max String Length	<i>i</i> 20 MB

```
[  
{  
  "operation": "modify-overwrite-beta",
```

```

"spec": {
  "**": {
    "steam_appid": "=toString(@(1,steam_appid))"
  }
}
}
]

```










**AnadirColumna**

UpdateRecord 2.2.0

org.apache.nifi - nifi-standard-nar

In	0 (0 bytes)	5 min
Read/Write	0 bytes / 0 bytes	5 min
Out	0 (0 bytes)	5 min
Tasks/Time	0 / 00:00:00.000	5 min


Edit Processor   UpdateRecord 2.2.0		
Settings	Scheduling	Properties
Required field <span>+</span>		
Property	Value	
Record Reader	 JsonTreeReader v1-1	
Record Writer	 JsonRecordSetWriter v1-2	
Replacement Value Strategy	 Literal Value	
/price_category	 \${price_usd:toNumber():gt(20):ifElse('no barato', 'barato')}	




```


${price_usd:toNumber():gt(20):ifElse('no barato', 'barato')}

```



	<b>GuardarFichero</b> PutFile 2.2.0 org.apache.nifi - nifi-standard-nar	
In	0 (0 bytes)	5 min
Read/Write	0 bytes / 0 bytes	5 min
Out	0 (0 bytes)	5 min
Tasks/Time	0 / 00:00:00.000	5 min

Edit Processor   PutFile 2.2.0	
Settings	Scheduling
Properties	
Required field	
Property	Value
Directory	 /home/hadoop/datasets-creados
Conflict Resolution Strategy	 replace
Create Missing Directories	 true

	<b>PutDatabaseRecord</b> PutDatabaseRecord 2.2.0 org.apache.nifi - nifi-standard-nar	
In	0 (0 bytes)	5 min
Read/Write	0 bytes / 0 bytes	5 min
Out	0 (0 bytes)	5 min
Tasks/Time	0 / 00:00:00.000	5 min

**Edit Processor** | PutDatabaseRecord 2.2.0









Settings

Scheduling










Properties

Required field +

Property	Value
Record Reader	<span>i</span> JsonTreeReader v1-2 <span>⋮</span>
Database Type	<span>i</span> Generic
Statement Type	<span>i</span> INSERT
Data Record Path	<span>i</span> No value set
Database Connection Pooling Servi...	<span>i</span> DBCPConnectionPool v1-1 <span>⋮</span>
Catalog Name	<span>i</span> No value set
Schema Name	<span>i</span> No value set
Table Name	<span>i</span> steam_games

Table Name	 steam_games
Binary String Format	 UTF-8
Translate Field Names	 true
Column Name Translation Strategy	 Remove Underscore
Unmatched Field Behavior	 Ignore Unmatched Fields
Unmatched Column Behavior	 Fail on Unmatched Columns
Quote Column Identifiers	 false
Quote Table Identifiers	 false









Configuración del reader.



Controller Service Details	
Settings	Properties
Required field	
Property	Value
Schema Access Strategy	 Use 'Schema Text' Property
Schema Text	 { "name": "steam_games_schema", "... 
Starting Field Strategy	 Root Node
Max String Length	 20 MB
Allow Comments	 false
Date Format	 No value set
Time Format	 No value set
Timestamp Format	 No value set









```
{
  "name": "steam_games_schema",
```

```
"namespace": "nifi",
"type": "record",
"fields": [
  { "name": "steam_appid", "type": "string" },
  { "name": "name", "type": "string" },
  { "name": "genres", "type": "string" },
  { "name": "release_date", "type": "string" },
  { "name": "is_free", "type": "boolean" },
  { "name": "price_usd", "type": "double" },
  { "name": "price_category", "type": "string" }
]
```









Configuración del DBCPConnectionPool.

Property	Value
Database Connection URL	 jdbc:mysql://localhost:3306/nifi_db
Database Driver Class Name	 com.mysql.cj.jdbc.Driver
Database Driver Location(s)	 /home/hadoop/nifi-2.2.0/lib/mysql-c...
Kerberos User Service	 No value set
Database User	 nifi_user
Password	 Sensitive value set
Max Wait Time	 500 millis
Max Total Connections	 8

	 <b>GuardarHDFS</b> PutHDFS 2.2.0 org.apache.nifi - nifi-hadoop-nar	
In	0 (0 bytes)	5 min
Read/Write	0 bytes / 0 bytes	5 min
Out	0 (0 bytes)	5 min
Tasks/Time	0 / 00:00:00.000	5 min

Edit Processor   PutHDFS 2.2.0		
Settings	Scheduling	Properties
Required field		+
Property	Value	
Hadoop Configuration Resources		/usr/local/hadoop/bin/hadoop/core-sit...
Kerberos User Service		No value set 
Additional Classpath Resources		No value set
Directory		/user/hadoop/data-create
Conflict Resolution Strategy		replace
Writing Strategy		Write and rename
Block Size		No value set

/usr/local/hadoop/etc/hadoop/core-site.xml, /usr/local/hadoop/etc/hadoop/hdfs-site.xml

Required field		+
Property	Value	
IO Buffer Size	 No value set	
Replication	 No value set	
Permissions umask	 No value set	
Remote Owner	 No value set	
Remote Group	 No value set	
Compression codec	 NONE	
Ignore Locality	 false	
Resource Transfer Source	 FlowFile Content	

