Tema 3 Tarea 2 NiFi

Juan Manuel García Moyano
IABD
Informática y comunicaciones

Índice

1. Instalación de NiFi

1. Descargamos Apache NiFi.

```
hadoop@dhcp2:~/Descargas$ wget https://dlcdn.apache.org/nifi/2.2.0/nifi-2.2.0-bi
n.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... conecta
do.
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 [729771289/72977
1289]

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 intalado 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.apach
e.nifi.bootstrap.BootstrapProcess
        java.lang.UnsupportedClassVersionError: org/apache/nifi/bootstrap/Bootst
rapProcess 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.

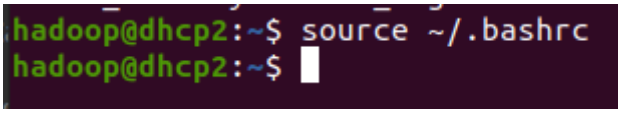
10. Nos metemos en el siguiente fichero.

```
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
```

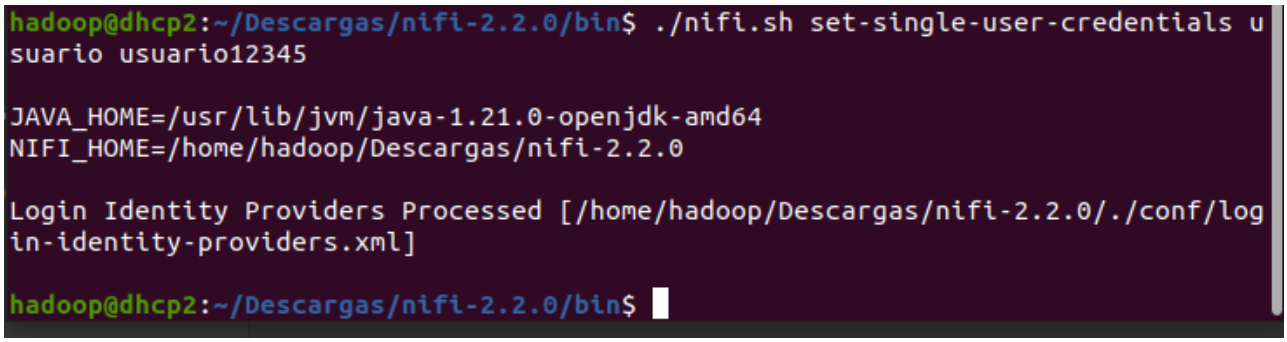
12. Guardamos y reiniciamos el bashrc. Si queremos comprobar que se ha realizado correctamente podemos hacerlo con “echo $JAVA_HOME”.



2. Trabajando con NiFi

1. Ahora si podemos iniciar nifi.

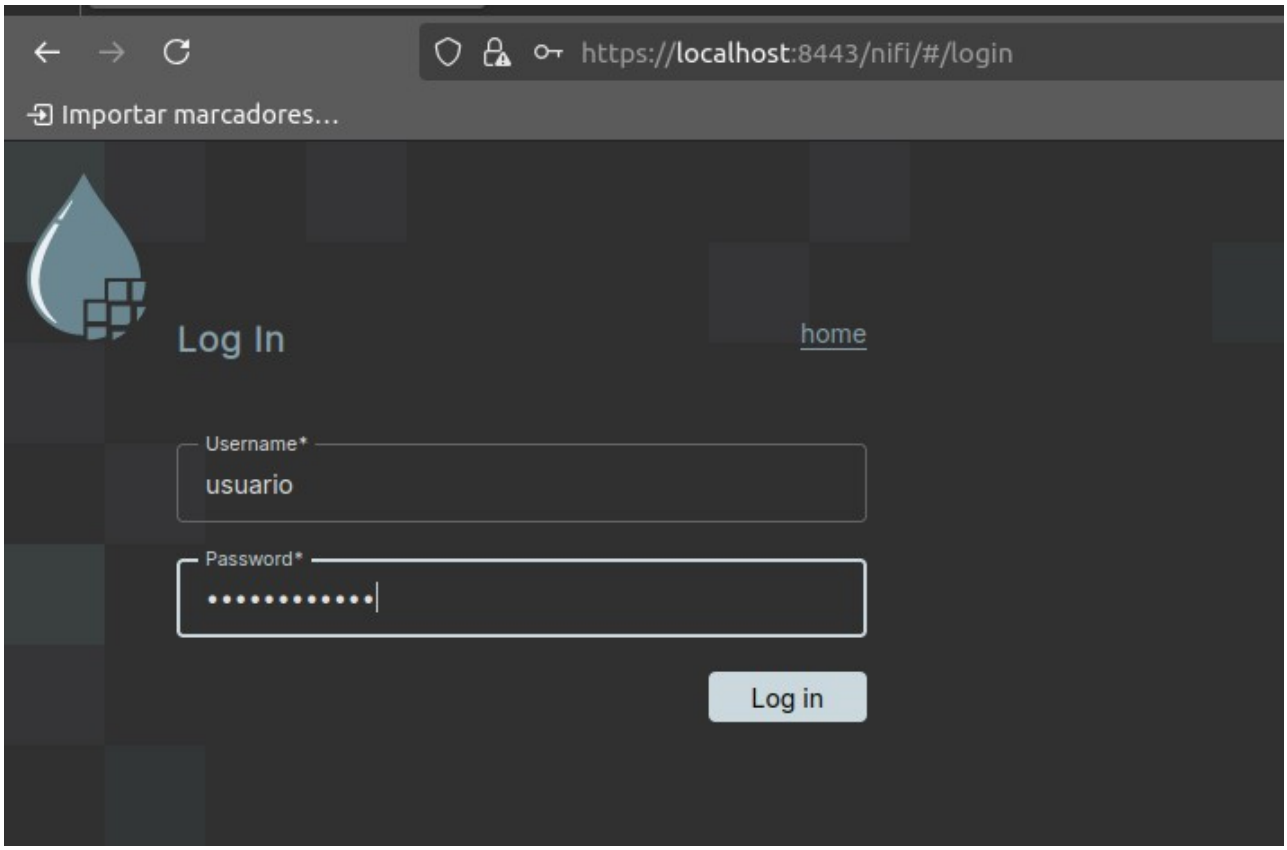
2. Ejecutamos el siguiente comando: ./nifi.sh set-single-user-credentials <usuario> <contraseña>. Nuestra función es cambiar donde pone usuario y contraseña por los valores que queramos.



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.

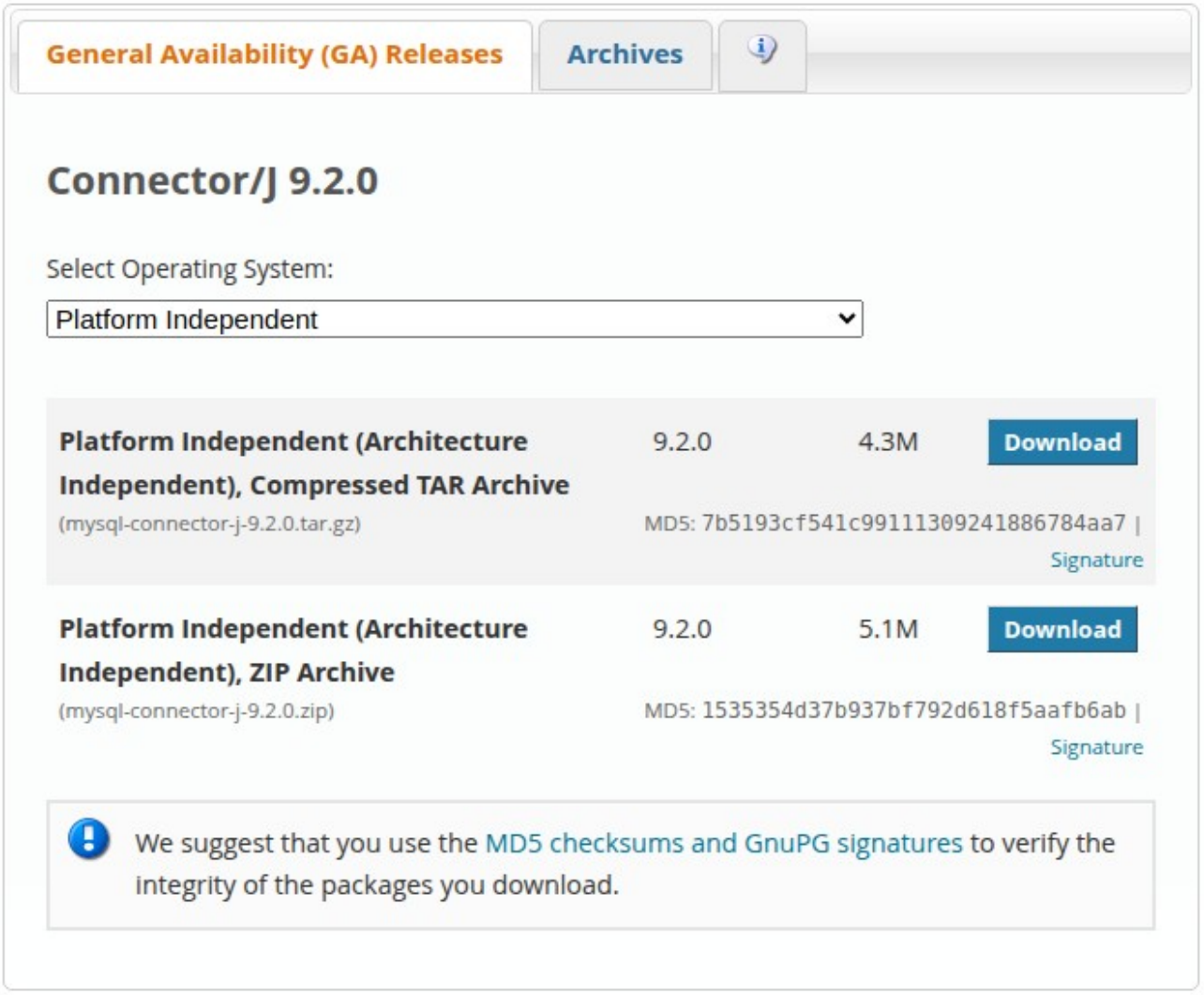
Ejecutamos los siguientes comando en phpmyadmin.

```
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:



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/
```

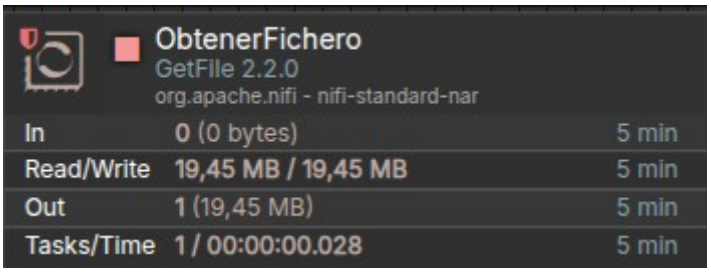
```
hadoop@kaju-VMware-Virtual-Platform:~/nifi-2.2.0/lib$ cp /home/hadoop/hadoop/sha
re/hadoop/common/*.jar /home/hadoop/nifi-2.2.0/lib/
```

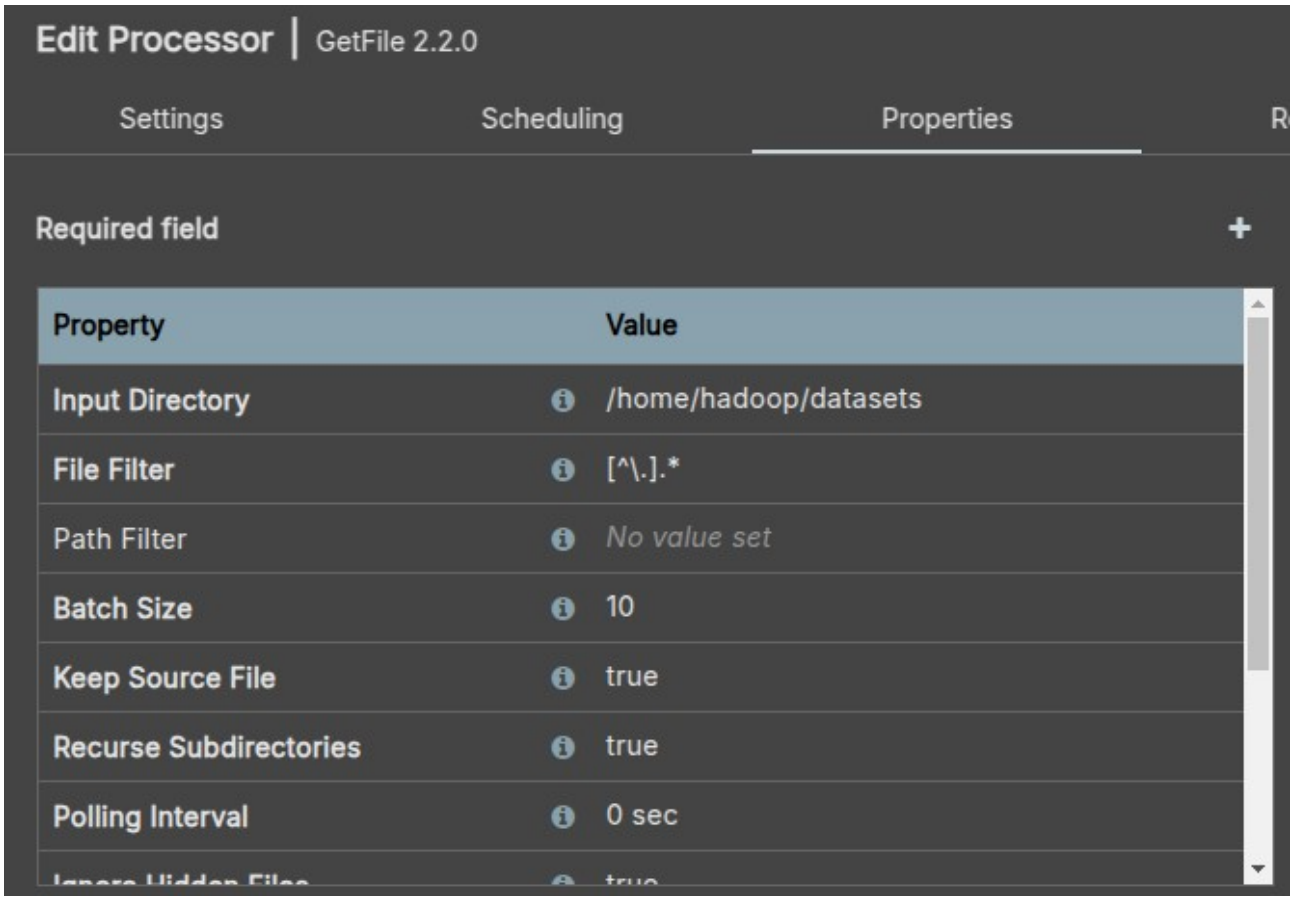
Descargar los .nar:

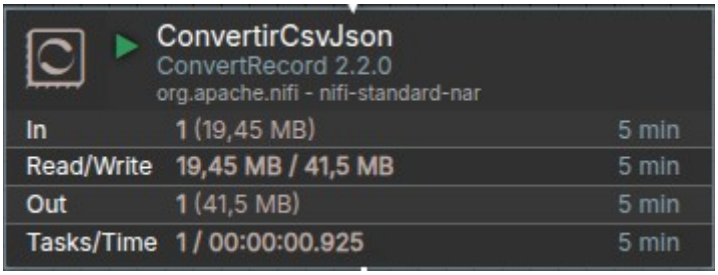
https://mvnrepository.com/artifact/org.apache.nifi/nifi-hadoop-nar/2.2.0
https://mvnrepository.com/artifact/org.apache.nifi/nifi-hadoop-libraries-nar/2.2.0

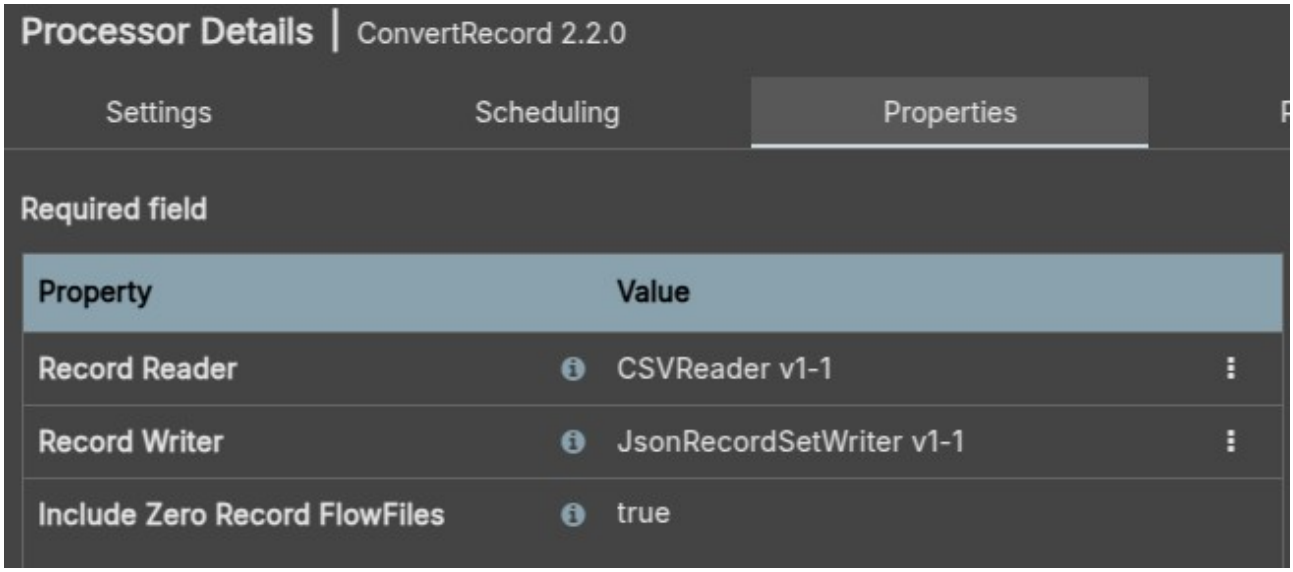
```
hadoop@kaju-VMware-Virtual-Platform:~/nifi-2.2.0/bin$ ./nifi.sh restart
```

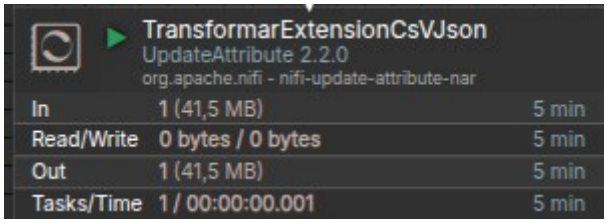
1. Básico: Un CSV o JSON cualquiera. Previamente descargamos desde Kaggle.

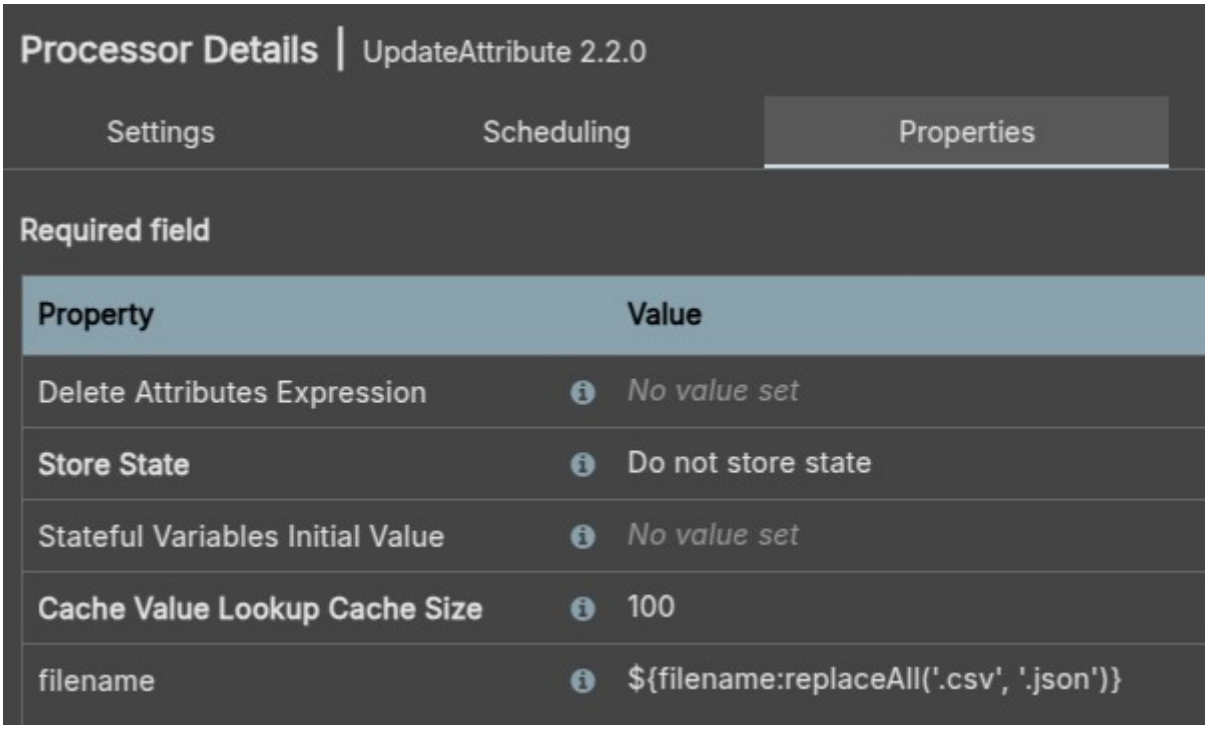


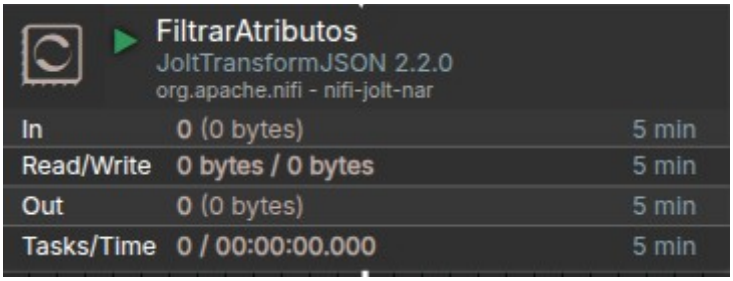


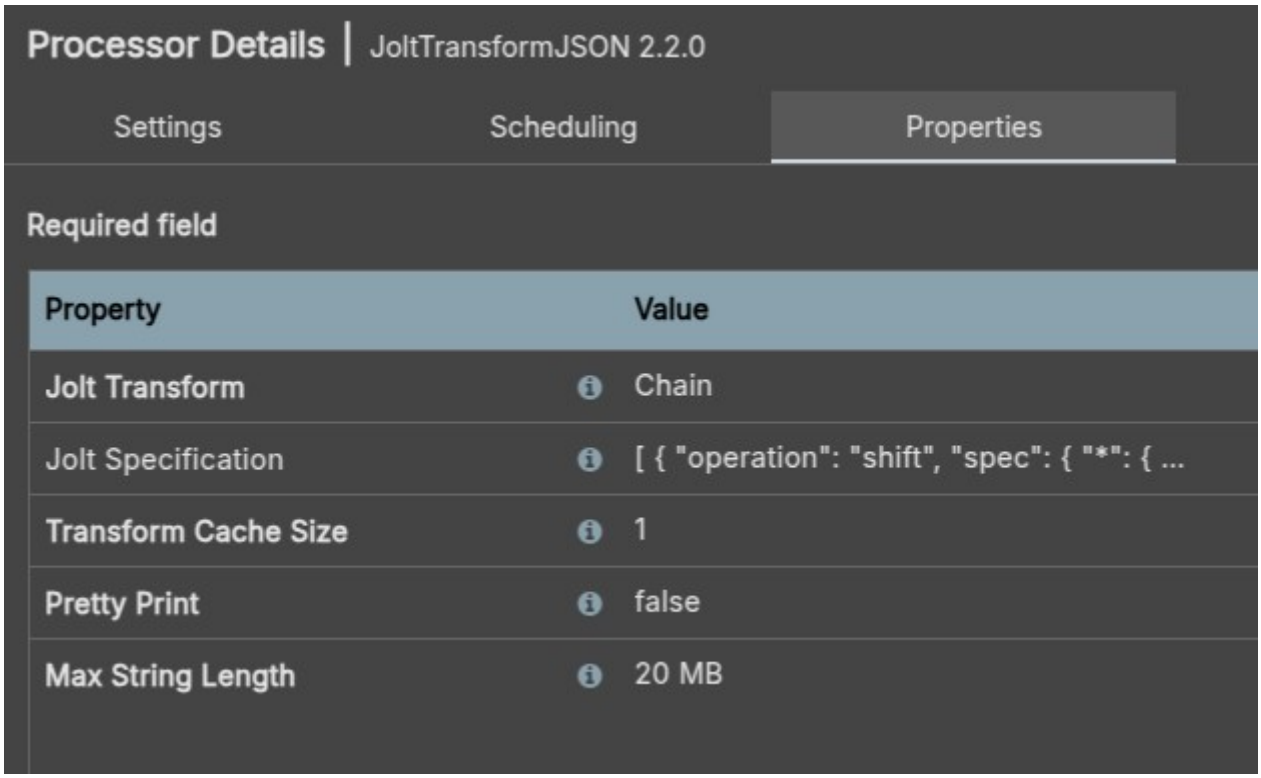

ConvertirCsvJson
ConvertRecord 2.2.0
org.apache.nifi - nifi-standard-nar
In 1 (19,45 MB) 5 min
Read/Write 19,45 MB / 41,5 MB 5 min
Out 1 (41,5 MB) 5 min
Tasks/Time 1 / 00:00:00.925 5 min


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


TransformarExtensionCsVJson
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

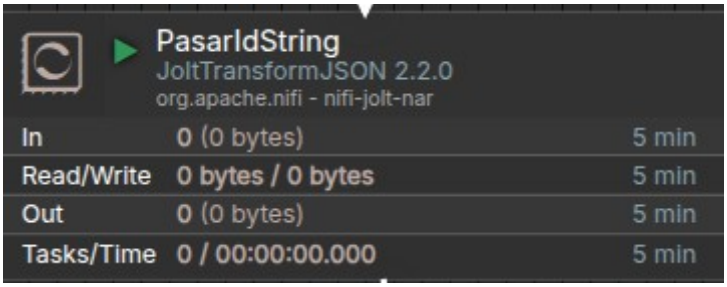

Processor Details | UpdateAttribute 2.2.0
Settings
Scheduling
Properties
Required field
Property
Value
Delete Attributes Expression
No value set
Store State
Do not store state
Stateful Variables Initial Value
No value set
Cache Value Lookup Cache Size
100
filename
${filename:replaceAll('.csv', '.json')}

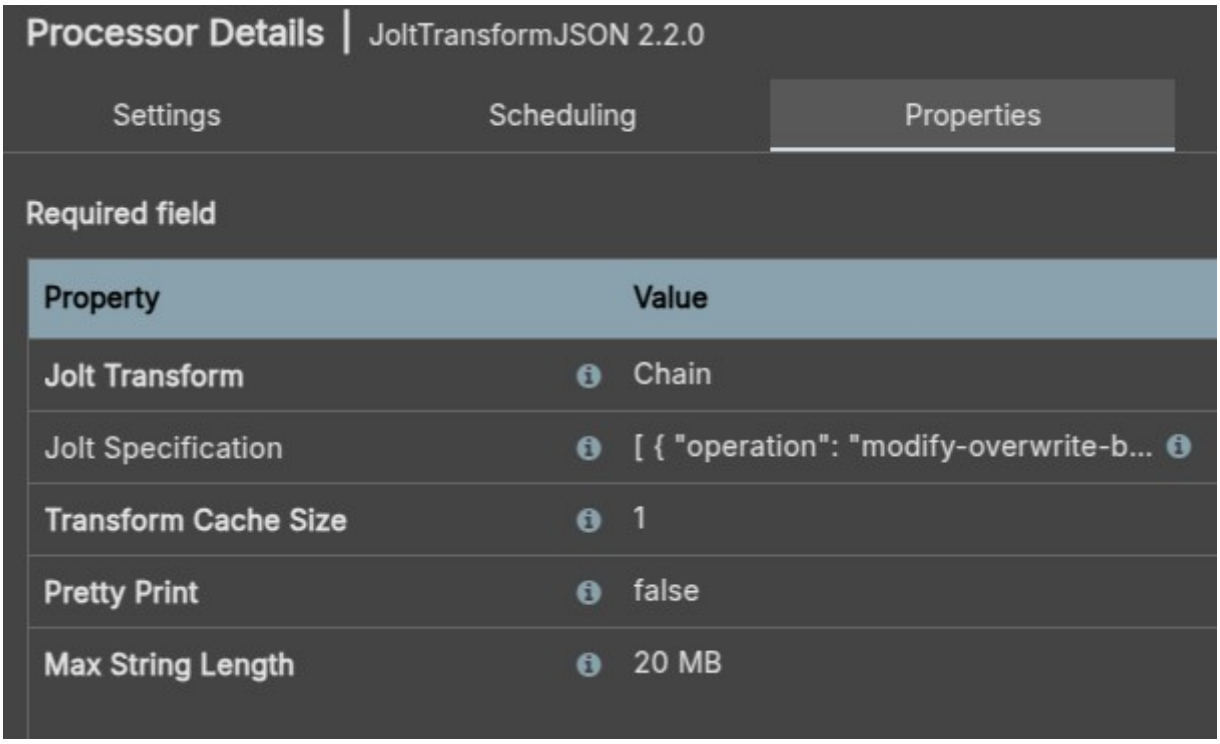

FiltrarAtributos
JoltTransformJSON 2.2.0
org.apache.nifi - nifi-jolt-nar
In 0 (0 bytes) 5 min
Read/Write 0 bytes / 0 bytes 5 min
Out 0 (0 bytes) 5 min
Tasks/Time 0 / 00:00:00.000 5 min




```
[
  {
    "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"
      }
    }
```

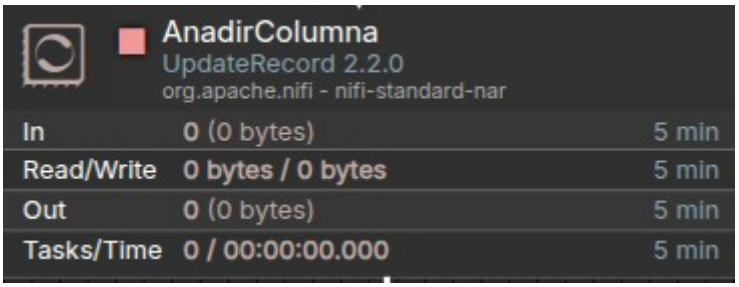
```
  }
]
```

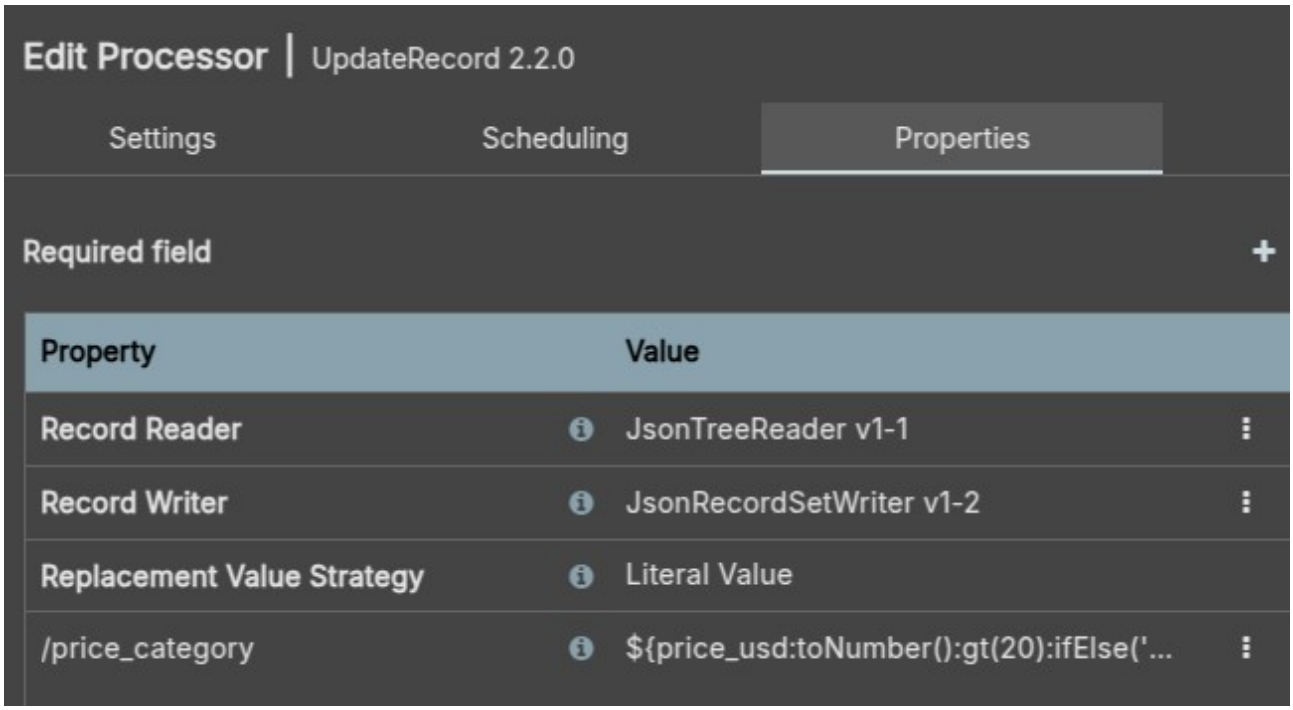




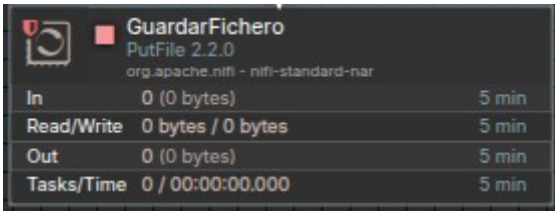

```
[
  {
    "operation": "modify-overwrite-beta",
```

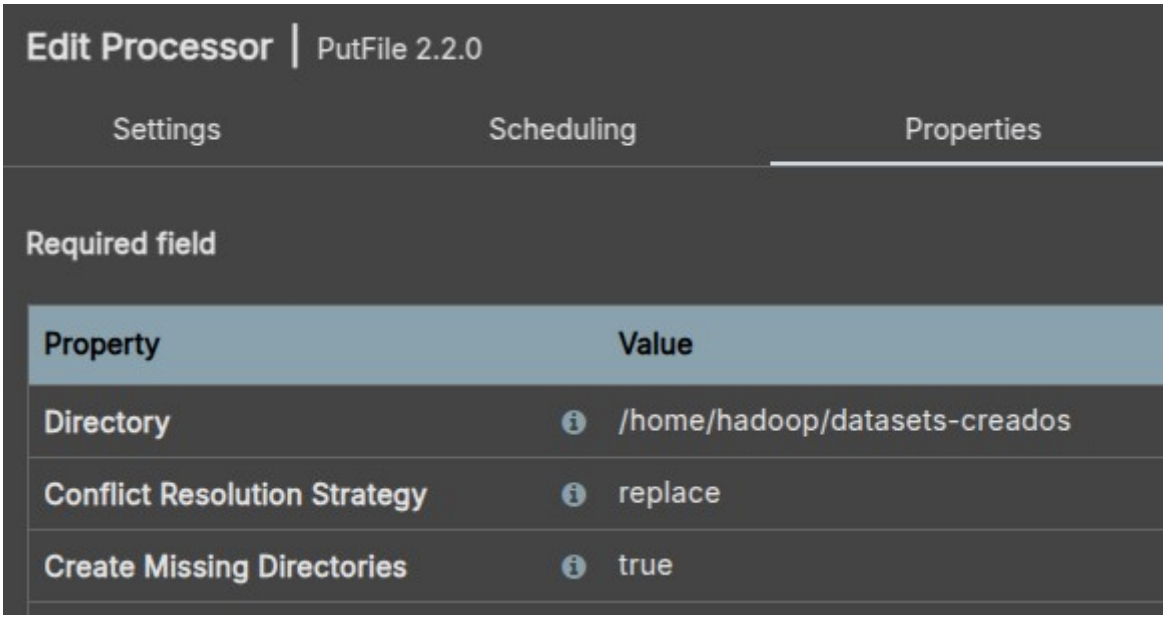
```
  "spec": {
    "*": {
      "steam_appid": "=toString(@(1,steam_appid))"
    }
  }
 }
]
```

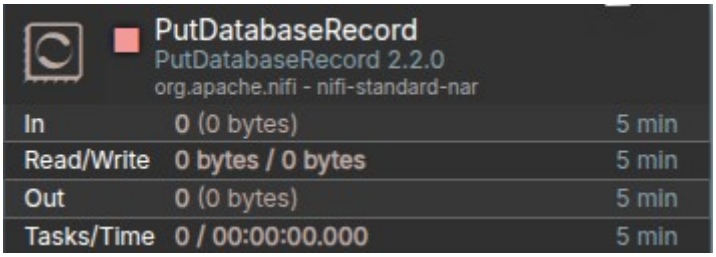


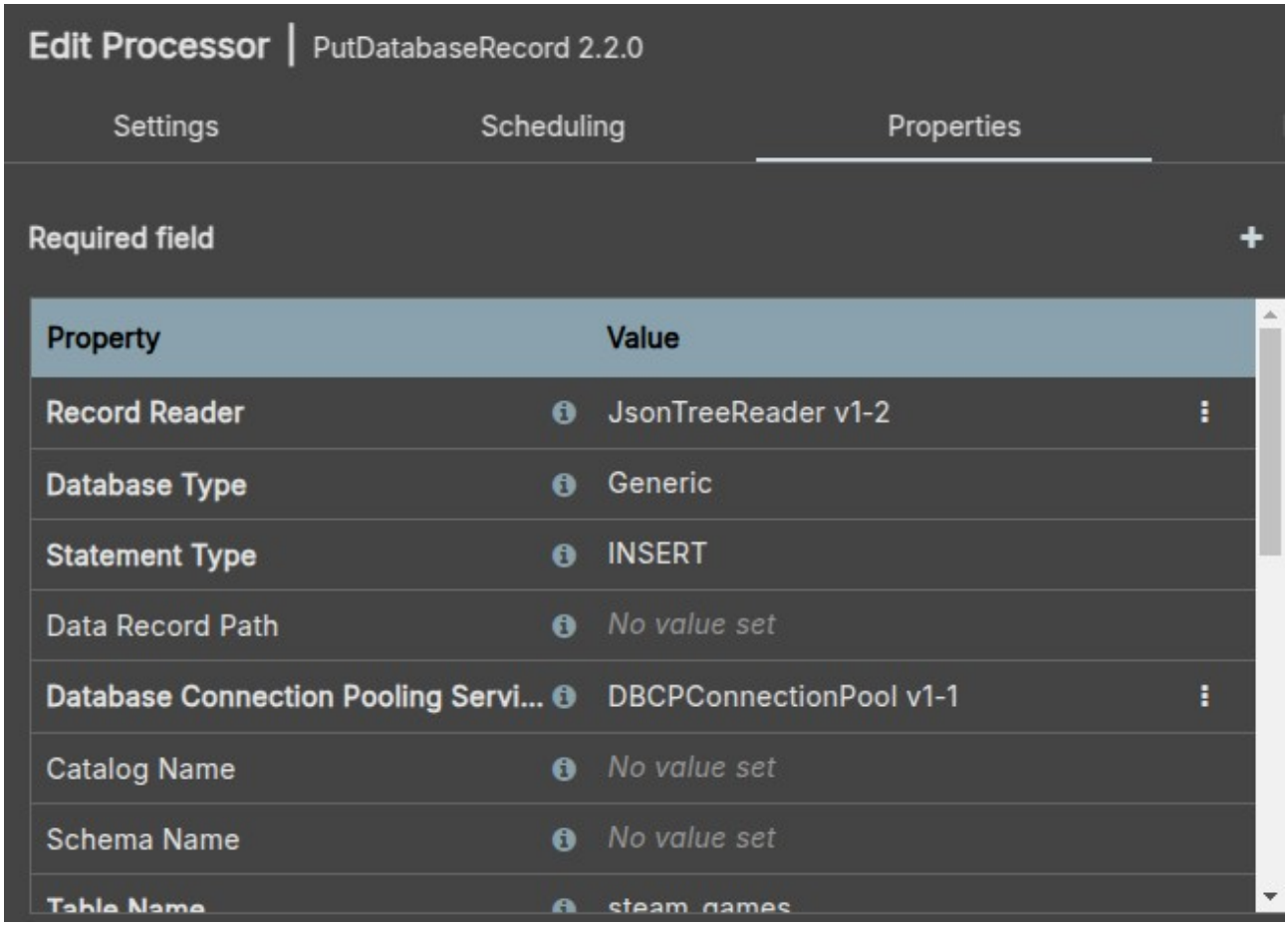


${price_usd:toNumber():gt(20):ifElse('no barato', 'barato')}

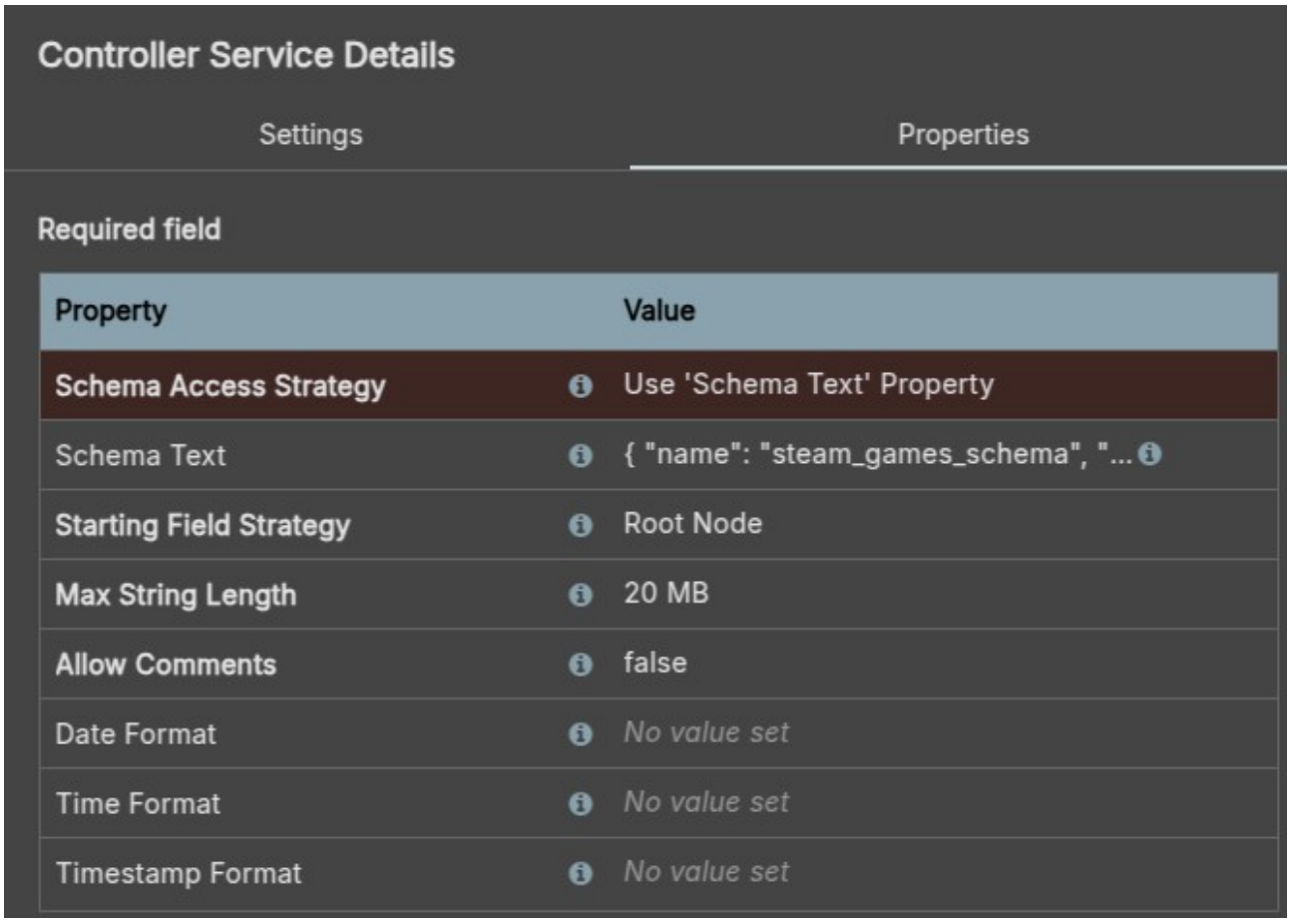

GuardarFichero
PutFile 2.2.0
org.apache.nifi - nifi-standard-nar
In 0 (0 bytes) 5 min
Read/Write 0 bytes / 0 bytes 5 min
Out 0 (0 bytes) 5 min
Tasks/Time 0 / 00:00:00.000 5 min


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


PutDatabaseRecord
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
JsonTreeReader v1-2
Database Type
Generic
Statement Type
INSERT
Data Record Path
No value set
Database Connection Pooling Servi...
DBCPConnectionPool v1-1
Catalog Name
No value set
Schema Name
No value set
Table Name
steam_games

Configuración del reader.

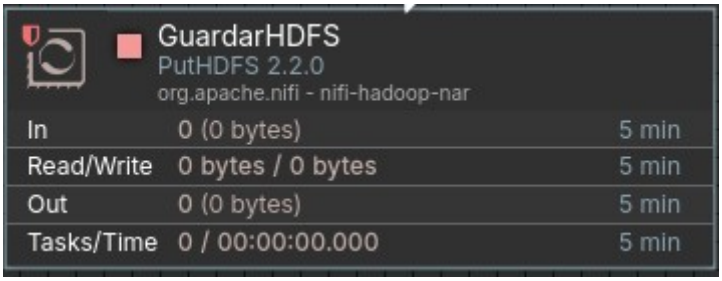


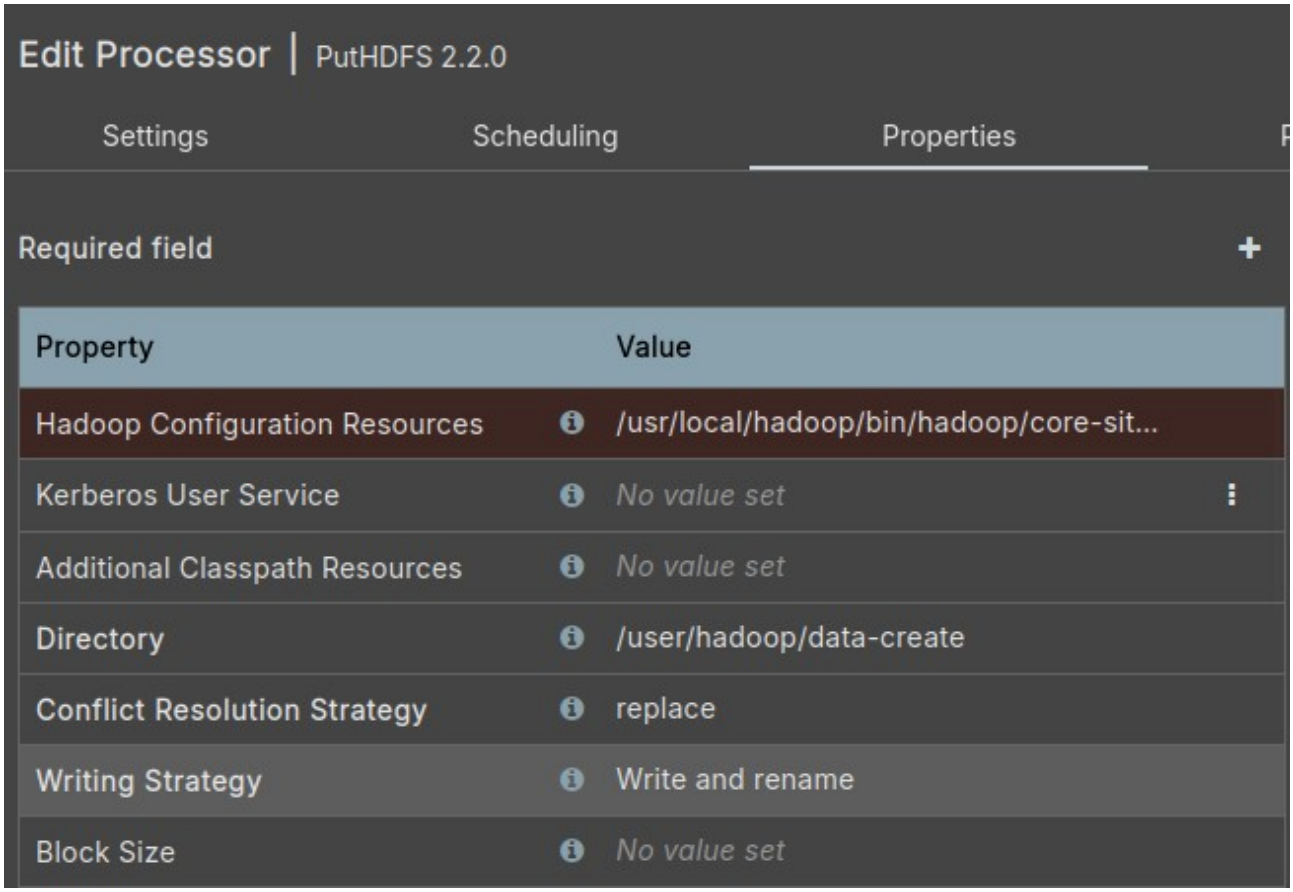

```
{
  "name": "steam_games_schema",
```

```
  "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" }
  ]
}
```

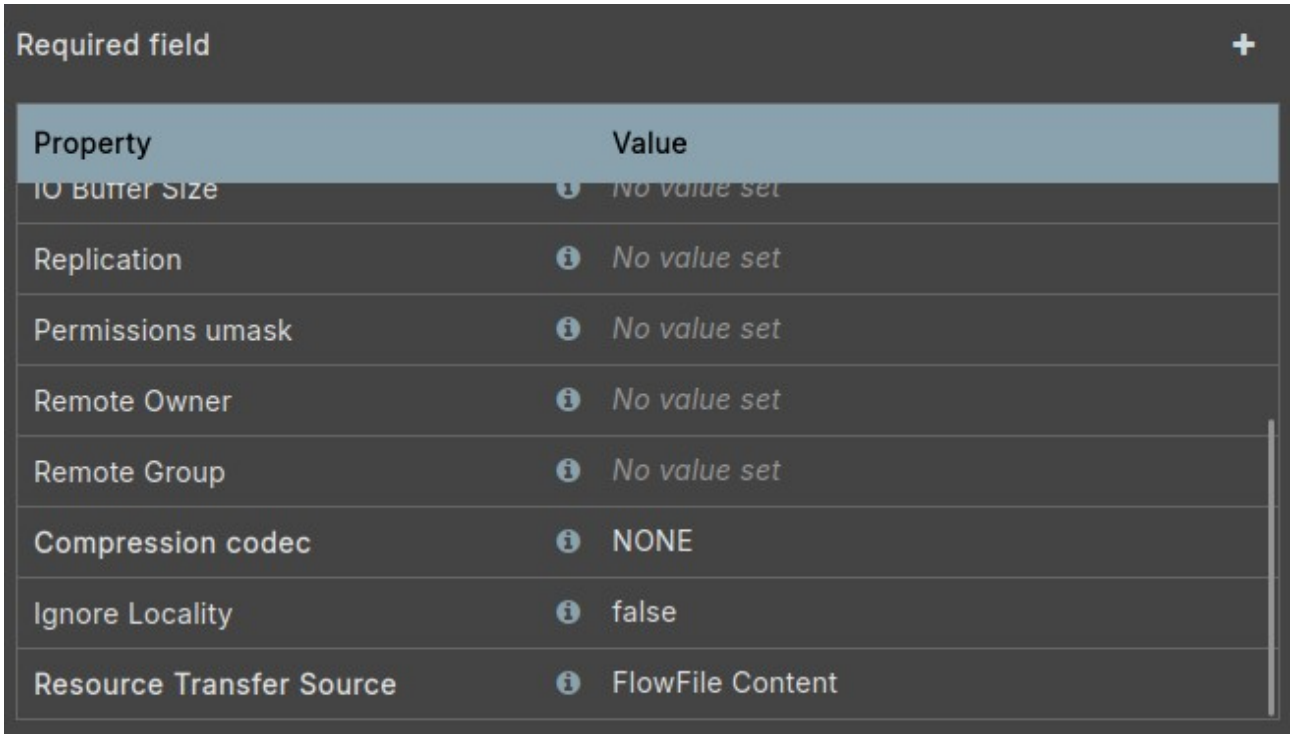
Configuración del DBCPConnectionPool.

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





/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