

## NIFI Docker

`docker pull apache/nifi`

### Correr el docker:

`docker run --name nifi -p 8443:8443 -d apache/nifi:latest`

Si necesitan correrlo con DNS público

`docker run --name nifi -p 8443:8443 --dns=8.8.8.8 -d apache/nifi:latest`

## UI WEB

`https://localhost:8443/nifi`

### Usr y pass

#buscar el usr y la pass en los logs

`docker logs nifi`

Linux: `docker logs nifi | grep Generated`

Windows: `docker logs nifi | findstrGenerated`

Ejemplo de lo que se debería ver:

Generated Username [5daf9eec-62d4-42cd-ad9c-acac425ee7f9]

Generated Password [2nYLm8aFtpszu+Ft0txirYOpXNGbrvXx]

Cambiar pass (opcional)

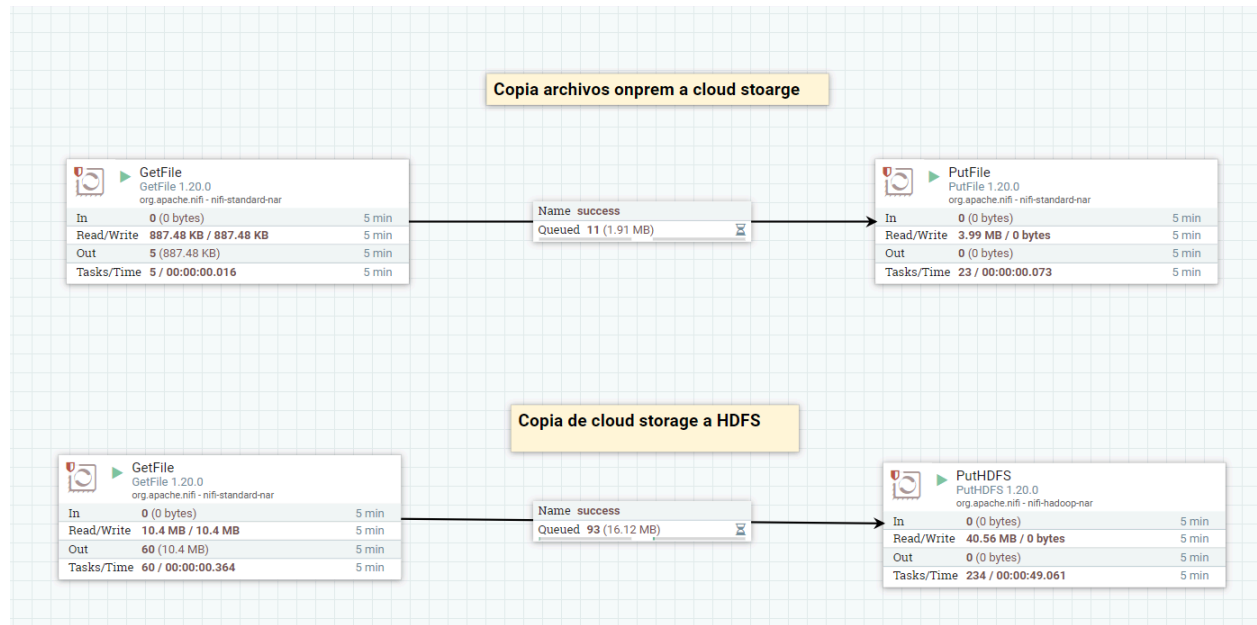
`./bin/nifi.sh set-single-user-credentials <username> <password>`

`/opt/nifi/nifi-current/bin nifi.sh set-single-user-credentials nifi <password>`

### Agregar processor:

Get file → Putfile

Get file → PutHDFS















## Configuración GetFile

SETTINGS	SCHEDULING	PROPERTIES	RELATIONSHIPS	COMMENTS
Required field <span>🔍</span> <span>+</span>				
Property		Value		
Input Directory		?	/home/nifi/ingest	
File Filter		?	starwars.csv	
Path Filter		?	No value set	
Batch Size		?	10	
Keep Source File		?	true	
Recurse Subdirectories		?	true	
Polling Interval		?	0 sec	
Ignore Hidden Files		?	true	
Minimum File Age		?	0 sec	
Maximum File Age		?	No value set	
Minimum File Size		?	0 B	
Maximum File Size		?	No value set	

SETTINGS	SCHEDULING	PROPERTIES	RELATIONSHIPS	COMMENTS
----------	------------	------------	---------------	----------

Required field

Property	Value	
Input Directory	 /home/nifi/bucket	
File Filter	 starwars.csv	
Path Filter	 No value set	
Batch Size	 10	
Keep Source File	 true	
Recurse Subdirectories	 true	
Polling Interval	 0 sec	
Ignore Hidden Files	 true	
Minimum File Age	 0 sec	
Maximum File Age	 No value set	
Minimum File Size	 0 B	
Maximum File Size	 No value set	

Si queremos solo procesar CSV:

```
[^\.]*\.*\.csv
```

En el shell de nifi:









1. Crear una carpeta llamada ingest en "/home/nifi"
2. Bajar starwars.csv con wget  
<https://github.com/fpineyro/homework-0/blob/master/starwars.csv>

## Configuración PutFile

SETTINGS	SCHEDULING	PROPERTIES	RELATIONSHIPS	COMMENTS
----------	------------	------------	---------------	----------

Required field



Property	Value	
Directory	 /home/nifi/bucket	
Conflict Resolution Strategy	 replace	
Create Missing Directories	 true	
Maximum File Count	 No value set	
Last Modified Time	 No value set	
Permissions	 No value set	
Owner	 No value set	
Group	 No value set	

En el shell de nifi:

1. Crear una carpeta llamada bucket en "/home/nifi"

## Configuración PutHDFS

Copiar los archivos (o el contenido) de core-site.xml y hdfs-site.xml a una carpeta determinada, las copie en /home/nifi/hadoop.

Luego en Hadoop Configuration Resources configurar lo siguiente:  
/home/nifi/hadoop/core-site.xml, /home/nifi/hadoop/hdfs-site.xml

Directory: donde queremos que se copie la info

Darle permisos a la carpeta /nifi con el siguiente comando desde el shell de hadoop:

Hdfs dfs -chmod 777 /nifi

**Configure Processor** | PutHDFS 1.20.0

Stopped

SETTINGS | SCHEDULING | **PROPERTIES** | RELATIONSHIPS | COMMENTS

Required field 🔍 +

Property	Value
Hadoop Configuration Resources	🔍 /home/nifi/hadoop/core-site.xml, /home/nifi/hadoop...
Kerberos Credentials Service	🔍 No value set
Kerberos User Service	🔍 No value set
Kerberos Principal	🔍 No value set
Kerberos Keytab	🔍 No value set
Kerberos Password	🔍 No value set
Kerberos Rlogin Period	🔍 4 hours
Additional Classpath Resources	🔍 No value set
<b>Directory</b>	🔍 /nifi
<b>Conflict Resolution Strategy</b>	🔍 replace
<b>Writing Strategy</b>	🔍 Write and rename
Block Size	🔍 No value set

Fuente:

<https://github.com/fpineyro/homework-0/blob/master/starwars.csv>

## Core-site.xml de ejemplo

```
<?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://172.17.0.2:9000</value>
  </property>
</configuration>
```

## Hdfs-site.xml de ejemplo

```
<?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/hadoop/hadoopdata/hdfs/namenode</value>
  </property>

  <property>
    <name>dfs.data.dir</name>
    <value>file:///home/hadoop/hadoopdata/hdfs/datanode</value>
  </property>
</configuration>
```