



Ejemplo 1 - Moviendo datos

Vamos a hacer un pequeño ejercicio con *Nifi* para familiarizarnos con el entorno desarrollando un flujo de datos sencillo que mueva un fichero de un directorio a otro.

A continuación detallamos los pasos a realizar:

1. Seleccionamos un procesador (primer icono grande) y lo arrastramos en nuestra área de trabajo.
2. Nos aparece un diálogo con tres partes diferenciadas:

The screenshot shows the 'Add Processor' dialog box in Apache NiFi. It features a search bar at the top with the text 'getf'. Below the search bar, there is a table displaying two processors: 'GetFTP' and 'GetFile'. The 'GetFile' processor is highlighted in yellow. To the left of the table, there is a list of tags including 'amazon', 'attributes', 'avro', 'aws', 'azure', 'consume', 'csv', 'database', 'delete', 'fetch', 'get', 'hadoop', 'ingest', 'insert', 'json', 'listen', 'logs', 'message', 'pubsub', 'put', 'record', 'restricted', 'source', 'text', and 'update'. Below the table, there is a description for the 'GetFile 1.14.0' processor, stating that it creates FlowFiles from files in a directory and that NiFi will ignore files it doesn't have at least read permissions for. At the bottom right, there are 'CANCEL' and 'ADD' buttons.

Type	Version	Tags
GetFTP	1.14.0	input, FTP, get, fetch, retrieve, fil...
GetFile	1.14.0	ingress, input, restricted, get, fil...

GetFile 1.14.0 org.apache.nifi - nifi-standard-nar

Creates FlowFiles from files in a directory. NiFi will ignore files it doesn't have at least read permissions for.

Diálogo de elección de procesador

- A la izquierda una nube de etiquetas para poder filtrar los procesadores.
- Arriba a la derecha tenemos un buscador para buscar procesadores por su nombre
- La parte central con el listado de procesadores, desde donde lo podemos seleccionar.

3. Así pues, buscamos el procesador *GetFile* y lo añadimos al flujo, el cual permite recuperar un fichero desde una carpeta.
4. Damos doble click sobre el elemento gráfico que representa nuestro procesador, y en la pestaña *properties* indicamos el directorio de entrada de donde tendrá que recoger el fichero mediante la propiedad *Input Directory*. En nuestro caso le pondremos el valor `/Users/miguelangelronda/Downloads/in:`

Configure Processor | GetFile 2.0.0-M1

Invalid

SETTINGS

SCHEDULING

PROPERTIES

RELATIONSHIPS

COMMENTS

Required field

Property		Value
Input Directory	?	/Users/miguelangelronda/Downloads/in
File Filter	?	[^\\].*
Path Filter	?	No value set

Propiedades de GetFile

5. Antes de cambiar de procesador, en la pestaña *Settings*, en la caja *Name* la rellenamos con *ObtenerFichero*.

6. Ahora añadimos un nuevo procesador de tipo *PutFile*, y en las propiedades indicamos el directorio de salida con la propiedad *directory* a */Users/miguelangelronda/Downloads/out*.
7. Si visualizamos la pestaña *Settings*, una buena práctica es ponerle nombre a los procesadores. Así pues, la caja *Name* la rellenamos con *PonerFichero*.
8. En la pestaña *Relationships*, podemos configurar el comportamiento a seguir si el procesador se ejecuta correctamente (*success*) o falla (*failure*). Como vamos a hacer que este procesador sea el paso final, vamos a configurar que *autoterminen* marcando ambos casos:

Configure Processor | PutFile 2.0.0-M1

Invalid

SETTINGS

SCHEDULING

PROPERTIES

RELATIONSHIPS

COMMENTS

Automatically Terminate / Retry Relationships ?

failure

☒ terminate

☐ retry

Files that could not be written to the output directory for some reason are transferred to this relationship

success

☒ terminate

☐ retry

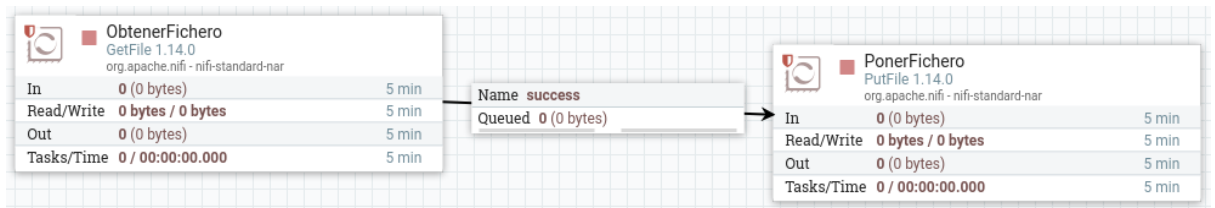
Files that have been successfully written to the output directory are transferred to this relationship

Finalización de PutFile

Nota: Terminar las relaciones

Si nos olvidamos de autoterminar las relaciones, o tenemos conexiones sin conectar, no podremos iniciar los procesadores implicados. Esto lo tenemos que realizar para todos los procesadores que tengamos en nuestro flujo de datos.

9. Unimos ambos procesadores creando una conexión. Para ello, tras pulsar sobre el icono de la flecha que aparece al dejar el ratón sobre el primer procesador y lo arrastramos hasta el segundo.



Create Connection

DETAILS

SETTINGS

From Processor

Obtener Fichero
GetFile

Within Group

NiFi Flow

For Relationships

☒ **success**

To Processor

Poner Fichero
PutFile

Within Group

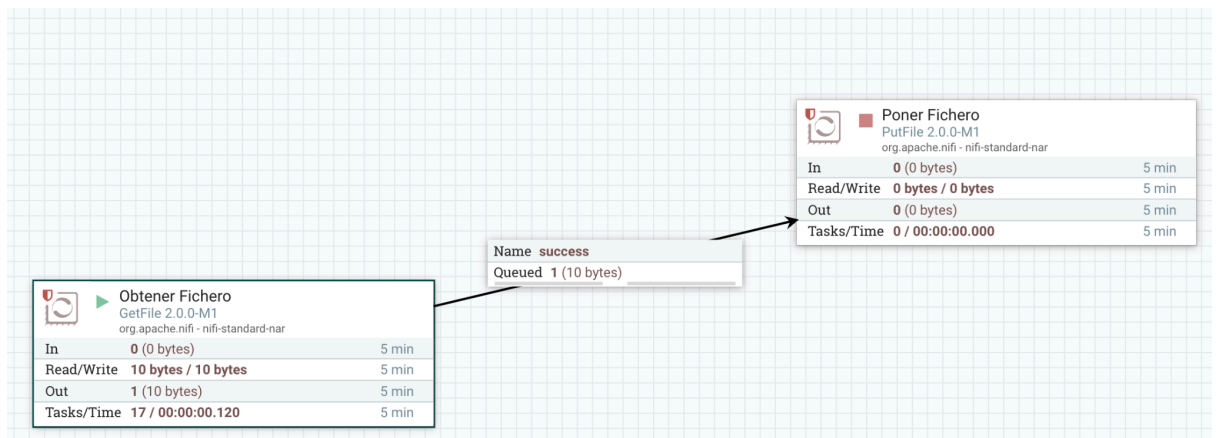
NiFi Flow

Conexión mediante un conector entre procesadores

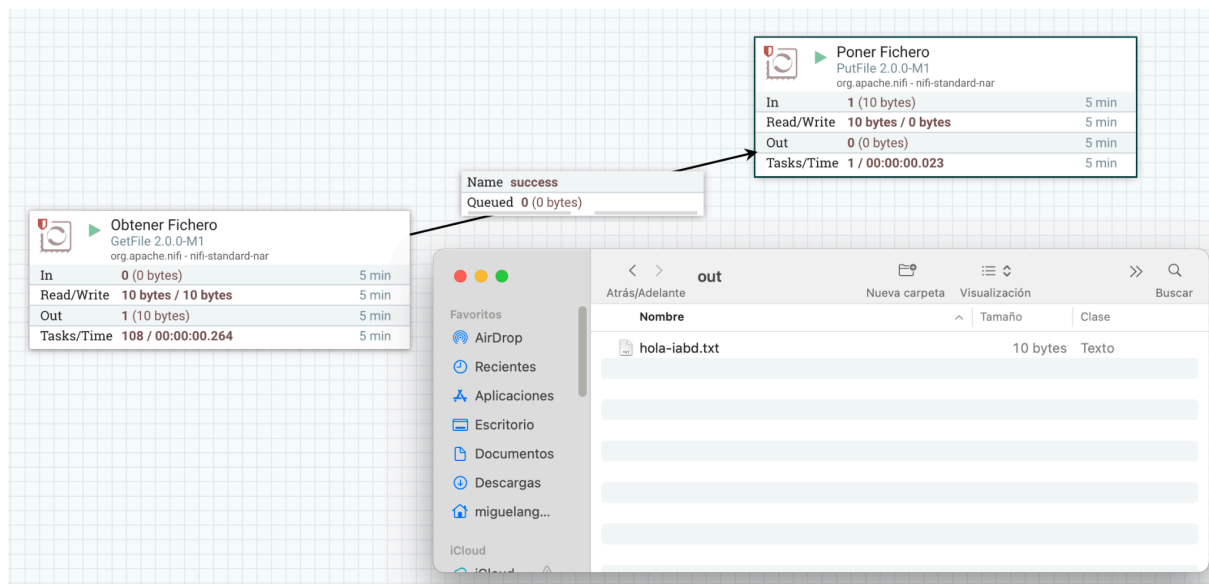
10. Antes de arrancar el primer procesador, creamos un pequeño fichero en el directorio que hemos puesto como entrada:

```
echo "Hola IABD" > hola-iabd.txt
```

11. Arrancamos el procesador mediante el botón derecho y la opción *Start*, y comprobamos que el fichero ya no está en la carpeta *in*, y que sí aparece en la cola (*Queued 1*). También podemos comprobar como tampoco está en la carpeta *out*.



12. Finalmente, arrancamos el procesador de *Poner Fichero*, y veremos como la cola se vacía y el archivo aparece en la carpeta *out*.



¡Ya hemos creado nuestro primer flujo de datos!