



Ejemplo 2 - Trabajando con atributos

Cada vez que se generan FF (FlowFile, representa un registro de datos que viaja por el flujo) estos van a tener asignados ciertos atributos por defecto. Entre estos atributos están el UUID o identificador único, su *timestamp* y el tamaño del fichero. Además, mediante el uso de procesadores podremos modificar estos o añadir nuevos atributos.

Generando contenido

Vamos a ver cómo hacerlo realizando los siguientes pasos:

1. Vamos a añadir un procesador del tipo [GenerateFlowFile](#) (este procesador crea FF con datos aleatorios o contenidos personalizados, lo cual es muy útil para testear y depurar nuestros flujos de datos).

En las opciones del procesador vamos a la pestaña de propiedades y completamos los campos:

- **Flow Size:** 10 bytes
- **Batch Size:** 1 para que nos genere un FF por cada ejecución
- **Data Format:** Text
- **Unique Flowfiles:** true e indicamos que los FF van a ser únicos.

Configure Processor | GenerateFlowFile 1.15.0

Invalid

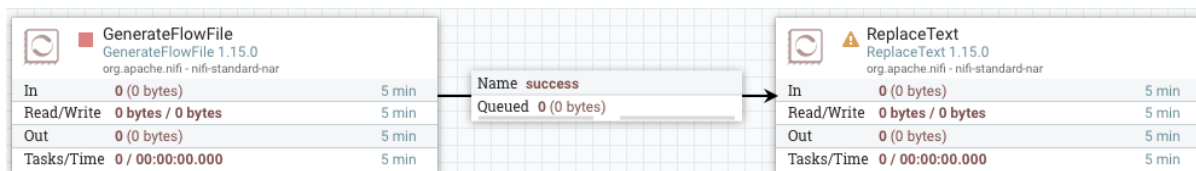
SETTINGS | SCHEDULING | **PROPERTIES** | COMMENTS

Required field

Property	Value
File Size	10B
Batch Size	1
Data Format	Text
Unique FlowFiles	true
Custom Text	No value set
Character Set	UTF-8
Mime Type	No value set

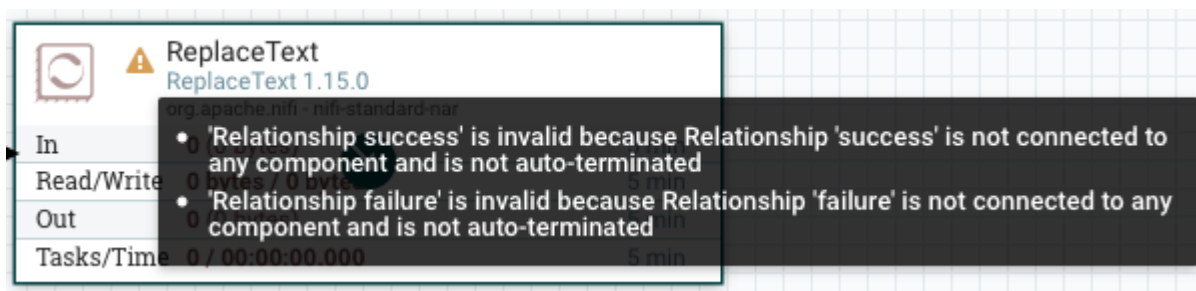
Configuración del procesador GenerateFlowFile

2. A continuación, en la configuración de planificación (*Scheduling*) de este procesador vamos a indicar que se ejecute cada 3 segundos (en el campo *Run Schedule* le ponemos como valor **3s**).
3. Una vez tengamos listo el generador, vamos a añadir el procesador **ReplaceText** con el que cambiaremos el texto. Tras ello, conectamos ambos procesadores.



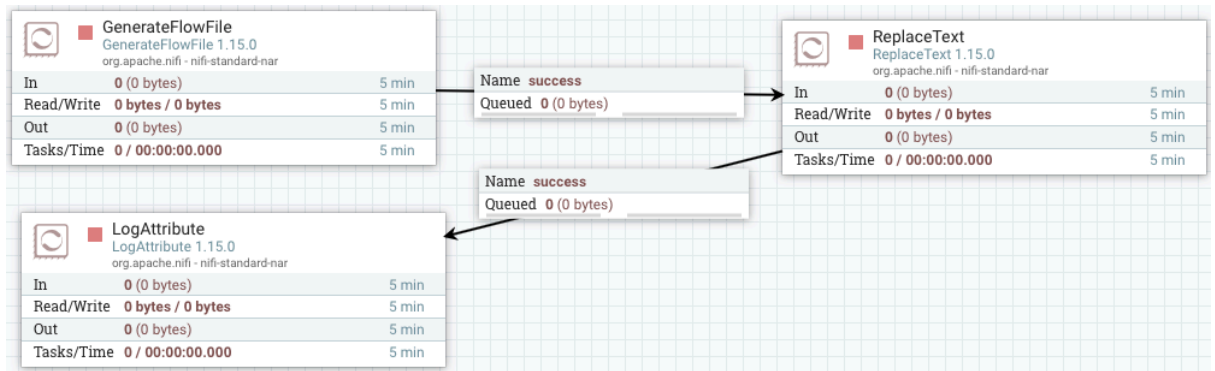
Conexión con ReplaceText

4. Si nos fijamos, a la izquierda del nombre del procesador, aparece un icono de aviso, el cual nos indica que necesitamos configurar el nuevo procesador, además de indicarnos que ambas relaciones no están conectadas o que faltan por autocompletar.



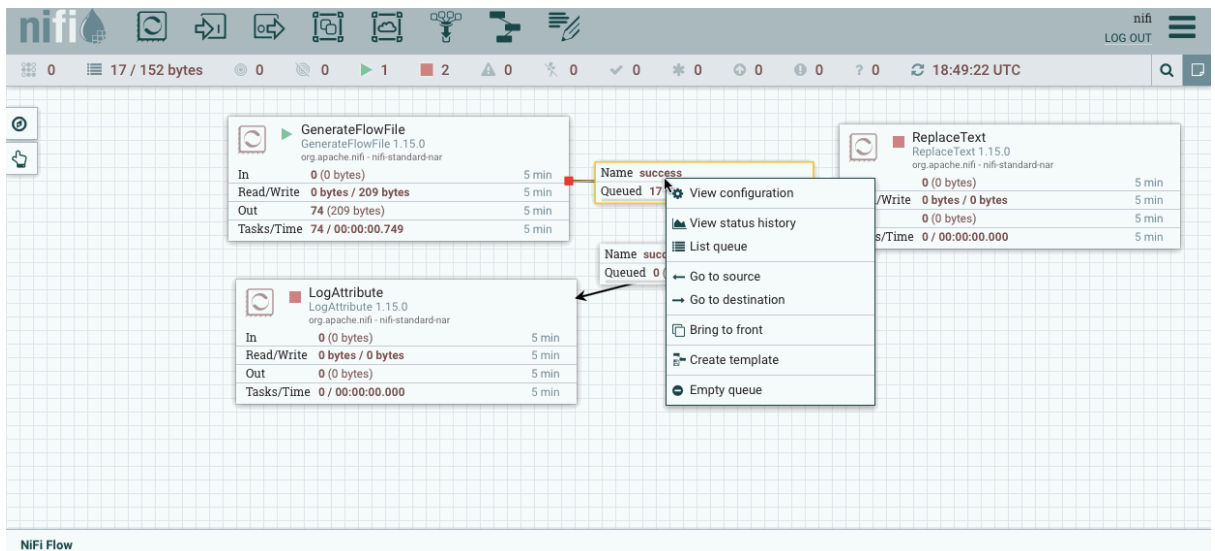
Avisos que aparecen en un Procesador

5. Para ello, configuramos la estrategia de reemplazo para que lo haga siempre (en el campo *Replacement Value* seleccionamos *Always Replace*), y al hacer esto el campo *Search Value* se invalida. Además, en el *Replacement Value* vamos a indicar simplemente **prueba**. Finalmente, marcamos para que autotermine la conexión **failure**.
6. Añadimos un procesador de tipo **LogAttribute** para mostrar en el log los atributos del FF, y conectamos el procesador anterior (*ReplaceText*) a éste mediante la relación **success**.



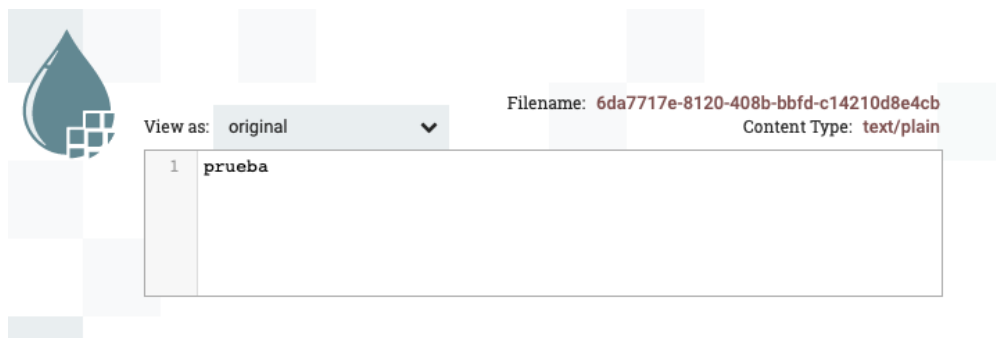
Log con los atributos

- Arrancamos el primer procesador y visualizamos la cola para comprobar qué ha generado. Para ello, sobre la cola elegimos la opción *list queue* para ver su contenido, y tras elegir uno, sobre el icono del ojo, visualizamos su contenido y comprobado que ha generado datos aleatorios:



Acceso y visualización de la cola

- Si ejecutamos el siguiente procesador, vemos que saca el FF de la cola anterior y aparecerá en la siguiente. Si comprobamos su valor, veremos que ha cambiado el valor original por **prueba**.

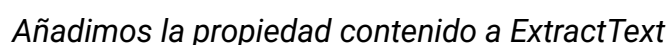


Resultado de visualizar la cola tras ReplaceText

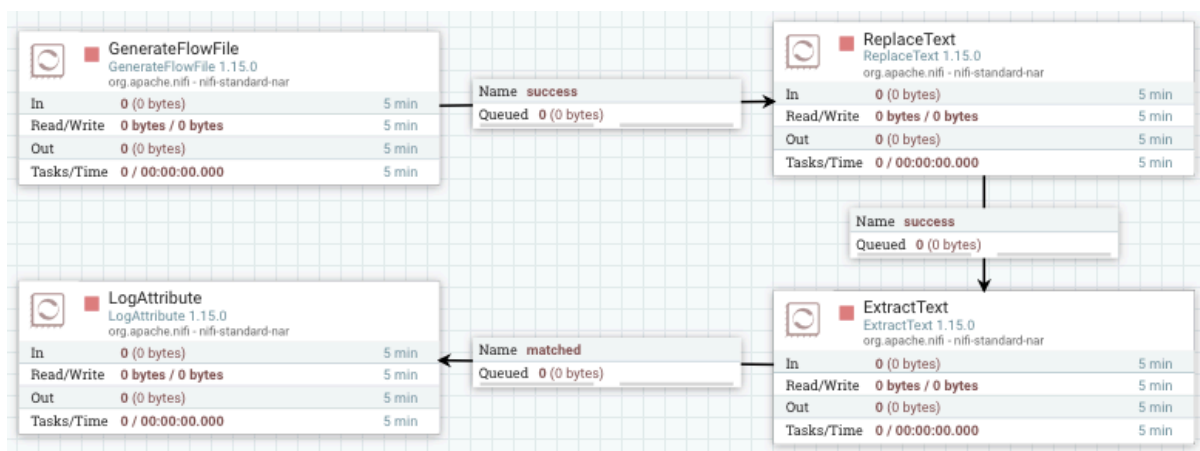
```
2023-11-17 19:12:37,446 INFO [Timer-Driven Process Thread-2]
o.a.n.processors.standard.LogAttribute LogAttribute
[id=2f300ff5-017d-1000-9020-2744e67e8d04]
logging for flow file StandardFlowFileRecord
[uuid=f4181825-f996-40c0-9e3c-a78326837d60,claim=StandardContentClaim
  [resourceClaim=StandardResourceClaim
    [id=1637174667410-1, container=default, section=1],
    offset=7953, length=7],
  offset=0,name=f4181825-f996-40c0-9e3c-a78326837d60,size=7]
```

Añadiendo un atributo

1. En las propiedades, creamos una nueva propiedad (botón + de la esquina superior derecha) que llamaremos `contenido`, y en cuyo valor vamos a poner la expresión `.*` que indica que queremos que coincida con todo.



- Una vez creado, vamos a colocar este procesador entre los dos anteriores (para el segundo con el caso **matched**, que es cuando ha coincidido con la expresión regular). En la conexión **unmatched** la marcamos para que autotermine, y comprobamos que no tenemos ninguna advertencia en ningún procesador.



Flujo completo del caso 2

- Finalmente, ejecutamos todos los procesadores y comprobamos como en el log aparece el nuevo atributo creado. También podemos acceder a la cola, y en la parte izquierda de cada flujo, en el icono de la **i**, pulsar y comprobar la pestaña **Attributes**.

matched
 Displaying 24 of 24 (159)

	Position	UUID
	1	a
	2	1
	3	7
	4	9
	5	0
	6	4
	7	e
	8	7
	9	7
	10	e
	11	2

FlowFile

DETAILS

ATTRIBUTES

Attribute Values

contenido.0
prueba
 filename
ad32e547-2d81-45cc-b474-87c9810df813
 path
 ./

uuid
ad32e547-2d81-45cc-b474-87c9810df813

Comprobación de los atributos de un FF

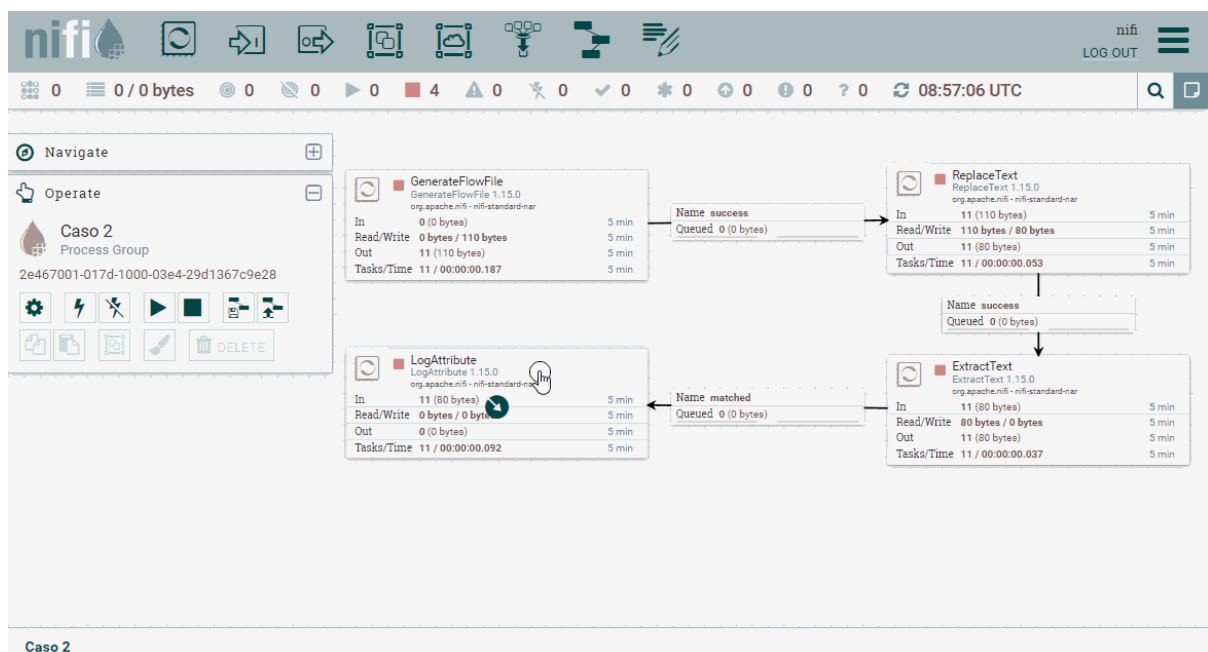
Linaje de los datos

Para comprobar el dato final, es muy útil utilizar la opción de *Data provenance*, la cual nos ofrece un linaje de los datos.

Linaje de los datos

El linaje de los datos describe el origen, movimientos, características y calidad de los datos, aportando visibilidad de punto a punto para incrementar la calidad de los datos. Se puede considerar como el historial de los datos, facilitando la documentación y gobernanza de los datos.

Para ello, sobre el procesador final, con el botón derecho, elegimos la opción *View data provenance*. Si elegimos uno de los flujos, a la derecha de cada flujo, podemos pulsar sobre el primer icono podremos ver un grafo y un *slider* que modifica el grafo respecto al instante temporal (en cada uno de los pasos, podemos hacer doble click y ver la información y el contenido del FF en dicho momento exacto):



Linaje de los datos