



Módulo Profesional: Big Data Aplicado

Ingestión de datos con NiFi

NiFi Inport Port y Out Port



Puerto de
entrada



Puerto de
salida

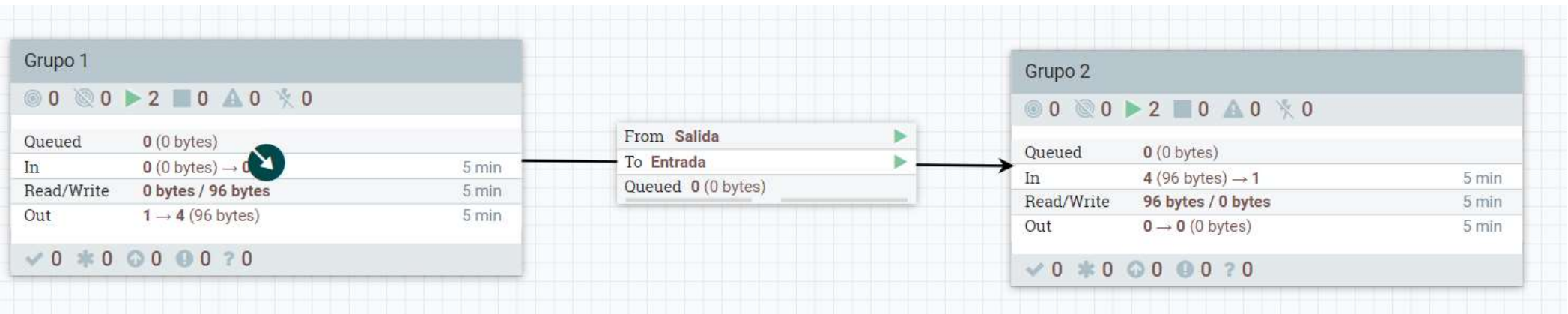


Grupo

Trabajando con grupos

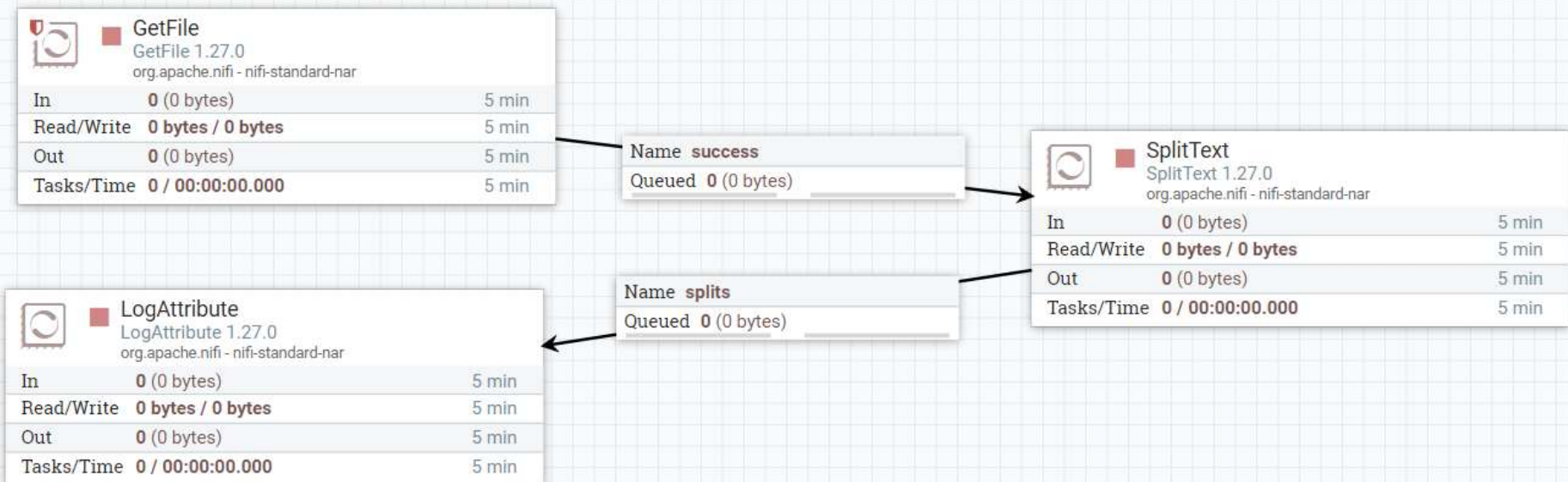
Dentro de los grupos, para indicar como entran los datos se utiliza un **Input port**, el cual va a definir un puerto de entrada al grupo. Del mismo modo, para indicar cómo salen los datos, se utiliza un **Output port** como puerto de salida para transferir la información fuera del grupo. Así pues, los grupos de procesos nos van a permitir refactorizar nuestros flujos para agruparlos y poder reutilizarlos en otros flujos, o bien mejorar la legibilidad de nuestro flujo de datos

NiFi Inport Port y Out Port



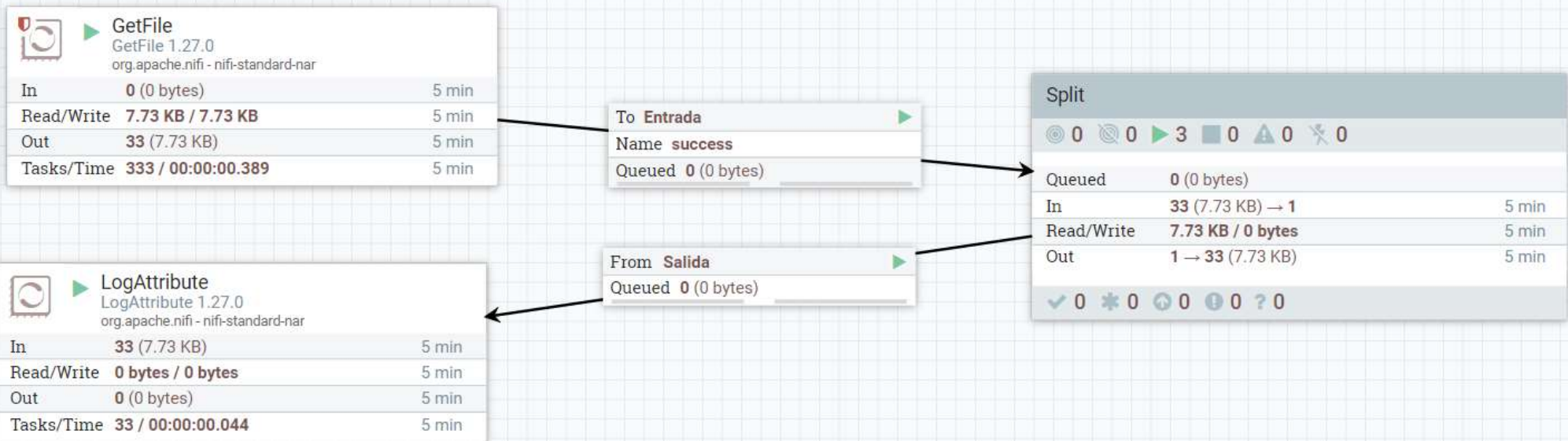
Ejercicio 5. Inport Port y Out Port

DIVIDIMOS ARCHIVOS EN FRAGMENTOS

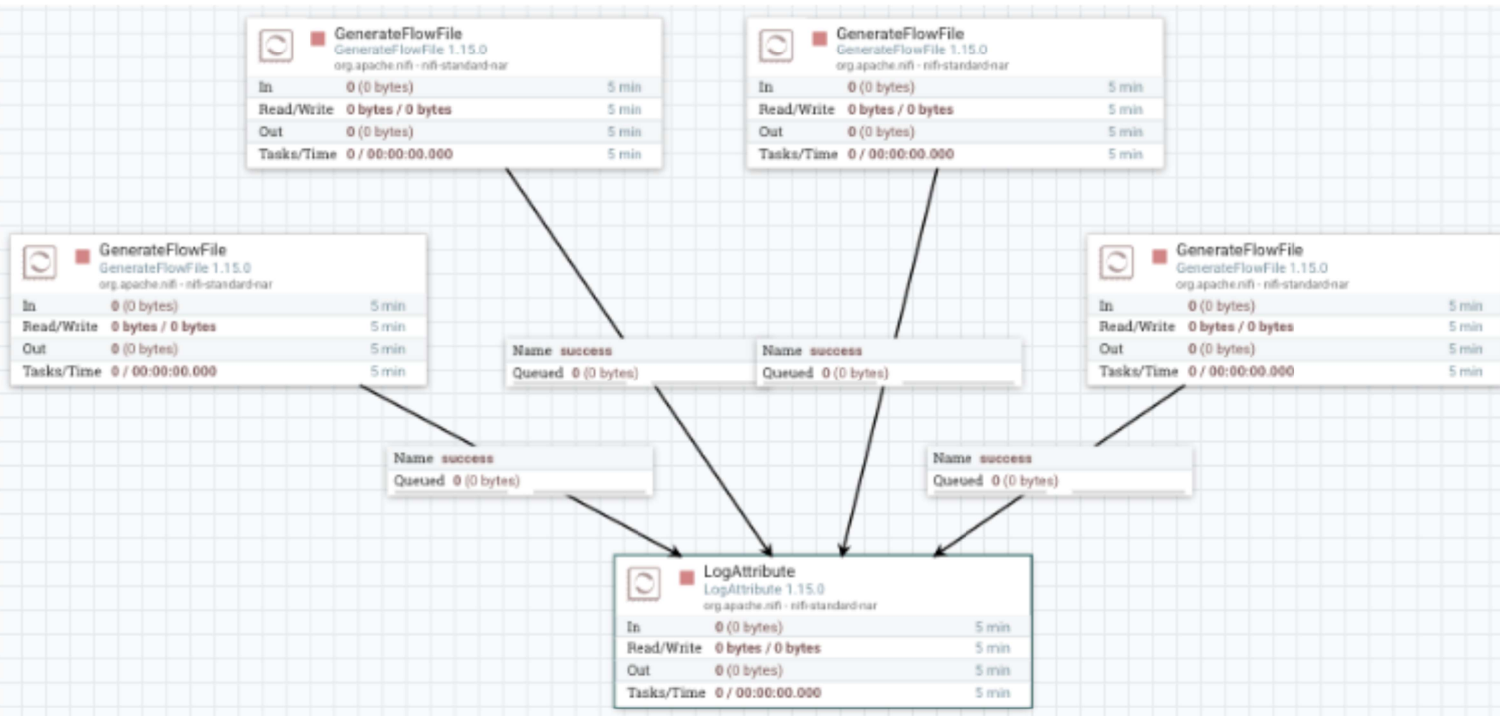


Ejercicio 5. Inport Port y Out Port

HEMOS ENCAPSULADO EL PROCESADOR SPLITTEXT DENTRO DEL GRUPO Y HEMOS HABILITADO LOS PUERTOS DE ENTRADA Y SALIDA PARA FACILITAR LA INTERACCIÓN CON EL FLUJO DE DATOS DENTRO DE ESTE GRUPO. ESTO MEJORA LA ORGANIZACIÓN Y LA MODULARIZACIÓN DE TU FLUJO DE DATOS EN NiFi.



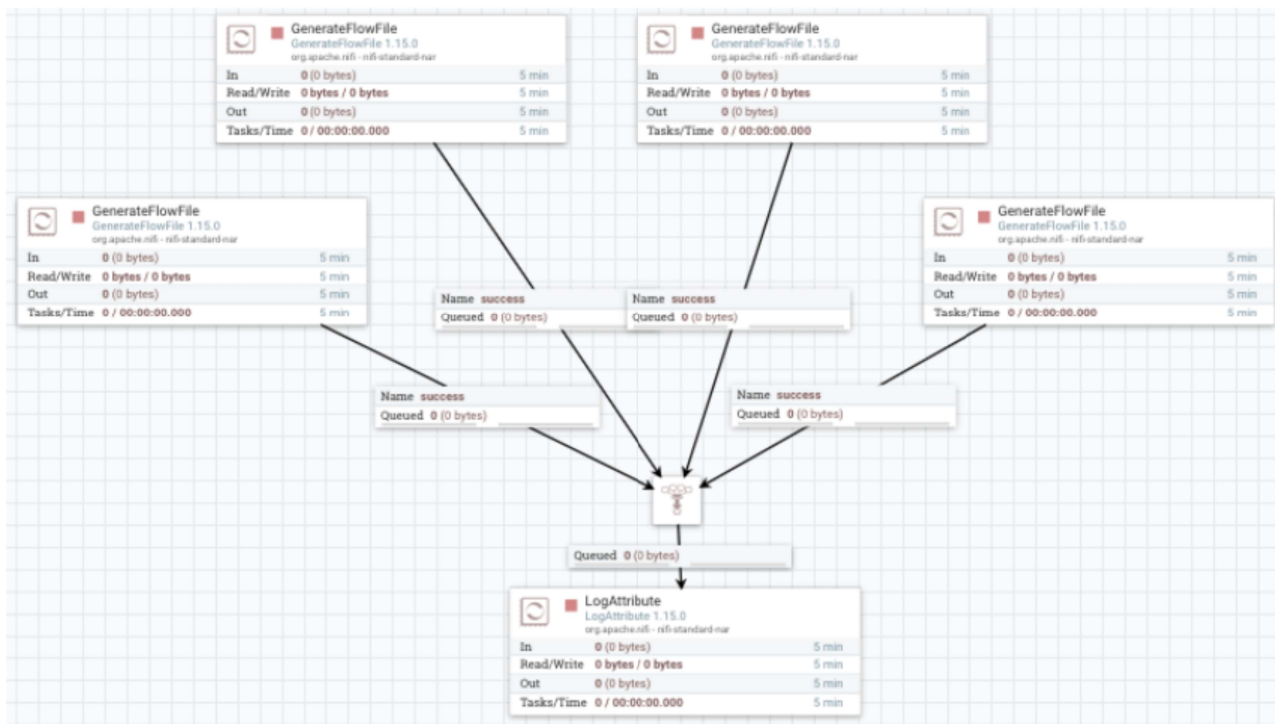
NiFi Funnel



Los funnels son un tipo de componente que permite trabajar en paralelo y después unir los diferentes flujos en un único flujo de ejecución, además de poder definir su propia prioridad de forma centralizada.

Varios procesadores que apuntan a uno

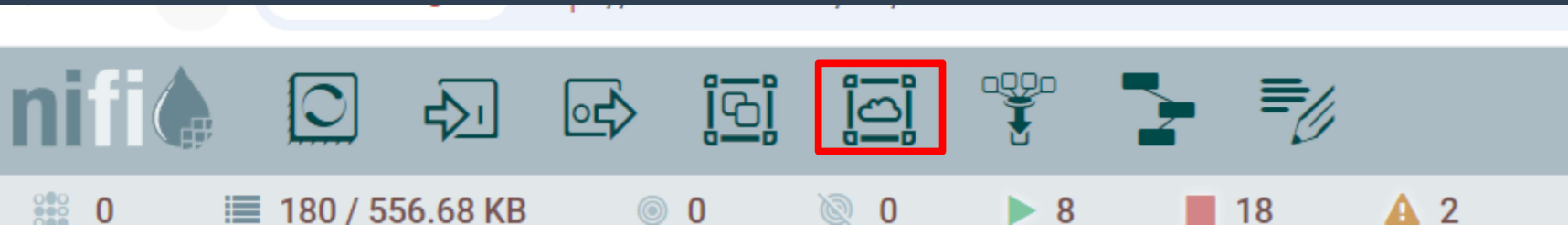
NiFi Funnel



El Funnel agrupa las conexiones

Si quisiéramos cambiar el procesador de LogAttribute por otro tipo de procesador, deberíamos borrar todas las conexiones y volver a conectarlo todo. Para evitar esto añadimos un Funnel que va a centralizar todas las conexiones en un único punto.

NiFi Add Remote Process Group



Se utiliza para integrar y enviar datos entre diferentes instancias de NiFi que están desplegadas en ubicaciones distintas. Este componente permite a un flujo de trabajo de NiFi comunicarse con otra instancia de NiFi que se ejecuta en un servidor remoto.

Funciones principales:

Interconexión entre nodos NiFi que pueden estar en diferentes redes, ubicaciones físicas o incluso diferentes centros de datos) se conecten entre sí para intercambiar datos.

Envío de datos: Puedes configurar este grupo para enviar datos desde el flujo de una instancia de NiFi hacia un "Input Port" (puerto de entrada) en la instancia remota.


Recepción de datos: También permite recibir datos de la instancia remota mediante un "Output Port" (puerto de salida).

Seguridad: Soporta comunicación segura entre instancias, utilizando HTTPS y otras configuraciones de seguridad como autenticación mediante certificados y encriptación de datos.


Optimización del flujo de trabajo: Esto es útil para crear flujos distribuidos, donde diferentes tareas o procesos se pueden ejecutar en diferentes servidores NiFi, repartiendo la carga de trabajo.


NiFi Add Remote Process Group


Add Remote Process Group


URLs 


https://remotehost:8443/nifi


Transport Protocol 


HTTP 


Local Network Interface 

HTTP Proxy Server Hostname 


HTTP Proxy Server Port 

HTTP Proxy User 

HTTP Proxy Password 

Communications Timeout 

30 sec

Yield Duration 

10 sec

CANCELADD

NiFi Add Remote Process Group

URLs ?

https://remotehost:8443/nifi

Transport Protocol ?

HTTP



Local Network Interface ?

URLs: Es la lista de URLs de destino donde se encuentra el otro NiFi. Se especifica el protocolo y la dirección (por ejemplo, http://otro-nifi:8080). Estas URLs son utilizadas por el RPG para conectarse con la instancia remota de NiFi.

Transport Protocol: Define el protocolo de transporte que se va a utilizar para la comunicación con la instancia remota. Puede ser HTTP o RAW. El protocolo RAW es más eficiente para grandes volúmenes de datos, mientras que HTTP es más fácil de integrar con firewalls y proxies.

Local Network Interface: Especifica la interfaz de red local (como una dirección IP) desde la que NiFi debe comunicarse. Se utiliza para casos donde el servidor tiene múltiples interfaces y se quiere forzar una en particular para la conexión.

NiFi Add Remote Process Group

HTTP Proxy Server Hostname ?

HTTP Proxy Server Port ?

HTTP Proxy User ?

HTTP Proxy Password ?

HTTP Proxy Server Hostname: Si se usa un proxy HTTP para conectarse a la instancia remota de NiFi, aquí se especifica el nombre de host o dirección IP del servidor proxy.

HTTP Proxy Server Port: El puerto del servidor proxy HTTP que se utiliza para la conexión. Por ejemplo, 8080 o 3128, dependiendo de la configuración del proxy.

HTTP Proxy User: El nombre de usuario necesario para autenticarse en el proxy HTTP (si se requiere autenticación).

HTTP Proxy Password: La contraseña correspondiente al usuario del proxy HTTP. Se usa junto con el nombre de usuario para la autenticación.

NiFi Add Remote Process Group

Communications Timeout ?

30 sec

Yield Duration ?

10 sec

Communication Timeout: Define el tiempo máximo que NiFi esperará para una respuesta de la instancia remota antes de dar por fallida la conexión. Se especifica en un formato como 30 sec o 1 min.

Yield Duration: Es el tiempo que NiFi esperará antes de volver a intentar una conexión fallida con la instancia remota. Sirve para evitar sobrecargar el sistema con intentos continuos. Se configura en un formato similar, por ejemplo, 10 sec.

