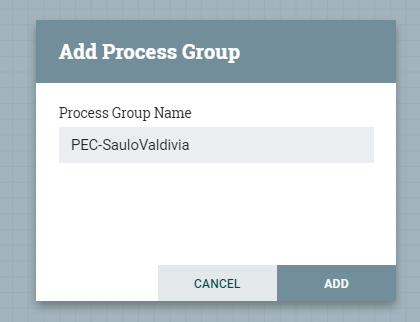
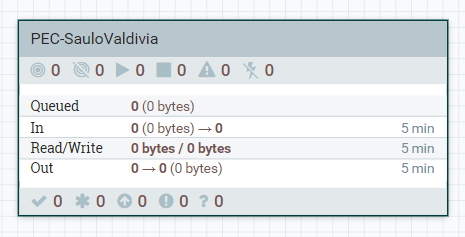
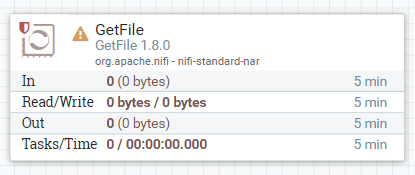
1. **Ejercicio NiFi**
   1. Creación de un grupo (Process Group) dentro de NiFi con el nombre PEC-<NombreAlumno>

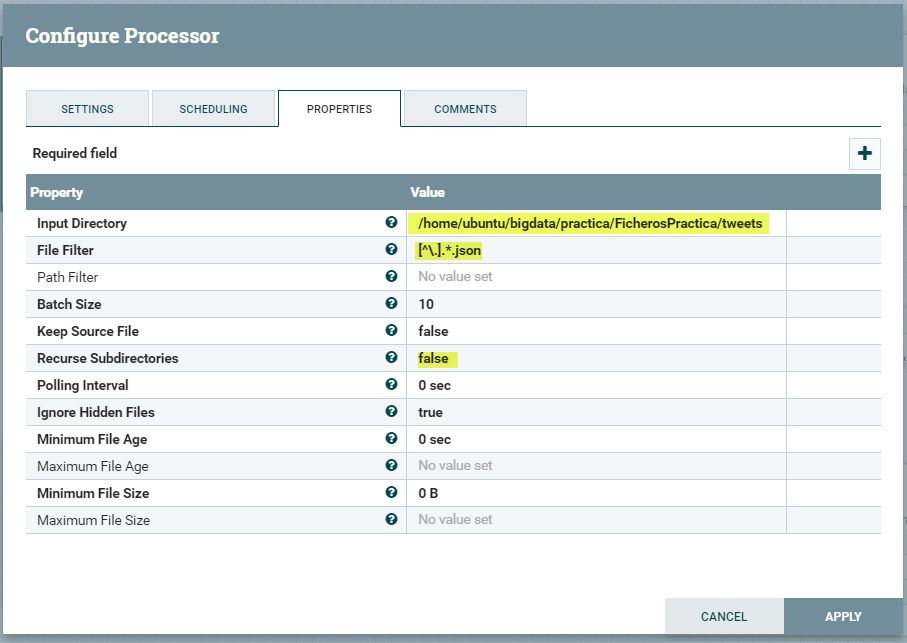




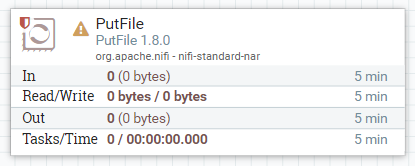
* 1. Crear un flujo en NiFi que coja solo los ficheros de la carpeta tweets (del zip  
     proporcionado) y no de sus subcarpetas y los deje en el directorio tweets/ejercicio
     1. Creación de un proceso GetFile para recuperar los tweets



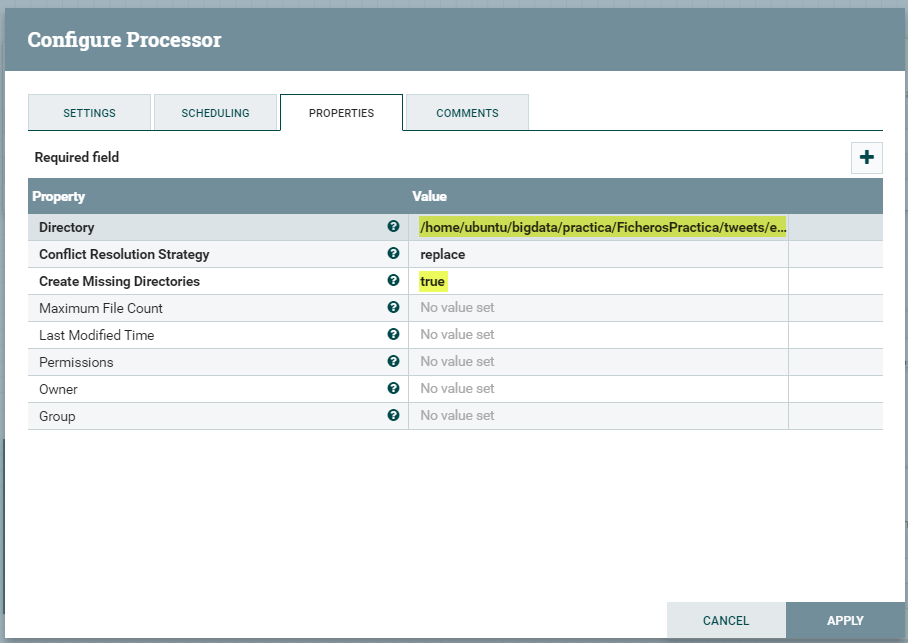
* + 1. Configuración de proceso GetFile



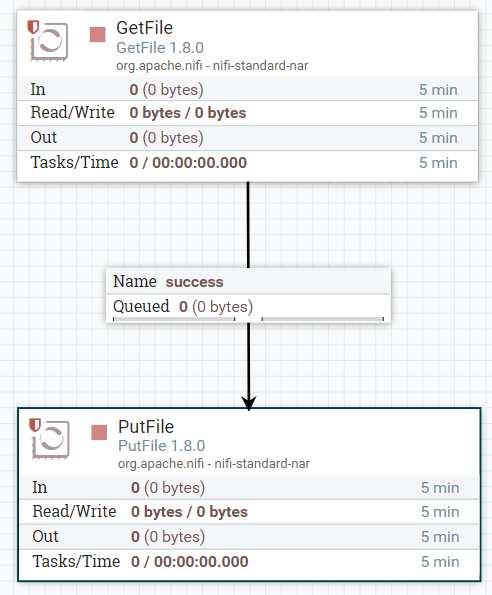
* + 1. Creación de proceso PutFile



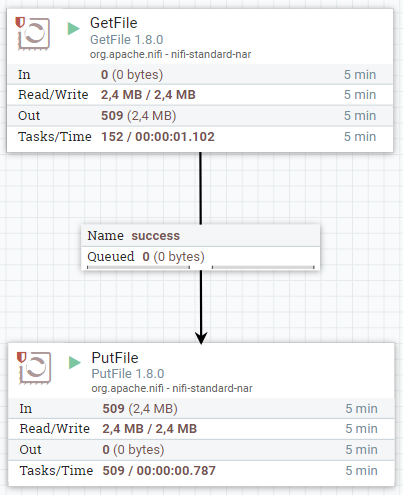
* + 1. Configuración de proceso PutFile



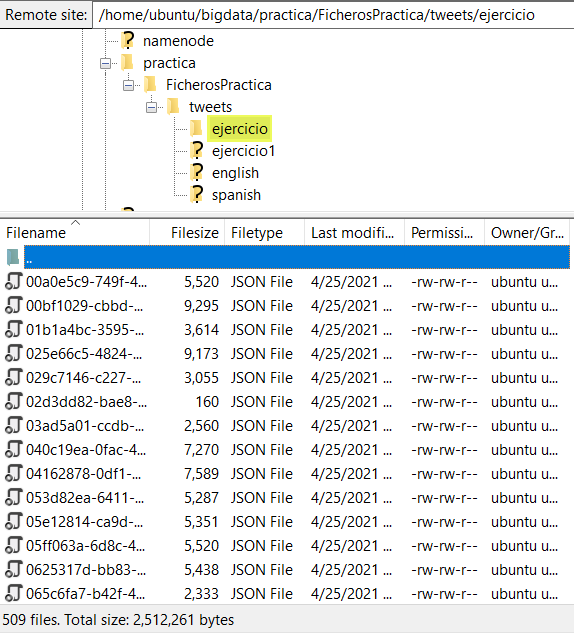
* + 1. Flujo completo



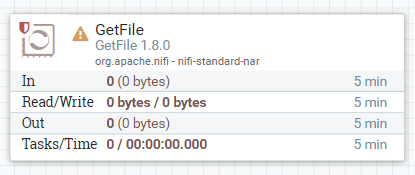
* + 1. Resultados



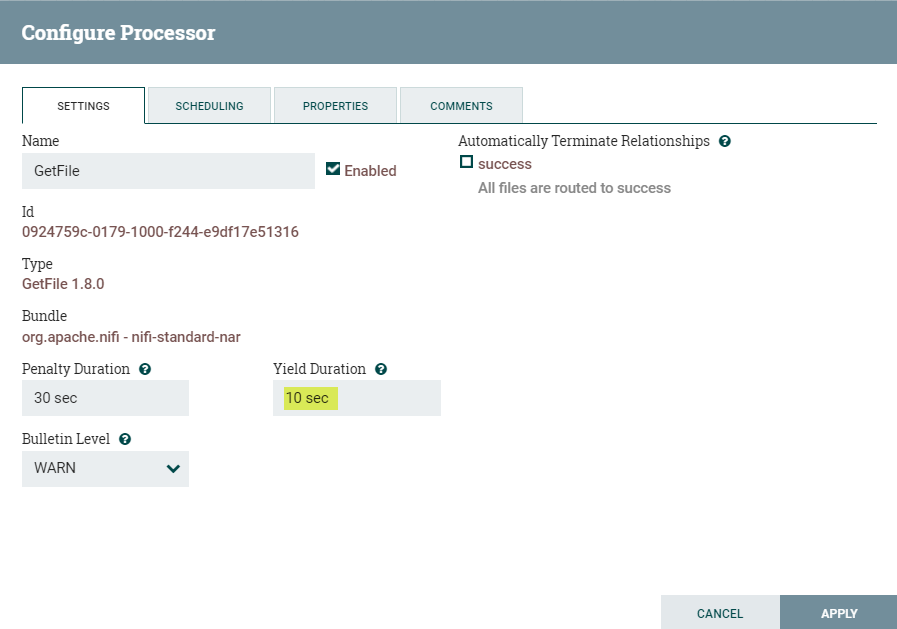
* + 1. Resultados en Filezilla

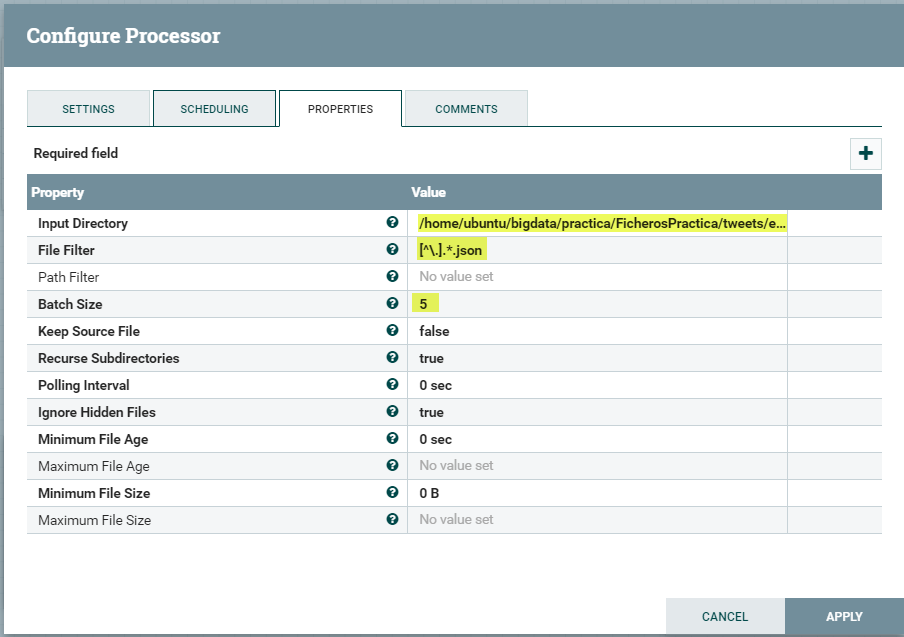


* 1. Crear un nuevo flujo que coja los tweets de 5 en 5 de la carpeta tweets/english  
     cada 10 segundos y los deje de nuevo en el directorio tweets/ejercicio
     1. Creación de proceso GetFile

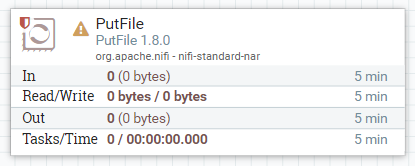


* + 1. Configuración de proceso GetFile

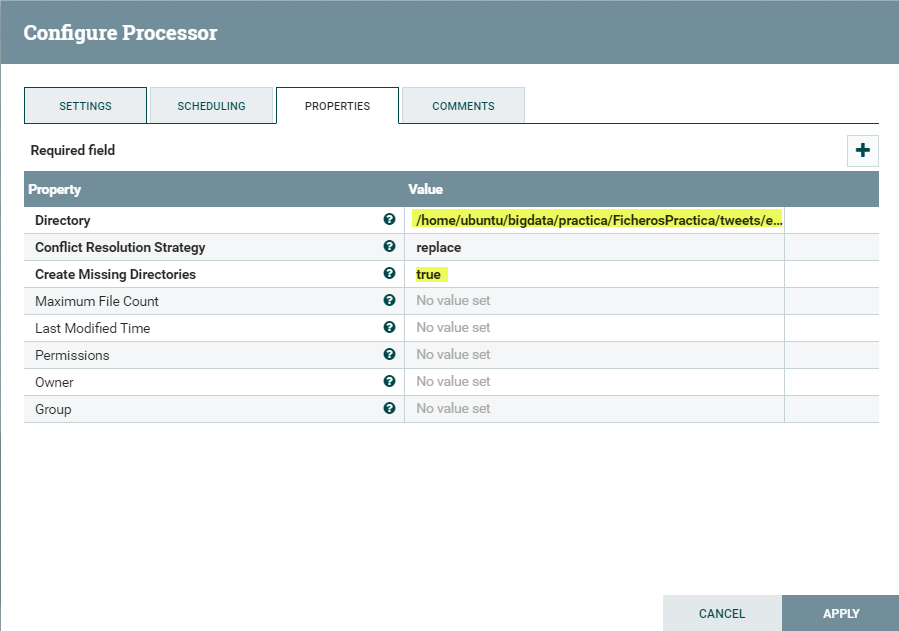




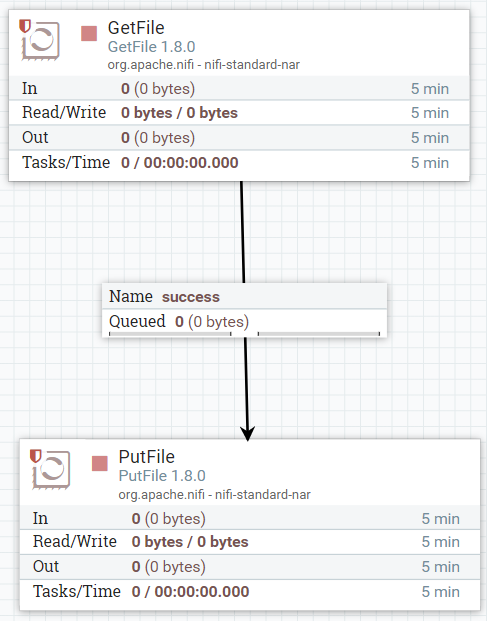
* + 1. Creación de proceso PutFile



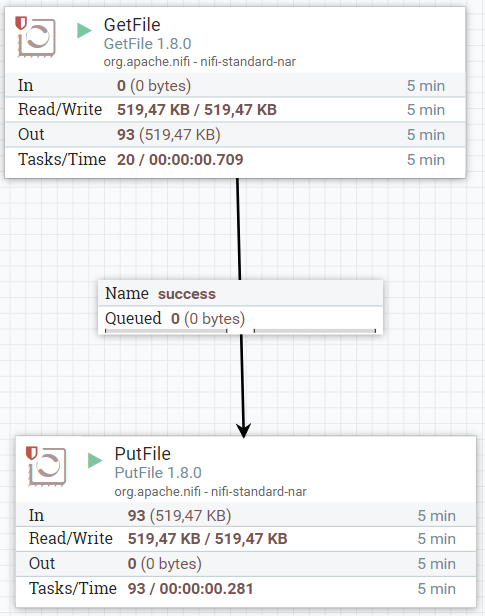
* + 1. Configuración de proceso PutFile



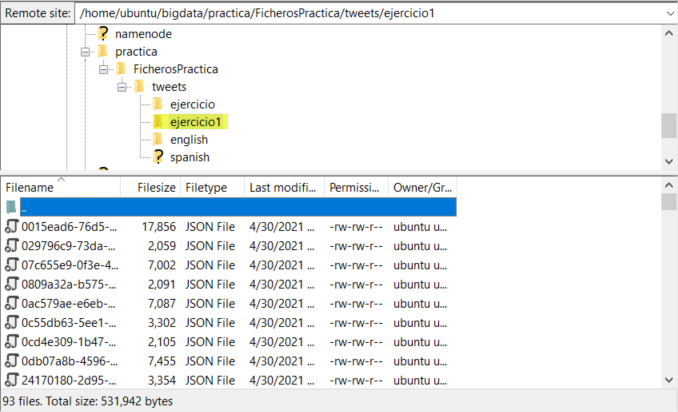
* + 1. Flujo Completo



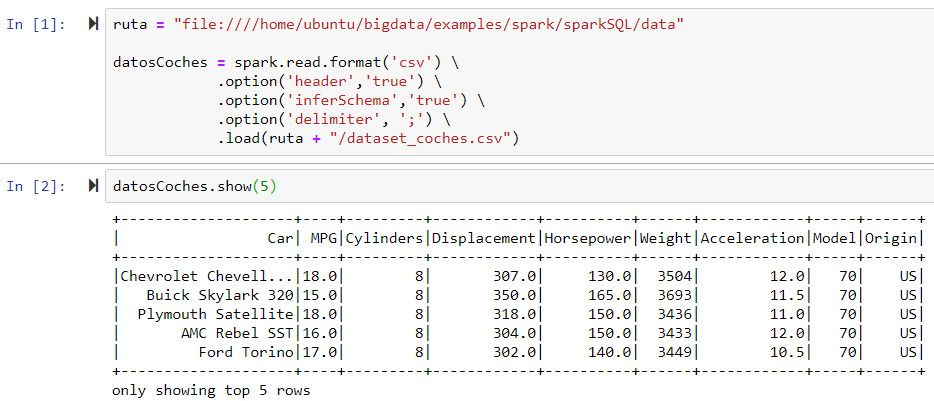
* + 1. Resultados



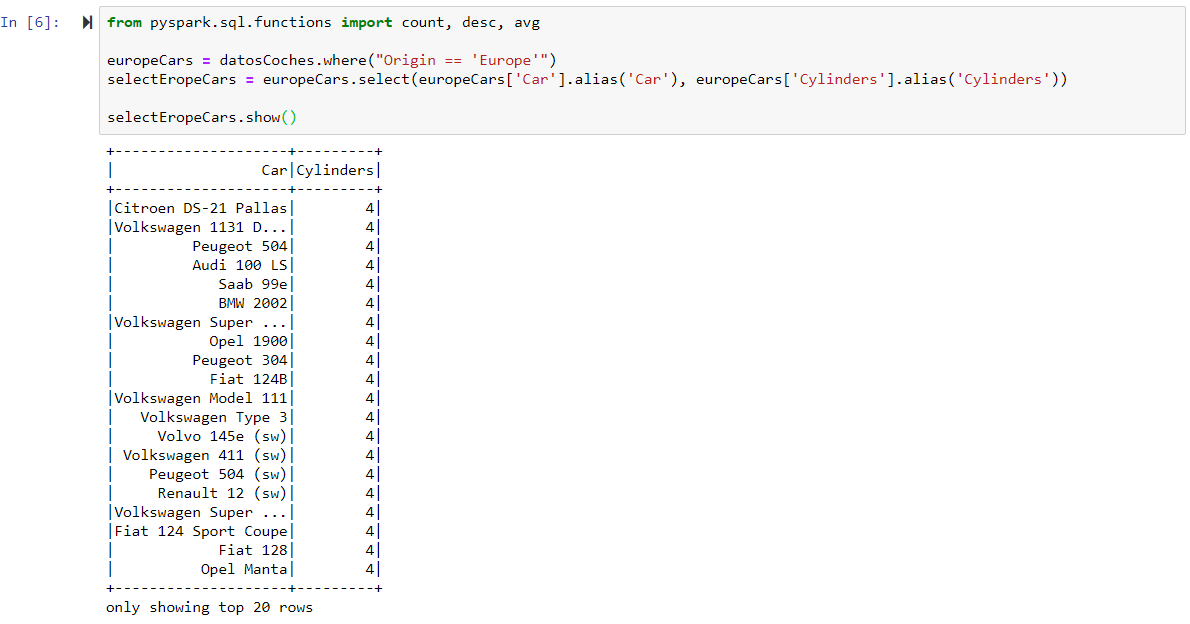
* + 1. Resultados Filezilla



1. **Ejercicio SparkSQL**
   1. Cargar los datos del fichero dataset\_coches.csv en un dataframe y mostrar las  
      primeras 5 líneas del dataframe.



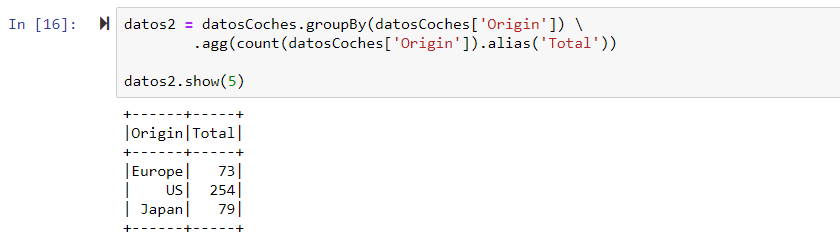
* 1. Usando tanto SQL como el API de dataframes de Spark obtener las columnas “Car” y “Cylinders” de todos los que sean de Europa
     1. SparkSQL Dataframe API



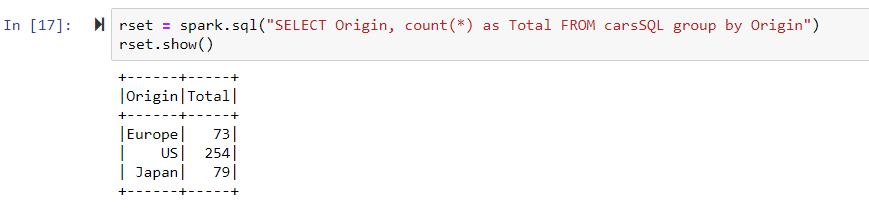
* + 1. SQL



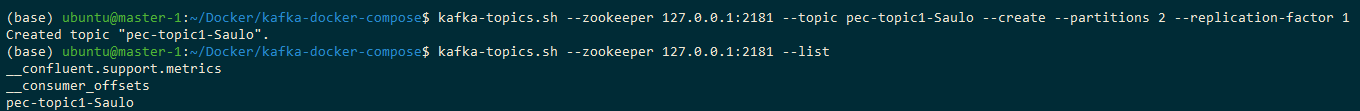
* 1. Usando tanto SQL como el API de dataframes de Spark obtener el número de vehículos por “Origin”
     1. SparkSQL Dataframe API

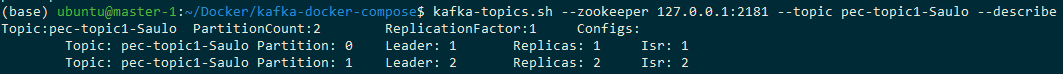


* + 1. SQL



1. **Ejercicio Kafka**
   1. Creación de topics
      1. Crear un nuevo topic denominado pec-topic1-<NombreAlumno> con dos  
         particiones y con factor de replicación 1

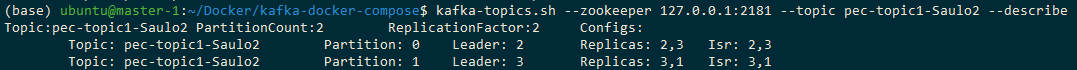




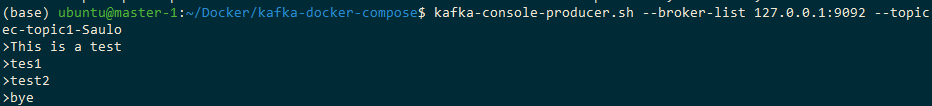
* + 1. ¿Sería posible crearlo en el entorno que tienes con un factor de replicación  
       de 2? ¿Por qué?

Nuestra arquitectura contiene 3 Brokers/Nodos y podria soportar un factor de replicacion de 2.

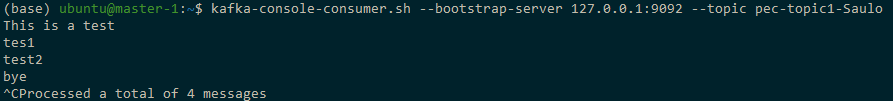




* 1. Crea un productor que empiece a pasar información desde la línea de comandos  
     al topic creado anteriormente y simular el envío de información



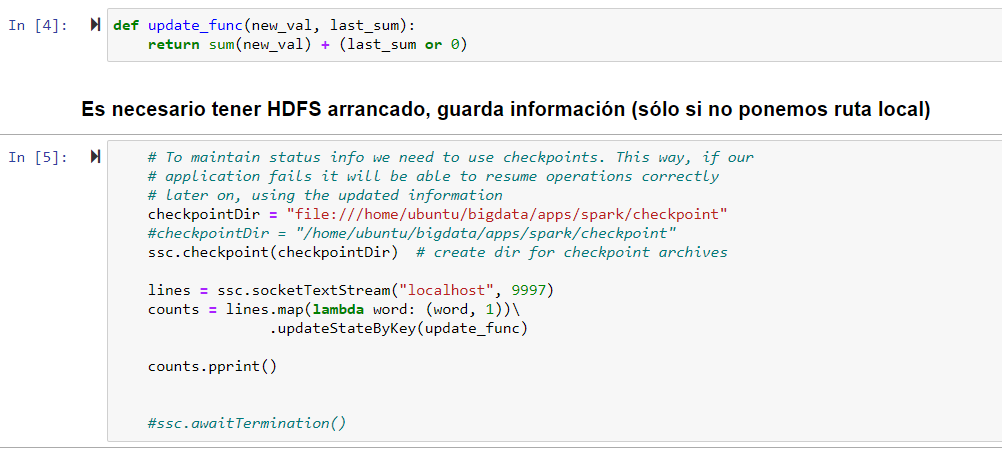
* 1. Creación de consumidores
     1. Crear un consumidor que sea capaz de leer *toda* la información que  
        contenga el topic creado



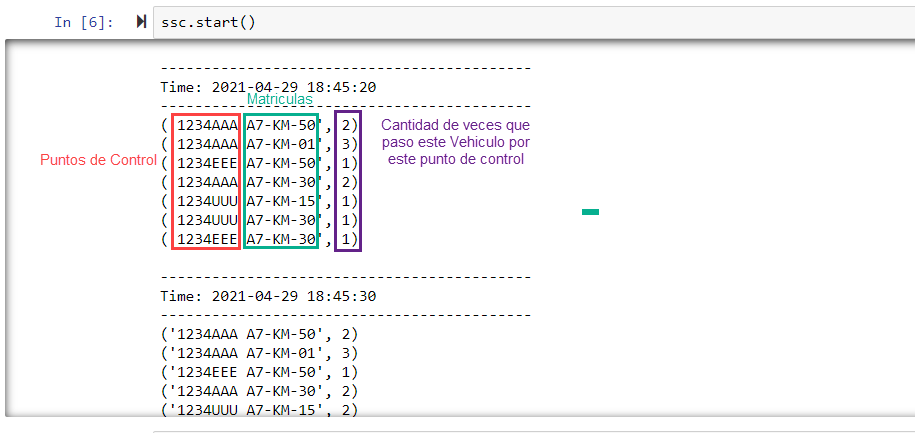
1. **Ejercicio Spark Streaming**
   1. Saber en tiempo real cuantas veces pasa un vehículo por cada punto del sistema desde el día que se pone en marcha el mecanismo.
      1. Iniciar el Producer



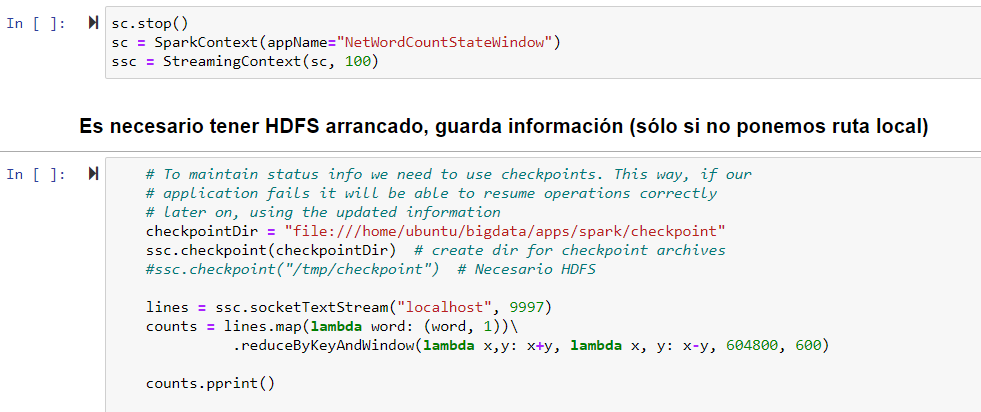
* + 1. Script para conectarse al stream y procesar las lineas



* + 1. Resultados



* 1. Saber en tiempo real cuántas veces pasa un vehículo por un determinado punto, pero solo teniendo en cuenta los últimos 7 días de información.
     1. Script para conectarse y procesar las líneas.



* + 1. Resultados

Los resultados son similares al apartado anterior. Sin embargo, esperaríamos que transcurrido el tiempo de la ventana temporal, estos datos se actualicen.

