

1. Consideraciones

- La puntuación total de la práctica será de 10 puntos repartidos tal y como se va indicando en cada ejercicio
- La entrega de la práctica debe incluir el código para resolver el ejercicio, así como los pantallazos solicitados donde se soliciten. IMPORTANTE: las imágenes tienen que ser claras y con un tamaño suficiente para leer claramente el comando o el resultado.
- Cualquier duda que tengáis podéis preguntarme a través del correo joseluis.llorenteperales@campus.eae.es

2. Ficheros proporcionados

- Para la realización de la práctica serán necesarios los ficheros que se encuentran en los documentos “*FicherosPractica.zip*”
- Los ficheros y carpetas a los que se hace referencia están en el zip *ficherosPractica.zip*
- Tanto la carpeta como el fichero que se encuentran en el zip se deberán pasar a la máquina virtual

3. Ejercicios

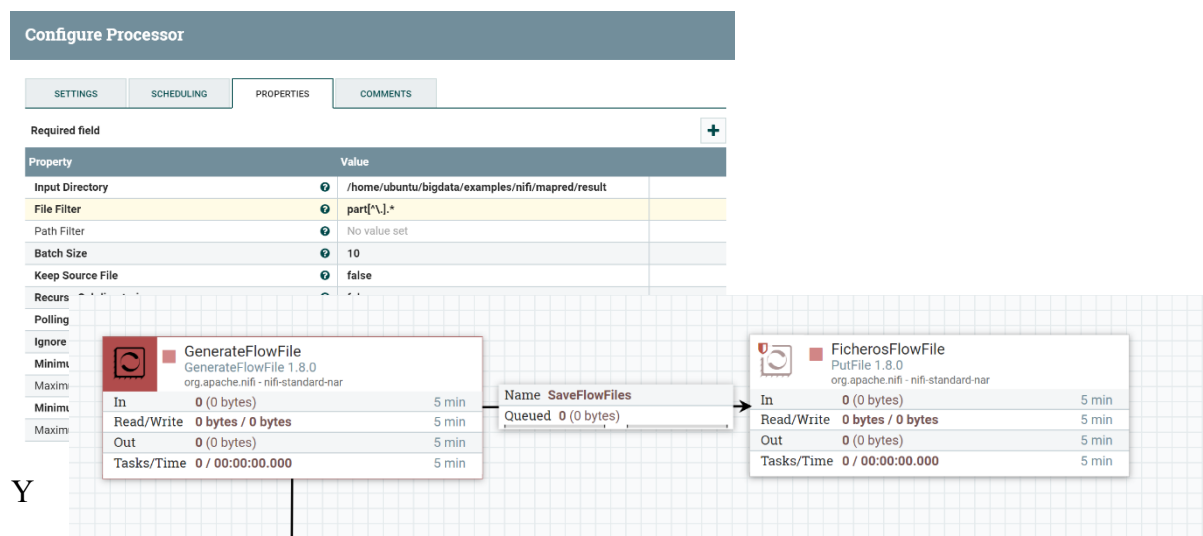
1. Ejercicio NiFi (2,5 puntos)

En este punto se trata de evaluar los conocimientos adquiridos sobre esta herramienta Apache NiFi y se deberán adjuntar pantallazos de cada punto

- Creación de un grupo (Process Group) dentro de NiFi con el nombre PEC-
<NombreAlumno> **(0,5 puntos)**
- 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 punto)**
- 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)**

En ninguno de los casos anteriores debe quedar información en la carpeta de origen

- Ejemplos de pantallazos:



2. Ejercicio SparkSQL (2,5 puntos)

- a) Cargar los datos del fichero dataset_coches.csv en un dataframe y mostrar las primeras 5 líneas del dataframe **(0,5 puntos)**
- b) Usando tanto SQL como el API de dataframes de Spark obtener las columnas “Car” y “Cylinders” de todos los que sean de Europa **(1 punto)**
- c) Usando tanto SQL como el API de dataframes de Spark obtener el número de vehículos por “Origin” **(1 punto)**

3. Ejercicio Kafka (2,5 puntos)

En este punto se trata de evaluar los conocimientos adquiridos sobre la herramienta Apache Kafka

- a) Creación de topics
 - Crear un nuevo topic denominado pec-topic1-<NombreAlumno> con dos particiones y con factor de replicación 1 **(0,5 puntos)**
 - ¿Sería posible crearlo en el entorno que tienes con un factor de replicación de 2? ¿Por qué? **(0,5 puntos)**
- b) Crea un productor que empiece a pasar información desde la línea de comandos al topic creado anteriormente y simular el envío de información **(0,5 puntos)**
- c) Creación de consumidores
 - Crear un consumidor que sea capaz de leer *toda* la información que contenga el topic creado **(1 puntos)**