### Ejercicio Spark Streaming (2,5 puntos)

**Instrucciones**:

Para poder realizar el siguiente ejercicio se necesita lo siguiente:

Dejar el fichero (script de python) vehiculos\_producer.py en el directorio /home/ubuntu/bigdata/examples/spark/sparkStreaming de la máquina virtual

Para realizar la práctica se deberá ejecutar el script desde el mismo directorio donde se encuentra el fichero a través de la siguiente instrucción:

python vehiculos\_producer.py --port 9997 --interval 5

Esto hará que se emita información al puerto 9997 cada 5 segundos.

Podeis comprobar que información se genera abriendo otra consola y escribiendo

nc -d localhost 9997

**Práctica:**

Una empresa que controla el movimiento de vehículos ha implementado un sistema con varios puntos de control en una ciudad. Cada vez que un vehículo pasa por uno de los puntos se recoge su matrícula y, en tiempo real, se envía la información de la matricula junto al punto por donde pasó ese vehículo.

Nos solicita que hagamos el siguiente control en tiempo real con Spark:

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 puntos)**
2. Saber en tiempo real cuántas veces pasa un vehículo por un determinado punto, pero solo teniendo en cuenta los últimos 7 días de información. **(1,5 puntos)**

¿Cómo desarrollarías ambos mecanismos con Spark Streaming? (utilizar Jupyter Notebook para implementarlo recogiendo la información que se genera desde el script comentado anteriormente)