Sección #6 Diseño de modelo de captación de datos

**1. Seleccionad un tema de vuestro interés y que nos permita extraer información a través del tiempo. Debería ser un sistema de indicadores que pudieran aportarnos información valiosa sobre la temática escogida (por ejemplo: sensores de tráfico, indicadores atmosféricos, índices bursátiles…)**

***Tráfico vehicular en una ciudad***:

* Nos permite recopilar datos como la cantidad de vehículos, velocidad promedio, tiempos de congestión, entre otros.
* Permite monitorear patrones de tráfico, identificar horas pico y ayudar a la planificación urbana.

**2. De este sistema extraeremos unos datos. Pensad como mínimo en 8 atributos distintos que podríamos guardar, teniendo en cuenta que tiene que contener todos los siguientes tipos de datos: Con base en el tema elegido, diseñen un conjunto de atributos que representen información relevante y que cumplan con los tipos de datos solicitados. Deben incluirse todos los tipos de datos indicados: entero, decimal, texto, fecha o tiempo, booleano y mapa (map).**

1. ***Cantidad de vehículos*** - Entero: Número total de vehículos detectados.
2. ***Velocidad promedio*** - Decimal: Velocidad media de los vehículos (en km/h).
3. ***Duración de congestión*** - Entero: Tiempo en minutos que duró la congestión.
4. ***Fecha y hora de registro*** - Fecha o tiempo: Momento en que se capturaron los datos.
5. ***Ubicación*** - Texto: Nombre de la calle o zona donde se encuentra el sensor.
6. ***Intensidad de tráfico*** - Entero: Nivel de tráfico en una escala numérica.
7. ***Día laboral*** - Booleano: Indica si el día es laboral (1) o no (0).
8. ***Coordenadas de ubicación*** - Mapa: Coordenadas en latitud y longitud.

**3. Cread un documento .csv con datos ficticios de vuestra estructura escogida. Insertad 100 registros como mínimo. (mostrad un pantallazo del documento) csv**

**4. Escribid la sentencia de creación de una tabla externa en Hive para vuestros datos, teniendo en cuenta dichos atributos. Vamos a suponer que los datos estarán el bucket de S3 siguiente s3://retoBD/input/**

**CREATE EXTERNAL TABLE** trafico\_vehicular ( cantidad\_vehiculos **INT**, velocidad\_promedio **DOUBLE**, duracion\_congestion **INT**, fecha\_hora **STRING**, ubicacion **STRING**, intensidad\_trafico **INT**, dia\_laboral **BOOLEAN**, latitud **DOUBLE**, longitud **DOUBLE**)

**ROW** FORMAT DELIMITED

FIELDS TERMINATED **BY** ','

STORED **AS** TEXTFILE

**LOCATION** 's3://retoBD/input/';