

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

```
Cantidad de vehiculos,Velocidad promedio,Duracion de congestion,Fecha y hora,Ubicacion,Intensidad de trafico,Dia laboral,Latitud,Longitud
291,69.3,15,2024-11-06 00:00:00,Calle A,3,0,-33.3445,178.8583
390,41.7,52,2024-11-06 00:15:00,Calle A,3,0,-47.7508,160.559
327,31.1,22,2024-11-06 00:30:00,Calle C,1,0,-75.1774,-88.8736
282,78.4,4,2024-11-06 00:45:00,Intersección B,1,1,58.1945,-148.2057
69,63.6,36,2024-11-06 01:00:00,Intersección B,2,0,-41.6781,-170.3558
147,64.8,44,2024-11-06 01:15:00,Intersección B,3,1,71.5599,120.9535
303,86.5,42,2024-11-06 01:30:00,Calle A,2,0,-72.1941,-125.0468
405,34.5,22,2024-11-06 01:45:00,Plaza E,1,0,-4.7007,-73.8957
325,36.4,21,2024-11-06 02:00:00,Calle A,2,0,50.1024,-56.4052
57,99.4,33,2024-11-06 02:15:00,Plaza E,2,1,50.7829,-162.6668
317,92.8,25,2024-11-06 02:30:00,Avenida D,3,1,27.7839,28.0845
```

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/';
```