

### Caso de Estudio 3: Ciudad Inteligente (Smart City)

**Caso de Uso:** Gestión del tráfico urbano y seguridad pública.

**Descripción Ampliada:** Las ciudades modernas están usando Big Data para mejorar la vida de sus ciudadanos. Un ejemplo es la gestión del tráfico y la seguridad:

- ★ **Gestión del Tráfico:** Sensores en las calles y cámaras de tráfico recopilan datos de flujo vehicular en tiempo real. Esta información se procesa para ajustar los semáforos de forma dinámica y re-enrutar el tráfico en caso de congestión o accidentes.
- ★ **Seguridad Pública:** Las cámaras y otros sensores recopilan datos para analizar patrones de movimiento y comportamiento. Utilizando análisis predictivos, las fuerzas de seguridad pueden identificar áreas de alto riesgo para desplegar recursos y prevenir delitos antes de que ocurran. Los datos de llamadas de emergencia, redes sociales y reportes ciudadanos también se integran en este análisis

### Trabajo

#### 1. Las 5 V's del Big Data:

- ★ **Volumen:** ¿Qué tipo de datos masivos se generan o manejan? ¿De qué escala hablamos (terabytes, petabytes)?
  - Muchísimos datos
  - Petabytes
- ★ **Velocidad:** ¿Los datos se procesan en tiempo real o por lotes? ¿Por qué esa velocidad es crucial para el éxito de la empresa?
  - en tiempo real, como los semáforos → prevenir el tráfico y accidentes
  - por lotes, para identificar áreas de alto riesgo en la seguridad pública
- ★ **Variedad:** ¿Qué tipos de datos se utilizan (estructurados, no estructurados, semi-estructurados)?
  - estructurados en cuanto a datos de sensores
  - no estructurados para las imágenes y videos de las cámaras
- ★ **Veracidad:** ¿Qué desafíos de calidad y confiabilidad de datos podrían enfrentar?
  - camaras mal iluminadas, retrasos de transmisión
- ★ **Valor:** ¿Cuál es el beneficio de negocio (ganancias, eficiencia, satisfacción del cliente) que se obtiene del Big Data en este caso?

- El beneficio que se obtiene de usar big data en una ciudad es que se puede usar para mejorar la calidad de vida de los ciudadanos, ayudando a prevenir delitos, tránsito, accidentes, etc.

## 2. Almacenamiento:

- ★ ¿Dónde se almacenarán estos datos? ¿Creen que sería un sistema de archivos distribuido como HDFS, un Data Lake o una base de datos más tradicional? ¿Qué desafíos de escalabilidad y costo enfrentarían al almacenar estos datos?
  - Almacenaremos los datos en un sistema oncloud interno
  - Utilizaremos un sistema de archivos distribuido como Hadoop para los datos procesados por lotes y Spark para los procesados en tiempo real
  - Escalabilidad alta y un costo alto.

## 3. Procesamiento y Análisis:

- ★ ¿Qué tipo de procesamiento se necesita (por lotes o en streaming)?
  - ambos, creemos que se necesita de procesamiento por lotes para tareas donde se debe analizar patrones o áreas de riesgo para tomar medidas de seguridad
  - también en tiempo real para situaciones donde las cámaras deben reaccionar en tiempo real para ajustar los semáforos o enrutar el tráfico en caso de congestión o tráfico.
- ★ ¿Qué herramientas de análisis serían las más adecuadas (ej. SQL, Python, machine learning)?
  - Machine learning, para que detecte patrones

## 4. Gobernanza y Seguridad:

- ★ ¿Qué datos sensibles o personales podrían estar manejando? (ej. datos personales de clientes, historial de navegación)?
  - Imágenes de personas sin su consentimiento, imágenes violentas, imágenes de patentes
- ★ ¿Qué desafíos de seguridad y privacidad tendrían que considerar para proteger la información?
  - Proteger la información de los ciudadanos