

Proyecto Sistemas Distribuidos: Plataforma de Análisis de Tráfico en Región Metropolitana

Profesor: Nicolás Hidalgo

Ayudantes: Sofía Belmar, César Muñoz Rivera, J. Tomás Silva, Joaquín Villegas y Marcelo Yáñez

Introducción y Contexto

Las grandes ciudades enfrentan desafíos significativos en la gestión de la movilidad urbana debido al incremento del parque automotor y la densidad poblacional. En este contexto, el uso de plataformas colaborativas que integran información en tiempo real sobre el estado del tránsito, incidentes y otros eventos viales se ha convertido en una herramienta esencial para la toma de decisiones y la planificación urbana.

Considerando el creciente aumento del parque automotriz de la región metropolitana (RM), los alcaldes, en un esfuerzo mancomunado de colaboración, decidieron generar una plataforma regional para el monitoreo del estado del tráfico basado en Waze. La plataforma Waze¹, reconocida a nivel mundial, ofrece información actualizada y colaborativa sobre las condiciones del tráfico, alertas de accidentes y otros incidentes viales. Esta plataforma permitirá obtener y visualizar datos relevantes que permitan apoyar la toma de decisiones de la Unidad de Control de Tránsito (UCT) como también de los mismos municipios a través de su oficina de dirección de tránsito.

El aprovechamiento de estos datos permite la optimización de rutas, la prevención de accidentes y una mejor distribución de recursos para la gestión del tráfico. Sin embargo, para explotar plenamente esta información, es indispensable contar con sistemas capaces de extraer, procesar y visualizar información actualizada y relevante. Durante este semestre, usted y su compañero/a de trabajo deberán abordar este desafío diseñando un sistema que permita abordar esta problemática. El sistema debe ser implementado considerando su posible extensión a otras regiones, por lo que su rendimiento y capacidad de tolerar fallos son cruciales.

Metodología

El presente proyecto es de carácter semestral y se abordará por medio de 3 entregas o módulos. Cada una de estas entregas posee un objetivo particular, el cual debe ser **integrable** con los entregables restantes y previos. Cada entrega será calificada con una nota, la cual consta de un informe técnico del proceso de diseño, así como de las pruebas realizadas para validar su funcionamiento. Toda decisión debe ser justificada y el comportamiento del sistema respaldado en datos provenientes del proceso de evaluación de cada módulo.

Las entregas abordan 3 hitos clave del desarrollo como lo son:

1. Datos y Cache: abocados a la recuperación y manejo de datos/eventos.
2. Procesamiento: abocados a la preparación de los datos para su posterior análisis.
3. Visualización: abocado a proveer una vista agregada de métricas relevantes para los tomadores de decisiones.

Para desarrollar y desplegar los diferentes módulos y sus tecnologías se trabajará con Docker² como tecnología de virtualización de recursos. No se impone un stack tecnológico específico; la elección de herramientas, lenguajes y metodologías es libre, siempre que se justifiquen las decisiones tomadas de manera apropiada.

¹<https://www.waze.com>

²<https://www.docker.com/>

1. Entregable 1: Datos y Cache

El objetivo de la presente entrega es el desarrollo de las bases para poder desarrollar y validar el proyecto en cuestión. Acá buscamos definir un modelo de almacenamiento y caching para nuestro futuro sistema de análisis. Para ello se solicita

El sistema deberá cumplir los siguientes objetivos:

- Extraer de forma automatizada la información proporcionada por Waze³ (Web scraping). El scraping debe realizarse sobre la región metropolitana.
- Guardar en un sistema de almacenamiento los eventos obtenidos por el sistema de scrapping. Se debe contar con al menos 10.000 eventos.
- Implementar un sistema de generación automática de tráfico (consultas) que obtenga los datos del sistema de almacenamiento y los envíe siguiendo al menos dos distribuciones de tasa de arribo a definir los alumnos/as.
- Implementar un sistema de cache que consuma los eventos del generador de tráfico y almacene aquellos eventos repetitivos. El sistema de cache debe ser parametrizado de manera experimental.

Requerimientos del Sistema

El sistema a desarrollar debe cumplir con los siguientes requisitos funcionales:

■ Diseño Modular

- Estructurar el sistema en módulos independientes que faciliten su mantenimiento y futuras expansiones. Debe permitir la posterior integración con otros sistemas o componentes que potencien la funcionalidad del sistema a diseñar.
- Los módulos son: 1) scraper, 2) almacenamiento, 3) generador tráfico, y 4) cache. Estos se deben comunicar de manera secuencial. El scraper recupera datos que pone en almacenamiento. El almacenamiento es el repositorio de los datos que es consumido posteriormente por el generador de tráfico sintético para simular consultas con tasas de arribo. Finalmente el sistema de cache recibe consultas desde el generador de tráfico.

■ Extracción y almacenamiento

- Debe automatizar la extracción de información desde la plataforma Waze por medio de su live map (<https://www.waze.com/es-419/live-map/>)
- Debe implementar un generador de tráfico que aplique dos modelos de llegada de eventos.
- Debe definir un sistema de almacenamiento que pueda cumplir con respuestas rápidas y actualizaciones frecuentes. Se requieren al menos 10.000 eventos almacenados.

■ Cache

- Diseñe y parametrize un sistema de caching para consultas frecuentes. Evalúe diferentes políticas de remoción y tamaños de caché, considerando al menos dos políticas distintas. Analice también el comportamiento del sistema bajo diferentes distribuciones para la generación de tráfico.
- Parametrize y justifique sus decisiones en base al comportamiento experimental del cache.

■ Distribución de servicios

- Debe permitir que el sistema se pueda desplegar y ejecutar en contenedores Docker, facilitando su portabilidad y escalabilidad.

■ Documentación y Buenas Prácticas:

- Documente detalladamente el código y la funcionalidad implementada, incluyendo una justificación de las decisiones de diseño, la elección de tecnologías y las metodologías utilizadas.

³<https://www.waze.com/es-419/live-map/>

2. Análisis y Discusión

Responda, en el contexto del proyecto desarrollado, las siguientes preguntas:

1. **Generador de Tráfico:** ¿Por qué utiliza la distribución seleccionada en el generador de tráfico, en que se basa dicha decisión? ¿Es realista la distribución utilizada? ¿En que escenarios vemos este tipo de comportamientos?
2. **Almacenamiento:** ¿Por que utiliza este sistema de almacenamiento y no otro? ¿Qué comportamiento o funcionalidades son las que justifican su decisión? ¿Qué limitantes tiene el sistema utilizado?
3. **Métricas:** ¿Qué métricas considera esenciales para evaluar la eficiencia de su sistema de cache? ¿Qué política de remoción le resulta más eficiente para este sistema y sus eventos? ¿Tiene un efecto relevante sobre el rendimiento del cache la distribución de tráfico que emplea en el generador de trafico implementado?
4. **Pruebas de Rendimiento:** ¿Qué buscan los escenarios de prueba generados? ¿Es realista el escenario de pruebas implementado? ¿Provee alta disponibilidad y escalabilidad el diseño propuesto?

Resultados esperados entrega 1

- Un informe en formato \LaTeX (PDF) que describa de forma concisa la arquitectura y el enfoque del sistema distribuido, incluyendo:
 - La descripción de los componentes y funcionalidades implementadas.
 - Las justificaciones de las decisiones de diseño, tecnologías y metodologías utilizadas.
 - Un análisis acabado basado en los resultados y las métricas definidas por los alumnos/as.
- El código fuente completo del proyecto, alojado en un repositorio público en GitHub o GitLab. El enlace al repositorio deberá incluirse en el informe y en la sección de comentarios de la plataforma CANVAS.
- Un archivo `Dockerfile` o `docker-compose` que permita la ejecución de los servicios implementados, junto con instrucciones claras en el archivo `README.md` del repositorio.

Reglas y consideraciones de la entrega

- **Fecha de Entrega:** La fecha límite para la entrega de esta tarea es el día 29/04/2025 hasta las 23:59 hrs. Se recomienda gestionar el tiempo de forma efectiva. La entrega es vía Canvas del curso.
- **Integrantes:** La tarea debe ser realizada en grupos de hasta **2** alumnos/as.
- **Formato de Entrega:**
 - El informe debe estar en formato PDF generado a partir de \LaTeX .
 - El repositorio debe estar bien documentado y organizado, incluyendo instrucciones para la configuración y ejecución del proyecto.
- **Ética y Autoría:** Se debe respetar el reglamento de la universidad en cuanto a plagio y autoría. Cualquier evidencia de copia o falta de autoría conllevará sanciones según lo establecido.