

## Objetivo

El objetivo de este proyecto es poner en práctica los conceptos aprendidos en clase acerca de las estructuras de datos lineales: Lista, Pila y Cola, así como sobre algoritmos de ordenamiento y búsquedas eficientes de información.

## Contexto

Debido al crecimiento continuo de las ciudades en población, infraestructura y servicios, las organizaciones y agencias que intervienen en su administración y funcionamiento tienen como preocupación contar con información actualizada en los aspectos que afectan la vida de sus ciudadanos: educación, salud, transporte, vivienda, infraestructura, entretenimiento, seguridad, economía, entre otras. Esta información permite mantener informados a sus ciudadanos en sus actividades comunes y a los administradores y autoridades locales tomar decisiones que mejoren la calidad de vida de sus ciudadanos.

El transporte es una de las problemáticas importantes en una ciudad. En particular, el transporte público debe ofrecer soluciones eficientes para transportar un alto volumen de la población, con buena calidad, para así aumentar la productividad de la ciudad y la calidad de vida de sus ciudadanos. Además, un buen servicio de transporte público puede incentivar la reducción del transporte privado. En este aspecto, el servicio de taxis se viene transformando y mejorando con la prestación del servicio a través de plataformas tecnológicas. Entre las empresas que vienen impulsando este cambio a nivel mundial están Uber, Lyft y Hailo, entre otras.

El tema del proyecto está relacionado con el manejo de información de la prestación del servicio de taxi en la ciudad de Chicago (USA). Esta ciudad está a la vanguardia en el registro de información en diferentes aspectos de su funcionamiento (portal oficial de datos *Chicago Data Portal* <https://data.cityofchicago.org>). Para el análisis de su servicio de taxis utilizaremos como fuente de información, el sitio web de consulta de información disponible en <https://data.cityofchicago.org/Transportation/Taxi-Trips/wrvz-psew>.

Como parte de los servicios ofrecidos por el Portal se tiene el API SODA (*Socrata Open Data API*) que permite hacer consultas sobre la información disponible usando como punto de consulta la URL <https://data.cityofchicago.org/resource/wrvz-psew.json>. Para información del uso del API se puede consultar la URL <https://dev.socrata.com/>. En este proyecto, les suministraremos la información obtenida a través del API, por lo que **no es necesario** hacer uso directo de las APIs, pero es importante entender de donde proviene la información y para qué es utilizada.

En este proyecto, vamos a construir una aplicación que le permita entender a los administradores y autoridades de la ciudad de Chicago un conjunto de consultas importantes sobre el servicio de taxis.

## Las Fuentes de Datos

A continuación, se presenta una descripción de las fuentes de datos que se utilizarán en el proyecto.

1. *taxi-trips-wrvz-psew-subset-small.json*
2. *taxi-trips-wrvz-psew-subset-medium.json*
3. *taxi-trips-wrvz-psew-subset-large.json*

Cada fuente de datos contiene el detalle de un subconjunto de servicios de taxi (carreras) de periodos de longitud de tiempo diferentes. Cada servicio de taxi (carrera) se describe a partir de 23 datos entre los cuales se tienen: el identificador del taxi, el identificador del servicio, la empresa de afiliación del taxi, la ubicación geográfica de recogida y de terminación, la zona de la ciudad de recogida y de terminación, la fecha y la hora estimada de recogida, la fecha y hora estimada de terminación, su duración (en segundos), su distancia (millas), su costo total (US\$), entre otros datos. Consulte el sitio Web <https://data.cityofchicago.org/Transportation/Taxi-Trips/wrvz-psew> para conocer el detalle de la información que se incluye en cada servicio.

## Carga de Información

Para responder a los requerimientos presentados más adelante, usted deberá cargar la información de todos los archivos .JSON y/o .CSV. Solo es permitido leer una vez la información de los archivos.

## Requerimientos - Parte A (estudiante 1 de cada grupo)

- 1A-** Generar una Cola con todos los servicios de taxi que se prestaron en un periodo de tiempo dado por una fecha/hora inicial y una fecha/hora final de consulta. El inicio y terminación del servicio debe estar incluido dentro del periodo de consulta. Los servicios deben mostrarse en orden cronológico de su fecha/hora inicial.
- 2A-** Buscar el taxi de una compañía dada que más servicios inició en un periodo de tiempo dado por una fecha/hora inicial y una fecha/hora final de consulta.
- 3A-** Buscar la información completa de un taxi, a partir de su identificador, en un periodo de tiempo dado por una fecha/hora inicial y una fecha/hora final de consulta. Incluye el nombre de su compañía y los valores totales de plata ganada, de servicios prestados, de distancia recorrida y de tiempo total de servicios.
- 4A-** Retornar una lista de rangos de distancia recorrida, en la que se encuentran todos los servicios de taxis servidos por las compañías, en una fecha dada y en un rango de horas especificada. La información debe estar ordenada por la distancia recorrida, así la primera posición de la lista tiene a su vez una lista con todos los servicios cuya distancia recorrida esta entre [0 y 1) milla. En la segunda posición, los recorridos entre [1 y 2) millas, y así sucesivamente.

## Parte B (estudiante 2 de cada grupo)

- 1B-** Mostrar la información de las compañías de taxi consistente en: El total de compañías que tienen al menos un taxi inscrito y el total de taxis que prestan servicio para al menos

una compañía. Adicionalmente, generar la lista alfabética de compañías a las cuales aparecen inscritos los servicios de taxi de la fuente de datos de consulta. Por cada compañía debe informarse su nombre y el número de taxis que tiene registrados.

**2B-** Buscar el taxi de una compañía dada que mayor facturación ha generado en un periodo de tiempo dado por una fecha/hora inicial y una fecha/hora final de consulta.

**3B-** Buscar la información completa de una zona de la ciudad en un periodo de tiempo dado por una fecha/hora inicial y una fecha/hora final de consulta. El número total de servicios que se recogieron en la zona de consulta y terminaron en otra zona y el valor total pagado por los usuarios; el número total de servicios que se recogieron en otra zona y terminaron en la zona de consulta y el valor total pagado por los usuarios, y el total de servicios que iniciaron y terminaron en la misma zona de consulta y el valor total pagado por los usuarios.

**4B-** Retornar una lista con todas las zonas de la ciudad (ordenadas por su identificador). Cada zona debe tener el total de servicios iniciados en dicha zona en un rango de fechas. Por ejemplo, la primera posición de la lista tiene todos los servicios de la primera zona, en dicha posición, se tiene una lista de fechas (ordenadas cronológicamente) con el total de servicios asociados a dicha fecha.

## Parte C (trabajo en grupo)

**1C-** Cargar toda la información de una fuente de datos seleccionada por el usuario (día, semana, mes o trimestre) y generar los VO y las estructuras de datos necesarias.

**2C-** Identificar el top X de compañías que más servicios iniciaron en un periodo de tiempo dado por una fecha/hora inicial y una fecha/hora final. El valor X es un dato de consulta. El resultado debe mostrar el Top X de compañías ordenadas por el número de servicios de mayor a menor. Por cada compañía debe informarse su nombre y su número de servicios de respuesta.

**3C-** Buscar el taxi más rentable de cada compañía. El taxi más rentable de una compañía es aquel cuya relación de plata ganada y distancia recorrida en los servicios prestados es mayor.

**4C-** Dada la gran cantidad de datos que requiere el proyecto, se desea poder compactar información asociada a un taxi particular. Para ello usted debe guardar en una pila todos los servicios generados por el taxi en orden cronológico, entre una hora inicial y una hora final, en una fecha determinada. Para comprimir la información, usted debe ir guardando los servicios reportados por el taxi, siempre y cuando la distancia acumulada de dichos servicios sea inferior a un valor dado (por ejemplo 10 millas). Cuando la sumatoria de las distancias de los servicios acumulados en la pila sobrepasen el valor límite, usted deberá retirar los servicios de la pila y generar uno nuevo, que resuma los servicios analizados. Este servicio resumen, tiene como hora inicial, la menor hora de los servicios, como hora final la mayor hora de los servicios, como distancia la sumatoria de todas las distancias de los servicios; como duración, la sumatoria de las duraciones de los servicios; y como valor total de la carrera, la sumatoria de todos los valores de los servicios. Los demás campos deben quedar en el valor por defecto dependiendo del tipo de dato.

## Restricciones

- Los datos contenidos en los archivos sólo se pueden leer una vez
- Se deberá trabajar en Java 7
- El proyecto se debe implementar en Eclipse
- La entrada/salida de información adicionales se debe realizar por consola
- **No usar las colecciones del API Java.**