



**Instituto Tecnológico de Buenos Aires**

## **MAPREDUCING PARKING TICKETS**

*72.42 Programación Orientada a Objetos - 2024Q1*

### **Alumnos:**

**Tomás Santiago Marengo, 61587**

**Abril Occhipinti, 61159**

**Santino Ranucci, 62092**

**Agustin Zakalik, 62068**

### **Profesores:**

**Ing. Marcelo Turrín**

**Ing. Franco Román Meola**

## **Índice**

<b>1. Introducción</b>	<b>2</b>
<b>2. Anotaciones</b>	<b>2</b>

## 1. Introducción

- Cómo se diseñaron los componentes de cada trabajo MapReduce, qué decisiones se tomaron y con qué objetivos. Además alguna alternativa de diseño que se evaluó y descartó, comentando el porqué.
- El análisis de los tiempos para la resolución de cada query: En caso de poder, analizar la diferencia de tiempos de correr cada query aumentando la cantidad de nodos (hasta 5 nodos) en una red local. De no poder, intentar predecir cómo sería el comportamiento.
- Potenciales puntos de mejora y/o expansión.
- La comparación de los tiempos de las queries ejecutándose con y sin Combiner.
- Otro análisis de tiempos de ejecución de las queries utilizando algún otro elemento de optimización a elección por el grupo.
- Para todos los puntos anteriores, no olvidar de indicar el tamaño de los archivos utilizados como entrada para las pruebas (cantidad de registros).

## 2. Anotaciones

- **Hablar del parallel read.**

Reading a single file at multiple positions concurrently wouldn't let you go any faster (but it could slow you down considerably).

Instead of reading the file from multiple threads, read the file from a single thread, and parallelize the processing of these lines. A single thread should read your CSV line-by-line, and put each line in a queue. Multiple working threads should then take the next line from the queue, parse it, convert to a request, and process the request concurrently as needed. The splitting of the work would then be done by a single thread, ensuring that there are no missing lines or overlaps.

- **Hablar del .json que recibe las columnas del CSV**
- **Hablar del DateFormats que recibe todos los tipos de formatos de fecha con los que se trabaja**