



# Sistemas Distribuidos I (75.74)

## Yelp Reviews Analysis

TP2: Middleware y Coordinación de Procesos

### Docentes

- Pablo D. Roca
- Ezequiel Torres Feyuk
- Guido Albarello



# Requerimientos Funcionales

- Se solicita un sistema distribuido que procese el detalle de críticas sobre el servicio de comercios.
- El sistema debe recibir la transmisión de los datos a ser procesados y retornar:
  - Usuarios con 50 o más *reviews*.
  - Usuarios con 50 o más *reviews* con comentarios 5 estrellas únicamente.
  - Usuarios con 5 o más *reviews* que utilizan siempre el mismo texto.
- A su vez, se desea obtener la siguiente información estadística:
  - Histograma de cantidad de comentarios por día de la semana (Lu, Ma, ..., Do)
  - Listado de las 10 ciudades más divertidas: con mayor Cant. de *reviews* 'funny'.
- Como origen de datos se definen los archivos de ingreso registrados en
  - <https://www.kaggle.com/pabloroca/yelp-review-analysis>



# Requerimientos No Funcionales

- El sistema debe estar optimizado para entornos multicomputadoras
- El sistema debe ser invocado desde un nodo que transmite los datos a ser procesados.
- Se debe soportar el escalamiento de los elementos de cómputo
- De ser necesaria una comunicación basada en grupos, se requiere la definición de un *middleware*
- El diseño debe permitir la adaptación para eventuales procesamiento en *streaming*, es decir, a medida que se reciben los nuevos comentarios.
- Debido a restricciones en el tiempo de implementación, se permite la construcción de un sistema acoplado al modelo de negocio. No es un requerimiento la creación de una plataforma de procesamiento de datos



Se espera del alumno:

- Empleo del tiempo de consultas en clase para resolver dudas y clarificar el negocio del sistema a construir previo a su diseño
- Exposición y verificación en clase de la arquitectura propuesta antes de iniciar su implementación
- Empleo del grupo de correos para realizar consultas que no pudieran ser resueltas en clase
- Consideración de prácticas distribuidas según lo estudiado en clase para elaborar una arquitectura flexible, escalable y robusta
- Aprobación del cuerpo docente para el uso de cualquier librería.
- Demo del sistema en funcionamiento previamente ensayada



- Fecha de entrega:
  - 24/11/2020
- Fecha de re-entrega:
  - 10/12/2020
- Formato de entrega:
  - Demostración del sistema.
  - Entrega digital mediante correo personal.
  - Entrega impresa:
    - Carátula
    - Documento 4+1 *Views* o C4Model incluyendo al menos un diagrama de: robustez, despliegue, actividades, paquetes y secuencia.
    - Código impreso