



# Sistemas Distribuidos I

## (75.74)

## Reddit Memes Analyzer HA

TP Final: Alta Disponibilidad y Tolerancia a Fallos

### Docentes

- Pablo D. Roca
- Ezequiel Torres Feyuk
- Ana Czarnitzki
- Cristian Raña



# Requerimientos Funcionales

- Se solicita un sistema distribuido que realice análisis de *posts* de Reddit sobre memes *#me\_irl* y sus *comments*.
- Los *posts* a analizar cuentan con URL del meme, *score* y muchos *comments* que poseen texto y un análisis de sentimiento (*sentiment*).
- Los *comments* son ingestados en el sistema a medida que se recolectan de Reddit. Aquellos con texto vacío o eliminado deben ser ignorados.
- Se debe obtener:
  - Promedio de *score* de todos los *posts*
  - URLs de memes que gustan a estudiantes (con *comments* sobre university, college, student, teacher, professor y con *score* mayor al promedio)
  - Descarga del meme con mejor *sentiment* promedio



# Requerimientos No Funcionales

- Para construir una simulación realista se define la serie de datos:
  - <https://www.kaggle.com/datasets/pavellexyr/the-reddit-irl-data-set>
- El análisis de sentimiento será provisto dentro de cada *comment*
- El sistema debe estar optimizado para entornos multicomputadoras
- Se debe soportar el incremento de los elementos de cómputo para escalar los volúmenes de información a procesar
- De ser necesaria una comunicación basada en grupos, se requiere la definición de un *middleware*
- Se debe soportar una única ejecución del procesamiento y proveer *graceful quit* frente a señales SIGTERM.



# Nuevos Requerimientos

- El sistema debe mostrar alta disponibilidad hacia los clientes
- El sistema debe ser tolerante a fallos como caída de procesos
- El sistema debe permitir ser re-ejecutado una vez finalizado el procesamiento sin necesidad de reiniciar el servidor.
- En caso de usar un algoritmo de consenso, el mismo tiene que ser implementado por los alumnos
- Se puede utilizar docker-in-docker para levantar procesos caídos
- No se puede utilizar la API de docker para verificar si un nodo está up o no



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:
  - 07/07/2022
- Formato de entrega:
  - Demostración del sistema.
  - Entrega digital mediante correo personal.
  - Entrega impresa:
    - Carátula
    - Documento de arquitectura 4+1 *Views* o C4Model con los siguientes diagramas como mínimo: robustez, despliegue, actividades, paquetes y secuencia.