



Sistemas Distribuidos I (75.74)

Movies Analysis

TP Escalabilidad: Middleware y Coordinación de Procesos

Docentes

- Pablo D. Roca
- Gabriel Robles
- Franco Barreneche
- Nicolás Zulaica
- Franco Papa
- Manuel Reberendo



Requerimientos Funcionales

- Se solicita un sistema distribuido que analice la información de películas y los *ratings* de sus espectadores en plataformas como IMDb.
- Los *ratings* son un valor numérico de 1 al 5. Las películas tienen información como género, fecha de estreno, países involucrados en la producción, idioma, presupuesto e ingreso.
- Se debe obtener:
 1. Películas y sus géneros de los años 2000 con producción Argentina y Española.
 2. Top 5 de países que más dinero han invertido en producciones sin colaborar con otros países.
 3. Película de producción Argentina estrenada a partir del 2000, con mayor y con menor promedio de rating.
 4. Top 10 de actores con mayor participación en películas de producción Argentina con fecha de estreno posterior al 2000
 5. Average de la tasa ingreso/presupuesto de peliculas con overview de sentimiento positivo vs. sentimiento negativo



Requerimientos No Funcionales

- 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
- Se requiere del desarrollo de un Middleware para abstraer la comunicación basada en grupos.
- Se debe soportar una única ejecución del procesamiento y proveer *graceful quit* frente a señales SIGTERM.



Datasets, notebook patrón y librerías

- Para construir una simulación realista, se trabajará sobre el siguiente dataset:
 - <https://www.kaggle.com/datasets/rounakbanik/the-movies-dataset>
 - Serán necesario los files "movies_metadata.csv", "ratings.csv" y "credits.csv".
- Se usarán los valores del siguiente notebook como resultados patrón:
 - <https://www.kaggle.com/code/gabrielrobes/fiuba-distribuidos-1-the-movies>
- Se deberán utilizar herramientas de procesamiento de lenguaje natural (PNL) para la 5ta consulta. (Sugerencia: [transformers de Hugging Face en Python](#))



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: 24/04/2025
- Formato de entrega:
 - Entrega presencial.
 - Documento de arquitectura **actualizado** 4+1 Views o C4Model incluyendo al menos: diagramas de robustez, despliegue, actividades, paquetes, secuencia y DAG.
 - Listado de tareas a ejecutar y división entre integrantes.
 - Demo en vivo del sistema funcionando con al menos una porción del dataset.
 - **No se evaluará código.**