

Obligatorio 2025

Contexto

La empresa UMovie es un nuevo servicio que provee películas por streaming en internet y está pensando en expandir el catálogo en Uruguay. La compañía viene trabajando desde hace un tiempo en el mercado y cuenta con información de sus abonados y le contratan a usted y su equipo para desarrollar un software que ayude a la compañía en la decisiones estratégicas y en algunas operativas.

Luego de cada reproducción de una película, el usuario puede realizar una calificación sobre la película. Esta calificación va desde 0 a 5 puntos.

El equipo de tecnologías de UMovie le proporcionará un dataset con los datos de las calificaciones de películas de distintos abonados, un dataset con la metadata de cada película y un tercer dataset con la definición de los participantes de cada película (actores, director, escritor, etc...). Usted y su equipo deberán diseñar un modelo de datos el cual sea eficiente y eficaz para almacenar los datos procesados del dataset teniendo como base las funciones que se solicitan y detallan a continuación.

1. Tomando en cuenta las películas y las evaluaciones que los usuarios realizaron sobre las mismas, indicar el Top 5 de las películas en idioma original Inglés, Francés, Italiano, Español y Portugués con más evaluaciones por parte de los usuarios.

Datos esperados del resultado:

- Id de la película.
- Título de la película.
- Total de evaluaciones.
- Idioma Original

2. Con las mismas consideraciones que en el caso anterior, indicar el Top 10 de las películas que mejor calificación media tienen por parte de los usuarios.

Datos esperados del resultado:

- Id de la película.
- Título de la película.
- Calificación media.

3. Una película puede pertenecer a una saga (o colección de películas). Agrupando las películas por saga, indicar el Top 5 de las colecciones que más ingresos generaron. Si una película no es parte de una saga, ella misma se debe considerar como una saga

Datos esperados del resultado:

-
- Id de la colección (o película).
 - Título de la colección (o película).
 - Cantidad de películas.
 - Conjunto de Ids de películas que la componen.
 - Ingresos generados.
4. Una película tiene un director. En general, los directores tienen cierto renombre y sus películas suelen tener una calidad similar. Considerando todas las películas de cada director, indicar el Top 10 de los directores que mejor calificación tienen. Únicamente se deben considerar directores con más de 1 película y más de 100 evaluaciones. Se deberá utilizar la mediana como métrica de comparación.
- Datos esperados del resultado:
- Nombre del director.
 - Cantidad de películas.
 - Mediana de su calificación.
5. Es sabido que hay épocas en las que ciertas películas son más vistas que otras. Para promocionar mejor estas películas, se desea saber qué actor es más común en ellas. Puntualmente, se desea saber cual es el actor con más calificaciones recibidas en cada mes del año.
- Datos esperados del resultado:
- Mes del año.
 - Nombre del actor.
 - Cantidad de películas (que fueron vistas en ese mes).
 - Cantidad de calificaciones.
6. Finalmente, la empresa quiere premiar a los usuarios que más participen evaluando películas, para ello necesita saber que usuarios evalúan más películas por cada género. Pero al haber tantos géneros solo necesitan saber los top 10 géneros más vistos.
- Datos esperados del resultado:
- Id del usuario
 - Género
 - Cantidad de evaluaciones de ese usuario sobre ese género.

Se debe realizar una aplicación Java que a través de la entrada estándar permita:

1. Realizar la carga masiva de datos a través de un archivo CSV. Puede asumir que el mismo se encuentra en la raíz del proyecto.
2. Realizar un menú a través de la entrada estándar, que permita la ejecución de los reportes.

A su vez, se debe realizar un documento, que contenga:

1. Diagrama UML de clases de la solución.
2. Breve descripción de los procesos de carga y realización de reportes.
3. Medición de eficiencia de la aplicación.
 - a. Para ello se deberá indicar, tanto para la carga como para la ejecución de los reportes el tiempo de ejecución promedio en milisegundos.
 - b. Se deberá realizar un análisis de memoria usando por operación, presentando gráficas donde en el eje de las X se presenta el tiempo y en el eje de las Y la cantidad de memoria consumida. Para ello se van usar herramientas de profiling, como JConsole.

Consideraciones generales:

- El proyecto se debe llamar “grupoX-p2-tads”, donde X debe sustituirse con el número de grupo asignado.
- Se debe utilizar los TADs que se proporcionan en el proyecto y NO es posible en esta instancia utilizar TADs de terceros (brindados por la máquina virtual de Java u otros).
- Se debe crear un repositorio GIT para la entrega del obligatorio. Se va a verificar que todos los participantes hayan realizado commits sobre los proyectos. Se va a penalizar en caso de que esto no suceda.
- Se deberá elegir la estructura más adecuada para resolver cada problema.
- No se admiten estructuras intermedias que almacene los resultados de las consultas. Se deben realizar los cálculos en el momento en que se solicita correr una consulta.
- Para descargar la fuente de datos puede acceder a un link que se compartirá vía moodle. Todos los archivos son de tipo ‘csv’, con codificación UTF-8.
 - El mismo contiene 3 archivos con la siguientes características:
 - i. movies_metadata.csv: contiene 45.5K líneas con las películas y su metadata
 - ii. ratings.csv: contiene 1MM de líneas con las calificaciones de los usuarios.
 - iii. credits.csv: contiene 45.5K de líneas con metadata de los participantes de las películas.
- Fecha de entrega: Viernes 27 de junio hasta las 23:59. Se debe entregar vía moodle un zip que contenga:
 - Código fuente de la solución.
 - Informe
 - Binario ejecutable en formato Jar (va a ser explicado como generarlo en la clase de práctico).

Dado que la corrección se realizará de manera automática se deberán respetar los siguientes menús, sin dejar de lado ninguna de las opciones mencionadas. La implementación incorrecta de la numeración, orden de las opciones o

incumplimiento del formato de salida será penalizado.

Menú principal:

Seleccione la opción que desee:

1. Carga de datos
2. Ejecutar consultas
3. Salir

En caso de ingresar 1, se deberán cargar los datos y se deberá imprimir en pantalla lo que se detalla a continuación:

Carga de datos exitosa, tiempo de ejecución de la carga:

Por otro lado, en caso de ingresar 2 se verá el Menú que se detalla a continuación.

1. Top 5 de las películas que más calificaciones por idioma.
2. Top 10 de las películas que mejor calificación media tienen por parte de los usuarios.
3. Top 5 de las colecciones que más ingresos generaron.
4. Top 10 de los directores que mejor calificación tienen.
5. Actor con más calificaciones recibidas en cada mes del año.
6. Usuarios con más calificaciones por género
7. Salir

Algo importante para tener en cuenta, es que una vez que el usuario seleccione una de las consultas, por ejemplo, ingrese el valor 1, luego de mostrar el resultado que se indica más abajo se deberá volver al menú de las consultas (el listado recién mencionado). A continuación se muestran los resultados esperados a partir de la interacción con el menú de consultas

1. Top 5 de las películas que más calificaciones por idioma.
Al seleccionar dicha opción se deberán mostrar los datos de la siguiente manera:

*<id_pelicula>, <titulo_pelicula>, <total_calificaciones>, <idoma>
Tiempo de ejecución de la consulta: <tiempo_ejecucion>*

2. Top 10 de las películas que mejor calificación media tienen por parte de los usuarios.
Al seleccionar dicha opción se deberán mostrar los datos de la siguiente manera:

*<id_pelicula>, <titulo_pelicula>, <calificacion_media>
Tiempo de ejecución de la consulta: <tiempo_ejecucion>*

3. Top 5 de las colecciones que más ingresos generaron..
Al seleccionar dicha opción se deberán mostrar los datos de la siguiente manera:

*<id_coleccion>,
<titulo_coleccion>,<cantidad_peliculas>,[id_pelicula_1,id_pelicula_2]<ingreso_generado>
Tiempo de ejecución de la consulta: <tiempo_ejecucion>*

4. Top 10 de los directores que mejor calificación tienen..
Al seleccionar dicha opción se deberán mostrar los datos de la siguiente manera:

*<nombre_director>,<cantidad_peliculas>,<mediana_calificacion>
Tiempo de ejecución de la consulta: <tiempo_ejecucion>*

5. Actor con más calificaciones recibidas en cada mes del año.
Al seleccionar dicha opción se deberán mostrar los datos de la siguiente manera:

*<mes>,<nombre_actor>,<cantidad_peliculas>,<cantidad de calificaciones>
Tiempo de ejecución de la consulta: <tiempo_ejecucion>*

6. Usuarios con más calificaciones por género
Al seleccionar dicha opción se deberán mostrar los datos de la siguiente manera:

*<id_usuario>,<genero>,<cantidad de calificaciones sobre ese género>
Tiempo de ejecución de la consulta: <tiempo_ejecucion>*