

Práctica 3. PyFlink

Enunciado:

Estás trabajando en el área de análisis de una aplicación de video en streaming. Cuentas con datos de usuarios y visualizaciones de contenido, y necesitas procesarlos para obtener métricas relevantes para el negocio.

Tabla de User (Usuarios)

user_id	name	country
1	Alice	USA
2	Bob	Canada
3	Carla	USA
4	David	Mexico
5	Eva	Canada

Tabla de View (visualizaciones)

view_id	user_id	video_id	watch_time_minutes	date
101	1	A1	10	2024-12-01
102	1	A2	30	2024-12-02
103	2	A2	15	2024-12-03
104	3	A1	5	2024-12-03
105	4	A3	20	2024-12-04
106	5	A1	50	2024-12-04
107	2	A4	25	2024-12-05
108	5	A2	35	2024-12-05
109	1	A3	40	2024-12-06
110	4	A1	60	2024-12-06

Ejercicio 1.

Configurar PyFlink en modo batch y crear dos tablas temporales a partir de datos simulados de la tabla de usuarios y de la tabla views (visualizaciones)

Ejercicio 2.

Calcular el tiempo total de visualización por usuario usando Table API

Ejercicio 3.

Obtener el país con mayor tiempo de visualización total usando SQL

Ejercicio 4.

Promedio de tiempo de visualización por video usando Table API

Ejercicio 5.

Contar cuántos usuarios únicos vieron cada video (SQL)

Ejercicio 6.

Obtener los 3 usuarios con más minutos vistos (SQL)

Ejercicio 7.

Determinar qué día tuvo más visualizaciones en total (Table API)