

Práctica 2. PyFlink

Enunciado:

En esta práctica debes realizar un análisis financiero de una empresa internacional que opera en varios países de Europa. El equipo de Contabilidad ha detectado que los ingresos (ingresos) de ciertas regiones fueron reportados con un error donde algunos países sus valores han sido multiplicados por 2.

Tenéis una tabla de pedidos (ordenes) que incluye los nombres de los clientes, su país y el ingreso por pedido donde debéis de realizar el ajuste a esos ingresos antes de que los datos se trasladen a otro sistema.

Pista: Debéis de realizar UDF definida como función lamda en PyFlink que corrija los ingresos.

Los puntos a desarrollar en la práctica son los siguientes:

Ejercicio 1. Crea un entorno en modo batch para procesamiento por lotes.

Ejercicio 2. Define la tabla orders con los siguientes datos:

- Alicia, España, 1500
- Roberto, Alemania, 2300
- Esther, Francia, 1239

Ejercicio 3. Crea una UDF tipo Pandas, definida con lambda:

- columnas nombre, país e ingreso
- multiplica por 2 el ingreso
- devuelve un DataFreame con nombre e ingreso ajustado

Ejercicio 4. Aplica esta función a la tabla

Ejercicio 5. Imprime el resultado final con los ingresos corregidos.