

CUBOS OLAP

Supuestos y alcance

- Solo usarás archivos CSV, PDI y pasos estándar de Spoon.
 - El flujo simula un "cubo": manipulas hechos y dimensiones, y luego puedes hacer pivoteo/agregación manual.
-

1. Estructura de los archivos CSV

Supón que analizas ventas simples:

1. ventas.csv (hechos)

```
id_venta,fecha,id_cliente,id_producto,cantidad,importe
1,2025-10-25,1,101,2,30
2,2025-10-25,2,102,1,40
3,2025-10-26,1,102,3,120
```

2. clientes.csv (dimensión)

```
id_cliente,nombre,ciudad
1,Ana,León
2,Juan,Managua
```

3. productos.csv (dimensión)

```
id_producto,producto,categoria
101,Maní Natural,SNACK
102,Frutas Deshidratadas,FRUTA
```

2. Diseño del flujo en Pentaho PDI

Paso 1: Leer los CSV

- Crea una nueva transformación.
- Agrega tres pasos "CSV File Input" e importa cada uno de los tres archivos según corresponda.
- Configura los campos y tipos correctamente en cada uno.

Paso 2: Unir hechos y dimensiones

- Usa dos pasos "Join Rows (Inner join)":
 - Primer join: ventas con clientes, usando id_cliente.
 - Segundo join: resultado anterior con productos, usando id_producto.

Paso 3: Agregar dimensiones y medidas

- Tendrás un resultado con todos los campos (fecha, cliente, producto, cantidad, importe, etc.)
- Puedes hacer un paso de "Group By" para simular la agregación OLAP, por ejemplo:
 - Group by: ciudad, producto, fecha
 - Aggregates: SUM cantidad, SUM importe

Paso 4: Exportar resultados

- Usa un paso "CSV File Output" para guardar el resultado ya "agregado", que simula la consulta OLAP deseada.
-

3. Ejemplo de configuración paso a paso

3.1 Preparando los pasos CSV

- Paso 1: "CSV File Input - ventas"
 - Archivo: ventas.csv
 - Campos: id_venta, fecha, id_cliente, id_producto, cantidad, importe
- Paso 2: "CSV File Input - clientes"
 - Archivo: clientes.csv
 - Campos: id_cliente, nombre, ciudad
- Paso 3: "CSV File Input - productos"
 - Archivo: productos.csv
 - Campos: id_producto, producto, categoria

3.2 Unir tablas

- Paso 4: "Join ventas-clientes"
 - Join: ventas.id_cliente = clientes.id_cliente
- Paso 5: "Join resultado-productos"
 - Join: resultado.id_producto = productos.id_producto

3.3 Simular consulta OLAP (agregación)

- Paso 6: "Group By"
 - Group fields: ciudad, producto, fecha
 - Aggregate fields: SUM cantidad, SUM importe

3.4 Resultados

- Paso 7: "CSV Output - ventas_olap.csv"
 - Exporta los resultados ya agregados.
-

4. Ejecuta y valida

- Ejecuta la transformación en Spoon.
- El archivo ventas_olap.csv es tu "vista OLAP". Puedes abrirlo en Excel/Pivot Table para ver los datos cruzados por dimensiones.