

GUIA PASO A PASO CUBOS OLAP SIMPLE

Ejercicio Práctico: Simulación de Cubo OLAP con Pentaho (Interfaz en Inglés)

Este ejercicio te guiará paso a paso para simular un cubo OLAP usando solo Pentaho Data Integration (versión 8) y un archivo CSV, sin necesidad de bases de datos ni complementos.

Aprenderás a simular las operaciones OLAP clave:

- Drill down
- Roll up
- Slice
- Dice

Todo desde la interfaz de Pentaho Spoon en inglés, con instrucciones precisas de lo que debes hacer en cada paso.

Paso 1: Preparar el archivo CSV

Crea un archivo llamado ventas.csv con el siguiente contenido:

```
id_venta,producto,categoría,region,mes,trimestre,anio,ventas,cantidad
1,Carrito,Miniaturas,Norte,Enero,1,2023,150,10
2,Muñeca,Juguetes,Sur,Enero,1,2023,200,8
3,Carrito,Miniaturas,Sur,Febrero,1,2023,100,5
4,Juego,Mesa,Norte,Febrero,1,2023,180,6
5,Muñeca,Juguetes,Norte,Marzo,1,2023,220,11
6,Carrito,Miniaturas,Sur,Marzo,1,2023,250,12
7,Juego,Mesa,Norte,Abril,2,2023,300,15
8,Muñeca,Juguetes,Sur,Mayo,2,2023,240,9
9,Carrito,Miniaturas,Norte,Mayo,2,2023,170,7
10,Juego,Mesa,Sur,Junio,2,2023,320,16
11,Carrito,Miniaturas,Norte,Julio,3,2023,210,10
12,Muñeca,Juguetes,Sur,Agosto,3,2023,260,13
13,Juego,Mesa,Norte,Septiembre,3,2023,350,14
14,Carrito,Miniaturas,Sur,Octubre,4,2023,280,14
15,Muñeca,Juguetes,Norte,Noviembre,4,2023,310,15
16,Juego,Mesa,Sur,Diciembre,4,2023,380,19
```

Guarda este archivo en la carpeta de ejercicios.

Paso 2: Abrir Pentaho

1. Abre Pentaho Data Integration (Spoon).

2. Cuando se inicie, crea una nueva transformación:

- Ve a [File > New > Transformation](#)

Paso 3: Leer el archivo CSV

1. En la barra de herramientas izquierda, busca:
 - Input → CSV file input
2. Arrástralos al lienzo principal.
3. Haz doble clic sobre el ícono CSV file input.
4. En la ventana que aparece:

Filename	Seleccione el archivo ventas.csv
Delimiter	,
Enclosure	(dejar vacío)
<input checked="" type="checkbox"/> Header	Marca esta casilla
<input checked="" type="checkbox"/> No empty lines	Marca esta casilla

5. Haz clic en "Preview rows" para confirmar que los datos se leen correctamente.
6. Haz clic en OK.

Paso 4: Simular las operaciones OLAP

Vamos a crear cuatro ramas diferentes desde el mismo origen ([CSV file input](#)) para simular cada operación OLAP.

☞ Importante: Usa "Send rows to multiple targets" para dividir el flujo.

Haz clic derecho sobre el ícono [CSV file input](#) → "Send rows to multiple targets" → marca las 4 opciones que crearemos.

❖ Operación 1: Drill Down (Profundizar)

Ir de un nivel general a uno más detallado.

Ejemplo: de trimestre a mes.

Pasos en Pentaho:

1. Agrega un nuevo paso:

- Transform → [Group by](#)

2. Conéctalo al [CSV file input](#) (una de las salidas múltiples).

3. Haz doble clic en [Group by](#).

Group field

mes

Group field

producto

Subject field

ventas

Type of calculation

Sum

Resultfieldname

ventas_por_mes

4. Haz clic en OK.

5. Agrega un paso: "Dummy" o "Write to log" para ver resultados:

- Output → [Write to log](#)

6. Conéctalo al [Group by](#).

7. Ejecuta la transformación (flecha verde).

8. Haz doble clic en [Write to log](#) en la ventana de ejecución para ver los resultados.

❖ Verás las ventas agrupadas por mes y producto → ¡Has hecho un drill down!

❖ **Operación 2: Roll Up (Resumir)**

Subir de nivel, resumiendo datos.

Ejemplo: de mes a trimestre.

Pasos en Pentaho:

1. Agrega un nuevo paso:

- Transform → Group by

2. Conéctalo al CSV file input.

3. Haz doble clic en Group by.

Group field

trimestre

Group field

categoria

Subject field

ventas

Type of calculation

Sum

Resultfieldname

ventas_trimestre

4. Haz clic en OK.

5. Agrega Write to log y conéctalo.

6. Ejecuta y visualiza.

❖ Resultado: ventas resumidas por trimestre y categoría → ¡Roll up completado!

❖ **Operación 3: Slice (Cortar una rebanada)**

Filtrar por una sola dimensión.

Ejemplo: solo datos del mes "Mayo".

Pasos en Pentaho:

1. Agrega un nuevo paso:

- Flow → Filter rows

2. Conéctalo al CSV file input.

3. Haz doble clic en Filter rows.

Configura la condición:

Condition

Field [mes] = String 'Mayo'

Puedes usar el constructor gráfico:

- Clic en **[+]** → Elige **mes** → Operador **=**, Valor **'Mayo'**

4. Haz clic en OK.

5. Agrega un **Group by** después del filtro:

- Group field: **producto**, **region**
- Subject field: **ventas**
- Calculation: **Sum**
- Result fieldname: **ventas_mayo**

6. Agrega **Write to log** y ejecuta.

❖ Verás solo los datos de Mayo → ¡Has hecho un slice!

❖ Operación 4: Dice (Cortar en varias dimensiones)

Filtrar por múltiples dimensiones.

Ejemplo: **mes = 'Mayo'** + **region = 'Norte'** + **producto = 'Carrito'**

Pasos en Pentaho:

1. Agrega un nuevo paso:

- Flow → **Filter rows**

2. Conéctalo al **CSV file input**.

3. Haz doble clic en **Filter rows**.

Configura la condición:

1
[mes] = 'Mayo' AND [region] = 'Norte' AND [producto] = 'Carrito'

Usa el constructor:

- **[+]** → **mes = 'Mayo'**
- **[+]** → **AND** → **region = 'Norte'**
- **[+]** → **AND** → **producto = 'Carrito'**

4. Haz clic en OK.

5. Agrega **Write to log** y conéctalo directamente (solo hay una fila).

6. Ejecuta.

❖ Resultado: solo 1 fila con 170 en ventas → ¡Has hecho un dice!

Paso 5: Ejecutar y visualizar

1. Haz clic en el botón verde "Run" (en la barra superior).
2. Se abrirá la ventana de ejecución.
3. Espera a que termine.
4. Haz doble clic en cada "Write to log" para ver los resultados de cada operación.

 Puedes etiquetar cada [Write to log](#) con comentarios:

- Haz clic derecho → [Add a comment](#)
- Escribe: "Drill Down", "Roll Up", etc.

Paso 6: Guardar la transformación

1. Ve a [File > Save as...](#)

¿Qué has aprendido?

Con este ejercicio, has simulado un [cubo OLAP](#) usando solo:

- Un archivo CSV
- Pentaho Spoon en inglés
- Transformaciones básicas

Y has aplicado las 4 operaciones clave del análisis OLAP:

Drill down	Pasaste de trimestre a mes (más detalle)
Roll up	Resumiste datos por trimestre y categoría
Slice	Filtraste solo por "Mayo"
Dice	Filtraste por tres condiciones a la vez

Recomendaciones

- Usa [comentarios](#) en el lienzo para organizar las ramas.

- Puedes reemplazar [Write to log](#) por "Text file output" para guardar cada resultado en un CSV.
- Este flujo es ideal para [preparar datos](#) antes de cargarlos a un verdadero cubo OLAP (con Mondrian).

INTERPRETACION DE RESULTADOS

Objetivo del Drill Down

En este caso, el drill down consiste en profundizar desde un nivel más general (trimestre) hacia un nivel más detallado: el mes.

Estamos agrupando las ventas por:

- mes
- producto

Y calculando la suma de ventas para cada combinación.

Operación realizada en Pentaho

Usamos el paso:

Group by

- Group fields: mes, producto
 - Sum of ventas → as ventas_por_mes
-

Resultado del Drill Down

Este es el resultado que verás al hacer clic en "Write to log" después del Group by:

Enero	Carrito	150
Enero	Muñeca	200
Febrero	Carrito	100
Febrero	Juego	180

Marzo	Muñeca	220
Marzo	Carrito	250
Abril	Juego	300
Mayo	Muñeca	240
Mayo	Carrito	170
Junio	Juego	320
Julio	Carrito	210
Agosto	Muñeca	260
Septiembre	Juego	350
Octubre	Carrito	280
Noviembre	Muñeca	310
Diciembre	Juego	380

Interpretación del resultado

Este resultado representa un análisis detallado por mes y producto, lo cual es un claro ejemplo de drill down:

- En lugar de ver solo el total del trimestre 1 (por ejemplo, \$500 en ventas), ahora puedes ver:

- ¿Qué producto se vendió cada mes?
- ¿En qué mes tuvo más éxito cada producto?

👉 Por ejemplo:

- El Carrito tuvo su mejor mes en Octubre (\$280)
 - El Juego creció constantemente, llegando a \$380 en Diciembre
-

💡 ¿Por qué es un "Drill Down"?

Porque estás bajando de nivel de agregación:

- Nivel alto (resumen): ventas por trimestre
- Nivel bajo (detalle): ventas por mes

➡ Al hacer esto, obtienes más detalle, como si estuvieras "acercándote" a los datos.

👉 **Conclusión**

El drill down te permite:

- Analizar tendencias mes a mes
- Detectar picos o caídas en ventas
- Tomar decisiones más precisas (por ejemplo: "¿en qué mes promocionar más carritos?")

Operación 2: Roll Up (Resumir)

🎯 Objetivo

Subir un nivel en la jerarquía de datos, resumiendo información de un nivel detallado (como el mes) a uno más general (como el trimestre).

En este caso:

De mes → a trimestre, agrupando también por categoría.

📊 Operación en Pentaho

Usamos el paso:

Group by

- Group fields: trimestre, categoría
- Sum of ventas → as ventas_trimestre

✓ Resultado del Roll Up

1	Miniaturas	500
1	Juguetes	420
1	Mesa	180
2	Mesa	970
2	Juguetes	240
3	Miniaturas	210

3	Mesa	350
3	Juguetes	260
4	Miniaturas	280
4	Juguetes	310
4	Mesa	380

Interpretación

Este resultado muestra cuánto se vendió en cada categoría por trimestre.

Ejemplos:

- En el trimestre 1, las Miniaturas (Carritos) generaron \$500.
- El trimestre 2 fue el mejor para Juegos de mesa (\$970).
- Las Muñecas (Juguetes) tuvieron su mejor desempeño en el trimestre 4 (\$310).

¿Por qué es un "Roll Up"?

Porque estás resumiendo datos:

- De mes → trimestre
- De producto → categoría

 Es como alejarte con un telescopio para ver el panorama general.

 Roll Up = Subir de nivel, resumir, generalizar.

Operación 3: Slice (Cortar una rebanada)

🎯 Objetivo

Analizar una sola dimensión, "cortando" el cubo para ver solo una parte.

En este caso:

Queremos ver solo las ventas del mes de Mayo.

📊 Operación en Pentaho

1. Filter rows: mes = 'Mayo'
 2. Luego: Group by → producto, region → Suma de ventas → ventas_mayo
-

✓ Resultado del Slice

Carrito	Norte	170
Muñeca	Sur	240

📊 Interpretación

Solo estamos viendo lo que se vendió en Mayo, sin importar el trimestre ni el año.

📌 Información clave:

- En Mayo, las muñecas en el Sur fueron lo más vendido (\$240).
 - Solo se vendió un carrito en el Norte.
-

💡 ¿Por qué es un "Slice"?

Porque has cortado el cubo por una dimensión (el mes), como si cortaras una rebanada de pan.

➡ El resto del cubo sigue existiendo, pero solo estás viendo una parte.

- 📌 Slice = Filtrar por un solo valor de una dimensión.

Operación 4: Dice (Cortar en varias dimensiones)

🎯 Objetivo

Hacer un "corte profundo" en varias dimensiones al mismo tiempo.

En este caso:

Queremos ver solo:

- mes = 'Mayo'
- region = 'Norte'
- producto = 'Carrito'

💻 Operación en Pentaho

1. Filter rows con condición:

1

[mes] = 'Mayo' AND [region] = 'Norte' AND [producto] = 'Carrito'

2. No necesitas agrupar, porque solo hay una fila que cumple.

✅ Resultado del Dice

9	Carrito	Miniaturas	Norte	Mayo	2	2023	170	7
---	---------	------------	-------	------	---	------	-----	---

📊 Interpretación

Este es el dato más específico posible en nuestro cubo:

- Solo un tipo de producto
- En una región específica

- En un solo mes

👉 Es como buscar una celda exacta en una tabla gigante.

💡 ¿Por qué es un "Dice"?

Porque cortaste el cubo en 3 dimensiones a la vez, como si lanzaras 3 cuchillos y solo quedara un cubito.

👉 Dice = Filtrar por múltiples dimensiones simultáneamente.

📊 Resumen Comparativo de las 4 Operaciones

Drill Down	Más detalle	De trimestre → mes	16 filas agrupadas por mes y producto
Roll Up	Más resumen	De mes → trimestre	11 filas resumidas por trimestre y categoría
Slice	Una dimensión	Solo "Mayo"	2 filas (ventas en Mayo)
Dice	Varias dimensiones	"Mayo + Norte + Carrito"	1 fila exacta