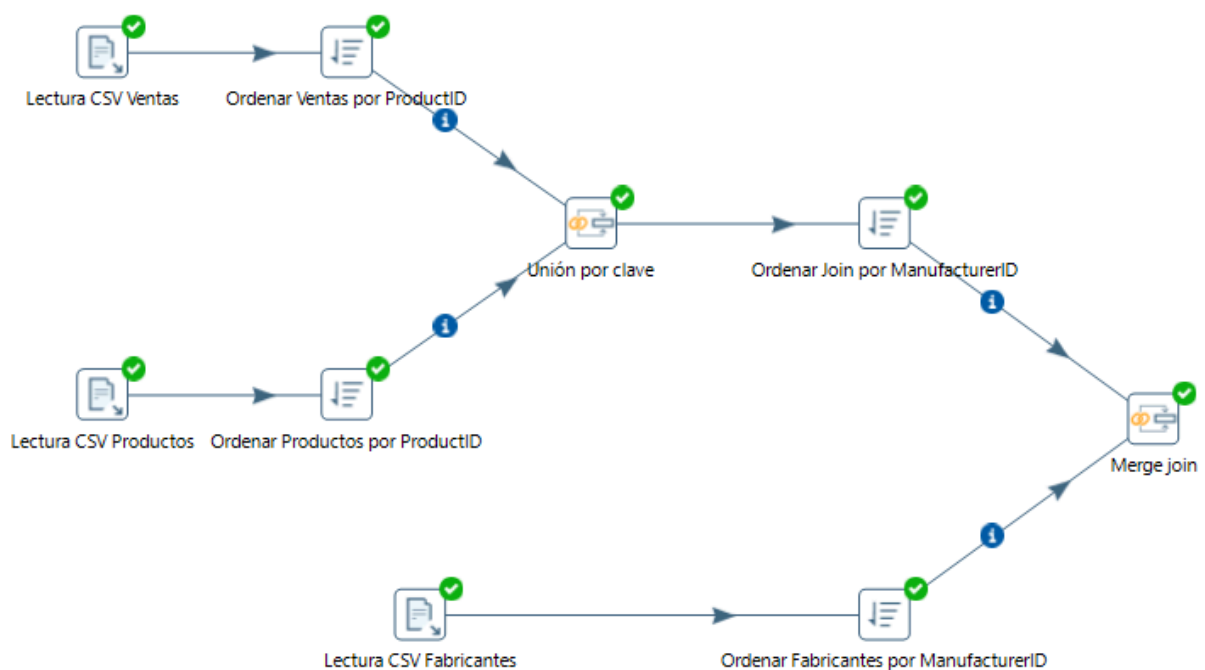


## Ejercicio 5 - Informe fabricantes en S3

A partir del Ejercicio 3 que hemos realizado previamente, vamos a generar un informe sobre ventas de los fabricantes, y a persistir el informe en S3.

### Unir fabricantes

Así pues, vamos a abrir el Ejercicio 3, y a partir de ahí, vamos a fusionar los datos con los de **fabricantes** (fichero CSV adjunto) para obtener el nombre de éstos. Para ello, creamos una nueva lectura de CSV, ordenación y posterior *Merge*. Cabe destacar que antes de hacer el segundo *join*, debemos ordenar ambas entradas por la clave de unión (**ManufacturerID**):



*Caso de Uso 4 - Merge fabricantes*

### Aplicar fórmulas

Tras unir los datos de las ventas con los fabricantes, para poder preparar el informe mediante group-by, del mismo modo que antes, necesitamos ordenar los datos (en este caso por **ManufacturerID**).

En el informe queremos mostrar para cada fabricante, además de su código y nombre (campos de agrupación), queremos obtener el total de unidades vendidas y la recaudación total de dicho fabricante. Para ello creamos una nueva agrupación.

**Group by**

---

Step name:  Agrupamos por Fabricante

Include all rows? ☐

Temporary files directory:  %%java.io.tmpdir%%

TMP-file prefix:  grp

Add line number, restart in each group ☐

Line number field name:

Always give back a result row ☐

The fields that make up the group:

#	Group field
1	ManufacturerID
2	Manufacturer

Aggregates :

#	Name	Subject	Type
1	Unidades	Units	Sum
2	Recaudacion	Revenue	Sum

< >

? Help   OK   Cancel

### Caso de Uso 4 - Agrupamos por fabricante

Además, queremos que nos muestre un nuevo campo que muestre el precio medio obtenido de dividir la recaudación obtenida entre el total de unidades. Para ello, dentro de la categoría *Transform*, elegimos el paso *Calculator*. Una vez abierto el diálogo para editar el paso, si desplegamos la columna *Calculation* puedes observar todas las posibles operaciones y transformaciones (tanto numéricas, como de campos de texto e incluso fecha) que podemos realizar. En nuestro caso, sólo necesitamos la división entre *A* y *B*:

Calculator

Step name  
Precio Medio

☒ Throw an error on non existing files

Fields:

#	New field	Calculation	Field A	Field B	Field C	Value type
1	PrecioMedio	A / B	Recaudacion	Unidades		

Help OK Cancel

Caso de Uso 4 - Fórmula entre dos campos

## Guardar en S3

Para almacenar los datos en S3, primero crearemos un *bucket* público (en mi caso lo he denominado `ejercicio5pdi`) y para facilitar el trabajo, vamos a crear una política que permita todas las operaciones:

```
{
  "Version": "2012-10-17",
  "Id": "Policy1635323698048",
  "Statement": [
    {
      "Sid": "Stmt1635323796449",
      "Effect": "Allow",
      "Principal": "*",
      "Action": "s3:*",
      "Resource": "arn:aws:s3:::ejercicio5pdi"
    }
  ]
}
```

El siguiente paso es configurar las credenciales de acceso en nuestro sistema. Recuerda que lo haremos mediante las variables de entorno

(AWS\_ACCESS\_KEY\_ID, AWS\_SECRET\_ACCESS\_KEY y AWS\_SESSION\_TOKEN) o almacenando las credenciales en el archivo ~/.aws/credentials.

Finalmente, añadimos el paso *S3 file output* indicando:

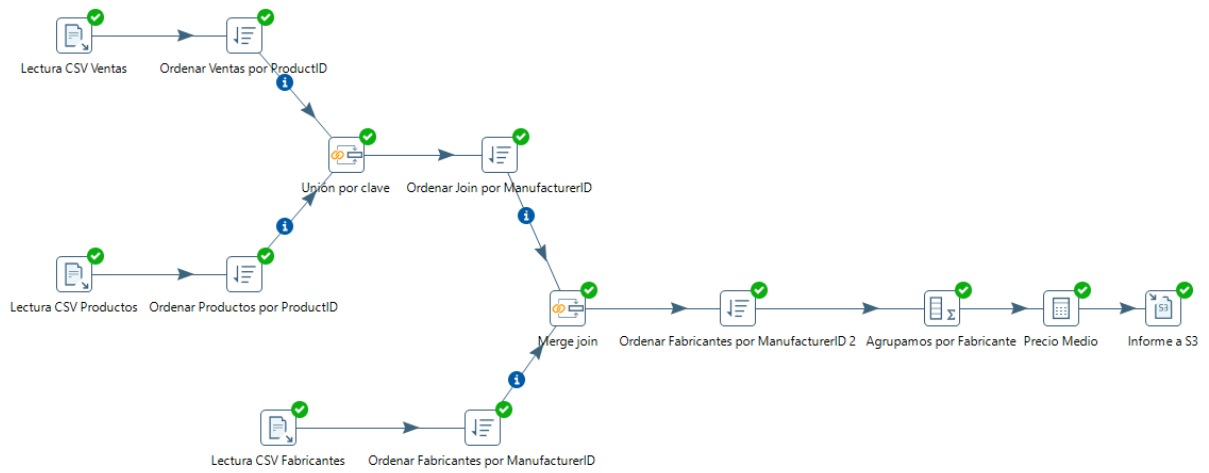
- La ruta del archivo, la cual se indica de forma similar a `s3n://s3n/bucket/nombreArchivo`. En nuestro caso, hemos utilizado el nombre `s3n://s3n/ejercicio5pdi/informeFabricantes`.
- La extensión: nosotros hemos elegido el formato csv

The screenshot shows the 'S3 file output' configuration window. The 'Step name' is 'Informe a S3'. The 'File' tab is selected, showing the 'Filename' field with the value 's3n://s3n/ejercicio5pdi/informeFabricantes' and a 'Browse...' button. Below this are several checkboxes: 'Do not create file at start', 'Accept file name from field', 'Include stepnr in filename?', 'Include partition nr in filename?', 'Include date in filename?', 'Include time in filename?', and 'Specify Date time format'. The 'Extension' is set to 'csv'. There is a 'Date time format' dropdown menu and a 'Show filename(s)...' button. The 'Add filenames to result' checkbox is checked. At the bottom, there are 'Help', 'OK', and 'Cancel' buttons.

#### *Caso de Uso 4 - Salida a S3*

Una vez todo unido, tras ejecutarlo podremos acceder a nuestra consola de AWS y comprobar cómo ha aparecido el fichero.

En resumen, en este caso de uso hemos generado la siguiente transformación:



### *Caso de Uso 4 - Resultado final*