Recurso complementario

Ejemplo 3: integración automatizada de múltiples fuentes

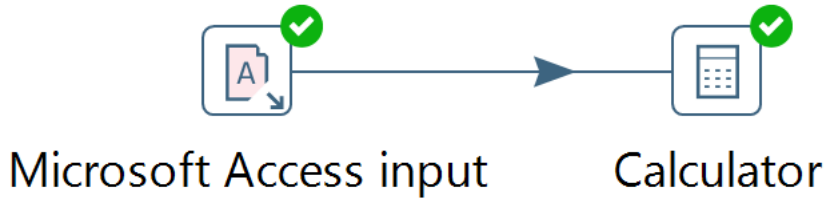
El tercer ejemplo de aplicación consiste en desarrollar un proceso *ETL* para leer datos desde un archivo de base de datos de Microsoft Access (.ccdb) y enviar los datos hacia dos componentes particulares: la calculadora y el concatenador de campos. La calculadora funciona como un paso particular que puede transformar matemáticamente el valor de ciertos datos en campos de entrada, mientras que el concatenador puede unir automáticamente los valores de dos o más campos de datos (en un único nuevo campo, a la salida).

Para construir la nueva transformación (que llamaremos **transformación03**), se deben cumplir los siguientes pasos:

1. Crea la transformación propiamente tal, incluyendo los pasos indicados en el diagrama de la figura 1:

Figura 1.

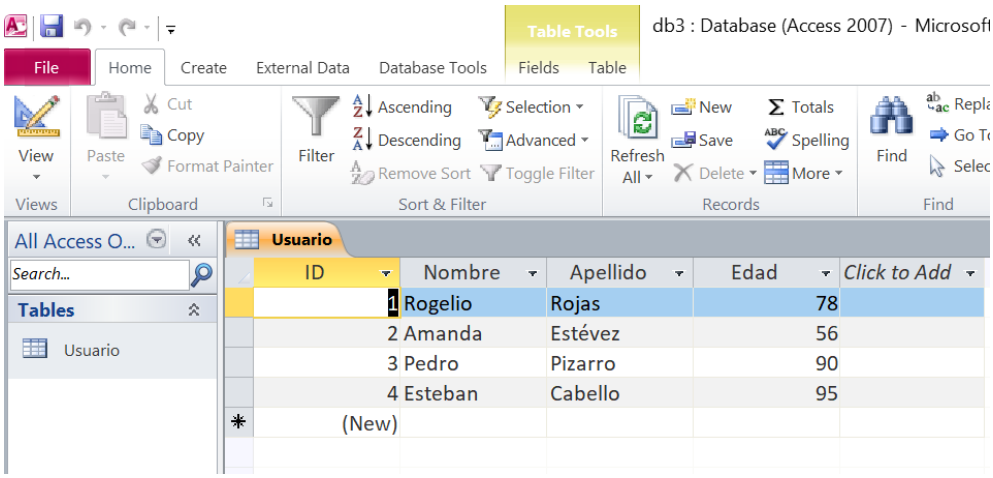
Esquema de la transformación a construir en el ejercicio 3



2. Crea un archivo de bases de datos de Microsoft Access en algún directorio vacío, con nombre **db3.accdb**, que contenga una tabla relacional, de nombre Usuario, con los datos mostrados en la figura 2

Figura 2.

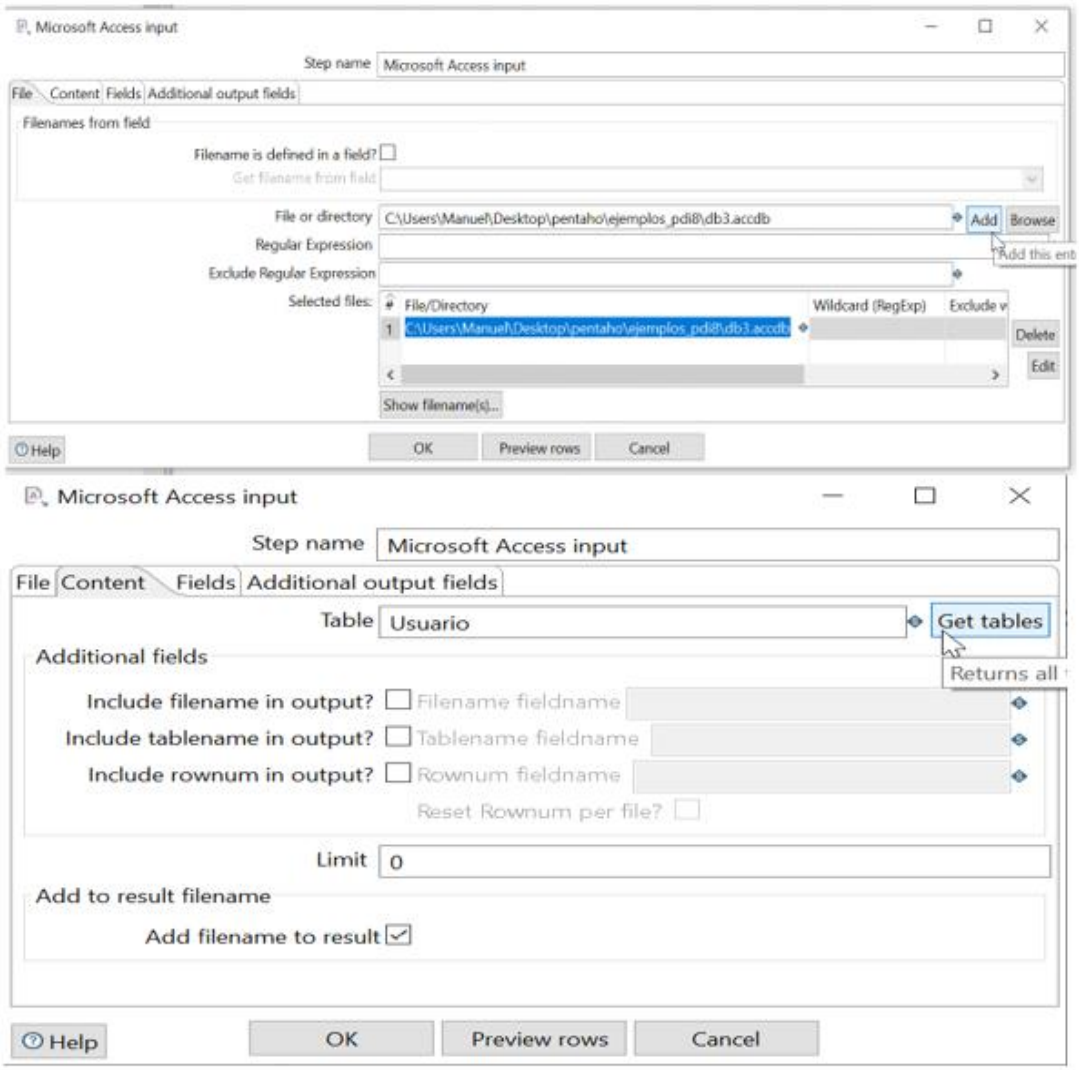
Contenidos del archivo de bases de datos de Microsoft Access



3. Dentro de la configuración del paso, indica el nombre de la tabla a leer (figura 3).

Figura 3.

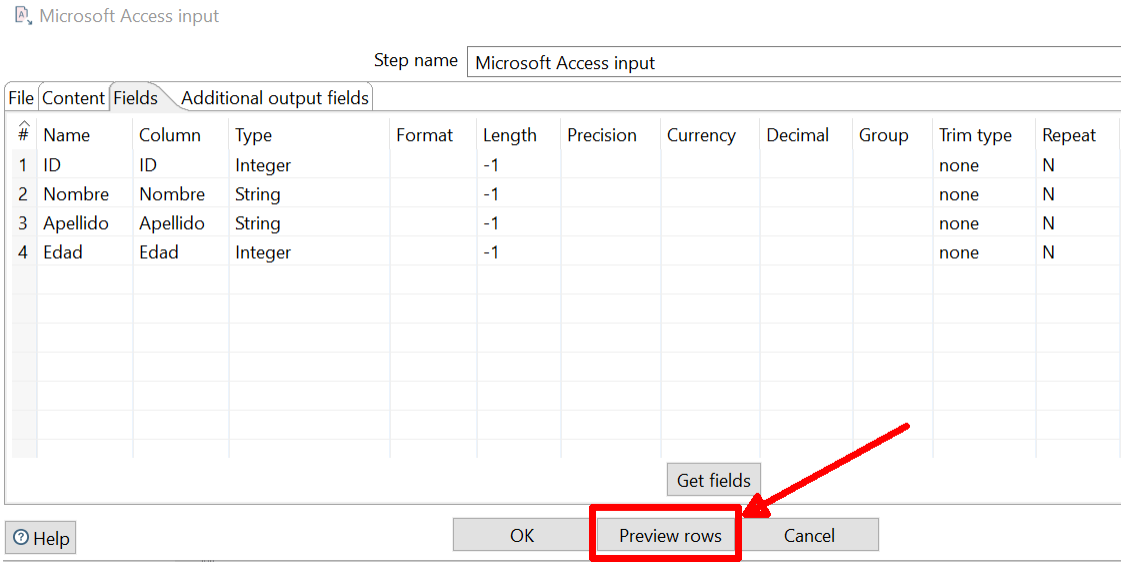
*Configuración general del paso **Microsoft Access input** (arriba) y configuración del nombre de tabla de base de datos a utilizar (abajo)*



4. Lee los nombres de columna y los datos que contiene la tabla Usuario a leer con ETL (figura 4), usando los botones **Get fields** y **Preview Rows** dentro de la pestaña de configuración **Fields**.

Figura 4.

*Uso de los botones **Get fields** y **Preview Rows** para leer los campos y datos de la tabla relacionan Usuario*

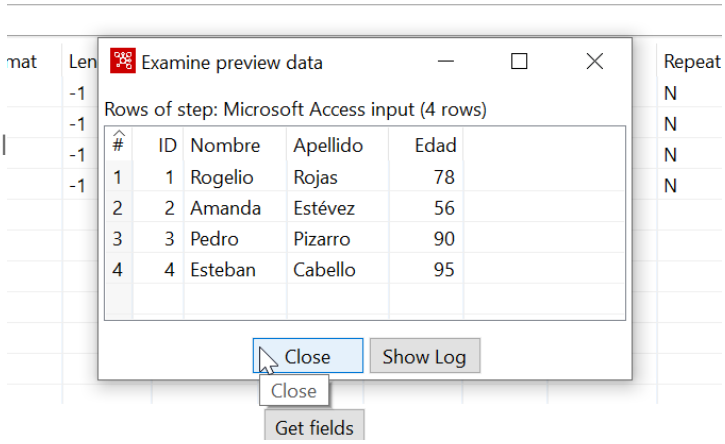


5. Al leer los datos desde la tabla (como se indicó en el paso anterior) para corroborar que la configuración completa esté correcta, mostrará lo indicado en la figura 5.

Figura 5.

Datos leídos desde la tabla relacional Usuario

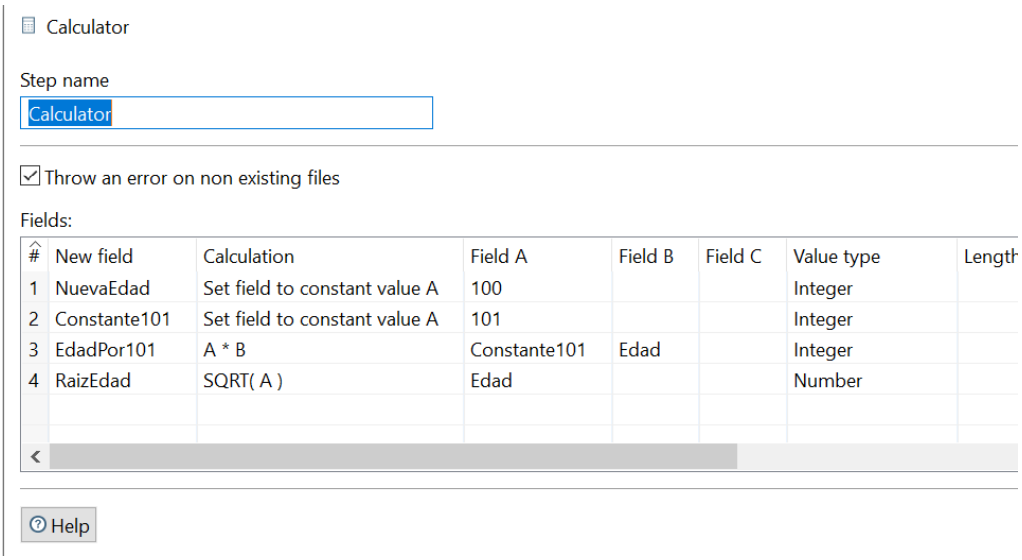
6. Ahora, configura el segundo paso,



denominado **Calculator**, con la configuración de la figura 6. Observa que la tabla de campos (*fields*) permitirá crear nuevos campos a la salida con los resultados de los cálculos matemáticos. Observa también que se pueden crear constantes, usar fórmulas predefinidas y combinar valores de varios campos (A, B y C, mapeados a las columnas de datos que requieran) para hacer cálculos más complejos.

Figura 6.

*Configuración general para el paso **Calculator** que realizará algúnos cálculos matemáticos simples sobre los datos de la tabla*

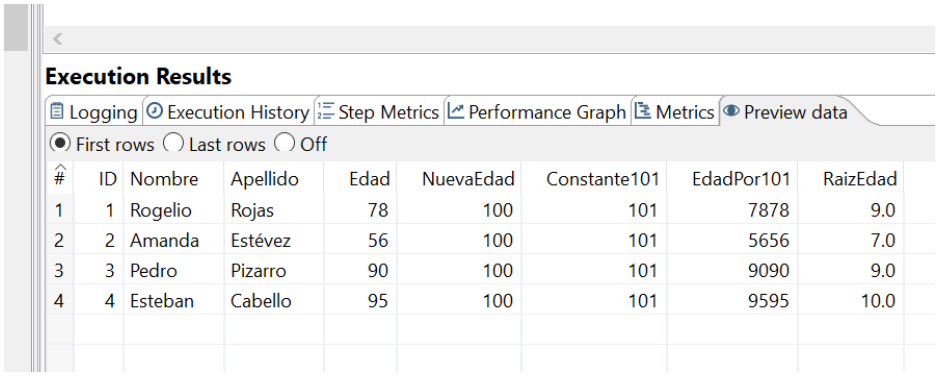


#	New field	Calculation	Field A	Field B	Field C	Value type	Length
1	NuevaEdad	Set field to constant value A	100			Integer	
2	Constante101	Set field to constant value A	101			Integer	
3	EdadPor101	A * B	Constante101	Edad		Integer	
4	RaizEdad	SQRT(A)	Edad			Number	

7. Luego, ejecuta la transformación, usando el botón **Run** correspondiente. Como te habrás dado cuenta, no hay un paso que escriba los datos algún archivo o base de datos de salida, sin embargo, luego de que termines la ejecución, si haces clic sobre el paso **Calculator**, revisa el panel inferior de la interfaz gráfica (llamada **Execution Results**) y selecciona la pestaña **Preview Results**, para ver los resultados de la transformación completa (figura 7).

Figura 7.

Resultados de la ejecución anterior



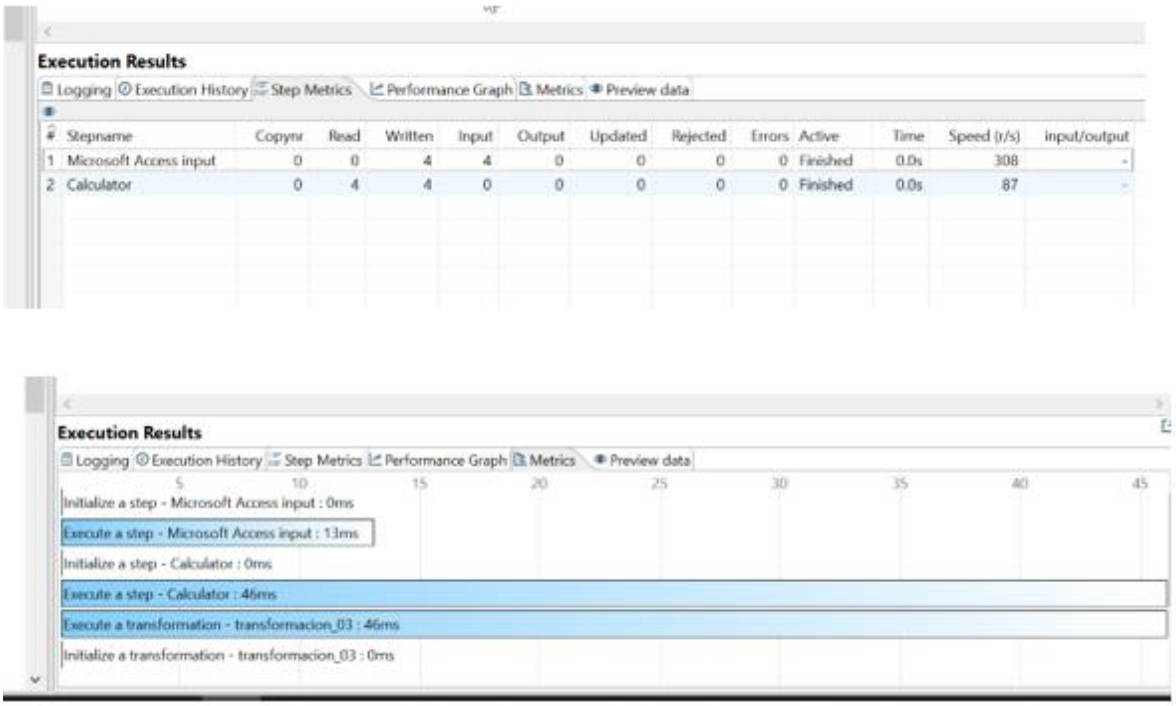
#	ID	Nombre	Apellido	Edad	NuevaEdad	Constante101	EdadPor101	RaizEdad
1	1	Rogelio	Rojas	78	100	101	7878	9.0
2	2	Amanda	Estévez	56	100	101	5656	7.0
3	3	Pedro	Pizarro	90	100	101	9090	9.0
4	4	Esteban	Cabello	95	100	101	9595	10.0

8. De esta forma, cada vez que hagas clic en un paso de una transformación larga, podrás ver los datos intermedios generados para cada paso, permitiendo corroborar que el proceso *ETL* funciona como lo requieres.

9. Finalmente, en el mismo panel inferior **Execution Results**, puedes también leer las métricas de cantidad de registros leídos/escritos en este paso (pestaña **Step Metrics**) y el tiempo de ejecución (pestaña **Metrics**), así como otra información relevante al proceso completo (figura 8).

Figura 8.

Panel de información de métricas para el paso actual (arriba) y para la transformación completa en segundos (abajo)



Para finalizar, es importante mencionar que este ejemplo permite comprender el valor de la integración automatizada de múltiples fuentes dentro de un proceso ETL, destacando cómo Pentaho Data Integration facilita no solo la extracción de datos desde bases relacionales, sino también su transformación mediante operaciones aritméticas y combinaciones lógicas a través de componentes especializados como *Calculator* y *Concatenator*. La posibilidad de visualizar datos intermedios y métricas de ejecución refuerza el control y la validación del proceso, promoviendo buenas prácticas en la gestión de datos y fortaleciendo competencias analíticas esenciales para el diseño de soluciones de inteligencia de negocios y apoyo a la toma de decisiones en entornos organizacionales complejos.