Recurso complementario

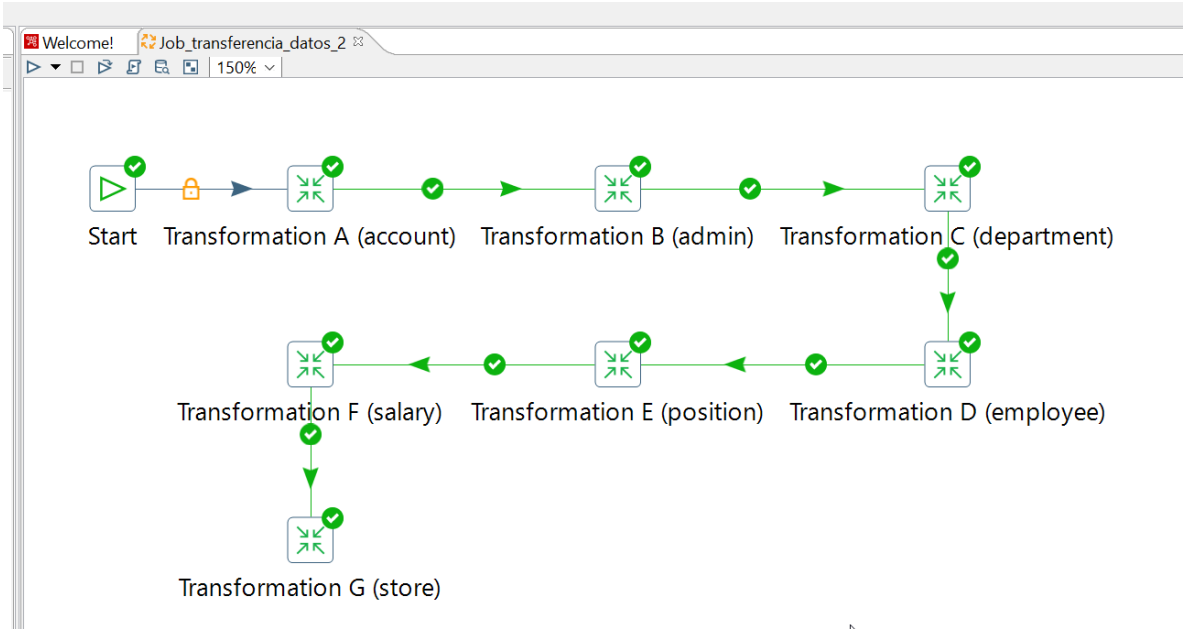
Ejemplo: integración del almacén de datos a partir de procesos *ETL* automatizados

La idea en este nuevo ejemplo es que continúe con el procesamiento de datos del ejemplo anterior (ejemplo 1: automatización del flujo *ETL* mediante trabajos en Pentaho) para lograr completar la copia de datos de todas las tablas relacionales del archivo **FoodMart.mdb** hacia la base de datos **foodmart** en MariaDB. El trabajo para este ejemplo tiene la forma mostrada en la figura 1. Procura crear todas las transformaciones extras que faltan, probar que los datos leídos sean los datos correctos y que el trabajo total funciona sin problemas.

Considera eso sí sucede lo siguiente: cada vez que vuelvas a ejecutar el trabajo completo, **todos los datos desde FoodMart.mdb** se leerán y se escribirán en la base de datos **foodmart** de MariaDB, por lo que deberás hacer algún proceso de limpieza manual de tablas (eliminar todas las filas de datos) antes de ejecutar el trabajo nuevamente, o los datos se escribirán duplicados y se mantendrán almacenados de forma indefinida en la base de datos.

Figura 1.

Trabajo (job) completo de Pentaho Data Integration para migrar todos los datos del archivo ***FoodMart.mdb*** *hacia MariaDB*



Fuente: elaboración propia.

En conclusión, este ejercicio permite afianzar los conceptos de automatización *ETL* en Pentaho Data Integration, destacando la importancia de validar los datos migrados y gestionar adecuadamente la limpieza de tablas para evitar duplicidades. Una correcta ejecución del trabajo garantiza la integridad y confiabilidad de la información cargada en MariaDB, fortaleciendo las buenas prácticas en integración de datos.