

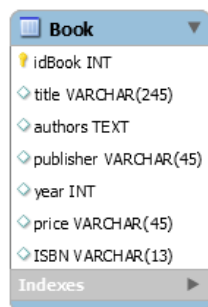
Mapeado LAV

Práctica de integración de datos (v2.0)

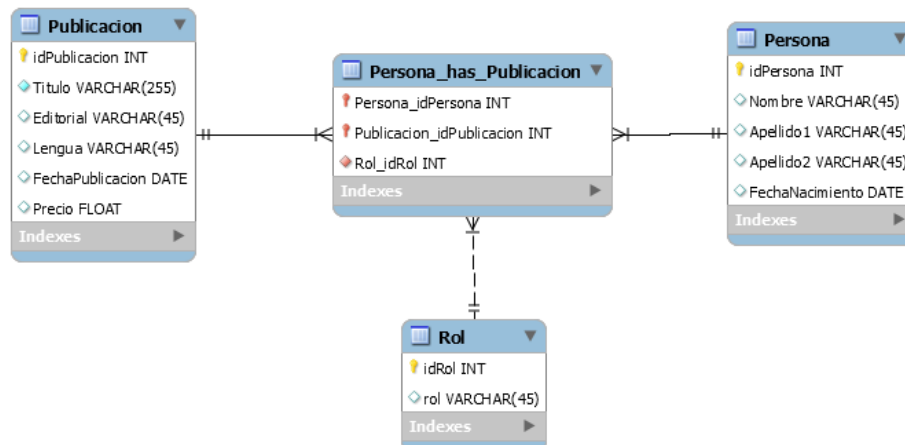
Se quieren agregar los datos sobre disponibilidad de libros procedentes de distintas librerías (fuentes). Cada fuente es independiente y el objetivo es implementar un proceso ETL que tome los datos de cada una de ellas y los integre, con los necesarios cambios debidos a las diferencias de esquema, en la base de datos objetivos que usará el esquema global.

Se disponen de las siguientes fuentes:

- Una librería británica que usa el siguiente esquema con una sola tabla:



- Una librería española con un esquema relacional algo más complejo:



- Y una librería de EEUU de la que se han extraído datos en formato JSON (archivo USA_books.json). No es necesario implementar una base de datos relacional para esta fuente si no se desea, se puede hacer como almacenamiento temporal de los datos antes de transformarlos si resulta más cómodo así.

Se ha propuesto el siguiente esquema global para la base de datos objetivo:



1. Implementar un proceso de ETL para cada una de las fuentes que debe:
 - a. Extraer los datos necesarios de la fuente (consulta).
 - b. Efectuar las transformaciones necesarias (incluida limpieza) de los datos extraídos.
 - c. Insertar los datos en la base de datos objetivo o global.
2. Explicar las decisiones tomadas.
3. Discutir la idoneidad del esquema global propuesto, y sugerir posibles mejoras.