**1.Data Set Description – Gravity Book Store OLTP Dataset**

The dataset captures the crucial information needed to oversee daily business operations and serves as an example of an online bookstore's operational data. It contains thorough details about books, their publishers, authors, and the languages in which they are published. The collection also keeps track of shipping details, client records, and the complete order lifecycle, from placement to delivery.  
  
Every element of the dataset represents a practical facet of bookshop operations. For example, it preserves the history and status updates of every order over time, allows various delivery options, and maintains track of customers' addresses and related statuses. The information is organized to provide a comprehensive picture of all the parties and transactions involved in running a sizable online bookshop.

**Tables:**

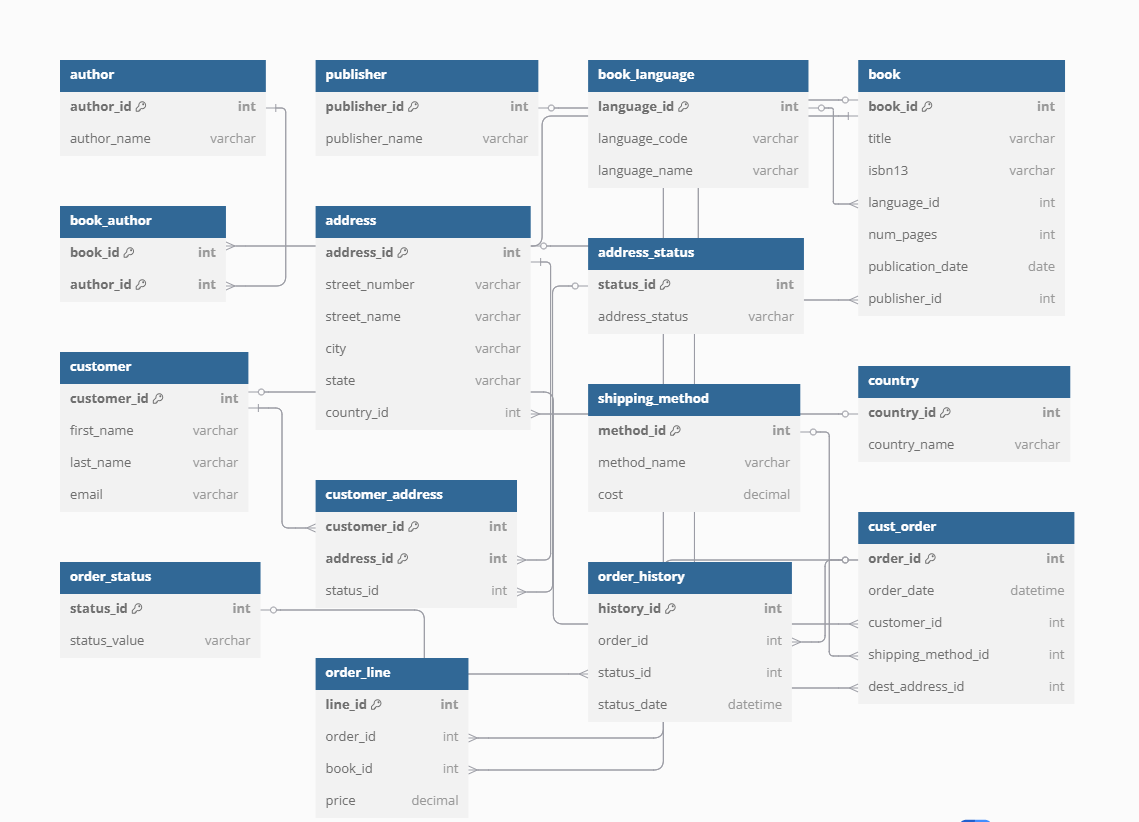
Book - Keeps track of each book's primary information, such as its title, ISBN, cost, and mentions of its publisher and language.  
  
author - Provides the author's name, biography, and other professional and personal information.

book\_author - junction table that associates books with one or more authors.  
  
Publisher - Contains details on the publishing houses that do the book releases.  
  
book\_language - Provides a list of the languages used in book publications.

Customer - Keeps track of names and contact information for patrons of the bookshop.  
  
address - Provides actual addresses linked to clients or shipping information.  
  
address\_status - Shows if an address is active, inactive, or removed at the moment.  
  
country - The nations used in the address table are listed in this reference table.

customer\_order - Indicates a customer's order, along with the shipping method and order status as of right now.  
  
order\_line - Lists each item in a customer's order, including the books they bought and how much of each.  
  
shipping\_method - Provides a list of the available shipping options (such as standard and express) together with the related metadata.

order\_status - Specifies the many phases that an order may experience, such as pending, shipped, and delivered.  
  
order\_history - Enables order monitoring over time by keeping a record of each order's status changes.



**2.Preparation of data sources**

|  |  |
| --- | --- |
| **Table Name** | **Source Type** |
| Author | DB |
| Publisher | DB |
| Book\_language | DB |
| Book | DB |
| Book\_author | DB |
| Address\_status | DB |
| Country | CSV |
| Address | CSV |
| Customer | DB |
| Customer\_order | DB |
| Shipping\_method | DB |
| Order\_line | DB |
| Order\_history | DB |
| Order\_status | DB |

**3.Solution Architecture**

**Source**   
  
the original information gathered from the bookshop system and stored in OLTP databases, CSV files, or Excel files.  
  
comprises client information, personnel data, book inventories, and sales transactions.

**ETL (SSIS)**  
  
Extract: Gathers information from several source systems.  
  
Transform: Converts data types, fixes naming errors, and cleans, normalises, and integrates the data.  
  
The converted data is first loaded into the staging area and subsequently into the data warehouse.

**Staging**  
  
SQL Server temporary storage for raw extracted data prior to treatment.  
  
guarantees data integrity and acts as a fallback for diagnosing ETL errors.

**Data Ware House (SQL Server)**  
  
A star schema was used to design the centralised repository.  
  
includes dimension tables (such as Books\_Dim, Customer\_Dim, Staff\_Dim, and Time\_Dim) and fact tables (such as Sales\_Fact).  
  
supports analytical queries and historical data.