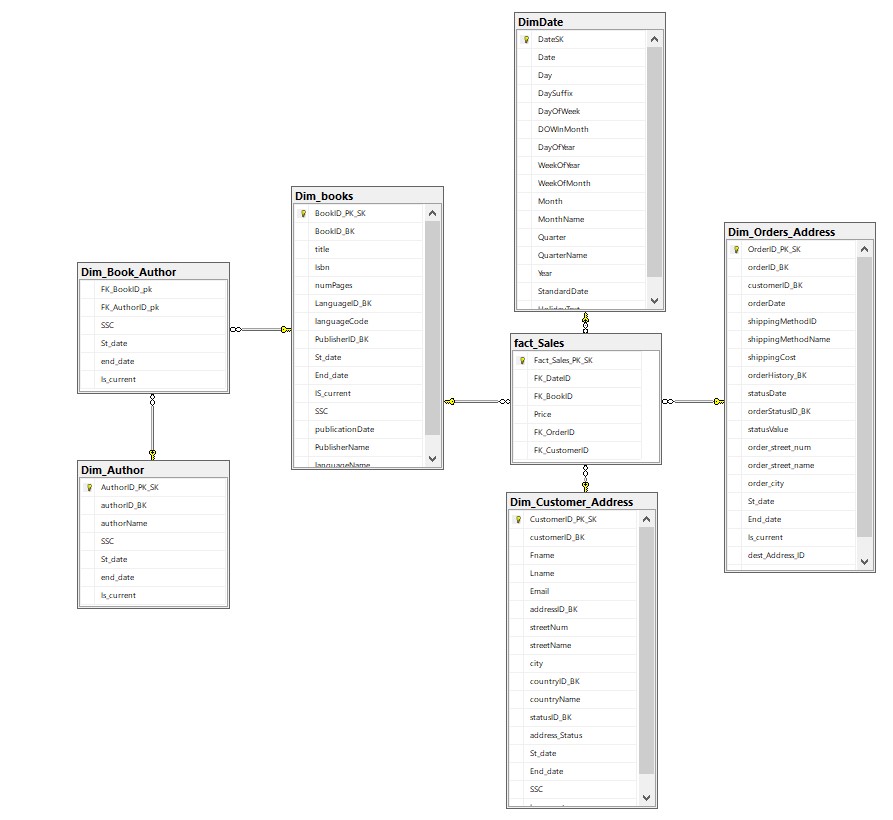
Gravity Books DWH Documentation

Team Members:

* Belal Mohammad Eltaiby
* Haidy Hossam
* Mohammad Ghareeb
* Medhat Sobhy

DHW Schema Diagram:



**DWH DDL Statement & Mapping:**

* **Link to the schema mapping Excel sheet:**

[**https://docs.google.com/spreadsheets/d/1k30Djd6lgteSnprG79aXEDHXA1-8ZTdH/edit?usp=sharing&ouid=102121981167089002487&rtpof=true&sd=true**](https://docs.google.com/spreadsheets/d/1k30Djd6lgteSnprG79aXEDHXA1-8ZTdH/edit?usp=sharing&ouid=102121981167089002487&rtpof=true&sd=true)

* **Link to the DDL SQL file:**

https://drive.google.com/file/d/1ngxah20lyObCiIy2gvn0Y0nQVKmDIRY\_/view?usp=sharing

**DWH Schema Approach:**

* **Snowflake Schema:**

Due to the Many-Many relationship between Books and Authors Tables, We decided to normalize this part of the schema.

Schema Components:

1. Dim\_Customer\_Address
2. Dim\_Order\_Address
3. Dim\_books
4. Dim\_Authors
5. Dim\_Book\_Author (Bridge)
6. Fact\_sales (fact table)

**Methods for integrity:**

* Adding attributes: Start\_date, End\_date, Is\_current and SSc will maintain the integrity and the ability for Slowly Changing dimensions in addition to adding data sources without fear of Duplicates (Surrogate key also helps in this).
* Incremental loading of data into Fact table.
* Adding Date dimension to track historical changes.

SSIS Packages:

1. **Dim\_books:**

*Source Data:*

select B.book\_id, B.title, B.isbn13,B.num\_pages,B.publication\_date,

BL.language\_id,BL.language\_name,BL.language\_code,

PB.publisher\_id,PB.publisher\_name

from dbo.book B

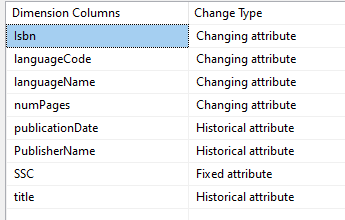
left join book\_language BL

on B.language\_id = BL.language\_id

left join publisher PB

on B.publisher\_id = PB.publisher\_id

*Types of changes:*

**

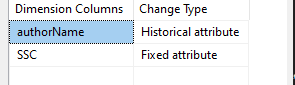
1. **Dim\_Authors:**

*Source Data:*

select author\_id, author\_name

from author

*Types of changes:*

**

1. **Dim\_book\_Author:**

*Source Data:*

SELECT

DB.BookID\_PK\_SK,

DA.AuthorID\_PK\_SK

FROM

[DESKTOP-PKR2U6N].[gravity\_books].[dbo].[book\_author] GB\_Src

JOIN Dim\_books DB ON GB\_Src.book\_id = DB.BookID\_BK

JOIN Dim\_Author DA ON GB\_Src.author\_id = DA.authorID\_BK;

*Types of Changes:*

**

1. **Dim\_Customer\_Address:**

*Source Data:*

select C.customer\_id,C.first\_name,C.last\_name,C.email,

ADDR.address\_id,ADDR.street\_number,ADDR.street\_name,ADDR.city,

Cntry.country\_id,Cntry.country\_name,

ADDS.status\_id,ADDS.address\_status

from customer C

left join customer\_address CAdd

on C.customer\_id = CAdd.customer\_id

left join address\_status ADDS

on CAdd.status\_id = ADDS.status\_id

left join address ADDR

on CAdd.address\_id = ADDR.address\_id

left join country Cntry

on ADDR.country\_id = Cntry.country\_id

order by C.customer\_id asc

*Types of Changes:*

1. **Dim\_Order\_Address:**

*Source Data:*

select Cord.order\_id,cord.order\_date,

ShipM.method\_id,ShipM.method\_name,ShipM.cost,

OrdHis.history\_id,OrdHis.status\_date,

OrdSt.status\_id,OrdSt.status\_value,

Cus.customer\_id,

addr.street\_number,addr.street\_name,addr.city,Addr.address\_id

from cust\_order Cord

left join shipping\_method ShipM

on Cord.shipping\_method\_id = ShipM.method\_id

left join order\_history OrdHis

on Cord.order\_id = OrdHis.order\_id

left join order\_status OrdSt

on OrdHis.status\_id = OrdSt.status\_id

left join customer Cus

on Cord.customer\_id = Cus.customer\_id

left join address Addr

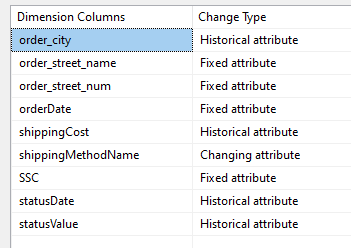
on Cord.dest\_address\_id = Addr.address\_id

left join country Cntry

on Addr.country\_id = Cntry.country\_id

order by order\_id asc

*Types of changing:*



1. Fact\_sales Package:

*Source data:*

select Cord.order\_id,cord.order\_date,

Cus.customer\_id,

B.book\_id,

ordlin.price

from cust\_order Cord

left join shipping\_method ShipM

on Cord.shipping\_method\_id = ShipM.method\_id

left join customer Cus

on Cord.customer\_id = Cus.customer\_id

left join order\_line ordlin

on Cord.order\_id = ordlin.order\_id

left join book B

on ordlin.book\_id = B.book\_id