Gravity Books DWH Documentation

DHW Schema Diagram:

DimDate
- DateSK
- Date
- Day
- DaySuffix
- DayOfWeek
- DOWInMonth
- DayOfYear
- WeekOfYear
- WeekOfMonth
- Month
- MonthName
- Quarter
- QuarterName
- Year
- StandardDate

Dim_books
- BookID_PK_SK
- BookID_BK
- title
- Isbn
- numPages
- LanguageID_BK
- languageCode
- PublisherID_BK
- St_date
- End_date
- iS_current
- SSC
- publicationDate
- PublisherName
- languageName

Dim_Book_Author
- FK_BookID_pk
- FK_AuthorID_pk
- SSC
- St_date
- end_date
- Is_current

Dim_Author
- AuthorID_PK_SK
- authorID_BK
- authorName
- SSC
- St_date
- end_date
- Is_current

fact_Sales
- Fact_Sales_PK_SK
- FK_DateID
- FK_BookID
- Price
- FK_OrderID
- FK_CustomerID

Dim_Orders_Address
- OrderID_PK_SK
- orderID_BK
- customerID_BK
- orderDate
- shippingMethodID
- shippingMethodName
- shippingCost
- orderHistory_BK
- statusDate
- orderStatusID_BK
- statusValue
- order_street_num
- order_street_name
- order_city
- St_date
- End_date
- Is_current
- dest_Address_ID

Dim_Customer_Address
- CustomerID_PK_SK
- customerID_BK
- Fname
- Lname
- Email
- addressID_BK
- streetNum
- streetName
- city
- countryID_BK
- countryName
- statusID_BK
- address_Status
- St_date
- End_date
- SSC

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

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:

Dimension Columns	Change Type
Isbn	Changing attribute
languageCode	Changing attribute
languageName	Changing attribute
numPages	Changing attribute
publicationDate	Historical attribute
PublisherName	Historical attribute
SSC	Fixed attribute
title	Historical attribute

2) Dim_Authors:

Source Data:

> select author_id, author_name
> from author

Types of changes:

Dimension Columns	Change Type
authorName	Historical attribute
SSC	Fixed attribute

3) 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;
```
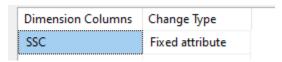
Types of Changes:

Dimension Columns	Change Type
SSC	Fixed attribute

4) 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
```

5) 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:

Dimension Columns	Change Type
order_city	Historical attribute
order_street_name	Fixed attribute
order_street_num	Fixed attribute
orderDate	Fixed attribute
shippingCost	Historical attribute
shippingMethodName	Changing attribute
SSC	Fixed attribute
statusDate	Historical attribute
statusValue	Historical attribute

6) 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
```

Power Bi Dashboard:

https://app.powerbi.com/links/K4bo-gCk1z?ctid=a2c31985-cc3b-4e19-8fa2-59fa488f0c27&pbi_source=linkShare