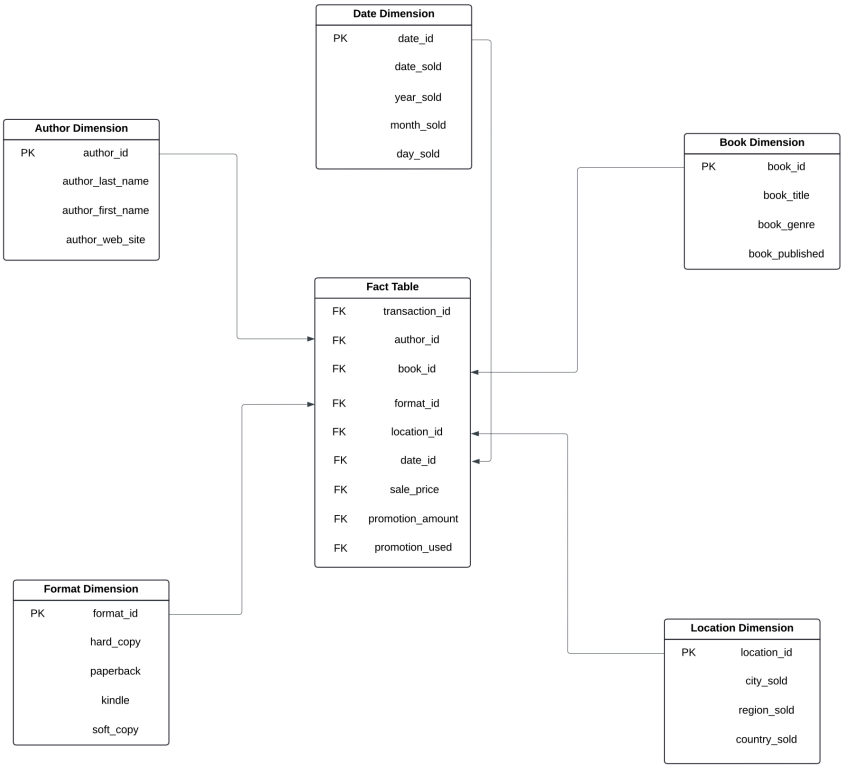
Georges Bitar
Professor Parsian
OMIS 113
29 May 2024

Building a Star Schema for a Bookseller

1. What is a Star Schema
    a. A star schema is a multidimensional data model that is used to organize data in a database so that it is easier to understand and analyze. Star schemas can be applied to data warehouses, databases, data marts, and other tools. They are efficient at storing data, maintaining history, and updating data by reducing the duplication of repetitive business definitions; these benefits make the process of aggregating and filtering data in a data warehouse to be quicker.
2. How would you design a Star Schema for these given tables



a.

3. What are your ETL programs to create Star Schema from these 2 operational tables
    a. Data is first extracted from the operation/transaction-based tables: “Sales” and “Books”.Then, data is transformed/loaded onto the fact table and the dimension tables: “Book Dimension”,”Author Dimension”, “Date Dimension”, “Location Dimension”, Format Dimension”. These tables can be created by loading information onto the dimension tables and their primary keys will be information as foreign keys for the fact table.
4. What are 3 Business Intelligence SQL queries to give an insight to book sales:
    1. What are the top 5 cities that have the most book purchases over the last 3 years?
    SELECT L.city_sold, COUNT(*) AS total_purchases
    FROM FactTable F

```
JOIN LocationDimension L ON F.location_id = L.location_id
JOIN DateDimension D ON F.date_id = D.date_id
WHERE D.year_sold >= YEAR(CURRENT_DATE) - 3
GROUP BY L.city_sold
ORDER BY total_purchases DESC
LIMIT 5;
```

2. What are the top 3 most popular authors this year based on purchases?

```
SELECT year_sold, author_first_name, author_last_name, total_purchases
FROM (
    SELECT D.year_sold, A.author_first_name, A.author_last_name, COUNT(*) AS total_purchases, ROW_NUMBER() OVER (PARTITION BY D.year_sold ORDER BY COUNT(*) DESC) AS rank
    FROM FactTable F
    JOIN AuthorDimension A ON F.author_id = A.author_id
    JOIN DateDimension D ON F.date_id = D.date_id
    GROUP BY D.year_sold, A.author_first_name, A.author_last_name
) AS ranked_authors
WHERE rank <= 3
ORDER BY year_sold, rank;
```

3. What are the top 5 cities that buy kindle books?

```
SELECT l.city_sold, COUNT(f.transaction_id) AS total_purchases
FROM Fact_Table f
JOIN Location_Dimension l ON f.location_id = l.location_id
JOIN Format_Dimension fr ON f.format_id = fr.format_id
WHERE fr.kindle = TRUE
GROUP BY l.city_sold
ORDER BY total_purchases DESC
LIMIT 5;
```