DWH Assignment

Table of Contents

zzgl slowly changing dimensions

ACHTUNG: KAPITEL 3 Reihenfolge beachen!!!!!!!

pro DB: Diagramm und TSQL-Skripst!!!!
ETL: inkl. Mapping Sheet , SChedule und TSQL

1.1 Project

“Create a data warehouse for a bookstore so that it can extract knowledge from its data and use it to for further growth and development.”

1.2 Abstract - Requiment Engineering

A bookstore data warehouse is a centralized repository that combines data from various sources within a bookstore, such as point-of-sale transactions, customer information, and inventory data. The purpose of a bookstore data warehouse is to provide a single source of truth for data analysis and reporting, enabling bookstore managers to make informed decisions about operations, marketing, and sales. The data in a bookstore data warehouse is typically organized into a series of dimensional tables, which can be queried using SQL or other tools to extract insights and trends. For example, a bookstore manager might use a data warehouse to track sales by book genre, customer demographics, or store location. By analyzing this data, the manager can identify best-selling titles, target marketing efforts to specific customer segments, and optimize inventory levels to meet demand.

1.3 Motivation

There are several key motivations for implementing a bookstore data warehouse:

1. Improved decision making: By consolidating data from multiple sources and providing a single source of truth, a bookstore data warehouse can help managers make more informed decisions about operations, marketing, and sales.

2. Increased efficiency: A data warehouse can streamline the process of accessing and analyzing data, allowing managers to retrieve the information quickly and easily they need to make decisions.

3. Enhanced customer experience: By analyzing customer data, a bookstore can better understand their customers' needs and preferences, and tailor their marketing and sales efforts accordingly.

4. Increased profitability: By optimizing operations and sales strategies based on data analysis, a bookstore can improve its bottom line and drive profitability.

5. Competitive advantage: A data warehouse can give a bookstore a competitive advantage by providing insights and intelligence that its competitors may not have access to.

6. Scalability: As a bookstore grows, it may become increasingly difficult to manage and analyze data using traditional methods. A data warehouse can scale with the business, allowing it to continue to grow and thrive.

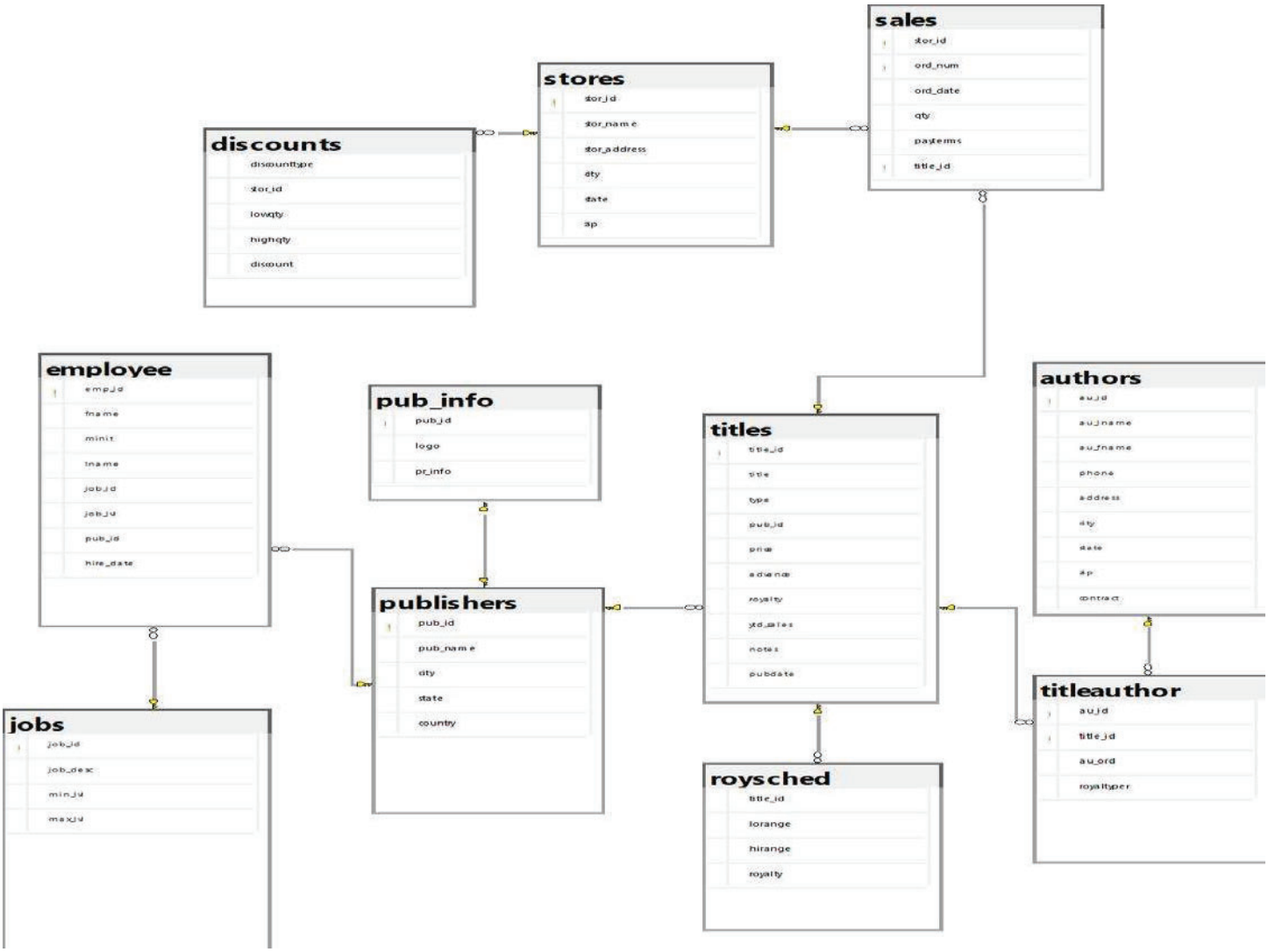
System Level Design
(OLTP System)
Business DB

2.1 Tables Structure

Our Book Store System consists of: Authors, Discounts, Employee, Jobs, pub_info, Publishers, roysched, Sales, Stores, Titles, TitleAuthor. These tables were created by Sql Script in SQL management server.

2.2 Database Design

Following figure shows the attributes of each entity that we had defined in the database of the bookstore system.



Data Warehouse Design ROLAP

3.1 Business Intelligence Questions

Following are the business intelligence questions which we will be trying answer:

- What is the average sales revenue per store location?
- How has the sales revenue of the store changed over time?
- What are the best-selling book in each store location?
- How does the book author demographic (age, gender, etc.) vary by store location?
- What is the store's average profit margin per book sold?
- How does the store's inventory turnover rate compare to industry benchmarks?
- Who is the best-selling publisher?

3.2 Fact Table

The fact table of bookstore management system consists of only **sales** table.

3.3 Dimensions Table

The dimensions table of bookstore management system consists of following table:

- **Time:** A dimension that tracks dates and times, often used to track events or transactions over time.
- **Location:** A dimension that tracks geographical location, often used to track sales or other data by store or region.
- **Author:** A dimension that tracks information about authors, such as name, genre, and nationality.
- **Title:** A dimension that tracks information about book titles, such as genre, publisher, and release date.
- **Store:** A dimension that tracks data related to individual stores, such as store location, size, and performance.
- **Publisher:** A dimension that tracks data related to publishers, such as publisher name, location, and the types of books they publish.

3.4 Star Schema

In a star schema, data is organized into a central fact table surrounded by several dimension tables. The fact table contains the key performance indicators (KPIs) that the business wants to track, and the dimension tables provide context and additional details about those KPIs.

3.4.1 Schema Table Structure

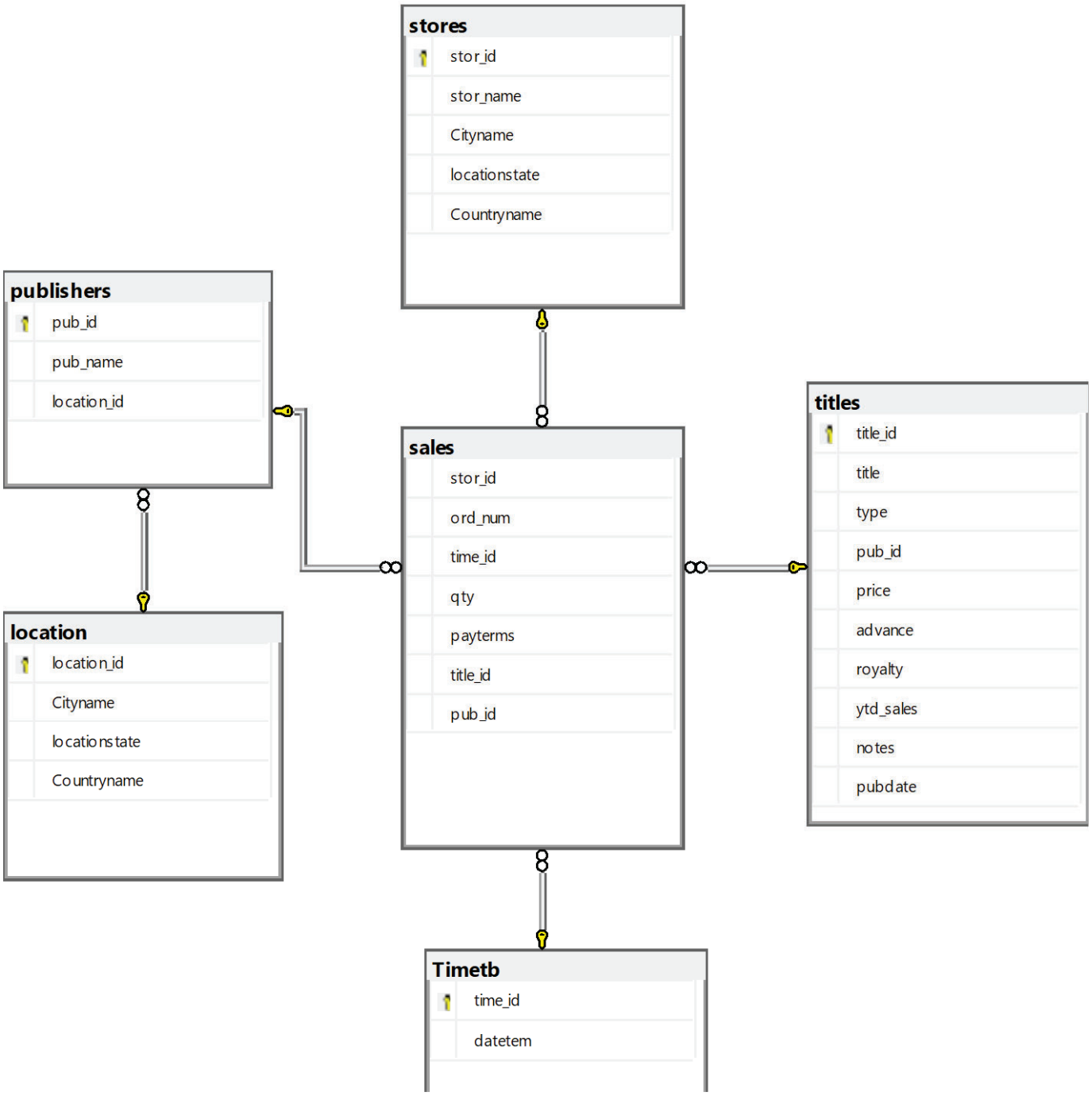
Following is the file directory for a star schema in a bookstore data warehouse:

- Sales Fact Table: Contains data about book sales, such as sales volume, revenue, and margin.
- Time Dimension Table: Contains data about dates and times, such as year, month, and day.
- Location Dimension Table: Contains data about store locations, such as store name and region.
- Author Dimension Table: Contains data about authors, such as demographics and purchasing history.
- Titles Dimension Table: Contains data about titles, such as book title and genre.

Achtung: Namensgebung

3.4.2 Star Schema

In this schema, the Sales Fact Table would contain foreign keys to the Time, Location, and titles dimension tables, allowing the data to be analyzed and reported on in various combinations. For example, a publisher might use the schema to track sales by store location and time period, or to analyze author demographics.



Extract Transform Load(ETL)

4.1 ETL Pipeline

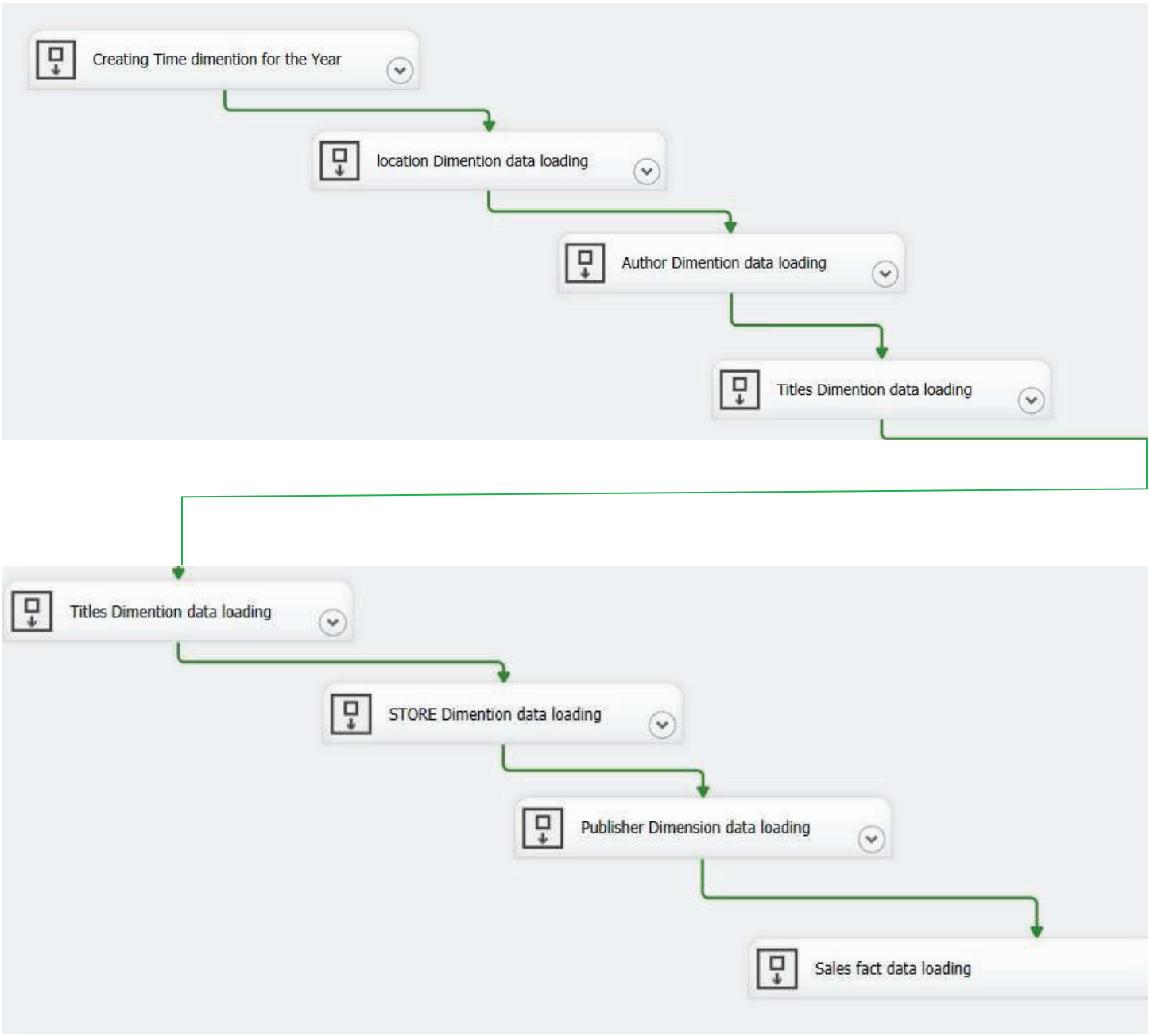
An ETL (extract, transform, load) pipeline is a set of processes that moves data from one or more sources, transforms it into a format that is suitable for analysis, and loads it into a destination, such as a data warehouse. In the bookstore management system, an ETL pipeline is used to extract data from various sources such as point-of-sale systems, inventory databases, and customer relationship management systems, and load it into a data warehouse for analysis and reporting.

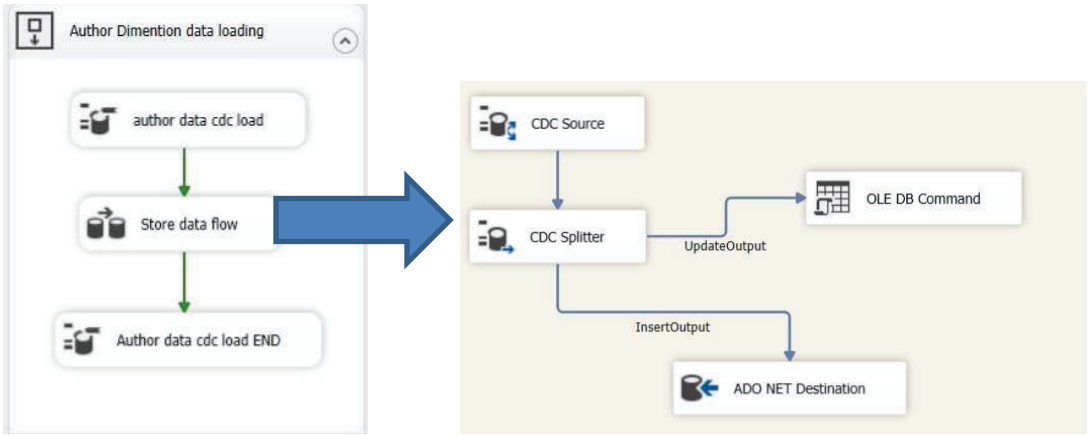
4.1.1 Data Loading

Data is being loaded in the following order:

1. Dimension of Location is loaded.
2. Then the author dimension loaded.
3. Afterwards titles dimension is loaded.
4. Similarly store dimension is loaded.
5. Also, publisher dimension is loaded.
6. Then sales dimension is loaded.

The ETL pipeline is scheduled to run regularly, such as daily or weekly, to ensure that the data warehouse is kept up-to-date with the latest data from the source systems.


Creating Time dimention for the Year
location Dimention data loading
Author Dimention data loading
Titles Dimention data loading
Titles Dimention data loading
STORE Dimention data loading
Publisher Dimension data loading
Sales fact data loading


Author Dimention data loading
author data cdc load
Store data flow
Author data cdc load END
CDC Source
CDC Splitter
OLE DB Command
UpdateOutput
InsertOutput
ADO NET Destination

4.2 Change data Capture (CDC)

Change Data Capture is a software process that identifies and tracks changes to data in a database. CDC provides real-time or near-real-time movement of data by moving and processing data continuously as new database events occur.

Following figure shows the code for the CDC:

```
exec sys.sp_cdc_enable_table
    @source_schema =N'dbo',
    @source_name=N'author' ,
    @Role_Name=N'cdc_admin',
    @capture_instance =  'newstore',
    @supports_Net_Changes=1
    <--Enabling store table
```