Data Warehouse – HW#2

Dimensional Modeling Design
Second Semester 2024/2025
Student: Hala Khalifeh
Student ID: **12112858**
Dataset Used: Rental Film Company.
DBMS Used: MySQL Workbench

Introduction

This assignment focuses on building a dimensional model based on transactional data for a rental film company. The provided dataset was imported into MySQL Workbench, and dimensional modeling techniques were applied to design and implement two analytical fact tables. These tables serve different business intelligence purposes, each supported by appropriate dimension tables derived from the original dataset.

The goal of the assignment is to design a schema that supports efficient OLAP (Online Analytical Processing) queries by organizing data into fact and dimension tables. The dimensional model must also be evaluated for its overall structure (Star vs. Snowflake schema) based on normalization level and query performance.

Objective

- Identify and create the required dimension tables for each fact table **from scratch**.
- Design two main fact tables:
    - Monthly payment per staff per rent.
    - Daily inventory per film per store.
- Analyze the dimensional model structure and determine whether it follows a Star Schema or Snowflake Schema, considering performance and normalization.

Design the dimensional model:

1. fact_monthly_payment_per_staff_per_rent

Tracks the total payment collected for each rental by each staff member per month.

Column Name	Description
fact_id	Primary key (auto-incremented)
date_id	Foreign key to dim_date
staff_id	Foreign key to dim_staff
rental_id	Foreign key to dim_rental
total_payment	Decimal value of the payment amount

```
mysql> DESCRIBE fact_monthly_payment_per_staff_per_rent;
+--------------+--------------+------+-----+---------+----------------+
| Field        | Type         | Null | Key | Default | Extra          |
+--------------+--------------+------+-----+---------+----------------+
| fact_id      | int          | NO   | PRI | NULL    | auto_increment |
| date_id      | int          | YES  | MUL | NULL    |                |
| staff_id     | smallint     | YES  | MUL | NULL    |                |
| rental_id    | int          | YES  | MUL | NULL    |                |
| total_payment | decimal(10,2) | YES  |     | NULL    |                |
+--------------+--------------+------+-----+---------+----------------+
```

Dimension Tables of fact_monthly_payment_per_staff_per_rent:

1. dim_date: Includes date attributes to support time-based aggregation.

Column Name	Description
date_id	Primary key
full_date	Full calendar date
day	Numeric day
month	Numeric month
year	Four-digit year
day_name	Name of the day (e.g., Monday)
month_name	Name of the month (e.g., March)
month_start	1st day in the month (SK)

```
mysql> DESCRIBE dim_date;
+------------+------------+------+-----+---------+----------------+
| Field      | Type       | Null | Key | Default | Extra          |
+------------+------------+------+-----+---------+----------------+
| date_id    | int        | NO   | PRI | NULL    | auto_increment |
| full_date  | date       | NO   |     | NULL    |                |
| day        | int        | YES  |     | NULL    |                |
| month      | int        | YES  |     | NULL    |                |
| year       | int        | YES  |     | NULL    |                |
| day_name   | varchar(10) | YES  |     | NULL    |                |
| month_name | varchar(10) | YES  |     | NULL    |                |
| month_start | date       | YES  |     | NULL    |                |
+------------+------------+------+-----+---------+----------------+
```

2. dim_staff: describes staff members responsible for rental transactions

Column Name	Description
staff_id	Primary key. Unique identifier for each staff member
full_name	Concatenation of first and last name
email	Staff member's email
store_name	Name or identifier of the store the staff belongs to
store_address	Address of the store (includes street details)
city	City where the store is located
country	Country where the store is located
active	Boolean value indicating whether the staff member is currently active

```
mysql> DESCRIBE dim_staff;
+---------------+-------------+------+-----+---------+-------+
| Field         | Type        | Null | Key | Default | Extra |
+---------------+-------------+------+-----+---------+-------+
| staff_id      | smallint    | NO   | PRI | NULL    |       |
| full_name     | varchar(100)| YES  |     | NULL    |       |
| email         | varchar(100)| YES  |     | NULL    |       |
| store_name    | varchar(100)| YES  |     | NULL    |       |
| store_address | varchar(255)| YES  |     | NULL    |       |
| city          | varchar(100)| YES  |     | NULL    |       |
| country       | varchar(100)| YES  |     | NULL    |       |
| active        | tinyint(1)  | YES  |     | NULL    |       |
+---------------+-------------+------+-----+---------+-------+
```

3. dim_rental: describes each rental transaction tied to a payment

Column Name	Description
rental_id	Primary key. Unique identifier for each rental transaction
rental_date	Date and time when the rental was made
return_date	Date and time when the film was returned
customer_full_name	Concatenation of customer’s first and last name
customer_email	Customer’s email address
customer_city	City where the customer resides
customer_country	Country where the customer resides

```
mysql> DESCRIBE dim_rental;
+--------------------+--------------+------+-----+---------+-------+
| Field              | Type         | Null | Key | Default | Extra |
+--------------------+--------------+------+-----+---------+-------+
| rental_id          | int          | NO   | PRI | NULL    |       |
| rental_date        | datetime     | YES  |     | NULL    |       |
| return_date        | datetime     | YES  |     | NULL    |       |
| customer_full_name | varchar(100) | YES  |     | NULL    |       |
| customer_email     | varchar(100) | YES  |     | NULL    |       |
| customer_city      | varchar(100) | YES  |     | NULL    |       |
| customer_country   | varchar(100) | YES  |     | NULL    |       |
+--------------------+--------------+------+-----+---------+-------+
```

2. fact_daily_inventory_per_film_per_store

Tracks the number of inventory items (copies) available per film in each store on a daily basis.

Column Name	Description
fact_id	Primary key (auto-incremented)
date_id	Foreign key to dim_date
film_id	Foreign key to dim_film
store_id	Foreign key to dim_store
inventory_count	Total number of copies of a specific film in a store on that date

```
mysql> DESCRIBE fact_daily_inventory_per_film_per_store;
+-----------------+----------+------+-----+---------+----------------+
| Field           | Type     | Null | Key | Default | Extra          |
+-----------------+----------+------+-----+---------+----------------+
| fact_id         | int      | NO   | PRI | NULL    | auto_increment |
| date_id         | int      | YES  | MUL | NULL    |                |
| film_id         | smallint | YES  | MUL | NULL    |                |
| store_id        | tinyint  | YES  | MUL | NULL    |                |
| inventory_count | int      | YES  |     | NULL    |                |
+-----------------+----------+------+-----+---------+----------------+
```

Dimension Tables of fact_daily_inventory_per_film_per_store:

1. dim_date (same as above)

2. dim_film: describes the films available in inventory.

Column Name	Description
film_id	Primary key. Unique identifier for each film
title	Title of the film
release_year	Year the film was released
language_name	Name of the language
rental_duration	Duration the film can be rented (in days)

`rental_rate`	Cost of renting the film
`length`	Length of the film (in minutes)
`rating`	Film rating (e.g., G, PG, PG-13, R)
`category`	Genre/category of the film

```
mysql> DESCRIBE dim_film;
+-----------------+-------------+------+-----+---------+-------+
| Field           | Type        | Null | Key | Default | Extra |
+-----------------+-------------+------+-----+---------+-------+
| film_id         | smallint    | NO   | PRI | NULL    |       |
| title           | varchar(255)| YES  |     | NULL    |       |
| release_year    | year        | YES  |     | NULL    |       |
| language_name   | varchar(50) | YES  |     | NULL    |       |
| rental_duration | tinyint     | YES  |     | NULL    |       |
| rental_rate     | decimal(4,2)| YES  |     | NULL    |       |
| length          | smallint    | YES  |     | NULL    |       |
| rating          | varchar(10) | YES  |     | NULL    |       |
| category        | varchar(50) | YES  |     | NULL    |       |
+-----------------+-------------+------+-----+---------+-------+
```

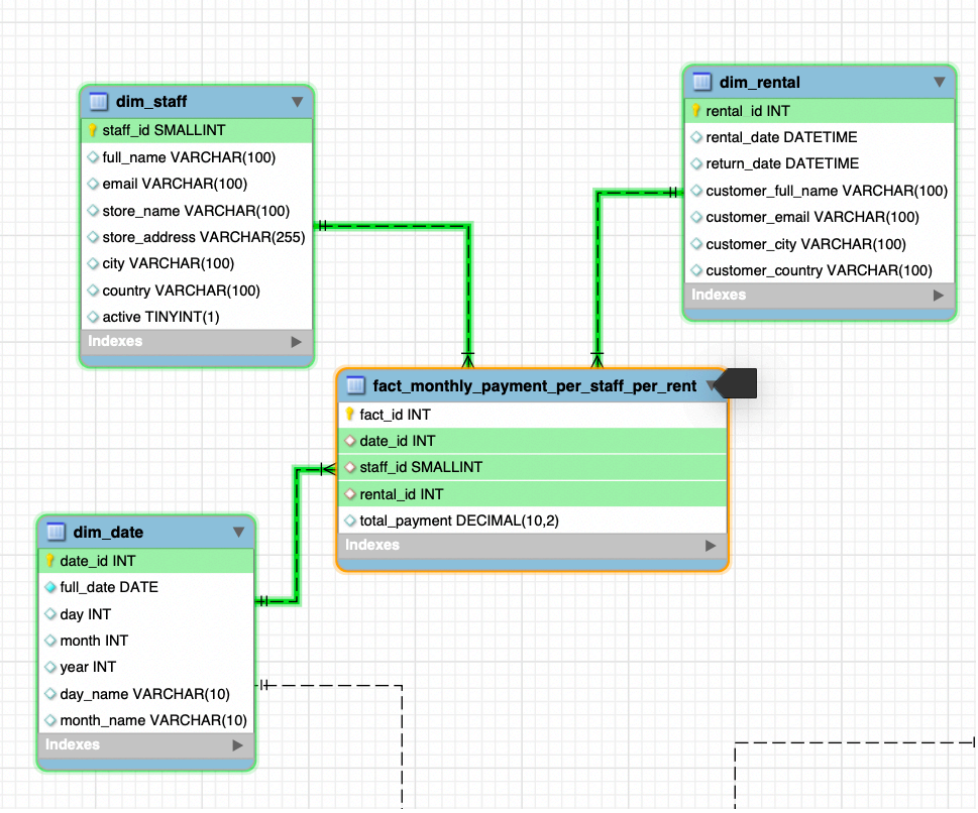
3. dim_store: describes the physical store location

Column Name	Description
`store_id`	Primary key. Unique identifier for each store
`store_name`	Optional store name or identifier (e.g., Store 1, Store 2)
`address`	Full address (street + district + postal code)
`city`	City where the store is located
`country`	Country where the store is located
`manager_name`	Full name of the store manager (denormalized from `staff`)

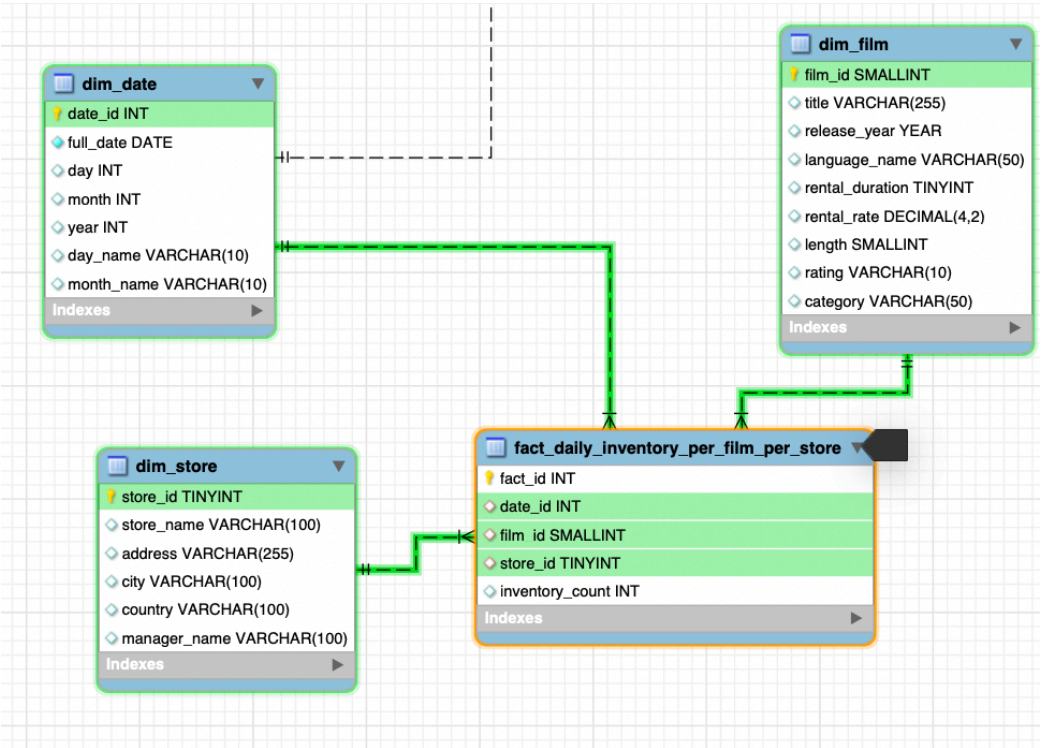
```
mysql> DESCRIBE dim_store;
+--------------+-------------+------+-----+---------+-------+
| Field        | Type        | Null | Key | Default | Extra |
+--------------+-------------+------+-----+---------+-------+
| store_id     | tinyint     | NO   | PRI | NULL    |       |
| store_name   | varchar(100)| YES  |     | NULL    |       |
| address      | varchar(255)| YES  |     | NULL    |       |
| city         | varchar(100)| YES  |     | NULL    |       |
| country      | varchar(100)| YES  |     | NULL    |       |
| manager_name | varchar(100)| YES  |     | NULL    |       |
+--------------+-------------+------+-----+---------+-------+
```

EERD (from MySQLWorkbench)

1. fact_monthly_payment_per_staff_per_rent



2. fact_daily_inventory_per_film_per_store



Which scheme fits the above dimension tables?

Answer: **Star** Schema

Justification:

The dimensional model follows a **Star Schema** because each fact table —

- `fact_monthly_payment_per_staff_per_rent`
- `fact_daily_inventory_per_film_per_store`

— is directly linked to a set of **fully denormalized dimension tables**. These dimension tables contain all necessary descriptive attributes and do **not rely on foreign key references to other normalized tables**, which is a defining characteristic of the **Star Schema**.

Specific Examples from Our Schema:

dim_staff

- Combines staff details **with store and address information** into a single table.
- Instead of referencing `store_id` or `address_id`, the store name, address, city, and country are **directly included**.

dim_rental

- Flattens rental and customer data into one table.
- Includes customer name, email, city, and country without referencing separate customer, address, or city tables.

dim_film

- Merges film details with language and category information.
- The language_name and category fields are stored directly instead of using foreign keys to the language or category tables.

dim_store

- Includes full address, city, country, and manager's name directly, rather than referencing address, city, country, or staff tables.

Conclusion:

The model is a **true Star Schema** because:

- All dimension tables are fully denormalized.
- There are **no snowflake-style foreign key relationships** between dimensions and other lookup tables.
- This structure minimizes joins and supports **faster OLAP queries** for analytical processing, which improves **query performance** and **simplifies report generation**.