**Data Warehouse Design Documentation for Sakila**

**Introduction**

A data warehouse is a specialized database designed to store historical data from various sources, primarily for analytical purposes. In our approach, we will construct our data warehouse using dimensional modeling, specifically utilizing a star schema.

To ensure the effectiveness of our design, we will follow specific steps to achieve a well-structured architecture. Firstly, we'll identify the business use cases, understanding the specific analytical needs and objectives of the organization. Next, we'll determine the grain of our data, which essentially defines the level of detail at which we will store our measures. This involves identifying the most detailed level of data relevant to our analytical objectives.

Following this, we'll proceed to identify the dimensions that provide context to our measures, answering questions such as What, When, Where, and by Whom. These dimensions will form the backbone of our dimensional model, providing the necessary context for meaningful analysis. Lastly, we'll pinpoint the measures we need to track and analyze. These measures represent the key metrics or performance indicators that will populate our fact table(s), capturing the quantitative data points essential for analysis and reporting.

By systematically following these steps, we aim to create a data warehouse design that is optimized for analytical purposes, providing the organization with valuable insights and facilitating informed decision-making.

**Define business use cases**  
To discern our business use cases, it's imperative to comprehend the business processes that warrant tracking within our data warehouse. Our business operations unfold in distinct steps: a customer visits a particular store, requests to rent a specific DVD from a designated staff member, subsequently returns the DVD after a defined duration, and settles the rental payment to same/another staff member.

In order to ascertain the requirements for tracking and storing these processes effectively, we must gather pertinent information regarding the customer, store, staff member, film, as well as the commencement and conclusion dates of the rental period, and the payment transaction date. This comprehensive dataset will equip us with the necessary insights to accurately capture and analyze the intricacies of our business operations.

**Identify the grain**

In our business operations, the most granular level of detail lies within the transaction where a customer rents a DVD film.

**Design Dimensions**   
Dimensions primarily aim to address key questions such as "what?", "where?", "when?", and "by whom?" In our context:

* "What?" pertains to the film of the rented DVD.
* "Where?" denotes the specific store where the transaction occurred.
* "When?" signifies both the start and end dates of the rental period.
* "Who?" encompasses both the customer and staff involved in the transaction.

**Design fact table**In our business case, the main measures we need to track are payments and the accumulated transactions occurring between rental and payment. Typically, this would necessitate two separate fact tables: one for tracking business transactions and another for accumulated processes. However, in our scenario, we can consolidate these into a single fact table. This unified fact table can effectively track both measures, as payments align with the granularity of the entire business process. Since payments cannot be measured at a lower level of granularity, merging the two fact tables simplifies the design.

However, this integration presents a challenge: some transactions in the accumulated fact may not yet have a corresponding payment, resulting in null payment amounts. This is acceptable because aggregation functions handle nulls as expected. Additionally, we can introduce a binary column indicating whether the rental has been paid for, facilitating filtering and payment count calculations