Mi Gao

migao@bu.edu

Assignment #4

MET CS 779

MET CS 779 Assignment #4

Table of Contents

[1. Introduction 2](#_Toc47485980)

[2. Business questions of the dimensional data mart will focus on 2](#_Toc47485981)

[1) Date Preference 2](#_Toc47485982)

[2) Weather Preference 2](#_Toc47485983)

[3) Preferred viewing period 2](#_Toc47485984)

[4) Preferred rental period 2](#_Toc47485985)

[5) Repurchase preference 2](#_Toc47485986)

[6) Quarter preference 2](#_Toc47485987)

[7) Media type preference 2](#_Toc47485988)

[3. ERD 2](#_Toc47485989)

[4. Slowly changing dimensions 3](#_Toc47485990)

[5. The measures I have selected 3](#_Toc47485991)

[6. Indexes and brief explanation 4](#_Toc47485992)

[5. Revision History 5](#_Toc47485993)

# Introduction

Create an initial design for a dimensional data mart for an online DVD rental business that is similar to a simplification of the business pioneered by *NetFlix*. They have two ways to rent the DVD for the customer which are digital copy online and hard copy in local store or shipment.

# Business questions of the dimensional data mart will focus on

## Date Preference

To check of what date the customers more like to watch the movies. For example, weekdays or weekends.

## Weather Preference

To check of what date the customers more like to watch what movies. Rainy days for some kind of movies.

## Preferred viewing period

To check of how often the customers would like to watch a movie. Watching one per day or per week.

## Preferred rental period

To check of how often the customers would like to keep a movie for a rental. Fast returning or keep for a while.

## Repurchase preference

To check of how likely the customer will repurchase the movies they have watched. Never or any.

## Quarter preference

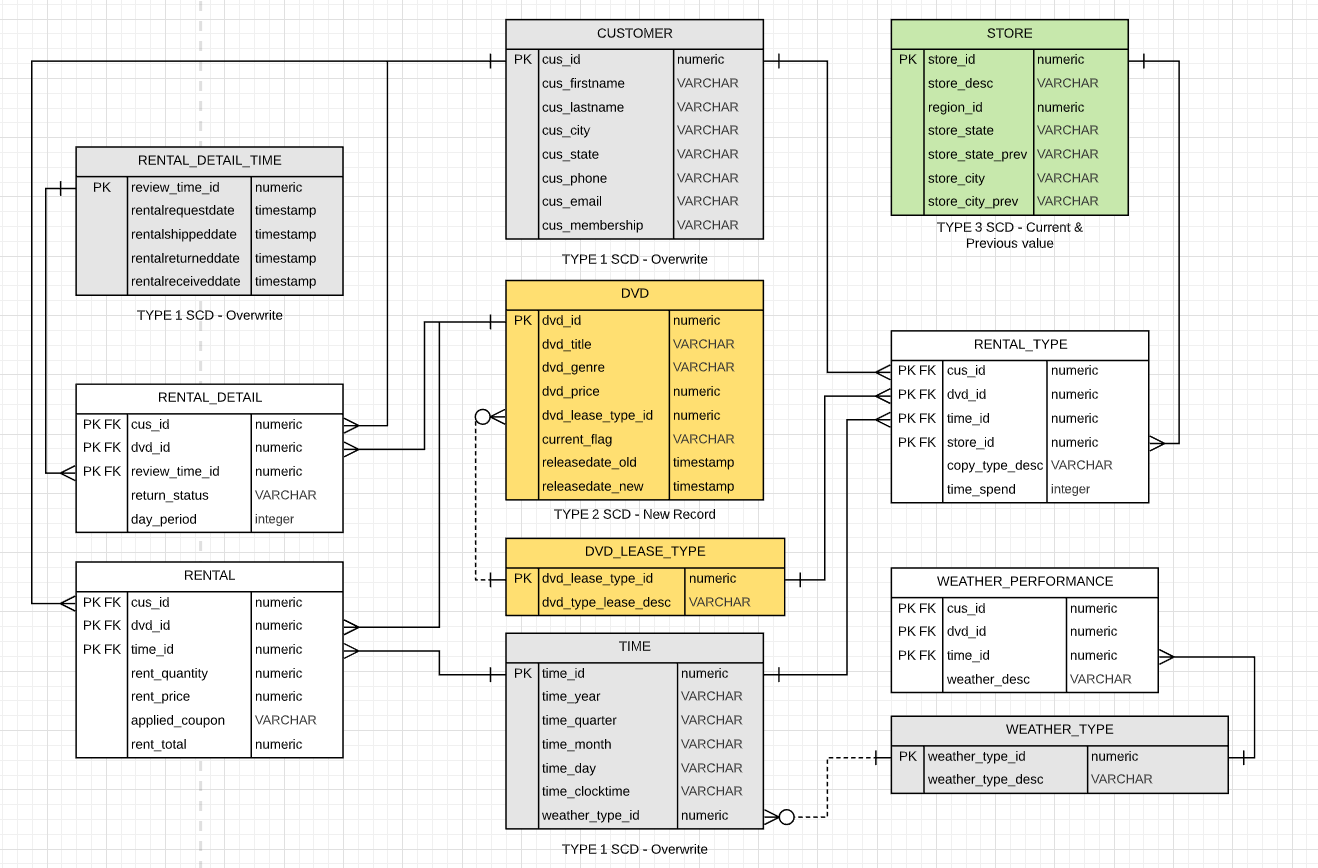
To check of which quarter has the highest purchase amount and highest purchase rate. Summer or Winter.

## Media type preference

To check if users prefer hard copy or digital version of the movie.

# ERD

My schema is constellation type. And you will see the implementations of the key business questions above in the fact tables I created. Notice that there will have on row for only digital copy store in STORE table as the record for supporting the record in fact table.



# Slowly changing dimensions

CUSTOMER table should be type 1 SCD because everything can be overwrite on it but will not change the unique of this person. REVIEW\_TIME is also should be type 1 with similar reason because if targeting to specific one record; they will only record one for each column otherwise they will be redundant. Time table should be type 1 and has same reason with REVIEW\_TIME table; I give the WEATHER\_TYPE table related to TIME because I do not want make the value in TIME table for several specific reasons.

DVD table should be type 2 SCD because retains the full history of values. When the value of a chosen attribute changes, the current record is closed. A new record is created with the changed data values and this new record becomes the current record. Each record contains the effective time and expiration time to identify the time period between which the record was active. For example, it will have several editions such as Cinema version, Director's Cut, or Collector's Edition with different prices.

STORE table is type 3 SCD because the store location may change but we do not really need to record all history of each re-location.

# The measures I have selected

RENTAL table: rent\_quantity, rent\_price, and rent\_total are additive for all of dimensions. Applied\_coupon is non-additive because it does not depend on any of the dimensions present in the fact table.

RENTAL\_DETAIL table: return\_status and day\_period are all additive facts because they are all can be summed up through all of the dimensions in the fact table.

WEATHER\_PERFORMANCE table: weather\_desc is semi-additive because it is meaningful for adding up to TIME and DVD level but not for CUSTOMER.

RENTAL\_TYPE table: copy\_type\_desc and time\_spend are semi-additive facts because they are meaningful for adding up to DVD and Store level but not for CUSTOMER and Time.

# Indexes and brief explanation

I will give those three indexes for those three attributes:

* 1. Weather table: weather\_type\_id
  2. DVD table: current\_flag, and dvd\_lease\_type\_id

For index of weather\_type\_id, I would like to use hash index for precise searching specific weather. Unlike the BTree index that needs to go from the root node to the branch node, the query efficiency of the Hash index is much higher than that of the B-Tree index. Also, I do not need to order the result so hash index should be good for it.

For the index of current\_flag, implementing a bitmap index so that we could quickly find the current instances. First, there are a large number of rows in the table. Then, the type of value is less. Bitmap index is particularly helpful in complex queries or aggregate queries (including SUM, COUNT or other aggregate functions) with lengthy WHERE conditions. We can quickly know which the flag type of this DVD.

For the index of dvd\_lease\_type\_id, I would also use bitmap index. Because only have the type of digital or hard copy in here; there does not have too many values in this column. It takes up very little space. Also, update, insert, or delete query does usually not effect on this column.

# 5. Revision History

|  |  |  |  |
| --- | --- | --- | --- |
| **Name** | **Date** | **Version** | **Description** |
| Mi Gao | 08/04/19 | 1.0 | Initial Document Creation |
| Mi Gao | 08/05/19 | 1.1 | Added more details |