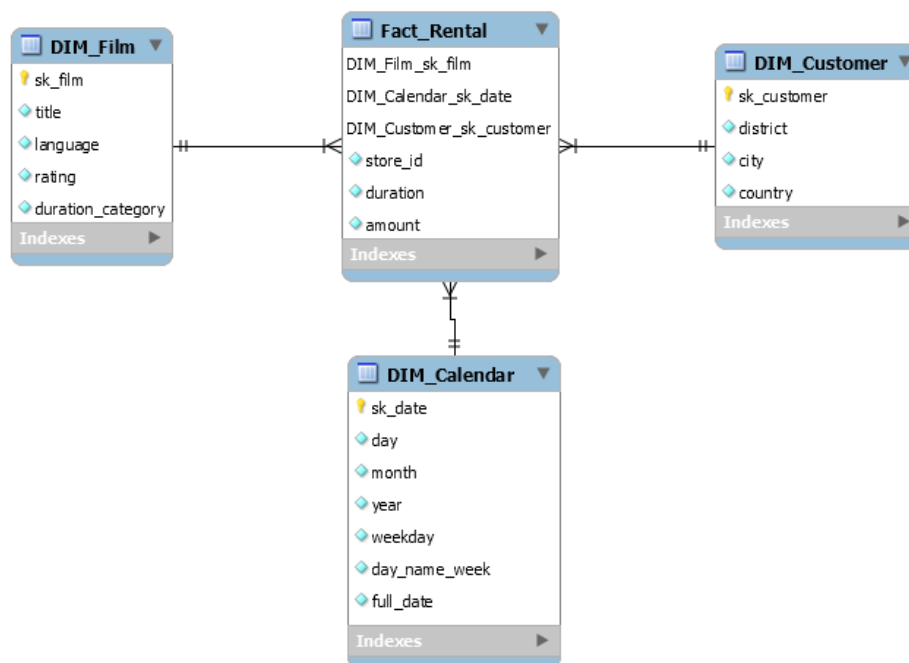


Considere o contexto do caso de estudo: “Sakila”, abordado anteriormente.

Tenha em consideração a existência de duas fontes (lojas) com a mesma estrutura da base de dados já apresentada no caso de estudo *Sakila* (disponíveis como recurso do caso de estudo). As bases de dados possuem nos dados relacionados com atividade comercial de cada loja.

Considere ainda o seguinte esquema de um *Data Mart* especialmente desenhado para suportar atividades de suporte à decisão da empresa:



1. Apresente um mapa *source-to-target* para suportar o processo de povoamento do *Data Mart* apresentado. Considere que:
 - a. Na dimensão “DIM_Film”, o atributo: “duration_category” representa a categoria de duração do filme. Considere o número de minutos de duração do filme (*length*) para a definição da categoria:
 - i. se < 80, então a categoria de duração será: “short”
 - ii. se estiver compreendida no intervalo 80 e 120 (inclusive) a categoria é “regular”
 - iii. se for superior a 120, então a categoria é “long”
 - b. O povoamento da dimensão calendário (“DIM_Calendar”) poderá ser realizada utilizando o ficheiro: *Date_Dim_2000-2020*.
2. Utilizando o Pentaho Data Integration, explore os dados dos objetos de base de dados identificados de forma a identificar possíveis problemas que comprometam o povoamento do Data Mart.
3. Desenvolva vistas (*views*) em cada uma das fontes de forma a simplificar o processo de extração de dados das fontes;
4. Identifique uma estrutura para a Área de Retenção (DSA) para suportar o processo de povoamento. Idealize um cenário em que o volume de dados, complexidade de transformação e disponibilidade das fontes, exija a

utilização de uma Área de Retenção de forma a suportar a preparação dos dados antes do seu posterior carregamento no *Data Mart*.

5. Apresente um conjunto de modelos conceptuais utilizando *Business Process Model and Notation* (BPMN) que representem as principais etapas de povoamento do Data Mart apresentado. Considere as especificidades do povoamento inicial e incremental.
6. Implemente o processo de povoamento idealizado utilizando a ferramenta: *Pentaho Data integration*¹.

Notas:

É disponibiliza uma máquina virtual com o sistema operativo Lubuntu que contém as ferramentas: MySQL, MySQL Workbench e Pentaho Data Integration já instalados. Podem também encontrar os *scripts* para criar e povoar cada um dos repositórios utilizados. É ainda disponibilizado um ficheiro: *Date_Dim_2000-2020.xlsx*, que pode ser utilizado para suportar o povoamento da dimensão temporal.

Para a modelação BPMN pode utilizar o website: <https://www.lucidchart.com/>, <https://bpmn.io> ou a ferramenta Microsoft Visio.

¹ <https://community.hitachivantara.com/docs/DOC-1009855-data-integration-kettle>