

Implementação de Data Mart (P01)

LICENCIATURA EM ENGENHARIA DE SISTEMAS INFORMÁTICOS

SISTEMAS DE APOIO À DECISÃO – 2020/2021

Grupo Nº14

José Antunes 11582, João Paiva 14154, Gabriela Condez 14844

Introdução

O presente relatório serve de suporte ao primeiro projeto da unidade curricular de Sistemas de Apoio à Decisão, lecionada pelo Professor Joaquim Silva. A elaboração deste trabalho consiste na implementação de um *Data Mart* sobre um conjunto de dados disponibilizados pelo docente.

Denominada de *Sakila*, a base de dados utilizada tem como fonte de dados uma loja online de aluguer de DVDs. O principal objetivo passa então por conhecer e organizar os seus dados, implementando um *Data Mart* através do processo ETL (*Extract, Transform, Load*) com base no software PDI (*Pentaho Data Integration*) *Kettle*.

Fontes de Dados

O seguinte diagrama representa o Modelo ER (modelo entidade – relação) em estudo, relativo a uma loja de aluguer de filmes, como já referido anteriormente. Estes tipos de diagramas têm o intuito de descrever, de forma abstrata, a informação referente ao processo de negócio.

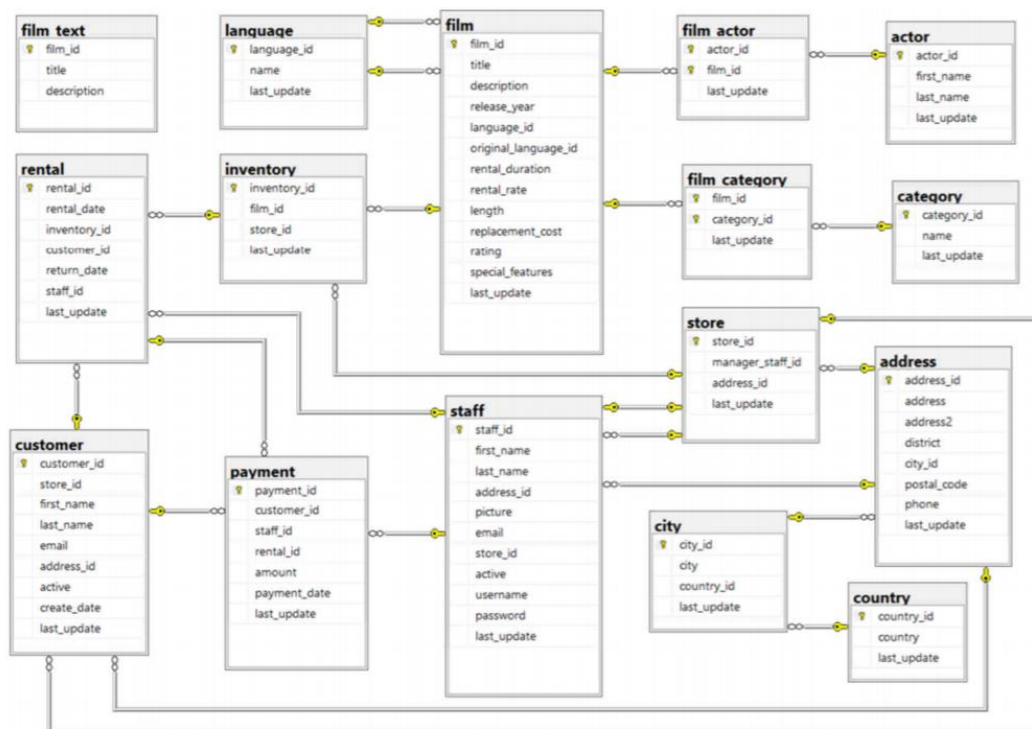


Figura 1 - Modelo ER

Com o apoio da ferramenta *PowerBI* foram construídas as seguintes **Table Profile** referentes a todas as dimensões. Contudo, apenas apresentamos aquelas que consideramos mais importantes bem como uma análise apenas dos dados essenciais para o processo de negócio.

Primeiramente, como se pode analisar na figura abaixo, temos representada a tabela *Customer* que dispõe de 599 registos, cada um deles com informações referentes ao nome, email, entre outros dados.

Column	Min	Max	Average	StandardDeviation	Count	NullCount
active	0	1	null	null	599	0
address	null	null	null	null	599	0
address_id	5	605	304,7245409	173,6986085	599	0
create_date	14/02/2006 22:04:36	14/02/2006 22:04:37	14/02/2006 22:04:37	null	599	0
customer_id	1	599	300	173,060683	599	0
email	AARON.SELBY@sakilacustom...	ZACHARY.HITE@sakilacusto...	null	null	599	0
first_name	AARON	ZACHARY	null	null	599	0
last_name	ABNEY	YOUNG	null	null	599	0
last_update	15/02/2006 04:57:20	15/02/2006 04:57:20	15/02/2006 04:57:20	null	599	0
payment	null	null	null	null	599	0
rental	null	null	null	null	599	0
store	null	null	null	null	599	0
store_id	1	2	1,455759599	0,498455188	599	0

Figura 2 – Table Profile Customer

Relativamente à tabela *Film* podemos observar na figura 3 que dispõe de 1000 registos, sendo que cada registo representa um filme diferente com características particulares, como por exemplo o título e a sua descrição.

Column	Min	Max	Average	StandardDeviation	Count	NullCount	DistinctCount
description	A Action-Packed Character St...	An Epic Yarn of a Teacher An...	null	null	1000	0	1000
film_actor	null	null	null	null	1000	0	null
film_category	null	null	null	null	1000	0	null
film_id	1	1000	500,5	288,8194361	1000	0	1000
inventory	null	null	null	null	1000	0	null
language(language_id)	null	null	null	null	1000	0	null
language(original_language_id)	null	null	null	null	1000	1000	null
language_id	1	1	1	0	1000	0	1
last_update	15/02/2006 05:03:42	15/02/2006 05:03:42	15/02/2006 05:03:42	null	1000	0	1
length	46	185	115,272	40,42633182	1000	0	140
original_language_id	null	null	null	null	1000	1000	1
rating	G	R	null	null	1000	0	5
release_year	2006	2006	null	null	1000	0	1
rental_duration	3	7	4,985	1,411654266	1000	0	5
rental_rate	0,99	4,99	2,98	1,646393213	1000	0	3
replacement_cost	5,99	29,99	19,984	6,050832718	1000	0	21
special_features	Behind the Scenes	Trailers, Deleted Scenes, Behin...	null	null	1000	0	15
title	ACADEMY DINOSAUR	ZORRO ARK	null	null	1000	0	1000

Figura 3 - Table Profile Film

Na seguinte figura está representada a tabela *Payment* dispoendo de 16049 registos diferentes, cada um relativo a um pagamento. Esta tabela fornece dados capazes de identificar o funcionário que efetuou o registo de pagamento (*staff_id*), o cliente que efetuou o pagamento (*customer_id*), o valor e a data.

Column	ABC 123 Min	ABC 123 Max	ABC 123 Average	1.2 StandardDeviation	1.2 Count	ABC 123 NullCount
<div> <div>Válido</div> <div>Erro</div> <div>Vazio</div> </div> <div> <div>100%</div> <div>0%</div> <div>0%</div> </div> <div> <div>13 distintos, 13 exclusivos</div> </div>	<div> <div>Válido</div> <div>Erro</div> <div>Vazio</div> </div> <div> <div>69%</div> <div>0%</div> <div>31%</div> </div>	<div> <div>Válido</div> <div>Erro</div> <div>Vazio</div> </div> <div> <div>69%</div> <div>0%</div> <div>31%</div> </div>	<div> <div>Válido</div> <div>Erro</div> <div>Vazio</div> </div> <div> <div>38%</div> <div>0%</div> <div>62%</div> </div>	<div> <div>Válido</div> <div>Erro</div> <div>Vazio</div> </div> <div> <div>23%</div> <div>0%</div> <div>77%</div> </div>	<div> <div>Válido</div> <div>Erro</div> <div>Vazio</div> </div> <div> <div>100%</div> <div>0%</div> <div>0%</div> </div>	<div> <div>Válido</div> <div>Erro</div> <div>Vazio</div> </div> <div> <div>100%</div> <div>0%</div> <div>0%</div> </div>
1 amount	0	11,99	4,200667331	2,362993854	16049	0
2 customer	null	null	null	null	16049	0
3 customer_id	1	599	297,1626893	172,4680997	16049	0
4 last_update	15/02/2006 22:12:30	15/02/2006 22:24:13	15/02/2006 22:16:23	null	16049	0
5 payment_date	24/05/2005 22:53:30	14/02/2006 15:16:03	23/07/2005 08:07:03	null	16049	0
6 payment_id	1	16049	8025	4633,09157	16049	0
7 rental	null	null	null	null	16049	5
8 rental_id	1	16049	8025,371478	4632,777249	16049	5
9 staff	null	null	null	null	16049	0
10 staff_id	1	2	1,497974952	0,500011477	16049	0

Figura 4 - Table Profile Payment

Por último apresentamos a tabela *Rental*, correspondente ao aluguer. Esta dispõe de 16044 registos, sendo possível saber qual o *rental_id* do aluguer, a data do aluguer, o *inventory_id* do item, o *customer_id* do cliente que realizou o aluguer, a data da devolução, o *staff_id* do funcionário e a última modificação que teve.

Column	ABC 123 Min	ABC 123 Max	ABC 123 Average	1.2 StandardDeviation	1.2 Count	ABC 123 NullCount
<div> <div>Válido</div> <div>Erro</div> <div>Vazio</div> </div> <div> <div>100%</div> <div>0%</div> <div>0%</div> </div> <div> <div>13 distintos, 13 exclusivos</div> </div>	<div> <div>Válido</div> <div>Erro</div> <div>Vazio</div> </div> <div> <div>69%</div> <div>0%</div> <div>31%</div> </div>	<div> <div>Válido</div> <div>Erro</div> <div>Vazio</div> </div> <div> <div>69%</div> <div>0%</div> <div>31%</div> </div>	<div> <div>Válido</div> <div>Erro</div> <div>Vazio</div> </div> <div> <div>38%</div> <div>0%</div> <div>62%</div> </div>	<div> <div>Válido</div> <div>Erro</div> <div>Vazio</div> </div> <div> <div>23%</div> <div>0%</div> <div>77%</div> </div>	<div> <div>Válido</div> <div>Erro</div> <div>Vazio</div> </div> <div> <div>100%</div> <div>0%</div> <div>0%</div> </div>	<div> <div>Válido</div> <div>Erro</div> <div>Vazio</div> </div> <div> <div>100%</div> <div>0%</div> <div>0%</div> </div>
1 customer	null	null	null	null	16044	0
2 customer_id	1	599	297,1431688	172,4531365	16044	0
3 inventory	null	null	null	null	16044	0
4 inventory_id	1	4581	2291,842558	1322,210643	16044	0
5 last_update	15/02/2006 21:30:53	23/02/2006 04:12:08	15/02/2006 21:31:32	null	16044	0
6 payment	null	null	null	null	16044	0
7 rental_date	24/05/2005 22:53:30	14/02/2006 15:16:03	23/07/2005 08:12:53	null	16044	0
8 rental_id	1	16049	8025,371478	4632,777249	16044	0
9 return_date	25/05/2005 23:55:21	02/09/2005 02:35:22	25/07/2005 23:58:03	null	16044	183
10 staff	null	null	null	null	16044	0
11 staff_id	1	2	1,498878085	0,500014324	16044	0

Figura 5 - Table Profile Rental

Visão Geral

A seguinte tabela representa uma visão geral e simplificada do modelo de negócio a ser analisado, constituído por quinze tabelas. Referente aos Atores dos filmes temos as tabelas *Actor* e *Film_actor* com 200 e 5462 registos, respetivamente. Relativamente à tabela *Rental*, que representa o aluguer de filmes, foram identificadas 16044 linhas. As tabelas que representam as categorias, *Category* e *Film_category* constituem 16 e 1000 linhas de registo, respetivamente. Temos ainda as tabelas que representam os funcionários, *Staff*, e os clientes, *Customer*, com 2 e 599 registos, bem como as tabelas que representam as moradas dos mesmos, tendo a tabela *Address* 603, a tabela *City* 600 e a tabela *Country* 109 registos. Na tabela que possui informações sobre os filmes, a tabela *Film*, estão presentes 1000 linhas. Na tabela referente ao seu idioma, tabela *Language*, temos 6 registos. Por último foram analisadas as tabelas *Store* e *Inventory*, referentes às Lojas, existindo 2 e 4581 registos, respetivamente e a tabela *Payment*, com 16049 registos de pagamentos.

Tabela 1 - Tabela de Dados

Evento/Objecto	Tabela	Nr. Registos
Actores	<i>Actor</i>	200
	<i>Film_actor</i>	5462
Aluguer	<i>Rental</i>	16044
Categorias	<i>Category</i>	16
	<i>Film_Category</i>	1000
Clientes	<i>Customer</i>	599
Filmes	<i>Film</i>	1000
Idioma	<i>Language</i>	6
Lojas	<i>Store</i>	2
	<i>Inventory</i>	4581
Moradas	<i>Address</i>	603
	<i>City</i>	600
	<i>Country</i>	109
Pagamentos	<i>Payment</i>	16049
Funcionários	<i>Staff</i>	2

Modelo Dimensional

Os objetivos/questões, além das já solicitadas no enunciado, que foram elaboradas para serem respondidas através da modelação do *data mart* são as seguintes:

- ✓ consultar diretamente as moradas do pessoal do staff, dos clientes e das lojas, evitando consultar mais do que uma tabela;
- ✓ detalhar a informação relativa às datas (ano, mês, semana e dia) apenas numa tabela;
- ✓ detalhar a informação relativamente ao tempo (horas, minutos e segundos) apenas numa tabela;
- ✓ colocar o aluguer (*rental*), que se trata do “ponto central” da base de dados, na tabela de factos (*fact table*).

Tabela 2 - Matriz do armazém de dados

<div> <div></div> <div>DIMENSÕES</div> </div> <div>PROCESSOS DE NEGÓCIO</div>	Dim Actor	Dim Category	Dim Customer	Dim Date	Dim Film	Dim Staff	Dim Store	Dim Time
Efetuar registo de um cliente			X					
Adicionar um filme ao catálogo					X			
Organizar um filme consoante critérios	X	X						
Adicionar uma nova loja							X	
Adicionar novo membro ao staff						X		
Registar novo pagamento				X				X

Construção do Modelo Dimensional

Para a implementação de todo o projeto é essencial fazer uma análise de todo o processo de negócio previamente de modo a conseguir identificar as possíveis dimensões a serem criadas bem como a tabela de factos. Analisando a base de dados *Sakila* e o respetivo Modelo ER, obtivemos o seguinte diagrama já estruturado por dimensões. É também importante referir que, sublinhado a cor-de-rosa, se encontram os registos referentes à dimensão *Date* e à dimensão *Time*, tratando-se de duas dimensões distintas.

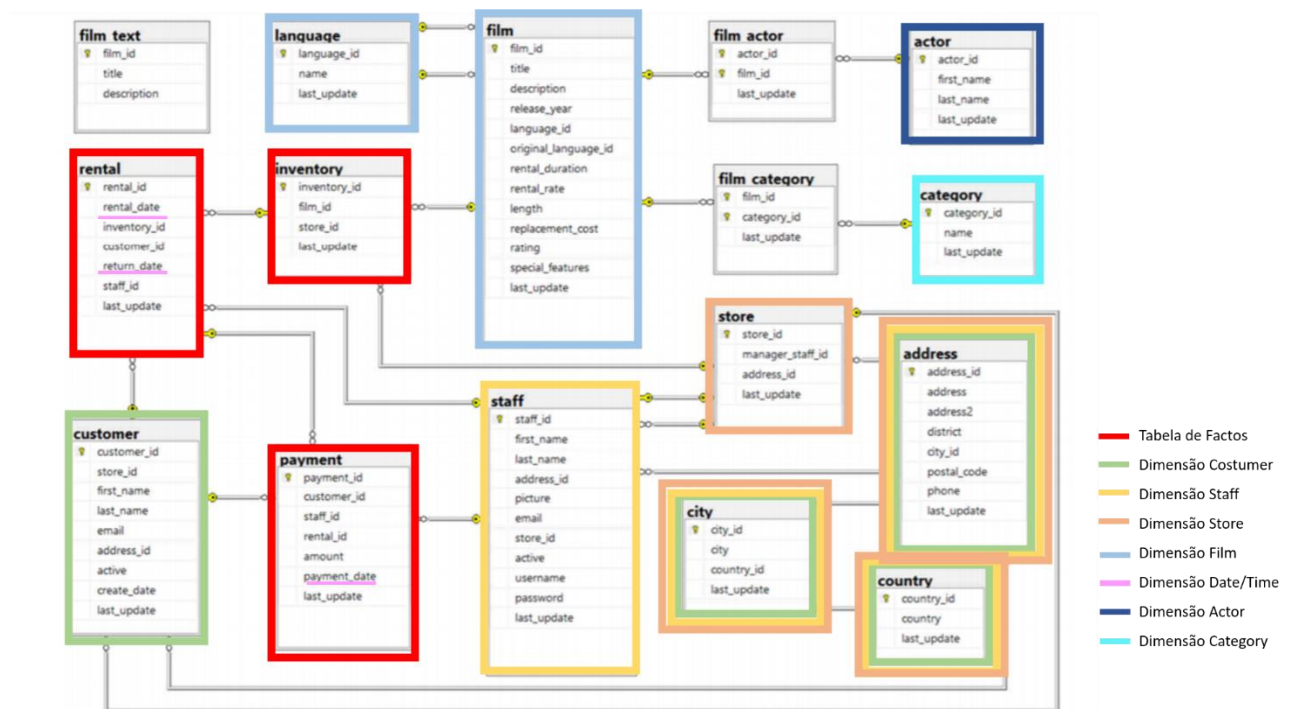


Figura 6 - Construção do Data Mart

Data Mart

O modelo relacional a nível dimensional representa o modelo apresentado inicialmente, de forma atualizada.

Esta atualização representa então as dimensões definidas ao longo do processo, assim como a respetiva Tabela de Factos. Podemos, na figura seguinte, identificar oito dimensões: *Dim_actor*, *Dim_category*, *Dim_customer*, *Dim_Date*, *Dim_Time*, *Dim_Film*, *Dim_Staff*, *Dim_Store*, assim como a Tabela de Factos, *fact_rent_payment*. É ainda possível verificar a existência de duas Bridge tables denominadas por *bridge_actor_film* e *bridge_film_category*.

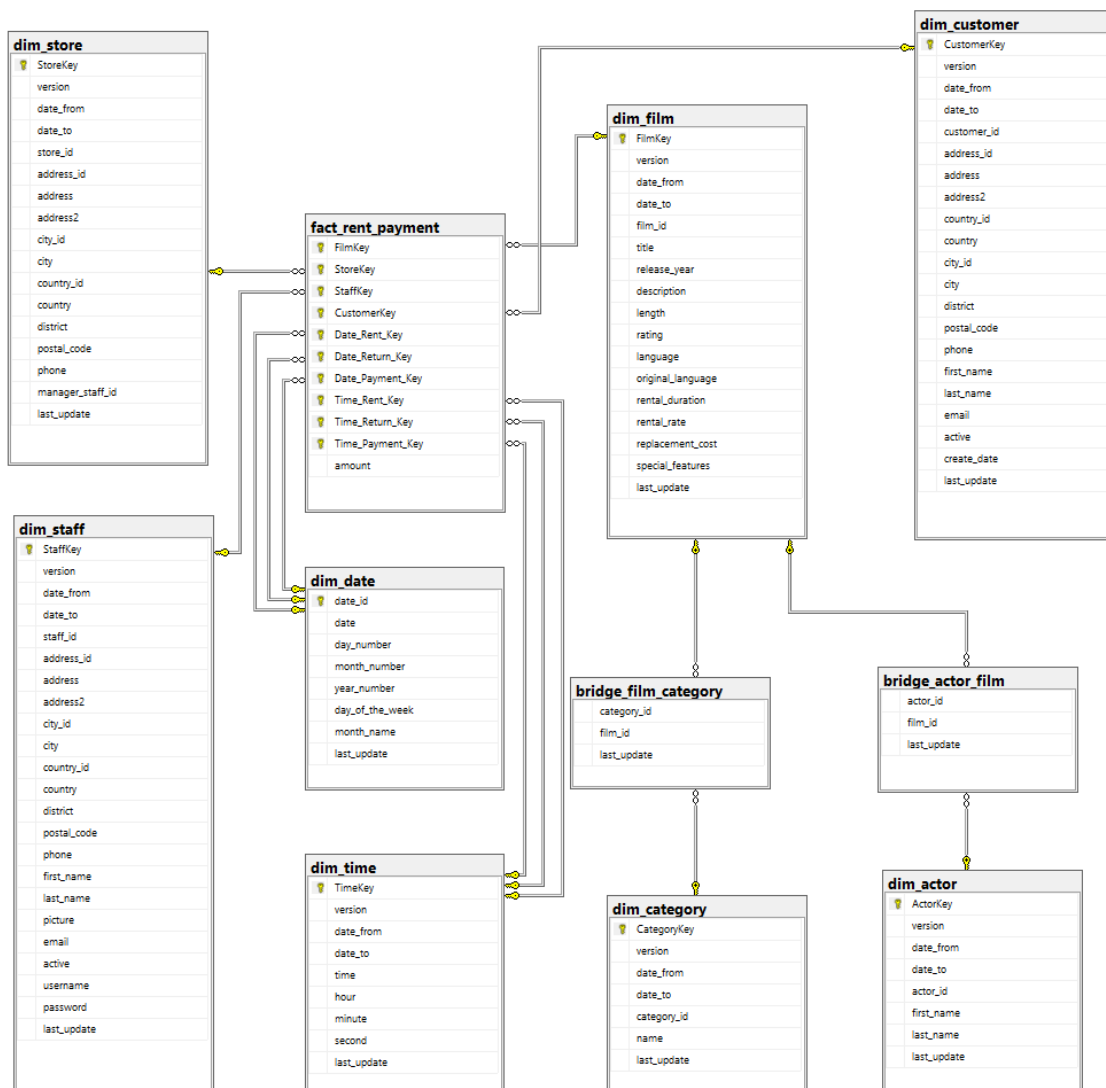


Figura 7 - Data Mart

Implementação do *Data Mart*

Através do software *Kettle* realizamos os processos ETL que têm por base a extração de dados, a sua transformação e a apresentação dos respetivos dados resultantes.

Dimensão *Dim_actor*

A criação desta dimensão tem como principal objetivo gerar uma nova dimensão '*actor*' no Data Mart. Para tal, iniciou-se o processo com um *Table input*, para efetuar uma leitura da tabela já existente na base de dados *Sakila*. Seguidamente, adicionou-se um campo de data de atualização e recorreu-se a um *Dimension lookup/update* para a acrescentar no Data Mart.



Figura 8 - Dimensão *Dim_actor*

Dimensão *Dim_category*

A criação desta dimensão tem como principal objetivo gerar uma nova dimensão '*category*' no Data Mart. Para tal, iniciou-se o processo com um *Table input*, para efetuar uma leitura da tabela já existente na base de dados *Sakila*. Seguidamente, adicionou-se um campo de data de atualização e recorreu-se a um *Dimension lookup/update* para a acrescentar no Data Mart.



Figura 9 - Dimensão *Dim_category*

Dimensão Dim_film

A criação desta dimensão tem como principal objetivo gerar uma nova dimensão 'film' no Data Mart. Para tal, iniciou-se o processo com um *Table input* da tabela 'film' original e da tabela 'language' da base de dados *Sakila*. Seguidamente, integrou-se as linguagens com os filmes através de *lookups*, adicionou-se um campo de data de atualização e recorreu-se a um *Dimension lookup/update* para preencher a dimensão 'film'.

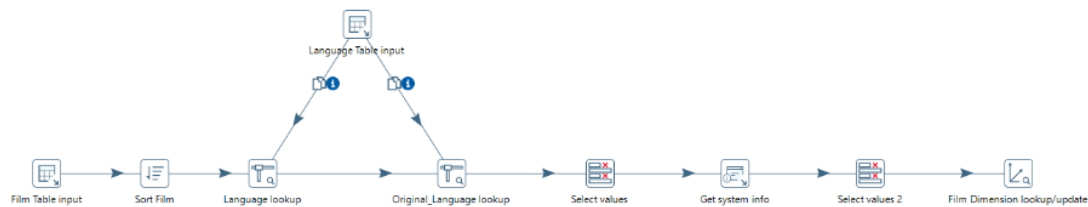


Figura 10 - Dimensão Dim_film

Dimensão Dim_date

A criação da dimensão *date* passa por criar 10000 índices (cada índice corresponde a um dia, equivale mais ou menos do ano de 2005 até 2030). Através dos comandos *generate rows*, *add sequence* e *calc*, "separa" cada data em vários campos. Por fim, adiciona um campo de data de atualização e preenche a *dim_date* no Data Mart através de um *insert/update*.

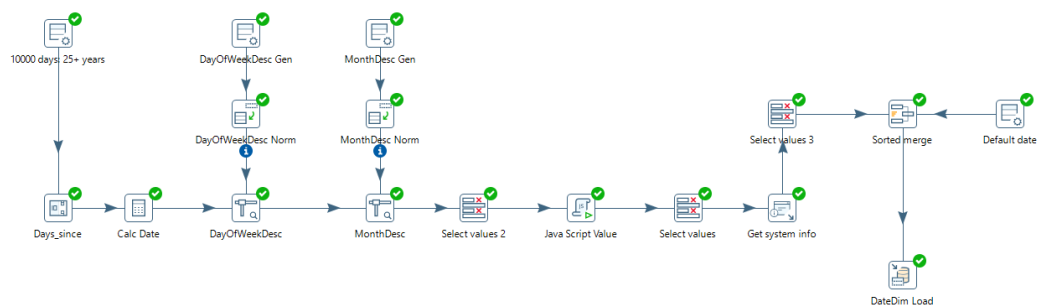


Figura 11 - Dimensão Dim_date

Dimensão Dim_staff

A criação desta dimensão tem como principal objetivo gerar uma nova tabela 'staff' no Data Mart. Para tal, iniciou-se o processo com um *Table input* das tabelas 'Adress', 'City' e 'Country' da base de dados *Sakila*. Seguidamente, integrou-se com a tabela 'staff', adicionou-se um campo de data de atualização e recorreu-se a um *Dimension lookup/update* para a acrescentar no Data Mart.

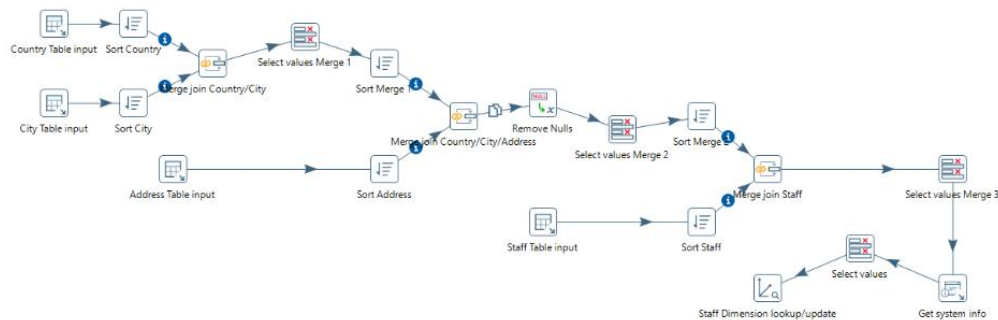


Figura 14 - Dimensão Dim_staff

Dimensão Dim_store

A criação desta dimensão tem como principal objetivo gerar uma nova tabela 'store' no Data Mart. Para tal, iniciou-se o processo com um *Table input* das tabelas 'Adress', 'City' e 'Country' da base de dados *Sakila*. Seguidamente, integrou-se com a tabela 'store', adicionou-se um campo de data de atualização e recorreu-se a um *Dimension lookup/update* para a acrescentar no Data Mart.

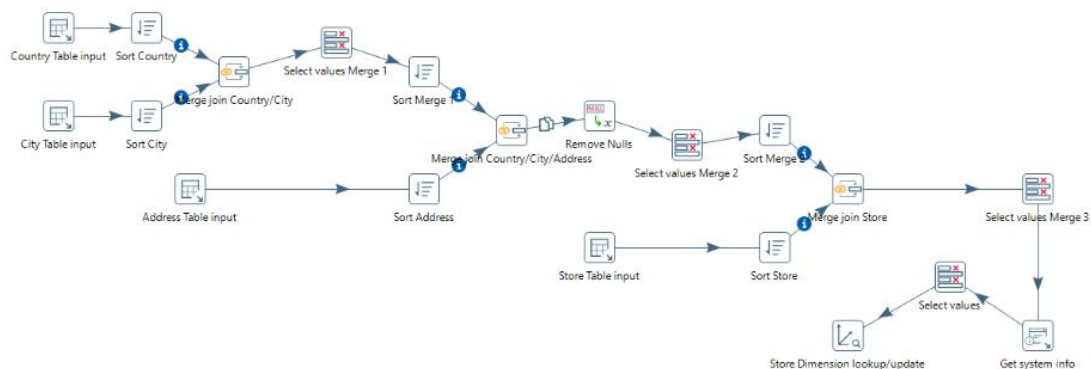


Figura 15 - Dimensão Dim_store

Bridge_film_actor

A criação da 'bridge_film_actor' iniciou-se o processo com um *Table input* da tabela 'film_actor', para efetuar uma leitura da tabela já existente na base de dados *Sakila* e também das tabelas 'actor' e 'film' para extrair os id's através de *lookups*. Seguidamente, adicionou-se um campo de data de atualização e recorreu-se a um *output* para preencher no Data Mart.

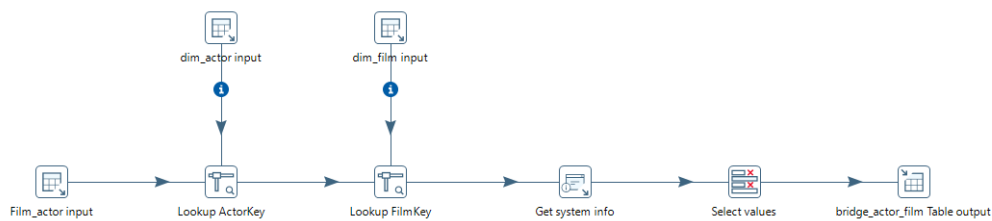


Figure 16 - Bridge_film_actor

Bridge_film_category

A criação da 'bridge_film_category' iniciou-se o processo com um *Table input* da tabela 'film_category', para efetuar uma leitura da tabela já existente na base de dados *Sakila* e também das tabelas 'category' e 'film' para extrair os id's através de *lookups*. Seguidamente, adicionou-se um campo de data de atualização e recorreu-se a um *output* para preencher no Data Mart.

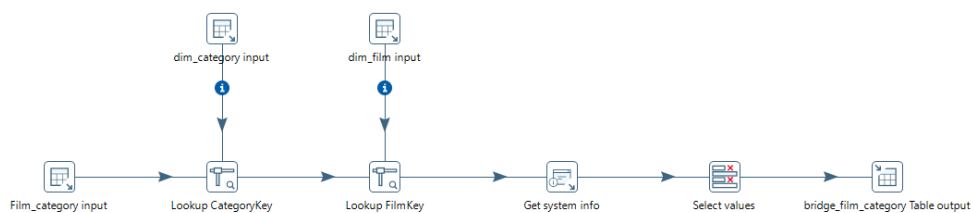


Figure 17 - Bridge_film_category

Fact_table

A criação da *fact table* consiste num *Table input* da tabela 'payment' já existente na base de dados, integrando-a com a tabela 'rental' através de um *lookup*. Seguidamente normalizou-se as 'rows' com valores nulos, transformou-se as três datas (*rental_date*, *return_date* e *payment_date*) em registos para as dimensões *dim_date* e *dim_time* e armazenou-se os respetivos valores das *keys*. Posteriormente juntou-se os dados de *film* e *store* através de *lookups* recorrendo à tabela *inventory*. Juntou-se também *staff* e *customer* das respetivas dimensões. Finalmente, normalizou-se novamente as *rows* com valores nulos e guardou-se, através de um *table output*, os dados na *fact_table*.

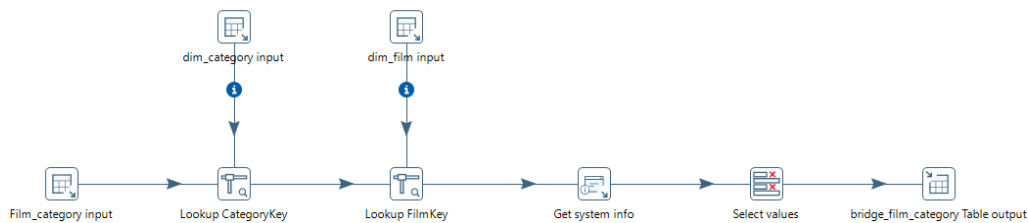


Figure 18 - Fact_table

Visão geral do *Data Mart*

Representada na tabela seguinte, após criadas todas as dimensões necessárias, a nova Base de Dados ficou criada depois da implementação do Data Mart.

Com oito dimensões: *Dim_actor* com 200 registos, *Dim_category* com 16 registos, *Dim_customer* com 599 registos, *Dim_date* com 10000 registos, *Dim_time* com 86400 registos, *Dim_film* com 1000 registos, *Dim_store* com 2 registos e *Dim_staff* com 2 registos.

Relativamente à Tabela de Factos, *Fact_rent_payment*, esta conta com 16049 registos.

Por último, as Bridge tabela, *Bridge_actor_film* com 5462 registos e *Bridge_film_category* com 1000 registos.

Tabela 3 - Conteúdo do Data Mart

Evento/Objecto	Tabela	Nr. Registos
Actores	<i>Dim_actor</i>	200
	<i>Bridge_actor_film</i>	5462
Aluguer e pagamento	<i>Fact_rent_payment</i>	16049
Categorias	<i>Dim_category</i>	16
Clientes	<i>Dim_customer</i>	599
Data/Tempo	<i>Dim_date</i>	10000
	<i>Dim_time</i>	86400
Filmes	<i>Dim_Film</i>	1000
	<i>Bridge_film_category</i>	1000
Lojas	<i>Dim_Store</i>	2
Funcionários	<i>Dim_Staff</i>	2

Conclusão

A realização deste projeto, bem como do seu relatório de suporte, foi bastante vantajoso para todos os elementos do grupo uma vez que se mostrou muito importante para a nossa aquisição de conhecimentos na área prática da unidade curricular em que se encontra inserido.

Tratando-se de um projeto com um elevado volume de dados e, dados esses, baseados na realidade, apresentados na base de dados *Sakila*, aprimorou a nossa capacidade e percepção dos Processos de Negócio bem como no trabalho em equipa. Foi então possível consolidar as matérias abordadas ao longo deste semestre e também do nosso prévio conhecimento do funcionamento de Bases de Dados.

Também as ferramentas utilizadas foram bastante úteis e práticas, possibilitando a realização com sucesso de todo o trabalho prático.

Anexos

Dimensão Actor

Name	Type of table	Nr. Records		Description				
Dim_Actor	Dimensão	200		Informação referente aos atores				
Target (Data mart)				Source (OLTP)				
Column	Description	Data type	SCD	Table	Column	Data type	ETL rules	Example values of
ActorKey	Chave de ator (Surrogate Key)	Int	-	-	-	-	Surrogate Key	1
Version	Versão	Int	-	-	-	-	-	1
Date_from	Data inicial dos dados (validade)	Varchar(20)	-	-	-	-	-	-
Date_to	Data final dos dados	Varchar(20)	-	-	-	-	-	-
Actor_id	Identificador ator	Int	-	Actor	Actor_id	Int	-	1
First_name	Nome	Varchar(50)	2	Actor	First_name	Varchar(50)	-	Sérgio
Last_name	Apelido	Varchar(50)	2	Actor	Last_name	Varchar(50)	-	Pereira
Last_update	Data da atualização	Varchar(20)	1	Actor	Last_update	Varchar(20)	-	05/05/2021

Figure 19 - Anexo A, dim_actor

Dimensão Category

Name	Type of table	Nr. Records		Description				
Dim_Category	Dimensão	16		Informação referente às categorias de filmes				
Target (Data mart)				Source (OLTP)				
Column	Description	Data type	SCD	Table	Column	Data type	ETL rules	Example values of
CategoryKey	Chave de categoria (Surrogate Key)	Int	-	-	-	-	Surrogate Key	1
Version	Versão	Int	-	-	-	-	-	1
Date_from	Data inicial dos dados (validade)	Varchar(20)	-	-	-	-	-	-
Date_to	Data final dos dados	Varchar(20)	-	-	-	-	-	-
Category_id	Identificador categoria	Int	-	Category	Actor_id	Int	-	1
Name	Nome	Varchar(50)	2	Category	Name	Varchar(50)	-	Ação
Last_update	Data atualização da	Varchar(20)	1	Category	Last_update	Varchar(20)	-	05/05/2021

Figure 20 - Anexo B, dim_category

Dimensão Customer

Name	Type of table	Nr. Records		Description				
Dim_Customer	Dimensão	599		Informação referente aos clientes				
Target (Data mart)				Source (OLTP)				
Column	Description	Data type	SCD	Table	Column	Data type	ETL rules	Example values of
CustomerKey	Chave de cliente (Surrogate Key)	Int	-	-	-	-	Surrogate Key	1
Version	Versão	Int	-	-	-	-	-	1
Date_from	Data inicial dos dados (validade)	DateTime	-	-	-	-	-	1900-01-01
Date_to	Data final dos dados	DateTime	-	-	-	-	-	2200-01-01
Customer_id	Identificador cliente	Int	-	Customer	Customer_id	Int	-	1
Address	Linha 1 do endereço	Varchar(50)	1	Address	Address	Varchar(50)	-	Rua do Rossio
Address2	Linha 2 do endereço	Varchar(50)	1	Address	Address2	Varchar(50)	-	Nº5, 1º esq.
City	Cidade	Varchar(50)	1	City	City	Varchar(50)	-	Cidade
Country	País	Varchar(50)	1	Country	Country	Varchar(50)	-	Portugal

District	Concelho	Varchar(20)	1	Address	District	Varchar(20)	-	Lisboa
Postal_Code	Código Postal	Varchar(10)	1	Address	Postal_Code	Varchar(10)	-	4000-001
Phone	Telefone	Varchar(20)	1	Address	Phone	Varchar(20)	-	912345678
First_name	Nome	Varchar(45)	1	Staff	First_name	Varchar(45)	-	José
Last_name	Apelido	Varchar(45)	1	Staff	Last_name	Varchar(45)	-	Ferreira
Email	E-mail	Varchar(50)	1	Customer	Email	Varchar(50)	-	123@gmail.com
Active	Flag: ativo	bit	2	Customer	Active	bit	-	1
Create_date	Data de criação	DateTime	-	Customer	Create_date	DateTime	-	01/01/2021
Last_update	Data atualização da	DateTime	1	-	-	-	-	05/05/2021

Figure 21 - Anexo C, dim_customer

Dimensão Film

Name	Type of table	Nr. Records		Description				
Dim_film	Dimensão	10000		Informação referente aos filmes				
Target (Data mart)				Source (OLTP)				
Column	Description	Data type	SCD	Table	Column	Data type	ETL rules	Example of values
FilmKey	Chave de film (Surrogate Key)	Int	-	-	-	-	Surrogate Key	1
Version	Versão	Int	-	-	-	-	-	1
Date_from	Data inicial dos dados (validade)	DateTime	-	-	-	-	-	1900-01-01
Date_to	Data final dos dados	DateTime	-	-	-	-	-	2200-01-01
Film_ID	Identificador Filme	Int	-	Film	Film_ID	Int	-	1
Title	Título do filme	Varchar(255)	1	Film	Title	Varchar(255)	-	Titanic
Release_year	Ano em que saiu	Varchar(4)	1	Film	Release_year	Varchar(4)	-	1997
Description	Descrição	Text	1	Film	Description	Text	-	Navio afunda

Length	Duração	Int	1	Film	Length	Int	-	194
Rating	Classificação	Varchar(10)	1	Film	Rating	Varchar(10)	-	M/12
Language	Linguagem	Varchar(20)	1	Language	Name	Varchar(20)	-	Portugues
Original_Language	Linguagem Original	Varchar(20)	1	Language	Name	Varchar(20)	-	Ingles
Rental_Duration	Duração aluguer	Int	1	Film	Rental_Duration	Int	-	10
Rental_Rate	Custo aluguer	Decimal(4, 2)	1	Film	Rental_Rate	Decimal(4, 2)	-	4,99
Replacement_cost	Custo substituição	Decimal(5, 2)	1	Film	Replacement_cost	Decimal(5, 2)	-	49,99
Special_features	Caraterísticas especiais	Varchar(255)	1	Film	Special_features	Varchar(255)	-	3D
Last_update	Data da atualização	DateTime	1	-	-	-	-	05/05/2021

Figure 22 - Anexo D, dim_film

Dimensão Staff

Name	Type of table	Nr. Records		Description				
Dim_staff	Dimensão	2		Informação referente aos funcionários				
Target (Data mart)				Source (OLTP)				
Column	Description	Data type	SCD	Table	Column	Data type	ETL rules	Example values of
StaffKey	Chave de staff (Surrogate Key)	Int	-	-	-	-	Surrogate Key	1
Version	Versão	Int	-	-	-	-	-	1
Date_from	Data inicial dos dados (validade)	DateTime	-	-	-	-	-	1900-01-01
Date_to	Data final dos dados	DateTime	-	-	-	-	-	2200-01-01
Staff_id	Identificador staff	Int	-	Staff	Staff_id	Int	-	1
Address	Linha 1 do endereço	Varchar(50)	1	Address	Address	Varchar(50)	-	Rua do Rossio
Address2	Linha 2 do endereço	Varchar(50)	1	Address	Address2	Varchar(50)	-	Nº5, 1º eq.
City	Cidade	Varchar(50)	1	City	City	Varchar(50)	-	Cidade
Country	País	Varchar(50)	1	Country	Country	Varchar(50)	-	Portugal


District	Concelho	Varchar(20)	1	Address	District	Varchar(20)	-	Lisboa
Postal_Code	Código Postal	Varchar(10)	1	Address	Postal_Code	Varchar(10)	-	4000-001
Phone	Telefone	Varchar(20)	1	Address	Phone	Varchar(20)	-	912345678
First_name	Nome	Varchar(45)	1	Staff	First_name	Varchar(45)	-	José
Last_name	Apelido	Varchar(45)	1	Staff	Last_name	Varchar(45)	-	Ferreira
Picture	Foto	Binary(MAX)	1	Staff	Picture	Binary(MAX)	-	
Email	E-mail	Varchar(50)	1	Staff	Email	Varchar(50)	-	123@gmail.com
Active	Flag: ativo	bit	2	Staff	Active	bit	-	1
Username	Nome de utilizador	Varchar(16)	1	Staff	Username	Varchar(20)	-	Staff1
Password	Password	Varchar(40)	1	Staff	Password	Varchar(20)	-	Password1
Last_update	Data da atualização	DateTime	1	-	-	-	-	05/05/2021

Figure 23 - Anexo E, dim_staff

Dimensão Store

Name	Type of table	Nr. Records		Description				
Dim_store	Dimensão	2		Informação referente às lojas				
Target (Data mart)				Source (OLTP)				
Column	Description	Data type	SCD	Table	Column	Data type	ETL rules	Example values of
StoreKey	Chave de store (Surrogate Key)	Int	-	-	-	-	Surrogate Key	1
Version	Versão	Int	-	-	-	-	-	1
Date_from	Data inicial dos dados (validade)	DateTime	-	-	-	-	-	1900-01-01
Date_to	Data final dos dados	DateTime	-	-	-	-	-	2200-01-01
Store_id	Identificador loja	Int	-	Store	Store_id	Int	-	1
Address	Linha 1 do endereço	Varchar(50)	1	Address	Address	Varchar(50)	-	Rua do Rossio
Address2	Linha 2 do endereço	Varchar(50)	1	Address	Address2	Varchar(50)	-	Nº5, 1º esq.
City	Cidade	Varchar(50)	1	City	City	Varchar(50)	-	Cidade

Country	País	Varchar(50)	1	Country	Country	Varchar(50)	-	Portugal
District	Concelho	Varchar(20)	1	Address	District	Varchar(20)	-	Lisboa
Postal_Code	Código Postal	Varchar(10)	1	Address	Postal_Code	Varchar(10)	-	4000-001
Phone	Telefone	Varchar(20)	1	Address	Phone	Varchar(20)	-	912345678
Manager_Staff ID	Identificador do Gerente	Int	1	Store				
Last_update	Data da atualização	DateTime	1	-	-	-	-	05/05/2021

Figure 24 - Anexo F, dim_store