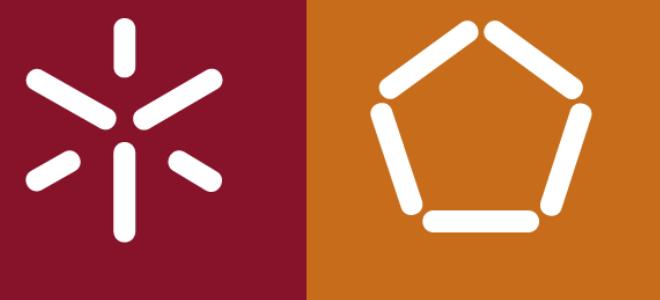


Análise de Dados

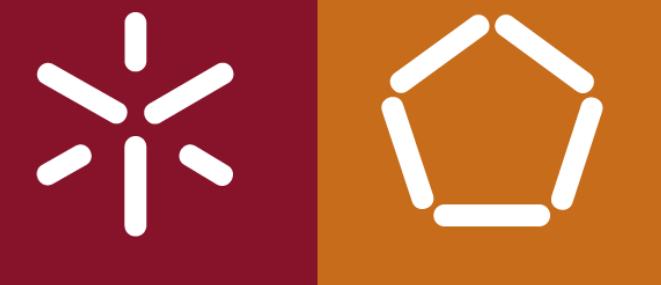
TP4

Daniela Oliveira



TP4

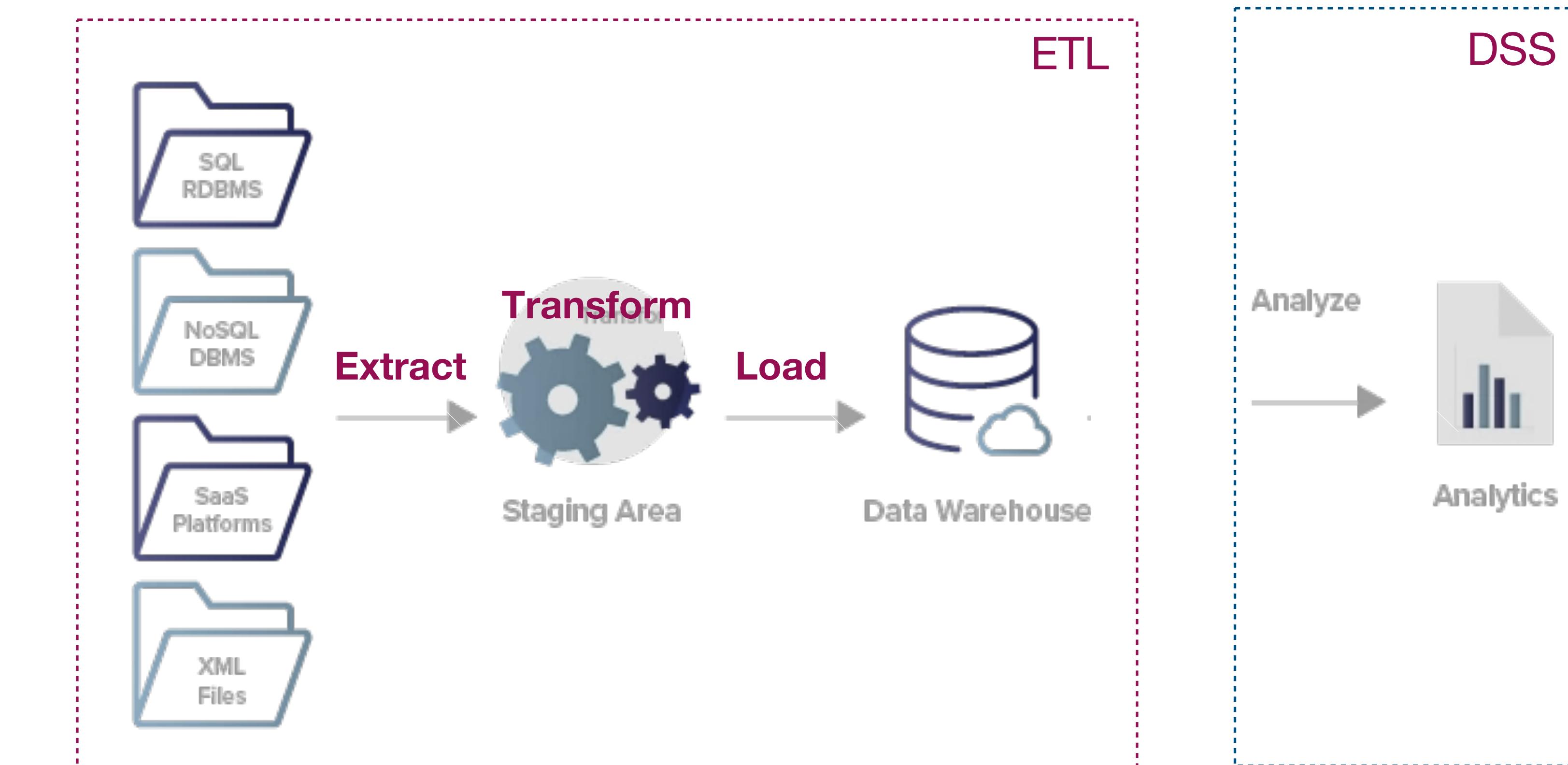
- Revisão dos conceitos de ETL e *Data warehousing*;
- Contextualização do software *Tableau Desktop*;
- Resolução da 4^a ficha TP individual.



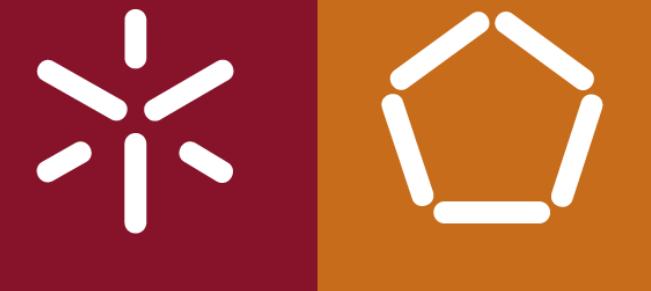
ETL

(Extract, Transform, Load)

Definição



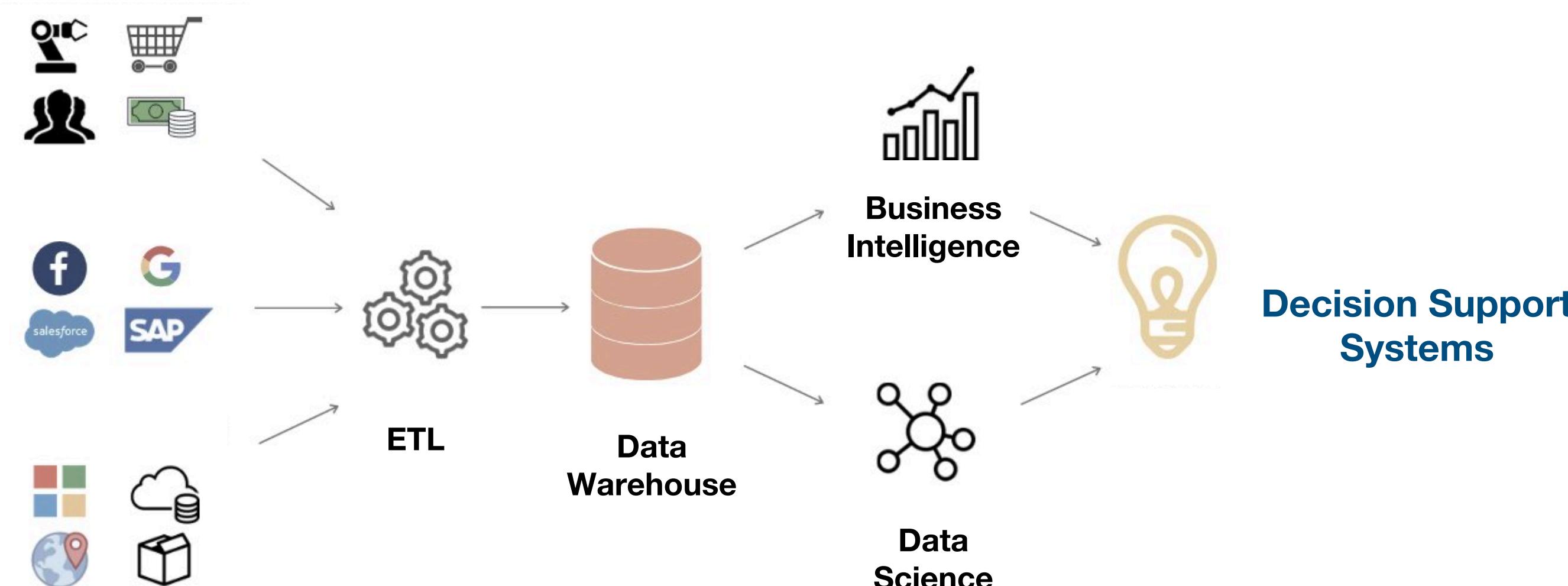
- Diferentes fontes de informação
- Constante aumento do volume de dados
- Diferentes níveis de estruturação de dados

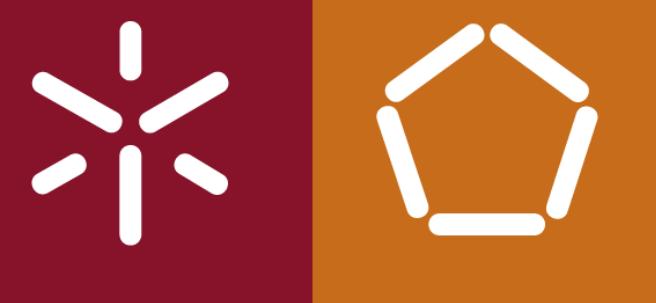


Data Warehouse

Definição

- Define-se como um sistema de agregação de dados derivados de diferentes fontes de informação, para suporte à decisão num determinado contexto;
- Estabelece correlação entre os dados das diferentes fontes;
- Revela-se o núcleo dos processos de Business Intelligence e Data Science;
- Permite a implementação de sistemas DBMS (Database Management System);

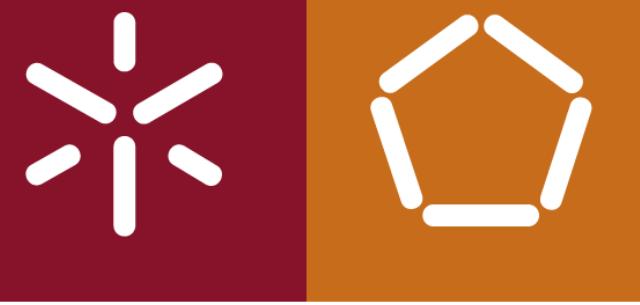




Data Warehouse

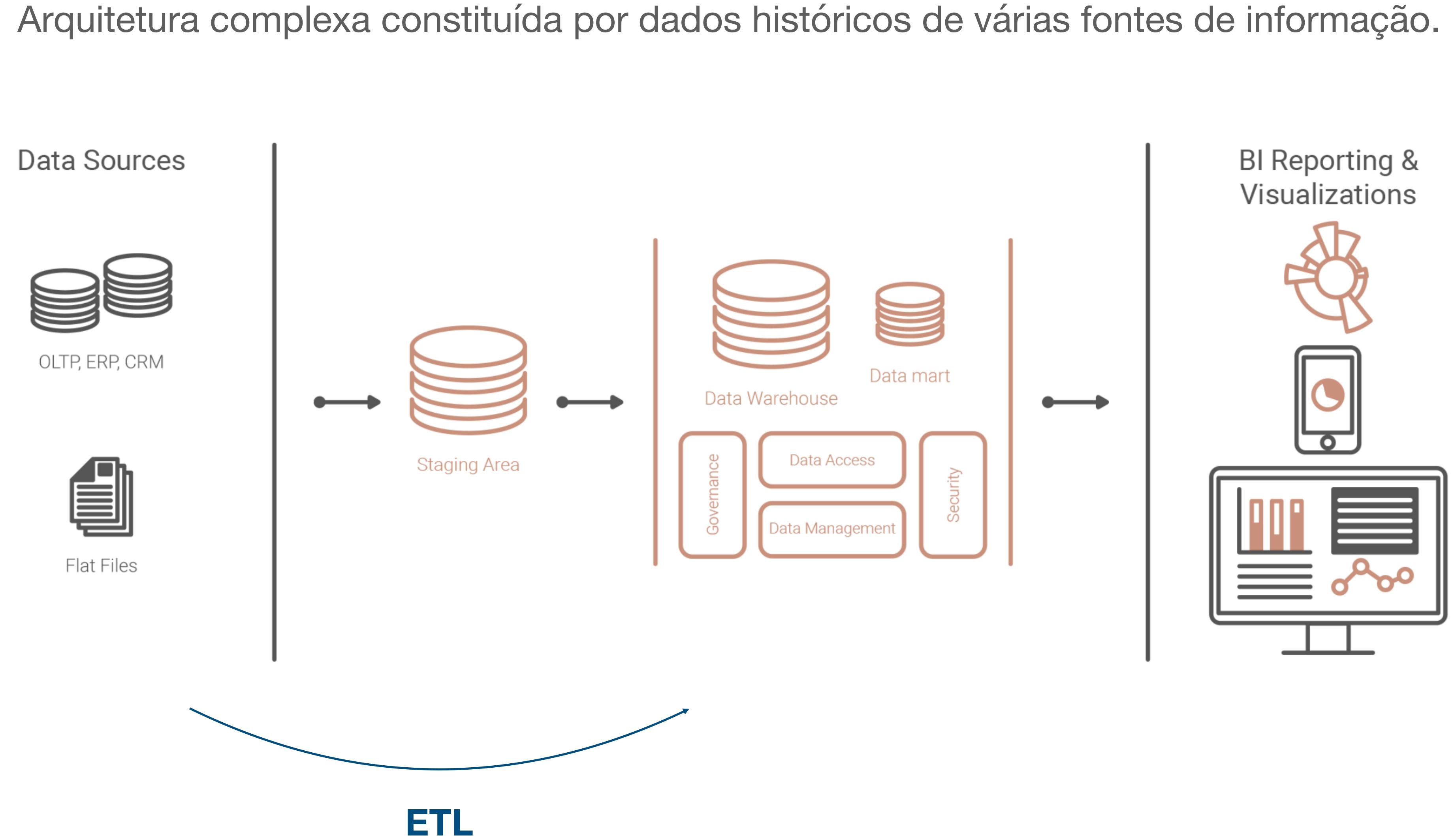
Características

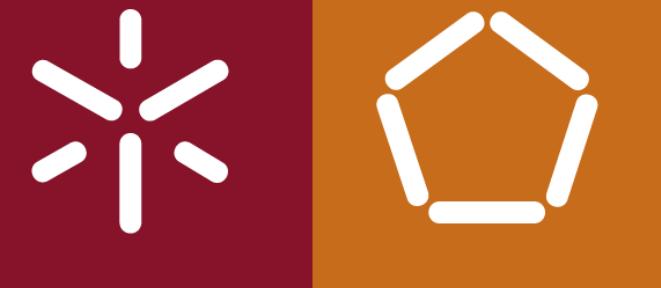
- **Subject-Oriented**
 - Orientado à modelação e análise de dados para a tomada de decisões;
 - Visão simples e concisa sobre um assunto específico para apoiar o processo de decisão.
- **Integrated**
 - Unidade de medida comum para dados provenientes de diferentes bases de dados;
 - Consistência de nomenclaturas, formatos e codificação dos dados;
- **Time-variant**
 - Dados estão relacionados com um determinado período de tempo, revelando informações do ponto de vista histórico;
 - Uma vez inseridos os dados no *data warehouse*, estes não sofrem mutações.
- **Non-volatile**
 - Os dados não são apagados aquando inserção de novos;
 - Apenas operações de *loading* e de *access* são permitidas sobre um data warehouse.



Data Warehouse

Arquitetura





Data Warehouse

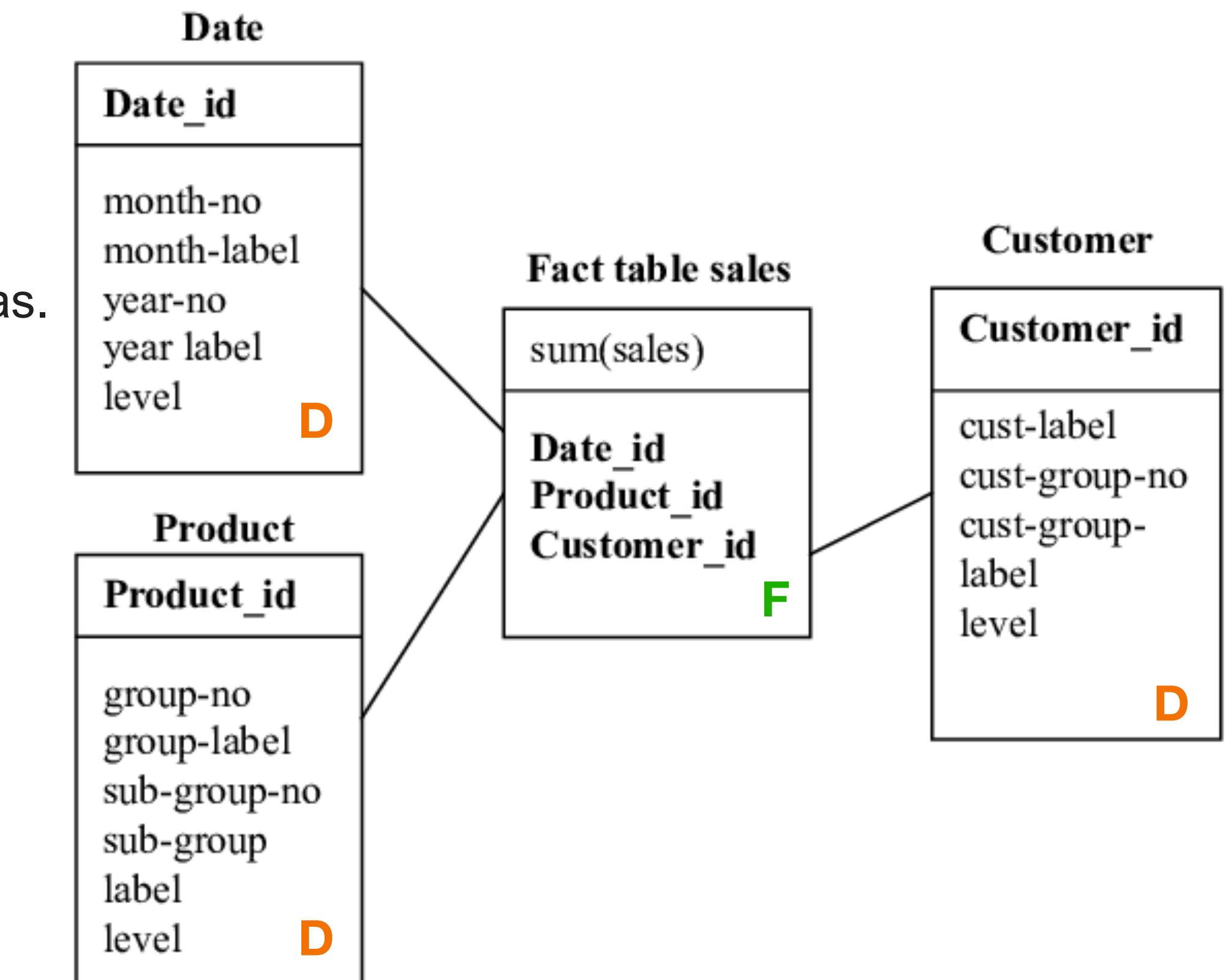
Modelação Dimensional

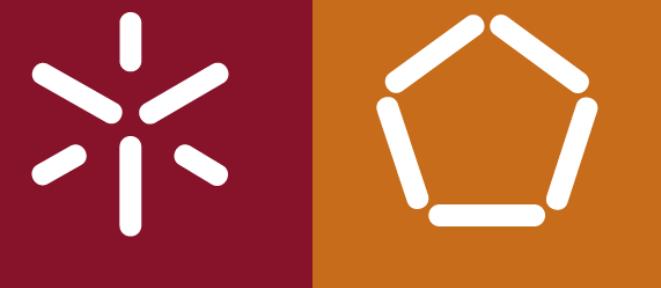
Fact table:

- Medidas, métrica e factos sobre um determinado processo de negócio;
- *Foreign keys* das correspondentes tabelas de dimensão;
- Não contém hierarquia.

Dimension table:

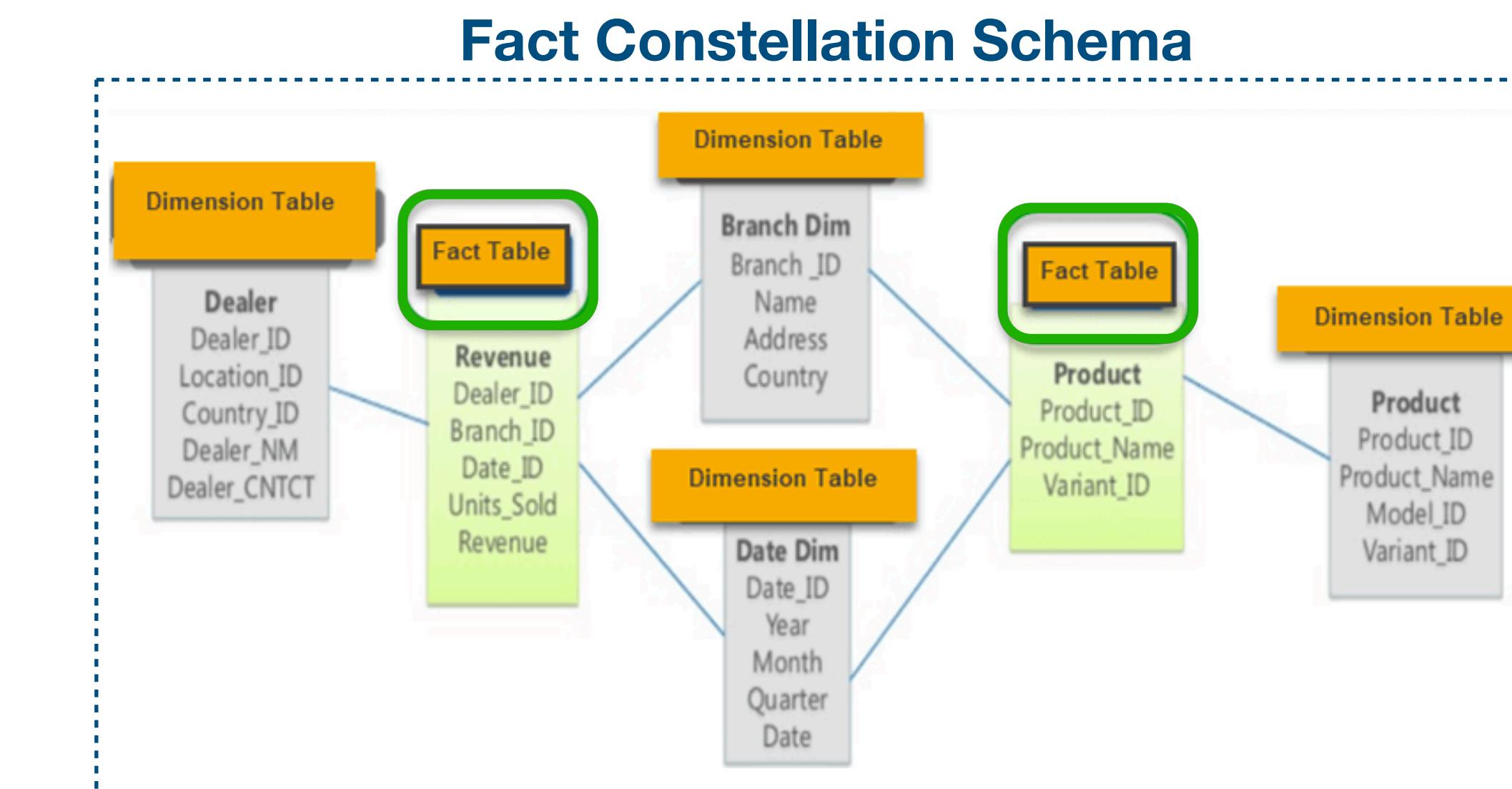
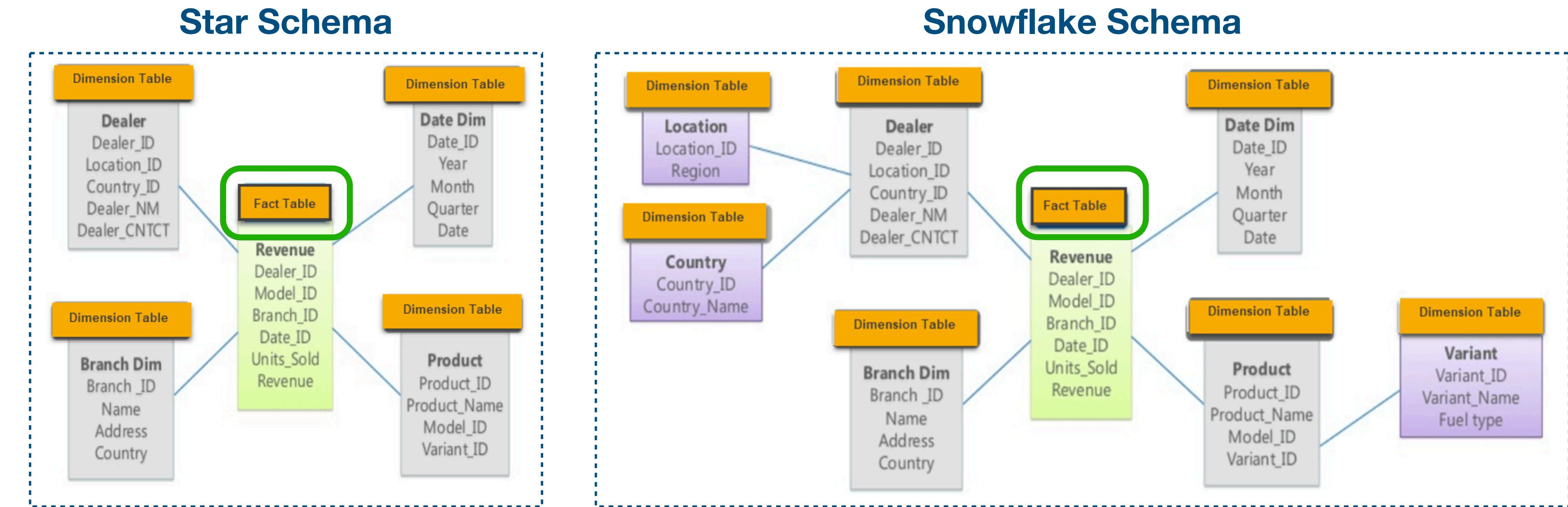
- Características descritivas sobre os factos;
- Pode conter uma ou mais relações hierárquicas.

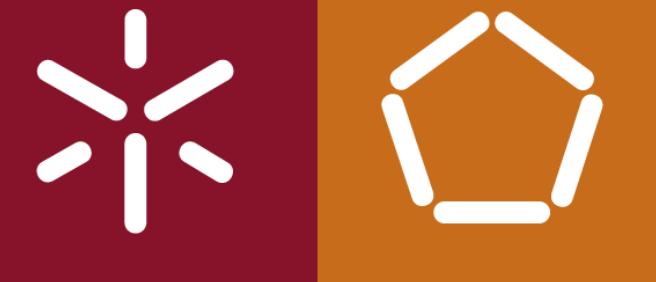




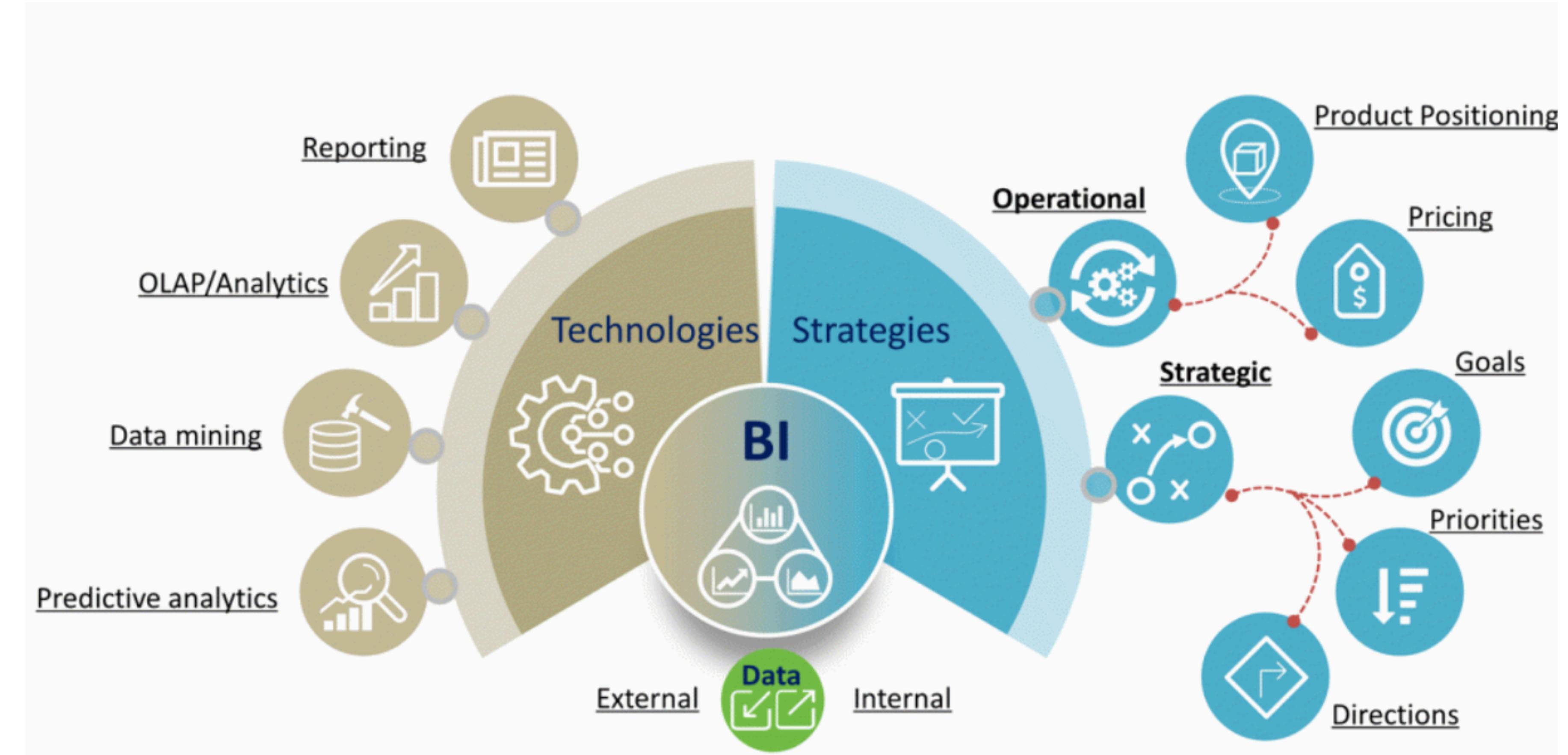
Data Warehouse

Modelação Dimensional

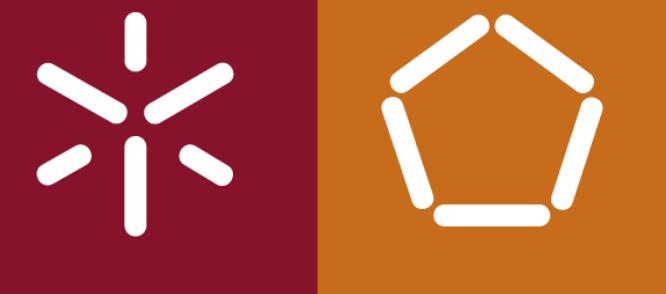




Business Intelligence



<https://yourfreetemplates.com/business-intelligence-infographic/>



Business Intelligence



CAMADA LÓGICA

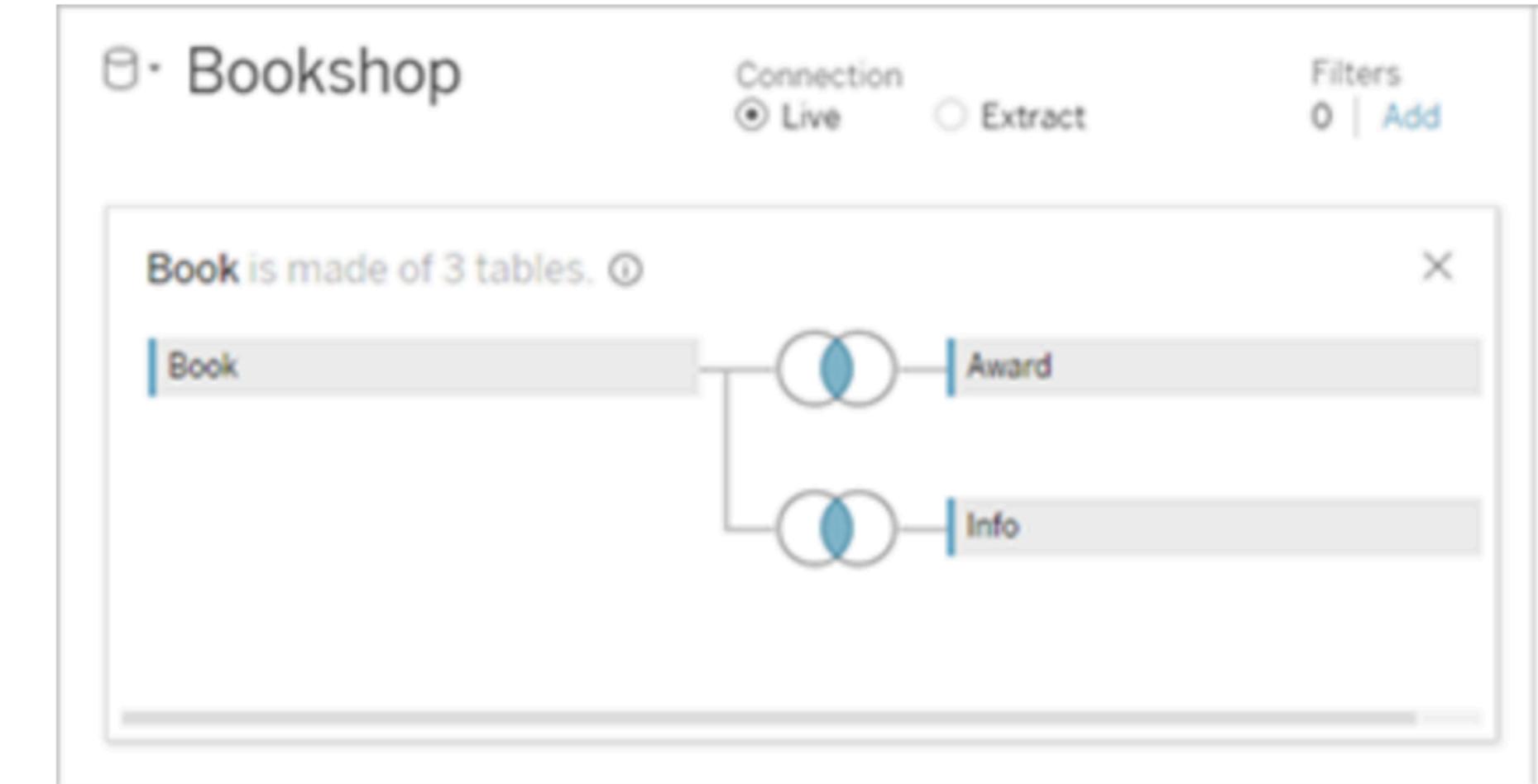
Noodles = Relações



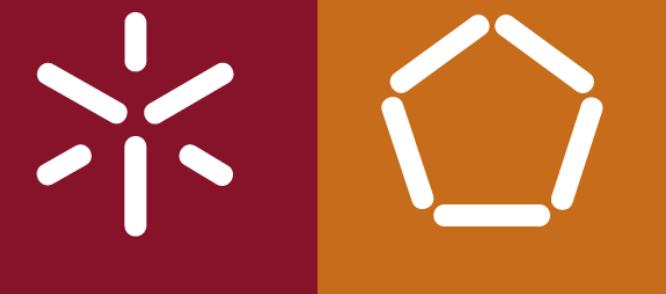
A exibição de nível superior de uma fonte de dados com várias tabelas relacionadas. Esta é a camada lógica. As tabelas lógicas podem ser combinadas usando relações (noodles). Elas não usam tipos de união. Elas atuam como contêineres de tabelas físicas.

CAMADA FÍSICA

Diagrama de Venn = União de colunas



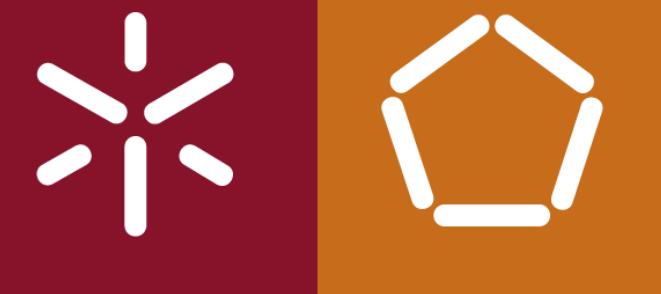
Clique duas vezes em uma tabela lógica para abri-la e ver as tabelas físicas. As tabelas físicas podem ser combinadas usando uniões de colunas e de linhas. Neste exemplo, a tabela lógica Livro consiste em três tabelas físicas com união de colunas (Livro, Prêmio, Informação).



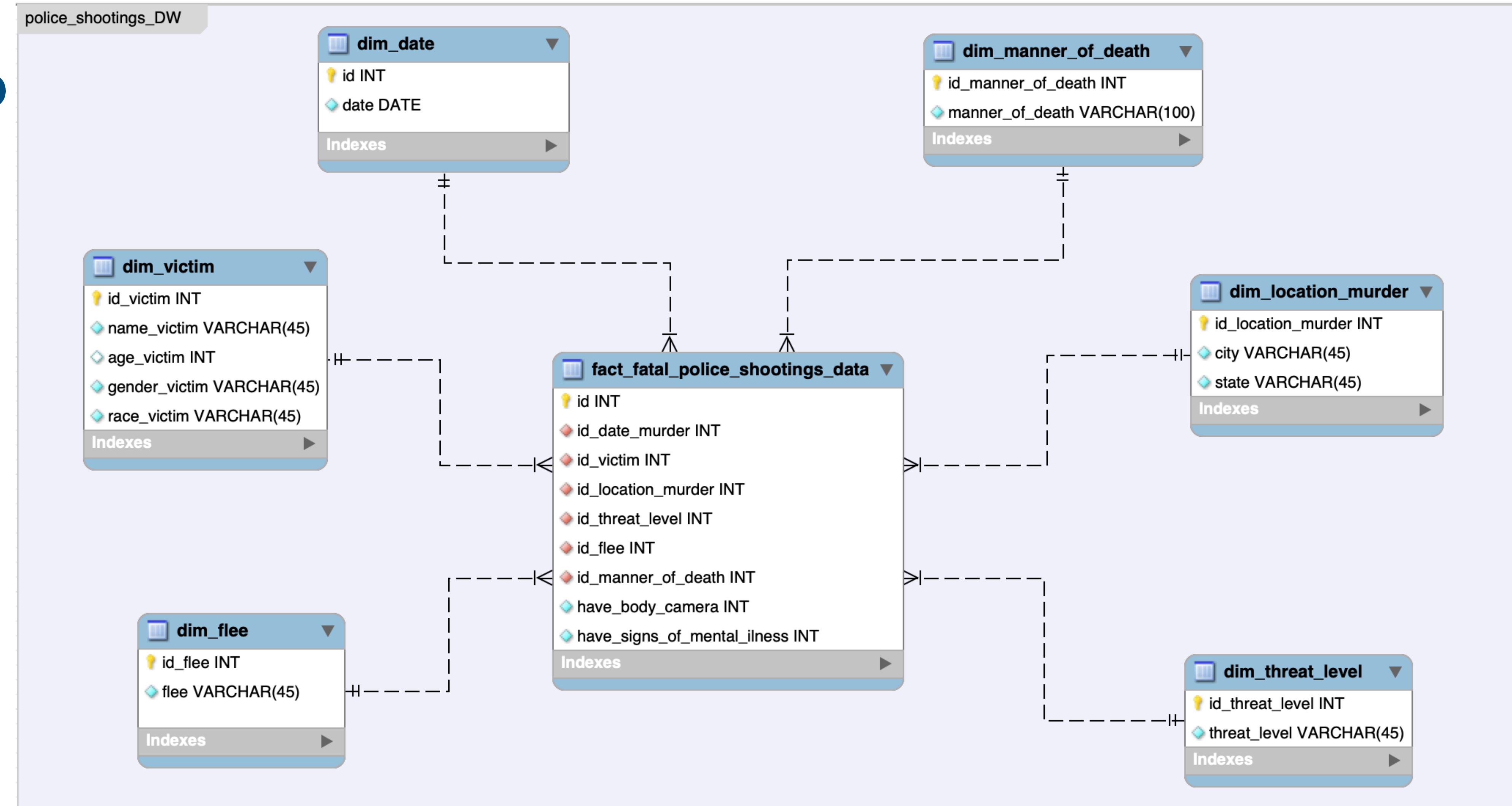
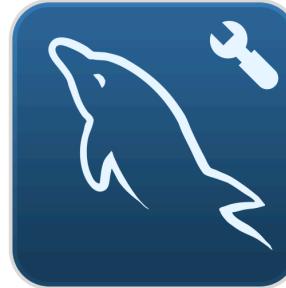
Business Intelligence

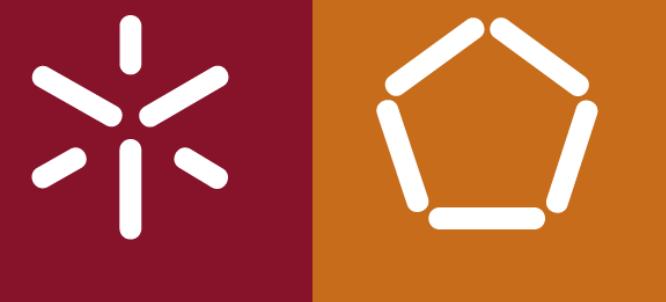


CAMADA LÓGICA	CAMADA FÍSICA
Tela das Relações na página Fonte de dados	Tela da União de colunas/União de linhas na página Fonte de dados
Tabelas lógicas	Tabelas físicas
As tabelas lógicas podem estar relacionadas a outras tabelas lógicas	As tabelas físicas podem unir colunas ou linhas de outras tabelas físicas
As tabelas lógicas são semelhantes a containers de tabelas físicas	Duplo click sobre a tabela lógica para ver as tabelas físicas
As tabelas lógicas permanecem distintas (normalizadas) na fonte de dados	As tabelas físicas são podem juntar-se em apenas uma única tabela simples que define a tabela lógica



Exemplo





Mestrado Integrado em Engenharia Informática, 4º ano

Exemplo



Screenshot of the Tableau Data Source interface for the "pol_usa_murders" database.

Conexões: localhost MySQL

Banco de dados: pol_usa_murders

Tabela:

- dim_date
- dim_flee
- dim_location_murder
- dim_manner_of_death
- dim_threat_level
- dim_victim
- fact_fatal_police_shootings_data
- Novo SQL personalizado
- Nova união

Diagrama de Relacionamento:

```

    graph LR
        fact[fact_fatal_police_shootings_data] --> dim_date
        fact --> dim_flee
        fact --> dim_location_murder
        fact --> dim_manner_of_death
        fact --> dim_threat_level
        fact --> dim_victim
    
```

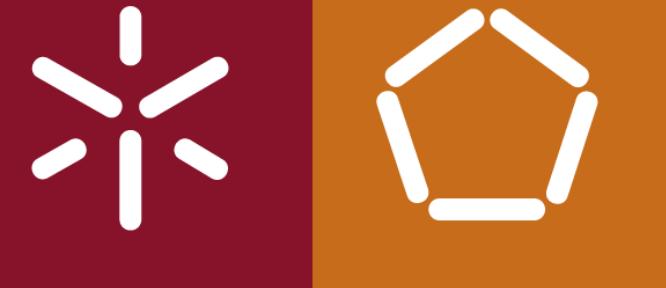
Filtros: 0 | Adicionar

Conexão: Em tempo real

Visualização de Dados:

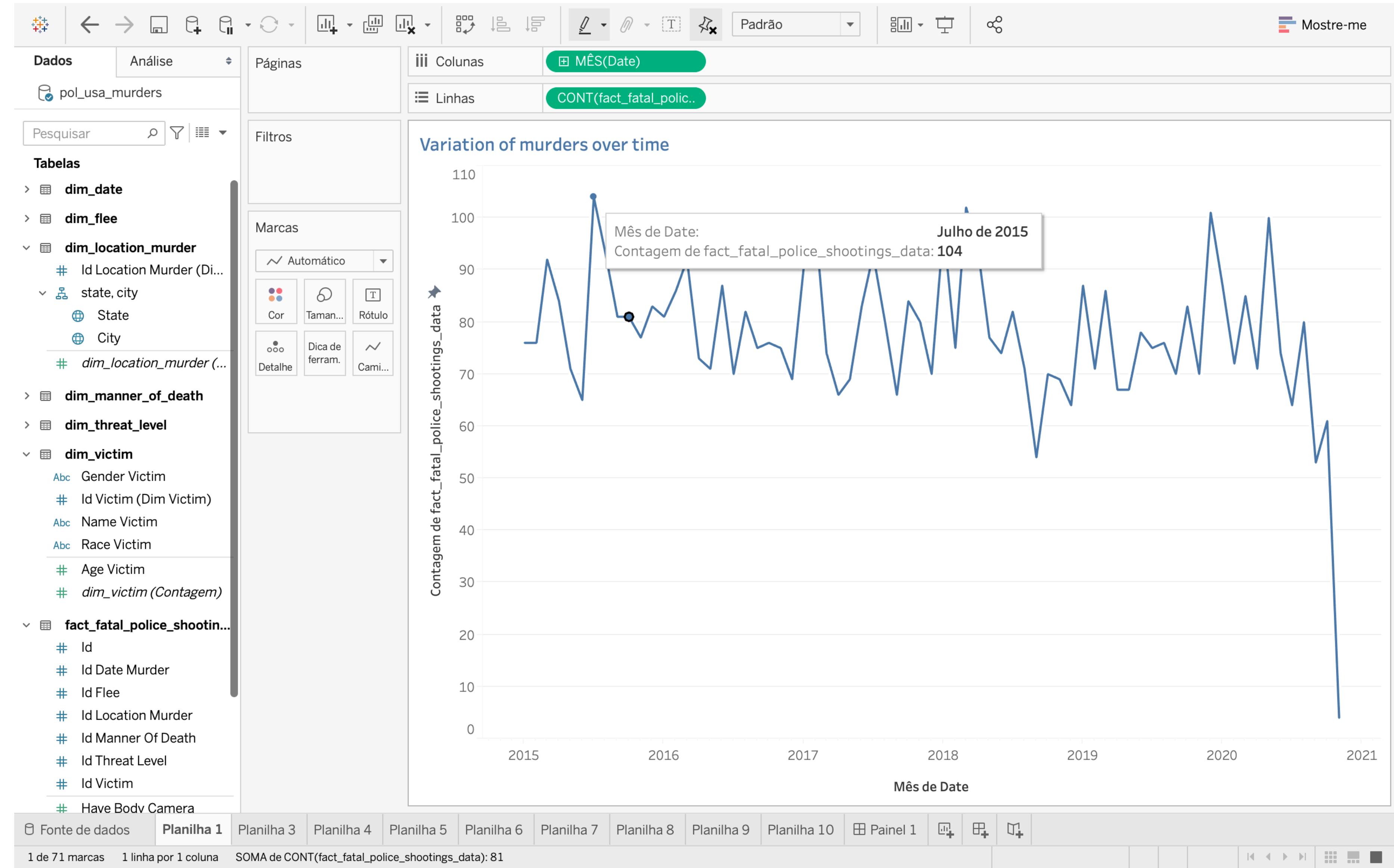
#	fact_fatal_... Id	fact_fatal_police_shootin... Id Date Murder	fact_fatal_police_... Id Victim	fact_fatal_police_shootings_data Id Location Murder	fact_fatal_police_shootin... Id Threat Level	fact_fatal_pol... Id Flee	fact_fatal_police_shootings_data Id Manner Of Death	fact_fatal_police_shootings_d... Have Body Camera	fact_... Ha
1		2	3	1	1	1	1	No	
2		2	4	2	1	1	1	No	
3		3	5	3	2	1	2	No	

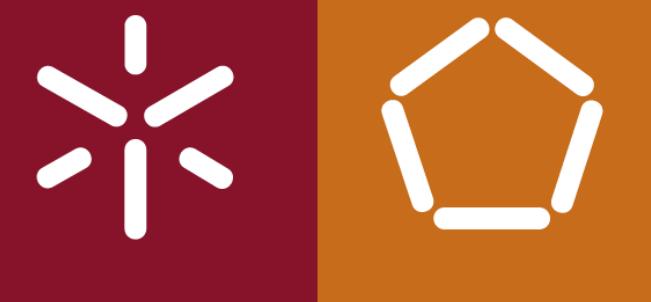
Opções: Classificar campos, Ordem das fontes de dados, Mostrar aliases, Mostrar campos ocultos, 1 000 linhas.



Mestrado Integrado em Engenharia Informática, 4º ano

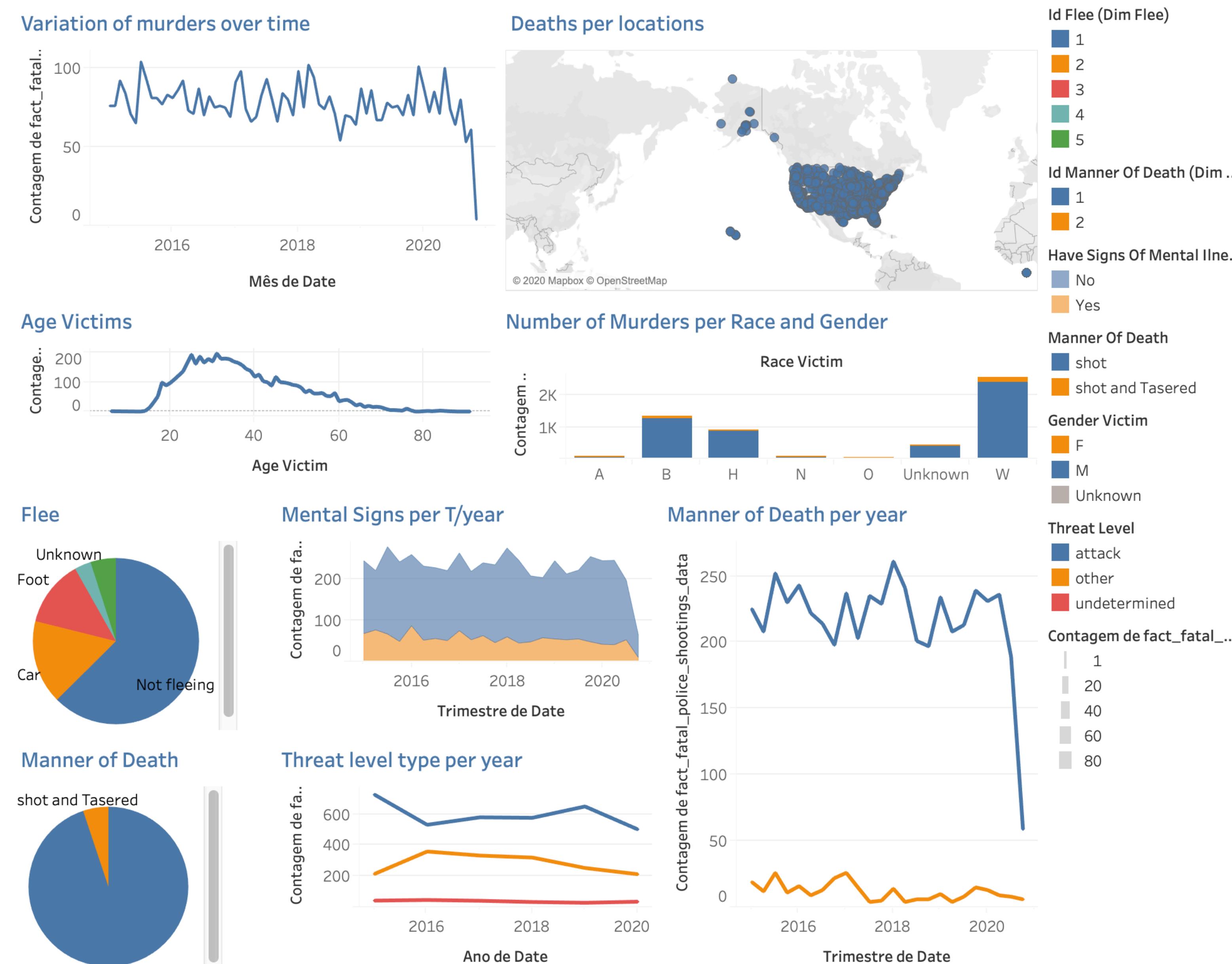
Exemplo

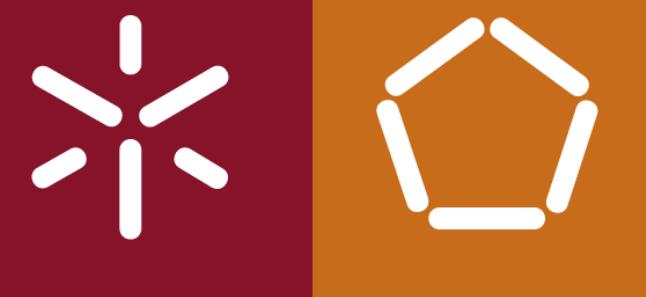




Mestrado Integrado em Engenharia Informática, 4º ano

Exemplo Painel





Send to: daniela.oliveira@algoritmi.uminho.pt with subject: “AD/TP4/pgXXXX”

Ficha N.º 4

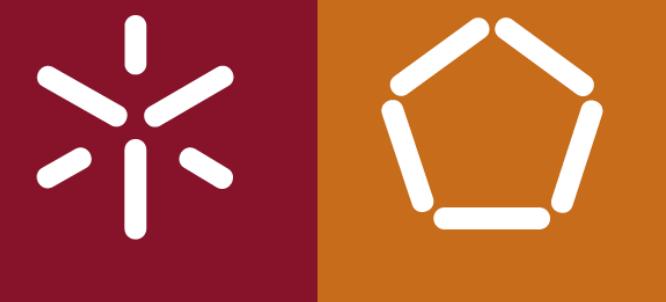
1 *Data Warehousing e Business Intelligence - Netflix*

1.1 Parte I - *Data Warehousing*

Desenvolva um processo de *data warehousing*, passando pelas etapas do processo ETL e convertendo os dados presentes no ficheiro *netflix_dataset.csv* para um modelo dimensional.

Ferramentas: MySQL e MySQL Workbench

1. Analisar o *dataset netflix_dataset.csv*;
2. Criar um modelo lógico a partir da análise dos dados disponibilizados, com a sua tabela de factos e respetivas tabelas de dimensões;
3. Converter o modelo lógico para o seu modelo físico através da opção *Forward Engineer*;
4. Povoar as tabelas com os dados presentes nos ficheiro *netflix_dataset.csv*.

**ETL****E**

CSV File

T

Null values treatment
Split column values
Create Functions/Procedures to describe values or acronyms
Analyze ratings and number of votes

L

INSERT INTO SELECT



Send to: daniela.oliveira@algoritmi.uminho.pt with subject: “AD/TP4/pgXXXX”

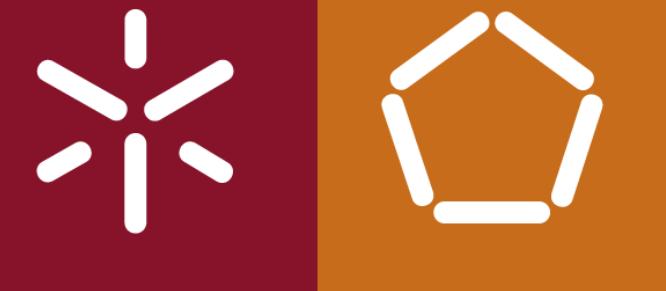
Ficha N.º 4

1.2 Parte II - *Business Intelligence*

Ferramentas: Tableau Desktop

É necessária a instalação do software *Tableau Desktop* e os seus respetivos *Drivers* para estabelecer uma nova *Connection MySQL*.

1. Selecionar a fonte de dados criada na parte I;
2. Estabelecer as relações entre as tabelas no modelo lógico do *Tableau*;
3. Criar indicadores relativos aos dados em estudo;
4. Criar um Painel com os diferentes indicadores.



Bibliografia

- Kimball, Ralph, and Margy Ross. ***The data warehouse toolkit: The definitive guide to dimensional modeling.*** John Wiley & Sons, 2013.
- <https://www.mysqltutorial.org/>
- <https://www.tableau.com/learn/get-started>