



# Análise de Dados

## TP5

Daniela Oliveira

## TP5

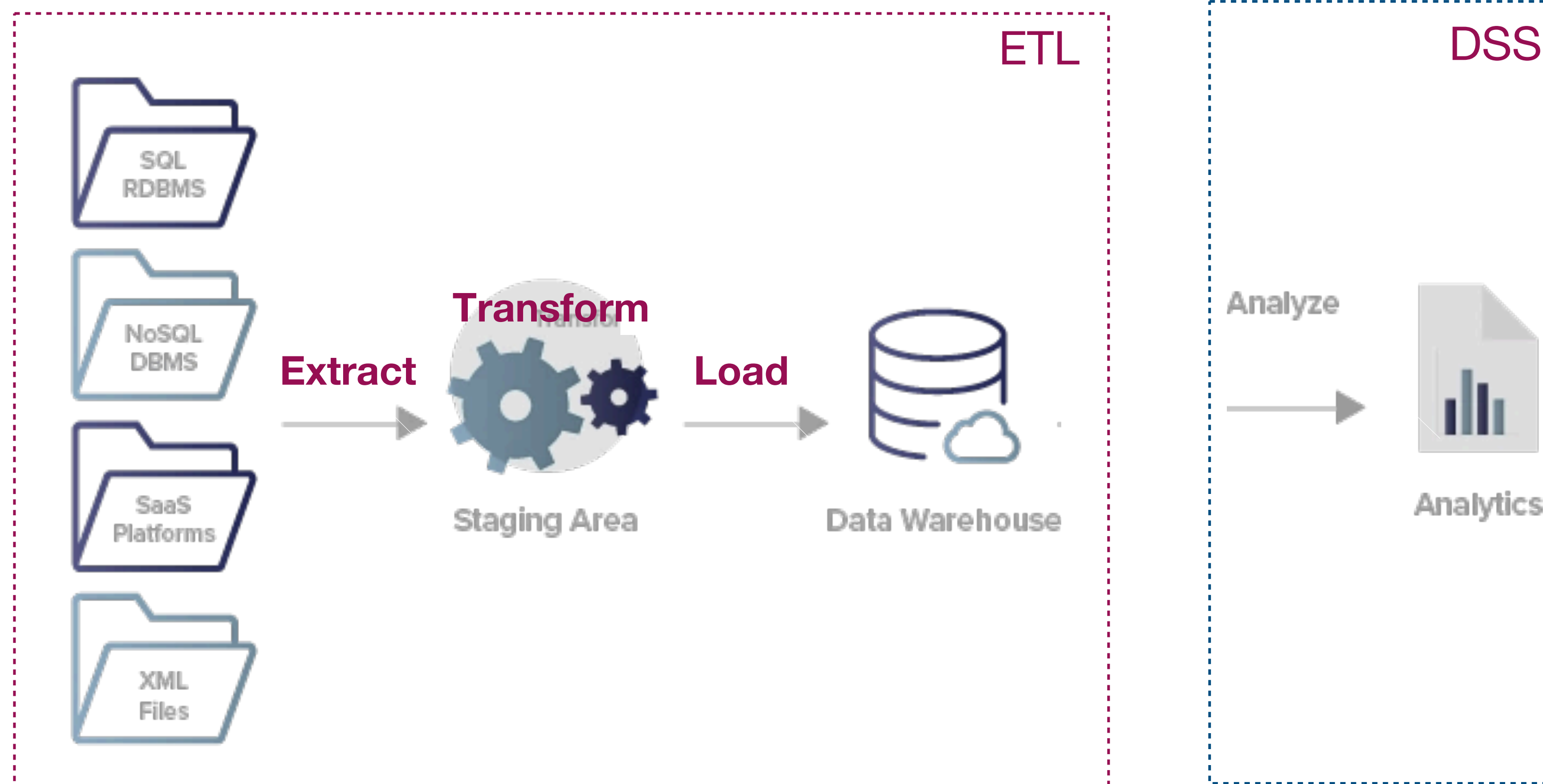
- Revisão dos conceitos de ETL;
- Contextualização do software *PowerBI Desktop*;
- Resolução da 5ª ficha TP individual.



# ETL

(Extract, Transform, Load)

## Definição



Diferentes fontes de informação

Constante aumento do volume de dados

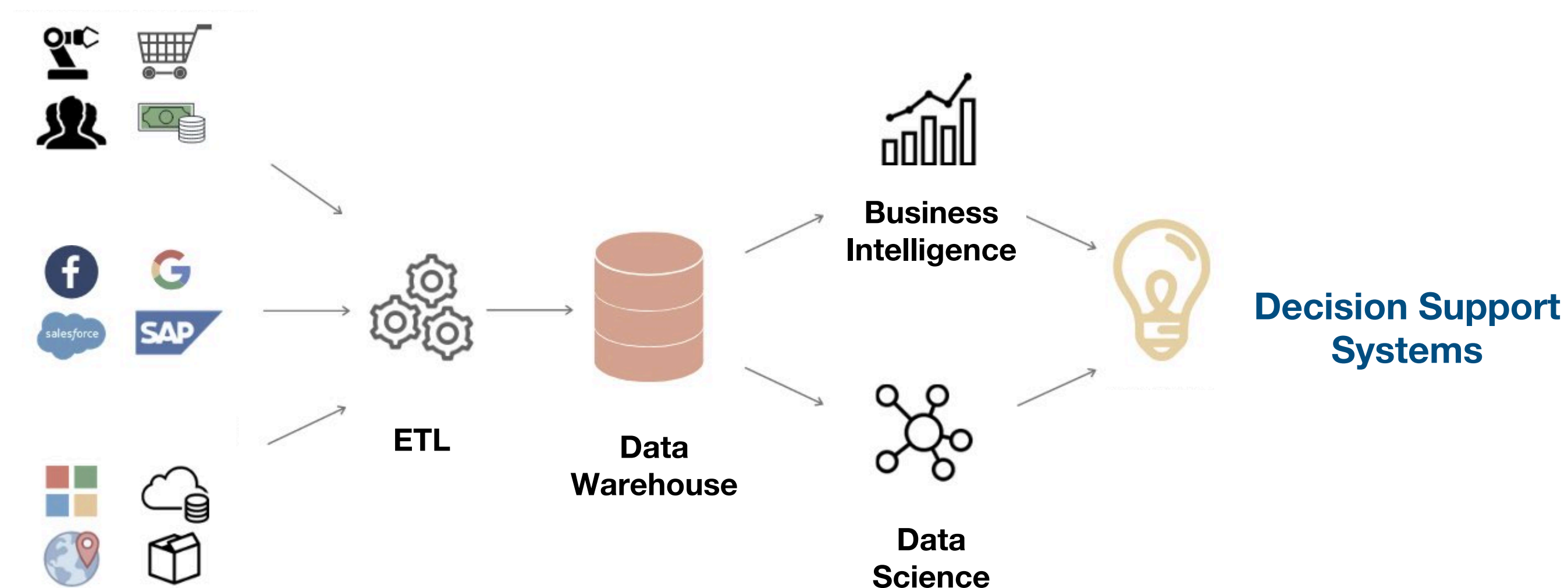
Diferentes níveis de estruturação de dados

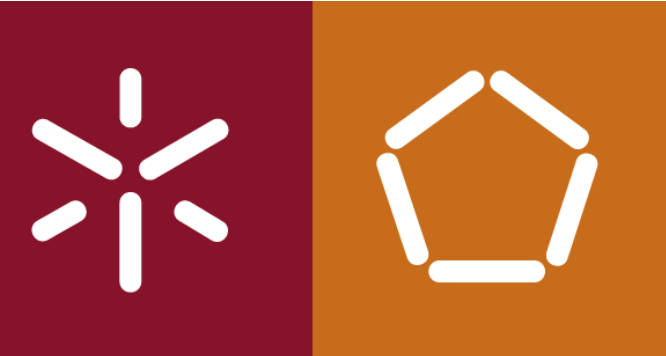


# Data Warehouse

## 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.

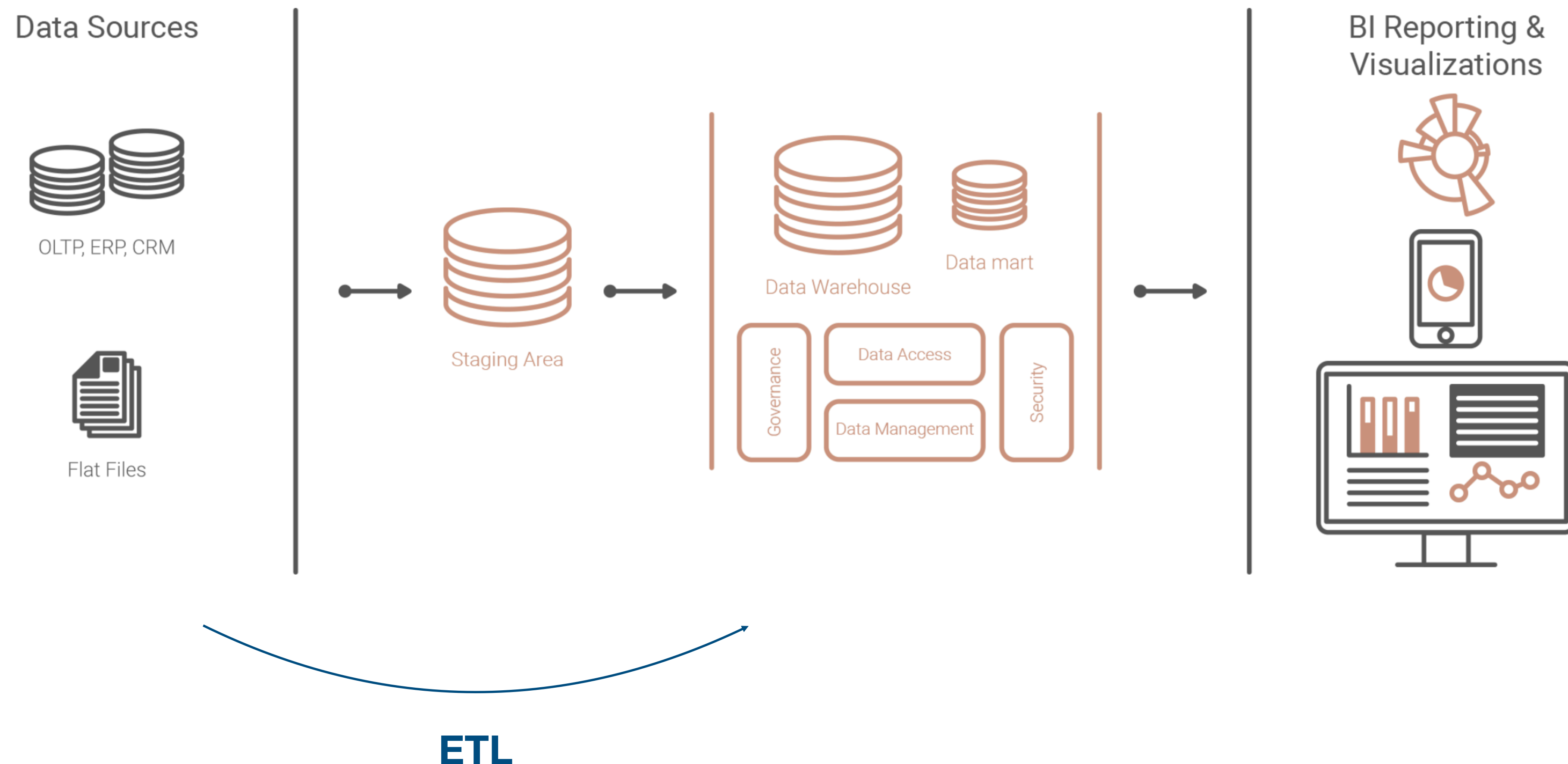


## Mestrado Integrado em Engenharia Informática, 4º ano

Arquitetura complexa constituída por dados históricos de várias fontes de informação.

# Data Warehouse

## Arquitetura







# Data Warehouse

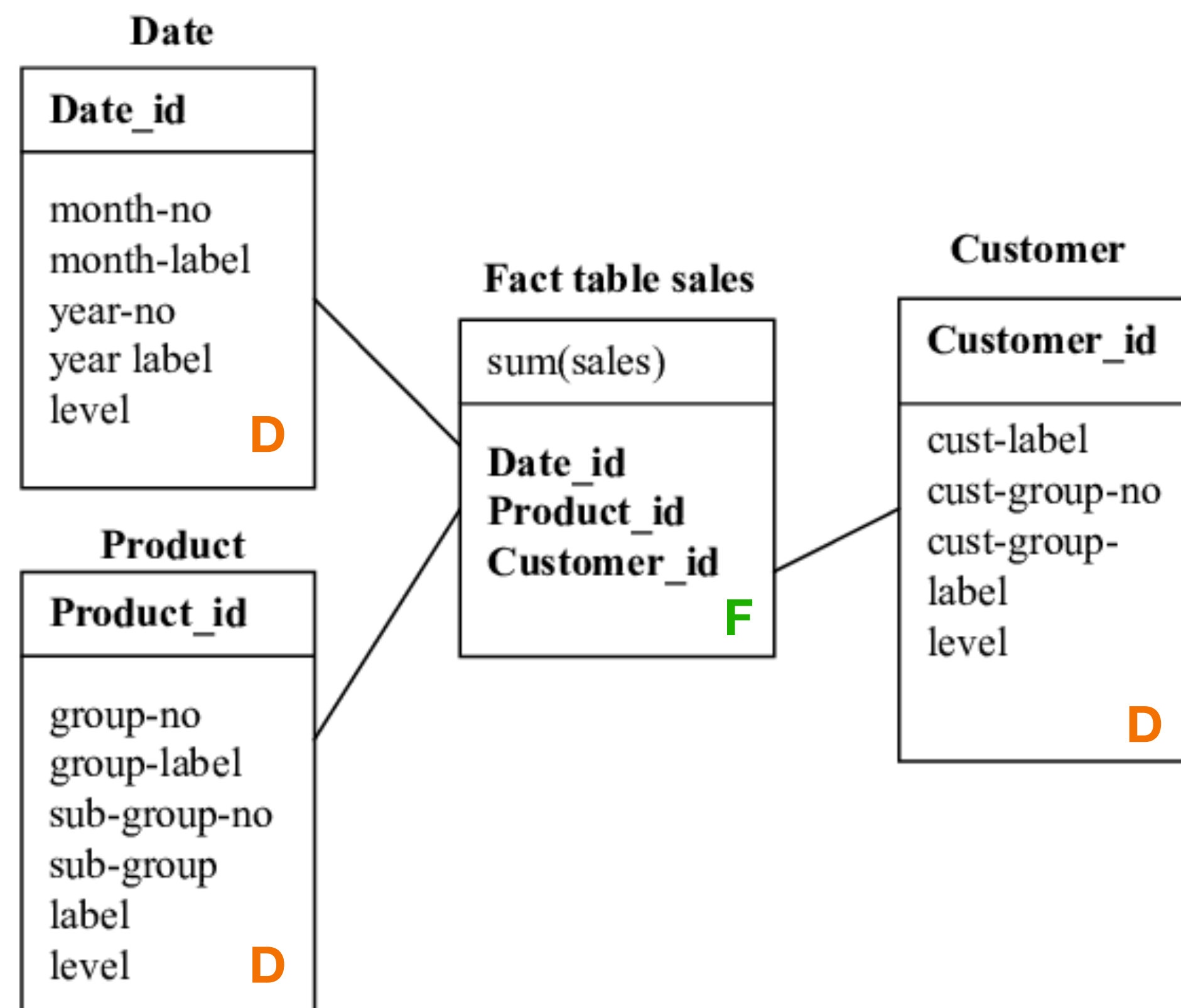
## Modelação Dimensional

### 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.

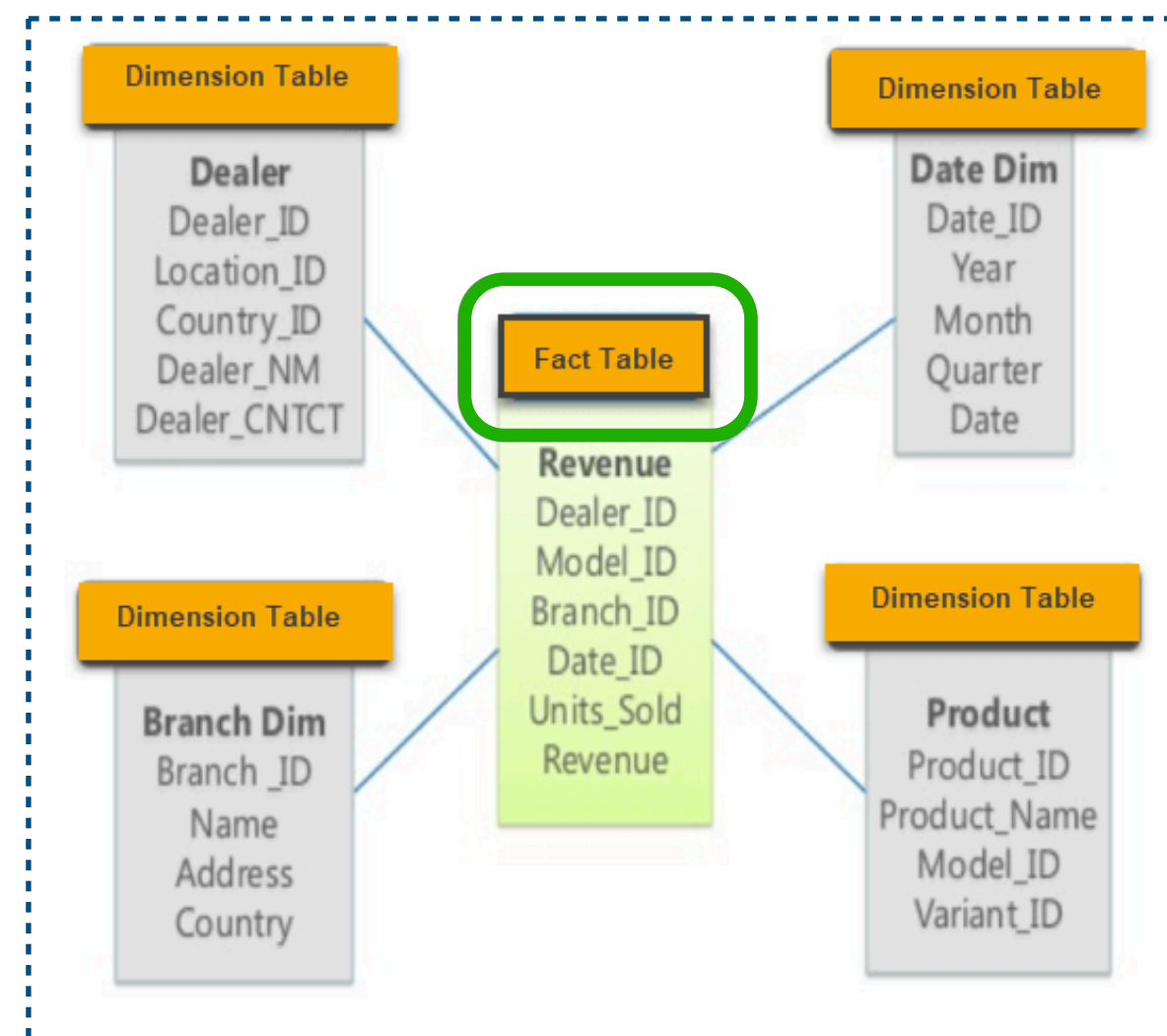




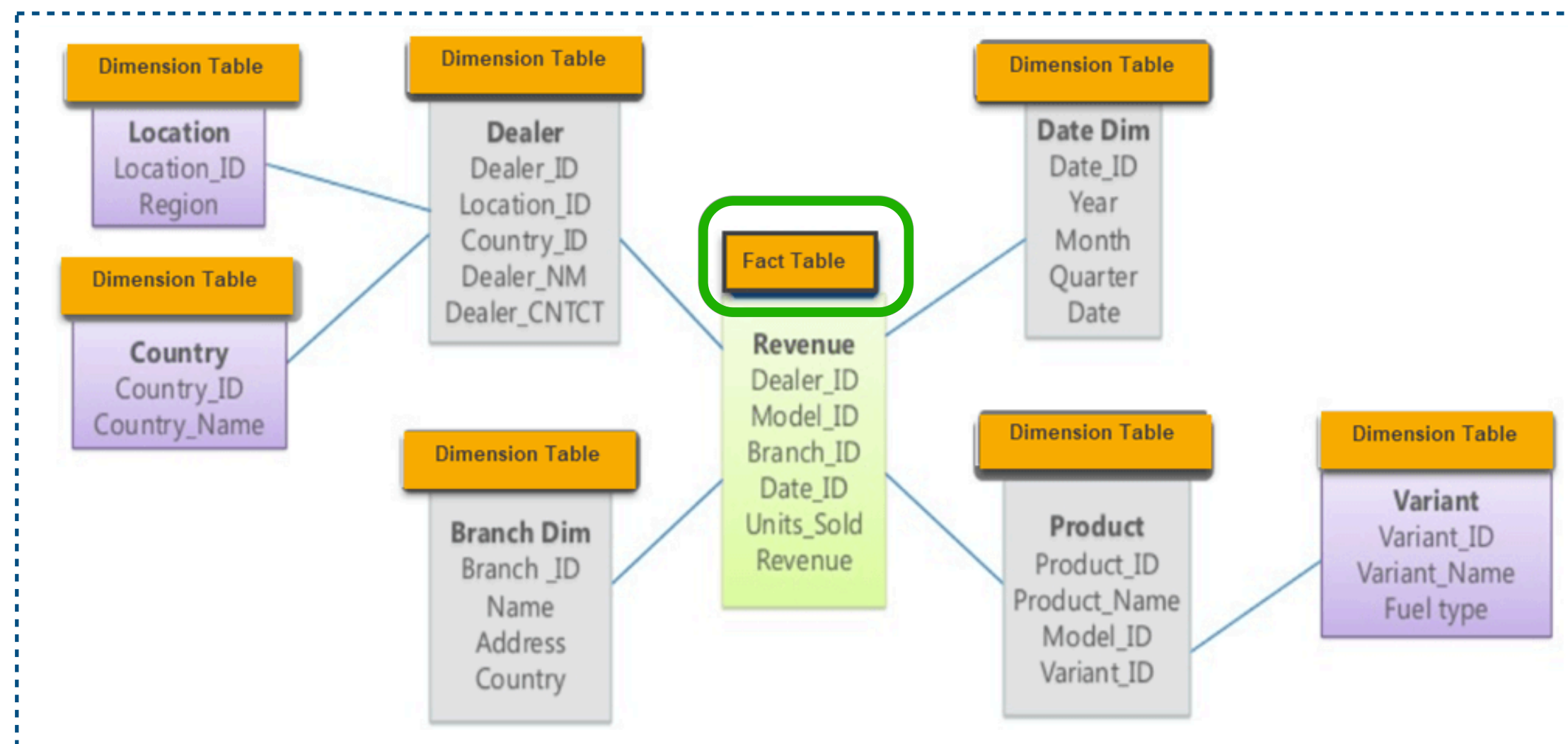
# Data Warehouse

## Modelação Dimensional

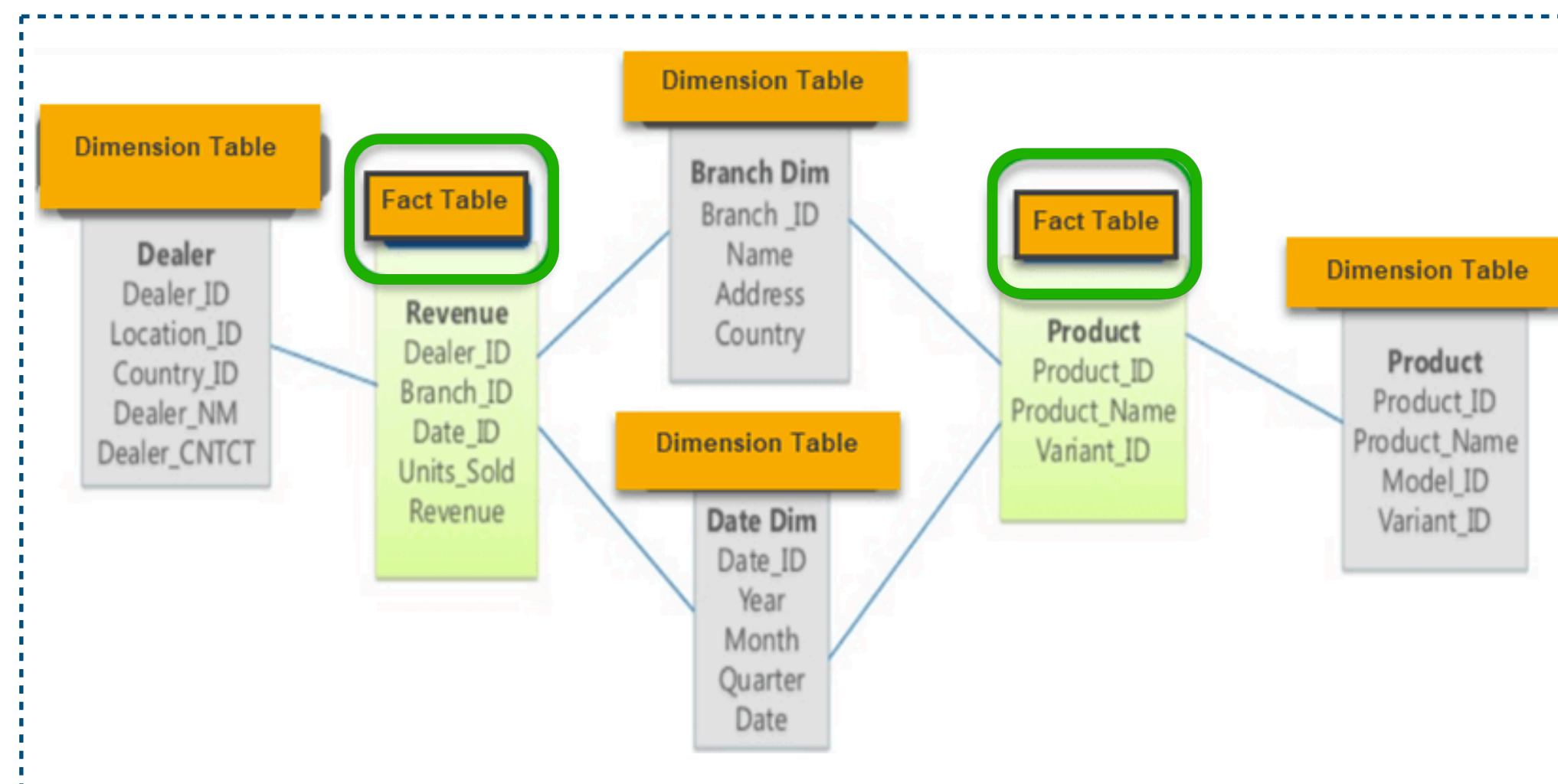
### Star Schema



### Snowflake Schema



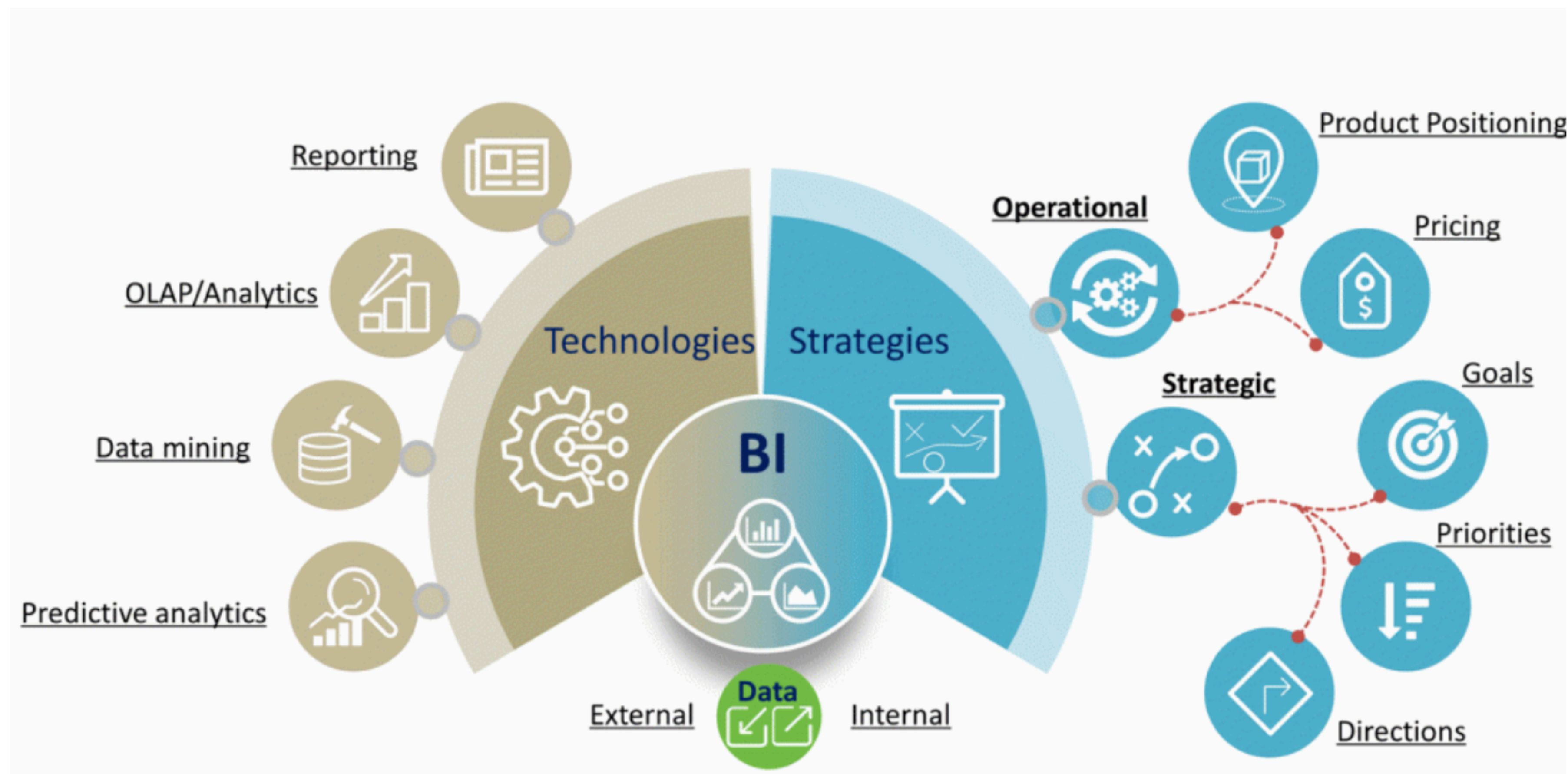
### Fact Constellation Schema







# Business Intelligence



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



# Business Intelligence

## Magic Quadrant for Analytics and BI



Source: Gartner (February 2020)

# Business Intelligence



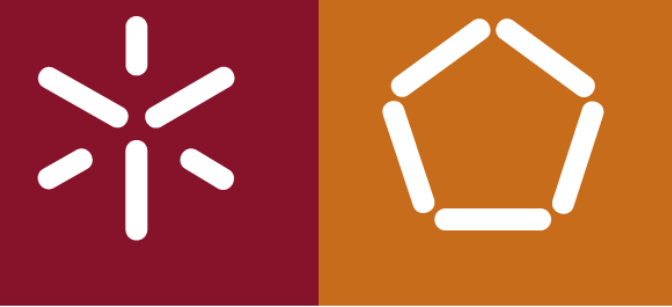
- Ferramenta para análise de negócios;
- Permite a visualização de dados e partilha de informação gerada;
- Permite a partilha de informação com outros utilizadores e através da sua incorporação aplicacional;
- Permite a ligação a diversos tipos de fontes de informação;
- Permite criar novo conhecimento através de *dashboards* e relatórios.

[Criar conta com o e-mail institucional](#)

[Login Power BI Online](#)

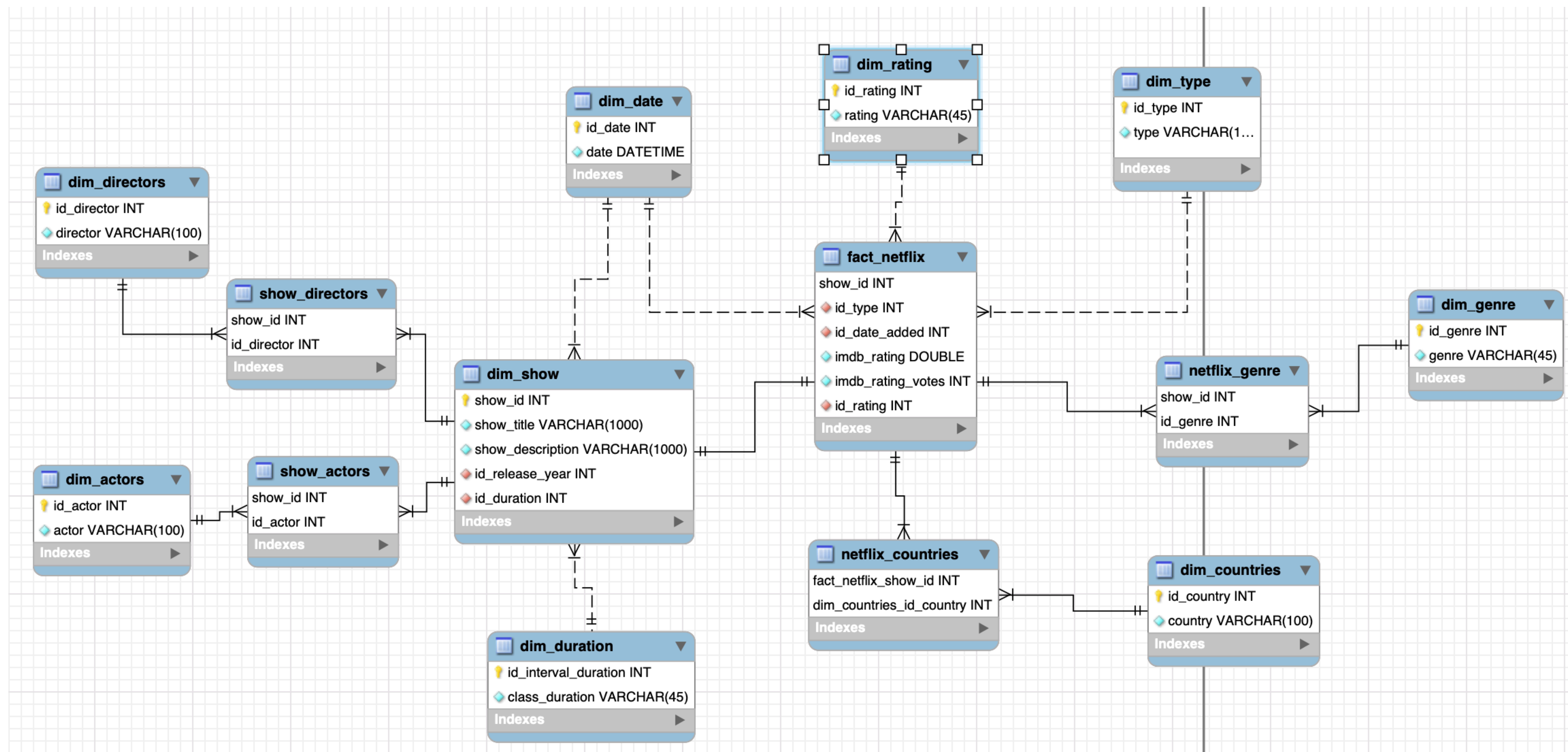
[Download Power BI Desktop](#)

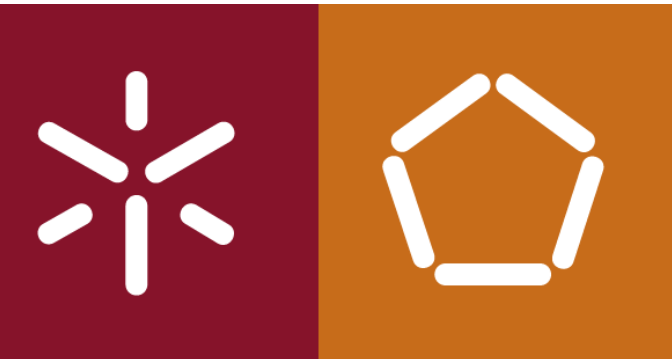




# Mestrado Integrado em Engenharia Informática, 4º ano

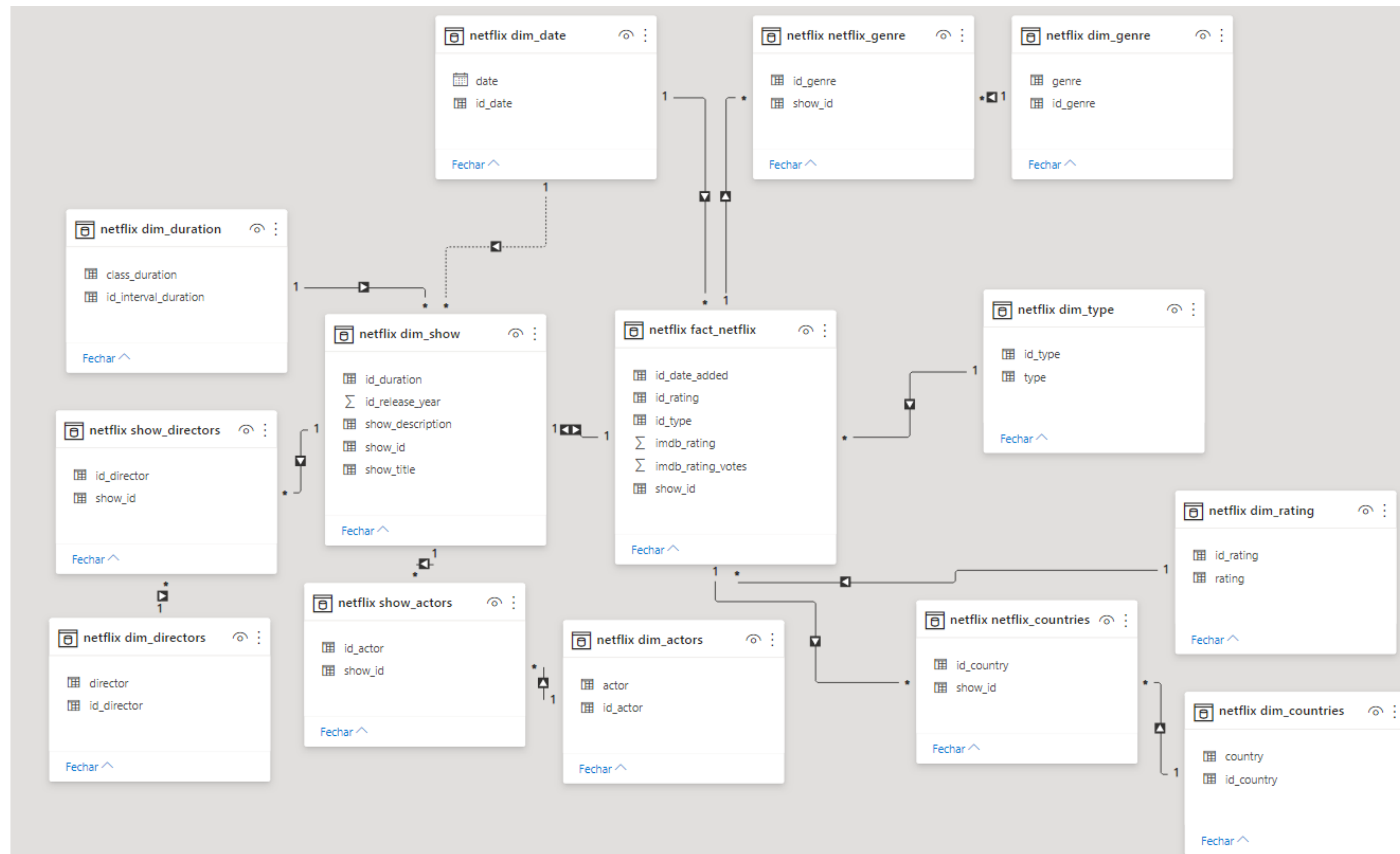
## Exemplo



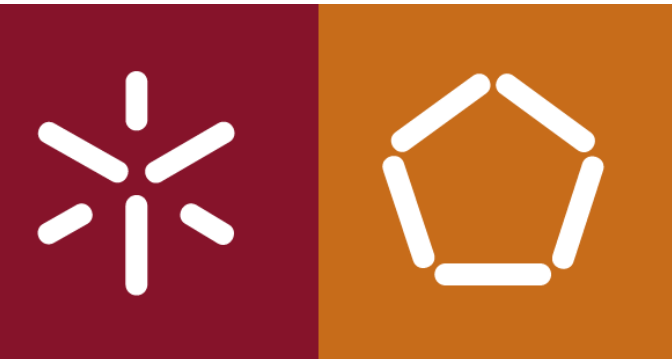


# Mestrado Integrado em Engenharia Informática, 4º ano

## Exemplo





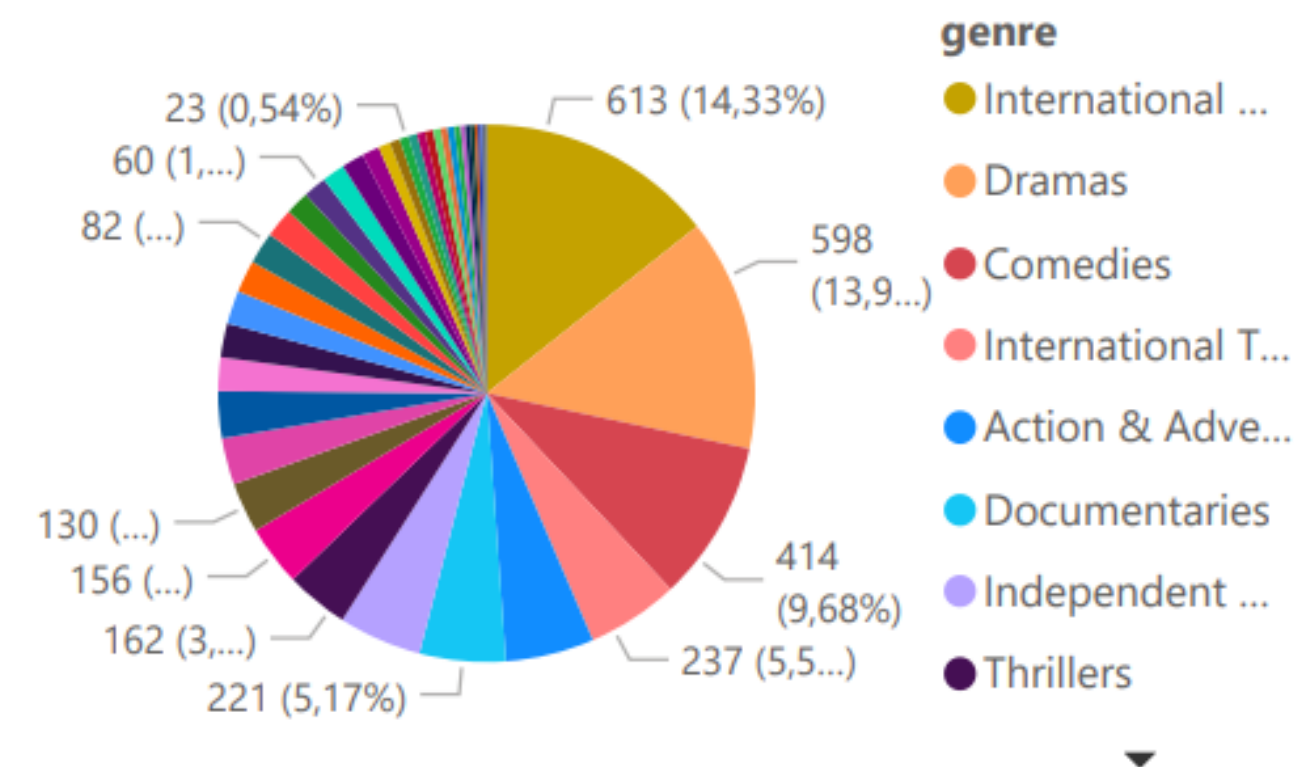


# Mestrado Integrado em Engenharia Informática, 4º ano

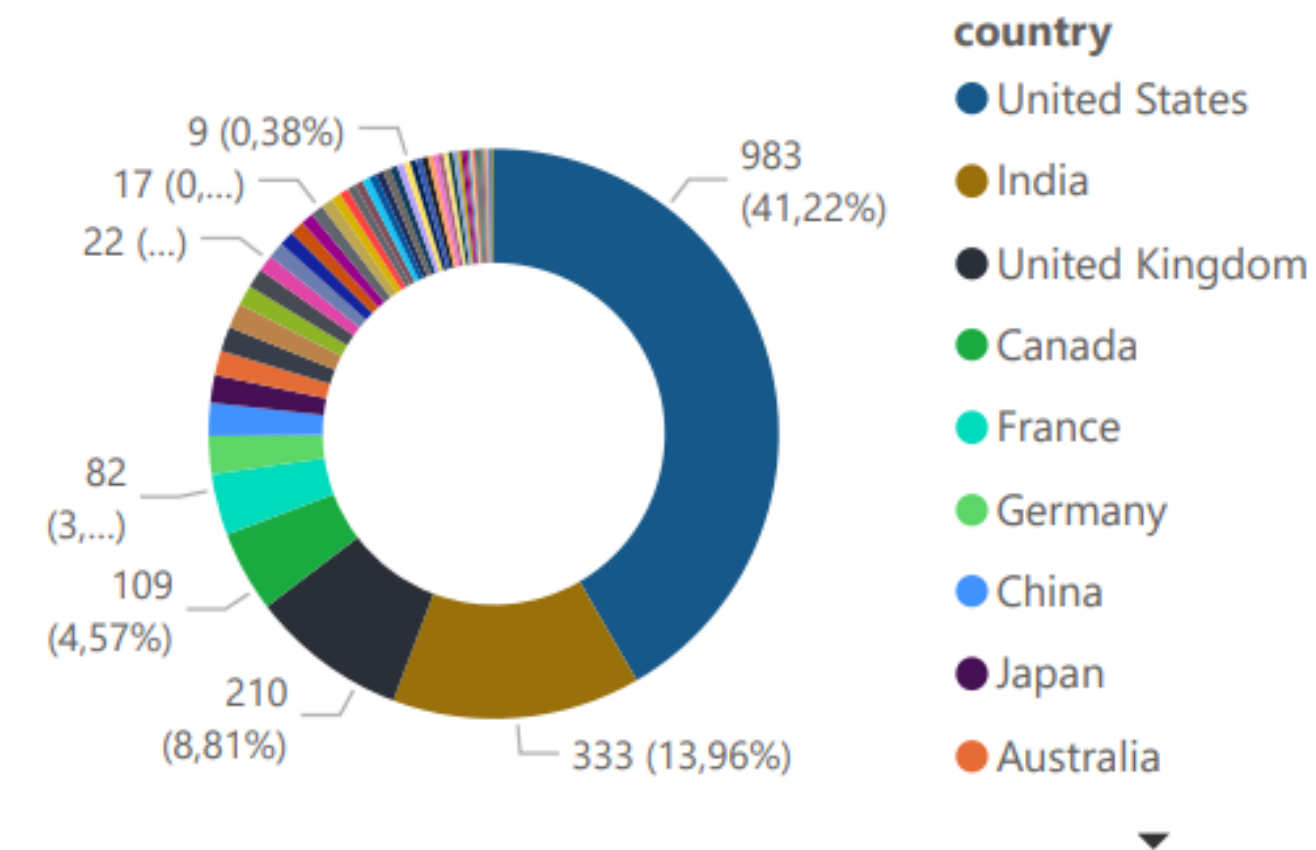
## Exemplo



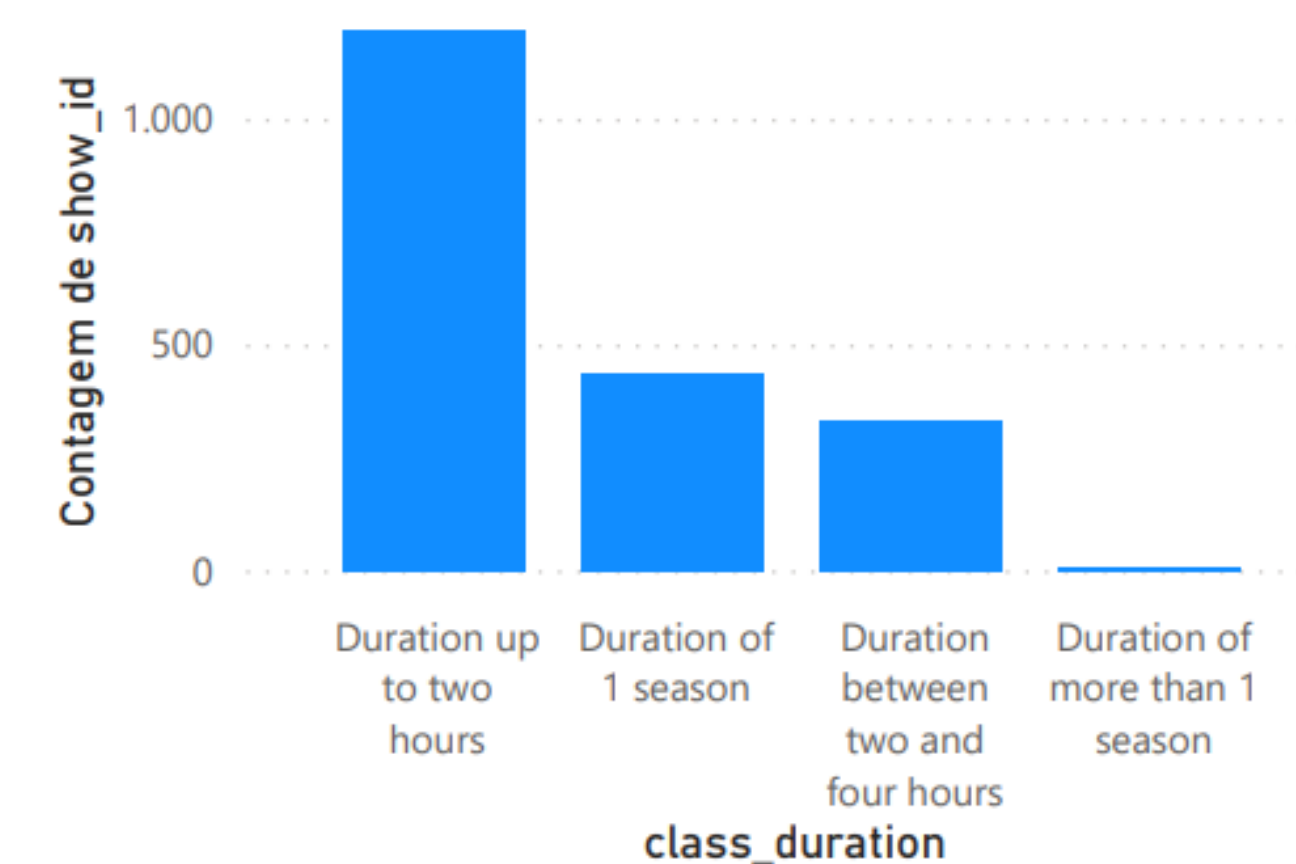
Contagem de show\_id por genre



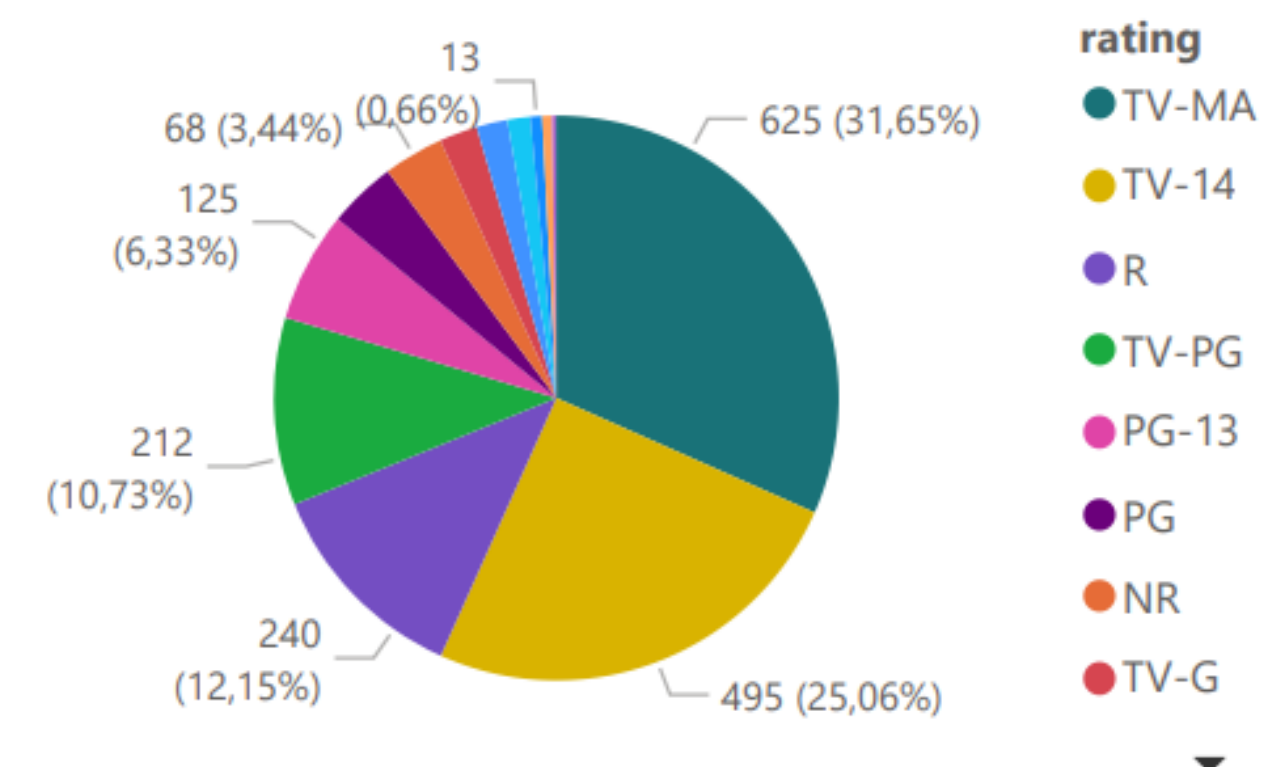
Contagem de show\_id por country



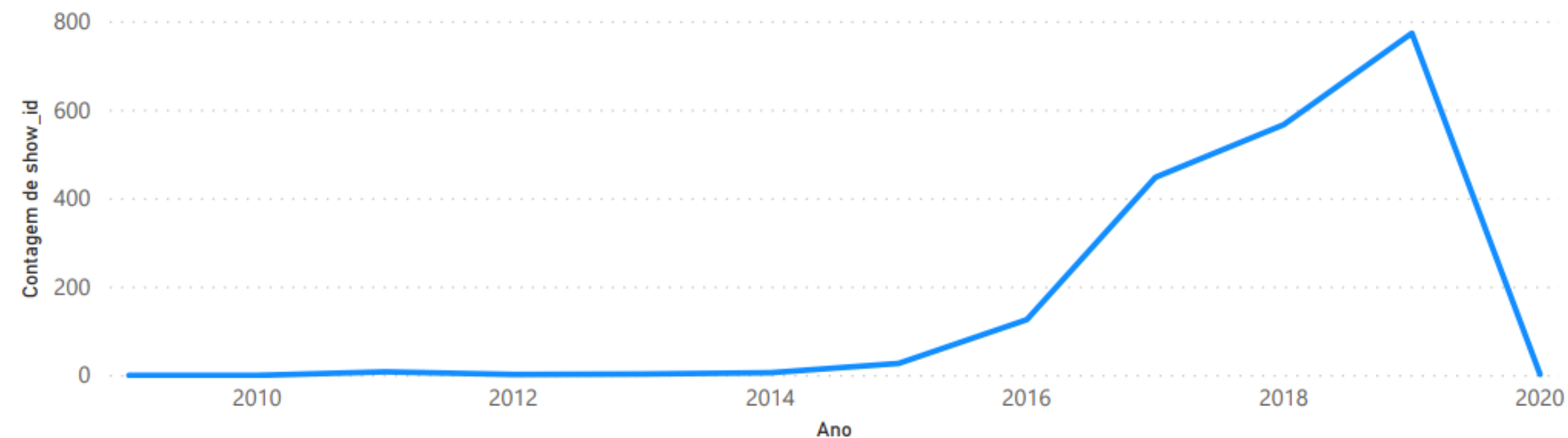
Contagem de show\_id por class\_duration



Contagem de show\_id por rating



Contagem de show\_id por Ano





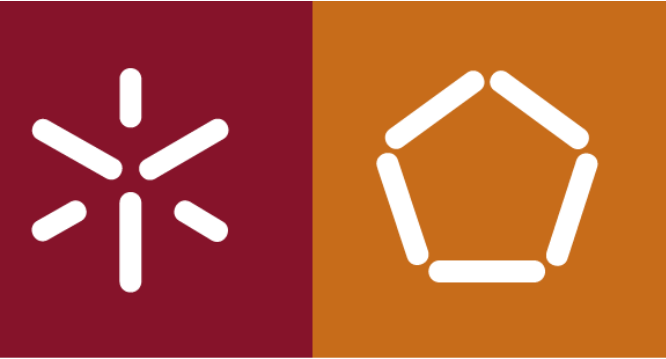
## Mestrado Integrado em Engenharia Informática, 4º ano

Send to: [daniela.oliveira@algoritmi.uminho.pt](mailto:daniela.oliveira@algoritmi.uminho.pt) with subject: “AD/TP5/pgXXXX”

### Ficha N.º 5

#### 1 *Business Intelligence* - Mental Health in IT Companies

O estudo apresentado trata-se de um questionário que avalia as atitudes dos trabalhadores em empresas de Tecnologias da Informação (ITs) relativamente à saúde mental, bem como a frequência de transtornos mentais em ambientes de trabalho. O dataset é constituído pelos seguintes atributos:



## 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://powerbi.microsoft.com/en-us/get-started/>