



Análise de Dados

TP3

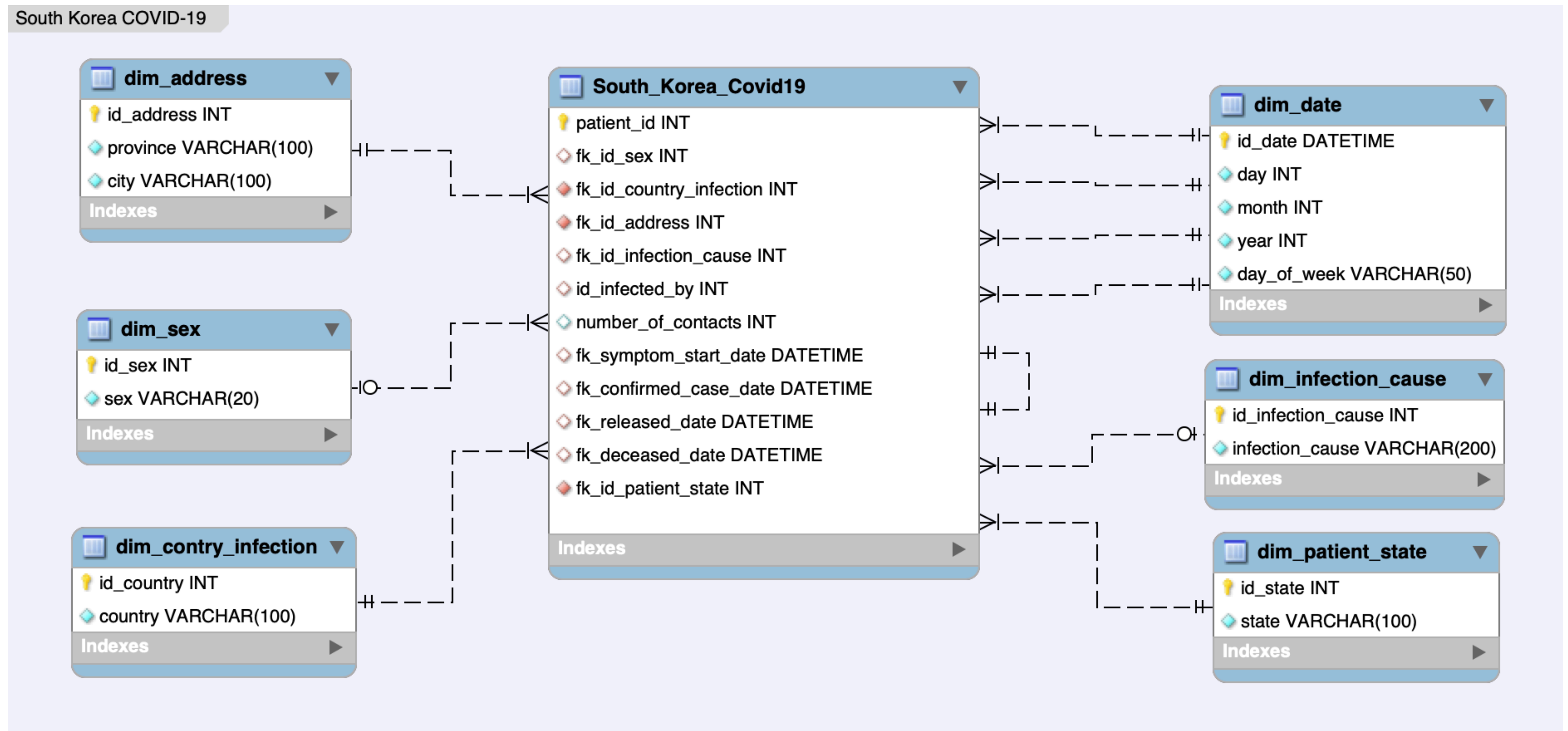
Daniela Oliveira



Mestrado Integrado em Engenharia Informática, 4º ano

TP3

Proposta Star Schema



TP2

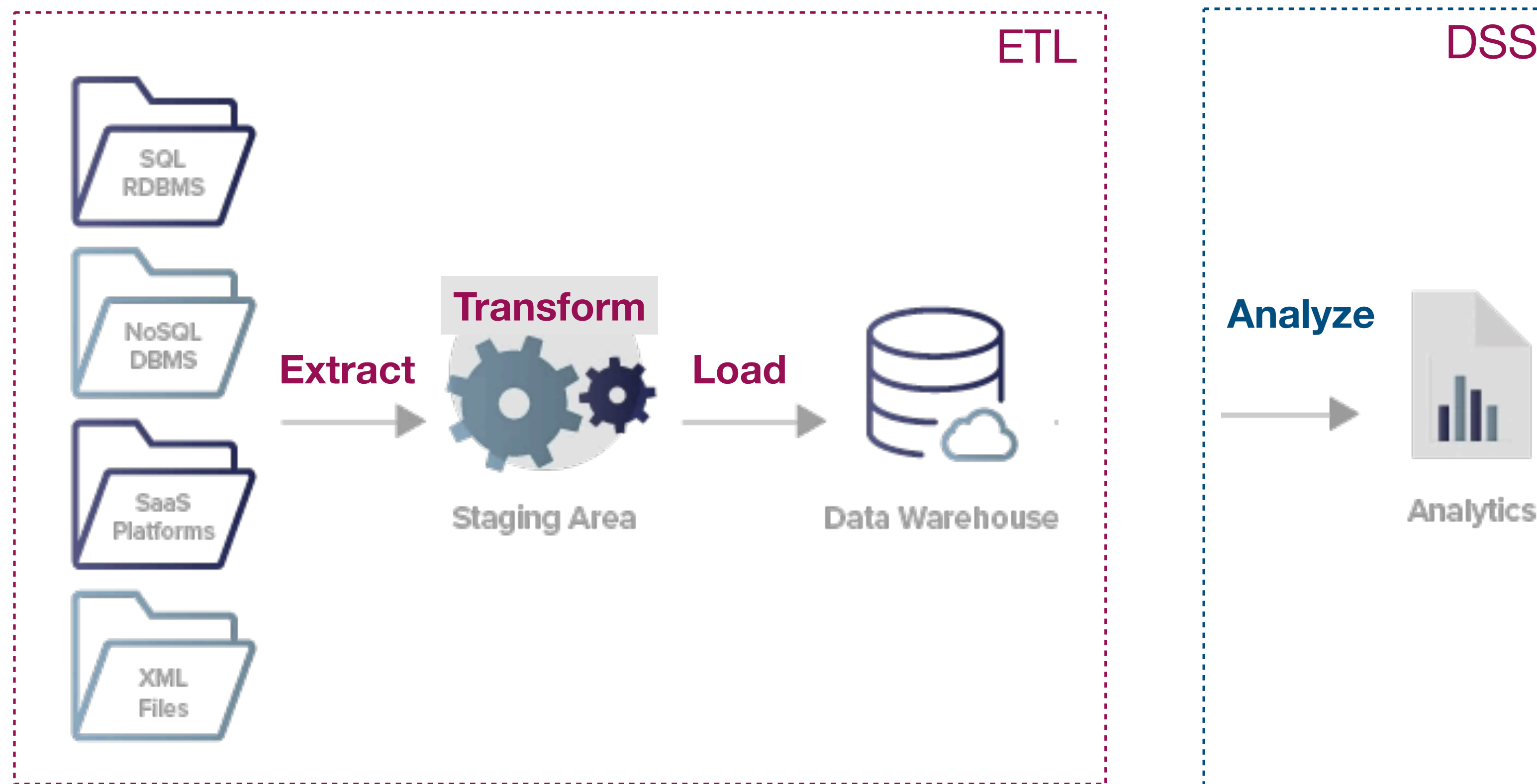
- Correção da 2ª ficha TP individual;
- Contextualização sobre os conceitos de ETL e *Data warehousing*;
- Resolução da 3ª ficha TP individual.



ETL

(Extract, Transform, Load)

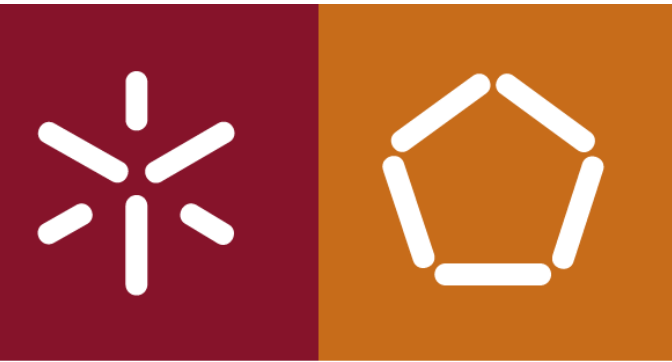
Definição



Diferentes fontes de informação

Constante aumento do volume de dados

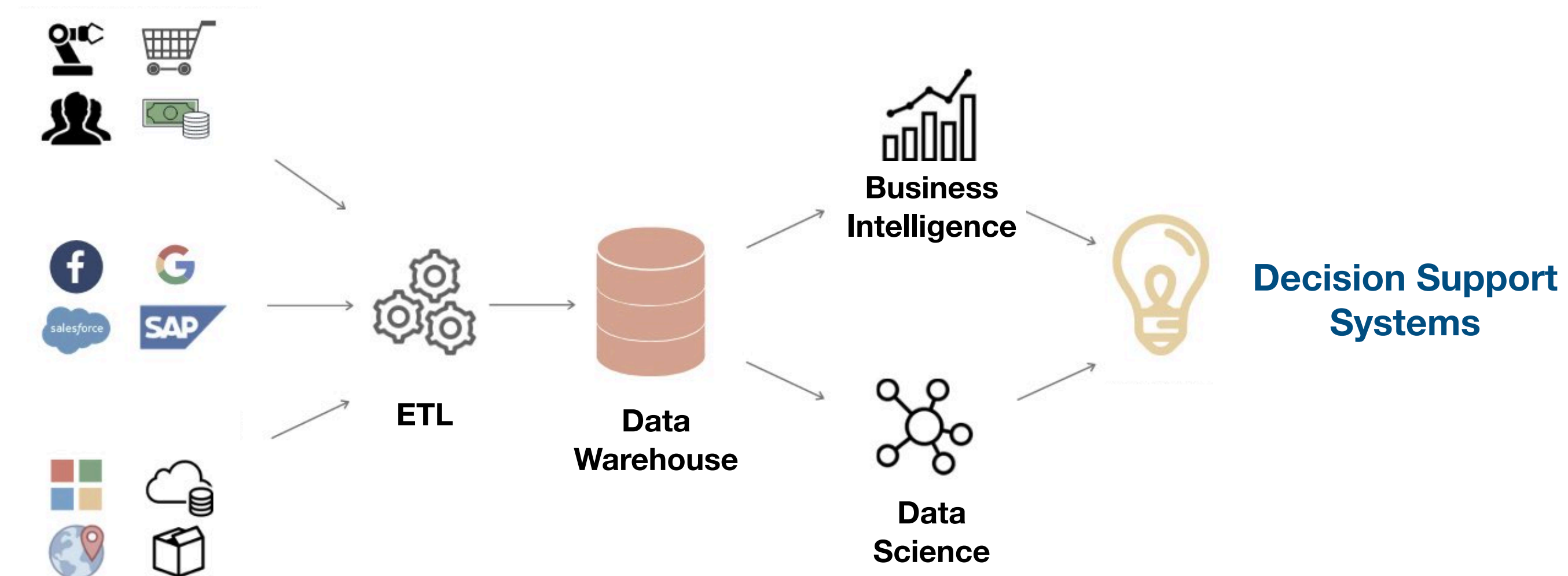
Diferentes níveis de estruturação de dados

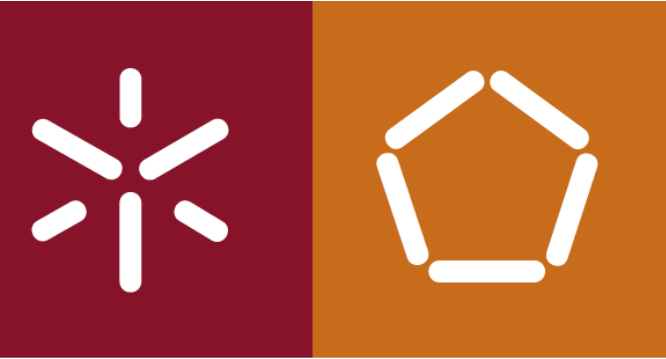


Data Warehouse

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.

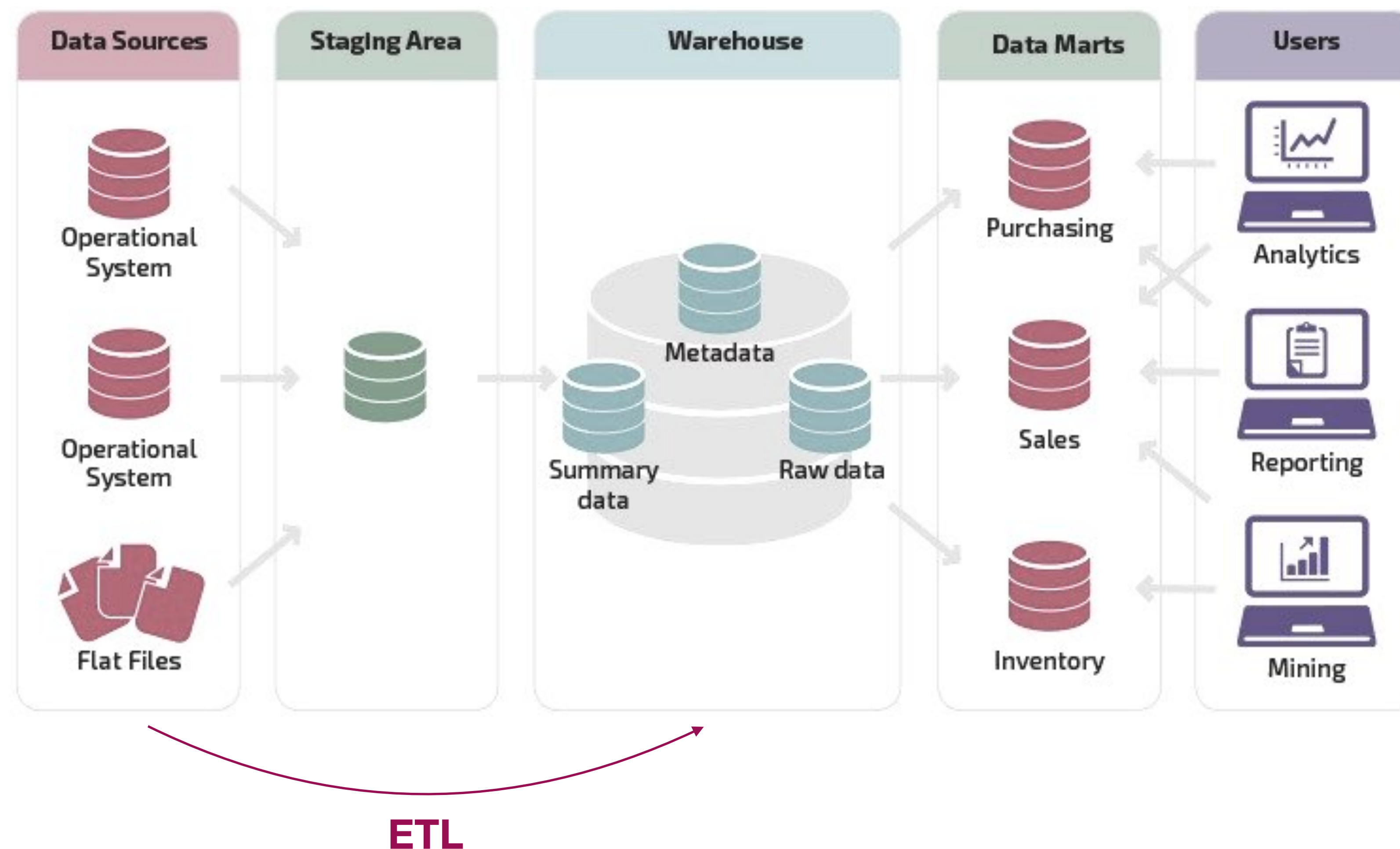


Mestrado Integrado em Engenharia Informática, 4º ano

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

Data Warehouse

Arquitetura





Data Warehouse

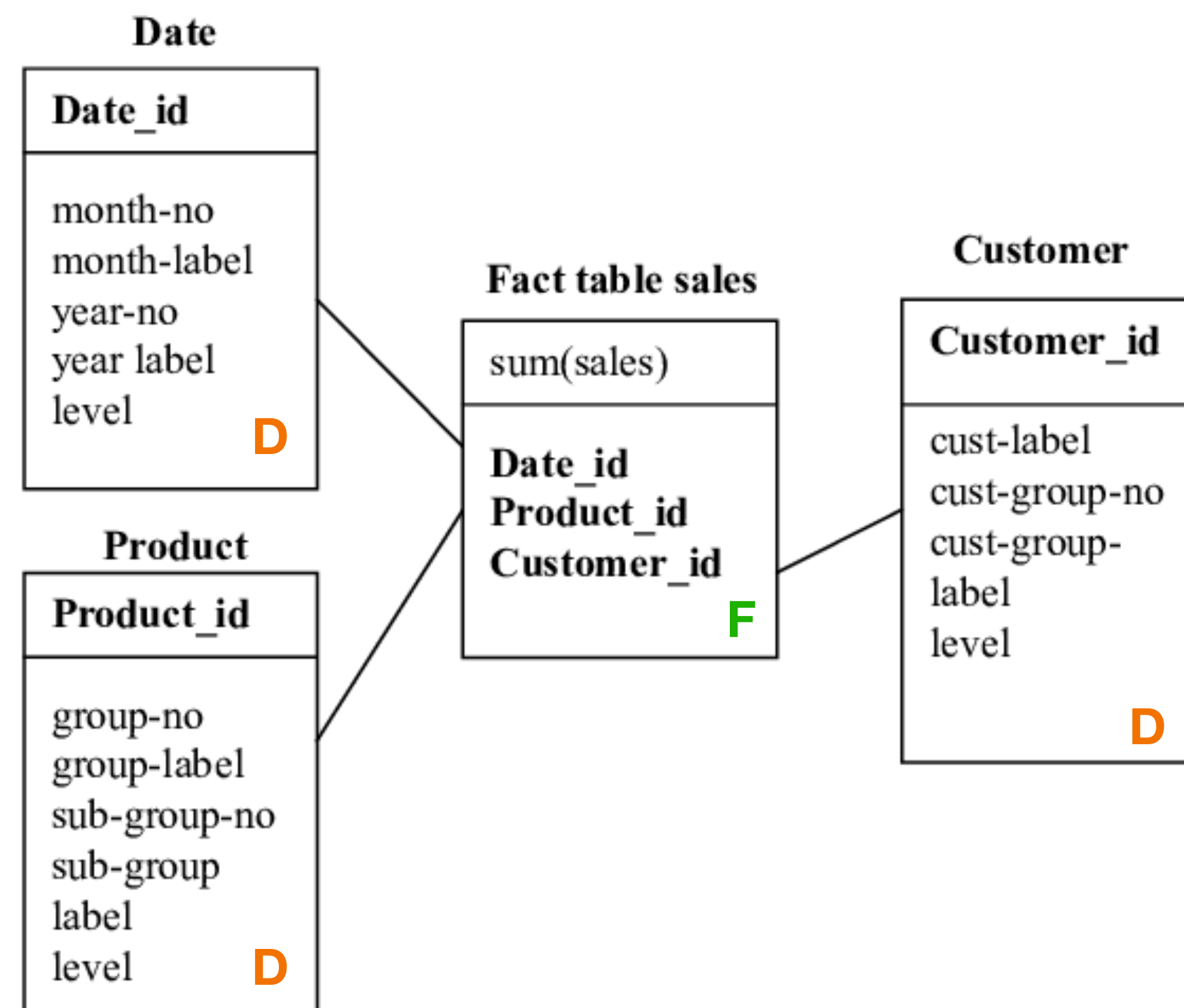
Modelação Dimensional

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.

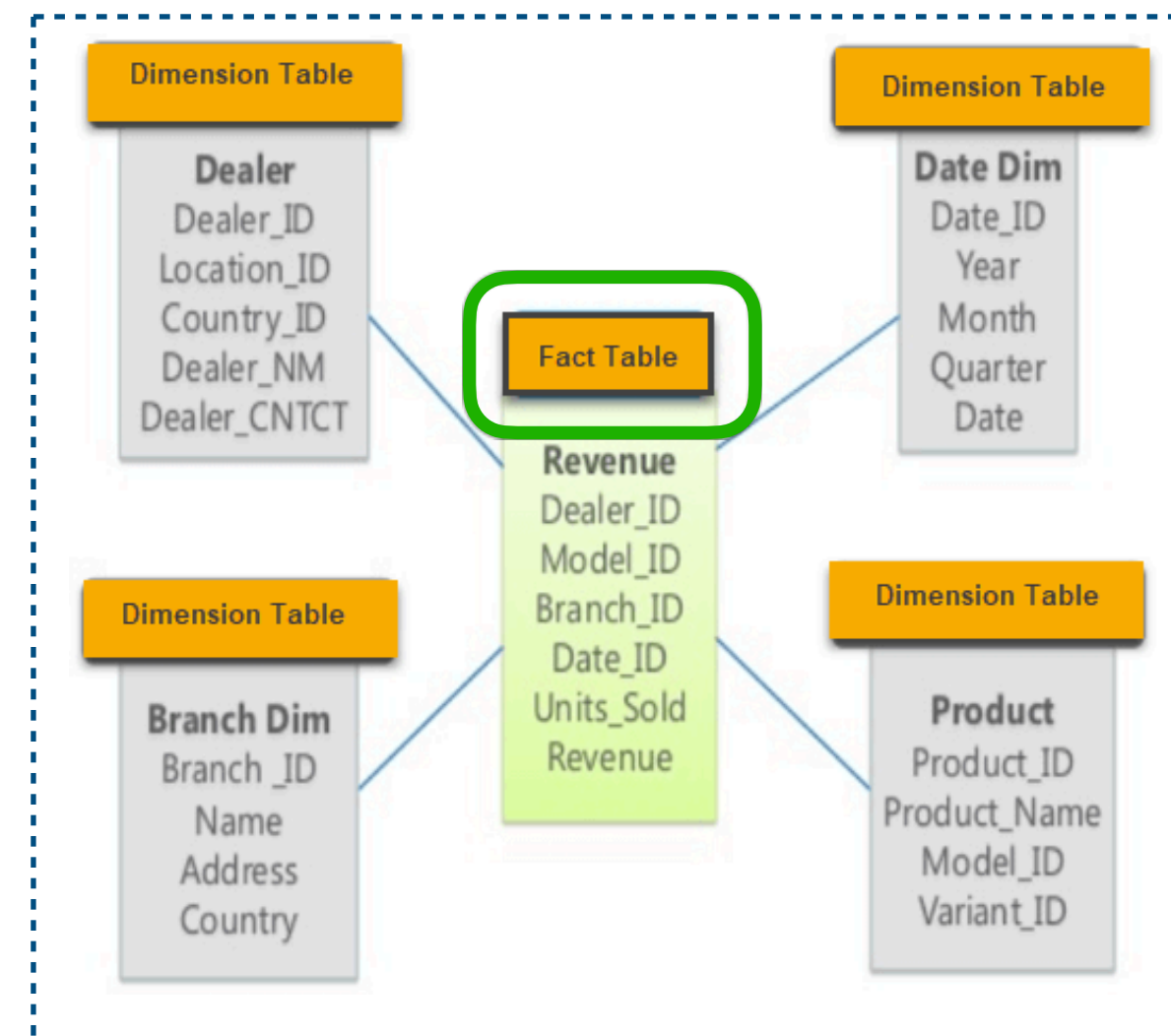




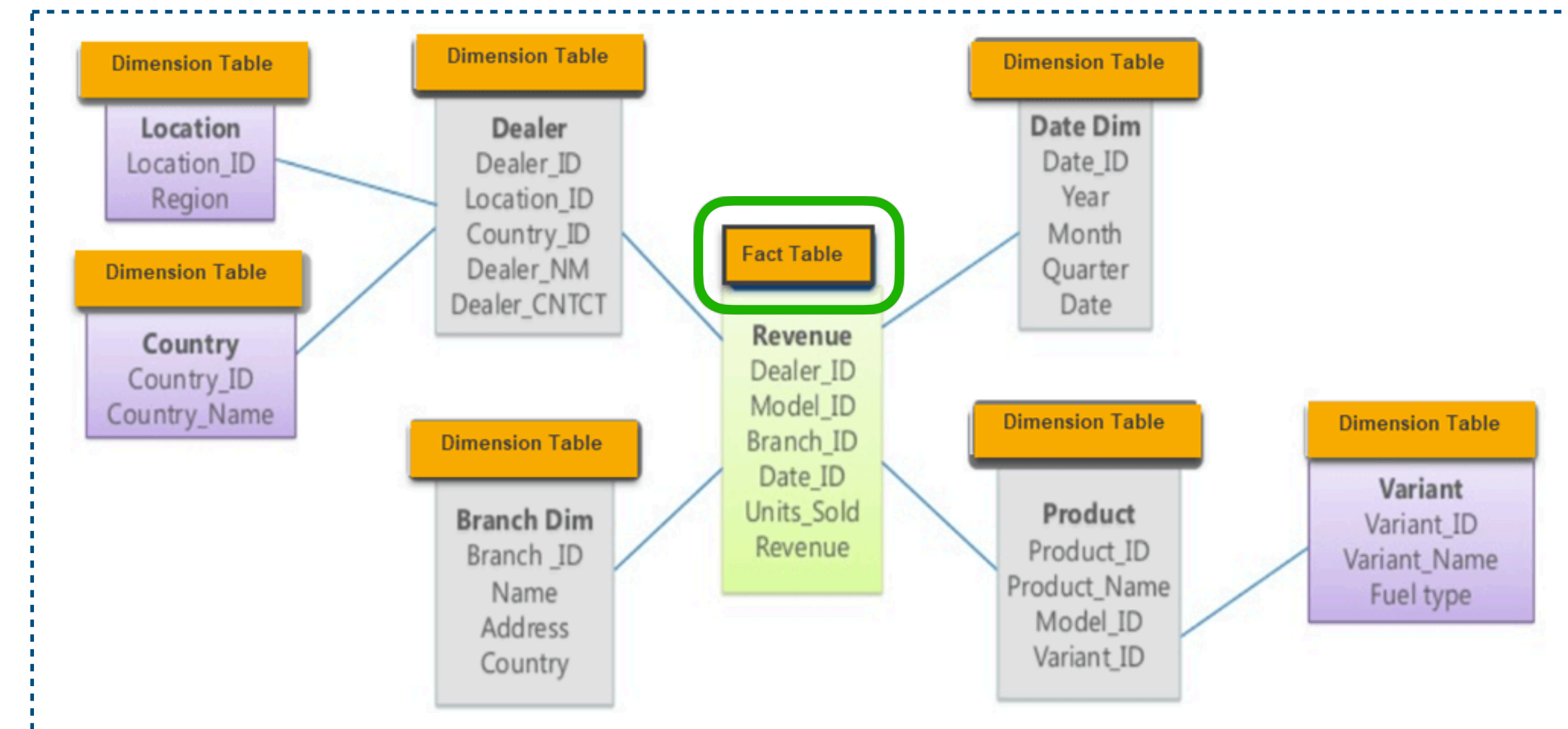
Data Warehouse

Modelação Dimensional

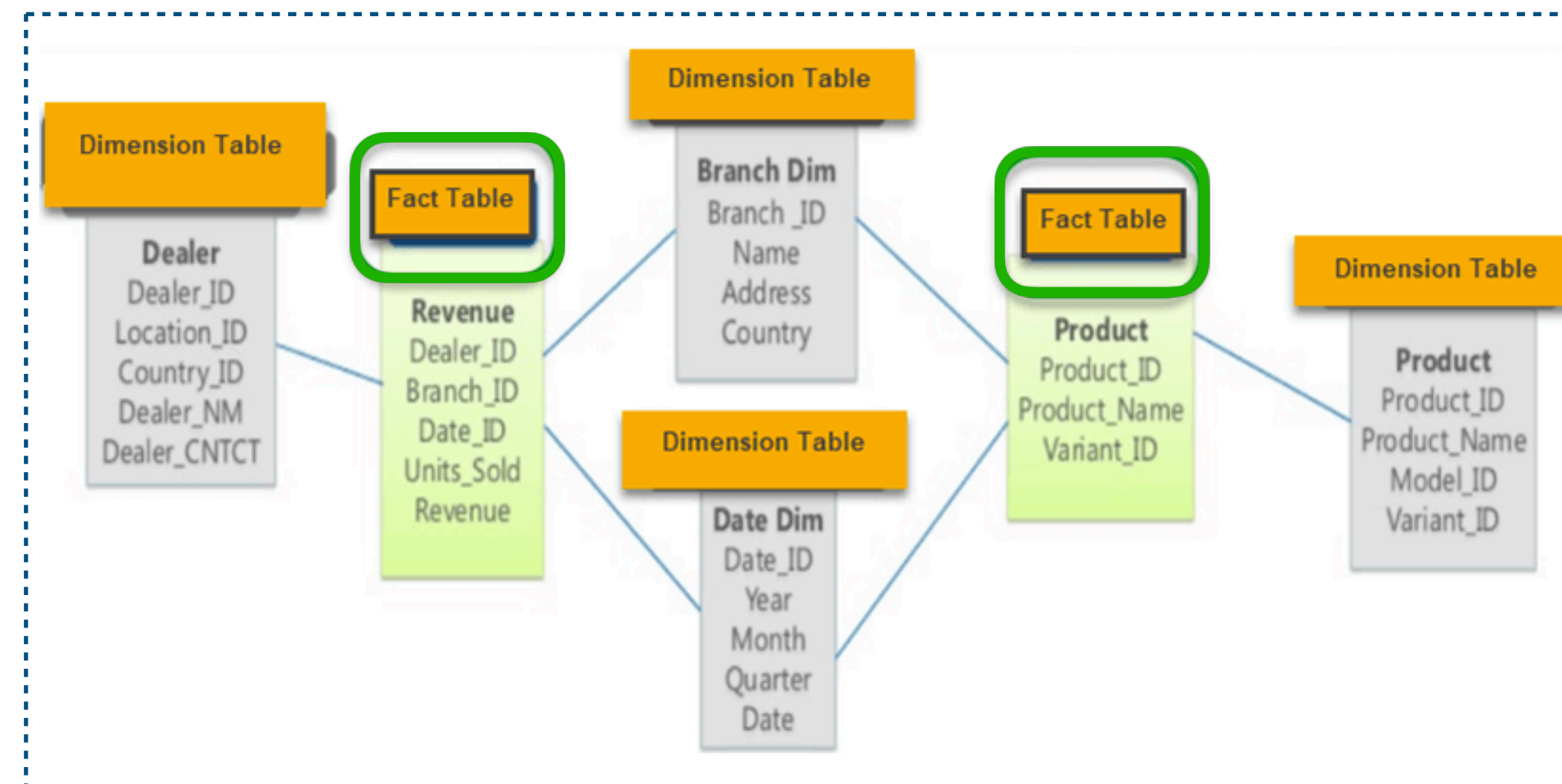
Star Schema



Snowflake Schema



Fact Constellation Schema





Mestrado Integrado em Engenharia Informática, 4º ano

Send to: daniela.oliveira@algoritmi.uminho.pt with subject: “**AD/TP3/pgXXXX**”

Ficha N.º 3

1 Modelação Dimensional - Disparos fatais policiais nos EUA

O presente exercício tem como objetivo analisar dados referentes às mortes nos EUA por disparos fatais de polícias, em cumprimento legal.

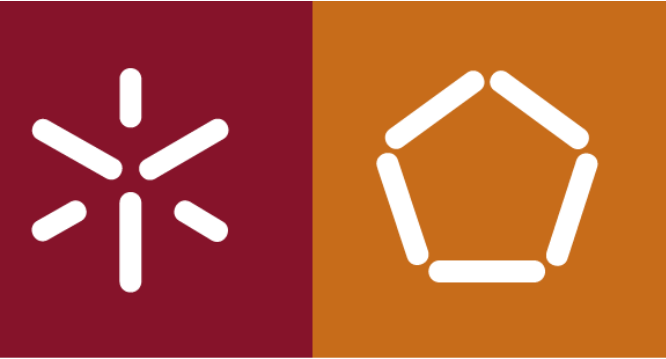
No *dataset* em estudo, apenas existem dados desde 2015, um ano depois do assassinato do Michael Brown, um homem negro desarmado, pela mão de um polícia em Ferguson, Missouri, EUA. Desde esse ano, o *The Washington Post* mantém a base de dados constantemente atualizada e disponível, contra qualquer tentativa de ocultação de informação.

O *fatal_police_shootings_data.csv* é constituído por 14 atributos descritivos de cada assassinato, nomeadamente o estado mental da vítima, a sua raça, a sua idade, o seu género, entre outros atributos relevantes para o estudo.

Ferramentas: MySQL e MySQL Workbench

Requisitos a desenvolver

1. Analisar o *dataset* fornecido e definir um modelo dimensional com esquema em estrela;
2. Analisar o *dataset* fornecido e definir um modelo dimensional com esquema em floco de neve;
3. Criar cada *EER Diagram* dos respetivos modelos dimensionais anteriormente definidos;
4. Povoar as tabelas com os dados presentes no ficheiro *fatal_police_shootings_data.csv*;
5. Desenvolver 5 *queries* cujo o seu significado represente um determinado indicador sobre o tema em estudo.



Bibliografia

- <https://www.washingtonpost.com/graphics/investigations/police-shootings-database/>
- Kimball, Ralph, and Margy Ross. ***The data warehouse toolkit: The definitive guide to dimensional modeling***. John Wiley & Sons, 2013.
- <https://www.mysqltutorial.org/>