



# PL04 – Explorar o Weka

AEC - Mestrado em Engenharia Biomédica

<https://hpeixoto.me/class/aec>

# Plano de Aula - PL03

---



Base de Dados, Data Warehouse, Data Mart, Dataset



Algoritmo ZeroR - Base Line



Ficha Exercícios (fe04)

# Base de Dados, Data Warehouse, Data Mart, Dataset

# Base de Dados

	A	B	C	D
1	3989.408	3989.408	140.4029	2654.278
2	140.4029	4125.044	4125.044	1335.467
3	2654.278	1335.467	2789.76	2789.76
4	5777.168	1788.068	5912.553	3123.153
5	2050.529	6039.689	1915.155	4704.363
6	1435.265	2554.287	1571.295	1219.56
7	4006.104	7994.156	3872.258	6659.535
8	671.2763	3318.277	807.9208	1983.314
9	2622.699	1367.091	2758.56	43.64889
10	8364.031	12353.06	8229.223	11018.06

## ROWS

DATA BASES: tuples or records

DATA WAREHOUSES: observations, examples or cases

## COLUMNS

DATA BASES: fields

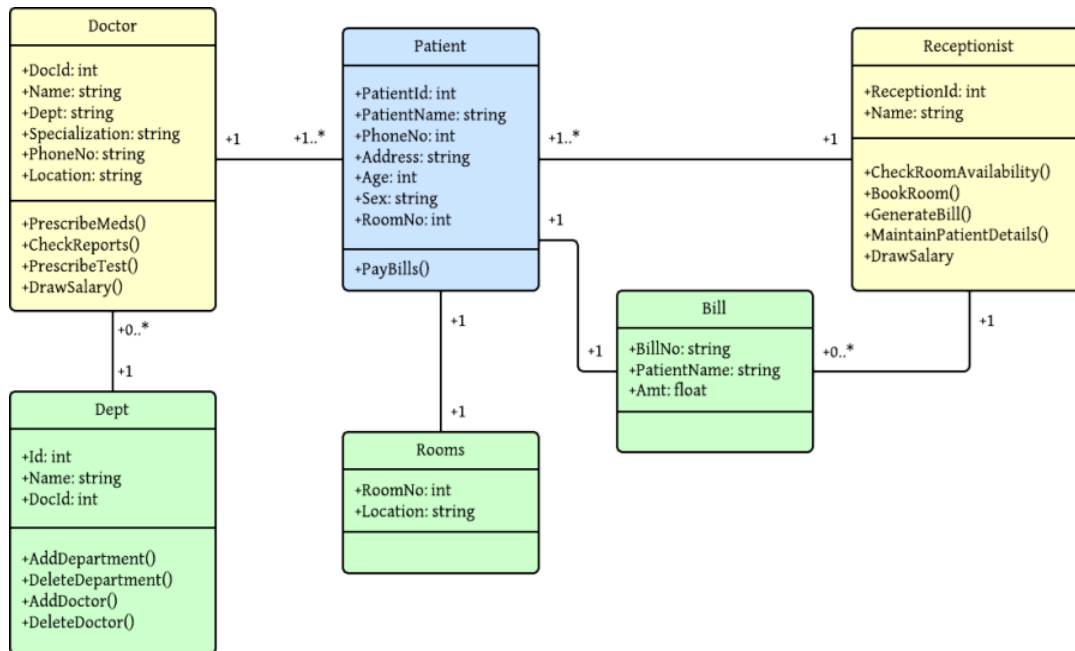
DATA WAREHOUSES: variables or attributes

# Base de Dados

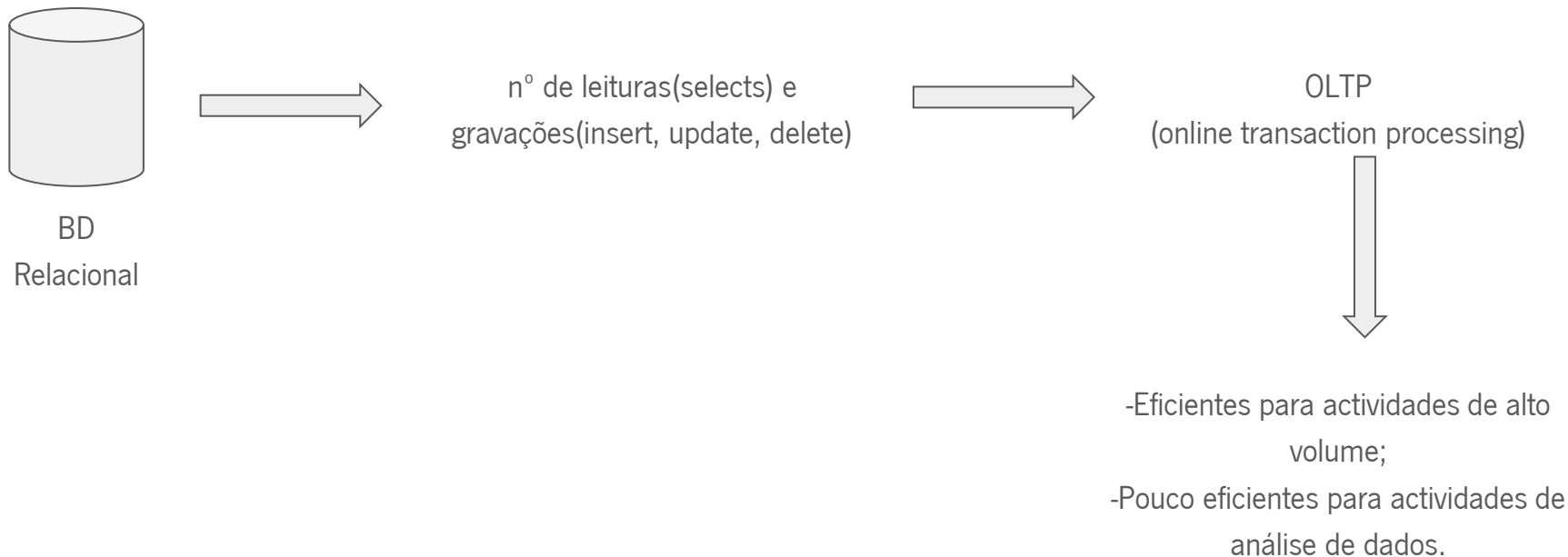
Owner_ID	Owner Name
1	Jim
2	Joan

Pet_ID	Pet Name	Owner_ID
1	Fifi	2
2	Butch	1
3	Clover	2
4	Animal	1
5	Tank	1



# Base de Dados





# Base de Dados vs Data Wharehouse

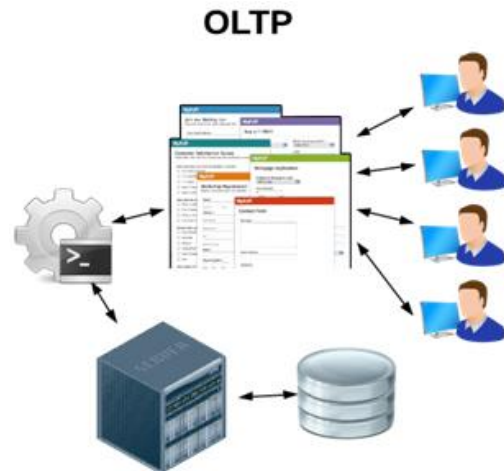
Owner_ID	Owner Name
1	Jim
2	Joan

Pet ID	Pet Name	Owner_ID
1	Fifi	2
2	Butch	1
3	Clover	2
4	Animal	1
5	Tank	1

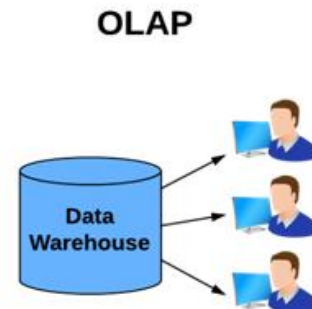
Pet_ID	Pet_Name	Owner_Name
1	Fifi	Joan
2	Butch	Jim
3	Clover	Joan
4	Animal	Jim
5	Tank	Jim

# OLTP vs OLAP

**On-line Transaction Processing (OLTP):** normalmente usado para caracterizar os sistemas transacionais, ou seja, os sistemas operacionais das organizações. São utilizados no processamento dos dados de rotina que são gerados diariamente através dos sistemas de informação da empresa e dão suporte às funções de negócio organizacional. **Dados Operacionais.**



**On-line Analytical Processing (OLAP):** trata da capacidade de analisar grandes volumes de informações nas mais diversas perspectivas dentro de um Data Warehouse (DW). O OLAP também faz referência às ferramentas analíticas utilizadas no BI para a visualização das informações gerenciais e dá suporte para as funções de análises do negócio organizacional. **Dados Organizacionais.**







# Base de Dados vs Data Warehouse

**DATA WAREHOUSE** – dados arquivados – copiados de uma base de dados transacional



**DESNORMALIZAÇÃO** ocorre no momento em que os dados são copiados para fora do sistema transacional



Processo de construção de Datawarehouse reflete normalmente um processo de ETL (extraction transformation and loading)

# Data set e Data mart

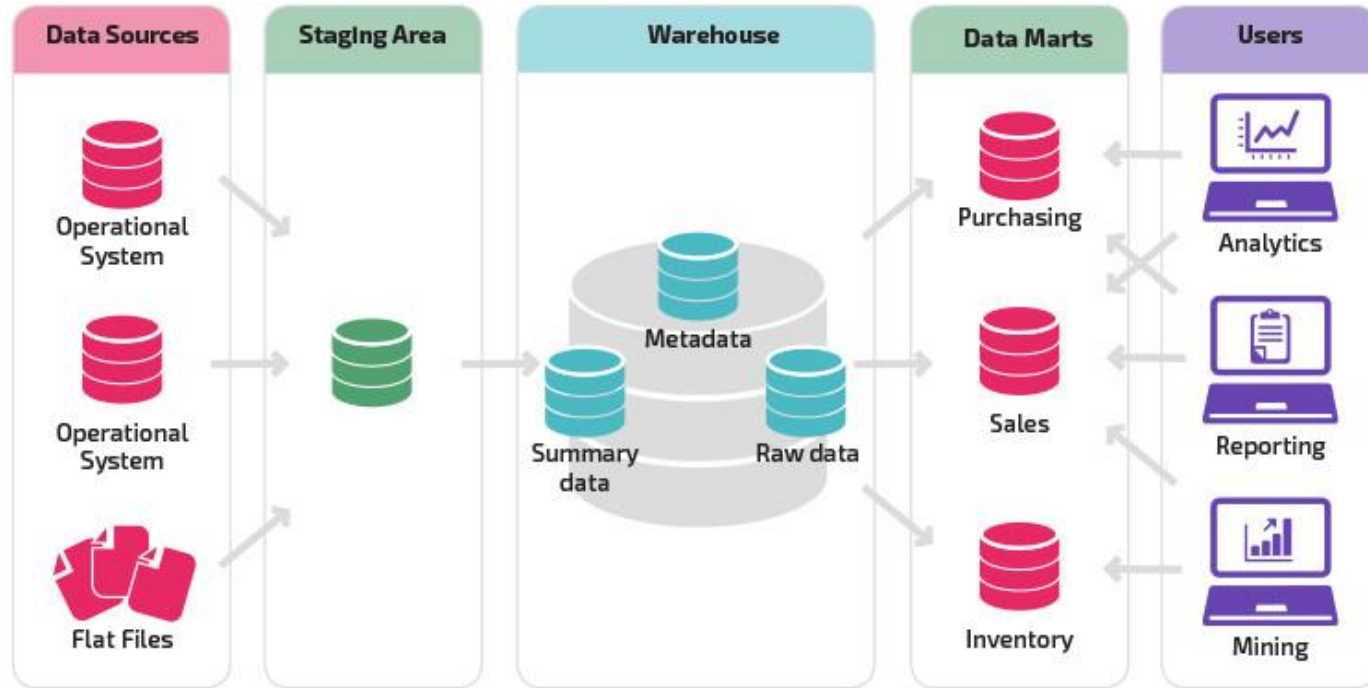
## **Data mart:**

- Armazenamento de dados organizacionais, semelhante a um datawarehouse, mas geralmente criado em conjunto com as necessidades das unidades de negócios, como o Marketing ou Atendimento ao Cliente, para fins de realização de relatórios de gestão;
- Ajudam a melhorar a resposta aos utilizadores devido à redução do volume de dados;
- Melhoram o acesso a dados frequentemente acedidos;
- Os dados são armazenados permitindo um controlo mais granular de privilégios de acesso.

## **Data set:**

- Subset de uma base de dados, data warehouse ou datamart.
- Criar um data set inclui: Anexar, combinar e simplificar algumas expressões de dados

# Data set e Data mart





# Algoritmo ZeroR - Base Line

# Algoritmo ZeroR - Base Line

## Exemplo:

Escolhe a classe mais comum e classifica todas as instâncias com esse valor.

No.	1: outlook	2: temperature	3: humidity	4: windy	5: play
	Nominal	Nominal	Nominal	Nominal	Nominal
1	sunny	hot	high	FALSE	no
2	sunny	hot	high	TRUE	no
3	overcast	hot	high	FALSE	yes
4	rainy	mild	high	FALSE	yes
5	rainy	cool	normal	FALSE	yes
6	rainy	cool	normal	TRUE	no
7	overcast	cool	normal	TRUE	yes
8	sunny	mild	high	FALSE	no
9	sunny	cool	normal	FALSE	yes
10	rainy	mild	normal	FALSE	yes
11	sunny	mild	normal	TRUE	yes
12	overcast	mild	high	TRUE	yes
13	overcast	hot	normal	FALSE	yes
14	rainy	mild	high	TRUE	no





# Ficha de Exercícios 04



# PL04 – Explorar o Weka

AEC - Mestrado em Engenharia Biomédica

<https://hpeixoto.me/class/aec>