



PL03 - Explorar o WEKA

SAEC - Mestrado Integrado em Engenharia Biomédica

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

Hugo Peixoto

hpeixoto@di.uminho.pt

2020/2021

Plano de Aula - PL02

 Base de Dados, Data Warehouse, Data Mart, Dataset

 Algoritmo ZeroR - Base Line

 Ficha Exercícios (fe03)

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

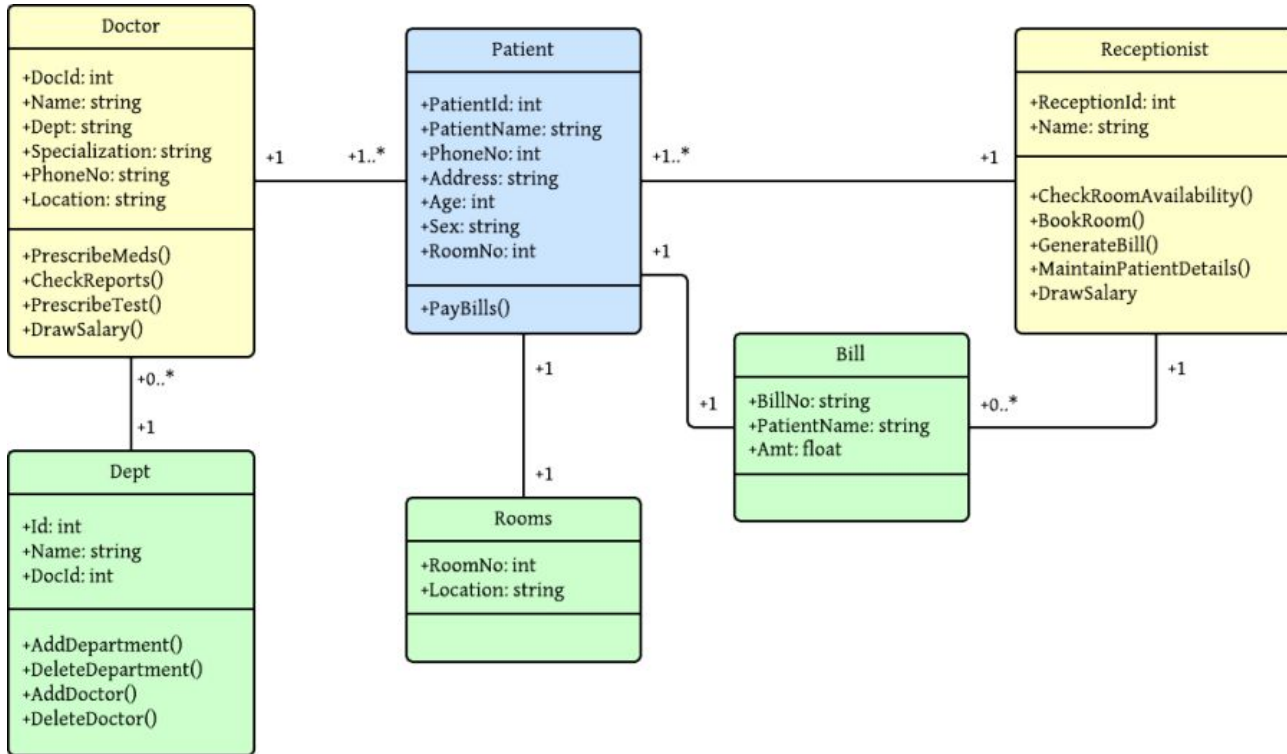
DATA WAREHOUSES: variables or attributes

Base de Dados

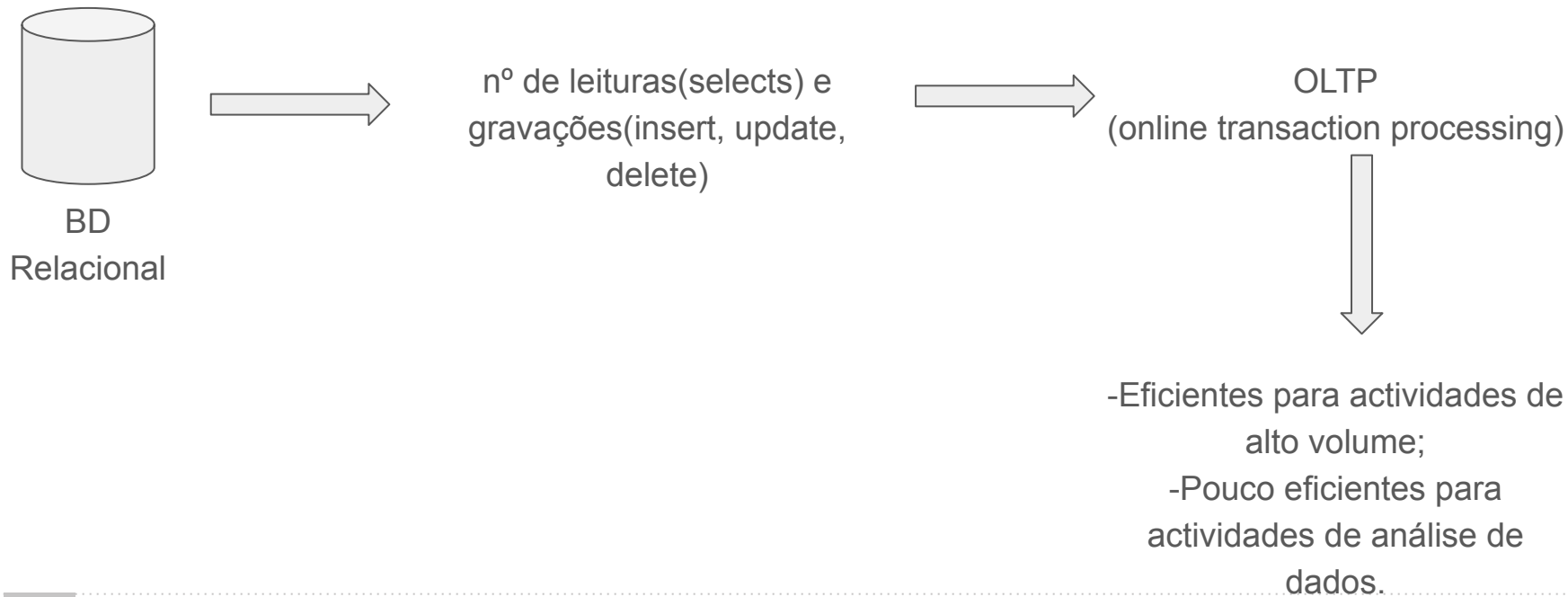




Base de Dados



Base de Dados





Base de Dados vs Data Warehouse

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

Base de Dados

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



As técnicas de Data Mining são geralmente mais eficazes em data sets, extraídos e OLAP, em vez de sistema sOLTP

Dataset

Subset de uma base de dados ou data warehouse

Criar um data set inclui:

Anexar, combinar e simplificar algumas expressões de dados

✓ Por exemplo mudar o formato de datas

Tipos de Dados

Dados Operacionais: tipo mais elementar de dados, vêm de sistemas transacionais que registam as atividades quotidianas.

Dados Organizacionais: ajudam a proteger a privacidade das pessoas, ao mesmo tempo em que são úteis para os *dataminers* que procurem tendências numa determinada população.

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

Segurança e Privacidade

Devemos estar cientes de que, quando reunimos, organizamos e analisamos dados, esses dados pertencem normalmente a utilizadores reais.

As pessoas têm direitos relativamente à privacidade e proteção contra crimes como roubo de identidade

Implementação do RGPD

Algoritmo ZeroR - Base Line

Algoritmo ZeroR - Base Line

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 03



PL03 - Explorar o WEKA

SAEC - Mestrado Integrado em Engenharia Biomédica

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

Hugo Peixoto

hpeixoto@di.uminho.pt

2020/2021