

Universidade do Minho
Escola de Engenharia

D escoberta do C onhecimento

José Machado
Cristiana Neto

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

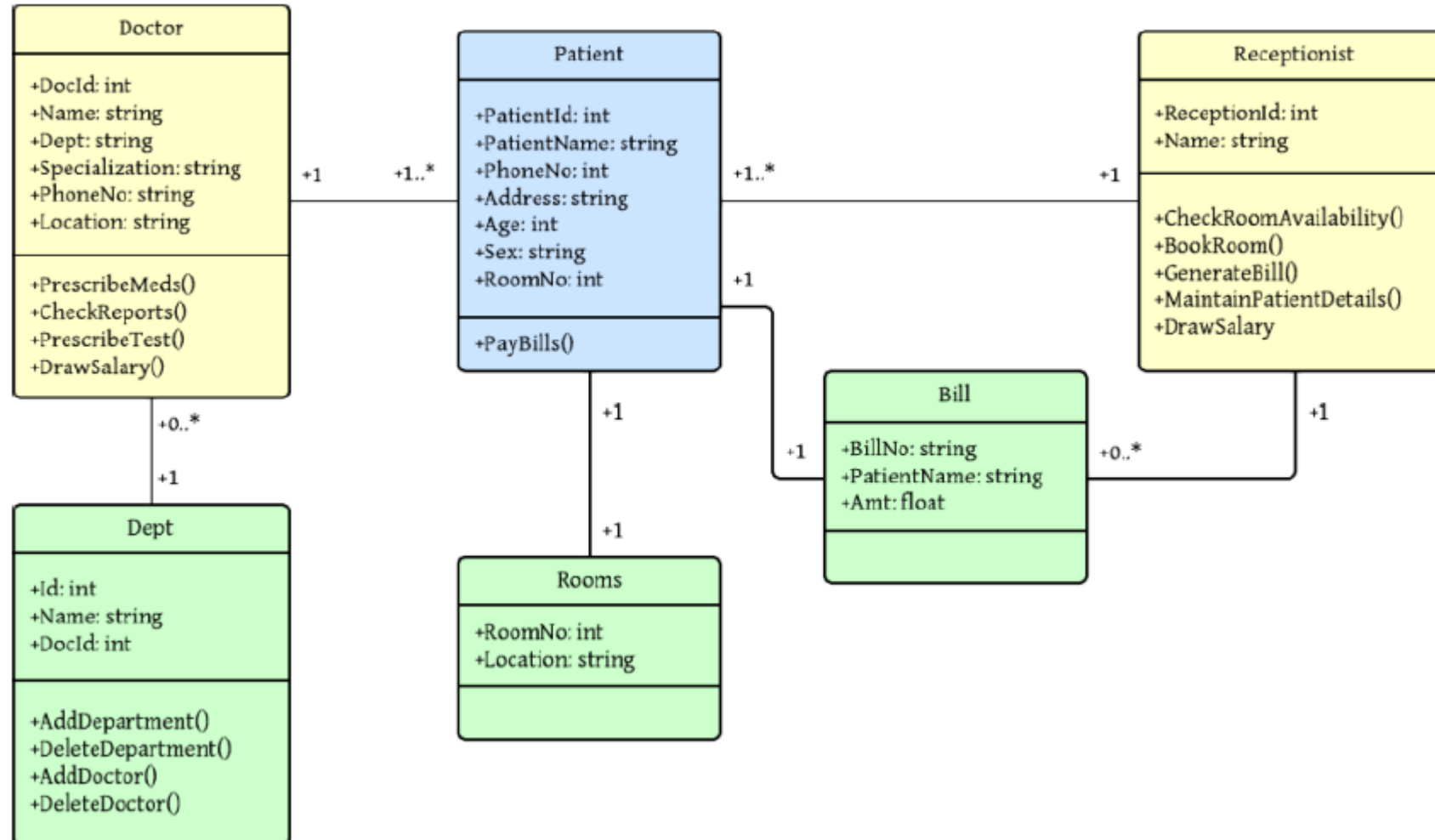
BASE DE DADOS



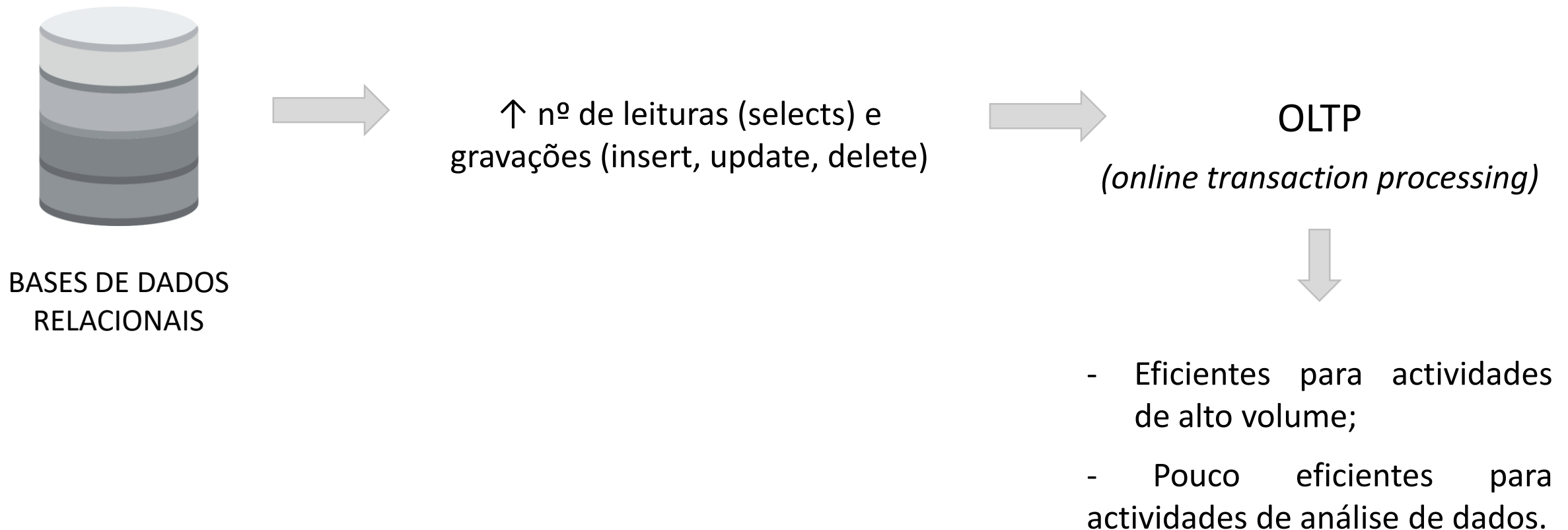
Owner_ID	Owner Name
1	Jim
2	Joan

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



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

OLTP

(Online Transaction Processing)



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

OLAP

(Online Analytical Processing)

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



As técnicas de Data Mining são geralmente mais eficaz em data sets, extraídos de OLAP,
em vez de sistemas OLTP

DATA SET



- Subset de uma database 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 data miners que procurem tendências numa determinada população.



DATA MART



Armazenamento de dados organizacionais, semelhante a um data warehouse, mas geralmente criado em conjunto com as necessidades das unidades de negócios, como Marketing ou Atendimento ao Cliente, para fins de realização de relatórios e gestão.

SEGURANÇA E PRIVACIDADE



- Devemos estar cientes de que, quando reunimos, organizamos e analisamos dados, há pessoas reais por trás.
- As pessoas têm certos direitos relativamente à privacidade e proteção contra crimes como roubo de identidade.



“BASE LINE”

- ZeroR

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