Knowledge Discovery

José Machado

Cristiana Neto

Universidade do Minho
Escola de Engenharia

DATA BASE

	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

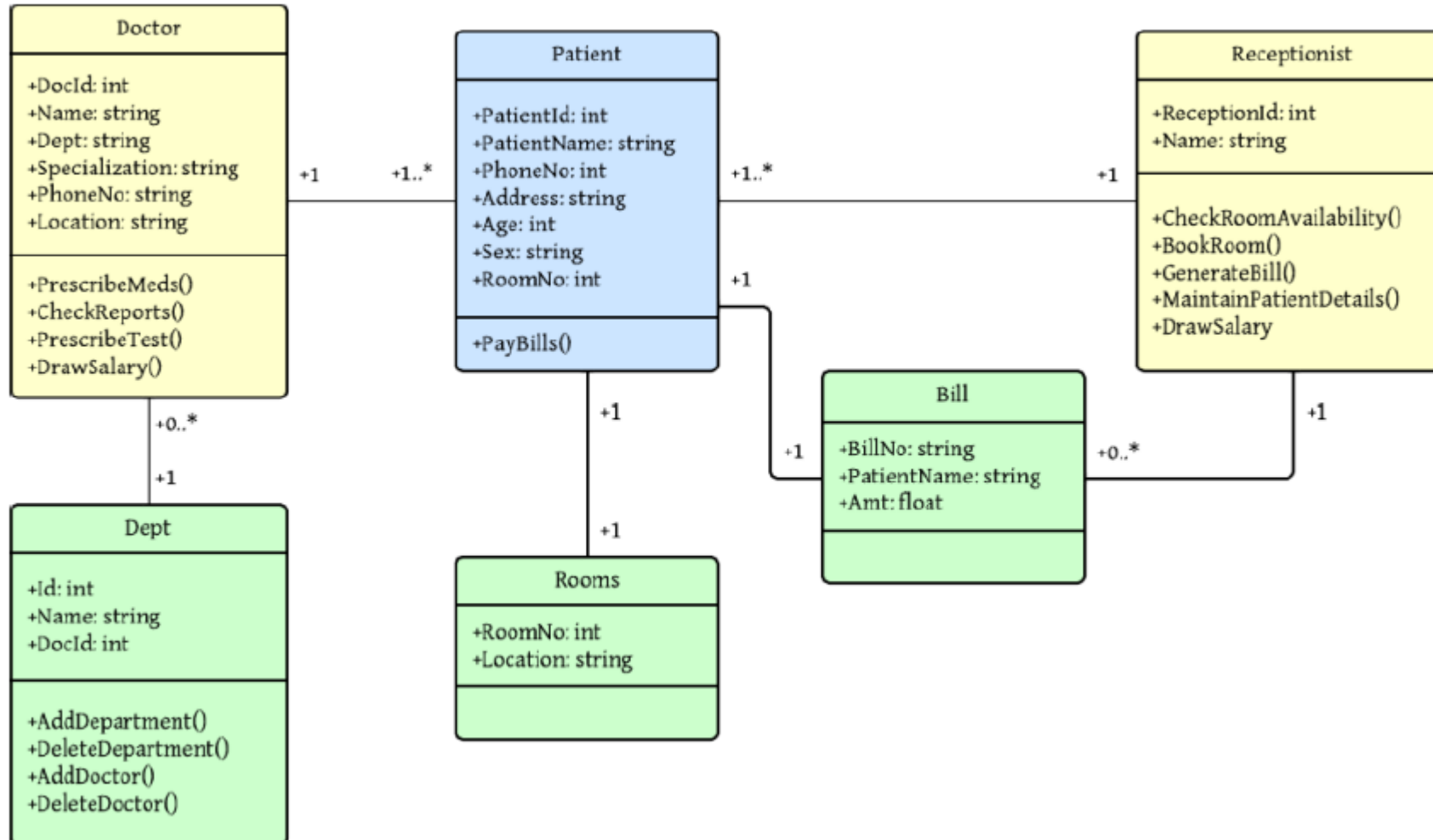DATA BASES: fields
DATA WAREHOUSES: variables or attributes

DATA BASE

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

DATA BASE

DATA BASE

RELATIONAL
DATABASES

↑ number of reads (selects) and writes (insert, update, delete)

OLTP

(online transaction processing)

- Efficient for high volume activities;

- Inefficient for data analysis activities.

DATA BASES 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

OLTP

(Online Transaction Processing)

OLAP

(Online Analytical Processing)

DATA WAREHOUSE

DATA WAREHOUSE – archived data **–** copied from a transactional database

⬇

DENORMALIZATION
occurs at the time the data is copied out of the transactional system

⬇

Data Mining techniques are generally more effective on data sets, extracted from OLAP,
rather than OLTP systems

DATA SET

- Subset of a data base or data warehouse

- The creation of a data set may contain several steps including appending or combining tables from source database tables, or simplifying some data expressions.

TYPE OF DATA

- Operational Data: the most elemental type of data, operational data, comes from transactional systems which record everyday activities

- Organizational Data: help to protect peoples' privacy, while still proving useful to data miners watching for trends in a given population

DATA MART

Organizational data storage, similar to a data warehouse, but usually created in conjunction with the needs of business units, such as Marketing or Customer Service, for reporting and management purposes.

SECURITY & PRIVACY

- We should be aware that as we gather, organize and analyze data, there are real people behind the figures

- People have certain rights to privacy and protection against crimes such as identity theft

DATA MINING

"BASE LINE"

- ZeroR

Choose the most common class and classify all instances with that value.