# Introduction to Relational Databases BD

Lic. Informatique, Université Lille1, Premier cours.

C. Kuttler (groupe 1)

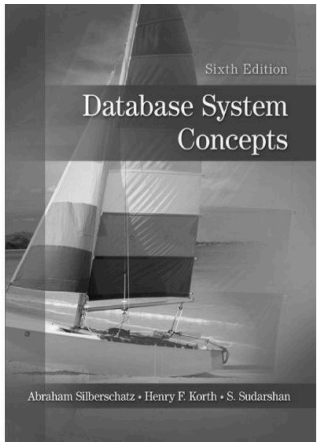
M.Bilasco (groupe 3)

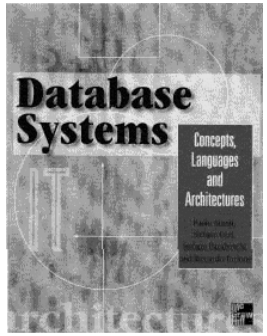
JC Marti (groupe 2)

A Bonifati (groupe 4)



# Books (suggestions...you can choose your own)



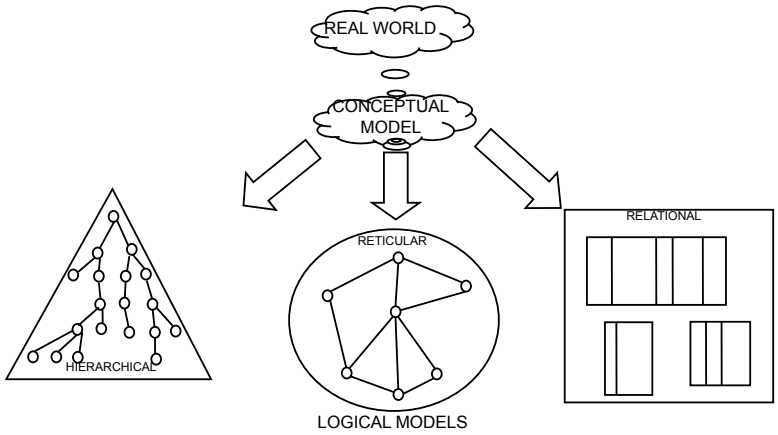






# 1st Class: Relational model

# DATA MODELS

DATA MODELS ENCODE A STRUCTURING OF THE REALITY THAT INCLUDES SPECIFIC FEATURES AND ALLOWS TO BETTER UNDERSTAND IT



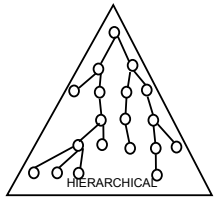
## Models for data representation

- Hierarchical model (1960)
- Reticular model (1970)
- Relational model (1980)
- O-O model (1990)
- XML model (2000)

# LOGICAL DATA MODELS

HIERARCHICAL

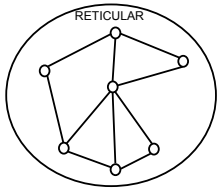
- DATA ARE ENCODED AS RECORDS
- THE LOGICAL ASSOCIATIONS BETWEEN DATA ARE REPRESENTED VIA POINTERS INSIDE A TREE STRUCTURE




# LOGICAL DATA MODELS

RETICULAR (CODASYL)

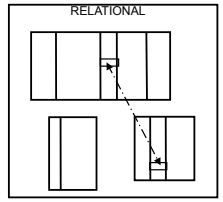
- DATA ARE ENCODED AS RECORDS
- THE LOGICAL ASSOCIATIONS BETWEEN DATA ARE REPRESENTED VIA POINTERS INSIDE A COMPLEX GRAPH STRUCTURE




# LOGICAL DATA MODELS

RELATIONAL

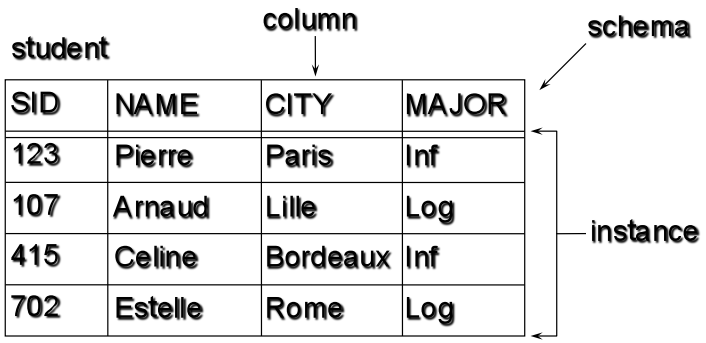
- DATA ARE ENCODED AS TABLES
- THE LOGICAL ASSOCIATIONS BETWEEN DATA ARE REPRESENTED VIA DIFFERENT VALUES OF ATTRIBUTES IN DIFFERENT TABLES



# History and Development of the Relational Model

- Invented by T. Codd, 1970
  (IBM Research in Santa Teresa, California)
- Related research projects:
  SYSTEM R (IBM), Ingres (Berkeley Un.)
- Commercial systems:
  beginning of the 80s (Oracle, IBM-SQL DS and DB2, Ingres, Informix, Sybase)
- Commercial success: from 1985 on.

# Informal definition of a Table




# Formal definition

- **Domain D:**
  **an arbitrary set of values**
- **Cartesian product on n domains**
  **$D_1$ x $D_2$ x … $D_n$ (not necessarily distinct): set of tuples $< d_1 , d_2 , \ldots d_n >$, with $d_i \in D_i$ , $1 \le i \le n$**
- **Relation R on $D_1$ x $D_2$ x... $D_n$ : an arbitrary subset of**
  **$D_1$ x $D_2$ x ... $D_n$ .**

# Example

- **$D_1$ = (a,b)**
- **$D_2$ = (1,2,3)**
- **$D_1$ x $D_2$ = ( <a,1>, <b,1>, <a,2>, <b,2>, <a,3>, <b,3> )**
- **R1 = ( <a,1>, <b,3> )**
- **R2 = ( <a,1>, <b,3>, <a,2> )**
- **R3 = ( )**
- **R4 = ( <a,1>, <b,1>, <a,2>, <b,2>, <a,3>, <b,3> )**

# Properties

- Arity of the relation:
  number of domains (n)
- Cardinality of the relation:
  number of tuples
- Attribute:
  name given to a domain in a relation
  [Names of attributes in a relation have to be all distinct one from another]

# Properties

Schema (of a relation):

table (attribute1,… attributeN)

[Name of relations in a schema have to be all distinct one from another]

R1(A,B) R2(C,D)

| A | B |
|---|---|
| a | 1 |
| b | 3 |

| C | D |
|---|---|
| c | 1 |
| b | 3 |
| a | 2 |

# Comparison of the terminology

| FORMAL DEFINITION | INFORMAL DEFINITION |
|---|---|
| relation | table |
| attribute | column |
| tuple | row |
| domain | data type |
| cardinality | number of rows |
| arity | number of columns |

A significant difference

FORMAL DEFINITION
lack of duplicates
INFORMAL DEFINITION
possible duplicates

# Example : university exams

course

| CID | TITLE | TEACHER |
|---|---|---|
| 1 | Maths | Leguichet |
| 2 | CS | Duchat |

# Example : university exams

exams

| SID | CID | DATE | GRADE |
|---|---|---|---|
| 123 | 1 | 7-9-03 | 10 |
| 123 | 2 | 8-1-03 | 8 |
| 702 | 2 | 7-9-03 | 5 |

# Example : university exams

student

| SID | NAME | CITY | MAJOR |
|---|---|---|---|
| 123 | Pierre | Paris | Inf |
| 415 | Celine | Bordeaux | Inf |
| 702 | Estelle | Rome | Log |

exams

| SID | CID | DATE | GRADE |
|---|---|---|---|
| 123 | 1 | 7-9-03 | 10 |
| 123 | 2 | 8-1-03 | 8 |
| 702 | 2 | 7-9-03 | 5 |

course

| CID | TITLE | TEACHER |
|---|---|---|
| 1 | Maths | Leguichet |
| 2 | CS | Duchat |

# Queries

- which teachers have graded Pierre?

student

| SID | NAME | CITY | MAJOR |
|---|---|---|---|
| 123 | Pierre | Paris | Inf |
| 415 | Celine | Bordeaux | Inf |
| 702 | Estelle | Rome | Log |

exams

| ID | CID | DATE | GRADE |
|---|---|---|---|
| 123 | 1 | 7-9-03 | 10 |
| 123 | 2 | 8-1-03 | 8 |
| 702 | 2 | 7-9-03 | 5 |

course

| CID | TITLE | TEACHER |
|---|---|---|
| 1 | Maths | Leguichet |
| 2 | CS | Duchat |

# Queries: examples

- which students have got grade 10 in Maths?

student

| SID | NAME | CITY | MAJOR |
|---|---|---|---|
| 123 | Pierre | Paris | Inf |
| 415 | Celine | Bordeaux | Inf |
| 702 | Estelle | Rome | Log |

exams

| ID | CID | DATE | GRADE |
|---|---|---|---|
| 123 | 1 | 7-9-03 | 10 |
| 123 | 2 | 8-1-03 | 8 |
| 702 | 2 | 7-9-03 | 5 |

course

| CID | TITLE | TEACHER |
|---|---|---|
| 1 | Maths | Leguichet |
| 2 | CS | Duchat |

## Incomplete information: it exists in real world

| Firstname | Forename | Lastname |
|---|---|---|
| Franklin | Delano | Roosevelt |
| Winston | | Churchill |
| Charles | | De Gaulle |
| Josip | | Stalin |



## Solution for incomplete information in the relational model

- Naive yet effective technique:
  - null value (NULL): indicates the lack of a value in the domain (as NULL does not belong to the domain!)
- Thus, each attribute may either have a value within the domain or a null value instead.



## Possible semantics of NULLs

- (at least) three different cases
  - unknown value
  - there exists no value
  - no information value
- Database systems consider the above values equivalent



## Integrity Constraints

- There exist database instances that are syntactically correct but do not correspond to the reality they need to represent.
- Therefore, instances need to be enriched with constraints, in order to guarantee the integrity of the data.

# Keys

- set of attributes that uniquely identifies the tuples of a relation r
- More formally:
  - a set K of attributes is a superkey for r if r does not contain two distinct tuples t1 and t2 with t1[K] = t2[K]
  - K is a key for r if it is a minimal superkey for r
    (i.e. it does not contain another superkey as a proper subset)



# Example of keys

| SID | NAME | CITY | MAJOR |
|---|---|---|---|
| 123 | Pierre | Paris | Inf |
| 107 | Arnaud | Lille | Log |
| 415 | Celine | Bordeaux | Inf |
| 702 | Estelle | Rome | Log |

- SID is a key:
  - is a superkey
  - is minimal as it contains only one attribute



# A key cannot be NULL

- If it was null, it would not be a superkey
- The goal in the relational model is to reduce the NULL values and provide unambiguous information



# Summary

- What we have seen today:
  - what is a relation
  - what is an attribute
  - what is a domain for an attribute
  - what is NULL
  - what is a key and a superkey in a relation