

# Basi di Dati

## Il Modello Relazionale

Raffaella Gentilini

September 30, 2020

## Introduzione al Modello Relazionale

### Definizione Formale del Modello Relazionale

Nozioni Preliminari

Tuple, Relazioni e Basi di Dati Relazionali

## Il modello relazionale

- Introdotto nel 1970 da E.F. Codd (un ricercatore dell'IBM) allo scopo di favorire l' **indipendenza dei dati**
- **Adottato dai sistemi commerciali** all'inizio degli **anni '80**
- Il modello relazionale e' **oggi il piu' diffuso modello di basi di dati sotteso ai DBMS commerciali**
- Tra le ragioni del suo successo:
  - **semplicita':** utente percepisce BD come insieme di tabelle
  - **carattere dichiarativo** dei linguaggi di manipolazione ed **interrogazione** associati
  - ...

## Il modello relazionale

Information Retrieval

P. BAXENDALE, Editor

### A Relational Model of Data for Large Shared Data Banks

E. F. CODD

IBM Research Laboratory, San Jose, California

Future users of large data banks must be protected from having to know how the data is organized in the machine (the internal representation). A prompting service which supplies such information is not a satisfactory solution. Activities of users at terminals and most application programs should remain unaffected when the internal representation of data is changed and even when some aspects of the external representation are changed. Changes in data representation will often be needed as a result of changes in query, update, and report traffic and natural growth in the types of stored information.

Existing noninferential, formatted data systems provide users with tree-structured files or slightly more general network models of the data. In Section 1, inadequacies of these models are discussed. A model based on  $n$ -ary relations, a normal form for data base relations, and the concept of a universal data sublanguage are introduced. In Section 2, certain operations on relations (other than logical inference) are discussed and applied to the problems of redundancy and consistency in the user's model.

**KEY WORDS AND PHRASES:** data bank, data base, data structure, data organization, hierarchies of data, networks of data, relations, derivability, redundancy, consistency, composition, join, retrieval language, predicate calculus, security, data integrity

**CR CATEGORIES:** 3.70, 3.73, 3.75, 4.20, 4.32, 4.39

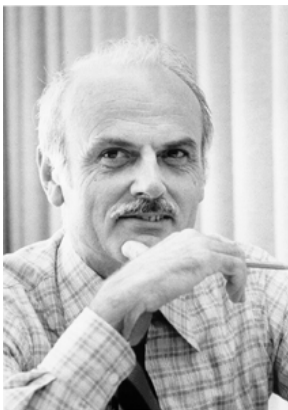
The relational view (or model) of data described in Section 1 appears to be superior in several respects to the graph or network model [3, 4] presently in vogue for non-inferential systems. It provides a means of describing data with its natural structure only—that is, without superimposing any additional structure for machine representation purposes. Accordingly, it provides a basis for a high level data language which will yield maximal independence between programs on the one hand and machine representation and organization of data on the other.

A further advantage of the relational view is that it forms a sound basis for treating derivability, redundancy, and consistency of relations—these are discussed in Section 2. The network model, on the other hand, has spawned a number of confusions, not the least of which is mistaking the derivation of connections for the derivation of relations (see remarks in Section 2 on the “connection trap”).

Finally, the relational view permits a clearer evaluation of the scope and logical limitations of present formatted data systems, and also the relative merits (from a logical standpoint) of competing representations of data within a single system. Examples of this clearer perspective are cited in various parts of this paper. Implementations of systems to support the relational model are not discussed.

#### 1.2. DATA DEPENDENCIES IN PRESENT SYSTEMS

The provision of data description tables in recently developed information systems represents a major advance toward the goal of data independence [5, 6, 7]. Such tables facilitate changing certain characteristics of the data representation stored in a data bank. However, the variety of data representation characteristics which can be changed without logically impairing some application programs is still quite limited. Further, the model of data with which users interact in these systems is not a representationally prop-



The adverse impact on development productivity of requiring programmers to navigate along access paths to reach target data [...] was enormous.

In addition, it was not possible to make slight changes in the layout in storage without simultaneously having to revise all programs that relied on the previous structure. [...]

As a result, far too much manpower was being invested in continual (and avoidable) maintenance of application programs.

E. Codd *Turing Award Lecture, 1981*

# Modello relazionale vs Modelli Reticolare/Gerarchico

Le **principali differenze** tra modello relazionale e modelli gerarchico/reticolare sono:

- modo in cui si rappresentano le associazioni tra record:
  - gerarchico e reticolare usano **puntatori**
  - **valori**, nel modello relazionale
- A differenza dei modelli gerarchico/reticolare, **il modello relazionale è formalmente definito** e trae fondamento nella teoria degli insiemi e nella logica dei predicati al primo ordine



## Nozioni Preliminari

Il modello relazionale si basa sui concetti di relazione e tupla, variazione delle nozioni matematiche di relazione ed  $n$ -upla.

### Definizione [Relazioni Matematiche]

Si considerino  $n$  insiemi  $D_1, \dots, D_n$  (non necessariamente distinti):

- il prodotto cartesiano  $D_1 \times \dots \times D_n$  e' l'insieme di tutte le  $n$ -uple ordinate  $(d_1 \dots d_n)$  tali che  $d_1 \in D_1, \dots, d_n \in D_n$
- una relazione (matematica)  $R$  su  $D_1, \dots, D_n$  e' un sottoinsieme del prodotto cartesiano  $R \subseteq D_1 \times \dots \times D_n$
- il grado di  $R \subseteq D_1 \times \dots \times D_n$  e'  $n$
- La cardinalita'  $|R|$  di  $R \subseteq D_1 \times \dots \times D_n$  e' il numero di elementi di  $R$ .



# Dalle Relazioni Matematiche alle Relazioni nel Modello Relazionale

- Gli elementi di una relazione matematica  $R$  sono  $n$ -uple ordinate nella forma  $(d_1 \dots d_n)$ ,  $d_i \in D_i, i \in \{1, \dots, n\}$
- Nel contesto delle basi di dati l'ordine degli elementi in una  $n$ -upla non e' importante.
- Risulta piuttosto conveniente essere in grado di nominare un particolare elemento  $d_i$ , piuttosto che individuarlo mediante il suo indice  $i$ .





## Domini ed Attributi

Diamo dunque le seguenti definizioni:

### Domini ed Attributi

Un **dominio** e' un **insieme non vuoto di valori atomici** (indivisibili) .

Un **attributo** e' un **nome associato ad un dominio**.

Indichiamo con  $Dom(A)$  il dominio associato all'attributo  $A$ .

### Example

$Dom(Nazione)$ : insieme di stringhe di caratteri che rappresentano nomi di nazione.

# Tuple

Siamo ora in grado di introdurre il concetto fondamentale di **tupla**, variazione della nozione di  $n$ -upla dove l'ordine degli elementi non ha rilevanza.

## Tupla

Si consideri un insieme di attributi  $X = \{A_1, \dots, A_n\}$ .

Una tupla  $t$  su  $X$  è una funzione che associa ad ogni  $A_i \in X$  un valore in  $Dom(A_i)$ , oppure uno speciale valore Null.

Indicheremo con  $t[A_i]$  il valore della tupla  $t$  sull'attributo  $A_i$ .



## Relazioni: Schemi ed Istanze

Precediamo definendo le nozioni di **schema di relazione** ed **istanza di relazione** nel modello relazionale.

### Schema di Relazione

Dato l'insieme di attributi  $X = \{A_1, \dots, A_n\}$ , uno schema di relazione su  $X$  e' dato da: un nome (di relazione)  $R$  e dall'insieme di attributi  $X$ .

Utilizzeremo la notazione  $R(X)$  (oppure  $R(A_1 \dots A_n)$  in luogo di  $R(\{A_1, \dots, A_n\})$ ) per indicare uno schema di relazione  $R$  su  $X$ .  
Sara' inoltre a volte conveniente specificare i domini di ogni attributo, scrivendo  $R(A_1 : Dom(A_1), \dots, A_n : Dom(A_n))$ .

### Istanza di Relazione

Dato l'insieme di attributi  $X = \{A_1, \dots, A_n\}$ , un'istanza di relazione su  $X$  e' un insieme di tuple su  $X$ .

## Basi di Dati: Schemi ed Istanze

Possiamo ora dare la seguente definizione parziale della nozione di schema/istanza di basi di dati.

La definizione completa include anche la specifica di un insieme di vincoli di integrita', di cui parleremo nella prossima lezione.

### Schema di Basi di Dati

Uno **schema di basi di dati** e' un **insieme di schemi di relazioni con nomi diversi**.

### Istanza di Basi di Dati

Un' **istanza di basi di dati** sullo schema :

$$\mathcal{B} = \{R_1(X_1), \dots, R_n(X_n)\}$$

e' un **insieme di istanze di relazioni**  $\{r_1, \dots, r_n\}$  tali che  $r_i$  e' un'istanza di  $R_i$ , per ogni  $i \in \{1 \dots n\}$ .