

# I modelli logici dei dati

I modelli logici *tradizionali* sono tre:

- **gerarchico**
- **reticolare**
- **relazionale**

I modelli gerarchico e reticolare sono più vicini alle strutture fisiche di memorizzazione. Quello relazionale è più astratto: vengono rappresentati solo valori. Anche i riferimenti tra record in relazioni diverse sono espressi tramite valori, e non si fa uso di riferimenti espliciti (puntatori)

Recentemente è stato introdotto il modello ad **oggetti**

# Il modello relazionale

Fu proposto da E.F. Codd nel 1970 per favorire l'indipendenza dei dati e reso disponibile come modello logico in DBMS reali nel 1981 (solo nella metà degli anni ottanta prende piede sul mercato)

Prende il nome dal concetto matematico di relazione, cui si ispira

Le relazioni hanno una rappresentazione naturale per mezzo di tabelle

# Tre accezioni di “relazione”

Relazione matematica: definizione rigorosa derivante dalla teoria degli insiemi (la richiameremo nel seguito)

Relazione nel senso di associazione o correlazione (in inglese *relationship*) utilizzata dai modelli concettuali per descrivere un collegamento tra due entità reali. Per esempio: l'entità **corso** è in relazione con il **docente** che lo svolge

Relazione secondo il modello relazionale dei dati: costruttore che permette di definire insiemi di record omogenei

# Relazione matematica

$D_1, D_2, \dots, D_n$  sono insiemi (non necessariamente distinti)

Il **prodotto cartesiano**  $D_1 \times D_2 \times \dots \times D_n$  è l'insieme di tutte le possibili ennuple ordinate  $(d_1, d_2, \dots, d_n)$  tali che  $d_i \in D_i$

Una **relazione matematica** sugli insiemi  $D_1, D_2, \dots, D_n$  è un sottoinsieme del loro prodotto cartesiano

La **cardinalità** della relazione è il numero di n-ple (nelle applicazioni reali sempre finito)

# Esempio di relazione matematica

$$D_1 = \{a, b\}$$

$$D_2 = \{x, y, z\}$$

prodotto cartesiano è

$$D_1 \times D_2 = \{ (a, x), (a, y), (a, z), \\ (b, x), (b, y), (b, z) \}$$

una relazione  $r \subseteq D_1 \times D_2$  può essere

$$r = \{ (a, x), (a, z), (b, y), (b, z) \}$$

la sua cardinalità è quattro

# Proprietà della relazione matematica

- non è definito alcun ordinamento tra le ennuple della relazione
- le ennuple di una relazione sono distinte l'una dall'altra (la relazione è un insieme e non un multiinsieme)
- le ennuple sono ordinate al loro interno

partite  $\subseteq$  squadra x squadra x punti x punti

Rossi	Verdi	3	1
Verdi	Gialli	2	0
Rossi	Blu	1	2
Blu	Gialli	0	1

*La struttura è posizionale*

# Relazione nel modello relazionale dei dati

A ciascun dominio è associato un nome (attributo), unico nella relazione, che *descrive* il ruolo nel dominio

Ne consegue che l'ordinamento degli attributi è irrilevante (la struttura è non posizionale)

Nella rappresentazione tabellare gli attributi sono generalmente usati come intestazioni delle colonne

Casa	Fuori	Reti1	Reti2
Rossi	Verdi	3	1
Verdi	Gialli	2	0
Rossi	Blu	1	2
Blu	Gialli	0	1

# Definizione formale

Definiamo una funzione **dom** che dato un attributo  $A$  ne ritorna il dominio

Es:      $\text{dom}(\text{Casa}) = \text{stringa}$   
           $\text{dom}(\text{Reti1}) = \text{intero}$

Una **ennupla** su un insieme di attributi  $X$  associa ad ogni attributo  $A$  un valore del suo dominio  $\text{dom}(A)$

Una **relazione** su un insieme di attributi  $X$  è un insieme di ennuple su quegli attributi (nota: *insieme* e non *multiinsieme*)



# Notazioni

Con lettere maiuscole iniziali  
dell'alfabeto indichiamo attributi:

A, B, C, ...

Con lettere maiuscole finali  
indichiamo insiemi di attributi:

X, Y, Z, ...

Se  $t$  è una ennupla su  $X$  e  $A \in X$ ,  
allora  $t[A]$  indica il valore di  $t$  su  $A$

Es:  $t =$

Casa	Fuori	Reti1	Reti2
Rossi	Verdi	3	1

$t[\text{Fuori}] = \text{Verdi}$

Lo stesso vale per insiemi  $X$  di  
attributi:  $t[\text{Fuori}, \text{Reti2}]$  è una ennupla  
sui due attributi

# Tabelle e relazioni

Una tabella rappresenta una relazione se:

- i valori di ciascuna colonna sono tra loro omogenei (stesso dominio)
- le righe sono diverse tra loro
- le intestazioni delle colonne sono diverse tra loro

Inoltre, se una tabella rappresenta una relazione:

- l'ordinamento delle righe interne alla tabella è irrilevante
- l'ordinamento delle colonne è irrilevante

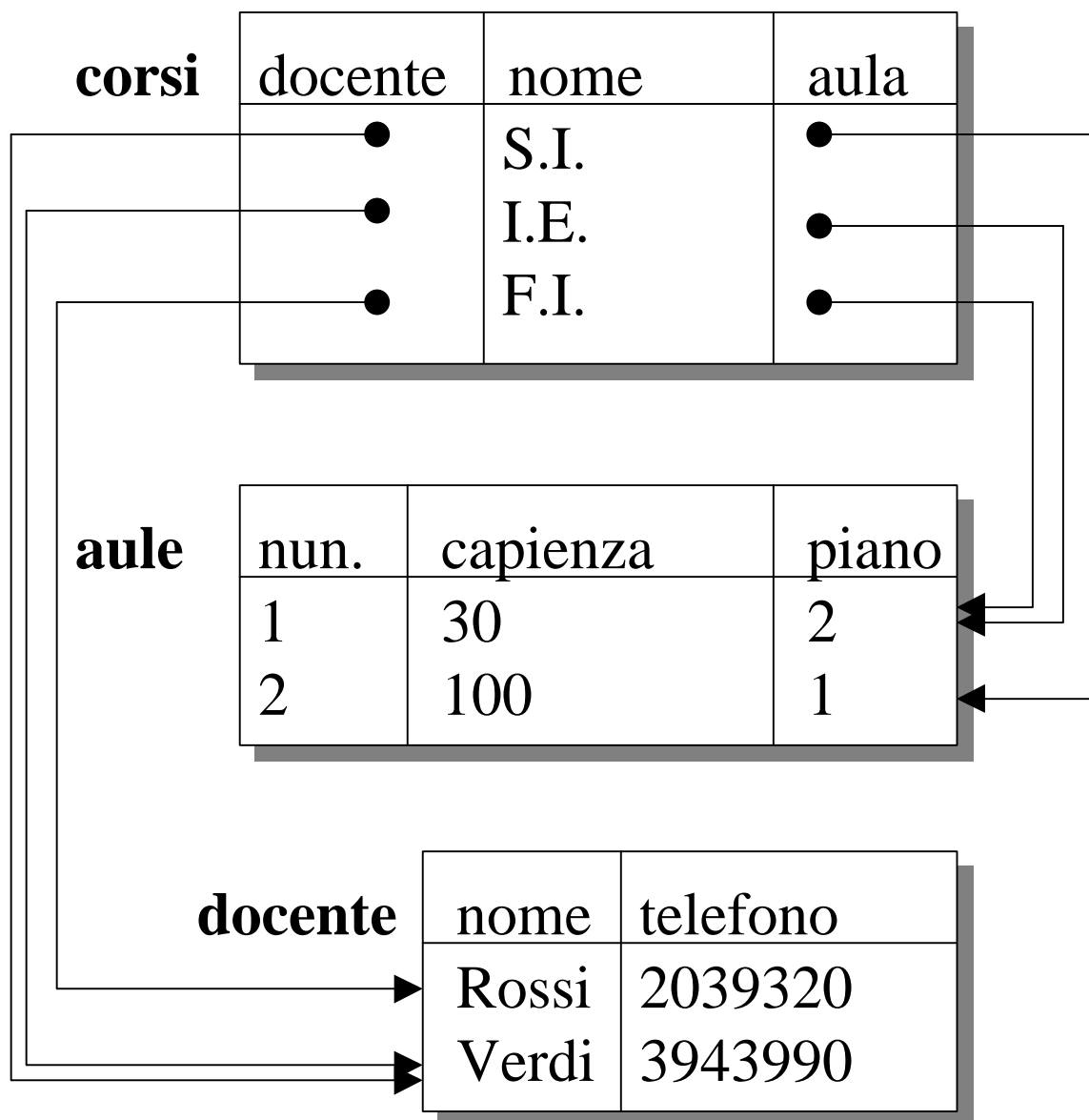
# Esempio di riferimenti per valore

<b>corsi</b>	corso	docente	aula
	S.I.	Rossi	2
	I.E.	Rossi	1
	F.I.	Verdi	1

<b>aule</b>	nun.	capienza	piano
	1	30	2
	2	100	1

<b>docente</b>	nome	telefono
	Rossi	2039320
	Verdi	3943990

# Esempio di riferimenti con puntatori



# Vantaggi dell'uso di valori

- indipendenza dalle strutture fisiche che possono cambiare arbitrariamente
- si rappresenta solo ciò che è rilevante dal punto di vista dell'applicazione e dell'utente (i puntatori sono meno comprensibili per l'utente finale)
- i dati sono portabili più facilmente da un sistema ad un altro
- i puntatori sono direzionali e dunque meno espressivi di una coincidenza di valori

## NOTE:

- i puntatori possono continuare ad esistere a livello fisico
- nel modello ad oggetti esistono puntatori “ad alto livello” che consentono una indipendenza dalle strutture fisiche

# Definizioni

**Schema di relazione:** un nome di relazione  $R$  con un insieme di attributi distinti  $A_1, \dots, A_n$

$$R(A_1, \dots, A_n)$$

**Schema di base di dati:** un insieme di schemi di relazioni con nomi diversi

$$\mathbf{R} = \{R_1(X_1), \dots, R_m(X_m)\}$$

**Istanza di relazione** su uno schema  $R(X)$ : insieme  $r$  di ennuple su  $X$

**Istanza di base di dati** su uno schema

$$\mathbf{R} = \{R_1(X_1), \dots, R_m(X_m)\}: \text{insieme di relazioni}$$
$$\mathbf{r} = \{r_1, \dots, r_p\}$$

dove  $r_i$  è una relazione su  $R_i(X_i)$

# Esempio

<b>studenti</b>	nome	matricola
	Rossi	2039320
	Verdi	3943990

<b>studenti</b>	matricola
<b>lavoratori</b>	2039320
	3943990

*sono possibili relazioni su un solo attributo*

# Esempio con strutture nidificate

DA FILIPPO PIZZERIA	
Ricevuta 2456 Del 12-05-2000	
3 coperti	6000
2 antipasti	12000
3 primi	27000
2 bistecche	36000
<b>Totale</b>	<b>81000</b>

DA FILIPPO PIZZERIA	
Ricevuta 2457 Del 12-05-2000	
2 coperti	4000
1 antipasti	6000
2 primi	15000
2 orate	50000
2 caffè	3000
<b>Totale</b>	<b>78000</b>



# Schema logico

## ricevute

numero	data	totale
2456	12-05-2000	81000
2457	12-05-2000	78000

## dettaglio

numero	quantità	descrizione	importo
2456	3	coperti	6000
2456	2	antipasti	12000
2456	3	primi	27000
2456	2	bistecche	36000
2457	2	coperti	4000
2457	1	antipasti	6000
2457	2	primi	15000
2457	2	orate	50000
2457	2	caffè	3000

# Verifica dello schema logico

Abbiamo rappresentato veramente tutti gli aspetti delle ricevute?

Dipende da che cosa ci interessa realmente!

- l'ordine delle righe è rilevante?
- possono esistere linee ripetute in una ricevuta?

Sono possibili rappresentazioni diverse

# Schema logico alternativo

## ricevute

numero	data	totale
2456	12-05-2000	81000
2457	12-05-2000	78000

## dettaglio

numero	riga	quantità	descrizione	importo
2456	1	3	coperti	6000
2456	2	2	antipasti	12000
2456	3	3	primi	27000
2456	4	2	bistecche	36000
2457	1	2	coperti	4000
2457	2	1	antipasti	6000
2457	3	2	primi	15000
2457	4	2	orate	50000
2457	5	2	caffè	3000

# Informazioni incomplete

Il modello relazionale impone ai dati una struttura rigida:

- le informazioni sono rappresentate per mezzo di ennuple
- solo alcuni formati di ennuple sono ammessi, quelli che corrispondono agli schemi di relazione

All'atto pratico i dati possono non corrispondere esattamente al formato previsto!

# Informazioni incomplete: esempio

Città	nome	prefettura
	Roma	Via IV Novembre
	Firenze	
	Tivoli	
	Prato	

- Firenze è provincia ma non conosciamo l'indirizzo della prefettura
- Tivoli non è provincia e non ha prefettura
- Prato è “nuova” provincia: ha la prefettura?

# Informazioni incomplete: possibili soluzioni

Non conviene (anche se spesso si fa) utilizzare valori ordinari del dominio (lo zero, la stringa nulla, il valore massimo, valori incoerenti, ecc.) per indicare informazioni incomplete:

- potrebbero non esistere valori non utilizzati o privi di significato
- in fase di utilizzo sarebbe necessario tener conto di quali valori, per quel determinato attributo, segnalano informazioni incomplete

# Informazione incompleta nel modello relazionale

Nel modello relazionale si adotta una tecnica rudimentale ma efficace: si definisce un **valore nullo** NULL che denota l'assenza di un valore del dominio (e non è un valore del dominio)

Formalmente è sufficiente estendere il concetto di ennupla:  $t[A]$ , per ogni attributo  $A$ , è un valore del dominio  $\text{dom}(A)$  oppure il valore nullo NULL

Si possono (e debbono) imporre restrizioni sulla presenza di valori nulli

# Eccesso di valori nulli

<b>corsi</b>	corso	docente	codice
	S.I.	Gialli	C12
	I.E.	NULL	C22
	F.I.	Verdi	NULL

<b>esami</b>	corso	matricola	voto
	NULL	2039320	27
	C22	NULL	28

<b>studenti</b>	nome	matricola
	Rossi	2039320
	NULL	3943990



# Tipi di valori nulli

Almeno tre casi differenti:

- **valore sconosciuto:** esiste un valore nel dominio, ma non è noto al momento dell'inserimento dei dati
- **valore inesistente:** non esiste un valore del dominio (lo schema della base di dati non riflette la complessità della realtà)
- **valore senza informazione:** non è noto se esista o meno un valore del dominio

I DBMS non distinguono tra questi tipi di valore nullo (e quindi implicitamente adottano il **valore senza informazione**)