MODELLI DI DATI PER SISTEMI DI DATABASE

I modelli logici dei dati



Un *modello dei dati* è un insieme di concetti utilizzati per organizzare i dati e descriverne la struttura in modo che essa risulti comprensibile ad un elaboratore

Modelli logici

- Gerarchico e reticolare
 - utilizzano riferimenti espliciti (puntatori) fra record
 - Vietata la ridondanza (reticolare)
- Relazionale "è basato su valori"
 - o anche i riferimenti fra dati in strutture (relazioni) diverse sono rappresentati per mezzo dei valori stessi
- A oggetti (di recente introduzione)

Il modello relazionale



- Proposto da E. F. Codd nel 1970 per favorire l'indipendenza dei dati
- Disponibile in DBMS reali nel 1981 (non è facile implementare l'indipendenza con efficienza e affidabilità!)
- Si basa sul concetto matematico di relazione (con una variante)
- Le relazioni hanno naturale rappresentazione per mezzo di tabelle

Relazione: tre accezioni



- relazione matematica: come nella teoria degli insiemi
- relazione (dall'inglese relationship) che rappresenta una classe di fatti, nel modello Entity-Relationship; tradotto anche con associazione o correlazione
- relazione secondo il modello relazionale dei dati

Relazione matematica



- $D_1, ..., D_n$ (n insiemi anche non distinti)
- prodotto cartesiano $D_1 \times ... \times D_n$: l'insieme di tutte le n-uple $(d_1, ..., d_n)$ tali che $d_1 \in D_1, ..., d_n \in D_n$
- relazione matematica su D_1 , ..., D_n : un sottoinsieme di $D_1 \times ... \times D_n$.
- $D_1, ..., D_n$ sono i domini della relazione

Relazione matematica, esempio



•
$$D_1 = \{a, b\}$$

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

prodotto cartesiano $D_1 \times D_2$

a y a z b x

b y

una relazione r

$$r \subseteq D_1 \times D_2$$

a x a z b y

Relazione matematica, proprietà



Una relazione matematica è un insieme di n-uple ordinate:

$$(d_1, ..., d_n)$$
 tali che $d_1 \in D_1, ..., d_n \in D_n$

Oss: una relazione è un insieme; quindi:

- o non c'è ordinamento fra le *n*-uple;
- o le *n*-uple sono distinte
- o ciascuna *n*-upla è ordinata: l' *i*-esimo valore proviene dall' *i*-esimo dominio

Relazione matematica, esempio



 $Partite \subseteq string \times string \times int \times int$

```
Juve Lazio 3 1
Lazio Milan 2 0
Juve Roma 0 2
Roma Milan 0 1
```

- Ciascuno dei domini ha due ruoli diversi, distinguibili attraverso la posizione:
 - La struttura è posizionale

Struttura non posizionale

10

A ciascun dominio si associa un nome (*attributo*), che ne descrive il "ruolo"

Casa	Fuori	RetiCasa	RetiFuori
Juve	Lazio	3	1
Lazio	Milan	2	0
Juve	Roma	0	2
Roma	Milan	0	1

Una tupla su un insieme di attributi X è una funzione t che associa a ciascun attributo A un valore del suo dominio. Una tupla valore tupla val

Tabelle e relazioni



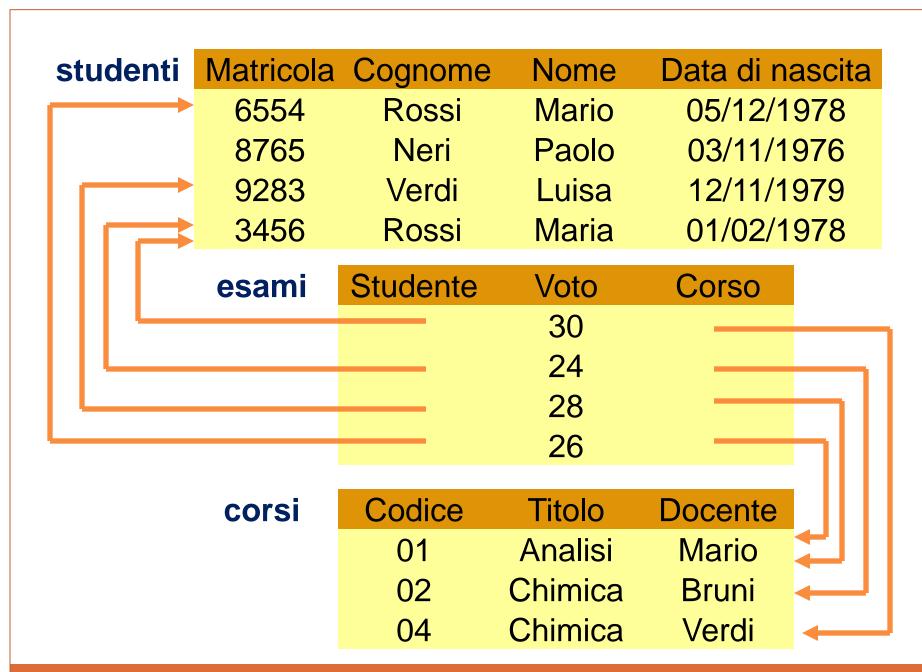
- Una tabella rappresenta una relazione se
 - o i valori di ogni colonna sono fra loro omogenei
 - o le righe sono diverse fra loro
 - o le intestazioni delle colonne sono diverse tra loro
- In una tabella che rappresenta una relazione
 - o l'ordinamento tra le righe è irrilevante
 - o l'ordinamento tra le colonne è irrilevante

Il modello è basato su valori



Il modello relazionale è basato su valori. Ciò significa che i riferimenti fra dati in relazioni diverse sono rappresentati per mezzo di valori dei domini che compaiono nelle ennuple.

studenti	Matricola	Cognome	Nome	Data di na	ascita
	6554	Rossi	Mario	05/12/1	978
	8765	Neri	Paolo	03/11/1	976
	9283	Verdi	Luisa	12/11/1	979
	3456	Rossi	Maria	01/02/1	978
	esami	Studente	Voto	Corso	
		3456	30	04	
		3456	24	02	
		9283	28	01	
		6554	26	01	
	corsi	Codice	Titolo	Docente	
		01	Analisi	Mario	
		02	Chimica	Bruni	
		04	Chimica	Verdi	



Vantaggi della struttura basata su valori



- indipendenza dalle strutture fisiche (si potrebbe avere anche con puntatori di alto livello) che possono cambiare dinamicamente. La rappresentazione logica dei dati (costituita dai soli valori) non fa riferimento a quella fisica
- si rappresenta solo ciò che è rilevante dal punto di vista dell'applicazione
- i dati sono portabili più facilmente da un sistema ad un altro
- i puntatori sono direzionali

Definizioni



• Schema di relazione: un nome R con un insieme di attributi $A_1, ..., A_n$:

$$R(A_1,...,A_n)$$

 Schema di base di dati: insieme di schemi di relazione:

$$R = \{R_1(X_1), ..., R_k(X_k)\}$$

Definizioni, 2



• Una tupla su un insieme di attributi *X*, denotata con *t*, è una funzione che associa a ciascun attributo *A* in *X* un valore del dominio di *A*

• *t*[*A*] denota il valore della tupla *t* sull'attributo *A*

Definizioni, 3



• Un'istanza di relazione o relazione su uno schema R(X) è l'insieme r di tuple su X

• Un'istanza di base di dati su uno schema $R = \{R_1(X_1), ..., R_n(X_n)\}$ è l'insieme delle relazioni $r = \{r_1, ..., r_n\}$ (con r_i relazione su R_i)

Relazioni su singoli attributi



studenti

Matricola	Cognome	Nome	Data di nascita
6554	Rossi	Mario	05/12/1978
8765	Neri	Paolo	03/11/1976
9283	Verdi	Luisa	12/11/1979
3456	Rossi	Maria	01/02/1978

studenti lavoratori Matricola 6554 3456

Strutture nidificate

	Da Filippo Via Roma 2, Roma					
	Ricevuta Fiscale 1235 del 12/10/2000					
3	Coperti	3,00				
2	Antipasti	6,20				
3	Primi	12,00				
2	Bistecche	18,00				
	<i>Totale</i> 39,20					

Da Filippo Via Roma 2, Roma Ricevuta Fiscale 1240 del 13/10/2000 2,00 Coperti **Antipasti** 7,00 Primi 8,00 Orate 20,00 2,00 Caffè 39,00 Totale

Relazioni che rappresentano strutture nidificate

21

Ricevute

Nume	ero	Data	Tot	ale
123	5 1	2/10/200	39	,20
124	0 1	3/10/200	39	,00

Dettaglio

Numero	Qtà	Coperti	Prezzo
1235	3	Coperti	3,00
1235	2	Antipasti	6,20
1235	3	Primi	12,00
1235	2	Bistecche	18,00
1240	2	Coperti	2,00

Strutture nidificate, riflessione



- Abbiamo rappresentato veramente tutti gli aspetti delle ricevute?
- Dipende da che cosa ci interessa!
 - o l'ordine delle righe è rilevante?
 - o possono esistere linee ripetute in una ricevuta?
- Sono possibili rappresentazioni diverse

Rappresentazione alternativa per strutture nidificate

(23)

Ricevute

Numero	Data	Totale
1235	12/10/2000	39,20
1240	13/10/2000	39,00

Dettaglio

Numero	Riga	Qtà	Descrizione	Importo
1235	1	3	Coperti	3,00
1235	2	2	Antipasti	6,20
1235	3	3	Primi	12,00
1235	4	2	Bistecche	18,00
1240	1	2	Coperti	2,00

Informazione incompleta



- ll modello relazionale impone ai dati una struttura rigida:
 - o le informazioni sono rappresentate per mezzo di ennuple
 - o solo alcuni formati di ennuple sono ammessi: quelli che corrispondono agli schemi di relazione
- I dati disponibili possono non corrispondere al formato di tupla previsto

Informazione incompleta: motivazioni



Nome	SecondoNome	Cognome
Franklin	Delano	Roosevelt
Winston		Churchill
Charles		De Gaulle
Josip		Stalin

Informazione incompleta: soluzioni?



- non conviene (anche se spesso si fa) usare valori del dominio (o, stringa nulla, "99", ...):
 - o potrebbero non esistere valori "non utilizzati"
 - o valori "non utilizzati" potrebbero diventare significativi
 - o in fase di utilizzo (nei programmi) sarebbe necessario ogni volta tener conto del "significato" di questi valori

Informazione incompleta nel modello relazionale



- Si adotta una tecnica rudimentale ma efficace:
 - o valore nullo: denota l'assenza di un valore del dominio (e non è un valore del dominio)
- t[A], per ogni attributo A, è un valore del dominio dom(A) oppure il valore nullo NULL
- Si possono (e debbono) imporre restrizioni sulla presenza di valori nulli

Troppi valori nulli

(28)

4		4.5
Cti	$1 \alpha \alpha$	ntı
่อแ	ıde	71 I L I

Matricola	Cognome	Nome	Data di nascita
6554	Rossi	Mario	05/12/1978
9283	Verdi	Luisa	12/11/1979
NULL	Rossi	Maria	01/02/1978

esam	
	ı
COGIII	

Studente	Voto	Corso
NULL	30	NULL
NULL	24	02
9283	28	01

corsi

Codice	Titolo	Docente
01	Analisi	Mario
02	Chimica	NULL
NULL	Chimica	Verdi

Significato dei valori nulli



Città	Indirizzo prefettura	
Roma	Via Quattro Novembre	sconosciuto
Firenze	NULL	
Tivoli	NULL	> inesistente
Crotone	NULL	
		Senza
		informazione

Tipi di valore nullo



- (almeno) tre casi differenti
 - o valore sconosciuto
 - o valore inesistente
 - o valore senza informazione
- I DBMS non distinguono i tipi di valore nullo

Vincoli di integrità

31

Esistono istanze di basi di dati che, pur sintatticamente corrette, non rappresentano informazioni possibili per l'applicazione di interesse

Una base di dati "scorretta"



Esami	Studente	Voto	Lode	Corso
	276545	32		01
	276545	30	e lode	02
	787643	27	e lode	03
	739430	24		04

0		- 1			
St	11	М	Δ	n	tı.
Uι	u	u	し	11	u

Matricola	Cognome	Nome
276545	Rossi	Mario
787643	Neri	Piero
787643	Bianchi	Luca

Vincolo di integrità



Un vincolo d'integrità è una proprietà che deve essere soddisfatta dalle istanze che rappresentano informazioni corrette per l'applicazione

Un vincolo è una funzione booleana (un predicato): associa ad ogni istanza il valore vero o falso

Vincoli di integrità, perché?



- Descrizione più accurata della realtà
- Contributo alla "qualità dei dati"
- Utili nella progettazione (vedremo)
- Usati dai DBMS nella esecuzione delle interrogazioni
- Non tutte le proprietà di interesse sono rappresentabili per mezzo di vincoli formulabili in modo esplicito

Tipi di vincoli



o vincoli intrarelazionali:

sono i vincoli che devono essere rispettati dai valori contenuti nella relazione considerata

- ×vincoli su valori (o di dominio)
- ×vincoli di ennupla

o vincoli interrelazionali

sono i vincoli che devono essere rispettati da valori contenuti in relazioni diverse

Esami	Studente	Voto	Lode	Corso
	276545	32		01
	276545	30	e lode	02
	787643	27	e lode	03
	739430	24		04

Studenti

Matricola	Cognome	Nome
276545	Rossi	Mario
787643	Neri	Piero
787643	Bianchi	Luca

Vincoli di ennupla



- I Vincoli di ennupla esprimono condizioni sui valori di ciascuna ennupla, indipendentemente dalle altre ennuple
- Caso particolare:
 - **▼Vincoli di dominio:** coinvolgono un solo attributo

Sintassi ed esempi



- Una possibile sintassi:
 - espressione booleana di atomi che confrontano valori di attributo o espressioni aritmetiche su di essi

$$(Voto \ge 18) AND (Voto \le 30)$$

$$(Voto = 30) OR NOT (Lode = "e lode")$$

Vincoli di ennupla, esempio



Stipendi

Impiegato	Lordo	Ritenute	Netto
Rossi	55.000	12.500	42.500
Neri	45.000	10.000	35.000
Bruni	47.000	11.000	36.000

Lordo = (Ritenute + Netto)

Identificazione delle ennuple



Matricola	Cognome	Nome	Corso	Nascita
27655	Rossi	Mario	Ing Inf	5/12/78
78763	Rossi	Mario	Ing Inf	3/11/76
65432	Neri	Piero	Ing Mecc	10/7/79
87654	Neri	Mario	Ing Inf	3/11/76
67653	Rossi	Piero	Ing Mecc	5/12/78

- non ci sono due ennuple con lo stesso valore sull'attributo Matricola
- o non ci sono due ennuple uguali su tutti e tre gli attributi Cognome, Nome e Data di Nascita

Chiave



• Una chiave è un insieme di attributi che identificano le ennuple di una relazione

Formalmente:

- un insieme K di attributi è superchiave per r se r non contiene due ennuple distinte t_1 e t_2 con $t_1[K] = t_2[K]$
- K è chiave per r se è una superchiave minimale per r (cioè non contiene un'altra superchiave)

Una chiave



Matricola	Cognome	Nome	Corso	Nascita
27655	Rossi	Mario	Ing Inf	5/12/78
78763	Rossi	Mario	Ing Inf	3/11/76
65432	Neri	Piero	Ing Mecc	10/7/79
87654	Neri	Mario	Ing Inf	3/11/76
67653	Rossi	Piero	Ing Mecc	5/12/78

- Matricola è una chiave:
 - o è superchiave
 - o contiene un solo attributo e quindi è minimale

Un'altra chiave



Matricola	Cognome	Nome	Corso	Nascita
27655	Rossi	Mario	Ing Inf	5/12/78
78763	Rossi	Mario	Ing Inf	3/11/76
65432	Neri	Piero	Ing Mecc	10/7/79
87654	Neri	Mario	Ing Inf	3/11/76
67653	Rossi	Piero	Ing Mecc	5/12/78

- Cognome, Nome, Nascita è un'altra chiave:
 - o è superchiave
 - o minimale

Un'altra chiave?

(44)	
11		_

Matricola	Cognome	Nome	Corso	Nascita
27655	Rossi	Mario	Ing Inf	5/12/78
78763	Rossi	Mario	Ing Civile	3/11/76
65432	Neri	Piero	Ing Mecc	10/7/79
87654	Neri	Mario	Ing Inf	3/11/76
67653	Rossi	Piero	Ing Mecc	5/12/78

- Non ci sono ennuple uguali su Cognome e Corso:
 - o Cognome e Corso formano una chiave
- Ma è sempre vero?

Vincoli, schemi e istanze

- i vincoli corrispondono a proprietà del mondo reale modellato dalla base di dati
- interessano a livello di schema (con riferimento cioè a tutte le istanze possibili)
- ad uno schema associamo un insieme di vincoli e consideriamo corrette (valide, ammissibili) le istanze che soddisfano tutti i vincoli
- un'istanza può soddisfare altri vincoli ("per caso")

Studenti

Matricola Cognome Nome Corso Nascita

• chiavi:

Matricola Cognome, Nome, Nascita

Matricola	Cognome	Nome	Corso	Nascita
27655	Rossi	Mario	Ing Inf	5/12/78
78763	Rossi	Mario	Ing Civile	3/11/76
65432	Neri	Piero	Ing Mecc	10/7/79
87654	Neri	Mario	Ing Inf	3/11/76
67653	Rossi	Piero	Ing Mecc	5/12/78

- È corretta: soddisfa i vincoli
- Ne soddisfa anche altri ("per caso"):
 - o Cognome, Corso è chiave

Esistenza delle chiavi



- Una relazione non può contenere ennuple distinte ma con valori uguali (una relazione è un sottoinsieme del prodotto cartesiano)
- Ogni relazione ha sicuramente come superchiave l'insieme di tutti gli attributi su cui è definita
- e quindi ha (almeno) una chiave

Importanza delle chiavi



- l'esistenza delle chiavi garantisce l'accessibilità a ciascun dato della base di dati
- le chiavi permettono di correlare i dati in relazioni diverse:
 - o il modello relazionale è basato su valori

Chiavi e valori nulli



La presenza di valori nulli fra i valori di una chiave non permette

- o di identificare le ennuple
- o di realizzare facilmente i riferimenti da altre relazioni

Matricola	Cognome	Nome	Corso	Nascita
NULL	NULL	Mario	Ing Inf	5/12/78
78763	Rossi	Mario	Ing Civile	3/11/76
65432	Neri	Piero	Ing Mecc	10/7/79
87654	Neri	Mario	Ing Inf	NULL
NULL	Neri	Mario	NULL	5/12/78

La presenza di valori nulli nelle chiavi deve essere limitata

Chiave primaria

- Una chiave primaria è una chiave su cui non sono ammessi valori nulli
- Notazione: sottolineatura

Matricola	Cognom	e Nome	Corso	Nascita
86765	NULL	Mario	Ing Inf	5/12/78
78763	Rossi	Mario	Ing Civile	3/11/76
65432	Neri	Piero	Ing Mecc	10/7/79
87654	Neri	Mario	Ing Inf	NULL
43289	Neri	Mario	NULL	5/12/78

Integrità referenziale



• Nel modello relazionale le informazioni in relazioni diverse sono correlate attraverso valori comuni

• in particolare, vengono spesso presi in considerazione i valori delle chiavi (primarie).

• le correlazioni debbono essere "coerenti"

Infrazioni

<u>Codice</u>	Data	Vigile	Prov	Numero
34321	1/2/95	3987	MI	39548K
53524	4/3/95	3295	TO	E39548
64521	5/4/96	3295	PR	839548
73321	5/2/98	9345	PR	839548

Vigili

Matricola	Cognome	Nome
3987	Rossi	Luca
3295	Neri	Piero
9345	Neri	Mario
7543	Mori	Gino

Infrazioni

<u>Codice</u>	Data	Vigile	Prov	Numero
34321	1/2/95	3987	MI	39548K
53524	4/3/95	3295	TO	E39548
64521	5/4/96	3295	PR	839548
73321	5/2/98	9345	PR	839548

A	u [.]	to
	•	_

<u>Prov</u>	<u>Numero</u>	Cognome	Nome
MI	39548K	Rossi	Mario
TO	E39548	Rossi	Mario
PR	839548	Neri	Luca

Vincolo di integrità referenziale



Un vincolo di integrità referenziale ("foreign key") fra gli attributi X di una relazione R₁ e un'altra relazione R₂ impone ai valori su X in R₁ di comparire come valori della chiave primaria di R₂

Vincoli di integrità referenziale



- vincoli di integrità referenziale fra:
 - o l'attributo Vigile della relazione INFRAZIONI e la relazione VIGILI
 - o gli attributi Prov e Numero di INFRAZIONI e la relazione AUTO

Violazione del vincolo di integrità referenziale

Infrazioni

<u>Codice</u>	Data	Vigile	Prov	Numero
34321	1/2/95	3987	MI	39548K
53524	4/3/95	3295	TO	E39548
64521	5/4/96	3295	PR	839548
73321	5/2/98	9345	PR	839548

Δ	u	to
	u	LU

Prov	<u>Numero</u>	Cognome	Nome
MI	E39548	Rossi	Mario
TO	F34268	Rossi	Mario
PR	839548	Neri	Luca

Vincoli di integrità referenziale: commenti



- Giocano un ruolo fondamentale nel concetto di "modello basato su valori."
- Sono possibili meccanismi per il supporto alla loro gestione ("azioni" compensative a seguito di violazioni)



Attenzione ai vincoli su più attributi



Azioni compensative



• Esempio:

 Viene eliminata una ennupla dalla tabella riferita causando così una violazione

Azioni

- o Rifiuto dell'operazione
- o Eliminazione in cascata
- o Introduzione di valori nulli

Eliminazione in cascata

Impiegati

Matricola	Progetto	
34321	Rossi	IDEA
53524	Neri	XYZ
64521	Verdi	NULL
73032	Bianchi	IDEA

Progetti

Codice	Inizio	Durata	Costo
IDEA	01/2000	36	200
XYZ	07/2001	24	120
BOH	09/2001	24	150

Introduzione di valori nulli

Impiegati

Matricola	Cognome	Progetto
34321	Rossi	IDEA
53524	Neri	NULL
64521	Verdi	NULL
73032	Bianchi	IDEA

Progetti

<u>Codice</u>	Inizio	Durata	Costo
IDEA	01/2000	36	200
XYZ	07/2001	24	120
ВОН	09/2001	24	150



Vincoli multipli su più attributi



Incidenti

Codice	Data	ProvA	NumeroA	ProvB	NumeroB
34321	1/2/95	TO	E39548	MI	39548K
64521	5/4/96	PR	839548	TO	E39548

Auto

<u>Prov</u>	<u>Numero</u>	Cognome	Nome
MI	39548K	Rossi	Mario
TO	E39548	Rossi	Mario
PR	839548	Neri	Luca

Vincoli multipli su più attributi



- vincoli di integrità referenziale fra:
 - o gli attributi ProvA e NumeroA di INCIDENTI e la relazione AUTO
 - o gli attributi ProvB e NumeroB di INCIDENTI e la relazione
- L'ordine degli attributi è significativo



Rappresentare per mezzo di una o più relazioni le informazioni per la gestione delle Prenotazioni Mediche dei Pazienti di uno Studio Medico Associato



Definire uno schema di basi di dati per organizzare le informazioni di una azienda che ha impiegati (ognuno con un codice fiscale, cognome, nome e data di nascita), filiali (con codice, sede e direttore (che è impiegato)). Ogni impiegato lavora presso una filiale. Indicare le chiavi e i vincoli di integrità referenziale dello schema. Mostrare un'instanza della base di dati e verificare che soddisfa i vincoli



Si considerino le informazioni per la gestione dei prestiti di una biblioteca personale. Il proprietario presta i libri ai propri amici, che indica con i loro nomi o soprannomi, e i cui numeri di telefono sono contenuti in una rubrica. I libri sono individuati attraverso i titoli. Quando si presta un libro si prende nota della data presunta di restituzione. Definire uno schema per rappresentare queste informazioni indicandone i vincoli.



Rappresentare con una o più relazioni le informazioni contenute nell'orario dei treni in partenza di una stazione ferroviaria: numero, orario, destinazione finale, categoria, fermate intermedie, di tutti i treni in partenza



Definire uno schema di base di dati che organizzi i dati necessari a generare la pagina dei programmi radiofonici di un quotidiano, con stazioni, ora e titoli dei programmi. Per ogni stazione sono memorizzati, oltre al nome anche la frequenza di trasmissione e la sede.