

Database relazionali: un'introduzione

- **Database**: collezione di fatti, registrabili e con un ben preciso significato, relazionati fra di loro
 - Rappresentazione astratta di aspetti del mondo reale (*Universe of Discourse, Miniworld*)
 - Cambiamenti avvenuti nella realtà implicano cambiamenti nel *miniworld*
 - Collezione di dati logicamente coerenti, con un significato implicito
 - Progettato, realizzato, e popolato con un ben preciso intento
 - È utilizzato da un gruppo di persone definito (utenti)
 - Assolve determinati compiti di interesse degli utenti

Database = modello della realtà

Un *database* trae dati da un'origine,
interagisce con eventi del mondo reale e
ha un pubblico attivamente interessato ai contenuti

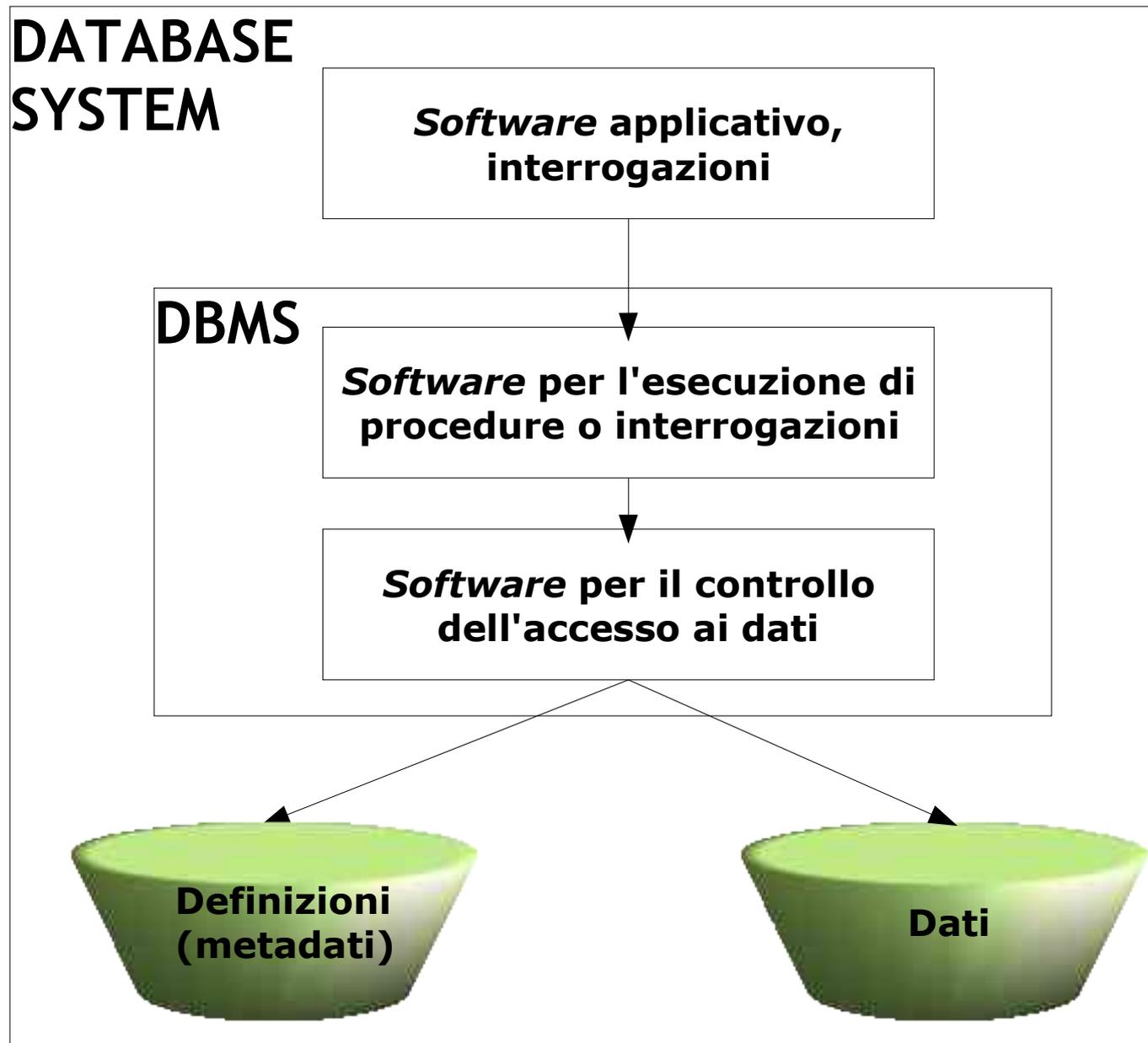
I contenuti di un *database* possono essere
inseriti
cancellati
modificati
oltre che consultati...

Altre definizioni...

- ***Database Management System (DBMS)***
 - Collezione di programmi che consente di creare, gestire ed utilizzare un *database*
- Definizione di un *database*
 - Specificazione tramite DBMS dei tipi di dato, delle strutture dati e dei vincoli
- Costruzione di un *database*
 - Immagazzinamento dei dati su un supporto, controllato dal DBMS
- Manipolazione di un *database*
 - Interrogazioni del DBMS allo scopo di ritrovare dati, di inserirne, o di modificarne di esistenti in risposta al verificarsi di eventi esterni

Database System: dati + DBMS

In pratica...



Definire un *database*

STUDENTE	Nome	Cognome	Matricola	AnnoDiCorso	CorsoDiLaurea
	Carmelo	Russo	475483	1	F54
	Giuseppe	Rossi	483729	2	F54

CORSO	Nome	Codice	Crediti
	Papirologia assira	F540321	5
	Avunculo GRATULAZIONE	F520352	6

Si tratta di specificare:

- Quali entuple compongono il database
 - Quali sono gli oggetti da rappresentare?
- Qual è la struttura di ciascuna entupla
 - Qual è il tipo di dati di ciascun elemento di una data entupla?
 - Qual è l'insieme dei valori possibili per un elemento di una entupla?

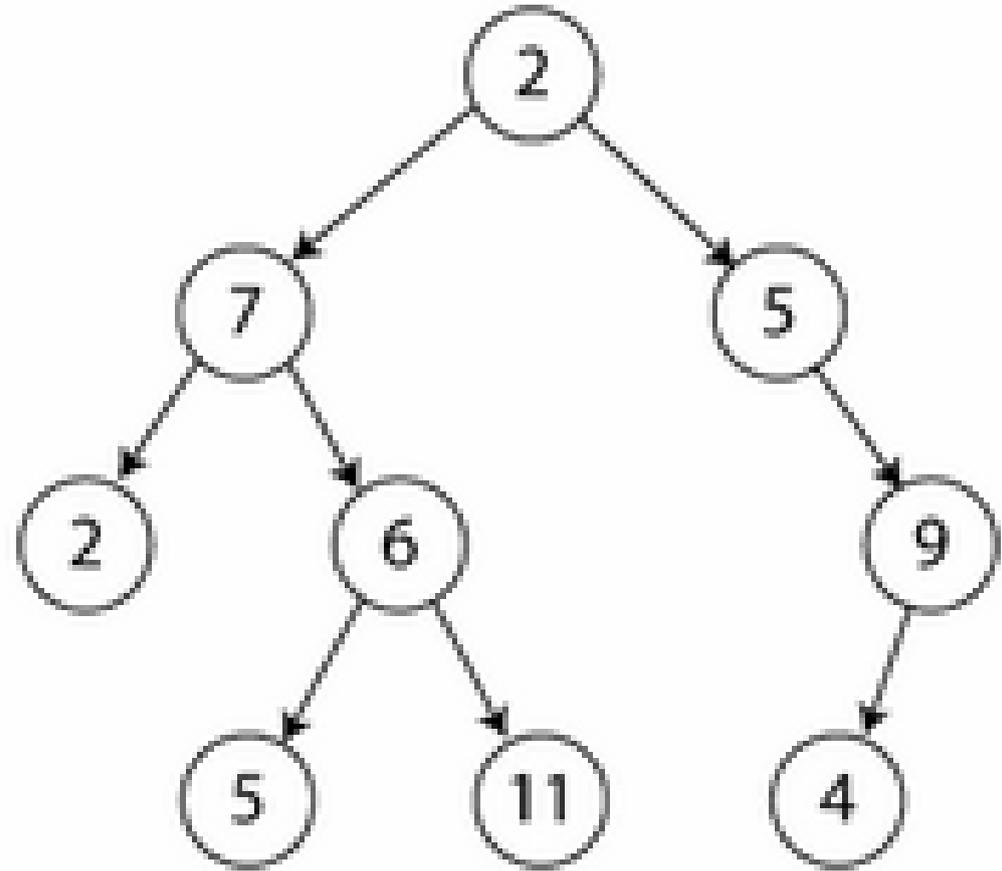
Il modello logico

- Insieme di costrutti utile all'organizzazione dei dati ed alla loro descrizione
 - meccanismi di strutturazione dei dati
 - costruttori di tipo
- Schema
 - rappresentazione invariante nel tempo che descrive la struttura dei dati
- Istanza (di un dato schema)
 - rappresentazione dei valori a un dato tempo t , può cambiare molto rapidamente

Modelli concettuali: totalmente indipendenti dal sistema fisico (un modello può essere sia logico che concettuale)

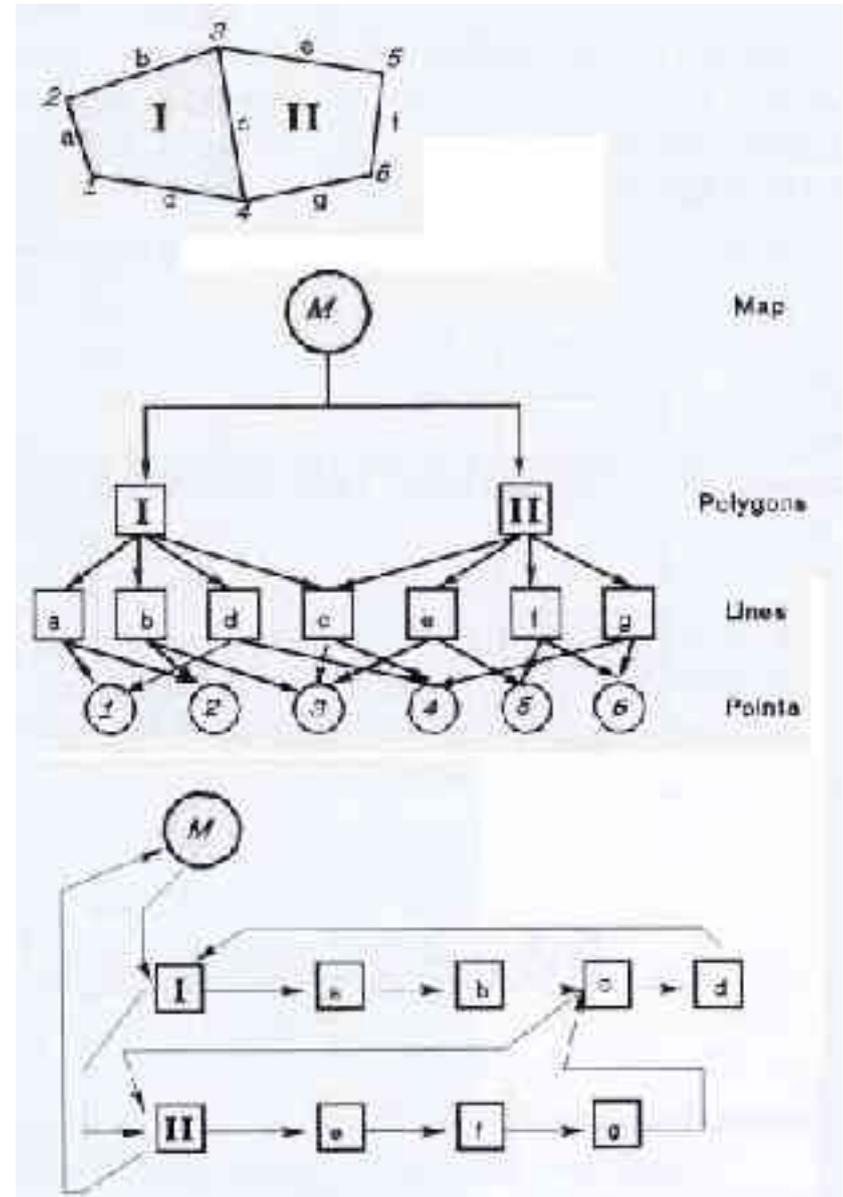
Modelli gerarchici

- le ennuple sono collegate fra loro in modo gerarchico, secondo una struttura ad albero, tale per cui una ennupla ha uno e un solo "genitore". Tipico dei primi DBMS basato su *mainframe*.
- Poco efficiente per descrivere situazioni reali (es. relazioni uno-a-molti).



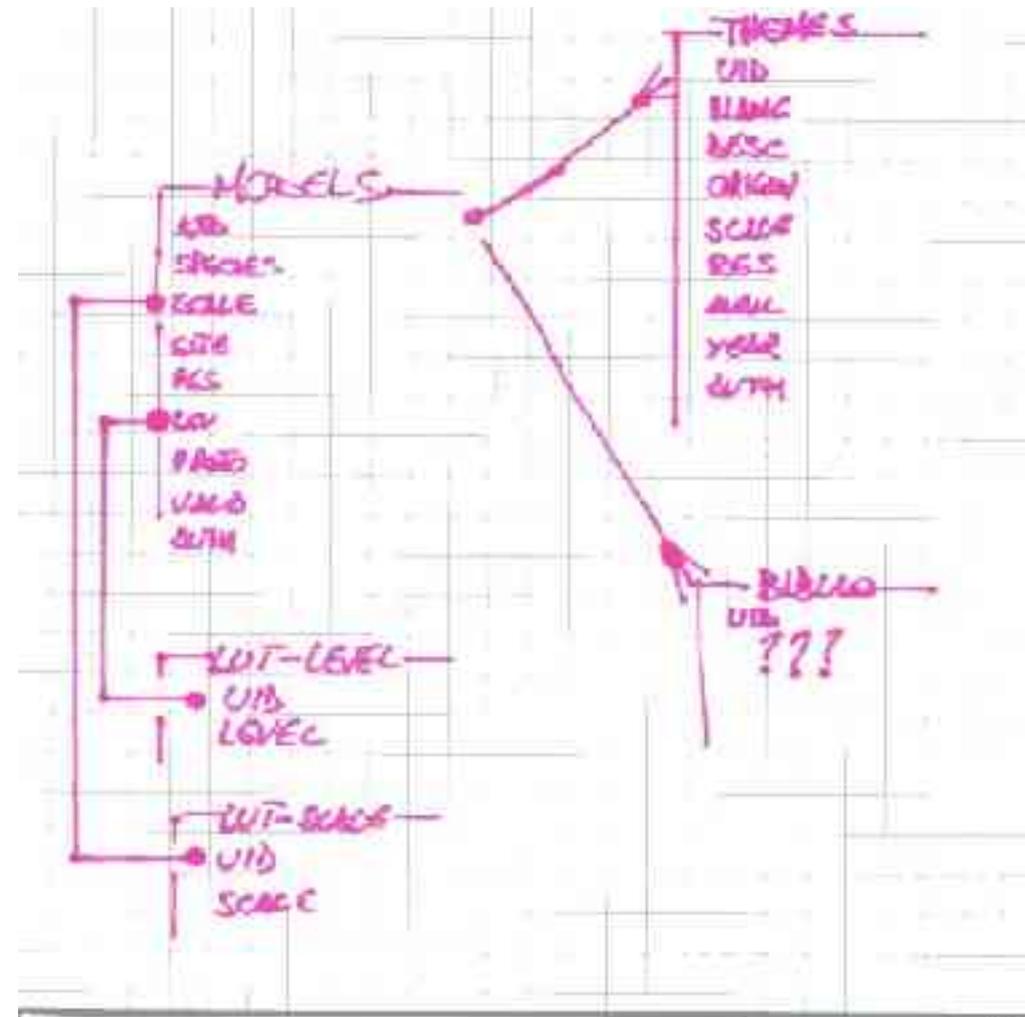
Modelli reticolari (*network*)

- Più versatili rispetto ai modelli gerarchici
- Forte enfasi sulle relazioni tra oggetti
- Tendenza a immagazzinare più relazioni che informazioni
- Superato, ma "riscoperto" per gestire sistemi complessi a oggetti



Modello relazionale

- Si rappresentano solo dei valori
 - i riferimenti tra relazioni sono anch'essi valori
- Elevato livello di astrazione
- Elevato grado di indipendenza dal livello fisico



Codd E.F. (1970) *A relational model of data for large shared data banks*. Communications of the ACM 13(6): 377 – 387

Prima implementazione: 1981

- D_1, D_2, \dots, D_n sono n insiemi (non necessariamente distinti)
- $D_1 \times D_2 \times \dots \times D_n$ prodotto cartesiano, definisce $\{d_1, d_2, \dots, d_n\} / d_1 \in D_1, d_2 \in D_2, \dots, d_n \in D_n$
- non è definito alcun ordinamento tra le ennuple
- le ennuple di una relazione sono distinte fra loro
- a ciascun *dominio* D è associato un nome (*attributo*), unico

Modello relazionale

- non è definito alcun ordinamento degli attributi
- per ogni attributo X e ogni dominio D
 $\exists dom: X \in D$
- Notazione:
 - se t è una ennupla di X e $A \in X$
 $\Rightarrow t[A]$ oppure $t.A$ indica il valore di t per l'attributo A
 - estensibile su più attributi:
 $t[nome, cognome] = \{\text{"Carmelo"}, \text{"Russo"}\}$

Rappresentazione tabulare

Una tabella è una rappresentazione di una relazione se

- i valori in una colonna appartengono allo stesso dominio
- le righe sono tutte diverse fra loro
- i titoli delle colonne sono tutti diversi fra loro
- l'ordine delle righe non è importante
- l'ordine delle colonne non è importante
- i riferimenti tra dati appartenenti a relazioni diverse sono rappresentati tramite valori di domini che appaiono nelle ennuple

Riferimenti tra relazioni

studenti

Matricola	Cognome	Nome	Data di Nascita
6554	Rossi	Mario	5/12/1978
8765	Neri	Paolo	3/11/1976
9283	Verdi	Luisa	12/11/1979
3456	Rossi	Maria	1/2/1978

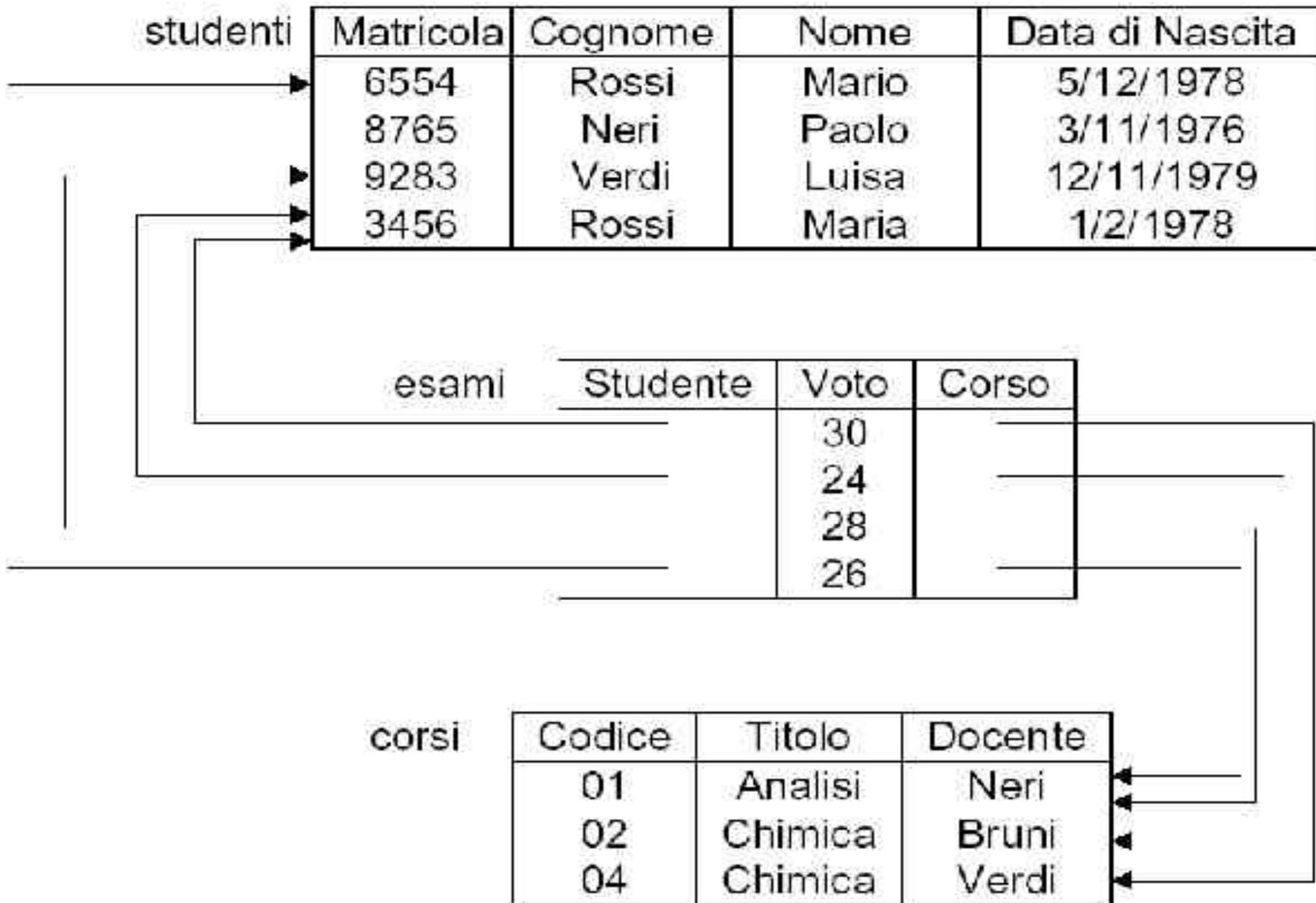
esami

Studente	Voto	Corso
3456	30	04
3456	24	02
9283	28	01
6554	26	01

corsi

Codice	Titolo	Docente
01	Analisi	Neri
02	Chimica	Bruni
04	Chimica	Verdi

Riferimenti tra relazioni



Il modello relazionale impone una strutturazione rigida dei dati

- è possibile utilizzare solo e soltanto ennuple
- non tutti i possibili formati di ennupla sono ammissibili
- sono ammissibili solo i tipi di ennupla che corrispondono a quanto è definito dagli schemi di relazione
- i dati disponibili non possono corrispondere a quanto stabilito negli schemi di relazione

Conseguenza: possono esistere delle ennuple "incomplete"

- pseudocodifiche: "0", "-9999"...
- NULL: $\forall t; \mathbf{A} \exists t[\mathbf{A}] / t[\mathbf{A}] \in \text{dom}(\mathbf{A}) \mid t[\mathbf{A}] \text{ è NULLO}$

Uso del valore nullo

- Valore sconosciuto
 - esiste un valore, ma non è noto al momento
- Valore inesistente
 - non esiste un valore entro il dominio
- Assenza di informazione
 - non è noto se esiste un valore entro il dominio

Normalmente, un DBMS non opera la distinzione sopra esposta...

Studenti	Matricola	Cognome	Nome	Nascita
	276545	Rossi	Maria	23/04/1968
	276545	Neri	Anna	23/04/1972
	788854	Verdi	Fabio	12/02/1972

Esami	Studente	Voto	Lode	Corso
	276545	28	e lode	01
	276545	32		02
	788854	23		03
	200768	30	e lode	03

Corsi	Codice	Titolo	Docente
	01	Analisi	Giani
	03	NULL	NULL
	02	Chimica	Belli

Possono esistere determinate istanze che sono sintatticamente corrette (rispettano lo schema) ma non rappresentano informazioni "possibili"

Il problema viene risolto specificando dei vincoli di integrità dei dati

- Qualità del dato
- Ausilio alla progettazione

Vincoli di integrità

- Proprietà dei dati che deve essere soddisfatta da un'istanza, perché questa rappresenti una informazione corretta
- È una funzione booleana
- Vincoli su valori (vincoli di dominio)
 - Sesso: {'M';'F';'I'}
 - (Voto \geq 18) AND (Voto \leq 30)
- Vincoli di ennupla
 - t[sesso,gravido] : t[gravido]=NULL se t[sesso]='M'
 - (Voto=30) OR NOT (Lode=TRUE)
- Vincoli relazionali
 - definiscono le “regole di validità” nei riferimenti tra relazioni

Vincoli di chiave

Garantiscono l'unicità di una ennupla

- Chiave

- insieme di attributi che identifica in modo univoco una ennupla
- un insieme $K = \{d_1, d_2, \dots, d_n\}$ è *superchiave* per r se
$$\forall t_1, t_2: t_1[K] \neq t_2[K]$$
- K è detto *chiave* se è minimale, ovvero non contiene una superchiave o è costituito da un solo attributo

Matricola	Cognome	Nome	Nascita	Corso
6554	Rossi	Mario	5/12/1978	Informatica
8766	Rossi	Mario	3/11/1976	Informatica
4723	Verdi	Laura	10/7/1979	Meccanica
9283	Verdi	Mario	3/11/1976	Informatica
3456	Rossi	Laura	5/12/1978	Meccanica

Perché vincolare?

- i vincoli sono indispensabili per rappresentare proprietà del mondo reale
- i vincoli operano a livello di schema
 - interessano qualunque istanza di un dato schema
 - sono corrette solo e soltanto le istanze che soddisfano tutti i vincoli

In particolare:

- una relazione è un insieme di ennuple uguali tra loro
- esiste sempre una superchiave coincidente con l'insieme degli attributi sui quali la relazione è definita
- l'insieme di attributi è finito, quindi dato uno schema di relazione esiste sempre almeno una chiave

Importanza delle chiavi

L'esistenza delle chiavi garantisce unicità e accessibilità:

- ogni valore è accessibile specificando
 - il nome della relazione
 - il valore della chiave
 - il nome dell'attributo del quale si vuole conoscere il valore
- le chiavi sono essenziali per correlare ennuple appartenenti a relazioni diverse
- le chiavi non dovrebbero ammettere valore uguale a NULL: *chiave primaria*
- *Integrità referenziale*: A è l'insieme di attributi di una relazione R_1 , che fa riferimento a R_2 . Ogni ennupla di R_1 deve contenere in A valori (chiavi esterne) ricavati dalla chiave primaria K di R_2 .