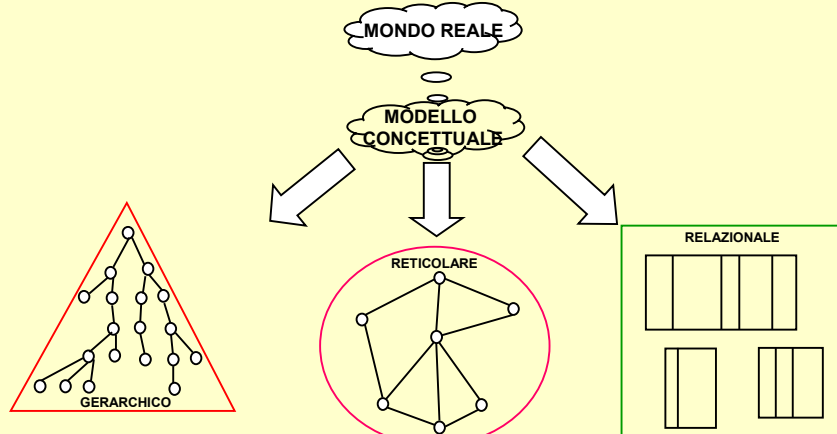


Atzeni, Ceri, Paraboschi, Torlone
Basi di dati
McGraw-Hill, 1999

**Capitolo 2: Il modello relazionale:
strutture e vincoli**

MODELLI DEI DATI

I MODELLI COSTITUISCONO UNA STRUTTURAZIONE SEMPLIFICATA DELLA REALTA' CHE NE ACCOGLIE ASPETTI SPECIFICI E AIUTA A COMPRENDERLA MEGLIO



MODELLI LOGICI

Atzeni-Ceri-Paraboschi-Torlone,
Basi di dati, Capitolo 2

MODELLI LOGICI DEI DATI

- SUPPORTANO UNA DESCRIZIONE DEI DATI CHE PUO' ESSERE **ELABORATA DAL SISTEMA** (DBMS)
- VENGONO **MAPPATI FACILMENTE** SULLE STRUTTURE FISICHE DI MEMORIZZAZIONE

Atzeni-Ceri-Paraboschi-Torlone,
Basi di dati, Capitolo 2

3

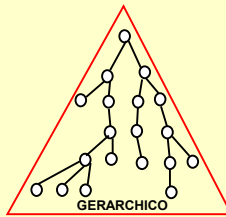
Cronologia dei modelli per la rappresentazione dei dati

- **Modello gerarchico (anni 60)**
- **Modello reticolare (anni 70)**
- **Modello relazionale (anni 80)**
- **Modello a oggetti (anni 90)**
- **Modello XML (anni 00)**

MODELLI LOGICI DEI DATI

GERARCHICO

- I DATI SONO RAPPRESENTATI COME **RECORD**
- LE ASSOCIAZIONI TRA I DATI SONO RAPPRESENTATE CON **PUNTATORI** IN UNA STRUTTURA AD **ALBERO**



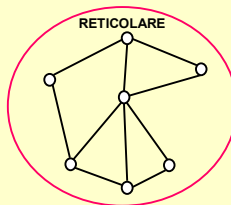
Atzeni-Ceri-Paraboschi-Torlone,
Basi di dati, Capitolo 2

5

MODELLI LOGICI DEI DATI

RETICOLARE (CODASYL)

- I DATI SONO RAPPRESENTATI COME **RECORD**
- LE ASSOCIAZIONI TRA I DATI SONO RAPPRESENTATE CON **PUNTATORI** IN UNA STRUTTURA A **GRAFO** COMPLESSO



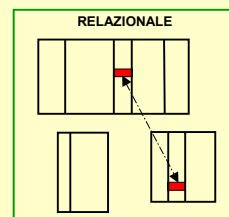
Atzeni-Ceri-Paraboschi-Torlone,
Basi di dati, Capitolo 2

6

MODELLI LOGICI DEI DATI

RELAZIONALE

- I DATI SONO RAPPRESENTATI COME SEQUENZE DI VALORI DI ATTRIBUTI
- DATI CARATTERIZZATI DALLE STESSSE SEQUENZE DI ATTRIBUTI SONO RAGGRUPPATI IN TABELLE
- LE ASSOCIAZIONI TRA I DATI SONO OTTENUTE ASSOCIANDO VALORI DI ATTRIBUTI IN TABELLE DIVERSE



Atzeni-Ceri-Paraboschi-Torlone,
Basi di dati, Capitolo 2

7

Cronologia del modello relazionale

- **Inventato da T. Codd, 1970**
(IBM Research di Santa Teresa, Cal)
- **Primi progetti:**
SYSTEM R (IBM), Ingres (Berkeley Un.)
- **Principali scoperte tecnologiche: 1978-1980**
- **Primi sistemi commerciali:**
inizio anni '80 (Oracle, IBM-SQL DS e DB2, Ingres, Informix, Sybase)
- **Successo commerciale: dal 1985.**

Definizione informale

studente

colonna

schema

| MATR | NOME | CITTA' | C-DIP |
|------|----------|---------|-------|
| 123 | Carlo | Bologna | Inf |
| 307 | Giovanni | Milano | Log |
| 415 | Paola | Torino | Inf |
| 702 | Antonio | Roma | Log |

istanza

riga

Relazione: tre accezioni

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

Relazione matematica

- D_1, D_2, \dots, D_n (n insiemi anche non distinti)
- il prodotto cartesiano $D_1 \times D_2 \times \dots \times D_n$, è l'insieme di tutte le n -uple ordinate (d_1, d_2, \dots, d_n) tali che $d_1 \in D_1, d_2 \in D_2, \dots, d_n \in D_n$.
- una relazione matematica su D_1, D_2, \dots, D_n è un sottoinsieme del prodotto cartesiano $D_1 \times D_2 \times \dots \times D_n$.
- D_1, D_2, \dots, D_n sono i domini della relazione. Una relazione su n domini ha grado n .
- il numero di n -uple è la cardinalità della relazione. Nelle applicazioni reali, la cardinalità è sempre finita.

Relazione matematica, esempio

- $D_1 = \{a, b\}$
- $D_2 = \{x, y, z\}$
- prodotto cartesiano $D_1 \times D_2$

| | |
|---|---|
| a | x |
| a | y |
| a | z |
| b | x |
| b | y |
| b | z |

- una relazione $r \subseteq D_1 \times D_2$

| | |
|---|---|
| a | x |
| a | z |
| b | y |
| b | z |

Relazione matematica, proprietà

- In base alle definizioni, una relazione matematica è un **insieme** di n -uple **ordinate**:
(d_1, d_2, \dots, d_n) tali che $d_1 \in D_1, d_2 \in D_2, \dots, d_n \in D_n$
- una relazione è un **insieme**; quindi:
 - non è definito alcun ordinamento fra le n -uple;
 - le n -uple di una relazione sono distinte l'una dall'altra;
- le n -uple sono **ordinate**: l' i -esimo valore di ciascuna proviene dall' i -esimo dominio; è cioè definito un ordinamento fra i domini.

Relazione matematica, esempio

$Partite \subseteq string \times string \times integer \times integer$

| | | | |
|-------|-------|---|---|
| Juve | Lazio | 3 | 1 |
| Lazio | Milan | 2 | 0 |
| Juve | Roma | 1 | 2 |
| Roma | Milan | 0 | 1 |

- Ciascuno dei domini ha due **ruoli** distinti, distinguibili attraverso la posizione: il primo e il terzo dominio si riferiscono a nome e reti della squadra ospitante; il secondo e il quarto a nome e reti della squadra ospitata.
- La struttura è **posizionale**

Relazioni nel modello relazionale dei dati

- A ciascun dominio associamo un nome (**attributo**), unico nella relazione, che “descrive” il ruolo del dominio.
- Nella rappresentazione tabellare, gli attributi possono essere usati come intestazioni delle colonne.

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

- L'ordinamento fra gli attributi è irrilevante:
la struttura è **non posizionale**

Formalizzando

- L'associazione fra domini e attributi è definita da una funzione *dom*: $X \in D$ che associa a ciascun attributo un dominio.
- Una **ennupla** su un insieme di attributi X è una funzione che associa a ciascun attributo A in X un valore del dominio $\text{dom}(A)$
- Una **relazione** su X è un insieme di ennuple su X

Notazioni

- Se t è una ennupla su X e $A \in X$, allora $t[A]$ (o $t.A$) indica il valore di t su A .
- Nell'esempio, se t è la prima ennupla della tabella
 $t[\text{Fuori}] = \text{Lazio}$
- La stessa notazione è estesa anche ad insiemi di attributi, nel qual caso denota ennuple: $t[\text{Fuori}, \text{RetiF}]$ è una ennupla su due attributi.

Tabelle e relazioni

- Una tabella rappresenta una relazione se
 - i valori di ciascuna colonna sono fra loro omogenei (dallo stesso dominio)
 - le righe sono diverse fra loro
 - le intestazioni delle colonne sono diverse tra loro
- Inoltre, in una tabella che rappresenta una relazione
 - l'ordinamento tra le righe è irrilevante
 - l'ordinamento tra le colonne è irrilevante

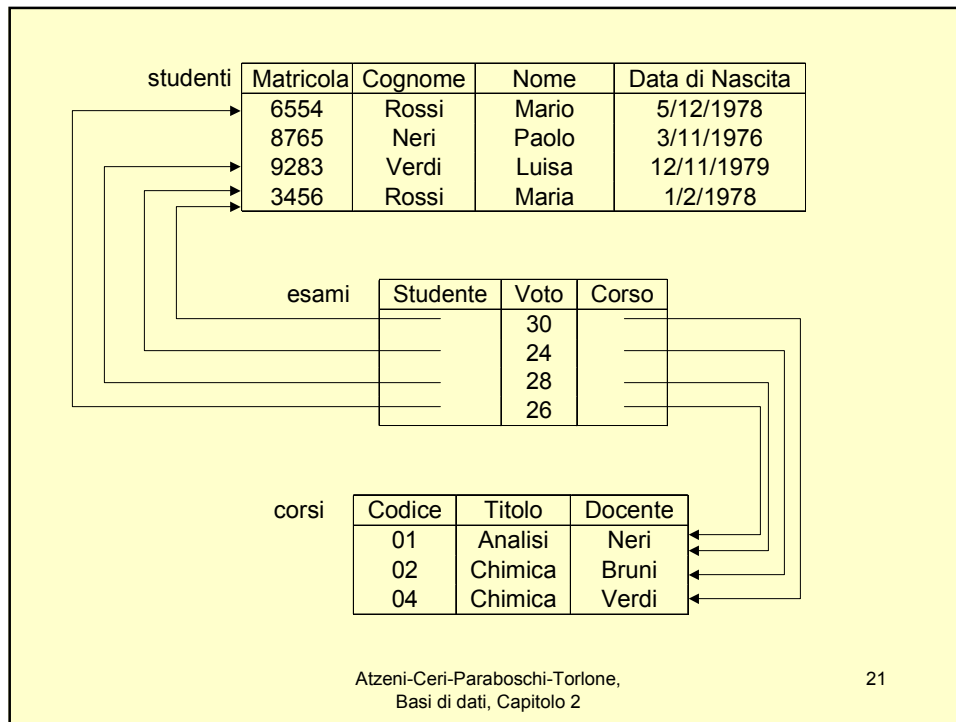
Il modello relazionale è basato su valori

- 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 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 |



Vantaggi della struttura basata su valori

- indipendenza dalle strutture fisiche (si potrebbe avere anche con puntatori di alto livello) che possono cambiare anche dinamicamente
- si rappresenta solo ciò che è rilevante dal punto di vista dell'applicazione (dell'utente); i puntatori sono meno comprensibili per l'utente finale (senza, l'utente finale vede gli stessi dati dei programmatori)
- i dati sono portabili più facilmente da un sistema ad un altro
- i puntatori sono direzionali

Note:

- i puntatori possono esistere a livello fisico
- nel modello a oggetti esistono "una specie di puntatori", ad alto livello

Definizioni

Schema di relazione:

un **nome di relazione** R con un insieme di **attributi** A_1, \dots, A_n
 $R(A_1, \dots, A_n)$

Schema di base di dati:

insieme di schemi di relazione con nomi diversi:

$$\mathbf{R} = \{R_1(X_1), \dots, R_n(X_n)\}$$

(Istanza di) relazione su uno schema $R(X)$:

insieme r di enuple su X

(Istanza di) base di dati su uno schema $\mathbf{R} = \{R_1(X_1), \dots, R_n(X_n)\}$:

insieme di relazioni $\mathbf{r} = \{r_1, \dots, r_n\}$ (con r_i relazione su R_i)

Notazione

attributi: lettere iniziali dell'alfabeto, maiuscole:

A, B, C, A', A_1, \dots

insiemi di attributi: lettere finali dell'alfabeto, maiuscole:

X, Y, Z, X', X_1, \dots

giustapposizione dei nomi degli attributi: $X=ABC$
(anziché $X=\{A,B,C\}$)

unioni di insiemi: XY anziché $X \cup Y$

nomi di relazione: R e lettere circostanti, maiuscole, anche con
indici e pedici: R_1, S, S', \dots

relazione: come il nome della relazione, ma in minuscolo

schema di base di dati: lettera maiuscola in grassetto $\mathbf{R}, \mathbf{S}, \dots$

base di dati: stesso simbolo dello schema, ma in minuscolo

Esempio

- sono possibili relazioni su un solo attributo

| 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 |

| Studenti lavoratori | Matricola |
|---------------------|-----------|
| | 6554 |
| | 8765 |

Atzeni-Ceri-Paraboschi-Torlone,
Basi di dati, Capitolo 2

25

Strutture nidificate

| "Da Filippo" Via Roma 23, Chissadove | | |
|--|-----------|-------|
| Ricevuta 2369 del 12/05/1997 | | |
| 3 | Coperti | 6000 |
| 2 | Antipasti | 12000 |
| 3 | Primi | 27000 |
| 2 | Bistecche | 36000 |
| Totale | | 81000 |

| "Da Filippo" Via Roma 23, Chissadove | | |
|--|-----------|-------|
| Ricevuta 2456 del 16/05/1997 | | |
| 2 | Coperti | 4000 |
| 1 | Antipasti | 6000 |
| 2 | Primi | 15000 |
| 2 | Orate | 50000 |
| 2 | Caffè | 3000 |
| Totale | | 78000 |

Atzeni-Ceri-Paraboschi-Torlone,
Basi di dati, Capitolo 2

26

Rappresentazione di strutture nidificate per mezzo di relazioni

ricevute

| Numero | Data | Totale |
|--------|------------|--------|
| 2369 | 12/05/1997 | 81000 |
| 2456 | 16/05/1997 | 78000 |

dettaglio

| Numero | Quantità | Descrizione | Importo |
|--------|----------|-------------|---------|
| 2369 | 3 | Coperti | 6000 |
| 2369 | 2 | Antipasti | 12000 |
| 2369 | 3 | Primi | 27000 |
| 2369 | 2 | Bistecche | 36000 |
| 2456 | 2 | Coperti | 4000 |
| 2456 | 1 | Antipasti | 6000 |
| 2456 | 2 | Primi | 15000 |
| 2456 | 2 | Orate | 50000 |
| 2456 | 2 | Caffè | 3000 |

Atzeni-Ceri-Paraboschi-Torlone,
Basi di dati, Capitolo 2

27

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

Atzeni-Ceri-Paraboschi-Torlone,
Basi di dati, Capitolo 2

28

Rappresentazione alternativa

ricevute

| Numero | Data | Totale |
|--------|------------|--------|
| 2369 | 12/05/1997 | 81000 |
| 2456 | 16/05/1997 | 78000 |

dettaglio

| Numero | Riga | Quantità | Descrizione | Importo |
|--------|------|----------|-------------|---------|
| 2369 | 1 | 3 | Coperti | 6000 |
| 2369 | 2 | 2 | Antipasti | 12000 |
| 2369 | 3 | 3 | Primi | 27000 |
| 2369 | 4 | 2 | Bistecche | 36000 |
| 2456 | 1 | 2 | Coperti | 4000 |
| 2456 | 2 | 1 | Antipasti | 6000 |
| 2456 | 3 | 2 | Primi | 15000 |
| 2456 | 4 | 2 | Orate | 50000 |
| 2456 | 5 | 2 | Caffè | 3000 |

Informazione incompleta

- 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
- I dati disponibili possono non corrispondere esattamente al formato previsto, per varie ragioni.

Informazione incompleta: motivazioni

- Firenze è provincia, ma non conosciamo l'indirizzo della prefettura
- Tivoli non è provincia: non ha prefettura
- Prato è "nuova" provincia: ha la prefettura?

| Città | Prefettura |
|---------|-----------------|
| Roma | Via IV novembre |
| Firenze | |
| Tivoli | |
| Prato | |

Informazione incompleta: soluzioni?

- non conviene (anche se spesso si fa) utilizzare valori ordinari del dominio (0, stringa nulla, "99", etc), per vari motivi:
 - potrebbero non esistere valori "non utilizzati"
 - valori "non utilizzati" potrebbero diventare significativi
 - in fase di utilizzo (ad esempio, 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:
 - **valore nullo**: denota l'assenza di un valore del dominio (e non è un valore del dominio)
- Formalmente, è sufficiente estendere il concetto di enupla: $t[A]$, 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

Troppi valori nulli!

| Studenti | Matricola | Cognome | Nome | Nascita |
|----------|-----------|---------|-------|------------|
| | 276545 | Rossi | Maria | NULL |
| | NULL | Neri | Anna | 23/04/1972 |
| | NULL | Verdi | Fabio | 12/02/1972 |

| Esami | Studente | Voto | Corso |
|-------|----------|------|-------|
| | 276545 | 28 | 01 |
| | NULL | 27 | NULL |
| | 200768 | 24 | NULL |

| Corsi | Codice | Titolo | Docente |
|-------|--------|---------|---------|
| | 01 | Analisi | Giani |
| | 03 | NULL | NULL |
| | NULL | Chimica | Belli |

Tipi di valore nullo

- (almeno) tre casi differenti
 - **valore sconosciuto**: esiste un valore del dominio, ma non è noto (Firenze)
 - **valore inesistente**: non esiste un valore del dominio (Tivoli)
 - **valore senza informazione**: non è noto se esista o meno un valore del dominio (Prato)
- I DBMS non distinguono i tipi di valore nullo (e quindi implicitamente adottano il valore **senza informazione**)

Vincoli di integrità

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

| 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 |

Vincolo di integrità

- Definizione
 - proprietà che deve essere soddisfatta dalle istanze che rappresentano informazioni corrette per l'applicazione ogni vincolo può essere visto come una funzione booleana (o un predicato) che associa ad ogni istanza il valore **vero** o **falso**.
- Tipi di vincoli:
 - vincoli intrarelazionali; casi particolari:
 - vincoli su valori (o di dominio)
 - vincoli di ennumera
 - vincoli interrelazionali

Vincoli di integrità, motivazioni

- risultano utili al fine di descrivere la realtà di interesse in modo più accurato di quanto le strutture permettano;
- forniscono un contributo verso la "qualità dei dati"
- costituiscono uno strumento di ausilio alla progettazione (vedremo la "normalizzazione")
- sono utilizzati dal sistema nella scelta della strategia di esecuzione delle interrogazioni

Nota:

- non tutte le proprietà di interesse sono rappresentabili per mezzo di vincoli esprimibili direttamente

Vincoli di ennupla

- Esprimono condizioni sui valori di ciascuna ennupla, indipendentemente dalle altre ennuple.
- Una possibile sintassi: espressione booleana (con AND, OR e NOT) di atomi che confrontano valori di attributo o espressioni aritmetiche su di essi.
- Un vincolo di ennupla è un **vincolo di dominio** se coinvolge un solo attributo
- Esempi:

(Voto ≥ 18) AND (Voto ≤ 30)
(Voto =30) OR NOT (Lode = "e lode")

Lordo = (Ritenute + Netto)

Atzeni-Ceri-Paraboschi-Torlone,
Basi di dati, Capitolo 2

39

Identificazione delle ennuple

| Matricola | Cognome | Nome | Corso | Nascita |
|-----------|---------|-------|-------------|-----------|
| 6554 | Rossi | Mario | Informatica | 5/12/1978 |
| 8765 | Rossi | Mario | Informatica | 3/11/1976 |
| 4723 | Verdi | Laura | Meccanica | 10/7/1979 |
| 9283 | Verdi | Mario | Informatica | 3/11/1976 |
| 3456 | Rossi | Laura | Meccanica | 5/12/1978 |

- il numero di matricola identifica gli studenti:
 - non ci sono due ennuple con lo stesso valore sull'attributo Matricola
- i dati anagrafici identificano gli studenti:
 - non ci sono due ennuple uguali su tutti e tre gli attributi Cognome, Nome e Data di Nascita

Atzeni-Ceri-Paraboschi-Torlone,
Basi di dati, Capitolo 2

40

Vincoli di chiave

- **chiave:**
 - insieme di attributi che identificano univocamente le ennuple di una relazione
- più precisamente:
 - un insieme K di attributi è **superchiave** per una relazione 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 (cioè non contiene un'altra superchiave) per r

| Matricola | Cognome | Nome | Corso | Nascita |
|-----------|---------|-------|-------------|-----------|
| 6554 | Rossi | Mario | Informatica | 5/12/1978 |
| 8765 | Rossi | Mario | Informatica | 3/11/1976 |
| 4723 | Verdi | Laura | Meccanica | 10/7/1979 |
| 9283 | Verdi | Mario | Informatica | 3/11/1976 |
| 3456 | Rossi | Laura | Meccanica | 5/12/1978 |

- Matricola è una chiave:
 - Matricola è superchiave
 - contiene un solo attributo e quindi è minimale
- Cognome, Nome, Nascita è un'altra chiave:
 - l'insieme Cognome, Nome, Nascita è superchiave
 - nessuno dei suoi sottoinsiemi è superchiave

| Matricola | Cognome | Nome | Corso | Nascita |
|-----------|---------|-------|-------------|-----------|
| 6554 | Rossi | Mario | Informatica | 5/12/1978 |
| 8765 | Rossi | Mario | Elettronica | 3/11/1976 |
| 4723 | Verdi | Laura | Meccanica | 10/7/1979 |
| 9283 | Verdi | Mario | Informatica | 3/11/1976 |
| 3456 | Rossi | Laura | Meccanica | 5/12/1978 |

- la relazione non contiene ennuple fra loro uguali su Cognome e Corso:
 - in ogni corso di studio gli studenti hanno cognomi diversi;
 - l'insieme { Cognome, Corso } è superchiave minimale e quindi chiave
- possiamo dire che questa proprietà è sempre soddisfatta?
 - No! In generale ci possono essere in un corso di studio studenti con lo stesso cognome

Chiavi, schemi e istanze

- i vincoli corrispondono a proprietà del mondo reale modellato dalla base di dati;
- quindi interessano a livello di schema (con riferimento cioè a tutte le istanze):
 - ad uno schema associamo un insieme di vincoli e consideriamo **corrette** (lecite, valide, ammissibili) solo le istanze che soddisfano tutti i vincoli;
 - singole istanze possono soddisfare ulteriori vincoli (“per pura coincidenza”)

Individuazione delle chiavi

- definendo uno schema di relazione, associamo ad esso i vincoli di chiave che vogliamo siano soddisfatti dalle sue istanze (corrette)
- li individuiamo
 - considerando le proprietà che i dati soddisfano nell'applicazione (il "frammento di mondo reale di interesse");
 - notando quali insiemi di attributi permettono di identificare univocamente le ennuple;
 - e individuando i sottoinsiemi minimali di tali insiemi che conservano la capacità di identificare le ennuple.

Individuazione delle chiavi, esempio

- Allo schema di relazione
STUDENTI(Matricola, Cognome, Nome, Corso, Nascita)
associamo i vincoli che indicano come chiavi gli insiemi di attributi Matricola e Cognome, Nome, Nascita

- La relazione

| Matricola | Cognome | Nome | Corso | Nascita |
|-----------|---------|-------|-------------|-----------|
| 6554 | Rossi | Mario | Informatica | 5/12/1978 |
| 8765 | Rossi | Mario | Elettronica | 3/11/1976 |
| 4723 | Verdi | Laura | Meccanica | 10/7/1979 |
| 9283 | Verdi | Mario | Informatica | 3/11/1976 |
| 3456 | Rossi | Laura | Meccanica | 5/12/1978 |

è corretta, perché soddisfa i vincoli associati allo schema.

- Ne soddisfa anche altri. Ad esempio, Cognome, Corso è chiave per essa.

Esistenza delle chiavi

- poiché le relazioni sono insiemi, ogni relazione non può contenere ennuple distinte ma uguali fra loro:
 - ogni relazione ha come superchiave l'insieme degli attributi su cui è definita;
- poiché l'insieme di tutti gli attributi è una superchiave per ogni relazione, ogni schema di relazione ha tale insieme come superchiave;
- poiché l'insieme di attributi è finito, ogni schema di relazione ha (almeno) una chiave

Importanza delle chiavi

- l'esistenza delle chiavi garantisce l'accessibilità a ciascun dato della base di dati
- ogni singolo valore è univocamente accessibile tramite:
 - nome della relazione
 - valore della chiave
 - nome dell'attributo
- le chiavi sono lo strumento principale attraverso il quale vengono correlati i dati in relazioni diverse ("il modello relazionale è basato su valori")

Chiavi e valori nulli

- In presenza di valori nulli, i valori degli attributi che formano la chiave
 - non permettono di identificare le ennuple come desiderato
 - né permettono di realizzare facilmente i riferimenti da altre relazioni

| Matricola | Cognome | Nome | Nascita | Corso |
|-----------|---------|-------|----------|-------------|
| NULL | Rossi | Luca | NULL | Informatica |
| 8765 | Rossi | Mario | 01/05/61 | Civile |
| 4856 | Neri | Mario | NULL | NULL |
| NULL | Neri | Mario | 05/03/63 | Civile |

Chiave primaria

- La presenza di valori nulli nelle chiavi deve essere limitata
- Soluzione pratica: per ogni relazione scegliamo una chiave (la chiave primaria) su cui non ammettiamo valori nulli.
- Notazione per la chiave primaria: gli attributi che la compongono sono sottolineati

| <u>Matricola</u> | Cognome | Nome | Nascita | Corso |
|------------------|---------|-------|----------|-------------|
| 6554 | Rossi | Luca | NULL | Informatica |
| 8765 | Rossi | Mario | 01/05/61 | Civile |
| 4856 | Neri | Mario | NULL | NULL |
| 6590 | Neri | Mario | 05/03/63 | Civile |

Vincoli di integrità referenziale ("foreign key")

- informazioni in relazioni diverse sono correlate attraverso valori comuni
- in particolare, valori delle chiavi (primarie, di solito)
- un **vincolo di integrità referenziale** fra un insieme di attributi X di una relazione R_1 e un'altra relazione R_2 impone ai valori su X di ciascuna ennupla dell'istanza di R_1 di comparire come valori della chiave (primaria) dell'istanza di R_2

Atzeni-Ceri-Paraboschi-Torlone,
Basi di dati, Capitolo 2

51

Base di dati con vincoli di integrità referenziale

infrazioni

| Codice | Data | Vigile | Prov | Numero |
|--------|-----------|--------|------|--------|
| 65524 | 3/9/1997 | 343 | MI | 3K9886 |
| 87635 | 4/12/1997 | 476 | MI | 6D5563 |
| 82236 | 4/12/1997 | 343 | RM | 7C5567 |
| 35632 | 6/1/1998 | 476 | RM | 7C5567 |
| 76543 | 5/3/1998 | 548 | MI | 6D5563 |

vigili

| Matricola | Cognome | Nome |
|-----------|----------|------|
| 343 | Rossi | Luca |
| 476 | Neri | Pino |
| 548 | Nicolosi | Gino |

automobili

| Prov | Numero | Proprietario | ... |
|------|--------|--------------|-----|
| MI | 3K9886 | Nestore | ... |
| MI | 6D5563 | Nestore | ... |
| RM | 7C5567 | Menconi | ... |
| RM | 1A6673 | Mussone | ... |
| MI | 5E7653 | Marchi | ... |

Atzeni-Ceri-Paraboschi-Torlone,
Basi di dati, Capitolo 2

52

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

Base di dati che viola vincoli di integrità referenziale

infrazioni

| Codice | Data | Vigile | Prov | Numero |
|--------|-----------|--------|------|--------|
| 65524 | 3/9/1997 | 343 | MI | 3K9886 |
| 87635 | 4/12/1997 | 476 | MI | 6D5563 |
| 82236 | 4/12/1997 | 343 | RM | 7C5567 |
| 35632 | 6/1/1998 | 476 | RM | 7C5567 |
| 76543 | 5/3/1998 | 548 | MI | 6D5563 |

vigili

| Matricola | Cognome | Nome |
|-----------|----------|------|
| 343 | Rossi | Luca |
| 548 | Nicolosi | Gino |

automobili

| Prov | Numero | Proprietario | ... |
|------|--------|--------------|-----|
| MI | 3K9886 | Nestore | ... |
| RM | 6D5563 | Nestore | ... |
| MI | 7C5567 | Menconi | ... |
| RM | 1A6673 | Mussone | ... |
| MI | 5E7653 | Marchi | ... |

Vincoli di integrità referenziale: commenti

- I vincoli di integrità referenziale giocano un ruolo fondamentale nel concetto “modello relazionale basato su valori.”
- Sono possibili meccanismi per il supporto alla gestione dei vincoli di integrità referenziale (“azioni” da svolgere in corrispondenza a violazioni).
- In presenza di valori nulli i vincoli possono essere resi meno restrittivi
- Attenzione ai vincoli su più attributi

Base di dati con vincoli di integrità referenziale

incidenti

| Codice | Data | ProvA | NumeroA | ProvB | NumeroB |
|--------|-----------|-------|---------|-------|---------|
| 65524 | 3/9/1997 | MI | 3K9886 | RM | 7C5567 |
| 87635 | 4/12/1997 | RM | 6D5563 | RM | 1A6673 |
| 82236 | 6/12/1997 | MI | 7C5567 | RM | 6D5563 |

automobili

| Prov | Numero | Proprietario | ... |
|------|--------|--------------|-----|
| MI | 3K9886 | Nestore | ... |
| RM | 6D5563 | Nestore | ... |
| MI | 7C5567 | Menconi | ... |
| RM | 1A6673 | Mussone | ... |
| MI | 5E7653 | Marchi | ... |