



Il modello relazionale

Un po' di storia...

| Anni 70

- * Definizione del modello
- * Prima versione del linguaggio SQL
- * Studi fondamentali sulla tecnologia relazionale
- * Primi prototipi di DBMS relazionali (RDBMS)

| Anni 80

- * Prima standardizzazione di SQL
- * Primi prototipi commerciali:
 - » Oracle
 - » IBM DB2

| Anni 90

- * Standard ISO-ANSI SQL-2

| Oggi

- * Esiste già lo standard ISO-ANSI SQL-3...
- * ...e sono allo studio diverse altre estensioni del linguaggio

Sul termine “relazione”

- | Il termine **relazione** può essere usato con diverse accezioni, che non vanno confuse tra loro:
 - * nel linguaggio comune denota un *legame* di qualche tipo
 - * nella teoria degli insiemi denota una *relazione matematica*
 - * nel modello relazionale è una generalizzazione della relazione matematica
 - * nei DBMS relazionali è usato spesso, a volte erroneamente, come sinonimo di *tabella*
- | Per introdurre il modello relazionale è quindi opportuno innanzitutto riesaminare il concetto di relazione matematica

Relazione matematica n-aria

- | Si considerino $n > 0$ insiemi $D_1, D_2, \dots, D_n$, non necessariamente distinti
- | Il prodotto Cartesiano $D_1 \times D_2 \times \dots \times D_n$ è l'insieme di tutte le n-ple 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 qualunque sottoinsieme del prodotto Cartesiano $D_1 \times D_2 \times \dots \times D_n$

Esempio:

$D_1 = \{\text{mela}, \text{pera}\}, D_2 = \{1, 2, 3\}; D_3 = \{1, 2, 3, 4\};$

$r = \{(\text{mela}, 1, 1), (\text{mela}, 3, 1), (\text{pera}, 1, 3), (\text{pera}, 2, 2), (\text{pera}, 3, 4)\}$

è una relazione su D_1, D_2, D_3 ($r \subseteq D_1 \times D_2 \times D_3$)

- * $D_1, D_2, \dots, D_n$ sono i **domini** della relazione
- * Il valore di n è detto **grado** (o **arità**) della relazione
- * Il numero di n-ple di una relazione è la sua **cardinalità**

Rappresentazione di relazioni

| rappresentazione insiemistica

- * non adeguata per relazioni di grado $n > 2$ e/o con cardinalità superiore a qualche unità

| rappresentazione tabellare

- * molto efficace e intuitiva

mela	1	1
mela	3	1
pera	1	3
pera	2	2
pera	3	4

L'importanza della posizione

- Nel caso di domini ripetuti, l'interpretazione dei dati si complica e la posizione assume un ruolo determinante

partite $\subseteq$ String $\times$ String $\times$ Integer $\times$ Integer

Benetton TV	Poliform Cantù	100	71
Kinder BO	MontePaschi SI	90	51
Paf BO	Adr RM	62	97
Adr RM	Kinder BO	80	62

- Il primo e il terzo dominio si riferiscono alla squadra ospitante (nome e numero di punti), mentre il secondo e il quarto dominio si riferiscono alla squadra ospitata

È SCOMODO E POCO CHIARO!!

Relazione nel modello relazionale

- Ad ogni occorrenza di dominio (ripetuto o meno) si associa un nome univoco nella relazione, detto *attributo*, il cui compito è specificare il ruolo che quel dominio svolge nella relazione (“cosa significa”)
- Nella rappresentazione tabellare, gli attributi sono le intestazioni delle colonne

TeamCasa	TeamOspite	PuntiCasa	PuntiOspite
Benetton TV	Poliform Cantù	100	71
Kinder BO	MontePaschi SI	90	51
Paf BO	Adr RM	62	97
Adr RM	Kinder BO	80	62

- La struttura non è più posizionale, ovvero *l'ordine degli attributi non ha più rilevanza!*

Relazione: una definizione formale

- | Si indichi con $\text{dom}(A)$ il dominio dell'attributo A e si consideri un insieme di attributi $X = \{A_1, A_2, \dots, A_n\}$
- | Una **relazione** su X è un sottoinsieme del prodotto cartesiano tra i domini degli attributi:
$$r(X) \subseteq \text{dom}(A_1) \times \text{dom}(A_2) \times \dots \times \text{dom}(A_n)$$
- | Ogni elemento di $r(X)$ è detto **tupla** su X
- | Lo **schema** di una relazione su X è dato da un nome R e dall'insieme di attributi X , scritto $R(X)$

Partite

TeamCasa	TeamOspite	PuntiCasa	PuntiOspite
Benetton TV	Poliform Cantù	100	71
Kinder BO	MontePaschi SI	90	51
Paf BO	Adr RM	62	97
Adr RM	Kinder BO	80	62

Notazione di base

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

Partite t	TeamCasa	TeamOspite	PuntiCasa	PuntiOspite
	Benetton TV	Poliform Cantù	100	71
	Kinder BO	MontePaschi SI	90	51
	Paf BO	Adr RM	62	97
	Adr RM	Kinder BO	80	62

$t[\text{TeamOspite}] = \text{"MontePaschi SI"}$

Databasi relazionale

- | Lo **schema di un DB** relazionale è un insieme di schemi di relazioni con nomi distinti

$$R = \{R_1(X_1), R_2(X_2), \dots, R_m(X_m)\} \quad (R_i \neq R_j \ \forall i \neq j)$$

- | L'**istanza** di un DB con schema $R = \{R_1(X_1), R_2(X_2), \dots, R_m(X_m)\}$ è un insieme di (istanze) di relazioni

$$r = \{r_1, r_2, \dots, r_m\}$$

con r_i istanza su $R_i(X_i)$

Esempio: DATI_AZIENDA =

{ IMPIEGATI(Matricola, Cognome, Nome, Livello, Stipendio),
FILIALI(CodFiliale, Nome, Indirizzo, Direttore),
FORNITORI(RagSoc, Indirizzo, PartitaIVA) }

Un semplice DB relazionale

Studenti	Matricola	Cognome	Nome	DataNascita
	29323	Bianchi	Giorgio	21/06/1978
	35467	Rossi	Anna	13/04/1978
	39654	Verdi	Marco	20/09/1979
	42132	Neri	Lucia	15/02/1978

Corsi	CodCorso	Titolo	Docente	Anno
	483	Analisi	Biondi	1
	729	Analisi	Neri	1
	913	Sistemi Informativi	Castani	2

Esami	Matricola	CodCorso	Voto	Lode
	29323	483	28	no
	39654	729	30	sì
	29323	913	26	no
	35467	913	30	sì

Tabelle vs Relazioni

- | In realtà, i termini "tabella" e "relazione" non sono affatto sinonimi; una relazione del modello relazionale può essere vista come un particolare tipo di tabella.
- | Una tabella rappresenta una relazione se:
 - * I valori di ciascuna colonna sono tra loro omogenei (definiti sullo stesso dominio)
 - * Le righe sono tra loro diverse
 - * Le intestazioni delle colonne sono diverse tra loro
- | In una tabella che rappresenta una relazione, l'ordinamento delle righe e delle colonne è irrilevante
- | Il linguaggio SQL consente di gestire tabelle che non sono relazioni e che ammettono righe duplicate

Prima forma normale

- | Il modello relazionale non permette di usare domini arbitrari per la definizione delle relazioni; in particolare non è in generale possibile usare domini strutturati (array, set, liste, ...)
 - * Vi sono eccezioni notevoli (ad esempio le date e le stringhe)
- | Concisamente, una relazione in cui ogni dominio è *atomico* (non ulteriormente decomponibile) si dice che è in **Prima Forma Normale, o 1NF (1st Normal Form)**
- | In molti casi è pertanto richiesta un'attività di normalizzazione dei dati che dia luogo a relazioni in 1NF e che preservi l'informazione originale

Normalizzazione

Ricevuta n. 231 del 12/02/2002		
Coperti	2	3,00
Antipasti	1	5,80
Primi	2	11,45
Secondi	2	22,30
Caffè	2	2,20
Vino	1	8,00
Totale (Euro)		52,75

Ricevuta n. 352 del 13/02/2002		
Coperti	1	1,50

Ricevute

Numero	Data	Totale
231	12/02/2002	52,75
352	13/02/2002	...
...	...	...

Dettaglio

Numero	Quantità	Descrizione	Prezzo
231	2	Coperti	3,00
231	1	Antipasti	5,80
231	2	Primi	11,45
231	2	Secondi	22,30
231	2	Caffè	2,20
231	1	Vino	8,00
352	1	Coperti	1,50

Informazione incompleta

- Le informazioni che si vogliono rappresentare mediante relazioni non sempre corrispondono pienamente allo schema prescelto, in particolare per alcune tuple e alcuni attributi **potrebbe non essere possibile specificare**, per diversi motivi, **un valore del dominio**

Fisico	DataMorte
Enrico Fermi	29/11/1954
Franco Rasetti	
Bruno Pontecorvo	
Ettore Majorana	

- * Franco Rasetti è ancora vivo (valore **non applicabile**)
- * Bruno Pontecorvo è deceduto, ma non conosciamo la data di morte (**applicabile ma ignoto**)
- * Ettore Majorana, non è noto se sia deceduto o se sia ancora vivo (**ignota l'applicabilità**)

Quale soluzione?

- | In diversi casi, in mancanza di informazione, si tende a usare un “**valore speciale**” del dominio (0, “”, “-1”, “9999”, ecc.) che non si utilizza per altri scopi
- | Questa pratica è **fortemente sconsigliata**, in quanto, anche dove possibile:
 - * valori inutilizzati potrebbero successivamente diventare significativi
 - * le applicazioni dovrebbero sapere cosa significa in realtà il valore usato allo scopo
 - » Esempio (reale!): nel 1998, analizzando i clienti di un'assicurazione, si scoprì una strana concentrazione di ultra-novantenni... tutte le date di nascita ignote erano state codificate con “01/01/00”!!
- | Nel modello relazionale si opera in maniera pragmatica: si adotta il concetto di **valore nullo** (NULL), che denota assenza di un valore nel dominio (e non è un valore del dominio). Pertanto

$$t[A] \in \text{dom}(A) \cup \{\text{NULL}\}$$

Fisico	DataMorte
Enrico Fermi	29/11/1954
Franco Rasetti	NULL
Bruno Pontecorvo	NULL
Ettore Majorana	NULL

Valori nulli: restrizioni

- La presenza di valori nulli non può essere sempre tollerata, ovvero è necessario imporre delle restrizioni al loro uso; si consideri ad esempio il caso della registrazione di esami:

Esami	Matricola	CodCorso	Voto	Lode
	29323	483	28	no
	NULL	729	30	sì
	29323	913	NULL	no
	35467	913	30	no

- * Il valore nullo per Matricola non permette di sapere chi ha sostenuto l'esame
- * Il valore nullo per Voto non è ammissibile nel contesto considerato



Istanze di questo tipo non sono accettabili!

Vincoli di integrità

- La “correttezza sintattica” di un'istanza non è condizione sufficiente affinché i dati rappresentino informazione possibile nel contesto reale considerato

Studenti	Matricola	Cognome	Nome	DataNascita
	35467	Bianchi	Giorgio	21/06/1978
	35467	Rossi	Anna	13/04/1978
	39654	Rossi	Anna	13/04/1978

- * La prima e la seconda tupla hanno la stessa Matricola (!?)
 - * La seconda e la terza tupla hanno gli stessi valori per Nome, Cognome e DataNascita, ma questo in linea di principio è possibile
-
- Un *vincolo di integrità* è una proprietà che deve essere soddisfatta da ogni possibile istanza; ogni vincolo può quindi essere descritto da una funzione booleana che associa a ogni istanza il valore VERO o FALSO

Vincoli di dominio

- Un vincolo che si riferisce ai valori ammissibili per un singolo attributo viene detto *vincolo di dominio* (o sui valori)

Esami	Matricola	CodCorso	Voto	Lode
	29323	483	28	no
	39654	729	30	sì
	29323	913	31	no
	35467	913	30	forse

- * Il voto deve essere compreso tra 18 e 30
(Voto $\geq$ 18) AND (Voto $\leq$ 30)
- * La lode può solo assumere i valori `sì' o `no'
(Lode = `sì') OR (Lode = `no')

Vincoli di tupla

- I vincoli di dominio sono un caso particolare dei *vincoli di tupla*, ovvero vincoli che esprimono condizioni su ciascuna tupla, indipendentemente dalle altre

Esami	Matricola	CodCorso	Voto	Lode
	29323	483	28	no
	39654	729	30	sì
	29323	913	26	sì
	35467	913	30	no

- * La Lode si può assegnare solo se il Voto è 30:
(Voto = 30) OR NOT(Lode = `sì')

Vincoli di chiave: intuizione

- I **vincoli di chiave**, che giocano un ruolo molto importante nel modello relazionale, vietano la presenza di tuple distinte che hanno lo stesso valore su uno o più attributi

Studenti

Matricola	CodiceFiscale	Cognome	Nome	DataNascita
210629323	BNCGRG78L21A944K	Bianchi	Giorgio	21/07/1978
216635467	RSSNNA78A53A944V	Rossi	Anna	13/01/1978
160239654	VRDMRC79H20F839X	Verdi	Marco	20/06/1979
214842132	VRDMRC79H20G125J	Verdi	Marco	20/06/1979

- * Il valore di Matricola identifica univocamente uno studente
- * Analogamente per il CodiceFiscale...e per ogni insieme di attributi che includa Matricola o CodiceFiscale: {Matricola, Cognome}, {CodiceFiscale, Nome},...
- * Viceversa, possono esistere due tuple uguali su {Cognome, Nome, DataNascita}

Chiavi e superchiavi

- | Dato uno schema $R(X)$, un insieme di attributi $K \subseteq X$ è
 - * una **superchiave** se e solo se in ogni istanza ammissibile r di $R(X)$ non esistono due tuple distinte $t1$ e $t2$ tali che $t1[K] = t2[K]$
 - * una **chiave** se e solo se è una superchiave minimale, ovvero non esiste $K' \subset K$ con K' superchiave
- | Una chiave è pertanto un *identificatore minimale* per ogni r su $R(X)$
 - * Nella relazione Studenti:
 - » $\{\text{Matricola}\}$ e $\{\text{CodiceFiscale}\}$ sono due chiavi
 - » $\{\text{Matricola}, \text{Cognome}\}$ e $\{\text{CodiceFiscale}, \text{Nome}\}$ sono solo superchiavi
 - » $\{\text{Cognome}, \text{Nome}, \text{DataNascita}\}$ non è superchiave

Importanza delle chiavi

- | L'esistenza delle chiavi garantisce l'**accessibilità a ciascun dato del DB**, in quanto ogni singolo valore è univocamente individuato da:
 - * Nome della relazione (individua una relazione del DB)
 - * Valore della chiave (individua una tupla della relazione)
 - * Nome dell'attributo (individua il valore desiderato)

- | Le chiavi sono lo strumento principale attraverso il quale vengono **correlati i dati in relazioni diverse** (il modello relazionale è basato su valori, non esistono puntatori)

Importanza delle chiavi

<u>codL</u>	nome
L001	TWA
L002	ALITALIA
....	

LINEE

<u>codDest</u>	nomedest	nazione
FCO	ROMA	ITALIA
JFK	NEW YORK	USA
.....		

DESTINAZIONI

<u>numvolo</u>	codL	codDest	giorno	ora	durata	attivo
TW056	L001	JFK	LUN	9:00	2	SI
AZ854	L002	FCO	MER	22:30	8	SI
....						

VOLI

Chiavi e valori nulli

- In presenza di valori nulli entrambe le funzioni svolte dalle chiavi possono venire a mancare

Studenti

Matricola	CodiceFiscale	Cognome	Nome	DataNascita
NULL	NULL	Bianchi	Giorgio	21/07/1978
216635467	RSSNNA78A53A944V	Rossi	Anna	13/01/1978
NULL	VRDMRC79H20F839X	Verdi	Marco	20/06/1979
214842132	NULL	Verdi	Marco	20/06/1979

- * La prima tupla non è identificabile in alcun modo, pertanto:
È necessario specificare il valore di almeno una chiave!
- * La terza e quarta tupla potrebbero riferirsi allo stesso studente, pertanto:
Non è sufficiente specificare il valore di una chiave!

Chiave primaria

- Per evitare i problemi visti è necessario scegliere una chiave, detta *chiave primaria*, su cui non si ammettono valori nulli
- Convenzionalmente, gli attributi della chiave primaria vengono sottolineati

Studenti

<u>Matricola</u>	CodiceFiscale	Cognome	Nome	DataNascita
210629323	NULL	Bianchi	Giorgio	21/07/1978
216635467	RSSNNA78A53A944V	Rossi	Anna	13/01/1978
160239654	VRDMRC79H20F839X	Verdi	Marco	20/06/1979
214842132	NULL	Verdi	Marco	20/06/1979



Nei casi in cui per nessuna chiave si possa garantire la disponibilità di valori è necessario introdurre un nuovo attributo (*chiave surrogata*), un codice che svolga le funzioni di chiave primaria (si pensi ad esempio al caso in cui non si riesca a identificare un paziente al pronto soccorso ospedaliero)

Vincoli di integrità referenziale

- * I vincoli sinora visti sono tutti di tipo intra-relazionale, in quanto interessano una relazione alla volta
 - * Viceversa, i vincoli di integrità referenziale sono importanti tipi di vincoli inter-relazionali, che enfatizzano come le correlazioni tra le tuple siano spesso ottenute usando i valori delle chiavi
- I Si considerino due schemi $R_1(X_1)$ e $R_2(X_2)$ di un DB R , e sia Y un insieme di attributi in X_2 . Un **vincolo di integrità referenziale** su Y impone che in ogni istanza $r = \{r_1, r_2, \dots\}$ del DB l'insieme dei valori di Y in r_2 sia un sottoinsieme dell'insieme dei valori della chiave primaria di $R_1(X_1)$ presenti nell'istanza r_1
- I L'insieme Y viene detto una **foreign key** (o *chiave importata* o *chiave esterna*)

Esempi di foreign key

Studenti

<u>Matricola</u>	Cognome	Nome	DataNascita
29323	Bianchi	Giorgio	21/06/1978
35467	Rossi	Anna	13/04/1978
39654	Verdi	Marco	20/09/1979
42132	Neri	Lucia	15/02/1978

Corsi

<u>CodCorso</u>	Titolo	Docente	Anno
483	Analisi	Biondi	1
729	Analisi	Neri	1
913	Sistemi Informativi	Castani	2

Esami

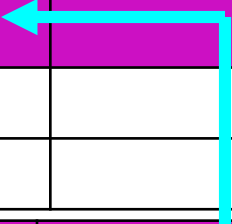
<u>Matricola</u>	<u>CodCorso</u>	Voto	Lode
29323	483	28	no
39654	729	30	sì
29323	913	26	no
35467	913	30	sì

In Esami, {Matricola} è una foreign key, così come {CodCorso}

Foreign key: precisazioni

- In generale la foreign key Y e la primary key K di $R_1(X_1)$ possono includere attributi con nomi diversi

Corsi	<u>Codice</u>		Titolo	Docente	Anno
	483		Analisi I	Biondi	1
	729		Analisi I	Neri	1
Esami	<u>NumMatricola</u>	<u>CodCorso</u>	Voto	Lode	
	29323	483	28	no	



- Foreign key e primary key possono far parte della stessa relazione, ovviamente con $Y \neq K$

Personale	<u>Codice</u>	Cognome	...	CodResponsabile
	123	Rossi	...	325
	134	Verdi	...	325
	325	Neri	...	...



Sommario

- | Il **modello relazionale** è basato sul concetto di **relazione**, che estende quello di relazione matematica tra n domini associando a ciascuna occorrenza di dominio un nome, detto **attributo**.
- | Lo **schema** di una relazione consiste di un **nome** e di un **insieme di attributi**; l'**istanza** di una relazione è un **insieme di tuple**, ovvero funzioni che associano a ogni attributo dello schema un valore del corrispondente dominio.
- | In **assenza di informazioni** si fa uso di un particolare valore, detto **valore nullo (NULL)**, che non appartiene a nessun dominio.
- | Per garantire l'**integrità dei dati** si possono specificare diversi tipi di **vincoli**, che definiscono quali sono le istanze legali (ammissibili).
- | I **vincoli intra-relazionali** includono quelli sui **domini**, sulle **tuple** e i vincoli di **chiave**; i **vincoli inter-relazionali** quelli di **integrità referenziale**. Questi ultimi permettono di stabilire le principali correlazioni tra i dati di diverse relazioni.