

Modello Relazionale

Università degli Studi del Sannio

Facoltà di Ingegneria

Corso di Laurea in Ingegneria Informatica

Corso di Basi di Dati

Anno Accademico 2005/2006

docente: ing. Corrado Aaron Visaggio

email: visaggio@unisannio.it

ricevimento: mercoledì 11.00-13.00.

Storia...

Tutto ha inizio negli **anni 60**, ad opera di **Charles W. Bachman**, che dirigeva il “**Database Task Group**” in un programma di standardizzazione del linguaggio Cobol, chiamato “CodaSyl”.

- Navigazione manuale
- Dati organizzati a rete, con puntatori

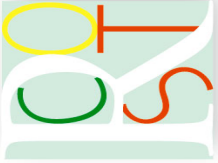
Nel **1968** la **IBM** sviluppa un proprio sistema DBMS, chiamato **IMS**, un programma utilizzato nella missione Apollo sui Sistemi/360: il sistema era gerarchico. Nasce così la definizione di “database navigazionali”. Bachman nel **1973** fu insignito del Premio Turing.

Il modello relazionale è stato introdotto nel **1970** da E.F. Codd (un ricercatore dell’IBM di San Jose, CA) allo scopo di favorire **l’efficienza e l’indipendenza dei dati**.

L’articolo di riferimento si chiama “**A Relational Model of Data for Large shared Data Banks**”, ed introduce i concetti di tabella, record e tupla.

...Storia

- **Anni 70**: definizione del modello, prima versione del linguaggio SQL (allora SEQUEL), studi fondamentali sulla tecnologia relazionale (ottimizzazione, transazione, recovery,...) e primi prototipi di DBMS relazionali:
 - System R (IBM, laboratorio di ricerca IBM, San Jose, CA). Nel '74 il sistema era monotabella, poi i dati furono divisi in più tabelle; nel '78 si inserì la **multiutenza**
 - Ingres (Università di Berkeley, CA, USA). Si intendeva creare un database geografico. Al progetto lavorano più di 30 persone, e nascono diverse spin-off, tra cui Informix, Sybase e NonStop SQL.
- **Anni 80**: prima standardizzazione di SQL, primi prototipi commerciali:
 - SQL/DS (derivato da System R)
 - Oracle
 - IBM DB2 (derivato da SQL/DS).



...Storia

- **Anni 90**: standard ISO-ANSI SQL-2 (quello di riferimento , anche noto come SQL-92).
- **Anni 2000**: Datawarehouse, OLTP, Sistemi Informativi Geografici, Sistemi a Supporto delle Decisioni.

Relazioni e Tabelle...

Il modello relazionale fu proposto per permettere, a *livello logico*, una descrizione efficace e significativa del *modello fisico*.

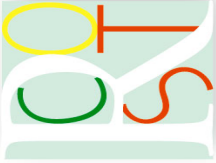
Tre accezioni per il termine *relazione*:

-
- relazione matematica
 - relazione secondo il modello relazionale (*tabella*)
 - relazione nel senso di relationship nel diagramma E-R.

Dati due insiemi D_1 e D_2 si definisce **prodotto cartesiano** $D_1 \times D_2$ l'insieme delle **coppie ordinate** (v_1, v_2) tali che $v_1 \in D_1$ e $v_2 \in D_2$.

Una **relazione matematica** sugli insiemi D_1 e D_2 (**domini della relazione**) è un **sottoinsieme proprio del prodotto cartesiano** $D_1 \times D_2$.

Le **relazioni** (nelle basi di dati) devono essere **insiemi finiti** per ovvi limiti fisici, nonostante i **domini** di definizione sono spesso (realmente) **infiniti**.



...Relazioni e Tabelle...

Dati $D_1, D_2, D_3, \dots, D_n$ insiemi non necessariamente distinti, con $n > 0$, il prodotto cartesiano $D_1 \times D_2 \times D_3 \times \dots \times D_n$ è costituito dalle n -uple $(v_1, v_2, v_3, \dots, v_n)$ tali che $v_i \in D_i$ per $1 \leq i \leq n$.

Una relazione matematica sui domini $D_1, D_2, D_3, \dots, D_n$ è un sottoinsieme del prodotto cartesiano $D_1 \times D_2 \times D_3 \times \dots \times D_n$, ove n è detto **grado** del prodotto cartesiano (relazione).

Il numero di elementi della relazione, cioè di n -uple è chiamato **cardinalità**.

Una relazione matematica è un insieme di **n -uple ordinate** $(v_1, v_2, v_3, \dots, v_n)$ con $v_i \in D_i$ per $1 \leq i \leq n$.

Relazione matematica  Tabella:

- Ciascuna n -upla contiene dati tra loro collegati
- Non è definito alcun ordinamento fra le n -uple
- Le n -uple di una relazione sono distinte l'una dall'altra.
- Ciascuna n -upla è ordinata al suo interno.

...Relazioni e Tabelle

La funzione $DOM : X \rightarrow D$ associa a ciascun attributo $A \in X$ un dominio $DOM(A) \in D$.

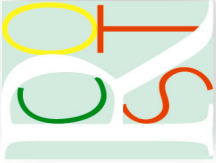
Una **tupla** su un insieme di attributi X è una funzione che associa a ciascun attributo $A \in X$ un valore del dominio $DOM(A)$.

Una relazione su X è un insieme di tuple su X .

Nella relazione matematica gli elementi sono individuati grazie ad una **notazione posizionale**. Nella tabella gli elementi sono individuati grazie agli attributi.

Se t è una tupla su X e $A \in X$, allora $t[A]$ o $t.A$ indica il valore di t su A . Se t è la prima tupla, allora $t[SquadraOspitata] = Lazio$.

SquadraDiCasa	SquadraOspitata	RetiD	RetiO
Juventus	Lazio	3	1
Inter	Fiorentina	5	2
Bari	Roma	2	1
Benevento	Milan	3	0



Relazioni e Basi di Dati

Il modello relazionale è **basato su valori**: i riferimenti fra dati in relazioni diverse sono rappresentati dai valori dei domini che compaiono nelle tuple.

Il modello reticolare e gerarchico realizzano le corrispondenze in modo esplicito attraverso puntatori e vengono detti **basati su record e puntatori**.

Vantaggi del modello relazionale:

- rappresenta solo ciò che serve all'utente: dati non puntatori!
- La rappresentazione logica non fa riferimento a quella fisica
- l'informazione è contenuta nei valori: è semplice trasferire i dati da un contesto all'altro.

I modelli gerarchico e reticolare preferivano gli **aspetti di efficienza** a quelli di portabilità, manutenibilità ed usabilità.

Il modello relazionale, a differenza degli altri due è formalmente definito: esiste una **teoria relazionale!**

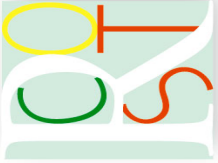
Definizioni

Uno **schema di relazione** è costituito da un simbolo R , detto nome della relazione e da un insieme di (nomi di) attributi $X = \{A_1, A_2, A_3, \dots, A_n\}$, il tutto di solito indicato con $R(X)$.

Uno **schema di base di dati** è un insieme di schemi di relazione con nomi diversi $R = \{R_1(X_1), R_2(X_2), \dots, R_n(X_n)\}$

Una **istanza di relazione** r su uno schema $R(X)$ è un insieme r di tuple su X .

Un' **istanza di base di dati** su uno schema $R = \{R_1(X_1), R_2(X_2), \dots, R_n(X_n)\}$ è un insieme di relazioni $r = \{r_1, r_2, r_3, \dots, r_n\}$, dove ogni r_i per $1 \leq i \leq n$ è una relazione dello schema $R_i(X_i)$.



Informazione incompleta e valori nulli

In molti casi i dati a disposizione possono non corrispondere al formato previsto.

Non sarebbe corretto utilizzare un elemento del dominio per rappresentare l'assenza di informazione: si potrebbe generare confusione.

I problema è risolto con il valore nullo (NULL), che si usa:

- quando il valore è sconosciuto;
- quando il valore è inesistente;
- quando non ci sono informazioni, vale a dire, non so se esiste e se esiste non so quale sia.

Vincoli di Integrità

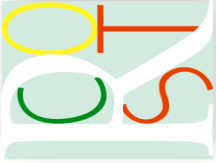
Un vincolo di integrità è una **proprietà** che deve essere soddisfatta dalle istanze che rappresentano informazioni corrette per l'applicazione.

Un vincolo è un **predicato** che associa ad ogni istanza il valore **vero** o **falso**. Se il predicato assume il balore vero, allora l'istanza soddisfa il vincolo.

Un vincolo è **intrarelazionale** se il suo soddisfacimento è definito rispetto alle singole relazioni della base di dati.

- Un **vincolo di tupla** è un vincolo che può essere valutato su ciascuna tupla indipendentemente dalle altre
- Un **vincolo su valori** o di dominio è definito con riferimento a singoli valori.

Un vincolo è **interrelazionale** se coinvolge più relazioni.



Vincoli di tupla e Chiavi...

Due esempi:

- un solo attributo: (VOTO \geq 18) AND (VOTO \leq 30)
- Più attributi: PAGAMENTI (DATA IMPORTO, RITENUTE, NETTO)
 NETTO= IMPORTO-RITENUTE.

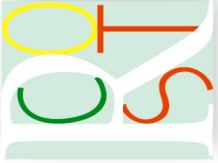
Un insieme K di attributi è **superchiave** di una relazione r se r non contiene due tuple distinte t_1 e t_2 con $t_1[K]=t_2[K]$.

K è **chiave** di r se è una superchiave minimale di r (cioè non esiste un'altra superchiave K' di r che sia contenuta in K come sottinsieme proprio).

...Vincoli di tupla e Chiavi...

Matricola	Cognome	Nome	Nascita	Corso
6789	Rossi	Dario	29/04/59	Ing. Informatica
1234	Rossi	Luca	01/05/61	Ing. Civile
2345	Neri	Luca	01/05/61	Ing. Informatica
3456	Neri	Luca	05/03/58	Ing. Civile

{MATRICOLA} è superchiave ed anche superchiave minimale
 {COGNOME; NOME; NASCITA} è superchiave, ma non lo è
 nessuno suo sottoinsieme;
 {MATRICOLA; CORSO} è superchiave, ma non minimale.
 {NOME; CORSO} non è superchiave.



...Vincoli di tupla e Chiavi

Ciascuna relazione e ciascuno schema di relazione hanno sempre una chiave.

L'insieme X di tutti gli attributi su cui è definita è senz'altro una superchiave per essa.

L'esistenza della chiave garantisce la **accessibilità a tutti i valori** di una base di dati e la loro **univoca identificabilità**.

La proliferazione di valori nulli rende impossibile capire se le due tuple facciano riferimento allo stesso record o meno: su una chiave, definita chiave primaria, si vieta la presenza di valori nulli.

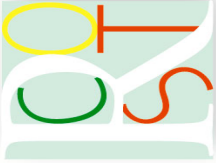
Molti codici identificativi sono stati introdotti, nel passato, proprio per consentire l'identificazione dei soggetti all'interno di un dominio.

Vincoli di integrità referenziale

Un vincolo di integrità referenziale fra un insieme di attributi X di una relazione R_1 ed un'altra relazione R_2 è soddisfatto se i valori su X di ciascuna tupla dell'istanza di R_1 compaiono come valori della chiave primaria dell'istanza di R_2 .

Se la chiave di R_2 è unica e composta da un solo attributo B allora il vincolo è soddisfatto se per ogni tupla t_1 in R_1 per cui $t_1[A]$ non è nullo, esiste una tupla t_2 in R_2 tale che $t_1[A] = t_2[B]$.

Nel caso più generale, ciascuno degli attributi in X deve corrispondere ad un preciso attributo della chiave primaria K in R_2 , ove $X = A_1A_2...A_p$ e $K = B_1B_2...B_p$. Il vincolo è soddisfatto se per ogni tupla t_1 in R_1 senza valori nulli su X esiste una tupla t_2 in R_2 con $t_1[A_i] = t_2[B_i]$, per ogni i compreso tra 1 e p .



1NF

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).

Una relazione in cui ogni dominio è **atomico**, si dice che è in prima forma normale, **1st Normal Form (1NF)**.

In molti casi è richiesta un'attività di normalizzazione dei dati che dia luogo a relazioni in 1NF e che preservi l'informazione originale.

Esempio Normalizzazione 1NF...

	Nome	Cognome	Data Di Nascita	Indirizzo
CF 1234	Gio'	Vanni	01/01/01	Via Taglia la Zazzera n°20, Bari
2345	Anna	Lacapa	29/02/04	Via Taglia la Zazzera n°18, Bari
3456	Natale	di Pasqua	24/12/00	Piazza Santo Stefano n°1, Salerno
4567	Marco	Antonio	15/03/03	Via Salaria n°4, Ostia (Roma)

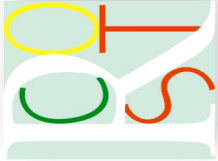


Id	Den	Valore	Numero Civico	Città	Provincia
1	Via	Taglia la Zazzera	20	Bari	Null
2	Via	Taglia la Zazzera	18	Bari	Null
3	Piazza	Santo Stefano	1	Salerno	Null
4	Via	Salaria	4	Ostia	Roma

...Esempio Normalizzazione 1NF



CF	Nome	Cognome	Data Di Nascita	Indirizzo
1234	Gio'	Vanni	01/01/01	1
2345	Anna	Lacapa	29/02/04	2
3456	Natale	di Pasqua	24/12/00	3
4567	Marco Antonio	Romano	15/03/03	4



Any question?

