

MODELLO RELAZIONALE

Introduzione

E' stato proposto agli inizi degli anni '70 da Codd finalizzato alla realizzazione dell'indipendenza dei dati, unisce concetti derivati dalla teoria degli insiemi (relazioni) con una rappresentazione dei dati di tipo tabellare.

E' stato teorizzato per separare il più possibile il livello logico dal livello fisico della descrizione dei dati, basato su un rigoroso modello matematico e permette un elevato grado di astrazione.

Un database relazionale e' un database costruito basandosi sul modello relazionale, proposto da E.F. Codd , in cui le entità di interesse vengono rappresentate mediante una struttura dati chiamata "relazione". I sistemi di gestione di dati relazionali vengono chiamati Relational DataBase Management System (RDBMS).

La rappresentazione del modello è semplice ed intuitiva.

Le relazioni ed i risultati delle operazioni su di esse sono facilmente rappresentabili ed interpretabili dagli utenti.

Prodotto Cartesiano

def.: **Prodotto Cartesiano**

Dati due insiemi $D1$ e $D2$ si definisce Prodotto Cartesiano di $D1$ e $D2$, e si indica con $D1 \times D2$, l'insieme delle coppie ordinate $(v1, v2)$ tali che $v1$ sia un elemento di $D1$ e $v2$ sia un elemento di $D2$.

Es. Dati gli insiemi $FORME = \{\text{cubo, cono}\}$ e $COLORI = \{\text{rosso, verde, blu}\}$ il loro prodotto cartesiano è $\{(\text{cubo,rosso}),(\text{cono,rosso}),(\text{cubo,verde}),(\text{cono,verde}),(\text{cubo,blu}),(\text{cono,blu})\}$

Una relazione matematica su due insiemi $D1$ e $D2$ è un sottoinsieme di $D1 \times D2$.

NOTA: a livello formale gli insiemi possono essere infiniti, a livello pratico non possiamo però considerare relazioni infinite.

Es. dati gli insiemi visti, una possibile relazione è $\{(\text{cubo,rosso}),(\text{cono,rosso}),(\text{cubo,blu})\}$ o, in forma tabellare,

FORME	COLORI
cubo	rosso
cono	rosso
cono	blu

Le definizioni viste per due insiemi possono essere generalizzate a n insiemi.

Attributo, tupla, relazione

def.: un **attributo** è l'identificatore di un dominio su cui è definito.

def.: una **tupla** è un insieme di valori, uno per attributo, ordinati secondo lo schema della relazione e definiti ciascuno su un proprio dominio; ogni riga della tabella è una **tupla** ordinata di elementi

def.: una **relazione** è una serie di tuple definite sul dominio della relazione (insieme ordinato dei domini dei singoli attributi).

def.: n è detto **grado** del prodotto cartesiano e quindi della relazione.

def.: Il numero di elementi della relazione è detto **cardinalità** della relazione.

Osservazioni

- Non esiste ordinamento fra le tuple
- Non possono esistere 2 tuple uguali
- Ogni tupla è internamente ordinata quindi se si scambiano due righe, la relazione non cambia
- Se si scambiano 2 colonne può cambiare o perdere consistenza
- Un insieme può apparire più volte in una relazione.

Esempio: La relazione *Risultati Schedina* è un sottoinsieme del prodotto cartesiano SQUADRA \times SQUADRA \times RISULTATO

N.B. Per spiegare il significato del termine "relazione" occorre inquadrarlo nel giusto contesto. Infatti, nell'ambito dei database esso viene utilizzato almeno in due accezioni. Se si sta parlando del modello relazionale, una relazione è la struttura principale del modello e corrisponde alla definizione sopra esposta.

In altri contesti (ad esempio considerando diagrammi Entity-Relationship), una relazione è un legame fra due entità.

Utilizzando la lingua inglese questa ambiguità di solito non sussiste, infatti per il primo significato viene usato il termine "relation", mentre per il secondo viene usato "relationship".

Modello sui valori

Il modello relazionale è basato su valori, i riferimenti fra dati in relazioni diverse avvengono attraverso i valori dei domini corrispondenti che appaiono nelle tuple.

Gli altri modelli (gerarchico, reticolare) utilizzano puntatori per le corrispondenze e sono detti basati su record e puntatori.

I vantaggi dell'approccio per valori sono:

- Si inseriscono nella base di dati solo valori significativi per l'applicazione (i puntatori sono dati aggiuntivi relativi alla sola implementazione);
- Il trasferimento dei dati da un ambiente ad un altro è più semplice (i puntatori hanno validità solo locale);
- La rappresentazione logica dei dati non fa riferimento a quella fisica e quindi si ottiene l'indipendenza dei dati

Schemi

Se t è una tupla definita sul dominio X della relazione e A è uno dei domini di X $t[A]$ (o $t.A$) è il valore di t relativo al dominio A

Esempio: (relazione COLOREFORMA) se t è la prima tupla della relazione $t.forme = \text{cubo}$

Uno **schema di relazione** $R(X)$ è costituito da un simbolo (nome della relazione) R e da una serie di attributi $X = \{A_1, A_2, \dots, A_n\}$

Esempio: Corsi (Codice, NomeCorso, Docente)

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

Una 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 $r = \{r_1, r_2, \dots, r_n\}$ dove ogni r è una relazione sullo schema $R_i(X_i)$.

Vincoli d'integrità

Sono condizioni, sotto forma di predicati logici, che sono inserite nella base di dati per garantirne la consistenza.

Ogni istanza della base di dati deve soddisfare i vincoli di integrità.

Una istanza che soddisfi tutti i vincoli è detta corretta (o lecita o ammissibile)

I vincoli di integrità possono essere di vario tipo:

- Intrarelazionale se coinvolge attributi della stessa relazione
- Vincoli di tupla possono essere valutati su ciascuna tupla indipendentemente dalle altre
- Vincoli di dominio sono definiti su singoli valori
- Interrelazionale se coinvolge più relazioni

Le tuple che compongono la base di dati devono essere omogenee.

Quindi ad ogni attributo deve essere associato un valore in ogni tupla. Non sempre questo è possibile.

Esempio: Persone(Cognome, Nome, Indirizzo, Telefono). Potrebbe esserci una persona che non ha telefono, o di cui non conosciamo l'indirizzo.

In questo caso è definito un valore convenzionale, detto valore nullo, che indica la non disponibilità dell'informazione.

Il valore nullo può rappresentare 3 tipi di informazione:

- sconosciuta
- inesistente
- indeterminata (nei DBMS disponibili si considera in genere questo caso)

Non tutte le combinazioni possibili di valori dei domini su cui è definita una relazione sono accettabili.

Alcuni attributi possono assumere valori in un certo range

Alcuni attributi devono essere diversi in ogni tupla della stessa relazione

Esempio: valori dell'attributo Matricola in una relazione del tipo Studenti (Matricola, Cognome, Nome, DataNascita)

Alcuni valori possono essere incompatibili con altri all'interno della stessa relazione

Esempio: data la relazione Esami (Matricola, Voto, Lode, CodCorso) una stessa coppia Matricola, Corso può apparire una sola volta. Il valore Vero per l'attributo Lode è corretto solo se Voto=30

Alcuni valori possono essere incompatibili con i valori di un'altra relazione

Esempio: Data la relazione Esami e la relazione Corsi (CodCorso, Titolo, Docente) ogni valore di CodCorso in Esami deve essere un valore esistente di CodCorso in Corsi

I vincoli di integrità possono essere definiti attraverso operatori booleani

Esempio: (Voto >= 18) AND (Voto <= 30) | (NOT (lode=Vero)) OR (Voto=30)

oppure, data la relazione Pagamenti(Data, Importo, Ritenute, Netto) -> Netto = Importo – Ritenute

Chiavi

Una chiave è un insieme minimale di attributi utilizzato per identificare univocamente le tuple di una relazione.

Formalmente: Un insieme di attributi K è superchiave per una relazione r se r non contiene due tuple t_1 e t_2 tali che $t_1[K] = t_2[K]$

Un insieme di attributi K è chiave per r se è superchiave minimale, cioè se non esiste un'altra superchiave K' che sia sottoinsieme di K

Una chiave è tale se soddisfa la definizione per tutte le possibili tuple appartenenti alla relazione, e non solo per quelle che effettivamente appaiono come istanze della relazione stessa.

Quindi la chiave è legata allo schema della relazione.

Ogni relazione, per definizione, possiede una chiave. Infatti l'insieme X su cui è definita è sicuramente superchiave.

La presenza di valori nulli in una chiave può creare può vanificare la proprietà di unicità delle tuple identificate da una certa chiave.

Si impone quindi che almeno una chiave non contenga valori nulli. Tale chiave è detta chiave primaria.

Di solito la chiave primaria è sottolineata nello schema di una relazione.

Esempio: Studenti (Matricola, Cognome, Nome, Nascita, Corso)