

Capitolo 1

Introduzione alle Basi di Dati

1.1 Sistemi informativi, informazioni e dati

Ogni organizzazione è dotata di un *sistema informativo*, che organizza e gestisce le informazioni necessarie per perseguire gli scopi dell'organizzazione stessa.

Per indicare la porzione automatizzata del sistema informativo di solito viene utilizzato il termine *sistema informatico*. Nei sistemi informatici le informazioni vengono rappresentate per mezzo di *dati*.

Una *base di dati* è una collezione di dati, utilizzati per rappresentare le informazioni di interesse per un sistema informativo.

1.2 Basi di dati e sistemi di gestione di basi di dati

Un *sistema di gestione di basi di dati* (DBMS) è un sistema software in grado di gestire collezioni di dati che siano

- *grandi*: in termini di occupazione di memoria
- *condivise*: applicazioni e utenti diversi devono poter accedervi
- *persistenti*: persistono anche dopo l'esecuzione del programma che le utilizza

assicurando la loro

- *affidabilità*: mantengono intatti i dati
- *privatezza*: mantengono sicuri e privati i dati

ed essendo

- *efficiente*: le operazioni vengono svolte rapidamente
- *efficace*: rendono produttive le attività dei loro utenti

1.3 Modelli dei dati

Un *modello di dati* è un insieme di concetti utilizzati per organizzare i dati di interesse e descriverne la struttura in modo che essa risulti comprensibile a un elaboratore.

Il *modello relazionale* dei dati permette di definire tipi per mezzo del costruttore *relazione*, che consente di organizzare i dati in insiemi di record a struttura fissa.

I *modelli concettuali* vengono utilizzati per descrivere i dati in maniera indipendente dal modello logico. Un tipo di modello concettuale è il modello *Entità-Relazione*.

1.3.1 Schemi e istanze

Nelle basi di dati esiste una parte sostanzialmente invariante nel tempo, detta *schema* della base di dati, costituita dalle caratteristiche dei dati, e una parte variabile nel tempo, detta *istanza* o *stato* della base di dati, costituita dai valori effettivi.

Lo schema di una relazione è costituito dalla sua intestazione, cioè dal nome della relazione seguito dai nomi dei suoi attributi, ad esempio:

Docenza(Corso, NomeDocente)

L'*istanza di una relazione* è costituita dall'insieme, variante nel tempo, delle sue righe.

Capitolo 2

Il modello relazionale

2.1 Il modello relazionale: strutture

2.1.1 Relazioni e tabelle

Dati due insiemi D_1 e D_2 , si chiama *prodotto cartesiano* di D_1 e D_2 l'insieme di coppie ordinate (v_1, v_2) tali che v_1 è un elemento di D_1 e v_2 è un elemento di D_2 . Il numero n delle componenti del prodotto cartesiano viene detto *grado* del prodotto cartesiano e della relazione. Il numero degli elementi (n -uple) della relazione viene chiamato *cardinalità* della relazione.

2.1.2 Relazioni con attributi

Nelle basi di dati, ciascuna n -upla contiene dati fra loro collegati. Inoltre, 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, in quanto tra gli elementi di un insieme non ce ne possono essere presenti due uguali tra loro

Ciascuna n -upla è, al proprio interno, ordinata: l' i -esimo valore di ciascuna proviene dall' i -esimo dominio.

Indichiamo con D l'insieme dei domini e specifichiamo la corrispondenza tra attributi e domini per mezzo della funzione $dom : X \rightarrow D$, che associa a ciascun attributo $A \in X$ un dominio $dom(A) \in D$. Diciamo che una *tupla* su un insieme di attributi X è una funzione t 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 .

2.1.3 Relazioni e basi di dati

Uno *schema di relazione* è costituito da un simbolo R , detto *nome della relazione*, e da un insieme di *attributi* $X = \{A_1, A_2, \dots, A_n\}$, indicato con $R(X)$. A ciascun attributo è associato un dominio.

Uno *schema di base di dati* è un insieme di schemi di relazione con nomi diversi:

$$R = \{R_1(X_1), R_2(X_2), \dots, R_n(X_n)\}$$

I nomi di relazione hanno come scopo principale quello di distinguere le varie relazioni nella base di dati.

Un *istanza di relazione* 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 dove ogni relazione è una relazione sullo schema $R_i(X_i)$.

2.2 Vincoli d'integrità

Il *vincolo d'integrità* è una proprietà che deve essere soddisfatta dalle istanze che rappresentano informazioni corrette per l'applicazione. Ogni vincolo può essere visto come un *predicato* che associa a ogni istanza il valore *vero* o *falso*. Se il predicato assume il valore vero, allora diciamo che l'istanza *soddisfa* il vincolo. Sono presenti due categorie di vincoli:

- Un vincolo è *intrarelazionale* se il suo soddisfacimento è definito rispetto a 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 definito con riferimento a singoli valori viene detto *vincolo su valori* o *vincolo di dominio*
- Un vincolo è *interrelazionale* se coinvolge più relazioni

2.2.1 Vincoli di tupla

I vincoli di tupla esprimono condizioni sui valori di ciascuna tupla, indipendentemente dalle altre tuple.

Chiavi

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

- 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 sottoinsieme proprio

Ciascuna relazione e ciascuno schema di relazione hanno sempre una chiave. Una relazione è un insieme e quindi è costituita da elementi fra loro diversi; di conseguenza, per ogni relazione $r(X)$, l'insieme X di tutti gli attributi su cui è definita è senz'altro una superchiave per essa. O tale insieme è anche chiave, nel qual caso si conferma l'esistenza della chiave stessa, oppure non è chiave, perchè esiste un'altra superchiave in esso contenuta.

Il fatto che su ciascuno schema di relazione possa essere definita almeno una chiave garantisce l'accessibilità a tutti i valori di una base di dati e la loro univoca identificabilità.

Chiavi e valori nulli

Su una delle chiavi, detta *chiave primaria* si vieta la presenza di valori nulli; sulle altre, i valori nulli sono generalmente ammessi.

Vincoli d'integrità referenziale

Un *vincolo d'integrità referenziale* fra un insieme di attributi X di una relazione R_1 e 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 , il vincolo di integrità referenziale fra l'attributo A di R_1 e la relazione R_2 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 $t_1[A] = t_2[B]$.