

Appunti Basi di Dati

Luca Seggiani

28 Febbraio 2024

1 Introduzione alla base di dati relazionale

Una base di dati (in inglese "database") è una mole di dati organizzata in modo da favorirne la consultazione attraverso determinate interrogazioni ("query"). Formalmente, una base di dati è un archivio di dati contenente informazioni ben strutturate organizzate secondo un modello logico (nel caso di MySQL un modello relazionale).

L'elemento fondamentale di una base di dati fondata sul modello relazionale è la **Tabella**, formata da colonne (associate a determinati attributi), e righe (che rappresentano record dei suddetti attributi). Una colonna particolare, che deve differire per ogni record della tabella, viene detta chiave primaria e serve a distinguere tra di loro record (che potrebbero essere altrimenti stati tra di loro identici!).

L'interrogazione della base di dati si effettua attraverso le cosiddette query, ovvero richieste compilate in linguaggio SQL (Structured Query Language), un linguaggio dichiarativo sviluppato appositamente per questo scopo. La query più semplice si basa su tre istruzioni: SELECT, FROM e WHERE. Prendiamo l'esempio in lingua naturale di un'interrogazione a una base di dati contenente dati su diverse persone (nello specifico, nome, cognome, età e codice fiscale come chiave.):

Riportare nomi e cognomi di tutte le persone con età maggiore di 40 anni.

In questo caso gli attributi di interesse sono i nomi e cognomi (in quest'ordine) delle persone valide, e la condizione è un'età maggiore di 40 anni. Di fronte a questa richiesta, la base di dati risponderà fornendo un insieme risultato (result set) contenente una tabella formata esattamente dalle informazioni richieste. In SQL potremmo definire tale query come segue:

```
1 SELECT Cognome, Nome  
2 FROM Persona  
3 WHERE Eta > 40
```

in questo caso, cognome e nome sono gli attributi di interesse che verrano inseriti nel nostro insieme risultato (attraverso un processo chiamato proiezione). La tabella persona sarà la fonte dei nostri dati, e età > 40 la nostra condizione.

Le condizioni in SQL supportano anche i classici operatori logici (AND, OR, ecc...). Un operatore degno di nota, applicabile ai soli dato di tipo numerico o su cui è comunque stabilita una relazione d'ordine, è il BETWEEN, che permette di selezionare (estremi inclusi) tutti i valori in un dato range.

Duplicati 2 righe risultano uguali se tutti gli attributi del record hanno valori identici. Non si possono avere righe uguali in tabelle SQL (la chiave distingue), ma è possibile avere duplicati tra gli attributi non chiave, con quindi risultanti problemi dati dalla proiezione di tali attributi su insiemi risultato privi di attributo chiave. Nel caso si rendesse necessario eliminare i duplicati da una data query, l'SQL offre il costrutto SELECT DISTINCT. Si nota che il costo per l'eliminazione dei duplicati è computazionalmente $O(n^2)$, e quindi da evitare quando possibile.

Valori NULL I valori NULL corrispondono solitamente a informazioni mancanti, non disponibili o comunque sconosciute. Qualsiasi condizione che coinvolge valori NULL è sempre falsa, compreso il confronto fra stessi NULL. È possibile utilizzare i valori NULL assegnandoli un valore semanticamente definito (e.g. una data non definita perché non ancora determinata, ecc...). In questo caso, l'SQL offre i costrutti IS NOT NULL e IS NULL, che restituiscono rispettivamente falso e vero posti accanto ad un valore NULL. Si nota inoltre che la chiave di un record non può essere NULL.

Gestione delle date Come per tutti i sistemi UNIX, le date in SQL vengono rappresentate come una distanza in secondi dalla mezzanotte del primo gennaio 1970.