

Le basi di dati

Informatica

Caratteristiche
generali

Prof. Luigi De Panfili

Sistema informativo e sistema informatico

Per sistema informativo si intende quel sistema che permette la disponibilità e la gestione delle informazione. Esso può essere organizzato in vario modo e da secoli è impiegato per la conservazione dei dati.

Un sistema informativo automatizzato diventa un sistema informatico, un sistema controllato da un computer e in cui l'informazione fluisce in maniera automatizzata.

Dati e informazioni

I dati costituiscono gli elementi di una base di dati. Essi da soli non hanno significato.

Interpretati in un particolare contesto e correlati tra di loro in modo opportuno, forniscono informazioni su una certa realtà modellizzata nel sistema informativo.

I dati costituiscono quindi una "risorsa" per l'organizzazione che li gestisce, un patrimonio significativo da sfruttare e proteggere da cui far scaturire, con opportune procedure, le informazioni.

Base di dati (Database)

Un database è una raccolta di elementi organizzati e gestiti da un computer

Questi elementi si dividono in due categorie :

- I metadati
- I dati

Modelli di dati

I dati vengono organizzati in base a *modelli di dati* che forniscono *meccanismi di strutturazione* per la loro organizzazione e per l'assegnazione di relazioni in modo che siano comprensibili ad un elaboratore.

I modelli di dati più utilizzati sono:

- ◆ modello relazionale;
- ◆ modello gerarchico;
- ◆ modello reticolare;
- ◆ modello ad oggetti.

Basi di dati relazionali

Tra tutti i modelli di base di dati si prende in considerazione il modello relazionale che è il più diffuso e si basa, nell'organizzazione dei dati, sulla tabella che è un modo semplice e a noi familiare di presentare i dati.

DBMS (data base management system)

Il software di gestione delle basi di dati è organizzato in un *Sistema di Gestione di Base di Dati* (DBMS).

Questo sistema ha il controllo dei dati e li rende accessibili agli utenti.

Consente di:

- definire i dati del database;
- scegliere le strutture per la memorizzazione e l'accesso ai dati;
- memorizzare, recuperare e modificare i dati, rispettando i vincoli definiti nello schema.

Linguaggi del DBMS

Il DBMS utilizza linguaggi strutturati in sottolinguaggi:

- Data Definition Language – DDL, utilizzato dal progettista della base di dati;
- Data Manipulation Language – DML, utilizzato dal gestore della base di dati (Data Base Administrator);
- Language – QL, utilizzato dall'utente per trarre informazioni dalla base di dati

Relazioni e basi di dati relazionali

Il fondamento teorico alla base del modello di dati relazionale è dato dal concetto matematico di *relazione* tra insiemi:

Dati n insiemi $D_1, \dots, D_n$ si dice relazione su di essi un sottoinsieme R del loro prodotto cartesiano

$$R \subseteq D_1 \times \dots \times D_n$$

Dove - $D_1, \dots, D_n$ sono i *domini* della relazione
- n è il *grado* della relazione

La relazione è un oggetto matematico che associa, fissata una legge, gli elementi degli n insiemi.

Relazioni e basi di dati relazionali

Ciascun elemento di R è un'ennupla ordinata di elementi ognuno appartenente ai domini $D_1, \dots, D_n$. Nell'organizzazione dei data base relazionali l'ennupla corrisponde ad una riga di una tabella (*record*) le cui colonne rappresentano gli *attributi* (detti anche *campi*).

nome	data nascita	residenza	
n1	d1	r1	
n2	d2	r2	
.....			
nk	dk	rk	

Una base di dati relazionale è un insieme di relazioni

Vincoli di integrità

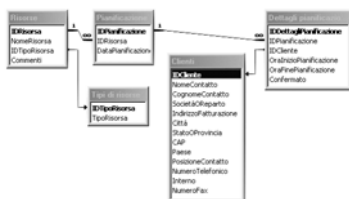
Perché le ennuple appartengano correttamente ad una base di dati devono essere verificate alcune condizioni, dette “*vincoli di integrità*”.

Dal punto di vista semantico, i vincoli si possono dividere in:

- ☐ **Vincoli di dominio**: impongono particolari valori ad un attributo;
- ☐ **Vincoli strutturali**: derivano dalle proprietà del database che vanno rispettate.

Relazioni e basi di dati relazionali

Una base di dati può essere formata da più tabelle non indipendenti. Tra di esse sono stabilite le *relazioni tra tabelle*.



Relazioni e basi di dati relazionali

Un concetto fondamentale nelle relazioni è il concetto di *chiave*.

Un insieme di attributi è detto *chiave* di una relazione se soddisfa alle seguenti condizioni:

- i valori non sono mai nulli
- identifica univocamente una ennupla
- non è ridondante, cioè ciascuno dei suoi valori è unico

La chiave può essere costituita da un solo attributo o può essere una *chiave multipla*. Una relazione può avere più chiavi distinte.

Relazioni e basi di dati relazionali

Ogni tabella è caratterizzata da una e da una sola *chiave primaria*, cioè una chiave che caratterizza la tabella ed è opportunamente scelta per metterla in relazione con le altre tabelle.

Un attributo (o insieme di attributi) di una tabella T' è *chiave esterna* (con riferimento ad un'altra tabella T) se è *chiave primaria* di T

Una relazione fra due tabelle T e T' si stabilisce tra due attributi a (chiave primaria) di T e b di T' (chiave esterna) in modo che si corrispondano.

Gli attributi a e b devono essere omogenei e contenere coppie di valori uguali.

Relazioni tra tabelle

Fra due tabelle possono esistere i seguenti tipi di relazioni:

- Relazione uno a uno
- Relazione uno a molti
- Relazione molti a molti

Normalizzazione

La definizione di una base di dati deve rispondere a criteri ben precisi onde evitare anomalie e malfunzionamenti dell'applicazione.

La *teoria della normalizzazione* ha come scopo quello di fornire metodi per progettare basi di dati senza anomalie

Dopo la definizione dello schema del database, è necessario seguire delle regole che rendano le sue tabelle in *Forma normale*.

Teoria della normalizzazione

Principali regole per costruire tabelle in *Forma normale*:

- ogni tabella deve avere una chiave primaria
- ogni campo deve contenere un solo valore
- i campi di una tabella non devono dipendere da altri campi (esclusa la chiave primaria)
- evitare le ripetizioni e la ridondanza dei dati

Progettazione di una base di dati

La progettazione si occupa della costruzione dello schema della base di dati, cioè dei metadati.

La creazione di un database può essere suddivisa in tre fasi:

1. ANALISI DEI REQUISITI: si stabilisce cosa si vuole rappresentare esattamente;
2. PROGETTO DEL SISTEMA: si progetta un sistema basandosi sulle informazioni ricavate dall'analisi dei requisiti;
3. REALIZZAZIONE DEL SISTEMA: si utilizza un DBMS (DataBase Management System).

Gestione della base di dati

Dopo la fase di progettazione e realizzazione la base di dati può essere utilizzata dall'utente finale.

Le principali attività che si possono effettuare con un database sono:

- Caricamento dei dati
- Elaborazioni occasionali
- Elaborazione del DB
- Riorganizzazione
- Transazioni applicative
- Salvataggio periodico
- Interrogazioni
- Protezione del DB

Fine presentazione

Metadati

I METADATI sono lo schema della base di dati:

- definizioni che descrivono la struttura dei dati ;
- restrizioni sui valori ammissibili dei dati (vincoli di integrità) ;
- relazioni esistenti fra gli insiemi di dati (tabelle);
- operazioni eseguibili sui dati.



Dati

I DATI sono le informazioni memorizzate nell'archivio.

Principali caratteristiche:

- sono organizzati in insiemi omogenei (tabelle), fra i quali sono definite delle relazioni;
- sono in genere molti e non possono essere gestiti tutti contemporaneamente nella memoria temporanea (RAM);
- sono permanenti: una volta creati continuano ad esistere finché non vengono esplicitamente rimossi;
- sono condivisibili: più utenti possono utilizzarli anche nello stesso momento.



Relazione uno a uno

Fra due tabelle T e T' esiste una relazione uno a uno se per ciascun valore della chiave primaria a di T esiste al più un valore uguale della chiave esterna b di T' .

La relazione uno a uno è una relazione banale : indica che due tabelle potrebbero essere una tabella unica



Relazione uno a molti

Fra due tabelle T e T' esiste una relazione uno (T) a molti (T') se per ciascun valore della chiave primaria a di T possono esistere più valori uguali della chiave esterna b di T' .

La relazione uno a molti è la relazione più comune. In una relazione uno-a-molti un record della tabella T può avere molti record corrispondenti nella tabella T' , ma un record della tabella T' non ha più di un record corrispondente nella tabella T .



Relazione molti a molti

Fra due tabelle T e T' esiste una relazione molti a molti se per ciascun valore della chiave primaria a di T esistono più valori uguali della chiave esterna b di T' e viceversa.

In una relazione molti-a-molti un record della tabella T può avere molti record corrispondenti nella tabella T' e viceversa. Questo tipo di relazione è possibile solo definendo una terza tabella, chiamata tabella di congiunzione, la cui chiave primaria consiste almeno di due campi, vale a dire le chiavi esterne di entrambe le tabelle T e T' .



Chiave primaria

Ogni tabella deve avere una chiave primaria. Questa è un campo che non contiene ripetizioni e che rappresenta ogni record nelle altre tabelle.

Se non esiste un campo con tali caratteristiche, se ne deve aggiungere un nuovo campo che svolga la funzione di rappresentare univocamente ogni record. Solitamente un campo di tipo Contatore assolve a questa funzione.

Chiave primaria



Data	Quantità	Prezzo	IdCliente	Pagamento
21/12/02	100	L. 50.000	1	contanti
21/12/02	150	L. 35.000	2	contanti
21/12/02	100	L. 35.000	5	rateale
21/12/02	150	L. 50.000	25	bonifico
30/12/02	100	L. 50.000	1	contanti

IdOrdine	Data	Quantità	Prezzo	IdCliente	Pagamento
1	21/12/02	100	L. 50.000	1	contanti
2	21/12/02	150	L. 35.000	2	contanti
3	21/12/02	100	L. 35.000	5	rateale
4	21/12/02	150	L. 50.000	25	bonifico
5	30/12/02	100	L. 50.000	1	contanti



Campo con un solo valore

PRIMA FORMA NORMALE O FORMA ATOMICA:

ogni campo deve contenere un solo valore.

I principali motivi per cui non si devono avere campi che contengono più di un valore sono:

- i database relazionali non sono in grado di fare ricerche efficienti se il campo considerato contiene più di un valore
- le relazioni non possono esistere su campi contenenti più valori

IdPersona	Cognome	Nome	Indirizzo	Città
1	Rossi	Mario	via Roma, 4	PADOVA
2	Bianchi	Marco	via Trieste, 12	TREVISO
3	Verdi	Mara	via Aurora, 32	VENEZIA

IdPersona	Cognome	Nome	Indirizzo	Nome	Numero	Città
1	Rossi	Mario	via	Roma,	4	PADOVA
2	Bianchi	Marco	via	Trieste	12	TREVISO
3	Verdi	Mara	piazza	Aurora	32	VENEZIA



Indipendenza dei campi

SECONDA FORMA NORMALE:

tutti i campi, diversi dalla chiave primaria, devono dipendere solo dal valore della chiave primaria.

I campi delle tabelle non devono dipendere da altri campi, esclusa la chiave primaria. Questo garantisce che la chiave primaria rappresenti in modo univoco ogni record della tabella.

Le tabelle non possono contenere valori calcolati. Per avere valori calcolati si devono utilizzare le query.



Ripetizione dei dati

Quando in una tabella i valori di alcuni campi sono ripetuti, è necessario creare due nuove tabelle che contengano i campi della prima. Le nuove tabelle devono poi essere messe in relazione tra di loro.

Cliente	Indirizzo	Città	DataOrdine	Importo
Rossi	Via Roma, 4	Lecce	12-lug	€ 7.500,00
Rossi	Via Roma, 4	Lecce	15-ago	€ 5.478,00
Rossi	Via Roma, 4	Lecce	17-ago	€ 4.700,00
Rossi	Via Roma, 4	Lecce	12-set	€ 7.500,00
Rossi	Via Roma, 4	Lecce	15-ott	€ 5.478,00
Rossi	Via Roma, 4	Lecce	17-ott	€ 4.700,00
Rossi	Via Roma, 4	Lecce	12-nov	€ 7.500,00
Rossi	Via Roma, 4	Lecce	15-nov	€ 5.478,00
Rossi	Via Roma, 4	Lecce	17-dic	€ 4.700,00

Clienti

Ditta	Indirizzo	Città
Rossi	Via Roma, 4	Lecce

Ordini

DataOrdine	Ordine
12-lug	7.500.000
15-ago	5.478.000
17-ago	4.700.000
12-set	7.500.000
15-ott	5.478.000



Analisi dei requisiti

L'analisi dei requisiti è lo studio preliminare che si deve fare prima di creare la base di dati, per stabilire:

- gli obiettivi dell'archivio
- i "campi di interesse"
- gli oggetti da rappresentare
- le azioni che si vorranno svolgere sull'archivio
- la durata che avrà l'archivio
- l'ambito in cui dovrà lavorare la base di dati
- tutto ciò che riguarda l'archivio stesso.



Progetto del sistema

A partire dal risultato dell'analisi dei requisiti, si struttura l'archivio, definendo:

- gli insiemi omogenei di dati (tabelle),
- i "campi di interesse" di ogni insieme (colonne delle tabelle),
- le caratteristiche di ogni campo (si tratta di un testo, di una data, di un numero, ecc.)
- le relazioni tra gli insiemi (i collegamenti tra le tabelle).

Il risultato di questa fase è lo schema della base di dati, "rappresentato" con carta e penna. Questo schema è indipendente dal DBMS che poi sarà utilizzato per realizzare il database.



Realizzazione del sistema

La realizzazione del sistema progettato è effettuata con un *DataBase Management System* (ad esempio Access).

La scelta del DBMS è determinata dalle esigenze degli utenti e dalla struttura che si vuole dare ai dati.

La realizzazione consiste nel partire dallo schema della base di dati realizzato nella fase precedente e crearlo, inserirlo, nel computer (si creano i metadati). La successiva fase sarà quella dell'inserimento dei dati.



Caricamento iniziale dei dati

E' la fase preliminare di registrazione di dati. Essa consiste nell'immissione delle ennuple definite dalla relazione che si accodano a quelle già esistenti.

Durante l'immissione vengono effettuati controlli di coerenza e compatibilità sui dati detti "*verifiche dei vincoli di integrità del database*".

E' possibile automatizzare l'immissione iniziale con tecniche basate u lettori ottici.



Elaborazione ordinaria del database

La fase di gestione ordinaria di un database consta di due classi fondamentali di operazioni:

- ✓ interrogazione, che non altera lo stato della base di dati (query);
- ✓ aggiunta, modifica, aggiornamento e cancellazione dei dati.

Per queste ultime, analogamente al caricamento iniziale, sono effettuate verifiche dei vincoli di integrità.



Transazioni applicative

Nella gestione di un database vengono definite alcune *transazioni*, cioè operazioni che alterano lo stato del database, per mezzo di procedure (programmi) che agiscono su di esso in modalità *batch* (automatica) o *interattiva*.

Questi programmi possono essere generati da appositi software associati al DBMS, o scritti in linguaggi di programmazione general purpose (*cobol*, *C++*, ecc.) nei quali è incorporato SQL (*Structured Query Language*).



Interrogazione della Base di Dati

Oltre alle interrogazioni programmate, se ne possono effettuare di estemporanee. In questi casi l'interrogazione del DB avviene attraverso linguaggi visuali di facile impiego direttamente dall'utente autorizzato alla selezione di informazioni che lo interessano. In Access ci si riferisce alle *funzioni di trova* e ai *filtri*.



Elaborazioni occasionali

Esistono alcune operazioni occasionali o periodiche, quali la visita completa di una tabella con relativa elaborazione, la *stampa di tabulati* per controlli o verifiche, la *produzione di elenchi* completi o parziali (ad esempio la verifica di un anagrafico oppure quella delle giacenze di magazzino).

Sono queste tipiche operazioni eseguite dal DBA (*Data Base Administrator*).



Riorganizzazione della Base di Dati

Viene effettuata per mantenere le caratteristiche di efficienza del DB.
Sono importanti le seguenti attività:

- riorganizzazione in relazione all'accrescimento dell'archivio;
- ampliamento delle funzionalità (nuovo indice, nuove relazioni, nuovi programmi di elaborazione, ecc.);
- compattamento del database per cancellare fisicamente gli oggetti cancellati logicamente;
- riordinamento degli archivi fisici (sort): ad esempio far coincidere l'ordine fisico con quello della visita ordinata fatta più di frequente per ridurre i tempi di accesso all'archivio.

Queste operazioni sono dispendiose in ordine al tempo di elaborazione.



Salvataggio periodico

I dati di una base di dati sono un patrimonio prezioso da conservare. Per questo è necessario effettuare periodici operazioni di *salvataggio* dei dati per salvaguardarli da possibili danneggiamenti fisici dei supporti di memorizzazione.

Il salvataggio è un'operazione essenziale e richiede anche molto tempo per cui è effettuata anche di notte.



Protezione del database

I data base sono di solito organizzati in modo da prevedere diverse classi di utenti, con *autorizzazioni di accesso* a tutto o a parte dei dati e per tutte o parte delle operazioni effettuabili sul database (ad esempio archivi disponibili in “*sola lettura*”).

Per questo il DBMS effettua i necessari controlli in fase di apertura di una seduta di lavoro (password, livello di autorizzazione) ed inibisce le operazioni non autorizzate.



Vincoli di dominio

I vincoli di dominio vanno dichiarati all'atto della definizione del database e vengono poi controllati dal DBMS in fase di esercizio del programma. Essi sono:

- *vincoli di appartenenza al tipo*: riguarda il valore dell'attributo (campo) dipendente dal tipo di dato;
- *vincoli di attributo*: è un vincolo che coinvolge il solo attributo in esame ed è espresso adoperando funzioni, operatori e costanti. Sono vincoli di attributo:
 - Controllo per fasce di valori (es. $0 < x < 100$)
 - Controllo per valori anomali
 - Controlli di appartenenza a liste di valori (es. $x \text{ IN } (a, b, c)$)
 - Controllo sui codici (es. codice fiscale)
 - Vincoli di tabella (es. paga > trattenute)



Vincoli strutturali

I vincoli strutturali sono quelli tipici di un database, cioè quelli che derivano dalla stessa teoria dei database. Essi scaturiscono in particolare dai concetti di chiave e di relazione:

- vincolo di chiave primaria: deve essere non nulla e individuare univocamente il record ;
- vincolo di unicità: ad un attributo può essere imposto di presentare molteplicità al più pari ad uno;
- vincolo di non nullità: ad un attributo può essere imposto di non assumere valore nullo;
- vincolo di integrità referenziale: se due tabelle sono legate da una relazione 1 – molti, un valore presente nel lato molti deve preesistere nel lato 1. Questo è un vincolo molto importante e va verificato dal DBMS.



Tipo di dati testo

Accetta dati lunghi fino a 255 caratteri, incluse combinazioni di testo e numeri, come nomi e indirizzi. Si usa questo tipo di dati per rappresentare numeri che non richiedono calcoli, come i numeri telefonici o i codici di avviamento postale.



Tipo di dati memo

Accetta dati lunghi fino a 65535 caratteri, incluse combinazioni di testo e numeri, commenti e lunghe descrizioni.



Tipo di dati numerico

Accetta valori numerici da usare nei calcoli.

Dimensione campo	Intervallo	Posizioni decimali
Byte	da 0 a 255	nessuna
Intero	da -32.768 a 32.767	nessuna
Intero lungo	da $-2,1 \times 10^9$ a $2,1 \times 10^9$	nessuna
Precisione singola	da $-3,4 \times 10^{-38}$ a $3,4 \times 10^{38}$	7
Precisione doppia	da $-1,8 \times 10^{-38}$ a $1,8 \times 10^{38}$	15
Decimale	da -10^{28} a 10^{28}	28



Formati di valuta e numerici

Numero generico	3456,789
Valuta	L. 3.457
Euro	€3,457
Fisso	3456,79
Standard	3.456,79
Percentuale	123,00%
Notazione scientifica	3,46E+03



Tipo di dati data/ora

Accetta solo valori di data e ora.



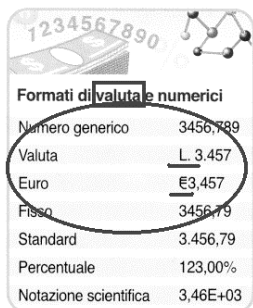
Formati di data e ora

Data generica	19/06/99 17.34.23
Data estesa	sabato 19 giugno 1994
Data breve	19-giu-99
Data in cifre	19/06/94
Ora estesa	17.34.23
Ora breve 12h	5.34 PM
Ora breve 24h	17.34



Tipo di dati valuta

Accetta solo dati per valuta.



Tipo di dati Contatore

Numera automaticamente ogni record inserito.

Viene utilizzato per il campo chiave quando nessun altro campo soddisfa alle condizione di essere non vuoto e individuare univocamente i record.



Tipo di dati Si/No

Accetta solo uno dei due valori nei formati:

Si/No

Vero/Falso

On/Off



Tipo di dati OLE

Accetta solo oggetti OLE. Un oggetto OLE è creato con un'altra applicazione: può essere un documento, in foglio elettronico, un'immagine, ecc.



Tipo di dati collegamento ipertestuale

Accetta collegamenti ipertestuali da selezionare per accedere ad un documento o a una pagina Web..



Tipo di dati ricerca guidata

Avvia ricerca guidata per creare un elenco di voci dal quale scegliere mentre si inseriscono i dati.

E' utile creare un tipo dati "ricerca guidata" se si usano spesso gli stessi valori da immettere in un campo.