

II. Le basi di dati

II.1 Gli archivi

La necessità di conservare dati e informazioni in modo permanente, poiché potranno essere utili in momenti successivi, è un problema molto evidente nel mondo moderno e riguarda un gran numero di persone. Il discorso è analogo sia per dati di tipo personale (rubrica telefonica), sia ai documenti utili per la vita di uno stato, di un ente, di un'impresa.

In un'azienda, l'esecuzione delle quotidiane attività sia amministrative che operative, la definizione e la scelta delle politiche commerciali, di quelle finanziarie e di quelle relative al personale, sono strettamente legate all'elaborazione di insiemi di dati che vengono denominati **archivi**.

L'elenco telefonico è un esempio molto comune di archivio di dati; in esso le informazioni riguardano gli abbonati al telefono di una provincia: per ogni abbonato sono riportati nell'ordine, generalità, indirizzo e numero di telefono, e tutte queste informazioni sono stampate su fogli di carta.

In generale possiamo definire un archivio come un insieme organizzato di informazioni caratterizzate da alcuni aspetti fondamentali:

- tra esse esiste un nesso logico (Ex. abbonati di una stessa provincia)
- sono rappresentate secondo un formato che ne rende possibile l'interpretazione (disposizione delle informazioni in righe nello stesso ordine per tutti gli abbonati)
- sono registrate su un supporto in cui è possibile scrivere e rileggere informazioni anche a distanza di tempo (elenco ristampato e aggiornato annualmente);
- sono organizzate in modo da permettere una facile consultazione (ordine alfabetico per comuni e cognomi).

II.2 Operazioni sugli archivi

In un archivio le informazioni, in genere, sono raggruppate secondo un'unità logica: nel caso dell'elenco telefonico i dati relativi ad ogni abbonato, in un archivio notarile i dati contenuti nell'incartamento relativo ad un cliente.

Questi insiemi di informazioni logicamente organizzate e riferite ad un unico soggetto vengono chiamati con il termine **record** (che in inglese significa *registrazione*). Le singole informazioni che compongono il record si chiamano **campi**; l'elenco dei campi viene detto **tracciato del record**.

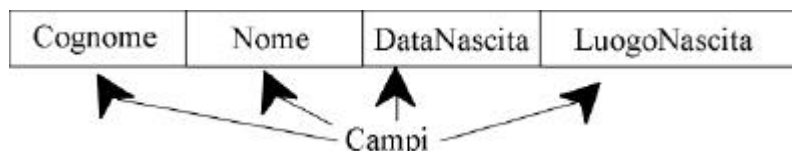


Figura 1 Tracciato del record

Per esempio il tracciato del record per l'archivio costituito dall'anagrafe, può essere descritto con i campi di figura 1:

- cognome;
- nome;
- data di nascita;
- luogo di nascita.

La creazione di un archivio richiede la definizione preliminare delle seguenti specifiche:

- *nome dell'archivio*, che lo identifica e serve a ricordarne il contenuto; per esempio “archivio fornitori” oppure “archivio anagrafico”
- *tracciato record*, in altre parole quali informazioni compongono il record
- *supporto da usare per archiviare i dati* (fogli di carta, dischi magnetici, dischi ottici)
- *dimensione massima dell'archivio*: per esempio il numero massimo di scaffali occupati in un archivio cartaceo
- *il modo in cui i dati sono strutturati e collegati fra loro*, cioè l'organizzazione dell'archivio

La decisione su queste specifiche fa parte dell'attività di analisi necessaria per la realizzazione di un'applicazione informatica e deve precedere la fase di realizzazione fisica dell'archivio. Tale fase può consistere nella sistemazione delle informazioni già esistenti in un archivio ben organizzato oppure nella predisposizione del supporto per la registrazione delle nuove informazioni che verranno successivamente inserite.

Prima di essere usato, un archivio deve essere creato. Inoltre la sua gestione viene realizzata mediante le seguenti operazioni che possono essere classificate in due categorie:

a) operazioni di manipolazione

- **l'inserimento** di nuovi dati dopo che l'archivio è stato creato: un esempio di aggiunta di nuovi dati è l'inserimento di un nuovo abbonato nell'elenco telefonico;
- **la modifica o aggiornamento dei dati** già presenti nell'archivio, ma non più corrispondenti ai dati reali: se un utente telefonico trasferisce la propria residenza, rimanendo nella medesima città, nella successiva edizione dell'elenco telefonico viene stampato il nuovo indirizzo;
- **la cancellazione** di informazioni che non si vogliono più conservare perché non esiste più un nesso logico rispetto alle altre informazioni presenti nell'archivio: è il caso di un abbonato al telefono che si trasferisce in un'altra città e quindi deve essere tolto dall'elenco telefonico.

b) operazioni di interrogazione

- **la consultazione o interrogazione**, cioè il reperimento all'interno dell'archivio delle informazioni necessarie per l'elaborazione desiderata. È ovvio infatti che la possibilità di reperire le informazioni volute è il motivo principale dell'esistenza di un archivio. Ad esempio l'archivio della contabilità di un'azienda viene interrogato per conoscere quali sono le fatture emesse che non sono ancora state pagate dai clienti.

Il risultato dell'interrogazione può essere rappresentato mediante una **visualizzazione** sul monitor del computer oppure con una **stampa** su carta. Gli output su stampante vengono correntemente indicati con il termine **report**.

II.3 File e periferiche

Per ragioni di velocità, nella ricerca e nell'elaborazione, e di spazio, nella memorizzazione dei dati, si è passati negli anni da archivi registrati su supporti cartacei contenuti in armadi (schedari), adatti al reperimento manuale da parte dell'uomo, a supporti ideati per essere trattati in modo automatico dai computer.

Gli archivi memorizzati su tali supporti vengono detti **file**, perché in inglese la parola archivio viene tradotta con “file”, anche se in informatica questo termine serve ad indicare in modo generico qualsiasi informazione che può essere registrata sui supporti di memoria, come un testo, un programma, un grafico, ...

Nelle applicazioni per computer, un archivio può contenere qualsiasi tipo di informazione, non solo di tipo testuale, ma anche di tipo multimediale (immagini, video, audio).

Nelle applicazioni informatiche, comunque, gli archivi sono, nella maggior parte dei casi, costituiti da insiemi di record omogenei, nel senso che ciascun archivio possiede un tracciato predefinito e uguale per tutti i record in esso contenuti (**file di record**).

I dispositivi che permettono di leggere e scrivere le informazioni contenute nei file, vengono denominati **unità periferiche di memoria**. Esse sono esterne al calcolatore.

Ogni trasferimento dati dalla periferica verso la memoria centrale rappresenta un'operazione di **input**, come pure ogni trasferimento dati dalla memoria centrale verso la periferica rappresenta un'operazione di **output** (si osservi figura 2). Ogni operazione di I/O (*input/output*) riguarda non un singolo carattere, bensì un insieme di caratteri, detto **blocco**. Il blocco è l'unità fisica di memorizzazione dei dati sulla memoria di massa.

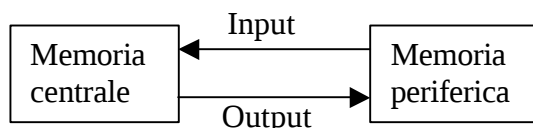


Figura 2 Operazioni di I/O

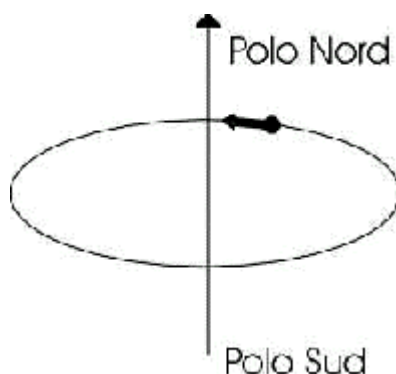
II.4 I dischi

Per la registrazione degli archivi si usano supporti di grandi capacità, che prendono il nome di **memorie di massa** perché possono contenere notevoli quantità di dati. Si chiamano anche memorie ausiliarie perché costituiscono in un certo senso un'estensione della memoria centrale di un computer e consentono, a differenza della RAM (che è volatile), la permanenza delle registrazioni nel tempo.

Sulle memorie di massa possono essere effettuate operazioni di lettura (input) o di scrittura (output): esse sono quindi **unità di I/O**.

II.4.1 Dispositivi magnetici: i dischi

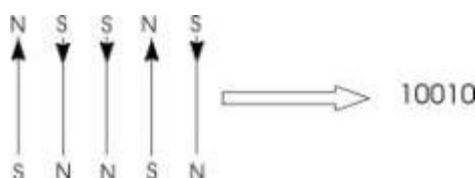
Il magnetismo è la proprietà di alcune sostanze di orientare frammenti di ferro. Diciamo brevemente cosa si intende per campo elettrico e per campo magnetico. Un campo elettrico è un sistema di forze esercitato da una o più cariche elettriche capace di influenzare altre cariche nelle vicinanze, così anche un campo magnetico è un sistema di forze capace di esercitare la sua influenza in una certa regione di spazio.



Vi è una stretta connessione tra le proprietà elettriche e le proprietà magnetiche: la variazione di un campo elettrico produce un campo magnetico, la variazione di un campo magnetico produce un campo elettrico. In termini qualitativi queste sono proprio le due leggi fondamentali dell'elettromagnetismo.

Un campo magnetico è caratterizzato da una direzione. Negli atomi la rotazione degli elettroni attorno al nucleo genera dei campi magnetici, ciascuno con una certa direzione nello spazio, ovvero con un polo sud ed un polo nord. La direzione dei campi dipende dal verso di percorrenza degli elettroni.

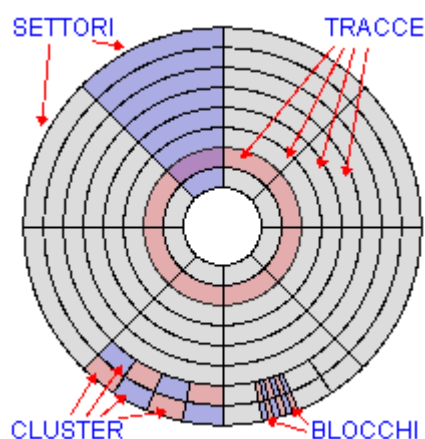
In natura è possibile costruire piccoli magneti, ovvero dispositivi magnetizzati, usando dei materiali detti ferromagnetici. Nei dispositivi magnetici digitali vi sono dei piccoli magneti che vengono fatti orientare in una direzione od in quella opposta (cioè facendo ruotare le particelle in una delle due direzioni) ottenendo così due diversi orientamenti interpretabili come segnale binario, in altri termini a ciascuno si può associare un bit di informazione. Una magnetizzazione verso l'alto (come in figura) darà un segnale binario 1, una magnetizzazione verso il basso darà un segnale binario 0.



Dato che non è possibile costruire interruttori magnetici, cioè sistemi magnetici che producano la variazione di un segnale magnetico, i segnali binari magnetici non vengono utilizzati per l'elaborazione dell'informazione digitale: è necessario trasformare un segnale digitale magnetico in uno elettronico prima di poterlo elaborare.

I segnali binari magnetici vengono utilizzati solo per la memorizzazione, infatti è possibile ottenere la magnetizzazione di zone molto piccole e conservare la direzione dei campi magnetici (nei materiali ferromagnetici) per molti anni.

Nei moderni calcolatori l'informazione digitale magnetica viene depositata su dischi in continua rotazione. Su questi dischi la testina di lettura e di scrittura è libera di muoversi a distanze diverse dal centro del disco percorrendo cerchi più o meno ampi detti tracce. Su ogni traccia di un disco vengono memorizzati moltissimi bit, è quindi necessario suddividere una traccia in tante parti chiamate settori.



Secondo la convenzione normalmente utilizzata, la faccia superiore di un disco (o di un elemento in una pila di dischi dell'hard), viene considerata come la faccia 0, mentre la inferiore è la faccia 1. In un hard disk la faccia 0 del primo disco della pila e la 1 dell'ultimo, non vengono generalmente utilizzate per la collocazione dei dati.

La formattazione è un'operazione che rende utilizzabile un disco, in quanto vi inserisce gli elementi per individuare le tracce, i settori ed i dati.

Durante la formattazione vengono definite 4 zone:

1. Zona di avviamento (boot record): occupa un settore, precisamente il settore 1 della traccia 0; reca le informazioni per la individuazione del sistema operativo dell'elaboratore che gestisce la memoria.

2. FAT (File Allocation Table): contiene gli elementi della formattazione (numero tracce e settori) e l'indicazione della distribuzione dei dati nel disco (settore d'inizio), individuati dal nome di raggruppamento dei dati medesimi (nome del file).

3. Directory: rappresenta l'indice dei dati contenuti sul disco e li individua in funzione del nome dato al file, della lunghezza del file,

4. Spazio dati: i settori non dedicati alle tre zone precedenti, vengono riservati ai dati.

Il settore é un blocco di bit che viene letto e scritto contemporaneamente. Infatti su di un disco non conviene leggere il singolo bit ma il gruppo di bit compreso in un settore. Questa tecnica condiziona, tra l'altro, l'intero funzionamento del calcolatore.

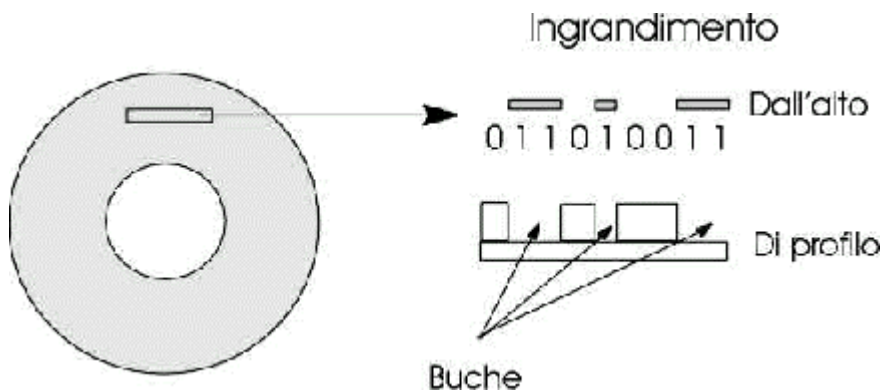
II.4.2 Ottica: laser e dischi ottici

Una delle forme più semplici di rappresentazione binaria é quella ottica.

La luce é un'onda elettromagnetica senza peso ma può essere considerata comunque composta da particelle (quanti di luce o fotoni).

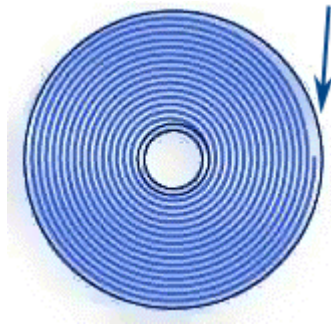
Naturalmente la luce non può essere facilmente conservata, cioè conservata in un luogo, ma può essere usata per cambiare lo stato di un altro dispositivo binario di tipo elettronico. Sappiamo infatti che la luce può essere assorbita da un elettrone. Un elettrone la cui energia aumenta per effetto del fotone assorbito può rendersi libero del legame con i protoni del nucleo ed essere quindi libero di muoversi e produrre un flusso di corrente.

I dispositivi ottici per eccellenza sono oggi dei dischi sui quali l'elemento binario é di fatto una piccola buca (denominata pit) scavata da un laser nella stagnola racchiusa nella plastica del disco.



Se esiste una buca lo stato dell'elemento binario é 1 se la buca non é stata fatta lo stato é 0.

Naturalmente nel CD le buche sono disposte lungo un percorso a spirale che viene letto da un fascio luminoso molto sottile prodotto da un laser a bassa potenza.



Durante la riproduzione il fascio luminoso percorre la spirale se incontra una buca non viene riflesso, se invece la buca non c'è il fascio viene riflesso. In questo modo la presenza o meno della riflessione rappresenta un sistema binario il cui stato può essere rilevato da un sensore di luce che trasforma il segnale luminoso binario in un sistema elettrico binario.

Anche nel caso di segnali ottici si ha dunque la trasformazione in segnali elettronici prima di poter essere elaborati. Sono quindi difficilmente costruibili calcolatori ottici.

II.5 Archivi e database

La registrazione dei dati su un supporto di memoria di massa deve essere fatta in modo da organizzare i dati nel migliore modo possibile e consentire il ritrovamento dei dati in modo sicuro e in tempi brevi: si deve cioè ottimizzare l'uso dello spazio disponibile sul supporto, in relazione alle sue capacità, e ridurre i tempi di accesso e di ritrovamento dei dati contenuti nella memoria di massa.

La progettazione e l'organizzazione degli archivi deve avere, inoltre due obiettivi fondamentali:

- l'integrazione degli archivi, per evitare ridondanze nelle registrazioni;
- l'indipendenza dallo schema logico dei dati dalle implementazioni software, dalle applicazioni che usano quei dati e dal modo con il quale i dati sono fisicamente registrati su memoria di massa.

Tutto questo rende naturalmente più facile il lavoro di **manutenzione degli archivi** nel tempo. Le attività più frequenti nella manutenzione di archivi sono:

- eliminazione degli errori presenti negli archivi o nel software che li gestiscono;
- necessità di ampliare gli archivi in corrispondenza dell'aumento delle informazioni per un'azienda che cresce nelle sue dimensioni;
- adeguamento dei dati a disposizione del software applicativo a nuove norme legislative o fiscali;
- conversione e adeguamento degli archivi per nuove piattaforme hardware, più potenti nell'elaborazione, oppure di maggiore capacità e prestazioni nelle unità di memorizzazione.

In generale gli archivi possono essere organizzati in tre differenti maniere:

1. **Organizzazione sequenziale:** i record sono registrati uno di seguito all'altro. In fase di lettura i record possono essere ritrovati scorrendo tutti i record del file dal primo fino all'ultimo, nello stesso ordine con il quale sono stati registrati. Questo tipo di organizzazione, semplice da gestire, consente l'uso di record aventi anche lunghezza diversa uno dall'altro (in questo caso il record è preceduto da un primo campo contenente la lunghezza del record). Tuttavia è evidente che tale organizzazione degli archivi presenta dei limiti in fase di ritrovamento dei dati quando il numero dei record diventa elevato. I sistemi operativi trattano l'insieme dei dati inviati ad una stampante come un particolare file ad organizzazione sequenziale, sul quale si può solamente scrivere.

2. **Organizzazione ad accesso diretto.** Se si utilizzano record di lunghezza fissa, si può pensare di associare ad ogni record un numero d'ordine e ritrovare facilmente all'interno del file il numero del primo carattere del record richiesto, considerando il file come una sequenza di byte. In generale se la lunghezza del record è uguale a LR e il numero del record è NR , ci si dovrà posizionare sul carattere corrispondente a : $(NR-1) \times LR + 1$. Ad esempio, se ogni record ha lunghezza uguale a 130, per leggere o scrivere il record 82, occorre posizionarsi sul carattere del file corrispondente al numero $(82-1) \times 130 + 1 = 10531$.

In definitiva, in un archivio ad organizzazione diretta, ciascun record viene identificato dalla posizione che esso occupa all'interno del file. Tale organizzazione risulta essere convenientemente utilizzata negli archivi per i quali i record vengono identificati attraverso un numero d'ordine progressivo.

3. **Organizzazione ad indici.** Nelle applicazioni informatiche molto spesso le entità che devono essere rappresentate con i record degli archivi vengono identificate attraverso un elemento caratteristico denominato **chiave**. Quest'ultimo non deve essere necessariamente un numero, ma è una variabile alfanumerica, per esempio la matricola di un dipendente di un'azienda, il codice di un articolo, il codice del cliente. L'organizzazione dell'archivio in questo caso, per facilitare il ritrovamento dei dati, deve essere simile a quella di un libro tecnico dotato di indice analitico: in tale indice sono elencate in ordine alfabetico le parole chiave, che richiamano gli argomenti e i concetti trattati nel testo, affiancate ciascuna dalla pagina corrispondente del libro.

La stessa cosa avviene nei file con organizzazione **ad indici** per i quali accanto alla zona dove sono registrati i record viene gestita una tabella delle chiavi: la ricerca del record avviene leggendo le chiavi e non i record come avviene nei file sequenziali. Ad ogni chiave è associato il puntatore al record, cioè la posizione che il record occupa all'interno del file.

La chiave può essere:

primaria, quando è necessariamente univoca e costituisce il criterio fondamentale per ritrovare un record;

secondaria anche non univoca.

Con il termine **basi di dati** si indicano in informatica gli archivi di dati, organizzati in modo integrato attraverso tecniche di *modellizzazione dei dati* e gestiti sulle memorie di massa dei computer attraverso appositi software, con l'obiettivo di raggiungere una grande efficienza nel trattamento e nel ritrovamento dati.

A grandi linee, si può affermare che il database è una collezione di archivi di dati ben organizzati e ben strutturati, in modo che possano costituire una base di lavoro per utenti diversi con programmi diversi. Per esempio i dati relativi agli articoli del magazzino di un'azienda possono essere utilizzati dal programma che stampa le fatture, oppure dal programma che stampa i listini di magazzino.

I prodotti software per la gestione di database vengono indicati con il termine di **DBMS (DataBase Management System)**. Occorre anche sottolineare l'importanza che l'uso dei database assume nella gestione degli archivi di dati, favorendo l'utente del computer nel suo modo di vedere i dati, e liberandolo dagli aspetti riguardanti la collocazione fisica delle registrazioni sui supporti magnetici degli archivi. In definitiva, i DBMS consentono all'utente di collocarsi in una posizione più lontana dall'hardware, dalle memorie di massa e dal Sistema Operativo e più vicina all'applicazione che utilizza i dati contenuti negli archivi.

II.6 Il modello relazionale

L'informatica offre metodologie e strumenti per la costruzione di **modelli** di situazioni reali, con l'obiettivo di lavorare sul modello per riprodurre o predire l'evoluzione della situazione oggetto di studio.

I modelli ricorrono ampiamente nella tecnologia e in ogni campo che richiede un'attività di progettazione. Essi permettono di riprodurre le caratteristiche essenziali di fenomeni reali, omettendo quei dettagli che, ai fini dello studio che ci si prefigge, costituirebbero un'inutile complicazione.

Nel campo delle scienze si impiegano modelli di natura diversa, che per semplicità qui si raggruppano in due categorie:

- **modelli in scala**, cioè la riproduzione in dimensioni ridotte di macchine, edifici, strutture che hanno grandi dimensioni e che vanno studiate, prima di passare alla loro realizzazione in scala reale, in condizioni il più possibile simili a quelle che esisteranno nella realtà;
- **modelli astratti** (matematici o simbolici), definiti come “la rappresentazione formale di idee e conoscenze relative a un fenomeno”.

Questa definizione evidenzia tre aspetti fondamentali di un modello astratto:

- a) è la rappresentazione di alcuni fatti di un fenomeno;
- b) la rappresentazione è data con un linguaggio formale;
- c) il modello è il risultato di un processo di interpretazione di un fenomeno, guidato dalle idee e conoscenze già possedute dal soggetto che interpreta.

L'informatica consente di costruire modelli astratti diversi:

- modelli per l'analisi del problema;
- modelli per la progettazione della soluzione;
- modelli per la realizzazione del progetto.

La differenza fra le varie categorie di modelli sta nel diverso livello di astrazione a cui si opera, e quindi nei tipi di fatti che si prendono in considerazione: per costruire il modello da utilizzare a regime si usano linguaggi formali che consentono di descrivere la soluzione del problema in tutti i dettagli richiesti da uno specifico sistema di elaborazione; per costruire invece il modello per la progettazione, e per l'analisi, si usano linguaggi che consentono di ignorare dei dettagli del sistema di elaborazione, per concentrarsi su alcune caratteristiche essenziali al fine di una più agevole comprensione degli aspetti cruciali del problema.

Un'analogia con l'atlante stradale aiuta a comprendere il ruolo dei diversi livelli di astrazione usati nella costruzione di un modello, in funzione del problema da risolvere: quando si pianifica un viaggio, uno sguardo al quadro d'unione delle singole tavole con la cartografia stradale consente di stabilire l'itinerario e le distanze da percorrere; un esame delle singole tavole consente di programmare deviazioni per strade panoramiche e le vie di accesso alle città da visitare; infine, l'esame delle piante delle città consente di fissare dettagliati itinerari di visita.

I principali modelli per le basi di dati sono: il **modello gerarchico**, il **modello reticolare** e il **modello relazionale**.

Prima di passare alla descrizione di questi modelli, occorre introdurre i concetti di *entità*, *associazioni* e *attributi*.

L'**entità** è un oggetto (concreto o astratto) che ha un significato anche quando viene considerato in modo isolato ed è di interesse per la realtà che si vuole modellare. Esempi di entità sono: una persona, un modello di automobile, una prova sostenuta da uno studente.

L'**associazione** è un legame che stabilisce un'interazione tra le entità. (Ex. Auto posseduta da una persona).

Le proprietà delle entità e delle associazioni vengono descritte attraverso gli **attributi**. Esempi di attributi per l'entità automobile sono: *produttore*, *cilindrata*.

Il **modello gerarchico** (utilizzato dai primi DBMS commerciali, il capostipite dei quali è l'IMS sviluppato dalla IBM nel 1969), utilizza uno schema del database vincolato a essere costituito da strutture ad albero. Tale modello dei dati, ormai superato, non era stato sottoposto ad alcuna forma di standardizzazione: quello dell'IMS era stato adottato come base comune da diverse case produttrici, ma le evoluzioni che ne seguirono erano anche molto diverse. Per accedere ai dati posti a un certo livello dell'albero, si deve iniziare la ricerca dalla radice procedendo verso il basso e utilizzando per la navigazione i puntatori unidirezionali.

Il **modello reticolare**, fu elaborato intorno ai primi degli anni Settanta segnando il superamento della strutturazione necessariamente gerarchica dello schema logico. Questo modello nasce da una proposta per la standardizzazione della definizione e l'utilizzo dei database, elaborata nel 1971 dal gruppo di lavoro DBTG (Data Base Task Group) del comitato CODASYL. Sebbene molti sistemi (IDS/II della Honeywell, DMS-1100 dell'Univac, DBMS- 10 della Digital) abbiano adottato gli elementi fondamentali dello standard, problemi di tipo commerciale e di tipo tecnologico, quali la complessità del linguaggio di accesso ai dati e i requisiti hardware, hanno fatto sì che non sia stata sviluppata nessuna implementazione rispondente totalmente alle regole DBTG-CODASYL.

Tuttavia la proposta ha avuto il ruolo fondamentale, se non di standardizzare totalmente un modello di dati, almeno di introdurre alcuni concetti basilari della teoria dei database, quali quelli di schema e sottoschema e di indipendenza dei dati dalle applicazioni. Tale modello è un'estensione del precedente modello gerarchico, e permette di rappresentare tutte le possibili connessioni.

Negli stessi anni in cui veniva sviluppato il modello reticolare iniziò la definizione del modello relazionale; ci vollero tuttavia quasi altri dieci anni perché tale modello potesse essere utilizzato per l'implementazione di database commerciali.

Il **modello relazionale**. Storicamente i DBMS relazionali (RDBMS, Relational Database Management System) furono i primi a essere basati su un modello dei dati precedentemente formalizzato. Tale modello venne sviluppato nel 1970 da Edward Codd presso l'IBM San José Research Laboratory. L'approccio adottato è basato sulla teoria matematica delle relazioni tra insiemi. Esso prescinde da specifiche tecniche realizzative e da considerazioni relative alla facilità di realizzazione e all'efficienza delle prestazioni.

Le prime implementazioni - Oracle presso la Relational Software Inc., poi Oracle Corporation, System R e successivamente SQL/DS e DB2 presso la IBM, Ingres presso l'Università di Berkeley - risalgono agli anni compresi tra il 1977 e il 1983. L'indipendenza del modello da specifiche tecnologie informatiche, se da una parte ne ha frenato inizialmente l'applicazione, dall'altra ha costituito il suo punto di forza e ne ha permesso successivamente un'enorme diffusione su tutte le piattaforme e architetture.

L'uso di un modello matematico dei dati consente l'applicazione di linguaggi e metodologie formali per l'accesso ai dati. In particolare, le due metodologie su cui si basano i linguaggi per l'accesso ai dati di un database relazionale sono l'algebra relazionale e il calcolo relazionale.

La rappresentazione dei dati nel modello logico relazionale è basata sul concetto di relazione; questa va intesa in termini algebrici, e non va confusa con le relazioni tra i dati del modello concettuale. Il concetto di relazione algebrica è così definito:

- dati gli insiemi finiti di valori $D_1, D_2, \dots, D_n$, una relazione K tra questi n insiemi è un insieme di n -uple $\langle d_1, d_2, \dots, d_n \rangle$, dove d_1 appartiene a D_1 , d_2 appartiene a D_2 , .. d_n appartiene a D_n ;
- $D_1, D_2, \dots, D_n$ sono detti domini, i loro nomi sono detti attributi, n è detto grado della relazione R .

Per esempio, considerati i due domini

CODICE CATEGORIA={01, 02, 03, 04}

e

NOME CATEGORIA={Autovettura, Furgone, Motociclo, Rimorchio }

una possibile relazione tra essi è

CATEGORIE=<(01,Autovettura), (02,Rimorchio), (03,Motociclo), (04, Furgone)>

Per comodità di rappresentazione è conveniente descrivere una relazione algebrica in forma di tabella, e in base a questa rappresentazione, spesso i concetti di relazioni algebriche, di n -uple e di attributi vengono indicati con i termini più familiari di tabelle, di righe e di colonne.

Indicheremo quindi con il termine "tabella" il concetto di "relazione algebrica", mentre il termine "relazione" indicherà un'associazione tra i dati. Dalla definizione di tabella come insieme derivano due conseguenze fondamentali:

- in una tabella non possono esistere due righe uguali;
- l'ordine tra le righe di una tabella non è significativo.

Dalla prima osservazione consegue che è possibile, e anche necessario, individuare in ciascuna tabella un insieme di attributi (colonne) in base ai quali vengono identificate le singole righe, che rappresentano quindi una chiave di accesso univoca alle informazioni contenute nella tabella stessa. Questo insieme di colonne, che va definito in fase di creazione dello schema logico, è detto chiave primaria (primary key) della tabella.

Per la creazione di uno schema logico relazionale, partendo da un determinato schema concettuale, è necessario applicare le regole seguenti.

1. le entità (entity set) dello schema concettuale divengono tabelle nello schema logico.
2. le relazioni (relation set) dello schema concettuale vengono rappresentate nello schema logico facendo uso delle cosiddette chiavi esterne. Una chiave esterna (foreign key) di una tabella è un insieme di attributi che corrispondono a quelli che costituiscono la chiave primaria di un'altra tabella, e stabiliscono quindi un riferimento tra le righe delle due tabelle.

Si possono stabilire tre tipi di relazioni individuate nella definizione dello schema concettuale:

- uno a uno: si tratta di relazioni tra elementi che hanno una corrispondenza univoca: ad un elemento di una tabella ne corrisponde uno soltanto in un'altra e viceversa;
- uno a molti: sono relazioni che si stabiliscono tra un record di una tabella e più record di un'altra tabella, ma non il contrario;
- molti a molti: un record può essere relazionato a più di un record di un'altra tabella e viceversa; questo tipo di relazione è normalmente definita tramite una terza tabella che costituisce un "ponte" tra le due da relazionare.

II. 7 Il modello a oggetti

I **database orientati agli oggetti** (OODB Object Oriented DataBase) sono database che gestiscono oggetti e classi. Per **classe** si intende uno schema di creazione che definisca i possibili stati e i comportamenti di **oggetti** appartenenti a tali classi. Sono diversi dai tradizionali database perché accettano sottoclassi di oggetti, ogni oggetto ha una propria identità (al contrario dei database tradizionali orientati al valore degli attributi) e utilizzano i concetti tipici della programmazione ad oggetti, quali il *metodo* e l'*ereditarietà*. Per metodo si intende l'invocazione di operazioni prestabilite e note agli utenti in modo astratto. L'ereditarietà stabilisce la presenza di relazioni tra classi di oggetti.

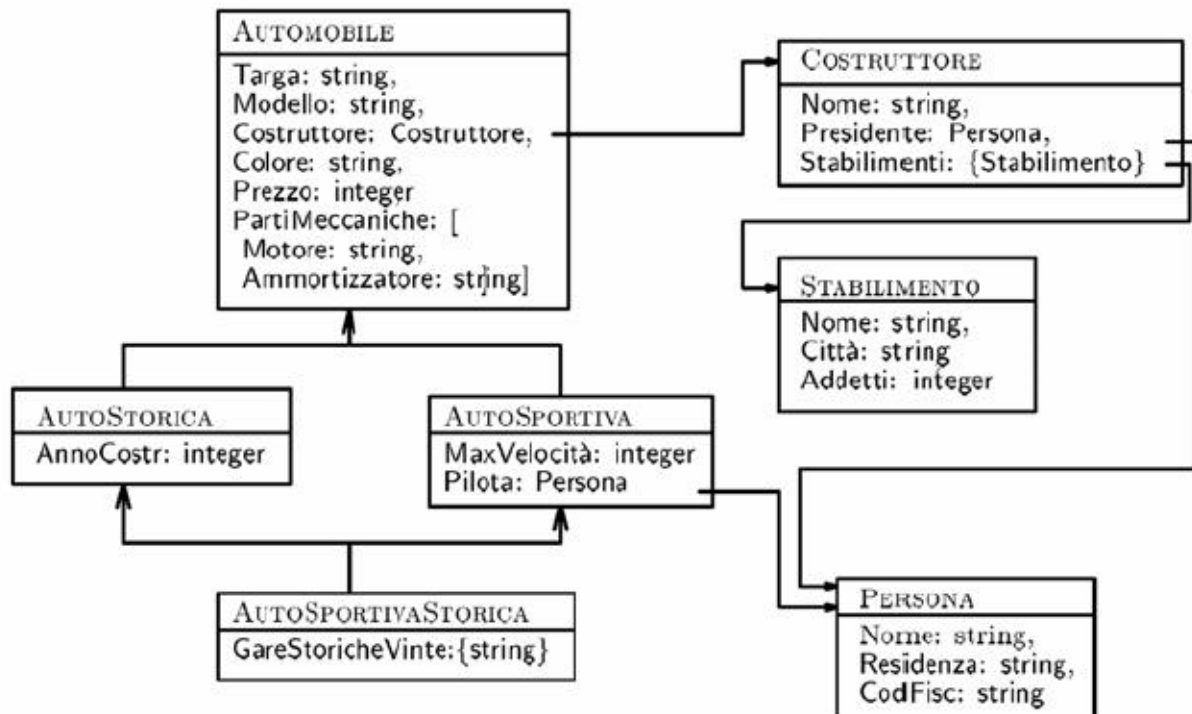
Di solito, i programmi rappresentano l'aspetto elaborativo di un'applicazione e i database l'aspetto informativo: i dati vengono organizzati nei database a partire dal modello dei dati in modo indipendente dai programmi.

La soluzione dei database object-oriented favorisce un avvicinamento tra i programmi applicativi ed i database, con la possibilità di usare tipi di dati definiti dall'utente (cosa che non è possibile nel database relazionale) e di poter associare ai dati routine di codice che rappresentano le modalità di accesso ai dati. In questo modo viene applicato il concetto di oggetto, con i dati e i metodi, al database.

Gli OODB sono stati introdotti nel tentativo di superare alcune caratteristiche dei database relazionali che apparivano inadeguate per alcuni ambiti applicativi, quali la medicina, l'elaborazione dei documenti, l'ingegneria meccanica, la multimedialità. In queste e in molte altre applicazioni esiste l'esigenza di gestire dati complessi in modo vicino alla realtà, non solo testi, ma anche grafici, audio, video, documenti e risorse Internet.

Nel modello dei dati a oggetti le entità del dominio vengono modellate con oggetti e relazioni fra oggetti. Un **ODBMS** (Object DataBase Management System) offre le funzioni necessarie a rendere persistenti collezioni di oggetti e le relazioni fra di loro. La naturalezza nel rappresentare dati complessi e l'abilità ad accedervi in modo efficiente è il principale punto a favore di un ODBMS rispetto a un DBMS relazionale.

A differenza che nel modello relazionale, nel modello a oggetti gli attributi di un oggetto possono essere di qualunque tipo: in particolare, un oggetto può contenere altri oggetti. La figura sottostante rappresenta un esempio di schema di un OODB che chiarisce il concetto di *ereditarietà*.



L'ereditarietà si riscontra nella presenza di più classi (Automobile, Costruttore, Stabilimento, ...) che sono in relazione tra loro. Per esempio la classe automobile può avere un certo numero di oggetti, i cui attributi sono *Targa*, *Modello*, *Colore*, ... Tali oggetti contengono altri oggetti appartenenti alla classe *Costruttore*, caratterizzata da membri del tipo *Nome*, *Presidente*, ... In definitiva, il modello object-oriented consente di utilizzare tecnologie più avanzate insieme alle prestazioni tipiche dei linguaggi utilizzati nella programmazione ad oggetti, come il C++.