

03. Basi di dati e sistemi di gestione delle basi di dati

database

basi di dati e sistemi di gestione delle basi di dati

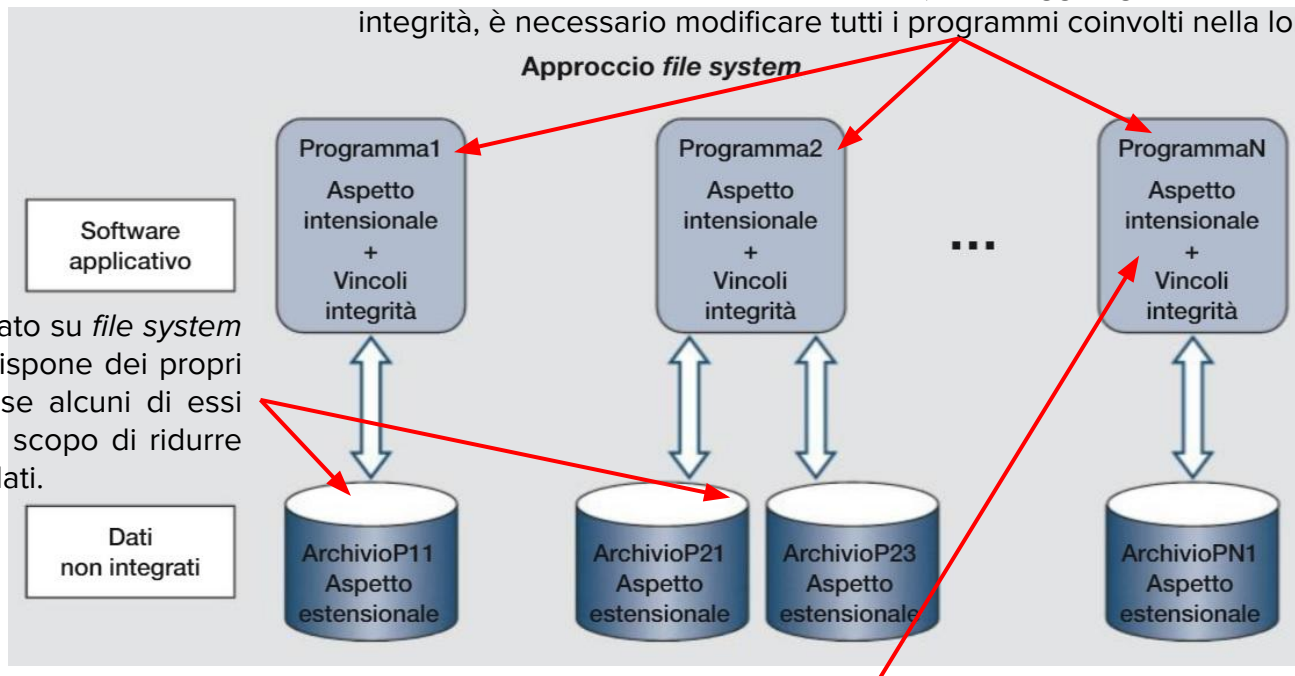
Un **DBMS** (*DataBase Management System* o Sistema di Gestione delle Basi di Dati) è un sistema software in grado di gestire grandi collezioni di dati **integrati**, **condivisi** e **persistenti** assicurando loro affidabilità e privatezza. Come ogni prodotto informatico un DBMS deve essere efficiente ed efficace.

Una **base di dati** (DB o *database*) è una collezione di dati gestita da un DBMS (un DBMS può gestire diverse basi di dati distinte).

L'introduzione di questi sistemi ha portato ad un approccio differente, rispetto a quello precedente basato su file system, nella gestione dei database, che ha introdotto una rivoluzione nel modo di «pensare» al problema dell'organizzazione dei dati.

basi di dati e sistemi di gestione delle basi di dati

Dovendo modificare la struttura dei dati (ad es. aggiungere nuovi campi) o le regole di integrità, è necessario modificare tutti i programmi coinvolti nella loro elaborazione



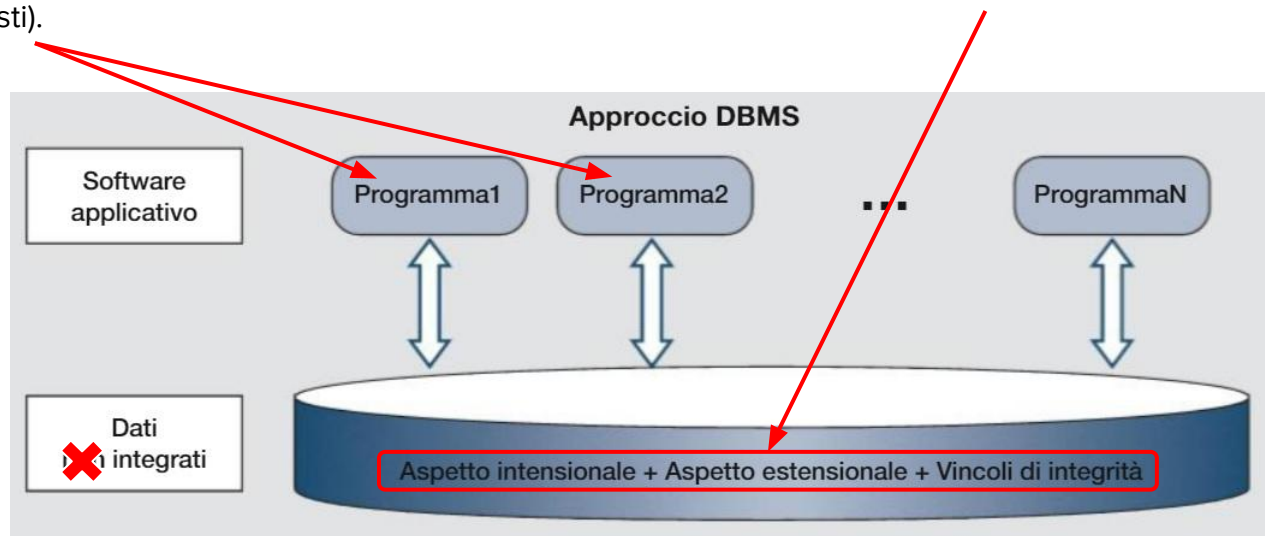
Nell'approccio basato su *file system* ogni programma dispone dei propri file di dati anche se alcuni di essi sono condivisi allo scopo di ridurre la ridondanza dei dati.

Il problema risiede nel fatto che **la definizione degli schemi e delle strutture di memorizzazione dei dati (*aspetto intensionale*) nonché le regole di integrità che i dati stessi devono soddisfare sono implementati dal codice dei programmi che gestiscono i file.**

basi di dati e sistemi di gestione delle basi di dati

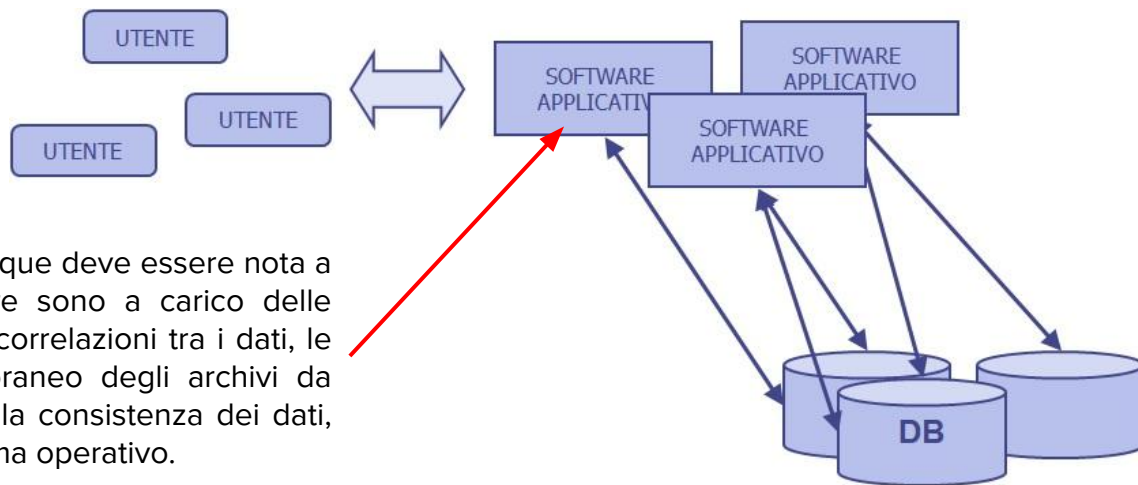
I programmi non prevedono al loro interno la definizione della struttura dei dati, ma fanno riferimento ai soli nomi dei campi che devono elaborare: in questo modo **è possibile modificare la struttura dei dati senza che questo implichi modifiche ai programmi** applicativi (o, eventualmente, solo a un numero limitato di questi).

Nell'approccio basato su *DBMS*, invece, i dati sono memorizzati una sola volta in modo integrato: **il DB contiene infatti sia la definizione degli schemi di memorizzazione dei dati (aspetto intensionale) sia i dati stessi (aspetto estensionale), compresi i vincoli di integrità.**



basi di dati e sistemi di gestione delle basi di dati

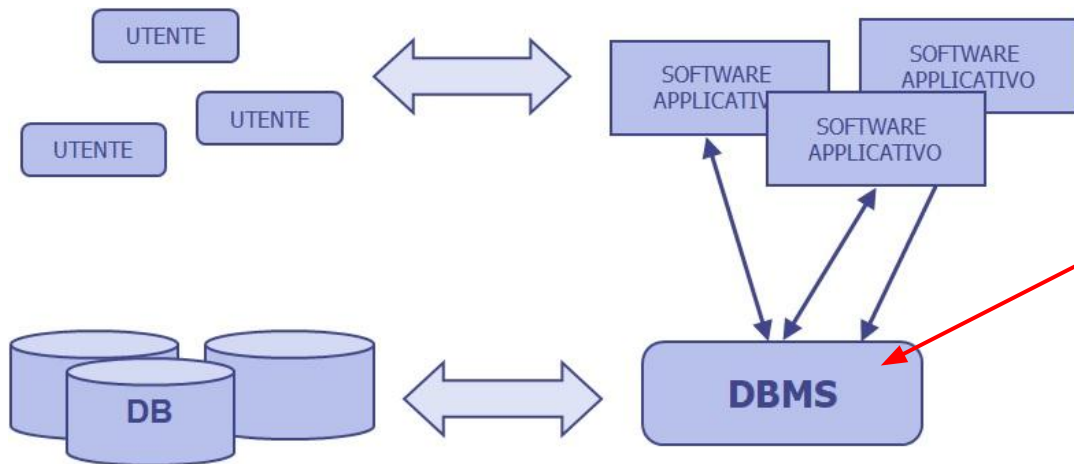
I primi DBMS furono sviluppati alla fine degli anni 60 per calcolatori di fascia alta (mainframe) di grandi organizzazioni. Precedentemente, l'approccio tradizionalmente applicato al problema dell'archiviazione dei dati era quello basato su file system in cui le singole applicazioni software accedono direttamente agli archivi dei dati di proprio interesse.



Nello sviluppo di tali programmi applicativi dunque deve essere nota a priori la struttura interna degli archivi; inoltre sono a carico delle applicazioni stesse la rappresentazione delle correlazioni tra i dati, le tecniche che permettono l'utilizzo contemporaneo degli archivi da parte di più utenti e quelle che garantiscono la consistenza dei dati, funzionalità spesso delegate a carico del sistema operativo.

basi di dati e sistemi di gestione delle basi di dati

Nell'approccio mediante DBMS invece, le richieste di accesso ai dati da parte delle singole applicazioni vengono rivolte direttamente al DBMS, il quale essendo preposto alla gestione completa della base di dati, svincola le applicazioni software da questo onere.

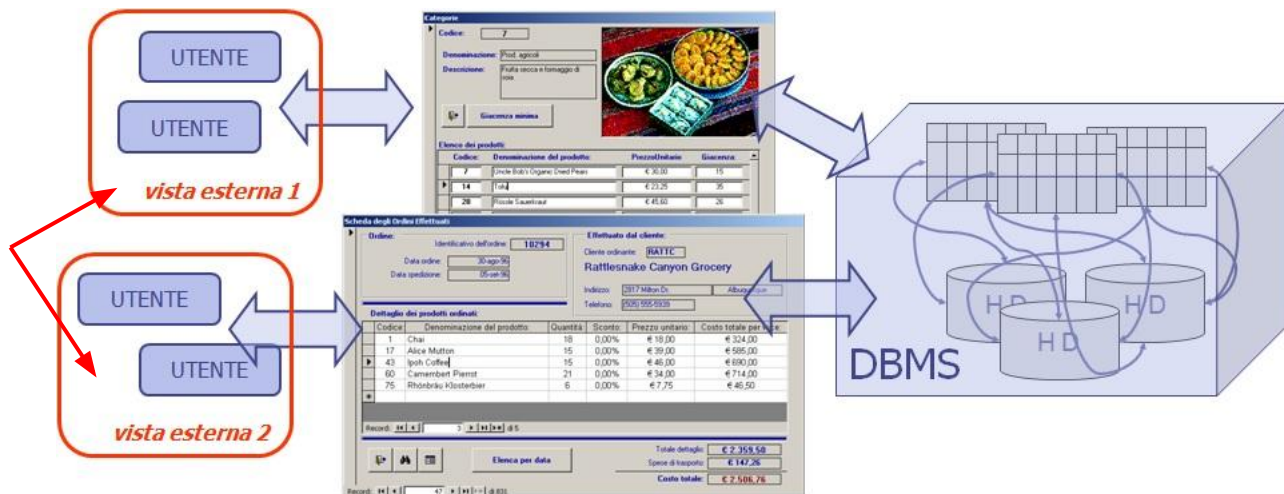


Come già accennato, quindi, **il DBMS si pone come interfaccia software tra il database fisico, memorizzato sulle memorie di massa degli elaboratori elettronici, e gli utenti finali**, che non accedono direttamente ai dati, ma solo attraverso l'uso di opportuni programmi applicativi, spesso forniti dal DBMS stesso.

basi di dati e sistemi di gestione delle basi di dati

In effetti, un **DBMS** realizza sempre un certo livello di astrazione dei dati registrati in memoria, dal momento che i dettagli sulla memorizzazione fisica degli stessi o sulla loro organizzazione logica restano comunque nascosti all'utente finale.

In questo modo è possibile fornire una prospettiva differente della stessa base di dati che può anche essere molto diversa per ogni tipologia di utente che accede al sistema (**viste esterne**), o comunque presentare una visione anche parziale delle informazioni presenti nel database, alcune delle quali possono essere derivate da altri dati registrati in memoria, anche se non esplicitamente memorizzati.



basi di dati e sistemi di gestione delle basi di dati

L'utente interagisce con il DBMS tramite specifici linguaggi:

- **DDL** (*Data Definition Language*): per la definizione della struttura dei dati; sono usati principalmente dall'amministratore del sistema e sono costituiti da comandi per definire e modificare la struttura del database e dei vincoli di integrità;
- **DML** (*Data Manipulation Language*): per la gestione e l'uso dei dati contenuti in una base di dati; sono costituiti da comandi che permettono di accedere ai dati per effettuare interrogazioni o manipolazioni (inserimento, eliminazione, modifica).
- **Query Language** (*Linguaggio di interrogazione*): col quale è possibile dare i comandi necessari per estrapolare informazioni dalla base di dati.

SQL (*Structured Query Language*) è un linguaggio di interrogazione usato spesso anche come DDL e DML.

basi di dati e sistemi di gestione delle basi di dati

Tutto ciò suggerisce una proposta di standardizzazione dell'architettura dei DBMS articolata su **tre diversi livelli di astrazione** (*interno-fisico, logico ed esterno*) ed altrettanti schemi descrittivi dei livelli stessi (**modello ANSI-SPARC**); si parla dunque di:

- **schema esterno**: costituito da più viste esterne ciascuna delle quali descrive quella parte di dati di interesse per l'utente (o per uno specifico gruppo di utenti) corrispondente;
- **schema logico**: descrive la base di dati al livello funzionale utilizzando il modello logico adottato dal DBMS in uso (relazionale, gerarchico, reticolare, ad oggetti), eventualmente derivato da un modello concettuale della stessa base di dati;
- **schema interno**: descrive la base di dati dal punto di vista delle strutture fisiche interne di memorizzazione (tipologie dei file o di indici adottati, ecc.).

basi di dati e sistemi di gestione delle basi di dati

Si tratta di un risultato molto importante, perché questa architettura su tre distinti livelli presuppone una sostanziale **indipendenza dei dati dai programmi applicativi** che li andranno infine ad utilizzare.

Secondo l'approccio su file system, i dati hanno senso e valore solo se interpretati dai programmi che li devono elaborare: la struttura dei dati e le applicazioni che gestiscono i dati stessi sono strettamente correlate. Le applicazioni devono conoscere il formato fisico dei dati (la struttura record), la loro localizzazione in memoria, le tecniche utilizzate per ottimizzare i tempi di ricerca, ecc.

Al contrario, **in un ambiente che utilizza un DBMS i dati hanno struttura fisica e organizzazione logica che prescinde dalle applicazioni**, per cui queste non devono più fare riferimento a tali strutture ed organizzazioni.

basi di dati e sistemi di gestione delle basi di dati

Ciò comporta notevoli vantaggi nella manutenzione del sistema informatico, poiché le modifiche alla struttura fisica dei dati o alla loro organizzazione logica, non implicano la necessità di modificare le applicazioni preesistenti; si parla in particolare di:

- **Indipendenza *fisica* dei dati:** nel senso che è possibile modificare la parte fisica, i supporti fisici di memorizzazione, le modalità con cui i dati vengono registrati sui dispositivi di memorizzazione di massa e le modalità di accesso alle memorie di massa senza doverne modificarne la descrizione logica o i programmi applicativi: **il livello interno (livello fisico) è dunque indipendente dal livello logico.**

basi di dati e sistemi di gestione delle basi di dati

Ciò comporta notevoli vantaggi nella manutenzione del sistema informatico, poiché le modifiche alla struttura fisica dei dati o alla loro organizzazione logica, non implicano la necessità di modificare le applicazioni preesistenti; si parla in particolare di:

→ **Indipendenza *logica* dei dati**: i programmi applicativi, e dunque anche le modalità con cui possono essere organizzate le *viste esterne* che realizzano di fatto l'interfaccia verso gli utenti, risultano essere indipendenti dalla struttura della progettazione logica con cui i dati sono organizzati.

Quindi è ad es. possibile aggiungere o modificare una vista esterna in base alle esigenze di un nuovo utente senza modificare il livello logico: **il livello esterno è dunque indipendente dal livello logico.**

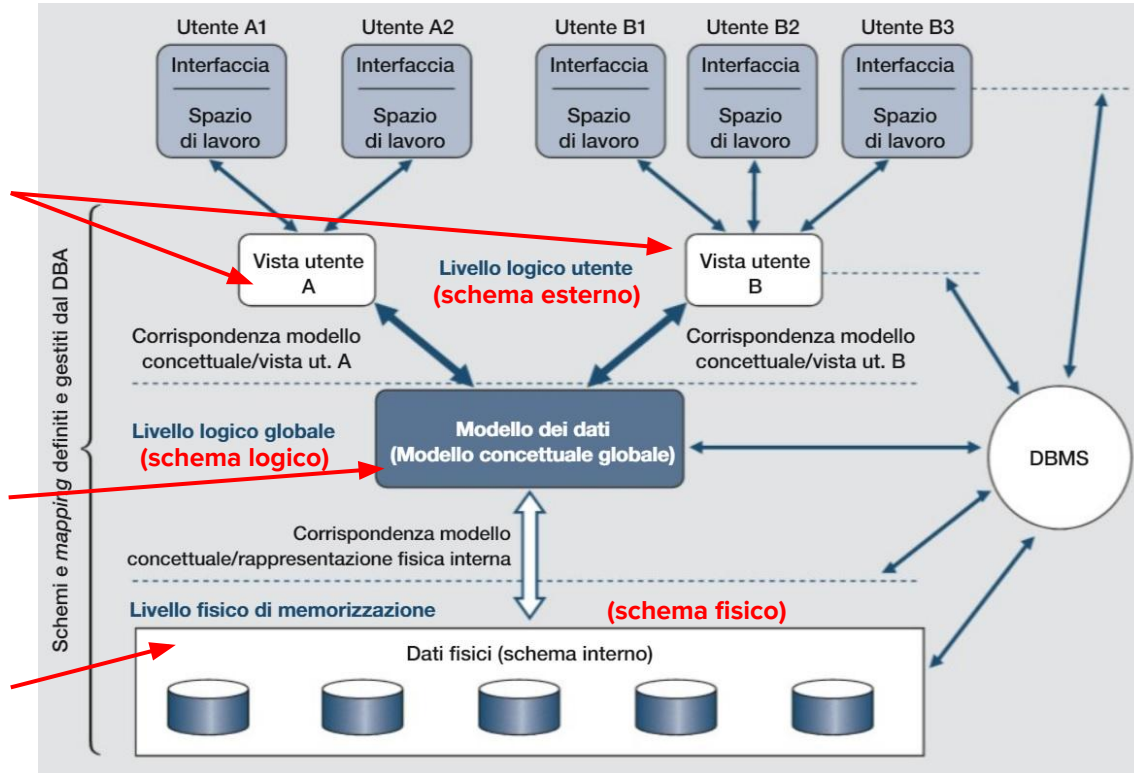
architettura logica di un sistema di gestione delle basi di dati

Livello logico utente: ogni utente ha un proprio spazio di lavoro che si interfaccia col database mediante un'applicazione, o direttamente tramite il linguaggio DML.

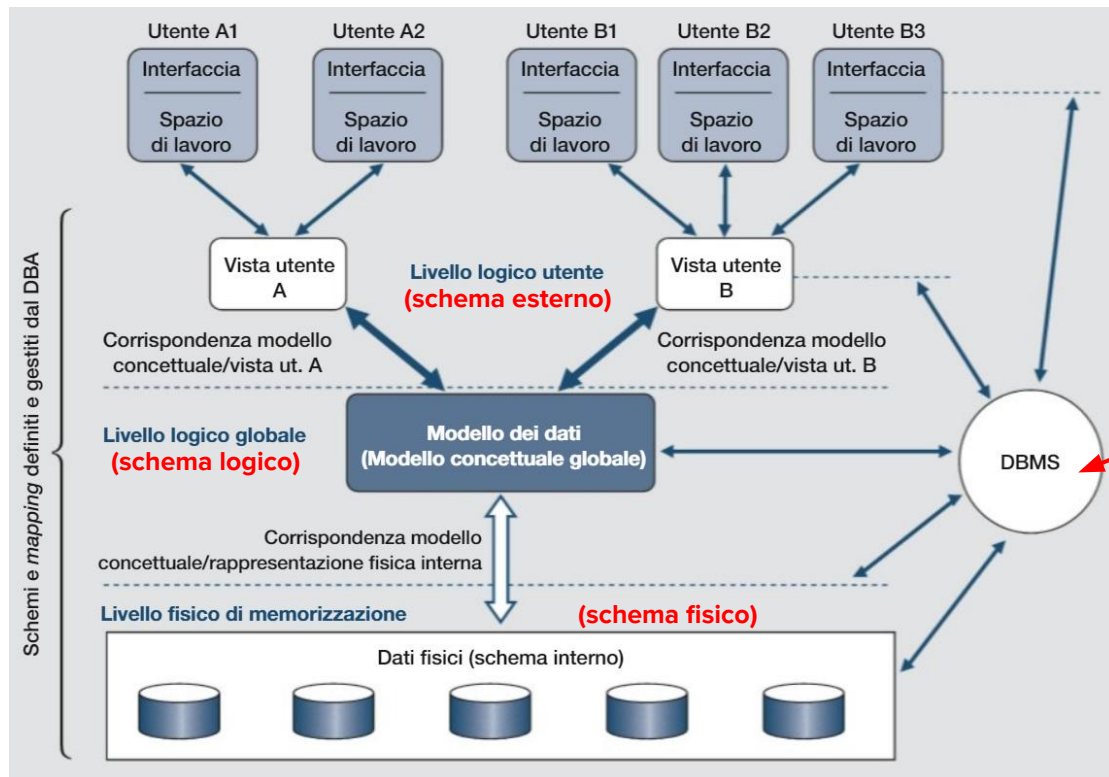
Mediante il linguaggio DDL, l'amministratore definisce le viste utente (viste esterne), ovvero sottoinsiemi del modello logico globale dell'intero database: le viste utente realizzano i meccanismi fondamentali della privacy dei dati.

Livello logico globale: relativo alla definizione della struttura logica (aspetto intensionale dei dati, vincoli di integrità, ecc.) generale del DB: viene definito dall'amministratore utilizzando il DDL.

Livello fisico di memorizzazione: è il livello più basso dove si trovano i dati fisici e pertanto relativo ai dispositivi di memorizzazione, all'organizzazione fisica dei dati e alle relative modalità di accesso.



architettura logica di un sistema di gestione delle basi di dati



Nello schema i collegamenti tra i livelli sono realizzati dal DBMS che gestisce l'intera struttura mediante tecniche di corrispondenza (mapping), cioè procedure software che associano gli elementi di un livello ai corrispondenti di quello contiguo.