

# BASI DI DATI E CONOSCENZA

## GESTIONE DEI DATI E DELLA CONOSCENZA 1

### BASI DI DATI E SISTEMI *WEB-BASED*

INTRODUZIONE

---

Roberto Basili

a.a. 2021/22

# Obiettivi Formativi (*refresh*)

- **Scenario**

- Le grandi quantità di dati accumulate nelle reti di calcolatori costituiscono una risorsa enorme con requisiti critici rispetto alla efficienza dei metodi di **accesso**, **localizzazione** della informazione, **persistenza** e **riutilizzo** in applicazioni eterogenee
- Le applicazioni sono sempre più esigenti dal punto di vista della indipendenza tra logiche dei dati e loro integrazione nei processi di business (vedi WATSON slides)

# Obiettivi Formativi (*refresh*)

- Approfondire i problemi della **modellazione concettuale** delle informazioni ed acquisire una significativa capacità di modellazione dei dati in applicazioni distribuite e complesse
- Conoscere le principali **metodologie e tecnologie di gestione e progettazione delle basi di dati** (Data Base Management).
- Applicare tali tecnologie nelle **applicazione delle BdD** relazionali in scenari tradizionali e Web.
- Introdurre la **relazione tra le Basi di Dati e la rappresentazione della conoscenza**.

# Scopi del Corso: *(refresh)*

- Introduzione alla ***modellazione logico-concettuale dei dati*** nelle applicazioni del software
- Introduzione ai concetti di base dei ***Sistemi di Gestione delle Basi di Dati (DBMS)***.
  - Uso dei **DBMS relazionali** per la progettazione delle Basi di Dati
  - **Modelli dei dati** a livello logico e fisico
- Introduzione alle architetture software per lo sfruttamento dei DBMS nelle moderne applicazioni distribuite, e.g. **Multi-Tier Architectures**
- Introdurre le tendenze e le prassi avanzati nella **gestione dei dati nelle reti di comunicazione**

# Scopi di un Database

- Esempio: [www.amazon.it](http://www.amazon.it)
- Unificare e generalizzare l'accesso ai dati
- Consentire un accesso semplificato e
  - ... efficiente
- Proteggere i dati
  - Integrità e Riservatezza (Sicurezza)
  - Tolleranza a malfunzionamenti/guasti
- Supportare la concorrenza
- Facilitare lo sviluppo dei programmi "utente"

The screenshot shows the Amazon Music interface. At the top, there's a search bar with 'springsteen' entered. Below the search bar, there are navigation options like 'Tutte', 'Supermercato', 'Libri', etc. The main content area displays search results for 'springsteen', showing two identical listings for 'Live 1979 - 2LP' by Bruce Springsteen & The E Street Band. Each listing includes a cover image, the price (37,17€), and a 'Vinite' button. To the left of the listings is a sidebar with various filters and categories, including 'prime Consegna Prime', 'Giorno di consegna', 'Amazon AutoRip', and a list of music genres like 'CD e Vinili', 'Blues', 'Folk', etc.



# The ... Amazon DB

amazon.it Invia a Roberto Roma 00176 CD e Vinili springsteen

Ciao Roberto Account e liste Resi e ordini

Tutte Supermercato Libri Acquista di nuovo Buoni Regalo Salute e cura della casa Amazon Basics Spedizione Gratuita Cronologia di navigazione Idee regalo Prime music 2 milioni di brani c Amazon Prime

amazon music Amazon Music Unlimited Amazon Music Prime CD e Vinili Store Musica Digitale

1-16 dei 773 risultati in "springsteen"

Ordina per: Suggeriti

prime Consegna Prime

- Oggi entro le 22:00  
 Consegna Prime

Giorno di consegna

- Ricevi oggi  
 Ricevi domani

Amazon AutoRip

- AutoRip >>

Categoria

< Tutte le categorie

CD e Vinili

- Altro
- Blues
- Colonne sonore
- Musica Country
- Dance ed Elettronica
- Musica Easy Listening
- Folk
- Musica Hard Rock e Metal
- Musica Indie e Alternativa
- Jazz
- Musica Classica
- Musica internazionale
- Musica italiana
- Musica e storie per bambini
- Musical e cabaret
- New Age e meditazione
- Pop
- Musica R&B e Soul
- Rap e Hip-hop
- Requae



Sponsorizzato

Live 1979 - 2LP

di Bruce Springsteen & The E Street Band | 2021

Vinile

37,17€

Prenotazione al prezzo minimo garantito

prime Consegna GRATUITA

Data di uscita: 19 novembre 2021.



Live 1979 - 2LP

di Bruce Springsteen & The E Street Band | 2021

Vinile

37,17€

Prenotazione al prezzo minimo garantito

prime Consegna GRATUITA

Data di uscita: 19 novembre 2021.

CD audio

27,44€

Prenotazione al prezzo minimo garantito

prime Consegna GRATUITA

Data di uscita: 19 novembre 2021.



Bruce Springsteen - Darkness On The Edge Of Town - CBS

di Bruce Springsteen

★★★★☆ 4

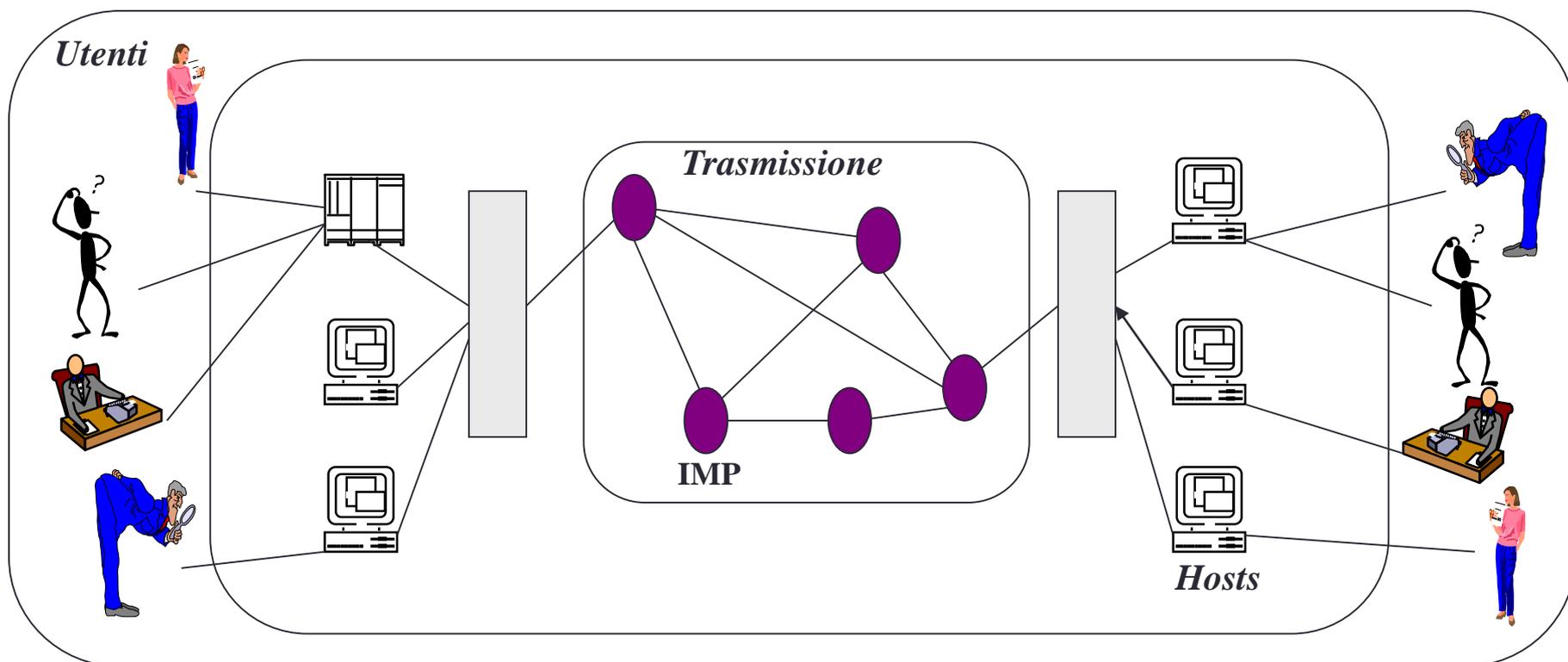
Vinile

# DBMS

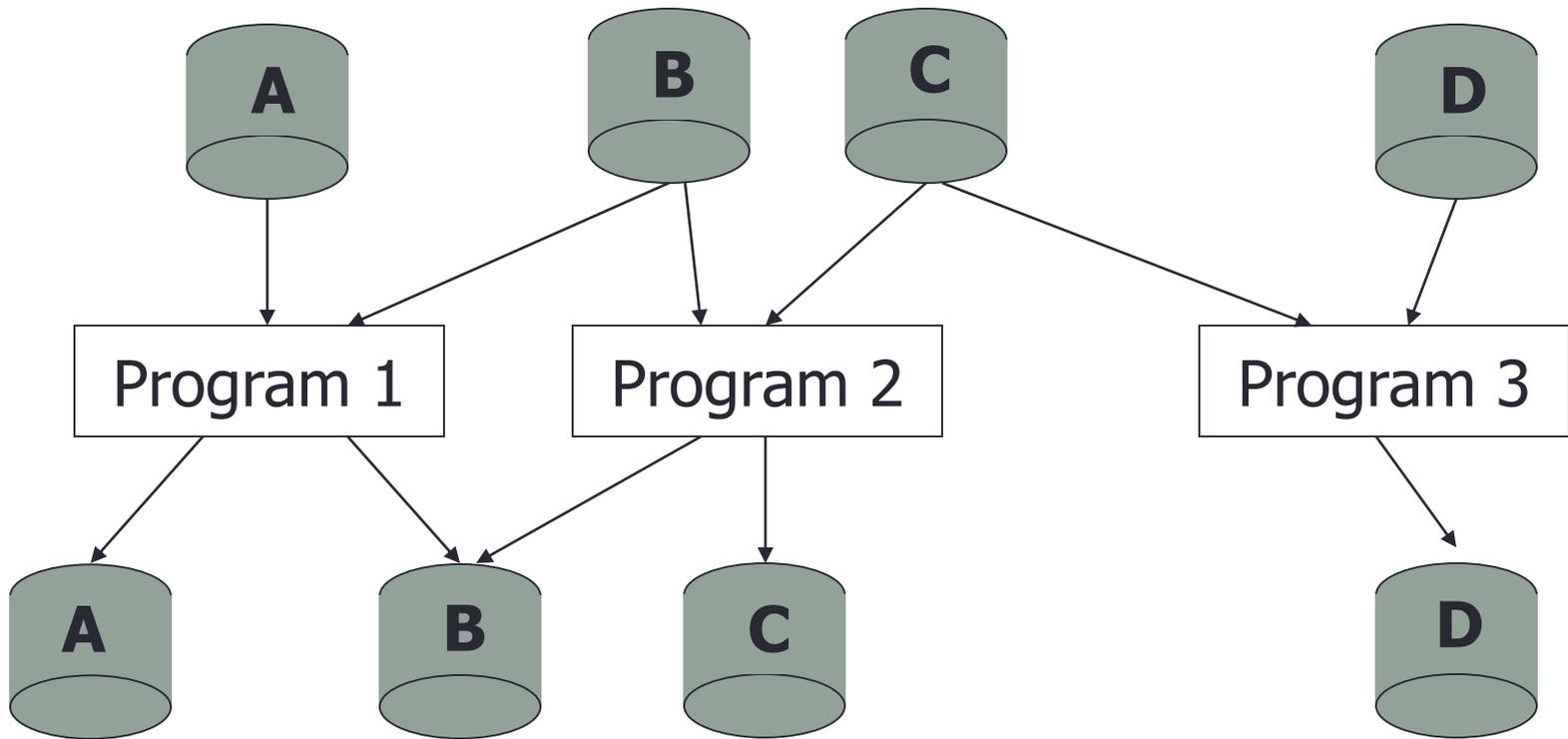
- Un **Data Base Management System** (DBMS) è un sistema software per lo storage, la gestione e la fornitura efficiente di dati secondo logiche eterogenee e verso applicazioni diverse.
- Vantaggi:
  - Indipendenza ed accesso efficiente ai dati
  - Riduzione del *tempo di sviluppo*
  - *Garanzia di Integrità e sicurezza* dei dati
  - Gestione degli accessi ai dati e Supporto all'*accesso concorrente*
  - Gestione delle *transazioni*
  - Gestione dei *Crash* e supporto al *Recovery*

# Applicazioni e DB

- ..... uno scenario Web

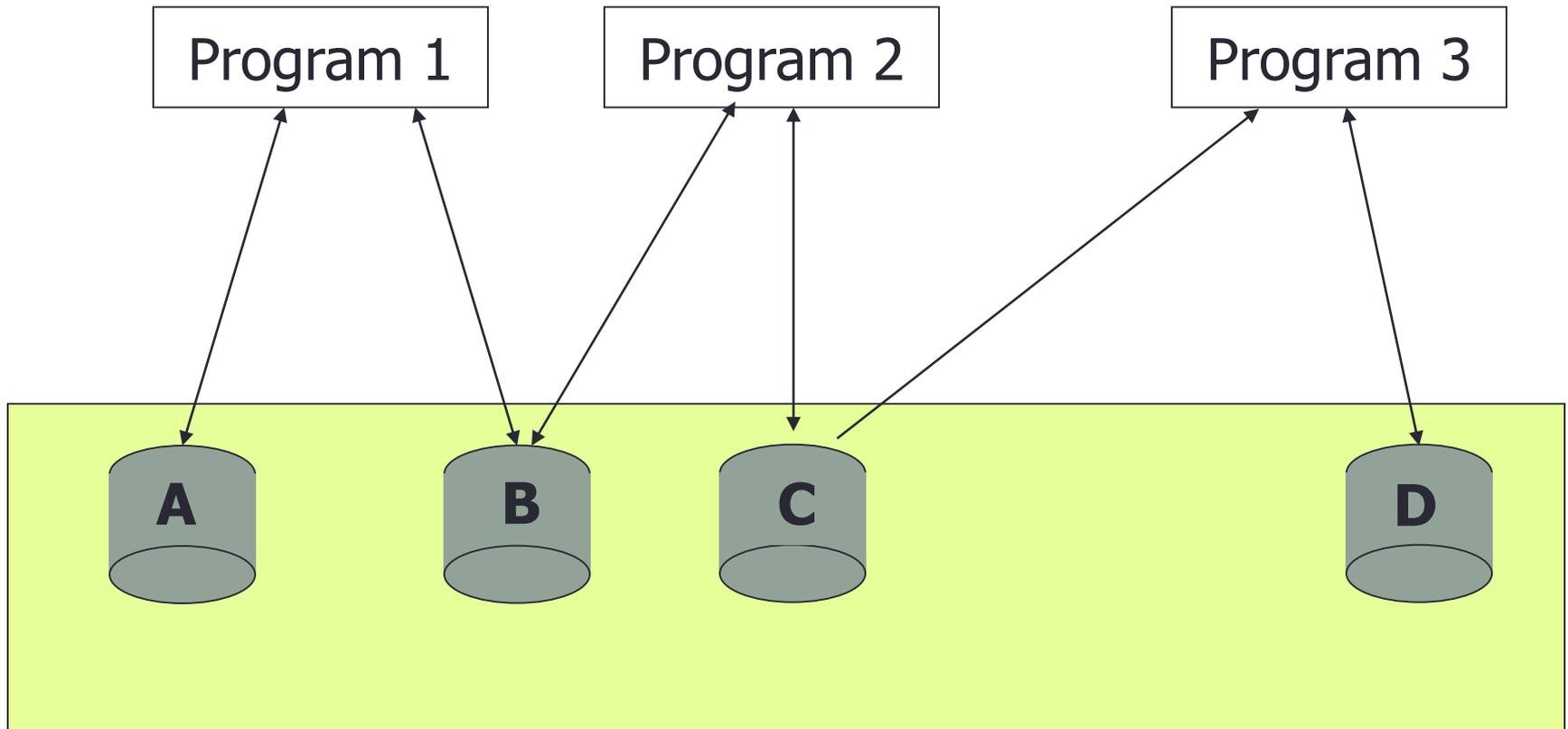


# Archivi e Basi di Dati



Archivi separati

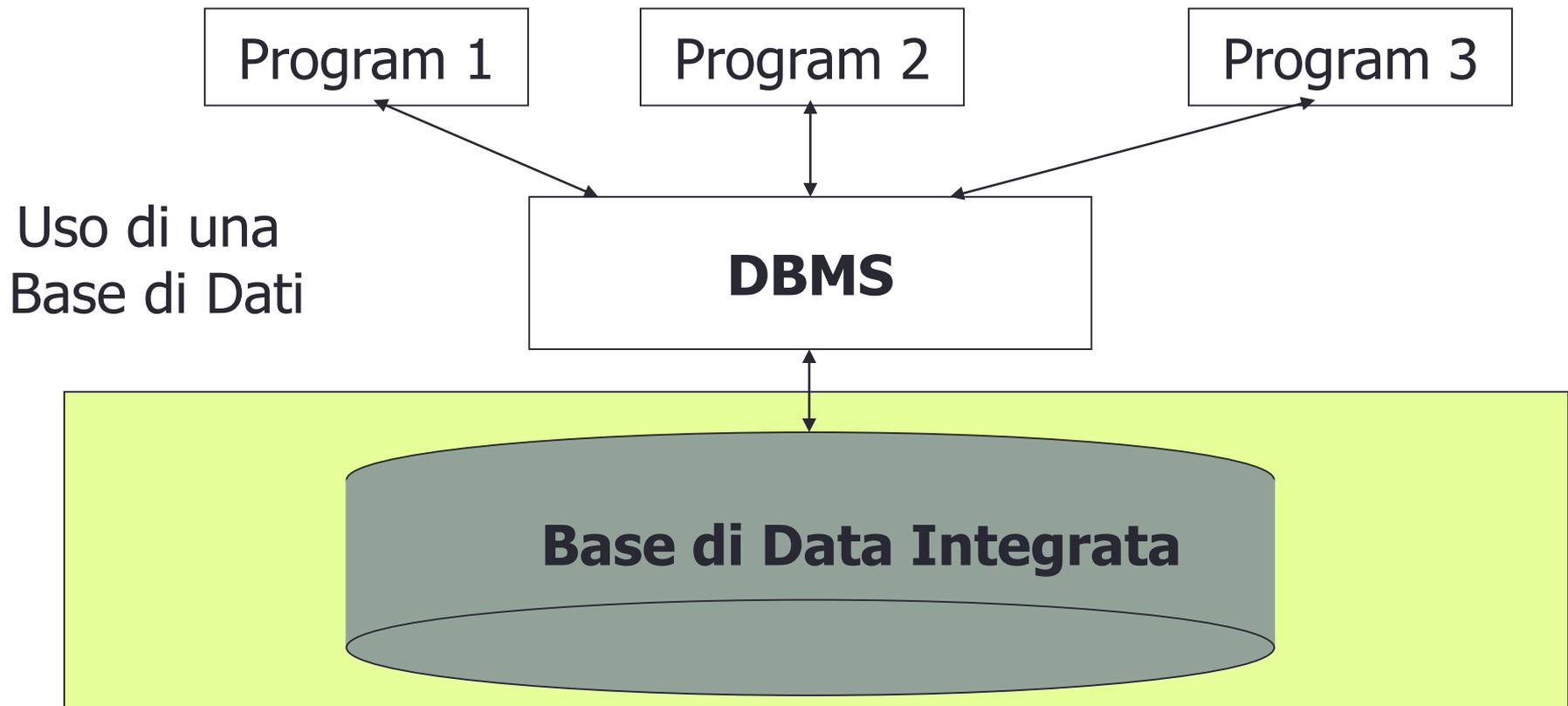
# Archivi e Basi di Dati



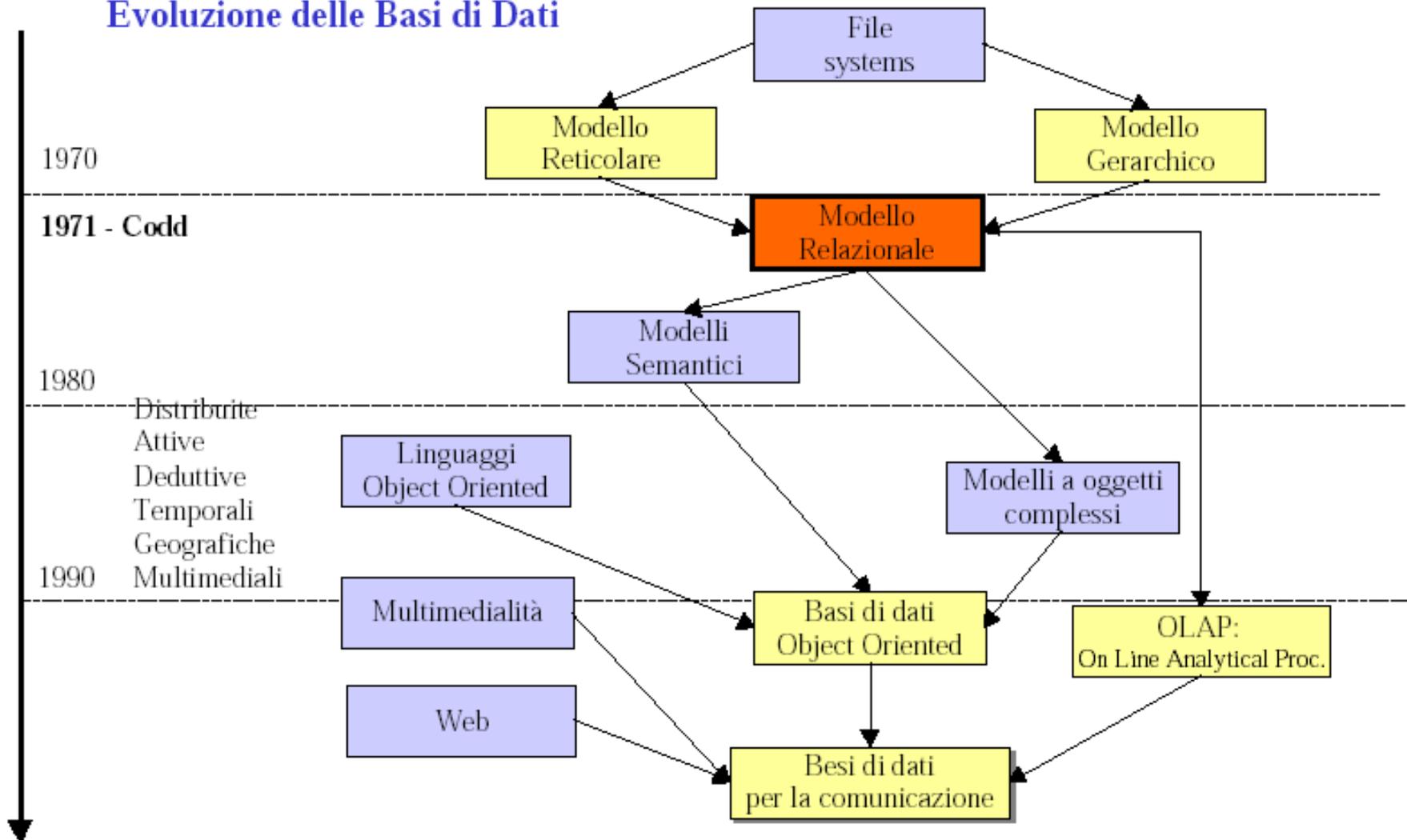
# Applicazioni e Database

- Requisiti:
  - Diverse applicazioni usano porzioni diverse dei dati (pensiamo una anagrafica clienti che può essere usata per la *fatturazione* ma anche per la *pianificazione di marketing*)
  - L'astrazione sui dati e' diversa ma "coerente" tra le applicazioni diverse (i clienti non cambiano, solo che vengono usate informazioni diverse)
    - Fatturazione: nome cognome, codice fiscale, indirizzo per la spedizione, ...
    - Marketing: Nome cognome, fascia di età, sesso, tipologia di acquisiti ... per le statistiche su acquisti e preferenze
  - L'accesso ai dati e'
    - dipendente dalla astrazione (modello logico, risponde a: *cos'è un cliente?*)
    - vincolato dalla natura fisica della memorizzazione (come ne memorizzo i diversi dati, in modo da poterli recuperare a seconda della applicazione)
- Conseguenza: E' necessario svincolare la *natura logica dei dati* dalla loro *rappresentazione in memoria*, cioè dalle forme di memorizzazione

# Archivi e Basi di Dati



# Evoluzione delle Basi di Dati

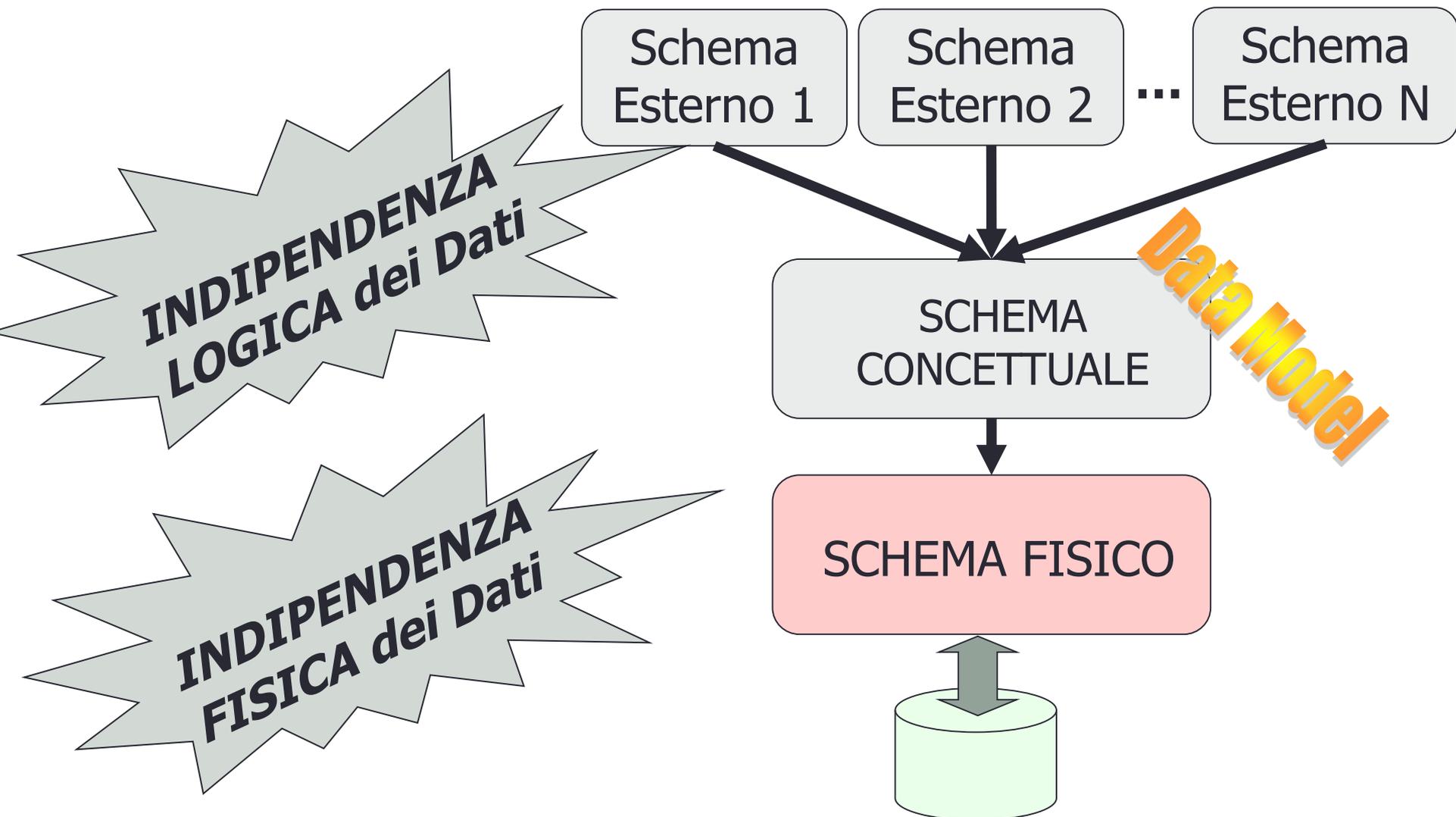


# Data Model

- **Schema Fisico**: definisce i dettagli legati alla memorizzazione dei dati (file, indici e ecc).
- **Schema Logico/Concettuale**: Definisce i dati in base alla loro natura concettuale. E' basato su un Modello Logico di definizione dei dati (ad esempio Modello Relazionale)
- **Schema esterno**: Rappresentazione, basata sul modello logico, dei dati che sono utili/necessari e legali per un gruppo di utenti. Sottoinsieme delle informazioni descritte nello schema concettuale (spesso detto view)

E' necessario svincolare la natura **logica** dei dati dalla rappresentazione in memoria, o meglio dalle forme di memorizzazione

# Livelli di Astrazione



# Schema Fisico

- Definisce i dettagli legati alla memorizzazione
  - Es.
    - liste per rappresentare insiemi
    - liste per rappresentare alberi
    - alberi per rappresentare insiemi con chiavi
- Determina la struttura dei *files* dedicati a contenere i diversi dati
- Definisce forme utili al ritrovamento (es. indici)

# Schema Concettuale ( o Logico)

- Definisce i dati relativamente alla loro natura concettuale, cioè legata al mondo della/e loro applicazione/i
- Viene realizzato all'interno di modello logico di definizione dei dati (per es. quello relazionale)
- In generale, descrive tutte le entità e le relazioni tra di loro che si rendono necessarie per descrivere completamente i dati in modo logico

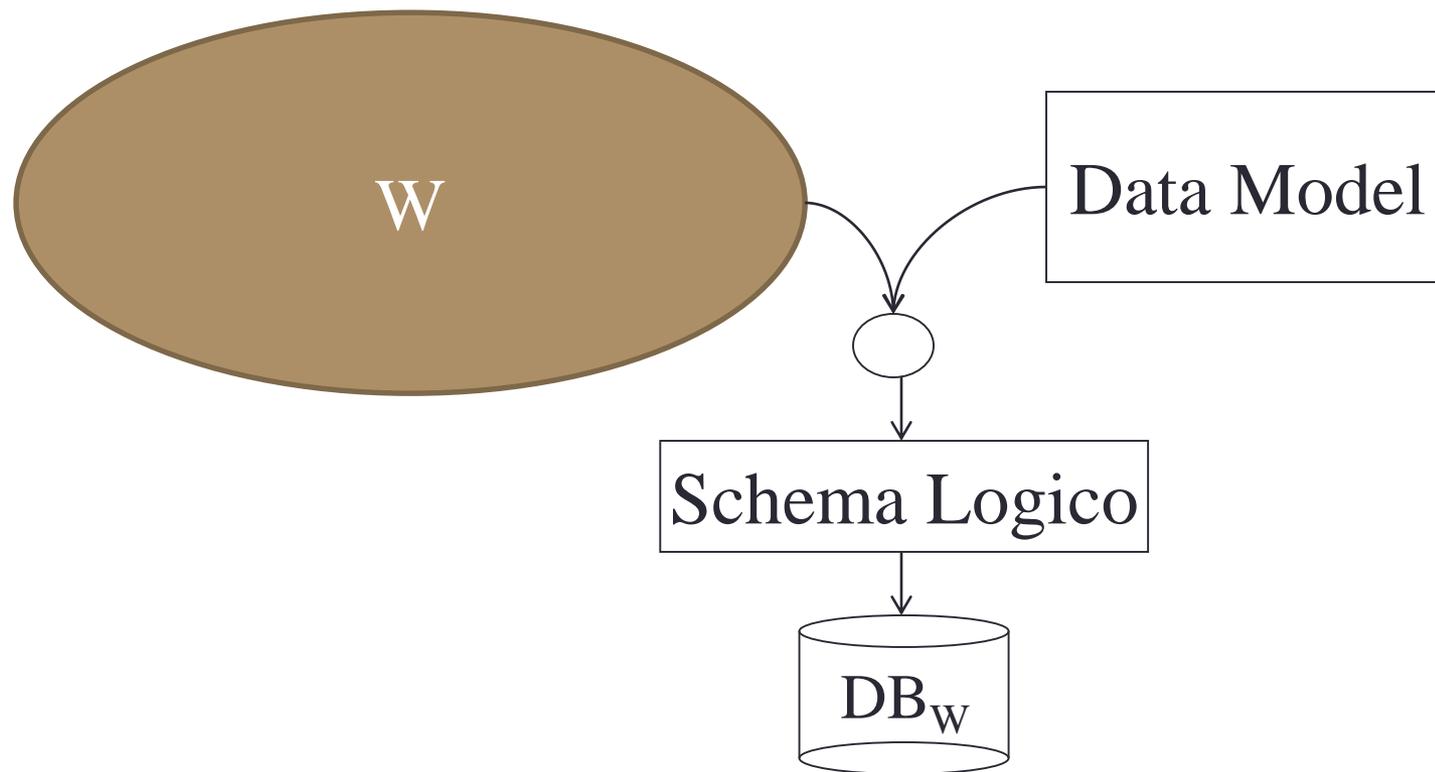
# Schema Esterno

- Descrive i dati che sono utili/necessari e legali per gruppi di utenti
- Utilizza il modello logico di descrizione dei dati
- E' rappresentato da un sottoinsieme delle informazioni dello schema concettuale, spesso detto "vista"

# Progettazione Concettuale

- Ha lo scopo di modellare il mondo in cui le applicazioni utente operano e richiede un linguaggio per la sua specificazione
- Tale linguaggio e' detto MODELLO dei DATI (*Data Model*)
  - e' una astrazione di tutte le possibili informazioni che i diversi mondi coinvolti dalle applicazioni richiedono
  - deve quindi essere *universale*, valido per tutti i domini/mondi
- Nella progettazione concettuale, la applicazione di un MODELLO dei DATI al mondo  $W$ , cioè al dominio di una applicazione, dà come risultato lo SCHEMA LOGICO del mondo e consente la creazione di una base di dati (relativa a  $W$ )

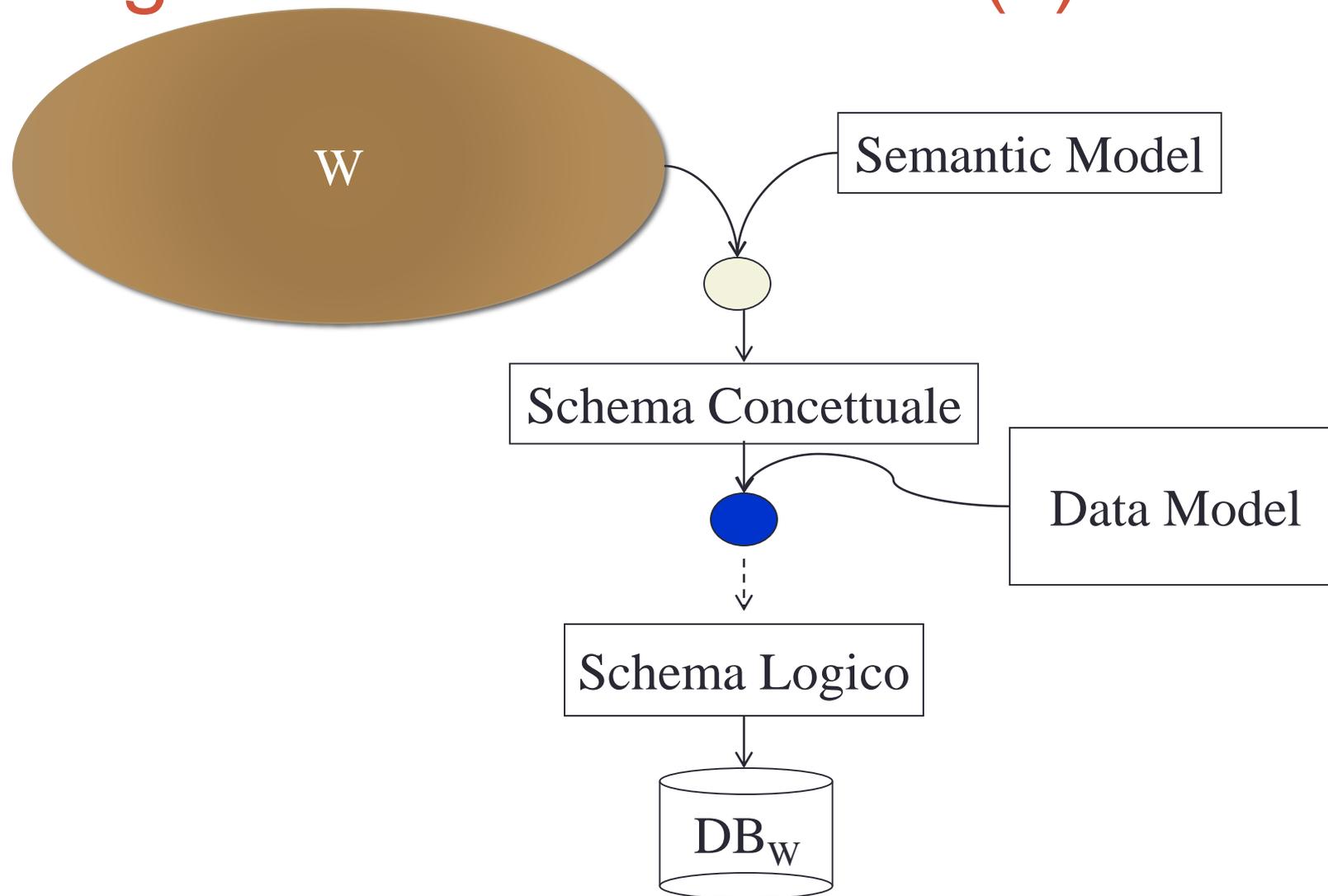
# Progettazione Concettuale (2)



# Progettazione Concettuale

- Il modello dei dati richiede competenze e linguaggi già propri dell'informatica ma ...
- la progettazione logica parte dalle descrizioni che gli utenti finali (non informatici) fanno del loro mondo operativo
  - Ad esempio le informazioni utili alla gestione degli studenti all'Università sono definite dalle segreterie didattiche e non coinvolgono (solo) il progettista del DB, ma soprattutto i manager delle segreterie
- => e' necessario quindi un modello (ed un sottostante linguaggio) più adatto a catturare la semantica del dominio applicativo, cioè un modello semantico (
- Tale modello semantico è indipendente, a sua volta dal modello dei dati)
- Il modello semantico poi sarà reso mappabile nel corrispondente modello logico tramite processi semplici e per i quali esistono prassi standard (il processo poi sarà deterministico)

# Progettazione Concettuale(3)



# Indipendenza dei Dati

- I livelli logici di progettazione supportano la Indipendenza Logica dei Dati ...
- Cambiamenti del mondo  $W$  si rifletteranno solo come estensioni/ridefinizioni dello schema logico minimizzando i cambiamenti nelle applicazioni utente (tramite adattamento degli schemi esterni)

# Indipendenza dei Dati (2)

- ... e l'Indipendenza Fisica dei Dati
- Cambiamenti nella struttura fisica delle rappresentazioni (ad es. nuovi dispositivi di memorizzazione) potranno essere resi invisibili (cioè trasparenti) alle applicazioni utente, mantenendo invariato lo schema logico

# La manipolazione dei Dati

- Gli utenti di un database hanno in generale necessita' di:
  - INSERIMENTO (ad es. per aggiungere nuovi studenti)
  - AGGIORNAMENTO (ad es. per correggere i dati di uno studente)
  - CANCELLAZIONE (ad es. per eliminare studenti che lasciano l'università)
  - INTERROGAZIONE (ad es. per recuperare i dati di uno studente)
- Tali operazioni possono operare su informazioni complete del dominio o su loro porzioni (cioè singole “parti” di una informazione, ad es. la matricola di uno studente, o la data di un esame)

# Le transazioni

- La realizzazione delle operazioni di accesso ai dati da luogo ad una molteplicità di azioni più elementari sui dati
  - es. per accedere alla modifica di un dato X
  - verifica dell'esistenza del dato X
    - < aggiornamento del dato X
    - < preparazione dell'output (es. feedback nuovo valore di X)
    - < avviso di terminazione
- La successione delle operazioni determinate da una (sola) manipolazione e' detta *transazione*

# Transazioni (2)

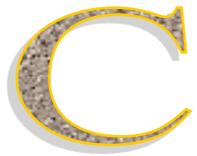
- Diverse operazioni sui dati (pur se specificate nello stesso modo) determinano in genere *transazioni diverse*
- La esecuzione della transazione non e' completa se non alla fine della sequenza delle (micro)operazioni richieste
- Quando eseguite nello stesso intervallo temporale, transazioni diverse si dicono tra loro *concorrenti* poichè la esecuzione dell'una può influenzare il risultato delle altre
  - in regime di concorrenza, sequenze di operazioni corrette per le singole transzioni, quindi legali per la singola operazione, possono produrre risultati complessivi diversi (es. due prelievi Bancomat sullo stesso conto)
- Osserviamo che deve essere applicato un rigido controllo delle conseguenze di esecuzioni concorrenti di transazioni diverse per mantenere (garantire) l'integrita' dei dati  
(vedi es. Bancomat!!)

# Transazioni

Una **Transazione** quindi è un raggruppamento di operazioni che deve avere le seguenti proprietà:

A large, stylized letter 'A' with a yellow outline and a grey, textured interior, serving as a decorative element for the first property.

**ATOMICITA'** - La transazione viene eseguita **completamente** o non viene eseguita affatto

A large, stylized letter 'C' with a yellow outline and a grey, textured interior, serving as a decorative element for the second property.

**CONSISTENZA** - Le transazioni debbono conservare la consistenza dei dati nel database

A large, stylized letter 'I' with a yellow outline and a grey, textured interior, serving as a decorative element for the third property.

**ISOLAMENTO** - L'effetto di una transazione eseguita in ambiente concorrente deve essere lo stesso della medesima transazione eseguita singolarmente

A large, stylized letter 'D' with a yellow outline and a grey, textured interior, serving as a decorative element for the fourth property.

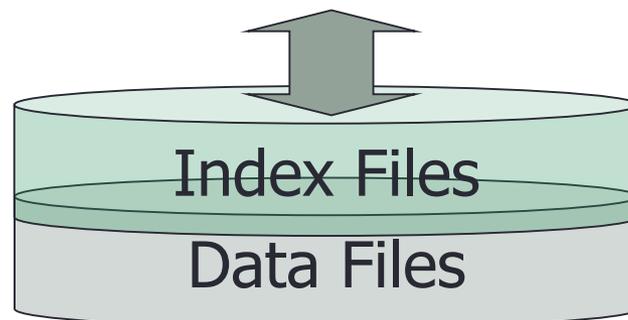
**DURABILITA'** - L'effetto di una transazione in un Database deve essere persistente anche in caso di crash del sistema

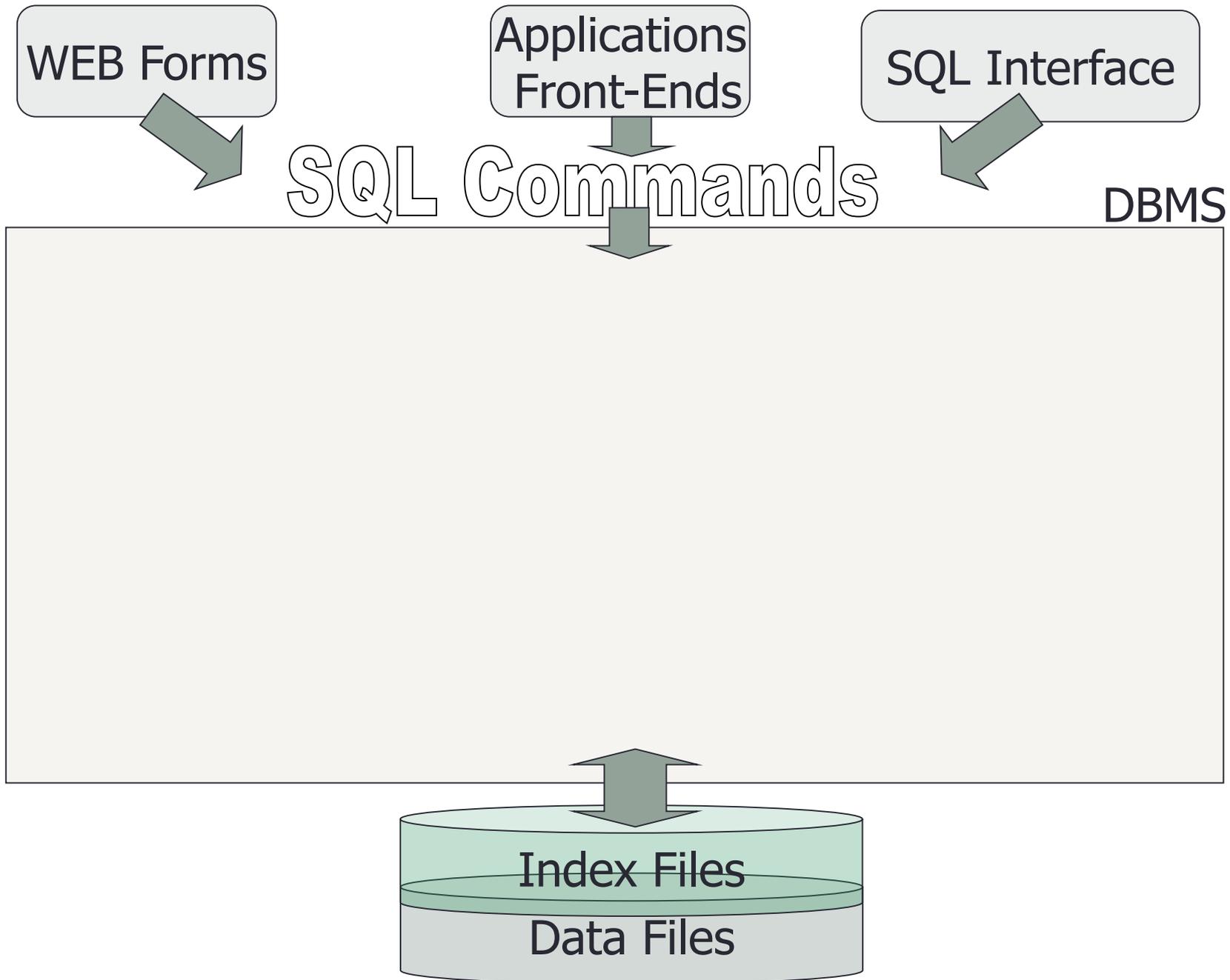
WEB Forms

Applications  
Front-Ends

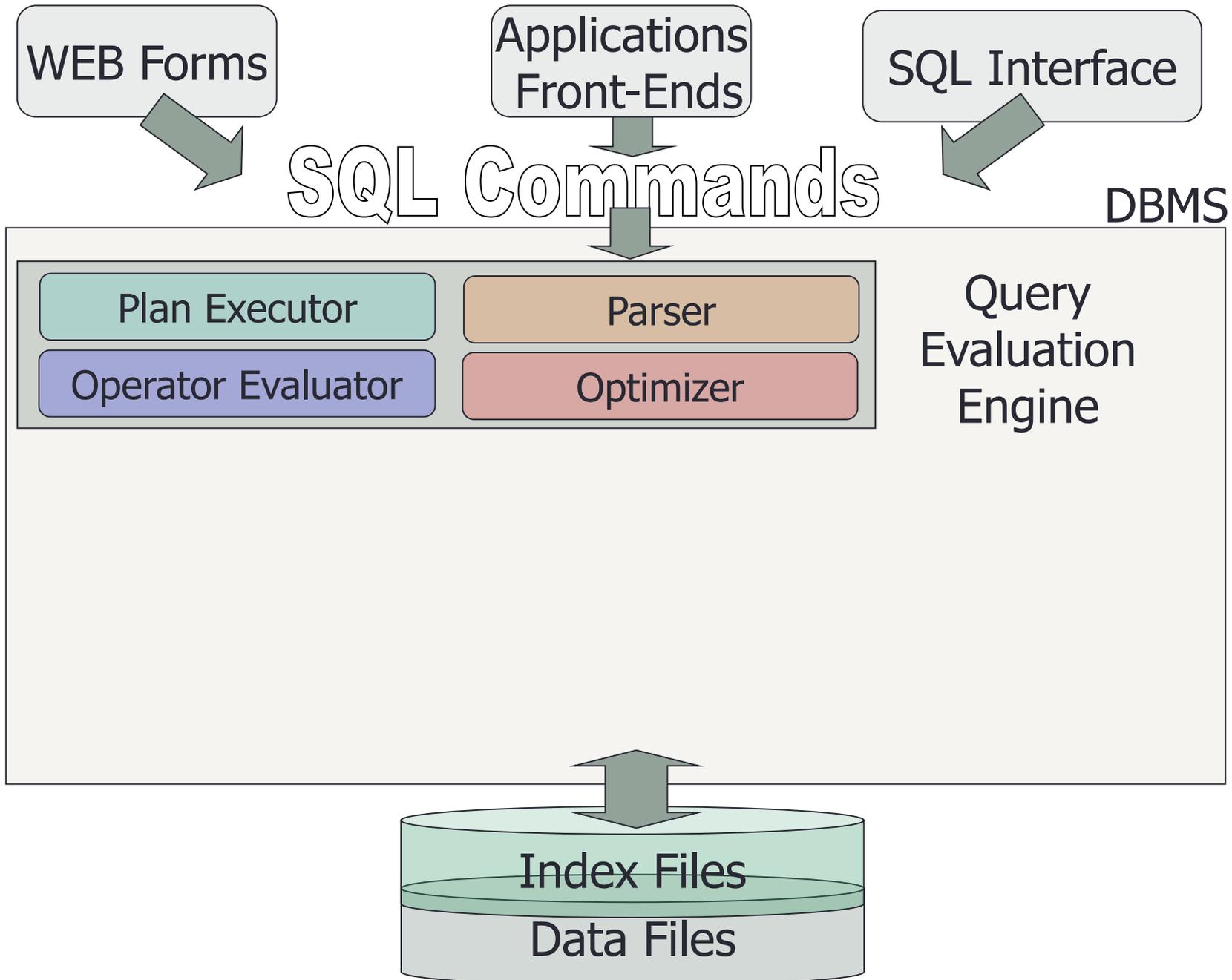
SQL Interface

Architettura di un Database

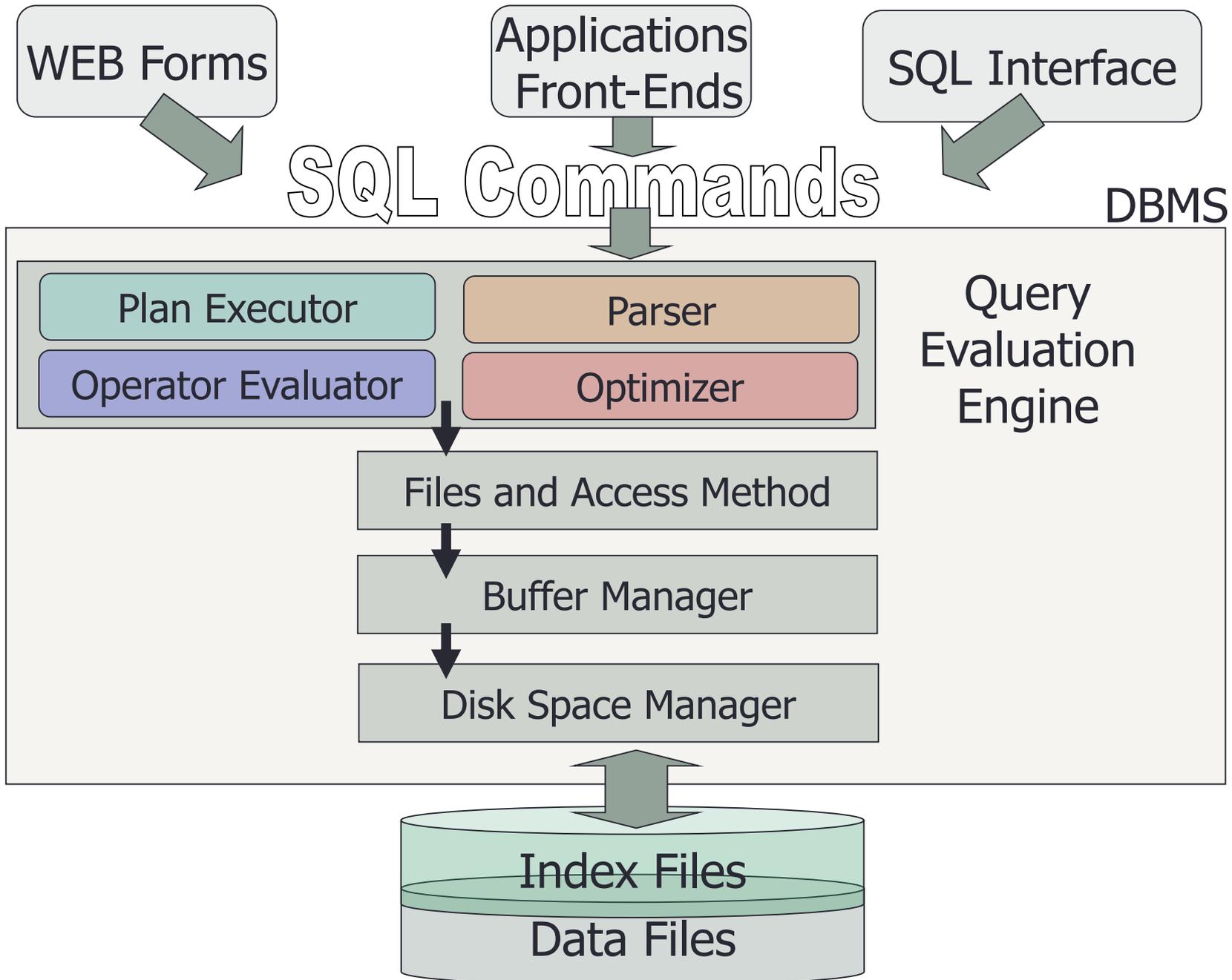




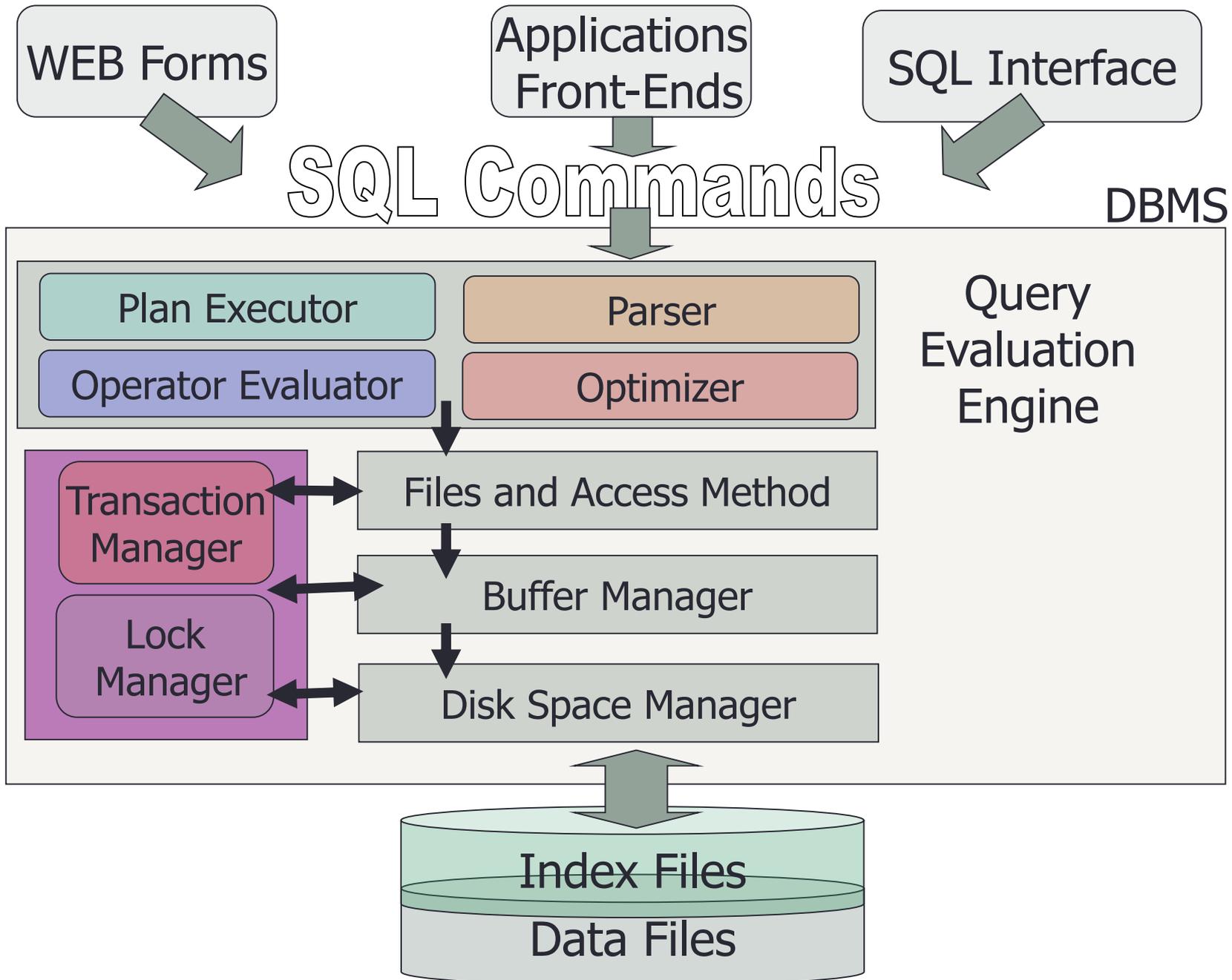
Architettura di un Database



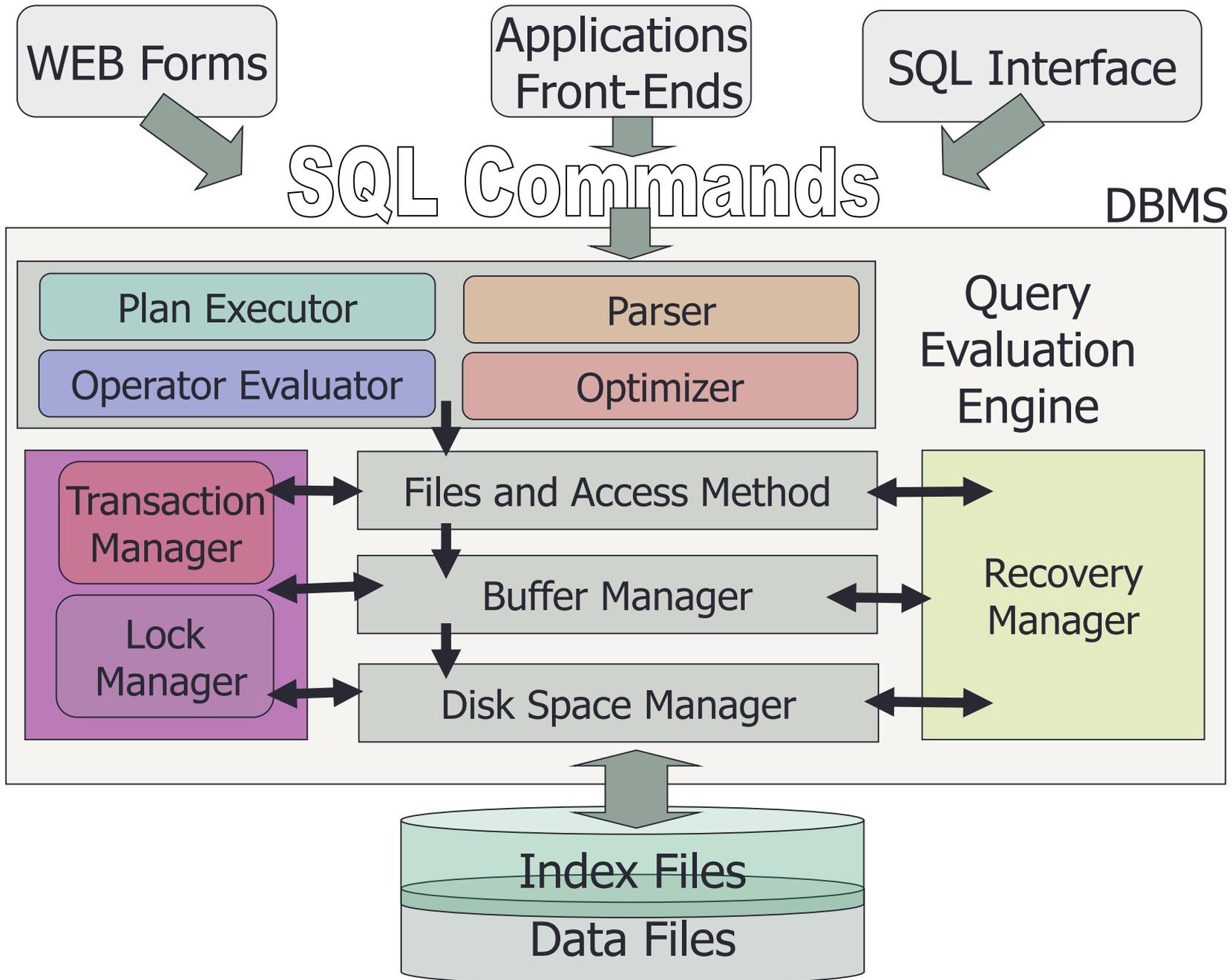
Architettura di un Database



Architettura di un Database



Architettura di un Database



Architettura di un Database

# Utenti di un Database

- Programmatori del DBMS
- Utenti (diretti) esperti e non esperti
- Programmi applicativi
- Amministratore
  - progetto degli schemi fisici e logici
  - gestione della sicurezza e dei criteri di autorizzazione
  - Manutenzione e prevenzione dei malfunzionamenti
  - Customizzazione

# Oltre le basi di dati

- Basi di dati e Web
  - Client vs. server, data layer vs. business logic layer
  - Load balancing: Data vs. problem solving computation
- Dati multimediali e multimodalità
  - Audio, Immagini e Video
  - Integrazione multimodalità (video storage & retrieval)
- Approcci No SQL
- Data warehousing
- DB search vs. Information Retrieval
  - Querying vs. Questioning
  - Content logic vs. Data logic
- Intelligenza Artificiale e DBs: basi di conoscenza

# Sommario

- Perché l'indipendenza tra dati e applicazioni è rilevante
- L'indipendenza si realizza (almeno) a tre livelli:
  - Concettuale
  - Logico
  - Fisico
- I DBMS rispondono metodologicamente e tecnologicamente a queste esigenze

## Sommario (2)

- I modelli dei dati consentono di esprimere tutti i vincoli di dominio, delle logiche di processo e delle tecnologie di storage utilizzate
- Essi sono meta-modelli per specializzare i processi del DBMS alla specifica applicazione/scenario
- Gli utenti dei DB formano una comunità complessa che include anche le applicazioni eterogenee tra loro
- La architettura complessiva di un DBMS risolve i diversi problemi attraverso la loro distribuzione tra processi indipendenti diversi