

# **Basi di Dati**

## **Introduzione**

Corso B

# Basi di Dati – Corso B

Progettazione, programmazione e architetture delle  
basi di dati

a.a. 2018-2019

Ruggero G. Pensa ([pensa@di.unito.it](mailto:pensa@di.unito.it))

Ricevimento: giovedì 16-18

**Sito Moodle:** Basi di Dati (2018/19), Corso B (BD-B-18-19)

**TESTO:**

- **Atzeni, Ceri, Fraternali, Paraboschi, Torlone**  
"Basi di dati", McGraw-Hill, Quinta edizione, 2018.

# Contenuti

- **Basi di Dati relazionali**
  - Modello Relazionale
  - Algebra e calcolo relazionale
  - SQL
- **Progettazione di Basi di Dati**
  - Metodologie e modello Entità Associazioni
  - Progettazione concettuale e logica
  - La normalizzazione
- **Sistemi per la gestione dei dati**
  - Metodi di accesso e indici
  - Ottimizzatore logico e fisico
  - Gestione della concorrenza
  - Gestione del ripristino

# Contenuti - Roadmap

Lab (progettazione)	Corso di Teoria	Lab (SQL)
<ul style="list-style-type: none"><li>• Metodologie e modello Entità Associazioni</li><li>• Progettazione concettuale e logica</li></ul>	<ul style="list-style-type: none"><li>• Modello Relazionale</li><li>• Algebra e calcolo relazionale</li><li>• Ottimizzazione logica</li><li>• La normalizzazione</li><li>• Metodi di accesso e indici</li><li>• Ottimizzatore fisico</li><li>• Gestione della concorrenza</li><li>• Gestione del ripristino</li></ul>	<ul style="list-style-type: none"><li>• Linguaggio SQL</li></ul>

```
graph LR; L1[Metodologie e modello Entità Associazioni] --> T1[Modello Relazionale]; L2[Progettazione concettuale e logica] --> T2[Algebra e calcolo relazionale]; T2 --> L3[Linguaggio SQL]; T3[Ottimizzazione logica] --> L3; T4[La normalizzazione] --> L3;
```

# **Dove contestualizzare gli argomenti del corso:**

## **Sistema informativo di una organizzazione**

Componente (sottosistema) di una organizzazione che gestisce (acquisisce, elabora, conserva, produce) le informazioni di interesse (cioè le informazioni utilizzate per il perseguimento degli scopi dell'organizzazione )

# Sistema informativo, commenti

- Ogni organizzazione ha un sistema informativo, eventualmente non esplicitato.
- Quasi sempre, il sistema informativo è di supporto ad altri sottosistemi (amministrazione, produzione, magazzino,...), va quindi studiato nel contesto in cui è inserito.
- Il sistema informativo è di solito suddiviso in sottosistemi (in modo gerarchico o decentrato), più o meno fortemente integrati.
- Il concetto di “sistema informativo” è indipendente da qualsiasi automatizzazione.

# Sistema organizzativo

- Insieme di **risorse** e regole per lo svolgimento coordinato delle attività (processi) al fine del perseguimento degli scopi.
- Risorse di un'azienda (ente, amministrazione,...):
  - risorse umane
  - risorse finanziarie
  - materiali
  - **informazioni**

# Sistema organizzativo e sistema informativo

- Il sistema informativo è parte del sistema organizzativo
- Il sistema informativo esegue/gestisce processi informativi (cioè i processi che coinvolgono varie risorse: personale, flussi finanziari, materiale, **informazioni**)



# Sistema Informatico

- Porzione automatizzata del sistema informativo:

la parte del sistema informativo che gestisce informazioni con tecnologia informatica.

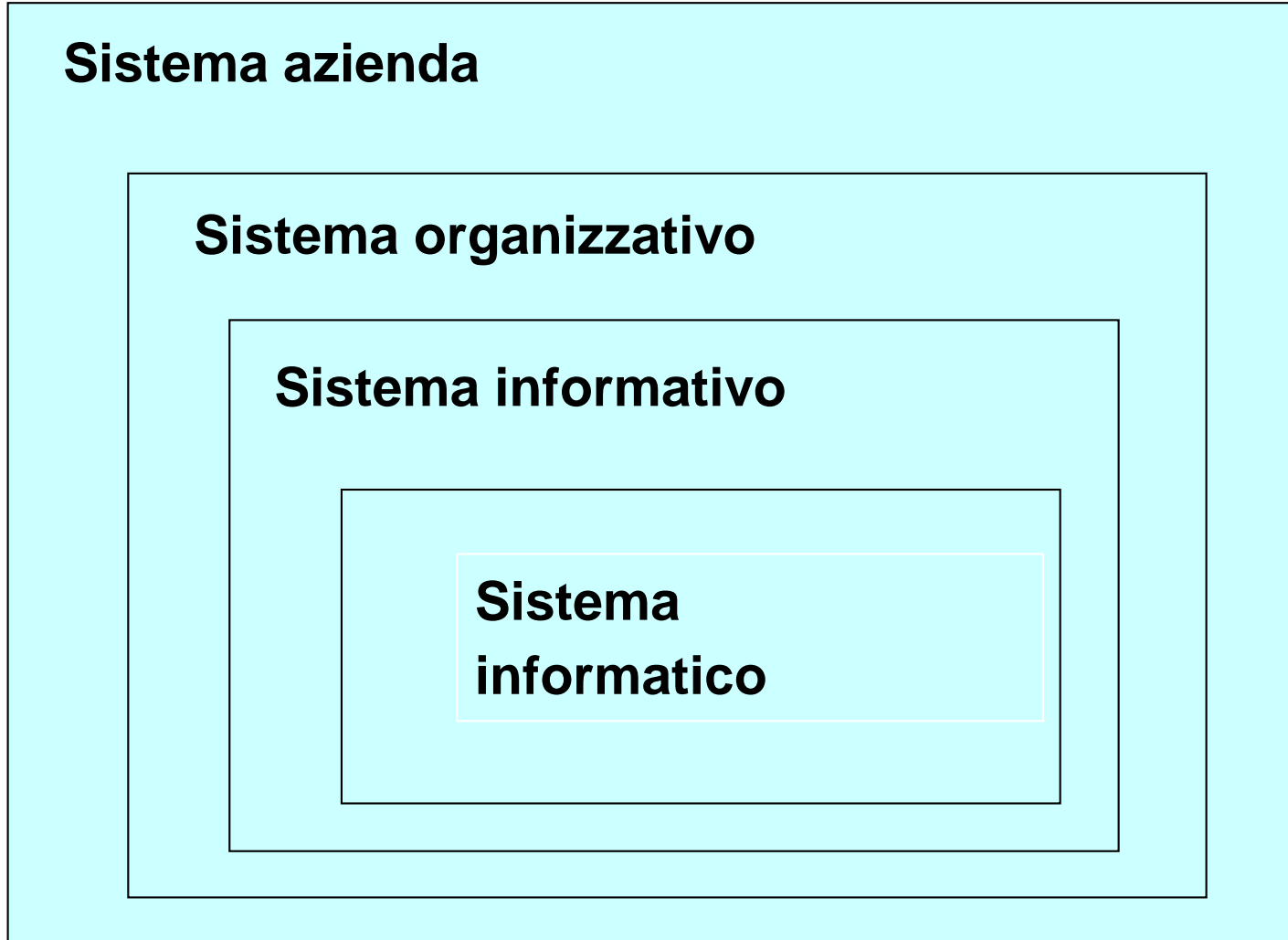
# Sistema informatico

**Sistema azienda**

**Sistema organizzativo**

**Sistema informativo**

**Sistema  
informatico**



# Paradigma generale di riferimento

- **Visione sistemica:** ogni organizzazione ha degli input (ordini, richieste di servizi, ecc...) e degli output (prodotti venduti, utile, erogazione di servizi, ecc...)
- **Focalizzazione sui dati**, o meglio ancora sulle **informazioni**: tutto ciò che non è espresso nei dati non esiste o meglio non interessa all'organizzazione. Es.: Quando lo studente si immatricola non dichiara il colore dei capelli.
- **Strutturare il sistema in sottosistemi interconnessi da interfacce.**  
Es.: La didattica comprende studenti e docenti con relativi sottosistemi.
- **Disporre di modelli formali per rappresentare dati e funzioni**  
Ci occuperemo dei modelli dei dati rinviando l'analisi delle funzioni ai corsi di ingegneria del software.

# Gestione delle informazioni

- *Raccolta*, acquisizione
- *Archiviazione*, conservazione
- *Elaborazione*, trasformazione, produzione
- *Distribuzione*, comunicazione, scambio

# Informazioni e dati

- Nei sistemi informatici (e non solo), le informazioni vengono rappresentate in modo essenziale, spartano: attraverso i **dati**.
- Nei sistemi informatici sono informazioni i dati che assumono un significato specifico in un determinato contesto. In altre parole i dati **interpretati e correlati tra loro diventano informazioni**.

# Dati e informazioni

- Un esempio:

Mario 275

su un foglio di carta sono due dati e non significano molto.

Se il foglio di carta viene fornito in risposta alla domanda “A chi mi devo rivolgere per il problema X; qual è il suo numero di telefono?”, allora i dati possono essere interpretati per fornire informazione e arricchire la conoscenza.

# Basi di dati

- Accezione generica, metodologica:
  - Insieme organizzato di dati utilizzati per il supporto allo svolgimento delle attività di un ente pubblico o privato (azienda, società, associazione, ecc..., ma anche solo un ufficio)
- Accezione specifica, metodologica e tecnologica:
  - Insieme di dati gestito da un DBMS

# Base di dati

Approfondiamo l'accezione generica, metodologica.

- Una **base di dati** è un insieme di dati atomici strutturati e persistenti, raggruppati in **insiemi omogenei in relazione tra loro** organizzati con la minima ridondanza per essere utilizzati da applicazioni diverse in modo sicuro e controllato.



# Classificazione

- I dati vengono prima di tutto raggruppati secondo il sistema della **classificazione** ovvero per omogeneità
- Esempio: in un ospedale abbiamo dottori, pazienti, reparti, attrezzature mediche, cancelleria, ecc.
- Le basi di dati offrono gli strumenti per classificare la realtà del dominio di interesse (e al relativo sistema informativo)

# Relazioni tra classi

- Le classi devono poter essere messe in relazione fra loro
- Esempio: correlazione tra medici e reparti ma non tra attrezzature mediche e pazienti

# Le basi di dati sono ... condivise

- Ogni organizzazione (specie se grande) è divisa in settori o comunque svolge diverse attività
- Ciascun settore/attività ha un (sotto)sistema informativo (non necessariamente disgiunto dagli altri sottosistemi)

# Un esempio /1



CCS Ingegneria Informatica Orario - Netscape

File Edit View Go Communicator Help

**UNIVERSITA' DEGLI STUDI DI CHISSADOVE**

**Corso di Studi in Ingegneria Informatica**

---

**ORARIO DELLE LEZIONI PER L'ANNO  
ACCADEMICO 1999-2000**

INSEGNAMENTO	Docente	Aula	Orario
Analisi matematica I	Luigi Neri	N1	8:00-9:30
Basi di dati	Piero Rossi	N2	9:45-11:15
Chimica	Nicola Mori	N1	9:45-11:30
Fisica I	Mario Bruni	N1	11:45-13:00
Fisica II	Mario Bruni	N3	9:45-11:15
Sistemi informativi	Piero Rossi	N3	8:00-9:30

Document: Done

# Un esempio /2



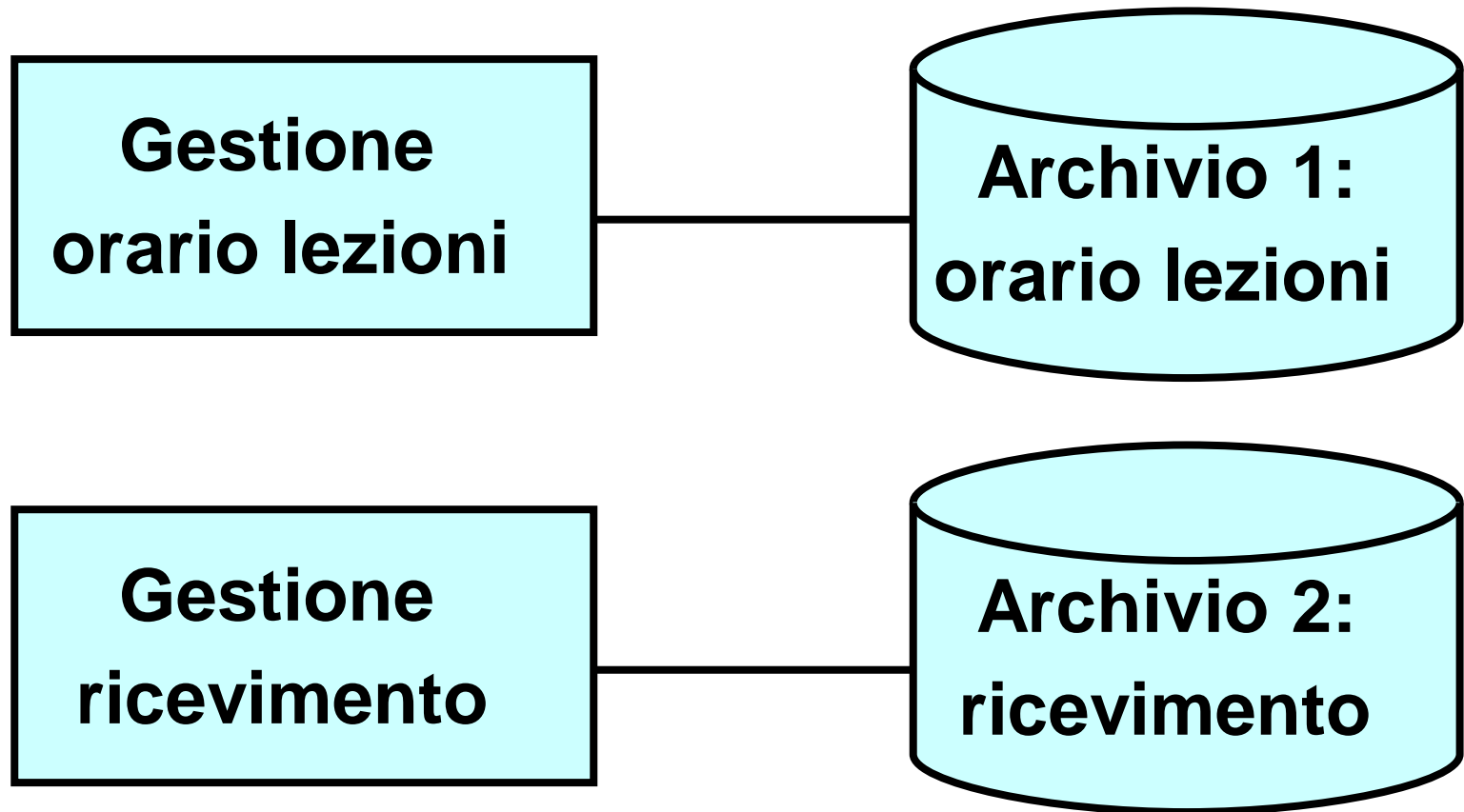
**UNIVERSITA' DEGLI STUDI CHISSADOVE**

**Corso di Studi in Ingegneria Informatica**

**Orario di ricevimento dei docenti**

DOCENTE	INSEGNAMENTI	ORARIO
Mario BRUNI	Fisica I Fisica II	Martedì' 10-12
Luigi NERI	Analisi matematica I	Lunedì' 12-13
Piero ROSSI	Basi di dati Sistemi informativi	Giovedì' 11-13
Nicola MORI	Chimica	Martedì' 16-18

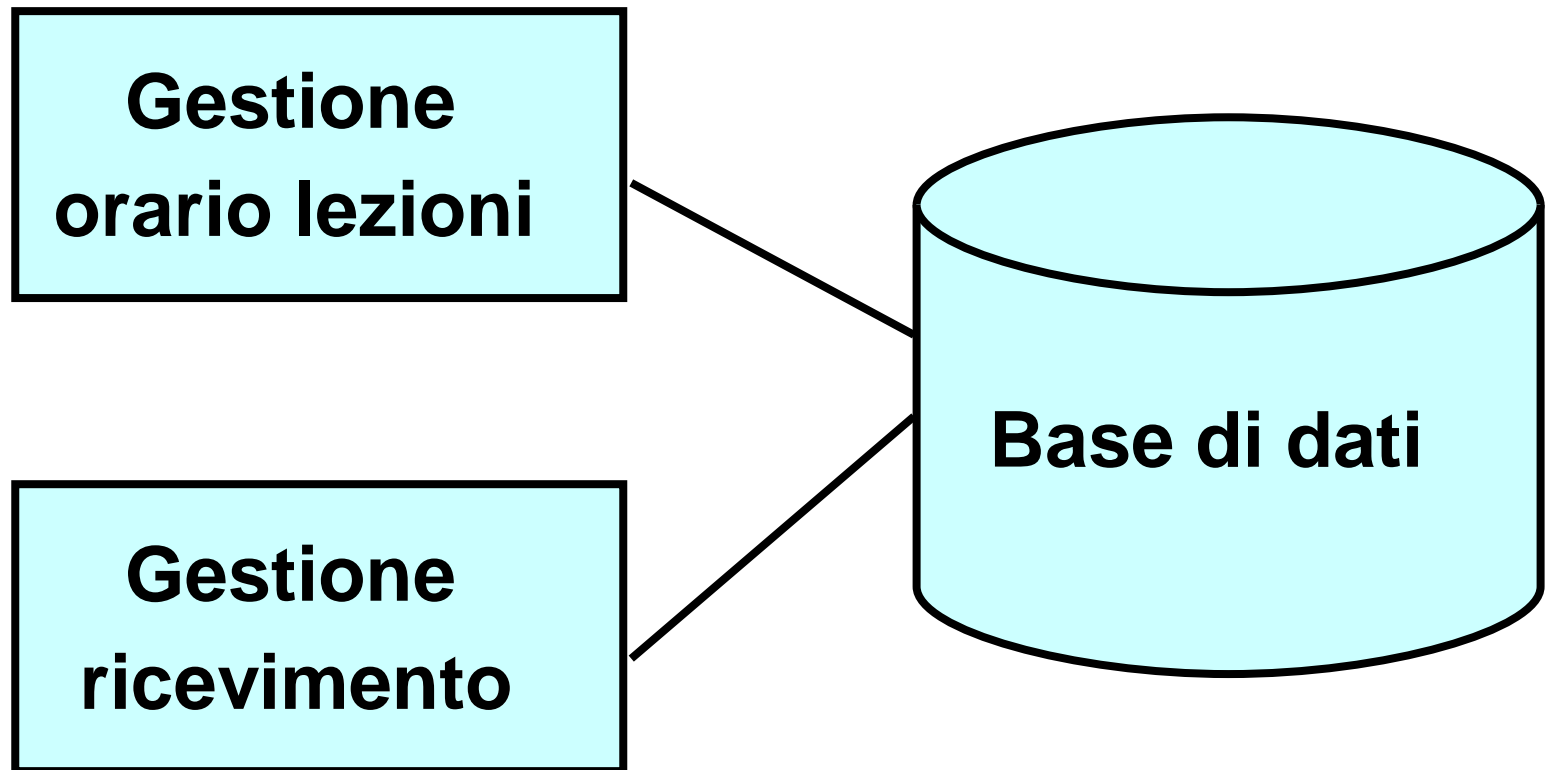
# Archivi vs basi di dati



# Problemi

- Ridondanza:
  - informazioni ripetute
- Rischio di incoerenza:
  - versioni diverse delle stesse informazioni possono non coincidere

# Basi di dati





# Le basi di dati sono condivise

- Una base di dati è una risorsa integrata, condivisa fra applicazioni.

Conseguenze:

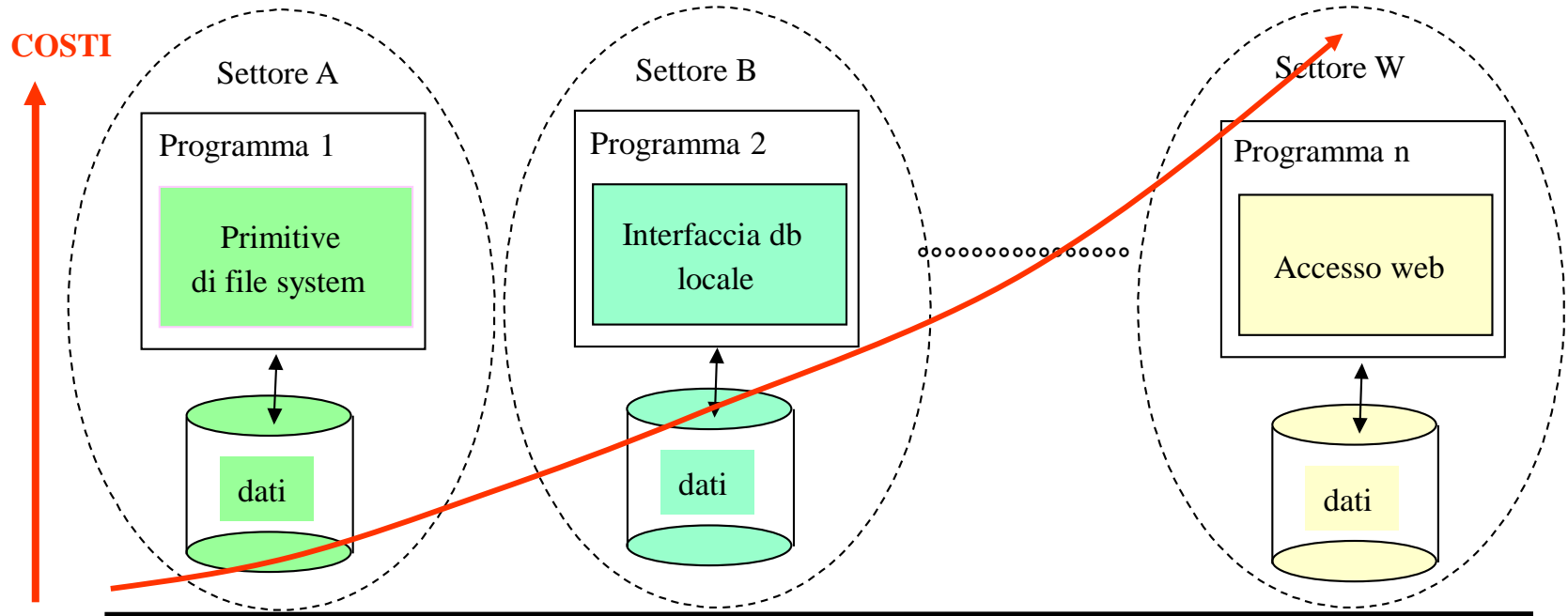
- Attività svolte da soggetti diversi su dati condivisi:

**meccanismi di autorizzazione**

- Accessi da parte di più soggetti ai dati condivisi:

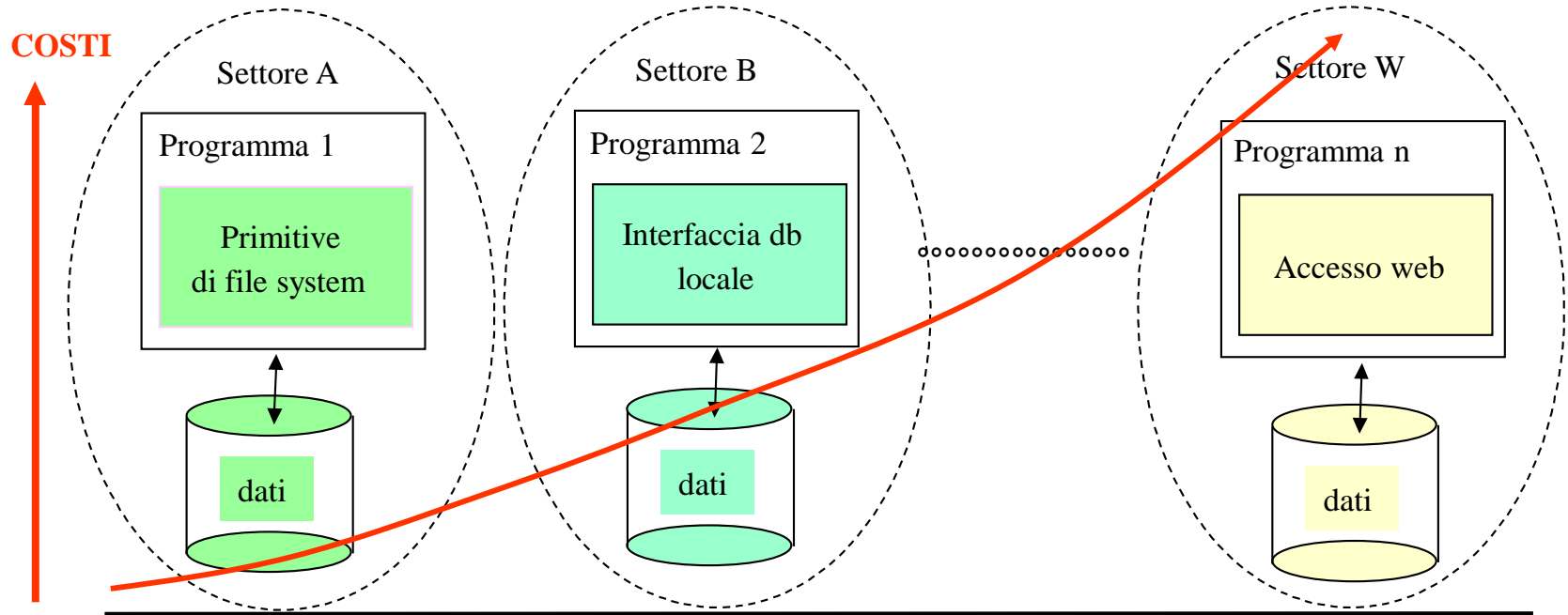
**controllo della concorrenza**

# Sviluppo basato su archivi indipendenti



**Si giunge rapidamente al collasso del sistema informatico e quindi informativo!**

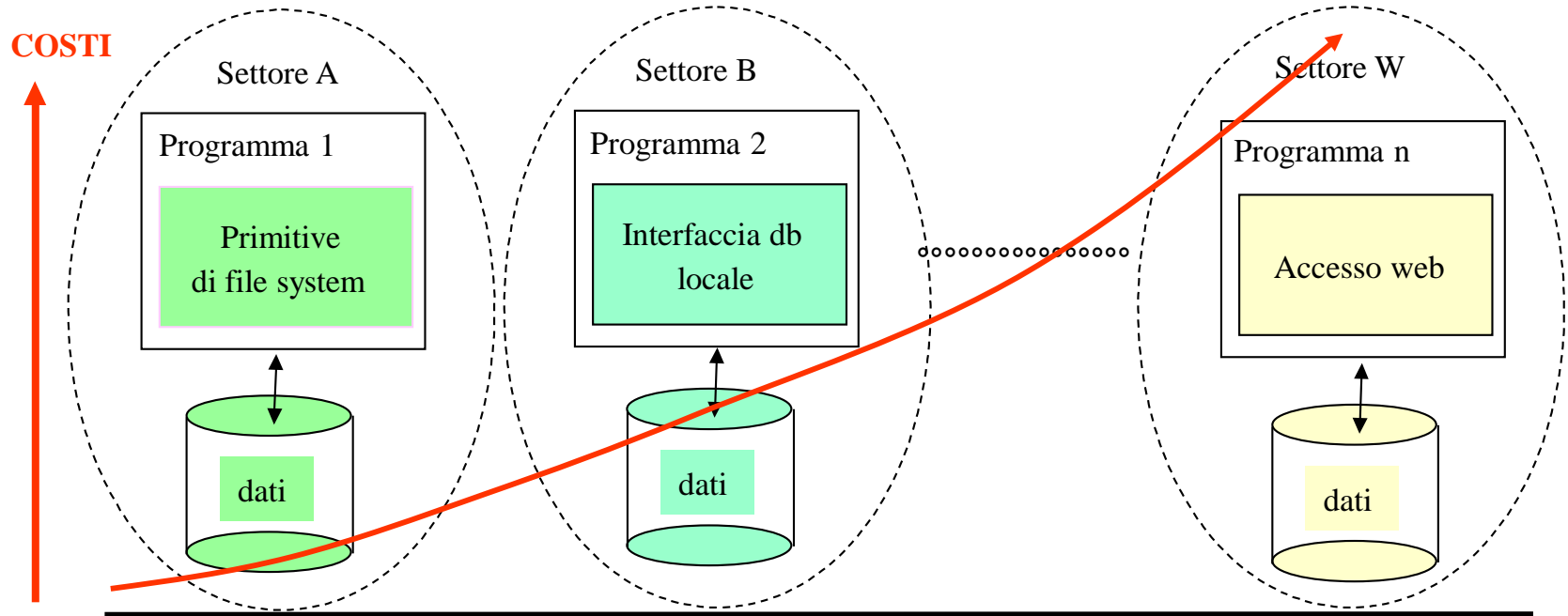
# Sviluppo basato su archivi indipendenti



## Criticità

- *Eterogeneità dei sistemi*: fornitori e tecnologie diverse che inducono costi di gestione
- *Difficoltà nell'interscambio di dati*: incompatibilità di formati e incompletezze
- *Ridondanza e incoerenza* delle informazioni
- *Assenza di una visione unitaria*

# Sviluppo basato su archivi indipendenti



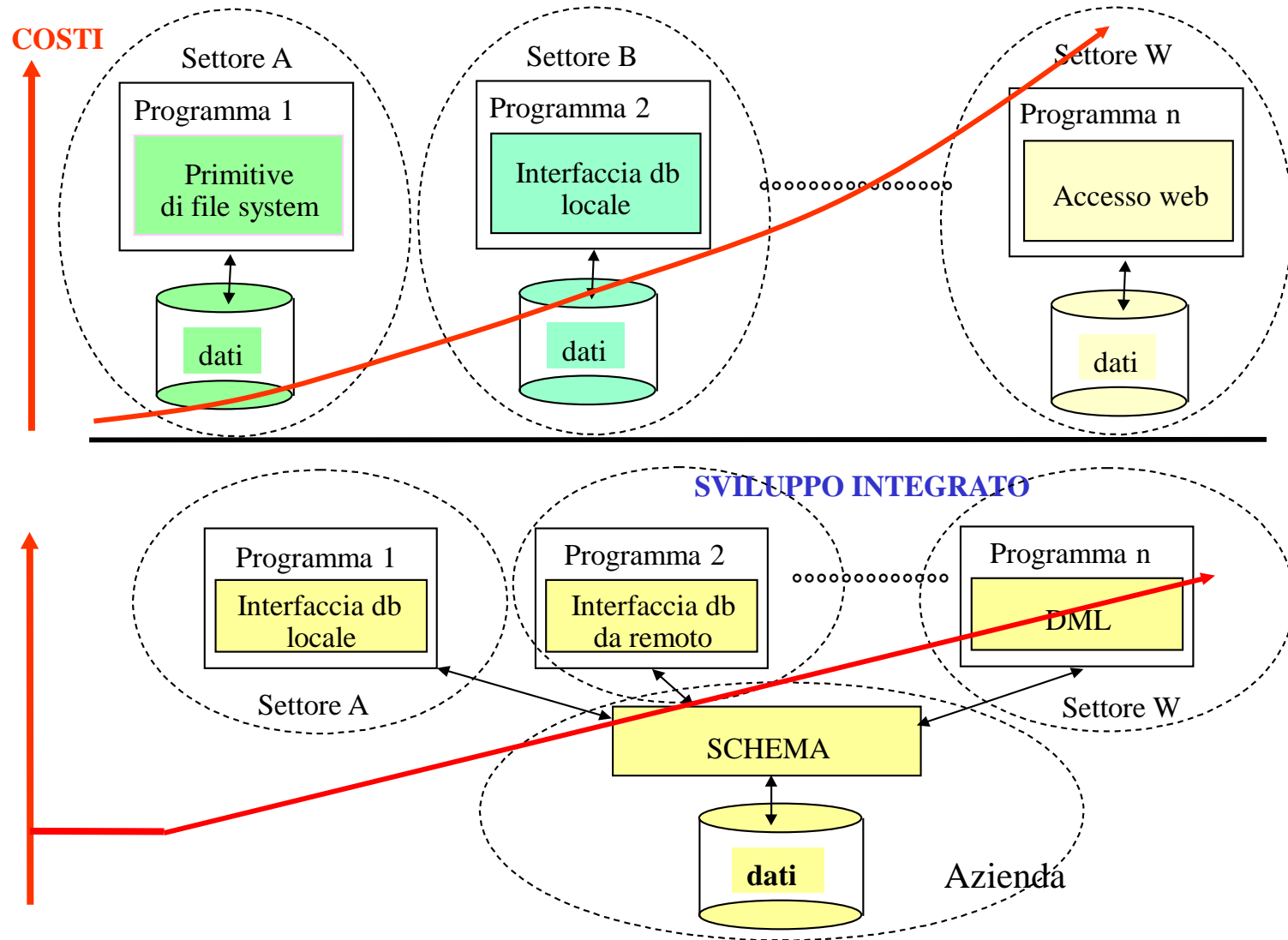
**Crescita quadratica della complessità!**

# Sviluppo integrato

1. Data la struttura organizzativa si inizia con un'analisi attenta e il più possibile esaustiva del sistema informativo indipendentemente dalle applicazioni
2. Definiti lo schema e i dati e le informazioni di interesse dell'organizzazione si sviluppa l'applicazione
3. L'applicazione si interfacerà con il DBMS (sistema di gestione della basi di dati) in base alle sue esigenze

**Prima i dati, poi le applicazioni!**

# Sviluppo integrato



# Data Base Management System (DBMS)

**DBMS:** sistema centralizzato o distribuito che consente di:

- **definire schemi** di basi di dati (necessari per interpretare i dati),
- specificare i **vincoli di integrità** dei dati,
- **scegliere le strutture** dati per la memorizzazione e l'accesso ai dati,
- **memorizzare, recuperare e modificare** i dati (interattivamente o da programma) da utenti **autorizzati**.

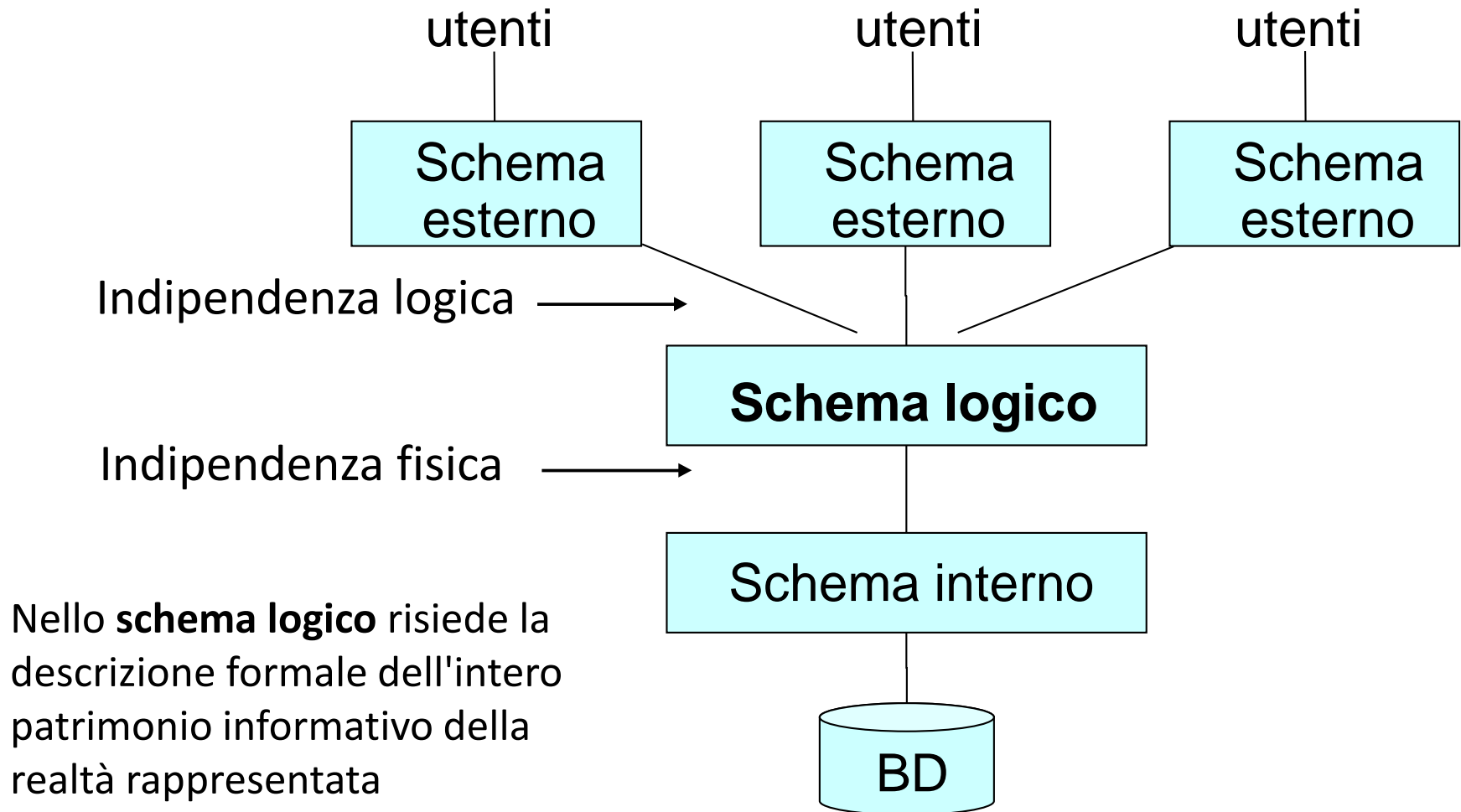
# Data Base Management System (DBMS)

## Componenti tipici:

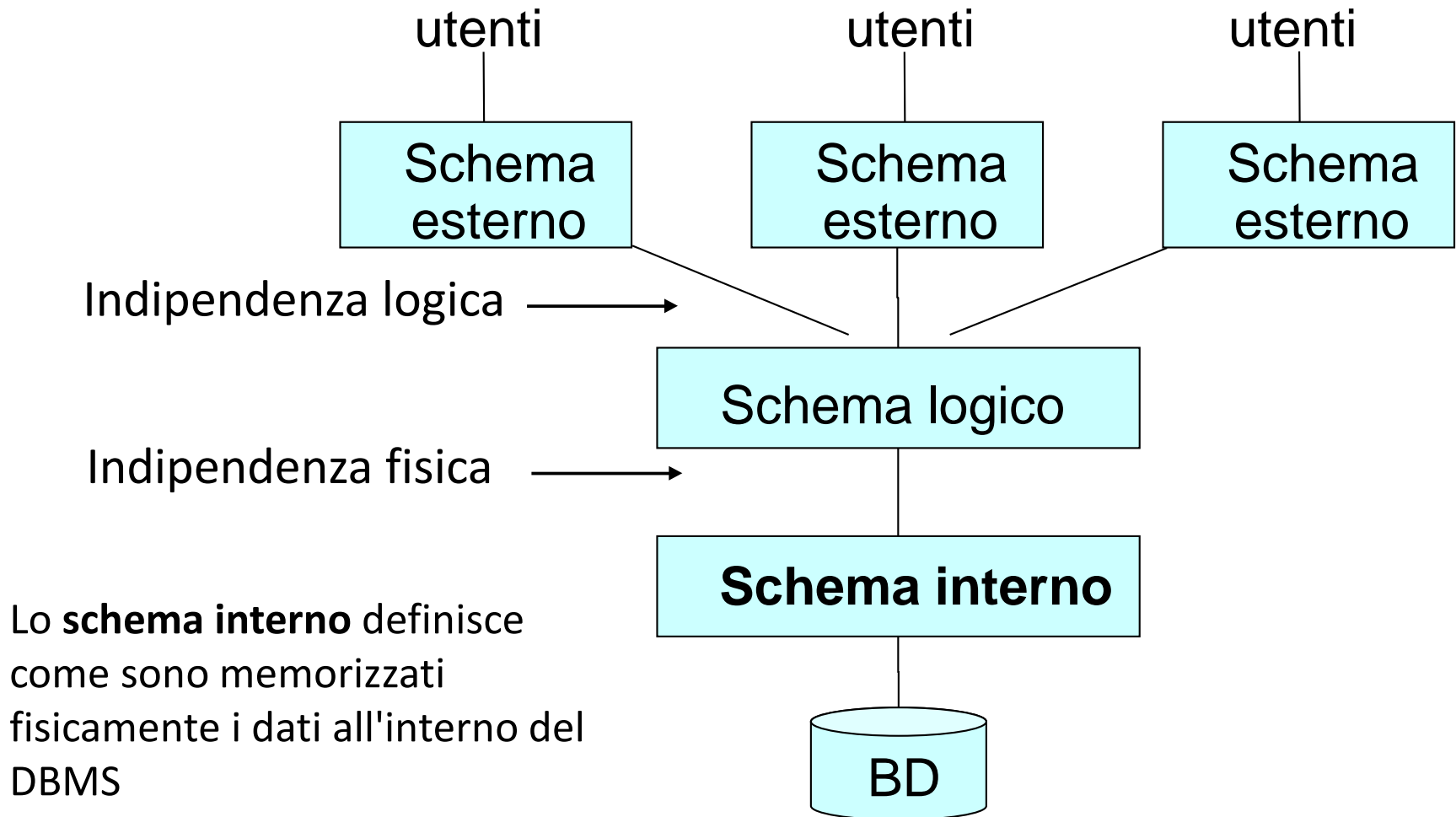
- Linguaggi per definire gli schemi logici dei dati (DDL, data definition language);
- Vincoli d'integrità;
- Linguaggi per manipolare i dati (DML, data manipulation language): interrogare, inserire, cancellare, modificare;
- Strutture per un accesso efficiente;
- Gestione automatica delle **transazioni**;
- Meccanismi di autorizzazione;
- Concorrenza;
- Ripristino da guasti;
- *Meccanismi di protezione dei dati personali (privacy).*



# Architettura standard (ANSI/SPARC) a tre livelli per DBMS - 1975



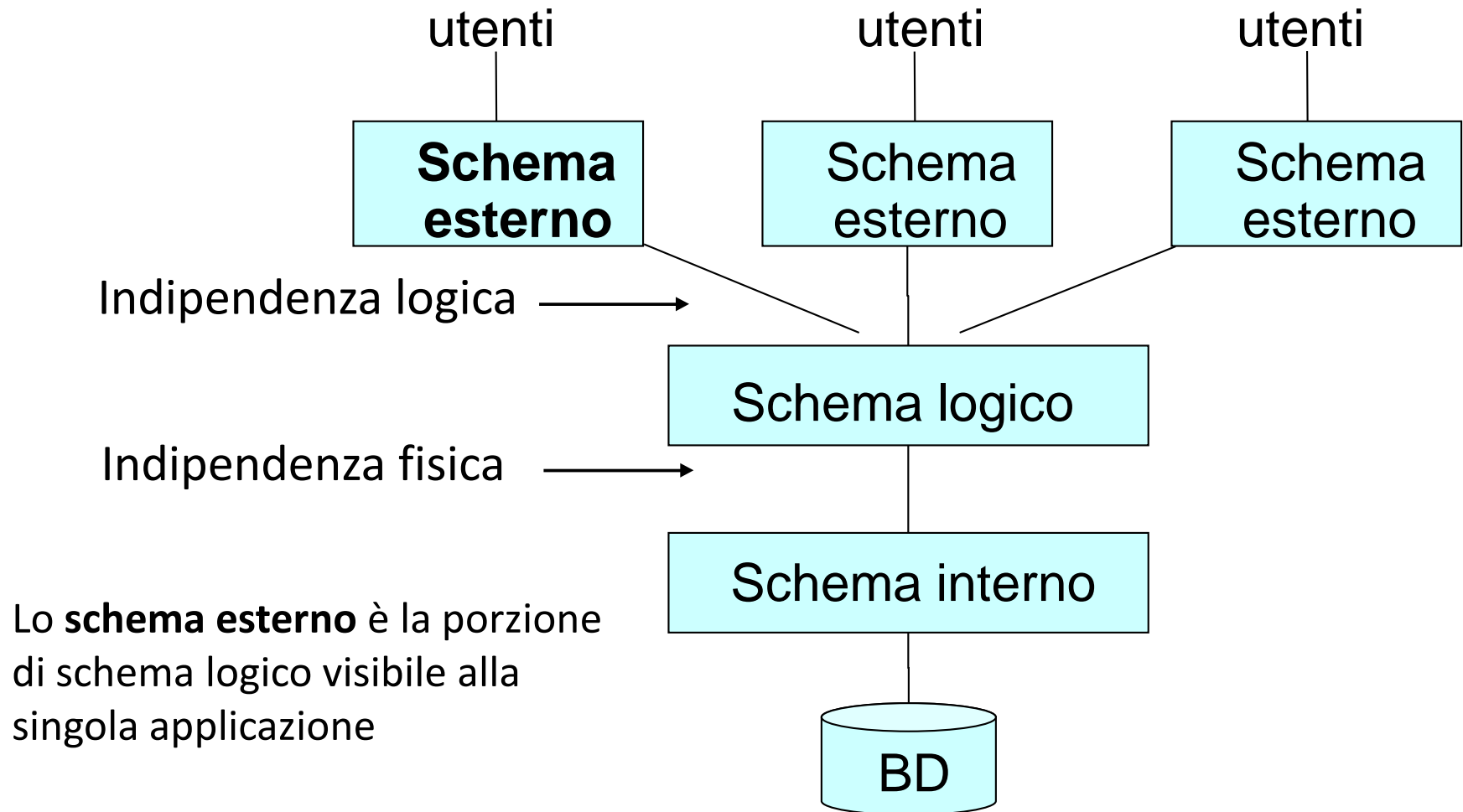
# Architettura standard (ANSI/SPARC) a tre livelli per DBMS - 1975



# Architettura standard (ANSI/SPARC) a tre livelli per DBMS

- **Indipendenza fisica** (garantita nei sistemi attuali)  
I programmi applicativi o le attività interattive devono rimanere logicamente inalterate a fronte di eventuali modifiche delle strutture di memorizzazione dei dati o delle vie di accesso.

# Architettura standard (ANSI/SPARC) a tre livelli per DBMS - 1975



# Dinamicità del patrimonio informativo

Lo schema logico è la descrizione del patrimonio informativo dell'organizzazione

In generale il patrimonio informativo è mediamente stabile, ma può essere rivisto e ampliato

- Nuove normative
- Nuovi requisiti
- ...

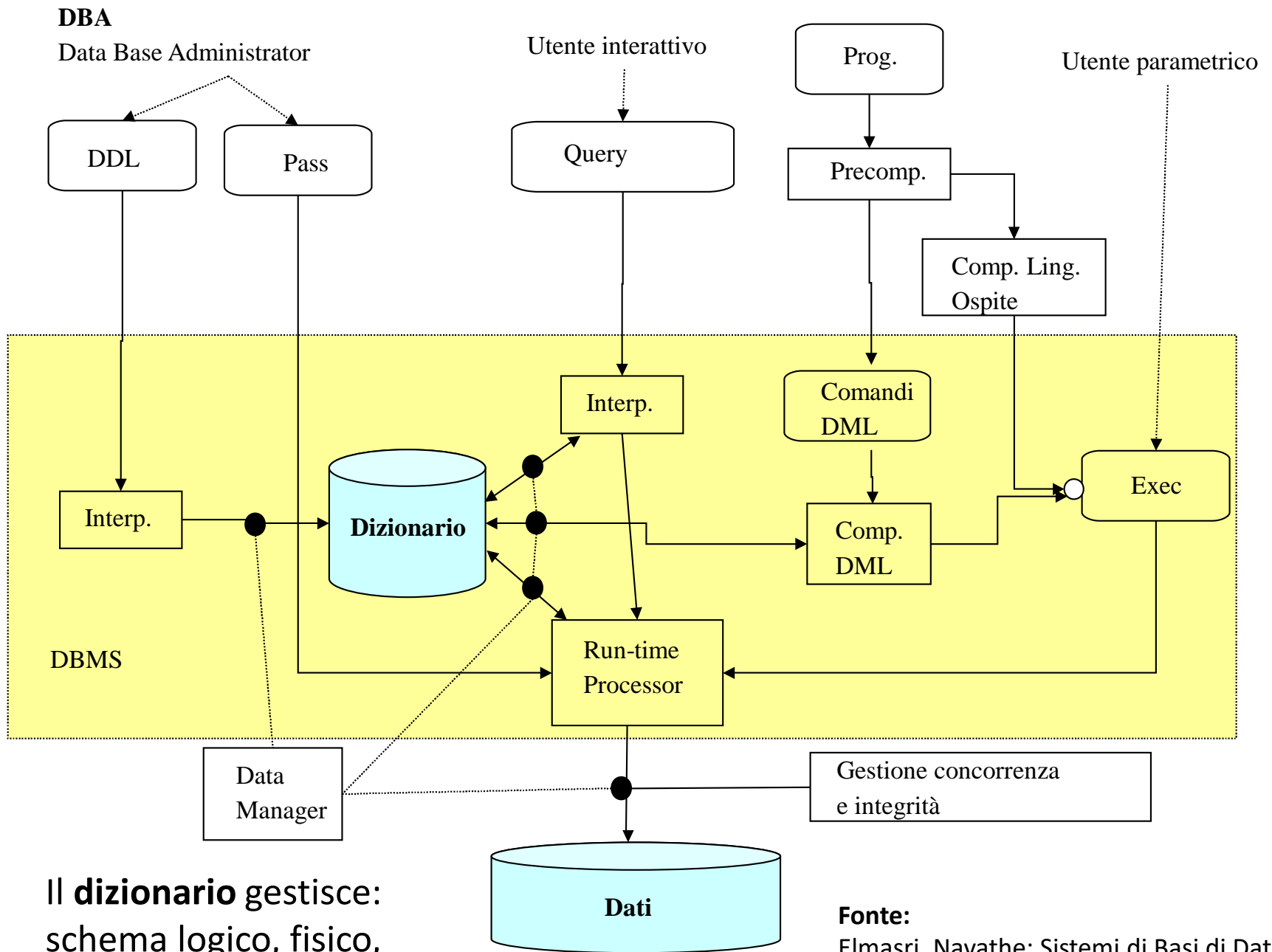
Se uno schema esterno non fa riferimento a porzioni modificate dello schema logico, la relativa applicazione continua a funzionare regolarmente

# Architettura standard (ANSI/SPARC) a tre livelli per DBMS

- **Indipendenza fisica** (garantita nei sistemi attuali)  
I programmi applicativi o le attività interattive devono rimanere logicamente inalterate a fronte di eventuali modifiche delle strutture di memorizzazione dei dati o delle vie di accesso.
- **Indipendenza logica** (garantita solo in parte)  
I programmi applicativi o le attività interattive devono rimanere logicamente inalterate a fronte di variazioni dello schema logico che preservano le informazioni originarie.

# Modelli logici

- Modello dei dati: un insieme di meccanismi di astrazione per definire una base di dati, con associato un insieme predefinito di operatori e di vincoli d'integrità.
  - Il modello è **l'intermediario** che permette di cogliere e rappresentare la realtà
- Adottati nei DBMS esistenti per l'organizzazione dei dati
  - utilizzati dai programmi
  - indipendenti dalle strutture fisiche
- esempi: **relazionale, reticolare, gerarchico, a oggetti, modelli NoSQL** (chiave-valore, a grafo, document-oriented, tabulare)



Il **dizionario** gestisce:  
schema logico, fisico,  
mapping tra i due

**Fonte:**  
Elmasri, Navathe: Sistemi di Basi di Dati  
Addison – Wesley, 2001



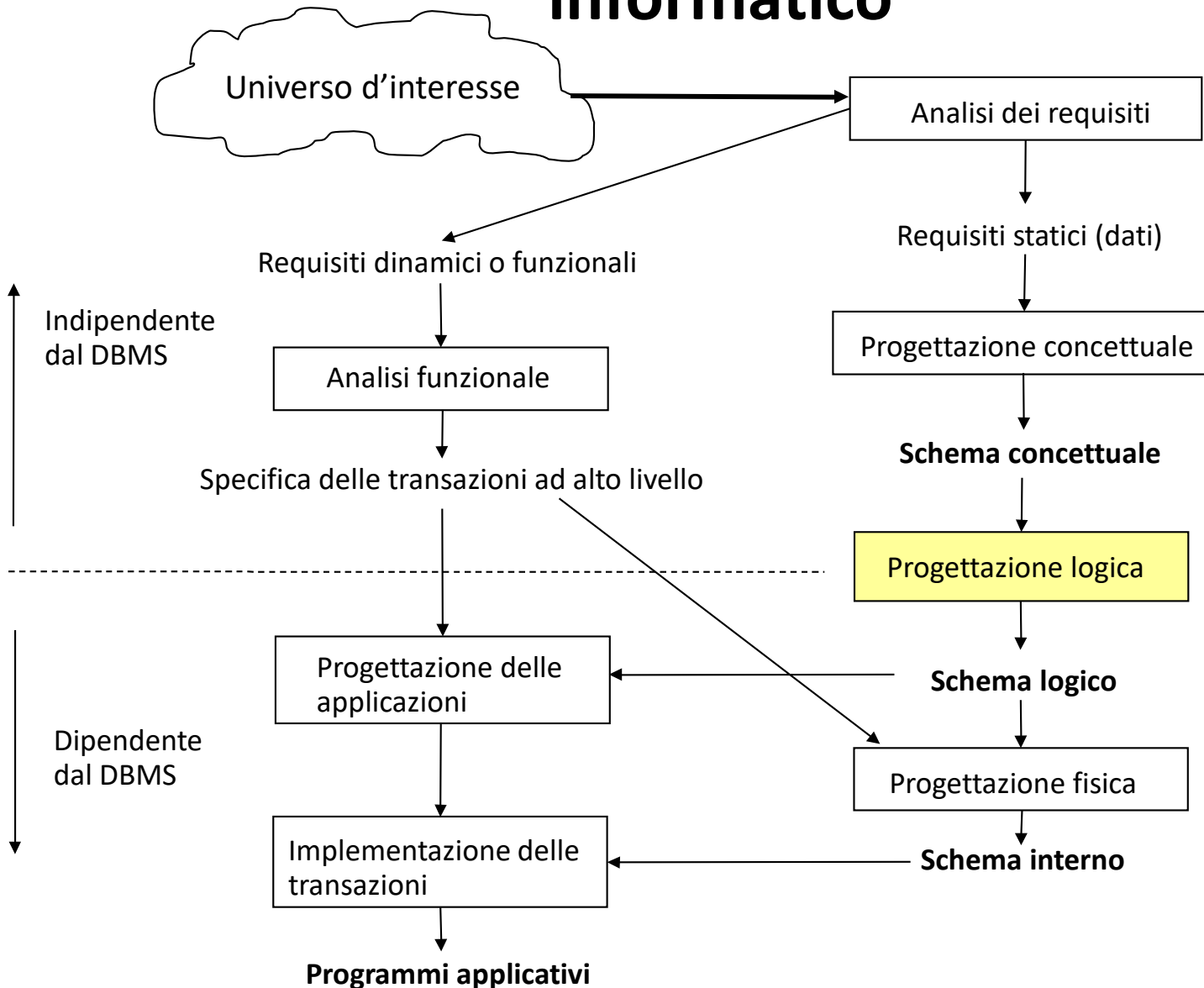
# File vs Basi di dati /1

FILE: CRITICITA'	DBMS: VANTAGGI	STRUMENTI
Applicazioni + archivi = dati duplicati, ridondanza	Integrazione	Dati condivisi Schemi concettuali/logici
Sinonimie e omonimie	Standardizzazione	Nomi definiti dallo schema Schemi concettuali/logici
Integrità controllata da ogni applicazione	Controllo centralizzato dell'Integrità	Vincoli
Stesse informazioni aggiornate in tempi diversi	Consistenza: aggiornamenti visti subito da tutte le applicazioni	Normalizzazione; Transazioni ACID
Procedure empiriche di gestione dei malfunzionamenti	Gestione centralizzata dell'Affidabilità	Ripristino

# File vs Basi di dati /2

FILE: CRITICITA'	DBMS: VANTAGGI	STRUMENTI
Variazioni di formato o di tipo devono essere riportati in ogni applicazione	Indipendenza logica	Architettura ANSI-SPARC Solo applicazioni interessate subiscono modifiche
Modalità specifiche di accesso alle periferiche	Indipendenza fisica	Architettura ANSI-SPARC Vie di accesso/indici
Accesso ai dati solo via applicazione	Facilità d'uso	Algebra e Calcolo relazionale; SQL
Dati usati esclusivamente da ogni applicazione	Concorrenza	Accesso simultaneo Serializzatore
Sicurezza gestita dalle singole applicazioni	Sicurezza centralizzata	Privilegi (Grant)

# Schema metodologico di sviluppo di un sistema informatico



# Modelli concettuali

- Permettono di rappresentare i dati in modo indipendente da ogni sistema
  - cercano di descrivere i concetti del mondo reale
  - sono utilizzati nelle fasi preliminari di progettazione
- Il più diffuso è il modello **Entity-Relationship (ER)** tradotto in italiano con **Entità-Associazione**

# Oltre il modello relazionale

## **Nuove applicazioni:**

- CASE: Computer Aided Software Engineering;
- CAD: Computer Aided Design (meccanico, elettronico, ...);
- CAM: Computer Aided Manufacturing;
- CAP: Computer Aided Publishing;
- Office Automation;
- Grafica / Multimedia;
- GIS;
- Controllo della produzione in tempo reale;
- Basi di conoscenza per applicazioni di A.I.;
- Software di sistema.

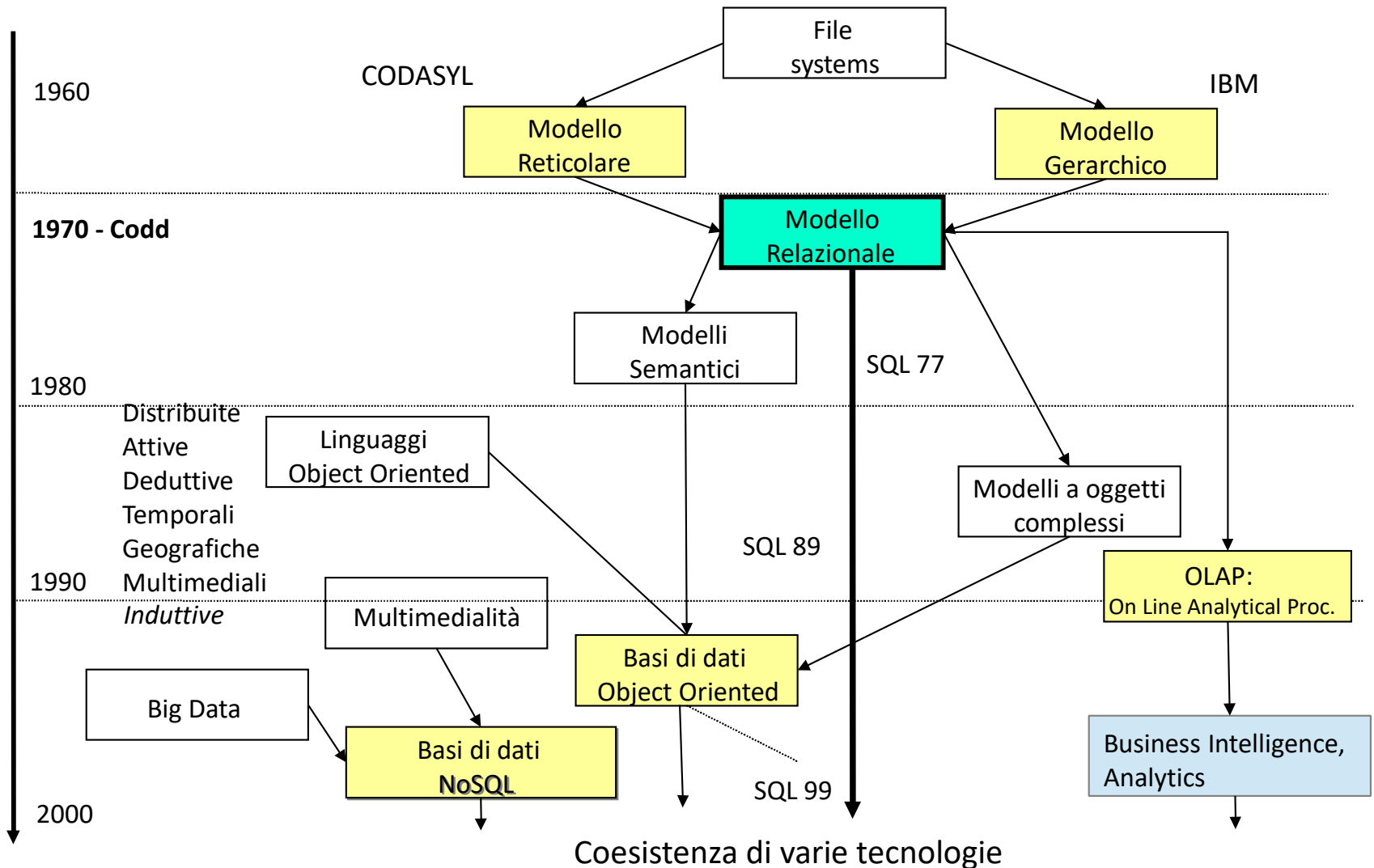
# Tipologie di applicazioni

Le applicazioni dei DBMS in generale possono essere classificate in tre categorie:

- **Applicazioni gestionali**  
caratterizzate da grandi quantità di dati di struttura semplice.
- **Applicazioni navigazionali complesse**  
quali applicazioni CAD e di telecomunicazioni che manipolano dati legati da relazioni complesse da attraversare in modo efficiente.
- **Applicazioni multimediali**  
che richiedono la manipolazione di immagini, testi e dati spaziali con operazioni specifiche dell'applicazione.

*Allo stato attuale né i DBMS relazionali né altri DBMS soddisfano pienamente tutte e tre le tipologie*

# Evoluzione delle basi di dati



**Stampare l'esempio di basi di dati  
presente in Moodle e portarlo a lezione!**