



**InformaticaUmanistica**

# Basi di dati e Laboratorio Progettazione Web

Docenti:

Giuseppe Amato e Chiara Renso

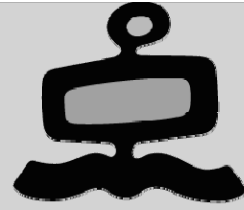
ISTI-CNR

a.a. 2009 – 2010

[http://www.cli.di.unipi.it/doku/doku.php/bdd-  
infuma/start](http://www.cli.di.unipi.it/doku/doku.php/bdd-infuma/start)



**UNIVERSITÀ DI PISA**



Informatica **U**manistica

# Basi di dati

*Giuseppe Amato*

*ISTI-CNR*

*a.a. 2009 – 2010*



UNIVERSITÀ DI PISA

# Docente

## ◆ Giuseppe Amato

- Ricercatore presso Istituto di Scienze e Tecnologie dell'Informazione "A. Faedo" del Consiglio Nazionale delle Ricerche
- [Giuseppe.amato@isti.cnr.it](mailto:Giuseppe.amato@isti.cnr.it)
- <http://www.nmis.isti.cnr.it/amato/>

## ◆ Pagina web del corso

- <http://www.cli.di.unipi.it/doku/doku.php/bdd-infuma/start>

**Ci sono dei fondamenti dell'informatica che dovrebbero stare nel bagaglio dell'umanista?**

- ◆ **Si danno tante risposte a questa domanda, molte di tipo tecnologico-pratico**
  - ◆ **La risposta che questo corso cerca di dare è metodologica**
  - ◆ **La nostra convinzione è che un(a) umanista dovrebbe padroneggiare i principi per la rappresentazione dell'informazione e della conoscenza**
-

# La tecnologia evolve rapidamente, ma ..

- ◆ **La tecnologia avanzata di oggi sarà la tecnologia superata di domani**
- ◆ **Per comprendere (e magari generare) i cambiamenti tecnologici occorre una profonda conoscenza metodologica e scientifica di base**

- ◆ ***Quelli che s'innamora di pratica senza scienza son come 'l nocchieri ch'entra in navilio senza timone o bussola, che mai ha certezza dove si vada.***
  - ◆ **Leonardo da Vinci, *Frammenti letterari e filosofici*, Barbera, Firenze, 1809**
-

## Obiettivi del corso

- ◆ **Fornire le conoscenze e le competenze necessarie alla rappresentazione e al trattamento consapevole delle informazioni pertinenti alle scienze umane.**
  - ◆ **Il corso prevede un mix di fondamenti di portata generale e di esercitazioni pratiche relative all'uso di tecnologie informatiche per la rappresentazione dell'informazione.**
-

# Quale informazione?

- ◆ **Informazione = Contenuto + Struttura**
- ◆ **Contenuto: l'oggetto che si intende comunicare/rappresentare attraverso**
  - Testo, linguaggio naturale
  - Immagine, linguaggio visuale
  - Suono, linguaggio audiovisivo
- ◆ **Struttura: la organizzazione della comunicazione, che esplicita, attraverso astrazioni, i tratti salienti della rappresentazione**



# Informazione = contenuto + struttura

## ◆ **Informazione non strutturata**

- Molto contenuto, poca struttura
- Un romanzo e il suo indice (magari solo capitoli numerati)

## ◆ **Informazione strutturata**

- Molta struttura, contenuto rigidamente codificato
- I sistemi informativi delle organizzazioni, le basi di dati

# Informazione strutturata vs. non strutturata

## ◆ Due situazioni estreme

### ◆ Informazione strutturata

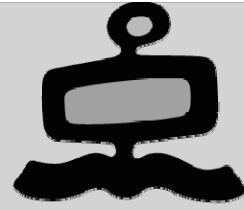
- - Rigida a costruirsi e ad alimentarsi
- + Adatta e flessibile ad interrogarsi per estrarre conoscenza

### ◆ Informazione non strutturata (testo)

- + Flessibile a costruirsi e ad alimentarsi
- - Difficile ad interrogarsi per estrarre conoscenza

## Fra i due estremi ...

- ◆ **L'informazione semi-strutturata, vari possibili compromessi fra contenuto e struttura**
  - ◆ **Rappresentazioni dell'informazione, dei documenti multimediali, che cercano di combinare i vantaggi dei due estremi.**
  - ◆ **Sullo sfondo, il Web, una sorgente sterminata di informazione semi-strutturata (a struttura crescente) da cui è difficile estrarre conoscenza**
-



**InformaticaUmanistica**

# Contenuti (e struttura ...) del corso



**UNIVERSITÀ DI PISA**

## BDD – obiettivi

### ◆ **Formare le conoscenze e le competenze di base per**

- la rappresentazione,
- l'organizzazione,
- l'interrogazione ed
- il recupero

**dell'informazione mediante la tecnologia delle basi di dati**

---

# BDD – contenuti

- ◆ **Le basi di dati relazionali**
  - ◆ **L'algebra relazionale e il linguaggio di interrogazione SQL**
  - ◆ **Il modello entità-relazioni e la progettazione concettuale**
-

## Testi di consultazione

- ◆ **Atzeni, Ceri, Paraboschi, Torlone. *Basi di dati*. McGraw-Hill, 2002.**
  - ◆ **Albano, Ghelli, Orsini. *Basi di dati relazionali e a oggetti*. Zanichelli, 1997.**
  - ◆ **El Masri, Navathe. *Sistemi di basi di dati*. Addison Wesley, 1999.**
  - ◆ **Albano. *Sistemi per l'archiviazione e il recupero delle informazioni*. dispensa, 1999.**
-

# Materiale didattico

## ◆ **Lucidi: grazie a**

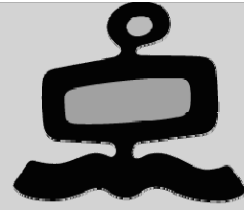
- Gianni Mecca, Università della Basilicata
- Antonio Albano e Paolo Manghi, Università di Pisa
- Domenico Saccà, Università della Calabria

## ◆ **Dispense ed esercizi**

## ◆ **Materiale reso disponibile on-line**

---





**InformaticaUmanistica**

# Una rapida panoramica sul corso

*Dall'informazione strutturata alla  
informazione semi-strutturata*



**UNIVERSITÀ DI PISA**

# Modello dei Dati

- ◆ **Insieme di meccanismi di astrazione per la rappresentazione di informazioni**

# Modello relazionale dei dati

- ◆ **i dati sono organizzati in “tabelle”**
  - ◆ **la tabella è un insieme di “record” (ennuple)**
  - ◆ **con un insieme di attributi**
  - ◆ **di tipi opportuni (numeri, stringhe, date ...)**
  - ◆ **i dati in tabelle diverse sono correlati sulla base dei valori**
-

**studenti**

<b>Matricola</b>	<b>Cognome</b>	<b>Nome</b>	<b>Data di nascita</b>
<b>6554</b>	<b>Pinco</b>	<b>Pallino</b>	<b>05/12/1978</b>
<b>8765</b>	<b>Neri</b>	<b>Paolo</b>	<b>03/11/1976</b>
<b>9283</b>	<b>Verdi</b>	<b>Luisa</b>	<b>12/11/1979</b>
<b>3456</b>	<b>Rossi</b>	<b>Maria</b>	<b>01/02/1978</b>

**esami**

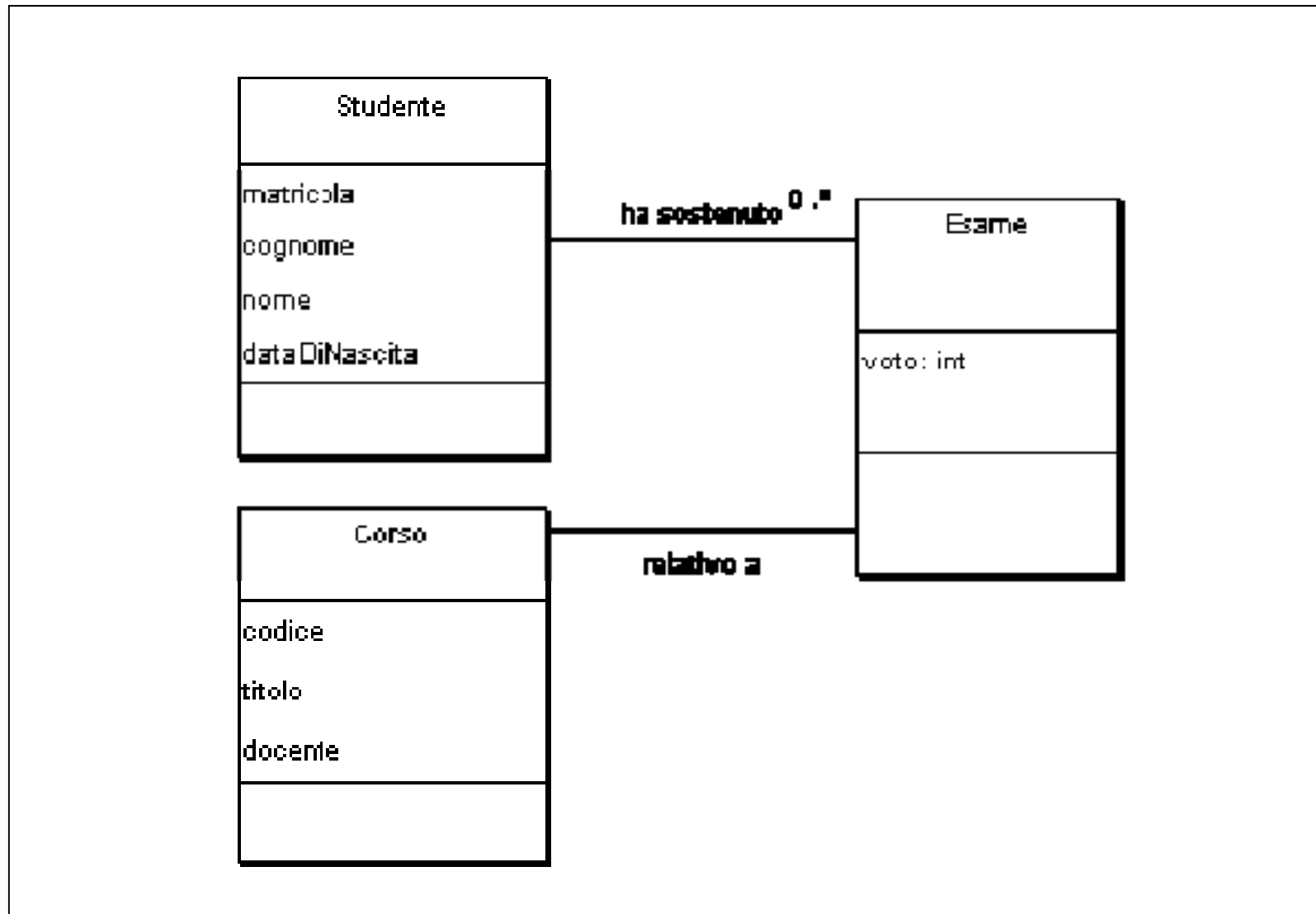
<b>Studente</b>	<b>Voto</b>	<b>Corso</b>
<b>3456</b>	<b>30</b>	<b>04</b>
<b>3456</b>	<b>24</b>	<b>02</b>
<b>9283</b>	<b>28</b>	<b>01</b>
<b>6554</b>	<b>26</b>	<b>01</b>

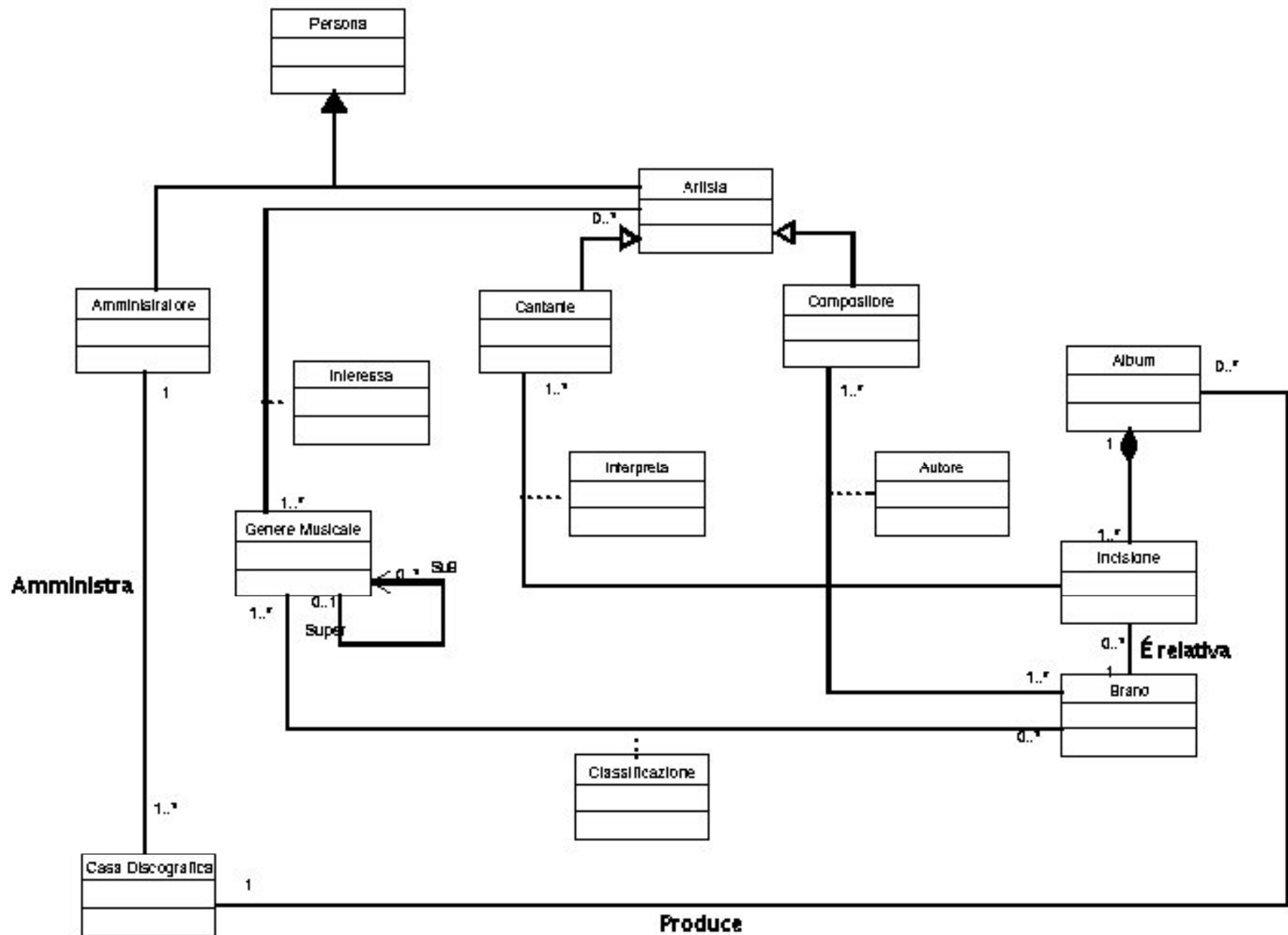
**corsi**

<b>Codice</b>	<b>Titolo</b>	<b>Docente</b>
<b>01</b>	<b>Analisi</b>	<b>Sempronio</b>
<b>02</b>	<b>Chimica</b>	<b>Bruni</b>
<b>04</b>	<b>Chimica</b>	<b>Verdi</b>

# Modello concettuale dei dati

- ◆ **i dati sono organizzati in classi**
  - ◆ **ogni classe rappresenta un insieme di oggetti**
  - ◆ **con un insieme di proprietà**
  - ◆ **le classi sono organizzate in ontologie**
  - ◆ **oggetti diversi sono correlati sulla base di identificatori**
-

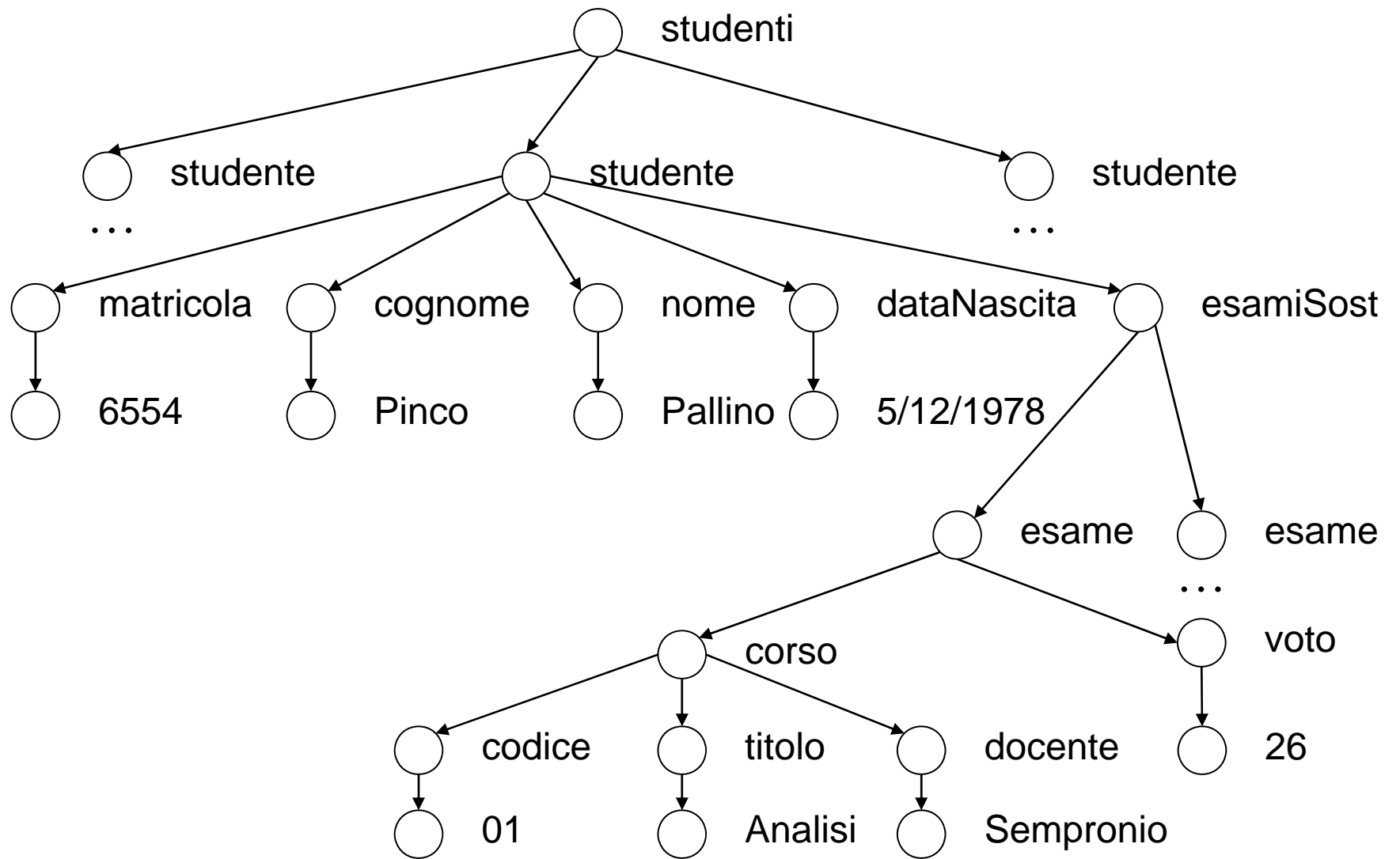




## Modello semi-strutturato: XML

- ◆ **i dati sono organizzati in strutture gerarchiche (alberi)**
  - ◆ **ogni albero ha un insieme di nodi (elementi)**
  - ◆ **oggetti diversi sono correlati sulla base di relazioni di contenimento**
-





```
<xml version="1.0" ?>
<studenti>
  <studente>
    <matricola>6554</matricola>
    <cognome>Pinco</cognome>
    <nome>Pallino</nome>
    <dataDiNascita>15/12/1978</dataDiNascita>
    <esamiSostenuti>
      <esame>
        <corso>
          <codice>01</codice>
          <titolo>Analisi</titolo>
          <docente>Giacomo</docente>
          <voto>26</voto>
        </corso>
      </esame>
      ...
    </esamiSostenuti>
  </studente>
</studenti>
```

# In ogni modello esistono

## ◆ Lo schema: **la descrizione della struttura**

- stabile nel tempo

## ◆ L'istanza: **i valori (cioè i dati)**

- variabile nel tempo

<b>studenti</b>	<b>Matricola</b>	<b>Cognome</b>	<b>Nome</b>	<b>Data di nascita</b>
	<b>6554</b>	<b>Pinco</b>	<b>Pallino</b>	<b>05/12/1978</b>
	<b>8765</b>	<b>Neri</b>	<b>Paolo</b>	<b>03/11/1976</b>
	<b>9283</b>	<b>Verdi</b>	<b>Luisa</b>	<b>12/11/1979</b>
	<b>3456</b>	<b>Rossi</b>	<b>Maria</b>	<b>01/02/1978</b>

# Linguaggio di interrogazione

- ◆ **Linguaggio per richiedere al DBMS (Database Management System) il recupero di informazione**
  - ◆ **Ogni modello ha i suoi linguaggi**
  - ◆ **DBMS relazionali: SQL, QBE**
  - ◆ **DBMS a oggetti: OQL**
  - ◆ **XML: XPath, XQuery**
-

# SQL: Un esempio

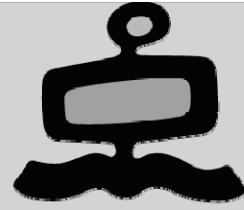
**SELECT** Matricola, Cognome, AVG(Voto)

**FROM** Studenti, Esami

**WHERE** Matricola = Studente

**GROUP BY** Matricola, Cognome

<b>Matricola</b>	<b>Cognome</b>	<b>AVG(voto)</b>
<b>6554</b>	<b>Pinco</b>	<b>26</b>
<b>8765</b>	<b>Neri</b>	<b>28</b>
<b>3456</b>	<b>Rossi</b>	<b>27</b>



**InformaticaUmanistica**

# Basi di Dati

*Sistemi per Basi di Dati:  
Introduzione*



**UNIVERSITÀ DI PISA**

# Sommario

## ◆ Introduzione

- Sistema Informativo e Sistema Informatico

## ◆ Definizione di DBMS

## ◆ Modello logico e modello fisico

- Panoramica sui modelli
- Schema e Istanza
- Linguaggi per basi di dati

## ◆ Indipendenza dei dati

---

# Database Management Systems

## ◆ **Oggetto di studio del corso**

- Sistemi di Gestione di Basi di Dati o “Database Management Systems” (DBMS)

## ◆ **Sono tipicamente utilizzati nelle organizzazioni complesse**

- molti dati da gestire
  - processi complessi
  - i dati sono cruciali per lo svolgimento delle attività
-



# DBMS

## ◆ In particolare

- i DBMS sono una componente fondamentale del sistema informatico di una grande organizzazione
  - che è una componente fondamentale del sistema informativo
  - che è una componente fondamentale del sistema organizzativo
-

# Sistema Organizzativo

## ◆ **Organizzazione complessa**

(es: azienda, banca, ente pubblico, ... ma anche fantacalcio...)

## ◆ **Sistema Organizzativo**

**complesso delle strutture, delle regole e delle procedure che regolano lo svolgimento delle attività**

**Es: banca – regolamento apertura c/c**

**anagrafe – procedure per la reg. nascite**

# Sistema Informativo

◆ **La parte del sistema organizzativo che presiede ai processi di raccolta, conservazione e utilizzo delle informazioni**

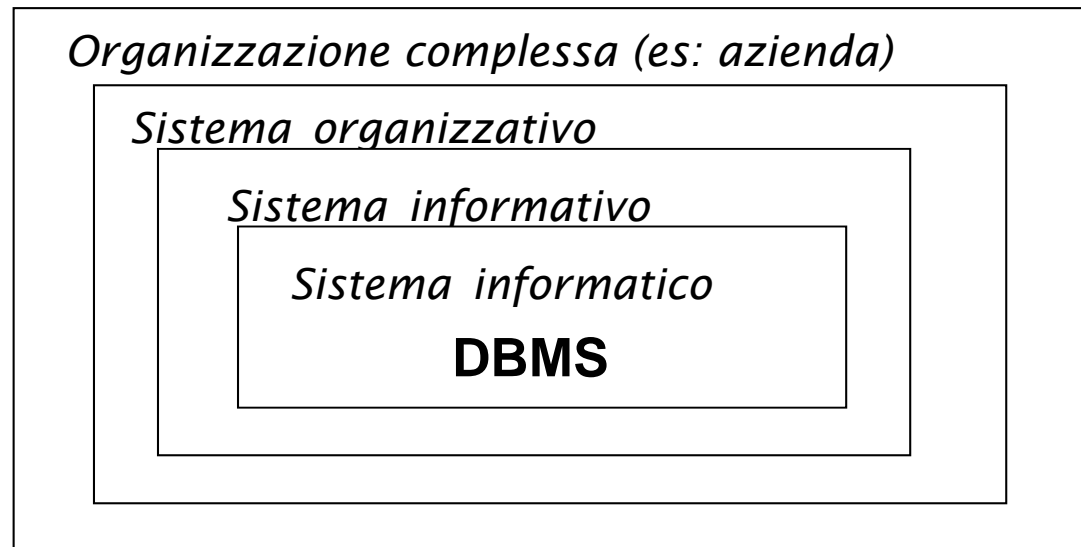
**Es: banca – registrazione prelievi su c/c  
emissione estratto conto  
anagrafe – registrazione nascita  
emissione certificato di nascita**

# Sistema Informatico

- ◆ **La porzione informatizzata del sistema informativo**
  - ◆ **Collezione di applicazioni che gestiscono i dati e i flussi informativi**
  - ◆ **In pratica, in molti casi:**
    - sistema informativo = sistema informatico
    - ma non è sempre così
-

# Sistema Informatico

## ◆ Struttura a livelli



obiettivo dei DBMS: il trattamento dei dati per ottenere informazione

# Dati e Informazioni (dal vocabolario)

- ◆ **Informazione: notizia, dato o elemento che consente di avere conoscenza più o meno esatta di fatti, situazioni, modi di essere.**
- ◆ **Dato: ciò che è immediatamente presente alla conoscenza, prima di ogni elaborazione; (in informatica) elementi di informazione costituiti da simboli che debbono essere elaborati.**

# Dati e Informazioni

- ◆ **I dati devono essere interpretati per diventare informazioni (conoscenza)**

**Esempio: 1 3 10 12 17 20 21 30**

- ◆ **Risultati del Totogol concorso 28 del 3 marzo 2002**

**1 3 10 12 17 20 21 30**

**Quote 8pt: € 215.594,00**

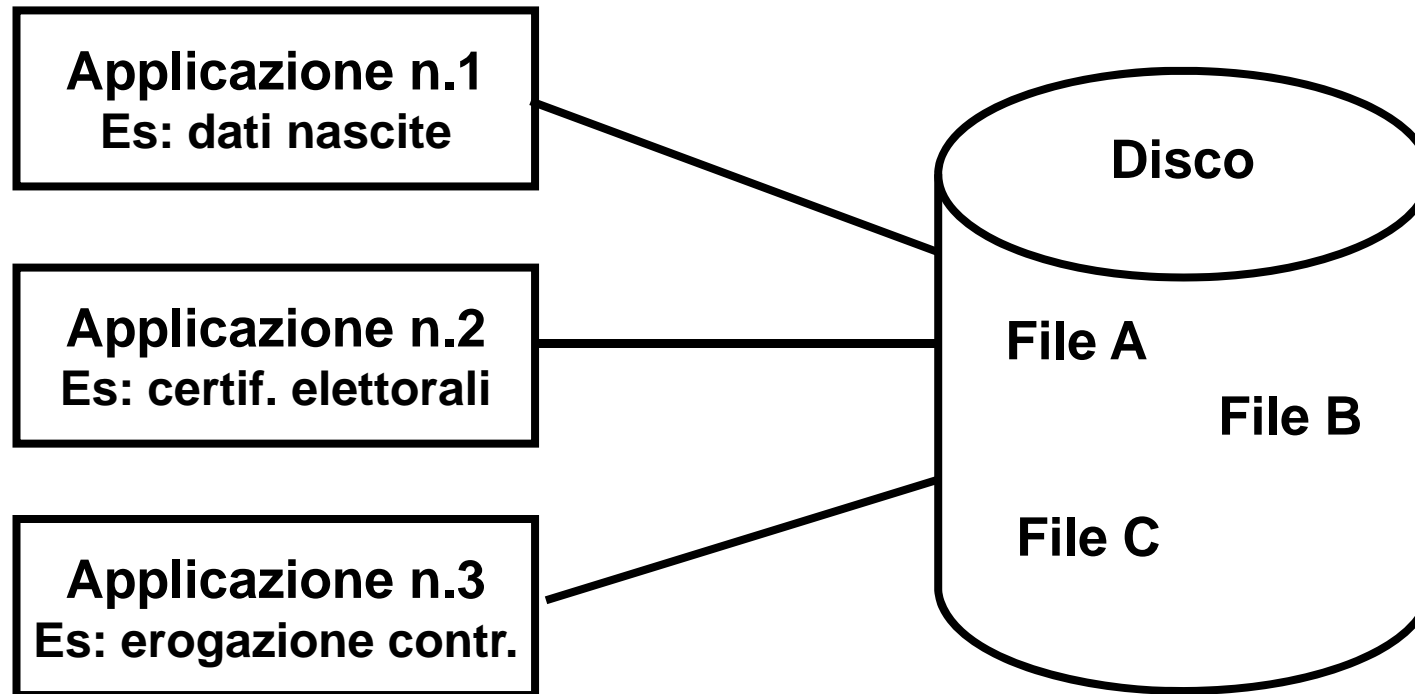
# Architettura dei Sistemi Informatici

- ◆ **I sistemi informatici sono presto stati considerati fondamentali nelle grandi organizzazioni**
  - ◆ **Esistono da prima che esistessero i DBMS**
  - ◆ **Erano basati su un'architettura diversa rispetto a quella attuale**
-



# Architettura Tradizionale dei S. I.

## ◆ Applicazioni e file



# Problemi ed Esigenze

- ◆ **Dimensioni dei dati**
  - ◆ **Efficienza**
  - ◆ **Condivisione dell'accesso**
    - ridondanza
  - ◆ **Affidabilità**
  - ◆ **Sicurezza**
-



# Database Management System (DBMS)

## ◆ Sistema per gestire

- grandi collezioni di dati persistenti

### in modo

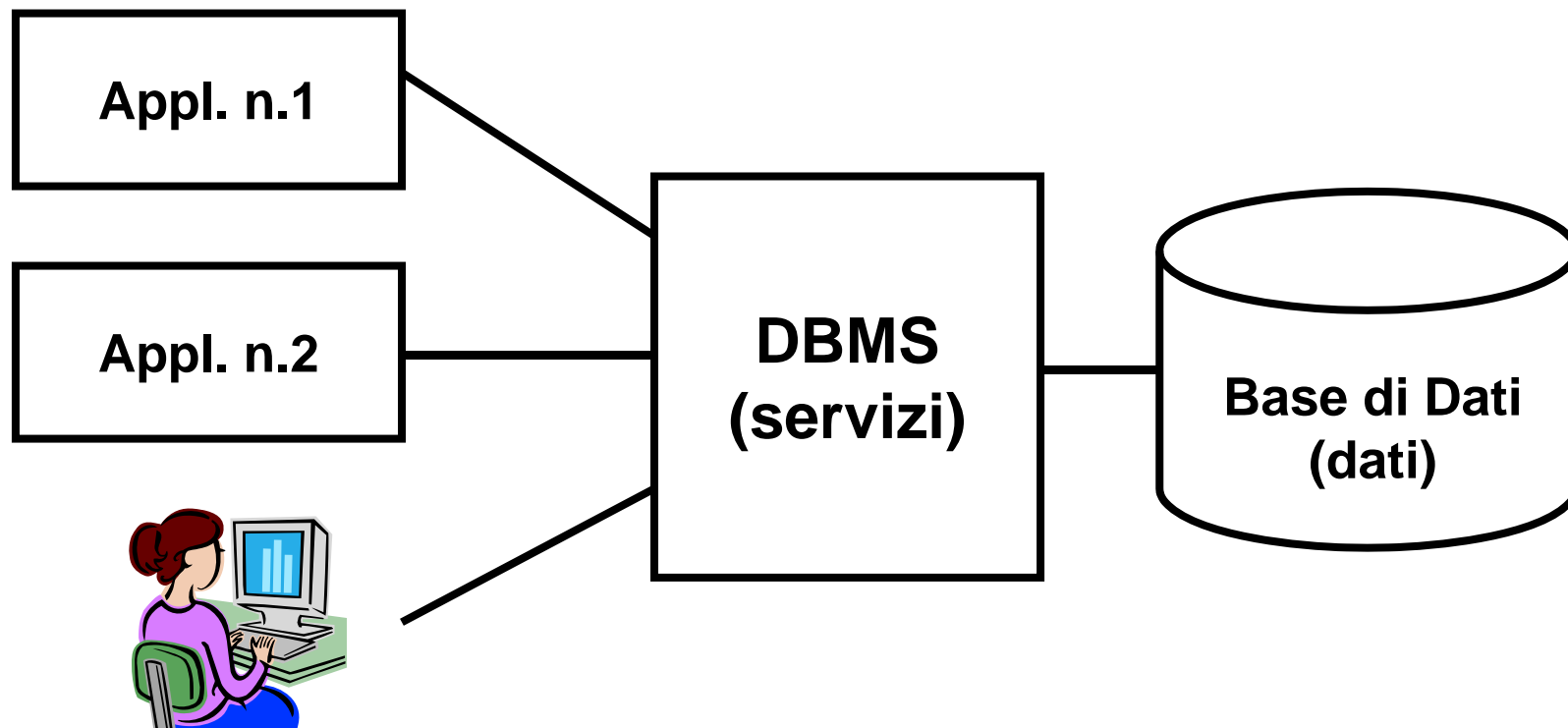
- efficiente

garantendo servizi di:

- condivisione
- affidabilità
- sicurezza

# Architettura Moderna dei S.I.

## ◆ Architettura basata su DBMS



# Servizi di un DBMS

## ◆ **Condivisione**

- riduzione di ridondanze (inconsistenze)
- gestione della concorrenza

## ◆ **Esempio:**

- un'unica collezione di dati ("cittadini") a cui accedono tutte le applicazioni

# Servizi di un DBMS

## ◆ Affidabilità

- recupero dei dati in caso di guasti

## ◆ Esempio:

- in caso di malfunzionamento (alimentazione, rottura del disco) i dati possono essere recuperati

# Servizi di un DBMS

## ◆ Sicurezza

- meccanismi di protezione dell'accesso
- meccanismi di autorizzazione

## ◆ Esempio:

- l'accesso ai dati dei cittadini è protetto agli utenti non autorizzati
- gli utenti autorizzati (impiegati) hanno diversi livelli di autorizzazione

# Caratteristiche di un DBMS

## ◆ **Efficienza**

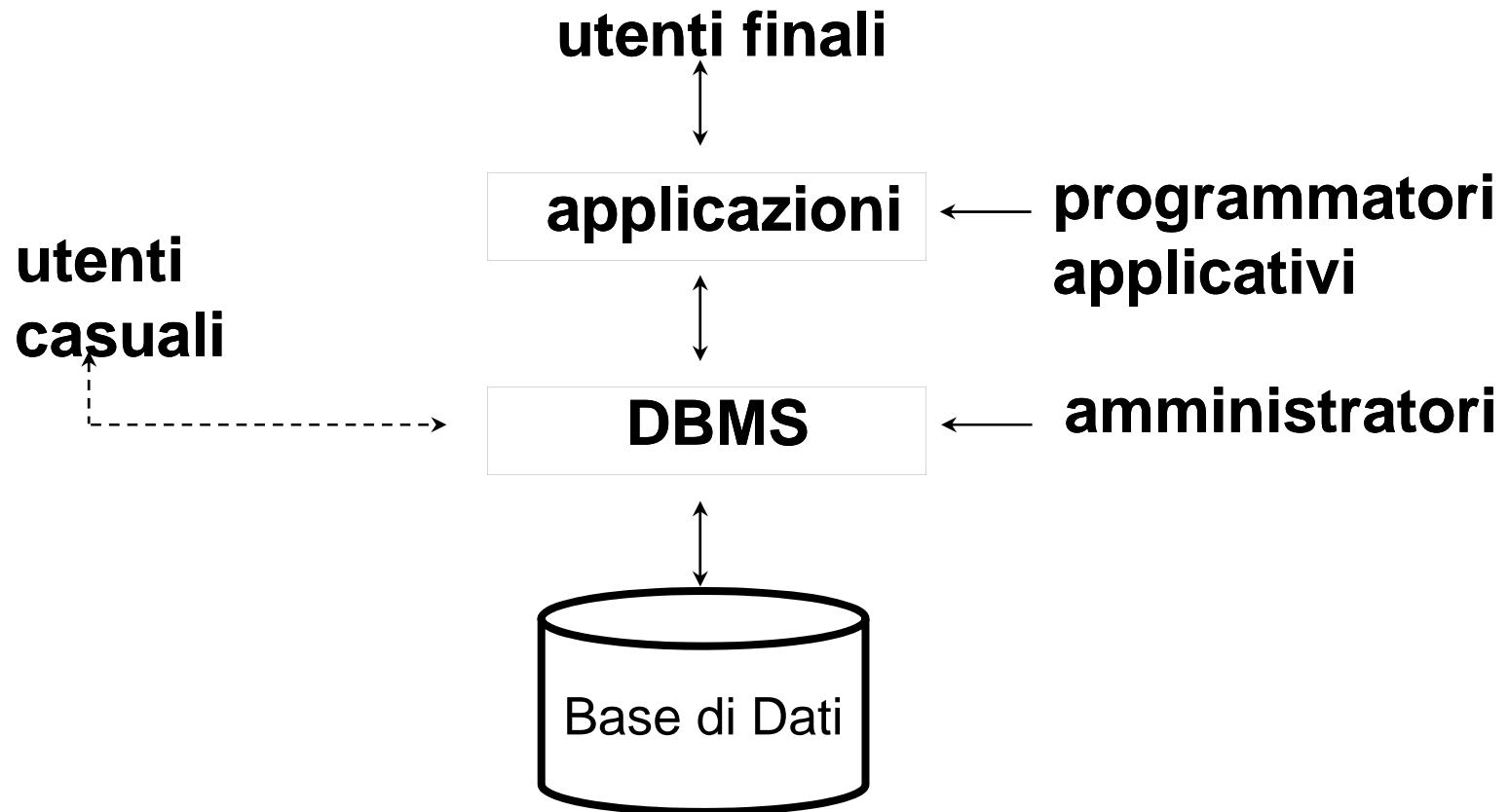
- utilizzo di risorse di calcolo
- relativa alla complessità dei servizi

## ◆ **Efficacia**

- centralizzazione dei servizi
  - miglioramento della produttività
  - semplicità della scrittura di applicazioni
-



# Figure Coinvolte



# Alcuni Esempi di DBMS

- ◆ **Commerciali, Fascia alta**  
IBM DB2, Oracle, Microsoft SQL Server, Sybase

**Commerciali, Fascia bassa**  
Microsoft Access, FileMaker

- ◆ **Open Source**  
MySQL ([www.mysql.com](http://www.mysql.com))  
PostgreSQL ([www.postgresql.org](http://www.postgresql.org))

# Interazione tra DBMS e Applicazioni

## ◆ Come fanno le applicazioni e gli utenti a interagire con il DBMS ?

- devono conoscere l'organizzazione dei dati
- devono sapere come comunicare con il DBMS

## ◆ Devono conoscere:

- Il modello dei dati del DBMS
  - Il linguaggio del DBMS
-

# Modello dei Dati

- ◆ **Insieme di strutture e di regole per la rappresentazione di informazioni**
  - ◆ **Modello logico dei dati**
    - astrazione per il programmatore
  - ◆ **Modello fisico dei dati**
    - al livello della macchina
-

# Modello dei Dati

## ◆ Esempi di modello logico

- In linguaggio C  
tipi base, array, strutture, puntatori
- In linguaggio Java  
tipi base, array, classi, oggetti, ereditarietà

## ◆ Esempi di modello fisico

- strutture per la rappresentazione dei bit  
(registri della memoria, file su disco)

# Modelli di Dati in un DBMS

- ◆ **Ogni DBMS è basato su precisi modelli**
- ◆ **Modello logico**
  - descrizione dei dati visibili alle applicazioni
- ◆ **Modello fisico**
  - strutture per la gestione della persistenza su disco

# DBMS Relazionale

## ◆ Modello Logico

- i dati sono organizzati in “tabelle”
- la tabella è un insieme di “record” (ennuple)
- con un insieme di attributi
- di tipi opportuni (numeri, stringhe, date...)
- i dati in tabelle diverse sono correlati sulla base dei valori

## ◆ Modello Fisico

- file, pagine e record (proprietario)
-

**studenti**

<b>Matricola</b>	<b>Cognome</b>	<b>Nome</b>	<b>Data di nascita</b>
<b>6554</b>	<b>Pinco</b>	<b>Pallino</b>	<b>05/12/1978</b>
<b>8765</b>	<b>Neri</b>	<b>Paolo</b>	<b>03/11/1976</b>
<b>9283</b>	<b>Verdi</b>	<b>Luisa</b>	<b>12/11/1979</b>
<b>3456</b>	<b>Rossi</b>	<b>Maria</b>	<b>01/02/1978</b>

**esami**

<b>Studente</b>	<b>Voto</b>	<b>Corso</b>
<b>3456</b>	<b>30</b>	<b>04</b>
<b>3456</b>	<b>24</b>	<b>02</b>
<b>9283</b>	<b>28</b>	<b>01</b>
<b>6554</b>	<b>26</b>	<b>01</b>

**corsi**

<b>Codice</b>	<b>Titolo</b>	<b>Docente</b>
<b>01</b>	<b>Analisi</b>	<b>Sempronio</b>
<b>02</b>	<b>Chimica</b>	<b>Bruni</b>
<b>04</b>	<b>Chimica</b>	<b>Verdi</b>



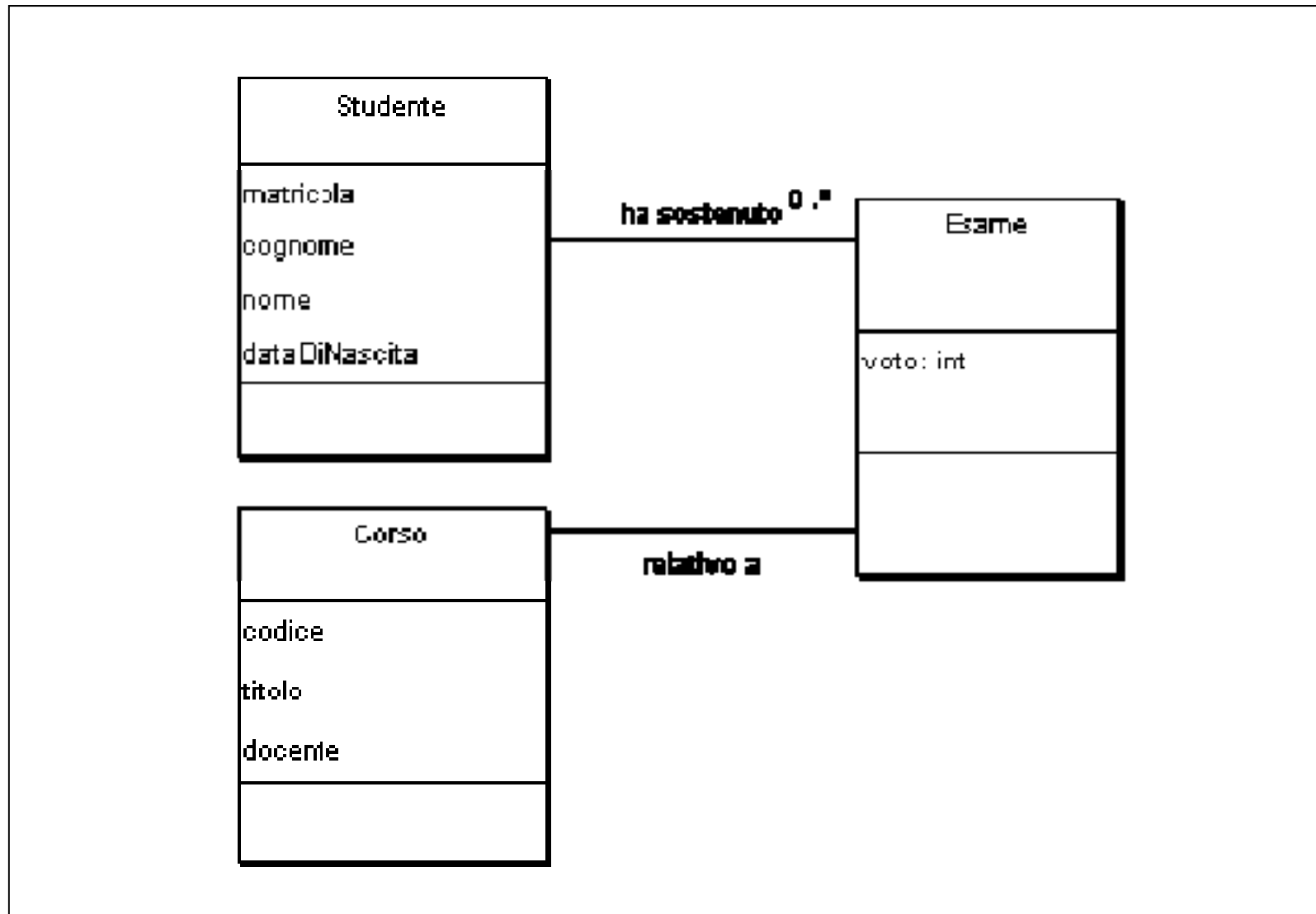
# DBMS Orientati agli Oggetti

## ◆ Modello Logico

- i dati sono organizzati in classi
- ogni classe genera un insieme di oggetti
- con un insieme di proprietà e di metodi
- oggetti diversi sono correlati sulla base di identificatori

## ◆ Modello Fisico

- file, pagine, record, puntatori (proprietario)



# DBMS Relazionali a Oggetti

- ◆ **Sono un compromesso tra DBMS relazionali e DBMS a oggetti**
- ◆ **I dati sono organizzati in tabelle... ma**
  - Limitate funzionalità per la definizione di tipi (classi)
  - Limitate funzionalità per la definizione di oggetti (identificatori)
  - Limitate funzionalità per la creazione di gerarchie

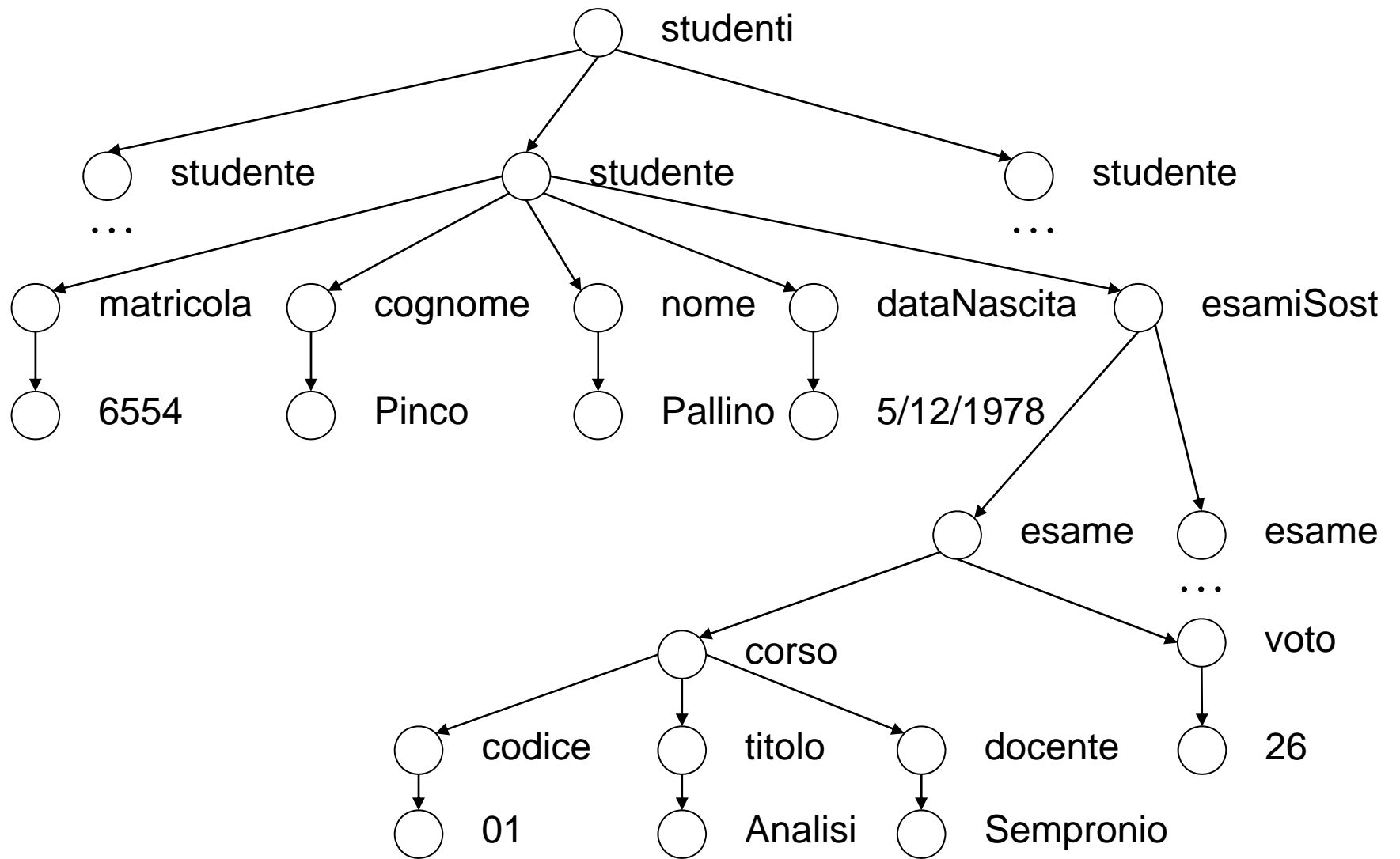
# Sistemi basati su XML

## ◆ Modello Logico

- i dati sono organizzati in strutture gerarchiche (alberi)
- ogni albero ha un insieme di nodi (elementi)
- oggetti diversi sono correlati sulla base di relazioni di contenimento

## ◆ Modello Fisico

- vari possibili modelli
  - standard: file di testo basato su marcatori
-



```
<xml version="1.0" ?>
<studenti>
  <studente>
    <matricola>6554</matricola>
    <cognome>Pinco</cognome>
    <nome>Pallino</nome>
    <dataDiNascita>15/12/1978</dataDiNascita>
    <esamiSostenuti>
      <esame>
        <corso>
          <codice>01</codice>
          <titolo>Analisi</titolo>
          <docente>Giacomo</docente>
          <voto>26</voto>
        </corso>
      </esame>
      ...
    </esamiSostenuti>
  </studente>
</studenti>
```

# Modelli Storici

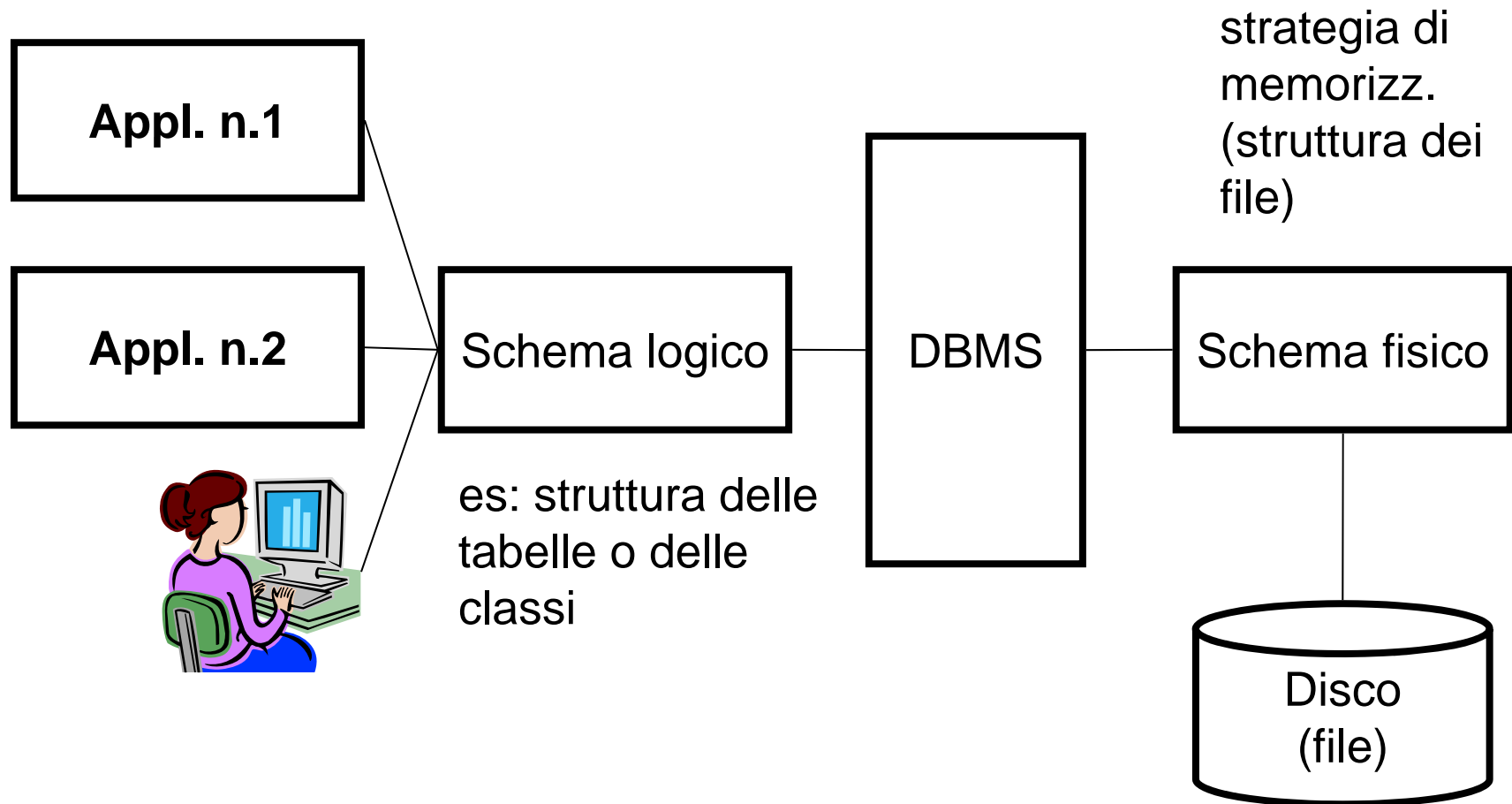
## ◆ **Modello gerarchico**

- anni '60
- IBM IMS
- COBOL, Pascal

## ◆ **Modello reticolare**

- anni '70
  - IDS, IDMS
  - Codasyl, COBOL
-

# Visibilità degli Schemi





# Schema e Istanza

## ◆ Che cosa del modello deve conoscere lo sviluppatore dell'applicazione ?

- il modello logico (e non il modello fisico)
- la struttura dei dati e non necessariamente i dati veri e propri

## ◆ In altri termini

- lo "schema"
  - e non necessariamente l'"istanza"
-

# In ogni modello esistono

## ◆ Lo schema: **la descrizione della struttura**

- stabile nel tempo

## ◆ L'istanza: **i valori (cioè i dati)**

- variabile nel tempo

<b>studenti</b>	<b>Matricola</b>	<b>Cognome</b>	<b>Nome</b>	<b>Data di nascita</b>
	<b>6554</b>	<b>Pinco</b>	<b>Pallino</b>	<b>05/12/1978</b>
	<b>8765</b>	<b>Neri</b>	<b>Paolo</b>	<b>03/11/1976</b>
	<b>9283</b>	<b>Verdi</b>	<b>Luisa</b>	<b>12/11/1979</b>
	<b>3456</b>	<b>Rossi</b>	<b>Maria</b>	<b>01/02/1978</b>

# Schemi di un DBMS

## ◆ Schema logico

- descrizione delle strutture secondo cui i dati sono organizzati nel modello logico (es: tabelle)

## ◆ Schema fisico

- descrizione delle strutture secondo cui i dati sono memorizzati su disco

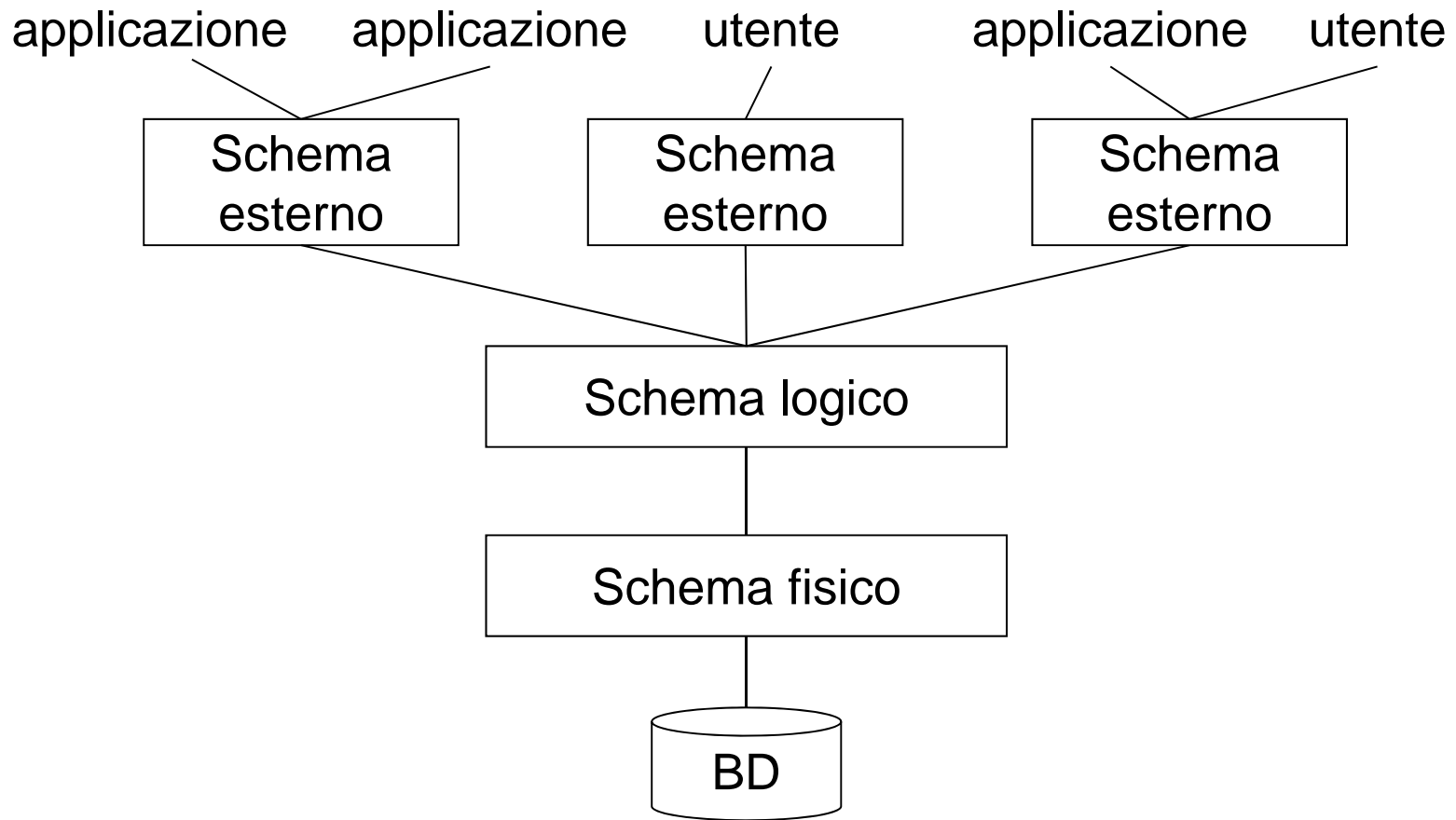
## ◆ Schema esterno

- relativo alla gestione della sicurezza
-

# Schemi Esterni

- ◆ **Non tutti gli utenti hanno gli stessi privilegi**
- ◆ **Schema esterno per un utente: porzione della base di dati che l'utente è autorizzato a vedere**
- ◆ **Viste: strumento attraverso cui si definisce lo schema esterno >>**
- ◆ **Es: base di dati comunale dei cittadini**
  - tabella "cittadino", attributo "imponibile a fini fiscali"
  - vista "cittadinoSenzaDatiFiscali"

# Architettura Standard a Tre Livelli



# Linguaggio per Basi di Dati

- ◆ **Sintassi e semantica per richiedere servizi al DBMS**
- ◆ **Ogni modello ha i suoi linguaggi**
- ◆ **DBMS relazionali: SQL-92, QBE**
- ◆ **DBMS a oggetti: OQL**
- ◆ **DBMS relazionali a oggetti: SQL-99**
- ◆ **XML: XPath. XQuerv**

# SQL: Un esempio

**SELECT** Matricola, Cognome, AVG(Voto)

**FROM** Studenti, Esami

**WHERE** Matricola = Studente

**GROUP BY** Matricola, Cognome

<b>Matricola</b>	<b>Cognome</b>	<b>AVG(voto)</b>
<b>6554</b>	<b>Pinco</b>	<b>26</b>
<b>8765</b>	<b>Neri</b>	<b>28</b>
<b>3456</b>	<b>Rossi</b>	<b>27</b>

# Indipendenza

- ◆ **E' la caratteristica fondamentale dei DBMS**
  - ◆ **Indipendenza dei dati rispetto alle applicazioni**
  - ◆ **Indipendenza dello schema logico rispetto allo schema fisico**
  - ◆ **Indipendenza degli schemi esterni rispetto allo schema logico**
-



# DBMS e Indipendenza

## ◆ Vantaggi

- centralizzazione dei servizi
- semplificazione delle applicazioni
- flessibilità nella manutenzione

## ◆ Svantaggi

- complessità dell'architettura
  - costi
  - minore modularità
-

# Sommario

## ◆ Introduzione

- Sistema Informativo e Sistema Informatico

## ◆ DBMS

## ◆ Modello logico e modello fisico

- Panoramica sui modelli
- Schema e Istanza
- Linguaggi per basi di dati

## ◆ Indipendenza dei dati

---