



Informatica**U**manistica

Basi di dati e Laboratorio Progettazione Web

Anna Monreale &
Giovanna Rosone

Università di Pisa

Andrea Marchetti

IIT-CNR, Pisa

Docenti

- **Anna Monreale**
 - **Email:** anna.monreale@unipi.it
 - **Ricevimento:** Lunedì, 14:00-15:45, Dipartimento di Informatica
- **Giovanna Rosone**
 - **Email:** giovanna.rosone@unipi.it
 - **Ricevimento:** Mercoledì, 14:30-16:30, Dipartimento di Informatica
- **Andrea Marchetti**
 - **Email:** andrea.marchetti@iit.cnr.it
 - **Ricevimento:** IIT-CNR, Pisa

Organizzazione Corso

- **Primo Modulo: BDD**
 - Primo semestre
 - **Esame scritto:** compiti o appelli
 - Agli appelli puoi recuperare solo una parte
 - **Esame Orale:** non obbligatorio ma necessario per avere la lode
- **Secondo Modulo: LPW**
 - Progetto
 - Discussione del progetto
- **Voto finale:** media dei due moduli
 - Lode solo se avete fatto l'orale del primo modulo

Materiale Didattico

- **Pagina corso:**
<http://didawiki.di.unipi.it/doku.php/bdd-infuma/start>
- **Slides** delle lezioni e altro materiale nella pagina del corso
ATTENZIONE: LE SLIDES NON SONO ESAUSTIVE!
- **Libro di testo:** Paolo Atzeni, Stefano Ceri, Stefano Paraboschi, Riccardo Torlone “***Basi di dati: Modelli e Linguaggi di Interrogazione***”, McGraw-Hill Libri Italia, Nuova edizione (2006).

Modalità di verifica

- **Basi di Dati**

- Prova Scritta:

- Compitini di verifica intermedi, indipendenti
 - Scritto per il recupero di un compitino, o di entrambi

- Prova Orale:

- Nella prova scritta sono incluse domande orali

- **Laboratorio Progettazione Web**

- Progetto concordato con il docente

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

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

Contenuti (e struttura ...) del corso

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

Una rapida panoramica sul corso

*Dall'Informazione strutturata
all'informazione semi-strutturata*

Modello dei Dati

- Insieme di meccanismi di astrazione per la **rappresentazione di informazioni**
- L'obiettivo è **organizzare** i dati e descriverne la **struttura** in modo comprensibile da un elaboratore

Modello Relazionale dei Dati

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

Esempio Modello Relazionale

STUDENTI

Matricola	Cognome	Nome	Data di nascita
276545	Rossi	Maria	25/11/1991
485745	Neri	Anna	23/04/1992
200768	Verdi	Fabio	12/02/1992
587614	Rossi	Luca	10/10/1991
937653	Bruni	Mario	01/12/1991

ESAMI

Studente	Voto	Corso
276545	28	01
276545	27	04
937653	25	01
200768	24	04

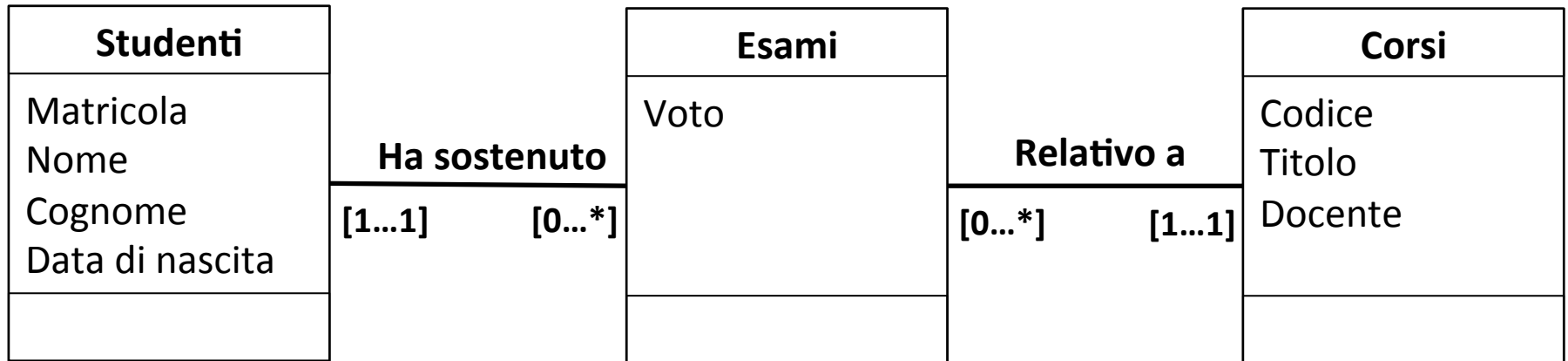
CORSI

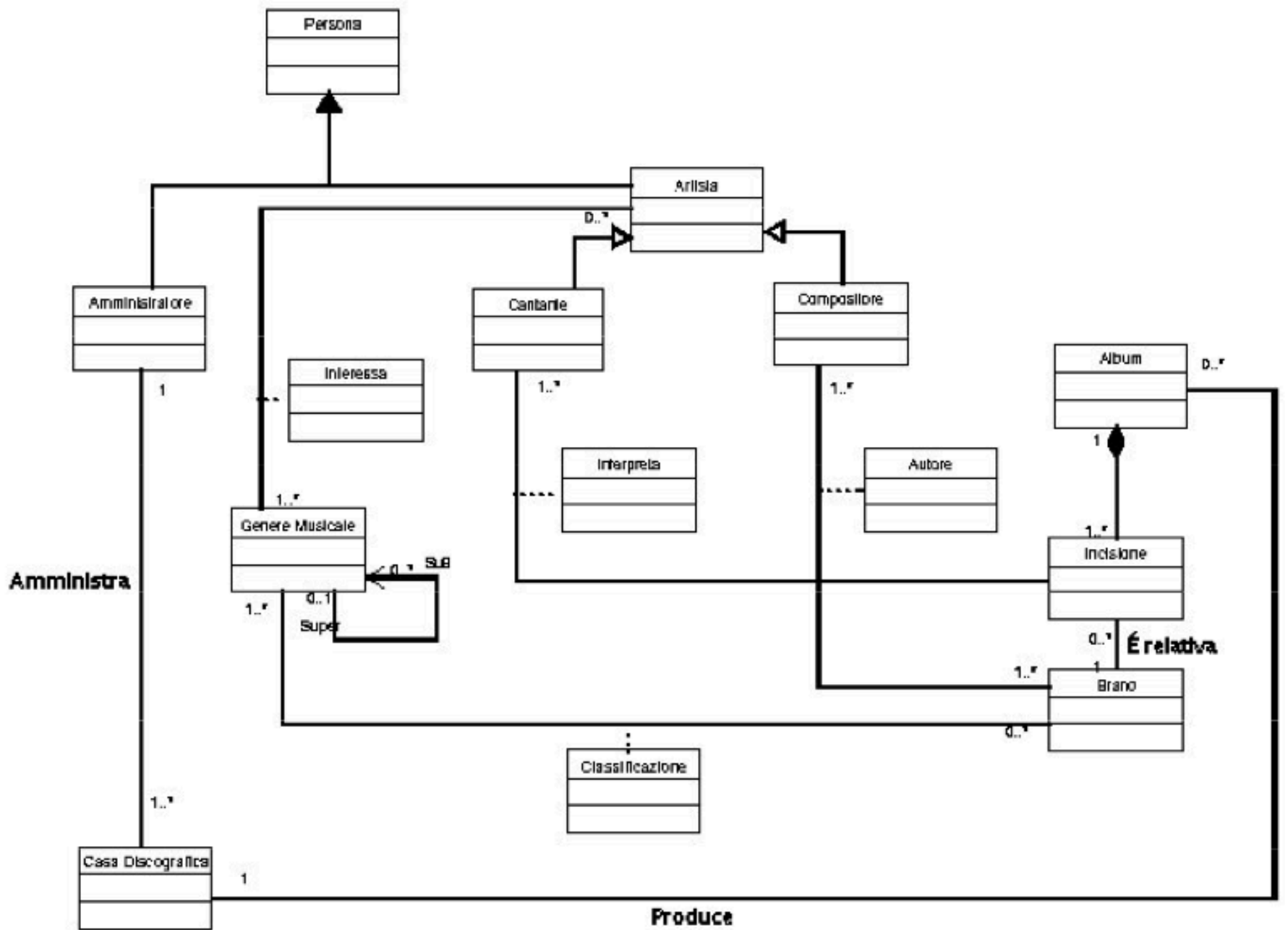
Codice	Titolo	Docente
01	Analisi	Giani
03	Chimica	Melli
04	Chimica	Belli

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**

Esempio Modello Concettuale

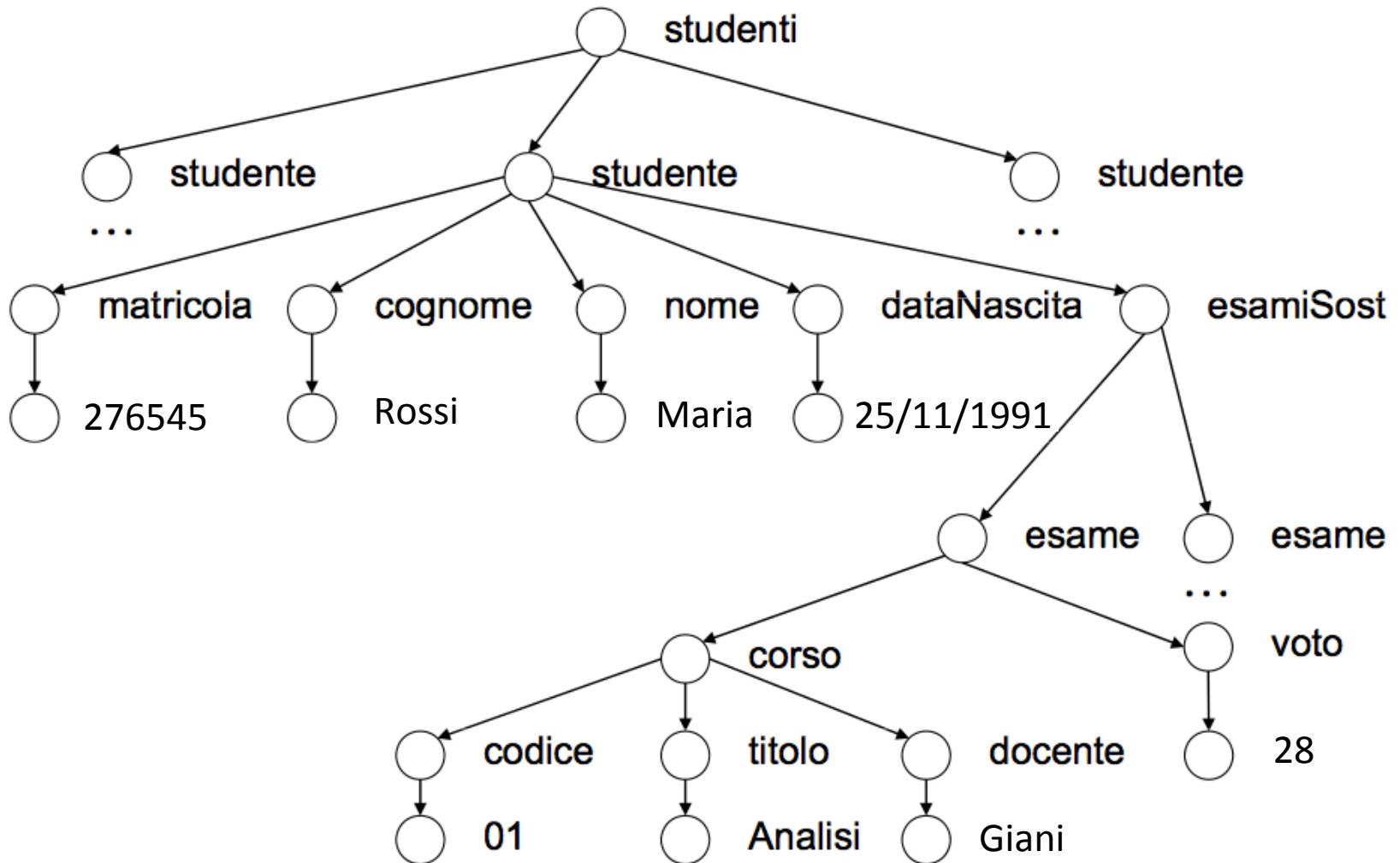




Modello semi-strutturato: XML

- 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**

Albero




```
<xml version="1.0" ?>
<studenti>
  <studente>
    <matricola>276545</matricola>
    <cognome>Rossi</cognome>
    <nome>Maria</nome>
    <dataDiNascita>25/11/1991</dataDiNascita>
    <esamiSostenuti>
      <esame>
        <corso>
          <codice>01</codice>
          <titolo>Analisi</titolo>
          <docente>Giani</docente>
          <voto>28</voto>
        </corso>
      </esame>
      ...
    </esamiSostenuti>
  </studente>
</studenti>
```

Componenti Modello

- In ogni modello esistono
 - Lo schema: **la descrizione della struttura**
 - stabile nel tempo
 - L'istanza: **i valori (cioè i dati)**
 - variabile nel tempo

STUDENTI

Matricola	Cognome	Nome	Data di nascita
276545	Rossi	Maria	25/11/1991
485745	Neri	Anna	23/04/1992
200768	Verdi	Fabio	12/02/1992
587614	Rossi	Luca	10/10/1991
937653	Bruni	Mario	01/12/1991

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**

Esempio Query SQL

```
SELECT Matricola, Cognome, AVG(Voto)
FROM Studenti, Esami
WHERE Matricola = Studente
GROUP BY Matricola, Cognome
```

Matricola	Cognome	AVG(Voto)
276545	Rossi	27,5
937653	Bruni	25
200768	Verdi	24



Informatica**Umanistica**

Basi di dati

Introduzione

Anna Monreale
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**

L'informatica

E' la ***scienza*** che si occupa della ***rappresentazione dell'informazione*** e della sua ***gestione***

Sistema Organizzativo

- **Organizzazione complessa**
(es: azienda, banca, ente pubblico, ... ma anche fantacalcio...)
- **Sistema Organizzativo**
 - complesso delle strutture, delle risorse, delle regole e delle procedure che regolano lo svolgimento delle attività
 - Es: banca – regolamento apertura c/c
 - Es: anagrafe – procedure per la registrazione nascite

Risorse

- Le risorse di una organizzazione(ente, amministrazione, azienda) sono:
 - persone
 - denaro
 - materiali
 - informazioni

Sistema Informativo

- Componente di una organizzazione **che gestisce le informazioni** utili ai fini dell'organizzazione stessa
 - acquisisce, elabora, conserva, produce informazioni
- Ogni organizzazione ha un sistema informativo, eventualmente non esplicitato nella struttura
- Il sistema informativo è di **supporto ad altri sottosistemi**, e va quindi studiato nel contesto in cui è inserito

Sistema informativo e automazione

- Il concetto di **sistema informativo** è indipendente da qualsiasi automatizzazione
- Esistono organizzazioni la cui ragion d'essere è la gestione di informazioni come servizi anagrafici e banche, e che operano da secoli



Biblioteca reale di Alessandria d'Egitto
IV-I secolo a.C - 400000 rotoli presenti.

Sistema informativo e automazione

- Il concetto di **sistema informativo** è indipendente da qualsiasi automatizzazione
- Esistono organizzazioni la cui ragion d'essere è la gestione di informazioni come servizi anagrafici e banche, e che operano da secoli

Censimenti e Registro Anagrafe



Nell'Antica Roma, i **censimenti** venivano effettuati dalla fine del IV secolo a.c. Gli **elenchi dei censiti**, distinti secondo il possesso o meno dei diritti civili e politici, la classe patrimoniale e l'età, venivano utilizzati come liste elettorali e per determinare la ruoli per l'esenzione dei tributi e le liste di leva.

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



obiettivo dei DBMS: il trattamento dei dati per ottenere informazione

Dato vs Informazione

- Nei sistemi informatici (e non solo) le **informazioni** vengono rappresentate attraverso i **dati**
- **Informazione** (def): notizia, o elemento che consente di avere conoscenza piu' o meno esatta di fatti, situazioni e modi d'essere.
- **Dato** (def): elementi di informazione costituiti da simboli che devono essere elaborati.

NOTA: **Senza interpretazione, il dato non e' molto utile!**

FERRARI,8



Base di Dati

- **Insieme organizzato di dati** utilizzati per il supporto allo svolgimento di attività di
 - enti
 - aziende
 - uffici
 - Persone
- Insieme di dati gestito da un **DBMS**

Database Management Systems

- 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

Database Management Systems

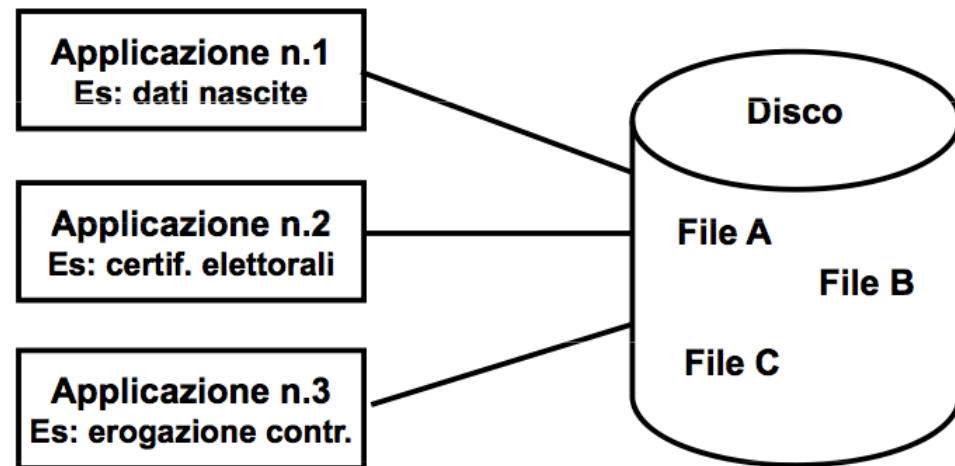
- DBMS sono una componente fondamentale del **sistema informatico** di una grande organizzazione
- DBMS sono una componente fondamentale del **sistema informativo**
- DBMS sono una componente fondamentale del **sistema organizzativo**

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

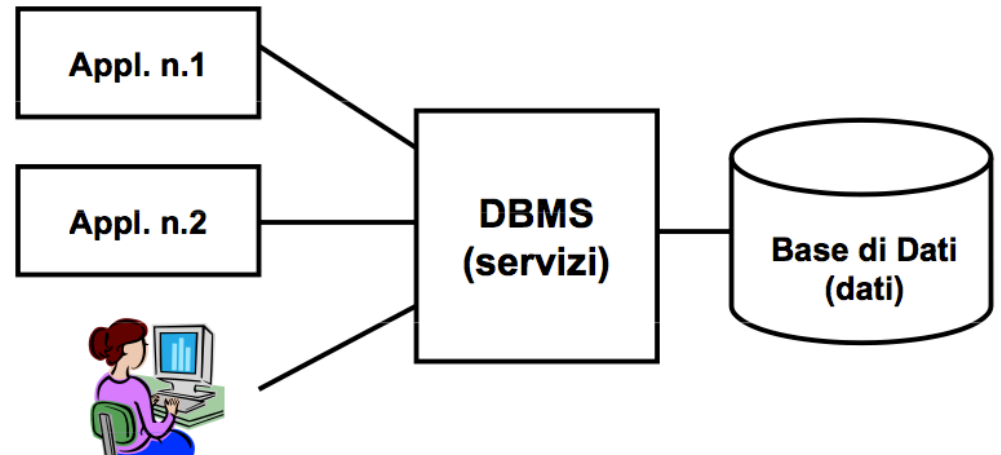
Architettura Tradizionale

- Sistema informatico organizzato in più archivi (*files*) disgiunti
- **Problemi:**
 - Dimensione dei dati
 - Ridondanza
 - Rischio di incoerenza
 - Efficienza
 - Condivisione
 - Sicurezza



Architettura basata su DBMS

- DBMS: Sistema **per gestire** grandi collezioni di dati persistenti, condivisi
- **in modo**
 - efficiente
 - efficace
- e garantendo servizi di:
 - Condivisione
 - Affidabilità
 - Sicurezza



Servizi di un DBMS

- **Condivisione:**
 - riduzione di ridondanze (inconsistenze)
 - gestione della concorrenza nell'accesso

- **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

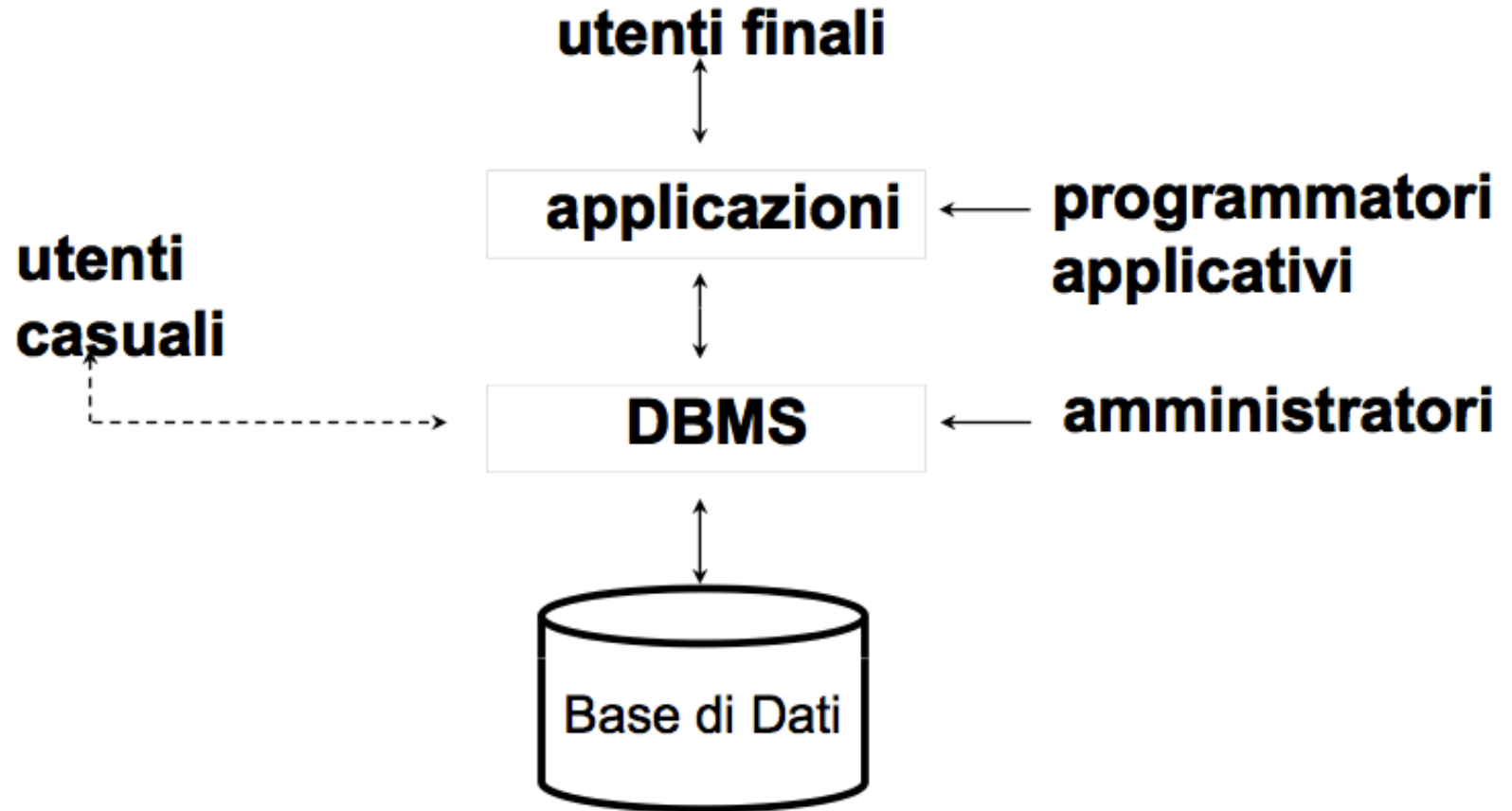
- **Efficienza**

- 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)
 - PostgreSQL (www.postgresql.org)

Descrizioni dei dati nei DBMS

- Nei programmi tradizionali che accedono a file, ogni programma contiene una descrizione della struttura del file stesso
 - rischi di incoerenza fra le descrizioni (ripetute in ciascun programma)
- Nei DBMS, esiste una porzione della base di dati (**catalogo** o **dizionario**) che contiene una descrizione centralizzata dei dati, che può essere utilizzata dai vari programmi

Descrizioni dei dati nei DBMS

- Rappresentazioni dei dati **a livelli diversi**
 - permettono **l'indipendenza dei dati** dalla rappresentazione fisica:
 - i programmi fanno riferimento alla struttura a livello più alto
 - le rappresentazioni sottostanti possono essere modificate senza necessità di modifica dei programmi
 - precisiamo attraverso il concetto di
 - **modello dei dati**

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 regole per la rappresentazione di informazioni
- L'obiettivo è **organizzare i dati** e **descrivere la struttura** in modo comprensibile da un elaboratore
- **Modello logico dei dati**
 - astrazione per il programmatore
 - adottati nei DBMS esistenti per l'organizzazione dei dati
 - utilizzati dai programmi
 - indipendenti dalle strutture fisiche
- **Modello fisico dei dati**
 - al livello della macchina
 - Strutture fisiche di memorizzazione

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)
- i record hanno 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

Esempio Modello Relazionale

STUDENTI

Matricola	Cognome	Nome	Data di nascita
276545	Rossi	Maria	25/11/1991
485745	Neri	Anna	23/04/1992
200768	Verdi	Fabio	12/02/1992
587614	Rossi	Luca	10/10/1991
937653	Bruni	Mario	01/12/1991

ESAMI

Studente	Voto	Corso
276545	28	01
276545	27	04
937653	25	01
200768	24	04

CORSI

Codice	Titolo	Docente
01	Analisi	Giani
03	Chimica	Melli
04	Chimica	Belli

DBMS Orientati agli Oggetti

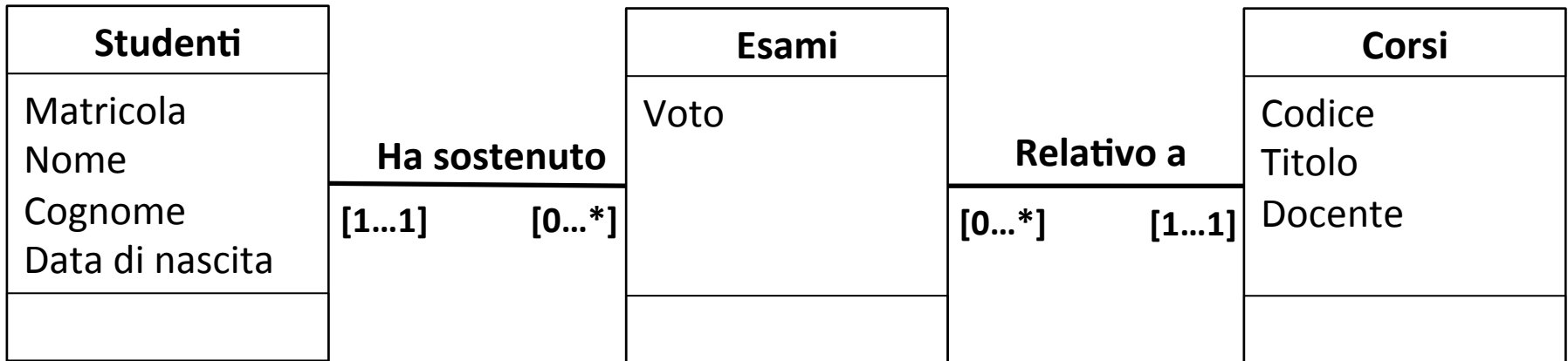
- **Modello Logico**

- 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

Esempio



DBMS Orientati agli 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

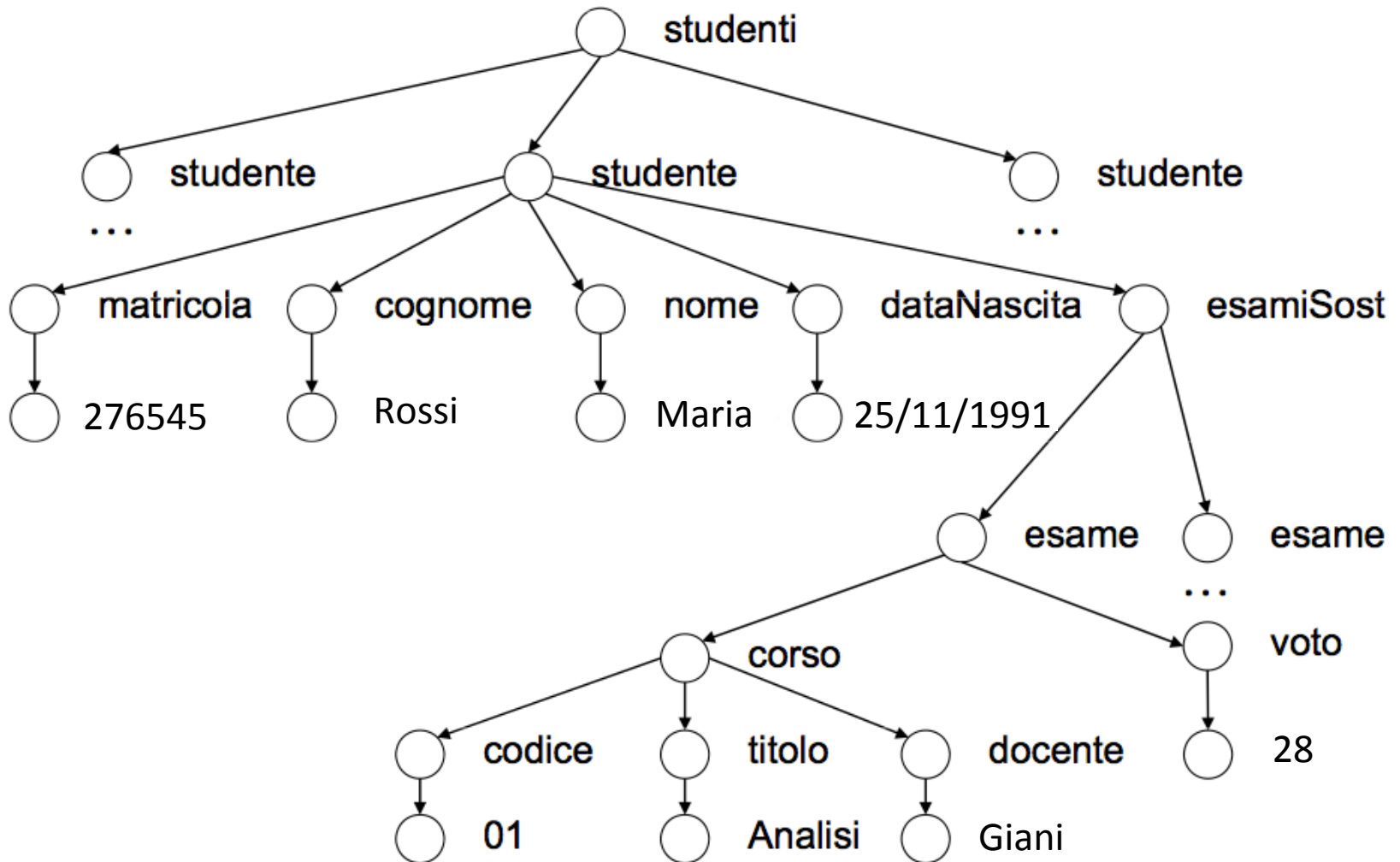
- **Modello Logico**

- 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

Albero

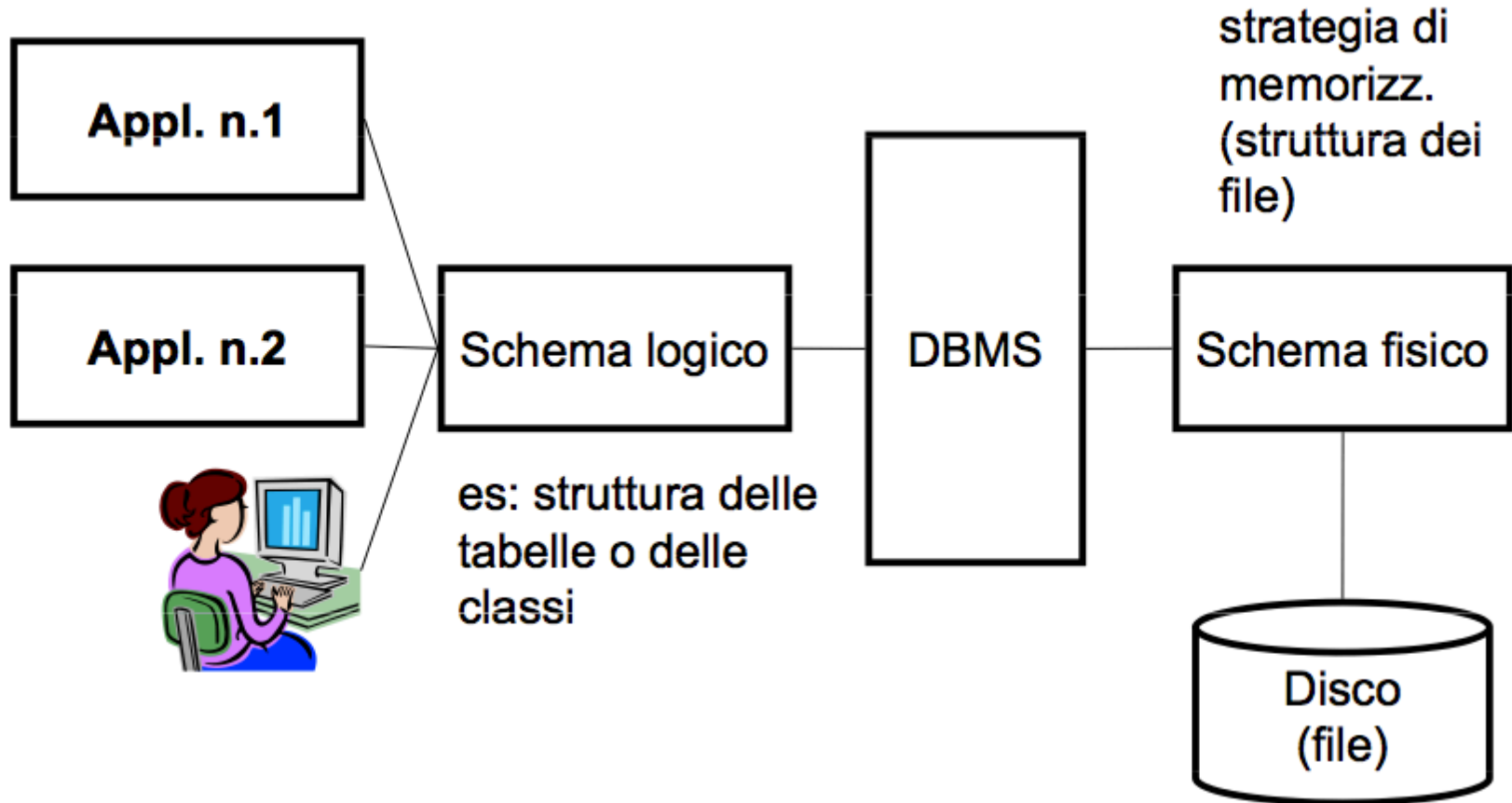



```
<xml version="1.0" ?>
<studenti>
  <studente>
    <matricola>276545</matricola>
    <cognome>Rossi</cognome>
    <nome>Maria</nome>
    <dataDiNascita>25/11/1991</dataDiNascita>
    <esamiSostenuti>
      <esame>
        <corso>
          <codice>01</codice>
          <titolo>Analisi</titolo>
          <docente>Giani</docente>
          <voto>28</voto>
        </corso>
      </esame>
      ...
    </esamiSostenuti>
  </studente>
</studenti>
```

Modelli Storici

- **Modello gerarchico**
 - Basato sull'uso di strutture ad albero
 - Definito negli anni '60
 - IBM IMS
 - COBOL, Pascal
- **Modello reticolare**
 - Basato sull'uso di grafi
 - Definito negli anni '70
 - IDS, IDMS
 - Codasyl, COBOL

Architettura di un DBMS: 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**

Schema e Istanza

- In ogni modello esistono
 - Lo schema: **la descrizione della struttura**
 - stabile nel tempo
 - L'istanza: **i valori (cioè i dati)**
 - variabile nel tempo

STUDENTI

Matricola	Cognome	Nome	Data di nascita
276545	Rossi	Maria	25/11/1991
485745	Neri	Anna	23/04/1992
200768	Verdi	Fabio	12/02/1992
587614	Rossi	Luca	10/10/1991
937653	Bruni	Mario	01/12/1991

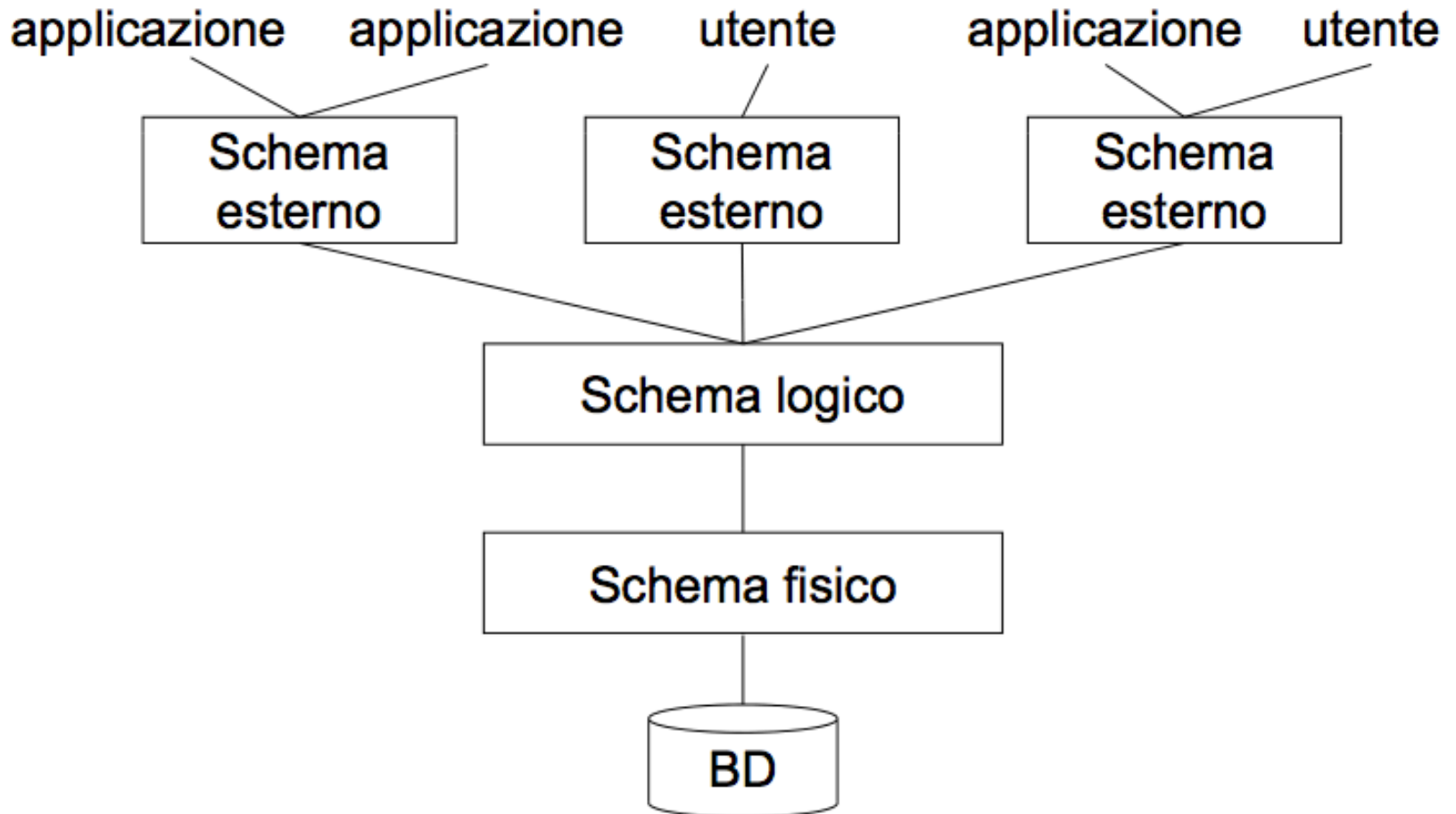
Schemi di un DBMS

- In gran parte dei DBMS moderni, i dati sono articolati su **tre livelli** di astrazione:
 - **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**
 - descrizione di una porzione della base di dati che prevede organizzazioni dei dati diverse rispetto a quelle dello schema logico
 - 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 “Cittadino Senza Dati Fiscali”

Architettura Standard a Tre Livelli



Indipendenza

- E' la caratteristica fondamentale dei DBMS
- **Indipendenza fisica**
 - Interagire con DBMS in modo indipendente dalla struttura fisica dei dati
 - Modificare le strutture fisiche senza influire sulla descrizione dei dati e sui programmi che li usano
- **Indipendenza logica**
 - Interagire con il livello esterno della BD in modo indipendente dallo schema logico
 - Modificare lo schema esterno senza modificare lo schema logico e viceversa

Indipendenza

- Indipendenza dei dati rispetto alle applicazioni
- Indipendenza dello schema logico rispetto allo schema fisico
- Indipendenza degli schemi esterni rispetto allo schema logico

Linguaggi per basi di dati

- L'accesso ai dati può avvenire mediante:
 1. linguaggi **testuali interattivi** (es. SQL)
 2. comandi (come quelli del linguaggio interattivo) **immersi in un linguaggio ospite** (Pascal, C, Cobol, etc.)
 3. comandi (come quelli del linguaggio interattivo) **immersi in un linguaggio ad hoc**, con anche altre funzionalità (es. per grafici o stampe strutturate), anche con l'ausilio di strumenti di sviluppo (p. es. per la gestione di maschere)
 4. **interfacce amichevoli** (es. linguaggi grafici)

Due tipi di linguaggi

- **Data Manipulation Language (DML)**
 - per l'interrogazione e l'aggiornamento di (istanze di) basi di dati
- **Data Definition Language (DDL)**
 - per la definizione di schemi (logici, esterni, fisici) e altre operazioni generali
- Alcuni linguaggi come SQL integrano le funzionalità di entrambi

Un'operazione DDL (sullo schema)

CREATE TABLE orario (

insegnamento CHAR(20),

docente CHAR(20),

aula CHAR(4),

ora CHAR(5))

Esempio Query SQL

```
SELECT Matricola, Cognome, AVG(Voto)
FROM Studenti, Esami
WHERE Matricola = Studente
GROUP BY Matricola, Cognome
```

Matricola	Cognome	AVG(Voto)
276545	Rossi	27,5
937653	Bruni	25
200768	Verdi	24

Linguaggi per basi di dati

- Ogni modello ha i suoi linguaggi
 - DBMS relazionali: **SQL, QBE**
 - DBMS a oggetti: **OQL**
 - XML: **XPath, XQuery**

Vantaggi dei DBMS

- Dati come risorsa comune
- Base di dati come modello della realtà
- Gestione centralizzata con possibilità di standardizzazione ed “economia di scala”
- Disponibilità di servizi integrati
- Riduzione di ridondanze e inconsistenze
- Indipendenza dei dati (favorisce lo sviluppo e la manutenzione delle applicazioni)

Svantaggi dei DBMS

- **Prestazioni**

- In alcuni sistemi con richieste di efficienza sull'elaborazione (es. real-time), l'overhead computazionale introdotto dal DBMS può essere eccessivo ...

- **Costo**

- Spese per l'acquisto di DBMS, formazione del personale, amministrazione del DB, etc

- **Complessità**

- Applicazioni/sistemi di dimensioni ridotte, single-user e con pochi dati da gestire ...