



INTRODUZIONE AL MODULO DI BASI DI DATI

Patrizio Dazzi
a.a. 2017 - 2018

UMANISTI E BASI DI DATI

Quali sono i fondamenti dell'informatica che dovrebbero stare nel bagaglio del bravo umanista?

Quelli necessari a padroneggiare i principi per la rappresentazione dell'informazione e della conoscenza



OBIETTIVI DEL CORSO

- Fornire le conoscenze e le competenze metodologiche 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.**

MATERIALE DIDATTICO

- Slides delle lezioni e altro materiale nella pagina del corso (DidaWiki)
- **Libro di testo:** Paolo Atzeni, Stefano Ceri, Stefano Paraboschi, Riccardo Torlone
“Basi di dati: Modelli e Linguaggi di Interrogazione”
McGraw-Hill Libri Italia.
- **NOTA BENE: LE SLIDES NON SONO ESAUSTIVE!!!**

MAILING LIST

- Spedire una mail a `patrizio.dazzi@isti.cnr.it` con:
 - **Oggetto:** [BDD-INFUMA-2017-18]
 - **Testo:** *Nome, Cognome, Matricola, Email universitaria*

Sarete inseriti in una mailing list che sarà usata per le comunicazioni utili e urgenti

INFORMAZIONE E STRUTTURA



INFORMAZIONE ???

- **Informazione = Contenuto + **Struttura****
- Contenuto: l'entità 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 E 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

STRUTTURIAMO O NON STRUTTURIAMO ?

- **Informazione non strutturata**
 - Flessibile a costruirsi e ad alimentarsi
 - Difficile ad interrogarsi per estrarre conoscenza

- **Informazione strutturata**
 - Rigida a costruirsi e ad alimentarsi
 - Adatta e flessibile ad interrogarsi per estrarre conoscenza

STRUTTURIAMO!!! IN CHE MODO ? IL 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

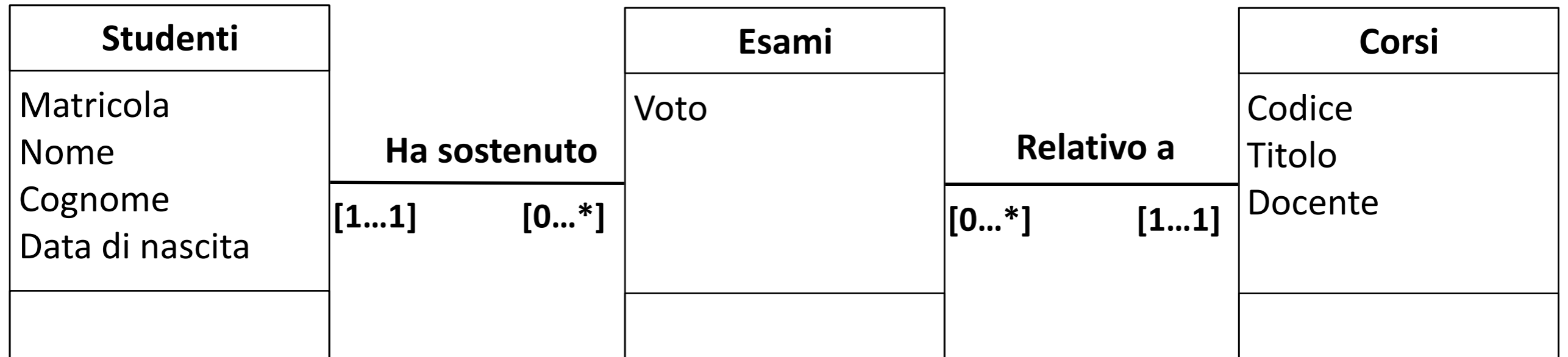
COS'È IL MODELLO DEI DATI ?



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 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 possono essere **correlati**

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

COMPONENTI MODELLO

- In ogni modello esistono
 - Lo schema: **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
 - Nel corso vedremo SQL
 - Consente di rappresentare lo schema
 - Di generare le istanze
 - Di interrogare la base di dati

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

SISTEMA INFORMATIVO E SISTEMA INFORMATICO





DATABASE

SOMMARIO

- Informativo o informatico ?
 - 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

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: DEFINIZIONE

Fatto: le organizzazioni gestiscono informazioni al fine di migliorare il loro funzionamento.

- **Definizione** Un sistema informativo di un'organizzazione e' una combinazione di risorse, umane e materiali, e di procedure organizzate per
 - la raccolta,
 - l'archiviazione,
 - l'elaborazione e
 - lo scambio delle informazioni necessarie
 - alle attività operative (informazioni di servizio),
 - alle attività di gestione (informazioni di gestione),
 - alle attività di programmazione, controllo e valutazione dell'organizzazione (informazioni di governo)

SISTEMA INFORMATIVO: UN PAIO DI ESEMPI

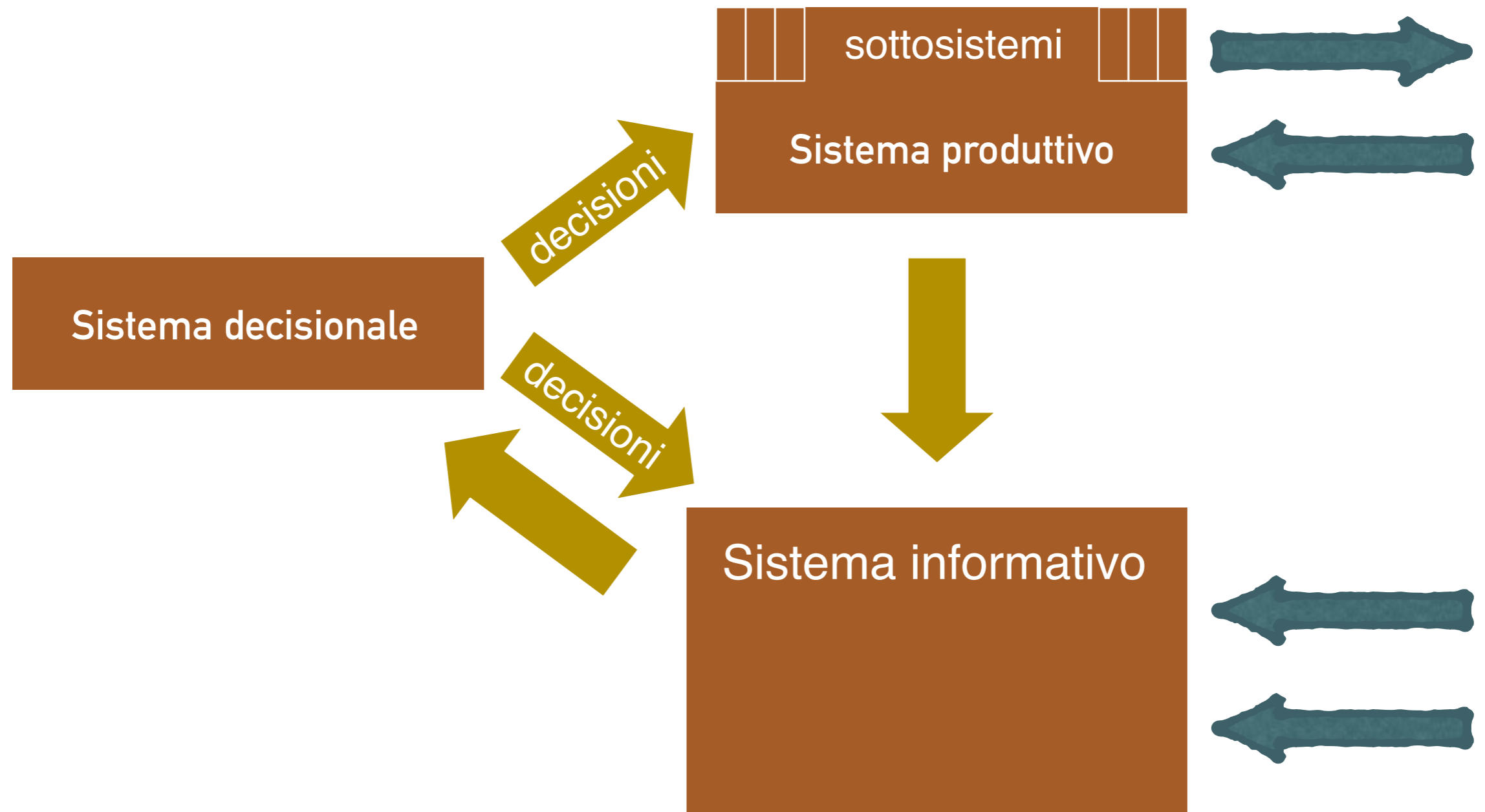
➤ Azienda manifatturiera

- Gestione degli ordini dei clienti
- Gestione degli ordini e dei pagamenti di materiali dei fornitori
- Gestione del magazzino
- Pianificazione della produzione e controllo dei costi

➤ Un Comune

- Gestione dei servizi demografici (anagrafe, stato civile, servizio elettorale e vaccinale) e della rete viaria
- Gestione del personale per il calcolo della retribuzione in base alla normativa contrattuale

SISTEMA INFORMATIVO NELLE ORGANIZZAZIONI



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

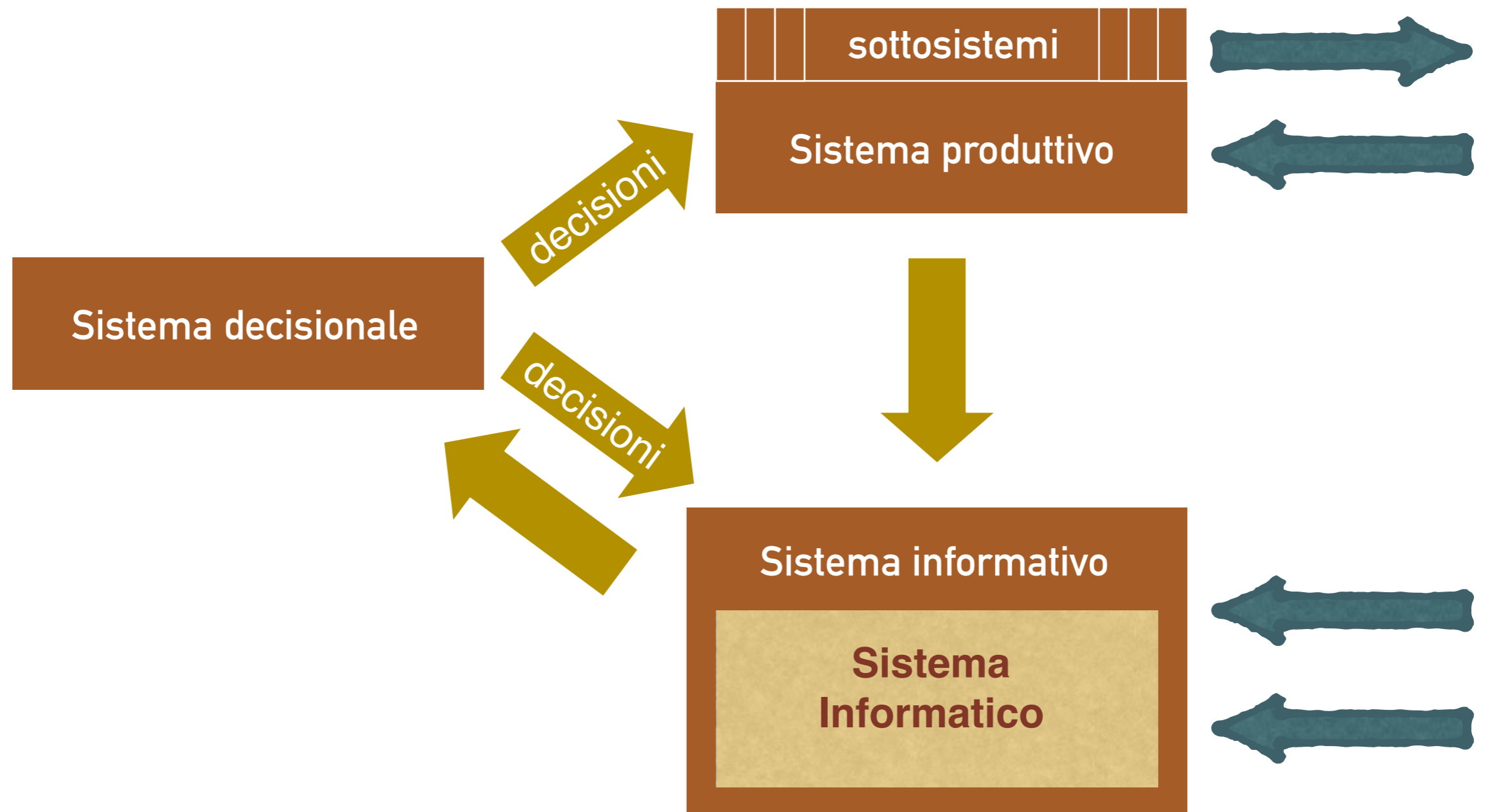
- La porzione **informatizzata** del sistema informativo
- Collezione di applicazioni che gestiscono i dati e i flussi informativi
- **Spesso il sistema informativo è il sistema informatico**
 - ma non è sempre così

SISTEMI INFORMATICI: **DEFINIZIONE**

Un sistema informatico e' l'insieme degli **strumenti informatici** impiegati per il **trattamento automatico** delle informazioni di un'organizzazione al fine di agevolare le funzioni del suo sistema informativo.

Delle informazioni trattate da un'organizzazione interessa prendere in considerazione solo quelle **strutturate**, con un formato **predeterminato** e di carattere prevalentemente globale, cioè di utilità a più reparti dell'organizzazione.

SISTEMA INFORMATICO NELLE ORGANIZZAZIONI



EVOLUZIONE DEI SISTEMI INFORMATICI

Prima fase

Utilizzo della tecnologia dei sistemi di archiviazione per applicazioni settoriali, soprattutto nell'amministrazione, con l'obiettivo di automatizzare quelle attività che richiedono l'elaborazione sistematica e ripetitiva di grandi quantità di dati.

Seconda fase

A partire dagli anni '70, utilizzo della tecnologia dei DBMS che hanno consentito la realizzazione di sistemi informatici di supporto ad ogni livello delle organizzazioni.

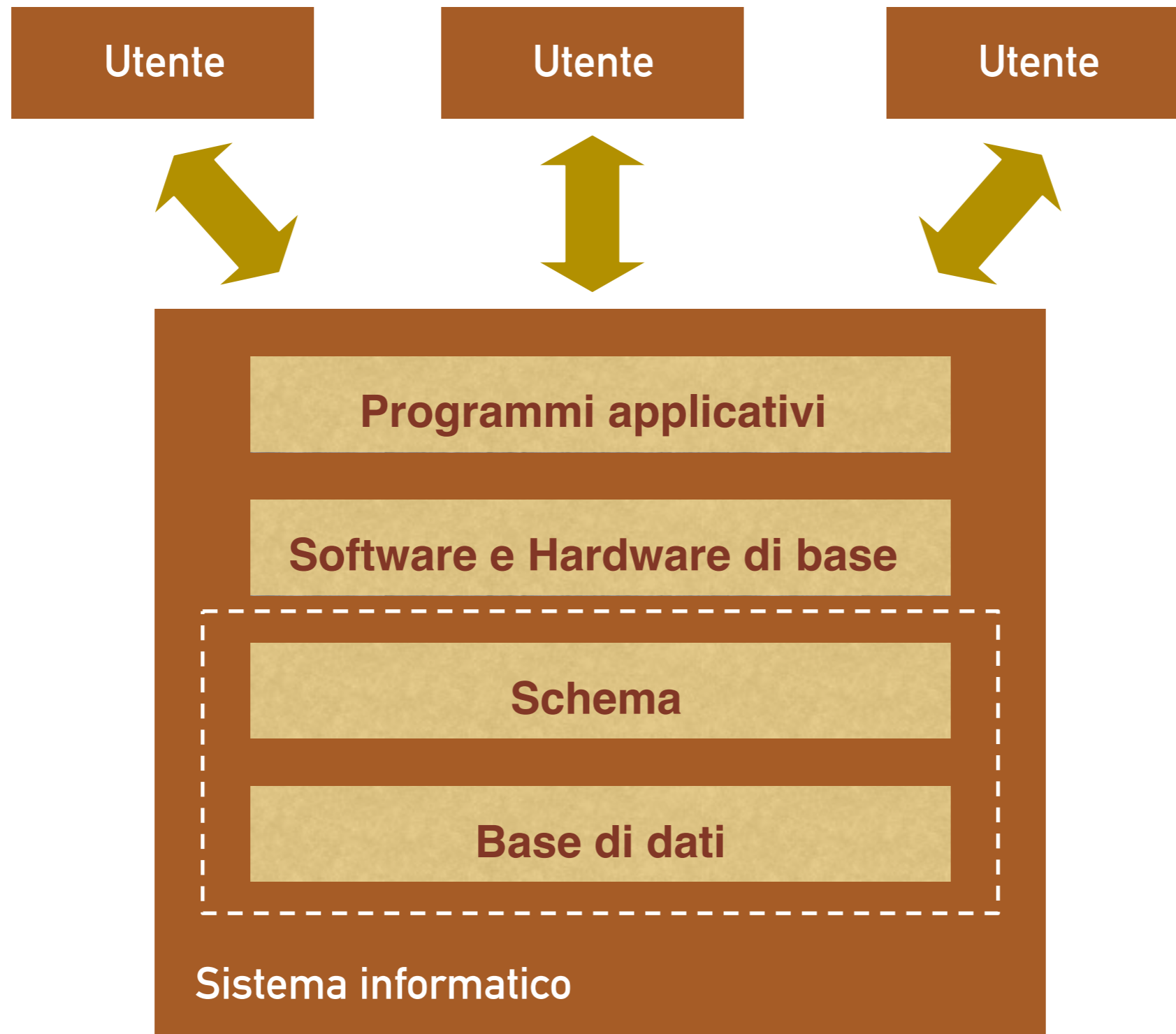
Terza fase

Alla fine degli anni '80 utilizzo della tecnologia dei DBMS distribuiti, su rete locale o geografica.

Quarta fase

A partire dagli anni '90, utilizzo della tecnologia Web.

COMPONENTI DI UN SISTEMA INFORMATICO



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 più o meno esatta di fatti, situazioni e modi d'essere.
- **Dato** (def): elemento di informazione costituito 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
 - software
 - ecc.
- Insieme di dati gestito da un **DBMS**

DBMS



DATABASE MANAGEMENT SYSTEMS

- Sistemi di Gestione di Basi di Dati o “Database Management Systems” (DBMS)
- Sono utilizzati laddove:
 - ci siano molti dati da gestire
 - siano utili alla realizzazione di processi complessi
 - i dati siano cruciali per lo svolgimento delle attività

DATABASE MANAGEMENT SYSTEMS

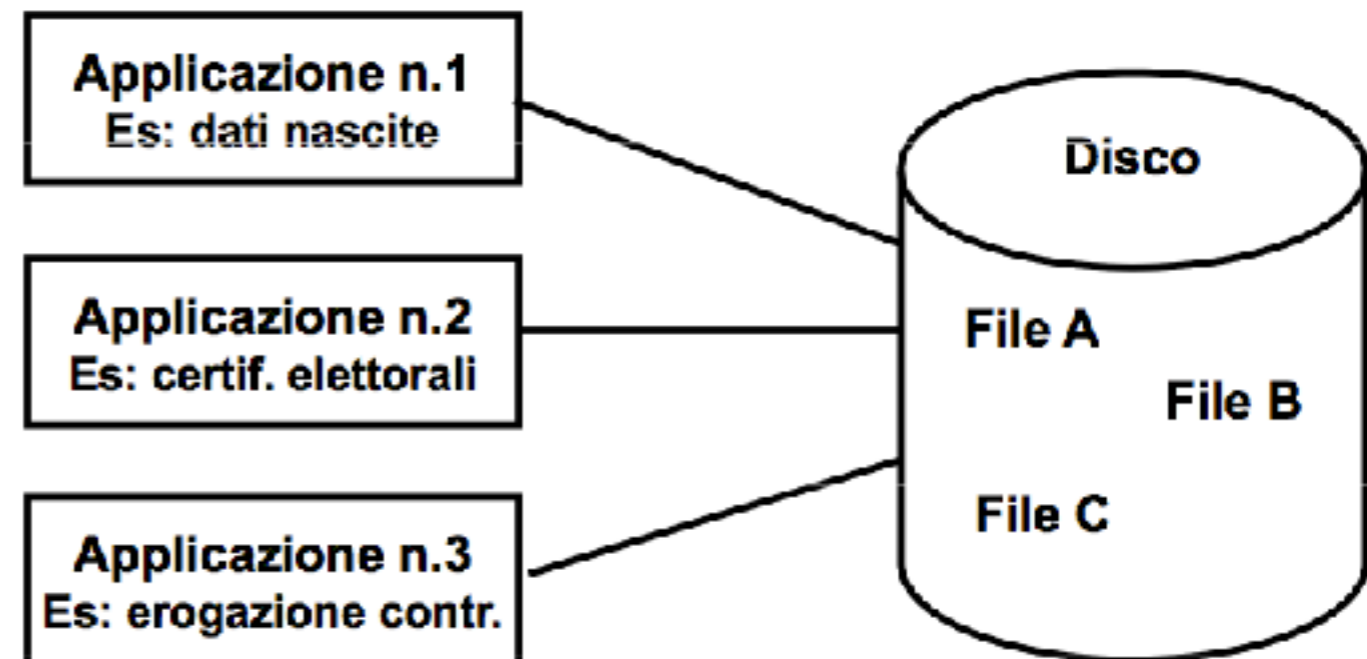
- DBMS sono una componente fondamentale del **sistema informatico** di una organizzazione
- DBMS sono una componente fondamentale del **sistema informativo**
- DBMS sono una componente fondamentale del **sistema organizzativo**

ARCHITETTURA DEI SISTEMI INFORMATICI

- I sistemi informatici sono elementi fondamentali di ogni organizzazione
- Esistono da prima che esistessero i DBMS
- Erano basati su un'architettura diversa rispetto a quella attuale

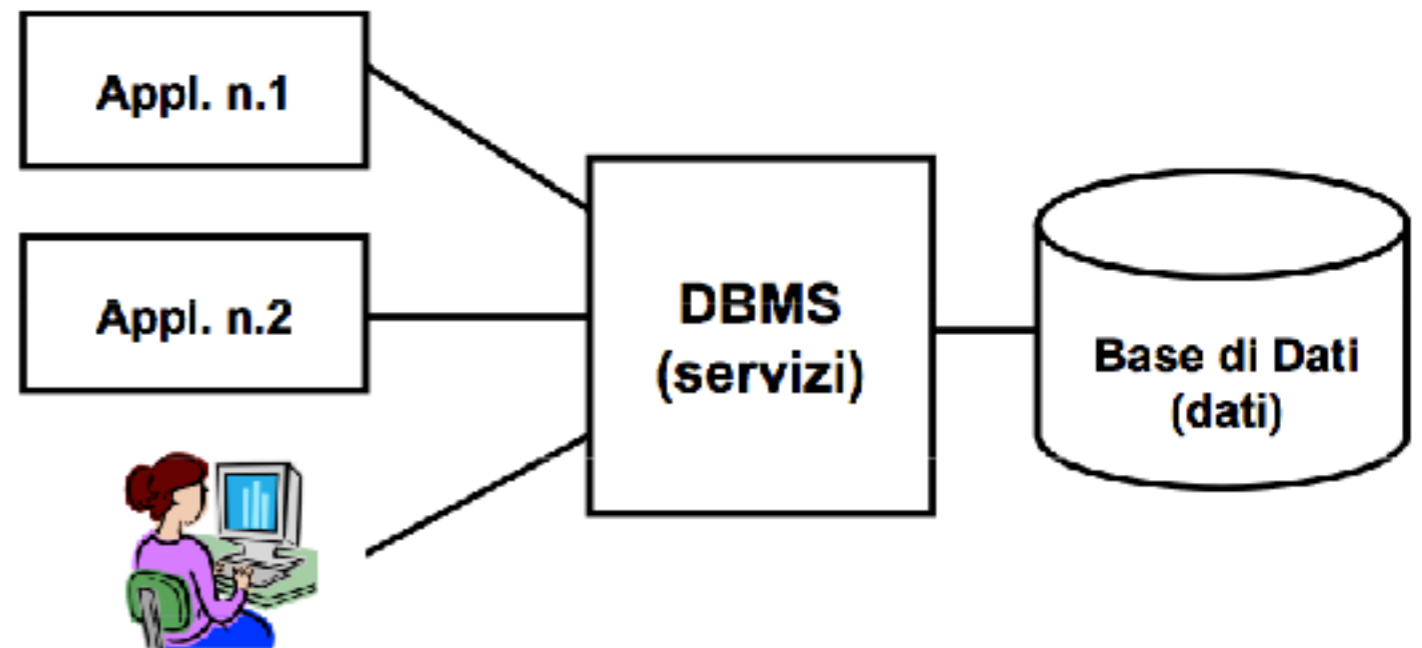
ARCHITETTURA TRADIZIONALE: BASATA SU FILESYSTEM

- 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 e condivisi**
- **in modo**
 - efficiente
 - efficace
- **e garantendo servizi di:**
 - Condivisione
 - Affidabilità
 - Sicurezza
 - Coerenza dei dati



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

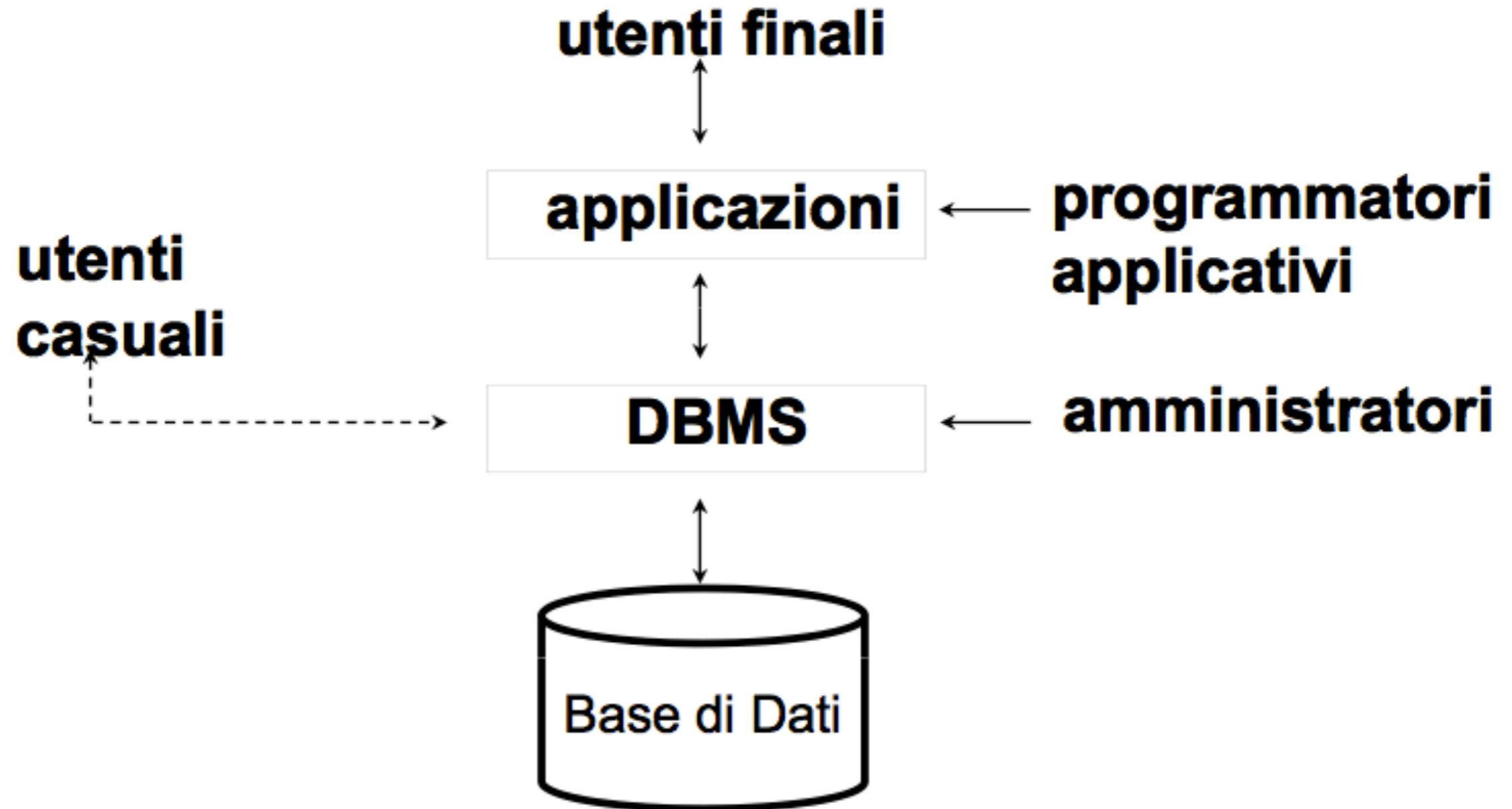
➤ **Efficienza**

- 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

- PostgreSQL (www.postgresql.org)
- MySQL (www.mysql.com)
- SQLite (www.sqlite.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**) contenente la descrizione dei dati, che può essere utilizzata dai programmi che vi accedono

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 ad alto livello
 - le rappresentazioni sottostanti possono essere modificate senza necessità di modifica dei programmi e della struttura logica della base di dati
- 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 NEI DBMS

- 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

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

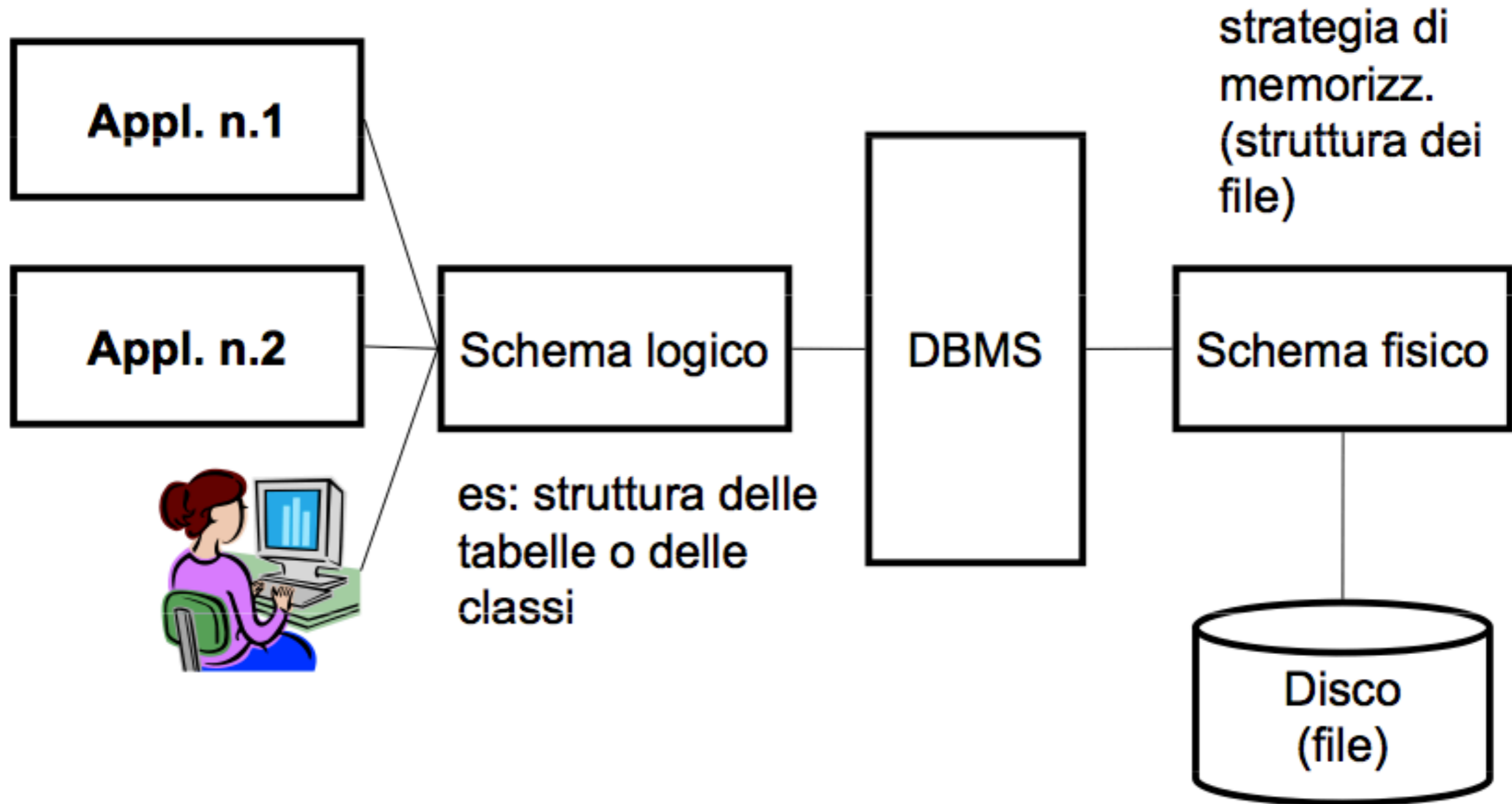
ESAMI

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

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

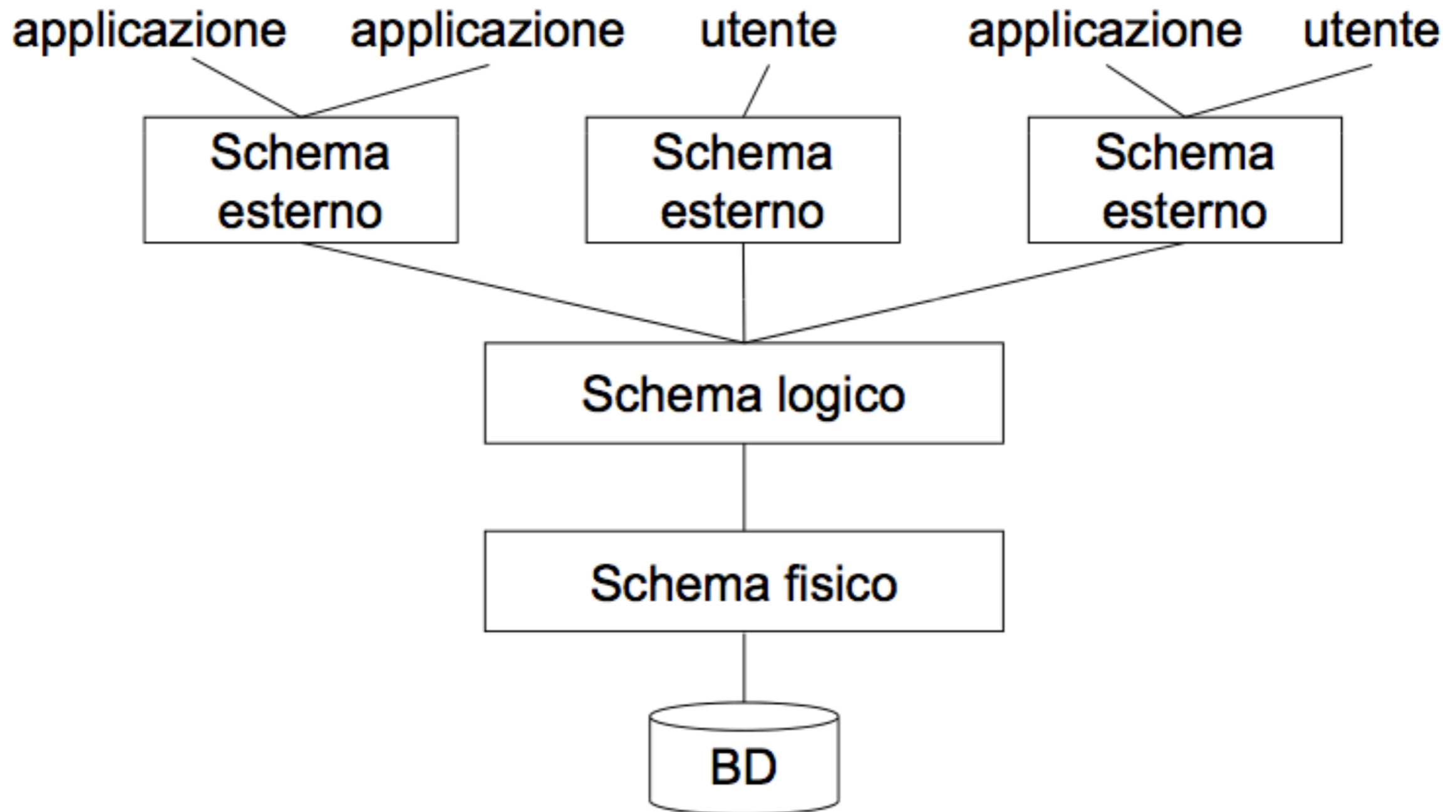
NESSUN ALTRO LIVELLO ? IN REALTÀ SÌ!

- 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' una delle caratteristiche fondamentali 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 TIPOLOGIE 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

- SQL integra le funzionalità di entrambi

UN'OPERAZIONE DDL (SULLO SCHEMA)

CREATE TABLE orario (

 insegnamento CHAR(20),

 docente CHAR(20),

 aula CHAR(4),

 ora CHAR(5))

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 ...

FINE DELLA LEZIONE