

Corso di Laurea in Ingegneria Informatica

# **Basi di Dati**

a.a. 2023-2024

Docenti: Nicola Tonellotto

Francesco Pistolesi

# Principali Obiettivi del Corso

- Imparare a portare a termine un **buon progetto** di base di dati, sia **concettuale** che **logico**
- Imparare ad **analizzare un progetto**, sia concettuale che logico, per verificarne la **consistenza**
- Imparare ad impostare **interrogazioni** (*queries*) per una base di dati relazionale; conoscere il modo in cui il **sistema di gestione** della base di dati (*database management systems*, DBMS) esegue una interrogazione in modo **efficiente**

# Bibliografia

- Libro di testo:
  - Atzeni, Ceri, Fraternali, Paraboschi, Torlone. ***Basi di Dati***.
    - Quinta Edizione. McGraw-Hill Italia, 2018.
    - Sesta Edizione. McGraw-Hill Italia, 2023.

# Comunicazioni coi Docenti

- Email: [nicola.tonellotto@unipi.it](mailto:nicola.tonellotto@unipi.it)
- Ricevimento su appuntamento (online o in presenza)
- Email: [francesco.pistolesi@unipi.it](mailto:francesco.pistolesi@unipi.it)
- Ricevimento iscrivendosi sulla pagina

<http://www.iet.unipi.it/f.pistolesi>

# Nota

- Queste diapositive e le successive sono state rielaborate da N. Tonellotto e G. Vaglini a partire dalle diapositive del libro di testo.

# Introduzione

# Che cos'è l'informatica?

- Scienza del **trattamento razionale**,  
specialmente **per mezzo di macchine automatiche**, **dell'informazione**, considerata  
come supporto alla conoscenza umana e alla  
comunicazione (Académie Française)
- Due anime:
  - **metodologica**
  - **tecnologica**

# Sistema Informativo

- Il **sistema organizzativo** è costituito da **risorse** e **regole** per lo **svolgimento coordinato** di **attività** (processi) per perseguire gli **scopi** propri di **un'organizzazione** (azienda o ente).
  - le risorse possono essere persone, denaro, materiali, informazioni.
- Il **sistema informativo** è la componente del sistema organizzativo che **acquisisce, elabora, conserva, produce le informazioni di interesse** (cioè utili al perseguimento degli scopi); inoltre **esegue/gestisce i processi informativi** (cioè i processi che coinvolgono informazioni)



# Gestione delle Informazioni

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

# 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 (p. es. servizi anagrafici e banche) e che operano da secoli senza impiegare automatizzazioni.
- La parte del sistema informativo che gestisce informazioni con tecnologia informatica è il **sistema informativo automatizzato** (o **sistema informatico**)

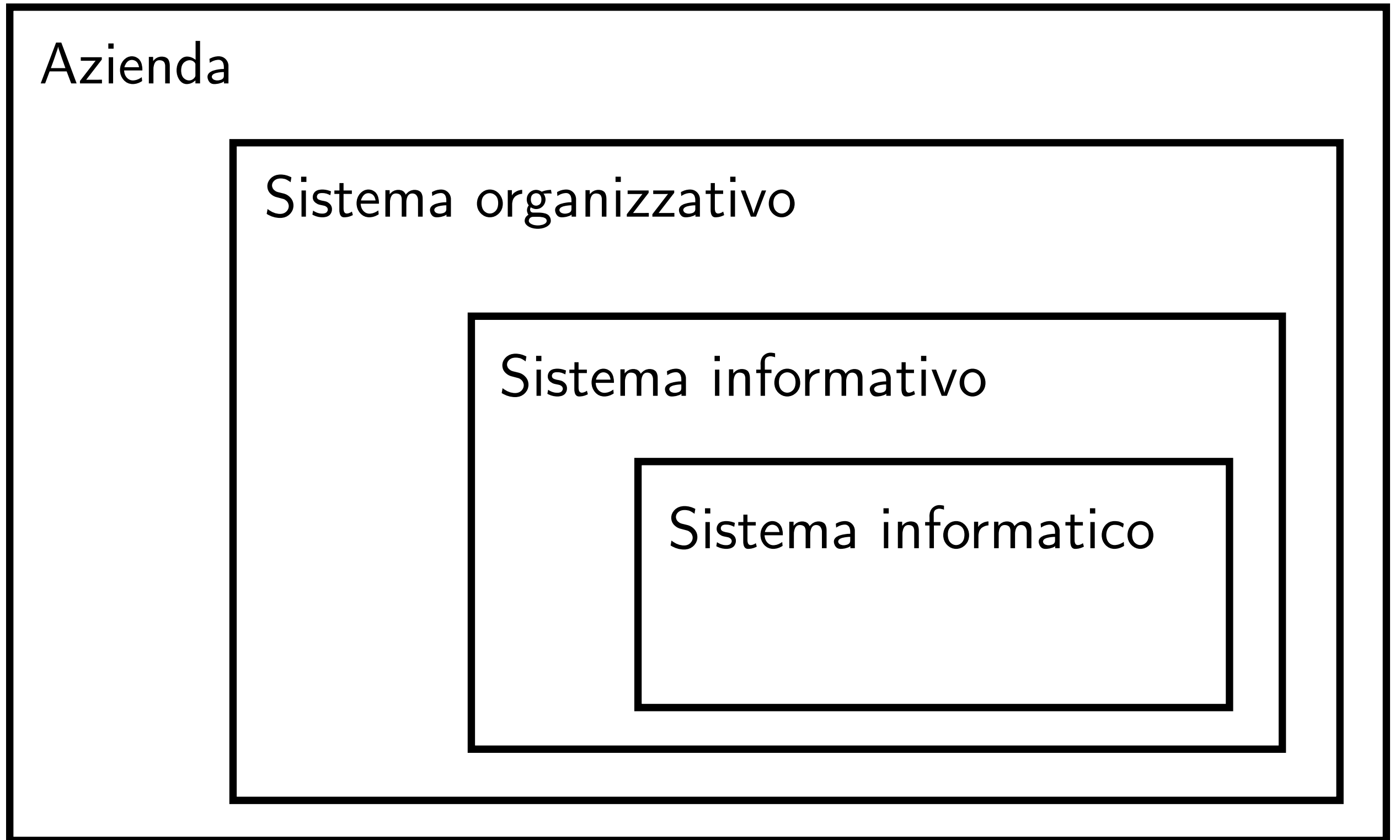
# Sistema Informatico

Azienda

Sistema organizzativo

Sistema informativo

Sistema informatico



# Gestione delle Informazioni

- Nelle attività umane, le informazioni vengono gestite in forme diverse:
  - idee informali
  - linguaggio naturale (scritto o parlato, formale o colloquiale, in varie lingue)
  - disegni, grafici, schemi
  - numeri e codici
- Su vari supporti
  - mente umana, carta, dispositivi elettronici

# Informazioni e Dati

*(definizioni dal Vocabolario della lingua italiana 1987)*

- **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

# Dati e Informazioni

**8 - 17**

# Dati e Informazioni

**8 – 17**

**(8 – 13)**



# Dati e Informazioni

**8 - 17**

**(8 - 13)**

**8 - 14**

# Dati e Informazioni

**8 - 17**

**(8 - 13)**

**8 - 14**

- che cosa significano questi numeri?

# Dati e Informazioni



- che cosa significano questi numeri?
- cartelli stradali, in Finlandia; sono orari!

# Dati e Informazioni



- che cosa significano questi numeri?
- cartelli stradali, in Finlandia; sono orari!
- ma la differenza?

# Dati e Informazioni



- che cosa significano questi numeri?
- cartelli stradali, in Finlandia; sono orari!
- ma la differenza?

# Dati e Informazioni



Lun-Ven



Sabato



- che cosa significano questi numeri?
- cartelli stradali, in Finlandia; sono orari!
- ma la differenza?

# Dati e Informazioni



- che cosa significano questi numeri?
- cartelli stradali, in Finlandia; sono orari!
- ma la differenza?

# Dati e Informazioni

		
<b>8 – 17</b>	<b>(8 – 13)</b>	<b>8 – 14</b>
Lun-Ven	Sabato	Festivo

- che cosa significano questi numeri?
- cartelli stradali, in Finlandia; sono orari!
- ma la differenza?
- senza "interpretazione" il dato serve a ben poco!



# Gestione delle Informazioni

- I dati sono spesso il risultato di forme di organizzazione e codifica delle informazioni
- Ad esempio, nei servizi anagrafici e nel riferimento a persone
  - descrizioni discorsive
  - nome e cognome
  - estremi anagrafici
  - codice fiscale

# Perché i dati?

- La rappresentazione precisa di forme più ricche di informazione e conoscenza è difficile
- I dati costituiscono spesso una risorsa strategica, perché più stabili nel tempo di altre componenti (processi, tecnologie, ruoli umani):
  - ad esempio, i dati delle banche o delle anagrafi

# Basi di Dati

- Il cuore di un sistema informativo automatizzato è la **base di dati** (*database*), cioè un insieme organizzato di dati utilizzati per rappresentare le informazioni di interesse
- Accezione generica, **metodologica**
  - **insieme organizzato di dati utilizzati per il supporto allo svolgimento delle attività di un ente** (azienda, ufficio, persona)
- Accezione specifica, **metodologica** e **tecnologica**
  - **insieme di dati gestito da un DBMS**

# Basi di Dati

- Il cuore di un sistema informativo automatizzato è la **base di dati** (*database*), cioè un insieme organizzato di dati utilizzati per rappresentare le informazioni di interesse
- Le Basi di Dati
  - hanno dimensioni (molto) maggiori della memoria centrale dei sistemi di calcolo utilizzati
  - hanno un tempo di vita indipendente dalle singole esecuzioni dei programmi che le utilizzano (**persistenza** dei dati)

# Sistema di Gestione di Basi di Dati

- Il Sistema di Gestione di Basi di Dati (*database management system*, DBMS) gestisce collezioni di dati:
  - grandi
  - persistenti
  - condivise
- garantendo:
  - privacy
  - affidabilità
  - efficienza
  - efficacia

# Grandi

- Dimensioni (molto) maggiori della memoria centrale dei sistemi di calcolo utilizzati
- Il limite deve essere solo quello fisico dei dispositivi
- Esempi di dimensioni molto grandi
  - 500 Gigabyte (dati transazionali)
  - 10 Terabyte (dati decisionali)
  - 500 Terabyte (dati scientifici)
  - 2.25 miliardi di pagine Web

# Persistenti

- Hanno un tempo di vita indipendente dalle singole esecuzioni dei programmi che le utilizzano

# 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)



# Esempio

## 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

## 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

# Problemi

- **Ridondanza:**
  - informazioni ripetute
- Rischio di **incoerenza:**
  - le versioni possono non coincidere

# Archivi e Basi di Dati

# Archivi e Basi di Dati

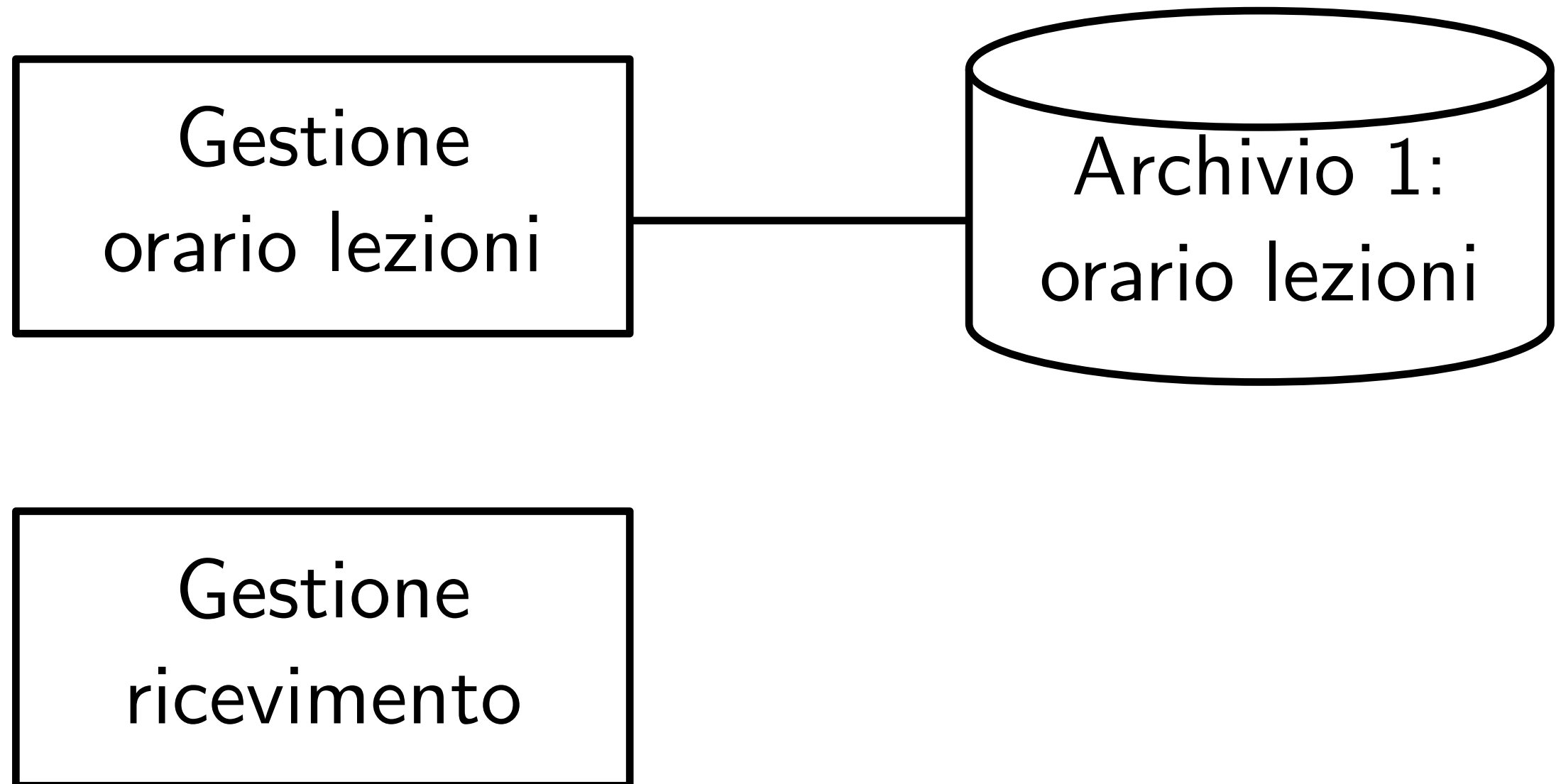
Gestione  
orario lezioni

# Archivi e Basi di Dati

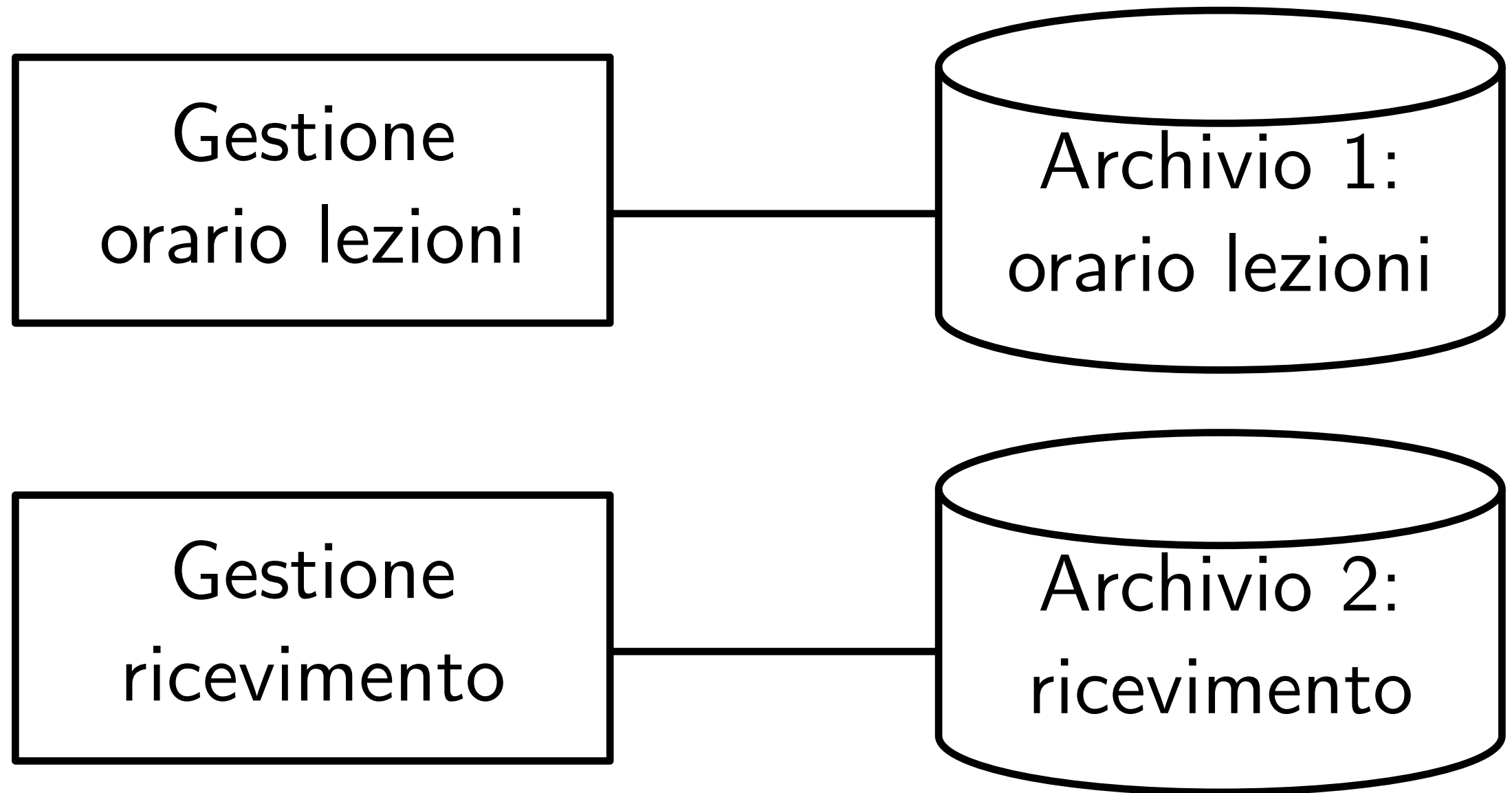
Gestione  
orario lezioni

Gestione  
ricevimento

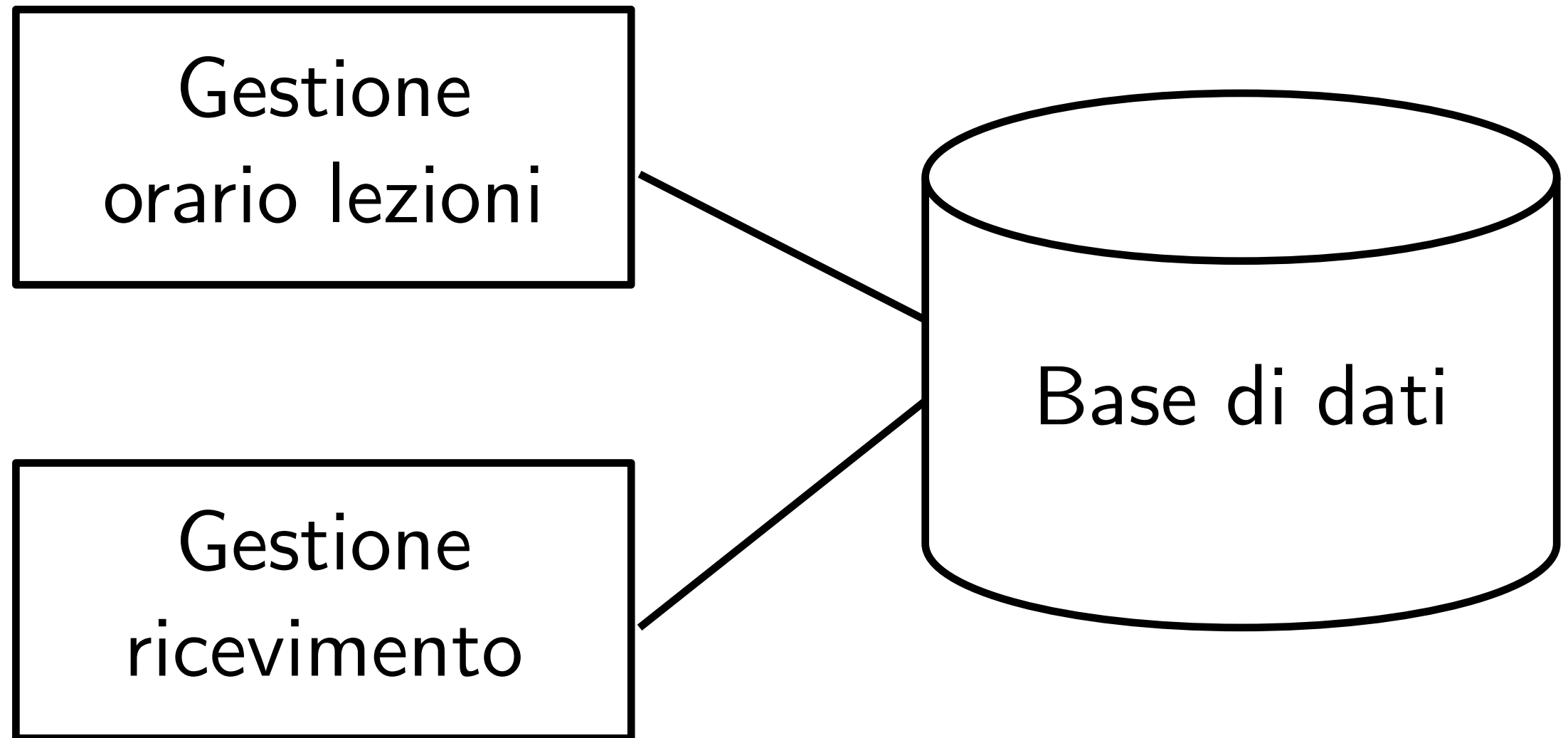
# Archivi e Basi di Dati



# Archivi e Basi di Dati



# Archivi e Basi di Dati





# Le Basi di Dati sono Condivise

- Una base di dati è una risorsa **integrata, condivisa** fra applicazioni
- Conseguenze
  - Attività diverse su dati condivisi:
    - meccanismi di **autorizzazione**
  - Accessi di più utenti ai dati condivisi:
    - controllo della **concorrenza**

# I DBMS garantiscono ... privacy

- Si possono definire meccanismi di **autorizzazione**
  - l'utente  $A$  è autorizzato a leggere tutti i dati e a modificare  $X$
  - l'utente  $B$  è autorizzato a leggere i dati  $X$  e a modificare  $Y$

# I DBMS garantiscono ... affidabilità

- **Affidabilità** (per le basi di dati):
  - resistenza a malfunzionamenti hardware e software
- Una base di dati è una risorsa pregiata e quindi deve essere conservata a lungo termine
- Tecnica fondamentale:
  - gestione delle **transazioni**

# Transazione

- **Insieme di operazioni** da considerare **indivisibile** ("*atomico*"), **corretto** anche in presenza di **concorrenza** e con **effetti definitivi**

# Le transazioni sono ... atomiche

- Una sequenza di operazioni correlate:
  - trasferimento di fondi da un conto  $A$  ad un conto  $B$ :
    - o si fanno il prelevamento da  $A$  e il versamento su  $B$
    - o nessuno dei due
- Deve essere eseguita per intero o per niente:
  - o si fanno il prelevamento da  $A$  e il versamento su  $B$
  - o nessuno dei due

# Le transazioni sono ... concorrenti

- L'effetto di transazioni concorrenti deve essere coerente
  - se due assegni emessi sullo stesso conto corrente vengono incassati contemporaneamente
    - ... si deve evitare di trascurarne uno!
- se due agenzie richiedono lo stesso posto (libero) su un treno
  - ... si deve evitare di assegnarlo due volte!

# I risultati delle transazioni sono permanenti

- La **conclusione positiva** di una transazione corrisponde ad un **impegno** (*commit*) a **mantenere traccia** del risultato in modo definitivo, anche in presenza di guasti e di esecuzione concorrente

# **I DBMS debbono essere ... efficienti**

- Cercano di utilizzare al meglio le risorse di spazio di memoria (principale e secondaria) e tempo (di esecuzione e di risposta)
- I DBMS, con tante funzioni, rischiano l'inefficienza e per questo ci sono grandi investimenti e competizione
- L'efficienza è anche il risultato della qualità delle applicazioni



# I DBMS debbono essere ... efficaci

- Cercano di rendere produttive le attività dei loro utilizzatori, offrendo funzionalità articolate, potenti e flessibili
- Il sistema informatico deve essere adeguatamente dimensionato e la base di dati ben progettata (e realizzata)

# DBMS vs File System

- La gestione di insiemi di dati grandi e persistenti è possibile anche attraverso sistemi più semplici — gli ordinari *file system* dei sistemi operativi
- I *file system* prevedono forme rudimentali di condivisione: “tutto o niente”
- I DBMS estendono le funzionalità dei file system, fornendo più servizi ed in maniera integrata

# Descrizione dei Dati

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

# Descrizione dei Dati nei DBMS

- I programmi fanno riferimento ai dati, ma la loro struttura in memoria deve poter essere modificata senza dover modificare i programmi
- Viene introdotto il concetto di
  - **modello dei dati** : insieme di costrutti utilizzati per organizzare i dati di interesse e descriverne la dinamica
  - il modello dei dati fornisce ai programmi applicativi una **vista astratta dei dati**

# Schema e Istanza

- In ogni base di dati esistono:
  - lo **schema**, sostanzialmente invariante nel tempo, che ne descrive la struttura
    - es.: le intestazioni delle tabelle (cfr. slide successiva)
  - l'**istanza**, i valori attuali, che possono cambiare anche molto rapidamente
    - es.: le “righe” di ciascuna tabella (cfr. slide successiva)

# Esempio

Insegnamento	Docente	Aula	Ora
Analisi Matem. I	Luigi Neri	N1	8:00
Basi di Dati	Piero Rossi	N2	9:45
Chimica	Nicola Mori	N1	9:45
Fisica I	Mario Bruni	N1	11:45
Fisica II	Mario Bruni	N3	9:45
Sistemi Inform.	Piero Rossi	N3	8:00

# Esempio

# Esempio

Insegnamento	Docente	Aula	Ora
--------------	---------	------	-----



# Esempio

Lo **schema** della base di dati

Insegnamento	Docente	Aula	Ora
--------------	---------	------	-----

# Esempio

Lo **schema** della base di dati

Insegnamento	Docente	Aula	Ora
Analisi Matem. I	Luigi Neri	N1	8:00
Basi di Dati	Piero Rossi	N2	9:45
Chimica	Nicola Mori	N1	9:45
Fisica I	Mario Bruni	N1	11:45
Fisica II	Mario Bruni	N3	9:45
Sistemi Inform.	Piero Rossi	N3	8:00

# Esempio

Lo **schema** della base di dati

Insegnamento	Docente	Aula	Ora
Analisi Matem. I	Luigi Neri	N1	8:00
Basi di Dati	Piero Rossi	N2	9:45
Chimica	Nicola Mori	N1	9:45
Fisica I	Mario Bruni	N1	11:45
Fisica II	Mario Bruni	N3	9:45
Sistemi Inform.	Piero Rossi	N3	8:00

L'**istanza** della base di dati

# Modelli dei Dati

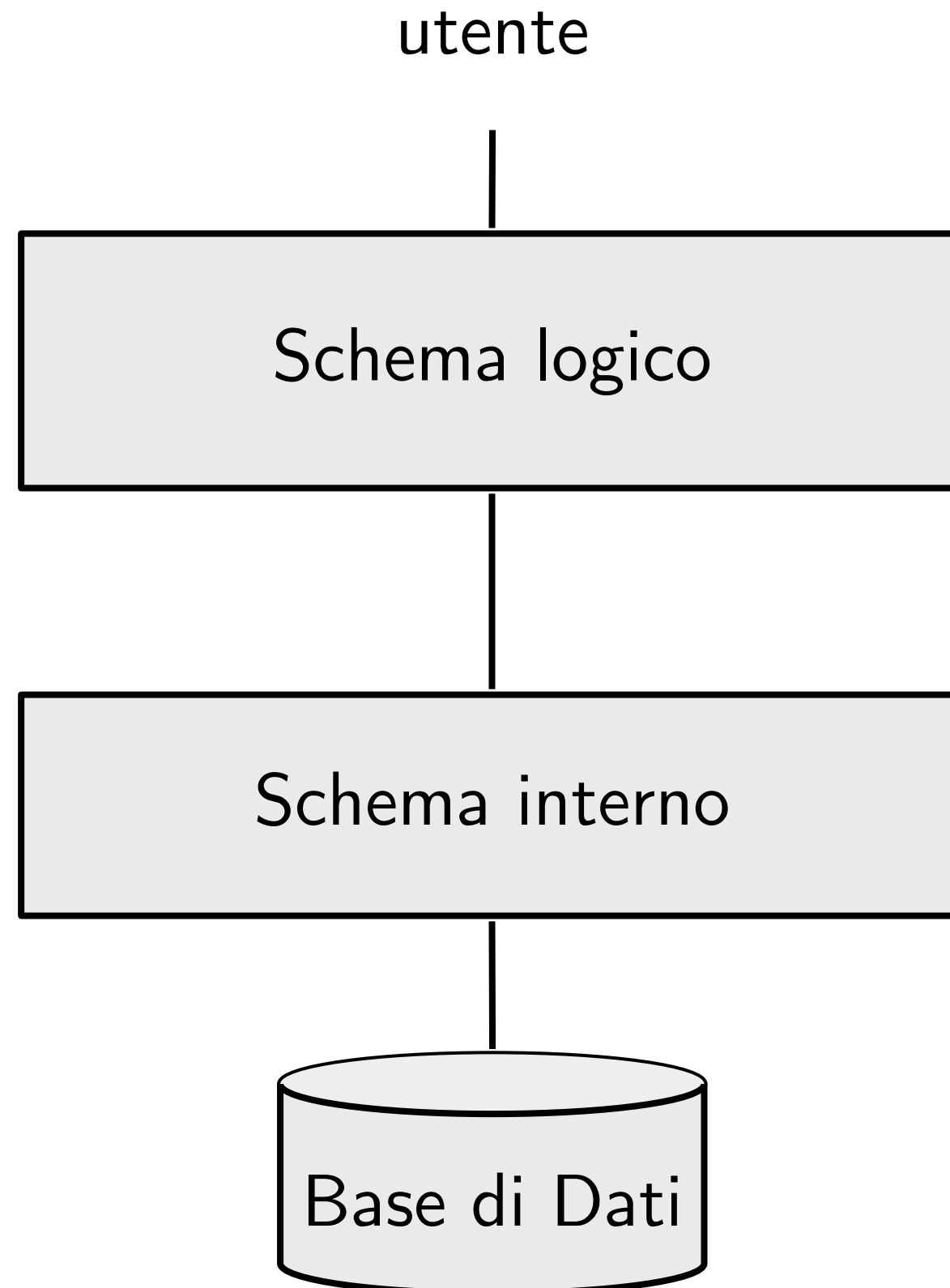
- **Modelli Logici**

- Adottati nei DBMS esistenti per l'organizzazione dei dati
  - utilizzati dai programmi
  - indipendenti dalle strutture fisiche
- Esempi: **relazionale**, reticolare, gerarchico, a oggetti, basato su XML

- **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)

# Architettura Semplificata di un DBMS



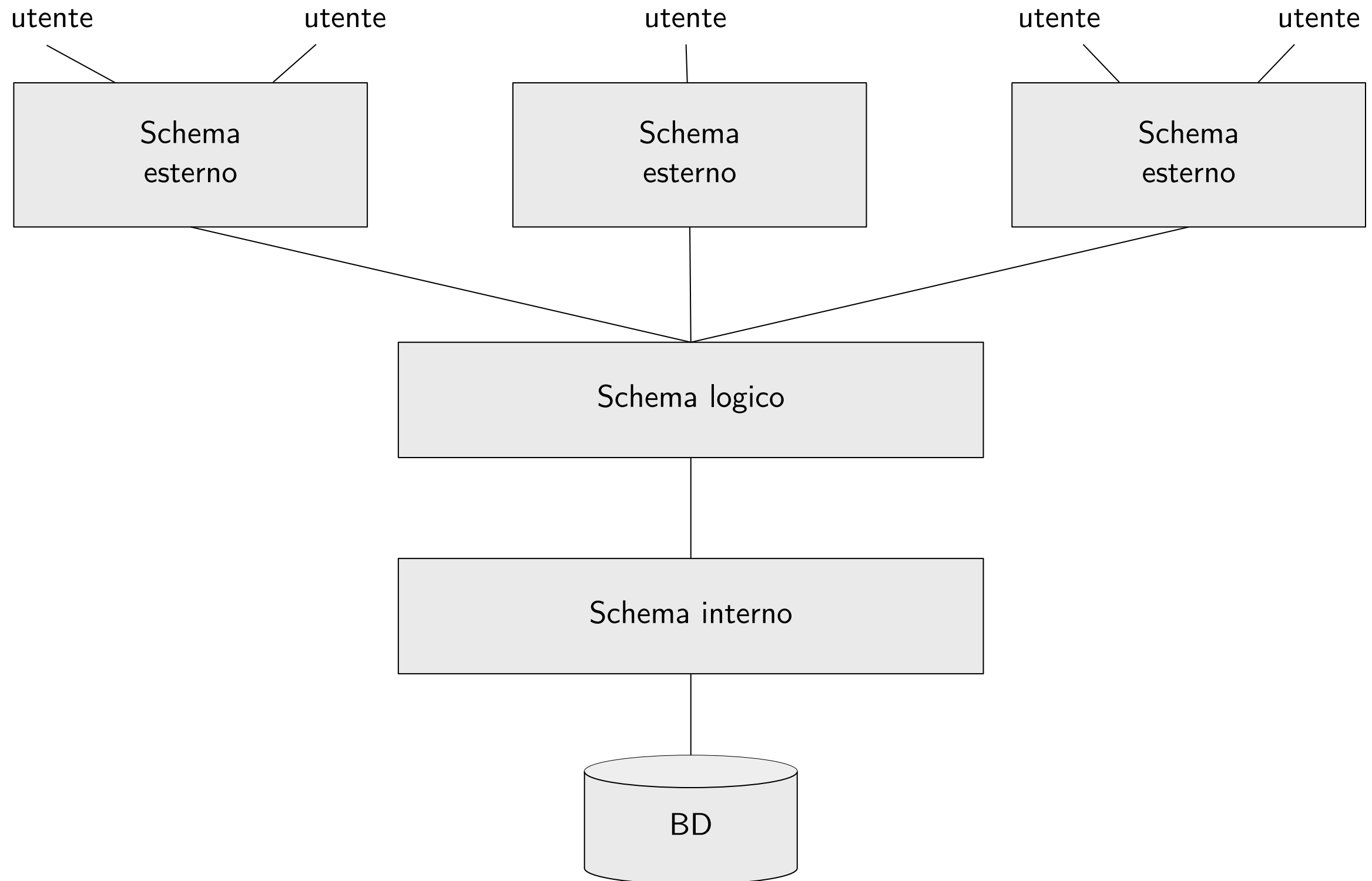
# Architettura Semplificata di un DBMS

- **Schema logico:**
  - descrizione della base di dati nel modello logico
    - ad esempio, la struttura della tabella
- **Schema interno** (o fisico):
  - rappresentazione dello schema logico per mezzo di strutture memorizzazione (*file*)
    - ad esempio, record con puntatori, ordinati in un certo modo
- Il livello logico è **indipendente** da quello fisico:
  - una tabella è utilizzata nello stesso modo qualunque sia la sua realizzazione fisica (che può anche cambiare nel tempo)
  - In questo corso vedremo solo il livello logico e non quello fisico

# Linguaggi per Basi di Dati

- La disponibilità di vari linguaggi e interfacce per la definizione di schemi e per la lettura/modifica di istanze contribuisce all'efficacia del DBMS
  - **Linguaggi testuali** interattivi (SQL)
  - **Comandi** (SQL) immersi in un **linguaggio ospite** (Java, C++, ...)
  - Con **interfacce** amichevoli (senza linguaggio testuale come Access)
- Una distinzione terminologica
  - ***data definition language*** (DDL) per la definizione di schemi (logici, fisici)
  - ***data manipulation language*** (DML) per l'interrogazione e l'aggiornamento di (istanze di) basi di dati

# Architettura a tre livelli per DBMS





# Architettura a tre livelli per DBMS

- **Schema logico:**
  - descrizione dell'intera base di dati nel modello logico “principale” del DBMS
- **Schema interno (o fisico):**
  - rappresentazione dello schema logico per mezzo di strutture fisiche di memorizzazione
- **Schema esterno:**
  - descrizione di parte della base di dati in un modello logico (“viste” parziali, derivate, anche in modelli diversi)

# Indipendenza dei Dati

- L'accesso ai dati avviene solo tramite il livello esterno (che può coincidere con il livello logico)
- **Indipendenza fisica:**
  - il livello logico e quello esterno sono indipendenti da quello fisico
    - una tabella è utilizzata nello stesso modo qualunque sia la sua realizzazione fisica
    - la realizzazione fisica può cambiare senza che debbano essere modificati i programmi
- **Indipendenza logica:**
  - il livello esterno è indipendente da quello logico
  - aggiunte o modifiche alle viste non richiedono modifiche al livello logico
  - modifiche allo schema logico che lascino inalterato lo schema esterno sono trasparenti

# Personaggi

- **Progettisti e realizzatori di DBMS**
- **Progettisti della base di dati e amministratori della base di dati**
- **Progettisti e programmatori di applicazioni**
- **Utenti:**
  - **utenti finali:** eseguono applicazioni predefinite (transazioni)
  - **utenti casuali:** eseguono operazioni non previste a priori, usando linguaggi interattivi

# 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

- Costo dei prodotti e della transizione verso di essi
- Non scorporabilità delle funzionalità (con riduzione di efficienza)