

Corso di Laurea in Ingegneria Informatica

Basi di Dati

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
- Ricevimento su appuntamento (online o in presenza)

- Email: francesco.pistolesi@unipi.it
- Ricevimento su appuntamento (online o in presenza)

Modalità di Esame

- L'esame si compone di:
 - **pretest**: test a risposta multipla il cui superamento (punteggio superiore al 60%) permette di accedere alla prova pratica e alla prova orale
 - **prova pratica**: esercizio di progettazione di un database + esercizio di programmazione MySQL sul database usato a lezione
 - **prova orale**: domande a risposta aperta ed esercizi sulla parte di teoria
- Il **pretest** può essere sostenuto ad **ogni appello**
- La prova **pratica** e la prova **orale** devono essere svolte nello **stesso appello** (ma non il pretest)
- La prova pratica e la prova orale possono essere sostenute **solo dopo aver superato** il pretest e l'esame di Fondamenti di Programmazione
- **Non si può usare alcun tipo di materiale** durante il pretest, né durante la prova pratica, né durante la prova orale
- Il voto complessivo dell'esame è calcolato come la **media dei voti della prova pratica e della prova orale**, ciascuno espresso in trentesimi, **arrotondata all'intero più vicino**

Prove In Itinere (1)

- Date potenziali
 - Venerdì 11 Aprile 2025
 - Venerdì 23 Maggio 2025
- Per partecipare a ciascuna prova in itinere, è **obbligatorio iscriversi online** tramite il portale esami.unipi.it **entro tre giorni dalla data prevista** da ciascuna prova

Prove In Itinere (2)

- Ogni prova in itinere è composta da **20 domande a scelta multipla** da compilare online
- Ogni **risposta corretta** vale **1,5 punti**, ogni **risposta sbagliata o incompleta** vale **0 punti**
- Il **voto minimo** per il superamento di ciascuna prova in itinere è **15**, corrispondente a 10 domande corrette
- Il **non raggiungimento del voto minimo** (15) alla prima prova in itinere **preclude alla partecipazione** alla seconda prova in itinere
- Il **voto finale** delle prove in itinere è calcolato come **media dei voti conseguiti nelle due prove in itinere**, arrotondata all'intero più vicino

Prove In Itinere (3)

- Se il **voto finale delle prove** in itinere è:
 - **maggiore o uguale a 18**: lo studente, in sede di appello di esame, potrà optare tra una delle due seguenti alternative:
 - la **convalida del voto** finale delle prove in itinere come voto finale del corso;
 - il sostenimento unicamente della **prova orale**, che concorrerà a formare, insieme al voto finale delle prove in itinere, il voto finale del corso. In caso di **valutazione insufficiente della prova orale**, sarà necessario sostenere l'esame per intero, in un appello successivo.
 - **minore di 18 e maggiore o uguale a 15**: lo studente, in sede di appello di esame, potrà **sostenere unicamente la prova orale**, che concorrerà a formare, insieme al voto finale delle prove in itinere, il voto finale del corso. In caso di **valutazione insufficiente della prova orale**, sarà necessario sostenere l'esame per intero, in un appello successivo.
- **minore di 15**: lo studente dovrà sostenere l'esame per intero.

Prove In Itinere (4)

- È sempre possibile sostenere l'esame per intero ignorando il risultato conseguito nelle prove in itinere
- Ogni partecipante alle prove in itinere **deve portare il proprio laptop** (no tablet, no smartphone), con **almeno di 2 ore di autonomia**
- **Non si può usare alcun tipo di materiale** durante le prove in itinere

Introduzione

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 (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**)

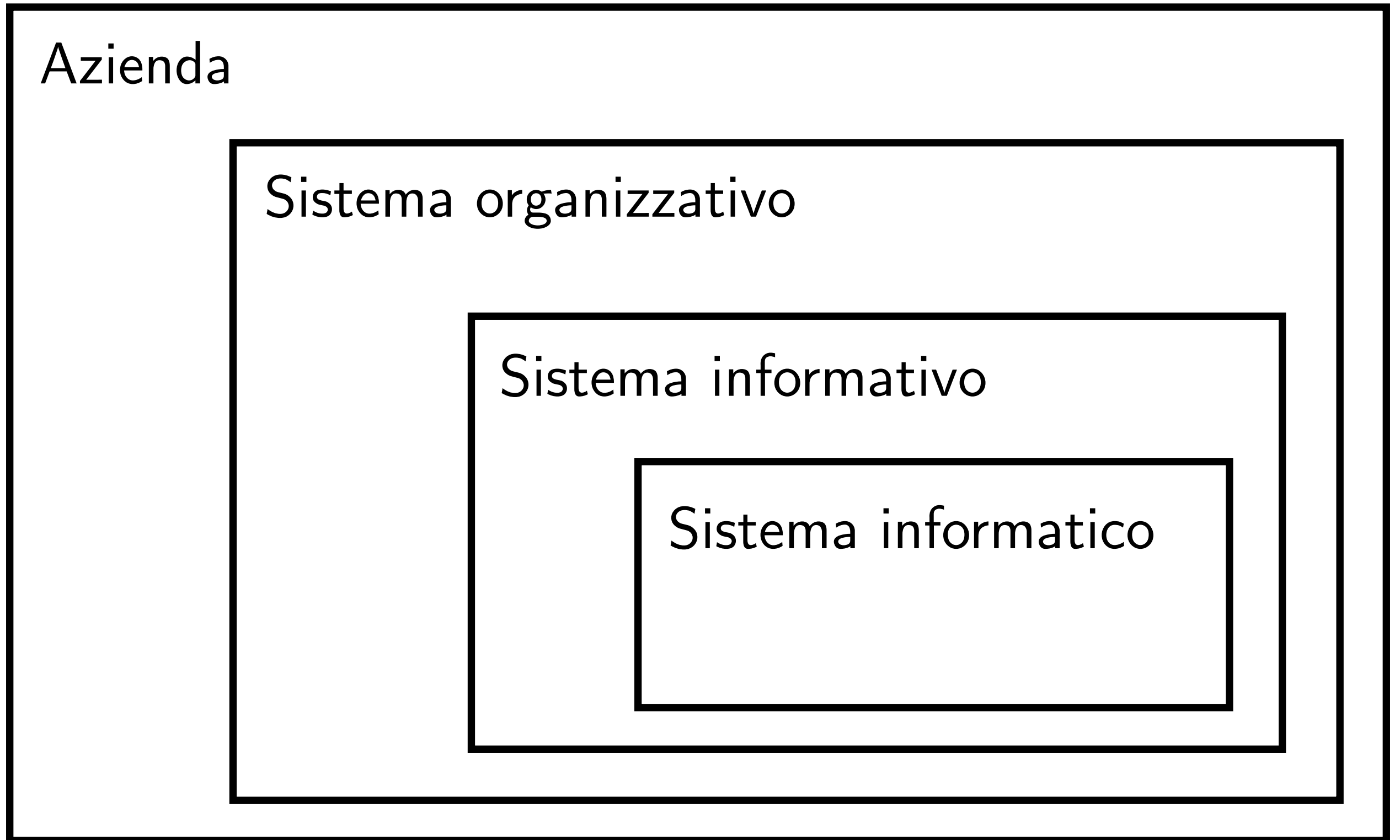
Sistema Informatico

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

Informazioni e Dati

- **informazione:**

- La temperatura di fusione dell'oro è 1064 °C

- **dato:**

- 01000001 01110101 00000100 00101000 01000011

au

1064

C

Dati e Informazioni



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
 - **insieme organizzato di dati utilizzati per il supporto allo svolgimento delle attività di un ente** (azienda, ufficio, persona)
- Accezione specifica
 - **insieme di dati gestito da un programma di gestione specializzato**

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 enormi**, (molto) maggiori della memoria principale dei sistemi di calcolo utilizzati
 - hanno un **tempo di vita indipendente** dalle singole esecuzioni dei programmi che le utilizzano
 - hanno **diverse tipologie** di informazioni e **diversi utenti**

Sistema di Gestione di Basi di Dati

- Il sistema di gestione di basi di dati (*DataBase Management System*, DBMS) è il programma specifico che gestisce collezioni di dati:
 - grandi
 - persistenti
 - condivise
- garantendo:
 - privacy
 - affidabilità
 - efficienza
 - efficacia

Volume dei dati

- Dimensioni (molto) maggiori della memoria principale dei sistemi di calcolo utilizzati
- Il limite deve essere solo quello fisico dei dispositivi
- Esempi di dimensioni molto grandi
 - 500 Gigabyte (dati bancari)
 - 10 Terabyte (dati societari)
 - 500 Terabyte (dati scientifici)
 - Netflix: Archivia **miliardi di ore di film e serie** con i dati degli utenti
 - Amazon: Registra **ogni acquisto, recensione**, e interazione degli utenti
 - Google: Indicizza e aggiorna **miliardi di pagine web**

Persistenza dei dati

- Hanno un tempo di vita indipendente dalle singole esecuzioni dei programmi che le utilizzano
- I dati non vengono persi dopo la chiusura dell'applicazione o lo spegnimento del computer
- I dati rimangono salvati in modo permanente finché non vengono modificati o cancellati esplicitamente
- Esempio:
 - Gli esami degli studenti in un database
 - Anche se il sistema viene spento di notte, i voti restano salvati e accessibili il giorno dopo 😊💧

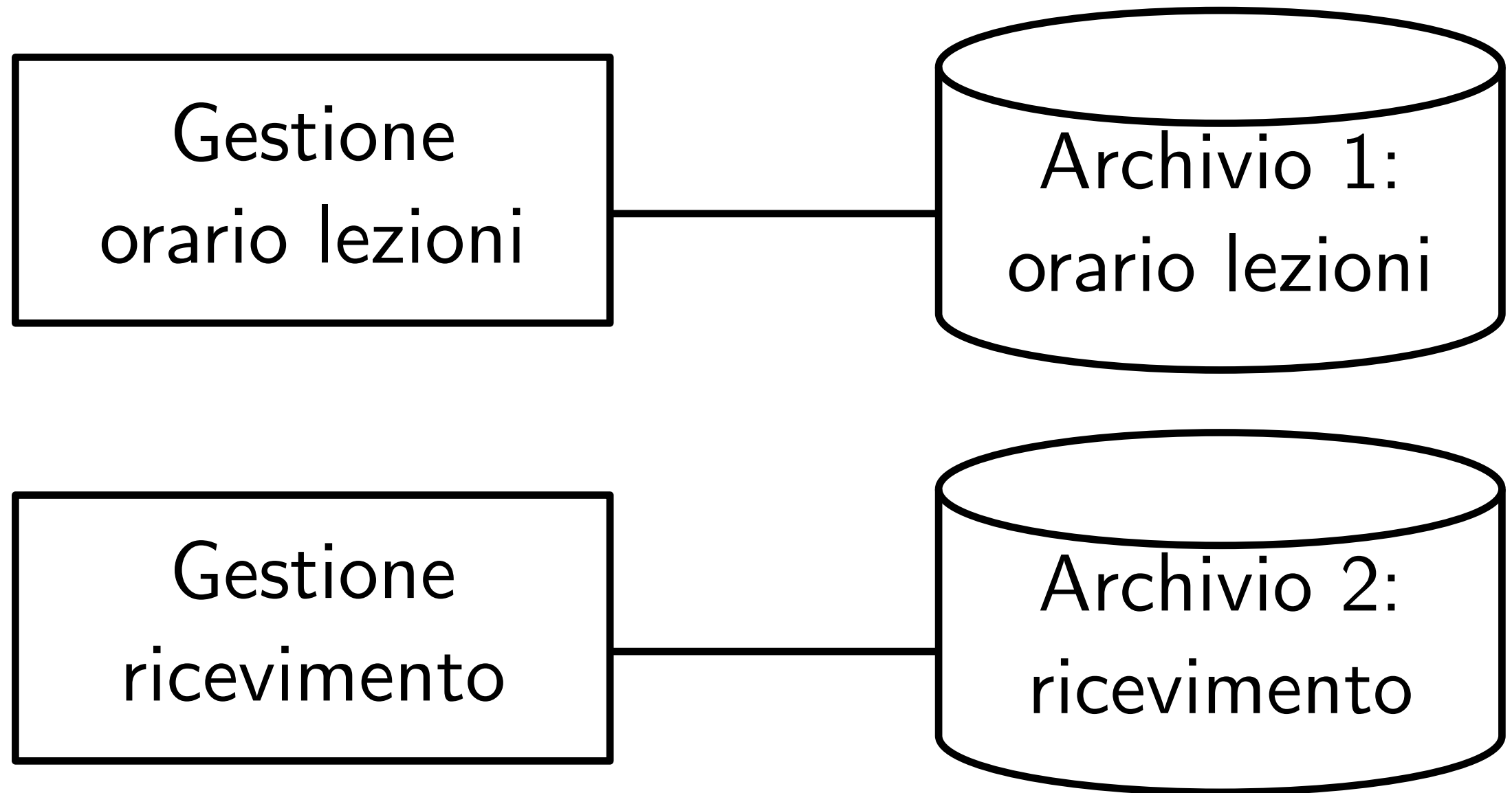
Condivisione dei dati

- Ogni organizzazione (specie se grande) è divisa in settori o comunque svolge diverse attività
- Ciascun settore/attività ha un (sotto)sistema informativo (non necessariamente distinto e indipendente)
- Esempio:
 - Il database degli studenti è condiviso tra segreteria, docenti e studenti
 - Uno studente può vedere i suoi voti, ma non modificarli
 - Un docente può inserire nuovi voti per i suoi corsi
 - La segreteria può gestire le iscrizioni agli esami

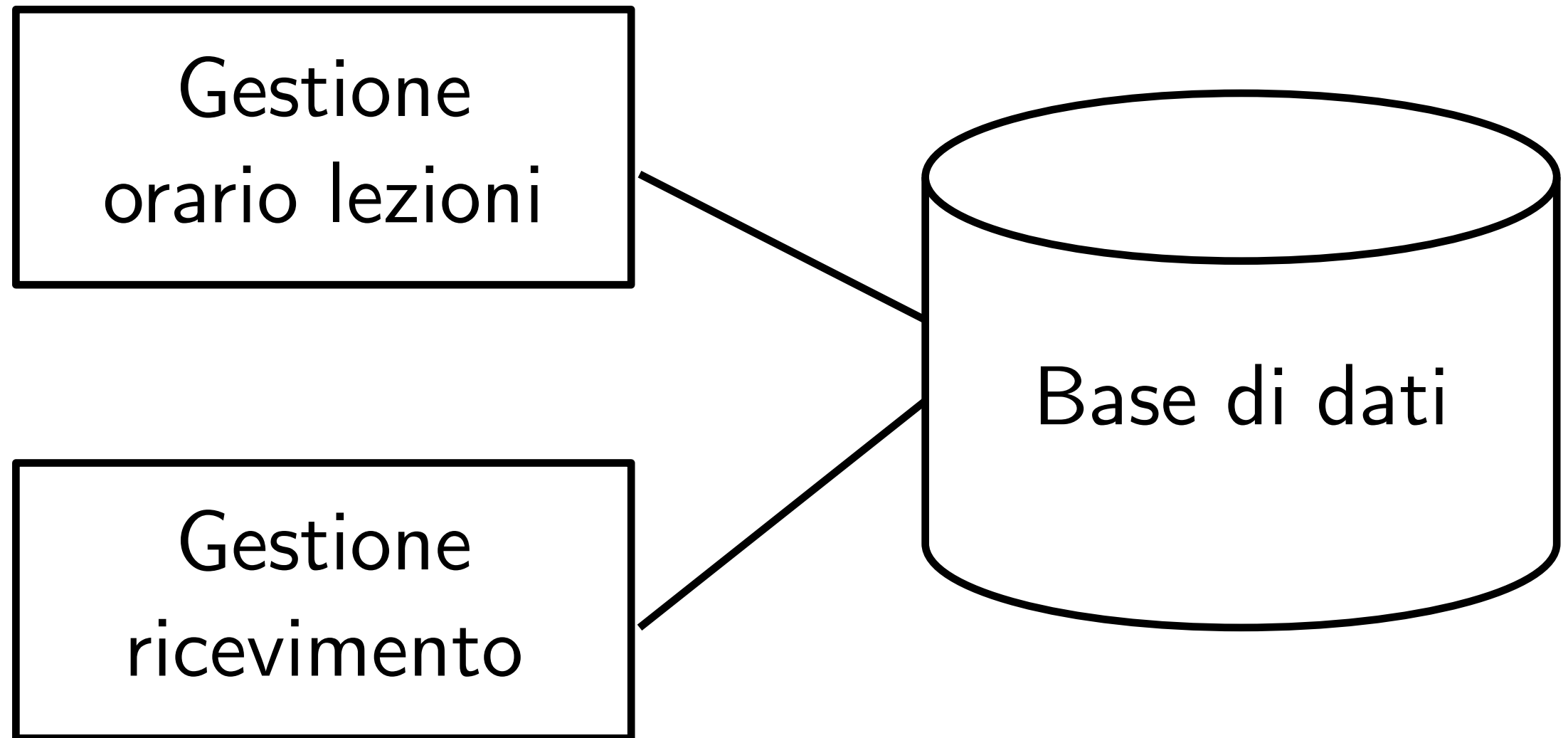
Problemi della condivisione

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

Archivi e Basi di Dati



Archivi e Basi di Dati



Condivisione dei dati

- 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

- Strumenti per **proteggere i dati** sensibili e **impedire accessi** non autorizzati
- 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
- Esempi:
 - Database ospedaliero
 - Database bancario

I DBMS garantiscono affidabilità

- **Resistenza a malfunzionamenti** hardware e software
- Un database affidabile **non perde dati, non si corrompe e garantisce operazioni coerenti**, anche in **situazioni critiche** come crash del sistema o blackout
- Esempi:
 - in un bonifico bancario, se il sistema si blocca mentre i soldi vengono trasferiti, il DBMS assicura che il denaro non sparisca nel nulla
 - se due persone prenotano lo stesso biglietto aereo nello stesso momento, il DBMS gestisce la priorità evitando doppie prenotazioni

Transazione

- Le **transazioni** sono uno degli strumenti principali con cui i DBMS garantiscono affidabilità
- **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 a mantenere traccia** del risultato in modo definitivo, anche in presenza di guasti e di esecuzione concorrente

I DBMS debbono essere efficienti

- Devono essere in grado di gestire **grandi quantità di dati e operazioni nel minor tempo possibile**, utilizzando le risorse del sistema in modo ottimale.
- Miglior compromesso tra spazio di memoria e tempo di esecuzione
- I DBMS, con tante funzionalità da garantire, rischiano l'inefficienza e per questo ci sono grandi investimenti e competizione
- Esempi:
 - Un sito di e-commerce deve recuperare i prodotti in pochi istanti, altrimenti gli utenti abbandonano il carrello
 - Un sistema sanitario deve archiviare milioni di cartelle cliniche e permettere di trovarle rapidamente
 - Un servizio di streaming video deve scalare da pochi utenti a milioni senza interruzioni del servizio

I DBMS debbono essere efficaci

- Devono fornire **risultati corretti e completi**, soddisfacendo le necessità degli utenti e delle applicazioni **in modo affidabile e funzionale**
- Esempio: database universitario
 - Efficiente: Trova i voti di uno studente rapidamente
 - Efficace: Mostra tutti i voti corretti

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

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

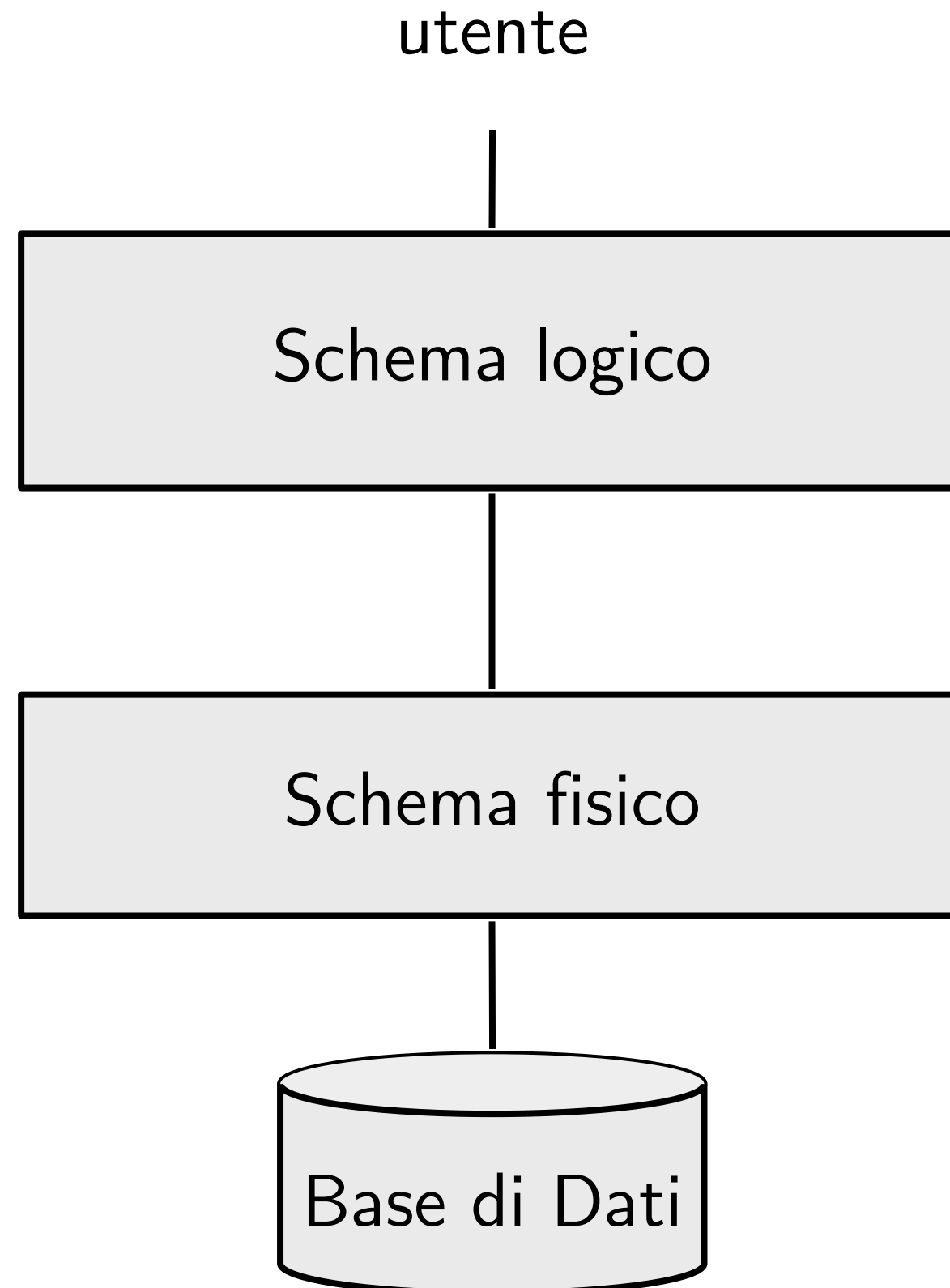
- **Modelli Logici**

- 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



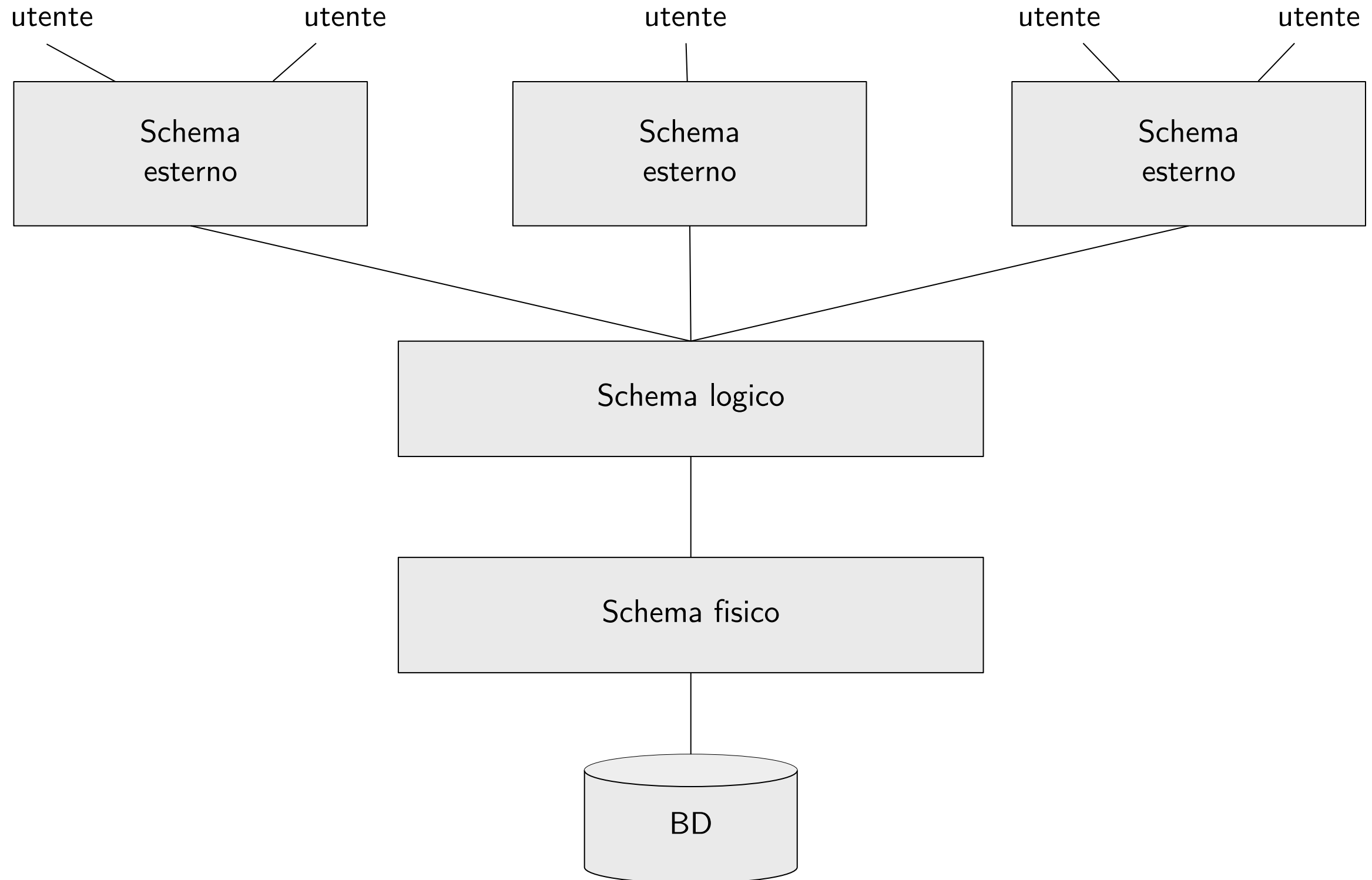
Architettura Semplificata di un DBMS

- **Schema logico:**
 - descrizione della base di dati nel modello logico
 - ad esempio, la struttura della tabella
- **Schema fisico** (o interno):
 - 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 fisico (o interno):**
 - 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)