

Informatica

Introduzione alle basi di dati

Prof. Giovanni Giuffrida

e-mail: giovanni.giuffrida@dmi.unict.it

Materiale didattico

- Atzeni,Ceri,Paraboschi,Torlone, Basi di Dati,Modelli e linguaggi di interrogazione, terza edizione, McGraw-Hill 2002.
- Albano-Ghelli-Orsini, Basi di Dati Relazionali e a Oggetti,Zanichelli,1997
- Ullman, Basi di Dati e Basi di Conoscenza
- Access: Manuale utente
- Slides del corso
 - Scaricabili dal sito del corso
 - Consiglio: utilizzate per prendere appunti

Corso di Informatica
Prof. Giovanni Giuffrida

Introduzione alle Basi di Dati

Sistema informativo

- Componente (sottosistema) di qualsiasi organizzazione
 - Azienda privata, pubblica, info private (PIM), famiglia, etc.
- Gestisce (acquisisce, elabora, conserva, produce) le informazioni di interesse (i.e., utilizzate per il perseguimento degli scopi dell'organizzazione)
 - ogni organizzazione ha un sistema informativo, eventualmente non esplicitato nella struttura;
 - quasi sempre, il sistema informativo è di supporto ad altri sottosistemi
 - il sistema informativo è di solito suddiviso in sottosistemi (in modo gerarchico o decentrato), più o meno fortemente integrati

Sistema organizzativo

- Insieme di risorse e regole per lo svolgimento coordinato delle attività al fine del perseguimento degli scopi
 - il sistema informativo è parte del sistema organizzativo
 - il sistema informativo esegue/gestisce processi informativi (cioè i processi che coinvolgono informazioni)

Risorse aziendali

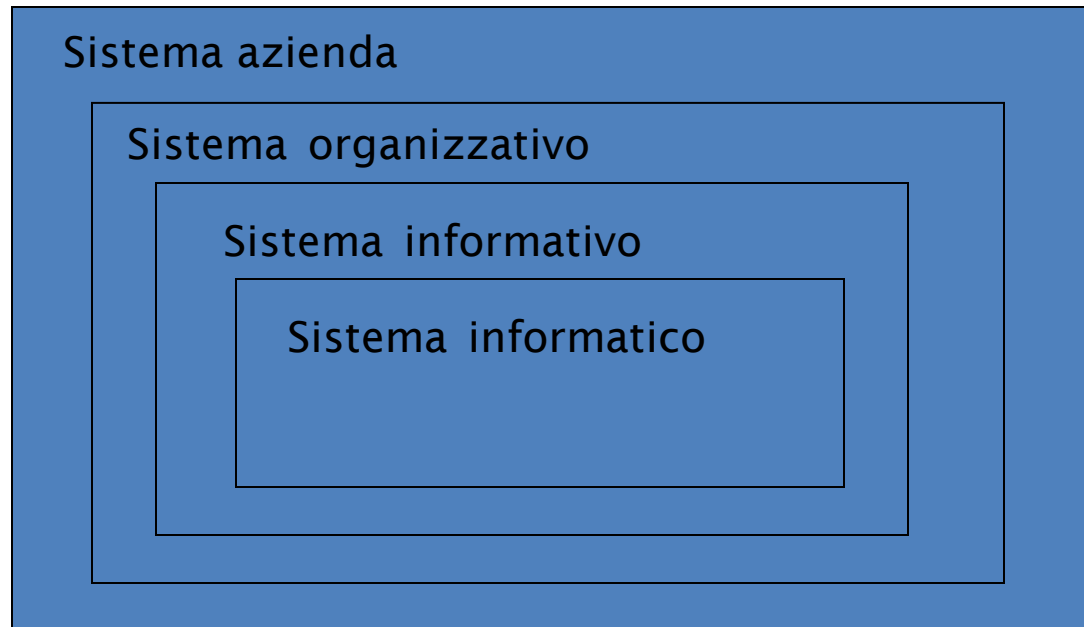
- le risorse di una azienda:
 - persone
 - denaro
 - materiali
 - **informazioni**
- *Oggi viviamo nell'era della conoscenza*
- *Il capitale principale delle organizzazioni è rappresentato dalla conoscenza (basata sui dati)*

Sistemi informativi e automazione

- Il concetto di “sistema informativo” è indipendente da qualsiasi automazione
- esistono organizzazioni la cui ragion d’essere è la gestione di informazioni
 - Es.: servizi anagrafici, tribunali, biblioteche, etc
 - Operano da secoli
 - La gestione delle loro informazioni si è evoluta nel tempo
 - Ma ciò non altera l’obiettivo dell’organizzazione!

Sistema Informatico

- porzione automatizzata del sistema informativo:
la parte del sistema informativo che gestisce informazioni con tecnologia informatica



Sistema informativo e sistema informatico

- Anche prima di essere informatizzati, molti sistemi informativi si sono evoluti verso una razionalizzazione e standardizzazione delle procedure e dell'organizzazione delle informazioni
 - Esempi:
 - Uffici anagrafe
 - Contabilità familiare
 - Ordinazione al ristorante

Gestione delle informazioni

- Nelle attività umane, le informazioni vengono gestite (registrate e scambiate) in forme diverse:
 - idee informali
 - linguaggio naturale (scritto o parlato, formale o colloquiale, in una lingua o in un'altra)
 - disegni, grafici, schemi
 - numeri e codici
- e su vari supporti
 - memoria umana, carta, dispositivi elettronici

Gestione delle informazioni

- Nelle attività standardizzate dei sistemi informativi complessi, sono state introdotte col tempo forme di organizzazione e codifica delle informazioni
- Ad esempio, nei servizi anagrafici si è iniziato con registrazioni discorsive e poi
 - nome e cognome
 - estremi anagrafici
 - codice fiscale

Informazioni e dati

- Nei sistemi informatici (e non solo), le informazioni vengono rappresentate in modo essenziale, spartano: attraverso i **dati**
- 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.
- *Informazione può essere vista come un'interpretazione semantica dei dati*
 - *Dipendente dal contesto*

Dati e informazioni

- I dati hanno bisogno di essere interpretati

Esempio

'Mario' '275' su un foglio di carta sono due **dati**.

Se il foglio di carta viene fornito in risposta alla domanda “A chi mi devo rivolgere per il problema X; qual è il suo interno?”, allora i dati possono essere interpretati per fornire **informazione** e arricchire la **conoscenza**.

Oppure, soldi che Mario mi deve

Oppure, giorni lavorativi di Mario

Etc. Etc.

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)

Base di dati

- 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**

Sistema di gestione di basi di dati

DataBase Management System — DBMS

Sistema (**prodotto software**) in grado di gestire **collezioni di dati** che siano (anche):

- **Grandi**, di dimensioni (molto) maggiori della memoria centrale dei sistemi di calcolo utilizzati
- **Persistenti**, con un periodo di vita indipendente dalle singole esecuzioni dei programmi che le utilizzano
- **Condivise**, utilizzate da applicazioni diverse anche in località geografiche diverse
- **Affidabili**, resistenza a malfunzionamenti hardware e software, blackout, etc.
- **Private**, con una disciplina e un controllo degli accessi
- Come ogni prodotto informatico, un DBMS deve essere
 - **Efficiente**, utilizzando al meglio le risorse di spazio e tempo del sistema
 - **Efficace**, rendendo produttive le attività dei suoi utilizzatori

Alcuni DBMS in commercio

- Business estremamente grosso
- Alcuni sistemi
 - Microsoft Access
 - IBM-DB2
 - Oracle
 - Informix
 - Sybase
 - Microsoft SQLServer
 - Ingres
 - MySql (open-source)

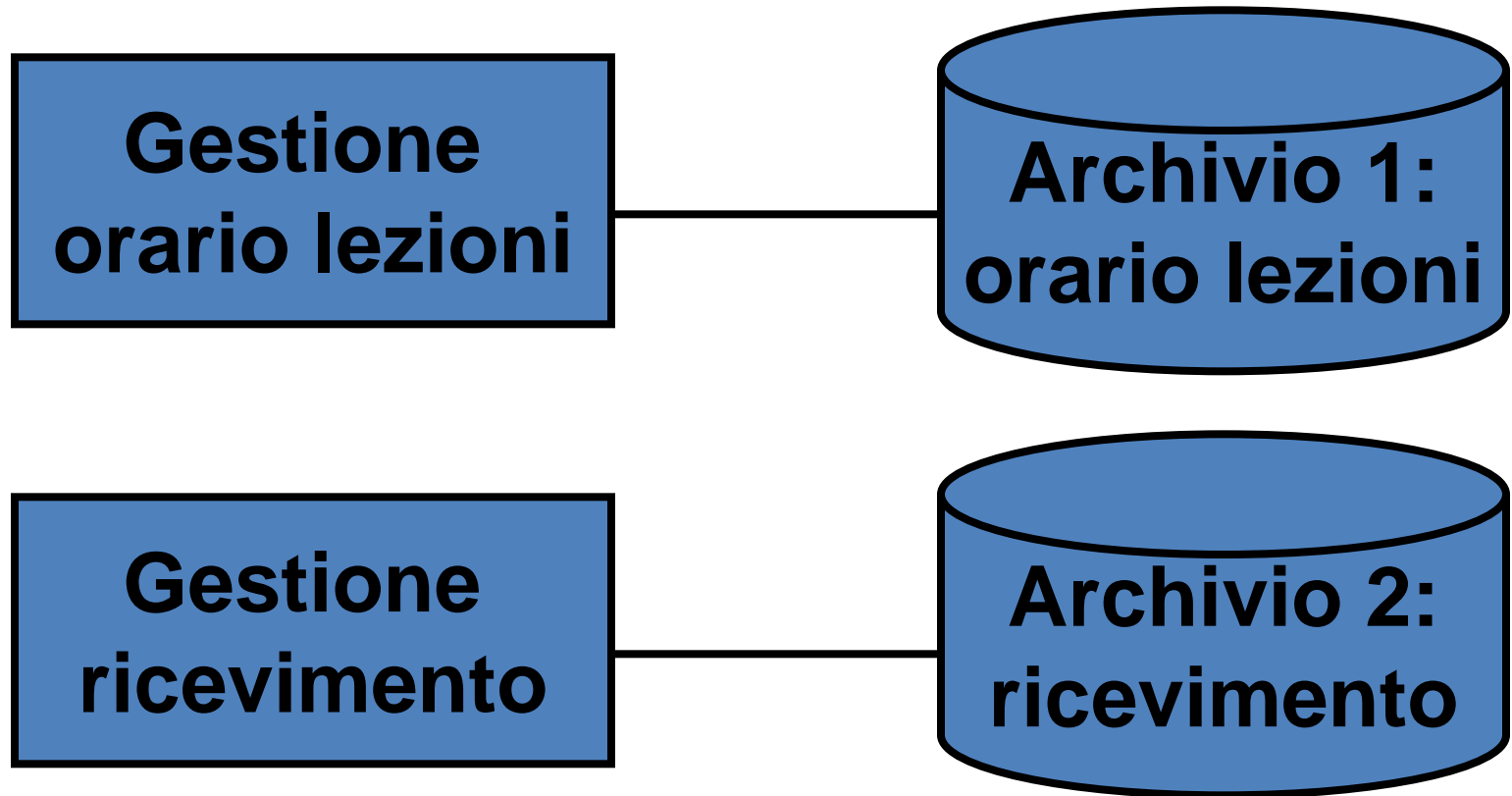
Condivisione dei dati

- Ogni organizzazione (specie se grande) è divisa in settori o comunque svolge diverse attività
- A ciascun settore o attività corrisponde un (sotto)sistema informativo
- Possono esistere sovrapposizioni fra i dati di interesse dei vari settori
- Una base di dati è una risorsa **integrata**, condivisa fra i vari settori

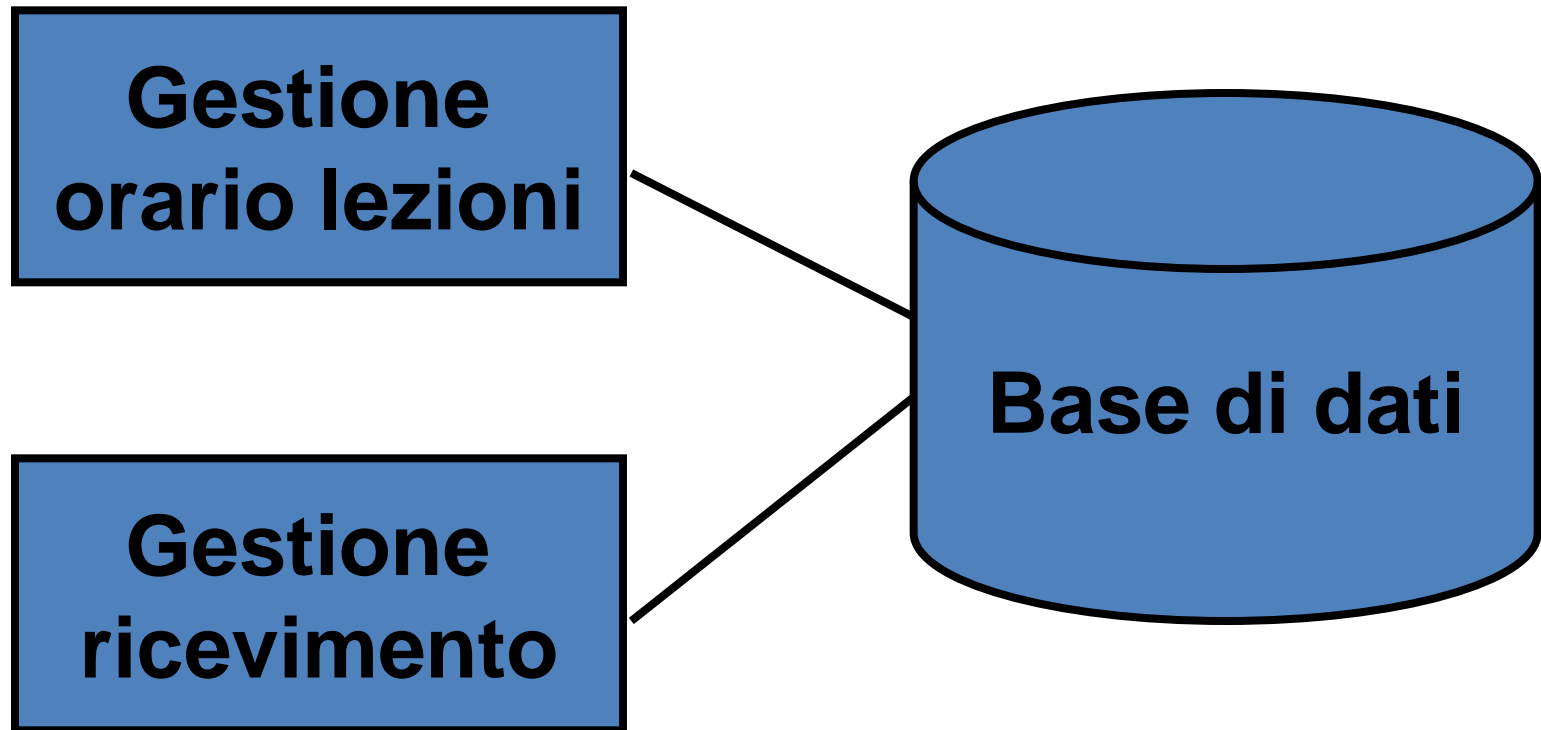
Possibili problemi

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

Archivi e basi di dati



Archivi e basi di dati



Le basi di dati sono condivise

- Una base di dati e' una risorsa **integrata**, **condivisa** fra le varie applicazioni
- conseguenze
 - Attivita' diverse su dati in parte condivisi:
 - meccanismi di **autorizzazione**
 - Attivita' multi-utente su dati condivisi:
 - controllo della **concorrenza**

Efficienza

- Si misura (come in tutti i sistemi informatici) in termini di tempo di esecuzione (tempo di risposta) e spazio di memoria (principale e secondaria).
- L'efficienza è il risultato della qualità del DBMS e delle applicazioni che lo utilizzano.

Descrizioni dei dati nei DBMS

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

Modello dei dati

- insieme di costrutti utilizzati per organizzare i dati di interesse e descriverne la dinamica
- componente fondamentale: **meccanismi di strutturazione** (o **costruttori di tipo**)
- così ogni modello dei dati prevede alcuni costruttori
- ad esempio, il **modello relazionale** prevede il costruttore **relazione**, che permette di definire insiemi di record omogenei

Organizzazione dei dati in una base di dati

Orario

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

Schemi e istanze

Lo **schema** della base di dati

Orario

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

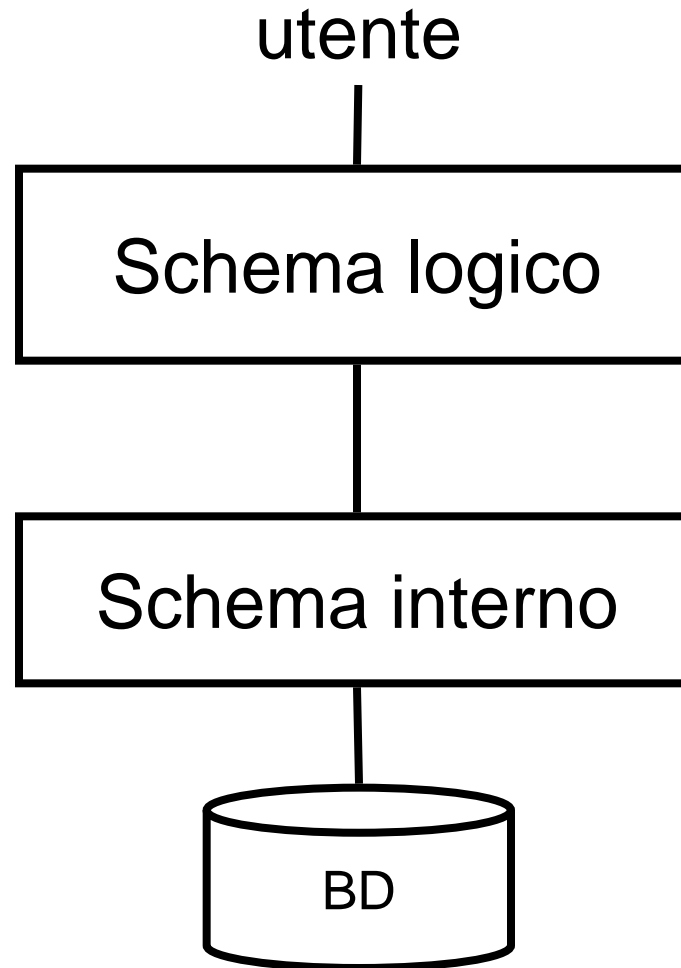
Schemi e istanze

- In ogni base di dati esistono:
 - lo **schema**, sostanzialmente invariante nel tempo, che ne descrive la struttura (aspetto intensionale)
 - nell'esempio, le intestazioni delle tabelle
 - l'**istanza**, i valori attuali, che possono cambiare anche molto rapidamente (aspetto estensionale)
 - nell'esempio, il “corpo” di ciascuna tabella

Due tipi (principali) di modelli

- **modelli logici**: utilizzati nei DBMS esistenti per l'organizzazione dei dati
 - utilizzati dai programmi
 - indipendenti dalle strutture fisicheesempi: **relazionale**, reticolare, gerarchico, a oggetti
- **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 progettazioneil più noto è il modello **Entity-Relationship**

Architettura (semplificata) di un DBMS



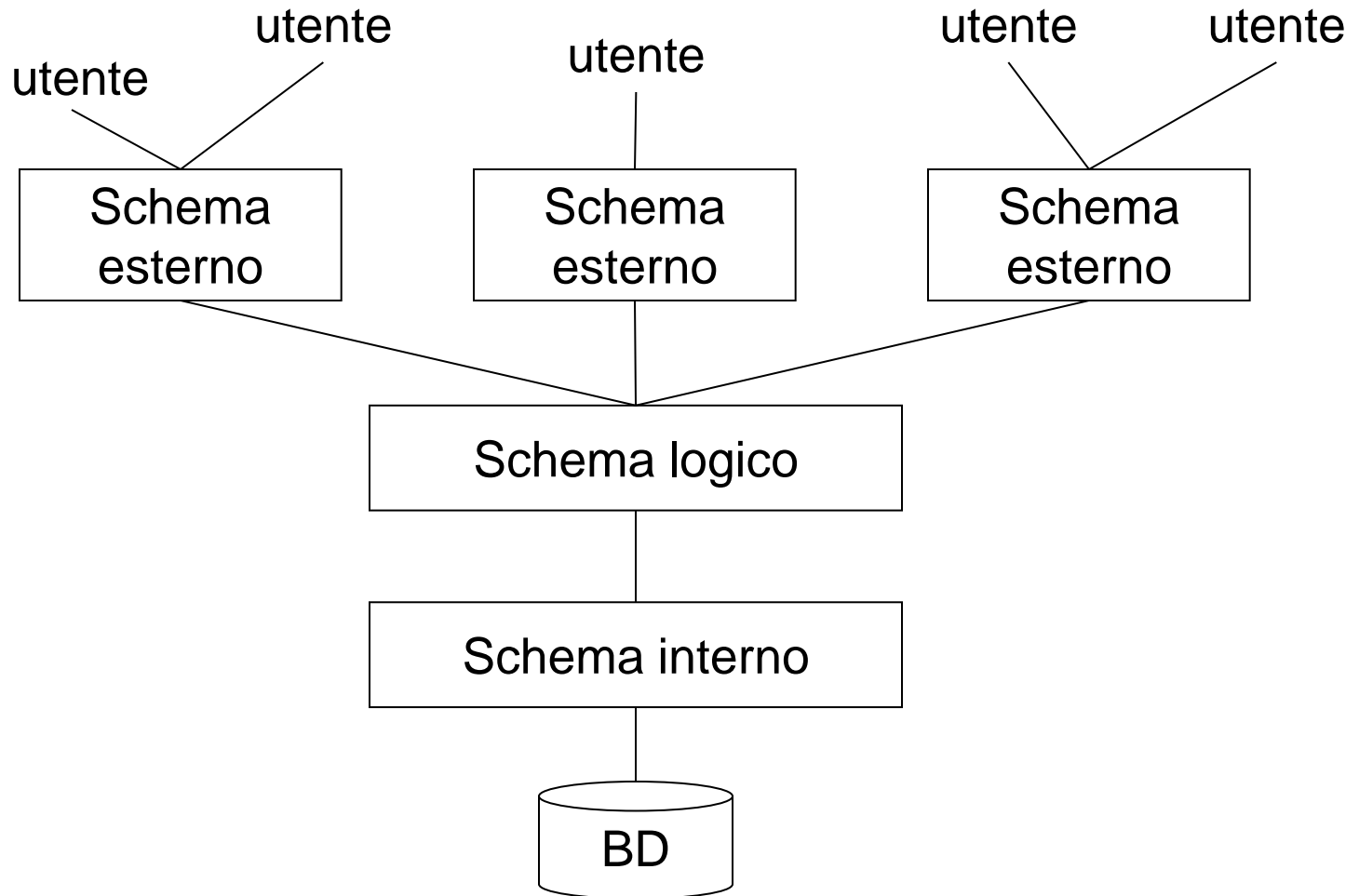
Architettura semplificata di un DBMS: schemi

- **schema logico**: descrizione della base di dati nel modello logico (ad esempio, la struttura della tabella)
- **schema fisico**: rappresentazione dello schema logico per mezzo di strutture memorizzazione (file)

Indipendenza dei dati

- 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)

Architettura standard (ANSI/SPARC) a tre livelli per DBMS



Architettura ANSI/SPARC: schemi

schema fisico/interno: rappresentazione dei dati per mezzo di strutture fisiche di memorizzazione

schema logico: descrizione dell'intera base di dati nel modello logico "principale" del DBMS

schema esterno: descrizione di parte della base di dati del modello logico ("viste" parziali, derivate, anche in modelli diversi)

Una vista

Corsi

Corso	Docente	Aula
Basi di dati	Rossi	DS3
Sistemi	Neri	N3
Reti	Bruni	N3
Controlli	Bruni	G

Aule

Nome	Edificio	Piano
DS1	OMI	Terra
N3	OMI	Terra
G	Pincherle	Primo

CorsiSedi

Corso	Aula	Edificio	Piano
Sistemi	N3	OMI	Terra
Reti	N3	OMI	Terra
Controlli	G	Pincherle	Primo

Indipendenza dei dati

- conseguenza della articolazione in livelli
- l'accesso avviene solo tramite il livello esterno (che può coincidere con il livello logico)
- due forme:
 - indipendenza fisica
 - indipendenza logica

Indipendenza fisica

- il livello logico e quello esterno sono indipendenti da quello fisico
 - una relazione 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

Linguaggi per basi di dati

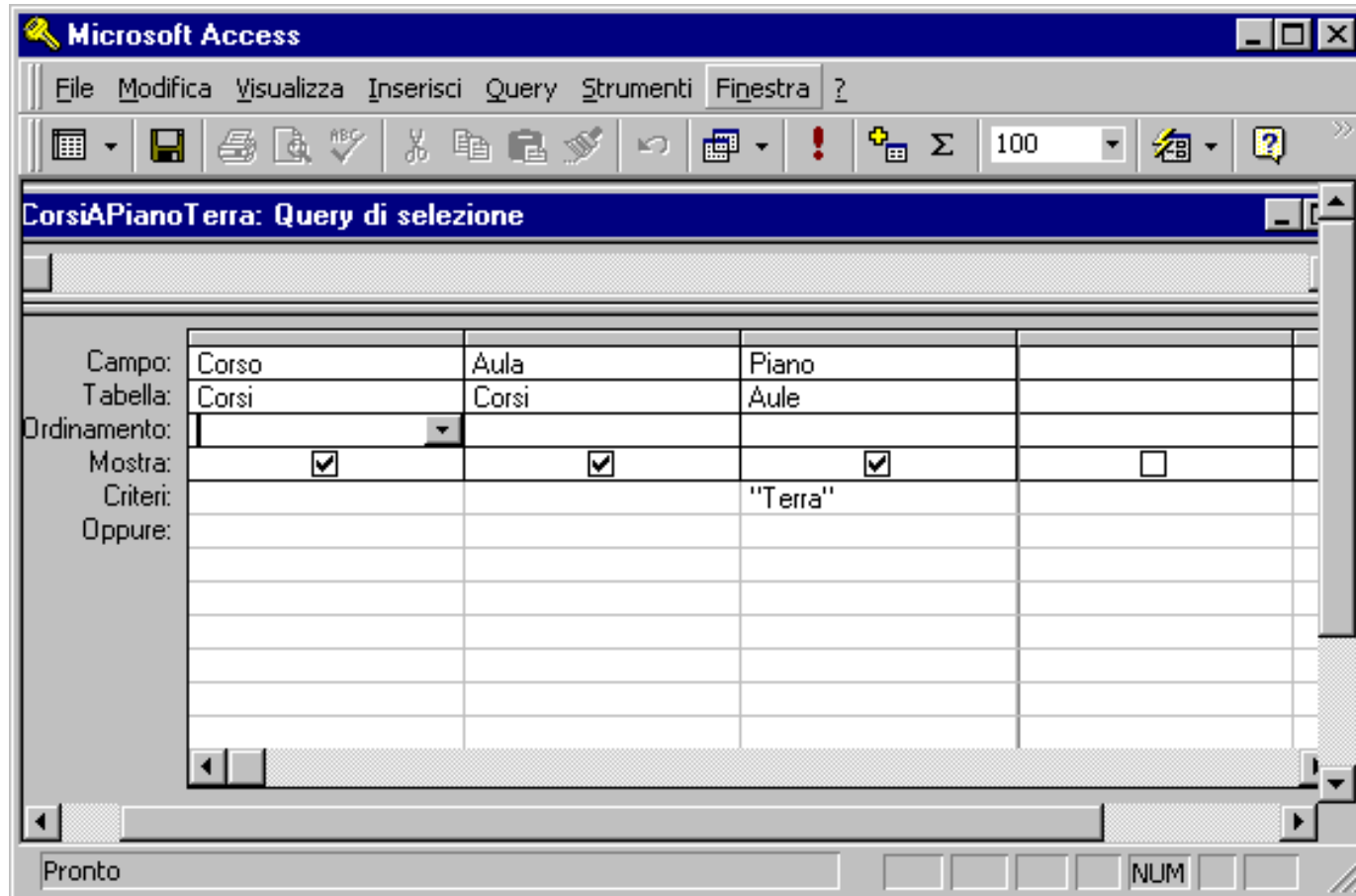
- Un altro contributo all'efficacia: disponibilità di vari linguaggi e interfacce diverse
 - ⇒ linguaggi testuali interattivi (**SQL**)
 - ⇒ comandi (come quelli del linguaggio interattivo) immersi in un linguaggio **ospite** (Pascal, C, Cobol, etc.)
 - ⇒ comandi (come quelli del linguaggio interattivo) immersi in un linguaggio ad hoc, con anche altre funzionalità (p.es. per grafici o stampe strutturate), anche con l'ausilio di strumenti di sviluppo (p. es. per la gestione di maschere)
 - ⇒ con interfacce amichevoli (senza linguaggio testuale)

SQL, un linguaggio interattivo

```
SELECT Corso, Aula, Piano  
FROM Aule, Corsi  
WHERE Nome = Aula  
AND Piano="Terra"
```

Corso	Aula	Piano
Sistemi	N3	Terra
Reti	N3	Terra

Interazione non testuale (in [Access](#))



Una distinzione terminologica (separazione fra dati e programmi)

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

Un'operazione DDL (sullo schema)

```
CREATE TABLE orario (  
  insegnamento      CHAR(20) ,  
  docente            CHAR(20) ,  
  aula                CHAR(4) ,  
  ora                 CHAR(5) )
```

Personaggi e interpreti

- **progettisti** e realizzatori di **DBMS**
- **progettisti della base di dati** e amministratori della base di dati ([DBA](#))
- **progettisti** e programmatori **di applicazioni**
- **utenti**
 - utenti **finali** (terminalisti): eseguono applicazioni predefinite ([transazioni](#))
 - utenti **casuali**: eseguono operazioni non previste a priori, usando linguaggi interattivi

Database administrator (DBA)

- Persona o gruppo di persone responsabile del controllo centralizzato e della gestione del sistema, delle prestazioni, dell'affidabilità, delle autorizzazioni
- Le funzioni del DBA includono quelle di progettazione, anche se in progetti complessi ci possono essere distinzioni

Transazioni

- Programmi che realizzano attività frequenti e predefinite, con poche eccezioni, previste a priori.
- Esempi:
 - versamento presso uno sportello bancario
 - emissione di certificato anagrafico
 - dichiarazione presso l'ufficio di stato civile
 - prenotazione aerea
- Le transazioni sono di solito realizzate con programmi in linguaggio ospite (tradizionale o ad hoc)
- **N. B.:** il termine **transazione** ha un'altra accezione, più specifica: sequenza indivisibile di operazioni (o vengono eseguite tutte o nessuna).

Vantaggi e svantaggi dei DBMS

Pro

- 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)

Contro

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

Esercitazioni

- Indicare quali delle seguenti affermazioni sono vere:
 - l'indipendenza dei dati permette di scrivere programmi senza conoscere le strutture fisiche dei dati
 - l'indipendenza dei dati permette di modificare le strutture fisiche dei dati senza dover modificare i programmi che accedono alla base di dati
 - l'indipendenza dei dati permette di scrivere programmi conoscendo solo lo schema concettuale della BD
 - l'indipendenza dei dati permette di formulare interrogazioni senza conoscere le strutture fisiche

Esercitazioni

- Indicare quali delle seguenti affermazioni sono vere:
 - il fatto che le basi di dati siano condivise favorisce l'efficienza dei programmi che le utilizzano
 - il fatto che le basi di dati siano condivise permette di ridurre ridondanze e inconsistenze
 - il fatto che le basi di dati siano persistenti ne garantisce l'affidabilità
 - il fatto che le basi di dati siano persistenti favorisce l'efficienza dei programmi
 - il fatto che le basi di dati siano condivise rende necessaria la gestione della privacy e delle autorizzazioni

Esercitazioni

- Indicare quali delle seguenti affermazioni sono vere:
 - la distinzione fra DDL e DML corrisponde alla distinzione fra schema e istanza
 - le istruzioni DML permettono di interrogare la base di dati ma non di modificarla
 - le istruzioni DDL permettono di specificare la struttura della base di dati ma non di modificarla
 - non esistono linguaggi che includono sia istruzioni DDL sia istruzioni DML
 - SQL include istruzioni DML e DDL
 - le istruzioni DML permettono di interrogare la base di dati e di modificarla

Esercitazioni

- Indicare quali delle seguenti affermazioni sono vere:
 - gli utenti casuali utilizzano transazioni predefinite
 - i terminalisti utilizzano transazioni predefinite
 - gli utenti casuali progettano la base di dati
 - i progettisti del DBMS realizzano le transazioni che saranno utilizzate dai terminalisti
 - i progettisti della base di dati realizzano il DBMS
 - i progettisti delle applicazioni utilizzano la base di dati come progettata dal progettista del DBMS
 - i progettisti delle applicazioni utilizzano la BD come progettata dal progettista della BD