

Basi di dati

INTRODUZIONE

Base di dati

- Insieme organizzato di dati utilizzati per il supporto allo svolgimento di attività (di un ente, azienda, ufficio, persona)

Punti di vista

- Metodologico
- Tecnologico

Contenuti

- modelli per l'organizzazione dei dati
- linguaggi per l'utilizzo dei dati
- sistemi per la gestione dei dati
- metodologie di progettazione di basi di dati

Il corso

Metodo di studio

- studio individuale, con riflessione sui concetti e riferimento alle esperienze personali
- svolgimento di esercizi
- sviluppo di progetti o almeno esercitazioni pratiche, con realizzazione con opportuno strumento (DB2, SQLServer, Oracle, PostgreSQL, MySQL, MS Access, ...)

Sistema informativo

- Componente di una organizzazione che gestisce le informazioni di interesse (cioè utilizzate per il perseguimento degli scopi dell'organizzazione)
- 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

Gestione delle informazioni

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

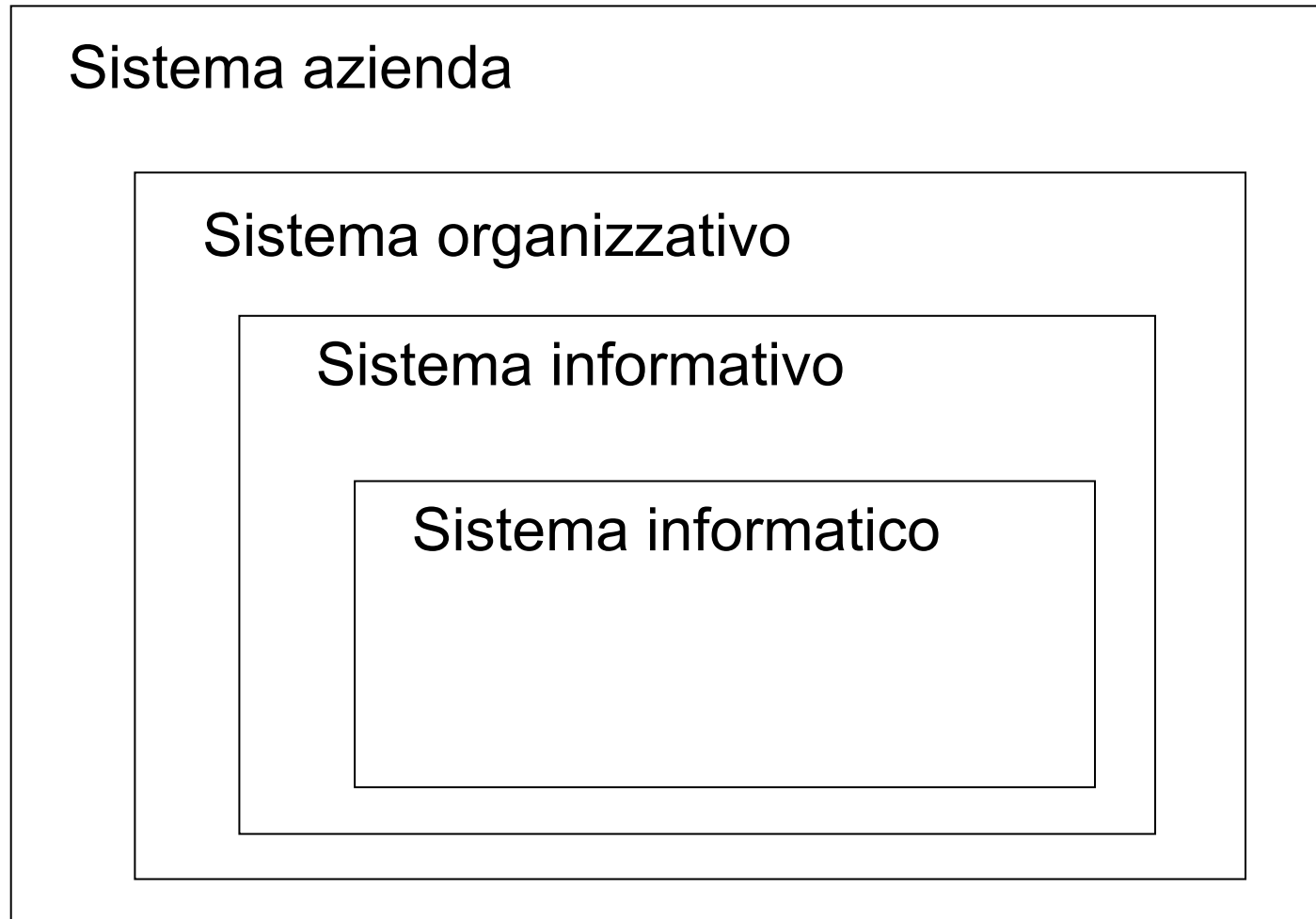
Sistemi informativi e automazione

- 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

Sistema Informatico

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

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
- e su vari supporti
 - mente umana, carta, dispositivi elettronici

Informazioni e dati

- Nei sistemi informatici (e non solo), le **informazioni** vengono rappresentate in modo essenziale, spartano: attraverso i **dati**

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



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

Base di dati

(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 DBMS

Sistema di gestione di basi di dati

DataBase Management System (DBMS)

- Sistema che gestisce collezioni di dati:
 - grandi
 - persistenti
 - condivisegarantendo
 - privatezza
 - affidabilità
 - efficienza
 - efficacia

DBMS

- Prodotti e servizi software (complessi) disponibili sul mercato; esempi:
 - DB2
 - Oracle
 - SQLServer
 - MySQL
 - PostgreSQL
 - Access
 - BigQuery

Le basi di dati sono ... 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)
 - 100 miliardi di record

Le basi di dati sono ... persistenti

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

Le basi di dati sono ... condivise

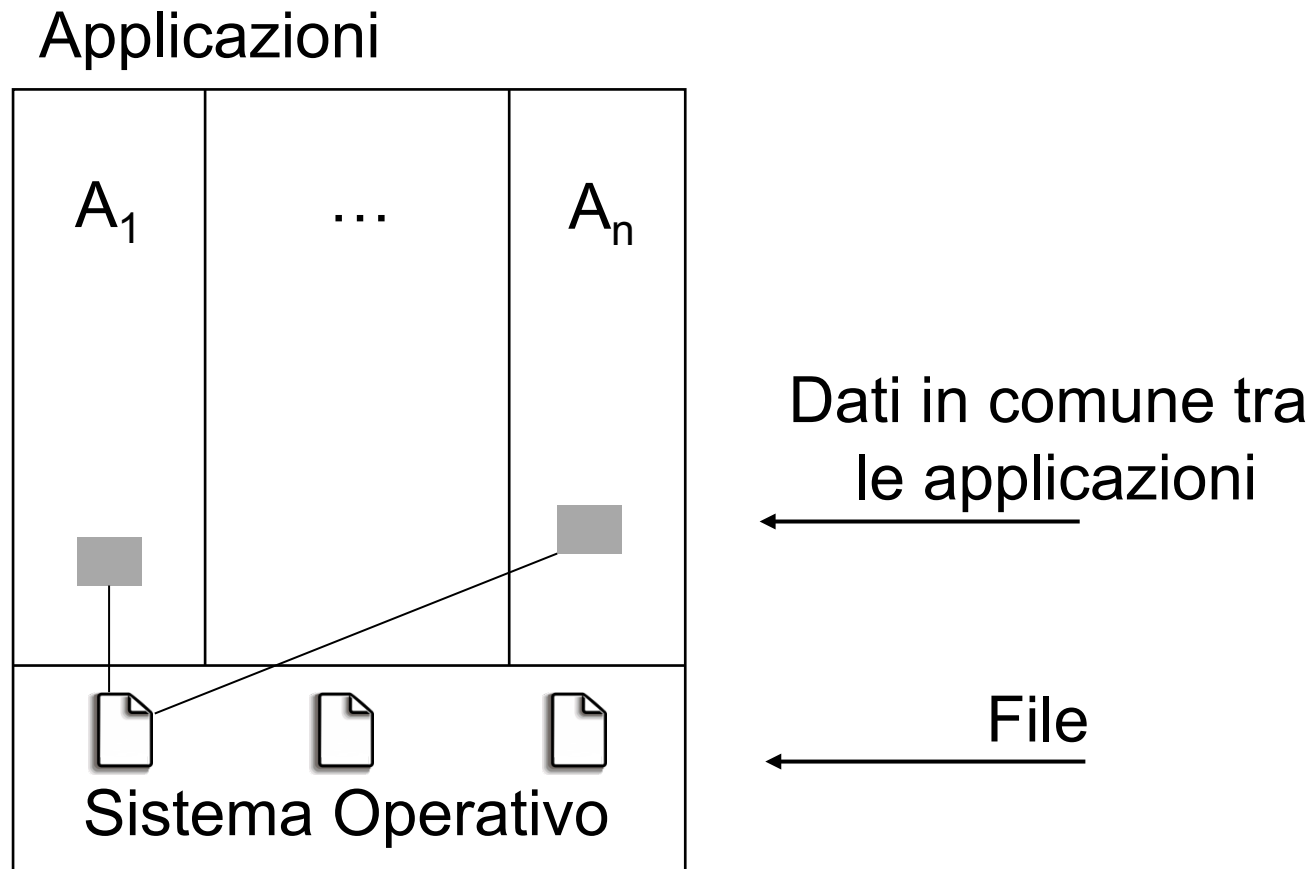
- Ogni organizzazione (specie se grande) è divisa in settori o comunque svolge diverse attività
- Ciascun settore/attività ha un (sotto)sistema informativo (non necessariamente disgiunto)

Problemi

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

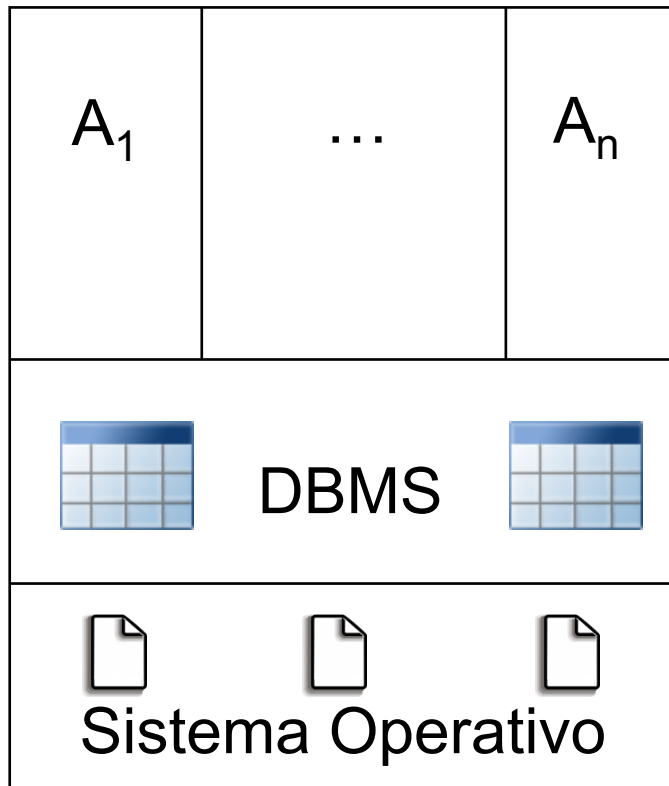
Vediamo ora l'evoluzione dei Database Management System

Anni 70: senza DBMS



Anni 80: i DBMS vengono introdotti

Applicazioni



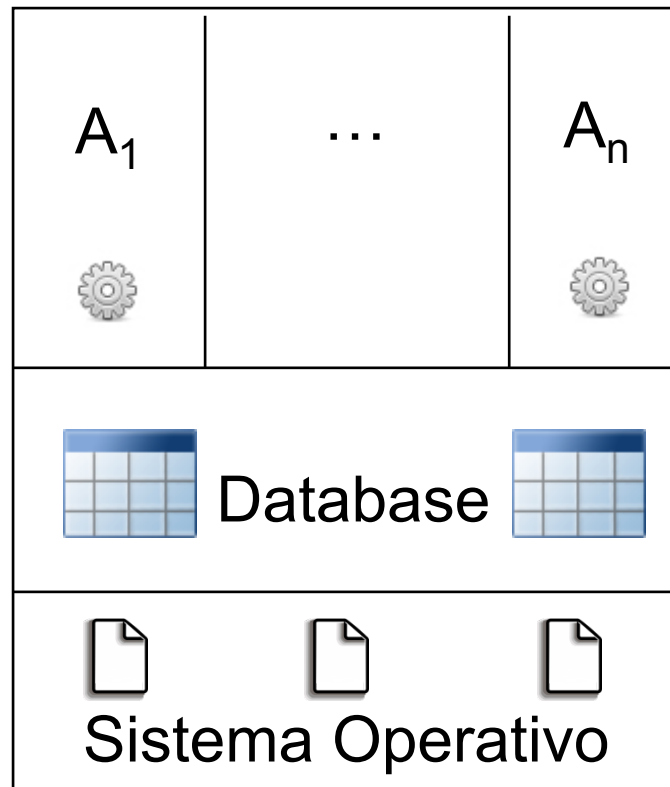
← Scritte in C, Pascal ...

← Tabelle

← File

Anni 90: comportamento procedurale condiviso nelle applicazioni

Applicazioni



Comportamento
Procedurale
in comune
←

← Tabelle

← File

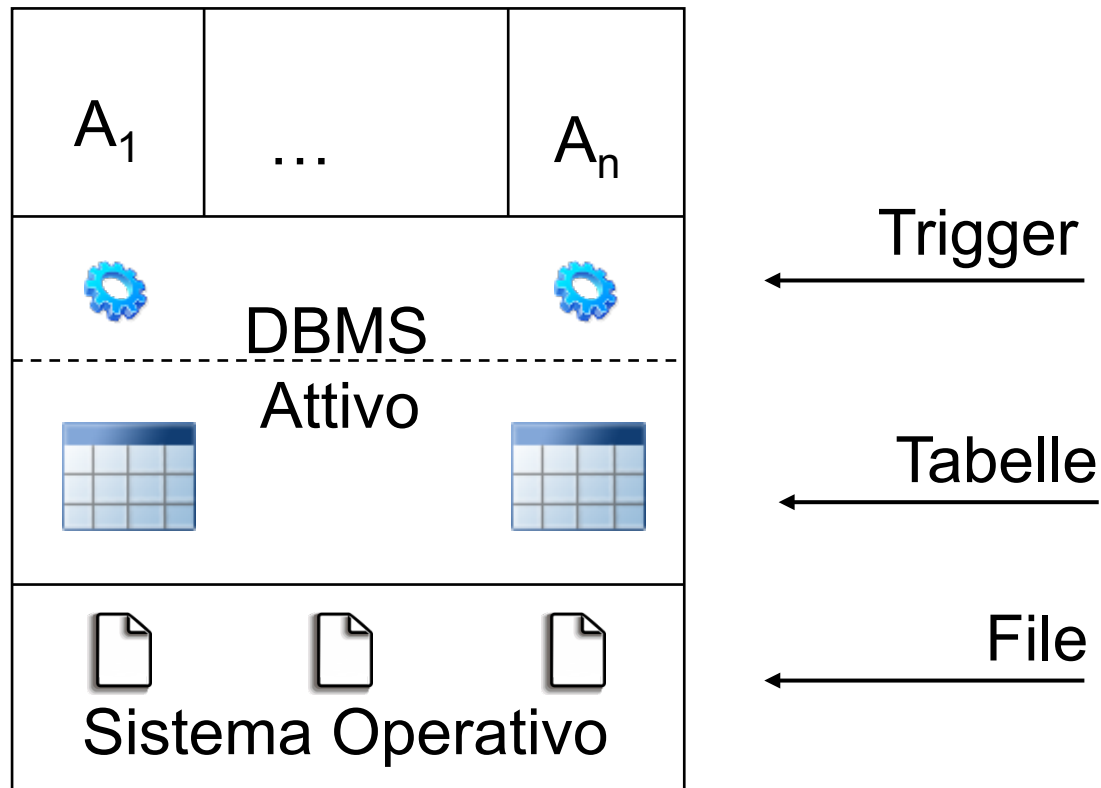
Evoluzioni dei DBMS:

Store procedure

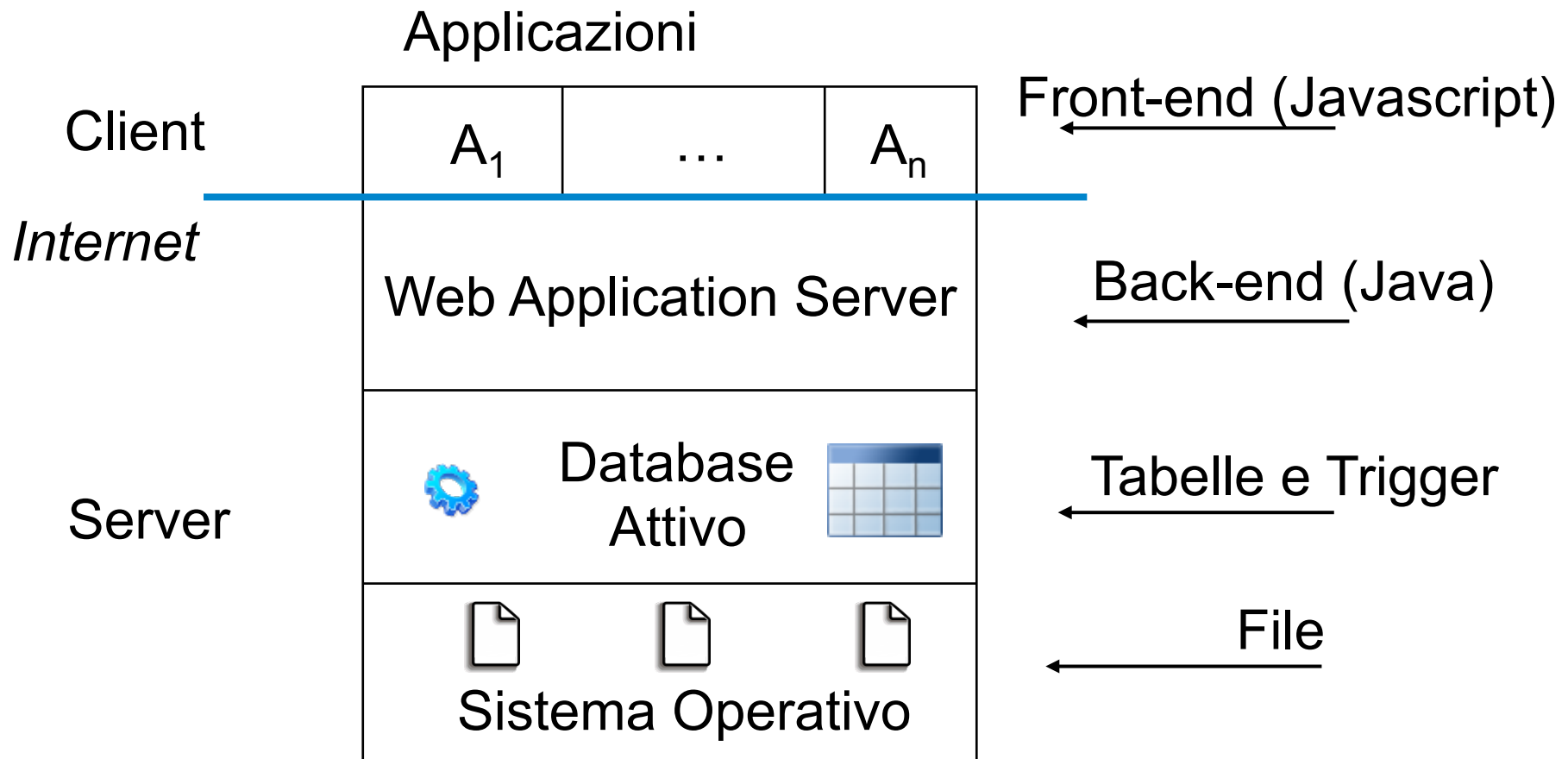
- Le stored procedure sono state introdotte per esprimere il comportamento procedurale messo a fattor comune tra le varie applicazioni utilizzando il DBMS
- Le stored procedure non aderiscono a nessuno standard ed hanno il problema dell' impedance mismatch (che vedremo) con il linguaggio usato per definirle
- Il risultato è l' introduzione dei trigger (o regole) per modellare il comportamento procedurale messo a fattor comune tra le varie applicazioni e sono gestite dal DBMS

Anni 90: DBMS Attivo

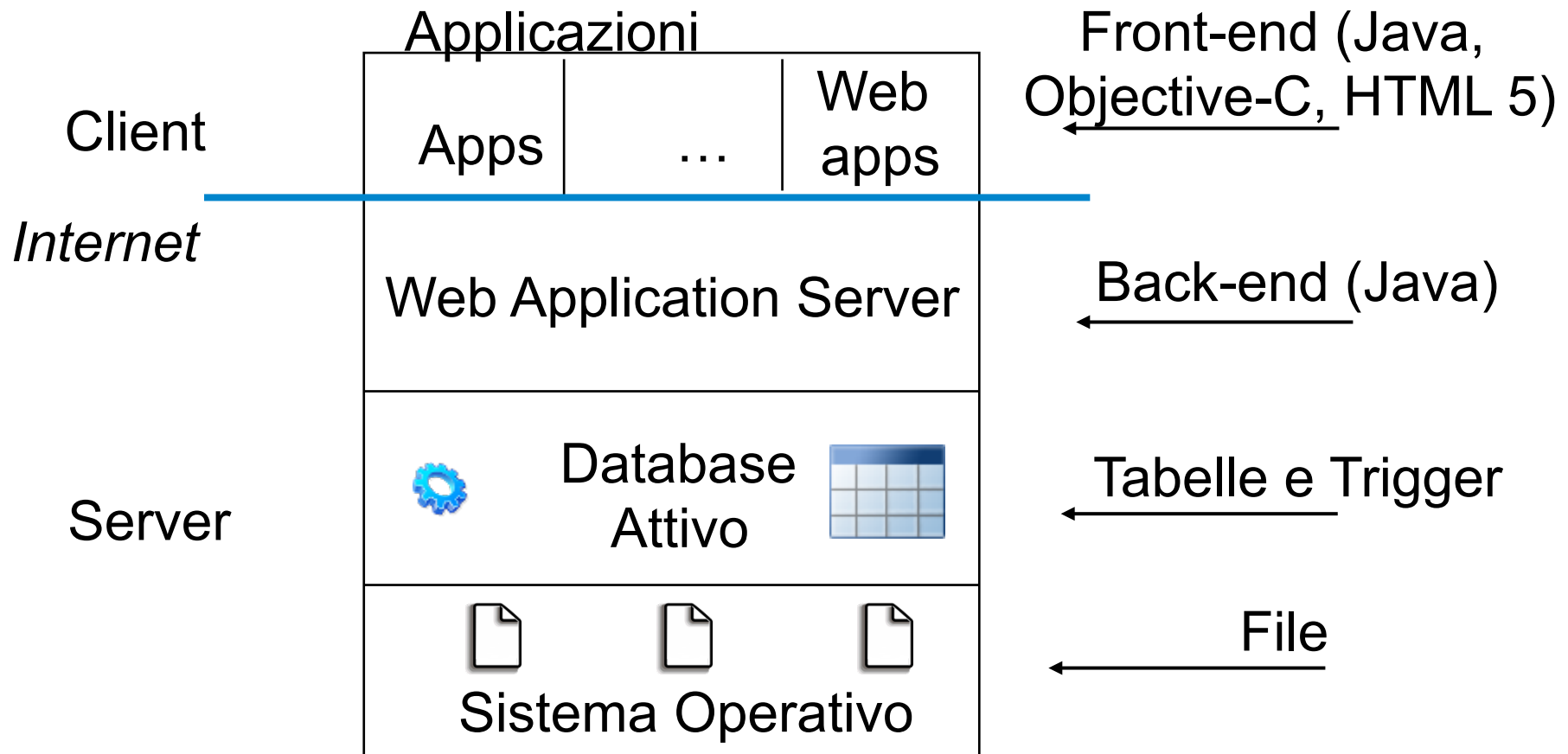
Applicazioni



Evoluzione anni 00



Evoluzione anni 10: in mobilità



Le basi di dati sono condivise

- Una base di dati e' una risorsa **integrata, condivisa** fra applicazioni
- conseguenze
 - Attivita' diverse su dati condivisi:
 - meccanismi di **autorizzazione**
 - Accessi di più utenti ai dati condivisi:
 - controllo della **concorrenza**

I DBMS garantiscono ... privatezza

- 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 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 rischiedono 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 (in inglese **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 corso è in buona parte dedicato ad illustrare come i DBMS perseguono l'efficacia

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 (il **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, 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**)
- Come nei linguaggi di programmazione esistono meccanismi che permettono di definire nuovi tipi, così ogni modello dei dati prevede alcuni costruttori
- Esempio: il **modello relazionale** prevede il costruttore **relazione**, che permette di definire insiemi di record omogenei

CCS Ingegneria Informatica Orario - Netscape

File Edit View Go Communicator Help

UNIVERSITA' DEGLI STUDI DI CHISSADOVE

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

Document: Done

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

Basi di dati: schema e istanza

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

Schema e istanza

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

Due tipi (principali) di modelli

- modelli logici
- modelli concettuali

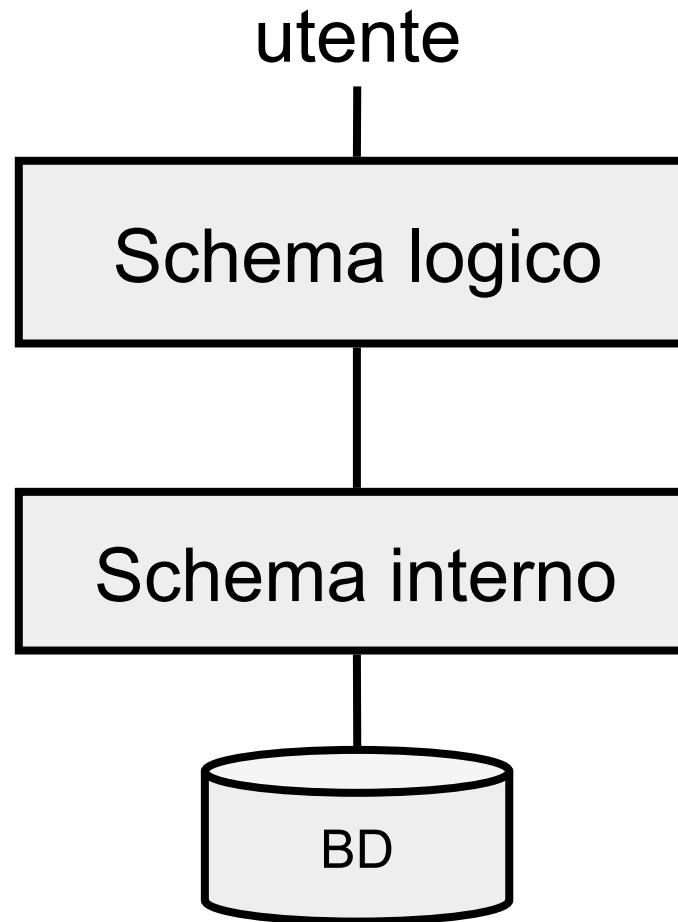
Modelli logici

- 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**

Architettura (semplificata) di un DBMS



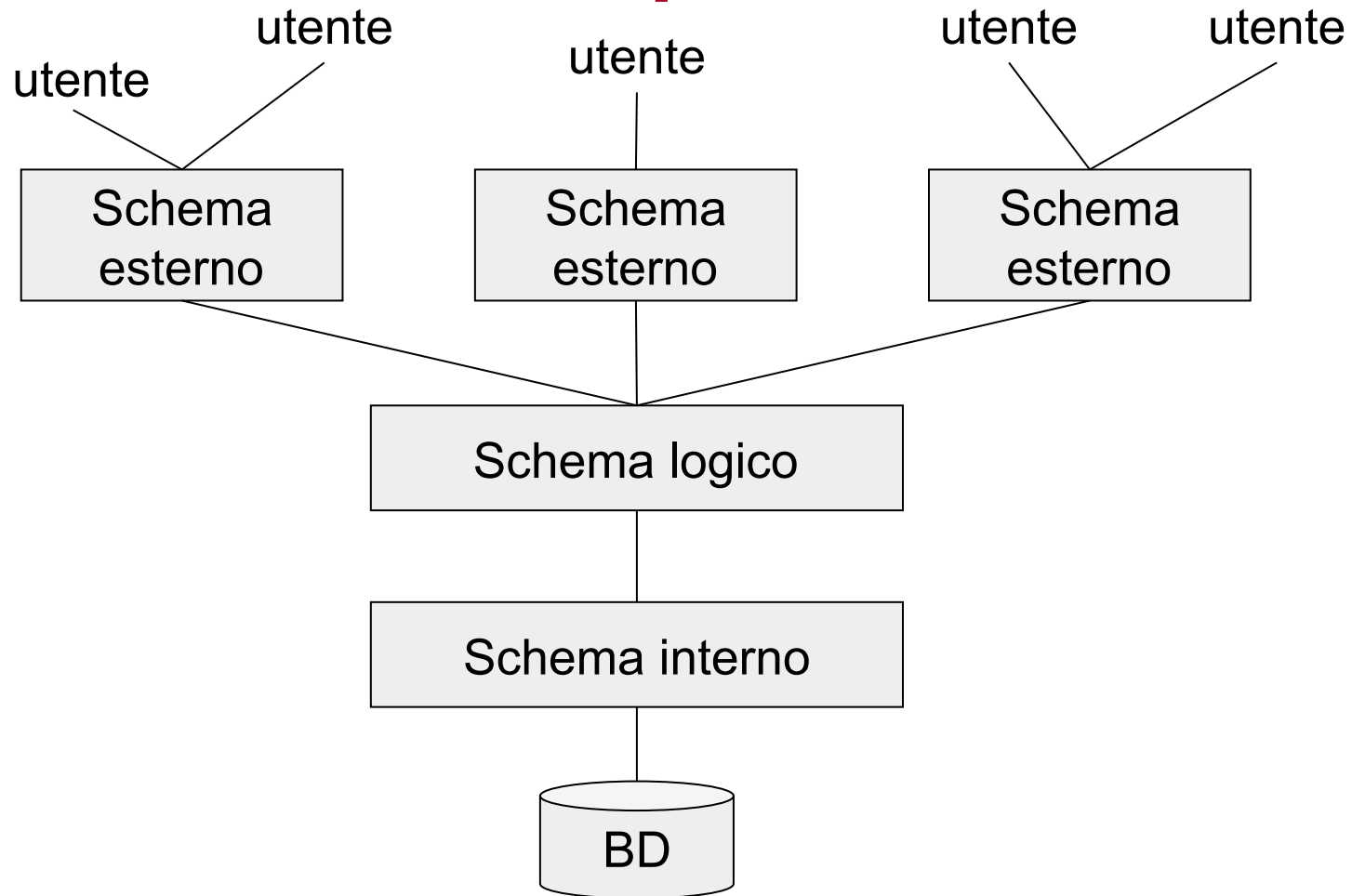
Architettura semplificata di un DBMS: schemi

- **schema logico**: descrizione della base di dati nel modello logico (ad esempio, la struttura della tabella)
- **schema interno** (o **fisico**): rappresentazione dello schema logico per mezzo di strutture memorizzazione (file; ad esempio, record con puntatori, ordinati in un certo modo)

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)
- Perciò in questo corso vedremo solo il livello logico e non quello fisico

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



Architettura ANSI/SPARC: schemi

- **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)

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
 - linguaggi testuali interattivi (**SQL**)
 - comandi (SQL) immersi in un linguaggio **ospite** (Pascal, Java, C ...)
 - comandi (SQL) immersi in un linguaggio ad hoc, con anche altre funzionalità (p.es. per grafici o stampe strutturate)
 - con interfacce amichevoli (senza linguaggio testuale)

SQL, un linguaggio interattivo

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

- "Trovare i corsi tenuti in aule a piano terra"

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

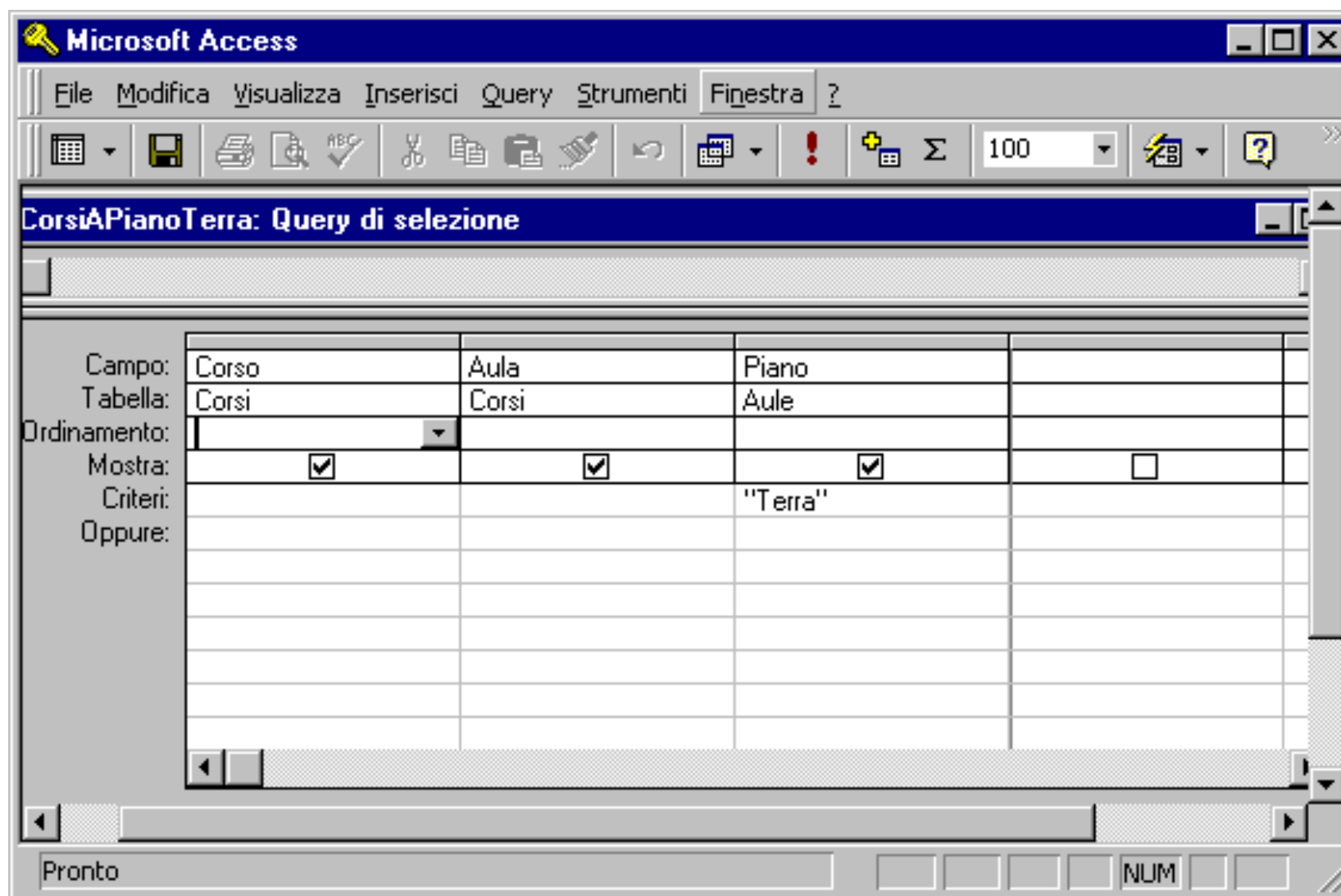
SQL immerso in linguaggio ospite

```
write('nome della citta"?'); readln(citta);
EXEC SQL DECLARE P CURSOR FOR
    SELECT NOME, REDDITO
    FROM PERSONE
    WHERE CITTA = :citta ;
EXEC SQL OPEN P ;
EXEC SQL FETCH P INTO :nome, :reddito ;
while SQLCODE = 0 do begin
    write('nome della persona:', nome, 'aumento?');
    readln(aumento);
    EXEC SQL UPDATE PERSONE
        SET REDDITO = REDDITO + :aumento
        WHERE CURRENT OF P
    EXEC SQL FETCH P INTO :nome, :reddito
end;
EXEC SQL CLOSE CURSOR P
```


SQL in linguaggio ad hoc (Oracle PL/SQL)

```
declare Stip number;  
begin  
    SELECT STIPENDIO INTO STIP FROM IMPIEGATO  
    WHERE MATRICOLA = '575488' FOR UPDATE OF STIPENDIO;  
    if Stip > 30 then  
        UPDATE IMPIEGATO SET STIPENDIO = STIPENDIO * 1.1  
        WHERE MATRICOLA = '575488';  
    else  
        UPDATE IMPIEGATO SET STIPENDIO = STIPENDIO * 1.15  
        WHERE MATRICOLA = '575488';  
    end if;  
    commit;  
exception  
    when no_data_found then  
        INSERT INTO ERRORI  
        VALUES('MATRICOLA INESISTENTE',SYSDATE);  
end;
```

Interazione non testuale (Access)



Una distinzione (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 (per l'utente)

- 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 in linguaggio ospite (tradizionale o ad hoc)

Transazioni, due accezioni

- Per l'utente:
 - programma a disposizione, da eseguire per realizzare una funzione di interesse
- Per il sistema:
 - sequenza indivisibile di operazioni (cfr. affidabilità)

Vantaggi 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)

Svantaggi dei DBMS

Contro

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