



# Modulo 2 Data Base

Università degli Studi di Salerno

Corso di Laurea in Scienze della comunicazione

**Informatica generale**

Docente: Angela Peduto

A.A. 2004/2005



# Testi di riferimento

✦ **Atzeni, Ceri, Paraboschi, Torlone**

*"Basi di Dati II edizione"*

Mc Graw Hill – 1999

**Altri testi di consultazione e approfondimento sui  
DBMS:**

✦ **Ullman, J. D. – Widom J.**

*"Introduzione ai Database"*

Jackson Libri Università, 2000



# Struttura dell'Argomento



## Concetti generali

- Concetti generali. Sistemi informativi, informazioni e dati. Basi di dati e DBMS. Modelli dei dati, schemi e istanze, livelli di astrazione, indipendenza dei dati. Linguaggi e utenti.
- Atzeni, cap. 1 + Curtin, cap 8.4,8.6-8.8



## Il modello dei dati relazionale

- Modelli dei dati. Relazioni e tabelle. Relazioni e Basi di Dati. Informazione incompleta e valori nulli. Vincoli d'integrità.
- Atzeni, cap. 2



## SQL (modulo 3)

- Il linguaggio SQL per creare e usare basi di dati. Interrogazioni semplici, giunzioni, quantificazioni esistenziali ed universali, raggruppamento
- Lucidi dalle lezioni

# 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 (Accad. di Francia)*

Due anime:

- metodologica
- tecnologica



# Sistema informativo

- ✦ Componente (sottosistema) di una organizzazione che gestisce (acquisisce, elabora, conserva, produce) le informazioni di interesse (cioè utilizzate per il perseguimento degli scopi dell'organizzazione )



# Sistema informativo, commenti

- ✦ Ogni organizzazione ha un sistema informativo, eventualmente non esplicitato, nella struttura
- ✦ Quasi sempre, il sistema informativo è di supporto ad altri sottosistemi, e va quindi studiato nel contesto in cui è inserito
- ✦ Il sistema informativo è di solito suddiviso in sottosistemi (in modo gerarchico o decentrato), più o meno fortemente integrati



# Sistema organizzativo e sistema informativo

- ✦ Insieme di risorse e regole per lo svolgimento coordinato delle attività (processi) 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)



# 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

**Sistema 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
- ☀ e su vari supporti
  - mente 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**



# 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

✦ Un esempio:

Mario 275

su un foglio di carta sono due **dati** e non significano molto

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



# 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

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

- ✦ Punti di vista:
- Metodologico
  - Tecnologico



# 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



# DBMS

Un **Database Management System (DBMS)** è un sistema software che si interpone fra le applicazioni e la memoria di massa in cui si trovano collezioni di dati, per consentirne la gestione, in modo indipendente dalle applicazioni stesse.

I dati non appartengono ad una specifica applicazione, ma le diverse applicazioni vi accedono attraverso il DBMS.



# DBMS

Normalmente le applicazioni accedono a dati **locali** gestendoli attraverso file, che sono proprietà delle applicazioni stesse.

La finalità dei DBMS è di estendere le funzionalità del file system, offrendo:

- nuove modalità di accesso ai dati
- condivisione dei dati
- gestione più sofisticata dei file



# DBMS

Le basi di dati gestite dai DBMS sono collezioni di dati:

- **Grandi**  
possono avere notevoli dimensioni (fino a migliaia di Gbyte) e devono quindi necessariamente risiedere nella memoria secondaria
- **Condivise**  
applicazioni ed utenti diversi devono potere accedere ai dati
- **Persistenti**  
Il tempo di vita dei dati va oltre la durata dell'esecuzione delle singole applicazioni



# DBMS

Un DBMS deve garantire:

- **Affidabilità**
- **Privatezza dei dati**
- **Efficienza**
- **Efficacia**



# Affidabilità

Un DBMS deve garantire di poter mantenere intatto il suo contenuto, anche in caso di malfunzionamento.

L'integrità dei dati è affidata a procedure di **backup** (salvataggio) e **recovery** (recupero) dei dati, o alla loro **duplicazione** nei casi più critici.



# Privatezza dei dati

Ogni utente, abilitato a utilizzare la base di dati attraverso una procedura di riconoscimento, può accedere ad insiemi limitati di dati e compiere solo certe operazioni su di essi.





# Efficienza

Un DBMS deve operare e fornire risposte agli utenti in tempi accettabili, utilizzando una quantità il più possibile limitata di risorse.

L'efficienza di un DBMS dipende essenzialmente dalle tecniche utilizzate per la sua implementazione e dalla buona progettazione della base di dati.

Si misura (come in tutti i sistemi informatici) in termini di tempo di esecuzione (tempo di risposta) e spazio di memoria (principale e secondaria) occupato.



# Efficacia

Capacità di un DBMS di rendere produttive le attività degli utenti, cioè di consentire la realizzazione di basi di dati che risolvano in modo efficace i problemi degli utenti.

Concetto generico, qualitativo e non legato a specifiche funzionalità del DBMS. Non esistono criteri oggettivi per valutarla.



# Modelli dei dati

I DBMS possono essere classificati in base al *modello di dati* che utilizzano.

Un modello di dati è costituito dai concetti sulla base dei quali i dati sono strutturati e codificati.

✦ Un **Modello di Dati** è un **insieme di concetti** utilizzati per organizzare i dati e descriverne la struttura in modo che essa risulti comprensibile ad un elaboratore



# Modello di dati

- ✦ Un **Modello di Dati** è un **insieme di concetti** utilizzati per organizzare i dati e descriverne la struttura in modo che essa risulti comprensibile ad un elaboratore
- ✦ (definizione alternativa) Un **Modello di dati** è un insieme di **notazioni** e di **operazioni** atti a rappresentare in modo semplificato un frammento di realtà in termini
  - delle strutture in cui i dati sono raggruppabili
  - delle proprietà dei dati e delle loro aggregazioni
  - delle correlazioni anche complesse esistenti fra gli stessi dati



# Modello di dati

- ✦ Un Modello di Dati permette di
  - **Organizzare i dati di interesse e descriverne la struttura** in maniera comprensibile per un elaboratore
  - **rappresentare la realtà esterna all'interno di un computer** in termini di proprietà elementari, ma rappresentative, di tale realtà
  - **descrivere all'utente umano le informazioni presenti in un DataBase** (e quindi all'interno di un Computer), in quanto gli si presentano
    - non come sequenza di bits (così come sono realmente memorizzate all'interno del computer)
    - ma in termini a lui più comprensibili

# Tipi di Modelli

## **Modello Concettuale**

- tende a **descrivere i concetti** del mondo reale e le loro relazioni piuttosto che i dati
- Modello vicino alla **mentalità dell'uomo**
- Fornisce una **visione astratta dei dati**
- Tra i Modelli Concettuali il più diffuso è il
  - **MODELLO ENTITY-RELATIONSHIPS**

## **Modello Logico**

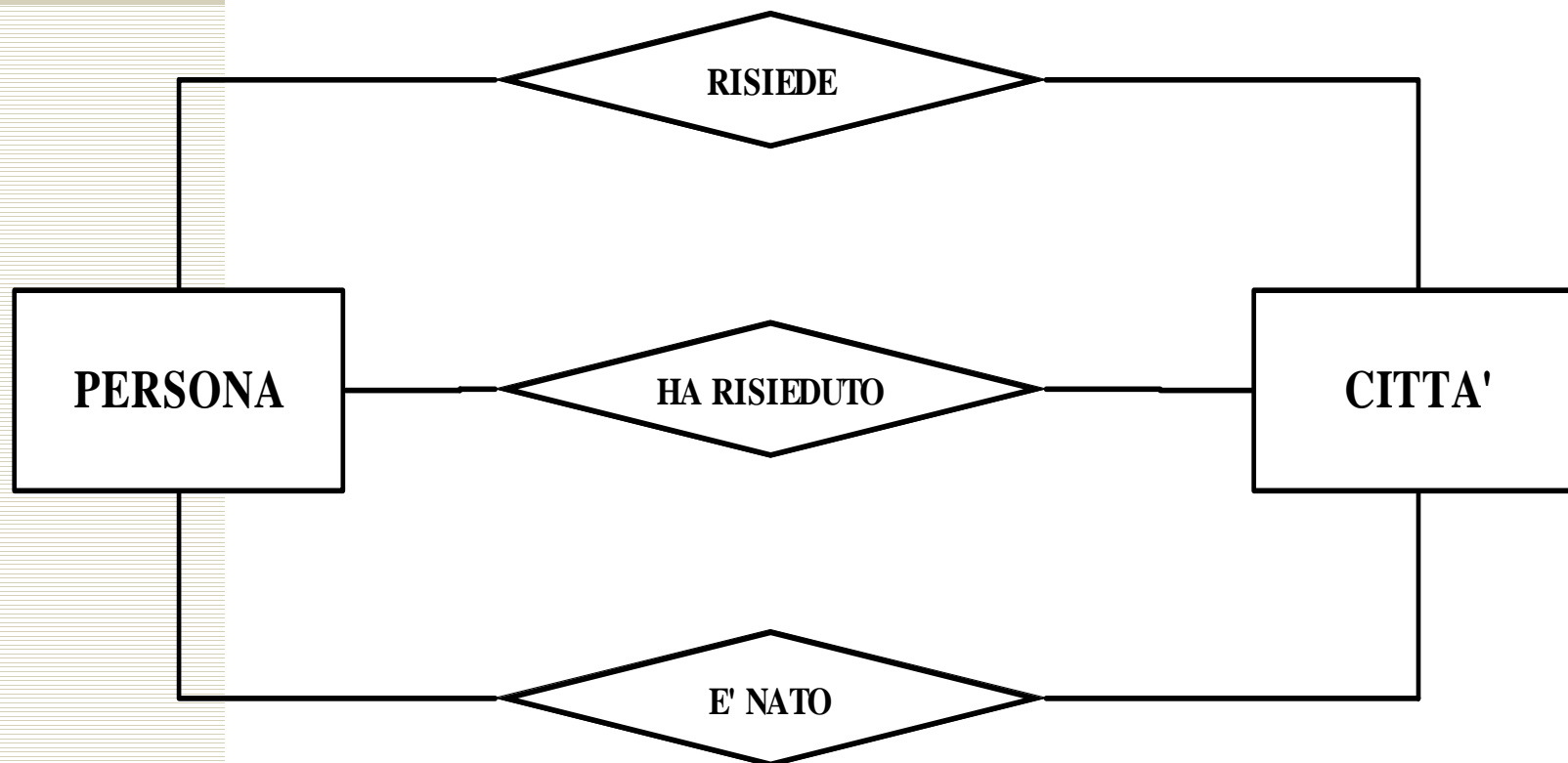
- è **legato ad una specifica organizzazione dei dati**, quale la tabella, anche se è indipendente dalla sua implementazione fisica.
- Tra i Modelli Logici tratteremo il Modello Relazionale per il quale un Database è visto come un insieme di tabelle



# Modelli Concettuali

- ✦ Un Modello Concettuale tende a **descrivere i concetti del mondo reale piuttosto che i dati**.
- ✦ **Modello Concettuale**, è indipendente dalle modalità di organizzazione dei dati.
- ✦ Esso viene utilizzato nella **fase preliminare della progettazione**, per analizzare la realtà di interesse senza dover pensare alle modalità di effettiva sua realizzazione pratica.

# Modello Entity-Relationships







# Modelli dei dati

- **Relazionale**

Il più diffuso, basato su un modello tabellare dei dati

- **Gerarchico**

Usato nei primi DBMS negli anni 60 (ma tuttora utilizzato), basato su strutture ad albero

- **Reticolare**

Estensione del modello gerarchico, basato su grafi

- **A oggetti**

Estensione del modello relazionale basato sui paradigmi di programmazione ad oggetti

Questi modelli sono detti *modelli logici*, in quanto, seppure astratti, riflettono la struttura con cui i dati sono organizzati.

I *modelli concettuali* si collocano ad un livello di astrazione superiore, si svincolano dalla rappresentazione dei dati e rappresentano solo concetti del mondo reale.



# Modelli dei dati

✦ basato sulla teoria matematica delle relazioni tra insiemi (algebra relazionale) ed è basato su un unico concetto fondamentale:

## la relazione

- ✦ che viene rappresentata per mezzo di una tabella:
- le cui righe rappresentano specifici record (informazioni)
  - le cui colonne corrispondono ai campi (dati);

# Modello Relazionale

Definisce tipi attraverso il costruttore *relazione*, che organizza i dati secondo **record a struttura fissa**, rappresentabili attraverso tabelle. ( **un Database è visto come un insieme di tabelle**)

Es. (relazioni INSEGNAMENTO e MANIFESTO)

Corso	Titolare
Basi di Dati	Cagnoni
Reti di Calcolatori	Poggi
Ingegneria Del Software	Bergenti

CdD	Materia	Anno
II	Basi di Dati	3
IE	Reti di Calcolatori	3
II	Ingegneria del Software	3



# Schemi e istanze

- In ogni base di dati si possono distinguere:
  - lo **schema**, sostanzialmente invariante nel tempo, che ne descrive la struttura (aspetto intensionale)
    - nell'esempio, le intestazioni delle tabelle
  - le **istanze**, cioè i valori attuali, che possono cambiare anche molto rapidamente (aspetto estensionale)
    - nell'esempio, il "contenuto" di ciascuna tabella



# Schemi e istanze

## SCHEMA

- ✎ parte invariante nel tempo
- ✎ SCHEMA di un Database definisce, per ogni tabella:
  - Il **Nome**
  - Le **Colonne** in essa presenti (che non variano, a meno di una ridefinizione del DB)
  - La **Tipologia dei dati**, compresi eventuali vincoli, che possono essere inseriti in ogni sua colonna

## ISTANZA

**Insieme dei dati effettivamente presenti nel DataBase in un dato momento** (parte variabile nel tempo)



# Schemi

Lo schema di una base di dati è la parte dichiarativa ed invariante della base di dati e ne definisce la struttura.

Nel modello relazionale lo schema di una relazione è paragonabile alla definizione del prototipo di una funzione in C.

**INSEGNAMENTO(Corso, Titolare)**

è lo *schema* della relazione **INSEGNAMENTO**.

Le effettive n-ple di attributi appartenenti alla relazione sono dette *istanze* della relazione.

Basi di dati Cagnoni è una istanza di **INSEGNAMENTO**.



# Schemi

Gli schemi possono operare a diversi livelli di astrazione

- **Schema logico**  
descrive la struttura dell'intera base di dati mediante il modello logico adottato dal DBMS (reticolare, gerarchico, relazionale)
- **Schema interno**  
implementa lo schema logico per mezzo di strutture fisiche di memorizzazione cioè come insieme di files su disco, con la loro distribuzione nelle varie unità di memorizzazione e la loro struttura (es. come file sequenziali o come file hash o come B-albero ...)

# Schemi

## Schema esterno

descrive la struttura di una porzione della base di dati attraverso il modello logico, riflettendo il punto di vista di una classe di utenti. Generalmente è realizzato per mezzo di *viste*, relazioni derivate da quelle che costituiscono lo schema logico.

ES. (i soli corsi di Ingegneria Informatica)

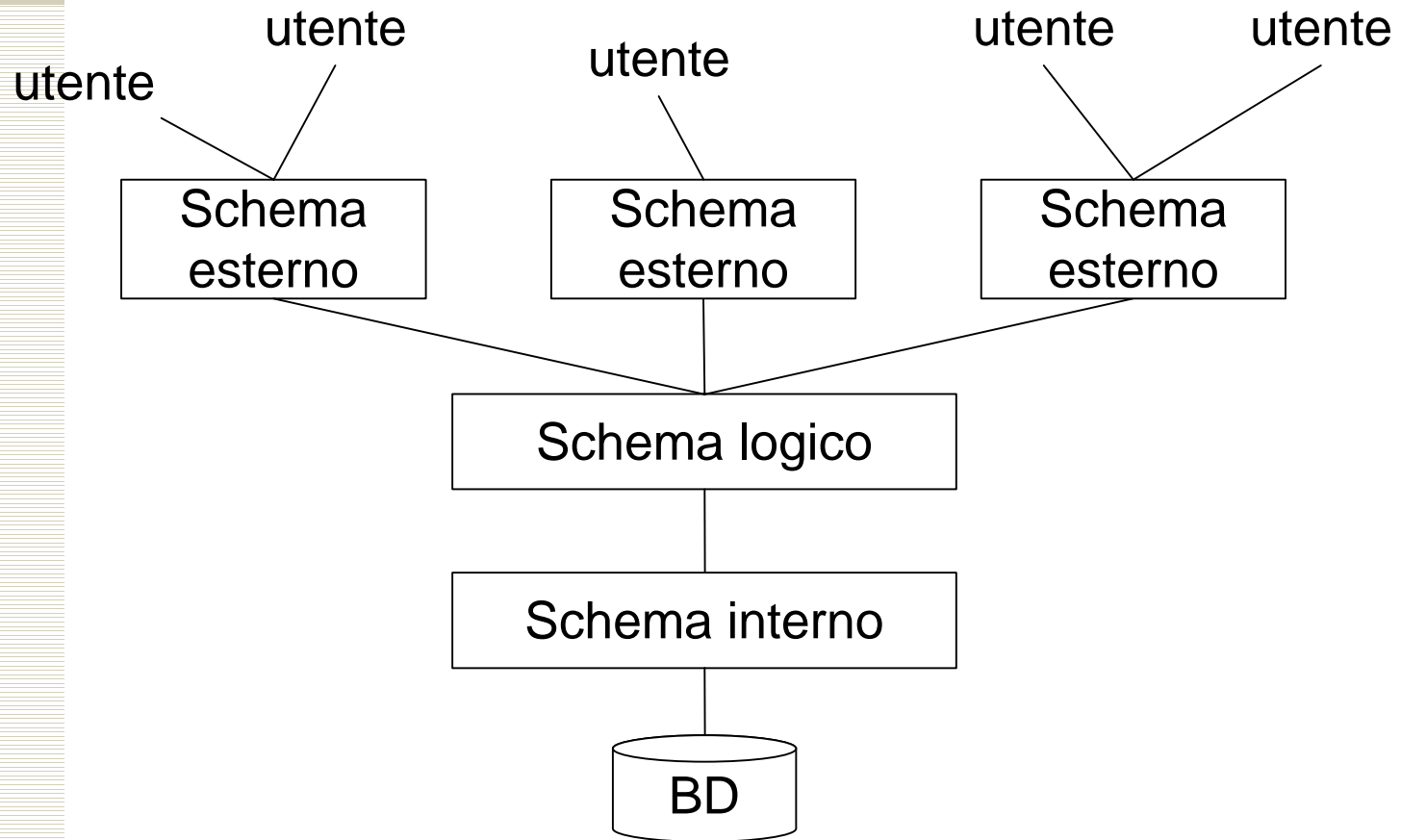
CdL	Materia	Anno
II	Basi di Dati	3
II	Ingegneria del Software	3



# Schemi

- ✦ Lo Schema Logico del Database dell'applicazione **Prenotazione posti aereo** contiene dati su:
  - Personale (di terra e di volo)
  - Passeggeri
  - Bagagli
  - Voli
  - Aeroplani
  - Aeroporti
- ✦ **La VISTA (il sottosistema) dell'Ufficio prenotazioni posti**
  - non comprende dati dagli archivi del personale
  - non comprende dati sui turni dei piloti
- ✦ **La VISTA (il sottosistema) dell'Ufficio spedizioni e bagagli**
  - comprende informazioni su voli e aerei
  - comprende specializzazione piloti (se sono cioè abilitati al trasporto bagagli)
  - non comprende informazioni sullo stipendio del personale

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





# Indipendenza dei dati

Permette agli utenti di operare a livello astratto, indipendentemente dai dettagli realizzativi del DBMS

- **Indipendenza fisica**

Consente di mantenere inalterata la struttura logica dei dati al variare della realizzazione fisica del sistema. Consente di utilizzare basi di dati su piattaforme diverse, o di distribuire una base di dati su più macchine.

- **Indipendenza logica**

Rende indipendente lo schema esterno da quello logico, consentendo di inserire nuove viste senza alterarlo, o di alterarlo mantenendo inalterate le viste definite in precedenza.



# Indipendenza fisica

- ✦ **Il livello logico e quello esterno sono indipendenti da quello fisico**
- ✦ **Possibilità di ridefinire il LIVELLO FISICO senza modificare il LIVELLO LOGICO (e quindi anche le VISTE degli utenti)**
- ✦ **In sostanza:**
  - una tabella è utilizzata nello stesso modo qualunque sia la sua realizzazione fisica (che può anche cambiare nel tempo)
  - è possibile modificare la configurazione dei dati su memoria di massa senza dover modificare i programmi degli utenti
  - Es.: Nel caso dell'array, le procedure che effettuano calcoli su di esso funzionano, senza alcuna modifica, sia che esso venga memorizzato per righe sia che venga memorizzato per colonne.

# Indipendenza logica

✦ Possibilità di modificare il **LIVELLO ESTERNO** (una o più **Viste**) senza dover alterare il **LIVELLO LOGICO**

oppure

✦ Possibilità di definire nuovi **SCHEMI ESTERNI**

oppure

✦ Possibilità di modificare lo **SCHEMA LOGICO** mantenendo inalterate le **Viste** (modificandone ovviamente la definizione in termini delle strutture logiche)

# Linguaggi per DBMS

## 🔧 DDL - Data Description Language

- per la definizione di **schemi** (logici, esterni, fisici) e altre operazioni generali che
- **Agisce sulla struttura del DB**

## 🔧 DML - Data Manipulation Language

- per l'interrogazione e l'aggiornamento di (**istanze** di) basi di dati
- **Agisce sul contenuto del DB.**

🔧 Nei comuni linguaggi di programmazione i due linguaggi coesistono

# DDL

## DATA DESCRIPTION LANGUAGE

Linguaggio che descrive

**tipi di entità e**

**relazioni** intercorrenti tra essi

☛ Usato solo per CREARE o MODIFICARE la struttura del DataBase, non per ottenere o modificare i dati

☛ Es.:

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

# DML

## DATA MANIPULATION LANGUAGE

- ✦ Linguaggio che descrive le operazioni di:
- ✦ **inserimento, cancellazione e modifica** di dati
- ✦ **interrogazione** dei dati (per questo spesso il DML è detto Query Language)
- ✦ Esempi:
  - Ritrovare il numero dei posti disponibili sul volo 999 del 24 luglio
  - Decrementare di 4 il numero di posti disponibili sul volo 123 del 31 Agosto
  - Trovare tutti i voli in partenza da ORD (Chicago) con destinazione JFK (New York) il 20 Agosto
  - Aggiungere il volo 456 con 100 posti in partenza da ORD con destinazione JFK il 21 Agosto



# Linguaggi per basi di dati

- ✿ L'accesso ai dati può essere effettuato utilizzando 4 diverse modalità (4 linguaggi)
- ✿ **linguaggi testuali interattivi (SQL)**
- ✿ **comandi (come quelli del linguaggio interattivo) immersi in un linguaggio ospite** (Pascal, C, Cobol, etc.)
  - Molti DBMS definiscono delle primitive DDL e DML da utilizzare all'interno di programmi scritti con normali linguaggi di programmazione (Cobol, C, Pascal). In questo caso tale linguaggio è detto HOST.
  - Il programma "C" prende decisioni, visualizza domande, legge risposte, ecc. Interagisce con il DBMS solo per accedere ai suoi dati, senza avere la necessità di definirne di propri



# Linguaggi per basi di dati

- ✦ **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), come ad esempio il QBE (Query By Example) principale linguaggio usato dal DBMS Access.

# 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 ad alto livello

```
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
```





# Utenti e progettisti

- **Amministratore della base di dati**  
responsabile della progettazione controllo e manutenzione della base di dati
- **Progettisti e programmatori**  
realizzano i programmi che accedono alla base di dati
- **Utenti**
  - **utenti finali**  
accedono alla base di dati frequentemente e attraverso procedure predefinite
  - **utenti casuali**  
interrogano o aggiornano la base di dati con procedure e modalità variabili utilizzando i linguaggi forniti dal DBMS



# Vantaggi dei DBMS

- Disponibilità dei dati a tutta una comunità
- Modello unificato e preciso della realtà di interesse
- Controllo centralizzato dei dati
- Condivisione
- Indipendenza dei dati



# Svantaggi dei DBMS

- Prodotti costosi, complessi, che richiedono investimenti in hardware, software, personale.
- Forniscono un numero elevato di servizi, in modo integrato e difficilmente scorporabile se le esigenze dell'utente sono inferiori alle caratteristiche offerte