

Lezione 1 - Introduzione alle basi di dati e ai sistemi di gestione di basi di dati

Prof.ssa Maria De Marsico
demarsico@di.uniroma1.it



SAPIENZA
UNIVERSITÀ DI ROMA

Parte del materiale è tratto dal materiale associato al testo Atzeni, Ceri, Paraboschi, Torlone
Basi di dati - Modelli e linguaggi di interrogazione, McGraw-Hill, 2002.

Argomenti del modulo

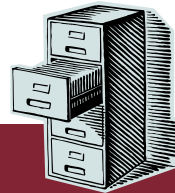


SAPIENZA
UNIVERSITÀ DI ROMA
DIPARTIMENTO DI INFORMATICA

Il primo modulo del corso di Basi di Dati tratterà i seguenti argomenti principali:

- Algebra relazionale
- Progettazione di una base di dati: Terza Forma Normale (3NF)
- Organizzazione fisica dei dati
- Controllo della concorrenza

- Componente di una organizzazione che viene utilizzata per gestire (acquisire, processare, memorizzare, comunicare) le informazioni di interesse
- Normalmente il Sistema Informativo opera a supporto delle altre componenti dell'organizzazione
- La nozione di Sistema Informativo è indipendente dalla sua computerizzazione
- Esempi di Sistema Informativo ...



3

Cosa succedeva prima?

- Ogni applicazione aveva il suo file privato
 - **file**: organizzazione sequenziale
 - **applicazione**: scritta in un linguaggio orientato alla gestione di file (Cobol, PL/1)
 - **gestione dei dati**: file system
-

4

Cosa succedeva prima?

- Svantaggi:
 - **ridondanza**: se due applicazioni usavano gli stessi dati, questi venivano replicati
 - **inconsistenza**: l'aggiornamento di un dato poteva riguardare una sola copia del dato
 - **dipendenza dei dati**: ogni applicazione organizzava i dati tenendo conto dell'uso che doveva farne
-

5

DBMS

- Una **Base di Dati** (Database – DB) è un insieme di file mutuamente connessi.
 - Gli insiemi di dati sono organizzati in diverse *strutture di dati* che ne facilitano la creazione, l'accesso e l'aggiornamento ed ottimizzano la gestione delle risorse fisiche.
- I **Sistemi di Gestione di Basi di Dati** (Database Management System – DBMS) sono strumenti software per la gestione di grandi masse (strutturate, processabili, condivise) di dati residenti su memoria secondaria

6

Registrazione dell'informazione in formato elettronico

- **Dati strutturati:** gli oggetti sono rappresentati da brevi stringhe di simboli e da numeri
- **Dati non strutturati:** testi scritti in un linguaggio naturale

7

- La struttura dell'informazione dipende dal suo utilizzo e può essere modificata nel tempo
- Esempio: per memorizzare dati su una persona, nel corso del tempo:
 - Nome e cognome (fino a qualche secolo fa non era ovvio neppure questo)
 - Nome, cognome, data di nascita e luogo di nascita
 - Codice Fiscale
 - ...

8

- **Obiettivo:** facilitare l'elaborazione dei dati sulla base delle loro relazioni.
- **Dati strutturati**
 - Si può accedere singolarmente agli elementi della struttura tramite «interrogazioni» per recuperare informazioni o effettuare calcoli
 - Le relazioni tra i dati individuali sono rappresentate nella struttura dei record

9

- In una organizzazione ogni componente è interessata ad una porzione del Sistema Informativo
- Queste porzioni possono sovrapporsi
- Una base di dati è una risorsa integrata condivisa da diverse componenti
- L'integrazione e la condivisione permettono di ridurre ridondanze (dati parzialmente o totalmente replicati) e conseguenti inconsistenze

10

- La condivisione non è mai completa: controllo della privacy e regolamentazione degli accessi
- La condivisione comporta la necessità di gestire accessi contemporanei agli stessi dati: controllo della concorrenza

11

- Un **Sistema Informativo** è un complesso di dati organizzati fisicamente in una memoria secondaria e gestiti in maniera tale da consentirne la creazione, l'aggiornamento e l'interrogazione.

12

- I dati sono organizzati concettualmente in aggregati di informazioni omogenee che costituiscono le componenti del sistema informativo, e ogni operazione di aggiornamento ha per oggetto un singolo aggregato mentre un'interrogazione può coinvolgerne uno o più.
- Nelle basi di dati:
 - aggregati di informazioni omogenee: file
 - indici: file che permettono di recuperare velocemente le informazioni dei «file principali»

13

- Nei sistemi computerizzati l'informazione è rappresentata sotto forma di dati
 - Dati: fatti grezzi che devono essere interpretati e correlati per fornire informazione
- Esempio:
 - “Maria De Marsico” e 0649918312 sono una stringa e un numero, ossia due dati
 - Se sono restituiti in risposta alla domanda “Chi è il docente del corso e qual è il suo numero di telefono” allora costituiscono informazione

14

- strutture da utilizzare per organizzare i dati di interesse e le loro relazioni
- componente fondamentale: costruttori di tipo
 - esempio: il modello relazionale prevede il costruttore di relazione: organizza i dati come insieme di record (tipi) omogenei

15

Due tipi principali di modelli

- **modelli logici**: indipendenti dalle strutture fisiche ma disponibili nei DBMS: es. reticolare (network), gerarchico (hierarchical), relazionale (relational), ad oggetti (object)
- **modelli concettuali**: indipendenti dalle modalità di realizzazione, hanno lo scopo di rappresentare le entità del mondo reale e le loro relazioni nelle prime fasi della progettazione: es. entità-relazioni (entity-relationship)

16

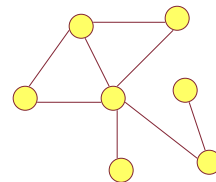
metà anni ' 60: primi sistemi

- **gerarchici**
 - Generalized Update Access Method
 - (IBM, Progetto Apollo, 1964)
 - DL/1 (Data Language 1)
 - (IBM, in commercio nel 1966)
 - **IMS** (Information Management System)
- **a rete**
 - I-D-S (Integrated Data Store)
 - (General Electric)
 - sistemi CODASYL / DBTG / a rete

17

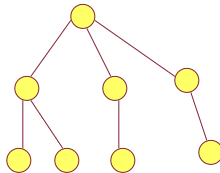
Modello reticolare

- I dati sono rappresentati come una collezione di *record* di tipo omogeneo
- Le relazioni binarie sono rappresentate come *link* (implementati come puntatori = dipendenza dalla struttura fisica della base di dati)
- Il modello è rappresentato come una struttura a grafo dove :
 - Nodi = record
 - Archi = link
- Il più popolare modello reticolare: CODASYL



18

- Tipo ristretto di modello reticolare:
 - Gerarchia = reticolo composto da una collezione di alberi (foresta)
 - Ogni nodo ha un solo genitore



19

Cenni storici

- **1970:** E.F. Codd (IBM) introduce il **modello relazionale**
- **anni ' 70:** prototipi sistemi relazionali (System R, IBM)
- **anni ' 80:** sistemi relazionali commerciali (Ingress, Oracle, ...)

20

- Dati e relazioni sono rappresentati come **valori**
- Non ci sono riferimenti espliciti, cioè puntatori come nei modelli reticolare e gerarchico
- => rappresentazione di livello più alto

21

- Oggetto = Record
- Campi = Informazioni di interesse



- Oggetto = “Membro dello Staff”
- Informazioni di interesse = Codice, Cognome, Nome, Ruolo, Anno di assunzione

CODICE	COGNOME	NOME	RUOLO	ASSUNZIONE
COD1	Rossi	Mario	Analista	1995

22

Modello relazionale

- Tabella = Insieme di record di tipo omogeneo



- Tabella STAFF = Insieme di record di tipo “Membro dello Staff”

CODICE	COGNOME	NOME	RUOLO	ASSUNZIONE
COD1	Rossi	Mario	Analista	1995
COD2	Bianchi	Pietro	Analista	1990
COD3	Neri	Paolo	Amministratore	1985

23

Esempio di DB relazionale

STUDENTI

Matric	Cognome	Nome	DataNasci
276545	Smith	Mary	25/11/1980
485745	Black	Anna	23/04/1981
200768	Verdi	Paolo	12/02/1981
587614	Smith	Lucy	10/10/1980
937653	Brown	Mavis	01/12/1980

CORSI

Codice	Titolo	Tutor
01	Physics	Grant
03	Chemistry	Beale
04	Chemistry	Clark

ESAMI

Stud	Voto	Corso
276545	C	01
276545	B	04
937653	B	01

24

Esempio di DB relazionale

STUDENTI

Matric	Cognome	Nome	DataNasci
276545	Smith	Mary	25/11/1980
485745	Black	Anna	23/04/1981
200768	Verdi	Paolo	12/02/1981
587614	Smith	Lucy	10/10/1980
937653	Brown	Mavis	01/12/1980

CORSI

Codice	Titolo	Tutor
01	Physics	Grant
03	Chemistry	Beale
04	Chemistry	Clark

ESAMI

Stud	Voto	Corso
276545	C	01
276545	B	04
937653	B	01

STUDENTI

Matric	Cognome	Nome	DataNasci
276545	Smith	Mary	25/11/1980
485745	Black	Anna	23/04/1981
200768	Verdi	Paolo	12/02/1981
587614	Smith	Lucy	10/10/1980
937653	Brown	Mavis	01/12/1980

ESAMI

Stud	Voto	Corso
	C	
	B	
	B	
	B	

CORSI

Codice	Titolo	Tutor
01	Physics	Grant
03	Chemistry	Beale
04	Chemistry	Clark

Esempio di DB reticolare

- **metà anni ' 80:** primi sistemi **orientati agli oggetti**
 - (O2, inizialmente INRIA e successivamente O2 Technology)
- **a partire dal ' 93:** definizione di uno standard (Object Data Management Group)

27

Modello a oggetti

- Modello basato su oggetti, classi, ecc.
- Attributi: descrivono lo stato di un oggetto
- Metodi (azioni) descrivono il comportamento di un oggetto
- L'oggetto incapsula sia stato che comportamento
- Non esiste ancora un modello universalmente riconosciuto

28

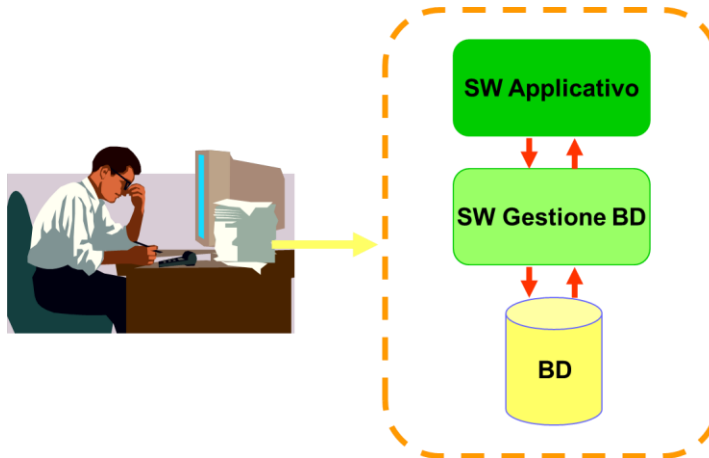
- DB (Database = Base di Dati): collezione di dati logicamente correlati di interesse per il Sistema Informativo
- DBMS (Database Management System = Sistema di gestione della Base di Dati): componente software che interagisce con la Base di Dati da una parte e con i programmi applicativi degli utenti dall'altra
- **Oggetti nella Base di Dati:** Memorizzano proprietà di "Oggetti" e Relazioni tra Oggetti nel dominio di interesse

29

Componenti di un sistema informativo informatizzato

- Base di dati (BD)
- Sistema di gestione della Base di Dati (DBMS)
- Software applicativo
- Hardware del computer (es. dispositivi di memorizzazione)
- Personale che sviluppa, gestisce o usa il sistema

30



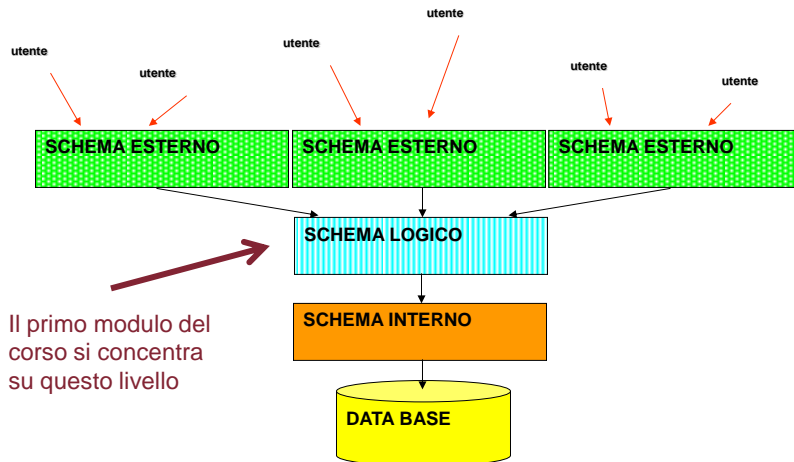
Memorizzazione

I sistemi di basi di dati utilizzano file in formati proprietary per memorizzare i dati

ma

offrono agli utenti una vista astratta dei dati, in modo da rendere trasparenti i dettagli di memorizzazione e manipolazione

I tre livelli di astrazione di un DB



33

I tre livelli di astrazione di un DB

- **Schema esterno:** descrizione di una porzione della base di dati in un modello logico attraverso “viste” parziali, o derivate, che possono prevedere organizzazioni dei dati diverse rispetto a quelle utilizzate nello schema logico, e che riflettono esigenze e privilegi di accesso di particolari tipologie di utenti; ad uno schema logico si possono associare più schemi esterni
- **Schema logico:** descrizione dell'intera base di dati nel modello logico “principale” del DBMS, ad esempio la struttura delle tabelle
- **Schema fisico:** rappresentazione dello schema logico per mezzo di strutture fisiche di memorizzazione, cioè i file

34

Una vista (schema esterno)

Corsi

Corso	Docente	Aula
Basi di dati	Rossi	DS1
Sistemi	Neri	N3
Reti	Bruni	N3
Controlli	Bruni	G

Aule

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

SCHEMA LOGICO

CorsiSedi

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

VISTA

35

Accesso ai dati

- Gli accessi alla base di dati avvengono solamente attraverso lo schema esterno, che può coincidere completamente con lo schema logico.

36

- **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 (organizzazione dei file e loro allocazione fisica)
- la realizzazione fisica può cambiare senza che debbano essere modificati i programmi

37

- **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

38

- In ogni base di dati esistono:
 - lo schema, sostanzialmente invariante nel tempo, che ne descrive la struttura (aspetto intensionale): nel modello relazionale, le intestazioni delle tabelle = lista di nomi di attributi e loro tipi
 - l'istanza, i valori attuali, che possono cambiare anche molto rapidamente (aspetto estensionale): nel modello relazionale, il “corpo” di ciascuna tabella

39

ANAGRAFICA

NOME	COGNOME	DATAN	LUOGON
Piero	Napoli	22-10-63	Bari
Marco	Bianchi	01-05-54	Roma
Maria	Rossi	09-02-68	Milano
Maria	Bianchi	07-12-70	Bari
Paolo	Sossi	15-03-75	Palermo

40

Schemi e istanze

ANAGRAFICA

NOME	COGNOME	DATAN	LUOGON
Piero	Napoli	22-10-63	Bari
Marco	Bianchi	01-05-54	Roma
Maria	Rossi	09-02-68	Milano
Maria	Bianchi	07-12-70	Bari
Paolo	Sossi	15-03-75	Palermo

SCHEMA

41

Schemi e istanze

ANAGRAFICA

NOME	COGNOME	DATAN	LUOGON
Piero	Napoli	22-10-63	Bari
Marco	Bianchi	01-05-54	Roma
Maria	Rossi	09-02-68	Milano
Maria	Bianchi	07-12-70	Bari
Paolo	Sossi	15-03-75	Palermo

ISTANZA

42

- **data definition language (DDL)**
 - per la definizione di **schemi** (logici, esterni, fisici) e altre operazioni generali
- **data manipulation language (DML)**
 - per l'interrogazione e l'aggiornamento di (**istanze di**) basi di dati
- **SQL** (Structured Query Language) è un linguaggio standardizzato per database basati sul modello relazionale (RDBMS)
- In SQL i due tipi di funzionalità sono integrate in un unico linguaggio di comandi

43

Le basi di dati: ricapitolando

- **caratteristiche**
 - multiuso
 - integrazione
 - indipendenza dei dati
 - controllo centralizzato (DBA: database administrator)
- **vantaggi**
 - minima ridondanza
 - indipendenza dei dati
 - integrità
 - sicurezza

44

Integrità

- I dati devono soddisfare dei “vincoli” che esistono nella realtà di interesse
 - uno studente risiede in una sola città (**dipendenze funzionali**)
 - la matricola identifica univocamente uno studente (**vincoli di chiave**)
 - un voto è un intero positivo compreso tra 18 e 30 (**vincoli di dominio**)
 - lo straordinario di un impiegato è dato dal prodotto del numero di ore per la paga oraria
 - lo stipendio di un impiegato non può diminuire (**vincoli dinamici**)

45

Sicurezza

- I dati devono essere protetti da accessi non autorizzati; il DBA deve considerare
 - **valore corrente** dell'informazione **per l'organizzazione**
 - **valore corrente** dell'informazione **per chi vuole violare la privacy**
 - **chi** può accedere a **quali dati** e in **quale modalità**
- e quindi decidere
- **regolamento di accesso**
 - **effetti di una violazione**

46

-
- I dati devono essere protetti da malfunzionamenti dell' hardware e del software e dall' accesso concorrente alla base di dati

47

Transazione

- **Transazione:** sequenza di operazioni che costituiscono un' unica operazione logica
 - “Trasferisci €1000 dal c/c c1 al c/c c2”
 - cerca c1
 - modifica *saldo* in *saldo*-1000
 - cerca c2
 - modifica *saldo* in *saldo*+1000
- Una transazione deve essere eseguita completamente (**committed**) o non deve essere eseguita affatto (**rolled back**)

48

Ripristino

- Per ripristinare un valore corretto della base di dati:
 - **transaction log** (contiene i dettagli delle operazioni: valori precedenti e seguenti la modifica)
 - **dump** (copia periodica della base di dati)

49

Concorrenza

- **Transazione 1:** “Accredita €1000 sul c/c c1”
 - **Transazione 2:** “Accredita € 500 sul c/c c1”
- NOTA!!!** Una volta letto un valore, ogni transazione lo modifica nel proprio spazio di memoria

Transazione 1	Tempo	Transazione 2
cerca c1	t1	
	t2	cerca c1
modifica <i>saldo</i> in <i>saldo</i> +1000	t3	
	t4	modifica <i>saldo</i> in <i>saldo</i> +500

valore iniziale *saldo*: 2000

valore finale *saldo*: 2500

50

Compiti del DBA

- **Progettazione**

definizione e descrizione di

- schema logico
- schema fisico
- sottoschemi e/o viste

(Data Definition Language)

- **Mantenimento**

- modifiche per nuove esigenze o per motivi di efficienza
- (routine: caricamento, copia e ripristino, riorganizzazione, statistiche, analisi)

51

Programma del Modulo 1

Ricapitolando

Il primo modulo del corso di Basi di Dati tratterà i seguenti argomenti principali:

- Algebra relazionale: Linguaggio di interrogazione procedurale
- Progettazione di una base di dati: come garantire/verificare la Terza Forma Normale (3NF), come decomporre uno schema preservando le dipendenze e l'informazione
- Organizzazione fisica dei dati
- Controllo della concorrenza

52