

Basi di dati

Capitolo 1: Introduzione

1/73

Punti di vista

- Metodologico
- Tecnologico

3/73

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](#)

5/73

Base di dati

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

2/73

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

4/73

Metodo di studio

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

6/73

Sistema informativo

- Componente di una organizzazione che gestisce le informazioni di interesse (cioé 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



7/73

Gestione delle informazioni

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



8/73

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

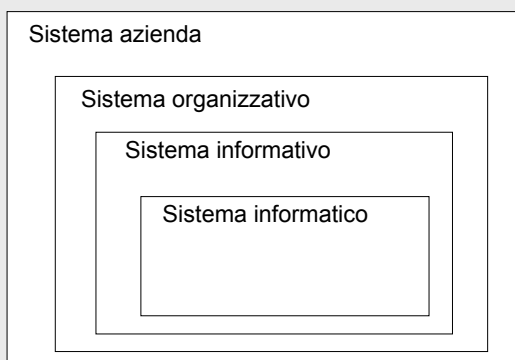
9/73

Sistema Informatico

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

10/73

Sistema Informatico



11/73

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

12/73

Informazioni e dati

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

13/73

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

14/73

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

15/73

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

16/73

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

17/73

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**

18/73

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*

19/73

Sistema di gestione di basi di dati DataBase Management System (DBMS)

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

21/73

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



23/73

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

20/73

DBMS

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



22/73

Le basi di dati sono ... persistenti

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



24/73

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)



25/73

Esempio di non condivisione: dati per orario lezioni

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

26/73

Esempio di non condivisione: ... e dati per orario ricevimento

UNIVERSITA' DEGLI STUDI CHISSADOVE		
Corso di Studi in Ingegneria Informatica		
Orario di ricevimento dei docenti		
DOCENTE	INSEGNAMENTI	ORARIO
Mario BRUNI	Fisica I Fisica II	Martedì' 10-12
Luigi NERI	Analisi matematica I	Lunedì' 12-13
Piero ROSSI	Basi di dati Sistemi informativi	Giovedì' 11-13
Nicola MORI	Chimica	Martedì' 16-18

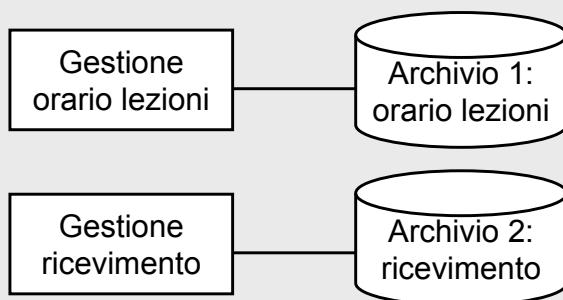
27/73

Problemi

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

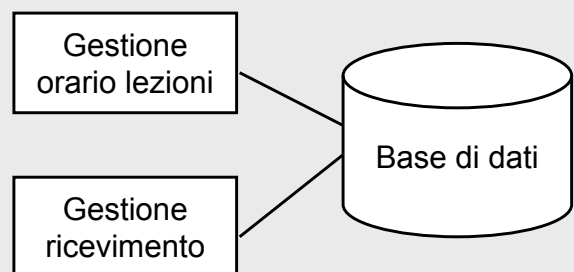
28/73

Archivi e basi di dati



29/73

Archivi e basi di dati



30/73

Le basi di dati sono condivise

- 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**



31/73

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



32/73

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**

33/73

Transazione

- Insieme di operazioni da considerare indivisibile ("**atomico**"), corretto anche in presenza di **concorrenza** e con effetti **definitivi**



34/73

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



35/73

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



36/73

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



37/73

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



38/73

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

39/73

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

40/73

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

41/73

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**

42/73

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

43/73

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

45/73

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

Ex

47/73



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

44/73

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

46/73

Due tipi (principali) di modelli

- modelli logici
- modelli concettuali

48/73

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



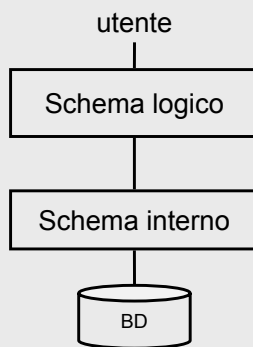
49/73

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**

50/73

Architettura (semplificata) di un DBMS



51/73

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)

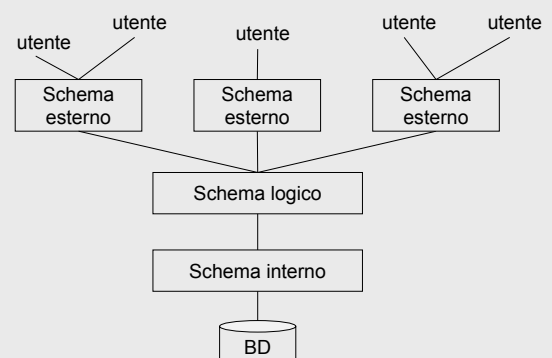
52/73

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

53/73

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



54/73

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)

55/73

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

57/73

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

59/73

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

Corsi Sedi

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

56/73

Indipendenza fisica

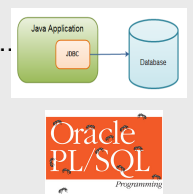
- il livello logico e quello esterno sono indipendenti da quello fisico
 - una relazione è utilizzata nello stesso modo qualunque sia la sua realizzazione fisica
 - la realizzazione fisica può cambiare senza che debbano essere modificati i programmi



58/73

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, C, Java, ..)
 - 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)



60/73

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"

61/73

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



62/73

SQL immerso in linguaggio ospite

```
write('nome della città?'); readln(città);
EXEC SQL DECLARE P CURSOR FOR
  SELECT NOME, REDDITO
  FROM PERSONE
  WHERE CITTA = :città ;
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
```



63/73

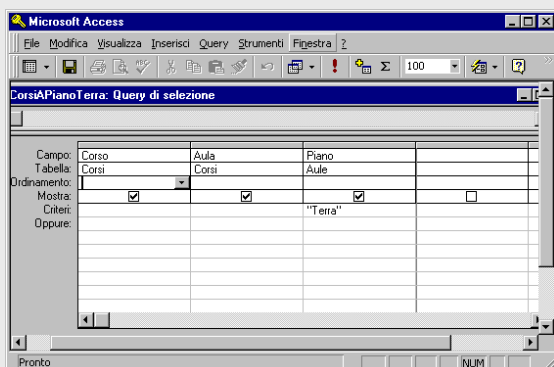
SQL in linguaggio ad hoc (Oracle PL/SQL)

```
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;
```



64/73

Interazione non testuale (Access)



65/73

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

66/73

Un'operazione DDL (sullo schema)

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

67/73

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



68/73

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



69/73

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)

70/73

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**)



71/73

Vantaggi e svantaggi dei DBMS, 1

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)

72/73

Vantaggi e svantaggi dei DBMS, 2

Contro

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

73/73

- 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

75/73

- 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

77/73

Esercizi su

Basi di dati e DBMS

74/73

- 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

76/73

- 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

78/73

- Illustrare, in modo sintetico ma chiaro, supponendo di rivolgersi ad un non esperto, le caratteristiche fondamentali delle basi di dati e il ruolo che esse giocano nei sistemi informativi.
- Discutere brevemente (meno di mezza pagina) la seguente affermazione: "i dati sono una risorsa per una organizzazione, e come tali vanno considerati anche separatamente dalle applicazioni che li utilizzano."
- Illustrare brevemente (non più di mezza pagina) il concetto di indipendenza dei dati.