Corso di Laurea in Informatica Anno 2013-2014 Corso di Basi di Dati



Docente

Prof. Sabrina Mantaci

Obiettivi del corso



- 1. Fornire i concetti e le metodologie fondamentali sulle basi di dati e sui sistemi per la loro gestione, con particolare riguardo alle basi di dati relazionali.
- 2. Strumenti concettuali per l'interrogazione di una base di dati (Algebra relazionale). Costruzione, interrogazione, accesso a basi di dati mediante il linguaggio SQL con strumenti DBMS relazionali (Oracle o MySQL).
- 3. Progettazione di basi di dati.

Programma Previsto

- Introduzione alle basi di dati: obiettivi e servizi di un DBMS. Modelli dei dati.
- Il modello relazionale: Concetti base. Vincoli di integrità.
- Algebra relazionale.
- Il linguaggio SQL. Definizione dei dati. Interrogazioni in SQL. Manipolazione dei dati in SQL. Viste in SQL. Sicurezza delle basi di dati. Modelli e politiche per il controllo dell'accesso. Applicazioni su un DBMS reale (Oracle o mySQL).
- Il modello Entità-Relazione. Concetti base. Vincoli di integrità. Gerarchie di generalizzazione.
- Metodologie di progettazione di basi di dati. Fasi della progettazione. Strategie per la progettazione concettuale. Integrazione di viste. Progettazione logica di basi di dati relazionali.
- La normalizzazione (forse...). Concetto di dipendenza funzionale. Forme normali. Cenni sulla decomposizione di relazioni e relative proprietà.

Base di dati

4

Una base di dati è un insieme organizzato di dati utilizzati per il supporto allo svolgimento di attività (di un ente, azienda, ufficio, persona).

Che cos'è l'informatica?



• L'informatica è la scienza del trattamento razionale, specialmente per mezzo di macchine automatiche, dell'informazione, considerata come supporto alla conoscenza umana e alla comunicazione (Accademia di Francia)

• Due anime:

- ✓ Tecnologica: i calcolatori elettronici e i sistemi che li utilizzano.
- ✓ Metodologica: i metodi per la risoluzione di problemi e la gestione delle informazioni.

Sistema informativo



- Un Sistema Informativo è un componente (sottosistema) di una organizzazione che si occupa di gestire (acquisisce, elabora, conserva, produce) le informazioni di interesse (cioè utilizzate per il perseguimento degli scopi dell'organizzazione)
 - o ogni organizzazione ha un sistema informativo, eventualmente non esplicitato nella struttura;
 - o quasi sempre, il sistema informativo è di supporto ad altri sottosistemi, e va quindi studiato nel contesto in cui è inserito
 - o il sistema informativo è di solito suddiviso in sottosistemi (in modo gerarchico o decentrato), più o meno fortemente integrati

Sistema organizzativo



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

Le Risorse di un Sistema organizzativo



Le risorse di una sistema organizzativo (azienda, amministrazione, etc...) sono:

- Persone
- Denaro
- Attrezzature/materiali
- Sede
- Informazioni

Sistemi informativi e automazione



- Il concetto di "sistema informativo" è indipendente da qualsiasi automatizzazione:
 - o esistono organizzazioni la cui ragion d'essere è la gestione di informazioni (p. es. servizi anagrafici e banche) e che operano da secoli.
 - Anche prima di essere automatizzati, molti sistemi informativi si sono evoluti verso una razionalizzazione e standardizzazione delle procedure e dell'organizzazione delle informazioni

Funzioni di un Sistema informativo

- raccolta, acquisizione delle informazioni
- archiviazione, conservazione delle informazioni
- elaborazione delle informazioni
- distribuzione, scambio di informazioni

Sistema Informatico



• Un sistema informatico è la porzione automatizzata del sistema informativo, ossia la parte del sistema informativo che gestisce informazioni con tecnologia informatica

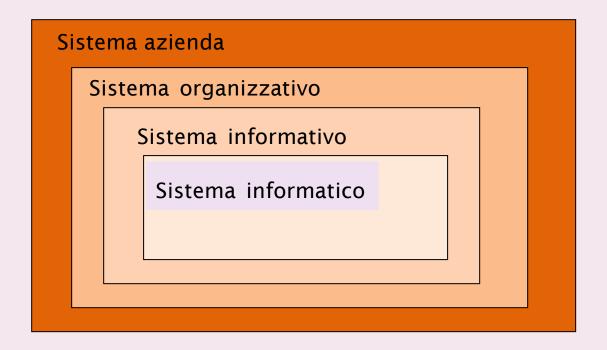
Componenti di un sistema informatico



- Software e Hardware di base
- Una *base di dati*, che contiene una rappresentazione del patrimonio informativo dell'organizzazione
- Uno *schema*, che descrive la struttura della base di dati, le operazioni per agire su di essa e le restrizioni sui valori memorizzabili nella base di dati e sui modi in cui essi si possono evolvere nel tempo (*vincoli di integrità*). Lo schema della base di dati viene usato dal sistema per garantire un uso corretto della base di dati
- I *programmi applicativi*, che forniscono servizi agli utenti eseguendo un certo insieme di operazioni sulla base di dati
- La comunicazione, che permette l'accesso ai servizi del sistema informatico ad utenti e programmi

Complessivamente...





Gestione delle informazioni



- Nelle attività umane, le informazioni vengono gestite (registrate e scambiate) in forme diverse:
 - o idee informali
 - o linguaggio naturale (scritto o parlato, formale o colloquiale, in una lingua o in un'altra)
 - o disegni, grafici, schemi
 - o numeri e codici
- e su vari supporti
 - o 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 informazioni schematizzate dai valori di particolari voci, come:
 - nome e cognome
 - o estremi anagrafici
 - o 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
 - **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



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 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)

Base di dati



Una base di dati è:

(accezione generica, metodologica)

Insieme organizzato di dati utilizzati per il supporto allo svolgimento delle attività di un ente (azienda, ufficio, persona) o meglio del sistema informativo

(accezione specifica, metodologica e tecnologica) insieme di dati gestito da un DBMS

Sistema di gestione di basi di dati DataBase Management System — DBMS

Un DataBase Management System (DBMS) è un sistema (prodotto software) in grado di gestire collezioni di dati che siano (anche):

- o **Grandi** (di dimensioni (molto) maggiori della memoria centrale dei sistemi di calcolo utilizzati)
- o **Persistenti** (con un periodo di vita indipendente dalle singole esecuzioni dei programmi che le utilizzano)
- o **Condivise** (utilizzate da applicazioni diverse)

garantendo **affidabilità** (resistenza a malfunzionamenti hardware e software-recovery) e **privatezza** (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) ed **efficace** (rendendo produttive le attività dei suoi utilizzatori).

Alcuni DBMS in commercio



- Access
- DB2
- Oracle
- Informix
- Sybase
- SQLServer
- Postgres

Condivisione



- 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

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



Basi di dati 1 23

Possibili problemi



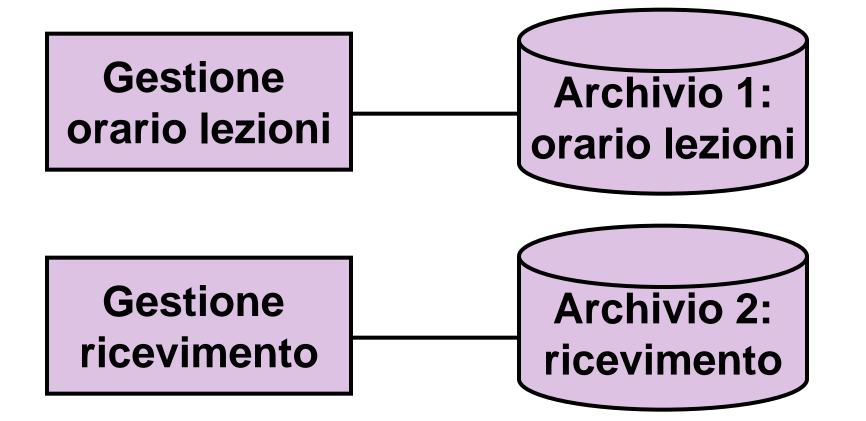
Se i dati non vengono gestiti in maniera organica, possiamo incorrere in problemi di:

Ridondanza: informazioni ripetute

Rischio di **incoerenza**: le versioni possono non coincidere

Archivi e basi di dati





Archivi e basi di dati





Gestione ricevimento

Base di dati

Le basi di dati sono condivise



- Una base di dati è una risorsa integrata e condivisa fra le varie applicazioni
- conseguenze
 - Attività diverse su dati in parte condivisi:
 - x meccanismi di autorizzazione
 - Attività multi-utente su dati condivisi:
 - x 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).
- I DBMS, a causa della varietà di funzioni, non sono necessariamente più efficienti dei file system.
- L'efficienza è il risultato della qualità del DBMS e delle applicazioni che lo utilizzano.

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". Nei DBMS, c'è maggiore flessibilità
- I DBMS estendono le funzionalità dei file system, fornendo più servizi ed in maniera integrata (cfr. Efficacia)

DBMS vs file system (2)



- 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



- Descrizioni e rappresentazioni dei dati a livelli diversi
 - o 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



- Un modello di dati è un 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
- ad esempio, il modello relazionale prevede il costruttore relazione, che permette di definire insiemi di record omogenei

Due tipi (principali) di modelli



- modelli logici: utilizzati nei DBMS esistenti per l'organizzazione dei dati
 - o utilizzati dai programmi
 - o indipendenti dalle strutture fisiche

Esempi: relazionale, reticolare, gerarchico, a oggetti, modello XML.

- modelli concettuali: permettono di rappresentare i dati in modo indipendente da ogni sistema e dalla scelta del modello logico
 - o cercano di descrivere i concetti del mondo reale
 - o sono utilizzati nelle fasi preliminari di progettazione
 - il più noto è il modello Entity-Relationship

Architettura semplificata di un DBMS: schemi

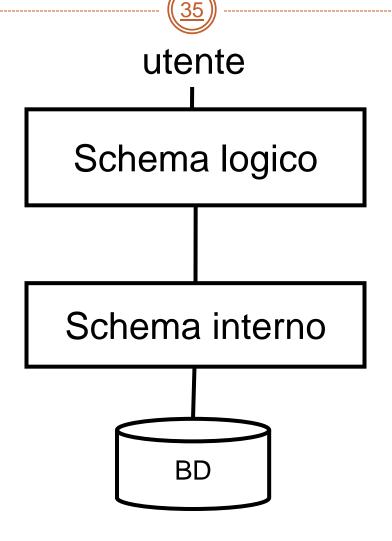


- o **schema logico**: descrizione della base di dati nel modello logico (ad esempio, la struttura della tabella)
- o **schema fisico**: rappresentazione dello schema logico per mezzo di strutture memorizzazione (file sequenziale, file hash, file sequenziale con indici...)

Oss: il livello logico è indipendente da quello fisico. Si parla di Indipendenza dei Dati:

una tabella è utilizzata nello stesso modo qualunque sia la sua realizzazione fisica (che può anche cambiare nel tempo)

Architettura (semplificata) di un DBMS



Architettura ANSI/SPARC: schemi

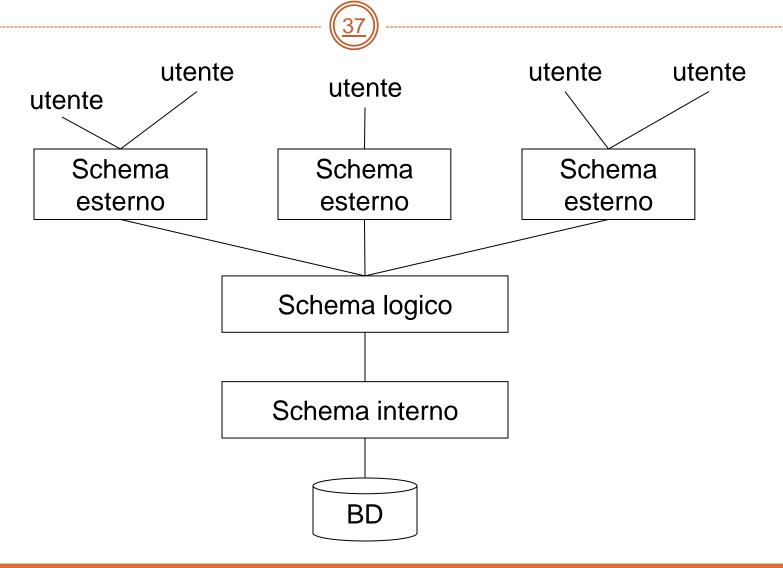


schema logico: descrizione dell'intera base di dati nel modello logico "principale" del DBMS (relazionale, gerarchico, reticolare, a oggetti...)

schema fisico: rappresentazione dello schema logico per mezzo di *strutture fisiche* di memorizzazione (file sequenziali, file hash, file sequenziali con indici)

schema esterno: descrizione di *parte* della base di dati in un modello logico ("viste" parziali, derivate, anche in modelli diversi). Possono esserci anche schemi esterni diversi, secondo il tipo di utente.

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



Esempio



Nell'organizzazione di una banca, lo schema logico conterrà tutte le tabelle e i dati relativi ai conti correnti, ma anche al personale. Lo schema logico conserva tutte le informazioni della banca.

Nello schema esterno, ogni correntista potrà accedere solo ad alcune informazioni che sono per lui di interesse: quelle relative al PROPRIO conto corrente. Questi meccanismi di selezione delle informazioni, ma anche di riservatezza (i dati di un correntista non sono visibili agli altri) vengono gestiti al livello esterno, con meccanismi di autorizzazione, ma anche per esempio tramite le viste

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 di indipendenza:
 - o indipendenza fisica
 - o indipendenza logica

Indipendenza fisica



- il livello logico e quello esterno sono indipendenti da quello fisico
 - o 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 all'utente finale

Schemi e istanze



- In ogni base di dati si distinguono:
 - o lo schema, sostanzialmente invariante nel tempo, che ne descrive la struttura (aspetto intensionale)
 - × le intestazioni delle tabelle
 - o l'istanza, i valori attuali, che possono cambiare anche molto rapidamente (aspetto estensionale)
 - × il "corpo" di ciascuna tabella

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

Lo schema della base di dati

Orario

Insegnamento Docente Aula Ora

L'istanza 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

Basi di dati 1 45

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)

Database



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

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 = o 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
```

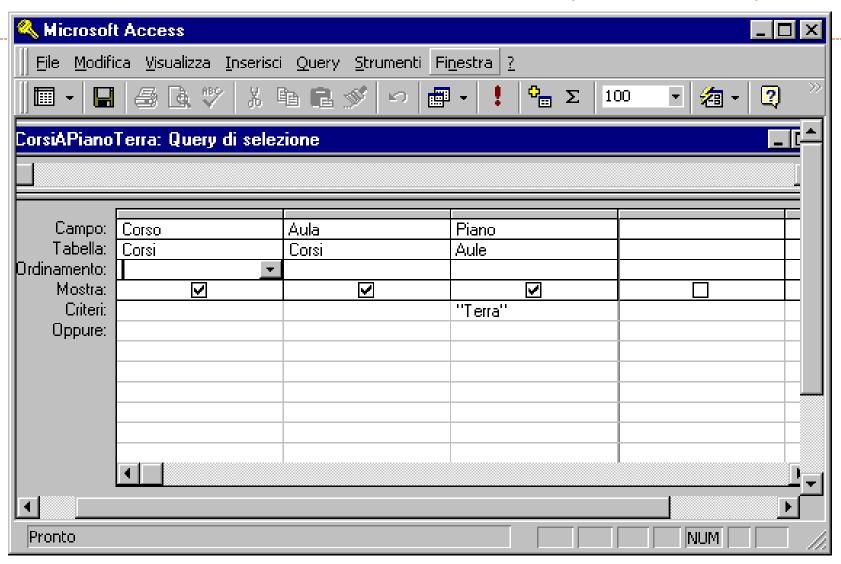


SQL immerso in linguaggio ad hoc (Oracle PL/SQL)

```
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('Non esiste la matricola specificata', sysdate);
 end;
```



Interazione non testuale (in Access)



Una distinzione terminologica (separazione fra dati e programmi)

Nei linguaggi di interrogazione di basi di dati distinguiamo tra

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))
```

Un'operazione DML (sull'istanza)

Select insegnamento

From Orario

Where docente="Mario Rossi"

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
 - o utenti finali (terminalisti): eseguono applicazioni predefinite (transazioni)
 - o utenti casuali: eseguono operazioni non previste a priori, usando linguaggi interattivi

<u>Skip</u>

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



- Le transazioni sono programmi che realizzano attività frequenti e predefinite, con poche eccezioni, previste a priori.
- Esempi:
 - o versamento presso uno presso sportello bancario
 - o emissione di certificato anagrafico
 - o dichiarazione presso l'ufficio di stato civile
 - o 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)

Basi di dati

ESERCIZI SU

BASI DI DATI E DBMS



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



• Indicare quali delle seguenti affermazioni sono vere:

- o il fatto che le basi di dati siano condivise favorisce l'efficienza dei programmi che le utilizzano
- o il fatto che le basi di dati siano condivise permette di ridurre ridondanze e inconsistenze
- o il fatto che le basi di dati siano persistenti ne garantisce l'affidabilità
- o il fatto che le basi di dati siano persistenti favorisce l'efficienza dei programmi
- o il fatto che le basi di dati siano condivise rende necessaria la gestione della privatezza e delle autorizzazioni



• Indicare quali delle seguenti affermazioni sono vere:

- o la distinzione fra DDL e DML corrisponde alla distinzione fra schema e istanza
- o le istruzioni DML permettono di interrogare la base di dati ma non di modificarla
- o le istruzioni DDL permettono di specificare la struttura della base di dati ma non di modificarla
- o 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



• Indicare quali delle seguenti affermazioni sono vere:

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