

Basi di Dati

prof. Letizia Tanca
Politecnico di Milano

Introduzione al corso

Corso di “Basi di dati”

- Titolare del corso: Prof. Letizia Tanca
 - ricevimento: v. sito Poli
 - attualmente, su appuntamento, il mercoledì alle 10:30
 - email: tanca@elet.polimi.it
 - tel: 02-2399-3531, fax: 02-2399-3411
 - Sito del corso:
<http://home.dei.polimi.it/tanca/BasiDati0708.html>
- Esercitazioni: dott. Elisa Quintarelli,
quintare@elet.polimi.it
- Laboratorio: ing. G. Orsi (una sola lezione da 4 ore)

Distribuzione indicativa dell'attività didattica

- Ore di lezione: 30
- Ore di esercitazione: 20
- Ore di laboratorio: 4
- 5 ore alla settimana non sempre tutte utilizzate, alcune usate per seminari di progetto
- Le date del laboratorio saranno pubblicate sul sito

1 ottobre 2007

Basi di dati

3

Modalità di verifica

- ***L'esame di basi di dati consiste in un'unica prova scritta.***
- ***Non sono previste prove in itinere.***

1 ottobre 2007

Basi di dati

4

Programma del corso

1. **INTRODUZIONE AL CORSO: Il sistema informativo nell'azienda; Caratteristiche dei DBMS; Utenti, strumenti e moduli di un DBMS**
2. **TEORIA RELAZIONALE: Il modello relazionale dei dati; Algebra relazionale; Interrogazioni in algebra relazionale e loro ottimizzazione; Calcolo relazionale: definizione, equivalenza fra algebra relazionale e calcolo relazionale; Datalog: interrogazioni ricorsive, vincoli di integrità, equivalenza fra tutti i linguaggi formali**
3. **LINGUAGGI PER BASI DI DATI: Introduzione a SQL; SQL come DDL: integrità referenziale, definizione di schemi e loro modifica, cataloghi; Interrogazioni SQL: semplici, con ordinamenti e raggruppamenti, complesse; Comandi di modifica e viste in SQL; Aspetti evoluti del DDL: indici e vincoli di integrità generici, controllo dell'accesso, viste e controllo dell'accesso, transazioni (cenni); SQL per la programmazione: funzioni SQL, SQL Embedded, ODBC/ADO e JDBC, procedure; Integrazione Web e Basi di Dati (Cenni)**
4. **PROGETTAZIONE DI BASI DI DATI: Fasi e astrazioni nella progettazione dei dati; Il modello Entità-Relazioni: definizione, associazioni, identificatori e gerarchie, proprietà delle gerarchie; Il ciclo di progettazione: fasi di progettazione, strategie, qualità di un progetto concettuale; Progettazione logica: fasi del progetto logico, cenni di normalizzazione; Progettazione fisica (cenni)** 5

Libri di testo

- **Libro di testo:** Paolo Atzeni, Stefano Ceri, Stefano Paraboschi, Riccardo Torlone “Basi di dati: Modelli e Linguaggi di Interrogazione”, McGraw-Hill Libri Italia, Nuova edizione (2006).
- Esiste anche la versione in inglese, degli stessi autori, titolo “Database systems”, McGraw-Hill (1999)
- **Eserciziari:**
 - Luciano Baresi, Chiara Francalanci, Fabio Schreiber, Letizia Tanca, “Progetto di dati e funzioni”, Esculapio, Bologna 2004.
 - D. Braga, M. Brambilla, A. Campi: Eserciziario di Basi di Dati, progetto leonardo, Esculapio, Bologna

Altri materiali

- Lucidi delle lezioni nella pagina del corso
- **ATTENZIONE: I LUCIDI NON SONO ESAUSTIVI!**
- Soluzioni di tutti gli esercizi del libro sul sito della versione inglese del libro <http://www.mcgraw-hill.co.uk/atzeni>
- Esercizi risolti sul sito corsionline <http://corsionline.como.polimi.it>, sotto “Basi di dati 1”
- Altre informazioni **e avvisi** nella pagina del corso
- **Altra bibliografia di approfondimento**
 - **R. Elmasri, S.B. Navathe:** *Fundamentals of Database Systems*, Benjamin-Cummings, quarta edizione
 - **R. Ramakrishnan, J. Gehrke:** *Sistemi di Basi di dati*, Mc Graw –Hill, 2004
 - **C. Batini, S. Ceri, S.B. Navathe:** *Conceptual Database Design: an Entity-Relationship Approach*, Benjamin-Cummings, 1991

L'informatica

E' la **scienza** che si occupa della
rappresentazione dell'informazione
e della sua **gestione**

L'informatica e le basi di dati

- **Informazione:** e' parte di qualsiasi attivita' umana
- **Rappresentazione:** il problema di astrarre i concetti importanti da quelli trascurabili per poter **modellare** la realta' di interesse: **ci serve studiare metodi di rappresentazione appropriati all'elaborazione da parte di una macchina digitale**
- **Gestione:** uso e trasformazione dell'informazione in maniera funzionale agli obiettivi : **programmazione**

1 ottobre 2007

Basi di dati

9

Sistema Informativo

- Elemento del sistema organizzativo dell'azienda
- Insieme di risorse umane, strumenti di elaborazione, scambio, acquisizione di informazioni, procedure per il loro trattamento, regole organizzative, *che consentono il funzionamento del sistema organizzativo*
- Evoluzione dei sistemi informativi da settoriali a integrati

1 ottobre 2007

Basi di dati

10

Sistema informativo

- Il sistema informativo organizza e gestisce le informazioni necessarie per perseguire gli scopi di un'organizzazione
Se è gestito in modo automatizzato
⇒ sistema informatico
- Il concetto di sistema informativo è indipendente dalla sua gestione automatica ed esiste da secoli (servizi anagrafici, banche)

Informazioni e dati

Nei sistemi informatici (e non solo in essi), le informazioni vengono rappresentate attraverso i **dati**.

- **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.

Base di dati

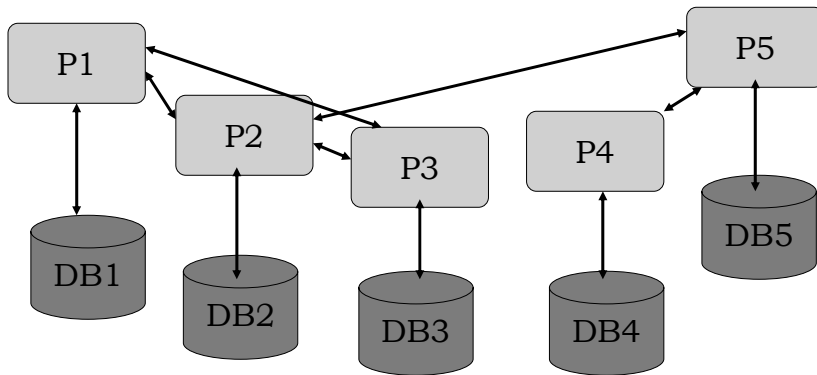
Collezione di dati, utilizzati per rappresentare le informazioni di interesse per una o più applicazioni di una organizzazione.

Particolarmente efficace se **gestita da un sistema di gestione di basi di dati (SGBD)** o, in inglese, Data Base Management System (**DBMS**)

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 (privato o porzione di un sistema più grande).
- Possono esistere sovrapposizioni fra i dati di interesse dei vari settori.

Da settoriali a integrati

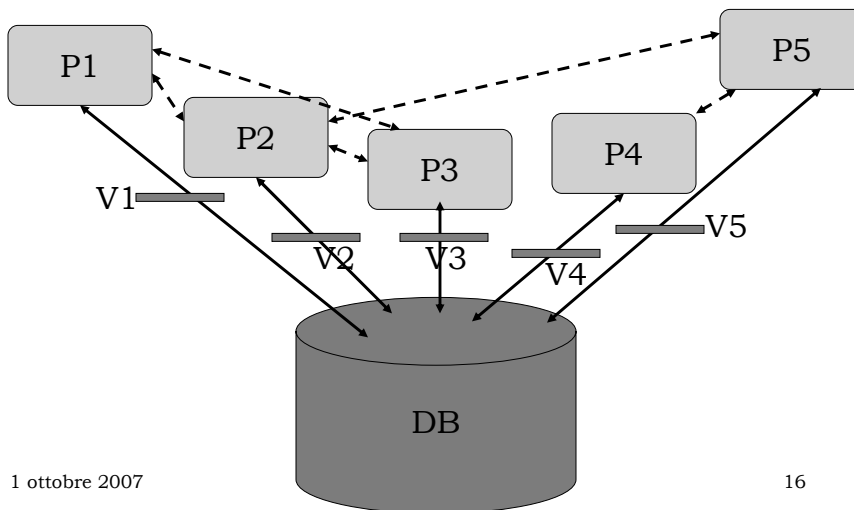


1 ottobre 2007

Basi di dati

15

Da settoriali a integrati



1 ottobre 2007

16

Una base di dati...

- è una risorsa **integrata**
- è **condivisa** tra i vari settori
- l'integrazione e la condivisione permettono di ridurre la **ridondanza**
- evitando le possibilità di **inconsistenza**
- occorrono meccanismi di definizione della **privatezza** dei dati e di limitazioni all'accesso (**autorizzazioni**).
- La condivisione richiede **controllo della concorrenza**.

1 ottobre 2007

Basi di dati

17

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.

1 ottobre 2007

Basi di dati

18

SGBD (DBMS) e file systems

Un **SGBD** (Data Base Management System = DBMS):

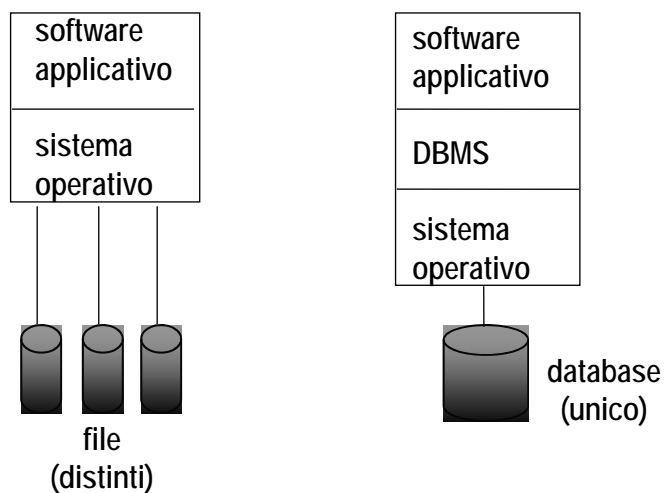
- permette di accedere in modo efficiente ai dati con una **granularità più fine che il file system**
- permette di accedere in modo diretto ai dati **basandosi sulle loro proprietà**
- effettua il controllo della concorrenza **alla granularità del singolo record**
- realizza meccanismi sofisticati per il **controllo dell'affidabilità**
- realizza meccanismi di **controllo della privacy**
- realizza **l'atomicità delle transazioni**

1 ottobre 2007

Basi di dati

19

File system e DBMS



1 ottobre 2007

Basi di dati

20

Una Base di Dati è un sistema OLTP

On Line Transaction Processing:

Elaborazione di transazioni, che realizzano i processi operativi dell'azienda-ente

- Operazioni spesso predefinite e relativamente semplici
- Ogni operazione coinvolge "pochi" dati
- Dati di dettaglio, aggiornati
- Le proprietà "acide" (Atomicità, Correttezza, Isolamento, Durabilità) delle transazioni sono essenziali

Le dimensioni delle basi di dati sono dell'ordine dei gbyte

La principale metrica di prestazione è il throughput delle transazioni (n. di transazioni al secondo)

1 ottobre 2007

Basi di dati

21

Sistemi OLAP

On Line Analytical Processing:

Elaborazione di operazioni per il supporto alle decisioni

- Operazioni complesse e casuali
- Ogni operazione può coinvolgere molti dati
- Dati aggregati, storici, anche non attualissimi
- Le proprietà "acide" non sono rilevanti, perché le operazioni sono di sola lettura

Le dimensioni del warehouse raggiungono facilmente i terabyte

Le prestazioni considerate sono il throughput delle interrogazioni e il loro tempo di risposta

1 ottobre 2007

Basi di dati

22

Descrizioni dei dati nei DBMS

- descrizioni e rappresentazioni dei dati a **livelli diversi**
- **indipendenza** dei dati dalla rappresentazione fisica
- concetto di **modello dei dati**.

1 ottobre 2007

Basi di dati

23

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)
- ad esempio, il **modello relazionale** prevede il costruttore **relazione (tabella)**, che permette di definire insiemi di record **omogenei**

1 ottobre 2007

Basi di dati

24

Tabelle: rappresentazione di relazioni

CORSI	Corso	Docente	Aula
	Basi di dati	Rossi	DS3
	Sistemi	Neri	N3
	Reti	Bruni	N3
	Controlli	Bruni	G

AULE	Nome	Edificio	Piano
	DS1	Ex-OMI	Terra
	N3	Ex-OMI	Terra
	G	Pincherle	Primo

1 ottobre 2007

Basi di dati

25

Schemi e istanze

- **schema**, descrive la struttura (aspetto **intensionale**):
nell'esempio, le **intestazioni delle tabelle**
- **istanza**, costituita dai valori attuali (aspetto **estensionale**):
nell'esempio, il "**contenuto**" di ciascuna **tabella**

1 ottobre 2007

Basi di dati

26

Esempi

- **Modello dei dati:** collezione di concetti che possono essere usati per rappresentare la realta'.
Es: alberi, grafi, tabelle
- **Schema dei dati:** rappresentazione di una specifica parte della realta', che usa un modello dei dati. **Es. albero genealogico, sistema delle strade e delle citta', tabella degli studenti e dei voti nei singoli esami**
- **Istanza:** collezione di valori dei dati che rispetta la struttura dello schema. **Es: albero genealogico della mia famiglia oggi, rete viaria della Germania il giorno 22 febbraio 2000, dati sugli studenti e gli esami in questo preciso istante**

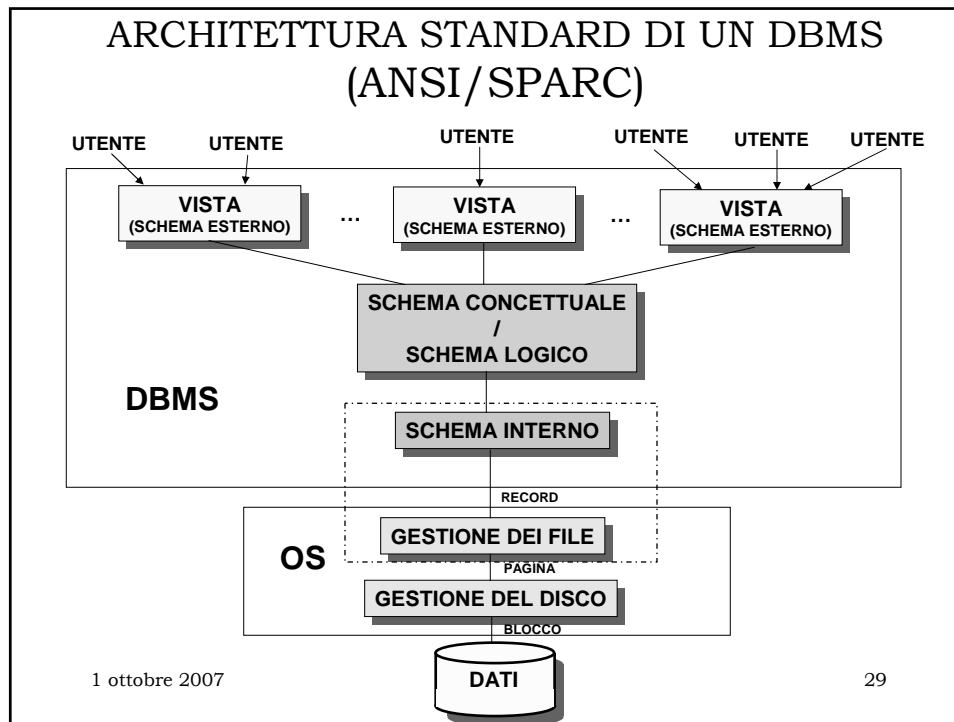
1 ottobre 2007

Basi di dati

27

Tipi (principali) di modelli

- **modelli fisici:**
 - utilizzati nei DBMS esistenti;
 - ad essi fa riferimento solo il DBMS;
 - Realizzano le strutture fisiche di memorizzazione;
esempi: **strutture indicizzate come alberi bilanciati (B-trees, B+-trees), file sequenziali, file ordinati, file organizzati con strutture hash**
- **modelli logici:**
 - utilizzati nei DBMS esistenti;
 - ad essi fanno riferimento i programmi;
 - indipendenti dalle strutture fisiche;
esempi: **relazionale, reticolare, gerarchico, a oggetti**
- **modelli concettuali:**
 - permettono di rappresentare i dati in modo indipendente da ogni sistema, cercando di descrivere i concetti del mondo reale;
 - utilizzati nelle fasi preliminari di progettazione;
il più noto è il modello **Entità-Relazione**



Architettura ANSI/SPARC: schemi

schema logico: descrizione dell'intera base di dati nel modello logico "principale" del DBMS

schema esterno: descrizione di parte della base di dati in un modello logico ("viste" parziali, derivate, anche in modelli diversi)

schema fisico: rappresentazione dello schema logico per mezzo di strutture fisiche di memorizzazione.

1 ottobre 2007 Basi di dati 30

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	Ex-OMI	Terra
N3	Ex-OMI	Terra
G	Pincherle	Primo

CorsiSedi

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

1 ottobre 2007

Basi di dati

31

Indipendenza dei dati

L'accesso avviene solo tramite il livello esterno (che può coincidere con il livello logico)

- **indipendenza fisica**
- **indipendenza logica**
 - aggiunte o modifiche alle viste non richiedono modifiche al livello logico;
 - modifiche allo schema logico che lascino inalterato lo schema esterno sono trasparenti

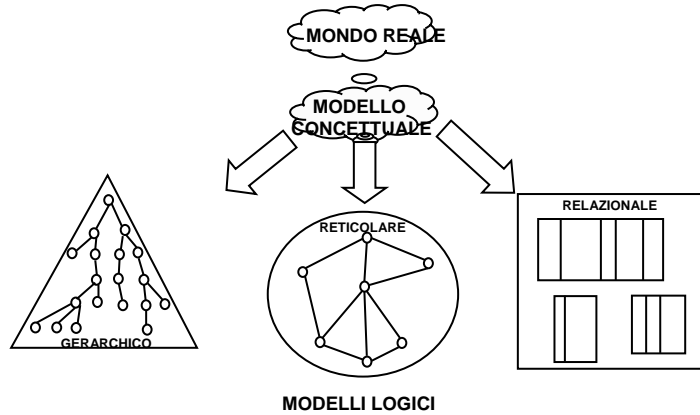
1 ottobre 2007

Basi di dati

32

Modelli dei dati e progettazione

i modelli costituiscono una strutturazione semplificata della realtà che ne accoglie aspetti specifici e aiuta a comprenderla meglio



1 ottobre 2007

Basi di dati

33

Modelli dei dati e progettazione

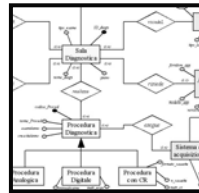
• Fatture

• Ordini dei clienti

• Indirizzi dei clienti

• Archivio prodotti in magazzino

Requisiti



Schema concettuale (COSA)

Fatture

Ordini

Clienti

Magazzino

Schema logico (COME)

1 ottobre 2007

Basi di dati

34

Linguaggi di interrogazione

- Permettono di trovare un dato basandosi sulle sue proprietà.
- **Es:** tabella STUDENTI--> *trovare tutti gli studenti che hanno preso più di 28 a Fondamenti di Informatica*
- Permettono di trovare dati basandosi su confronti tra i contenuti di più tabelle.
- **Es:** tabella ELEZIONI, tabella PRESIDENTI --> *Trovare gli anni in cui è stato eletto un presidente repubblicano proveniente dall'Illinois.*

1 ottobre 2007

Basi di dati

35

Linguaggi per basi di dati

L'accesso ai dati può avvenire mediante:

1. linguaggi **testuali interattivi** (es. SQL)
2. comandi (come quelli del linguaggio interattivo) **immersi in un linguaggio ospite** (Pascal, C, Cobol, etc.)
3. 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)
4. **interfacce amichevoli** (es. linguaggi grafici)

1 ottobre 2007

Basi di dati

36

SQL, un linguaggio interattivo

```
SELECT Corso, Aula, Piano
FROM Aule, Corsi
WHERE Nome = Aula
AND Piano="Terra"
```

Corso	Aula	Piano
Reti	N3	Terra
Sistemi	N3	Terra

1 ottobre 2007

Basi di dati

37

SQL immerso in Pascal (o altro 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
```

1 ottobre 2007

Basi di dati

38

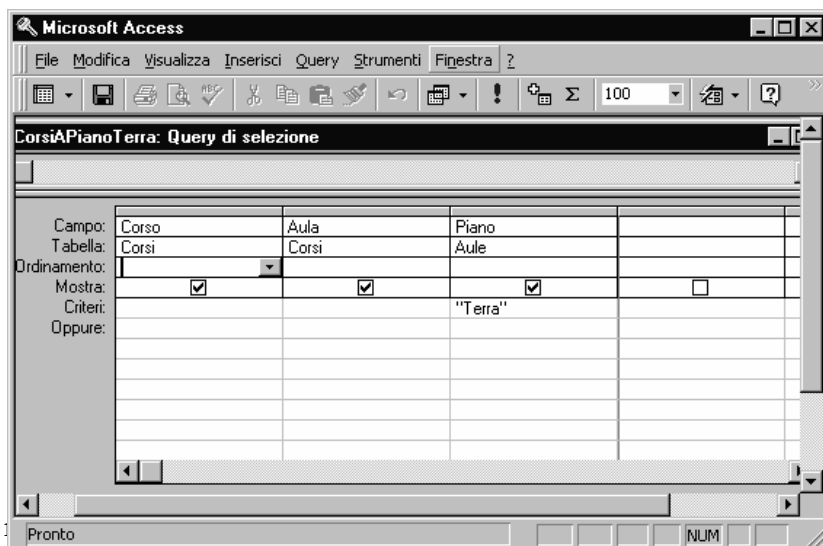
SQL immerso in linguaggio ad hoc (Oracle PL/SQL)

```

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); 39
end;

```

Access: query by example



Transazione

- Una **transazione**:
 - Rappresenta la tipica unità di lavoro elementare di un Database Server
 - Identifica una unità di lavoro svolta da una applicazione, cui si vogliono associare particolari caratteristiche di correttezza, robustezza e isolamento
- Tutti i DBMS (distribuiti) sono **sistemi transazionali**: mettono a disposizione un meccanismo per la definizione e l'esecuzione di transazioni
- Nell'esecuzione di una transazione devono essere garantite le proprietà ACIDE

1 ottobre 2007

Basi di dati

41

Proprietà ACIDE

- **Atomicità**: Una transazione è una unità indivisibile di esecuzione
- **Consistenza**: l'esecuzione di una transazione non deve violare i vincoli di integrità definiti sulla base di dati
- **Isolamento**: l'esecuzione di una transazione non è influenzata dalla esecuzione di altre transazioni concorrenti
- **Persistenza (Durability)**: gli effetti di una transazione andata a buon fine devono essere permanenti

1 ottobre 2007

Basi di dati

42

Componenti di un database server

- **Ottimizzatore** - seleziona la strategia d'accesso ai dati
- **Access Methods Manager** - esegue la strategia
 - RSS (Relational Storage System)
 - OM (Object Manager)
- **Buffer Manager** - gestisce gli accessi alle pagine
- **Gestore di affidabilità** - gestisce i malfunzionamenti
- **Controllo di concorrenza** - gestisce le interferenze dovute alla concorrenza nell'accesso ai dati

1 ottobre 2007

Basi di dati

43

Architettura e modello di esecuzione **single tier**

Mainframe + terminali intelligenti

- *Pro*: Facilmente gestibile da un amministratore centrale
- *Contro*: Interfaccia grafica richiede potenza di calcolo, sottratta alla gestione del BD



1 ottobre 2007

44

Architetture distribuite

- Quasi tutte le basi di dati hanno **un'architettura distribuita** (caratterizzata dalla presenza di una rete)
- **Architettura client-server**
 - E' la più semplice e diffusa
 - **Server**: elaboratore per la gestione dei dati
 - **Client**: elaboratore per la gestione del dialogo con l'utente
 - **Rete**: collega i client con il server
- **Architettura distribuita**
 - Almeno **DUE server** che, oltre a lavorare autonomamente, in alcuni casi devono **interagire**

1 ottobre 2007

Basi di dati

45

Architettura client-server

(1)

- Più in generale, modello di interazione **tra processi software** in cui i processi si dividono in Client e Server
- **Processi Client**
 - **Richiedono i servizi**
 - **Dedicati a interagire con l'utente finale**
 - **Ruolo attivo: genera richieste**
- **Processi Server**
 - **Offrono i servizi**
 - **Ruolo reattivo: si limita a rispondere alle richieste dei diversi client**
- **Nota Si parla di processi, generalmente i processi client e server risiedono su macchine diverse, collegate via rete**
- **L'interazione fra client e server richiede una interfaccia di servizi: elenco dei servizi messi a disposizione dal server**

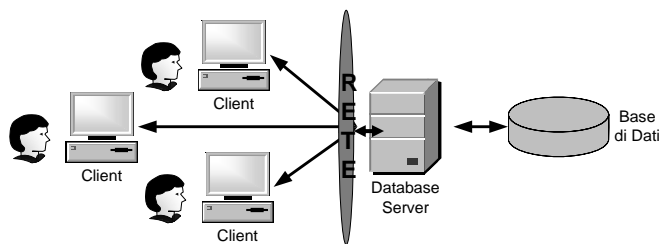
1 ottobre 2007

Basi di dati

46

Architettura client-server (2)

- **Un processo client può richiedere alcuni (pochi) servizi a vari processi server (se sono più di uno è una architettura distribuita)**
- **Ogni processo server risponde a (molte) richieste da parte di molti processi client gestendo in modo opportuno le relative transazioni**



1 ottobre 2007

Basi di dati

47

Motivi uso Client-Server

- **Le funzioni di client e server sono ben definite**
 - **C'è una corrispondenza con gli utenti**
 - **Programmatore applicativo:** ha la responsabilità di gestire il software relativo al client
 - **Amministratore:** Deve organizzare la base dati sul server per garantire prestazioni ottimali a tutti i client
- **Diverse esigenze di hardware**
 - **Il client è un elaboratore adatto alla interazione con l'utente**
 - **Strumenti di produttività**
 - **Applicazioni "amichevoli" che accedono alla base dati**
 - **Grafica: non possiamo più usare terminali non intelligenti**
 - **Il server è dimensionato in base ai servizi che deve offrire e al carico transazionale (grande memoria centrale, grande memoria di massa, ...)**

1 ottobre 2007

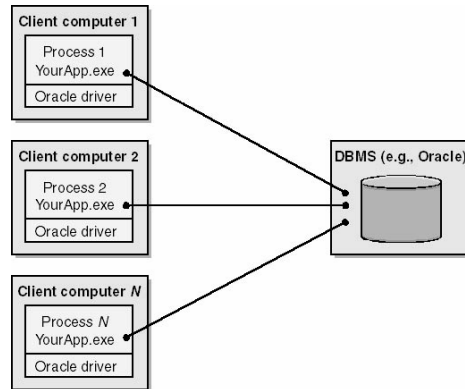
Basi di dati

48

Architettura a 2 livelli

(two tier)

- Il client ha funzioni solo di interfaccia o anche di gestione dell'applicazione
- Il server ha funzione di gestione dei dati
- Più diffusa l'architettura *thin client*

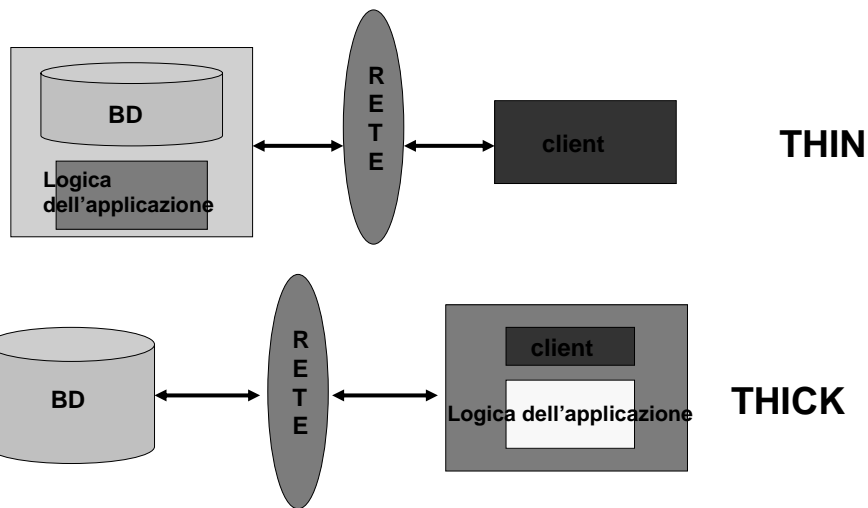


1 ottobre 2007

Basi di dati

49

Thin client e Thick client



1 ottobre 2007

Basi di dati

50

Svantaggi thick client

- Non c'è un posto centralizzato dove aggiornare e mantenere la logica dell'applicazione
- È necessaria **fiducia** tra il server e i client
 - Un modo per risolvere il problema: **stored procedures**
- **Non scalabile** (non più di poche centinaia di client), poiché:
 - richiede altissima capacità di elaborazione da parte del server, che deve rispondere a un elevato numero di interrogazioni
 - se si aggiungono ulteriori server, o DBMS, troppe connessioni aperte

1 ottobre 2007

Basi di dati

51

Strumenti di sviluppo client server

- **Microsoft Visual Basic**
- **Sybase Powerbuilder**
- **etc.**
- **Soprattutto usati per progettare nella versione *thin client***

1 ottobre 2007

Basi di dati

52

Architettura a 3 livelli

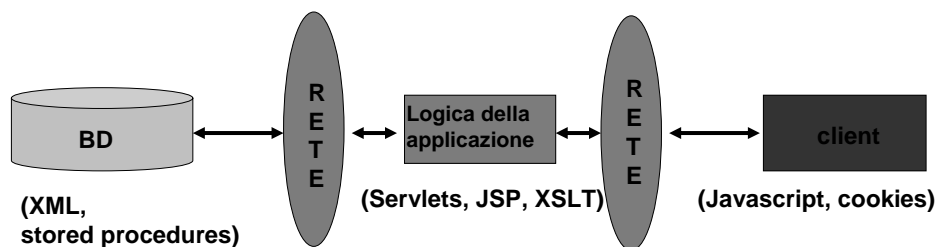
- **Nell'architettura three-tier è presente un secondo server, il server applicativo, responsabile di gestire la logica applicativa comune a più client**
 - **Il client è più semplice e si occupa solo dell'interfacciamento con l'utente finale (thin client)**
 - Il client invia le richieste al server applicativo
 - Il server applicativo dialoga con il server per la gestione dei dati
 - **Questa architettura è predisposta per Internet**
- CLIENT = BROWSER WEB**

1 ottobre 2007

Basi di dati

53

Architetture three tier



1 ottobre 2007

Basi di dati

54

Vantaggi three tier

- Si possono interconnettere **sistemi eterogenei**
- I client sono **thin** (browser web)
- **Integrazione** dei dati gestita nel livello intermedio (resta il problema metodologico del “*come??*”)
- **Scalabilità** sul numero di client (si possono appropriatamente moltiplicare anche le macchine a livello intermedio, impossibile nel caso thick client)
- **Riuso**

1 ottobre 2007

Basi di dati

55

Personaggi del mondo delle BD

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

1 ottobre 2007

Basi di dati

56

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 scorparabilità (spesso) delle funzionalità (con riduzione di efficienza)**

1 ottobre 2007

Basi di dati

57