

Corso di Informatica

Modulo T2

A2 - Introduzione ai database

M.Malatesta A2 - Introduzione ai database-19

1
12/01/2012

Prerequisiti

- Concetto di sistema
- File system
- Archivi
- File e record

M.Malatesta A2 - Introduzione ai database-19

2
12/01/2012

Introduzione

Nella gestione di una attività, ad esempio un'azienda, la mole di informazioni da gestire è rilevante. Perciò possiamo chiederci:

Quali strumenti esistono, più evoluti rispetto agli archivi, per trattare queste situazioni ?

Come con questi strumenti si garantisce l'efficiente, razionale e sicuro trattamento dei dati ?

In questa Unità e nelle successive affrontiamo lo studio di questi strumenti, detti **DBMS** che rispondono in modo esauriente alle domande poste.

Informazioni generali

Fino agli anni '70 circa, le attività aziendali erano automatizzate con programmi di gestione di dati *mediante archivi*.

Successivamente, lo studio di *modelli matematici* ha reso possibile la nascita di tecnologie informatiche note con il nome di **database** che, rispetto alla gestione basata su archivi di una attività aziendale, offrono grandi vantaggi in termini di **efficienza** ed **affidabilità**.

In questa Unità introduciamo il concetto di **Base di Dati** (*database*) e mostriamo come questo sia una naturale evoluzione della superata gestione ad archivi.

La gestione ad archivi

La gestione di grandi insiemi di dati è possibile attraverso sistemi semplici come gli ordinari *file system* dei sistemi operativi, che prevedono strumenti per effettuare:

- *condivisioni di dati*
- *operazioni sui dati*

anche se in modo *molto elementare* (a basso livello);

M.Malatesta A2 - Introduzione ai database-19

5
12/01/2012

La gestione ad archivi

La gestione ad archivi però presenta numerosi problemi. Il tracciato record seguente mostra la gestione mediante archivi dei conti correnti bancari di una *banca centrale* avente diverse filiali.

Sede centrale della banca

N° Conto	Nominativo	Indirizzo	Città	Cod. filiale	Dati filiale	Descrizione
1020	Rossi Mario	v. Roma, 34	Roma	10	v. Capri-Roma	Privato
1021	Carli Anna	v. Rho,27	Roma	10	v. Capri-Roma	Privato
1022	Soc.“Acque”	v.le Marconi,	Napoli	22	p.zza Brasile-Napoli	Società
1135	Regione Molise	p.zza Adua, 7	Milano	30	C.so Sempione-Milano	Pubblico
1125	Telecom	v. Rossimi	Bergamo	55	v. Einaudi-Bergamo	Società

M.Malatesta A2 - Introduzione ai database-19

6
12/01/2012

La gestione ad archivi

Nelle *single filiali* devono inoltre essere registrati i movimenti dei singoli clienti. Ad esempio per la **filiale 10 di Roma** si avrà:

Filiale 10 di Roma

N° Cliente	Nominativo	Indirizzo	Cod. Filiale	Data	Descrizione	Importo	Saldo
1020	Rossi Mario	v. Roma, 35-Roma	10	10-07-2002	Versamento	100	10000
1021	Carli Anna	v. Rho,27-Roma	10	08-08-2002	Prelievo	150	20500
1020	Rossi Mario	v. Roma, 35-Roma	10	09-08-2002	Prelievo	50	9950
1121	Carli Anna	v. Rho,27-Roma	10	12-08-2002	Versamento	200	20700
1120	Rossi Mario	v. Roma, 35-Roma	10	13-08-2002	Prelievo	50	9900

M.Malatesta A2 - Introduzione ai database-19

7
12/01/2012

La gestione ad archivi

Cosa possiamo osservare su n°Conto e N° Cliente?

- sono campi che pur avendo diverso nome, rappresentano lo stesso dato

Cosa possiamo osservare su Indirizzo e Nominativo?

- i campi **Indirizzo e Nominativo** sono duplicati con conseguente spreco di spazio su disco

Cosa possiamo osservare sul campo Descrizione?

- il campo **Descrizione** rappresenta nel primo la categoria del cliente (*Privato, Società, Pubblico*), mentre nel secondo rappresenta il tipo di operazione (*Prelievo, Versamento*).

Come si possono risolvere questi problemi?

M.Malatesta A2 - Introduzione ai database-19

8
12/01/2012

La gestione ad archivi

In molte situazioni, i problemi sui dati si risolvono attribuendo **specifici codici** come indicato in figura ed utilizzando più file.

N° Conto	Nominativo	Indirizzo	Città	Cod. filiale	Categoria
1020	Rossi Mario	v. Roma, 34	Roma	10	Privato
1021	Carli Anna	v. Rho, 27	Roma	10	Privato
1022	Soc. "Acque"	v. le Marconi,	Napoli	22	Società
1135	Regione Molise	p.zza Adua, 7	Milano	30	Pubblico
1125	Telecom	v. Rossini	Bergamo	55	Società

In questo modo si evitano le **ridondanze** (informazioni ripetute)

Esempio: in quale filiale è stato eseguito il movimento **13**?

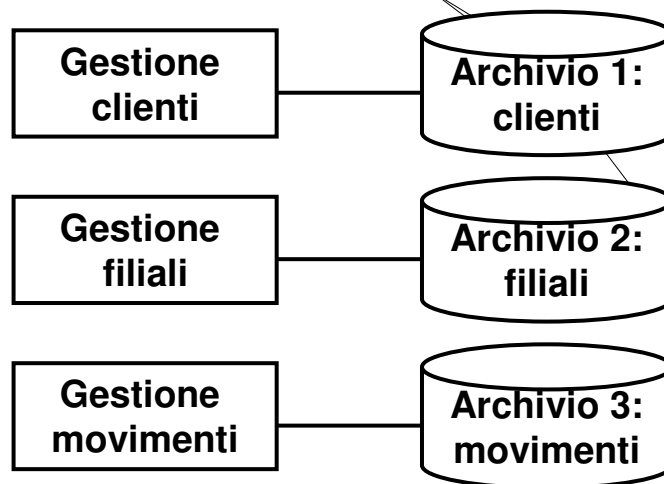
Cod. Filiale	Dati Filiale
10	v. Capri-Roma
22	p.zza Brasile-Napoli
30	C.so Sempione-Milano
55	v. Einaudi-Bergamo

N° Movim.	N° Cliente	Data	Descrizione	Importo	Saldo
10	1020	10-07-2002	Versamento	100	10000
11	1021	08-08-2002	Prelievo	150	20500
12	1020	09-08-2002	Prelievo	50	9950
13	1021	12-08-2002	Versamento	200	20700
14	1020	13-08-2002	Prelievo	50	9900

M.Malatesta A2 - Introduzione ai database-19

9
12/01/2012

La gestione ad archivi



M.Malatesta A2 - Introduzione ai database-19

10
12/01/2012

Problemi nella gestione ad archivi

Tuttavia, restano ancora molti problemi nel gestire i dati mediante archivi, alcuni dei quali sono riassunti di seguito:

- rischio di **ridondanza** (informazioni ripetute)
- necessità di **apposite routine** di codice per la manipolazione dei dati
- **elevato numero di file** che rende pesante l'elaborazione;
- codice del programma sorgente risulta **prolisso e rigido**;
- **inconsistenza**: la modifica di un dato in un archivio, può comportare modifiche in altri archivi, altrimenti si corre il rischio di avere archivi con dati non compatibili tra loro.

In sostanza, occorre un sistema che consenta lo sviluppo del software (gestione di **dati e programmi**) in base ai seguenti requisiti.

Requisiti dei dati

I dati devono essere gestiti correttamente (protetti e mantenuti riservati); inoltre devono essere razionalmente progettati per evitare qualunque malfunzionamento nella loro gestione

Garantire	Evitare
<u>Integrità</u> Protezione da malfunzionamento dei programmi	<u>Ridondanza</u> Uno stesso dato è presente più volte nei vari archivi
<u>Sicurezza</u> Protezione da accessi non autorizzati che possono compromettere l'integrità dei dati	<u>Incongruenza</u> A causa della ridondanza, uno stesso campo presente in più archivi ha contenuto diverso
<u>Riservatezza</u> Protezione della privacy, anche nei diversi settori dell'azienda	<u>Inconsistenza</u> A causa dell'incongruenza, il dato non è affidabile, ossia si perde il dato vero aggiornato

Requisiti dei programmi

I programmi devono essere gestibili in multiutenza, facili da usare, facili da realizzare e flessibili, ossia il più possibile indipendenti dai dati

Garantire	Evitare
<p><u>Multiutenza</u> Accesso contemporaneo di più utenti</p>	<p><u>Dipendenza dalla struttura logica dei dati</u> Un aggiornamento di un tracciato record (ad esempio aggiunta, modifica o eliminazione di un campo) comporta un aggiornamento di tutti i programmi connessi (correzione, compilazione, messa a punto)</p>
<p><u>Facilità di uso</u> La presenza di interfacce amichevoli rende <i>friendly</i> l'ambiente</p>	<p><u>Dipendenza dalla struttura fisica dei dati</u> Modifiche alle tecniche di accesso, di organizzazione e di collocazione degli archivi comportano aggiornamento dei programmi di gestione.</p>
<p><u>Facilità di realizzazione</u> Mediante appositi linguaggi (SQL)</p>	

M.Malatesta A2 - Introduzione ai database-19

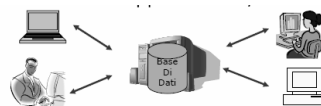
13
12/01/2012

Base di dati (DB)

Una **Base di dati (DB)** è un *insieme integrato di dati* utilizzati per lo svolgimento delle attività di un ente (azienda, ufficio, persona, organizzazione).

INTEGRAZIONE DEI DATI

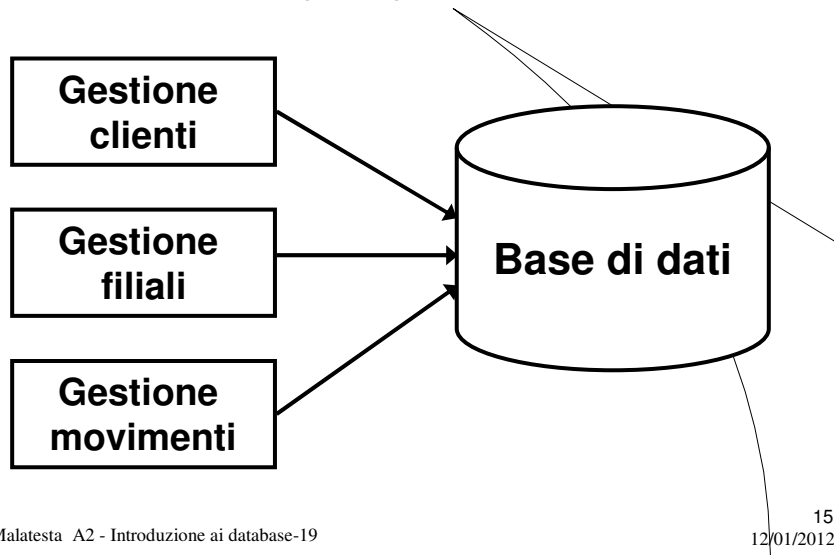
- i dati presenti negli archivi sono memorizzati senza ridondanze, sotto forma di **tabelle**.
- i dati sono strutturati in modo più semplice rispetto agli archivi;
- la memoria di massa è occupata in modo razionale ed efficiente;
- i programmi applicativi sono più semplici
- si evitano eventuali inconsistenze.



M.Malatesta A2 - Introduzione ai database-19

14
12/01/2012

Base di dati (DB)



M.Malatesta A2 - Introduzione ai database-19

Caratteristiche di un DB

- **Monoutente**
 - Utilizzo: un solo utente
 - Semplice
 - Piccole dimensioni
- **Multiutente**
 - Utilizzo: più utenti contemporaneamente
 - Complesso, grandi dimensioni
 - Rete (server e workstation)

M.Malatesta A2 - Introduzione ai database-19

16
12/01/2012

Caratteristiche di un DB

- **Centralizzato**
 - Residente su di un unico server
- **Distribuito**
 - Residente su di una rete di calcolatori (o insieme di database residenti su sistema distribuito)

Caratteristiche di un DB

	Centralizzato	Distribuito
Funzionamento	Unico DB per ogni query viene consultato il DB globale.	Un DB globale; molti DB locali: 1) query su DB locale; 2) replicazione dei dati; 3) aggiornamento DB globale.
Limiti	<ul style="list-style-type: none"> ● Elevati Costi Hardware ● Unico punto d'accesso → maggiori tempi d'attesa ● Minore affidabilità 	<ul style="list-style-type: none"> ● Elevati Costi Software ● Rete → maggiore difficoltà di protezione ● Maggiore complessità
Vantaggi	<ul style="list-style-type: none"> ● Unico punto d'accesso → Migliore protezione degli accessi ● Memoria di massa → minore difficoltà di protezione 	<ul style="list-style-type: none"> ● Migliore aderenza del DB al sistema informativo ● Autonomia locale dei dati: maggior controllo ● Rete → Maggiore velocità di accesso ● Maggiore affidabilità

I DBMS

I **DBMS** (*Data Base Management System*) sono **applicazioni complesse** che consentono in modo **semplice** (astratto) ma **efficiente** di operare su un **DB**, *estendendo le funzionalità dei file system* e fornendo più servizi ed in maniera integrata.

Esempi di prodotti software **DBMS** disponibili sul mercato sono:

- Access
- DB2
- Oracle
- Informix
- Sybase
- SQLServer
- mySql

I DBMS

In particolare, un DBMS è costituito da un insieme di programmi che consentono:

- una semplice *rappresentazione dei dati* (record e campi)
- *razionale gestione dei dati* su memoria di massa (memorizzazione e reperimento)
- un *controllo sugli accessi* ai dati da parte degli utenti, che consente sicurezza e integrità del database
- una agevole gestione dei dati anche in *multiutenza*;
- un facile *collaudo e manutenzione*
- una semplice *interfaccia* tra utenti e il DB

I DBMS

Un DBMS fornisce linguaggi come:

- **DDL** (*Data Definition Language*), linguaggio che consente la definizione dei dati (ad esempio creare tabelle, modificare il nome degli attributi, il loro tipo e la loro dimensione)
- **DML** (*Data Manipulation Language*), linguaggio mediante il quale è possibile operare sui dati (aggiornamenti e ricerche)
- **DCL** (*Data Control Language*) □ istruzioni che ci permettono di controllare gli accessi ed i diritti (permessi) dei vari utenti

Inoltre prevede una descrizione centralizzata dei dati detta **catalogo** o **dizionario** che, che può essere utilizzata dai vari programmi; in questo modo tutti i termini e gli elementi utilizzati sono considerati standard per l'applicazione.

Caratteristiche di un DBMS

Gestisce collezioni di dati...

...grandi

- le dimensioni di un DB sono in genere molto maggiori (anche migliaia di Gb) rispetto alla memoria centrale dei sistemi di calcolo utilizzati
- il limite deve essere solo quello fisico dei dispositivi

...persistenti

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

...condivise

- meccanismi di **autorizzazione** in caso di attività diverse su dati condivisi
- controllo della **concorrenza**, in caso di accesso di più utenti ai dati condivisi

Caratteristiche di un DBMS

Conferisce garanzie di...

...**privatezza** (autorizzazione)

- l'utente A può leggere tutti i dati, ma può modificare solo i dati di un insieme X
- l'utente B può leggere tutti i dati dell'insieme X, ma può modificare solo quelli di un insieme Y

...**affidabilità del software**

- resistenza a guasti hw e sw durante le *transazioni*
- funzioni di backup e recovery

Le **transazioni** sono sequenze di operazioni da considerare indivisibili e che coinvolgono più insiemi di dati, come aggiornamenti, inserimenti, ricerche di dati. La sequenza di operazioni sulla base di dati viene eseguita **per intero** o **per niente** (ad es. trasferimento di fondi da un conto A ad un conto B).

Caratteristiche di un DBMS

Conferisce garanzie di...

...**efficienza**

- l'ottimale utilizzo delle risorse:
 - spazio di memoria (principale e secondaria);
 - tempo (di esecuzione e di risposta)

...**efficacia**

- capaci di offrire molteplici servizi (definizione, inserimento, consultazione, aggiornamento, ecc... dei dati), grazie a funzionalità articolate, potenti, flessibili e semplici

Indipendenza dei dati

Una delle caratteristiche più importanti di un **DBMS** è quella di essere una potente ma semplice interfaccia tra l'utente e i dati, cosa che offre le seguenti proprietà, accennate in precedenza:

- **indipendenza fisica**
- **indipendenza logica**

Modifiche agli archivi fisici (tecniche di accesso, di organizzazione e collocazione) non comportano aggiornamenti dei programmi di gestione.

Aggiornamenti del tracciato record non comportano un aggiornamento dei programmi di gestione.

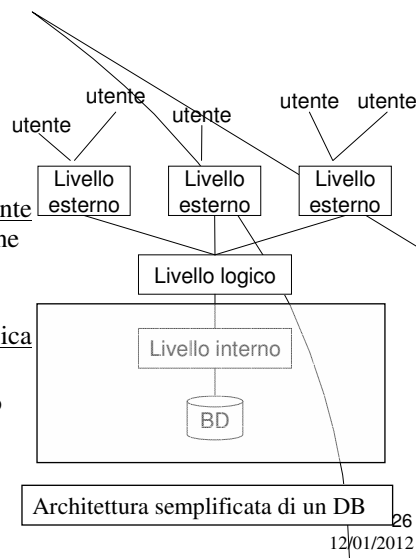
dei dati, *rispetto ai programmi* per la loro gestione.

Indipendenza fisica dei dati

Il **livello logico** (e quello esterno) sono **indipendenti da quello fisico**

Ciò significa che:

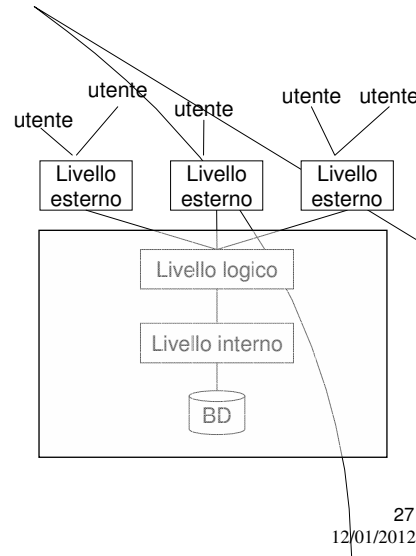
- la gestione dei dati sui supporti (aggiunta, eliminazione, modifica) è curata direttamente dal file system, che gestisce organizzazione ed accesso ai file coinvolti
- i dati sono utilizzati nello stesso modo qualunque sia la loro implementazione fisica (memorizzazione dati sui supporti)
- l'implementazione fisica delle tabelle può cambiare senza che debbano essere modificati i programmi che le gestiscono



Indipendenza logica dei dati

Il **livello esterno** è indipendente da quello logico

- aggiunte o modifiche alle **viste** (livello esterno) non richiedono modifiche al **livello logico** (tabelle)
- modifiche al **livello logico** che lascino inalterato il livello esterno sono trasparenti all'utente



M.Malatesta A2 - Introduzione ai database-19

Argomenti

- Informazioni generali
- La gestione ad archivi
- Problemi nella gestione ad archivi
- Requisiti dei dati
- Requisiti dei programmi
- Base di dati (DB)
- Caratteristiche di un DB
- I DBMS
- Esempi di DBMS
- Caratteristiche di un DBMS

M.Malatesta A2 - Introduzione ai database-19

28
12/01/2012

Altre fonti di informazione

- Atzeni, Ceri, Paraboschi, Torlone Basi di dati- McGraw-Hill, 2002
- A.Lorenzi-D.Rossi, Le basi di dati e il linguaggio SQL - ed. ATLAS