

Gestione dati e archivi

Barbara Carminati
barbara.carminati@uninsubria.it



Aspetti Introduttivi

1

Introduzione al corso

- Barbara Carminati, email:
barbara.carminati@uninsubria.it
- Ricevimento: in coda alle lezioni, o concordato tramite email
- Materiale didattico e informazioni:
<http://elearning.uninsubria.it>



2

Orari lezioni



- Mercoledì:
 - 14/15.30, aula xxx
- Giovedì:
 - 14/15.30, aula xxx

3

Obiettivo del corso



- Conoscenza dei principi fondamentali che stanno alla base dei principali strumenti utilizzati per la creazione e gestione di una base di dati

4

Organizzazione del corso



- Lezioni frontali ed esercitazioni: 36 ore
- Laboratorio (in via Ravasi): 12 ore

5

Programma del corso



- Basi di dati: aspetti introduttivi
- Modello relazionale
- Il linguaggio SQL
- Progettazione di basi di dati

- Laboratorio: MS access

6

Libri di testo



- Libro di testo: B. Catania, E. Ferrari, e G. Guerrini. *Sistemi di Basi di Dati: Concetti e Architetture*, De Agostini, 2007.
- Lucidi online su <http://elearning.uninsubria.it>

7

Modalità d'esame



- Compitini durante il corso:
 - 1° compitino circa a metà-fine aprile:
 - Parte introduttiva (domande teoriche)
 - Esercizi su progettazione di basi dati
 - 2° compitino a fine corso (fine maggio- inizio giugno):
 - Esercizi SQL
- Appelli

8

Aspetti Introduttivi



Aspetti Introduttivi

9

Organizzazione/Azienda



- Utilizza un insieme di **risorse** e **regole** per lo svolgimento coordinato delle attività al fine del perseguimento degli scopi
- le risorse di una azienda (o amministrazione):
 - persone
 - denaro
 - materiali
 - **informazioni**

10

Sistema informativo



- Una delle principali esigenze di ogni organizzazione e' quella di gestire e rendere disponibili le informazioni



Il sistema preposto a tale compito prende il nome di **sistema informativo**

11

Sistema informativo



- Componente di una organizzazione che gestisce le informazioni di interesse
 - Produce, acquisisce, elabora, conserva e distribuisce le informazioni
- E` costituito da strumenti, procedure e strutture, sia automatizzate sia manuali
- La sua definizione e` del tutto indipendente dal grado di automazione in essere

12

Sistema Informatico



- Un sistema informativo non presuppone necessariamente l'ausilio di un supporto informatico
- **Sistema informatico:** parte del sistema informativo che gestisce l'informazione mediante l'ausilio di strumenti e tecnologie informatiche
- Spesso *sistema informatico* viene usato come sinonimo di *sistema informativo*

13

Componenti di un sistema informativo



- Strumenti per la rappresentazione dell'**informazione**, mediante una qualche codifica
- Programmi applicativi e di sistema che, operando su tale rappresentazione, realizzano tutte le funzioni necessarie alla gestione delle informazioni

14

Cosa sono le informazioni?



Tutto ciò che produce variazioni nel patrimonio conoscitivo di un soggetto detto percettore dell'informazione



L'informazione deve essere utile per il percettore dell'informazione



Un sistema informativo deve fornire una chiave di lettura mediante cui interpretare l'informazione che gestisce

15

I dati



- Nei sistemi informatici, le informazioni sono rappresentate sotto forma di dati
- **Dato**: registrazione della descrizione di una qualsiasi caratteristica del dominio di interesse su un supporto che ne garantisca la conservazione e, mediante un insieme di simboli, ne garantisca la **comprensibilità** e la **reperibilità**

16



I dati

- I dati hanno bisogno di un contesto interpretativo che permetta di estrarre da essi le informazioni di interesse per gli utenti
 - Se considero il dato rappresentato dal numero 4, tale dato non fornisce alcuna informazione
 - Viceversa, dire che 4 è il numero di film noleggiati da Anna Rossi nell'ultimo mese fornisce un'informazione



Uno degli obiettivi fondamentali di un sistema informativo è fornire tale contesto interpretativo ai dati

17



Base di dati

- **Base di dati:**
 - collezione di dati tra loro correlati, utilizzati per rappresentare le informazioni di interesse in un sistema informativo
- **Sistema di gestione di basi di dati** (DBMS – Data Base Management System):
 - sistema software, centralizzato o distribuito, che fornisce gli strumenti necessari a gestire le informazioni



- **Base di dati:** collezione di dati gestita da un DBMS

18

Dai SO ai DBMS



- A partire dagli anni '60, i DBMS sono diventati una componente essenziale nella realizzazione di qualsiasi sistema informativo
 - I primi sistemi informativi erano basati sull'uso di archivi separati, gestiti dal sistema operativo, utilizzando direttamente i servizi del *file system*
 - Si è passati in seguito ad un approccio in cui i dati vengono organizzati in un unico insieme logicamente integrato, la base di dati, gestito dal DBMS ed in grado di soddisfare il fabbisogno informativo di tutte le applicazioni

19

Un esempio per capire: la videoteca



- Una videoteca vuole mantenere informazioni relative ai propri clienti ed ai noleggi che questi hanno effettuato
- Le applicazioni usano direttamente i servizi del file system per la memorizzazione e l'accesso a tali dati
- In base a tale approccio, i dati relativi ai clienti ed ai noleggi sono mantenuti in record memorizzati in vari file su memoria secondaria
- In aggiunta ai file, esiste un insieme di programmi applicativi:
 - un programma di modifica della residenza di un dato cliente
 - un programma per l'inserimento di un nuovo noleggio
 - un programma per l'inserimento e la cancellazione di un cliente
 - un programma che stampa la lista di tutti i clienti della videoteca in ordine alfabetico

20

Ridondanza ed inconsistenza



- I file di dati ed i programmi sono creati in tempi diversi da progettisti software e programmatori diversi
- Non esiste una descrizione ad alto livello e centralizzata dei dati (replicazione)
- Determinare se ed in che file una certa informazione è memorizzata è estremamente difficile



- La presenza dello stesso dato in file diversi è detta **ridondanza** e può causare alti costi di memorizzazione ed inconsistenze nei dati

21

Ridondanza ed inconsistenza



- Nella videoteca:
 - Il nome, il cognome ed il numero di tessera di un cliente possono essere memorizzati sia nel file che contiene le informazioni sui clienti sia nel file che contiene le informazioni sui noleggi
 - Se il numero di tessera di un cliente viene modificato, a seguito dello smarrimento della vecchia tessera, è necessario riportare tale cambiamento non solo nel file dei clienti ma anche in tutti i record relativi ai noleggi effettuati dal cliente in questione

22

Ridondanza ed inconsistenza




- Nel corso vedremo quali strumenti gli attuali DBMS forniscono per limitare inconsistenze e ridondanze dei dati:
 - a livello di rappresentazione dei dati
 - a livello di strumenti di progettazione della base di dati

23

Difficolta` nell'accesso ai dati



- La mancanza di una descrizione ad alto livello e centralizzata dei dati ne rende estremamente difficoltoso l'utilizzo al fine di rispondere a nuove esigenze applicative
- 
- Nuove esigenze applicative possono richiedere lo sviluppo di nuovi programmi applicativi o la modifica di quelli esistenti

24

Difficolta` nell'accesso ai dati



- Nella videoteca:
 - supponiamo che sia necessario stampare periodicamente la lista dei clienti VIP ordinata in ordine alfabetico (cliente VIP = cliente che ha accumulato un certo numero di punti sulla base dei noleggi effettuati)
 - non esiste un programma per la generazione della lista dei clienti VIP
 - esiste pero` un programma che stampa la lista di tutti i clienti in ordine alfabetico

25

Difficolta` nell'accesso ai dati



- Per risolvere il problema si puo`:
 - stampare la lista di tutti i clienti ed estrarre manualmente da tale lista quella relativa ai clienti VIP
 - Tempi molto lunghi in presenza di molti clienti
 - richiedere che sia sviluppato un nuovo programma
 - Supponiamo che dopo qualche tempo sia necessario stampare la lista di tutti i clienti VIP ordinata per numero di punti accumulati
 - Tale programma non esiste
 - Si torna al problema di partenza

26

Difficolta` nell'accesso ai dati



- Un DBMS supera queste limitazioni, mettendo a disposizione:
 1. linguaggi che facilitano l'accesso ai dati secondo modalita` non necessariamente note a priori
 2. strutture ausiliarie di accesso che consentono di rispondere piu` efficientemente alle richieste di utenti ed applicazioni

27

Problemi nell'accesso concorrente ai dati



- Per migliorare le prestazioni, tutti i sistemi permettono a piu` utenti di eseguire accessi concorrenti ai dati
- I SO forniscono meccanismi per garantire la mutua esclusione nella modifica dei dati
- Tali meccanismi devono pero` essere estesi in un ambiente di basi di dati



- L'interazione di modifiche concorrenti, se non opportunamente controllata, puo` causare inconsistenze dei dati e puo` portare utenti ed applicazioni ad agire su dati non corretti

28

Problemi nell'accesso concorrente ai dati



- Nella videoteca:
 - Supponiamo che esista un'applicazione che, ad intervalli regolari, conti il numero di noleggi effettuati da ogni cliente della videoteca e generi un report con queste informazioni
 - Supponiamo che, mentre l'applicazione è in esecuzione, un cliente per cui il conteggio è già stato effettuato, noleggi un video
 - In questo caso, l'esecuzione concorrente delle due applicazioni genera un conteggio non esatto
 - Non permettere l'esecuzione concorrente di applicazioni degraderebbe invece in modo non accettabile le prestazioni del sistema

29

Problemi nell'accesso concorrente ai dati



- **Transazione:** porzione di programma applicativo a cui il DBMS assicura particolari proprietà durante la sua esecuzione tra cui la consistenza dei dati in presenza di transazioni concorrenti, senza che il programmatore debba preoccuparsi della gestione di tali problematiche

30

Problemi nell'accesso concorrente ai dati



- Il corso non introdurrà il concetto di transazione

31

Problemi di protezione dei dati



- Non tutti gli utenti devono poter accedere a tutti i dati presenti nel sistema
- I file system tipicamente non hanno meccanismi di controllo dell'accesso adeguati a supportare tali requisiti di protezione



- Questi controlli devono essere implementati a livello di programmi applicativi ma difficile verificare che tali controlli siano effettivamente e correttamente incorporati in tutti i programmi applicativi

32

Problemi di protezione dei dati



- Nella videoteca:
 - un commesso della videoteca potrebbe essere autorizzato a conoscere la data in cui un cliente ha effettuato un noleggio
 - per motivi di privacy, potrebbe non potere conoscere il titolo del film che ha noleggiato

33

Problemi di protezione dei dati



- E' necessario disporre di meccanismi che consentano un accesso selettivo ai dati:
 - solo a determinati record
 - solo a determinati campi di ogni record
- Un DBMS offre tali meccanismi per gestire efficacemente il controllo dell'accesso senza dover ricorrere allo sviluppo di programmi ad-hoc

34

Problemi di protezione dei dati



- Non vedremo questi aspetti nel corso

35

Problemi di integrità dei dati



- Per riflettere correttamente una certa realtà applicativa, è necessario che i dati rispettino determinate condizioni, note come **vincoli di integrità semantica**
- Un insieme di dati è semanticamente corretto se verifica tutti i vincoli di integrità semantica ad esso associati
- I file system non consentono la specifica e la verifica automatica di tali vincoli
- Devono quindi essere implementati come codice applicativo



- L'aggiunta di nuovi vincoli o la modifica di un qualche vincolo richiede costose modifiche ai programmi applicativi

36

Problemi di integrità dei dati



- Nella videoteca:
 - Esempio di vincolo:
 - la data di restituzione di un video non deve essere antecedente alla data di noleggio
 - Il vincolo deve essere implementato in ogni programma che gestisce inserimenti e modifiche ai noleggi

37

Problemi di integrità dei dati



- I DBMS offrono adeguati strumenti per la specifica di vincoli e la loro verifica automatica
- I vincoli vengono specificati dall'utente una sola volta
- E' poi il sistema a preoccuparsi di verificarli sui dati

- Nel corso vedremo come questo sia possibile

38

La carta vincente dei DBMS



- Il meccanismo fondamentale di un DBMS che permette di risolvere le problematiche prima delineate è lo **schema** (o schema logico) della base di dati
- Uno schema logico descrive il contenuto della base di dati tramite un formalismo ad alto livello che esula dai dettagli della sua effettiva implementazione fisica, detto **modello dei dati**

39

Vantaggi



- Unica rappresentazione centralizzata e condivisa
 - No ridondanze e incoerenze
 - consente di introdurre un controllo centralizzato sugli stessi, minimizzando quindi i problemi di protezione dei dati
- Il DBMS mette a disposizione linguaggi (dichiarativi) per agire sui dati rappresentati secondo il modello
 - Possibilità di interagire con i dati secondo modalità non anticipate al momento della definizione e realizzazione della base di dati
 - Ottimizza le operazioni e quindi garantisce efficienza (l'utente non si preoccupa di come i dati sono memorizzati e gestiti su disco)
 - Garantisce affidabilità

40

Servizi di un DBMS

Esterni (invocabili dagli utenti)



Descrizione dei dati	Per specificare i dati da memorizzare nella base di dati
Manipolazione dei dati	Per: <ul style="list-style-type: none">- accedere ai dati- inserire nuovi dati- modificare dati esistenti- cancellare dati esistenti

Interni al DBMS

Controllo di integrità	Per evitare di memorizzare dati non corretti
Strutture di memorizzazione	Per rappresentare in memoria secondaria i costrutti del modello dei dati
Ottimizzazione di interrogazioni	Per determinare la strategia più efficiente per accedere ai dati
Protezione dei dati	Per proteggere i dati da accessi non autorizzati
Ripristino della base di dati	Per evitare che errori e malfunzionamenti: <ul style="list-style-type: none">- determinino una base di dati inconsistente- provochino perdite di dati
Controllo della concorrenza	Per evitare che accessi concorrenti alla base di dati provochino inconsistenze dei dati

Modello dei dati



- Un modello dei dati è un insieme di strumenti concettuali, o formalismo, che consta di tre componenti fondamentali:
 - un insieme di strutture dati
 - un linguaggio per specificare i dati tramite le strutture previste dal modello, per aggiornare tali strutture e per specificare vincoli su tali strutture
 - un linguaggio per manipolare i dati

Modello dei dati: concetti di base



- Le strutture dati devono poter rappresentare i seguenti concetti:
- **Entità**. Insieme di “oggetti” della realtà applicativa di interesse, aventi caratteristiche comuni
- **Istanza di entità**. Singolo oggetto della realtà applicativa di interesse, modellato da una certa entità

43

Modello dei dati: concetti di base



- **Attributo**. Proprietà significativa di un'entità, ai fini della descrizione della realtà applicativa di interesse
 - Ogni entità è caratterizzata da uno o più attributi
 - Un attributo di un'entità assume uno o più valori per ciascuna delle istanze dell'entità, detti valori dell'attributo, in un insieme di possibili valori, detto dominio dell'attributo

44

Modello dei dati: concetti di base



- **Associazione.** Corrispondenza tra un certo numero di entita`
 - Anche le associazioni possono avere degli attributi che corrispondono a proprieta` dell'associazione
- **Istanza di associazione.** Corrispondenza tra le istanze di un certo numero di entita`

45

Modello dei dati: esempio videoteca



- **Entita`:** *Cliente, Noleggio, Video e Film.*
 - L'entita` Cliente rappresenta l'insieme dei clienti della videoteca
- **Istanza di entita`:**
 - anna rossi istanza di Cliente,
 - pulp fiction istanza di Film
- **Attributi:**
 - Il *nome di un cliente* attributo dell'entita` Cliente
 - la *data di restituzione del video noleggiato* attributo dell'entita` Noleggio
 - il *regista* di un film attributo dell'entita` Film
- **Associazione:** legame tra i clienti ed i film che consigliano (*Consiglia*)
 - potrebbe avere un attributo giudizio, che modella il giudizio espresso dai clienti della videoteca sui film da loro consigliati
- **Istanza di associazione:**
 - relazione che lega l'istanza di Film relativa al film "Pulp fiction" di Quentin Tarantino con l'istanza di Cliente relativa al cliente con numero di tessera 6610 e` un'istanza dell'associazione Consiglia

46

Modello dei dati



- Qualsiasi modello dei dati deve rispondere a due domande fondamentali:
 - (a) come rappresentare le entità e i loro attributi
 - (b) come rappresentare le associazioni e i loro attributi
- la maggioranza dei modelli usa strutture come il record, in cui ogni componente rappresenta un attributo
- i modelli differiscono notevolmente in questo aspetto; la rappresentazione può avvenire tramite:
 - strutture, valori, puntatori

47

Il modello relazionale



- E' basato su una singola struttura dati - la *relazione*
- una relazione può essere vista come una tabella con righe, dette *tuple*, e colonne contenenti dati di tipo specificato (ad esempio interi e stringhe)
- Relazione = insieme di entita`
- Tupla = entita`

48

Esempio



Relazione

Attributo

Film				
titolo	regista	anno	genere	valutaz
underground	emir kusturica	1995	drammatico	3.20
edward mani di forbice	tim burton	1990	fantastico	3.60
nightmare before christmas	tim burton	1993	animazione	4.00
ed wood	tim burton	1994	drammatico	4.00
mars attacks	tim burton	1996	fantascienza	3.00
il mistero di sleepy hollow	tim burton	1999	horror	3.50
big fish	tim burton	2003	fantastico	3.10
la sposa cadavere	tim burton	2005	animazione	3.50
la fabbrica di cioccolato	tim burton	2005	fantastico	4.00
io non ho paura	gabriele salvatores	2003	drammatico	3.50
nirvana	gabriele salvatores	1997	fantascienza	3.00
mediterraneo	gabriele salvatores	1991	commedia	3.80
pulp fiction	quentin tarantino	1994	thriller	3.50
le iene	quentin tarantino	1992	thriller	4.00

Tupla

49

Il modello relazionale



- Il modello relazionale utilizza una rappresentazione cosiddetta per valore delle associazioni
 - non sono rappresentate esplicitamente (mediante costrutti quali puntatori) ma implicitamente tramite
 - La replica di alcune colonne di una tabella in quella con cui vogliamo stabilire un'associazione
 - creando una tabella ad-hoc per modellare l'associazione contenente alcune colonne delle tabelle che vogliamo mettere in relazione

50

Esempio

Film

titolo	regista	anno	genere	valutaz
underground	emir kusturica	1995	dramatico	3.20
edward mani di forbice	tim burton	1990	fantastico	3.60
nightmare before christmas	tim burton	1993	animazione	4.00
ed wood	tim burton	1994	drammatico	4.00
mars attacks	tim burton	1996	fantascienza	3.00
il mistero di sleepy hollow	tim burton	1999	horror	3.50
big fish	tim burton	2003	fantastico	3.10
la sposa cadavere	tim burton	2005	animazione	3.50
la fabbrica di cioccolato	tim burton	2005	fantastico	4.00
io non ho paura	gabriele salvatores	2003	drammatico	3.50
nirvana	gabriele salvatores	1997	fantascienza	3.00
mediterraneo	gabriele salvatores	1991	comedia	3.80
pulp fiction	quentin tarantino	1994	thriller	3.50
le iene	quentin tarantino	1992	thriller	4.00

Video

colloc	titolo	regista	tipo
1111	underground	emir kusturica	v
1112	underground	emir kusturica	d
1113	big fish	tim burton	v
1114	big fish	tim burton	d
1115	edward mani di forbice	tim burton	d
1116	nightmare before christmas	tim burton	v
1117	nightmare before christmas	tim burton	d
1118	ed wood	tim burton	d
1119	mars attacks	tim burton	d
1120	il mistero di sleepy hollow	tim burton	d
1121	la sposa cadavere	tim burton	d
1122	la fabbrica di cioccolato	tim burton	d
1123	la fabbrica di cioccolato	tim burton	d
1124	io non ho paura	gabriele salvatores	d
1125	nirvana	gabriele salvatores	d
1126	mediterraneo	gabriele salvatores	d
1127	pulp fiction	quentin tarantino	v
1128	pulp fiction	quentin tarantino	d
1129	le iene	quentin tarantino	d

51

Il modello relazionale

- **Chiave:** insieme di attributi che identificano univocamente le tuple di una relazione
 - (Titolo, regista) in Film
 - Colloc in Video
- **Chiave esterna:** gli attributi corrispondenti ad una chiave in un'altra relazione, inseriti per modellare l'associazione tra le tuple delle due relazioni
 - (Titolo, regista) in Video

52

Esempio

chiave

Film

titolo	regista	anno	genere	valutaz
underground	emir kusturica	1995	dramatico	3.20
edward mani di forbice	tim burton	1990	fantastico	3.60
nightmare before christmas	tim burton	1993	animazione	4.00
ed wood	tim burton	1994	dramatico	4.00
mars attacks	tim burton	1996	fantascienza	3.00
il mistero di sleepy hollow	tim burton	1999	horror	3.50
big fish	tim burton	2003	fantastico	3.10
la sposa cadavere	tim burton	2005	animazione	3.50
la fabbrica di cioccolato	tim burton	2005	fantastico	4.00
io non ho paura	gabriele salvatores	2003	dramatico	3.50
nirvana	gabriele salvatores	1997	fantascienza	3.00
mediterraneo	gabriele salvatores	1991	comedia	3.50
pulp fiction	quentin tarantino	1994	thriller	3.50
le iene	quentin tarantino	1992	thriller	4.00

Video

colloc	titolo	regista	tipo
1111	underground	emir kusturica	v
1112	underground	emir kusturica	d
1113	big fish	tim burton	v
1114	big fish	tim burton	d
1115	edward mani di forbice	tim burton	d
1116	nightmare before christmas	tim burton	v
1117	nightmare before christmas	tim burton	d
1118	ed wood	tim burton	d
1119	mars attacks	tim burton	d
1120	il mistero di sleepy hollow	tim burton	d
1121	la sposa cadavere	tim burton	d
1122	la fabbrica di cioccolato	tim burton	d
1123	la fabbrica di cioccolato	tim burton	d
1124	io non ho paura	gabriele salvatores	d
1125	nirvana	gabriele salvatores	d
1126	mediterraneo	gabriele salvatores	d
1127	pulp fiction	quentin tarantino	v
1128	pulp fiction	quentin tarantino	d
1129	le iene	quentin tarantino	d

chiave esterna

53

Il modello relazionale

Concetto	Costrutto relazionale
Entità	Relazione
Istanza di entità	Tupla
Attributo	Attributo di una relazione/associazione
Associazione	Relazione/attributi di una relazione
Istanza di associazione	Tupla Valori uguali per chiavi e chiavi esterne

54



Schema e istanza

- Indipendentemente dal modello dei dati, in un DBMS distinguiamo:
 - **Schema della base di dati**: descrizione dei dati, specificata tramite il modello dei dati
 - Cambia raramente
 - Descrizione intensionale
 - **Istanza della base di dati**: insieme dei dati presenti in un dato momento in una base di dati
 - Cambia molto spesso nel tempo
 - Descrizione estensionale
- In una base di dati relazionale:
 - Schema della base di dati = insieme degli schemi delle relazioni in esso presenti
 - Schema di relazione: nome della relazione + nome e dominio per ogni attributo

55



Esempio

Video

colloc	titolo	regista	tipo
1111	underground	emir kusturica	v
1112	underground	emir kusturica	d
1113	big fish	tim burton	v
1114	big fish	tim burton	d
1115	edward mani di forbice	tim burton	d
1116	nightmare before christmas	tim burton	v
1117	nightmare before christmas	tim burton	d
1118	ed wood	tim burton	d
1119	mars attacks	tim burton	d
1120	il mistero di sleepy hollow	tim burton	d
1121	la sposa cadavere	tim burton	d
1122	la fabbrica di cioccolato	tim burton	d
1123	la fabbrica di cioccolato	tim burton	d
1124	io non ho paura	gabriele salvatores	d
1125	nirvana	gabriele salvatores	d
1126	mediterraneo	gabriele salvatores	d
1127	pulp fiction	quentin tarantino	v
1128	pulp fiction	quentin tarantino	d
1129	le iene	quentin tarantino	d

Schema

Istanza

Film

titolo	regista	anno	genere	valutaz
underground	emir kusturica	1995	drammatico	3.20
edward mani di forbice	tim burton	1990	fantastico	3.60
nightmare before christmas	tim burton	1993	animazione	4.00
ed wood	tim burton	1994	drammatico	4.00
mars attacks	tim burton	1996	fantascienza	3.00
il mistero di sleepy hollow	tim burton	1999	horror	3.50
big fish	tim burton	2003	fantastico	3.10
la sposa cadavere	tim burton	2005	animazione	3.50
la fabbrica di cioccolato	tim burton	2005	fantastico	4.00
io non ho paura	gabriele salvatores	2003	drammatico	3.50
nirvana	gabriele salvatores	1997	fantascienza	3.00
mediterraneo	gabriele salvatores	1991	commedia	3.80
pulp fiction	quentin tarantino	1994	thriller	3.50
le iene	quentin tarantino	1992	thriller	4.00

Schema e istanza



- Il primo passo nello sviluppo di una base di dati è la definizione dello schema della base di dati
- successivamente vengono immessi i dati veri e propri che devono conformarsi alla definizione data dallo schema

57

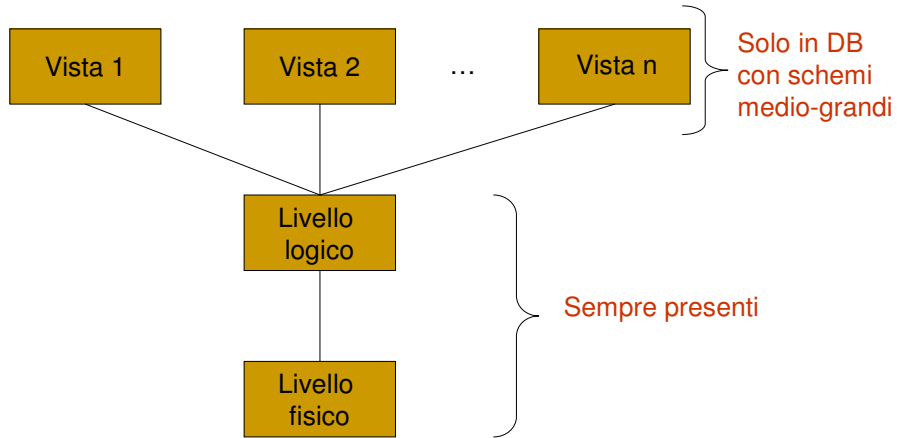
Livelli nella rappresentazione dei dati



- **Livello fisico.** E' il livello piu' basso in cui viene definito lo **schema fisico** della base di dati, precisando come i dati sono effettivamente memorizzati tramite strutture di memorizzazione (file, record, ecc.)
- **Livello logico.** E' il secondo livello di astrazione in cui viene descritto lo **schema logico**
 - quali sono i dati memorizzati nella base di dati
 - eventuali associazioni tra di essi
 - vincoli di integrita` semantica e di autorizzazioneLa base di dati e` descritta tramite le strutture del modello dei dati
- **Livello esterno o livello delle viste.** E' il livello di astrazione piu' alto
 - descrive una porzione dell'intero schema della base di dati (**vista**)
 - possono essere definite piu' viste di una stessa base di dati

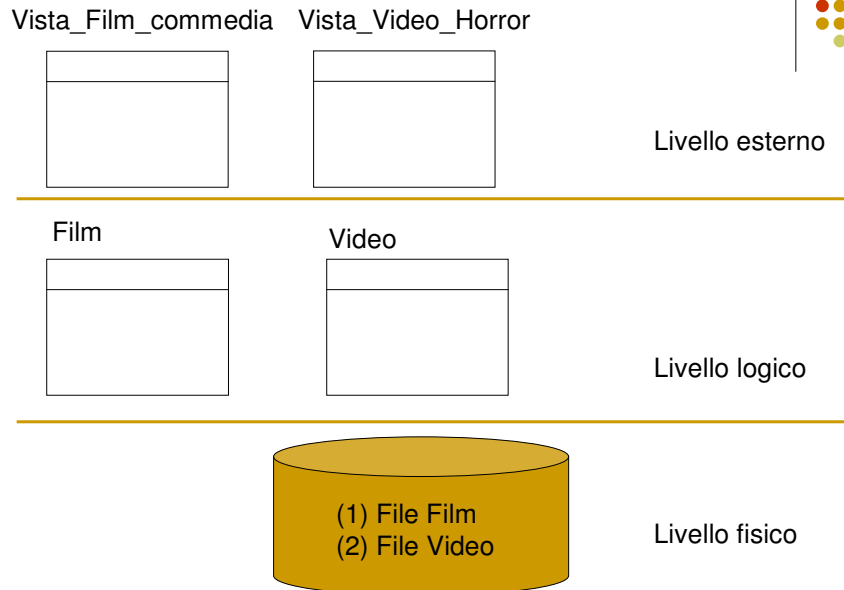
58

Livelli nella rappresentazione dei dati



59

Esempio



60

Indipendenza



- L'introduzione di questi tre livelli assicura alcune importanti proprietà ai dati, che facilitano l'accesso ai dati e lo sviluppo di applicazioni
 - **indipendenza fisica:** utenti ed applicazioni che accedono alla rappresentazione logica dei dati sono indipendenti da qualsiasi modifica a livello di rappresentazione fisica dei dati in quanto modifiche a tale livello non influenzano la rappresentazione dei dati al livello logico
 - **Indipendenza logica:** la presenza delle viste permette di nascondere (entro certi limiti) modifiche alla rappresentazione dei dati al livello logico alle applicazioni/utenti che vedono la rappresentazione esterna (tramite vista) dei dati

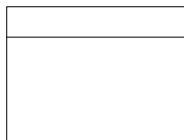
61

Indipendenza fisica: esempio

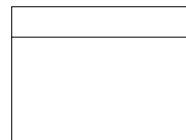


Tempo T1

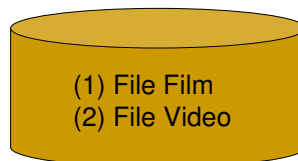
Film



Video



Livello logico



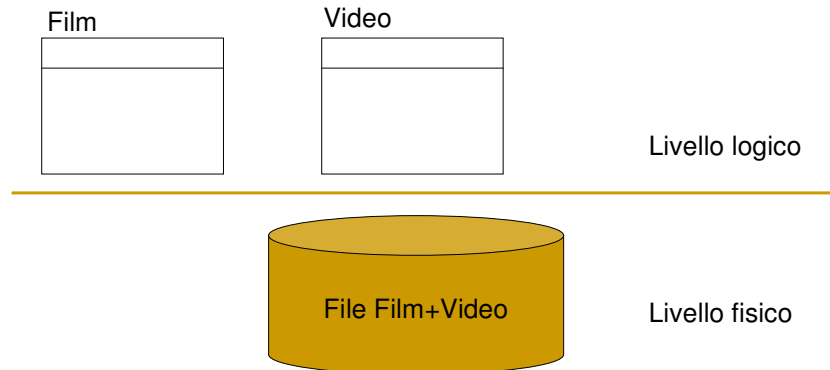
Livello fisico

62

Indipendenza fisica: esempio



Tempo T2:
cambiano le strutture di memorizzazione
il livello logico non cambia

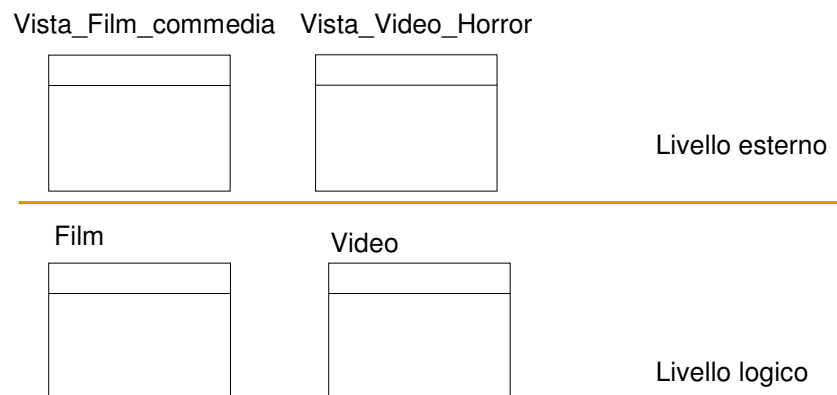


63

Indipendenza logica: esempio



Tempo T1



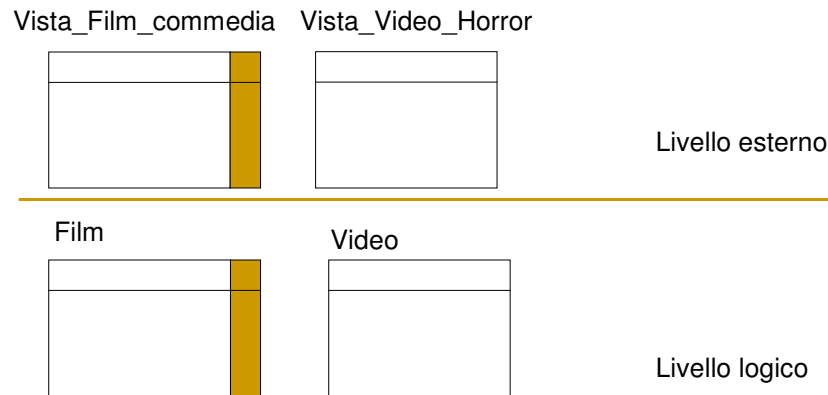
64

Indipendenza logica: esempio



Tempo T2:

cambia lo schema logico (aggiunta di un attributo)
solo una parte delle viste cambia



65

Il modello dei dati: linguaggi



- **Data Definition Language (DDL)**
 - Permette di specificare e modificare lo schema della base di dati e lo schema delle viste
 - Livello logico ed esterno
- **Data Manipulation Language (DML)**
 - Permette di creare, modificare e interrogare l'istanza della base di dati
 - Livello logico
- **Storage Definition Language (SDL)**
 - Definisce lo schema fisico del DB
 - Livello fisico

66

Data Definition Language



- Concretizza il modello dei dati fornendo la notazione che permette di specificarne le strutture
- Permette la specifica e la modifica di:
 - nome della base di dati
 - tutte le unità logiche elementari della base di dati (ad esempio relazioni e colonne nel caso del modello relazionale)
 - eventuali vincoli di integrità semantica
 - viste

67

Data Manipulation Language



- Fornisce le seguenti operazioni di base:
 - **Inserimento**: per l'immissione di nuovi dati
 - **Ricerca**: per il ritrovamento dei dati di interesse
 - Un'operazione di ricerca è detta **interrogazione** da cui il nome linguaggio di interrogazione del DML che permette di esprimere le operazioni di ricerca
 - **Cancellazione**: per l'eliminazione di dati obsoleti
 - **Aggiornamento**: per la modifica di dati

68

Data Manipulation Language



- Può essere:
 - **Procedurale** (operazionale)
 - è responsabilità del programmatore specificare sia il risultato che vuole ottenere sia come ottenerlo
 - **Non procedurale** (dichiarativo)
 - l'utente deve solo specificare quali caratteristiche devono avere i dati su cui vuole operare, tipicamente indicando condizioni sugli attributi
 - Viene lasciata al DBMS la responsabilità di decidere come accedere effettivamente a tali dati

69

Esempio



- Interrogazione dichiarativa:
 - Voglio conoscere il nome del film contenuto nel video con codice di collocazione 1116
- Interrogazione procedurale:
 - Per ogni tupla della tabella Video, guarda se il codice è uguale a 1116.
 - Se sì, restituisci in output il nome del film
 - Altrimenti, passa ad analizzare la tupla successiva

70

Linguaggi dichiarativi



- Vantaggi:
 - facile utilizzo, anche per utenti poco esperti di informatica
 - possibilità di attuare strategie di ottimizzazione
 - non specificare il modo operativo con cui ottenere il risultato consente al DBMS di applicare tutta una serie di strategie per eseguire in modo ottimizzato l'interrogazione limitando il numero di accessi a disco necessari

71

Storage Definition Language



- Le corrispondenze fra le strutture logiche dei dati, specificate nello schema logico della base di dati, e le strutture di memorizzazione devono essere opportunamente definite
- Nella maggior parte dei DBMS attuali la definizione di tali corrispondenze è eseguita automaticamente dal DBMS stesso una volta che lo schema è definito
- Tuttavia l'utente esperto può influenzare le scelte operate dal DBMS per velocizzare determinati accessi ai dati
- Tali richieste vengono effettuate tramite i comandi dell'SDL

72

Utenti di un DBMS



- Amministratore della base di dati (DBA)
- Progettista di basi di dati
- Programmatore applicativo
- Progettista, sviluppatore di DBMS
- Utenti semplici

73

Database administrator (DBA)



- Amministrazione e controllo della base di dati
- Deve stabilire, in accordo alle politiche dell'organizzazione, le regole per l'utilizzo dei dati e le procedure che assicurino la protezione e l'integrità dei dati
- Deve occuparsi delle opportune procedure di ripristino della base di dati
- Migliorare, ove possibile, l'efficienza e la redditività del sistema
- Mantenere i contatti con gli utenti

74

Progettista di basi di dati



- Provvede, in base ai requisiti dell'organizzazione e delle varie applicazioni, alla progettazione della base di dati
- Definisce lo schema logico e fisico della base di dati, ed eventuali viste sugli stessi

75

Programmatore applicativo



- E` preposto allo sviluppo dei programmi applicativi che operano sulla base di dati ed effettua il mantenimento di tali programmi
- Spesso piu` di uno
- I programmatori che hanno inizialmente sviluppato i programmi non sono necessariamente gli stessi che eseguono la manutenzione

76

Progettista di DBMS



- Un DBMS e` un software
- In quanto tale, deve essere progettato, sviluppato e mantenuto cosi` come una qualsiasi applicazione software
- A causa della complessita` dei DBMS, servono adeguate professionalita` per svolgere tale compito

77

Utenti semplici



- **Utenti parametrici**
 - usano sistematicamente la base di dati, come parte integrante delle loro mansioni
 - l'uso che tali utenti fanno dei dati puo` essere anche complesso ma e` fondamentalmente noto a priori
 - E` possibile sviluppare appositi programmi applicativi per supportare le funzioni necessarie a tali utenti
- **Utenti occasionali**
 - Utenti con richieste non predicibili a priori e che tuttavia nella maggior parte dei casi fanno un uso molto semplice della base di dati
- Per tali utenti sono spesso predisposte interfacce interattive molto semplici da utilizzare

78