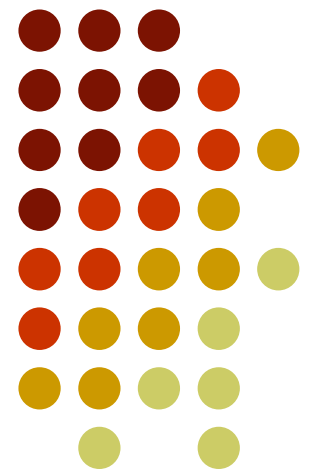


Basi di dati: aspetti introduttivi

Elena Ferrari
Basi di Dati
A.A. 2018/2019





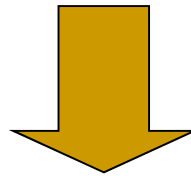
Un po' di storia

- I computer sono stati originariamente sviluppati per applicazioni di calcolo numerico
- Col tempo, la capacità di memorizzare e gestire informazioni è diventata una delle funzionalità primarie
- La quantità e la tipologia di informazioni sono aumentate:
 - Immagini/audio/video
 - Dati biomedici/transazioni cliente
 - Big Data
- **Necessità di sistemi software che rispondano a queste esigenze**



Sistema informativo

- Una delle principali esigenze è quella di gestire e rendere disponibili le informazioni



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

Esempi



- Uno studio medico mantiene, per ragioni fiscali, informazioni sui pazienti, sulle visite fatte e sulle parcelle richieste



Esempi

- Un'industria manifatturiera gestisce informazioni per svolgere attività che includono:
 - gestione degli ordini e dei pagamenti
 - dei prodotti venduti
 - ai fornitori di materiali
 - gestione del magazzino
 - pianificazione della produzione
 - controllo di gestione

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



Dati

- Nei sistemi informatici, le informazioni sono rappresentate sotto forma di dati
- I dati sono simboli grezzi che devono essere interpretati e correlati per fornire informazioni



Dati

- I dati hanno bisogno di un contesto interpretativo che permetta di estrarre da essi le informazioni di interesse per gli utenti:
- Esempio
 - “Mario Rossi” e 347424242
 - informazione: risultato della ricerca di un numero di telefono sulla propria agenda telefonica

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



Caratteristiche dei dati

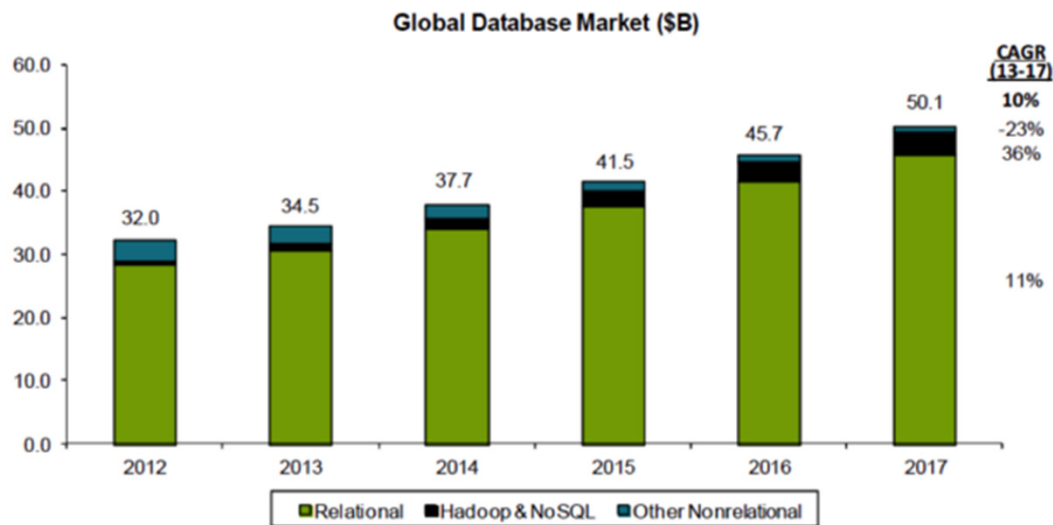
- I dati sono molto più stabili nel tempo dei processi che li gestiscono
- Esempio:
 - i dati relativi alle applicazioni bancarie hanno una struttura invariata da decenni
 - le procedure che li gestiscono variano di anno in anno
- I dati sono una risorsa importante dell'organizzazione che li gestisce



Base di dati

- **Base di dati (def. generale):**
 - 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 (def. tecnica):** collezione di dati gestita da un DBMS

Mercato dei DBMS



Source: IDC, Bernstein analysis

Mercato dei DBMS

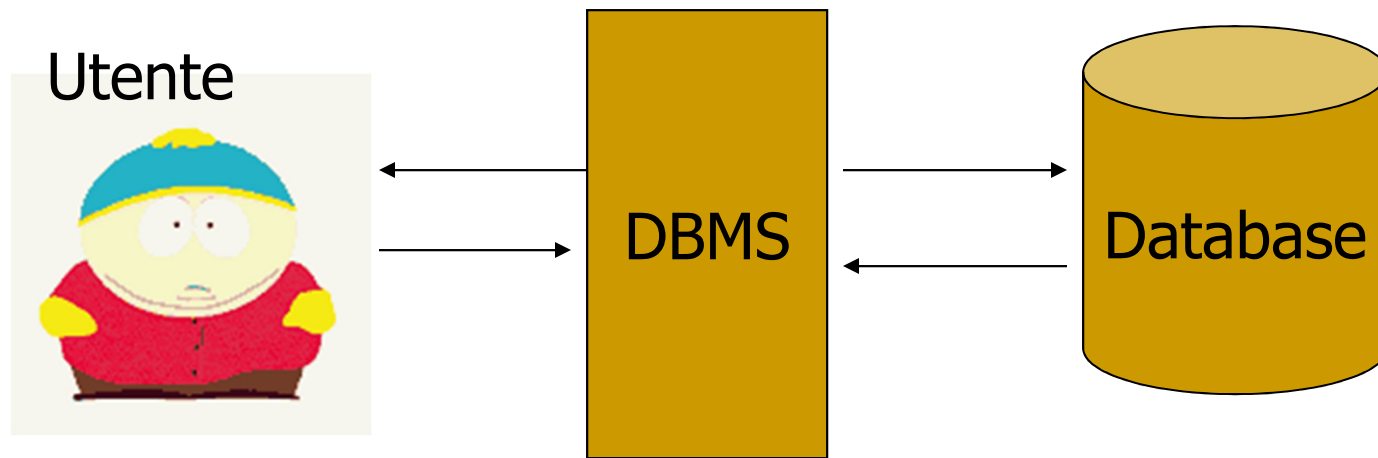
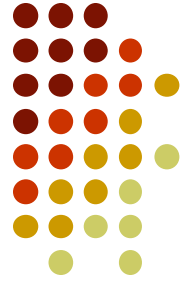


343 systems in ranking, September

Rank			DBMS	Database Model	Score	
Sep 2018	Aug 2018	Sep 2017			Sep 2018	Aug 2018
1.	1.	1.	Oracle +	Relational DBMS	1309.12	-2.91
2.	2.	2.	MySQL +	Relational DBMS	1180.48	-26.33
3.	3.	3.	Microsoft SQL Server +	Relational DBMS	1051.28	-21.37
4.	4.	4.	PostgreSQL +	Relational DBMS	406.43	-11.07
5.	5.	5.	MongoDB +	Document store	358.79	+7.81
6.	6.	6.	DB2 +	Relational DBMS	181.06	-0.78
7.	↑ 8.	↑ 10.	Elasticsearch +	Search engine	142.61	+4.49
8.	↓ 7.	↑ 9.	Redis +	Key-value store	140.94	+2.37
9.	9.	↓ 7.	Microsoft Access	Relational DBMS	133.39	+4.30
10.	10.	↓ 8.	Cassandra +	Wide column store	119.55	-0.02

<https://db-engines.com/en/ranking>

DBMS



- Le vecchie bd richiedevano grandi server dedicati
- Oggi è possibile memorizzare gigabyte in una chiave usb
- Un DBMS può essere installato in un PC



DBMS relazionali

- Alcuni RDBMS:
 - Oracle
 - IBM DB2
 - Microsoft SQL Server
 - MySQL, PostgreSQL (open source)



Caratteristiche dei DBMS

- Dimensioni molto maggiori della memoria centrale disponibile
 - gestione dei dati in memoria secondaria
- Condivisione dei dati tra applicazioni ed utenti: una base di dati è una risorsa **integrata**, condivisa da più settori aziendali
 - riduzione della ridondanza dei dati
 - riduzione delle inconsistenze tra i dati
 - meccanismo di controllo dell'accesso concorrente



Caratteristiche dei DBMS

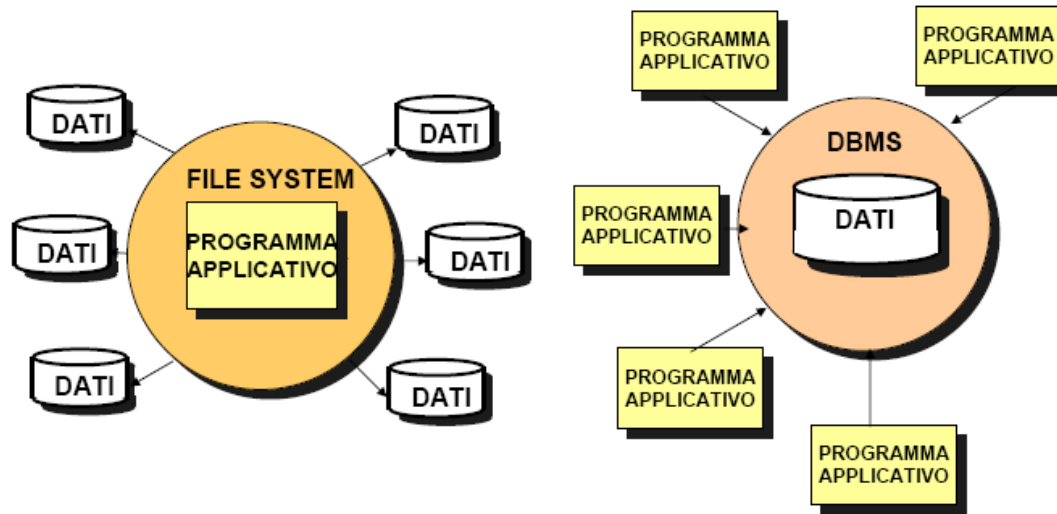
- **Persistenza dei dati**
 - tempo di vita non limitato a quello dell'esecuzione dei programmi che li utilizzano
- **Affidabilità dei dati in caso di malfunzionamenti hardware/software**
 - funzionalità di salvataggio (backup) e ripristino (recovery)
- **Sicurezza dei dati**
 - meccanismi di autorizzazione



Caratteristiche dei DBMS

- **Efficienza**
 - capacità di svolgere le operazioni utilizzando un insieme di risorse (tempo e spazio) accettabile per gli utenti
- **Efficacia**
 - capacità di rendere produttive le attività degli utenti

Base di dati vs. file system



- I DBMS estendono le funzionalità dei file system, fornendo più servizi ed in maniera **integrata**



La carta vincente dei DBMS

- Il meccanismo fondamentale di un DBMS è lo **schema** (o schema logico) della base di dati
- Lo 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**

Esempio



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



Modello dei dati

- Un modello dei dati è un insieme di concetti, o formalismo, utilizzato per rappresentare i dati di interesse. Consta di tre componenti fondamentali:
 - Un insieme di strutture dati
 - Linguaggi per:
 - specificare i dati tramite le strutture previste dal modello
 - aggiornare tali strutture
 - specificare vincoli su tali strutture
 - manipolare/ricercare i dati



Il modello relazionale

- Modello dei dati più diffuso
- 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)



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



Altri modelli dei dati

- Prima del modello relazionale, erano utilizzati modelli più vicini alle strutture fisiche di memorizzazione (poco astratti)
 - gerarchico
 - reticolare
- Dopo il modello relazionale:
 - object-oriented, object-relational
 - XML
 - NoSQL DB



Schema e istanza

- Indipendentemente dal modello dei dati, in un DBMS sono definiti:
 - **Schema (logico) della base di dati:** descrizione dei dati, specificata tramite il modello dei dati: *(come impostare la tabella)*
 - Cambia raramente
 - **Istanza della base di dati:** insieme dei dati presenti in un dato momento in una base di dati: *(Righe della tabella)*
 - Cambia molto spesso nel tempo

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

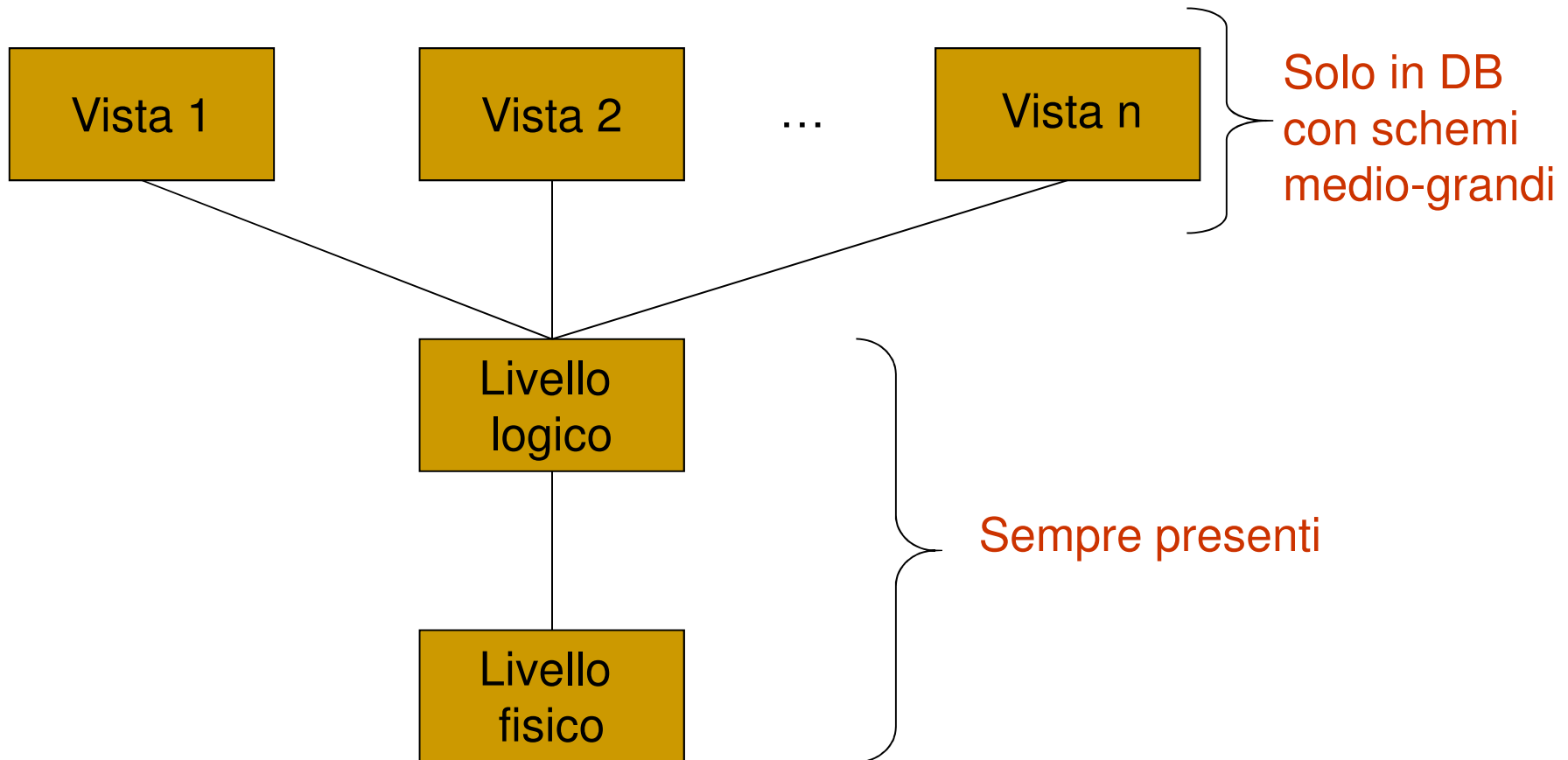
titolo	regista	anno	genere	valutaz
underground	emir kusturica	1995	drammatico	3.20
edward mani di forbice	tim burton	1990	fantastico	3.60
nightmare before christmas	tim burton	1993	animazione	4.00
ed wood	tim burton	1994	drammatico	4.00
mars attacks	tim burton	1996	fantascienza	3.00
il mistero di sleepy 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

Livelli nella rappresentazione dei dati



Livelli nella rappresentazione dei dati



- **Livello logico:**

- descrizione della base di dati mediante il modello logico del DBMS (**schema logico**):
 - Quali sono i dati memorizzati nella base di dati
 - Eventuali associazioni tra di essi
 - Vincoli di integrità semantica e di autorizzazione

- **Livello fisico:**

- E' il livello più basso in cui viene definito lo **schema fisico** della base di dati, precisando come i dati nello schema logico sono effettivamente memorizzati tramite strutture di memorizzazione (file, record, ecc.)

Livelli nella rappresentazione dei dati



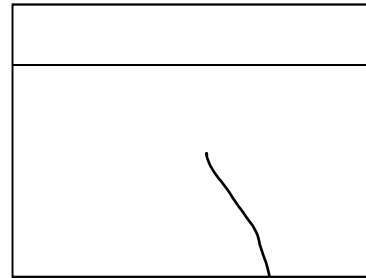
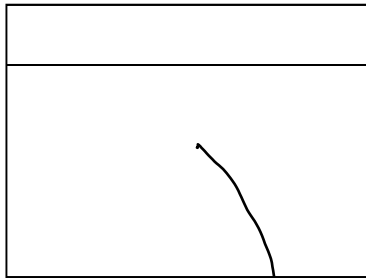
- **Livello esterno o livello delle viste:**
 - E' il livello di astrazione più alto:
 - Descrive una porzione dell'intero schema della base di dati (**vista**)
 - Possono essere definite più viste di una stessa base di dati
 - E' definito sullo schema logico

Esempio



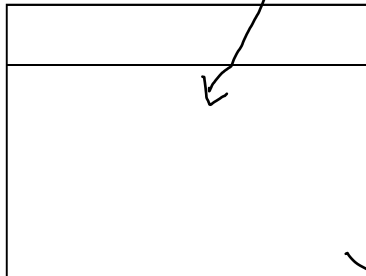
Vista_Film_commedia

Vista_Video_Horror

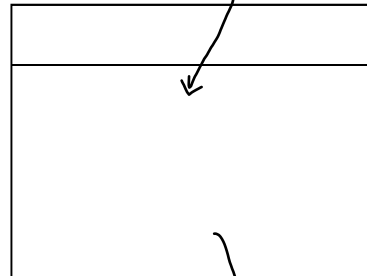


Livello esterno

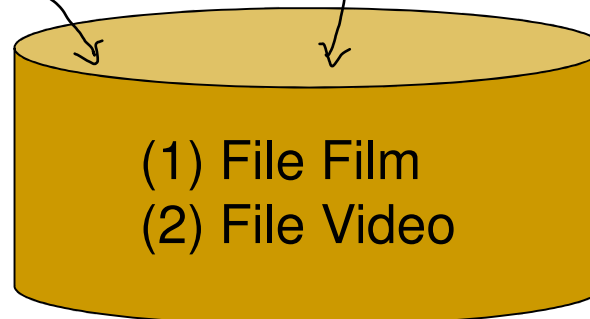
Film



Video



Livello logico



Livello fisico

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**
 - **Indipendenza logica**



Indipendenza fisica

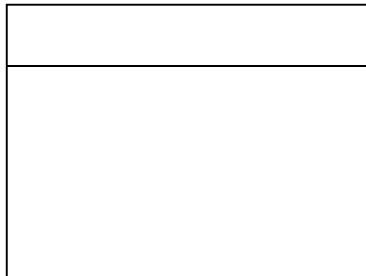
- Utenti ed applicazioni che accedono alla rappresentazione logica dei dati sono indipendenti da qualsiasi modifica a livello di rappresentazione fisica dei dati
- L'accesso a una relazione (livello logico o esterno) avviene sempre nello stesso modo, indipendentemente dalla modalità di memorizzazione fisica

Indipendenza fisica: esempio

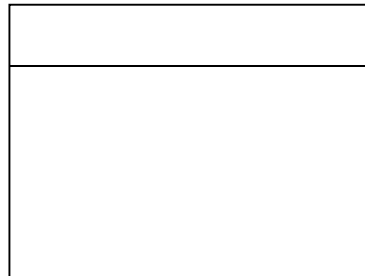


Tempo T1

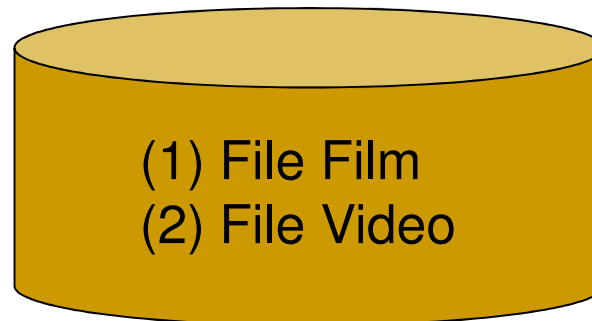
Film



Video



Livello logico



Livello fisico

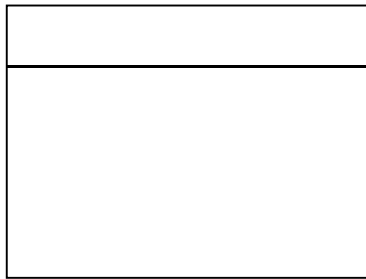


Indipendenza fisica: esempio

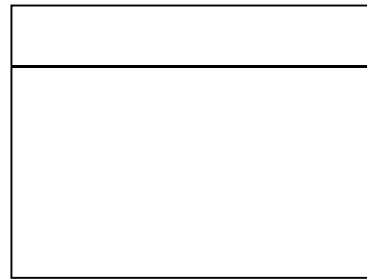
Tempo T2:

**cambiano le strutture di memorizzazione
il livello logico non cambia**

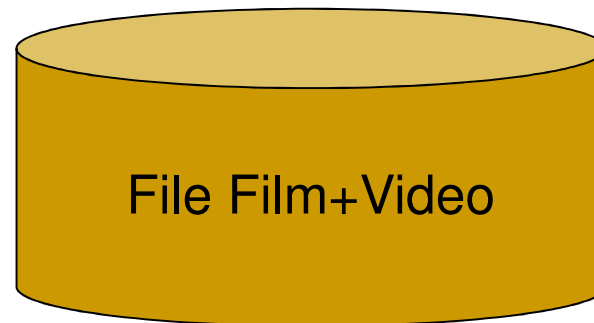
Film



Video



Livello logico



Livello fisico

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

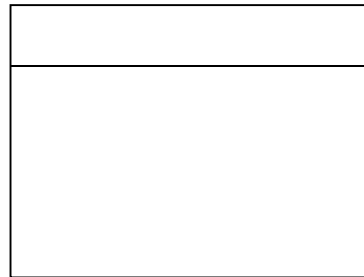
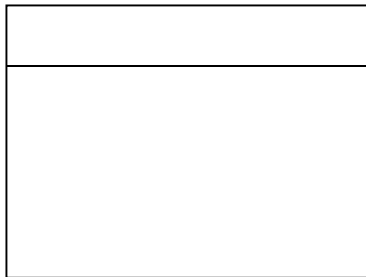
Indipendenza logica: esempio



Tempo T1

Vista_Film_commedia

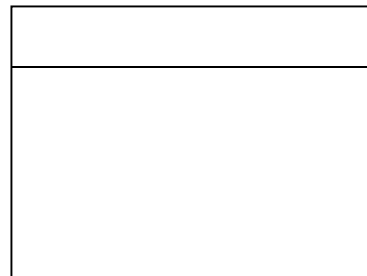
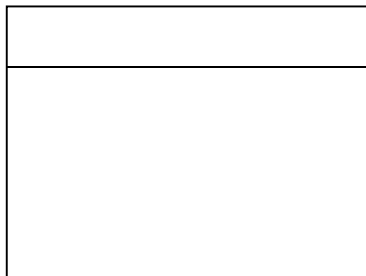
Vista_Video_Horror



Livello esterno

Film

Video



Livello logico

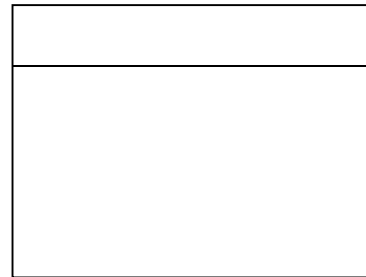
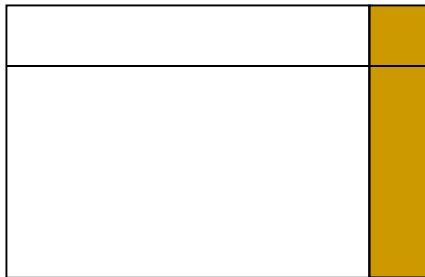


Indipendenza logica: esempio

Tempo T2:

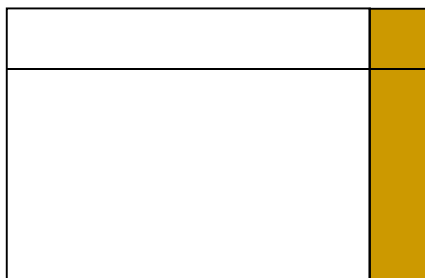
cambia lo schema logico (aggiunta di un attributo)
solo una parte delle viste cambia

Vista_Film_commedia Vista_Video_Horror

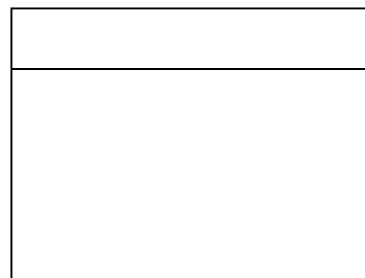


Livello esterno

Film



Video



Livello logico



Linguaggi di un DBMS

- **Data Definition Language (DDL)**

- Permette di specificare e modificare lo schema della base di dati, lo schema delle viste ed i vincoli di integrità
- Livello logico ed esterno

- **Data Manipulation Language (DML)**

- Permette di creare, modificare ed interrogare l'istanza della base di dati
- Livello logico ed esterno

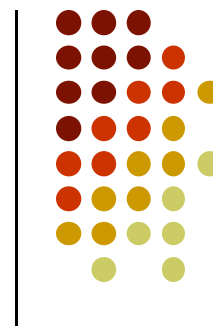
- **Storage Definition Language (SDL)**

- Definisce lo schema fisico del DB
- Livello fisico

Perché c'è bisogno di esperti di DB & DBMS?



- C'è bisogno di qualcuno che progetti la base di dati
- C'è bisogno di qualcuno che amministri (effettui il recovery ed il backup dei dati, gestisca gli utenti, ecc.) mantenga ed ottimizzi la base di dati
- C'è bisogno di qualcuno che sviluppi applicazioni che si interfacciano ad una base di dati
- C'è bisogno di qualcuno che progetti le interrogazioni
- C'è bisogno di qualcuno che progetti e realizzi i DBMS



ESERCIZI

Quali delle seguenti affermazioni sono vere?



- l'indipendenza dei dati permette:
 - di scrivere programmi senza conoscere le strutture fisiche dei dati
 - di modificare le strutture fisiche dei dati senza dover modificare i programmi che accedono alla base di dati
 - di formulare interrogazioni senza conoscere le strutture logiche dei dati

Quali delle seguenti affermazioni sono vere?



- il fatto che le basi di dati siano condivise favorisce:
 - l'efficienza dei programmi che le utilizzano
 - permette di ridurre ridondanze e inconsistenze
 - rende necessaria la gestione del controllo dell'accesso
- il fatto che le basi di dati siano persistenti:
 - ne garantisce l'affidabilità
 - favorisce l'efficienza dei programmi

Quali delle seguenti affermazioni sono vere?



- la distinzione fra DDL e DML corrisponde alla distinzione fra schema e istanza
- le istruzioni del DML permettono:
 - di interrogare la base di dati ma non di modificarla
 - di interrogare la base di dati e di modificarla
- le istruzioni del DDL permettono di specificare la struttura della base di dati ma non di modificarla
- non esistono linguaggi che includono sia istruzioni DDL sia istruzioni DML
- SQL include istruzioni DML e DDL



Esercizi

- Illustrare, in modo sintetico ma chiaro, supponendo di rivolgersi ad un non esperto, le caratteristiche fondamentali delle basi di dati e il ruolo che esse giocano nei sistemi informativi.
- Illustrare brevemente (non più di mezza pagina) il concetto di indipendenza dei dati