19 ottobre 2020

Basi di Dati + Laboratorio - Informatica Triennale - Corso A

94

Breve storia dei Database

- Prime applicazioni: I modelli gerarchico e reticolare sono stati introdotto nella metà degli anni '60, dominando durante gli anni settanta. Ancora oggi una buona parte di applicazioni funzionanti sono presenti in tutto il mondo
- Sistemi basati sul modello relazionale: modello introdotto nel 1970 da Codd pesantemente studiato e realizzato da IBM e dalle università. I primi prodotti per il modello relazionale sono nati negli anni '80

Basi di Dati + Laboratorio - Informatica Triennale - Corso A

Breve storia dei Database

- Applicazioni Object Oriented: l'OODBMS fu introdotto tra la fine degli anni '80 e i primi anni '90 per far fronte alle complesse esigenze dei processi di CAD e di altre applicazioni
- Dati nel Web e in applicazioni di e-commerce: il Web contiene dati in HTML (Hypertext markup language) e collegamenti tra pagine. Questo ha dato vita a nuove appicazioni (anche) nell'ambito dell'e-commerce utilizzando nuovi standard come XML (eXtended Markup Language)

Basi di Dati + Laboratorio - Informatica Triennale - Corso A

100

Hierarchical Model

ADVANTAGES:

- Hierarchical Model is simple to construct and operate on
- Corresponds to a number of natural hierarchically organized domains - e.g., assemblies in manufacturing, personnel organization in companies
- Language is simple; uses constructs like GET, GET UNIQUE, GET NEXT, GET NEXT WITHIN PARENT etc.

DISADVANTAGES:

- Navigational and procedural nature of processing
- Database is visualized as a linear arrangement of records
- Little scope for "query optimization"

Basi di Dati + Laboratorio - Informatica Triennale - Corso A

Network Model

ADVANTAGES:

- Network Model is able to model complex relationships and represents semantics of add/delete on the relationships
- Can handle most situations for modeling using record types and relationship types
- Language is navigational; uses constructs like FIND, FIND member, FIND owner, FIND NEXT within set, GET etc. Programmers can do optimal navigation through the database.

DISADVANTAGES:

- Navigational and procedural nature of processing
- Database contains a complex array of pointers that thread through a set of records. Little scope for automated "query optimization"

Basi di Dati + Laboratorio - Informatica Triennale - Corso A

102

History of Data Models



 Relational Model: proposed in 1970 by E.F. Codd (IBM), first commercial system in 1981-82.

Now in several commercial products (DB2, ORACLE, SQL Server, SYBASE, INFORMIX, MySql)

Basi di Dati + Laboratorio - Informatica Triennale - Corso A

History of Data Models

- Object-oriented Data Model(s): several models have been proposed for implementing in a database system. One set comprises models of persistent OO Programming Languages such as C++ (e.g., in OBJECTSTORE or VERSANT), and Smalltalk (e.g., in GEMSTONE). Additionally, systems like O2, ORION (at MCC then ITASCA), IRIS (at H.P.- used in Open OODB)
- Object-Relational Models: Most Recent Trend. Started with Informix Universal Server. Exemplified in the latest versions of Oracle-10i, DB2







Basi di Dati + Laboratorio - Informatica Triennale - Corso A

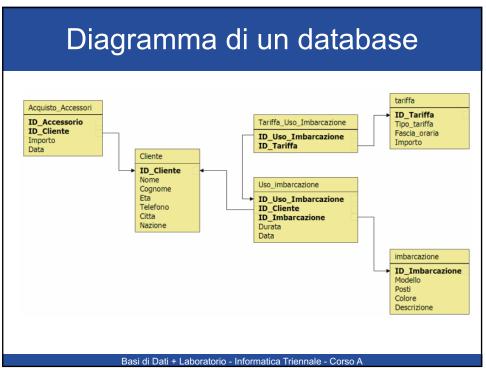
104

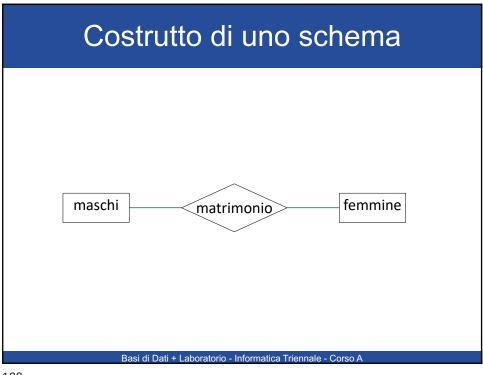
Schemas versus Instances

- Database Schema: The description of a database.
 Includes descriptions of the database structure and the constraints that should hold on the database
- Schema Diagram: A diagrammatic display of (some aspects of) a database schema
- Schema Construct: A component of the schema or an object within the schema, e.g., STUDENT, COURSE
- Database Instance: The actual data stored in a database at a particular moment in time. Also called database state (or occurrence)

Basi di Dati + Laboratorio - Informatica Triennale - Corso A

Schema di un database
veicoli targa nome_modello cilindrata cod_combustibile cavalli_fiscali posti velocità immatricolazione
modelli cod_modello nome_modello cod_fabbrica numero_versione
costruttori cod_costruttore nome_costruttore
categorie cod_categoria <mark>nome_categoria</mark>
combustibili cod_combustibile nome_combustibile
Basi di Dati + Laboratorio - Informatica Triennale - Corso A





Database Schema Vs. Database State

- Database State
 - the content of a database at a moment in time
- Initial Database State
 - · the database when it is loaded
- Valid State
 - satisfies the structure and constraints of the database
- Distinction
 - The database schema changes very infrequently. The database state changes every time the database is updated
 - Schema is also called intension, whereas state is called extension

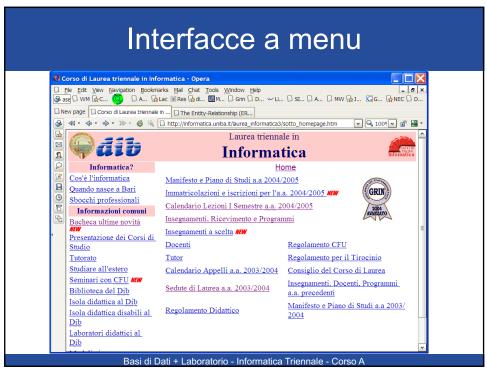
Basi di Dati + Laboratorio - Informatica Triennale - Corso A

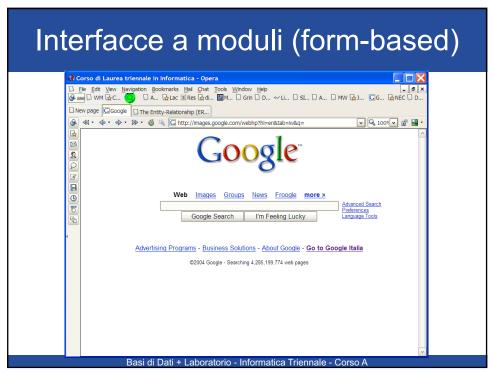
110

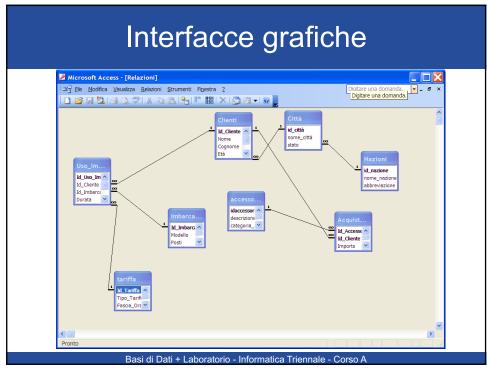
DBMS Interfaces

- Stand-alone query language interfaces
- Programmer interfaces for embedding DML in programming languages:
 - Pre-compiler Approach
 - Procedure (Subroutine) Call Approach
- User-friendly interfaces
 - Menu-based, popular for browsing on the web
 - Forms-based, designed for naïve users
 - Graphics-based (Point and Click, Drag and Drop etc.)
 - Natural language: requests in written English
 - Combinations of the above

Basi di Dati + Laboratorio - Informatica Triennale - Corso A







Other DBMS Interfaces

- Speech as Input (?) and Output
- · Web Browser as an interface
- Parametric interfaces (e.g., bank tellers) using function keys
- Interfaces for the DBA:
 - · Creating accounts, granting authorizations
 - Setting system parameters
 - · Changing schemas or access path

Basi di Dati + Laboratorio - Informatica Triennale - Corso A



Modelli logici dei dati

- Tre modelli logici tradizionali
 - gerarchico
 - reticolare
 - relazionale
- Modelli più recenti
 - a oggetti
 - XML

Basi di Dati + Laboratorio - Informatica Triennale - Corso A

Modelli logici, caratteristiche

- · Gerarchico e reticolare
 - · utilizzano riferimenti espliciti (puntatori) fra record
- Relazionale
 - · basato su valori
 - anche i riferimenti fra dati in strutture (relazioni) diverse sono rappresentati per mezzo dei valori stessi

Basi di Dati + Laboratorio - Informatica Triennale - Corso A

118

Il modello relazionale

- Proposto da E. F. Codd nel 1970 per favorire l'indipendenza dei dati e disponibile in DBMS reali nel 1981
- Si basa sul concetto matematico di relazione (con una variante)
- Le relazioni hanno naturale rappresentazione per mezzo di tabelle

Basi di Dati + Laboratorio - Informatica Triennale - Corso A

Relazione: tre accezioni

- relazione matematica
 - · come nella teoria degli insiemi
- relazione (dall'inglese relationship)
 - · rappresenta una classe di fatti, nel modello Entity-Relationship; tradotto anche con associazione o correlazione
- relazione
 - · secondo il modello relazionale dei dati

120

Relazione matematica, esempio

- $D_1 = \{a,b\}$
- $D_2 = \{x, y, z\}$
- prodotto cartesiano D₁ × D_2

а	Χ
а	У
а	Z
b	Χ
b	У
b	Z

una relazione $r \subseteq D_1 \times D_2$

$$r \subseteq D_1 \times D_2$$

а	Χ
а	Z
b	V

Relazione matematica

- D₁, ..., D_n
 - n insiemi, anche non distinti
- prodotto cartesiano D₁×...×D_n
 - insieme di tutte le n-uple $(d_1, \, ..., \, d_n)$ tali che $d_1 \! \in \! D_1, \, ..., \, d_n \in D_n$
- relazione matematica su D₁, ..., D_n:
 - un sottoinsieme di $D_1 \times ... \times D_n$.
- D₁, ..., D_n domini della relazione

Basi di Dati + Laboratorio - Informatica Triennale - Corso A

122

Relazione matematica, proprietà

- una relazione matematica
 - · insieme di n-uple ordinate
 - $(d_1, ..., d_n)$ tali che $d_1 \in D_1, ..., d_n \in D_n$
- una relazione è un insieme
 - non c'è ordinamento fra le n-uple
 - · le n-uple sono distinte
 - · ciascuna n-upla è ordinata
 - · l'i-esimo valore proviene dall'i-esimo dominio

Basi di Dati + Laboratorio - Informatica Triennale - Corso A

Relazione matematica, esempio

Partite ⊆ squadre × squadre × reti × reti

Juve Lazio 3 1 Lazio Milan 2 0 Juve Roma 0 2 Roma Milan 0 1

- Ciascuno dei domini ha due ruoli diversi, distinguibili attraverso la posizione
 - la struttura è posizionale

Basi di Dati + Laboratorio - Informatica Triennale - Corso A

124

Struttura non posizionale

 A ciascun dominio si associa un nome unico nella tabella (attributo), che ne descrive il "ruolo"

Casa	Fuori	RetiCasa	RetiFuori
Juve	Lazio	3	1
Lazio	Milan	2	0
Juve	Roma	0	2
Roma	Milan	0	1

Basi di Dati + Laboratorio - Informatica Triennale - Corso A

Tabelle e relazioni

- Una tabella rappresenta una relazione se
 - · le righe sono diverse fra loro
 - · le intestazioni delle colonne sono diverse tra loro
 - i valori di ogni colonna sono fra loro omogenei
- In una tabella che rappresenta una relazione
 - · l'ordinamento tra le righe è irrilevante
 - · l'ordinamento tra le colonne è irrilevante

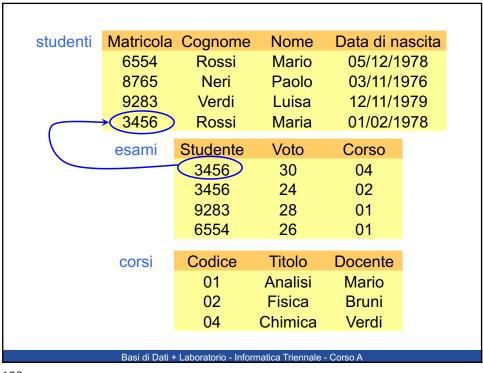
Basi di Dati + Laboratorio - Informatica Triennale - Corso A

126

Il modello è basato su valori

 I riferimenti fra dati in relazioni diverse sono rappresentati per mezzo di valori dei domini che compaiono nelle ennuple

Basi di Dati + Laboratorio - Informatica Triennale - Corso A

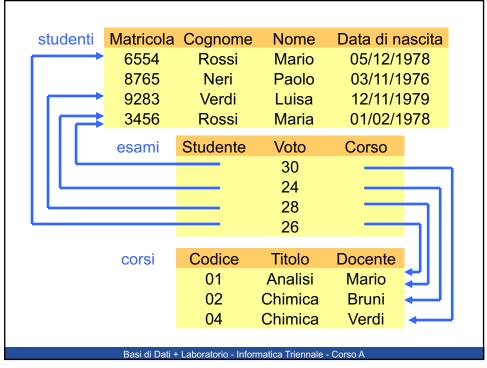


Alternativa

 Altri modelli (sia quelli "storici", reticolare e gerarchico, sia quello a oggetti) prevedono riferimenti espliciti, gestiti dal sistema

Basi di Dati + Laboratorio - Informatica Triennale - Corso A

studenti	Matricola	Cognome	Nome	Data di na	ascita
	6554	Rossi	Mario	05/12/1	978
	8765	Neri	Paolo	03/11/19	976
	9283	Verdi	Luisa	12/11/19	979
	3456	Rossi	Maria	01/02/1	978
	esami	Studente	Voto	Corso	
		3456	30	04	
		3456	24	02	
		9283	28	01	
		6554	26	01	
	,				
	corsi	Codice	Titolo	Docente	
		01	Chimica	Mario	
		02	Analisi	Bruni	
		04	Chimica	Verdi	
	,				
	Basi di Dati +	- Laboratorio - Infor	matica Triennale ·	- Corso A	



Struttura basata su valori: vantaggi

- indipendenza dalle strutture fisiche (si potrebbe avere anche con puntatori di alto livello) che possono cambiare dinamicamente
- si rappresenta solo ciò che è rilevante dal punto di vista dell'applicazione
- l'utente finale vede gli stessi dati dei programmatori
- i dati sono portabili piu' facilmente da un sistema ad un altro
- i puntatori sono direzionali

Basi di Dati + Laboratorio - Informatica Triennale - Corso A

132

Definizioni

Schema di relazione:

un nome R con un insieme di attributi

$$A_1, ..., A_n$$

$$R(A_1,..., A_n)$$

Studente(Matricola, Cognome, Nome)

Schema di base di dati.

insieme di schemi di relazione:

$$R = \{R_1(X_1), ..., R_k(X_k)\}$$

Basi di Dati + Laboratorio - Informatica Triennale - Corso A

Definizioni, 2

- Una tupla su un insieme di attributi X è una funzione che associa a ciascun attributo A in X un valore del dominio di A
- t[A] denota un valore della tupla t sull'attributo A
- t[cognome, nome] = "Rossi", "Mario"

Basi di Dati + Laboratorio - Informatica Triennale - Corso A

134

Definizioni, 3

- (Istanza di) relazione su uno schema R(X):
 - insieme r di tuple su X
- $r=\{t_1[X], ..., t_n[X]\}$
- (Istanza di) base di dati su uno schema R= {R₁(X₁), ..., R_k(X_k)}: insieme di relazioni r = {r₁,..., r_k} (con r_i relazione su R_i)

Basi di Dati + Laboratorio - Informatica Triennale - Corso A

Relazioni su singoli attributi studenti Matricola Cognome Nome Data di nascita 6554 Rossi Mario 05/12/1978 8765 Neri Paolo 03/11/1976 9283 Verdi 12/11/1979 Luisa 01/02/1978 3456 Rossi Maria studenti lavoratori Matricola 6554 3456

136

Strutture nidificate

DA FILIPPO					
	VIA ROMA 2, R	ОМА			
	RICEVUTA FISC	CALE			
•	1235 <i>DEL</i> 12/10/	2002			
3	Coperti	3,00			
2	Antipasti	6,20			
3	Primi	12,00			
2	Bistecche	18,00			
	<i>Totale</i> 39,20				

DA FILIPPO			
	VIA ROMA 2, R	OMA	
	RICEVUTA FISC	CALE	
•	1240 <i>DEL</i> 13/10/	2002	
2	Coperti	2,00	
2	Antipasti	7,00	
2	Primi	8,00	
2	Orate	20,00	
2	Caffè	2,00	
<i>Totale</i> 39,00			

Basi di Dati + Laboratorio - Informatica Triennale - Corso A

Strutture nidificate

	DA FILIPPO VIA ROMA 2, ROMA			
	<i>RICEVUTA FISCALE</i> 1235 <i>DEL</i> 12/10/2002			
3	Coperti	3,00		
2	Antipasti	6,20		
3	Primi	12,00		
2	Bistecche	18,00		
	TOTALE	39,20		

DA FILIPPO VIA ROMA 2, ROMA				
RICEVUTA FISCALE 1240 DEL 13/10/2002				
2	Coperti	2,00		
2	Antipasti	7,00		
2	Primi	8,00		
2	Orate	20,00		
2 Caffè		2,00		
<i>Totale</i> 39,00				

Basi di Dati + Laboratorio - Informatica Triennale - Corso A

138

Strutture nidificate

	DA FILIPPO VIA ROMA 2, ROMA				
	<i>RICEVUTA FISCALE</i> 1235 <i>DEL</i> 12/10/2002				
3	Coperti	3,00			
2	Antipasti	6,20			
3	Primi	12,00			
2	Bistecche	18,00			
	TOTALE	39,20			

DA FILIPPO VIA ROMA 2, ROMA				
RICEVUTA FISCALE 1240 DEL 13/10/2002				
2	Coperti	2,00		
2	Antipasti	7,00		
2	Primi	8,00		
2	Orate	20,00		
2	Caffè	2,00		
<i>Totale</i> 39,00				

Basi di Dati + Laboratorio - Informatica Triennale - Corso A

Strutture nidificate					
Ricevute)				
Numero	Data	Qtà	Descrizione	Importo	Totale
1235	12/10/2002	3	Coperti	3,00	39,20
		2	Antipasti	6,20	
		3	Primi	12,00	
		2	Bistecche	18,00	
1240	13/10/2002	2	Coperti	2,00	39,00
• Ma	i valori dev	ono e	ssere semplio	ci, non rel	lazioni!
	Basi di Dati	+ Laborato	rio - Informatica Triennale -	· Corso A	

Relazioni che rappresentano strutture nidificate						
Ricevute	Numero		Data	Totale		
	1235	12/	10/2002	39,20		
	1240	13/	10/2002	39,00		
Dettaglio	Numero	Qtà	Descrizio	ono In	nnorto	
Dottagno	1235	3			nporto	
			Coper		3,00	
	1235	2	Antipa	stı	6,20	
	1235	3	Primi	i	12,00	
	1235	2	Bistecc	he	18,00	
	1240	2	Coper	ti	2,00	

Strutture nidificate, riflessione

- Abbiamo rappresentato veramente tutti gli aspetti delle ricevute?
- Dipende da che cosa ci interessa!
 - l'ordine delle righe e' rilevante?
 - possono esistere linee ripetute in una ricevuta?
- Sono possibili rappresentazioni diverse

Basi di Dati + Laboratorio - Informatica Triennale - Corso A

142

Rappresentazione alternativa per strutture nidificate							
Ricevu	te	Nume	ro	Data	3	Totale	
		1235	5	12/10/2	002	39,20	
		1240)	13/10/2	002	39,00	
Dettaglio	N	umero	Riga	Qtà	Des	crizione	Importo
		1235	1	3	С	operti	3,00
		1235	2	2	Ar	ntipasti	6,20
		1235	3	3	ı	Primi	12,00
		1235	4	2	Bis	stecche	18,00
		1240	1	2	С	operti	2,00
Basi di Dati + Laboratorio - Informatica Triennale - Corso A							

Informazione incompleta

- Il modello relazionale impone ai dati una struttura rigida:
 - le informazioni sono rappresentate per mezzo di ennuple
 - solo alcuni formati di ennuple sono ammessi: quelli che corrispondono agli schemi di relazione
- I dati disponibili possono non corrispondere al formato previsto

Basi di Dati + Laboratorio - Informatica Triennale - Corso A

144

Informazione incompleta: motivazioni

Nome	SecondoNome	Cognome
Franklin	Delano	Roosevelt
Winston		Churchill
Charles		De Gaulle
Josip		Stalin

Basi di Dati + Laboratorio - Informatica Triennale - Corso A

Informazione incompleta: soluzioni?

- Non conviene (anche se spesso si fa) usare valori del dominio (0, stringa nulla, "99", ...):
 - potrebbero non esistere valori "non utilizzati"
 - valori "non utilizzati" potrebbero diventare significativi
 - in fase di utilizzo (nei programmi) sarebbe necessario ogni volta tener conto del "significato" di questi valori

Basi di Dati + Laboratorio - Informatica Triennale - Corso A

146

Informazione incompleta nel modello relazionale

- Tecnica rudimentale ma efficace:
 - valore nullo: denota l'assenza di un valore del dominio (e non è un valore del dominio)
- t[A], per ogni attributo A, è un valore del dominio dom(A) oppure il valore nullo NULL
- Si possono (e debbono) imporre restrizioni sulla presenza di valori nulli

Basi di Dati + Laboratorio - Informatica Triennale - Corso A

Tipi di valore nullo

- (Almeno) tre casi differenti
 - valore sconosciuto
 - valore inesistente
 - valore senza informazione
- I DBMS non distinguono i tipi di valore nullo

Basi di Dati + Laboratorio - Informatica Triennale - Corso A

148

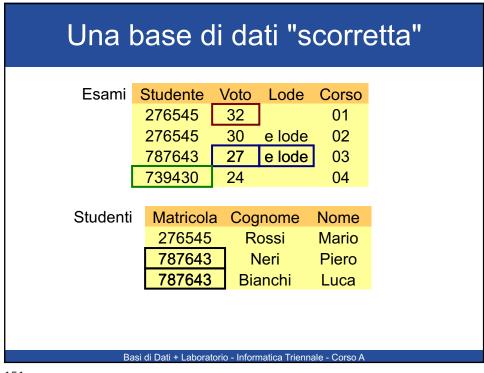
Studenti Matricola Cognome Nome Data di nascita 6554 Rossi Mario 05/12/1978 9283 Verdi Luisa 12/11/1979 NULL Rossi Maria 01/02/1978 esami Studente Voto Corso NULL 30 NULL NULL 24 02 9283 28 01 corsi Codice Titolo Docente 01 Analisi Mario 02 NULL NULL 04 Chimica Verdi	Troppi valori nulli						
esami Studente Voto Corso NULL 30 NULL NULL 24 02 9283 28 01 corsi Codice Titolo Docente 01 Analisi Mario 02 NULL NULL	studenti	6554 9283	Rossi Verdi	Mario Luisa	05/12/1 12/11/1	978 979	
01 Analisi Mario 02 NULL NULL			Studente NULL NULL	Voto 30 24	Corso NULL 02	970	
		corsi	01 02	Analisi NULL	Mario NULL		

Vincoli di integrità

· Esistono istanze di basi di dati che, pur sintatticamente corrette, non rappresentano informazioni possibili per l'applicazione di interesse

Basi di Dati + Laboratorio - Informatica Triennale - Corso A

150



esercizio...

struct tab1{

char cognome struct tab3{
char nome int codice

int matricola int matricola }studente; char data struct tab2{ int voto

char nome }esame;

int codice

int anno Typedef struct

int semestre database.studente

studente;

}corso;

Basi di Dati + Laboratorio - Informatica Triennale - Corso A

152

Vincolo di integrità

- Proprietà che deve essere soddisfatta dalle istanze che rappresentano informazioni corrette per l'applicazione
- Un vincolo è una funzione booleana (un predicato):

associa ad ogni istanza il valore vero o falso

Basi di Dati + Laboratorio - Informatica Triennale - Corso A

Vincoli di integrità, perché?

- descrizione più accurata della realtà
- contributo alla "qualità dei dati"
- utili nella progettazione (vedremo)
- usati dai DBMS nella esecuzione delle interrogazioni

Basi di Dati + Laboratorio - Informatica Triennale - Corso A

154

Vincoli di integrità, nota

- alcuni tipi di vincoli (ma non tutti) sono "supportati" dai DBMS:
 - possiamo quindi specificare vincoli di tali tipi nella nostra base di dati e il DBMS ne impedisce la violazione
- per i vincoli "non supportati", la responsabilità della verifica è dell'utente o del programmatore

Basi di Dati + Laboratorio - Informatica Triennale - Corso A

Tipi di vincoli

- vincoli intrarelazionali
 - vincoli su valori (o di dominio)
 - vincoli di tupla
- vincoli interrelazionali

Basi di Dati + Laboratorio - Informatica Triennale - Corso A

156

Lode Corso Esami Studente Voto 276545 32 01 276545 30 e lode 02 27 787643 e lode 03 739430 24 04 Matricola Cognome Studenti Nome 276545 Rossi Mario 787643 Piero Neri 787643 Bianchi Luca

Vincoli di tupla

- Esprimono condizioni sui valori di ciascuna tupla, indipendentemente dalle altre ennuple
- Caso particolare:
 - Vincoli di dominio: coinvolgono un solo attributo

Basi di Dati + Laboratorio - Informatica Triennale - Corso A

158

Sintassi ed esempi

- Una possibile sintassi:
 - espressione booleana di atomi che confrontano valori di attributo o espressioni aritmetiche su di essi

(Voto
$$\geq$$
 18) AND (Voto \leq 30)

Basi di Dati + Laboratorio - Informatica Triennale - Corso A

Vincoli di tupla, altro esempio

Stipendi

Impiegato	Lordo	Ritenute	Netto
Rossi	55.000	12.500	42.500
Neri	45.000	10.000	35.000
Bruni	50.000	11.000	36.000

Lordo = (Ritenute + Netto)

Basi di Dati + Laboratorio - Informatica Triennale - Corso A

160

Vincoli di tupla, violazione

Stipendi

Impiegato	Lordo	Ritenute	Netto
Rossi	55.000	12.500	42.500
Neri	45.000	10.000	35.000
Bruni	50.000	11.000	36.000

Lordo = (Ritenute + Netto)

Basi di Dati + Laboratorio - Informatica Triennale - Corso A