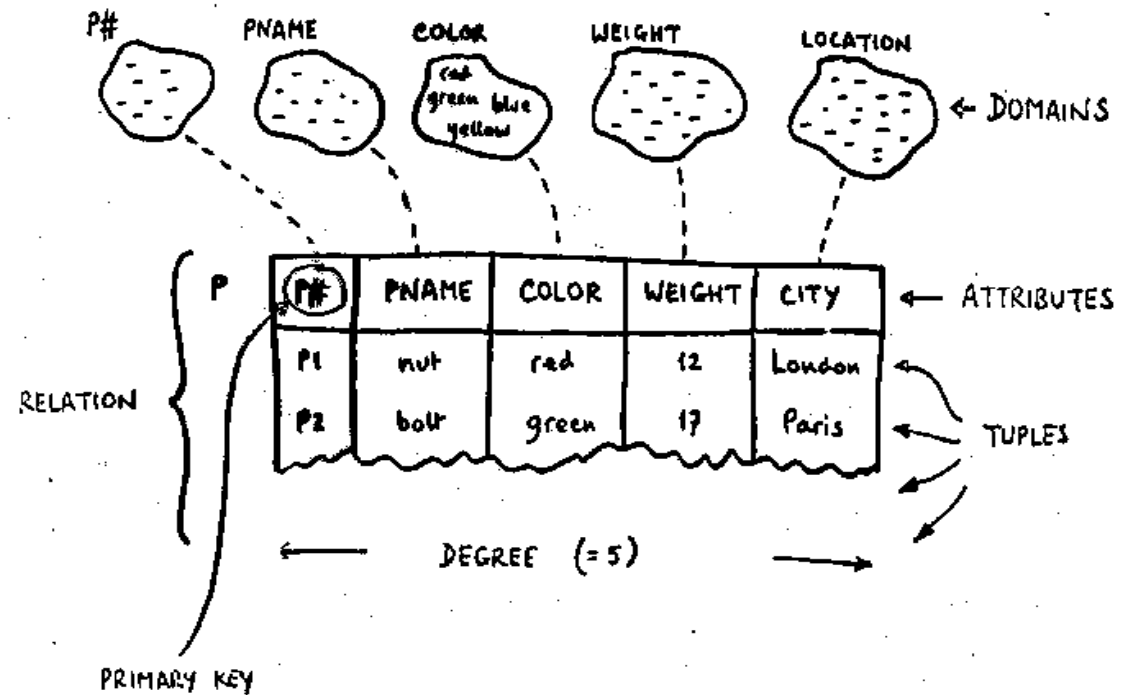


TERMINOLOGY



Il modello relazionale

Annalisa Franco, Dario Maio
Università di Bologna

Introduzione al modello relazionale

- Il modello relazionale fu introdotto nel 1970 da E.F. Codd (presso i laboratori IBM di San Jose, CA) allo scopo di favorire l'indipendenza dei dati.
- I modelli preesistenti, gerarchico (hierarchical model) e reticolare (network model) erano fortemente influenzati da considerazioni di natura fisica, che enfatizzavano maggiormente gli aspetti di efficienza piuttosto che la semplicità d'uso.
- Il sistema di gestione IMS, basato sul modello gerarchico, fu sviluppato originariamente nel 1966 da IBM insieme a Rockwell e Caterpillar per il programma Apollo. È tuttora operativo per ambienti OLTP industriali.
- Il modello reticolare è stato introdotto da Charles Bechman e standardizzato dal consorzio CODASYL nel 1969.
- Rispetto ai modelli gerarchico e reticolare il modello relazionale si caratterizza per:
 - la totale assenza di legami costruiti con puntatori; nel modello relazionale, infatti, si fa uso solo di valori;
 - la presenza di una teoria utile per la progettazione di DB, per la definizione di linguaggi d'interrogazione e per l'ottimizzazione di query.

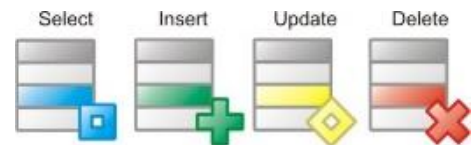
Un po' di storia...

- Lo sviluppo di efficienti DBMS basati sul modello relazionale è stato piuttosto lento, a causa dell'elevato livello d'astrazione offerto rispetto ai modelli precedenti. Le prime soluzioni commerciali risalgono alla prima metà degli anni 80.

Anni 70: definizione del modello, prima versione del linguaggio **SQL** (Structured Query Language, inizialmente denominato **SEQUEL**), studi fondamentali sulla tecnologia relazionale (ottimizzazione, transazioni, recovery, ...) e primi prototipi di DBMS relazionali (RDBMS): System R (IBM, laboratorio di ricerca di San Jose, CA, USA), Ingres (Università di Berkeley, CA, USA)

Anni 80: prima standardizzazione di SQL, primi prototipi commerciali: SQL/DS (derivato da System R), Oracle, IBM DB2

Anni 90: standard ISO-ANSI SQL-2 (anche noto come SQL-92).



- Esiste oggi lo standard ISO-ANSI SQL-3 (o SQL:1999) e sono state definite ulteriori estensioni: SQL 2003, SQL 2008, ecc.
- Nonostante il nome, SQL non è un semplice linguaggio di interrogazione perché alcune istruzioni sono dedicate alla creazione, alla gestione e all'amministrazione del database.

Modello logico relazionale: note

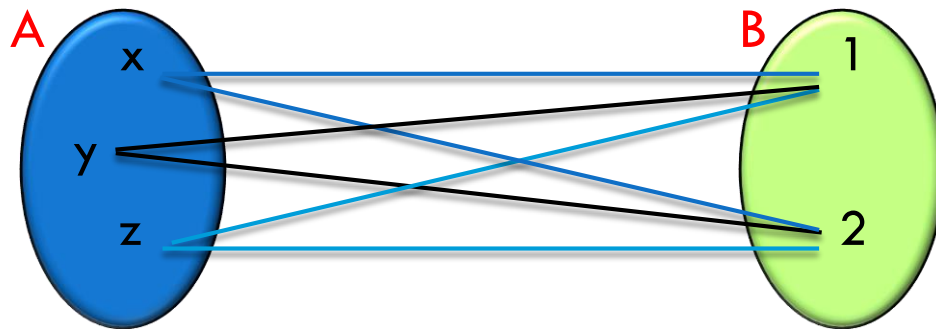
- Il modello relazionale è un **modello logico** nel senso che risponde al requisito di indipendenza dalla particolare rappresentazione dei dati adottata a livello fisico.
- Nel contesto di un DB relazionale, gli utenti che accedono ai dati e i programmatori che sviluppano applicazioni fanno riferimento **solo al livello logico**, senza specificare i percorsi di accesso per eseguire le operazioni.
- I modelli logici **gerarchico** (rappresentazione con **strutture ad albero**) e **reticolare** (rappresentazione con **strutture a grafo**), ricordano più da vicino tecniche di organizzazione dei dati a livello fisico e richiedono al programmatore di esplicitare i cammini per accedere ai dati.
- Nel modello relazionale l'unica astrazione è il concetto di **relazione**; non vi sono costrutti concettuali di alto livello in grado di descrivere entità, associazioni, generalizzazioni, specializzazioni, aggregazioni.
- Tuttavia, una relazione può rappresentare opportunamente ed efficientemente i suddetti concetti astratti. Ciò motiva la necessità di ricorrere, nella fase di progettazione logica, a una traduzione da schemi concettuali E/R a schemi logici relazionali.

Sul termine “relazione”

- Il termine “**relazione**” può essere usato con diverse accezioni, che non devono essere confuse tra loro:
 - ▣ nel linguaggio comune denota un “legame” di qualche tipo;
 - ▣ nella teoria degli insiemi denota una “**relazione matematica**”;
 - ▣ nel modello relazionale è una **generalizzazione della relazione matematica**;
 - ▣ nel modello Entity-Relationship denota una **classe di legami fra entità** (sono sinonimi “**associazione**” e “**correlazione**”);
 - ▣ nei DBMS relazionali è usato spesso, a volte erroneamente, come sinonimo di “**tabella**”.
- Per introdurre il modello relazionale è quindi opportuno innanzitutto riesaminare il concetto di relazione matematica.

Richiami: prodotto cartesiano

- Si considerino due insiemi A e B , non vuoti e non necessariamente distinti; si definisce **prodotto cartesiano** $A \times B$ come l'insieme delle coppie ordinate (a,b) con $a \in A$ e $b \in B$. Per definizione $A \times \emptyset = \emptyset \times A = \emptyset$.
- Se A e B sono insiemi distinti il prodotto $A \times B$ è formalmente diverso da $B \times A$, anche se i due prodotti sono in naturale corrispondenza biunivoca.



Esempio:

$$A = \{x, y, z\}, B = \{1, 2\};$$

$$A \times B = \{(x,1), (x,2), (y,1), (y,2), (z,1), (z,2)\}.$$

- La definizione può essere estesa considerando $n > 0$ insiemi D_1, D_2, \dots, D_n , non necessariamente distinti.
- Il **prodotto cartesiano** $D_1 \times D_2 \times \dots \times D_n$ è l'insieme di tutte le ennuple (n-ple) ordinate (d_1, d_2, \dots, d_n) tali che $d_1 \in D_1, d_2 \in D_2, \dots, d_n \in D_n$.
- Il prodotto cartesiano di n copie dell'insieme D si indica con D^n ed è detto **potenza cartesiana**.

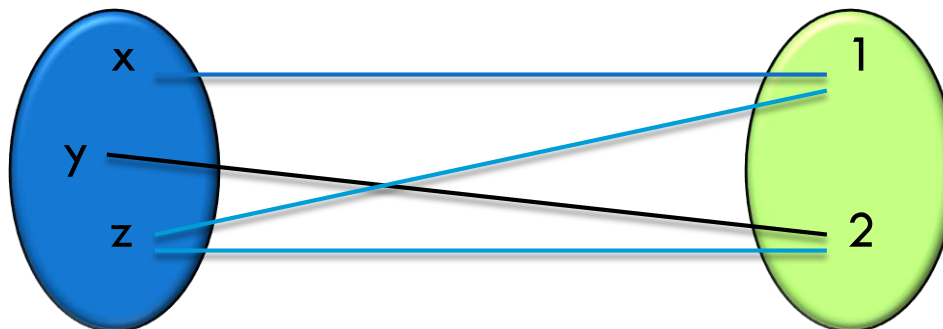
Relazione matematica (binaria)

- Si considerino due insiemi A e B , non vuoti e non necessariamente distinti; ogni sottoinsieme non vuoto del prodotto cartesiano $A \times B$ è detto relazione da A a B . Se $B = A$ allora un sottoinsieme non vuoto del prodotto cartesiano A^2 è detto anche relazione in A o su A .
- Data una relazione $r \subseteq A \times B$, si dice che l'elemento $a \in A$ è in relazione con l'elemento $b \in B$ se la coppia $(a,b) \in r$.

Esempio: $A = \{x,y,z\}$, $B = \{1,2\}$;

$A \times B = \{(x,1),(x,2),(y,1),(y,2),(z,1),(z,2)\}$

$r = \{(x,1),(y,2),(z,1),(z,2)\}$ è una relazione su A e B .

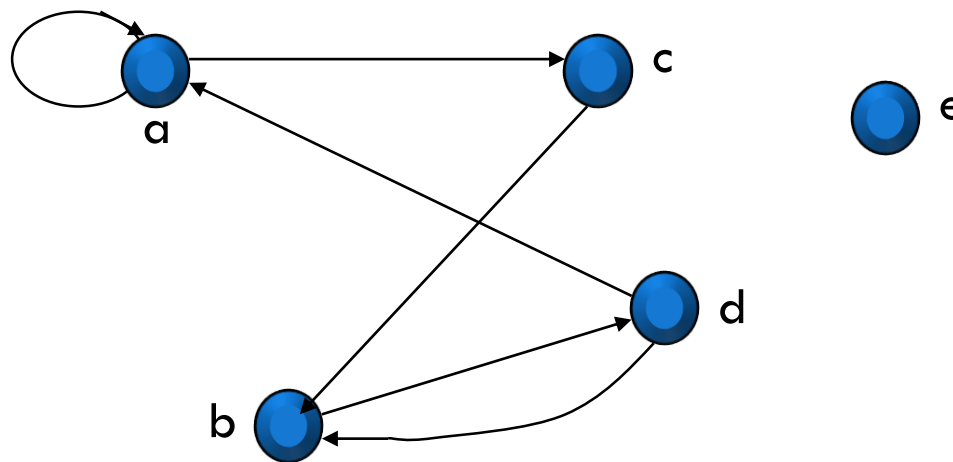


Un esempio d'uso

- Un **grafo orientato** è una coppia $(A; r)$ dove A è un insieme finito e non vuoto e r è una **relazione in A** .

Esempio: $A = \{a, b, c, d, e\}$;

$r = \{(a, a), (a, c), (b, d), (c, b), (d, a), (d, b)\}$



Relazione matematica n-aria

- Si considerino $n > 0$ insiemi D_1, D_2, \dots, D_n , non necessariamente distinti.
- Una **relazione (matematica)** su D_1, D_2, \dots, D_n è un qualunque **sottoinsieme del prodotto cartesiano** $D_1 \times D_2 \times \dots \times D_n$.

Esempio:

$D_1 = \{\text{mela}, \text{pera}\}; D_2 = \{1, 2, 3\}; D_3 = \{1, 2, 3, 4\};$

$r = \{(\text{mela}, 1, 1), (\text{mela}, 3, 1), (\text{pera}, 1, 3), (\text{pera}, 2, 2), (\text{pera}, 3, 4)\}$

è una **relazione** su D_1, D_2, D_3 ($r \subseteq D_1 \times D_2 \times D_3$)

- D_1, D_2, \dots, D_n sono i **domini** della relazione;
- il valore di n è detto **grado** (o “**arity**”) della relazione;
- il numero di n-ple di una relazione è la sua **cardinalità**.

Relazione matematica: proprietà

- Una relazione è un insieme di n-ple...:
 - tutte le n-ple sono distinte tra loro;
 - non è definito alcun ordinamento tra le diverse n-ple;
 - $D_1 = \{a,b,c\}$; $D_2 = \{1,2,3\}$;
 - $\{(a,1),(b,2),(c,1),(c,2)\} = \{(a,1),(b,2),(c,2),(c,1)\} = \{(b,2),(c,2),(c,1),(a,1)\}$
- ... ciascuna considerata al proprio interno ordinata rispetto ai domini ...:
 - l'ordine in cui si considerano i domini è rilevante;
 - $(D_1 \times D_2) \neq (D_2 \times D_1)$
 - $\{(a,1),(c,1),(c,2)\} \neq \{(1,a),(1,c),(2,c)\}$
- ...su domini non necessariamente distinti:
 - uno stesso dominio può essere usato in più posizioni;
 - $\{(2,a,1),(1,c,1),(1,c,2)\} \subseteq D_2 \times D_1 \times D_2$.

Relazione matematica: note

- Una relazione è un insieme di n-ple...: la definizione estesa di prodotto cartesiano contempla anche il caso di $n = 1$, sottintendendo che il prodotto cartesiano coincida con l'insieme $D1$.

Esempio

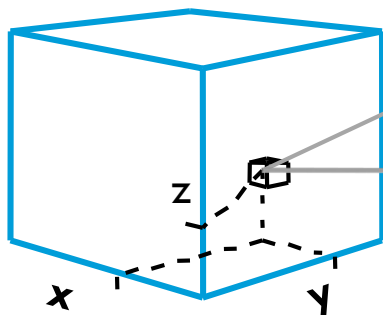
$D1 = \{\text{Carlo, Mario, Giacomo, Marco, Giorgio}\};$
 $r = \{(\text{Carlo}), (\text{Giorgio})\}$ è dunque una relazione!

- La definizione contempla anche:
 - relazioni con un numero infinito di n-ple anche se ai fini pratici, a causa della dimensione finita della memoria di un elaboratore, le relazioni sono necessariamente costituite da un numero finito di n-ple; tuttavia è possibile a volte “gestire” anche relazioni con numero infinito di n-ple, se descrivibili attraverso un algoritmo finito;
 - domini infiniti, e ciò è utile per definire valori ammissibili anche se non presenti nella base di dati.

Rappresentazione di relazioni

- Rappresentazione **insiemistica**; non adeguata per relazioni di grado $n > 2$ e/o con cardinalità superiore a qualche unità.
- Rappresentazione **tabellare**; efficace e intuitiva.

mela	1	1
mela	3	1
pera	1	3
pera	2	2
pera	3	4



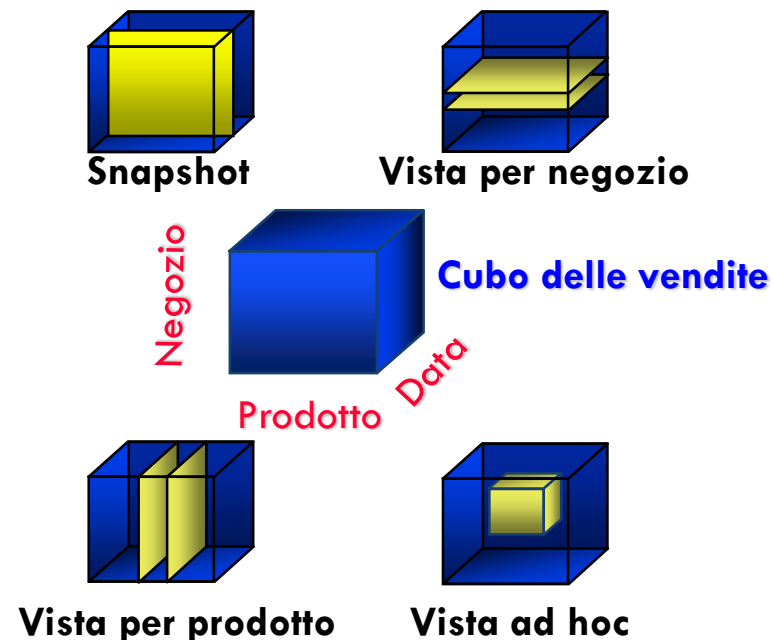
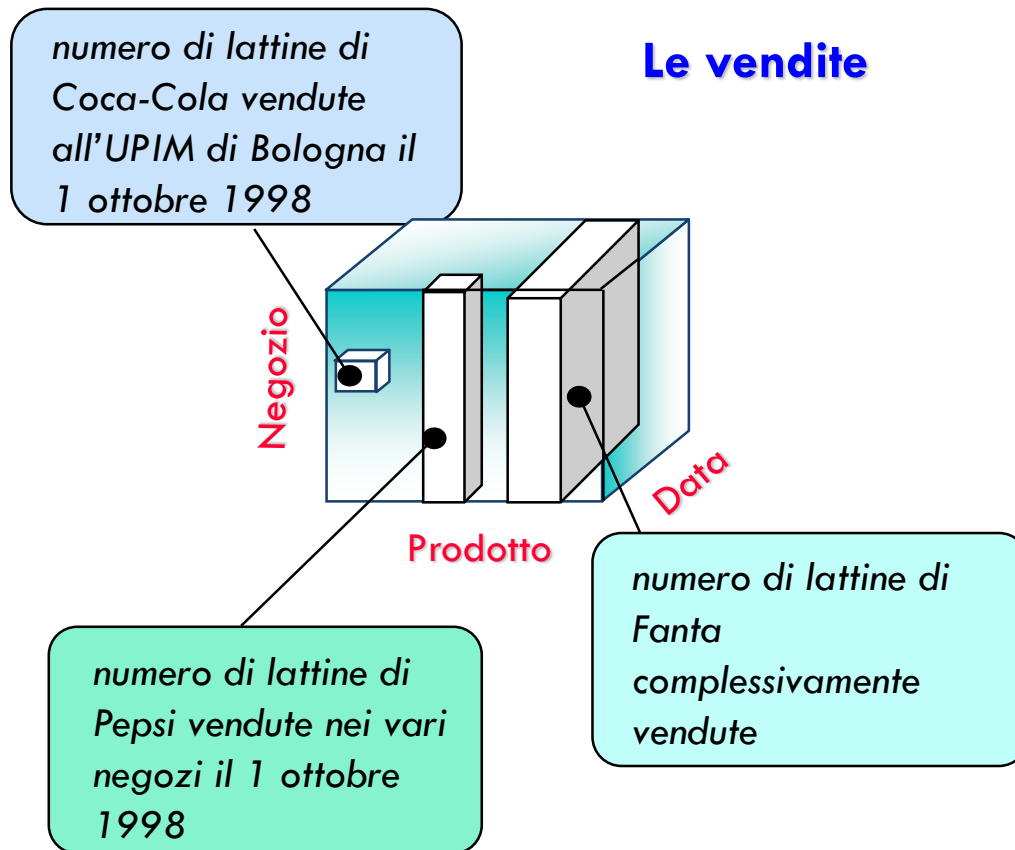
1 se la tripla (x,y,z) appartiene alla relazione,
0 se non vi appartiene.

- Rappresentazione multi-dimensionale; adatta se il grado n è minore o uguale a 3:

Rappresentazione multidimensionale: note

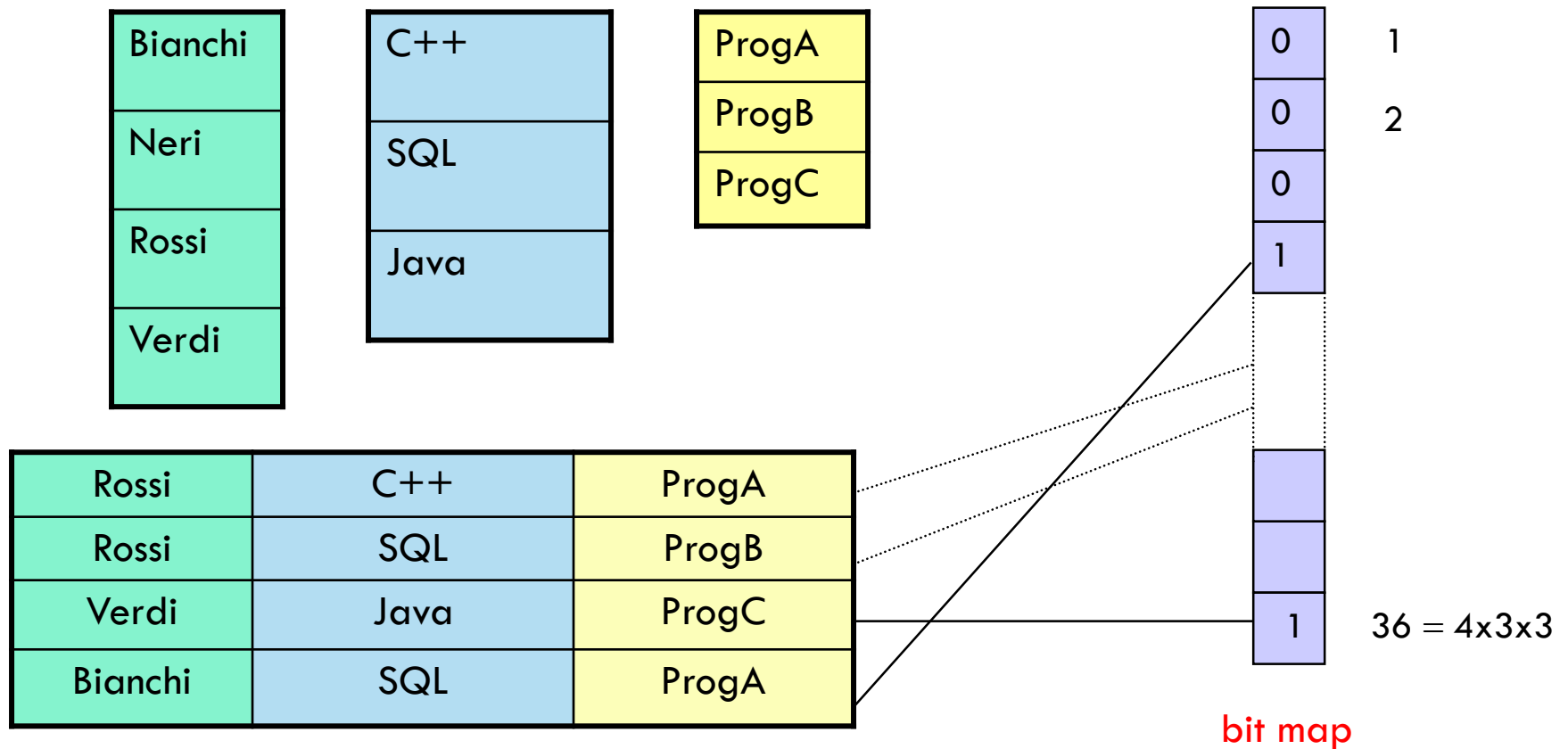
- Una rappresentazione multi-dimensionale può essere utile per evidenziare alcune viste sui dati; molto usata nei data warehouse.

Le vendite



E ancora altre rappresentazioni...

- Ad esempio: una **bit map**



L'importanza della posizione

- Nel caso di domini ripetuti, l'interpretazione dei dati si complica e la posizione assume un ruolo determinante. La relazione indicata in figura è un sottoinsieme del prodotto cartesiano:

String \times String \times Integer \times Integer

Enel Brindisi	Sidigas Avellino	92	88
MIA Cantù	Virtus Bologna	94	87
Fiat Torino	Vanoli Cremona	88	80
The Flex Pistoia	Consultinvest Pesaro	86	83

- Il primo e il terzo dominio si riferiscono alla squadra ospitante (nome e numero di punti), mentre il secondo e il quarto dominio si riferiscono alla squadra ospitata.
- Questa notazione è scomoda e poco chiara.

Relazione nel modello relazionale

- A ogni occorrenza di dominio (ripetuto o meno) si associa un nome univoco nella relazione, detto attributo, il cui compito è specificare il ruolo che quel dominio svolge nella relazione (“cosa significa”).
- Nella rappresentazione tabellare, gli attributi sono le intestazioni delle colonne (e in quella multi-dimensionale sono i nomi degli assi).

TeamCasa	TeamOspite	PuntiCasa	PuntiOspite
Enel Brindisi	Sidigas Avellino	92	88
MIA Cantù	Virtus Bologna	94	87
Fiat Torino	Vanoli Cremona	88	80
The Flex Pistoia	Consultinvest Pesaro	86	83

- La struttura non è più posizionale, ovvero l'ordine degli attributi non ha più rilevanza. In questo modo si supera il problema della non commutatività del prodotto cartesiano.

Relazione: una definizione formale

- Si indichi con $\text{dom}(A)$ il dominio dell'attributo A e si consideri un insieme di attributi $X = \{A_1, A_2, \dots, A_n\}$.
- Una **tupla** t su X è una funzione che associa a ogni $A_i \in X$ un valore di $\text{dom}(A_i)$.
- Uno **schema di relazione** su X è definito da un **nome** (della relazione) R e **dall'insieme di attributi** X , e si indica con $R(X)$.
- Uno **stato** (o **estensione**) di relazione su X è un insieme r di tuple su X , che è anche denominato semplicemente "**relazione**".



PARTITE

TeamCasa	TeamOspite	PuntiCasa	PuntiOspite
Enel Brindisi	Sidigas Avellino	92	88
MIA Cantù	Virtus Bologna	94	87
Fiat Torino	Vanoli Cremona	88	80
The Flex Pistoia	Consultinvest Pesaro	86	83

Notazione di base

- Per denotare insiemi di attributi si usa la notazione semplificata:
 - A in luogo di $\{A\}$ e XY in luogo di $X \cup Y$
 - ...e si scrive ABC (o A,B,C) anziché $\{A,B,C\}$
 - ...e quindi R(ABC) o R(A,B,C) anziché $R(\{A, B, C\})$.
- Se t è una tupla su X e $A \in X$, allora $t[A]$ o t.A è il valore di t su A.

PARTITE

	TeamCasa	TeamOspite	PuntiCasa	PuntiOspite
	Enel Brindisi	Sidigas Avellino	92	88
	MIA Cantù	Virtus Bologna	94	87
t	Fiat Torino	Vanoli Cremona	88	80
	The Flex Pistoia	Consultinvest Pesaro	86	83

$t[\text{TeamOspite}] = t.\text{TeamOspite} = \text{'Vanoli Cremona'}$

- La stessa notazione si usa per insiemi di attributi, e denota una tupla
 - $t[\text{TeamOspite}, \text{PuntiOspite}]$ è una tupla su $\{\text{TeamOspite}, \text{PuntiOspite}\}$

Livelli intensionale ed estensionale

- Uno schema $R(X)$ definisce a livello intensionale una relazione.
Esempio:

PARTITE(TeamCasa, TeamOspite, PuntiCasa, PuntiOspite)
(con le opportune definizioni dei domini e dei vincoli)

- Se è necessario a **livello estensionale**, per riferirsi a un generico stato (estensione) di relazione con schema $R(X)$, si usa semplicemente il nome dello schema in minuscolo, ovvero **r**.
- A volte si usa la notazione $r(X)$ per indicare una relazione su X , descrivendo così al tempo stesso lo schema e l'estensione come insieme di tuple.

Ulteriori precisazioni sulla terminologia

- Nella terminologia relazionale il termine “istanza” è sinonimo di “estensione” o “stato” di relazione e non di tupla. Vi è dunque una differenza rispetto alle definizioni adottate per il modello E/R.
- Per questo motivo in questa sede si preferisce usare il termine “stato” o “estensione” .
- Poiché è scomodo precisare tutte le volte la distinzione fra livello intensionale e livello estensionale, a eccezione delle definizioni formali matematiche, si usa lo stesso nome per indicare sia lo schema di relazione sia l'estensione (stato) corrente.
- Ad esempio:
STUDENTI indica l'insieme corrente di tuple nella relazione di schema STUDENTI(Matricola, Cognome, Nome, ...).

Data Base relazionale

- Lo schema di un DB relazionale è un insieme di schemi di relazioni con nomi distinti:

$$R = \{R_1(X_1), R_2(X_2), \dots, R_m(X_m)\} \quad (R_i \neq R_j \quad \forall i \neq j)$$

- Uno stato (o estensione) di un DB con schema

$$R = \{R_1(X_1), R_2(X_2), \dots, R_m(X_m)\}$$

è un insieme di stati di relazioni $r = \{r_1, r_2, \dots, r_m\}$
con r_i stato di relazione su $R_i(X_i)$.

- Esempio: AZIENDA =

{
 IMPIEGATI(Matricola, Cognome, Nome, Livello, Stipendio),
 FILIALI(CodiceFiliale, Nome, Indirizzo, Direttore),
 FORNITORI(RagioneSociale, Indirizzo, PartitaIVA) }

N.B. In realtà la definizione di uno schema di relazione e di uno schema di DB comprende anche l'indicazione di un insieme di vincoli d'integrità.

Uno stato di un semplice DB relazionale

STUDENTI

Matricola	Cognome	Nome	DataNascita
29323	Bianchi	Giorgio	21/06/1978
35467	Rossi	Anna	13/04/1978
39654	Verdi	Marco	20/09/1979
42132	Neri	Lucia	15/02/1978

CORSI

CodCorso	Titolo	CodDocente	Anno
483	Analisi	0201	1
729	Analisi	0021	1
913	Sistemi Informativi	0123	2

ESAMI

Matricola	CodCorso	Voto	Lode
29323	483	28	no
39654	729	30	sì
29323	913	26	no
35467	913	30	sì

DOCENTI

CodDocente	Cognome	Nome	DataNascita
0021	Biondi	Carlo	21/06/1958
.....

Modello basato sui valori: vantaggi

- Nella rappresentazione relazionale i legami fra i dati non sono stabiliti con puntatori ma per mezzo dei valori dei domini che compaiono nelle tuple. I vantaggi, rispetto ad altri modelli (ad es.: gerarchico, reticolare) sono appresso elencati.
 - ▣ Indipendenza dalle strutture fisiche che possono cambiare anche dinamicamente.
 - ▣ Si rappresenta solo ciò che risulta rilevante dal punto di vista dell'applicazione utente; l'uso di puntatori non è molto comprensibile all'utente finale.
 - ▣ Maggiore portabilità dei dati da un sistema all'altro.
 - ▣ I puntatori sono direzionali e pertanto stabiliscono un percorso di navigazione all'interno dei dati.
- Si noti, tuttavia, che a livello fisico l'implementazione di un insieme di relazioni può prevedere l'uso di puntatori, invisibili comunque all'utente applicativo.

Tabelle vs Relazioni

- In realtà i termini “**tabella**” e “**relazione**” non sono affatto sinonimi; una relazione del modello relazionale può essere vista come un particolare tipo di tabella (che Codd chiama R-table).
- Una **tabella** rappresenta una relazione se:
 - ▣ i valori di ciascuna colonna sono tra loro omogenei (definiti sullo stesso dominio);
 - ▣ le righe sono tra loro diverse;
 - ▣ le intestazioni delle colonne sono diverse tra loro.
- In **una tabella che rappresenta una relazione**:
 - ▣ l'ordinamento delle righe è irrilevante;
 - ▣ l'ordinamento delle colonne è irrilevante.
- Il linguaggio SQL nei DBMS commerciali consente di gestire tabelle che non sono relazioni, e che ammettono righe duplicate.

Il problema dei duplicati


- ❑ Esempio di derivazione in SQL di una tabella da una relazione.

STUDENTI

Matricola	Cognome	Nome	DataNascita
2106103423	Bianchi	Giorgio	21/06/1978
2106111021	Rossi	Anna	13/04/1978
1602042312	Rossi	Anna	11/03/1978

- ❑ SELECT Cognome, Nome

FROM STUDENTI ;


una tabella che
non è una relazione

Cognome	Nome
Bianchi	Giorgio
Rossi	Anna
Rossi	Anna

- ❑ SELECT DISTINCT Cognome, Nome

FROM STUDENTI ;


una tabella che
è una relazione

Cognome	Nome
Bianchi	Giorgio
Rossi	Anna

Nome dello schema di una relazione (1)

- Vi sono diverse “scuole di pensiero” a riguardo della convenzione da adottare per il nome di uno schema di relazione. Molto spesso si ricorre all’uso di un sostantivo e alcuni preferiscono indicare il nome al singolare, altri al plurale: **PERSONA** o **PERSONE** , **CORSO** o **CORSI**?
- Naturalmente sono possibili, se opportuno, anche altre categorie grammaticali. Esistono inoltre casi dove il problema “singolare o plurale” non si pone, laddove una parola di per sé indica una pluralità di elementi (es. STAFF, PERSONALE,...).
- La motivazione principale per il **singolare** risiede nel fatto che il nome dello schema esprime un’asserzione che denota il tipo di ogni tupla della relazione in qualunque stato si trovi.
- Chi propende per il **plurale** pensa invece che il nome debba indicare un insieme ovvero una pluralità di elementi, e in questo modo concentra l’attenzione sull’aspetto estensionale della relazione e sulle operazioni che si effettuano. Entrambe le opinioni hanno pro e contro.

Nome dello schema di una relazione (2)

- ❑ Per conformità alla convenzione adottata per gli schemi E/R, si dovrebbe privilegiare il singolare, facilitando il processo di traduzione da E/R verso il relazionale.
- ❑ È doveroso tuttavia citare che alcune linee guida di RDBMS commerciali prediligono fortemente l'uso del plurale.
- ❑ Nei linguaggi a oggetti l'iterazione su una collezione che deriva da una tabella è espressa meglio se la collezione è denominata al plurale e se la classe, che definisce il singolo oggetto (corrispondente a una tupla), ha un nome al singolare.
- ❑ E comunque ciò che è davvero importante è la scelta di nomi che diano immediatamente il significato di “cosa contiene” una relazione, evitando sigle prive di significato o denominazioni non chiare.

Nome di una SQL table

- Qual è la conseguenza del nome ai fini pratici in SQL?

CodImpiegato	Nome	Cognome	Ruolo
E001	Carlo	Rossi	Analista
E003	Mario	Bianchi	Programmatore
E006	Giorgio	Grigi	Sistemista
E007	Carlo	Verdi	Programmatore

IMPIEGATI.Nome

oppure

IMPIEGATO.NOME ?

Select Cognome, Nome
From IMPIEGATI

oppure

Select Cognome, Nome
From IMPIEGATO ?

- In SQL lo statement **CREATE TABLE** definisce al contempo gli attributi di ogni tupla e alloca un certo spazio per ospitare i record che saranno inseriti. Si tratta dunque di una vera e propria **collezione di record**.
- Nel seguito, laddove possibile e sensato, si utilizzeranno sostantivi al plurale, ma ciò comporta altre conseguenze nella fase di progettazione logica

Progettazione logica e nomi delle tabelle

- A livello di schema E/R, ponendo l'accento sui concetti rappresentati abbiamo privilegiato sostantivi al singolare, laddove possibile e sensato. Operando, come vedremo, i passi di progettazione logica, produrremo schemi relazionali e sarebbe naturale riutilizzare in gran parte gli stessi nomi dati per le entità e per le associazioni.
- Se si utilizza un tool automatico di progettazione logica, si potranno avere due comportamenti, a seconda del tipo di software utilizzato: a) sono conservati i nomi usati a livello concettuale; b) i nomi sono trasformati al plurale, con effetti non sempre desiderabili.
- Se s'interagisce con un DB relazionale attraverso un linguaggio a oggetti, avvalendosi eventualmente di strumenti ORM (Object-Relational Mapping) che generano automaticamente le definizioni delle classi, allora si deve operare una scelta che non dipende solo da considerazioni stilistiche.
- In conclusione una soluzione standardizzata e universalmente accettata non esiste, e spesso ci si deve adeguare alle consuetudini utilizzate in un certo ambiente di produzione del software. E, infine, non è sempre così semplice essere consistenti in un settore dove si opera con diversi strumenti di progettazione e implementazione.

1 NF, ovvero solo domini semplici

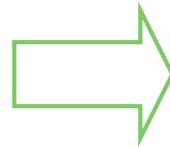
- Il modello relazionale non permette di usare domini arbitrari per la definizione delle relazioni; in particolare **non è in generale possibile usare domini strutturati** (array, set, liste, ...).
 - ▣ Vi sono **eccezioni** notevoli (esempi: **date** e **stringhe**).
- Concisamente, una relazione in cui ogni dominio è “**atomico**” (non ulteriormente decomponibile) si dice che è in
Prima Forma Normale, o **1NF** (First Normal Form)
- In molti casi è pertanto richiesta preliminarmente un’attività di **normalizzazione dei dati** che dia luogo a relazioni in 1NF e che preservi l’informazione originale.

Strutture nidificate: normalizzazione in 1NF

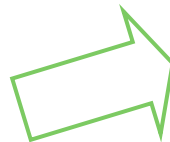
Ricevuta n. 231 del 12/02/2002		
Coperti	2	3,00
Antipasti	1	5,80
Primi	2	11,45
Secondi	2	22,30
Caffè	2	2,20
Vino	1	8,00
Totale (Euro)		52,75

Ricevuta n. 352 del 13/02/2002		
Coperti	1	1,50

RICEVUTE



DETTAGLI



Numero	Data	Totale
231	12/02/2002	52,75
352	13/02/2002	...
...

Numero	Quantità	Descrizione	Prezzo
231	2	Coperti	3,00
231	1	Antipasti	5,80
231	2	Primi	11,45
231	2	Secondi	22,30
231	2	Caffè	2,20
231	1	Vino	8,00
352	1	Coperti	1,50

Considerazioni sulla normalizzazione in 1NF

- Il fatto che una rappresentazione normalizzata sia adeguata o meno dipende (molto) dal contesto.
 - Ad esempio: l'ordine delle righe nella ricevuta è rilevante o meno?
- Analogamente per eventuali ridondanze che si possono osservare.
 - Ad esempio: il coperto e il caffè hanno un prezzo che non varia da ricevuta a ricevuta?
- In generale è bene ricordare che ogni caso presenta una sua specificità e pertanto non deve essere trattato “automaticamente”.
- Normalizzare in 1NF è, a tutti gli effetti, un'attività di progettazione (logica), e in quanto tale può essere solo oggetto di “regole guida” che però non hanno validità assoluta.

Relazione matematica versus Relazione nel modello relazionale

Relazione matematica	Relazione Modello Relazionale
domini arbitrari	domini atomici
colonne senza nome	colonne con nomi
colonne distinte in base alla posizione	nomi univoci per le colonne
di solito costante nel tempo	di solito variabile nel tempo

Informazione incompleta

- Le informazioni che si vogliono rappresentare mediante relazioni non sempre corrispondono pienamente allo schema prescelto; in particolare, per alcune tuple e per alcuni attributi potrebbe non essere possibile specificare, per diversi motivi, un valore del dominio.

PERSONE

Codice	Cognome	Nome	DataMorte
A001	Rossi	Mario	20/02/1954
A002	Verdi	Paolo	
A003	Bianchi	Bruno	
A004	Grigi	Carlo	

- Paolo Verdi è ancora vivo (**valore non applicabile**);
- Bruno Bianchi è deceduto, ma non conosciamo la data di morte (**valore applicabile ma ignoto**);
- Carlo Grigi è scomparso misteriosamente, non sappiamo se è vivo o se è deceduto (**ignota l'applicabilità**).

Quale soluzione?

- In diversi casi, in mancanza di informazione, si tende a usare un “valore speciale” del dominio (0, “”, “-1”, “9999”, ecc.) che non si utilizza per altri scopi.
- Questa pratica è **fortemente sconsigliata**, in quanto, anche dove possibile:
 - ▣ valori inutilizzati potrebbero successivamente diventare significativi;
 - ▣ le applicazioni dovrebbero sapere “che cosa significa in realtà” il valore usato allo scopo.
- Esempio (reale!): nel 1998, analizzando i clienti di un’assicurazione, si scoprì una strana concentrazione di ultra-novantenni... tutte le date di nascita ignote erano state codificate con “01/01/00”!
- Nel modello relazionale si opera in maniera pragmatica: si adotta il concetto di **valore nullo (NULL)**, che denota assenza di un valore nel dominio (e **non è un valore del dominio**);
- ...pertanto $t[A] \in \text{dom}(A) \cup \{\text{NULL}\}$.

Valori nulli: considerazioni

PERSONE

Codice	Cognome	Nome	DataMorte
A001	Rossi	Mario	20/02/1954
A002	Verdi	Paolo	NULL
A003	Bianchi	Bruno	NULL
A004	Grigi	Carlo	NULL

- La presenza di un valore nullo non fornisce alcuna informazione sull'applicabilità o meno.
- È importante ricordare che NULL non è un valore del dominio; in particolare, se due tuple hanno entrambe valore NULL per un attributo, non si può inferire che esse abbiano lo stesso valore per quell'attributo, ovvero:

NULL ≠ NULL

- **N.B.** Tuttavia, ai fini della verifica di assenza di tuple duplicate è opportuno che i NULL siano considerati come gli altri valori e quindi uguali tra loro: (NULL = NULL).

Valori nulli: restrizioni

- La presenza di valori nulli non può essere sempre tollerata, ovvero è necessario imporre delle **restrizioni** al loro uso; si consideri ad esempio il caso della registrazione di esami:

ESAMI

Matricola	CodCorso	Voto	Lode
29323	483	28	no
NULL	729	30	sì
29323	913	NULL	no
35467	913	30	no

- Un valore nullo per Matricola non permetterebbe di sapere quale studente ha sostenuto l'esame.
- Un valore nullo per Voto non è proprio ammissibile nel contesto considerato.



Istanze di questo tipo non sono accettabili!

Vincoli di integrità

- La “**correttezza sintattica**” di uno stato di una relazione non è condizione sufficiente affinché i dati rappresentino un'informazione possibile nel contesto reale considerato.

STUDENTI	Matricola	Cognome	Nome	DataNascita
	35467	Bianchi	Giorgio	21/06/1978
	35467	Rossi	Anna	13/04/1978
	39654	Rossi	Anna	13/04/1978

- La prima e la seconda tupla hanno la stessa Matricola?
- La seconda e la terza tupla hanno gli stessi valori per Nome, Cognome e DataNascita, ma questo in linea di principio è possibile!
- Un **vincolo di integrità** è una proprietà che deve essere soddisfatta da ogni possibile stato osservabile di una relazione; ogni vincolo può quindi essere descritto da una funzione booleana che associa a ogni stato il valore VERO o FALSO.

Vincoli di dominio

- Un vincolo che si riferisce ai valori ammissibili per un singolo attributo è detto **vincolo di dominio** (o sui **valori**).

ESAMI

Matricola	CodCorso	Voto	Lode
29323	483	28	no
39654	729	30	sì
29323	913	31	no
35467	913	30	forse

- Il Voto deve essere compreso tra 18 e 30 :

$(Voto \geq 18) \text{ AND } (Voto \leq 30)$

- La Lode può solo assumere i valori 'sì' o 'no' :

$(Lode = \text{'sì'}) \text{ OR } (Lode = \text{'no'})$

Vincoli di tupla

- I vincoli di dominio sono un caso particolare dei **vincoli di tupla**, ovvero vincoli che esprimono **condizioni su ciascuna tupla**, indipendentemente dalle altre.

ESAMI

Matricola	CodCorso	Voto	Lode
29323	483	28	no
39654	729	30	sì
29323	913	26	sì
35467	913	30	no

- La **Lode** si può assegnare solo se il **Voto** è 30:
 $(\text{Voto} = 30) \text{ OR NOT}(\text{Lode} = \text{'sì'})$
- Nello schema **PAGAMENTI**(Data, ImportoLordo, Ritenute, Netto) si ha:

$$\text{ImportoLordo} = \text{Netto} + \text{Ritenute}$$

Vincoli di chiave: intuizione

- I vincoli di chiave, che giocano un ruolo molto importante, vietano la presenza di tuple distinte che hanno lo stesso valore su uno o più attributi.

STUDENTI

Matricola	CodiceFiscale	Cognome	Nome	DataNascita
210629323	BNCGRG78L21A944Z	Bianchi	Giorgio	21/07/1978
216635467	RSSNNA78A53A944N	Rossi	Anna	13/01/1978
160239654	VRDMRC79H20F839U	Verdi	Marco	20/06/1979
214842132	VRDMRC79H20G125T	Verdi	Marco	20/06/1979

- Un valore di Matricola **identifica univocamente** uno studente;
- analogamente per il CodiceFiscale
- e ogni insieme di attributi che includa Matricola o CodiceFiscale
 - {Matricola, Cognome}, {CodiceFiscale, Nome}, ...;
- ma possono esistere due tuple uguali su {Cognome, Nome, DataNascita}.

Chiavi e superchiavi

- Dato uno schema $R(X)$, un insieme di attributi $K \subseteq X$ è:
 - ▣ una **superchiave** se e solo se
in ogni stato ammissibile r di $R(X)$ non esistono due tuple distinte t_1 e t_2 tali che $t_1[K] = t_2[K]$;
 - ▣ una **chiave** se e solo se
è una superchiave minimale, ovvero non esiste $K' \subset K$ con K' superchiave.
- Una chiave è pertanto un **identificatore minimale** per ogni r su $R(X)$.
- Nella relazione STUDENTI:
 - $\{\text{Matricola}\}$ e $\{\text{CodiceFiscale}\}$ sono due **chiavi**;
 - $\{\text{Matricola}, \text{Cognome}\}$ e $\{\text{CodiceFiscale}, \text{Nome}\}$ sono solo **superchiavi**;
 - $\{\text{Cognome}, \text{Nome}, \text{DataNascita}\}$ **non è superchiave**.

Esistenza di chiavi e superchiavi

- Poiché ogni stato r su $R(X)$ è un insieme, ne consegue che
 - ▣ l'insieme X di tutti gli attributi dello schema è senz'altro una **superchiave** per $R(X)$.
- Poiché il numero di attributi, n , è finito, è sempre possibile arrivare a individuare (almeno) una **chiave** $K \subseteq X$

```
K = X;  
for i = 1 to n  
{  
    if K - {Ai} è superchiave  
        then K = K - {Ai};  
}
```



In casi (molto) particolari il numero di chiavi può essere esponenziale in n .

Vincoli espressi a livello di schema

- I vincoli di chiave si esprimono a livello di schema, sulla base di un'analisi della realtà che si vuole modellare mediante relazioni, e **limitano l'insieme di estensioni legali** (o “ammissibili”, “corrette”, “valide”, ecc.).
- Una specifica estensione può soddisfare altri vincoli (di chiave), ma ciò non autorizza a generalizzare.

ESAMI

Matricola	CodCorso	Voto	Lode
29323	483	28	no
39654	729	30	sì
29323	913	26	no
35467	913	30	sì

- La (unica) **chiave** è {Matricola, CodiceCorso}.
- Questa particolare estensione soddisfa anche altri vincoli, ad esempio {Matricola, Voto} è un identificatore, **ma ciò è puramente casuale**.

Importanza delle chiavi

- L'esistenza delle chiavi garantisce l'accessibilità a ciascun dato del DB, in quanto ogni singolo valore è univocamente individuato da:



- Le chiavi sono lo strumento principale attraverso il quale vengono correlati i dati in relazioni diverse (“il modello relazionale è basato su valori”).

Legami basati sui valori

LINEE_AEREE

<u>CodLinea</u>	NomLinea
L001	TWA
L002	ALITALIA
....

DESTINAZIONI

<u>CodDest</u>	NomeDest	Nazione
FCO	ROMA	ITALIA
JFK	NEW YORK	USA
.....

VOLI

<u>NumVolo</u>	<u>CodLinea</u>	<u>CodDest</u>	Giorno	Ora	Durata	Attivo
TW056	L001	JFK	LUN	9:00	2	SI
AZ854	L002	FCO	MER	22:30	8	SI
....			

Chiavi e valori nulli

- In presenza di valori nulli entrambe le funzioni svolte dalle chiavi (identificazione e correlazione) possono venire a mancare.

STUDENTI

Matricola	CodiceFiscale	Cognome	Nome	DataNascita
NULL	NULL	Bianchi	Giorgio	21/07/1978
216635467	RSSNNA78A53A944N	Rossi	Anna	13/01/1978
NULL	VRDMRC79H20F839U	Verdi	Marco	20/06/1979
214842132	NULL	Verdi	Marco	20/06/1979

- La prima tupla non è identificabile in alcun modo, pertanto:
 - è necessario specificare il valore di almeno una chiave!
- La terza e quarta tupla potrebbero riferirsi allo stesso studente, pertanto:
 - non è sufficiente specificare il valore di una chiave!

Chiave primaria

- Per evitare i problemi visti è necessario scegliere una chiave, detta **chiave primaria (primary key)**, su cui non si ammettono valori nulli.
- Convenzionalmente si sottolineano gli attributi che costituiscono la chiave primaria.

STUDENTI

<u>Matricola</u>	CodiceFiscale	Cognome	Nome	DataNascita
210629323	NULL	Bianchi	Giorgio	21/07/1978
216635467	RSSNNA78A53A944N	Rossi	Anna	13/01/1978
160239654	VRDMRC79H20F839U	Verdi	Marco	20/06/1979
214842132	NULL	Verdi	Marco	20/06/1979



Nei casi in cui per nessuna chiave si possa garantire la disponibilità di valori, è necessario introdurre un nuovo attributo (un “**codice**”) che svolga le funzioni di chiave primaria; si pensi ad esempio al caso in cui non si riesca a identificare un paziente all’arrivo a un pronto soccorso ospedaliero.

Vincoli di integrità referenziale

- I vincoli sinora visti sono tutti di tipo **intra-relazionale**, in quanto interessano una relazione alla volta.
- Viceversa, i **vincoli di integrità referenziale** sono importanti tipi di vincoli **inter-relazionali** che enfatizzano come le correlazioni tra le tuple siano spesso ottenute usando i valori delle chiavi.
- Si considerino due schemi $R_1(X_1)$ e $R_2(X_2)$ di un DB R , e sia Y un insieme di attributi in X_2 .
- Un vincolo di integrità referenziale su Y impone che in ogni stato $r = \{r_1, r_2, \dots\}$ del DB l'insieme dei valori di Y in r_2 sia un sottoinsieme dell'insieme dei valori della chiave primaria di $R_1(X_1)$ presenti nello stato r_1 .
- L'insieme Y viene detto una **foreign key** (o “**chiave importata**”).

Esempi di foreign key

STUDENTI

<u>Matricola</u>	Cognome	Nome	DataNascita
29323	Bianchi	Giorgio	21/06/1978
35467	Rossi	Anna	13/04/1978
39654	Verdi	Marco	20/09/1979
42132	Neri	Lucia	15/02/1978

CORSI

<u>CodCorso</u>	Titolo	CodDocente	Anno
483	Analisi	0201	1
729	Analisi	0021	1
913	Sistemi Informativi	0123	2

ESAMI

foreign key

<u>Matricola</u>	<u>CodCorso</u>	Voto	Lode
29323	483	28	no
39654	729	30	sì
29323	913	26	no
35467	913	30	sì

In **CORSI**, {CodDocente} è una foreign key.

In **ESAMI**, {Matricola} è una foreign key, così come {CodCorso}.

Foreign key: precisazioni

- In generale la foreign key Y e la primary key K di $R1(X1)$ possono includere attributi con nomi diversi:

CORSI

<u>Codice</u>		Titolo	CodDocente	Anno
483		Analisi I	0201	1
729		Analisi I	0021	1

ESAMI

<u>NumMatricola</u>	<u>CodCorso</u>	Voto	Lode
29323	483	28	no

- Foreign key e primary key possono far parte della stessa relazione, ovviamente con $Y \neq K$.


PERSONALE

<u>Codice</u>	Cognome	...	CodResponsabile
123	Rossi	...	325
134	Verdi	...	325
325	Neri

Foreign key: valori nulli

- In presenza di **valori nulli**, i vincoli di integrità referenziale si possono parzialmente rilassare:

PERSONALE



<u>Codice</u>	Cognome	...	CodResponsabile
123	Rossi	...	325
134	Verdi	...	325
325	Neri	...	NULL

- Nei DBMS un vincolo di integrità referenziale può anche esprimersi con riferimento a una generica chiave (quindi anche non primaria):

STUDENTI

<u>Matricola</u>	CodiceFiscale	Cognome	Nome	DataNascita
29323	BNCGRG78L21A944Z	Bianchi	Giorgio	21/07/1978
35467	RSSFLV78M53G125O	Rossi	Flavia	13/08/1978

REDDITI

<u>CF</u>	Imponibile
BNCGRG78L21A944Z	45300

Foreign key: notazioni negli schemi

- Si usano, nei vari testi sui DB relazionali, diverse notazioni per indicare nella definizione di uno schema le foreign key. Ad esempio:

AGENZIE(Agenzia, Luogo)

IMPIEGATI(CodImpiegato, Cognome, Nome, CodAgenzia)

FK: CodAgenzia REFERENCES AGENZIE(Agenzia)

o più semplicemente

FK: CodAgenzia REFERENCES AGENZIE

oppure

AGENZIE(Agenzia, Luogo)

IMPIEGATI(CodImpiegato, Cognome, Nome, CodAgenzia: AGENZIE)

Queste notazioni sono da intendersi come semplificazioni rispetto alla sintassi del linguaggio SQL.

Valori nulli e altri vincoli: notazioni negli schemi

- Per quanto riguarda gli attributi e le foreign key che ammettono valori nulli, nella definizione di schemi relazionali a volte si usa il simbolo * proprio per denotare che è ammesso NULL come valore.
- Per denotare che un attributo A ammette solo valori unici, ovvero non ripetuti, a volte si usa scrivere: Unique(A).

AGENZIE(CodAgenzia, Nome, Sede, Direttore: IMPIEGATI)

Unique(Direttore)

IMPIEGATI(Codice, Nome, Cognome, Agenzia*: AGENZIE, DataInizio*)

- **N.B.** Non sempre negli esempi, per semplicità, questa notazione è rispettata lasciando al lettore il compito di interpretare di volta in volta.

Domande?

