



# BASI DI DATI

## IL MODELLO RELAZIONALE

Polese G. Caruccio L. Breve B.

a.a. 2023/2024

# Modelli logici dei dati

- Tre modelli logici tradizionali
  - gerarchico
  - reticolare
  - relazionale
- Più recente (e poco diffuso)
  - a oggetti

# Modelli logici, caratteristiche

- **Gerarchico e reticolare**
  - Utilizzano riferimenti espliciti (puntatori) fra record
- **Relazionale** "è basato su valori"
  - Anche i riferimenti fra dati in strutture (relazioni) diverse sono rappresentati per mezzo di valori

# Il modello relazionale

- Proposto da E. F. Codd nel 1970 per favorire l'indipendenza dei dati
- Disponibile in DBMS reali nel 1981 (non era facile implementare l'indipendenza con efficienza ed affidabilità!)
- Si basa sul concetto matematico di relazione (con una variante)
- Le relazioni hanno naturale rappresentazione per mezzo di tabelle

# Relazione: 2 accezioni

- **Relazione matematica:** sottoinsieme del prodotto cartesiano di insiemi
- **Relazione** secondo il modello relazionale dei dati

# Relazione matematica

- $D_1, \dots, D_n$  ( $n$  insiemi anche non distinti)
- **Prodotto cartesiano**  $D_1 \times \dots \times D_n$ :
  - L'insieme di tutte le  $n$ -uple  $(d_1, \dots, d_n)$  tali che  $d_1 \in D_1, \dots, d_n \in D_n$
- **Relazione matematica** su  $D_1, \dots, D_n$ :
  - Un sottoinsieme di  $D_1 \times \dots \times D_n$ .
- $D_1, \dots, D_n$  sono i **domini** della relazione

# Relazione matematica, esempio

- $D_1 = \{a, b\}$
- $D_2 = \{x, y, z\}$
- Prodotto cartesiano  $D_1 \times D_2$

a	x
a	y
a	z
b	x
b	y
b	z

- Una relazione  $r \subseteq D_1 \times D_2$

a	x
a	z
b	y

# Relazione matematica, proprietà

- Essendo la relazione matematica un insieme, esso gode delle seguenti proprietà:
  - Non c'è ordinamento fra le  $n$ -uple  $(d_1, \dots, d_n)$
  - Le  $n$ -uple sono distinte
  - Ciascuna  $n$ -upla è ordinata: l'  $i$ -esimo valore proviene dall'  $i$ -esimo dominio



# Relazione matematica, esempio

***Partite*  $\subseteq$  *string*  $\times$  *string*  $\times$  *int*  $\times$  *int***

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

- Ciascuno dei domini (string ed int) ha due **ruoli** diversi, distinguibili attraverso la posizione:
  - La struttura è **posizionale**

# Struttura non posizionale

- A ciascun dominio si associa un nome (**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

# Tabelle e Relazioni

- Una tabella rappresenta una relazione se
  - I valori di ogni colonna sono fra loro omogenei (dello stesso tipo)
  - Le righe sono diverse fra loro (altrimenti non è un insieme)
  - Le intestazioni delle colonne sono diverse tra loro
- In una tabella che rappresenta una relazione
  - L'ordinamento tra le righe è irrilevante
  - L'ordinamento tra le colonne è irrilevante

# Il modello relazionale è basato su valori

- I riferimenti fra dati in relazioni diverse (es. i dati di uno studente e quelli dei suoi esami) sono rappresentati per mezzo di valori dei domini che compaiono nelle ennuple

# Riferimenti tra ennuple, esempio

## Studenti

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

## Esami

Studente	Voto	Corso
3456	30	04
3456	24	02
9283	28	01
6554	26	01

## Corsi

Codice	Titolo	Docente
01	Analisi	Mario
02	Chimica	Bruni
04	Chimica	verdi

## Studenti

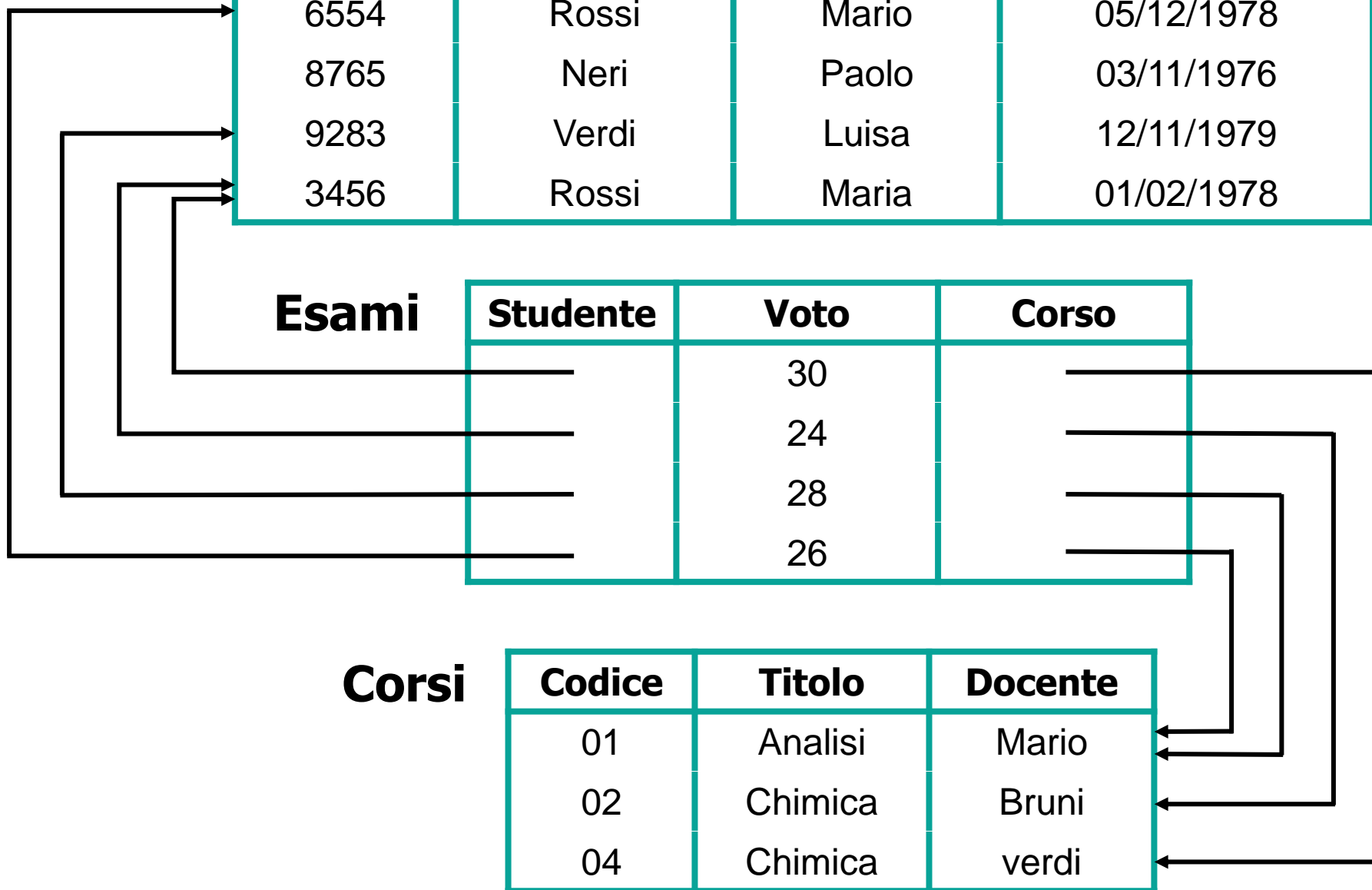
Matricola	Cognome	Nome	Data di nascita
6554	Rossi	Mario	05/12/1978
8765	Neri	Paolo	03/11/1976
9283	Verdi	Luisa	12/11/1979
3456	Rossi	Maria	01/02/1978

## Esami

Studente	Voto	Corso
	30	
	24	
	28	
	26	

## Corsi

Codice	Titolo	Docente
01	Analisi	Mario
02	Chimica	Bruni
04	Chimica	verdi



# 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 più facilmente da un sistema ad un altro
- I puntatori sono direzionali, i valori no

# Schema di Relazione e di BD

- Nel **modello relazionale** i concetti di **Schema** (intensione) ed **Istanza** (estensione) **di Base di Dati** si traducono in:

- **Schema di relazione**

Un nome **R** con un insieme di attributi  $A_1, \dots, A_n$ :

$$R(A_1, \dots, A_n)$$

- **Schema di base di dati**

Insieme di schemi di relazione:

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

dove  $X_1, \dots, X_k$  sono **insiemi di attributi**



# Ennuple (Tuple) di Relazioni

- Una **ennupla** (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 il valore della ennupla **t** sull'attributo **A**
- Le **ennuple** sono anche dette **tuple**

# Istanze di Relazione e di BD

- (Istanza di) **relazione** su uno schema  $R(X)$ :
  - insieme  $r$  di ennuple su  $X$
- (Istanza di) **base di dati** su uno schema  $R = \{R_1(X_1), \dots, R_n(X_n)\}$ :
  - insieme di relazioni  
 $r = \{r_1, \dots, r_n\}$  (con  $r_i$  relazione su  $R_i(X_i)$ )

# 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	Luisa	12/11/1979
3456	Rossi	Maria	01/02/1978

## Studenti lavoratori

Matricola
6554
3456

# Informazione incompleta

- Il modello relazionale impone ai dati una struttura rigida.
- Solo alcuni formati di ennuple sono ammessi: quelli che corrispondono agli schemi di relazione
- I dati disponibili possono non corrispondere al formato previsto

# Informazione incompleta: esempio

<b>Nome</b>	<b>SecondoNome</b>	<b>Cognome</b>
Franklin	Delano	Roosevelt
Winston		Churchill
Charles		De Gaulle
Josip		Stalin

# 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

# 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  $A$ , diventa un valore del dominio  $\text{dom}(A)$  oppure il valore nullo NULL
- Si possono (e si debbono) imporre restrizioni sulla presenza di valori nulli

# Eccesso di valori nulli

## 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



# Tipi di valore nullo

- Esistono varie interpretazioni di un valore nullo, ad esempio
  - Valore **sconosciuto**
  - Valore **inesistente**
- I DBMS non distinguono i diversi tipi di valore nullo

# Vincoli di integrità

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

# Una base di dati "scorretta"

## Esami

Studente	Voto	Lode	Corso
276545	32		01
276545	30	e lode	02
787643	27	e lode	03
739430	24		04

## Studenti

Matricola	Cognome	Nome
276545	Rossi	Mario
787643	Neri	Piero
787643	Bianchi	Luca

# 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**

# Vincoli di integrità, perché?

- Descrizione più accurata della realtà
- Contributo alla “qualità dei dati”
- Utili nella progettazione (vedremo)
- Usati dai DBMS nell'esecuzione delle interrogazioni

# Vincoli di integrità, nota

- Non tutte le proprietà di interesse sono rappresentabili per mezzo di vincoli formulabili in modo esplicito

# Tipi di vincoli

- Vincoli **intrarelazionali**
  - Vincoli su valori (o di **dominio**)
  - Vincoli di **ennupla**
- Vincoli **interrelazionali**

# Violazione di vincoli, esempi

## Esami

Studente	Voto	Lode	Corso
276545	32		01
276545	30	e lode	02
787643	27	e lode	03
739430	24		04

## Studenti

Matricola	Cognome	Nome
276545	Rossi	Mario
787643	Neri	Piero
787643	Bianchi	Luca



# Vincoli di ennupla

- Esprimono condizioni sui valori di ciascuna ennupla, in modo indipendente dalle altre ennuple
- Caso particolare:
  - Vincoli di dominio che coinvolgono un solo attributo

# Sintassi ed esempi

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

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

$(\text{Voto} = 30) \text{ OR NOT } (\text{Lode} = \text{"e lode"})$

# Vincoli di ennupla, esempio

## Stipendi

Impiegato	Lordo	Ritenute	Netto
Rossi	55.000	12.500	42.500
Neri	45.000	10.000	35.000
Bruni	47.000	11.000	36.000

$$\text{Lordo} = (\text{Ritenute} + \text{Netto})$$

# Identificazione di ennuple

Matricola	Cognome	Nome	Corso	Data di nascita
27655	Rossi	Mario	Ing Inf	05/12/1978
78763	Rossi	Mario	Ing Inf	03/11/1976
65432	Neri	Piero	Ing Mecc	10/07/1979
87654	Neri	Mario	Ing Inf	03/11/1976
67653	Rossi	Piero	Ing Mecc	05/12/1978

- Non ci sono due ennuple con lo stesso valore sull'attributo Matricola
- Non ci sono due ennuple uguali su tutti e tre gli attributi Cognome, Nome e Data di Nascita

# Chiave

- Insieme di attributi che identificano le ennuple di una relazione

Formalmente:

- Un insieme  $K$  di attributi è **superchiave** per  $r$  se  $r$  non contiene due ennuple distinte  $t_1$  e  $t_2$  con  $t_1[K] = t_2[K]$
- $K$  è **chiave candidata** per  $r$  se è una superchiave minimale per  $r$  (cioè non contiene un'altra superchiave)

# Una chiave

Matricola	Cognome	Nome	Corso	Data di nascita
27655	Rossi	Mario	Ing Inf	05/12/1978
78763	Rossi	Mario	Ing Inf	03/11/1976
65432	Neri	Piero	Ing Mecc	10/07/1979
87654	Neri	Mario	Ing Inf	03/11/1976
67653	Rossi	Piero	Ing Mecc	05/12/1978

- Matricola è una chiave candidata:
  - È superchiave
  - Contiene un solo attributo e quindi è minimale

# Un'altra chiave

Matricola	Cognome	Nome	Corso	Data di nascita
27655	Rossi	Mario	Ing Inf	05/12/1978
78763	Rossi	Mario	Ing Inf	03/11/1976
65432	Neri	Piero	Ing Mecc	10/07/1979
87654	Neri	Mario	Ing Inf	03/11/1976
67653	Rossi	Piero	Ing Mecc	05/12/1978

- Cognome, Nome, Data di nascita è un'altra chiave candidata:
  - È superchiave
  - È minimale

# Un'altra chiave?

Matricola	Cognome	Nome	Corso	Data di nascita
27655	Rossi	Mario	Ing Inf	05/12/1978
78763	Rossi	Mario	Ing Civile	03/11/1976
65432	Neri	Piero	Ing Mecc	10/07/1979
87654	Neri	Mario	Ing Inf	03/11/1976
67653	Rossi	Piero	Ing Mecc	05/12/1978

- Non ci sono ennuple uguali su Cognome e Corso:
  - Cognome e Corso formano una chiave?
  - Ma è sempre vero?



# Vincoli, schemi e istanze

- I vincoli corrispondono a proprietà del mondo reale modellato dalla base di dati
- Interessano a livello di schema (con riferimento cioè a tutte le istanze)
- Ad uno schema associamo un insieme di vincoli e consideriamo **corrette** (valide, ammissibili) solo le istanze che soddisfano tutti i vincoli

## Studenti

Matricola	Cognome	Nome	Corso	Data di nascita
-----------	---------	------	-------	-----------------

- chiavi candidate:

**Matricola**

**Cognome, Nome, Nascita**

<b>Matricola</b>	<b>Cognome</b>	<b>Nome</b>	<b>Corso</b>	<b>Data di nascita</b>
27655	Rossi	Mario	Ing Inf	05/12/1978
78763	Rossi	Mario	Ing Civile	03/11/1976
65432	Neri	Piero	Ing Mecc	10/07/1979
87654	Neri	Mario	Ing Inf	03/11/1976
67653	Rossi	Piero	Ing Mecc	05/12/1978

- È corretta: soddisfa i vincoli

# Esistenza delle chiavi

- Una relazione non può contenere ennuple distinte ma uguali
- Ogni relazione ha come superchiave l'insieme degli attributi su cui è definita....
- ....e quindi ha (almeno) una chiave

# Importanza delle chiavi

- L'esistenza delle chiavi garantisce l'accessibilità a ciascun dato della base di dati
- Le chiavi permettono di correlare i dati in relazioni diverse (modello basato su valori)

# Chiavi e valori nulli

- In presenza di valori nulli i valori della chiave non permettono
  - Di identificare le ennuple
  - Di realizzare facilmente i riferimenti da altre relazioni

<b>Matricola</b>	<b>Cognome</b>	<b>Nome</b>	<b>Corso</b>	<b>Data di nascita</b>
NULL	NULL	Mario	Ing Inf	05/12/1978
78763	Rossi	Mario	Ing Civile	03/11/1976
65432	Neri	Piero	Ing Mecc	10/07/1979
87654	Neri	Mario	Ing Inf	NULL
NULL	Rossi	Piero	NULL	05/12/1978

- La presenza di valori nulli nelle chiavi non deve essere permessa

# Chiave primaria

- Chiave candidata su cui non sono ammessi valori nulli
- Notazione: sottolineatura

<u>Matricola</u>	Cognome	Nome	Corso	Data di nascita
86765	NULL	Mario	Ing Inf	05/12/1978
78763	Rossi	Mario	Ing Civile	03/11/1976
65432	Neri	Piero	Ing Mecc	10/07/1979
87654	Neri	Mario	Ing Inf	NULL
43289	Neri	Mario	NULL	05/12/1978



# Integrità referenziale

- Informazioni in relazioni diverse sono correlate attraverso valori comuni
- In particolare, valori delle chiavi (primarie)
- Le correlazioni debbono essere "coerenti"

# Infrazioni

<u>Codice</u>	Data	Vigile	Prov	Numero
34321	01/02/1995	3987	MI	39548K
53524	04/03/1995	3295	TO	E39548
64521	05/04/1996	3295	PR	839548
73321	05/02/1998	9345	PR	839548

## Vigili

<u>Matricola</u>	Cognome	Nome
3987	Rossi	Luca
3295	Neri	Piero
9345	Neri	Mario
7543	Mori	Gino

# Infrazioni

<u>Codice</u>	Data	Vigile	Prov	Numero
34321	01/02/1995	3987	MI	39548K
53524	04/03/1995	3295	TO	E39548
64521	05/04/1996	3295	PR	839548
73321	05/02/1998	9345	PR	839548

# Auto

<u>Prov</u>	<u>Numero</u>	Cognome	Nome
MI	39548K	Rossi	Mario
TO	E39548	Rossi	Mario
PR	839548	Neri	Luca

# Vincolo di integrità referenziale

- Un vincolo di **integrità referenziale** (“**Chiave esterna**”) fra gli attributi  $X$  di una relazione  $R_1$  e un'altra relazione  $R_2$  impone ai valori su  $X$  in  $R_1$  di comparire come valori della chiave primaria di  $R_2$

# Chiave Esterna (foreign key)

- In altre parole una chiave esterna è un gruppo di attributi di una relazione che coincide con la chiave primaria di un'altra relazione dello schema di database.

# Vincolo di integrità referenziale, esempio

- Vincoli di integrità referenziale fra:
  - L'attributo Vigile della relazione INFRAZIONI e la relazione VIGILI
  - Gli attributi Prov e Numero di INFRAZIONI e la relazione AUTO

# Violazione di vincolo di integrità referenziale




## Infrazioni

<u>Codice</u>	Data	Vigile	Prov	Numero
34321	01/02/1995	3987	MI	39548K
53524	04/03/1995	3295	TO	E39548
64521	05/04/1996	3295	PR	839548
73321	05/02/1998	9345	PR	839548

## Auto

<u>Prov</u>	<u>Numero</u>	Cognome	Nome
MI	E39548	Rossi	Mario
TO	F34268	Rossi	Mario
PR	839548	Neri	Luca

# Vincoli di integrità referenziale: commenti

- Giocano un ruolo fondamentale nel concetto di "modello basato su valori"
- In presenza di valori nulli i vincoli possono essere resi meno restrittivi 
- Sono possibili meccanismi per il supporto alla loro gestione ("azioni" compensative a seguito di violazioni) 
- Attenzione ai vincoli su più attributi 



# Integrità referenziale e valori nulli

## Impiegati

<u>Matricola</u>	Cognome	Progetto
34321	Rossi	IDEA
53524	Neri	XYZ
64521	Verdi	NULL
73321	Bianchi	IDEA

## Progetti

<u>Codice</u>	Inizio	Durata	Costo
IDEA	01/2000	36	200
XYZ	07/2001	24	120
BOH	09/2001	24	150



# Azioni compensative del DBMS

- Esempio:
  - Viene eliminata una ennupla causando una violazione
- Comportamento "standard" del DBMS:
  - Rifiuto dell'operazione
- Azioni compensative del DBMS:
  - Eliminazione in cascata
  - Introduzione di valori nulli

# Eliminazione in cascata

## Impiegati

<u>Matricola</u>	Cognome	Progetto
34321	Rossi	IDEA
53524	Neri	XYZ
64521	Verdi	NULL
73321	Bianchi	IDEA

## Progetti

<u>Codice</u>	Inizio	Durata	Costo
IDEA	01/2000	36	200
XYZ	07/2001	24	120
BOH	09/2001	24	150

# Introduzione di valori nulli

## Impiegati

<u>Matricola</u>	Cognome	Progetto
34321	Rossi	IDEA
53524	Neri	NULL
64521	Verdi	NULL
73321	Bianchi	IDEA

## Progetti

<u>Codice</u>	Inizio	Durata	Costo
IDEA	01/2000	36	200
XYZ	07/2001	24	120
BOH	09/2001	24	150



# Vincoli multipli su più attributi

## Incidenti

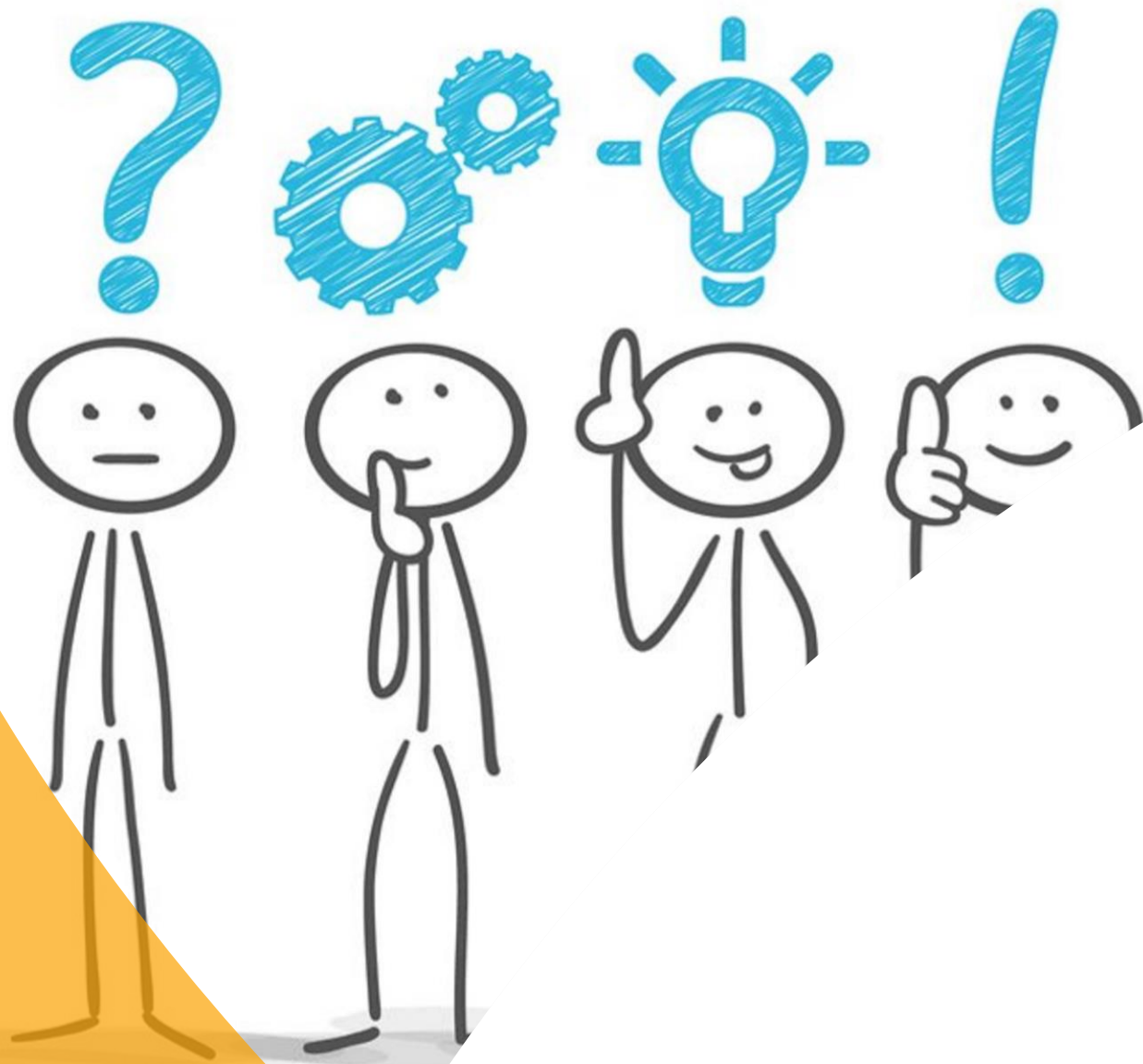
<u>Codice</u>	Data	ProvA	NumeroA	ProvB	NumeroB
34321	01/02/1995	TO	E39548	MI	39548K
64521	05/04/1996	PR	839548	TO	E39548

## Auto

<u>Prov</u>	<u>Numero</u>	Cognome	Nome
MI	39548K	Rossi	Mario
TO	E39548	Rossi	Mario
PR	839548	Neri	Luca

# Ordine Attributi in Vincoli Multipli

- Vincoli di integrità referenziale fra:
  - Gli attributi ProvA e NumeroA di INCIDENTI e la relazione AUTO
  - Gli attributi ProvB e NumeroB di INCIDENTI e la relazione AUTO
- L'ordine degli attributi è significativo



Grazie per  
l'attenzione