

Il Modello Relazionale

Il Modello Relazionale

- Modello proposto nel 1970 da E.F. Codd in:
 - “A relational model for large shared data banks”
Communications of the ACM Vol. 13, n. 6, pagg. 377-387
- Prime apparizioni nel mercato solo nel 1981
- Caratterizzato da un alto livello di astrazione
 - proposto per superare le limitazioni precedenti
 - caratterizzato da una elevata indipendenza dei dati
 - ha richiesto l'individuazione di realizzazioni efficienti e di hardware adeguato.

I fattori del successo

Si basa su due concetti:

- Tabella (concetto primitivo)
 - Semplice ed intuitiva
 - Rappresentazione grafica
 - Utile nella comunicazione con gli utenti
- Relazione
 - Matematica
 - Secondo la definizione del modello relazionale
 - Con riferimento al modello concettuale entity-relationship

La forza del modello è dovuta al suo essere basato su un concetto estremamente intuitivo (tabella) e su un concetto estremamente ben formalizzato dal punto di vista matematico (relazione).

Prodotto Cartesiano

- Dati due insiemi $D1$ e $D2$ si dice prodotto cartesiano di $D1$ per $D2$ ($D1 \times D2$) l'insieme formato dalle coppie ordinate $(v1, v2)$ con $v1 \in D1$ e $v2 \in D2$
- Esempio:
 - $A = \{1,2,3\}$ $B = \{a,b,c\}$
 - $A \times B = \{(1,a),(1,b),(1,c), (2,a),(2,b),(2,c) (3,a),(3,b),(3,c)\}$

Relazione Matematica

- Una relazione matematica R sugli insiemi $D1$ e $D2$ (domini della relazione) è definita come un sottoinsieme di $(D1 \times D2)$.
 - $\{(1,a),(2,a),(3,b)\}$ relazione su A e B .
- Gli insiemi e quindi le relazioni possono essere rappresentati in forma tabellare:

1
2
3

D1

a
b
c

D2

1	a
2	a
3	b

R

- Dati n insiemi non necessariamente distinti

$$D_1, D_2, D_3, \dots, D_n \quad \text{con } n > 0$$

Il Prodotto Cartesiano

$$D_1 \times D_2 \times D_3 \times \dots \times D_n$$

è costituito dall'insieme di n -uple

$$(v_1, v_2, v_3, \dots, v_n)$$

con v_i appartenente a D_i per $1 \leq i \leq n$

- Una relazione matematica **R** sugli insiemi D_i è un sottoinsieme del prodotto cartesiano

-

$$D_1 \times D_2 \times D_3 \times \dots \times D_n$$

Gli insiemi D_i sono detti i domini della relazione

Siano dati i seguenti domini:

CODICE = {001, 004, 005}
NOME = {Mel, Pedro, Federico}
COGNOME = {Almodovar, Gibson, Fellinig},
NAZIONALITA = {Italia, Spagna, Australia}.

Una relazione su questi domini è una generica :

$$r \subseteq \text{CODICE} \times \text{NOME} \times \text{COGNOME} \times \text{NAZIONALITA}$$

Possibili relazioni sono dunque:

$$r1 = \{(001, Pedro, Almodovar, Spagna)\}$$

$$r2 = \{(001, Pedro, Amodovar, Spagna); \\ (004, Mel, Gibson, Australia)\}$$

$$r3 = \{\phi\}$$

Una relazione è quindi definita da:

- un numero di domini elementari n
 - grado della relazione
- dal numero di ennuple che la compongono
 - cardinalità della relazione

Dalle coppie quando $n=2$ si passa alle ennuple dette anche tuple

- Il valore di n è finito
 - Rappresentazione finita delle informazioni
- La cardinalità dei domini può essere considerata infinita
 - Si pensi all'insieme dei cognomi delle persone
- I domini possono essere
 - Tutti dello stesso tipo
 - Di tipo diverso
 - Non tutti dello stesso tipo

Relazioni con Attributi

- In una relazione non è definito alcun ordinamento tra le tuple che la compongono.
 - due tabelle che differiscono solo nell'ordine delle righe rappresentano la stessa relazione.
- Le tuple di una relazione sono distinte l'una dall'altra
 - Perché si possa parlare di relazione non devono esistere due tuple con valori uguali.
- Esiste un legame di tipo posizionale tra i valori interni ad una ennupla ed i corrispondenti domini

Ordine posizionale

- In ogni tupla esiste un ordine posizionale

juventus	lazio	0	1
milan	inter	2	0
roma	napoli	2	0

- Se scambiamo la colonna 3 con la 4 si altera l'informazione
- Tale caratteristica è insoddisfacente in quanto si tende, in informatica, a privilegiare notazioni non posizionali.

Per risolvere il problema:

- ad ogni occorrenza di dominio di una relazione si associa un **attributo** che permette di identificare e qualificare il ruolo del dominio stesso

Da quanto detto segue che le **tuple** diventano

$$t = \{ \langle v_1, A_1 \rangle, \langle v_2, A_2 \rangle, \dots, \langle v_n, A_n \rangle \}$$

Non esiste alcun ordinamento tra le tuple .

Relazioni con Attributi

SCasa	SOspite	RSC	RSO
juventus	lazio	0	1
milan	inter	2	0
roma	napoli	0	0

- In maniera formale:
 - Una relazione è assimilabile ad un insieme di record omogenei.
 - Nel caso dei record ad ogni campo è associato un nome.
 - Analogamente, si associa ad ogni occorrenza di dominio della relazione un attributo che ne descrive il ruolo
 - Con l'introduzione degli attributi, l'ordine delle colonne non è più significativo.

La funzione t

SCasa	SOspite	RSC	RSO
juventus	lazio	0	1
milan	inter	2	0
roma	napoli	0	0

- Sull'insieme di attributi X si definisce la funzione t che associa all'attributo A_i il valore del dominio $DOM_i(A_i)$ sulla tupla indicata
- Con riferimento alla seconda tupla:

$$t_2[SCasa] = t.SCasa = milan$$

Rappresentazione di una relazione

- Una relazione può essere rappresentata naturalmente attraverso le **tabelle** dove
 - ogni riga è una tupla
 - ogni colonna riporta i valori relativo ad uno specifico attributo
 - la prima riga elenca gli attributi

**Non tutte le tabelle possono essere delle relazioni:
solo quelle con righe diverse tra loro**

Tabelle e Relazioni

- Dalla relazione dell'esempio precedente otteniamo la seguente tabella:

<i>Codice</i>	<i>Nome</i>	<i>Cognome</i>	<i>Nazionalità</i>
<i>001</i>	<i>Pedro</i>	<i>Almodovar</i>	<i>Spagna</i>
<i>004</i>	<i>Federico</i>	<i>Fellini</i>	<i>Italia</i>
<i>005</i>	<i>Mel</i>	<i>Gibson</i>	<i>Australia</i>

Schema di Relazione

Dati un insieme di nomi di attributi

$$X = \{A1, A2, \dots, An\},$$

si definisce $R(X)$ schema di relazione:

R seguito dall'insieme dei nomi di attributi di X

$$R(X) = R(A1, A2, \dots, An)$$

- *Sono esempi di schemi di relazione:*

R1 = AUTORI(CODICE, NOME, COGNOME, NAZIONALITA)

R2 = FILM(AUTORE, TITOLO, ATTORE)

Istanza di una relazione

- *Si definisce relazione r sullo schema $R(X)$ una istanza di $R(X)$.*

Nome	Cognome	Telefono
Angelo	Chianese	0817683827
Antonio	Picariello	0817683826



In altre parole l'insieme delle tuple

Informazioni incomplete

- Il modello relazionale impone strutture rigide alle informazioni:
 - Una relazione è un insieme di tuple omogenee (stessa struttura)
 - Per alcune tuple può accadere che non sia definito il valore di alcuni campi
- Come gestire l'assenza di informazione?
 - Riempendo i campi con valori opportuni?
 - E come sceglierli?

Il valore NULL

- In questi casi, si è soliti estendere i domini delle relazione con un valore speciale, detto NULL

$$D_i = D_i \cup \text{NULL}$$

Con il valore NULL si intende prendere in considerazione una assenza di informazione che può essere dovuta a diversi fattori:

1. il dato c'è ma non lo conosco (valore sconosciuto)
2. il dato non può esserci, in quanto non è applicabile ad una tupla (valore inesistente)
3. non so dire nulla, il dato è sconosciuto o inesistente.

- **Ipotesi:** ogni studente è dotato di un cellulare, mentre il professore è dotato di un cellulare e di un telefono di ufficio.

Elenco telefonico

Nome	Ruolo	Cellulare	Ufficio
Giacomo	STUD	335/123123	NULL
Antonio	PROF	335/123131	081/7683826
Marta	STUD	NULL	NULL
Annarita	NULL	334/123123	NULL

- valore sconosciuto: la terza tupla, presenta un valore NULL nel campo Cellulare pur essendo Marta una *STUD*
- Non applicabilità dell'informazione: la prima tupla riporta il valore NULL al telefono di ufficio di uno studente
- Per l'ultima tupla non si può dire nulla non potendo stabilire se Annarita è *STUD* o *PROF*

Vincoli di integrità

Un vincolo di integrità o *integrity constraint*, è una regola che ogni istanza della base di dati deve rispettare affinché i suoi dati siano corrispondenti al modello della realtà che un DB rappresenta.

L'insieme di queste regole è chiamato **IC**.

- **Schema Basi di Dati:**

- *Uno schema di base di dati BD è dato dal nome della base di dati BD seguito dall'insieme di schemi di relazione $R1(X1)$, $R2(X2)$, ..., $Rn(Xn)$, più un insieme di regole IC*

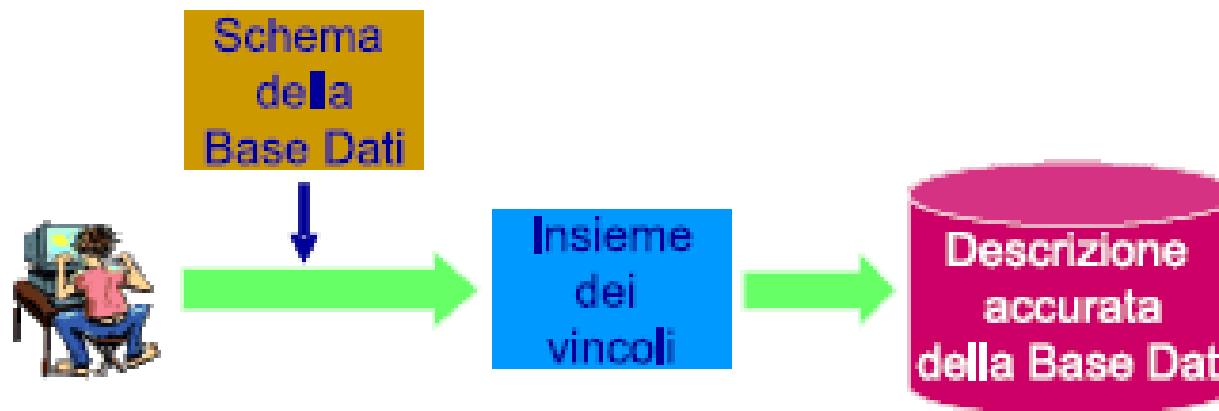
- **Basi di Dati Relazionale:**

- *Si definisce base di dati relazionale una istanza di uno schema di basi di dati che soddisfa le regole contenute in IC*

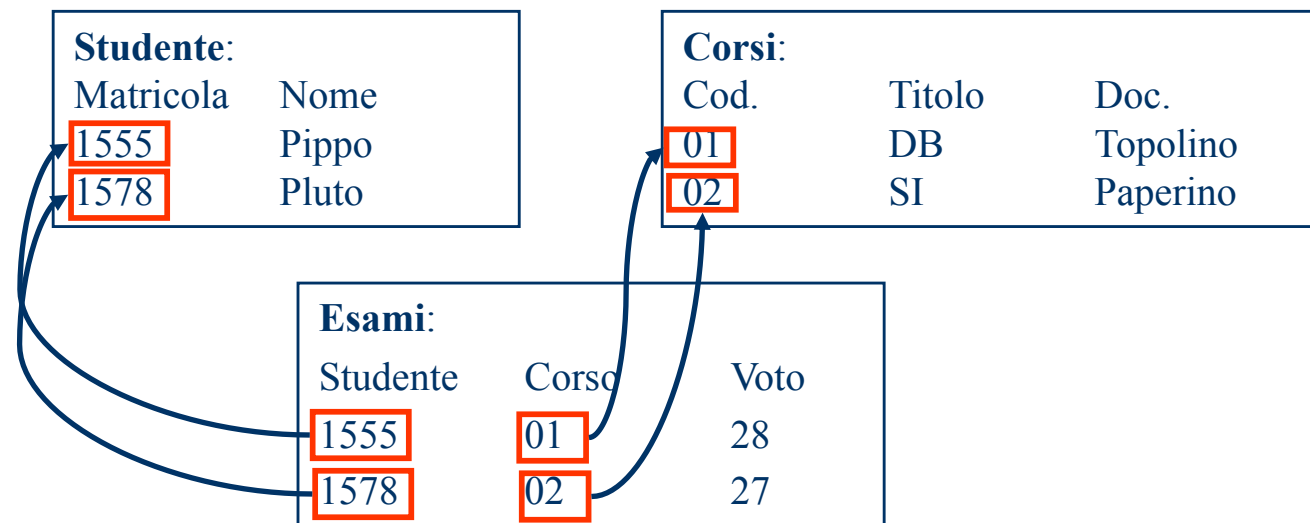
Istanza legale

- Se una istanza soddisfa tutti i vincoli di integrità specificati nello schema della base di dati, si parla allora di

istanza legale della base di dati



Relazioni e Basi di Dati



- Nel modello relazionale, una Base di Dati è costituita da più relazioni collegate, ove necessario e/o opportuno, attraverso valori comuni sulle tuple.

Relazioni e Basi di Dati (continua)

- L'esempio mostra una caratteristica fondamentale del modello relazionale:
 - Il modello relazionale è basato su valori
 - I riferimenti fra dati in relazioni diverse sono cioè rappresentati attraverso valori uguali dei domini che compaiono nelle tuple.
 - I modelli gerarchico e reticolare sono basati su “record e puntatori”.
- Vantaggi del modello basato su valori:
 - Rappresenta solo ciò che è rilevante.
 - Tutta l'informazione è contenuta nei valori ⇒ è facile trasferire informazioni tra vari contesti.
 - Non vi è alcun legame tra rappresentazione logica e fisica.
- A livello fisico anche un modello relazionale può essere (e, ovviamente, è) basato su puntatori.
 - Il punto è che essi non sono né visibili né necessari a livello logico.

- Uno schema di relazione è costituito da un simbolo (nome della relazione) e da un insieme di attributi
Studente (Matricola, Nome)
- Uno schema di basi di dati è (per ora) un insieme di schemi di relazione
 $R = \{ \text{Studente (Matricola, Nome)}$
 $\quad \text{Corsi (Codice, Titolo, Docente)}$
 $\quad \text{Esami (Studente, Corso, Voto)} \}$
- Una istanza di relazione (o relazione) è un insieme di tuple.
- Una istanza di base di dati sullo schema $R\{R_1, \dots, R_n\}$ è un insieme di relazioni $r\{r_1, \dots, r_n\}$ dove ogni r_i è una istanza di relazione sullo schema R_i .

Vincoli di Integrità

<u>Studente:</u>	<u>Matricola</u>	<u>Nome</u>
	1555	Pippo
	1578	Pluto
	1555	Gastone

<u>Corsi:</u>	<u>Titolo</u>	<u>Corso</u>	<u>Docente</u>
	01	DB	Topolino
	02	SI	Paperino

<u>Esami:</u>	<u>Studente</u>	<u>Corso</u>	<u>Voto</u>	<u>Lode</u>
	1556	01	28	Si
	1578	02	36	No
	?????			

- A chi affido il compito di controllare che tali situazioni non si verificano?
 1. A chi inserisce dati (utenti tramite applicazioni o direttamente)
 2. Direttamente al DBMS => vincolo di integrità

Vincoli di Integrità

- Si è introdotto, per evitare le situazioni appena viste, il concetto di Vincolo di Integrità, inteso come proprietà che devono essere soddisfatta dalle istanze
- Un vincolo è un predicato che associa alla generica istanza il valore vero\falso. Se il predicato è vero, il vincolo è soddisfatto.
- Uno schema di base dati è un insieme di schemi di relazioni e di vincoli.
- Una istanza è lecita (o corretta, o consistente) se soddisfa tutti i vincoli.

Vincoli di Integrità

- Si possono imporre vincoli sulle singole relazioni:
 - Vincoli **intra**relazionali
 - Sui singoli valori dei campi (vincoli di dominio)
 - Su più elementi della tupla o sull'intera tupla (vincoli di tupla)
 - Interessano tutte le tuple, l'una indipendentemente dalle altre
 - Sulle chiavi
- Oppure imporre vincoli che coinvolgono più relazioni
 - Vincoli **inter**relazionali

Vincoli Intrarelazionali

- vincoli di integrità intra-relazionale sono vincoli espressi attraverso condizioni logiche che devono essere soddisfatte all'interno di una singola relazione.
- In genere si parla di:
 - vincoli di dominio
 - vincoli di tupla
 - Vincoli di chiave

Esempio

NomeStudente	NomeEsame	Voto	Lode
Pippo	TSI	30	SI
Paolo	TM	28	NO
Marta	EI	18	SI
Maria	A1	33	NO

$18 \leq \text{voto} \leq 30$	$(\text{voto} \geq 18) \text{ AND } (\text{voto} \leq 30)$	Valore
La lode è ammissibile se voto è uguale a 30	$\text{NOT } ((\text{lode} == \text{'lode'}) \text{ AND } (\text{voto} != 30))$	Tupla
NOME STUDENTE	$\text{NOT}(\text{Cod_fiscale} = \text{NULL})$	Valore

$$IC_2 \leftarrow \neg(\text{Lode} = \text{'SI'} \wedge \text{Voto} \neq 30).$$

Funzione t estesa

Ricordiamo che data una tupla t_i definita su un insieme di attributi X , con la funzione $t_i[A]$ si ottiene il valore della tupla t_i relativamente al solo attributo A .

La stessa notazione è possibile estenderla ad un sottinsieme di attributi Y : in questo caso con il termine $t_i[Y]$ indichiamo la tupla t_i ristretta ai soli attributi Y .

NomeStudente	NomeEsame	Voto	Lode
Pippo	TSI	30	SI
Paolo	TM	28	NO
Marta	EI	18	NO
Maria	A1	30	NO

$$t_2[\text{NomeStudente}] = \text{Paolo}$$
$$t_2[\text{NomeStudente}, \text{Voto}] = (\text{Paolo}, 28).$$

- *Sia dato uno schema di relazione $R(X)$, e sia SK un sottinsieme di attributi di X . Diciamo che SK è una superchiave di una relazione r sullo schema $R(X)$ se **per ogni istanza**:*

$$\forall t_i, t_j \in r, i \neq j \Rightarrow t_i[SK] \neq t_j[SK]$$

- Un insieme di attributi K si dice superchiave per la relazione r se in r non esistono (e non devono esistere) due tuple t_1 e t_2 tali che $t_1[K]=t_2[K]$.

- Un insieme di attributi K si dice chiave se è una chiave minimale
 - se non esiste un'altra superchiave K_1 contenuta propriamente in K
 - *Oppure togliendo un qualsiasi attributo da K , K non è più chiave*
 - *Sono potenziali chiavi minimali chiavi composte da un unico attributo*
- *Un sottinsieme K di attributi X è chiave per r se è una superchiave minimale di r .*

Consideriamo lo schema di relazione:

STUDENTE(MATRICOLA,NOME,COGNOME,NASCITA,CDS)

- **Superchiavi**

- *MATRICOLA, NOME*
- *MATRICOLA, NOME, COGNOME*
- *NOME, COGNOME, NASCITA*
- *MATRICOLA*

- **Superchiavi minimali (chiavi)**

- *NOME, COGNOME, NASCITA*
- *MATRICOLA*

- **Chiave primaria**

- *MATRICOLA*

- In una generica relazione esiste sempre almeno una superchiave
 - la tupla intera è sicuramente una superchiave per la proprietà di unicità delle tuple (definizione di relazione)
- In generale, in una relazione è possibile individuare chiavi differenti
- Tra tutte le possibili chiavi, si sceglie sempre una chiave detta ***chiave primaria della relazione.***

Vincolo di chiave

- ***La chiave primaria di una qualunque relazione r deve essere***
 - *non nulla*
 - *ovviamente unica (non possono esistere due chiavi uguali in una stessa relazione)*
- *DBMS relazionale in presenza di una chiave primaria nulla **non permette** l'inserimento di una nuova tupla in una relazione.*
- *Dall'esempio precedente abbiamo:*
STUDENTE(MATRICOLA, NOME, COGNOME, NASCITA, CDS)

Scelta della chiave

- Esempio:

Matricola	Cognome	Nome	Nascita	Corso
0001	Rossi	Mario	1963	Inf
0002	Rossi	Luca	1963	DB
0003	Neri	Luca	1962	Analisi
0004	Neri	Luca	1960	DB

- {Corso, Cognome} è una superchiave minimale (chiave).
- Questo fatto probabilmente non è vero in generale.
- In altri termini, si tratta di una chiave reale o di una chiave “casuale” legata alla particolare istanza della relazione?
 - Si tratta di una chiave casuale legata al fatto, ad esempio, che Neri Luca del 1960 non segue il corso di Analisi.

Chiavi aggiunte

- **Id numerico**
 - Un codice numerico incorporato nella tupla
 - Codice di magazzino
 - Numero di inventario
- **Codice generico**
 - Matricola studente
 - ISBN dei libri
- **Codice chiamante**
 - Con parti che hanno significato rispetto alla tupla
 - Codice fiscale

Vincoli Interrelazionali

- In una base di dati solitamente si distribuisce l'informazione su relazioni differenti, in modo da evitare ridondanze dei dati
- La distribuzione delle informazioni richiede un meccanismo semplice che permetta di mettere in relazione i dati

STUDENTI

<u>matricola</u>	nome	cognome	indirizzo
150	Alex	Del Piero	via dei Palloni, 30
151	Martina	Stellina	via del Cielo, 40
142	Giovanni	SenzaTerra	via delle Crociate, 30

CORSI

<u>codice</u>	corso
TSI	Tecnologia dei Sistemi Informatici: Basi di Dati e Reti
EI	Elementi di Informatica
A1	Analisi Matematica 1

ESAMI

<u>matStudente</u>	<u>codiceCorso</u>	data	voto
150	TSI	10/10/04	30
150	A1	10/9/05	28
151	TSI	10/10/05	28

Provate a inserire tutte le informazioni in una unica tabella

Soluzione

Nome	Cognome	Indirizzo	Corso	Data	voto
Alex	Del Piero	Via dei palloni, 30	Tecnologia Sistemi informatici	10/10/04	30
Alex	Del Piero	Via dei palloni, 30	Analisi Matematica	10/09/05	28
Martina	Stellina	Via del Cielo, 40	Tecnologia Sistemi informatici	10/10/05	28
NULL	NULL	NULL	Elementi di informatica	NULL	NULL
Giovanni	SenzaTerra	Via delle Crociate, 30	NULL	NULL	NULL

- Ridondanza dei nomi degli studenti
- Valori nulli per assenza di informazioni

Vincoli di Integrità Referenziale

- La più importante classe di vincoli interrelazionali.
- Una tupla in una relazione che fa riferimento ad una tupla di un'altra relazione, deve far riferimento ad un tupla esistente.

<u>Studente:</u>	<u>Matricola</u>	<u>Nome</u>
	1555	Pippo
	1578	Pluto

<u>Esame:</u>	<u>Studente</u>	<u>Corso</u>	<u>Voto</u>	<u>Lode</u>
	1556	01	28	Si
	1578	02	36	No

<- errore

- Esame tabella interna, Studente tabella esterna.
- Si rappresenta come una freccia dalla tabelle interna verso la tabella esterna.

- **matStudente** (da *ESAMI*) è definita sullo stesso dominio dell'attributo **matricola** (chiave primaria di *STUDENTI*)
- **codiceCorso** (da *ESAMI*) è definito sullo stesso dominio dell'attributo **codice** (chiave primaria di *CORSI*).

**matStudente e codiceCorso sono dette
chiavi esterne**

- **NOTA:**
 - un valore di una **matStudente** in *ESAMI* deve essere anche presente come valore di **matricola** in *STUDENTE*
 - *Lo stesso vale per codicecorso e corso*

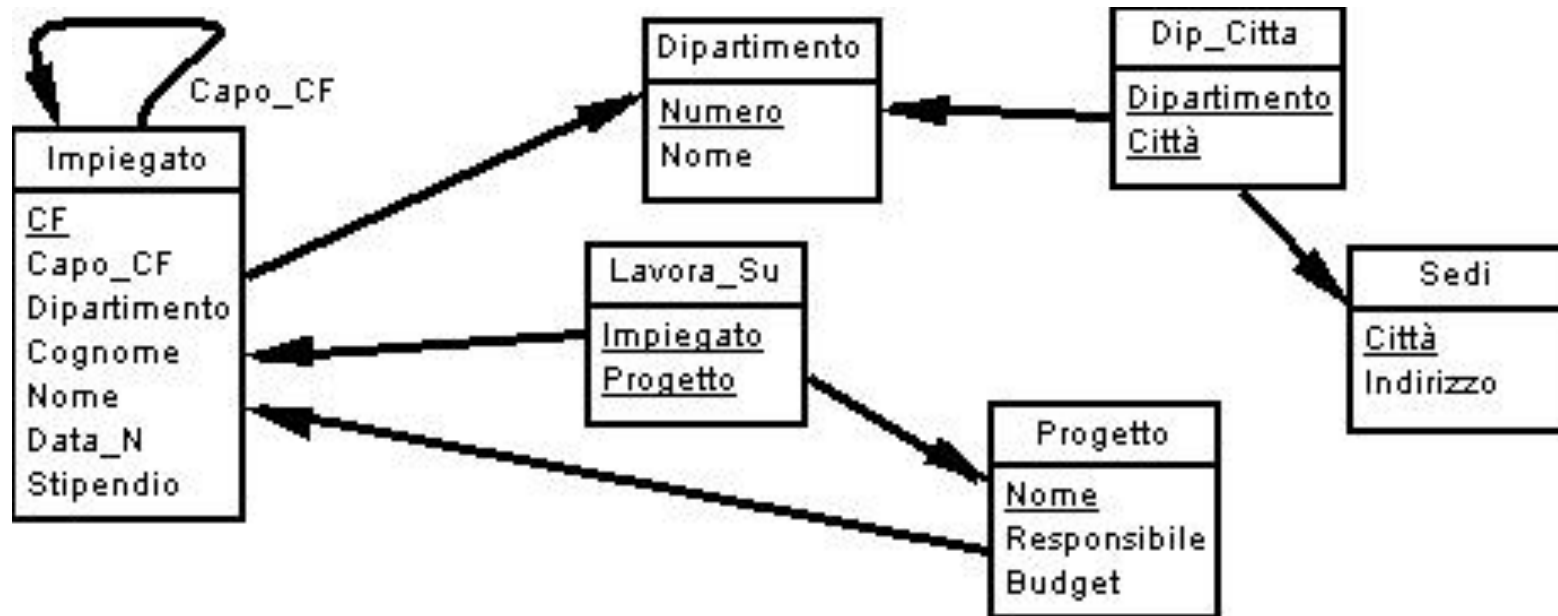
Integrità referenziale

- *Date due relazioni r_1 ed r_2 , con r_1 dotata di chiave esterna FK relativa alla chiave primaria PK della relazione r_2 .*
Si dice che tra r_1 ed r_2 sussiste un vincolo di integrità referenziale se, ogni occorrenza di FK in t_1 appartenente a r_1
 - *è NULL oppure*
 - *esiste una tupla $t_2 \in r_2 | t_1[FK] = t_2[PK]$*
- *Si noti che FK non è chiave primaria per r_1*

Concetto di integrità referenziale

- Per ogni occorrenza non nulla della chiave esterna nella tabella *referente* è presente un valore di chiave (primaria) nella tabella *riferita*.

Un primo esempio



- I vincoli rappresentati graficamente sono solo un parte dei vincoli che possiamo immaginare esistere
 - Lo stipendio di un impiegato deve essere \leq dello stipendio del proprio capo
 - Il responsabile di un progetto lavora sul progetto
 -

Esempio di Base di Dati Relazionale

Giocatore

CodTesserera	Nome	Cognome	Ruolo	Età	Squadra
--------------	------	---------	-------	-----	---------

Squadra

Nome	Colori-Sociali	Anno di Fondazione	Stadio
------	----------------	--------------------	--------

$IC_1^{Giocatore} \leftarrow Ruolo \in \{attaccante, difensore, portiere, centrocampista\}$

$IC_2^{Giocatore} \leftarrow Eta \leq 100 \wedge Eta \geq 0$

$IC_1^{Squadra} \leftarrow Nome \in \{Ascoli, Cagliari, Chievo, \dots, Udinese\}$

$IC_2^{Squadra} \leftarrow AnnodiFondazione \leq 2005 \wedge AnnodiFondazione \geq 1800$

- Seconda Tupla NON VALIDA (valore di ruolo)

CodTesserera	Nome	Cognome	Ruolo	Età	Squadra
00001	Francesco	Totti	attaccante	29	Roma
00002	Antonio	Picariello	professore	18	Avellino

- Seconda Tupla NON VALIDA (valore di età)

CodTesserera	Nome	Cognome	Ruolo	Età	Squadra
00001	Francesco	Totti	attaccante	29	Roma
00002	Antonio	Picariello	portiere	180	Avellino

- Seconda Tupla NON VALIDA (valore di NOME)

Nome	Colori Sociali	Anno di Fondazione	Stadio
Juventus	Bianco-Nero	1910	Delle Alpi
Napoli	Azzurro	1926	San Paolo

- Seconda Tupla NON VALIDA (valore di Anno di Fondazione)

Nome	Colori Sociali	Anno di Fondazione	Stadio
Juventus	Bianco-Nero	1910	Delle Alpi
Milan	Rosso-Nero	1200	San Siro/Meazza

Scelta della chiavi

Giocatore

CodTesserera	Nome	Cognome	Ruolo	Età	Squadra
00001	Francesco	Totti	attaccante	29	Roma
00002	Fabio	Cannavaro	difensore	30	Juventus
00003	Alessandro	Del Piero	attaccante	30	Juventus
00004	Francesco	Coco	difensore	29	Livorno

Squadra

Nome	Colori Sociali	Anno di Fondazione	Stadio
Juventus	Bianco-Nero	1910	Delle Alpi
Ascoli	Bianco-Nero	1920	Del Duca
Inter	Nero-Azzurro	1910	San Siro/Meazza
Milan	Rosso-Nero	1920	San Siro/Meazza

- **CHIAVI:**
 - CODTESSERA (**Primaria**) e NOME, COGNOME per Giocatore
 - NOME (Primaria) per Squadra
- **CHIAVE ESTERNA**
 - campo **Squadra** della relazione **Giocatore** e il campo **Nome** della relazione **Squadra**.