

VINCOLI DI INTEGRITA'

Esistono schemi relazionali sintatticamente corretti, ovvero il database è stato progettato correttamente, ma le informazioni inseribili possono non essere pertinenti. Per esempio in un campo *età* sarebbe un errore associare un valore negativo. Pertanto occorre stabilire dei criteri che permettano di evitare l'inserimento di informazioni non ammissibili: si parla quindi di **vincoli**.

I vincoli in una base di date sono necessari per garantire l'integrità dei dati.

MODELLO RELAZIONALE

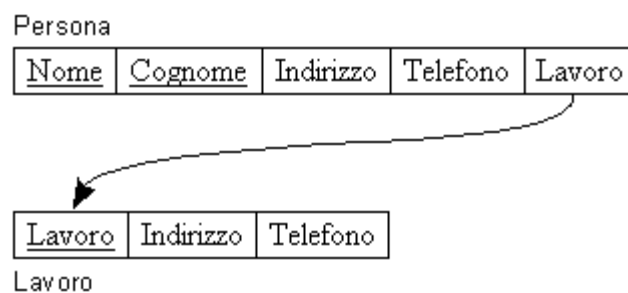
Uno schema relazionale è costituito da un insieme di relazioni fra loro associate che consentono di descrivere dal punto di vista logico una base di dati. Con il modello relazionale dei dati ci avviciniamo molto all'organizzazione che avrà la base dati nell'RDBMS. Una relazione, identificata da un nome, è una tabella di uno schema relazionale costituita da un insieme di tuple. Una tupla è una riga della tabella, ovvero, un record (istanza) della relazione. Ogni tupla è costituita da un insieme di campi, le colonne della tabella, che conservano il valore degli attributi della relazione. Come nel Modello ER, anche per i campi di una relazione va specificato un dominio di appartenenza. Inoltre uno o più campi di una relazione formano la chiave primaria (Primary Key) della tabella, nello stesso modo visto per le entità del Modello ER. E' importante notare che, anche se la relazione viene comunemente identificata con una tabella l'analogia non è totale. Infatti le tuple di una relazione non sono ordinate, non esiste la prima o l'ultima tupla, contrariamente a ciò che accade invece per le righe di una comune tabella.

In una relazione possono esserci particolari campi, che prendono il nome di chiavi esterne (Foreign Key), il cui compito è quello di realizzare legami logici tra le relazioni. Una chiave esterna è un attributo (o un insieme di attributi) di una relazione che, facendo riferimento alla chiave primaria di un'altra relazione, realizza l'associazione tra le due tabelle. Sappiamo infatti che attraverso la chiave primaria possiamo individuare una precisa tupla di una relazione. La chiave esterna di una relazione contiene dei dati che si riferiscono alla chiave primaria di un'altra relazione, creando così un legame logico tra le tuple delle due tabelle.

Il Modello Relazionale che rappresenta il mini-modo di interesse può essere ricavato direttamente dal Modello ER, attraverso una sequenza di operazioni di conversioni. Ecco i passi fondamentali da seguire.

Esempio

Come esempio ecco una rappresentazione grafica di quanto appena detto, che si riferisce al caso della rubrica personale trattato nella prima lezione. Le relazioni vengono rappresentate tramite rettangoli divisi in celle, ogni cella rappresenta un campo della relazione. Le chiavi primarie (PK) vengono sottolineate, mentre dalle chiavi esterne (FK) parte una freccia verso le chiavi primarie a cui si riferiscono.



Per convincerci dell'efficacia della chiave esterna nella creazione di legami logici tra relazioni, basta vedere come sarebbero strutturati i dati delle due relazioni che costituiscono la nostra rubrica personale. Osserviamo in particolare come i dati della colonna "Lavoro" (FK) delle prima relazione referenziano quelli presenti nell'omonima colonna della seconda relazione.

<u>Nome</u>	<u>Cognome</u>	<u>Indirizzo</u>	<u>Telefono</u>	<u>Lavoro</u>
Michele	Brambilla	via delle Puglie 9	0332 7827161	Banca Europe
Alessandro	De Magistris	p.zza Matteotti 5	2123 2112231	KL Elettronix
Paperon	De Paperoni	viale delle Libertà 19	2311 2311231	NULL
Luca	Rossi	via Merano 1	3445 7843924	Banca Europe

<u>Lavoro</u>	<u>Indirizzo</u>	<u>Telefono</u>
Banca Europe	via del Lavoro 7	099 4567899
KL Elettronix	via Nova 134	123 45221123

Il Modello Relazionale spesso va accompagnato, oltre che dalla sua rappresentazione grafica, anche da un dizionario dei dati. Tale dizionario è costituito da una descrizione testuale che ha il compito di indicare i domini di tutti gli attributi di ogni tabella con i relativi domini di appartenenza, individuare le chiavi primarie e le chiavi esterne e a volte riporta anche le descrizioni delle relazioni.

Persona

Nome String PK
Cognome String PK
Indirizzo String
Telefono String
Lavoro String FK (referenzia il campo Lavoro della relazione Lavoro)

Lavoro

Lavoro String PK
Indirizzo String
Telefono String

TIPOLOGIA DI VINCOLI

I vincoli possono essere così classificati:

A. **Vincoli intra-relazionali** (coinvolgono una o più tuple della stessa relazione)

1. Vincoli di dominio
2. Vincoli di tupla
3. Vincolo NOT NULL
4. Vincoli di chiave

B. **Vincoli interrelazionali** (coinvolgono più tuple di relazioni diverse)

1. Integrità referenziale

A. VINCOLI INTRA-RELAZIONALI

1. VINCOLO DI DOMINIO

Per ogni attributo è necessario specificare il tipo di dato. Per esempio età deve essere del tipo **numerico**.

2. VINCOLO DI TUPLA

Vincolo che coinvolge più valori (associati ai campi dello stesso record) della stessa tupla.

Per esempio, nella relazione:

ESAMI(Studente, Materia, Data, Voto, Lode)

- La Lode può essere assegnata solo se il Voto è 30.
- Il Voto deve essere compreso tra 18 e 30 ($Voto \geq 18$ AND $Voto \leq 30$).

3. Vincolo NOT NULL

Particolare vincolo di tupla. Ad un campo con tale vincolo occorre obbligatoriamente assegnare un valore. Pertanto una colonna della tabella NOT NULL deve essere sempre valorizzata.

4. VINCOLI DI CHIAVE

Vincoli che coinvolgono più tuple della stessa relazione.

La chiave di una relazione non consente valori nulli (NOT NULL) e valori duplicati. La chiave infatti identifica in modo univoco una tupla della relazione.

B. VINCOLI INTERAZIONALI

1. VINCOLI DI INTEGRITA' REFERENZIALE

Detti anche vincoli di Foreign Key (chiave esterna), Referential Integrity Constraints.

Gli attributi di una data tabella (slave, secondaria) possono assumere soltanto dei valori specificati in un'altra tabella (master, primaria).

NB I vari linguaggi per basi dati solitamente implementano appositi costrutti per questo vincolo (ad esempio la foreign key o chiave esterna in SQL).

Esempio

Album(TitoloAlbum, Anno, Prezzo)

Brani(TitoloBranco, Durata, TitoloAlbum^{Album})

Fra Album e Brani esiste un'associazione 1:N. Un Album, infatti, può contenere uno o più brani; un brano è contenuto in un solo album (nel nostro minimondo, per semplicità, supponiamo che lo stesso brano non sia contenuto in differenti album).

Album	
Nome campo	Tipo dati
TitoloAlbum	Testo breve
Anno	Numerico
Prezzo	Valuta

Brani	
Nome campo	Tipo dati
TitoloBranco	Testo breve
Durata	Numerico
TitoloAlbum	Testo breve

Generale	Ricerca
Visualizza controllo	Casella di riepilogo
Tipo origine riga	Tabella/query
Origine riga	Album
Colonna associata	1
Numero colonne	3
Intestazioni colonne	No
Larghezza colonne	
Consenti più valori	No
Consenti modifiche a ele	Sì
Maschera di modifica voc	
Mostra solo valori origine	No

Il vincolo può essere espresso in questo modo:

Modifica relazioni

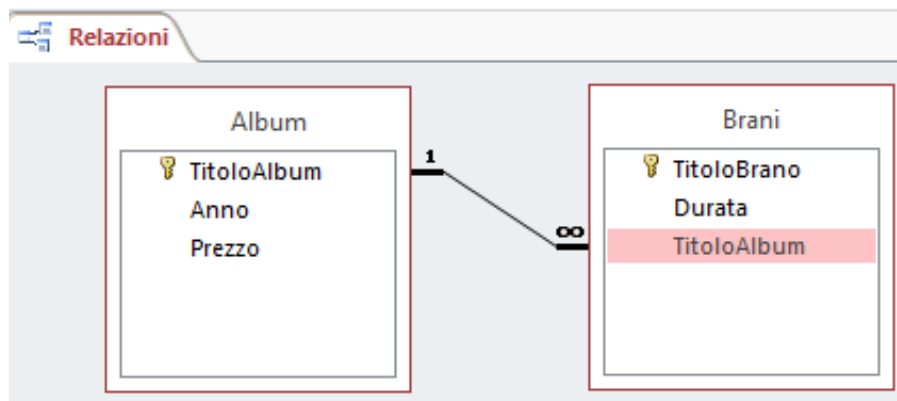
Tabella/query: Album Tabella/query correlata: Brani

TitoloAlbum TitoloAlbum

Applica integrità referenziale
 Aggiorna campi correlati a catena
 Elimina record correlati a catena

Tipo relazione: Uno-a-molti

Buttons: Crea, Annulla, Tipo join..., Crea nuova..



QUERY

SELECT *

FROM Album, Brani

WHERE Album.TitoloAlbum=Brani.TitoloAlbum;

Analizziamo in dettaglio questo fondamentale vincolo:

DEFINIZIONE DI INTEGRITA' REFERENZIALE

L'**integrità referenziale** (referential integrity) è un insieme di regole del modello relazionale che garantiscono l'integrità dei dati quando si hanno relazioni associate tra loro attraverso la **chiave esterna**: queste regole servono per rendere valide le associazioni tra le tabelle e per eliminare gli errori di inserimento, cancellazione o modifica di dati collegati tra loro.

I vincoli d'integrità referenziale

- Servono a garantire che i riferimenti tra tabelle siano possibili.
- Un vincolo di integrità referenziale fra insiemi di attributi X di una relazione R1 ed una relazione R2 è soddisfatto se i valori su X di ciascun record di R2 (eccetto potenzialmente i valori NULL) compaiono come valori della chiave primaria di R2. In altre parole **R1 non può contenere valori per X che non esistano in R2 (eccetto NULL)**.



Consideriamo la seguente base di dati nel modello relazionale:

infrazioni	Codice	Data	Vigile	Prov	Numero
	65524	3/9/1997	343	MI	3K9886
	87635	4/12/1997	476	MI	6D5563
	82236	4/12/1997	343	RM	7C5567
	35632	6/1/1998	476	RM	7C5567
	76543	5/3/1998	548	MI	6D5563

vigili	Matricola	Cognome	Nome
	343	Rossi	Luca
	476	Neri	Pino
	548	Nicolosi	Gino

automobili	Prov	Numero	Proprietario	...
	MI	3K9886	Nestore	...
	MI	6D5563	Nestore	...
	RM	7C5567	Menconi	...
	RM	1A6673	Mussone	...
	MI	5E7653	Marchi	...

Le informazioni della relazione INFRAZIONI sono rese significative e complete attraverso il riferimento alle altre due relazioni. I riferimenti sono significativi in quanto i valori della relazione

INFRAZIONI sono uguali ai valori effettivamente presenti nelle alte due. Informazioni in relazioni diverse sono quindi correlate attraverso valori comuni, in particolare, valori delle chiavi primarie solitamente. Un vincolo di integrità referenziale fra un insieme di attributi X (per esempio Prov e Numero) di una relazione R1 (INFRAZIONI) e un'altra relazione R2 (AUTOMOBILI), impone ai valori su X di ciascuna enupla dell'istanza di R1 di comparire come valori della chiave (primaria) dell'istanza di R2. Nell'esempio, esistono vincoli di integrità referenziale fra l'attributo Vigile della relazione INFRAZIONI e la relazione VIGILI e tra gli attributi Prov e Numero di INFRAZIONI e la relazione AUTO. Garantendo il vincolo di integrità referenziale, dalla relazione INFRAZIONI possiamo risalire univocamente a qualsiasi valore delle relazione AUTOMOBILI e VIGILI.

Consideriamo le seguenti relazioni:

ORDINI		
NumOrdine	DataOrdine	CodCliente
1	10/02/2014	Tech
2	10/02/2014	Giga
3	12/03/2014	
4	01/03/2014	Eboo
5	20/03/2014	Tech
6	32/03/2014	Stor
7	15/04/2014	Giga
8	16/04/2014	Acer

CLIENTI		
<u>CodCliente</u>	RagioneSociale	Città
Tech	Technology System	Milano
Giga	GigaBit	Varese
Stor	StoreOne	Milano
Eboo	EbookStore	Torino
	InfoTech	Varese
Link	LinkOffice	Milano
Tech	TechBit	Torino



In esse sono stati evidenziati alcuni campi che invalidano i corrispondenti record e quindi l'intero schema relazionale risulta inconsistente.

- L'ordine 3 non è riferito ad alcun cliente.
- La data dell'ordine 6 è assurda.
- L'ordine 8 è riferito a un cliente inesistente.
- Al cliente con ragione sociale InfoTech non è abbinato alcun valore della chiave e, di conseguenza, non è possibile collegarla ad alcun ordine.

- Il cliente riportato come ultima tupla contiene un valore di chiave ripetuto.

Risulta invece possibile l'esistenza nella relazione CLIENTI di un cliente che non ha fatto alcun ordine (finora).

Per prevenire queste situazioni sono state definite opportune regole di integrità sui dati tra le quali ricordiamo l'integrità di chiave (è la regola d'integrità violata in CLIENTI).

I vincoli di tupla esprimono condizioni che devono essere soddisfatte dai valori di ciascuna tupla indipendentemente dalle altre. La data dell'ordine 6 è un esempio di violazione di un vincolo di dominio, in quanto il valore 32/03/2014 non ricade nell'insieme delle date ammissibili.

Anche per la chiave esiste il vincolo NOT NULL come riportato nella quinta tupla della relazione CLIENTI; allo stesso modo non possono esistere valori ripetuti per la chiave come indicato nell'ultima tupla della stessa relazione CLIENTI.

L'integrità referenziale è un insieme di regole del modello relazionale che garantiscono l'integrità dei dati quando si hanno relazioni associate tra loro attraverso la chiave esterna (foreign key).

Queste regole servono per rendere valide le associazioni tra le tabelle e per eliminare gli errori di inserimento, cancellazione o modifica dei dati tra loro collegati.

L'integrità referenziale viene rispettata quando per ogni valore non nullo della chiave esterna, esiste un valore corrispondente della chiave primaria nella tabella associata.

Quando viene applicata l'integrità referenziale, è necessario osservare le seguenti regole pratiche:

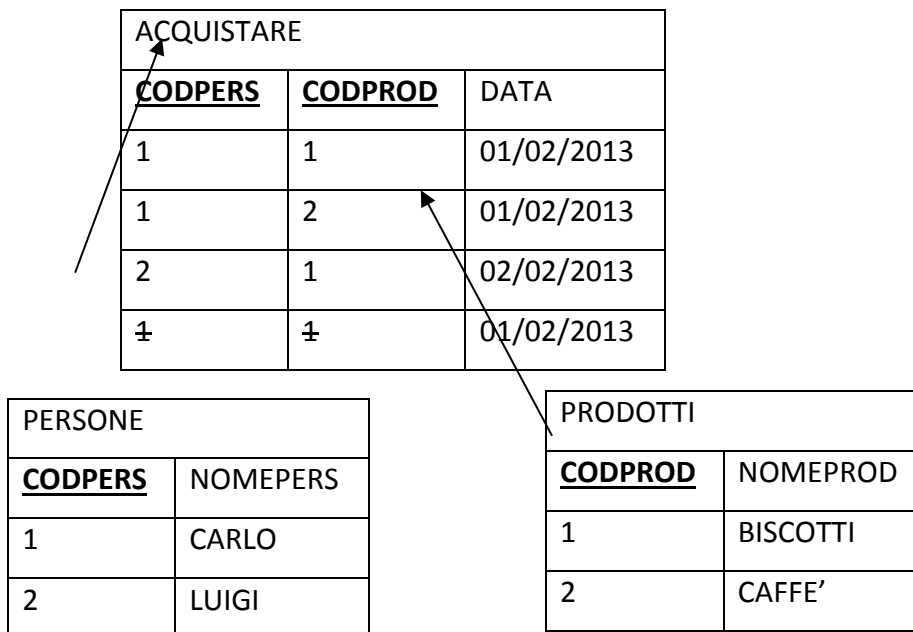
- Non è possibile immettere un valore nella chiave esterna della tabella associata, se tale valore non esiste tra le chiavi della tabella primaria.
- Non è possibile eliminare una tupla dalla tabella primaria, se esistono righe legate ad essa attraverso la chiave esterna nella tabella correlata.
- Non si può modificare, come è ovvio, il valore della chiave nella tabella primaria, se ad essa corrispondono righe nella tabella correlata.

Esempio: associazioni N:N

PERSONE(CODPERS, NOME PERS)

PRODOTTI(CODPROD, NOME PROD)

ACQUISTARE(CODPERS^{PERSONE}, CODPROD^{PRODOTTI}, DATA)



Applicare l'integrità referenziale al precedente schema relazionale significa garantire che ogni valore presente nella tabella ACQUISTARE per la chiave esterna CODPERS abbia un corrispondente valore di chiave CODPRES in una delle righe della tabella PERSONE, e analogamente, che ogni valore presente nella tabella ACQUISTARE per la chiave esterna CODPROD abbia un corrispondente valore di chiave CODPROD in una delle righe della tabella PRODOTTI.

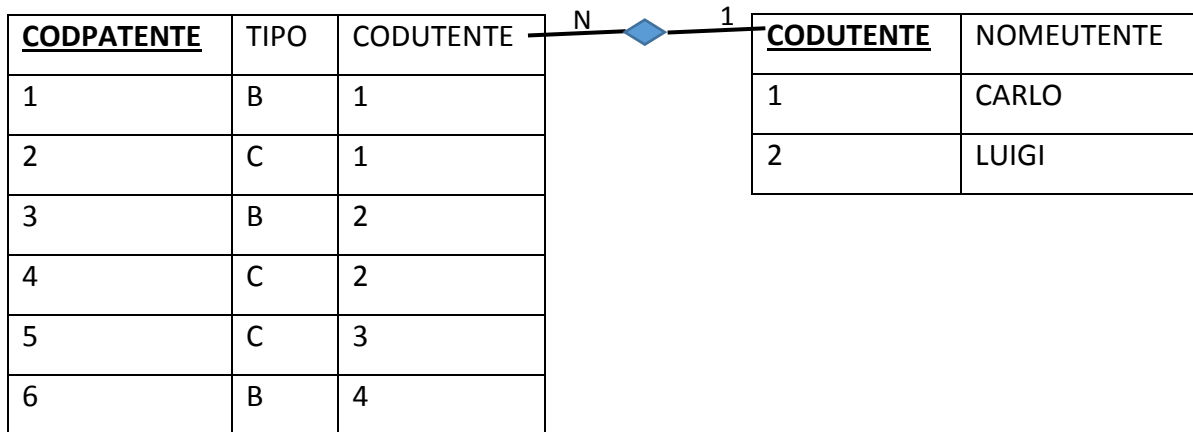
Inoltre non si deve consentire la cancellazione di una persona o di un prodotto rispettivamente dalle tabelle PERSONE e PRODOTTI se ci sono righe nella tabella ORDINI che si riferiscono ad essi.

L'integrità referenziale, se applicata, non permette che si presentino situazioni come quelle dell'ordine 8 descritto nel primo esempio proposto, riferito ad un cliente inesistente, di cancellare la riga relativa al cliente Eboo per la presenza di un ordine collegato, oppure di modificare il valore della chiave Giga per la medesima ragione.

Esempio: associazioni 1:N

UTENTI(CODUTENTE, NOMEUTENTE)

PATENTI(CODPATENTE, TIPO CODUTENTE^{UTENTI})



In questo schema l'integrità referenziale viene violata per esempio nei seguenti casi:

- Presenza nella tabella Patenti dei record con chiave primaria 5 e 6 che hanno come chiave esterna un valore non esiste nella tabella Utenti come chiave primaria. CodUtente 3 e 4 infatti non esistono nella tabella Utenti.

CHIAVE ESTERNA

Una chiave esterna chiamata generalmente "**foreign key**" (FK) specifica un **vincolo di integrità referenziale**, questa chiave punta a una chiave di un'altra relazione.

Una FK soddisfa due regole:

- Gli attributi presenti nella FK hanno gli stessi domini degli attributi di chiave primaria (PK).
Una FK di una tupla in R_1 fa riferimento a una chiave di una tupla in R_2 .
- Un valore FK di una tupla in R_1 o è presente a un valore di una tupla in R_2 o non esiste quindi è nullo.

Nel primo caso si ha: $t_1[FK] = t_2[FK]$ e si dice che t_1 "riferisce o "fa riferimento" alla tupla t_2 .
Se questa condizione esiste si dice che **sussiste un vincolo di integrità referenziale**, in questo caso R_1 è detta relazione referenziante (che fa riferimento), mentre R_2 è detta relazione riferita (alla quale si fa riferimento).

Vincolo di integrità dell'entità

Il **vincolo di integrità dell'entità** stabilisce che nessun valore chiave può essere nullo.

In precedenza abbiamo parlato di una terza categoria di vincoli, quelli basati sulle applicazioni. Questi tipi di vincoli, sono anche chiamati **vincoli di integrità semantici**, ovvero sono dei vincoli legati al significato, che vengono posti non nel DDL ma negli applicativi con cui si manipolano le basi di dati.

Per esempio, se io ho una relazione STUDENTE con un campo ANNO e uno CREDITI è normale che se uno studente ha come valore nel campo anno=3 i crediti non potranno mai essere 0, perché per entrare nel terzo anno bisogna aver maturato un certo numero di crediti.

Questi vincoli quindi, o vengono imposti dalle applicazioni per gestire le basi di dati, oppure viene utilizzato un "linguaggio di specifica dei vincoli" di uso generale.

Di questa categoria di vincoli ne fanno parte i **vincoli di stato**, che definiscono i vincoli per cui uno stato sia dichiarato valido, questi vincoli sono a loro volta definiti da **vincoli di transizione** ad esempio se un campo di una relazione ha come valore il numero delle volte che un libro viene prestato, questo valore può solo aumentare e mai diminuire, al massimo rimane costante.

Vincolo di integrità referenziale

L'integrità referenziale (referential integrity) è un insieme di regole del modello relazionale che garantiscono l'integrità dei dati quando si hanno relazioni associate tra loro attraverso la chiave esterna: queste regole servono per rendere valide le associazioni tra le tabelle e per eliminare gli errori di inserimento, cancellazione o modifica di dati collegati tra loro.

L'integrità referenziale viene rispettata quando per ogni valore non nullo della chiave esterna, esiste un valore corrispondente della chiave primaria nella tabella associata.

Quando viene applicata l'integrità referenziale, è necessario osservare le seguenti regole pratiche:

- Non è possibile immettere un valore nella chiave esterna della tabella associata, se tale valore non esiste tra le chiavi della tabella primaria.
- Non è possibile eliminare una n-upla dalla tabella primaria, se esistono righe legate ad essa attraverso la chiave esterna nella tabella correlata
- Non si può modificare, come è ovvio, il valore della chiave nella tabella primaria, se ad essa corrispondono righe nella tabella correlata.

Un vincolo di integrità referenziale ("Foreign key") fra gli attributi X di una relazione R1 e un'altra relazione R2, impone su X in R1 di comparire come valori della chiave primaria di R2.

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

Nella tabella "Esami" vi sono dei valori di FK che non sono presenti nelle tabelle "Studenti" e "Corsi". È, infatti, opportuno definire vincoli di integrità referenziale fra:

- L'attributo CodCorso della relazione Esami e la relazione Corsi, e l'attributo Matricola della relazione Esami e la relazione Studenti.

Se uno studente ha conseguito un voto in un esame, deve comparire anche nella relazione Studenti.

Se un corso compare nella relazione Esami, deve comparire anche nella relazione Corsi.

L'integrità referenziale assicura la consistenza e la correttezza dei record presenti in un database, evitando che ci siano record orfani, ovvero che non hanno riferimento nella tabella collegata.