

Modello Relazionale

RELAZIONE: dati n domini D_1, D_2, \dots, D_n , (non necessariamente distinti), essa è un sottoinsieme del prodotto cartesiano $D_1 \times D_2 \times \dots \times D_n$

- n è detto grado di R ;
- il numero di tuple che si viene a creare è detta cardinalità di R ;

Esempio: $D_1 = \{v_{11}, v_{12}\}$ $D_2 = \{v_{21}, v_{22}, v_{23}, v_{24}\}$

$R \subseteq D_1 \times D_2 = \{(v_{11}, v_{21}), (v_{11}, v_{22}), (v_{11}, v_{23}), (v_{11}, v_{24}), (v_{12}, v_{21}), (v_{12}, v_{22}), (v_{12}, v_{23}), (v_{12}, v_{24})\}$

tale relazione è di grado 2 e cardinalità 8

$R_1 = \{(v_{11}, v_{21}), (v_{11}, v_{22})\}$ --> grado 2 e cardinalità 2

$R_1 = \{\}$ --> grado e cardinalità = 0

NB: ogni relazione la si può rappresentare in una tabella avente per caratteristiche:

- colonne = grado;
- righe = cardinalità;

Attributo: nome dato ad un dominio in una relazione;

Schema di relazione: coppia costituita dal nome di R e da un insieme di nomi degli attributi $X = (A_1, A_2, \dots, A_n)$ e si indica con $R(X)$;

Istanza di relazione: dato $R(X)$, è un insieme r di tuple (A_1, A_2, \dots, A_n)

Schema di Basi di Dati: è un insieme di schemi di relazione $R = \{R_1(X)_1, R_2(X)_2, \dots, R_n(X)_n\}$

Istanza di Basi di Dati: è un insieme di istanze di relazione $r = \{r_1, r_2, \dots, r_n\}$ dove r_i è una relazione su R_i , $\forall 1 \leq i \leq n$

Insieme di attributi: Y di $R(X)$ cioè $Y \subseteq X$ oppure $R.Y$, A denota $\{A\}$ e XY denota $\{X\} \cup \{Y\}$

Tuple: dati, t su $R(X)$, $A \in X$ e $Y \subseteq X$, si ha che:

- $t[A] = t.A$ denota il valore t su A ;
- $t[Y]$ denota la tupla t ottenuta considerando i valori degli attributi in Y ;

Chiavi

Per chiave di una relazione si intende un sottoinsieme di suoi attributi che identifica univocamente ogni tupla della relazione stessa.

Dati $R(X)$, $K \subseteq X$ si dice chiave per $R(X)$ se e solo se per ogni r su $R(X)$ valgono le seguenti proprietà:

- Univocità: $\forall t_1, t_2 \in r, t_1[K] = t_2[K] \rightarrow t_1 \equiv t_2$ ossia non esistono 2 tuple di r con lo stesso valore della chiave
- Minimalità: $\forall A_i \in K, K - A_i$ non verifica la proprietà di unicità (ossia non esiste un sottoinsieme proprio di K con la proprietà di univocità)

Superchiave: chiave che soddisfa solo la condizione di univocità

NB: ogni schema di relazione $R(X)$ ha almeno una chiave

Uno schema $R(X)$ può avere più chiavi, dette chiavi candidate. Tra esse, ne viene scelta una detta chiave primaria, mentre le altre vengono dette chiavi alternative. Per indicare la chiave primaria, si sottolineano gli attributi che la identificano, mentre per identificare le chiavi alternative, si scrivono dopo AK gli attributi che la identificano.

Vincolo di Entity Integrity: gli attributi che costituiscono la chiave primaria di una relazione non possono assumere valori nulli $\forall t \in r, t[K_i] \neq \text{null}, \forall 1 \leq i \leq m$

Vincolo di Integrità Referenziale: assicura che quando in una tupla si utilizza il valore di un attributo per riferirsi ad un'altra tupla, quest'ultima sia una tupla esistente. Tale vincolo viene dichiarato specificando:

- Chiave esterna (FK): insieme di attributi $FK = \{FK_1, FK_2, \dots, FK_m\}$ di uno schema di relazione $R_1 \in R$
- Chiave della relazione riferita: schema di relazione $R_2 \in R$, non necessariamente distinti da R_1 con chiave $K = \{K_1, K_2, \dots, K_m\}$

Formalmente, un'istanza $r = \{r_1, r_2, \dots\}$ su R , soddisfa il vincolo di integrità referenziale se e solo se:

$\forall t_1 \in r_1, (t_1[FK_i] = \text{null} \vee \exists t_2 \in r_2 / t_1[FK_i] = t_2[FK_i]) \forall 1 \leq i \leq n$

Un'istanza è detta "Istanza legale di Base di Dati", se dati $R = \{R_1(X)_1, R_2(X)_2, \dots, R_n(X)_n\}$ e un'istanza $r = \{r_1, r_2, \dots, r_n\}$ su R , si ha:

- ogni relazione r_i soddisfa il vincolo di entity integrity (primary key \neq null);
- r soddisfa tutti i vincoli di integrità referenziale su R ;

Forme Normali

L'obiettivo della forma normale è quello di definire formalmente la qualità degli schemi. La qualità di uno schema viene formalmente definita con l'assenza di:

- Ridondanza nei dati;
- Anomalie di aggiornamento nei dati;

Dipendenza Funzionale: dato uno schema di relazione $R(X)$, una dipendenza funzionale (FD) su R , è un vincolo di integrità espresso nella formula $Y \rightarrow Z$, dove Y e Z sono sottoinsiemi di X ; in tal caso si dice che Y determina funzionalmente Z

Dato uno schema $\langle R(T), F \rangle$, un attributo $A \in T$ e un insieme di attributi $Y \subseteq T$, si dice che A dipende da Y se $Y \rightarrow A \in F$ e non esiste nessun sottoinsieme proprio di Y , $Z \subset Y, Z \rightarrow A$.

1NF) Uno schema $R(X)$ è in 1NF se e solo se i valori di tutti i domini degli attributi $A \in X$ sono atomici. Generalmente nel modello relazionale, si assume come implicito il vincolo di avere domini atomici per gli attributi.

Es: considero lo schema di relazione FREQUENZA
FREQUENZA (MATR, CODCOR, NUMEROORE, CODDOC)

In questo caso $CODCOR \rightarrow CODDOC$ ($CODCOR$ determina funzionalmente $CODDOC$), così abbiamo RIDONDANZA di dati in ogni tupla, che determina varie anomalie, ossia si hanno delle dipendenze parziali di un attributo dalla chiave. Per questo definiamo la 2NF che serve per definire schemi esenti da tali problemi.

2NF) Uno schema $\langle R(X), F \rangle$ è in 2NF se e solo se ogni attributo non primo $A \in T$ dipende completamente da ognuna delle chiavi di R .

Si dice che un attributo di $R(X)$ è primo se e solo se appartiene ad almeno una chiave di $R(X)$

Es: considero lo schema di relazione CORSO
CORSO (CODCOR, NOME, CODDOC, CODDIP)

In questo caso, CODDOC --> CODDIP (CODDOC determina funzionalmente CODDIP), così abbiamo RIDONDANZA di dati in ogni tupla, che determina varie anomalie, ossia si ha che un attributo NON superchiave determina funzionalmente un altro attributo. Per questo motivo definiamo la 3NF che serve per definire schemi esenti da tali problemi.

3NF) Uno schema $\langle R(X), F \rangle$ è in 3NF se e solo se per ogni dipendenza funzionale non banale, $X \twoheadrightarrow A \in F$ o X è una superchiave o A è primo.

Teorema: uno schema $\langle R(X), F \rangle$ che è in 3NF è anche in 2NF

La 3NF definisce schemi con un buon livello di qualità e costituisce quindi lo standard da noi adottato per il progetto logico.

Progetto logico di alto livello con il modello E/R

1. Ristrutturazione dello schema E/R: è indipendente dal modello logico e consiste nell'eliminazione di gerarchie e identificazioni esterne, normalizzazione di attributi composti o multipli, scelta di chiavi primarie;
2. Traduzione verso il modello logico: è riferita al modello relazionale e consiste nella traduzione di entità e associazioni in schemi relazionali;

Dopo la prima fase, lo schema E/R è costituito soltanto da entità associazioni e attributi semplici. Ogni trasformazione impoverisce semanticamente lo schema; la semantica persa resterà sotto forma di vincoli di integrità che governeranno l'uso delle relazioni.

Bisogna tenere conto anche che l'incidenza di valori nulli per attributi opzionali va limitata.

Il carico di lavoro sul DB è rappresentato sia dalla dimensione dei dati che dalle operazioni più significative che si stima saranno eseguite sul DB. Inoltre per la regola 20-80 si considera che il 20% delle operazioni produce l'80% del carico.

Volume dei dati: - numero medio di istanze di ogni entità e associazione
 - cardinalità e dimensioni di ciascun attributo
 - percentuali di copertura di gerarchia

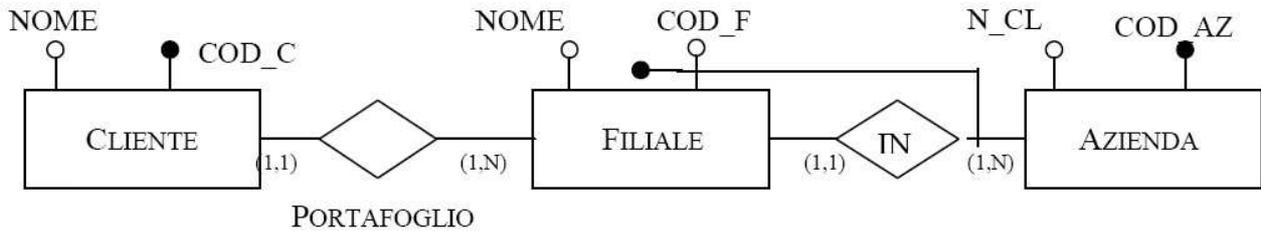
Descrizione delle operazioni: - tipo: iterativa o batch;
 - frequenza: numero medio di esecuzioni in un certo periodo di tempo;
 - schema di operazione: frammento dello schema E-R interessato all'operazione;
 - schema di navigazione: schema di operazione con l'indicazione del "cammino logico" da percorrere per accedere alle informazioni di interesse;

Il peso degli accessi in scrittura è il doppio del peso degli accessi in lettura

Definizione: un dato derivato è un dato che può essere ottenuto attraverso una serie di operazioni da altri dati

Esempio Numerico

Sia dato il seguente schema E/R, volume dei dati e operazioni, nel quale l'attributo derivato N_CL è la somma di tutti i clienti delle filiali di una certa azienda.



Siano date le seguenti operazioni e volumi dei dati

Operazione 1

Inserimento di un nuovo CLIENTE (in questa operazione si suppongono noti e validi il COD_C, il NOME ed il COD_AF della FILIALE);

Operazione 2

Visualizzazione di tutti gli attributi di un'AZIENDA (compreso il N_CL).

Tavola dei volumi

CONCETTO	TIPO	VOL.
AZIENDA	E	10
FILIALE	E	100
CLIENTE	E	2000

Tavola delle operazioni

OPER.	TIPO	FREQ.
Oper. 1	I	200/giorno
Oper. 2	I	50/giorno

Decidere se è conveniente conservare nello schema l'attributo derivato. Si trascuri l'occupazione di memoria del dato derivato.

Senza il dato derivato

	Concetto	Acc.	Tipo
Operazione 1 2 accessi in scrittura. 4*200 = 800/Giorno	CLIENTE	1	S
	PORTAFOGLIO	1	S
Operazione 2 211 accessi in lettura 411*50 = 10550/Giorno	AZIENDA	1	L
	IN	10	L
	PORTAFOGLIO	200	L
Totale: 11350/Giorno			

Con il dato derivato

	Concetto	Acc.	Tipo
Operazione 1 2 accessi in lettura 3 accessi in scrittura 8*200 = 1600/Giorno	CLIENTE	1	S
	PORTAFOGLIO	1	S
	IN	1	L
	AZIENDA	1	L
Operazione 2 1 accesso in lettura 1*50 = 50/Giorno	AZIENDA	1	S
	AZIENDA	1	L
Totale: 1650/Giorno			

La soluzione più conveniente è quella di mantenere il dato derivato

Eliminazione delle Gerarchie

Le alternative che si possono presentare sono 3:

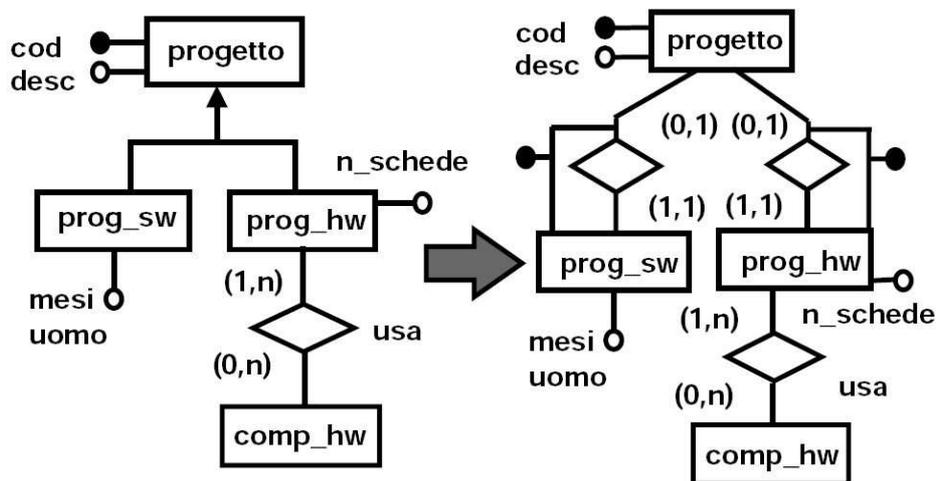
- mantenimento delle entità con associazioni (sempre possibile);
- collasso verso l'alto;
- collasso verso il basso;

Le trasformazioni delle gerarchie possono modificare le cardinalità di attributi ed associazioni;

L'applicabilità e la convenienza delle soluzioni dipendono dalle proprietà di copertura della gerarchia e delle operazioni previste;

a) *Mantenimento delle entità con associazioni*

Le entità figlie sono in associazione con l'entità padre e sono identificate esternamente

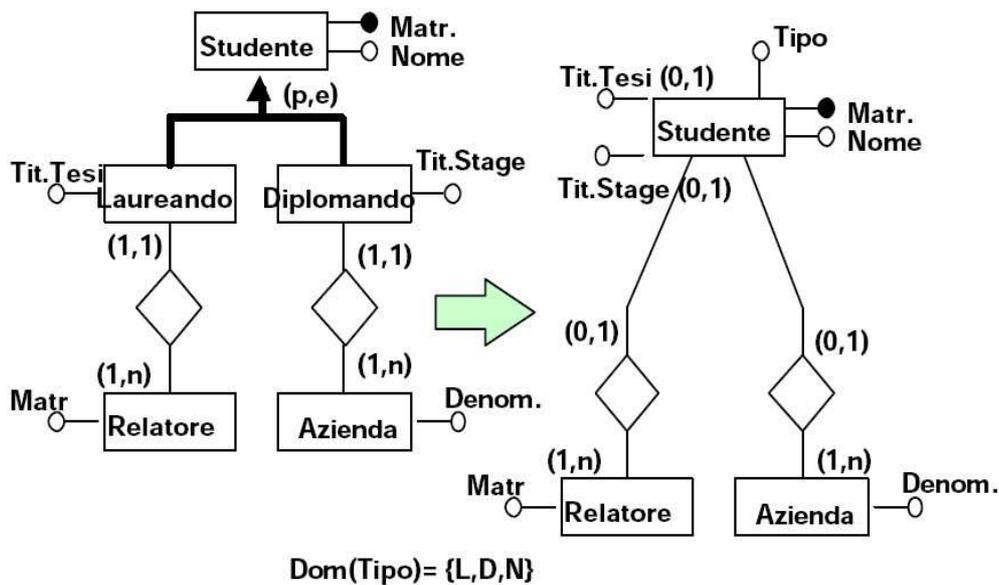


NB: padre verso i figli cardinalità (0, 1), mentre i figli verso il padre hanno cardinalità (1, 1)

b) *Collasso verso l'alto*

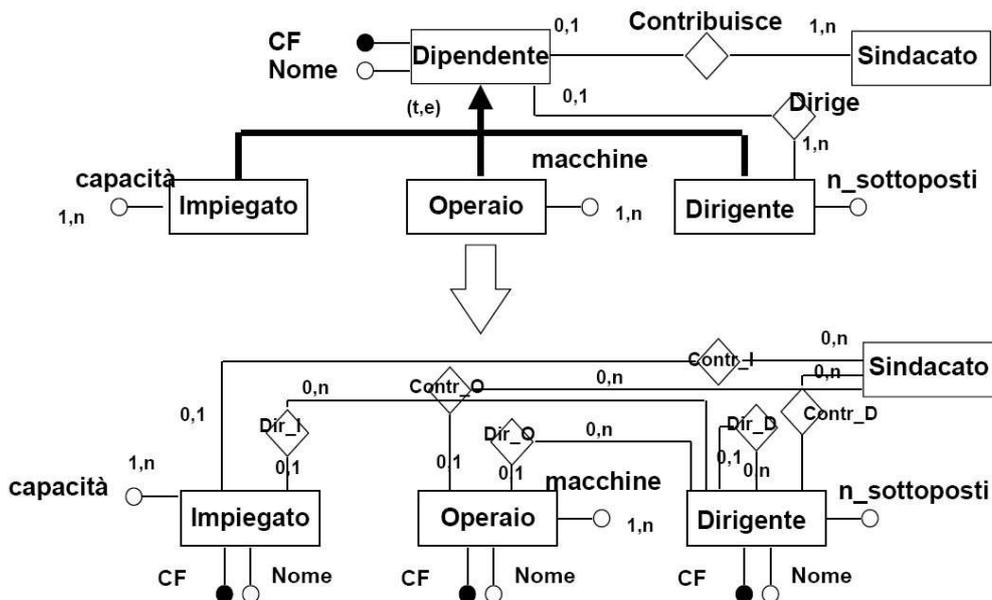
Riunisce tutte le entità figlie nell'entità padre, compare Sel che è un attributo selettore che specifica se una singola istanza di entità appartiene ad una delle N sottoentità.

- copertura (t, e) dove Sel ha N valori, quante sono le sottoentità;
- coperture (p, e) dove Sel ha N + 1 valori, il valore in più serve per istanze che non appartengono a nessuna sottoentità;
- copertura sovrapposta (o) dove occorrono tanti selettori Sel_i quante sono le sottoentità, ciascuno a valore booleano;
- gli attributi obbligatori per le entità figlie diventano opzionali per l'entità padre (0, 1) se sono 0 il loro valore è null;
- le eventuali associazioni connesse alle sottoentità si trasportano su E, le eventuali cardinalità minime diventano 0;



c) *Collasso verso il basso*

Elimina l'entità padre trasferendone attributi e associazioni su tutte le entità figlie. Tale collasso si può effettuare solamente se la copertura è totale, poiché non si saprebbe dove mettere le istanze di E che non sono istanze né di E₁ né di E₂. Se la copertura non è esclusiva, introduce ridondanza.



Ulteriori Trasformazioni

Partizionamento di entità: un'entità può essere partizionata al fine di meglio rispondere alle operazioni previste.

Accorpamenti di entità: due entità partecipanti ad un'associazione uno a uno, possono essere accorpate in un'unica entità contenente gli attributi di entrambi. Per le associazioni uno a molti e molti a molti, gli accorpamenti di entità generano ridondanza.

Trasformazione di attributi composti:

- eliminare l'attributo composto e considerare i suoi componenti come attributi semplici;
- eliminare i componenti e considerare l'attributo come semplice;

Una componente di identificazione esterna di una entità E_2 da una entità E_1 tramite un'associazione R , comporta il trasporto dell'identificatore di E_1 su E_2 .

Scelta della chiave primaria: è necessario che tra i diversi identificatori di una entità sia disegnata la chiave primaria. Criteri di scelta euristici:

- scegliere la chiave che è usata più frequentemente per accedere direttamente alle istanze dell'entità;
- preferire chiavi semplici a chiavi composte;
- eventualmente introdurre una chiave surrogata;
- tutti gli identificatori diversi dalla chiave primaria, sono denominati chiavi alternative;

Traduzione di entità e associazioni

Ogni entità è tradotta con una relazione con gli stessi attributi, la chiave primaria è quella dell'entità stessa, ogni altro identificatore dell'entità è chiave alternativa della relazione.

Ogni associazione avente degli attributi, è tradotta con una relazione con gli stessi attributi, cui si aggiungono le chiavi primarie di tutte le entità che essa collega, le chiavi primarie delle entità collegate, sono chiavi straniere riferite alle corrispondenti entità.

NB: la traduzione standard è sempre possibile ed è praticamente l'unica possibilità per le associazioni in cui tutte le entità partecipano con molteplicità maggiore di 1.

Associazione binaria uno a uno tradotta con una relazione:

- se l'associazione è obbligatoria per entrambe le entità, la chiave primaria può essere indifferentemente K_1 o K_2 .
- se l'associazione è parziale per un'entità E_1 e obbligatoria per l'altra E_2 , la chiave deve essere K_1 , ossia la chiave primaria è quella dell'entità che partecipa parzialmente.
- se l'associazione è parziale per entrambe, occorre che entrambe le entità abbiano lo stesso identificatore, la chiave può essere indifferentemente K_1 o K_2 , l'altra sarà FK.

Associazione binaria uno a uno tradotta con due relazioni:

- l'associazione si può compattare in una delle entità che includerà gli attributi dell'associazione e la chiave primaria dell'altra associazione come FK.
- per evitare i valori nulli, è preferibile compattare l'associazione in un'entità che partecipa obbligatoriamente.

Associazione binaria uno a molti tradotta con due relazioni:

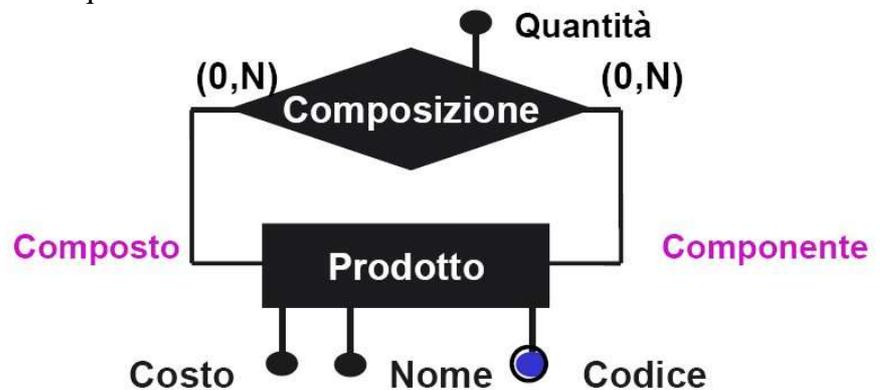
- l'associazione si può compattare nell'entità che partecipa con molteplicità unitaria, che includerà gli attributi dell'associazione e la chiave primaria dell'altra associazione come FK.

Associazione binaria uno a molti tradotta con 3 relazioni:

- se la partecipazione di un'entità è parziale, per evitare valori nulli, si può optare per la traduzione standard con tre relazioni.

Associazioni unarie:

- anello molti a molti: è tradotto con 2 relazioni, una per l'entità e una per l'associazione, la chiave della relazione che modella l'associazione è composta da 2 attributi, i cui nomi riflettono il diverso ruolo dell'entità. Ognuno di questi 2 attributi è anche FK.



Prodotto(Codice, Nome, Costo)

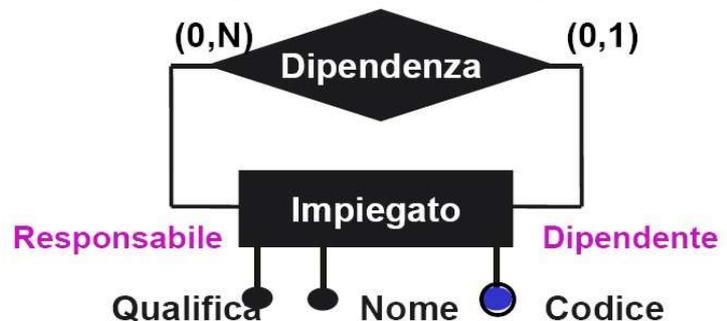
Composizione(Composto, Componente, Quantità)

FK: Composto REFERENCES Prodotto

FK: Componente REFERENCES Prodotto

- anello uno a molti: oltre che con due relazioni è traducibile con una sola relazione che contiene due volte l'attributo identificatore, una volta come chiave primaria e una volta come chiave esterna con un nome che riflette il ruolo dell'entità (Responsabile*=Codice_Responsabile)

■



1 relazione:

Impiegato(Codice, Nome, Qualifica, Responsabile*)

FK: Responsabile REFERENCES Impiegato

Associazione n-aria: segue la traduzione standard. Se ci sono vincoli particolari bisogna valutarli