

Basi di Dati

UniShare

Davide Cozzi
@dlcgold

Gabriele De Rosa
@derogab

Federica Di Lauro
@f_dila

Indice

1	Introduzione	2
2	Introduzione al Corso	3
3	Progettazione Concettuale:ER	11
4	Modello Relazionale	27
4.1	Vincoli d'integrità	29
4.1.1	In pratica	34
4.1.2	Relazione Matematica	35
4.1.3	Relazioni nel Modello Relazionale	36
5	SQL	43
5.0.1	Interrogazioni in SQL	50

Capitolo 1

Introduzione

Questi appunti sono presi a ldurante le esercitazioni in laboratorio. Per quanto sia stata fatta una revisione è altamente probabile (praticamente certo) che possano contenere errori, sia di stampa che di vero e proprio contenuto. Per eventuali proposte di correzione effettuare una pull request. Link: <https://github.com/dlcgold/Appunti>.

Grazie mille e buono studio!

Capitolo 2

Introduzione al Corso

Il corso di Basi di Dati affronta gli aspetti e i metodi per lo sviluppo di un database in maniera efficiente, aspetto fondamentale per un informatico e per lo sviluppo ottimale di software.

Il corso si divide in 7 parti:

1. introduzione generale
2. metodologie e modelli per il progetto delle basi di dati
3. progettazione concettuale
4. modello razionale
5. progettazione logica
6. linguaggio SQL
7. algebra relazionale

Le **informazioni** fanno parte delle risorse di un'azienda, soprattutto negli ultimi anni di clima globale, per cui la gestione efficiente ed ottimale dei dati è fondamentale, come ad esempio Facebook ed Amazon fanno ampio uso dei nostri dati, sia a scopi pubblicitari sia a scopi di marketing.

Un sistema informativo è una componente di un'organizzazione che gestisce le informazioni d'interesse, non per forza attraverso un'automatizzazione e/o supporto di un calcolatore, infatti sin dall'antichità le banche tenevano traccia dei depositi tramite un archivio cartaceo.

Una porzione automatizzata del sistema informativo si chiama sistema informatico, che si divide in:

- acquisizione e memorizzazione

- aggiornamento
- interrogazione
- elaborazione

Le informazioni vengono gestite in vari modi, attraverso il linguaggio naturale, graficamente con schemi e/o numeri, con il tempo si è arrivati a codifiche standard per quasi tutte le tipologie di informazioni, e sono rappresentate nei sistemi informatici dai *dati*, la cui differenza è che i dati sono valori senza alcun valore mentre le informazioni stabilisce un'interpretazione attribuendo una semantica ai valori.

I dati sono una risorsa strategica in quanto sono stabili nel tempo, infatti solitamente i dati sono immutati durante una migrazione tra un sistema e un altro, per questo lo sviluppo progettazione di un database rimane stabile in teoria, senza notevoli cambiamenti durante la durata di un sistema informativo.

Un **Data Base** è una collezione di dati usati per rappresentare le informazioni di interesse di un sistema informativo, definite solo una volta a cui un insieme di applicazioni ed utenti può accedere ad essi, mentre un **DBMS** è un software per la gestione di un database.

I dati presenti in un database sono molti, indipendenti dal programma in cui vengono utilizzati e cui si cerca di evitare la ridondanza, per garantire la consistenza delle informazioni.

Per la creazione di un database, al fine di garantire privacy, affidabilità, efficienza ed efficacia, sono presenti le seguenti tre fasi:

1. definizione
2. creazione e popolazione
3. manipolazione

Un'organizzazione è divisa in vari settori e ogni settore ha un suo sottosistema informativo, non necessariamente disgiunto, e i database solitamente sono condivisi, al fine di ridurre la ridondanza delle informazioni, per cui sono presenti meccanismi di autorizzazione e di controllo della concorrenza.

Un database deve essere conservato a lungo termine e si ha una gestione delle **transazioni**: insieme di operazioni da considerare indivisibile, atomico, corretto anche in presenza di concorrenza e con effetti definitivi.

La sequenza di operazioni nel database deve essere eseguita nella sua interezza, per cui l'effetto di transazioni concorrenti deve essere coerente, infatti la conclusione positiva di una transazione corrisponde ad un impegno,

commit, a mantenere traccia del risultato, anche in presenza di guasti e di esecuzioni concorrente.

Come tutti i software, i DBMS devono essere efficienti, utilizzando al meglio memoria e tempo, efficaci e produttivi.

Si hanno delle caratteristiche nell'approccio alla base dati:

- natura autodescrittiva di un sistema di basi di dati: il sistema di basi di dati memorizza i dati con anche una descrizione completa della sua struttura (metadati), per consentire ai DBMS di lavorare con qualsiasi applicazione.
- separazione tra programmi e dati, infatti è possibile cambiare la struttura dati senza cambiare i programmi.
- astrazione dei dati: si usa un modello dati per nascondere dettagli e presentare all'utente una visione concettuale del database.
- supporto di viste multiple dei dati, per cui ogni utente può usare una vista (view) differente del database, contenente solo i dati di interesse per quell'utente.
- condivisione dei dati e gestione delle transazioni con utenti multipli.

I DBMS estendono le funzionalità dei file system, fornendo più servizi ed in maniera integrata.

In ogni base di dati si ha:

- lo **schema**, sostanzialmente invariante nel tempo, che ne descrive la struttura, l'aspetto intensionale.
- l'**istanza**, i valori attuali, che possono cambiare anche molto rapidamente, l'aspetto estensionale "concreto".

Per lo sviluppo delle basi di dati si hanno due tipi di modelli, ambedue importanti:

- **modelli logici**, adottati nei DBMS esistenti per l'organizzazione dei dati, utilizzati dai programmi e sono indipendenti dalle strutture fisiche
- **modelli concettuali** che permettono di rappresentare i dati in modo indipendente da ogni sistema, con il fine di descrivere i concetti del mondo reali; sono usati nelle fasi preliminari di progettazione e il più diffuso è il modello **Entity-Relationship**.

Un database è organizzato solitamente attraverso i tre schemi dell'architettura ANSI/SPARC:

- schema logico: descrizione dell'intera base di dati nel modello logico “principale” del DBMS, ossia si definisce la struttura concettuale del database, senza considerare l'implementazione fisica nel DBMS.
- schema fisico: rappresentazione dello schema logico per mezzo di strutture fisiche di memorizzazione
- schema esterno: descrizione di parte della base di dati in un modello logico (“viste” parziali, derivate, anche in modelli diversi)

L'accesso ai dati avviene solo mediante il livello esterno, il quale a volte coincide con quello logico, e si hanno 2 forme di indipendenza: quella fisica, in cui è possibile interagire con il DBMS senza conoscere la struttura fisica e quella logica, in cui si può accedere al livello esterno senza interagire con lo schema logico.

Per la definizione dei database ci sono quattro tipologie di linguaggi:

- **DLL**(Data Manipulation Languages) linguaggio per definire i dati
- **DML**(Data Manipulation Languages) linguaggio per la manipolazione dei dati
- **DCL**(Data Control Languages) linguaggio per il controllo degli accessi al database
- **DQL**(Data Query Languages) linguaggio per effettuare delle interrogazioni al database

Noi vedremo SQL per definire i database, linguaggio basato sull'algebra relazionale che implementa tutti e 4 le tipologie di linguaggi per la gestione di un database.

Si hanno due tipi di utenti:

1. **utenti finali (terminalisti)**: eseguono applicazioni predefinite (transazioni)
2. **utenti casuali**: eseguono operazioni non previste a priori, usando linguaggi interattivi

inoltre si hanno:

- progettisti e realizzatori di DBMS

- progettisti della base di dati e amministratori della base di dati (DBA), persona o gruppo di persone responsabile del controllo centralizzato del database, in tutti i suoi aspetti.
- progettisti e programmatori di applicazioni



Come visto anche nel corso di analisi e progettazione del software e nella figura X, durante lo sviluppo di un sistema informatico si hanno le seguente fasi:

- **studio di fattibilità**, definizione di costi, priorità e competenze per un progetto
- **raccolta e analisi dei requisiti**, ovvero lo studio delle proprietà del sistema
- **progettazione** di dati e funzioni
- **implementazione**
- **validazione e collaudo**, che comprendono anche test da parte del cliente
- **funzionamento**, ovvero lo stadio finale dove il sistema diventa effettivamente operativo

In questo corso ci occuperemo soltanto delle fasi di progettazione ed implementazione di base di dati, lasciando l'analisi delle altre fasi al corso ed ai libri di Ingegneria del Software.

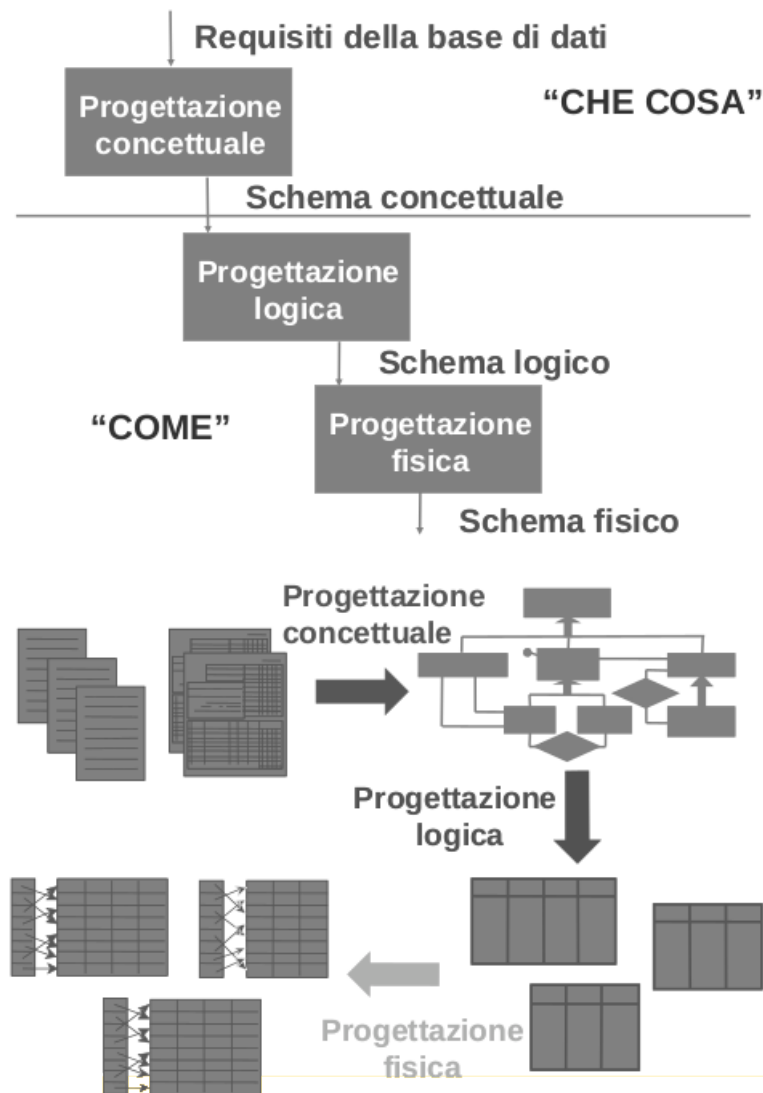
In questo processo a rimanere stabili sono prettamente i **dati** infatti prima si progetta la base dati, con una **metodologia di progetto**, e poi il software.

Ciclo di vita (modello a spirale)



Una **metodologia** è un'articolazione in fasi di guida ad un'attività di progettazione, in caso di una base dati la metodologia, con il fine di separare il cosa rappresentare dal come, è la seguente:

- suddivida la progettazione in fasi indipendenti
- fornisca strategie e criteri di scelta in caso di alternative
- fornisca modelli di riferimento (i linguaggi)
- garantisca generalità rispetto al problema
- garantisca qualità e facilità d'uso



Come si può notare nella figura X1, la metodologia corrente di modellazione di un database prevede la definizione e l'esecuzione delle seguenti fasi:

- la **progettazione concettuale** consiste nel tradurre i requisiti del sistema informatico in una descrizione formale, integrata e indipendente dalle scelte implementative (DBMS, SW e HW)
- la **progettazione concettuale** consiste nella traduzione dello schema concettuale nel modello dei dati scelto per la modellazione, ottenendo uno schema logico, espresso nel DDL del DBMS. In questa fase si considerano anche aspetti legati ai vincoli ed all'efficienza. Si hanno due sotto-fasi:

- ristrutturazione dello schema concettuale
- traduzione verso il modello logico
- la **progettazione fisica** completa lo schema logico ottenuto con le specifiche proprie del DBMS scelto. Il risultato è lo schema fisico che descrive le strutture di memorizzazione ed accesso ai dati

Incominciamo nel prossimo capitolo a considerare la progettazione concettuale, per poi analizzare nei successivi capitoli in dettaglio anche la fase di progettazione concettuale e il linguaggio SQL.

Capitolo 3

Progettazione Concettuale:ER

Come abbiamo già introdotto nel precedente capitolo, affrontiamo ora la progettazione concettuale di una base dati attraverso il modello **ER**(Entity Relationships), modello concettuale che fornisce una serie di strutture atte a descrivere in maniera semplice e facile la realtà di interesse da modellare.

Si hanno dei vantaggi con la progettazione concettuale, prevale infatti l'aspetto intensionale indipendente dalla tecnologia ed è una rappresentazione grafica ed è utile per la documentazione, in quanto facilmente comprensibile anche da persone poco avvezze alla tecnologia e ai database.

In uno schema ER si hanno i seguenti costrutti, come si nota nella figura Y, la cui rappresentazione effettiva varia in quanto vi sono più versioni di ER:

- **entità**: classe di oggetti con proprietà comune ed esistenza "autonoma", della quale si vogliono specificare fatti specifici; ogni entità ha un nome univoco, espressivo e al singolare.

- **relazione**: rappresentano legami logici, significativi per la realtà da modellare, tra due o più entità.

Un'occorrenza di relazione è un n-upla costituita da occorrenze di entità, una per ciascuna delle entità coinvolte, e ogni relazione ha un nome univoco, in cui è preferibile assegnare un sostantivo per evitare di stabilire un verso alla relazione.

Essendo una relazione matematica tra le entità coinvolte non è possibile avere delle n-uple identiche, con conseguenze per la realtà da rappresentare, per esempio non è possibile attraverso una relazione il fatto che è possibile ripetere un esame, obbligando a rappresentare l'esame come entità e non più come relazione.

- **attributo semplice:** associa ad ogni istanza di entità o associazione un valore, definito su un dominio di valori, specificato nella documentazione associata, con il fine di descrivere le proprietà elementari di entità e/o relazioni disegnate per rappresentare la realtà d'interesse.
- **attributo composto:** raggruppamento di attributi di una medesima entità/relazione con affinità di significato e/o uso come ad esempio possiamo raggruppare gli attributi Via, Numero Civico e Cap dall'entità persona per formare l'attributo composto Indirizzo.
- **cardinalità delle relazioni:** vengono specificate per ogni relazione e descrivono il numero minimo e massimo di occorrenze di relazione, a cui una occorrenza dell'entità può partecipare alla relazione, ossia quante volte un'occorrenza di un'entità può essere legata ad occorrenze delle altre entità coinvolte.
È possibile assegnare un qualunque intero non negativo, con l'unico vincolo che la cardinalità minima sia minore o uguale alla cardinalità massima e di solito si usano i valori $0, 1eN$, indicanti zero, una o molte occorrenze, senza preoccuparsi in caso di N del numero effettivo di occorrenze.
- **cardinalità di un attributo:** descrivono il numero minimo e massimo di valori dell'attributo associati all'entità e/o relazione, con la cardinalità $(1, 1)$ stabilita come default, che può essere vista come funzione che associa ad ogni occorrenza di entità un solo valore dell'attributo; si hanno le stesse consuetudini delle cardinalità delle relazioni.
- **identificatore interno:** permette di identificare in maniera univoca un'entità ed un identificatore è interno in caso sia uno o più attributi di un'entità, tutti con cardinalità $(1, 1)$.
- **identificatore esterno:** un identificatore è esterno, in caso un'entità E viene identificata da un'attributo di un'entità F , cui esiste una relazione uno a uno tra l'entità E e F .
È possibile, anche se molto raro, avere la definizione dell'identificatore usando entità di entità, ossia l'identificatore dell'entità E viene definito nell'entità G , cui esiste una relazione con l'entità F

che a sua volta ha una relazione con l'entità E , ma si può capire da quanto è contorto il ragionamento qual'è la sua percentuale d'utilizzo nella modellazione dello schema ER.

- **generalizzazione:** rappresentano legami logici tra un'entità E , detta padre, e una serie di entità E_1, E_2, \dots, E_n , dette figlie, di cui l'entità E rappresenta un caso generale della serie di entità figlie.

Tra le entità coinvolte in una generalizzazione valgono le seguenti proprietà:

- ogni occorrenza di un'entità figlia è anche un'occorrenza dell'entità genitore.
- ogni proprietà dell'entità genitore è anche una proprietà delle entità figlie, quindi nello schema ER non devono essere rappresentate, e ciò prende il nome di ereditarietà.




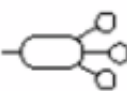



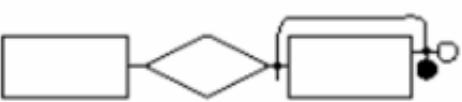
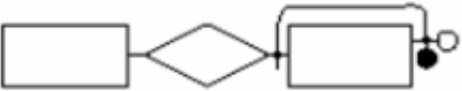


Le generalizzazioni possono essere classificate sulla base di due proprietà ortogonali:

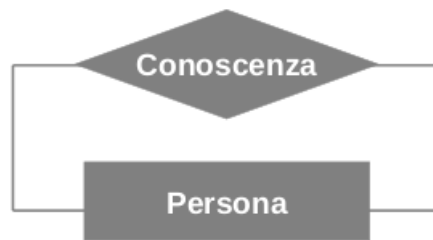
- una generalizzazione è totale se ogni occorrenza del genitore è un'occorrenza di almeno uno dei figli, altrimenti è parziale.
 - una generalizzazione è esclusiva se ogni occorrenza del genitore è al più un'occorrenza di una delle entità figlie, altrimenti è sovrapposta.
- **sottoinsieme:** generalizzazione con soltanto un'entità figlia, di cui solitamente rappresenta una parte dell'entità genitore come ad esempio gli studenti sono un sottoinsieme delle persone.

Un'occorrenza, o istanza, di un'entità, è un oggetto della classe che l'entità rappresenta, ma noi rappresentiamo le entità dello schema concettuale non le singole istanze, in quanto esse sono variabili nel tempo a differenza della struttura, concetto fondamentale per definire un modello funzionale.

Nessuno impone un tipo per un certo attributo, possono essere anche complessi di cui però non ci interessano le informazioni che lo rappresentano, come ad esempio una foto può essere un attributo, ma se ho bisogno dell'autore della foto, non sarà più un attributo ma un'altra entità.

Un'**istanza di associazione** è una combinazione o aggregazione di istanze di entità che prendono parte all'associazione (per esempio "prof. Schettini")

Construct	Graphical representation
Entity	
Relationship	
Simple attribute	
Composite attribute	
Cardinality of a	
Cardinality of an attribute	
Internal identifier	 
External identifier	
Generalization	
Subset	

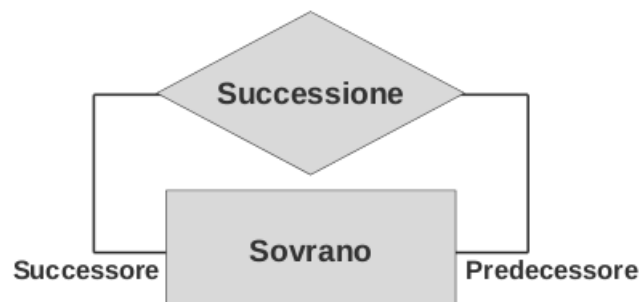


è istanza di associazione per l'entità docente).

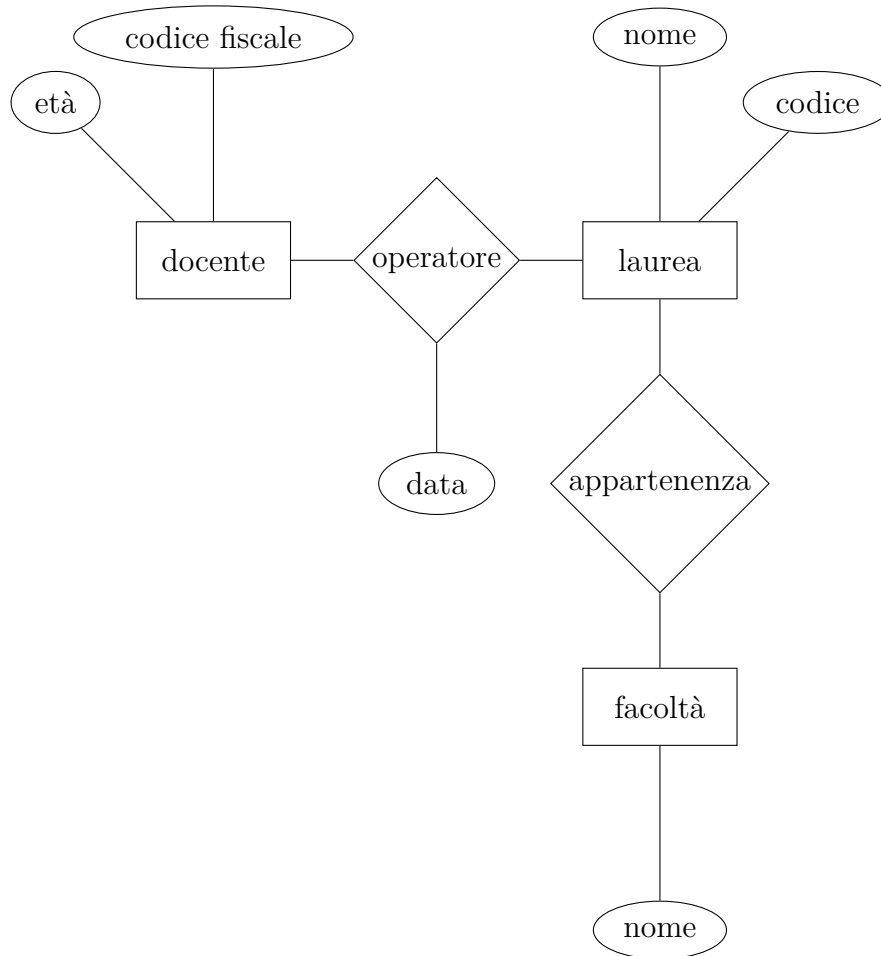
Le relazioni possono avere attributi, con valori specificati in un certo dominio, che modella una proprietà del legame tra tutte le entità rappresentato dalla relazione.

Una associazione può coinvolgere “due o più volte” la stessa entità e ciò si chiama *associazione ricorsiva o ad anello*, come si nota nella figura XXXX, ed è molto utile per definire le relazioni subordinazione e/o comunicazioni tra istanze di una stessa entità, come nel caso della relazione sovrano, indicante la lista dei sovrani con i predecessori e successori; in queste associazioni è necessario aggiungere la specifica dei **ruoli**, come nell'esempio SDERR, per specificare in maniera chiara e non ambigua quale è il verso della relazione.

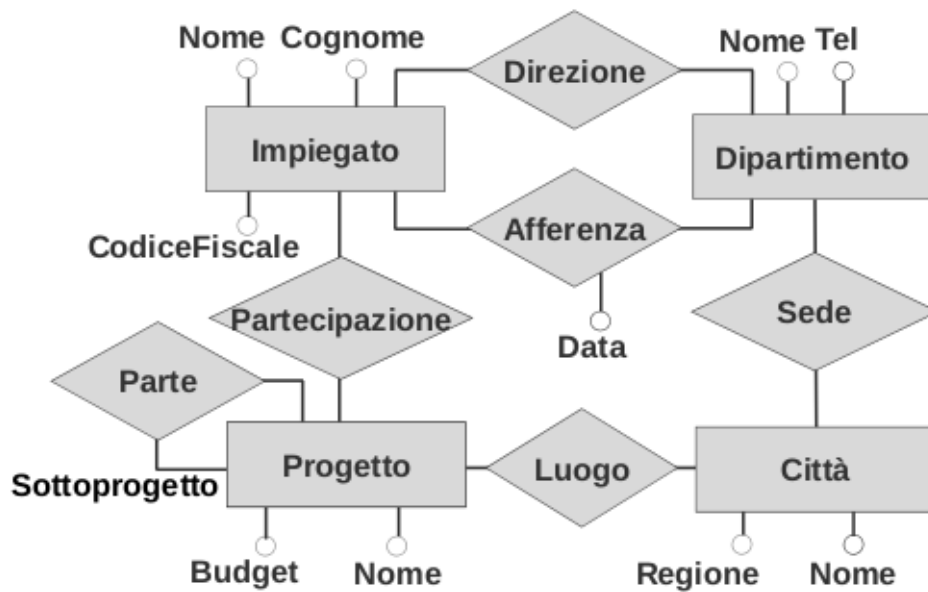
Un'associazione ad anello, ovviamente essendo una relazione, può avere delle proprietà come ad esempio la relazione della figura SJDFHJF è irreflessiva, intransitiva e asimetrica.



Descrivere lo schema concettuale della seguente realtà: I docenti hanno un codice fiscale ed una età. I docenti operano nei corsi di laurea (si dice che afferiscono ai corsi di laurea). Interessa la data di afferenza dei docenti ai corsi di laurea. I corsi di laurea hanno un codice ed un nome, ed appartengono alle facoltà. Ogni facoltà ha un nome



Esempio 1. Descrivere lo schema concettuale della seguente realtà: Degli impiegati interessa il codice fiscale, il nome, il cognome, i dipartimenti ai quali afferiscono (con la data di afferenza), ed i progetti ai quali partecipano. Dei progetti interessa il nome, il budget, e la città in cui hanno luogo le corrispondenti attività. Alcuni progetti sono parti di altri progetti, e sono detti loro sottoprogetti. Dei dipartimenti interessa il nome, il numero di telefono, gli impiegati che li dirigono, e la città dove è localizzata la sede. Delle città interessa il nome e la regione:



Nella scelta del costrutto ideale, per rappresentare in maniera fedele e corretta la realtà, si usano le seguenti "regole":

- **entità:** si usa un'entità in caso vengano rispettate le seguenti proprietà: le sue istanze sono significative indipendentemente dalle altre istanze, se si ha o si può avere delle proprietà indipendenti dagli altri concetti oppure se il concetto da rappresentare è importante per la realtà da modellare.
- **attributo:** si usa un'attributo in caso succedono i seguenti avvenimenti: non ha senso considerare una sua istanza a se stante ad altre istanze, le istanze non sono concettualmente significative oppure serve soltanto rappresentare una proprietà locale di un altro concetto.
- **relazione:** si dovrebbe usare una relazione in caso non ha senso pensare alla partecipazione delle sue istanze ad altre relazioni e

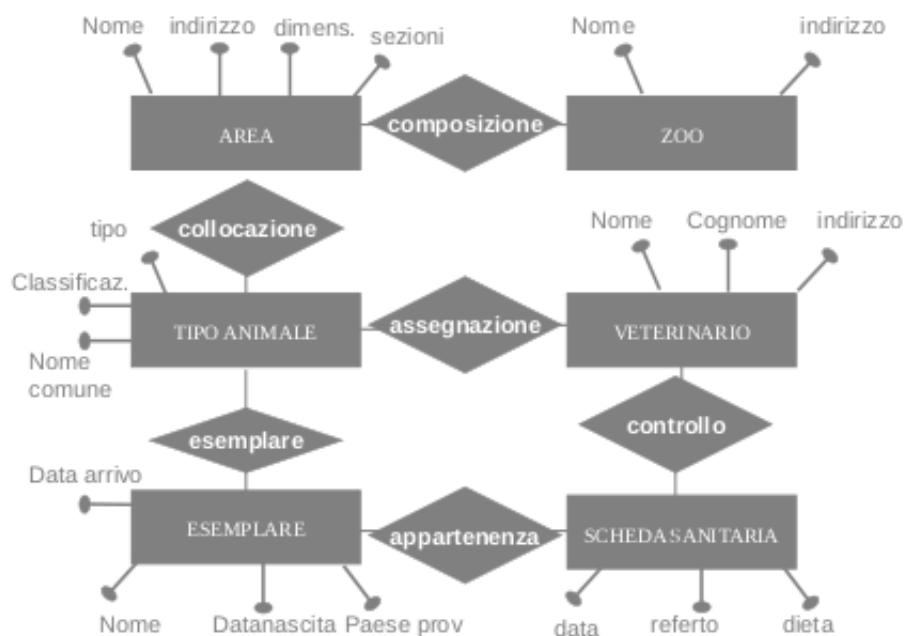
le sue istanze non sono significative indipendentemente da altre istanze, ossia una sua istanza ha senso se messa in relazione con altre istanze, provenienti da altri costrutti.

•Calcio: un giocatore gioca in una squadra

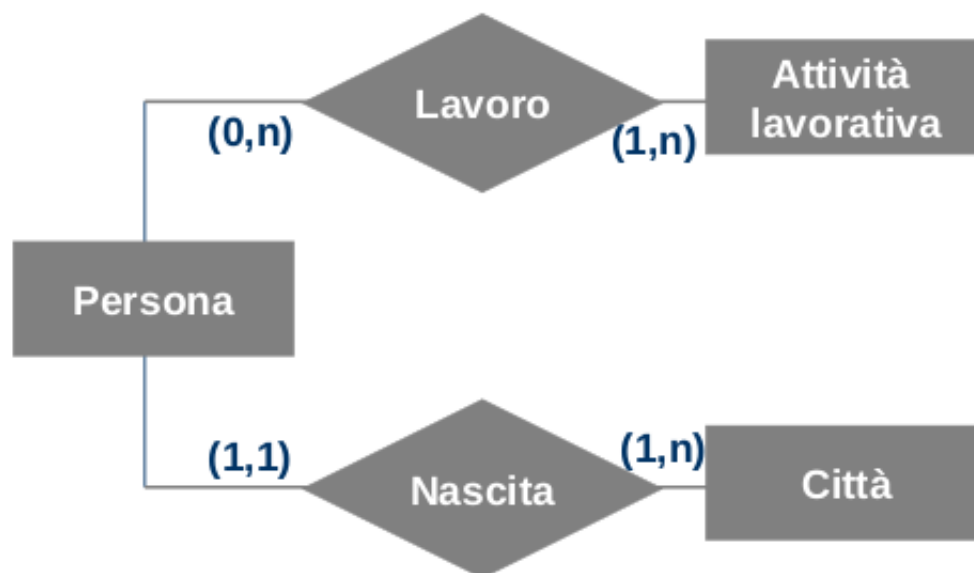


Esempio 2. • ogni zoo è diviso in aree diverse a seconda che si tratti di rettili, pesci, uccelli, scimmie, grandi mammiferi, ... Ogni area è dotata di: nome, indirizzo, dimensione, numero di sezioni.

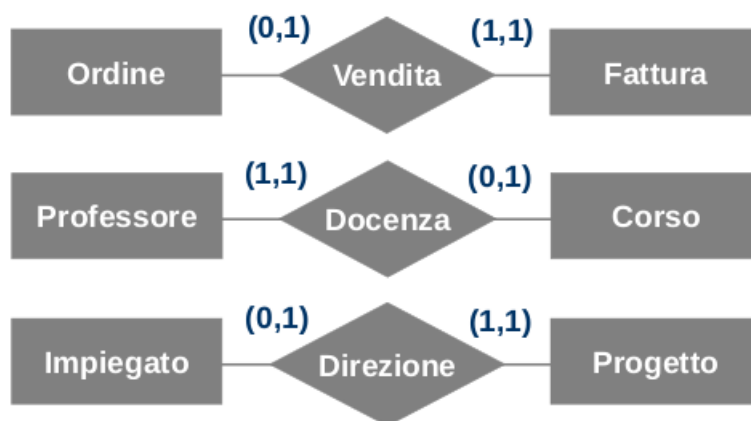
- per ogni tipo di animale ci sono informazioni che riguardano: classificazione zoologica, nome comune (giraffa, elefante, serpente, tartaruga, ...), habitat, alimentazione, ... Per ogni tipo di animale c'è un diverso veterinario specialista, dipendente dello zoo.
- ogni tipo di animale è rappresentato da esemplari e relativi dati anagrafici: nome proprio (giraffa Enrico, giraffa Giulia, ...), data di nascita, Paese di provenienza, data di arrivo allo zoo, ...
- ogni esemplare è dotato di più schede sanitarie contenenti ognuna: la data della visita, referto, dieta, nome del veterinario, ...



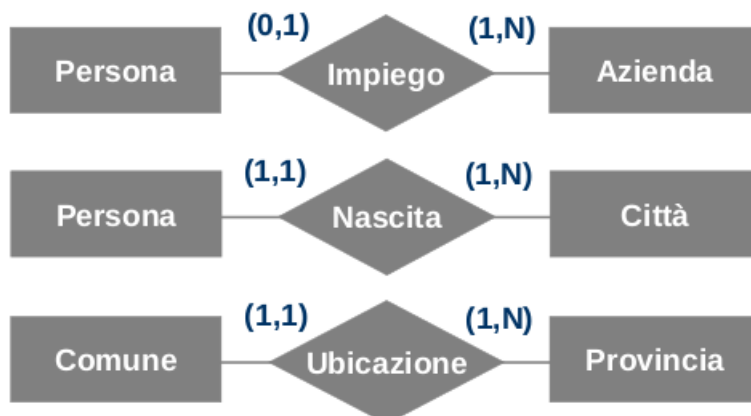
lo zoo ha degli attributi che vanno indicati anche se non richiesti dalla traccia per evitare ambiguità. Dato che ogni scheda è collegata ad un veterinario, di cui ho un'entità la collego a veterinario con una relazione e non ne faccio un attributo



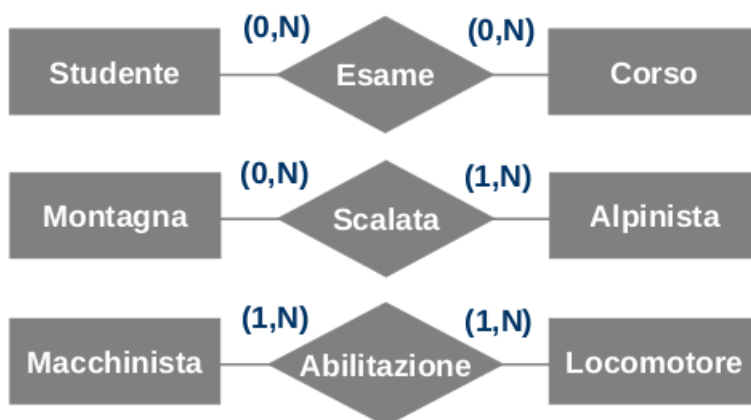
Relazioni “uno a uno”



Relazioni “uno a molti”



Relazioni “molti a molti”



vediamo un esempio:

Esempio 3. *si ha:*



- $\text{min-card}(\text{Automobile}, \text{Proprietario}) = 0$: esistono automobili non possedute da alcuna persona
- $\text{min-card}(\text{Persona}, \text{Proprietario}) = 0$: esistono persone che non posseggono alcuna automobile
- $\text{max-card}(\text{Persona}, \text{Proprietario}) = n$: ogni persona può essere proprietaria di un numero arbitrario di automobili
- $\text{max-card}(\text{Automobile}, \text{Proprietario}) = 1$: ogni automobile può avere al più un proprietario



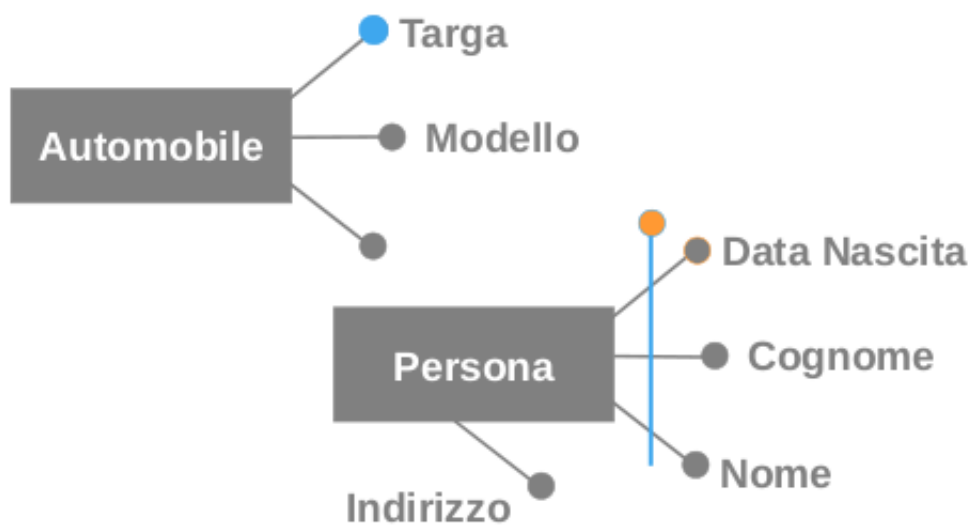
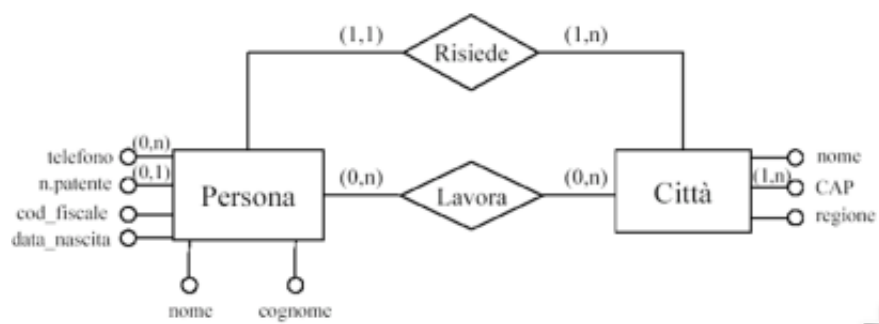
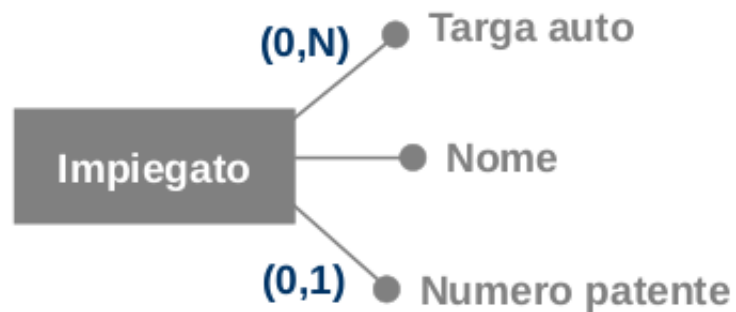
Istanza dello schema:

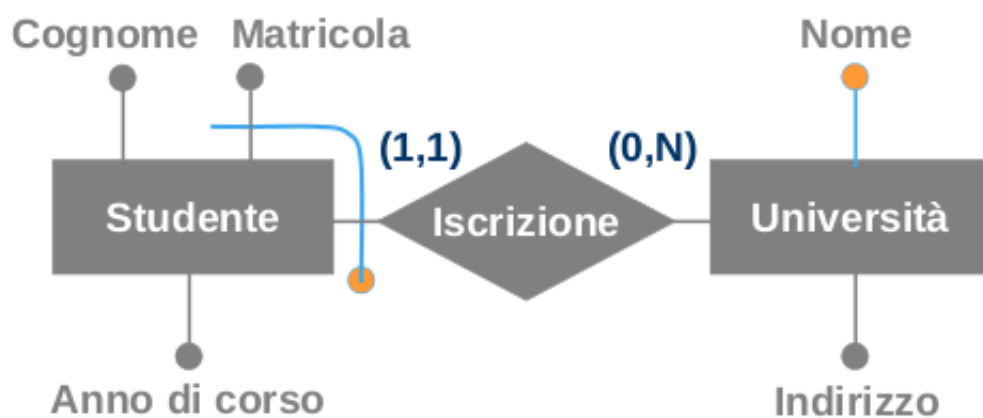
Istanze(Impiegato) = { a,b,c }

istanze(Progetto) = { x,y,v,w,z }

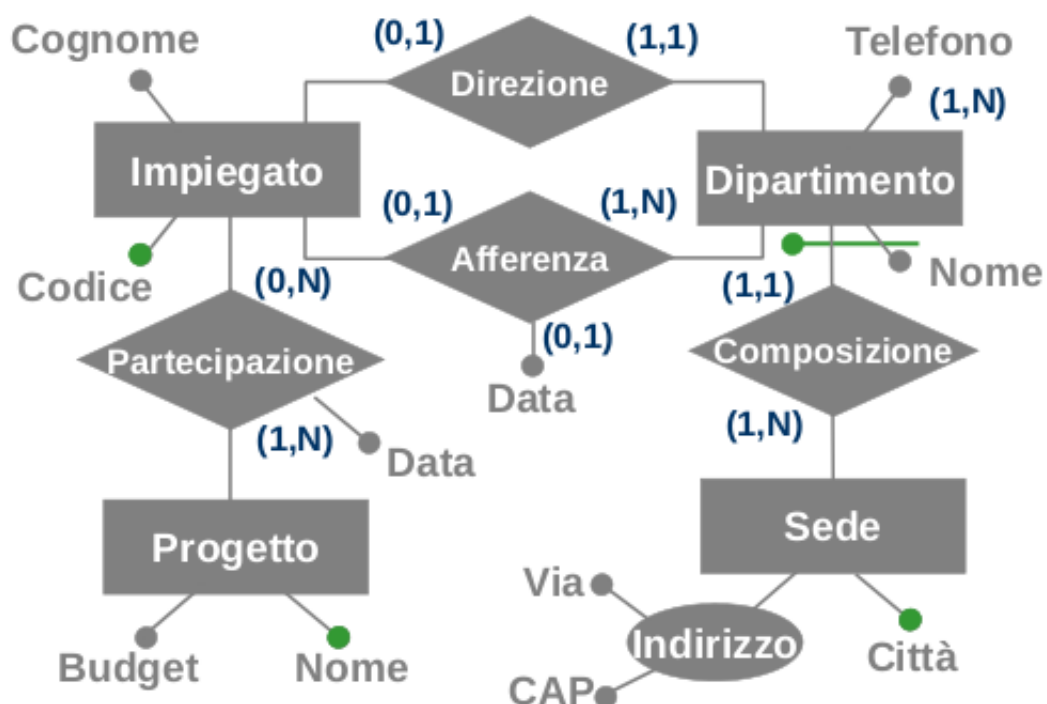
istanze(Assegnazione) = { (a,w), (b,v), (b,w), (c,y), (c,w), (c,z) }

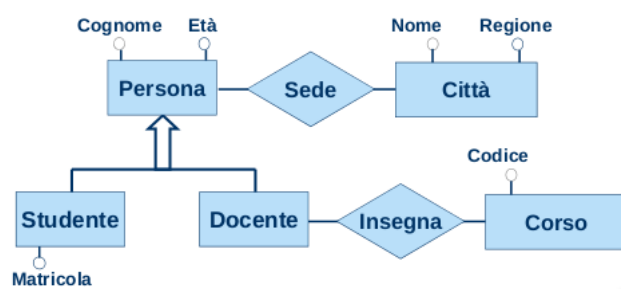
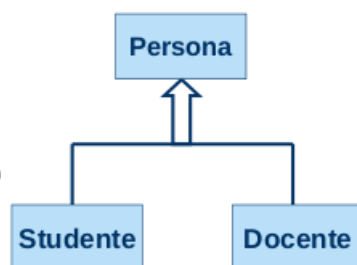
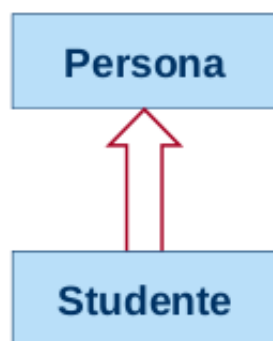
Ad ogni impiegato sono assegnati da 1 a 5 progetti e ogni progetto è assegnato ad al più 50 impiegati. a,b,c compaiono in almeno una istanza di Assegnazione. x non compare nelle istanze di Assegnazione. Infine ci sono progetti (ad esempio lanciati da poco tempo) che possono non essere assegnati a nessun impiegato

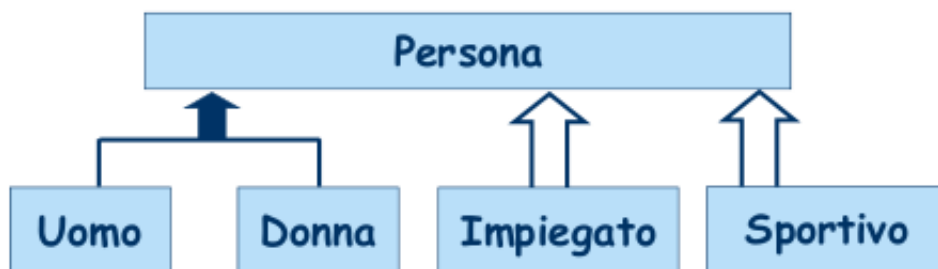
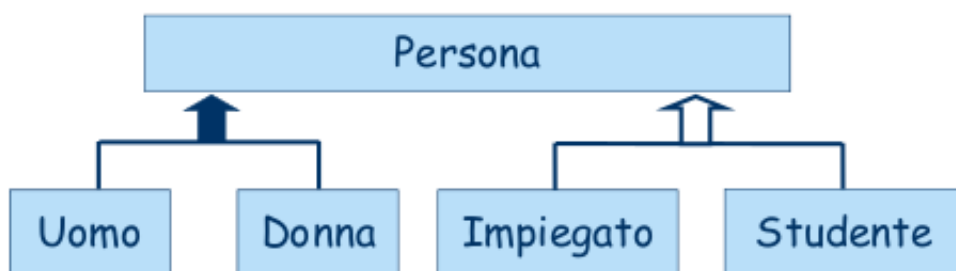
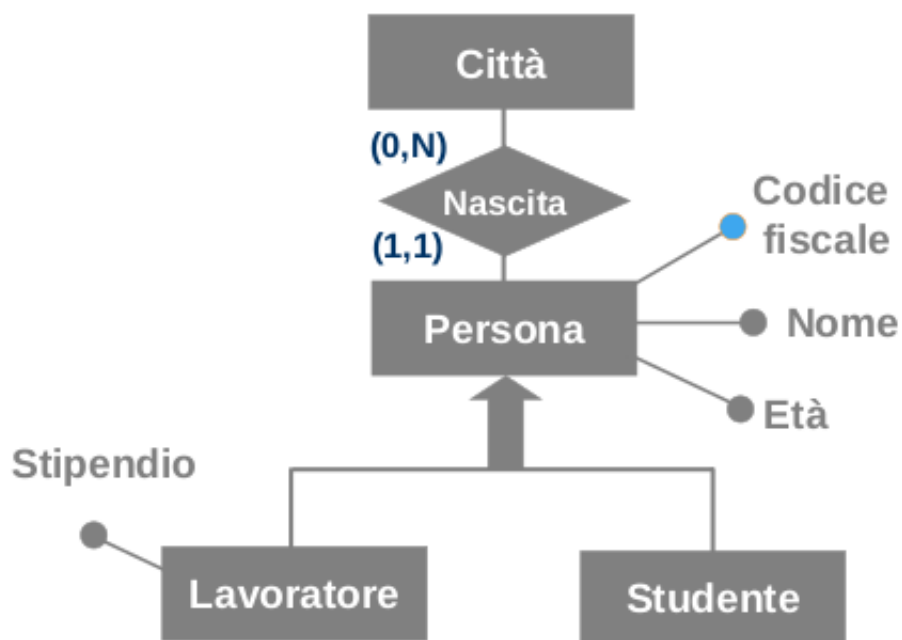


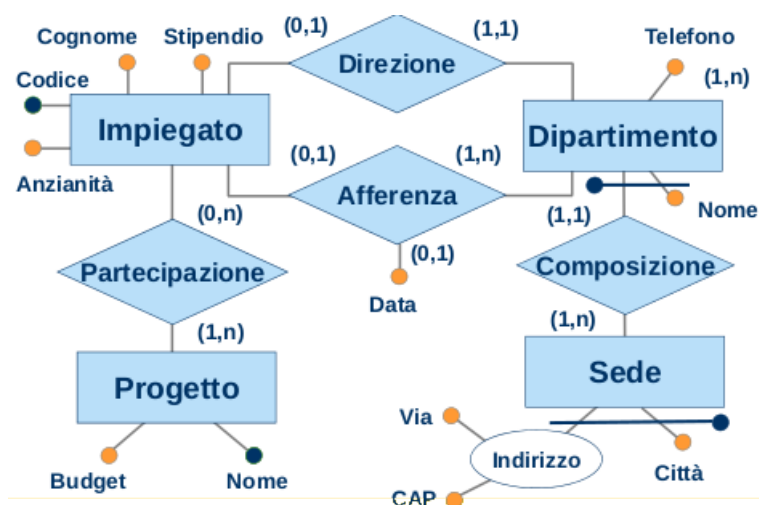


vediamo un esempio complesso, che è preso da un vecchio esempio:









Vincoli di integrità esterni
(1) Il direttore di un dipartimento deve afferire a tale dipartimento da almeno 5 anni
(2) Un impiegato non deve avere uno stipendio maggiore del direttore del dipartimento al quale afferisce
(3) Un dipartimento con sede a Roma deve essere diretto da un impiegato con più di dieci anni di anzianità
(4) Un impiegato non può partecipare ad un numero di progetti maggiore di due volte il numero di dipartimenti ai quali afferisce



Capitolo 4

Modello Relazionale

Dopo aver introdotto il modello concettuale, specifichiamo ed analizziamo ora il modello logico necessario per rappresentare in maniera efficace ed efficiente la base dati modellata.

Analizziamo il modello relazionale, il più diffuso dei modelli dei dati, in cui si utilizzano le tabelle e le relazioni per ottenere la rappresentazione voluta, introdotto negli anni '70 con l'obiettivo di rendere indipendenti i dati, ossia separare la modellazione “concettuale” dall'effettiva implementazione fisica sul DBMS.

Il modello relazionale, presenta una notevole differenza dai modelli precedenti, dato che i riferimenti fra dati in strutture diverse sono rappresentati per mezzo dei valori stessi, garantendo così l'indipendenza dei dati.

Il concetto di relazione, ovviamente proviene dalla teoria degli insiemi, rappresenta un sottoinsieme del prodotto cartesiano ma nel modello relazionale si utilizza una variante di esso, ossia ad ogni dominio, su cui è definito il prodotto cartesiano, viene associato un nome, chiamato *attributo*, al fine di superare la struttura posizionale della relazione matematica.

Questo aspetto di utilizzare un'attributo con una notazione non posizionale ci garantisce di poter accedere ai campi di una tupla mediante un nome e soprattutto ci evita di stabilire un ordine tra i diversi campi, ad esempio chi stabilisce che la matricola sia prima del nome e non viceversa.

Forniamo ora un pò di formalismo al fine di definire in maniera non ambigua il concetto di base di dati:

Definizione 1. *Si definisce dominio una funzione $dom : X \rightarrow D$ in cui si associa ogni attributo $A \in X$ un dominio $dom(A) \in D$ mentre si definisce tupla su un insieme di attributi X una funzione t che associa ogni elemento di X un elemento appartenente al dominio $dom(X)$.*

Definizione 2. Si definisce schema di relazione $R(X)$ formato da R indicante il nome della relazione mentre X indica un insieme di attributi, a cui ad ogni attributo viene associato un dominio, come abbiamo definito nella definizione precedente.

Si definisce invece istanza di relazione su uno schema $R(X)$ è un insieme r di tuple definite su X , a cui si può accedere ad un attributo della tupla attraverso r.a, con a nome di un attributo.

Definizione 3. Si definisce schema di base di dati un insieme $R = \{R_1(X_1), R_2(X_2), \dots, R_n(X_n)\}$ di schemi di relazione in cui risulta $R_i \neq R_j$ con $i \neq j$.

Si definisce inoltre istanza di base di dati su uno schema $R = \{R_1(X_1), \dots, R_n(X_n)\}$, un insieme di relazioni $r = \{r_1, r_2, \dots, r_n\}$ dove ogni r_i , con $1 \leq i \leq n$, è una relazione sullo schema $R_i(X_i)$.

Nel modello relazione i riferimenti tra i dati di relazioni diverse viene fatto attraverso i valori delle tuple, a differenza di altri modelli, come ad esempio il modello reticolare, in cui i riferimenti si effettuano attraverso i puntatori; il vantaggio di questo modo di gestire i riferimenti permette di ottenere l'indipendenza dei dati, evitando i riferimenti alla rappresentazione fisica ma ciò non significa che nel livello fisico non si possa usare i puntatori per gestire i riferimenti ma solo che si evita di renderlo visibile agli utenti e/o al livello logico.

Secondo le definizioni formali fatte degli schemi di relazioni e/o basi di dati, sono ammissibili relazioni su un solo attributo e ciò ha senso soprattutto in basi di dati con molte relazioni, in cui la relazione su un solo attributo contiene valori che appaiono come valori di un attributo di un'altra relazione.

La struttura del modello relazionale discussa ed analizzata nei precedenti paragrafi è indubbiamente molto semplice e potente ma essa impone un certo grado di rigidità, in quanto le informazioni devono essere rappresentate attraverso tuple appartenenti allo schema di relazione stesso ma ciò non sempre può avvenire.

Ad esempio in una relazione *Persona*(Cognome, Nome, Indirizzo, Telefono) non sempre il valore del telefono risulta definito per cui sarebbe scorretto usare un valore del dominio per rappresentare questa assenza di informazione in quanto esso genererebbe confusione ed ambiguità.

Per risolvere sto problema e rappresentare in maniera semplice e chiara la mancanza di valori si è deciso di estendere il concetto di relazione, permettendo di usare il valore speciale *null*, non facente parte il dominio degli attributi.

Questo valore nullo può rappresentare tre diverse tipologie di informazioni:

- indicare un valore sconosciuto, in cui si sa che il valore esiste ma non si conosce il suo valore.

- indicare un valore inesistente, in cui il valore dell'attributo non esiste.
- indicare l'assenza di informazione, in cui non si sa se il valore esiste e soprattutto che valore può rappresentare.

Nei sistemi di basi di dati relazionali non si accentrano sulla differenza tra le diverse tipologie ma suppone in maniera semplicista che si rappresenta sempre il caso di assenza di informazione.

Non sempre il valore nullo può essere sensato, infatti come si nota nella figura 2.13 l'assenza di valore di matricola e codice di corso crea notevoli problemi in quanto non permette di stabilire correlazioni fra tuple di relazioni diverse ed inoltre la presenza di molti valori nulli può generare dubbi sull'effettivo significato delle tuple, in quanto è impossibile riconoscere in maniera univoca le istanze della relazione.

Come vedremo nel prossimo paragrafo, sui vincoli d'integrità, si può indicare in quali attributi sono ammessi valori nulli e in quali no.

4.1 Vincoli d'integrità

Le strutture del modello relazione ci permettono di organizzare le informazioni d'interesse ma non è vero che tutti gli insiemi di tuple rappresentano dati corretti per l'applicazione.

Come ad esempio non è ammissibile che il voto assuma il valore 36, in quanto nel sistema universitario italiano i voti vanno da 0 a 30, poi inoltre compaiono nella relazione esami delle matricole non presenti nella relazione studente.

In un database è opportuno evitare le situazioni appena descritte, per cui a questo scopo sono stati introdotti i *vincoli di integrità*, ossia delle proprietà da soddisfare da tutte le istanze delle relazioni altrimenti lo schema di database risulta non ammissibile.

È possibile classificare i vincoli a seconda degli elementi coinvolti:

- **vincolo intrarelazionale:** sono i vincoli definiti sulle singole relazioni e su cui a sua volta si divide in 2 sottocategorie:
 1. *vincoli di tupla:* sono delle limitazioni effettuate su ciascuna tupla indipendentemente dalle altre e vengono definite tramite delle espressioni booleane, i cui atomi sono valutati attraverso dei confronti, come ad esempio un vincolo sui voti degli esami può essere $(Voto \geq 0) \text{ and } (Voto \leq 30)$.

2. *vincoli di chiavi*: sono i più importanti vincoli intra-relazionali, che permettono di avere la certezza dell'unicità di uno schema di relazione, come ad esempio si garantisce che non possono esistere due studenti con la stessa matricola.

La definizione formale di una chiave è la seguente:

Definizione 4. *Un insieme K di attributi è superchiave di una relazione r se r non contiene 2 tuple distinte t_1 e t_2 , con $t_1[K] = t_2[K]$.*

Si dice che K è chiave di r se è una superchiave minimale di r .

Per la relazione, presente nella figura 2.16, l'insieme $\{\text{Cognome}, \text{Corso}\}$ è una superchiave e poiché vi sono delle tuple uguali su cognomi e su corsi, tale insieme è pure una chiave ma questo vale sulle istanze presenti nella figura 2.16 e non in maniera generale, cosa che quanto sviluppiamo uno schema di relazione cerchiamo di ottenere l'unicità delle tuple su qualsiasi elementi ammissibili, per esempio nella relazione *Studente* la chiave indicata sarebbe la matricola, indice univoco usato da tutte le università per identificare gli studenti. Si può notare che in ogni relazione e relativo schema si ha la presenza di una chiave, infatti una relazione è un insieme, costituita come si sa bene da elementi diversi, da cui di conseguenza per ogni relazione $r(X)$ l'insieme X di tutti gli attributi su cui è definita è senz'altro una superchiave che può essere di due diversi tipi:

- tale insieme è anche una chiave, per cui risulta dimostrato l'esistenza della chiave
- tale insieme non è una chiave, ossia esiste un'altra superchiave, sottoinsieme di quella considerata, da cui si può applicare ricorsivamente le stesse considerazioni fatte, usando un sottoinsieme di relazione fino ad arrivare alla superchiave minimale in una sequenza finita di passi, dato che il numero di attributi è una quantità finita.

Lo stesso ragionamento può essere fatto anche a livello di schema di relazione, in cui l'insieme di tutti gli

attributi è una superchiave per ciascuna relazione e la ricerca della chiave si effettua alla stessa maniera di quella vista per l'istanza di una relazione.

Il fatto appena dimostrato, ossia che ogni relazione possieda una chiave, ci garantisce l'accessibilità e l'unicità di tutti i valori della base di dati ed inoltre permette di stabilire efficacemente le corrispondenze fra dati contenuti in relazioni diverse.

Come già notato nel paragrafo sui valori nulli, si deve avere dei valori senza alcuna informazione e per risolvere codesto problema si adotta una soluzione semplice ma efficace ossia su una delle chiavi, detta *chiave primaria*, si vieta la presenza di valori nulli mentre sugli altri attributi non si ha solitamente nessuna problematica riguardo ai valori null.

Gli attributi che costituiscono la chiave primaria vengono evidenziati attraverso una sottolineatura ed inoltre la maggior parte dei riferimenti avviene tramite chiave primaria.

Solitamente in quasi tutti i casi reali è possibile trovare degli attributi, i cui attributi sono identificativi e sempre disponibili ma in caso ciò non è possibile si introduce un codice non significativo per l'applicazione.

- *vincoli di integrità referenziale* (Foreign Key) fra un insieme di attributi X di una relazione R_1 e una relazione R_2 risulta soddisfatto se i valori di X su ciascuna tupla di R_1 compaiono come valori della chiave dell'istanza di R_2 . Vediamo un esempio:

Infrazioni

<u>Codice</u>	Data	Vigile	Prov	Numero
34321	1/2/95	3987	MI	39548K
53524	4/3/95	3295	TO	E39548
64521	5/4/96	3295	PR	839548
73321	5/2/98	9345	PR	839548

Vigili

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

La definizione precisa e formale richiede un pochino di attenzione in più, in particolare nel caso in cui la chiave della relazione riferita sia composta e nel caso in cui vi siano più chiavi, per cui procediamo per gradi, incominciando prima con il caso di chiave unica:

Definizione 5. Sia R_2 una relazione con una sola chiave B e sia a sua volta $X = \{A\}$ l'insieme di attributi, allora il vincolo di integrità referenziale fra l'attributo A di R_1 e la relazione R_2 risulta soddisfatto se per ogni tupla t_1 in R_1 , in cui $t_1[A]$ non è nullo, esiste una tupla t_2 in R_2 tale che $t_1[A] = t_2[B]$.

Nel caso più generale bisogna stare attenti al fatto che ogni attributo di X deve corrispondere ad uno specifico attributo della chiave primaria di R_2 per questo la definizione è la seguente:

Definizione 6. Siano $X = \{A_1, A_2, \dots, A_n\}$ e $K = \{B_1, B_2, \dots, B_n\}$ due insiemi ordinati di attributi per le relazioni R_1 e R_2 , il vincolo risulta soddisfatto se per ogni tupla t_1 in R_1 , senza attributi nulli

su X , esiste una tupla t_2 in R_2 tale per cui $t_1[A_i] = t_2[B_i]$, per ogni i compreso tra 1 e p .

Per analizzare i vincoli di integrità referenziale consideriamo la figura in cui le informazioni della relazione Infrazioni sono rese significative attraverso il collegamento con la relazione Agenti, con l'attributo agente, e alla relazione Auto, per mezzo degli attributi stato e numero di targa:

Infrazioni

<u>Codice</u>	Data	Vigile	Prov	Numero
34321	1/2/95	3987	MI	39548K
53524	4/3/95	3295	TO	E39548
64521	5/4/96	3295	PR	839548
73321	5/2/98	9345	PR	839548

Vigili

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

In questa situazione analizzata abbiamo deciso di inserire i seguenti vincoli referenziali:

- fra l'attributo *Agente* della relazione Infrarossi e la relazione Agenti.
- fra gli attributi Stato e Numero di Infrarossi e la relazione Auto, in cui l'ordine degli attributi nella chiave preveda Stato e poi Numero.

Relativamente al secondo vincolo, notiamo come il ragionamento sull'ordine degli attributi può risultare pesante, visto che la corrispondenza può, almeno in questo caso, essere realizzata per mezzo dei nomi degli attributi ma in generale ciò non può accedere, quindi l'ordinamento è essenziale.

4.1.1 In pratica

Una relazione è spesso rappresentata da una tabella ove le righe rappresentano specifici record e le colonne corrispondono ai campi dei record, l'ordine di righe e colonne è sostanzialmente irrilevante:

nome	cognome	Data di nascita	professione	tel
mario	rossi	21/10/80	impiegato	02 345678
sara	bianchi	17/03/77	avvocato	031 45678
marco	verdi	11/11/67	medico	06 789052

In una base di dati relazionale ci sono più relazioni:

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

Si hanno 2 componenti:

1. lo **schema**, invariante nel tempo, che ne descrive la struttura (aspetto intensionale). Si rappresenta con le intestazioni delle tabelle
2. l'**istanza**, i valori attuali, che possono cambiare anche molto rapidamente (aspetto estensionale). Si rappresenta col corpo di ciascuna tabella

Si hanno 3 accezioni del concetto di relazione:

1. **relazione matematica**: come nella teoria degli insiemi
2. **associazione/correlazione** che rappresenta una classe di fatti, nel modello *Entity-Relationship*
3. **relazione** secondo il modello relazionale dei dati

quindi ad esempio:

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

4.1.2 Relazione Matematica

Una relazione matematica sugli insiemi D_1 e D_2 , detti **domini**, è un sottoinsieme del prodotto cartesiano $D_1 \times D_2$. Le relazioni si possono visualizzare efficacemente con una tabella in cui ogni colonna corrisponde ad un dominio e ogni riga a un elemento della relazione. Quindi una relazione matematica è un insieme di n-uple ordinate (d_1, \dots, d_n) tali che $d_i \in D_i$.

Si hanno delle proprietà:

- non c'è ordinamento fra le n-uple
- le n-uple sono distinte
- ciascuna n-upla è ordinata: l'i-esimo valore proviene dall'i-esimo dominio

Dato che ciascuno dei domini ha un ruolo definito dalla posizione nella tabella si ha una **struttura posizionale**

4.1.3 Relazioni nel Modello Relazionale

Si ha che:

- a ciascun dominio si associa un nome (attributo), che ne descrive il "ruolo"
- gli attributi possono essere usati come intestazione
- struttura non posizionale

Ovvero:

Casa	Fuori	RetiCasa	RetiFuori
Juve	Lazio	3	1
Lazio	Milan	2	0
Juve	Roma	0	2
Roma	Milan	0	1

Si hanno delle regole per poter identificare una relazione con una tabella:

- i valori di ogni colonna sono fra loro omogenei
- le righe sono diverse fra loro
- le intestazioni delle colonne sono diverse tra loro

Inoltre l'ordinamento di righe e colonne è irrilevante.

Si possono creare corrispondenze fra le tuple di relazioni distinte per mezzo di valori degli attributi che compaiono nelle ennuple:

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

dove matricola e studente rappresentano la stessa informazione, così come corso e codice.

Si hanno diversi vantaggi nell'uso della struttura basata su valori:

- 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
- i dati sono portabili più facilmente da un sistema ad un altro
- per accedere ai dati non serve sapere come sono memorizzati fisicamente

vediamo quindi un paio di esempi:

STUDENTI(Matricola,Cognome,Nome,Data di Nascita)**Istanza della relazione studenti**

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
tupla i →	3456	Rossi	Maria	01/02/1978

Da una tupla possiamo estrarre

il valore degli attributi

$t_i[\text{matricola}]=3456$

PARTITE (Casa, Fuori, RetiCasa, RetiFuori)**Istanza della relazione partite**

Partita	Casa	Fuori	RetiCasa	RetiFuori
t1	Juve	Lazio	3	1
t2	Lazio	Milan	2	0
t3	Juve	Roma	0	2
t4	Roma	Milan	0	1

$t1[\text{casa}] = \text{Juve}$

$t4[\text{casa}] = \text{Roma}$

$t1[\text{RetiCasa}] = 3$

$t2[\text{Fuori}] = \text{Milan}$

$t4[\text{Fuori}] = \text{Milan}$

$t3[\text{RetiFuori}] = 2$

e possono esistere relazioni su un solo attributo:

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

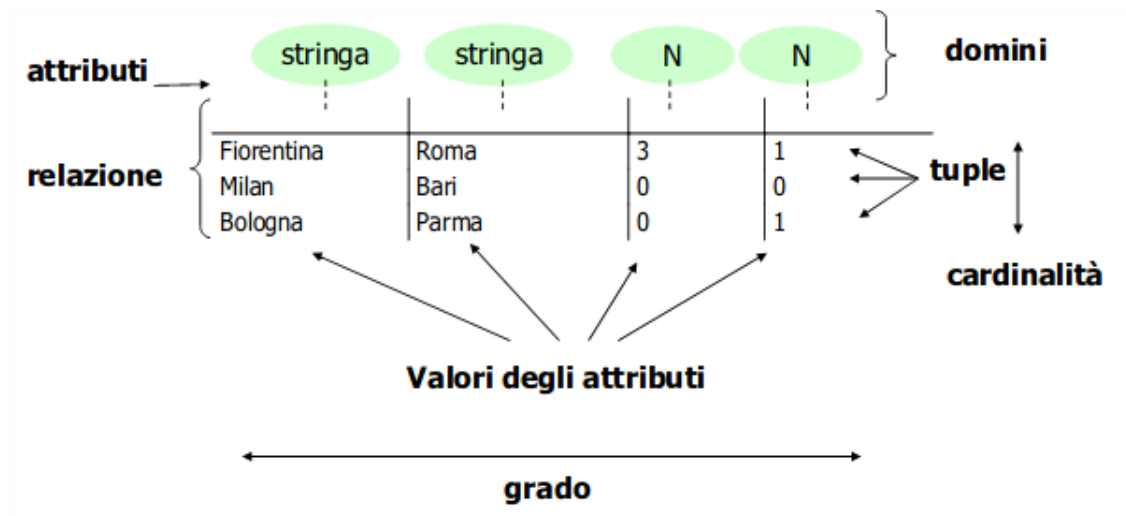
studenti lavoratori

Matricola
6554
3456

Si hanno delle regole di notazione:

- **attributi:** lettere iniziali dell'alfabeto, maiuscole: $A, B, C, A', A1, \dots$
- **insieme di attributi:** lettere finali dell'alfabeto, maiuscole: $X, Y, Z, X', X1, \dots$
- **unioni di insiemi:** XY anziché $X \cup Y$
- **nomi di relazioni:** R e lettere circostanti, maiuscole, anche con indici e pedici: $R1, S, S', \dots$
- **relazione:** come il nome della relazione, ma in minuscolo
- **schema di base di dati:** lettera maiuscola in grassetto $\mathbf{R}, \mathbf{S}, \dots$
- **base di dati:** stesso simbolo dello schema, ma in minuscolo

Vediamo un'immagine riassuntiva:



Si possono rappresentare strutture nidificate, diverse a seconda della necessità:

Ricevute	Numero	Data	Totale
	1235	12/10/2000	39,20
	1240	13/10/2000	39,00

Dettaglio	Numero	Qtà	piatto	costo
	1235	3	Coperti	3,00
	1235	2	Antipasti	6,20
	1235	3	Primi	12,00
	1235	2	Bistecche	18,00
	1240	2	Coperti	2,00

Ricevute	Numero	Data	Totale
	1235	12/10/2000	39,20
	1240	13/10/2000	39,00

Dettaglio	Numero	Riga	Qtà	Descrizione	Importo
	1235	1	3	Coperti	3,00
	1235	2	2	Antipasti	6,20
	1235	3	3	Primi	12,00
	1235	4	2	Bistecche	18,00
	1240	1	2	Coperti	2,00

Si definisce la **chiave** come *l'insieme di attributi che identificano le tuple di una relazione*.

Più formalmente un insieme di K attributi è una **superchiave** per r se r non contiene due tuple distinte r_1 e t_2 con $t_1[K] = t_2[K]$. Inoltre è chiave

per r se è una **superchiave minimale** per r (cioè non contiene un'altra superchiave). Vediamo un esempio:

Matricola	Cognome	Nome	Corso	Nascita
27655	Rossi	Mario	Ing Inf	5/12/78
78763	Rossi	Mario	Ing Inf	3/11/76
65432	Neri	Piero	Ing Mecc	10/7/79
87654	Neri	Mario	Ing Inf	3/11/76
67653	Rossi	Piero	Ing Mecc	5/12/78

Nella tabella Matricola è una superchiave ed è minimale (contiene un solo attributo). L'insieme dato da Cognome, Nome e Nascita è una superchiave minimale. Volendo sarebbe anche una chiave la coppia Cognome-Corso ma noi però interessano le chiavi corrispondenti a vincoli di integrità soddisfatti da tutte le relazioni lecite dello schema (sarebbe una chiave senza senso).

Il modello relazionale ha una **struttura rigida**, dove le informazioni sono rappresentate da tuple e solo alcuni formati di tuple sono ammessi (quelli che corrispondono agli schemi di relazione). **I dati disponibili possono non corrispondere al formato previsto.**

Può capitare di avere un'informazione incompleta. In tal caso non conviene usare valori del dominio (0, stringa nulla) perché potrebbero diventare significativi. Si usa quindi il valore nullo **NULL**, che denota l'assenza di un valore del dominio (e non è un valore del dominio). Si hanno almeno tre casi di valore nullo anche se i DBMS non li distinguono:

1. valore sconosciuto
2. valore inesistente
3. valore senza informazione

Ovviamente l'uso del valore NULL deve essere sensato (per esempio di uno studente non si può avere data di nascita o matricola NULL). Esistono istanze di basi di dati che, pur sintatticamente corrette, non rappresentano informazioni possibili per l'applicazione di interesse. Una chiave in cui non sono ammessi NULL è detta **chiave primaria** e come notazione si usa la sottolineatura

Capitolo 5

SQL

Il linguaggio SQL è un linguaggio per la definizione e la manipolazione dei dati in database relazionali, sviluppato originariamente presso il laboratorio IBM a San Jose' (California) e adottato nel sistema System R. L'SQL è stato poi adottato da molti altri DBMS e quindi è stato soggetto ad un'intensa attività di standardizzazione. È un linguaggio con varie funzionalità che contiene:

- **DDL:** definizione di domini, tabelle, autorizzazioni, vincoli, procedure, ecc.
- **DML:** linguaggio di query, modifica, ...

Si studierà una versione vecchia di SQL in quanto è la più implementata. Nello specifico studieremo SQL-2. SQL-3 è comunque compatibile con SQL-2 e introduce il concetto di oggetto. Noi useremo SQL per definire, aggiornare e interrogare la base di dati. Sono sempre più frequenti in realtà sistemi dotati di interfacce più facili da usare. Questi programmi generano le istruzioni SQL corrispondenti.

Esistono diverse implementazioni di SQL, per esempio il DBMS di Oracle, Access, Mysql...

SQL è relazionale completo e ogni espressione dell'algebra relazionale può essere tradotta in SQL. Il modello dei dati di SQL è basato su tabelle anziché relazioni e possono essere presenti righe duplicate (le tuple) e adotta la logica a 3 valori dell'algebra relazionale (gestendo il null e aggiungendo il valore di verità U, *unknown*).

Le tabelle di verità diventano quindi:

Tabelle di verità

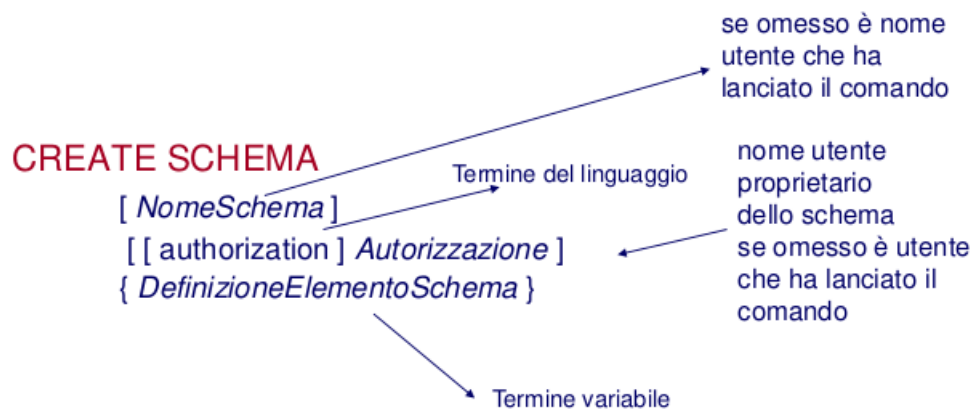
<i>not</i>		<i>and</i>	F T U	<i>or</i>	F T U
T	F	T	F T U	T	T T T
F	T	F	F F F	F	F T U
U	U	U	F U U	U	U T U

Vediamo la notazione per definire la sintassi di SQL:

- $\langle x \rangle$ per isolare un termine x
- $[x]$ indicano che un termine x è opzionale
- $\{x\}$ indicano che un termine x può essere ripetuto 0 volte o un numero arbitrario di volte
- $|$ separa opzioni alternative

Le parentesi tonde () appartengono al linguaggio SQL e non alla notazione sopra descritta.

Uno **schema di base di dati** è una collezione di oggetti e ogni schema ha un nome e un proprietario. SI ha la seguente sintassi:



Dopo il comando CREATE SCHEMA compaiono le definizioni dei vari elementi. Non è necessario che la definizione di tutti gli elementi avvenga contemporaneamente alla creazione dello schema. Può avvenire in più fasi successive.

Una **tabella** è costituita da una collezione ordinata di attributi e da un insieme (eventualmente vuoto) di vincoli:

```
CREATE TABLE NomeTabella (  
    NomeAttributo Dominio [ValoreDiDefault] [Vincoli]  
    {,NomeAttributo Dominio [ValoreDiDefault] [Vincoli] }  
    [AltriVincoli])
```

questa istruzione quindi definisce uno schema di relazione e ne crea un'istanza vuota. Per ogni attributo va specificato il dominio, un eventuale valore di default (per questo è comodo il null, che è il valore di default di base) ed eventuali vincoli. Possono essere espressi altri vincoli a livello di tabella (tra più attributi).

I **domini** specificano i valori ammissibili per gli attributi di una relazione. SQL ha 6 domini predefiniti:

1. bit SQL-2 poi eliminato e sostituito parzialmente da BOOLEAN in SQL-3
2. carattere
3. numerico Esatto
4. numerico Approssimato
5. data/Ora
6. intervallo Temporale

più ovviamente i domini definiti dall'utente.

Vediamo il **dominio di tipo Bit**. Può essere di lunghezza fissa o variabile. Se la lunghezza non è specificata corrisponde ad 1 singolo valore. Corrisponde ad attributi che possono assumere solo due valori (0,1). Attributi di questo tipo (*flag*) indicano se l'oggetto rappresentato possiede o meno una certa proprietà:

```
bit [varying] [(lunghezza)]  
bit  
bit(lunghezza)  
bit varying (lunghezza)  
varbit (lunghezza)
```

Esempio 4. *Si definisce l'attributo `Lavoratore` nella relazione `STUDENTI` per indicare se lo studente è o meno lavoratore:*

Lavoratore **bit**

Vediamo il **tipo carattere**. Rappresenta singoli caratteri alfanumerici oppure stringhe di lunghezza fissa o variabile:

character [**varying**] [(lunghezza)]
character o **char**
character(lunghezza)
varchar(lunghezza)
character varying (lunghezza)

Esempio 5. *Si definisce l'attributo `Nome` della relazione `IMPIEGATI` come sequenza di caratteri di lunghezza massima 20:*

Nome **char**(20)
Nome **varchar**(20)

con il primo che genera `Paolo_Bianchi_ _ _ _ _` e il secondo `Paolo_Bianchi`

Vediamo i **tipi numerici esatti**. Rappresentano numeri interi o numeri decimali in virgola fissa (con un numero prefissato di decimali, ad esempio i valori monetari). Precision è numero di cifre significative, scala il numero di cifre dopo la virgola:

integer
smallint
numeric **numeric**(precisione) **numeric**(precisione,scala)
decimal **decimal**(precisione) **decimal**(precisione,scala)

Esempio 6. *Si definisce l'attributo `Eta` nella relazione `IMPIEGATI`:*

Eta **decimal**(2) // Rappresenta tutti i numeri fra -99 e +99

Si definisce l'attributo `Cambio` nella relazione `PAGAMENTO` per il valore del cambio di una certa moneta preciso al centesimo

Cambio **numeric**(6,2) // Rappresenta tutti i numeri fra -9999,99 e +9999,99

quindi `INTEGER` / `SMALLINT` rappresentano valori interi. La precisione (numero totale di cifre) varia a seconda della specifica implementazione di SQL, non è specificata nello standard. `SMALLINT` richiede minore spazio di

memorizzazione.

NUMERIC / *DECIMAL* rappresentano i valori decimali. La differenza tra *NUMERIC* e *DECIMAL* è che il primo deve essere implementato esattamente con la precisione richiesta, mentre il secondo può avere una precisione maggiore, la precisione segnalata è requisito minimo. Se la precisione non è specificata si usa il valore caratteristico dell'implementazione. Per la scala si usa valore 0. Si hanno ovviamente:

```
float float(precisione)
double precision
real
```

che sono utili per rappresentare valori reali approssimati, ad esempio grandezze fisiche (rappresentazione in virgola mobile, in cui a ciascun numero corrisponde una coppia di valori: mantissa e esponente).

Per la data si ha:

```
date
time time(precisione)
timestamp timestamp(precisione)
```

Ciascuno di questi domini è strutturato e decomponibile in un insieme di campi (anno, mese, giorno, ora, minuti, secondi):

```
DataDiNascita date // 1999-09-18
OraDiConsegna time // 19.24.16
Arrivo timestamp // 2000-09-18 21.15.20
```

SI possono avere intervalli temporali:

```
interval PrimaUnitàDiTempo
interval PrimaUnitàDiTempo to UltimaUnitàDiTempo
```

Infine si hanno i tipi **blob**. Permettono di includere direttamente nel database oggetti molto grandi. *Binary Large Object (BLOB)* e *Character Large Object (CLOB)*:

```
fotografia BLOB(10M)
descrizione CLOB(100k)
```

Definiti solo in SQL-3, ma implementati in diversi DBMS commerciali. Il sistema garantisce solo di memorizzarne il valore. Non possono essere usati come criterio di selezione per le interrogazioni.

Per creare un dominio uso la primitiva *create domain*:


```
CREATE DOMAIN NomeDominio AS DominioElementare
        [ValoreDefault] [Constraints]
```

Un nuovo dominio è caratterizzato dalle seguenti informazioni: nome, dominio elementare, valore di default, insieme di vincoli (constraints). Vediamo un esempio:

```
CREATE DOMAIN Voto AS SMALLINT
    DEFAULT 0
    NOT NULL
```

Il nuovo dominio Voto è definito come uno SMALLINT con valore di default e che non deve essere null.

La definizione di "nuovi domini" è utile perché permette di associare dei vincoli a un nome di dominio: questo è importante quando si deve ripetere la stessa definizione di attributo su diverse tabelle: ad esempio, modifiche alla definizione di Voto si ripercuotono in tutte le occorrenze di questo dominio nello schema del Database.

Esempio 7. *Vediamo un esempio concreto:*

```
CREATE TABLE NomeTabella (
    NomeAttributo Dominio [ValoreDiDefault] [Vincoli]
    {,NomeAttributo Dominio [ValoreDiDefault] [Vincoli] }
    [AltriVincoli] )
```

```
CREATE TABLE Impiegato (
    Matricola CHAR(6) PRIMARY KEY,
    Nome CHAR(20) NOT NULL,
    Cognome CHAR(20) NOT NULL,
    Dipart CHAR(15)
    Stipendio NUMERIC(9) DEFAULT 0,
    UNIQUE (Cognome, Nome)
)
```

UNIQUE richiede che non ci siano coppie cognome-nome ripetute

Un vincolo (constraint) è una regola che specifica delle condizioni sui valori di un elemento dello schema del database. Un vincolo può essere associato ad una tabella, ad un attributo, ad un dominio. Sono di due tipi:

1. **vincoli intrarelazionali** che si applicano all'interno di una relazione. Possono essere:

```
NOT NULL // (Il valore deve essere non null)
UNIQUE // (I valori devono essere non ripetuti)
PRIMARY KEY // (Chiave primaria)
CHECK // (Condizioni complesse)
```

Il vincolo PRIMARY KEY può essere definito una sola volta all'interno della relazione. In alcune implementazioni di SQL potrebbe essere necessario specificare comunque anche il vincolo NOT NULL per tutti gli attributi coinvolti.

2. **vincoli interrelazionali** che si applicano tra relazioni diverse. Possono essere definiti attraverso i costrutti sintattici:

```
REFERENCES // Permettono di definire vincoli di integrità
FOREIGN KEY // referenziale
CHECK // (Vincoli complessi)
```

e si hanno sintassi per singoli o più attributi. È possibile definire politiche di reazione alle violazioni. Si ha l'**integrità referenziale** che esprime un legame gerarchico (padre / figlio) fra tabelle. Alcuni attributi della tabella figlio sono definiti FOREIGN KEY e si devono riferire (REFERENCES) ad alcuni attributi della tabella padre che costituiscono una chiave (devono essere UNIQUE e NOT NULL oppure PRIMARY KEY). I valori contenuti nella FOREIGN KEY devono essere sempre presenti nella tabella padre. Si hanno due sintassi:

- (a) nella parte di definizione degli attributi con il costrutto sintattico REFERENCES:

```
AttrFiglio CHAR(3) REFERENCES TabellaPadre(AttrPadre)
```

- (b) oppure dopo le definizioni degli attributi con i costrutti FOREIGN KEY e REFERENCES:

```
FOREIGN KEY (AttrFiglio)
REFERENCES TabellaPadre(AttrPadre)
```

quindi:

```
CREATE TABLE Impiegato (
  Matricola CHAR(6) PRIMARY KEY,
  Nome CHAR(20) NOT NULL,
  Cognome CHAR(20) NOT NULL,
  Dipart CHAR(15) REFERENCES Dipartimento(NomeDip) ,
```

```

        Stipendio NUMERIC(9) DEFAULT 0,
        UNIQUE (Cognome, Nome)
    )

```

Se si omettono gli attributi destinazione, vengono assunti quelli della chiave primaria.

```

AttrFiglio CHAR(3) REFERENCES TabellaPadre

```

Quando si hanno più attributi da riferire, si utilizza sempre FOREIGN KEY:

```

FOREIGN KEY (AttrFiglio1 { }, AttrFiglio2 { })
    REFERENCES TabellaPadre (AttrPadre1 { }, AttrPadre2 { })

```

Esempio 8. Definiamo tre tabelle con le informazioni degli esami sostenuti dagli studenti: *Studente*, *Esame*, *Corso*:

```

CREATE TABLE Studente (
    Matr CHAR(6)
    Nome VARCHAR(30)
    Città VARCHAR(20),
    CDip CHAR(3)
)

```

<u>Matr</u>	Nome	Città	CDip
34321	Luca	Mi	Inf
53524	Giovanni	To	Mat
64521	Emilio	Ge	Ing
73321	Francesca	Vr	Mat

```

CREATE TABLE Esame (
    Matr CHAR(6),
    CodCorso CHAR(6),
    Data DATE NOT NULL,
    Voto Voto,
    PRIMARY KEY (Matr, CodCorso)
)

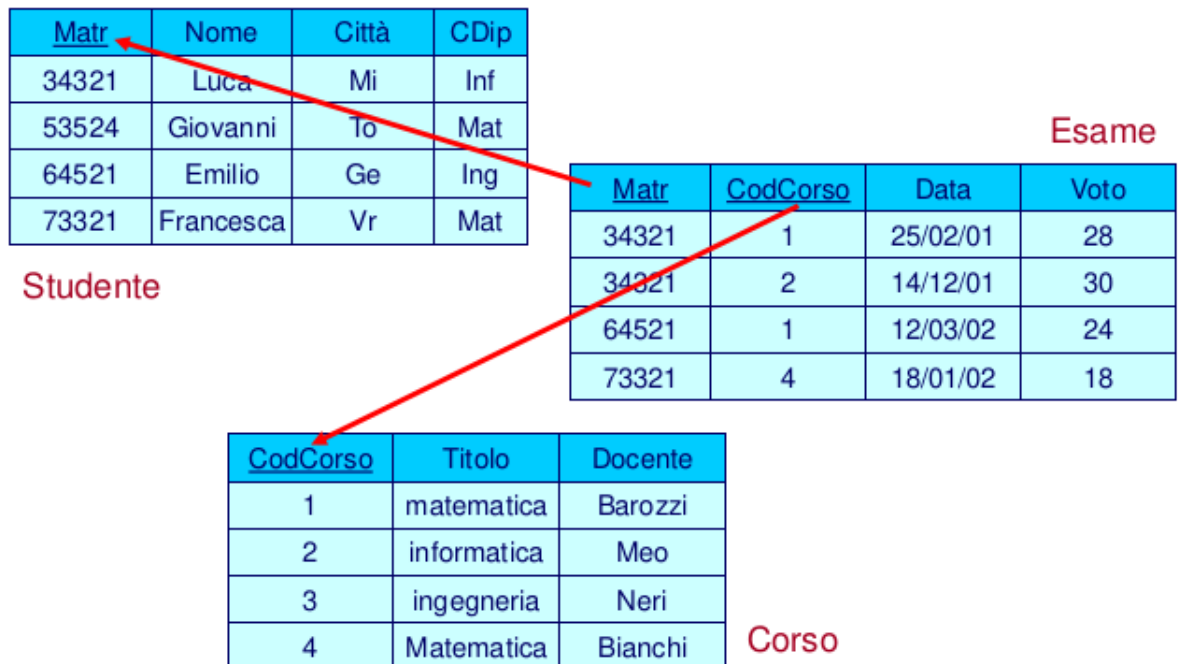
```

<u>Matr</u>	<u>CodCorso</u>	Data	Voto
34321	1	25/02/01	28
34321	2	14/12/01	30
64521	1	12/03/02	24
73321	4	18/01/02	18

```
CREATE TABLE Corso (
  CodCorso CHAR(6) PRIMARY KEY,
  Titolo VARCHAR(30) NOT NULL,
  Docente VARCHAR(20)
)
```

<u>CodCorso</u>	Titolo	Docente
1	matematica	Barozzi
2	informatica	Meo
3	ingegneria	Neri
4	Matematica	Bianchi

con le relazioni:



aggiungiamo un vincolo a esame:

```
CREATE TABLE Esame (
  Matr CHAR(6),
  CodCorso CHAR(6),
  Data DATE NOT NULL,
  Voto Voto,
  PRIMARY KEY (Matr, CodCorso)
  FOREIGN KEY (Matr) REFERENCES Studente
  FOREIGN KEY (CodCorso) REFERENCES Corso
```

oppure:

```
CREATE TABLE Esame (
  Matr CHAR(6), REFERENCES Studente,
  CodCorso CHAR(6), REFERENCES Corso,
  Data DATE NOT NULL,
  Voto Voto,
  PRIMARY KEY (Matr, CodCorso)
```

Si introduce il **problema delle violazioni**.

Per tutti gli altri vincoli visti fino ad ora, a seguito di una violazione, il comando di aggiornamento viene rifiutato segnalando l'errore all'utente.

Per i vincoli di integrità referenziale invece SQL permette di scegliere delle reazioni da adottare in caso di violazioni.

Si possono introdurre violazioni operando sulle righe della tabella padre (tabella esterna) o sulle righe della tabella figlio (tabella interna). Si hanno le seguenti modifiche sulla tabella interna (figlio):

- inserimento di una nuova riga
- modifiche della *foreign key*

Inoltre non vengono proposte reazioni, solo il rifiuto dell'operazione. Per la tabella esterna (padre) si hanno le seguenti modifiche:

- cancellazione di una riga
- modifica dell'attributo riferito

inoltre vengono proposte diverse reazioni.

Si ha anche il problema dell'aggiornamento. Se rimuovi una matricola alla tabella Esame si aggiunge un problema, il **problema degli orfani**.

Si hanno le reazioni alle violazioni, che si introducono con una modifica (update) dell'attributo cui si fa riferimento o con la cancellazione di tuple. Le reazioni operano sulla tabella figlio (es Esami), in seguito a modifiche alla tabella padre (es Studente):

CASCADE: propaga la modifica

SET NULL: annulla l'attributo che fa riferimento

SET DEFAULT: assegna il valore di **default** all'attributo

NO ACTION: impedisce che la modifica possa avvenire

A fini diagnostici (e di documentazione) è spesso utile sapere quale vincolo è stato violato a seguito di un'azione sul DB. A tale scopo è possibile associare dei nomi ai vincoli, ad esempio:

```
tipendio INTEGER CONSTRAINT StipendioPositivo  
CHECK (Stipendio > 0),
```

```
CONSTRAINT ForeignKeySedi  
FOREIGN KEY (Sede) REFERENCES Sedi
```

Per le modifiche degli schemi si hanno:

- **ALTER** per la modifica
- **DROP** che cancella oggetti dallo schema

```
ALTER DOMAIN NomeDominio <
  SET DEFAULT ValoreDefault |
  DROP DEFAULT |
  ADD CONSTRAINT DefVincolo |
  DROP CONSTRAINT NomeVincolo >
```

```
ALTER TABLE NomeTabella <
  ALTER COLUMN NomeAttributo <
    SET DEFAULT NuovoDefault |
    DROP DEFAULT > |
  DROP COLUMN NomeAttributo |
  ADD COLUMN DefAttributo |
  DROP CONSTRAINT NomeVincolo
  ADD CONSTRAINT DefVincolo >
```

DROP cancella oggetti DDL, si applica su domini, tabelle, indici, view, asserzioni, procedure,...

```
DROP < schema, domain, table, view, ... > NomeElemento
[ RESTRICT | CASCADE ]
```

- RESTRICT: impedisce drop se gli oggetti comprendono istanze non vuote
- CASCADE applica drop agli oggetti collegati. Potenziale pericolosa reazione a catena

5.0.1 Interrogazioni in SQL

Le interrogazioni si fanno con la SELECT:

```
SELECT ListaAttributi
FROM ListaTabelle
[WHERE Vondizioni]
```

```
SELECT AttrEspr [ [ AS ] Alias ] {, AttrEspr [ [ AS ] Alias ] }
FROM Tabella [ [ AS ] Alias ] {, Tabella [ [ AS ] Alias ] }
[ WHERE Condizione ]
```

con gli AS setto gli alias. Vediamo qualche esempio:

```
-- Individuare lo stipendio degli impiegati di cognome
"Rossi":
```

```
SELECT Stipendio
FROM Impiegato
WHERE Cognome="Rossi"
```

```
-- Selezionare nome, cognome e mansione degli
impiegati dell'ufficio 10
```

```
SELECT Nome, Cognome, Mansione
FROM Impiegato
WHERE ufficio='10'
```

```
-- Individuare le informazioni degli impiegati con
stipendio > 40
Individuare le informazioni degli impiegati con
stipendio > 40
```

```
-- Selezionare tutte le informazioni sui dipendenti:
```

```
SELECT id_impiegato, Nome, Cognome, Dipart, Ufficio,
stipendio, premioprod, Mansione, Città, idcapo
FROM Impiegato
```

```
SELECT *
FROM Impiegato
```

```
-- Selezionare cognome, stipendio e premio di
produzione per gli impiegati che hanno uno
stipendio compreso tra 38 e 60:
SELECT cognome, stipendio, premioprod AS premio_di_produzione
FROM Impiegato
WHERE ( stipendio>=38 AND stipendio <=60 )
```

```
-- Selezionare i nomi, i cognomi e le città di provenienza
degli impiegati
```

```
SELECT Nome, Cognome, città AS città_provenienza
FROM Impiegato
```

```
-- Selezionare i nomi e i cognomi degli impiegati, le città
di provenienza e le città in cui lavorano:
```

```
SELECT Nome, Cognome, CittàDip AS città_lavoro,
Città AS città_provenienza
FROM Impiegato, Dip
WHERE Dipart = NomeDip
```


/ Specificare più tabelle su cui operare per determinare il risultato significa: eseguire il prodotto cartesiano delle tabelle e poi selezionare le righe che soddisfano clausola WHERE.
Join di algebra relazionale*/*

-- Selezionare i nomi e i cognomi degli impiegati, le città di provenienza e le città in cui lavorano:

```
SELECT Nome, Cognome, CittàDip AS città_lavoro,  
Città AS città_provenienza  
FROM Impiegato, Dip  
WHERE Dipart = NomeDip
```

-- Selezionare i nomi e i cognomi degli impiegati, le città di provenienza e le città in cui lavorano:

```
SELECT Nome, Cognome, Città AS città_lavoro,  
Città AS città_provenienza  
FROM Impiegato, Dip  
WHERE Dipart = Nome
```

/ Notazione punto: si utilizza il nome della tabella cui l'attributo fa riferimento */*

```
SELECT Impiegato.Nome, Cognome, Dip.Città AS città_lavoro,  
Impiegato.Città AS città_provenienza  
FROM Impiegato, Dip  
WHERE Dipart = Dip.Nome
```

/ Si usano Alias per ridenominare le tabelle:*

1) Abbreviano riferimento a tabelle

2) Risolvono ambiguità di riferimento/*

```
SELECT I.Nome, Cognome, D.Città AS città_lavoro,  
I.Città AS città_provenienza  
FROM Impiegato I, Dip D  
WHERE Dipart = D.Nome
```

pagina 18