

Progettare una base di dati Il modello Entità Relazione

Gestione informatica dei dati

Progettazione di basi di dati

- È una delle attività del processo di sviluppo dei sistemi informativi
- va quindi inquadrata in un contesto più generale:
- **il ciclo di vita dei sistemi informativi:**
 - Insieme e sequenzializzazione delle attività svolte da analisti, progettisti, utenti, nello sviluppo e nell'uso dei sistemi informativi
 - attività iterativa, quindi ciclo



Gestione informatica dei dati

Fasi della progettazione



Gestione informatica dei dati

La progettazione di un sistema informativo riguarda due aspetti:

- ▶ **progettazione dei dati**

- ▶ i dati hanno un ruolo centrale
- ▶ i dati sono più stabili

- ▶ **progettazione delle applicazioni**

- Per garantire prodotti di buona qualità è opportuno seguire una **metodologia di progetto**, ovvero:
 - articolazione delle attività in fasi
 - criteri di scelta
 - modelli di rappresentazione
 - generalità e facilità d'uso



Gestione informatica dei dati

Progettazione
concettuale

“CHE COSA”:
analisi

Schema concettuale

Progettazione
logica

Schema logico

“COME”:
progettazione

Progettazione
fisica

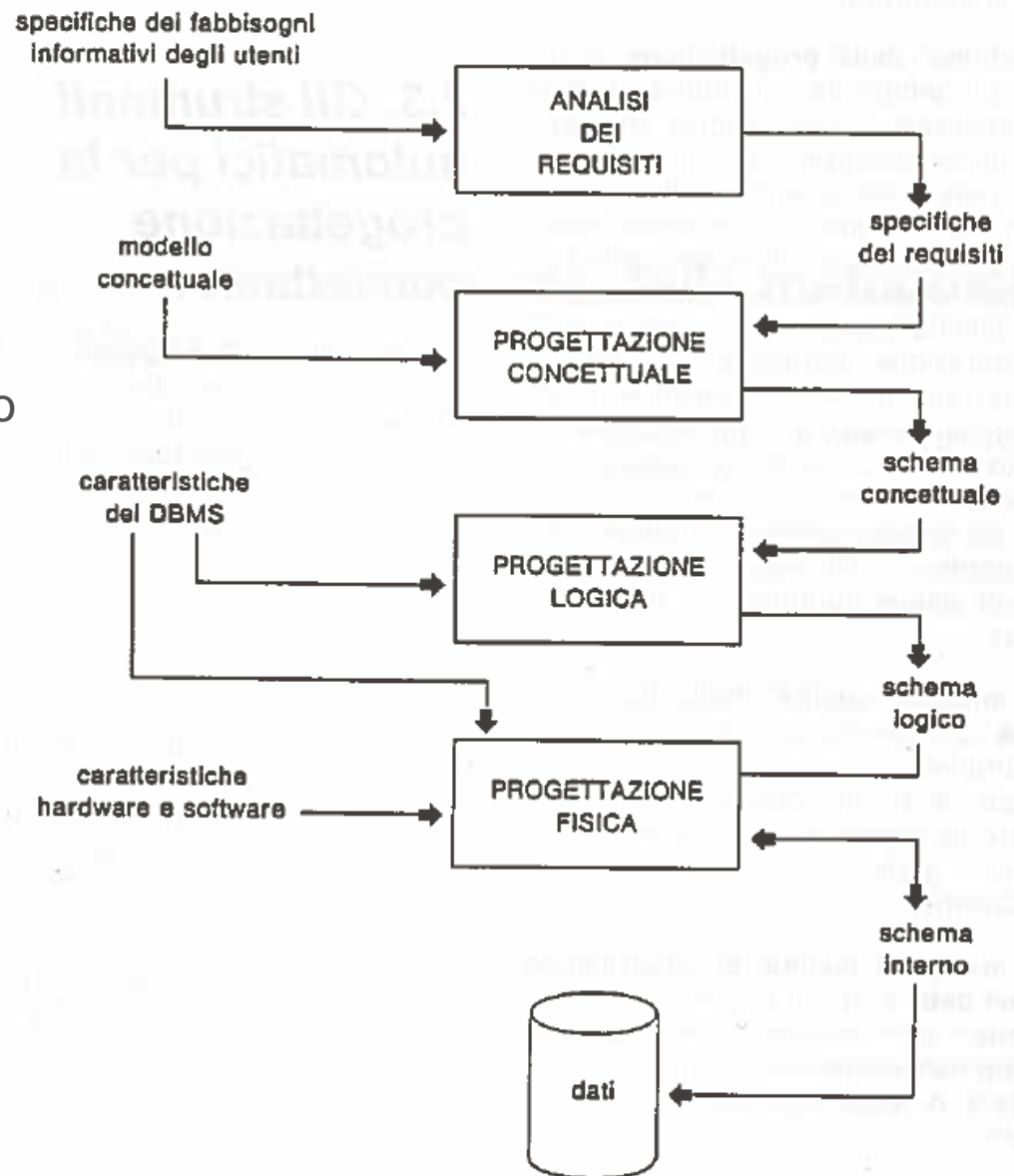
Schema fisico



La progettazione

La **progettazione concettuale** per la costruzione di un sistema informativo offre una serie di vantaggi rilevanti. L'uso degli schemi concettuali permette:

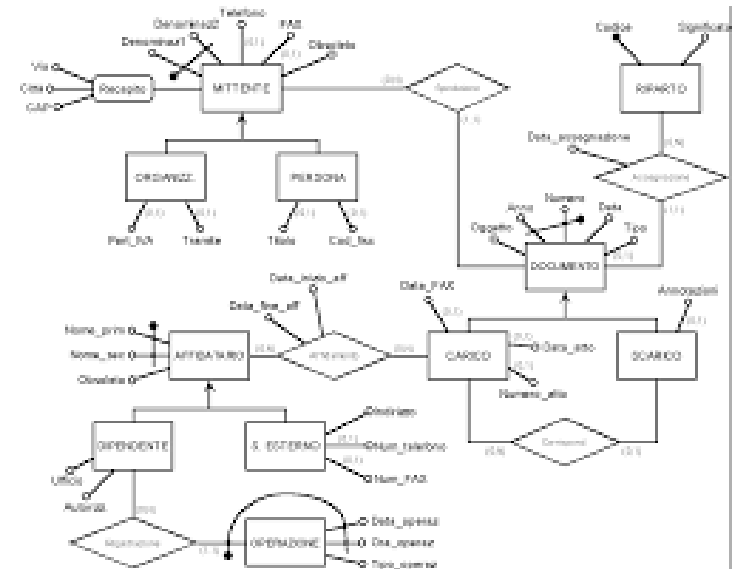
1. il coinvolgimento e il controllo da parte del committente del risultato,
2. maggiore chiarezza degli obiettivi,
3. stabilità della progettazione,
4. maggiore facilità di integrazione di nuovi dati,
5. migliore possibilità di confronto fra sistemi informativi



Progettazione concettuale

Modelli concettuali: perché sono utili?

- servono per ragionare sulla realtà di interesse, indipendentemente dagli aspetti realizzativi
- permettono di rappresentare le classi di dati di interesse e le loro correlazioni
- prevedono efficaci rappresentazioni grafiche (utili anche per documentazione e comunicazione)



Progettazione concettuale

Modello Entità – Relazione (E-R)

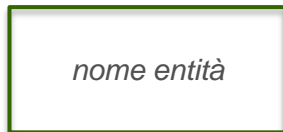
Il modello di rappresentazione Entità-Relazione è stato proposto da P. P. Chen nel 1976 in ambito informatico allo scopo di facilitare la progettazione di una base di dati.

Mediante tale modello è possibile costruire un grafo, ossia uno schema che costituisce una fotografia del fenomeno. Ciascuno schema è il risultato della composizione logica di cinque tipi di strutture di rappresentazione che si avvalgono di regole formali e di una simbologia grafica molto semplice ma estremamente chiara.

Tali strutture sono: l'entità, la relazione, l'attributo, il sottoinsieme, la gerarchia di generalizzazione.



Modello Entità – Relazione (E-R)



Il rettangolo rappresenta una **entità**, ossia un insieme di oggetti di cui sono note alcune caratteristiche. In termini statistici, il rettangolo rappresenta un collettivo di unità statistiche.

Occorrenza (o **istanza**) di entità è un elemento della classe - l'oggetto, la persona, lo studente, ecc. -, non sono i dati! (aspetto intensionale)

Nello schema concettuale vengono rappresentate le entità, non le singole istanze (il modello è astratto o concettuale)

Ogni entità ha un nome che la identifica univocamente nello schema. È conveniente utilizzare nomi espressivi e al singolare.

Modello Entità – Relazione (E-R)

Studente

Materia

Prodotto

Impiegato



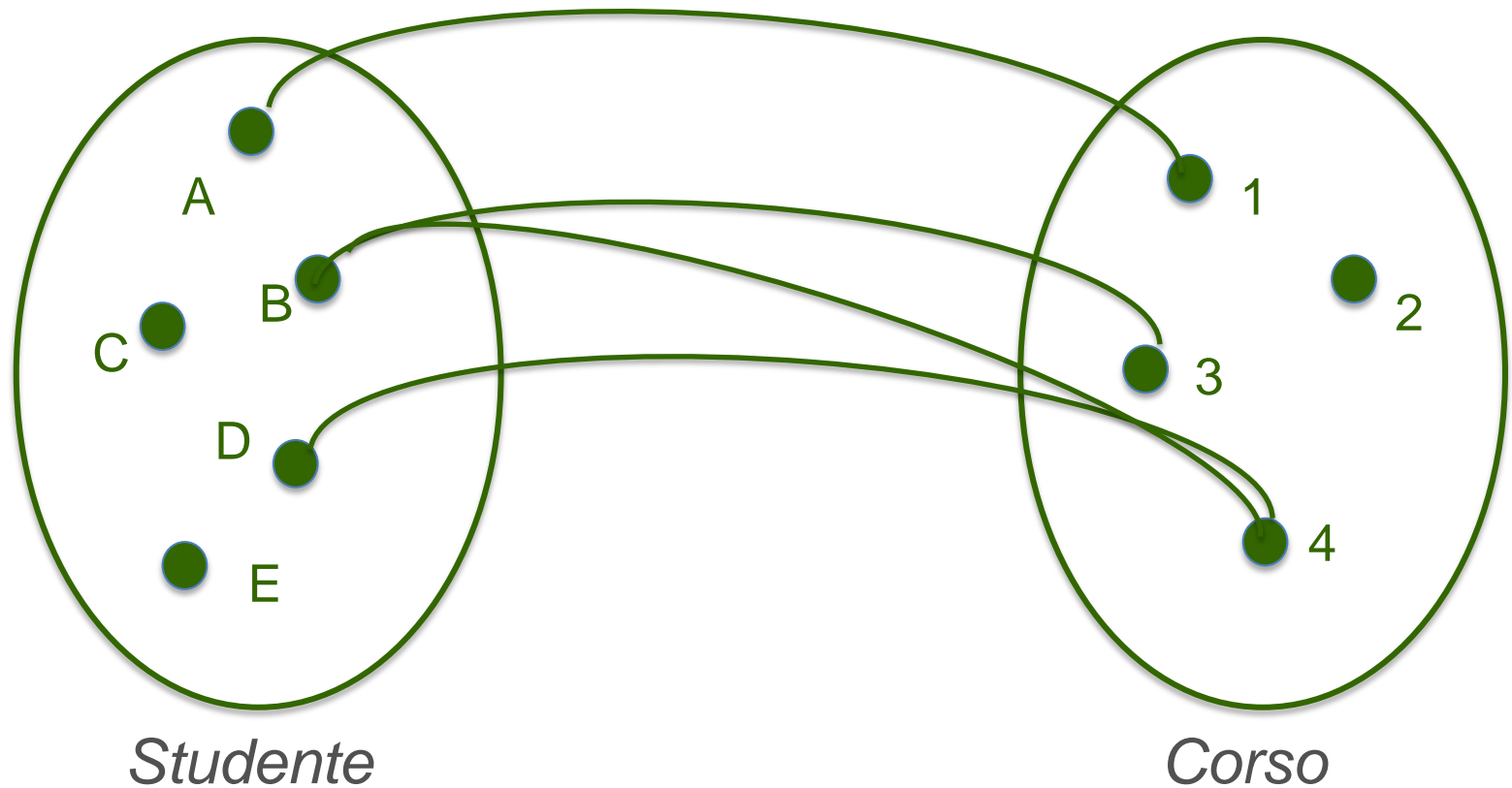
Modello Entità – Relazione (E-R)



Il rombo rappresenta una **relazione**, ovvero un legame logico tra due o più entità.
Si trova indicata anche come correlazione o associazione.



Modello Entità – Relazione (E-R)



Modello Entità – Relazione (E-R)

Ogni entità ha un nome che la identifica univocamente nello schema. È conveniente utilizzare nomi espressivi , al singolare e usando sostantivi invece che verbi.

Nell'ambito di una relationship non ci possono essere occorrenze (coppie, ennuple) ripetute.

Le relazioni possono essere di tre tipi:

Uno a molti,

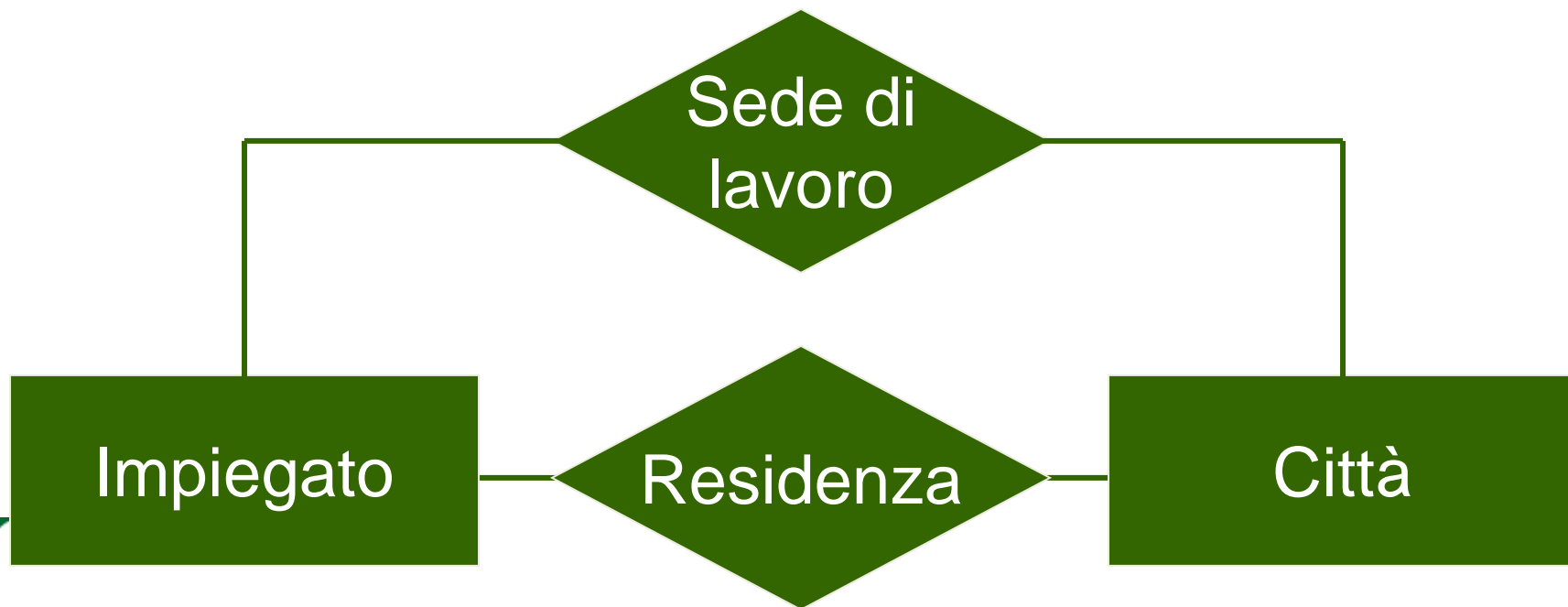
Molti a molti,

Uno a uno



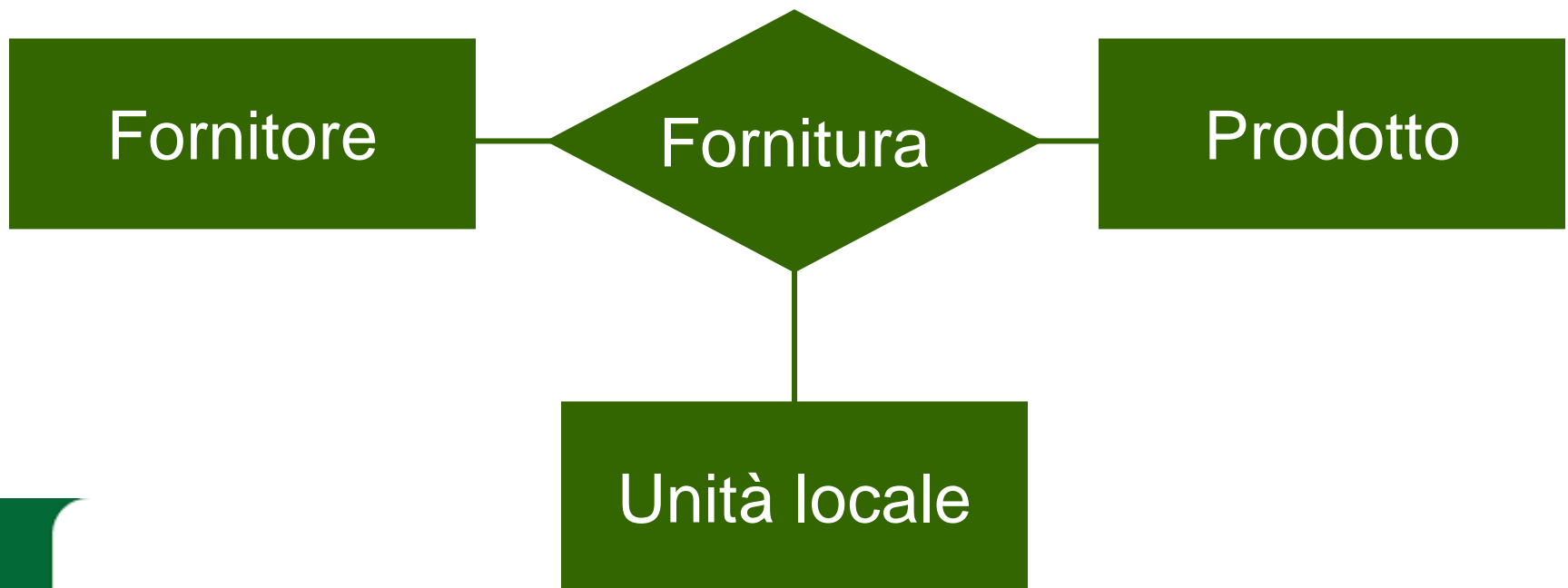
Modello Entità – Relazione (E-R)

Due entità possono essere coinvolte in più relazioni

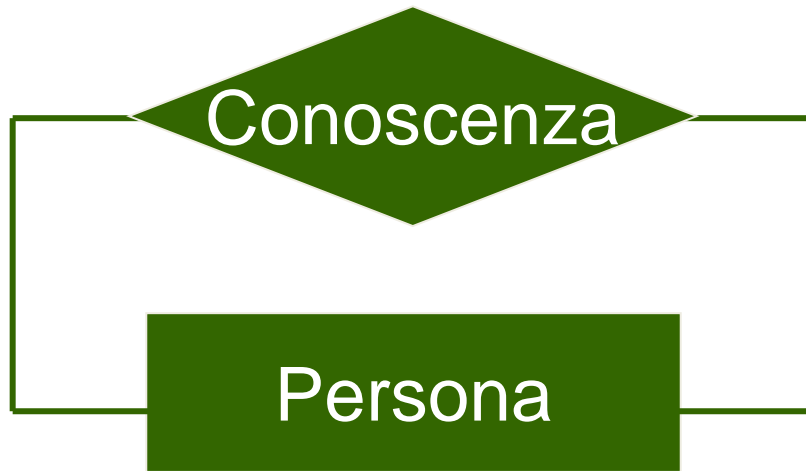


Modello Entità – Relazione (E-R)

Le relazioni possono coinvolgere più di due entità



Modello Entità – Relazione (E-R)



Una relazione può coinvolgere due volte la stessa entità (relazione ricorsiva). In alcuni casi è necessario specificare i “ruoli” delle dell’entità.



Modello Entità – Relazione (E-R)



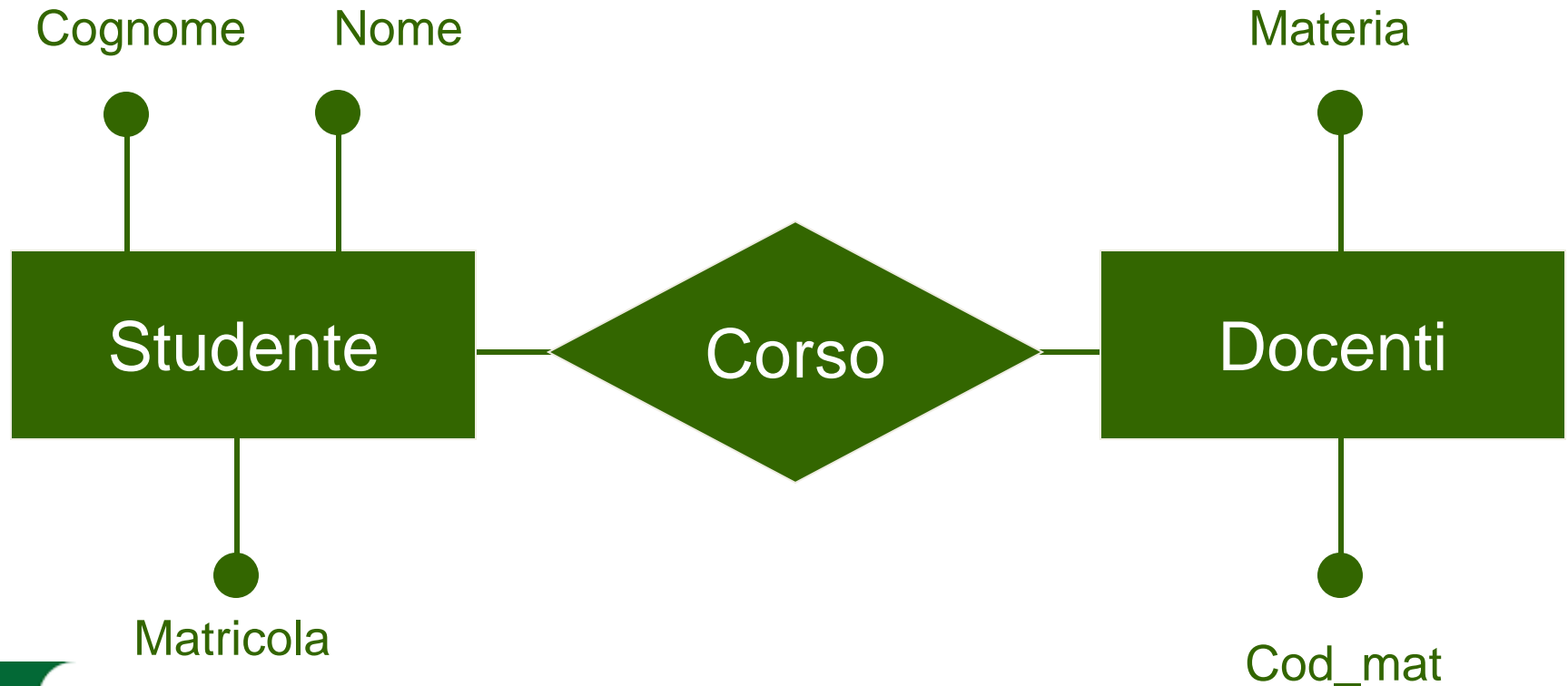
Una riga con un pallino finale rappresenta gli **attributi** di una entità o di una relazione.

Associa ad ogni occorrenza di entità o di relationship un valore appartenente a un insieme detto **dominio** dell'attributo.

Un attributo può realizzare un solo valore (ad esempio lo stato civile) o essere un attributo multiplo (ad esempio le lingue straniere conosciute).

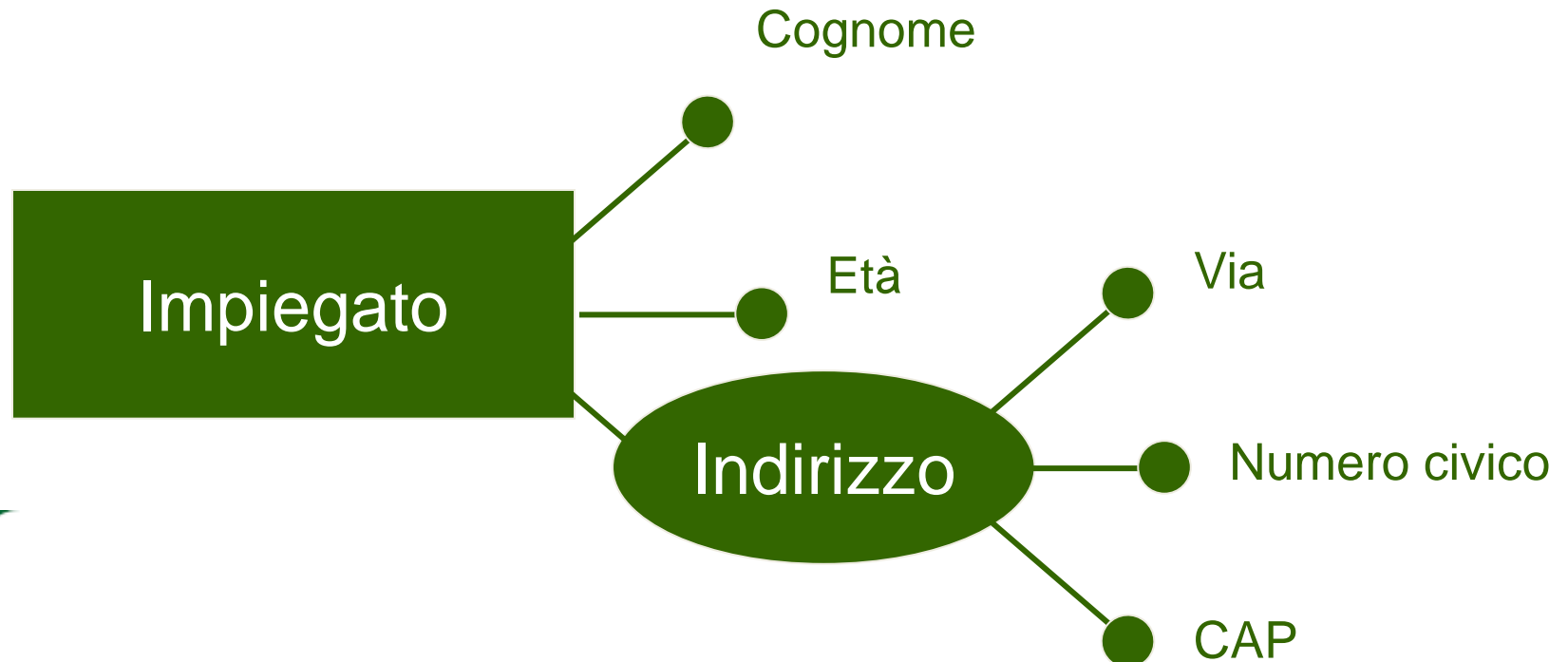


Modello Entità – Relazione (E-R)



Modello Entità – Relazione (E-R)

Attributi semplici e composti (rappresentati da un ovale) qualora è definibile dall'unione di due attributi: ad esempio l'indirizzo dell'abitazione di residenza.

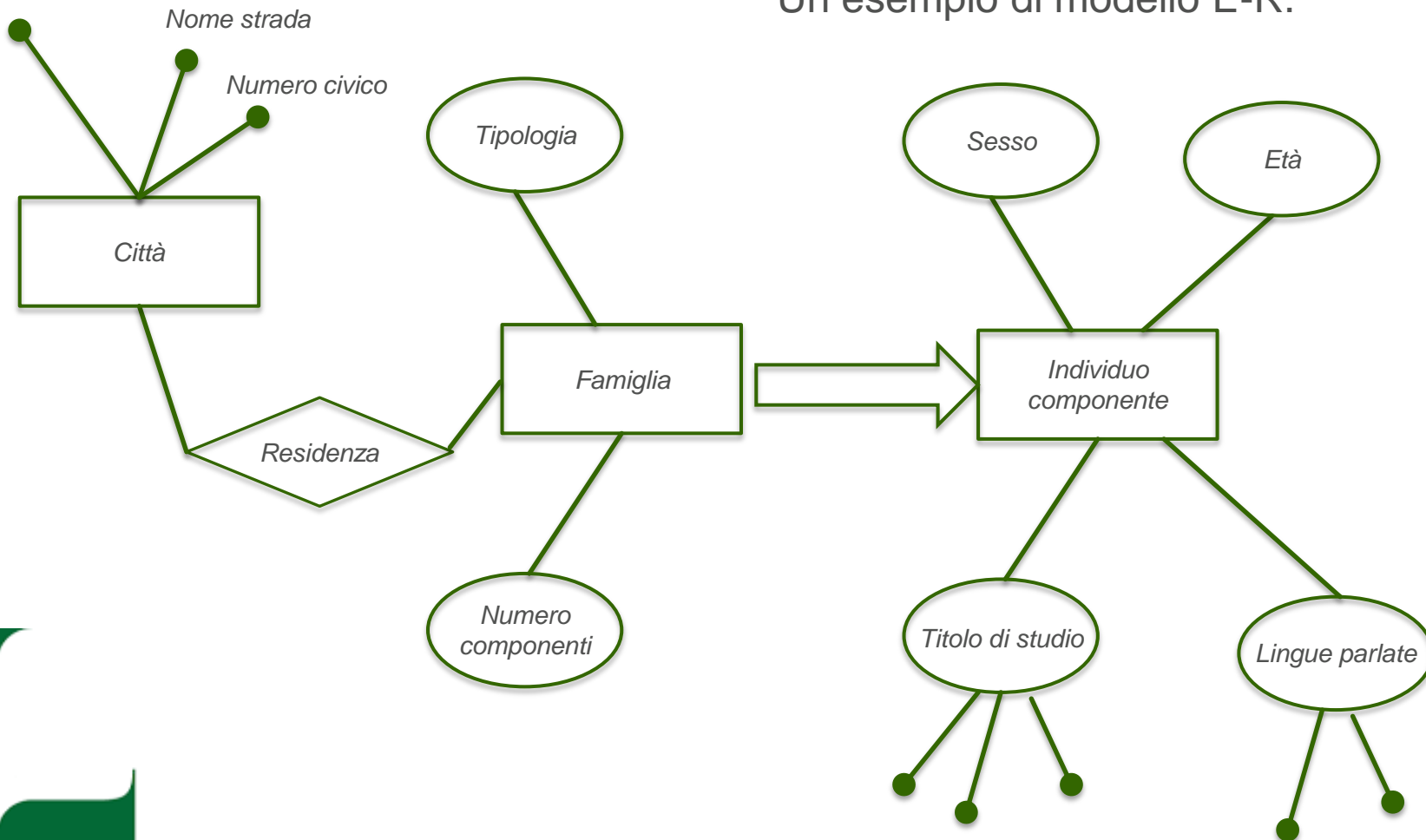


Modello Entità – Relazione (E-R)

Area circolazione

Nome strada

Numero civico



Un esempio di modello E-R.

Modello Entità – Relazione (E-R)



sottoinsieme

Una freccia rappresenta una **relazione di sottoinsieme**. Due unità si trovano in una relazione di sottoinsieme se ogni istanza dell'entità dipendente (detta entità figlia) è anche istanza dell'entità superiore (detta entità padre) mentre non è vero il contrario.

Se la relazione è su più entità si parla di **generalizzazione**.

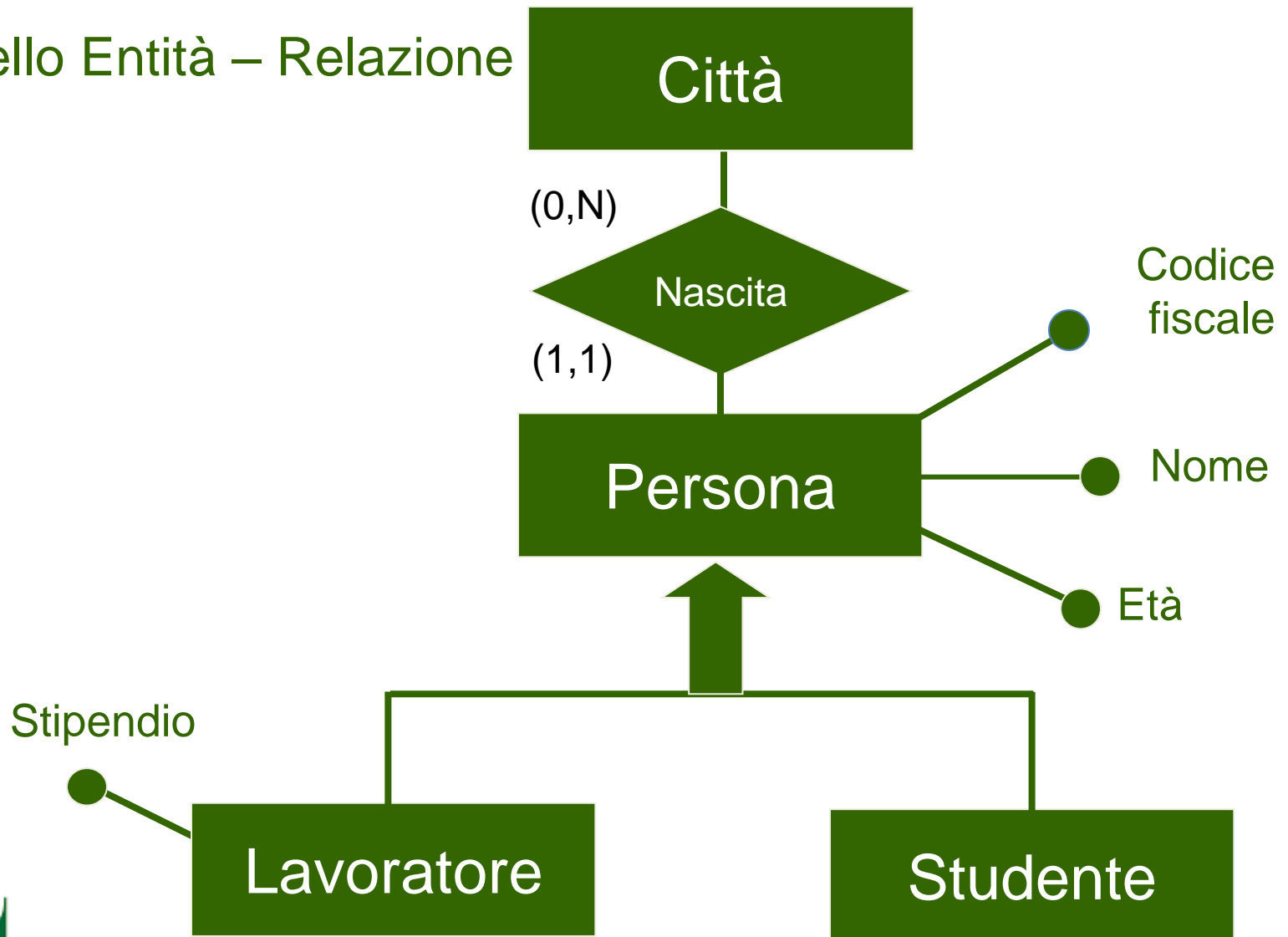
Se E (padre) è generalizzazione di E1, E2, ..., En (figlie):

- ogni proprietà di E è significativa per E1, E2, ..., En
- ogni occorrenza di E1, E2, ..., En è occorrenza anche di E

Ereditarietà

- tutte le proprietà (attributi, relationship, altre generalizzazioni) dell'entità padre vengono ereditate dalle entità figlie e non rappresentate esplicitamente.

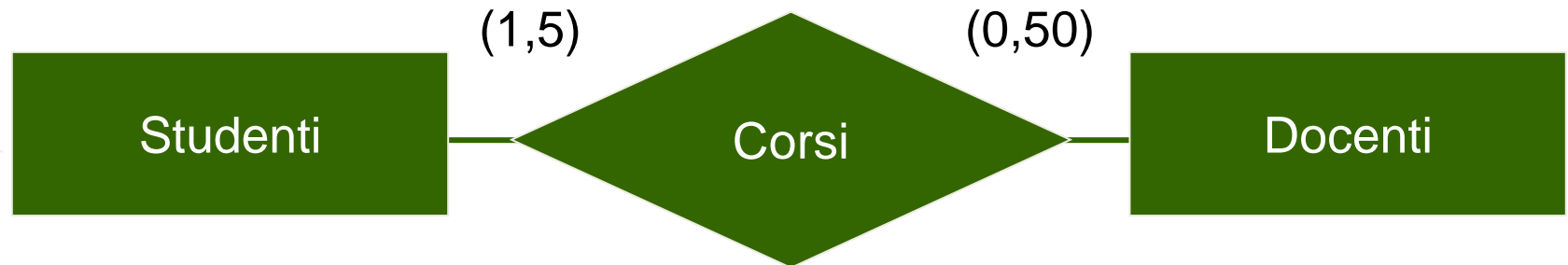
Modello Entità – Relazione



Modello Entità – Relazione (E-R)

La **cardinalità** è una coppia di valori associati a ogni entità che partecipa a una relationship.

Le cardinalità specificano il numero minimo e massimo di occorrenze delle relationship cui ciascuna occorrenza di una entità può partecipare



Modello Entità – Relazione (E-R)

Per semplicità vengono utilizzati solamente tre simboli:

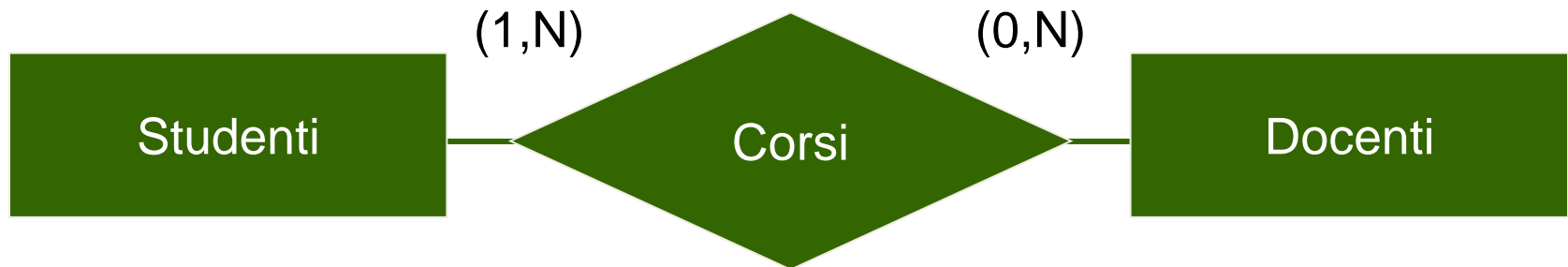
0 e 1 per la cardinalità minima:

0 = “partecipazione opzionale”

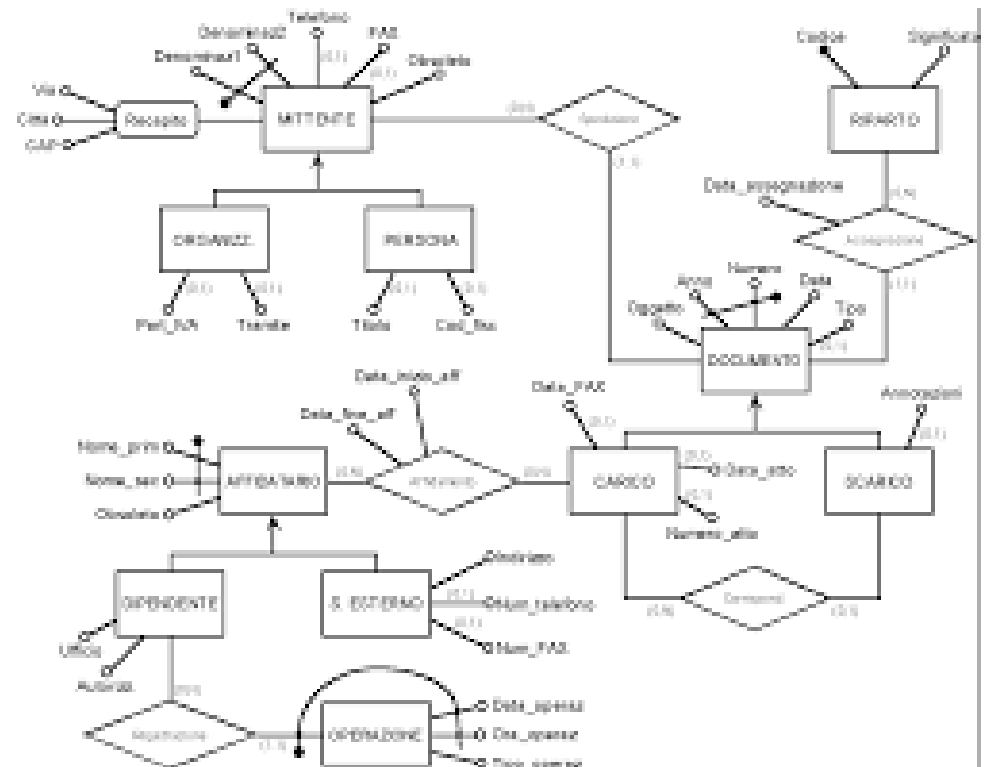
1 = “partecipazione obbligatoria”

1 e “N” per la massima:

“N” non pone alcun limite

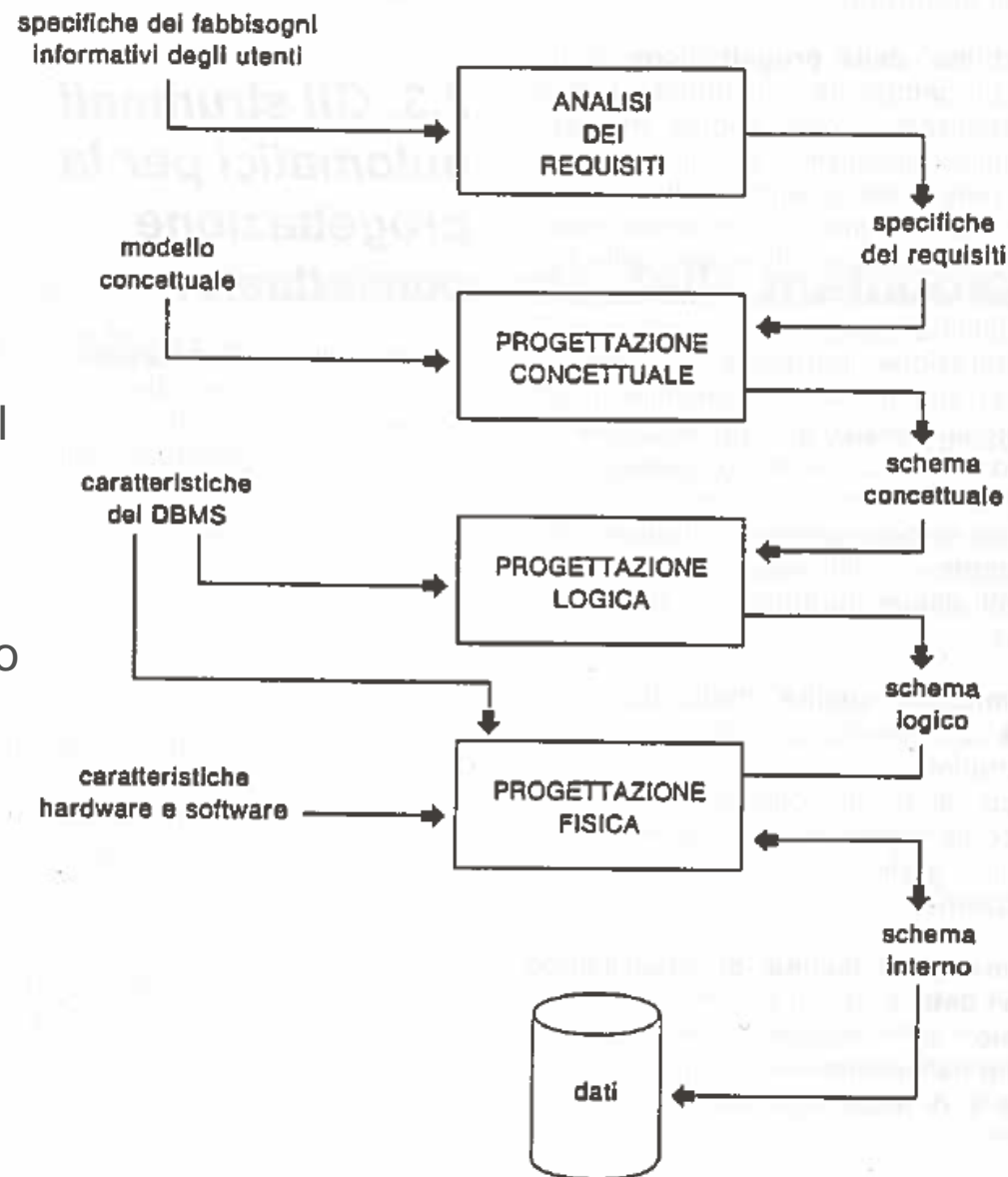


Guardando le cardinalità massime coinvolte nella relationship è possibile stabilire il tipo di relazione (uno-uno, uno-molti, molti-molti).



La progettazione

La **progettazione logica** della base di dati consiste nella traduzione dello schema concettuale dei dati in uno schema logico che rispecchia il modello dei dati scelto, cioè, nel nostro caso, il modello relazionale. Lo schema logico risultante è indipendente dallo specifico DBMS che verrà scelto al termine della progettazione logica. Inoltre vengono definiti i vincoli di integrità sui dati. Infine durante la progettazione logica si definiscono eventuali schemi esterni (viste) per le specifiche applicazioni.

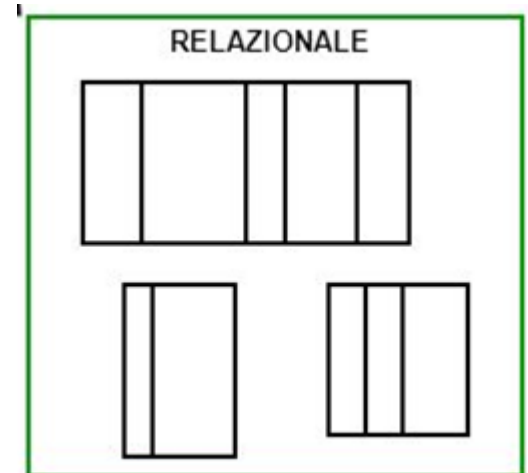
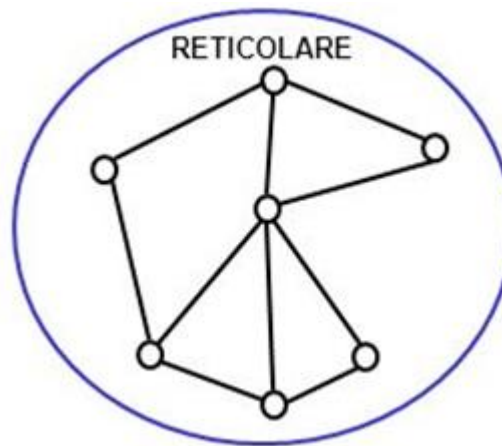
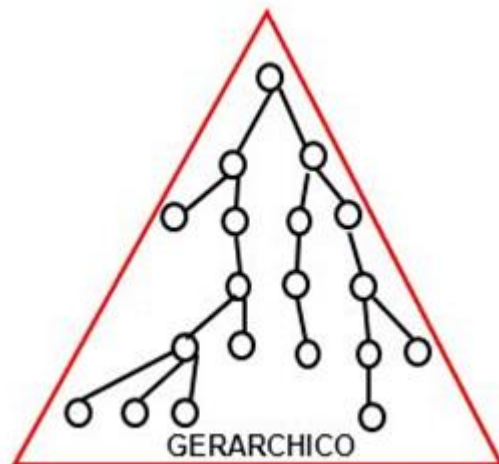


Progettazione logica

I modelli logici

I modelli logici forniscono le regole di derivazione che permettono di passare dalla fase di livello concettuale al quella del livello realizzazione di un database, cioè per trasformare lo schema E-R nello schema delle relazioni.

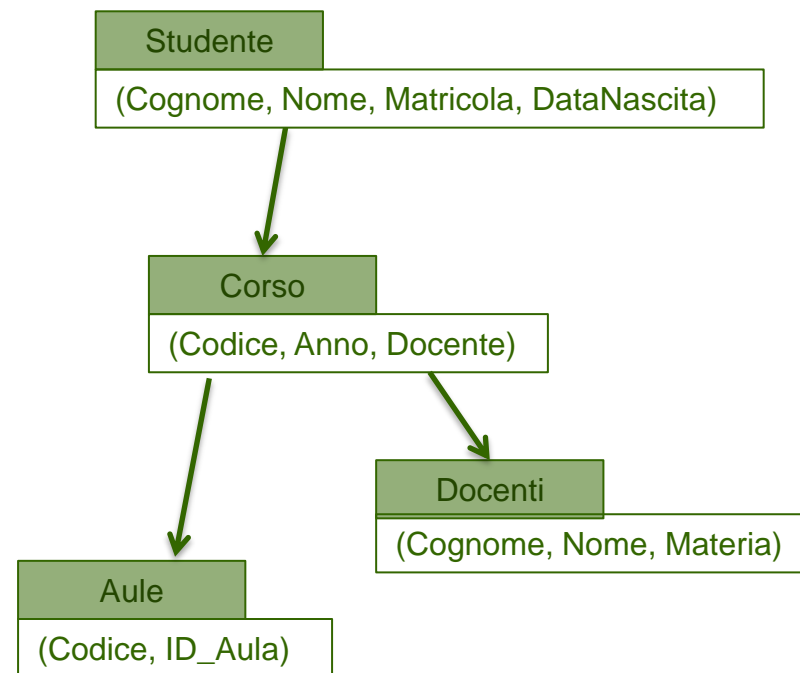
I modelli logici più utilizzati presentano tre strutture di relazione, e sono:



Il modello gerarchico

La struttura ad albero che caratterizza il modello gerarchico si basa sulla possibilità di individuare un segmento principale, il padre o la radice, dal quale dipendono n segmenti figli, che a loro volta si trasformano in padri per altri figli e così via.

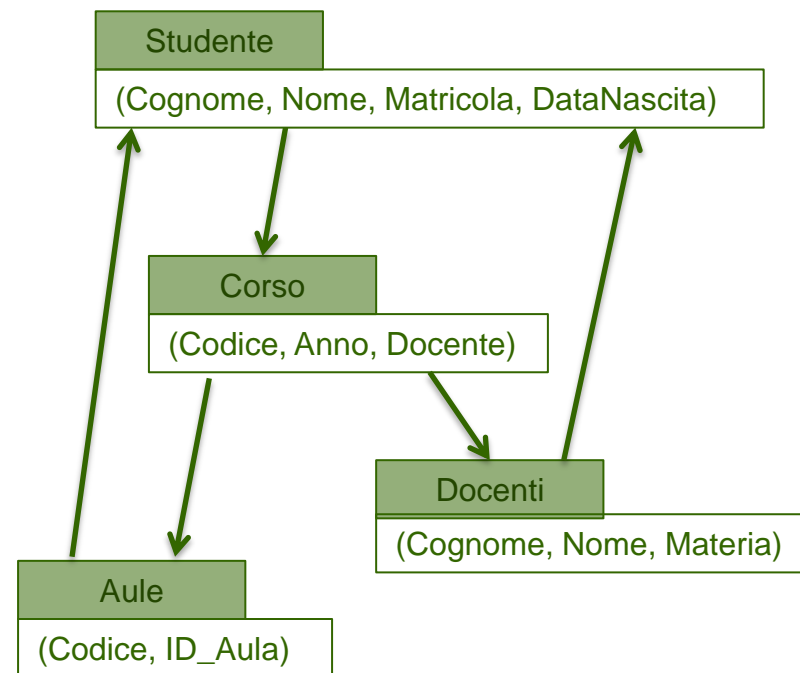
A questi, in virtù della totale dipendenza dal padre, è possibile fare riferimento solo attraverso il passaggio dal nodo principale. Non è possibile dal figlio risalire al padre.



Il modello reticolare

A differenza del modello gerarchico, nel modello reticolare ad ogni record può essere associato un numero qualsiasi di record subordinati e di record precedenti e le correlazioni vengono espresse attraverso record particolari, chiamati record di collegamento (member), che formano delle catene tra le varie parti del sistema.

Le strutture utilizzate nel modello reticolare sono due, il record (si pensi ai comuni file) e il set, che permette di correlare i record, per mezzo di catene di puntatori. Dunque una base di dati reticolare è definita con riferimento ad uno schema, che contiene tipi record collegati fra loro da tipo set.



Il modello relazionale

È stato proposto nel 1970 da Edgar F. Codd nel suo articolo *A Relational Model for Large Shared Data Banks* apparso sulla rivista Communications of the ACM.

Codd metteva in risalto i limiti dei modelli utilizzati in quegli anni, ovvero i modelli reticolare e gerarchico, e in particolare il fatto che tali modelli non distinguessero il livello logico e quello fisico dei dati. Tali modelli, infatti, operavano attraverso puntatori e l'accesso ai dati avveniva sfruttando la rappresentazione fisica dei dati. Le difficoltà emergevano vuoi per una variazione nella struttura fisica, ad esempio per obsolescenza della macchina, vuoi nella copia dei dati da un elaboratore ad un altro.

Il modello relazionale opera una distinzione tra livello fisico e livello logico, ottenendo l'indipendenza dall'organizzazione fisica dei dati.



Il modello relazionale

Nel modello relazionale l'accesso ai dati avviene a livello logico astraendo dalla loro rappresentazione fisica.

Nonostante ciò il successo del modello relazionale è stato lento, in quanto la proprietà di indipendenza fisica dei dati rese difficile una implementazione efficiente delle basi di dati relazionali. Solo dagli anni '80 le basi di dati relazionali sono diventate una realtà e oggi è il modello più diffuso.

Il modello relazionale può essere suddiviso in due componenti principali:

- le **strutture** che permettono di organizzare i dati;
- i **vincoli di integrità** che permettono di inserire solo dati corretti per la realtà modellata.



Il modello relazionale

Il modello relazionale si fonda sul concetto di relazione (da non confondere con la relazione concettuale del modello ER), la cui rappresentazione è una tabella.

Mentre il concetto di tabella è intuitivo ed è usato in vari contesti che prescindono dalle basi di dati il concetto di relazione è formale e proviene dalla teoria degli insiemi, una parte della matematica.



Il modello relazionale

- relazione matematica: come nella teoria degli insiemi
- relazione (dall'inglese relationship) che rappresenta una classe di fatti, nel modello Entity-Relationship; tradotto anche con associazione o correlazione
- relazione secondo il modello relazionale dei dati



Il modello relazionale

Relazione matematica

- D_1, \dots, D_n
- **prodotto cartesiano** $D_1 \times \dots \times D_n$:
 - l'insieme di tutte le n -uple (d_1, \dots, d_n) tali che $d_1 \in D_1, \dots, d_n \in D_n$
 -
- **relazione matematica** su D_1, \dots, D_n :
 - un sottoinsieme di $D_1 \times \dots \times D_n$.
- D_1, \dots, D_n sono i **domini** della relazione



Il modello relazionale

Relazione matematica: un esempio

- $D1=\{1,2,4\}$
- $D2=\{a,b\}$
- prodotto cartesiano $D1 \times D2$

1	a
1	b
2	a
2	b
4	a
4	b

1	b
2	b
4	b

• una relazione $r \subseteq D1 \times D2$

Il modello relazionale

Relazione matematica: proprietà

- una relazione matematica è un insieme di n -uple ordinate:
 - (d_1, \dots, d_n) tali che $d_1 \in D_1, \dots, d_n \in D_n$
- una relazione è un insieme; quindi:
 - 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



Il modello relazionale

Partite \subseteq string \times string \times int \times int

Juventus	Palermo	0	1
Lazio	Milan	2	0
Juventus	Roma	1	0
Roma	Palermo	1	2

La struttura è posizionale.

A ciascun dominio si associa un nome (attributo), che ne descrive il "ruolo"

Casa	Ospite	RetiCasa	RetiOspite
Juventus	Palermo	0	1
Lazio	Milan	2	0
Juventus	Roma	1	0
Roma	Palermo	1	2



Il modello relazionale

Casa	Ospite	RetiCasa	RetiOspite
Juventus	Palermo	0	1
Lazio	Milan	2	0
Juventus	Roma	1	0
Roma	Palermo	1	2

- Una tabella rappresenta una relazione se
 - i valori di ogni colonna sono fra loro omogenei
 - le righe sono diverse fra loro
 - le intestazioni delle colonne sono diverse tra loro
- In una tabella che rappresenta una relazione
 - l'ordinamento tra le righe è irrilevante
 - l'ordinamento tra le colonne è irrilevante



Il modello relazionale

Il modello è basato su valori, ovvero i riferimenti fra dati in relazioni diverse sono rappresentati per mezzo di valori dei domini che compaiono nelle ennuple (tuple).

Struttura basata su valori: vantaggi

- indipendenza dalle strutture fisiche che possono cambiare dinamicamente
- si rappresenta solo ciò che è rilevante dal punto di vista dell'applicazione
- l'utente finale vede gli stessi dati dei programmatori
- i dati sono portabili più facilmente da un sistema ad un altro



Il modello relazionale

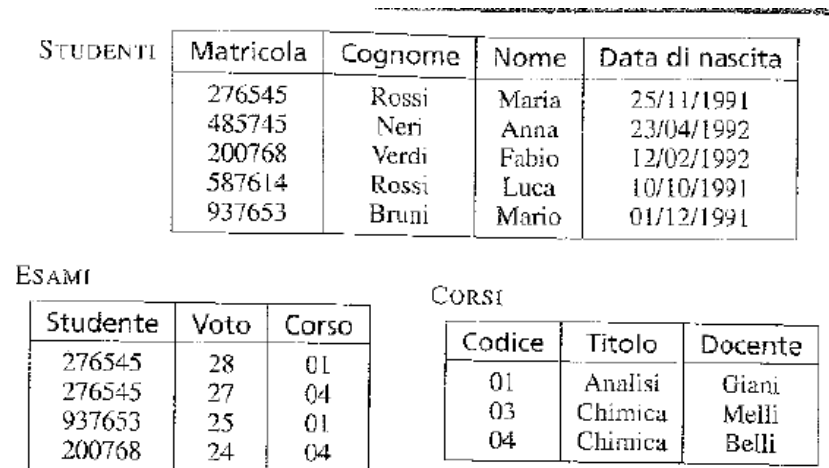


Figura 2.6 Una base di dati relazionale.

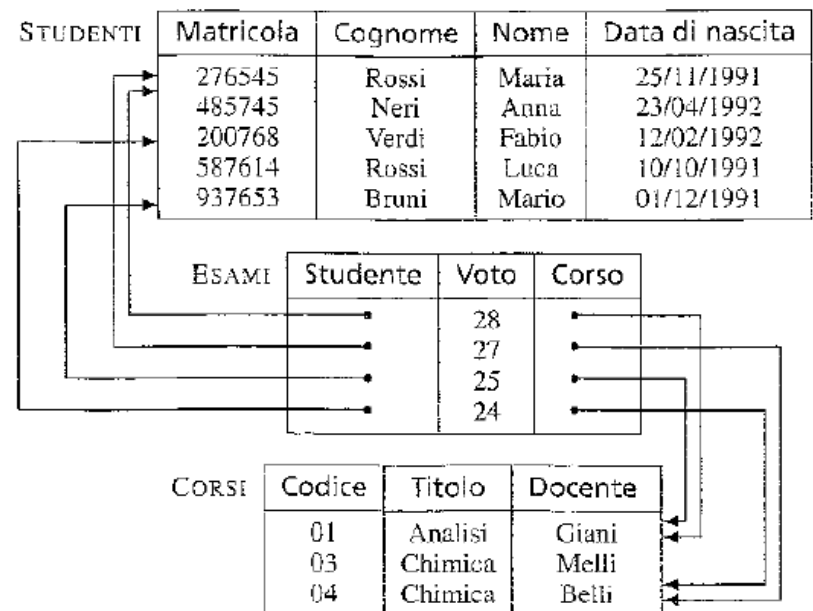


Figura 2.7 Una base di dati con puntatori.

La progettazione

La **progettazione fisica** consiste nella traduzione dello schema logico nel cosiddetto schema interno, cioè in una rappresentazione che tenga conto delle particolari caratteristiche hardware e software del sistema informatico esistente.

