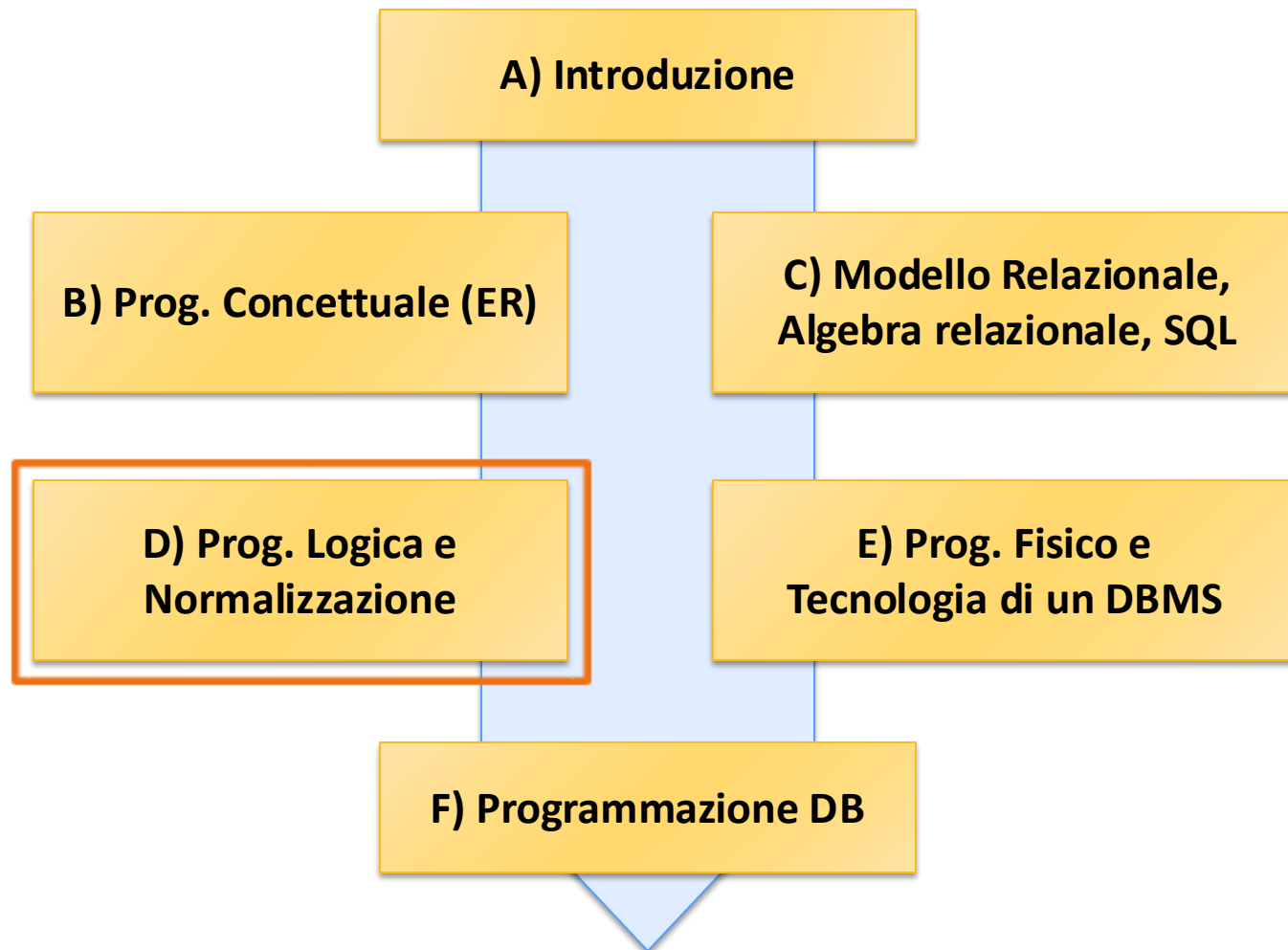


Basi di Dati

Progetto Logico Relazionale (Parte 1)

Basi di Dati – Dove ci troviamo?



Progetto logico

- ▶ Lo schema **E/R** descrive un **dominio applicativo** ad un dato livello di astrazione
- ▶ Lo schema E/R è molto utile per:
 - ▶ fornire una **descrizione** sintetica e visiva
 - ▶ rappresentare buona parte della **semantica** dell'applicazione
 - ▶ scambiare **informazioni** sull'attività progettuale tra i membri del team di progetto e mantenere una **documentazione**

Progetto logico

- ▶ Non esistono **DBMS** in grado di operare direttamente sui concetti di schemi E/R
 - ▶ è quindi necessario tradurli in **altri schemi** di dati (**logico relazionale** in queste lezioni)
 - ▶ questa traduzione può essere eseguita in modo **semi-automatico**
 - ▶ le **scelte alternative** devono tenere conto dell'efficienza dello **schema logico** risultante e delle operazioni da effettuare (derivanti da flussi e processi)

Progettazione base di dati

Requisiti della base di dati

**Progettazione
concettuale**

Schema concettuale

**Progettazione
logica**

Schema logico

**Progettazione
fisica**

Schema fisico

Obiettivo

"tradurre" lo schema concettuale in uno schema logico che rappresenti gli stessi dati in maniera corretta ed efficiente

Dati di ingresso e uscita

▶ Ingresso:

- ▶ schema concettuale
- ▶ informazioni sul carico applicativo
- ▶ modello logico

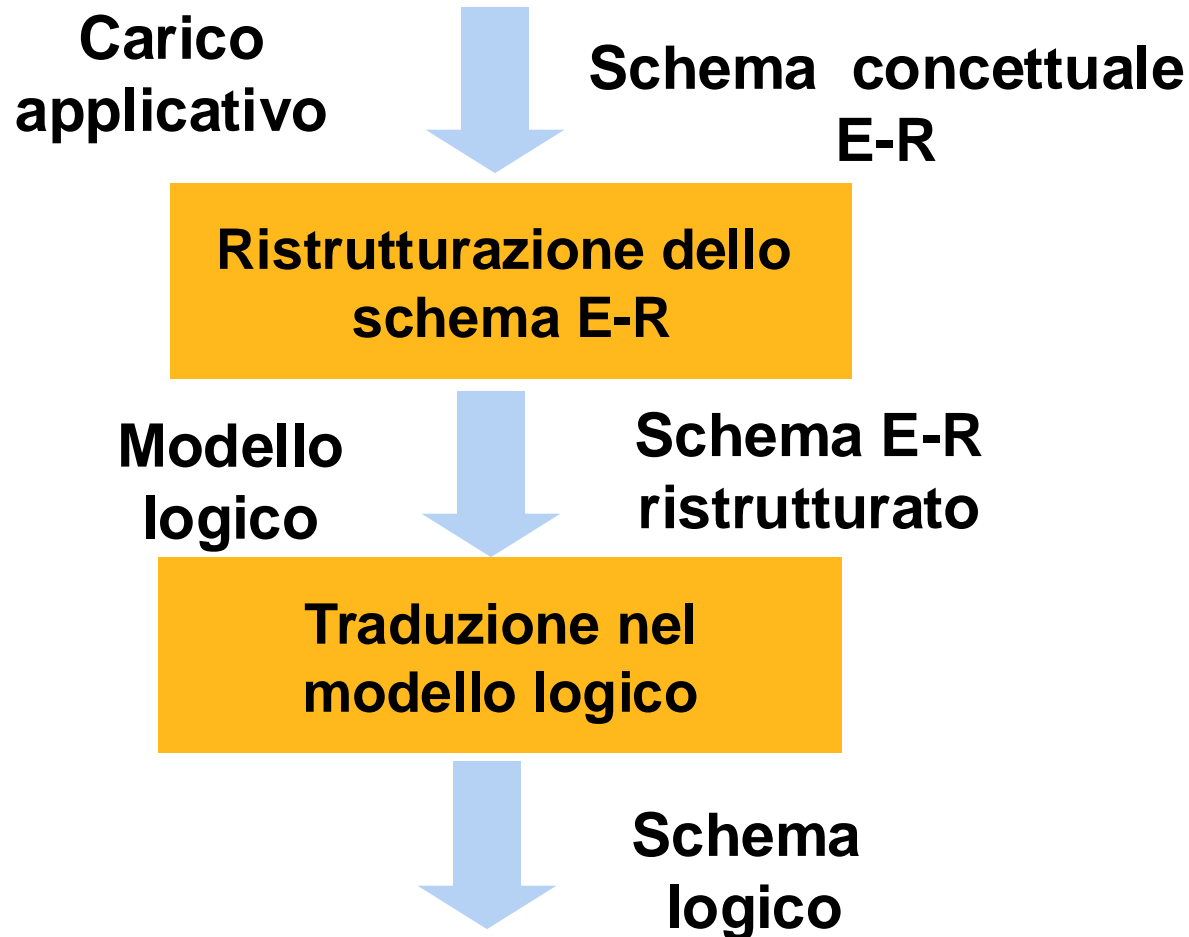
▶ Uscita:

- ▶ schema logico
- ▶ documentazione associata

Non si tratta di una pura e semplice traduzione

- ▶ alcuni aspetti non sono direttamente rappresentabili
- ▶ è necessario considerare le prestazioni

Fasi della progettazione logica



Ristrutturazione schema E-R

- ▶ Motivazioni:

- ▶ semplificare la traduzione
- ▶ "ottimizzare" le prestazioni

- ▶ Osservazione:

- ▶ uno schema E-R ristrutturato non è (più) uno schema concettuale nel senso stretto del termine

Prestazioni?

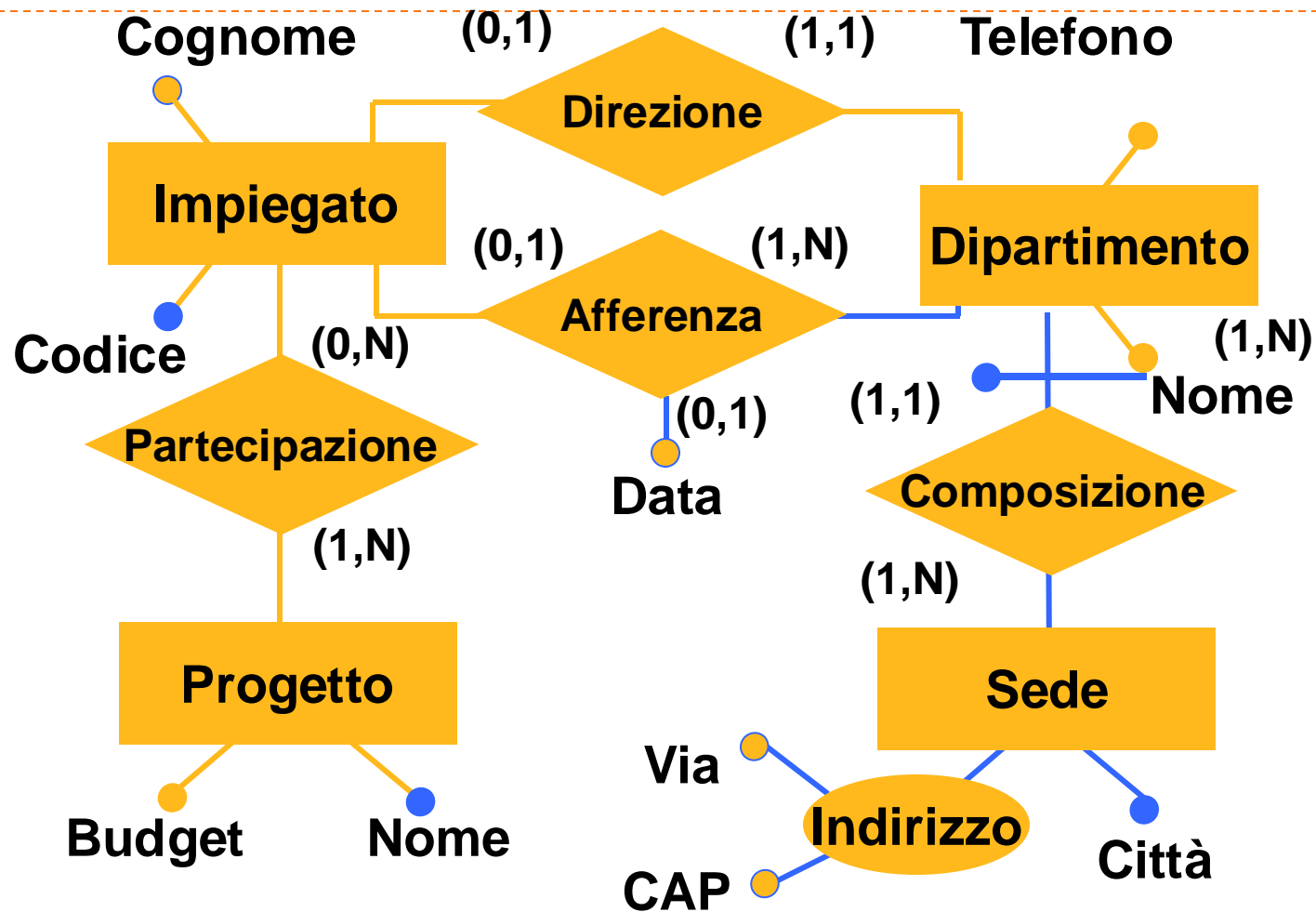
- ▶ Per ottimizzare il risultato abbiamo bisogno di analizzare le prestazioni a questo livello
- ▶ Ma:
 - ▶ le prestazioni non sono valutabili con precisione su uno schema concettuale!

Prestazioni, approssimate

- ▶ Consideriamo:
 - ▶ “indicatori” dei parametri che regolano le prestazioni
- ▶ spazio:
 - ▶ numero di occorrenze previste
- ▶ tempo:
 - ▶ numero di occorrenze (di entità e relationship) visitate durante un'operazione

Un esempio di calcolo di prestazioni

Schema E-R



Un esempio di calcolo di prestazioni :

Tavola dei volumi

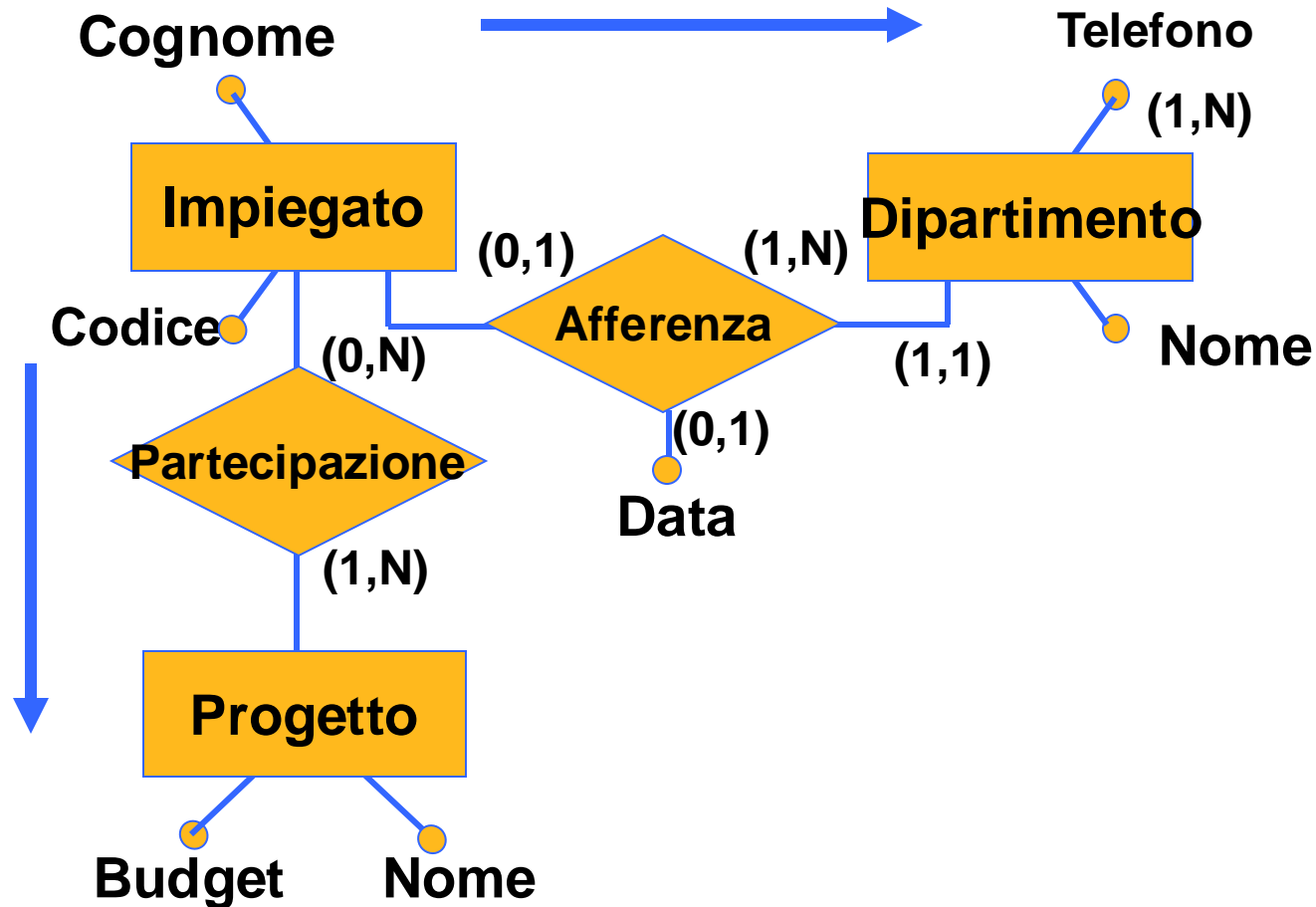
Concetto	Tipo	Volume
Sede	E	10
Dipartimento	E	80
Impiegato	E	2000
Progetto	E	500
Composizione	R	80
Afferenza	R	1900
Direzione	R	80
Partecipazione	R	6000

Un esempio di calcolo di prestazioni:

Esempio di valutazione di costo

- ▶ Operazione:
 - ▶ trova tutti i dati di un impiegato, del dipartimento nel quale lavora e dei progetti ai quali partecipa
- ▶ Si costruisce una **tavola degli accessi** basata su uno **schema di navigazione**

Un esempio di calcolo di prestazioni: Schema di navigazione



Un esempio di calcolo di prestazioni: Tavola degli accessi

Concetto	Costrutto	Accessi	Tipo
Impiegato	Entità	1	L
Afferenza	Relationship	1	L
Dipartimento	Entità	1	L
Partecipazione	Relationship	3	L
Progetto	Entità	3	L

Attività della ristrutturazione

- ▶ Analisi delle ridondanze
- ▶ Eliminazione delle generalizzazioni
- ▶ Partizionamento/accorpamento di entità e relationship
- ▶ Scelta degli identificatori primari

Analisi delle ridondanze

- ▶ Una ridondanza in uno schema E-R è una informazione significativa ma derivabile da altre
- ▶ in questa fase si decide se eliminare le ridondanze eventualmente presenti o mantenerle (o anche di introdurne di nuove)

Ridondanze

▶ Vantaggi

- ▶ semplificazione delle interrogazioni

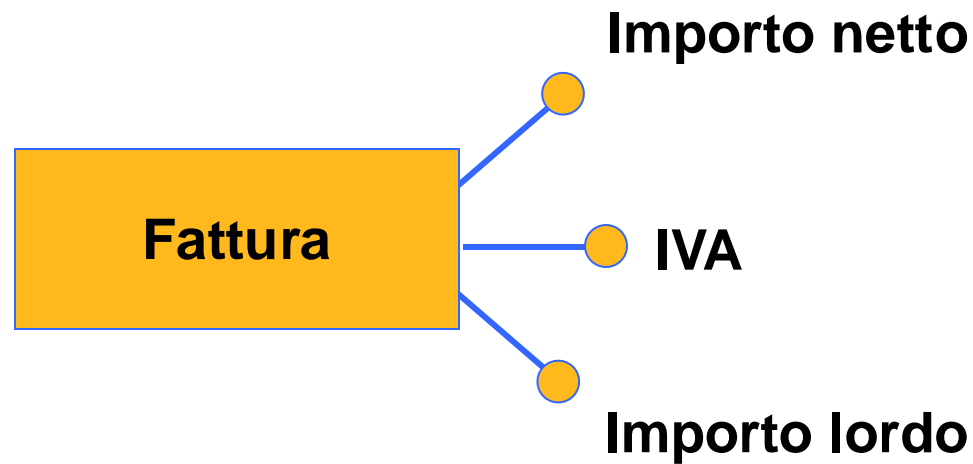
▶ Svantaggi

- ▶ appesantimento degli aggiornamenti
- ▶ maggiore occupazione di spazio

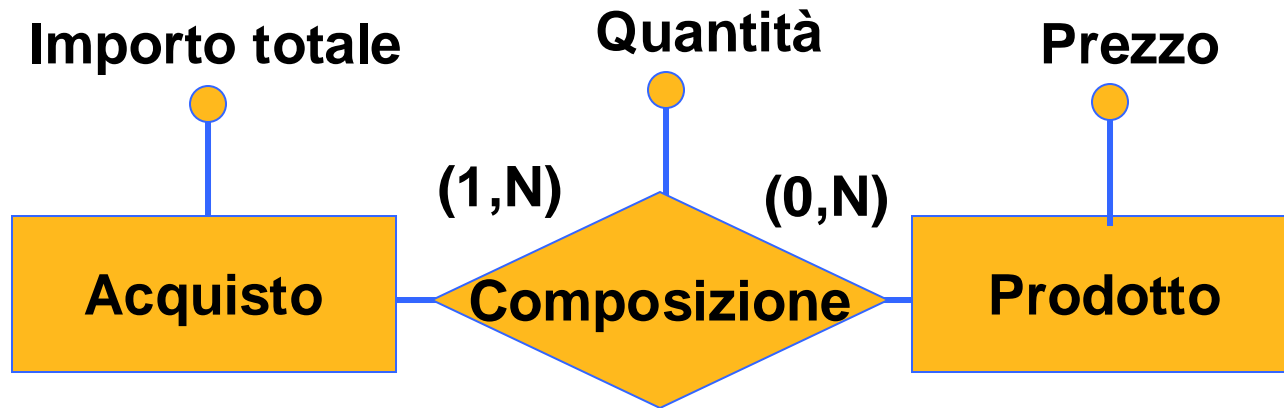
Forme di ridondanza in uno schema E-R

- ▶ attributi derivabili:
 - ▶ da altri attributi della stessa entità (o relationship)
 - ▶ da attributi di altre entità (o relationship)
- ▶ relationship derivabili dalla composizione di altre (più in generale: cicli di relationship)

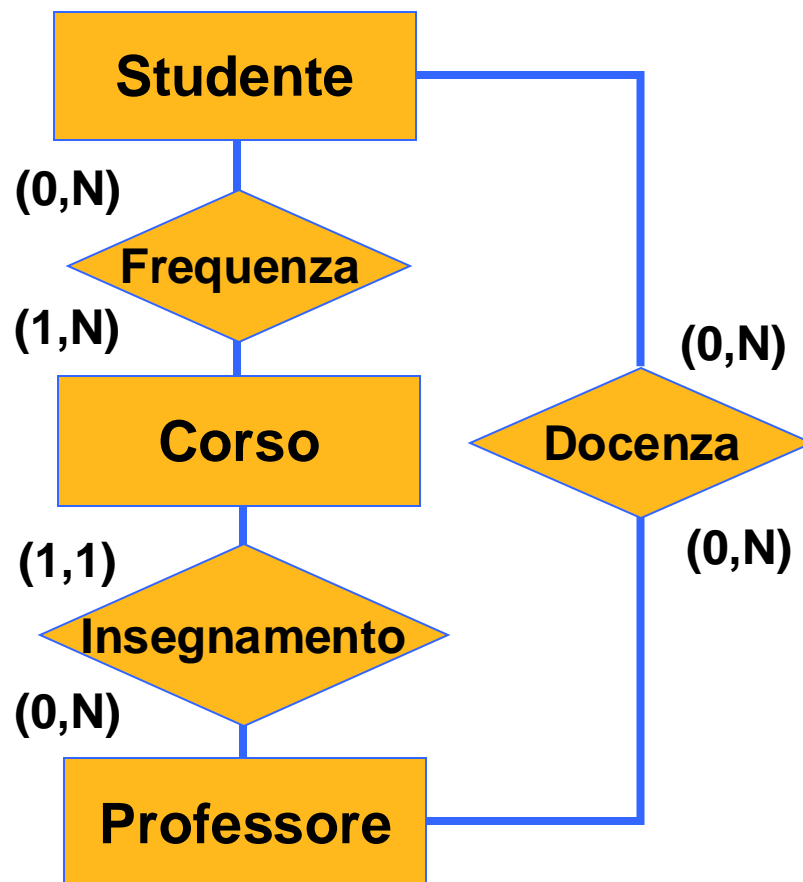
Attributo derivabile dalla stessa entità



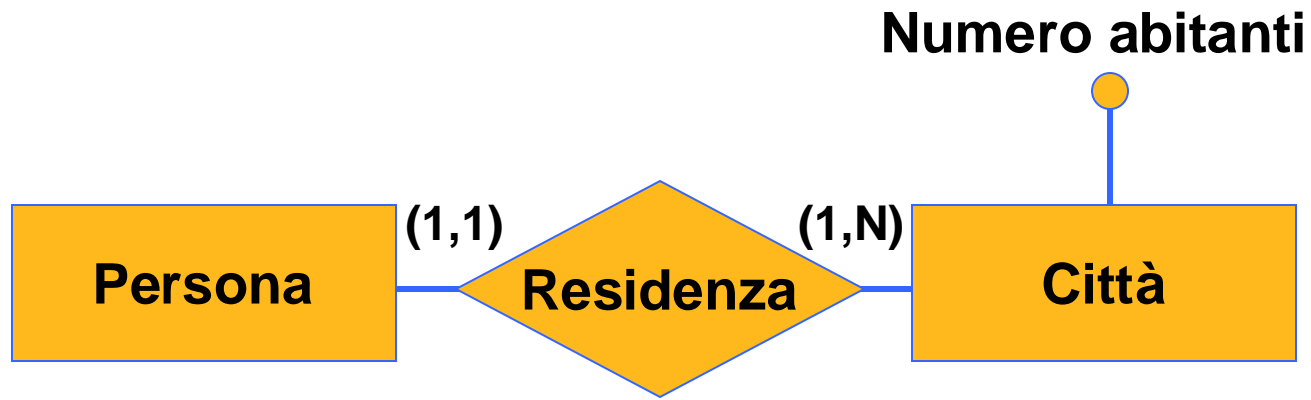
Attributo derivabile da altra entità



Ridondanza dovuta a ciclo



Esempio di analisi di una ridondanza



Esempio di analisi di una ridondanza: Spazio occupato e carico di lavoro

Concetto	Tipo	Volume
Città	E	200
Persona	E	1000000
Residenza	R	1000000

- ▶ **Spazio occupato** da Numero abitanti: 4 byte x 200 istanze = 800 byte
- ▶ **Costo operazioni**
 - ▶ **Operazione 1**: memorizza una nuova persona con la relativa città di residenza (500 volte al giorno)
 - ▶ **Operazione 2**: stampa tutti i dati di una città (incluso il numero di abitanti) (2 volte al giorno)

Esempio di analisi di una ridondanza:

Tavola degli accessi in presenza di ridondanza

Operazione 1

Concetto	Costrutto	Accessi	Tipo
Persona	Entità	1	S
Residenza	Relazione	1	S
Città	Entità	1	L
Città	Entità	1	S

Costo
aggiornamento



Operazione 2

Concetto	Costrutto	Accessi	Tipo
Città	Entità	1	L

Esempio di analisi di una ridondanza:

Tavola degli accessi in assenza di ridondanza

Operazione 1

Concetto	Costrutto	Accessi	Tipo
Persona	Entità	1	S
Residenza	Relazione	1	S

Operazione 2

Concetto	Costrutto	Accessi	Tipo
Città	Entità	1	L
Residenza	Relazione	5000	L

Esempio di analisi di una ridondanza:

Costi di accesso in presenza di ridondanza

- ▶ Costi:
 - ▶ Operazione 1: 1500 accessi in scrittura e 500 accessi in lettura al giorno
 - ▶ Operazione 2: trascurabile.
- ▶ Contiamo doppi gli accessi in scrittura
 - ▶ Totale di 3500 accessi al giorno

Esempio di analisi di una ridondanza:

Costi di accesso in assenza di ridondanza

- ▶ Costi:
 - ▶ Operazione 1: 1000 accessi in scrittura
 - ▶ Operazione 2: 10000 accessi in lettura al giorno
- ▶ Contiamo doppi gli accessi in scrittura
 - ▶ Totale di 12000 accessi al giorno

Esempio di analisi di una ridondanza:

Conclusioni finali

Costi in presenza di ridondanza

- ▶ Costi di accesso:
3500 accessi al giorno
- ▶ Costi di memorizzazione:
800 byte

Costi in assenza di ridondanza

- ▶ #accessi al giorno: 12000

Si decide di mantenere la ridondanza

Attività della ristrutturazione

- ▶ Analisi delle ridondanze
- ▶ Eliminazione delle generalizzazioni
- ▶ Partizionamento/accorpamento di entità e relationship
- ▶ Scelta degli identificatori primari

Eliminazione delle gerarchie

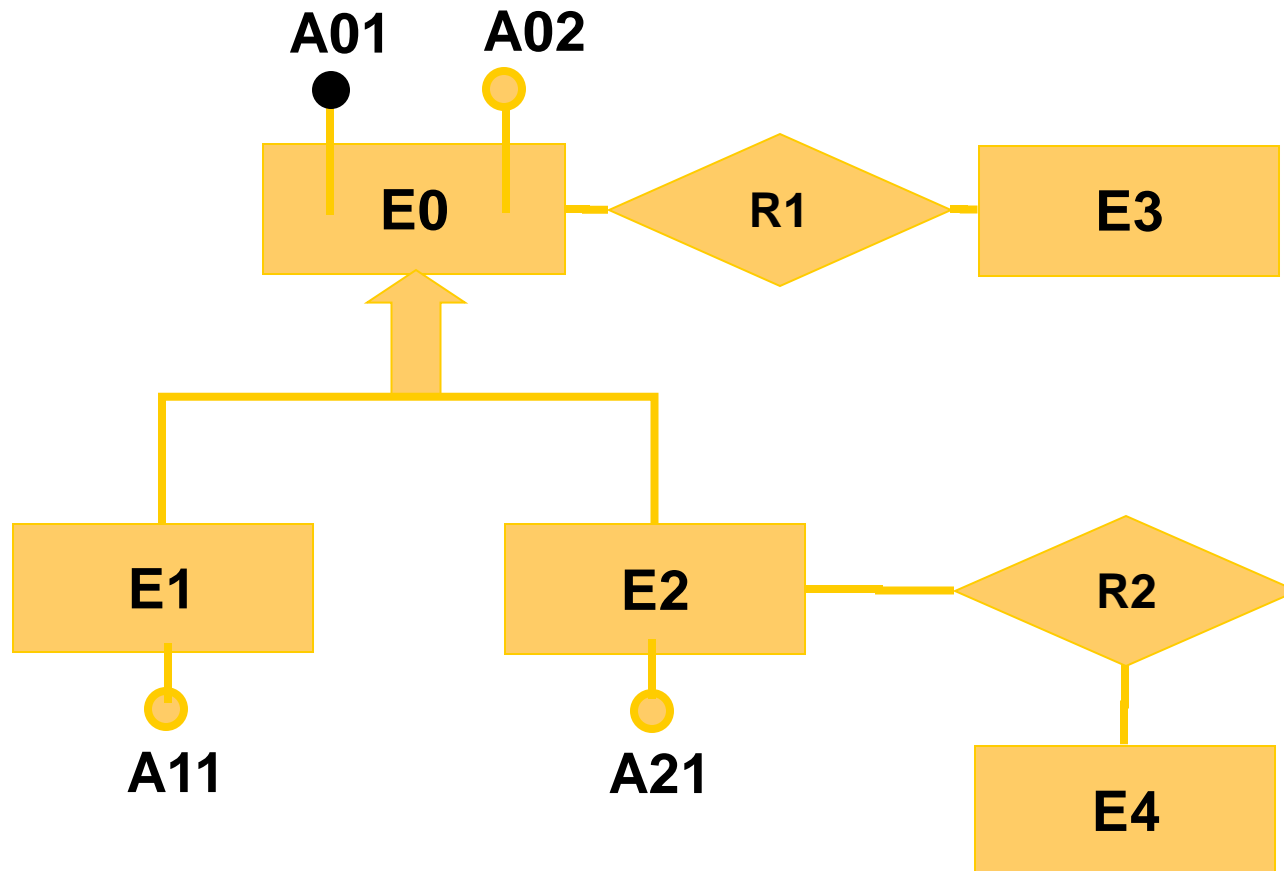
- ▶ il modello relazionale non può rappresentare direttamente le generalizzazioni
- ▶ entità e relationship sono invece direttamente rappresentabili

si eliminano perciò le gerarchie, sostituendole con entità e relationship

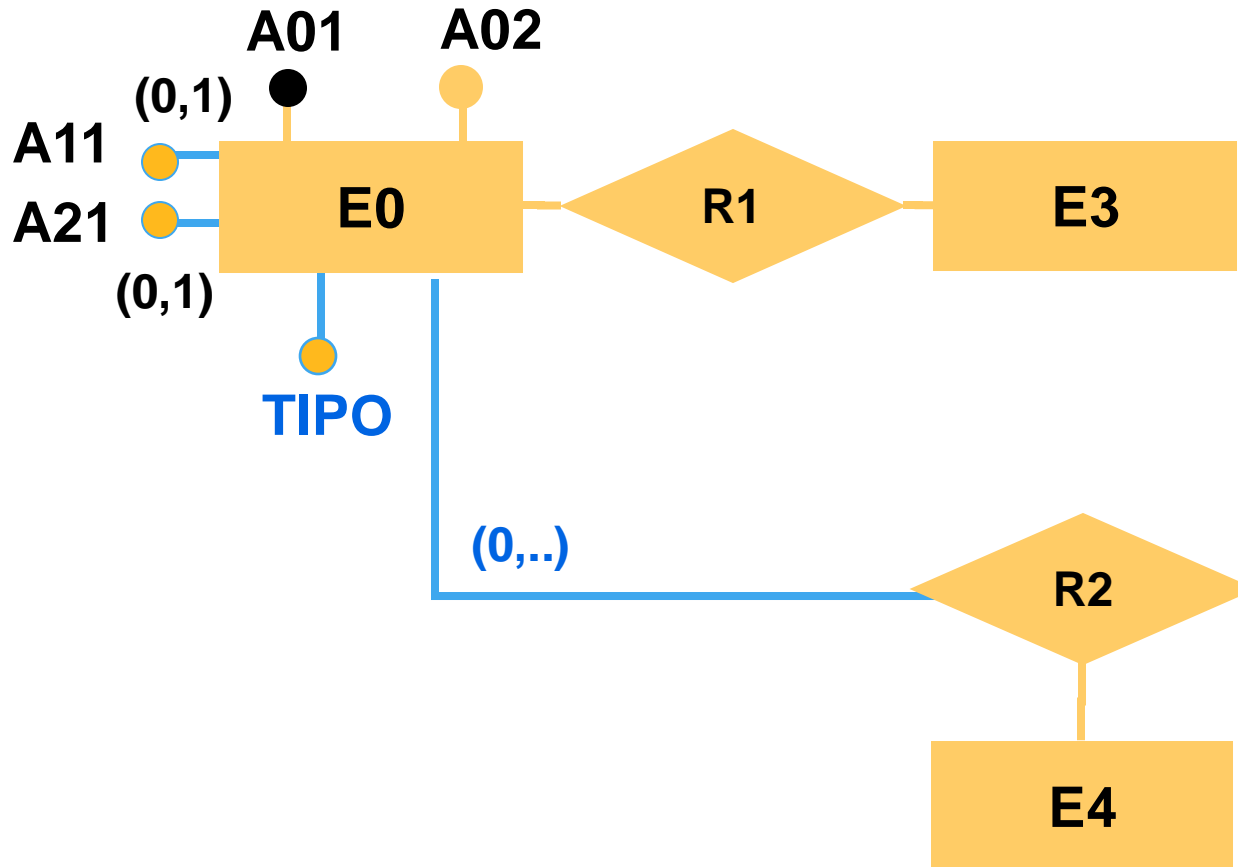
Tre possibilità

1. accorpamento delle figlie della generalizzazione nel genitore
2. accorpamento del genitore della generalizzazione nelle figlie
 - ✓ Possibile solo se la gerarchia è totale
3. sostituzione della generalizzazione con relationship

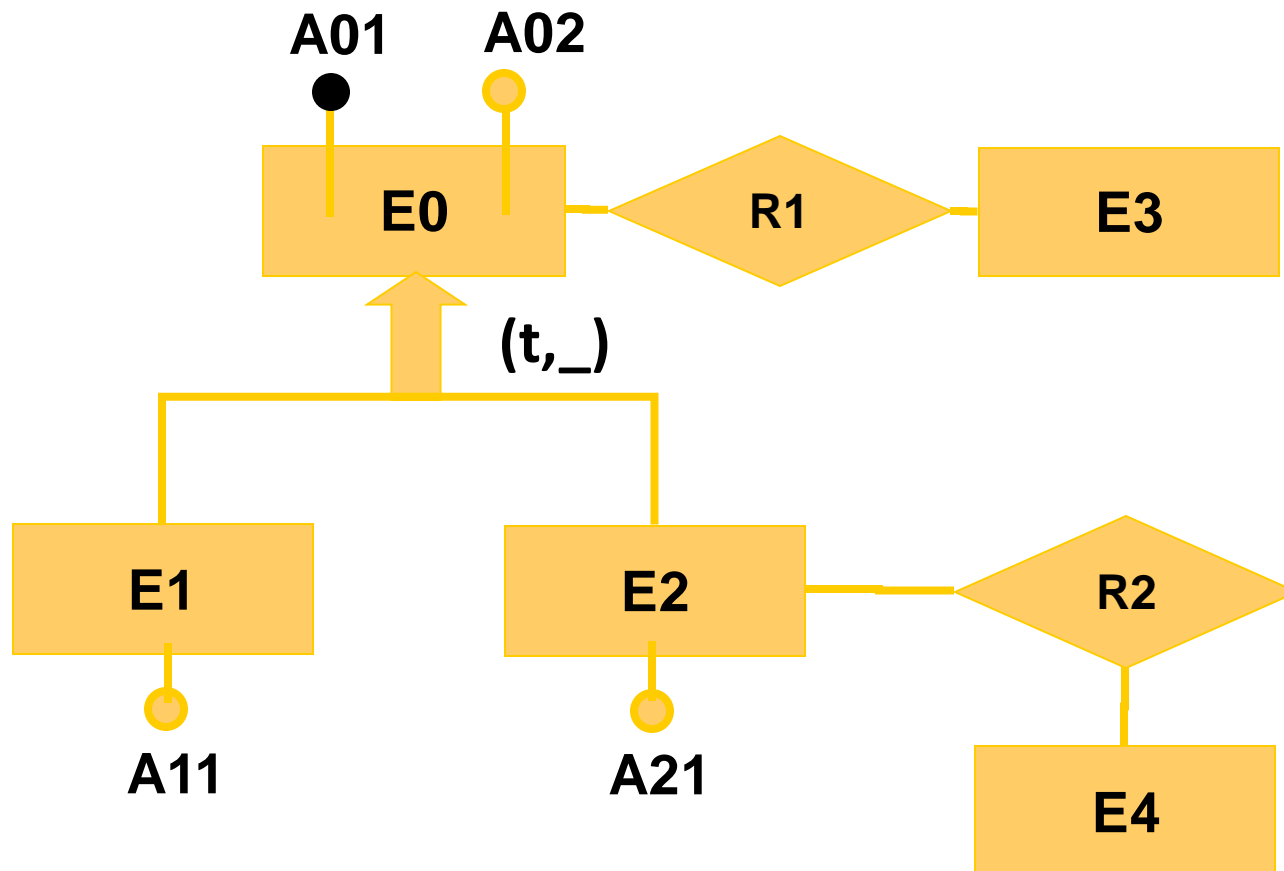
Schema con gerarchia



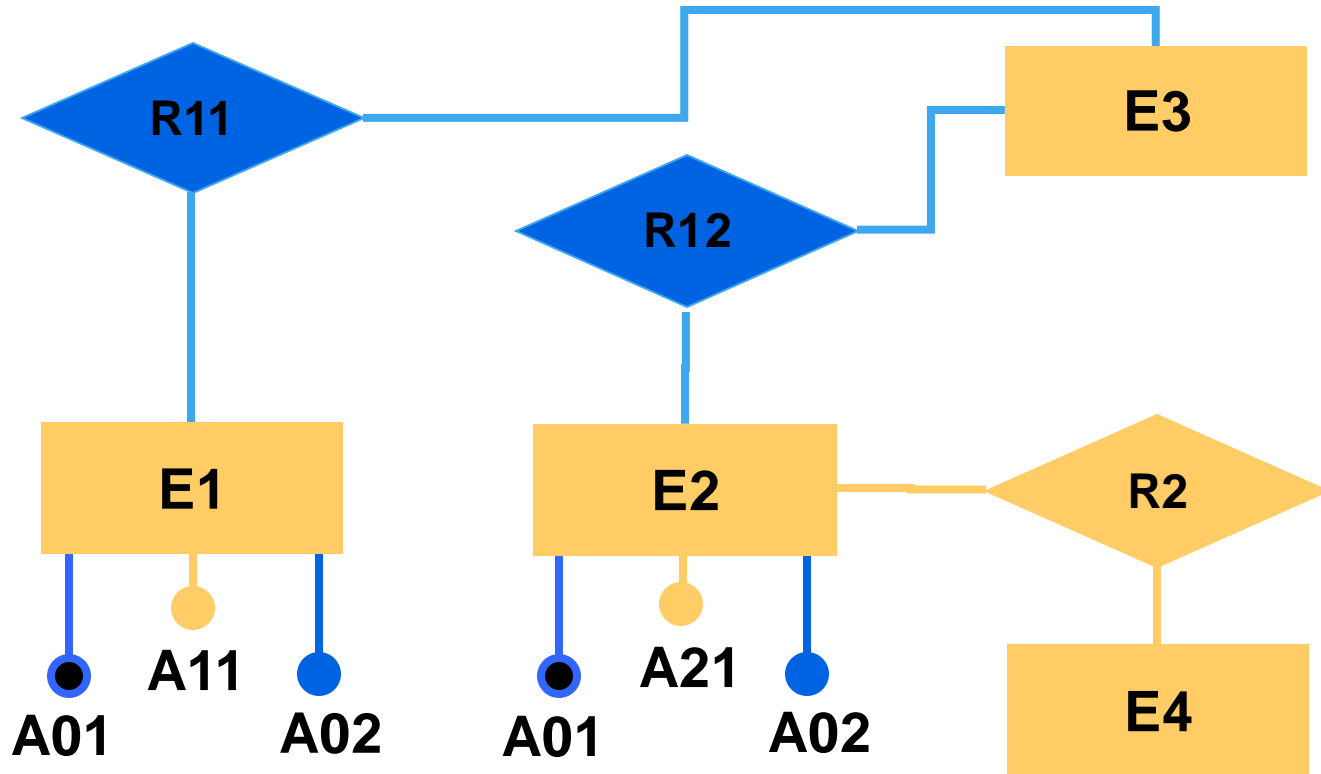
Accorpamento delle figlie nel genitore



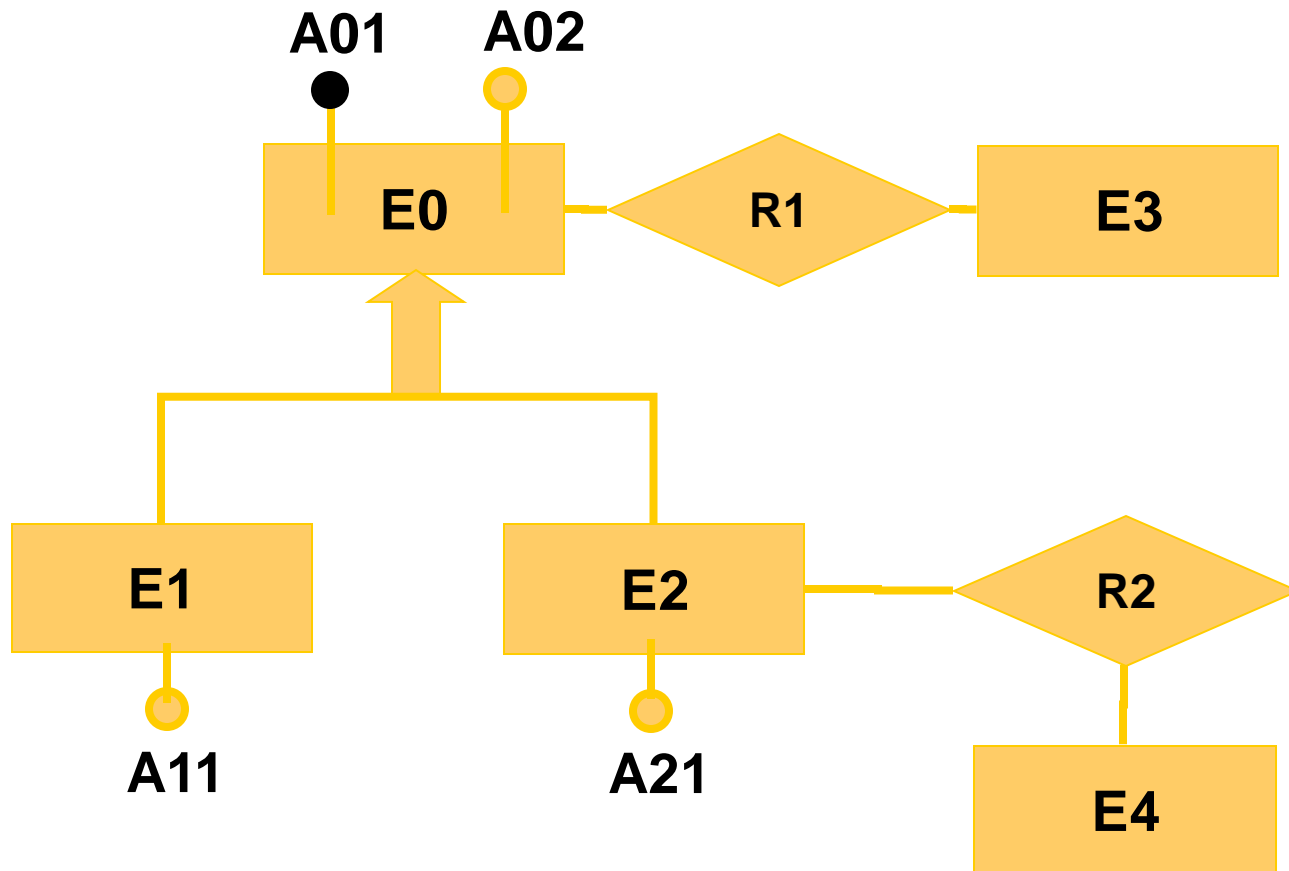
Schema con gerarchia



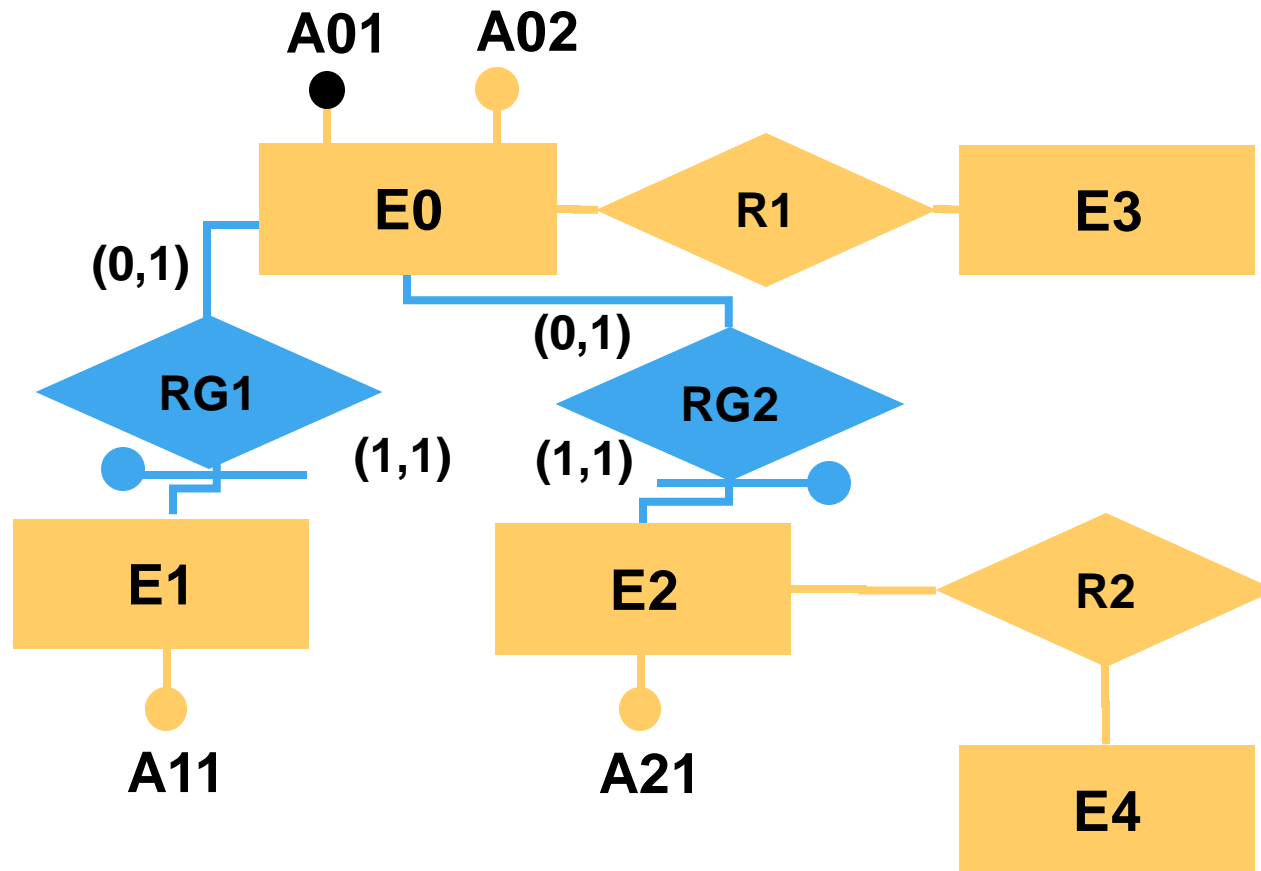
Accorpamento del genitore nelle figlie



Schema con gerarchia



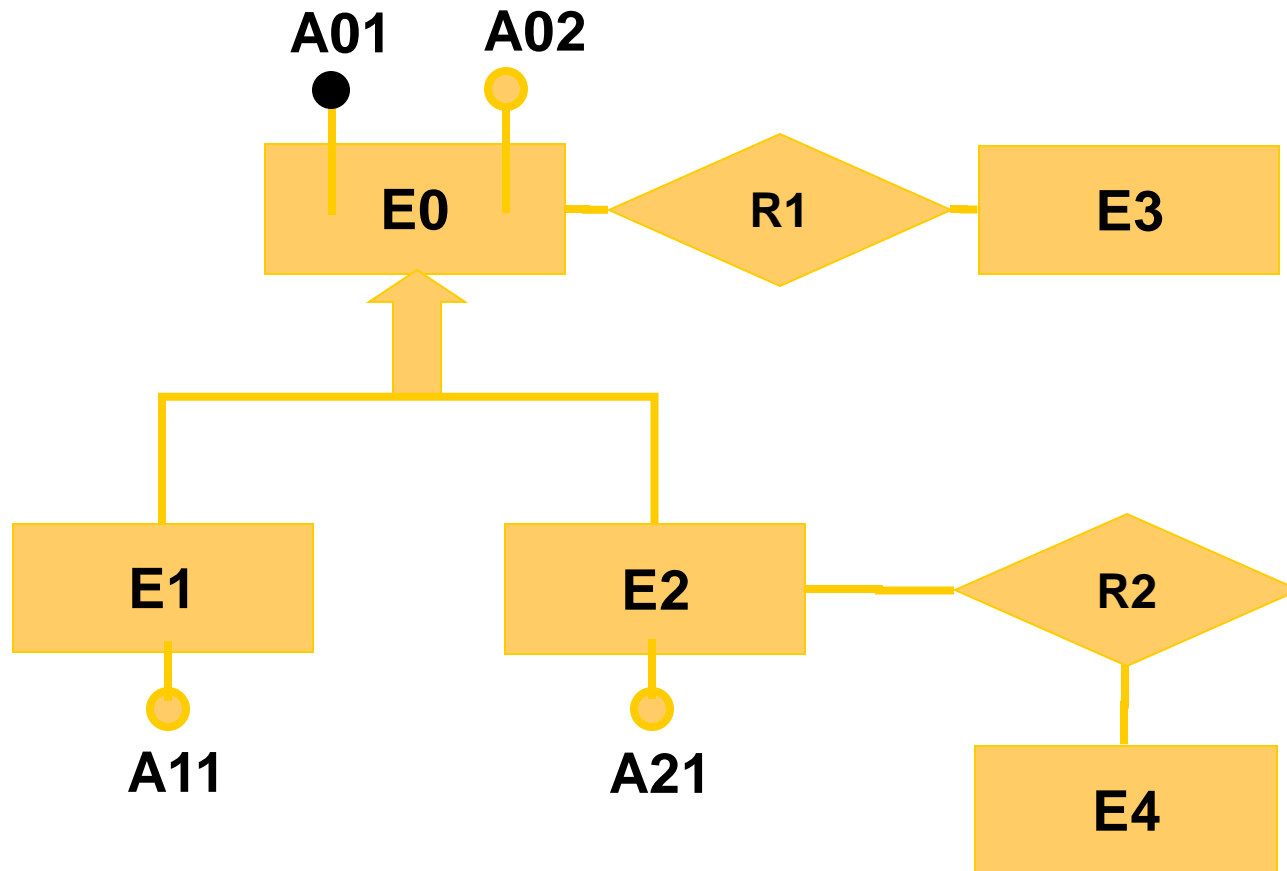
Sostituzione della generalizzazione con relationship



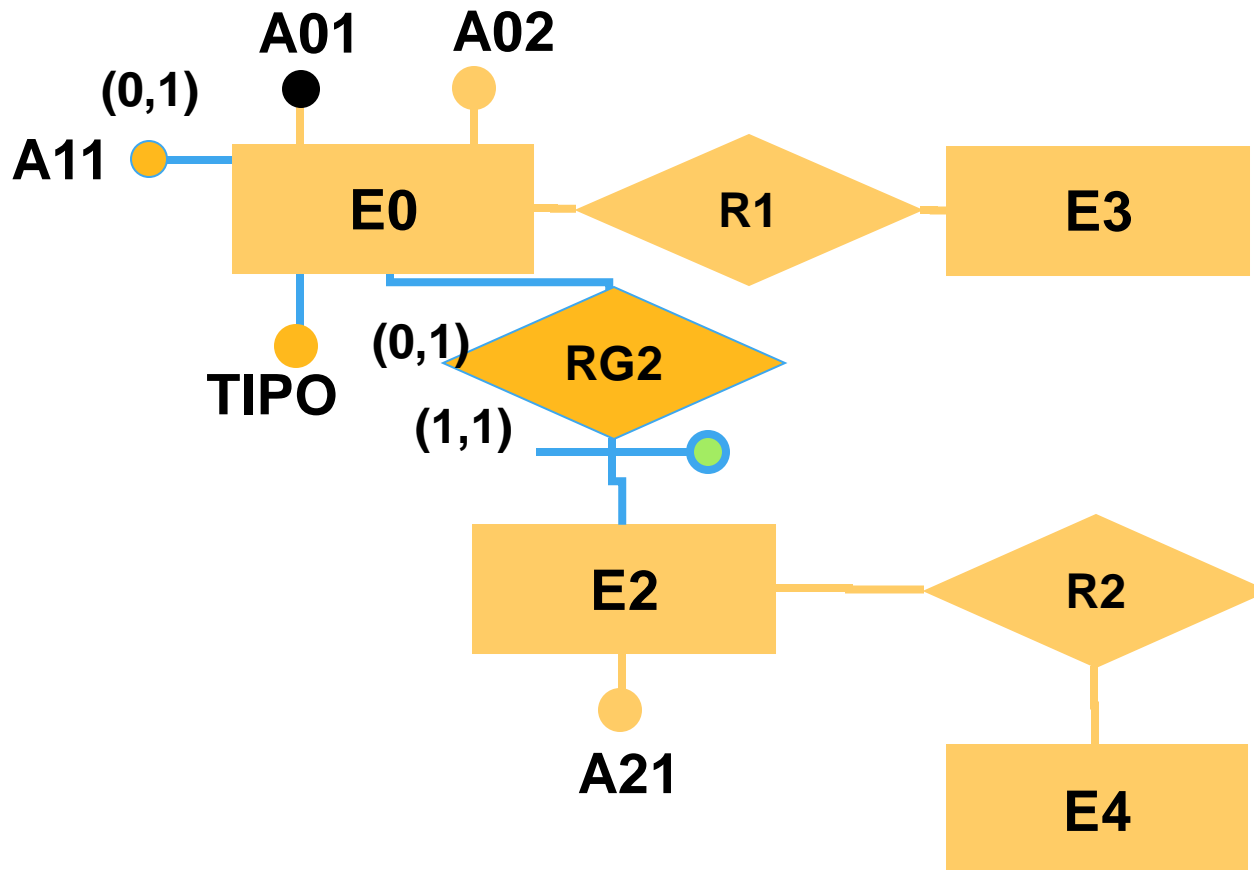
Eliminazione delle gerarchie

- ▶ la scelta fra le alternative si può fare con metodo simile a quello visto per l'analisi delle ridondanze (però non basato solo sul numero degli accessi)
- ▶ è possibile seguire alcune semplici regole generali
 1. conviene se gli accessi al padre e alle figlie sono contestuali
 2. conviene se gli accessi alle figlie sono distinti
 3. conviene se gli accessi alle entità figlie sono separati dagli accessi al padre
- ▶ sono anche possibili soluzioni “ibride”, soprattutto in gerarchie a più livelli

Soluzione ibrida



Soluzione ibrida



Attività della ristrutturazione

- ▶ Analisi delle ridondanze
- ▶ Eliminazione delle generalizzazioni
- ▶ Partizionamento/accorpamento di entità e relationship
- ▶ Scelta degli identificatori primari

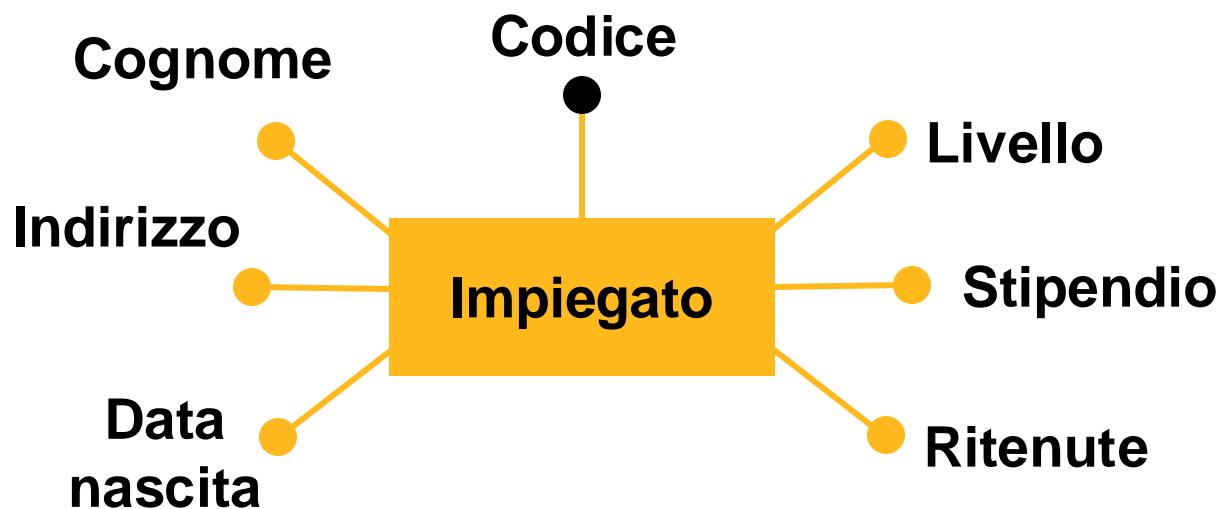
Partizionamento/accorpamento di entità e relationship

- ▶ Ristrutturazioni effettuate per rendere più efficienti le operazioni in base a un semplice principio
- ▶ Gli accessi si riducono:
 - ▶ separando attributi di un concetto che vengono acceduti separatamente
 - ▶ raggruppando attributi di concetti diversi acceduti insieme

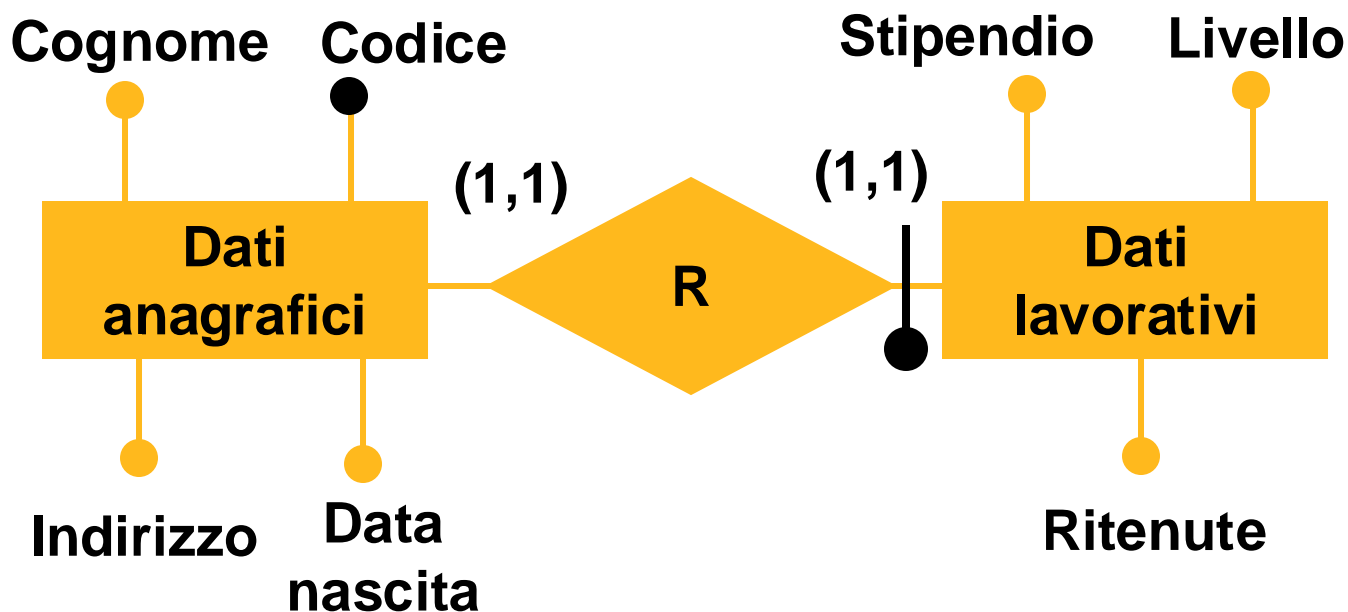
Ristrutturazioni, casi principali

- ▶ partizionamento verticale di entità
- ▶ partizionamento orizzontale di relationship
- ▶ eliminazione di attributi multivalore
- ▶ accorpamento di entità/ relationship

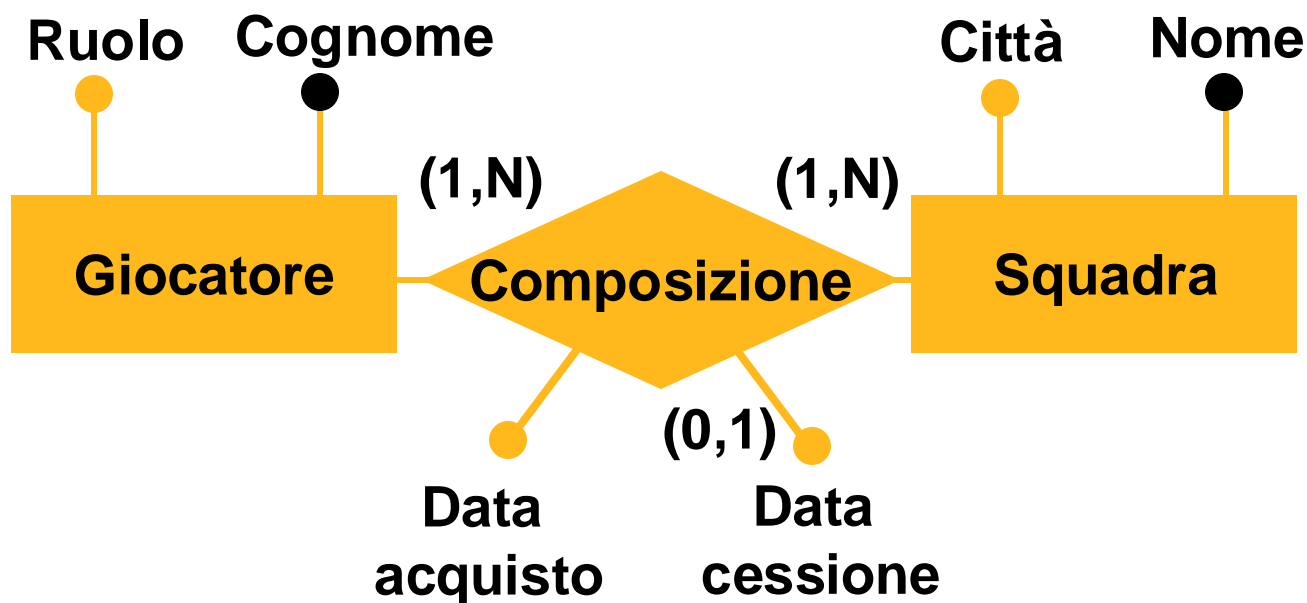
Partizionamento verticale di entità



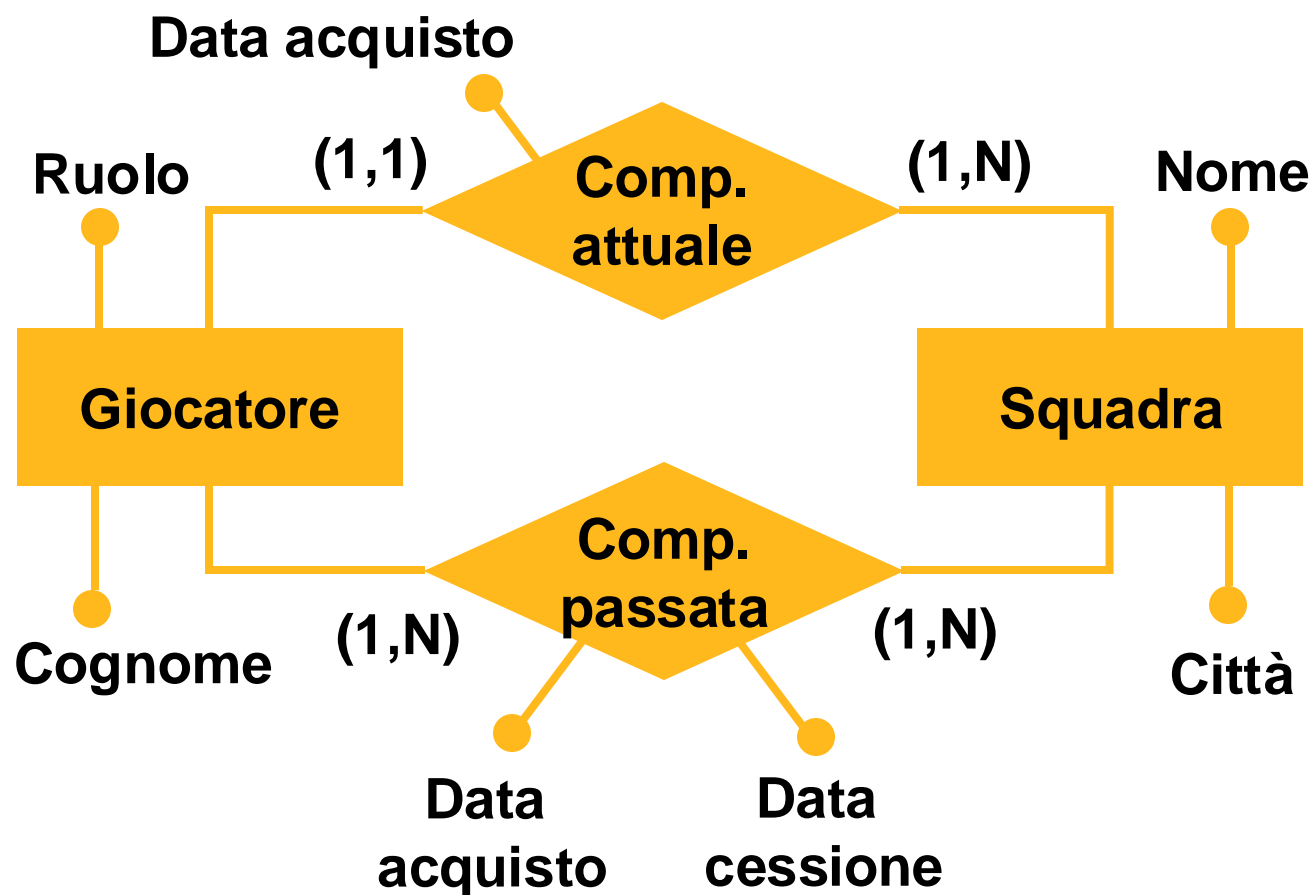
Partizionamento verticale di entità



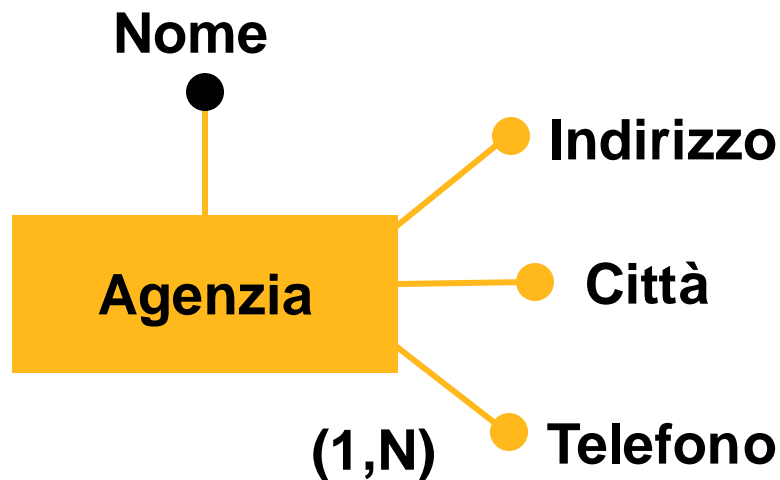
Partizionamento orizzontale di relationship



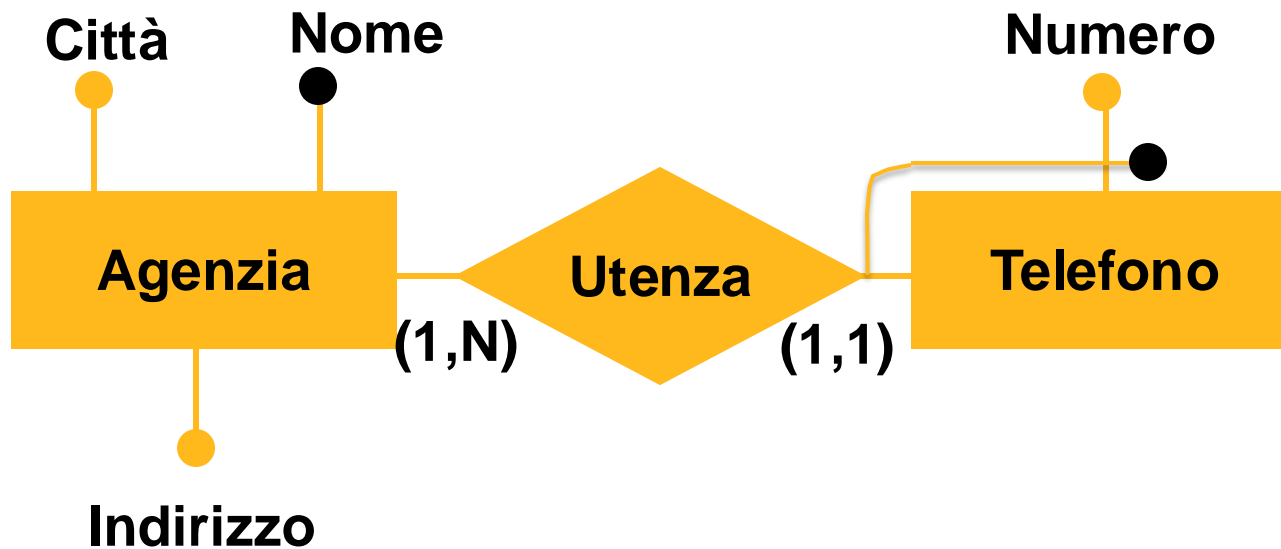
Partizionamento orizzontale di relationship



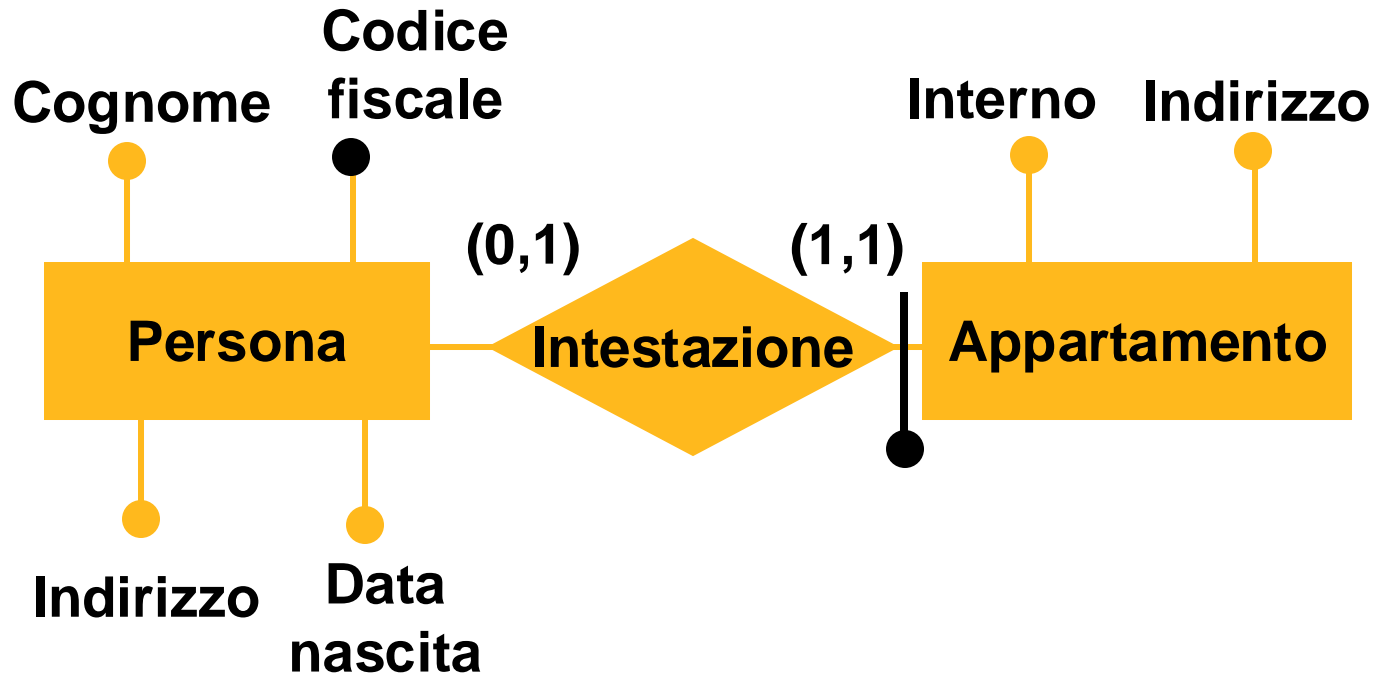
Eliminazione di attributi multivalore



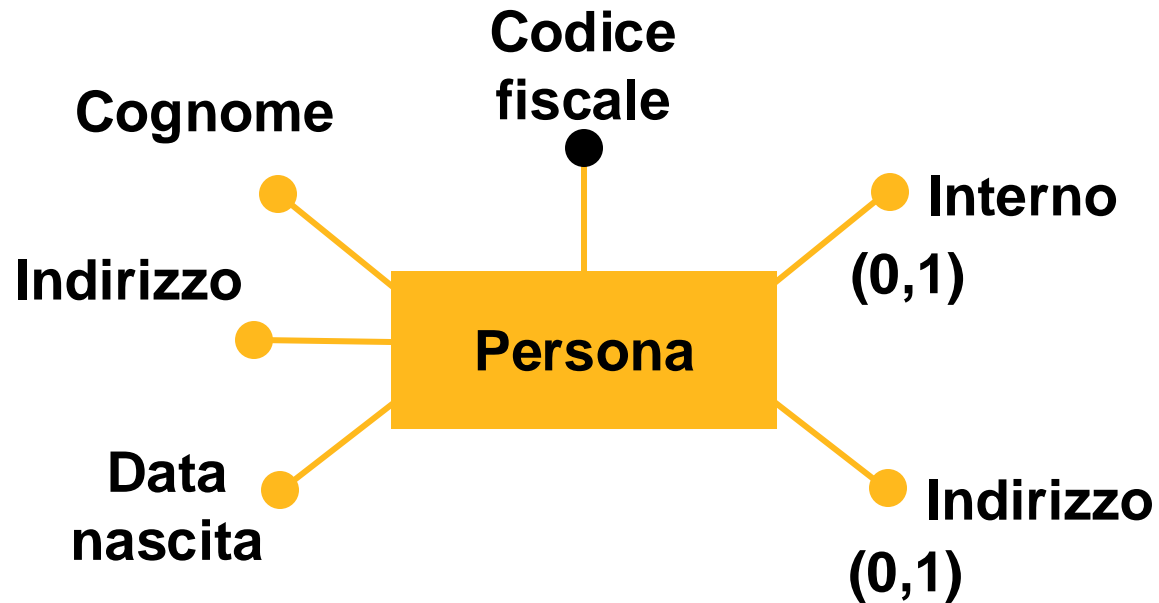
Eliminazione di attributi multivalore



Accorpamento di entità



Accorpamento di entità



Attività della ristrutturazione

- ▶ Analisi delle ridondanze
- ▶ Eliminazione delle generalizzazioni
- ▶ Partizionamento/accorpamento di entità e relationship
- ▶ Scelta degli identificatori principali

Scelta degli identificatori principali

- ▶ operazione indispensabile per la traduzione nel modello relazionale
- ▶ Criteri
 - ▶ assenza di opzionalità
 - ▶ semplicità
 - ▶ utilizzo nelle operazioni più frequenti o importanti
- ▶ Se nessuno degli identificatori soddisfa i requisiti visti?

Si introducono nuovi attributi (codici**)
contenenti valori speciali generati
appositamente per questo scopo**