

Basi di Dati

Progettazione Logica delle Basi di Dati

Prof. Giuseppe Polese

↓ **Requisiti della base di dati**

**Progettazione
concettuale**

↓ **Schema concettuale**

**Progettazione
logica**

↓ **Schema logico**

**Progettazione
Fisica**

↓ **Schema fisico**

Obiettivo della progettazione logica

- "tradurre" lo schema concettuale in uno schema logico che rappresenti gli stessi dati nel formato di un modello intermedio (**modello logico**), ad es. il modello relazionale.

Dati di ingresso e uscita

- Ingresso:

- schema concettuale
- informazioni sul carico applicativo
- modello logico

- Uscita:

- schema logico (memorizzabile tramite il DBMS)
- documentazione associata

Traduzione ER-Relazionale

- Non si tratta di una semplice trascrizione tra i due modelli
- Alcuni aspetti dello schema concettuale non sono direttamente rappresentabili nello schema logico
- In questa fase è opportuno anche valutare le prestazioni

**Carico
applicativo**

Schema E-R



```
graph TD; A[Carico applicativo] --> B[Ristrutturazione dello schema E-R]; B --> C[Modello logico]; C --> D[Traduzione nel modello logico]; D --> E[Schema logico];
```

**Ristrutturazione dello
schema E-R**

**Modello
logico**

**Schema E-R
ristrutturato**

**Traduzione nel
modello logico**

**Schema
logico**

Ristrutturazione schema E-R

- Motivazioni:
 - semplificare la traduzione
 - "ottimizzare" le prestazioni
- Per ottimizzare il risultato abbiamo bisogno di analizzare le prestazioni a questo livello
- Le prestazioni non sono valutabili con precisione su di uno schema concettuale!

Parametri per valutare le prestazioni

- numero di occorrenze previste
- numero di accessi ad occorrenze (di entità ed associazioni) durante un'operazione

Principio di Pareto (80:20)

- Regola empirica secondo la quale un sistema dedica l'80% delle sue risorse per elaborare il 20% delle operazioni più frequenti.
- Sfruttando questo principio calcoliamo gli accessi totali per il 20% di operazioni più frequenti.

Tavole di Carico

- Pertanto, per stimare le prestazioni sviluppiamo 3 tipi di tavole:
 - **Tavola Volumi**, contenente una stima delle occorrenze per entità ed associazioni
 - **Tavola operazioni**, riporta tipo e frequenza per il 20% di operazioni più frequenti
 - **Tavole accessi**, numero accessi in lettura e scrittura su entità ed associazioni per il 20% di operazioni più frequenti

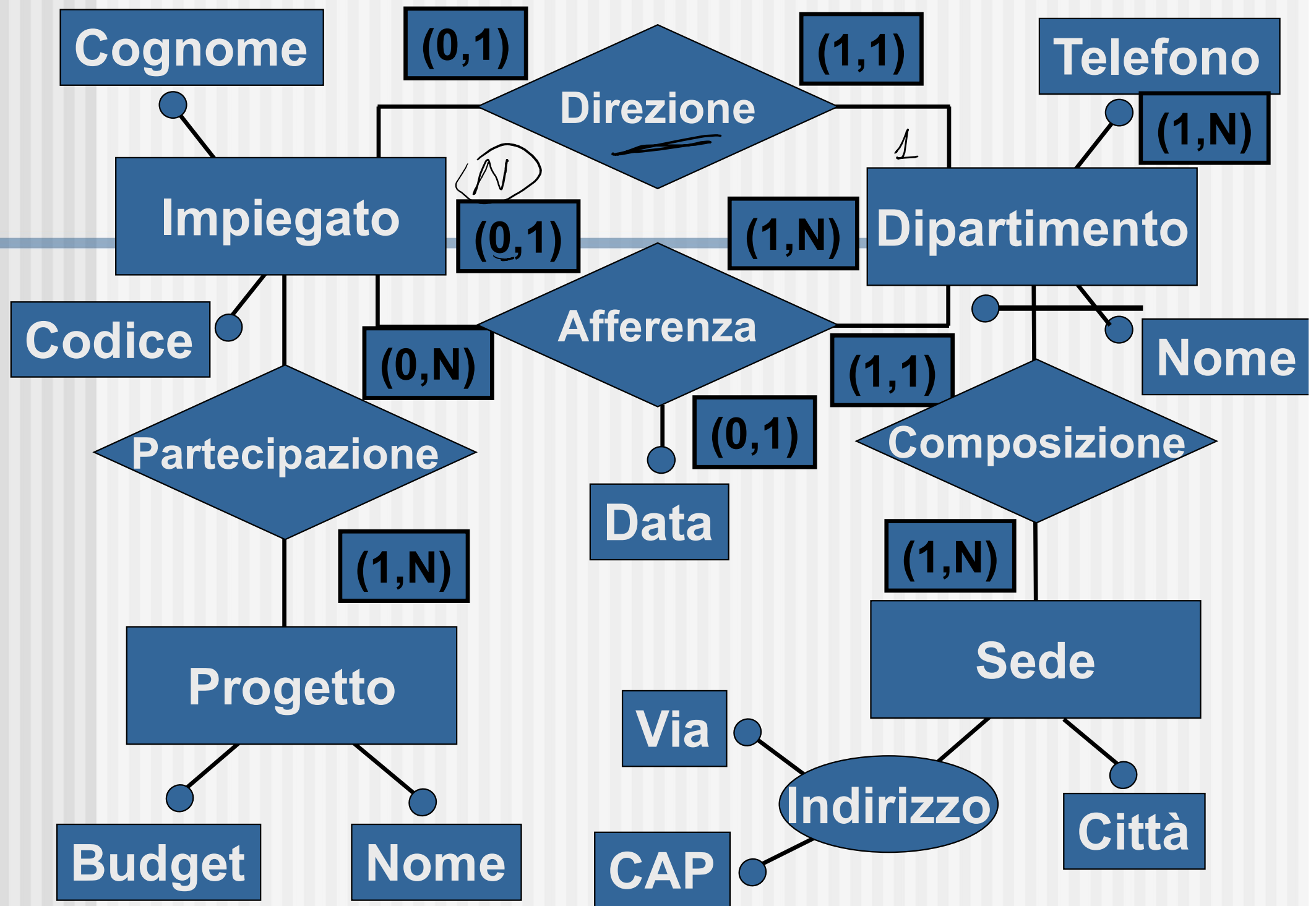


Tavola dei volumi

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

Tavola delle operazioni

Operazione	Tipo	Frequenza
Operazione 1	I	1 volta/giorno
Operazione 2	B	1 volta/mese

- **I**: Operazione Interattiva
- **B**: Operazione Batch

In questo caso sono state previste 10 operazioni, quindi la stima si concentra sul 20% (2) con maggiore frequenza

Esempio di valutazione di costo

- Operazione frequente:
 - 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 di uno schema di navigazione

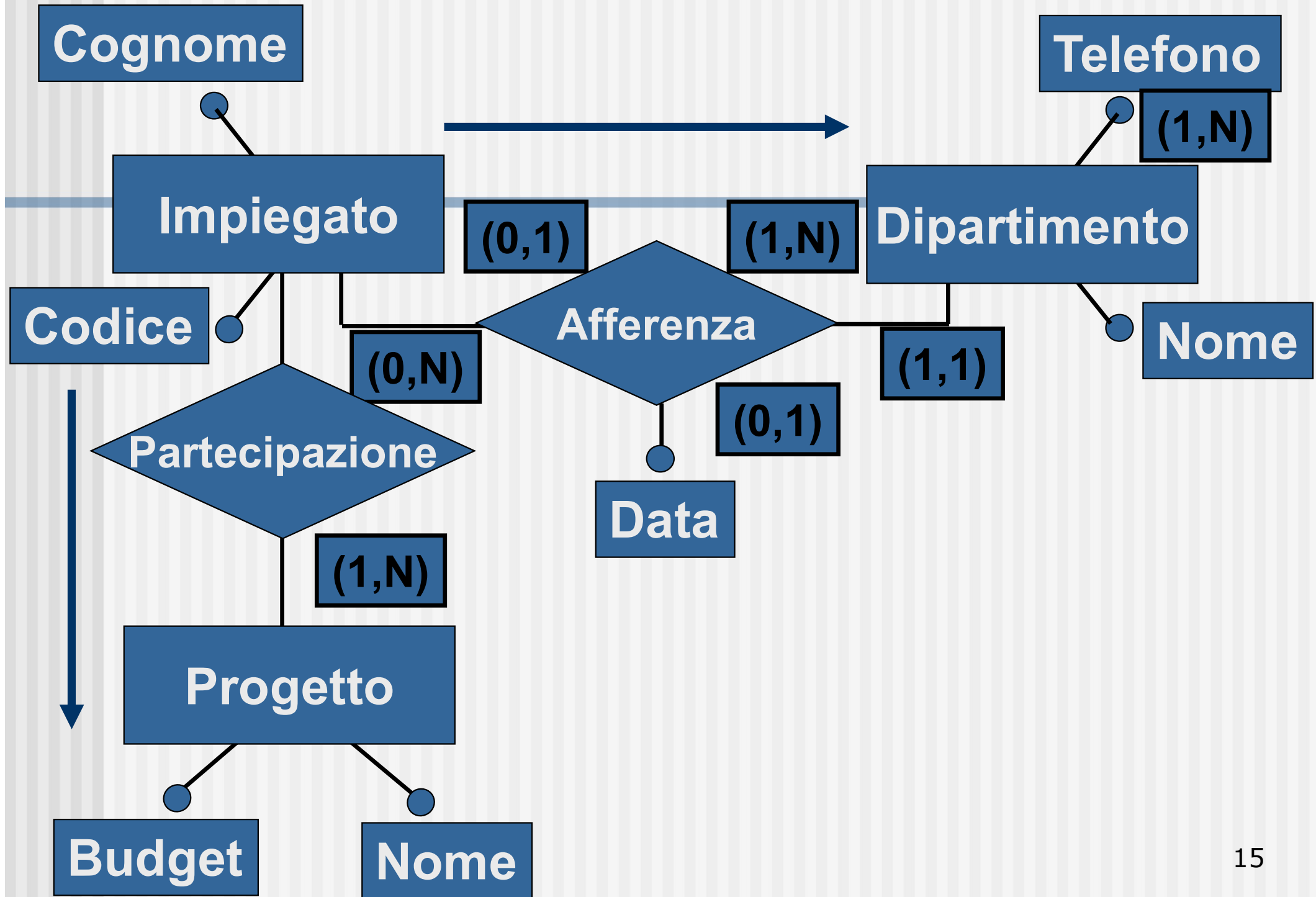


Tavola degli accessi

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

Attività della ristrutturazione

- **Analisi delle ridondanze**
- Eliminazione delle generalizzazioni
- Partizionamento/accorpamento di entità e associazioni
- 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 di mantenerle, in base al loro impatto sul numero di accessi per il 20% di operazioni più frequenti

Ridondanze

■ Vantaggi

- semplificazione delle interrogazioni

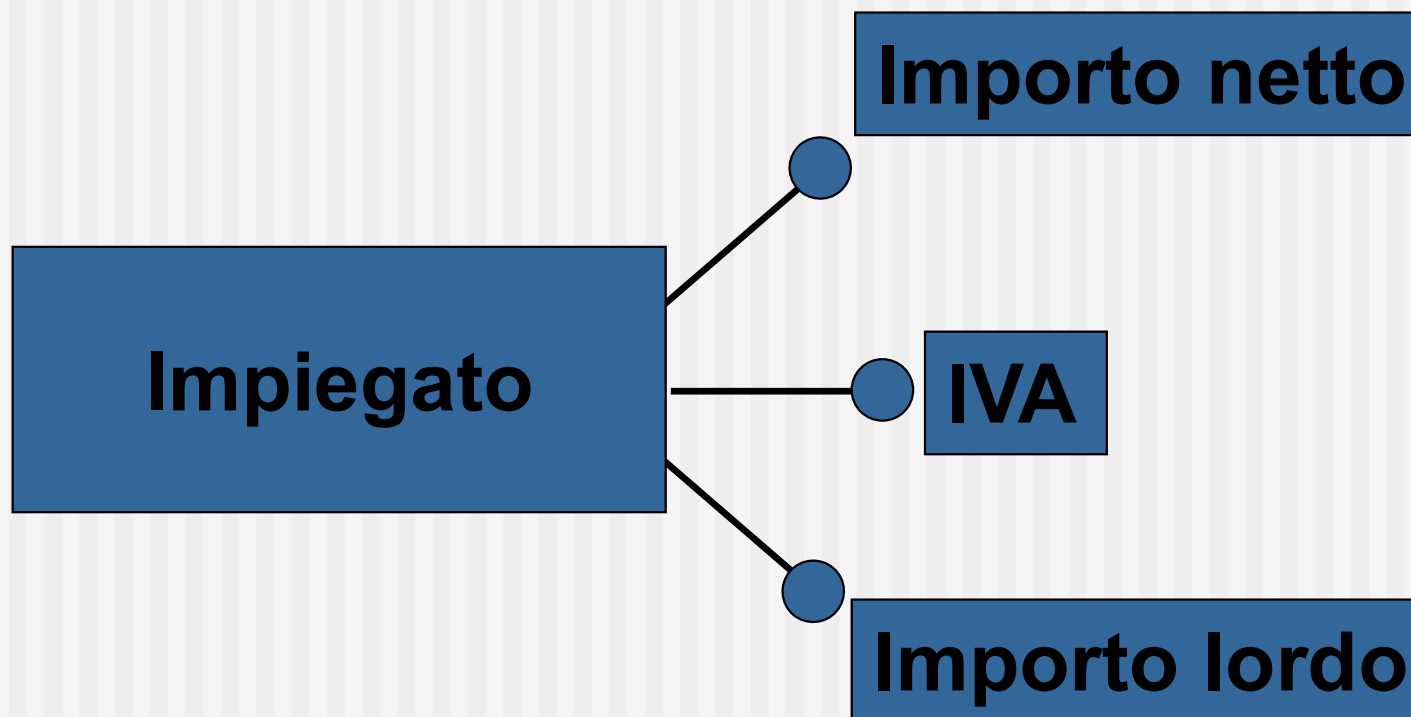
■ Svantaggi

- appesantimento degli aggiornamenti
- maggiore occupazione di spazio

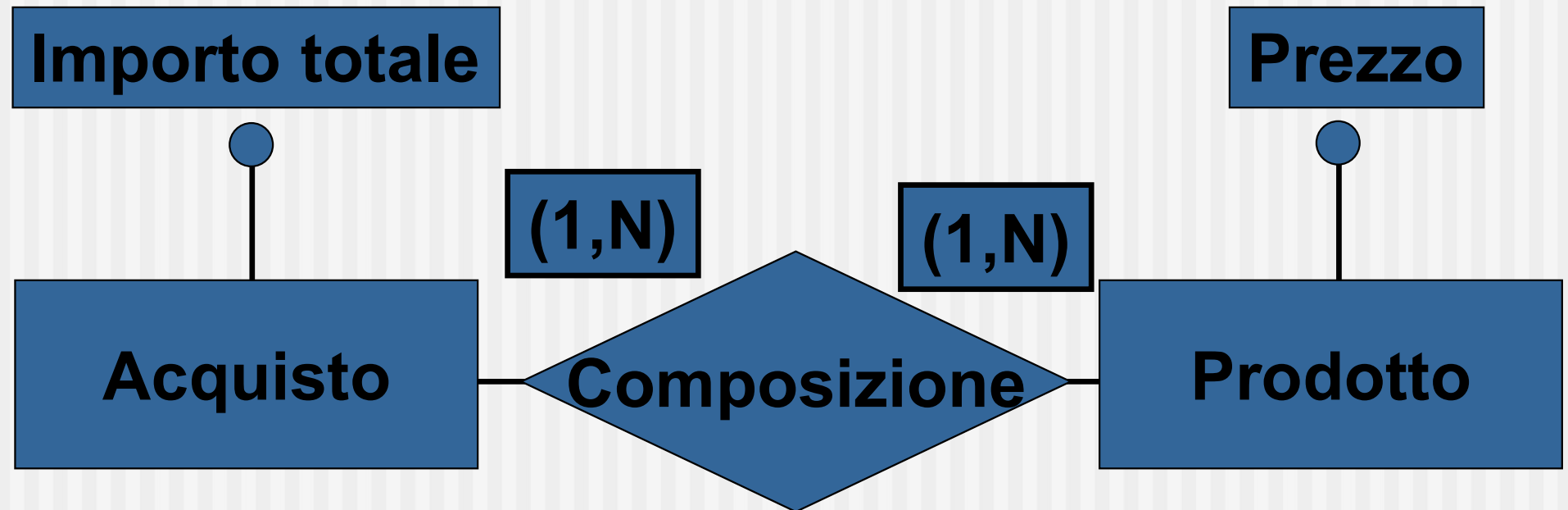
Forme di ridondanza in uno schema E-R

- attributi derivabili:
 - da altri attributi della stessa entità (o associazione)
 - da attributi di altre entità (o associazioni)
- Associazioni derivabili dalla composizione di altre associazioni in presenza di cicli

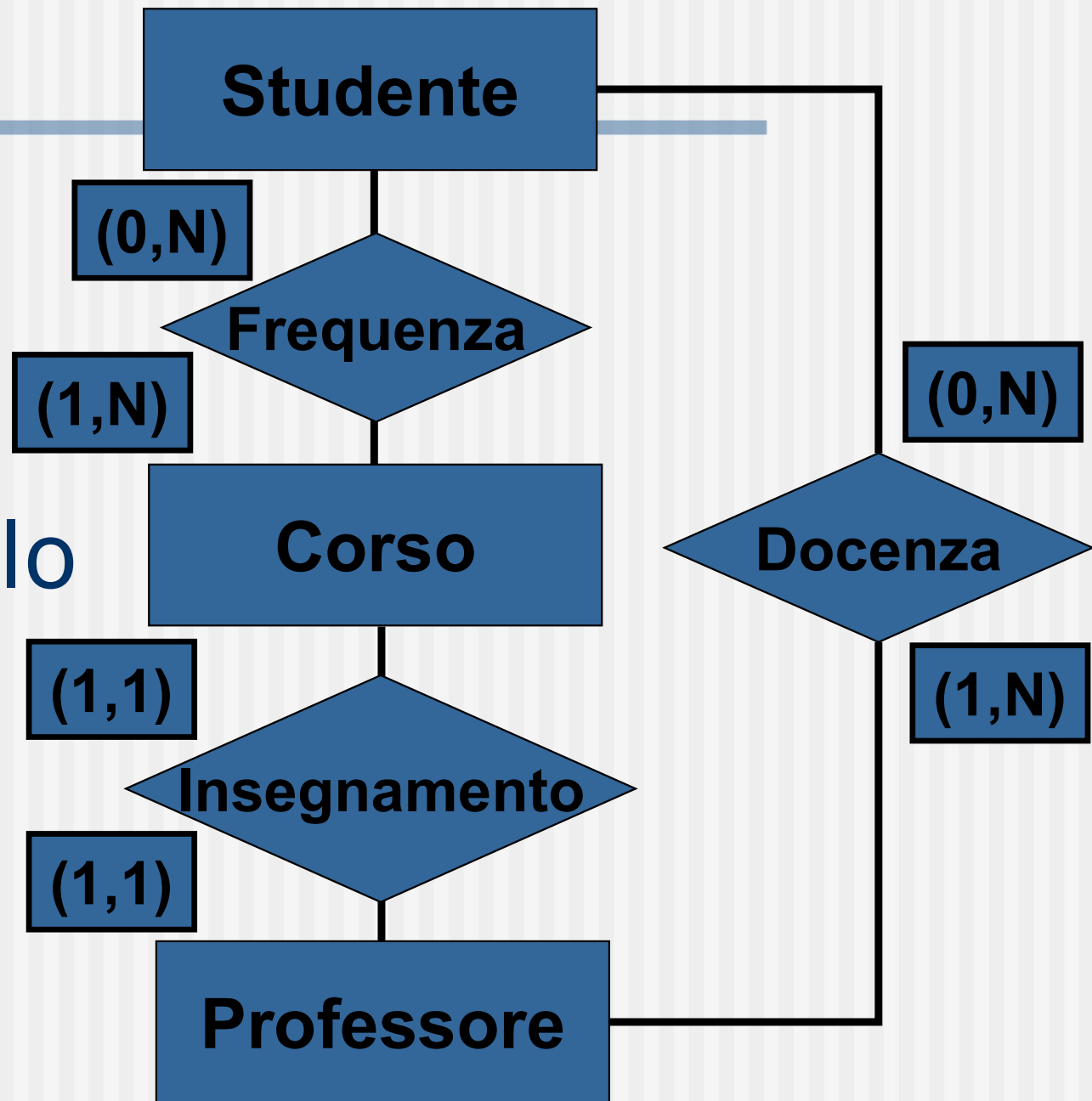
Attributo derivabile



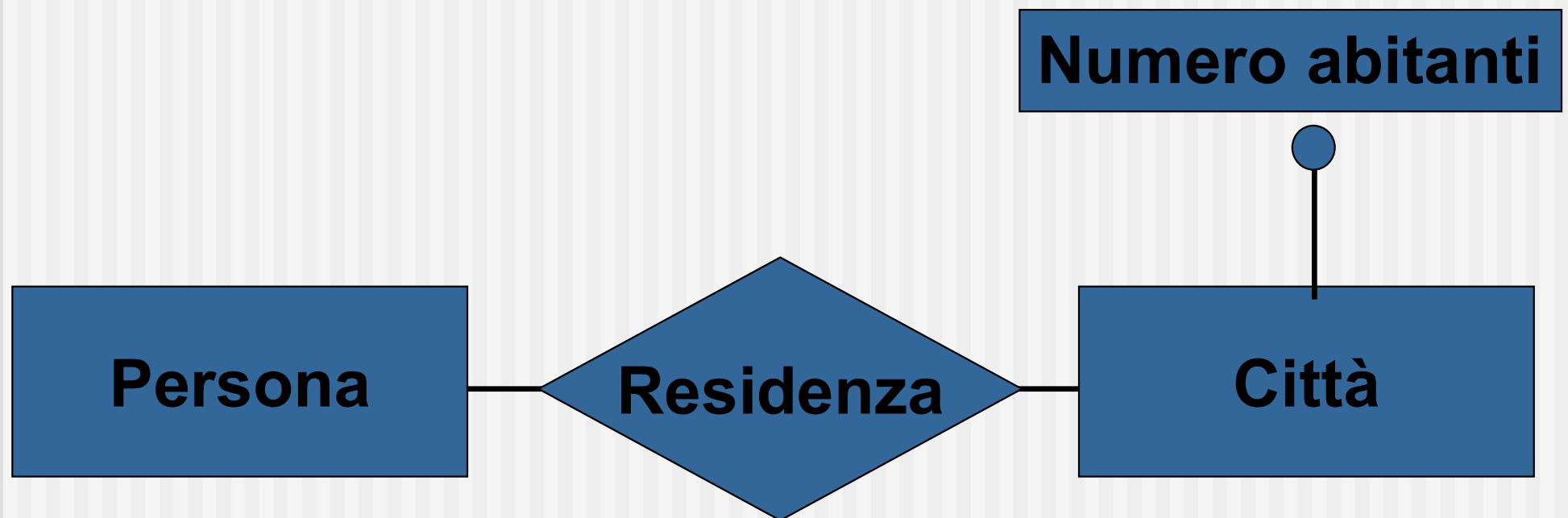
Attributo derivabile da altra entità



Ridondanza
dovuta a ciclo



Analisi di una ridondanza



Ipotesi di Tavola dei volumi

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

- Inoltre, se una città può avere fino a milioni di abitanti, occorrono circa **3 byte** per città per memorizzare il dato ridondante, totale **600 byte**.

Ipotesi di Tavola Operazioni

Concetto	Tipo	Volume
Operazione 1	I	500 volte/giorno
Operazione 2	B	2 volte/giorno

- **Operazione 1**: memorizza una nuova persona e relativa città di residenza
- **Operazione 2**: stampa i dati di una città (incluso il numero di abitanti)

Tavole 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

Operazione 2

Concetto	Costrutto	Accessi	Tipo
Città	Entità	1	L

Tavole 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

Numero totale accessi (In presenza di ridondanza)

- Costi:
 - Operazione 1: 1500 accessi in scrittura e 500 accessi in lettura al giorno
 - Operazione 2: 2 accessi in lettura.
- Contiamo doppi gli accessi in scrittura
- Totale di 3502 accessi al giorno e 600 byte per il dato ridondante

Assenza di ridondanza

■ Costi:

- Operazione 1: 1000 accessi in scrittura
- Operazione 2: 10002 accessi in lettura al giorno

■ Contando doppi gli accessi in scrittura si hanno **12002 accessi al giorno**

Attività della ristrutturazione

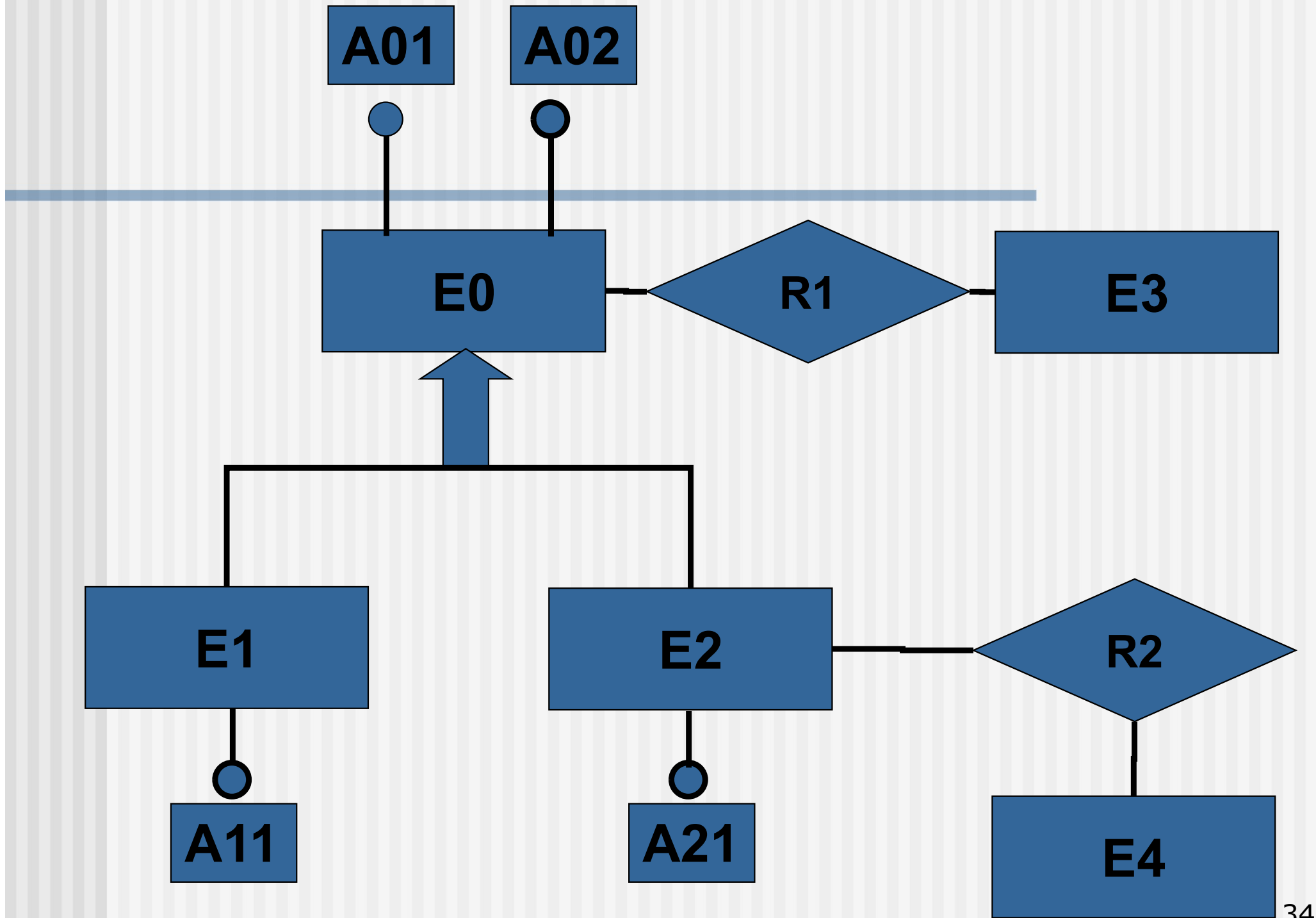
- Analisi delle ridondanze
- Eliminazione delle generalizzazioni
- Partizionamento/accorpamento di entità e relazioni
- Scelta degli identificatori primari

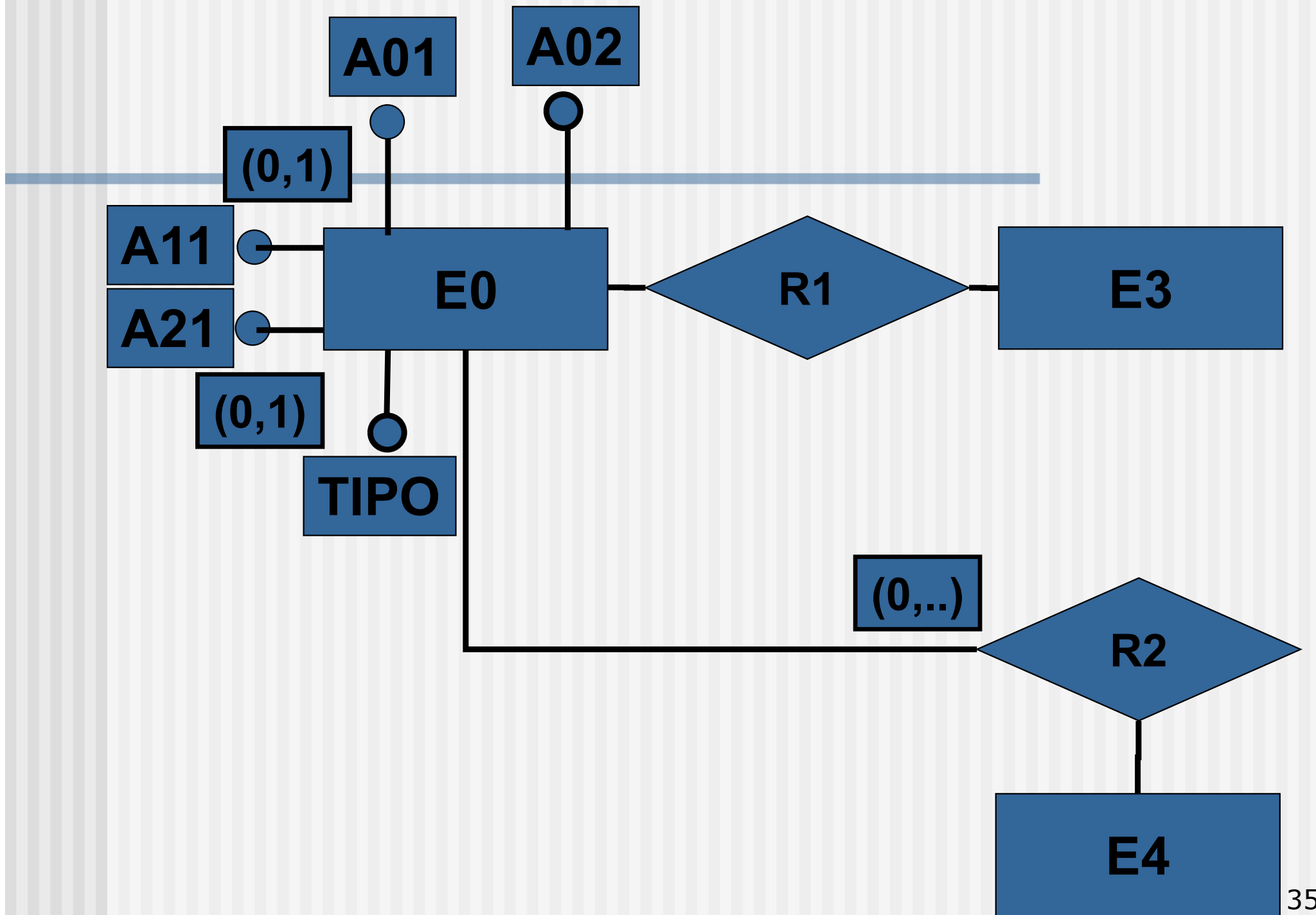
Eliminazione delle gerarchie

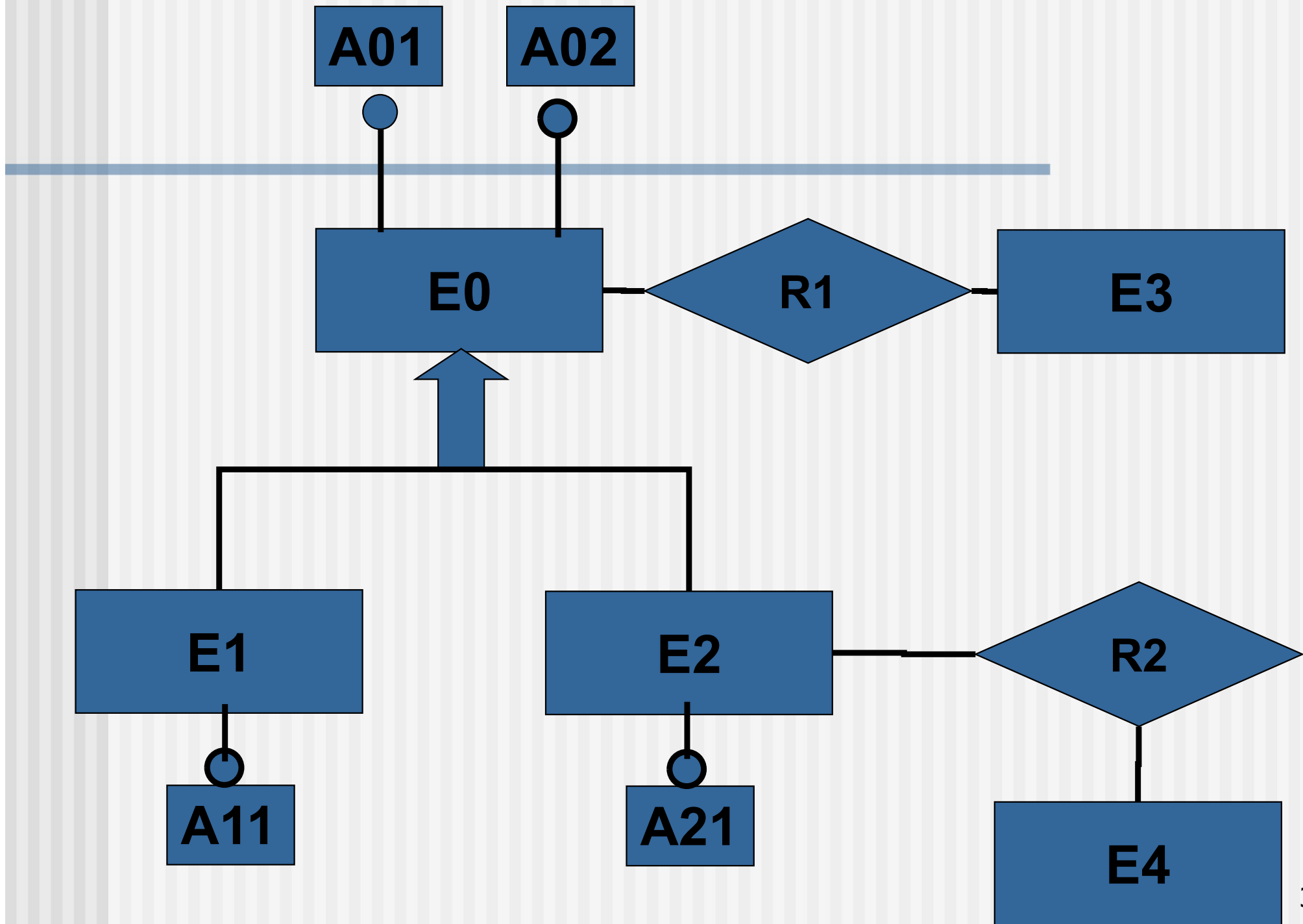
- il modello relazionale non può rappresentare direttamente le generalizzazioni
- entità e associazioni sono invece direttamente rappresentabili
- si eliminano perciò le gerarchie, sostituendole con entità e associazioni

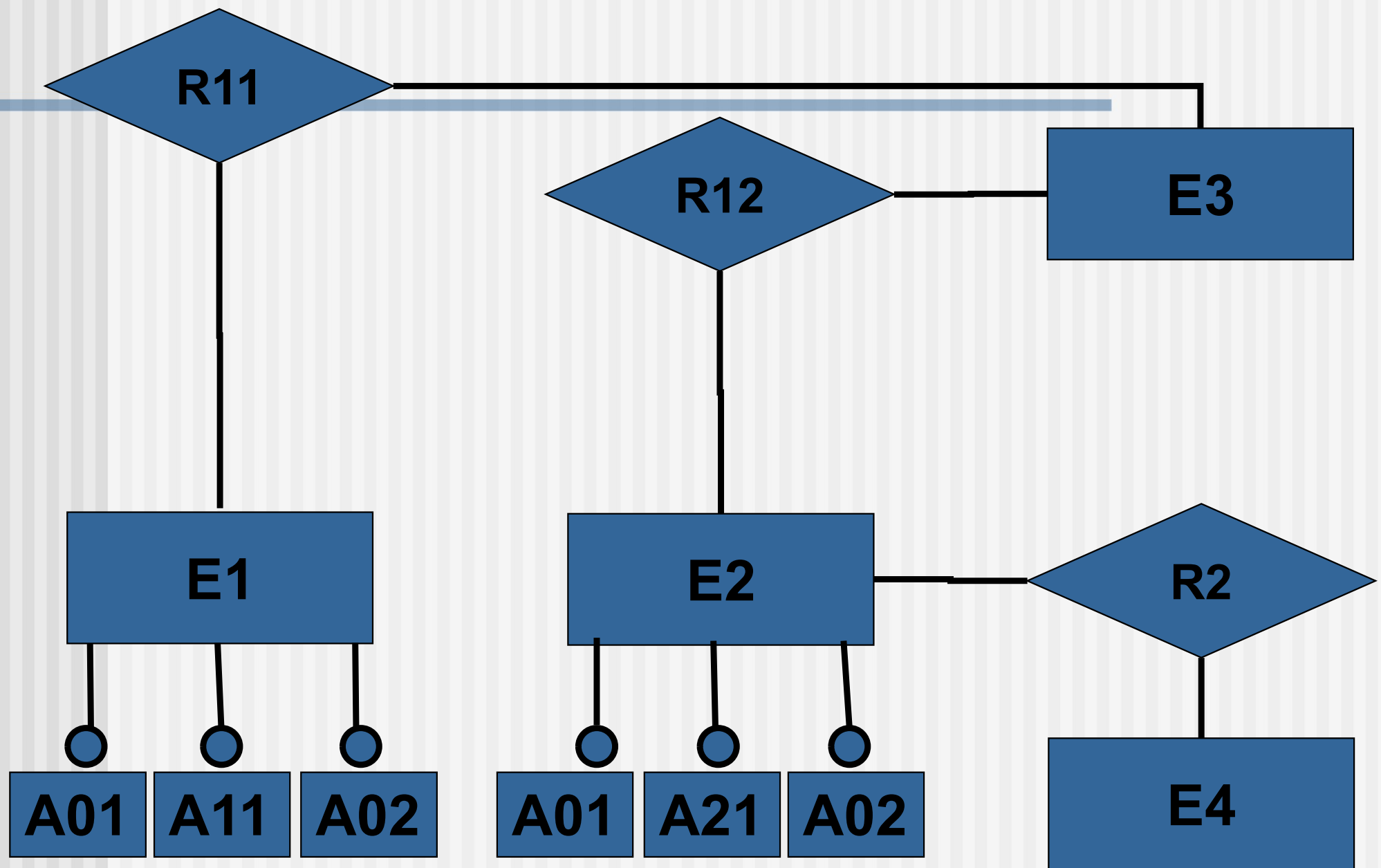
Tre possibilità

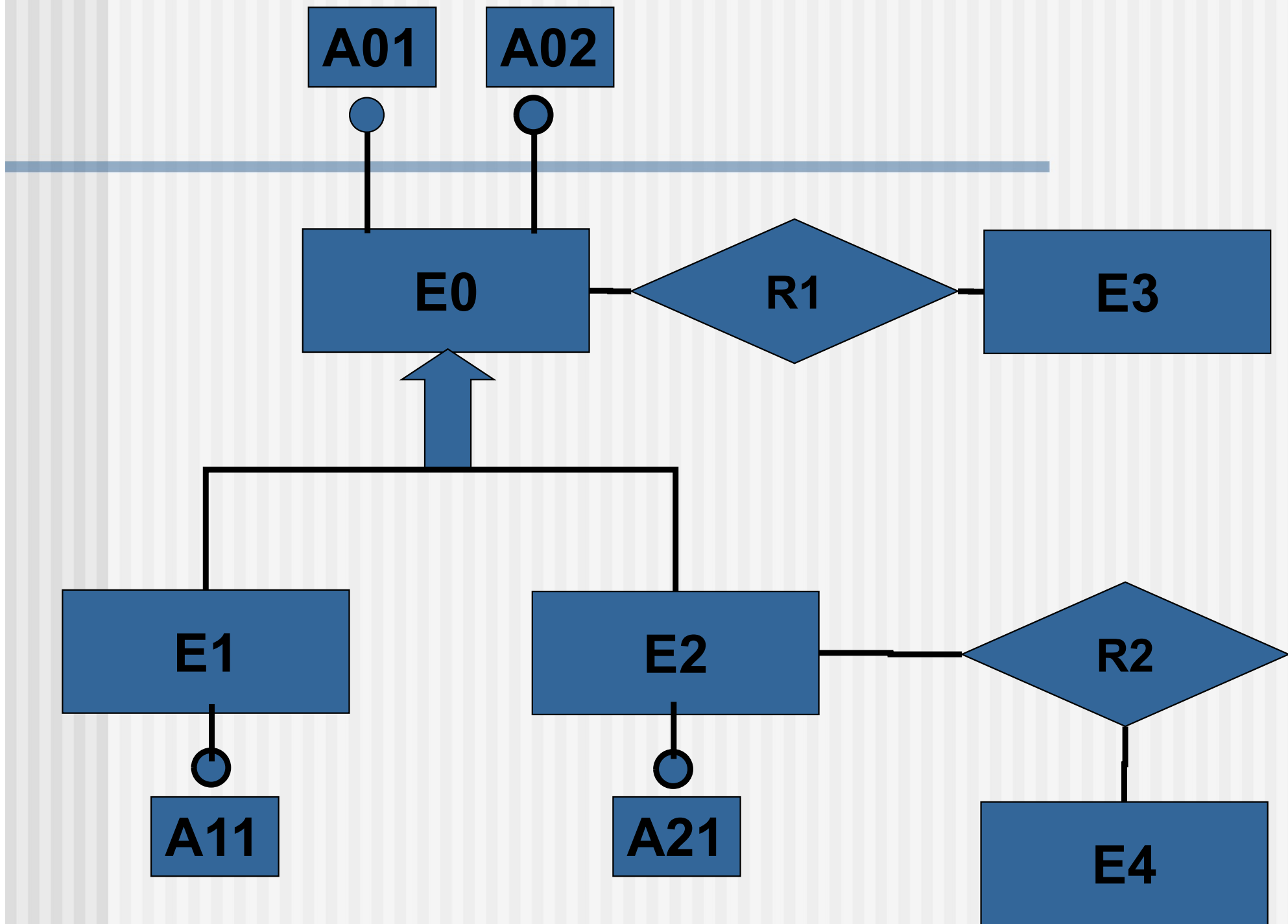
- accorpamento delle figlie della generalizzazione nel genitore
- accorpamento del genitore della generalizzazione nelle figlie
- sostituzione della generalizzazione con associazioni

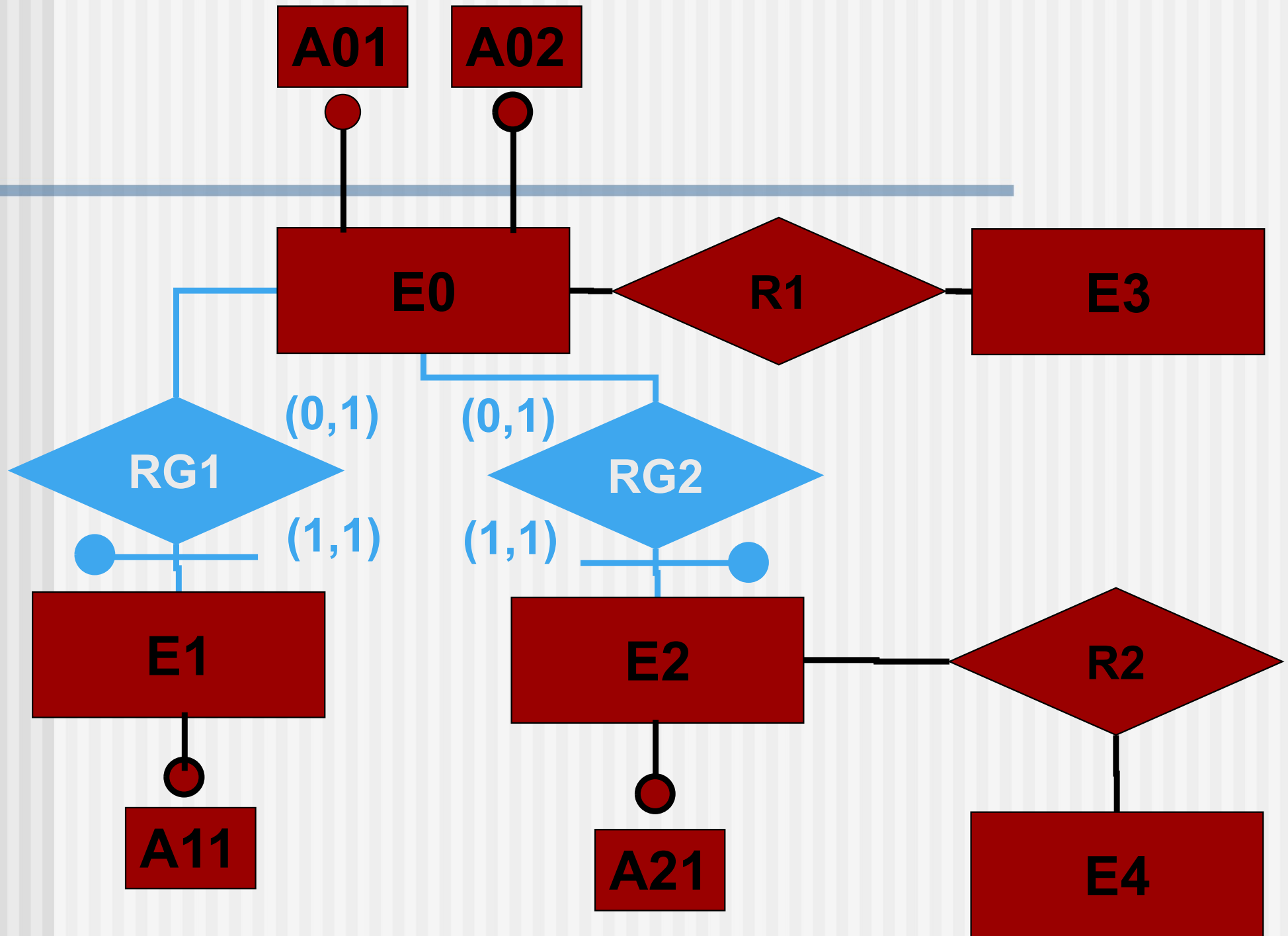










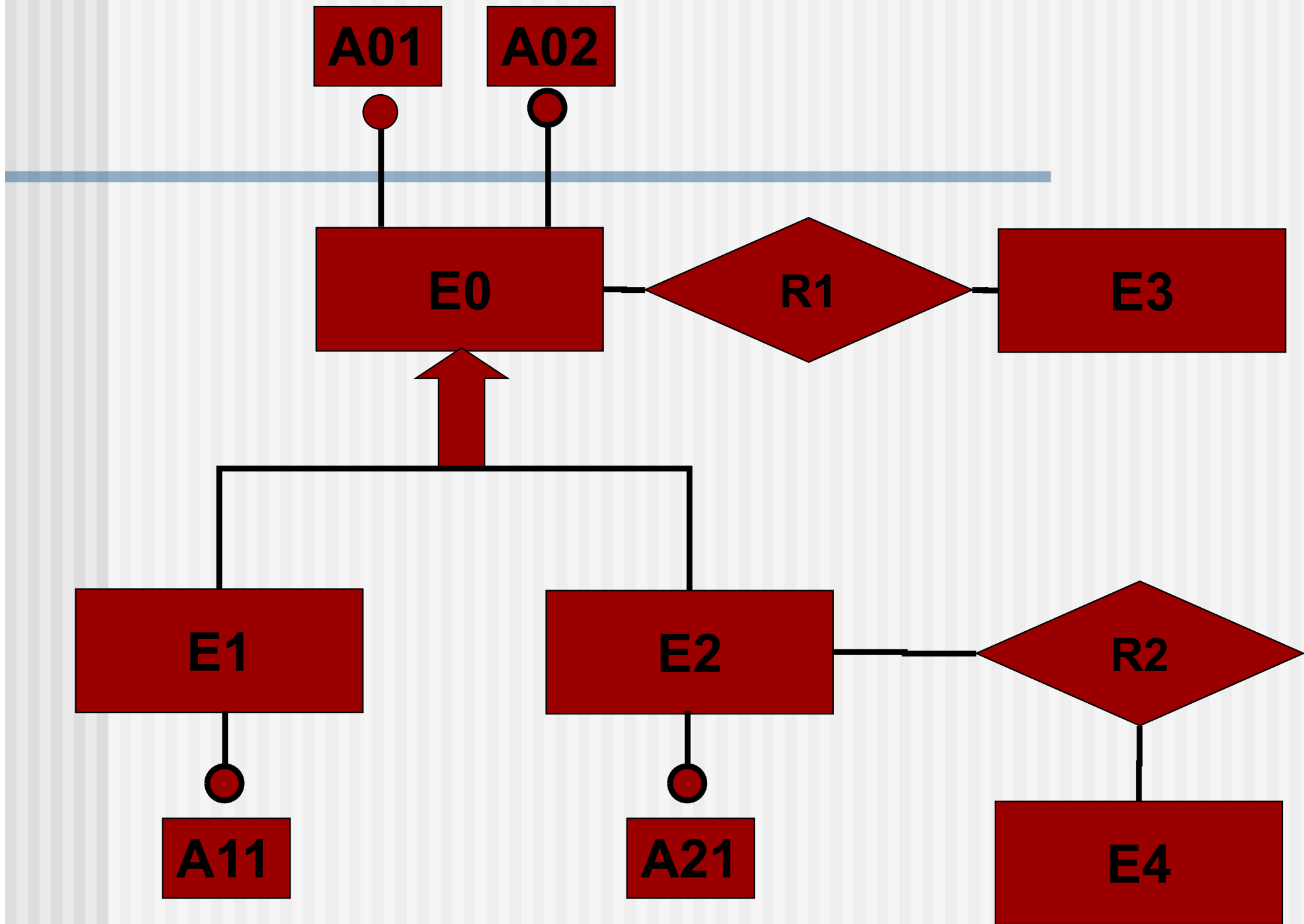


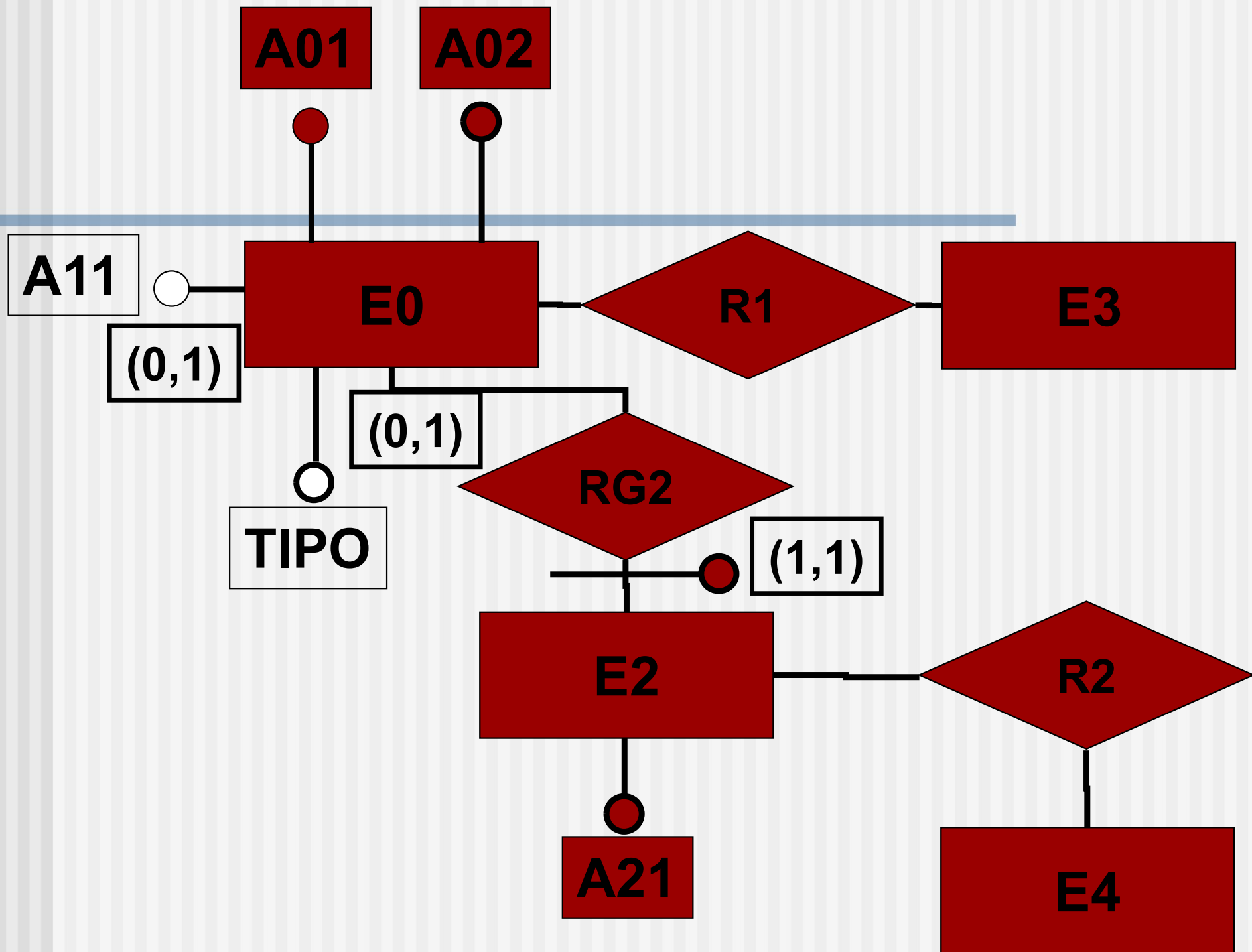
Scelte progettuali

- 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 però seguire alcune semplici regole generali

Criteri di scelta

- conviene se gli accessi al padre e alle figlie sono contestuali
- conviene se gli accessi alle figlie sono distinti
- conviene se gli accessi alle entità figlie sono separati dagli accessi al padre
- sono anche possibili soluzioni “ibride”, soprattutto in gerarchie a più livelli





Attività della ristrutturazione

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

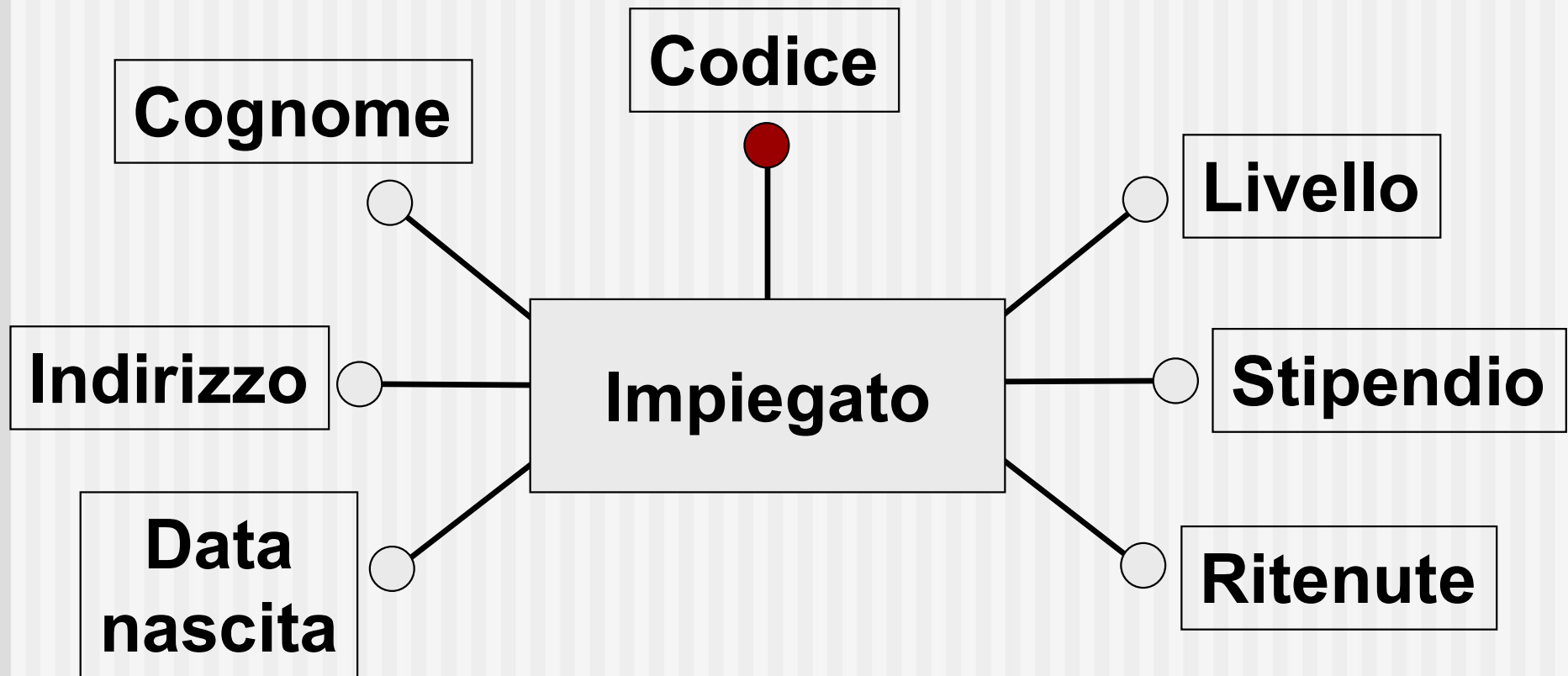
Partizionamenti e Accorpamenti

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

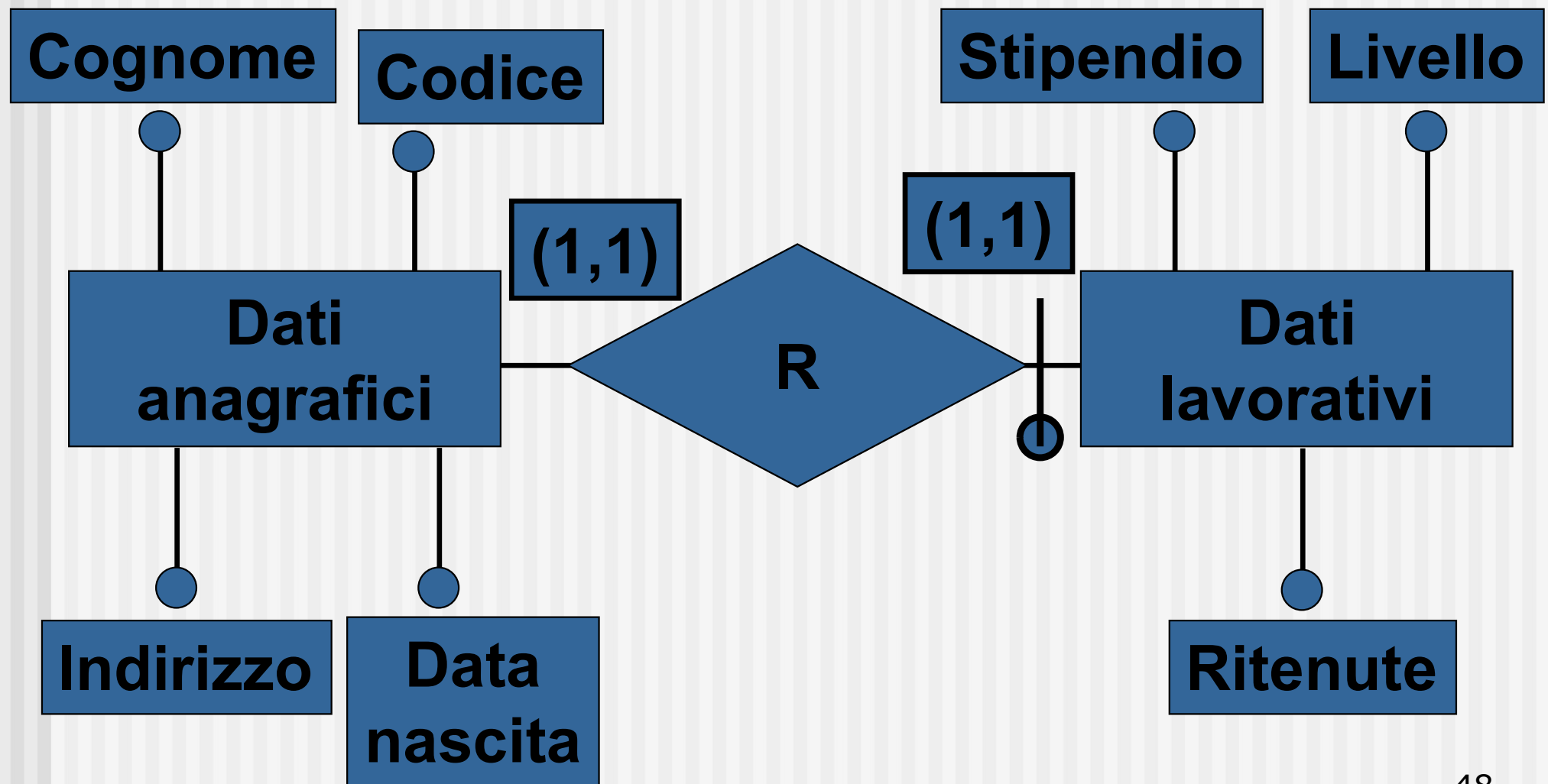
Ristrutturazioni, casi principali

- partizionamento verticale di entità
- partizionamento orizzontale di associazioni
- eliminazione di attributi multivalore
- accorpamento di entità/associazioni

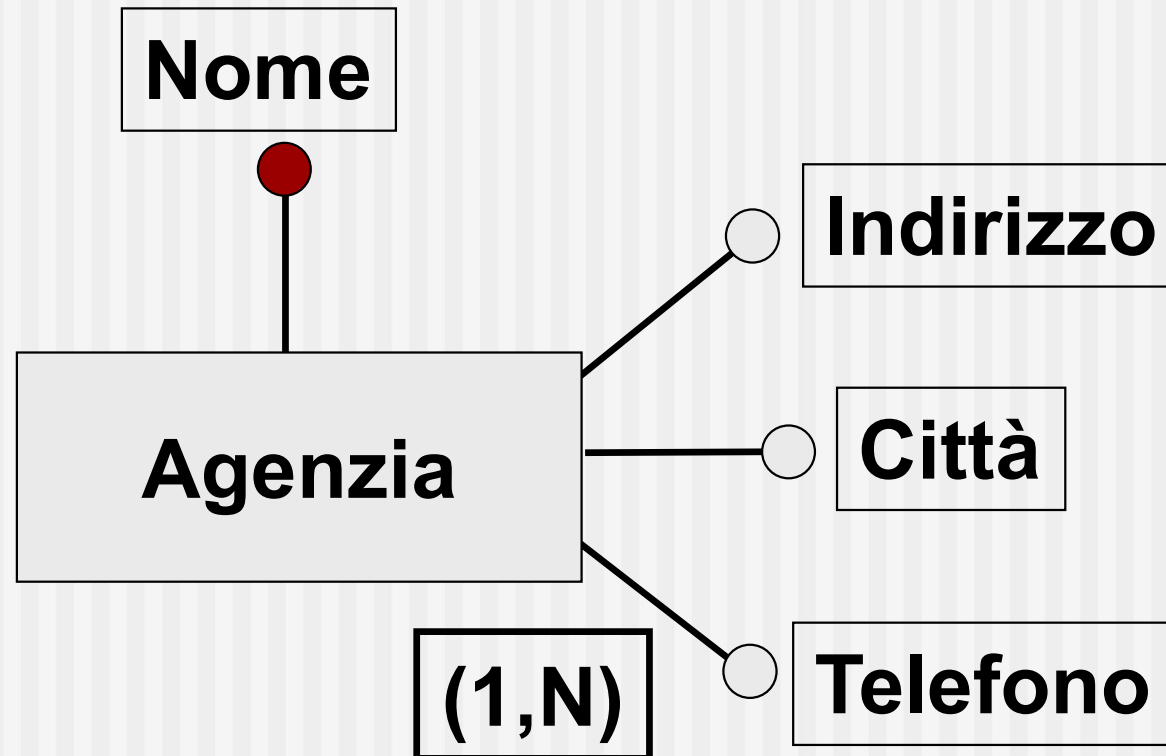
Partizionamento: Esempio 1



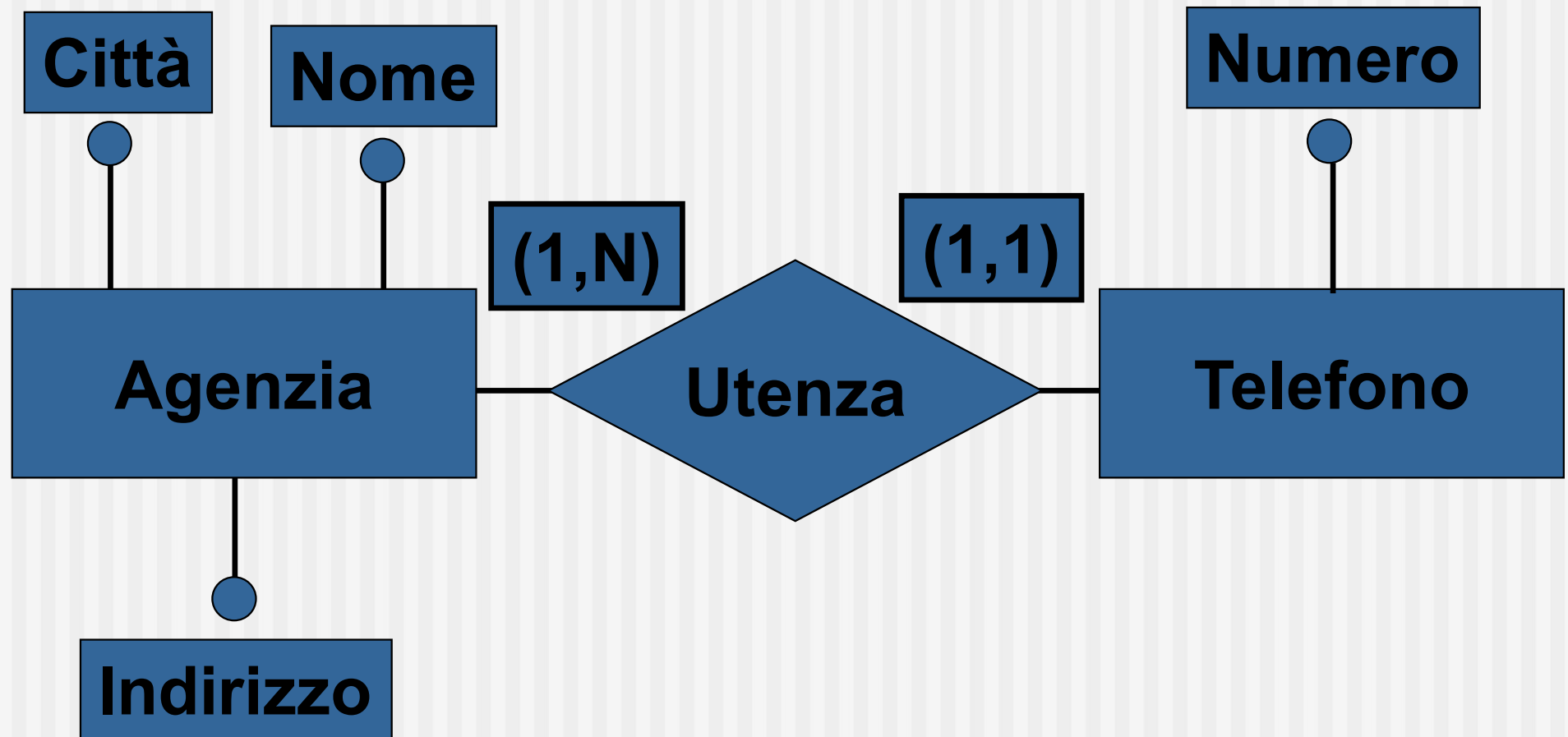
Partizionamento: Esempio 1



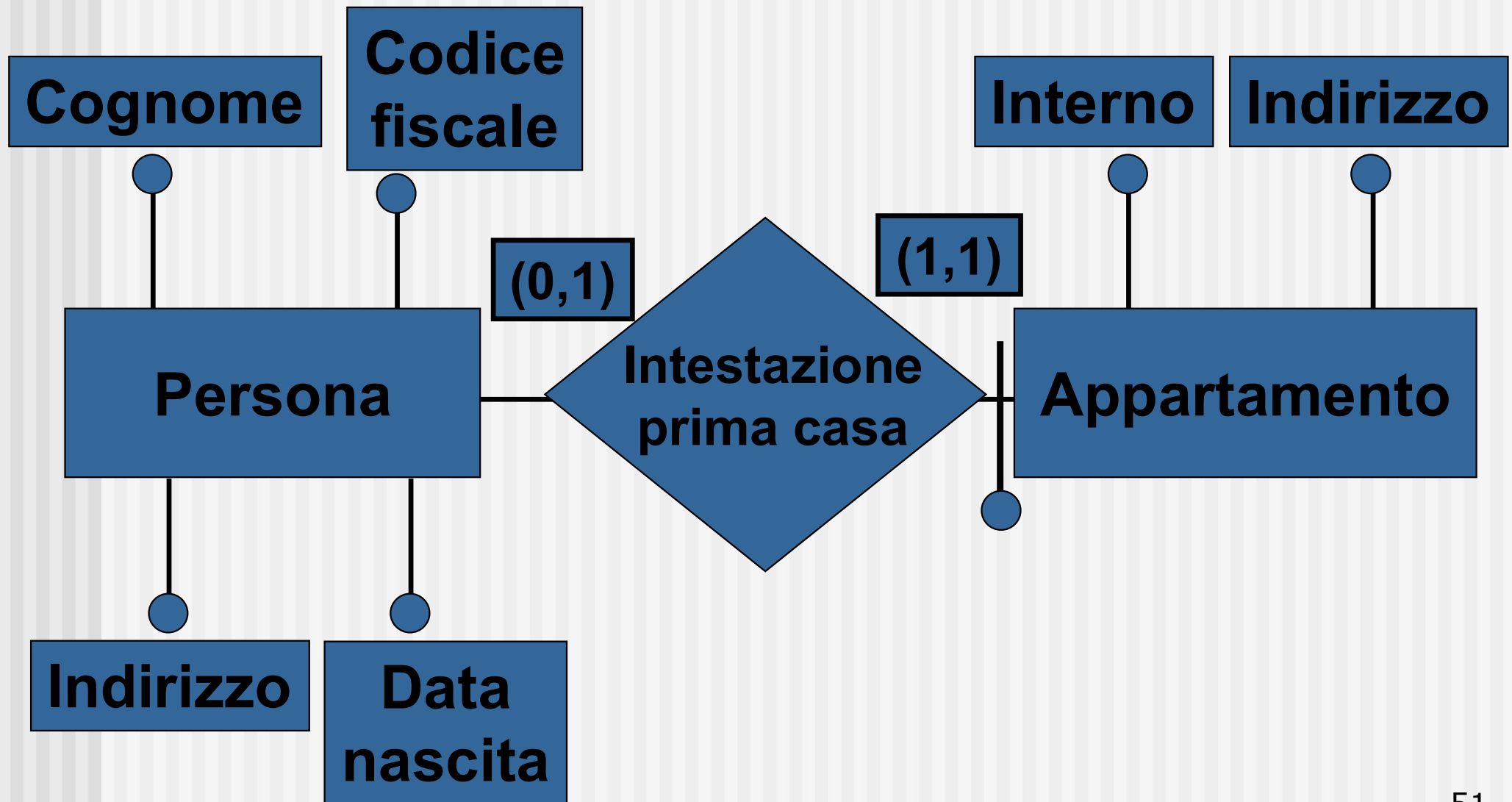
Partizionamento: Esempio2



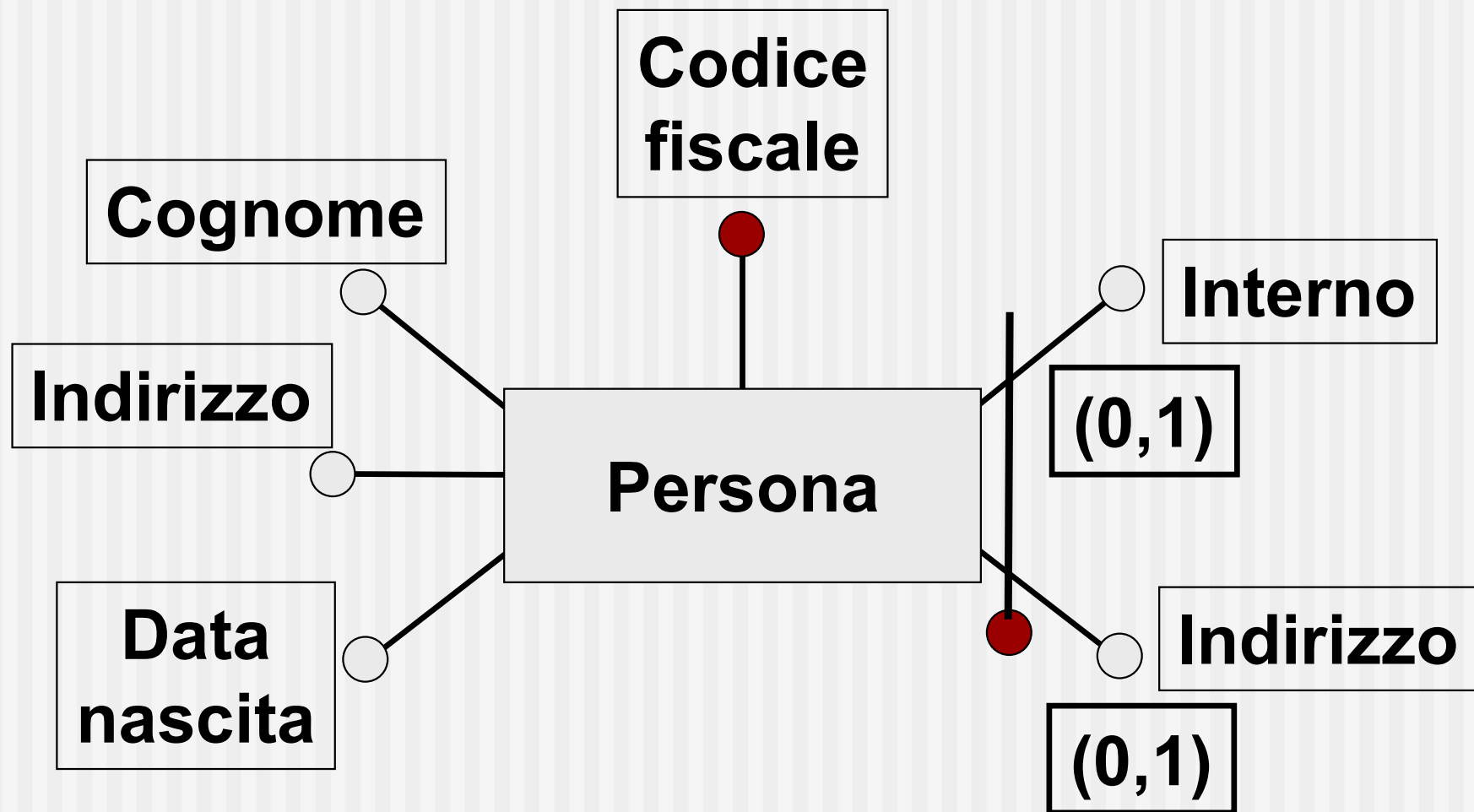
Partizionamento: Esempio2



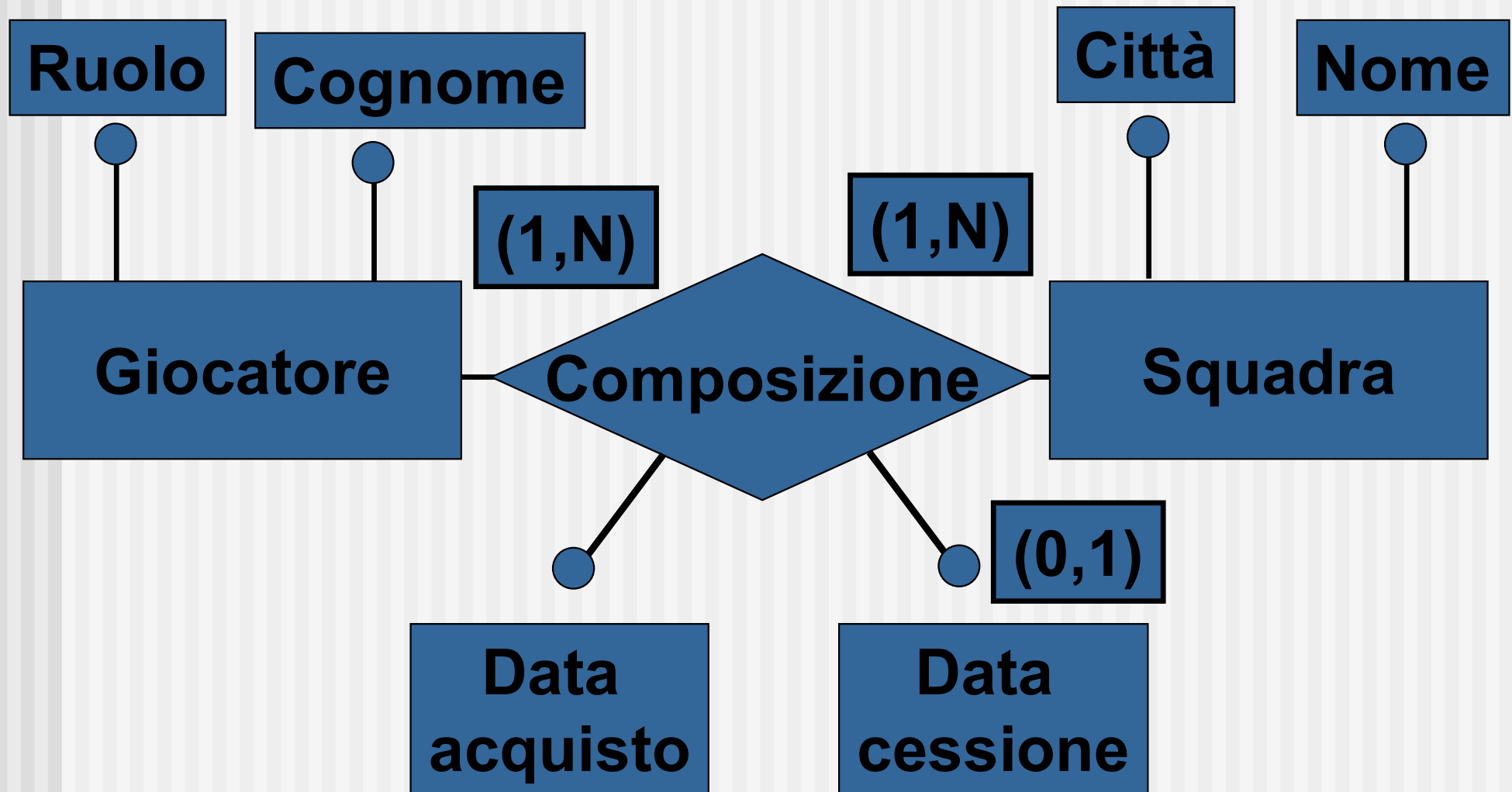
Accorpamento: Esempio 1

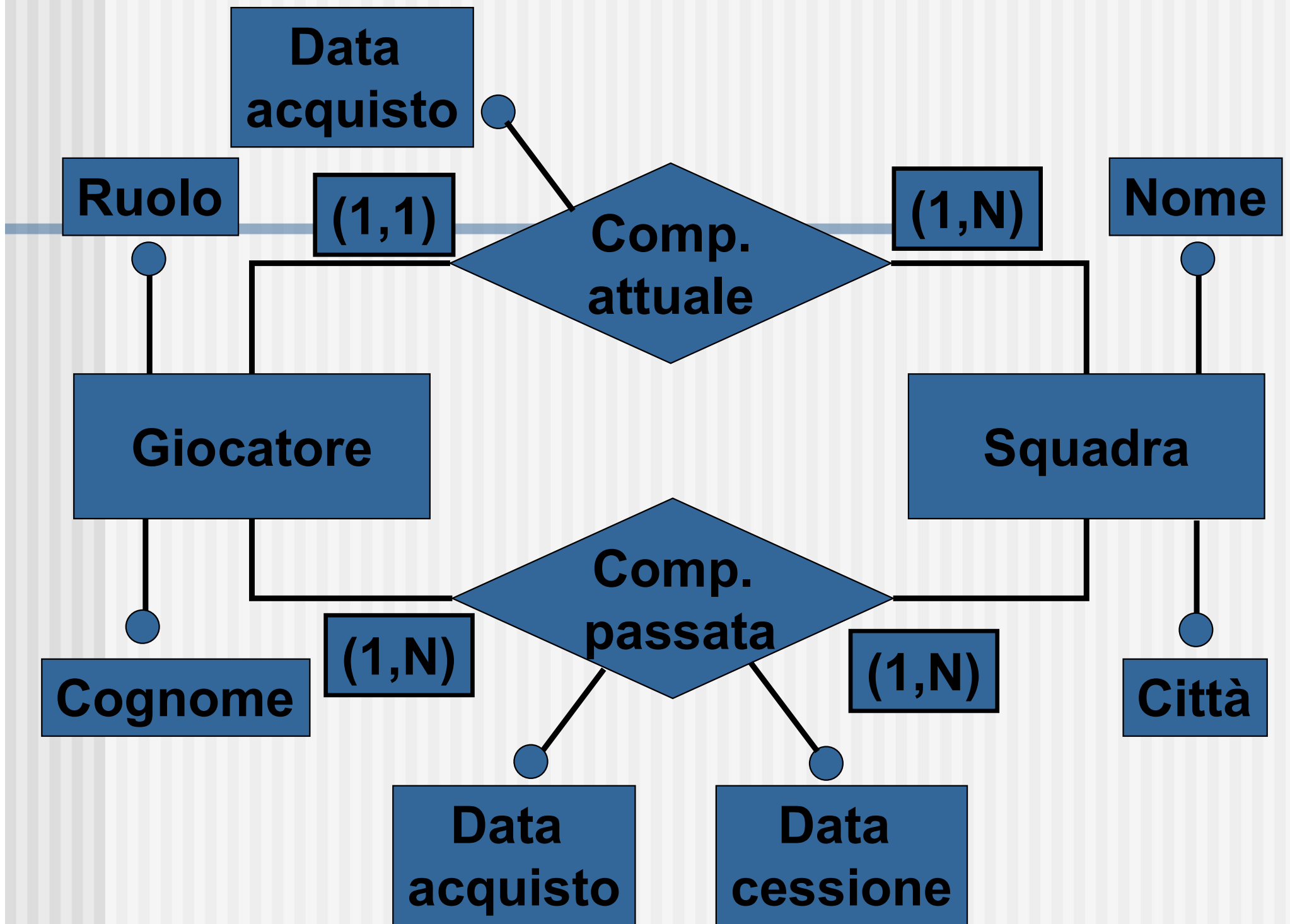


Accorpamento: Esempio 1



Partizionamento Orizzontale





Attività della ristrutturazione

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

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 per questo scopo