



Capitolo 8

Progettazione logica

Requisiti della base di dati

**Progettazione
concettuale**

Schema concettuale

**Progettazione
logica**

Schema logico

**Progettazione
fisica**

Schema fisico



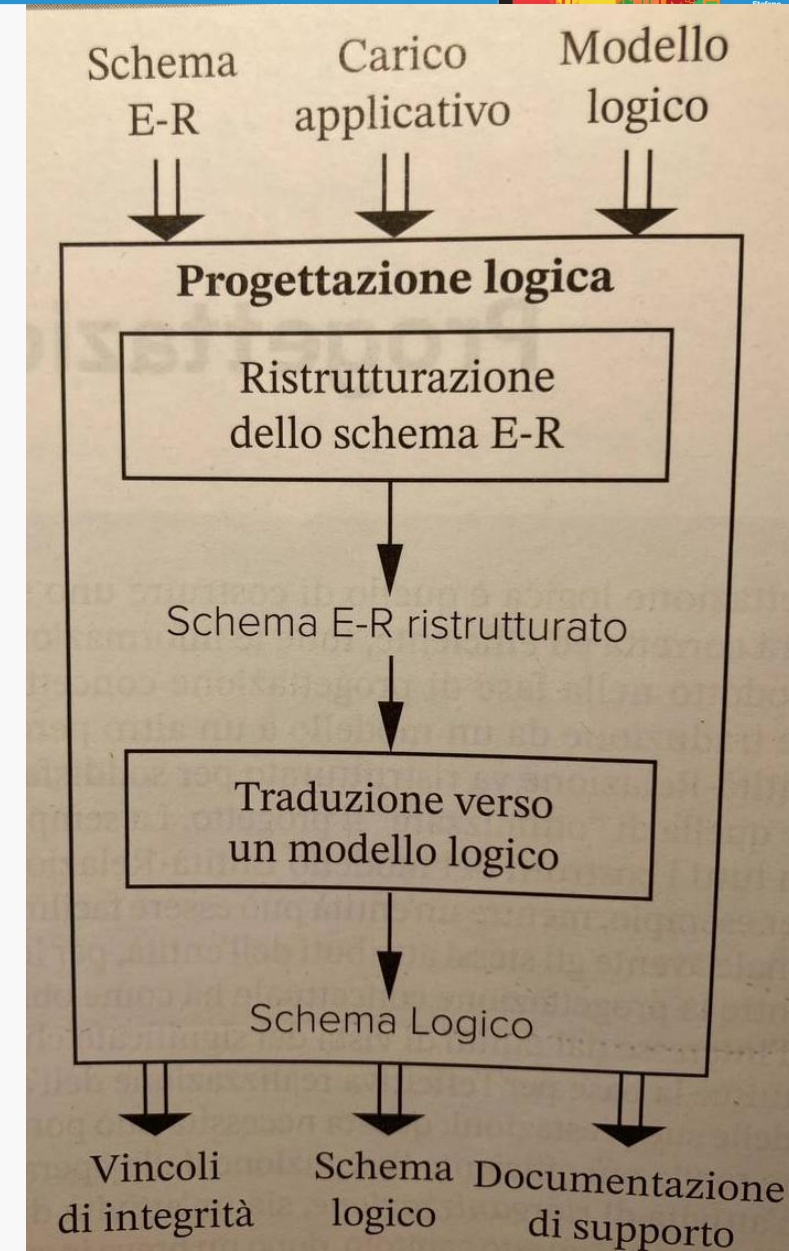
Obiettivo della progettazione logica

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



Dati di ingresso e uscita

- Ingresso:
 - schema concettuale
 - informazioni sul carico applicativo
 - modello logico
- Uscita:
 - schema logico
 - documentazione associata
 - Vincoli di integrità



Non si tratta di una pura e semplice traduzione

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



Ristrutturazione schema E-R

- Motivazioni:
 - semplificare la traduzione verso il modello relazionale
 - "ottimizzare" le prestazioni
- Osservazione:
 - uno schema E-R ristrutturato non è (più) uno schema concettuale nel senso stretto del termine, perchè costituisce una rappresentazione dei dati che tiene conto degli aspetti realizzativi (i.e. le prestazioni)



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!



Analisi delle prestazioni su schemi E-R

Consideriamo “**indicatori**” dei parametri che regolano le prestazioni

- **spazio:**
 - numero di occorrenze previste
- **tempo:**
 - numero di occorrenze (di entità e relazioni) visitate durante un'operazione

Per studiare questi parametri abbiamo bisogno di conoscere, oltre allo schema, le seguenti informazioni:

- **Volume dei dati:** (---> **tavola dei volumi**)
 - Numero di occorrenze di ogni entità e relazione dello schema
 - Dimensioni di ciascun attributo
- **Caratteristiche delle operazioni:** (---> **tavola delle operazioni**)
 - Tipo dell'operazione (interattiva o batch)
 - Frequenza (numero medio di esecuzioni in un certo intervallo di tempo)
 - Dati coinvolti (entità e/o associazioni)



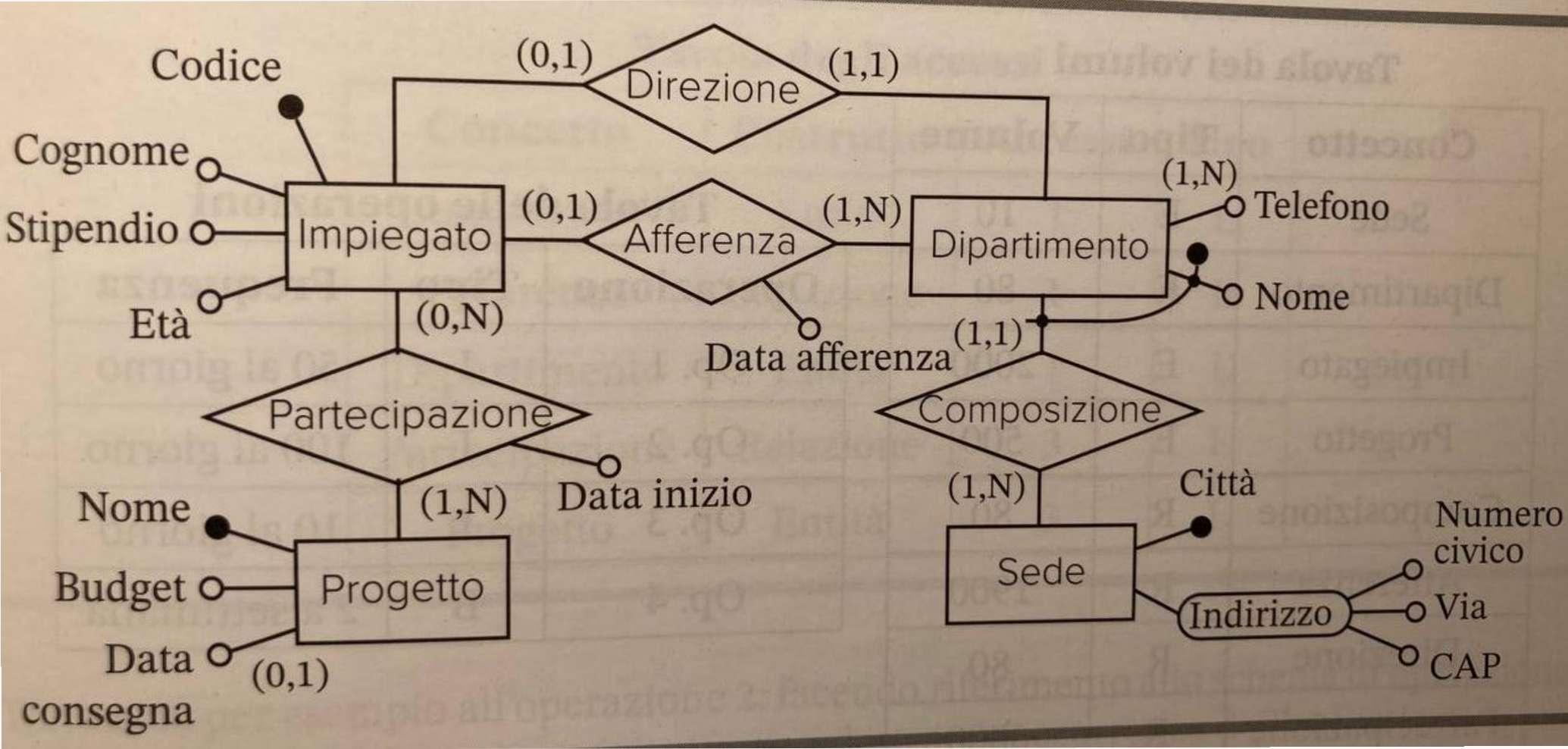


Tavola dei volumi

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

Attenzione ad alcuni aspetti:

- Vincoli 1-1:
 - Il numero di occorrenze nella relazione “Composizione” è pari al numero di Dipartimenti.
- Vincoli 0-1:
 - Il numero di occorrenze della relazione “Afferenza” è poco meno del numero degli impiegati.
- Vincoli 0-N:
 - Assumendo che un impiegato partecipi **in media** a 3 progetti, abbiamo $2000 \times 3 = 6000$ occorrenze per la relazione “Partecipazione”.

Esempio di valutazione di costo di un'operazione

Per ogni operazione, possiamo **descrivere graficamente** i dati coinvolti con uno ***schema di navigazione***:

- Consiste nel frammento di schema E-R interessato dall'operazione, sul quale viene disegnato il “**cammino logico**” da percorrere per accedere all'informazione di interesse.
- Esempio:
 - Operazione: “*trova tutti i dati di un impiegato, del dipartimento nel quale lavora e dei progetti ai quali partecipa*”
- Sulla base di questo, si costruisce una **tavola degli accessi** basata su uno **schema di navigazione**



Esempio di schema di navigazione

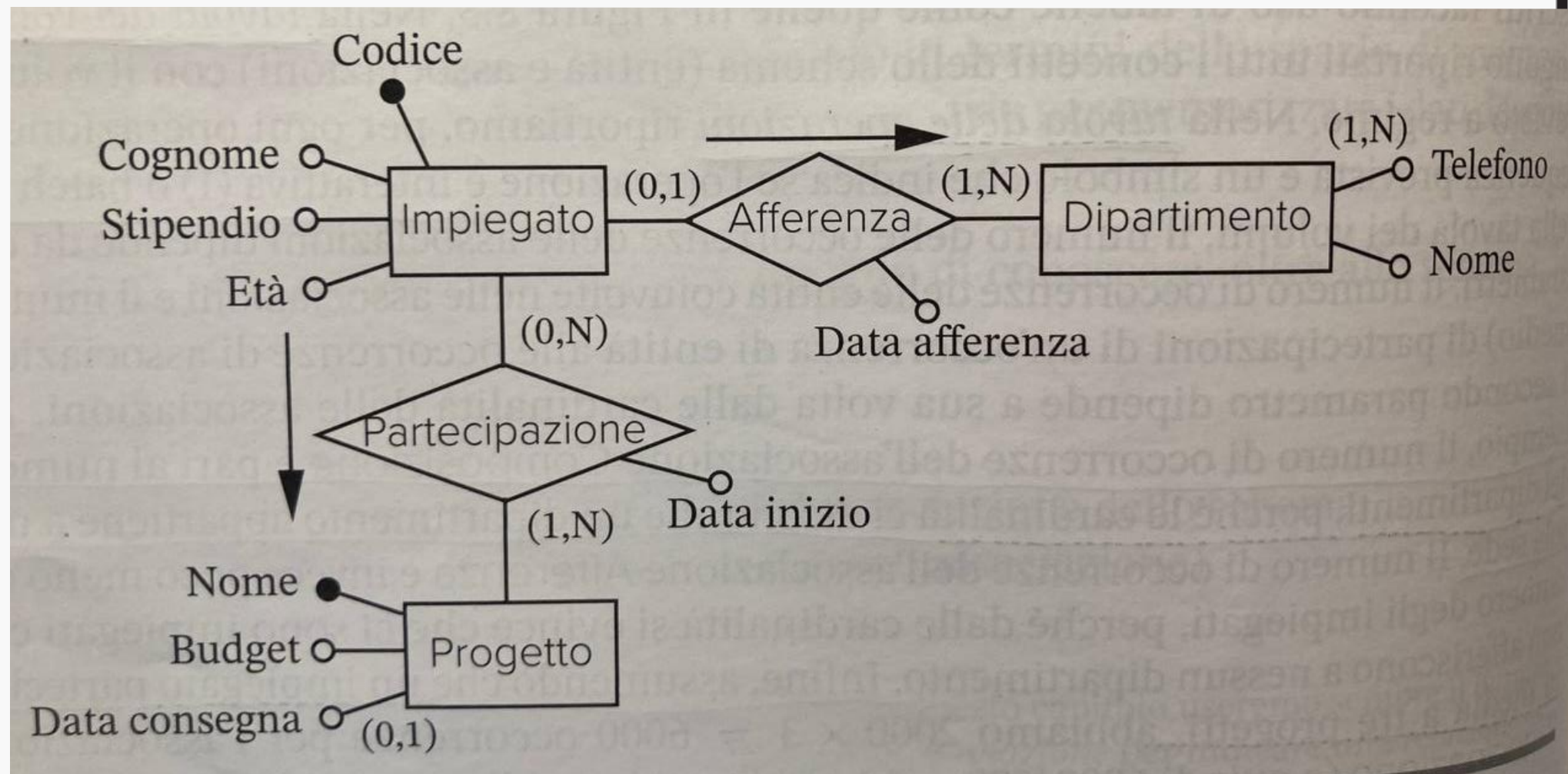


Tavola degli accessi

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



Perchè abbiamo detto che **in media** un impiegato lavora su tre progetti.

Attenzione: il numero degli accessi è contato come il numero di istanze dell’entità o della relazione che si vuole recuperare. **NON** si devono fare considerazioni del tipo: “devo fare uno scan di tutta la tabella ecc.”: i DBMS implementano meccanismi sofisticati che permettono di evitare questo.

Attenzione: supponiamo, da qui in avanti, che una Lettura costi 1 unità di costo (i.e. un accesso) e che una Scrittura costi 2 unità di costo (i.e. 2 accessi).

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
 - Pericolo di mantenere l'integrità
- 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
 - **integrità dei dati:** ci vogliono delle operazioni aggiuntive per mantenere il dato derivato aggiornato.

La fase di **analisi delle ridondanze** risponde alla domanda “*conviene o meno tenere/eliminare una ridondanza*”?

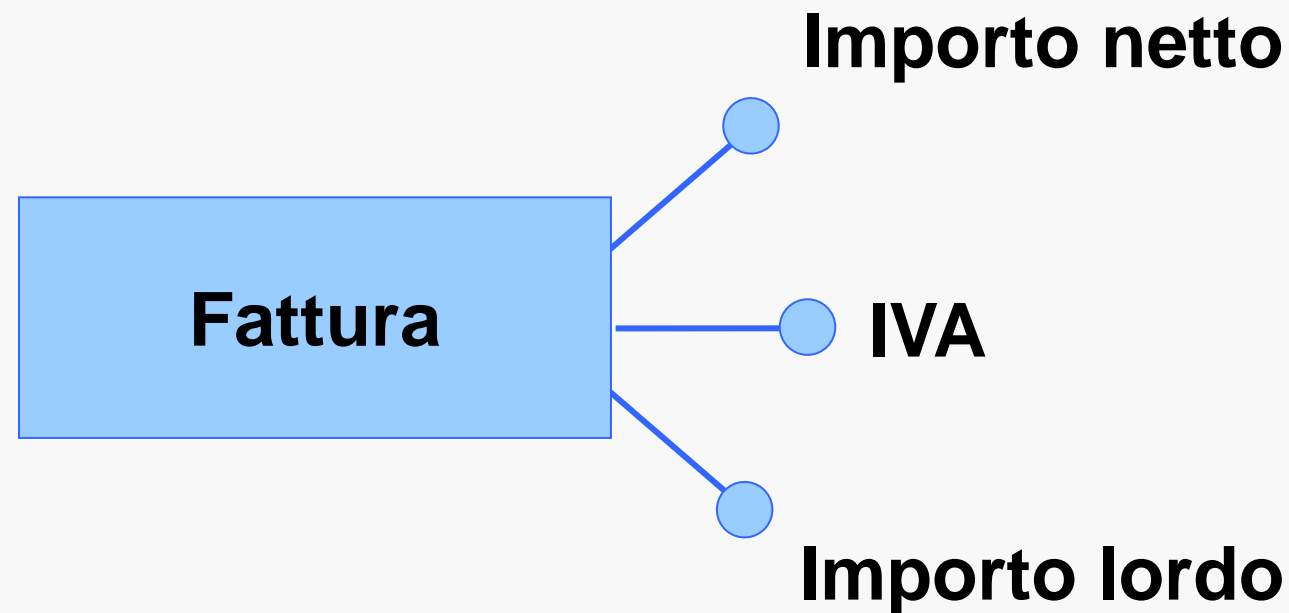


Forme di ridondanza in uno schema E-R

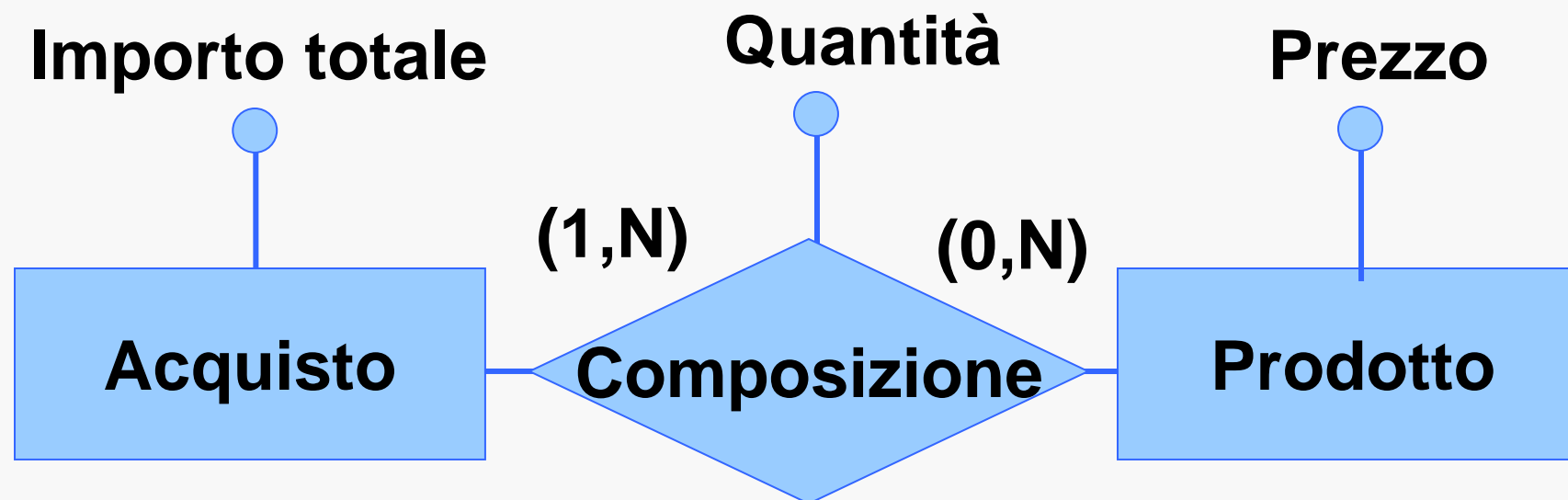
- attributi derivabili:
 - da altri attributi della stessa entità (o relazioni)
 - da attributi di altre entità (o relazioni), **di solito tramite funzioni aggregative**
 - da operazioni di **conteggio di occorrenze**.
- relazioni derivabili dalla composizione di altre (più in generale: cicli di relazioni)



Attributo derivabile da attributi della stessa entità

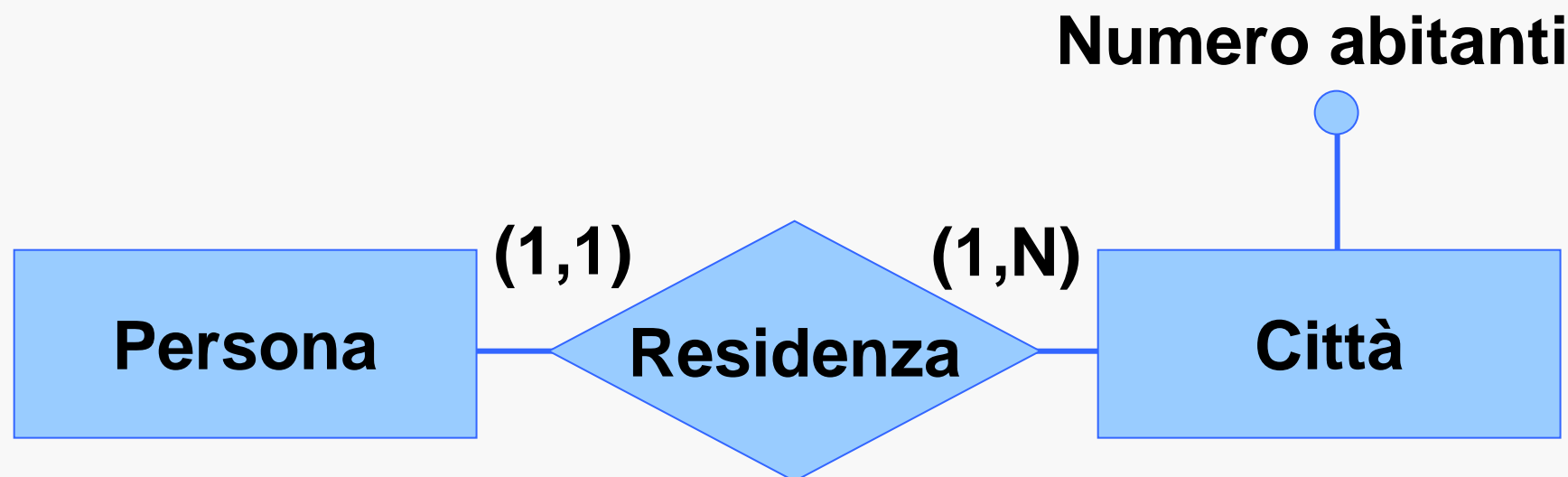


Attributo derivabile da altra entità (funzioni aggregative)



L'importo totale dell'entità "Acquisto" si può derivare, attraverso la relazione "Composizione", dall'attributo Prezzo dell'entità "Prodotto", **sommando i prezzi dei prodotti di cui un acquisto è composto.**

Attributo derivabile da altra entità (operazioni di conteggio)

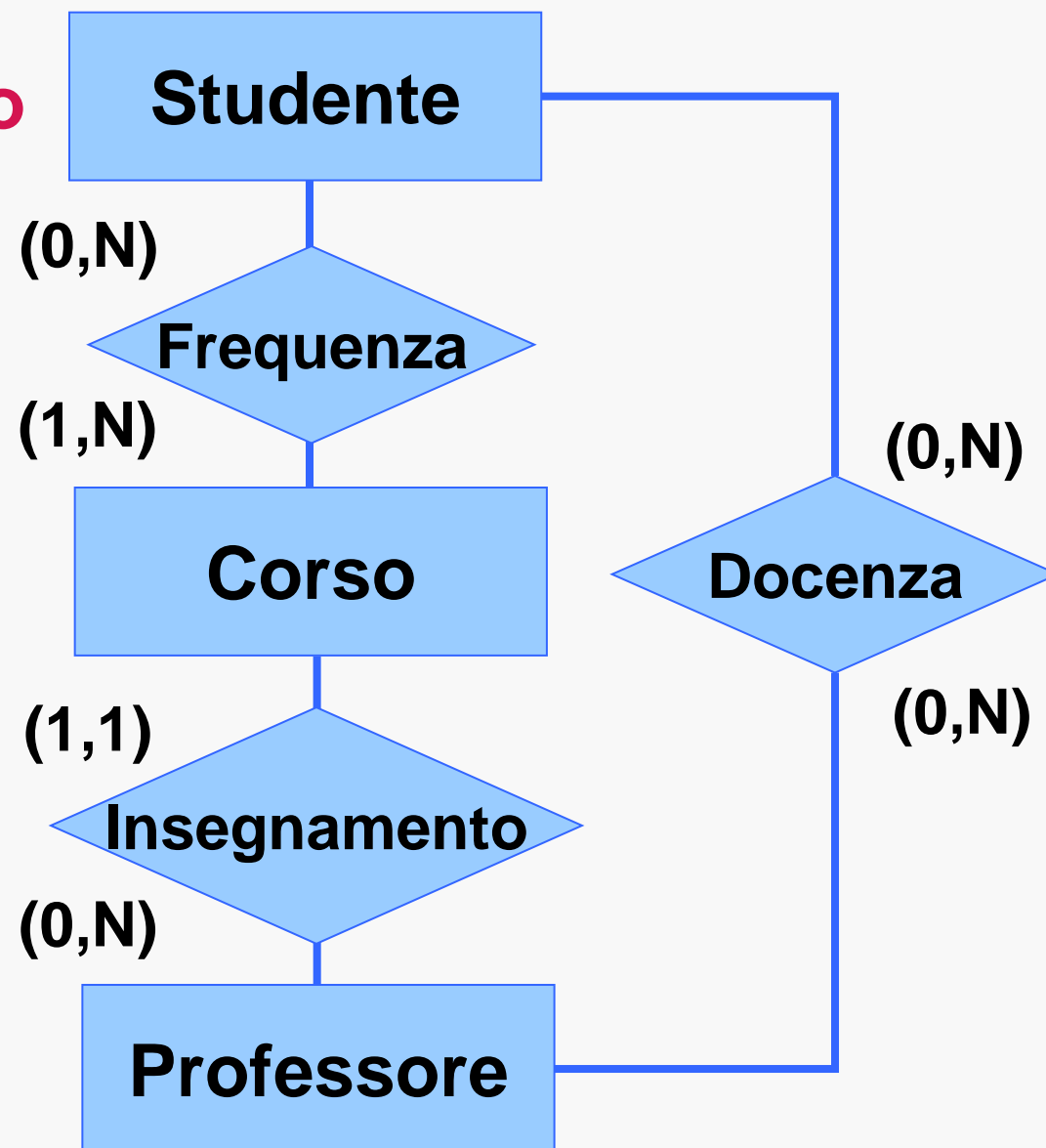


Ridondanza dovuta a ciclo

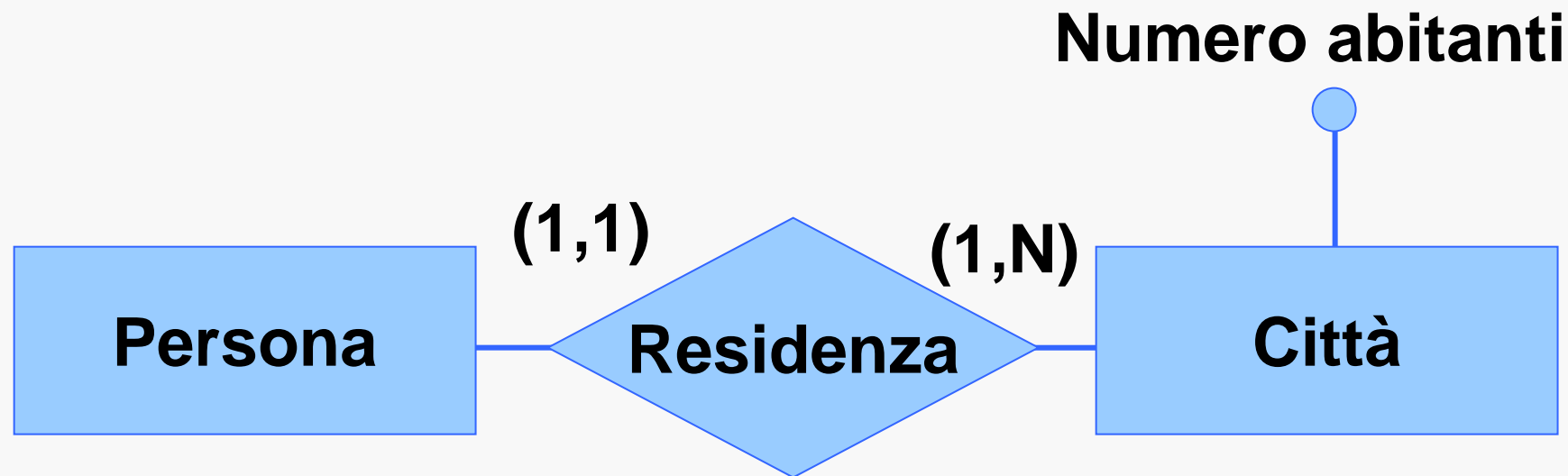
La relazione “Docenza” tra studenti e professori può essere derivata dalle relazioni “Frequenza” e “Insegnamento”.

Va comunque precisato che la presenza di **cicli** non genera necessariamente ridondanze.

- e.g. se al posto di “Docenza” ci fosse stata una relazione “Tesi”, allora lo schema non sarebbe stato ridondante.



Analisi di una ridondanza



Vediamo ora come eseguire l'**analisi delle ridondanze** su questo esempio.

Tavola dei volumi

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

Tavola delle 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**)



Analisi delle ridondanze

A questo punto, valutiamo le prestazioni in caso di:

- 1) Presenza del dato ridondante
- 2) Assenza del dato ridondante

L'analisi delle ridondanze offre un modo **quantitativo e formale** (confronto tra due numeri reali) per decidere quando conviene o meno tenere/eliminare la ridondanza.





Tavole degli accessi in presenza di ridondanza

- Memorizzo la nuova persona
- Memorizzo la coppia persona-città
- Cerco la città di interesse
- Incremento di 1 il numero di abitanti di quella città

Il tutto ripetuto per 500 volte/giorno.

Totale: 500 x 3 accessi in scrittura e 500 x 1 accessi in lettura.

Totale: $(500 \times 3) \times 2 + (500 \times 1) \times 1 = \mathbf{3500}$ accessi

Trascurabile perchè richiede solo un accesso in lettura all'entità "Città" (per leggere l'attributo "numero di abitanti"). Da ripetere 2 volte/giorno.

Totale: $(2 \times 1) \times 1 = \mathbf{2}$ accessi

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 degli accessi in assenza di ridondanza

Non c'è bisogno di accedere all'entità "Città" per aggiornare il dato derivato.

Totale: $(500 \times 2) \times 2 = \mathbf{2000}$ accessi

Per calcolare il numero di abitanti dobbiamo accedere alla relazione "Residenza" un numero **medio** volte pari al numero **medio di residenti in una città** =

$$\begin{aligned} &\text{volume di "Persona" / volume di "Città"} \\ &= 1.000.000 / 200 = \mathbf{5000} \end{aligned}$$

Totale: $(2 \times 5001) \times 1 = 10.002$

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



Presenza di ridondanza

Costo:

- OP 1: 3500
- OP 2: 2

Totale: 3502

Assenza di ridondanza

Costo:

- OP 1: 2000
- OP 2: 10.002

Totale: 12.002

Ci sono **8500** accessi in più (in media) nel caso in cui la ridondanza non sia presente.

==> in questo caso **conviene mantenere la ridondanza**

Questo è dovuto al fatto che gli accessi in lettura necessari per calcolare il dato derivato sono molti di più rispetto agli accessi in scrittura necessari per mantenerlo aggiornato.



Attività della ristrutturazione

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



Eliminazione delle gerarchie

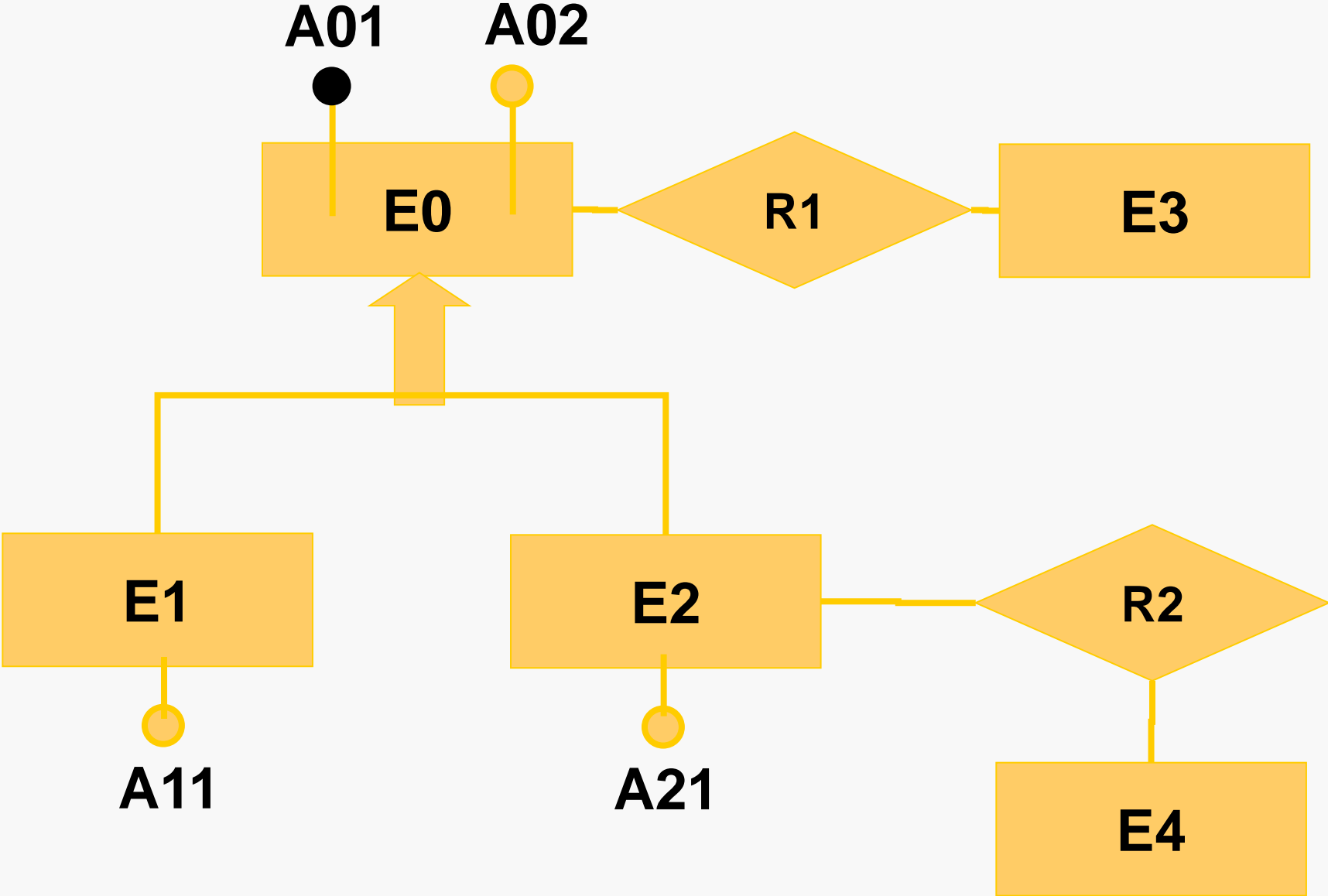
- il modello relazionale non può rappresentare direttamente le generalizzazioni
 - entità e relationship sono invece direttamente rappresentabili
-
- si eliminano perciò le gerarchie, sostituendole con entità e relationship

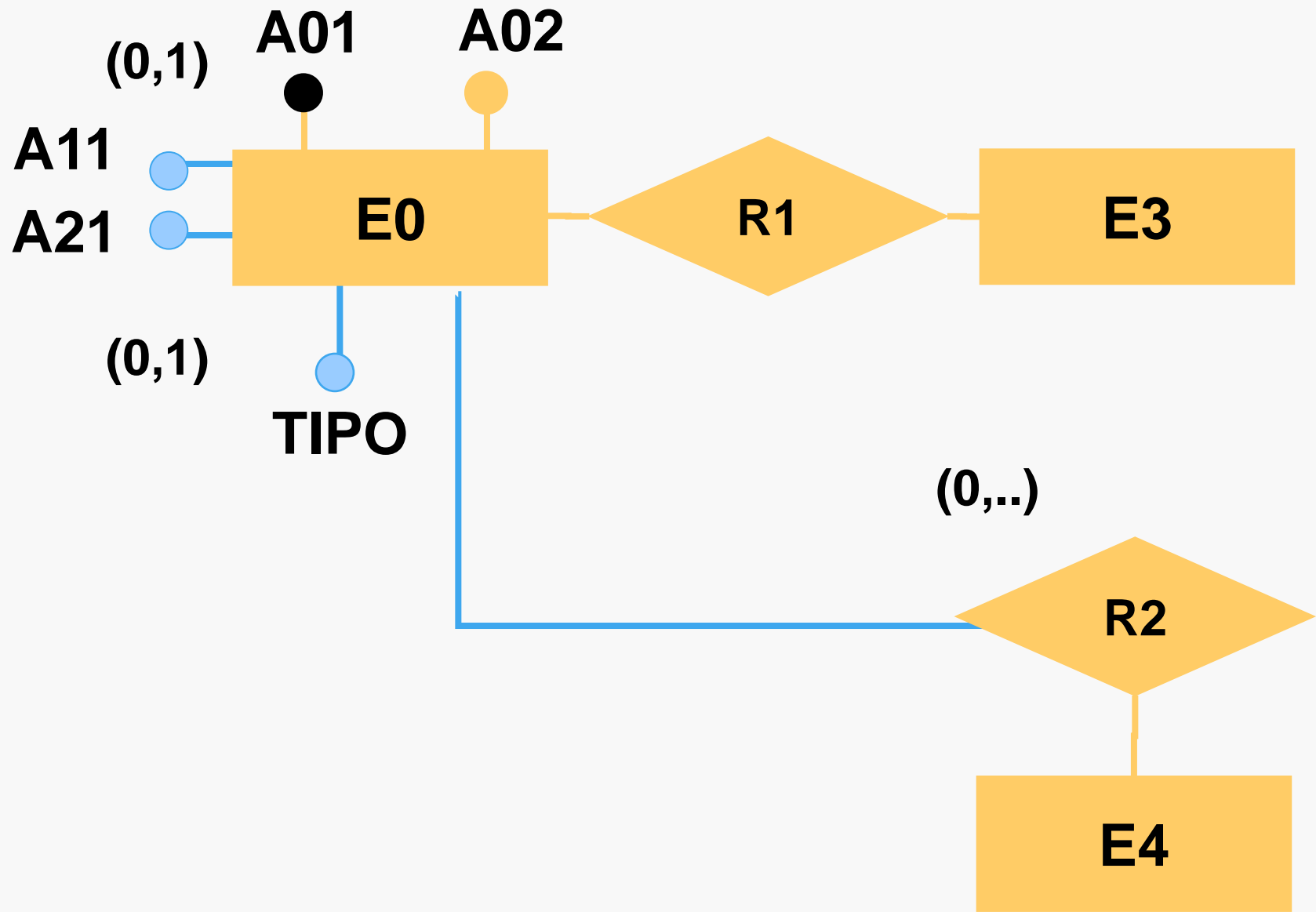


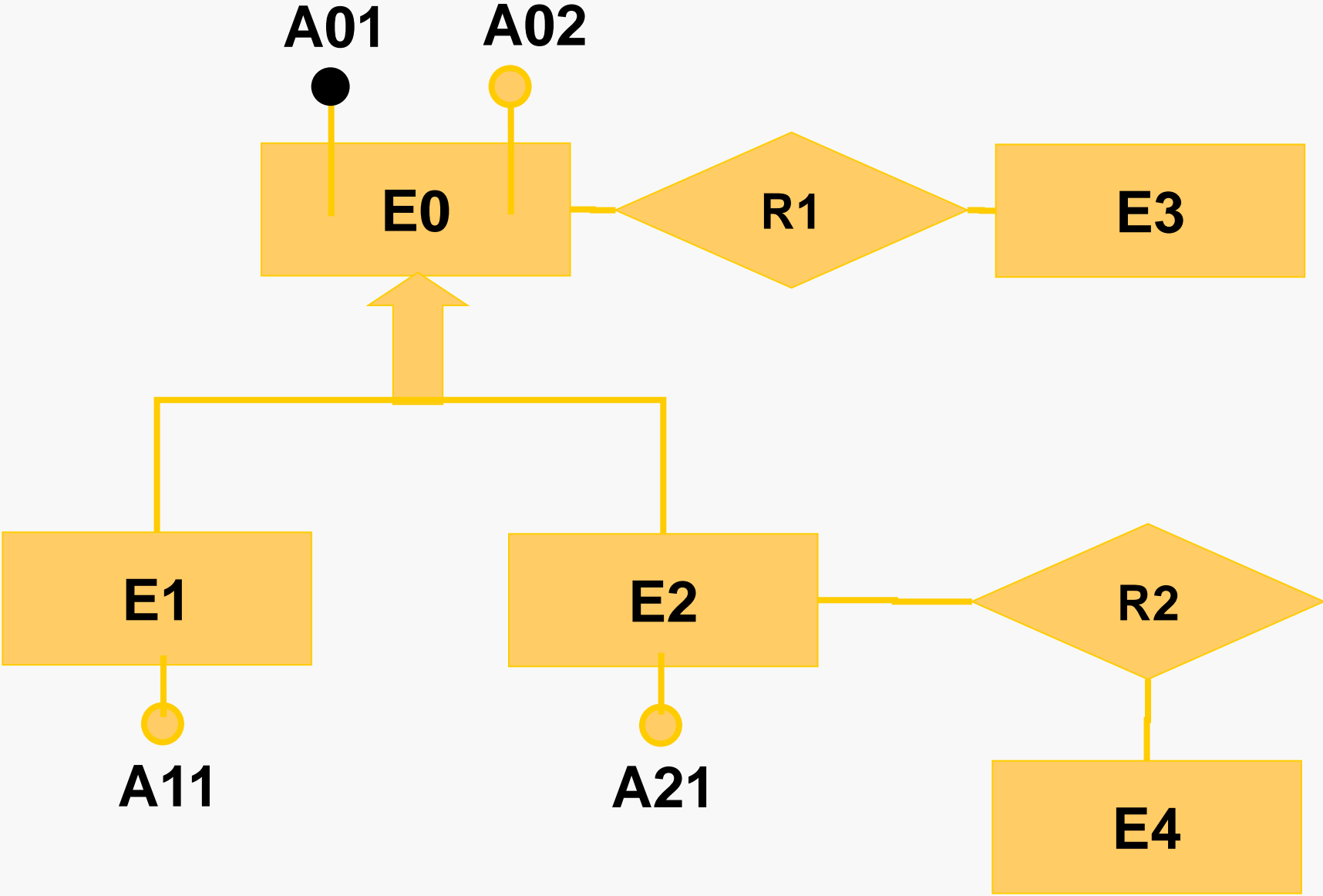
Tre possibilità

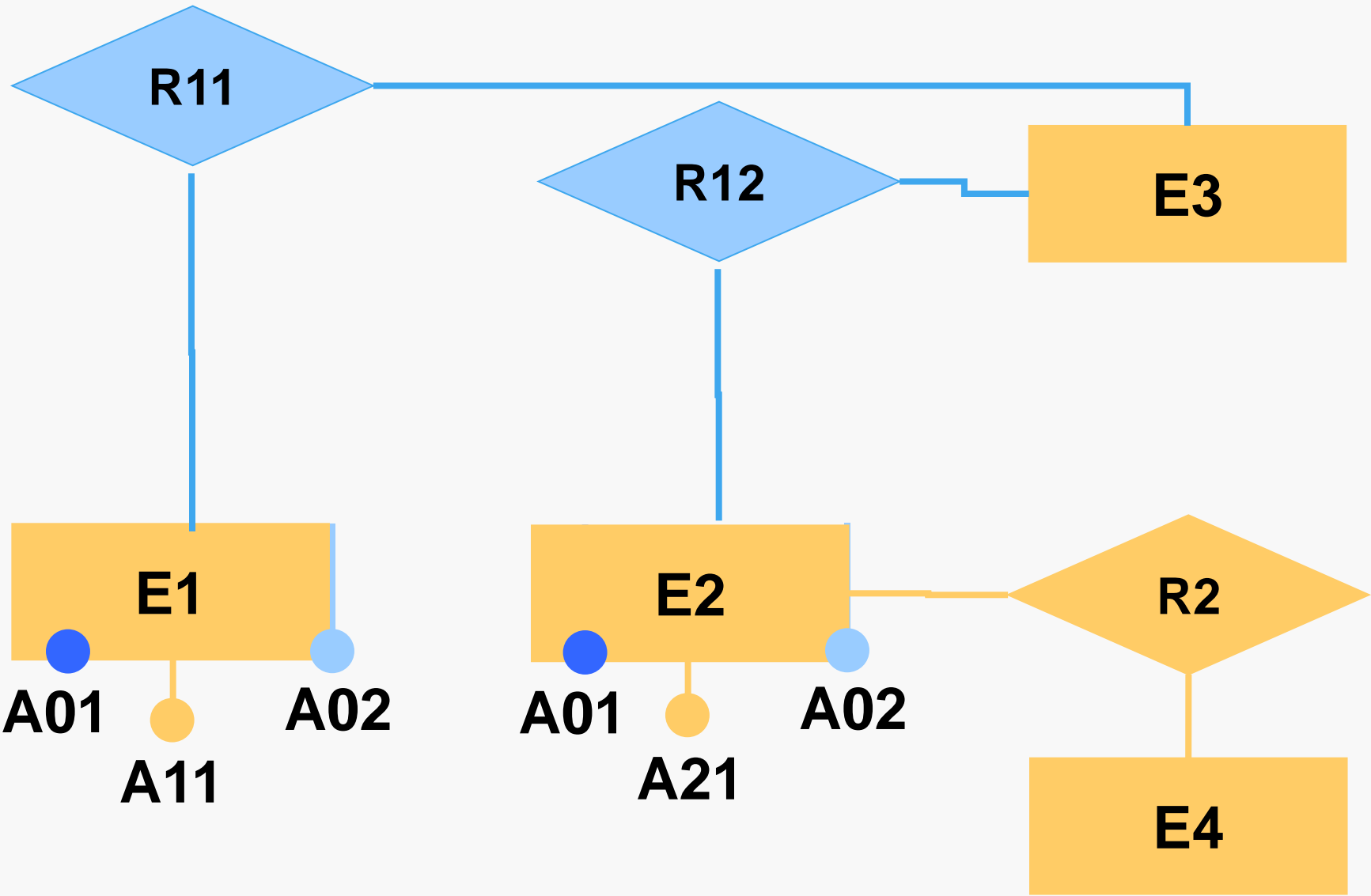
1. accorpamento delle figlie della generalizzazione nel genitore
2. accorpamento del genitore della generalizzazione nelle figlie
3. sostituzione della generalizzazione con relationship

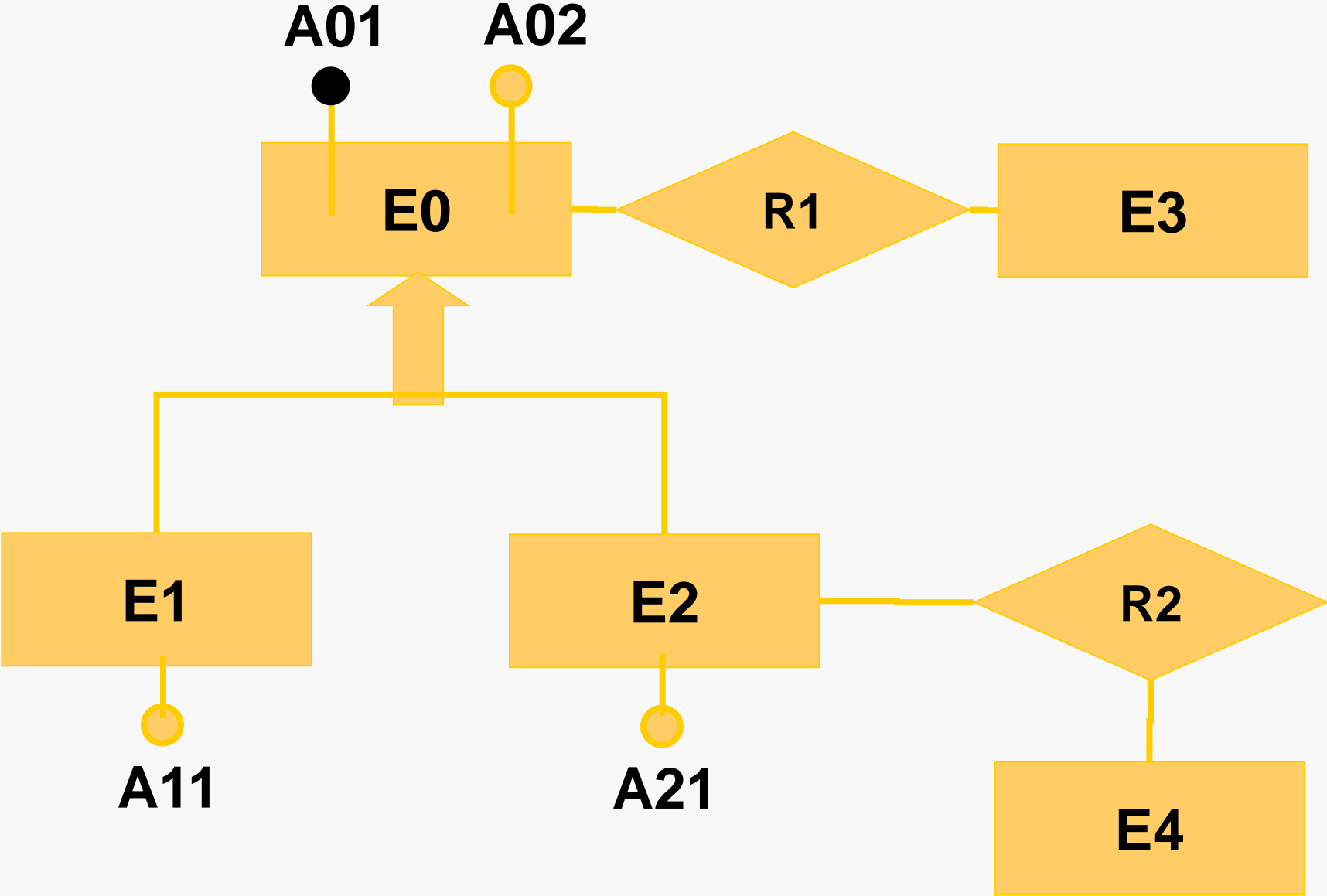


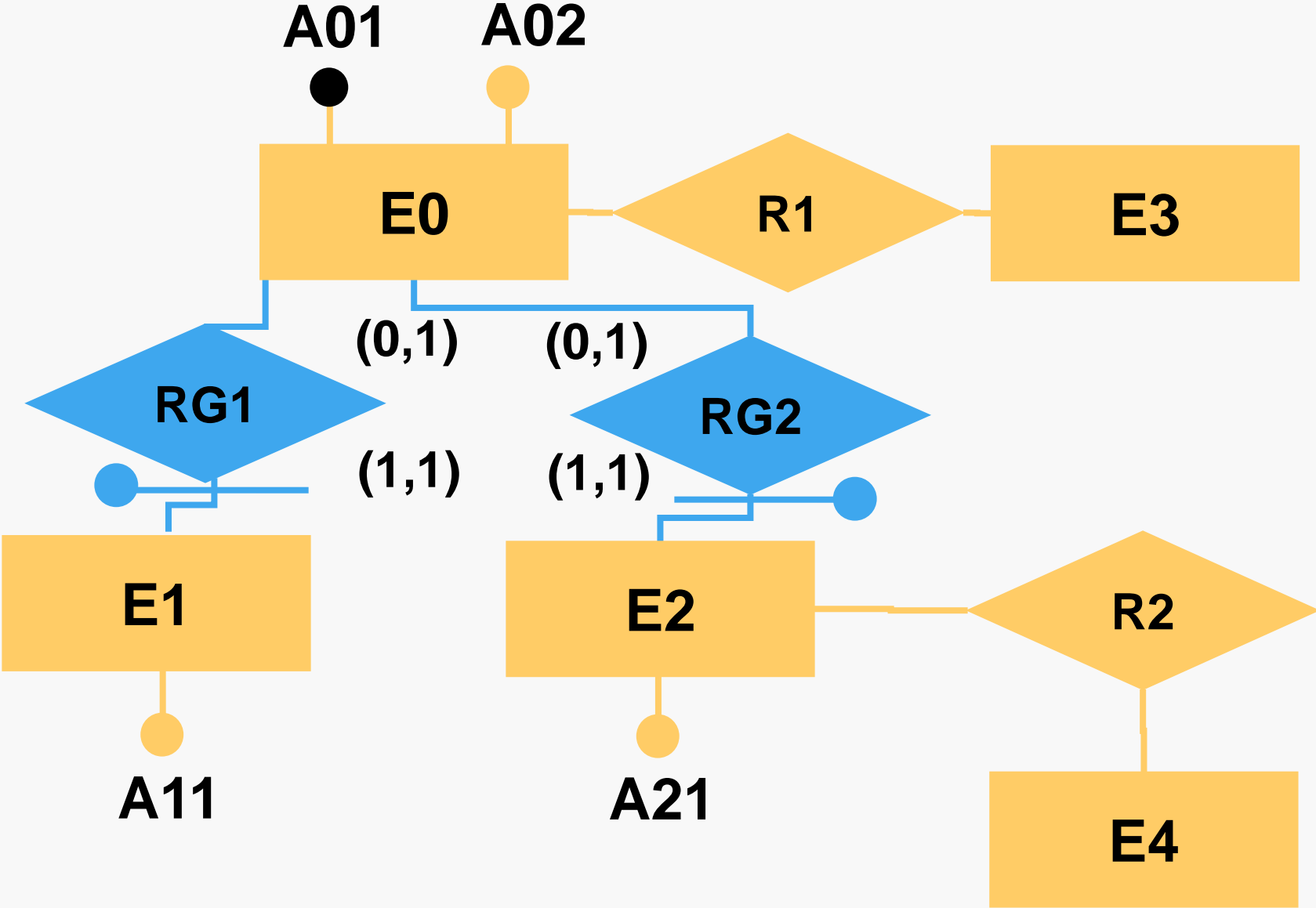






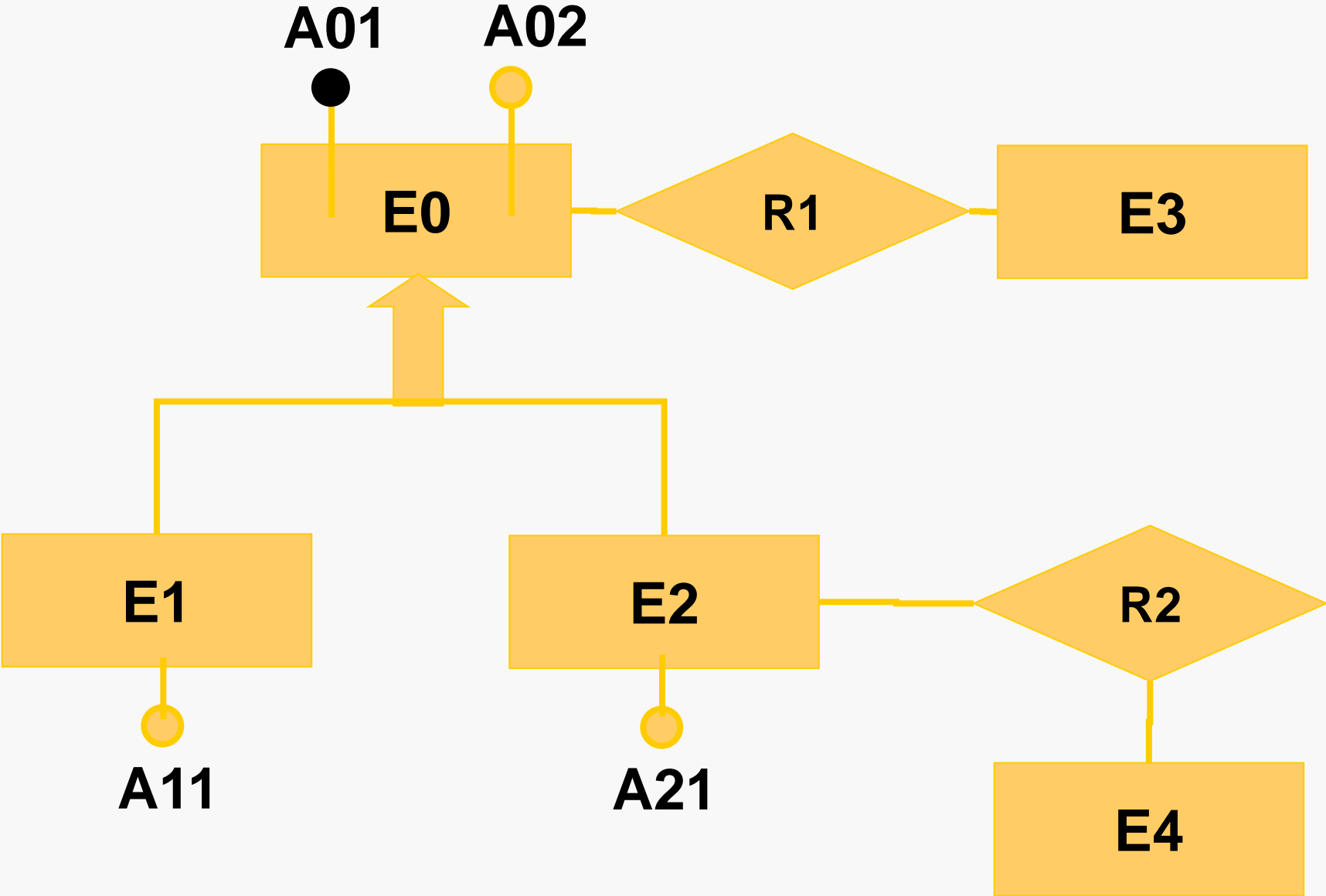


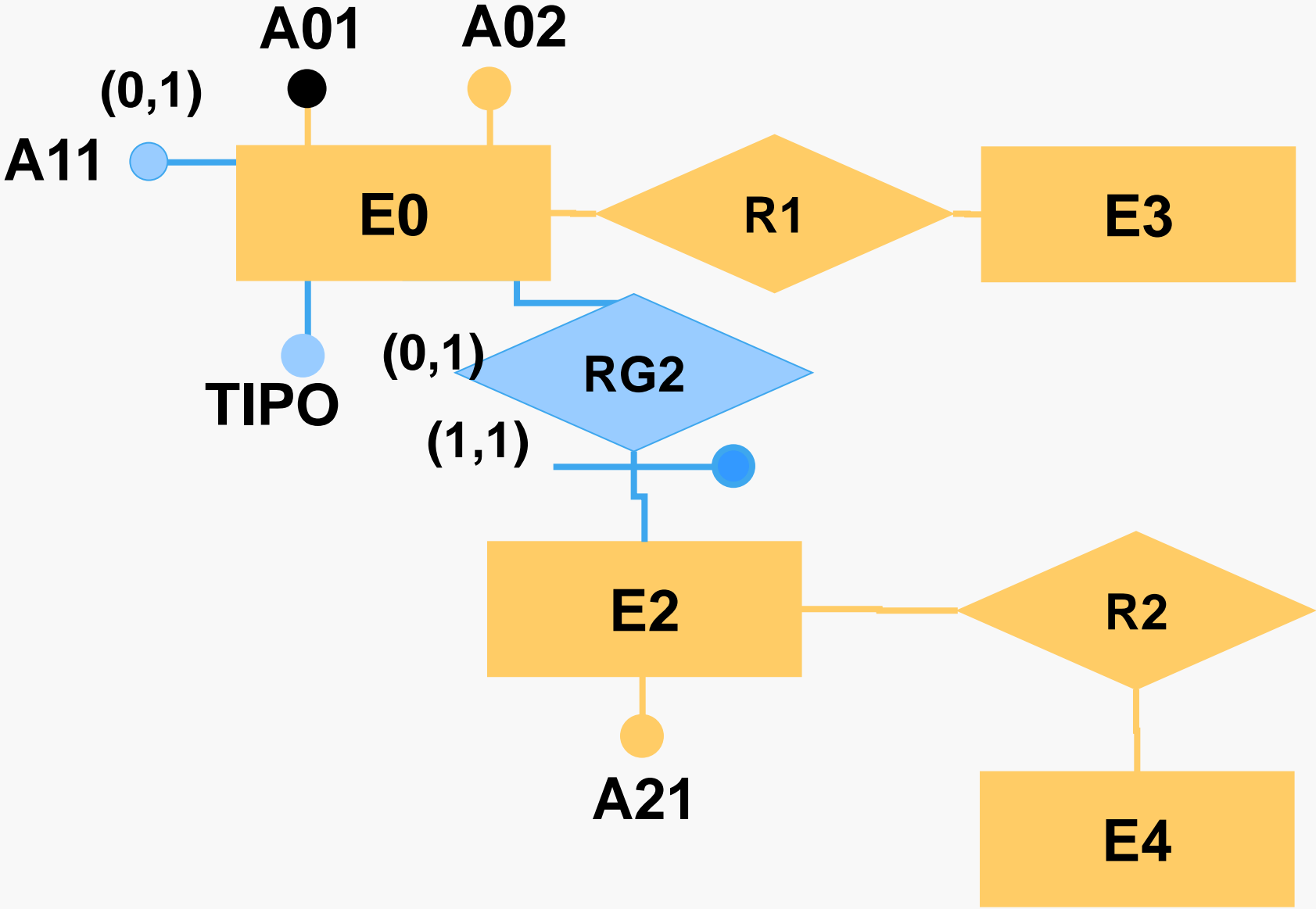




- la scelta fra le alternative si può fare con metodo simile a quello visto per l'analisi delle ridondanze (però non basato solo sul numero degli accessi)
 - è possibile seguire alcune semplici regole generali

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





Attività della ristrutturazione

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

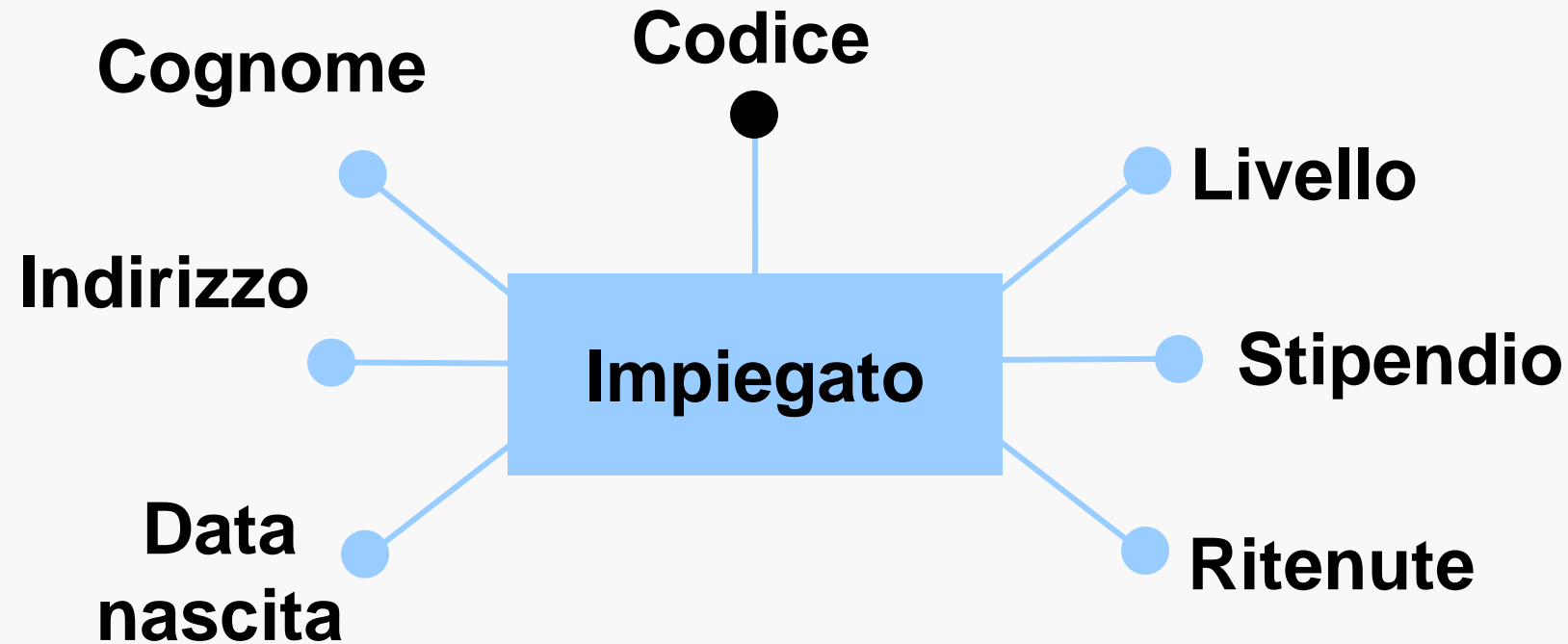


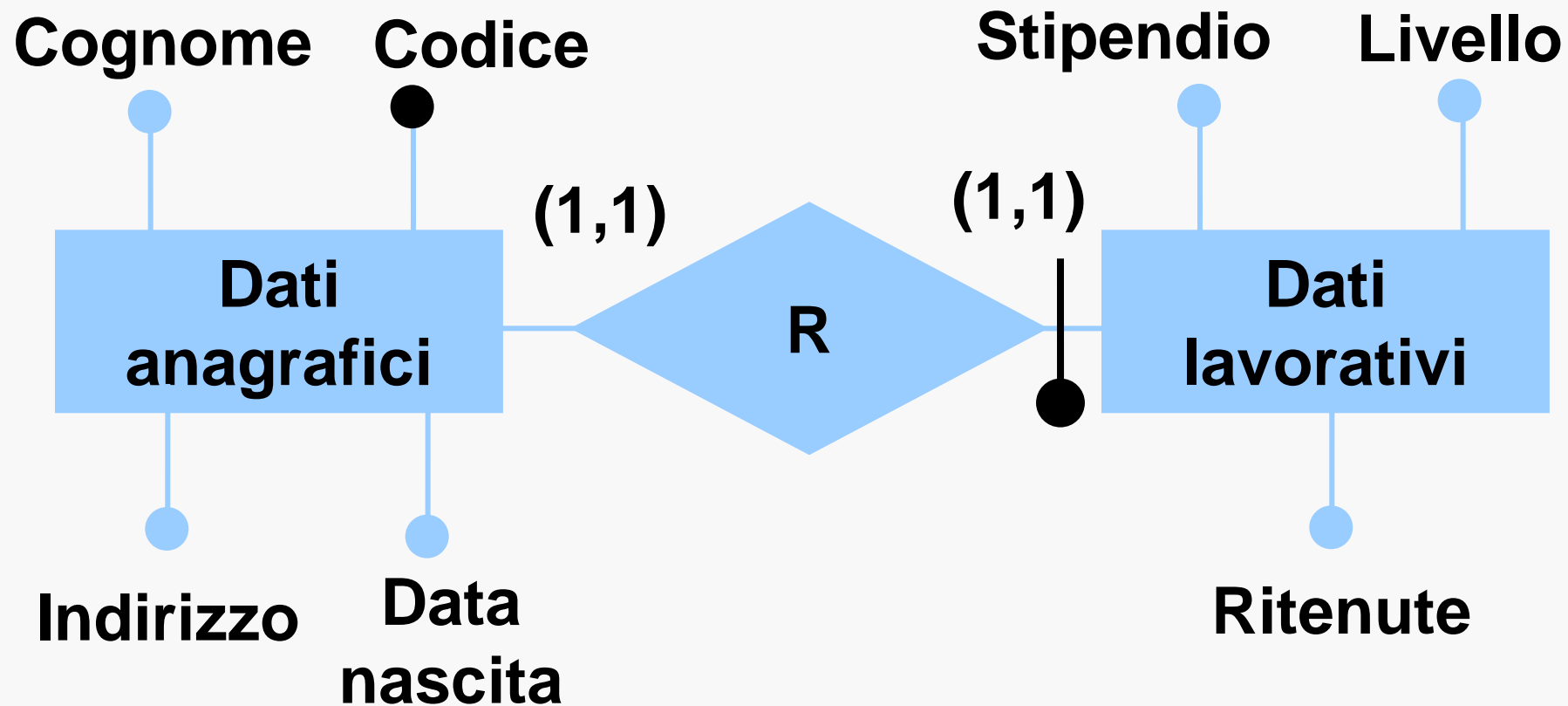
- Ristrutturazioni effettuate per rendere più efficienti le operazioni in base a un semplice principio
- Gli accessi si riducono:
 - separando attributi di un concetto che vengono acceduti separatamente
 - raggruppando attributi di concetti diversi acceduti insieme

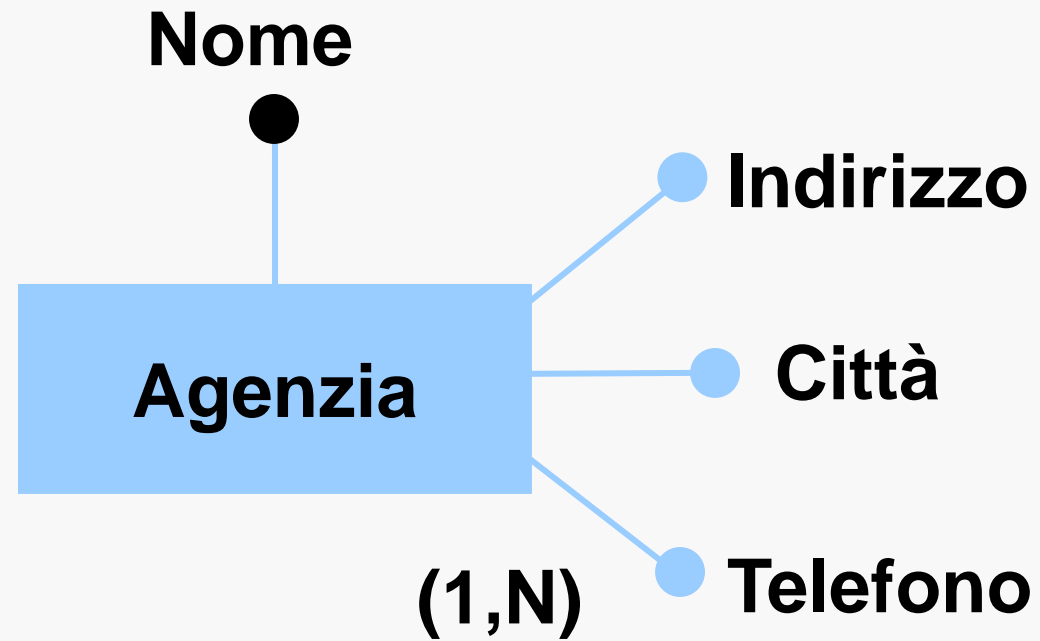
Ristrutturazioni, casi principali

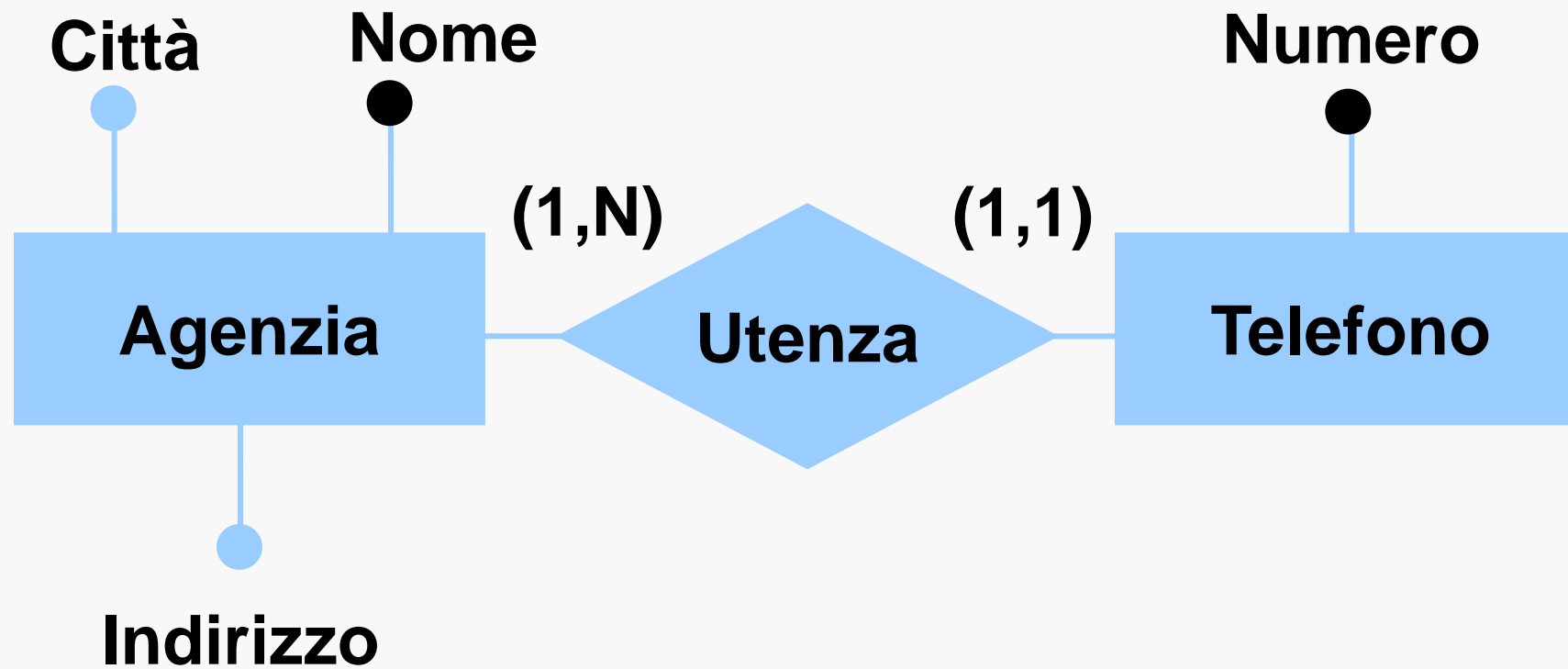
- partizionamento verticale di entità
- partizionamento orizzontale di relationship
- eliminazione di attributi multivalore
- accorpamento di entità/ relationship

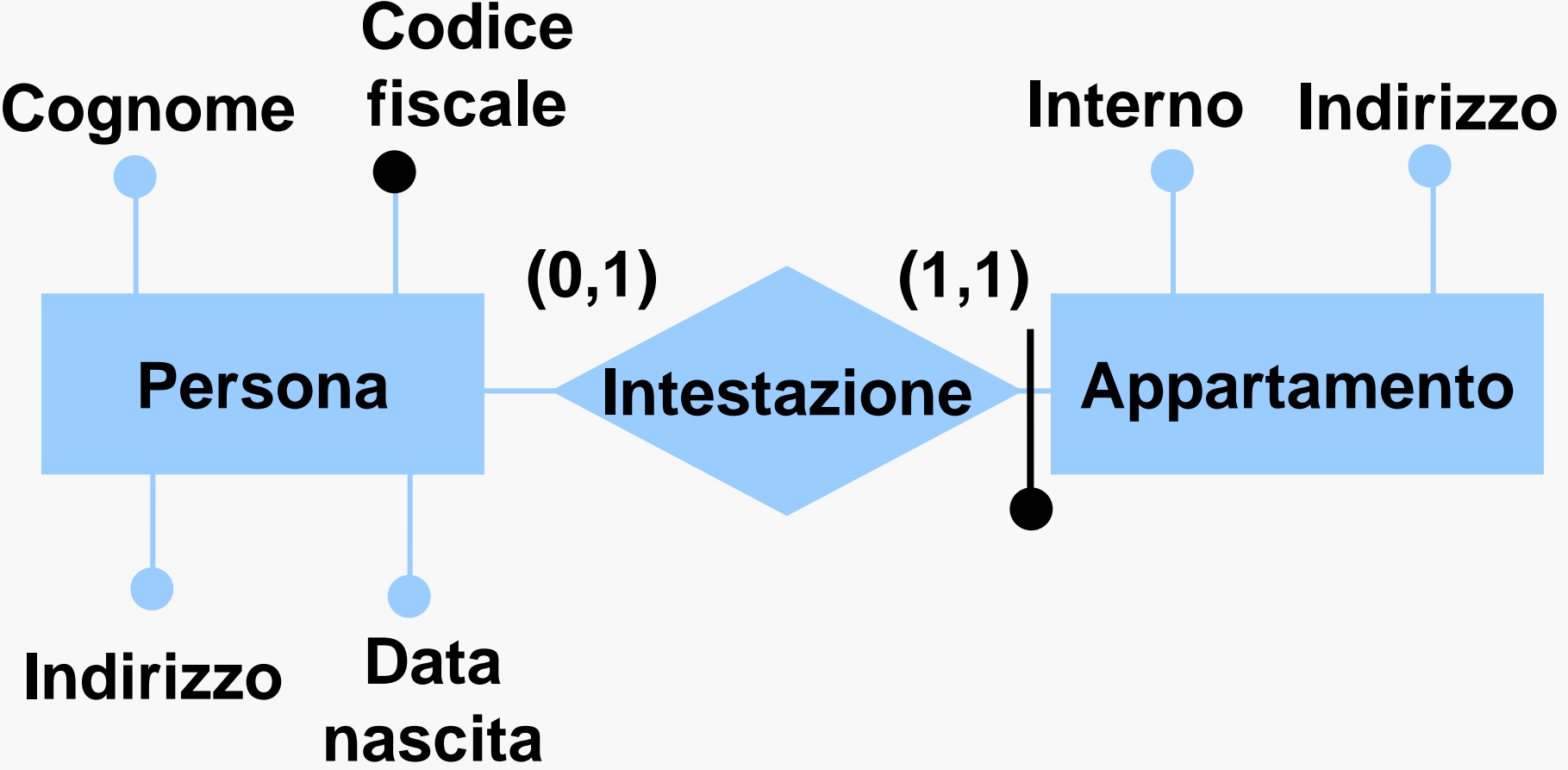


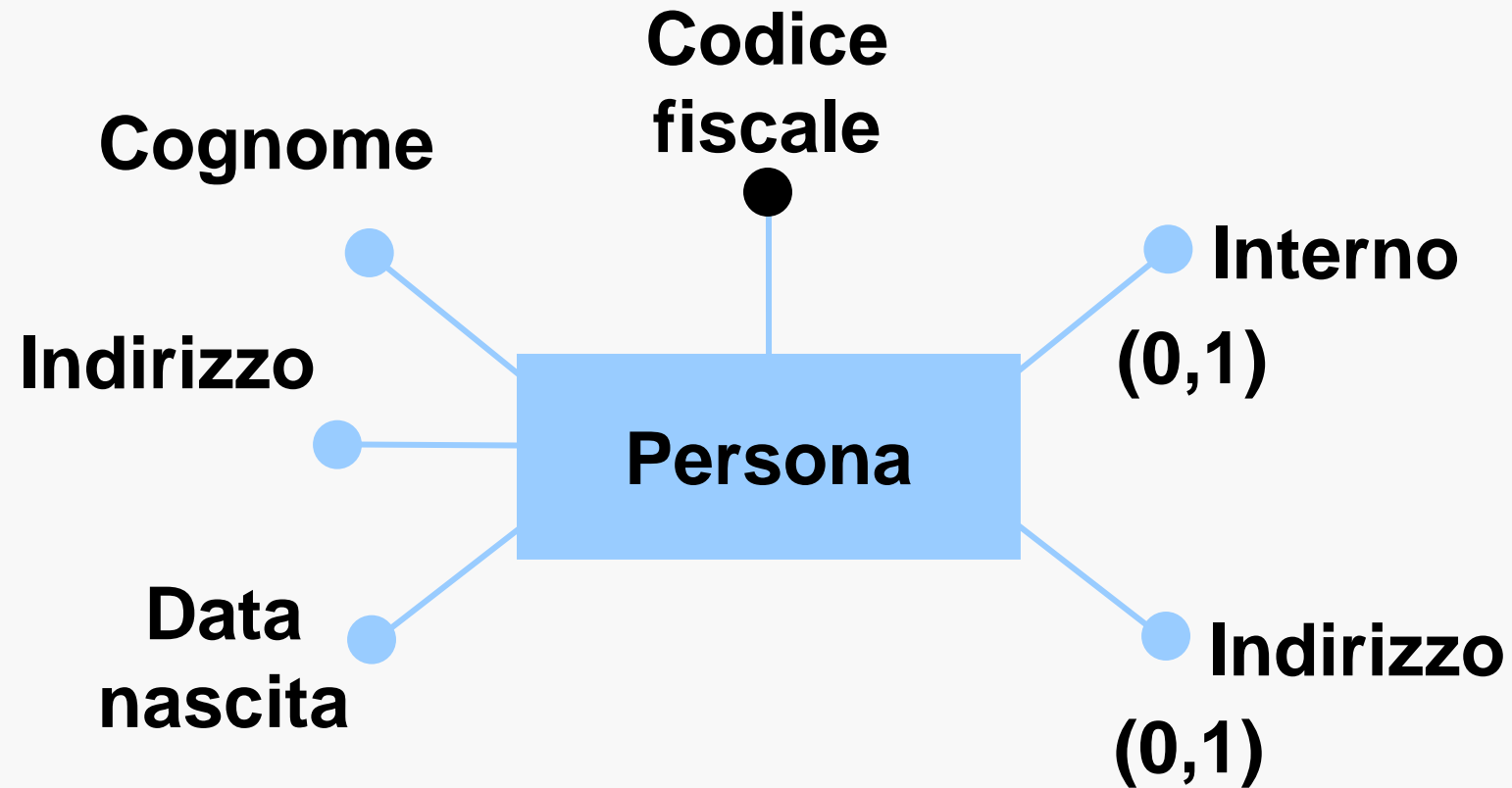


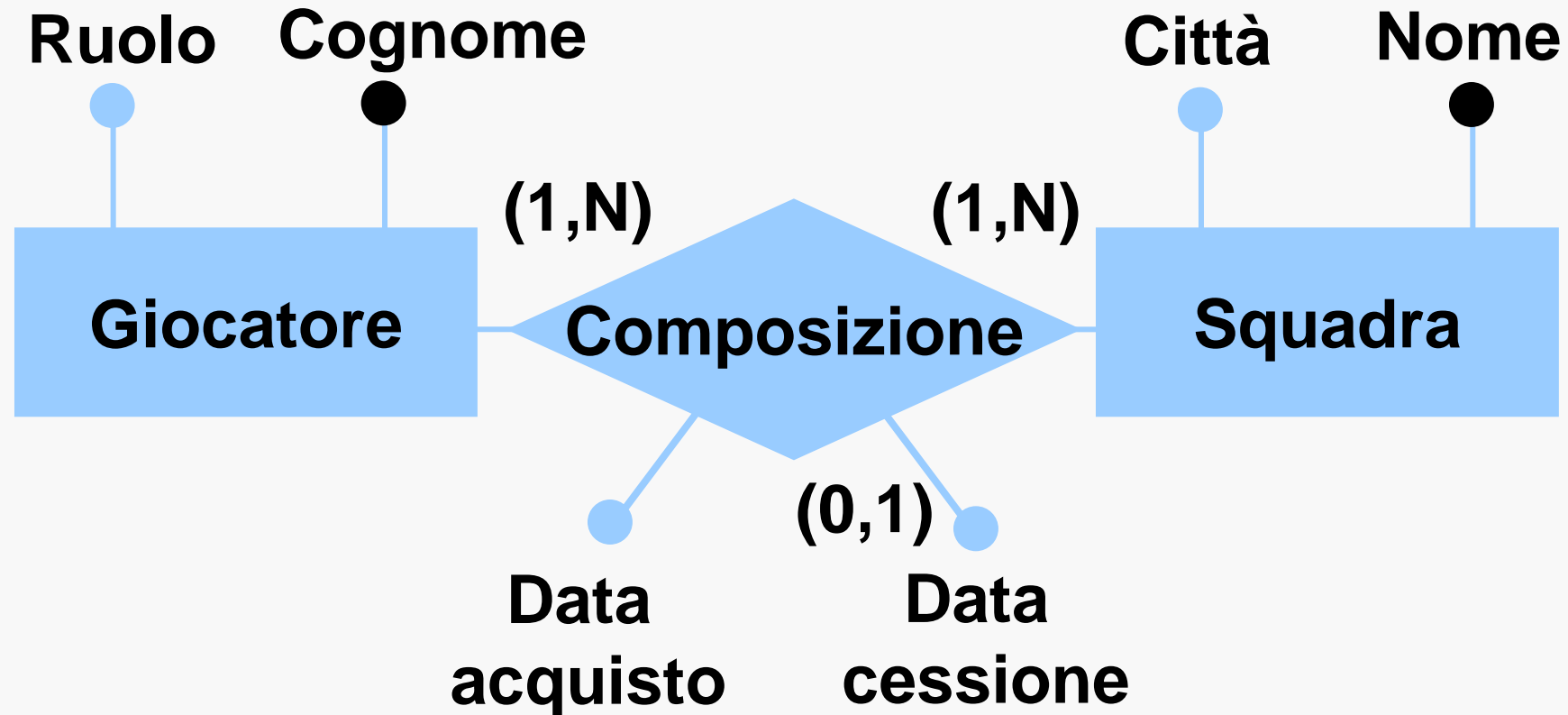


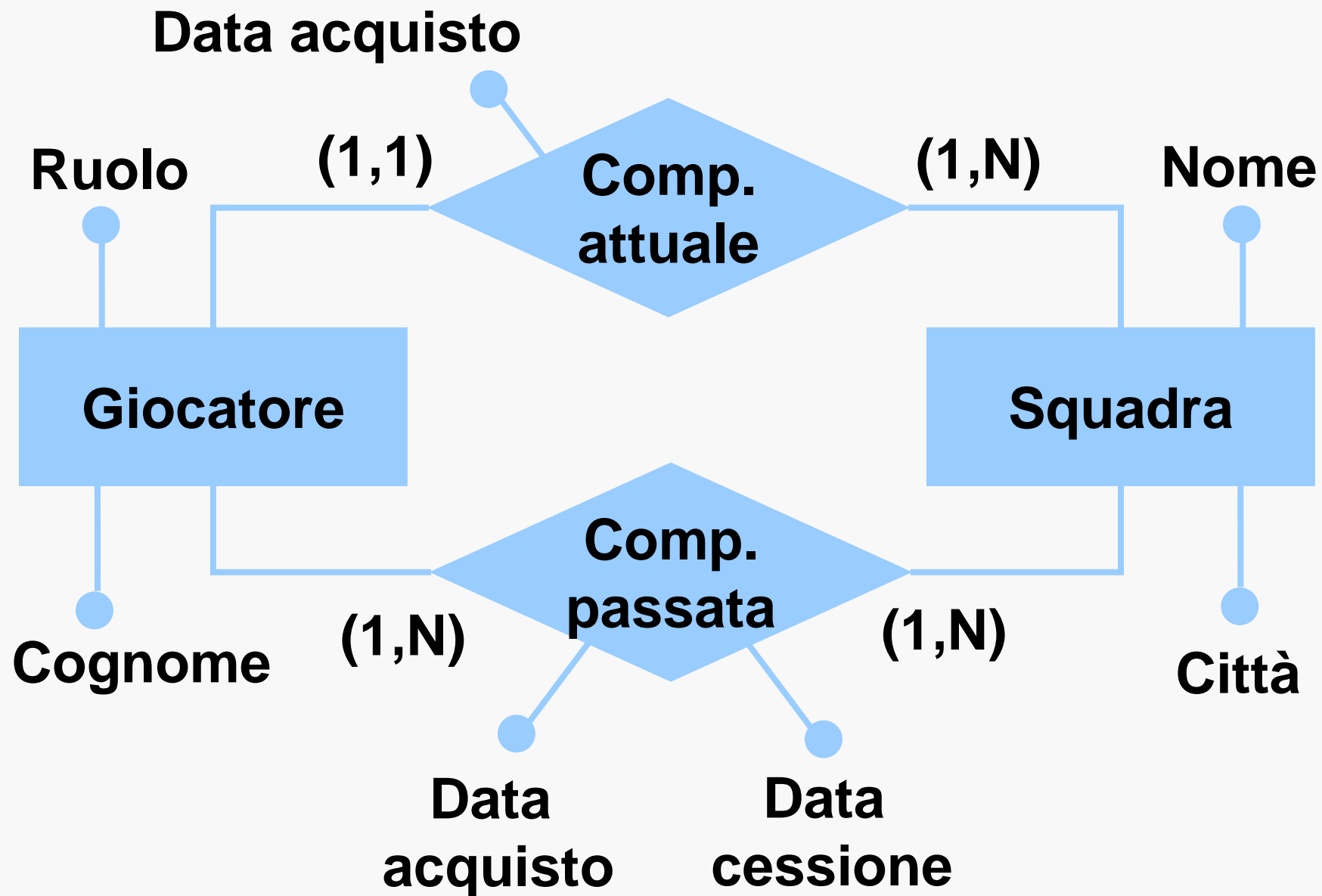












Attività della ristrutturazione

- 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?

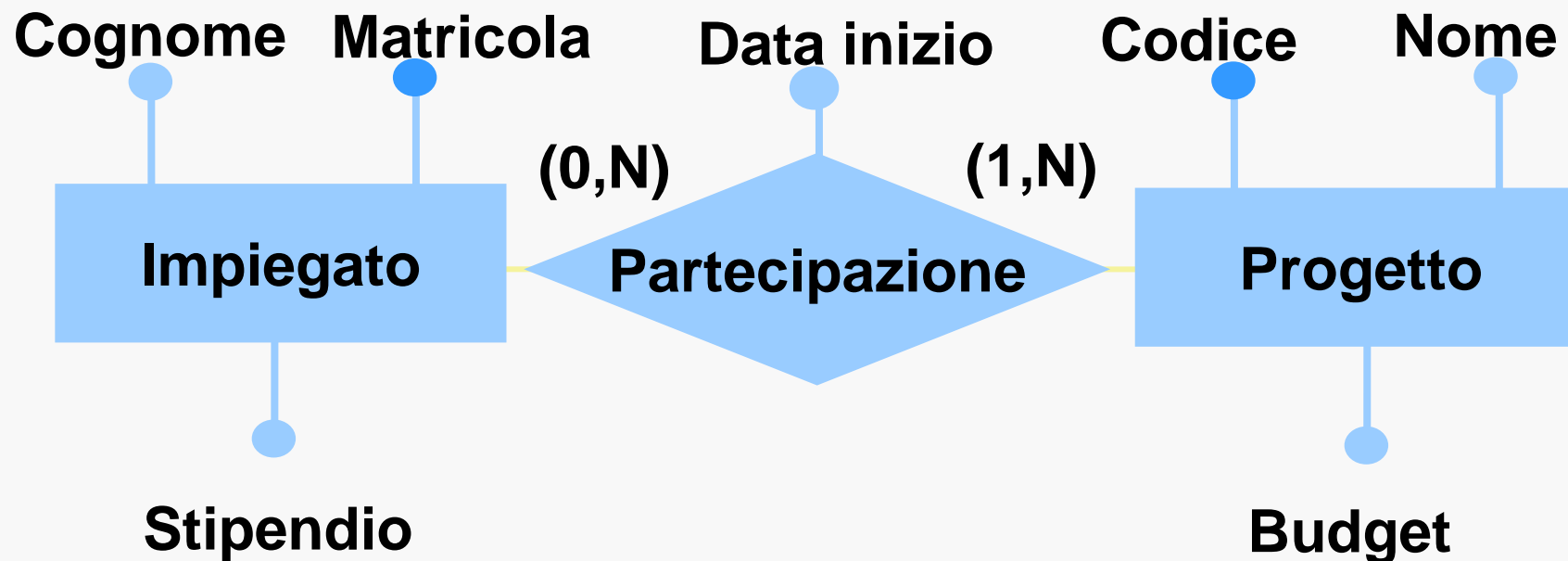
Si introducono nuovi attributi (codici**) contenenti valori speciali generati appositamente per questo scopo**

Traduzione verso il modello relazionale

- idea di base:
 - le entità diventano relazioni sugli stessi attributi
 - le relationship diventano relazioni sugli identificatori delle entità coinvolte (più gli attributi propri)



Entità e relationship molti a molti



Impiegato(Matricola, Cognome, Stipendio)

Progetto(Codice, Nome, Budget)

Partecipazione(Matricola, Codice, DataInizio)

Entità e relationship molti a molti

Impiegato(Matricola, Cognome, Stipendio)

Progetto(Codice, Nome, Budget)

Partecipazione(Matricola, Codice, DataInizio)

- con vincoli di integrità referenziale fra
 - Matricola in Partecipazione e (la chiave di) Impiegato
 - Codice in Partecipazione e (la chiave di) Progetto



Nomi più espressivi per gli attributi della chiave della relazione che rappresenta la relationship

Impiegato(Matricola, Cognome, Stipendio)

Progetto(Codice, Nome, Budget)

Partecipazione(Matricola, Codice, DataInizio)

Partecipazione(Impiegato, Progetto, DataInizio)

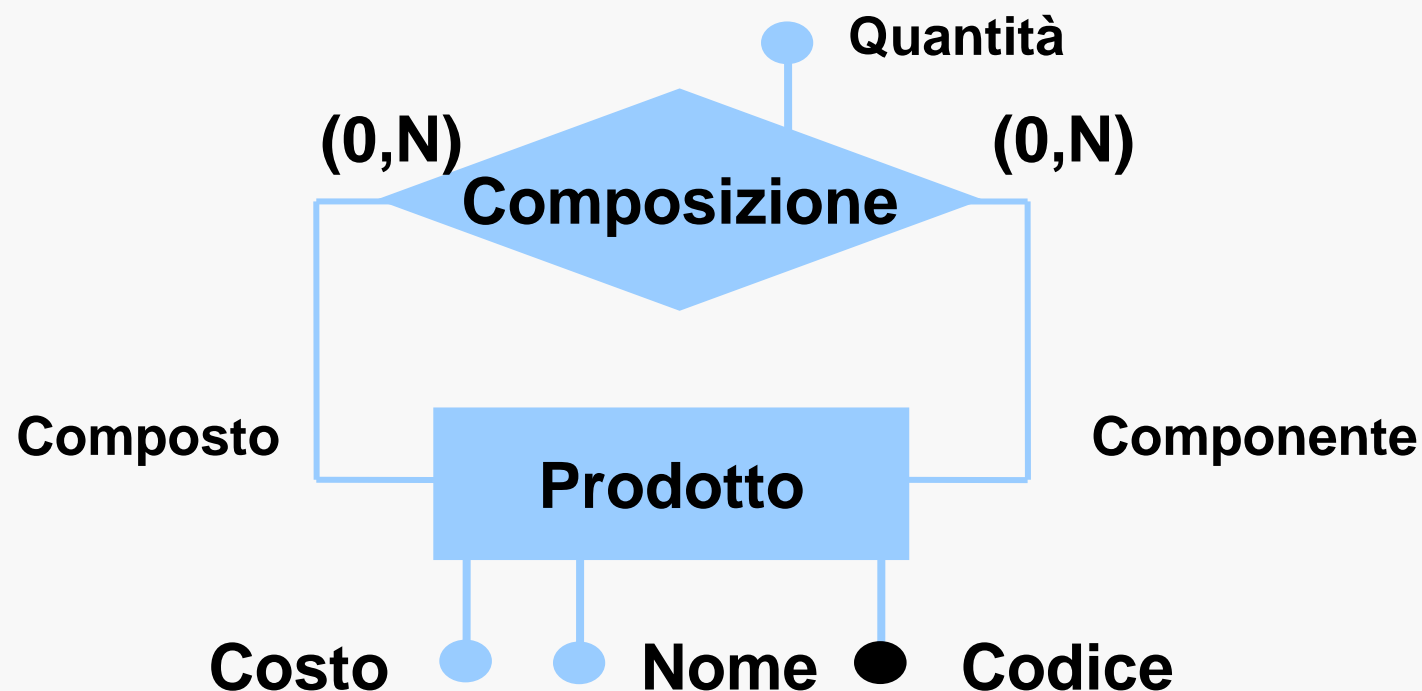


Nota

- La traduzione non riesce a tener conto delle cardinalità minime delle relationship molti a molti (se non con vincoli di CHECK complessi e poco usati)



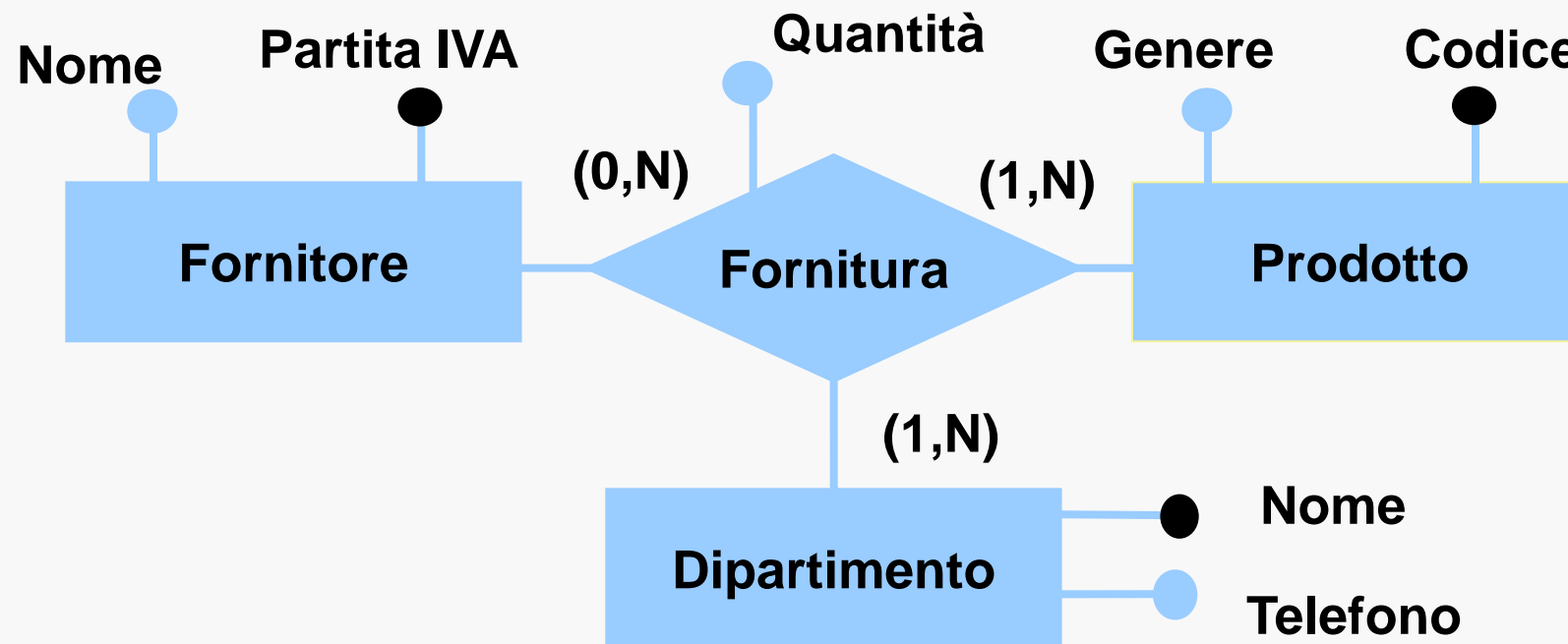
Relationship ricorsive



Prodotto(Codice, Nome, Costo)

Composizione(Composto, Componente, Quantità)

Relationship n-arie



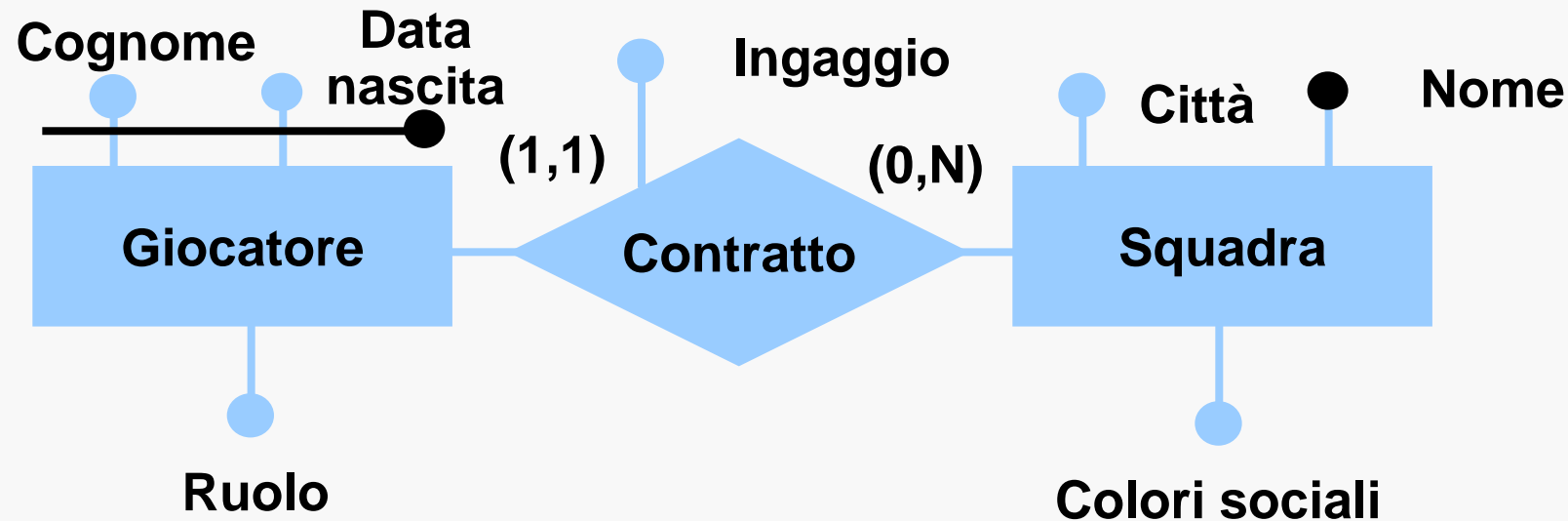
Fornitore(PartitaIVA, Nome)

Prodotto(Codice, Genere)

Dipartimento(Nome, Telefono)

Fornitura(Fornitore, Prodotto, Dipartimento, Quantità)

Relationship uno a molti



Giocatore(Cognome, DataNascita, Ruolo)

Contratto(CognGiocatore, DataNascG, Squadra, Ingaggio)

Squadra(Nome, Città, ColoriSociali)

- **corretto?**

Soluzione più compatta

Giocatore(Cognome, DataNascita, Ruolo)
Contratto(CognGiocatore, DataNascG, Squadra, Ingaggio)
Squadra(Nome, Città, ColoriSociali)

Giocatore(Cognome, DataNasc, Ruolo, Squadra, Ingaggio)
Squadra(Nome, Città, ColoriSociali)

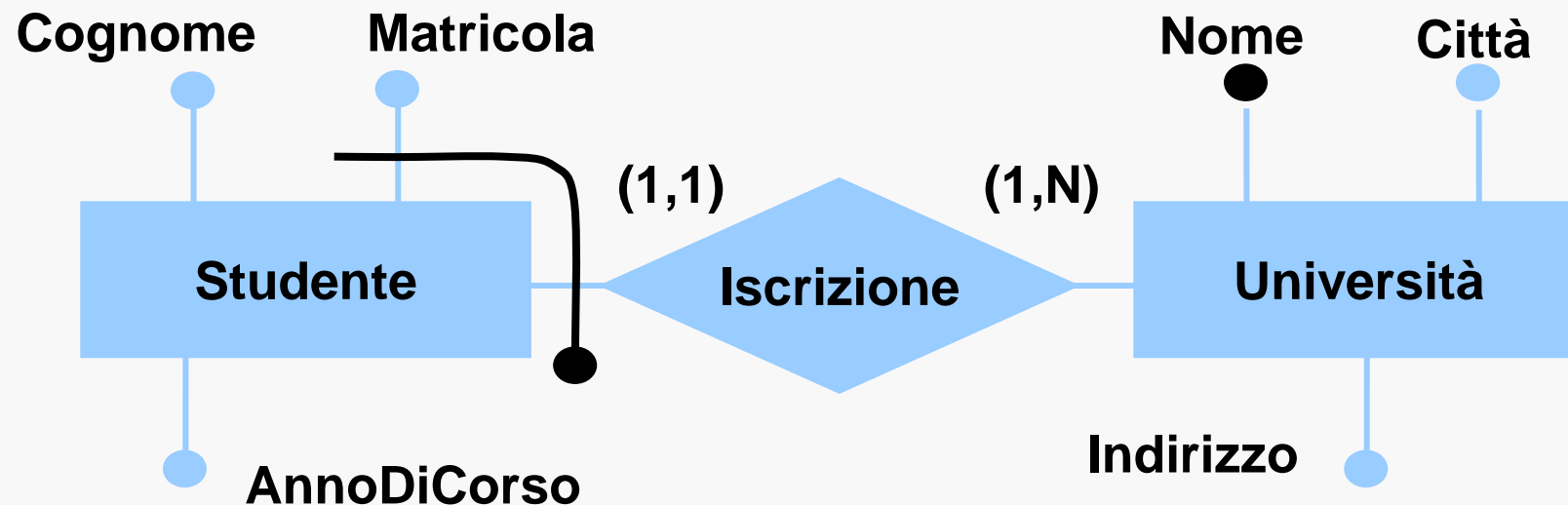
- con vincolo di integrità referenziale fra **Squadra** in **Giocatore** e la chiave di **Squadra**
- se la cardinalità minima della relationship è 0, allora **Squadra** in **Giocatore** deve ammettere valore nullo

Nota

- La traduzione riesce a rappresentare efficacemente la cardinalità minima della partecipazione che ha 1 come cardinalità massima:
 - 0 : valore nullo ammesso
 - 1 : valore nullo non ammesso



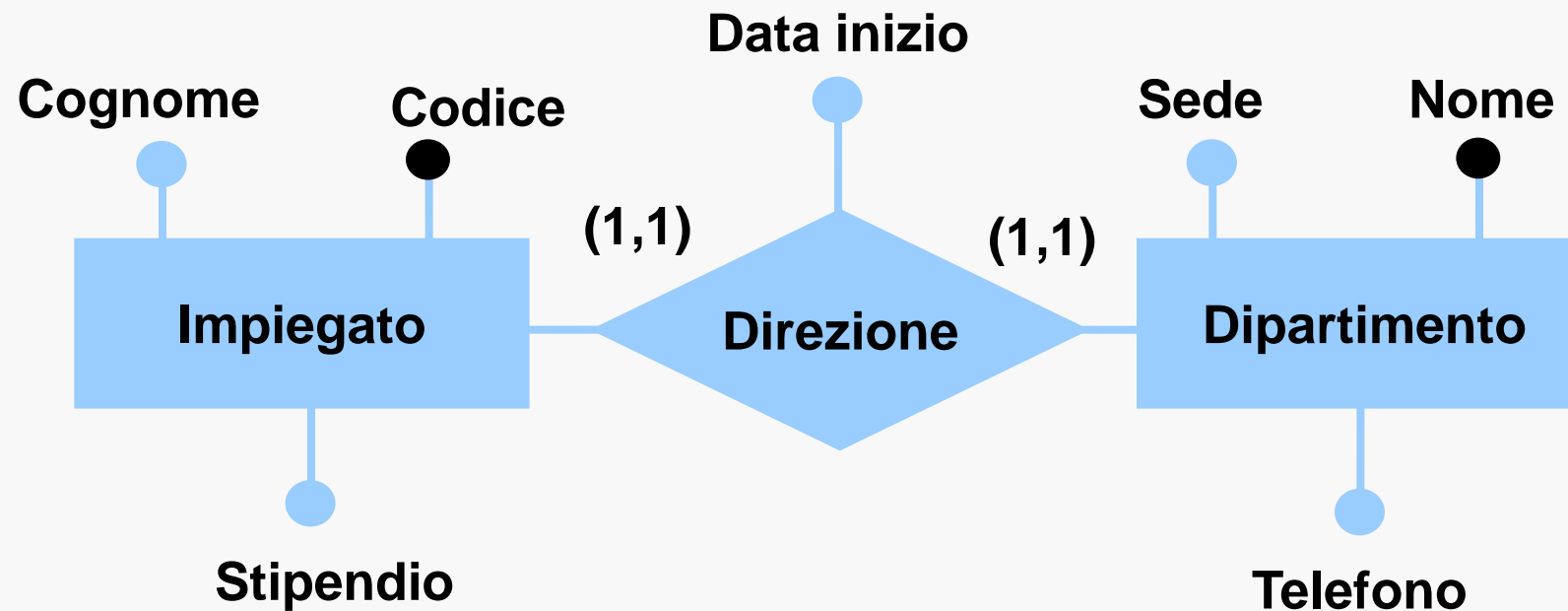
Entità con identificazione esterna



Studente(Matricola, Università, Cognome, AnnoDiCorso)
Università(Nome, Città, Indirizzo)

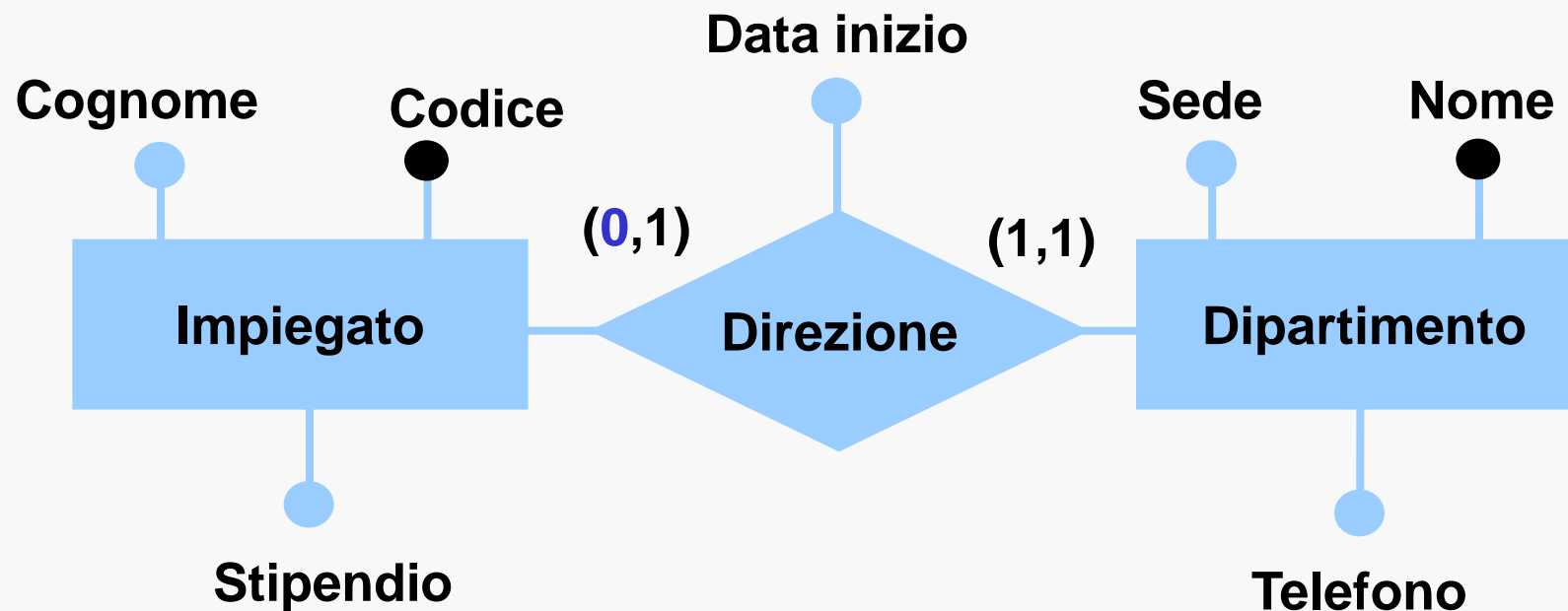
- con vincolo ...

Relationship uno a uno



- varie possibilità:
 - fondere da una parte o dall'altra
 - fondere tutto?

Una possibilità privilegiata

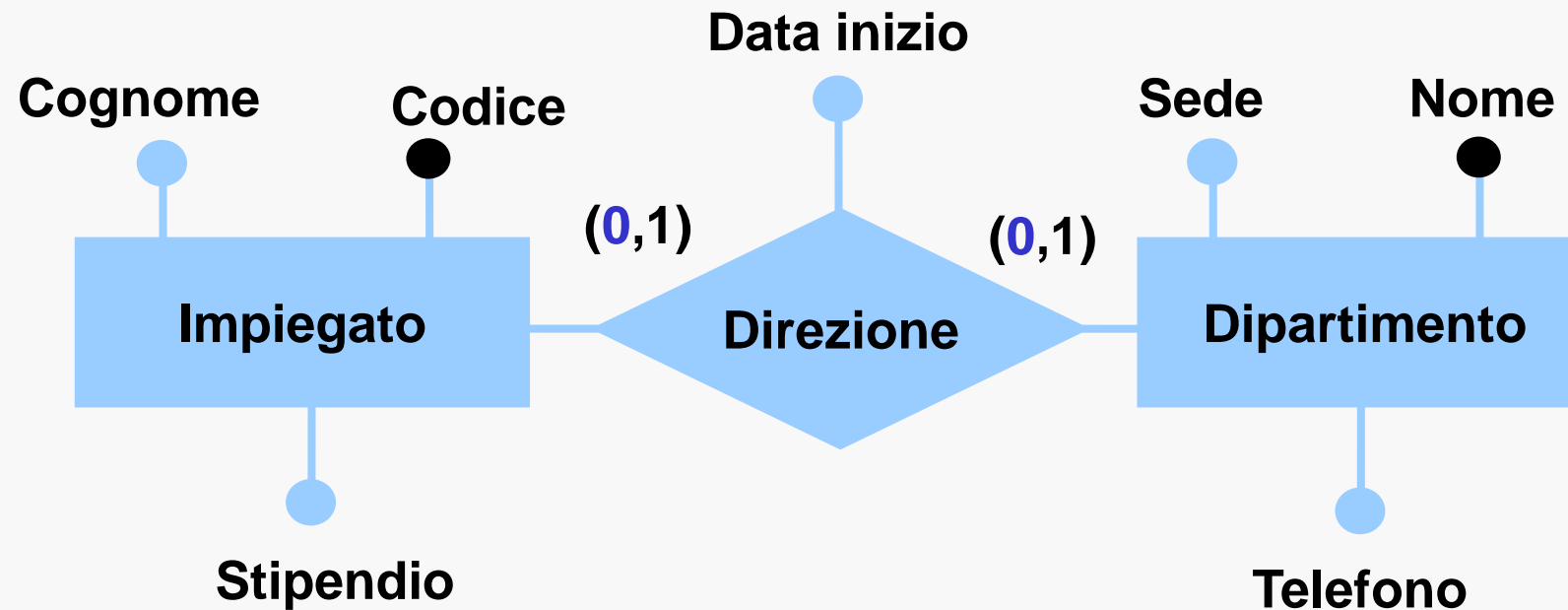


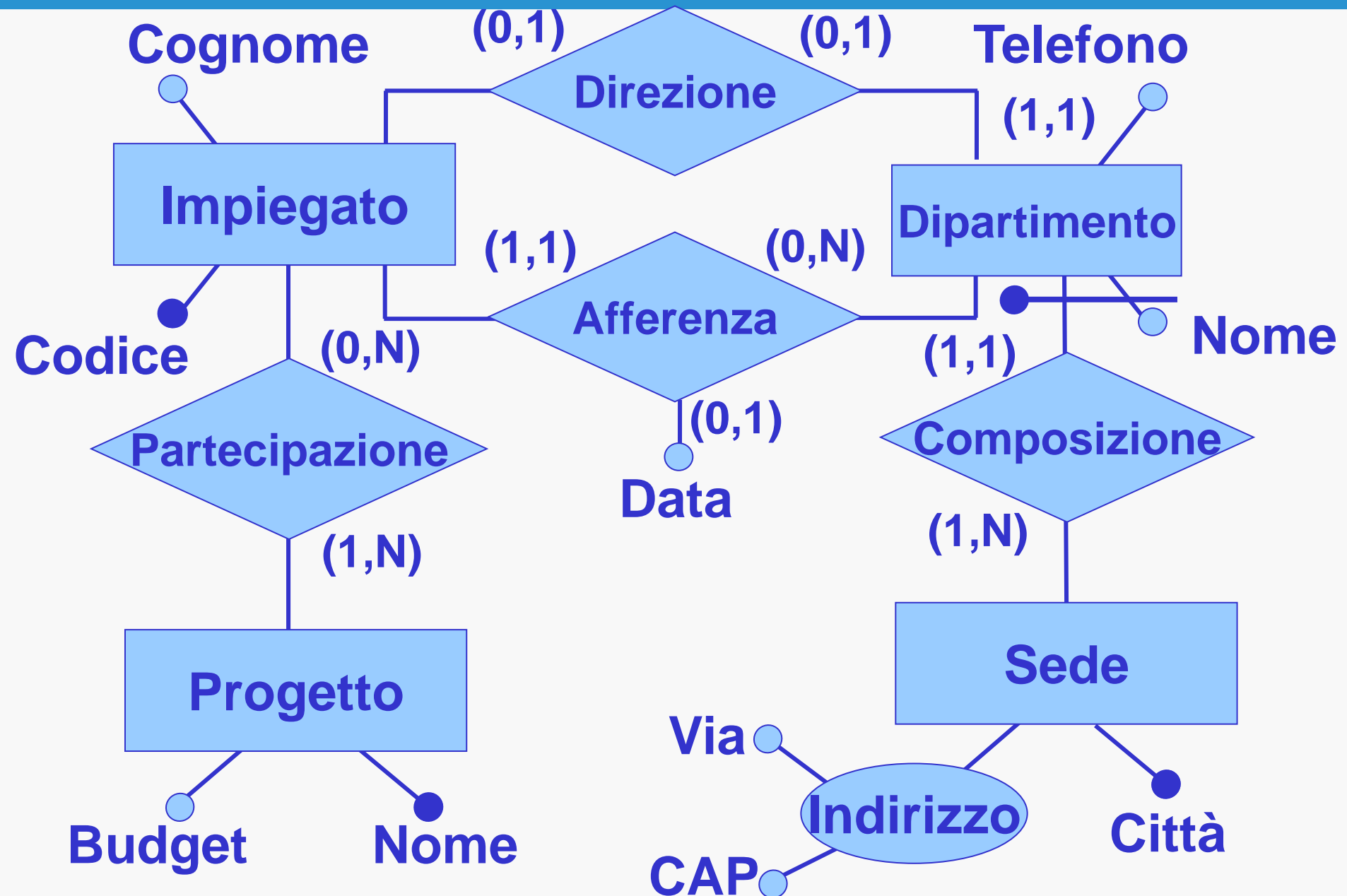
Impiegato (Codice, Cognome, Stipendio)

Dipartimento (Nome, Sede, Telefono, Direttore, InizioD)

- con vincolo di integrità referenziale, senza valori nulli

Un altro caso





Schema finale

**Impiegato(Codice, Cognome,
Dipartimento, Sede, Data*)**

Dipartimento(Nome, Città, Telefono, Direttore*)

Sede(Città, Via, CAP)

Progetto(Nome, Budget)

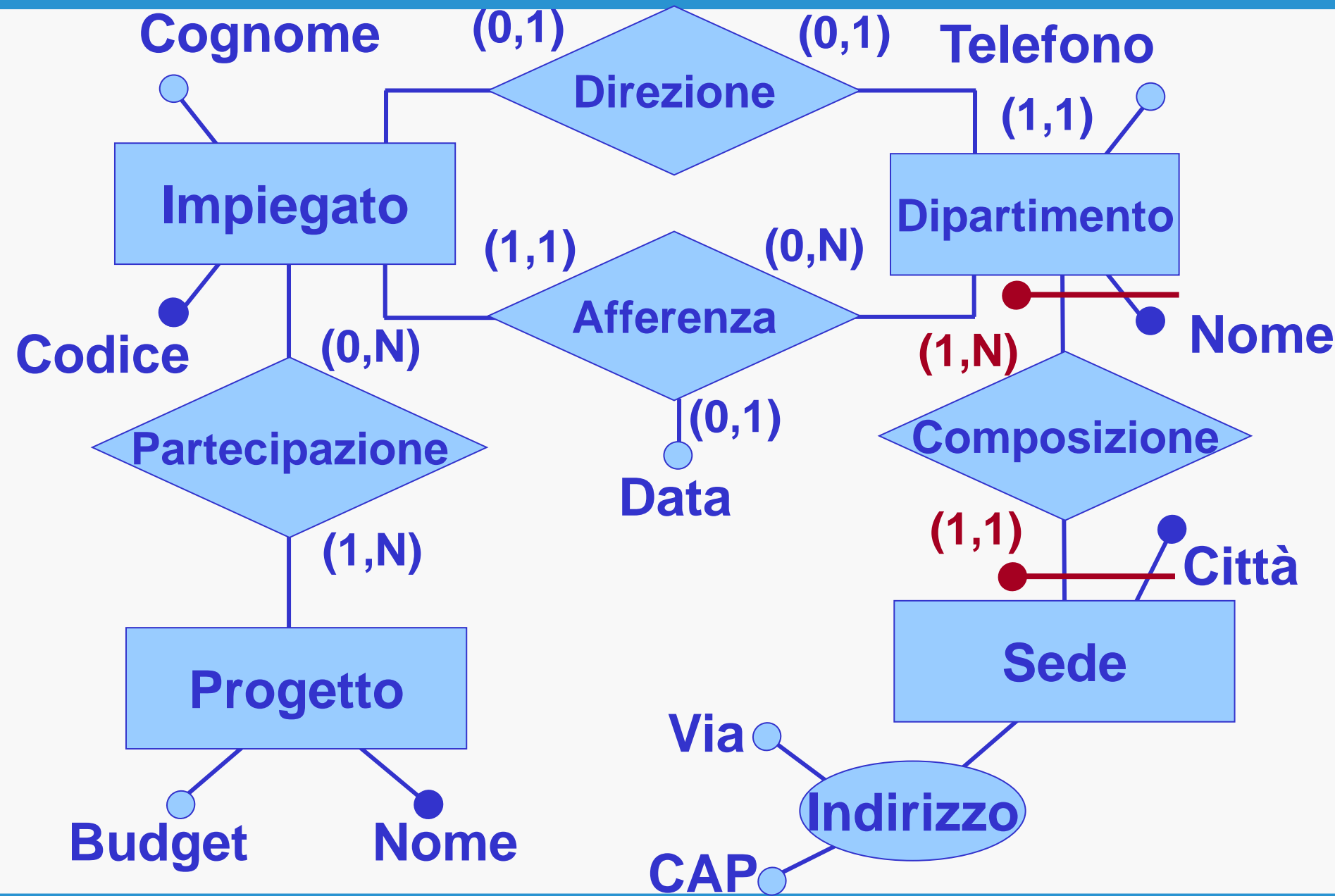
Partecipazione(Impiegato, Progetto)



Attenzione

- Differenze apparentemente piccole in cardinalità e identificatori possono cambiare di molto il significato ...





Strumenti di supporto

- Esistono sul mercato prodotti CASE che forniscono un supporto a tutte le fasi della progettazione di basi di dati



