

Basi di Dati

La teoria della normalizzazione

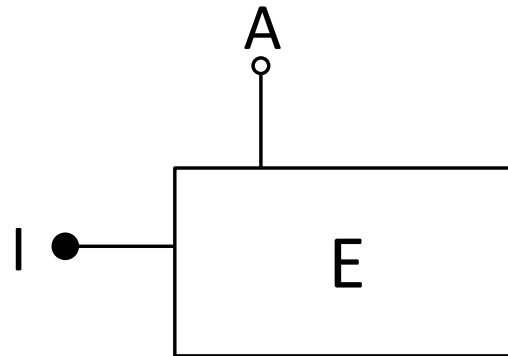
-- quarta parte --

Entity-Relationship e normalizzazione

Normalizzazione e verifica qualità

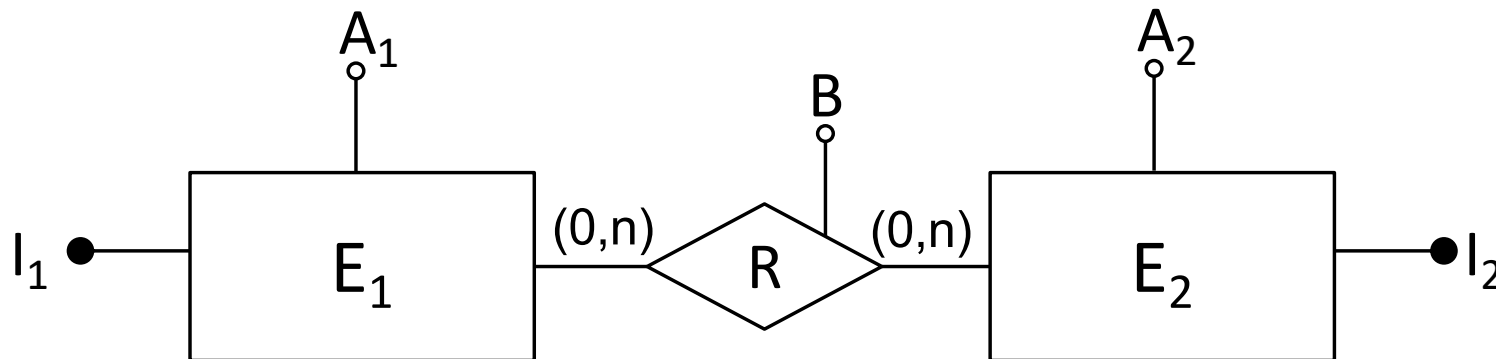
- Qual è il rapporto tra la normalizzazione e la progettazione con il modello Entity-Relationship?
- Possiamo usare la normalizzazione per verificare la qualità dei progetti concettuali: consideriamo se l'ER rispetta le dipendenze funzionali.
- Ora esaminiamo i costrutti ER e discutiamo quali dipendenze funzionali sono sottintese negli schemi concettuali.

Entità



- Entità E con identificatore I e attributo A.
- L'entità E esprime la dipendenza funzionale $I \rightarrow A$, infatti due occorrenze di E con lo stesso valore per I devono avere lo stesso valore per A.
- La traduzione in relazionale è $R_E(I, A)$ in cui la d.f. $I \rightarrow A$ è sottintesa e in BCNF.

Associazione molti a molti

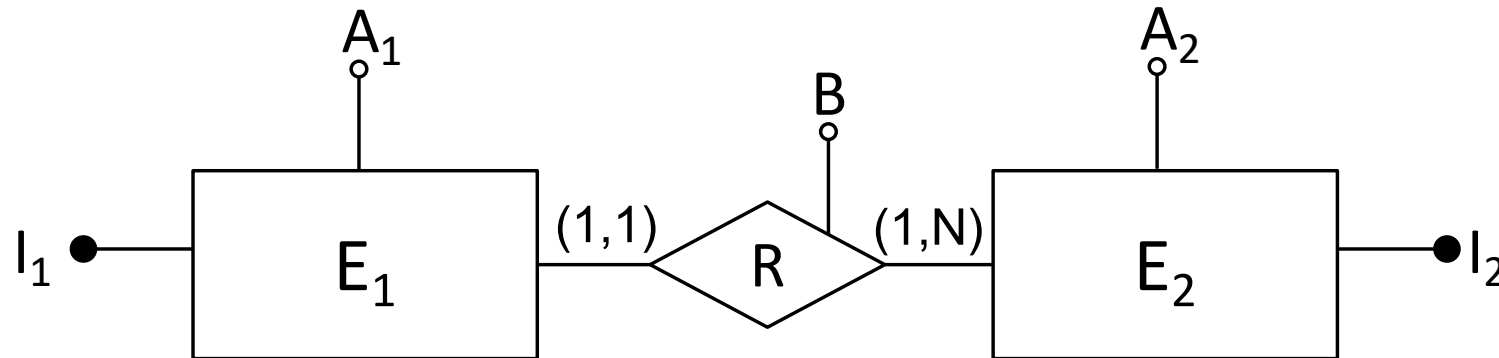


- Da E_1 abbiamo $I_1 \rightarrow A_1$ e da E_2 abbiamo $I_2 \rightarrow A_2$.
- Da R abbiamo che ogni occorrenza dell'associazione (e quindi dell'attributo B) è individuata da una coppia di occorrenze di E_1 e di E_2 .

Quindi $I_1 I_2 \rightarrow B$.

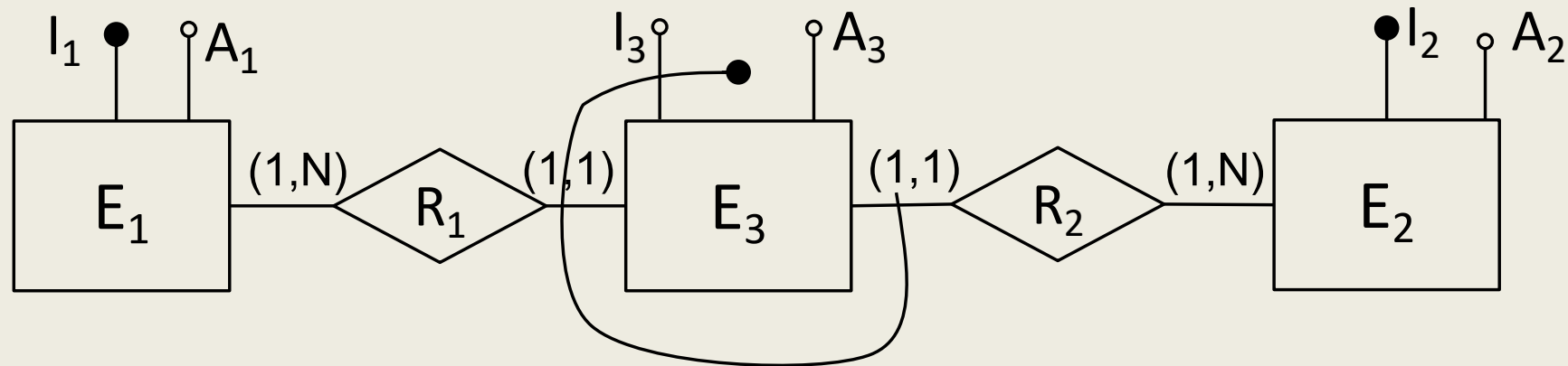
- La traduzione in relazionale è $R_{E_1}(I_1, A_1)$, $R_{E_2}(I_2, A_2)$, $R_A(I_1, I_2, B)$.
- Le d.f. $I_1 \rightarrow A_1$, $I_2 \rightarrow A_2$ e $I_1 I_2 \rightarrow B$ sono sottintese e BCNF.

Associazione uno a molti



- Da E_1 abbiamo $I_1 \rightarrow A_1$ e da E_2 abbiamo $I_2 \rightarrow A_2$.
- Da R abbiamo che ogni occorrenza dell'associazione è determinata da una coppia di occorrenze di E_1 e E_2 . Quindi $I_1 I_2 \rightarrow B$.
- A causa della cardinalità $(1,1)$, ogni occorrenza di E_1 è associata, tramite R , a un'unica occorrenza di E_2 . Quindi abbiamo $I_1 \rightarrow B I_2$.
- Calcolando l'insieme di copertura minimale e raggruppando per antecedente, rimangono le d.f. $I_1 \rightarrow A_1 B I_2$ e $I_2 \rightarrow A_2$.
- La traduzione in relazionale è $R_{E1}(I_1, A_1, B, I_2)$, $R_{E2}(I_2, A_2)$, quindi le d.f. sono BCNF.

Identificazione esterna

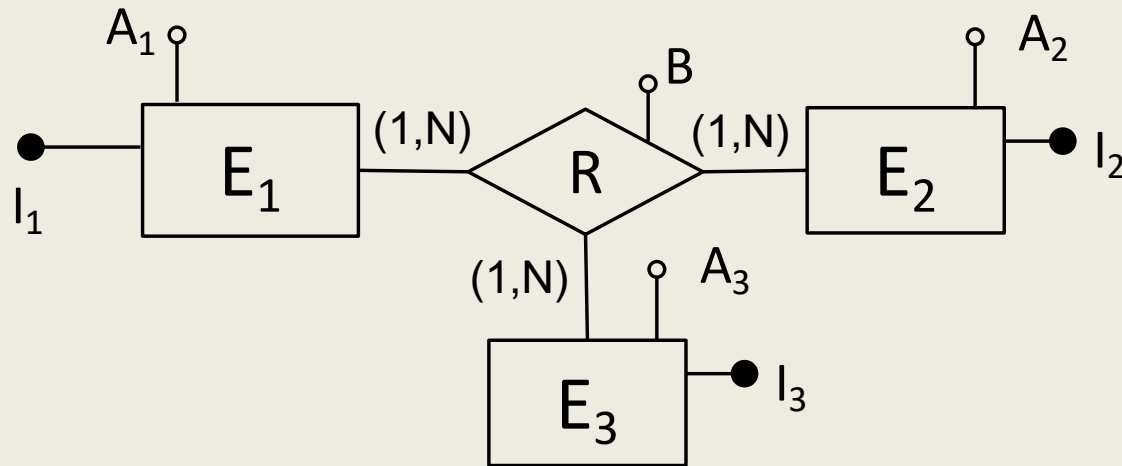


Da E_1 e E_2 abbiamo $I_1 \rightarrow A_1$ e $I_2 \rightarrow A_2$.

Da E_3 abbiamo che, considerando l'identificazione esterna, preso un determinato valore di I_3 abbinato a una coppia di occorrenze di E_1 e di E_2 , troviamo un determinato valore di A_3 . Quindi $I_3 I_1 I_2 \rightarrow A_3$.

La traduzione in relazionale è $R_{E1}(I_1, A_1)$, $R_{E2}(I_2, A_2)$, $R_{E3}(I_3, I_1, I_2, A_3)$, che è BCNF.

Associazione molti a molti a molti

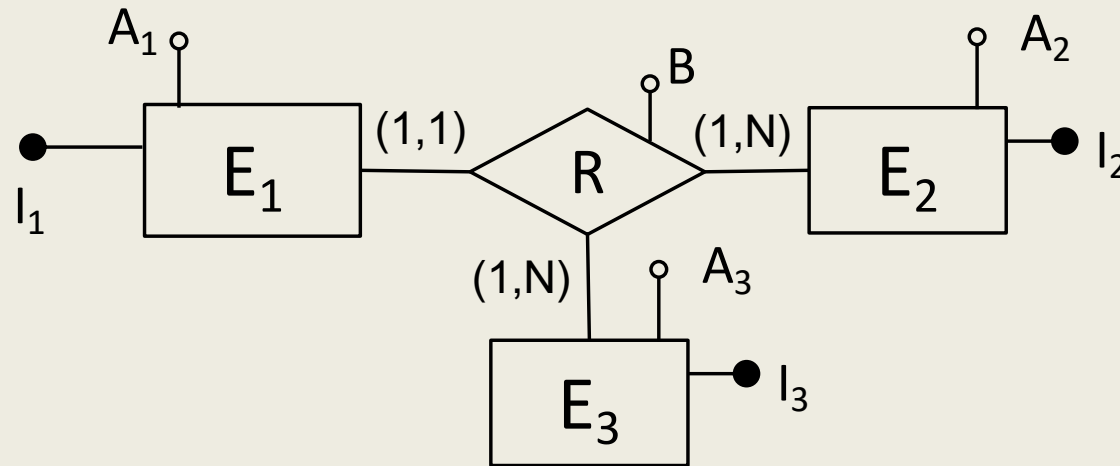


Da E_1 , E_2 e E_3 abbiamo $I_1 \rightarrow A_1$, $I_2 \rightarrow A_2$ e $I_3 \rightarrow A_3$.

Da R abbiamo che ogni occorrenza dell'associazione è determinata da una tripletta di occorrenze di E_1 , di E_2 e di E_3 . Quindi abbiamo $I_1 I_2 I_3 \rightarrow B$.

La traduzione in relazionale è $R_{E1}(I_1, A_1)$, $R_{E2}(I_2, A_2)$, $R_{E3}(I_3, A_3)$, $R_A(I_1, I_2, I_3, B)$, che è BCNF.

Associazione uno a molti a molti



Da E_1 , E_2 e E_3 abbiamo $I_1 \rightarrow A_1$, $I_2 \rightarrow A_2$ e $I_3 \rightarrow A_3$.

Da R abbiamo che ogni occorrenza dell'associazione è relativa a una tripletta di occorrenze di E_1 , di E_2 e di E_3 . Quindi abbiamo $I_1 I_2 I_3 \rightarrow B$.

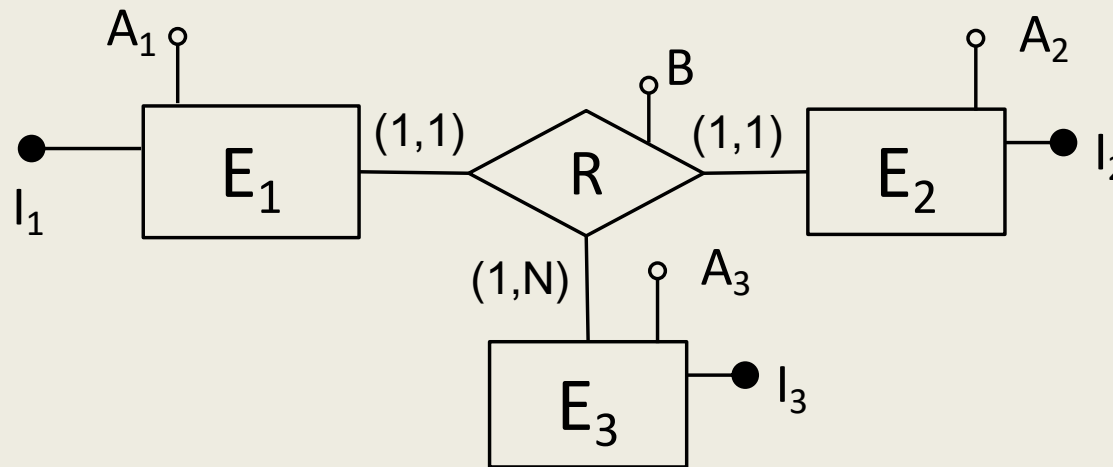
Inoltre per la cardinalità (1,1) ogni occorrenza di E_1 è associata, tramite l'associazione, a un'unica coppia di occorrenze di E_2 e E_3 , quindi abbiamo $I_1 \rightarrow B I_2 I_3$. Dalle cardinalità (1,N) non ricaviamo d.f.

Calcolando l'insieme di copertura minimale, possiamo ottenere le d.f.

$I_1 \rightarrow A_1 B I_2 I_3$, $I_2 \rightarrow A_2$ e $I_3 \rightarrow A_3$.

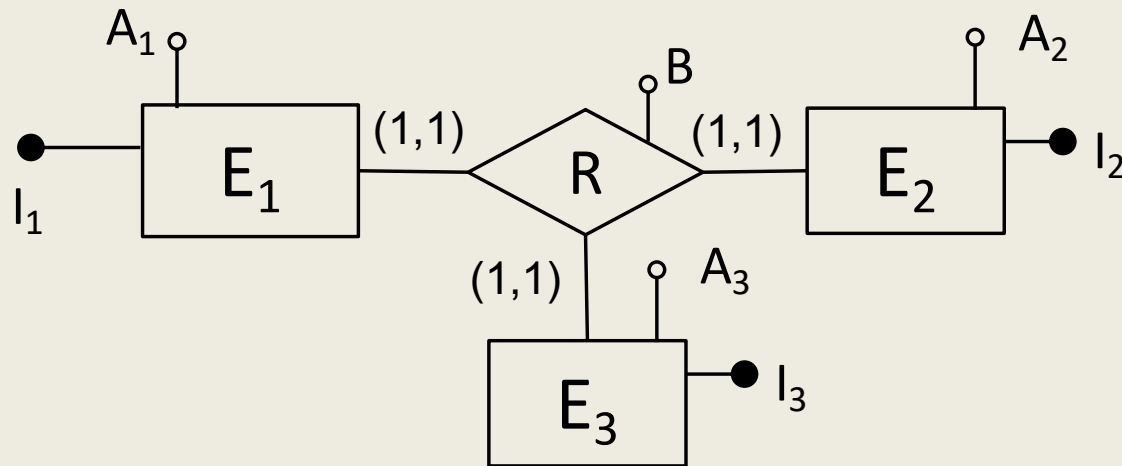
La traduzione in relazionale è $R_{E1}(I_1, A_1, B, I_2, I_3)$, $R_{E2}(I_2, A_2)$, $R_{E3}(I_3, A_3)$, che è BCNF.

Associazione uno a uno a molti



- Da E_1 , E_2 e E_3 abbiamo $I_1 \rightarrow A_1$, $I_2 \rightarrow A_2$ e $I_3 \rightarrow A_3$.
- Da R abbiamo che ogni occorrenza dell'associazione è relativa a una tripletta di occorrenze di E_1 , di E_2 e di E_3 . Quindi abbiamo $I_1 I_2 I_3 \rightarrow B$.
- Inoltre, per la cardinalità (1,1) ogni occorrenza di E_1 è associata a un'unica coppia di occorrenze di E_2 e E_3 , quindi abbiamo $I_1 \rightarrow B I_2 I_3$. Analogamente abbiamo $I_2 \rightarrow B I_1 I_3$, mentre dalla cardinalità (1,N) non ricaviamo d.f.
- Calcolando uno dei possibili insiemi di copertura minimale, possiamo ottenere le d.f. $I_1 \rightarrow A_1 B I_2 I_3$, $I_2 \rightarrow A_2$ e $I_3 \rightarrow A_3$.
- La traduzione in relazionale è $R_{E_1}(I_1, A_1, B, I_2, I_3)$, $R_{E_2}(I_2, A_2)$, $R_{E_3}(I_3, A_3)$, che è BCNF.
- Alternativamente, possiamo ottenere le d.f. $I_2 \rightarrow A_2 B I_1 I_3$, $I_1 \rightarrow A_1$ e $I_3 \rightarrow A_3$ e la traduzione in relazionale è $R_{E_2}(I_2, A_2, B, I_1, I_3)$, $R_{E_1}(I_1, A_1)$, $R_{E_3}(I_3, A_3)$, che è BCNF.

Associazione uno a uno a uno



Da E_1 , E_2 e E_3 abbiamo $I_1 \rightarrow A_1$, $I_2 \rightarrow A_2$ e $I_3 \rightarrow A_3$.

Da R abbiamo che ogni occorrenza dell'associazione è determinata da una tripletta di occorrenze di E_1 , E_2 e E_3 . Quindi abbiamo $I_1 I_2 I_3 \rightarrow B$.

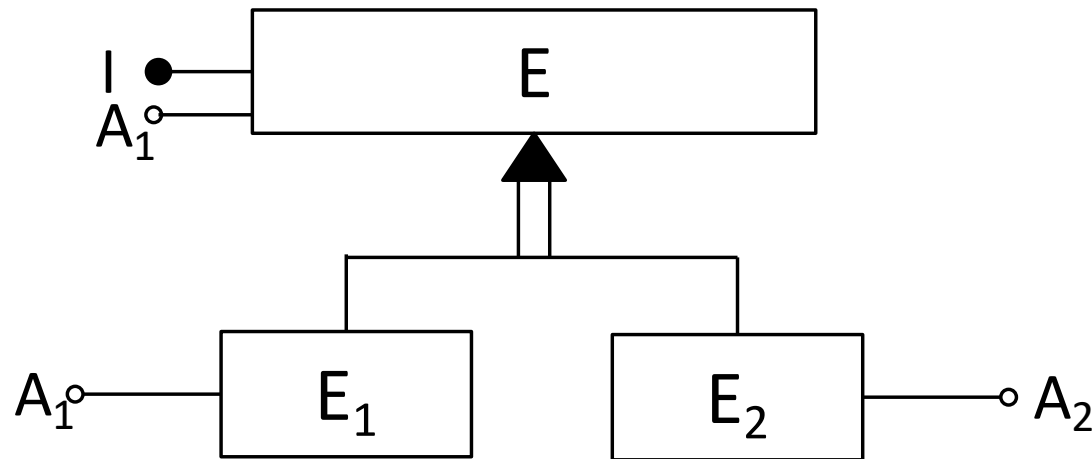
Inoltre, per la cardinalità (1,1) ogni occorrenza di E_1 è associata, tramite R , a un'unica coppia di occorrenze di E_2 e E_3 , quindi abbiamo $I_1 \rightarrow B I_2 I_3$. Analogamente abbiamo $I_2 \rightarrow B I_1 I_3$ e $I_3 \rightarrow B I_1 I_2$.

Calcolando uno dei possibili insiemi di copertura minimale e raggruppando per antecedente, possiamo ottenere la d.f. $I_1 \rightarrow A_1 B I_2 I_3$, $I_2 \rightarrow A_2$ e $I_3 \rightarrow A_3$.

La traduzione in relazionale è $R_{E_1}(I_1, A_1, B, I_2, I_3)$, $R_{E_2}(I_2, A_2)$, $R_{E_3}(I_3, A_3)$, che è BCNF.

Abbiamo altri due insiemi di copertura minimale con le relative traduzioni (BCNF).

Generalizzazione



Non consideriamo esplicitamente le gerarchie perché possiamo assumere che vengano ristrutturare in uno degli altri costrutti dell'ER.

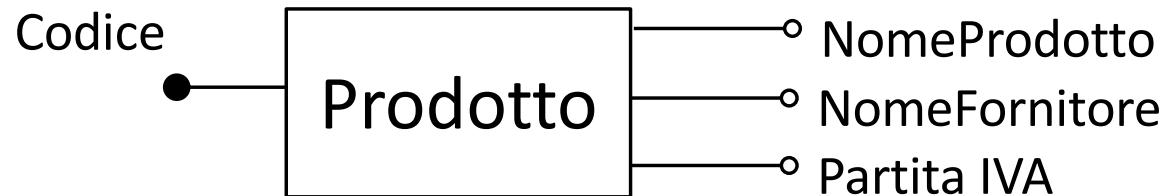
ER e BCNF

- Tutte le d.f. viste espresse nell'ER sono rappresentate nello schema relazionale sotto forma di chiave, cioè sono in BCNF.
- Però BCNF ha una limitazione: non garantisce la conservazione delle dipendenze.
- Quindi anche in ER non tutte le d. f. possibili possono essere espresse localmente.
- Queste vengono espresse come regole aziendali che nella traduzione in relazionale coinvolgono relazioni diverse.
- Possono causare degradi in performance e possiamo valutare se rinunciare alla BCNF accontentandoci della 3NF.

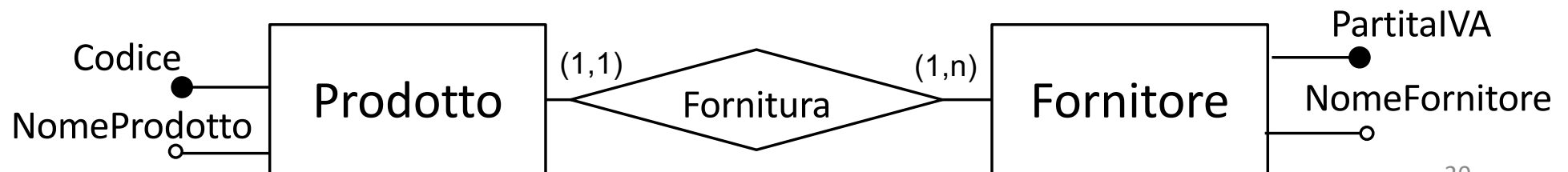
Verifica della qualità

- Confrontiamo le dipendenze funzionali espresse dall'ER con le dipendenze funzionali individuate dall'analisi del dominio che vogliamo rappresentare.

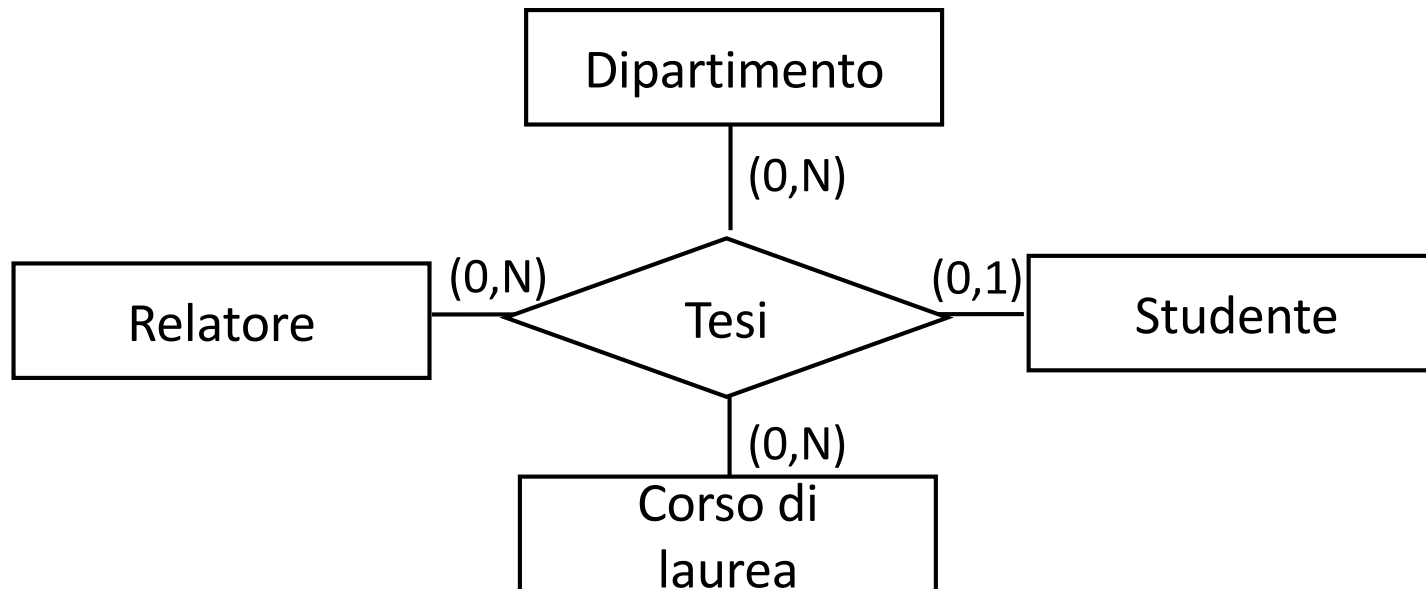
Verifica di normalizzazione su entità



- L'ER rappresenta
Codice \rightarrow NomeProdotto, PartitaIVA, NomeFornitore,
- Dall'analisi della realtà invece sappiamo che sono valide le d.f.
Codice \rightarrow NomeProdotto e PartitaIVA \rightarrow NomeFornitore.
- Ma l'ER permette ad esempio di avere due fornitori con la stessa partita IVA ma nomi diversi.
- Questo ci fa capire che usare una sola entità è errato e occorre separare il prodotto dal fornitore.



Verifica di normalizzazione su associazione



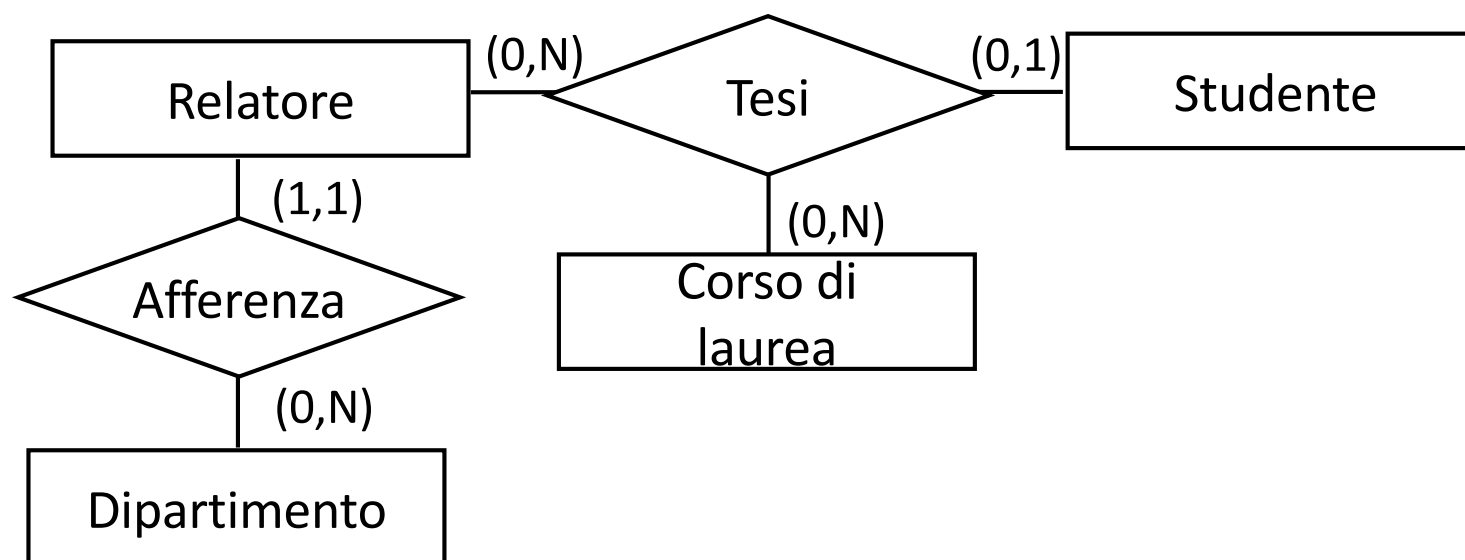
Supponiamo di avere l'ER qui sopra e di conoscere, tramite l'analisi della realtà, le d.f. (per semplicità non abbiamo incluso gli attributi)

Studente → *Corso di laurea*; *Studente* → *Relatore*; *Relatore* → *Dipartimento*.

L'ER è coerente con le prime due d.f., ma non con *Relatore* → *Dipartimento*: infatti l'ER non impedisce di avere un relatore che segue tesi per dipartimenti diversi.

Questo controllo ci fa capire che in questo caso l'associazione quaternaria è errata. Separiamo Dipartimento in modo da rappresentare correttamente *Relatore* → *Dipartimento*...

Verifica di normalizzazione su associazione



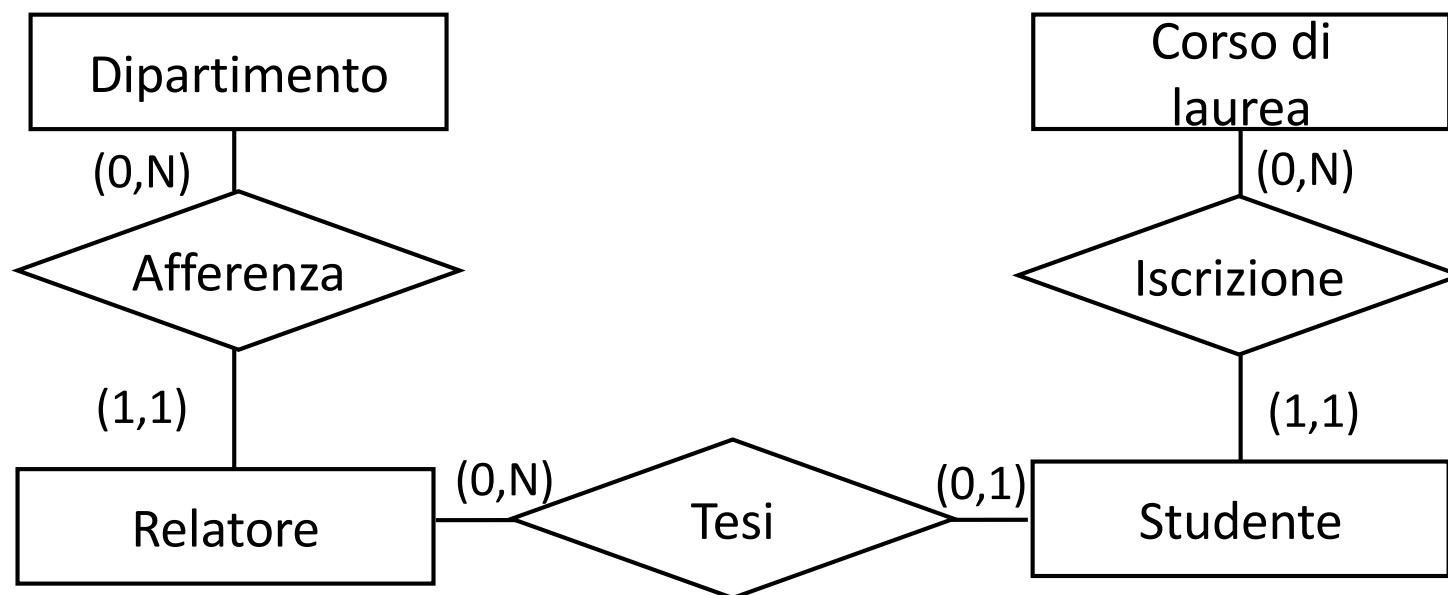
Questo schema rispetta le d.f.

Studente → *Corso di laurea*; *Studente* → *Relatore*; *Relatore* → *Dipartimento*

Però uno studente può iscriversi a un corso di laurea anche se non è ancora un tesista, mentre nell'ER l'associazione ternaria Tesi obbliga a rappresentare triplete (Studente, Corso di laurea, Relatore).

Quindi è opportuno decomporre ulteriormente l'associazione Tesi...

Verifica di normalizzazione su associazione



In questo modo conserviamo tutte le d.f. e inoltre possiamo rappresentare il vincolo (non rappresentato dalle d.f.) che uno studente può iscriversi a un corso di laurea anche se non sta ancora svolgendo una tesi.

Conclusioni

- Abbiamo introdotto le *dipendenze funzionali* come strumento per analizzare gli schemi delle basi di dati.
- Abbiamo definito due *forme normali*, la *BCNF* e la *3NF*, e discusso del modo in cui permettono di evitare le anomalie.
- Abbiamo illustrato un algoritmo per normalizzare uno schema di una base di dati.
- Abbiamo illustrato il rapporto tra forme normali e ER.

Conclusioni

- Nella pratica è sempre desiderabile avere una base di dati normalizzata?
- DB normalizzato riduce le ridondanze e evita anomalie di inserimento/modifica/cancellazione,
- ma richiede frequenti join in fase di lettura, che sono costosi
- Quindi dipende dallo scopo!
- **Sistemi OLTP** (*online transaction processing*): normale gestione operativa quotidiana
 - i dati vengono frequentemente inseriti/modificati e vengono letti pochi record per volta
 - è desiderabile avere *DB normalizzati* per evitare le ridondanze e le anomalie.

Conclusioni

- **Sistemi OLAP** (*online analytical processing*) e **data warehouse**: analisi dati per supporto alle decisioni
 - i dati vengono raramente scritti/modificati (di solito caricati «a blocchi» partendo dagli OLTP) e vengono letti molti record per volta
 - le anomalie di aggiornamento, inserimento e cancellazione si presenterebbero raramente perché i dati vengono modificati di rado
 - è desiderabile avere *DB denormalizzati*: le ridondanze permettono di aumentare le performance in lettura/interrogazione perché si eseguono meno join