

La normalizzazione



Forme normali

- # Una forma normale è una proprietà di una base di dati relazionale che ne garantisce la “qualità”, cioè l'assenza di determinati difetti
 - # Quando una relazione non è normalizzata:
 - presenta ridondanze,
 - si presta a comportamenti poco desiderabili durante gli aggiornamenti
 - # Le forme normali sono di solito definite sul modello relazionale, ma hanno senso in altri contesti, ad esempio il modello E-R
-

Normalizzazione

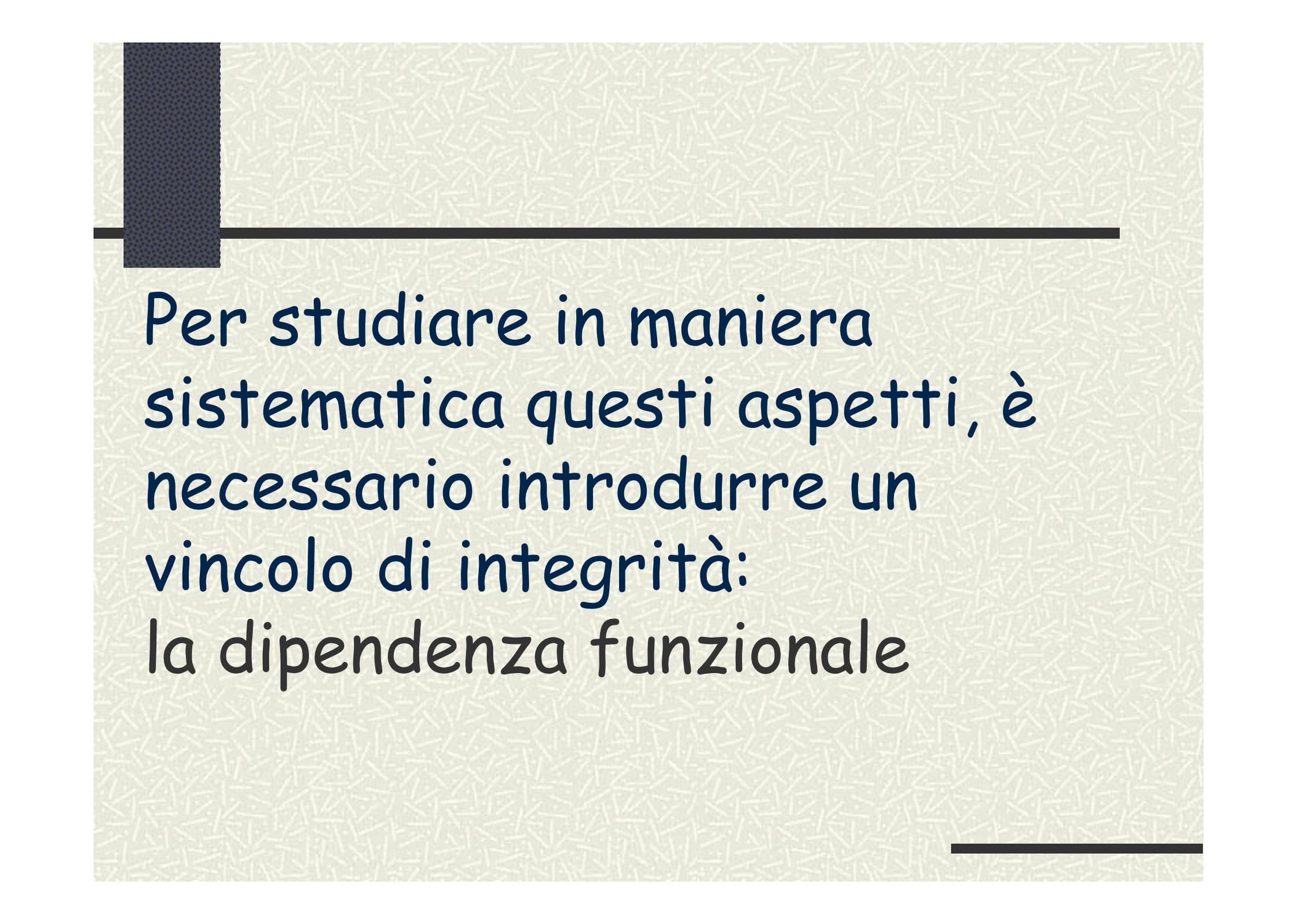
- # Procedura che permette di trasformare schemi non normalizzati in schemi che soddisfano una forma normale
 - # La normalizzazione va utilizzata come tecnica di verifica dei risultati della progettazione di una base di dati
 - # Non costituisce una metodologia di progettazione
-

Una relazione con anomalie

<u>Impiegato</u>	Stipendio	<u>Progetto</u>	Bilancio	Funzione
Rossi	20	Marte	2	tecnico
Verdi	35	Giove	15	progettista
Verdi	35	Venere	15	progettista
Neri	55	Venere	15	direttore
Neri	55	Giove	15	consulente
Neri	55	Marte	2	consulente
Mori	48	Marte	2	direttore
Mori	48	Venere	15	progettista
Bianchi	48	Venere	15	progettista
Bianchi	48	Giove	15	direttore

Anomalie

- # Lo stipendio di ciascun impiegato è ripetuto in tutte le ennuple relative
 - ridondanza
 - # Se lo stipendio di un impiegato varia, è necessario andarne a modificare il valore in diverse ennuple
 - anomalia di aggiornamento
 - # Se un impiegato interrompe la partecipazione a tutti i progetti, dobbiamo cancellarlo
 - anomalia di cancellazione
 - # Un nuovo impiegato senza progetto non può essere inserito
 - anomalia di inserimento
-



Per studiare in maniera
sistematica questi aspetti, è
necessario introdurre un
vincolo di integrità:
la dipendenza funzionale

Proprietà

- # Ogni impiegato ha un solo stipendio (anche se partecipa a più progetti)
 - # Ogni progetto ha un bilancio
 - # Ogni impiegato in ciascun progetto ha una sola funzione (anche se può avere funzioni diverse in progetti diversi)
-

Dipendenza funzionale

- # relazione R su $R(X)$
 - # due sottoinsiemi non vuoti Y e Z di X
 - # esiste in R una dipendenza funzionale (FD) da Y a Z se, per ogni coppia di ennuple t_1 e t_2 di R con gli stessi valori su Y , risulta che t_1 e t_2 hanno gli stessi valori anche su Z
-

Notazione

$$X \rightarrow Y$$

Esempi:

Impiegato \rightarrow Stipendio

Progetto \rightarrow Bilancio

Impiegato Progetto \rightarrow Funzione

Altre FD

- # Impiegato Progetto \rightarrow Progetto
 - # Si tratta però di una FD “banale” (sempre soddisfatta)
 - # $Y \rightarrow A$ è non banale se A non appartiene a Y
 - # $Y \rightarrow Z$ è non banale se nessun attributo in Z appartiene a Y
-

Le anomalie derivano da alcune FD

gli impiegati hanno un unico stipendio

Impiegato \rightarrow Stipendio

i progetti hanno un unico bilancio

Progetto \rightarrow Bilancio

Non tutte le FD causano anomalie

- # In ciascun progetto, un impiegato svolge una sola funzione

Impiegato Progetto → Funzione

- # Il soddisfacimento è più "semplice"
-

Una differenza fra FD

causano anomalie

Impiegato → Stipendio

Progetto → Bilancio

non causa anomalie

Impiegato Progetto → Funzione

Perché?

<u>Impiegato</u>	Stipendio	<u>Progetto</u>	Bilancio	Funzione
Rossi	20	Marte	2	tecnico
Verdi	35	Giove	15	progettista
Verdi	35	Venere	15	progettista
Neri	55	Venere	15	direttore
Neri	55	Giove	15	consulente
Neri	55	Marte	2	consulente
Mori	48	Marte	2	direttore
Mori	48	Venere	15	progettista
Bianchi	48	Venere	15	progettista
Bianchi	48	Giove	15	direttore

Impiegato → Stipendio
Progetto → Bilancio
Impiegato Progetto → Funzione

FD e anomalie

- # La terza FD corrisponde ad una chiave e non causa anomalie
 - # Le prime due FD non corrispondono a chiavi e causano anomalie
 - # La relazione contiene alcune informazioni legate alla chiave e altre ad attributi che non formano una chiave
 - # È stata usata un'unica relazione per rappresentare informazioni eterogenee
-

Forma normale di Boyce e Codd (BCNF)

- # Una relazione R è in forma normale di Boyce e Codd se, per ogni dipendenza funzionale (non banale) $X \rightarrow Y$, X contiene una chiave K di R
- # La forma normale richiede che i concetti in una relazione siano omogenei (solo proprietà direttamente associate alla chiave)

Violazione della BCNF

- # Si sostituisce la relazione con altre che soddisfano la BCNF, decomponendo le FD che violano la forma normale

<u>Impiegato</u>	Stipendio	<u>Progetto</u>	Bilancio	Funzione
Rossi	20	Marte	2	tecnico
Verdi	35	Giove	15	progettista
Verdi	35	Venere	15	progettista
Neri	55	Venere	15	direttore
Neri	55	Giove	15	consulente
Neri	55	Marte	2	consulente
Mori	48	Marte	2	direttore
Mori	48	Venere	15	progettista
Bianchi	48	Venere	15	progettista
Bianchi	48	Giove	15	direttore

Impiegato	Stipendio
Rossi	20
Verdi	35
Neri	55
Mori	48
Bianchi	48

Impiegato	Progetto	Funzione
Rossi	Marte	tecnico
Verdi	Giove	progettista
Verdi	Venere	progettista
Neri	Venere	direttore
Neri	Giove	consulente
Neri	Marte	consulente
Mori	Marte	direttore
Mori	Venere	progettista
Bianchi	Venere	progettista
Bianchi	Giove	direttore

Progetto	Bilancio
Marte	2
Giove	15
Venere	15

Non è sempre così facile

Impiegato	Progetto	Sede
Rossi	Marte	Roma
Verdi	Giove	Milano
Verdi	Venere	Milano
Neri	Saturno	Milano
Neri	Venere	Milano

Impiegato → Sede
Progetto → Sede

Decomposizione con perdita

Impiegato	Progetto	Sede
Rossi	Marte	Roma
Verdi	Giove	Milano
Verdi	Venere	Milano
Neri	Saturno	Milano
Neri	Venere	Milano

Impiegato	Sede
Rossi	Roma
Verdi	Milano
Neri	Milano

Progetto	Sede
Marte	Roma
Giove	Milano
Saturno	Milano
Venere	Milano

Ricostruzione

Impiegato	Sede
Rossi	Roma
Verdi	Milano
Neri	Milano

Progetto	Sede
Marte	Roma
Giove	Milano
Saturno	Milano
Venere	Milano

Impiegato	Progetto	Sede
Rossi	Marte	Roma
Verdi	Giove	Milano
Verdi	Venere	Milano
Neri	Saturno	Milano
Neri	Venere	Milano
Verdi	Saturno	Milano
Neri	Giove	Milano

Diversa dalla relazione di partenza!

Decomposizione senza perdita

- # Una relazione R si decompone senza perdita su X_1 e X_2 se il join delle proiezioni di r su X_1 e X_2 è uguale a r stessa (cioè non contiene ennuple spurie)
 - # La decomposizione senza perdita è garantita se gli attributi comuni contengono una chiave per almeno una delle relazioni decomposte
-

Decomposizione senza perdita

Impiegato	Progetto	Sede
Rossi	Marte	Roma
Verdi	Giove	Milano
Verdi	Venere	Milano
Neri	Saturno	Milano
Neri	Venere	Milano

Impiegato	Sede
Rossi	Roma
Verdi	Milano
Neri	Milano

Impiegato	Progetto
Rossi	Marte
Verdi	Giove
Verdi	Venere
Neri	Saturno
Neri	Venere

Impiegato → **Sede**
Progetto → **Sede**

Un altro problema

- Supponiamo di voler inserire una nuova ennupla che specifica la partecipazione dell'impiegato Neri, che opera a Milano, al progetto Marte

Impiegato	Sede
Rossi	Roma
Verdi	Milano
Neri	Milano

Impiegato	Progetto
Rossi	Marte
Verdi	Giove
Verdi	Venere
Neri	Saturno
Neri	Venere

Impiegato → Sede
Progetto → Sede

Impiegato	Sede
Rossi	Roma
Verdi	Milano
Neri	Milano

Impiegato	Progetto
Rossi	Marte
Verdi	Giove
Verdi	Venere
Neri	Saturno
Neri	Venere
Neri	Marte

Impiegato	Progetto	Sede
Rossi	Marte	Roma
Verdi	Giove	Milano
Verdi	Venere	Milano
Neri	Saturno	Milano
Neri	Venere	Milano
Neri	Marte	Milano

Conservazione delle dipendenze

- # Una decomposizione conserva le dipendenze se ciascuna delle dipendenze funzionali dello schema originario coinvolge attributi che compaiono tutti insieme in uno degli schemi decomposti
 - # Progetto → Sede non è conservata
-

Qualità delle decomposizioni

- # Una decomposizione dovrebbe sempre soddisfare:
 - la decomposizione senza perdita, che garantisce la ricostruzione delle informazioni originarie
 - la conservazione delle dipendenze, che garantisce il mantenimento dei vincoli di integrità originari
-

Una relazione non-normalizzata

Dirigente	<u>Progetto</u>	<u>Sede</u>
Rossi	Marte	Roma
Verdi	Giove	Milano
Verdi	Marte	Milano
Neri	Saturno	Milano
Neri	Venere	Milano

Progetto Sede → Dirigente
Dirigente → Sede

La decomposizione è problematica

- # Progetto Sede → Dirigente coinvolge tutti gli attributi e quindi nessuna decomposizione può preservare tale dipendenza
 - # quindi in alcuni casi la BCNF “non è raggiungibile”
-

Una nuova forma normale

- # Una relazione R è in terza forma normale se, per ogni FD (non banale) $X \rightarrow Y$ definita su R , è verificata almeno una delle seguenti condizioni:
 - X contiene una chiave K di R
 - ogni attributo in Y è contenuto in almeno una chiave di R

BCNF e terza forma normale

- # La terza forma normale è meno restrittiva della forma normale di Boyce e Codd (e ammette relazioni con alcune anomalie)
- # Ha il vantaggio di essere sempre “raggiungibile”

Decomposizione in 3NF

- # Si crea una relazione per ogni gruppo di attributi coinvolti in una dipendenza funzionale
 - # Si verifica che alla fine una relazione contenga una chiave della relazione originari
-

Una possibile strategia

- # Se la relazione non è normalizzata si decompone in terza forma normale
 - # Alla fine si verifica se lo schema ottenuto è anche in BCNF
 - # Se una relazione ha una sola chiave allora le due forme normali coincidono
-

Uno schema non decomponibile in BCNF

Dirigente	<u>Progetto</u>	<u>Sede</u>
Rossi	Marte	Roma
Verdi	Giove	Milano
Verdi	Marte	Milano
Neri	Saturno	Milano
Neri	Venere	Milano

Dirigente → Sede
Progetto Sede → Dirigente

Una possibile riorganizzazione

Dirigente	<u>Progetto</u>	<u>Sede</u>	Reparto
Rossi	Marte	Roma	1
Verdi	Giove	Milano	1
Verdi	Marte	Milano	1
Neri	Saturno	Milano	2
Neri	Venere	Milano	2

Dirigente → Sede Reparto

Sede Reparto → Dirigente

Progetto Sede → Reparto

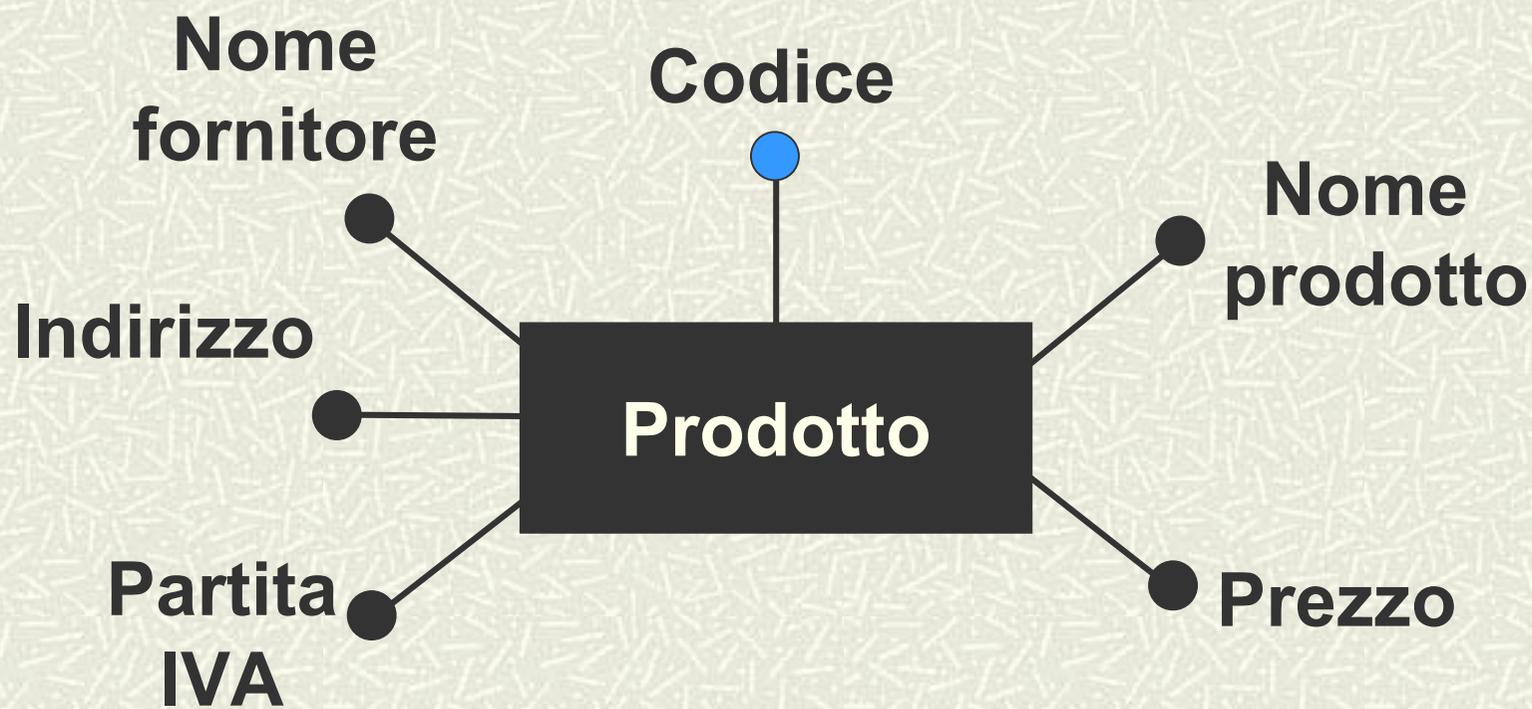
Decomposizione in BCNF

<u>Dirigente</u>	<u>Sede</u>	<u>Reparto</u>
Rossi	Roma	1
Verdi	Milano	1
Neri	Milano	2

<u>Progetto</u>	<u>Sede</u>	<u>Reparto</u>
Marte	Roma	1
Giove	Milano	1
Marte	Milano	1
Saturno	Milano	2
Venere	Milano	2

Progettazione e normalizzazione

- # La teoria della normalizzazione può essere usata nella progettazione logica per verificare lo schema relazionale finale
 - # Si può usare anche durante la progettazione concettuale per verificare la qualità dello schema concettuale
-



PartitaIVA → NomeFornitore Indirizzo

Analisi dell'entità

- L'entità viola la terza forma normale a causa della dipendenza:

PartitaIVA → NomeFornitore Indirizzo

**Nome
prodotto** **Codice**

**Partita
IVA** **Nome
fornitore**



(1,1)

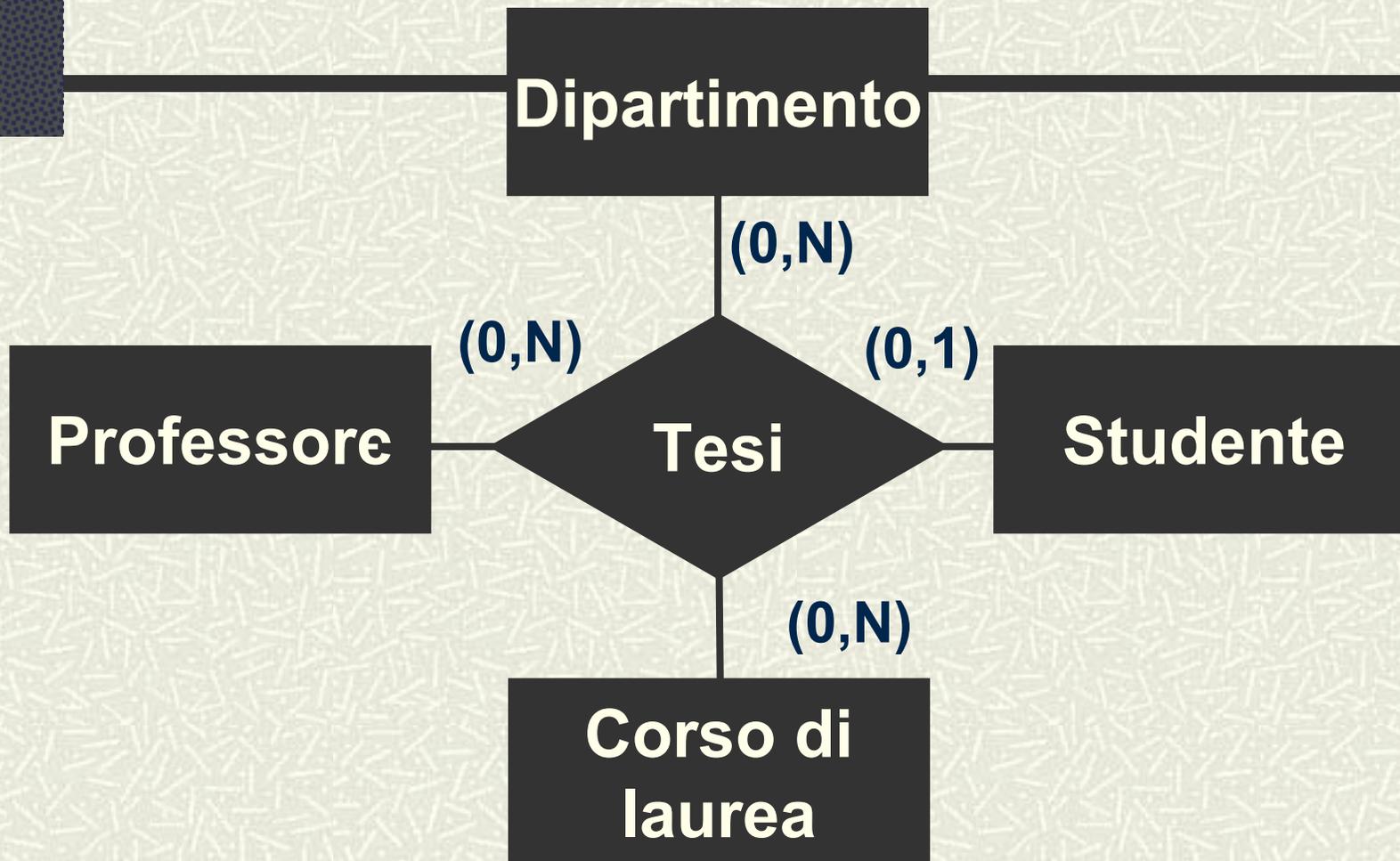
(0,N)

Fornitura



Prezzo

Indirizzo

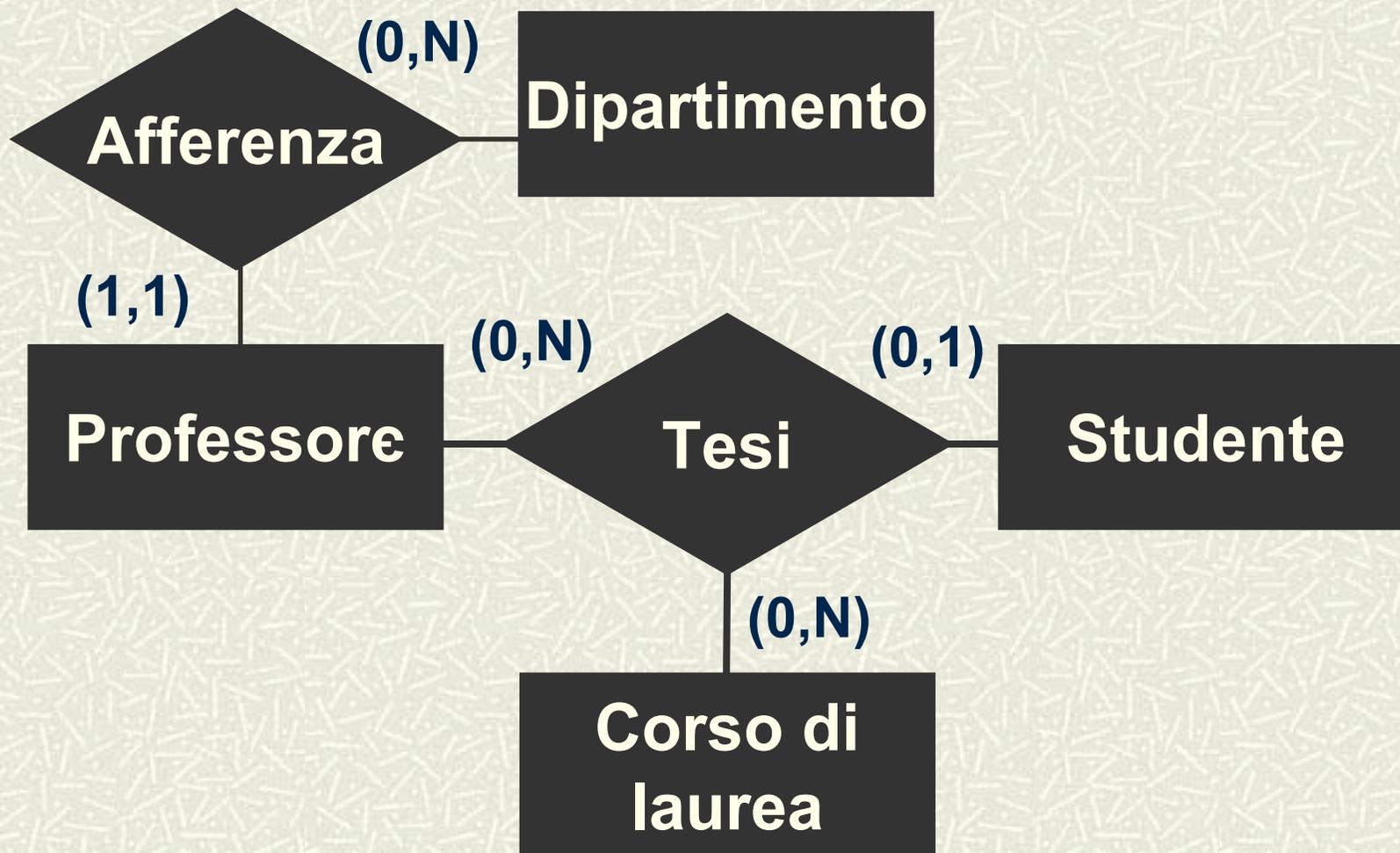


Studente → Corso di laurea
Studente → Professore
Professore → Dipartimento

Analisi della relationship

La relationship viola la terza forma normale a causa della dipendenza:

Professore \rightarrow Dipartimento



Ulteriore analisi sulla base delle dipendenze

- # La relationship Tesi è in BCNF sulla base delle dipendenze

Studente \rightarrow CorsoDiLaurea

Studente \rightarrow Professore

- # le due proprietà sono indipendenti
 - # questo suggerisce una ulteriore decomposizione
-

