

Basi di dati

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

Perché questi fenomeni indesiderabili?

- abbiamo usato un'unica relazione per rappresentare informazioni eterogenee
 - gli impiegati con i relativi stipendi
 - i progetti con i relativi bilanci
 - le partecipazioni degli impiegati ai progetti con le relative funzioni

Dipendenza funzionale

Per studiare in maniera sistematica i concetti introdotti precedentemente, è necessario far uso della **dipendenza funzionale** ovvero di un particolare vincolo di integrità per il modello relazionale che descrive legami di tipo funzionale tra gli attributi di una relazione

Dipendenza funzionale: definizione

Data una relazione r su uno schema $R(X)$ e due sottoinsiemi di attributi non vuoti Y e Z di X , diremo che esiste su r una **dipendenza funzionale** (FD) tra Y e Z se, per ogni coppia di ennuple t_1 e t_2 di r aventi gli stessi valori sugli attributi di Y , risulta che t_1 e t_2 hanno gli stessi valori anche sugli attributi Z

Dipendenza funzionale: notazione

$$Y \rightarrow Z$$

- Esempi:

Impiegato \rightarrow Stipendio

Progetto \rightarrow Bilancio

Impiegato Progetto \rightarrow Funzione

Dipendenza funzionale: esempio

<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

Un inciso: altre FD, particolari

- 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 sono legate ad alcune FD

- gli impiegati hanno un unico stipendio

Impiegato → Stipendio

- i progetti hanno un unico bilancio

Progetto → Bilancio

Non tutte le FD causano anomalie

- In ciascun progetto, un impiegato svolge una sola funzione

Impiegato Progetto → Funzione

- Il soddisfacimento è più “semplice”, perché Impiegato Progetto è chiave

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
- Le anomalie sono causate dalla presenza di concetti eterogenei:
 - proprietà degli impiegati (lo stipendio)
 - proprietà di progetti (il bilancio)
 - proprietà della chiave **Impiegato Progetto**

Forma normale di Boyce e Codd (BCNF)

Alla luce di quanto detto sulle dipendenze funzionali, l'idea fondamentale è quella di introdurre delle proprietà, chiamate forme normali, definite con riferimento alle dipendenze funzionali che sono soddisfatte quando non risultano anomalie. La più importante forma normale è quella di Boyce e Codd.

Definizione:

Una relazione r è in forma normale di Boyce e Codd se, per ogni dipendenza funzionale (non banale) $X \rightarrow Y$ definita su di essa, X contiene una chiave K di r cioè X è superchiave per r

Se una relazione non soddisfa la BCNF

- La rimpiazziamo con altre relazioni che soddisfano la BCNF
- Decomponendo sulla base delle dipendenze funzionali, al fine di separare i concetti

Decomposizione in forma normale di Boyce e Codd

- Data una relazione che non soddisfa la forma normale di Boyce e Codd è possibile, in molti casi, sostituirla con due o più relazioni normalizzate attraverso un processo detto di normalizzazione.
- Questo processo si fonda su un semplice criterio: se una relazione rappresenta più concetti dipendenti, allora va decomposta in relazioni più piccole, una per concetto

Decomposizione

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

Altra decomposizione (ma con perdita)

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

Impiegato → Sede

Impiegato	Sede
Rossi	Roma
Verdi	Milano
Neri	Milano

Progetto → Sede

Progetto	Sede
Marte	Roma
Giove	Milano
Saturno	Milano
Venere	Milano

Proviamo a ricostruire

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

Definizione:

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

Esempio: inserimento ennupla (1)

- 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

Esempio: inserimento ennupla (2)

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

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”

Terza forma normale

Definizione:

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 però di essere sempre “raggiungibile”
- se una relazione ha una sola chiave, allora essa è in BCNF se e solo se è in 3NF

Decomposizione in terza forma normale

- 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 originaria
- Dipende dalle dipendenze individuate

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

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>	Sede	Reparto
Rossi	Roma	1
Verdi	Milano	1
Neri	Milano	2

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

Teoria delle dipendenze

Tutti i concetti precedentemente discussi possono essere formalizzati attraverso un processo di normalizzazione realizzabile in modo algoritmico

Ovvero:

- Data una relazione e un insieme di dipendenze funzionali definite su di essa, generare una decomposizione della relazione che contenga solo relazioni in forma normale e soddisfi le qualità di decomposizione senza perdita e conservazione delle dipendenze.

Implicazione di dipendenze funzionali

Dalle dipendenze funzionali riconosciute valide, se ne possano identificare delle altre. Diremo che quest'ultime sono logicamente implicate dalle prime

Un insieme di dipendenze funzionali F *implica* f se ogni relazione che soddisfa tutte le dipendenze in F soddisfa anche f

Un esempio di implicazione

Impiegato	Categoria	Stipendio
Neri	3	30.000
Verdi	3	30.000
Rossi	4	50.000
Mori	4	50.000
Bianchi	5	72.000

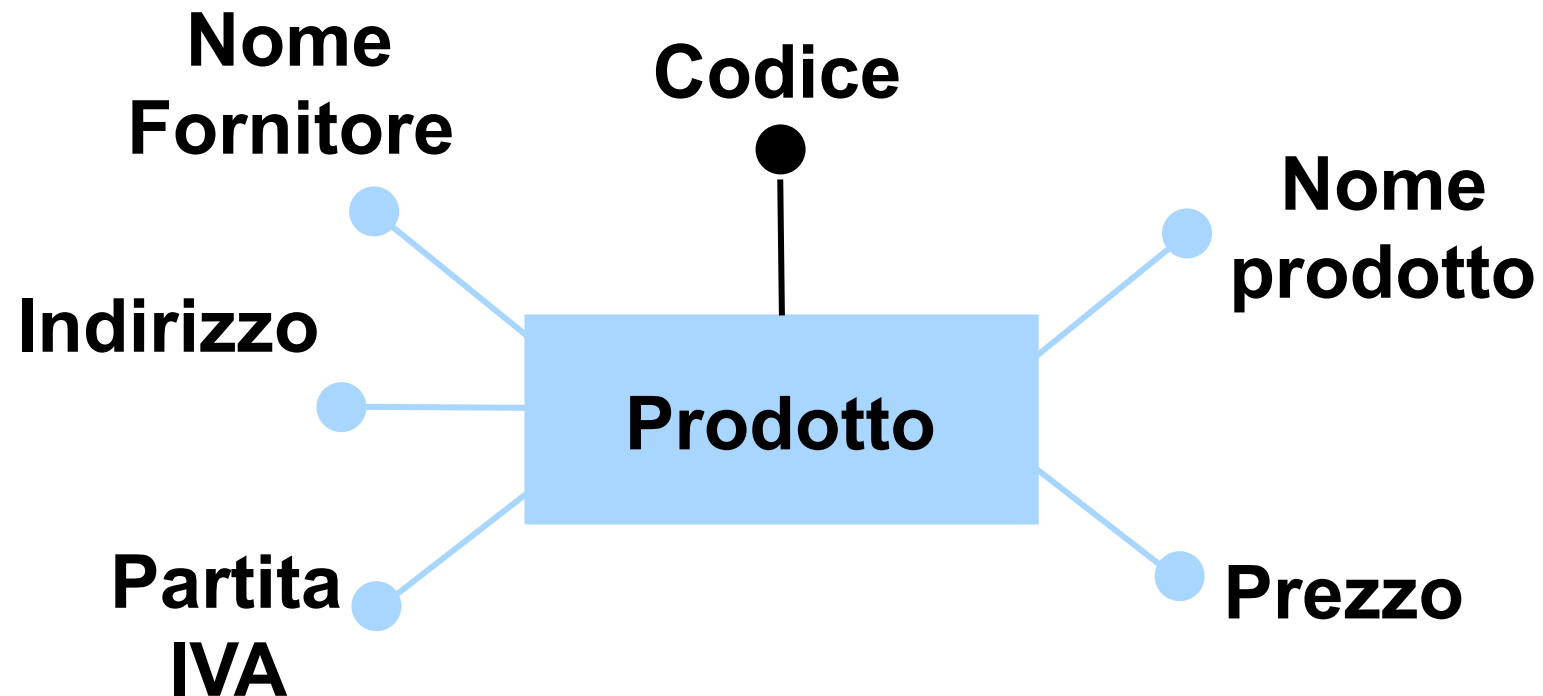
Impiegato \rightarrow Categoria e **Categoria \rightarrow Stipendio**
implicano la dipendenza **Impiegato \rightarrow Stipendio**

Ovvero ogni relazione che soddisfa le prime due,
soddisfa anche la terza

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

Verifica di normalizzazione su entità



Partita IVA → Nome Fornitore Indirizzo

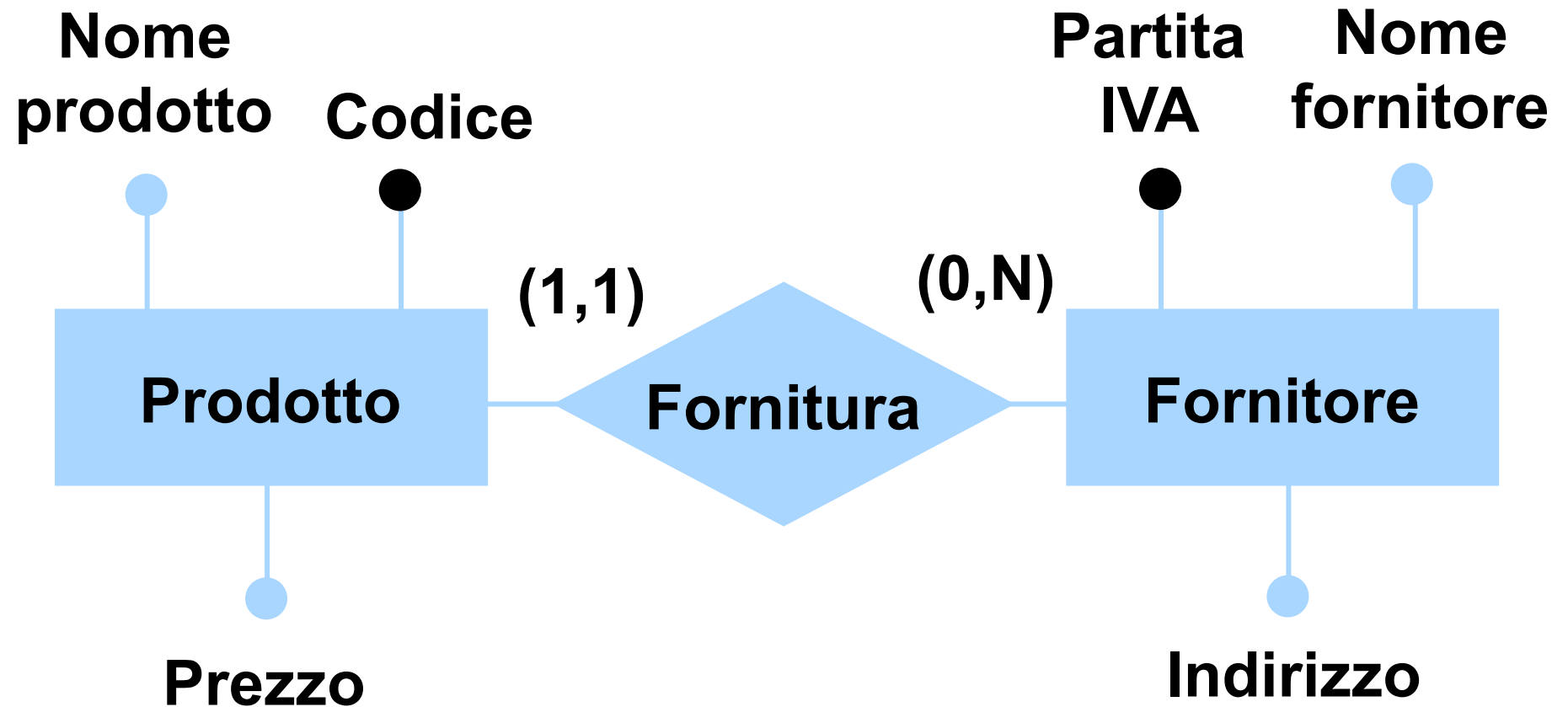
Analisi dell'entità

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

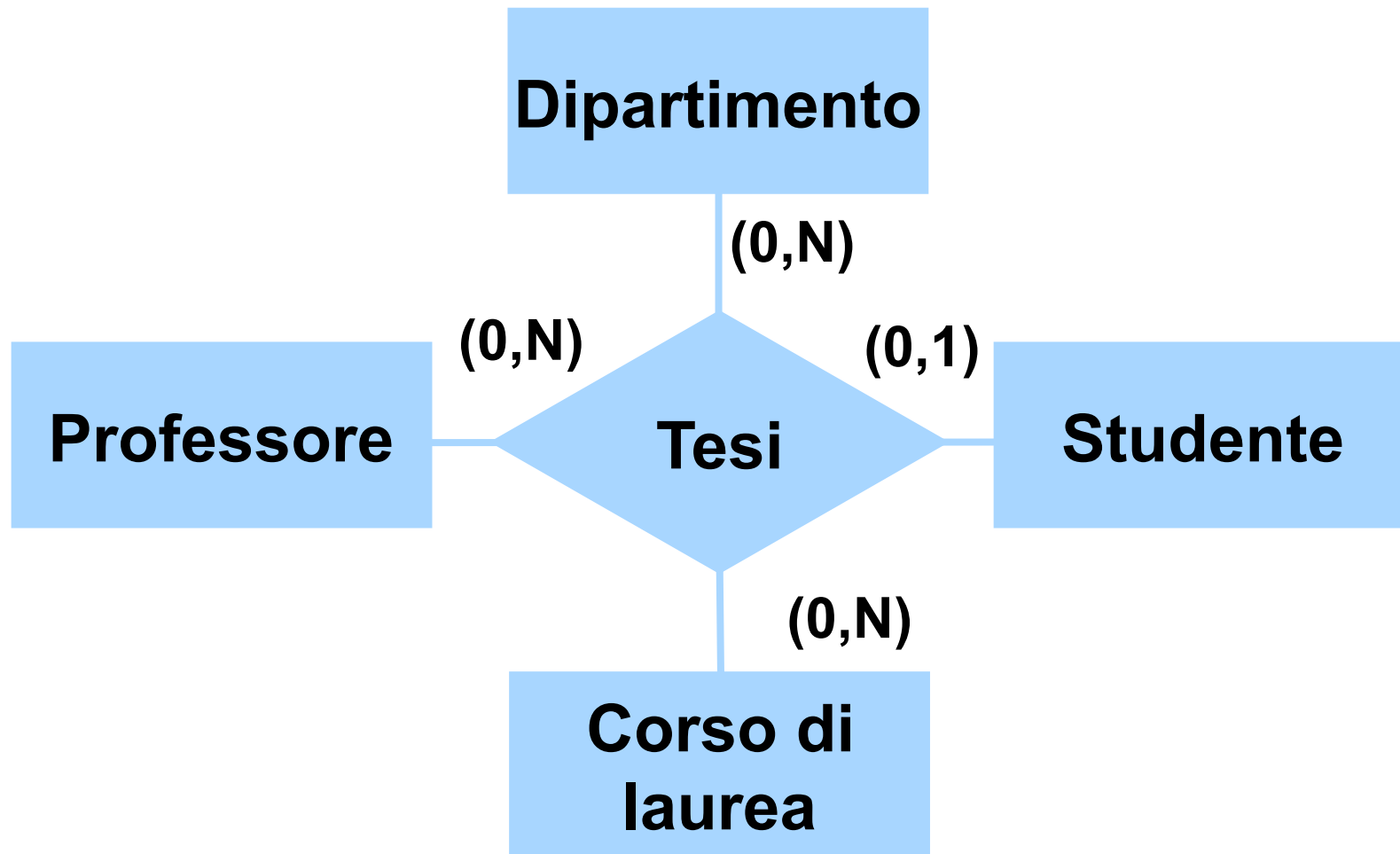
PartitaIVA → NomeFornitore Indirizzo

- Possiamo decomporre sulla base di questa dipendenza

Risultato decomposizione entità



Verifica di normalizzazione su associazioni



Studente → Corso di laurea
Studente → Professore
Professore → Dipartimento

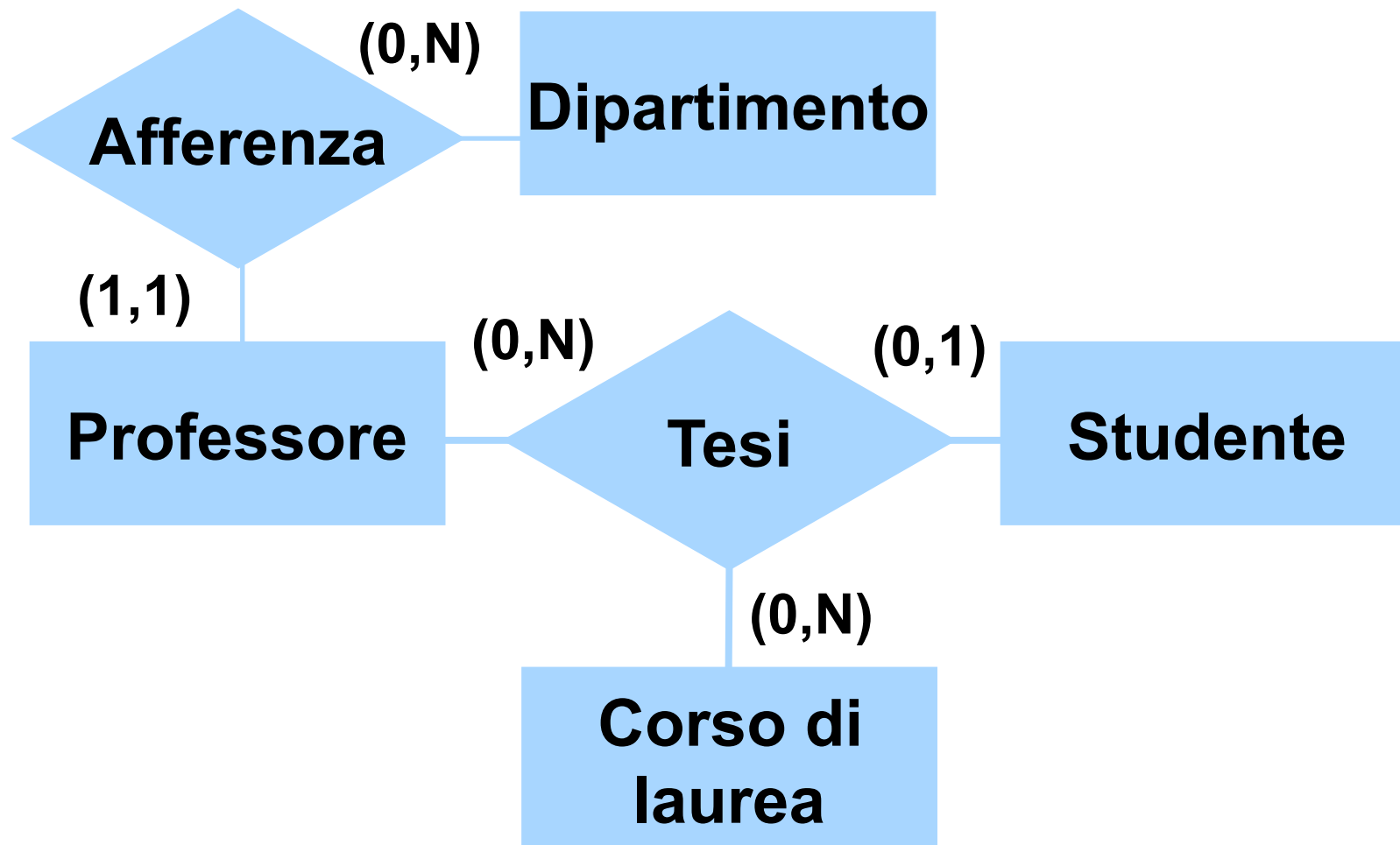
Analisi della relationship

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

Professore → Dipartimento

- Possiamo decomporre sulla base di questa dipendenza

Risultato decomposizione associazione



Ulteriore analisi sulla base delle dipendenze

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

Studiante \rightarrow **CorsoDiLaurea**

Studiante \rightarrow **Professore**

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

Risultato decomposizione associazione (2)

