



**Atzeni, Ceri, Fraternali,
Paraboschi, Torlone**
Basi di dati
Quarta edizione
McGraw-Hill, 2013

Capitolo 9:

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

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

$$Y \rightarrow Z$$

- Esempi:

Impiegato \rightarrow Stipendio

Progetto \rightarrow Bilancio

Impiegato Progetto \rightarrow Funzione

<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)

- 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
- La forma normale richiede che i concetti in una relazione siano omogenei (solo proprietà direttamente associate alla chiave)

Che facciamo se una relazione non soddisfa la BCNF?



- La rimpiazziamo con altre relazioni che soddisfano la BCNF

Come?

- Decomponendo sulla base delle dipendenze funzionali, al fine di separare i concetti

<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

Procedura intuitiva di normalizzazione

- Non valida in generale, ma solo nei "casi semplici"
 - Per ogni dipendenza $X \rightarrow Y$ che viola la BCNF, definire una relazione su XY ed eliminare Y dalla relazione originaria

Non 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

Decomponiamo sulla base delle dipendenze

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

Proviamo a ricostruire



Impiegato	Sede
Rossi	Roma
Verdi	Milano
Neri	Milano

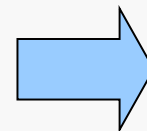
Progetto	Sede
Marte	Roma
Giove	Milano
Saturno	Milano
Venere	Milano

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

- 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



Proviamo a decomporre 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 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>	<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

Teoria della normalizzazione

- I concetti visti possono essere formalizzati in maniera precisa
- Problema: data una relazione r e un insieme di dipendenze funzionali definite su r , generare una decomposizione di r che:
 - Sia senza perdita e conservi le dipendenze
 - Contenga solo relazioni normalizzate
- Faremo riferimento alla 3NF

Implicazione dipendenze funzionali



- Un insieme F di FD **implica** un'altra FD f se ogni relazione che soddisfa tutte le FD in F soddisfa anche f .
- Esempio:
 - $R(\text{Impiegato}, \text{Categoria}, \text{Stipendio})$
 - Le FD $\text{Impiegato} \rightarrow \text{Categoria}$ e $\text{Categoria} \rightarrow \text{Stipendio}$ implicano la FD $\text{Impiegato} \rightarrow \text{Stipendio}$.

Chiusura di un insieme di attributi

- Dati uno schema di relazione $R(U)$, un insieme F di FD definite su U e un insieme di attributi X contenuti in U (cioè $X \subseteq U$): la **chiusura** di X rispetto ad F , indicata con X_F^+ , è l'insieme degli attributi che dipendono funzionalmente da X :

$$X_F^+ = \{ A \mid A \in U \text{ e } F \text{ implica } X \rightarrow A \}$$

- Se A appartiene a X_F^+ allora $X \rightarrow A$ è implicata da F

Calcolo di X_F^+



Input: un insieme X di attributi e un insieme F di dipendenze funzionali

Output: un insieme X_p di attributi.

1. Inizializziamo X_p con l'insieme di input X .
2. Se esiste una FD $Y \rightarrow A$ in F con $Y \subseteq X_p$ e $A \notin X_p$, allora aggiungiamo A a X_p .
3. Ripetiamo il passo (2) fino a quando non ci sono ulteriori attributi che possono essere aggiunti a X_p .

Chiusura e chiave



- Un insieme di attributi K è chiave per uno schema di relazione $R(U)$ su cui è definito un insieme di dipendenze funzionali F se F implica $K \rightarrow U$.
- L'algoritmo appena mostrato può essere utilizzato per verificare se un insieme di attributi è chiave.

Coperture di dipendenze funzionali



- Due insiemi di dipendenze funzionali F_1 ed F_2 sono **equivalenti** se F_1 implica ciascuna dipendenza in F_2 e viceversa.
- Se due insiemi sono equivalenti diciamo anche che ognuno è una **copertura** dell'altro.
- Questa proprietà consente di utilizzare, dato un insieme di dipendenze, un altro, a esso equivalente, ma più semplice.

Proprietà desiderabili di FD

- Un insieme di dipendenze F è:
 - **non ridondante** se non esiste dipendenza $f \in F$ tale che $F - \{f\}$ implica f ;
 - **ridotto** se (i) è non ridondante e (ii) non esiste un insieme F' equivalente a F ottenuto eliminando attributi dai primi membri di una o più dipendenze di F .
- Esempio:

$$F_1 = \{A \rightarrow B; AB \rightarrow C; A \rightarrow C\}$$

$$F_2 = \{A \rightarrow B; AB \rightarrow C\}$$

$$F_3 = \{A \rightarrow B; A \rightarrow C\}$$

Calcolo copertura ridotta

1. Sostituiamo l'insieme dato con quello equivalente che ha tutti i secondi membri costituiti da singoli attributi;
2. Eliminiamo le dipendenze ridondanti;
3. Per ogni dipendenza verifichiamo se esistono attributi eliminabili dal primo membro

In pratica, per ogni dipendenza $Y \rightarrow A \in F$, verifichiamo se esiste $Y \subseteq X$ tale che F è equivalente a $F - \{X \rightarrow A\} \cup \{Y \rightarrow A\}$.

Sintesi di schemi in 3NF



Dati uno schema $R(U)$ e un insieme di dipendenze F su U

1. Viene calcolata una copertura ridotta G di F ;
2. G viene partizionato in sottoinsiemi tali che a ogni insieme appartengono dipendenze che hanno primi membri con la stessa chiusura;
3. Viene costruito un insieme \mathbf{U} di sottoinsiemi di U , uno per ciascuna partizione di dipendenze, con tutti gli attributi coinvolti nella partizione;
4. Se un elemento di \mathbf{U} è propriamente contenuto in un altro, allora esso viene eliminato da \mathbf{U} ;
5. Viene costruito uno schema di relazione $R_i(U_i)$ per ciascun elemento $U_i \in \mathbf{U}$ con associate le dipendenze in G i cui attributi sono tutti contenuti in U_i ;
6. Se nessuno degli U_i è chiave per $R(U)$, allora viene calcolata una chiave K di $R(U)$ e viene aggiunto allo schema generato uno schema di relazione sugli attributi K , senza dipendenze.

Esempio

Schema: R(MCGRDSPA)

FD: $M \rightarrow RSDG$, $MS \rightarrow CD$, $G \rightarrow R$, $D \rightarrow S$, $S \rightarrow D$, $MPD \rightarrow AM$.

• Al passo 1, si ottiene la copertura ridotta:

$M \rightarrow D$, $M \rightarrow G$, $M \rightarrow C$, $G \rightarrow R$, $D \rightarrow S$, $S \rightarrow D$, $MP \rightarrow A$.

• Al passo 2, si partiziona la copertura negli insiemi:

$G_1 = \{ M \rightarrow D; M \rightarrow G; M \rightarrow C \}$, $G_2 = \{ G \rightarrow R \}$,

$G_3 = \{ D \rightarrow S; S \rightarrow D \}$, $G_4 = \{ MP \rightarrow A \}$

• I passi 3, 4 e 5 costruiscono uno schema di relazione per ciascuna partizione (senza eliminazioni), con le dipendenze corrispondenti.

• Il passo 6 non ha effetti, perché MP è chiave per la R.

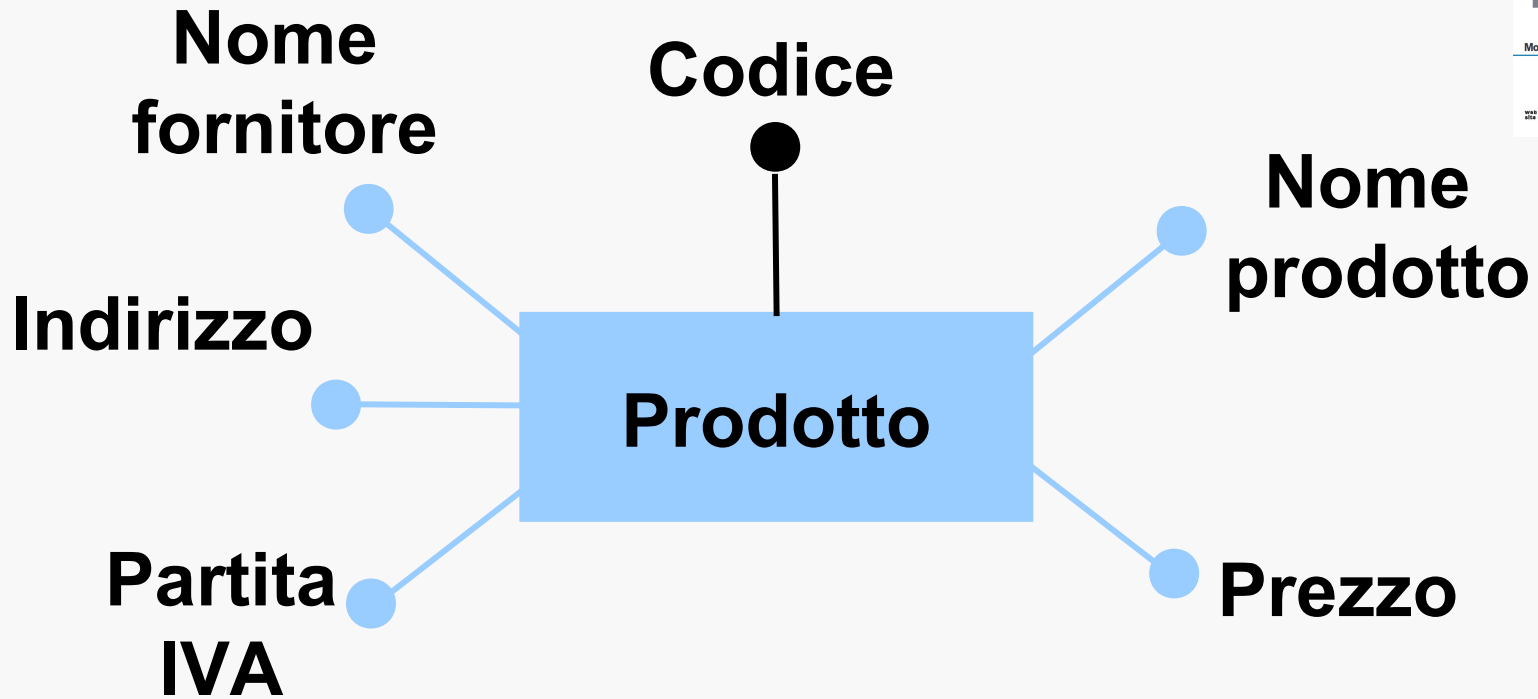
• Quindi, viene generato lo schema con le relazioni:

- $R_1(MDGC)$, con le dipendenze $\{M \rightarrow D; M \rightarrow G; M \rightarrow C\}$
- $R_2(GR)$ con $\{G \rightarrow R\}$
- $R_3(DS)$ con $\{D \rightarrow S; S \rightarrow D\}$
- $R_4(MPA)$ con $\{MP \rightarrow A\}$

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



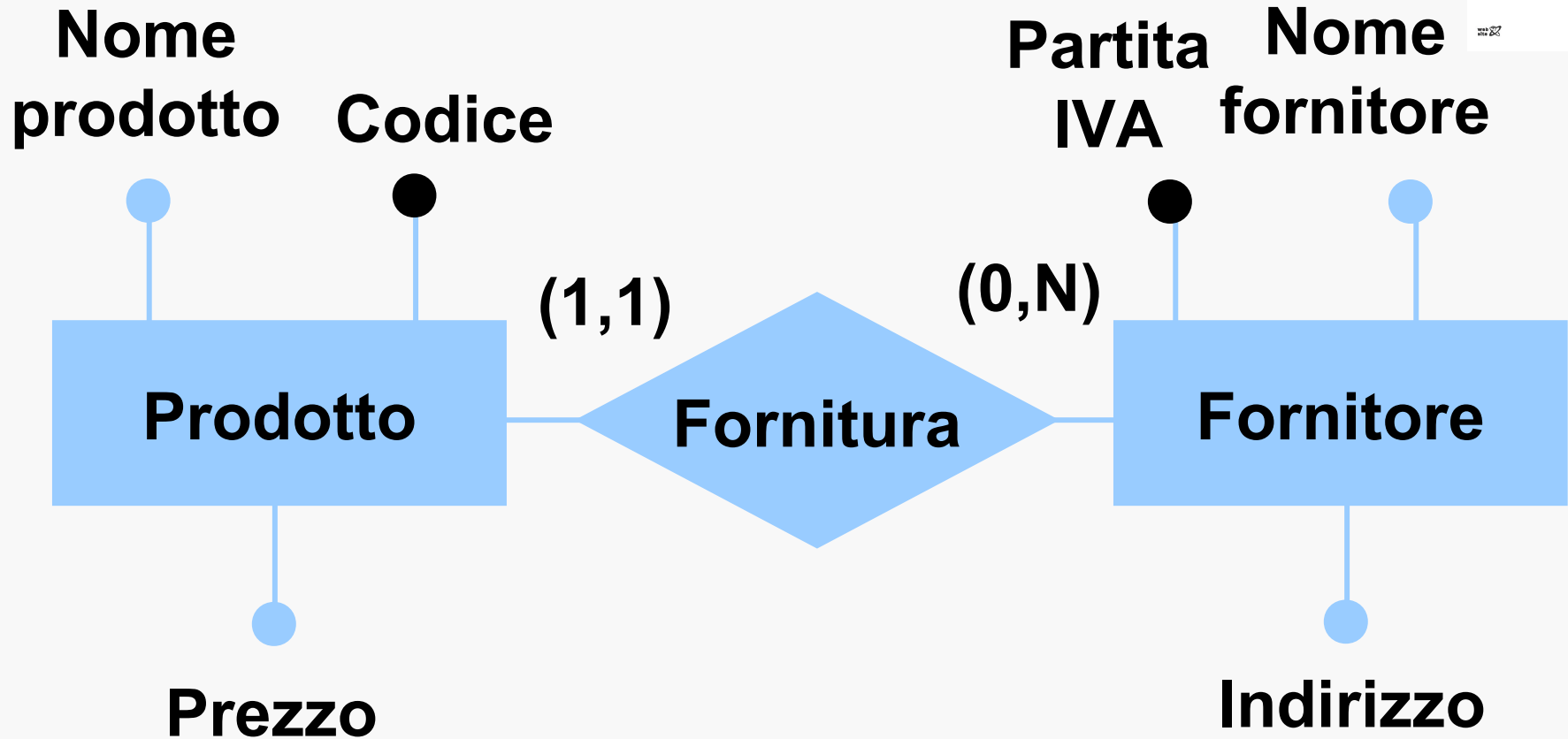
PartitaIVA → NomeFornitore Indirizzo

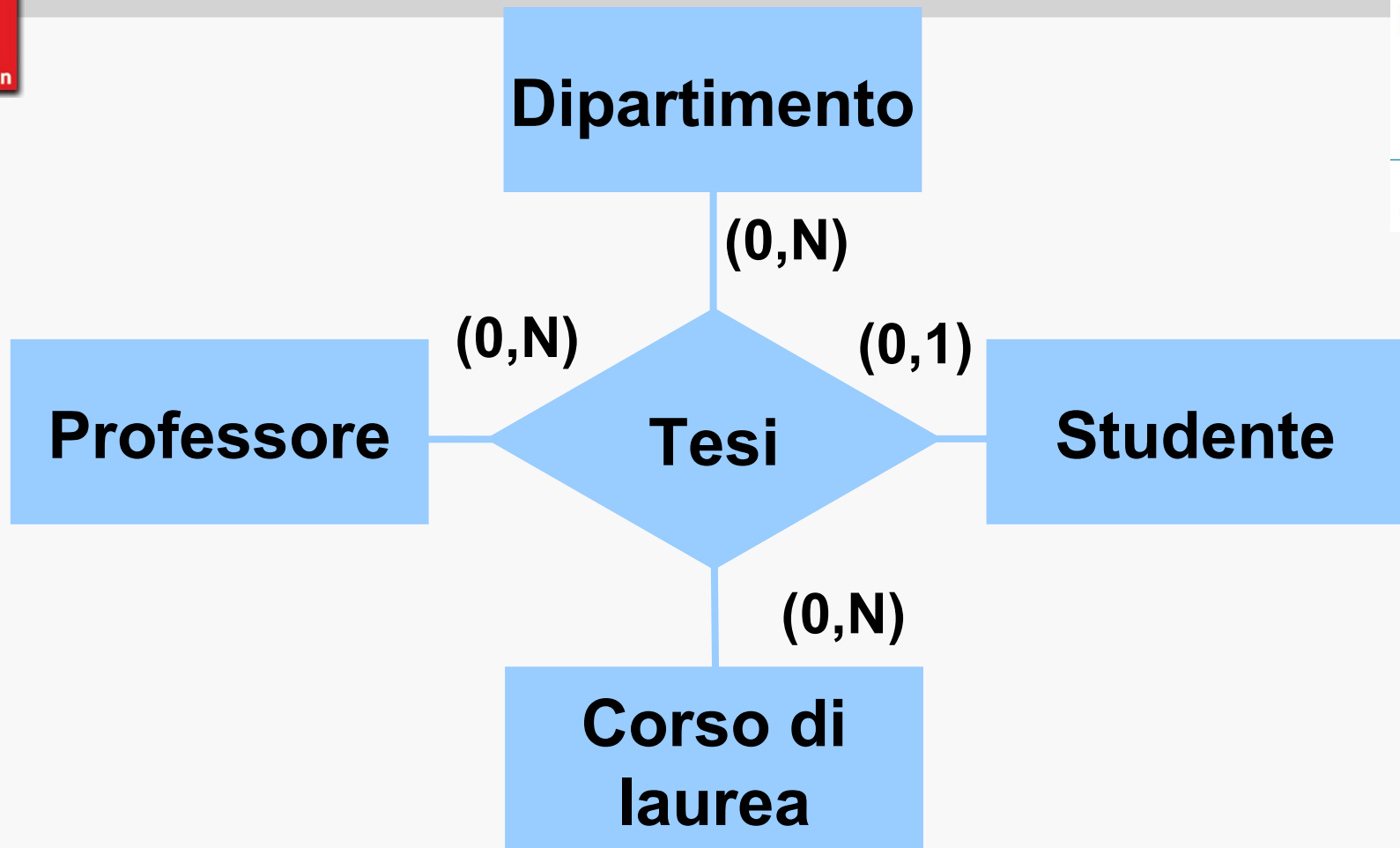
Analisi dell'entità

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

PartitaIVA → NomeFornitore Indirizzo

- Possiamo decomporre sulla base di questa dipendenza





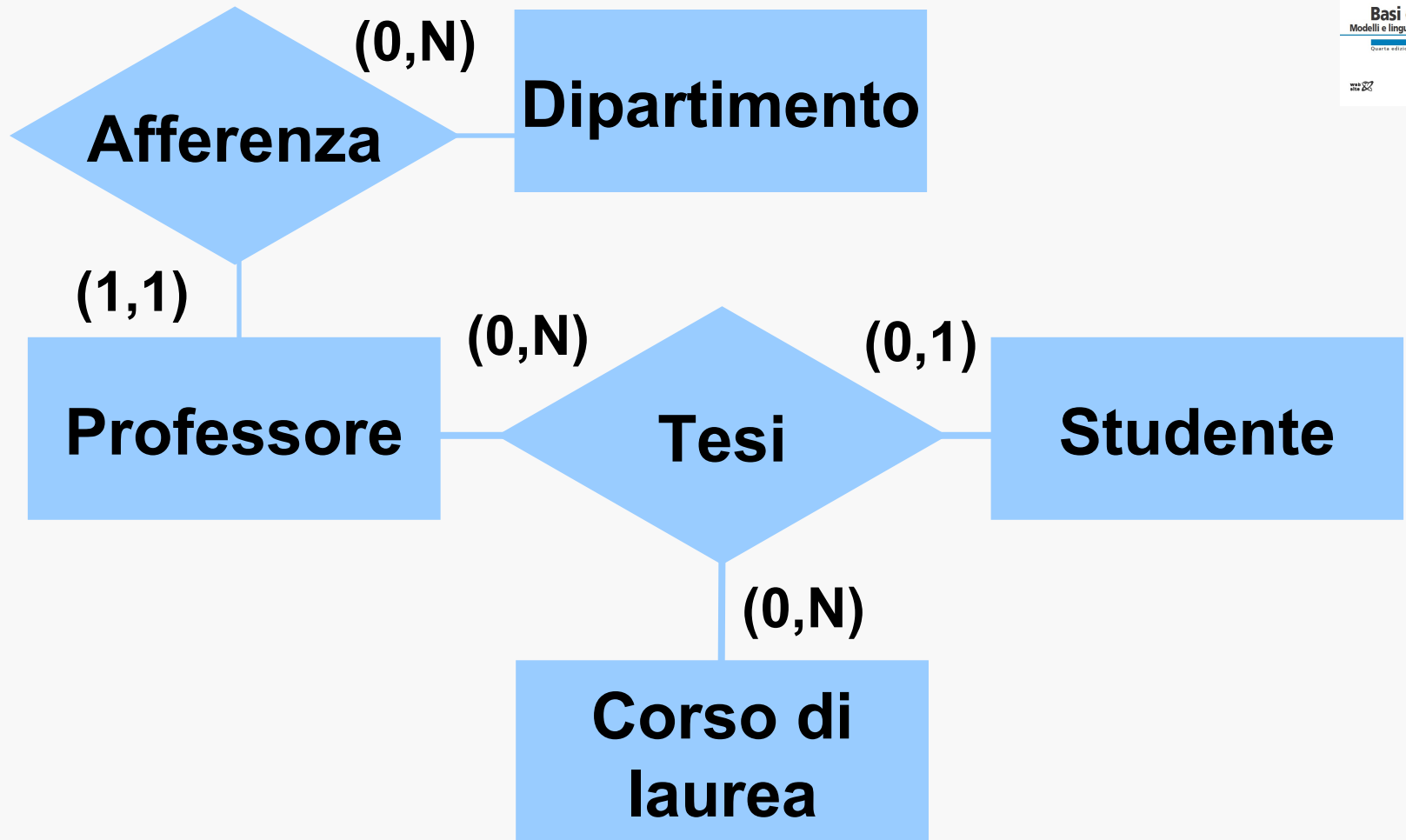
Studente → Corso di laurea
Studente → Professore
Professore → Dipartimento

Analisi della relationship

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

Professore → Dipartimento

- Possiamo decomporre sulla base di questa dipendenza



Ulteriore analisi sulla base delle dipendenze

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

Studiante → CorsoDiLaurea

Studiante → Professore

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

