



Basi Di Dati e di conoscenza

Normalizzazione



Contenuti della lezione

- Anomalie di uno schema
- Normalizzazione
- Dipendenze funzionali
- Forme Normali
 - Prima forma Normale (1NF)
 - Seconda forma Normale (2NF)
 - Terza forma Normale (3NF)
 - Boyce Codd (BCNF)
 - Quarta e quinta forma normale
- Normalizzazione e decomposizione

Semantica di un buon schema

- **Linea guida 1** : Informalmente, ogni tupla in una relazione dovrebbe rappresentare un'entità o un'istanza di relazione. (Si applica alle relazioni individuali e ai loro attributi).
- Gli attributi di diverse entità (dipendenti, dipartimenti, progetti) **non dovrebbero essere mescolati** nella stessa relazione
- Per riferirsi ad altre entità dovrebbero essere usate solo le chiavi esterne.
- **Gli attributi di entità e di relazione diverse** dovrebbero essere tenuti il più possibile separati.
- **Progettare uno schema che possa essere spiegato facilmente relazione per relazione. La semantica degli attributi dovrebbe essere facile da interpretare**

Ridondanze e anomalie di aggiornamento

- La mischiare attributi di più entità può causare problemi
- **Le informazioni vengono memorizzate in modo ridondante spreca spazio di archiviazione**
- Problemi con le anomalie di aggiornamento
 - Anomalie di inserimento
 - Anomalie di eliminazione
 - Anomalie di modifica

Esempio di anomalia di aggiornamento

Si consideri la relazione:

EMP_PROJ (Emp#, Proj#, Ename, Pname, No_hours)

- **Anomalia dell'aggiornamento:** la modifica del nome del numero di progetto P1 da "Fatturazione" a "Customer-Accounting" può causare l'aggiornamento per tutti i 100 dipendenti che lavorano al progetto P1.

Esempio di anomalia di aggiornamento

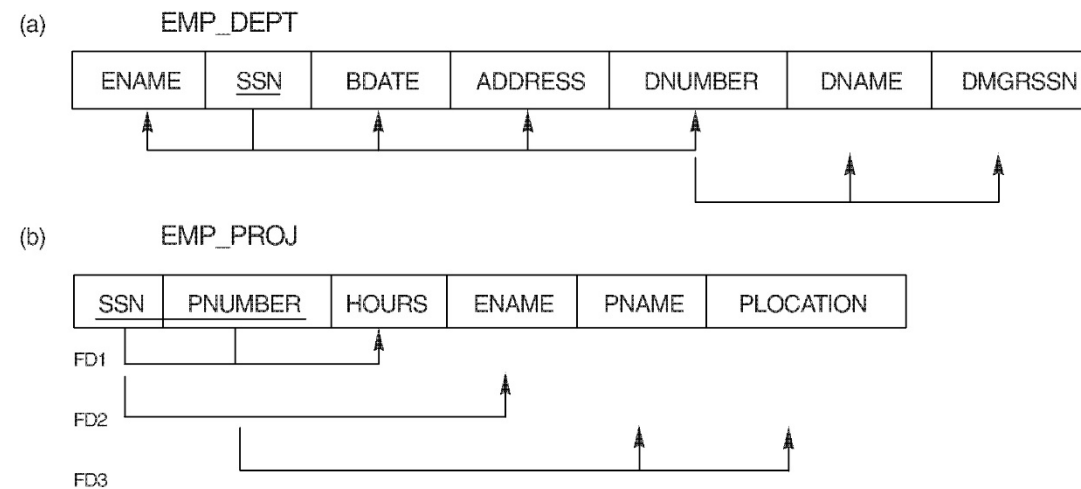
Anomalia d'inserimento: non è possibile inserire un progetto a meno che non abbia un dipendente assegnato:

Al contrario - Impossibile inserire un dipendente a meno che un lui/ lei è assegnato a un progetto.

Anomalia di cancellazione: quando un progetto viene eliminato, ciò comporterà l'eliminazione di tutti i dipendenti che lavorano su quel progetto. In alternativa, se un dipendente è l'unico dipendente di un progetto, l'eliminazione di tale dipendente comporterebbe l'eliminazione del progetto corrispondente.

Schema esempio

Figure 14.3 Two relation schemas and their functional dependencies. Both suffer from update anomalies. (a) The EMP_DEPT relation schema. (b) The EMP_PROJ relation schema.



Schema esempio

Figure 14.4 Example relations for the schemas in Figure 14.3 that result from applying NATURAL JOIN to the relations in Figure 14.2. These may be stored as base relations for performance reasons.

EMP_DEPT

ENAME	SSN	BDATE	ADDRESS	DNUMBER	DNAME	DMGRSSN
Smith, John B.	123456789	1965-01-09	731 Fondren, Houston, TX	5	Research	333445555
Wong, Franklin T.	333445555	1955-12-08	638 Voss, Houston, TX	5	Research	333445555
Zelaya, Alicia J.	999887777	1968-07-19	3321 Castle, Spring, TX	4	Administration	987654321
Wallace, Jennifer S.	987654321	1941-06-20	291 Berry, Bellaire, TX	4	Administration	987654321
Narayan, Ramesh K.	666884444	1962-09-15	975 FireOak, Humble, TX	5	Research	333445555
English, Joyce A.	453453453	1972-07-31	5631 Rice, Houston, TX	5	Research	333445555
Jabbar, Ahmad V.	987987987	1969-03-29	980 Dallas, Houston, TX	4	Administration	987654321
Borg, James E.	888665555	1937-11-10	450 Stone, Houston, TX	1	Headquarters	888665555

EMP_PROJ

SSN	PNUMBER	HOURS	ENAME	PNAME	PLOCATION
123456789	1	32.5	Smith, John B.	ProductX	Bellaire
123456789	2	7.5	Smith, John B.	ProductY	Sugarland
666884444	3	40.0	Narayan, Ramesh K.	ProductZ	Houston
453453453	1	20.0	English, Joyce A.	ProductX	Bellaire
453453453	2	20.0	English, Joyce A.	ProductY	Sugarland
333445555	2	10.0	Wong, Franklin T.	ProductY	Sugarland
333445555	3	10.0	Wong, Franklin T.	ProductZ	Houston
333445555	10	10.0	Wong, Franklin T.	Computerization	Stafford
333445555	20	10.0	Wong, Franklin T.	Reorganization	Houston
999887777	30	30.0	Zelaya, Alicia J.	Newbenefits	Stafford
999887777	10	10.0	Zelaya, Alicia J.	Computerization	Stafford
987987987	10	35.0	Jabbar, Ahmad V.	Computerization	Stafford
987987987	30	5.0	Jabbar, Ahmad V.	Newbenefits	Stafford
987654321	30	20.0	Wallace, Jennifer S.	Newbenefits	Stafford
987654321	20	15.0	Wallace, Jennifer S.	Reorganization	Houston
888665555	20	null	Borg, James E.	Reorganization	Houston

© Addison Wesley Longman, Inc. 2000, Elmasri/Navathe, Fundamentals of Database Systems, Third Edition

Anomalie e valori nulli

- **Linea guida 2:** Progettare uno schema che non risenta delle anomalie di inserimento, cancellazione e aggiornamento. Se sono presenti, annotarle in modo che le applicazioni possano tenerne conto.
- **Linea guida 3:** Le relazioni dovrebbero essere progettate in modo tale che le loro tuple abbiano il minor numero possibile di **valori NULL**
- Gli attributi che sono spesso NULL potrebbero essere collocati in relazioni separate (con la chiave primaria)
- Motivi per i nulli:
 - **attributo non applicabile o non valido**
 - **attributo valore sconosciuto (può esistere)**
 - **valore noto per esistere, ma non disponibile**

Tuple spurie

- Un progetto non corretto per un database relazionale può causare risultati errati per alcune operazioni JOIN
- La **proprietà "lossless join"** viene utilizzata per garantire risultati significativi per le operazioni di join
- **Linea guida 4:** Le relazioni dovrebbero essere progettate per soddisfare la condizione di lossless join. Non si dovrebbero creare tuple spurie facendo un natural-join di tutte le relazioni.

Contenuti della lezione

- Anomalie di uno schema
- Normalizzazione
- Dipendenze funzionali
- Forme Normali
 - Prima forma Normale (1NF)
 - Seconda forma Normale (2NF)
 - Terza forma Normale (3NF)
 - Boyce Codd (BCNF)
 - Quarta e quinta forma normale
- Normalizzazione e decomposizione

Normalizzazione

- La **normalizzazione** è una formalizzazione teorica di alcuni problemi che possono emergere durante l'utilizzo, l'interrogazione e la gestione dei dati in un database e che possono impedire o rendere complicato l'uso delle informazioni. Non sempre è applicabile.
- Permette di costruire un DB corretto e ben definito.

Normalizzazione

- La normalizzazione quindi è un procedimento utile per **l'eliminazione della ridondanza** delle informazioni e per ridurre il rischio di inconsistenza della base dati.
- Di fatto riduce la dimensione delle relazioni a partire da relazioni con concetti tra loro indipendenti.

Normalizzazione

- La normalizzazione dei dati può essere considerata come un processo di analisi degli schemi forniti, basato sulle loro dipendenze funzionali e chiavi primarie, per raggiungere le proprietà desiderate di
 1. **Minimizzazione della ridondanza**
 2. **Minimizzazione delle anomalie di inserimento, cancellazione, modifica**

Contenuti della lezione

- Anomalie di uno schema
- Normalizzazione
- Dipendenze funzionali
- Forme Normali
 - Prima forma Normale (1NF)
 - Seconda forma Normale (2NF)
 - Terza forma Normale (3NF)
 - Boyce Codd (BCNF)
 - Quarta e quinta forma normale
- Normalizzazione e decomposizione

Dipendenza funzionale

- Le **dipendenze funzionali** (FD) sono usate per specificare misure formali della "bontà" dei progetti relazionali
- Le FD e le chiavi sono usati per definire le **forme normali** per le relazioni
- Le FD sono vincoli che derivano dal significato e dalle interrelazioni degli attributi dei dati
- Un insieme di attributi X **determina funzionalmente** un insieme di attributi Y se il valore di **X determina un valore univoco per Y**

Dipendenze funzionali

- Si ha **dipendenza funzionale** tra attributi quando il valore di un insieme di attributi A determina un singolo valore dell'attributo B e si indica con $A \rightarrow B$. Si dice anche che B dipende da A o che A è un determinante per B
- Se un attributo è chiave candidata di una relazione allora è un determinante di ogni attributo della relazione e viceversa, **un attributo che determina tutti gli altri attributi è chiave candidata.**
- Si ha **dipendenza transitiva** quando A determina B e B determina C. Si dice allora che C dipende transitivamente da A

Dipendenze funzionali: esempio

- Schema 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

Esempio: 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)

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

Dipendenze funzionali

- Un FD è una **proprietà degli attributi** nello schema R
- Il vincolo deve valere su ogni istanza della relazione $r(R)$
- **Se K è una chiave di R, allora K determina funzionalmente tutti gli attributi in R (poiché non abbiamo mai due tuple distinte con $t1[K]=t2[K]$)**

Contenuti della lezione

- Anomalie di uno schema
- Normalizzazione
- Dipendenze funzionali
- Forme Normali
 - Prima forma Normale (1NF)
 - Seconda forma Normale (2NF)
 - Terza forma Normale (3NF)
 - Boyce Codd (BCNF)
 - Quarta e quinta forma normale
- Normalizzazione e decomposizione

Prima Forma Normale (1NF)

- Uno schema di relazione $R(X)$ con X insieme di attributi, è in **1NF** se **ogni attributo appartenente ad X è un attributo semplice.**
- Un attributo è un **attributo semplice** se il suo valore è unico e indivisibile in una ennupla

Prima Forma Normale (1NF)

Esempio

- Tabella non in 1NF: l'attributo Figli a carico contiene più valori

Impiegati

Codice	Cognome	Nome	Data Nascita	Figli a carico
001	Rossi	Mario	01/01/1978	Luca Serena
002	Verdi	Luca	02/04/1959	Marzia Ilaria

Prima Forma Normale -decomposizione

Esempio

Tabella in 1NF: scomposizione in due tabelle

Codice	Cognome	Nome	Data Nascita
001	Rossi	Mario	01/01/1978
002	Verdi	Luca	02/04/1959

Codice	Codice Figlio	Nome
001	01	Luca
001	02	Serena
002	01	Marzia
002	02	Ilaria

Seconda Forma Normale (2NF)

- **Uno schema di relazione $R(X)$ è in 2NF** se è in 1NF e se ogni **attributo non primo (non facente parte della chiave)** di $R(X)$ dipende funzionalmente e completamente da ogni chiave di $R(X)$.

Seconda Forma Normale (2NF)

Esempio

Tabella non in 2NF: tutte le colonne corrispondenti agli attributi no chiave non dipendono dall'intera chiave primaria

Inventario

<u>CodArticolo</u>	<u>CodMagazzino</u>	DescArticoli	Quantità	IndirizzoMagazzino
--------------------	---------------------	--------------	----------	--------------------

Seconda Forma Normale (2NF)

Esempio

Tabella in 2NF

Inventario

<u>CodArticolo</u>	<u>CodMagazzino</u>	Quantità
--------------------	---------------------	----------

Articoli

<u>CodArticolo</u>	DescArticoli
--------------------	--------------

Magazzino

<u>CodMagazzino</u>	IndirizzoMagazzino
---------------------	--------------------

Terza Forma Normale (3NF)

- **Uno schema di relazione $R(X)$ è in 3NF** se è in 1NF e se ogni attributo **non primo** (non facente parte della chiave) di $R(X)$ è dipendente in modo non transitivo da ogni chiave di $R(X)$.

Terza Forma Normale (3NF)

Esempio

Tabella non in 3NF:

- telefono del Reparto ripetuto per ogni Impiegato di quel Reparto (ridondanza)
- se il telefono cambia, occorre modificare molte righe
- con errori di aggiornamento, si avrebbero telefoni differenti
- se un Reparto non ha impiegati, non si può conoscere il suo telefono

Impiegati

<u>CodImpiegato</u>	Nome	Reparto	TelefonoReparto
---------------------	------	---------	-----------------

Terza Forma Normale (3NF)

Esempio

Tabella non in 3NF, dipendenze funzionali:

- $\text{CodImpiegato} \rightarrow \text{Reparto}$
- $\text{Reparto} \rightarrow \text{TelefonoReparto}$

Impiegati

<u>CodImpiegato</u>	Nome	Reparto	TelefonoReparto
---------------------	------	---------	-----------------

Terza Forma Normale (3NF)

Esempio

Tabella in 3NF,

Impiegati

<u>CodImpiegato</u>	Nome	Reparto
---------------------	------	---------

Reparto

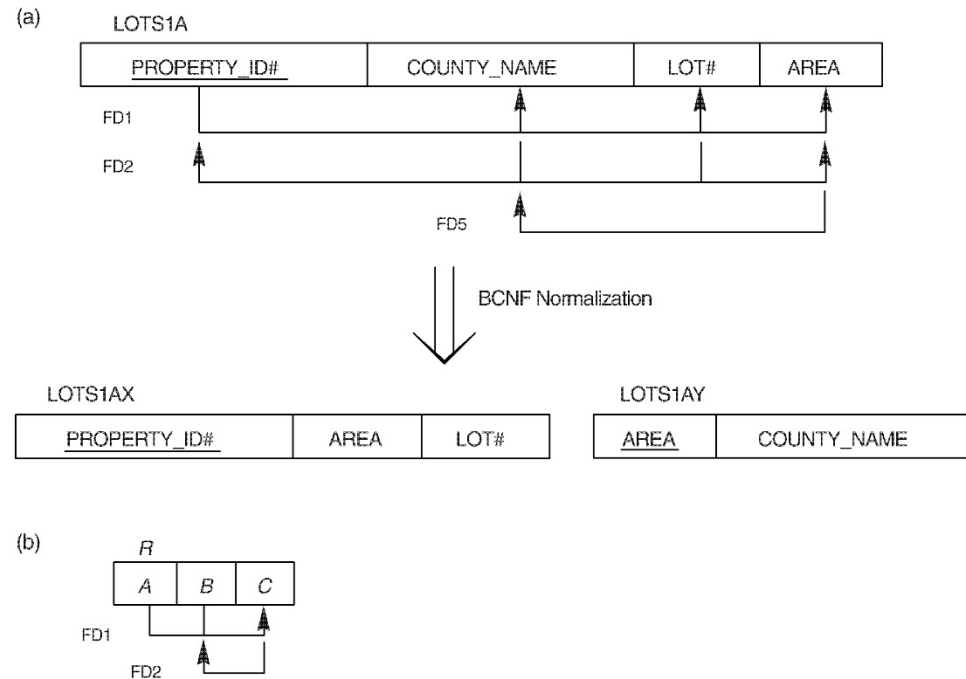
<u>Reparto</u>	TelefonoReparto
----------------	-----------------

Forma normale di Boyce-Codd

- Una relazione è in **forma normale di Boyce-Codd (BCNF)** se è in 1NF se ogni volta che vale FD $X \rightarrow A$ in R, allora X è una superchiave di R
- Ogni forma normale è strettamente più forte della precedente
 - Ogni relazione 2NF è in 1NF
 - Ogni relazione 3NF è in 2NF
 - Ogni relazione BCNF è in 3NF
- Esistono relazioni che sono in 3NF ma non in BCNF
- L'obiettivo è avere ogni relazione in BCNF (o 3NF)

Forma normale di Boyce-Codd

Figure 14.12 Boyce-Codd normal form. (a) BCNF normalization with the dependency of FD2 being “lost” in the decomposition.
(b) A relation R in 3NF but not in BCNF.



© Addison Wesley Longman, Inc. 2000, Elmasri/Navathe, Fundamentals of Database Systems, Third Edition

Quarta e quinta forma normale

- La quarta e la quinta forma normale risolvono i problemi che si possono creare quando nella relazione sono presenti attributi multivalore, cioè attributi che possono assumere più valori in corrispondenza dello stesso valore di un altro attributo.

Quale forma Normale

- E' sufficiente rappresentare le relazioni in 3FN che, come si può dimostrare, ha il pregio di essere sempre ottenibile senza **perdita di informazioni** e senza **perdita di dipendenze funzionali**.
- Non è così invece per la forma normale di Boyce-Codd: ci sono relazioni che non possono essere normalizzate nella forma di Boyce-Codd senza perdita di dipendenze funzionali

Forma normale di Boyce-Codd

- Relazione in 3NF ma non in BCNF

STUDENT, COURSE → INSTRUCTOR
STUDENT, INSTRUCTOR → COURSE
INSTRUCTOR → COURSE

STUDENT, COURSE, INSTRUCTOR sono tutti attributi **primi** quindi la 3° FD non viola la 3NF ma la BCNF si

Figure 14.13 A relation TEACH that is in 3NF but not in BCNF.

TEACH		
STUDENT	COURSE	INSTRUCTOR
Narayan	Database	Mark
Smith	Database	Navathe
Smith	Operating Systems	Ammar
Smith	Theory	Schulman
Wallace	Database	Mark
Wallace	Operating Systems	Ahamad
Wong	Database	Omicinski
Zelaya	Database	Navathe

Contenuti della lezione

- Anomalie di uno schema
- Normalizzazione
- Dipendenze funzionali
- Forme Normali
 - Prima forma Normale (1NF)
 - Seconda forma Normale (2NF)
 - Terza forma Normale (3NF)
 - Boyce Codd (BCNF)
 - Quarta e quinta forma normale
- Normalizzazione e decomposizione

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

Decomposizione: 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

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 è possibile



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

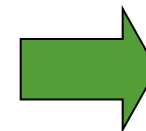
Diversa dalla relazione di partenza!



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



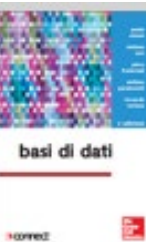
Proviamo a decomporre senza perdita

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



Perdita di una FD

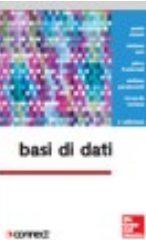
Impiegato	Progetto	Sede
Rossi	Marte	Roma
Verdi	Giove	Milano
Verdi	Venere	Milano
Neri	Saturno	Milano
Neri	Venere	Milano
Neri	Marte	Milano

~~Progetto → Sede~~



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 originali

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 3NF



- 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 (**tutti gli attributi sono primi**)

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 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:

$F_1 = \{A \rightarrow B; AB \rightarrow C; A \rightarrow C\}$ ridondante e equivalente a F_2

$F_2 = \{A \rightarrow B; AB \rightarrow C\}$ Non ridondante ma non ridotto

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

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:

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

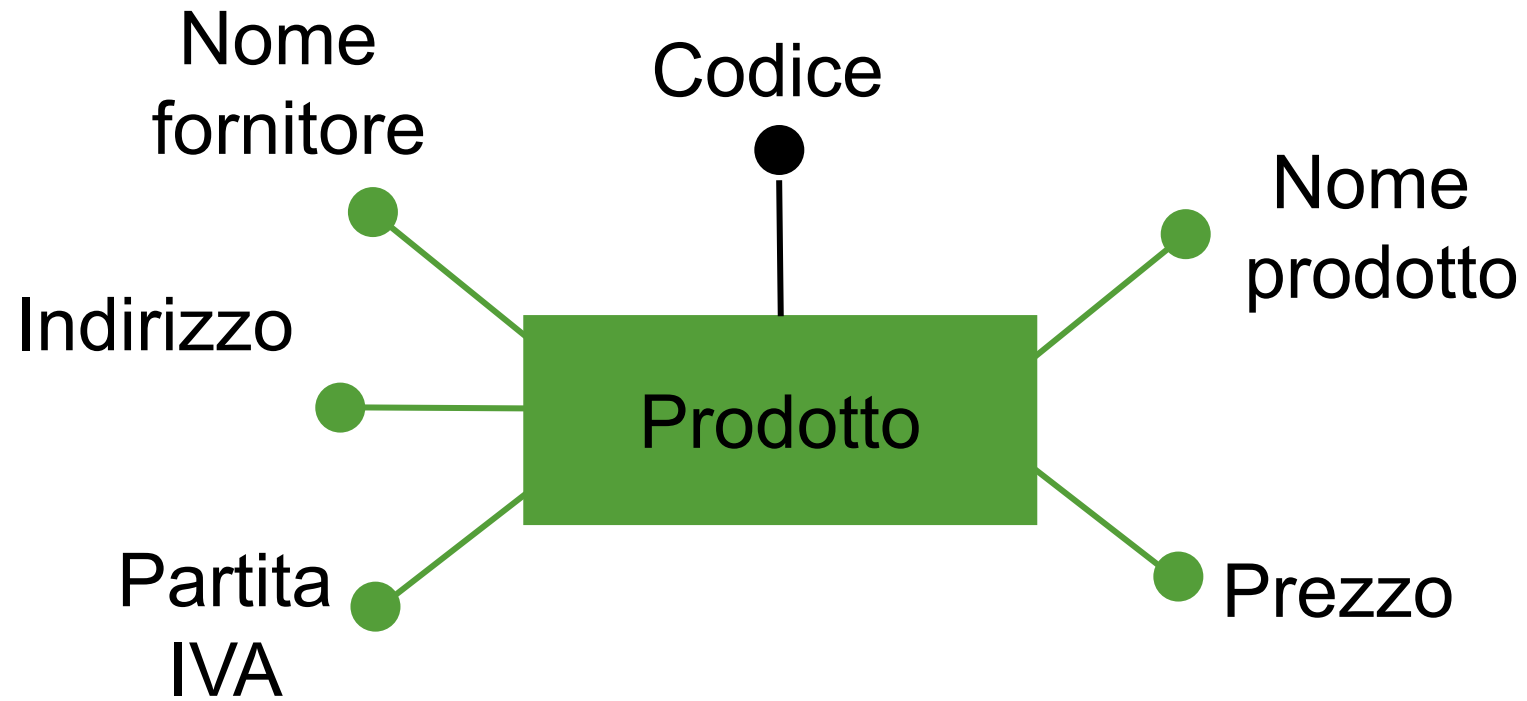
$G_3 = \{ D \rightarrow S; S \rightarrow D \}$, $G_4 = \{ MP \rightarrow A \}$

- 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

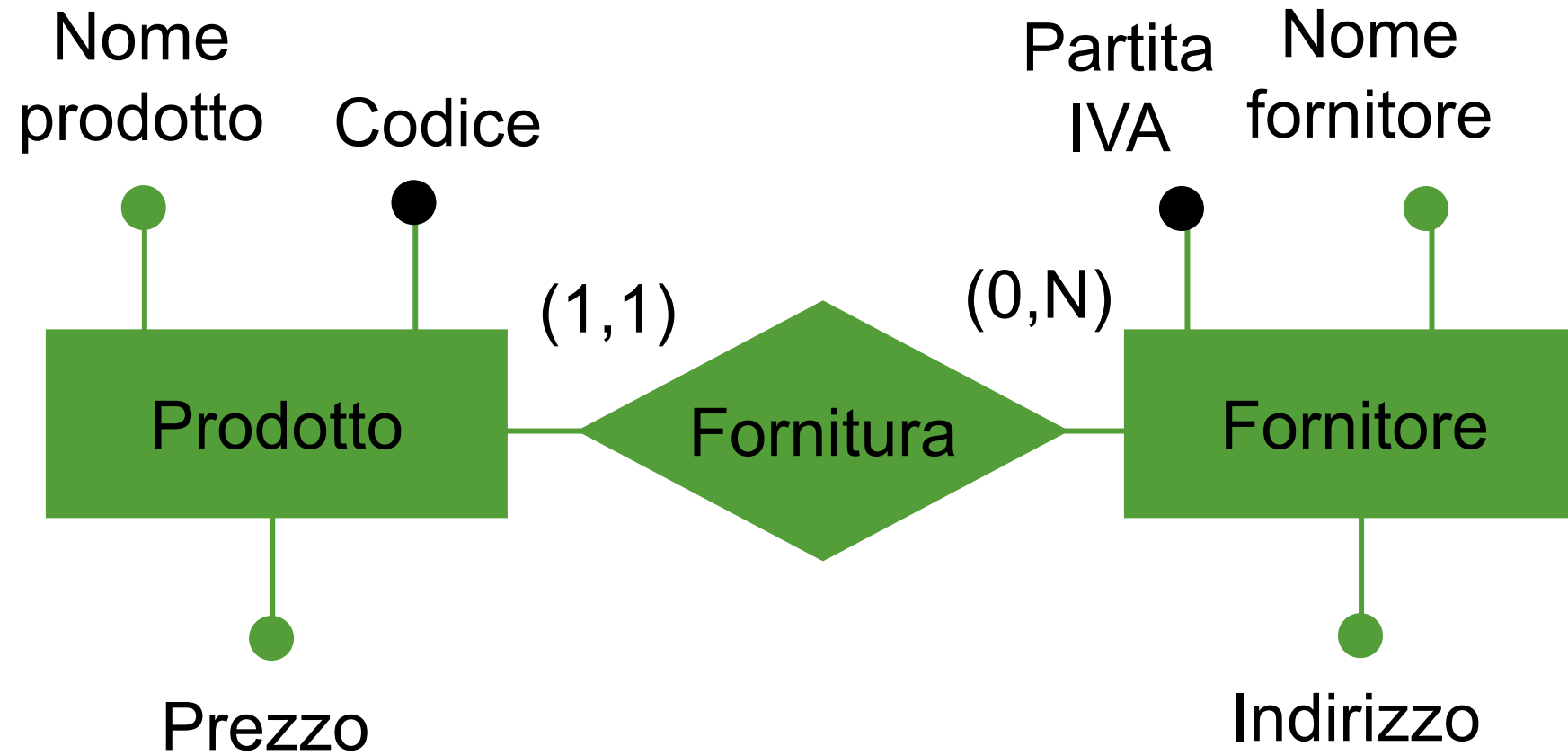
Analisi dell'entità

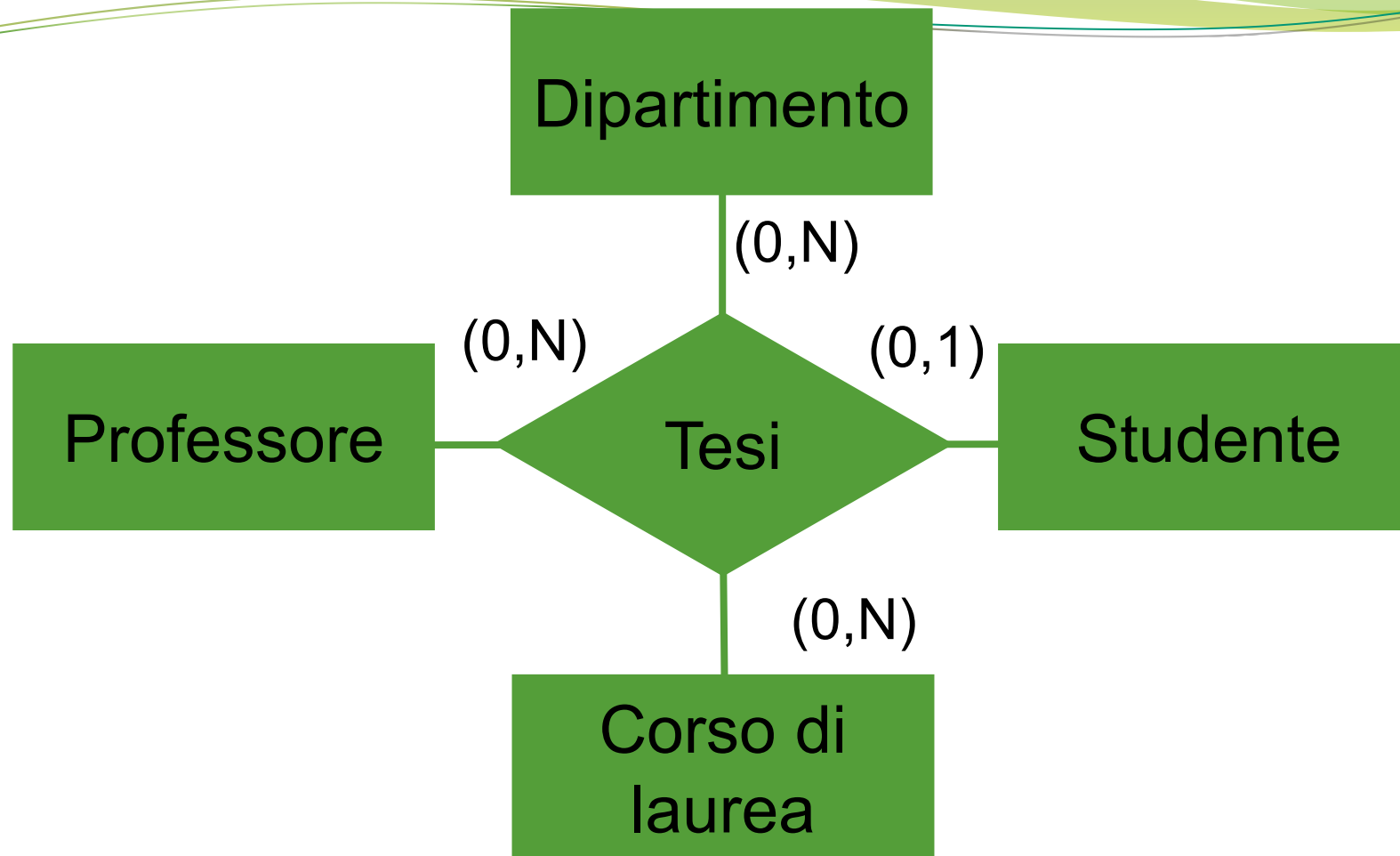


- L'entità viola la forma normale a causa della dipendenza:

PartitaIVA → NomeFornitore Indirizzo

- Possiamo decomporre sulla base di questa dipendenza





Studente → Corso di laurea
Studente → Professore
Professore → Dipartimento

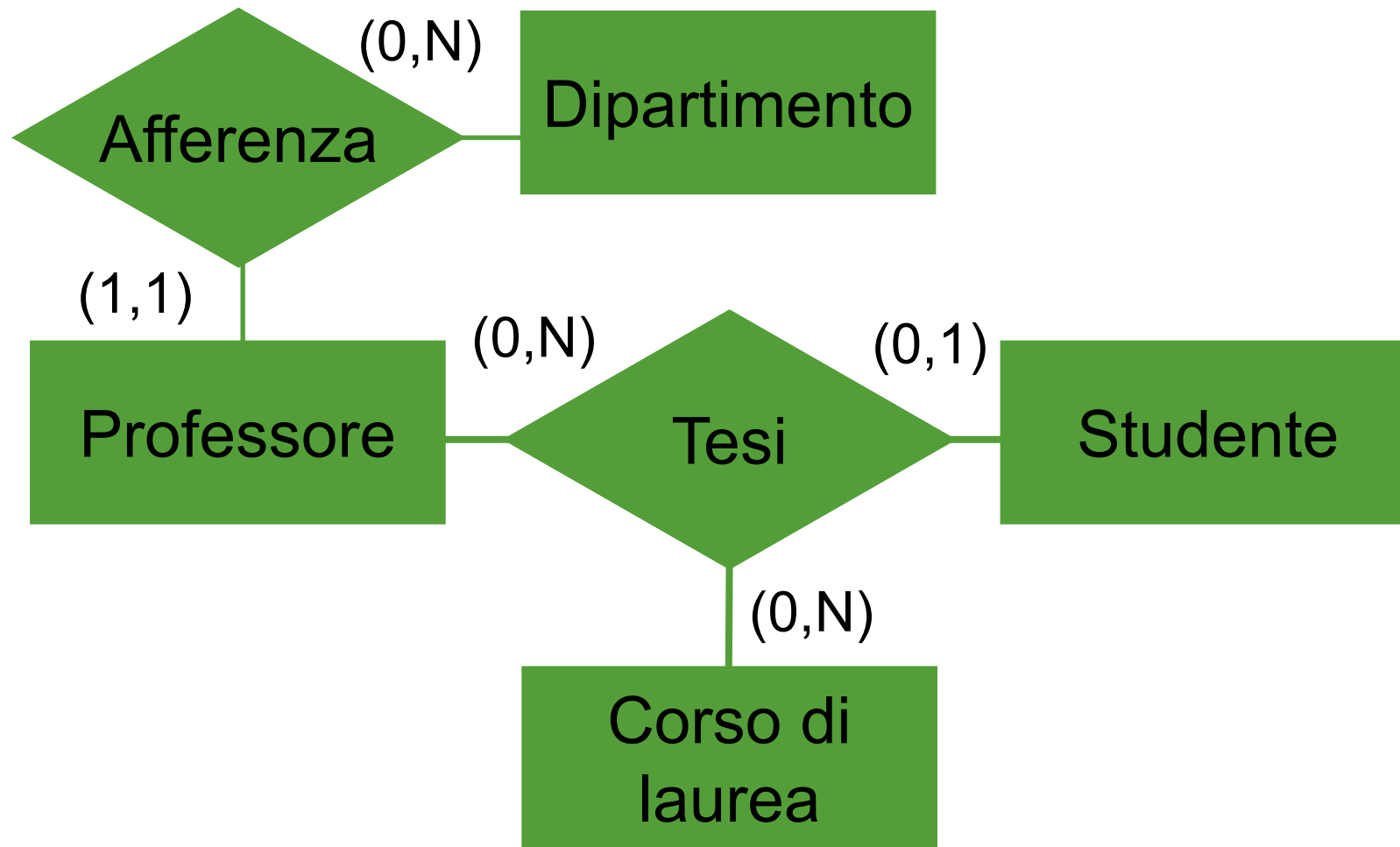
Analisi della relationship



- 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

