

Forme normali

1NF, 2NF, 3NF, BCNF

Prima forma normale (1NF)

Uno schema di relazione $R(X)$ è in prima forma normale se ogni attributo di X è un attributo semplice, ovvero con dominio atomico.

Nel modello relazionale una relazione $R(X)$ è per definizione in 1NF.

Ne segue che ogni attributo di $R(X)$ non primo (cioè che non fa parte di alcuna chiave) dipende funzionalmente da una chiave di $R(X)$.

Relazione non in 1NF

Impiegato	Progetto	Stipendio	Indirizzo	Funzione
Silvietti	Marte Saturno	50	via:Roma; città:Napoli; nc: 20	Direttore
Marcuccio	Marte Giove	30	Via:Claudio; città: Milano; nc=10	Progettista

Progetto è un attributo multivalore

Indirizzo è un attributo strutturato

Seconda forma normale (2NF)

Uno schema di relazione $R(X)$ è in 2NF se:

- è in 1NF
- Ogni attributo non primo dipende completamente da ogni chiave di $R(X)$

Per verificare il soddisfacimento della 2NF

- Esaminare le parti sinistre delle DF che contengano attributi primi.

Se ogni chiave di $R(X)$ ha un solo attributo allora $R(X)$ è in seconda forma normale

Tabella in 1NF e non in 2NF

Magazzini

Articolo	Magazzino	Quantità	Indirizzo
scarpe	NA1	2500	v. Leopardi 17, Napoli
scarpe	RM1	4500	v. S. Maria Maggiore 3, Napoli
pantaloni	NA1	3000	v. Leopardi 17, Napoli

DF1: Articolo, Magazzino → Indirizzo

DF2: Magazzino → Indirizzo

Indirizzo dipende anche da parte della chiave Articolo, Magazzino

NORMALIZZAZIONE

Decomposizione

Uno schema di relazione $R(X)$ non in 2NF si normalizza decomponendo le relazioni di partenza in relazioni che soddisfano la 2NF.

La relazione viene decomposta sulla base di "attributi comuni" in modo tale che :

- Le relazioni ottenute soddisfino la 2NF
- la decomposizione non alteri il contenuto informativo della base di dati

NORMALIZZAZIONE

Decomposizione in 2NF

ArticoliInMagazzino

Articolo	Magazzino	Quantità
scarpe	NA1	2500
scarpe	RM1	2500
pantaloni	NA1	3000

IndirizzoMagazzino

Magazzino	Indirizzo
NA1	v. Leopardi 17, Napoli
RM1	v. S. Maria Maggiore 3, Napoli

NORMALIZZAZIONE

Decomposizione senza perdita

Si introduce il seguente requisito

- Uno schema $R(X)$ si decompone senza perdita sugli attributi comuni Y_C negli schemi $R_1(X_1)$ e $R_2(X_2)$ se, per ogni istanza legale r su $R(X)$, il join naturale delle proiezioni di r su X_1 e X_2 è uguale a r stessa

$$\Pi_{X_1}(r) \bowtie \Pi_{X_2}(r) = r$$

$$\text{con } X_1 \cup X_2 = X$$

$$\text{e } X_1 \cap X_2 = Y_C$$

NORMALIZZAZIONE

Decomposizione senza perdita

Una decomposizione con perdita può generare tuple spurie

Per decomporre senza perdita è necessario e sufficiente che il **join naturale** sia eseguito su una **superchiave** di uno dei due sottoschemi, ovvero che valga una delle due dipendenze funzionali:

$$X1 \cap X2 \rightarrow X1 \text{ oppure } X1 \cap X2 \rightarrow X2$$

Cioè la decomposizione senza perdita è garantita se gli **attributi comuni** contengono una chiave per almeno una delle relazioni decomposte.

Normalizzazione

Decomposizione: esempio Magazzini

Nell'esempio precedente la decomposizione:
 ArticoliInMagazzino (Articolo, Magazzino, Quantità)
 IndirizzoMagazzino (Magazzino, Indirizzo)

è senza perdita.

Infatti la decomposizione è effettuata sull'attributo comune **Magazzino** che risulta chiave della relazione **IndirizzoMagazzino** e pertanto è senza perdita.

Normalizzazione

Decomposizione con perdita

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

Impiegato → Sede

Progetto → Sede

Decomposizione su attributo comune **Sede**

<u>Impiegato</u>	Sede
Rossi	Roma
Verdi	Milano
Neri	Milano

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

Normalizzazione

Sintesi che denota la perdita

<u>Impiegato</u>	Sede
Rossi	Roma
Verdi	Milano
Neri	Milano

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

Diversa dalla relazione di partenza!

Tuple spurie

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

Normalizzazione

Una decomposizione senza perdita

13

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

Decomposizione su
attributo comune
Impiegato

Impiegato	Sede
Rossi	Roma
Verdi	Milano
Neri	Milano

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

Normalizzazione



Conservazione delle dipendenze

14

Una decomposizione preserva le dipendenze se ciascuna delle dipendenze funzionali dello schema originario coinvolge attributi che compaiono tutti insieme in uno degli schemi decomposti

- Nell'esempio Magazzini la dipendenza **Magazzino → Indirizzo** viene mantenuta

Normalizzazione



Tabella Magazzini

15

Magazzini

Articolo	Magazzino	Quantità	Indirizzo
scarpe	NA1	2500	v. Leopardi 17, Napoli
scarpe	RM1	4500	v. S. Maria Maggiore 3, Napoli
pantaloni	NA1	3000	v. Leopardi 17, Napoli

DF: Magazzino → Indirizzo

Normalizzazione



Decomposizione su Magazzino

16

ArticoliInMagazzino

Articolo	Magazzino	Quantità
scarpe	NA1	2500
scarpe	RM1	2500
pantaloni	NA1	3000

IndirizzoMagazzino

Magazzino	Indirizzo
NA1	v. Leopardi 17, Napoli
RM1	v. S. Maria Maggiore 3, Napoli

Normalizzazione



Esempio che non conserva le DF

17

DF1: Impiegato → Sede
DF2: Progetto → Sede

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

Decomposizione su
attributo comune
Impiegato

Impiegato	Sede
Rossi	Roma
Verdi	Milano
Neri	Milano

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

NORMALIZZAZIONE



... non conserva le DF(2/4)

18

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

Ricordiamo che le dipendenze sullo schema
originario sono

- Impiegato → Sede
- Progetto → Sede

Ossia un impiegato deve operare su una sola
sede e anche i progetti devono insistere su
una sola sede

NORMALIZZAZIONE



... non conserva le DF(3/4)

19

Aggiunta della coppia { Neri, Marte } ad
(Impiegato, Progetto)

Impiegato	Sede
Rossi	Roma
Verdi	Milano
Neri	Milano

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

NORMALIZZAZIONE



... non conserva le DF(4/4)

20

Proviamo a ricostruire

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

La dipendenza

Progetto → Sede non è preservata

NORMALIZZAZIONE



Conservazione delle dipendenze

Una istanza legale nello schema decomposto genera sullo schema ricostruito una soluzione non ammissibile

Ogni singola istanza è ("localmente") legale, ma il DB ("globalmente") non lo è

- Infatti il progetto "Marte" risulta essere assegnato a due sedi, in violazione del vincolo:

Progetto → Sede

Problemi di consistenza dei dati si hanno quando la decomposizione "separa" gli attributi di una DF. Per verificare che la DF sia rispettata si rende necessario far riferimento a entrambe le relazioni.

Normalizzazione



Le DF di Progettazioni

Impiegato	Stipendio	Progetto	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

DF1: Impiegato Progetto → Funzione

DF2: Impiegato → Stipendio

DF3: Progetto → Bilancio

Stipendio dipende da parte della chiave Impiegato Progetto
Bilancio dipende da parte della chiave Impiegato Progetto

Normalizzazione



Impiegato Progetto → Funzione

La DF riporta che in ogni progetto ogni impiegato svolge una ed una sola funzione.

Si noti che, essendo (Impiegato, Progetto) chiave di Progettazione

- non esistono due tuple con stessa chiave e Funzione (non esistono ridondanze)
- La modifica interviene su una sola tupla
- La cancellazione non comporta perdita di informazioni che potrebbero essere ancora utili
- L'inserimento è possibile anche attribuendo valori nulli agli attributi diversi da Impiegato e Progetto

Normalizzazione



Impiegato → Stipendio

La DF riporta che lo stipendio di ciascun impiegato dipende solo dall'impiegato stesso indipendentemente dai progetti a cui partecipa e quindi presenta

- Ridondanza
- Anomalia di aggiornamento
- Anomalia di cancellazione
- Anomalia di inserimento

Normalizzazione



Progetto → Bilancio

25

La DF riporta che il bilancio di ciascun progetto dipende solo dal progetto stesso indipendentemente dalla partecipazione degli impiegati ai progetti e quindi presenta

- Ridondanza
- Anomalia di aggiornamento
- Anomalia di cancellazione
- Anomalia di inserimento

NORMALIZZAZIONE



DF ed anomalie

26

Le anomalie viste si riconducono alla presenza delle DF:

- Impiegato → Stipendio
- Progetto → Bilancio

Viceversa la DF

- Impiegato, Progetto → Funzione

non causa problemi

Le anomalie sono causate dalla presenza di concetti eterogenei:

- proprietà degli impiegati (lo stipendio)
- proprietà dei progetti (il bilancio)
- proprietà della chiave **Impiegato Progetto**

NORMALIZZAZIONE



La causa delle anomalie

27

Gli attributi non primi **Stipendio** e **Bilancio** non dipendono completamente dalla chiave **Impiegato Progetto** e quindi causano anomalie.

La terza DF è completa cioè l'attributo non primo **Funzione** dipende completamente dalla chiave **Impiegato Progetto** e non causa anomalie.

Impiegato → Stipendio

- **Impiegato** è parte di una chiave

Progetto → Bilancio

- **Progetto** è parte di una chiave

Impiegato Progetto → Funzione

- **Impiegato Progetto** è una chiave

NORMALIZZAZIONE



Un esempio di decomposizione

28

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

NORMALIZZAZIONE



Proprietà della decomposizione

29

E' facile mostrare che la decomposizione della precedente slide :

- E' senza perdita
- Mantiene le dipendenze funzionali

Pertanto è una decomposizione legale.

Normalizzazione



Dipendenza transitiva

30

Dato uno schema di relazione $R(X)$ un attributo A dipende transitivamente dall'insieme di attributi Y se esiste un altro insieme di attributi Z tale che:

- $Y \rightarrow Z$ e $Z \not\rightarrow Y$
- $Z \rightarrow A$ e $A \not\rightarrow Z$
- $A \notin Y \cup Z$

Normalizzazione



Terza forma normale

31

Uno schema di relazione $R(X)$ è in 3NF se:

- è in 2NF

e

- Ogni attributo non primo di $R(X)$ non dipende in modo transitivo dalla chiave di R - cioè dipende solo dalla chiave.

cioè

- Non vi sono dipendenze funzionali tra attributi non primi

Normalizzazione



3NF : definizione alternativa

32

Uno schema di relazione $R(X)$ è in 3NF se per ogni DF non banale $Y \rightarrow Z$ definita su $R(X)$:

- Y è una chiave di $R(X)$ (1)

o

- Z è un attributo primo (2)

Per la (1) ogni dipendenza funzionale è completa.
Per la (2) se Y è un attributo non primo per la precedente definizione di 3NF Z deve essere un attributo primo.

Normalizzazione



Normalizzazione in 3F: Esempio

Consideriamo la relazione in 2NF

IMPIEGATI

Codice	Nome	Reparto	Caporeparto
001B	Rossi	Produzione	Maggi
001A	Marat	Marketing	Marini
001C	Mattei	Produzione	Maggi
01AB	Nero	Marketing	Marini

Normalizzazione



Anomalie

Con riferimento alla tabella **IMPIEGATI**:

- **Aggiornamento**: se cambia il Caporeparto della Produzione devo modificare più volte la tabella.
- **Cancellazione**: se cancello il Caporeparto Maggi cancellerò tutti gli impiegati del reparto Produzione al quale Maggi appartiene
- **Inserimento**: non posso inserire un Caporeparto se non esiste almeno un impiegato nel reparto

Normalizzazione



Decomposizione in 3NF

ALLOCAIMPIEGATI

Codice	Nome	Reparto
001B	Rossi	Produzione
001A	Marat	Marketing
001C	Mattei	Produzione
01AB	Nero	Marketing

REPARTI

Reparto	Caporeparto
Produzione	MAGGI
Marketing	MARINI

Normalizzazione



Ancora anomalie

Lo schema

TEL(Prefisso, Numero, Località, Abbonato, Indirizzo) ha vincoli

- **Prefisso, Numero** → **Località, Abbonato, Indirizzo**
- **Località** → **Prefisso**

Lo schema è in 3NF, in quanto **Prefisso** è primo (non c'è dipendenza transitiva)

Nella seguente istanza legale l'informazione sul prefisso viene replicata per ogni abbonato:

Prefisso	Numero	Località	Abbonato	Indirizzo
051	457856	Bologna	Rossi	Via Roma 8
059	452332	Modena	Verdi	Via Bari 16
051	987856	Bologna	Bianchi	Via Napoli 77
051	552346	Castenaso	Neri	Piazza Borsa 12
059	387654	Vignola	Mori	Via Piave 65

Normalizzazione



Anomalie in una relazione in 3NF

37

Se devo inserire un nuovo prefisso debbo conoscere la Località.

Allora può essere utile non ammettere dei determinanti del tipo:

$$Y \rightarrow Z$$

con Y attributo non primo e Z attributo primo

In questo caso si introduce una nuova forma normale detta di Boice e Codd.

NORMALIZZAZIONE



Forma Normale di Boyce e Codd

38

Uno schema di relazione R(X) è in forma normale di Boyce e Codd se:

- per ogni dipendenza funzionale (non banale)

$$Y \rightarrow Z$$

definita su R(X), Y contiene una chiave K di R(X) cioè Y è una superchiave di R(X).

NORMALIZZAZIONE



Decomposizione in BCNF

39

Una soluzione consiste nel decomporre lo schema in

- NUM_TEL(Numero,Località,Abbonato,Indirizzo)
- PREF_TEL(Località, Prefisso)

La decomposizione è lossless perché

- $(NUM_TEL \bowtie PREF_TEL) = TEL$

NUM_TEL

Numero	Località	Abbonato	Indirizzo
457856	Bologna	Rossi	Via Roma 8
452332	Modena	Verdi	Via Bari 16
987856	Bologna	Bianchi	Via Napoli 77
552346	Castenaso	Neri	Piazza Borsa 12
387654	Vignola	Mori	Via Piave 65

PREF_TEL

Prefisso	Località
051	Bologna
059	Modena
051	Castenaso
059	Vignola

NORMALIZZAZIONE



Relazione tra BCNR e 3NF

40

- Se uno schema di relazione R(X) è in BCNF allora è in 3NF.
- Se uno schema di relazione R(X) è in 3NF non è detto che sia in BCNF.
- Uno schema di relazione R(X) in BCNF è anche in 3NF se e solo se ogni chiave di R(X) è costituita da un solo attributo.

- Infatti per tutte le DF del tipo: $Y \rightarrow Z$ con Z attributo primo si ha che Z deve essere una chiave e quindi vale anche $Z \rightarrow Y$ e quindi Y deve anch'esso essere una chiave per R(x).

NORMALIZZAZIONE



Progettazione e normalizzazione

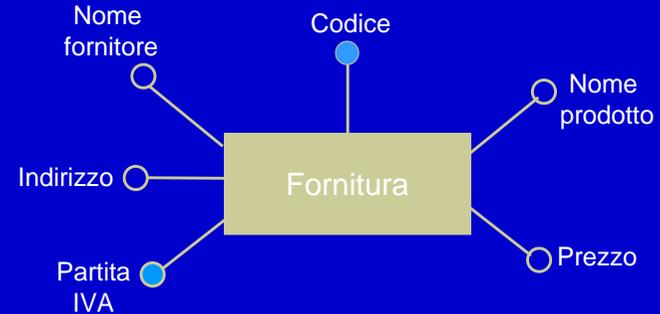
Normalizzazione

La teoria della normalizzazione può essere usata anche durante la progettazione concettuale per verificare la qualità dello schema concettuale stesso



Es.: entità non normalizzata

Normalizzazione



Chiave: PartitaIVA, Codice
DF1: PartitaIVA → NomeFornitore, Indirizzo
DF2: Codice → NomeProdotto, Prezzo



Analisi dell'entità

Normalizzazione

L'entità viola la seconda forma normale a causa delle dipendenze parziali dalla chiave:

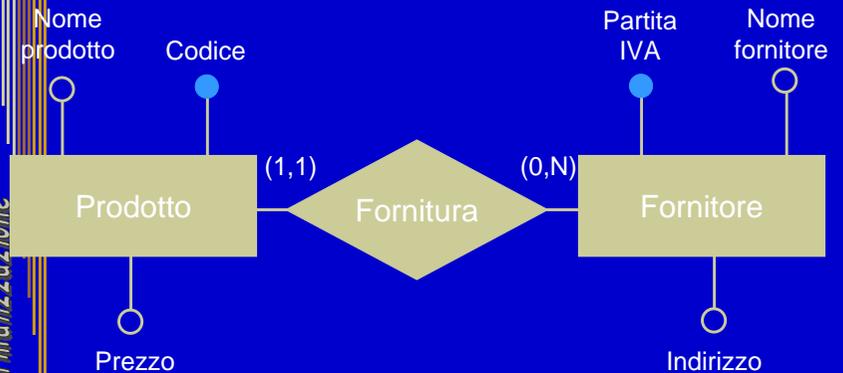
- PartitaIVA → NomeFornitore Indirizzo
- Codice → NomeProdotto Prezzo

Possiamo decomporre sulla base di queste dipendenze

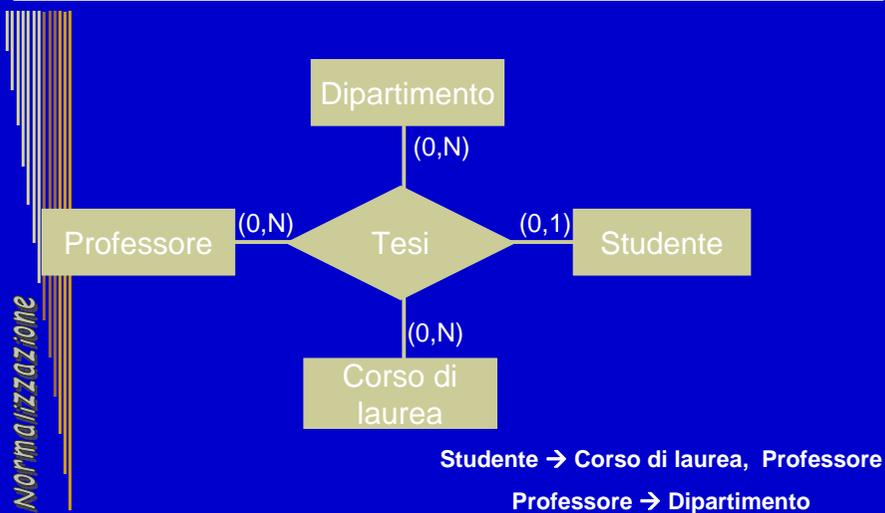


Decomposizione

Normalizzazione



Es.: associazione non normalizzata



Analisi dell'associazione

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

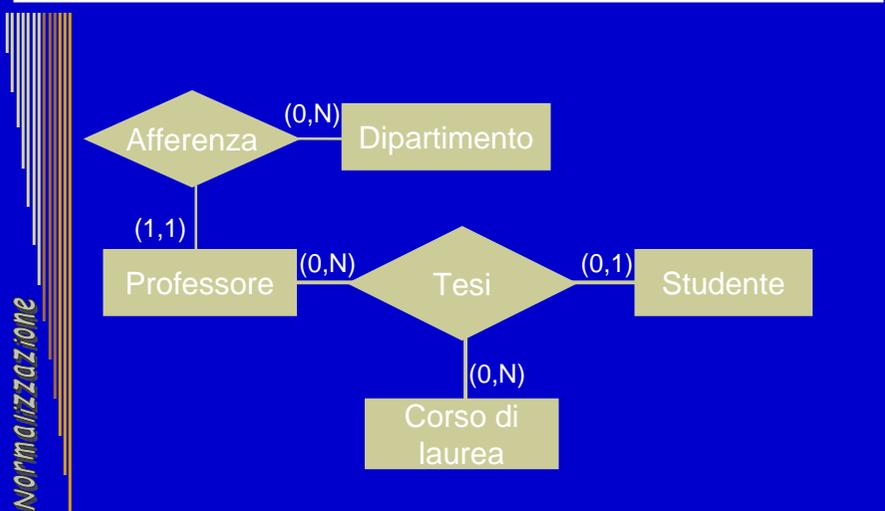
- Professore → Dipartimento

Possiamo decomporre sulla base di questa dipendenza

Normalizzazione



Decomposizione



Ulteriore analisi sulla base delle DF

L'associazione Tesi in BCNF sulla base delle dipendenze

- Studente → CorsoDiLaurea
- Studente → Professore

Le due proprietà sono indipendenti

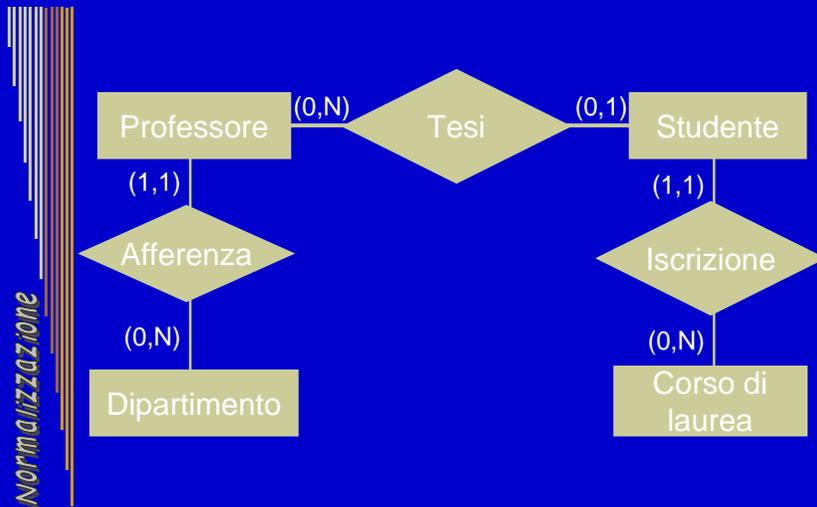
Questo suggerisce una ulteriore decomposizione

Normalizzazione



Ulteriore decomposizione

49



Normalizzazione vs. performance

50

Potremmo voler utilizzare schemi non normalizzati per aumentare le performance
Ad es. collegare e mostrare informazioni memorizzate in due tabelle differenti richiede il join delle tabelle

Vertical text on the left: *Normalizzazione*

Normalizzazione vs. performance

51

- **Alternativa 1: usare schemi denormalizzati** che contengono gli attributi di entrambe le relazioni
 - accesso più veloce
 - spazio e tempo di esecuzione superiore per gestire le modifiche
 - maggiore sforzo di programmazione per gestire la ridondanza, con conseguente maggiore incidenza degli errori di programmazione
 - **Alternativa 2: usare una vista materializzata**
 - stessi vantaggi e svantaggi della alternativa 1, eccetto il maggiore sforzo di programmazione
- Vertical text on the left: *Normalizzazione*

Riferimenti

52

- Chianese, Moscato, Picariello, Sansone - *Basi di Dati per la gestione dell'informazione* - McGraw-Hill, 2007
 - Atzeni, Ceri, Paraboschi, Torlone - *Basi di dati* - McGraw-Hill, 2006
 - Atzeni, Batini, De Antonellis - *Teoria relazionale dei dati* - Boringhieri, Torino, 1985
- Vertical text on the left: *Normalizzazione*