

La normalizzazione

La teoria della normalizzazione

- Una forma normale è una proprietà di una base dati relazionale che ne garantisce la “qualità”
 - assenza di difetti
- Quando una relazione non è normalizzata
 - presenta ridondanze
 - creando non poche difficoltà in operazioni di aggiornamento

La teoria della normalizzazione

- Le forme normali sono state definite sul modello relazionale quando non esistevano le metodologie di progettazione
 - Non tengono quindi conto del processo di progettazione
- L'applicazione delle metodologie produce infatti schemi già in forma normale
 - anche se nei complessi processi di progetti reali possono prodursi situazioni che richiedono una verifica a posteriori della qualità dello schema prodotto
- Non devono sostituirsi alle metodologie, ma possono essere applicate sul modello E-R finale per l'analisi della qualità del processo di progettazione

La normalizzazione: cosa è

- Procedura che permette di trasformare tabelle non normalizzate in tabelle che soddisfano le forme normali (almeno tre)
- La normalizzazione va utilizzata come tecnica di verifica dei risultati della progettazione di una base dati
- Non è una metodologia di progettazione

Un esempio

- La relazione:

**Progettazione(Impiegato, Stipendio,
Progetto, Bilancio, Funzione)**

- In cui:
 - Lo stipendio di ciascun impiegato è unico e gli è legato indipendentemente dai progetti a cui partecipa
 - Il bilancio di ciascun progetto è unico e dipende dal solo progetto indipendentemente dagli impiegati che vi partecipano

Una relazione con anomalie

PROGETTAZIONE

<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

- Ad esempio lo stipendio di ciascun impiegato è ripetuto
 - **Ridondanza**: presenza di dati ripetuti in diverse tuple, senza aggiunta di informazioni significative
- Se lo stipendio di un impiegato varia, è necessario andarne a modificare il valore in tutte le sue tuple
 - **anomalia di aggiornamento**: necessità di estendere l'aggiornamento di un dato a tutte le tuple in cui esso compare con inutili perdite di tempo
- Se un impiegato interrompe la partecipazione a tutti i progetti, dobbiamo cancellarlo
 - **anomalia di cancellazione**: l'eliminazione di una tupla può comportare l'eliminazione di dati importanti
- Un nuovo impiegato senza progetto non può essere inserito
 - **anomalia di inserimento**: l'inserimento di informazioni relative a uno solo dei concetti di pertinenza di una relazione è impossibile se non esiste un intero insieme di concetti in grado di costituire una tupla completa

Anomalia di aggiornamento

Un'altra relazione:

carriera(Studente, Corso, Fascia, Docente, Voto)

Se la fascia di reddito di uno studente varia, bisogna cabiare tutte le tuple relative allo stesso studente

Studente	Corso	Fascia	Docente	Voto
Moscato	Basi di Dati	I	Chianese	30
Moscato	Informatica	I	Picariello	30
Penta	Basi di Dati	IV	Chianese	18
Capasso	Sistemi Informativi	III	Sansone	18

Anomalia di cancellazione

Se viene cancellata la seconda tupla, ovvero l'unica con riferimento all'esame di Informatica, si perdono informazioni circa il docente del corso

Studente	Corso	Fascia	Docente	Voto
Moscato	Basi di Dati	I	Chianese	30
Moscato	Informatica	I	Picariello	30
Penta	Basi di Dati	IV	Chianese	18
Capasso	Sistemi Informativi	III	Sansone	18

Anomalia di inserimento

Non è possibile inserire un nuovo studente che non ha sostenuto alcun esame

Studente	Corso	Fascia	Docente	Voto
Moscato	Basi di Dati	I	Chianese	30
Moscato	Informatica	I	Picariello	30
Penta	Basi di Dati	IV	Chianese	18
Capasso	Sistemi Informativi	III	Sansone	18

La causa delle anomalie

- Uso di una sola tabella 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
- La fusione di concetti disomogenei in una unica tabella comporta:
 - Ridondanza
 - Anomalie di aggiornamento
 - Anomalie di cancellazione
 - Anomalie di inserimento

Le dipendenze funzionali

Definizione Preliminare

- Sia R una relazione con chiave primaria K
- Ogni attributo A appartenente allo schema R si dice
 - **primo** se fa parte di K
 - **non primo** se non appartiene a K

Persona(nome, cognome, datanascita, via, CAP, comune, provincia)

- nome, cognome, datanascita sono primi
- via, CAP, comune, provincia sono non primi

Dipendenze funzionali

- Per scoprire e rimuovere le anomalie in uno schema del modello logico si devono innanzitutto individuare legami di tipo funzionale tra gli attributi di una relazione
- Lo strumento è la dipendenza funzionale
 - fissa vincoli di integrità nello schema di relazione $R(X)$ consistenti proprio nei legami funzionali esistenti tra gli attributi
- La notazione per indicare una dipendenza funzionale da Y a Z , con Y e Z attributi di una relazione $R(X)$, con $Y \cup Z$ contenuto in X , è

$$Y \rightarrow Z$$

Si legge

- Y determina Z
- Z dipende funzionalmente da Y

Dipendenza funzionale

- Consideriamo
 - uno schema di relazione $R(X)$
 - due sottoinsiemi non vuoti Y e Z di X
- Diremo che in $R(X)$ esiste una dipendenza funzionale (DF) da Y a Z
 $Y \rightarrow Z$
sse:
 $\forall t_1, t_2 \in r : t_1[Y] = t_2[Y] \Rightarrow t_1[Z] = t_2[Z]$

Dipendenze Funzionali

PROGETTAZIONI

<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

Esempi

- Ogni impiegato ha un solo stipendio (anche se partecipa a più progetti)
 - Lo stipendio “dipende” dall’impiegato
 - **Impiegato → Stipendio**
- Ogni progetto ha un bilancio
 - Il bilancio “dipende” dal progetto
 - **Progetto → Bilancio**
- Ogni impiegato in ciascun progetto ha una sola funzione (anche se può avere funzioni diverse in progetti diversi)
 - La funzione “dipende” dal progetto e dall’impiegato
 - **Impiegato, Progetto → Funzione**

Tipi di dipendenze funzionali

- $Y \rightarrow Z$ non implica (ovviamente) $Z \rightarrow Y$
- DF banale
 - una DF sempre soddisfatta
 - Impiegato Progetto \rightarrow Progetto
 - $Z \subseteq Y$
 - Quando tutti gli attributi Z o un suo sottoinsieme sono un sottoinsieme di Y
- DF piena (o anche completa):
 - $Y \rightarrow Z$ è una dipendenza funzionale piena sse:
rimuovendo un qualsiasi attributo da Y essa non vale più.
 - Impiegato Progetto \rightarrow Stipendio (non è piena)
 - Impiegato \rightarrow Stipendio (continua a essere una DF)
 - sono piene tutte quelle in cui Y è un solo attributo
- DF si esprime in modo minimale quando:
 - $Y \rightarrow A, B \Rightarrow Y \rightarrow A$ e $Y \rightarrow B$
- Le DF che interessano sono quelle non banali, piene ed espresse in maniera minimale.

Sulle dipendenze funzionali banali

- Si osservi la DF:

Progetto \rightarrow Progetto Bilancio

è ovvio che un Progetto determini se stesso.

- Una DF banale si può ricondurre ad una DF non banale eliminando gli attributi del secondo membro che compaiono anche al primo

- Infatti si può dimostrare che se vale

$$X \rightarrow Z$$

allora vale anche

$$X \rightarrow W$$

con W sottoinsieme di Z che non contiene attributi di X

Vincoli e dipendenze funzionali

- Una dipendenza funzionale è
 - una caratteristica dello schema
 - non della particolare istanza dello schema
 - è dettata dalla semantica degli attributi di una relazione e non può essere inferita da una particolare istanza dello schema
 - deve essere definita esplicitamente da chi conosce la semantica degli attributi
- Poiché la DF è un vincolo, una relazione è corretta quando soddisfa la DF

Esempi di indipendenza dalle istanze

<u>Articolo</u>	<u>Magazzino</u>	Quantità
maglie	NA3	4000
scarpe	RM1	2500
pantaloni	NA1	3000

Magazzini con
articoli diversi

Realtà che non presenta ridondanze

Magazzini con gli
stessi articoli

<u>Articolo</u>	<u>Magazzino</u>	Quantità
scarpe	NA1	2500
scarpe	RM1	2500
pantaloni	NA1	3000

Realtà che invece presenta ridondanze

Vincoli di chiave e DF

- Se K è una superchiave in uno schema $R(X)$ allora ogni attributo A di $R(X)$ non contenuto in K dipende funzionalmente da K
- Se K è superchiave di $R(X)$, dalla definizione di superchiave si ha che
$$t_1[K] = t_2[K] \Rightarrow t_1 = t_2, \text{ e quindi } t_1[A] = t_2[A]$$
- Quindi il vincolo di DF generalizza il vincolo di chiave
$$K \rightarrow X$$

O equivalentemente

- Il vincolo di dipendenza funzionale **generalizza** il vincolo di chiave:
 - Se esiste la dipendenza funzionale $Y \rightarrow Z$ definita su $R(X)$
 - e $X = Y \cup Z$
ossia Y e Z non hanno attributi in comune
 - allora Y è super chiave (o chiave) per $R(X)$
 - cioè la DF degenera nel vincolo di chiave

Esempio

<u>Articolo</u>	<u>Magazzino</u>	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

- Articolo, Magazzino → Quantità, Indirizzo

Dipendenza funzionale transitiva

- Dato uno schema di relazione $R(X)$
 - Si ha che 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 $A \not\rightarrow Z$
 - $A \notin Y \cup Z$
- Quindi
 - se $Y \rightarrow Z$ e $Z \rightarrow A$ allora $Y \rightarrow A$

Esempio

- **Dato lo schema di relazione**

AUTOMOBILI (TARGA, MARCA, DATA, COLORE, MODELLO, CILINDRATA)

L'attributo (**non primo**) **CILINDRATA**, dipende transitivamente da **TARGA** in quanto:

MODELLO dipende da **TARGA**
CILINDRATA dipende da **MODELLO**

In conclusione le FD cosa indicano?

- **Le FD indicano concetti autonomi della realtà che stiamo analizzando**
- Ossia le entità del modello ER
- Per questo sono proprietà dello schema e non delle istanze

Le forme normali (NF)

Il processo di normalizzazione

- Fu proposto da Codd (inventore del modello relazionale) per:
 - sottoporre uno schema di relazione a una serie di test
 - capaci di *certificare* il soddisfacimento di una data **forma normale**.
- Esistono:
 - Prima forma normale (1NF)
 - Seconda forma normale (2NF)
 - Terza forma normale (3NF)
 - Forma normale di Boyce e Codd (BCNF)
 - 4FN
 - 5FN

Processo di analisi

- La normalizzazione è un processo di analisi degli schemi di relazione, basato sulle loro dipendenze funzionali e chiavi primarie, per ottenere le proprietà desiderate di
 - Minimizzazione della ridondanza
 - Minimizzazione delle anomalie di inserimento, cancellazione e modifica
- Obiettivo:
 - Gli schemi di relazione che non soddisfano le NF
 - devono essere decomposti in schemi di relazione più piccoli che possiedono le proprietà desiderate.

Prima Forma Normale

- **Una schema di relazione $R(X)$ è in 1NF se i suoi attributi X sono tutti atomici**
- Ogni dominio degli attributi X deve comprendere solo valori atomici. Vanno cioè evitati:
 - attributi multivalore
 - attributi composti
- Nel modello relazionale una relazione $R(X)$ deve essere per definizione in 1NF

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

Riduzione in prima forma normale

- Una tabella non è una relazione se possiede attributi multivalore o strutturati
- Per ridurre la tabella in prima forma normale occorre:
 - Sviluppare gli attributi multivalore
 - Esplodere i singoli attributi della struttura

Dipendente(Impiegato, Progetto, Stipendio, Via, Citta, Nc, Funzione)

Seconda forma normale(2NF)

- **Uno schema di relazione $R(X)$ è in 2NF se:**
 - **è in 1NF**
 - **ogni DF del tipo $Y \rightarrow Z$ ha attributi non primi Z dipendenti funzionalmente in maniera piena da ogni chiave di $R(X)$**
- Se la chiave di $R(X)$ ha un solo attributo allora $R(X)$ è già in seconda forma normale
- Negli altri casi per verificare il soddisfacimento della 2NF
 - Basta esaminare se le parti sinistre delle DF contengono attributi primi e non sono superchiavi ma chiavi

Tabella non in 2NF

<u>Articolo</u>	<u>Magazzino</u>	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

- Articolo, Magazzino → Quantità
- Magazzino → Indirizzo
 - Indirizzo dipende solo dall'attributo primo Magazzino che è parte della chiave
 - **Quantità dipende invece da una superchiave**

Decomposizione: modalità

- Uno schema di relazione $R(X)$ non in 2NF si normalizza decomponendo le relazioni di partenza in relazioni che soddisfano la 2NF
- Ogni relazione in 2NF prodotta deve avere
 - gli attributi non primi associati solo alla parte della chiave primaria da cui sono funzionalmente dipendenti in modo pieno
 - La decomposizione è senza perdita se tra gli **attributi comuni** ne esiste qualcuno che è chiave per almeno una delle relazioni decomposte

Decomposizione senza perdita

<u>Articolo</u>	<u>Magazzino</u>	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

<u>Articolo</u>	<u>Magazzino</u>	Quantità
scarpe	NA1	2500
scarpe	RM1	4500
pantaloni	NA1	3000

<u>Magazzino</u>	Indirizzo
NA1	v. Leopardi 17, Napoli
RM1	v. S. Maria Maggiore 3,Napoli

La decomposizione è senza perdita perché è stata effettuata con l'attributo comune Magazzino che risulta chiave della relazione (Magazzino, Indirizzo)

Proprietà delle decomposizioni

- Una decomposizione deve essere senza perdita informativa
- Come si verifica
 - Ricomponendo le relazioni ottenute dalla decomposizione
 - ottenendo la relazione iniziale

Decomposizione senza perdita

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

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

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



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

Decomposizione con perdita informativa

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

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

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

Tuple
spurie

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

Conservazione delle dipendenze

- Una decomposizione preserva le dipendenze se non genera relazioni che separano gli attributi delle dipendenze funzionali
 - Nell'esempio Magazzini le dipendenze
Articolo, Magazzino -> Quantità
Magazzino-> Indirizzo
vengono mantenute

Esempio in cui DF non sono preservate

- Un impiegato deve operare su una sola sede
 - Impiegato → Sede
- Un progetto deve insistere su una sola sede
 - Progetto → Sede

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



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

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

Gli attributi di Progetto → Sede
sono stati separati

Implicazioni

- Aggiungiamo {Neri, Marte} ad (Impiegato, Progetto)

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

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

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

Progetto → Sede
non è preservata

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

Terza forma normale

- 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 da ogni chiave di R
 - cioè dipende
 - dalla chiave (1NF) e
 - da tutta la chiave (2NF) e
 - solo dalla chiave (3NF)

Terza forma normale (altra definizione)

Dalla definizione precedente discende:

- Uno schema di relazione $R(X)$ è in 3NF se:
 - è in seconda forma normale
 - per ogni FD non banale $Y \rightarrow Z$ definita su $R(X)$
 - o Y è superchiave di $R(X)$
 - quindi la dipendenza è non transitiva e completa
 - o Y è un attributo non primo per cui Z deve essere un attributo primo
 - Infatti se Z fosse anch'esso non primo, avremmo una FD transitiva (per la 2FD) violando la 3NF, quindi Z deve necessariamente essere primo

3NF più semplice da ricordare

- Uno schema di relazione $R(X)$ è in 3NF se:
 - è in seconda forma normale
 - per tutte le FD non banali $Y \rightarrow Z$ definite su $R(X)$
 - Y è una chiave di R
 - o
 - Z è parte di una chiave di R

Terza forma normale: esempio

Relazione in seconda forma normale ma non in 3NF

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

Per la definizione di chiave

Codice \rightarrow Nome

Codice \rightarrow Reparto

Codice \rightarrow Caporeparto

Per una FD

Reparto \rightarrow Caporeparto

Esiste FD transitiva: Codice \rightarrow Reparto e Reparto \rightarrow Caporeparto

Che genera anomalie:

- **Aggiornamento**
 - se cambia il Caporeparto della Produzione devo modificare più righe della 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

Decomposizione in 3NF

<u>Codice</u>	Nome	Reparto
001B	Rossi	Produzione
001A	Marat	Marketing
001C	Mattei	Produzione
01AB	Nero	Marketing

<u>Reparto</u>	Caporeparto
Produzione	MAGGI
Marketing	MARINI

- Decomposizione senza perdite
 - Reparto chiave delle seconda relazione
- Mantiene le dipendenze
 - Sulla prima
 - Codice → Nome
 - Codice → Reparto
 - Codice → Caporeparto
 - Sulla seconda
 - Reparto → Caporeparto

Ancora anomalie

- Prefisso, Numero → Località, Abbonato, Indirizzo
- Località → Prefisso
- Lo schema è in 3NF, in quanto Prefisso è primo.
- Nella seguente istanza legale l'informazione sul prefisso viene replicata per ogni abbonato

<u>Prefisso</u>	<u>Numero</u>	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

Anomalie in una 3NF

- Allora può essere utile
 - non ammettere dei determinanti del tipo:
 $Y \rightarrow Z$
 - con Y attributo non primo e Z attributo primo
- In conclusione se R è in 3NF qualche ridondanza è possibile
- In questo caso si introduce una nuova forma normale detta di Boice e Codd.

FN di Boyce e Codd

- Uno schema di relazione $R(X)$ è in forma normale di Boyce e Codd se:
 - è in seconda forma normale
 - per ogni dipendenza funzionale $Y \rightarrow Z$ (non banale) definita su $R(X)$
 - Y è chiave di $R(X)$
 - Ovvero Y è una superchiave di $R(X)$

FN di Boyce e Codd

- Una relazione è in forma normale di Boyce-Codd (BCNF) quando
 - è in (2FN)
 - e in essa tutti i determinanti possono essere chiavi candidate
 - cioè ogni attributo Y dal quale dipendono altri attributi Z può svolgere la funzione di chiave.
- Una relazione che soddisfa la BCNF è anche in seconda e in terza forma normale, in quanto la BCNF esclude che
 - un determinante Y possa essere composto solo da una parte della chiave,
 - come avviene per le violazioni alla 2FN
 - o che possa essere esterno alla chiave
 - come avviene per le violazioni alla 3FN

Relazione tra BCNF e 3NF

- 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
- Partendo da uno schema in 2FN
 - Esiste sicuramente una decomposizione (o una serie di decomposizioni) che sono senza perdita e conservano le dipendenze che ci portano a uno schema in 3FN
 - Non è detto che si possa arrivare ad uno schema in BCNF

Esempio

- Consideriamo una relazione che descrive l'allocazione delle sale operatorie di un ospedale
 - Le sale operatorie sono prenotate, giorno per giorno, in orari previsti, per effettuare interventi su pazienti ad opera dei chirurghi dell'ospedale.
 - Nel corso di una giornata una sala operatoria è occupata sempre dal medesimo chirurgo che effettua più interventi, in ore diverse.
 - Noti i valori di Paziente e DataIntervento, sono noti anche:
 - ora dell'intervento, chirurgo, e sala operatoria utilizzata.

Interventi

Paziente	DataIntervento	OraIntervento	Chirurgo	Sala
Bianchi	25/10/2005	8.00	De Bakey	Sala1
Rossi	25/10/2005	8.00	Romano	Sala2
Negri	26/10/2005	9.30	Veronesi	Sala1
Viola	25/10/2005	10.30	De Bakey	Sala1
Verdi	25/10/2005	11.30	Romano	Sala2

Le FD dell'esempio

- Nella relazione Interventi valgono le dipendenze funzionali:
 - a) $\{\text{Paziente, DataIntervento}\} \rightarrow \text{OrtoIntervento, Chirurgo, Sala}$
 - b) $\{\text{Chirurgo, DataIntervento, OrtoIntervento}\} \rightarrow \text{Paziente, Sala}$
 - c) $\{\text{Sala, DataIntervento, OrtoIntervento}\} \rightarrow \text{Paziente, Chirurgo}$
 - d) $\{\text{Chirurgo, DataIntervento}\} \rightarrow \text{Sala}$
- Ci sono tre insiemi di attributi che possono svolgere la funzione di chiave:
 - $\{\text{Paziente, DataIntervento}\},$
 - $\{\text{Chirurgo, DataIntervento, OrtoIntervento}\},$
 - $\{\text{Sala, DataIntervento, OrtoIntervento}\}$

Interventi (Paziente, DataIntervento, OrtoIntervento, Chirurgo, Sala)

Analisi

- La BCNF è soddisfatta per le FD a, b, c
 - Infatti i determinati possono svolgere la funzione di chiave
- La BCNF non è invece soddisfatta dalla FD d
 - Perché il suo determinante ha un insieme di attributi non chiave
- **Però la relazione Interventi è invece in 3FN**
 - Infatti le FD a, b, c hanno tutte determinanti che sono chiavi
 - l'attributo Sala nella DF d è un attributo che fa parte della chiave candidata {Sala, DataIntervento, OraIntervento}
 - quindi Sala è un attributo primo

Decomposizione in BCNF

- Dipendenze funzionali
 - Prefisso, Numero \rightarrow Località, Abbonato, Indirizzo
 - Località \rightarrow Prefisso

<u>Prefisso</u>	<u>Numero</u>	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

- Non è in BCF
 - Località \rightarrow Prefisso
perché Località non è chiave
ma è in 3FN in quanto prefisso è primo

Decomposizione in BCNF

- Una soluzione consiste nel decomporre lo schema in
 - NUMERO(Numero,Località,Abbonato,Indirizzo)
 - PREFISSO(Località, Prefisso)

<u>Numero</u>	<u>Località</u>	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

Prefisso	<u>Località</u>
051	Bologna
059	Modena
051	Castenaso
059	Vignola

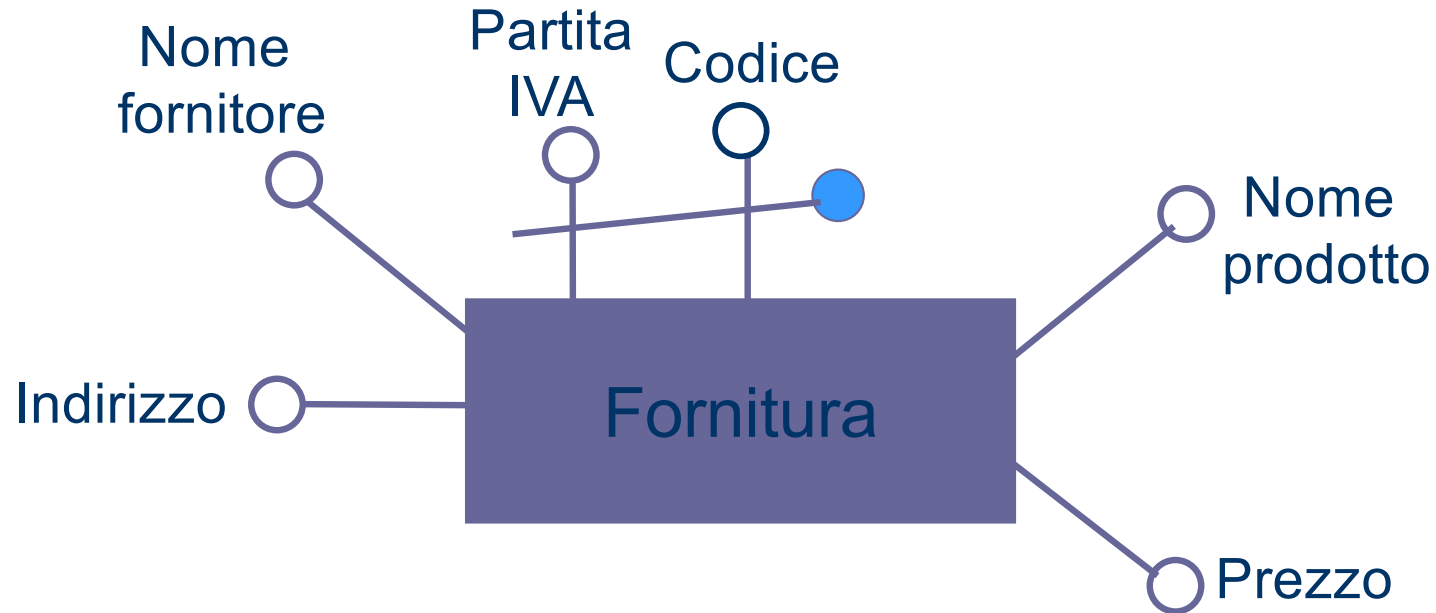
Normalizzazione vs. performance

- Potremmo voler utilizzare schemi non normalizzati per aumentare le performance
- Infatti per collegare e mostrare informazioni memorizzate in due tabelle differenti si richiede il join delle tabelle
 - Usare schemi denormalizzati che contengono gli attributi di entrambe le relazioni
 - Accesso più veloce
 - Spazio e tempo di esecuzione superiore per gestire le modifiche

Progettazione e normalizzazione

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

Entità non normalizzata



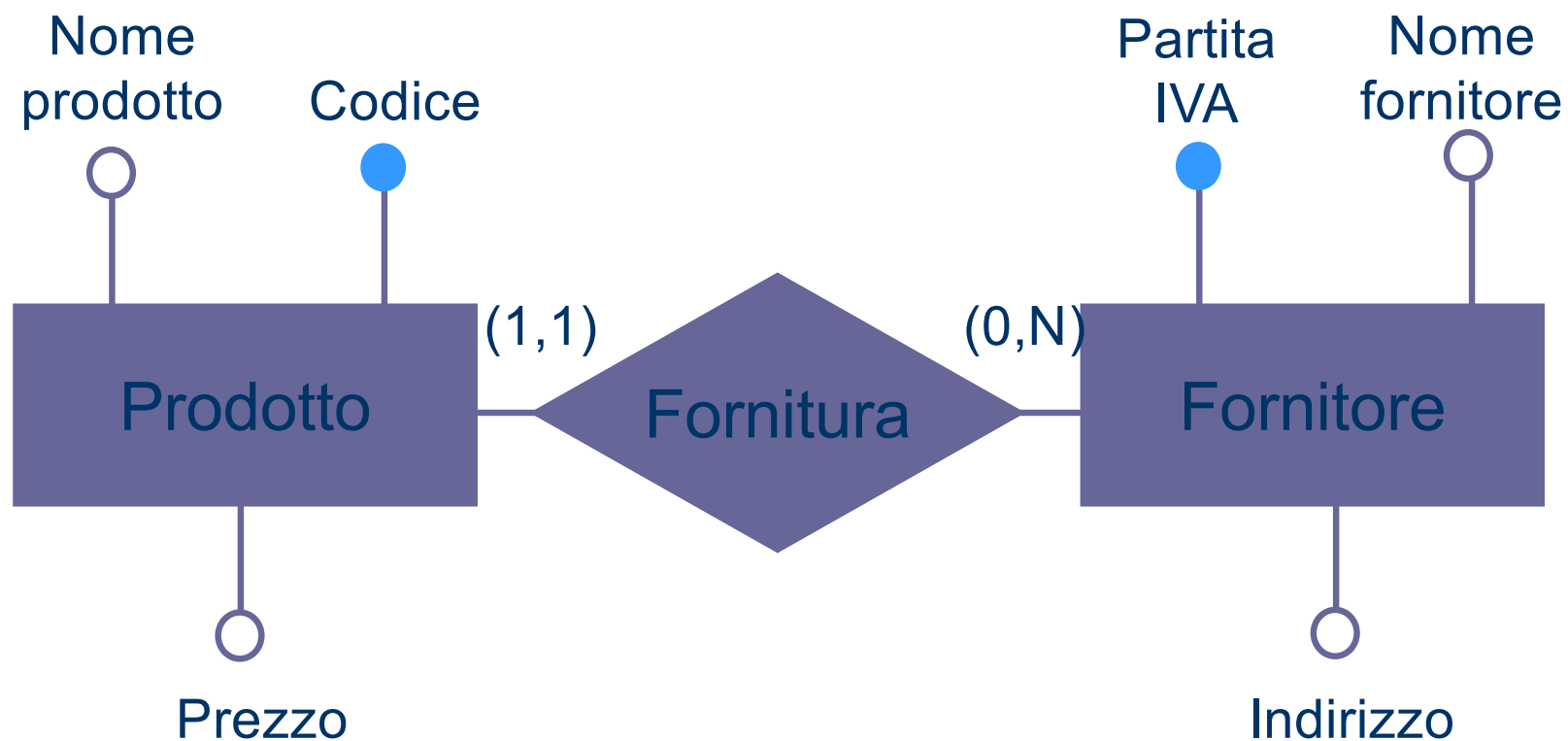
FD1: PartitaIVA \rightarrow NomeFornitore, Indirizzo

FD2: Codice \rightarrow NomeProdotto, Prezzo

Analisi dell'entità

- 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



Esercizio 1

- Siano assegnati le due FD:
 - CodiceAutore, TitoloLibro \rightarrow Prezzo
 - CodiceAutore \rightarrow CognomeAutore

e la relazione:

Libro(CodiceAutore, TitoloLibro, CognomeAutore, Prezzo)

- Dire se la relazione è in 2NF
- In caso contrario normalizzarla senza perdita

Esercizio 2

- Siano assegnati le FD:
 - Matricola \rightarrow Stipendio
 - Matricola \rightarrow NomeDipartimento
 - Matricola \rightarrow CittaDipartimento
 - NomeDipartimento \rightarrow CittaDipartimento

e la relazione:

Impiegato(Matricola, Stipendio, NomeDipartimento, CittaDipartimento)

- Dire se la relazione è in una NF
- In caso contrario normalizzarla

Esercizio 3

- Siano assegnati le FD:
 - $\text{CodiceDocente} \rightarrow \text{CognomeDocente}$
 - $\text{CodiceCorso} \rightarrow \text{NomeCorso}$
 - $\text{CodiceCorso} \rightarrow \text{Crediti}$

e la relazione:

Docenza(CodiceDocente, CodiceCorso, CognomeDocente, NomeCorso, Crediti)

- Dire se la relazione è in 2NF
- In caso contrario normalizzarla