
Elementi di Normalizzazione

Prof. Crescenzo Gallo
c.gallo@unifg.it

Introduzione

- ❑ La **normalizzazione** puo' essere vista come un processo sistematico basato sull'applicazione ripetuta dell'operatore relazionale di proiezione
- ❑ Obiettivo: eliminare le ridondanze e le anomalie

Concetto di normalizzazione

Una forma normale è una proprietà di uno schema relazionale che ne garantisce la “**qualità**”, cioè **l'assenza di determinati difetti**

Una relazione non normalizzata:

- **presenta ridondanze**
- **si presta a comportamenti poco desiderabili durante gli aggiornamenti**

Concetto di normalizzazione

Le forme normali sono di solito definite sul modello relazionale, ma hanno senso anche in altri contesti, ad esempio nel modello E/R

L'attività che permette di trasformare schemi non normalizzati in schemi che soddisfano una forma normale è detta

normalizzazione

Concetto di normalizzazione

La normalizzazione va utilizzata come **tecnica di verifica dei risultati della progettazione** di una base di dati.

Non costituisce quindi una metodologia di progettazione.

Una relazione con anomalie

In un'unica relazione sono rappresentati gli impiegati, i loro stipendi, i progetti, i bilanci dei progetti, e l'appartenenza degli impiegati ai progetti

| <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 |

Analisi della relazione

Se analizziamo il contenuto della relazione, possiamo osservare che:

1. Ogni impiegato ha un solo stipendio (anche se partecipa a più progetti)
2. Ogni progetto ha un (solo) bilancio
3. Ogni impiegato in ciascun progetto ha una sola funzione (anche se può avere funzioni diverse in progetti diversi)

Analisi della relazione

Ma abbiamo usato un'unica relazione per rappresentare tutte queste informazioni eterogenee:

- gli impiegati con i relativi stipendi
- i progetti con i relativi bilanci
- le partecipazioni degli impiegati ai progetti con le relative funzioni

Anomalie

| Impiegato | Stipendio | Progetto | Bilancio | Funzione |
|------------------|---------------|------------------|--------------|--------------------|
| Rossi | 20 | Marte | 2 | tecnico |
| Verdi | 30 | Giove | 15 | progettista |
| Verdi | 30 | Venere | 15 | progettista |
| Neri | 50 | Venere | 15 | direttore |
| Neri | 50 | Giove | 15 | consulente |
| Neri | 50 | Marte | 2 | consulente |
| Mori | 40 | Marte | 2 | direttore |
| Mori | 40 | Venere | 15 | progettista |
| Bianchi | 40 | Venere | 15 | progettista |
| Bianchi | 40 | Giove | 15 | direttore |

- Se lo stipendio di un impiegato varia, è necessario modificare il valore in diverse righe: anomalia di aggiornamento
- Lo stipendio di ciascun impiegato è ripetuto in tutte le righe relative: ridondanza
- 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

Definizione delle anomalie

Ridondanza

presenza di dati ripetuti in diverse righe, senza aggiungere informazioni significative

Anomalia di aggiornamento

necessità di estendere l'aggiornamento di un dato a tutte le righe in cui esso compare

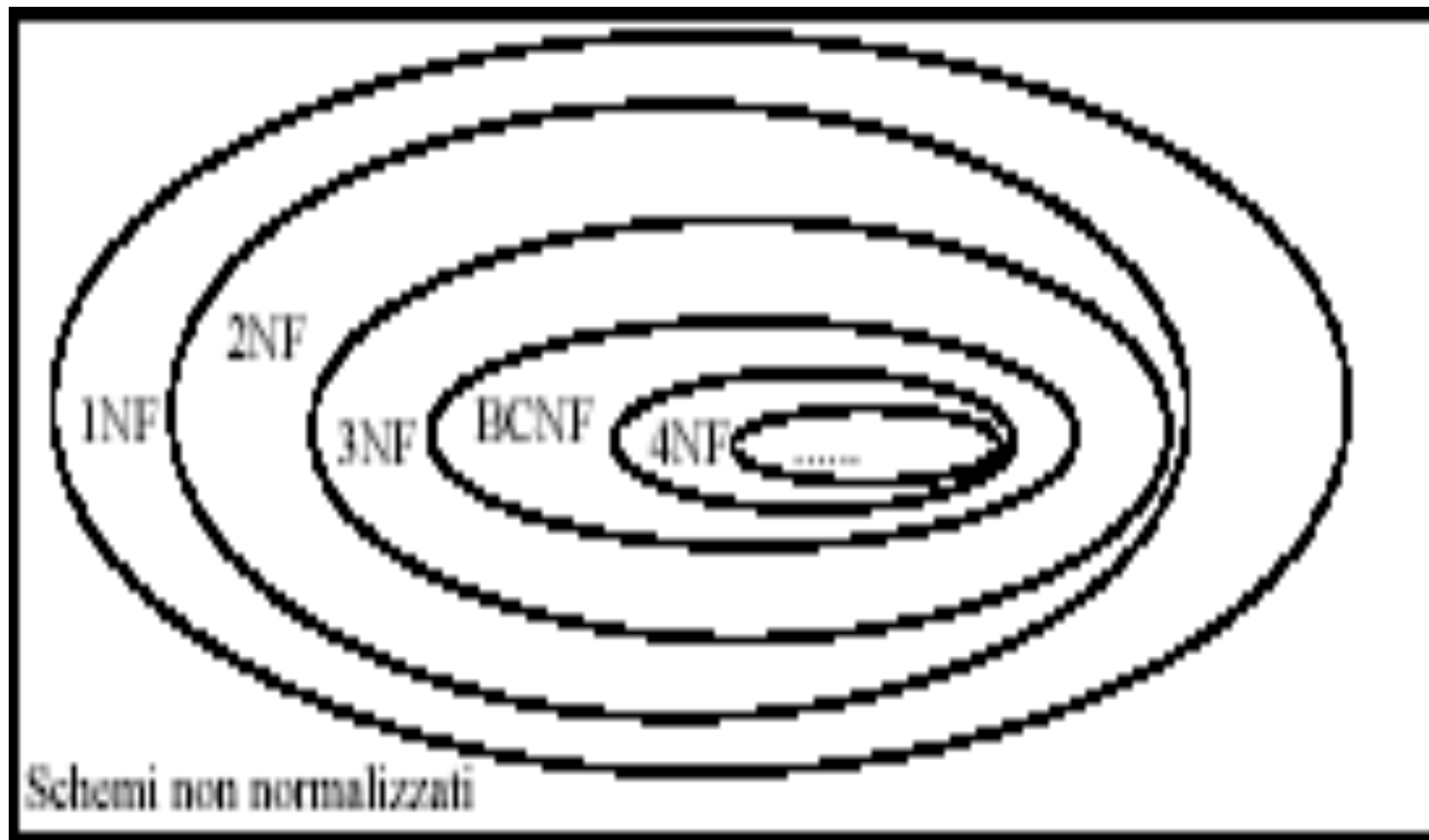
Anomalia di cancellazione

l'eliminazione di una riga motivata dal fatto che non è più valido l'insieme dei concetti in essa espressi, può comportare l'eliminazione di dati che conservano la loro validità

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 un tuple completo.

Livelli di normalizzazione



Prima Forma Normale

- Una tabella è in **Prima Forma Normale (1NF)** se e solo se ogni componente di ogni riga (cella) ha un valore atomico, non ulteriormente decomponibile

| | Attr A | Attr B | Attr C |
|---------|--------|--------|--------|
| Tuple 1 | | | |
| Tuple 2 | | X | |
| Tuple 3 | | | |
| Tuple 4 | | | |

- Tutte le relazioni normalizzate sono almeno in **1NF**

Seconda Forma Normale

➤ Uno schema $R(T)$ con vincoli funzionali è in **Seconda Forma Normale (2NF)** se e solo se ogni attributo **non-primo** dipende completamente (non parzialmente) da ogni chiave candidata dello schema,

ovvero se

non c'è dipendenza parziale di un attributo **non-primo** da una chiave.

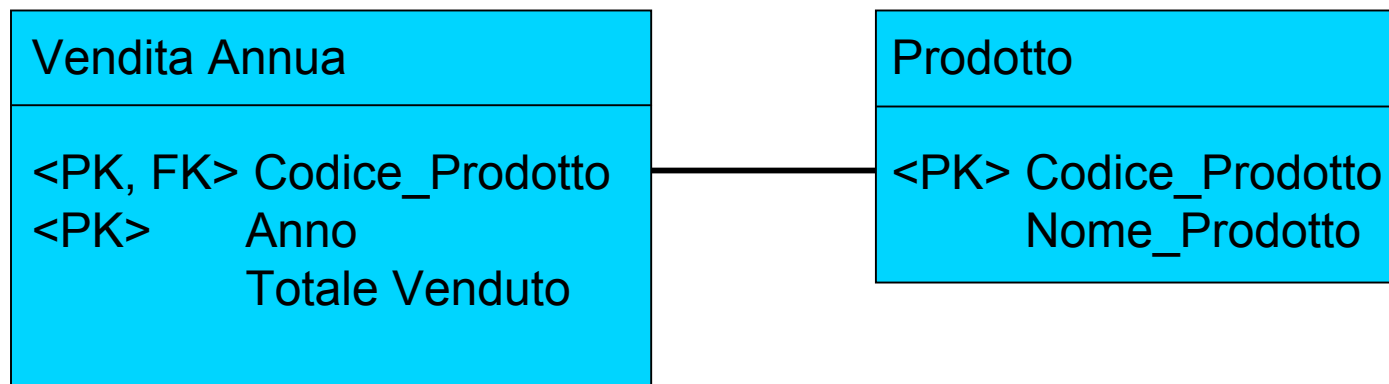
Attributo primo: dato uno schema $R(T)$, un attributo A è **primo** se e solo se fa parte di almeno una chiave dello schema. In caso contrario A è **non-primo**.

Uno schema in 1NF le cui chiavi siano tutte “semplici”, ovvero formate da un singolo attributo, è anche in 2NF.

Seconda Forma Normale

| Vendita Annua |
|----------------------|
| <PK> Codice_Prodotto |
| <PK> Anno |
| Nome_Prodotto |
| Totale Venduto |

- ❑ Una relazione e' in 2NF se e' in 1NF e se ogni suo attributo non-chiave dipende funzionalmente dalla chiave completa

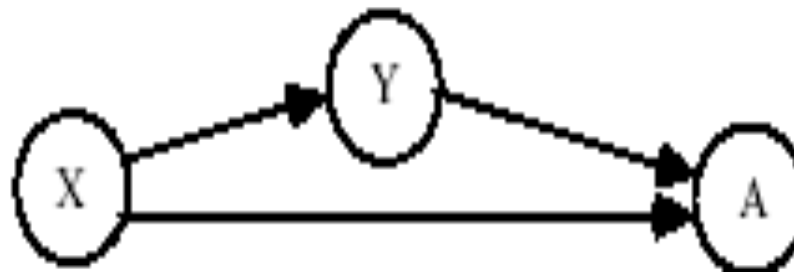


Terza Forma Normale

□ Uno schema $R(T)$ con vincoli funzionali è in **Terza Forma Normale (3NF)** se e solo se ogni attributo non-primario non dipende transitivamente da nessuna chiave

ovvero se

non c'è **dipendenza transitiva** di un attributo non-primario da una chiave



Dipendenza transitiva: data una relazione con attributi A, X e Y, e con PK A;
se X dipende funzionalmente da Y, a sua volta dipendente da A
se X non è una chiave candidata della relazione, alternativa ad A
allora X ha una **dipendenza funzionale transitiva** da A

Esempio di normalizzazione

| <u>Impiegato</u> | Stipendio | <u>Progetto</u> | Bilancio | Funzione |
|------------------|-----------|-----------------|----------|----------|
|------------------|-----------|-----------------|----------|----------|

Impiegato → Stipendio

Impiegato, Progetto → Funzione

Progetto → Bilancio

| <u>Impiegato</u> | Stipendio |
|------------------|-----------|
| Rossi | 20 |
| Verdi | 35 |
| Neri | 55 |
| Mori | 48 |
| Bianchi | 48 |

| <u>Impiegato</u> | <u>Progetto</u> | 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 |

| <u>Progetto</u> | Bilancio |
|-----------------|----------|
| Marte | 2 |
| Giove | 15 |
| Venere | 15 |

Verifica della normalizzazione

| <u>Impiegato</u> | <u>Stipendio</u> | <u>Progetto</u> | <u>Bilancio</u> | <u>Funzione</u> |
|------------------|------------------|-----------------|-----------------|-----------------|
| 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 |

| <u>Impiegato</u> | <u>Stipendio</u> |
|------------------|------------------|
| Rossi | 20 |
| Verdi | 35 |
| Neri | 55 |
| Mori | 48 |
| Bianchi | 48 |

| <u>Impiegato</u> | <u>Progetto</u> | <u>Funzione</u> |
|------------------|-----------------|-----------------|
| 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 |

| <u>Progetto</u> | <u>Bilancio</u> |
|-----------------|-----------------|
| Marte | 2 |
| Giove | 15 |
| Venere | 15 |

Osservazioni

Benché gli schemi in **3NF** non siano esenti da problemi, tale livello di normalizzazione è comunemente accettato nella pratica.

Osservazioni

Nel caso generale, problemi di complessità computazionale rendono improponibile affrontare l'attività di normalizzazione mediante tecniche di "analisi".

Tutti i seguenti problemi sono **NP-completi**:

- Determinare se un attributo è primo
- Verificare se esiste una chiave di grado minore di k
- Verificare se uno schema è in 3NF rispetto a un insieme di FD

Osservazioni

L'approccio adottato è di tipo costruttivo, ovvero *anziché verificare se uno schema è al livello di normalizzazione richiesto, si progettano schemi che siano a tale livello di normalizzazione.*

La qualità di una decomposizione (ottenibile con algoritmi di normalizzazione):

- deve essere **senza perdita**, per garantire la ricostruzione delle informazioni originarie
- dovrebbe **conservare le dipendenze**, per semplificare il mantenimento dei vincoli di integrità originari

Normalizzare o non normalizzare?

La normalizzazione non deve essere intesa come un obbligo, in quanto in alcune situazioni le anomalie che si riscontrano in schemi non normalizzati sono un male minore rispetto alla situazione che si verrebbe a creare normalizzandoli.

Normalizzare o non normalizzare?

In particolare, si ricordano le considerazioni seguenti:

- Normalizzare elimina le anomalie, ma può appesantire l'esecuzione di certe operazioni (*join* tra gli schemi normalizzati).
- La frequenza con cui i dati vengono modificati incide su qual è la scelta più opportuna (relazioni “*non-volatili*”, quasi statiche, danno meno problemi se non normalizzate).
- La ridondanza presente in relazioni non normalizzate va quantificata per capire quanto essa incida sull'occupazione di memoria e sui costi da pagare quando le repliche di una stessa informazione devono essere aggiornate.

Riassumendo....

- Una forma normale è una proprietà di uno schema relazionale che ne garantisce la “qualità”, cioè l’assenza di determinati difetti.
- Una relazione non normalizzata presenta ridondanze e dà luogo a comportamenti poco desiderabili durante gli aggiornamenti.
- La definizione delle forme normali (2NF, 3NF - e BCNF) si basa sul vincolo di dipendenza funzionale (FD).
- Normalizzare uno schema significa decomporlo in sottoschemi.
- Ogni decomposizione deve essere senza perdita, ovvero deve permettere di ricostruire esattamente la relazione originaria non decomposta.
- È anche opportuno che la decomposizione preservi le FD, al fine di evitare (o ridurre la complessità di) query di verifica che garantiscano che i vincoli siano rispettati.
- La normalizzazione deve essere reversibile.