



Normalizzazione

Forme normali

- ❑ 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.**
- ❑ 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**.
- ❑ La normalizzazione deve essere utilizzata come **tecnica di verifica dei risultati della progettazione** di una base di dati.

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 |

- ❑ In un'unica relazione sono rappresentati gli impiegati con i relativi stipendi, i progetti con i relativi bilanci e la partecipazione degli impiegati ai progetti

Analizziamo la relazione...

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

Ridondanze e anomalie

| Impiegato | Stipendio | Progetto | Bilancio | Funzione |
|-----------|-----------|----------|----------|-------------|
| Rossi | 20 | Marte | 2 | tecnico |
| Verdi | 35 | Giove | 15 | progettista |
| Verdi | 35 | Venere | 15 | progettista |
| Neri | 58 | Venere | 15 | direttore |
| Neri | 58 | Giove | 15 | consulente |
| Neri | 58 | Marte | 2 | consulente |
| Mori | 48 | Marte | 2 | direttore |
| Mori | 48 | Venere | 15 | progettista |
| Bianchi | 48 | Venere | 15 | progettista |
| Bianchi | 48 | Giove | 15 | direttore |

← ???

Gialli

- ❑ Lo stipendio di ciascun impiegato è ripetuto in tutte le tuple relative: **ridondanza**.
- ❑ Se lo stipendio di un impiegato varia, è necessario modificare il valore in diverse tuple: **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**.

Ridondanze e anomalie

- ❑ **Ridondanza**: presenza di dati ripetuti in diverse tuple senza aggiungere informazioni significative.
- ❑ **Anomalia di aggiornamento**: necessità di estendere l'aggiornamento di un dato a tutte le tuple in cui esso compare.
- ❑ **Anomalia di cancellazione**: l'eliminazione di una tupla 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 una tupla completa.

Ridondanze: una precisazione

- In una base dati l'informazione può essere duplicata in modo :

NON RIDONDANTE :

la duplicazione dei dati è **necessaria**, l'eliminazione delle duplicazioni comporta **perdita di informazione**.

STUDENTE

| Matr | Tutor |
|--------|-------|
| 125233 | Mario |
| 127988 | Carlo |
| 150444 | Carlo |
| 190787 | Mario |

← ← duplicazione
di dati **non**
ridondante

RIDONDANTE :

la duplicazione dei dati **non è necessaria**, comporta spreco di memoria, è causa di possibili **anomalie e inconsistenze**.

STUDENTE

| Matr | Tutor | Tel |
|--------|-------|------|
| 125233 | Mario | 7575 |
| 127988 | Carlo | 5566 |
| 150444 | Carlo | 5566 |
| 190787 | Mario | 7575 |

← ← duplicazione
di dati
ridondante

Scomposizione di schemi

- Le ridondanze si possono eliminare mediante scomposizione degli schemi.

STUDENTE

| N-mat | Tutor | Tel |
|--------|-------|------|
| 125233 | Mario | 7575 |
| 127988 | Carlo | 5566 |
| 150444 | Carlo | 5566 |
| 190787 | Mario | 7575 |

STUDENTE

| N-mat | Tutor |
|--------|-------|
| 125233 | Mario |
| 127988 | Carlo |
| 150444 | Carlo |
| 190787 | Mario |

TUTOR

| Tutor | Tel |
|-------|------|
| Mario | 7575 |
| Carlo | 5566 |

Dipendenza funzionale

- Per formalizzare i problemi visti si introduce un nuovo tipo di vincolo, la **dipendenza funzionale (FD)**.

Si considerino:

- un'istanza r di uno schema $R(X)$;
- due sottoinsiemi (non vuoti) di attributi Y e Z di X .
- Si dice che **in r vale la dipendenza funzionale (FD) $Y \rightarrow Z$** (Y determina funzionalmente Z) se

$$\forall t_1, t_2 \in r : t_1[Y] = t_2[Y] \rightarrow t_1[Z] = t_2[Z]$$

per ogni coppia di tuple t_1 e t_2 di r con gli stessi valori su Y , t_1 e t_2 hanno gli stessi valori anche su Z

Esempi di FD

- Nella relazione

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

 si hanno diverse FD, tra cui:
 - $\text{Impiegato} \rightarrow \text{Stipendio}$
 - $\text{Progetto} \rightarrow \text{Bilancio}$
 - $\text{Impiegato, Progetto} \rightarrow \text{Funzione}$
- Altre FD sono “meno interessanti” (“banali”), poiché sempre soddisfatte, ad esempio:
 - $\text{Impiegato, Progetto} \rightarrow \text{Progetto}$
 - Se Z è un sottoinsieme Y allora sicuramente $Y \rightarrow Z$.
 - FD di questo tipo sono dette **FD banali**.
 - $Y \rightarrow Z$ è non banale se nessun attributo in Z appartiene a Y .

FD - Precisazioni

- ❑ Una dipendenza funzionale è una caratteristica dello schema, **aspetto intensionale**, e non della particolare istanza dello schema, **aspetto estensionale**.
- ❑ Una dipendenza funzionale è **dettata dalla semantica** degli attributi di una relazione e non può essere inferita da una particolare istanza dello schema.
- ❑ Una istanza di uno schema che rispetti una data dipendenza funzionale è detta **istanza legale** dello schema rispetto alla data dipendenza funzionale.
- ❑ Se X è una **chiave** in uno schema R allora ogni altro attributo di R dipende funzionalmente da X .
- ❑ Dire che $X \rightarrow Y$ significa asserire che i valori della componente Y dipendono (sono determinati) dai valori della componente X .
- ❑ Se $X \rightarrow Y$ non necessariamente risulta anche $Y \rightarrow X$

Anomalie e FD

- ❑ Le anomalie viste si riconducono alla presenza delle FD:
 - Impiegato → Stipendio
 - Progetto → Bilancio
- ❑ viceversa la FD
 - Impiegato, Progetto → Funzionenon causa problemi.
- ❑ Motivazioni:
 - la terza FD ha sulla sinistra una chiave e non causa anomalie;
 - le prime due FD non hanno sulla sinistra una chiave e causano anomalie.
- ❑ La relazione contiene alcune informazioni legate alla chiave e altre ad attributi che non formano una chiave.

Esercizio

- Identificare le FD e le chiavi per questa relazione:

LIBRI_CLASSICI (Titolo, Autore, Genere, Copia, Scaffale)

| Titolo | Autore | Genere | Copia | Scaffale |
|-------------------------|-----------|----------|-------|----------|
| Decamerone | Boccaccio | Racconti | 1 | A75 |
| Divina Commedia | Dante | Poema | 1 | A90 |
| Divina Commedia | Dante | Poema | 2 | A90 |
| I Malavoglia | Verga | Romanzo | 1 | A90 |
| I Malavoglia | Verga | Romanzo | 2 | A75 |
| I Promessi Sposi | Manzoni | Romanzo | 1 | B10 |
| Adelchi | Manzoni | Tragedia | 1 | B20 |

N.B. Si assume che due libri non possono avere lo stesso titolo

Esercizio - soluzione

- Identificare le FD e le chiavi per questa relazione:

LIBRI_CLASSICI (Titolo, Autore, Genere, Copia, Scaffale)

| Titolo | Autore | Genere | Copia | Scaffale |
|------------------|-----------|----------|-------|----------|
| Decamerone | Boccaccio | Racconti | 1 | A75 |
| Divina Commedia | Dante | Poema | 1 | A90 |
| Divina Commedia | Dante | Poema | 2 | A90 |
| I Malavoglia | Verga | Romanzo | 1 | A90 |
| I Malavoglia | Verga | Romanzo | 2 | A75 |
| I Promessi Sposi | Manzoni | Romanzo | 1 | B10 |
| Adelchi | Manzoni | Tragedia | 1 | B20 |

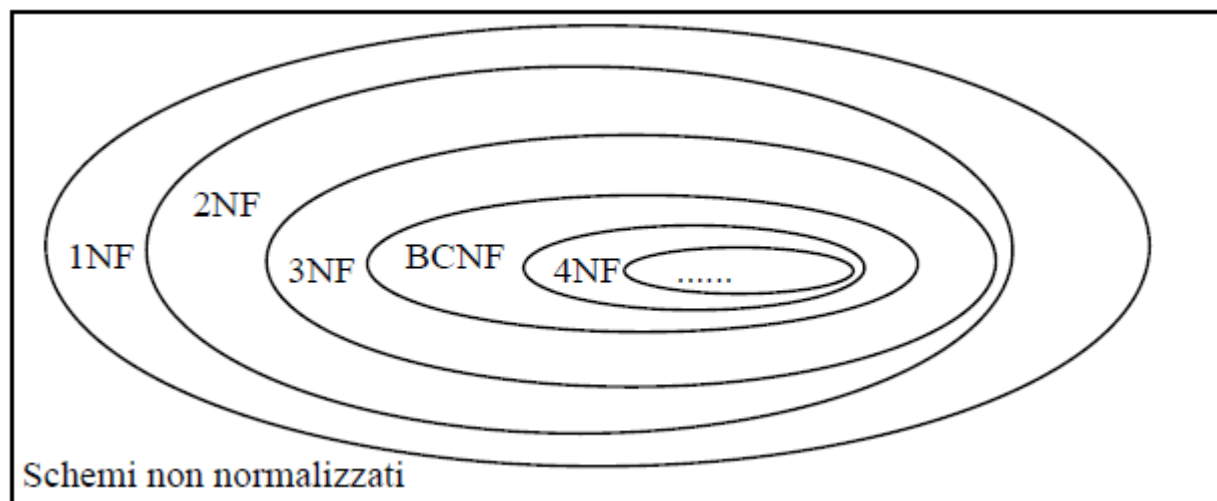
FDs:

Titolo → Genere, Autore

Titolo, Copia → Scaffale

Chiave: Titolo, Copia

Evitare le anomalie: schemi normalizzati



- ❑ Il processo di normalizzazione fu inizialmente introdotto da Codd (1972) con la definizione delle prime tre forme normali (**1NF**, **2NF**, **3NF**). In seguito Boyce e Codd definirono una forma più restrittiva di 3NF denominata **BCNF**. Tutte queste forme normali si basano sulle dipendenze funzionali tra gli attributi di una relazione.
- ❑ Più tardi furono definite altre forme normali (**4NF**, **5NF**) basate sulle dipendenze multivalore e sulle dipendenze di join.

1a Forma Normale

□ Una relazione è in 1NF se e solo se:

- tutte le tuple della relazione hanno lo **stesso numero di attributi**
- tutti i valori di un attributo sono dello **stesso tipo** (appartengono allo stesso dominio)
- **non presenta attributi che assumono valori composti** (ossia ciascun attributo è definito su un dominio con valori atomici)
 - Vedi slide su Progettazione Logica per gestire la presenza di attributi multivalore
- **esiste una chiave primaria** (ossia esiste un insieme di attributi che identifica in modo univoco ogni tupla della relazione)
- **l'ordine delle righe è irrilevante** (non è portatore di informazioni)

Si consideri lo schema **DIPARTIMENTI**(CodDip, Nome, CodDir, SediDip) e lo stato:

| <u>CodDip</u> | Nome | CodDir | SediDip |
|---------------|-----------------|--------|------------------------|
| D0001 | Amministrazione | 33301 | (Milano, Napoli, Roma) |
| D0005 | Produzione | 18007 | Aprilia |
| D0003 | Ricerca | 33010 | Napoli |

La relazione **non è in 1NF** a causa dell'attributo SediDip.

1a Forma Normale: Soluzione

- Si rimuove l'attributo SediDip e lo si pone in un'altra relazione separata con chiave combinazione di CodDip e SedeDip.

DIPARTIMENTI(CodDip, Nome, CodDir)

| <u>CodDip</u> | Nome | CodDir |
|---------------|-----------------|--------|
| D0001 | Amministrazione | 33301 |
| D0005 | Produzione | 18007 |
| D0003 | Ricerca | 33010 |

SEDI(CodDip:DIPARTIMENTI, SedeDip)

| <u>CodDip</u> | <u>SedeDip</u> |
|---------------|----------------|
| D0001 | Milano |
| D0001 | Napoli |
| D0001 | Roma |
| D0005 | Aprilia |
| D0003 | Napoli |

Questa soluzione non presenta ridondanze ed è completamente generale, non presentando limiti sul massimo numero di sedi per un dipartimento.

2a Forma Normale

- ❑ Una relazione è in 2NF se e solo se
 - è in 1NF
 - tutti gli attributi non-chiave dipendono funzionalmente dall'intera chiave composta (ovvero la relazione non ha attributi che dipendono funzionalmente da una parte della chiave)
- ❑ L'esempio:

| <u>Articolo</u> | <u>Magazzino</u> | Quantità | Indirizzo |
|-----------------|------------------|----------|------------------------|
| scarpe | VR1 | 25000 | v. Albere 17 - Verona |
| pantaloni | VR1 | 18000 | v. Albere 17 - Verona |
| scarpe | BO1 | 4500 | v. Agucchi 3 - Bologna |
| camicie | VR2 | 7000 | v. Monti 6 - Verona |

viola la 2NF perché

➤ **Magazzino** → **Indirizzo**

ossia, l'indirizzo **dipende solo parzialmente** dalla chiave.

Normalizzazione in 2NF

- ❑ La soluzione consiste nell'estrarre la FD che crea i problemi, generando gli schemi:

➤ MAG_ART (Articolo, Magazzino, Quantità) (AM → Q)

➤ MAGAZZINI (Magazzino, Indirizzo) (M → I)

| <u>Articolo</u> | <u>Magazzino</u> | Quantità |
|-----------------|------------------|----------|
| scarpe | VR1 | 25000 |
| pantaloni | VR1 | 18000 |
| scarpe | BO1 | 4500 |
| camicie | VR2 | 7000 |

| <u>Magazzino</u> | Indirizzo |
|------------------|------------------------|
| VR1 | v. Albere 17 - Verona |
| BO1 | v. Agucchi 3 - Bologna |
| VR2 | v. Monti 6 - Verona |

- ❑ L'informazione originale si può ricostruire eseguendo un join tra le due tabelle.

3a Forma Normale

- Una relazione è in 3NF se e solo se
 - è in 2NF
 - tutti gli attributi non-chiave dipendono dalla chiave soltanto, ossia non esistono attributi che dipendono da altri attributi non-chiave
- L'esempio:

| <u>Imp_cod</u> | Nome | Reparto | Caporeparto |
|----------------|---------|-----------|-------------|
| 001 | Rossi | Vendite | Marchi |
| 002 | Verdi | Acquisti | Stefani |
| 003 | Bianchi | Magazzino | Bielli |
| 004 | Neri | Vendite | Marchi |

viola la 3NF perché

- $\text{Imp_cod} \rightarrow \text{Nome, Reparto, Capo_reparto}$ ($I \rightarrow \text{NRC}$)
- $\text{Reparto} \rightarrow \text{Capo_reparto}$ ($R \rightarrow C$)

ossia, C **dipende transitivamente** dalla chiave I.

Normalizzazione in 3NF

- Anche in questo caso la soluzione consiste nell'estrarre la FD che crea i problemi, generando gli schemi:

- REP_IMP (Imp_cod, Nome, Reparto) ($I \rightarrow NR$)
- REP_CAPO (Reparto, Capo_reparto) ($R \rightarrow C$)

| <u>Imp_cod</u> | Nome | Reparto |
|----------------|---------|-----------|
| 001 | Rossi | Vendite |
| 002 | Verdi | Acquisti |
| 003 | Bianchi | Magazzino |
| 004 | Neri | Vendite |

| <u>Reparto</u> | Caporeparto |
|----------------|-------------|
| Vendite | Marchi |
| Acquisti | Stefani |
| Magazzino | Bielli |

- L'informazione originale si può ricostruire eseguendo un join tra le due tabelle.

Esempio di normalizzazione

- Anche lo schema di riferimento non è normalizzato (non è in 3NF né in 2NF); la soluzione consiste nel “decomporlo” sulla base delle FD.

| <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 |