



# **Vincoli di integrità Normalizzazione**

**A. Lorenzi, E. Cavalli**

**INFORMATICA PER ISTITUTI TECNICI TECNOLOGICI**

---

## **Integrità dei dati**

# Integrità dei dati

## Studenti

<u>Matricola</u>	Cognome	Nome	DataNascita	CodScuola
545	Rossi	Maria	NULL	125
653	Neri	Anna	20-set-1994	125
768	Verdi	Giuseppe	30-ott-1996	NULL
653	Rossi	Franco	32-ott-1994	180
314	Bruni	Enrico	27-ott-1995	185

## Scuole

<u>Codice</u>	NomeScuola
125	ITC Manzoni
180	Liceo Dante
190	Liceo Fermi
NULL	ITIS Galvani

E' un'anomalia solo se il campo è obbligatorio

- Ci sono situazioni anomale che invalidano i dati
  - Chiave con valore **nullo** e chiavi **duplicate**
  - Valore di una data **scorretto**
  - Riferimento alla scuola **mancante**
  - Riferimento a scuola **inesistente**

# Regole di integrità

---

- **Vincoli di chiave**

- La chiave primaria non può avere **valore nullo**
- La chiave primaria non può essere **duplicata**

- **Vincoli di tupla**

- **Obbligatorietà** dei dati
- Vincoli di **dominio** dei dati (specifici del problema)
- **Correlazione interna** fra campi del singolo record

- **Vincoli di Integrità referenziale**

- Riguarda **FK** di una tabella e **PK** della tabella correlata
- L'**integrità referenziale** richiede che per ogni valore **non nullo** della chiave esterna, esista un valore corrispondente della chiave primaria nella tabella associata

# Integrità referenziale

OK

Studenti

<u>Matricola</u>	Cognome	Nome	DataNascita	CodScuola
545	Rossi	Maria	NULL	125
653	Neri	Anna	20-set-1994	125
768	Verdi	Giuseppe	30-ott-1996	NULL
654	Rossi	Franco	31-ott-1994	180
314	Bruni	Enrico	27-ott-1995	185

Scuole

<u>Codice</u>	NomeScuola
125	ITC Manzoni
180	Liceo Dante
190	Liceo Fermi
185	ITIS Galvani

E' un'anomalia solo se il campo è obbligatorio

- Non si deve poter **inserire** una riga in **Studenti** con valore di *CodScuola* che non compare fra quelli di *Codice* in **Scuole**
- Non è possibile **cancellare** una scuola dalla tabella **Scuole** se ci sono righe nella tabella **Studenti** che si riferiscono ad essa
- Non si possono **modificare** i valori di *Codice* o di *CodScuola* se sono violate le regole di integrità referenziale

# Integrità referenziale e DBMS

Modifica relazioni

Tabella/query: Agenti    Tabella/query correlata: Clienti

Crea

Annulla

IDAgente    IDAgente

☒ Applica integrità referenziale

☐ Aggiorna campi correlati a cate

☐ Elimina record correlati a cater

Tipo relazione: Uno-a-molti

Applicando l'integrità referenziale il **DBMS**:

- Impedisce di **immettere** (o di **modificare**) un valore nella chiave esterna della tabella associata, se tale valore non esiste tra le chiavi della tabella primaria. È possibile invece immettere un valore nullo nella chiave esterna, per rappresentare il fatto che le righe non sono correlate.
- Impedisce di **eliminare** una n-upla dalla tabella primaria, se esistono righe legate ad essa attraverso la chiave esterna nella tabella correlata.
- Impedisce di **modificare** il valore alla chiave nella tabella primaria, se ad essa corrispondono righe nella tabella correlata.

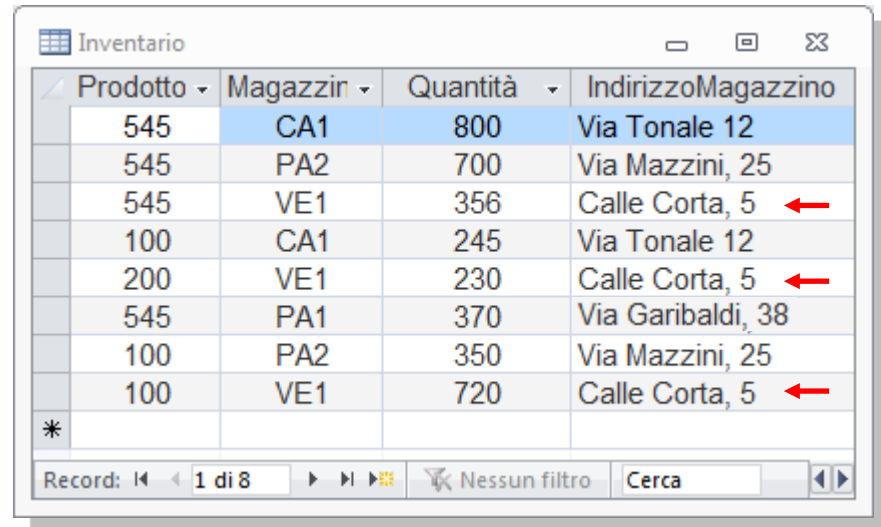
---

# **Normalizzazione**

# Ridondanza e anomalie (1)

**Inventario** è una buona tabella?

- Il campo *IndirizzoMagazzino* è ripetuto per ogni codice magazzino in **Inventario** e ci sono dati ridondanti



Prodotto	Magazzino	Quantità	IndirizzoMagazzino
545	CA1	800	Via Tonale 12
545	PA2	700	Via Mazzini, 25
545	VE1	356	Calle Corta, 5 ←
100	CA1	245	Via Tonale 12
200	VE1	230	Calle Corta, 5 ←
545	PA1	370	Via Garibaldi, 38
100	PA2	350	Via Mazzini, 25
100	VE1	720	Calle Corta, 5 ←
*			

Record: 1 di 8 Nessun filtro Cerca

- La ridondanza spreca spazio ma, peggio, è causa di **anomalie**:
  - Anomalia di **aggiornamento**
  - Anomalia di **inserimento**
  - Anomalia di **cancellazione**



# Ridondanza e anomalie (2)

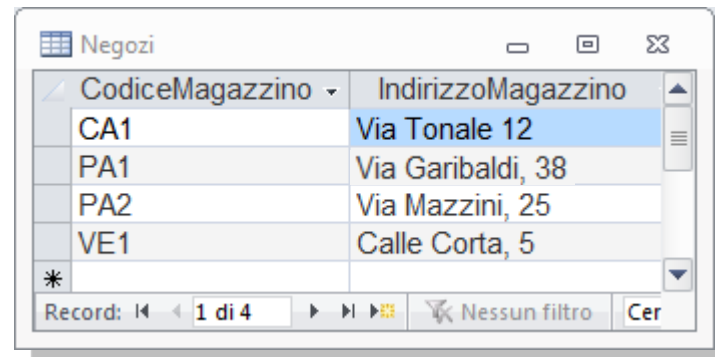
- Per evitare la ridondanza si scompone la tabella originale in:

**Inventario** (Prodotto, Magazzino, Quantità)

**Negozi** (CodiceMagazzino, IndirizzoMagazzino)



Magazzino	Prodotto	Quantità
CA1	100	245
PA2	100	350
VE1	100	720
VE1	200	230
CA1	545	800
PA1	545	370
PA2	545	700
VE1	545	356

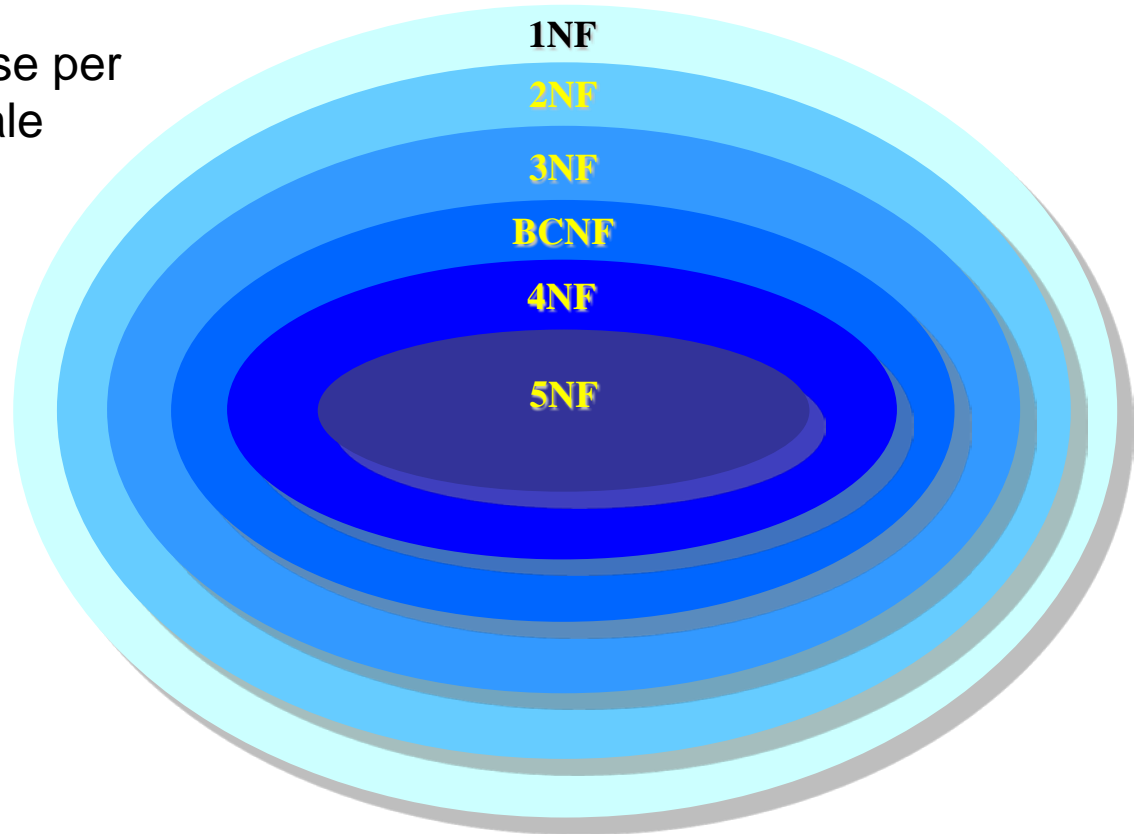


CodiceMagazzino	IndirizzoMagazzino
CA1	Via Tonale 12
PA1	Via Garibaldi, 38
PA2	Via Mazzini, 25
VE1	Calle Corta, 5

- Le due tabelle sono ricavate per **proiezione** dalla tabella di partenza
- La informazioni nella tabella originaria sono ricostruibili con:  
**Inventario Join Negozi** per *Magazzino = CodiceMagazzino*
- C'è un modo sistematico per capire se una tabella è una buona tabella?

# Forme normali

- Criteri che definiscono le condizioni che devono essere soddisfatte per evitare situazioni anomale: **forme normali**
- **1NF** richieste di base per il modello relazionale
- **2NF, 3NF, BCNF** anomalie da dipendenze funzionali
- **4NF** anomalie per dipendenze multivalore
- ...



**\*\* Ci interessano: 1NF, 2NF, 3NF**

# Definizioni

- **Chiave** o **chiave primaria** è l'insieme di uno o più attributi che identificano in modo univoco una riga della tabella
- **Chiave candidata** è ogni insieme minimale di attributi che possono svolgere la funzione di chiave (ci possono essere molte chiavi candidate, ma una sola chiave primaria)
- **Attributo non-chiave** è un campo che non fa parte della chiave primaria
- Esempio:  
**Inventario**( Numero, Prodotto, Magazzino, Quantità, IndirizzoMagazzino )
  - *Numero* è chiave candidata
  - { *Prodotto*, *Magazzino* } è chiave candidata
  - { *Prodotto*, *Magazzino*, *Quantità* } non è chiave candidata; un insieme di attributi come questo si dice *superchiave*

# Dipendenze funzionali (1)

Si ha **dipendenza funzionale** tra attributi quando il valore di uno o più attributi **A** determina univocamente il valore di un attributo **B** e si indica con:  $A \rightarrow B$ . Si dice che **A** determina funzionalmente **B**, o che:

- **B** dipende funzionalmente da **A**
- **A** è un determinante per **B**

**Inventario**( Numero, Prodotto, Magazzino, Quantità, IndirizzoMagazzino )

Numero  $\rightarrow$  ( Prodotto, Magazzino, Quantità, IndirizzoMagazzino )

{Prodotto, Magazzino}  $\rightarrow$  ( Numero, Quantità, IndirizzoMagazzino )

Magazzino  $\rightarrow$  IndirizzoMagazzino

- Una **chiave candidata** di una relazione è determinante per ogni attributo della relazione. Viceversa un **determinante** per ogni attributo di una relazione, è chiave candidata per la relazione stessa.

# Dipendenze funzionali (2)

IDProdotto	Descrizione	Reparto	Compratore
1001	Pinne piccole	Mare	Enrico
1002	Pinne medie	Mare	Enrico
1010	Maschera media	Mare	Mary
1011	Maschera grande	Mare	Mary
2010	Tenda igloo	Camping	Gianni
2012	Tenda doppio igloo	Camping	Gianni
2050	Sci fondo	Montagna	Damiano
2051	Sci discesa	Montagna	Damiano

Scoprire le dipendenze funzionali non è facile!  
Si tratta di un'attività simile alla progettazione concettuale dei dati

IDProdotto → Descrizione

IDProdotto → Reparto

IDProdotto → Compratore

- **IDProdotto** → ( Descrizione, Reparto, Compratore )
- **Descrizione** → ( IDProdotto, Reparto, Compratore )
- **Compratore** → Reparto

# Dipendenze funzionali (3)

**Dipendenti**( Matricola, Nome, Cognome, Stipendio, CodiceCapo, NomeCapo )

Matricola  $\rightarrow$  ( Nome, Cognome, Stipendio, CodiceCapo )

CodiceCapo  $\rightarrow$  NomeCapo

Matricola  $\rightarrow$  NomeCapo **transitivamente**

Matricola  $\rightarrow$  ( Nome, Cognome, Stipendio, CodiceCapo, NomeCapo )

Matricola è chiave per **Dipendenti**

Si ha **dipendenza transitiva** tra attributi quando un attributo **A** determina **B** e **B** determina **C**; si dice allora che **C** dipende transitivamente da **A**:

**A  $\rightarrow$  B e B  $\rightarrow$  C allora A  $\rightarrow$  C transitivamente**

# Prima Forma Normale 1NF

---

## Prima forma normale

- Tutte le righe hanno lo stesso numero di attributi e contengono informazioni di una (sola) entità
- I valori delle colonne rappresentano informazioni elementari (non ci sono informazioni di gruppo)
- Tutti i valori in una colonna sono del medesimo tipo
- Non ci possono essere righe duplicate, quindi ci deve essere un attributo o un insieme di attributi con la funzione di chiave primaria
- L'ordine delle colonne non è rilevante (perché le colonne hanno un'intestazione)
- L'ordine delle righe non è rilevante

# Seconda Forma Normale 2NF

Una relazione è in **seconda forma normale** (2FN) quando è in prima forma normale e tutti i suoi attributi non-chiave dipendono dall'intera chiave e non da una parte della chiave

**Inventario** (Prodotto, Magazzino, Quantità, IndirizzoMagazzino)  
{Prodotto, Magazzino} è chiave per **Inventario**

Magazzino → IndirizzoMagazzino ☒☒

Violazione alla 2NF

1) Si costruisce una tabella con gli attributi di ☒☒

**Magazzini** ( Magazzino, IndirizzoMagazzino )

2) Si costruisce una seconda tabella con gli attributi di **Inventario** togliendo quelli di ☒☒ e integrati con l'attributo *Magazzino* per avere un collegamento con **Magazzini**

**Inventario1** ( Prodotto, Magazzino, Quantità )



# Algoritmo di scomposizione

## Algoritmo di scomposizione di una relazione **T**

1. Si identificano tutte le dipendenze funzionali e le chiavi candidate di **T**
2. Se c'è una dipendenza funzionale che viola le regole di normalizzazione:
  - a. si costruisce una nuova tabella con tutti gli attributi della dipendenza funzionale considerata
  - b. il determinante della dipendenza funzionale è la chiave primaria della tabella definita in a.
  - c. si costruisce una seconda tabella rimuovendo dalla tabella di partenza tutti gli attributi che sono determinati dalla dipendenza funzionale in esame
3. Si ripete il passo 2. sino a che non ci sono più violazioni alle regole di normalizzazione

# Esempio di scomposizione 2NF

**Inventario** (Prodotto, Magazzino, Qta, NomeProdotto, IndirizzoMagazzino)

Dipendenze funzionali:

1. {Prodotto, Magazzino} → (Qta, NomeProdotto, IndirizzoMagazzino)
2. Prodotto → NomeProdotto ⊗
3. Magazzino → IndirizzoMagazzino ⊗⊗

Violazioni alla 2NF

{Prodotto, Magazzino} è chiave per **Inventario**

⊗ **Prodotti** ( Prodotto, NomeProdotto )

**Inventario1** ( Prodotto, Magazzino, Qta, IndirizzoMagazzino )

Finito?

⊗⊗ **Magazzini** ( Magazzino, IndirizzoMagazzino )

**Inventario2** ( Prodotto, Magazzino, Qta )

**Prodotti** ( Prodotto, NomeProdotto )

**Magazzini** ( Magazzino, IndirizzoMagazzino )

**Inventario2** ( Prodotto, Magazzino, Qta )

# Terza Forma Normale 3NF

Una relazione è in **terza forma normale** (3FN) quando è in seconda forma normale e tutti gli attributi non-chiave dipendono direttamente dalla chiave, quindi non hanno dipendenze **transitive** dalla chiave

**Studenti** ( Nome, Scuola, TelefonoScuola )

Nome	Scuola	TelefonoScuola
Bedoni	Liceo Fermi	045-2345613
Bianchi	ITC Paleari	02-2538469
Galli	ITIS Leonardo	011-3377659
Leoni	Liceo Dante	06-3845287
Lorenzini	ITIS Galvani	081-5416875
Negri	Liceo Dante	06-3845287
Rossi	ITC Paleari	02-2538469
Verdi	Liceo Dante	06-3845287

Dipendenze funzionali:

1. Nome → ( Scuola, TelefonoScuola )

2. Scuola → TelefonoScuola ⊗

Violazione alla 3NF

Nome è chiave per **Studenti**

⊗ **Scuole** ( Scuola, TelefonoScuola )

**Studenti1** ( Nome, Scuola )

# Esempio di scomposizione 3NF

**Dipendenti** (Matr, Cognome, Nome, CodRep, NomeRep, CodCittà, NomeCittà)

Dipendenze funzionali:

1. Matr  $\rightarrow$  ( Cognome, Nome, CodRep, NomeRep, CodCittà, NomeCittà )
2. CodRep  $\rightarrow$  NomeRep  $\otimes$
3. CodCittà  $\rightarrow$  NomeCittà  $\otimes\otimes$

Violazioni alla 3NF

Matr è chiave per **Dipendenti**

$\otimes$  **Reparti** ( CodRep, NomeRep )

**Dipendenti1** ( Matr, Cognome, Nome, *CodRep*, CodCittà, NomeCittà )

$\otimes\otimes$  **Città**( CodCittà, NomeCittà )

**Dipendenti2** ( Matr, Cognome, Nome, *CodRep*, *CodCittà* )

Finito?

**Reparti** ( CodRep, NomeRep )

**Città**( CodCittà, NomeCittà )

**Dipendenti2** ( Matr, Cognome, Nome, *CodRep*, *CodCittà* )