



UNIVERSITA DEGLI STUDI DI FOGGIA

Dipartimento di Agraria

Cdl in Ingegneria dei Sistemi Logistici per l'Agroalimentare

Corso integrato di Sistemi di Elaborazione

Modulo I

Prof. Crescenzo Gallo

crescenzo.gallo@unifg.it

La normalizzazione di una base di dati

La normalizzazione delle tabelle

- Una **forma normale** è una proprietà di uno schema relazionale che ne garantisce la qualità misurata in assenza di determinati difetti.
- Uno schema non normalizzato può presentare **anomalie** durante le operazioni di inserimento, modifica e cancellazione dei dati.
- La **normalizzazione** è un procedimento che trasforma uno schema che presenta delle anomalie in uno equivalente normalizzato.

La normalizzazione delle tabelle

Consideriamo la tabella **Magazzino** di una concessionaria auto:

<u>CodCli</u>	Indirizzo	Città	Cap	<u>CodAcc</u>	Descrizione	Prezzo	Quantità
C01	via Po, 23	Pisa	56100	M03	Batteria	100,00	3
C01	via Po, 23	Pisa	56100	M12	Radiatore	1200,00	1
C01	via Po, 23	Pisa	56100	M04	Antenna	25,00	3
C02	via Moro, 2	Pisa	56100	M03	Batteria	100,00	2
C02	via Moro, 2	Pisa	56100	M12	Radiatore	1200,00	1
C03	via Roma, 1	Lucca	55100	M03	Batteria	100,00	2

Possiamo evidenziare le seguenti anomalie:

- **anomalia di inserimento** (non si può inserire un nuovo cliente senza aver prima inserito i dati relativi all'auto e agli accessori ordinati);

La normalizzazione delle tabelle

Consideriamo la tabella **Magazzino** di una concessionaria auto:

<u>CodCli</u>	Indirizzo	Città	Cap	<u>CodAcc</u>	Descrizione	Prezzo	Quantità
C01	via Po, 23	Pisa	56100	M03	Batteria	100,00	3
C01	via Po, 23	Pisa	56100	M12	Radiatore	1200,00	1
C01	via Po, 23	Pisa	56100	M04	Antenna	25,00	3
C02	via Moro, 2	Pisa	56100	M03	Batteria	100,00	2
C02	via Moro, 2	Pisa	56100	M12	Radiatore	1200,00	1
C03	via Roma, 1	Lucca	55100	M03	Batteria	100,00	2

Possiamo evidenziare le seguenti anomalie:

- **anomalia di aggiornamento** (l'indirizzo del cliente è ripetuto; se varia, va modificato in tutte le righe, altrimenti si provoca una inconsistenza nei dati);

La normalizzazione delle tabelle

Consideriamo la tabella **Magazzino** di una concessionaria auto:

<u>CodCli</u>	Indirizzo	Città	Cap	<u>CodAcc</u>	Descrizione	Prezzo	Quantità
C01	via Po, 23	Pisa	56100	M03	Batteria	100,00	3
C01	via Po, 23	Pisa	56100	M12	Radiatore	1200,00	1
C01	via Po, 23	Pisa	56100	M04	Antenna	25,00	3
C02	via Moro, 2	Pisa	56100	M03	Batteria	100,00	2
C02	via Moro, 2	Pisa	56100	M12	Radiatore	1200,00	1
C03	via Roma, 1	Lucca	55100	M03	Batteria	100,00	2

Possiamo evidenziare le seguenti anomalie:

- **anomalia di cancellazione** (cancellando ad es. la riga con chiave C03-M03 si perdono le informazioni relative al cliente C03).

La normalizzazione delle tabelle

Queste anomalie si verificano poiché abbiamo rappresentato informazioni eterogenee in un'unica tabella:

- gli *articoli* presenti in magazzino;
- i dati anagrafici dei *clienti*;
- gli *ordini* dei clienti relativi a determinati articoli.

Il processo di **normalizzazione** elimina tali anomalie, suddividendo la tabella originaria in una serie di tabelle collegate.

Esistono molte forme normali; per i nostri scopi è sufficiente considerare le prime tre.

Prima forma normale (1FN)

- La **prima forma normale** di una tabella è quella più semplice e spesso viene trascurata, poiché è quella in cui si trovano la maggior parte delle tabelle del modello (logico) relazionale.
- Diremo che una tabella T è in **prima forma normale (1FN)** quando rispetta i requisiti fondamentali del modello relazionale:
 - *i valori di un attributo (colonna) sono dello stesso tipo, cioè appartengono allo stesso dominio;*
 - *i valori di una n-upla (riga) sono diversi per almeno un attributo, cioè non possono esistere due righe identiche;*
 - *l'ordine delle righe (e delle colonne) è irrilevante;*
 - *gli attributi sono di tipo elementare (non composti da altri attributi).*

Prima forma normale (1FN)

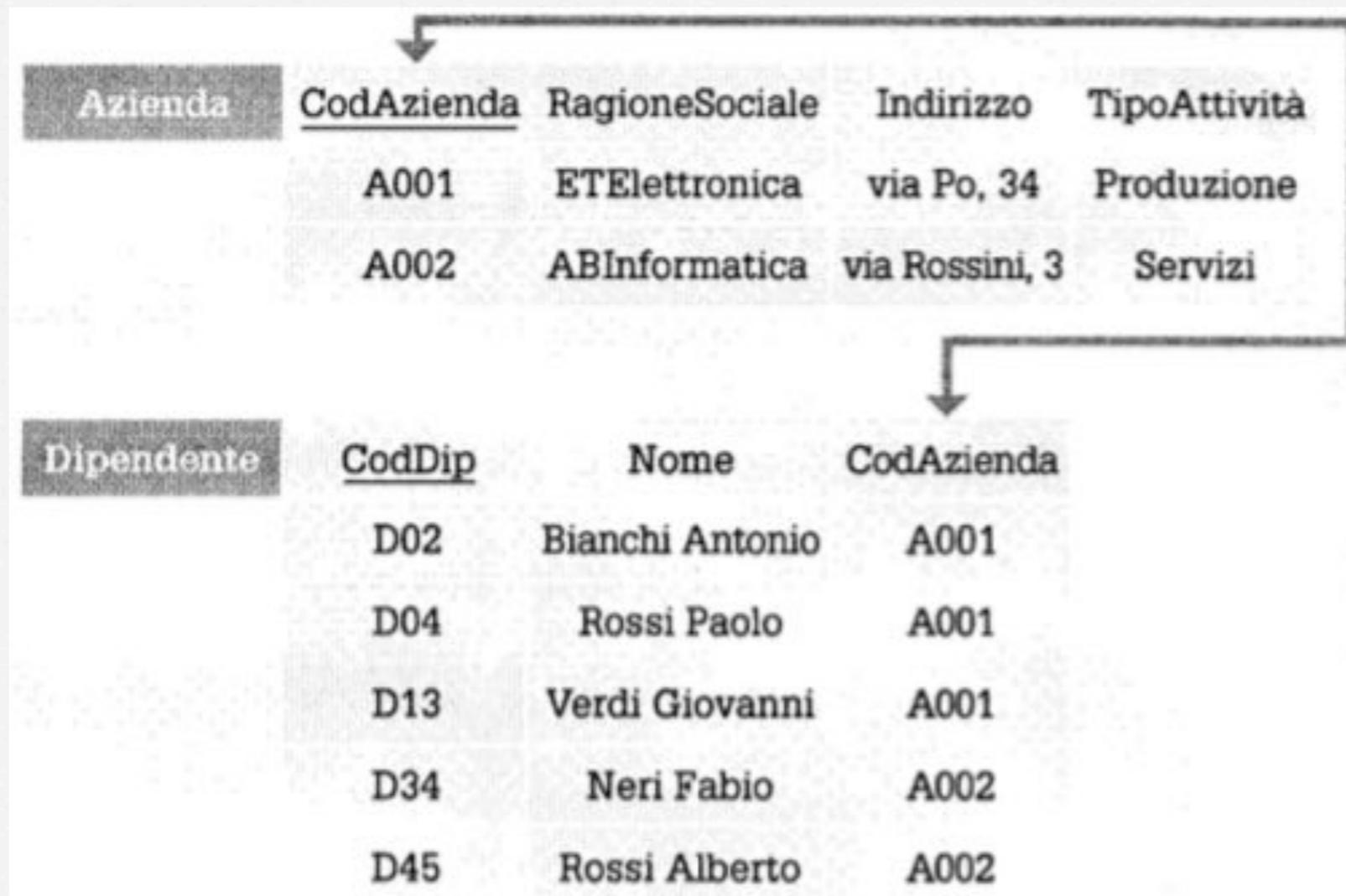
- Consideriamo la seguente tabella, contenente i dati di alcune aziende e relativi dipendenti:

<u>CodAzienda</u>	RagioneSociale	Indirizzo	TipoAttività	Dipendenti
A001	ETElettronica	via Po, 34	Produzione	Bianchi Antonio Rossi Paolo Verdi Giovanni
A002	ABInformatica	via Rossini, 3	Servizi	Neri Fabio Rossi Alberto

- La tabella non è in 1FN poiché l'attributo Dipendenti è composto da un gruppo di attributi ripetuti (una lista di nomi di persone).

Prima forma normale (1FN)

- Decomponendo la tabella originaria otteniamo:



Dipendenze funzionali

- Data una tabella T e un insieme $X = \{X_1, X_2, \dots, X_n\}$ di attributi di T , si dice che un attributo Y di T **dipende funzionalmente** da X e si scrive:

$$X_1, X_2, \dots, X_n \rightarrow Y$$

se e solo se i valori degli attributi di X determinano univocamente il valore dell'attributo Y per ogni riga della tabella T .

- Si dice anche che X **determina** Y .
- La dipendenza funzionale di un attributo da uno o più attributi è quindi una proprietà che non dipende da una singola riga della tabella, è cioè *indipendente dal tempo* (**proprietà semantica**).

Dipendenze funzionali

Nella tabella *Magazzino* possiamo individuare le dipendenze funzionali:

<u>CodCli</u>	Indirizzo	Città	Cap	<u>CodAcc</u>	Descrizione	Prezzo	Quantità
C01	via Po, 23	Pisa	56100	M03	Batteria	100,00	3
C01	via Po, 23	Pisa	56100	M12	Radiatore	1200,00	1
C01	via Po, 23	Pisa	56100	M04	Antenna	25,00	3
C02	via Moro, 2	Pisa	56100	M03	Batteria	100,00	2
C02	via Moro, 2	Pisa	56100	M12	Radiatore	1200,00	1
C03	via Roma, 1	Lucca	55100	M03	Batteria	100,00	2

- CodCli → Indirizzo, Città, CAP
- CodAcc → Descrizione, Prezzo
- CodCli, CodAcc → Quantità

Seconda forma normale (2FN)

- Diremo che una tabella T è in **seconda forma normale (2FN)** se essa è in 1FN e non esistono attributi dipendenti solo da una *parte* della chiave.
- La tabella Magazzino(CodCli, Indirizzo, Città, CAP, CodAcc, Descrizione, Prezzo, Quantità) non è in 2FN poiché:
 - la chiave è composta dai due attributi *CodCli* e *CodAcc*;
 - gli attributi *Indirizzo*, *Città* e *CAP* dipendono solo da *CodCli*;
 - gli attributi *Descrizione* e *Prezzo* dipendono solo da *CodAcc*.

Seconda forma normale (2FN)

Questo comporta la suddivisione della tabella *Magazzino* nelle tabelle:

- Cliente(CodCli, Indirizzo, Città, CAP)
- Accessorio(CodAcc, Descrizione, Prezzo)
- Ordine(CodCli, CodAcc, Quantità)

<u>CodCli</u>	Indirizzo	Città	Cap
C01	via Po, 23	Pisa	56100
C01	via Po, 23	Pisa	56100
C01	via Po, 23	Pisa	56100
C02	via Moro, 2	Pisa	56100
C02	via Moro, 2	Pisa	56100
C03	via Roma, 1	Lucca	55100

<u>CodAcc</u>	Descrizione	Prezzo	Quantità
M03	Batteria	100,00	3
M12	Radiatore	1200,00	1
M04	Antenna	25,00	3
M03	Batteria	100,00	2
M12	Radiatore	1200,00	1
M03	Batteria	100,00	2

Terza forma normale (3FN)

- Le tabelle in 2FN possono ancora essere esposte ad anomalie.
- Infatti, osservando la tabella Cliente(CodCli, Indirizzo, Città, CAP) possiamo osservare che esiste la dipendenza funzionale:

Città → CAP

in cui il CAP dipende dalla città di residenza del cliente.

- Ma poiché vale anche:

CodCli → Città

abbiamo la seguente catena di dipendenze:

CodCli → Città → CAP

dalla quale si evince che l'attributo CAP dipende transitivamente (e non direttamente) dalla chiave.

Terza forma normale (3FN)

- Diremo quindi che una tabella T è in **terza forma normale (3FN)** se essa è in 2FN e ogni attributo non facente parte della chiave dipende direttamente dalla chiave (cioè non esistono *attributi non chiave* che dipendono da altri attributi non chiave).
- Per trasformare una tabella in 3FN si crea una nuova tabella per ogni gruppo di attributi non chiave coinvolti nella dipendenza funzionale da altri attributi non chiave.
- Ad es. la tabella Cliente(CodCli, Indirizzo, Città, CAP) con la dipendenza funzionale Città → CAP deve decomporsi nelle tabelle:
 - Cliente(CodCli, Indirizzo, Città)
 - Città(Città, CAP)

Considerazioni

- La *normalizzazione* risolve molte anomalie nelle operazioni sui dati, ma rende più complesse le operazioni di interrogazione.
- Naturalmente ciò perché sono più importanti l'integrità e la consistenza della base di dati piuttosto che la velocità di risposta alle interrogazioni (infatti talvolta per aumentare la velocità si tollera talvolta una qualche forma di “denormalizzazione”, cioè di ridondanza controllata dei dati).
- La decomposizione in più tabelle è eseguita in modo da poter ricomporre l'informazione attraverso un'operazione di *natural join* sugli attributi comuni.