

Basi di dati relazionali

Gianluca Amato

Idoneità Informatica per Lauree Magistrali
Università “G. D'Annunzio” di Chieti-Pescara
Facoltà di Economia
anno accademico 2010/2011
ultimo aggiornamto: 4/18/11

Licenza D'Uso

- © 2011 Gianluca Amato
- Queste slide sono distribuite in accordo alla licenza

Creative Commons Attribution-ShareAlike 3.0
(CC BY-SA 3.0)

<http://creativecommons.org/licenses/by-sa/3.0/>

Basi di Dati

Archivi

- Un **archivio** è una **raccolta strutturata di informazioni**.
 - Dizionario
 - Pagine gialle
 - Elenco libri di una biblioteca
- Le informazioni sono **strutturate**.. per ogni nominativo nelle pagine gialle troviamo una serie di dati fissi
 - Nome, cognome, telefono, indirizzo
- Un archivio può essere cartaceo o elettronica
 - Un archivio elettronico è spesso chiamato **database** o **base di dati**.

Basi di dati

- Le basi di dati hanno numerosi vantaggi rispetto agli archivi tradizionali
 - Ricerche più versatili e veloci
 - Sulle pagine gialle posso fare solo ricerche per nome
 - Sulle pagine gialle elettroniche posso fare ricerche anche per vicinanza o numeri di telefono
 - Possibilità di ordinare e filtrare i dati
 - Minore occupazione di spazio
- I programmi che si occupano di gestire le basi di dati prendono il nome di
 - **Sistemi di Gestione di Basi di Dati** (SGBD)
 - oppure **DataBase Management System** (DBMS).

Esempi di basi di dati

- Sistemi di prenotazione aerea
- Pubblica amministrazione
 - PRA
 - Catasto
- Conti correnti bancari
- Cartelle cliniche dei pazienti ricoverati in ospedale
- Dati degli studenti immatricolati in una università
- Pagine gialle/bianche elettroniche

Dati e informazioni 1

- Per poter rappresentare le **informazioni** in un database occorre prima capirne la **struttura** in modo da distinguere i **dati** che la compongono.
 - Informazioni:
 - Il libro *Harry Potter e la pietra filosofale*, dell'autrice Joanne K. Rowling, è di genere fantasy, si compone di 293 pagine e il suo prezzo è di € 16.80.
 - Il libro *Dieci Piccoli Indiani*, di Agata Christie, è un giallo, si compone di 210 pagine, e costa € 7.70.
 - Struttura
 - Ogni libro è caratterizzato da un titolo, un autore, un genere, un numero di pagine e un prezzo.
 - Inoltre... numero di pagine è un valore numerico intero, il prezzo è un valore numerico con due cifre dopo la virgola, titolo e autori sono sequenze di caratteri

Dati e informazioni 2

- Informazione
 - Il libro *Dieci Piccoli Indiani*, di Agata Christie, è un giallo, si compone di 210 pagine, e costa € 7.70.
- Dati corrispondenti
 - Dieci Piccoli Indiani
 - Agata Christie
 - Giallo
 - 210
 - 7.70

Dati e informazioni 3

- Detto in altri termini
 - Un dato è una descrizione elementare che strutturata con le altre, è in grado di rappresentare una informazione.
- Ci sono varie metodologie con cui estrarre dati e struttura dalle informazioni, che danno origine a tipi diversi di database
 - Database relazionali
 - Database gerarchici
 -
- Noi ci occuperemo esclusivamente dei database più comuni, i **database relazionali**.

Basi di dati relazionali

Tabelle

- Nei database relazionali i dati sono strutturati in tabelle.
 - Ogni **riga** (chiamata anche **record** o **istanza**) della tabella è una singola informazione che vogliamo memorizzare.
 - Ogni **colonna** (o **campo**) è un singolo dato che vogliamo memorizzare per tutte le informazioni
- Ecco la rappresentazione tabellare delle informazioni di prima

Titolo	Genere	Pagine	Prezzo	Autore
Harry Potter e la pietra filosofale	Fantasy	293	16,80	J. K. Rowling
Dieci piccoli indiani	Giallo	530	29,90	A. Christie

Ridondanza 1

- Supponiamo di aggiungere anche delle informazioni sull'autore, come nazionalità e anno di nascita

Titolo	Genere	Pagine	Prezzo	Autore	Nazionalità	Anno nascita
Harry Potter e la pietra filosofale	Fantasy	293	16,80	J. K. Rowling	Britannica	1965
Dieci piccoli indiani	Giallo	530	29,90	A. Christie	Britannica	1890
Harry Potter e il prigioniero di Azkaban	Fantasy	366	14,50	J. K. Rowling	Britannica	1965
Io non ho paura	Romanzo	219	9,50	N. Ammanniti	Italiana	1966

- Notare che c'è **ridondanza**
 - Tutte le volte che il campo autore contiene *J. K. Rowling*, la nazionalità dovrà sempre essere *Britannica* e l'anno di nascita sempre *1965*.

Ridondanza 2

- **Equazione Fondamentale** delle basi di dati

Ridondanza = MALE

- Perché?
 - Possibilità di avere basi di dati inconsistenti
 - Più righe con lo stesso autore hanno date di nascita diversa.. qual'è quella giusta?
 - Quando l'informazione è duplicata, è più facile commettere errori!!
 - Necessaria più memoria per registrare più volte gli stessi dati

Ridondanza 3

- Perché abbiamo ridondanza?
 - Stiamo cercando di rappresentare nella stessa tabella informazioni di natura differente!!!
 - *Genere, prezzo, pagine, autore* sono caratteristiche di un libro..
 - ... ma *nazionalità e anno di nascita* sono caratteristiche di un autore, non di un libro!
- Come risolviamo il problema?
 - Usiamo più di una tabella.

Ridondanza 4

- Tabella libri

Titolo	Genere	Pagine	Prezzo	Autore
Harry Potter e la pietra filosofale	Fantasy	293	16,80	J. K. Rowling
Dieci piccoli indiani	Giallo	530	29,90	A. Christie
Harry Potter e il prigioniero di Azkaban	Fantasy	366	14,50	J. K. Rowling
Io non ho paura	Romanzo	219	9,50	N. Ammanniti

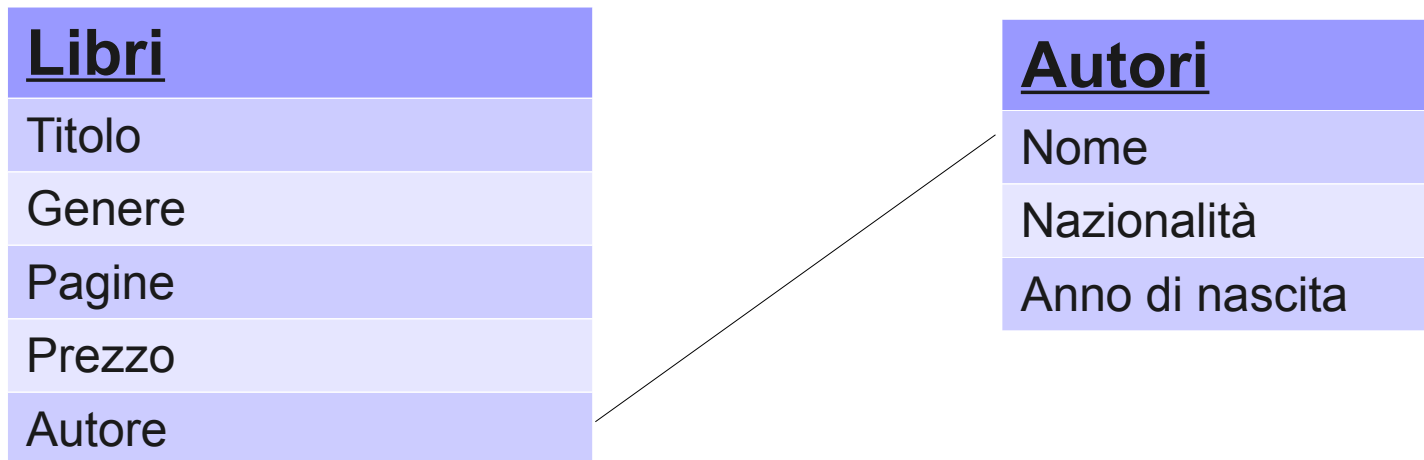
- Tabella autori



Nome	Nazionalità	Anno nascita
J. K. Rowling	Britannica	1965
A. Christie	Britannica	1980
N. Ammanniti	Italiana	1966

Schema logico

- Si crea una relazione tra la tabella libri e quella autori
 - Il campo autore della tabella libri si riferisce al campo nome della tabella autori
- Nomi delle tabelle, campi e loro relazioni si possono rappresentare in un disegno chiamato **schema logico**.



- Quando in una base di dati non ci sono ridondanze, si dice che è **normalizzata**.

Interrogazioni 1

- Un database è inutile se non si possono recuperare da esso le informazioni che servono
 - l'operazione di recupero informazioni da un database si chiama **interrogazione**.
- Le interrogazioni possono essere più o meno semplici:
 - Dammi tutte le informazioni sul libro *Harry Potter e la Pietra Filosofale*
 - Dammi l'elenco degli autori che hanno scritto più di uno libro e per il quale la differenza di prezzo tra il libro più economico e quello più costoso è superiore a 20 €.
- È possibile avere risposta ad una interrogazione solo se il database contiene le informazioni rilevanti

Interrogazioni 2

- A quali di queste interrogazioni si può rispondere nel nostro caso?
 - Tutti i libri scritti da J. K. Rowling
 - Il numero di libri scritti da ogni autore
 - Gli autori che hanno scritto più di uno libro e per il quale la differenza di prezzo tra il libro più economico e quello più costoso è superiore a 20 €
 - Gli autori che hanno scritto almeno un libro ogni 5 anni

Nome	Nazionalità	Anno nascita
J. K. Rowling	Britannica	1965
A. Christie	Britannica	1980
N. Ammanniti	Italiana	1966

Titolo	Prezzo	Autore
Harry Potter e la pietra filosofale	16,80	J. K. Rowling
Dieci piccoli indiani	29,90	A. Christie
Harry Potter e il prigioniero di Azkaban	14,50	J. K. Rowling
Io non ho paura	9,50	N. Ammanniti

Linguaggi di Interrogazione

- Ovviamente non possiamo fornire le interrogazioni al computer scrivendole in italiano (o in inglese)
 - Il computer è stupido, e non capisce.
- Esistono dei linguaggi speciali che il computer capisce col quale possiamo interrogare una base di dati
 - **Structured Query Language (SQL)**
 - Più potente e versatile, ma più complesso.
 - **Query By Example (QBE)**
 - Più semplice e intuitivo, ma meno versatile.. è quello che studieremo noi.
 - Implementato in Microsoft Access, OpenOffice, LibreOffice, etc...

Chiavi primarie ed esterne

Chiavi primarie 1

- Cosa succede nel caso di omonimia?
 - Quale delle due Rowling ha scritto i romanzi di Harry Potter?

Nome	Nazionalità	Anno nascita	Titolo	Prezzo	Autore
J. K. Rowling	Britannica	1965	Harry Potter e la pietra filosofale	16,80	J. K. Rowling
A. Christie	Britannica	1980	Dieci piccoli indiani	29,90	A. Christie
N. Ammanniti	Italiana	1966	Harry Potter e il prigioniero di Azkaban	14,50	J. K. Rowling
J. K. Rowling	Statunitense	1983	Io non ho paura	9,50	N. Ammanniti

- Bisogna avere un campo (o un gruppo di campi) che identifica univocamente una riga: la **chiave primaria**.
 - Non possono esistere due righe con lo stesso valore di chiave primaria.

Chiavi primarie 2

- Possibili chiavi primarie
 - Codice fiscale
 - Telaio automobile
 - Codice ISBN per un libro
- E quando non esiste nessun campo che possa andare bene?
 - Nel caso degli autori, non ne conosciamo certo il codice fiscale.
 - Si crea un campo nuovo (tipicamente numerico) da utilizzare esclusivamente per identificare le righe.

Chiavi primarie 3

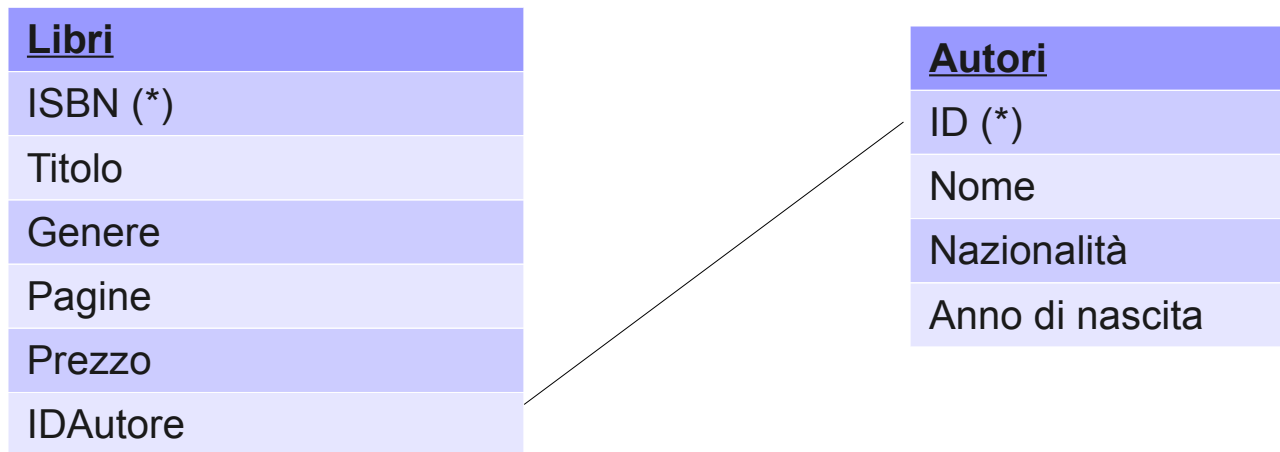
- Ecco la base di dati di prima con l'aggiunta di due chiavi primarie per le tabelle autori e libri

ID	Nome	Nazionalità	Anno nascita
1	J. K. Rowling	Britannica	1965
2	A. Christie	Britannica	1980
3	N. Ammanniti	Italiana	1966
4	J. K. Rowling	Statunitense	1983

ISBN	Titolo	Prezzo	IDAutore
3858-38-4923	Harry Potter e la pietra filosofale	16,80	1
xxxxx	Dieci piccoli indiani	29,90	2
xxxxx	Harry Potter e il prigioniero di Azkaban	14,50	1
xxxxx	Io non ho paura	9,50	3

Chiavi esterne

- I campi di una tabella che si riferiscono alle chiavi primarie di un'altra tabella prendono il nome di **chiavi esterne**.
- Il nuovo schema logico è



- I campi con l'asterisco sono le chiavi primarie.
- IDAutore è una chiave esterna che si riferisce alla tabella Autore.

Riassumendo

- Una chiave primaria è un campo (o una combinazione di campi) che identifica in maniera univoca una riga di una tabella.
- Una chiave esterna è un campo (o una combinazione di campi) che si riferisce alla chiave primaria di un'altra tabella
 - Serve a collegare le informazioni di diverse tabelle
- Il disegno che mostra le tabelle, i campi, le chiavi primarie e le chiavi esterne prende il nome di schema logico.

Tipi di dato

Campi

- Cosa può essere usato come valore per un campo?
 - Per ora siamo stati molto vaghi a proposito
- Puntualizziamo di nuovo un aspetto fondamentale:
 - Ogni campo può contenere un solo dato elementare.
 - Il campo Titolo contiene solo il titolo del libro
 - Il campo Prezzo contiene solo il prezzo
 - Violare questa regola porta solo guai
 - Se un solo campo contiene sia il titolo che il prezzo, può poi diventare impossibile ricercare tutti i libri che costano meno di 10 €

Tipi di dato

- Cosa può essere un dato elementare?
- In fase di creazione della tabella, i sistemi DBMS obbligano a scegliere un **tipo di dato**
 - Il tipo di dato limita i valori che possono essere messi in un campo.
 - Ci sono 5 categorie principali per i tipi di dato
 - Numerico: a sua volta divisi in
 - Intero, Virgola mobile e Virgola fissa
 - Stringa
 - Temporale
 - Booleano

Tipi Numerici (1)

- **INTEGER**: valori interi
 - Non si possono rappresentare tutti i numeri, solo quelli non troppo piccoli né troppo grandi. Di solito, da $-2^{31}-1$ a 2^{31} .
 - Spesso disponibili varianti come **SMALLINT** o **BIGINT** che supportano intervalli più piccoli o più grandi.
 - Esempio: 12, -45
- **DECIMAL/NUMERIC**: valori non interi, con un numero di cifre dopo la virgola fissato
 - Ottimo per esempio per un campo destinato a contenere dei prezzi, poiché di solito si considerano solo 2 o 3 cifre dopo la virgola
 - Esempio con 2 cifre dopo la virgola: 12.45, -9.10

Tipi Numerici (2)

- **REAL**: valori non interi, in notazione scientifica.
 - Esempio: 3.5E-4 che vuol dire $3.5 \cdot 10^{-4}$ ovvero 0.00035
 - Il numero di cifre significative e l'intervallo di valori possibili per l'esponente è fissato.
 - I calcoli sono approssimati, non esatti e questo può causare spiacevoli conseguenze.
 - Non usare a meno che non sia assolutamente necessario.

Tipi Stringa

- Per **stringa** si intende una sequenza di caratteri
- **CHAR**: stringa a lunghezza fissa
 - Se si tenta di inserire una stringa più corta, viene allungata con spazi
 - Utile per campi come il codice fiscale che hanno una lunghezza costante
- **VARCHAR**: stringa a lunghezza variabile
 - Bisogna comunque specificare un lunghezza massima
 - Meno efficiente (il computer è lento.. per modo di dire.. a trattare questi dati)
 - Utile per campi che possono essere molto corti o molto lunghi (un documento, per esempio)

Altri tipi

- Tipi temporali
 - **DATE**: una data dal 1 Gennaio 99 al 1 Gennaio 9999
 - **TIME**: una ora del giorno, in ore, minuti e secondi
 - **TIMESTAMP**: ora e data assieme
- Tipi booleani
 - **BOOLEAN**: può assumere solo due valori (sì/no, vero/falso, 0/1 a seconda dei casi)

Il valore NULL

- In alcuni casi, il valore di un campo può non essere noto.
 - Per rappresentare un valore non noto si potrebbero usare dei valori particolari per ogni tipi di dato:
 - numero civico: 0
 - codice fiscale: stringa vuota
 - data di nascita: 1/1/9999
- **Possibilità di confusione!**
- Esiste un valore speciale, chiamato **NULL**, che indica proprio la mancanza di dati
 - Quando si specifica un campo, oltre al tipo bisogna indicare se accetta o meno valori NULL.

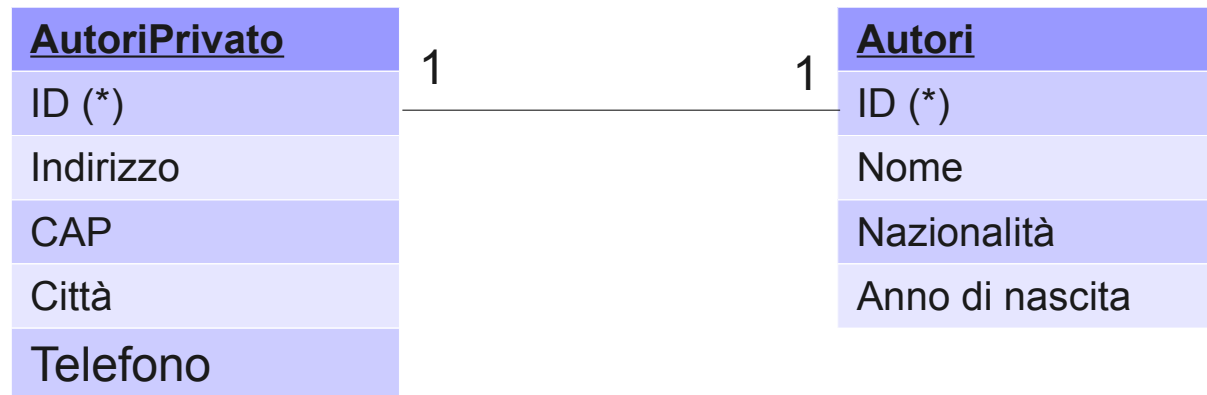
Ancora sulle relazioni tra tabelle

Tipi di relazioni

- Chiavi esterne e chiavi primarie creano delle relazioni
- Ci sono vari tipi di relazione
 - Relazioni uno ad uno
 - Relazioni uno a molti
 - Relazioni molti a molti
- Alcune si possono realizzare direttamente in un database, altre necessitano di **tabelle ausiliari**.
- Nello schema logico si possono annotare le linee tra le tabelle per chiarire di che tipo di relazione si tratti.

Relazioni uno ad uno

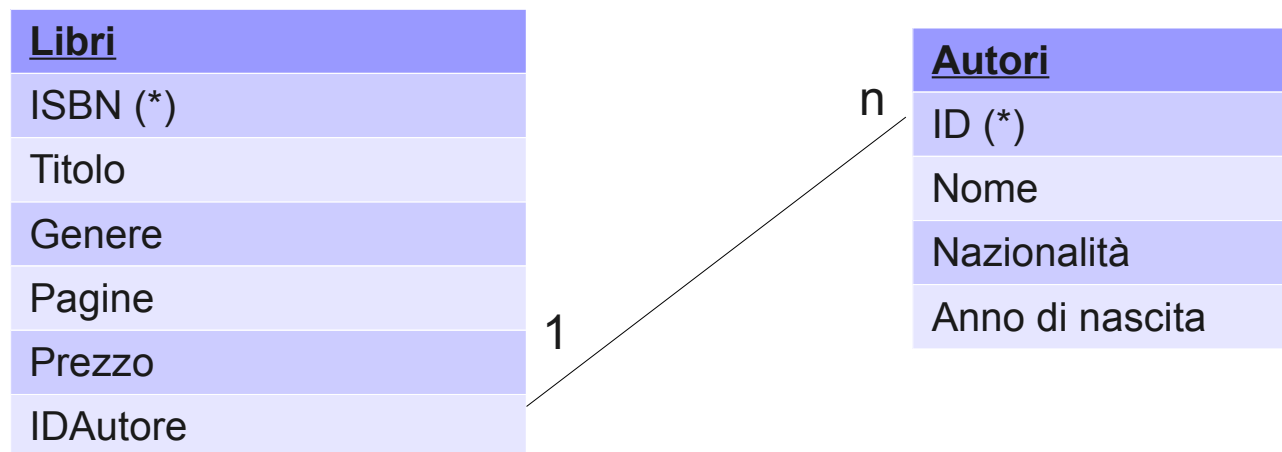
- Ad ogni riga di una tabella corrisponde una sola riga dell'altra e viceversa
 - Esempio: se vogliamo memorizzare dei dati privati degli autori, per dividerli meglio da quelli pubblici possiamo usare una sola tabella.



- Non molto comune.
- La chiave esterna coincide con la chiave primaria.

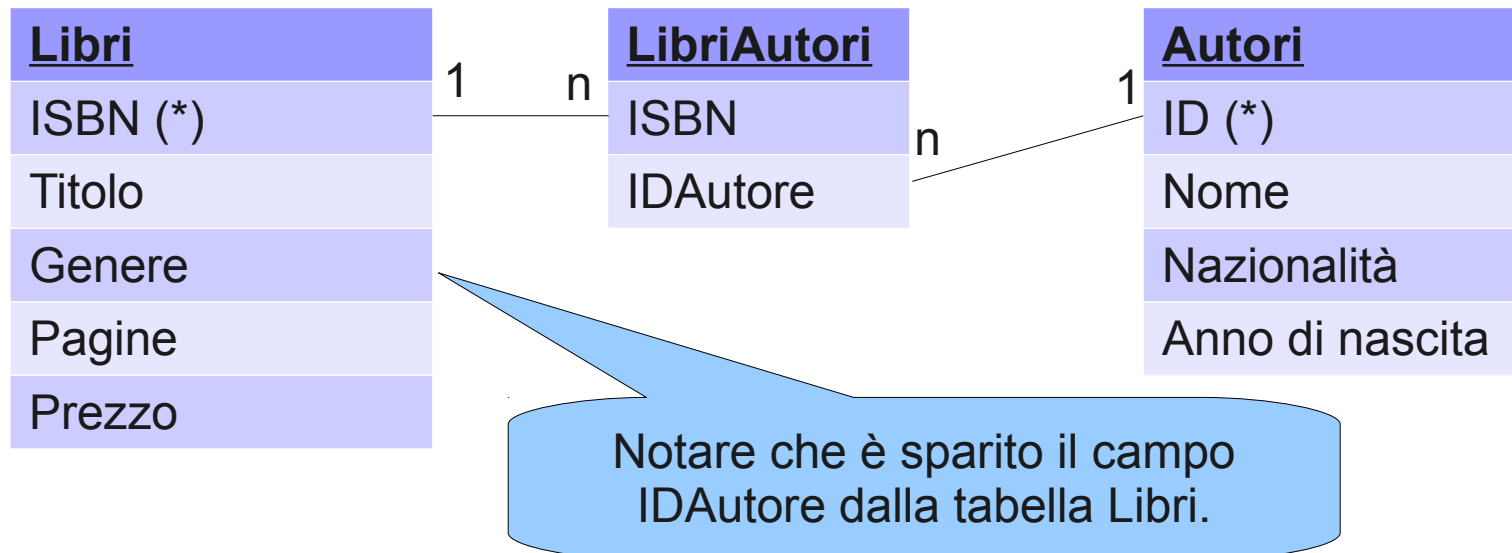
Relazioni uno a molti

- Ad ogni riga della prima tabella corrisponde una riga della seconda tabella, ma nella direzione opposta, ad ogni riga della seconda corrispondono più righe della prima.
- Esempio: la relazione tra autori e libri che abbiamo visto prima. Ad ogni libro corrisponde un autore (ovviamente stiamo semplificando la realtà), ma ad ogni autore corrispondono molti libri.



Relazioni molti a molti (1)

- Ad ogni riga della prima tabella corrispondono più righe della seconda tabella e viceversa.
 - Non si può realizzare direttamente, serve una tabella ausiliaria.
 - Esempio: nel mondo reale, un libro può essere scritto da più autori.



Relazioni molti a molti (2)

- Chi ha scritto il “Libro inesistente” ?

Autori

ID	Nome	Nazionalità	Anno nascita
1	J. K. Rowling	Britannica	1965
2	A. Christie	Britannica	1980
3	N. Ammanniti	Italiana	1966
4	J. K. Rowling	Statunitense	1983

LibriAutori

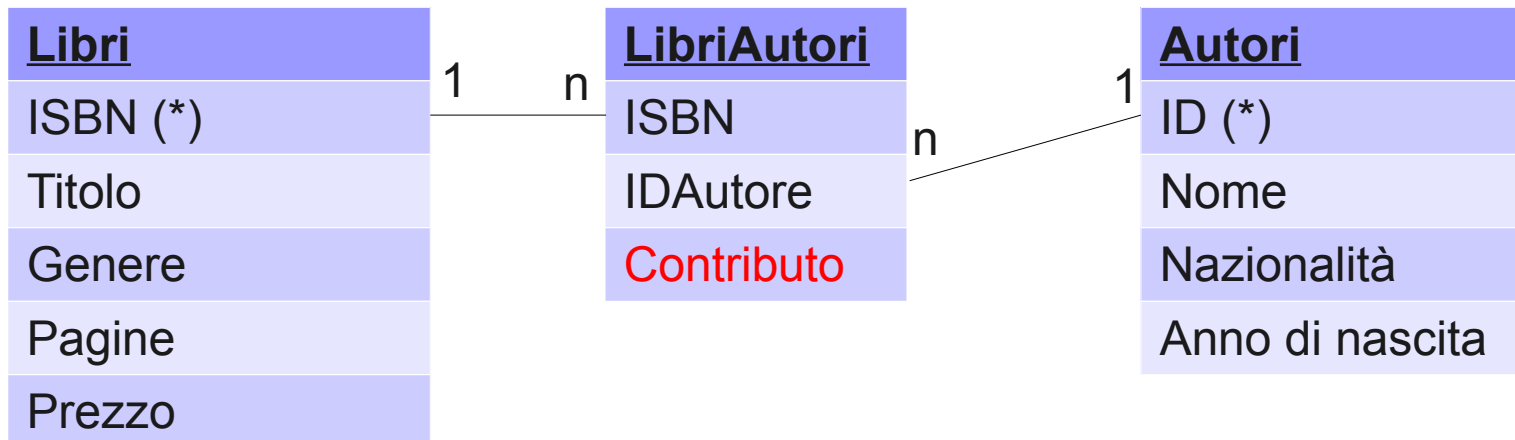
ISBNs	IDAutore
3858-38-4923	1
x1	2
x2	1
x3	3
x3	4

Libri

ISBN	Titolo	Prezzo
3858-38-4923	Harry Potter e la pietra filosofale	16,80
x1	Dieci piccoli indiani	29,90
x2	Harry Potter e il prigioniero di Azkaban	14,50
x3	Libro inesistente	0,4

Dati aggiuntivi di una relazione (1)

- Nel caso di relazioni molti a molti, è possibile specificare dei dati aggiuntivi nella tabella ausiliaria.
 - Utile se ci sono informazioni inerenti proprio la relazione tra le due tabelle principali.
 - Esempio: supponiamo di voler memorizzare, il contributo di ogni autore ad ogni libro che ha scritto.
 - Non ha senso mettere questa informazione né nella tabella autori, né in quella libri.



Dati aggiuntivi di una relazione (2)

Autori

ID	Nome	Nazionalità	Anno nascita
1	J. K. Rowling	Britannica	1965
2	A. Christie	Britannica	1980
3	N. Ammanniti	Italiana	1966
4	J. K. Rowling	Statunitense	1983

LibriAutori

ISBNs	IDAutore	Contributo
3858-38-4923	1	tutto
x1	2	tutto
x2	1	tutto
x3	3	Cap. 1-5
x3	4	Cap 6-12

Libri

ISBN	Titolo	Prezzo
3858-38-4923	Harry Potter e la pietra filosofale	16,80
x1	Dieci piccoli indiani	29,90
x2	Harry Potter e il prigioniero di Azkaban	14,50
x3	Lbro inesistente	0,4

Relazioni gerarchiche

- Talvolta una chiave esterna fa riferimento alla sua stessa tabella.
 - Utile per rappresentare relazione gerarchiche (capo di... , progetto derivato da... , etc..)
 - Esempio: tabella con il personale di una azienda



Integrità referenziale

Relazioni e integrità

- In una coppia chiave esterna – chiave primaria, tutti i valori per la chiave esterna dovrebbero corrispondere a valori validi per la chiave primaria.

ID	Nome	Nazionalità	Anno nascita
1	J. K. Rowling	Britannica	1965
2	A. Christie	Britannica	1980
3	N. Ammanniti	Italiana	1966
4	J. K. Rowling	Statunitense	1983

ISBN	Titolo	Prezzo	ID
3858-38-4923	Harry Potter e la pietra filosofale	16,80	1
xxxxx	Dieci piccoli indiani	29,90	2
xxxxx	Harry Potter e il prigioniero di Azkaban	14,50	1
xxxxx	Io non ho paura	9,50	5

Chi è l'autore di questo libro?

Integrità referenziale

- Per **integrità referenziale** si intende la proprietà delle base di dati “buone” per cui ogni valore per una chiave esterna corrisponde ad un valore della relativa chiave primaria.
- Il DBMS tenta di mantenere la integrità referenziale evitando di effettuare operazioni che la violino.
 - Per far ciò il DBMS deve conoscere quali sono le coppie chiavi esterne – chiavi private.
 - Ogni DBMS ha degli strumenti che servono a istruirlo sulle relazioni presenti nel database.

Cancellazione (1)

- Cosa fare se si vuole cancellare l'autore 4?

ID	Nome	Nazionalità	Anno nascita
1	J. K. Rowling	Britannica	1965
2	A. Christie	Britannica	1980
3	N. Ammanniti	Italiana	1966
4	J. K. Rowling	Statunitense	1982

ISBN	Titolo	Prezzo	IDAutore
3858-38-4923	Harry Potter e la pietra filosofale	16,80	1
xxxxx	Dieci piccoli indiani	29,90	2
xxxxx	Harry Potter e il prigioniero di Azkaban	14,50	1
xxxxx	Io non ho paura	9,50	4



Cancellazione (2)

- 1^a possibilità: annullare l'operazione

ID	Nome	Nazionalità	Anno nascita
1	J. K. Rowling	Britannica	1965
2	A. Christie	Britannica	1980
3	N. Ammanniti	Italiana	1966
4	J. K. Rowling	Statunitense	1983

ISBN	Titolo	Prezzo	IDAutore
3858-38-4923	Harry Potter e la pietra filosofale	16,80	1
xxxxx	Dieci piccoli indiani	29,90	2
xxxxx	Harry Potter e il prigioniero di Azkaban	14,50	1
xxxxx	Io non ho paura	9,50	4

Cancellazione (3)

- 2^a possibilità: **cancellazione a cascata**

ID	Nome	Nazionalità	Anno nascita
1	J. K. Rowling	Britannica	1965
2	A. Christie	Britannica	1980
3	N. Ammanniti	Italiana	1966
4	J. K. Rowling	Statunitense	1982

ISBN	Titolo	Prezzo	IDAutore
3858-38-4923	Harry Potter e la pietra filosofale	16,80	1
xxxxx	Dieci piccoli indiani	29,90	2
xxxxx	Harry Potter e il prigioniero di Azkaban	14,50	1
xxxxx	Io non ho paura	9,50	4

Cancellazione (4)

- 3^a soluzione: impostare a NULL i corrispondenti valori delle chiavi esterne (se il valore NULL è ammesso)

ID	Nome	Nazionalità	Anno nascita
1	J. K. Rowling	Britannica	1965
2	A. Christie	Britannica	1980
3	N. Ammanniti	Italiana	1966
4	J. K. Rowling	Statunitense	1965

ISBN	Titolo	Prezzo	IDAutore
3858-38-4923	Harry Potter e la pietra filosofale	16,80	1
xxxxx	Dieci piccoli indiani	29,90	2
xxxxx	Harry Potter e il prigioniero di Azkaban	14,50	1
xxxxx	Io non ho paura	9,50	NULL

Cancellazione (5)

- Notare che non c'è problema a cancellare una riga dal lato “molti” della relazione.

ID	Nome	Nazionalità	Anno nascita
1	J. K. Rowling	Britannica	1965
2	A. Christie	Britannica	1980
3	N. Ammanniti	Italiana	1966
4	J. K. Rowling	Statunitense	1983

ISBN	Titolo	Prezzo	IDAutore
3858-38-4923	Harry Potter e la pietra filosofale	16,80	1
xxxxx	Dieci piccoli indiani	29,90	2
xxxxx	Harry Potter e il prigioniero di Azkaban	14,50	1
xxxxx	Io non ho paura	9,50	1

Modifiche (1)

- Problemi analoghi nel caso di modifiche della chiave primaria

4 rimpiazzato con 5

ID	Nome	Nazionalità	Anno nascita
1	J. K. Rowling	Britannica	1965
2	A. Christie	Britannica	1980
3	N. Ammanniti	Italiana	1966
5	J. K. Rowling	Statunitense	1983

ISBN	Titolo	Prezzo	IDAutore
3858-38-4923	Harry Potter e la pietra filosofale	16,80	1
xxxxx	Dieci piccoli indiani	29,90	2
xxxxx	Harry Potter e il prigioniero di Azkaban	14,50	1
xxxxx	Io non ho paura	9,50	4



Modifiche (2)

- Soluzioni
 - Annullare l'operazione
 - Effettuare le modifiche in cascata sulla chiave esterna
 - Mettere a NULL i valori corrispondenti per la chiave esterna.