



Decomposing a table

Splitting up a table into tables -
towards normalization



Too much information

Consider a table for books:

```
sqlite> .schema book
CREATE TABLE book(title TEXT, author TEXT, publisher TEXT,
                   isbn TEXT PRIMARY KEY NOT NULL);
```

```
sqlite> SELECT * from book;
```

title	author	publisher	isbn
Databases	Manda Tory	IT-Literature Inc	0000001
SQL	Manda Tory	IT-Literature Inc	0000002
Java	D. Fault Values	IT-Literature Inc	0000003
More Java	M.T. Objects	IT-Literature Inc	0000004
Swedish Flowers	Lynn E. Uss	Biology Books AB	0000005
Java direkt med Swin	Jan Skansholm	Studentlitteratur	978914410431
Databasteknik	Thomas Padron-Mccart	Studentlitteratur	978914404449
Programming in Java	Henrik and Rikard	Juneday	1234567

Too much information

- What if we want to store information about a new publisher (without any books yet)?
- What if we stop selling some books from a publisher, what to do?

```
sqlite> .schema book
CREATE TABLE book(title TEXT, author TEXT, publisher TEXT,
                   isbn TEXT PRIMARY KEY NOT NULL);

sqlite> INSERT INTO book(....???) -- add new publisher
sqlite> UPDATE book SET ....???. -- remove books but keep publisher
```

Duplicate information

- How is the publisher IT Literature Inc stored?

```
sqlite> SELECT * FROM book WHERE publisher = 'IT-Literature Inc';
title          author          publisher       isbn
-----
Databases      Manda Tory      IT-Literature  0000001
SQL            Manda Tory      IT-Literature  0000002
Java           D. Fault Values IT-Literature  0000003
More Java      M.T. Objects    IT-Literature  0000004

-- The publisher's name is repeated four times...
```

Rule of thumb

- One type of “thing” per table - The book table should be about books
- One such thing per row - One book per row
- One row per thing - The “publisher” is actually in four rows...

title	author	publisher	isbn
Databases	Manda Tory	IT-Literature Inc	0000001
SQL	Manda Tory	IT-Literature Inc	0000002
Java	D. Fault Values	IT-Literature Inc	0000003
More Java	M.T. Objects	IT-Literature Inc	0000004

See for instance <http://www.databasteknik.se/webbkursen/normalisering/> (Swedish)

We can't store new publishers without books

- One type of “thing” per table - The book table contains also publishers
- One such thing per row
- One row per thing - A publisher needs at least one book...

title	author	publisher	isbn
Databases	Manda Tory	IT-Literature Inc	0000001
SQL	Manda Tory	IT-Literature Inc	0000002
Java	D. Fault Values	IT-Literature Inc	0000003
More Java	M.T. Objects	IT-Literature Inc	0000004
		Infological Books Ltd	

We need a table for publishers, if we want to be able to store information about publishers apart from books

We have to spell publisher's name correctly

- Adding more than one book requires careful spelling of the publisher

title	author	publisher	isbn	
Databases	Manda Tory	IT-Literature Inc	0000001	
SQL	Manda Tory	IT-Literature Inc	0000002	
Java	D. Fault Values	IT-Literature Inc	0000003	
More Java	M.T. Objects	IT-Litterature Inc	0000004	--OOPS!
		Infological Books Ltd		

We need a table for publishers, to save the name about a publisher on one row (and spell it correctly once)

Create a publisher table

- One type of “thing” per table - The table should be about publishers
- One such thing per row - One publisher per row
- One row per thing - One publisher per row as well
- We can store/keep publishers without books!

```
CREATE TABLE publisher(publisher_id INTEGER PRIMARY KEY NOT NULL,  
                        name TEXT UNIQUE NOT NULL);
```

```
sqlite> SELECT * FROM publisher;
```

publisher_id	name
1	Studentlitteratur
2	Juneday
3	Mayday! Mayday!
4	Oh Really
5	IT-Literature Inc
6	Biology Books AB

Change the book table - reference publisher

- A book still “has” a publisher, only we refer to it by the ID of the other tbl
- The publisher_id column is called a “foreign key”
- The column in the foreign table must have unique values - Why?

```
CREATE TABLE book (title TEXT, author TEXT, isbn TEXT PRIMARY KEY NOT NULL,  
                    publisher_id INTEGER);
```

```
sqlite> SELECT * FROM book;
```

title	author	isbn	publisher_id
Java direkt med Swing	Jan Skansholm	9789144104317	1
Databasteknik	Thomas Padron-Mccarthy	9789144044491	1
Programming in Java	Henrik and Rikard	1234567	2
Swedish Flowers	Lynn E. Uss	0000005	6
More Java	M.T. Objects	0000004	5
Java	D. Fault Values	0000003	5
SQL	Manda Tory	0000002	5
Databases	Manda Tory	0000001	5

Retrieving book info including publisher

- We have to JOIN the two tables
- Both tables have column "publisher_id" - NATURAL JOIN is possible

```
sqlite> SELECT title, author, isbn, name AS publisher FROM  
...> book NATURAL JOIN publisher;
```

title	author	isbn	publisher
Java direkt med Swing	Jan Skansholm	9789144104317	Studentlitteratur
Databasteknik	Thomas Padron-Mccarthy	9789144044491	Studentlitteratur
Programming in Java	Henrik and Rikard	1234567	Juneday
Swedish Flowers	Lynn E. Uss	0000005	Biology Books AB
More Java	M.T. Objects	0000004	IT-Literature Inc
Java	D. Fault Values	0000003	IT-Literature Inc
SQL	Manda Tory	0000002	IT-Literature Inc
Databases	Manda Tory	0000001	IT-Literature Inc

We still have duplicate authors

- We do the same with the author column - move it to its own table
- Authors “know” nothing about books - one thing per table, per row

```
CREATE TABLE author(author_id INTEGER PRIMARY KEY NOT NULL, name TEXT NOT NULL);
```

```
sqlite> SELECT * FROM author;
```

author_id	name
1	Jan Skansholm
2	Thomas Padron-Mccarthy
3	Henrik and Rikard
4	Lynn E. Uss
5	M.T. Objects
6	D. Fault Values
7	Manda Tory
8	Selma Lagerlöf

Change book table - refer also to author table

- author → author_id
- author_id FK in book, must be unique in author table

```
CREATE TABLE book(title TEXT, author_id INTEGER, isbn TEXT PRIMARY KEY NOT NULL,  
                  publisher_id INTEGER);
```

```
sqlite> SELECT * FROM book;
```

title	author_id	isbn	publisher_id
Java direkt med Swing	1	9789144104317	1
Databasteknik	2	9789144044491	1
Programming in Java	3	1234567	2
Swedish Flowers	4	0000005	6
More Java	5	0000004	5
Java	6	0000003	5
SQL	7	0000002	5
Databases	7	0000001	5

Combining the three tables

- author → author_id, publisher → publisher_id
- author_id, publisher_id FKs in book, must be unique in other tables

```
CREATE TABLE book(title TEXT, author_id INTEGER, isbn TEXT PRIMARY KEY NOT NULL,  
                  publisher_id INTEGER);
```

```
sqlite> SELECT title, author.name AS author, isbn, publisher.name AS publisher FROM  
...> book NATURAL JOIN author JOIN publisher  
...> ON publisher.publisher_id = book.publisher_id;
```

title	author	isbn	publisher
Java direkt med Swing	Jan Skansholm	9789144104317	Studentlitteratur
Databasteknik	Thomas Padron-Mccarthy	9789144044491	Studentlitteratur
Programming in Java	Henrik and Rikard	1234567	Juneday
Swedish Flowers	Lynn E. Uss	0000005	Biology Books AB
More Java	M.T. Objects	0000004	IT-Literature Inc
Java	D. Fault Values	0000003	IT-Literature Inc
SQL	Manda Tory	0000002	IT-Literature Inc
Databases	Manda Tory	0000001	IT-Literature Inc

Combining the three tables (alternative version)

- author → author_id
- author_id FK in book, must be unique in author table

```
CREATE TABLE book(title TEXT, author_id INTEGER, isbn TEXT PRIMARY KEY NOT NULL,  
                  publisher_id INTEGER);
```

```
sqlite> SELECT title, isbn, publisher, author FROM book  
...> NATURAL JOIN (SELECT publisher_id, name AS publisher from publisher)  
...> NATURAL JOIN (SELECT author_id, name AS author FROM author);
```

title	author	isbn	publisher
Java direkt med Swing	Jan Skansholm	9789144104317	Studentlitteratur
Databasteknik	Thomas Padron-Mccarthy	9789144044491	Studentlitteratur
Programming in Java	Henrik and Rikard	1234567	Juneday
Swedish Flowers	Lynn E. Uss	0000005	Biology Books AB
More Java	M.T. Objects	0000004	IT-Literature Inc
Java	D. Fault Values	0000003	IT-Literature Inc
SQL	Manda Tory	0000002	IT-Literature Inc
Databases	Manda Tory	0000001	IT-Literature Inc

Pause - part 1

Alternative design (bonus info)

- book is about the book title and isbn
- author is about the author name
- publisher is about the publisher name
- book_publisher is about who publishes what book(s)
- book_author is about who authored what book

```
sqlite> .schema book
CREATE TABLE book(title TEXT, isbn TEXT PRIMARY KEY NOT NULL);
sqlite> .schema book_author
CREATE TABLE book_author(isbn TEXT, author_id INTEGER);
sqlite> .schema book_publisher
CREATE TABLE book_publisher(isbn TEXT, publisher_id INTEGER);
```


Query gets a little more complicated, though

```
sqlite> SELECT title, book.isbn, a.name AS author_name, publisher.name AS pub FROM book
NATURAL JOIN book_author NATURAL JOIN author a NATURAL JOIN book_publisher JOIN publisher
ON book_publisher.publisher_id = publisher.publisher_id;
```

title	isbn	author_name	pub
Java direkt med Swing	9789144104317	Jan Skansholm	Studentlitteratur
Databasteknik	9789144044491	Thomas Padron-	Studentlitteratur
Programming in Java	1234567	Henrik and Rik	Juneday
Swedish Flowers	0000005	Lynn E. Uss	Biology Books AB
More Java	0000004	M.T. Objects	IT-Literature Inc
Java	0000003	D. Fault Value	IT-Literature Inc
SQL	0000002	Manda Tory	IT-Literature Inc
Databases	0000001	Manda Tory	IT-Literature Inc

--OR (handle the fact that there are two "name" columns):

```
sqlite> SELECT title, isbn, name AS author, publisher FROM book
...> NATURAL JOIN book_author NATURAL JOIN author a NATURAL JOIN book_publisher
...> NATURAL JOIN (SELECT publisher_id, name AS publisher FROM publisher);
```

We can always create a “view” (bonus info!)

```
sqlite> CREATE VIEW ALL_ABOUT_BOOK AS
...> SELECT title, book.isbn, a.name AS author_name, publisher.name as pub FROM book
...> NATURAL JOIN book_author NATURAL JOIN author a NATURAL JOIN book_publisher
...> JOIN publisher ON book_publisher.publisher_id = publisher.publisher_id;
```

```
sqlite> SELECT * FROM ALL_ABOUT_BOOK;
```

title	isbn	author_name	pub
Java direkt med Swing	9789144104317	Jan Skansholm	Studentlitteratur
Databasteknik	9789144044491	Thomas Padron-	Studentlitteratur
Programming in Java	1234567	Henrik and Rik	Juneday
Swedish Flowers	0000005	Lynn E. Uss	Biology Books AB
More Java	0000004	M.T. Objects	IT-Literature Inc
Java	0000003	D. Fault Value	IT-Literature Inc
SQL	0000002	Manda Tory	IT-Literature Inc
Databases	0000001	Manda Tory	IT-Literature Inc

Pause part 2

Another alternative (bonus info!)

```
-- We can make a table of the connection between books, authors and publishers
CREATE TABLE publication(isbn TEXT, publisher_id INTEGER, author_id INTEGER);
```

```
sqlite> SELECT isbn, title, publisher, author FROM book
...> NATURAL JOIN publication
...> NATURAL JOIN (SELECT publisher_id, name AS publisher FROM publisher)
...> NATURAL JOIN (SELECT author_id, name AS author FROM author);
```

isbn	title	publisher	author
9789144104317	Java direkt med Swing	Studentlitteratur	Jan Skansholm
1234567	Programming in Java	Juneday	Henrik and Rikard
0000005	Swedish Flowers	Biology Books AB	Lynn E. Uss
0000004	More Java	IT-Literature Inc	M.T. Objects
0000003	Java	IT-Literature Inc	D. Fault Values
0000002	SQL	IT-Literature Inc	Manda Tory
0000001	Databases	IT-Literature Inc	Manda Tory
9789144044491	Databasteknik	Studentlitteratur	Thomas Padron-Mccarthy

Another alternative (bonus info!)

```
-- This makes for a nice model - book, author, publisher and publication
CREATE TABLE book(title TEXT, isbn TEXT PRIMARY KEY NOT NULL);
CREATE TABLE author(author_id INTEGER PRIMARY KEY NOT NULL, name TEXT NOT NULL);
CREATE TABLE publisher(publisher_id INTEGER PRIMARY KEY NOT NULL, name TEXT NOT NULL);
CREATE TABLE publication(isbn TEXT, publisher_id INTEGER, author_id INTEGER);
```

```
sqlite> SELECT isbn, title, publisher, author FROM book
...> NATURAL JOIN publication
...> NATURAL JOIN (SELECT publisher_id, name AS publisher FROM publisher)
...> NATURAL JOIN (SELECT author_id, name AS author FROM author);
```

isbn	title	publisher	author
9789144104317	Java direkt med Swing	Studentlitteratur	Jan Skansholm
1234567	Programming in Java	Juneday	Henrik and Rikard
0000005	Swedish Flowers	Biology Books AB	Lynn E. Uss
0000004	More Java	IT-Literature Inc	M.T. Objects
0000003	Java	IT-Literature Inc	D. Fault Values
0000002	SQL	IT-Literature Inc	Manda Tory
0000001	Databases	IT-Literature Inc	Manda Tory
9789144044491	Databasteknik	Studentlitteratur	Thomas Padron-Mccarthy

Further reading

- <http://www.databasteknik.se/webbkursen/er/index.html> (Swedish)
- <http://www.databasteknik.se/webbkursen/er2relationer/index.html> (Swedish)
- <http://www.databasteknik.se/webbkursen/normalisering/index.html> (Swedish)