

Assignment 7

Normalization

Introduction

Some of the most difficult decisions that you face as a database developer are what tables to create and what columns to place in each table, as well as how to relate the tables that you create. Normalization is the process of applying a series of rules to ensure that your database achieves optimal structure. Normal forms are a progression of these rules. Each successive normal form achieves a better database design than the previous form did. Although there are several levels of normal forms, we mainly discussed about Boyce Codd Normal Form and 4th Normal Form.

First Normal Form (1NF)

To achieve first normal form the data must be in the form of a relation. The following are the criteria for a relation:

- Rows contain data about a single entity
- Columns contain data about attributes of the entity
- Cells of the table hold a single value
- All entries in a column are of the same kind
- Each column has a unique name
- The order of the columns is unimportant
- The order of the rows is unimportant
- No two rows may be identical

1NF Exercise

Bubba runs a small video store and decides to implement a database system. Bubba's view of the business is that he has customers and rentals. Customers usually rent 1 to 3 videos. His database design uses two tables, *Customers* and *Rentals* as shown below.

CustID	CustName	CustCity	CustState	CustZip
1	John Doe	Pittsburgh	PA	15136
2	Matt Smith	Columbia	SC	29210
3	Melinda Black	Stevensville	MD	21666
4	Tom Johnson	Annapolis	MD	21210

RentalID	CustID	CheckOutDate	Title1	Title2	Title3
1	1	3/3/2004	Die Hard	The last man standing	Wedding Crashers
2	2	3/4/2005	Dodgeball	Die Hard	Forest Gump
3	3	6/7/2006	As good as it gets	The thin red line	The Bachelor
4	1	6/7/2006	Forest Gump	The Godfather	Saving private Ryan

```
create table Customers(CustID integer IDENTITY(1,1),
    CustName char(30) not null,
    CustCity char(30) not null,
    CustState char(2) not null,
```

```

    CustZip integer not null,
    primary key (CustID));

create table Rentals(RentalID integer IDENTITY(1,1),
    CustID integer null,
    CheckOutDate datetime not null,
    Title1 char(50) not null,
    Title2 char(50) null,
    Title3 char(50) null,
    primary key (RentalID),
    foreign key (CustID) references Customers(CustID)
        on delete no action on update cascade);

insert into Customers values('John Doe', 'Pittsburgh', 'PA', 15136);
insert into Customers values('Matt Smith', 'Columbia', 'SC', 29210);
insert into Customers values('Melinda Black', 'Setevensville', 'MD', 21666);
insert into Customers values('Tom Johnson', 'Annapolis', 'MD', 21210);

insert into Rentals values(1, '3/3/2004', 'Die Hard', 'The last man standing', 'Wedding
Crashers');
insert into Rentals values(2, '3/4/2005', 'Dodgeball', 'Die Hard', 'Forest Gump');
insert into Rentals values(3, '6/7/2006', 'As good as it gets', 'The thin red line', 'The
Bachelor');
insert into Rentals values(1, '6/7/2006', 'Forest Gump', 'The Godfather', 'Saving private
Ryan');

```

1. Bubba wants to find all the customers who rented the ‘Die Hard’ movie in order to send them information about other Bruce Willis movies. Write an SQL query to list all customers who rented the ‘Die Hard’ movie.
 - What makes this query cumbersome?
 - Is it easy to determine how many times Die Hard was rented?
2. What do you do if a customer wants to rent 4 movies (or more)?
3. The difficulty with accomplishing these simple queries and tasks is the “Rentals” table is not in 1NF. Title1, Title2, and Title3 are unique names for column headings, but really describe the same thing: a “Movie”. Redesign the Rentals table (new name Rentals_2) so that it meets the criteria of 1NF with no repeating attributes. What is the primary key of the new table?
4. Write the SQL statement to create the Rentals_2 table in PostgreSQL.
5. Write SQL statements to insert all the data from Rentals table in the new Rentals_2 table. (remember bulk insert)

6. Re-write and rerun query 1. You will see that query1 is now easy to write and the problem with a customer renting a variable number of movies is eliminated.

Boyce-Codd Normal Form (BCNF)

To achieve Boyce-Codd normal form, all the determinants have to be candidate keys (either primary key or alternate key). (If you do not understand functional dependency then review functional dependency starting on page 74 in the textbook). Below are tables with a slight variation of the video store. The two tables meet 1NF requirements.

Customers:

CustID	CustName	CustCity	CustState	CustZip
1	John Doe	Pittsburgh	PA	15136
2	Matt Smith	Columbia	SC	29210
3	Melinda Black	Stevensville	MD	21666
4	Tom Johnson	Annapolis	MD	21210

Rentals3

RentalID	Title	CustID	CheckOutDate	Director	ReleaseType	Price
1	Die Hard	1	3/3/2004	John McTiernan	Old	\$4.25
1	The last man standing	1	3/3/2004	Walter Hill	Old	\$4.25
1	Wedding Crashers	1	3/3/2004	David Dobkin	New	\$5.50
2	Dodgeball	2	3/4/2005	Rawson Marshall Thurber	New	\$5.50
2	Die Hard	2	3/4/2005	John McTiernan	Old	\$4.25
3	As good as it gets	3	6/7/2006	James Brooks	Old	\$4.25
4	Forest Gump	1	6/7/2006	Robert Zemeckis	Old	\$4.25

The primary key in the Rentals3 table is the composite key (***RentalID, Title***).

7. Find the functional dependencies for Rentals3.

8. Is Rentals3 in BCNF? The answer is No, but you should justify it.

9. What problems are manifested by not having the tables in BCNF? The text discusses ***insertion anomalies, deletion anomalies and modification anomalies***. Give one example of each of these anomalies in the above table.

10. Decompose the Rentals3 table into BCNF.

4NF

Suppose we have the following Courses table with columns Course, Teacher, Book. The meaning of a row (C, T, B) in the table is that teacher T can teach course C and B is a recommended book for course C. The recommended book is independent of teacher. Here is an example of instantiation for this table:

Course	Teacher	Book
IT420	Crainiceanu	Kroenke
IT420	Crainiceanu	Ramakrishnan
IT420	McKenna	Kroenke
IT420	McKenna	Ramakrishnan
SI440	Crainiceanu	Kroenke
SI440	Crainiceanu	Ramakrishnan
SI440	Crainiceanu	Stonebraker

11. What is a multivalued dependency?

12. Is the above table in 4th normal form?

13. If answer to 12 is yes, say why. If not, transform the table to be in 4th normal form.

Normalization: yes or no?

14. Assuming that city and state determine zip code, is Customers table in BCNF?

15. If Customers table is not in BCNF, would you or would you not normalize it to BCNF? Give at least one reason for the choice you make.