

COPENHAGEN BUSINESS ACADEMY



- Normalisation

Why Normalisation

- Redundant data from update anomalies results introduces inconsistencies into a database.
- The rules associated with the most commonly used normal forms, namely first (1NF), second (2NF), and third (3NF)
- The identification of various types of update anomalies such as insertion, deletion, and modification anomalies can be found when tables that break the rules of 1NF, 2NF, and 3NF
- Technique for producing a set of tables with desirable properties that support the requirements of a user or company
- Aim of relational database design is to group columns into tables to minimize data redundancy and reduce file storage space

Data redundancy and update anomalies

- Types of update anomalies include
 - Insertion – when you have to add extra data
 - Deletion – when deleting a row inadvertently causes other data to be deleted
 - Modification – when you have to change multiple rows

Data redundancy and update anomalies

empno	ename	job	mgr	hiredate	sal	deptno	dname	loc
7369	SMITH	CLERK	7902	12/17/1980	800	20	RESEARCH	DALLAS
7499	ALLEN	SALESMAN	7698	02/20/1981	1600	30	SALES	CHICAGO
7521	WARD	SALESMAN	7698	02/22/1981	1250	30	SALES	CHICAGO
7566	JONES	MANAGER	7839	04-02-1981	2975	20	RESEARCH	DALLAS
7654	MARTIN	SALESMAN	7698	09/28/1981	1250	30	SALES	CHICAGO
7698	BLAKE	MANAGER	7839	05-01-1981	2850	30	SALES	CHICAGO
7782	CLARK	MANAGER	7839	06-09-1981	2450	10	ACCOUNTING	NEW YORK
7788	SCOTT	ANALYST	7566	04/19/1987	3000	20	RESEARCH	DALLAS
7839	KING	PRESIDENT		11/17/1981	5000	10	ACCOUNTING	NEW YORK
7844	TURNER	SALESMAN	7698	09-08-1981	1500	30	SALES	CHICAGO
7876	ADAMS	CLERK	7788	05/23/1987	1100	20	RESEARCH	DALLAS
7900	JAMES	CLERK	7698	12-03-1981	950	30	SALES	CHICAGO
7902	FORD	ANALYST	7566	12-03-1981	3000	20	RESEARCH	DALLAS
7934	MILLER	CLERK	7782	01/23/1982	1300	10	ACCOUNTING	NEW YORK

- *Problem might occur due to redundancy (anomalies)*
 - Insert a new employee in research dept.
 - Insert a new department
 - Delete the last employee in a department
 - Change location for a dept.

Data redundancy and update anomalies

empno	ename	job	mgr	hiredate	sal	deptno
7369	SMITH	CLERK	7902	12/17/1980	800	20
7499	ALLEN	SALESMAN	7698	02/20/1981	1600	30
7521	WARD	SALESMAN	7698	02/22/1981	1250	30
7566	JONES	MANAGER	7839	04-02-1981	2975	20
7654	MARTIN	SALESMAN	7698	09/28/1981	1250	30
7698	BLAKE	MANAGER	7839	05-01-1981	2850	30
7782	CLARK	MANAGER	7839	06-09-1981	2450	10
7788	SCOTT	ANALYST	7566	04/19/1987	3000	20
7839	KING	PRESIDENT		11/17/1981	5000	10
7844	TURNER	SALESMAN	7698	09-08-1981	1500	30
7876	ADAMS	CLERK	7788	05/23/1987	1100	20
7900	JAMES	CLERK	7698	12-03-1981	950	30
7902	FORD	ANALYST	7566	12-03-1981	3000	20
7934	MILLER	CLERK	7782	01/23/1982	1300	10

deptno	dname	loc
10	ACCOUNTING	NEW YORK
20	RESEARCH	DALLAS
30	SALES	CHICAGO
40	OPERATIONS	BOSTON

Data redundancy and update anomalies

*Give examples of
potential anomalies!*

staffNo	name	position	salary	branchNo	branchAddress	telNo
S1500	Tom Daniels	Manager	46000	B001	8 Jefferson Way, Portland, OR 97201	503-555-3618
S0003	Sally Adams	Assistant	30000	B001	8 Jefferson Way, Portland, OR 97201	503-555-3618
S0010	Mary Martinez	Manager	50000	B002	City Center Plaza, Seattle, WA 98122	206-555-6756
S3250	Robert Chin	Supervisor	32000	B002	City Center Plaza, Seattle, WA 98122	206-555-6756
S2250	Sally Stern	Manager	48000	B004	16 – 14th Avenue, Seattle, WA 98128	206-555-3131
S0415	Art Peters	Manager	41000	B003	14 – 8th Avenue, New York, NY 10012	212-371-3000

Data redundancy and update anomalies

Give examples of potential anomalies!

staffNo	name	position	salary	branchNo	branchAddress	telNo
S1500	Tom Daniels	Manager	46000	B001	8 Jefferson Way, Portland, OR 97201	503-555-3618
S0003	Sally Adams	Assistant	30000	B001	8 Jefferson Way, Portland, OR 97201	503-555-3618
S0010	Mary Martinez	Manager	50000	B002	City Center Plaza, Seattle, WA 98122	206-555-6756
S3250	Robert Chin	Supervisor	32000	B002	City Center Plaza, Seattle, WA 98122	206-555-6756
S2250	Sally Stern	Manager	48000	B004	16 – 14th Avenue, Seattle, WA 98128	206-555-3131
S0415	Art Peters	Manager	41000	B003	14 – 8th Avenue, New York, NY 10012	212-371-3000

- Insert new staff we must add branch details e.g if it is for B002 and it must be consistent with other records
- Insert a new branchNo that may not have any staff yet. It has to be filled up with null value

Data redundancy and update anomalies

Give examples of potential anomalies!

staffNo	name	position	salary	branchNo	branchAddress	telNo
S1500	Tom Daniels	Manager	46000	B001	8 Jefferson Way, Portland, OR 97201	503-555-3618
S0003	Sally Adams	Assistant	30000	B001	8 Jefferson Way, Portland, OR 97201	503-555-3618
S0010	Mary Martinez	Manager	50000	B002	City Center Plaza, Seattle, WA 98122	206-555-6756
S3250	Robert Chin	Supervisor	32000	B002	City Center Plaza, Seattle, WA 98122	206-555-6756
S2250	Sally Stern	Manager	48000	B004	16 – 14th Avenue, Seattle, WA 98128	206-555-3131
S0415	Art Peters	Manager	41000	B003	14 – 8th Avenue, New York, NY 10012	212-371-3000

- Where would deletion anomalies occur?

Data redundancy and update anomalies

Give examples of potential anomalies!

staffNo	name	position	salary	branchNo	branchAddress	telNo
S1500	Tom Daniels	Manager	46000	B001	8 Jefferson Way, Portland, OR 97201	503-555-3618
S0003	Sally Adams	Assistant	30000	B001	8 Jefferson Way, Portland, OR 97201	503-555-3618
S0010	Mary Martinez	Manager	50000	B002	City Center Plaza, Seattle, WA 98122	206-555-6756
S3250	Robert Chin	Supervisor	32000	B002	City Center Plaza, Seattle, WA 98122	206-555-6756
S2250	Sally Stern	Manager	48000	B004	16 – 14th Avenue, Seattle, WA 98128	206-555-3131
S0415	Art Peters	Manager	41000	B003	14 – 8th Avenue, New York, NY 10012	212-371-3000

- Modification what will be the issue if we change the telephone number of B001
- Inconsistency if we don't manage to change all the telNo of B001

Solution

Staff

staffNo	name	position	salary	branchNo
S1500	Tom Daniels	Manager	46000	B001
S0003	Sally Adams	Assistant	30000	B001
S0010	Mary Martinez	Manager	50000	B002
S3250	Robert Chin	Supervisor	32000	B002
S2250	Sally Stern	Manager	48000	B004
S0415	Art Peters	Manager	41000	B003

Branch

branchNo	branchAddress	telNo
B001	8 Jefferson Way, Portland, OR 97201	503-555-3618
B002	City Center Plaza, Seattle, WA 98122	206-555-6756
B003	14 – 8th Avenue, New York, NY 10012	212-371-3000
B004	16 – 14th Avenue, Seattle, WA 98128	206-555-3131

Normalization in database

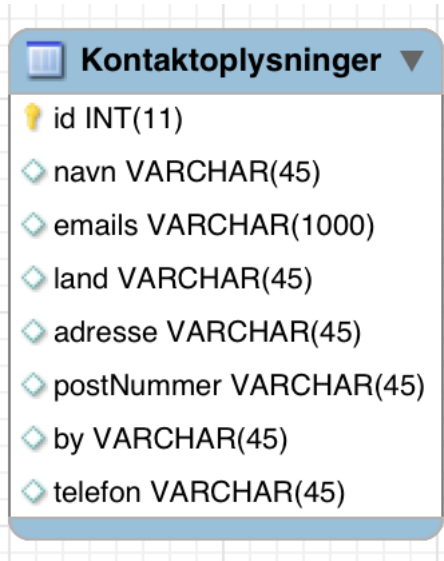
- Abstract concept
- Quality control for the design of the database
 - apply a set of rules called normal forms
- Set of checklist to make sure your database has
 - reduce data redundancy/minimum duplication
 - improved data integrity

First Normal Form(1NF)

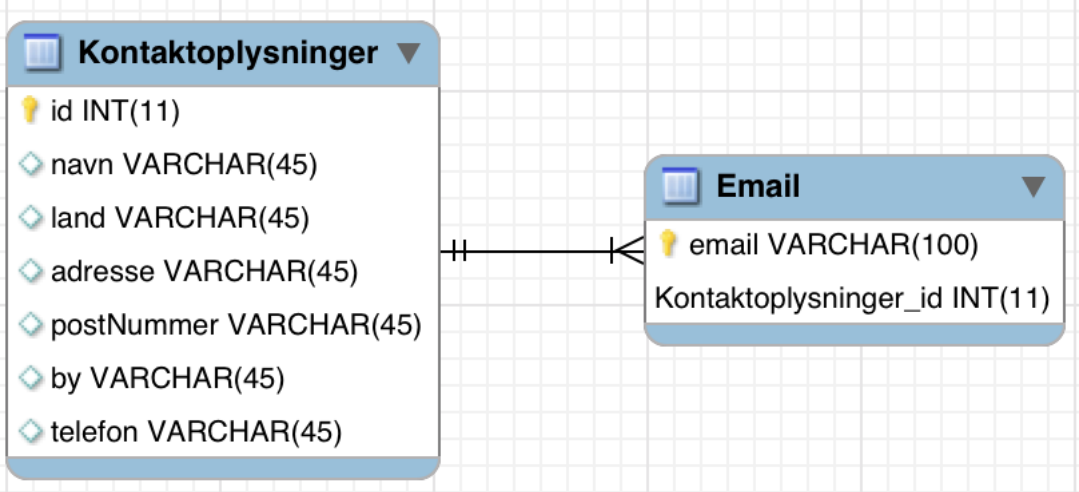
- No repeating groups
- Data values are atomic, each column in each row can store only one value
- The table has a primary key
- The order of the row does not matter

Normalisation of the 1st Normal Form

If you have the following table:



You can have more than one email address.



Branch table is not in 1NF

branchNo	branchAddress	telNos
B001	8 Jefferson Way, Portland, OR 97201	503-555-3618, 503-555-2727, 503-555-6534
B002	City Center Plaza, Seattle, WA 98122	206-555-6756, 206-555-8836
B003	14 – 8th Avenue, New York, NY 10012	212-371-3000
B004	16 – 14th Avenue, Seattle, WA 98128	206-555-3131, 206-555-4112

↑
Primary key

More than
one value, so
not in 1NF

Converting Branch table to 1NF

Branch (Not 1NF)

branchNo	branchAddress	telNos
B001	8 Jefferson Way, Portland, OR 97201	503-555-3618, 503-555-2727, 503-555-6534
B002	City Center Plaza, Seattle, WA 98122	206-555-6756, 206-555-8836
B003	14 – 8th Avenue, New York, NY 10012	212-371-3000
B004	16 – 14th Avenue, Seattle, WA 98128	206-555-3131, 206-555-4112

Take copy of branchNo column to new table to become foreign key

Remove telNos column and create new column called telNo in the new table

Branch (1NF)

branchNo	branchAddress
B001	8 Jefferson Way, Portland, OR 97201
B002	City Center Plaza, Seattle, WA 98122
B003	14 – 8th Avenue, New York, NY 10012
B004	16 – 14th Avenue, Seattle, WA 98128

Primary key

BranchTelephone (1NF)

branchNo	telNo
B001	503-555-3618
B001	503-555-2727
B001	503-555-6534
B002	206-555-6756
B002	206-555-8836
B003	212-371-3000
B004	206-555-3131
B004	206-555-4112

Becomes foreign key

Second Normal Form(2NF)

- Must satisfy 1NF
- Only applies to tables with **composite primary key**
- All about relationship between columns that are key and columns that are not keys
- All (non-key) attributes/columns dependent on the full primary key
- Should not be dependent on partial of the key
- Each **non-primary key column** can be worked out from the values in ALL the column that make up the primary key
- A table/entity that has one primary key with one attribute is already in 2NF
- Formel definition:
A table is in 2NF if it is in 1NF and no *non-prime attribute* is *dependent* on any *proper subset* of any *candidate key* of the table.

TempStaffAllocation table is not in 2NF

staffNo	branchNo	branchAddress	name	position	hoursPerWeek
S4555	B002	City Center Plaza, Seattle, WA 98122	Ellen Layman	Assistant	16
S4555	B004	16 – 14th Avenue, Seattle, WA 98128	Ellen Layman	Assistant	9
S4612	B002	City Center Plaza, Seattle, WA 98122	Dave Sinclair	Assistant	14
S4612	B004	16 – 14th Avenue, Seattle, WA 98128	Dave Sinclair	Assistant	10

Composite
primary key

Values in branchAddress column can be worked out from only branchNo, so table *not* in 2NF

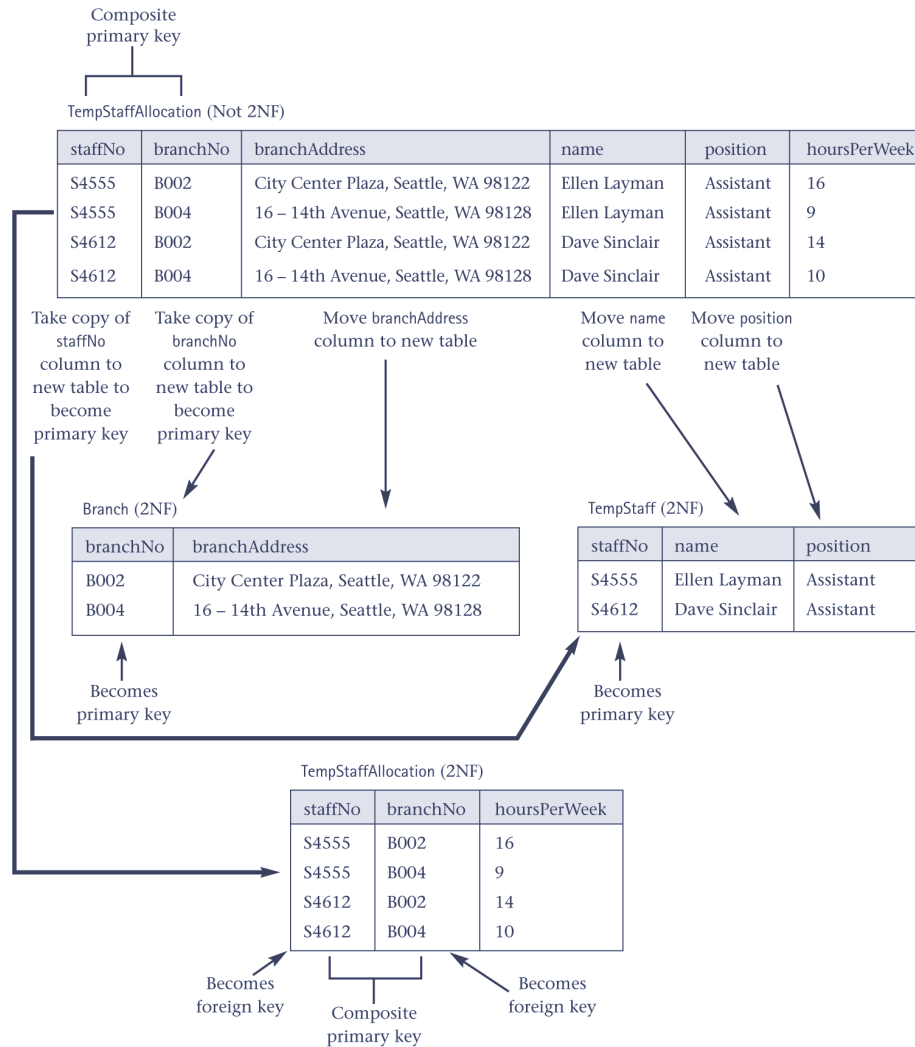
Values in branchNo column can be worked out from branchAddress

Values in name and position columns can be worked out from only staffNo, so table *not* in 2NF

Values in hoursPerWeek column can only be worked out from staffNo and branchNo

- © Pearson Education Limited, 2004

Converting TempStaffAllocation table to 2NF



Third normal form

- all the attributes in a table are determined only by the candidate keys of that table (and not by any non-prime attributes)
- All non-primary key attribute provides a fact that is independent of other non-key attributes/column and depends only on the primary key
- All non-primary-key column can be worked out from only the primary key column(s) and no other columns

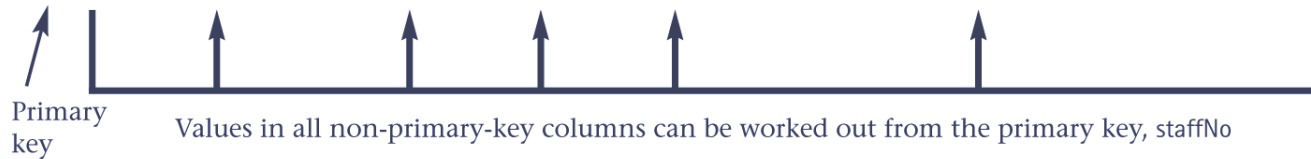
What is wrong with this?

Customer
customer_id
customer_name
balance
creditLimit
repNumber
repFName
repLName

StaffBranch table is not in 3NF

StaffBranch (Not 3NF)

staffNo	name	position	salary	branchNo	branchAddress	telNo
S1500	Tom Daniels	Manager	46000	B001	8 Jefferson Way, Portland, OR 97201	503-555-3618
S0003	Sally Adams	Assistant	30000	B001	8 Jefferson Way, Portland, OR 97201	503-555-3618
S0010	Mary Martinez	Manager	50000	B002	City Center Plaza, Seattle, WA 98122	206-555-6756
S3250	Robert Chin	Supervisor	32000	B002	City Center Plaza, Seattle, WA 98122	206-555-6756
S2250	Sally Stern	Manager	48000	B004	16 – 14th Avenue, Seattle, WA 98128	206-555-3131
S0415	Art Peters	Manager	41000	B003	14 – 8th Avenue, New York, NY 10012	212-371-3000



Values in branchAddress and telNo columns can be worked out from branchNo, so table *not* in 3NF



Values in branchNo and telNo columns can be worked out from branchAddress, so table *not* in 3NF



Values in branchNo and branchAddress columns can be worked out from telNo, so table *not* in 3NF



- © Pearson Education Limited, 2004

Converting the StaffBranch table to 3NF

StaffBranch (Not 3NF)

staffNo	name	position	salary	branchNo	branchAddress	telNo
S1500	Tom Daniels	Manager	46000	B001	8 Jefferson Way, Portland, OR 97201	503-555-3618
S0003	Sally Adams	Assistant	30000	B001	8 Jefferson Way, Portland, OR 97201	503-555-3618
S0010	Mary Martinez	Manager	50000	B002	City Center Plaza, Seattle, WA 98122	206-555-6756
S3250	Robert Chin	Supervisor	32000	B002	City Center Plaza, Seattle, WA 98122	206-555-6756
S2250	Sally Stern	Manager	48000	B004	16 – 14th Avenue, Seattle, WA 98128	206-555-3131
S0415	Art Peters	Manager	41000	B003	14 – 8th Avenue, New York, NY 10012	212-371-3000

