
Normalization

Objectives

- **The purpose of normalization.**
- **The problems associated with redundant data.**
- **The identification of various types of update anomalies such as insertion, deletion, and modification anomalies.**
- **How to recognize the appropriateness or quality of the design of relations.**

Objectives

- **The concept of functional dependency, the main tool for measuring the appropriateness of attribute groupings in relations.**
- **How functional dependencies can be used to group attributes into relations that are in a known normal form.**
- **How to define normal forms for relations.**

Objectives

- **How to undertake the process of normalization.**
- **How to identify the most commonly used normal forms, namely first (1NF), second (2NF), and third (3NF) normal forms, and Boyce–Codd normal form (BCNF).**
- **How to identify fourth (4NF), and fifth (5NF) normal forms.**

Objectives

- **How to undertake the process of normalization.**
- **How to identify the most commonly used normal forms, namely first (1NF), second (2NF), and third (3NF) normal forms, and Boyce–Codd normal form (BCNF).**
- **How to identify fourth (4NF), and fifth (5NF) normal forms.**

Normalization

- **Main objective in developing a logical data model for relational database systems is to create an accurate representation of the data, its relationships, and constraints.**
- **To achieve this objective, we must identify a suitable set of relations.**

Normalization

- A technique for producing a set of relations with desirable properties, given the data requirements of an enterprise.
- Developed by E.F. Codd (1972).
- Often performed as a series of tests on a relation to determine whether it satisfies or violates the requirements of a given normal form.

Normalization

- Four most commonly used normal forms are first (1NF), second (2NF), third (3NF) and Boyce-Codd (BCNF) normal forms.
- Based on functional dependencies among the attributes of a relation.
- A relation can be normalized to a specific form to prevent the possible occurrence of update anomalies.

Data Redundancy

- Major aim of relational database design is to group attributes into relations to minimize data redundancy and reduce file storage space required by base relations.
- Problems associated with data redundancy are illustrated by comparing the Staff and Branch relations with the Staff_Branch relation.

Data Redundancy

Staff Relation

<i>Staff_No</i>	<i>SName</i>	<i>SAddress</i>	<i>Position</i>	<i>Salary</i>	<i>Branch_No</i>
SL21	John White	19 Taylor St, London	Manager	30000	B5
SG37	Ann Beech	81 George St, Glasgow	Snr Asst	12000	B3
SG14	David Ford	63 Ashby St, Glasgow	Deputy	18000	B3
SA9	Mary Howe	2 Elm Pl, Aberdeen	Assistant	9000	B7
SG5	Susan Brand	5 Gt Western Rd, Glasgow	Manager	24000	B3
SL41	Julie Lee	28 Malvern St, Kilburn	Assistant	9000	B5

Branch Relation

<i>Branch_No</i>	<i>BAddress</i>	<i>Tel_No</i>
B5	22 Deer Rd, London	0171-886-1212
B7	16 Argyll St, Aberdeen	01224-67125
B3	163 Main St, Glasgow	0141-339-2178

Staff_Branch Relation

<i>Staff_No</i>	<i>SName</i>	<i>SAddress</i>	<i>Position</i>	<i>Salary</i>	<i>Branch_No</i>	<i>BAddress</i>	<i>Tel_No</i>
SL21	John White	19 Taylor St, London	Manager	30000	B5	22 Deer Rd, London	0171-886-1212
SG37	Ann Beech	81 George St, Glasgow	Snr Asst	12000	B3	163 Main St, Glasgow	0141-339-2178
SG14	David Ford	63 Ashby St, Glasgow	Deputy	18000	B3	163 Main St, Glasgow	0141-339-2178
SA9	Mary Howe	2 Elm Pl, Aberdeen	Assistant	9000	B7	16 Argyll St, Aberdeen	01224-67125
SG5	Susan Brand	5 Gt Western Rd, Glasgow	Manager	24000	B3	163 Main St, Glasgow	0141-339-2178
SL41	Julie Lee	28 Malvern St, Kilburn	Assistant	9000	B5	22 Deer Rd, London	0171-886-1212

Data Redundancy

- **Staff_Branch relation has redundant data; the details of a branch are repeated for every member of staff.**
- **In contrast, the branch information appears only once for each branch in the Branch relation and only the branch number (Branch_No) is repeated in the Staff relation, to represent where each member of staff is located.**

Update Anomalies

- **Relations that contain redundant information may potentially suffer from update anomalies.**
- **Types of update anomalies include**
 - **Insertion**
 - **Deletion**
 - **Modification**

Lossless-join and Dependency Preservation Properties

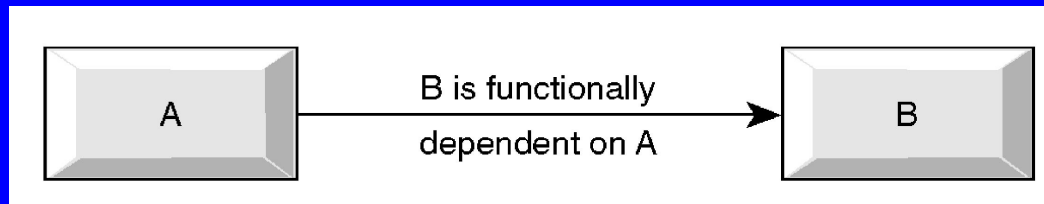
- Two important properties of decomposition.
 - *Lossless-join property* enables us to find any instance of the original relation from corresponding instances in the smaller relations.
 - *Dependency preservation property* enables us to enforce a constraint on the original relation by enforcing some constraint on each of the smaller relations.

Functional Dependency

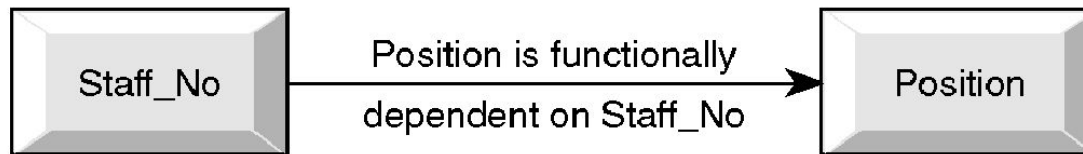
- Main concept associated with normalization.
- Functional Dependency
 - Describes the relationship between attributes in a relation.
 - For example, if A and B are attributes of relation R, B is functionally dependent on A (denoted $A \twoheadrightarrow B$), if each value of A in R is associated with exactly one value of B in R.

Functional Dependency

- Property of the meaning or semantics of the attributes in a relation.
- The *determinant* of a functional dependency refers to the attribute or group of attributes on the left-hand side of the arrow.
- Diagrammatic representation.

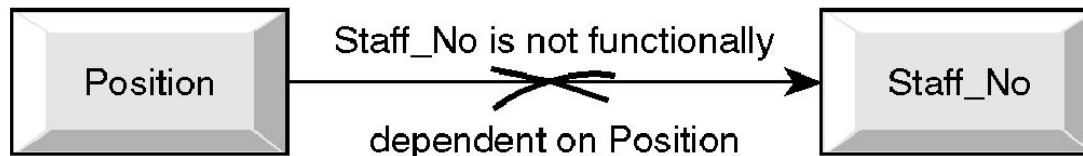


Example - Functional Dependency



Staff number SL21 → Manager

(a)



Manager → Staff number SL21
Manager → Staff number SG5

(b)

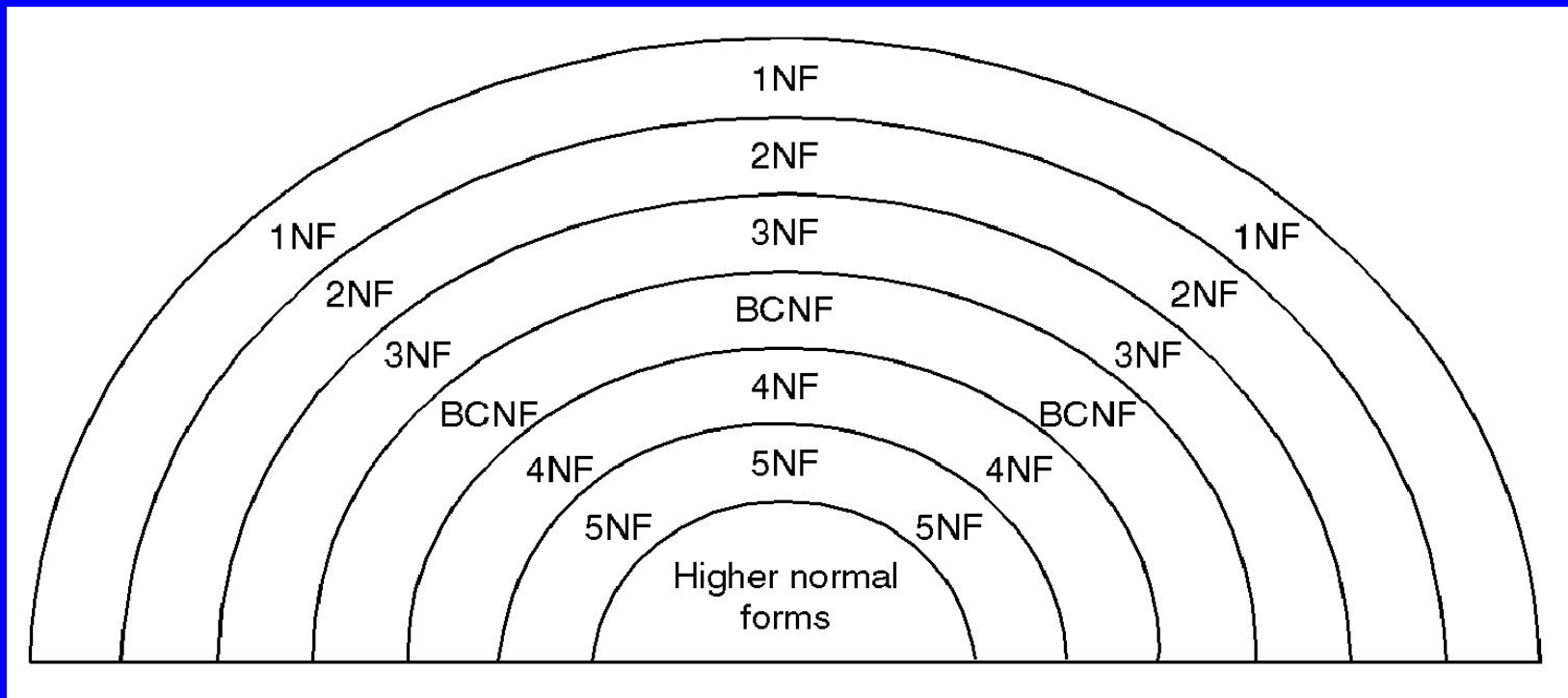
The Process of Normalization

- **Formal technique for analyzing a relation based on its primary key and the functional dependencies between the attributes of that relation.**
- **Often executed as a series of steps. Each step corresponds to a specific normal form, which has known properties.**

The Process of Normalization

- As normalization proceeds, the relations become progressively more restricted (stronger) in format and also less vulnerable to update anomalies.

Relationship Between Normal Forms



Unnormalized Form (UNF)

- A table that contains one or more repeating groups.
- To create an unnormalized table
 - Transform the data from the information source (e.g. form) into table format with columns and rows.

First Normal Form (1NF)

- A relation in which the intersection of each row and column contains one and only one value.

UNF to 1NF

- **Nominate an attribute or group of attributes to act as the key for the unnormalized table.**
- **Identify the repeating group(s) in the unnormalized table which repeats for the key attribute(s).**

UNF to 1NF

- **Remove the repeating group by**
 - **Entering appropriate data into the empty columns of rows containing the repeating data ('flattening' the table).**
 - **Or by**
 - **Placing the repeating data along with a copy of the original key attribute(s) into a separate relation.**

Second Normal Form (2NF)

- **Based on the concept of full functional dependency.**
- **Full functional dependency indicates that if**
 - **A and B are attributes of a relation,**
 - **B is fully dependent on A if B is functionally dependent on A but not on any proper subset of A.**

Second Normal Form (2NF)

- A relation that is in 1NF and every non-primary-key attribute is fully functionally dependent on the primary key.

1NF to 2NF

- Identify the primary key for the 1NF relation.
- Identify the functional dependencies in the relation.
- If partial dependencies exist on the primary key remove them by placing them in a new relation along with a copy of their determinant.

Third Normal Form (3NF)

- Based on the concept of transitive dependency.
- Transitive Dependency is a condition where
 - A, B and C are attributes of a relation such that if $A \twoheadrightarrow B$ and $B \twoheadrightarrow C$,
 - then C is transitively dependent on A through B. (Provided that A is not functionally dependent on B or C).

Third Normal Form (3NF)

- A relation that is in 1NF and 2NF and in which no non-primary-key attribute is transitively dependent on the primary key.

2NF to 3NF

- Identify the primary key in the 2NF relation.
- Identify functional dependencies in the relation.
- If transitive dependencies exist on the primary key remove them by placing them in a new relation along with a copy of their determinant.

Boyce-Codd Normal Form (BCNF)

- Based on functional dependencies that takes into account all candidate keys in a relation.
- For a relation with only one candidate key, 3NF and BCNF are equivalent.
- A relation is in BCNF, if and only if every determinant is a candidate key.

Boyce-Codd Normal Form (BCNF)

- Violation of BCNF may occur in a relation that
 - contains two (or more) composite keys
 - which overlap and share at least one attribute in common.

3NF to BCNF

- **Identify all candidate keys in the relation.**
- **Identify all functional dependencies in the relation.**
- **If functional dependencies exists in the relation where their determinants are not candidate keys for the relation, remove the functional dependencies by placing them in a new relation along with a copy of their determinant.**

Example 1 - Normalization Report

Page 1		DreamHome Customer Rental Details			Date 7-Oct-98	
Customer Name John Kay Customer Number CR76						
Property Number	Property Address	Rent Start	Rent Finish	Rent	Owner Number	Owner Name
PG4	6 Lawrence St, Glasgow	1-Jul-94	31-Aug-96	350	CO40	Tina Murphy
PG16	5 Novar Dr, Glasgow	1-Sep-96	1-Sep-98	450	CO93	Tony Shaw

Example 1 - Normalization

UNF to 1NF

Customer Rental Table

<i>Customer_No</i>	<i>CName</i>	<i>Property_No</i>	<i>PAddress</i>	<i>RentStart</i>	<i>RentFinish</i>	<i>Rent</i>	<i>Owner_No</i>	<i>OName</i>
CR76	John Kay	PG4	6 Lawrence St, Glasgow	1-Jul-94	31-Aug-96	350	CO40	Tina Murphy
		PG16	5 Novar Dr, Glasgow	1-Sep-96	1-Sep-98	450	CO93	Tony Shaw
CR56	Aline Stewart	PG4	6 Lawrence St, Glasgow	1-Sep-92	10-June-94	350	CO40	Tina Murphy
		PG36	2 Manor Rd, Glasgow	10-Oct-94	1-Dec-95	375	CO93	Tony Shaw
		PG16	5 Novar Dr, Glasgow	1-Jan-96	10-Aug-96	450	CO93	Tony Shaw



Customer_Rental Relation

<i>Customer No</i>	<i>Property No</i>	<i>CName</i>	<i>PAddress</i>	<i>RentStart</i>	<i>RentFinish</i>	<i>Rent</i>	<i>Owner No</i>	<i>OName</i>
CR76	PG4	John Kay	6 Lawrence St, Glasgow	1-Jul-94	31-Aug-96	350	CO40	Tina Murphy
CR76	PG16	John Kay	5 Novar Dr, Glasgow	1-Sep-96	1-Sep-98	450	CO93	Tony Shaw
CR56	PG4	Aline Stewart	6 Lawrence St, Glasgow	1-Sep-92	10-Jun-94	350	CO40	Tina Murphy
CR56	PG36	Aline Stewart	2 Manor Rd, Glasgow	10-Oct-94	1-Dec-95	375	CO93	Tony Shaw
CR56	PG16	Aline Stewart	5 Novar Dr, Glasgow	1-Jan-96	10-Aug-96	450	CO93	Tony Shaw

Example 1 - Normalization

UNF to 1NF (Alternative)

Customer Rental Table

<i>Customer_No</i>	<i>CName</i>	<i>Property_No</i>	<i>PAddress</i>	<i>RentStart</i>	<i>RentFinish</i>	<i>Rent</i>	<i>Owner_No</i>	<i>OName</i>
CR76	John Kay	PG4	6 Lawrence St, Glasgow	1-Jul-94	31-Aug-96	350	CO40	Tina Murphy
		PG16	5 Novar Dr, Glasgow	1-Sep-96	1-Sep-98	450	CO93	Tony Shaw
CR56	Aline Stewart	PG4	6 Lawrence St, Glasgow	1-Sep-92	10-June-94	350	CO40	Tina Murphy
		PG36	2 Manor Rd, Glasgow	10-Oct-94	1-Dec-95	375	CO93	Tony Shaw
		PG16	5 Novar Dr, Glasgow	1-Jan-96	10-Aug-96	450	CO93	Tony Shaw



Customer Relation

<i>Customer_No</i>	<i>CName</i>
CR76	John Kay
CR56	Aline Stewart

Prop_Rental_Owner Relation

<i>Customer_No</i>	<i>Property_No</i>	<i>PAddress</i>	<i>RentStart</i>	<i>RentFinish</i>	<i>Rent</i>	<i>Owner_No</i>	<i>OName</i>
CR76	PG4	6 Lawrence St, Glasgow	1-Jul-94	31-Aug-96	350	CO40	Tina Murphy
CR76	PG16	5 Novar Dr, Glasgow	1-Sep-96	1-Sep-98	450	CO93	Tony Shaw
CR56	PG4	6 Lawrence St, Glasgow	1-Sep-92	10-Jun-94	350	CO40	Tina Murphy
CR56	PG36	2 Manor Rd, Glasgow	10-Oct-94	1-Dec-95	375	CO93	Tony Shaw
CR56	PG16	5 Novar Dr, Glasgow	1-Jan-96	10-Aug-96	450	CO93	Tony Shaw

Example 1 - Normalization

FDs for Customer_Rental Relation

Primary Key

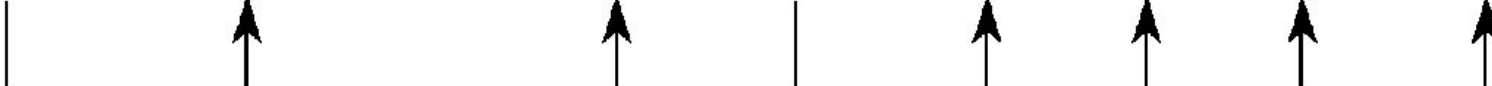
<i>Customer_No</i>	<i>Property_No</i>	<i>CName</i>	<i>PAddress</i>	<i>RentStart</i>	<i>RentFinish</i>	<i>Rent</i>	<i>Owner_No</i>	<i>OName</i>
--------------------	--------------------	--------------	-----------------	------------------	-------------------	-------------	-----------------	--------------

fd1  (Primary key)

fd2  (Partial dependency)

fd3  (Partial dependency)

fd4  (Transitive dependency)

fd5  (Candidate key)

fd6  (Candidate key)

Example 1 - Normalization

Customer_Rental to 2NF Relations

Customer Relation

<i>Customer_No</i>	<i>CName</i>
CR76	John Kay
CR56	Aline Stewart

Rental Relation

<i>Customer_No</i>	<i>Property_No</i>	<i>RentStart</i>	<i>RentFinish</i>
CR76	PG4	1-Jul-94	31-Aug-96
CR76	PG16	1-Sep-96	1-Sep-98
CR56	PG4	1-Sep-92	10-Jun-94
CR56	PG36	10-Oct-94	1-Dec-95
CR56	PG16	1-Jan-96	10-Aug-96

Property_Owner Relation

<i>Property_No</i>	<i>PAddress</i>	<i>Rent</i>	<i>Owner_No</i>	<i>OName</i>
PG4	6 Lawrence St, Glasgow	350	CO40	Tina Murphy
PG16	5 Novar Dr, Glasgow	450	CO93	Tony Shaw
PG36	2 Manor Rd, Glasgow	375	CO93	Tony Shaw

Example 1 - Normalization

Property_Owner to 3NF Relations

Property_for_Rent Relation

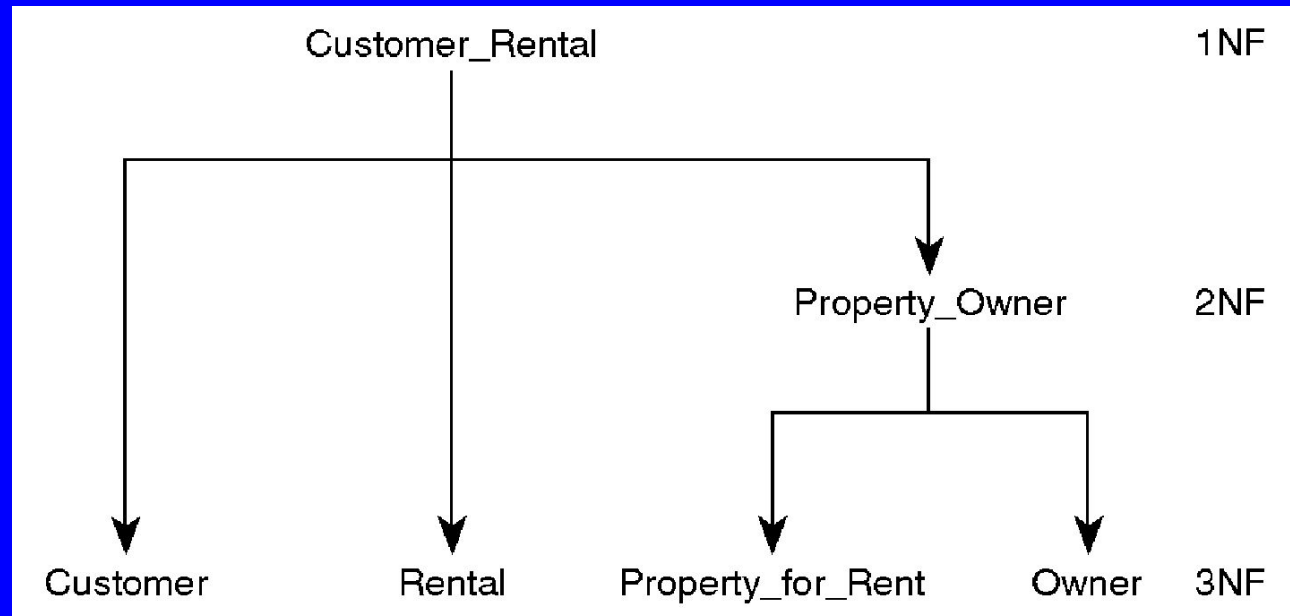
<i>Property_No</i>	<i>PAddress</i>	<i>Rent</i>	<i>Owner_No</i>
PG4	6 Lawrence St, Glasgow	350	CO40
PG16	5 Novar Dr, Glasgow	450	CO93
PG36	2 Manor Rd, Glasgow	375	CO93

Owner Relation

<i>Owner_No</i>	<i>OName</i>
CO40	Tina Murphy
CO93	Tony Shaw

Example 1 - Normalization

Process of Decomposition



Example 1 - Normalization

Summary of 3NF Relations

Customer Relation

<i>Customer_No</i>	<i>CName</i>
CR76	John Kay
CR56	Aline Stewart

Rental Relation

<i>Customer_No</i>	<i>Property_No</i>	<i>RentStart</i>	<i>RentFinish</i>
CR76	PG4	1-Jul-94	31-Aug-96
CR76	PG16	1-Sep-96	1-Sep-98
CR56	PG4	1-Sep-92	10-Jun-94
CR56	PG36	10-Oct-94	1-Dec-95
CR56	PG16	1-Jan-96	10-Aug-96

Property_for_Rent Relation

<i>Property_No</i>	<i>PAddress</i>	<i>Rent</i>	<i>Owner_No</i>
PG4	6 Lawrence St, Glasgow	350	CO40
PG16	5 Novar Dr, Glasgow	450	CO93
PG36	2 Manor Rd, Glasgow	375	CO93

Owner Relation

<i>Owner_No</i>	<i>OName</i>
CO40	Tina Murphy
CO93	Tony Shaw

Example 2 - Normalization

3NF to BCNF Relations

Client_Interview Relation

<i>Client_No</i>	<i>Interview_Date</i>	<i>Interview_Time</i>	<i>Staff_No</i>	<i>Room_No</i>
CR76	13-May-98	10.30	SG5	G101
CR56	13-May-98	12.00	SG5	G101
CR74	13-May-98	12.00	SG37	G102
CR56	1-Jul-98	10.30	SG5	G102



Interview Relation

<i>Client_No</i>	<i>Interview_Date</i>	<i>Interview_Time</i>	<i>Staff_No</i>
CR76	13-May-98	10.30	SG5
CR56	13-May-98	12.00	SG5
CR74	13-May-98	12.00	SG37
CR56	1-Jul-98	10.30	SG5

Staff_Room Relation

<i>Staff_No</i>	<i>Interview_Date</i>	<i>Room_No</i>
SG5	13-May-98	G101
SG37	13-May-98	G102
SG5	1-Jul-98	G102

Example 3 - Normalization Report

Page 1		DreamHome Property Inspection Report		Date 1-Oct-98	
Property Number PG4			Property Address 6 Lawrence St, Glasgow		
Inspection Date	Inspection Time	Comments	Staff Number	Staff Name	Car Reg
18-Oct-96	10.00	Need to replace crockery	SG37	Ann Beech	M231 JGR
22-Apr-97	09.00	In good order	SG14	David Ford	M533 HDR
1-Oct-98	12.00	Damp rot in bathroom	SG14	David Ford	N721 HFR

Example 3 - Normalization

UNF to 1NF Relation

Property_Inspection Table

<i>Property_No</i>	<i>PAddress</i>	<i>IDate</i>	<i>ITime</i>	<i>Comments</i>	<i>Staff_No</i>	<i>SName</i>	<i>Car_Reg</i>
PG4	6 Lawrence St, Glasgow	18-Oct-96	10.00	need to replace crockery	SG37	Ann Beech	M231 JGR
		22-Apr-97	09.00	in good order	SG14	David Ford	M533 HDR
		1-Oct-98	12.00	damp rot in bathroom	SG14	David Ford	N721 HFR
PG16	5 Novar Dr, Glasgow	22-Apr-96	13.00	replace living room carpet	SG14	David Ford	M533 HDR
		24-Oct-97	14.00	good condition	SG37	Ann Beech	N721 HFR



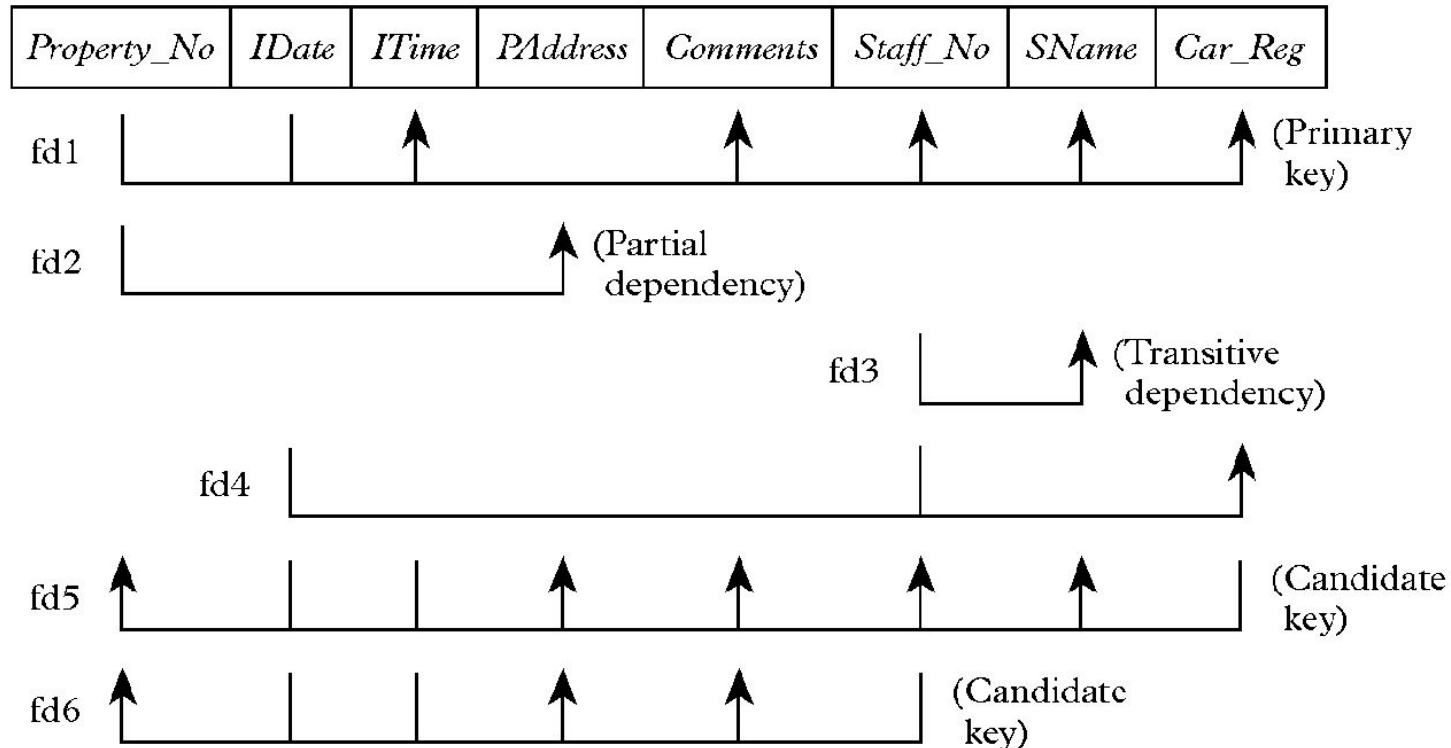
Property_Inspection Relation

<i>Property_No</i>	<i>IDate</i>	<i>ITime</i>	<i>PAddress</i>	<i>Comments</i>	<i>Staff_No</i>	<i>SName</i>	<i>Car_Reg</i>
PG4	18-Oct-96	10.00	6 Lawrence St, Glasgow	need to replace crockery	SG37	Ann Beech	M231 JGR
PG4	22-Apr-97	09.00	6 Lawrence St, Glasgow	in good order	SG14	David Ford	M533 HDR
PG4	1-Oct-98	12.00	6 Lawrence St, Glasgow	damp rot in bathroom	SG14	David Ford	N721 HFR
PG16	22-Apr-96	13.00	5 Novar Dr, Glasgow	replace living room carpet	SG14	David Ford	M533 HDR
PG16	24-Oct-97	14.00	5 Novar Dr, Glasgow	good condition	SG37	Ann Beech	N721 HFR

Example 3 - Normalization

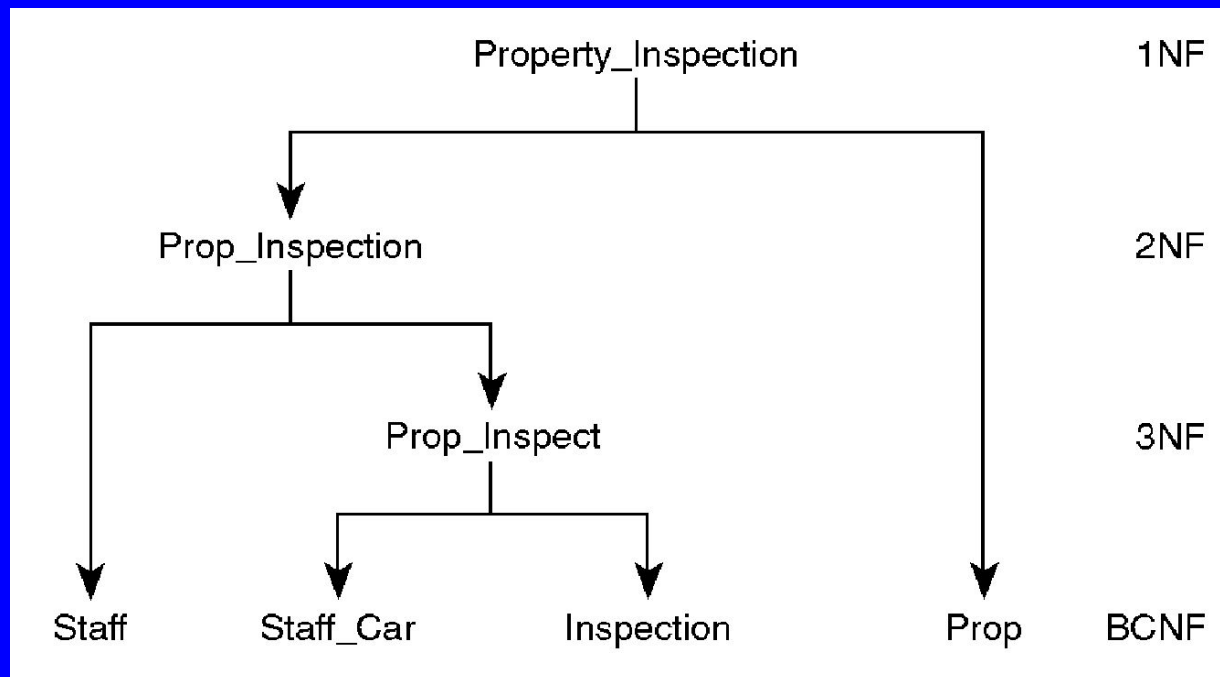
FDs of Property_Inspection Relation

Primary Key



Example 3 - Normalization Review

Process of Decomposition



Fourth Normal Form (4NF)

- Associated with a dependency called **multi-valued dependency (MVD)**.
- **MVDs in a relation are due to first normal form (1NF), which disallows an attribute in a row from having a set of values.**

MVD

- Represents a dependency between attributes (for example, A, B, and C) in a relation, such that for each value of A there is a set of values for B, and a set of values for C. However, the set of values for B and C are independent of each other.
- MVD between attributes A, B, and C in a relation using the following notation:

A \twoheadrightarrow B

A \twoheadrightarrow C

Fourth Normal Form (4NF)

- A relation that is in Boyce-Codd Normal Form and contains no MVDs.
- BCNF to 4NF involves the removal of the MVD from the relation by placing the attribute(s) in a new relation along with a copy of the determinant(s).

Example 4 - Normalization

BCNF to 4NF Relations

Branch_Staff_Client relation

<i>Branch_No</i>	<i>SName</i>	<i>CName</i>
B3	Ann Beech	Aline Stewart
B3	David Ford	Aline Stewart
B3	Ann Beech	Mike Richie
B3	David Ford	Mike Richie



Branch_Staff relation

<i>Branch_No</i>	<i>SName</i>
B3	Ann Beech
B3	David Ford

Branch_Client relation

<i>Branch_No</i>	<i>CName</i>
B3	Aline Stewart
B3	Mike Richie

Fifth Normal Form (5NF)

- **Lossless-join property refers to when we decompose a relation into two relations - we can rejoin the resulting relations to produce the original relation.**
- **However, sometimes there is the requirement to decompose a relation into more than two relations. Although rare, these cases are managed by join dependency and 5NF.**

5NF and Lossless-join Dependency

- **Lossless-join Dependency**
 - A property of decomposition, which ensures that no spurious rows are generated when relations are reunited through a natural join operation.
- **5NF**
 - A relation that has no join dependency.

Example 4 - Normalization

4NF to 5NF Relations

Property_Item_Supplier relation

<i>Property_No</i>	<i>Item_Description</i>	<i>Supplier_No</i>
PG4	Bed	S1
PG4	Chair	S2
PG16	Bed	S2
PG16	Table	S1
PG36	Chair	S3



Property_Item relation

<i>Property_No</i>	<i>Item_Description</i>
PG4	Bed
PG4	Chair
PG16	Bed
PG16	Table
PG32	Chair

Item_Supplier relation

<i>Item_Description</i>	<i>Supplier_No</i>
Bed	S1
Chair	S2
Bed	S2
Table	S1
Chair	S3

Property_Supplier relation

<i>Property_No</i>	<i>Supplier_No</i>
PG4	S1
PG4	S2
PG16	S2
PG16	S1
PG36	S3

The Process of Normalization up to 5NF

