Chapter 7,
DATABASE SYSTEMS Thomas Connolly et al.,

Normalization Transparencies

Naranbat Bayanmunkh

Chapter 7 - Objectives

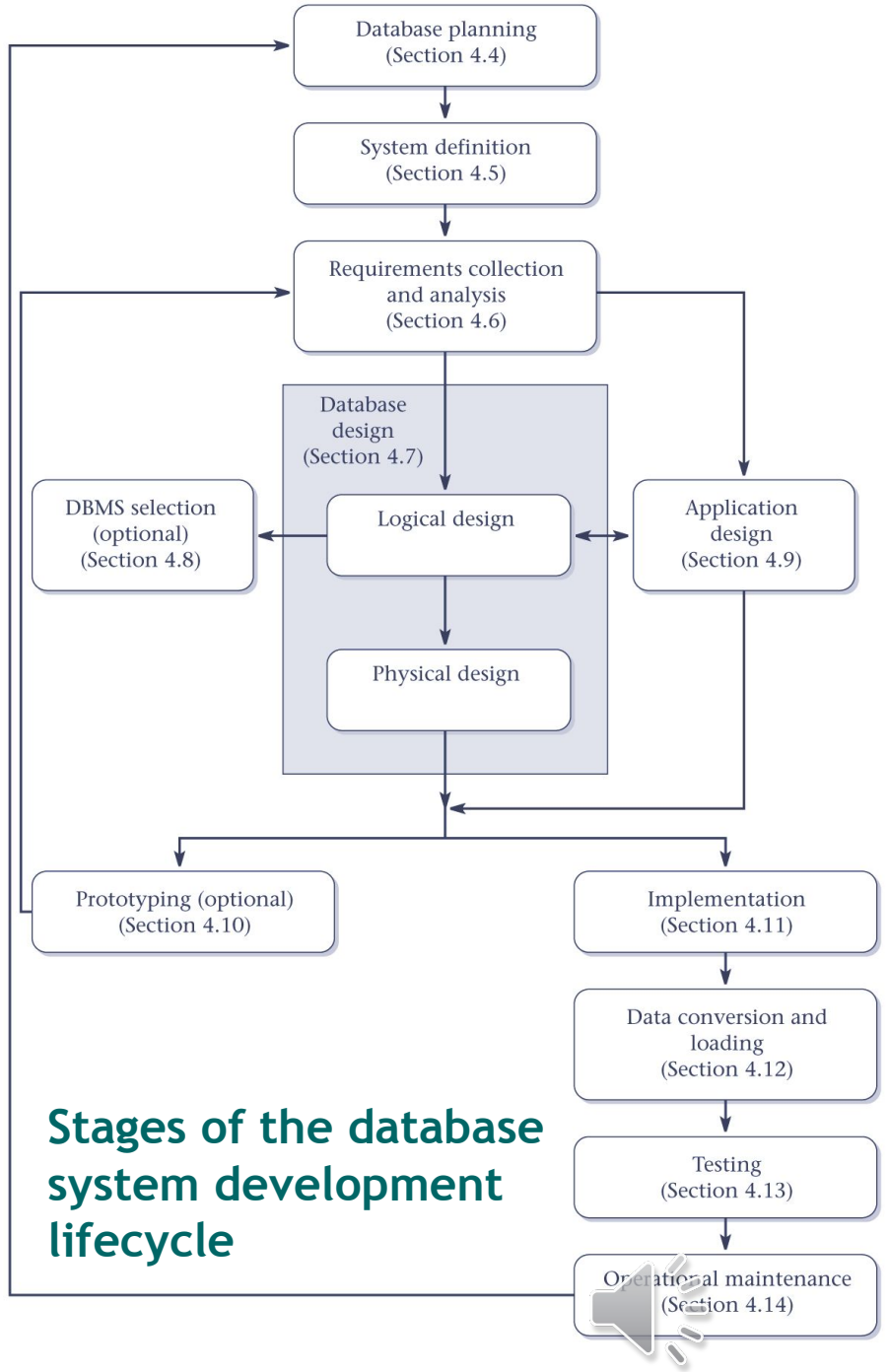
- Purpose of normalization.
- Problems associated with redundant data.
- Identification of various types of update anomalies such as insertion, deletion, and modification anomalies.
- How to recognize appropriateness or quality of the design of relations.

Chapter 7 - Objectives

- How functional dependencies can be used to group attributes into relations that are in a known normal form.
- How to undertake process of normalization.
- How to identify most commonly used normal forms, namely 1NF, 2NF, 3NF, 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, must identify a suitable set of relations.



Stages of the database system development lifecycle

Normalization

- Four most commonly used normal forms are first (1NF), second (2NF) and third (3NF) normal forms, and Boyce–Codd normal form (BCNF).
- Based on functional dependencies among the attributes of a relation.
- A relation can be normalized to a specific form to prevent 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 following Staff and Branch relations with the StaffBranch relation.**

Data Redundancy

Staff

staffNo	sName	position	salary	branchNo
SL21	John White	Manager	30000	B005
SG37	Ann Beech	Assistant	12000	B003
SG14	David Ford	Supervisor	18000	B003
SA9	Mary Howe	Assistant	9000	B007
SG5	Susan Brand	Manager	24000	B003
SL41	Julie Lee	Assistant	9000	B005

Branch

branchNo	bAddress
B005	22 Deer Rd, London
B007	16 Argyll St, Aberdeen
B003	163 Main St, Glasgow

Staff Branch

staffNo	sName	position	salary	branchNo	bAddress
SL21	John White	Manager	30000	B005	22 Deer Rd, London
SG37	Ann Beech	Assistant	12000	B003	163 Main St, Glasgow
SG14	David Ford	Supervisor	18000	B003	163 Main St, Glasgow
SA9	Mary Howe	Assistant	9000	B007	16 Argyll St, Aberdeen
SG5	Susan Brand	Manager	24000	B003	163 Main St, Glasgow
SL41	Julie Lee	Assistant	9000	B005	22 Deer Rd, London

Data Redundancy

- **StaffBranch relation has redundant data: details of a branch are repeated for every member of staff.**

- **In contrast, branch information appears only once for each branch in Branch relation and only branchNo is repeated in Staff relation, to represent where each member of staff works.**

Update Anomalies

- **Relations that contain redundant information may potentially suffer from update anomalies.**

- **Types of update anomalies include:**
  - **Insertion,**
  - **Deletion,**
  - **Modification.**

Data Redundancy

Staff

staffNo	sName	position	salary	branchNo
SL21	John White	Manager	30000	B005
SG37	Ann Beech	Assistant	12000	B003
SG14	David Ford	Supervisor	18000	B003
SA9	Mary Howe	Assistant	9000	B007
SG5	Susan Brand	Manager	24000	B003
SL41	Julie Lee	Assistant	9000	B005

Branch

branchNo	bAddress
B005	22 Deer Rd, London
B007	16 Argyll St, Aberdeen
B003	163 Main St, Glasgow

Staff Branch

staffNo	sName	position	salary	branchNo	bAddress
SL21	John White	Manager	30000	B005	22 Deer Rd, London
SG37	Ann Beech	Assistant	12000	B003	163 Main St, Glasgow
SG14	David Ford	Supervisor	18000	B003	163 Main St, Glasgow
SA9	Mary Howe	Assistant	9000	B007	16 Argyll St, Aberdeen
SG5	Susan Brand	Manager	24000	B003	163 Main St, Glasgow
SL41	Julie Lee	Assistant	9000	B005	22 Deer Rd, London

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 original relation from corresponding instances in the smaller relations.

  - *Dependency preservation property* enables us to enforce a constraint on original relation by enforcing some constraint on each of the smaller relations.

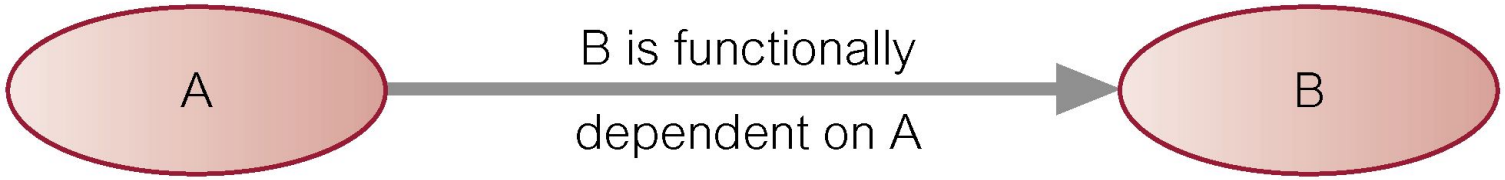
Functional Dependency

- **Main concept associated with normalization.**

- **Functional Dependency**
  - **Describes relationship between attributes in a relation.**
  - **If A and B are attributes of relation R, B is functionally dependent on A (denoted A → B), if each value of A in R is associated with exactly one value of B in R.**

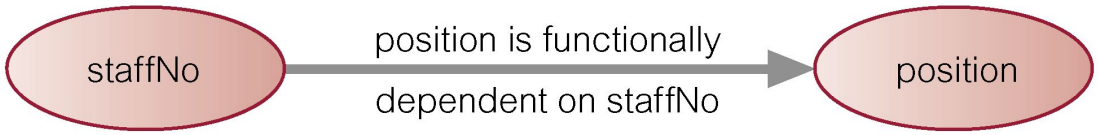
Functional Dependency

- **Property of the meaning (or semantics) of the attributes in a relation.**
- **Diagrammatic representation:**

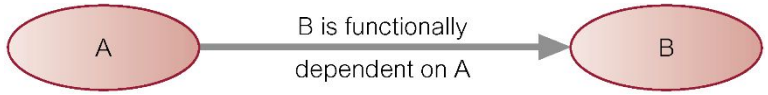


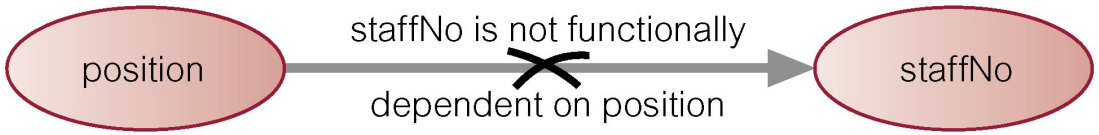
  - ***Determinant*** **of a functional dependency refers to attribute or group of attributes on left-hand side of the arrow.**

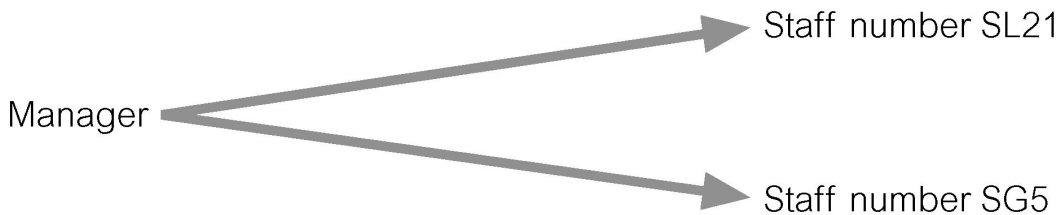
Example - Functional Dependency



(a)







(b)

Functional Dependency

- **Main characteristics of functional dependencies used in normalization:**
  - **have a 1:1 relationship between attribute(s) on left and right-hand side of a dependency;**
  - **hold for all time;**
  - **are nontrivial.**

Functional Dependency

- **Complete set of functional dependencies for a given relation can be very large.**
- **Important to find an approach that can reduce set to a manageable size.**
- **Need to identify set of functional dependencies (X) for a relation that is smaller than complete set of functional dependencies (Y) for that relation and has property that every functional dependency in Y is implied by functional dependencies in X.**

Staff Branch

staffNo	sName	position	salary	branchNo	bAddress
SL21	John White	Manager	30000	B005	22 Deer Rd, London
SG37	Ann Beech	Assistant	12000	B003	163 Main St, Glasgow
SG14	David Ford	Supervisor	18000	B003	163 Main St, Glasgow
SA9	Mary Howe	Assistant	9000	B007	16 Argyll St, Aberdeen
SG5	Susan Brand	Manager	24000	B003	163 Main St, Glasgow
SL41	Julie Lee	Assistant	9000	B005	22 Deer Rd, London

Staff_No->Sname,
Staff_No->Position,
Staff_No->Salary,
Staff_No->Branch_No,
Staff_No->BAddress,
Branch_No->BAddress,
BAddress->Branch_No,
Branch_No, Position-> salary
Badress, position->salary

Staff

staffNo	sName	position	salary	branchNo
SL21	John White	Manager	30000	B005
SG37	Ann Beech	Assistant	12000	B003
SG14	David Ford	Supervisor	18000	B003
SA9	Mary Howe	Assistant	9000	B007
SG5	Susan Brand	Manager	24000	B003
SL41	Julie Lee	Assistant	9000	B005

Branch

branchNo	bAddress
B005	22 Deer Rd, London
B007	16 Argyll St, Aberdeen
B003	163 Main St, Glasgow

Staff_No->Sname,Position,Salary,Branch_No, Baddress
Branch_No->BAddress,
BAddress->Branch_No
Branch_No, Position-> salary
Badress, position->salary

Functional Dependency

- **Set of all functional dependencies implied by a given set of functional dependencies X called closure of X (written X^+).**

- **Set of inference rules, called Armstrong's axioms, specifies how new functional dependencies can be inferred from given ones.**

Functional Dependency

- **Let A, B, and C be subsets of the attributes of relation R. Armstrong's axioms are as follows:**

  1. Reflexivity

  If B is a subset of A, then A → B

  2. Augmentation

  If A → B, then A,C → B,C

  3. Transitivity

  If A → B and B → C, then A → C

Functional Dependency

4. Self-determination:

$$A \rightarrow A$$

5. Decomposition:

If $A \rightarrow B,C$, then $A \rightarrow B$ and $A \rightarrow C$

6. Union:

If $A \rightarrow B$ and $A \rightarrow C$, then $A \rightarrow B,C$

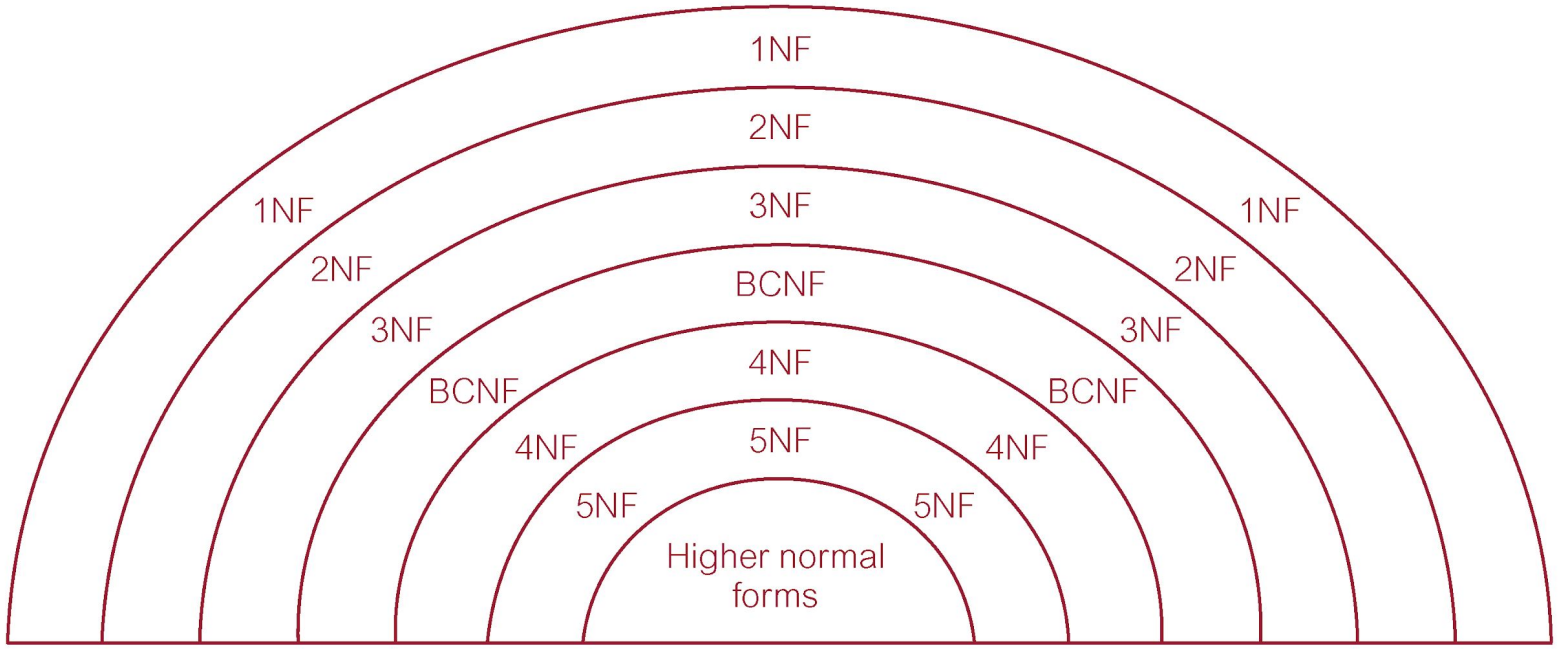
7. Composition:

If $A \rightarrow B$ and $C \rightarrow D$ then $A,C \rightarrow B,D$

The Process of Normalization

- Formal technique for analyzing a relation based on its primary key and functional dependencies between its attributes.
- Often executed as a series of steps. Each step corresponds to a specific normal form, which has known properties.
- As normalization proceeds, 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 data from information source (e.g. form) into table format with columns and rows.**

First Normal Form (1NF)

- **A relation in which 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 repeating group(s) in unnormalized table which repeats for the key attribute(s).**

UNF to 1NF

- **Remove repeating group by:**
  - **entering appropriate data into the empty columns of rows containing repeating data ('flattening' the table).**

  Or by

  - **placing repeating data along with copy of the original key attribute(s) into a separate relation.**

Second Normal Form (2NF)

- Based on concept of **full functional dependency:**
  - 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.

NOT Staff_No, Sname->Branch_No

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

3. Transitivity

If A → B and B → C, then A → C

Third Normal Form (3NF)

- Based on concept of transitive dependency:
  - A, B and C are attributes of a relation such that if A □ B and B □ C,
  - then C is transitively dependent on A through B. (Provided that A is not functionally dependent on B or C).

- 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 copy of their determinant.**

General Definitions of 2NF and 3NF

- **Second normal form (2NF)**
  - **A relation that is in 1NF and every non-primary-key attribute is fully functionally dependent on *any candidate key*.**

- **Third normal form (3NF)**
  - **A relation that is in 1NF and 2NF and in which no non-primary-key attribute is transitively dependent on *any candidate key*.**

Boyce–Codd Normal Form (BCNF)

- **Based on functional dependencies that take into account all candidate keys in a relation, however BCNF also has additional constraints compared with general definition of 3NF.**

- **BCNF - A relation is in BCNF if and only if every determinant is a candidate key.**

Boyce–Codd normal form (BCNF)

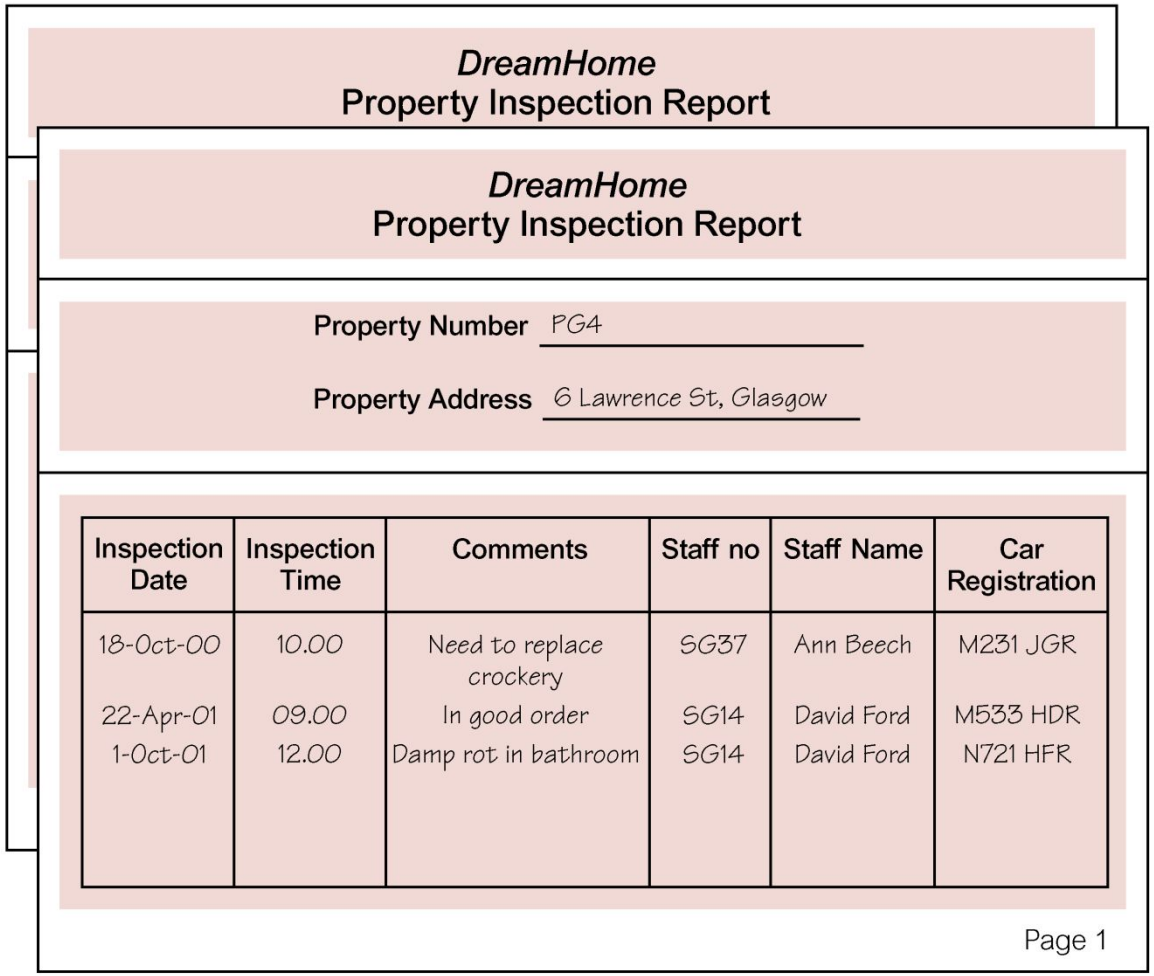
- **Difference between 3NF and BCNF is that for a functional dependency A → B, 3NF allows this dependency in a relation if B is a primary-key attribute and A is not a candidate key.**
- **Whereas, BCNF insists that for this dependency to remain in a relation, A must be a candidate key.**
- **Every relation in BCNF is also in 3NF. However, relation in 3NF may not be in BCNF.**

Boyce–Codd normal form (BCNF)

- **Violation of BCNF is quite rare.**

- **Potential to violate BCNF may occur in a relation that:**
  - **contains two (or more) composite candidate keys;**
  - **the candidate keys overlap (i.e. have at least one attribute in common).**

Review of Normalization (UNF to BCNF)



DreamHome
Property Inspection Report

DreamHome
Property Inspection Report

Property Number PG4

Property Address 6 Lawrence St, Glasgow

Inspection Date	Inspection Time	Comments	Staff no	Staff Name	Car Registration
18-Oct-00	10.00	Need to replace crockery	SG37	Ann Beech	M231 JGR
22-Apr-01	09.00	In good order	SG14	David Ford	M533 HDR
1-Oct-01	12.00	Damp rot in bathroom	SG14	David Ford	N721 HFR

Page 1

Review of Normalization (UNF to BCNF)

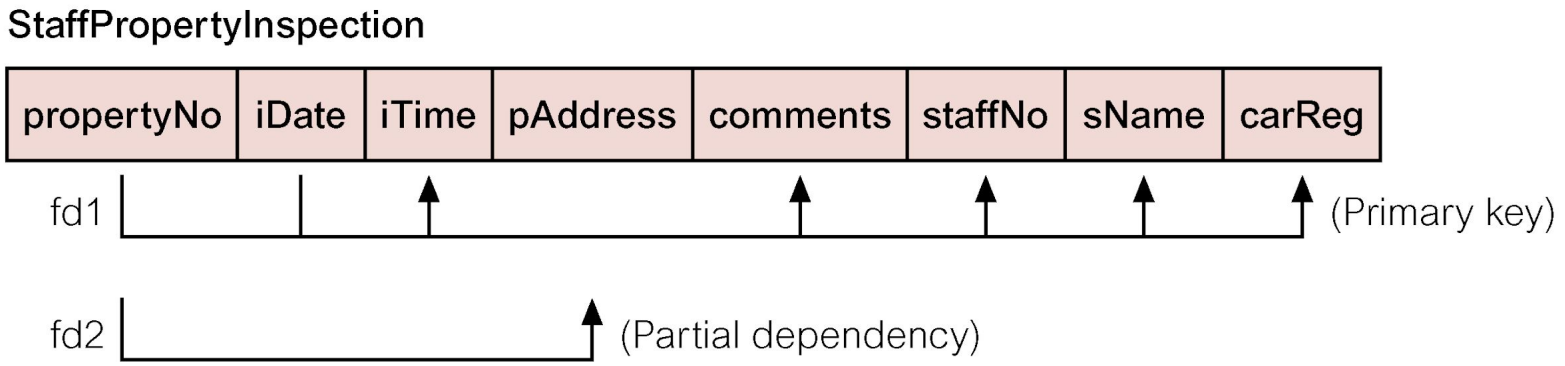
StaffPropertyInspection

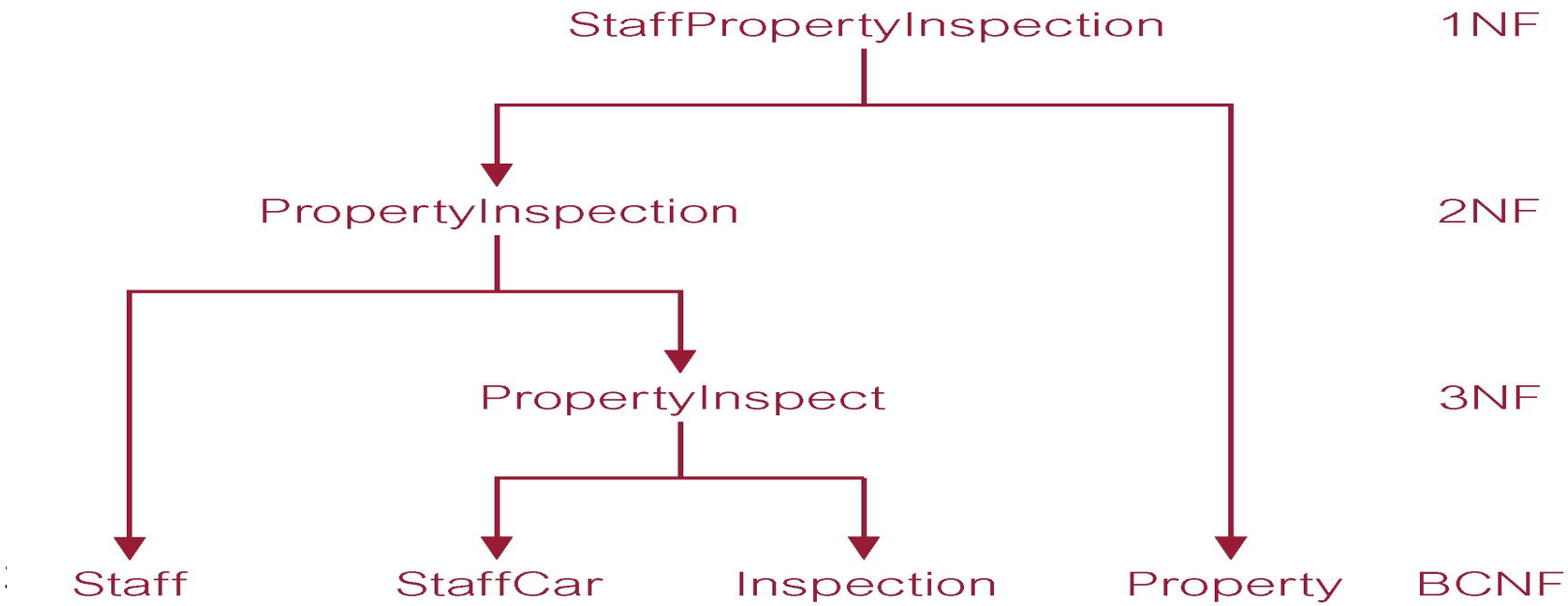
propertyNo	pAddress	iDate	iTime	comments	staffNo	sName	carReg
PG4	6 Lawrence St, Glasgow	18-Oct-00 22-Apr-01 1-Oct-01	10.00 09.00 12.00	Need to replace crockery In good order Damp rot in bathroom	SG37 SG14 SG14	Ann Beech David Ford David Ford	M231 JGR M533 HDR N721 HFR
PG16	5 Novar Dr, Glasgow	22-Apr-01 24-Oct-01	13.00 14.00	Replace living room carpet Good condition	SG14 SG37	David Ford Ann Beech	M533 HDR N721 HFR

StaffPropertyInspection

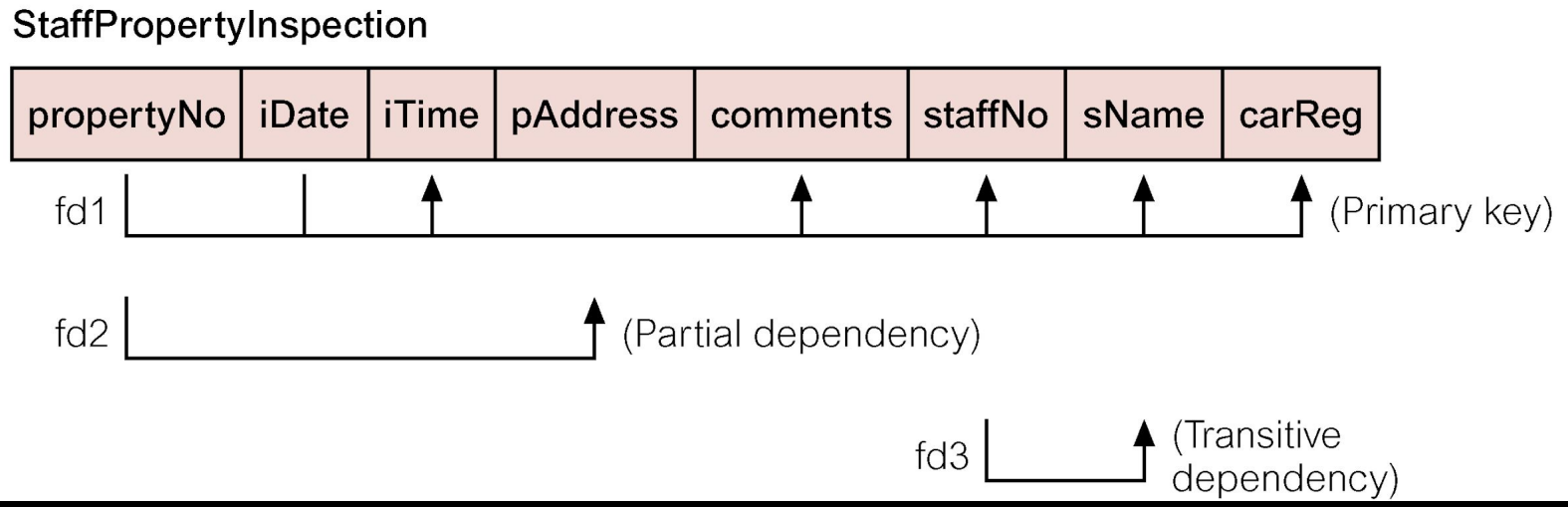
propertyNo	iDate	iTime	pAddress	comments	staffNo	sName	carReg
PG4	18-Oct-00	10.00	6 Lawrence St, Glasgow	Need to replace crockery	SG37	Ann Beech	M231 JGR
PG4	22-Apr-01	09.00	6 Lawrence St, Glasgow	In good order	SG14	David Ford	M533 HDR
PG4	1-Oct-01	12.00	6 Lawrence St, Glasgow	Damp rot in bathroom	SG14	David Ford	N721 HFR
PG16	22-Apr-01	13.00	5 Novar Dr, Glasgow	Replace living room carpet	SG14	David Ford	M533 HDR
PG16	24-Oct-01	14.00	5 Novar Dr, Glasgow	Good condition	SG37	Ann Beech	N721 HFR

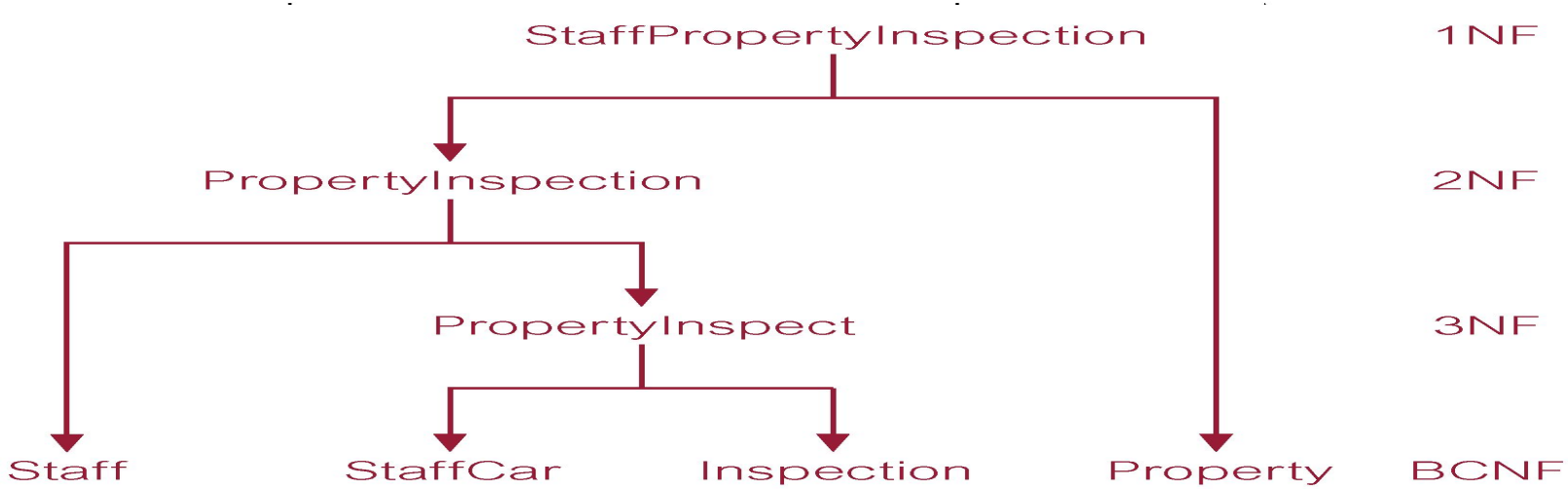
Review of Normalization (UNF to BCNF)



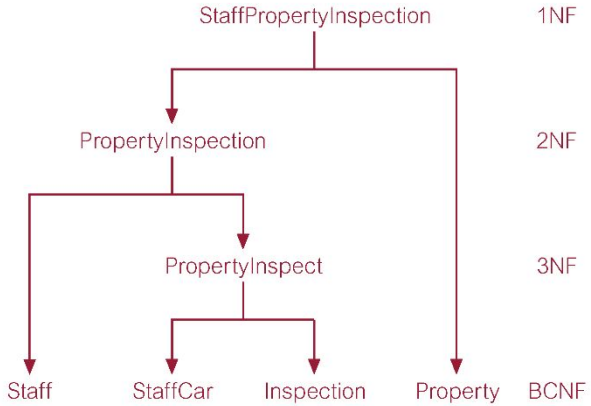


Review of Normalization (UNF to BCNF)





Review of Normalization (UNF to BCNF)



StaffPropertyInspection

propertyNo	iDate	iTime	pAddress	comments	staffNo	sName	carReg

fd1 (Primary key)

fd2 (Partial dependency)

fd3 (Transitive dependency)

fd4

fd5 (Candidate key)

fd6 (Candidate key)