Normalisation

Exercise

Customer No	Property No	Cname	Paddress	RentStart	RentFinish	Rent	Owner_No	Oname
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

- Identify the functional dependencies in the above relation
- What is the relationship between the number of attributes and number of FDs?

The Process of Normalisation

- 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

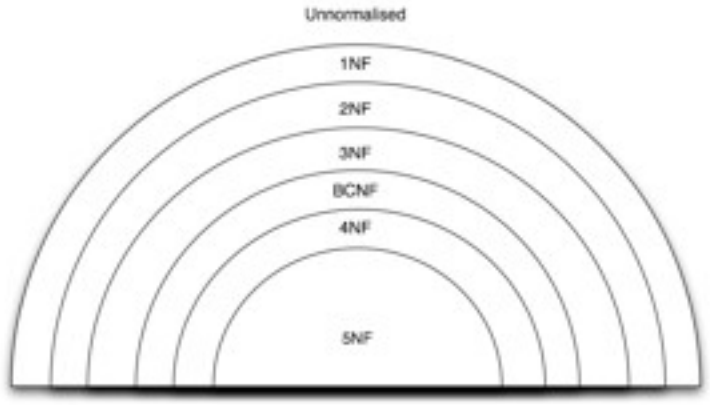
Normalisation

- 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 normalised to a specific form to prevent the possible occurrence of update anomalies

The Process of Normalisation

- As normalisation proceeds, we say that the relations become progressively more restricted (stronger)
- They also become less vulnerable to update anomalies

Relationship Between Normal Forms



Unnormalised Form (UNF)

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

Unnormalised example

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

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 unnormalised table
- Identify the repeating group(s) in the unnormalised 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

Exercise

Customer No	Property No	Cname	Paddress	RentStart	RentFinish	Rent	Owner_No	Oname
CR76	PG4	John Kay	6 Lawrence St, Glasgow	1-Jul-94	31-Aug-96	350	CO40	Tina Murphy
CR76	PG16	John Kay	3 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	3 Novar Dr, Glasgow	1-Jan-96	10-Aug-96	450	CO93	Tony Shaw

- Identify Superkeys, Candidate Keys and then an appropriate Primary key for this table
- Redefine the term Superkey in terms of functional dependencies

Normalisation UNF to 1NF

Customer No	Cname	Property No	Paddress	RentStart	RentFinish	Rent	Owner_No	Oname
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-Jun-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 No	Property No	Cname	Paddress	RentStart	RentFinish	Rent	Owner_No	Oname
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

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 then in a new relation along with a copy of their determinant

1NF to 2NF

Customer No	Property No	Cname	Paddress	RentStart	RentFinish	Rent	Owner_No	Oname
CR76	PG4	John Kay	6 Lawrence St, Glasgow	1-Jul-94	31-Aug-96	350	CO40	Tina Murphy
CR76	PG16	John Kay	3 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	3 Novar Dr, Glasgow	1-Jan-96	10-Aug-96	450	CO93	Tony Shaw

- CNo, PNo is primary key
- But not all attributes are dependant on both
- Which are dependant on CNo?
- Which are dependant on PNo alone?

1NF to 2NF

Rental

Customer No	Property No	RentStart	RentFinish
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

Customer

Customer No	Cname
CR76	John Kay
CR56	Aline Stewart

Property

Property No	Paddress	Rent	Owner_No	Oname
PG4	6 Lawrence St, Glasgow	350	CO40	Tina Murphy
PG36	2 Manor Rd, Glasgow	375	CO93	Tony Shaw
PG16	3 Novar Dr, Glasgow	450	CO93	Tony Shaw

- Check for DPD and LJD at this stage

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 ->B and B->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

2NF to 3NF

Rental

Customer No	Property No	RentStart	RentFinish
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

Customer

Customer No	Cname
CR76	John Kay
CR56	Aline Stewart

Property

Property No	Paddress	Rent	Owner_No	Oname
PG4	6 Lawrence St, Glasgow	350	CO40	Tina Murphy
PG36	2 Manor Rd, Glasgow	375	CO93	Tony Shaw
PG16	5 Novar Dr, Glasgow	450	CO93	Tony Shaw

- Check for transitive dependencies in Customer, Property and Rental

2NF to 3NF

- Property is split

Property

Property No	Paddress	Rent	Owner_No	Oname
PG4	6 Lawrence St, Glasgow	350	CO40	Tina Murphy
PG36	2 Manor Rd, Glasgow	375	CO93	Tony Shaw
PG16	5 Novar Dr, Glasgow	450	CO93	Tony Shaw

Property

Property No	Paddress	Rent	Owner_No
PG4	6 Lawrence St, Glasgow	350	CO40
PG36	2 Manor Rd, Glasgow	375	CO93
PG16	5 Novar Dr, Glasgow	450	CO93

Owner

Owner_No	Oname
CO40	Tina Murphy
CO93	Tony Shaw

- Again - check for LJD and DPD

Normalisation
Summary of 3NF Relations

Customer

Customer No	Cname
CR76	John Kay
CR56	Aline Stewart

Rental

Customer No	Property No	RentStart	RentFinish
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

Property No	Paddress	Rent	Owner_No
PG4	6 Lawrence St, Glasgow	350	CO40
PG36	2 Manor Rd, Glasgow	375	CO93
PG16	5 Novar Dr, Glasgow	450	CO93

Owner

Owner_No	Oname
CO40	Tina Murphy
CO93	Tony Shaw

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

3NF to BCNF Example

Client_Interview Relation

Client_No	*Interview_Date*	*Interview_Time*	*Staff_No*	*Room_No*
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

Client_No	*Interview_Date*	*Interview_Time*	*Staff_No*
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

Staff_No	*Interview_Date*	*Room_No*
SG5	13-May-98	G101
SG37	13-May-98	G102
SG5	1-Jul-98	G102

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).

BCNF to 4NF Relations

Branch_Staff_Client relation

Branch_No	*SName*	*CName*
B3	Ann Beech	Aline Stewart
B3	David Ford	Aline Stewart
B3	Ann Beech	Mike Richie
B3	David Ford	Mike Richie

↓

Branch_Staff relation

Branch_No	*SName*
B3	Ann Beech
B3	David Ford

Branch_Client relation

Branch_No	*CName*
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

4NF to 5NF Relations

Property_Item_Supplier relation:

Property_No	Item_Description	Supplier_No
PG4	Bed	S1
PG4	Chair	S2
PG16	Bed	S2
PG16	Table	S1
PG36	Chair	S3

Property_Item relation:

Property_No	Item_Description
PG4	Bed
PG4	Chair
PG16	Bed
PG16	Table
PG32	Chair

Item_Supplier relation:

Item_Description	Supplier_No
Bed	S1
Chair	S2
Bed	S2
Table	S1
Chair	S3

Property_Supplier relation:

Property_No	Supplier_No
PG4	S1
PG4	S2
PG16	S2
PG16	S1
PG36	S3