Normalisation

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

Definition

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

  – For 3NF, it allows **A** → **B**
    - **B** is a primary key attribute and **A** is not a candidate key
  – For BCNF, it insists **A** → **B**
    - **B** is a primary key attribute and **A** must be 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

Example

Client interviews by members of staff (ClientInterview relation)

clientNo	interviewDate	interviewTime	staffNo	roomNo
CR76	13-May-02	10.30	SG5	G101
CR56	13-May-02	12.00	SG5	G101
CR74	13-May-02	12.00	SG37	G102
CR56	1-Jul-02	10.30	SG5	G102

Assumptions

Client interviews by members of staff (ClientInterview relation)

clientNo	interviewDate	interviewTime	staffNo	roomNo
CR76	13-May-02	10.30	SG5	G101
CR56	13-May-02	12.00	SG5	G101
CR74	13-May-02	12.00	SG37	G102
CR56	1-Jul-02	10.30	SG5	G102

- Members of staff involved in interviewing clients are allocated to a specific room on the day of interview

Assumptions

Client interviews by members of staff (ClientInterview relation)

clientNo	interviewDate	interviewTime	staffNo	roomNo
CR76	13-May-02	10.30	SG5	G101
CR56	13-May-02	12.00	SG5	G101
CR74	13-May-02	12.00	SG37	G102
CR56	1-Jul-02	10.30	SG5	G102

- A room may be allocated to several members of staff as required throughout a working day

Assumptions

Client interviews by members of staff (ClientInterview relation)

clientNo	interviewDate	interviewTime	staffNo	roomNo
CR76	13-May-02	10.30	SG5	G101
CR56	13-May-02	12.00	SG5	G101
CR74	13-May-02	12.00	SG37	G102
CR56	1-Jul-02	10.30	SG5	G102

- A client is only interviewed once on a given date, but may be requested to attend further interviews at later dates

FDs

clientNo	interviewDate	interviewTime	staffNo	roomNo
CR76	13-May-02	10.30	SG5	G101
CR56	13-May-02	12.00	SG5	G101
CR74	13-May-02	12.00	SG37	G102
CR56	1-Jul-02	10.30	SG5	G102

- fd1 (primary key)

  clientNo, interviewDate → interviewTime, staffNo, roomNo

FDs

clientNo	interviewDate	interviewTime	staffNo	roomNo
CR76	13-May-02	10.30	SG5	G101
CR56	13-May-02	12.00	SG5	G101
CR74	13-May-02	12.00	SG37	G102
CR56	1-Jul-02	10.30	SG5	G102

- fd2 (candidate key)

  staffNo, interviewDate, interviewTime → clientNo

FDs

clientNo	interviewDate	interviewTime	staffNo	roomNo
CR76	13-May-02	10.30	SG5	G101
CR56	13-May-02	12.00	SG5	G101
CR74	13-May-02	12.00	SG37	G102
CR56	1-Jul-02	10.30	SG5	G102

- fd3 (candidate key)

  roomNo, interviewDate, interviewTime → staffNo, clientNo

FDs

clientNo	interviewDate	interviewTime	staffNo	roomNo
CR76	13-May-02	10.30	SG5	G101
CR56	13-May-02	12.00	SG5	G101
CR74	13-May-02	12.00	SG37	G102
CR56	1-Jul-02	10.30	SG5	G102

- fd4

  staffNo, interviewDate → roomNo

FDs

- fd1 (primary key)
  clientNo, interviewDate → interviewTime, staffNo, roomNo
- fd2 (candidate key)
  staffNo, interviewDate, interviewTime → clientNo
- fd3 (candidate key)
  roomNo, interviewDate, interviewTime → staffNo, clientNo
- fd4
  staffNo, interviewDate → roomNo

- In 3NF - why?

FDs

- fd1 (primary key)
  clientNo, interviewDate → interviewTime, staffNo, roomNo
- fd2 (candidate key)
  staffNo, interviewDate, interviewTime → clientNo
- fd3 (candidate key)
  roomNo, interviewDate, interviewTime → staffNo, clientNo
- fd4
  staffNo, interviewDate → roomNo

Not BCNF - staffNo, interviewDate is not a candidate key

Update Anomaly found

clientNo	interviewDate	interviewTime	staffNo	roomNo
CR76	13-May-02	10.30	SG5	G101
CR56	13-May-02	12.00	SG5	G101
CR74	13-May-02	12.00	SG37	G102
CR56	1-Jul-02	10.30	SG5	G102

- To change the room number for staff number SG5 on the 13-May-02
  - we must update two tuples because of redundancy

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

- We create two new relations :

  Interview (clientNo, interviewDate, interviewTime, staffNo)

  StaffRoom (staffNo, interviewDate, roomNo)

- However, we lost fd3

  roomNo, interviewDate, interviewTime → staffNo, clientNo

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

Desirable to get to BCNF?

- Whether it is better to stop the normalization at 3NF or progress to BCNF?
- Depends on the amount of redundancy resulting from the presence of fd4 and the significance of the loss in fd3

Suggestions

- Stay in 3NF if
  - members of staff conduct only one interview per day
- Transform to BCNF if
  - members of staff conduct numerous interviews per day;
  - lost of fd3 is not significant

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