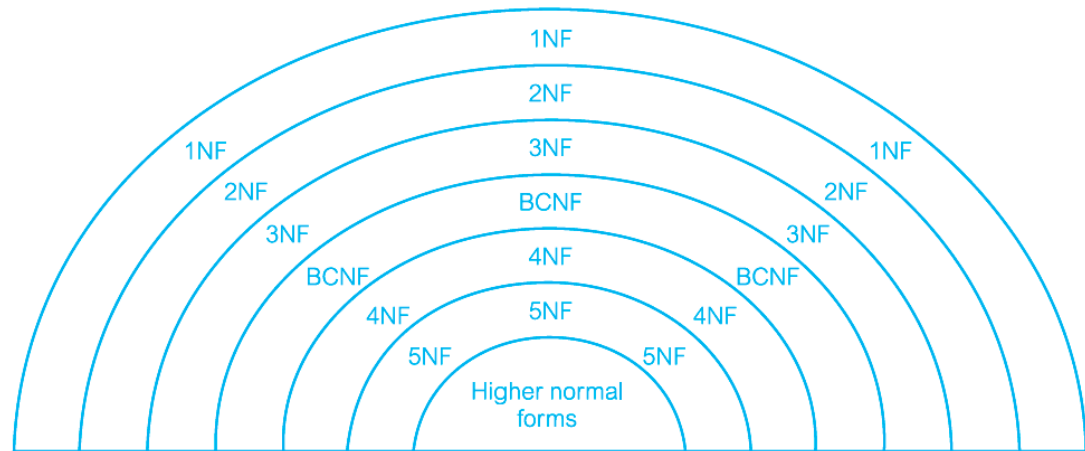# Design: Normalization

Textbook (new edition), Chapter 14 & 15

Normalization

- Analysis of functional dependencies can help us to appropriately decompose relations, a process known as *normalization*

- These decompositions should:
  - Reduce update anomalies
  - Reduce redundancy

- Progressive levels of decomposition we can employ (normal forms)

1NF: No Multi-Valued Attributes

- First normal form:
  - A relation in which the intersection of each row and column contains one and only one value
  - Essentially, no multi-valued attributes

- We have already required this in our definition of a relation; how do we enforce it?

- Sample cases: two majors, two telephone numbers…

1NF: Example

Student

studentID	lastName	firstName	year	major	GPA
1123	Smith	Robert	4	CSC, MTH	3.5
1129	Jones	Douglas	3	MTH, POL	2.9
1145	Brady	Susan	4	CSC	3.8

An approach to get in 1NF (resolve the multi-valued attributes):
- Have multiple rows with repeated entries

Student

studentID	lastName	firstName	year	major	GPA
1123	Smith	Robert	4	CSC	3.5
1123	Smith	Robert	4	MTH	3.5
1129	Jones	Douglas	3	MTH	2.9
1129	Jones	Douglas	3	POL	2.9
1145	Brady	Susan	4	CSC	3.8

1NF: Example

Student

studentID	lastName	firstName	year	major	GPA
1123	Smith	Robert	4	CSC, MTH	3.5
1129	Jones	Douglas	3	MTH, POL	2.9
1145	Brady	Susan	4	CSC	3.8

A second (preferred) approach to get in 1NF (resolve the multi-valued attributes):
-   Break off multi-valued attribute and original key attribute(s) into separate table

Student

studentID	lastName	firstName	year	GPA
1123	Smith	Robert	4	3.5
1129	Jones	Douglas	3	2.9
1145	Brady	Susan	4	3.8

studentID	major
1123	CSC
1123	MTH
1129	MTH
1129	POL
1145	CSC

2NF: Primary Key Driven

- 2NF transformations apply to relations with composite primary keys

- If a relation has a primary key that is a single attribute, it is already in 2NF

- Note, the primary key attribute is the attribute set such that all other attributes are functionally dependent on the attribute set

2NF: Primary Key Driven

- A relation is in 2NF if
  - It is in 1NF   AND
  - All non primary key attributes are *fully* functionally dependent on the primary key

- An example of a relation in 1NF but not in 2NF

staffNo	dentistName	patNo	patName	appointment date time	surgeryNo
S1011	Tony Smith	P100	Gillian White	12-Sep-04 10.00	S15
S1011	Tony Smith	P105	Jill Bell	12-Sep-04 12.00	S15
S1024	Helen Pearson	P108	Ian MacKay	12-Sep-04 10.00	S10
S1024	Helen Pearson	P108	Ian MacKay	14-Sep-04 14.00	S10
S1032	Robin Plevin	P105	Jill Bell	14-Sep-04 16.30	S15
S1032	Robin Plevin	P110	John Walker	15-Sep-04 18.00	S13

PK: staffNo, date, time   OTHERS partial on pk?
dentistName only relies on staffNo so not ffd on pk

2NF Transformation

- To transform to 2NF,
  - Remove attributes that are partially dependent on the PK, and their determinant, to separate tables

staffNo	dentistName
S1011	Tony Smith
S1024	Helen Pearson
S1032	Robin Plevin

staffNo	patientNo	patientName	date	time
S1011	P100	Gillian White	12-Sep-04	10.00
S1011	P015	Jill Bell	12-Sep-04	12.00
S1024	P108	Ian MacKay	12-Sep-04	10.00
S1024	P108	Ian MacKay	14-Sep-04	14.00
S1032	P105	Jill Bell	14-Sep-04	16.30
S1032	P110	John Walker	15-Sep-04	18.00

staffNo	Date	surgeryNo
S1011	12-Sep-04	S15
S1024	12-Sep-04	S10
S1024	14-Sep-04	S10
S1032	14-Sep-04	S15
S1032	15-Sep-04	S13

Transitive Dependencies

- *Transitive dependency definition*: If A, B, and C are attributes of a relation such that A→B and B→C, then A→ C (via B), provided A is not functionally dependent on B or C

StaffBranch

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

Functional Dependencies:
staffNo → branchNo
staffNo → branchAddress
branchNo → branchAddress
branchAddress → branchNo

Transitive dependencies:

staffNo →$_{transitive}$ branchAddress (via branchNo)

staffNo →$_{transitive}$ branchNo (via branchAddress)
given that branchNo and branchAddress do not 1:1 map to staffNo (which they don't)

Transitive Dependencies

- What, if any, arguments can we make about transitive dependencies that exist in this table?

staffNo	patientNo	patientName	date	time
S1011	P100	Gillian White	12-Sep-04	10.00
S1011	P015	Jill Bell	12-Sep-04	12.00
S1024	P108	Ian MacKay	12-Sep-04	10.00
S1024	P108	Ian MacKay	14-Sep-04	14.00
S1032	P105	Jill Bell	14-Sep-04	16.30
S1032	P110	John Walker	15-Sep-04	18.00

staffNo, date, time → patientName (via patientNo)   *sanity check patientInfo does not*
staffNo, date, time → patientNo (via patientName)   *1:1 map to staffNo, date, time*

3NF

- A relation is in 3NF if it is:
    - In 1NF and 2NF   AND
    - No non-primary-key attribute is transitively dependent on the primary key

- Transformation:
    - Move the transitively dependent attribute and its determinant (the "via") to another table

3NF Transformation

- Convert this 2NF version to 3NF version

staffNo	dentistName
S1011	Tony Smith
S1024	Helen Pearson
S1032	Robin Plevin

staffNo	patientNo	patientName	date	time
S1011	P100	Gillian White	12-Sep-04	10.00
S1011	P015	Jill Bell	12-Sep-04	12.00
S1024	P108	Ian MacKay	12-Sep-04	10.00
S1024	P108	Ian MacKay	14-Sep-04	14.00
S1032	P105	Jill Bell	14-Sep-04	16.30
S1032	P110	John Walker	15-Sep-04	18.00

staffNo	Date	surgeryNo
S1011	12-Sep-04	S15
S1024	12-Sep-04	S10
S1024	14-Sep-04	S10
S1032	14-Sep-04	S15
S1032	15-Sep-04	S13

3NF Transformation

Primary keys in red boxes
Many foreign key references

Patient

patientNo	patientName
P100	Gillian White
P015	Jill Bell
P108	Ian MacKay
P110	John Walker

Staff

staffNo	dentistName
S1011	Tony Smith
S1024	Helen Pearson
S1032	Robin Plevin

staffNo	patientNo	date	time
S1011	P100	12-Sep-04	10.00
S1011	P015	12-Sep-04	12.00
S1024	P108	12-Sep-04	10.00
S1024	P108	14-Sep-04	14.00
S1032	P105	14-Sep-04	16.30
S1032	P110	15-Sep-04	18.00

Appointment

staffNo	Date	surgeryNo
S1011	12-Sep-04	S15
S1024	12-Sep-04	S10
S1024	14-Sep-04	S10
S1032	14-Sep-04	S15
S1032	15-Sep-04	S13

StaffAssignment

Practice

- Given the following relation and domain description (and reasonable assumptions)
  - Identify functional dependencies
  - Identify and verify a primary key
  - Decompose the table as appropriate to 3NF

id	contract	hours Worked	name	hotel Number	hotel Location
1135	C1024	16	Smith J	H25	East Kilbride
1057	C1024	24	Hocine D	H25	East Kilbride
1068	C1024	28	White T	H4	Glasgow
1135	C1025	15	Smith J	H4	Glasgow

Context: *Database of a temp-staffing agency that staffs hotels. Provides national employee id, contract working under, hours worked under that contract, name, hotel worked at, and location of hotel*

- Functional Dependencies?
  - id,contract → hoursWorked
  - id → name
  - contract → hotelNumber
  - contract → hotelLocation
  - hotelNumber → hotelLocation

id	contract	hours Worked	name	hotel Number	hotel Location
1135	C1024	16	Smith J	H25	East Kilbride
1057	C1024	24	Hocine D	H25	East Kilbride
1068	C1025	28	White T	H4	Glasgow
1135	C1025	15	Smith J	H4	Glasgow

Context: *Database of a temp-staffing agency that staffs hotels. Provides national employee id, contract working under, hours worked under that contract, name, hotel worked at, and location of hotel*

Practice

- Primary Key?
  - id,contract → hoursWorked
  - id → name
  - contract → hotelNumber
  - contract → hotelLocation
  - hotelNumber → hotelLocation

PK: id, contract

id	contract	hours Worked	name	hotel Number	hotel Location
1135	C1024	16	Smith J	H25	East Kilbride
1057	C1024	24	Hocine D	H25	East Kilbride
1068	C1025	28	White T	H4	Glasgow
1135	C1025	15	Smith J	H4	Glasgow

Context: *Database of a temp-staffing agency that staffs hotels. Provides national employee id, contract working under, hours worked under that contract, name, hotel worked at, and location of hotel*

Practice

- 3NF Decomposition: 2NF First! (no partial primary key fds)

id	contract	hours Worked
1135	C1024	16
1057	C1024	24
1068	C1024	28
1135	C1025	15

id	name
1135	Smith J
1057	Hocine D
1068	White T

contract	hotel Number	hotel Location
C1024	H25	East Kilbride
C1025	H4	Glasgow

Context: *Database of a temp-staffing agency that staffs hotels. Provides national employee id, contract working under, hours worked under that contract, name, hotel worked at, and location of hotel*

Practice

- 3NF Decomposition: no transitive primary key fds

id	contract	hours Worked
1135	C1024	16
1057	C1024	24
1068	C1024	28
1135	C1025	15

id	name
1135	Smith J
1057	Hocine D
1068	White T

contract	hotel Number
C1024	H25
C1025	H4

hotel Number	hotel Location
H25	East Kilbride
H4	Glasgow

Context: *Database of a temp-staffing agency that staffs hotels. Provides national employee id, contract working under, hours worked under that contract, name, hotel worked at, and location of hotel*

Generalized 2NF & 3NF

- Technically, should extend beyond functional dependencies with just primary key

- Generalized 2NF: In 1NF and every non-candidate-key attribute is fully functionally dependent on any candidate key *(a candidate key attribute is part of any candidate key, so non-candidate key attributes are those not involved in any candidate keys)*

- Generalized 3NF: In 1NF, 2NF, and no non-candidate-key attributes are transitively dependent on any candidate key

- Rare, though not impossible, to have redundancy that stem from candidate keys beyond the primary key

Beyond 3NF

- Normalizing to 3NF is often "very good", though there exist additional cases where redundancy/update anomalies occur.

- We'll start with Boyce-Codd Normal Form (BCNF)
  - Generalized 3NF   AND
  - Every determinant (LHS of a FD) must be a candidate key

BCNF

- What will a functional dependency look like that could make it past 3NF but violate the BCNF principle?
  - (2NF) Every non-candidate-key attribute is fully functionally dependent on any (some) candidate key.
  - (3NF) No non-candidate-key attributes are transitively dependent on any candidate key
  - (BCNF) Every determinant (LHS of a FD) must be a candidate key

The 2NF and 3NF rules make no constraints on how candidate key attributes could be functionally dependent on other attributes. An example?

BCNF

Staff interview clients. On a given day, a staff member will interview in one room, though multiple staff could hold interviews in that room. A client is interviewed at most once a day, but could have multiple interviews on different days.

clientNo	date	time	staffNo	roomNo
CR76	13-May-09	10.30	SG5	G101
CR56	13-May-09	12.00	SG5	G101
CR74	13-May-09	12.00	SG37	G102
CR56	1-Jul09	10.30	SG5	G102

What are the candidate keys (unique tuple identifiers)?

(clientNo, date)   // let's use as PK
(staffNo, date, time)
(roomNo, date, time)

BCNF

Candidate keys:
(clientNo, date)
(staffNo, date, time)
(roomNo, date, time)

Staff interview clients. On a given day, a staff member will interview in one room, though multiple staff could hold interviews in that room. A client is interviewed at most once a day, but could have multiple interviews on different days.

clientNo	date	time	staffNo	roomNo
CR76	13-May-09	10.30	SG5	G101
CR56	13-May-09	12.00	SG5	G101
CR74	13-May-09	12.00	SG37	G102
CR56	1-Jul09	10.30	SG5	G102

What are the functional dependencies?

clientNo, date → time, staffNo, roomNo   // property of PK/CK
staffNo, date, time → clientNo   // property of PK/CK
staffNo, date → roomNo   // subset of PK/CK
roomNo, date, time → staffNo, clientNo   // property of PK/CK

I took the (accepted) liberty here of writing multiple FDs on the same line
(RHS can be split, left can't)

BCNF

Candidate keys:
(clientNo, date)
(staffNo, date, time)
(roomNo, date, time)

clientNo	date	time	staffNo	roomNo
CR76	13-May-09	10.30	SG5	G101
CR56	13-May-09	12.00	SG5	G101
CR74	13-May-09	12.00	SG37	G102
CR56	1-Jul09	10.30	SG5	G102

clientNo, date → time, staffNo, roomNo   // property of PK/CK
staffNo, date, time → clientNo   // property of PK/CK
staffNo, date → roomNo   // subset of PK/CK
roomNo, date, time → staffNo, clientNo   // property of PK/CK

(2NF) Every non-candidate-key attribute is fully functionally dependent
on any (some) candidate key: *trivially holds* – there are no non-candidate-key attributes
(3NF) No non-candidate-key attributes are transitively dependent on any candidate key:
trivially holds for same reason
(BCNF) Every determinant (LHS of a FD) must be a candidate key: *does not hold* for
    staffNo, date → roomNo   *(staffNo, date is not a candidate key)*

BCNF

Candidate keys:
(clientNo, date)
(staffNo, date, time)
(roomNo, date, time)

clientNo	date	time	staffNo	roomNo
CR76	13-May-09	10.30	SG5	G101
CR56	13-May-09	12.00	SG5	G101
CR74	13-May-09	12.00	SG37	G102
CR56	1-Jul09	10.30	SG5	G102

(BCNF) Every determinant (LHS of a FD) must be a candidate key: *does not hold* for staffNo, date → roomNo   *(staffNo, date is not a candidate key)*

This means you have a (set of) attribute values that may appear across multiple rows [(staffNo, date) is not unique] but which, given a value, will map to the same value, requiring multiple rows to be updated if a change is made

In this case, changing where a staff member is to hold interviews on a given date

BCNF

- Such a violation can only occur under the following conditions:
  - The relation contains two or more composite candidate keys
  - The candidate keys overlap with at least one attribute in common

  Thus, leaving the door open for a candidate key attribute being functionally dependent on *parts* of another candidate key

BCNF Transformation

- Remove the violating functional dependency into its own table (staffNo, date → roomNo)

clientNo	date	time	staffNo	roomNo
CR76	13-May-09	10.30	SG5	G101
CR56	13-May-09	12.00	SG5	G101
CR74	13-May-09	12.00	SG37	G102
CR56	1-Jul09	10.30	SG5	G102

InterviewAppointments

clientNo	date	time	staffNo
CR76	13-May-09	10.30	SG5
CR56	13-May-09	12.00	SG5
CR74	13-May-09	12.00	SG37
CR56	1-Jul09	10.30	SG5

StaffLocations

date	staffNo	roomNo
13-May-09	SG5	G101
13-May-09	SG37	G102
1-Jul09	SG5	G102

BCNF Practice

- Is this 3NF decomposition of the hotel-contract-workers relation in BCNF?

id	contract	hours Worked
1135	C1024	16
1057	C1024	24
1068	C1024	28
1135	C1025	15

id	name
1135	Smith J
1057	Hocine D
1068	White T

contract	hotel Number
C1024	H25
C1025	H4

hotel Number	hotel Location
H25	East Kilbride
H4	Glasgow

Yes, by default, as no relation contains more than one candidate key.

BCNF Practice

Patient

patientNo	patientName
P100	Gillian White
P015	Jill Bell
P108	Ian MacKay
P110	John Walker

Primary keys in red boxes
Many foreign key references

Staff

staffNo	dentistName
S1011	Tony Smith
S1024	Helen Pearson
S1032	Robin Plevin

staffNo	patientNo	date	time
S1011	P100	12-Sep-04	10.00
S1011	P015	12-Sep-04	12.00
S1024	P108	12-Sep-04	10.00
S1024	P108	14-Sep-04	14.00
S1032	P105	14-Sep-04	16.30
S1032	P110	15-Sep-04	18.00

Appointment

staffNo	Date	surgeryNo
S1011	12-Sep-04	S15
S1024	12-Sep-04	S10
S1024	14-Sep-04	S10
S1032	14-Sep-04	S15
S1032	15-Sep-04	S13

Is this 3NF in BCNF?

StaffAssignment

BCNF Practice

- All tables except the one below only have one candidate key

staffNo	patientNo	date	time
S1011	P100	12-Sep-04	10.00
S1011	P015	12-Sep-04	12.00
S1024	P108	12-Sep-04	10.00
S1024	P108	14-Sep-04	14.00
S1032	P105	14-Sep-04	16.30
S1032	P110	15-Sep-04	18.00

Appointment

Two candidate keys:

--staffNo, date, time
--patientNo, date, time

Two functional dependencies:

staffNo, date, time → patientNo
patientNo, date, time → staffNo

For this relation, all determinants are candidate keys – so it is in BCNF

Next Week

- Tuesday: I hope to have
  - Grading done (at least of CREATE/ER project)
  - A sample test made
  - A study guide made
- Thursday: I won't be here (SIGCSE)
  - Work during class time on project as needed
    - ER diagram, ER description, 5-10 English queries
    - ER diagram, ER description: Similar to what was given you/you developed for the first Hotel assignment
    - An English Query: "List the title of all of Prof. Turkett's books that have been loaned out and the names of who has them".