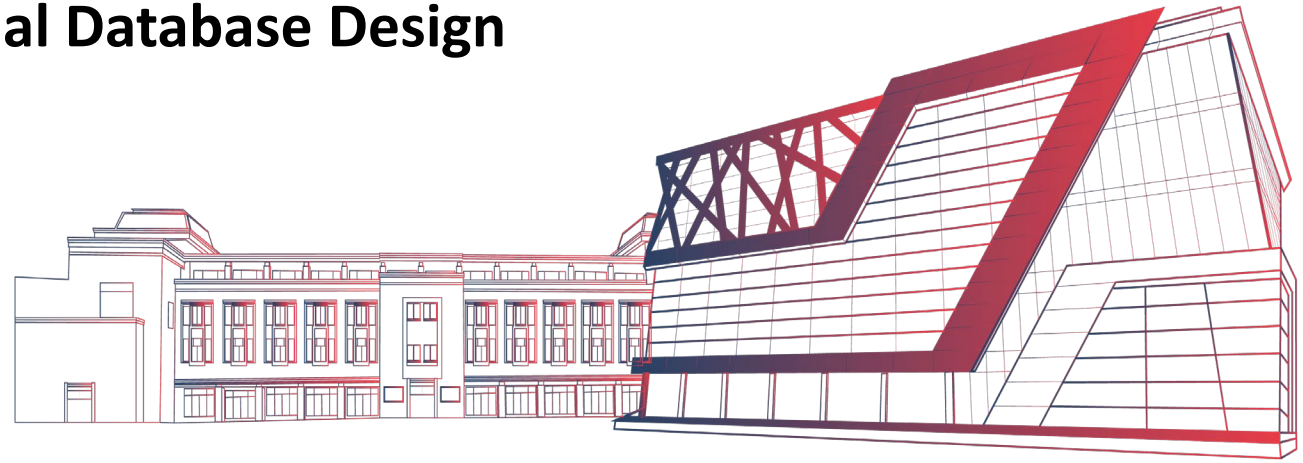
THE NORTHCAP UNIVERSITY

UNIT III

Relational Database Design



30 YEARS OF EXCELLENCE

NAAC Accredited

Lecture 17

THE NORTHCAP UNIVERSITY

Normalization & Normal Forms

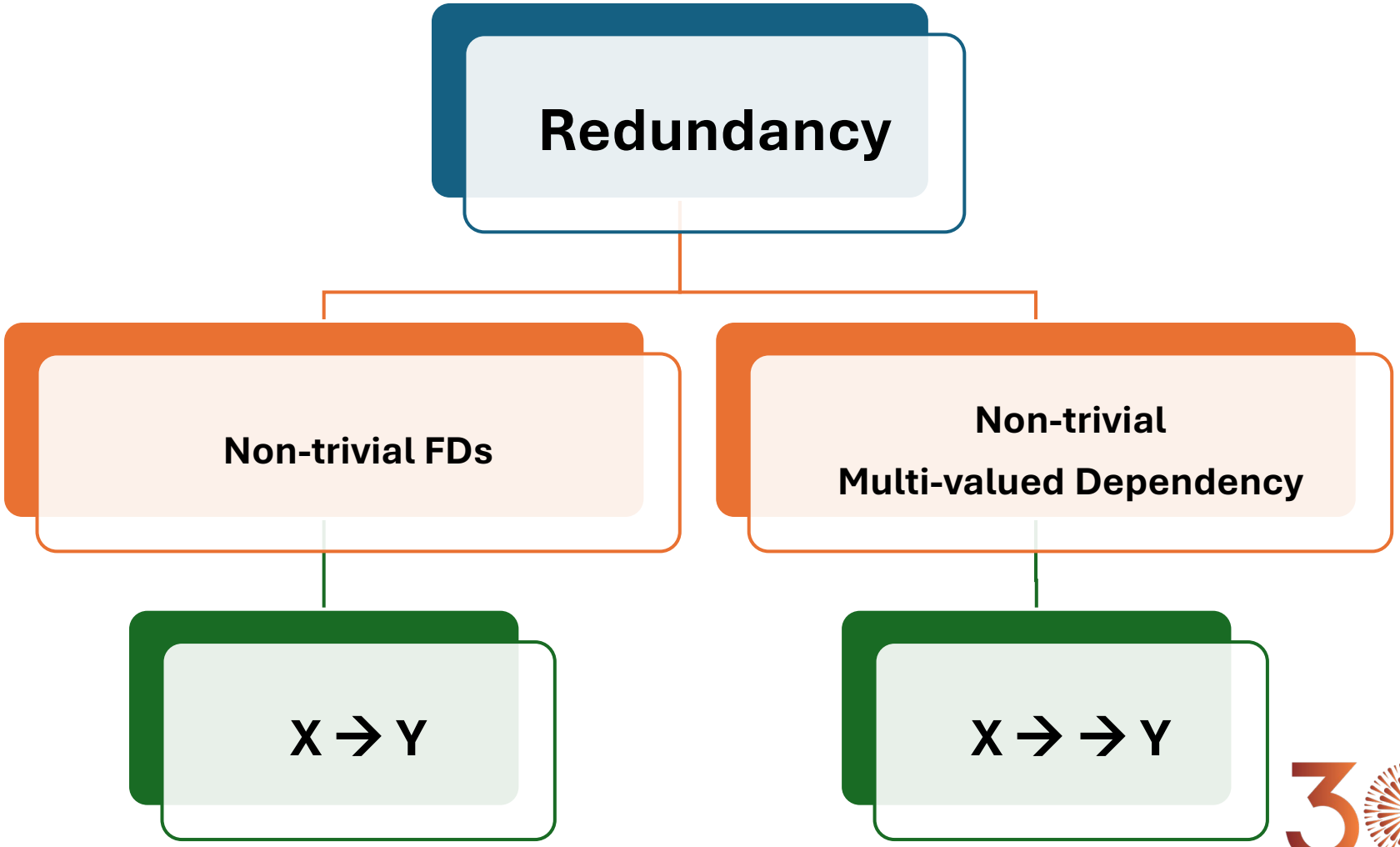
30 YEARS OF EXCELLENCE

NAAC Accredited

Normalization

- **Normalization:** The process of decomposing unsatisfactory "bad" relations by breaking up their attributes into smaller relations
- **Denormalization:** the process of storing the join of higher normal form relations as a base relation—which is in a lower normal form

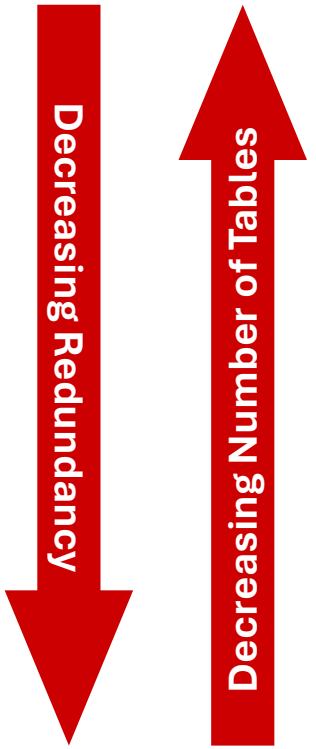
Normalization (Cont.)



Normal Forms

THE NORTHCAP UNIVERSITY

- Levels of normalization based on the amount of redundancy in the database.
- Various levels of normalization are:
  - **First Normal Form (1NF)**
  - **Second Normal Form (2NF)**
  - **Third Normal Form (3NF)**
  - **Boyce-Codd Normal Form (BCNF)**
  - **Fourth Normal Form (4NF)**
  - **Fifth Normal Form (5NF)**

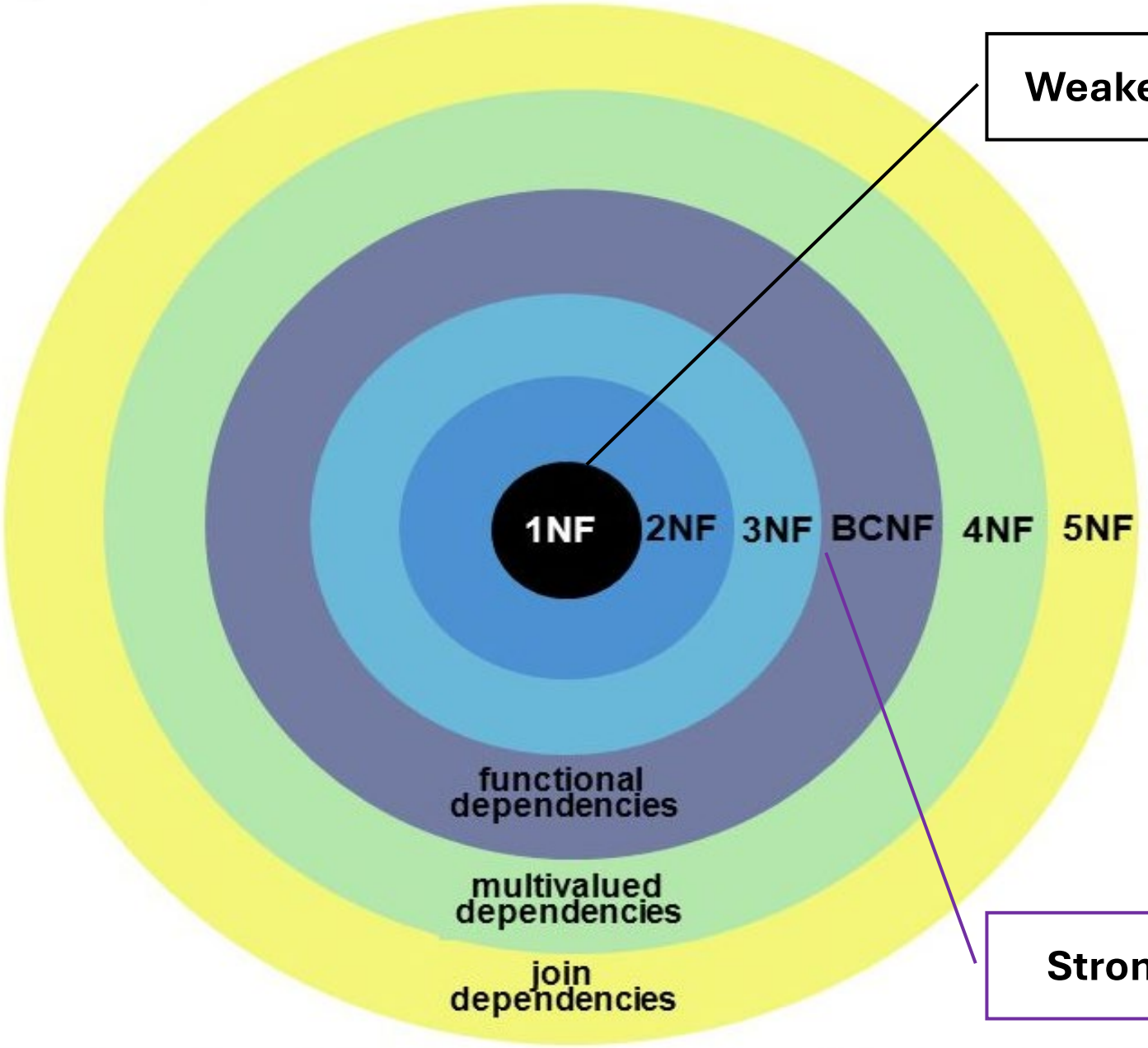


Most databases should be 3NF or BCNF in order to avoid the database anomalies.

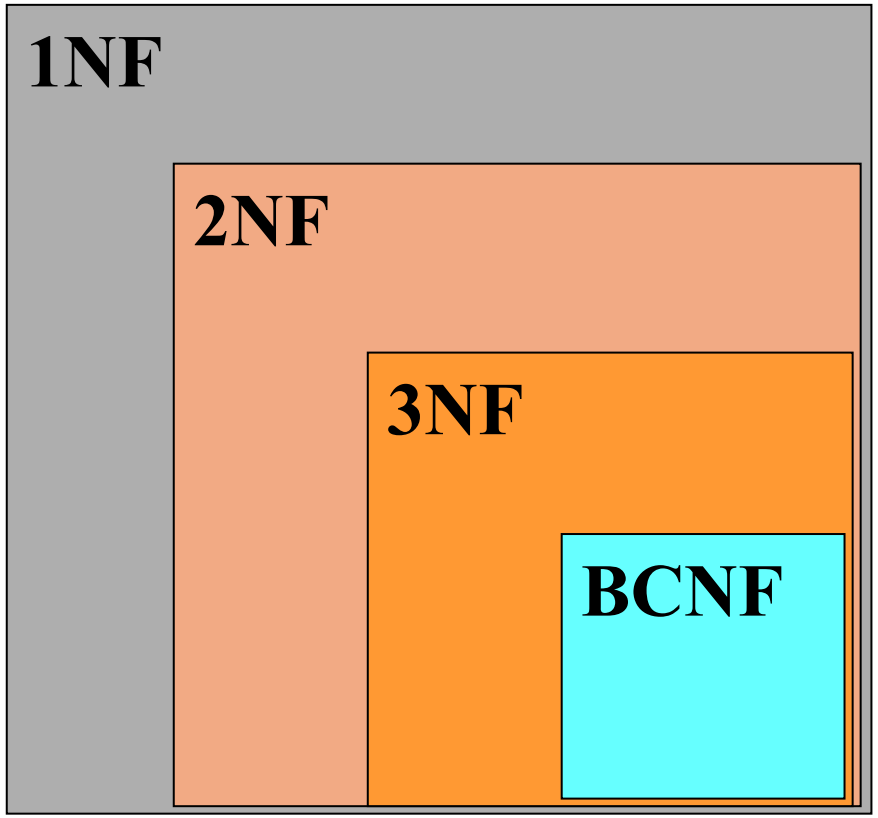
30 YEARS OF EXCELLENCE

NAAC Accredited

Hierarchy of Normal Forms



Normalization



- *a relation in BCNF, is also in 3NF*
- *a relation in 3NF is also in 2NF*
- *a relation in 2NF is also in 1NF*

First Normal Form

- A relation is in **1NF** if all values stored in the relation are single-valued and atomic.

i.e. 1NF places restrictions on the structure of relations. Values must be simple.

- It implies no multi-valued dependencies allowed.

All single valued attributes

DEPARTMENT

Dname	Dnumber	Mgr_ssn	Mgr_start_date
Research	5	333445555	1988-05-22
Administration	4	987654321	1995-01-01
Headquarters	1	888665555	1981-06-19

Normalization into 1NF

THE NORTHCAP UNIVERSITY

DEPARTMENT

Dname	Dnumber	Dmgr_ssn	Dlocations
Research	5	333445555	{Bellaire, Sugarland, Houston}
Administration	4	987654321	{Stafford}
Headquarters	1	888665555	{Houston}

DEPARTMENT

Dname	Dnumber	Dmgr_ssn	Dlocation
Research	5	333445555	Bellaire
Research	5	333445555	Sugarland
Research	5	333445555	Houston
Administration	4	987654321	Stafford
Headquarters	1	888665555	Houston

Multi – valued entries changed to single valued

30 YEARS OF EXCELLENCE

NAAC Accredited

Second Normal Form

- A relation is in **2NF** if it is in 1NF, and every non-key attribute is fully dependent on each candidate key i.e. no partial FD.

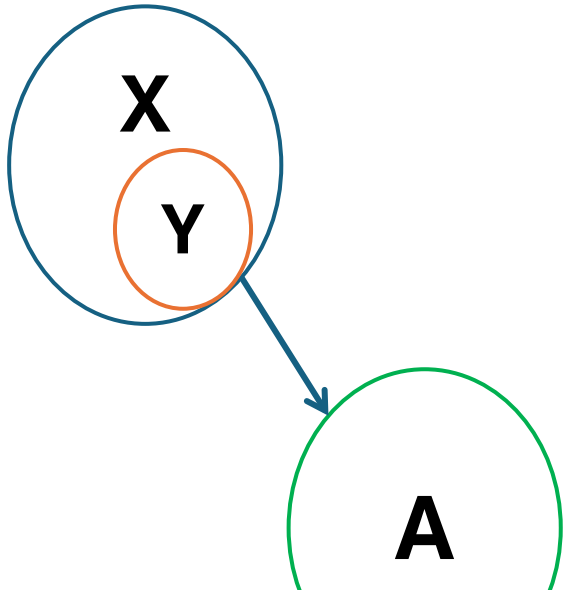
- **Partial Dependency**

  X: any candidate key

  Y: proper subset of candidate key

  A: non-prime attribute

  Y → A is partial dependency



Second Normal Form Example

1. **{SSN, PNUMBER} → HOURS is a full FD**

   ∵ neither SSN → HOURS

   nor PNUMBER → HOURS hold

2. **{SSN, PNUMBER} → ENAME is PD**

   ∵ SSN → ENAME holds

Third Normal Form

- A relation R is in 3NF iff every non-trivial FD X → Y in R

  a) **X is superkey (or)**

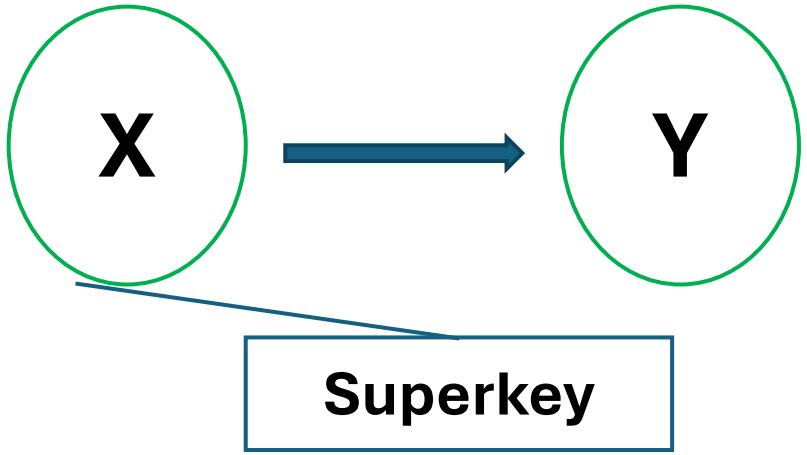
  b) **Y is prime attribute**

Superkey (or) Prime Attribute

BCNF Normal Form

(Boyce-Codd Normal Form)

- A relation schema R is in **Boyce-Codd Normal Form** (**BCNF**) if whenever an FD X → A holds in R, then **X is a superkey of R**



- There exist relations that are in 3NF but not in BCNF
- The goal is to have each relation in BCNF (or 3NF)

Practice Drill

- Consider the following four relational schemas. For each schema, all non-trivial functional dependencies are listed, The underlined attributes are the respective primary keys.

Schema I:

Registration (rollno, courses)

rollno → courses

Schema II:

Registration (rollno, coursid, email)

rollno, courseid → email

email → rollno

Practice Drill (Cont.)

Schema III:

Registration (rollno, courseid, marks, grade)

rollno, courseid, → marks, grade

marks → grade

Schema IV:

Registration (rollno, courseid, credit)

rollno, courseid → credit

courseid → credit

Which one of the relational schemas above is in 3NF but not in BCNF?

Solution Practice Drill

THE NORTHCAP UNIVERSITY

Schema I:

Registration(rollno, courses)
rollno → courses

Since, rollno is primary key, so this **relation is in BCNF as well as 3 NF.**

Schema II:

Registration (rollno, coursid, email)
rollno, courseid → email
email → rollno

- {rollno, coursid} is primary key so rollno and coursid are prime attributes. email is non-prime attribute.
- FD *rollno, courseid → email* is in BCNF and 3NF,
- FD *email → rollno* violates the rule of BCNF as email is not superkey, but 3NF is satisfied because rollno is prime attribute.

Thus, overall this relation is in 3 NF but not in BCNF.

30 YEARS OF EXCELLENCE | NAAC Accredited

Solution Practice Drill (Cont.)

THE NORTHCAP UNIVERSITY

Schema III:

Registration (rollno, courseid, marks, grade)
Non-trivial functional dependencies:
rollno, courseid, → marks, grade
marks → grade

- rollno, courseid is primary key, so rollno and courseid are prime attributes and marks and grade are non-prime attributes.
- FD *rollno, courseid, → marks, grade* satisfies BCNF as well as 3NF.
- FD *marks → grade* does not satifies 3 NF because nither marks is superkey nor grade is prime-attribute..

So, overall this relation is not in 3 NF and not in BCNF but it does not violate rule of 2 NF, so can be only in 2 NF.

30 YEARS OF EXCELLENCE | NAAC Accredited

Solution Practice Drill (Cont.)

THE NORTHCAP UNIVERSITY

Schema IV:

Registration (rollno, courseid, credit)

rollno, courseid → credit

courseid → credit

- rollno, courseid is primary key, so rollno and courseid are prime-attributes and credit is non-prime attribute.
- FD *rollno, courseid → credit* satifies BCNF as well as 3 NF.
- FD *courseid → credit* violates rule of 2 NF, so can not be in 2NF.

So, overall this is not in 2 NF, 3 NF, and BCNF. But it is only in 1 NF.

30 YEARS OF EXCELLENCE | NAAC Accredited

Practice Drill

Relation R has eight attributes ABCDEFGH. Fields of R contain only atomic values.

F = { CH→G, A→BC, B→CFH, E→A, F→EG } is a FD set

i) **How many candidate keys does the relation R have?**

(a) 3 (b) 4 (c) 5 (d) 6

ii) **The relation R is in which normal form?**

(a) in 1NF, but not in 2NF.

(b) in 2NF, but not in 3NF.

(c) in 3NF, but not in BCNF.

(d) in BCNF.

Practice Drill Solution

THE NORTHCAP UNIVERSITY

Ans i): Option (b)

A+:ABCFHGE

B+: BCFHEGA

C+:C

D+:D

E+: EABCFHG

F+:FEGABCH

G+:G

H+ : H

A+,B+,E+,F+ contains all attributes except D. Thus there are 4 candidate keys DA,DB,DE and DF.

Ans ii) Option (a)

- F → G; G is a non-prime attribute and F is a proper subset of a candidate key. This is a case of partial dependency. Hence 2NF condition is violated.
- A →C and B → CH also violates 2NF condition, hence R is not in 2NF.

Since attributes of relation R has only atomic values, R is in 1NF.

30 YEARS OF EXCELLENCE

NAAC Accredited

Practice Drill

THE NORTHCAP UNIVERSITY

1. Find the highest normal form of a relation R(A,B,C,D,E) with FD set as

   { BC → D, AC → BE, B → E }

1. Find the highest normal form of a relation R(A,B,C,D) with FD set as

   { AB → C, BC → D }

1. Find the highest normal form of a relation R(A,B,C,D,E,F) with FD set as

   { AB → C, C → D, D → AE, E → F, F → B }

30 YEARS OF EXCELLENCE | NAAC Accredited

Solution Practice Drill

1. **F = { BC → D, AC → BE, B → E }**

- (AC)+ ={A,C,B,E,D} but none of its subset can determine all attribute of relation, So AC will be candidate key. A or C can't be derived from any other attribute of the relation, so there will be only 1 candidate key {AC}.
- {A,C} – Prime and {B,D,E} – Non-prime
- The relation is in 2nd normal form because BC → D is in 2nd normal form (BC is not proper subset of candidate key AC)
- AC → BE is in 2nd normal form (AC is candidate key)
- B → E is in 2nd normal form (B is not a proper subset of candidate key AC).
- The relation is not in 3rd normal form because in BC → D (neither BC is a super key nor D is a prime attribute) and in B → E (neither B is a super key nor E is a prime attribute).

  So the highest normal form of relation will be 2nd Normal form.

Solution Practice Drill (Cont.)

2. **{ AB → C, BC → D }**

- AB+ = A, B, C, D
- A+ = A
- B+ = B
- BC+ = B, C
- Prime – A, B Non-prime – C, D
- Thus, because of BC → D this relation is not in BCNF and neither in 3NF
- **Thus, the relation is in 2NF**

Solution Practice Drill (Cont.)

3. **{ AB → C, C → D, D → AE, E → F, F → B }**

- AB+ = A, B, C, D, E, F A+ = A B+ = B
- AF+ = A, B, C, D, E, F F+ = F, B
- AE+ = A, B, C, D, E, F E+ = E, F, B
- AD+ = A, B, C, D, E, F D+ = A, B, C, D, E, F
- C+ = A, B, C, D, E, F
- Prime - A, B, C, D, E, F
- Because of E → F and F → B relation is not in BCNF
- **Thus, the relation is in 3NF**

THE NORTHCAP UNIVERSITY

Thanks!!

30 YEARS OF EXCELLENCE

NAAC Accredited