RDBMS and SQL
Normal Forms

Venkatesh Vinayakarao
venkateshv@cmi.ac.in
http://vvtesh.co.in

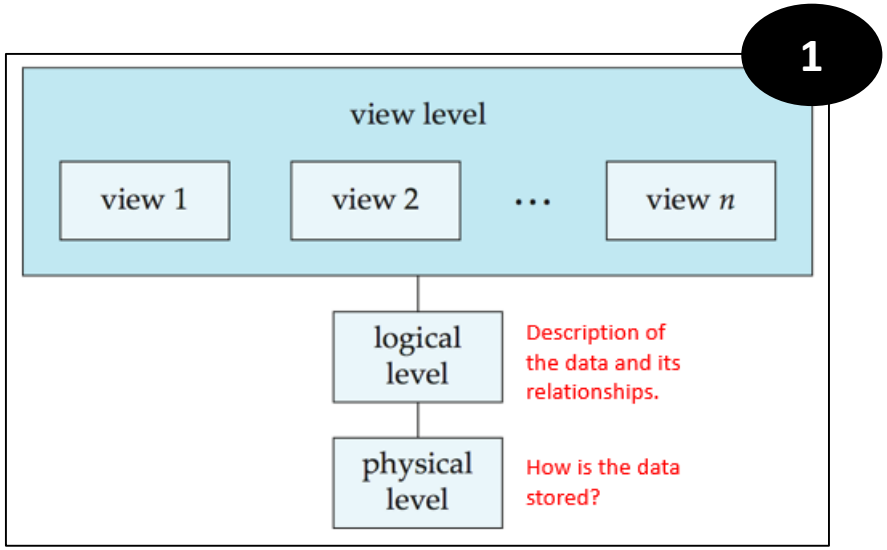
Chennai Mathematical Institute

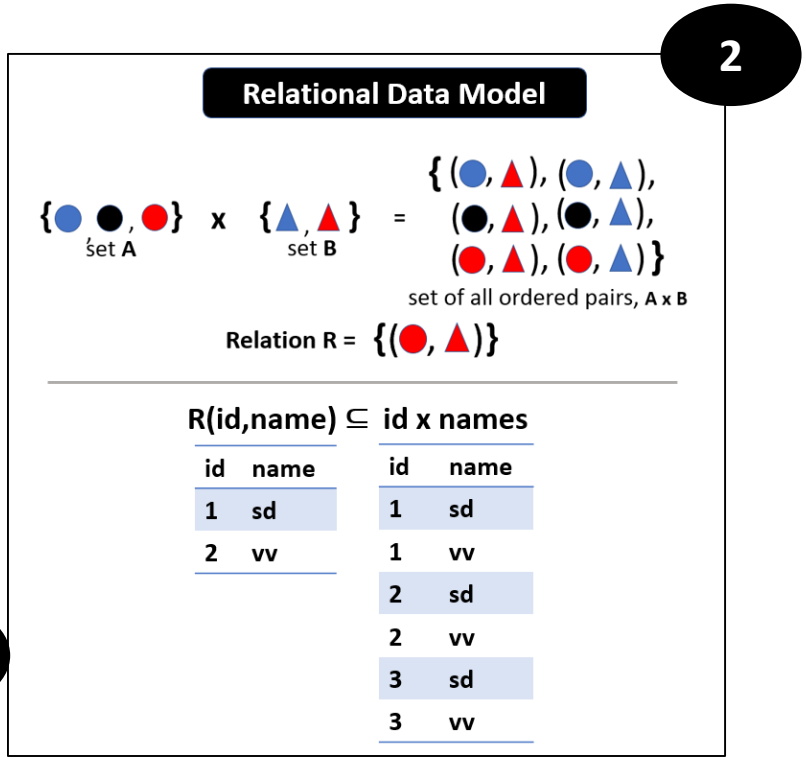
Slide contents are borrowed from the course text. For the authors' original version of slides, visit: https://www.db-book.com/db6/slide-dir/index.html.

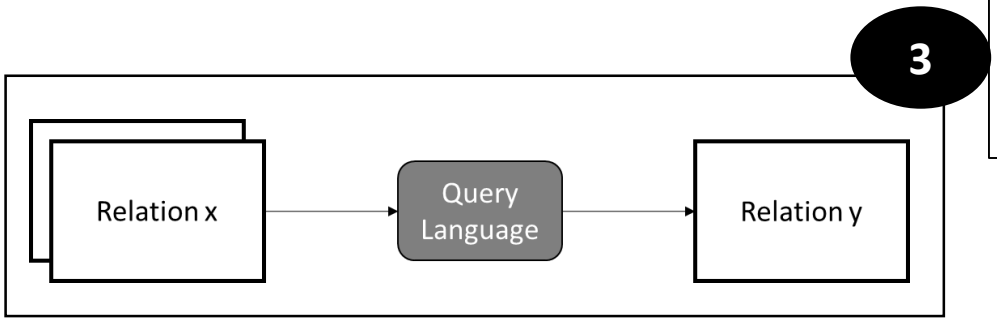
Agenda

- Review
- Normal Forms
  - BCNF
  - 4th Normal Form
  - Decomposition

Story So Far…

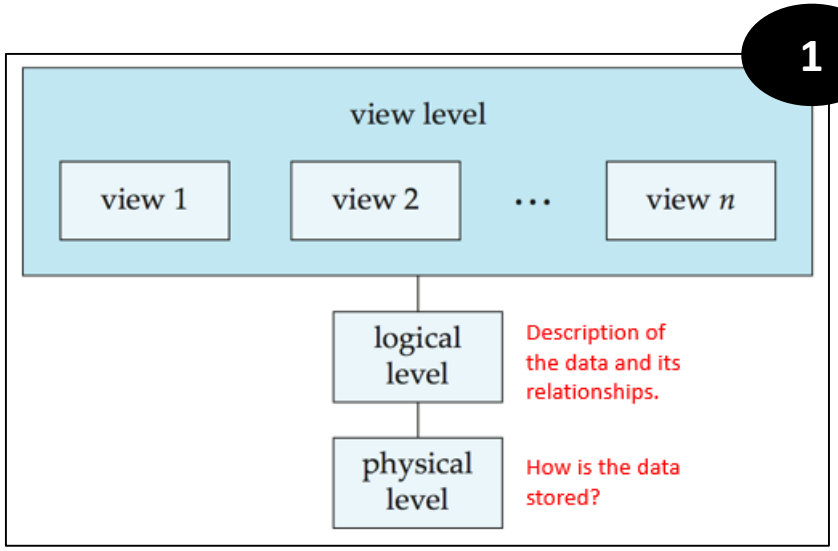


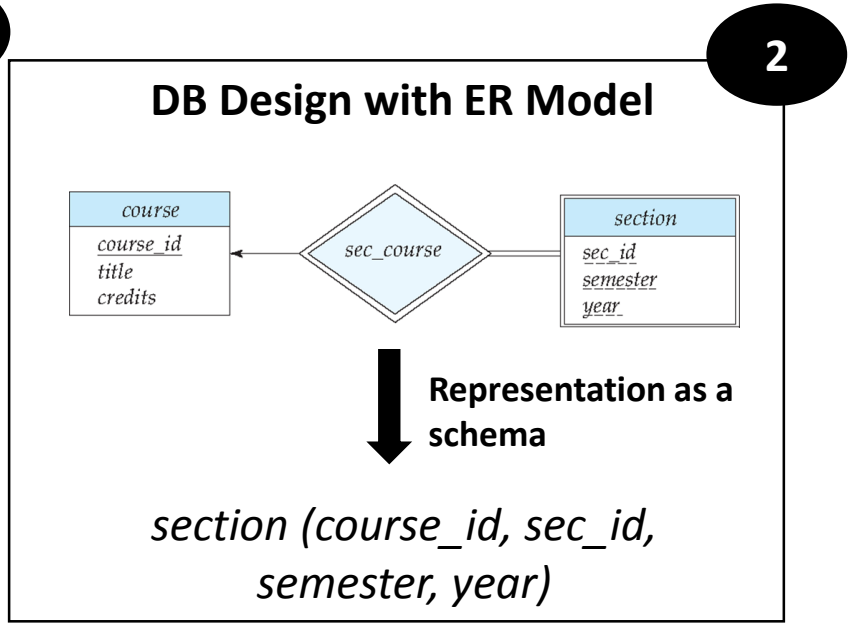




Relational Algebra
SQL

Story So Far…





Functional Dependency and Normal Forms

Prime Attribute

- All attributes that form the key are called prime attributes.

R1(ABCD)

AB is the key.
A and B are prime attributes.

Quiz

- Let R(A,B,C) be a relation where the candidate keys are AB and AC. We chose AB as the primary key. What are the prime attributes?

Quiz

- Let R(A,B,C) be a relation where the candidate keys are AB and AC. We chose AB as the primary key. What are the prime attributes?
  - Answer: A, B and C

Review: Second Normal Form (2NF)

- A relation is in second normal form iff
  - it is in 1nf, and
  - no non-prime attribute is functionally dependent on any proper subset of any candidate key.

No Partial Dependency!

Electric toothbrush models

Manufacturer	Model	Model full name	Manufacturer country
Forte	X-Prime	Forte X-Prime	Italy
Forte	Ultraclean	Forte Ultraclean	Italy
Dent-o-Fresh	EZbrush	Dent-o-Fresh EZbrush	USA
Brushmaster	SuperBrush	Brushmaster SuperBrush	USA
Kobayashi	ST-60	Kobayashi ST-60	Japan
Hoch	Toothmaster	Hoch Toothmaster	Germany
Hoch	X-Prime	Hoch X-Prime	Germany

Manufacturer → Manufacturer Country
{Manufacturer, Model} → Model full name

Quiz

- Is the relation R(A,B,C) in 2NF,
  - if B → C is a functional dependency?
- Assume AB and AC are candidate keys.
- Assume AB is the primary key.

Quiz

- Is the relation R(A,B,C) in 2NF,
  - if B → C is a functional dependency?
- Assume AB and AC are candidate keys.
- Assume AB is the primary key.

R is in 2NF

2NF requires that every ***non-prime*** attribute is fully dependent on every candidate key.

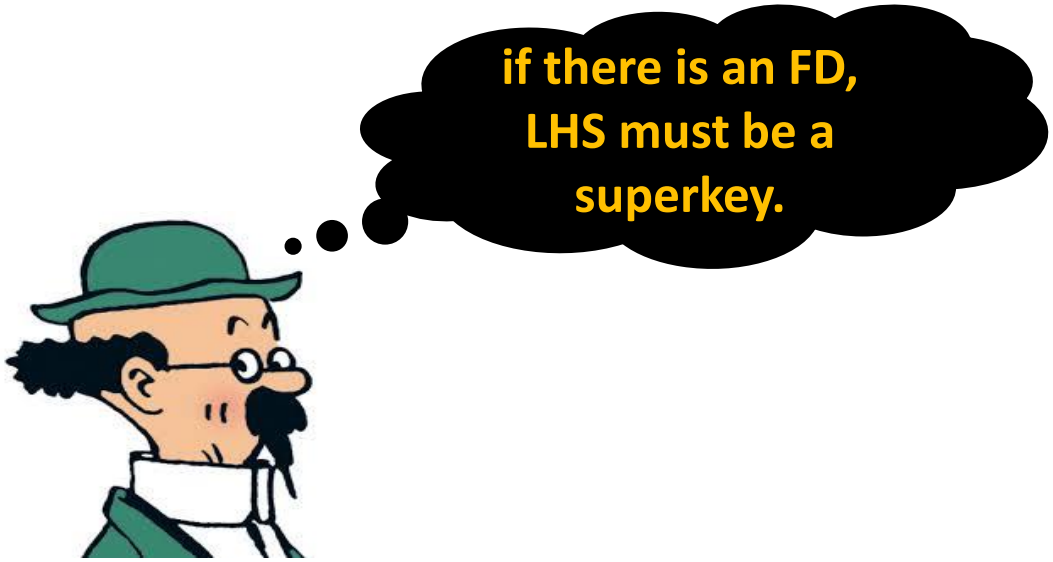
A,B and C are prime attributes! So, R is in 2NF.

Anomalies

- Delete
  - When deleting a value of an attribute, we should not lose the value of some other attribute.
- Insert
  - To insert the value of an attribute, we need to know the value of another attribute which we might not know.
  - Storing the same data at multiple relations may lead to consistency issues.
- Update
  - Same as insert, but for record updates.

Boyce-Codd Normal Form

- A relation schema *R* is in BCNF with respect to a set *F* of functional dependencies if for all functional dependencies in F^+ of the form $\alpha \rightarrow \beta$ where $\alpha \subseteq R$ and $\beta \subseteq R$, at least one of the following holds:
  - $\alpha \rightarrow \beta$ is trivial (i.e., $\beta \subseteq \alpha$)
  - α is a superkey for *R*



BCNF Violation

Today's court bookings

Court	Start time	End time	Rate type
1	09:30	10:30	SAVER
1	11:00	12:00	SAVER
1	14:00	15:30	STANDARD
2	10:00	11:30	PREMIUM-B
2	11:30	13:30	PREMIUM-B
2	15:00	16:30	PREMIUM-A

Relation is not in BCNF

- FD is non-trivial
- In FD, Rate type is not a superkey

Candidate Keys

S_1 = {Court, Start time}

S_2 = {Court, End time}

S_3 = {Rate type, Start time}

S_4 = {Rate type, End time}

Assume the FD is

Rate type → Court

RateTypes

- **SAVER**: for Court 1 bookings made by members
- **STANDARD**: for Court 1 bookings made by non-members
- **PREMIUM-A**: for Court 2 bookings made by members
- **PREMIUM-B**: for Court 2 bookings made by non-members

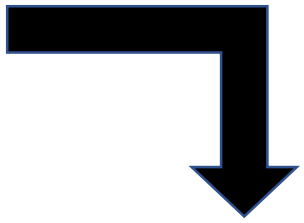
Is this relation in 2NF?

Yes, no non-prime attributes

Decompose to BCNF

Today's court bookings

Court	Start time	End time	Rate type
1	09:30	10:30	SAVER
1	11:00	12:00	SAVER
1	14:00	15:30	STANDARD
2	10:00	11:30	PREMIUM-B
2	11:30	13:30	PREMIUM-B
2	15:00	16:30	PREMIUM-A



Rate types

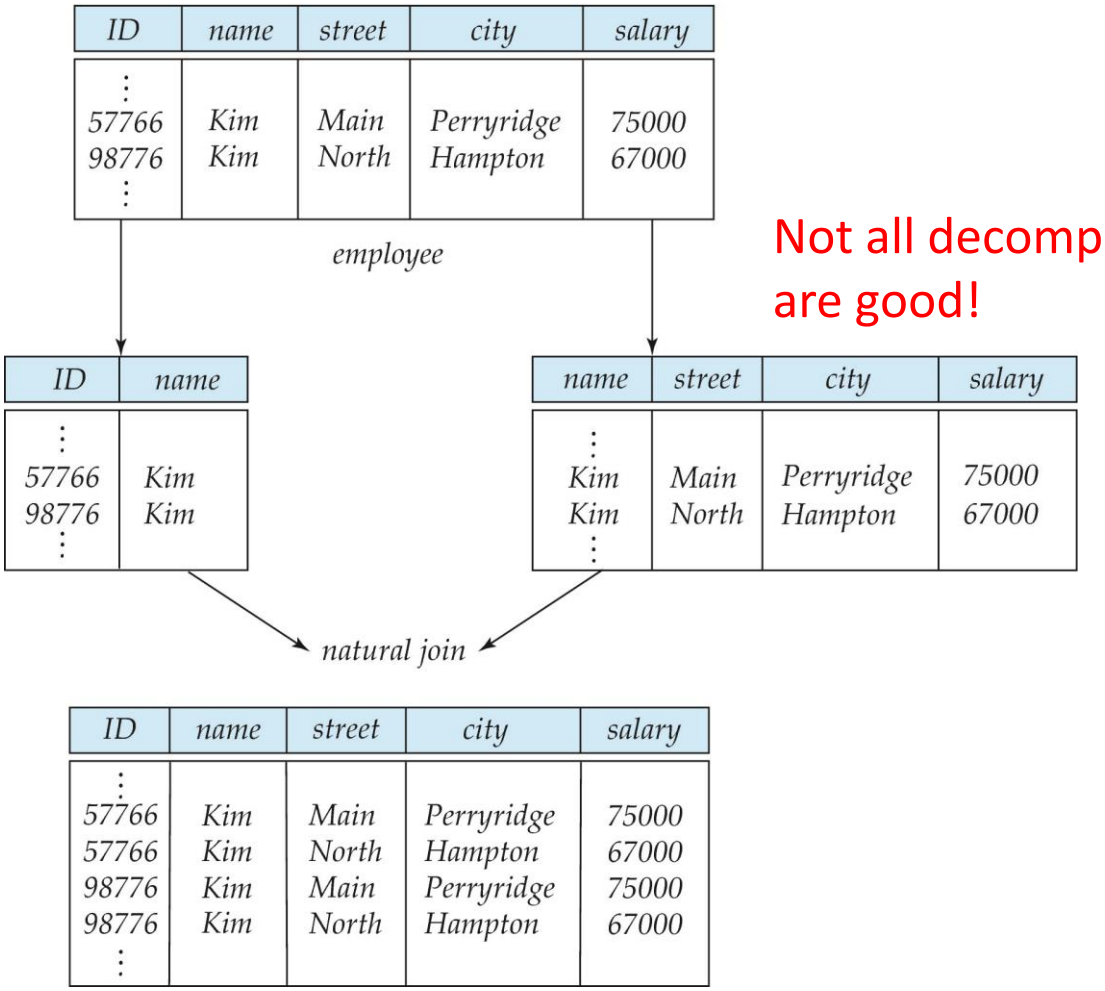
Rate type	Court	Member flag
SAVER	1	Yes
STANDARD	1	No
PREMIUM-A	2	Yes
PREMIUM-B	2	No

Today's bookings

Member flag	Court	Start time	End time
Yes	1	09:30	10:30
Yes	1	11:00	12:00
No	1	14:00	15:30
No	2	10:00	11:30
No	2	11:30	13:30
Yes	2	15:00	16:30

Example Source: Wikipedia

A Lossy Decomposition



Not all decompositions are good!

Example of Lossless-Join Decomposition

- **Lossless join and dependency preserving too!**

Assume,
F = {A→B, B→C}

$R = (\underline{A}, B, C)$

$\underline{A}$	B	C
α	1	A
β	2	B

r

$R_1 = (A, B)$

A	B
α	1
β	2

$\prod_{A,B}(r)$

$R_2 = (B, C)$

B	C
1	A
2	B

$\prod_{B,C}(r)$

$\prod_{A,B}(r) \bowtie \prod_{B,C}(r)$

A	B	C
α	1	A
β	2	B

Natural join (on B)

Joined on columns with common name.

Decomposition and Lossless Joins

- Consider a schema R(A, B, C, D) and functional dependencies A -> B and C -> D. Then the decomposition of R into R1(A, B) and R2(C, D) is
  (a) dependency preserving and lossless join
  (b) lossless join but not dependency preserving
  (c) dependency preserving but not lossless join
  (d) not dependency preserving and not lossless join

[GATE CS 2001]

Decomposition and Lossless Joins

- Consider a schema R(A, B, C, D) and functional dependencies A -> B and C -> D. Then the decomposition of R into R1(A, B) and R2(C, D) is
  - (a) dependency preserving and lossless join
  - (b) lossless join but not dependency preserving
  - (c) **dependency preserving but not lossless join**
  - (d) not dependency preserving and not lossless join

[GATE CS 2001]

Quiz

- Consider a relation *inst_info (ID, child_name, phone)*
  - where an instructor may have multiple phones and can have multiple children

ID	*child_name*	*phone*
99999	David	512-555-1234
99999	David	512-555-4321
99999	William	512-555-1234
99999	Willian	512-555-4321

inst_info

Is this in BCNF?

How good is BCNF? (Cont.)

- There are no non-trivial functional dependencies and therefore the relation is in BCNF
- Insertion anomalies exist– i.e., if we add a phone 981-992-3443 to 99999, we need to add two tuples

(99999, David, 981-992-3443)
(99999, William, 981-992-3443)

Redundancy in BCNF

- There are database schemas in BCNF that do not seem to be sufficiently normalized

ID	*child_name*	*phone*
99999	David	512-555-1234
99999	David	512-555-4321
99999	William	512-555-1234
99999	Willian	512-555-4321

inst_info

Quiz

- Let the set of functional dependencies F = {QR → S, R → P, S → Q} hold on a relation schema X = (PQRS). Suppose X is decomposed into two schemas Y and Z where Y = (PR) and Z = (QRS).
  - Is the decomposition dependency preserving?

Quiz

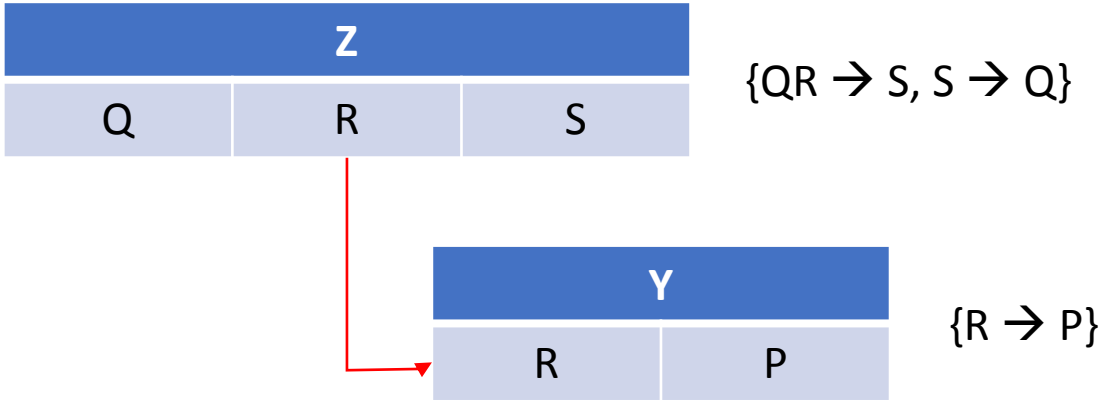
- Let the set of functional dependencies F = {QR → S, R → P, S → Q} hold on a relation schema X = (PQRS). Suppose X is decomposed into two schemas Y and Z where Y = (PR) and Z = (QRS).
  - Is the decomposition dependency preserving?
    - Y = (PR) : F = {R → P}
    - Z = (QRS) : F = {QR → S, S → Q}
    - We see that the decomposition is dependency preserving.

Quiz

- Let the set of functional dependencies F = {QR → S, R → P, S → Q} hold on a relation schema X = (PQRS). Suppose X is decomposed into two schemas Y and Z where Y = (PR) and Z = (QRS).
  - Is the decomposition lossless on join?

Quiz

- Let the set of functional dependencies F = {QR → S, R → P, S → Q} hold on a relation schema X = (PQRS). Suppose X is decomposed into two schemas Y and Z where Y = (PR) and Z = (QRS).
  - Is the decomposition lossless on join?



R is a key! So, the decomposition is lossless.

Quiz

- Let the set of functional dependencies F = {QR → S, R → P, S → Q} hold on a relation schema X = (PQRS). Suppose X is decomposed into two schemas Y and Z where Y = (PR) and Z = (QRS).
  - Is Y in BCNF?
  - Is Z in BCNF?

Quiz

- Let the set of functional dependencies F = {QR → S, R → P, S → Q} hold on a relation schema X = (PQRS). Suppose X is decomposed into two schemas Y and Z where Y = (PR) and Z = (QRS).
  - Is Y in BCNF?
  - Is Z in BCNF?

{R → P}

R is a super key.
So, Y is in BCNF.

Quiz

- Let the set of functional dependencies F = {QR → S, R → P, S → Q} hold on a relation schema X = (PQRS). Suppose X is decomposed into two schemas Y and Z where Y = (PR) and Z = (QRS).
  - Is Y in BCNF?
  - Is Z in BCNF?

Z		
Q	R	S

{QR → S, S → Q}

S is not a super key.
So, Z is not in BCNF.

Quiz

- Let the set of functional dependencies F = {QR → S, R → P, S → Q} hold on a relation schema X = (PQRS). Suppose X is decomposed into two schemas Y and Z where Y = (PR) and Z = (QRS).
  - What are the candidate keys of Z?

Z		
Q	R	S

{QR → S, S → Q}

Quiz

- Let the set of functional dependencies F = {QR → S, R → P, S → Q} hold on a relation schema X = (PQRS). Suppose X is decomposed into two schemas Y and Z where Y = (PR) and Z = (QRS).
  - What are the candidate keys of Z?

Z		
Q	R	S

{QR → S, S → Q}

Candidate Keys = RQ, RS

Quiz

- Let the set of functional dependencies F = {QR → S, R → P, S → Q} hold on a relation schema X = (PQRS). Suppose X is decomposed into two schemas Y and Z where Y = (PR) and Z = (QRS).
  - Is Z in 3NF?

Z		
Q	R	S

{QR → S, S → Q}

Candidate Keys = RQ, RS

Quiz

- Let the set of functional dependencies F = {QR → S, R → P, S → Q} hold on a relation schema X = (PQRS). Suppose X is decomposed into two schemas Y and Z where Y = (PR) and Z = (QRS).
  - Is Z in 3NF?

Z		
Q	R	S

{QR → S, S → Q}

Candidate Keys = RQ, RS

Yes! Z is in 3NF. No non-prime attributes.

Quiz

- Let the set of functional dependencies F = {QR → S, R → P, S → Q} hold on a relation schema X = (PQRS). Suppose X is decomposed into two schemas Y and Z where Y = (PR) and Z = (QRS).
  I. Both Y and Z are in BCNF
  II. Decomposition of X into Y and Z is dependency preserving and lossless
- Which of the above statements is/are correct?
  a) I only
  b) Neither I nor II
  c) Both I and II
  d) II only

Quiz

- Let the set of functional dependencies F = {QR → S, R → P, S → Q} hold on a relation schema X = (PQRS). Suppose X is decomposed into two schemas Y and Z where Y = (PR) and Z = (QRS).
  I. Both Y and Z are in BCNF
  II. Decomposition of X into Y and Z is dependency preserving and lossless
- Which of the above statements is/are correct?
  a) I only
  b) Neither I nor II
  c) Both I and II
  d) II only

[GATE CS 2019]

Fourth Normal Form (4NF)

- Motivation
  - The Pizza Delivery relation is in BCNF.
  - Yet there is redundancy!
- Restaurant does not functionally determine variety, but it multi-determines it.

{Restaurant} $\rightarrow$ {Pizza Variety}
{Restaurant} $\rightarrow$ {Delivery Area}

{Restaurant} $\rightarrow\!\!\rightarrow$ {Pizza Variety}
{Restaurant} $\rightarrow\!\!\rightarrow$ {Delivery Area}

Pizza Delivery Permutations

Restaurant	Pizza Variety	Delivery Area
A1 Pizza	Thick Crust	Springfield
A1 Pizza	Thick Crust	Shelbyville
A1 Pizza	Thick Crust	Capital City
A1 Pizza	Stuffed Crust	Springfield
A1 Pizza	Stuffed Crust	Shelbyville
A1 Pizza	Stuffed Crust	Capital City
Elite Pizza	Thin Crust	Capital City
Elite Pizza	Stuffed Crust	Capital City
Vincenzo's Pizza	Thick Crust	Springfield
Vincenzo's Pizza	Thick Crust	Shelbyville
Vincenzo's Pizza	Thin Crust	Springfield
Vincenzo's Pizza	Thin Crust	Shelbyville

4NF Decomposition

Pizza Delivery Permutations

Restaurant	Pizza Variety	Delivery Area
A1 Pizza	Thick Crust	Springfield
A1 Pizza	Thick Crust	Shelbyville
A1 Pizza	Thick Crust	Capital City
A1 Pizza	Stuffed Crust	Springfield

Varieties By Restaurant

Restaurant	Pizza Variety
A1 Pizza	Thick Crust
A1 Pizza	Stuffed Crust
Elite Pizza	Thin Crust
Elite Pizza	Stuffed Crust
Vincenzo's Pizza	Thick Crust
Vincenzo's Pizza	Thin Crust

Delivery Areas By Restaurant

Restaurant	Delivery Area
A1 Pizza	Springfield
A1 Pizza	Shelbyville
A1 Pizza	Capital City
Elite Pizza	Capital City
Vincenzo's Pizza	Springfield
Vincenzo's Pizza	Shelbyville

Review

Customer ID	First Name	Surname	Telephone Number
123	Pooja	Singh	555-861-2025, 192-122-1111
456	San	Zhang	(555) 403-1659 Ext. 53; 182-929-2929
789	John	Doe	555-808-9633

Not in 1NF

Manufacturer	Model	Model full name	Manufacturer country
Forte	X-Prime	Forte X-Prime	Italy
Forte	Ultraclean	Forte Ultraclean	Italy
Dent-o-Fresh	EZbrush	Dent-o-Fresh EZbrush	USA

Not in 2NF

Tournament	Year	Winner	Winner's date of birth
Indiana Invitational	1998	Al Fredrickson	21 July 1975
Cleveland Open	1999	Bob Albertson	28 September 1968
Des Moines Masters	1999	Al Fredrickson	21 July 1975

Not in 3NF

Court	Start time	End time	Rate type
1	09:30	10:30	SAVER
1	11:00	12:00	SAVER
1	14:00	15:30	STANDARD

Not in BCNF

Restaurant	Pizza Variety	Delivery Area
A1 Pizza	Thick Crust	Springfield
A1 Pizza	Thick Crust	Shelbyville
A1 Pizza	Thick Crust	Capital City

Not in 4NF

Quiz

- Consider a schema

  R(VOLUME, NUMBER, STARTPGE, ENDPAGE, TITLE, YEAR, PRICE)

- Functional dependencies are:

  (VOLUME, NUMBER, STARTPAGE, ENDPAGE) -> TITLE

  (VOLUME, NUMBER) -> YEAR

  (VOLUME, NUMBER, STARTPAGE, ENDPAGE) -> PRICE

- The database is redesigned to use the following schemas.

  R1(VOLUME, NUMBER, STARTPAGE, ENDPAGE, TITLE, PRICE)

  R2(VOLUME, NUMBER, YEAR)

- Which is the weakest normal form that the new database satisfies, but the old one does not?

[GATE CS 2016]

Quiz

- Consider a schema

  R(VOLUME, NUMBER, STARTPGE, ENDPAGE, TITLE, YEAR, PRICE)

- Functional dependencies are:

  (VOLUME, NUMBER, STARTPAGE, ENDPAGE) -> TITLE

  (VOLUME, NUMBER) -> YEAR

  (VOLUME, NUMBER, STARTPAGE, ENDPAGE) -> PRICE

- The database is redesigned to use the following schemas.

  R1(VOLUME, NUMBER, STARTPAGE, ENDPAGE, TITLE, PRICE)

  R2(VOLUME, NUMBER, YEAR)

- Which is the weakest normal form that the new database satisfies, but the old one does not? 2NF

[GATE CS 2016]