Chapter 9

Functional Dependencies and Normalization (from E&N,Silberschatz and my editing)

	Test	Remedy (Normalization)
1NF	Relation should have no nonatomic attributes or nested relations.	Form new relations for each nonatomic attribute or nested relation.
2NF	For relations where primary key contains multiple attributes, no nonkey attribute should be functionally dependent on a part of the primary key.	Decompose and set up a new relation for each partial key with its dependent attribute(s). Make sure to keep a relation with the original primary key and any attributes that are fully functionally dependent on it.
3NF	Relation should not have a nonkey attribute functionally determined by another nonkey attribute (or by a set of nonkey attributes.) That is, there should be no transitive dependency of a nonkey attribute on the primary key.	Decompose and set up a relation that includes the nonkey attribute(s) that functionally determine(s) other nonkey attribute(s).

Ex: (from Pak Wir's modul)

TABEL TIDAK NORMAL

ID_DOKTER	NAMA	ALAMAT	KOTA	ID_RS	NAMA_RS	KATEGORI_RS	DESKRIPSI_KATEGORI
101	Syamsulhadi	Jl. Pramuka 10	Klaten	1001	RS. Moewardi	A	RS. Daerah
102	Tunjung Sulaksono	Jl. Slamet Riyadi	Solo	**1002**	**RSI. Kustati**	B	RS. Swasta
				1003	RSU. Karima Utama	B	RS. Swasta
103	Achmad Subiyanto	Jl. Merpati 2	Solo	1002	RSI. Kustati	B	RS. Swasta
104	Tonang Dwi Ardiyanto	Jl. Pemuda Tengah 20	Klaten	1001	RS. Moewardi	A	RS. Daerah
105	Noor Rachma	Jl. Pisang 50	Solo	**1002**	**RSI. Kusta**	B	RS. Swasta
				1003	RSU. Karima Utama	B	RS. Swasta

-
- Problem?

ID_DOKTER	NAMA	ALAMAT	KOTA	ID_RS	NAMA_RS	KATEGORI_RS	DESKRIPSI_KATEGORI
101	Syamsulhadi	Jl. Pramuka 10	Klaten	1001	RS. Moewardi	A	RS. Daerah
102	Tunjung Sulaksono	Jl. Slamet Riyadi	Solo	1002	RSI. Kustati	B	RS. Swasta
NULL	**NULL**	**NULL**	**NULL**	1003	RSU. Karima Utama	B	RS. Swasta
103	Achmad Subiyanto	Jl. Merpati 2	Solo	1002	RSI. Kustati	B	RS. Swasta
104	Tonang Dwi Ardiyanto	Jl. Pemuda Tengah 20	Klaten	1001	RS. Moewardi	A	RS. Daerah
105	Noor Rachma	Jl. Pisang 50	Solo	1002	RSI. Kustati	B	RS. Swasta
NULL	**NULL**	**NULL**	**NULL**	1003	RSU. Karima Utama	B	RS. Swasta

- Nested
- NULL
- Anomaly INSERT → Input dokter, input RS juga, all
- Anomaly DELETE → 101 dihapus akan kehilangan RS.Moewardi
- Anomaly UPDATE → Ganti nama RS, ganti untuk seluruh dokter dll

- 1st N → hilangkan nested

TABEL DOKTER

ID_DOKTER	NAMA	ALAMAT	KOTA
101	Syamsulhadi	Jl. Pramuka 10	Klaten
102	Tunjung Sulaksono	Jl. Slamet Riyadi	Solo
103	Achmad Subiyanto	Jl. Merpati 2	Solo
104	Tonang Dwi Ardiyanto	Jl. Pemuda Tengah 20	Klaten
105	Noor Rachma	Jl. Pisang 50	Solo

TABEL DOKTER_RUMAH SAKIT

ID_DOKTER	ID_RS	NAMA_RS	KATEGORI_RS	DESKRIPSI_KATEGORI
101	1001	RS. Moewardi	A	RS. Daerah
102	1002	RSI. Kustati	B	RS. Swasta
103	1003	RSU. Karima Utama	B	RS. Swasta
103	1002	RSI. Kustati	B	RS. Swasta
104	1001	RS. Moewardi	A	RS. Daerah
105	1002	RSI. Kustati	B	RS. Swasta
105	1003	RSU. Karima Utama	B	RS. Swasta

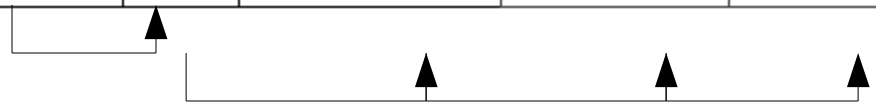
- FD

TABEL DOKTER

ID_DOKTER	NAMA	ALAMAT	KOTA
101	Syamsulhadi	Jl. Pramuka 10	Klaten
102	Tunjung Sulaksono	Jl. Slamet Riyadi	Solo
103	Achmad Subiyanto	Jl. Merpati 2	Solo
104	Tonang Dwi Ardiyanto	Jl. Pemuda Tengah 20	Klaten
105	Noor Rachma	Jl. Pisang 50	Solo

TABEL DOKTER_RUMAH SAKIT

ID_DOKTER	ID_RS	NAMA_RS	KATEGORI_RS	DESKRIPSI_KATEGORI
101	1001	RS. Moewardi	A	RS. Daerah
102	1002	RSI. Kustati	B	RS. Swasta
103	1003	RSU. Karima Utama	B	RS. Swasta
103	1002	RSI. Kustati	B	RS. Swasta
104	1001	RS. Moewardi	A	RS. Daerah
105	1002	RSI. Kustati	B	RS. Swasta
105	1003	RSU. Karima Utama	B	RS. Swasta



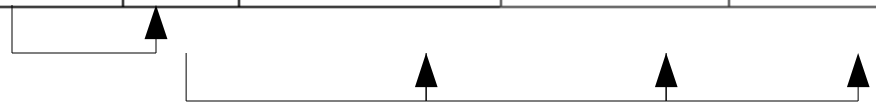
- FD

TABEL DOKTER

ID_DOKTER	NAMA	ALAMAT	KOTA
101	Syamsulhadi	Jl. Pramuka 10	Klaten
102	Tunjung Sulaksono	Jl. Slamet Riyadi	Solo
103	Achmad Subiyanto	Jl. Merpati 2	Solo
104	Tonang Dwi Ardiyanto	Jl. Pemuda Tengah 20	Klaten
105	Noor Rachma	Jl. Pisang 50	Solo

TABEL DOKTER_RUMAH SAKIT

ID_DOKTER	ID_RS	NAMA_RS	KATEGORI_RS	DESKRIPSI_KATEGORI
101	1001	RS. Moewardi	A	RS. Daerah
102	1002	RSI. Kustati	B	RS. Swasta
103	1003	RSU. Karima Utama	B	RS. Swasta
103	1002	RSI. Kustati	B	RS. Swasta
104	1001	RS. Moewardi	A	RS. Daerah
105	1002	RSI. Kustati	B	RS. Swasta
105	1003	RSU. Karima Utama	B	RS. Swasta



- 2nd N → hilangkan partial FD

TABEL DOKTER

ID_DOKTER	NAMA	ALAMAT	KOTA
101	Syamsulhadi	Jl. Pramuka 10	Klaten
102	Tunjung Sulaksono	Jl. Slamet Riyadi	Solo
103	Achmad Subiyanto	Jl. Merpati 2	Solo
104	Tonang Dwi Ardiyanto	Jl. Pemuda Tengah 20	Klaten
105	Noor Rachma	Jl. Pisang 50	Solo

TABEL PRAKTEK

ID_DOKTER	ID_RS
101	1001
102	1002
102	1003
103	1002
104	1001
105	1002
105	1001

TABEL RUMAH SAKIT

ID_RS	NAMA_RS	KATEGORI_RS	DESKRIPSI_KATEGORI
1001	RS. Moewardi	A	RS. Daerah
1002	RSI. Kustati	B	RS. Swasta
1003	RSU. Karima Utama	B	RS. Swasta

- FD

TABEL DOKTER

ID_DOKTER	NAMA	ALAMAT	KOTA
101	Syamsulhadi	Jl. Pramuka 10	Klaten
102	Tunjung Sulaksono	Jl. Slamet Riyadi	Solo
103	Achmad Subiyanto	Jl. Merpati 2	Solo
104	Tonang Dwi Ardiyanto	Jl. Pemuda Tengah 20	Klaten
105	Noor Rachma	Jl. Pisang 50	Solo

TABEL PRAKTEK

ID_DOKTER	ID_RS
101	1001
102	1002
102	1003
103	1002
104	1001
105	1002
105	1001

TABEL RUMAH SAKIT

ID_RS	NAMA_RS	KATEGORI_RS	DESKRIPSI_KATEGORI
1001	RS. Moewardi	A	RS. Daerah
1002	RSI. Kustati	B	RS. Swasta
1003	RSU. Karima Utama	B	RS. Swasta

- 3th N → hilangkan transitive FD X->Y, Y->Z, dimana Y adl non-prime attribute

TABEL DOKTER

ID_DOKTER	NAMA	ALAMAT	KOTA
101	Syamsulhadi	Jl. Pramuka 10	Klaten
102	Tunjung Sulaksono	Jl. Slamet Riyadi	Solo
103	Achmad Subiyanto	Jl. Merpati 2	Solo
104	Tonang Dwi Ardiyanto	Jl. Pemuda Tengah 20	Klaten
105	Noor Rachma	Jl. Pisang 50	Solo

TABEL PRAKTEK

ID_DOKTER	ID_RS
101	1001
102	1002
102	1003
103	1002
104	1001
105	1002
105	1001

TABEL RUMAH SAKIT

ID_RS	NAMA_RS	KATEGORI_RS
1001	RS. Moewardi	A
1002	RSI. Kustati	B
1003	RSU. Karima Utama	B

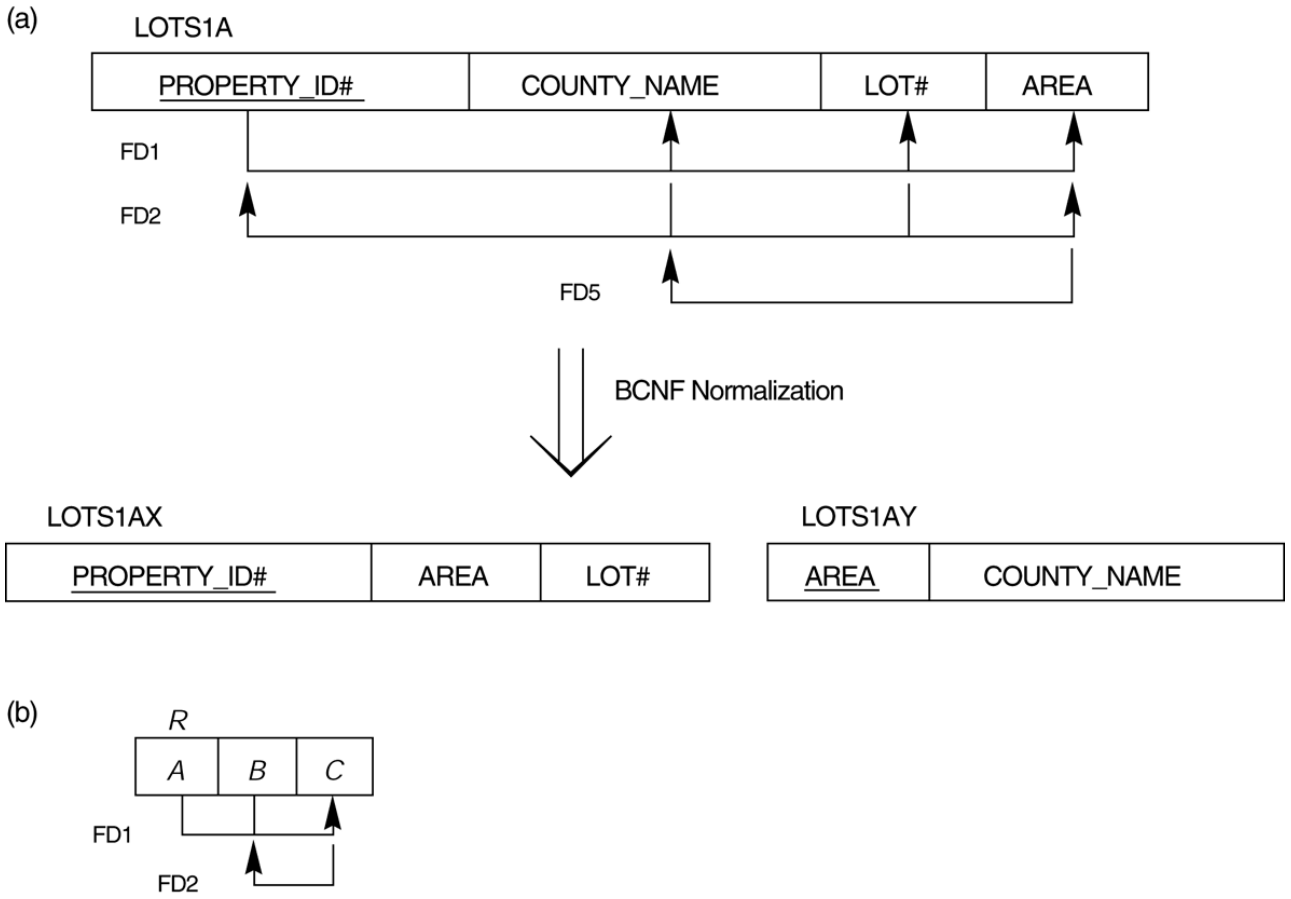
TABEL KATEGORI RS

KATEGORI_RS	DESKRIPSI_KATEGORI
A	RS. Daerah
B	RS. Swasta

BCNF

- 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
- Each normal form is strictly stronger than the previous one
  - Every 2NF relation is in 1NF
  - Every 3NF relation is in 2NF
  - Every BCNF relation is in 3NF
- There exist relations that are in 3NF but not in BCNF
- The goal is to have each relation in BCNF (or 3NF)

- (a) BCNF normalization of LOTS1A with the functional dependency FD2 being lost in the decomposition. (b) A schematic relation with FDs; it is in 3NF, but not in BCNF.



TEACH

STUDENT	COURSE	INSTRUCTOR
Narayan	Database	Mark
Smith	Database	Navathe
Smith	Operating Systems	Ammar
Smith	Theory	Schulman
Wallace	Database	Mark
Wallace	Operating Systems	Ahamad
Wong	Database	Omiecinski
Zelaya	Database	Navathe

BCNF,Decomp 1

- Two FDs exist in the relation TEACH:
  - fd1: { student, course} -> instructor
  - fd2: instructor ->course
- {student, course} is a candidate key for this relation and that the dependencies shown follow the pattern in. So this relation is in 3NF but not in BCNF

BCNF,Decomp 2

- Three possible decompositions for relation TEACH
  - {student, instructor} and {student, course}
  - {course, instructor } and {course, student}
  - {instructor, course } and {instructor, student}
- All three decompositions will lose fd1. We have to settle for sacrificing the functional dependency preservation. But we cannot sacrifice the non-additivity property after decomposition.
- Out of the above three, only the 3rd decomposition will not generate spurious tuples after join.(and hence has the non-additivity property).

Comparing the Normal Form

Poor Relational Schema Design
Developed as Stepping Stone

Eliminate the non-trivial functional dependencies of non-key attributes to key

1NF

↓

Eliminate partial FDs of non-key attributes to key

2NF

↓

Eliminate transitive FDs of non-key attributes to key

3NF

↓

Eliminate partial and transitive FDs of key attributes to key

BCNF

Recall Transitive FD

R = (U, F)

U = { S#, DName, DHead }

F = { S#→DName,

DName →DHead }

S#	DName	DHead
S1	D1	John
S2	D1	Jonh
S3	D2	Smith
S4	D3	Black

- S# → Dhead" is a Transitive FD
  - When S4 Graduates, Head Information of D3 Lost
  - Similarly, If D5 has No Students Yet, then the Head Information cannot be Stored in this Database
  - Update Head of Any Department Requires an Update to Every Student Enrolled in the Dept.

Possible Decompositions

R = (U, F) U = { S#, DName, DHead }

F = { S# → DName, DName → DHead }

S#
S1
S2
S3
S4

DName
D1
D1
D2
D3

DHead
John
John
Smith
Black

δ_1 **Information Based**

$\delta_1 = \{ R_1(S\#, \varnothing), R_2(DName, \varnothing), R_3(DHead, \varnothing)\}$

δ_1 **Neither Lossless nor FD-Preserving**

R = (U, F) U = { S#, DName, DHead }

F = { S#→DName, DName →DHead }

S#	DName
S1	D1
S2	D1
S3	D2
S4	D3

S#	DHead
S1	John
S2	John
S3	Smith
S4	Black

δ_2

- **Lossless Decomposition but not Dependency-Preserving**
- **DName→DHead is lost in the decomposition**

$\delta_2 = \{ R_1(\{S\#, DName\}, \{S\#\rightarrow DName\}),$

$R_2(\{S\#, DHead\}, \{S\#\rightarrow DHead\})\}$

δ_2 **Lossless but not FD-Preserving**

R = (U, F) U = { S#, DName, DHead }

F = { S#→DName, DName →DHead }

S#	DName
S1	D1
S2	D1
S3	D2
S4	D3

DName	DHead
D1	John
D1	John
D2	
D3	

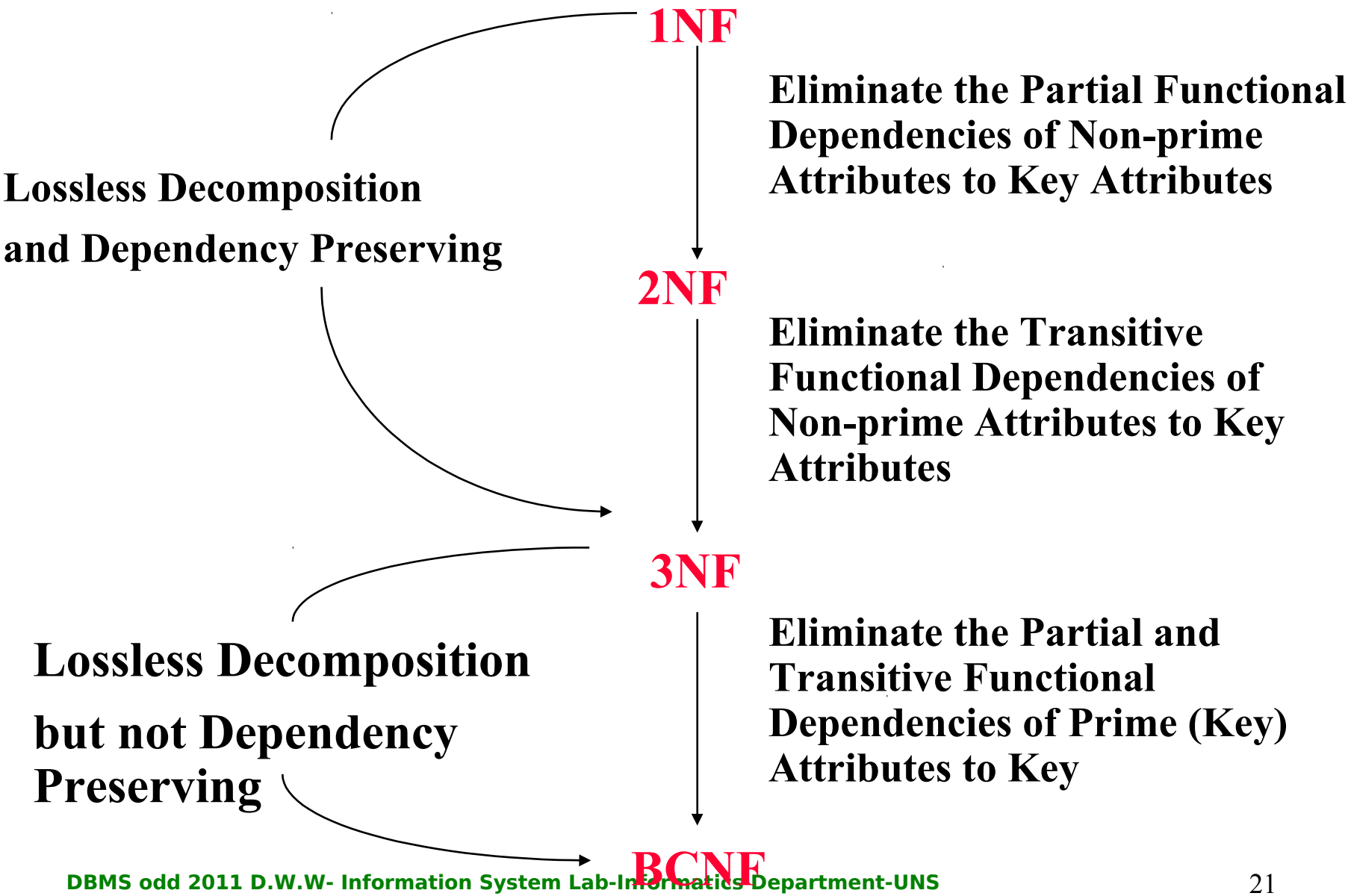
δ_3 **Lossless & dependency-preserving decomposition**

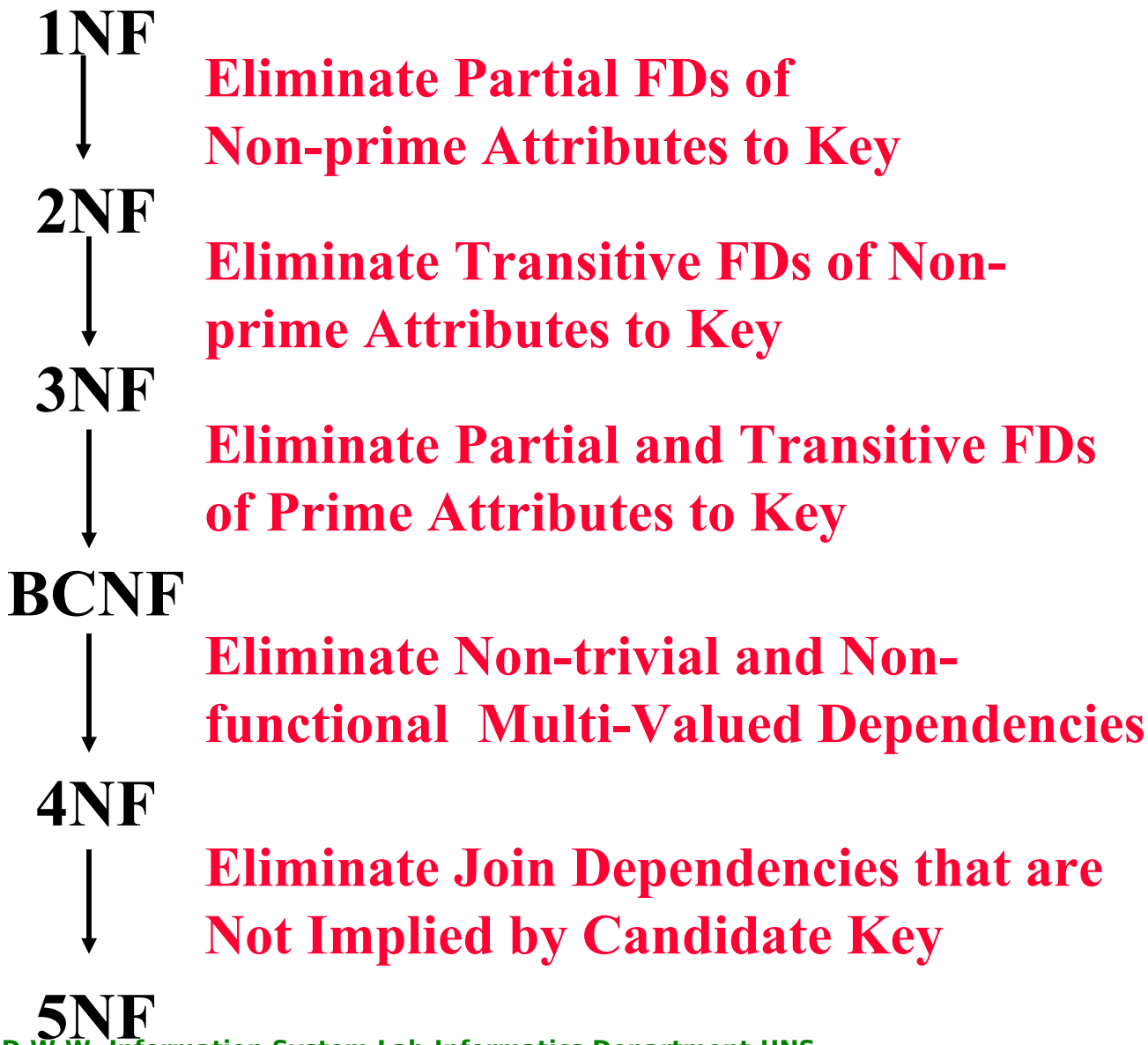
$\delta_3 = \{ R_1(\{S\#, DName\}, \{S\# \rightarrow DName\})$

$R_3(\{DName, DHead\}, \{Dname \rightarrow DHead\})\}$

δ_3 **is both Lossless and FD-Preserving**

Summary of Normlz




1NF
Eliminate Partial FDs of Non-prime Attributes to Key
2NF
Eliminate Transitive FDs of Non-prime Attributes to Key
3NF
Eliminate Partial and Transitive FDs of Prime Attributes to Key
BCNF
Eliminate Non-trivial and Non-functional Multi-Valued Dependencies
4NF
Eliminate Join Dependencies that are Not Implied by Candidate Key
5NF

Multivalued Dependences

- Focused on the Concept of Multi-Valued Dependencies
- A MVD X $\rightarrow\!\!\rightarrow$ Y Indicates that a Value of X Corresponds to Multiple Values of Y
- Consider EMP with MVDs:
  - ENAME $\rightarrow\!\!\rightarrow$ PNAME (E works on many Project)
  - ENAME $\rightarrow\!\!\rightarrow$ DNAME (E has many Dependents)

EMP

ENAME	PNAME	DNAME
Smith	X	John
Smith	Y	Anna
Smith	X	Anna
Smith	Y	John

4th Normal Form

- A Relation Schema R is in **Fourth Normal Form** (**4NF**) w.r.t Dependencies F (FD and MVD) if for every Non-Trivial MVD $X \rightarrow\rightarrow Y$ in F^+, X is a Superkey for R\
- MVD $X \rightarrow\rightarrow Y$ in R is called trivial if
  - Y is subset of X, or
  - X U Y = R
- Reconsider EMP with MVDs:
  - ENAME $\rightarrow\rightarrow$ PNAME (E works on many P)
  - ENAME $\rightarrow\rightarrow$ DNAME (E has many Dependents)
- ENAME is Not a Superkey of R since Need Triple of ENAME, PNAME, and DNAME to Distinguish
- We need to Decompose EMP!

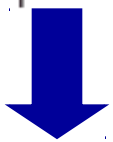
Note on FD

- *A* functional dependency is trivial if it is satisfied by all instances of a relation
- *E.g.*
  - *customer-name, loan-number* $\rightarrow$ *customer-name*
  - *customer-name* $\rightarrow$ *customer-name*
- In general, $\alpha \rightarrow \beta$ is trivial if $\beta \subseteq \alpha$

Decomp into 4NF

EMP

ENAME	PNAME	DNAME
Smith	X	John
Smith	Y	Anna
Smith	X	Anna
Smith	Y	John



EMP_PROJECTS

ENAME	PNAME
Smith	X
Smith	Y

EMP_DEPENDENTS

ENAME	DNAME
Smith	John
Smith	Anna

ENAME $\rightarrow\!\!\!\rightarrow$ PNAME is Trivial MVD: ENAME $\cup$ PNAME is Equal to EMP_PROJECTS (same for ENAME $\rightarrow\!\!\!\rightarrow$ DNAME)

Multivalued Dep and 4NF

Decomposing a relation state of EMP that is not in 4NF. (a) EMP relation with additional tuples. (b) Two corresponding 4NF relations EMP_PROJECTS and EMP_DEPENDENTS.

(a) **EMP**

ENAME	PNAME	DNAME
Smith	X	John
Smith	Y	Anna
Smith	X	Anna
Smith	Y	John
Brown	W	Jim
Brown	X	Jim
Brown	Y	Jim
Brown	Z	Jim
Brown	W	Joan
Brown	X	Joan
Brown	Y	Joan
Brown	Z	Joan
Brown	W	Bob
Brown	X	Bob
Brown	Y	Bob
Brown	Z	Bob

(b) **EMP_PROJECTS**

ENAME	PNAME
Smith	X
Smith	Y
Brown	W
Brown	X
Brown	Y
Brown	Z

EMP_DEPENDENTS

ENAME	DNAME
Smith	Anna
Smith	John
Brown	Jim
Brown	Joan
Brown	Bob