Database Management Systems

Vijaya Saradhi

IIT Guwahati

Tue, 06rd Feb 2020

Introduction

Example - 02

S(EMPLOYEE, CHILD, SALARY, YEAR)

EMPLOYEE	CHILD	SALARY	YEAR
E1	e1_c1	$35K	1976
E1	e1_c1	$40K	1976
E2	e2_c1	$40K	1975
E2	e2_c2	$40K	1975
E2	e2_c1	$50K	1976
E2	e2_c2	$50K	1976
E3	e3_c1	$15K	1975
E3	e3_c1	$20K	1976

MVD: Employee $\rightarrow\!\!\rightarrow$ Child

Consider Employee E_2

E_2	$e_2 c_1$	($40K, 1975)
E_2	$e_2 c_2$	($40K, 1975)
E_2	$e_2 c_1$	($50K, 1976)
E_2	$e_2 c_2$	($50K, 1976)

MVD: Employee $\rightarrow\!\!\rightarrow$ Child

Two rows $t, u \in R$

t	E_2	$e_2 c_1$	(\$40K, 1975)
	E_2	$e_2 c_2$	(\$40K, 1975)
	E_2	$e_2 c_1$	(\$50K, 1976)
u	E_2	$e_2 c_2$	(\$50K, 1976)

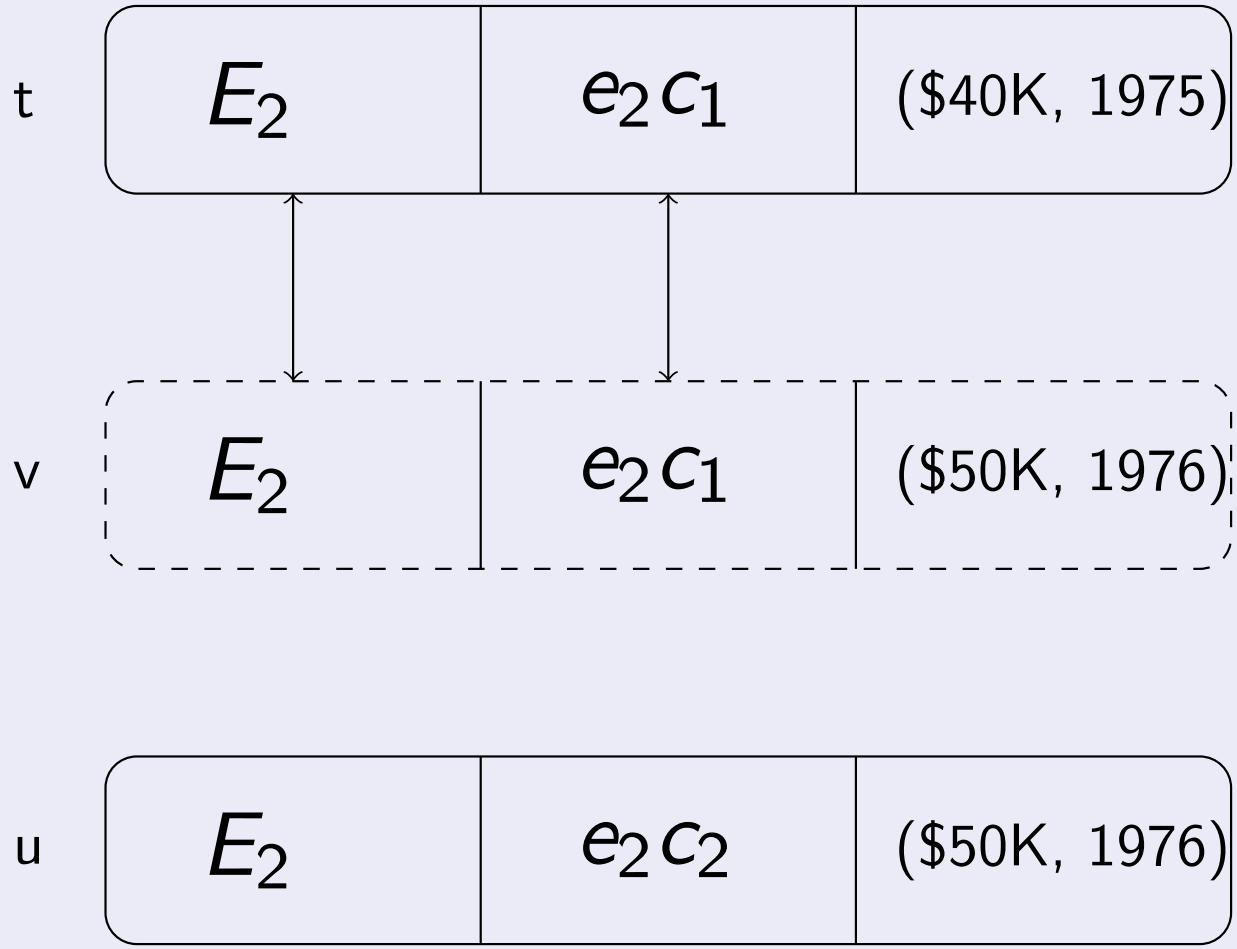
MVD: Employee $\rightarrow\!\!\rightarrow$ Child

There exists a $v \in R$

t	E_2	$e_2 c_1$	($40K, 1975)
v	E_2	$e_2 c_1$	($50K, 1976)
u	E_2	$e_2 c_2$	($50K, 1976)

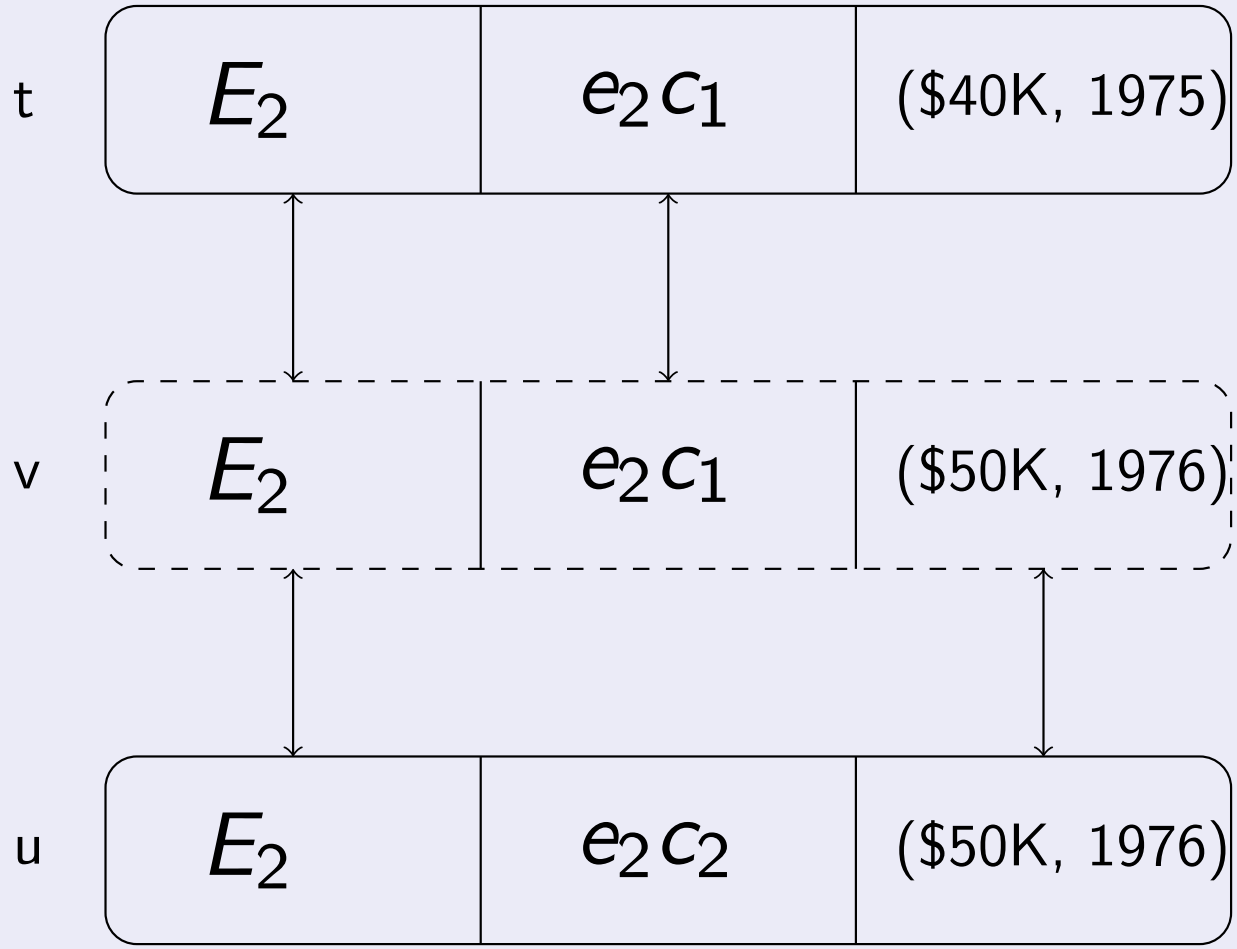
MVD: Employee $\rightarrow\!\!\rightarrow$ Child

$t[X] = v[X]$ $\&t[Y] = v[Y]$ $X = \{employee\}, Y = \{child\}, Z = \{salary, year\}$

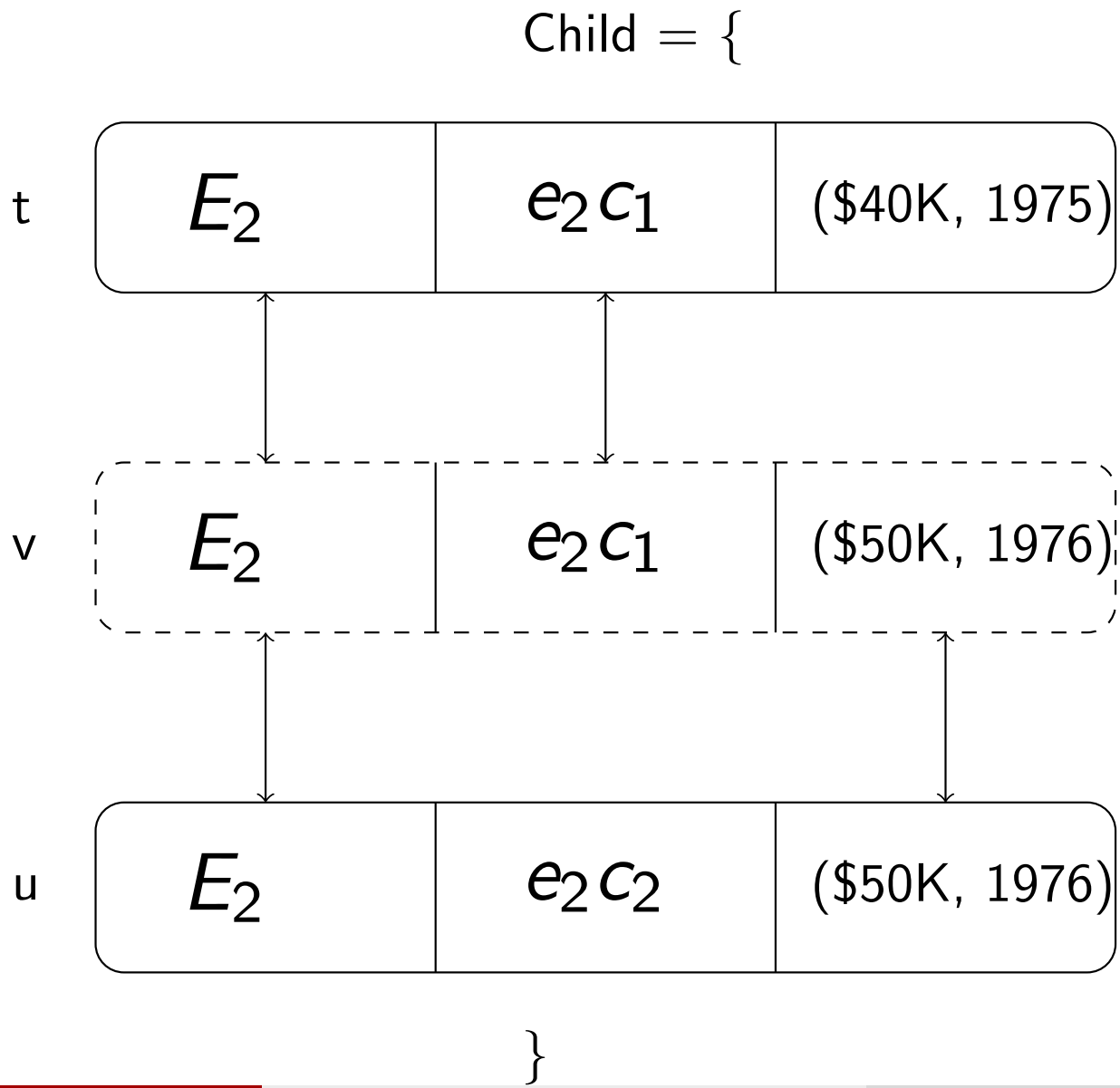


MVD: Employee $\rightarrow\!\!\rightarrow$ Child

$u[X] = v[X]$ &$u[Z] = v[Z]$ $X = \{employee\}, Y = \{child\}, Z = \{salary, year\}$



MVD: Employee $\rightarrow\!\!\rightarrow$ Child



MVD

Notation

- Let $R(X_1, \cdots, X_m, Y_1, \cdots, Y_n, Z_1, \cdots, Z_r)$ be a relation
- R has (m + n + r) column names
- Let $\mathbf{X} = \{X_1, \cdots, X_m\}$; so is **Y**, **Z**
- if $X_1 = x_1, \cdots, X_m = x_m$ then $\mathbf{x} = (x_1, \cdots, x_m)$; so is **y**, **z**

MVD

Define

$\mathbf{Y_{xz}} = \{\mathbf{y} : (\mathbf{x}, \mathbf{y}, \mathbf{z}) \in R\}$
is non-empty if and only if **x** and **z** appear together in a tuple of R

MVD

Definition

The MVD $\mathbf{X} \rightarrow\rightarrow \mathbf{Y}$ holds for R($\mathbf{X}$, $\mathbf{Y}$, $\mathbf{Z}$) if $\mathbf{Y_{x,z}}$ depends only on $\mathbf{x}$
That is $\mathbf{Y_{x,z}} = \mathbf{Y_{x,z'}}$ for each $\mathbf{x}$, $\mathbf{z}$, $\mathbf{z'}$ such that $\mathbf{Y_{x,z}}$ and $\mathbf{Y_{x,z'}}$ are nonempty

MVD

S(EMPLOYEE, CHILD, SALARY, YEAR)

EMPLOYEE	CHILD	SALARY	YEAR
E1	e1_c1	$35K	1976
E1	e1_c1	$40K	1976
E2	e2_c1	$40K	1975
E2	e2_c1	$50K	1976
E2	e2_c2	$40K	1975
E2	e2_c2	$50K	1976
E3	e3_c1	$15K	1975
E3	e3_c1	$20K	1976

- EMPLOYEE $\rightarrow\rightarrow$ CHILD holds for T
- $\text{CHILD}_{E2,\$40K,1975} = \{\text{e2\_c1, e2\_c2}\}$
- $\text{CHILD}_{E2,\$50K,1976} = \{\text{e2\_c1, e2\_c2}\}$
- Therefore $\text{CHILD}_{E2,\$40K,1975} = \text{CHILD}_{E2,\$50K,1976}$

MVD

S(EMPLOYEE, CHILD, SALARY, YEAR)

EMPLOYEE	CHILD	SALARY	YEAR
E1	e1_c1	$35K	1976
E1	e1_c1	$40K	1976
E2	e2_c1	$40K	1975
E2	e2_c1	$50K	1976
E2	e2_c2	$40K	1975
E2	e2_c2	$50K	1976
E3	e3_c1	$15K	1975
E3	e3_c1	$20K	1976

- Delete the row (E2, e2_c2, $50K, 1976) from T; Let this relation be T'
- Then EMPLOYEE $\rightarrow\rightarrow$ CHILD holds for T'
- $\text{CHILD}_{E2,\$40K,1975} = \{\text{e2\_c1, e2\_c2}\}$
- $\text{CHILD}_{E2,\$50K,1976} = \{\text{e2\_c1}\}$
- Therefore $\text{CHILD}_{E2,\$40K,1975} \neq \text{CHILD}_{E2,\$50K,1976}$

FDs in terms of MVD

Definition

- The FD $\mathbf{X} \rightarrow \mathbf{Y}$ holds for R($\mathbf{X}$, $\mathbf{Y}$, $\mathbf{Z}$) if $\mathbf{Y_{x,z}}$ depends only on $\mathbf{x}$ AND
- $\mathbf{Y_{xz}} = \{\mathbf{y} : (\mathbf{x}, \mathbf{y}, \mathbf{z}) \in R\}$ is non-empty if and only if $\mathbf{x}$ and $\mathbf{z}$ appear together in a tuple of R AND
- The set $\mathbf{Y_{x,z}}$ is nonempty and contains at most one member

FDs and MVDs

Proposition - 01

If **X** and **Y** are disjoint, and if the FD $\mathbf{X} \rightarrow \mathbf{Y}$ holds for a relation R, then the MVD $\mathbf{X} \rightarrow\rightarrow \mathbf{Y}$ also holds for R

FDs and MVDs

Discussion

Proposition 01 implies that if the FD $\mathbf{X} \rightarrow \mathbf{Y}$ holds for a relation every instance of R, then the MVD $\mathbf{X} \rightarrow\rightarrow \mathbf{Y}$ necessarily holds

Trivial MVD

Trivial

- The MVD $\mathbf{X} \twoheadrightarrow \phi$
- The MVD $\mathbf{X} \twoheadrightarrow \mathbf{Y}$ necessarily hold for R(**X, Y**)
- The MVD $\{A, B\} \twoheadrightarrow C$ holds for every relation R(A, B, C) with exactly three columns A, B, C
- These are trivial MVDs

4^{th} Normal Form

Definition

A relation schema R^* is in 4NF if, whenever a nontrivial MVD $\mathbf{X} \rightarrow\rightarrow \mathbf{Y}$ holds, then so does the *functional dependency* $\mathbf{X} \rightarrow$ A for every column A of R^*.

4^{th} Normal Form

Definition

Intuitively all dependencies are the result of the keys

4NF and BCNF

Theorem

If a relation schema R^* is in 4NF, then it is in BCNF

4NF and BCNF

Proof - 01

- Assume R^* is in 4NF and not in BCNF
- As R^* not in BCNF, there is a nontrivial FD $\mathbf{X} \rightarrow \mathbf{Y}$
- This FD holds for R^* and there is a column A such that the FD $\mathbf{X} \not\rightarrow$ A hold

4NF and BCNF

Proof - 02

- Let $\mathbf{Y}_1 = \mathbf{Y} - \mathbf{X}$
- $\mathbf{X} \rightarrow \mathbf{Y}_1$ holds as $\mathbf{X} \rightarrow \mathbf{Y}$ holds
- $\mathbf{X}$ and $\mathbf{Y}_1$ are disjoint
- The MVD $\mathbf{X} \rightarrow\rightarrow \mathbf{Y}_1$ holds from proposition 01

4NF and BCNF

Proof - 03

- The MVD is nontrivial
  - $\mathbf{Y}_1 \neq \phi$
  - $\mathbf{X}$ and $\mathbf{Y}_1$ do not partition the columns of R^*
  - As $A \notin \mathbf{X}$ or $A \notin \mathbf{Y}_1$

4NF and BCNF

Proof - 05

- By definition of 4NF, since the MVD $\mathbf{X} \rightarrow\rightarrow \mathbf{Y}_1$ holds for R^*
- so does the FD $\mathbf{X} \rightarrow A$
- This is a contradiction
- Therefore the theorem holds

4NF Examples

When two (or more) many-to-many relations are placed in single table

Example - 01

EMPLOYEE	SKILL	LANGUAGE
Smith	Cook	
Smith	Type	
Smith		French
Smith		German
Smith		Greek

4NF Examples

Storage Format - Disjoint

EMPLOYEE	SKILL	LANGUAGE
Smith	Cook	
Smith	Type	
Smith		French
Smith		German
Smith		Greek

4NF Examples

Storage Format - Random Mix with minimal number of records

EMPLOYEE	SKILL	LANGUAGE
Smith	Cook	French
Smith	Type	German
Smith	Type	Greek

4NF Examples

Storage Format - Random Mix with NULL values

EMPLOYEE	SKILL	LANGUAGE
Smith	Cook	French
Smith	Type	German
Smith	$\perp$	Greek

4NF Examples

Storage Format - Random Mix - Unrestricted

EMPLOYEE	SKILL	LANGUAGE
Smith	Cook	French
Smith	Type	$\perp$
Smith	$\perp$	German
Smith	Type	Greek

4NF Examples

Storage Format - Cross product

EMPLOYEE	SKILL	LANGUAGE
Smith	Cook	French
Smith	Cook	German
Smith	Cook	Greek
Smith	Type	French
Smith	Type	German
Smith	Type	Greek

4NF Examples

When two (or more) many-to-many relations are placed in single table

Example - 01

EMPLOYEE	SKILL	LANGUAGE
Smith	Cook	
Smith	Type	
Smith		French
Smith		German
Smith		Greek

- MVD: EMPLOYEE $\rightarrow\!\!\rightarrow$ SKILL
- MVD: EMPLOYEE $\rightarrow\!\!\rightarrow$ LANGUAGE

4NF Examples

Example - 01

Decompose the original relation into two

- EMP_SKILL(EMPLOYEE, SKILL)
- EMP_LANGUAGE(EMPLOYEE, LANGUAGE)

4NF Examples

Example - 02

R^*(class, section, student, major, exam, year, instructor, rank, salary, text, day, room)

4NF Examples

FDs

FD #	**X**	$\rightarrow$	**Y**
F1	{class, section}	$\rightarrow$	instructor
F2	{class, section, day}	$\rightarrow$	room
F3	student	$\rightarrow$	{major, year}
F4	instructor	$\rightarrow$	{rank, salary}

MVDs

MVD #	**X**	$\rightarrow\rightarrow$	**Y**
M1	{class, section}	$\rightarrow\rightarrow$	{student, major, exam, year}
M2	{class, section}	$\rightarrow\rightarrow$	{instructor, rank, salary}
M3	{class, section}	$\rightarrow\rightarrow$	text
M4	{class, section}	$\rightarrow\rightarrow$	{day, room}
M5	{class, section, student}	$\rightarrow\rightarrow$	exam
M6	class	$\rightarrow\rightarrow$	text

4NF Examples - 02

R^*(class, section, student, major, exam, year, instructor, rank, salary, text, day, room)
Use

MVD #	**X**	$\rightarrow\rightarrow$	**Y**
M1	{class, section}	$\rightarrow\rightarrow$	{student, major, exam, year}

to decompose R^*

R_1^*(class, section, student, major, exam, year)
R_2^*(class, section, instructor, rank, salary, text, day, room)

4NF Examples - 02

R_1^*(class, section, student, major, exam, year)
R_2^*(class, section, instructor, rank, salary, text, day, room)
Use

MVD #	**X**	$\rightarrow\rightarrow$	**Y**
M5	{class, section, student}	$\rightarrow\rightarrow$	exam

to decompose R_1^*

R_{11}^*(class, section, student, exam)
R_{12}^*(class, section, student, major, year)
R_2^*(class, section, instructor, rank, salary, text, day, room)

4NF Examples - 02

R_{11}^*(class, section, student, exam)
R_{12}^*(class, section, student, major, year)
R_2^*(class, section, instructor, rank, salary, text, day, room)
Use

FD #	**X**	$\rightarrow$	**Y**
F3	student	$\rightarrow$	{major, year}

to decompose R_{12}^*

R_{11}^*(class, section, student, exam)
R_{121}^*(student, major, year)
R_{122}^*(class, section, student)
R_2^*(class, section, instructor, rank, salary, text, day, room)

4NF Examples - 02

R_{11}^*(class, section, student, exam)
R_{121}^*(student, major, year)
R_{122}^*(class, section, student)
R_2^*(class, section, instructor, rank, salary, text, day, room)
Use

MVD #	**X**	$\rightarrow\rightarrow$	**Y**
M2	{class, section}	$\rightarrow\rightarrow$	{instructor, rank, salary}

to decompose R_2^*

R_{11}^*(class, section, student, exam)
R_{121}^*(student, major, year)
R_{122}^*(class, section, student)
R_{21}^*(class, section, instructor, rank, salary)
R_{22}^*(class, section, text, day, room)

4NF Examples - 02

R^*_{11}(class, section, student, exam)
R^*_{121}(student, major, year)
R^*_{122}(class, section, student)
R^*_{21}(class, section, instructor, rank, salary)
R^*_{22}(class, section, text, day, room)
Use

FD #	**X**	$\rightarrow$	**Y**
F4	instructor	$\rightarrow$	{rank, salary}

to decompose R^*_{21}

R^*_{11}(class, section, student, exam)
R^*_{121}(student, major, year)
R^*_{122}(class, section, student)
R^*_{211}(instructor, rank, salary)
R^*_{212}(class, section, instructor)
R^*_{22}(class, section, text, day, room)

4NF Examples - 02

R^*_{11}(class, section, student, exam)
R^*_{121}(student, major, year)
R^*_{122}(class, section, student)
R^*_{211}(instructor, rank, salary)
R^*_{212}(class, section, instructor)
R^*_{22}(class, section, text, day, room)
Use

MVD #	**X**	$\rightarrow\rightarrow$	**Y**
M6	{class, section, student}	$\rightarrow\rightarrow$	exam

to decompose R^*_{22}

R^*_{11}(class, section, student, exam)
R^*_{121}(student, major, year)
R^*_{122}(class, section, student)
R^*_{211}(instructor, rank, salary)
R^*_{212}(class, section, instructor)
R^*_{221}(class, text)
R^*_{222}(class, section, day, room)

4NF Examples - 02

R^*_{11}(class, section, student, exam)
R^*_{121}(student, major, year)
R^*_{122}(class, section, student)
R^*_{211}(instructor, rank, salary)
R^*_{212}(class, section, instructor)
R^*_{22}(class, section, text, day, room)
Remove R^*_{122} as it is part of R^*_{11}

R^*_{11}(class, section, student, exam)
R^*_{121}(student, major, year)
R^*_{211}(instructor, rank, salary)
R^*_{212}(class, section, instructor)
R^*_{221}(class, text)
R^*_{222}(class, section, day, room)