

Лекція 15

Нормалізація

S

S#	SNAME	STATUS	CITY
S1	Smith	20	London
S2	Jones	10	Paris
S3	Blake	30	Paris
S4	Clark	20	London
S5	Adams	30	Athens

J

J#	JNAME	CITY
J1	Sorter	Paris
J2	Display	Rome
J3	OCR	Athens
J4	Console	Athens
J5	RAID	London
J6	EDS	Oslo
J7	Tape	London

SPJ

S#	P#	J#	QTY
S1	P1	J1	200
S1	P1	J4	700
S2	P3	J1	400
S2	P3	J2	200
S2	P3	J3	200
S2	P3	J4	500
S2	P3	J5	600
S2	P3	J6	400
S2	P3	J7	800
S2	P5	J2	100
S4	P6	J3	300
S4	P6	J1	300
S5	P2	J2	200
S5	P2	J4	100
S5	P5	J5	500
S5	P5	J7	100
S5	P6	J2	200
S5	P1	J4	100
S5	P3	J4	200
S5	P4	J4	800
S5	P5	J4	400
S5	P6	J4	500

P

P#	PNAME	COLOR	WEIGHT	CITY
P1	Nut	Red	12.0	London
P2	Bolt	Green	17.0	Paris
P3	Screw	Blue	17.0	Oslo
P4	Screw	Red	14.0	London
P5	Cain	Blue	12.0	Paris
P6	Cog	Red	19.0	London

SCP

S#	CITY	P#	QTY
S1	London	P1	100
S1	London	P2	100
S2	Paris	P1	200
S2	Paris	P2	200
S3	Paris	P2	300
S4	London	P2	400
S5	London	P4	500
S6	London	P5	600

Нормалізація

S

S#	SNAME	STATUS	CITY
S1	Smith	20	London
S2	Jones	10	Paris
S3	Blake	30	Paris
S4	Clark	20	London
S5	Adams	30	Athens

SC

S#	SNAME	STATUS
S1	Smith	20
S2	Jones	10
S3	Blake	30
S4	Clark	20
S5	Adams	30

SP

S#	P#	QTY
S1	P1	100
S1	P2	100
S2	P1	200
S2	P2	200
S3	P2	300
S4	P2	400
S5	P4	500
S6	P5	600

SCP

S#	CITY	P#	QTY
S1	London	P1	100
S1	London	P2	100
S2	Paris	P1	200
S2	Paris	P2	200
S3	Paris	P2	300
S4	London	P2	400
S5	London	P4	500
S6	London	P5	600

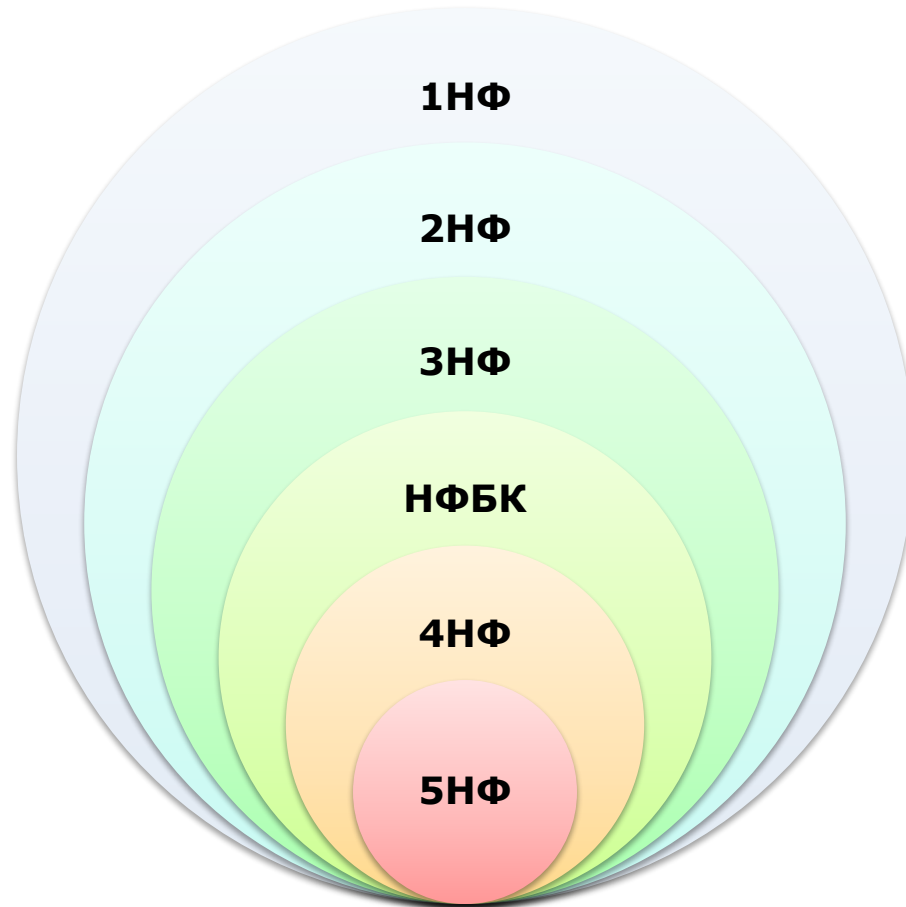
Нормалізація

- Нормалізація – формальний метод аналізу відношень на основі їх потенційних ключів та існуючих функціональних залежностей.
- Нормалізація - послідовне зведення заданого набору змінних відношень до деякої більш бажаної форми.

Нормальні форми

- Перша нормальна форма
- Друга нормальна форма
- Третя нормальна форма
- Нормальна форма Бойса-Кодда
- Четверта нормальна форма
- П'ята нормальна форма

Нормальні форми



Перша нормальна форма

S

SNAME	PHONES
Smith	555-0121
	555-2804
Ned Flanders	555-8904

S

SNAME	PHONES
Smith	555-0121
Smith	555-2804
Ned Flanders	555-8904

Перша нормальна форма

SC

S#	SNAME	COLORS
S1	Smith	Red, Blue, Green, Brown
S2	Jones	Fuchsia, Yellow
S3	Blake	Black, Lime, Blue
S4	Clark	Gray, Brown
S5	Adams	Olive, Silver, Red, Gold

SC

S#	SNAME	COLORS
S1	Smith	Red
S1	Smith	Blue
S1	Smith	Green
S1	Smith	Brown
S2	Jones	Fuchsia
S2	Jones	Yellow
S3	Blake	Black
S3	Blake	Lime
S3	Blake	Blue
S4	Clark	Gray
S4	Clark	Brown
S5	Adams	Olive
S5	Adams	Silver
S5	Adams	Red
S5	Adams	Gold

Перша нормальна форма

- Змінна відношення знаходиться у 1НФ тоді і тільки тоді, коли в будь-якому допустимому значенні цієї змінної відношення кожний її кортеж містить тільки одне значення для кожного з атрибутів.

Декомпозиція без втрат

S

S#	STATUS	CITY
S3	30	Paris
S5	30	Athens

SST

a)

S#	STATUS
S3	30
S4	30

SC

S#	CITY
S3	Paris
S5	Athens

SST

b)

S#	STATUS
S3	30
S4	30

STC

STATUS	CITY
30	Paris
30	Athens

Декомпозиція без втрат

S

S#	STATUS	CITY
S3	30	Paris
S5	30	Athens

$S\# \rightarrow \text{STATUS}$

$S\# \rightarrow \text{CITY}$

SST

a)

S#	STATUS
S3	30
S4	30

$S\# \rightarrow \text{STATUS}$

$S\# \rightarrow \text{CITY}$

SC

S#	CITY
S3	Paris
S5	Athens

SST

b)

S#	STATUS
S3	30
S4	30

$S\# \rightarrow \text{STATUS}$

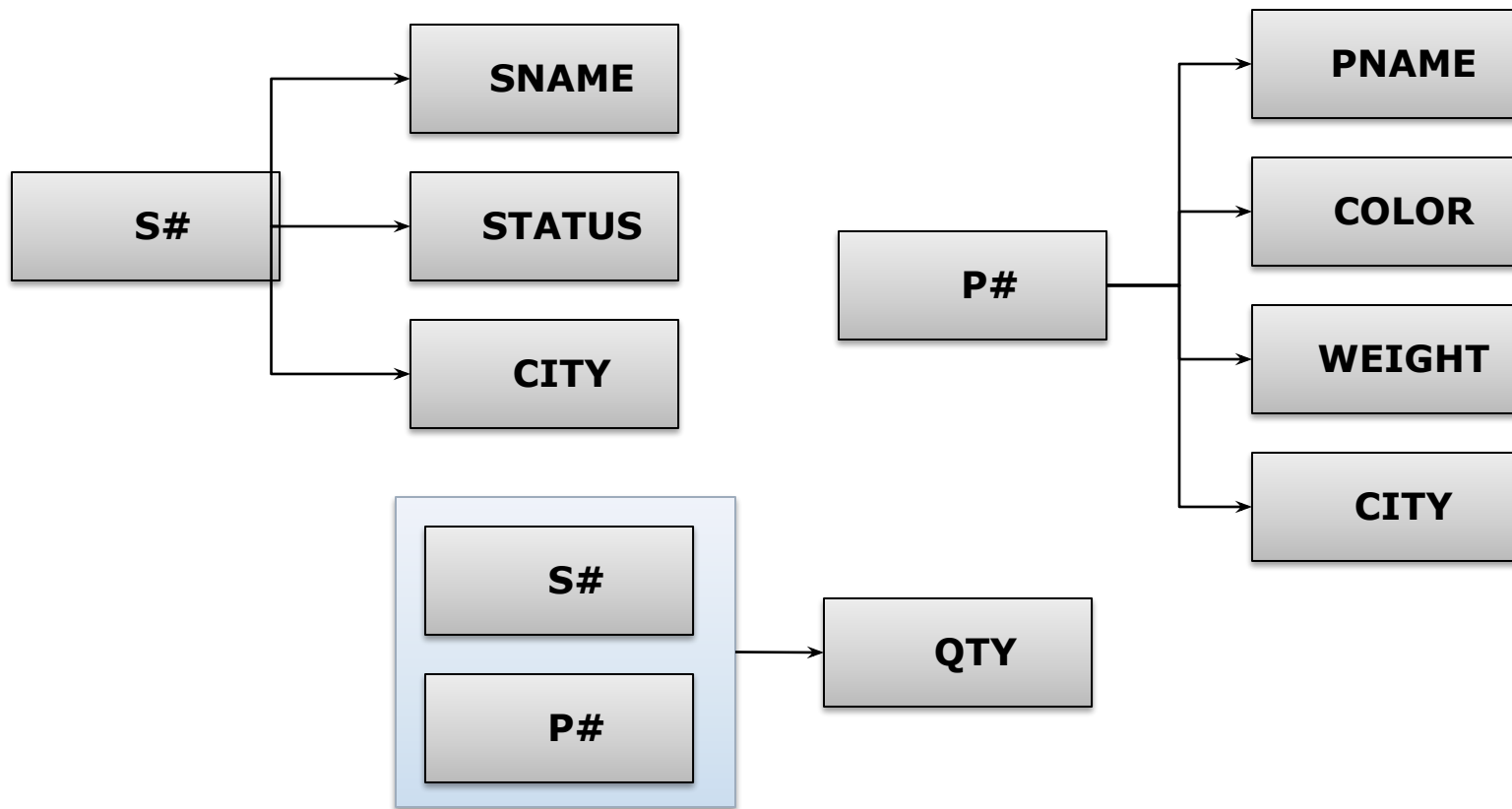
STC

STATUS	CITY
30	Paris
30	Athens

Теорема Хіта

- Нехай $R\{A, B, C\}$ є змінною відношенням, де A, B і C — множини атрибутів цієї змінної відношення. Якщо R задовольняє функціональній залежності $A \rightarrow B$, то R дорівнює з'єднанню її проєкцій по атрибутам $\{A, B\}$ і $\{A, C\}$.
- Декомпозиція змінної відношення R на проєкції R_1, R_2, \dots, R_n виконується без втрат, якщо R дорівнює з'єднанню R_1, R_2, \dots, R_n .

Діаграми функціональних залежностей

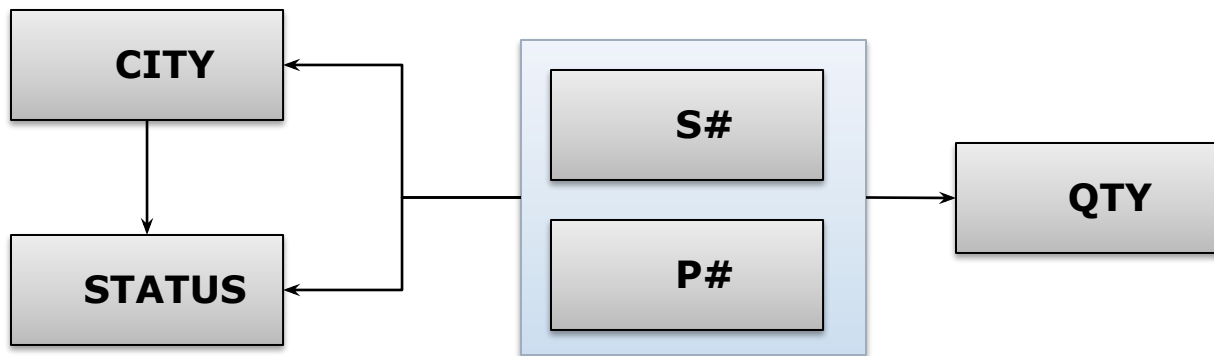


Перша нормальна форма

FIRST

S#	CITY	STATUS	P#	QTY
S1	London	20	P1	300
S1	London	20	P2	200
S1	London	20	P3	400
S1	London	20	P4	200
S1	London	20	P5	100
S1	London	20	P6	100
S2	Paris	10	P1	300
S2	Paris	10	P2	400
S3	Paris	10	P2	200
S4	London	20	P2	200
S4	London	20	P4	300
S4	London	20	P5	400

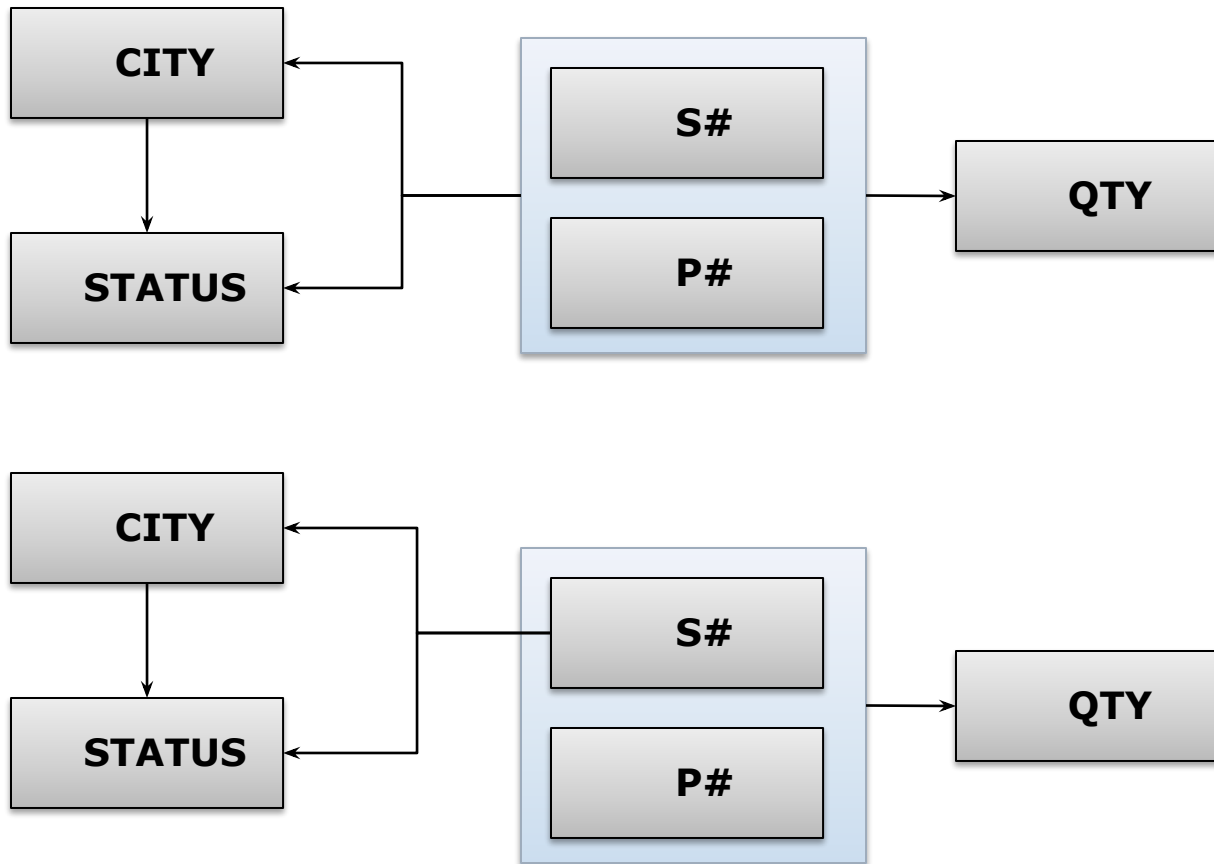
Перша нормальна форма



Аномалії оновлення

- Операція **INSERT**. Не можна помістити в змінну відношення FIRST інформацію про те, що деякий постачальник знаходиться в певному місті, до тих пір поки цей постачальник не здійснить поставку хоча б одної деталі.
- Операція **DELETE**. Якщо із змінної відношення FIRST видалити кортеж, який є єдиним для деякого постачальника, буде видалена не тільки інформація про поставку постачальником деякої деталі, але також інформація про те, що цей постачальник знаходиться в певному місті.
- Операція **UPDATE**. У загальному, назва міста для кожного постачальника повторюється в змінній відношенні FIRST декілька разів, і ця збитковість призводить до виникнення проблем при оновленні.

Перша нормальна форма



Друга нормална форма

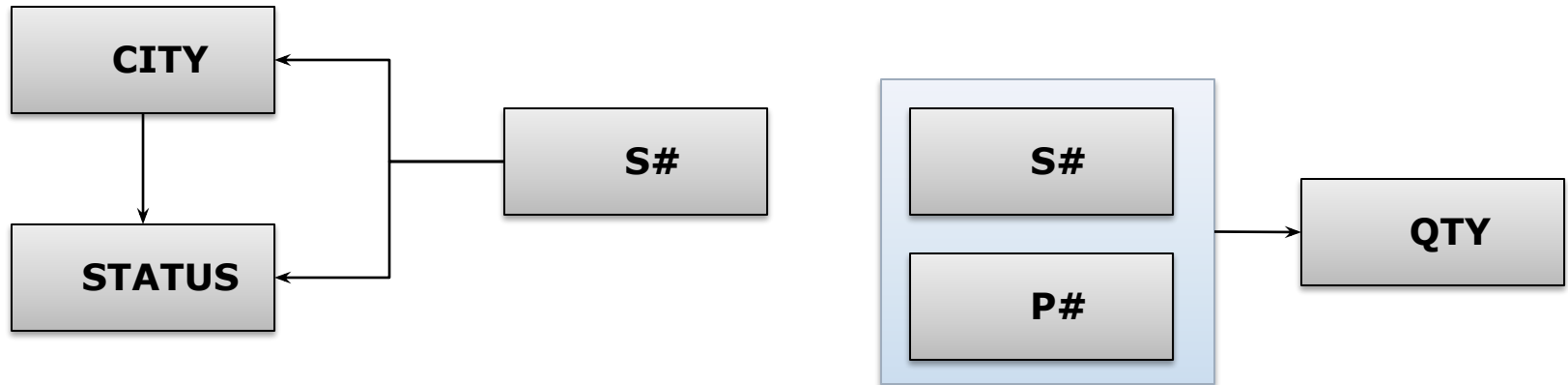
SECOND

S#	CITY	STATUS
S1	London	20
S2	Paris	10
S3	Paris	10
S4	London	20
S4	London	20
S5	Athens	30

SP

S#	P#	QTY
S1	P1	300
S1	P2	200
S1	P3	400
S1	P4	200
S1	P5	100
S1	P6	100
S2	P1	300
S2	P2	400
S3	P2	200
S4	P2	200
S4	P4	300
S4	P5	400

Друга нормална форма



Друга нормальна форма

- Операція **INSERT**. Інформацію про те, що постачальник з номером S5 знаходиться в Афінах, можна помістити в базу даних, додавши відповідний кортеж в змінну відношення SECOND, причому навіть в тому випадку, якщо він на даний момент не постачає жодної деталі.
- Операція **DELETE**. Тепер цілком можливо видалити інформацію про поставку деталі з номером P2 постачальником з номером S3.
- Операція **UPDATE**. У вдосконаленій структурі назва міста для кожного постачальника вказується всього один раз, оскільки існує тільки один кортеж для даного постачальника в змінній відношенні SECOND.

Друга нормальна форма

- Змінна відношення знаходиться в другій нормальній формі тоді й тільки тоді, коли вона перебуває в першій нормальній формі й кожний неключовий атрибут незвідно залежить від її первинного ключа.

Друга нормальна форма

- Перший етап процедури нормалізації полягає у створенні проєкцій, які дозволяють виключити функціональні залежності, що не є незвідні.

- Нехай дана змінна відносини R , що має наступний вид.
 $R \{A, B, C, D\}$ PRIMARY KEY $\{A, B\}$ /* $A \rightarrow D$ */

- Процедура нормалізації передбачає заміну цієї змінної відношення наступними двома проєкціями, $R1$ і $R2$.

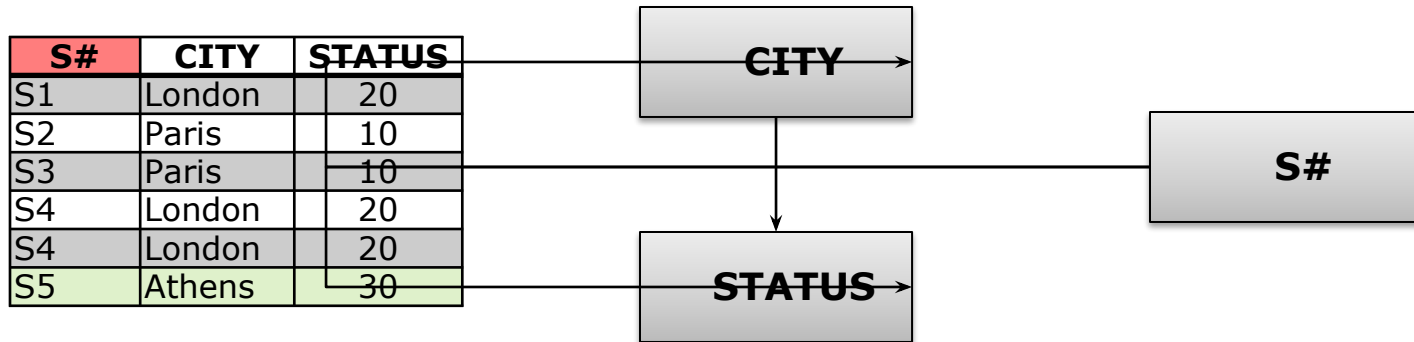
$R1 \{A, D\}$ PRIMARY KEY $\{A\}$

$R2 \{A, B, C\}$ PRIMARY KEY $\{A, B\}$ FOREIGN KEY $\{A\}$ REFERENCES $R1$

- Змінна відношення R завжди може бути відновлена за допомогою з'єднання змінних відношень $R1$ і $R2$ по зовнішньому ключу й відповідному йому первинному ключу цих змінних відношень.

Третя нормална форма

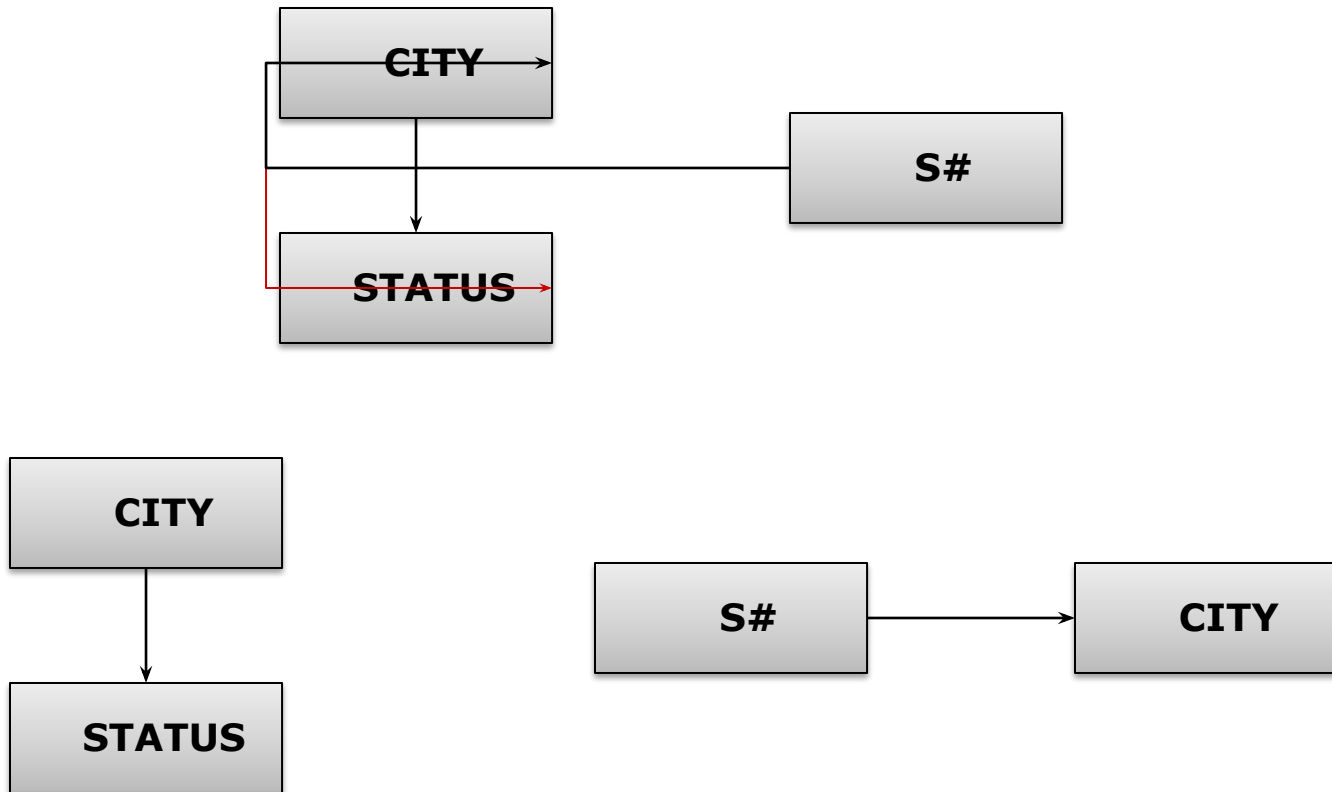
SECOND



Аномалії оновлення

- Операція **INSERT**. Не можна помістити в базу даних відомості про те, що певне місто характеризується деяким статусом, доти, поки в цьому місті не з'явиться певний постачальник.
- Операція **DELETE**. При видаленні зі змінної відношення SECOND кортежу для деякого міста, представленого в ній цим єдиним кортежем, будуть вилучені не тільки відомості про постачальника з даного міста, але й інформація про те, який статус відповідає самому місту.
- Операція **UPDATE**. У змінній відношенні SECOND значення статусу для кожного міста повторюється кілька раз (тому вона усе ще має деяку надмірність).

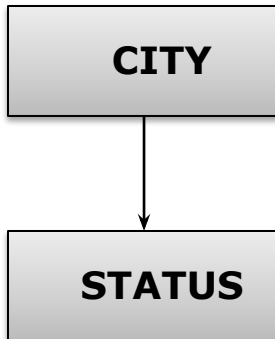
Третя нормална форма



Трета нормална форма

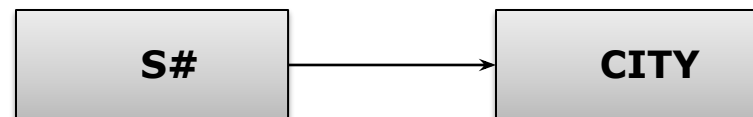
SC

S#	CITY
S1	London
S2	Paris
S3	Paris
S4	London
S5	Athens



CS

CITY	STATUS
London	20
Paris	10
Athens	30
Rome	50



Третя нормальна форма

- Змінна відношення знаходиться в третій нормальній формі тоді й тільки тоді, коли вона перебуває в другій нормальній формі й жоден неключовий атрибут не є транзитивно залежним від її первинного ключа.

Третя нормальна форма

- Другий етап нормалізації полягає в створенні проєкцій для усунення транзитивних залежностей.
- Нехай дана змінна відношення R , що має наступний вид.
 $R \{ A, B, C \}$ PRIMARY KEY $\{ A \}$ /* $B \rightarrow C$ */
- Процедура нормалізації передбачає заміну змінної відносини R наступними двома проєкціями, $R1$ і $R2$.

$R1 \{ B, C \}$ PRIMARY KEY $\{ B \}$

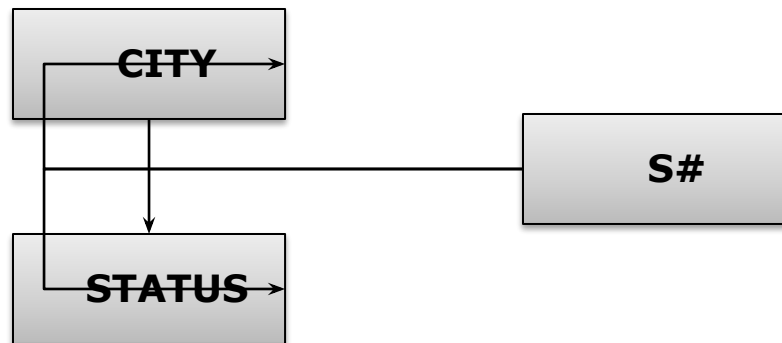
$R2 \{ A, B \}$ PRIMARY KEY $\{ A \}$ FOREIGN KEY $\{ B \}$ REFERENCES $R1$

- Змінна відношення R може бути відновлена за допомогою з'єднання змінних відношень $R1$ і $R2$ по зовнішньому ключу й відповідному йому первинному ключу цих змінних відношень.

Третя нормальна форма

- Рівень нормалізації змінної відношення визначається семантикою, а не конкретним значенням цієї змінної в певний момент часу.
 - Інакше кажучи, за конкретним значенням деякої змінної відносини неможливо визначити, чи перебуває вона, наприклад, у ЗНФ.
- Для цього необхідно також знати, які функціональні залежності визначені в розглянутій змінній відношенні.
 - Навіть знаючи про залежності в деякій змінній відношенні, не можна на підставі конкретного її значення довести, що вона перебуває в ЗНФ.

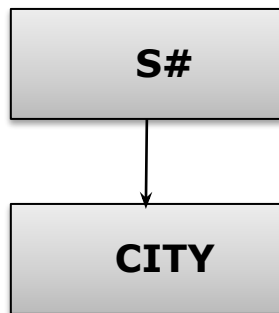
Збереження залежностей



Збереження залежностей

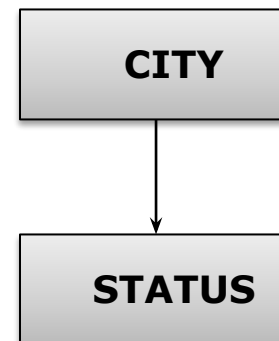
SC

S#	CITY
S1	London
S2	Paris
S3	Paris
S4	London
S5	Athens



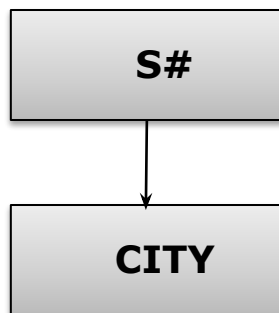
CS

CITY	STATUS
London	20
Paris	10
Athens	30
Rome	50



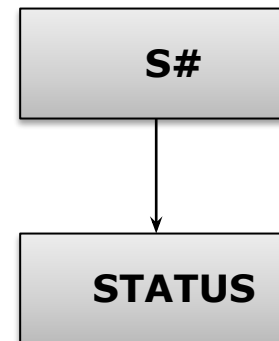
SC

S#	CITY
S1	London
S2	Paris
S3	Paris
S4	London
S5	Athens



CS

S#	STATUS
S1	20
S2	10
S3	10
S4	20
S5	30



Збереження залежностей

- Концепція незалежних проєкцій надає критерій вибору одного з можливих варіантів декомпозиції.

- Проєкції R_1 і R_2 змінної відношення R будуть незалежні тоді й тільки тоді, коли дотримуються наступні вимоги:
 - кожна функціональна залежність у змінній відношенні R є логічним наслідком функціональних залежностей у її проєкціях R_1 і R_2 ;
 - загальні атрибути проєкцій R_1 і R_2 утворюють потенційний ключ принаймні для однієї із цих двох проєкцій.

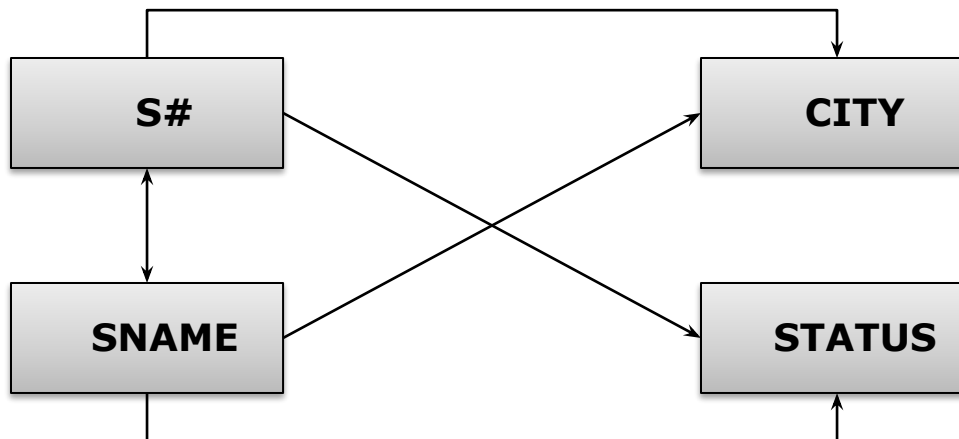
Нормальна форма Бойса-Кодда

- Змінна відношення знаходиться в нормальній форм Бойса-Кодда тоді й тільки тоді, коли кожна її нетривіальна й незвідна зліва функціональна залежність, має в якості детермінанта потенційний ключ.

Нормальна форма Бойса-Кодда

S

S#	SNAME	STATUS	CITY
S1	Smith	20	London
S2	Jones	10	Paris
S3	Blake	30	Paris
S4	Clark	20	London
S5	Adams	30	Athens



Нормальна форма Бойса-Кодда

SSP

S#	SNAME	P#	QTY
S1	Smith	P1	300
S1	Smith	P2	200
S1	Smith	P3	400
S1	Smith	P4	200
S1	Smith	P5	100
S1	Smith	P6	100
S2	Jones	P1	300
S2	Jones	P2	400
S3	Blake	P2	200
S4	Clark	P2	200
S4	Clark	P4	300
S4	Clark	P5	400

{S#, P#} і {SNAME, P#} – потенційні ключі SSP

Нормальна форма Бойса-Кодда

SS

S#	SNAME
S1	Smith
S2	Jones
S3	Blake
S4	Clark

SP

S#	P#	QTY
S1	P1	300
S1	P2	200
S1	P3	400
S1	P4	200
S1	P5	100
S1	P6	100
S2	P1	300
S2	P2	400
S3	P2	200
S4	P2	200
S4	P4	300
S4	P5	400

Нормальна форма Бойса-Кодда

SS

S#	SNAME
S1	Smith
S2	Jones
S3	Blake
S4	Clark

SP

SNAME	P#	QTY
Smith	P1	300
Smith	P2	200
Smith	P3	400
Smith	P4	200
Smith	P5	100
Smith	P6	100
Jones	P1	300
Jones	P2	400
Blake	P2	200
Clark	P2	200
Clark	P4	300
Clark	P5	400

Нормальна форма Бойса-Кодда

S – студент

T – викладач

J – предмет

Кожний студент вивчає певний предмет тільки в одного викладача.

Кожний викладач веде тільки один предмет, але кожний предмет ведуть декілька викладачів.

Нормальна форма Бойса-Кодда

STJ

S	J	T
Smith	Math	Prof. White
Smith	Physics	Prof. Green
Jones	Math	Prof. White
Jones	Physics	Prof. Brown

S – студент

T – викладач

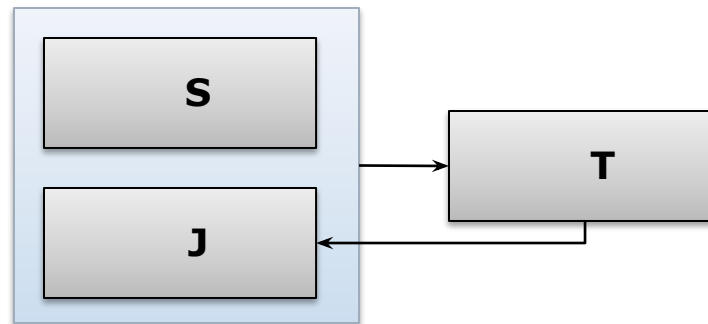
J – предмет

$\{ S, J \} \rightarrow \{ T \}$

$\{ T \} \rightarrow \{ J \}$

Кожний студент вивчає певний предмет тільки в одного викладача.

Кожний викладач веде тільки один предмет, але кожний предмет ведуть декілька викладачів.



Нормальна форма Бойса-Кодда

ST

S	T
Smith	Prof. White
Smith	Prof. Green
Jones	Prof. White
Jones	Prof. Brown

TJ

T	J
Prof. White	Math
Prof. Green	Physics
Prof. Brown	Physics

$\{ T \} \rightarrow \{ J \}$

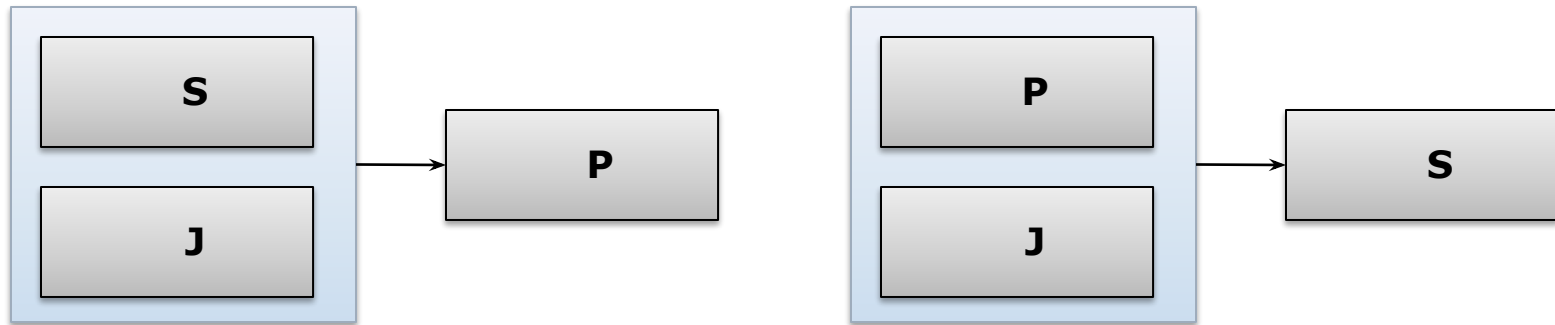
$\{ S, J \} \rightarrow \{ T \} ?$

Нормальна форма Бойса-Кодда

- Дано змінну відношення EXAM з атрибутами S (студент), J (предмет) і P (позиція).
- Кожний кортеж (s, j, p) змінної відношення EXAM відображає відомості про те, що деякий студент s екзаменується по певному предмету j і займає певну позицію p в екзаменаційній відомості.
- Крім того, ніякі два студенти не можуть займати ту саму позицію в екзаменаційній відомості, яка відноситься до того самого предмету.

Нормальна форма Бойса-Кодда

EXAM {S, J, P}



Нормальна форма Бойса-Кодда

- Змінна відношення знаходиться в нормальній форм Бойса-Кодда тоді й тільки тоді, коли кожна її нетривіальна й незвідна зліва функціональна залежність, має в якості детермінанта потенційний ключ.