

Лекция 10

**Нормальные формы.
Декомпозиция без потерь.**

МГТУ им. Н.Э.Баумана. ФН12

При проектировании базы данных решаются две основные проблемы.

1. Проблема логического проектирования баз данных.
2. Проблема физического проектирования баз данных.

Декомпозиция без потерь и ФЗ

Подход к проектированию реляционных БД на основе нормализации.

Процедура нормализации основывается на декомпозиции исходной переменной-отношения на другие переменные-отношения.

Процесс декомпозиции заключается в замене данной переменной-отношения некоторым набором ее проекций. Декомпозиция должна быть обратимой. (**декомпозиция без потерь**).

Определение. Пусть даны отношения $R_1(A,C)$ и $R_2(B,C)$, имеющие одинаковые атрибуты (т.е. атрибуты с одинаковыми именами и определенные на одинаковых доменах).

Естественным соединением отношений $R_1(A,C)$ и $R_2(B,C)$ называется отношение R с заголовком (A,B,C) и телом, содержащим множество кортежей (a_i, b_i, c_i) , таких, что $(a_i, c_i) \in R_1$ и $(b_i, c_i) \in R_2$.

Пример 7.1 Дана переменная-отношение S {S#, STATUS, CITY}

S поставщики		
S#	STATUS	CITY
S ₁	20	London
S ₂	30	Paris
S ₄	20	London
S ₅	30	Athens

Таблица 7.1 Отношение S

Рассмотрим два возможных варианта декомпозиции отношения S.

SSt		StC	
S#	STATUS	STATUS	CITY
S ₁	20	20	London
S ₂	30	30	Paris
S ₄	20	20	London
S ₅	30	30	Athens

SSt		SC	
S#	STATUS	S#	CITY
S ₁	20	S ₁	London
S ₂	30	S ₂	Paris
S ₄	20	S ₄	London
S ₅	30	S ₅	Athens

a)

b)

Таблица 7.2. Декомпозиции отношения S.

Корректные и некорректные декомпозиции отношений.

Теорема Хита.

Дано отношение $R \{A, B, C\}$, где A, B и C — множество атрибутов отношения R . Если R удовлетворяет ФЗ $A \rightarrow B$, тогда R равно соединению ее проекций $PROJECT[A, B]$ и $PROJECT[A, C]$.

$$\{ (R\{A, B, C\} \ \& \ A \rightarrow B) \Rightarrow R = (R \ PROJECT[A, B]) \ JOIN \ (R \ PROJECT[A, C]) \}.$$

Доказательство.

Пусть $R_1 = (R \ PROJECT[A, B]) \ JOIN \ (R \ PROJECT[A, C])$; Докажем, что $R_1 = R$.

По определению $(R_1 = R \Leftrightarrow R_1 \subseteq R \ \& \ R \subseteq R_1; \ R \subseteq R_1 \Leftrightarrow \forall r \in R \Rightarrow r \in R_1)$

1. $R \subseteq R_1$;

$$\forall (a, b, c) \in R \Rightarrow (a, b) \in (R \ PROJECT[A, B]) \ \& \ (a, c) \in (R \ PROJECT[A, C]) \Rightarrow (a, b, c) \in R_1$$

2. $R_1 \subseteq R$; Докажем от противного:

$$\begin{aligned} \exists (a, b, c') \in R_1 \ \& \ (a, b, c') \notin R; \ (a, b, c') \in R_1 \Rightarrow \exists (a, c') \in (R \ PROJECT[A, C]) \Rightarrow \exists b^* \\ | (a, b^*, c') \in R \Rightarrow \exists (a, b^*) \in (R \ PROJECT[A, B]); \ ((a, b^*) \in (R \ PROJECT[A, B]) \ \& \\ (a, b) \in (R \ PROJECT[A, B]) \ \& \ A \rightarrow B) \Rightarrow (a = a \Rightarrow b^* = b) \Rightarrow (a, b^*, c') = (a, b, c') \end{aligned}$$

$$\begin{aligned} (a, b', c) \in R_1 \ \& \ (a, b', c) \notin R; \ \exists (a, c) \in (R \ PROJECT[A, C]) \Rightarrow \exists b, (a, b, c) \in R \Rightarrow \\ (a, b) \in (R \ PROJECT[A, B]) \Rightarrow ((a, b) \in (R \ PROJECT[A, B]) \ \& \\ (a, b') \in (R \ PROJECT[A, B]) \ \& \ A \rightarrow B) \Rightarrow (b' = b) \Rightarrow (a, b', c) = (a, b, c). \# \end{aligned}$$

Assignment_Department		
<u>clientID</u>	<u>employeeID</u>	DepartmentID
1	2211	12
2	2213	10
3	2211	12

(a) Исходное отношение

ED		CE	
<u>employeeID</u>	DepartmentID	<u>clientID</u>	employeeID
2211	12	1	2211
2213	10	2	2213
		3	2211

(б) Декомпозиция без потерь по теореме Хита
Таблица 7.3.

Атрибут employeeID не является возможным ключом, но наличия ФЗ employeeID→DepartmentID достаточно для декомпозиции этого отношения без потерь.

Определение. Атрибут В неприводимо (минимально) зависит от атрибута А, если выполняется неприводимая слева ФЗ $A \rightarrow B$.

Диаграммы функциональных зависимостей

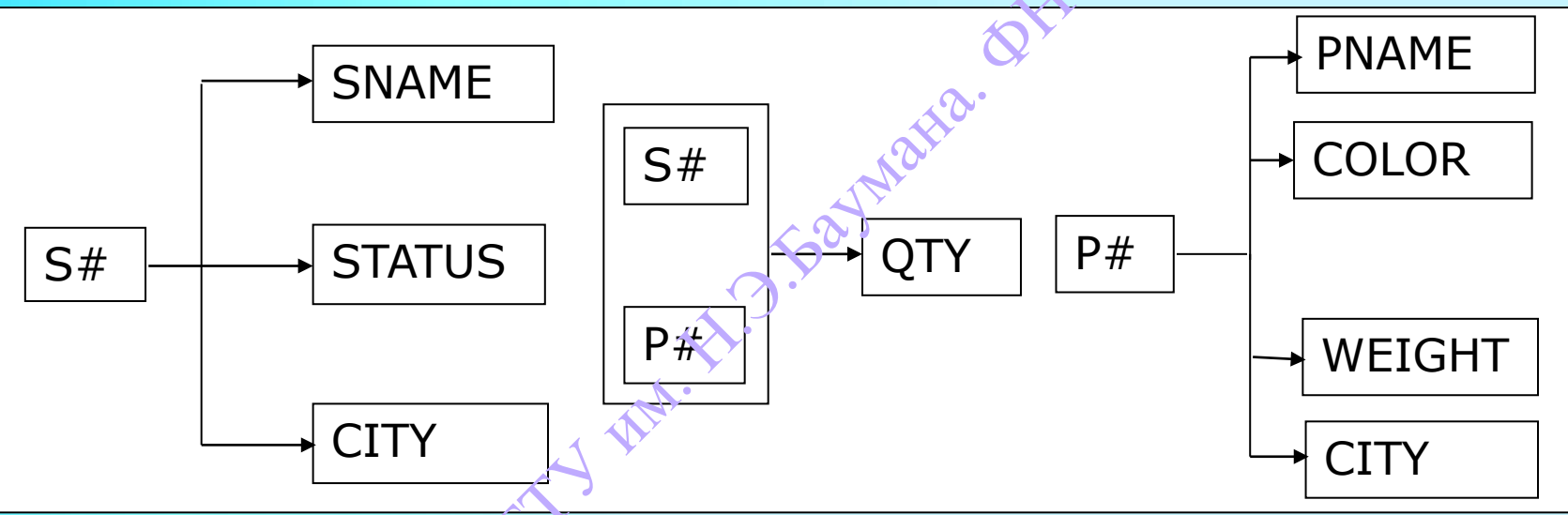


Рисунок 7.1. Диаграмма ФЗ для переменных отношения S#, P# и SP#.

Нормальные формы.

Обычно выделяется следующая последовательность нормальных форм:

- первая нормальная форма (1NF);
- вторая нормальная форма (2NF);
- третья нормальная форма (3NF);
- нормальная форма Бойса-Кодда (BCNF);
- четвертая нормальная форма (4NF);
- пятая нормальная форма, или нормальная форма проекции-соединения (5NF или PJ/NF)

Первая нормальная форма

Таблица находится в первой **нормальной форме**, если каждый её атрибут атомарен и все кортежи различны.

Неприводимые функциональные зависимости и вторая нормальная форма

Множество функциональных зависимостей S называется **неприводимым** тогда и только тогда, когда оно обладает следующими свойствами.

1. Правая (зависимая) часть каждой функциональной зависимости из множества S содержит только один атрибут (т.е. является одноэлементным множеством).
2. Функциональная зависимость является **неприводимой слева**.
3. Ни одна функциональная зависимость из множества S не может быть удалена из множества S без изменения его замыкания S^+ .

Пусть задана переменная-отношение FIRST{S#,STATUS,CITY,P#,QTY}
{S#,P#} — первичный ключ. FIRST находится в 1НФ.

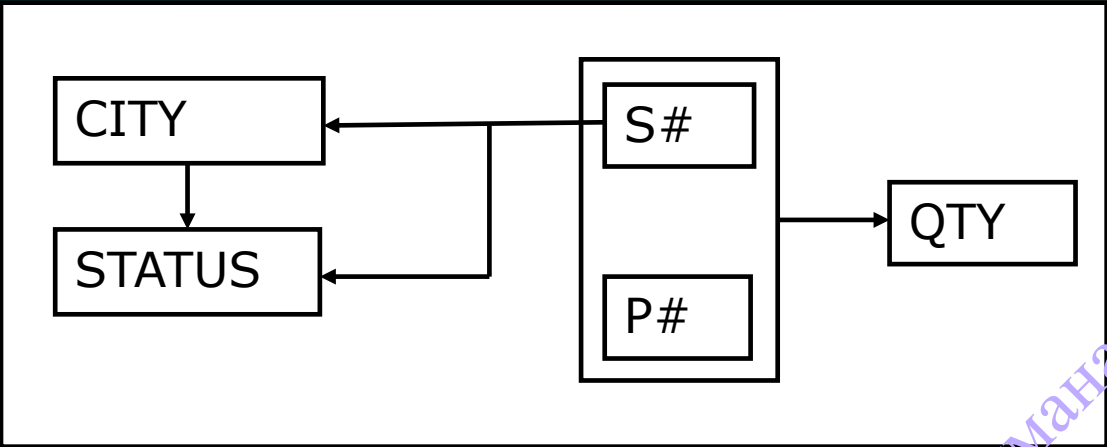


Рис. 7.2. Диаграмма ФЗ переменной-отношения FIRST.

ФЗ {S#,P#}→CITY не является неприводимой, вызывает аномалии обновления.

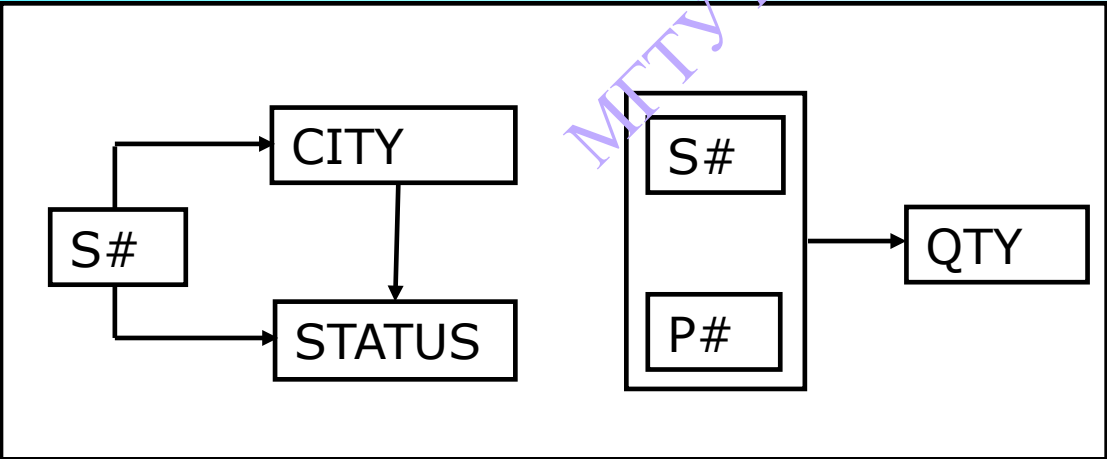


Рис. 7.3. Диаграммы ФЗ в переменных-отношениях SECOND и SP.

FIRST				
S#	STATUS	CITY	P#	QTY
S ₁	20	London	P ₁	300
S ₁	20	London	P ₂	200
S ₁	20	London	P ₃	400
S ₁	20	London	P ₄	200
S ₁	20	London	P ₅	100
S ₁	20	London	P ₆	100
S ₂	10	Paris	P ₁	300
S ₂	10	Paris	P ₂	400
S ₃	10	Paris	P ₂	200
S ₄	20	London	P ₂	200
S ₄	20	London	P ₄	300
S ₄	20	London	P ₅	400

SECOND		
S#	STATUS	CITY
S ₁	20	London
S ₂	10	Paris
S ₃	10	Paris
S ₄	20	London

SP		
S#	P#	QTY
S ₁	P ₁	300
S ₁	P ₂	200
S ₁	P ₃	400
S ₁	P ₄	200
S ₁	P ₅	100
S ₁	P ₆	100
S ₂	P ₁	300
S ₂	P ₂	400
S ₃	P ₂	200
S ₄	P ₂	200
S ₄	P ₄	300
S ₄	P ₅	400

Таблица 7.4
Возможное тело значения
переменной-отношения FIRST

Таблица 7.4
Возможные тела значений перемен-
ных - отношений SECOND и SP.

SECOND		
S#	STATUS	CITY
S ₁	20	London
S ₂	10	Paris
S ₃	10	Paris
S ₄	20	London
S ₅	30	Athens

SP		
S#	P#	QTY
S ₁	P ₁	300
S ₁	P ₃	400
S ₁	P ₄	200
S ₁	P ₅	100
S ₁	P ₆	100
S ₂	P ₁	300
S ₄	P ₄	300
S ₄	P ₅	400

SECOND		
S#	STATUS	CITY
S ₁	20	Rome
S ₂	10	Paris
S ₃	10	Paris
S ₄	20	London
S ₅	30	Athens

Таблица 7.5
Тело значения
переменной-отношения
SECOND после операции
INSERT (S₅;30;Athens)

Таблица 7.6
Тело значения
переменной-отношения
SP после операции
DELETE P₂

Таблица 7.7
Тело значения
переменной-отношения
SECOND после операции
UPDATE (S₁;Рим)

Вторая нормальная форма

Переменная отношения находится во **второй нормальной форме (2НФ)** тогда и только тогда, когда она находится в *первой нормальной форме*, и каждый неключевой атрибут неприводимо функционально зависит от первичного ключа.

Пусть дана переменная-отношение R , имеющая следующий вид.

$R\{A, B, C, D\}$

PRIMARY KEY $\{A, B\}$ (т.е. $\{A, B\} \rightarrow C, \{A, B\} \rightarrow D$)

Предполагается ФЗ $A \rightarrow D$

Процедура нормализации предусматривает замену переменной-отношения R следующими двумя проекциями R_1 и R_2 .

$R_1\{A, D\}$

PRIMARY KEY $\{A\}$

$R_2\{A, B, C\}$

PRIMARY KEY $\{A, B\}$

FORIGN KEY $\{A\}$ REFERENCES R_1

Нетранзитивные функциональные зависимости и третья нормальная форма

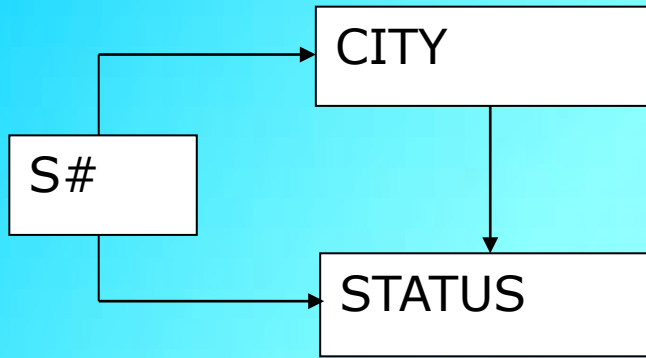


Рис. 7.4. Диаграмма ФЗ в переменной-отношения SECOND.

Пусть STATUS характеризует город, где находится поставщик, а не поставщика.

Транзитивная ФЗ $S\# \rightarrow STATUS$, через ФЗ $S\# \rightarrow CITY$ и $CITY \rightarrow STATUS$, вызывает аномалии обновления

Нельзя поместить в БД сведения о статусе города без данных о поставщике. При удалении кортежа с данными о единственном поставщике города, будет утеряна информация о статусе города. Для изменения статуса города надо изменить значения атрибута STATUS во всех кортежах для всех поставщиков данного города.

Возможная декомпозиция переменной-отношения SECOND на две переменные-отношения – SC(S#, CITY) и CS(CITY, STATUS).

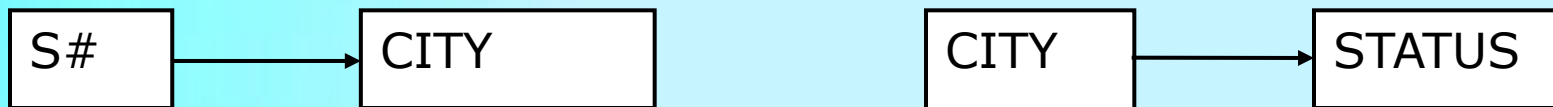


Рис. 7.5. Диаграммы ФЗ в отношениях SC и CS.

Третья нормальная форма

Переменная отношения находится в **третьей нормальной форме (ЗНФ)** тогда и только тогда, когда она находится во **второй нормальной форме**, и каждый неключевой атрибут нетранзитивно функционально зависит от первичного ключа.

Пусть дана переменная-отношение R с атрибутами A, B, C.

R {A, B, C} PRIMARY KEY {A} (т.е. $A \rightarrow B$, $A \rightarrow C$)
Имеется ФЗ $B \rightarrow C$

Процедура нормализации предусматривает декомпозицию этой переменной-отношения на переменные-отношения R_1 и R_2 .

R_1 {B, C}
PRIMARY KEY {B}

R_2 {A, B}
PRIMARY KEY {A}
FOREIGN KEY {B} REFERENCES R_1

Сохранение зависимостей. Независимые проекции отношений. Теорема Риссанена. (15.04.16)

Рассмотрим переменную-отношение SECOND с ФЗ $S\# \rightarrow CITY$ и $CITY \rightarrow STATUS$, и транзитивной зависимостью $S\# \rightarrow STATUS$.

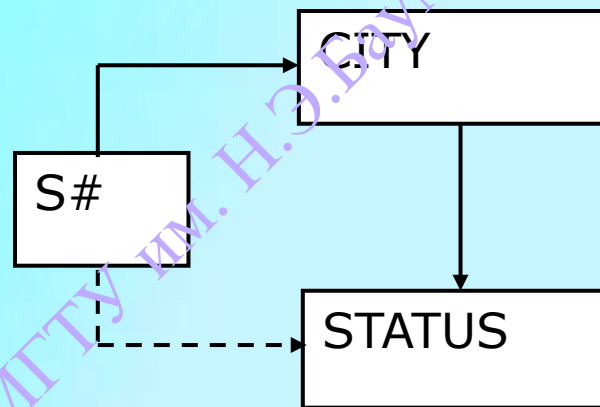


Рисунок 7.6 переменная-отношение SECOND

Аномалии обновления, сопровождающие переменную-отношение SECOND можно преодолеть посредством ее декомпозиции с последующей заменой двумя проекциями в ЗНФ.

Существуют три варианта декомпозиции

Вариант 1. Декомпозиция **A**

SC {S#, CITY}
CS {CITY, STATUS}

Вариант 2. декомпозиция **B**.

SC {S#, CITY}
SS {S#, STATUS}

Вариант 3. декомпозиция **C**.

SSt {S#, STATUS}
CSt {CITY, STATUS}

Ограничение БД — транзитивная ФЗ

$S\# \rightarrow STATUS$, выполняется автоматически, при выполнении ограничений $S\# \rightarrow CITY$ и $CITY \rightarrow STATUS$.

ФЗ $CITY \rightarrow STATUS$ превращается в ограничение БД, охватывающее две переменные-отношения.

Не является допустимой декомпозицией, т.к. сопровождается потерей информации.

Переменная- отношение SECOND не является естественным соединением значений отношений SSt и CSt по атрибуту STATUS.

SSt		CSt	
S#	STATUS	CITY	STATUS
S ₁	20	Rome	20
S ₂	10	Paris	10
S ₃	10	Paris	10
S ₄	20	London	20
S ₅	30	Athens	30

Таблица 7.9 декомпозиция C

Теорема Риссанена

Проекция R_1 и R_2 переменной-отношения R будут **независимы** тогда и только тогда, когда:

1. каждая ФЗ в переменной-отношении R является логическим следствием ФЗ в ее проекциях R_1 и R_2 ;
2. общие атрибуты проекций R_1 и R_2 образуют потенциальный ключ по хотя бы для одной из этих проекций.

Определение.

Переменная-отношение называется **атомарной**, если она не может быть подвергнута декомпозиции с получением независимых проекций.

Требование сохранения зависимостей. Нормализация всегда должна предусматривать декомпозицию переменных-отношений на независимые проекции.

Замечания.

Пусть дано отношение R , которое после выполнения всех этапов нормализации заменяется множеством отношений R_1, R_2, \dots, R_n (проекциями R).

Пусть также задано множество ФЗ S , имеющих место в исходной переменной-отношении R , и множество ФЗ S_1, S_2, \dots, S_n , выполняющихся в отношениях R_1, R_2, \dots, R_n .

Каждая ФЗ в множестве S_i будет иметь отношение только к атрибутам проекции R_i (где $i=1, 2, 3, \dots, n$). В результате реализация ограничений (устанавливаемых существующими ФЗ) для любого данного множества S_i представляется достаточно простой задачей.

Пусть S' является объединением множеств зависимостей S_1, S_2, \dots, S_n . В общем случае равенство $S'=S$ не выполняется. Для декомпозиции с сохранением зависимостей достаточно, чтобы были равны замыкания множеств S и S'