

Проектирование реляционных баз данных на основе принципов нормализации

Классический подход, представляет весь процесс проектирования базы данных в терминах реляционной модели данных методом последовательных приближений к удовлетворительному набору схем отношений.

Исходная точка - представление предметной области в виде одного или нескольких отношений с большим количеством атрибутов.

Шаг проектирования – модификация схем отношений, так, чтобы они обладали некоторым новым «улучшенными» свойствами.

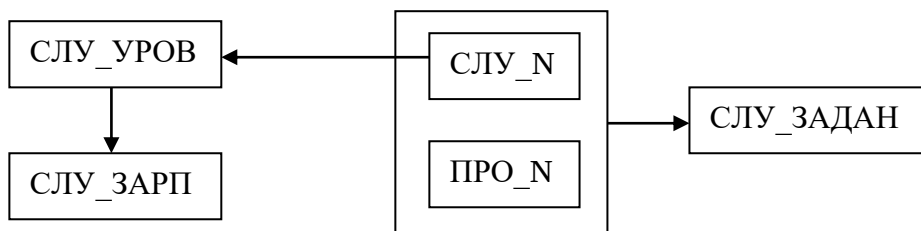
Процесс проектирования принято называть процессом **нормализации схем отношений**, или другими словами декомпозиции исходного отношения, находящегося в предыдущей нормальной форме, на два или более отношений, которые удовлетворяют требованиям следующей нормальной формы.

Нормальная форма представляет собой определенный набор ограничений. Считают, что отношение находится в некоторой нормальной форме, если оно удовлетворяет данному набору ограничений.

Каждая следующая нормальная форма обладает свойствами, в некотором смысле, лучшими, чем предыдущая. При этом свойства всех предыдущих форм сохраняются.

Минимальные функциональные зависимости и вторая нормальная форма

Диаграмма минимального множества FD отношения
СЛУЖАЩИЕ_ПРОЕКТЫ_ЗАДАНИЯ



СЛУЖАЩИЕ ПРОЕКТЫ ЗАДАНИЯ

СЛУ N	СЛУ УРОВ	СЛУ ЗАРП	ПРО N	СЛУ ЗАДАН
34	2	22000	1	A
36	1	18000	1	B
37	1	18000	1	D
34	2	22000	2	D
35	3	30000	2	C

Данное отношение удовлетворяет требованию первой нормальной формы 1NF: значения всех ее атрибутов **атомарны**.

Аномалии обновления, возникающие из-за наличия не минимальных функциональных зависимостей

Под **аномалиями обновления** мы будем понимать трудности, с которыми приходится сталкиваться при выполнении операций добавления кортежей в отношение (INSERT), удаления кортежей (DELETE) и модификации кортежей (UPDATE).

- **Добавление кортежей.** Мы не можем добавить кортеж с данными о служащем, который не участвует ни в одном проекте (ПРО_N - частью первичного ключа).
- **Удаление кортежей.** Мы не можем сохранить кортеж с данными о служащем, завершившем участие в своем последнем проекте.
- **Модификация кортежей.** Чтобы изменить разряд служащего, мы будем вынуждены модифицировать несколько кортежей с соответствующим значением атрибута СЛУ_N.

Неформально:

СЛУ_N → СЛУ_УРОВ - вызывает избыточность данных, описывая «служащего», а не «служащего в некотором проекте».

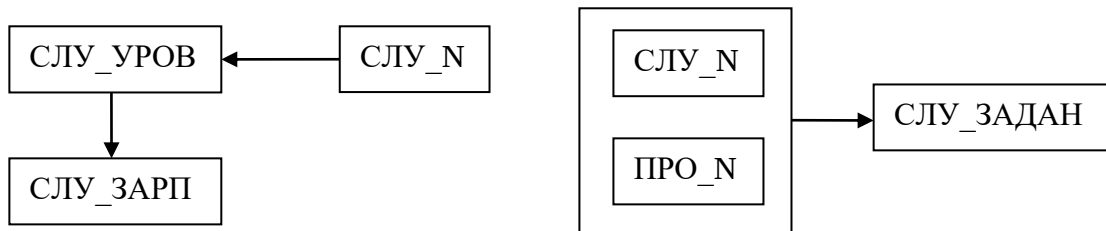
Формально:

СЛУЖАЩИЕ_ПРОЕКТЫ_ЗАДАНИЯ содержит FD, в которых детерминантом является не возможный ключ отношения, а его часть, в результате чего некоторые FD атрибутов от первичного ключа не являются минимальными.

(Например FD { СЛУ_N, ПРО_N } → СЛУ_УРОВ).

Возможная декомпозиция

СЛУЖ {СЛУ_Н, СЛУ_УРОВ, СЛУ_ЗАРП} и СЛУЖ_ПРО_ЗАДАН {СЛУ_Н, ПРО_Н, СЛУ_ЗАДАН}
 $\exists$ FD СЛУ_Н $\rightarrow$ {СЛУ_УРОВ, СЛУ_ЗАРП} $\Rightarrow$ можно декомпозировать без потерь



СЛУЖ

СЛУ_Н	СЛУ_УРОВ	СЛУ_ЗАРП
34	2	22000
36	1	18000
37	1	18000
35	3	30000

СЛУЖ_ПРО_ЗАДАН

СЛУ_Н	ПРО_Н	СЛУ_ЗАДАН
34	1	A
36	1	B
37	1	D
34	2	D
35	2	C

Отношение находится во **второй нормальной форме (2NF)** тогда и только тогда, когда оно находится в первой нормальной форме, и каждый неключевой атрибут, т.е. атрибут, не входящий ни в один возможный ключ, минимально функционально зависит от первичного ключа.

СЛУЖАЩИЕ_ПРОЕКТЫ_ЗАДАНИЯ не находится в 2NF

FD {СЛУ_Н, ПРО_Н} $\rightarrow$ СЛУ_УРОВ не является минимальной (ПРО_Н – можно удалить).

СЛУЖ и СЛУЖ_ПРО_ЗАДАН находятся в 2NF

все неключевые атрибуты минимально зависят от первичных ключей СЛУ_Н и {СЛУ_Н, ПРО_Н}.

Аномалии обновлений, возникающие из-за наличия транзитивных функциональных зависимостей в отношении СЛУЖ

Наличие транзитивной $СЛУ_N \rightarrow СЛУ_ЗАРП$ ($СЛУ_N \rightarrow СЛУ_УРОВ$ и $СЛУ_УРОВ \rightarrow СЛУ_ЗАРП$) вызывает избыточностью хранения атрибута $СЛУ_ЗАРП$ в каждом кортеже служащего:

- **Добавление кортежей.** Невозможно сохранить данные о новом разряде (и соответствующем ему размере зарплаты), пока не появится служащий с новым разрядом.
- **Удаление кортежей.** При увольнении последнего служащего с данным разрядом мы утратим информацию о наличии такого разряда и соответствующем размере зарплаты.
- **Модификация кортежей.** При изменении размера зарплаты, соответствующей некоторому разряду, мы будем вынуждены изменить значение атрибута $СЛУ_ЗАРП$ в кортежах всех служащих, которым назначен этот разряд.

Возможная декомпозиция:

$СЛУЖ1 \{СЛУ_N, СЛУ_УРОВ\}$ и $УРОВ \{СЛУ_УРОВ, СЛУ_ЗАРП\}$

($\exists FD \text{ } СЛУ_УРОВ \rightarrow СЛУ_ЗАРП \Rightarrow$ можно декомпозировать без потерь)



СЛУЖ1

СЛУ N	СЛУ УРОВ
34	2
36	1
37	1
35	3

УРОВ

СЛУ УРОВ	СЛУ ЗАРП
2	22000
1	18000
3	30000

Переменная отношения находится в **третьей нормальной форме (3NF)** в том и только в том случае, когда она находится во второй нормальной форме, и каждый неключевой атрибут нетранзитивно функционально зависит от первичного ключа.

Независимые проекции отношений. Теорема Риссанена

Возможные декомпозиции СЛУЖ {СЛУ_Н, СЛУ_УРОВ, СЛУ_ЗАРП}:

1. $A\{СЛУ_Н, СЛУ_УРОВ\}$ и $B\{СЛУ_УРОВ, СЛУ_ЗАРП\}$
2. $A\{СЛУ_Н, СЛУ_УРОВ\}$ и $B\{СЛУ_Н, СЛУ_ЗАРП\}$
3. $A\{СЛУ_Н, СЛУ_ЗАРП\}$ и $B\{СЛУ_УРОВ, СЛУ_ЗАРП\}$

3 – с потерями (по теореме Хита)

2 – не устраняет аномалий обновления (невозможно сохранить информацию о разряде)

1 – ок

В процессе нормализации декомпозиция отношения на независимые проекции является предпочтительной. Необходимые и достаточные условия независимости проекций отношения обеспечивает теорема Риссанена.

Теорема Риссанена

Проекции r_1 и r_2 отношения r являются **независимыми** тогда и только тогда, когда:

- каждая FD в отношении r логически следует (выводится по Армстронгу) из FD в r_1 и r_2 ; ($\forall FD \in r \Leftarrow \{FD \in r_1 \cup FD \in r_2\}$)
- общие атрибуты r_1 и r_2 образуют возможный ключ хотя бы для одного из этих отношений. ($Hr_1 \cap Hr_2$ – возм. ключ r_1 или r_2)

Пример 1: $СЛУ_Н \rightarrow СЛУ_ЗАРП \Leftarrow СЛУ_Н \rightarrow СЛУ_УРОВ$ и $СЛУ_УРОВ \rightarrow СЛУ_ЗАРП$

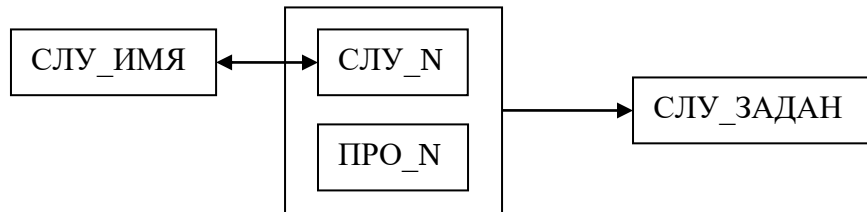
Пример 2: $СЛУ_УРОВ \rightarrow СЛУ_ЗАРП \not\Leftarrow СЛУ_Н \rightarrow СЛУ_УРОВ$ и $СЛУ_Н \rightarrow СЛУ_ЗАРП$.

Атомарным отношением называется отношение, которое невозможно декомпозировать на независимые проекции.

Далеко не всегда для неатомарных отношений требуется декомпозиция на атомарные проекции. При выборе способа декомпозиции нужно стремиться к получению независимых проекций, но не обязательно атомарных.

Аномалии обновлений, связанные с наличием перекрывающихся возможных ключей

СЛУЖ_ПРО_ЗАДАН1 {СЛУ_Н, СЛУ_ИМЯ, ПРО_Н, СЛУ_ЗАДАН} с множеством FD:



СЛУЖ ПРО ЗАДАН1

СЛУ N	СЛУ ИМЯ	ПРО N	СЛУ ЗАДАН
34	Иванов	1	A
35	Федоров	2	B
34	Иванов	2	B
35	Федоров	1	A

В отношении СЛУЖ_ПРО_ЗАДАН1 служащие уникально идентифицируются как по номерам удостоверений, так и по именам, каждый служащий может участвовать в нескольких проектах.

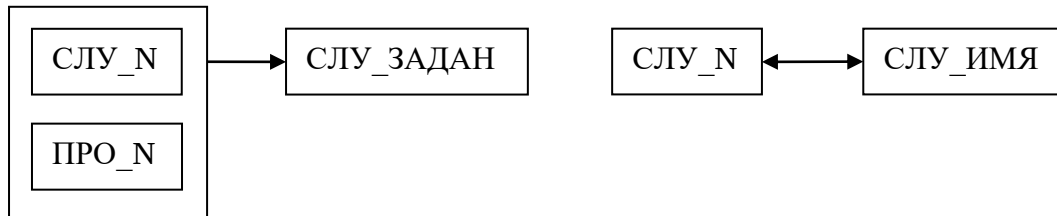
Условия 2NF и 3NF соблюдены, однако имеются **аномалии обновления**: в случае изменения имени служащего требуется обновить атрибут СЛУ_ИМЯ в нескольких карточках, иначе будет нарушена $\text{СЛУ_Н} \rightarrow \text{СЛУ_ИМЯ}$.

Нормальная форма Бойса-Кодда

Отношение находится в **нормальной форме Бойса-Кодда (BCNF)** в том и только в том случае, когда оно находится в **3NF** и любая нетривиальная и минимальная FD этого отношения имеет в качестве детерминанта некоторый возможный ключ.

СЛУЖ_ПРО_ЗАДАН1 может быть приведена к BCNF путем одной из двух декомпозиций:

1. СЛУЖ_ИМЯ {СЛУ_N, СЛУ_ИМЯ} и НОМ_ЗАДАН {СЛУ_N, ПРО_N, СЛУ_ЗАДАН},
2. СЛУЖ_ИМЯ {СЛУ_N, СЛУ_ИМЯ} и ИМЯ_ЗАДАН {СЛУ_ИМЯ, ПРО_N, СЛУ_ЗАДАН}



СЛУЖ_ИМЯ

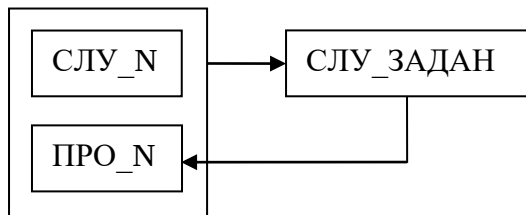
СЛУ_N	СЛУ_ИМЯ
34	Иванов
35	Федоров

НОМ_ЗАДАН

СЛУ_N	ПРО_N	СЛУ_ЗАДАН
34	1	A
35	2	B
34	2	B
35	1	A

Всегда ли следует стремиться к BCNF?

- служащий может участвовать в нескольких проектах
- в каждом проекте служащий выполняет одно задание
- задание может быть связано только с одним проектом.



СЛУЖ ПРО ЗАДАН

СЛУ N	ПРО N	СЛУ ЗАДАН
34	1	A
35	1	A
41	2	B
34	2	C
41	1	D

Возможные ключи:

{СЛУ_N, ПРО_N}

{СЛУ_N, СЛУ_ЗАДАН}

Находится в 3NF (нет неключевых атрибутов), но **НЕ** находится в BCNF ($\text{СЛУ_ЗАДАН} \rightarrow \text{ПРО_N}$)

Аномалии обновления: невозможно удалить данные о единственном служащем, выполняющем задание в некотором проекте, не утратив информацию об этом задании (в каком проекте оно выполняется).

Возможная декомпозиция для получения BCNF устраняющей аномалии:

СЛУЖ ЗАДАН

СЛУ N	СЛУ ЗАДАН
34	A
35	A
41	B
34	C
41	D

ПРО ЗАДАН

ПРО N	СЛУ ЗАДАН
1	A
2	B
2	C
1	D

Имеется **проблема**: добавление <34, D> в отношение СЛУЖ_ЗАДАН система должна запретить (D относится к проекту 1, в котором служащий 34 уже выполняет задание A).

$\{\text{СЛУ_N}, \text{ПРО_N}\} \rightarrow \text{СЛУ_ЗАДАН}$ не выводится из зависимостей результирующих отношений. Эта FD должна быть определена как ограничение целостности БД общего вида.

Т.о. проекции СЛУЖ_ЗАДАН и ПРО_ЗАДАН - не являются независимыми, а отношение СЛУЖ_ПРО_ЗАДАН атомарно, хотя и не находится в BCNF.

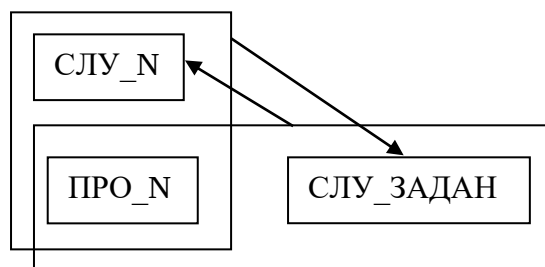
Отношение, с двумя перекрывающимися возможными ключами, находящееся в BCNF

Служащий может участвовать в нескольких проектах.

В каждом проекте служащий выполняет только одно задание.

Каждое задание может присутствовать в разных проектах.

В каждом проекте задание выполняться только одним служащим.



СЛУЖ ПРО ЗАДАН

СЛУ N	ПРО N	СЛУ ЗАДАН
34	1	A
35	1	B
41	2	B
34	2	A
41	1	D

Возможные ключи: {СЛУ_N, ПРО_N} и {ПРО_N, СЛУ_ЗАДАН},
но отношение находится в BCNF, поскольку эти ключи являются единственными
детерминантами, а значит эти зависимости минимальные.