

Базы данных и SQL

Нормальные формы

ПРОЕКТИРОВАНИЕ РЕЛЯЦИОННЫХ БАЗ ДАННЫХ С ИСПОЛЬЗОВАНИЕМ НОРМАЛИЗАЦИИ

- **Нормализация отношений** – это пошаговый обратимый процесс композиции или декомпозиции исходных отношений в отношения, обладающие лучшими свойствами при включении, изменении и удалении данных, назначение им ключей по определенным правилам нормализации и выявление всех возможных функциональных зависимостей.

НОРМАЛЬНЫЕ ФОРМЫ

- первая нормальная форма (1 НФ);
- вторая нормальная форма (2 НФ);
- третья нормальная форма (3 НФ);
- нормальная форма Бойса-Кодда (НФБК);
- четвертая нормальная форма (4 НФ);
- пятая нормальная форма, или нормальная форма проекции-соединения (5 НФ или ПС/НФ).

КЛЮЧИ

- **Ключ отношения** – это группа из одного или более атрибутов, которая уникальным образом идентифицирует кортеж. Таким образом, ключ имеет уникальные в рамках отношения значения. Если ключ состоит из нескольких атрибутов, он называется составным. Ключей может быть несколько; основной ключ – **первичный**, его значения не могут обновляться. Другие ключи называются **возможными** или **потенциальными** ключами.

ОТНОШЕНИЕ СЕКЦИЯ (простой ключ)

Код студента	Секция	Плата
123	Лыжи	3
324	Плавание	4
210	Волейбол	5
434	Плавание	4

ОТНОШЕНИЕ СЕКЦИЯ (составной ключ)

Код студента	Секция	Плата
210	Аэробика	7
324	Плавание	4
434	Аэробика	7
123	Лыжи	3
434	Плавание	4
210	Волейбол	5
123	Плавание	4

ФУНКЦИОНАЛЬНЫЕ ЗАВИСИМОСТИ

- В отношении R атрибут Y функционально зависит от атрибута X (X и Y могут быть составными) в том и только в том случае, если каждому значению X соответствует в точности одно значение Y : $R.X \rightarrow R.Y$.

Функциональные зависимости обозначаются следующим образом:

НомерЗачётнойКнижки \rightarrow ФИО

НомерЗачётнойКнижки \rightarrow Курс

БАЗОВЫЕ ОПРЕДЕЛЕНИЯ

- **Полная функциональная зависимость**
- **Детерминант**
- **Транзитивная функциональная зависимость**
- **Неключевой атрибут**
- **Взаимно независимые атрибуты**

АНОМАЛИЯМИ МОДИФИКАЦИИ

Аномалия удаления

Избыточность

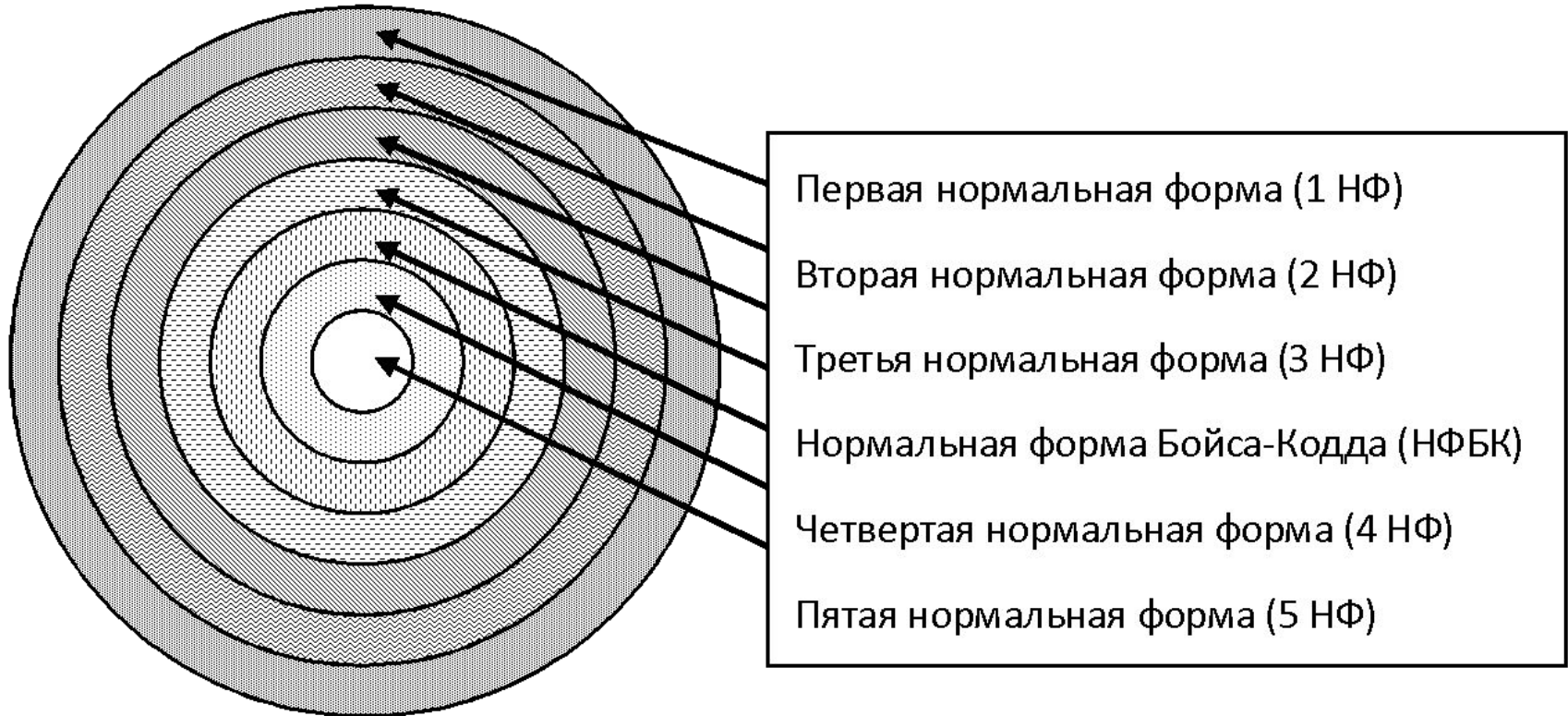
Аномалия
обновления

Аномалия вставки

Отношение Секция
(Код студента, Секция, Плата)
Ключ: (Код студента)

Код студента	Секция	Плата
123	Лыжи	3
324	Плавание	4
210	Волейбол	5
434	Плавание	4

КЛАССЫ ОТНОШЕНИЙ



Взаимосвязь нормальных форм

ПЕРВАЯ НОРМАЛЬНАЯ ФОРМА

- в отношении не может быть двух одинаковых кортежей;
- порядок следования кортежей несущественен;
- каждый атрибут должен иметь уникальное имя, но порядок следования атрибутов несущественен.
- все значения атрибутов содержат только одно значение из соответствующего домена, и в качестве значений не допускаются ни повторяющиеся группы, ни массивы;
- **каждый кортеж в отношении должен иметь одно и то же количество атрибутов; повторяющиеся атрибуты нужно перемещать в отдельные связанные отношения;**

ПЕРВАЯ НОРМАЛЬНАЯ ФОРМА

в отношении не может быть двух одинаковых кортежей

Фамилия	Рабочий телефон
Долгополова	2095570
Прохорова	2095045
Кремень	2095361
Кремень	2095361

ПЕРВАЯ НОРМАЛЬНАЯ ФОРМА

порядок следования кортежей несущественен

<u>Номер зачетной книжки</u>	Фамилия	Имя	Телефон
111111	Иванов	Иван	+375171234567, Белтелеком
222222	Петров	Петр	+375172345634, Белтелеком
			+375291234560, МТС
333333	Сидоров	Сидор	+375258745779, Life

ПЕРВАЯ НОРМАЛЬНАЯ ФОРМА

все значения атрибутов содержат только одно значение из соответствующего домена, и в качестве значений не допускаются ни повторяющиеся группы, ни массивы

<u>Номер зачетной книжки</u>	Фамилия	Имя	Телефон
111111	Иванов	Иван	+375171234567, Белтелеком
222222	Петров	Петр	+375172345634, Белтелеком +375291234560, МТС
333333	Сидоров	Сидор	+375258745779, Life

ПЕРВАЯ НОРМАЛЬНАЯ ФОРМА

каждый кортеж в отношении должен иметь одно и то же количество атрибутов;
повторяющиеся атрибуты нужно перемещать в отдельные связанные отношения

<u>Номер зачетной книжки</u>	Фамилия	Имя	Телефон1	Оператор1	Телефон2	Оператор2
111111	Иванов	Иван	+375171234567	Белтелеком		
222222	Петров	Петр	+375172345634	Белтелеком	+375291234560	МТС
333333	Сидоров	Сидор	+375258745779	Life		

ПЕРВАЯ НОРМАЛЬНАЯ ФОРМА

СТУДЕНТ(Номер зачетной книжки,
Фамилия, Имя)
Ключ: (Номер зачетной книжки)

<u>Номер зачетной книжки</u>	Фамилия	Имя
111111	Иванов	Иван
222222	Петров	Петр
333333	Сидоров	Сидор

ТЕЛЕФОН(Номер зачетной книжки,
Телефон, Оператор)
Ключ: (Номер зачетной книжки,
Телефон)

<u>Номер зачетной книжки</u>	<u>Телефон</u>	Оператор
111111	+375171234567	Белтелеком
222222	+375172345634	Белтелеком
222222	+375291234560	МТС
333333	+375258745779	Life

ВТОРАЯ НОРМАЛЬНАЯ ФОРМА

- Проверяйте составные ключи

<u>Код студента</u>	<u>Секция</u>	Плата
100	Футбол	3
100	Теннис	5
150	Плавание	4,5
150	Волейбол	3,5
200	Теннис	5
250	Волейбол	3,5
250	Плавание	4,5

Отношение СЕКЦИЯ, Ключ: (Код студента , Секция)

Нормализация отношения

СЕКЦИЯ

СТУДЕНТ(Код студента книжки,
Секция)

Ключ: (Код студента, Секция)

<u>Код студента</u>	<u>Секция</u>
100	Футбол
100	Теннис
150	Плавание
150	Волейбол
200	Теннис
250	Волейбол
250	Плавание

СЕКЦИЯ(Секция, Плата)

Ключ: (Секция)

<u>Секция</u>	<u>Плата</u>
Футбол	3
Теннис	5
Плавание	4,5
Волейбол	3,5

ВТОРАЯ НОРМАЛЬНАЯ ФОРМА

- *Отношение R находится во второй нормальной форме (2 НФ) в том и только в том случае, когда оно находится в 1 НФ, и каждый неключевой атрибут функционально полно зависит от любого возможного ключа этого отношения R .*
- Для приведения отношения во 2 НФ необходимо:
 - построить его проекцию, исключив атрибуты, которые не находятся в полной функциональной зависимости от составного ключа;
 - построить дополнительно одну или несколько проекций на часть составного ключа и атрибуты, функционально зависящие от этой части.

ТРЕТЬЯ НОРМАЛЬНАЯ ФОРМА

- Проверьте любые оставшиеся группы информации. Отношение не должно иметь транзитивных зависимостей.

<u>Код студента</u>	Общежитие	Плата
100	№ 7	18
150	№ 4	20
200	№ 1	23
250	№ 1	23
300	№ 2	20

Отношение ПРОЖИВАНИЕ (Код студента, Общежитие, Плата)

Устранение транзитивной зависимости в таблице ПРОЖИВАНИЕ

ПРОЖИВАНИЕ

(Код студента, Общежитие)

Ключ: (Код студента)

<u>Код студента</u>	Общежитие
100	№ 7
150	№ 4
200	№ 1
250	№ 1
300	№ 2

ОБЩЕЖИТИЕ

(Общежитие, Плата)

Ключ: (Общежитие)

<u>Общежитие</u>	Плата
№ 7	18
№ 4	20
№ 1	23
№ 2	20

ТРЕТЬЯ НОРМАЛЬНАЯ ФОРМА

- *Отношение R находится в третьей нормальной форме (3 НФ) в том и только в том случае, если находится во 2 НФ, и каждый неключевой атрибут не является транзитивно зависимым от какого-либо ключа R .*

Для приведения отношения в 3 НФ необходимо при наличии зависимости некоторых неключевых атрибутов от других неключевых атрибутов произвести декомпозицию отношения. Т. е. неключевые атрибуты, которые зависят от других неключевых атрибутов, вынести в отдельное отношение.

Нормальная форма Воиса- Кодда

Каждый кортеж означает, что некоторый студент изучает определенную дисциплину у указанного преподавателя

<u>Код студента</u>	<u>Предмет</u>	<u>Преподаватель</u>
100	Программирование	Расолько Г.А.
600	Психология	Егоров К.Д.
700	Программирование	Буяльская Ю.В.
200	Программирование	Расолько Г.А.
400	Психология	Егоров К.Д.
200	Психология	Белова Т.О.
600	Программирование	Буяльская Ю.В.

При этом существуют следующие ограничения:

- Каждый преподаватель может преподавать только по один предмет.
- Один предмет могут преподавать несколько преподавателей.
- Для студента один предмет ведет только один преподаватель
- Будем также предполагать, что у преподавателей не может быть одинаковых фамилий.

Нормальная форма Бойса-Кодда

<u>Код студента</u>	<u>Предмет</u>	Преподаватель
100	Программирование	Расолько Г.А.
600	Психология	Егоров К.Д.
700	Программирование	Буяльская Ю.В.
200	Программирование	Расолько Г.А.
400	Психология	Егоров К.Д.
200	Психология	Белова Т.О.
600	Программирование	Буяльская Ю.В.

- Отношение находится в 3 НФ.
- Наличие аномалий.
- В данном отношении есть два потенциальных ключа (Код студента, Предмет) и (Код студента, Преподаватель)
- Оба ключа являются составными.
- Ключи имеют общий атрибут Код студента, т.е. перекрываются

Отношение Учёба

<u>Код студента</u>	<u>Предмет</u>	<u>Преподаватель</u>
100	Программирование	Расолько Г.А.
600	Психология	Егоров К.Д.
700	Программирование	Буяльская Ю.В.
200	Программирование	Расолько Г.А.
400	Психология	Егоров К.Д.
200	Психология	Белова Т.О.
600	Программирование	Буяльская Ю.В.

В данном отношении существует функциональная зависимость Преподаватель \rightarrow Предмет и Преподаватель при этом не является потенциальным ключом, следовательно отношение не находится в НФБК

Отношение R находится в нормальной форме Бойса-Кодда (НФБК) в том и только в том случае, если каждый детерминант является ВОЗМОЖНЫМ КЛЮЧОМ

Приведение отношения Учёба к нормальной форме Бойса- Кодда

УЧЁБА

Ключ: (Код студента,
Преподаватель)

<u>Код студента</u>	<u>Преподаватель</u>
100	Расолько Г.А.
600	Егоров К.Д.
700	Буяльская Ю.В.
200	Расолько Г.А.
400	Егоров К.Д.
200	Белова Т.О.
600	Буяльская Ю.В.

ПРЕПОДАВАНИЕ

Ключ: Преподаватель

<u>Предмет</u>	<u>Преподаватель</u>
Программирование	Расолько Г.А.
Психология	Егоров К.Д.
Программирование	Буяльская Ю.В.
Психология	Белова Т.О.

КОГДА ЗНФ \leftrightarrow НФБК

Отношение имеет два (или больше) потенциальных ключа

Эти потенциальные ключи являются составными

Они перекрываются (т.е. имеют по крайней мере один общий атрибут)

На практике комбинация всех трёх условий встречается редко, и для отношений, в которых не выполняются все эти три условия, ЗНФ и НФБК полностью эквивалентны

ЧЕТВЁРТАЯ НОРМАЛЬНАЯ ФОРМА

Таблица не должна содержать полей для двух или более независимых, различных фактов

<u>Код студента</u>	<u>Язык программирования</u>	<u>Иностранный язык</u>
110011	C++	Английский
110011	Java	Английский
110011	C++	Немецкий
110011	Java	Немецкий
333333	Pascal	Английский
333333	Java	Английский
896700	C++	Немецкий

Отношение **НАВЫКИ**.

Ключ: (Код студента, Язык программирования, Иностранный язык)

Многозначная зависимость

- Атрибут X многозначно определяет атрибут Y в R (или Y многозначно зависит от X), если каждому значению атрибута X соответствует множество (возможно, пустое) значений атрибута Y , никак не связанных с другими атрибутами R . То есть для наличия в отношении многозначной зависимости необходимо иметь как минимум три атрибута.
- Многозначная зависимость существует, когда отношение имеет минимум три атрибута, причем два из них являются многозначными, а их значения зависят только от третьего атрибута. В отношении $R(A, B, C)$ существует многозначная зависимость, если A многозначным образом определяет B и C , а сами B и C не зависят друг от друга.

Устранение многозначной зависимости в отношении НАВЫКИ

ПРОГРАММИРОВАНИЕ(Код студента, Язык программирования)
Ключ: (Код студента, Язык программирования)

<u>Код студента</u>	<u>Язык программирования</u>
110011	C++
110011	Java
333333	Pascal
333333	Java
896700	C++

ИНОСТРАННЫЙ_ЯЗЫК(Код студента, Иностранный язык)
Ключ: (Код студента, Иностранный язык)

<u>Код студента</u>	<u>Иностранный язык</u>
110011	Английский
110011	Немецкий
333333	Английский
896700	Немецкий

Для приведения отношения в 4 НФ необходимо при обнаружении в отношении нетривиальных многозначных зависимостей провести декомпозицию для исключения таких зависимостей

ЧЕТВЕРТАЯ НОРМАЛЬНАЯ ФОРМА

Страны, которые посетил студент, и иностранные языки, которые знает студент

<u>Студент</u>	<u>Страна</u>	<u>Иностранный язык</u>
Иванов	Египет	Английский
Иванов	Турция	Английский
Иванов	Египет	Немецкий
Иванов	Турция	Немецкий
Петров	Болгария	Английский
Петров	Турция	Английский
Сидоров	Египет	Немецкий

<u>Студент</u>	<u>Страна</u>
Иванов	Египет
Иванов	Турция
Петров	Болгария
Петров	Турция
Сидоров	Египет

<u>Студент</u>	<u>Иностранный язык</u>
Иванов	Английский
Иванов	Немецкий
Петров	Английский
Сидоров	Немецкий

ПЯТАЯ НОРМАЛЬНАЯ ФОРМА

- Возможность восстановить исходные данные, которые были удалены из таблиц по причине их избыточности

ПЯТАЯ НОРМАЛЬНАЯ ФОРМА

Предположим, что нужно хранить данные об ассортименте нескольких дилеров, торгующих продукцией нескольких фирм (перечень товаров фирм может пересекаться):

- Дилеры (Иванов, Петров) представляют Производителей (Mercedes и Volvo).
- Производители выпускают Товары (автобусы и грузовики).
- Дилеры продают товары.

Наиболее общий

<u>Дилер</u>	<u>Производителе</u> <u>ль</u>	<u>Товар</u>
Иванов	Mercedes	автобус
Иванов	Mercedes	грузови к
Иванов	Volvo	автобус
Иванов	Volvo	грузови к
Петров	Mercedes	автобус
		грузови

А если 1000 Дилеров, 1000 Производителей, 1000

ПЯТАЯ НОРМАЛЬНАЯ ФОРМА

<u>Дилер</u>	<u>Производителе</u>
Иванов	Mercedes
Иванов	Volvo
Петров	Mercedes
Петров	Volvo

<u>Производителе</u>	<u>Товар</u>
Mercedes	автобус
	грузови к
Mercedes	
Volvo	автобус
	грузови к
Volvo	

ПЯТАЯ НОРМАЛЬНАЯ ФОРМА

Предположим, что нужно хранить данные об ассортименте нескольких дилеров, торгующих продукцией нескольких производителей (перечень товаров у разных производителей может пересекаться):

- Дилеры (Иванов, Петров) представляют Производителей (Mercedes и Volvo).
- Производители выпускают Товары (автобусы и грузовики).
- Дилеры продают товары.

В большинстве реальных случаев возникают дополнительные ограничения (бизнес-правила).

Пусть дилеру требуется лицензия на некоторые товары. Например, у Иванова есть лицензия на продажу автобусов и грузовиков Mercedes и Volvo, а у Петрова только лицензия на продажу автобусов:

<u>Дилер</u>	<u>Производитель</u> <u>ь</u>	<u>Товар</u>
Иванов	Mercedes	автобус
Иванов	Mercedes	грузовик
Иванов	Volvo	автобус
Иванов	Volvo	грузовик
Петров	Mercedes	автобус

ПЯТАЯ НОРМАЛЬНАЯ ФОРМА

<u>Дилер</u>	<u>Производителе</u> <u>ль</u>	<u>Товар</u>
Иванов	Mercedes	автобус
Иванов	Mercedes	грузовик
Иванов	Volvo	автобус
Иванов	Volvo	грузовик
Петров	Mercedes	автобус
Петров	Volvo	автобус

Нельзя выполнить декомпозицию на две таблицы

<u>Дилер</u>	<u>Производителе</u> <u>ль</u>
Иванов	Mercedes
Иванов	Volvo

<u>Производителе</u> <u>ль</u>	<u>Товар</u>
Mercedes	автобус
Mercedes	грузовик
Volvo	автобус
Volvo	грузовик

ФОРМА

<u>Дилер</u>	<u>Производитель</u>	<u>Производитель</u>	<u>Товар</u>
Иванов	Mercedes	Mercedes	автобус
Иванов	Volvo	Mercedes	грузовик
Петров	Mercedes	Volvo	автобус
Петров	Volvo	Volvo	грузовик
		<u>Производитель</u>	<u>Товар</u>
		Mercedes	автобус
		Mercedes	грузовик
		Volvo	автобус
	Иванов	Volvo	грузовик
	Петров	Mercedes	автобус
	Петров	Mercedes	грузовик
	Петров	Volvo	автобус
	Петров	Volvo	грузовик

При соединении двух таблиц появляются «лишние» строки

ПЯТАЯ НОРМАЛЬНАЯ ФОРМА

<u>Дилер</u>	<u>Производитель</u>	<u>Товар</u>
Иванов	Mercedes	автобус
Иванов	Mercedes	грузовик
Иванов	Volvo	автобус
Иванов	Volvo	грузовик
Петров	Mercedes	автобус
Петров	Volvo	автобус

Можно выполнить декомпозицию на три таблицы

<u>Дилер</u>	<u>Производитель</u>
Иванов	Mercedes
Иванов	Volvo

<u>Производитель</u>	<u>Товар</u>
Mercedes	автобус
Mercedes	грузовик
Volvo	автобус

<u>Дилер</u>	<u>Товар</u>
Иванов	автобус
Иванов	грузовик
Петров	автобус