

Курс «Базы данных»

Тема. Принципы нормализации реляционных БД

- Аномалии модификации данных
- Функциональные зависимости
- Принципы нормализации

Аномалии модификации данных

- После составления логической схемы БД ее надо проверить на отсутствие аномалий модификации данных.
- В неправильно спроектированной схеме БД могут возникнуть аномалии операций модификации данных:
 - *аномалии обновления*
 - *аномалии удаления*
 - *аномалии добавления*

Пример

Сессия					
№ зачетной книжки	Фамилия	Имя	Отчество	Предмет	Оценка
100	Тюряпина	Елизавета	Николаевна	Информатика	Удовлетворительно
...
1100	Михайлова	Анастасия	Александровна	Базы данных	Отлично
...
100	Тюряпина	Елизавета	Николаевна	Литература	Отлично

Цель даталогического проектирования

Цель даталогического проектирования:

разработка корректной **схемы БД** в терминах выбранной модели БД.

Основой анализа корректности схемы являются анализ **функциональных зависимостей** между атрибутами отношений БД.

Функциональная зависимость

Атрибут Y некоторого отношения **функционально зависит** от X (атрибуты могут быть составными), если в любой момент времени каждому значению X соответствует одно значение Y .

Функциональная зависимость

обозначается: $X \xrightarrow{Y}$

Пример: Номер зач.кн. ФИО

Нормализация базы данных

Нормализация — это процесс преобразования отношения в состояние, обеспечивающее лучшие условия выборки, добавления, изменения и удаления данных.

Главная цель нормализации: устранение избыточности и дублирования информации в базе данных

Нормальные формы

первая нормальная форма (1NF)



вторая нормальная форма (2NF)



третья нормальная форма (3NF)



нормальная форма Бойса—Кодда
(BCNF)



четвертая нормальная форма (4NF)



пятая нормальная форма (5NF)

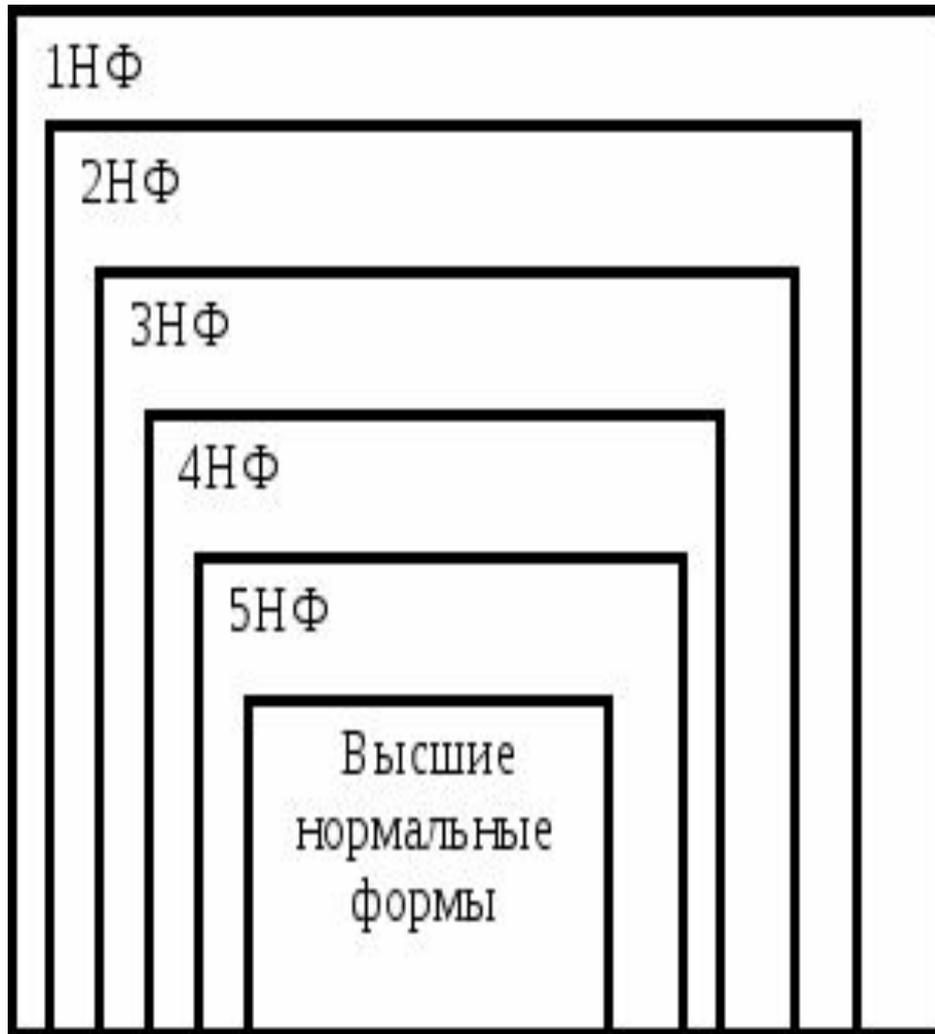
Свойства нормальных форм

Каждой нормальной форме соответствует определенный набор ограничений.

Основные свойства нормальных форм:

- каждая следующая нормальная форма ***улучшает свойства предыдущей.***
- при переходе к следующей нормальной форме ***свойства предыдущих нормальных форм сохраняются.***

Нормальные формы



← процесс начинается с 1НФ

Первая нормальная форма

Отношение находится в **первой нормальной форме**, если значения всех его атрибутов **атомарны**.

R1 - Ненормализованное отношение

КодПоставщика	КодПродукта	Продукт
P ₁	1	Сахар
	2	Соль
	13	Мука
P ₂	26	Рис
	58	Гречка
	130	Крупа манная
	162	Пшено
P ₃	474	Молоко
	891	Кефир

R2 - Нормализованное отношение

КодПоставщика	КодПродукта	Продукт
P ₁	1	Сахар
P ₁	2	Соль
P ₁	13	Мука
P ₂	26	Рис
P ₂	58	Гречка
P ₂	130	Крупа манная
P ₂	162	Пшено
P ₃	474	Молоко
P ₃	891	Кефир

Первая нормальная форма: пример

Преподаватель	День недели	Номер пары	Название дисциплины	Тип занятий	Группа
Петров В. И.	Понедельник	1	Теор. выч. проц.	Лекция	4906
	Вторник	1	Комп. графика	Лаб. раб.	4907
	Вторник	2	Комп. графика	Лаб. раб.	4906
Киров В. А.	Понедельник	2	Теория информ.	Лекция	4906
	Вторник	3	Пр-е на C++	Лаб. раб.	4907
	Вторник	4	Пр-е на C++	Лаб. раб.	4906
Серов А. А.	Понедельник	3	Защита инф.	Лекция	4944
	Среда	3	Базы данных	Лаб. раб.	4942
	Четверг	4	Базы данных	Лаб. раб.	4922 11

Первая нормальная форма:

пример

Преподаватель	День недели	Номер пары	Название дисциплины	Тип занятий	Группа
Петров В. И.	Понедельник	1	Теор. выч. проц.	Лекция	4906
Петров В. И.	Вторник	1	Комп. графика	Лаб. раб.	4907
Петров В. И.	Вторник	2	Комп. графика	Лаб. раб.	4906
Киров В. А.	Понедельник	2	Теория информ.	Лекция	4906
Киров В. А.	Вторник	3	Пр-е на C++	Лаб. раб.	4907
Киров В. А.	Вторник	4	Пр-е на C++	Лаб. раб.	4906
Серов А. А.	Понедельник	3	Защита инф.	Лекция	4944
Серов А. А.	Среда	3	Базы данных	Лаб. раб.	4942
Серов А. А.	Четверг	4	Базы данных	Лаб. раб.	4922

Недостатки первой нормальной формы

- **избыточность** — многократное повторение информации в столбцах данных
- **аномалии модификации** (обновления) данных
- **аномалии добавления** данных
- **аномалии удаления** данных

Пример:

Экзамены (ФИО, Номер зач.кн., Группа, Дисциплина, Дата экзамена, Оценка)

Избыточность данных: пример

ФИО	Номер ЗачКн	Группа	Название дисциплины	Дата	Оценка
Пупкин В. И.	323556	ММ-117	Управление данными	17/01/10	2
Пупкин В. И.	323556	ММ-117	Управление данными	25/01/10	3
Петров В. А.	156900	ММ-117	Управление данными	25/01/10	5
Сидоров А. А.	278001	ММ-119	Мат. анализ	21/01/10	5
Киров В. У.	777890	ММ-119	Мат. анализ	21/01/10	4
Хренова Г. П.	123456	ММ-334	Инф. менеджмент	21/01/10	3
Бобриков С. С.	998769	ММ-334	Инф. менеджмент	21/01/10	5
Хренова Г. П.	123456	ММ-334	Базы данных	24/01/10	2
Бобриков С. С.	998769	ММ-334	Базы данных	24/01/10	4

Полная функциональная зависимость

Неключевой атрибут **функционально полно** зависит от **составного ключа**, если он функционально зависит от всего ключа в целом, но не находится в функциональной зависимости от какого-либо из входящих в него атрибутов.

Пример:

Номер зач.кн., Дисциплина, Дата ►
Оценка

Вторая нормальная форма

Отношение (таблица) находится во 2НФ, если оно находится в 1НФ, и каждый неключевой атрибут функционально полно зависит от всего ключа.

Приводить ко 2НФ необходимо только отношения с составным

Вторая нормальная форма

Если какой-либо атрибут зависит от **части составного первичного ключа**, то необходимо:

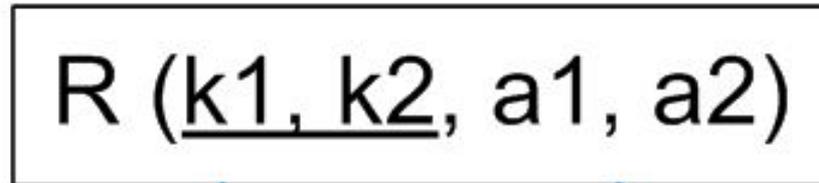
- создать новое отношение, атрибутами которого будут:
 - часть составного ключа (первичный ключ нового отношения)
 - атрибут, зависящий от нового ключа
- из исходного отношения исключить атрибут, включенный в новое отношение.

Вторая нормальная форма

Полная ФЗ: $\underline{k1, k2} \rightarrow a1$

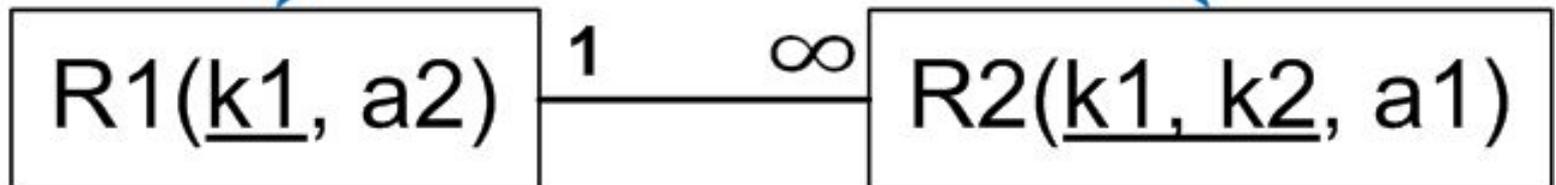
Неполная ФЗ: $\underline{k1} \rightarrow a2$

1 НФ



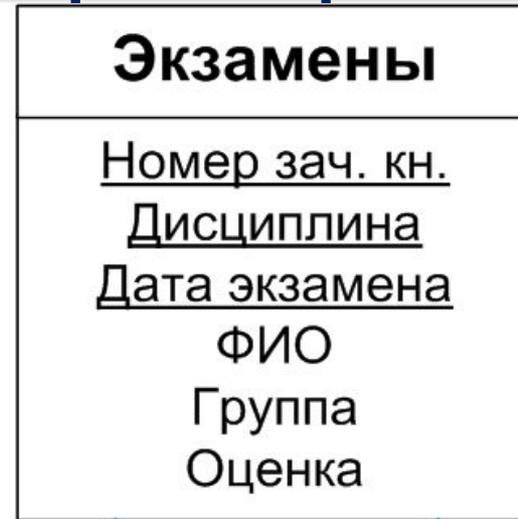
Проекция

2 НФ



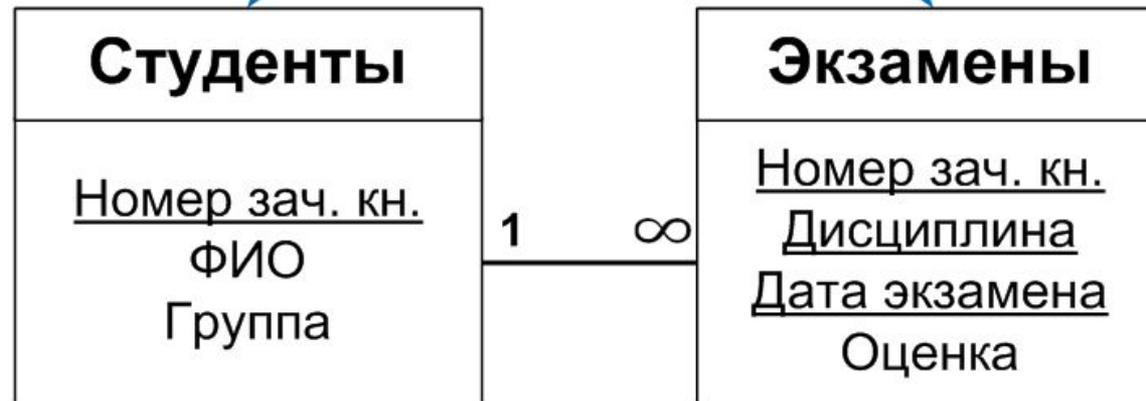
Вторая нормальная форма: пример

1 НФ



Проекция

2 НФ



Определение неполных ФЗ

Составление таблицы-опросника:

КЛ – ключевые атрибуты, НК – неключевые

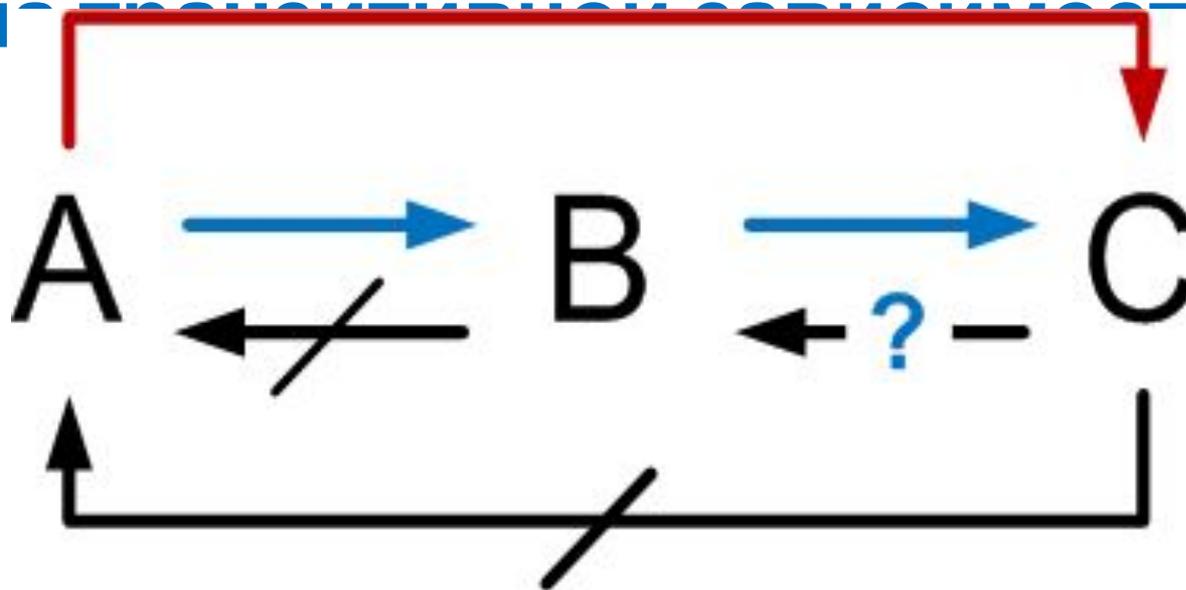
КЛ НК	КЛ1	КЛ2	...	КЛn
НК1	+	+	+	+
НК2		+		+
...	+	+	+	+
НКn	+			

Транзитивная зависимость

Транзитивная функциональная зависимость:

Пусть A, B, C – три атрибута некоторого отношения R.

Схема транзитивной зависимости:



Транзитивная зависимость

Транзитивная зависимость возникает, когда **неключевой атрибут функционально зависит от одного или более неключевых атрибутов.**

Пример: Книга(Ид, Код, Тема, Название, Год)

Атрибут Тема зависит от Код, а не от ключа Ид.

Здесь Код – это Код Темы:

11.21 Физика

12.01 Информатика

Третья нормальная форма

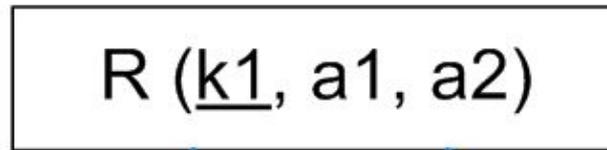
Отношение находится в 3НФ, если оно находится во 2НФ и **каждый неключевой атрибут нетранзитивно** зависит от первичного ключа.

Наличие транзитивной зависимости влечет за собой появление **аномалий обновления**.

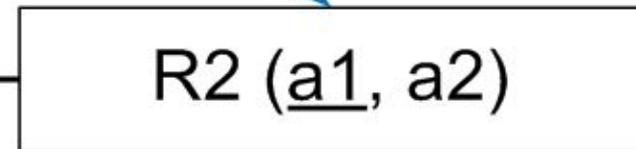
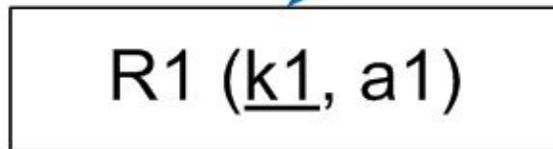
Третья нормальная форма

Транзитивная зависимость:
 $k1 \rightarrow a1$ ($a1 \not\rightarrow k1$)
 $a1 \rightarrow a2$ ($a2 \not\rightarrow a1$)

2 НФ



Проекция



∞

1

3 НФ

Третья нормальная форма: пример

**Транзитивная
зависимость:**

Группа → Специальность

Специальность → Факультет

2 НФ



Проекция

3 НФ



∞

1



Определение транзитивных ФЗ

Составление таблицы-опросника:

НК – неключевые атрибуты

НК	НК1	НК2	...	НКn
НК1		+		
НК2	+			
...				
НКn		+		

Нормальные формы высокого порядка

- На практике редко применяются НФ более высокого порядка, чем 3НФ.
- Самостоятельно изучить НФ высоких порядков:
 - нормальная форма Бойса-Кодда (НФБК)
 - 4 нормальная форма (4НФ)
 - 5 нормальная форма (5НФ)