

Четвёртая нормальная форма

2025-04-08

Рассмотрим таблицу на соответствие 1НФ...

Лекарство	Действующее вещество	Аптека	Цена
Антигриппин	парацетамол, хлорфенамин, аскорбиновая кислота	Алоэ	340
		Озерки	402
Терафлю	парацетамол, фенилэфрин, фениламин	Невис	375
		Лаванда	479
Колдрекс	парацетамол, фенилэфрин, аскорбиновая кислота	Озерки	402
Нурофен	ибупрофен	Алоэ	188
		Озерки	188

Разделим объединённые ячейки...

Лекарство	Действующее вещество	Аптека	Цена
Антигриппин	парацетамол, хлорфенамин, аскорбиновая кислота	Алоэ	340
Антигриппин	парацетамол, хлорфенамин, аскорбиновая кислота	Озерки	402
Терафлю	парацетамол, фенилэфрин, фениламин	Невис	375
Терафлю	паацетамол, фенилэфрин, фениламин	Лаванда	479
Колдрекс	парацетамол, фенилэфрин, аскорбиновая кислота	Озерки	402
Нурофен	ибупрофен	Алоэ	188
Нурофен	ибупрофен	Озерки	188

И затем разнесём в разные строки значения из ячеек, где имеется сразу несколько значений...

1НФ

Потенциальный ключ:
{Лекарство, Вещество, Аптека}

Неключевой атрибут:
Цена

Нарушение 2НФ:
{Лекарство, Аптека} → Цена

Лекарство	Вещество	Аптека	Цена
Антигриппин	парацетамол	Алоэ	340
Антигриппин	хлорфенамин	Алоэ	340
Антигриппин	аскорбиновая кислота	Алоэ	340
Антигриппин	парацетамол	Озерки	402
Антигриппин	хлорфенамин	Озерки	402
Антигриппин	аскорбиновая кислота	Озерки	402
Терафлю	парацетамол	Невис	375
Терафлю	фенилэфрин	Невис	375
Терафлю	фениламин	Невис	375
Терафлю	парацетамол	Лаванда	479
Терафлю	фенилэфрин	Лаванда	479
Терафлю	фениламин	Лаванда	479
Колдрекс	парацетамол	Озерки	402
Колдрекс	фенилэфрин	Озерки	402
Колдрекс	аскорбиновая кислота	Озерки	402
Нурофен	ибупрофен	Алоэ	188
Нурофен	ибупрофен	Озерки	188

Декомпозиция (приведение к 2НФ)

<u>Лекарство</u>	<u>Вещество</u>	<u>Аптека</u>
Антигриппин	парацетамол	Алоэ
Антигриппин	хлорфенамин	Алоэ
Антигриппин	аскорбиновая кислота	Алоэ
Антигриппин	парацетамол	Озерки
Антигриппин	хлорфенамин	Озерки
Антигриппин	аскорбиновая кислота	Озерки
Терафлю	парацетамол	Невис
Терафлю	фенилэфрин	Невис
Терафлю	фениламин	Невис
Терафлю	парацетамол	Лаванда
Терафлю	фенилэфрин	Лаванда
Терафлю	фениламин	Лаванда
Колдрекс	парацетамол	Озерки

<u>Лекарство</u>	<u>Аптека</u>	<u>Цена</u>
Антигриппин	Алоэ	340
Антигриппин	Озерки	402
Терафлю	Невис	375
Терафлю	Лаванда	479
Колдрекс	Озерки	402
Нурофен	Алоэ	188
Нурофен	Озерки	188

Оба отношения находятся в 2НФ
 (в первом отношении только один СК и нет неключевых атрибутов,
 а во втором только один СК и один неключевой атрибут, не зависящий от неполных частей СК).

Другой пример

Фильм	Актёр	Сценарист
День сурка	Билл Мюррей	Денни Рубин
День сурка	Билл Мюррей	Хэрольд Рэмис
День сурка	Энди МакДауэлл	Денни Рубин
День сурка	Энди МакДауэлл	Хэрольд Рэмис
Назад в будущее	Майкл Дж. Фокс	Боб Гейл
Назад в будущее	Майкл Дж. Фокс	Роберт Земекис
Назад в будущее	Кристофер Ллойд	Боб Гейл
Назад в будущее	Кристофер Ллойд	Роберт Земекис
Кто подставил кролика Роджера	Кристофер Ллойд	Роберт Земекис

Потенциальный ключ:
{Фильм, Актёр, Сценарист}.

Отношение находится в НФБК.

Но явно есть лишние дублирования.
И отношение может быть без потерь разделено на два отношения:
{Фильм, Актёр} и
{Фильм, Сценарист}

Многозначные зависимости

- ▶ В 1977 году (через 3 года после описания НФБК Бойсом и Коддом) **Рональд Фейджин** описал четвертую нормальную форму (4НФ), в основе проверки которой лежат не функциональные зависимости, а описанные Фейджином **многозначные зависимости**.
- ▶ Функциональная зависимость (ФЗ) обозначается: $A \rightarrow B$, а **многозначная зависимость (МЗ) обозначается: $A \twoheadrightarrow B$** , где A и B – это некоторые атрибуты или множества атрибутов отношения.
- ▶ В ФЗ каждому значению **A** соответствует ровно одно значение **B**, а в МЗ одному значению **A** может соответствовать несколько значений **B**, но при этом набор этих значений **B** будет всегда один и тот же для этого значения **A**, при любых значениях остальных атрибутов **C** в этом отношении.

Рассмотрим отношение $R(A,B,C)$

A	B	C
A1	B1	C1
A1	B2	C1
A2	B3	C2
A2	B4	C2
A2	B5	C2
A1	B1	C3
A1	B2	C3
A2	B3	C4
A2	B4	C4
A2	B5	C4

Сгруппируем кортежи по значениям атрибута A:

Группа кортежей **A1**

A1	B1	C1
A1	B2	C1
A1	B1	C3
A1	B2	C3

Группа кортежей **A2**

A2	B3	C2
A2	B4	C2
A2	B5	C2
A2	B3	C4
A2	B4	C4
A2	B5	C4

В каждой группе выделим подгруппы с совпадающими значениями C

Группа кортежей **A1**

A1	B1	C1
A1	B2	C1
A1	B1	C3
A1	B2	C3

Подгруппа **A1 C1**

A1	B1	C1
A1	B2	C1

Подгруппа **A1 C3**

A1	B1	C3
A1	B2	C3

Группа кортежей **A2**

A2	B3	C2
A2	B4	C2
A2	B5	C2
A2	B3	C4
A2	B4	C4
A2	B5	C4

Подгруппа **A2 C2**

A2	B3	C2
A2	B4	C2
A2	B5	C2

Подгруппа **A2 C4**

A2	B3	C4
A2	B4	C4
A2	B5	C4

МНОГОЗНАЧНАЯ ЗАВИСИМОСТЬ

Если в получившихся подгруппах, входящих в одну и ту же группу, набор значений атрибута **B** будет всегда в точности совпадать, то имеется многозначная зависимость **A** \twoheadrightarrow **B**

Группа кортежей **A1**

Подгруппа **A1 C1**

A1	B1	C1
A1	B2	C1

Подгруппа **A1 C3**

A1	B1	C3
A1	B2	C3

Группа кортежей **A2**

Подгруппа **A2 C2**

A2	B3	C2
A2	B4	C2
A2	B5	C2

Подгруппа **A2 C4**

A2	B3	C4
A2	B4	C4
A2	B5	C4

Если же возможна ситуация, когда хотя бы в одной подгруппе набор значений **B** будет отличаться от набора значений **B** в другой подгруппе той же группы, тогда **B** не зависит многозначно от **A**.

Определение МЗ

- ▶ Множество атрибутов **В** многозначно зависит от множества атрибутов **А**, если в любой подгруппе кортежей с одинаковыми значениями **А** и **С** (где **С** – множество всех атрибутов отношения кроме **А** и **В**) множество значений **В** будет таким же, как в любой другой подгруппе кортежей с таким же значением **А** и другим значением **С**.
- ▶ В общем случае множество атрибутов **С** может быть пустым, то есть отношение может состоять только из атрибутов **А** и **В** – тогда между **А** и **В** обязательно будет многозначная зависимость.
- ▶ ФЗ является частным случаем МЗ (если в каждой подгруппе с одинаковыми значениями **А** и **С** всегда только один кортеж).

Тривиальные многозначные зависимости

- ▶ Аналогично тому, как ФЗ называется тривиальной, если детерминант содержит в себе зависимую часть, так же и МЗ $A \twoheadrightarrow B$ в отношении $R(A, B, C)$ является тривиальной, если детерминант A содержит в себе B в качестве подмножества.
- ▶ Кроме того, МЗ $A \twoheadrightarrow B$ называют тривиальной, если в отношении нет других атрибутов, кроме A и B (иными словами, если множество C является пустым). Как уже говорилось, если множество C пустое, то между A и B обязательно будет многозначная зависимость, поэтому она является тривиальной.
- ▶ Из этого следует, что если в отношении имеется только два атрибута, то в этом отношении могут быть только тривиальные многозначные зависимости.

Четвёртая нормальная форма

- ▶ Отношение находится в 4НФ, если во всех его нетривиальных **многозначных** зависимостях детерминант является суперключом.
- ▶ Сравним с НФБК: отношение находится в НФБК, если во всех его нетривиальных **функциональных** зависимостях детерминант является суперключом.
- ▶ Поскольку ФЗ является частным случаем МЗ, то, если отношение находится в 4НФ, значит, в нём нет нарушений НФБК.

Если в некотором отношении есть нарушение НФБК, то ФЗ, нарушающая НФБК, является нетривиальной многозначной зависимостью (тривиальной МЗ она могла бы быть только если отношение не имеет атрибутов, кроме атрибутов, входящих в эту ФЗ, но тогда получалось бы, что детерминант этой ФЗ зависит от атрибутов отношения, значит детерминант этой ФЗ был бы тривиальным, следовательно ФЗ не нарушала бы НФБК).

Декомпозиция по правилу Фейджина

- ▶ Если в отношении $R(A, B, C)$ имеется многозначная зависимость $A \twoheadrightarrow B$, то декомпозиция такого отношения на два отношения $R1(A, B)$ и $R2(A, C)$ является декомпозицией без потерь.
- ▶ **Теорема Фейджина** о декомпозиции по МЗ является обобщением **теоремы Хеза** о декомпозиции по ФЗ.
- ▶ Как и при декомпозиции по правилу Хеза, можно ввести суррогатный ключ в момент декомпозиции: если в $R(A, B, C)$ есть $A \twoheadrightarrow B$, то получаем $R1(idA, A, B)$ и $R2(idA, C)$.
- ▶ Если до декомпозиции у нас уже есть отношение, в котором A является первичным ключом, то не нужно создавать новое отношение – просто переносим в него зависимые атрибуты.

Парность многозначных зависимостей

Фильм	Актёр	Сценарист
День сурка	Билл Мюррей	Денни Рубин
День сурка	Билл Мюррей	Хэрольд Рэмис
День сурка	Энди МакДауэлл	Денни Рубин
День сурка	Энди МакДауэлл	Хэрольд Рэмис
Назад в будущее	Майкл Дж. Фокс	Боб Гейл
Назад в будущее	Майкл Дж. Фокс	Роберт Земекис
Назад в будущее	Кристофер Ллойд	Боб Гейл
Назад в будущее	Кристофер Ллойд	Роберт Земекис
Кто подставил кролика Роджера	Кристофер Ллойд	Роберт Земекис

Если в отношении $R(A, B, C)$ имеется многозначная зависимость $A \twoheadrightarrow B$, то в нём обязательно также будет и МЗ $A \twoheadrightarrow C$ (если множество C не является пустым).

Фильм \twoheadrightarrow Актёр

Фильм \twoheadrightarrow Сценарист

После декомпозиции получаем

Фильм	Актёр
День сурка	Билл Мюррей
День сурка	Энди МакДауэлл
Назад в будущее	Майкл Дж. Фокс
Назад в будущее	Кристофер Ллойд
Кто подставил кролика Роджера	Кристофер Ллойд

Фильм	Сценарист
День сурка	Денни Рубин
День сурка	Хэрольд Рэмис
Назад в будущее	Боб Гейл
Назад в будущее	Роберт Земекис
Кто подставил кролика Роджера	Роберт Земекис

Если отношение состоит только из двух атрибутов, то в нём не может быть нарушений 4НФ (так как могут быть только тривиальные МЗ). Таким образом, оба полученных отношения находятся в 4НФ.

Вернёмся к первому примеру...

R1:

<u>Лекарство</u>	<u>Вещество</u>	<u>Аптека</u>
Антигриппин	парацетамол	Алоэ
Антигриппин	хлорфенамин	Алоэ
Антигриппин	аскорбиновая кислота	Алоэ
Антигриппин	парацетамол	Озерки
Антигриппин	хлорфенамин	Озерки
Антигриппин	аскорбиновая кислота	Озерки
Терафлю	парацетамол	Невис
Терафлю	фенилэфрин	Невис
Терафлю	фениламин	Невис
Терафлю	парацетамол	Лаванда
Терафлю	фенилэфрин	Лаванда
Терафлю	фениламин	Лаванда
Колдрекс	парацетамол	Озерки

R2:

<u>Лекарство</u>	<u>Аптека</u>	<u>Цена</u>
Антигриппин	Алоэ	340
Антигриппин	Озерки	402
Терафлю	Невис	375
Терафлю	Лаванда	479
Колдрекс	Озерки	402
Нурофен	Алоэ	188
Нурофен	Озерки	188

В отношении *R1* имеется нарушение 4НФ (МЗ **Лекарство** →→ **Вещество** и **Лекарство** →→ **Аптека**, при этом **Лекарство** не является суперключом). Нужно разделить его на два отношения. Но {Лекарство, Аптека} уже есть в *R2*.

Приведение к 4НФ

R1:

<u>Лекарство</u>	<u>Вещество</u>
Антигриппин	парацетамол
Антигриппин	хлорфенамин
Антигриппин	аскорбиновая кислота
Антигриппин	парацетамол
Антигриппин	хлорфенамин
Антигриппин	аскорбиновая кислота
Терафлю	парацетамол
Терафлю	фенилэфрин
Терафлю	фениламин
Терафлю	парацетамол
Терафлю	фенилэфрин
Терафлю	фениламин
Колдрекс	парацетамол

R2:

<u>Лекарство</u>	<u>Аптека</u>	<u>Цена</u>
Антигриппин	Алоэ	340
Антигриппин	Озерки	402
Терафлю	Невис	375
Терафлю	Лаванда	479
Колдрекс	Озерки	402
Нурофен	Алоэ	188
Нурофен	Озерки	188

В данном случае для приведения к 4НФ следует просто убрать столбец *Аптека* из отношения *R1*. При этом не происходит потери информации, потому что по правилу Фейджина все необходимые атрибуты имеются в отношениях *R1* и *R2*.

Рассмотрение соблюдения 4НФ

- ▶ На практике при проектировании БД методом нормализации обычно не занимаются доказательством соблюдения 4НФ для каждого отношения, так как это долгий процесс.
- ▶ Как правило, нарушения 4НФ в таблице, для которой доказано соблюдение НФБК, сразу заметны «невооружённым глазом», т. к. приводят к большому количеству нежелательных дублирований информации, поэтому нет необходимости искать их путём подробного рассмотрения возможных МЗ.
- ▶ Кроме того, если перед применением метода нормализации был использован другой метод проектирования БД, называемый методом сущностей и связей (ER-метод), то возникновение нарушений 4НФ в такой БД практически невозможно (если правильно применять ER-метод).
- ▶ Поэтому на практике соблюдение 4НФ проверяют только для тех отношений, которые вызывают сомнения.
- ▶ На следующих слайдах для примера показано строгое доказательство соблюдения 4НФ, но на практике такое доказательство обычно не требуется.

Доказательство соблюдения 4НФ

R2:

<u>Лекарство</u>	<u>Аптека</u>	<u>Цена</u>
Антигриппин	Алоэ	340
Антигриппин	Озерки	402
Терафлю	Невис	375
Терафлю	Лаванда	479
Колдрекс	Озерки	402
Нурофен	Алоэ	188
Нурофен	Озерки	188

- Докажем, что в отношении R2 нет нарушений 4НФ.
- Теоретически могли бы быть три пары нетривиальных МЗ: **Лекарство $\rightarrow\rightarrow$ Аптека** (и парная **Лекарство $\rightarrow\rightarrow$ Цена**), **Цена $\rightarrow\rightarrow$ Аптека** (и парная **Цена $\rightarrow\rightarrow$ Лекарство**), **Аптека $\rightarrow\rightarrow$ Лекарство** (и парная **Аптека $\rightarrow\rightarrow$ Цена**).
Докажем, что их на самом деле нет.
- Поскольку нетривиальные МЗ всегда парные, то, доказав отсутствие одной МЗ, мы тем самым доказываем и отсутствие парной ей МЗ. Поэтому достаточно доказать отсутствие трёх перечисленных МЗ.
- Рассмотрим доказательство отсутствия МЗ **Лекарство $\rightarrow\rightarrow$ Аптека**. Для этого приведём пример группы кортежей с одинаковыми значениями атрибута **Лекарство**, и покажем, что в ней могут быть две подгруппы, в каждой из которых **Цена** совпадает, а **Аптека** разная.

Доказательство отсутствия МЗ

Лекарство $\rightarrow\rightarrow$ Аптека

1. Например, отношение теоретически может состоять только из таких 6 кортежей:

<u>Лекарство</u>	<u>Аптека</u>	<u>Цена</u>
Нурофен	Алоэ	188
Нурофен	Озерки	210
Нурофен	Невис	188
Нурофен	Лаванда	210
Антигриппин	Озерки	402
Колдрекс	Озерки	210

3. Как видим, в подгруппах **188** и **210** не совпадают множества значений атрибута **Аптека**. Значит, нет МЗ **Лекарство $\rightarrow\rightarrow$ Аптека**.

2. Разделим группу кортежей со значением «Нурофен» на подгруппы с одинаковыми значениями цен:

Подгруппа **188**:

<u>Лекарство</u>	<u>Аптека</u>	<u>Цена</u>
Нурофен	Алоэ	188
Нурофен	Невис	188

Подгруппа **210**:

<u>Лекарство</u>	<u>Аптека</u>	<u>Цена</u>
Нурофен	Озерки	210
Нурофен	Лаванда	210

Доказательство отсутствия МЗ

Цена $\rightarrow\rightarrow$ Лекарство

Лекарство	Аптека	Цена
Нурофен	Алоэ	188
Нурофен	Озерки	210
Нурофен	Невис	188
Нурофен	Лаванда	210
Антигриппин	Озерки	402
Колдрекс	Озерки	210

- Разделим группу кортежей со значением цены «210» на подгруппы с одинаковыми значениями аптек:

Подгруппа **Озерки**:

Лекарство	Аптека	Цена
Нурофен	Озерки	210
Колдрекс	Озерки	210

Подгруппа **Лаванда**:

Лекарство	Аптека	Цена
Нурофен	Лаванда	210

- Как видим, в подгруппах **Озерки** и **Лаванда** не совпадают множества значений атрибута **Лекарство** (множество {Нурофен, Колдрекс} не равно множеству {Нурофен}).
Значит, нет многозначной зависимости **Цена $\rightarrow\rightarrow$ Лекарство**.
- Важное замечание: мы должны рассматривать все кортежи, входящие в выбранные две подгруппы!

Доказательство отсутствия МЗ

Аптека $\rightarrow\rightarrow$ Лекарство

Лекарство	Аптека	Цена
Нурофен	Алоэ	188
Нурофен	Озерки	210
Нурофен	Невис	188
Нурофен	Лаванда	210
Антигриппин	Озерки	402
Колдрекс	Озерки	210

- Разделим группу кортежей со значением «Озерки» на подгруппы с одинаковыми значениями цен:

Подгруппа **210**:

Лекарство	Аптека	Цена
Нурофен	Озерки	210
Колдрекс	Озерки	210

Подгруппа **402**:

Лекарство	Аптека	Цена
Антигриппин	Озерки	402

- Как видим, в подгруппах **210** и **402** не совпадают множества значений атрибута Лекарство. Значит, нет МЗ **Аптека $\rightarrow\rightarrow$ Лекарство**.
- Таким образом, мы доказали отсутствие МЗ **Лекарство $\rightarrow\rightarrow$ Аптека** (и парной **Лекарство $\rightarrow\rightarrow$ Цена**), **Цена $\rightarrow\rightarrow$ Аптека** (и парной **Цена $\rightarrow\rightarrow$ Лекарство**), **Аптека $\rightarrow\rightarrow$ Лекарство** (и парной **Аптека $\rightarrow\rightarrow$ Цена**).
- Следовательно, в отношении нет нетривиальных МЗ, и оно находится в 4НФ.

Домашнее задание

1. Если в ваших отношениях, приведённых к НФБК, явно присутствуют нарушения 4НФ, то выполнить декомпозицию отношения по правилу Фейджина.
2. Проверить, нет ли у вас в итоге отношений с одинаковыми потенциальными ключами – если такие обнаружатся, то их следует соединить в одно отношение.
3. Для каждого получившегося отношения придумать подходящее **название**.
Название отношениям принято давать во множественном числе, если каждый кортеж отношения хранит информацию о каком-то объекте и его свойствах, а всё отношение хранит информацию о нескольких таких объектах.
Выбирая название, лучше всего отталкиваться от первичного ключа отношения, например, если первичным ключом является фамилия студента (или id студента), то отношение следует назвать «Студенты». Если же первичным ключом является {студент, дисциплина}, а неключевым – оценка студента по этой дисциплине, то, по смыслу, такую таблицу логичнее назвать «Оценки студентов», или просто «Успеваемость» (в единственном числе).