

Основы проектирования баз данных

Нормализация

Поломошнов И.С.

преп. каф. ИБ, ИМФКН, БГУ

Определение

Нормальные формы БД (1НФ, 2НФ, 3НФ и др.) — это набор правил проектирования реляционных баз данных, направленных на устранение дублирования, минимизацию избыточности и предотвращение аномалий при обновлении, удалении или вставке данных.

- Процесс приведения базы данных к этим формам называется нормализацией.

Термины

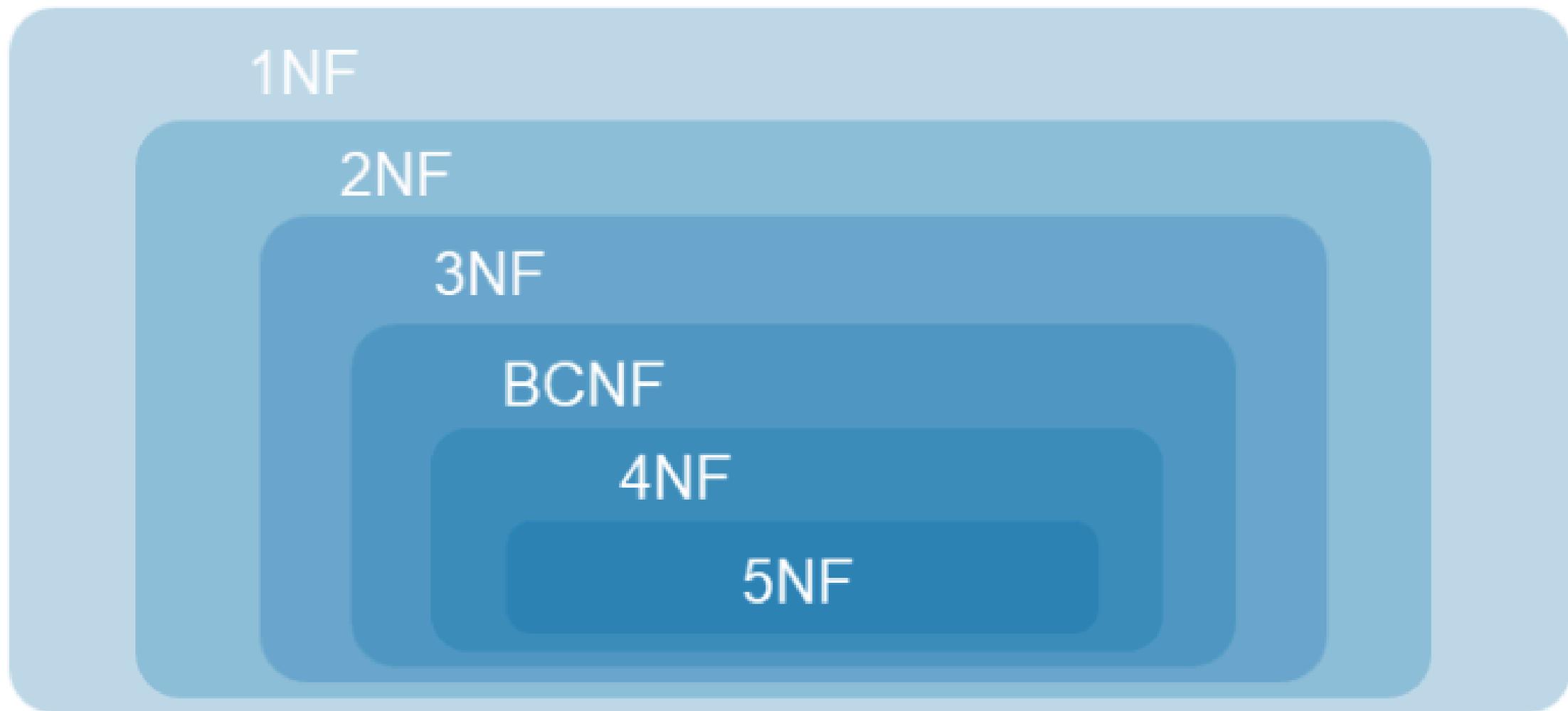
1. ключи - атрибуты столбцов, которые однозначно (уникально) определяют запись в БД;
2. функциональные зависимости - ограничения между двумя взаимосвязанными атрибутами;
3. нормальные формы - этапы для достижения определенного качества БД.

- Процесс проектирования БД с использованием метода НФ является итерационным и заключается в последовательном переводе отношения из 1НФ в НФ более высокого порядка по определенным правилам.
- Каждая следующая НФ ограничивается определенным типом функциональных зависимостей и устранением соответствующих аномалий при выполнении операций над отношениями БД, а также сохранении свойств предшествующих НФ.

Существующие нормальные формы

- Первая нормальная форма (1НФ)
- Вторая нормальная форма (2НФ)
- Третья нормальная форма (3НФ)
- Нормальная форма Бойса-Кодда (BCNF/НФБК)
- Четвертая (4НФ)
- Пятая (5НФ) формы

Normal Forms



Первая нормальная форма (1НФ)

Требует атомарности значений (в ячейке только одно значение, а не список), отсутствия повторяющихся групп и наличия первичного ключа.

- Существует таблица автомобили

Фирма	Модели
BMW	M5, X5M, M1
Nissan	GT-R

Нарушение нормализации 1НФ происходит в моделях BMW, т.к. в одной ячейке содержится список из 3 элементов: M5, X5M, M1, т.е. он не является атомарным.

- Исправленная версия

Фирма	Модели
BMW	M5
BMW	X5M
BMW	M1
Nissan	GT-R

Вторая нормальная форма (2НФ)

Соответствует 1НФ, и все неключевые атрибуты полностью зависят от всего первичного ключа, а не от его части. Устраняет частичные зависимости.

Цена машины зависит от модели и фирмы. Скидка зависит от фирмы, то есть зависимость от первичного ключа неполная.

- Исправляется это путем декомпозиции на два отношения, в которых не ключевые атрибуты зависят от ПК.

<u>Модель</u>	<u>Фирма</u>	Цена	Скидка
M5	BMW	5500000	5%
X5M	BMW	6000000	5%
M1	BMW	2500000	5%
GT-R	Nissan	5000000	10%

<u>Модель</u>	<u>Фирма</u>	Цена
M5	BMW	5500000
X5M	BMW	6000000
M1	BMW	2500000
GT-R	Nissan	5000000

<u>Фирма</u>	Скидка
BMW	5%
Nissan	10%

Третья нормальная форма (3НФ)

Соответствует 2НФ, и отсутствуют транзитивные зависимости (неключевые поля зависят только от ключа, а не от других неключевых полей).

- Таблица находится во 2НФ, но не в 3НФ.
- В отношении атрибут «Модель» является первичным ключом. Личных телефонов у автомобилей нет, и телефон зависит исключительно от магазина.

<u>Модель</u>	Магазин	Телефон
BMW	Риал-авто	87-33-98
Audi	Риал-авто	87-33-98
Nissan	Некст-Авто	94-54-12

1. Таким образом, в отношении существуют следующие функциональные зависимости:

Модель → Магазин, Магазин → Телефон, Модель → Телефон.

2. Зависимость Модель → Телефон является транзитивной, следовательно, отношение не находится в 3НФ.
3. В результате разделения исходного отношения получаются два отношения, находящиеся в 3НФ1

В результате разделения исходного отношения получаются два отношения, находящиеся в ЗНФ

<u>Магазин</u>	Телефон
Риал-авто	87-33-98
Некст-Авто	94-54-12

<u>Модель</u>	Магазин
BMW	Риал-авто
Audi	Риал-авто
Nissan	Некст-Авто

Нормальная форма Бойса-Кодда (BCNF/НФБК)

Более строгая версия ЗНФ, где каждый детерминант является потенциальным ключом.

Определение ЗНФ не совсем подходит для следующих отношений:

1. отношение имеет два или более потенциальных ключа;
 2. два и более потенциальных ключа являются составными;
 3. они пересекаются, т.е. имеют хотя бы один общий атрибут.
- Для отношений, имеющих один потенциальный ключ (первичный), НФБК является ЗНФ.
 - Отношение находится в НФБК, когда каждая нетривиальная и неприводимая слева функциональная зависимость обладает потенциальным ключом в качестве детерминанта.

Тариф имеет уникальное название и зависит от выбранной стоянки и наличия льгот, в частности:

1. «Бережливый»
2. «Стандарт»
3. «Премиум-А»
4. «Премиум-В»

Номер стоянки	Время начала	Время окончания	Тариф
1	09:30	10:30	Бережливый
1	11:00	12:00	Бережливый
1	14:00	15:30	Стандарт
2	10:00	12:00	Премиум-В
2	12:00	14:00	Премиум-В
2	15:00	18:00	Премиум-А

- Таким образом, возможны следующие составные первичные ключи:

{Номер стоянки, Время начала}, {Номер стоянки, Время окончания}, {Тариф, Время начала}, {Тариф, Время окончания}.

- Отношение находится в 3НФ. Требования второй нормальной формы выполняются, так как все атрибуты входят в какой-то из потенциальных ключей, а неключевых атрибутов в отношении нет.

Можно улучшить структуру с помощью декомпозиции отношения на два, получив отношения, удовлетворяющие НФБК

<u>Тариф</u>	Номер стоянки	Имеет льготы
Бережливый	1	Да
Стандарт	1	Нет
Премиум-А	2	Да
Премиум-В	2	Нет

<u>Тариф</u>	<u>Время начала</u>	Время окончания
Бережливый	09:30	10:30
Бережливый	11:00	12:00
Стандарт	14:00	15:30
Премиум-В	10:00	12:00
Премиум-В	12:00	14:00
Премиум-А	15:00	18:00

Четвертая (4НФ) и Пятая (5НФ) формы

Устраняют многозначные и JOIN-зависимости, используются реже, в специализированных случаях.

Четвертая нормальная форма

Отношение находится в 4НФ, если оно находится в НФБК и все нетривиальные многозначные зависимости фактически являются функциональными зависимостями от ее потенциальных ключей.

- В отношении $R(A, B, C)$ существует многозначная зависимость $R.A \twoheadrightarrow R.V$ в том и только в том случае, если множество значений B , соответствующее паре значений A и C , зависит только от A и не зависит от C .

Предположим, что рестораны производят разные виды пиццы, а службы доставки ресторанов работают только в определенных районах города.

- Составной первичный ключ соответствующей переменной отношения включает три атрибута: {Ресторан, Вид пиццы, Район доставки}.

Такая переменная отношения не соответствует 4НФ, так как существует следующая многозначная зависимость:
{Ресторан} → {Вид пиццы}
{Ресторан} → {Район доставки}

- То есть, например, при добавлении нового вида пиццы придется внести по одному новому кортежу для каждого района доставки.
- Возможна логическая аномалия, при которой определенному виду пиццы будут соответствовать лишь некоторые районы доставки из обслуживаемых рестораном районов.

Пятая нормальная форма

Отношения находятся в 5НФ, если оно находится в 4НФ и отсутствуют сложные зависимые соединения между атрибутами.

Если «Атрибут_1» зависит от «Атрибута_2», а «Атрибут_2» в свою очередь зависит от «Атрибута_3», а «Атрибут_3» зависит от «Атрибута_1», то все три атрибута обязательно входят в один кортеж.

- Это очень жесткое требование, которое можно выполнить лишь при дополнительных условиях.

Некоторые существующие НФ

- **Доменно-ключевая нормальная форма**
- - Ограничение ключа – ограничение, утверждающее, что некоторый атрибут или комбинация атрибутов является потенциальным ключом.
- **Шестая нормальная форма**
- - Переменная отношения находится в шестой нормальной форме тогда и только тогда, когда она удовлетворяет всем нетривиальным зависимостям соединения.

Пример 6 НФ

Работники

<u>Таб.№</u>	<u>Время</u>	Должность	Домашний адрес
6575	01-01-2000:10-02-2003	слесарь	ул.Ленина,10
6575	11-02-2003:15-06-2006	слесарь	ул.Советская,22
6575	16-06-2006:05-03-2009	бригадир	ул.Советская,22

- Переменная отношения «Работники» не находится в 6НФ и может быть подвергнута декомпозиции на переменные отношения «Должности работников» и «Домашние адреса работников».

Должности работников

<u>Таб.№</u>	<u>Время</u>	<u>Должность</u>
6575	01-01-2000:10-02-2003	слесарь
6575	16-06-2006:05-03-2009	бригадир

Домашние адреса работников

<u>Таб.№</u>	<u>Время</u>	<u>Домашний адрес</u>
6575	01-01-2000:10-02-2003	ул.Ленина,10
6575	11-02-2003:15-06-2006	ул.Советская,22

Источники

- Нормализация отношений БД <https://habr.com/ru/articles/254773/>
- Как привести данные в форму: что такое нормализация и зачем она нужна <https://practicum.yandex.ru/blog/что-такое-normalizaciya-dannyh/>
- Что такое нормализация базы данных?
<https://wiki.merionet.ru/articles/что-такое-normalizaciya-bazy-dannyh>

Спасибо за внимание

Вопросы?