

Нормализация и нормальные
формы. Связи в базе данных.

Определение

- **Нормализация** — процесс преобразования отношения (таблицы) базы данных к виду, отвечающему нормальным формам;
- **Нормальная форма** — свойство отношения (таблицы) в реляционной модели данных, характеризующее его с точки зрения избыточности, потенциально приводящей к логически ошибочным результатам выборки или изменения данных.
- Нормальная форма определяется как совокупность требований, которым должно удовлетворять отношение (таблица). Выделяют 8 нормальных форм.

Нормальные формы

Цель нормализации

- **Нормализация** предназначена для приведения структуры БД к виду, обеспечивающему **минимальную логическую избыточность**, и не имеет целью увеличение производительности работы или же уменьшение физического объёма базы данных.
- Конечной целью нормализации является **уменьшение потенциальной противоречивости** хранимой в базе данных информации.
- Устранение избыточности производится, как правило, за счёт декомпозиции отношений (таблиц) таким образом, чтобы в каждом отношении (таблице) хранились только первичные факты (то есть факты, не выводимые из других хранимых фактов).
- Нормализация **НЕ ВСЕГДА** полезна, особенно в Web приложениях, где на первую очень выходят требования производительности запросов и масштабируемости базы данных. Для таких систем получили широкое распространение NoSQL решения.

Первая нормальная форма

- Переменная отношения (таблица) находится в **первой нормальной форме** тогда и только тогда, когда в любом допустимом значении отношения каждый его кортеж содержит только одно значение для каждого из атрибутов:
- Исходная ненормализованная (то есть не являющаяся правильным представлением некоторого отношения) таблица:

Фамилия	Имя	Номер
Давыдов	Давид	+38 (067) 122-11-69
Иванов	Иван	+38 (050) 123-12-78
Петров	Петр	+38 (050) 123-12-34
		+38 (063) 345-67-98
Сергеев	Сергей	+38 (050) 123-12-45

- Таблица, приведённая к 1НФ, являющаяся правильным представлением некоторого отношения:

Фамилия	Имя	Номер
Давыдов	Давид	+38 (067) 122-11-69
Иванов	Иван	+38 (050) 123-12-78
Петров	Петр	+38 (050) 123-12-34
Петров	Петр	+38 (063) 345-67-98
Сергеев	Сергей	+38 (050) 123-12-45

Вторая нормальная форма

- Переменная отношения (таблица) находится во **второй нормальной форме** тогда и только тогда, когда она находится в первой нормальной форме и каждый неключевой атрибут полностью зависит от её потенциального ключа (всего ключа, а не части):
- **Потенциальный ключ** — в реляционной модели данных — подмножество атрибутов отношения, удовлетворяющее требованиям **уникальности** и **минимальности (несократимости)**:
- **Уникальность** означает, что нет и не может быть двух кортежей данного отношения, в которых значения этого подмножества атрибутов совпадают (равны).
- **Минимальность (несократимость)** означает, что в составе потенциального ключа отсутствует меньшее подмножество атрибутов, удовлетворяющее условию уникальности. Иными словами, если из потенциального ключа убрать любой атрибут, он утратит свойство уникальности.

Вторая нормальная форма

Категория	Название фотографии	Дата путешествия	Дата фотографии	Размер файла	Фотография
Свадебное путешествие	Поход в ресторан	10/7/2015 4:51:44 PM	10/7/2015 4:51:44 PM	3Мб	
Свадебное путешествие	Красивый закат	10/7/2015 4:51:44 PM	10/8/2015 4:51:44 PM	4Мб	
Свадебное путешествие	Прозрачная водичка	10/7/2015 4:51:44 PM	10/9/2015 4:51:44 PM	3,4Мб	
Отпуск в Египте	Верхом на верблюде	10/25/2016 4:51:44 PM	10/26/2016 4:51:44 PM	5Мб	
Отпуск в Египте	Путешествие на квадрациклах	10/25/2016 4:51:44 PM	10/28/2016 4:51:44 PM	4Мб	
Отпуск в Египте	Подводный мир	10/25/2016 4:51:44 PM	10/25/2016 4:51:44 PM	5Мб	
Отпуск в Египте	Поход в ресторан	10/25/2016 4:51:44 PM	10/25/2016 4:53:10 PM	8Мб	
Отпуск в Египте	Пирамиды	10/25/2016 4:51:44 PM	10/25/2016 4:51:44 PM	4,5Мб	

- В качестве потенциального ключа выбираем пару **<Категория, Название фотографии>**, тогда **Дата фотографии, Размер файла и Фотография** зависят от Категории и Названия фотографии, а вот **Дата путешествия** зависит только от Категории.
- Существует функциональная зависимость *Категория* → *Дата путешествия*, в которой левая часть является лишь частью первичного ключа, что нарушает условие полной зависимости и следовательно данное отношение (таблица) находится не во второй нормальной форме. (Пример логической ошибки данных: некорректная запись при некорректной Дате путешествия)

Категория	Название фотографии	Дата фотографии	Размер файла	Фотография
Свадебное путешествие	Поход в ресторан	10/7/2015 4:51:44 PM	3Мб	
Отпуск в Египте	Подводный мир	10/25/2016 4:51:44 PM	5Мб	
Отпуск в Египте	Пирамиды	10/25/2016 4:51:44 PM	4,5Мб	
Свадебное путешествие	Красивый закат	10/8/2015 4:51:44 PM	4Мб	
Свадебное путешествие	Прозрачная водичка	10/9/2015 4:51:44 PM	3,4Мб	
Отпуск в Египте	Верхом на верблюде	10/26/2016 4:51:44 PM	5Мб	
Отпуск в Египте	Путешествие на квадрациклах	10/28/2016 4:51:44 PM	4Мб	
Отпуск в Египте	Поход в ресторан	10/25/2016 4:53:10 PM	8Мб	

Категория	Дата путешествия
Свадебное путешествие	10/7/2015 4:51:44 PM
Отпуск в Египте	10/25/2016 4:51:44 PM

Понятие первичного ключа

- Первичный ключ (Primary key) — один из потенциальных ключей отношения (таблицы), выбранный в качестве основного ключа;
- Критерии выбора первичного ключа:
 1. Сохранение уникальности со временем;
 2. Имеет наименьший размер хранения данных;
 3. Состоит из меньшего числа атрибутов.
- Первичный ключ может быть **простым** и **составным**;
- Первичный ключ может быть **естественным** и **суррогатным**.
- Преимущества суррогатных ключей: **Неизменность** (ключ никогда не меняется), **Гарантированная уникальность**, **Гибкость** (из-за неинформативности ключ может быть заменен в любое время), **Эффективность** (запросы с суррогатными ключами более компактные).

Использование суррогатного ключа

ID	Категория	Дата путешествия
1	Свадебное путешествие	10/7/2015 4:51:44 PM
2	Отпуск в Египте	10/25/2016 4:51:44 PM

ID	ID_Категории	Название фотографии	Дата фотографии	Размер файла	Фотография
2	2	Подводный мир	10/25/2016 4:51:44 PM	5M6	
3	2	Пирамиды	10/25/2016 4:51:44 PM	4,5M6	
6	2	Верхом на верблюде	10/26/2016 4:51:44 PM	5M6	
7	2	Путешествие на квадрациклах	10/28/2016 4:51:44 PM	4M6	
8	2	Поход в ресторан	10/25/2016 4:53:10 PM	8M6	
1	1	Поход в ресторан	10/7/2015 4:51:44 PM	3M6	
4	1	Красивый закат	10/8/2015 4:51:44 PM	4M6	
5	1	Прозрачная водичка	10/9/2015 4:51:44 PM	3,4M6	

Понятие внешнего ключа

- **Внешний ключ** (англ. foreign key) — подмножество атрибутов некоторого отношения (таблицы), значения которых должны совпадать со значениями первичного ключа второго отношения (таблицы), для указания информационной целостности всей базы данных.
- Свойства внешнего ключа: RESTRICT, CASCADE, SET NULL, SET DEFAULT

Категория	Дата путешествия
Свадебное путешествие	10/7/2015 4:51:44 PM
Отпуск в Египте	10/25/2016 4:51:44 PM

Категория	Название фотографии	Дата фотографии	Размер файла	Фотография
Свадебное путешествие	Поход в ресторан	10/7/2015 4:51:44 PM	3Мб	
Отпуск в Египте	Подводный мир	10/25/2016 4:51:44 PM	5Мб	
Отпуск в Египте	Пирамиды	10/25/2016 4:51:44 PM	4,5Мб	
Свадебное путешествие	Красивый закат	10/8/2015 4:51:44 PM	4Мб	
Свадебное путешествие	Прозрачная водичка	10/9/2015 4:51:44 PM	3,4Мб	
Отпуск в Египте	Верхом на верблюде	10/26/2016 4:51:44 PM	5Мб	
Отпуск в Египте	Путешествие на квадрациклах	10/28/2016 4:51:44 PM	4Мб	
Отпуск в Египте	Поход в ресторан	10/25/2016 4:53:10 PM	8Мб	

ID	ID_Категории	Название фотографии	Дата фотографии	Размер файла	Фотография
2	2	Подводный мир	10/25/2016 4:51:44 PM	5Мб	
3	2	Пирамиды	10/25/2016 4:51:44 PM	4,5Мб	
6	2	Верхом на верблюде	10/26/2016 4:51:44 PM	5Мб	
7	2	Путешествие на квадрациклах	10/28/2016 4:51:44 PM	4Мб	
8	2	Поход в ресторан	10/25/2016 4:53:10 PM	8Мб	
1	1	Поход в ресторан	10/7/2015 4:51:44 PM	3Мб	
4	1	Красивый закат	10/8/2015 4:51:44 PM	4Мб	
5	1	Прозрачная водичка	10/9/2015 4:51:44 PM	3,4Мб	

ID	Категория	Дата путешествия
1	Свадебное путешествие	10/7/2015 4:51:44 PM
2	Отпуск в Египте	10/25/2016 4:51:44 PM

Третья нормальная форма

- Переменная отношения (таблица) находится в **третьей нормальной форме** тогда и только тогда, когда она находится во второй нормальной форме, и отсутствуют транзитивные функциональные зависимости неключевых атрибутов от ключевых. Использование суррогатных ключей решает проблему второй нормальной формы:

ID	Категория	Название фотографии	Дата путешествия	Дата фотографии	Размер файла	Фотография
1	Свадебное путешествие	Поход в ресторан	10/7/2015 4:51:44 PM	10/7/2015 4:51:44 PM	3M6	
2	Свадебное путешествие	Красивый закат	10/7/2015 4:51:44 PM	10/8/2015 4:51:44 PM	4M6	
3	Свадебное путешествие	Прозрачная водичка	10/7/2015 4:51:44 PM	10/9/2015 4:51:44 PM	3,4M6	
4	Отпуск в Египте	Подводный мир	10/25/2016 4:51:44 PM	10/25/2016 4:51:44 PM	5M6	
5	Отпуск в Египте	Пирамиды	10/25/2016 4:51:44 PM	10/25/2016 4:51:44 PM	4,5M6	
6	Отпуск в Египте	Верхом на верблюде	10/25/2016 4:51:44 PM	10/26/2016 4:51:44 PM	5M6	
7	Отпуск в Египте	Путешествие на квадраци	10/25/2016 4:51:44 PM	10/28/2016 4:51:44 PM	4M6	
8	Отпуск в Египте	Поход в ресторан	10/25/2016 4:51:44 PM	10/25/2016 4:53:10 PM	8M6	

В примере существуют следующие функциональные зависимости: *Дата путешествия* → *Категория*, *Категория* → *ID*, *Дата путешествия* → *ID*, следовательно отношения (таблицы) находятся не в третьей нормальной форме. Для простоты можно использовать принцип оценки принадлежности атрибута к данному отношению.

Третья нормальная форма

ID	ID_Категории	Название фотографии	Дата фотографии	Размер файла	Фотография
2	2	Подводный мир	10/25/2016 4:51:44 PM	5M6	
3	2	Пирамиды	10/25/2016 4:51:44 PM	4,5M6	
6	2	Верхом на верблюде	10/26/2016 4:51:44 PM	5M6	
7	2	Путешествие на квадрациклах	10/28/2016 4:51:44 PM	4M6	
8	2	Поход в ресторан	10/25/2016 4:53:10 PM	8M6	
1	1	Поход в ресторан	10/7/2015 4:51:44 PM	3M6	
4	1	Красивый закат	10/8/2015 4:51:44 PM	4M6	
5	1	Прозрачная водичка	10/9/2015 4:51:44 PM	3,4M6	

ID	Категория	Дата путешествия
1	Свадебное путешествие	10/7/2015 4:51:44 PM
2	Отпуск в Египте	10/25/2016 4:51:44 PM

Благодаря третьей нормальной форме исключаются ошибки неоднозначного удаления данных: (Хотим удалить только фотки или всю категорию?)

ID	Категория	Название фотографии	Дата путешествия	Дата фотографии	Размер файла	Фотография
1	Свадебное путешествие	Поход в ресторан	10/7/2015 4:51:44 PM	10/7/2015 4:51:44 PM	3M6	
2	Свадебное путешествие	Красивый закат	10/7/2015 4:51:44 PM	10/8/2015 4:51:44 PM	4M6	
3	Свадебное путешествие	Прозрачная водичка	10/7/2015 4:51:44 PM	10/9/2015 4:51:44 PM	3,4M6	
4	Отпуск в Египте	Подводный мир	10/25/2016 4:51:44 PM	10/25/2016 4:51:44 PM	5M6	
5	Отпуск в Египте	Пирамиды	10/25/2016 4:51:44 PM	10/25/2016 4:51:44 PM	4,5M6	
6	Отпуск в Египте	Верхом на верблюде	10/25/2016 4:51:44 PM	10/26/2016 4:51:44 PM	5M6	
7	Отпуск в Египте	Путешествие на квадраци	10/25/2016 4:51:44 PM	10/28/2016 4:51:44 PM	4M6	
8	Отпуск в Египте	Поход в ресторан	10/25/2016 4:51:44 PM	10/25/2016 4:53:10 PM	8M6	

Остальные нормальные формы

Нормальная форма Бойса — Кодда (BCNF)

Переменная отношения находится в нормальной форме Бойса — Кодда (иначе — в усиленной третьей нормальной форме) тогда и только тогда, когда каждая её нетривиальная и неприводимая слева функциональная зависимость имеет в качестве своего детерминанта некоторый потенциальный ключ.

Четвёртая нормальная форма (4NF)

Переменная отношения находится в четвёртой нормальной форме, если она находится в нормальной форме Бойса — Кодда и не содержит нетривиальных многозначных зависимостей.

Пятая нормальная форма (5NF)

Переменная отношения находится в пятой нормальной форме (иначе — в проекционно-соединительной нормальной форме) тогда и только тогда, когда каждая нетривиальная зависимость соединения в ней определяется потенциальным ключом (ключами) этого отношения.

Доменно-ключевая нормальная форма (DKNF)

Переменная отношения находится в ДКНФ тогда и только тогда, когда каждое наложенное на неё ограничение является логическим следствием ограничений доменов и ограничений ключей, наложенных на данную переменную отношения.

Шестая нормальная форма (6NF)

Переменная отношения находится в шестой нормальной форме тогда и только тогда, когда она удовлетворяет всем нетривиальным зависимостям соединения.

Нормальные формы

Связи между таблицами

Связи между таблицами (таблицей) базы данных:

- Один к одному;
- Один ко многим (Многие к одному);
- Один ко многим (Многие к одному) в одной таблице;
- Многие ко многим.

Один к одному

ID	Фамилия	Номер телефона
1	Иванов	+38(050) 123-45-67
2	Петров	+38(099) 677-34-00
3	Сергеев	+38(050) 987-65-23
4	Николаев	+38(050) 987-65-23
5	Васильев	+38(066) 987-11-23
9	Васильев	+38(050) 123-45-67

Таблица А

email
1 ivanov@gmail.com
2 petrov@gmail.com
3 serveev@gmail.com

Таблица В



Одна запись в таблице А может ссылаться только на одну запись из таблицы В, поэтому связь называется один к одному.

Р.С. Одна запись из таблицы А может ссылаться на ноль записей из таблицы В

Связь один к одному реализуется когда первичный ключ таблицы А становится внешним ключом к уникальному полю (в частности первичному ключу) таблицы В

Один ко многим (Многие к одному)

ID	ID_Категории	Название фотографии
2	2	Подводный мир
3	2	Пирамиды
6	2	Верхом на верблюде
7	2	Путешествие на квадрациклах
8	2	Поход в ресторан
1	1	Поход в ресторан
4	1	Красивый закат
5	1	Прозрачная водичка

Таблица А

ID	Категория	Дата путешествия
1	Свадебное путешествие	10/7/2015 4:51:44 PM
2	Отпуск в Египте	10/25/2016 4:51:44 PM

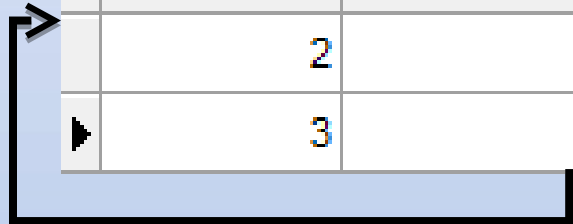
Таблица В



Несколько записей (ноль и больше) в таблице А могут ссылаться только на одну запись из таблицы В, поэтому связь называется многие к одному. Если рассмотреть связь с обратной стороны то будет один ко многим.

Связь один к многим (многие к одному) реализуется когда поле таблицы А, которое не является уникальным полем становится внешним ключом к первичному ключу таблицы В

Многие к одному в одной таблице



id	id_parent	folder_name
1	Null	Корневая директория
2	1	Поддиректория 1
3	1	Поддиректория 2

Таблица А

Несколько записей (ноль и больше) в таблице А могут ссылаться только на одну запись из этой же таблице А, поэтому связь называется многие к одному. Если рассмотреть связь с обратной стороны то будет один ко многим

Связь один к многим в одной таблице реализуется когда поле таблицы А, которое не является уникальным становится внешним ключом к первичному ключу этой же таблице А.

Многие ко многим

Таблица А

id	Фамилия
2	Петров
1	Иванов

Таблица В

id	Номер телефона
2	234-89-67
1	789-45-09

id_контакта	id_номера_телефона
1	2
1	1
2	2
2	1

Таблица АВ

Одна запись из таблицы А может ссылаться на несколько (ноль и больше) записей из таблицы В и одна запись из таблицы В может ссылаться на несколько (ноль и больше) записей из таблицы А.

Связь многие ко многим реализуется через вспомогательную таблицу как комбинация двух связей одни ко многим.

Выводы

1. **Нормализация** — процесс преобразования отношения (таблицы) базы данных к виду, отвечающему нормальным формам;
2. **Нормальная форма** — свойство отношения (таблицы) в реляционной модели данных, характеризующее его с точки зрения избыточности, потенциально приводящей к логически ошибочным результатам выборки или изменения данных;
3. Конечной целью нормализации является **уменьшение потенциальной противоречивости и избыточности** хранимой в базе данных информации;
4. Первичный ключ (Primary key) — один из потенциальных ключей отношения (таблицы), выбранный в качестве основного ключа. Первичный ключ может быть **простым и составным; естественным и суррогатным**.
5. **Внешний ключ** (англ. foreign key) — подмножество атрибутов некоторого отношения (таблицы), значения которых должны совпадать со значениями первичного ключа второго отношения (таблицы), для указания информационной целостности всей базы данных;
6. Использование третьей нормальной формы обычно достаточно во многих приложениях;
7. Свойства внешнего ключа: RESTRICT, CASCADE, SET NULL, SET DEFAULT;
8. Связи между таблицами (таблицей) базы данных могут быть: один к одному, один ко многим (многие к одному), один ко многим в одной таблице, многие ко многим;
9. Связь один к одному реализуется когда уникальное поле таблицы A становится внешним ключом к первичному ключу таблицы B;
10. Связь один к многим (многие к одному) реализуется когда поле таблицы A, которое не является уникальным становится внешним ключом к первичному ключу таблицы B;
11. Связь многие ко многим реализуется через вспомогательную таблицу как комбинация двух связей один ко многим.