

# Нормальные формы

- **Нормализация** – это процесс удаления избыточных данных.
- **Нормализация** – это метод проектирования базы данных, который позволяет привести базу данных к минимальной избыточности.

Цели:

- Устранение аномалий
- Повышение производительности
- Повышение удобства управления данными
- **Избыточность данных** – это когда одни и те же данные хранятся в базе в нескольких местах, именно это и приводит к аномалиям.

- **Нормальная форма базы данных** – это набор правил и критериев, которым должна отвечать база данных.
- **Процесс нормализации** – это последовательный процесс приведения базы данных к эталонному виду, т.е. переход от одной нормальной формы к следующей.

- Процесс проектирования БД с использованием метода нормальных форм заключается в последовательном переводе отношений из первой нормальной формы в нормальные формы более высокого порядка по определенным правилам. Каждая следующая нормальная форма ограничивает определенный тип функциональных зависимостей, устраняет соответствующие аномалии при выполнении операций над отношениями БД и сохраняет свойства предшествующих нормальных форм.

***Нормализация таблиц базы данных*** - первый шаг на пути проектирования структуры реляционной базы данных.

База данных считается нормализованной, если ее таблицы представлены как минимум в третьей нормальной форме. Часто многие таблицы нормализуются до четвертой нормальной формы, иногда, наоборот, производится денормализация.

***Главная цель нормализации базы данных*** - устранение избыточности и дублирования информации. В идеале при нормализации надо добиться, чтобы любое значение хранилось в базе в одном экземпляре, причем значение это не должно быть получено расчетным путем из других данных, хранящихся в базе.

Выделяют следующую последовательность нормальных форм:

- Первая нормальная форма (1НФ)
- Вторая нормальная форма (2НФ)
- Третья нормальная форма (3НФ)
- Усиленная нормальная форма, или нормальная форма Бойса-Кодда (БКНФ)
- Четвертая нормальная форма (4НФ)
- Пятая нормальная форма (5НФ).

По реляционной теории строки в таблицах не должны быть пронумерованы, т.е. порядок строк не имеет значения, так же как не имеет значения порядок столбцов. Т.е. например, если мы поменяем порядок столбцов, или порядок строк, ничего измениться не должно, это не должно ни на что повлиять. Таким образом по реляционной теории мы не можем обратиться к определённой строке или столбцу по ее номеру.

# Нулевая нормальная форма (UNF)

Порядковый номер строки	A	B
1	Иван	Иванов
2	Сергей	Сергеев
3	John	Smith
4	Иван	Иванов

Однако, к сожалению, подобные таблицы нельзя назвать реляционными, так как это пронумерованные двумерные массивы данных. И если мы поменяем местами строки, то наша нумерация просто нарушится.



❖ Каждой нормальной форме соответствует некоторый определенный набор ограничений, и отношение находится в некоторой нормальной форме, если удовлетворяет свойственному ей набору ограничений. Примером набора ограничений является ограничение ***первой нормальной формы*** - значения всех атрибутов отношения атомарны. Поскольку требование ***первой нормальной формы*** является базовым требованием классической реляционной модели данных, мы будем считать, что исходный набор отношений уже соответствует этому требованию.

# ***Первая нормальная форма.***

Отношение находится в 1НФ, если все его атрибуты являются простыми (имеют единственное значение). Исходное отношение строится таким образом, что было в 1НФ.

Перевод отношения в следующую нормальную форму осуществляется методом «декомпозиции без потерь». Такая декомпозиция должна обеспечивать то, что запросы к исходному отношению и к отношениям, получаемым в результате декомпозиции, дадут одинаковый результат.

# Требования первой нормальной формы (1NF)

Чтобы база данных находилась в 1 нормальной форме, необходимо чтобы ее таблицы соблюдали следующие реляционные принципы:

- В таблице не должно быть дублирующих строк
- В каждой ячейке таблицы хранится атомарное значение (одно не составное значение)
- В столбце хранятся данные одного типа
- Отсутствуют массивы и списки в любом виде

# Требования первой нормальной формы (1NF)

Строки, столбцы и ячейки в таблицах необходимо использовать строго по назначению.

- Назначение строк – хранить данные
- Назначение столбцов – хранить структурную информацию
- Назначение ячеек – хранить атомарное значение

# Первая нормальная форма:



Таб. №	ФИО	Должность	№ кабинета	Телефон	Дети
1	Иванов	Директор	32	212	Саша(1996) Маша(1995)
2	Зайцева	Бухгалтер	33	213	Петя(1992) Витя(1997) Лена(1999)
3	Волков	Менеджер	35	311	-

**Дети не являются элементарными данными, следовательно, таблицу нельзя вводить таким образом, её необходимо преобразовать**

# Первая нормальная форма

Таб. № РК	ФИО	Должность	№ кабинета	Телефон	Имя ребёнка	Год рождения	№ ребёнка РК
1	Иванов	Директор	32	212	Саша	1996	1
2	Зайцева	Бухгалтер	33	213	Петя	1992	1
3	Волков	Менеджер	35	311	-	-	-
1	Иванов	Директор	32	212	Маша	1995	2
2	Зайцева	Бухгалтер	33	213	Витя	1997	2
2	Зайцева	Бухгалтер	33	213	Лена	1999	3

Теперь все ячейки таблицы являются элементарными неделимыми (атомарными) данными и таблица может быть помещена в БД

## Аномалия вставки:

Таб. № РК	ФИО	Должность	№ кабинета	Телефон	Имя ребёнка	Год рождения	№ ребёнка РК
1	Иванов	Директор	32	212	Саша	1996	1
2	Зайцева	Бухгалтер	33	213	Петя	1992	1
3	Волков	Менеджер	35	311	-	-	-
1	Иванов	Директор	32	212	Маша	1995	2
2	Зайцева	Бухгалтер	33	213	Витя	1997	2
2	Зайцева	Бухгалтер	33	213	Лена	1999	3

Волков не может быть внесён в базу данных, т.к. у него нет детей, а эта информация входит в состав ключа

# Аномалия модификации

Таб. № РК	ФИО	Должность	№ кабинета	Телефон	Имя ребёнка	Год рождения	№ ребёнка РК
1	Иванов	Директор	32	212	Саша	1996	1
2	Петрова	Бухгалтер	33	213	Петя	1992	1
1	Иванов	Директор	32	212	Маша	1995	2
2	Зайцева	Бухгалтер	33	213	Витя	1997	2
2	Зайцева	Бухгалтер	33	213	Лена	1999	3

Зайцева сменила фамилию. Если не принять специальных мер, изменение базы может быть произведено не полностью, и в базе данных окажется противоречивая информация.



# Аномалия удаления

Таб. № РК	ФИО	Должность	№ кабинета	Телефон	Имя ребёнка	Год рождения	№ ребёнка РК
2	Зайцева	Бухгалтер	33	213	Витя	1997	2
2	Зайцева	Бухгалтер	33	213	Лена	1999	3

Дети, достигшие совершеннолетия, удаляются из БД. Однако, вместе с удалением Саши, ребенка Иванова, мы удалили и информацию о самом Иванове.

# ***Вторая нормальная форма***

- Отношение находится в 2НФ, если оно находится в 1НФ, и каждый неключевой атрибут функционально полно зависит от первичного ключа (составного).

# Требования второй нормальной формы (2NF)

- Чтобы база данных находилась во второй нормальной форме (2NF), необходимо чтобы ее таблицы удовлетворяли следующим требованиям:
- Таблица должна находиться в первой нормальной форме
- Таблица должна иметь ключ
- Все неключевые столбцы таблицы должны зависеть от полного ключа (*в случае если он составной*)

# Требования второй нормальной формы (2NF)

- **Ключ** – это столбец или набор столбцов, по которым гарантировано можно отличить строки друг от друга, т.е. ключ идентифицирует каждую строку таблицы. По ключу мы можем обратиться к конкретной строке данных в таблице.
- Таблица должна иметь правильный ключ, по которому можно идентифицировать каждую строку.

ФИО	Должность	Подразделение	Описание подразделения
Иванов И.И.	Программист	Отдел разработки	Разработка и сопровождение приложений и сайтов
Сергеев С.С.	Бухгалтер	Бухгалтерия	Ведение бухгалтерского и налогового учета финансово-хозяйственной деятельности
John Smith	Продавец	Отдел реализации	Организация сбыта продукции

Мы видим, что она удовлетворяет условиям первой нормальной формы, т.е. в ней нет дублирующих строк и все значения атомарны.

Табельный номер	ФИО	Должность	Подразделение	Описание подразделения
1	Иванов И.И.	Программист	Отдел разработки	Разработка и сопровождение приложений и сайтов
2	Сергеев С.С.	Бухгалтер	Бухгалтерия	Ведение бухгалтерского и налогового учета финансово-хозяйственной деятельности
3	John Smith	Продавец	Отдел реализации	Организация сбыта продукции

В результате, так как наш первичный ключ является простым, а не составным, наша таблица автоматически переходит во вторую нормальную форму.

## Вторая нормальная форма:

Таб № РК	№ ребёнка РК	Имя ребёнка	Год рожде ния
1	1	Саша	1996
2	1	Петя	1992
1	2	Маша	1995
2	2	Витя	1997
2	3	Лена	1999

Таб. № РК	ФИО	Должность	№ кабине та	Телефон
1	Иванов	Директор	32	212
2	Зайцев	Бухгалтер	33	213
3	Волков	Менеджер	35	311
4	Кошкина	Менеджер	35	311

# ***Третья нормальная форма***

Определение 1. Отношение находится в 3НФ, если оно находится в 2НФ, и каждый неключевой атрибут нетранзитивно зависит от первичного ключа.

Альтернативное определение.

Определение 2. Отношение находится в 3НФ в том и только в том случае, если все неключевые атрибуты отношения взаимно независимы и полностью зависят от ключа.



# ***Третья нормальная форма***

Требование третьей нормальной формы (3NF) заключается в том, чтобы в таблицах отсутствовала транзитивная зависимость.

**Транзитивная зависимость** — это когда неключевые столбцы зависят от значений других неключевых столбцов.

Таблица должна содержать правильные неключевые столбцы

В чем проблема таблицы?

Табельный номер	ФИО	Должность	Подразделение	Описание подразделения
1	Иванов И.И.	Программист	Отдел разработки	Разработка и сопровождение приложений и сайтов
2	Сергеев С.С.	Бухгалтер	Бухгалтерия	Ведение бухгалтерского и налогового учета финансово-хозяйственной деятельности
3	John Smith	Продавец	Отдел реализации	Организация сбыта продукции

**Декомпозиция** – это процесс разбиения одного отношения (таблицы) на несколько.

## Третья нормальная форма:

Таб. № РК	ФИО	Должность	№ кабинета	Телефон
1	Иванов	Директор	32	212
2	Зайцев	Бухгалтер	33	213
3	Волков	Менеджер	35	311
4	Кошкина	Менеджер	35	311

- ❖ Предположим, что в кабинете № 35 изменился телефон. Изменения требуется синхронно внести в две строки, следовательно, существует аномалия модификаций.
- ❖ Пусть есть кабинет, который ещё не занят, тогда невозможно (без сотрудника) внести информацию о телефоне в этом кабинете. Следовательно, существует аномалия вставки
- ❖ Существует аномалия удаления (если в кабинете капитальный ремонт, то необходимо убрать сотрудников, следовательно, теряются данные о телефоне).

## Третья нормальная форма:

- ❖ Причина существования аномалий в том, что № телефона не зависит от ключа (точнее, зависит от ключа транзитивно, через № кабинета). В отношении существует зависимость между неключевыми атрибутами.

## Третья нормальная форма:

<b>№ кабинета</b> РК	<b>Телефон</b>
32	212
33	213
35	311

<b>Таб. №</b>	<b>ФИО</b>	<b>Должность</b>	<b>№ кабинета</b>
1	Иванов	Директор	32
2	Зайцев	Бухгалтер	33
3	Волков	Менеджер	35
4	Кошкина	Менеджер	35

# ***Усиленная 3НФ или нормальная форма Бойса-Кодда (БКНФ)***

- Отношение находится в БКНФ, если оно находится в 3НФ и в нем отсутствуют зависимости ключей (атрибутов составного ключа) от неключевых атрибутов. У нас подобной зависимости нет, поэтому процесс проектирования на этом заканчивается. Результатом проектирования является БД, состоящая из следующих таблиц: R1, R3, R4, R5. В полученной БД имеет место необходимое дублирование данных, но отсутствует избыточное.

# Нормальная форма Бойса-Кодда (BCNF)

Требования нормальной формы Бойса-Кодда следующие:

- Таблица должна находиться в третьей нормальной форме. Здесь все как обычно, т.е. как и у всех остальных нормальных форм, первое требование заключается в том, чтобы таблица находилась в предыдущей нормальной форме, в данном случае в третьей нормальной форме;
- Ключевые атрибуты составного ключа не должны зависеть от неключевых атрибутов.



# Нормальная форма Бойса-Кодда (BCNF)

Требования нормальной формы Бойса-Кодда следующие:

- Таблица должна находиться в третьей нормальной форме.
- Ключевые атрибуты составного ключа не должны зависеть от неключевых атрибутов.
- Часть составного первичного ключа не должна зависеть от неключевого столбца.

# Нормальная форма Бойса-Кодда (BCNF)

Требования нормальной формы Бойса-Кодда следующие:

- Таблица должна находиться в третьей нормальной форме.
- Ключевые атрибуты составного ключа не должны зависеть от неключевых атрибутов.
- Часть составного первичного ключа не должна зависеть от неключевого столбца.

Проект	Направление	Куратор
1	Разработка	Иванов И.И.
1	Бухгалтерия	Сергеев С.С.
2	Разработка	Иванов И.И.
2	Бухгалтерия	Петров П.П.
2	Реализация	John Smith
3	Разработка	Андреев А.А.

Но в данном случае таблица не находится в нормальной форме Бойса-Кодда, дело в том, что зная куратора, мы можем четко определить, какое направление он курирует, иными словами, часть составного ключа, т.е. «*Направление*», зависит от неключевого атрибута, т.е. «*Куратора*».

Чтобы привести данную таблицу к нормальной форме Бойса-Кодда, необходимо, как всегда сделать декомпозицию данного отношения, т.е. разбить эту таблицу на несколько таблиц.

Таблица кураторов.

Идентификатор куратора	ФИО	Направление
1	Иванов И.И.	Разработка
2	Сергеев С.С.	Бухгалтерия
3	Петров П.П.	Бухгалтерия
4	John Smith	Реализация
5	Андреев А.А.	Разработка

Таблица связи кураторов и проектов.

Проект	Идентификатор куратора
1	1