

- ❖ 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.

- ❖ Процесс проектирования производится методом последовательных приближений к удовлетворительному набору схем отношений. Исходной точкой является представление предметной области в виде одного или нескольких отношений, и на каждом шаге проектирования производится некоторый набор схем отношений, обладающих лучшими свойствами. Процесс проектирования представляет собой процесс нормализации схем отношений, причем каждая следующая нормальная форма обладает свойствами лучшими, чем предыдущая.

СТРУКТУРА НОРМАЛИЗАЦИИ:

В теории реляционных баз данных обычно выделяется следующая последовательность нормальных форм:

- первая нормальная форма (1NF);
- вторая нормальная форма (2NF);
- третья нормальная форма (3NF);
- нормальная форма Бойса-Кодда (BCNF);
- четвертая нормальная форма (4NF);
- пятая нормальная форма, или нормальная форма проекции-соединения (5NF или PJ/NF).

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.

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.

Основные определения:

- ❖ Наиболее важные на практике нормальные формы отношений основываются на фундаментальном в теории реляционных баз данных понятии *функциональной зависимости*. Для дальнейшего изложения потребуются несколько определений.

Определение 1. *Функциональная зависимость*

В отношении R атрибут Y функционально зависит от атрибута X (X и Y могут быть составными) в том и только в том случае, если каждому значению X соответствует в точности одно значение Y :

$$R.X \twoheadrightarrow R.Y.$$

Определение 2. *Полная функциональная зависимость*

Функциональная зависимость $R.X \twoheadrightarrow R.Y$ называется *полной*, если атрибут Y не зависит функционально от любого точного подмножества X .

Основные определения:

Определение 3. *Транзитивная функциональная зависимость*

Функциональная зависимость $R.X \twoheadrightarrow R.Y$ называется транзитивной, если существует такой атрибут Z , что имеются функциональные зависимости $R.X \twoheadrightarrow R.Z$ и $R.Z \twoheadrightarrow R.Y$ и отсутствует функциональная зависимость $R.Z \twoheadrightarrow R.X$. (При отсутствии последнего требования мы имели бы "неинтересные" транзитивные зависимости в любом отношении, обладающем несколькими ключами.)

Определение 4. *Неключевой атрибут*

Неключевым атрибутом называется любой атрибут отношения, не входящий в состав первичного ключа.

Определение 5. *Взаимно независимые атрибуты*

Два или более атрибута взаимно независимы, если ни один из этих атрибутов не является функционально зависимым от других.

Вторая нормальная форма:

Определение:

Отношение R находится во второй нормальной форме (2NF) в том и только в том случае, когда оно находится в 1NF, и каждый неключевой атрибут функционально полно зависит от ключа R .

Вторая нормальная форма:

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

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

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:

- Определение:** Отношение R находится в третьей нормальной форме (3NF) в том и только в том случае, если находится во 2NF, и каждый неключевой атрибут не является транзитивно зависимым от какого-либо ключа R .
- ❖ На практике третья нормальная форма схем отношений достаточно в большинстве случаев, и приведением к третьей нормальной форме процесс проектирования реляционной базы данных обычно заканчивается.

Третья нормальная форма:

№ кабинета РК	Телефон
32	212
33	213
35	311


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

Основные свойства нормальных форм:

- каждая следующая нормальная форма в некотором смысле лучше предыдущей;
- при переходе к следующей нормальной форме свойства предыдущих нормальных свойств сохраняются.

НАЗНАЧЕНИЕ НОРМАЛИЗАЦИИ

- Устранить аномалии обновления отношений (При проведении нормализации базы данных происходит устранение или серьезное сокращение избыточности, дублирования данных):
 1. Аномалии вставки
 2. Аномалии удаления
 3. Аномалия обновления
- Значительно сокращается вероятность появления противоречивых данных.
- Облегчается администрирование базы и обновление информации в ней, сокращается объем занятого базой дискового пространства.



Зачастую, чтобы извлечь информацию из нормализованной базы данных, приходится конструировать очень сложные запросы, которые к тому же, бывает, работают довольно медленно - из-за, главным образом, большого количества соединений таблиц. Поэтому, чтобы увеличить скорость выборки данных и упростить программирование запросов, нередко приходится идти на выборочную денормализацию базы.