

Нормализация таблиц

Различия между дублированием и избыточным дублированием данных. Рассмотрим таблицу учета книг, взятых студентами в библиотеке:

Книга	Студент
Книга 1	Иванов
Книга 2	Петров
Книга 3	Иванов
Книга 4	Петров

Здесь студенты дублируются, но удалить их нельзя. Данные дублирующиеся, но не избыточные.

Книга	Студент	Телефон
Книга 1	Иванов	33-33-33
Книга 2	Петров	55-55-55
Книга 3	Иванов	33-33-33
Книга 4	Петров	55-55-55

Здесь данные по телефонам избыточные. Эту таблицу лучше разбить на две.

Книга	Студент
Книга 1	Иванов
Книга 2	Петров
Книга 3	Иванов
Книга 4	Петров

Студент	Телефон
Иванов	33-33-33
Петров	55-55-55

Из этого примера видно, что для таблиц очень важны проблемы удаления и обновления записей. Поэтому важно находить потенциально опасные таблицы и "нормализовать" их путем разбиения определенным образом на несколько таблиц. Нормализация – это формальная процедура, в ходе которой одна таблица разбивается на две или несколько по определенным правилам. Задачи нормализации:

1. Исключение из таблиц повторяющейся информации.
2. Создание структуры, в которой предусмотрена возможность ее будущих изменений.
3. Создание структуры, в которой влияние структурных изменений на приложения, использующие информацию из этой базы данных, сведено к минимуму.

Существует несколько нормальных форм: 1 НФ, 2 НФ, ..., 5 НФ, НФ Бойса-Кодда. При практической разработке важны 1-я, 2-я и 3-я НФ.

Первая НФ требует, чтобы каждое поле таблицы:

- было неделимым (не должно делиться на более мелкие значения)
- не содержало повторяющихся групп.

Повторяющиеся группы:

Студент	Телефон
Иванов	33-33-33; 44-44-44
Петров	55-55-55; 88005555501

Это ненормализованная таблица.

Студент	Телефон
Иванов	33-33-33
Иванов	44-44-44
Петров	55-55-55
Петров	88005555501

Таблица приведена к 1-й НФ.

Пример. Требуется автоматизировать процесс отпуска товара со склада.

Накладная №5

Дата: 03.03.2013 Покупатель: ООО "Друг" Адрес: г. Саранск

Наименование	Кол-во	Ед. изм.	Цена	Сумма
Тушенка	1000	Банка	50	50000
Сахар	100	Кг	30	3000

Итого: 53000 руб

Сведем данные в одну таблицу:

Отпуск товаров

Номер
Дата
Покупатель
Адрес
Наименование
Ед_изм
Цена
Количество
Сумма

Эта таблица приведена к 1-й НФ.

Вторая НФ требует, чтобы все поля таблицы зависели от первичного ключа, т.е. чтобы первичный ключ однозначно определял запись и не был избыточным. Те поля, которые зависят только от части первичного ключа, должны быть выделены в отдельные таблицы.

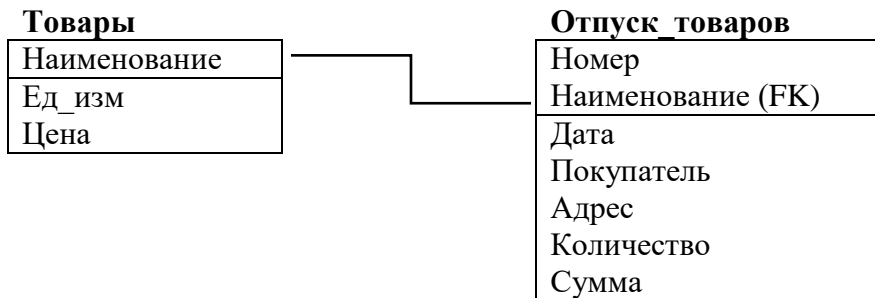
В нашем примере считаем, что номер накладной сквозной (в разные даты не может быть накладной с одним и тем же номером).

Отпуск_товаров

Номер
Наименование
Дата
Покупатель
Адрес
Наименование
Ед_изм
Цена
Количество

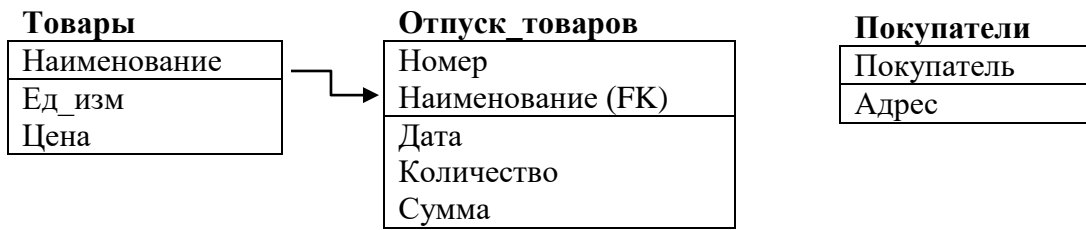
Сумма

Это таблица с избыточным первичным ключом, не приведенная к 2-й НФ. Здесь "Ед_изм" и "Цена" зависят только от "Наименование". Поэтому выделяем данные поля в отдельную таблицу "Товары" и определяем связь.



Выделение таблицы "Товары"

"Покупатель" и "Адрес" зависят только от номера накладной. Поэтому "Покупатель" и "Адрес" выделяем в отдельную таблицу "Покупатели" (пока без связей).



"Дата" зависит только от "Номер". Поэтому выделяем дату и номер накладной в таблицу "Накладные".



Установим теперь связи между таблицами.



Третья НФ требует, чтобы в таблице не имелось транзитивных зависимостей между неключевыми полями, т.е. чтобы значение любого поля таблицы, не входящего в первичный ключ, не зависело от значения другого поля, не входящего в первичный ключ.

В нашем примере "Отпуск_товаров"."Сумма"="Отпуск_товаров"."Количество"*"Товары"."Цена", поэтому поле "Сумма" можно удалить. В результате получим нормализованную базу данных.

Таким образом, таблицы нормализованной базы данных содержат только один элемент избыточных данных – это поле связи, присутствующее одновременно у родительской и дочерних таблиц.

Недостатки нормализованных баз данных:

1. Крупные БД могут состоять из огромного числа связанных таблиц – теряется восприятие системы в целом.
2. При выполнении запросов приходится считывать данные из разных таблиц. Это может приводить к более сложным запросам и потерям времени.