

# Организация баз данных и знаний

## Лекция 4. Нормализация базы данных

Лектор: Селиванова Алла Витальевна

# Нормализация

это метод создания набора отношений с заданными свойствами на основе требований, предъявляемых к данным в организации

# Аномалии

Отношения с избыточностью данных могут страдать от **аномалий обновления**, которые делятся на аномалии

- вставки,
- удаления
- обновления данных.

# ФУНКЦИОНАЛЬНАЯ

**ЗАВИСИМОСТЬ** описывает связь

между атрибутами отношения

- Пусть  $A$  и  $B$  — это атрибуты некоторого отношения  $K$ .
- Атрибут  $B$  функционально зависит от атрибута  $A$  ( $A \rightarrow B$ ), если каждое значение  $A$  связано с одним значением  $B$ .
- Причем каждый из атрибутов  $A$  и  $B$  может состоять из одного или нескольких атрибутов.

# Детерминантом

- называется любой атрибут, от которого полностью функционально зависит какой-то другой атрибут.
- В определении функциональной зависимости термин "детерминант" характеризует один или несколько атрибутов, расположенных с левой стороны от стрелки **A—»B**.

# Ненормализованной формой (ННФ)

- называется таблица, которая содержит одну или несколько повторяющихся групп атрибутов.

Повторяющиеся группы значений

ACCOUNTCD	STREETCD	STREETNM	HOUSENO	FLATNO	FIO	PHONE
005488		→3 ВОЙКОВ ПЕРЕУЛОК		4	1 АКСЕНОВ С.А.	556893
015527		→3 ВОЙКОВ ПЕРЕУЛОК		1	65 КОНОХОВ В.С.	761699
080047		8 МОСКОВСКОЕ ШОССЕ УЛИЦА		39	36 СЕРОВА Т.П.	257842
080270		6 МОСКОВСКАЯ УЛИЦА		35	6 ТИМОШКИНА Н.Г.	321002
080613		8 МОСКОВСКОЕ ШОССЕ УЛИЦА		35	11 ЛУКАШИНА Р.М.	254417
115705		→3 ВОЙКОВ ПЕРЕУЛОК		1	82 МИЦЕНКО Е.В.	769975
126112		4 ТАТАРСКАЯ УЛИЦА		7	11 МАРКОВА В.Л.	683301
136159		7 КУТУЗОВА УЛИЦА		39	1 СВЕРИНА Э.А.	350003
136160		4 ТАТАРСКАЯ УЛИЦА		9	15 ШМАКОВ С.В.	982222
136169		4 ТАТАРСКАЯ УЛИЦА		7	13 ДЕНИСОВА Е.К.	680305
443069		4 ТАТАРСКАЯ УЛИЦА		51	55 СТАРОДУБЦЕВ Е.В.	683014
443690		7 КУТУЗОВА УЛИЦА		5	1 ТУЛЬПОВА М.И.	pmbk.ru

# Первой нормальной формой

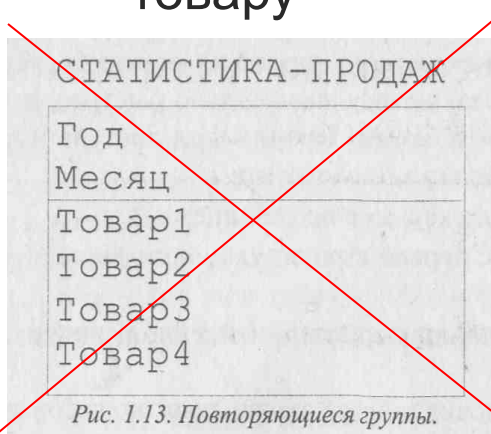
- **(1НФ)** называется отношение, в котором на пересечении каждой строки и каждого столбца располагается одно и только одно значение

# 1НФ

Первая нормальная форма требует, чтобы каждое поле таблицы БД было неделимым и не содержало повторяющихся групп

*Неделимость* поля означает, что содержащиеся в нем значения не должны делиться на более мелкие.

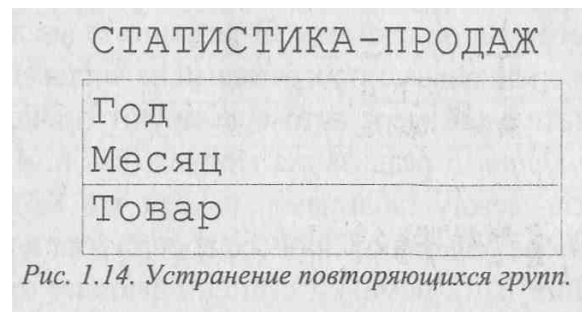
*Повторяющимися* являются поля, содержащие одинаковые по смыслу значения. Например, если требуется получить статистику продаж четырех товаров по месяцам, можно создать поля для хранения данных о продаже по каждому товару



СТАТИСТИКА-ПРОДАЖ

Год
Месяц
Товар1
Товар2
Товар3
Товар4

*Рис. 1.13. Повторяющиеся группы.*



СТАТИСТИКА-ПРОДАЖ

Год
Месяц
Товар

*Рис. 1.14. Устранение повторяющихся групп.*



# Второй нормальной формой

- **(2НФ)** называется отношение, которое находится в первой нормальной форме, а каждый атрибут, не входящий в первичный ключ, полностью функционально зависит от этого первичного ключа.
- **Полная функциональная зависимость** для атрибутов  $A$  и  $B$  некоторого отношения означает следующее: атрибут  $B$  полностью функционально зависит от атрибута  $A$ , если атрибут  $B$  функционально зависит от атрибута  $A$ , но не зависит ни от какого подмножества атрибута  $A$

# 2NF

- 2NF требует, чтобы все поля таблицы зависели от первичного ключа, то есть чтобы первичный ключ однозначно определял запись и не был избыточен.
- Те поля, которые зависят только от части первичного ключа, должны быть выделены в составе отдельных таблиц.

ОТПУСК-ТОВАРОВ-СО-СКЛАДА

Дата
Покупатель
Номер_накладной
Товар
Город
Адрес
Ед_измерения
Цена_за_ед_изм
Отпущено_ед
Общая_стоимость

Рис. 1.17. Таблица с избыточным первичным ключом.

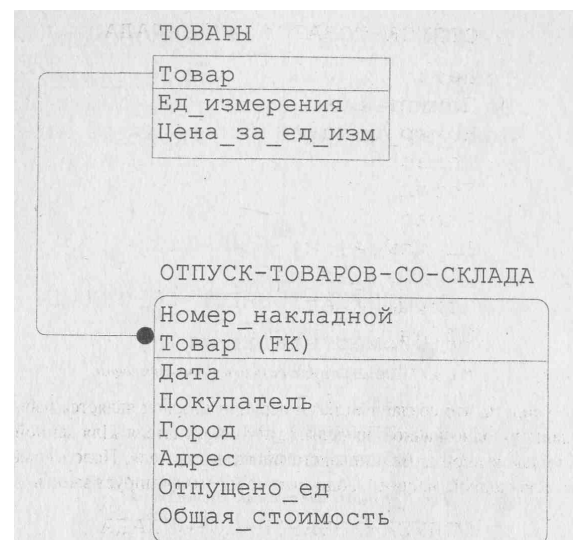


Рис. 1.19. Выделение таблицы «Товар».

# Третьей нормальной формой

- **(3НФ)** называется отношение, которое находится в первой и **во второй** нормальной форме, причем в нем нет атрибутов, не входящих в первичный ключ, которые транзитивно зависят от этого первичного ключа. **Транзитивная зависимость** для атрибутов  $A$ ,  $B$  и  $C$  некоторого отношения означает следующее: если  $A \rightarrow B$  и  $B \rightarrow C$ , то  $C$  транзитивно зависит от атрибута  $A$  через атрибут  $B$  (при условии, что  $A$  функционально не зависит от  $B$  или  $C$ ).

# ЗНФ

- требует, чтобы в таблице не имелось транзитивных зависимостей между неключевыми полями, то есть чтобы значение любого поля, не входящего в первичный ключ, не зависело от значения другого поля, также не входящего в первичный ключ.

# Нормальной формой Бойса-Кодда (НФБК)

- называется отношение, в котором каждый детерминант является потенциальным ключом

# Четвертой нормальной формой

- **(4НФ)** называется отношение, которое находится в нормальной форме Бойса-Кодда и не содержит нетривиальных многозначных зависимостей.
- **Многозначная зависимость** представляет такую зависимость между атрибутами А, В и С некоторого отношения, при которой для каждого значения атрибута А существуют соответствующие наборы значений атрибутов В и С, причем оба этих набора не зависят друг от друга.

# Пятой нормальной формой

- **(5НФ)** называется отношение, которое не содержит зависимостей соединения.  
**Зависимость соединения** — это такая ситуация при которой декомпозиция отношения может сопровождаться генерацией ложных строк при обратном соединении декомпозированных отношений посредством операции естественного соединения.

# (5НФ)

- Пятая нормальная форма - это последняя нормальная форма, которую можно получить путем декомпозиции.
- Ее условия достаточно нетривиальны, и на практике 5НФ не используется.
- Зависимость соединения является обобщением как многозначной зависимости, так и функциональной зависимости.



# Ненормализованная форма

Удаление повторяющихся групп

1 НФ

Удаление зависимости от части ключа

2 НФ

Удаление транзитивных зависимостей

3 НФ

Удаление из зависимостей оставшихся аномалий

НФ Бойса-Кодда

Удаление многозначных зависимостей

4 НФ

Удаление зависимостей соединения

5 НФ

Минимальный достаточный уровень нормализации

# Преимущества нормализации

- Нормализация таблиц БД призвана устранить из них избыточную информацию.
- таблицы нормализованной БД содержат только один элемент избыточных данных - это **поля связи**, присутствующие одновременно у родительской и дочерних таблиц.
- Поскольку избыточные данные в таблицах не хранятся, экономится дисковое пространство.

# Недостатки нормализации

- Чем шире число сущностей, охватываемых предметной областью, тем из большего числа таблиц будет состоять нормализованная БД.
- с увеличением числа нормализованных таблиц уменьшается целостное восприятие базы данных как *системы взаимосвязанных данных*.
- необходимость считывать связанные данные из нескольких таблиц при выполнении одного запроса.

# Нормализация за и против

- При работе с данными большого объема приходится искать компромисс между требованиями нормализации (то есть логичности данных и экономии места на носителях информации) и необходимостью улучшения быстродействия системы.

# Денормализация

- это процесс осознанного приведения базы данных к виду, в котором она не будет соответствовать правилам нормализации. Обычно это необходимо для повышения производительности и скорости извлечения данных, за счет увеличения избыточности данных.
- Прежде чем приступить к **денормализации**, необходимо убедиться, что ожидаемые результаты оправдывают издержки, с которыми придется столкнуться