

## Тема 22. Нормальные формы схем отношений

В рамках реляционной модели данных разработан аппарат нормализации отношений, позволяющий устранить некоторые из перечисленных ранее недостатков схем БД. Нормализация широко используется как в системах автоматизации проектирования, так и при ручном проектировании БД.

Определение. Простым атрибутом называется атрибут, значения которого неделимы (атомарны). Сложный (составной) – тот, который может иметь значения, состоящие из нескольких простых значений (конкатенация или множество).

### Первая нормальная форма

Схема отношения называется нормализованной, находится в первой нормальной форме (1НФ), если все ее атрибуты – простые. Приведение ненормализованной схемы к 1НФ достигается заменой сложного атрибута набором простых.

#### Пример.

Исходная схема:

Поставщики' (**Название**, Телефон, Адрес, **Поставка** (Товар, Цена)).

Атрибут Поставка сложный, является конкатенацией атрибутов Товар и Цена. Нормализация схемы достигается заменой сложного атрибута Поставка простыми атрибутами Товар и Цена. Получаем схему Поставщики'' в 1НФ:

Поставщики'' (**Название**, Телефон, Адрес, **Товар**, Цена).

Жирным шрифтом здесь выделены ключевые атрибуты. Замена сложного атрибута простыми может изменить ключ отношения. В данном случае ключ **Название, Поставка** отношения Поставщики' заменяется ключом **Название, Товар** отношения Поставщики''.

### Вторая нормальная форма

#### Определение.

Атрибут А в схеме R называется (ключевым), если он является элементом какого-либо ключа R. В противном случае он называется неключевым.

Неключевой атрибут функционально полно зависит от составного ключа, если он функционально зависит от ключа, но не находится в функциональной зависимости ни от какой его части. В противном случае имеет место частичная функциональная зависимость неключевого атрибута от составного ключа.

Например, в схеме "Поставщики" атрибуты Телефон и Адрес зависят только от атрибута Название и поэтому находятся в частичной функциональной зависимости от составного ключа **Название,Товар**. Неключевой атрибут Цена находится в полной функциональной зависимости от составного ключа.

Частичные зависимости приводят к дублированию информации.

### **Определение.**

Схема отношения находится во второй нормальной форме 2НФ, если она находится в 1НФ и каждый неключевой атрибут функционально полно зависит от составного ключа.

Для приведения схемы в 2НФ нужно:

- Исключить из схемы отношения атрибуты, не находящиеся в полной функциональной зависимости от составного ключа.
- Построить дополнительно одну или несколько схем отношений, включающие часть атрибутов составного ключа и неключевые атрибуты, функционально зависящие от этой части ключа.

Приведение схемы "Поставщики" в 2НФ достигается заменой ее двумя схемами (декомпозицией):

Поставщики'''(Название,Товар,Цена),

Пост\_Тел\_Адрес(Название,Телефон,Адрес).

### **Третья нормальная форма**

#### **Определение.**

Пусть  $R$  – схема,  $F$  - множество функциональных зависимостей схемы;  $X, Y, Z$  - подмножества множества атрибутов схемы  $R$ . Пусть  $X \rightarrow Y, Y \rightarrow Z \in F^+$ ; причем  $Y \rightarrow X \notin F^+$ . Тогда говорят, что  $Z$  транзитивно зависит от  $X$ . Множество  $Z$  строго транзитивно зависит от  $X$ , если  $Y \rightarrow X \notin F^+$ .

#### **Определение.**

Схема отношения находится в 3НФ, если она находится в 2НФ и каждый неключевой атрибут нетранзитивно зависит от каждого ключа R.

### **Пример.**

Пусть несколько поставщиков могут располагаться по одному адресу, и с каждым адресом связан единственный телефон. В этом случае в схеме отношения Пост\_Тел\_Адрес имеется транзитивная зависимость неключевого атрибута Телефон от ключа **Название**: Название → Адрес, Адрес → Телефон. Эта зависимость нежелательна, т.к. она приводит к избыточности (дублированию номера телефона для всех поставщиков с одним адресом), аномалиям включения и удаления – невозможности фиксации адреса поставщика, если нет телефона, потери адреса поставщика при удалении телефона.

### **Нормальная форма Бойса-Кодда**

Определение. Схема отношения R с зависимостями F находится в нормальной форме Бойса-Кодда, если всякий раз, когда в F имеется зависимость  $X \rightarrow A$  и  $A \notin X$ , X включает некоторый ключ R.

Иными словами, допускаются только такие нетривиальные зависимости, в которых ключ функционально определяет один или более других атрибутов.

Можно показать, что если схема отношения находится в нормальной форме Бойса-Кодда, то она находится и в 3НФ. Нормальная форма Бойса-Кодда обладает всеми достоинствами 3НФ и предотвращает некоторые аномалии, допускаемые 3НФ.

Можно показать, что любая схема отношения может быть приведена в нормальную форму Бойса-Кодда таким образом, что соответствующая декомпозиция обладает свойством соединения без потерь. Возможно также приведение в 3НФ с получением декомпозиции, обладающей свойством соединения без потерь и сохраняющей зависимости. Однако схема отношения может оказаться неприводимой в НФБК с сохранением зависимостей.

### ***Приведение в НФБК***

Опишем алгоритм приведения в НФБК с получением декомпозиции, обладающей свойством соединения без потерь.

Вход: схема  $R$ , множество функциональных зависимостей  $F$ .

Выход: декомпозиция схемы  $R$ , обладающая свойством соединения без потерь, такая, что каждая схема в декомпозиции находится в НФБК относительно проекции  $F$  на эту схему.

Метод: Итерации. В начале  $\rho = \{R\}$ . На каждом шаге выполняется проверка: если  $S$  – схема из  $\rho$  и  $S$  не находится в НФБК, то пусть  $X \rightarrow A$  – зависимость в  $S$ , где  $X$  не содержит ключа  $S$  и  $A \notin X$ . Заменяем  $S$  в  $\rho$  на  $S_1$  и  $S_2$ , где  $S_1$  состоит из  $A$  и атрибутов  $X$ , а  $S_2$  – из всех атрибутов  $S$  за исключением  $A$ .

В конце концов, достигнем момента, когда декомпозиция  $\rho$  будет состоять из схем в НФБК, причем  $\rho$  обладает свойством соединения без потерь. Последнее обусловлено тем, что начальное  $\rho$  обладает им, а все последующие шаги сохраняют свойство соединения без потерь.

#### **Пример.**

Пост-Тел\_Адрес(**Название**, Телефон, Адрес)

Схема отношения не находится в НФБК, так как имеется зависимость Адрес  $\rightarrow$  Телефон и Адрес не ключ отношения. Разбиваем схему на две: Адрес\_Телефон(**Адрес**, Телефон) и Пост\_Адрес(**Название**, Адрес).

Обе схемы находятся в НФБК, причем естественное соединение отношений с данными схемами восстанавливает отношение со схемой Пост\_Тел\_Адрес (декомпозиция обладает свойством соединения без потерь).

### ***Приведение в третью нормальную форму***

Произвольная схема отношения всегда может быть заменена сохраняющей зависимости декомпозицией в ЗНФ.

Алгоритм.

Вход. Схема отношения  $R$  и множество зависимостей  $F$ .

Выход. Сохраняющая зависимости декомпозиция  $R$ , все схемы которой находятся в ЗНФ относительно проекции  $F$  на эту схему.

Метод. Для каждой зависимости  $X \rightarrow A$  из  $F$  декомпозиция включает схему  $XA$ . Если в  $F$  имеются зависимости  $X \rightarrow A_1, \dots, X \rightarrow A_n$ , то можно использовать схему  $XA_1 \dots A_n$ . Если некоторые атрибуты не участвуют в зависимостях из  $F$ , то они образуют отдельную схему.

**Пример.**

Схема Пост\_Тел\_Адрес(**Название**, Телефон, Адрес).

Множество ф.з.  $F = \{\text{Название} \rightarrow \text{Адрес}, \text{Адрес} \rightarrow \text{Телефон}\}$ .

Для ф.з.  $\text{Название} \rightarrow \text{Адрес}$  получаем схему Пост\_Адрес(**Название**, Адрес).

Для ф.з.  $\text{Адрес} \rightarrow \text{Телефон}$  получаем схему Адрес\_Тел(**Адрес**, Телефон).

Для приведения схемы отношения в 3НФ с сохранением зависимостей и соединением без потерь надо к полученной декомпозиции добавить схему, состоящую из ключевых атрибутов  $R$ .

***Многозначные зависимости***

Кроме функциональных, имеются также многозначные зависимости. Пусть  $R$  – схема,  $X$  и  $Y$  – подмножества  $R$ . Говорят, что  $X$  мультиопределяет  $Y$  (существует многозначная зависимость от  $X$  к  $Y$  (обозначается  $X \twoheadrightarrow Y$ )), если при заданных значениях атрибутов из  $X$  существует ноль или более ассоциированных значений атрибутов из  $Y$ , и это множество значений не связано каким-либо образом со значениями атрибутов в  $R-X-Y$ .

В теории РБД доказывается, что в схеме  $R(X, Y, Z)$  существует многозначная зависимость  $X \twoheadrightarrow Y$  в том и только том случае, когда существует многозначная зависимость  $X \twoheadrightarrow Z$ .

Справедливо утверждение: отношение  $R(X, Y, Z)$  можно спроецировать без потерь в отношения  $R_1(X, Y)$  и  $R_2(X, Z)$  в том и только том случае, когда существуют многозначные зависимости  $X \twoheadrightarrow Y$  и  $X \twoheadrightarrow Z$ .

**Определение.**

Отношение R находится в 4НФ в том и только том случае, если при наличии многозначной зависимости  $X \twoheadrightarrow Y$  все остальные атрибуты R функционально зависят от X.