**10. Нормализация. Функциональные зависимости атрибутов. Нормальные формы**

**Нормализация таблиц** это формальный аппарат ограничений на формирование таблиц, описывающий разбиение таблиц на две или более части и обеспечивающий применение лучших методов добавление, изменения и удаления данных.

Можно также сказать, что нормализация – это процесс представления данных в виде простых двумерных таблиц, который позволяет устранить дублирование этих данных и обеспечивает непротиворечивость хранимой в базе информации. **>>***Цель нормализации* – получение проекта БД, в котором любая часть логически законченной информации хранится в одном месте, т.е. исключается избыточность информации.

Основа нормализации – аппарат нормализации отношений.

Всего существует 6 форм нормальных отношений. Но, как правило, необходимо и достаточно привести БД к третьей нормальной форме, чтобы исключить указанные аномалии при работе с БД.

Таблица считается нормализованной на определенном уровне, когда она удовлетворяет условиям, накладываемым соответствующей формой нормализации.

**Формы нормализации:**

-первая нормальная форма (First Normal Form – 1NF);

-вторая нормальная форма (Second Normal Form – 2NF);

-третья нормальная форма (Third Normal Form – 3NF);

-нормальная форма Бойса-Кодда (Boice-Codd Normal Form – BCNF);

-четвертая нормальная форма (Fourth Normal Form – 4NF);

-пятая нормальная форма или нормальная форма проекции-соединения (Fifth Normal Form – 5NF или PJ/NF);

Теория нормализации основывается на наличии той или иной зависимости между атрибутами отношения.

***Виды зависимостей:***

**Функциональные зависимости**

Для устранения указанных аномалий (а на самом деле *для правильного проектирования модели данных*!) применяется метод нормализации отношений. Нормализация основана на понятии функциональной зависимости атрибутов отношения^

1) Пусть *R* - отношение. Множество атрибутов *Y* **функционально зависимо**от множества атрибутов *X* (*X* **функционально определяет***Y*) тогда и только тогда, когда *для любого состояния* отношения *R* для любых кортежей r1,r2∈*R* из того, что *r1X=r2X* следует что *r1Y=r2Y* (т.е. во всех кортежах, имеющих одинаковые значения атрибутов *X*, значения атрибутов *Y* также совпадают *в любом состоянии* отношения *R*). Символически функциональная зависимость записывается *X*→*Y*.

Замечание. Если атрибуты *X* составляют потенциальный ключ отношения *R*, то *любой* атрибут отношения *R* функционально зависит от *X*.

**Полная функциональная зависимость**

*К* ключ, *А1*,*A2*,...*An* – некоторые атрибуты в отношении *R*. Пусть *А1*,*A2*,...*An*∈*К* (т.е. имеем составной ключ)

Полной функциональной зависимостью неключевых атрибутов *B1*, *B2*,… *Bm* называется такая зависимость, при которой каждый *B1*, *B2*,… *Bm* функционально зависит от *K* (т.е. от всей совокупности атрибутов ключа), но не находится в функциональной зависимости ни от какой части составного ключа.

**Пример**. В отношении **Университет** можно привести следующие примеры полных функциональных зависимостей:

**{*Caf*, *Disp, Cat*}**→**FIO**

**Транзитивная зависимость** наблюдается в том случае, если один из двух неключевых атрибутов функционально зависит от ключа, а другой неключевой атрибут зависит только от первого.

Нетранзитивная зависимость обратно понятию транзитивной зависимости. Т.е. **нетранзитивная зависимость** имеется в том случае, если ни один из неключевых атрибутов функционально не зависит от любого другого неключевого атрибута.

**Многозначная зависимость** Атрибут *A* однозначно определяет атрибут *B*, если для каждого значения атрибута *A* существует хорошо определенное множество соответствующих значений атрибута *B*.

**Первая нормальная форма** (**1НФ**) - это обычное отношение. Согласно нашему определению отношений, любое отношение автоматически уже находится в 1НФ. Напомним кратко свойства отношений (это и будут свойства 1НФ):

-В отношении нет одинаковых кортежей.

-Кортежи не упорядочены.

-Атрибуты не упорядочены и различаются по наименованию.

-Все значения атрибутов атомарны.

Отношение *R* находится во **второй нормальной форме** (**2НФ**) тогда и только тогда, когда отношение находится в 1НФ и все ее поля, не входящие а первичный ключ, связаны полной функциональной зависимостью с первичным ключом. Т.е. любое неключевое поле однозначно идентифицируется полым набором ключевых полей.

Замечание. Если потенциальный ключ отношения является простым, то отношение автоматически находится во 2НФ.

Таблица, находящаяся во 2НФ должна удовлетворять следующим требованиям:

1. Таблица должна содержать данные об одном типе объектов;
2. Каждая таблица должна содержать одно поле или несколько полей, образующих уникальный идентификатор или первичный ключ для каждой строки;
3. Все, не входящие в первичный ключ поля, должны однозначно определяться этим ключом;
4. Если в таблице есть хотя бы одно поле, не зависящее от первичного ключа, то в первичный ключ необходимо включить дополнительные столбцы. Если таких нет, то добавить его.

Если таблица не находится во 2НФ, то нужно выполнить ее *декомпозицию*.

Отношение *R* находится в **третьей нормальной форме** (3**НФ**) тогда и только тогда, когда отношение находится во 2НФ и **ни одно из ее неключевых полей функционально не зависит от любого другого неключевого поля.** Или другими словами, таблица находится в 3НФ, если она находится во 2НФ и каждое ее неключевое поле нетранзитивно зависит от первичного ключа.

Очевидно, что для нашего примера все таблицы удовлетворяют условию 3НФ, кроме отношения ***Disciplines***. Т.к. категория занятия, вообще говоря, зависит от вида дисциплины.

Отношение находится в **нормальной форме Бойса-Кодда** только в том случае, если любая функциональная зависимость между его атрибутами сводится к полной функциональной зависимости от ключа.

Отношение находится в **четвертой нормальной форме**, если оно находится в нормальной форме Бойса-Кодда и все его многозначные зависимости фактически являются функциональными зависимостями от потенциальных ключей.

Отношение находится в **пятой нормальной форме** тогда и только тогда, когда в каждой его полной декомпозиции все проекции содержат возможный ключ. Отношение, не имеющее ни одной сложной декомпозиции также находится в пятой нормальной форме.