

## Лекция 6

$R, S$  – отношения

$p(R, S)$

- предикат, который будем использовать для соединения

### Подходы к реализации соединения

- **Перебор во вложенных циклах**

Псевдокод:

```
for r in R
  for s in S
    if p(r, s)
      concatenate(r, s)
```

- Реализация для частного случая (экви-соединение)

$$\square p(R, S) = (R.a == S.b)$$

```
...
R.sort(a)
S.sort(b)
while ((not endof(R)) and (not endof(S)))
  if (R.a > S.b)
    s.next
  if (R.a < S.b)
    r.next
  if (R.a == S.b)
    concatenate(r, s)
    R.next
...
```

Что тратим по времени?

$$2 * n * \log(n) + n$$

Два логарифма на сортировки и линия на проиход по таблицам

Тако способ рекомендуется использовать, если одна из тиаблиц предварительно отсортирована. **Решение принимает оптимизатор.**

## Нормализация

Приводим отношения к виду, удовлетворяющему **нормальным формам**.

- В базах данных часто встречается дублирование данных, но не всегда это дублирование избыточно.

Таблица: Студент

ФИО (PK)	N Группы (PK)	Образовательная программа	Факультет	Форма обучения
	дублирование (не избыточное)	дублирование (избыточное)	дублирование (избыточное)	дублирование

Первичным ключом делаем ФИО + Номер группы (гарантируем, что в одной группе никогда не будет полных тезок)

- Представим, что одному студенту присвоили образовательную программу. Все его одноклассники имеют ту же ОП. А если его отчислили?

ФИО (PK)	N Группы (PK)	Образовательная программа	Факультет	Форма обучения
Студент1	M3214	09.03.02	ФИТиП	...
Студент2	M3214	NULL	ФИТиП	...
Студент3	M3214	NULL	ФИТиП	...

Здесь же (помимо избыточного выделения памяти) появляются 2 проблемы:

- Если номер ОП дублировать на все кортежи, то при изменении номера ОП в одном кортеже в других кортежах он не поменяется - нарушается целостность
- Если номер ОП хранить только в одном кортеже, то при поиске для другого кортежа придется искать именно тот кортеж - тратим больше времени

И в этом случае говорят, что у нашей модели появляется аномалия

Получается, что избыточность тяжело ликвидировать

Аномалия в базе данных - состояние базы данных, которое приводит к противоречию или существенно усложняет обработку данных

## Виды аномалий:

- **Аномалии модификаций** - изменение значения одной записи повлечет за собой просмотр всей таблицы и изменение значений некоторых других записей
- **Аномалия удаления** - при удалении записи может пропасть и другая информация
- **Аномалия добавление** - информацию в таблицу нельзя поместить, пока она не полна или требуется дополнительный просмотр таблицы

**Функциональная зависимость** - Y функционально зависит от X тогда и только тогда, когда каждому значению из X соответствует в точности одно значение из Y. X - детерминант, Y - зависимая часть. **Пример:** ОП → Факультет

**Частичная функциональная зависимость** - зависимость неключевого атрибута от части составного потенциального ключа. **Пример:** номер ОП частично зависит от первичного ключа (ФИО + N Группы) (потому что номер ОП зависит от номера группы)

**Полная функциональная зависимость** - зависимость неключевого атрибута от всего составного потенциального ключа.

**Транзитивная функциональная зависимость** - функциональная зависимость Z от X является транзитивной, если существует такое множество атрибутов Y, что Z зависит от Y, а Y от X ( $X \rightarrow Z \Rightarrow \exists Y : X \rightarrow Y, Y \rightarrow Z$ )

## Нормальные формы:

- Отношение находится в **первой нормальной форме** (1НФ), если все его атрибуты являются простыми (тут как раз отличие SQL от реляционной алгебры Кодда).
- Отношение находится во **второй нормальной форме** (2НФ), если оно находится в 1НФ и каждый неключевой атрибут функционально полно зависит от первичного ключа.

Видим нарушение 2НФ: **НомерГруппы → Образовательная Программа, Факультет**  
Факультет и ОП зависят только от **части ключа**

Чтобы привести таблицы из предыдущего примера в 2НФ делим ее на 2 таблицы:

Таблица **Студент**

ФИО	Номер Группы	Форма обучения
-----	--------------	----------------

Таблица **Группа**

Номер Группы	ОП	Факультет
--------------	----	-----------

Теперь оба отношения находятся во 2НФ.

Два варианта определения **третьей нормальной формы** (3НФ):

- Отношение находится в **третьей нормальной форме**(3НФ), если нах в 2НФ и все неключевые атрибуты взаимнонезависимы
- Отношение находится в **третьей нормальной форме** (3НФ), если оно находится в 2НФ и ни один неключевой атрибут не находится в транзитивной зависимости от потенциального ключа

Видим нарушение 3НФ: **\*\*Номер группы -> Образовательная программа -> Факультет\*\***  
Факультет транзитивно зависит от номера группы

Приведение к 3НФ обеспечит следующее действие:

Таблица **ОП**

ОП(РК)	Факультет
--------	-----------

Таблица **Группа**

Номер Группы (РК)	ОП
-------------------	----

Теперь для получения связи Студент-Факультет нужно объединить целых 3 таблицы

**Нормальная форма Бойесса-Кодда** (БКНФ):

Нам нужен новый пример!

Таблица **Проекты**

ИСУ	Паспорт	ПроектID	Роль
-----	---------	----------	------

Что тут может быть ключом?

- ИСУ - нет, может повторяться
- ПроектID - тоже нет, может повторяться
- **???** (ИСУ + ПроектID) или (Паспорт + ПроектID) **???** - возьмем первое

Определение **нормальной формы Бойса-Кодда** (БКНФ)

- Отношение находится в **БКНФ**, если оно находится в 3НФ и и детерминанты всех зависимостей являются потенциальными ключами

Видим нарушение БКНФ: паспорт зависит от номера ИСУ, при этом ИСУ - лишь часть потенциального ключа

Приведем к БКНФ:

Декомпозируем таблицы и сделаем две

- Таблица **Проекты**

ИСУ	Паспорт	ПроектID	Роль
-----	---------	----------	------

- Таблица **Паспорт**

ИСУ	Паспорт
-----	---------

**Четвертая нормальная форма (4НФ)**

Приведем другой пример - таблица лекторов и практиков:

ID Дисциплины	ID Лектора	ID Практика
---------------	------------	-------------

Здесь при замене лектора по причине болезни придется заменять его во всех кортежах соответствующей дисциплины.

Чтобы исправить это, будем использовать **4НФ**.

- Отношение находится в **4НФ**, если находится в **БКНФ** и не содержит нетривиальных многозначных зависимостей
- Что такое многозначное отношение?

- Пусть существует некоторое отношение  $r$  со схемой  $R$ , а также два произвольных подмножества атрибутов  $A, B \subseteq R$ . Пусть  $C = R \setminus (A \cup B)$ . В этом случае  $B$  **многозначно зависит** от  $A$ , тогда и только тогда, когда множество значений атрибута  $B$ , соответствующее заданной паре  $(a : A ; c : C)$ , зависит от  $a$  и не зависит от  $c$ .

Что такое тривиальная многозначная зависимость?

- Многозначная зависимость называется тривиальной, если выполняется хотя бы одно из условий:

1. Множество  $A$  является надмножеством  $B$

$$B \subseteq A$$

2. Объединение  $A$  и  $B$  образует весь заголовок отношения.

$$A \cup B = R$$

Символически выражается записью:

$$X \rightarrow Y$$

Видим нарушение 4НФ:

1. **ID\_Дисциплины  $\rightarrow$  ID\_Лектора**

- Зная дисциплину, мы можем определить множество лекторов, которые ее читают
- Это множество **не зависит** от того, какой практик ведет занятия

2. **ID\_Дисциплины  $\rightarrow$  ID\_Практика**

- Зная дисциплину, мы можем определить множество практиков
- Это множество **не зависит** от того, какой лектор читает теорию

**Решение:**

ID Дисциплины	ID Преподавателя	Роль
---------------	------------------	------

Здесь нет аномалии модификации, но при этом мы теряем связь лектор-практик (допускается, что некоторые лектора несовместимы с некоторыми практиками)

Заметим, что в отношении без неключевых атрибутов автоматически выполнена 2НФ и 3НФ.

Помимо этих нормальных форм выделяют 5НФ и 6НФ, но на этом курсе рассматриваться они не будут