

# МЕТОДИЧЕСКИЕ УКАЗАНИЯ И РЕКОМЕНДАЦИИ К ПРАКТИЧЕСКИМ ЗАНЯТИЯМ ПО ДИСЦИПЛИНЕ БАЗЫ ДАННЫХ

## Операции реляционной алгебры

### 1. Объединение

$Q = A \cup B$ . Атрибуты результирующего отношения не поименованы. Дублирующие кортежи не удваиваются. Основное формальное ограничение – одинаковая Арность отношений.

#### Пример 1

Отношение A:  
Арность отношения A=3  
Кол-во кортежей =3

A	B	C
a	b	c
a	a	c
d	b	a

Отношение B:  
Арность отношения B=3  
Кол-во кортежей =4

D	E	F
a	q	w
a	b	c
a	s	d
f	g	h

Отношение Q:  
Кол-во кортежей =6

a	b	c
a	a	c
d	b	a
a	q	w
a	s	d
f	g	h

#### Пример 2

Отношение A:  
Арность отношения A=4  
Кол-во кортежей =4

A	B	C	D
a	s	d	f
z	x	c	v
q	w	e	r
g	h	j	k

Отношение B:  
Арность отношения B=4  
Кол-во кортежей =5

E	F	G	I
c	v	b	n
a	s	d	f
h	j	k	l
q	w	e	r
y	u	i	o

Отношение Q:  
Арность отношения Q=4  
Кол-во кортежей =7

a	s	d	f
z	x	c	v

q	w	e	r
g	h	j	k
c	v	b	n
h	j	k	l
y	u	i	o

Пример 3 – по усмотрению самостоятельно

## 2. Разность

$Q = A - B$ . Атрибуты результирующего отношения не поименованы. Дублирующие кортежи не удваиваются. Основное формальное ограничение – одинаковая Арность отношений.

### Пример 1

Отношение A:  
Арность отношения A=3  
Кол-во кортежей =3

A	B	C
a	b	c
a	a	c
d	b	a

Отношение B:  
Арность отношения B=3  
Кол-во кортежей =4

D	E	F
a	q	w
a	b	c
a	s	d
f	g	h

Отношение Q:  
Кол-во кортежей =2

a	a	c
d	b	a

### Пример 2

Отношение A:  
Арность отношения A=4  
Кол-во кортежей =4

A	B	C	D
a	s	d	f
z	x	c	v
q	w	e	r
g	h	j	k

Отношение B:  
Арность отношения B=4  
Кол-во кортежей =5

E	F	G	I
c	v	b	n
a	s	d	f
h	j	k	l
q	w	e	r
y	u	i	o

Отношение Q:  
Арность отношения Q=4  
Кол-во кортежей =2

z	x	c	v
g	h	j	k

### 3. Декартово произведение

$Q = A * B$ . Атрибуты результирующего отношения поименованы. Мощности отношений перемножаются. Арности отношений могут быть разными.

#### Пример 1

Отношение A:  
Арность отношения A=3  
Кол-во кортежей =3  
 $N_A = 3$

A	B	C	
a	b	c	a1
a	a	c	a2
d	b	a	a3

Отношение B:  
Арность отношения B=3  
Кол-во кортежей =4  
 $N_B = 4$

D	E	F	
a	q	w	b1
a	b	c	b2
a	s	d	b3
f	g	h	b4

Отношение Q:  $N_Q = N_A * N_B = 12$

A	B	C	D	E	F
a1			b1		
a1			b2		
a1			b3		
a1			b4		
a2			b1		
a2			b2		
a2			b3		
a2			b4		
a3			b1		
a3			b2		
a3			b3		
a3			b4		

A	B	C	D	E	F
a	b	c	a	q	w
a	b	c	a	b	c
a	b	c	a	s	d
a	b	c	f	g	h
a	a	c	a	q	w
a	a	c	a	b	c
a	a	c	a	s	d
a	a	c	f	g	h
d	b	a	a	q	w
d	b	a	a	b	c
d	b	a	a	s	d
d	b	a	f	g	h

#### Пример 2

Отношение A:  
Арность отношения A=4  
Кол-во кортежей =3  
 $N_A = 3$

A	B	C	D
a	s	d	f
z	x	c	v
q	w	e	r

Отношение B:  
Арность отношения B=3  
Кол-во кортежей =5  
 $N_B = 5$

E	F	G
c	v	b
a	s	d
h	j	k
q	w	e
y	u	i

Отношение Q:  $N_Q = 15$

Арность отношения Q=7  
Кол-во кортежей =15

A	B	C	D	E	F	G
a	s	d	f	c	v	b

a	s	d	f	a	s	d
a	s	d	f	h	j	k
a	s	d	f	q	w	e
a	s	d	f	y	u	i
z	x	c	v	c	v	b
z	x	c	v	a	s	d
z	x	c	v	h	j	k
z	x	c	v	q	w	e
z	x	c	v	y	u	i
q	w	e	r	c	v	b
q	w	e	r	a	s	d
q	w	e	r	h	j	k
q	w	e	r	q	w	e
q	w	e	r	y	u	i

#### 4. Проекция (понижение арности) – «выборка по столбцам»

##### *Пример 1*

Отношение A:

Арность отношения A=3

Кол-во кортежей =3

A	B	C
a	b	c
a	a	c
d	b	a

$\pi_{B,C}(A)$  - Арность результирующего отношения = 2

B	C
b	c
a	c
b	a

##### *Пример 2*

Отношение B:

Арность отношения B=3

Кол-во кортежей =4

D	E	F
a	q	w
a	b	c
a	s	d
f	g	h

$\pi_{D,F}(B)$  - Арность результирующего отношения = 2

D	F
a	w
a	c
a	d
f	h

### Пример 3

Отношение A:

Арность отношения A=4

Кол-во кортежей =3

$$N_A = 3$$

A	B	C	D
a	s	d	f
z	x	c	v
q	w	e	r

$\pi_{1,2,4}(A)$  (то же, что и Проекция A,B,D) - Арность результирующего отношения = 3

A	B	D
a	s	f
z	x	v
q	w	r

### 5. Селекция (выбор по критерию) – «выборка по столбцам»

#### Пример 1

Отношение A:

Арность отношения A=3

Кол-во кортежей =3

A	B	C
a	b	c
a	a	c
d	b	a

Селекция  $\sigma A = a(A)$  - обозначает множество кортежей, принадлежащих отношению A, первый компонент которых (под именем A) совпадает с символом a.

Кол-во кортежей =2

A	B	C
a	b	c
a	a	c

#### Пример 2

Отношение B:

Арность отношения B=3

Кол-во кортежей =4

D	E	F
a	q	w
a	b	c
a	s	d
f	g	h

$\sigma E = q(B)$  - Кол-во кортежей =1

D	E	F
a	q	w

## Технология нормализации отношений реляционной модели данных

Одним из атрибутов отношения должен однозначно определять отдельный кортеж (т.е. по его значению можно найти один и только один кортеж). Такой атрибут называется **ключевым** или **первичным ключом** отношения. Иногда ключ может состоять из нескольких атрибутов, тогда он называется **составным**. В отношении не должно быть двух и более кортежей с одинаковыми значениями ключа. Это свойство называется **целостностью сущности**. *Ключ отношения можно определить как совокупность атрибутов, от которой функционально зависят все остальные атрибуты отношения. При этом в составной ключ должны входить только функционально независимые атрибуты.*

**Нормализация** – это пошаговый обратимый процесс разложения исходных отношений на более мелкие и простые.

В нашем примере в отношении **ВУЗЫ** на роль ключа претендует атрибут **Название**.

В примере атрибут **Регион** функционально зависит от атрибута **Город** (**Город→Регион**), поскольку в любой момент времени один город относится только к одному региону. Обратное утверждение (**Регион→Город**) неверно, поскольку одному региону можно поставить в соответствие более одного города, т.к. в регионе много городов.

### Приведение к 1 нормальной форме

Определение: Отношение находится в первой нормальной форме (1НФ), если все его атрибуты являются атомарными и соблюдается целостность сущности, т.е. имеется первичный ключ отношения.

Для приведения к 1НФ нужно:

1. Заполнить таблицы БД, максимально дублируя строки;
2. Избавиться от сложных атрибутов;
3. Определить ключ каждого отношения.

В нашем случае отношение **ВУЗЫ** можно представить в виде нижеследующей таблицы. Для облегчения процесса нормализации заполним ее, максимально дублируя данные (пункт 1). В таблице два разных ВУЗа имеют одинаковую аббревиатуру – МГТУ «СТАНКИН и МГТУ им.Баумана».

ВУЗЫ

Название ВУЗа	Город	Регион	Дисциплина	
			Предмет	Часы
МГТУ	Москва	Центр	Информатика	204
			Ин.яз.	150
УПИ	Томск	Урал	Информатика	150
МГТУ	Москва	Центр	Информатика	204

Избавимся от сложного атрибута **Дисциплина** (пункт 2), заменив его двумя атомарными - Предмет и Часы. Тогда отношение и таблица **ВУЗЫ** преобразуются к следующему виду:

ВУЗЫ

ВУЗЫ
Название
Город
Регион
Предмет
Часы

Название ВУЗа	Город	Регион	Предмет	Часы
МГТУ	Москва	Центр	Информатика	204
МГТУ	Москва	Центр	Ин.яз.	150
УПИ	Томск	Урал	Информатика	150
МГТУ	Москва	Центр	Информатика	204

В качестве ключа (пункт 3) предполагалось выбрать атрибут **Название**, но сейчас видно, что поскольку аббревиатуры ВУЗов могут повторяться, то в отношении **ВУЗЫ** необходимо добавить атрибут, однозначно идентифицирующий ВУЗ, например, атрибут **Код ВУЗа**.

√Код ВУЗа	Название ВУЗа	Город	Регион	√Предмет	Часы
1	МГТУ	Москва	Центр	Информатика	204
1	МГТУ	Москва	Центр	Ин.яз.	150
2	УПИ	Томск	Урал	Информатика	150
3	МГТУ	Москва	Центр	Информатика	204

Однако и новый атрибут не подходит в качестве первичного ключа, поскольку его значения повторяются в первых строках таблицы, а соответствующие предметы и часы – разные.

При внимательном рассмотрении можно увидеть, что не повторяется совокупность из двух атрибутов - **Код ВУЗа** и **Предмет**, т.е. в отношении **ВУЗЫ** ключ составной – **Код ВУЗа + Предмет**. Атрибут **Часы** зависит как от **Кода ВУЗа**, так и от **Предмета**, но не от каждого в отдельности. Ключевые атрибуты помечены символом √.

В итоге БД находится в 1НФ.

### Приведение ко 2 нормальной форме (2НФ)

*Определение:* Отношение находится во 2НФ, если оно находится в 1НФ и каждый неключевой атрибут функционально полно зависит от составного ключа.

Если отношение находится в 1НФ, то все неключевые атрибуты функционально зависят от ключа. Но степень формализации может быть различной.

**Частичная зависимость** – если неключевой атрибут зависит от части ключа.

**Полная функциональная зависимость** – когда неключевой атрибут зависит от всего ключа целиком.

В нашем примере **Название** зависит только от **Кода ВУЗа**, т.е. налицо частичная зависимость. Полная зависимость наблюдается между ключом **Код ВУЗа+Предмет** и атрибутом **Часы**.

Такая частичная зависимость приводит к следующим аномалиям:

- Дублирование данных о ВУЗе (поскольку ВУЗ обучает нескольким предметам);
- Проблема контроля из-за избыточности данных (изменение, например, названия ВУЗа влечет за собой необходимость поиска изменения этого названия во всех corteжах с данным ВУЗом).

Поэтому нужно устранить частичную зависимость неключевых атрибутов от ключа. Для этого необходимо:

4. указать письменно частичные зависимости неключевых атрибутов от ключа;
5. разложить исходное отношение на два:
  - построить отношение, в которое входит часть составного ключа и атрибуты, зависящие только от этой части. Повторяющиеся corteжи нужно удалить;
  - построить отношение, в которое входят полный ключ и оставшиеся атрибуты.
6. Указать ключи в каждом из получившихся отношений и установить от первого (главного – со стороны *один*) отношения ко второму (подчиненному – со стороны *многие*) связи **1:М** по внешнему ключу, которым является часть составного ключа во втором отношении.
7. *Еще раз рассмотреть каждое из отношений БД на наличие частичных зависимостей и проделать заново пункты 4-6, если такие найдены.*

*Определение:* **Внешним ключом** является один из неключевых атрибутов отношения, который в другом отношении является первичным ключом.

В нашем примере нужно письменно указать, от чего зависит каждый неключевой атрибут (пункт 4). Сделаем это.

**Код ВУЗа→Название, Код ВУЗа→Город, Код ВУЗа→Регион** (частичные зависимости). **Код ВУЗа →Предмет→Часы** (полная зависимость).

Разбиваем отношение **ВУЗЫ** на два (пункт 5). В первое входят часть составного ключа **Код ВУЗа** и атрибуты, зависящие только от этой части, во второе – полный ключ **Код ВУЗа+Предмет** и оставшиеся, т.е. **Часы**. Устанавливаем связь **1:M** между главным отношением **ВУЗЫ** и подчиненным отношением **Дисциплины** через внешний ключ **Код ВУЗа** (пункт 6).

ВУЗЫ	1	∞	Дисциплины
Код ВУЗа√			Код ВУЗа√
Название			Предмет√
Город			Часы
Регион			

В получившейся таблице убираем дублирующиеся строки и указываем ключи получившихся отношений.

ВУЗЫ

√Код ВУЗа	Название ВУЗа	Город	Регион
1	МГТУ	Москва	Центр
1	МГТУ	Москва	Центр
2	УПИ	Томск	Урал
3	МГТУ	Москва	Центр

Дисциплины

√Код ВУЗа	√Предмет	Часы
1	Информатика	204
1	Ин.яз.	150
2	Информатика	150
3	Информатика	204

Поскольку в первом отношении ключ простой, то частичных зависимостей в нем быть не может (пункт 7). Поскольку единственный неключевой атрибут во втором отношении зависит от полного ключа, то в нем тоже нет частичных зависимостей. В итоге БД находится во 2 НФ.

Определение: **Ссылочной целостностью** называется такое состояние Базы данных, при котором каждому значению внешнего ключа подчиненного отношения соответствует кортеж другого (главного) отношения с таким же значением первичного ключа. Обратного соответствия не требуется.

В нашем примере каждому значению внешнего ключа **Код ВУЗа** в подчиненном отношении **Дисциплины** (и значению 1, и 2, и 3) соответствует такое же значение этого атрибута в главном отношении **ВУЗЫ**. Если бы в подчиненном отношении появился кортеж со значением атрибута **Код ВУЗа** равным, например, 5, то ссылочная целостность была бы нарушена, поскольку ВУЗа с таким кодом в главном отношении **ВУЗЫ** не существует (т.е. дисциплина в ВУЗе есть, а самого ВУЗа нет!).

Наоборот, присутствие ВУЗа с кодом 5 в списке ВУЗов (главном отношении) при отсутствии его в списке дисциплин (подчиненном отношении) не нарушает ссылочной целостности (ВУЗ есть, но еще ничего не преподает).

Приведение к 3 нормальной форме



Определение: Отношение находится в 3НФ, если оно находится во 2НФ и в нем отсутствуют транзитивные зависимости неключевых атрибутов от ключа.

В полученном отношении **ВУЗЫ** имеются следующие аномалии:

- Дублированные информации о регионе для ВУЗов одного города;
- Проблема контроля из-за избыточности данных, поскольку изменение названия региона, в котором находится город, влечет за собой необходимость поиска и изменения этого названия у всех ВУЗов этого города;
- Нельзя включить данные о новом городе, если на данный момент в нем еще отсутствуют ВУЗы и, наоборот, при упразднении всех ВУЗов в городе данные о нем нельзя сохранить.

Такие аномалии связаны с тем, что в отношении имеется транзитивная зависимость атрибута **Регион** от ключа **Код ВУЗа**.

*Говорят, что атрибут **С** зависит от атрибута **А** транзитивно, если есть атрибут-посредник **В** такой, для которого выполняются зависимости  $A \rightarrow B$  и  $B \rightarrow C$ , и при этом обратные зависимости отсутствуют. Такую зависимость можно изобразить следующим образом:  $A \rightarrow B \rightarrow C$ .*

Необходимо удалить все транзитивные зависимости неключевых атрибутов от ключа:

8. Рассмотреть каждую из таблиц на наличие транзитивных зависимостей неключевых атрибутов от ключа. Указать письменно такие зависимости.
9. Если транзитивных зависимостей нет, доказать это письменно;
10. Разложить исходное отношение на два:
  - построить отношение, в которое входит зависимый атрибут и атрибут-посредник. Повторяющиеся кортежи нужно удалить;
  - построить отношение, в которое входят все атрибуты, кроме зависимого;
11. Указать ключи в каждом из получившихся отношений и установить от первого (главного) отношения ко второму (подчиненному) связь 1:М по внешнему ключу, которым является посредник во втором отношении.
12. Еще раз рассмотреть каждое из отношений на наличие транзитивных зависимостей и проделать заново пункты 8-11, если такие найдены.

Отметим сразу, что атрибут-посредник не может быть частью составного ключа рассматриваемого отношения, т.к. часть ключа всегда является независимой от других атрибутов. Поэтому для транзитивной зависимости от ключа всегда необходимо наличие двух неключевых атрибутов – зависимого и посредника.

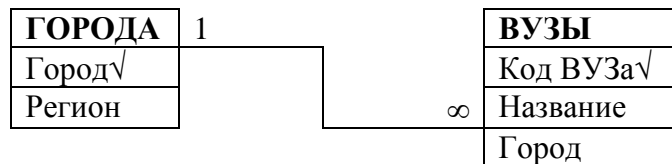
В нашем примере нужно письменно указать, какие варианты транзитивных зависимостей неключевых атрибутов от ключей возможны (пункт 8).

Сразу отметим, что в отношении **Дисциплины** не может быть транзитивных зависимостей, потому что в нем только один неключевой атрибут. Отметим также, что в отношении **ВУЗЫ** атрибут **Название** не может быть посредником, поскольку мы с самого начала условились, что аббревиатуры ВУЗов могут повторяться. Из этого следует, что утверждения **Название**→**Город** и **Название**→**Регион** неверны, поскольку одной аббревиатуре ВУЗа могут соответствовать несколько городов и регионов. Остаются возможные варианты:

**Код ВУЗа**→**Регион**→**Город**,

**Код ВУЗа**→**Город**→**Регион**.

Первое утверждение неверно, поскольку в одном регионе может находиться много городов, а второе будет истинным, если мы будем считать, что городов с одним названием в разных регионах нет. Для избавления от этой зависимости (пункт 10) разложим исходное отношение на два (в первое войдут зависимый атрибут **Регион** и посредник **Город**, а второе - все, кроме зависимого). Устанавливаем связь **1:М** через внешний ключ **Город** (пункт 11).



В получившейся таблице уберем дублирующие строки и укажем ключи получившихся отношений.

ГОРОДА

ВУЗЫ

✓Город	Регион
Москва	Центр
Томск	Урал
<del>Москва</del>	<del>Центр</del>

✓Код ВУЗа	Название	Город
1	МГТУ	Москва
2	УПИ	Томск
3	МГТУ	Москва

Еще раз рассмотрим получившиеся отношения (пункт 12). В отношении ГОРОДА транзитивных зависимостей быть не может, поскольку в нем только один неключевой атрибут. Во втором отношении (ВУЗЫ) транзитивная зависимость будет иметь место, если выполнится одно из утверждений:

**Код ВУЗа → Название → Город,**

**Код ВУЗа → Город → Название.**

Первое утверждение неверно, поскольку мы уже выяснили, что атрибут Город не зависит от аббревиатуры ВУЗа. Второе утверждение с одной аббревиатурой, как это и произошло в нашем примере. Получаем, что транзитивных зависимостей в отношении ВУЗЫ нет.

В итоге БД, состоящая из трех связанных отношений Дисциплины, ВУЗЫ, и ГОРОДА, находится в 3НФ.