

# Реляционная алгебра

Курс «Базы данных», лекция 5

## Реляционная модель данных

---

Реляционная модель данных была предложена в 1970-м году [Эдгаром Коддом](#).

Реляционная модель данных единственная из моделей данных для которой разработан теоретический базис: **РЕЛЯЦИОННАЯ АЛГЕБРА**

Базовой структурой реляционной модели данных является **ОТНОШЕНИЕ**.

**Сущность** — объект реального мира, о котором нужно хранить информацию.

## Отношение

---

Сущность: СТУДЕНТ ГРУППЫ

№ зачётной книжки	Фамилия	Имя	Дата рождения	Группа
15345	Иванов	Иван	15.04.2004	392
15256	Медведева	Анна	13.02.2003	591
15438	Петров	Михаил	12.11.2004	392
15123	Сидоров	Александр	26.01.2003	392
15348	Санин	Михаил	20.10.2004	591

## Отношение

---

**Домен** – **множество** значений, которые может принимать одна из характеристик сущности.

Декартово произведение множеств – множество всех возможных комбинаций элементов исходных множеств.

ПРИМЕР:

$D1 = \{1, 2\}; D2 = \{A, B, C\}$

$D1 \times D2 = \{(1, A), (1, B), (1, C), (2, A), (2, B), (2, C)\}$

**ОТНОШЕНИЕ** – подмножество декартова произведения доменов.

## Отношение

---

Домен Фамилия – 5 элементов

Домен Имя – 4 элемента

Домен Дата рождения – 5 элементов

Домен Группа – 2 элемента

Всего:  $5 \times 4 \times 5 \times 2 = 200$  комбинаций.

Отношение Студенты групп 392, 591 включает 5 комбинаций.

**КОРТЕЖ** – элемент отношения. Каждый кортеж соответствует описанию реализации одной сущности определённого типа.

**АТТРИБУТ** – элемент кортежа.

**МОЩНОСТЬ ОТНОШЕНИЯ** – количество кортежей в отношении.

**АРНОСТЬ** – количество атрибутов в кортеже.

## Свойства отношений

---

1. В отношении не должно быть одинаковых кортежей.
2. Порядок кортежей в отношении несущественен.

Атрибут определён на некотором домене. Домен задаётся типом данных, размером и ограничениями целостности. **NULL** может принадлежать любому типу данных.

**СХЕМА ОТНОШЕНИЯ** – перечень атрибутов с указанием домена, на котором они заданы.

**КЛЮЧ** – атрибут (группа атрибутов), значения которого классифицируют или идентифицируют кортеж.

**ПЕРВИЧНЫЙ КЛЮЧ** – ключ с уникальным значением, отличным от NULL.

## Связь между отношениями

---

**Внешний ключ** – атрибут отношения, значения которого ссылаются на атрибут со свойством UNIQUE в другой таблице. Атрибуты, задающие связь по внешнему ключу должны быть определены на одном и том же домене. Внешний ключ может состоять из нескольких атрибутов (составной внешний ключ).

### Зачётная ведомость

№ зачётной книжки	Предмет	Оценка	Номер записи
15345	Линейная алгебра	4	1
15256	Иностранный язык	5	2
15438	Экономика	5	3
15345	Иностранный язык	5	4

## Внешний ключ

---

Внешний ключ является **ограничением целостности**.

Ограничение целостности по внешнему ключу проверяется:

- при добавлении записи в подчинённую таблицу
- при удалении записи из главной таблицы



## Связи между сущностями

---

**Размерность связи** – количество сущностей, участвующих в связи.

Основные размерности: унарные, двойные, тройные.

Связи бывают полными и частичными.

Связь называется полной, если каждому экземпляру сущности одного отношения соответствует хотя бы один экземпляр сущности другого отношения.

**Мощность связи** - максимальное число экземпляров одной сущности, которое может быть связано с экземпляром другой сущности.

Основные разновидности мощности связи:

- один к одному;
- один ко многим;
- многие ко многим

## Связи между сущностями

---

Связи «**один к одному**» между сущностями X и Y — это такие связи, при которых каждый экземпляр сущности X может быть связан только с одним экземпляром сущности Y.

Связь «**один ко многим**»: один экземпляр сущности связан с нулевым числом, с одним или множеством экземпляров другой сущности.

Связь «**многие ко многим**» в реляционной модели реализовать невозможно.

## Замкнутость реляционной алгебры

---

- Реляционная алгебра представляет собой набор таких операций над отношениями, что результат каждой из операций также является отношением. Это свойство алгебры называется **замкнутостью**.
- Реляционный оператор  $f$  выглядит как функция с реляционными отношениями в качестве аргументов:  $R = f(R1, R2, \dots, Rn)$ .
- В реляционных выражениях *можно использовать вложенные выражения* сколь угодно сложной структуры.

# Типы операций реляционной алгебры

---

## **Теоретико-множественные операции реляционной алгебры:**

- объединение (union)
- пересечение (intersection)
- разность (set difference)
- декартово произведение (cartesian product)

## **Специальные реляционные операции:**

- выборка (selection)
- проекция (projection)
- соединение (join)
- деление (division)

## Совместимость отношений

---

Два отношения являются **совместимыми по типу**, если они имеют идентичные заголовки:

- множества имён атрибутов этих отношений совпадают
- атрибуты с одинаковыми именами определены на одном и том же домене.

Для приведения отношений к одному типу следует использовать операцию переименования.

## Объединение

Результатом операции **объединения** двух совместимых по типу отношений  $R1$  и  $R2$  является отношение с тем же заголовком, и телом, состоящим из кортежей, принадлежащих  $R1$  или  $R2$  или обоим отношениям.

В SQL это операция **union**:  
 $R1 \text{ union } R2$

$R1 \cup R2$

Название	Размер	Цена
Болт	M5	10
Гайка	M7	7
Шуруп	P21	9
Шайба	M5	3

$R1$

Название	Размер	Цена
Болт	M5	10
Гайка	M7	7
Шайба	M5	3

$R2$

Название	Размер	Цена
Болт	M5	10
Гайка	M7	7
Шуруп	P21	9

## Пересечение

Результатом операции *пересечения* двух совместимых по типу отношений *R1* и *R2* является отношение с тем же заголовком, и телом, состоящим из кортежей, принадлежащих *R1* и *R2* *одновременно*.

В SQL это операция **intersect**:  
R1 intersect R2

R1 ∩ R2

Название	Размер	Цена
Болт	M5	10
Гайка	M7	7

R1

Название	Размер	Цена
Болт	M5	10
Гайка	M7	7
Шайба	M5	3

R2

Название	Размер	Цена
Болт	M5	10
Гайка	M7	7
Шуруп	P21	9

## Разность

Результатом операции **разность** двух совместимых по типу отношений  $R1$  и  $R2$  является отношение с тем же заголовком, и телом, состоящим из кортежей, принадлежащих  $R1$  и не принадлежащих  $R2$ .

В SQL это операция **except**:  
 $R1$  except  $R2$

$R1 - R2$

Название	Размер	Цена
Шайба	M5	3

$R2 - R1$

Название	Размер	Цена
Шуруп	P21	9

$R1$

Название	Размер	Цена
Болт	M5	10
Гайка	M7	7
Шайба	M5	3

$R2$

Название	Размер	Цена
Болт	M5	10
Гайка	M7	7
Шуруп	P21	9



## Декартово произведение

Результатом операции *декартова произведения* двух отношений  $R1$  и  $R2$ , не имеющих общих имен атрибутов, является отношение с заголовком, который представляет собой сцепление заголовков  $R1$  и  $R2$ , и телом, состоящим из всех возможных кортежей, каждый из которых представляет собой сцепление кортежа из  $R1$  и кортежа из  $R2$ .

В SQL это операция, которая производится над таблицами из секции **from** оператора **select**:

```
select * from R1, R2;
```

## Декартово произведение

R1

Название	Размер	Цена
Болт	M5	10
Гайка	M7	7
Шайба	M5	3

R2

Поставщик	Страна
Метиз23	Россия
GoodMetall	China

R1 x R2

Название	Размер	Цена	Поставщик	Страна
Болт	M5	10	Метиз23	Россия
Гайка	M7	7	Метиз23	Россия
Шайба	M5	3	Метиз23	Россия
Болт	M5	10	GoodMetall	China
Гайка	M7	7	GoodMetall	China
Шайба	M5	3	GoodMetall	China

# Выборка

---

## Операция выборки:

Работает с одним отношением **R**.

**Определяет** результирующее отношение с тем же заголовком, что и отношение **R**, и телом, состоящим из кортежей, значения атрибутов которых при подстановке в условие (предикат) дают значение истина.

$X \theta Y$  - условие (предикат),

$\theta$  – один из операторов сравнения ( $=$ ,  $\neq$ ,  $<$ ,  $>$  и т.д.),

**X** и **Y** - атрибуты отношения **R** или скалярные значения.

Синтаксис операции выборки:

**R where  $\sigma$**  , или **R where ( $X \theta Y$ )**

Синтаксис на языке SQL:

**select \* from R where ( $X \theta Y$ )**

# Выборка

## Отношение R

№ зачётной книжки	Предмет	Оценка	Номер записи
15345	Линейная алгебра	4	1
15256	Иностранный язык	5	2
15438	Экономика	3	3

Результат выборки: R where Оценка < 5

№ зачётной книжки	Предмет	Оценка	Номер записи
15345	Линейная алгебра	4	1
15438	Экономика	3	3

## Проекция

---

**Проекция** выполняется на одном отношении  $R$ , которое имеет атрибуты  $X, Y, \dots, Z$  (и, возможно, другие атрибуты) и определяет отношение, которое содержит вертикальное подмножество исходного отношения  $R$ , извлекая значения указанных атрибутов и отбрасывая дубликаты полученных кортежей.

**Проекция** отношения  $R$  по атрибутам  $X, Y, \dots, Z$  определяется с помощью следующего выражения:  $\mathbf{R} \{ X, Y, \dots, Z \}$ .

Заголовок результата формируется из заголовка отношения  $R$  путем удаления всех атрибутов, не указанных в множестве  $\{ X, Y, \dots, Z \}$ .

Тело формируется из оставшихся кортежей после удаления дубликатов

# Проекция

Отношение R

Название	Размер	Цена	Поставщик	Страна
Болт	M9	15	Метиз23	Россия
Болт	M5	10	Метиз23	Россия
Гайка	M7	7	Метиз23	Россия
Шайба	M5	3	Метиз23	Россия
Болт	M5	10	GoodMetall	China
Гайка	M7	7	GoodMetall	China
Гайка	M11	8	GoodMetall	China
Шайба	M5	3	GoodMetall	China

R{Название, Поставщик, Страна}

Название	Поставщик	Страна
Болт	Метиз23	Россия
Гайка	Метиз23	Россия
Шайба	Метиз23	Россия
Болт	GoodMetall	China
Гайка	GoodMetall	China
Шайба	GoodMetall	China

## Соединение

---

### Операция соединения –

*комбинация декартового произведения и выборки.*

Эквивалентна операции выборки из декартового произведения двух отношений тех кортежей, которые удовлетворяют условию, указанному в предикате соединения в качестве формулы выборки.

**Соединением** отношений **R** и **S** по условию **F** выражается: **(R times S) where F**

Синтаксис на языке SQL:

```
Select R.*, S.* from R, S where F
```

# Соединение

---

## Типы операций соединения:

- тета-соединение
- соединение по эквивалентности (частный случай тета-соединения)
- естественное соединение
- внешнее соединение
- полу-соединение



# Соединение

---

## Тета-соединение

*определяет отношение, которое содержит кортежи из декартового произведения отношений **R** и **S**, удовлетворяющие предикату **F**.*

Предикат **F** имеет вид  $R.a_i \theta S.b_i$ , где  $\theta$  - один из операторов сравнения ( $<$ ,  $<=$ ,  $>$ ,  $>=$ ,  $=$  или  $\neq$ ).

$\theta$  -соединением отношения **R** по атрибуту **X** с отношением **S** по атрибуту **Y** называют отношение **(R times S) where (X  $\theta$  Y)**.

Синтаксис на языке SQL:

**Select \* from R, S where (R.X  $\theta$  S.Y)**

## Соединение

### Отношение R (преподаватели)

Табельны йномер	Фамилия	Статус преподавателя
4587	Бондаренко	4
2136	Воронин	1
5496	Анисимова	2

*Какие предметы имеет право преподавать каждый из преподавателей ?*

**(R time S) where (R.статус  $\geq$  S. статус)**

### Отношение S (предметы)

Код предмета	Название	Статус предмета
101	История	3
102	География	2
103	Философия	1

## Соединение

### Декартово произведение R x S

Табельный номер	Фамилия	Статус преподавателя	Код предмета	Название	Статус предмета
4587	Бондаренко	4	101	История	3
4587	Бондаренко	4	102	География	2
4587	Бондаренко	4	103	Философия	1
2136	Воронин	1	101	История	3
2136	Воронин	1	102	География	2
2136	Воронин	1	103	Философия	1
5496	Анисимова	2	101	История	3
5496	Анисимова	2	102	География	2
5496	Анисимова	2	103	Философия	1

## Соединение

---

$\theta$  - соединение

Табельный номер	Фамилия	Статус преподавателя	Код предмета	Название	Статус предмета
4587	Бондаренко	4	101	История	3
4587	Бондаренко	4	102	География	2
4587	Бондаренко	4	103	Философия	1
2136	Воронин	1	103	Философия	1
5496	Анисимова	2	102	География	2
5496	Анисимова	2	103	Философия	1

## Соединение

---

**Экви-соединение (соединение по эквивалентности)** – частный случай  $\theta$  - соединения, когда  $\theta$  может быть только равенством.

Синтаксис экви-соединения:

**$R[X=Y]S$**

Синтаксис на языке SQL:

**select R.\*, S .\* from R, S where (R.X = S.Y)**

# Соединение

---

## Естественное соединение

соединение по эквивалентности двух отношений  $R$  и  $S$ , выполненное по всем общим атрибутам, из результатов которого исключается по одному экземпляру каждого общего атрибута.

Синтаксис естественного соединения:  
 $R \text{ join } S$ .

### **ВАЖНО!**

Если отношения  $A$  и  $B$  не имеют общих имен атрибутов, то естественное соединение превращается в декартово произведение.

# Соединение

## Отношение R

A	B
a1	b1
a2	b2

## Отношение S

B	C
b1	c1
b2	c2
b3	c3
b1	c4

## R x S

A	B	B	C
a1	b1	b1	c1
a1	b1	b2	c2
a1	b1	b3	c3
a1	b1	b1	c4
a2	b2	b1	c1
a2	b2	b2	c2
a2	b2	b3	c3
a2	b2	b1	c4

## R [X = Y] S

A	B	C
a1	b1	c1
a1	b1	c4
a2	b2	c2

# Соединение

---

## Внешние соединения

**Левым соединением** отношений  $R1$  и  $R2$  называется естественное соединение котором:

если по одноимённым атрибутам кортежу из  $R1$  не соответствует ни одного кортежа из  $R2$ , то в результат добавляется этот кортеж из  $R1$ , дополненный пустыми значениями.

**Правым соединением** отношений  $R1$  и  $R2$  называется естественное соединение в котором:

если по одноимённым атрибутам кортежу из  $R2$  не соответствует ни одного кортежа из  $R1$ , то в результат добавляется этот кортеж из  $R2$ , дополненный пустыми значениями.

**Внешнее соединение** – объединение правил исключительных ситуаций для левого и правого соединений.



## Соединение

### Отношение R

ID	Имя
1	Иван
2	Пётр
3	Сидор

### Отношение S

ID	Фамилия
1	Иванов
1	Петров
3	Сидоров
4	Савельев

### Левое соединение

ID	Имя	Фамилия
1	Иван	Иванов
1	Иван	Петров
2	Пётр	
3	Сидор	Сидоров

### Правое соединение

ID	Имя	Фамилия
1	Иван	Иванов
1	Иван	Петров
3	Сидор	Сидоров
4		Савельев

### Внешнее соединение

ID	Имя	Фамилия
1	Иван	Иванов
1	Иван	Петров
2	Пётр	
3	Сидор	Сидоров
		Савельев

## Соединение

---

### Полу-соединение

выполняет естественное соединение отношений  $R$  и  $S$ , с последующим выполнением проекции по атрибутам отношения  $R$ .

## Деление

---

Пусть даны:

отношение  $A$  с заголовком  $XUY$ ;

отношение  $B$  с заголовком  $Y$ .

**Делением**  $A$  на  $B$  называется максимальное по включению отношение  $C$  с заголовком  $X$ , такое что  $B \times C \subseteq A$ .

Обозначение:  $A \div B$

# Деление

R			
A1	A2	A3	A4
2	S3	4	sun
3	X8	7	kab
3	R4	8	kab

S	
A2	A3
R4	8
X8	7



?			
A1	A2	A3	A4
2	S3	4	sun
3	X8	7	kab
3	R4	8	kab



?			
A1	A2	A3	A4
3	X8	7	kab
3	R4	8	kab



?	
A1	A4
3	kab



?	
A1	A4
3	kab
3	kab

## Реляционное исчисление кортежей

---

В реляционном исчислении кортежей задача состоит в нахождении таких кортежей, для которых **предикат является истинным**.

Это исчисление основано на **кортежных переменных**.

Значение кортежной переменной - один из кортежей отношения.

Например, факт того, что отношение Staff, используется в качестве области определения кортежной переменной S записывается в виде: Staff ( S )

**Запрос** «найти множество всех кортежей S, для которых F(S) является истинным»

записывается следующим образом: { S | F( S ) }

Здесь предикат F называется формулой (в математической логике, правильно построенной формулой — Well-Formed Formula, или сокращенно WFF).

Например, запрос «Создайте список всех менеджеров, зарплата которых превышает 25000 рублей» записывается следующим образом: { S.fName, S.lname | Staff( S ) ∧ S.position = 'Manager' ∧ S.salary > 25000 }

## Реляционное исчисление кортежей

---

Для указания **количества экземпляров**, к которым должен быть применен предикат, в формулах могут использоваться два типа кванторов:

**Квантор существования**  $\exists$  («существует») используется в формуле, которая должна быть истинной хотя бы для одного экземпляра, например:  $\text{Staff}(S) \wedge (\exists B) (\text{Branch}(B) \wedge (B.\text{branchNo} = S.\text{branchNo}) \wedge B.\text{city} = \text{'London'})$

**Квантор всеобщности**  $\forall$ , или символ «для всех», используется в выражениях, которые относятся ко всем экземплярам, например:  $(\forall B) (B.\text{city} \neq \text{'Paris'})$

Кортежные переменные называются **свободными переменными**, если они не квалифицируются кванторами  $\forall$  или  $\exists$ ; в противном случае они называются **связанными переменными**.

## Реляционное исчисление доменов

---

В реляционном **исчислении доменов** значения переменные берутся из области определения атрибутов.

Выражение в реляционном исчислении доменов имеет форму:  $\{ d_1, d_2, \dots, d_n \mid F(d_1, d_2, \dots, d_m) \}$

Здесь:

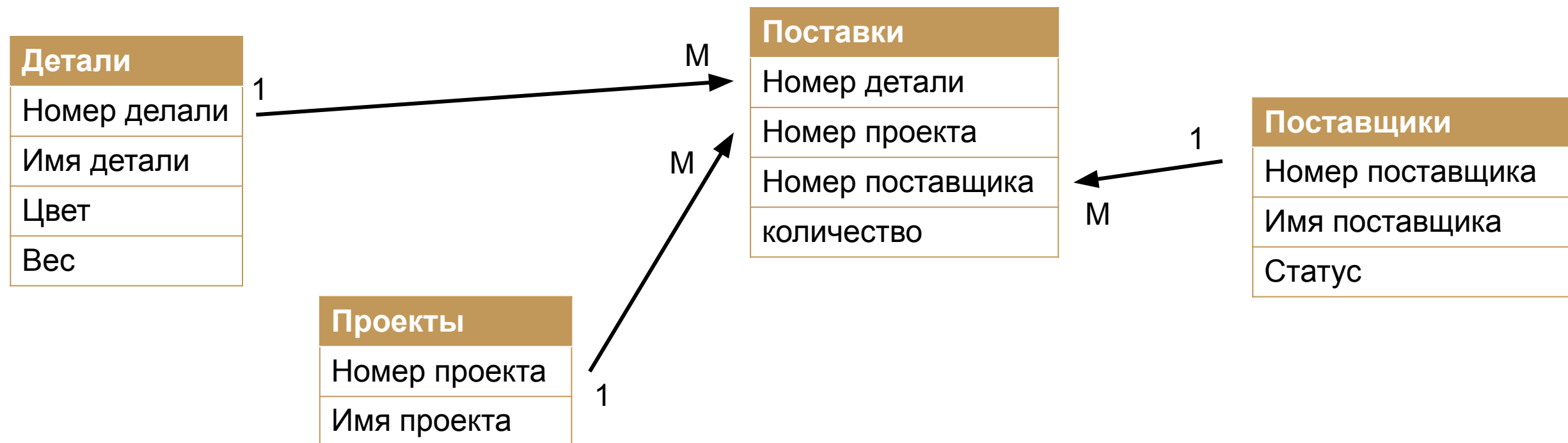
- $d_1, d_2, \dots, d_n$  — переменные области определения домена;
- $F(d_1, d_2, \dots, d_m)$  — формула.

Пример:

Найдите имена всех менеджеров, зарплата которых равна 25000 рублей:  $\{ fN, IN \mid \exists (sN, posn, sex, DOB, sal, bN) ( Staff(sN, fN, IN, posn, sex, DOB, sal, bN) \wedge posn = 'Manager' \wedge sal = 25000 ) \}$

•

# Примеры





## Примеры

---

1) Получить имена поставщиков, которые поставляют деталь Д2.

Требуется построить таблицу с одним полем **Имя поставика**, используя таблицы **Поставки** и **Поставщики**:

((Поставки JOIN Поставщики) WHERE Д№='Д2') [Имя\_П];

2) Получить имена поставщиков, которые поставляют по крайней мере одну черную деталь.

Здесь также надо построить унарную таблицу с атрибутом **Имя\_П**, но необходимые нам сведения находятся в трех таблицах – **Детали**, **Поставки** и **Поставщики**:

((Детали WHERE Цв = 'Черный') JOIN Поставки) JOIN Поставщики [Имя\_П]

или

((Детали WHERE Цв = 'Черный') [Д№] JOIN Поставки) [П№] JOIN Поставщики [Имя\_П].

*Во втором выражении предлагается делать проекции после операций выборки и соединения, уменьшая при этом степень промежуточных таблиц. Благодаря этому повышается эффективность алгоритма.*

## Примеры

---

3) Получить имена поставщиков, которые поставляют все детали.

Наиболее простое решение этой задачи получаем при использовании реляционной операции деления:

```
((Поставки [П№,Д№] DIVIDED BY Детали [Д№]) JOIN Поставщики) [Имя_П]
```

4) Получить номера поставщиков, которые поставляют по крайней мере все те детали, которые поставляет поставщик П2.

Алгоритм решения похож на алгоритм прошлой задачи:

```
Поставки [П№,Д№] DIVIDEDBY ((Поставки WHERE Имя_П='П2') [Д№])
```

5) Получить имена поставщиков, которые не поставляют деталь Д2.

В этой задаче находим номера поставщиков, поставляющих деталь д2, и вычитаем их из таблицы, содержащей все номера поставщиков:

```
((поставщики [п№] MINUS (поставки WHERE д№='д2') [п№]) JOIN (поставщики) [имя_п])
```

## Примеры

---

б) Получить номера поставщиков, у которых максимальное значение поставок как минимум в 2 раза превышает среднее значение поставок деталей остальными поставщиками.

Требуется построить унарную таблицу с атрибутом П№. Вся необходимая информация находится в таблице Поставки. Выбрать нужных поставщиков мы сможем, если построим таблицу, содержащую для каждого поставщика данные о величине его максимальной поставки и о средних поставках всех остальных. Тогда операцией выборки сможем отобрать поставщиков, у которых первая величина более чем в 2 раза превышает вторую.

На первом шаге строим таблицу с данными о максимальных поставках каждого поставщика:  
SUMMARIZE поставки BY(П№) ADD MAX(Кол) AS m

Получили бинарную таблицу с атрибутами П№ и m. Далее добавим в эту таблицу новый атрибут, в котором для каждого поставщика будут находиться значения средних поставок всех остальных поставщиков. Для этого удобно использовать оператор расширения EXTEND:

---

```
EXTEND (SUMMARIZE поставки BY(п№) ADD MAX(Кол) AS m) ADD AVG(((поставки RENAME П№ AS x)
WHERE x<>П№) [кол]) AS ср
```

Далее в полученной таблице выбираем строки, в которых  $m \geq 2 * ср$  и делаем проекцию по полю п№:

```
(EXTEND (SUMMARIZE поставки BY(п№) ADD MAX(кол) AS m) ADD AVG(((поставки RENAME п№ AS x)
WHERE x<>п№) [кол]) AS ср) WHERE m >= 2 * ср)[п№]
```