

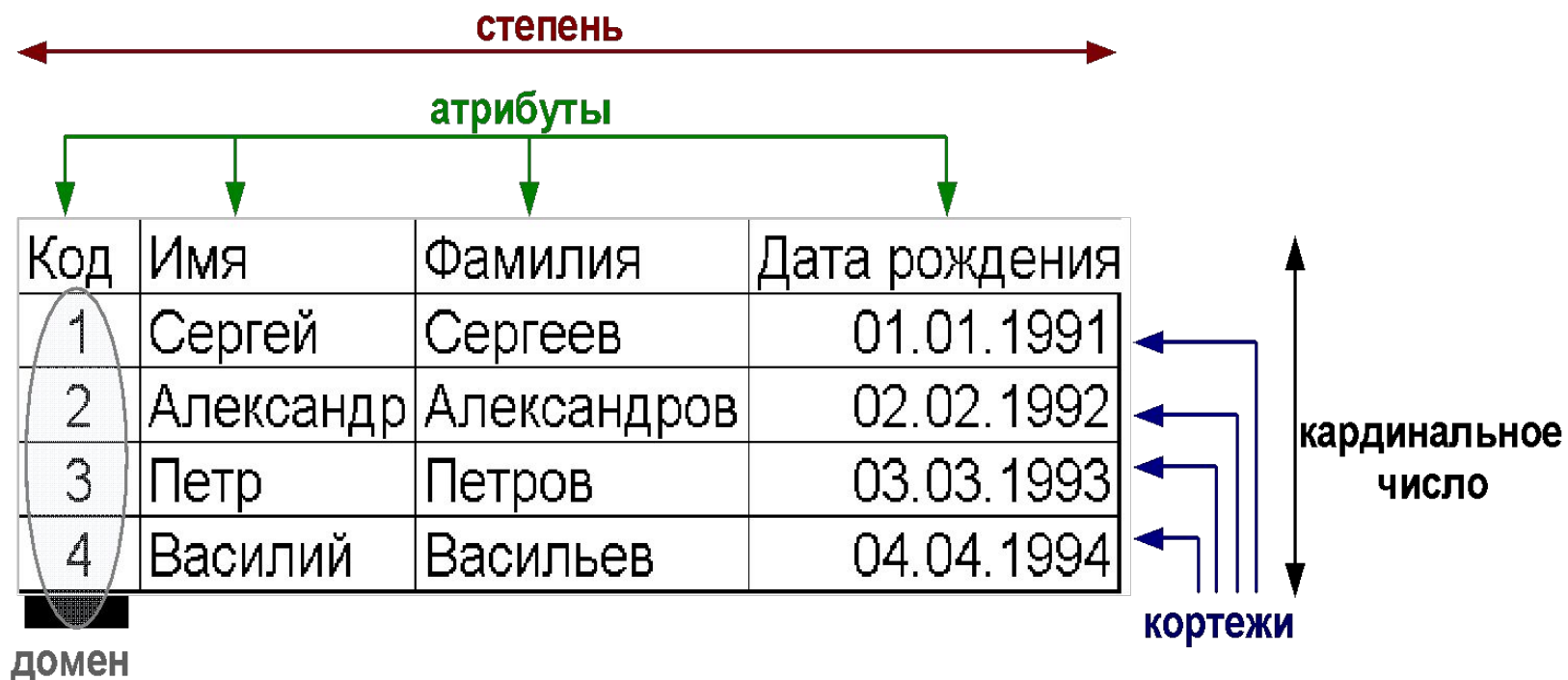
Основы реляционной алгебры

Появление теоретико-множественных моделей в системах баз данных было предопределено настоятельной потребностью пользователей в переходе от работы с элементами данных, как это делается в графовых моделях, к работе с некоторыми макрообъектами. Основной моделью в этом классе является реляционная модель данных.

Теоретической основой этой модели стала теория отношений, основу которой заложили два логика — американец Чарльз Содерс Пирс (1839-1914) и немец Эрнст Шредер (1841-1902).

Американский математик Эдгар Франк Кодд в 1970 году впервые сформулировал основные понятия и ограничения реляционной модели, ограничив набор операций в ней семью основными и одной дополнительной операцией. Предложения Кодда были настолько эффективны для систем баз данных, что за эту модель он был удостоен престижной премии Тьюринга в области теоретических основ вычислительной техники.

Основной структурой данных в модели является отношение, именно поэтому модель получила название **реляционной** (от английского *relation* — отношение).



Домен – это общее множество значений, из которых берутся конкретные значения определенного атрибута некоторого отношения (D)

Кортеж соответствует экземпляру записи

Степень (n – арность отношения)– количество атрибутов

Кардинальное число (кардинальность отношения) – количество кортежей содержащихся в отношении

Набор атрибутов, однозначно определяющий каждый объект, называют **КЛЮЧОМ**.

Для таблицы должны выполняться следующие правила:

- таблица имеет имя, отличное от имен других таблиц;
- каждая клетка таблицы содержит только атомарное значение;
- каждый столбец таблицы имеет уникальное имя;
- данные для столбца берутся из одного множества значений (домен);
- порядок следования столбцов не имеет значения;
- таблица не имеет повторяющихся строк;
- строки не имеют имен;
- порядок следования строк не имеет значения.

Если атрибуты принимают значения из одного и того же домена, то они называются *Q-сравнимыми*, где Q множество допустимых операций сравнения, заданных для данного домена.

Например, если домен содержит числовые данные, то для него допустимы все операции сравнения, тогда $Q = \{=, <>, >=, <=, <, >\}$. Однако и для доменов, содержащих символьные данные, могут быть заданы не только операции сравнения по равенству и неравенству значений. Если для данного домена задано лексикографическое упорядочение, то он имеет также полный спектр операций сравнения.

Схемы двух отношений называются ***эквивалентными***, если они имеют одинаковую степень и возможно такое упорядочение имен атрибутов в схемах, что на одинаковых местах будут находиться сравнимые атрибуты, то есть атрибуты, принимающие значения из одного домена.

В отличие от теоретико-графовых моделей в реляционной модели связи между отношениями поддерживаются неявным образом.

В каждой связи одно отношение может выступать как основное, а другое отношение выступает в роли подчиненного. Это означает, что один кортеж основного отношения может быть связан с несколькими кортежами подчиненного отношения. Для поддержки этих связей оба отношения должны содержать наборы атрибутов, по которым они связаны. В основном отношении это первичный ключ отношения (PRIMARY KEY), который однозначно определяет кортеж основного отношения. В подчиненном отношении для моделирования связи должен присутствовать набор атрибутов, соответствующий первичному ключу основного отношения. Данный набор атрибутов в подчиненном отношении принято называть внешним ключом (FOREIGN KEY).

Связи между отношениями

Любая таблица имеет один или несколько столбцов, значения которых однозначно идентифицируют каждую ее строку. Такой столбец (или совокупность столбцов) называется первичным ключом.

Взаимосвязи таблиц в реляционной модели поддерживаются внешними ключами. Внешний ключ – это столбец (или совокупность столбцов), значения которого однозначно характеризуют сущности, представленные строками некоторого другого отношения, т.е. задают значения их первичного ключа.

Реляционная алгебра – это набор операций, которые принимают отношения в качестве операндов и возвращают отношение в качестве результата.

Алгеброй называется множество объектов с заданной на нем совокупностью операции, замкнутых относительно этого множества, называемого *основным множеством*.

Основным множеством в реляционном алгебре является множество отношений.

Всего Коддом было предложено 8 операций.

В общем это множество избыточное, так как одни операции могут быть представлены через другие, однако множество операций выбрано из соображений максимального удобства при реализации произвольных запросов к БД.

Все множество операций можно разделить на две группы: теоретико-множественные операции и специальные операции.

В первую группу входят 4 операции. Три первые теоретико-множественные операции являются бинарными, то есть в них участвуют два отношения и они требуют **эквивалентных схем исходных отношений**.

Операции над отношениями

- Объединение
- Пересечение
- Разность
- Произведение
- Выборка (ограничение)
- Проекция
- Соединение
- Деление

Объединением двух отношении называется отношение, содержащее множество кортежей, принадлежащих либо первому, либо второму исходным отношениям, либо обоим отношениям одновременно (логическое «ИЛИ»).

Пусть заданы два отношения $R1 = \{ r1 \}$, $R2 = \{ r2 \}$. где $r1$ и $r2$ — соответственно кортежи отношений $R1$ и $R2$

R1	
Шифр детали	Название детали
00011073	Гайка М1
00011075	Гайка М2
00011076	Гайка М3
00011003	Болт М1
00011006	Болт М3
00013063	Шайба М1
00013066	Шайба М3

R2	
Шифр детали	Название детали
00011073	Гайка М1
00011076	Гайка М3
00011077	Гайка М4
00011004	Болт М2
00011006	Болт М3

R3	
Шифр детали	Название детали
00011073	Гайка М1
00011075	Гайка М2
00011076	Гайка М3
00011003	Болт М1
00011006	Болт М3
00013063	Шайба М1
00013066	Шайба М3
00011077	Гайка М4
00011004	Болт М2

Пересечением отношений называется отношение, которое содержит множество кортежей, принадлежащих одновременно и первому и второму отношениям R1 и R2 (логическое «И»)

R1	
Шифр детали	Название детали
00011073	Гайка М1
00011075	Гайка М2
00011076	Гайка М3
00011003	Болт М1
00011006	Болт М3
00013063	Шайба М1
00013066	Шайба М3

R2	
Шифр детали	Название детали
00011073	Гайка М1
00011076	Гайка М3
00011077	Гайка М4
00011004	Болт М2
00011006	Болт М3

R4	
Шифр детали	Название детали
00011073	Гайка М1
00011076	Гайка М3
00011006	Болт М3

Разностью отношений R_1 и R_2 называется отношение, содержащее множество кортежей, принадлежащих R_1 и не принадлежащих R_2

R1	
Шифр детали	Название детали
00011073	Гаика М1
00011075	Гайка М2
00011076	Гаика М3
00011003	Болт М1
00011006	Болт М3
00013063	Шайба М1
00013066	Шайба М3

R2	
Шифр детали	Название детали
00011073	Гайка М1
00011076	Гайка М3
00011077	Гайка М4
00011004	Болт М2
00011006	Болт М3

R5	
Шифр детали	Название детали
00011075	Гайка М2
00011003	Болт М1
00013063	Шайба М1
00013066	Шайба М3

R6	
Шифр детали	Название детали
00011077	Гайка М4
00011004	Болт М2

Пример

Исходными являются три отношения R_1 , R_2 и R_3 .

Все они имеют эквивалентные схемы.

$R_1 = (\text{ФИО}, \text{Паспорт}, \text{Школа});$

$R_2 = (\text{ФИО}, \text{Паспорт}, \text{Школа});$

$R_3 = (\text{ФИО}, \text{Паспорт}, \text{Школа}).$

Рассмотрим ситуацию поступления в высшие учебные заведения, которая была характерна для периода, когда были разрешены так называемые репетиционные вступительные экзамены, которые сдавались раньше основных вступительных экзаменов в вуз.

Отношение R_1 содержит список абитуриентов, сдававших репетиционные экзамены.

Отношение, R_2 содержит список абитуриентов, сдававших экзамены на общих условиях.

Отношение R_3 содержит список абитуриентов, принятых в институт. Будем считать, что при неудачной сдаче репетиционных экзаменов абитуриент мог делать вторую попытку и сдавать экзамены в общем потоке.

Ответим на следующие вопросы:

1. Полный список всех поступавших в вузы.
2. Список абитуриентов поступивших по репетиционному тестированию.
3. Список абитуриентов, которые поступали два раза и не поступили в вуз.
4. Список абитуриентов, которые поступили в вуз с первого раза, то есть они сдавали экзамены только один раз и сдали их так хорошо, что сразу были зачислены в вуз.
5. Список абитуриентов, которые поступили в вуз только со второго раза.
6. Список абитуриентов, которые поступали только один раз и не поступили.

Результатом **произведения** двух отношений является отношение, содержащее все возможные кортежи, которые являются сцеплением двух кортежей, принадлежащих двум отношениям.

Эта операция не накладывает никаких дополнительных условий на схемы исходных отношений, поэтому операция расширенного декартова произведения, обозначаемая $R_1 \otimes R_2$, допустима для любых двух отношений.

Сцеплением, или конкатенацией, кортежей $s = \langle c_1, c_2, \dots, c_n \rangle$ и $q = \langle q_1, q_2, \dots, q_m \rangle$ называется кортеж, полученный добавлением значений второго в конец первого.

Сцепление кортежей s и q обозначается как (s, q) .

$$(s, q) = \langle c_1, c_2, \dots, c_n, q_1, q_2, \dots, q_m \rangle$$

Здесь n — число элементов в первом кортеже s , m — число элементов во втором кортеже q .

Расширенным декартовым произведением отношения R ,

степени n со схемой

$$S_{R_1} = (A_1, A_2, \dots, A_n)$$

и отношения R_2 степени m со схемой

$$S_{R_2} = (B_1, B_2, \dots, B_m)$$

называется отношение R_3 степени $n+m$ со схемой

$$S_{R_3} = (A_1, A_2, \dots, A_n, B_1, B_2, \dots, B_m),$$

содержащее кортежи, полученные сцеплением каждого кортежа r отношения R_1 с каждым кортежем q отношения R_2 .

То есть если $R_1 = \{ r \}$, $R_2 = \{ q \}$

$$R_1 \textcircled{R} R_2 = \{(r, q) \mid r \in R_1 \wedge q \in R_2\}$$

Например, на производстве в отношении R7 задана обязательная номенклатура деталей для всех цехов, а в отношении R8 дан перечень всех цехов.

R7	
Шифр детали	Название детали
00011073	Гайка М1
00011006	Болт М6
00013062	Шайба М2

R8
Цех
Цех 1
Цех 2

R9		
Шифр детали	Название детали	Цех
00011073	Гайка М1	Цех 1
00011073	Гайка М1	Цех 2
00011006	Болт М6	Цех 1
00011006	Болт М6	Цех 2
00013062	Шайба М2	Цех 1
00013062	Шайба М2	Цех 2

Специальные операции реляционной алгебры

Первой специальной операцией реляционной алгебры является **горизонтальный выбор**, или **операция фильтрации**, или **операция ограничения отношений**.

Результатом выборки является отношение, содержащее все кортежи из исходного отношения, которые удовлетворяют определённому условию.

Пусть a — булевское выражение, составленное из термов сравнения с помощью связок И , ИЛИ , НЕ и, возможно, скобок. В качестве термов сравнения допускаются:

- а) терм $A \text{ ос } a$, где A — имя некоторого атрибута, принимающего значения из домена D ; a — константа, взятая из того же домена D ; ос — одна из допустимых для данного домена D операций сравнения;
- б) терм $A \text{ ос } B$, где A, B — имена некоторых Q -сравнимых атрибутов, то есть атрибутов, принимающих значения из одного и того же домена D .

Тогда результатом операции выбора, или фильтрации, заданной на отношении R в виде булевского выражения, определенного на атрибутах отношения R , называется отношение, включающее те кортежи из исходного отношения, для которых истинно условие выбора или фильтрации

Например, выбрать из R_9 детали с шифром «0011073». $R_{12} = R_9[\text{Шифр детали} = \text{«0011003»}]$

R12

Шифр детали	Название детали	Цех
00011073	Гайка M1	Цех 1
00011073	Гайка M1	Цех 3

Проекцией отношения R на набор атрибутов V , обозначаемой $R[V]$, называется отношение со схемой, соответствующей набору атрибутов V , содержащему кортежи, получаемые из кортежей исходного отношения R путем удаления из них значений, не принадлежащих атрибутам из набора V .

По определению отношений все дублирующие кортежи удаляются из результирующего отношения.

Результатом проекции является отношение, содержащее все кортежи после удаления из них некоторых атрибутов. возможно указание всех атрибутов исходного отношения для проекции – получится **тождественная проекция**; возможно указать пустой список атрибутов – получится **нулевая проекция**.

Например, выберем все цеха, которые изготавливают деталь «Болт М1». Для этого нам необходимо из отношения R_{10} выбрать детали с заданным названием, а потом полученное отношение спроектировать на столбец «Цех». Результатом выполнения этих операций будет отношение R_{14} :

$R_{13} = R_9$ [Название детали = «Болт М1»]

$R_{14} = R_{13}$ [Цех]

R13		
Шифр детали	Название	Цех
00011003	Болт М1	Цех 1
00011003	Болт М1	Цех3

R14
Цех
Цех 1
Цех3

R10		
Шифр	Название детали	Цех
00011073	Гайка М1	Цех 1
000 11 076	Гайка М3	Цех 1
000 11 003	Болт М1	Цех 1
000 11 004	Болт М2	Цех 1
00011077	Гайка М4	Цех 1
00011006	Болт М3	Цех2
00013063	Шайба М1	Цех 3
0013066	Шайба М3	Цех 3
00011077	Гайка М4	Цех 2
00011003	Болт М1	Цех 3

Операция условного соединения.

Результат соединения – это отношение, кортежи которого являются сочетанием двух кортежей, принадлежащих двум начальным отношениям и имеющих общие значения для одного или нескольких атрибутов (общее значение в результирующем отношении появляется только один раз).

Пусть $R = \{r\}$, $Q = \{q\}$ — исходные отношения,

S_R , S_Q — схемы отношений R и Q соответственно.

$S_R = (A_1, A_2, \dots, A_k)$; $S_Q = (B_1, B_2, \dots, B_m)$,

где A, B , — имена атрибутов в схемах отношений R и Q соответственно. При этом полагаем, что заданы наборы атрибутов A и B и эти наборы состоят из Q -сравнимых атрибутов.

Тогда соединением отношений R и Q при условии p будет подмножество декартова произведения отношений R и Q , кортежи которого удовлетворяют условию p

Например, рассмотрим следующий запрос. Пусть отношение R_{15} содержит перечень деталей с указанием материалов, из которых эти детали изготавливаются. В каком цеху производится деталь указано в отношении R_{10} .

R15		
Шифр детали	Название детали	Материал
00011073	Гайка М1	сталь-ст1
00011075	Гайка М2	сталь-ст2
00011076	Гайка М3	сталь-ст1
00011003	Болт М1	сталь-ст3
00011006	Болт М3	сталь-ст3
00013063	Шайба М1	сталь-ст1
00013066	Шайба М3	сталь-ст1
00011077	Гайка М4	сталь-ст2
00011004	Болт М2	сталь-ст1
00011005	Болт М5	сталь-ст3
00013062	Шайба М2	сталь-ст1

Получим перечень деталей, которые изготавливаются в цеху 1 из материала «сталь-ст1»

R16
Название детали
Гайка М1
Гайка М3
Болт М2

Операция деления . Результатом деления двух отношений (бинарного и унарного) является отношение, содержащее все значения атрибута первого бинарного отношения, которые соответствуют всем значениям унарного отношения.

R1	
A	X
A	Y
B	Z
B	X
C	Y

R2
X
Y

R
A

Пусть заданы два отношения:

A с заголовком $\{a_1, a_2, \dots, a_n, b_1, b_2, \dots, b_m\}$

B с заголовком $\{b_1, b_2, \dots, b_m\}$.

Будем считать, что атрибут b_i отношения A и атрибут b_i отношения B не только обладают одним и тем же именем, но и определены на одном и том же домене.

Назовем множество атрибутов $\{a_j\}$ составным атрибутом a, а множество атрибутов $\{b_j\}$ - составным атрибутом b.

После этого будем говорить о реляционном делении бинарного отношения A(a,b) на унарное отношение B(b).

Результатом деления A на B является унарное отношение C(a), состоящее из кортежей v таких, что в отношении A имеются кортежи $\langle v, w \rangle$ такие, что множество значений $\{w\}$ включает все множество значений атрибута b в отношении B.

Операция деления удобна тогда, когда требуется сравнить некоторое множество характеристик отдельных атрибутов.

Например, пусть у нас есть отношение A , которое содержит номенклатуру всех выпускаемых деталей на нашем предприятии, а в отношении B хранятся сведения о том, что и в каких цехах действительно выпускается.

Операция деления позволит определить перечень цехов, в которых выпускается вся номенклатура деталей.