

Реляционная модель данных

Артамонов Ю.Н.

Основные термины и определения

Реляционная модель данных некоторой предметной области представляет собой набор отношений, изменяющихся во времени. При создании информационных систем совокупность отношений позволяет хранить данные об объектах предметной области и моделировать связи между ними.

При использовании реляционной модели используется ряд понятий и терминов. Рассмотрим их.

Элемент реляционной модели	Форма представления
Отношение	Таблица
Схема отношения	Строка заголовков столбцов таблицы (заголовок таблицы)
Кортеж	Строка таблицы
Сущность	Описание свойств объекта
Атрибут	Заголовок столбца таблицы
Домен	Множество допустимых значений атрибута
Значение атрибута	Значение поля в записи
Первичный ключ	Один или несколько атрибутов
Тип данных	Тип значений элементов таблицы

Основные термины и определения

- Отношение является важнейшим понятием и представляет собой двумерную таблицу, содержащую некоторые данные.
- Сущность есть объект любой природы, данные о котором хранятся в базе данных. Данные о сущности хранятся в отношении.
- Атрибуты представляют собой свойства, характеризующие сущность. В структуре таблицы каждый атрибут именуется и ему соответствует заголовок некоторого столбца таблицы.
- Запись (кортеж) отношения представляет собой набор значений всех атрибутов отношения для некоторой сущности. В структуре таблицы кортеж соответствует строке таблицы.
- Домен представляет собой множество всех возможных значений определенного атрибута отношения.
- Схема отношения (заголовки отношения) представляет собой список имен атрибутов.
- Кардинальное число - это число кортежей в отношении.
- Степень отношения - это число атрибутов в отношении.

Математически отношение можно описать следующим образом.

Пусть даны n множеств $D_1, D_2, D_3, \dots, D_n$, тогда отношение R есть множество упорядоченных кортежей $\langle d_1, d_2, d_3, \dots, d_n \rangle$, где $d_k \in D_k$, d_k — атрибут, а D_k его домен.

Примеры введенных терминов

Отношение СОТРУДНИК (таблица)			
Атрибут Отдел (заголовок столбца)		Схема отношения (строка заголовков)	
ФИО	Отдел	Должность	Д_рождения
Иванов И.И.	002	Начальник	27.09.51
Петров П.П.	001	Заместитель	15.04.55
Сидоров И.П.	002	Инженер	13.01.70

Отношение СОТРУДНИК включает 4 домена. Домен 1 содержит фамилии всех сотрудников, домен 2 — номера всех отделов фирмы, домен 3 — названия всех должностей, домен 4 — даты рождения всех сотрудников. Каждый домен образует значения одного типа данных, например, числовые или символьные.

Отношение СОТРУДНИК содержит 3 кортежа. Кортеж рассматриваемого отношения состоит из 4 элементов, каждый из которых выбирается из соответствующего домена. Каждому кортежу соответствует строка таблицы.

Схема отношения имеет вид СОТРУДНИК (ФИО , Отдел, Должность, Д_Рождения). Множество собственно кортежей отношения часто называют содержимым (телом) отношения.

Отличия реляционного отношения от таблицы

Основное отличие реляционного отношения от таблицы связано с тем, что отношение это множество, в котором:

- не может быть одинаковых элементов (кортежей);
- кортежи не упорядочены (строки в таблице упорядочены сверху вниз);
- атрибуты в кортеже также не упорядочены (столбцы в таблице упорядочены слева направо);
- все значения атрибутов атомарны.

Ограничения целостности данных

Определение базы данных нуждается в расширении, включающем правила обеспечения целостности, назначение которых в том, чтобы информировать СУБД об ограничениях реального мира.

Ограничения целостности могут быть специфическими для конкретной базы данных. Например, группа, в которую входит студент, должна быть из определенного списка существующих групп. Но в реляционной модели существуют и общие правила обеспечения целостности, описываемые потенциальными и внешними ключами. Пусть R - реляционное отношение. Тогда потенциальным ключом (обозначим его K) будет называться подмножество атрибутов R , обладающее следующими свойствами:

- уникальность — в отношении R не может быть двух различных кортежей с одинаковым значением K ;
- неизбыточность — никакое из подмножеств K , отличных от него, не обладает свойством уникальности.

Потенциальный ключ, состоящий из одного атрибута, называется простым. Например, в отношении, представленном в таблице, можно ввести потенциальный ключ Номер_пропуска (считаем, что Номер_пропуска уникален). Потенциальный ключ, состоящий из нескольких атрибутов, называется составным (например, номер, серия паспорта).

Ограничения целостности данных

Потенциальные ключи важны из-за того, что они обеспечивают основной механизм адресации на уровне кортежей. Иначе говоря, единственный гарантируемый системой способ указать на кортеж указать значение его потенциального ключа или одного из таких ключей, если их несколько.

В реляционной модели принято один из потенциальных ключей выбирать в качестве первичного ключа. Остальные потенциальные ключи будут альтернативными ключами. С точки зрения реляционной модели, какой из ключей выбрать в качестве первичного непринципиально. Но многие СУБД при хранении данных упорядочивают записи по значению первичного ключа. В этом случае, очень важно, чтобы операция сравнения значений первичного ключа выполнялась максимально быстро. Поэтому предпочтительнее будет использовать простой ключ с целочисленными значениями. Если среди имеющихся атрибутов, подходящего на такую роль нет, нередко специально для создания первичного ключа вводят атрибут, которому присваивают значения целочисленного счетчика. Подобный первичный ключ называют суррогатным, в противоположность естественному, который формируется из атрибутов, характеризующих реальный объект.

Некоторые авторы вводят понятие вторичного ключа как подмножества атрибутов, которое существенным образом характеризует кортеж, но не обладает уникальностью. В таблице это может быть поле ФИО. При проектировании базы данных он может учитываться путем создания неуникального индекса.

Ограничения целостности данных

Отношений в БД может быть достаточно много. Причем некоторые из них могут быть связаны друг с другом. При операциях с ними большое значение имеют внешние ключи.

Пусть $R1$ – базовое отношение. Подмножество атрибутов FK отношения $R1$ будет называться внешним ключом, если выполняются следующие условия:

- 1 существует базовое отношение R (причем $R1$ и R не обязательно различны) с потенциальным ключом ;
- 2 каждое значение FK в текущем значении отношения $R1$ всегда совпадает со значением ключа некоторого кортежа в текущем значении отношения R .

Внешний ключ будет простым тогда и только тогда, когда соответствующий потенциальный ключ простой. Аналогично для составного внешнего ключа. Каждый атрибут, входящий в данный внешний ключ, должен быть определен в том же домене, что и соответствующий атрибут потенциального ключа.

Проблема обеспечения того, что база данных не включает неверных значений внешних ключей, называется проблемой обеспечения ссылочной целостности. Отношение, содержащее внешний ключ, называется ссылающимся отношением. Отношение, содержащее соответствующий потенциальный ключ, называется ссылочным отношением.

Ограничения целостности данных

Правило ссылочной целостности заключается в том, что база данных не должна содержать несогласованных значений внешних ключей. При реализации данного правила необходимо ответить как минимум на два вопроса:

- 1) Что должно случиться при попытке удалить объект ссылки внешнего ключа?
- 2) Что должно случиться при попытке обновить потенциальный ключ, на который ссылается внешний ключ?

В первом случае можно ограничить действие – запретить операцию удаления до момента пока существуют ссылающиеся значения внешних ключей. А можно «каскадировать» удаление, то есть удалить кортеж с соответствующим значением потенциального ключа и все кортежи со ссылающимися на него внешними ключами. Аналогично можно поступить и во втором случае – ограничить обновление или каскадировать изменения.

На практике надо очень аккуратно использовать каскадирование удаления записей. Единичная ошибка пользователя в этом случае может принести большой вред, так как вместе с удаляемой записью может быть удалено множество записей в связанных таблицах.

Реляционная алгебра

При разработке реляционной модели Эдгар Кодд ввел реляционную алгебру, которая состоит из набора операторов, использующих отношения в качестве операндов и возвращающих отношения в качестве результата.

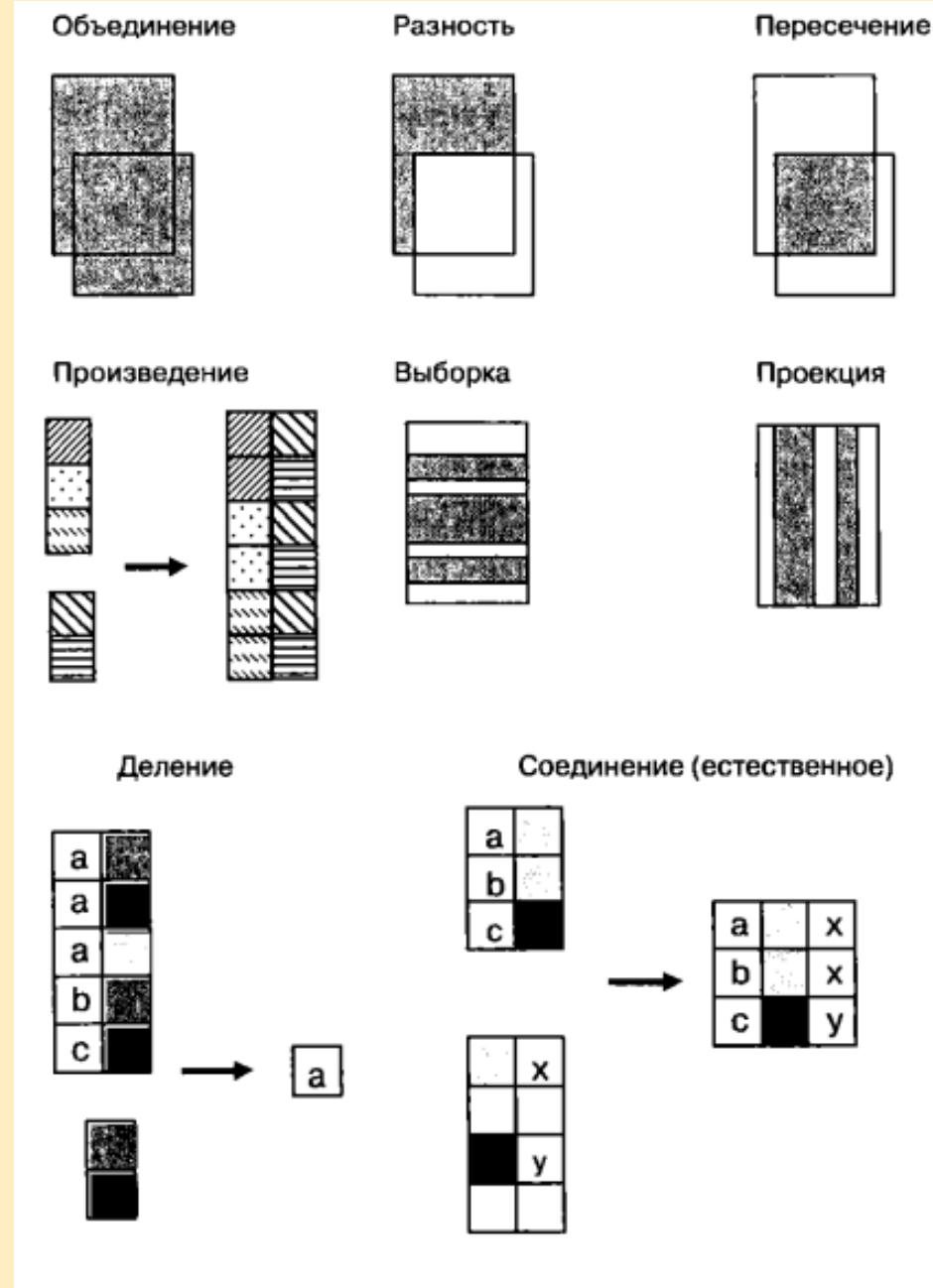
Реляционная алгебра включает восемь операций:

- 1 объединение;
- 2 пересечение;
- 3 вычитание;
- 4 декартово произведение;
- 5 выборка;
- 6 проекция;
- 7 соединение;
- 8 деление.

Операции: объединение, пересечение, вычитание, декартово произведение относятся к традиционным операциям над множествами. Остальные: выборка, проекция, соединение, деление - это специальные реляционные операции.

Говоря о реляционной алгебре, нужно указать на свойство замкнутости. Оно заключается в том, что результат реляционной операции над отношением также является отношением. Следовательно, результат одной операции может использоваться в качестве исходных данных для другой.

Основные операции реляционной алгебры



Основные операции реляционной алгебры

При выполнении бинарных операций участвующие в операциях отношения должны быть совместимы по структуре. Совместимость структур отношений означает совместимость имен атрибутов и типов соответствующих доменов. Частным случаем совместимости является идентичность (совпадение). Для устранения конфликтов имен атрибутов в исходных отношениях применяется операция переименования атрибутов. Структура результирующего отношения по определенным правилам наследует свойства структур исходных отношений. В большинстве рассматриваемых бинарных реляционных операций будем считать, что заголовки исходных отношений идентичны, так как в этом случае не возникает проблем с заголовком результирующего отношения (в общем случае, заголовки могут не совпадать, тогда нужно оговаривать правила формирования заголовка отношения-результата).

Объединением двух совместимых отношений R_1 и R_2 одинаковой размерности ($R_1 \text{ UNION } R_2$ ($R = R_1 \cup R_2$)) является отношение R , содержащее все элементы исходных отношений (с исключением повторений).

Вычитание совместимых отношений R_1 и R_2 одинаковой размерности ($R_1 \text{ MINUS } R_2$ ($R = R_1 \setminus R_2$)) есть отношение, тело которого состоит из множества кортежей, принадлежащих R_1 , но не принадлежащих отношению R_2 .

Пересечение двух совместимых отношений R_1 и R_2 одинаковой размерности ($R_1 \text{ INTERSECT } R_2$ ($R = R_1 \cap R_2$)) порождает отношение R с телом, включающим в себя кортежи, одновременно принадлежащие обоим исходным отношениям.

Пример операций объединения, вычитания, пересечения реляционной алгебры

Отношение А

<u>StudID</u>	ФИО	Группа
123	Иванов И.И.	382
124	Петров П.П.	382

Отношение В

<u>StudID</u>	ФИО	Группа
123	Иванов И.И.	382
127	Сидоров С.С.	383

Объединение $A \cup B$

<u>StudID</u>	ФИО	Группа
123	Иванов И.И.	382
124	Петров П.П.	382
127	Сидоров С.С.	383

Пересечение $A \cap B$

<u>StudID</u>	ФИО	Группа
123	Иванов И.И.	382

Вычитание $A \setminus B$

<u>StudID</u>	ФИО	Группа
124	Петров П.П.	382

Операция декартово произведение реляционной алгебры

Декартово произведение двух отношений R_1 и R_2 (где R_1 и R_2 не имеют общих имен атрибутов) определяется как отношение с заголовком, который представляет собой объединение («сцепление») двух заголовков исходных отношений R_1 и R_2 , и телом, состоящим из множества всех кортежей, сцепления для каждого кортежа R_1 всех возможных кортежей R_2 . Кардинальное число результата равно произведению кардинальных чисел исходных отношений, а степень равна сумме их степеней.

Отношение А

<u>StudID</u>	ФИО
123	Иванов И.И.
124	Петров П.П.
127	Сидоров С.С.

Отношение В

<u>Предмет</u>
Математика
Физика

Декартово произведение $A \times B$

<u>StudID</u>	Ф.И.О.	<u>Предмет</u>
123	Иванов И.И.	Математика
123	Иванов И.И.	Физика
124	Петров П.П.	Математика
124	Петров П.П.	Физика
127	Сидоров С.С.	Математика
127	Сидоров С.С.	Физика

Операция выборка реляционной алгебры

Выборка ($R \text{ WHERE } f$) отношения R по формуле f представляет собой новое отношение с таким же заголовком и телом, состоящим из таких кортежей отношения R , которые удовлетворяют истинности логического выражения, заданного формулой f .

Для записи формулы используются операнды — имена атрибутов (или номера столбцов), константы, логические операции (AND — И, OR — ИЛИ, NOT — НЕ), операции сравнения и скобки.

Отношение STUDENTS

<u>StudID</u>	ФИО	Группа
123	Иванов И.И.	382
124	Петров П.П.	382
127	Сидоров С.С.	383

Выборка STUDENTS [Группа = 382]

<u>StudID</u>	ФИО	Группа
123	Иванов И.И.	382
124	Петров П.П.	382

Операция проекция реляционной алгебры

Проекция отношения R на атрибуты X, Y, \dots, Z ($R[X, Y, \dots, Z]$) представляет собой отношение с заголовком X, Y, \dots, Z и телом, содержащим кортежи отношения R , за исключением повторяющихся кортежей.

Отношение STUDENTS

<u>StudID</u>	ФИО	Группа
123	Иванов И.И.	382
124	Петров П.П.	382
127	Сидоров С.С.	383
128	Иванов И.И.	384

Проекция STUDENTS [ФИО]

ФИО
Иванов И.И.
Петров П.П.
Сидоров С.С.

Операция соединения реляционной алгебры

Определены две разновидности операции соединения: естественное соединение и Θ -соединение.

Пусть отношения A и B имеют заголовки $\{X, Y\}$ и $\{Y, Z\}$, где атрибуты X, Y, Z могут быть составными, причем только атрибут Y (или подмножество атрибутов, обозначаемое Y) присутствует как в первом, так и во втором отношении. Атрибуты с одинаковыми именами определены на одних доменах. Тогда естественным соединением отношений A и B называется отношение с заголовком $\{X, Y, Z\}$ и телом, содержащим все кортежи, для которых в отношении A значение атрибута Y равно значению атрибута Y в отношении B .

Обозначается естественное соединение A и B , как $A \text{ JOIN } B$. Если отношения A и B не имеют общих имен атрибутов, то их естественное соединение эквивалентно декартовому произведению.

Пример естественного соединения реляционной алгебры

Ниже приведен пример естественного соединения двух реляционных отношений.

Отношение STUDENTS

<u>StudID</u>	ФИО	Группа
123	Иванов И.И.	382
124	Петров П.П.	382
127	Сидоров С.С.	383

Отношение GROUPS

Староста	<u>Группа</u>
Петров П.П.	382
Сергеев С.С.	384

Естественное соединение STUDENTS JOIN GROUPS

<u>StudID</u>	ФИО	Группа	Староста
123	Иванов И.И.	382	Петров П.П.
124	Петров П.П.	382	Петров П.П.

⊖ соединение реляционной алгебры

Пусть отношения A и B не имеют общих атрибутов, и оператор Θ определяется также как функция f в выборке. Тогда Θ -соединением отношения A по атрибуту X с отношением B по атрибуту Y называется отношение с тем же заголовком, что и при декартовом произведении отношений A и B , и с телом, содержащим множество всех кортежей декартова произведения A и B , для которых условие $X\Theta Y$ даст значение «истина».

Отношение A

<u>Прибор</u>	Цена_прибора
Прибор1	100
Прибор2	200

Отношение B

<u>Подставка</u>	Цена_подставки
Подставка1	50
Подставка2	150

⊖-соединение A [Цена_прибора ≥ Цена_подставки] B

<u>Прибор</u>	Цена_прибора	<u>Подставка</u>	Цена_подставки
Прибор1	100	Подставка1	50
Прибор2	200	Подставка1	50
Прибор2	200	Подставка2	150

Θ соединение эквивалентно нахождению декартова произведения двух отношений и последующей Θ -выборке, произведенной над результатом. Часто используется следующий вариант записи этой операции: $A[X\Theta Y]B$

Операция деления реляционной алгебры

Пусть отношения A и B имеют заголовки $\{X, Y\}$ и $\{Y\}$ соответственно, атрибуты X и Y могут быть составными. Одинаково названные атрибуты двух отношений определены на одних и тех же доменах.

Результатом деления отношения A на B будет отношение с заголовком $\{X\}$ и телом, содержащим множество всех кортежей из начений $x \in X$, таких, что в A существует кортеж $\{X : x, Y : y\}$, для всех кортежей $\{Y : y\} \in B$.

Обозначается эта операция как:

A/B

$A \text{ DIVIDE BY } B$

Отношение А

<u>StudID</u>	<u>Предмет</u>
123	Физика
123	Математика
124	Математика
127	Физика

Отношение В

<u>Предмет</u>
Физика
Математика

Деление А / В

<u>StudID</u>
123

Недостатки реляционной алгебры

По замечанию Дейта, реляционная алгебра Кодда обладает рядом недостатков. Во-первых, восемь перечисленных операций по охвату своих функций, с одной стороны, избыточны, так как минимально необходимый набор составляют пять операций: объединение, вычитание, произведение, проекция и выборка. Три другие операции (пересечение, соединение и деление) можно определить через пять минимально необходимых. Так, например, соединение — это проекция выборки произведения. Во-вторых, этих восьми операций недостаточно для построения реальной СУБД на принципах реляционной алгебры. Требуются расширения, включающие операции: переименования атрибутов, образования новых вычисляемых атрибутов, вычисления итоговых функций, построения сложных алгебраических выражений, присвоения, сравнения и т. д.