Глава
Организация реляционных баз данных.

1. Основные определения реляционных БД.

Работы над построением реляционных моделей данных ведутся с конца 60-х годов. Термин Реляционная СУБД впервые был использован Коддом (Codd E.F) в 1970 году. Кодд предложил использовать для обработки данных аппарат теории множеств (объединение, пересечение, разность, декартово произведение). Он показал, что любое представление данных сводится к совокупности двумерных таблиц особого вида, известного в математике как *отношение* – relation (англ.). в настоящее время благодаря своей простоте реляционные СУБД доминируют на рынке СУБД.

В соответствии с реляционной моделью данных база данных представляется в виде совокупности таблиц данных, над которыми могут выполняться операции реляционной алгебры или реляционного исчисления.

Различают бинарные реляционные модели и реляционные модели с произвольным числом арности отношения.

Наименьшая единица данных в реляционной модели – это *атомарное* (неразложимое) для данной модели значение данных. Так, в одной предметной области фамилия, имя и отчество могут рассматриваться как единое значение, а в другой – как три различных значения.

Доменом называется множество атомарных значений одного и того же типа, из которого извлекаются значения для атрибутов. Например, домен пунктов отправления (назначения) – множество названий населенных пунктов, а домен номеров рейса – множество целых положительных чисел.
Смысл доменов состоит в следующем. Если значения двух атрибутов берутся из одного и того же домена, то имеют смысл операции сравнения, использующие эти два атрибута. Например, можно задать запрос "*Выдать рейсы, в которых время вылета из Москвы в Сочи больше времени прибытия из Архангельска в Москву*".
Если же значения двух атрибутов берутся из различных доменов, то их сравнение лишено смысла. Например, нельзя сравнить номер рейса со стоимостью билета, даже несмотря на то, что эти атрибуты имеют числовой тип. Обозначаются домены так :

D1, D2, ..., Dn

Отношение – это двумерная таблица, наименования колонок которой совпадают с именами атрибутов, а значения элементов каждого из столбцов извлекаются из соответствующих доменов. Отношение на доменах D1, D2,

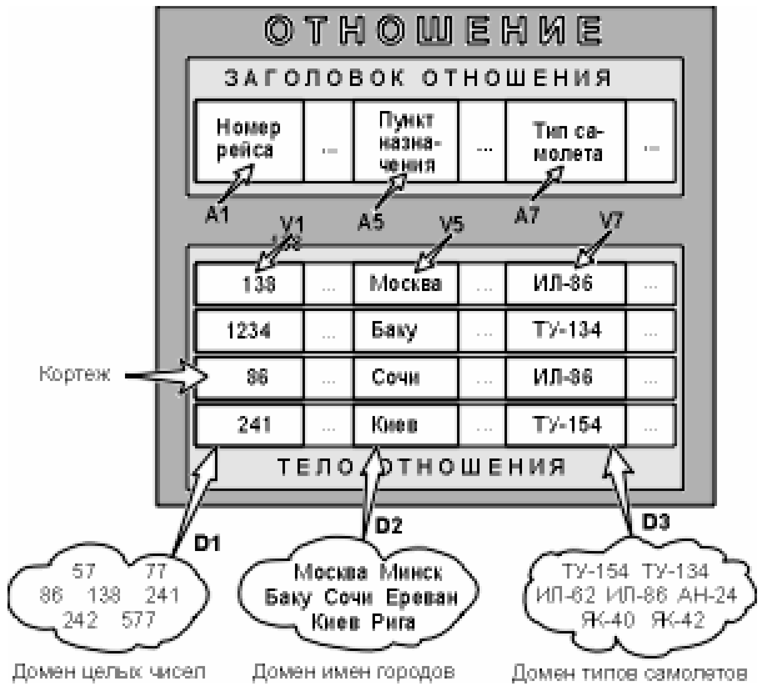
..., Dn (не обязательно, чтобы все они были различны) состоит из заголовка и тела.
Заголовок состоит из такого фиксированного множества атрибутов A1, A2, ..., An, что существует взаимно однозначное соответствие между этими атрибутами Ai и определяющими их доменами Di (i=1,2,...,n).
Тело состоит из меняющегося во времени множества *кортежей*, где каждый кортеж состоит в свою очередь из множества пар атрибут-значение (Ai:Vi), (i=1,2,...,n), по одной такой паре для каждого атрибута Ai в заголовке. Для любой заданной пары атрибут-значение (Ai:Vi) Vi является значением из единственного домена Di, который связан с атрибутом Ai.
Степень отношения – это число его атрибутов. Отношение степени один называют унарным, степени два – бинарным, степени три – тернарным, ..., а степени n – n-арным.
Кардинальное число или мощность отношения – это число его кортежей. Кардинальное число отношения изменяется во времени в отличие от его степени.
Поскольку отношение – это множество, а множества по определению не содержат совпадающих элементов, то никакие два кортежа отношения не могут быть дубликатами друг друга в любой произвольно-заданный момент времени.
Пусть R – отношение с атрибутами A1, A2, ..., An. Говорят, что множество атрибутов K=(Ai, Aj, ..., Ak) отношения R является возможным ключом R тогда и только тогда, когда удовлетворяются два независимых от времени условия:

a) Уникальность: в произвольный заданный момент времени никакие два различных кортежа R не имеют одного и того же значения для Ai, Aj, ..., Ak.
b) Минимальность: ни один из атрибутов Ai, Aj, ..., Ak не может быть исключен из K без нарушения уникальности.

Каждое отношение обладает хотя бы одним возможным ключом, поскольку по меньшей мере комбинация всех его атрибутов удовлетворяет условию уникальности. Один из возможных ключей (выбранный произвольным образом) принимается за его первичный ключ. Остальные возможные ключи, если они есть, называются альтернативными (вторичными) ключами.

Для пользователя реляционных СУБД можно использовать неформальные эквиваленты этих понятий:
Отношение – это таблица либо файл данных
Кортеж – это Строка (иногда Запись),
Атрибут – это Столбец либо Поле.
При этом принимается, что "запись" означает "экземпляр записи", а "поле" означает "имя и тип поля".

Рисунок – отношение с математической точки зрения



2. Фундаментальные свойства отношений.

Реляционная база данных – это совокупность отношений, содержащих всю информацию, которая должна храниться в БД.

Для каждого отношения реляционной БД должны выполняться следующие свойства :

1. . Каждая таблица состоит из однотипных строк и имеет уникальное имя.
2. Строки имеют фиксированное число полей (столбцов) и значений. Множественные поля и повторяющиеся группы полей недопустимы. Поэтому каждой позиции таблицы на пересечении строки и столбца всегда имеется в точности одно значение.
3. Строки таблицы обязательно отличаются друг от друга хотя бы единственным значением. Это позволяет однозначно идентифицировать любую строку такой таблицы.
4. Столбцам таблицы однозначно присваиваются имена, и в каждом из них размещаются однородные значения данных (даты, фамилии, целые числа или денежные суммы).
5. Полное информационное содержание базы данных представляется в виде явных значений данных, и такой метод представления является единственным.

6. При выполнении операций с таблицей ее строки и столбцы можно обрабатывать в любом порядке безотносительно к их информационному содержанию. Этому способствует наличие имен таблиц и их столбцов, а также возможность выделения любой их строки или любого набора строк с указанными признаками.
7. Отсутствие кортежей-дубликатов. То свойство, что отношения не содержат кортежей-дубликатов, следует из определения отношения как множества кортежей. В классической теории множеств по определению каждое множество состоит из различных элементов. Из этого свойства вытекает наличие у каждого отношения так называемого первичного ключа - набора атрибутов, значения которых однозначно определяют кортеж отношения. Для каждого отношения по крайней мере полный набор его атрибутов обладает этим свойством
8. Отсутствие упорядоченности кортежей. Это является следствием определения отношения как множества кортежей. В случае упорядоченности кортежей получается не отношение, а некоторый упорядоченный список кортежей. Отсутствие требования упорядоченности дает дополнительную гибкость СУБД при хранении баз данных и выполнении запросов.
9. Атрибуты отношений не упорядочены. Для ссылки на значение атрибута в кортеже отношения всегда используется имя атрибута. Это свойство теоретически позволяет модифицировать схемы существующих отношений путем добавления новых атрибутов и путем удаления существующих атрибутов. Однако в большинстве существующих систем такая возможность не допускается

Реляционная алгебра.

Поскольку отношения – это множества, то над ними применимы все операции, производимые над множествами.

Оерации реляционной алгебры делятся на два класса : теоретико-множественные и специальные.

В состав теоретико-множественных операций входят операции:

- объединения отношений;
- пересечения отношений;
- взятия разности отношений;
- прямого произведения отношений.

Специальные реляционные операции включают:

- ограничение отношения;
- проекцию отношения;
- соединение отношений;
- деление отношений.

Кроме того, в состав алгебры включается операция присваивания, позволяющая сохранить в базе данных результаты вычисления алгебраических выражений, и операция переименования атрибутов, дающая возможность корректно сформировать заголовок (схему) результирующего отношения.

Объединение отношений.

Пусть имеется два исходных отношения *R1 и R2*. Операция объединения отношений обозначается следующим образом : *R1 U R2*. Результат объединения - отношение, объединяющее кортежи, содержащиеся в исходных отношениях. Отношения называются совместимыми по объединению в том и только в том случае, когда они обладают одинаковыми заголовками. Это означает, что в заголовках обоих отношений содержится один и тот же набор имен атрибутов, и одноименные атрибуты определены на одном и том же домене. В случае, если два отношения частично совместимы по объединению, то до выполнения операции объединения эти отношения можно сделать полностью совместимыми по объединению путем применения операции переименования.

$$R_1 \cup R_2 = \{ r \mid r \in R_1 или R_2 \}$$

Пример *R1*

A_1	A_2	A_3
a1	b1	c1
a2	b2	c2
a3	b3	c3

R2

A_1	A_2	A_3
a3	b4	c2
a4	b5	c3
a2	b2	c2

R1 U R2

A_1	A_2	A_3
a1	b1	c1
a2	b2	c2
a3	b3	c3
a3	b4	c2
a4	b5	c3
a2	b2	c2

Пересечение отношений.

Результат операции – отношение, которое включает кортежи, общие для *R1 и R2.*

$$R_1 \cap R_2 = \{ r \mid r \in R_1 и R_2 \}$$

R1

A_1	A_2	A_3
a1	b1	c1
a2	b2	c2
a3	b3	c3

R2

A_1	A_2	A_3
a3	b4	c2
a4	b5	c3
a2	b2	c2

$R_1 \cap R_2$

A_1	A_2	A_3
a2	b2	c2

Разность отношений.

Результат операции – отношение, содержащее кортежи, являющиеся кортежами отношения *R1* и не являющиеся кортежами отношения *R2*.

$$R_1 \setminus R_2 = \{ r \mid r \in R_1, r \notin R_2$$

Для данных отношений *R1 и R2* в результате получим

A_1	A_2	A_3
a3	b3	c3
a1	b1	c1

Прямое декартово произведение

Результат - отношение, кортежи которого являются конкатенацией (сцеплением) кортежей первого и второго отношений. Путь имеется –m-местное отношение R1 и n-местное отношение R2. В результате операции получаем (m+n)- местное отношение, причем первые m элементов представляют собой кортежи из отношения R1, а последние n элементов – кортежи из отношения R2.

$$R_1 \times R_2 = \{\langle r_1, r_1 \rangle \mid r_1 \notin R_1, r_2 \in R_2\}$$

Два отношения совместимы по взятию произведения в том и только в том случае, если множества имен атрибутов этих отношений не пересекаются. Любые два отношения могут быть сделаны совместимыми по взятию прямого произведения путем применения операции переименования к одному из этих отношений.

Пример . Пусть имеются два отношения R1 и R2. Результат их произведения будет следующим :

A_1	A_2	A_3	A_1	A_2	A_3
a1	b1	c1	a3	b4	c2
a2	b2	c2	a3	b4	c2
a3	b3	c3	a3	b4	c2
a1	b1	c1	a4	b5	c3
2	b2	c2	a4	b5	c3
a3	b3	c3	a4	b5	c3
a1	b1	c1	a2	b2	c2
a2	b2	c2	a2	b2	c2
a3	b3	c3	a2	b2	c2

Проекция отношений.

При выполнении проекции отношения на заданный набор его атрибутов получается отношение, кортежи которого производятся путем взятия соответствующих значений из кортежей отношения-операнда. Операция проекции предназначена для изменения числа столбцов в отношении, т.е. в том случае, когда из кортежей необходимо исключить или выделить какие-либо атрибуты.

Обозначим через $j_1, j_2, \dots, j_n$ номера столбцов n-арного отношения.

Операцию проекции обозначим $\pi_{j_1, j_2, \dots, j_n}(R)$

Операция заключается в том, что из отношения R выбираются столбцы и компонуются в указанном порядке $j_1, j_2, \dots, j_n$.

Пример.

Рассмотрим отношение ТЕЛЕВИЗОР

Наименование	Индекс	Диагональ	Цена
LG	MX23VD	23	9864
Samsung	SVX89FT	89	28564
Grundig	GR45TX	45	14587
Sony	LPD67E	67	28564
Lg	MX24VD	24	13568

$\pi_{1,4}(R)$

Наименование	Цена
LG	9864
Samsung	28564
Grundig	14587
Sony	28564
Lg	13568

$\pi_{4,1,3}(R)$

Цена	Наименование	Диагональ
9864	LG	23
28564	Samsung	89
14587	Grundig	45
28564	Sony	67
13568	Lg	24

$\pi_{4}(R)$

Цена
9864
28564
14587
28564
13568

В последней проекции оказались одинаковыми две строки .

Ограничение отношений

Результатом ограничения отношения по некоторому условию является отношение, включающее кортежи отношения-операнда, удовлетворяющее этому условию. Для выполнения ограничения отношение просматривают по строкам и выделяют множество кортежей, удовлетворяющих заданным условиям. Пусть из отношения ТЕЛЕВИЗОР требуется выделить те марки телевизоров, которые имеют стоимость меньше 15000. В результате получим следующее отношение :

Наименование	Индекс	Диагональ	Цена
LG	MX23VD	23	9864
Grundig	GR45TX	45	14587
Lg	MX24VD	24	13568

Соединение отношений

Операция соединения отношений обратна операции проекции. При соединении двух отношений по некоторому условию образуется отношение, кортежи которого являются конкатенацией кортежей первого и второго отношений и удовлетворяют этому условию.

Рассмотрим два бинарных отношения R(A, B) и R(B,C). Соединением отношений *R1 и R2* называют операцию, при которой соединяют два отношения ,используя в качестве признака соединения какой-либо общий атрибут этих отношений.

$R_1 \rhd\lhd R_2 = \{\langle A,B,C\rangle \mid \langle A,B\rangle \in R_1, \langle B,C\rangle \in R_2\}.$

Пример

Такая операция соответствует случаю стыковки таблиц отношений. Поскольку в результирующей таблице атрибуты с одинаковыми именами присутствуют дважды, то один из столбцов исключают. Такую операцию называют *естественным соединением*.

Существует операция эквисоединения.
Операция соединения называется *операцией эквисоединения*, если условие соединения имеет вид (a = b), где a и b - атрибуты разных операндов соединения.

Деление отношений .
Пусть заданы два отношения - R1 с заголовком $\{A_1, A_2, ..., A_n, B_1, B_2, ..., B_m\}$ и R2 с заголовком $\{B_1, B_2, ..., B_m\}$. Будем считать, что атрибут B_i отношения R1 и атрибут B_i отношения R2 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 производить реляционное деление бинарного отношения R1(a,b) на унарное отношение R2(b).
Результатом деления R1 на R2 является унарное отношение C(a), состоящее из кортежей v таких, что в отношении R1 имеются кортежи <v, w> такие, что множество значений {w} включает множество значений атрибута b в отношении R2.

$$R_1[\{a,b\} \div b]R_2 = \pi_a(R_1) \setminus \pi_a((\pi_a(R_1) \times \pi_b(R_2)) \setminus R_1)$$

R1

Наименование	Индекс
LG	MX23VD
Samsung	SVX89FT
Grundig	GR45TX
Sony	LPD67E
Lg	MX24VD

R2

Индекс	Цена
MX23VD	9864
SVX89FT	28564
GR45TX	14587
LPD67E	28564
MX24VD	13568

Пример

R1

A_1	A_2	B_1	B_2
a1	b1	c3	d1
a2	b1	c1	d1
a3	b2	c2	d2
a1	b3	c1	d3
a3	b2	c1	d1
a3	b1	c2	d2

R2

B_1	B_2
c1	d1
c2	d2

Тогда a={ A_1, A_2 } b={ B_1, B_2}

$\pi_a(R_1)$

A_1	A_2
a1	b1
a2	b1
a3	b2
a1	b3
a3	b2
a3	b1

$\pi_b(R_2)$

B_1	B_2
C1	D1
C2	D2

$\pi_a(R_1)\times\pi_b(R_2)$

A_1	**A_2**	**B_1**	**B_2**
a1	b1	c1	d1
a1	b1	c2	d2
a2	b1	c1	d1
a2	b1	c2	d2
a3	b2	c1	d1
a3	b2	c2	d2
a1	b3	c1	d1
a1	b3	c2	d2
a3	b1	c1	d1
a3	b1	c2	d2

$\pi_a(R_1)\times\pi_b(R_2)$

A_1	**A_2**	**B_1**	**B_2**
a1	b1	c1	d1
a1	b1	c2	d2
a2	b1	c2	d2
a1	b3	c1	d1
a1	b3	c2	d2
a3	b1	c1	d1

$\pi_a(\pi_a(R_1)\times\pi_b(R_2)\setminus R_1)$

A_1	**A_2**
a1	b1
a2	b1
a1	b3
a3	b1

Результат -

A_1	**A_2**
a3	b2