ИСПОЛЬЗОВАНИЕ СВОЙСТВ ОПЕРАЦИЙ РЕЛЯЦИОННОЙ АЛГЕБРЫ ДЛЯ ОПТИМИЗАЦИИ ВЫЧИСЛЕНИЯ ВЫРАЖЕНИЙ

Д. В. Тимофеев

Санкт-Петербургский институт информатики и автоматизации РАН

СПИИРАН, 14-я линия ВО, д. 39, Санкт-Петербург, 199178

<tdv@sparm.com>

УДК 004.65

Тимофеев Д. В. **Использование свойств операций реляционной алгебры для оптимизации вычисления выражений** // Труды СПИИРАН. Вып. 4. — СПб.: Наука, 2007.

Аннотация. *В данной работе рассматриваются операции реляционной алгебры и их свойства, а также определяемые свойствами правила преобразования реляционных выражений, направленные на увеличение производительности вычисления выражений.* — Библ. 3 назв.

UDC 004.65

Timofeev D. V. **Using the relational algebra operation properties to optimize expression calculation** // SPIIRAS Proceedings. Issue 4. — SPb.: Nauka, 2007.

Abstract. *The article deals with the relational algebra operations and operation properties. Also we discuss rules for transformation of relational expressions. These rules are directed to increase the performance of expression calculation.* — Bibl. 3 items.

1. Введение

Реляционная модель данных предложена сотрудником компании IBM Е. Ф. Коддом в 1970 г. На сегодняшний день СУБД, поддерживающие реляционную модель, занимают доминирующее положение среди инструментальных средств разработки систем баз данных. Основными достоинствами, обеспечившими реляционной модели большую популярность, являются:

1) наличие простого и мощного математического аппарата;
2) простота используемой структуры данных и манипулятивных операций;
3) возможность ненавигационного манипулирования данными;
4) обеспечение физической и логической независимости данных.

Единственной существенной структурой реляционной модели является отношение степени n. Между отношениями поддерживаются информационные связи. Реляционная база данных представляет собой совокупность взаимосвязанных отношений.

В манипуляционной части реляционной модели утверждаются два базовых механизма манипулирования реляционными данными: основанная на теории множеств (с некоторыми уточнениями) реляционная алгебра и базирующееся на логическом аппарате исчисления предикатов первого порядка реляционное исчисление. Эти механизмы замкнуты относительно понятия отношения. Это означает, что выражения алгебры и формулы исчисления определяются над отношениями и результатом вычисления также являются отношения. Любое выражение или формула могут использоваться в других выражениях или формулах.

В данной работе рассматриваются операции реляционной алгебры, согласно [1, 2, 3], свойства операций, а также определяемые свойствами правила преобразования реляционных выражений, направленные на увеличения производительности вычисления выражений.

2. Структуры данных реляционной модели

Основными понятиями описания структуры данных реляционной модели являются: тип данных (домен), отношение, кортеж, атрибут, степень, мощность, первичный ключ (рис. 1).

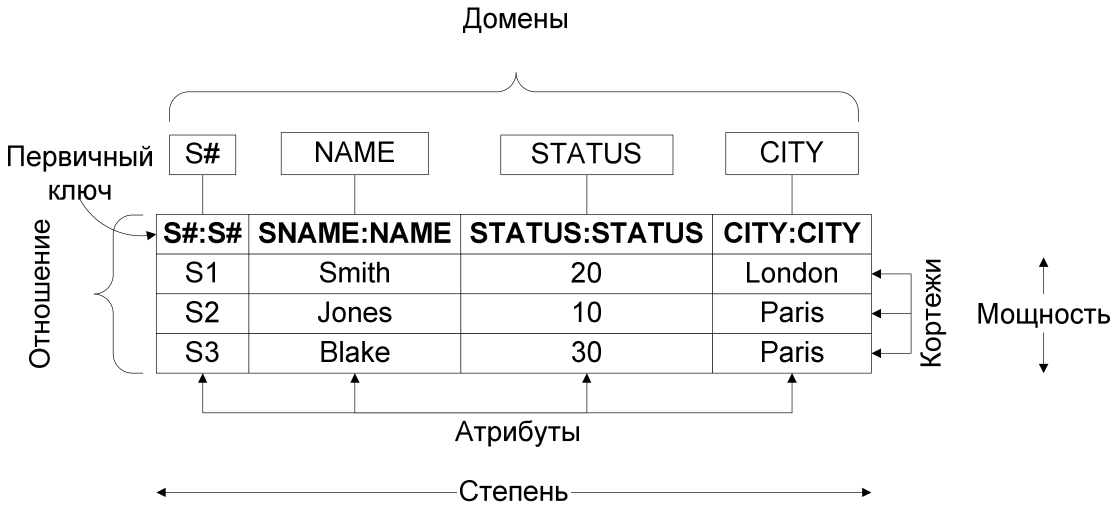

Рис. 1. Термины, используемые для описания структуры данных.

Домен — это тип данных, определённый в системе (встроенный) или определённый пользователем.

Пусть задано множество из n доменов T_i ($i = 1, 2, \ldots, n$), необязательно различных. Тогда r будет отношением, определённым на этих доменах, если оно состоит из двух частей — заголовка и тела, где:

- заголовок (схема отношения) — множество из атрибутов вида $A_i : T_i$; здесь A_i — имена атрибутов (которые должны отличаться друг от друга) отношения r, а T_i — соответствующие имена доменов ($i = 1, 2, \ldots, n$);
- тело — множество из m кортежей t; здесь t в свою очередь является множеством компонентов вида $A_i : v_i$, в которых v_i — значение типа T_i, т.е. значение атрибута для атрибута A_i в кортеже t ($i = 1, 2, \ldots, n$).

Отношение может быть определено следующим образом:

$$
\begin{aligned}
&RELATION\ \{A_1 : T_1, A_2 : T_2, \ldots, A_n : T_n\} \\
&\{ \\
&TUPLE\ \{A_1 : v_{11}, A_2 : v_{21}, \ldots, A_n : v_{n1}\} \\
&TUPLE\ \{A_1 : v_{12}, A_2 : v_{22}, \ldots, A_n : v_{n2}\} \\
&\ldots \\
&TUPLE\ \{A_1 : v_{1m}, A_2 : v_{2m}, \ldots, A_n : v_{nm}\} \\
&\}
\end{aligned}
$$

Труды СПИИРАН. 2007. Вып. 4. ISSN 2078-9181 (печ.), ISSN 2078-9599 (онлайн)
SPIIRAS Proceedings. 2007. Issue 4. ISSN 2078-9181 (print), ISSN 2078-9599 (online)
www.proceedings.spiiras.nw.ru

Значения m и n называются соответственно мощностью (кардинальностью) и степенью отношения r. Набор именованных схем отношений представляет из себя схему базы данных.

Потенциальным ключом отношения называется множество его атрибутов, обладающее свойствами уникальности и неизбыточности. Потенциальные ключи обеспечивают основной механизм адресации на уровне кортежей. По традиции один из потенциальных ключей должен быть выбран в качестве первичного ключа.

Внешним ключом отношения называется множество его атрибутов, такое что каждое значение внешнего ключа обязательно совпадает со значением потенциального ключа некоторого кортежа другого отношения.

Для отражения связи между кортежами разных отношений в реляционной модели используются значения их ключей. Реляционная база данных представляет собой совокупность взаимосвязанных отношений.

3. Реляционная алгебра

Реляционная алгебра — это коллекция операций, которые принимают отношения в качестве операндов и возвращают отношение в качестве результата. Основная идея реляционной алгебры состоит в том, что коль скоро отношения являются множествами, то средства манипулирования отношениями могут базироваться на традиционных теоретико-множественных операциях, дополненных некоторыми специальными операциями, специфичными для баз данных.

Начальный вариант алгебры, предложенный Коддом, состоит из восьми операций, которые делятся на два класса — теоретико-множественные операции и специальные реляционные операции.

В состав теоретико-множественных операций входят операции: объединение ($UNION$), пересечение ($INTERSECT$), разность ($MINUS$), произведение ($TIMES$).

Специальные реляционные операции включают операции: сокращение ($WHERE$), проекция ($\{\ \}$), соединение ($JOIN$), деление ($DIVIDEBY\ \ldots\ PER$).

Кроме того, в состав алгебры включается операция переименования атрибутов ($RENAME$), дающая возможность корректно сформировать заголовок (схему) результирующего отношения.

Оператор проекции $\{\ \}$ имеет более высокий приоритет, все остальные операторы имеют одинаковый приоритет.

3.1. Переименование

Оператор переименования ($RENAME$) принимает заданное отношение A и возвращает другое, идентичное заданному, за исключением того что один или несколько из его атрибутов имеют другие имена

$A\ RENAME\ \ X\ AS\ X_1, Y\ AS\ Y_1, \ldots, Z\ AS\ Z_1$.

Труды СПИИРАН. 2007. Вып. 4. ISSN 2078-9181 (печ.), ISSN 2078-9599 (онлайн)
SPIIRAS Proceedings. 2007. Issue 4. ISSN 2078-9181 (print), ISSN 2078-9599 (online)
www.proceedings.spiiras.nw.ru

3.2. Объединение, пересечение и разность

Отношения-операнды этих операций должны быть совместимы по типу, то есть должны обладать одинаковыми заголовками.

Если даны отношения A и B одного и того же типа, то объединением этих отношений (A *UNION* B) является отношение того же типа с телом, которое состоит из всех кортежей t, присутствующих в A или B или в обоих отношениях (рис. 2).

Если даны отношения A и B одного и того же типа, то пересечением этих отношений (A *INTERSECT* B) является отношение того же типа с телом, состоящим из всех кортежей t, таких что t присутствует одновременно в A и B (рис. 2).

Если даны отношения A и B одного и того же типа, то разностью этих отношений (A *MINUS* B) является отношение того же типа с телом, состоящим из всех кортежей t, таких что t присутствует в A, но не в B (рис. 2).

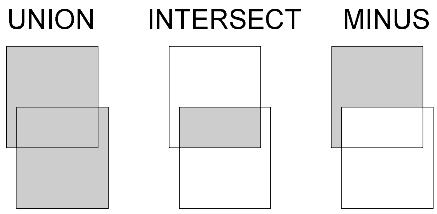

Рис. 2. Объединение, пересечение, разность.

3.3. Произведение

Реляционное декартово произведение отношений A и B, не имеющих общих имён атрибутов (A *TIMES* B), представляет собой новое отношение, заголовок которого является (теоретико-множественным) объединением заголовков отношений A и B, а тело состоит из всех кортежей t, таких что t является (теоретико-множественным) объединением кортежа, принадлежащего к отношению A, и кортежа, принадлежащего к отношению B (рис. 3).

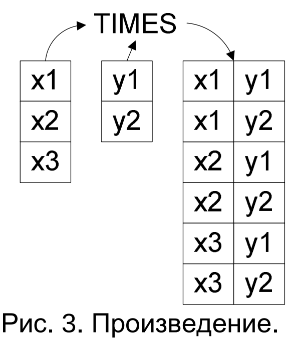

Рис. 3. Произведение.

Кардинальность результата равна произведению кардинальностей входных отношений, A и B, а степень результата — сумме степеней входных отношений.

Труды СПИИРАН. 2007. Вып. 4. ISSN 2078-9181 (печ.), ISSN 2078-9599 (онлайн)
SPIIRAS Proceedings. 2007. Issue 4. ISSN 2078-9181 (print), ISSN 2078-9599 (online)
www.proceedings.spiiras.nw.ru

3.4. Сокращение

Пусть отношение A имеет атрибуты $X, Y, \ldots, Z$ (и, возможно, другие атрибуты), а p является логической функцией, формальные параметры которой представляют собой строгое подмножество атрибутов $X, Y, \ldots, Z$. В таком случае сокращение A в соответствии с p ($A\, WHERE\ p$) является отношением с тем же заголовком, что и в A, и с телом, состоящим из всех кортежей отношения A, таких что функция p принимает логическое значение $TRUE$ для рассматриваемого кортежа.

Оператор сокращения, по сути, позволяет получить «горизонтальное» подмножество заданного отношения (рис. 4).

WHERE

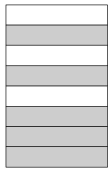

Рис. 4. Сокращение.

3.5. Проекция

Пусть отношение A имеет атрибуты $X, Y, \ldots, Z$ (и, возможно, другие атрибуты). В таком случае проекция отношения A по атрибутам $X, Y, \ldots, Z$ ($A\{X, Y, \ldots, Z\}$) является отношением, соответствующим описанным ниже требованиям:

- Его заголовок формируется из заголовка отношения А путем удаления всех атрибутов, не указанных во множестве $\{X, Y, \ldots, Z\}$.
- Тело состоит из всех кортежей $\{X : x, Y : y, \ldots, Z : z\}$, таких что в отношении A присутствует кортеж со значением x атрибута X, y атрибута Y ... и z атрибута Z.

Применение операции проекции фактически приводит к получению «вертикального» подмножества заданного отношения (рис. 5).

{ }

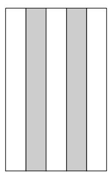

Рис. 5. Проекция.

На практике часто удобно иметь возможность указывать не атрибуты, по которым должна быть сформирована проекция, а скорее те атрибуты, которые «должны быть отброшены» (т.е. удалены) после выполнения операции проекции. Такая форма операции проекции записывается следующим образом:

$A\{ALL\ BUT\ X, Y, \ldots, Z\}$

Труды СПИИРАН. 2007. Вып. 4. ISSN 2078-9181 (печ.), ISSN 2078-9599 (онлайн)
SPIIRAS Proceedings. 2007. Issue 4. ISSN 2078-9181 (print), ISSN 2078-9599 (online)
www.proceedings.spiiras.nw.ru

3.6. Соединение

Пусть отношения A и B соответственно имеют следующие атрибуты:

X , Y

Y , Z

Здесь X — множество атрибутов $\{X_1, X_2, \ldots, X_m\}$, Y — множество атрибутов $\{Y_1, Y_2, \ldots, Y_n\}$, Z — множество атрибутов $\{Z_1, Z_2, \ldots, Z_p\}$.

Тогда естественное соединение отношений A и B (A *JOIN* B) представляет собой отношение с заголовком $\{X, Y, Z\}$ и телом, состоящим из всех таких кортежей $\{X : x, Y : y, Z : z\}$, для которых в отношении A значение атрибута X равно x, а значение атрибута Y равно y и в отношении B значение атрибута Y равно y, а значение атрибута Z равно z (рис. 6).

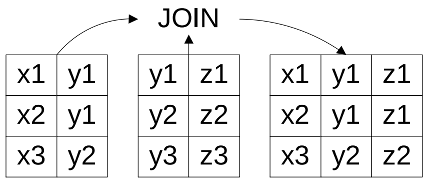

Рис. 6. Соединение.

3.7. Деление

Пусть отношения A и B соответственно имеют следующие атрибуты:

X

Y

Здесь X — множество атрибутов $\{X_1, X_2, \ldots, X_m\}$, Y — множество атрибутов $\{Y_1, Y_2, \ldots, Y_n\}$.

Пусть отношение C имеет следующие атрибуты:

X , Y

Это означает, что C имеет заголовок, представляющий собой (теоретико-множественное) объединение заголовков A и B. В таком случае результат деления A на B по C (A *DIVIDEBY* B *PER* C, где A — делимое, B — делитель, а C — посредник) представляет собой отношение с заголовком $\{X\}$ и телом, состоящим из всех кортежей $\{X : x\}$, присутствующих в A, причем таких что кортеж $\{X : x, Y : y\}$ присутствует в C для всех кортежей $\{Y : y\}$, присутствующих в B (рис. 7).

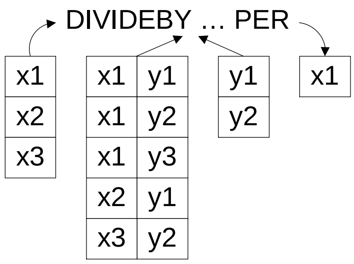

Рис. 7. Деление.

Труды СПИИРАН. 2007. Вып. 4. ISSN 2078-9181 (печ.), ISSN 2078-9599 (онлайн)
SPIIRAS Proceedings. 2007. Issue 4. ISSN 2078-9181 (print), ISSN 2078-9599 (online)
www.proceedings.spiiras.nw.ru

3.8. Минимальное множество операций

Восемь базовых операций Кодда не составляют минимально возможного множества, поскольку некоторые из них не являются примитивными и могут быть определены в терминах других операций.

Операция пересечения может быть выражена через операции разности и объединения:

$$A\ INTERSECT\ B \equiv A\ MINUS\ (A\ MINUS\ B) \equiv B\ MINUS\ (B\ MINUS\ A) \equiv$$
$$(A\ MINUS\ (A\ MINUS\ B))UNION\ (B\ MINUS\ (B\ MINUS\ A))$$

Операция соединения может быть выражена через операции произведения, сокращения, проекции и переименования атрибутов:

$$A\{X,Y\}JOIN\ B\{Y,Z\} \equiv$$
$$(A\ TIMES\ (B\ RENAME\ Y\ AS\ Q)WHERE\ Y = Q)\{ALL\ BUT\ Q\}$$

В частных случаях операция соединения $A\{X^m, Y^n\}JOIN\ B\{Y^n, Z^p\}$ вырождается в операции произведения и пересечения:

- Если $n = 0$ (отношения A и B не имеют общих атрибутов), то операция $A\ JOIN\ B$ вырождается в операцию $A\ TIMES\ B$.
  $$A\{X^m, Y^n\}JOIN\ B\{Y^n, Z^p\} \equiv A\{X^m\}JOIN\ B\{Z^p\} \equiv A\ TIMES\ B$$
- Если $m = p = 0$ (отношения A и B относятся к одинаковому типу), то операция $A\ JOIN\ B$ вырождается в операцию $A\ INTERSECT\ B$.
  $$A\{X^m, Y^n\}JOIN\ B\{Y^n, Z^p\} \equiv A\{Y^n\}JOIN\ B\{Y^n\} \equiv A\ INTERSECT\ B$$

Операция деления может быть выражена через операции соединения (произведения), разности и проекции:

$$A\{X\}DIVIDEBY\ B\{Y\}PER\ C\{X,Y\} \equiv$$
$$A\{X\}MINUS\ ((A\{X\}TIMES\ B\{Y\})MINUS\ C\{X,Y\})\{X\} \equiv$$
$$A\{X\}MINUS\ ((A\{X\}JOIN\ B\{Y\})MINUS\ C\{X,Y\})\{X\}$$

Таким образом, операции пересечения, соединения (произведения) и деления могут быть определены в терминах остальных пяти операций, поэтому их можно исключить без потери каких-либо функциональных возможностей. Но ни одна из оставшихся пяти операций не может быть определена в терминах остальных четырех, поэтому данные пять операций вместе с операцией *RENAME* могут рассматриваться как составляющие минимальное множество операций. Но на практике дополнительные операции являются настолько полезными, что обеспечивается их непосредственная поддержка.

3.9. Назначение реляционной алгебры

Основное назначение реляционной алгебры состоит в обеспечении возможности составления реляционных выражений. Эти выражения в свою очередь могут использоваться во многих разных операциях. Некоторые возможные области применения таких выражений:

- определение области действия операции выборки,
- определение области действия операции обновления,
- определение ограничений целостности,

Труды СПИИРАН. 2007. Вып. 4. ISSN 2078-9181 (печ.), ISSN 2078-9599 (онлайн)
SPIIRAS Proceedings. 2007. Issue 4. ISSN 2078-9181 (print), ISSN 2078-9599 (online)
www.proceedings.spiiras.nw.ru

- определение производных переменных отношения,
- определение требований устойчивости,
- определение ограничений защиты.

Вообще говоря, фактически реляционные выражения служат в качестве высокоуровневого символического описания намерений пользователя. А именно потому, что они являются высокоуровневыми и символическими, сами эти выражения могут стать предметом действия различных высокоуровневых, символических правил преобразования. В результате исходное выражение может быть преобразовано в логически эквивалентное, но, возможно, более эффективное выражение. Таким образом, алгебра служит удобной основой для выполнения оптимизации.

4. Оптимизация и свойства операций

В реляционных системам общее назначение процесса оптимизации состоит в выборе наиболее эффективной стратегии вычисления реляционного выражения.

Обычно реляционные языки позволяют сформулировать любые запросы несколькими разными способами. Производительность вычисления запроса не должна зависеть от формы записи запроса, которую выбрал пользователь. Поэтому необходимо преобразовать запрос в некоторую эквивалентную каноническую форму. Назначением данного преобразования является поиск представления запроса, в некотором смысле более эффективного по сравнению с исходным представлением.

Чтобы преобразовать запрос в некоторую эквивалентную, но более эффективную форму, оптимизатор использует правила или законы преобразования, которые основываются на свойствах операций реляционной алгебры.

Если принять в качестве меры оценки производительности количество операций ввода-вывода кортежей отношений, то основная задача процесса оптимизации — минимизировать количество сканируемых кортежей на каждом этапе вычисления и по возможности избежать выполнения дорогостоящих операций произведений и соединения.

Далее рассматриваются свойства операций реляционной алгебры и их применение для оптимизации вычисления выражений.

4.1. Операции сокращения и проекции

1. Последовательность операций сокращения одного и того же отношения может быть заменена единственной операцией сокращения этого отношения (причем условные выражения всех исходных операций с помощью операций AND объединяются в одно условное выражение).

$$(A\ WHERE\ p_1)WHERE\ p_2 \equiv A\ WHERE\ p_1\ AND\ p_2$$

<u>Доказательство:</u>

$$t \in (A\ WHERE\ p_1)WHERE\ p_2 \equiv (t \in A\ AND\ p_1(t))WHERE\ p_2 \equiv$$
$$(t \in A\ AND\ p_1(t))AND\ p_2(t) \equiv t \in A\ AND\ (p_1(t)\ AND\ p_2(t)) \equiv$$
$$t \in A\ WHERE\ (p_1\ AND\ p_2)$$

Труды СПИИРАН. 2007. Вып. 4. ISSN 2078-9181 (печ.), ISSN 2078-9599 (онлайн)
SPIIRAS Proceedings. 2007. Issue 4. ISSN 2078-9181 (print), ISSN 2078-9599 (online)
www.proceedings.spiiras.nw.ru

Такое преобразование желательно, поскольку первоначальный вариант требует двух проходов в процессе обработки переменной отношения A, тогда как преобразованный вариант требует только одного прохода.

Здесь также следует отметить следующие эквивалентности:

$(A\ WHERE\ p_1)UNION\ (A\ WHERE\ p_2) \equiv A\ WHERE\ p_1\ OR\ p_2$

$(A\ WHERE\ p_1)INTERSECT\ (A\ WHERE\ p_2) \equiv A\ WHERE\ p_1\ AND\ p_2$

$A\ MINUS\ (A\ WHERE\ p) \equiv A\ WHERE\ NOT\ (p)$

2. В последовательности операций проекции для одного и того же отношения можно игнорировать все проекции кроме последней.

$(A\{X, Y, Z\})\{Z\} \equiv A\{Z\}$

<u>Доказательство:</u>

$t\{Z\} \in (A\{X, Y, Z\})\{Z\} \equiv (t\{X, Y, Z\} \in A)\{Z\} \equiv t\{Z\} \in A\{Z\}$

Чтобы исходное выражение имело смысл, каждый атрибут, присутствующий во второй проекции ($\{Z\}$), обязательно должен присутствовать и в первой проекции ($\{X, Y, Z\}$).

3. Операцию сокращения для результата операции проекции можно преобразовать в операцию проекции для результата операции сокращения.

$(A\{C\})WHERE\ p \equiv (A\ WHERE\ p)\{C\}$

<u>Доказательство:</u>

$t\{C\} \in (A\{C\})WHERE\ p \equiv t\{C\} \in (A\{C\}\ AND\ p(t)\{C\}) \equiv$

$t\{C\} \in (A\ AND\ p(t))\{C\} \equiv\ t\{C\} \in (A\ WHERE\ p)\{C\}$

Операции сокращения целесообразно выполнять перед операциями проекции, поскольку операции сокращения обычно приводят к уменьшению объема тех данных, которые будут являться входными для операций проекции. Следовательно, в этом случае уменьшается количество данных, которые потребуется сортировать для исключения возможных дублирующихся записей, образующихся в процессе выполнения операций проекции.

4.2. Распределительный закон

Унарный оператор f распределяется по бинарной операции $\circ$ тогда и только тогда, когда для всех A и B выполняется следующее тождество:

$f(A \circ B) \equiv f(A) \circ f(B)$

В реляционной алгебре операция сокращения распределяется по операциям объединения, пересечения и разности.

$(A\ UNION\ B)WHERE\ p \equiv (A\ WHERE\ p)UNION\ (B\ WHERE\ p)$

$(A\ INERSECT\ B)WHERE\ p \equiv (A\ WHERE\ p)\ INERSECT\ (B\ WHERE\ p)$

$(A\ MINUS\ B)WHERE\ p \equiv (A\ WHERE\ p)MINUS\ (B\ WHERE\ p)$

<u>Доказательство (для операции объединения):</u>

$t \in (A\ UNION\ B)WHERE\ p \equiv (t \in A\ OR\ t \in B)WHERE\ p \equiv$

$(t \in A\ OR\ t \in B)AND\ p(t) \equiv (t \in A\ AND\ p(t))OR\ (t \in B\ AND\ p(t)) \equiv$

$(t \in A\ WHERE\ p)OR\ (t \in B\ WHERE\ p) \equiv$

$t \in (A\ WHERE\ p)UNION\ (B\ WHERE\ p)$

Она также распределяется по операции соединения, но лишь тогда и только тогда, когда условие операции сокращения состоит (в самом сложном случае) из двух простых условий (проверка такого условия может выполняться

Труды СПИИРАН. 2007. Вып. 4. ISSN 2078-9181 (печ.), ISSN 2078-9599 (онлайн)
SPIIRAS Proceedings. 2007. Issue 4. ISSN 2078-9181 (print), ISSN 2078-9599 (online)
www.proceedings.spiiras.nw.ru

для определенного кортежа t без исследования любых других кортежей, кроме t) сокращения, соединенных операцией AND — по одному условию сокращения для каждого операнда операции соединения.

$$(A\ JOIN\ B)\ WHERE\ p_1\ AND\ p_2 \equiv (A\ WHERE\ p_1)\ JOIN\ (B\ WHERE\ p_2)$$

В силу этого закона операция сокращения может быть выполнена раньше остальных операций. Предварительное выполнение операций сокращения почти всегда себя оправдывает, поскольку приводит к существенному уменьшению количества кортежей, обрабатываемых следующей операцией, а также, возможно, к уменьшению количества кортежей на выходе этой операции.

Операция проекции распределяется по операциям объединения и пересечения (но не по операции разности):

$$(A\ UNION\ B)\{C\} \equiv A\{C\}\ UNION\ B\{C\}$$

$$(A\ INTERSECT\ B)\{C\} \equiv A\{C\}\ INTERSECT\ B\{C\}$$

Здесь A и B должны иметь одинаковые типы.

Операция проекции распределяется также по операции соединения при условии, что в результате операции проекции сохраняются все атрибуты соединения, поэтому справедливо следующее тождество:

$$(A\ JOIN\ B)\{C\} \equiv A\{AC\}\ JOIN\ B\{BC\}$$

Здесь AC — объединение атрибутов соединения и тех атрибутов из C, которые присутствуют только в A, а BC — объединение атрибутов соединения и тех атрибутов из C, которые присутствуют только в B.

Закон можно использовать для организации предварительного выполнения операций проекции, что обычно себя оправдывает по тем же причинам, как и в случае операций сокращения.

Бинарная операция δ распределяется по бинарной операции o тогда и только тогда, когда для всех A, B и C истинно следующее тождество:

$$A\ \delta\ (B\ \text{o}\ C) \equiv (A\ \delta\ B)\ \text{o}\ (A\ \delta\ C)$$

В реляционной алгебре операция объединения распределяется по операции пересечения, а операция пересечения — по операции объединения:

$$A\ UNION\ (B\ INTERSECT\ C) \equiv (A\ UNION\ B)\ INTERSECT\ (A\ UNION\ C)$$

$$A\ INTERSECT\ (B\ UNION\ C) \equiv (A\ INTERSECT\ B)\ UNION\ (A\ INTERSECT\ C)$$

<u>Доказательство (для операции объединения):</u>

$$t \in A\ UNION\ (B\ INTERSECT\ C) \equiv (t \in A)\ OR\ (t \in (B\ INTERSECT\ C)) \equiv$$

$$(t \in A)\ OR\ (t \in B\ AND\ t \in C) \equiv (t \in A\ OR\ t \in B)\ AND\ (t \in A\ OR\ t \in C) \equiv$$

$$(t \in A\ UNION\ B)\ AND\ (t \in A\ UNION\ C) \equiv$$

$$t \in (A\ UNION\ B)\ INTERSECT\ (A\ UNION\ C)$$

В силу этого закона операция пересечения может быть выполнена раньше операции объединения, что может привести к существенному уменьшению количества кортежей, обрабатываемых следующей операцией.

4.3. Коммутативность и ассоциативность

Бинарная операция o является коммутативной тогда и только тогда, когда для всех A и B истинно следующее тождество:

$$A\ \text{o}\ B \equiv B\ \text{o}\ A$$

Труды СПИИРАН. 2007. Вып. 4. ISSN 2078-9181 (печ.), ISSN 2078-9599 (онлайн)
SPIIRAS Proceedings. 2007. Issue 4. ISSN 2078-9181 (print), ISSN 2078-9599 (online)
www.proceedings.spiiras.nw.ru

В реляционной алгебре коммутативными являются операции объединения, пересечения, произведения и соединения, а операции разности и деления таковыми не являются.

$A\ UNION\ B \equiv B\ UNION\ A$

$A\ INTERSECT\ B \equiv B\ INTERSECT\ A$

$A\ TIMES\ B \equiv B\ TIMES\ A$

$A\ JOIN\ B \equiv B\ JOIN\ A$

Доказательство (для операции соединения):

$t\{X, Y, Z\} \in A\ JOIN\ B \equiv (t\{X, Y\} \in A)\ AND\ (t\{Y, Z\} \in B) \equiv$

$(t\{Y, Z\} \in B)\ AND\ (t\{X, Y\} \in A) \equiv t\{Z, Y, X\} \in B\ JOIN\ A$

Если запрос включает соединение двух отношений, A и B, то коммутативность означает, что безразлично, какое из отношений A или B выбрано в качестве внешнего или внутреннего. Следовательно, при вычислении конкретного соединения системе предоставлено право выбора в качестве внешнего отношения, например меньшего из двух отношений.

Бинарная операция $\circ$ является ассоциативной тогда и только тогда, когда для всех A, B и C истинно следующее тождество:

$A \circ (B \circ C) \equiv (A \circ B) \circ C$

В реляционной алгебре ассоциативными являются операции объединения, пересечения, произведения и соединения, а операции разности и деления таковыми не являются.

$A\ UNION\ (B\ UNION\ C) \equiv (A\ UNION\ B)\ UNION\ C \equiv A\ UNION\ B\ UNION\ C$

$A\ INTERSECT\ (B\ INTERSECT\ C) \equiv (A\ INTERSECT\ B)\ INTERSECT\ C \equiv$

$A\ INTERSECT\ B\ INTERSECT\ C$

$A\ TIMES\ (B\ TIMES\ C) \equiv (A\ TIMES\ B)\ TIMES\ C \equiv A\ TIMES\ B\ TIMES\ C$

$A\ JOIN\ (B\ JOIN\ C) \equiv (A\ JOIN\ B)\ JOIN\ C \equiv A\ JOIN\ B\ JOIN\ C$

Доказательство (для операции объединения):

$t \in A\ UNION\ (B\ UNION\ C) \equiv t \in A\ OR\ (t \in (B\ UNION\ C)) \equiv$

$t \in A\ OR\ (t \in B\ OR\ t \in C) \equiv t \in A\ OR\ t \in B\ OR\ t \in C \equiv$

$t \in (A\ UNION\ B)\ UNION\ C \equiv t \in A\ UNION\ B\ UNION\ C$

Если в запросе используется соединение трех отношений, A, B и C, то из законов коммутативности и ассоциативности следует, что не имеет значения, в каком порядке будет выполняться соединение отношений. Поэтому в процессе вычисления подобного соединения системе предоставляется право выбора наиболее эффективной последовательности из всех существующих.

4.4. Идемпотентность и поглощение

Бинарную операцию $\circ$ называют идемпотентной тогда и только тогда, когда для любого A выполняется следующее тождество:

$A \circ A \equiv A$

В реляционной алгебре операции объединения, пересечения, произведения и соединения являются идемпотентными, а операции деления и разности — нет.

$A\ UNION\ A \equiv A$

$A\ INTERSECT\ A \equiv A$

Труды СПИИРАН. 2007. Вып. 4. ISSN 2078-9181 (печ.), ISSN 2078-9599 (онлайн)
SPIIRAS Proceedings. 2007. Issue 4. ISSN 2078-9181 (print), ISSN 2078-9599 (online)
www.proceedings.spiiras.nw.ru

$A\ TIMES\ A \equiv A$

$A\ JOIN\ A \equiv A$

Операции объединения и пересечения подчиняются законам поглощения:

$A\ UNION\ (A\ INTERSECT\ B) \equiv A$

$A\ INTERSECT\ (A\ UNION\ B) \equiv A$

Можно ожидать, что применение законов идемпотентности и поглощения окажется полезным в процессе преобразования выражений.

5. Заключение

Оптимизируемость (возможность оптимизации запросов) является сильной стороной реляционного подхода к организации данных. Свойства реляционных определяют правила преобразования, с помощью которых исходное выражение преобразуется в другое, улучшенное с точки зрения скорости вычисления.

Благодаря тому что реляционные выражения оптимизируются на достаточно высоком семантическом уровне (собственно на уровне операций реляционной алгебры), появляется возможность осуществления автоматической оптимизации. При этом пользователь может не задумываться над наилучшим способом выражения своих запросов. Начиная с исходного выражения, оптимизатор будет шаг за шагом применять различные правила преобразования до тех пор, пока будет получено выражение, которое, согласно встроенным в оптимизатор эвристическим правилам, будет считаться оптимальным для рассматриваемого запроса.

Литература

1. *Дейт К. Дж.* Введение в системы баз данных. 7-е изд. М.: Издательский дом «Вильямс», 2002. 1072 с.
2. *Дейт К. Дж.* Введение в системы баз данных. 8-е изд. М.: Издательский дом «Вильямс», 2005. 1328 с.
3. *Дейт К. Дж., Дарвен Х.* Основы будущих систем баз данных. Третий манифест. 2-е изд. М.: Янус-К, 2004. 656 с.