

УДК 004.658.3

С.Л. Зиноватная, канд.техн.наук,
 Сиран Мусанна, специалист,
 Одес. нац. политехн. ун-т,
 Н.А. Новикова, математик,
 Одес. нац. мор. ун-т

МОДЕЛЬ ИНДЕКСНОЙ СТРУКТУРЫ БАЗЫ ДАННЫХ

С.Л. Зиноватна, Сиран Мусанна, Н.О. Новикова. Модель індексної структури бази даних. Розглядається представлення індексів реляційної бази даних, яке дозволяє формалізувати пошук їх оптимальної структури з точки зору підвищення продуктивності інформаційної системи.

С.Л. Зиноватная, Сиран Мусанна, Н.О. Новикова. Модель индексной структуры базы данных. Рассматривается представление индексов реляционной базы данных, которое позволяет формализовать поиск их оптимальной структуры с точки зрения повышения производительности информационной системы.

S.L. Zinovatnaya, Siran Musanna, N.A. Novikova. Model of database index structure. Presentation of indexes of relational database is examined, which allows to formalize the search of their optimal structure from the point of view of the informative system productivity increase.

В последнее время информационные системы (ИС) стали одним из основных приложений вычислительной техники, от которых напрямую зависит успешная деятельность многих организаций. Объемы данных, хранимые в таких системах, часто достигают многих терабайт, при этом требования прикладных областей к их размерам постоянно растут [1, 2]. Способность системы сохранять громадные объемы данных вряд ли будет востребована, если производительность такой системы недостаточно высока [2]. Актуальной задачей является уменьшение времени реакции ИС на запросы пользователя. Одним из способов повышения производительности ИС является создание оптимальной индексной структуры для базы данных (БД), лежащей в основе ИС. Индекс — структура данных, которая помогает системе управления базами данных (СУБД) быстрее обнаружить отдельные записи и сократить время выполнения запросов пользователей.

При использовании индексов на больших объемах данных или при частых обращениях к БД наблюдается заметное, иногда даже многократное, ускорение работы СУБД. Однако обработка индексной структуры также требует ресурсов компьютера. Поэтому важной задачей при разработке БД является получение сбалансированной структуры индексов, которая гарантированно приведет к существенному повышению производительности ИС.

Достичь этой цели можно только в результате тщательного анализа запросов, поступающих к БД. Более того, со временем может изменяться состав запросов и их распределение во времени. Формальные методы получения индексной структуры в БД, оптимальной с точки зрения баланса между уменьшением времени обработки запросов и временем, затрачиваемым на поддержку самого индекса, отсутствуют.

Предлагается модель индексной структуры, которая позволит формализовать исследование ИС для принятия решения о создании индекса и проанализировать имеющуюся в БД индексную структуру. Рассмотрим множество индексов БД в виде набора кортежей

$$Ind = \langle R_i^I, \langle E_{ij}^I, Ord_{ij} \rangle, T_i^I \rangle, i = \overline{1, m_{Ind}},$$

где R_i^I — таблица БД, для которой определен индекс;

$\langle E_{ij}^I, Ord_{ij} \rangle$ — упорядоченный набор кортежей, каждый экземпляр которого представля-

ет одно поле индекса $j = \overline{1, n_{Ind}}$;

n_{Ind} — количество полей в представленном индексе;

E_{ij}^I — выражение над атрибутами таблицы R_i^I , соответствующее j -му полю представленного индекса;

Ord_{ij} — порядок сортировки, определенный для j -го поля индекса;

T_i^I — тип индекса;

m_{Ind} — количество индексов.

Известно представление запроса к БД, которое позволяет для выявления вариантов денормализации анализировать поведение ИС [3]. Выполним модификацию данного представления следующим образом. Запрос к БД представляется множеством кортежей в виде

$$Q = \langle R_i^Q, F_i^Q, E_i^Q \rangle,$$

где R_i^Q — таблица БД, участвующая в запросе Q ;

F_i^Q — множество атрибутов таблицы R_i^Q , использованных в запросе Q ;

E_i^Q — множество выражений над атрибутами таблицы R_i^Q в запросе Q .

В свою очередь F_i^Q также представляется множеством кортежей, каждый элемент множества описывает характеристики использования одного атрибута таблицы R_i^Q в запросе Q ,

$$F_i^Q = \langle F_{ij}, Ch_{ij}, QF_{ij} \rangle,$$

где F_{ij} — атрибут таблицы R_i^Q ;

Ch_{ij} — множество характеристик атрибута F_{ij} по ключу, каждая определяется из множества возможных значений $\{PK_{Full}, FK_{Full}, PK_{Part}, FK_{Part}\}$, здесь PK_{Full} означает, что атрибут является первичным ключом таблицы R_i^Q ; FK_{Full} — внешним ключом; PK_{Part} — атрибут является частью составного первичного ключа таблицы; FK_{Part} — частью составного ключа.

QF_{ij} — множество характеристик использования в запросе Q атрибута F_{ij} , каждая определяется из множества возможных значений $\{S_{Ord}OrdN, GrN, Sel_{Eqv}, Sel_{NoEqv}, Sel_D, Sel_{Comp}, L_{Begin}\}$, S_{Ord} — в запросе по атрибуту выполняется сортировка в порядке Ord ;

Gr — в запросе по атрибуту выполняется группировка;

Sel_{Eqv}, Sel_{NoEqv} — атрибут присутствует в условии отбора записей с условием по равенству значения или по неравенству значения, соответственно;

Sel_D — выполняется проверка на вхождение в диапазон значений;

Sel_{Comp} — выполняется сравнение больше (или равно) и/или меньше (или равно);

L_{Begin} — выполняется сравнение с шаблоном по начальной части значения атрибута;

N — порядковый номер атрибута в списке сортировки или группировки соответственно для характеристик S_{Ord} и Gr .

Функция пересечения I^{Ind} запроса Q и индексной структуры Ind определяет степень соответствия индексов БД запросу и может принимать следующие значения:

0 — для всех таблиц R_i^Q индексы не определены, $R_i^Q \cap R_i^I = \emptyset$;

1 — хотя бы для одной таблицы из множества R_i^Q существует индекс, который может быть целиком применен для ускорения обработки запроса;

2 — хотя бы для одной таблицы из множества R_i^Q существует индекс, который может быть частично применен для ускорения обработки запроса.

Для определения значений соответствия индексной структуры запросу введены также следующие функции.

Функция пересечения по группировке I_{Gr}^{Ind} запроса Q и индексной структуры Ind определяет, может ли быть использована индексная структура БД для реализации операции группировки в запросе. Функция $I_{Gr}^{Ind} = 1$, если $R^{QI} = R_i^Q \cap R_i^I \neq \emptyset$, и для хотя бы одной таблицы из множества R^{QI} существует такой F_{ij} , для которого $\{Gr\} \subset Ch_{ij}$, и при этом существует кортеж $\langle E_{ij}^I, Ord_{ij} \rangle$ такой, что $F_{ij} = E_{ij}^I$. Кроме того, если $|R_i^Q| > 1$, где R^Q — множество всех таблиц

R_i^Q , то в Q отсутствуют такие пары кортежей F_{i1}^Q и F_{i2}^Q , относящиеся к различным таблицам запроса, для которых существуют характеристики атрибутов $\{Gr\} \subset Ch_{i1j}$ и $\{Gr\} \subset Ch_{i2j}$. В остальных случаях $I_{Gr}^{Ind} = 0$.

Функция пересечения по сортировке I_S^{Ind} запроса Q и индексной структуры Ind определяет, могут ли существующие индексы быть использованы для реализации операции сортировки в запросе. Функция $I_S^{Ind} = 1$, если $R^Q = R_i^Q \cap R_i^I \neq \emptyset$, и для хотя бы одной таблицы из множества R^Q существует такой F_{ij} , для которого $\{S_{Ord}\} \subset Ch_{ij}$, и при этом существует кортеж $\langle E_{ij}^I, Ord_{ij} \rangle$ такой, что $F_{ij} = E_{ij}^I$ и $Ord_{ij} = Ord$ из S_{Ord} . Кроме того, если $|R^Q| > 1$, где R^Q — множество всех таблиц R_i^Q , то в Q отсутствуют такие пары кортежей F_{i1}^Q и F_{i2}^Q , относящиеся к различным таблицам запроса, для которых существуют характеристики атрибутов $\{S_{Ord}\} \subset Ch_{i1j}$ и $\{S_{Ord}\} \subset Ch_{i2j}$. В остальных случаях $I_S^{Ind} = 0$.

Также определены функции пересечения запроса Q и индексной структуры Ind по условию выбора данных I_{Comp}^{Ind} .

Кроме того, перечисленные функции расширены для проверки соответствия запроса и индексной структуры БД не только по одному атрибуту, но и по группе атрибутов.

Предлагается следующий механизм анализа поведения ИС при реализации индексной структуры.

Пусть вес запросов H^{RF} выборки относительно таблицы R и атрибута F — это время выполнения запросов на выборку с использованием таблицы R и атрибута F относительно общего времени выполнения всех запросов к БД на протяжении периода P .

Результаты сопоставления индексной структуры БД и множества запросов, определяющих поведение ИС, представляются в виде множества матриц $M' = \{M^{Gr}, M^S, M^{Comp}\}$. Строки каждой матрицы соответствуют таблицам R_i БД. Столбцы матрицы соответствуют атрибутам таблиц F_j БД. Элемент матрицы состоит из двух компонентов

$$m'_{ij} = \langle H^{R_i F_j}, \{0,1\} \rangle,$$

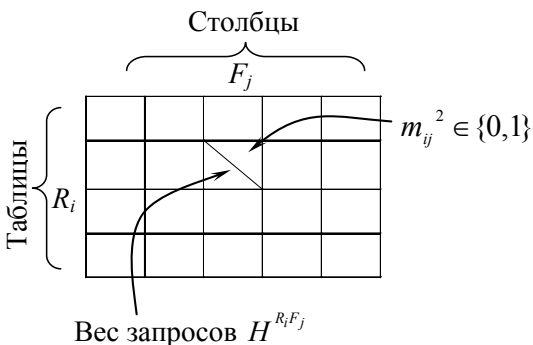
где $H^{R_i F_j}$ — вес запросов выборки для таблицы, соответствующей i -й строке матрицы, и атрибута, соответствующего j -му столбцу матрицы.

Второй компонент элемента матрицы M^{Gr} принимает значение 1, если значение $I_{Gr}^{Ind} = 1$ для атрибута F_j таблицы R_i , иначе — 0; для M^S и M^{Comp} используются соответствующие функции.

Далее формируется матрица $M = \sum M'$, второй компонент $m_{ij}^2 = 0$, если равны нулю все

вторые компоненты элементов m'_{ij} , иначе —

$$m_{ij}^2 = 1 \text{ (см. рисунок).}$$



Предполагается, что если $H^{R_i F_j}$ выше среднего значения из всех элементов матрицы, а $m_{ij}^2 = 0$, то, возможно, потребуется модификация индексной структуры.

Таким образом, анализ содержимого матрицы M позволяет оценить соответствие имеющейся в БД индексной структуры поведению ИС, используя эту БД, и/или получить первичный набор индексов для исследуемой БД. Полученный набор индексов является первичным, т.к. потребуются

Результаты сопоставления индексной структуры БД и множества запросов

дальнейшее уточнение его содержимого с учетом дополнительных характеристик БД и поведения ИС, в частности, размера таблиц, частоты обновления данных и т.д.

Литература

1. Дейт, К. Дж. Введение в системы баз данных / К. Дж. Дейт. 8-е издание. — М.: Издательский дом “Вильямс”, 2005. — 1328 с.
2. Гарсиа-Молина, Г. Системы баз данных. Полный курс / Г. Гарсиа-Молина, Дж. Ульман, Дж. Уидом. — М.: Издательский дом “Вильямс”, 2003. — 1088 с.
3. Зиноватная, С.Л. Представление запроса к реляционной БД / С.Л. Зиноватная // Сборник тезисов научного семинара “Математическое моделирование и информационные технологии”. — Одесса: ОГАХ, 2007. — С. 35 — 36.

Рецензент д-р техн. наук, проф. Одес. нац. политехн. ун-та Дрозд А.В.

Поступила в редакцию 12 мая 2010 г.