

УДК 681.3

В.В. Дрождин

ОРГАНИЗАЦИЯ АДАПТИВНОГО ИНДЕКСА

Рассматривается организация многоуровневого иерархического индекса, приспособляющегося к изменениям интенсивности использования элементов индекса, объема индекса и наличия свободных ресурсов в системе.

1. Назначение, задачи и критерий оптимальности индекса

Индекс представляет собой **управляющую структуру данных**, основным **назначением** которой является обеспечение быстрого доступа к данным. На индекс могут возлагаться также следующие задачи:

- обеспечение прямого и последовательного доступа к данным в возрастающем и убывающем порядке элементов индекса;
- возможность доступа к данным по отдельным полям, группам полей (включая разнотипные) и результатам вычисления выражений, а также по полям большой размерности (длинным строкам, текстам, изображениям и неструктурированным данным);
- осуществление внутренней идентификации данных в системе;
- поддержка взаимосвязей (отношений) между данными.

Основными **требованиями**, предъявляемыми к индексу, являются обеспечение очень высокого быстродействия, высокой надежности и стабильности поведения индекса в процессе всего времени эксплуатации независимо от изменений количества и вероятностей использования элементов индекса, а также доступных в каждый момент времени свободных ресурсов.

Таким образом, обладая большими функциональными возможностями, индекс должен удовлетворять очень высоким требованиям к быстродействию и эффективности функционирования в сильно изменяющихся внешних и внутренних условиях.

Отсюда можно сформулировать обобщенный критерий оптимальности индекса в аддитивной форме:

$$J(t_{cp}, \eta, v_r) = a \cdot \alpha \cdot J_1(t_{cp}) + b \cdot \beta \cdot J_2(\eta) + g \cdot \gamma \cdot J_3(v_r) \rightarrow \min,$$

где $J_1(t_{cp})$ - критерий учета среднего времени выполнения любой операции над индексом;

$J_2(\eta)$ - критерий учета надежности функционирования индекса;

$J_3(v_r)$ - критерий учета наличия свободных ресурсов;

a, b, g - коэффициенты перевода времени, надежности и объема свободных ресурсов в обобщенные единицы;

α, β, γ - коэффициенты важности временного фактора, надежности и наличия определенного объема свободных ресурсов.

Среднее время выполнения любой операции является интегральным критерием оценки степени организованности индекса и эффективности методов реализации операций его обработки.

Критерий надежности учитывает защищенность индекса от сбоев и выхода из строя части оборудования, потери информации вследствие некорректности программного обеспечения и неготовности к выполнению операций обработки индекса из-за выполнения динамической реорганизации индекса или высокой интенсивности его использования.

Необходимость наличия в системе определенного объема свободных ресурсов диктуется соображениями эффективности и надежности функционирования индекса.

Коэффициенты a, b, g обеспечивают сравнимость оцениваемых параметров. Они являются константами и задают объем соответствующего параметра, приходящегося на одну обобщенную единицу.

Коэффициенты α, β, γ отражают ценность соответствующего параметра для системы в каждый момент времени. По существу они являются динамическими приоритетами параметров и зависят от объема параметров в данный момент. Для согласованности коэффициентов наложим условие нормирования $\alpha + \beta + \gamma = 1$.

Другие требования к индексу и задачи, решаемые им, а также объем доступных ресурсов и максимально допустимое время выполнения операций обработки индекса, выступают в качестве ограничений.

2. Функционально-структурная организация индекса

В настоящее время известно большое число структур и методов быстрого доступа к данным. В подавляющем большинстве из них принимается равновероятное обращение ко всем значениям данных и обеспечивается одинаковое время доступа к ним. Однако известно, что обращения к данным базы данных осуществляются далеко не с равными вероятностями, особенно в разные периоды времени. Для учета распределения вероятностей использования данных Кнутом в [1] рассматривались самоорганизующиеся файлы, а Мартиным во [2] приводились адаптивные списковые структуры и метод размещения блоков данных в памяти с различным быстродействием. Но эти структуры носили преимущественно линейный характер и даже в оптимальном случае не могли конкурировать, например, с иерархическими структурами (B-деревьями). Кроме того, реализация их на основе регулярных структур приводила к большим затратам ресурсов и снижению эффективности структур.

Подавляющее большинство существующих систем управления базами данных (FoxPro, Paradox, Access) поддерживают индексы в форме B-дерева или близкие к ним со сжатием ключей и переменным числом элементов в блоках [2]. Это объясняется тем, что время доступа к данным и сложность обработки индекса пропорциональны $O(\log_m N)$, где N - общее число элементов в индексе, а m - число элементов в блоке. При достаточно больших m ($m > 10$) это очень

медленно растущая функция от N . Графическое изображение структуры В-дерева приведено на рис. 1.

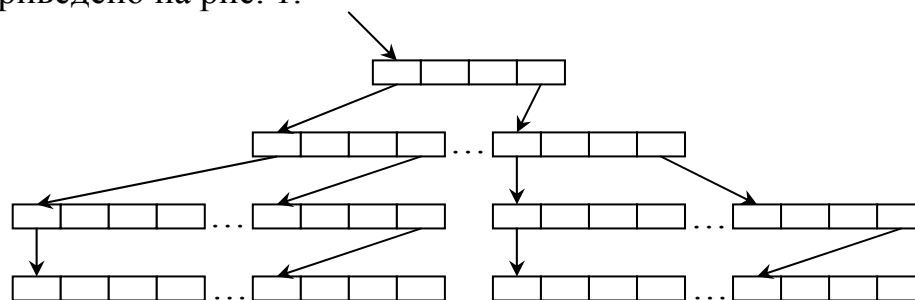


Рис. 1. Структура четырехуровневого В-дерева с блоками по четыре элемента.

Повысить эффективность многоуровневого иерархического индекса можно на основе учета распределения вероятностей использования индексных значений следующим образом:

- размещать элементы данных во всех элементах индекса, а не только в элементах нижнего уровня индекса;
- распределять элементы по уровням индекса в соответствии с интенсивностью их использования таким образом, чтобы на верхнем уровне размещались наиболее часто используемые элементы, а на нижнем - практически неиспользуемые;
- в процессе функционирования индекса элементы, к которым производится больше обращений должны последовательно перемещаться на более высокие уровни, а к которым меньше - на нижние.

Загрузив в быстродействующую память блок верхнего уровня индекса можно обеспечить выполнение большого числа операций обработки индекса без дополнительного обращения к внешней памяти, что делает практически постоянным среднее время доступа к данным независимо от объема индекса. Реорганизация индекса может производиться специальной процедурой в периоды времени, когда система не выполняет внешние запросы.

Однако такая функциональная организация приводит к сложностям в структурной организации, заключающимся в поддержке множеств элементов - потомков [3] существенно отличающихся по мощности (в десятки и сотни раз). Для повышения эффективности использования ресурсов целесообразно использовать блоки нескольких размеров, а для снятия ограничений на количество потомков использовать смежные блоки одного размера. Последовательность смежных блоков, в которой размещается множество потомков элемента, будем называть экстендом.

Наличие у элемента индекса нескольких блоков - потомков (экстенда) для поиска некоторого промежуточного элемента будет требовать нескольких обращений к внешней памяти. Для устранения этого недостатка посредством вычисления номера нужного блока целесообразно использовать некоторую интерполяционную функцию [4].

Кроме того, на нижнем уровне индекса могут использоваться эффективные методы сжатия данных, что обеспечит компактность индекса.

В процессе функционирования количество уровней в индексе и размеры блоков могут меняться в зависимости от изменения мощностей множеств элементов и критерия наличия свободных ресурсов в системе.

Графическое изображение структуры адаптивного индекса приведено на рис. 2.

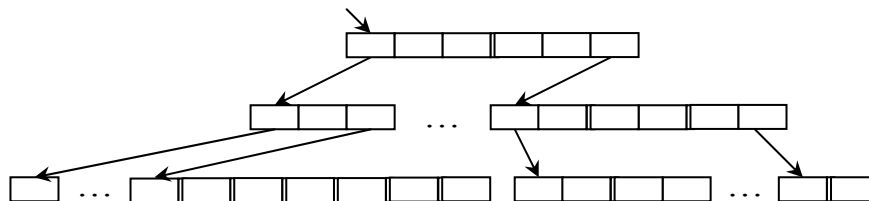


Рис. 2. Структура трехуровневого адаптивного индекса с блоками переменного размера на 1-3 элемента.

Рассмотренная функционально - структурная организация адаптивного индекса является очень гибкой и предоставляет широкие возможности для управления на основе текущей информации и самообучения.

Литература

1. Кнут Д. Искусство программирования на ЭВМ: сортировка и поиск. Т. 3. - М.: Мир, 1978. - 844 с.
2. Мартин Дж. Организация баз данных в вычислительных системах. - М.: Мир, 1980. - 662 с.
3. Евстигнеев В.А. Применение теории графов в программировании. - М.: Наука, 1985. - 352 с.
4. Новиков Б.А. Системы хранения баз данных и знаний // Программирование. - 1993. - N 2.- С. 3 - 30.