

# Средства ускоренного доступа к данным

Чтобы пользователь чувствовал себя комфортно, время ожидания ответа на запрос к БД не должно превышать нескольких секунд. В связи с этим специально разрабатываются методы ускорения выборки, позволяющие обойтись без полного перебора строк при выполнении реляционных операций модификации отношений и отбора данных.

Наиболее эффективны методы индексирования и хеширования значений ключей отношения.

Индексирование — логическая  
сортировка строк таблицы —  
заключается в создании  
вспомогательных файлов, содержащих  
упорядоченные списки значений ключей  
отношения со ссылками на строку  
отношения, в которой они находятся.

Индексные файлы занимают дополнительную память, но резко ускоряют поиск благодаря применению метода половинного деления. Для одного отношения может быть создано несколько индексов.

Кроме того, можно создать индекс для нескольких отношений, если они содержат одинаковые атрибуты, что позволит ускорить выполнение операций соединения этих отношений.

Хеширование (hashing) — использование хэш-функций, которые вычисляют вес строки таблицы по значению ее ключевых атрибутов. Результат вычисления хэш-функции — целое число в диапазоне физических номеров строк таблицы.

Идеальная хэш-функция должна давать  
разные значения веса для разных  
ключевых атрибутов. Но это не всегда  
ВОЗМОЖНО.



На практике обычно используют простые хэш-функции, например  $f(k) = k \bmod p$ , где  $k$  — целое число, первичный ключ отношения;  
 $p$  — простое целое число;  $\bmod$  — операция, вычисляющая остаток при целочисленном делении.

Если ключевой атрибут — строка символов, то для вычисления  $f(k)$  выбирается один из методов преобразования строки в число, например вычисление контрольной суммы.

Для организации доступа к данным при хешировании создается таблица с пустыми строками, которая заполняется следующим образом:

- по первичному ключу новой строки вычисляется значение хэш-функции  $f\{k\}$  и результат трактуется как номер строки в созданной таблице;
- если строка уже занята, производится проверка следующих строк по специальному алгоритму до тех пор, пока не будет обнаружено свободное место.

Аналогично производится поиск нужной строки:

- если после вычисления  $f(k)$  на месте в таблице, которое соответствует вычисленному значению, оказывается пустая строка, значит, искомой строки просто нет;
- если значение ключа совпало с искомым, поиск заканчивается;

- если же значение ключа не совпало с искомым, проверяются следующие строки таблицы до обнаружения строки с нужным ключом (в этом случае искомая строка найдена) или пустой строки (в этом случае искомая строка отсутствует).

Если таблица заполнена не более чем на 60 %, то для размещения новой или поиска существующей строки необходимо проверить в среднем не более двух ячеек. Хеширование используют для поиска строк по точному совпадению значения ключевого атрибута кортежа с нужным значением ключа.