

Описание Менеджера эвристик

Из сложности DPsize [Moe]:

$$I_{DPsize}^{chain}(n) = \begin{cases} \frac{1}{48}(5n^4 + 6n^3 - 14n^2 - 12n), & n \text{ even} \\ \frac{1}{48}(5n^4 + 6n^3 - 14n^2 - 6n + 11), & n \text{ odd} \end{cases}$$
$$I_{DPsize}^{cycle}(n) = \begin{cases} \frac{1}{4}(n^4 - n^3 - n^2), & n \text{ even} \\ \frac{1}{4}(n^4 - n^3 - n^2 + n), & n \text{ odd} \end{cases}$$
$$I_{DPsize}^{star}(n) = \begin{cases} 2^{2n-4} - \frac{1}{4} \binom{2n}{n-1} + q(n), & n \text{ even} \\ 2^{2n-4} - \frac{1}{4} \binom{2(n-1)}{n-1} + \frac{1}{4} \binom{n-1}{(n-1)/2} + q(n), & n \text{ odd} \end{cases}$$

with $q(n) = n2^{2n-1} - 5 \times 2^{n-3} + \frac{1}{2}(2^n - 5n + 4)$

$$I_{DPsize}^{clique}(n) = \begin{cases} 2^{2n-2} - 5 \times 2^{n-2} + \frac{1}{4} \binom{2n}{n} - \frac{1}{4} \binom{n}{n/2} + 1, & n \text{ even} \\ 2^{2n-2} - 5 \times 2^{n-2} + \frac{1}{4} \binom{2n}{n} + 1, & n \text{ odd} \end{cases}$$

Заметим, что для цепей, циклов, звёзд с небольшим количеством таблиц, DPsize обладает быстрым временем планирования, следовательно эвристики нужно использовать для крупных топологий.

Заметим, что плотные графы \sim clique обладают высокой алгоритмической сложностью планирования алгоритмом DPsize. Поэтому все такие графы от 6+ вершин и будем отдавать Geqo.

Заметим, что также крупные звёзды обладают высокой сложностью. Их можно разбить на цепи по центральной вершине. Посмотрим на сложность планирования цепей и циклов: при количестве таблиц ≤ 12 сложность циклов не более чем в 2 раза больше сложности цепи. Таким образом, при количестве таблиц ≤ 12 в цепи(цикле) будем применять стандартный DP.

Пусть дан большой (≥ 50 таблиц) аналитический запрос с внутренними соединениями. Для его планирования будем выполнять следующие шаги:

1. Представим запрос в виде набора связных компонент, любые две таблицы из разных компонент не имеют условия соединения между собой. Для каждой компоненты будем выполнять шаги 2-4.
2. Разобьём компоненту на топологии: сначала выделим из компоненты **плотные графы** (с 6+ вершинами), затем циклы, звёзды, останутся цепи. Звездой назовём группу вершин: центральная и связанные с ней цепи, каждые две цепи не связаны с друг другом в оставшейся топологии. Или если имеется вершина с количеством строк в 10+ раз больше чем у 2+ соседей, то данная вершина станет центральной, а её соседи будут концами цепочек.
3. Назовём топологию маленькой, если в ней $\leq geqo_threshold$ элементов, иначе топология большая. Спланируем каждую топологию графа:
 - (a) Для маленькой цепи будем использовать DPsize. Для большой будем брать минимальный по общей стоимости результат планирования GOO и Geqo.
 - (b) Для маленького цикла будем использовать DPsize. В большом цикле найдем соединение с наибольшей кардинальностью, и разобьём цикл по этом соединению - уберём одну таблицу в этом соединении. Спланируем цепь, и присоединим удалённую таблицу.
 - (c) Маленькую звезду спланируем DPsize. Для большой - сначала спланируем лучи как цепи, затем будем последовательно присоединять к центру используя GOO с центральной фиксированной вершиной.

(d) Плотный граф от ≤ 6 вершин будем планировать DPsize. Для больших плотных графом будем использовать Geoq.

4. Алгоритмом GOO соберём все планы компоненты в один результат.

5. Объединим планы компонент алгоритмом GOO с помощью декартового произведения.

Помимо этого если после выделения плотных графов и циклов компонента исходного графа останется связной, то для оставшейся части дополнительно выполним планирование с linDP. Тогда на 4 шаге у нас будет два варианта планов : плотные графы, циклы, результат linDP и плотные графы, циклы, звёзды, цепи. Выберем из них минимальный по стоимости.

Список литературы

[Moe] Guido Moerkotte. Dynamic programming for join ordering revisited. *School of Electrical Engineering and Computer Science, Faculty of Engineering University of Ottawa*, page 10.