



# Методы ускорения перебора соединений таблиц в планировании SQL запросов

НГУ, ММФ, 22126, Пьянзин Богдан

28 февраля 2025 г.

# Структура курсовой работы

## Введение

- Актуальность темы.
- Цель исследования.

## Основная часть

### Теоретические основы соединений таблиц в SQL запросах

- Определение и классификация соединений и планирование запроса.
- Задача выбора порядка соединений таблиц.
- Факторы, влияющие на производительность (размеры таблиц, статистики, индексы, типы данных и т.д.).

### Методы ускорения перебора соединений таблиц

- Классические подходы.
- Реализация в PostgreSQL.
- Современные подходы. %TODO

### Сравнительный анализ методов ускорения %TODO

- Преимущества и недостатки каждого метода %TODO

### Заключение %TODO

- Итоги работы.
- Рекомендации по применению методов.
- Перспективы исследований.

## Список литературы

# Введение

## Актуальность темы

Современные СУБД работают с большим объёмом информации и транзакций, используя для ввода запросов язык SQL. В процессе трансляции запрос превращается сначала в логическое представление в виде дерева, затем с помощью оптимизатора СУБД в физический план исполнения. Производительность СУБД напрямую зависит от качества сгенерированного физического плана. Для создания "хорошего" плана, нужно решить одну из NP-полных задач перебора соединений таблиц, так как он требует сложных вычислений и значительных затрат ресурсов.

Сложность задачи обусловлена тем, что количество возможных соединений (бинарных) с  $n$  таблицами равно количеству бинарных деревьев с  $n$  листьями - Числа Каталана, которые имеют экспоненциальную скорость роста. В классических подходах используется динамическое программирование (DP) и эвристический поиск для решения этой проблемы. Однако с ростом количества таблиц, потребление памяти (для хранения промежуточных результатов) и время планирования становятся слишком велики. Эвристические методы не гарантируют оптимальность решения, так как находят локально оптимальные планы, но не гарантируют глобально оптимальное решение.

Современные подходы предлагают использовать машинное обучение на основе статистик, которые могут динамически меняться в процессе выполнения запроса. И создавать на их основе оптимальные планы.

Таким образом, выбор неэффективного плана запроса может привести к значительному замедлению работы системы, а процесс поиска хорошего решения является нетривиальной задачей. В свою очередь оптимальный план позволяет эффективно исполнить запрос, с приемлемым потреблением CPU, памяти, I/O, сетевых ресурсов (особенно актуально для распределённых СУБД).

## Цель

Цель данного исследования – выявить факторы, влияющие на выбор порядка соединений в запросах к СУБД, а также проанализировать классические и современные подходы, применяемые для оптимизации данной задачи. В рамках работы планируется:

- Определить ключевые параметры, влияющие на производительность операций соединения (размер таблиц, наличие индексов, тип соединения и др.).
- Разобрать механизмы планирования запросов в PostgreSQL.
- Рассмотреть традиционные методы планирования порядка соединений, включая динамическое программирование, эвристические алгоритмы и генетические, используемые в PostgreSQL, изучить их эффективность и ограничения при увеличении числа соединений в запросе.
- Изучить современные подходы, основанные на машинном обучении, и их применение в автоматическом выборе порядка соединений.
- Провести сравнительный анализ эффективности различных методов и выявить сценарии нагрузок, при которых каждый из них показывает наилучшие результаты.
- Выработать критерии для выбора оптимального метода планирования соединений в зависимости от типа запроса и структуры данных.

Исследование позволит оценить, какие подходы обеспечивают наилучшую производительность SQL-запросов в различных условиях и предложить рекомендации по их применению.

## Основная часть

### Теоретические основы соединений таблиц в SQL запросах

#### Определение и классификация соединений в планировании запроса

**Соединение таблиц** – это операция, которая позволяет объединять данные из двух и более таблиц по определённому условию. Использование соединений необходимо, когда информация распределена между несколькими таблицами. В процессе работы оптимизатора, решается вопрос какой тип соединения использовать и в каком порядке сое. Существуют три типа соединений:

**Nested Loop Join** – выполняет вложенный перебор строк. Для каждой строки из левой таблицы проходится по всем строкам правой таблицы и ищет соответствующие строки по условию соединения. Пусть левая таблица занимает  $M$  страниц и имеет  $m$  кортежей, правая  $N$ ,  $n$  соответственно. Имеется варианты NLJ:

**Наивный NLJ** – перебирает каждый кортеж левой таблицы, сравнивая с каждым из правой. Сложность  $O(M + (m * N))$ . Работает медленно если таблицы большие, но прост в реализации, эффективен при малых таблицах.

**Индексированный NLJ** – правая таблица имеет индекс по соединяемому полю. Сложность  $O(M + (m * \log(n)))$ . Может значительно ускорить поиск, если правая таблица большая. Но будет всё ещё медленным если левая таблица большая.

**Блочный поиск** – улучшенная версия наивного NLJ, при ограничении памяти в буфере. Пусть доступно  $B$  буферов. Загружаем  $B-2$  блоков из левой таблицы, 1 буфер для правой таблицы, 1 под результат. Для каждой страницы из буферов для правой таблицы, проверяем условие с каждой строкой из буфера правой страницы, проходимся одним буфером по всей правой таблице. Сложность  $O(M + [M/(B - 2)] * N)$ .

**Merge Join** – таблицы сортируются по ключу соединения, происходит построчное слияние таблиц. Очень хорошо работает если таблицы отсортированы и/или есть индекс по ключу, лучше чем NLJ на больших таблицах. Недостатки: потребность в сортировке, нужна память для хранения отсортированных таблиц. Сложность  $O(M + N)$ .

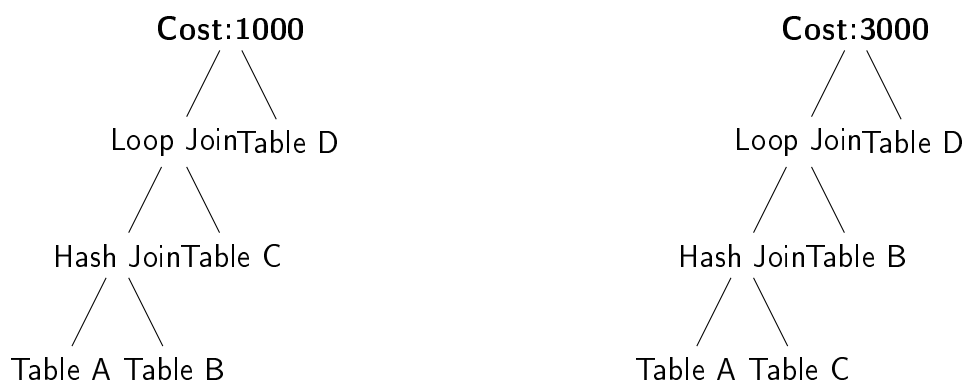
**Hash Join** – строится хэш-таблица для меньшей таблицы, перебирается каждая строка в большей таблице и проверяется соответствие в хэш таблице. Сложность  $O(M + N)$ . Работает хорошо одна таблица большая другая маленькая и условие на равенство(=). Недостатки: требует память для хэш-таблицы, плохо работает если условие на есть диапазон.

Перечисленные выше способы соединения имеют разную эффективность, напрямую зависящую от статистик. Выбор неоптимального типа соединения может привести к значительному ухудшению стоимости исполнения плана. Определение нужного типа происходит в процессе оптимизации, то есть перевода логического дерева в физическое

#### Задача выбора порядка соединений

С увеличением количества данных в СУБД увеличивается и время, необходимое для выполнения запросов. Сокращение времени выполнения запросов становится решающим фактором для повышения удобства и эффективности работы пользователей. СУБД преобразует запрос в набор планов. Каждый план представлен в виде дерева, узлами в которых являются данные из таблиц

или результат соединения таблиц по условию, в общем случае называется отношением. Рёбра — условия соединения отношений. При этом операция соединения отношений не ассоциативна, то есть  $(R1 \times R2) \times R3 \neq R1 \times (R2 \times R3)$  по стоимости. В статье [https://ruor.uottawa.ca/items/9898a486-a7ea-42af-a6d9-66b165a90e33] приводится пример разной стоимости для двух деревьев соединений.



Выбор в какой последовательности нужно соединить таблицы является задачей выбора порядка соединений. Различные планы запросов для одного и того же SQL возвращают одинаковый результат, но время и ресурсы (CPU, память, I/O обращения, возможно сетевые ресурсы), необходимые для выполнения запроса, сильно различаются. Поэтому выбор оптимального плана позволяет сократить время отклика, минимизировать потребление ресурсов и эффективно обрабатывать большие массивы данных, значительно повышая удобство работы пользователя.

Задача планировщика/оптимизатора — построить наилучший план выполнения. Если это не требует больших вычислений, оптимизатор запросов будет перебирать все возможные варианты планов, чтобы в итоге выбрать тот, который имеет наименьшую стоимость. Например, если для обрабатываемого отношения создан индекс, прочитать отношение можно двумя способами. Во-первых, можно выполнить простое последовательное сканирование, а во-вторых, можно использовать индекс. Затем оценивается стоимость каждого варианта и выбирается самый дешёвый. Затем выбранный вариант разворачивается в полноценный план, который сможет использовать исполнитель. Мы примем, что уже получена стоимость каждого оператора сканирования.

### Факторы влияющие на производительность

Исходя из классификации типов соединений при планировании запроса нужно иметь статистику: по размеру таблиц подаваемых на вход оператору, наличие индексов, наличие сортировки данных. Применение некоторых типов может потребовать дополнительной памяти. Помимо этого существуют следующие факторы:

**Размер отношений (кардинальность таблицы)** — количество отдельных строк в таблице базы данных. Если таблица большая, выполнение соединения с ней может быть дорогим, если же таблица маленькая, её можно поместить в память (in-memory), и не придётся обращаться к внешней памяти, что может повлиять на стоимость плана.

**Кардинальность условия (селективность соединения)** — количество строк, оставшихся после фильтрации. Наиболее селективные соединения следуют применять как можно раньше, так стоимость оператора соединения это стоимость дочерних операторов + собственная обработка.

**Селективность столбцов** — уникальность значений в столбце. Высокая селективность - много уникальных значений, низкая селективность - мало уникальных значений.

**Корреляция столбцов** – зависимость значений между столбцами. Если данные сильно коррелируют, оптимизатор может дать неправильную оценку оператору соединения. Тогда ошибка распространится на операторы стоящие выше, что приведёт к ошибке в оценке стоимости исполнения всего плана.

**Распределение данных по столбцам** – оказывает влияние корреляцию данных, селективность соединения. Зная распределение, оптимизатор может делать предположения, как данных хранятся в отношениях: есть ли отсортированность, использовать последовательное сканирование или индекс для поиска.

## Методы ускорения перебора соединений таблиц

СУБД применяют классические и современные подходы к оптимизации порядка соединений. Классические алгоритмы включают в себя:

- Методы ДП, выполняющие полный поиск среди всех возможных порядков соединения, но требуют больших затрат памяти и зачастую имеют плохую асимптотическую сложность.
- Эвристические методы, требуют меньших затрат ресурсов по сравнению с ДП, но находят приблизительно (локально) оптимальные планы.

Современные методы машинного обучения (ML) включают в себя:

- Обучение с подкреплением (Reinforcement Learning), позволяющее адаптивно находить оптимальный порядок соединений на основе исторических данных и дообучаться в процессе работы.
- Специальные модели СУБД, запоминающие часть истории запросов и использующие их для дообучения.

### Классические подходы

Данные подходы в большинстве своём представляют алгоритмы динамического программирования и эвристические.

Основная идея методов ДП в том чтобы разбить задачу поиска оптимального порядка на меньшие (по размеру соединяемых отношений) и решить оптимально их. Затем объединить в более крупные и так далее, пока не получим оптимальное решение для текущего запроса.

С другой стороны эвристический поиск по определённому предположению выбирает "выгодные" планы, отсеивая "плохие на каждом шаге. Зачастую это позволяет получить приблизительно оптимальный план.

В данной работе рассмотрим следующие алгоритмы ДП: DPsize, DPsub, DPccp, DPhyp и LinearizedDP. А также эвристические алгоритмы GOO (Greedy Operator Ordering) и Geqo.

### DPsize [<https://db.in.tum.de/teaching/ws2425/queryopt/main.pdf?lang=de>]

Строит оптимальное **ветвистое дерево** (такое дерево, что хотя бы у одной вершины, начиная с корня, оба потомка составные отношения, т.е. не являются таблицами) снизу вверх, начиная с маленьких соединений и расширяя их. Имеет существенное ограничение – не поддерживает работу с внешними соединениями, т.к. они не коммутативны.

Изначально В хранит каждое отношение  $R_i$  как лучший план для  $R_i$ . Затем начинается перебор всех подмножеств размера  $|s|$  возможных планов от 2 до  $n$ . Перебираются всевозможные разбиения  $S$  на непересекающиеся множества  $S_1, S_2$ , такие что существует хотя бы пара отношений в  $S_1$  и  $S_2$ , связанная между собой условием соединения. Так как  $|S_1|$  и  $|S_2|$  меньше  $|s|$ ,

то известны их оптимальные планы  $p_1$  и  $p_2$  соответственно. Стоимость объединения  $p_1$  и  $p_2$  меньше дешевле старой комбинации  $S_1$  и  $S_2$ , то происходит замена.

---

```

1: Input: A set of relations  $R = \{R_1, \dots, R_n\}$  to be joined
2: Output: An optimal bushy join tree
3:  $B \leftarrow$  an empty DP table  $2^R \rightarrow$  join tree
4: for each  $R_i \in R$  do
5:    $B[\{R_i\}] \leftarrow R_i$ 
6: end for
7: for each  $1 < s \leq n$  ascending do
8:   for each  $S_1, S_2 \subset R$  such that  $|S_1| + |S_2| = s$  do
9:     if (not cross products  $\wedge \neg S_1$  connected to  $S_2$ )  $\vee (S_1 \cap S_2 \neq \emptyset)$  then
10:      continue
11:    end if
12:     $p_1 \leftarrow B[S_1], p_2 \leftarrow B[S_2]$ 
13:    if  $p_1 = \epsilon$  or  $p_2 = \epsilon$  then continue
14:    end if
15:     $P \leftarrow \text{CreateJoinTree}(p_1, p_2)$ 
16:    if  $B[S_1 \cup S_2] = \epsilon$  or  $C(B[S_1 \cup S_2]) > C(P)$  then
17:       $B[S_1 \cup S_2] \leftarrow P$ 
18:    end if
19:  end for
20: end for

```

---

Сложность DPsize[<https://dsg.uwaterloo.ca/seminars/notes/Guido.pdf>]:

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

$$I_{\text{DPsize}}^{\text{cycle}}(n) = \begin{cases} \frac{1}{4}(n^4 - n^3 - n^2), & n \text{ even} \\ \frac{1}{4}(n^4 - n^3 - n^2 + n), & n \text{ odd} \end{cases}$$

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

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

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

Где **chain** - запросы соединения от  $n$  таблиц, имеющие вид цепочки. **Cycle, star, clique** - цикл, звезда (одна вершина связана со многими, все остальные между собой не имеют связей), полный граф.

**DPsub** [<https://db.in.tum.de/teaching/ws2425/queryopt/main.pdf?lang=de>]

Разбивает граф соединений на подграфы, строит оптимальный порядок для каждого компонента и соединяет их. Также не поддерживает работу с внешними соединениями. Перебирает всевозможные подмножества таблиц  $S$ . Разделение множество  $S$  на  $S_1$  и  $S_2$ , так же как в DPsize.



---

```

1: Input: A set of relations  $R = \{R_1, \dots, R_n\}$  to be joined
2: Output: An optimal bushy join tree
3:  $B \leftarrow$  an empty DP table  $2^R \rightarrow$  join tree
4: for each  $R_i \in R$  do
5:    $B[\{R_i\}] \leftarrow R_i$ 
6: end for
7: for each  $1 < i \leq 2^n - 1$  ascending do
8:    $S \leftarrow \{R_j \in R \mid (\lfloor i/2^{j-1} \rfloor \bmod 2) = 1\}$ 
9:   for each  $S_1, S_2 \subset S$  such that  $S_2 = S \setminus S_1$  do
10:    if (not cross products  $\wedge \neg S_1$  connected to  $S_2$ ) then
11:      continue
12:    end if
13:     $p_1 \leftarrow B[S_1], p_2 \leftarrow B[S_2]$ 
14:    if  $p_1 = \epsilon$  or  $p_2 = \epsilon$  then continue
15:    end if
16:     $P \leftarrow \text{CreateJoinTree}(p_1, p_2)$ 
17:    if  $B[S] = \epsilon$  or  $C(B[S]) > C(P)$  then
18:       $B[S] \leftarrow P$ 
19:    end if
20:  end for
21: end for

```

---

Аналогично, имеются лучшие порядки соединений  $p_1$  и  $p_2$ , и если дерево, построенное из  $p_1$  и  $p_2$  дешевле старого варианта то происходит замена.

$$I_{\text{DPsub}}^{\text{chain}}(n) = 2^{n+2} - n^2 - 3n - 4$$

$$I_{\text{DPsub}}^{\text{cycle}}(n) = n2^n + 2^n - 2n^2 - 2$$

$$I_{\text{DPsub}}^{\text{star}}(n) = 2 \times 3^{n-1} - 2^n$$

$$I_{\text{DPsub}}^{\text{clique}}(n) = 3^n - 2^{n+1} + 1$$

### DPccp

[<https://db.in.tum.de/teaching/ws2425/queryopt/main.pdf?lang=de>][<https://dsg.uwaterloo.ca/se>]

Пусть дан граф соединений  $G = (V, E)$ . Введём понятие csg-сmp-pair  $(S_1, S_2)$  – в графе запроса выдели  $S_1, S_2$  связные непересекающиеся подграфы, такие что  $S_1 \in V, S_2 \in V/S_1$  существует между ними существует ребро( условие соединения).

Заметим, что если  $(S_1, S_2)$  - csg-сmp-пара, то и  $(S_2, S_1)$  – тоже. Мы ограничим перечисление csg-сmp-пар теми  $(S_1, S_2)$ , которые удовлетворяют условию, что  $\min(S_1) < \min(S_2)$ , где  $\min(S) = s$  такое, что  $s \in S$  и  $\forall s' \in S : s \neq s' \implies s < s'$ . Поскольку это ограничение будет действовать для всех csg-сmp-пар, перечисляемых нашей процедурой, не будет вычислено ни одной дублирующей csg-сmp-пары. Как следствие, мы должны позаботиться о том, чтобы наша процедура динамического программирования была полной: если применяемый нами бинарный оператор коммутативен, то процедура построения плана для  $S_1 \cup S_2$  из планов для  $S_1$  и  $S_2$  должна учитывать коммутативность.

Однако это не представляет особой сложности. Очевидно, что для того, чтобы быть корректным, любой алгоритм динамического программирования должен учитывать все csg-сmp-пары. Таким

образом, минимальное количество вызовов функции стоимости любого алгоритма динамического программирования в точности равно количеству csg-cmp-пар для данного гиперграфа. Заметим, что число связанных подграфов намного меньше числа csg-cmp-пар. Теперь задача состоит в том, чтобы перечислить csg-cmp-pairs эффективно и в порядке, приемлемом для динамического программирования. Задача может быть выражена более конкретно. Перед перечислением csg-cmp-pair  $(S_1, S_2)$ , все csg-cmp-pairs  $(S'_1, S'_2)$ , где  $S'_1 \subseteq S_1$  и  $S'_2 \subseteq S_2$  должны быть пронумерованы.

---

```

1: Input: A set of relations  $R = \{R_1, \dots, R_n\}$ 
2: Output: An optimal bushy join tree
3:  $B \leftarrow$  an empty DP table  $2^R \rightarrow$  join tree
4: for each  $R_i \in R$  do
5:    $B[\{R_i\}] \leftarrow R_i$ 
6: end for
7: for each csg-cmp-pairs  $(S_1, S_2)$ ,  $S = S_1 \cup S_2$  do
8:    $p_1 \leftarrow B[S_1], p_2 \leftarrow B[S_2]$ 
9:    $P \leftarrow \text{CreateJoinTree}(p_1, p_2)$ 
10:  if  $B[S] = \epsilon$  or  $C(B[S]) > C(P)$  then
11:     $B[S] \leftarrow P$ 
12:  end if
13: end for
14: return  $B[\{R_0, \dots, R_{n-1}\}]$ 

```

---

Сначала алгоритм инициализирует начальные значения  $B[R_i] = R_i$ . Затем перебираются такие подмножества вершин  $S_1$  и  $S_2$ , что  $(S_1, S_2)$  – csg-cmp-pair и  $B[S_1], B[S_2]$  уже известны. Берутся оптимальные планы  $p_1$  и  $p_2$  от  $S_1$  и  $S_2$ , затем они соединяются. Если стоимость получившегося плана больше  $B[S]$ , то происходит замена. Ограничения:

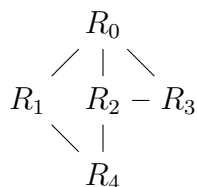
- Не поддерживает внешние соединения, из-за разбиения на csg-cmp-pairs.
- Сильно зависит от эффективности поиска csg-cmp-pair.

Данная реализация возможна, только если удастся перечислить все csg-cmp-pairs, т.к DPccp перебирает только связанные графы и их комплементарные пары. Нужно выполнить несколько задач:

- Эффективно пронумеровать все связанные графы (CSG).
- Не пересчитывать CSG дважды, пример  $(S_1, S_2)$  и  $(S_2, S_1)$ .
- Эффективно находить CCP для каждой CSG.

Предварительно проведём поиск в ширину и пронумеруем все узлы:

$V = R_0, \dots, R_{n-1}$  – избавимся от повторного добавления  $(S_1, S_2)$  и  $(S_2, S_1)$ . Пусть  $G = V, E$  – граф запроса. Введём множества  $B_i = v_j | j \leq i$  и функцию соседства  $N(V') = v' | v \in V' (v, v') \in E$



---

**Algorithm EnumerateCsg**

```

1: for  $i = n - 1$  to 0 do
2:   emit  $u_i$ ;
3:   EnumerateCsgRec( $G, u_i, B_i$ );
4: end for
5: EnumerateCsgRec( $G, S, X$ );
6:  $N \leftarrow \mathcal{N}(S) \setminus X$ ;
7: for each  $S' \subseteq N$  where  $S' \neq \emptyset$  do
8:   emit ( $S, S'$ )
9: end for
10: for each  $S' \subseteq N$  where  $S' \neq \emptyset$  do
11:   EnumerateCsgRec( $G, (S \cup S'), (X \cup N)$ )
12: end for

```

---

$S$	$X$	$N$	emit/ $S$
{4}	{0, 1, 2, 3, 4}	$\emptyset$	
{3}	{0, 1, 2, 3}	{4}	{3, 4}
{2}	{0, 1, 2}	{3, 4}	{2, 3} {2, 4} {2, 3, 4}
{1}	{0, 1}	{4}	{1, 4}
$\rightarrow$ {1, 4}	{0, 1, 4}	{2, 3}	{1, 2, 4} {1, 3, 4} {1, 2, 3, 4}
{0}	{0}	{1, 2, 3}	{0, 1} {0, 2} {0, 3}
$\rightarrow$ {0, 1}	{0, 1, 2, 3}	{4}	{0, 1, 4}
$\rightarrow$ {0, 2}	{0, 1, 2, 3}	{4}	{0, 2, 4}
$\rightarrow$ {0, 3}	{0, 1, 2, 3}	{4}	{0, 3, 4}

---

**Algorithm EnumerateCmp**

```

1: Input: A connected query graph  $G = (V, E)$ , a connected subset  $S_1$ 
2: Precondition: Nodes in  $V$  are numbered according to a BFS
3: Output: Emits all complements  $S_2$  for  $S_1$  such that  $(S_1, S_2)$  is a csg-cmp-pair
4:  $X \leftarrow B_{\min(S_1)} \cup S_1$ 
5:  $N \leftarrow \mathcal{N}(S) \setminus X$ 
6: for each  $u_i \in N$  by descending  $i$  do
7:   emit  $u_i$ 
8:   EnumerateCmpRec( $G, \{u_i\}, X \cup N$ )
9: end for

```

---

Вместе эти алгоритмы позволяют эффективно находить сср для каждой csg. Сложность:

$$I_{\text{DPchain}}^{\text{chain}}(n) = n^3$$

$$I_{\text{DPccp}}^{\text{cycle}}(n) = n^3$$

$$I_{\text{DPccp}}^{\text{star}}(n) = n2^n$$

$$I_{\text{DPccp}}^{\text{clique}}(n) = 3^n$$

Можно заметить, что также как и DPsize и DPsub, DPccp не поддерживает внешние соединения. Однако он на каждом типе графа имеет меньшую сложность и имеет более сложную реализацию чем DPsize и DPsub.