

Nội dung

- Xử lý truy vấn phân tán
 - Phân rã và cục bộ hóa truy vấn
 - Tối ưu hóa truy vấn phân tán
 - Trình tự kết nối
 - Xử lý truy vấn thích ứng

31

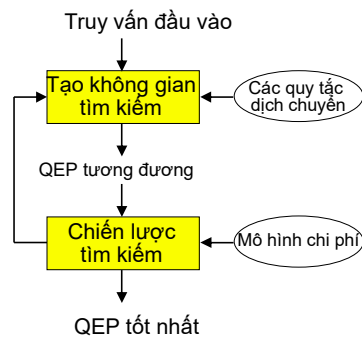
Bước 3 – Tối ưu hóa truy vấn toàn cục

Đầu vào: Truy vấn phân mảnh

- Tìm lịch biểu toàn cục *tốt nhất* (không cần thiết phải tối ưu)
 - Tối thiểu hóa hàm chi phí
 - Xử lý kết nối phân tán
 - Cây hỗn hợp với cây tuyến tính
 - Quan hệ nào được dịch chuyển tới đâu?
 - Dịch chuyển toàn bộ hay dịch chuyển theo nhu cầu
 - Quyết định việc sử dụng các phép nối nửa
 - Phép nối nửa tiết kiệm chi phí truyền thông và chi phí xử lý cục bộ nhiều hơn
 - Các phương thức kết nối
 - Vòng lặp lồng nhau, nối hợp nhất (merge join) hoặc nối băm (hash join)

32

Quá trình tối ưu hóa truy vấn



33

Các thành phần

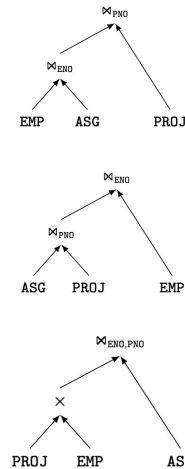
- Không gian tìm kiếm
 - Tập các biểu thức đại số tương đương (cây truy vấn)
- Mô hình chi phí
 - Chi phí I/O + chi phí CPU + chi phí truyền thông
 - Có thể có trọng số khác nhau trong các môi trường phân tán khác nhau (LAN so với WAN)
 - Cũng có thể tối đa hóa thông lượng
- Thuật toán tìm kiếm
 - Làm thế nào có thể đi vào bên trong không gian giải pháp?
 - Tìm kiếm toàn diện, các thuật toán heuristic (cải tiến lặp, ủ mô phỏng, di truyền,...)

34

Cây kết nối

- Đặc trưng hóa không gian tìm kiếm để tối ưu hóa
- Cho N quan hệ, Có thể có $O(N!)$ cây nối tương đương bằng cách áp dụng các quy tắc giao hoán và kết hợp

```
SELECT ENAME, RESP
FROM EMP
NATURAL JOIN ASG
NATURAL JOIN PROJ
```

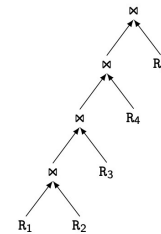


35

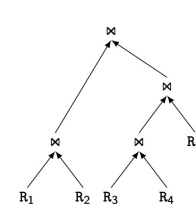
Cây kết nối

- Hai dạng chính
 - Cây tuyến tính và cây hỗn hợp

Cây kết nối tuyến tính



Cây kết nối hỗn hợp (Bushy)



36

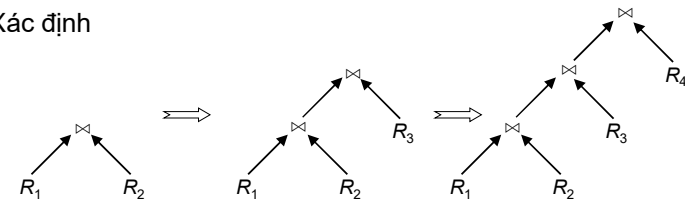
Chiến lược tìm kiếm

- Cách di chuyển trong không gian tìm kiếm
- Xác định
 - Bắt đầu từ các quan hệ cơ sở và xây dựng kế hoạch bằng cách thêm một quan hệ ở mỗi bước
 - Lập trình động: theo chiều rộng
 - Giải thuật tham lam: theo chiều sâu
- Ngẫu nhiên
 - Tìm kiếm sự tối ưu xung quanh một điểm xuất phát cụ thể
 - Thời gian tối ưu hóa giao dịch cho thời gian thực thi
 - Tốt hơn khi > 10 quan hệ
 - Ủ mô phỏng
 - Cải tiến lặp

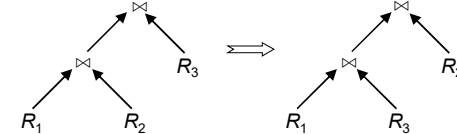
37

Chiến lược tìm kiếm

- Xác định



- Ngẫu nhiên



38