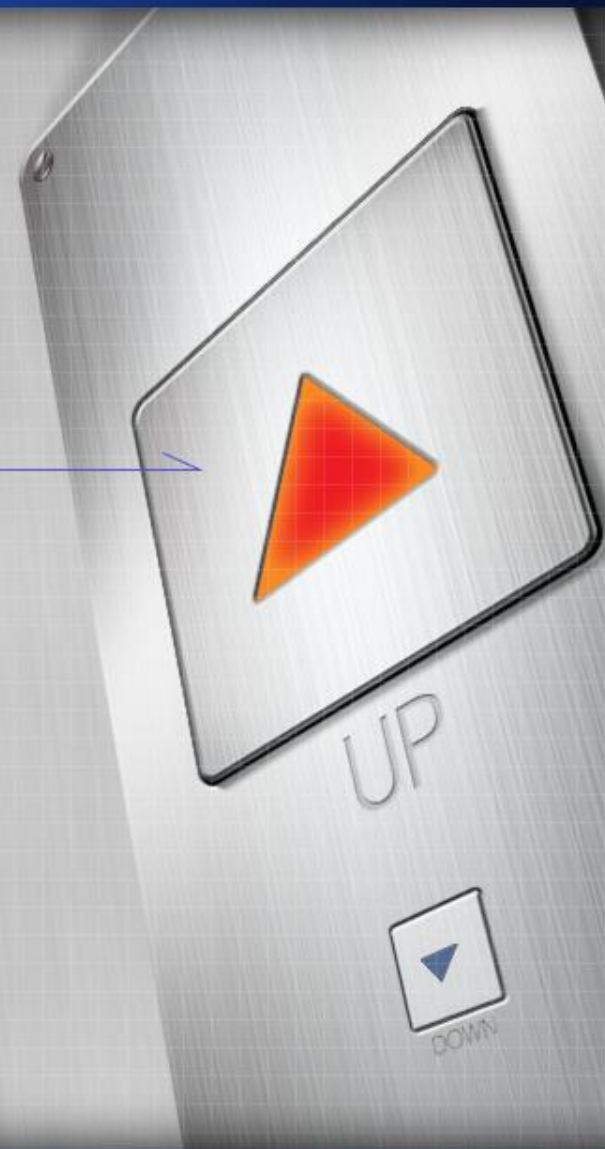


Nhóm 7

Đào Thế Quốc	B18DCCN497
Hoàng Công Thiện	B18DCCN637
Hoàng Văn Tú	B18DCCN548
Nguyễn Thị Hồng Duyên	B18DCCN116
Đào Thị Thanh	B18DCCN600
Đinh Vũ Long	B18DCCN339
Lê Trần Quang Huy	B18DCCN266
Nguyễn Xuân Bách	B18DCCN048
Nguyễn Thế Anh	B18DCCN026
Nguyễn Văn Duy	B18DCCN111
Đỗ Mạnh Hùng	B18DCCN252



LOGO

Chủ đề:

Tối ưu hóa truy vấn

1. Định nghĩa

2. Phân loại

3. Tính hiệu quả trong việc tối ưu hoá truy vấn

4. Các phương pháp tối ưu hoá truy vấn

5. Các kỹ thuật tối ưu hoá truy vấn

6. Các bước tối ưu hoá trong csdl tập trung

1. Định nghĩa

- Truy vấn là 1 biểu thức được biểu diễn bằng một ngôn ngữ thích hợp và dùng để xác định một phần dữ liệu được chứa trong cơ sở dữ liệu quan hệ.
- Một truy vấn có thể được biểu diễn bởi 1 cây toán tử. Chuỗi các phép toán đại số quan hệ sẽ được thực hiện từ nút lá đến nút gốc để tạo kết quả truy vấn.

*Ví dụ

$$\begin{array}{c} ||_{Tenlop} \\ | \\ \sigma_{malop='MT'} \\ | \\ lop \end{array}$$

1. Định nghĩa

- Trong hệ cơ sở dữ liệu phân tán, tối ưu hoá truy vấn nhằm mục đích lựa chọn 1 giải pháp tốt nhất trong toàn bộ các giải pháp truy vấn có thể thực hiện.



- Một phương pháp tối ưu hoá truy vấn có hiệu quả là tìm kiếm trong tập các giải pháp, dự đoán chi phí của chúng sau đó lựa chọn giải pháp có chi phí nhỏ nhất.

2. Phân Loại

Tối ưu hóa truy vấn tĩnh

Tối ưu hóa truy vấn động

Phương pháp truy vấn hỗn hợp

2. Phân Loại

Tối ưu hóa truy vấn tĩnh

- Việc tối ưu hoá truy vấn tĩnh được thực hiện tại thời điểm biên dịch truy vấn.
- Trước khi thực hiện truy vấn, kích thước của các quan hệ trung gian không được biết trước

2. Phân Loại

Tối ưu hóa truy vấn động

- Việc tối ưu hoá truy vấn động được tiến hành vào thời gian thực hiện truy vấn.
- Việc lựa chọn thao tác tiếp theo tốt nhất cho tối ưu hoá truy vấn dựa trên những thông tin chính xác về kết quả của các thao tác thực hiện trước đó.

2. Phân Loại

* Ưu điểm:

- Kích thước thực sự của các quan hệ trung gian phù hợp cho bộ xử lý truy vấn. Vì vậy nó giảm thiểu xác suất của việc lựa chọn một giải pháp tồi.

* Nhược điểm:

- Các thao tác tối ưu hoá có chi phí cao. Lặp lại nhiều lần cho mỗi thao tác.

2. Phân Loại

Phương pháp truy vấn hỗn hợp

- Phương pháp tối ưu hoá truy vấn hỗn hợp có các ưu điểm của tối ưu hoá truy vấn tĩnh, tránh được các vấn đề được tạo ra bởi các đánh giá không chính xác gây ra.
- Về cơ bản phương pháp này là tĩnh nhưng quá trình truy vấn động có thể diễn ra

3. Tính hiệu quả

- Tính hiệu quả của việc tối ưu hoá truy vấn dựa trên các số liệu thống kê về cơ sở dữ liệu.
- Tối ưu hoá truy vấn động cần đến các số liệu thống kê nhằm chọn các thao tác cần phải thực hiện trước tiên.
- Tối ưu hoá truy vấn tĩnh cần đến các số liệu thống kê nhằm chọn các thao tác cần phải thực hiện trước tiên.
- Các số liệu thống kê dành cho việc tối ưu hoá truy vấn có liên quan đến các mảnh, lực lượng và kích thước của mảnh, cũng như kích thước và số lượng các giá trị phân biệt của mỗi thuộc tính.
- Độ chính xác của số liệu thống kê phụ thuộc vào việc cập nhật theo chu kỳ.

4. Các phương pháp tối ưu hoá truy vấn

- Việc tối ưu hóa truy vấn có tầm ảnh hưởng quan trọng đến mức độ hiệu quả của cơ sở dữ liệu.
- Các phương pháp tối ưu hóa truy vấn:

Phương pháp 1 : Tối ưu hóa về chi phí.

- Chi phí thời gian CPU thực hiện truy vấn.
- Chi phí lưu giữ kết quả trung gian.
- Chi phí trao đổi giữa bộ nhớ trong và bộ nhớ ngoài.

Phương pháp 2 : Tối ưu hóa Heuristic.

- Biến đổi biểu thức đại số quan hệ.
- Biến đổi thứ tự thực hiện các phép toán của biểu thức đại số quan hệ

4. Các phương pháp tối ưu hoá truy vấn

So sánh giữa 2 phương pháp:

- Nhìn chung, phương pháp Heuristic sẽ tăng hiệu quả làm việc của hầu hết các truy vấn.
- Phép tối ưu hóa chi phí với ước tính từ thống kê chỉ phù hợp đối với 1 số trường hợp truy vấn cụ thể.
- Để tối ưu hóa truy vấn, thực tế, mọi người thường sử dụng kết hợp của 2 phương pháp trên với nhau.

5. Các kỹ thuật tối ưu hoá truy vấn

Kỹ thuật tối ưu hóa chi phí:

- Liệt kê các kế hoạch truy vấn có tiềm năng là tốt nhất
- Áp dụng các phép biến đổi để tạo ra tất cả các kế hoạch khả thi
- Tính toán giá dựa trên giá mẫu
- Chọn kế hoạch với giá ít nhất

5. Các kỹ thuật tối ưu hoá truy vấn

Kỹ thuật tối ưu hóa Heuristic.

- Đầu tiên là cần thực hiện các phép toán làm giảm kích thước của các kết quả trung gian.
- Đẩy phép chọn và phép chiếu xuống thực hiện sớm nhất có thể
- Nhóm dãy các phép chọn và chiếu:
- Kết hợp phép chọn và tích Đề các thành phép kết nối
- Tìm các biểu thức con chung trong biểu thức đại số quan hệ để đánh giá chỉ một lần.
- Ước lượng chi phí và lựa chọn thứ tự thực hiện

5. Các kỹ thuật tối ưu hoá truy vấn

- Một số các phép biến đổi có ích trong việc tối ưu hóa truy vấn

Các phép biến đổi tương đương

1. Phân phối σ

$$\sigma_{C_1 \text{ AND } C_2 \text{ AND } \dots \text{ AND } C_n}(R) \equiv \sigma_{C_1}(\sigma_{C_2}(\dots(\sigma_{C_n}(R))\dots))$$

2. Giao hoán của σ

$$\sigma_{C_1}(\sigma_{C_2}(R)) \equiv \sigma_{C_2}(\sigma_{C_1}(R))$$

3. Phân phối của Π

$$\pi_{list1}(\pi_{list2}(\dots(\pi_{listn}(R))\dots)) \equiv \pi_{list1}(R)$$

4. Giao hoán σ với Π

$$\pi_{A_1, A_2, \dots, A_n}(\sigma_C(R)) \equiv \sigma_C(\pi_{A_1, A_2, \dots, A_n}(R))$$

C involves only A_1, \dots, A_n

5. Các kỹ thuật tối ưu hoá truy vấn

Các phép biến đổi tương đương

7. Giao hoán π với \bowtie (or \times)

$$\pi_L (R \bowtie_C S) \equiv (\pi_{A_1, \dots, A_n} (R)) \bowtie_C (\pi_{B_1, \dots, B_m} (S))$$

$$L = \{ A_1, \dots, A_n, B_1, \dots, B_m \}$$

join condition C only involves L

General Form

$$\begin{aligned} \pi_L (R \bowtie_C S) \equiv \\ \pi_L ((\pi_{A_1, \dots, A_n, A_{n+1}, \dots, A_{n+k}} (R)) \bowtie \\ (\pi_{B_1, \dots, B_m, B_{m+1}, \dots, B_{m+p}} (S)) \end{aligned}$$

5. Các kỹ thuật tối ưu hoá truy vấn

Các phép biến đổi tương đương

8. Giao hoán của các phép toán tập hợp

\cup and \cap

9. Kết hợp các phép toán \bowtie , \cup , \cap

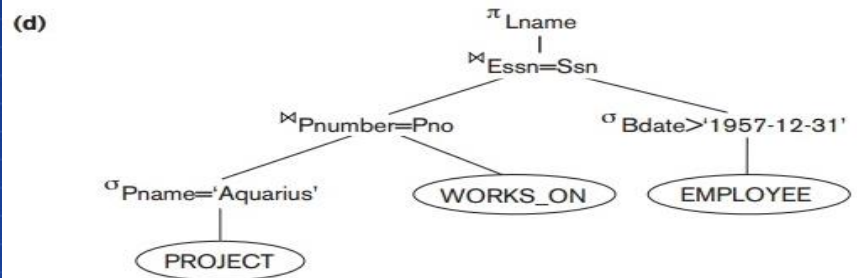
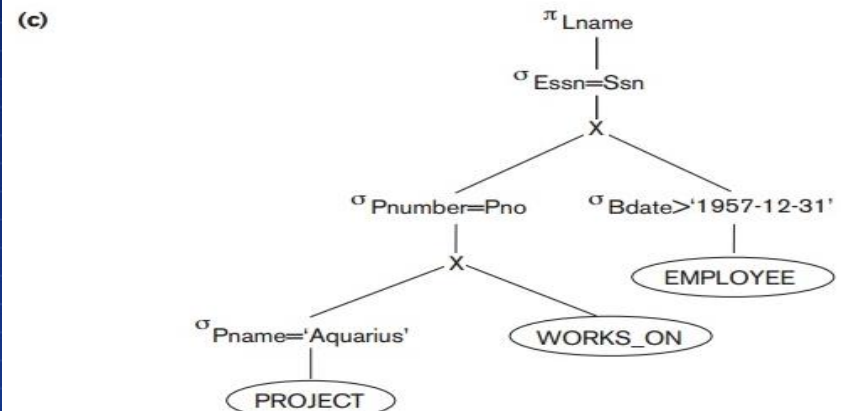
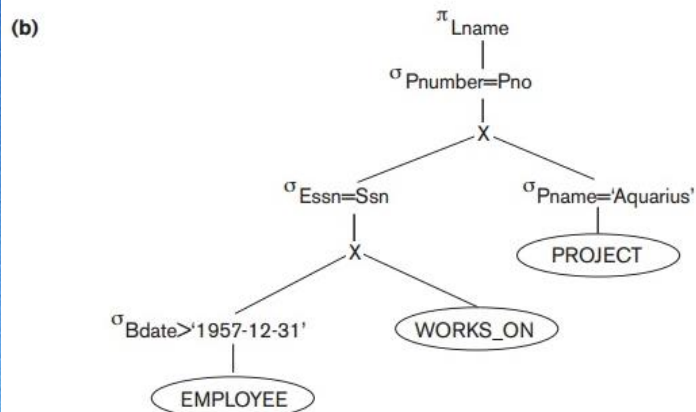
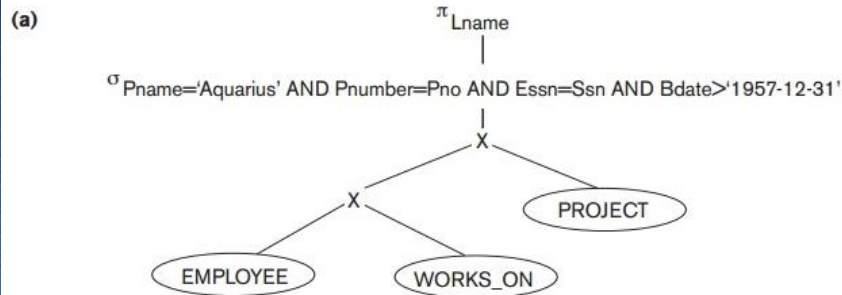
$$(R \bowtie S) \bowtie T \equiv R \bowtie (S \bowtie T)$$

10. Kết hợp σ với các phép toán tập hợp

$$\sigma_c (R \bowtie S) \equiv (\sigma_c (R)) \bowtie (\sigma_c (S))$$

\bowtie : \cup , \cap , $-$

Ví dụ tối ưu hoá truy vấn



a - Câu truy vấn chưa tối ưu

b, c, d - Câu truy vấn đã tối ưu

6. Các bước tối ưu hoá trong csdl tập trung

Bước 1: Kiểm tra ngữ pháp

VD: Xét truy vấn Q1:

SELECT ten FRPM truong;

Truy vấn này sai cú pháp (FROM not FRPM)

Bước 2: Kiểm tra sự hợp lệ

Kiểm tra sự tồn tại và hợp lệ của các đối tượng dữ liệu

VD: Xét truy vấn Q2:

SELECT ten FROM truong1;

Truy vấn này có bảng truong1 không tồn tại trong csdl.

Xét truy vấn Q3:

SELECT matruong FROM truong

WHERE matruong = 'a';

Truy vấn này không hợp lệ vì matruong phải là kiểu integer

6. Các bước tối ưu hoá trong csdl tập trung

Bước 3: Dịch truy vấn

- Trong bước này, hệ quản trị dữ liệu sẽ biến đổi truy vấn hợp lệ này thành 1 dạng biểu diễn bên trong hệ thống ở mức thấp hơn mà nó có thể sử dụng được.
- Thông thường ta thường sử dụng đại số quan hệ vì nó dễ dàng biến đổi thành tác vụ của hệ thống.

VD: Xét truy vấn Q4:

```
SELECT DISTINCT tentruong  
FROM truong  
WHERE matruong = '236'
```

Truy vấn này sẽ được biến đổi thành biểu thức đại số quan hệ như sau:

$$\Pi_{tentruong}(\sigma_{matruong = '236'}(truong))$$

6. Các bước tối ưu hoá trong csdl tập trung

Bước 4: Tối ưu hoá biểu thức đại số quan hệ

Các phép biến đổi tương đương

1. Phân phối σ
 $\sigma_{C_1 \text{ AND } C_2 \text{ AND } \dots \text{ AND } C_n}(R) \equiv \sigma_{C_1}(\sigma_{C_2}(\dots(\sigma_{C_n}(R))\dots))$
2. Giao hoán của σ
 $\sigma_{C_1}(\sigma_{C_2}(R)) \equiv \sigma_{C_2}(\sigma_{C_1}(R))$
3. Phân phối của π
 $\pi_{list1}(\pi_{list2} \dots (\pi_{listn}(R))\dots) \equiv \pi_{list1}(R)$
4. Giao hoán σ với π
 $\pi_{A1, A2, \dots, An}(\sigma_C(R)) \equiv \sigma_C(\pi_{A1, A2, \dots, An}(R))$
C involves only A1, ..., An

17

Hệ quản trị CSOL © BM HTTT

Các phép biến đổi...

5. Giao hoán của \bowtie (or \times)
 $R \bowtie_C S \equiv S \bowtie_C R$
meaning
6. Giao hoán σ với \bowtie (or \times)
- $\sigma_C(R \bowtie S) \equiv (\sigma_C(R)) \bowtie S$
attributes in C involve only attributes of R
- $\sigma_C(R \bowtie S) \equiv (\sigma_{C_1}(R)) \bowtie (\sigma_{C_2}(S))$
C1 (C2) involves only attribute of R(S)

18

Hệ quản trị CSOL © BM HTTT

Các phép biến đổi...

7. Giao hoán π với \bowtie (or \times)
 $\pi_L(R \bowtie_C S) \equiv (\pi_{A1, \dots, An}(R)) \bowtie_C (\pi_{B1, \dots, Bm}(S))$
 $L = \{A1, \dots, An, B1, \dots, Bm\}$
join condition C only involves L

General Form

$$\pi_L(R \bowtie_C S) \equiv \pi_L((\pi_{A1, \dots, An, An+1, \dots, An+k}(R)) \bowtie (\pi_{B1, \dots, Bm, Bm+1, \dots, Bm+p}(S)))$$

19

Hệ quản trị CSOL © BM HTTT

Các phép biến đổi...

8. Giao hoán của các phép toán tập hợp
 U and \cap
9. Kết hợp các phép toán \bowtie , U , \cap
 $(R \bowtie S) \sqcup T \equiv R \sqcup (S \bowtie T)$
10. Kết hợp σ với các phép toán tập hợp
 $\sigma_C(R \sqcup S) \equiv (\sigma_C(R)) \sqcup (\sigma_C(S))$
 $\sqcup : U, \cap, -$

20

Hệ quản trị CSOL © BM HTTT

6. Các bước tối ưu hoá trong csdl tập trung

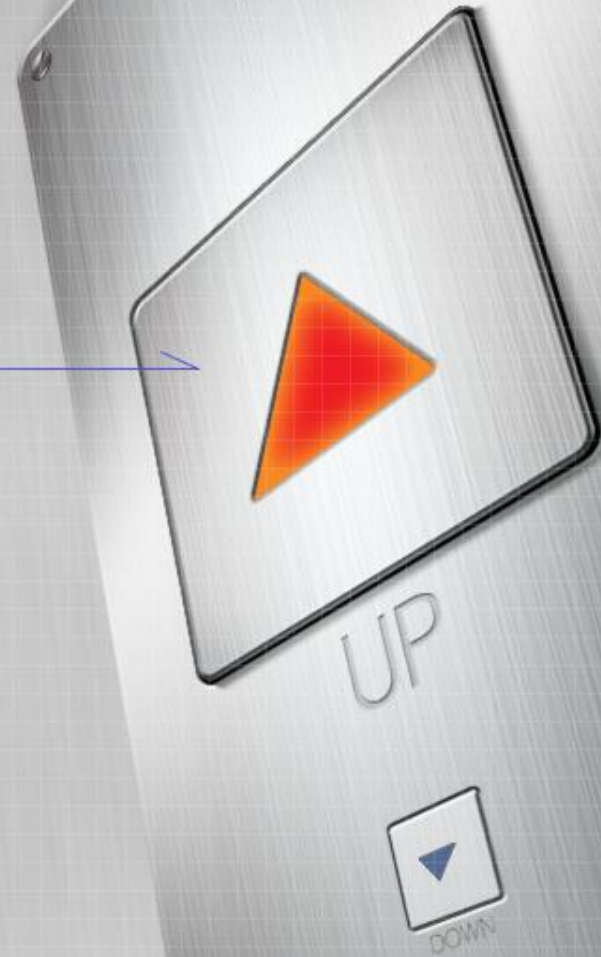
Bước 5: Chọn chiến lược truy xuất

- Hệ quản trị dữ liệu sử dụng các thông số về kích thước các bảng, các chỉ mục,... để xác định cách xử lý truy vấn
- Hệ quản trị dữ liệu sẽ đánh giá để chọn kế hoạch có chi phí ít nhất (thời gian xử lý và vùng nhớ trung gian nhỏ nhất)
- Thông số đánh giá chi phí: số lần và loại truy xuất đĩa, kích thước vùng nhớ, thời gian thực hiện.

Bước 6: Tạo sinh mã code

Ở bước này, kế hoạch thực hiện của truy vấn cuối bước 5 sẽ được mã hoá và thực hiện.

Thank you!



LOGO