TỐI ƯU TRUY VẤN VÀ CÁC LỰA CHỌN.

Muốn học piano có các cách sau: tìm 1 giáo viên piano giỏi và học từng bài 1, cách 2: tự học bằng sách, video, hoặc bắt đầu bằng chơi những bản nhạc đơn giản, vd: piano for dummies;

Chương này bàn về các khái niệm cơ bản của tối ưu hoá truy vấn mà cần hiểu sự tác động qua lại giữa thiết kế vật lý và quá trình truy vấn. ta bắt đầu với 1 ví dụ đơn giản của tối ưu truy vấn và minh hoạ các kỹ thuật ước lượng thời gian I/O được dùng để xác định hiệu suất của query nào tốt nhất.

1. Quá trình truy vấn và tối ưu

Các bước cơ bản của quá trình truy vấn:

* Duyệt tìm, phân tích ngữ pháp và phân rã của 1 câu truy vấn: bước này kiểm tra cú pháp của truy vấn, nếu cú pháp sai sẽ thông báo lỗi cho người dùng. Đầu ra của bước này là 1 dạng trung gian của truy vấn được biết như 1 cây truy vấn hoặc dự án xử lý truy vấn.
* Tối ưu truy vấn:bước này bao gồm tối ưu cục bộ và toàn cục. tối ưu toàn cục xác định loại kết và loại lựa chọn và phép chiếu liên quan với những phép kết. tối ưu cục bộ xác định phương pháp chỉ mục cho sự lựa chọn và phép kết.
* Tạo ra mã truy vẫn và xử lý: bước này dùng ngôn ngữ lập trình cổ điển và trình biên dịch để tạo mã xử lý.

1. Lợi ích của tính năng tối ưu trong hệ thống cơ sở dữ liệu.
2. Thay đổi ( hoặc viết lại) truy vấn.

Các quản trị dữ liệu sẽ viết lại các truy vấn để nó trở thành dạng truy vấn có hiệu suất hơn trước khi qua tối ưu. Điều này giúp sự lựa chọn thực thi rất nhiều. ví dụ các phép biến đổi thường dùng là: thay đổi các thành phần truy vấn thành các phép kết hoặc 1 phần là phép kết, việc đẩy các phép gom nhóm xuống dưới phép kết, việc loại trừ phép kết trên các khoá ngoại khi các bảng bao gồm các kết quả của phép kết mà không được dùng nhiều trong truy vấn, thay đổi kết ngoài thành kết trong khi chúng cho ra các kết quả như nhau, và thay thế view reference thành actual view.

1 phép biến đổi thường dùng nữa là viết lại khung nhìn một cụ thể. Nếu 1 vài phần của truy vấn mà tương đương với 1 khung nhìn cụ thể thì truy vấn đó được thay thế bằng khung nhìn đó..

Biến đổi truy vấn rát quen thuộc trong kho dữ liệu mà sử dụng dạng lược đồ sao.

1. Query execution plan viewing

Làm thế nào bạn biết được đề án nào được chọn cho những câu truy vấn thường dùng nhất trong cơ sở dữ liệu của bạn.? hầu hết các hệ quản trị đều cung cấp 1 vài công cụ cho người dùng biết được dự án truy cập (access plan). Công cụ này miêu tả tất cả các bước trong dự án, loại mà những bảng được truy cập cho truy vấn, và có cần dùng chỉ mục cho bảng này hay không. Việc tối ưu lựa chọn dự định tốt nhất trong số các dự định mà được ứng cử.

1. Tần số

Nhiều hệ thống cơ sở dữ liệu lưu tần số xuất hiện của dãy các giá trị thuộc tính từ dữ liệu thực để giúp ước lượng các tính chọn lọc của thao tác lựa chọn và kết.

1. Những lời gợi ý kế hoạch thực hiện truy vấn

Song song với việc giải thích, lời gợi ý kế hoạch trở thành vật cố định chính trong công nghiệp dữ liệu, việc chấp nhận ứng dụng người dùng để bắt sự lựa chọn dự định.

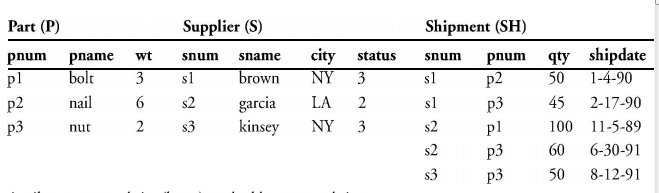
1. Tối ưu sâu.

Những quản trị dữ liệu khác nhau có độ sâu tìm kiếm khác nhau, đơn giản nhất là tìm kiếm kỹ.

1. Ước lượng chi phí truy vấn – ví dụ

Giả sử ta có 1 dự định thực thi 1 câu truy vấn đơn giản, và những điểm chính trên định lượng chi phí truy vấn đến thời gian truy xuất bé để xử lý câu truy ấn mà dự định của nó được để ra. Nó minh hoạ lợi ích của việc dùng chiến lược tối ưu truy vấn để giải quyết các vấn đề đơn giản ở ngoài đời thực. mặc dù vấn đề này còn cơ bản nhưng nó có lợi ích rất lớn trong thời gian truy vấn mà có thể sẽ được tiến đến việc dùng các luật, và định nghĩa của những luật này được minh hoạ rất rõ.

Ta giả sử có 3 bảng cơ sở dữ liệu được thể hiện cụ thể trong các bảng: part, supplier, intersection .



Tên và kích thước của thuộc tính , các bảng như sau:

* Supplier: snum(5), sname(20), city(10), status(2) 🡪 37 bytes trong 1 record của supplier.
* Part: pnum(8), pname(10), wt(5) --> 23 bytes trong 1 record của part
* Shipment: snum(5), pnum(8), qty(5), shipdate(8) 🡪 26 bytes trong 1 record trong shipment.

Chú ý: giả sử kích thước của block: 15000 bytes.

1. Ví dụ câu truy vấn cho cơ sở dữ liệu trê: những names nào của part được cung cấp bởi nhà cung cấp trong thành phố New York? Ta có truy vấn:

SELECT p.pname

FROM p, sh, s

WHERE P.PNUM = SH.PNUM

AND sh.snum = s.snum

AND s.city = ‘NY’;

* S kết SH kết P (1)
* SH kết S kết P (2)
* P kết SH kết S (3)
* SH kết P kết S (4)
* S X P kết SH (5)
* P X S kết SH (6)

Có 6 cách kết theo thứ tự kết khác nhau của: P, SH, S. phép kết ‘S kết SH kết P’ (1) và phép ‘ SH kết S kết P’ (2) thì tương đương nhau vì phép kết có tính chất giao hoán giữa 2 vế với nhau. Tương tự (3), (4) cũng tương đương nhau, (5) (6) tương đương nhau. Ta nên tránh lệnh (5) (6) nếu có thể vì chúng chứa các bộ cartesian của phép kết khi có sự chồng các cột. khi điều này xảy ra , kích thước của bảng kết quả sẽ tăng lên rất nhiều. và dữ liệu trong bảng kết quả cũng bị nối một cách tuỳ ý.

Xem lệnh (1) và (3) ta nên xét nên làm phép kết trước hay là pháp chọn trước. ta cũng xét nên tạo chỉ mục hay không. Giờ ta ước lượng chi phí cho 4 cách sau:

* Cách 1A: dùng lệnh 1 với kết trước, chọn sau va