ĐẠI HỌC ĐÀ NẴNG

TRƯỜNG ĐẠI HỌC SƯ PHẠM

KHOA TIN HỌC



CƠ SỞ DỮ LIỆU NÂNG CAO

ĐỀ TÀI: TỐI ƯU HÓA TRUY VẤN

GVHD: ThS. Phạm Dương Thu Hằng

Nhóm SVTH: Nhóm 3

Lê Văn Việt

Trần Thị Lệ Hằng

Trương Đức Thiệu

Nguyễn Gia Bảo

Nguyễn Trọng Hoàng

**Đà Nẵng 03/2024**

# 

# LỜI CẢM ƠN

Trong thời gian thực hiện đồ án với sự giúp đỡ tạo điều kiện của trường Đại học Sư Phạm – Đại học Đà Nẵng, sự góp ý của các bạn và đặc biệt là sự quan tâm hướng dẫn trực tiếp, chỉ bảo tận tình của cô giáo ThS. Phạm Dương Thu Hằng nhóm em đã hoàn thành đề tài “**TỐI ƯU HÓA TRUY VẤN**”

Mặc dù em đã rất cố gắng hoàn thành đồ án nhưng chắc chắn không thể tránh khỏi thiếu sót. Kính mong cô đóng góp ý kiến để em có thể hoàn thiện kiến thức của bản thân.

Nhóm em xin chân thành cảm ơn!

# NHẬN XÉT CỦA GIẢNG VIÊN HƯỚNG DẪN

*Đà Nẵng, ngày 05 tháng 03 năm 2024*

**Giáo viên hướng dẫn**

**ThS.Phạm Dương Thu Hằng**

# LỜI CAM ĐOAN

Nhóm em xin cam đoan:

Những nội dung trong báo cáo này là do nhóm em thực hiện dưới sự hướng dẫn trực tiếp của cô Phạm Dương Thu Hằng.

Mọi tham khảo dùng trong đồ án thực tập đều được trích dẫn rõ ràng và trung thực tên tác giả, tên công trình, thời gian, địa điểm công bố.

Mọi sao chép không hợp lệ, vi phạm quy chế đào tạo, hay gian trá, nhóm em xin chịu hoàn toàn trách nhiệm.

Đà Nẵng, ngày 05 tháng 03 năm 2024

**Nhóm sinh viên thực hiện**

**Nhóm 3**

# MỤC LỤC

[LỜI CẢM ƠN i](#_Toc160614598)

[NHẬN XÉT CỦA GIẢNG VIÊN HƯỚNG DẪN ii](#_Toc160614599)

[LỜI CAM ĐOAN iii](#_Toc160614600)

[MỤC LỤC iv](#_Toc160614601)

[DANH MỤC HÌNH ẢNH vi](#_Toc160614602)

[MỞ ĐẦU 1](#_Toc160614603)

[1. Lý do chọn đề tài 1](#_Toc160614604)

[2. Mục tiêu và nhiệm vụ đề tài 1](#_Toc160614605)

[2.1. Mục tiêu đề tài: 1](#_Toc160614606)

[2.2. Nhiệm vụ đề tài: 2](#_Toc160614607)

[3. Đối tượng và phạm vi nghiên cứu 2](#_Toc160614608)

[3.1. Đối tượng nghiên cứu 2](#_Toc160614609)

[3.2. Phạm vi nghiên cứu 2](#_Toc160614610)

[4. Nhiệm vụ thực hiện của đề tài 2](#_Toc160614611)

[4.1. Xây dựng cơ sở lý thuyết 2](#_Toc160614612)

[4.2. Phân tích thực trạng của vấn đề nghiên cứu 2](#_Toc160614613)

[4.3. Tiến hành nghiên cứu 2](#_Toc160614614)

[5. Phương pháp thực hiện 2](#_Toc160614615)

[6. Kết quả của đề tài 2](#_Toc160614616)

[6.1. Báo cáo đề tài 2](#_Toc160614617)

[6.2. Sản phẩm 2](#_Toc160614618)

[7. Bố cục của đề tài 3](#_Toc160614619)

[8. Kế hoạch thực hiện 3](#_Toc160614620)

[CHƯƠNG 1: Tổng Quan Về Xử Lý Truy Vấn 4](#_Toc160614621)

[1. Định nghĩa truy vấn và xử lý truy vấn 4](#_Toc160614622)

[2. Mục tiêu của tối ưu hóa truy vấn 4](#_Toc160614623)

[CHƯƠNG 2: Chuyển Đổi Truy Vấn Sang Biểu Thức Đại Số Quan Hệ 5](#_Toc160614624)

[1. Khái niệm biểu thức đại số quan hệ 5](#_Toc160614625)

[2. Phương pháp chuyển đổi sang đại số quan hệ 5](#_Toc160614626)

[3. Ưu và nhược điểm của việc sử dụng biểu thức đại số quan hệ 5](#_Toc160614627)

[CHƯƠNG 3: Tối Ưu Hóa Truy Vấn 6](#_Toc160614628)

[1. Cây truy vấn 6](#_Toc160614629)

[2. Quy tắc Heuristic 6](#_Toc160614630)

[3. Ước lượng chi phí 6](#_Toc160614631)

[CHƯƠNG 4: Tối Ưu Hóa Trong Quá Trình Viết Truy Vấn 7](#_Toc160614632)

[1. WHERE hay HAVING 7](#_Toc160614633)

[2. Thứ tự thực thi các phép nối 7](#_Toc160614634)

[TÀI LIỆU THAM KHẢO 8](#_Toc160614635)

# MỞ ĐẦU

## Lý do chọn đề tài

Tối ưu hóa truy vấn là một kĩ năng quan trọng đối với các Data Analyst, nguời khai thác dữ liệu từ CSDL với tần suất cao. Để cải thiện hiệu quả của các truy vấn, các lập trình viên và Data Analyst cần hiểu được công cụ tối ưu hóa truy vấn và và cách công cụ này chọn một lối truy cập và chuẩn bị kế hoạch tiến hành truy vấn. Việc điều chỉnh truy vấn có liên quan tới những kiến thức về các kĩ thuật như công cụ tối ưu theo chi phí và theo kinh nghiệm, cộng với những công cụ giúp phân tích kế hoạch thực thi truy vấn (execution plan) do CSDL cung cấp. Cách tốt nhất để điều chỉnh hiệu suất là tìm cách viết các truy vấn theo nhiều cách khác nhau, từ đó so sánh kế hoạch thực thi của mỗi truy vấn.

Là một lập trình viên, chắc hẳn mọi người không còn xa lạ với các câu lệnh SQL. Các câu lệnh SQL có thể giúp try xuất dữ liệu từ bất cứ CSDL nào. Tuy nhiên, trong quá trình truy xuất dữ liệu có thể gặp trường hợp những lệnh truy vấn rất chậm. Do vậy, chúng ta cần biết cách để tối ưu câu lệnh SQL nhanh hơn, mang lại hiệu suất làm việc cao hơn.

Đó là lý do nhóm em quyết định chọn đề tài “ **TỐI ƯU HÓA TRUY VẤN**” làm đề tài CƠ SỞ DỮ LIỆU NÂNG CAO .

## Mục tiêu và nhiệm vụ đề tài

### Mục tiêu đề tài:

Tìm ra cách hiệu quả nhất để thực hiện truy vấn, từ đó giảm thời gian phản hồi, giảm thiểu mức tiêu thụ tài nguyên và cải thiện hiệu suất của các ứng dụng cơ sở dữ liệu.

Cơ sở dữ liệu quan hệ xử lý một lượng lớn dữ liệu và việc làm như vậy một cách hiệu quả là rất quan trọng để duy trì ứng dụng có hiệu suất cao. Các truy vấn SQL được viết và thiết kế kém có thể ảnh hưởng đáng kể đến trải nghiệm người dùng vì chúng có thể làm chậm ứng dụng và tiêu tốn quá nhiều tài nguyên hệ thống. Hiểu và áp dụng các kỹ thuật tối ưu hóa truy vấn SQL có thể cải thiện đáng kể khả năng quản lý và truy xuất dữ liệu một cách hiệu quả và nhanh chóng.

### Nhiệm vụ đề tài:

* Tìm hiểu và nghiên cứu các phương pháp tối ưu hóa truy vấn
* Đánh giá các biểu thức Đại số quan hệ
* Tìm hiểu các chiến lược tối ưu và tổng quát

## Đối tượng và phạm vi nghiên cứu

### Đối tượng nghiên cứu

* Câu truy vấn SQL cho CSDL tập trung, và tối ưu hóa câu truy vấn đó để sinh ra các mảnh ở những vị trí khác nhau nhằm tối ưu hóa về chi phí thực hiện.
* Đối tượng nghiên cứu thuộc phạm vi nghiên cứu lý thuyết về tối ưu hóa, ứng dụng trong lĩnh vực giáo dục, đào tạo là chủ yếu.

### Phạm vi nghiên cứu

* Tập trung vào các phương pháp tối ưu hóa
* Kỹ thuật đồng bộ và tối ưu truy vấn trong cơ sỏ dữ liệu

## Nhiệm vụ thực hiện của đề tài

### Xây dựng cơ sở lý thuyết

### Phân tích thực trạng của vấn đề nghiên cứu

### Tiến hành nghiên cứu

## Phương pháp thực hiện

## Kết quả của đề tài

### Báo cáo đề tài

### Sản phẩm

## Bố cục của đề tài

* Chương 1: Tổng Quan Về Cơ Sở Dữ Liệu Tập Trung
* Tổng quan về xử lý truy vấn
* Chuyển đổi truy vấn sang biểu thức đại số quan hệ
* Chương 2: Tối Ưu Hóa Truy Vấn
* Cây truy vấn
* Quy tắc Heuristic
* Ước lượng chi phí
* Chương 3: Tối Ưu Hóa Trong Quá Trình Viết Truy Vấn
* WHERE hay HAVING
* Thứ tự thực thi các phép nối

## Kế hoạch thực hiện

# CHƯƠNG 1: Setup

## Xây dựng theo mô hình MVC

Xây dựng web theo mô hình mvc (model view controller)

Model: chịu trách nhiệm quản lý dữ liệu

View : lớp giao diện chính của ứng dụng

Controller :quản lí và điều phối luồng hoạt động của ứng dụng ( code chính)

## Database

A screenshot of a computer

Description automatically generated

Luồng dữ liệu:

**Người dùng đăng ký và đăng nhập**: Dữ liệu được chèn và truy vấn từ bảng users.

**Người dùng tạo thói quen**: Một bản ghi mới được thêm vào bảng habits, có thể kèm theo lịch hàng tuần hoặc hàng tháng trong habit\_week\_days và habit\_month\_days(lặp lại sẽ được xử lý qua hàm trong code theo quy tắc +7 hoặc +30 cho ngày hoặc tuần để tự động lặp lại).

**Các lần xuất hiện được theo dõi**: Khi thói quen diễn ra, các bản ghi được thêm vào habit\_week\_occurrences hoặc habit\_month\_occurrences(ở đây trường date chỉ lưu tuần bắt đầu nghĩa là hết 1 tuần thì nó mới thực hiện lưu dữ liệu khác).

**Người dùng đặt nhắc nhở**: Nhắc nhở được tạo và lưu trữ trong bảng reminders.

**Người dùng hoàn thành thói quen**: Chi tiết hoàn thành được ghi lại trong bảng completed\_habits( ở đây sẽ sử dụng luôn để thực hiện thống kê)

# CHƯƠNG 2: Code

## User

register :

login

logout

view information

## Habit

Create habit

Edit

Delete

## Ước lượng chi phí

Mục tiêu của trình tối ưu hóa truy vấn là tìm ra một chiến lược thực thi tốt nhất trong rất nhiều các chiến lược thực thi có thể có mà trình tối ưu hóa xem xét cho một truy vấn nhất định. Đối với một câu truy vấn đã từng được thực hiện trước đây, trình tối ưu hóa truy vấn sẽ thực thi câu truy vấn đó với chiến lược thực thi được cho là tốt nhất mà nó đã từng thực hiện. Ngược lại, đối với một câu truy vấn mới, nó phân tích tất cả các chiến lược thực thi có thể xảy ra, tạo ra chi tiết các bước thực hiện và thực hiện câu truy vấn. Các chiến lược thực thi mà trình tối ưu hóa phân tích được tiến hành dựa trên kết quả của biểu thức đại số quan hệ với các phép toán chọn, chiếu và tích Đề - các (là các toán tử quan hệ trong cây truy vấn). Việc tối ưu hóa một biểu thức đại số quan hệ như vậy bao gồm hai bước cơ bản và được mô tả như trên Hình 2.8:

Thứ nhất, liệt kê các kế hoạch truy vấn khả thi (nhiều các đường dẫn truy cập bảng và chỉ mục, các hoán vị nối (thứ tự nối), các phương thức nối, các phương thức gộp nhóm, v.v. khác nhau sẽ cung cấp các kế hoạch truy vấn khả thi khác nhau) để đánh giá biểu thức đại số quan hệ. Kế hoạch truy vấn là một cây các toán tử quan hệ và được thực hiện bằng cách gọi các toán tử theo thứ tự (có thể xen kẽ) nào đó. Mỗi toán tử có một hoặc nhiều đầu vào và một đầu ra, cũng là các nút trong của kế hoạch truy vấn và các bộ dữ liệu phải được chuyển giữa các toán tử theo cấu trúc cây của kế hoạch. Thông thường, một trình tối ưu hóa xem xét một tập con các kế hoạch khả thi vì số lượng toàn bộ các kế hoạch khả thi là rất lớn.

# CHƯƠNG 3: Tối Ưu Hóa Trong Quá Trình Viết Truy Vấn

## WHERE hay HAVING

## Thứ tự thực thi các phép nối

# TÀI LIỆU THAM KHẢO