**vb**



**TRƯỜNG ĐẠI HỌC SÀI GÒN**

**KHOA CÔNG NGHỆ THÔNG TIN**

A picture containing text, clipart

Description automatically generated

**TIỂU LUẬN**

**CƠ SỞ DỮ LIỆU NÂNG CAO**

MSSV: 3120410297

Họ tên: Trần Nguyên Lộc

Lớp môn học: DCT1201

Năm học: 2022-2023, Học kì: 1

Giảng viên giảng dạy: PGS. TS .Nguyễn Tuấn Đăng

***Thành phố Hồ Chí Minh - Tháng 05/2022***

**Mục lục**

[**1. MỤC TIÊU ĐỀ TÀI TIỂU LUẬN** 1](#_Toc121432320)

[1.1 Lí do chọn đề tài 1](#_Toc121432321)

[1.2 Đối tượng nghiên cứu 2](#_Toc121432322)

[1.3 Mục tiêu của việc nghiên cứu 2](#_Toc121432323)

[**2. CƠ SỞ LÝ THUYẾT** 3](#_Toc121432324)

[2.1 Lý thuyết tổng quát 3](#_Toc121432325)

[2.2 Primary Index 4](#_Toc121432326)

[2.3 Clustering Index 5](#_Toc121432327)

[2.4 Secondary Index 8](#_Toc121432328)

[**3. GIẢI QUYẾT CÁC BÀI TOÁN** 9](#_Toc121432329)

[3.1 Lời giải các bài toán 9](#_Toc121432330)

[3.2 Các lưu đồ thuật toán 9](#_Toc121432331)

[**4. CHƯƠNG TRÌNH BÀI TOÁN** 10](#_Toc121432332)

[**5. TÀI LIỆU THAM KHẢO** 12](#_Toc121432333)

# **1. MỤC TIÊU ĐỀ TÀI TIỂU LUẬN**

## 1.1 Lí do chọn đề tài

Ngày nay, Cơ sở dữ liệu là một phần cốt lõi thiết yếu của hầu hết các kiến trúc phần mềm, chúng ta cần cơ sở dữ liệu để lưu trữ thông tin ứng dụng, thông tin khách hàng, thông tin cá nhân người sử dụng hay là thông tin quan trọng của công ty và việc thiết kế một cơ sở dữ liệu tốt cũng đồng nghĩa với việc sử dụng chúng sẽ tốt hơn rất nhiều, một cơ sở dữ liệu được thiết kế tốt sẽ góp phần giảm bớt thời gian truy vấn dữ liệu, hạn chế rủi ro từ việc truy vấn và góp phần tạo thêm lợi ích cho cơ sở dữ liệu đó.

Cơ sở dữ liệu được thiết kế tốt không chỉ quan trọng mà một phần của quy trình thiết kế còn liên quan đến việc đảm bảo rằng cấu trúc của hệ thống được phát triển đủ hiệu quả để đáp ứng yêu cầu của người dùng hiện tại và trong tương lai. Việc điều chỉnh cơ sở dữ liệu một cách có kĩ thuật có thể cải thiện hiệu suất, do đó trong bài báo cáo này chúng ta sẽ xem xét các vấn đề về thiết kế chỉ mục và các cách tích hợp để sử dụng chỉ mục. Các chỉ mục trong cơ sở dữ liệu đóng một vai trò tương tự như một quyển sách và trong đó các chỉ mục được sử dụng để tăng tốc độ truy cập thông tin. Các cấu trúc tẹp có thể bị ảnh hưởng bởi các kĩ thuật lập chỉ mục khác nhau và chúng sẽ có ảnh hưởng đến hiệu suất cơ sở dữ liệu. Các chỉ mục có thể giúp các nhà phát triển cơ sở dữ liệu xây dựng cấu trúc tệp hiệu quả và đưa ra các phương pháp truy cập hiệu quả. Khi được sử dụng và điều chỉnh đúng cách, hiệu suất cơ sở dữ liệu có thể được cải thiện hơn nữa. Trên thực tế, các chỉ mục có lẽ là cơ chế quan trọng nhất có sẵn rõ ràng cho các nhà phát triển cơ sở dữ liệu và quản trị viên để điều chỉnh hiệu suất của cơ sở dữ liệu.

## 1.2 Đối tượng nghiên cứu

Đối tượng nghiên cứu của bài báo cáo này tập trung vào phần Single Level, trong đó bao gồm các thành phần chỉ mục sau:

* Primary Index
* Clustering Index
* Secondary Index

## 1.3 Mục tiêu của việc nghiên cứu

Mục tiêu của việc nghiên cứu nhằm giúp người đọc hiểu hơn về cách các chỉ mục hoạt động như thế nào trong CSDL và cũng như việc ứng dụng của nó trong thực tế như thế nào, cũng như sẽ có ví dụ cho việc ứng dụng bằng các bài toán kĩ thuật trong bài viết.

# **2. CƠ SỞ LÝ THUYẾT**

## 2.1 Lý thuyết tổng quát

Chỉ mục là các kĩ thuật được sử dụng để lưu trữ một lượng lớn dữ liệu có cấu trúc trên đĩa. Những kĩ thuật này rất quan trọng đối với các nhà phát triển thiết kế CSDL, ví dụ như: DBA – Quản trị viên CSDL, DBMS – Người triển khai CSDL

Một ứng dụng có sử dụng cơ sở dữ liệu điển hình luôn cần truy cập cơ sở dữ liệu và truy xuất một số dữ liệu để xử lý. Bất cứ khi nào cần dữ liệu nhất định, phần dữ liệu đó phải được đặt trên dĩa và được tải vào bộ nhớ chính để xử lý và sau đó được ghi trở lại dĩa nếu có sự thay đổi. Dữ liệu được lưu trữ trên đĩa được tổ chức dưới tẹp của bản ghi và mỗi bản ghi là một tập hợp các dữ liệu có thể hiểu là sự thật về các thực thể (entity), thuộc tính (atribute) và mối quan hệ (relationship) của chúng. Do đó các bản ghi nên được lưu trữ trên đĩa theo cách có thể định vị, và vì thế chúng ta có thể tìm chúng một cách dễ dàng.

Vậy trong thực tế, chỉ mục (Indexing) là gì và có công dụng gì? Indexing là một cấu trúc giúp xác định nhanh chóng các bản ghi (record) trong bảng (table). Hiểu một cách đơn giản thì nếu không có chỉ mục (indexing) thì CSDL phải scan toàn bộ bảng (table) để tìm được các bản ghi (record) có liên quan. Dữ liệu càng lớn tốc độ truy vấn (query) càng chậm

Do đó, cấu trúc truy cập còn được gọi là chỉ mục, chúng được sử dụng để tăng tốc độ truy xuất bản ghi nếu một số yêu cầu về điều kiện tìm kiếm được đáp ứng. Chỉ mục cho tệp bản ghi hoạt động giống như danh mục chỉ mục trong thư viện. Ví dụ, trong một môi trường thư viện bình thường, nên có các danh mục như chỉ mục tác giả và chỉ mục tên sách. Người dùng có thể sử dụng một trong các chỉ mục này để nhanh chóng tìm thấy vị trí của một cuốn sách được yêu cầu, nếu họ biết (các) tác giả hoặc tên sách. Mỗi chỉ mục cung cấp một đường dẫn truy cập tới các bản ghi. Một số loại chỉ mục, được gọi là đường dẫn truy cập thứ cấp, không ảnh hưởng đến vị trí vật lý của các bản ghi trên đĩa. Thay vào đó, chúng cung cấp các đường dẫn tìm kiếm thay thế để định vị các bản ghi một cách hiệu quả dựa trên các trường lập chỉ mục. Các loại chỉ mục khác chỉ có thể được xây dựng trên một tệp với một tổ chức chính cụ thể. Trong đó 3 phương pháp lập chỉ mục chính được đề cập đến trong bài viết này là:

1. Primary Index
2. Clustering Index
3. Secondary Index

## 2.2 Primary Index

Chỉ mục chính (primary index) được định nghĩa chủ yếu trên khóa chính (primary key) của tập tin dữ liệu (data file), trong đó tập tin dữ liệu đã được sắp xếp thứ tự dựa trên khóa chính.

Chỉ mục chính (primary index) là một tập tin có thứ tự có các bản ghi (record) có độ dài cố định với hai trường (field). Trường đầu tiên của chỉ mục sao chép (replicate) khóa chính của tập tin dữ liệu theo cách có thứ tự và trường thứ hai của chứa một con trỏ (pointer) trỏ đến khối dữ liệu (data block) nơi có bản ghi chứa khóa (key). Số khối (khối) trung bình sử dụng chỉ mục chính là = log2B + 1, trong đó B là số khối của chỉ mục.

A picture containing table

Description automatically generated

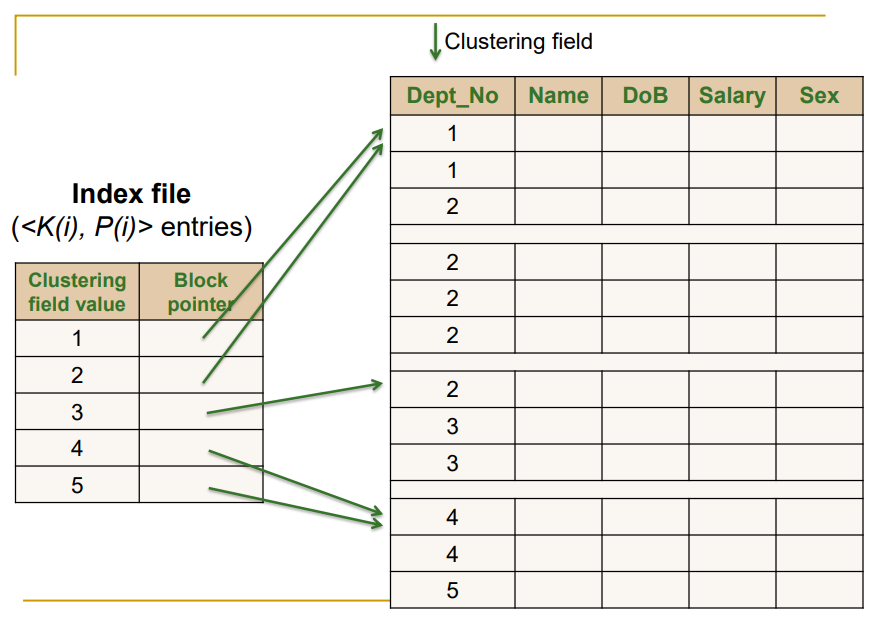
2.2.1 Hình minh họa cách hoạt động của Primary Index

## 2.3 Clustering Index

Nếu các bản ghi (record) của một tệp được sắp xếp theo thứ tự vật lý trên một trường không khóa (non-key field) có thể không có giá trị duy nhất cho mỗi bản ghi, thì trường đó được gọi là trường phân cụm (the clustering field). Dựa trên các giá trị của trường phân cụm, một chỉ mục phân cụm có thể được xây dựng để tăng tốc độ truy xuất các bản ghi có cùng giá trị cho trường phân cụm.

Chỉ mục phân cụm (clustering index) cũng là một tệp được sắp xếp gồm các bản ghi có độ dài cố định với hai trường. Trường đầu tiên cùng loại với trường phân cụm (the clustering field) của tệp dữ liệu và trường thứ hai là một con trỏ khối (a block pointer).

Ngoài ra, ta có 2 trường hợp thường thấy của Clustering index:



2.3.1 Trường hợp 1

Trường hợp 1, các bản ghi dữ liệu (record) được chứa chung trong 1 block data, điều đó khiến cho việc truy vấn dữ liệu từ index file đến data file hoạt động theo cách block pointer từ index file trỏ đến block có chứa dữ liệu trùng với giá trị clustering field value, và chúng luôn trỏ đến block anchoring ở đầu mỗi block data để tìm dữ liệu từ trên xuống.

Table

Description automatically generated

2.3.2 Trường hợp 2

Trường hợp 2, các bản ghi dữ liệu (record) được chia riêng biệt trong từng block, điều đó khiến cho việc truy vấn dữ liệu từ index file đến data file hoạt động theo cách block pointer từ index file trỏ đến đúng block dữ liệu trùng với giá trị clustering field value và chúng luôn trỏ đến block anchoring ở đầu mỗi block data để tìm dữ liệu từ trên xuống, nó làm dữ liệu được chia ra từng block tách biệt hơn nhưng đồng thời cũng làm tăng số lượng record thừa trong block nếu như dữ liệu đối chiếu quá ít.

## 2.4 Secondary Index

Chỉ mục phụ (Secodary index) là một tệp bản ghi được sắp xếp (có độ dài cố định hoặc độ dài thay đổi) với hai trường. Trường đầu tiên có cùng kiểu dữ liệu với trường lập chỉ mục (tức là trường không theo thứ tự mà chỉ mục được xây dựng trên đó). Trường thứ hai là con trỏ khối hoặc con trỏ bản ghi. Một tệp có thể có nhiều hơn một chỉ mục phụ.

Trong đó, chúng ta xem xét hai trường hợp của chỉ số phụ:

1. Cấu trúc truy cập chỉ mục được xây dựng trên một trường khóa (key).
2. Cấu trúc truy cập chỉ mục được xây dựng trên một trường không khóa (non-key).

A picture containing line chart

Description automatically generated

2.4.1 Hình minh họa cách hoạt động của Secondary Index

# **3. GIẢI QUYẾT CÁC BÀI TOÁN**

## 3.1 Lời giải các bài toán

* **Bài 1**
* **Bài 2**
* **Bài 3**

## 3.2 Các lưu đồ thuật toán

* **Bài 1**
* **Bài 2**
* **Bài 3**

# **4. CHƯƠNG TRÌNH BÀI TOÁN**

Sau khi hoàn tất quá trình giải quyết bài toán bằng lời giải lý thuyết, ta tiếp tục xây dựng chương trình và thực thi các bài toán trong phần phụ lục.

* **Bài tập 1**

Ta nhập vào input chương trình như sau:

Text

Description automatically generated

4.1 Đối số đầu vào của bài tập 1

Sau khi chương trình thực hiện và ta được kết quả như sau:

Text

Description automatically generated

4.2 Kết quả đáp án bài tập 1

* **Bài tập 2**

Ta nhập vào input chương trình như sau:

Graphical user interface, text, application, chat or text message

Description automatically generated

4.3 Đối số đầu vào của bài tập 2

Sau khi chương trình thực hiện, ta nhận được kết quả như sau:

Text

Description automatically generated

4.4 Kết quả đáp án bài tập 2

* **Bài tập 3**

Ta nhập vào input chương trình như sau:

Text

Description automatically generated

4.5 Đối số đầu vào của bài tập 3

Sau khi chương trình thực hiện và ta được kết quả như sau:

Text

Description automatically generated

4.6 Kết quả đáp án bài tập 3

# **5. TÀI LIỆU THAM KHẢO**

**- Sách, Slide bài giảng:**

* **Chapter 18. Indexing Structures for Files**
* **Chapter 2. Indexing Structures for Files (Slide thầy Đăng)**

**- Website:**

* **Chapter 11. File Organisation and Indexes.** <https://www.cs.uct.ac.za/mit_notes/database/htmls/chp11.html?fbclid=IwAR1n1ebfQa9HH3IlU241CaApp8ggPQb4nnwqt_e-PVyIQMNC7gwCumCzlBg#primary-indexes>
* **Primary Index là gì**

[Primary index là gì? - Từ điển CNTT (dictionary4it.com)](https://www.dictionary4it.com/term/primary-index-7812/)

* VTI TechBlog – Sử dụng index trong MySQL

[Sử dụng Index trong MySQL: Phần 1- Các loại index và cách đánh index - VTI TechBlog!](https://vtitech.vn/su-dung-index-trong-mysql-phan-1-cac-loai-index-va-cach-danh-index/#:~:text=Index%20l%C3%A0%20m%E1%BB%99t%20c%E1%BA%A5u%20tr%C3%BAc,%C4%91%E1%BB%99%20query%20s%E1%BA%BD%20c%C3%A0ng%20ch%E1%BA%ADm.)