**Trường Đại học Tài nguyên và Môi trường TPHCM**

****

**BÀI BÁO CÁO**

**Môn: Hệ quản trị cơ sở dữ liệu**

**Đề tài: Cơ sở dữ liệu SQL và NoSQL**

Lớp: 09CNTT2

Nhóm 9

Thành viên

Đỗ Phước Khải

Lý Huỳnh Thanh Hậu

Nguyễn Trung Kiên

**Thành phố Hồ Chí Minh, 11 tháng 12, năm 2022**

**LỜI MỞ ĐẦU**

Ngày nay, cùng với sự phát triển của xã hội, ngành CNTT cũng có nhiều bước phát triển đáng kể và đã dần đi vào cuộc sống của mọi người và được sử dụng hầu hết trong tất cả các ngành nghề trong xã hội. Với số lượng tài liệu trong các cơ quan, tổ chức tăng theo cấp số nhân theo từng năm, từng thời kỳ, thì việc lưu trữ số lượng tài liệu đó trở nên vô cùng khó khăn, đặc biệt khi nó là những tài liệu quan trọng mà lại được lưu trữ trên các thiết bị cứng. Yếu tố thời gian, các tác động bên ngoài có thể làm cho những tài liệu đó bị hỏng hóc, khó bảo quản. Do đó, nhu cầu sử dụng các phần mềm hỗ trợ khả năng lưu trữ các dữ liệu đảm bảo các yếu tố an toàn và tiện lợi trong thao tác với dữ liệu đó là vô cùng cần thiết. Và nhu cầu đó sẽ trở nên dễ dàng khi có sự vào cuộc của CNTT, đặc biệt là các hệ quản trị cơ sở dữ liệu (CSDL). Nói đến CNTT thì không thể không nói đến các hệ quản trị cơ sở dữ liệu. Đó là phần mềm hay hệ thống được thiết kế để quản trị một CSDL, nó hỗ trợ khả năng lưu trữ, sửa chữa, xóa và tìm kiếm trong tin trong một CSDL.

Với sự gợi ý của giáo viên hướng dẫn là thầy Phạm Trọng Huynh, chúng em đã tập trung tìm hiểu và nghiên cứu về Cơ sở dữ liệu SQL và NoSQL để nhận thấy sự khác biệt giữa hai nền tảng

Đề tài bao gồm các nội dung:

1. Giới thiệu
2. Nguồn gốc
3. Tuyến tính của SQL
4. Vai trò của kiến trúc dữ liệu SQL
5. Đặc điểm của NoSQL
6. Kết luận và công việc tương lai

**LỜI CẢM ƠN**

Lời đầu tiên, nhóm em xin gửi lời tri ân sâu sắc đến thầy Phạm Trọng Huynh. Trong quá trình tìm hiểu và học tập bộ môn Hệ quản trị cơ sở dữ liệu, nhóm em đã nhận được sự giảng dạy và hướng dẫn rất tận tình, tâm huyết của thầy. Thầy đã giúp nhóm em tích lũy thêm nhiều kiến thức hay và bổ ích. Từ những kiến thức mà thầy truyền đạt, nhóm em xin trình bày lại những gì mình đã tìm hiểu về đề tài Cơ sở dữ liệu SQL và NoSQL gửi đến thầy. Mặc dù nhóm em đã có sự cố gắng, nhưng trong khoảng thời gian cho phép cũng như hạn chế về kiến thức nên đồ án này của em không thể tránh khỏi những kiến thức thiếu sót. Chính vì vậy, nhóm em rất mong nhận được sự góp ý của các thầy cô giáo và bạn bè. Kính chúc thầy hạnh phúc và thành công hơn nữa trong sự nghiệp “trồng người”. Kính chúc thầy luôn dồi dào sức khỏe để tiếp tục dìu dắt nhiều thế hệ học trò đến những bến bờ tri thức.

Sinh Viên Thực Hiện

Nhóm 9

Tóm lược : NoSQL (Not only SQL) là một cơ sở dữ liệu được sử dụng để lưu trữ một lượng lớn dữ liệu. Cơ sở dữ liệu NoSQL được phân phối, không liên quan, nguồn mở và có thể mở rộng theo chiều ngang (theo cách tuyến tính). NoSQL không tuân theo thuộc tính của ACID như chúng ta tuân theo trong SQL. Trong tài liệu nghiên cứu này, chúng tôi đang khảo sát về NoSQL, nền tảng của nó, các nguyên tắc cơ bản như định lý ACID, BASE và CAP. Cũng trên cơ sở định lý CAP, nghiên cứu được thực hiện về các loại kho lưu trữ dữ liệu NoSQL khác nhau với các ví dụ, đặc điểm và ưu nhược điểm của NoSQL.

**MỤC LỤC**

[Chương 1: Giới thiệu](#gioithieu)

[Chương 2: Nguồn gốc](#nguon_goc)

[Chương 3: Tuyến tính của NoSQL](#chuong3)

[A.Không ACID](#A_khong_acid)

[B.BASE](#B_base)

[C.CAP](#C_cap)

[Chương 4: Vai trò của kiến trúc dữ liệu trong NoSQL](#chuong4)

[A.Thành phần](#vaitro_4a)

[B.Các kiểu lưu trữ dữ liệu NoSQL](#vaitro_4b)

[1.Cơ sở dữ liệu giá trị khóa](#vaitro4b1)

[1.1.Đặc điểm của cơ sở dữ liệu giá trị khóa](#vaitro4b1_1)

[1.2.Mục đích của cơ sở dữ liệu giá trị khóa](#vaitro4b2_2)

[2.Cơ sở dữ liệu trữ tài liệu](#vaitro4b2)

[2.1.Đặc điểm của cơ sở dữ liệu trữ tài liệu](#vaitro4b2_1)

[2.2.Ví dự về cơ sở dữ liệu trữ tài liệu](#vaitro4b2_2)

[3.Cơ sở dữ liệu dạng cột](#vaitro4b3)

[3.1.Đặc điểm của cơ sở dữ liệu dạng cột](#vaitro4b3_1)

[4.Cơ sở dữ liệu đồ thị](#vaitro4b4)

[4.1.Ví dụ về cơ sở dữ liệu đồ thị](#vaitro4b4_1)

[4.2.Đặc điểm của cơ sở dữ liệu đồ thị](#vaitro4b4_2)

[Chương 5: Đặc điểm của NoSQL](#chuong5)

[Chương 6: Kết luận và công việc tương lai](#chuong6)

**1.GIỚI THIỆU**

NoSQL là viết tắt của Not Only SQL. Nó được phát âm là Nô-si-qồ. Đó là một trong những loại lưu trữ dữ liệu khác ngoài cơ sở dữ liệu được sử dụng để lưu trữ lượng dữ liệu khổng lồ như dữ liệu trong facebook. NoSQL là một hệ quản trị cơ sở dữ liệu phi quan hệ (đôi khi được gọi là dẫn xuất từ cơ sở dữ liệu quan hệ), cơ sở dữ liệu truy xuất thông tin nhanh và có thể di chuyển được. NoSQL về cơ bản bắt nguồn từ hệ thống cơ sở dữ liệu RDB. Cơ sở dữ liệu này thường tương tác với hệ điều hành UNIX. Cơ sở dữ liệu NoSQL là những cơ sở dữ liệu không quan hệ, nguồn mở, được phân phối trong tự nhiên cũng như có hiệu suất cao theo cách tuyến tính có thể mở rộng theo chiều ngang. Cơ sở dữ liệu không quan hệ không tổ chức dữ liệu trong các bảng có liên quan (nghĩa là dữ liệu được lưu trữ theo cách không chuẩn hóa). Cơ sở dữ liệu NoSQL là mã nguồn mở; do đó, mọi người có thể tự do xem mã, cập nhật theo nhu cầu và biên dịch. Phân tán có nghĩa là dữ liệu được lan truyền đến các máy khác nhau và được quản lý bởi các máy khác nhau nên ở đây sử dụng khái niệm sao chép dữ liệu. NoSQL có thể được biểu diễn một cách tượng trưng như trong hình 1:

Hình 1: Biểu tượng đặc trưng của NoSQL

Hình 1 cho biết truy vấn đến cơ sở dữ liệu mà không có bất kỳ tương tác hoặc giao diện nào của ngôn ngữ SQL. Đường nghiêng trong hình cho thấy việc sử dụng cơ sở dữ liệu mà không sử dụng SQL (Ngôn ngữ truy vấn có cấu trúc). Vì vậy, để truy cập các cơ sở dữ liệu này, chúng ta có thể sử dụng một số định dạng khác như XML để lưu trữ và truy xuất thông tin từ cơ sở dữ liệu.

2.NGUỒN GỐC

Vào năm 1960, cơ sở dữ liệu đa giá trị được phát triển để thực hiện một số hoạt động cơ sở dữ liệu phân cấp bằng cách sử dụng lưu trữ cây B+. Theo Scott Jones, M[umps] được phát triển vào năm 1966 để hoạt động dựa trên các loại cơ sở dữ liệu như vậy (phân cấp). Năm 1977, M[umps] đã được chấp thuận trở thành ngôn ngữ chuẩn ANSI. Năm 1979, Ken Thompson đã phát triển DBM (Trình quản lý cơ sở dữ liệu) sử dụng kỹ thuật băm để cung cấp cho người dùng khả năng truy xuất dữ liệu nhanh hơn. Vào năm 1980, một số lượng lớn người kế nhiệm DBM đã được phát triển như sau:

∙ TDBM là cơ sở dữ liệu kiểu DBM hỗ trợ các giao dịch nguyên tử (tuân theo các thuộc tính ACID A: Atomicity, C:Consistency, I: Isolation, D: Durability). Những ưu điểm của TDBM bao gồm:

* Nếu ứng dụng đang sử dụng TDBM, trong trường hợp có hư hỏng hoặc sự cố, nội dung được lưu trữ trong cơ sở dữ liệu sẽ được an toàn.
* TDBM cũng hỗ trợ các giao dịch nguyên tử lồng nhau bằng cách sử dụng các hoạt động đa luồng.
* SDBM là cơ sở dữ liệu kiểu DBM được sử dụng vì một số lý do cấp phép.
* GT.M là một sản phẩm kế thừa khác của DBM, tập trung vào việc xử lý giao dịch có hiệu suất cao. Năm 2000, DBM này có nguồn mở.

Sự phát triển của cơ sở dữ liệu SQL bắt đầu vào cuối những năm 1990. Sau một vài năm, nó trở thành một đối thủ nặng ký với RDBMS (Hệ thống quản lý cơ sở dữ liệu quan hệ). Trong năm 2009 và 2010 đã có các hội nghị NoSQL được tổ chức như NoSQL live, NoSQL eu, NoSQL East,… Tên NoSQL này được Carlo Strozzi sử dụng lần đầu tiên vào năm 1998 làm tên tệp mà ông đang phát triển cho cơ sở dữ liệu của mình.

Ngày nay, NoSQL đang trở nên quá phổ biến do dung lượng lưu trữ cao và cũng vì các thuộc tính của nó tránh các tính năng cơ bản của SQL.

Cơ sở dữ liệu quan hệ được thiết kế để chạy trên một máy duy nhất, vì vậy, để làm được điều này, chúng ta cần một máy lớn để mở rộng quy mô. Một giải pháp để khắc phục điều này là mua nhiều máy (và biến chúng thành một cụm) để lưu trữ dữ liệu. Nó sẽ là một quy trình kinh tế hơn cũng như có thể mở rộng theo chiều ngang (theo cách tuyến tính). Ngay cả khi máy bị trục trặc, độ tin cậy tổng thể của cụm là khá cao. Do đó, ưu thế của SQL ngày càng giảm và NoSQL đang trở nên phổ biến.

Hình 2: Sự suy giảm ưu thế của SQL

3.TUYẾN TÍNH CỦA NoSQL

**A.Không ACID**

ACID là viết tắt của **A**tomicity(Tính nguyên tử), **C**onsistency(Nhất quán), **I**solation(Độc lập) và **D**urability(Bền vững). Khái niệm ACID về cơ bản xuất phát từ môi trường SQL. Nhưng trong NoSQL chúng ta sẽ không sử dụng khái niệm ACID vì tính năng nhất quán của SQL. Trong bài viết này, chúng ta sẽ thấy khái niệm ACID tạo ra các vấn đề đối với NoSQL như thế nào.

Như trong môi trường phân tán, dữ liệu được lan truyền đến các máy khác nhau, mỗi máy lưu trữ dữ liệu của mình và cần duy trì tính phù hợp. Ví dụ, nếu có thay đổi trong một bộ dữ liệu của bảng thì cần có thay đổi trong mỗi máy chứa dữ liệu cụ thể đó. Nếu thông tin liên quan đến một bản cập nhật lan truyền ngay lập tức, thì tính nhất quán sẽ được đưa ra; nếu không, thì sự không nhất quán được thực hiện.

**B. BASE**

BASE là viết tắt của “**B**asically **A**vailable (Tính khả dụng cơ bản), **S**oft state(trạng thái mềm), **E**ventual consistency(Tính nhất quán cuối cùng)”. BASE ngược lại với ACID . Cơ sở dữ liệu NoSQL được phân chia giữa con đường từ ACID đến BASE. Sau khi giao dịch một cách nhất quán, trạng thái mà chúng ta sẽ nhận được là trạng thái mềm không phải trạng thái rắn. Trọng tâm chính đằng sau BASE là tính khả dụng lâu dài.

Ví dụ, nghĩ về cơ sở dữ liệu trong ngân hàng, nếu hai người đang truy cập cùng một tài khoản ở các thành phố khác nhau thì việc cập nhật dữ liệu không chỉ cần kịp thời mà còn cần một số cơ sở dữ liệu thời gian thực. Những cập nhật đó cần được thực hiện thường xuyên trên tất cả các máy. Một số ví dụ khác là đặt chỗ đường sắt trực tuyến, mua bán sách trực tuyến, v.v.

**C.CAP**

CAP là viết tắt của **C**onsistency (Tính nhất quán), **A**vailable **(**Tính khả dụng) và **P**artition tolerance (Dung sai phân vùng). CAP về cơ bản là một định lý tuân theo ba nguyên tắc :

1) Dữ liệu có sẵn trên tất cả các máy phải giống nhau ở mọi khía cạnh và việc cập nhật thường xuyên được thực hiện trên tất cả các máy, tức là dữ liệu nhất quán.

2) Dữ liệu phải có sẵn vĩnh viễn và có thể truy cập mọi lúc, tức là có sẵn.

3) Trong thời gian máy bị lỗi hoặc bất kỳ lỗi nào trong cơ sở dữ liệu máy sẽ hoạt động tốt mà không làm ngừng công việc của chúng, tức là dung sai phân vùng.

4. VAI TRÒ CỦA KIẾN TRÚC DỮ LIỆU TRONG NoSQL

**A. Thành phần**: Có bốn thành phần trong khối xây dựng của nó.

[1.] Ngôn ngữ mô hình hóa: Nó mô tả cấu trúc của cơ sở dữ liệu và cũng xác định lược đồ dựa trên đó. Dữ liệu được lưu trữ dưới dạng hàng và cột sử dụng các định dạng XML. Và mỗi dữ liệu (giá trị) tương ứng với nó được gán một khóa duy nhất về bản chất. Để truy cập dữ liệu nhanh hơn, mô hình được xây dựng trong một môi trường thích hợp.

[2.] Cấu trúc cơ sở dữ liệu: Mỗi và mọi cơ sở dữ liệu trong khi xây dựng đều sử dụng cấu trúc dữ liệu riêng và lưu trữ dữ liệu bằng thiết bị lưu trữ vĩnh viễn.

[3.] Ngôn ngữ truy vấn cơ sở dữ liệu: Tất cả các hoạt động được thực hiện trên cơ sở dữ liệu đó là tạo, cập nhật, đọc và xóa (CURD).

[4.] Chuyển đổi: Trong bất kỳ chuyển đổi nào trong dữ liệu, có thể có bất kỳ loại lỗi hoặc lỗi nào; sau đó, máy sẽ không ngừng hoạt động.

**B. Các kiểu lưu trữ dữ liệu NoSQL**

Trên cơ sở định lý CAP, cơ sở dữ liệu NoSQL được chia thành số lượng cơ sở dữ liệu. Có bốn kiểu lưu trữ dữ liệu khác nhau mới trong NoSQL .

**1.Cơ sở dữ liệu giá trị khóa**: Bản thân tên của cơ sở dữ liệu giá trị khóa đã nói lên rằng nó là sự kết hợp của hai thứ là khóa và giá trị. Nó là một trong những hệ thống cơ sở dữ liệu cấu hình thấp (truyền thống). Cơ sở dữ liệu Key Value (KV) là mẹ của tất cả các cơ sở dữ liệu của NoSQL.

∙ Khóa là một định danh duy nhất cho một mục nhập dữ liệu cụ thể. Không nên lặp lại khóa nếu một khóa đã được sử dụng mà nó không bị trùng lặp về bản chất.

∙ Giá trị là một loại dữ liệu được trỏ bởi một khóa.

Cơ sở dữ liệu giá trị chính dường như là bảng băm hoặc bảng tra cứu. Trong loại cơ sở dữ liệu này, chỉ có một cách để truy vấn đó là với sự trợ giúp của khóa (duy nhất) và tất cả các khóa có thể đặt tên trong bất kỳ đối tượng dữ liệu nào và được sắp xếp theo thứ tự bảng chữ cái. Đối với tính khả dụng cao hơn của các cửa hàng dữ liệu, các đối tượng dữ liệu được nhân rộng.

|  |  |
| --- | --- |
| Cơ sở dữ liệu ngân hàng | |
| Khóa | Giá trị |
| 1 | ID:1  Ngày tham gia:15/7/1985  Chức danh: Thu ngân |
| 2 | ID:2  Ngày tham gia:19/3/1982  Chức danh: Trợ lý |
| 3 | ID:3  Ngày tham gia: 4/4/1988  Chức danh: Tiếp tân |

* 1. **Đặc điểm của cơ sở dữ liệu giá trị khóa**

∙ Số lượng khóa có thể có một tập hợp động các thuộc tính trong cơ sở dữ liệu giá trị khóa trong quá trình lưu trữ dữ liệu.

∙ Dữ liệu được lưu trữ trong cơ sở dữ liệu được lưu trữ theo thứ tự bảng chữ cái.

∙ Tất cả các hoạt động có thể được thực hiện trên dữ liệu, tức là CRUD ( Create, Read, and Update and Delete).

∙ Tất cả các mối quan hệ với dữ liệu được lưu trữ trong mã ứng dụng (không lan truyền rõ ràng).

* 1. **Mục đích cơ sở dữ liệu giá trị khóa (KV)**

∙ Đây là một trong những mô hình dữ liệu đơn giản trong số tất cả vì nó chỉ sử dụng khóa và một giá trị.

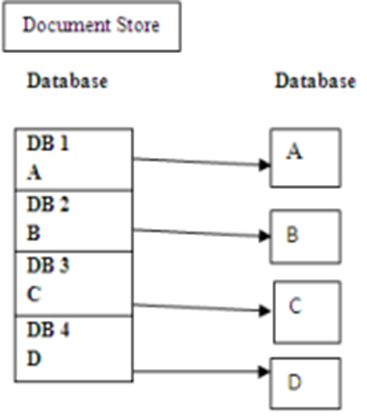
∙ Nó xử lý tải dữ liệu lớn.

∙ Nó mở rộng đến khối lượng lớn dữ liệu.

∙ Việc sao chép dữ liệu được thực hiện bằng cách sử dụng cơ sở dữ liệu ở dạng vòng. Dữ liệu sao chép được lưu trữ dưới dạng vòng cũng như theo thứ tự bảng chữ cái.

**2.** **Cơ sở dữ liệu trữ tài liệu**

Cơ sở dữ liệu trữ tài liệu là những cơ sở dữ liệu NoSQL sử dụng các bản ghi làm tài liệu. Loại cơ sở dữ liệu này lưu trữ các tài liệu phi cấu trúc (văn bản) hoặc bán cấu trúc (XML) thường có bản chất phân cấp. Ở đây mỗi tài liệu bao gồm một tập hợp các khóa và giá trị gần giống như trong cơ sở dữ liệu giá trị khóa. Mỗi cơ sở dữ liệu nằm trong kho lưu trữ tài liệu trỏ đến các trường của nó bằng cách sử dụng con trỏ vì nó sử dụng kỹ thuật băm. Cơ sở dữ liệu trữ tài liệu không có giản đồ và không cố định về bản chất.

Cấu trúc của Cơ sở dữ liệu trữ tài liệu như được minh họa trong hình dưới đây. Hình mô tả rằng nó bao gồm một số cơ sở dữ liệu trong kho lưu trữ tài liệu như cơ sở dữ liệu 1,2,3,4 và có id của nó là A, B, C, D nằm trong đó, nó đang trỏ đến cơ sở dữ liệu của nó có quan hệ nào đó. với nó. Cơ sở dữ liệu trỏ đến giá trị của nó bằng cách sử dụng một số khóa duy nhất nằm trong cơ sở dữ liệu của nó. Điều này bao gồm một mảng cơ sở dữ liệu (ở dạng nhóm).

**2.1 Đặc điểm của cơ sở dữ liệu trữ tài liệu**

∙ Tài liệu được xử lý trong cơ sở dữ liệu bằng cách sử dụng khóa (duy nhất) đại diện cho tài liệu đó.

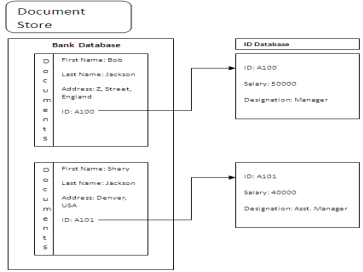
∙ Có một số loại để tổ chức dữ liệu đó là bộ sưu tập, thẻ, siêu dữ liệu không nhìn thấy được và phân cấp thư mục.

∙ Trong trường hợp này, chúng ta có thể sử dụng tra cứu giá trị khóa để truy xuất tài liệu.

**2.2** **Ví dụ về cơ sở dữ liệu trữ tài liệu**

Hình dưới minh họa một ví dụ về cơ sở dữ liệu lưu trữ tài liệu. Trong hình này, chúng ta đang lấy một ví dụ về cơ sở dữ liệu ngân hàng sử dụng ID làm khóa duy nhất. Và nó cũng bao gồm cơ sở dữ liệu ID chứa tất cả thông tin liên quan đến người đang có ID cụ thể đó. Một ID trong cơ sở dữ liệu ngân hàng trỏ đến cơ sở dữ liệu ID bằng cách sử dụng con trỏ. Cơ sở dữ liệu ngân hàng bao gồm hai tài liệu:

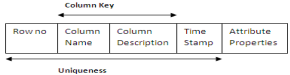
(1) Một người có tên: Bob; Họ: Jackson; Địa chỉ: Z Street, England; ID: A100. (2) Một người có tên: sherry; Họ: Hilton; Địa chỉ: Denver, Hoa Kỳ; ID: A101.

Bây giờ chúng ta có thể coi ID của cả hai tài liệu như một khóa được sử dụng để trao đổi thư từ. Khóa phải là duy nhất về bản chất và không rỗng. Chúng tôi không thể tạo một khóa trùng lặp. Chúng tôi chọn ID làm khóa vì tên, địa chỉ không thể là duy nhất và ID của một người luôn là duy nhất. Vì vậy, đó là lý do tại sao chúng tôi chọn ID làm khóa. Bây giờ chúng ta đang có một cơ sở dữ liệu khác là cơ sở dữ liệu ID. Cơ sở dữ liệu ID bao gồm mức lương và chỉ định của một người. Cơ sở dữ liệu ID bao gồm các cơ sở dữ liệu của ID’s A100, A101. Các tài liệu hiện có trong cơ sở dữ liệu ngân hàng bao gồm ID và trỏ đến dữ liệu của ID cụ thể. Đây là cách dữ liệu được tổ chức trong cơ sở dữ liệu lưu trữ tài liệu. Dữ liệu được tổ chức dưới dạng tập hợp, thẻ, siêu dữ liệu, v.v. Trong đó, chúng ta có thể truy xuất dữ liệu bằng cách sử dụng cặp khóa và giá trị. Điều này được minh họa rõ hơn trong ví dụ mà chúng ta có khóa A100 và các thuộc tính tương ứng của nó bao gồm mức lương kèm theo chỉ định.

**3. Cơ sở dữ liệu dạng cột**

Cơ sở dữ liệu dạng cột còn được gọi là cơ sở dữ liệu họ cột vì chúng là cơ sở dữ liệu hướng cột. Có hai loại cơ sở dữ liệu hướng cột có chi tiết như dưới đây:

(1) Kho dữ liệu cột rộng:

Nó là một loại cơ sở dữ liệu NoSQL. Kho dữ liệu Cột rộng là những cơ sở dữ liệu được sử dụng để xử lý web, truyền dữ liệu và tài liệu. Cấu trúc của kho dữ liệu Cột rộng được mô tả trong hình dưới đây:

Ý nghĩa của từng trường được cung cấp trong cấu trúc của cơ sở dữ liệu lưu trữ dữ liệu cột rộng như được mô tả dưới đây của hình trên

|  |  |
| --- | --- |
| **Thuộc tính** | **Ý nghĩa** |
| **Số dòng** | Đó là 1 khóa đặc biệt tự nhiên. Có thể là 1 chuỗi hoặc 1 số |
| **Tên cột** | Dữ liệu được lưu trữ dựa trên cơ bản của chuỗi cột |

|  |  |
| --- | --- |
| **Mô tả cột** | Miêu tả về mục dữ liệu được chứa. |
| **Dấu thời gian** | Đánh dấu thời gian hoàn thành |
| **Giá trị dữ liệu** | Giá trị hoặc thuộc tính liên quan tới khóa quan hệ. |

Ý nghĩa của các trường trong cấu trúc của cơ sở dữ liệu cột rộng

(2) Cơ sở dữ liệu hướng cột:

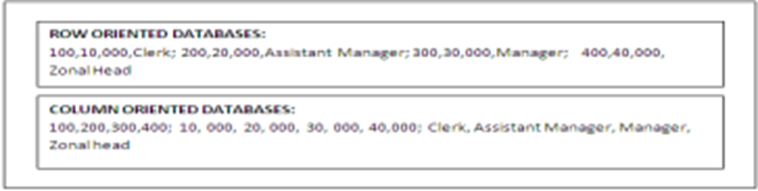
Để hiểu cơ sở dữ liệu hướng cột, chúng ta hãy lấy một ví dụ về cơ sở dữ liệu ngân hàng được cho trong hình bên dưới có các trường thuộc tính là EmpID, tiền lương và tuyến tính và các giá trị tương ứng với nó như được mô tả trong cơ sở dữ liệu.

|  |  |  |
| --- | --- | --- |
| **EmpID** | **Lương** | **Chức danh** |
| 100 | 10,000 | Nhân viên văn phòng |
| 200 | 20,000 | Trợ lý giám đốc |
| 300 | 30,000 | Quản lý |
| 400 | 40,000 | Trưởng ban |

Ví dụ về cơ sở dữ liệu ngân hàng

Biểu diễn cơ sở dữ liệu hướng hàng và cơ sở dữ liệu hướng cột:

Cơ sở dữ liệu hướng hàng là những cơ sở dữ liệu trong đó tất cả các hàng được ghép lại với nhau từng hàng một. ∙ Cơ sở dữ liệu hướng theo cột là những cơ sở dữ liệu trong đó tất cả các giá trị chứa các cột được ghép lại với nhau.

Với sự trợ giúp của cơ sở dữ liệu được đưa ra ở trên, chúng tôi sẽ đại diện cho cơ sở dữ liệu hướng hàng và cột như thể hiện trong hình bên dưới

**3.1 Đặc điểm của cơ sở dữ liệu dạng cột**

(1) Cơ sở dữ liệu cột nhanh hơn cơ sở dữ liệu dựa trên hàng trong khi truy vấn.

(2) Trong cơ sở dữ liệu cột, việc gán đơn vị lưu trữ được thực hiện cho mỗi và mọi cột.

(3) Trong DBMS cột chỉ có các cột bắt buộc được đọc, do đó, việc đọc nhanh hơn trong trường hợp này.

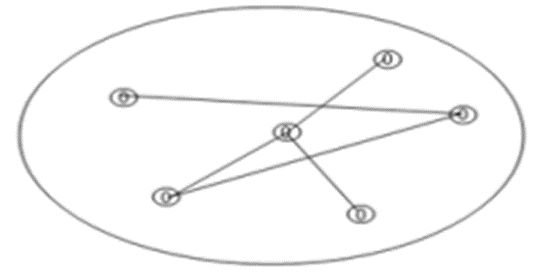
**4. Cơ sở dữ liệu đồ thị**

Cơ sở dữ liệu đồ thị dựa trên lý thuyết đồ thị. Nhìn chung, chúng ta thấy rằng đồ thị thường bao gồm các nút, các thuộc tính và các cạnh. Cơ sở dữ liệu NoSQL Graph bao gồm:

(1) Các nút đại diện cho các thực thể

(2) Thuộc tính đại diện cho các thuộc tính

(3) Các cạnh thể hiện các mối quan hệ.

Cấu trúc của cơ sở dữ liệu đồ thị như hình dưới đây:

**4.1 Ví dụ về cơ sở dữ liệu Đồ thị**

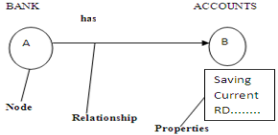
Ví dụ về cấu trúc của cơ sở dữ liệu đồ thị được thể hiện trong hình ở phần 4.2. Nó giải thích cho chúng ta cách biểu diễn các nút, mối quan hệ và thuộc tính. Ở đây có hai nút (được biểu diễn bằng vòng tròn) A và B đại diện cho NGÂN HÀNG và TÀI KHOẢN. Cả hai đang có mối quan hệ được thể hiện bằng một dòng trong đó một ngân hàng có số tài khoản. Tiếp theo là các thuộc tính / thuộc tính / giá trị đại diện cho các loại tài khoản trong ví dụ được đưa ra trong hình sau. Thuộc tính là gửi tiết kiệm, tiền tệ và RD.

**4.2 Đặc điểm của cơ sở dữ liệu đồ thị**

∙ Việc truyền qua đồ thị được thực hiện với tốc độ không đổi, không phụ thuộc vào tổng kích thước của đồ thị. Không có hoạt động tập hợp nào liên quan đến việc làm giảm hiệu suất như đã thấy với các hoạt động kết hợp trong RDBMS.

∙ Cơ sở dữ liệu đồ thị đang có hiệu suất cao trong bối cảnh với các đường dẫn sâu của chúng.

∙ Chúng được sử dụng để tính toán đường đi ngắn nhất.

∙ Có thể mở rộng. Nhưng độ phức tạp tăng lên.

V. ĐẶC ĐIỂM CỦA NoSQL

∙ NoSQL không sử dụng mô hình dữ liệu quan hệ do đó không sử dụng ngôn ngữ SQL.

∙ NoSQL lưu trữ khối lượng lớn dữ liệu.

∙ Trong môi trường phân tán (truyền dữ liệu đến các máy khác nhau), chúng tôi sử dụng NoSQL mà không có bất kỳ sự nhất quán nào. ∙ Nếu bất kỳ lỗi hoặc hỏng hóc nào tồn tại trong bất kỳ máy nào, thì trong máy này sẽ không có bất kỳ sự ngừng hoạt động nào. ∙ NoSQL là cơ sở dữ liệu mã nguồn mở, tức là mã nguồn của nó có sẵn cho tất cả mọi người và được sử dụng miễn phí mà không cần bất kỳ chi phí nào.

∙ NoSQL cho phép dữ liệu lưu trữ trong bất kỳ bản ghi nào mà nó không có bất kỳ lược đồ cố định nào.

∙ NoSQL không sử dụng khái niệm thuộc tính ACID.

∙ NoSQL có thể mở rộng theo chiều ngang dẫn đến hiệu suất cao theo cách tuyến tính.

∙ Nó có cấu trúc linh hoạt hơn.

VI. KẾT LUẬN VÀ CÔNG VIỆC TƯƠNG LAI

Mục đích chính của bài báo này là cung cấp một cái nhìn tổng quan về cơ sở dữ liệu NoSQL, về cách nó đã dập tắt ưu thế của SQL, với nền tảng và đặc điểm của nó. Nó cũng mô tả các nguyên tắc cơ bản tạo thành cơ sở của cơ sở dữ liệu NoSQL như ACID, BASE và định lý CAP. Thuộc tính ACID không được sử dụng trong cơ sở dữ liệu của cơ sở dữ liệu NoSQL vì tính nhất quán của dữ liệu, vì vậy chúng ta biết cách SQL làm ra tính nhất quán của dữ liệu. Sau đó, trên cơ sở định lý CAP, chúng tôi đã mô tả các loại cơ sở dữ liệu NoSQL khác nhau là cơ sở dữ liệu Giá trị khóa, cơ sở dữ liệu lưu trữ tài liệu, cơ sở dữ liệu dựa trên cột và cơ sở dữ liệu đồ thị với sự trợ giúp của một ví dụ. Ngoài tất cả những điều này, chúng tôi cũng đã mô tả các đặc điểm, độ phức tạp và hiệu suất của chúng. Nghiên cứu sâu hơn đang diễn ra về các công nghệ mới đang phát sinh cho hoặc sau NoSQL, đó là tính bền đa giác, v.v.

**THAM KHẢO**

[1.]SilvanWeber,“NoSQLDatabases”http://www.christofstrauch.de/nosqldbs.pdf

[2.]MartinFowler and Pramod Sadalage Rendered, “NoSQLdbs-“, February8,2012,11:26, http://martinfowler.com/articles/nosql-intro.pdf

[3.] An Oracle White Paper, “Oracle NoSQL Database”, September2011, http://www.oracle.com/technetwork/database/nosqldb/learnmore/nosql-database-498041.pdf

[4.] Luis Ferreira Universidade do Minho, “Bridging the gap between SQL and NoSQL”, httpsikhote.files.wordpress.com201105artigo-mi-star1.pdf

[5].Andrew J. Brust, Blue Badge Insights, Inc., “NoSQL and the Windows Azure platform”, April 25, 2011 [6.] DAMA - Philadelphia / Delaware Valley, the “Role of Data Architecture in NOSQL”, Wednesday January 11th, 2012, http://www.damaphila.org/HaugheyNOSQL.pdf