Đánh giá các mô hình và hệ thống cơ sở NOSQL Bhupinder Khoa học Máy tinh va Kỹ thuật AIT, Đại học Chandigarh , Mohali (Punjab), Ấn Độ

Tóm tắt— Với nhu cầu mới về quy mô ứng dụng internet,cần có một hệ quản trị cơ sở dữ liệucó thể mở rộng quy mô khi nhu cầu tăng lên. các hệ quản trị cơ sở dữ liệu truyền thống không thể đáp ứng nhu cầu về khả năng mở rộng và độ tin cậy của dữ liệu,mặc dù chúng rất phù hợp để quản lý các giao dịch.Cơ sở dữ liệu NoSQL cung cấp chức năng mà hầu hết cơ sở dữ liệu truyền thống thì không. Đối với các trường hợp sử dụng như giữ số lượng lớn khối lượng dữ liệu trong các cặp giá trị chính trong khoảng thời gian ngắn hoặc giữ dữ liệu phi cấu trúc với khối lượng lớn, không thể dễ dàng xử lý bằng cách sử dụng cơ sở dữ liệu quan hệ và cấu trúc ngôn ngữ truy vấn (SQL), cơ sở dữ liệu NoSQL được sử dụng. bài này giới thiệu với các cơ sở dữ liệu NoSQL khác nhau và giải thích về chúng với mục đích và chức năng sẽ giúp người đọc quyết định cơ sở dữ liệu NoSQL nào phù hợp cho một trường hợp sử dụng cụ thể.

Từ khóa—Cơ sở dữ liệu NoSQL, Kho giá trị khóa, Tài liệu

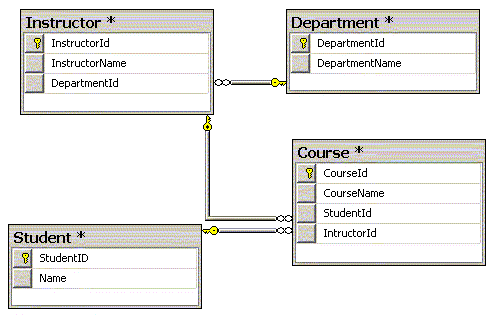
DB, đồ thị DB, cột DB

I.GIỚI THIỆU

Một hệ quản trị cơ sở dữ liệu (DBMS) là một phần mềm để tạo và quản lý cơ sở dữ liệu một cách có hệ thống. Nó hoạt động như một giao diện giữa lưu trữ dữ liệu và người dùng cuối hoặc các chương trình ứng dụng sử dụng nó và cung cấp một cái nhìn nhất quán Dữ liệu. DBMS giữ trách nhiệm quản lý ba các thành phần quan trọng: dữ liệu, công cụ cơ sở dữ liệu và lược đồ cơ sở dữ liệu. Dữ liệu là nội dung được lưu trữ. Công cụ cơ sở dữ liệu là phần mềm cho phép dữ liệu được được truy cập, khóa và sửa đổi và lược đồ cơ sở dữ liệu xác định cấu trúc logic của một cơ sở dữ liệu. Ba nền tảng này các yếu tố giúp cung cấp đồng thời, bảo mật, dữ liệu liêm chính và các thủ tục hành chính thống nhất. Hệ quản trị cơ sở dữ liệu rất hữu ích để cung cấp chế độ xem tập trung dữ liệu có thể được truy cập bởi nhiều người dùng, từ nhiều địa điểm, trong một Cách thức kiểm soát.

A. Hệ thống quản lý cơ sở dữ liệu quan hệ (RDBMS):

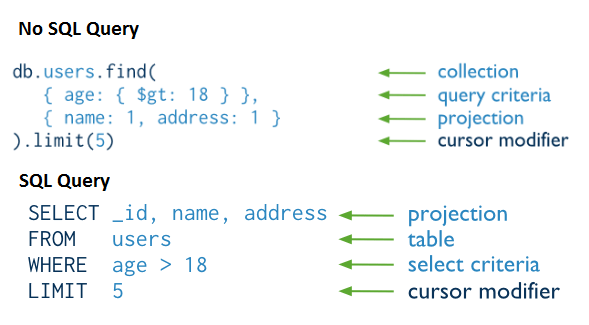
Cơ sở dữ liệu quan hệ hay còn gọi là cơ sở dữ liệu sql là một tập hợp các các bảng được mô tả chính thức duy trì dữ liệu trong một cách cấu trúc và các mối quan hệ trong dữ liệu. có cấu trúc ngôn ngữ truy vấn (SQL) được sử dụng để tạo và thao tác lược đồ của cơ sở dữ liệu cũng như xử lý dữ liệu cho người dùng cuối/ứng dụng tiêu thụ dữ liệu. EF Codd [1] đề xuất tổ chức dữ liệu trong các bảng ở dạng hàng và các cột thay vì lưu trữ theo thứ bậc hoặc các cấu trúc điều hướng. Mỗi bảng đang xác định thực thể với cột xác định các thuộc tính. Bảng cũng được sử dụng để lưu trữ các mối quan hệ giữa các thực thể. Một trường hợp duy nhất của thực thể được mô tả bởi mỗi hàng trong một bảng. Khóa chính, là duy nhất trên các bản ghi trong cơ sở dữ liệu xác định danh tính của trường hợp cụ thể. Mối quan hệ được duy trì sử dụng khóa ngoại liên kết với khóa chính của người khác bàn. Hình 1 cung cấp biểu diễn đơn giản của RDBMS / Cơ sở dữ liệu SQL. Cơ sở dữ liệu quan hệ cho phép người dùng dễ dàng phân loại và lưu trữ dữ liệu mà sau này có thể được truy vấn và được lọc để trích xuất thông tin cụ thể để báo cáo mục đích. Cơ sở dữ liệu quan hệ cũng dễ mở rộng và không phụ thuộc vào tổ chức vật lý. Thực thể và thuộc tính có thể được thêm mà không cần tất cả các ứng dụng hiện có sửa đổi.



Hình .Biểu diễn đơn giản các bảng SQL và các mối quan hệ.

B. Cơ sở dữ liệu NoSQL:

Cơ sở dữ liệu NoSQL và hệ thống quản lý, theo thiết kế, lược đồ ít hơn. Chúng không dựa trên một mô hình duy nhất (ví dụ: Các thực thể và các mối quan hệ ở dạng bảng trong quan hệ cơ sở dữ liệu) và mỗi cơ sở dữ liệu có thể áp dụng một lược đồ khác nhau tùy thuộc vào chức năng mục tiêu. Không có lược đồ xác định loại bỏ nhu cầu về SQL tiêu chuẩn và do đó được gọi là Cơ sở dữ liệu NoSQL.



Hình .Truy vấn NoSQL và Truy vấn SQL tương ứng.

Cơ sở dữ liệu NoSQL có thể lưu trữ dữ liệu ở dạng giá trị khóa cặp, cấu trúc dựa trên cột dưới dạng tài liệu xml hoặc json hoặc dưới dạng đồ thị. Hình 2 đưa ra một phương pháp NoSQL đơn giản của truy xét dữ liệu tương ứng với một câu hỏi về SQL.

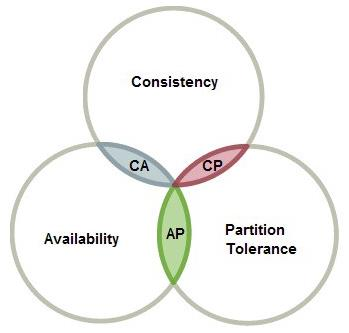
C. Định lý CAP:

Để xây dựng một DBMS có thể hoạt động tốt với quy mô internet , người ta phải chọn một cơ sở dữ liệu phân tán mà có thể được thu nhỏ theo chiều ngang trên các nút khác nhau.định lý CAP định lý [2] cung cấp cơ sở cho sự lựa chọn đảm bảo thuộc tính trong kho dữ liệu phân tán. định lý CAP, cũng được gọi là định lý của nhà sản xuất bia phát biểu rằng không thể có một lưu trữ dữ liệu phân tán để cung cấp đồng thời nhiều hơn hai tính nhất quán (C), tính khả dụng (A) và phân vùng dung sai (P).

consistency: Mọi lần đọc đều nhận được thông tin gần đây nhất viết hoặc một lỗi

availibiaty: Mọi yêu cầu đều nhận được phản hồi không phải là lỗi

Partition Tolerance: Hệ thống tiếp tục hoạt động bất chấp số lượng tin nhắn tùy ý bị rớt (hoặc trễ) bởi mạng giữa điểm giao



Hình .Định lí CAP cho cơ sở dữ liệu phân tán

Không có hệ thống phân tán nào an toàn trước sự cố mạng; do đó, phân vùng mạng thường phải được chấp nhận. Vì thế Định lý CAP ngụ ý rằng với sự có mặt của phân vùng sức chịu đựng người ta phải lựa chọn giữa tính nhất quán và khả dụng. Trong trường hợp, khi tính nhất quán được chọn trên sẵn có, nếu thông tin cụ thể không thể được đảm bảo được cập nhật do phân vùng mạng, hệ thống sẽ trả về lỗi hoặc hết thời gian chờ. Mặt khác, nếu tính khả dụng được chọn dựa trên tính nhất quán, truy xét được đảm bảo để trả về kết quả, ngay cả khi nó không được cập nhật do mạng phân vùng. Bài viết này được định dạng như sau. Phần 2 thảo luận chi tiết về các danh mục khác nhau của cơ sở dữ liệu NoSQL. Phần 3 đưa ra sự so sánh các loại khác nhau của cơ sở dữ liệu NoSQL và cuối cùng nó đưa ra kết luận.

II. 2. DANH MỤC CƠ SỞ DỮ LIỆU NOSQL:

Cơ sở dữ liệu NoSQL [3] có thể được phân loại thành các loại sau

danh mục chính:

Cơ sở dữ liệu dựa trên cặp khóa – giá trị

Cơ sở dữ liệu dựa trên cột

Cơ sở dữ liệu dựa trên tài liệu

Cơ sở dữ liệu dựa trên đồ thị

A. 2.1 Cơ sở dữ liệu NoSQL cặp khóa-giá trị:

Cơ sở dữ liệu cặp khóa-giá trị là cách đơn giản nhất để Cơ sở dữ liệu NoSQL. Điều này rất giống với một cuốn từ điển, nơi vòng lặp xảy ra bằng cách khớp các phím và giá trị tương ứng được trả về. Giá trị có thể là một giá trị duy nhất hoặc một tập hợp dữ liệu phức tạp.

B. 2.2 Cơ sở dữ liệu NoSQL dựa trên cột:

Cơ sở dữ liệu NoSQL dựa trên cột là phần mở rộng của cơ sở dữ liệu đơn giản cơ sở dữ liệu cặp giá trị khóa trong đó mỗi bản ghi là một tập hợp cột và số lượng cột có thể khác nhau trong mỗi bản ghi. Không giống như cơ sở dữ liệu quan hệ, không có lược đồ thiết lập cho Hồ sơ. Điều này về cơ bản tương tự như một mảng 2 chiều trong đó mỗi khóa có một hoặc nhiều cặp giá trị khóa được đính kèm nó.

C. 2.3 Cơ sở dữ liệu NoSQL dựa trên tài liệu:

Cơ sở dữ liệu dựa trên tài liệu hoạt động theo cách tương tự như cơ sở dữ liệu cột, tuy nhiên chúng cho phép sâu hơn nhiều làm tổ để đạt được. Tài liệu có thể được lưu trong một tệp phẳng hoặc được cấu trúc dưới dạng định dạng json hoặc xml giúp dễ dàng áp dụng các công cụ phân tích cú pháp json/xml tiêu chuẩn để trích xuất dữ liệu từ một tài liệu.

D. 2.4 Cơ sở dữ liệu NoSQL dựa trên đồ thị:

Các mô hình cơ sở dữ liệu dựa trên đồ thị biểu diễn dữ liệu một cách hoàn chỉnh cách khác so với 3 mô hình khác. Họ sử dụng cây giống như các cấu trúc với các nút và các cạnh để kết nối chúng với một quan hệ. Chủ yếu được sử dụng trong các nền tảng truyền thông xã hội, bởi vì một số hoạt động nhất định hoạt động tốt hơn với những hoạt động này do chúng bản chất của việc liên kết và nhóm các mẩu thông tin liên quan.

III. 3. SO SÁNH CƠ SỞ DỮ LIỆU NOSQL:

Bảng dưới đây cung cấp sự so sánh của các loại khác nhau của cơ sở dữ liệu NoSQL. Chúng tôi đã chọn hiệu suất, khả năng mở rộng, nỗ lực bảo trì và các trường hợp sử dụng áp dụng cho từng loại cơ sở dữ liệu để chúng tôi so sánh.

Bảng 1. cung cấp các kết quả so sánh.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Lưu trữ dữ liệu Key Value | Cơ sở dữ liệu cột | Cơ sở dữ liệu tài liệu | Cơ sở dữ liệu dựa trên đồ thị |
| Hiệu suất | Hiệu suất cao | Hiệu suất cao | Có hiệu quả | Hiệu suất cao |
| Khả năng mở rộng | Khả năng mở rộng | Khả năng mở rộng | Có thể mở rộng | Có thể mở rộng |
| Khả năng bảo trì | Cao linh hoạt là dễ dàng | Cao không linh hoạt | Linh hoạt với phương tiện nỗ lực duy trì | Linh hoạt với phương tiện nỗ lực duy trì |
| Các trường hợp sử dụng áp dụng | * Bộ nhớ đệm * Xếp hang * Triển khai pub-sub * Giữ thông tin | * Giữ biến động phi cấu trúc dữ liệu * Cập nhập thời gian thực | * Giữ thông tin lồng nhau * Javacript thân thiện bằng cách sử dụng json/xml, | * Xử lysd phức hợp quan hệ thông tin * Sự điều khiển phân loại và làm mẫu |
| Sản phẩm có sẵn | * Redis * Riak * Bộ nhớ đệm DB | * Cassendra * Hbase | * CouchBase   CouchDB  MongoDB   * Azure Document DB | * Orient DB * Neo4j |

IV. PHẦN KẾT LUẬN:

Cơ sở dữ liệu NoSQL phù hợp nhất với quy mô internet ứng dụng đòi hỏi hiệu suất gần thời gian thực và khả năng mở rộng. Các mô hình đơn giản như cặp Giá trị chính rất dễ sử dụng cho các giá trị dữ liệu đơn giản trong khi dựa trên cột và cơ sở dữ liệu dựa trên tài liệu có thể lưu trữ các tập dữ liệu phức tạp hơn. Hơn nữa, cơ sở dữ liệu dựa trên tài liệu có hiệu quả cao ở lưu trữ các mối quan hệ lồng nhau trong khi cơ sở dữ liệu dựa trên biểu đồ hiệu quả hơn trong việc lưu trữ các quan hệ phi cấu trúc. Sự lựa chọn lựa chọn cơ sở dữ liệu nào phụ thuộc nhiều vào ứng dụng và trường hợp sử dụng áp dụng. Lựa chọn có sẵn Sản phẩm NoSQL DBMS cho một trường hợp sử dụng cụ thể cũng là được thực hiện bởi thuộc tính được đảm bảo (sẵn có so với tính chắc chắn ) theo định lý CAP.