Nghiên cứu so sánh về Cơ sở dữ liệu phi quan hệ và cơ sở dữ liệu quan hệ

## Tóm tắt:

Relational Database (Cơ sở dữ liệu quan hệ) và NoSQL(Cơ sở dữ liệu phi quan hệ) là các loại mô hình cơ sở dữ liệu đang cạnh tranh với nhau. Relational Database tồn tại từ năm 1979 và NoSQL ra đời từ năm 2000. Bởi nhu cầu của các ứng dụng hiện đại, đặc biệt là các trang Web 2.0, 3.0, nhờ nguồn dữ liệu lớn đã biến NoSQL trở thành databases phổ biến được lựa chọn. Việc chọn một mô hình databases phù hợp để sử dụng là một quyết định quan trọng mà những developer phải làm dựa trên các tính năng của một mô hình databases. Bài báo này so sánh các tính năng của Relational Database và NoSQL nhằm củng cố databases nào hỗ trợ nhu cầu các ứng dụng hiện đại tốt hơn. Bài báo cũng đưa ra những thách thức của NoSQL. Cuối cùng, bài báo kết luận bằng cách xác định liệu Relational Database có thể hoàn toàn bị thay thế bởi mô hình databases NoSQL hay không. Những phát hiện cho thấy Relational Database dự trên mô hình ACID, mà mô hình này nhấn mạnh tính nhất quán, bảo mật và cung cấp ngôn ngữ truy vấn tiêu chuẩn tốt hơn. Tuy nhiên, Relational Databases có khả năng mở rộng kém, hiệu năng yếu, tốn nhiều chi phí hơn, gặp bất lợi khi hỗ trợ số lượng lớn người dùng và xử lí lượng dữ liệu lớn một cách khó khăn. Ngược lại, NoSQL dựa trên mô hình BASE, mô hình này chú trọng khả năng mở rộng và cung cấp lược đồ linh hoạt, cung cấp hiệu suất tốt hơn, chủ yếu mà mã nguồn mở, giá rẻ nhưng thiếu ngôn ngữ truy vấn tiêu chuẩn và không cung cấp đầy đủ các cơ chế bảo mật. Cả hai databases sẽ tiếp tục tồn tại song song với nhau mà không có cái nào tối ưu hơn. Việc lựa chọn databases để sử dụng sẽ phụ thuộc vào đặc tính của ứng dụng đang phát triển. Mỗi loại databases đều có những ưu và nhược điểm riêng, với relational database thì thiếu hỗ trợ cho dữ liệu phi cấu trúc trong khi NoSQL thiếu tiêu chuẩn hóa và bảo mật kém. Những ứng dụng hiện đại trong web 2.0, 3.0 và và dữ liệu lớn rất thích hợp để sử dụng NoSQL những vẫn nhiều các ứng dụng dựa vào Relational Databases.

## Giới thiệu

Cơ sở dữ liệu đã thay thế các flat file (tệp phẳng) làm một kho lưu trữ các nhóm dữ liệu lớn. Từ khi cơ sở dữ liệu ra đời, Relation Databases đã làm chủ hơn 30 năm cho đến năm 2000 khi mà NoSQL dần thay thế chúng trong một số ứng dụng. Bản chất cần thiết của ứng dụng là thay đổi theo thời gian và dẫn đến các ứng dụng có tương tác cao, xử lý được khối lượng dữ liệu lớn, chẳng hạn như các trang thương mại điện tử và phương tiện truyền thông xã hội. Việc cung cấp các tính năng tương tác với một cơ sở dữ liệu là một yêu cầu quan trọng đối với những cơ sở dữ liệu xây dựng nên ứng dụng web 2.0 và 3.0. Sự thay đổi trong nhu cầu ứng dụng đã khiến Relational Databases không ứng được nhu cầu của nhà phát triển và người dùng. Các công ty như Google, Facebook và Yahoo đã chuyển sang dùng NoSQL để giảm thiểu những thiếu sót của Relational Databases. Bất chấp xu hướng trên, nhiều công ty vẫn sử dụng Relational Databases. Các công ty này có đặc điểm là khối lượng dữ liệu hạn chế đồng thời đề cao tính chắc chắn.

Relational Databases dựa trên i.e (Internet Explorer) mô hình ACID. Atomicity (Tính nguyên tử), Consistency (Tính nhất quán), Isolation (Tính độc lập), và Durability (Tính bền vững). Atomicity đảm bảo tính hoàn chỉnh của giao dịch. Consistency mang lại sự ổn định của dữ liệu trong cơ sở dữ liệu, Isolation đảm bảo được sự độc lập của các giao dịch được thực hiện cùng lúc và Durability đảm bảo các giao dịch được lưu trữ dù có xảy ra sự cố. ACID cung cấp sự chắc chắn và tiện lợi như những đặc tính mạnh mẽ giúp Relational Databases trở nên phổ biến. Mặt khác, NoSLQ dựa trên mô hình BASE (Basically Available, Soft State and Eventually Consistent). Bản chất phân tán của NoSQL mang đến phần nào có ích của dữ liệu khi một số phần của cơ sở dữ liệu phân tán không hoạt động hoặc không thể truy cập vì thế, thuật ngữ Basically Available (Tính sẳn sàng ở mức cơ bản). Soft State (Trạng thái mềm) cho phép dữ liệu thay đổi theo thời gian hoặc không cần đầu vào. Cuối cùng, Consistent (Tính nhất quán) sẽ đảm bảo dữ liệu nhất quán trong tương lai chứ không phải sau một thao tác. BASE cung cấp cho NoSQL khả năng mở rộng quy mô dễ dàng, mang lại hiệu suất tốt hơn và mang lại sự tiện lời cao hơn cho người dùng.

Bài viết này dựa trên những đánh giá của những tài liệu trước đây và bắt đầu bằng việc mô tả về Relational Databases và mô hình cơ sở dữ liệu NoSQL. Sau đó, cuộc tranh luận chuyển sang so sánh cách tính năng của Relational Databases và NoSQL, sau đó là những thách thức của NoSQL. Dựa trên những tính năng, các tác giả đã cố gắng xác định xem NoSQL có tốt hơn Relational Databases trong việc hỗ trợ nhu cầu của các ứng dụng cơ sở dữ liệu hiện đại và liệu NoSQL có thể thay thế hoàn toàn Relational Databases hay không.

## RELATIONAL DATABASE và cac loại NoSQL

Chỉ có một dạng của Relation Database được dựa trên mô hình quan hệ. Nhiều tổ chức, đã thay đổi các yêu cầu ứng dụng của họ để phù hợp với sự nghiêm ngặt của lược đồ đã được cố định từ trước ở Relational Databases. Lược đồ nghiêm ngặt yêu cầu ứng dụng phải tuân theo yêu cầu của cơ sở dữ liệu thay vì ngược lại. Những ví dụ về Relational Databases là MySQL, Microsoft SQL Server và Postgres.

Cơ sở dữ liệu NoSQL có khá nhiều loại và từ những tài liệu đã được đánh giá, có bốn loại nổi bật. Chúng bao gồm Key-Value (Khóa – Giá trị), Document oriented (hướng tài liệu), Column databases (cơ sở dữ liệu dạng cột) and Graph databases (cơ sở dữ liệu đồ thị). Trong Key-Value, dữ liệu được lưu trữ dưới dạng tập hợp các khóa và giá trị, trong đó khóa là thành phần duy nhất trong cơ sở dữ liệu được xác định bởi thuộc tính của nó và giá trị là giá trị của thuộc tính. Key – Value dễ sử dụng nhưng nó không hỗ trợ xử lý mối quan hệ giữa các mục dữ liệu. Ví dụ về cơ sở dữ liệu key value bao gồm Memcached và Redis. Cơ sở dữ liệu Document oriented sử dụng khóa và tài liệu làm thuộc tính trong đó khóa đề cập đến toàn bộ tài liệu. Ví dụ về cơ sở dữ liệu Document oriented là MongoDB và CouchDB, rất phù hợp để xử lý các cấu trúc dữ liệu phức tạp những vẫn thiếu khả năng xử lý các mối quan hệ giữa các mục dữ liệu. Column databases chứa các hàng và cột tương tự như Relational Databases nhưng mỗi cột được lưu trữ trong một tệp riêng biệt. Một khóa trong cơ sở dữ liệu column oriented ám chỉ đến một cột. Các giá trị thuộc tính khác được lưu trữ bao gồm giá trị và thời gian đánh dấu. Bigtable và Cassandra là những ví dụ và theo thiết kế, chúng kém linh hoạt hơn nhưng cung cấp năng suất lớn hơn. Graph databases biểu thị dữ liệu dưới dạng đồ thị được kết nối và dựa trên lý thuyết đồ thị. Graph databases ít có khả năng mở rộng hơn nhưng hỗ trợ kết nối tốt hơn. Các ví dụ bao gồm GraphDB và OrientDB.