**Giới thiệu về NoSQL:**

Thuật ngữ “NoSQL” là một cái tên dùng cho các middeware như database hướng document và KVS(Key Value Store). Các công ty dịch vụ web hàng đầu như Twitter, Facebook,… trở thành khởi nguồn cho việc bùng nổ của NoSQL. Vậy ở đây NoSQL thực chức là gì? Mọi người thường nghỉ “NoSQL” loại bỏ RDBMS (Relational Database Management System- Hệ thống quản lý cơ sở dữ liệu quan hệ) và thay thế nó bằng KVS,thực chất nó chỉ thay thế những phần RDBMS làm không tốt bằng các middleware. Như vậy chính xác nó mang ý nghĩa đó là Not Only SQL.

**So sánh NoSQL với RDBMS:**

|  |  |  |
| --- | --- | --- |
|  | NoSQL | RDBMS |
| Điểm quan trọng | Mở rộng quy mô, tính khả dụng cao | Tính nhất quán |
| Performance | Mở rộng quy mô bằng cách sắp xếp các máy chủ cạnh nhau(side by side) | Chia tỷ lệ hoặc chia nhỏ dữ liệu theo chiều ngang |
| Query | Query với một key khóa đơn giản | Query bằng SQL |
| Tính nhất quán | Linh động, lỏng lẻo | Mạnh mẽ |
| Data model | Nhiều mô hình như mô hình định hướng cột, khóa và giá trị thuần túy,… | Mô hình quan hệ |

**Các đặc trưng chi tiết của NoSQL middeware:**

+ Hoạt động nhanh

+ Là data model chứ không phải là relation model

+ Kiến trúc tỷ lệ

+ Được xây dựng bởi commodity sever

+ Giản đồ miễn phí

+ Không có SPOF((một điểm gây tổn hại))

+ Tự động sao chép sang nhiều đơn vị

+ Cho phép lựa chọn tính nhất quán hoặc nhất quán cuối cùng

+ Không có ngôn ngữ truy vấn mạnh như SQL và chỉ có thể thực hiện các truy vấn đơn giản

**Cassandra là gì?**

Cassandra là một cơ sở dữ liệu phân tán kết hợp mô hình dữ liệu của Google Bigtable với thiết kế hệ thống phân tán như bản sao của Amazon Dynamo. Tất cả được thực hiện bởi Java và được cung cấp trong ASL 2 (Giấy phép Phần mềm Apache Phiên bản 2). Nó ban đầu được tạo ra bởi Facebook. Nhưng sau đó được tặng cho Quỹ Apache vào tháng 2 năm 2010 và được nâng cấp lên dự án hàng đầu của Apache.

**Các tính năng ưu việt mà Cassandra sở hữu đó là:**

1. Thích hợp để sử dụng thực tế
2. Khả năng chịu lỗi cao
3. Kiến trúc không có SPOF (một điểm gây tổn hại)
4. Mức độ tự do kiểm soát nhất quán
5. Mô hình dữ liệu phong phú
6. Có thể tăng cường cải thiện thông lượng cho tuyến tính
7. Tính khả dụng cao
8. Hỗ trợ các ngôn ngữ khác nhau dưới dạng client code
9. Dễ dàng nắm bắt trạng tháy bên trong của máy chủ bằng JMX/ Dễ giám sát

Các đặc điểm đặc biệt hấp dẫn là 3 đặc trừng 3,5 và 8.

Thứ nhất Kiến trúc không có SPOF nó sẻ khiến cho phần mềm sẻ không dừng khi có một chổ bị lỗi tráng hư hao và khả năng chịu lỗi của nó dẫn đến là rất cao.

Thứ hai Mô hình dữ liệu phong phú. Vì nó được kết hợp giữa Google Bigtable và Amazon Dynamo nên nó có nhiều mô hình dữ liệu của Bigtable, giúp cho việc hình dung hình ảnh mình đang tính toán dễ dàng hơn.

Thứ ba là Hồ trợ các ngôn ngữ khác nhau dưới dạng client code. Vì Cassandra thu thập dữ liệu bằng framework có tên là Thrift. Thirft có một cơ chế để giao tiếp với nhiều ngôn ngữ khác nhau vì vậy Cassandra có thể chạy bằng nhiều ngôn ngữ

Nút

Nút là nơi lưu trữ dữ liệu. Nó là thành phần cơ bản của Cassandra.

Trung tâm dữ liệu

Một tập hợp các nút được gọi là trung tâm dữ liệu. Nhiều nút được phân loại là một trung tâm dữ liệu.

Cụm

Các cụm là tập hợp của nhiều trung tâm dữ liệu.

Nhật ký cam kết

Mỗi thao tác ghi được ghi vào Commit Log. Nhật ký cam kết được sử dụng để phục hồi sự cố.

Bảng ghi nhớ

Sau khi dữ liệu được ghi trong Nhật ký cam kết, dữ liệu được ghi trong bảng Mem. Dữ liệu được ghi trong bảng Mem tạm thời.

SSTable

Khi bảng Mem đạt đến một ngưỡng nhất định, dữ liệu sẽ được chuyển sang tệp đĩa SSTable.