

Giới thiệu Apache Cassandra

Apache Cassandra là cơ sở dữ liệu dạng cột (column family) được phát triển bởi Facebook trong năm 2008, sau khi trở thành sản phẩm của Facebook trong một thời gian ngắn, Cassandra trở thành dự án open source tại Google code vào tháng 7 năm 2008. Tháng 3 năm 2009, Cassandra trở thành dự án tiềm năng của tổ chức Apache và không lâu sau, vào tháng 2 năm 2010 Cassandra đã trở thành một trong những project nổi bật nhất của Apache. (Theo wikipedia.org)



1. Là cơ sở dữ liệu hướng cột (column-family)

Cơ sở dữ liệu dạng cột (column-family) lưu trữ dữ liệu theo nhiều cột trong mỗi dòng (row key). Có thể coi column-family như là một table trong cơ sở dữ liệu quan hệ.

2. Khóa chính (Primary key)

Được thiết kế như partition key. Nếu primary key bao gồm nhiều column thì:

- Column đầu tiên sẽ là partition key: Chịu trách nhiệm phân tán dữ liệu tới các node trong hệ thống
- Các column phía sau là clustering key: Sắp xếp thứ tự dữ liệu trong mỗi partition key.
- Ưu điểm: So với cơ sở dữ liệu RDMS truyền thống (như MySQL, Oracle) dạng dòng (row), khi truy xuất dữ liệu cần thực hiện lấy nhiều row cần scan theo index key trên vùng rộng

lớn (full-scan), trong khi đó, Cassandra tổ dữ dữ liệu hướng cột (column), khi đó cần truy cập dữ liệu theo partition key, chỉ cần đọc block dữ liệu duy nhất, do đó sẽ giảm thời gian và tải hệ thống khi truy vấn dữ liệu.

3. Lưu trữ và xử lý phân tán

Trong Cassandra, dữ liệu được lưu trữ phân tán trên các máy chủ của cụm để đảm bảo tính sẵn sàng cao, đồng thời tối ưu việc xử lý truy vấn dữ liệu trên các máy chủ mà dữ liệu được lưu trữ.

4. Đảm bảo an toàn

Trong Cassandra, mỗi đối tượng dữ liệu có thể được nhân bản và lưu giữ trên nhiều máy chủ. Nếu một trong các máy chủ lưu một phiên bản dữ liệu bị lỗi hoặc không phải là phiên bản được cập nhật dữ liệu mới nhất, Cassandra có cơ chế đồng bộ để luôn đảm bảo các thao tác đọc sẽ luôn trả về dữ liệu mới nhất.

5. Hỗ trợ đa dạng mô hình

Hỗ trợ lưu trữ các dạng dữ liệu:

- Key-value
- Cơ sở dữ liệu cấu trúc: Structure
- Cơ sở dữ liệu đồ thị: Graph

6. Khả năng chịu lỗi cao

Nếu một trong các máy chủ lưu một phiên bản dữ liệu bị lỗi hoặc không phải là phiên bản được cập nhật dữ liệu mới nhất, Cassandra có cơ chế đồng bộ để luôn đảm bảo các thao tác đọc sẽ luôn trả về dữ liệu mới nhất. Đồng thời với việc này Cassandra tiến hành thao tác sửa lỗi đọc (read repair) là tiến trình ngầm để cập nhật trạng thái mới nhất cho tất cả các máy chủ lưu trữ nhân bản của dữ liệu.

7. Tính khả dụng, mở rộng cao

- Cassandra cho phép mở rộng cơ sở dữ liệu bằng cách bổ sung thêm node vào cluster một cách dễ dàng.
- Hỗ trợ nhiều ngôn ngữ lập trình ở client side: C++, Java, Python, PHP, Ruby, Erlang, C#...

Revision #1

Created 25 June 2021 16:25:57 by Laptrinh.vn

Updated 25 June 2021 16:31:21 by Laptrinh.vn