

# So sánh và đánh giá Cassandra và HBase

Cassandra và HBase là cơ sở dữ liệu phổ biến thuộc loại Column family.



## 1. Giống nhau

#	Thành phần	Nội dung
1	Database	Cả Cassandra và HBase đều là CSDL mã nguồn mở thuộc NoSQL. Có thể lưu trữ và xử lý lượng lớn dữ liệu bao gồm cả dữ liệu không có quan hệ.
2	Scalability (khả năng mở rộng)	Cả Cassandra và HBase đều có khả năng mở rộng cao, bằng cách tăng số lượng node trong cluster. Bởi vậy, cả hai đều là lựa chọn tốt trong việc xử lý dữ liệu lớn
3	Replication (nhân bản)	Luôn có khả năng xảy ra lỗi trong một chương trình hoặc ứng dụng, do đó, dữ liệu có thể bị mất. Nhưng cả Cassandra và HBase đều có giải pháp để bảo vệ và ngăn chặn việc mất dữ liệu ngay cả khi máy chủ bị lỗi. Điều này được thực hiện thông qua replication. Dữ liệu được ghi trên một node được sao chép đến nhiều node trong cluster. Do đó, nếu một node bị lỗi, luôn có một node dự phòng để truy cập dữ liệu.

## 2. Khác nhau

#	Thành phần	Cassandra	HBase
1	Infrastructure (Hạ tầng)	Cassandra được thiết kế để triển khai độc lập. Tuy nhiên, một số ứng dụng có thể kết hợp Cassandra với các DBMS khác hoặc cũng có thể triển khai tích hợp với Storm hoặc Hadoop.	HBase sử dụng cơ sở hạ tầng Hadoop. Cơ sở hạ tầng HBase-Hadoop này bao gồm một số thành phần như Zookeeper, HBase master, data node và name node.
2	Support	Cassandra hỗ trợ ordered partitioning (sắp xếp trên partition). Cassandra có một số hạn chế khi scan các row theo khoảng, và bộ đồng xử lý (coprocessor)	HBase không hỗ trợ ordered partitioning (sắp xếp trên partition) HBase cung cấp khả năng đồng xử lý. Khả năng này hỗ trợ các trình kích hoạt. Trong HBase, một row được phục vụ bởi chính xác một máy chủ tại một thời điểm. Do đó, nó không hỗ trợ cân bằng tải đọc đối với chỉ một row.
3	Node	Trong Cassandra, có các seed node. Các nút này đóng vai trò là các điểm giao tiếp giữa các cụm. Do đó, khả năng mở rộng và tính khả dụng cao trong Cassandra được đảm bảo bằng cách cho phép nhiều seed node trong một cụm.	Trong HBase, có các master node. Các nút chủ này giám sát và điều phối hoạt động của các máy chủ khu vực. Do đó, khả năng này được đảm bảo bởi các standby node trong HBase. Trong trường hợp master node bị lỗi, standby node sẵn sàng thế chỗ.
4	Giao tiếp giữa các node	Cassandra sử dụng Gossip Protocol để giao tiếp giữa các node, dữ liệu được sao chép từ node này sang node khác	HBase dựa vào giao thức Zookeeper để xác định các node, dữ liệu được sao chép, đọc ghi dữ liệu thông qua master node.
5	Query language (ngôn ngữ truy vấn)	Cassandra hỗ trợ sử dụng ngôn ngữ CQLSH tương tự ngôn ngữ SQL	HBase chỉ hỗ trợ HBase shell

Trong thực tế, nhà phát triển có thể kết hợp một hoặc một vài hệ cơ sở dữ liệu cho ứng dụng. Mỗi cơ sở dữ liệu sẽ có những điểm mạnh, điểm yếu riêng. Do đó, các nhà phát triển cần có sự hiểu biết về các hệ cơ sở dữ liệu và linh hoạt trong việc lựa chọn nền tảng nhằm đưa ra giải pháp tốt nhất cho bài toán.

Revision #2

Created 26 June 2021 04:03:49 by Laptrinh.vn

Updated 26 June 2021 04:08:39 by Laptrinh.vn