

So sánh hệ thống cơ sở dữ liệu hướng cột và hướng hàng

Một kho dữ liệu về cơ bản là nơi lưu trữ các tập dữ liệu được thu thập, chẳng hạn như cơ sở dữ liệu, hệ thống tệp hoặc thư mục. Trong hệ thống cơ sở dữ liệu, chúng có thể được lưu trữ theo những cách sau:

- Lưu trữ dữ liệu hướng theo hàng
- Lưu trữ dữ liệu hướng cột

Trong cơ sở dữ liệu quan hệ, dữ liệu luôn được truy cập theo hàng. Việc thay đổi cách truy cập từ hàng thành cột có thể cải thiện hiệu suất khi số lượng lớn dữ liệu được tổng hợp trong một vài cột. Bằng cách lưu trữ dữ liệu trong các cột thay vì các hàng, cơ sở dữ liệu có thể truy cập chính xác hơn dữ liệu cần trả lời truy vấn thay vì quét và loại bỏ dữ liệu không mong muốn trong các hàng. Hiệu suất truy vấn được tăng cho một số khối lượng công việc nhất định.

So sánh giữa cách lưu trữ dữ liệu hướng hàng và lưu trữ dữ liệu theo định hướng cột như sau:

dữ liệu hướng hàng	dữ liệu hướng cột
Dữ liệu được lưu trữ và lấy ra từ một hàng tại một thời điểm và do đó có thể sẽ đọc dữ liệu không cần thiết nếu chỉ có một số dữ liệu trong một hàng được yêu cầu.	Dữ liệu được lưu trữ và truy xuất trong các cột và do đó nó chỉ có thể đọc dữ liệu có liên quan nếu cần.
Các bản ghi trong lưu trữ dễ đọc và dễ viết.	hoạt động đọc và ghi chậm hơn so với định hướng hàng.
Các hệ thống lưu trữ dữ liệu định hướng hàng phù hợp nhất cho hệ thống giao dịch trực tuyến.	việc xử lý phân tích trực tuyến.
không hiệu quả trong việc thực hiện các hoạt động áp dụng cho toàn bộ các tập dữ liệu và do đó tập hợp theo định hướng hàng là một công việc khó khăn	hiệu quả trong việc thực hiện các hoạt động áp dụng cho toàn bộ các tập dữ liệu và vì vậy cho phép tập hợp qua nhiều hàng và cột.
Cơ chế nén điển hình cung cấp kết quả kém hiệu quả hơn so với những gì đạt được từ việc lưu trữ dữ liệu theo định hướng cột.	Các loại lưu trữ dữ liệu này về cơ bản cho phép tỷ lệ nén cao do ít giá trị khác biệt hoặc duy nhất trong cột.

Ví dụ:

- Về cơ chế lưu trữ dữ liệu hướng hàng là cơ sở dữ liệu quan hệ, là một lưu trữ dữ liệu có cấu trúc và cũng là một công cụ truy vấn phức tạp. Cần phải đầu tư hệ thống để cải thiện hiệu suất khi kích thước dữ liệu tăng lên.
- Với cơ chế lưu trữ dữ liệu hướng-cột là cơ sở dữ liệu HBase, được thiết kế cơ bản từ đầu để cung cấp khả năng mở rộng và phân vùng để cho phép tuần tự hóa, lưu trữ và truy xuất cấu trúc dữ liệu hiệu quả.

*HBase là một cơ sở dữ liệu định hướng cột được phân phối và xây dựng trên hệ thống tệp Hadoop. Nó là một dự án nguồn mở và có khả năng mở rộng theo chiều ngang.