

HBase là 1 CSDL phân tán, mã nguồn mở được xây dựng dựa trên Apache Software Foundation. Ban đầu, nó được gọi là Google Big Table, sau đó đổi thành HBase và được viết bởi ngôn ngữ Java. HBase có thể lưu trữ lượng lớn data.

HBase mang lại những lợi ích như có khả năng mở rộng. HBase được thiết kế để xử lý việc điều chỉnh quy mô trên hàng ngàn nghìn máy chủ quyền truy cập vào hàng petabyte dữ liệu.

- Nhanh: HBase cho phép truy cập đọc và ghi ngẫu nhiên với độ trễ thấp vào hàng petabyte dữ liệu bằng cách phân phối yêu cầu từ các ứng dụng trên một cụm máy chủ.

- Có khả năng chịu lỗi: HBase tách dữ liệu trong các bảng trên nhiều máy chủ trong cụm và được xây dựng để chịu được các lỗi máy chủ riêng

lẽ. Vì dữ liệu được lưu trữ trên HDFS hoặc S3 nên HBase sẽ tự động chọn máy chủ hoạt động tốt nhất để lưu trữ dữ liệu từng được máy chủ bị lỗi phân phối. Dữ liệu này sẽ tự động chuyển sang trạng thái trực tuyến.

Mô phỏng truy vấn phân tán trong HBase
Một hệ thống quản lý điểm danh sinh viên trong trường đại học sử dụng HBase
Mỗi sinh viên điểm danh sẽ tạo ra một bản ghi gồm:

- Mã sinh viên
- Thời gian điểm danh
- Trạng thái (Có mặt / Không)

Cách dữ liệu được phân tán

Dữ liệu trong HBase ~~được~~ được chia thành nhiều Region dựa trên RowKey
Mỗi Region được lưu trữ trên một Region Server khác nhau

Các bản ghi của sinh viên có mã nhỏ để
lưu ở Region Server 1

Các bản ghi tiếp theo được lưu ở Region
Server 2

Các bản ghi còn lại được lưu ở Region
Server 3

Mô phỏng truy vấn đọc dữ liệu phân tán
Khi người quản lý cần xem thông tin
điểm danh của một SV cụ thể:

Người dùng gửi yêu cầu truy vấn đến
hệ thống/HBase

HBase xác định Region nào đang lưu trữ
~~lưu~~ dữ liệu cần được chuyển đến đúng
Region Server chứa dữ liệu.

Region Server xử lý và trả kết quả về
cho người dùng.

Mô phỏng truy vấn quét dữ liệu trên
toàn hệ thống.

Khi cần thống kê điểm danh của tất

có sẵn hiện trong một ngày, hệ thống:
Giải yêu cầu truy vấn đến toàn bộ các
Region Server

Mỗi Region Server xử lý dữ liệu mà nó
đang lưu trữ.

Kết quả từ các Region Server được tổng hợp
lại

Hệ thống trả về kết quả cuối cùng cho
người dùng