

25 YEARS ANNIVERSARY
SOICT

HA NOI UNIVERSITY OF SCIENCE AND TECHNOLOGY
SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY



HA NOI UNIVERSITY OF SCIENCE AND TECHNOLOGY
SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

Chương 4

NoSQL - phần 1

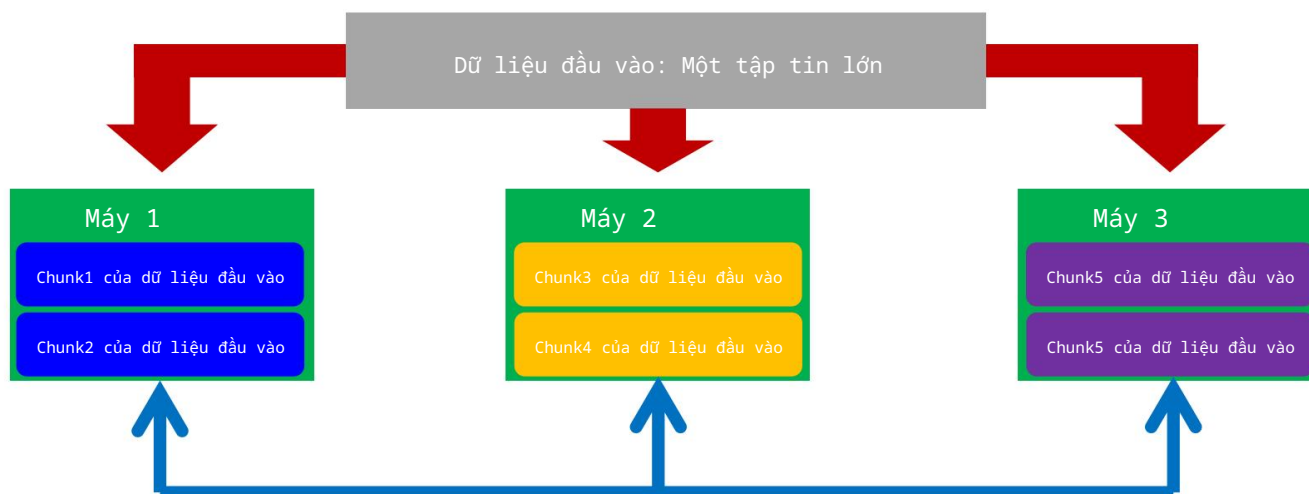
Định lý CAP

Mở rộng cơ sở dữ liệu truyền thống

- RDBMS truyền thống có thể được mở rộng theo:
 - Theo chiều dọc (hoặc Lên) • Có thể đạt được bằng cách nâng cấp phần cứng (ví dụ: CPU nhanh hơn, nhiều hơn bộ nhớ hoặc đĩa a lớn hơn) • Bị giới hạn bởi số lượng CPU, RAM và đĩa a có thể được cấu hình trên một máy duy nhất •
 - Theo chiều ngang (hoặc Ra ngoài)
 - Có thể đạt được bằng cách thêm nhiều máy hơn
 - Yêu cầu phân mảnh cơ sở dữ liệu và có thể sao chép • Bị giới hạn bởi tỷ lệ Đọc-Ghi và chi phí truyền thông

Phân mảnh dữ liệu

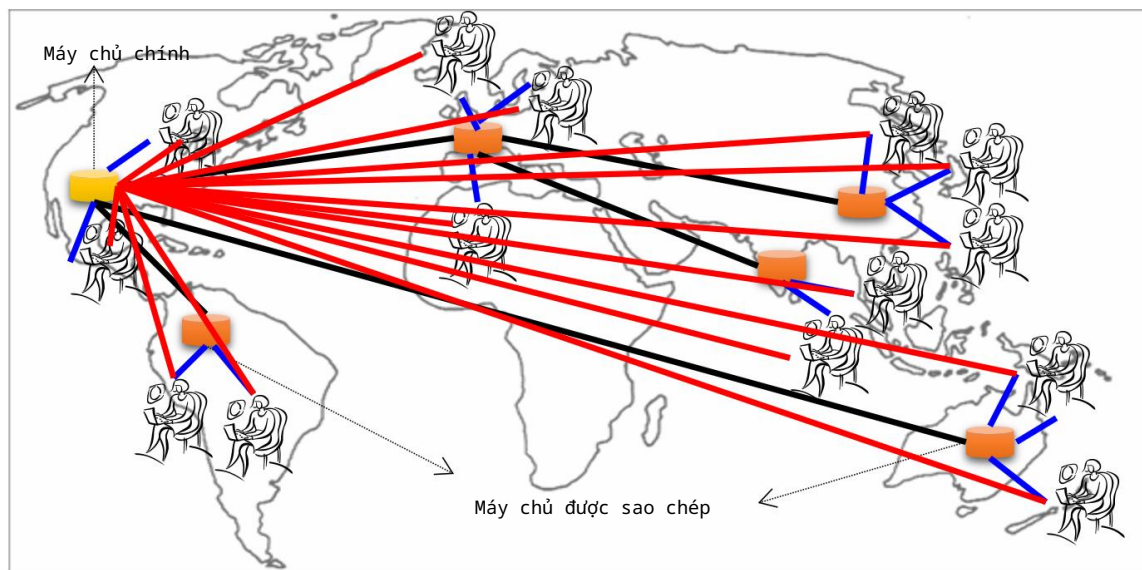
- Dữ liệu thường được phân mảnh (hoặc chia thành từng phần) để cho phép truy cập đồng thời/song song
- Liệu nó có khả năng mở rộng để xử lý truy vấn phức tạp không?



Ví dụ, các Chunks 1, 3 và 5 có thể được truy cập song song

Sao chép dữ liệu

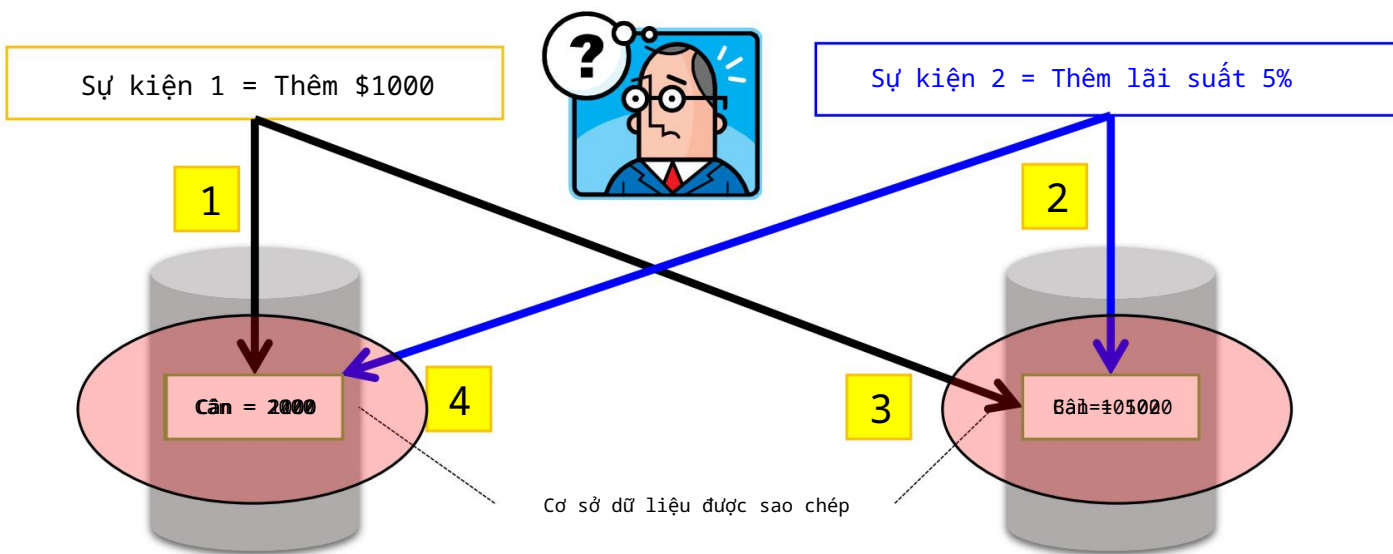
- Sao chép dữ liệu trên nhiều máy chủ giúp:
 - Tránh tình trạng tắc nghẽn hiệu suất
 - Tránh điểm lỗi đơn
 - Và do đó, tăng cường khả năng mở rộng và tính khả dụng



Nhưng, sự nhất quán trở thành một Thử thách

- Ví dụ:

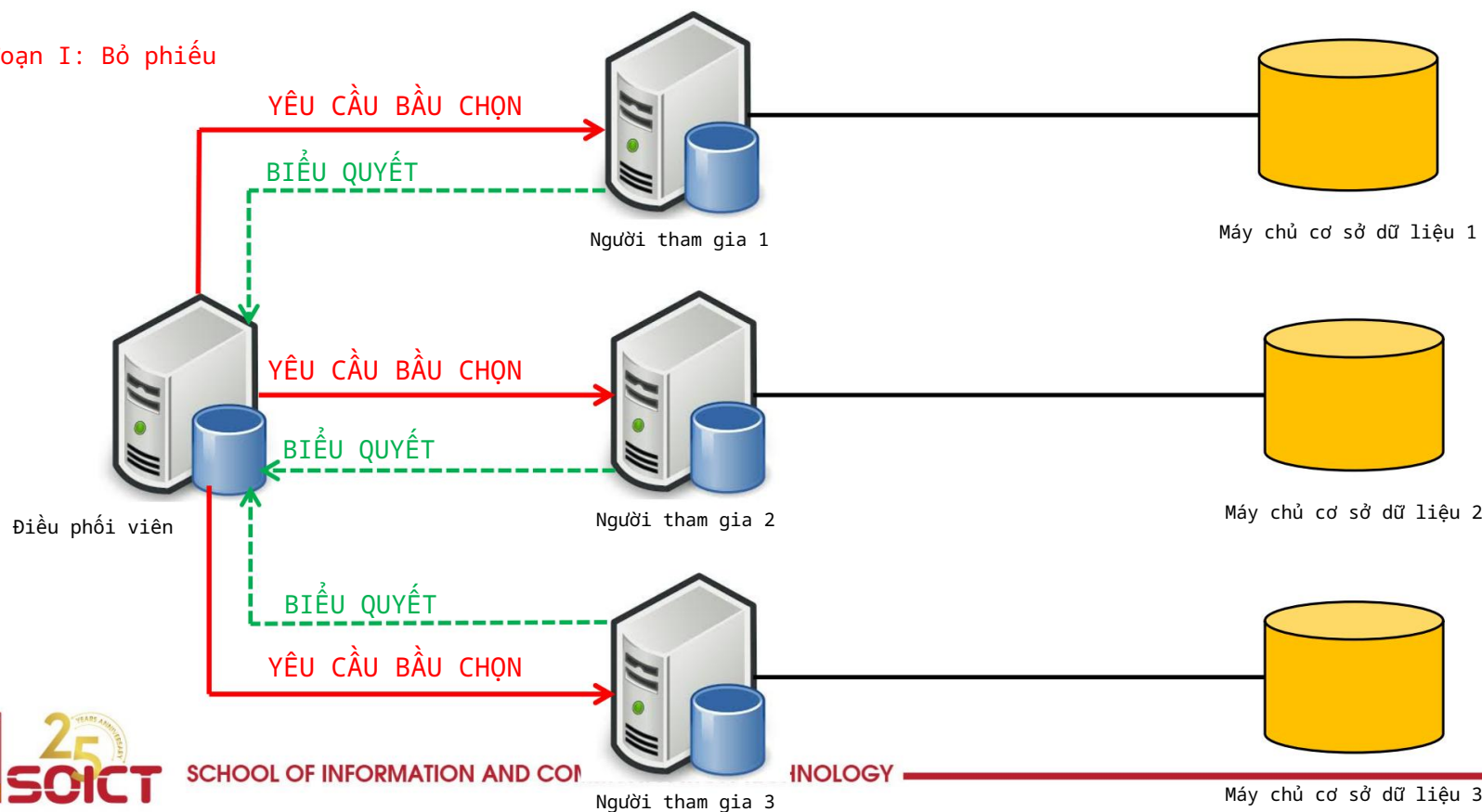
- Trong ứng dụng thương mại điện tử, cơ sở dữ liệu ngân hàng đã được sao chép trên hai máy chủ
- Duy trì tính nhất quán của dữ liệu được sao chép là một thách thức



Giao thức cam kết hai giai đoạn

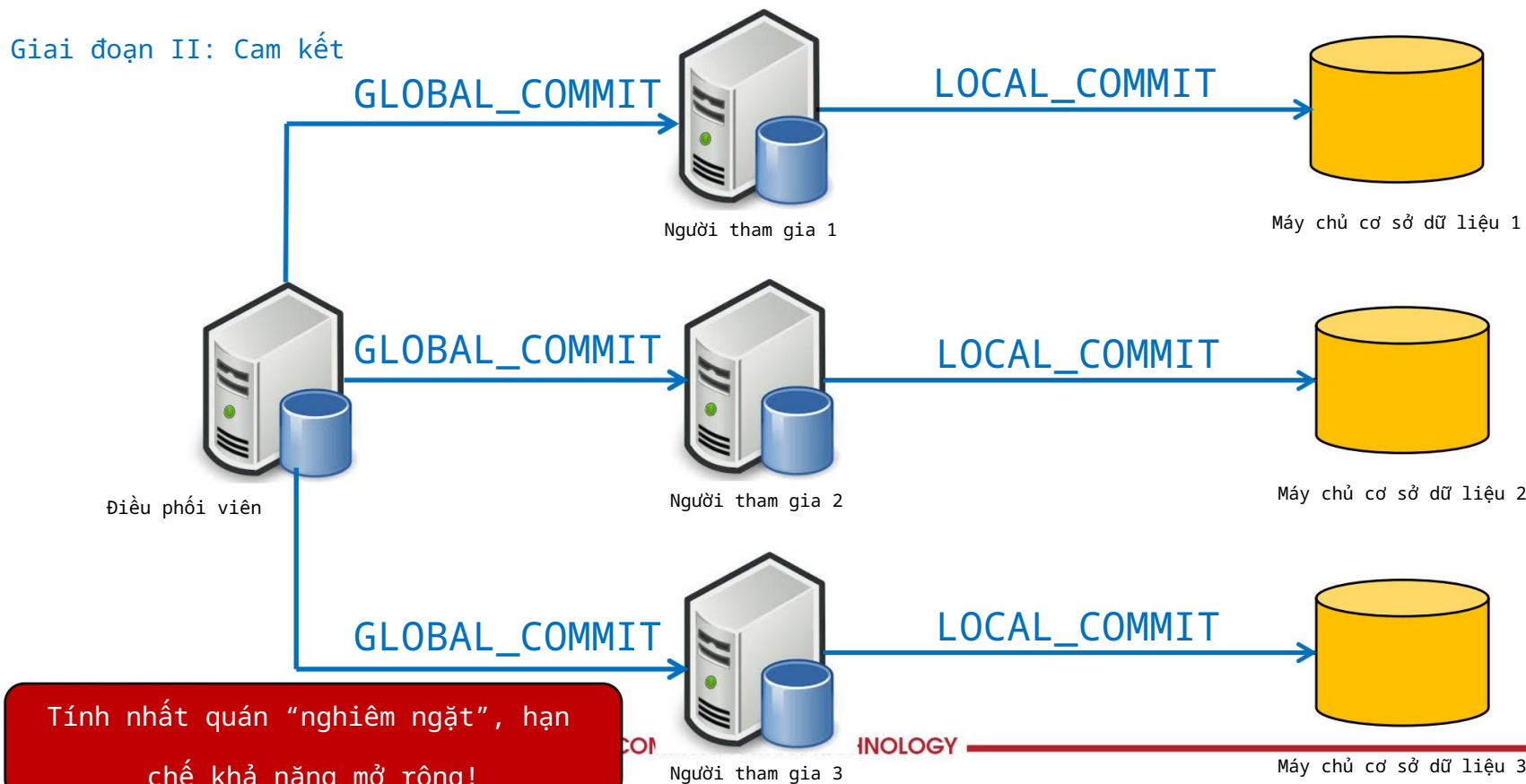
- Giao thức cam kết hai giai đoạn (2PC) có thể được sử dụng để đảm bảo tính nguyên tử và tính nhất quán

Giai đoạn I: Bỏ phiếu



Giao thức cam kết hai giai đoạn

- Giao thức cam kết hai giai đoạn (2PC) có thể được sử dụng để đảm bảo tính nguyên tử và tính nhất quán

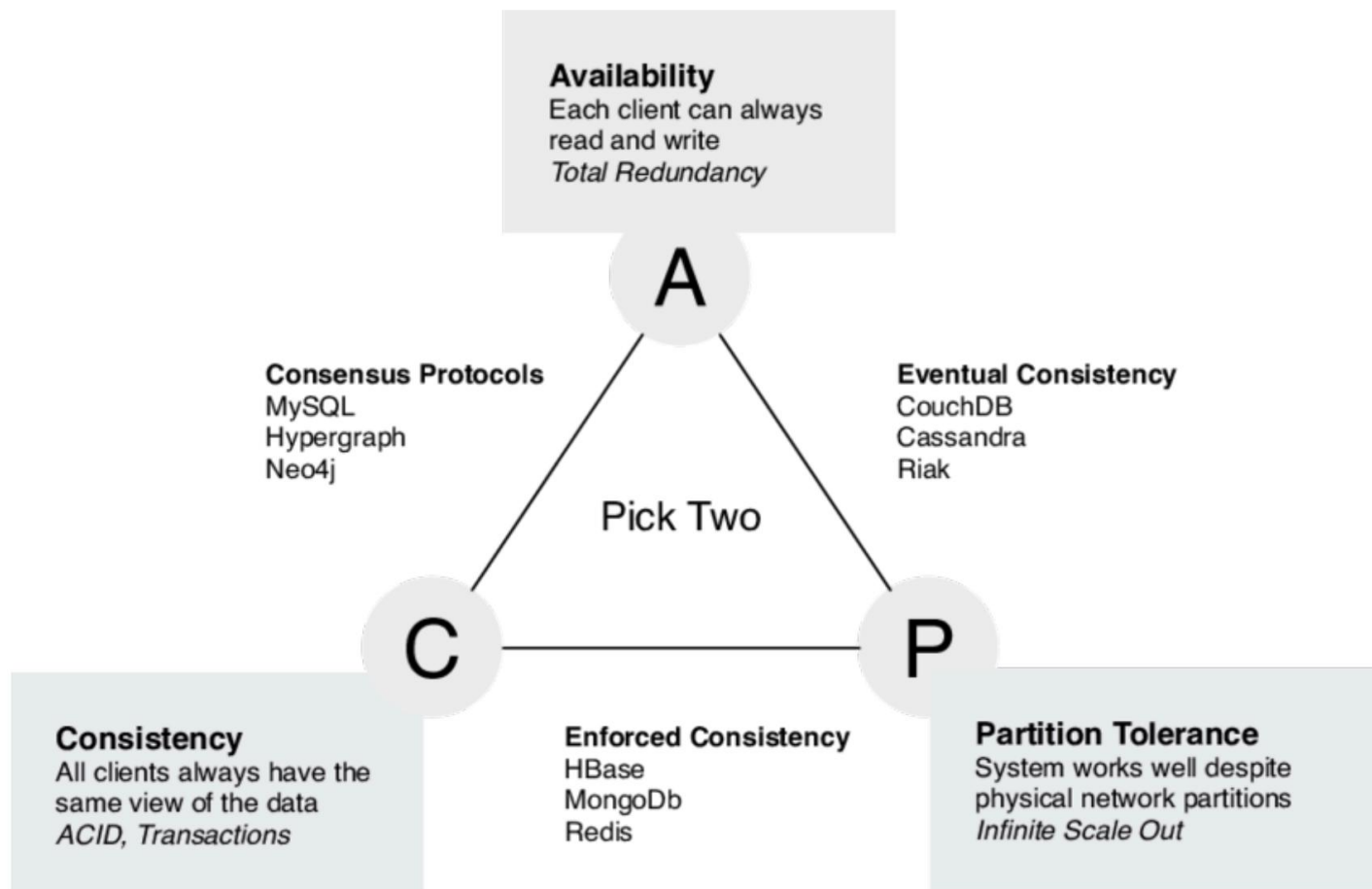


Định lý CAP

- Những hạn chế của cơ sở dữ liệu phân tán có thể được mô tả trong cái gọi là định lý CAP
 - Tính nhất quán: mọi nút luôn nhìn thấy cùng một dữ liệu tại bất kỳ trường hợp nào (tức là tính nhất quán nghiêm ngặt)
 - Tính khả dụng: hệ thống vẫn tiếp tục hoạt động, ngay cả khi các nút trong cụm bị sập hoặc một số phần cứng hoặc phần mềm ngừng hoạt động do nâng cấp
 - Dung sai phân vùng: hệ thống tiếp tục hoạt động trong sự hiện diện của các phân vùng mạng

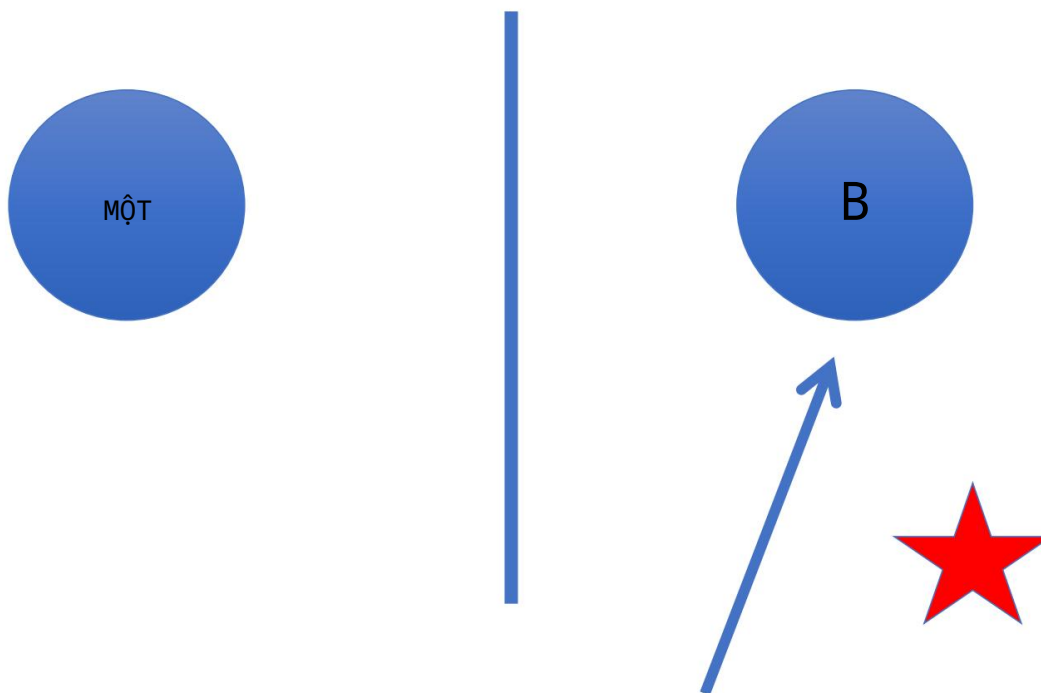
Định lý CAP: bất kỳ cơ sở dữ liệu phân tán nào có dữ liệu được chia sẻ đều có thể có tối đa hai trong ba thuộc tính mong muốn là C, A hoặc P. Đây là những sự đánh đổi liên quan đến hệ thống phân tán của Eric Brewer trong PODC 2000.

Định lý CAP



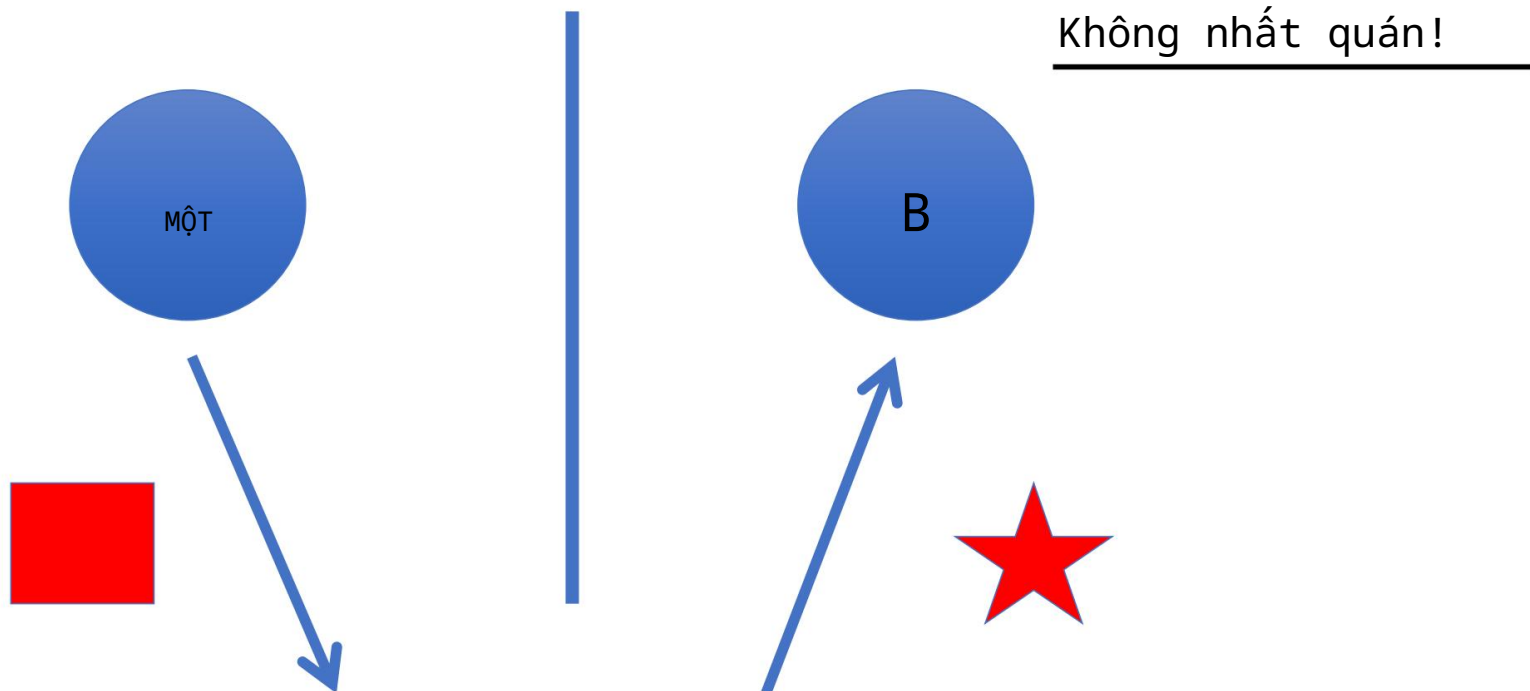
Định lý CAP: Bằng chứng

- Một bằng chứng đơn giản sử dụng hai nút:



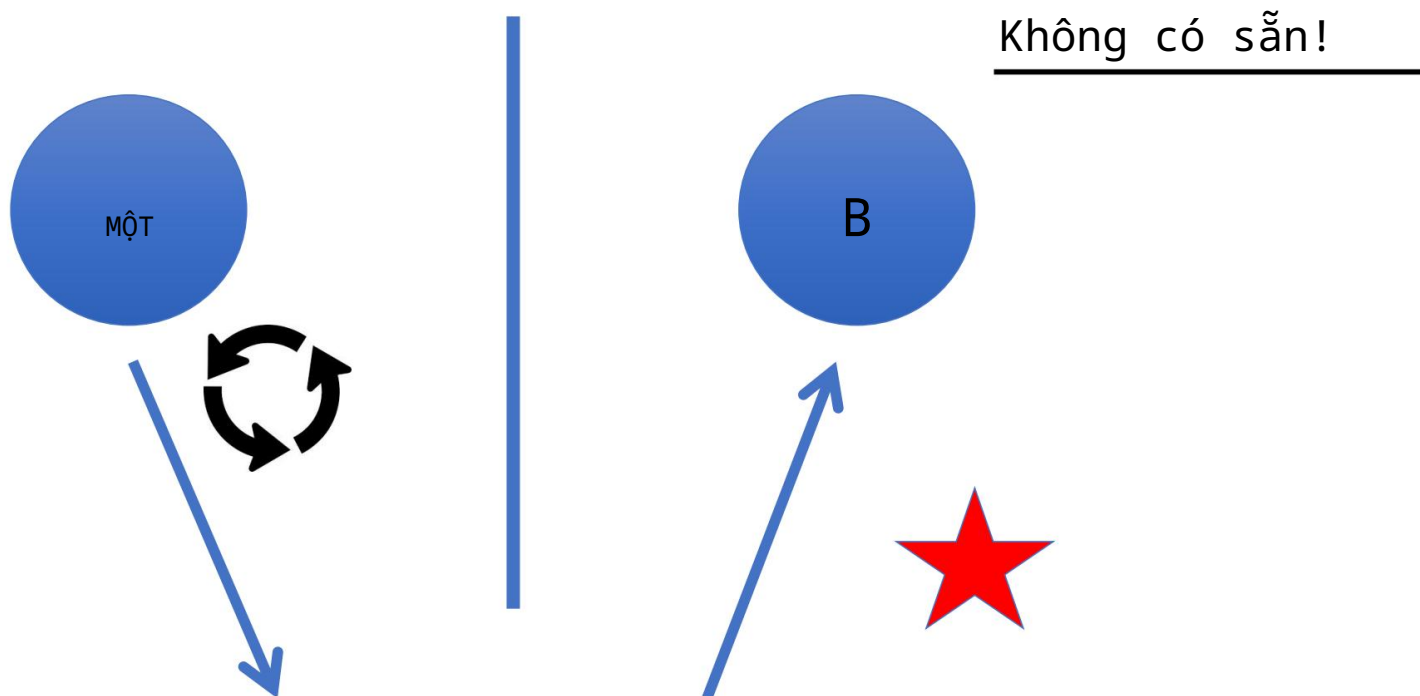
Định lý CAP: Bằng chứng

- Một bằng chứng đơn giản sử dụng hai nút:



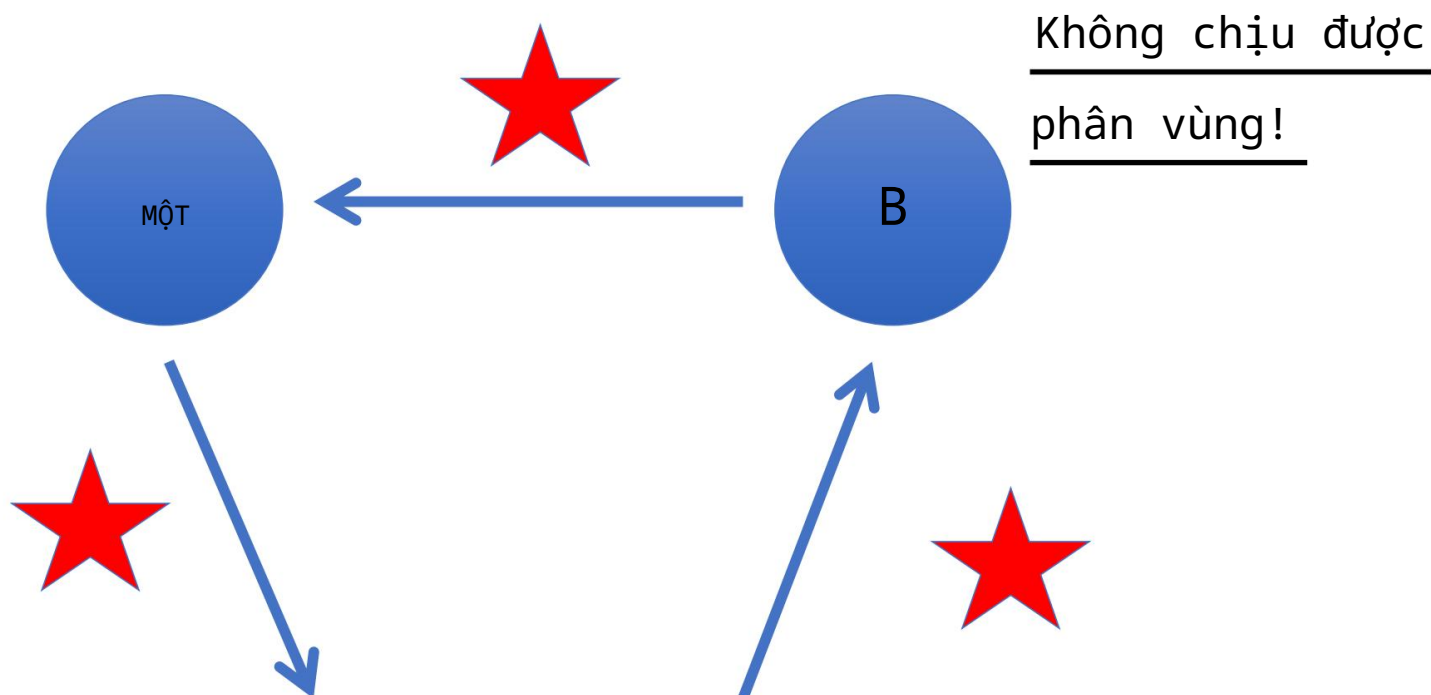
Định lý CAP: Bằng chứng

- Một bằng chứng đơn giản sử dụng hai nút:



Định lý CAP: Bằng chứng

- Một bằng chứng đơn giản sử dụng hai nút:



A được cập nhật từ B

Khả năng mở rộng của cơ sở dữ liệu quan hệ

- Cơ sở dữ liệu quan hệ được xây dựng trên nguyên tắc ACID (Độ nguyên tử, Độ đồng nhất, Độ cô lập, Độ bền)
- Nó ngụ ý rằng một cơ sở dữ liệu quan hệ phân tán thực sự phải có tính khả dụng, tính nhất quán và khả năng chịu đựng phân vùng.
- Thật không may là điều đó là không thể .

Cơ sở dữ liệu quy mô lớn

- Khi các công ty như Google và Amazon thiết kế cơ sở dữ liệu quy mô lớn, tính khả dụng 24/7 là yếu tố then chốt

- Vài phút ngừng hoạt động có nghĩa là mất doanh thu

- Khi mở rộng cơ sở dữ liệu theo chiều ngang lên tới 1000 máy móc, khả năng xảy ra lỗi của một nút hoặc mạng tăng lên rất nhiều • Do đó, để có

được sự đảm bảo chắc chắn về

Khả năng sẵn có và khả năng chịu đựng phân vùng, họ phải hy sinh tính nhất quán “ng nghiêm ngặt” (được ngụ ý bởi định lý CAP)

Sự nhất quán của sự cân bằng

- Duy trì tính nhất quán phải cân bằng giữa tính nghiêm ngặt của tính nhất quán so với tính khả dụng/khả năng mở rộng
 - Độ đặc vừa đủ phụ thuộc vào ứng dụng của bạn

Sự nhất quán của sự cân bằng

- Duy trì tính nhất quán phải cân bằng giữa tính nghiêm ngặt của tính nhất quán so với tính khả dụng/khả năng mở rộng
 - Độ đặc vừa đủ phụ thuộc vào ứng dụng của bạn

Sự nhất quán lỏng lẻo



Sự nhất quán nghiêm ngặt



Dễ dàng thực hiện hơn và
hiệu quả hơn

Nói chung khó thực hiện và không hiệu
quả

Các thuộc tính BASE

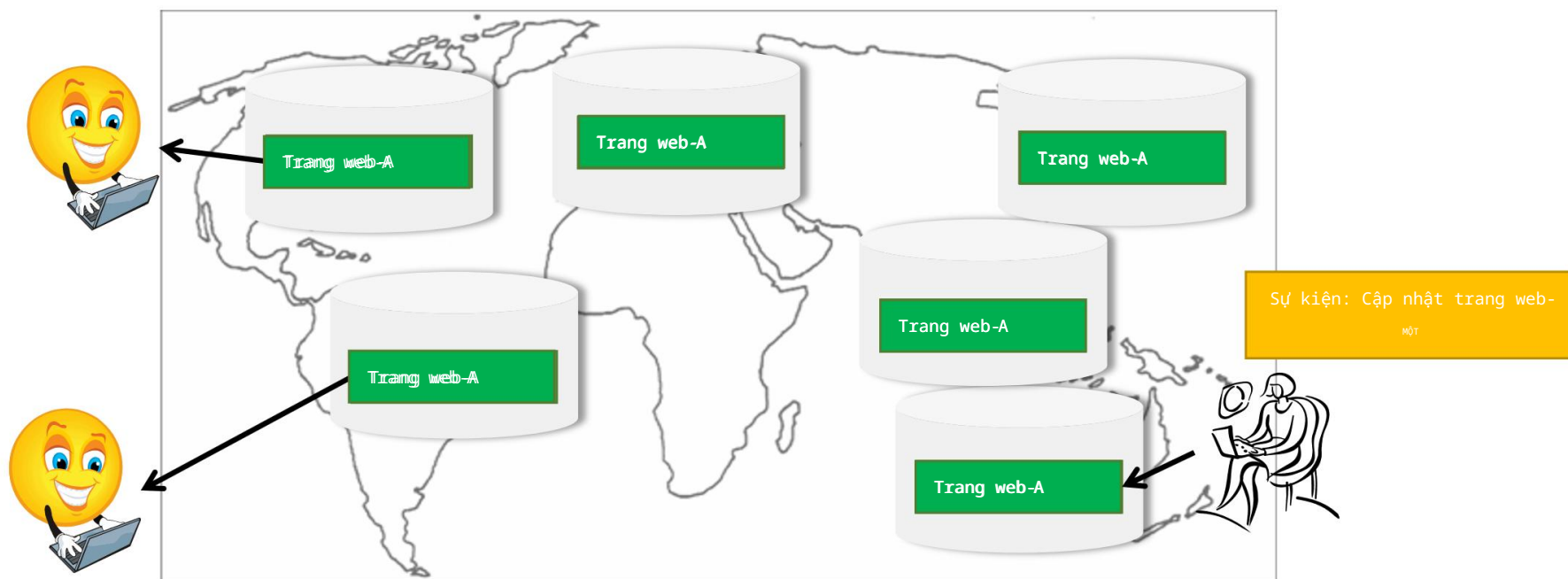
- Định lý CAP chứng minh rằng không thể đảm bảo tính nhất quán và tính khả dụng nghiêm ngặt trong khi vẫn có thể chấp nhận các phân vùng mạng
- Điều này dẫn đến cơ sở dữ liệu với ACID được nổi lên đảm bảo
- Đặc biệt, các cơ sở dữ liệu như vậy áp dụng các thuộc tính BASE:
 - Về cơ bản là có sẵn: hệ thống đảm bảo tính khả dụng
 - Trạng thái mềm: trạng thái của hệ thống có thể thay đổi theo thời gian
 - Sự nhất quán cuối cùng: hệ thống cuối cùng sẽ trở thành nhất quán

Sự nhất quán cuối cùng

- Một cơ sở dữ liệu được gọi là Nhất quán Cuối cùng nếu:
 - Tất cả các bản sao sẽ dần trở nên nhất quán khi không có bản cập nhật mới

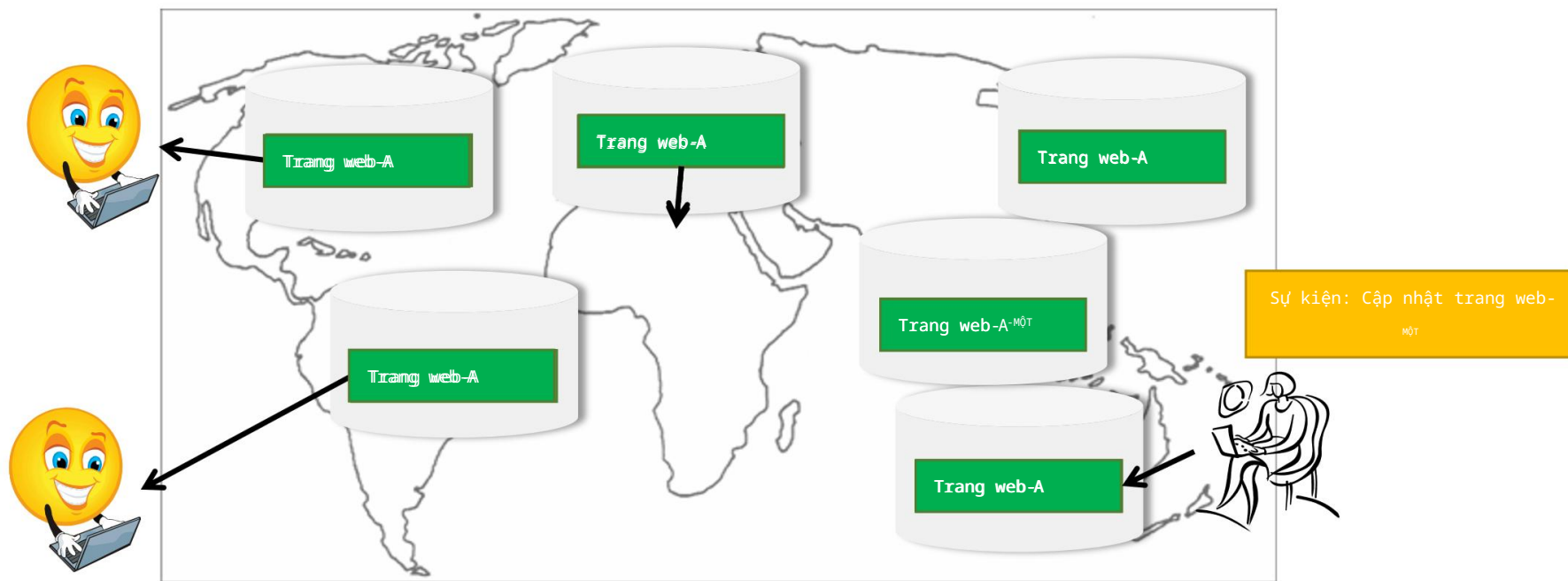
Sự nhất quán cuối cùng

- Một cơ sở dữ liệu được gọi là Nhất quán Cuối cùng nếu:
 - Tất cả các bản sao sẽ dần trở nên nhất quán khi không có bản cập nhật mới



Tính nhất quán đọc sau khi ghi (ví dụ: Amazon S3)

- Nhưng, điều gì sẽ xảy ra nếu khách hàng truy cập dữ liệu từ các nguồn khác nhau?
bản sao?



Có thể áp dụng các giao thức như Đọc lệnh ghi của riêng bạn (RYOW)!

Tài liệu tham khảo

- Gilbert, Seth và Nancy Lynch. "Quan điểm về CAP Định lý." Máy tính 45.2 (2012): 30-36.
- Kleppmann, Martin. "Một lời phê bình về Định lý CAP." arXiv bản in trước arXiv:1509.05393 (2015).
- Brewer, Eric. "CAP mười hai năm sau: "Các quy tắc" đã thay đổi như thế nào." Máy tính 45.2 (2012): 23-29.
- Chandra, Deka Ganesh. "Phân tích BASE của NoSQL cơ sở dữ liệu." Hệ thống máy tính thế hệ tương lai 52 (2015): 13-21.



VIỆN CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG
SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

Cảm ơn sự
chú ý
của bạn!!!



soict.hust.edu.vn/



fb.com/groups/soict

