

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

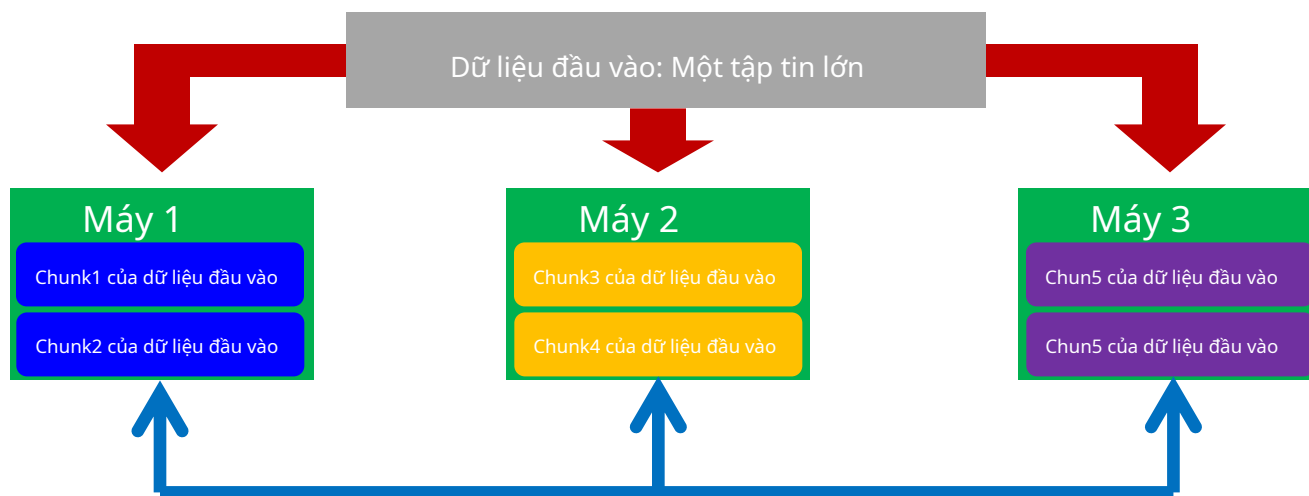
## định lý CAP

# Mở rộng cơ sở dữ liệu truyền thống

- RDBMS truyền thống có thể được mở rộng quy mô:
  - Theo chiều dọc (hoặc lên)
    - Có thể đạt được bằng cách nâng cấp phần cứng (ví dụ: CPU nhanh hơn, nhiều bộ nhớ hơn hoặc ổ đĩa lớn hơn)
    - Bị giới hạn bởi số lượng CPU, RAM và ổ đĩa có thể được cấu hình trên một máy
  - Theo chiều ngang (hoặc ra ngoài)
    - Có thể đạt được bằng cách bổ sung thêm máy móc
    - Yêu cầu phân chia cơ sở dữ liệu và có thể sao chép
    - Bị giới hạn bởi tỷ lệ Đọc-Ghi và chi phí liên lạc

# Phân mảnh dữ liệu

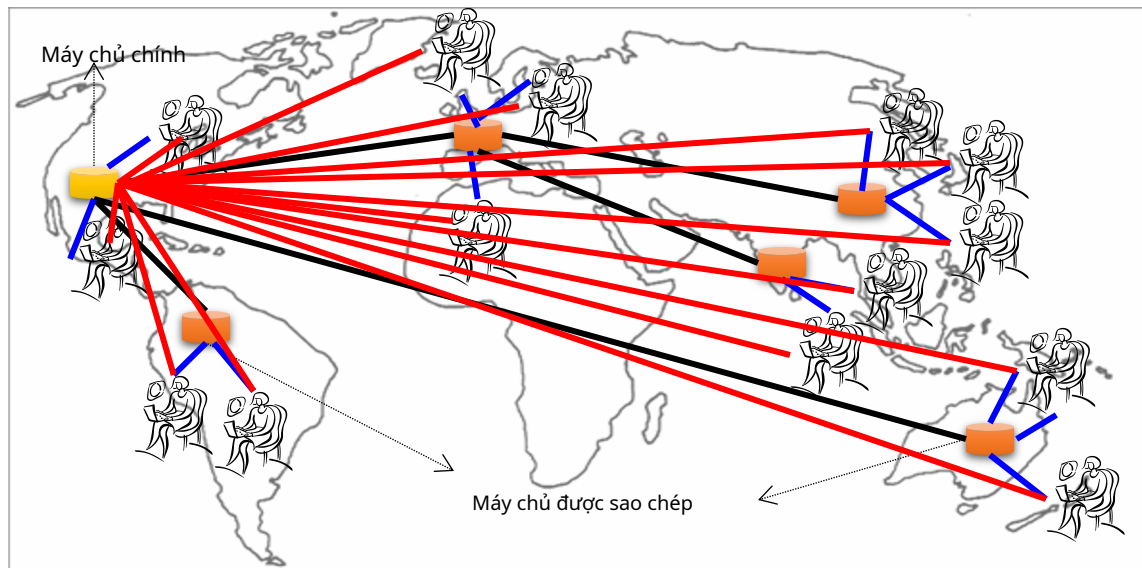
- Dữ liệu thường được phân chia (hoặc phân đoạn) để cho phép truy cập đồng thời/song song
- Liệu nó có mở rộng quy mô để xử lý truy vấn phức tạp không?



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

# Sao chép dữ liệu

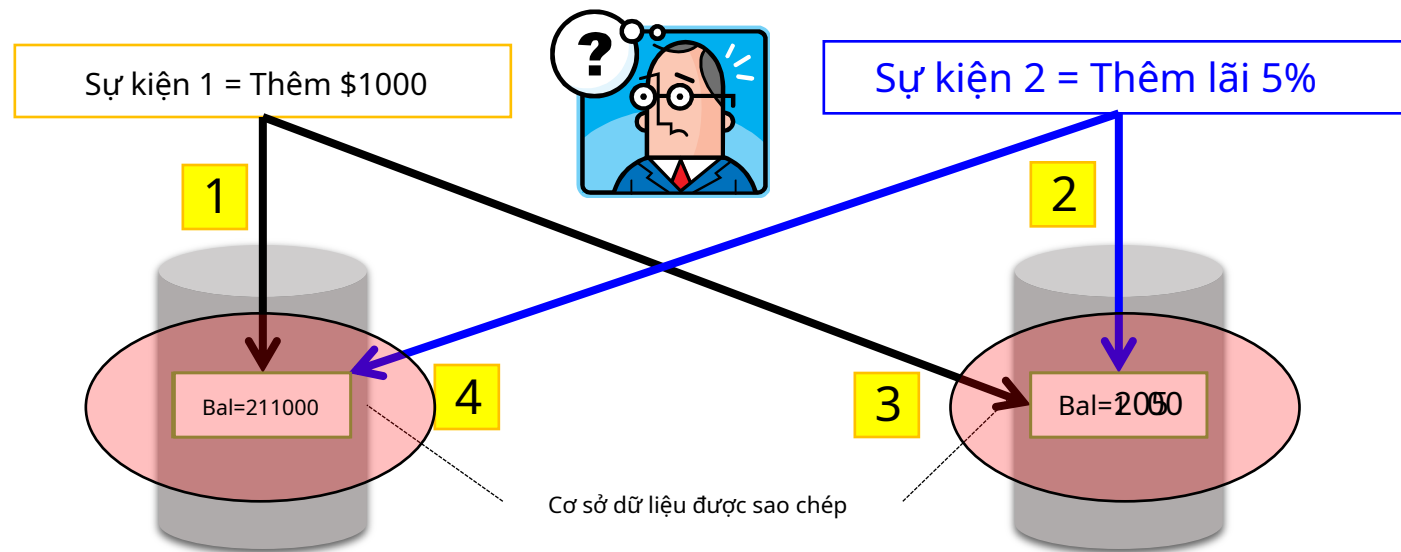
- Sao chép dữ liệu trên các máy chủ giúp:
  - Tránh tắc nghẽn hiệu suất
  - Tránh điểm lỗi duy nhất
  - Và do đó, nâng cao khả năng mở rộng và tính sẵn sàng



# Tuy nhiên, tính nhất quán trở thành một thách thức

- Một ví dụ:

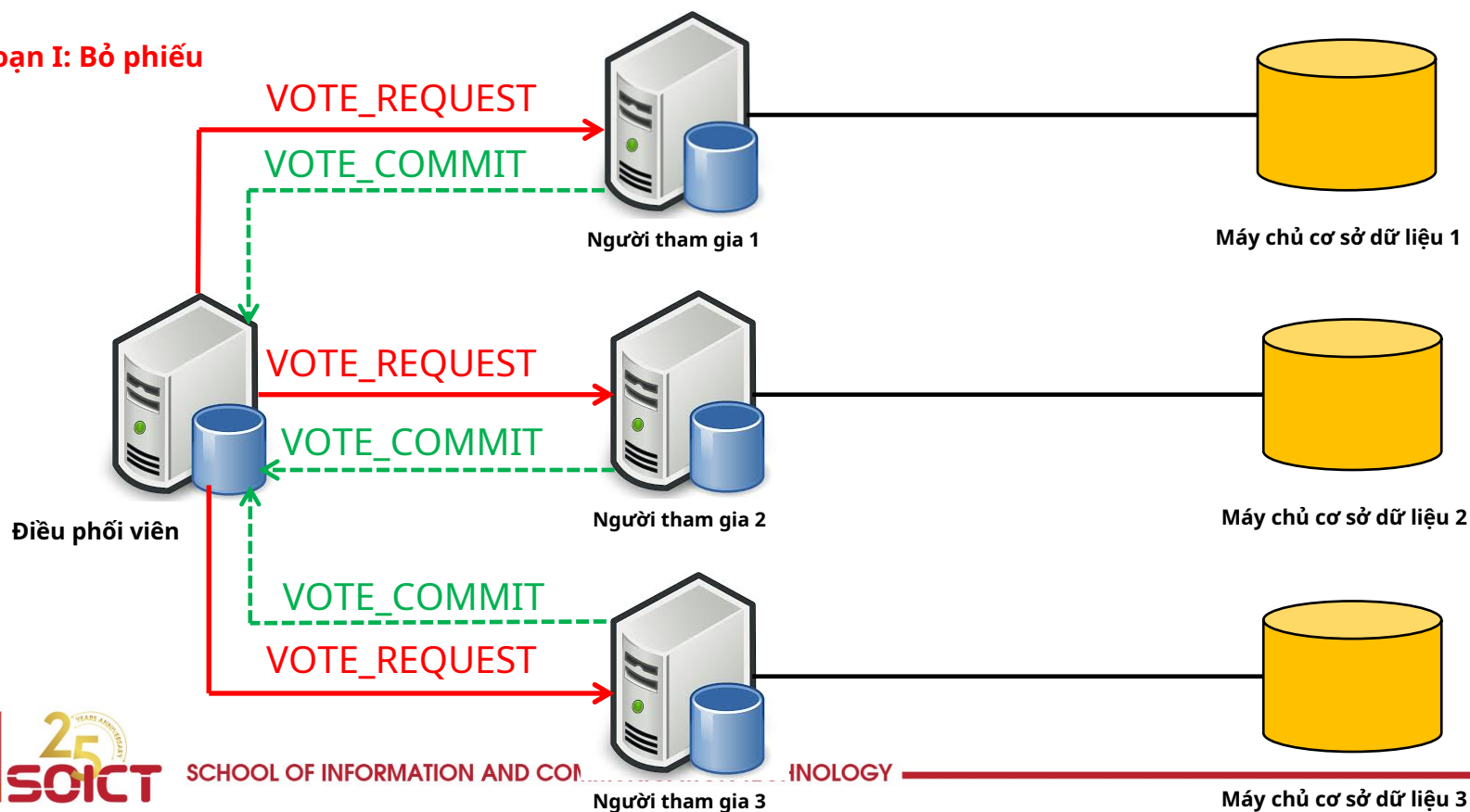
- Trong một ứng dụng thương mại điện tử, cơ sở dữ liệu ngân hàng được sao chép trên hai máy chủ
- Duy trì tính nhất quán của dữ liệu được sao chép là một thách thức



# Giao thức cam kết hai pha

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

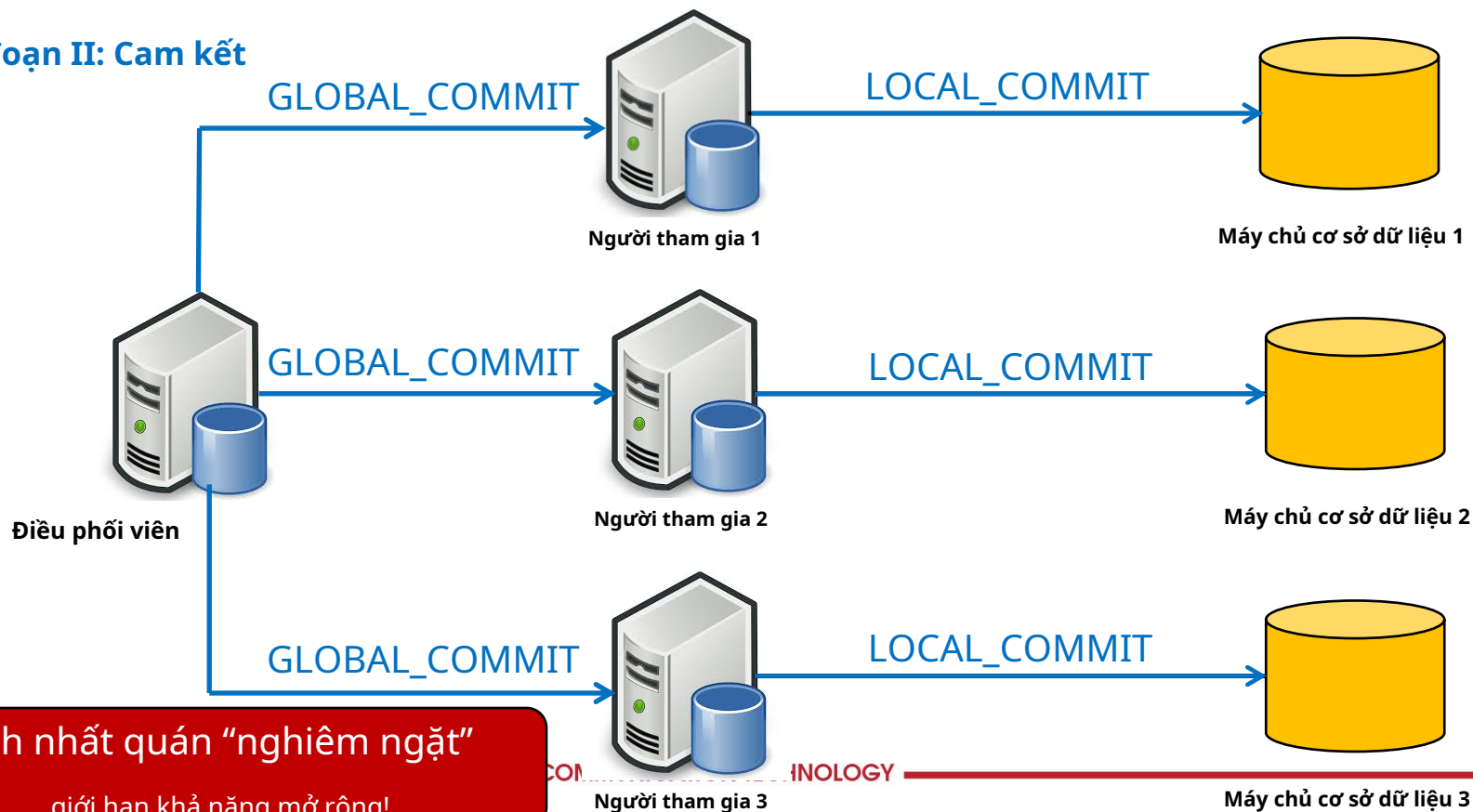
## Giai đoạn I: Bỏ phiếu



# Giao thức cam kết hai pha

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

## Giai đoạn II: Cam kết



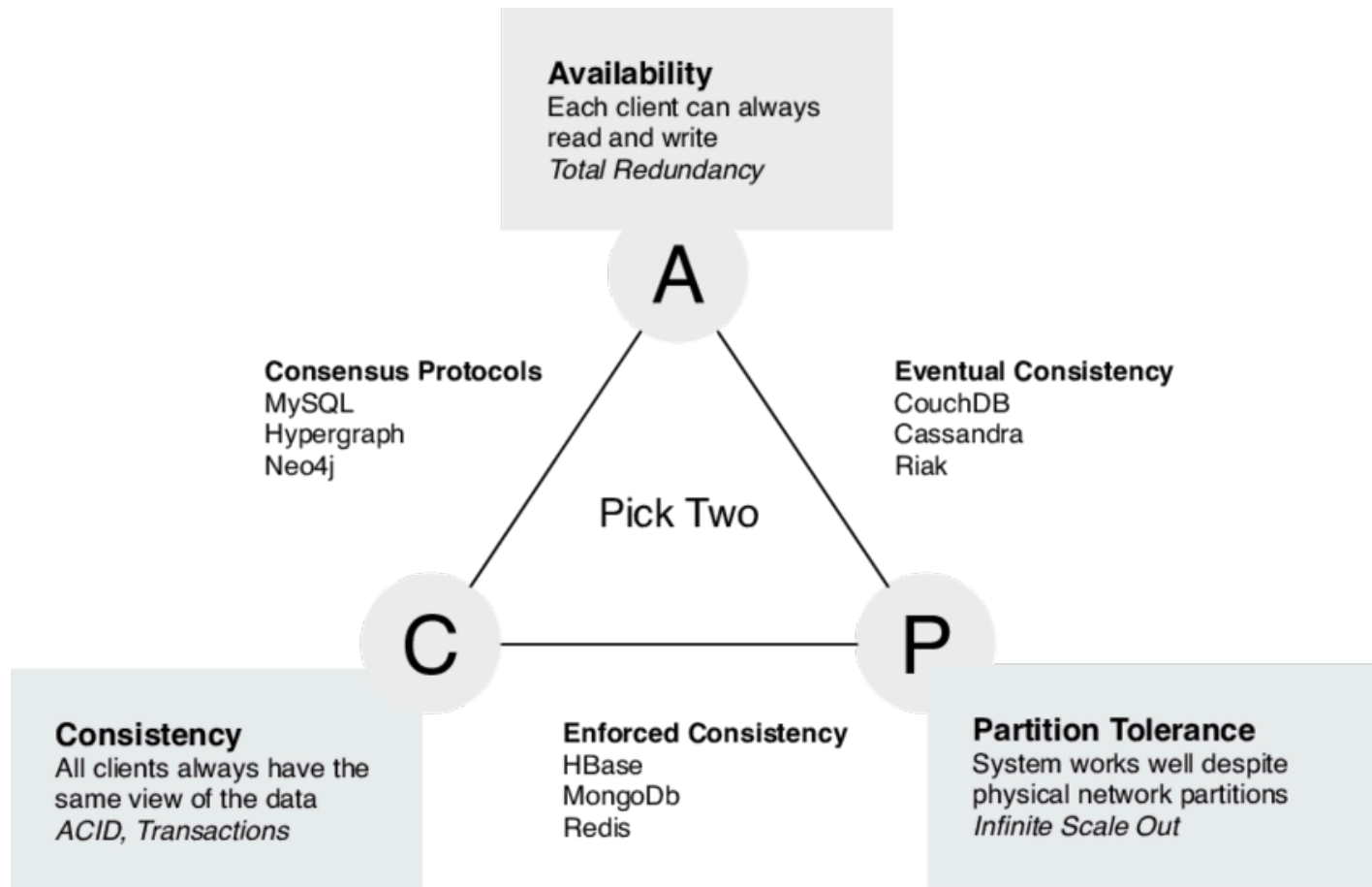


# Định lý CAP

- Những hạn chế của cơ sở dữ liệu phân tán có thể được mô tả trong định lý CAP
  - Tính nhất quán: mọi nút luôn nhìn thấy cùng một dữ liệu ở bất kỳ trường hợp cụ thể nào (nghĩa là tính nhất quán nghiêm ngặt)
  - Tính sẵn sàng: hệ thống tiếp tục hoạt động ngay cả khi các nút trong cụm gặp sự cố hoặc một số bộ phận phần cứng hoặc phần mềm không hoạt động do nâng cấp
  - Dung sai phân vùng: hệ thống tiếp tục hoạt động khi có phân vùng mạng

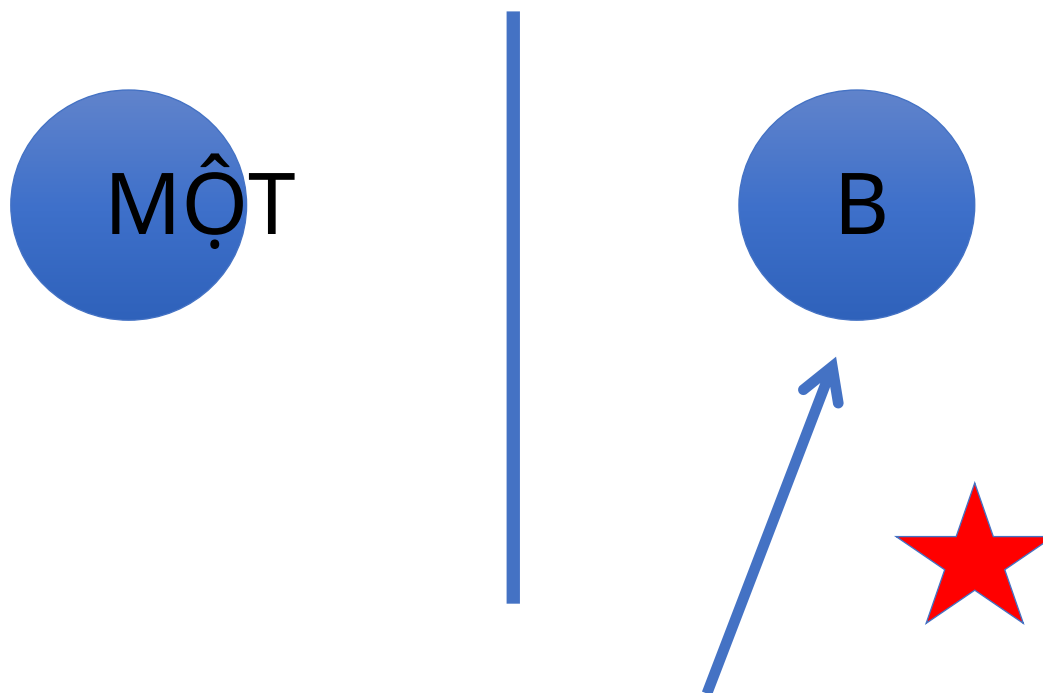
Định lý CAP: bất kỳ cơ sở dữ liệu phân tán nào có dữ liệu được chia sẻ đều có thể có nhều nhất là hai trong ba thuộc tính mong muốn C, A hoặc P. **Đây là những sự đánh đổi liên quan trong hệ thống phân tán của Eric Brewer trong PODC 2000.**

# Định lý CAP



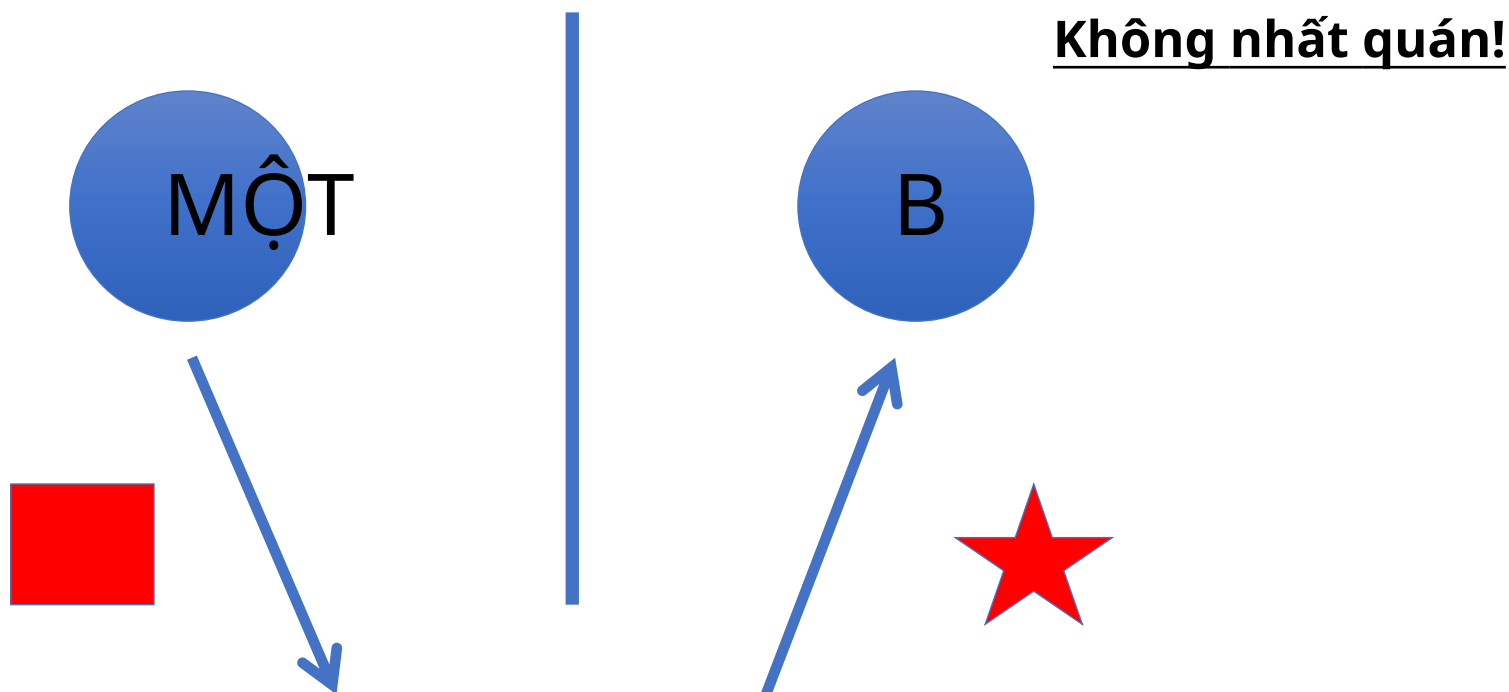
# Định lý CAP: Chứng minh

- Chứng minh đơn giản sử dụng hai nút:



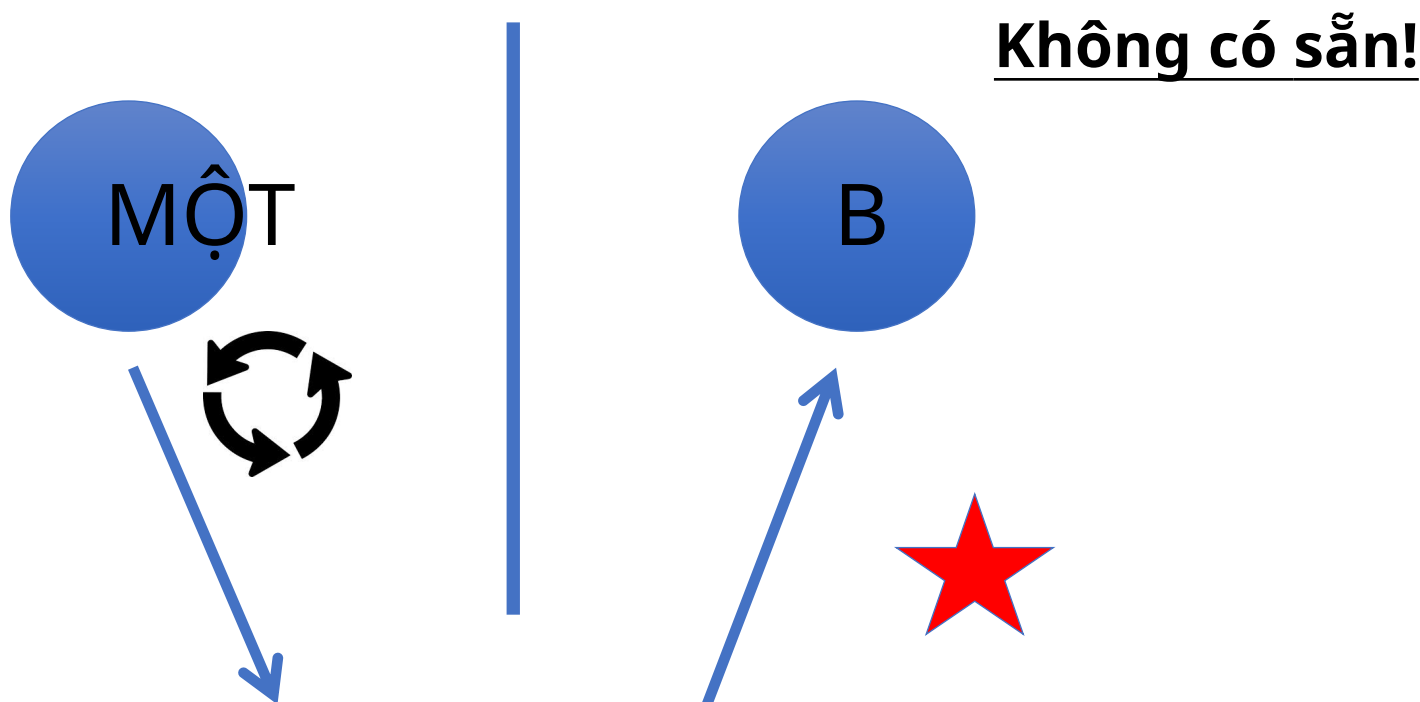
# Định lý CAP: Chứng minh

- Chứng minh đơn giản sử dụng hai nút:



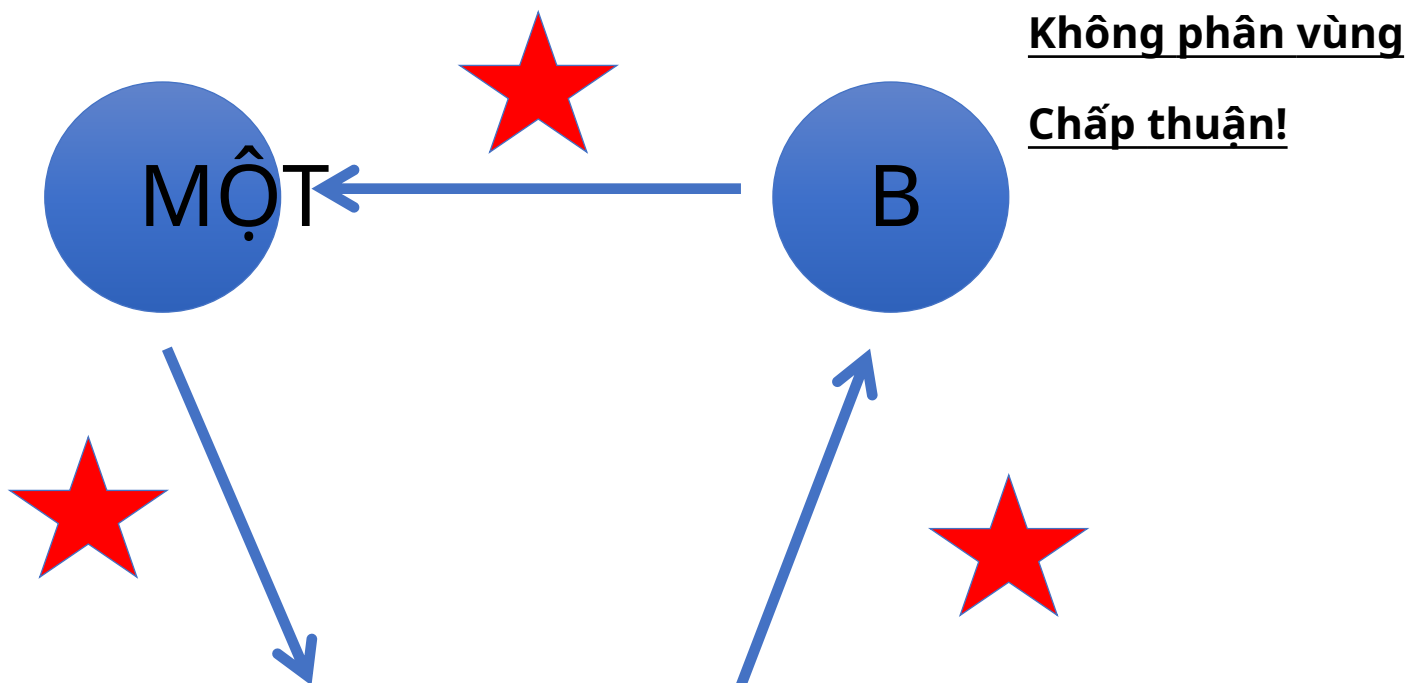
# Định lý CAP: Chứng minh

- Chứng minh đơn giản sử dụng hai nút:



# Định lý CAP: Chứng minh

- Chứng minh đơn giản sử dụng hai nút:



# Khả năng mở rộng của cơ sở dữ liệu quan hệ

- Cơ sở dữ liệu quan hệ được xây dựng trên nguyên tắc **axit**(Tính nguyên tử, tính nhất quán, sự cô lập, độ bền)
- Nó ngụ ý rằng một cơ sở dữ liệu quan hệ phân tán thực sự cần có **tính sẵn sàng, tính nhất quán và dung sai phân vùng**.
- Thật không may là **không thể nào...**

# 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 sẵn sàng 24/7 là chìa khóa
  - Một vài phút ngừng hoạt động đồng nghĩa với việc mất doanh thu
- Khi mở rộng quy mô cơ sở dữ liệu theo chiều ngang lên hàng nghìn máy, khả năng xảy ra lỗi nút hoặc mạng sẽ tăng lên rất nhiều
- Do đó, để có được sự đảm bảo chắc chắn về tính sẵn sàng và dung sai phân vùng, họ đã phải hy sinh tính nhất quán “ng nghiêm ngặt” (ngụ ý theo định lý CAP)



# Tính nhất quán trong giao dịch

- Duy trì tính nhất quán cần cân bằng giữa tính chặt chẽ của tính nhất quán và tính sẵn có/khả năng mở rộng
  - Tính nhất quán đủ tốt tùy thuộc vào ứng dụng của bạn

# Tính nhất quán trong giao dịch

- Duy trì tính nhất quán cần cân bằng giữa tính chặt chẽ của tính nhất quán và tính sẵn có/khả năng mở rộng
  - Tính nhất quán đủ tốt tùy thuộc vào ứng dụng của bạn

Tính nhất quán lỏng lẻo



Tính nhất quán chặt chẽ

Dễ thực hiện hơn,  
và có hiệu quả

Nói chung khó thực hiện  
và không hiệu quả

# Thuộc tính CƠ SỞ

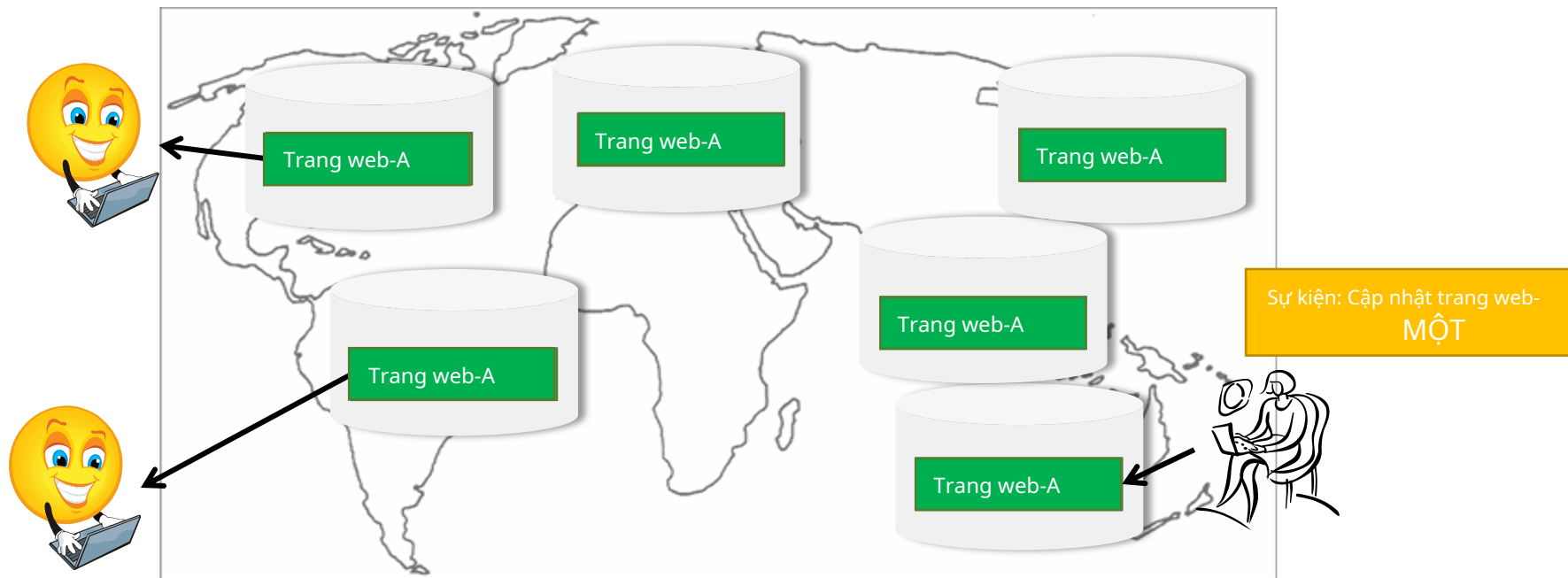
- Định lý CAP chứng minh rằng không thể đảm bảo tính nhất quán và tính sẵn sàng nghiêm ngặt trong khi vẫn có thể chịu đựng được các phân vùng mạng
- Điều này dẫn đến cơ sở dữ liệu có bảo đảm ACID thoải mái
- Đặc biệt, các cơ sở dữ liệu như vậy áp dụng các thuộc tính BASE:
  - Về cơ bản có sẵn: hệ thống đảm bảo tính sẵn sàng
  - Soft-State: trạng thái của hệ thống có thể thay đổi theo thời gian
  - Tính nhất quán cuối cùng: hệ thống cuối cùng sẽ trở nên nhất quán

# Tính 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 trong trường hợp không có bản cập nhật mới

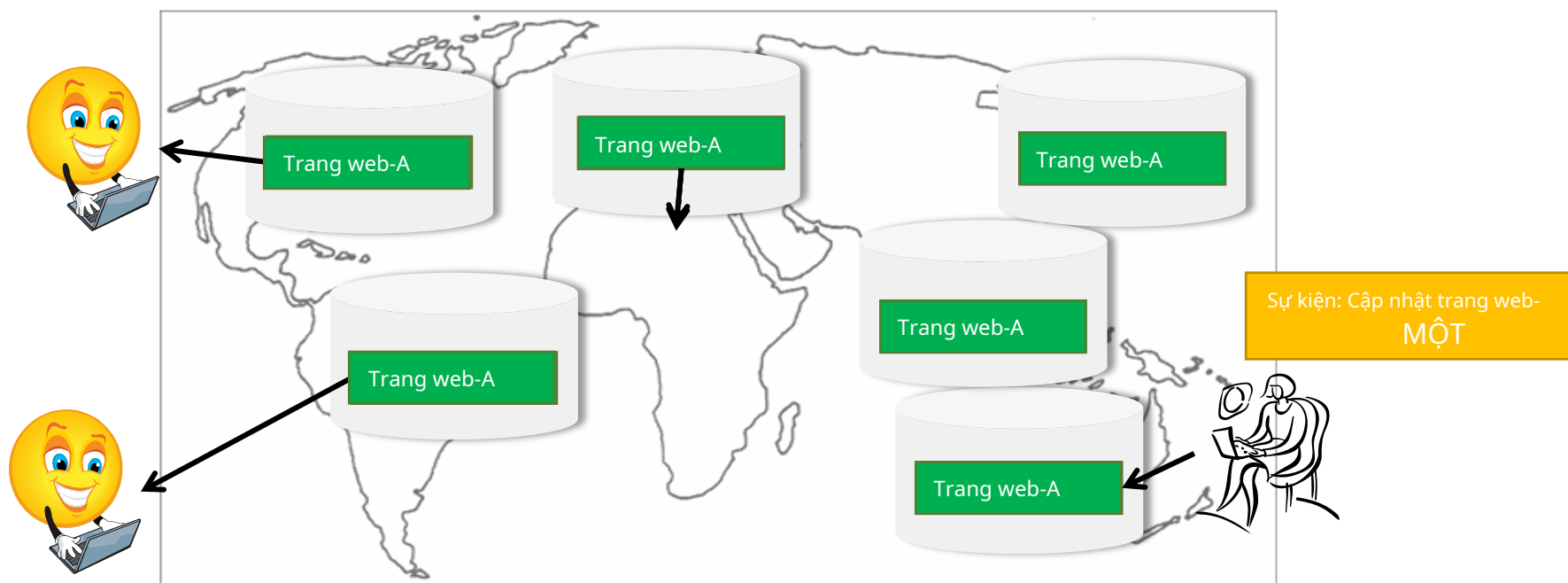
# Tính nhất quán cuối cùng

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



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

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



Các giao thức như Đọc bài viết của riêng bạn (RYOW) có thể được áp dụng!

# Người giới thiệu

- Gilbert, Seth và Nancy Lynch. "Quan điểm về Định lý CAP." *Máy tính* 45,2 (2012): 30-36.
- Kleppmann, Martin. "Một phê phán của Định lý CAP." *bản in trước arXiv arXiv:1509.05393* (2015).
- Nhà sản xuất bia, Eric. "CAP mười hai năm sau:" Quy tắc "đã thay đổi như thế nào." *Máy tính* 45,2 (2012): 23-29.
- Chandra, Deka Ganesh. "Phân tích BASE của cơ sở dữ liệu NoSQL." *Hệ thống máy tính thế hệ tương lai* 52 (2015): 13- 21.



25 YEARS ANNIVERSARY  
**SOICT**

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

Cảm ơn  
cho bạn  
chú ý!!!



[soict.hust.edu.vn/](http://soict.hust.edu.vn/)



[fb.com/groups/soict](https://fb.com/groups/soict)

