

BÀI TẬP TRÊN LỚP

MÔN HỌC: HỆ PHÂN TÁN

CHƯƠNG 7: SẠO LƯU VÀ THỐNG NHẤT DỮ LIỆU

HỌ TÊN SV: Mạc Quang Huy

MSSV: 20173169

MÃ LỚP: 118636

MÃ HỌC PHẦN: IT4611

Câu hỏi 1: Tại sao phải thực hiện sao lưu dữ liệu?

Trả lời:

- Dữ liệu được sao lưu ra các máy chủ bản sao để:
 - Cải thiện hiệu năng của hệ thống
 - Giảm lưu lượng mạng
 - Tăng tính chịu lỗi

Câu hỏi 2: 1. Xét một kho dữ liệu phân tán với 5 tiến trình độc lập P1, P2, P3, P4, và P5. Mỗi tiến trình chỉ tác động lên được bản sao cục bộ riêng của mình. Các bản sao cục bộ kết nối thành kho dữ liệu phân tán. Xét các tiến trình chỉ tương tác (ghi, đọc) lên thành phần dữ liệu x ở bản sao cục bộ riêng của mình. Hoạt động của mô hình ở các thời điểm t tương ứng, các thao tác được thực hiện như sau: t1: P1 ghi giá trị a t2: P3 đọc giá trị a

t3: P2 ghi giá trị b và P3 ghi giá trị c

t4: P5 đọc được giá trị b

t5: P4 và P5 đều đọc được giá trị a

t6: P4 đọc được giá trị b

t7: P4 và P5 đọc được giá trị c

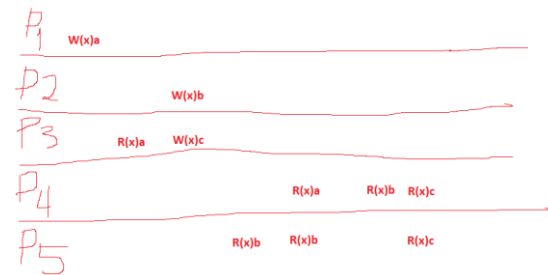
(biết rằng $t_i < t_{i+1}$ với $i=(0..6)$)

Câu hỏi:

a) Mô hình trên có thoả mãn thống nhất nhân quả không? Giải thích.

b) Mô hình trên có thoả mãn thống nhất tuần tự không? Giải thích.

Trả lời:



a. Thống nhất nhân quả: Các thao tác ghi có ràng buộc nhân quả cần được các tiến trình nhìn thấy theo cùng một thứ tự

- P1, P2 có 2 thao tác ghi có tính tương tranh -> Ở các tiến trình khác đọc sao cũng được
- P1, P3 có ràng buộc nhân quả -> Nhìn cùng 1 thứ tự ở các tiến trình khác => OK

⇒ Có tính nhân quả

b. Thống nhất tuần tự: Tất cả các tiến trình đều nhìn thấy một thứ tự của các thao tác ghi

- Thứ tự ghi: $W(x)a \rightarrow W(x)b, W(x)c$

- Không TM thứ tự đọc thỏa mãn tính tuần tự

⇒ Không có tính tuần tự

Câu hỏi 3: Conit là gì? Nếu đặt kích thước Conit quá lớn thì sẽ gây ra vấn đề gì? Tương tự với kích thước Conit quá nhỏ?

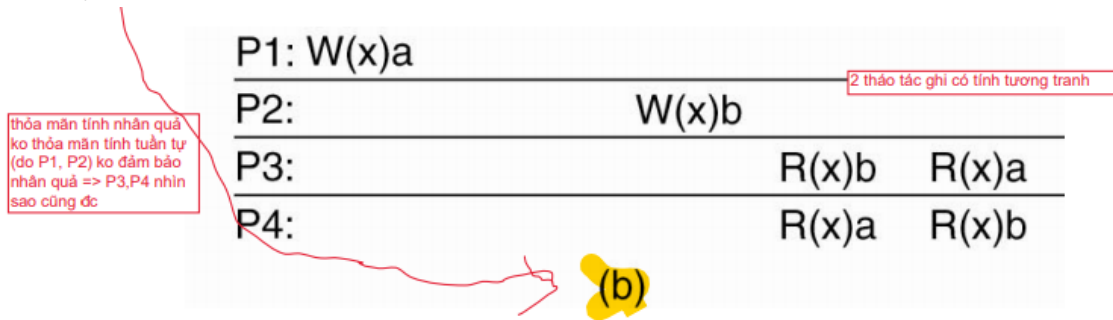
Trả lời:

- Conit (consistency unit): Đơn vị thống nhất (để giám sát đơn vị dữ liệu của các cặp replica)
- Conit quá lớn: Các bản sao sẽ sớm bị rơi vào trạng thái không thống nhất
- Conit quá nhỏ: Số lượng conit nhiều \rightarrow quản lý phức tạp

Câu hỏi 4: Tại sao *thống nhất nhân quả* có tính thống nhất yếu hơn *thống nhất tuần tự*? Cho ví dụ để làm rõ điều này.

Trả lời:

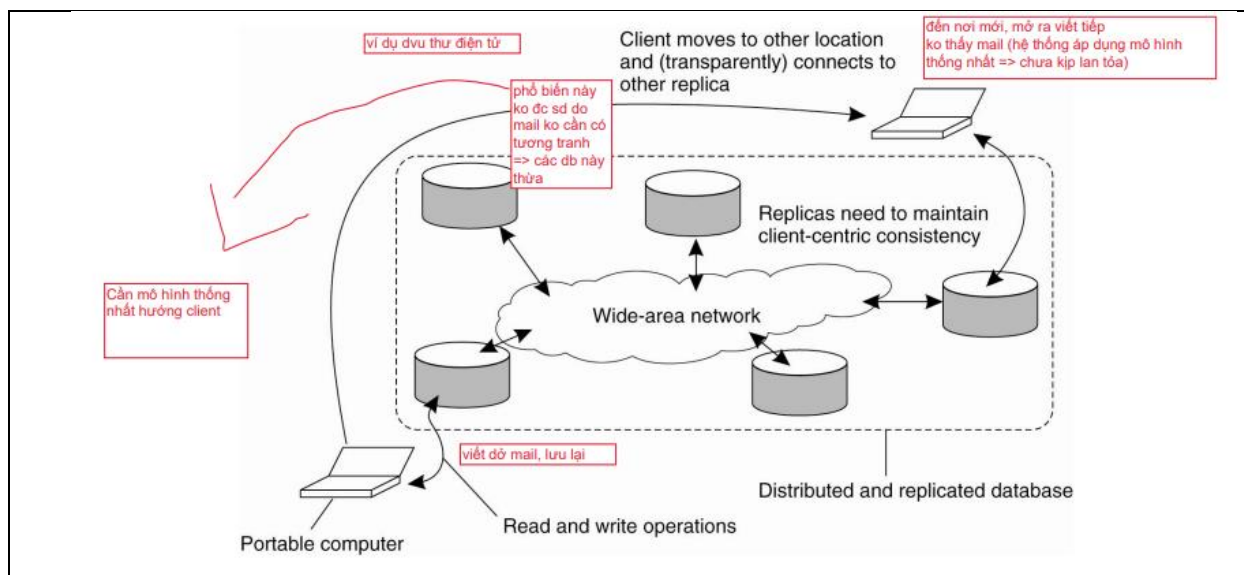
- Vì quan hệ thống nhất nhân quả chỉ đảm bảo $W(x)$ có ràng buộc nhân quả phải được nhìn cùng 1 thứ tự còn quan hệ thống nhất tuần tự đảm bảo *toàn bộ* $W(x)$ được nhìn cùng 1 thứ tự
- Ví dụ:



Câu hỏi 5: Vấn đề của mô hình Eventual Consistency là gì? Từ đó đưa ra định nghĩa mô hình thống nhất hướng client.

Trả lời:

- Ví dụ vấn đề viết mail, tài liệu,...



Câu hỏi 6: Một ngân hàng quyết định sử dụng dịch vụ CDN (Content Delivery Network) của một công ty mới khởi nghiệp cung cấp.

- Với bước đặt máy chủ, công ty chọn thuật toán chọn đặt các máy chủ bản sao (replica) dựa trên khoảng cách với các chi nhánh ngân hàng. Hãy đề xuất thuật toán chọn đặt k replica với N vị trí có thể đặt máy chủ. Biết rằng đây là thuật toán dựa trên khoảng cách và công ty biết trước các vị trí các chi nhánh ngân hàng.
- Với thuật toán để quản lý nội dung dữ liệu ở các replica, công ty quyết định chọn thuật toán dựa trên bản sao kích hoạt bởi server (server-initiated replicas). Hãy mô tả cơ chế đó với việc xem xét một đơn vị dữ liệu X là thông tin tài khoản một người dùng cùng với 2 ngưỡng là $del(X)$ và $rep(X)$.
- Liên quan đến giao thức đảm bảo thống nhất, công ty quyết định chọn giao thức ghi trên các bản sao (replicated write), tuy nhiên công ty băn khoăn giữa giao thức sao lưu tích cực và giao thức sao lưu dựa trên túc số. Bạn hãy giúp công ty lựa chọn giao thức phù hợp bằng việc so sánh 2 giao thức trên với việc chỉ ra ưu nhược điểm của chúng.

Trả lời:

- Do đề bài chưa cho biết khoảng cách giữa client và các bản sao nên giải pháp được đề xuất là : Chia thành các cell - autonomous systems

Sau đó Chọn AS lớn nhất (đặt server ở vị trí có nhiều link nhất)

Tiếp đó chúng ta tiếp tục với các AS nhỏ hơn

- .

Sao lưu tích cực	Sao lưu túc số
<ul style="list-style-type: none"> Một tiến trình chịu trách nhiệm phổ biến thao tác cập nhật đến tất cả các bản sao Cần có một cơ chế trật tự toàn cục (total ordered mechanism) 	<ul style="list-style-type: none"> Sao lưu theo vote-Giải thuật quorum (túc số) <ul style="list-style-type: none"> Để có thống nhất mạnh=> cần cập nhật tất cả các bản sao Sau khi cập nhật với chi phí tốn kém=> không phải tất cả

	<ul style="list-style-type: none"> - Có thể sử dụng thời gian logic (Lamport)=> không có giãn được - Có thể sử dụng sequencer đảm bảo trật tự toàn cục 	các bản sao đều được đọc=> chi phí vô ích -
--	---	---

Câu hỏi 7: Liên quan đến các mô hình thống nhất hướng dữ liệu và các mô hình thống nhất hướng người dùng:

- Giải thích vấn đề ý tưởng của 2 loại mô hình thống nhất hướng dữ liệu trên.
- Một công ty startup mới mở chuyên triển khai thương mại hóa dịch vụ CDN (Content Delivery Network) cho 2 loại hình dịch vụ là thư điện tử và WWW. Để đảm bảo thống nhất dữ liệu cho 2 loại dịch vụ đó thì tầng middleware sẽ áp dụng mô hình thống nhất dữ liệu nào (ở câu a) cho mỗi loại dịch vụ trên? Giải thích.
- Công ty đó triển khai 3000 server bản sao vật lý và chọn hình thức sao lưu dữ liệu dựa trên túc số (quorum) với $N_w = 1600$ và $N_r = 1100$. Vậy hệ thống đó sẽ tránh được xung đột đọc-ghi và xung đột ghi-ghi hay không? Giải thích.

Trả lời:

- Mô hình hướng dữ liệu thì Cam kết giữa các tiến trình và kho dữ liệu còn mô hình hướng client thì cung ứng đảm bảo thống nhất cho các truy cập của một client đơn vào kho dữ liệu
- Đối với dịch vụ thư điện tử thì dùng mô hình hướng dữ liệu. Còn WWW thì sử dụng mô hình hướng client
- Hệ thống này tránh được xung đột ghi-ghi do $N_w (1600) > N/2 (3000/2=1500)$
Hệ thống này không tránh được xung đột đọc-ghi do $N_w + N_r = 2700 < N = 3000$