

BÀI TẬP TRÊN LỚP
MÔN HỌC: HỆ PHÂN TÁN
CHƯƠNG 7: SAO LƯU VÀ THỐNG NHẤT DỮ LIỆU

HỌ TÊN SV:

MSSV:

MÃ LỚP:

MÃ HỌC PHẦN:

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

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.

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?

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.

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

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.

a) 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.

b) 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 $rep(X)$.

c) 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.

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

- a. Giải thích vấn đề ý tưởng của 2 loại mô hình thống nhất hướng dữ liệu trên.
- b. 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.
- c. 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.