**BÀI TẬP LÝ THUYẾT**

**MÔN : HỆ PHÂN TÁN**

**CHƯƠNG 5: NHÂN BẢN VÀ NHẤT QUÁN DỮ LIỆU TRONG HỆ PHÂN TÁN**

**Họ và tên: Lê Thị Yên**

**MSSV: 20183861**

**Câu hỏi 1:** Tại sao phải thực hiện nhân bản dữ liệu?

* Cải thiện hiệu năng hệ thống:

Tăng quy mô của hệ thống cả về số lượng lẫn phạm vi địa lý.

* + Xây dựng hệ thống theo mô hình tập trung thì chỉ phù hợp với các dự án vừa và nhỏ, tức là số lượng client đủ nhỏ. Khi số lượng client tăng lên đột biến và lượng request tới server tăng lên khiến cho server bị quá tải
  + Nếu số lượng máy tăng => tải trên mỗi máy trong hệ phân tán cũng ko tăng lên nhiều hơn, phạm vi địa lý tăng => có thể sử dụng các bản nhân bản gần khu vực địa lý đó
* Giảm tải mạng
* Cung cấp khả năng chịu lỗi tốt hơn

Trong quá trình đọc hoặc ghi dữ liệu, nếu dữ liệu đó bị lỗi hay vì một nguyên nhân nào đó mà không thể dùng được, ta có thể dùng ngay bản sao dữ liệu đó để hệ thống không phải dừng lại và tránh được tình trạng sử dụng các dữ liệu không chính xác.

 Tuy nhiên phải chắc chắn rằng khi một bản sao được cập nhật thì tất cả những bản còn lại cũng phải được cập nhật theo để đảm bảo tính nhất quán của dữ liệu. Điều này ảnh hưởng rất lớn đến hiệu năng của hệ thống…

**Câu hỏi 2:** 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 ti < ti+1 với i=(0..6)

Câu hỏi:

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

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

**Trả lời:**

**Mô hình:**

**t1 t2 t3 t4 t5 t6 t7**

**P1 W1(x)a**

**P2 W2(x)b**

**P3 R3(x)a W3(x)c**

**P4 R4(x)a R4(x)b R4(x)c**

**P5 R5(x)b R5(x)a R5(x)c**

1. Mô hình trên thỏa mãn tính nhất quán nhân quả vì:

+ Trước khi ghi giá trị W3(x)c thì P3 đã đọc được giá trị được thay đổi do thực hiện W1(x)a tại thời điểm t1=> R3(x)a ở t2 và W3(x)c ở t3 có quan hệ nhân quả với nhau

+ Một thao tác đọc có quan hệ nhân quả với thao tác ghi cung cấp dữ liệu cho nó

+ Nếu sự kiện b được gây ra hoặc bị tác động bởi một sự kiện a xảy ra sớm hơn thì tính nhân quả đòi hỏi mọi thực thể khác phải "nhìn" thấy a trước rồi mới thấy b sau:

Ghi giá trị W1(x)a tại t1 thì mọi tiến trình còn lại trong kho dữ liệu phân tán đều nhận được giá trị a trước ghi đọc được giá trị thay đổi khác. Ví dụ: P4 đọc được giá trị a rồi mới đọc được giá trị b do P2 thay đổi ở thời điểm t2, sau đó mới đọc được giá trị c được thay đổi bởi P3 cũng ở thời điểm t2 (Các thao tác ghi đồng thời có thể được nhận biết theo thứ tự khác nhau trên những máy khác nhau)

1. Mô hình trên không thỏa mãn nhất quán trình tự vì:

Khi các tiến trình chạy đồng thời trên các máy khác nhau thì cho phép sự đan xen của các thao tác nhưng tất cả các tiến trình đều phải nhận biết được sự đan xen của các thao tác đó là như nhau.

Trong mô hình trên thứ tự đọc được phải theo thứ tự chung thì mới được coi là nhất quán trình tự, nhưng ở P5 khi đọc giá trị tại các thời điểm t4, t5, t7 không giống thứ tự đọc giá trị ở P4 tại các thời điểm t5, t6, t7

**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?

* Conit là đơn vị dùng để đánh giá sự không nhất quán giữa 2 cơ sở dữ liệu bản sao
* Với kích thước lớn: Các bản sao sẽ sớm bị rơi vào trạng thái không nhất quán
* Với kích thước nhỏ: số lượng conit nhiều: quản lý phức tạp

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

**Giải thích:**

* Nhất quán nhân quả thì chỉ cần các sự kiện có quan hệ nhân quả đảm bảo thứ tự, còn các sự kiện khác thì không cần
* Nhất quán tuần tự thì thứ tự các thao tác cục bộ của một tiến trình không thay đổi trong thứ tự thực hiện chung trên kho dữ liệu =>Kết quả luôn luôn như nhau.

Tất cả các tiến trình đều nhìn thấy một thứ tự của các thao tác ghi

Ví dụ:

Table

Description automatically generated

Khi ghi giá trị W1(x)a tại t1, W2(x)b tại t2 thì:

+ Đối với nhất quán trình tự thì thứ tự đọc giá trị ở P3 và P4 phải cùng là đọc được b rồi a hoặc đọc được a rồi mới đọc được b

+ Đối với nhất quán nhân quả thì chỉ cần thứ tự đọc giá trị ở P3, P4 có thể khác nhau

**Câu hỏi 5:** Vấn đề của mô hình Eventual Consistency là gì? Từ đó định nghĩa mô hình nhất quán hướng client

* Vấn đề của mô hình Eventual Consistency:

 Khi client thực hiện cập nhật tại một bản sao và trong thời gian ngắn chuyển sang bản sao khác. Trường hợp client là thiết bị di động, việc thực hiện yêu cầu gặp khó khăn hơn => đảm bảo các bản sao luôn nhất quán khi mà client thay đổi về vị trí vật lý.Diagram

Description automatically generated

* Ví dụ hình trên:
  + Người dùng sử dụng máy tính kết nối mạng không dây, việc kết nối đến bản sao hoàn toàn trong suốt đối với người sử dụng.
  + Sau khi cập nhật tại một bản sao người dùng di chuyển sang vị trí khác và hoàn toàn có khả năng phần mềm ứng dụng có thể kết nối đến bản sao chưa kịp cập nhật.
  + Sau các thao tác cập nhật tất cả các bản sao đều phải giống nhau, yêu cầu này sẽ được thực hiện tốt nếu mỗi client luôn cập nhật các bản sao.
  + Việc cập nhật các bản sao ngay sau khi cập nhật bản chính có thể kéo dài thời gian thực hiện, do đó lập trình viên cần dự đoán thời gian thực hiện mỗi yêu cầu và lựa chọn phương án thích hợp.

Vấn đề trên có thể giảm bớt bằng cách sử nhất quán lấy client làm trung tâm, nó đảm bảo tính nhất quán cho một client chứ không đảm bảo nhất quán cho các client khác nhau.

* Mô hình nhất quán client: Cung ứng đảm bảo nhất quán cho các truy cập của một client đơn vào kho dữ liệu.

Chú ý: không đảm bảo nhất quán cho các truy cập cạnh tranh của các tiến trình khác.

**Câu hỏi 6:**

1. Có thể chia các chi nhánh ngân hang thành K cụm có vị trí gần nhau. Với mỗi cụm chọn 1 vị trí đặt máy chủ sao cho khoảng cách trung bình từ máy chủ bản sao đến các chi nhánh trong cụm là nhỏ nhất
2. Cơ chế bản sao kích hoạt bởi server:

Bản sao kích hoạt bởi server được sử dụng để làm tăng hiệu năng và được tạo từ yêu cầu của kho dữ liệu.Các bản sao này được xếp đặt động dựa vào yêu cầu của máy chủ khác.

* + Nếu số lượng yêu cầu người dùng gửi đến nhỏ hơn ngưỡng del(X) thì bản sao đó sẽ xóa đơn vị dữ liệu X đó đi.
  + Nếu số lượng yêu cầu của người dùng gửi đến lớn hơn ngưỡng rep(X) thì bản sao đó sẽ nhân bản dữ liệu X đó cho các bản sao gần mình để cùng phục vụ yêu cầu.
  + Nếu số lượng yêu cầu của người dùng nằm trong khoảng từ del(X) đến rep(X) thì nó sẽ di trú tài nguyên X đến bản sao gần client hơn.

c.So sánh 2 giao thức nhân bản tích cực và nhân bản dựa trên túc số:

* Nhân bản tích cực:
* Khi có một cập nhật đến một bản sao thì nó sẽ quảng bá cho tất cả các bản sao khác.
* Cần phải đảm bảo trật tự toàn cục.

🡪 Tốn kém tài nguyên.

* Nhân bản dựa trên túc số:
* Khi có một cập nhật đến một bản sao, cập nhật ngẫu nhiên Nw bản sao khác.
* Khi đọc dữ liệu sẽ đọc Nr bản sao khác.
* Để tránh mâu thuẫn đọc ghi: Nr + Nw > N
* Để tránh mâu thuẫn ghi ghi: Nw > N/2

🡪 khắc phục được tiêu tốn tài nguyên.

Công ty nên sử dụng nhân bản dựa trên túc số.

**Câu hỏi 7:**

1. **Ý tưởng:**

* Mô hình nhất quán hướng người dùng:
* Cung ứng đảm bảo thống nhất cho các truy cập của một client đơn vào kho dữ liệu.
* Giải quyết vấn đề nhiều tiến trình liên tục cập nhật tương tranh trên dữ liệu.
* Không đảm bảo thống nhất cho các truy cập cạnh tranh của các tiến trình khác.
* Mô hình nhất quán hướng dữ liệu:
* Trong kiểu mô hình này, bất kì người sử dụng nào truy cập vào kho dữ liệu cũng sẽ nhìn thấy các thao tác được sắp xếp theo mô hình.
* Kho dữ liệu có thể được đọc hay ghi bởi bất cứ tiến trình nào trong một hệ phân tán.
* Tuy nhiên dữ liệu ghi vào một bản sao cục bộ phải đảm bảo cũng được truyền tới tất cả các bản sao ở xa.
* Một mô hình thống nhất có thể được coi là một bản hợp đồng giữa một kho dữ liệu của hệ phân tán với các các tiến trình của nó. Nếu các tiến trình đồng ý với các điều khoản của hợp đồng thì kho dữ liệu sẽ hoạt động đúng như tiến trình mong muốn.

1. Để đảm bảo nhất quán dữ liệu cho 2 loại dịch vụ trên thì tầng Middleware sẽ áp dụng:

- Mô hình nhất quán dữ liệu với dịch vụ WWW vì chủ yếu là thao tác đọc, thao tác đọc-ghi hầu như không xảy ra

- Mô hình nhất quán client với dịch vụ Email vì dữ liệu phải được đảm bảo nhất quán cho các truy cập của một client đơn vào kho dữ liệu. Đảm bảo rằng khi một client thực hiện một thao tác đọc trên một bản sao rồi tiếp theo lại đọc trên một bản sao khác thì bản sao thứ hai kia ít nhất cũng phải được ghi giống với bản sao đầu tiên.

c.

- Xét xung đột đọc-ghi:

Nr + Nw = 2700 < 3000 (thỏa mãn Nr + Nw > N)

* Hệ thống tránh được xung đột đọc-ghi
* Xét xung đột ghi-ghi:

Hệ thống tránh được bởi 1600 > 3000/2 (thỏa mãn Nw > N/2)