



(12) BẢN MÔ TẢ SÁNG CHẾ THUỘC BẰNG ĐỘC QUYỀN SÁNG CHẾ

(19) Cộng hòa xã hội chủ nghĩa Việt Nam (VN)
CỤC SỞ HỮU TRÍ TUỆ

(11)



1-0053772

(51)^{2020.01} G06F 12/08

(13) B

(21) 1-2020-01147

(22) 28/02/2020

(45) 25/11/2025 452

(43) 25/08/2020 389A

(73) Tập đoàn Công nghiệp - Viễn thông Quân đội (VN)

Lô D26 khu đô thị mới Cầu Giấy, Phường Yên Hoà, Quận Cầu Giấy, Thành phố Hà Nội

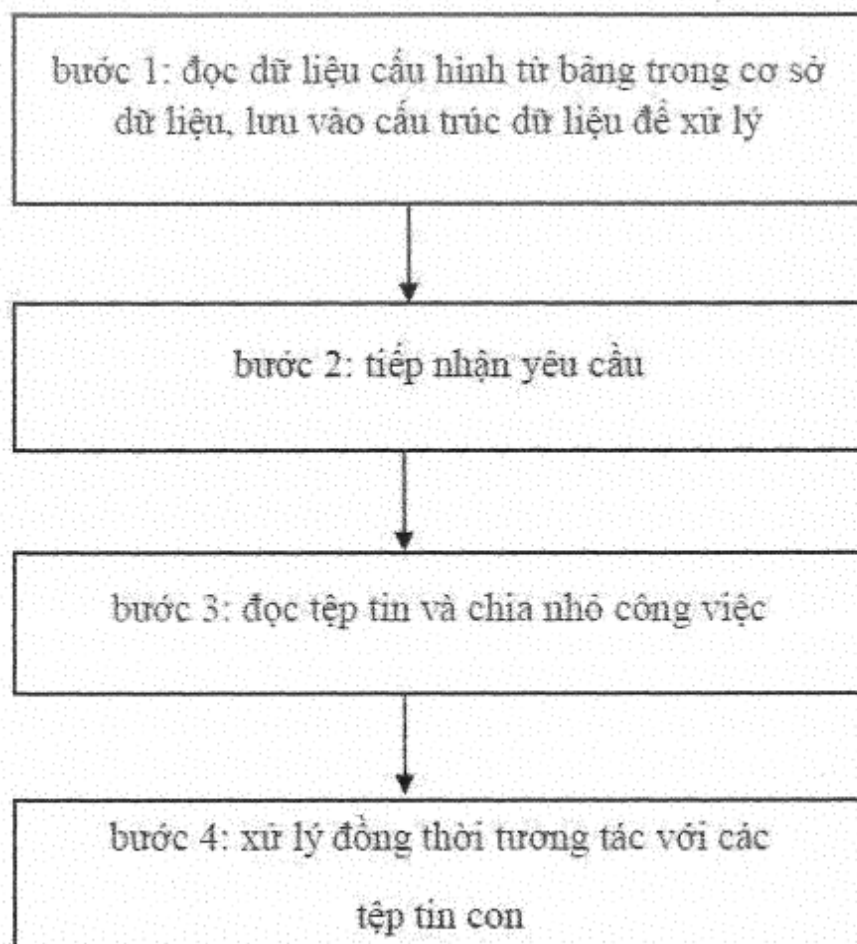
(72) TRỊNH VĂN CHUNG (VN).

(74) Công ty TNHH Tư vấn Quốc Dân (NACILAW)

(54) PHƯƠNG PHÁP PHÂN MẢNH TỆP TIN DUNG LƯỢNG LỚN ÁP DỤNG VỚI
BÀI TOÁN CẬP NHẬT DỮ LIỆU SỐ ĐIỆN THOẠI

(21) 1-2020-01147

(57) Phương pháp phân mảnh tệp tin dung lượng lớn áp dụng với bài toán cập nhật dữ liệu số điện thoại theo sáng chế đề xuất giúp giải quyết bài toán về tăng tốc độ của đọc/ghi/cập nhật/xóa tệp tin lớn, giải quyết được bài toán cao tải thi tương tác với một tệp tin có dữ liệu cỡ hàng triệu bản ghi bao gồm các bước: bước 1: đọc dữ liệu cấu hình từ bảng trong cơ sở dữ liệu, lưu vào cấu trúc dữ liệu để xử lý; bước 2: tiếp nhận yêu cầu; bước 3: đọc tệp tin và chia nhỏ công việc; bước 4: xử lý đồng thời tương tác với các tệp tin con.



Hình 4

Lĩnh vực kỹ thuật được đề cập

Sáng chế đề cập đến phương pháp phân mảnh tệp tin dung lượng lớn áp dụng với bài toán cập nhật dữ liệu số điện thoại phục vụ cho hệ thống chiến dịch tương tác khách hàng. Sáng chế được ứng dụng trong lĩnh vực công nghệ thông tin.

Tình trạng kỹ thuật của sáng chế

Nhiệm vụ quan trọng nhất của hệ thống chiến dịch tương tác khách hàng là kích thích tương tác trên một tập thuê bao cụ thể. Việc kích thích tương tác bao gồm: tạo các chương trình khuyến mại cho thuê bao, gửi lời mời đến thuê bao để tăng lượt tương tác.

Các bước thực hiện thêm mới như sau:

Bước 1: người vận hành khai thác tạo chương trình khuyến mại;

Bước 2: đẩy tệp tin chứa danh sách thuê bao được hưởng chương trình lên hệ thống;

Bước 3: hệ thống đọc tệp tin chứa danh sách thuê bao và thực hiện tương tác với từng thuê bao trong danh sách.

Trong quá trình thực hiện chiến dịch, có thể cập nhật thêm mới, sửa đổi, xóa một danh sách số điện thoại trong tệp tin ban đầu, quá trình này bao gồm các bước:

Bước 4: hệ thống đọc tệp danh sách thuê bao muốn thêm mới, sửa đổi, xóa;

Bước 5: cập nhật lại vào tệp tin ban đầu để lưu lại bản dữ liệu cuối cùng;

Bước 6: hệ thống thực hiện tương tác với tệp danh sách thuê bao được thay đổi.

Trong đó, bước 3 và bước 5 của quá trình thay đổi là bước quan trọng trong toàn bộ quá trình thực hiện chiến dịch tương tác. Quá trình này muốn đạt hiệu quả về tốc độ, cần giảm thời gian xử lý của bước thi vào tệp tin và bước cập nhật tệp tin dữ liệu lớn đến mức tối thiểu nhất.

Bản chất kỹ thuật của sáng chế

Mục đích của sáng chế là đề xuất phương pháp phân mảnh tệp tin dung lượng lớn áp dụng với bài toán cập nhật dữ liệu số điện thoại. Phương pháp giải quyết bài toán liên quan tới việc đọc, ghi mới và cập nhật tệp tin số điện thoại dữ liệu lớn (hàng triệu bản ghi), từ đó tối ưu thời gian xử lý của toàn bộ quá trình.

Để đạt được mục đích nêu trên, sáng chế đề xuất phương pháp phân mảnh tệp tin dung lượng lớn áp dụng với bài toán cập nhật dữ liệu số điện thoại với các bước thực hiện cụ thể như sau:

Bước 1: đọc dữ liệu cấu hình từ bảng trong cơ sở dữ liệu, lưu vào cấu trúc dữ liệu để xử lý.

Khi khởi tạo chiến dịch, người vận hành khai thác cần định nghĩa phương thức đọc, ghi tập tin số điện thoại. Phương thức này sẽ được lưu vào bảng dữ liệu trong cơ sở dữ liệu. Khi luồng xử lý đọc cấu hình (Campaign Cache) khởi chạy, thực hiện đọc bảng dữ liệu và lưu lại trên cấu trúc dữ liệu để xử lý.

Bước 2: tiếp nhận yêu cầu.

Khi nhận được yêu cầu xử lý tệp tin từ phía người vận hành, luồng tiếp nhận yêu cầu (Campaign Handler) thực hiện lấy thông tin định danh của chiến dịch trong yêu cầu, từ thông tin này lấy ra phương pháp đọc dữ liệu tệp tin trong cấu trúc dữ liệu được xây dựng ở bước 1.

Bước 3: đọc tệp tin và chia nhỏ công việc.

Luồng đọc tệp tin và phân phối (Campaign Read and Distribute) thực hiện truy cập vào tệp tin số điện thoại ban đầu, đọc từng số điện thoại. Dựa vào phương thức đọc ghi lấy được ra ở bước 2. Chuyển yêu cầu ghi dữ liệu sang các luồng xử lý song song.

Bước 4: xử lý đồng thời tương tác với các tệp tin con.

Tùy vào việc cấu hình, có thể phát sinh số luồng xử lý đồng thời tương ứng với số lượng cấu hình. Mỗi luồng xử lý một tập dữ liệu số điện thoại có chung một đặc tính cụ thể.

Giải pháp này dựa theo lý thuyết chia công việc lớn thành nhiều công việc nhỏ, từ đó có thể xử lý đồng thời các công việc nhỏ, giúp giảm tổng thời gian xử lý.

Mô tả vắn tắt các hình vẽ

Để sáng chế được mô tả một cách rõ ràng hơn, các hình vẽ bên dưới mô tả các phần trong sáng chế:

Hình 1 là hình vẽ mô tả mô hình luồng thêm mới tệp tin và xử lý đồng thời;

Hình 2 là hình vẽ mô tả mô hình luồng thêm mới tệp tin và xử lý cập nhật/xóa dữ liệu;

Hình 3 là hình vẽ mô tả cấu trúc dữ liệu lồng nhau ba cấp được sử dụng;

Hình 4 là hình vẽ phương pháp mà sáng chế đề xuất.

Mô tả chi tiết sáng chế

Trong hệ thống thực hiện chiến dịch tương tác khách hàng, nhiệm vụ chính là tăng tương tác với một tập lớn khách hàng trong một khoảng thời gian ngắn nhất. Do vậy, việc tối ưu tốc độ xử lý được ưu tiên hàng đầu.

Phương pháp phân mảnh tệp tin dung lượng lớn áp dụng với bài toán cập nhật dữ liệu số điện thoại giúp tối ưu khoảng thời gian đọc, ghi dữ liệu số điện thoại. Tham chiếu vào hình 4, phương pháp theo sáng chế đề xuất bao gồm các bước thực hiện tuần tự được nêu dưới đây:

Bước 1: đọc dữ liệu cấu hình từ bảng trong cơ sở dữ liệu, lưu vào cấu trúc dữ liệu để xử lý.

Khi khởi báo một chương trình chiến dịch tương tác khách hàng, thực hiện định nghĩa phương thức đọc, ghi khi có yêu cầu đọc/ghi/cập nhật tệp tin số điện thoại.

Phương thức đọc/ghi được hiểu là các luật cụ thể để chia tách tệp tin gốc thành các tệp tin nhỏ hơn. Các luật được lưu trên bảng quy định luật chia tệp tin, bảng có cấu trúc như sau:

Tên trường	Kiểu dữ liệu	Ý nghĩa
ID_LUAT	Số nguyên	Khóa chính của bảng, có kiểu dữ liệu số nguyên
IDCHIENDICH	Số nguyên	Định danh của chiến dịch

DIEUKIEN	Số nguyên	Mô tả điều kiện để chia tách file
MOTA	Chuỗi	Mô tả thêm thông tin luật

Sau khi có luật, dữ liệu bảng quy định danh sách tệp tin con được sinh dữ liệu, mô tả tên tệp tin con khi được chia tách.

Tên trường	Kiểu dữ liệu	Ý nghĩa
ID	Số nguyên	Khóa chính của bảng, có kiểu dữ liệu số nguyên
ID_LUAT	Số nguyên	Khóa ngoại, liên kết với bảng luật
TEN_TEP_TIN_CHA	Số nguyên	Tên tệp tin ban đầu được nhập vào
TEN_TEP_TIN_CON	Số nguyên	Tên tệp tin con được tách ra. Số lượng tệp tin con được tách ra bằng số bản ghi được sinh ra
MOTA	Chuỗi	Mô tả thêm thông tin

Luồng xử lý đọc cấu hình (Campaign Cache) thực hiện truy cập cơ sở dữ liệu và lấy thông tin từ hai bảng quy định luật chia tệp tin và bảng quy định danh sách tệp tin con dựa trên phương thức tương tác với cơ sở dữ liệu có sẵn trên cơ sở dữ liệu.

Sau khi lấy được dữ liệu, thông tin được lưu lại trên cấu trúc dữ liệu của chương trình. Sử dụng đồng thời hai cấu trúc dữ liệu: cấu trúc dữ liệu “khóa - nội dung đồng bộ” (ConcurrentHashMap) và cấu trúc dữ liệu danh sách (Array List).

Cấu trúc dữ liệu danh sách (Array List) dùng để lưu tập hợp danh sách các tệp tin con có trong bảng dữ liệu bảng quy định danh sách tệp tin con có cùng giá trị theo trường ID_LUAT. Mỗi bản ghi được biểu diễn bằng một đối tượng với các thuộc tính được biểu diễn như sau: TEPTIN{ ID_LUAT, TEN_TEP_TIN_CON, MOTA }

Tham chiếu hình 3, cấu trúc dữ liệu “khóa - nội dung đồng bộ” được sử dụng hai lớp:

Lớp phía trong sử dụng để mô tả một tệp tin ban đầu tách ra được bao nhiêu tệp tin con. Trong cấu trúc dữ liệu này, khóa được lấy từ thông tin TEN_TEP_TIN_CHA, hay tên tệp tin ban đầu được nhập vào, nội dung là cấu trúc dữ liệu danh sách bao gồm tập hợp những đối tượng có cùng giá trị này.

Lớp ngoài cùng sử dụng để mô tả một chiến dịch có bao nhiêu tệp tin cha và các tệp tin cha đó được chia thành bao nhiêu tệp tin con. Trong cấu trúc dữ liệu này, khóa được lấy từ thông tin định danh của chiến dịch (trường IDCHIENDICH theo bảng quy định luật chia tệp tin), nội dung là tập hợp các cấu trúc dữ liệu khóa-nội dung đồng bộ của lớp bên trong.

Kết thúc quá trình, kết quả đầu ra là cấu trúc dữ liệu khóa-nội dung đồng bộ hai lớp, phía bên trong lưu cấu trúc dữ liệu danh sách. Thông tin bao gồm danh sách các chiến dịch, tập hợp các tệp tin cha của từng chiến dịch, tập hợp các tệp tin con tương ứng với từng tệp tin cha.

Bước 2: tiếp nhận yêu cầu.

Luồng tiếp nhận yêu cầu (Campaign Handler) nhận được thông tin yêu cầu đọc/ghi/cập nhật dữ liệu từ tệp tin bên ngoài. Trong bản tin yêu cầu chứa thông tin định danh chiến dịch và đường dẫn đến tệp tin chứa danh sách số điện thoại.

Luồng xử lý lấy thông tin đường dẫn đến tệp tin trong yêu cầu, phương thức có sẵn hỗ trợ truy xuất tệp tin được sử dụng với tham số đầu vào là đường dẫn đến tệp tin.

Bước 3: đọc tệp tin và chia nhỏ công việc.

Luồng đọc tệp tin và phân phối (Campain Read and Distribute) thực hiện truy cập vào tệp tin số điện thoại ban đầu, đọc từng số điện thoại. Dựa vào phương thức đọc ghi lấy được ra ở bước 2. Chuyển yêu cầu ghi dữ liệu sang các luồng xử lý song song.

Sau khi truy xuất đến tệp tin đầu vào, sử dụng phương thức đọc tệp tin được ngôn ngữ lập trình hỗ trợ, thực hiện đọc từng dòng của tệp tin, mỗi dòng của tệp tin được lấy ra là một số điện thoại.

Luồng xử lý lấy thông tin định danh chiến dịch, đưa vào tìm khóa trong cấu trúc dữ liệu khóa – nội dung đồng bộ hai lớp, từ đó tìm ra phương pháp chia tách tệp

tin trong thuộc tính điều kiện của đối tượng, số lượng tệp tin con và nơi lưu trữ các tệp tin con.

Phương pháp tách tệp tin được hiểu là một luật, đưa số điện thoại có trong tệp tin ban đầu vào một trong cách tệp tin con.

Tham chiếu vào hình 1, nếu yêu cầu đầu vào là thêm mới dữ liệu, số điện thoại sẽ được ghi vào tệp tin con và được hệ thống thực hiện nghiệp vụ tương tác khách hàng.

Tham chiếu vào hình 2, nếu yêu cầu đầu vào là cập nhật hoặc xóa, luồng xử lý thực hiện tìm đến tệp tin con có chứa số điện thoại để thực hiện công việc cập nhật hoặc xóa.

Sau khi tìm được luồng xử lý phù hợp cho số điện thoại, thực hiện gửi yêu cầu đến các luồng xử lý tương ứng để thực hiện tiếp.

Ví dụ: Trường điều kiện sử dụng hàm: $SOSAUCUNG(2)\%2$. Được hiểu lấy hai số sau cùng của số điện thoại, chia lấy dư cho 2. Nếu kết quả là 0, số điện thoại này được luồng xử lý tệp tin con số 0 thực thi và lưu vào TỆP TIN CON 0. Nếu kết quả là 1, số điện thoại này được luồng xử lý tệp tin con số 1 thực thi và lưu vào TỆP TIN CON 1.

Bước 4: xử lý đồng thời tương tác với các tệp tin con.

Các luồng xử lý tệp tin con được hiểu là các luồng xử lý đồng thời, áp dụng cơ chế xử lý đa luồng của hệ điều hành (Multi Thread).

Mỗi tiến trình con xử lý một tệp tin con, do xử lý đồng thời nên tốc độ sẽ gần như tuyến tính theo số lượng tiến trình con xử lý.

Khi thực thi xong yêu cầu, các tiến trình con tự kết thúc phần đời của mình để giảm tải cho máy chủ và đợi các yêu cầu tiếp theo tiếp tục xử lý.

Hiệu quả đạt được của sáng chế

Giải pháp này giúp giải quyết bài toán về tăng tốc độ của đọc/ghi/cập nhật/xóa tệp tin lớn, giải quyết được bài toán cao tải thi tương tác với một tệp tin có dữ liệu cỡ hàng triệu bản ghi.

Yêu cầu bảo hộ

1. Phương pháp phân mảnh tệp tin dung lượng lớn áp dụng với bài toán cập nhật dữ liệu số điện thoại bao gồm các bước:

bước 1: đọc dữ liệu cấu hình từ bảng trong cơ sở dữ liệu, lưu vào cấu trúc dữ liệu để xử lý; cụ thể: khi khái báo một chương trình chiến dịch tương tác khách hàng, thực hiện định nghĩa phương thức đọc, ghi khi có yêu cầu đọc/ghi/cập nhật tệp tin số điện thoại (các luật) và được lưu trên bảng quy định luật chia tệp tin;

luồng xử lý đọc cấu hình thực hiện truy cập cơ sở dữ liệu và lấy thông tin từ hai bảng quy định luật chia tệp tin, và bảng quy định danh sách tên tệp tin dựa trên phương thức tương tác với cơ sở dữ liệu có sẵn trên cơ sở dữ liệu; sau khi lấy được dữ liệu, thông tin được lưu lại trên cấu trúc dữ liệu của chương trình; sử dụng đồng thời hai cấu trúc dữ liệu: cấu trúc dữ liệu khóa - nội dung đồng bộ và cấu trúc dữ liệu danh sách;

cấu trúc dữ liệu danh sách dùng để lưu tập hợp danh sách các tệp tin con có trong bảng dữ liệu danh sách tệp tin con có cùng giá trị theo trường khóa ngoại và liên kết với bảng quy định luật chia tệp tin; mỗi bản ghi được biểu diễn bằng một đối tượng với các thuộc tính được biểu diễn với ba thông số: giá trị theo trường khóa ngoại nêu trên; tên tệp tin con và mô tả;

bước 2: tiếp nhận yêu cầu; luồng tiếp nhận yêu cầu nhận được thông tin yêu cầu đọc/ ghi/cập nhật dữ liệu từ tệp tin bên ngoài; trong bản tin yêu cầu chứa thông tin định danh chiến dịch và đường dẫn đến tệp tin chứa danh sách số điện thoại; luồng xử lý lấy thông tin đường dẫn đến tệp tin trong yêu cầu, phương thức có sẵn hỗ trợ truy xuất tệp tin được sử dụng với tham số đầu vào là đường dẫn đến tệp tin;

bước 3: đọc tệp tin và chia nhỏ công việc; luồng đọc tệp tin và phân phối thực hiện truy cập vào tệp tin số điện thoại ban đầu, đọc từng số điện thoại; dựa vào phương thức đọc ghi lấy được ra ở bước 2, chuyển yêu cầu ghi dữ liệu sang các luồng xử lý song song; luồng xử lý lấy thông tin định danh chiến dịch, đưa vào tìm khóa trong cấu trúc dữ liệu khóa - nội dung đồng bộ hai lớp, từ đó tìm ra phương pháp chia tách tệp tin trong thuộc tính điều kiện của đối tượng, số lượng tệp tin con và nơi lưu

trữ các tệp tin con, đưa số điện thoại có trong tệp tin ban đầu vào một trong các tệp tin con;

bước 4: xử lý đồng thời; các luồng xử lý tệp tin con được hiểu là các luồng xử lý đồng thời, áp dụng cơ chế xử lý đa luồng của hệ điều hành; mỗi tiến trình con xử lý một tệp tin con, khi thực thi xong yêu cầu, các tiến trình con tự kết thúc phần đời của mình để giảm tải cho máy chủ và đợi các yêu cầu tiếp theo tiếp tục xử lý.

2. Phương pháp phân mảnh tệp tin dung lượng lớn áp dụng với bài toán cập nhật dữ liệu số điện thoại theo điểm 1, trong đó tại bước đọc dữ liệu cấu hình từ bảng trong cơ sở dữ liệu, các thông tin được lưu trên cấu trúc dữ liệu khóa – nội dung hai lớp để mô tả các quan hệ và điều kiện xử lý.

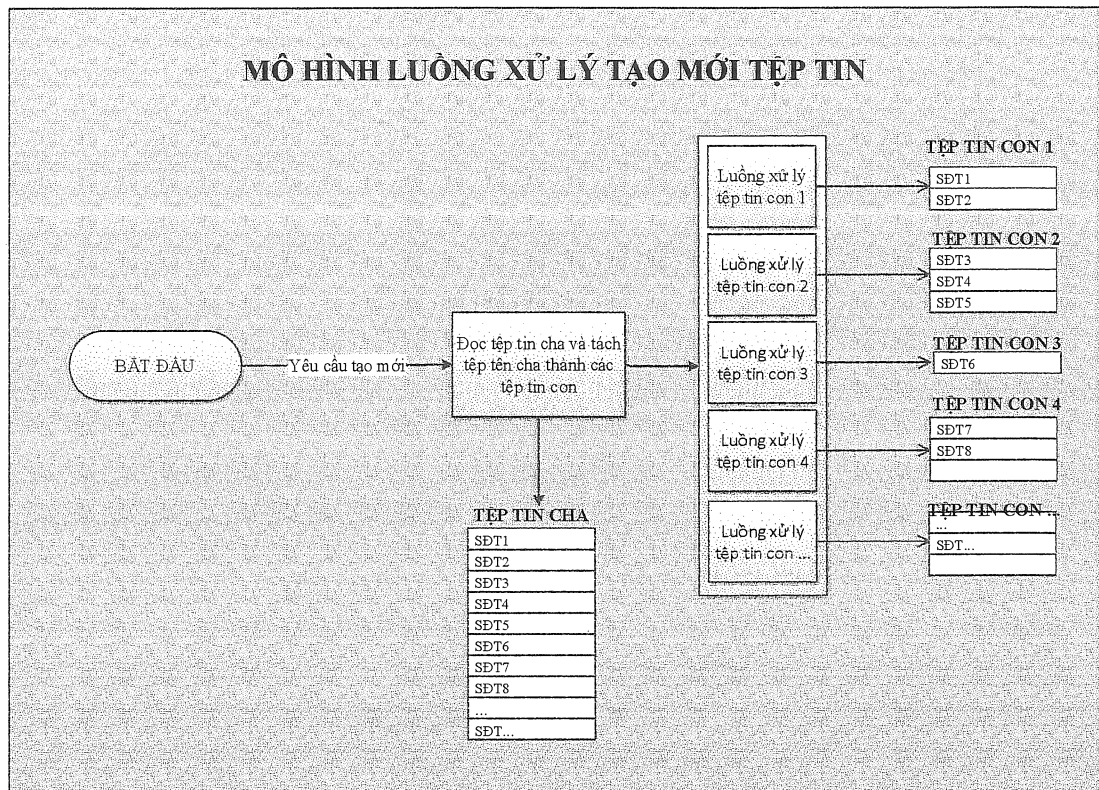
3. Phương pháp phân mảnh tệp tin dung lượng lớn áp dụng với bài toán cập nhật dữ liệu số điện thoại theo một trong số các điểm 1,2 trong đó:

nếu yêu cầu đầu vào là thêm mới dữ liệu, số điện thoại sẽ được ghi vào tệp tin con và được hệ thống thực hiện nghiệp vụ tương tác khách hàng;

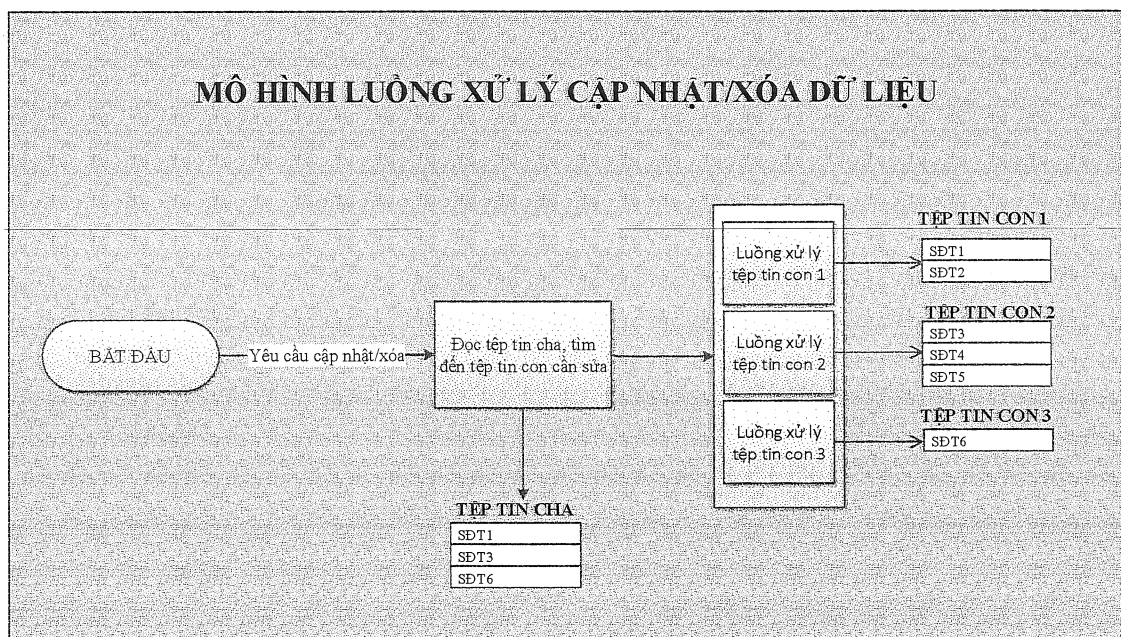
nếu yêu cầu đầu vào là cập nhật hoặc xóa, luồng xử lý thực hiện tìm đến tệp tin con có chứa số điện thoại để thực hiện công việc cập nhật hoặc xóa;

sau khi tìm được luồng xử lý phù hợp cho số điện thoại, thực hiện gửi yêu cầu đến các luồng xử lý tương ứng để thực hiện tiếp.

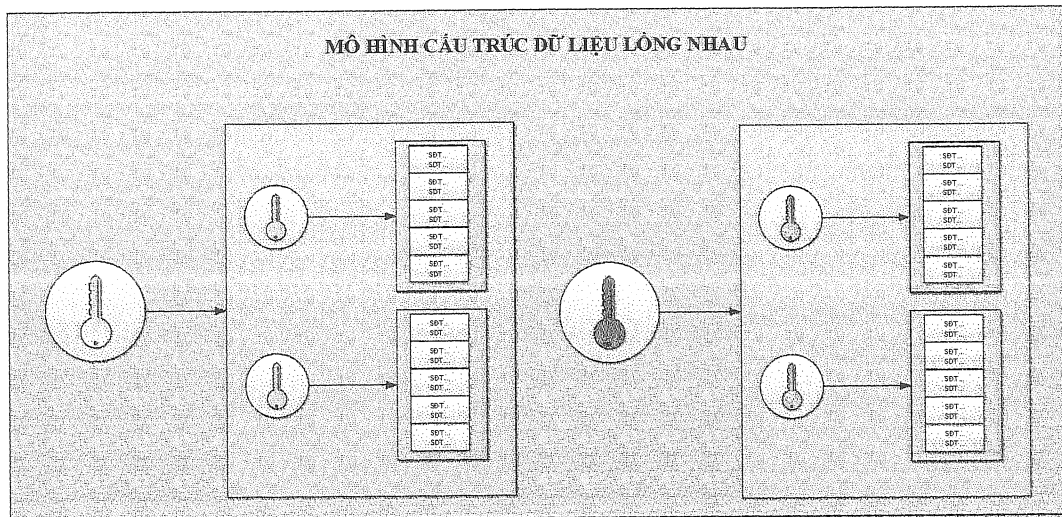
Hình vẽ



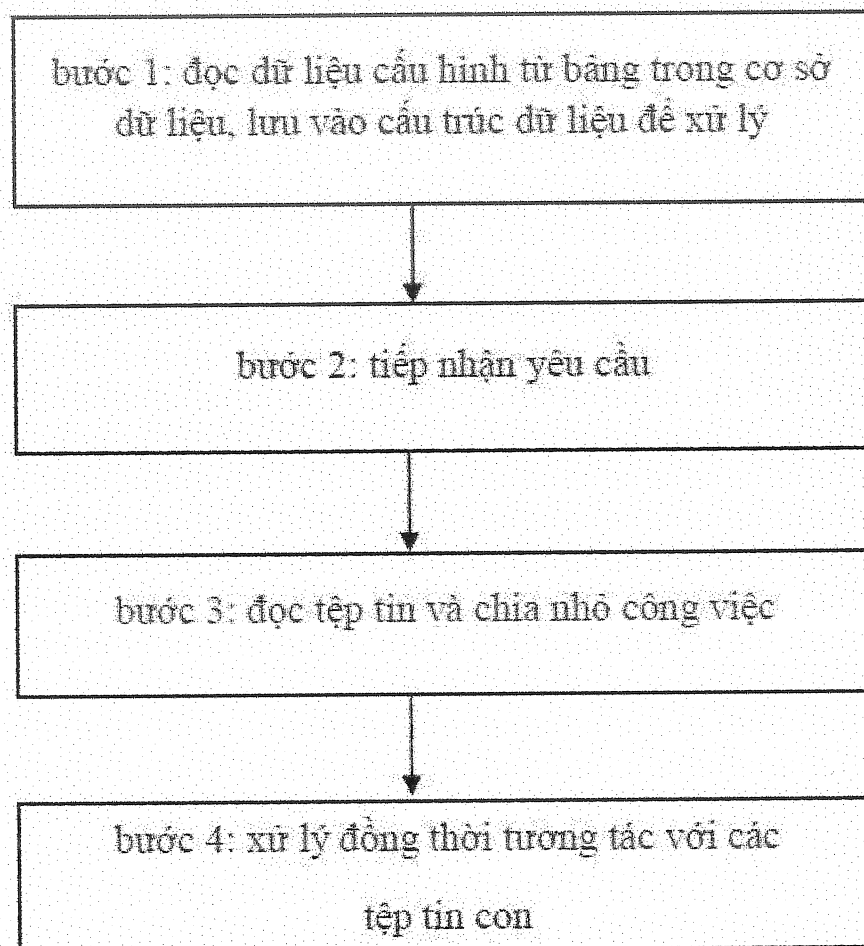
Hình 1



Hình 2



Hình 3



Hình 4