

Tổng quan bài toán

Giả sử GHTK có nhiều kho, mỗi kho có từ 5k-10k items (unique item), với mỗi item trên toàn hệ thống lại có trung bình 5k items có liên quan đến item đó. Chi tiết item nào liên quan đến item nào đã có và được tính toán trước

Tổng số items có trong hệ thống là **100M (100 triệu)**

Yêu cầu:

- Tính số lượng (count distinct) item có liên quan đến 1 kho bất kỳ
- Tốc độ phản hồi: ≤ 1 giây (1 second)

Dữ liệu đã cho là ma trận 100M x 5k, dạng

- item_1: <list item liên quan>
- item_2: <list item liên quan>
- item_3: <list item liên quan>
- ...
- item_100M: <list item liên quan>

Input: 1 kho bất kỳ kèm danh sách các unique items của kho

Output: Số nguyên thể hiện số lượng items liên quan đến toàn bộ items trong kho

NOTE: Vì tính bảo mật nên mô tả đã có sự thay đổi cho phù hợp

Ví dụ:

Ma trận cho:

1	2	3	4
2	1	3	5
3	1	2	
4	2	5	
5	1	4	

Input: Với kho X có 2 item (1,4)

Output: 4

Giải thích: Do kho X có 2 item là 1 và 4 mà:

- Item 1 có các item liên quan là (2,3,4)
- Item 4 có item liên quan là (2,5)
- => Số item liên quan của kho X là 4 - Bao gồm (2,3,4,5)

Định nghĩa thêm:

- **Liên quan:** Việc tính toán liên quan đã có mô hình tính toán và output ra dữ liệu sẵn

Thời gian

- Thời gian bắt đầu: 21/3/2024
- Kết thúc: 04/04/2024
- Thông báo kết quả và trao giải: 08/04/2024

Cách thức gửi giải pháp

- Tổng hợp thành 1 file zip: Bao gồm data model, hình ảnh miêu tả giải pháp, source code (Nếu có)
- Gửi thông tin về email: mannd@ghk.co, tuanla18@ghk.co, itlinhph@ghk.co

Giải thưởng

- Giải nhất: **3M** (Số lượng 01)
- Giải nhì: **800K** (Số lượng 01)
- Giải ba: **300k** (Số lượng 02)

Giải thưởng được BTC đồng ý đáp ứng được các yêu cầu

- Yêu cầu bắt buộc đã nêu
- Yêu cầu triển khai được với các phần cứng thực tế (không đòi các server chuyên dụng, siêu máy tính).
- Phương án đưa ra bao gồm data model/cấu trúc dữ liệu, database sử dụng, truy vấn (query) để count số lượng, giải pháp load data, update data

Xếp hạng giải thưởng dựa trên các tiêu chí

- Người đưa ra được giải pháp có tốc độ phản hồi nhanh nhất
- Sử dụng ít tài nguyên nhất
- Giải pháp đơn giản
- Các vấn đề về Scale, high availability
- Thời gian gửi phương án.

Liên hệ

- Nguyễn Đình Mẫn (itmannd)
- Phan Hồng Lĩnh (itlinhph)
- Lê Anh Tuấn (tuanla18)
- Địa chỉ: Tầng 8 GHTK Building