**IDEA**

1. **Ý tưởng xây dựng hệ thống**

- Đầu tiên user sẽ truy cập vào hệ thống của mình

- Hệ thống sẽ recommend cho user social trending (user chưa có vị trí) và CB, CF (dựa vào thao tác của user trước đó)

- Tiếp theo đó user sẽ thao tác với một địa danh bất kỳ hoặc search địa điểm cần đến (lúc này user đã có vị trí đứng). Hệ thống sẽ dựa vào thao tác đó để recommend:

+ Social trending (tương ứng với vị trí đứng của user): *xu hướng* dựa vào tập user đánh giá, rating, rank

+ CB (Content Base) dựa vào category, tính toán vector theo engine của từng địa điểm du lịch để tìm sự tương đồng giữa các địa điểm và recommend cho user

+ CF (Collarative Filtering) theo realtime, dựa vào tập data người dùng đến thời điểm hiện tại trên hệ thống *riêng* của mình và recommend cho user

1. **Database – MongoDB**

- Data crawl về dưới dạng row json, với các thuộc tính cần phân loại ở trước

Địa điểm 🡪 Category 🡪 Địa danh 🡪 Thuộc tính 🡪 Attraction

1. **Bài toán đặt ra**

- Nếu như user muốn du lịch nhiều loại hình du lịch thì sao??? Không lẽ chỉ recommend xung quanh 1 loại hình du lịch (CF giải quyết được 1 số, CB thì không, Social trending thì còn tùy)

1. **Đối tượng người dùng hướng đến**

- Đối tượng là những người có nhu cầu đi du lịch nhưng vẫn chưa biết về các điểm du lịch. Hệ thống cung cấp và recommend những điểm vui chơi, giải trí và tham quan cho khách

1. **Tập dữ liệu sử dụng**

- Sử dụng dữ liệu về các địa điểm du lịch, giải trí từ trang tripadvisor cho hệ thống.

- Data user lấy từ hệ thống mình tự xây dựng, lý do:

+ Các trang web hiện không cho lấy tập user, và nếu có đi chăng nữa thì chỉ nhận được các user có đánh giá và rating, bên cạnh đó có rất nhiều user từng xem qua địa điểm du lịch nhưng không đánh giá và rating thì không thể lấy được, bởi vậy data user sẽ không đủ và đúng

+ Trường hợp tự xây dựng tập data user, kiểm soát được các hành động của user với item, độ chính xác cao và xây dựng được tập data cho riêng mình.

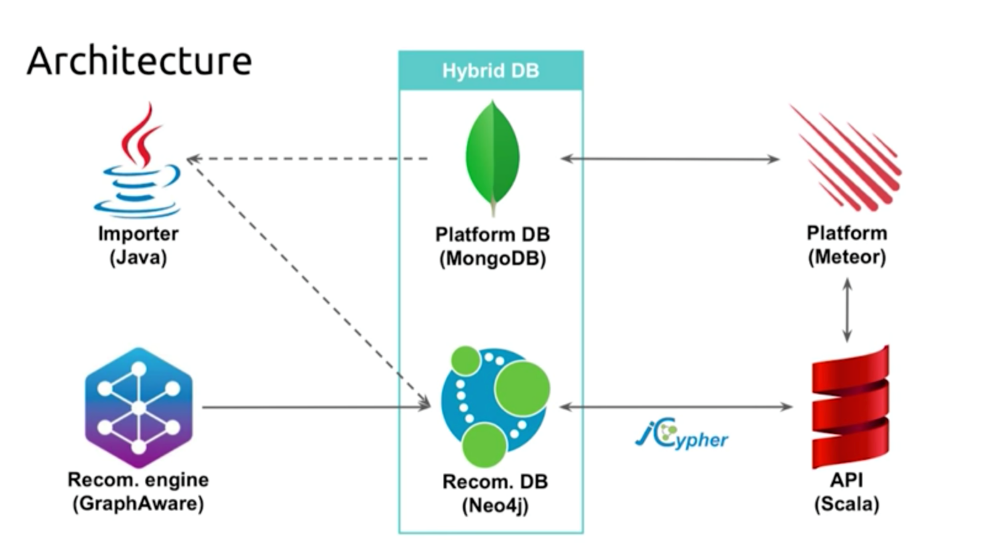
+ Trong thời gian ngắn ban đầu, tập data user không thể nhiều được, nhưng mình có thể dùng tập data user hiện tại để recommend để chứng minh thuật toán đúng, hệ thống đúng. Và tất nhiên khi lượng user tăng lên, hệ thống mình vẫn xử lý đúng thuật toán và và similar càng tăng cao.

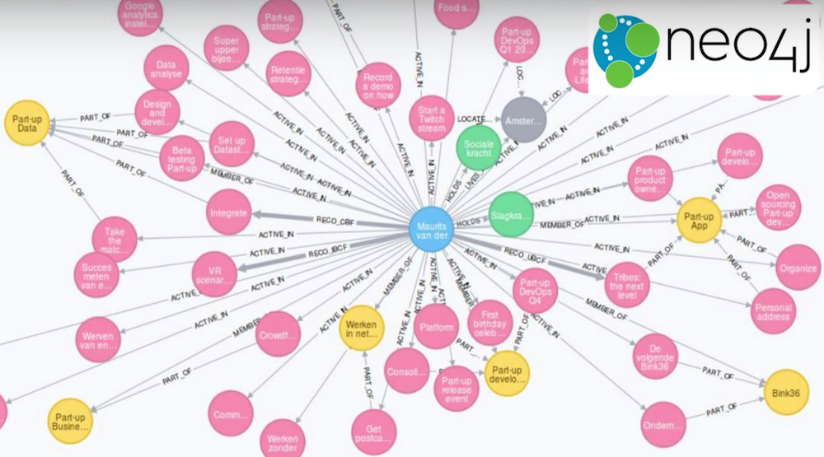
1. **Công cụ và thuật toán hỗ trợ:**

- MongoDB để lưu trữ data

- Neo4j để lưu data với dạng GraphDB

[*https://neo4j.com/blog/collaborative-filtering-creating-teams/*](https://neo4j.com/blog/collaborative-filtering-creating-teams/)





1. **Mô hình dữ liệu**

- Mô hình gồm có các Label (gọi là lớp):

+ Province

+ Catelory

+ Place

+ User

+ ….

- Mô hình Neo4j

1. **xxx**