Bài tập 1: (chương 1)

1. Chọn một hệ thống thông tin (information systems) mà phù hợp với việc lưu trữ dữ liệu bằng NoSQL – MongoDB

* Đưa ra tên của hệ thống
* Đưa ra phạm vi của hệ thống
* Nêu lý do tại sao lưu trữ bằng NoSQL – MongoDB là phù hợp nhất

1. Viết đặc tả sơ lược về hệ thống đó hoặc đưa ra yêu cầu của người dùng (user requirements) của hệ thống

* Hệ thống phần cứng của hệ thống (nếu có)
* Mô tả nhóm người dùng của hệ thống
* Mục đích chính của hệ thống này dùng để làm gì
* Cách thức thông tin dữ liệu được phát sinh (lấy về) từ người dùng hoặc hệ thống. Tần suất phát sinh dữ liệu.
* Những yếu tố nào có thể ảnh hưởng (không tốt) đến thông tin dữ liệu của hệ thống?
* Thông tin dữ liệu nào thường xuyên được xử lý, truy vấn. Thông tin/dữ liệu này được thực hiện bởi nhóm người dùng nào?
* Người dùng mong muốn lưu trữ thông tin dữ liệu trong khoảng thời gian bao lâu?
* Hiện tại người dùng đã có sẳn hệ thống phần cứng nào cần tận dụng lại?

Bài tập 2: (chương 2) Phân tích và đặc tả Workload cho hệ thống đã chọn - tương tự file pdf cô gửi “M320-workload-IOT”

1. Đưa ra danh sách các actor, và các trường hợp sử dụng (use case) của các actor đó
2. Đưa ra danh sách các CRUD của hệ thống, ứng với actor, Operation, Information, Frequency, criticality, ….
3. Phân tích chi tiết 2 thao tác (operation) nào mà quan trọng nhất.
4. Đưa ra mô hình lưu trữ dữ liệu ở mức logic: gồm các thực thể lưu trữ dữ liệu (tên thực thể, cấu trúc lưu trữ), mối quan hệ giữa chúng.
5. Đưa ra data size của document lưu trữ ứng với thao tác ghi (write) quan trọng
6. Đưa ra các câu truy vấn có thể có, đưa ra chỉ mục cần tạo

Bài tập 3: (Chương 3)

1. Thiết kế cấu trúc lưu trữ dữ liệu cho hệ thống ở mức vật lý. Giải thích lý do tại sao thiết kế như vậy
2. Đề xuất cấu hình phần cứng của hệ thống

Bài tập 4: (Chương 4)

Đưa ra một số pattern nên áp dụng trong thiết kế cấu trúc lưu trữ dữ liệu cho hệ thống ở mức vật lý. Giải thích lý do tại sao áp dụng các pattern đó. 🡪 MÔ HÌNH VẬT LÝ SAU CÙNG

Bài tập 5: (Chương 5)

Đề xuất 1 số giải pháp để giải quyết các vấn đề liên quan đến hiệu năng (performance) của hệ thống như: dung lượng đĩa cứng lưu trữ, indexes, Relication, Sharding…