**Bài tập chương 5**

**(môn: Mô hình hóa dữ liệu MongoDB)**

Nội dung 1: Để tăng hiệu năng cho hệ thống, hãy optimize từng câu truy vấn sau và đề xuất câu tạo chỉ mục tương ứng

1. db.users.find({numoflike: {$gte:50000}, gender: 'female', age: 29}).sort(name:1);
2. db.users.find({'Addresses.Country': 'Norfolk'}).skip(10).limit(20).pretty()
3. db.products.find({price: { $gt: 1, $lt: 5 }, in\_stock: true }).sort({ name: 1 })
4. db.user.find( { numoflike : { $gte : NumberDecimal(100000.00) }, city: "TP HCM" } ) .sort( { lastName: 1, firstName: 1 } )
5. db.products.find({'details.manufacturer': 'Acme','pricing.sale': { $lt: 7500}})
6. db.myColl.find( { score: 5, price: { $gt: NumberDecimal( "10" ) } } ).sort( { price: 1 } )

Nội dung 2:

Công ty bào hiểm ABC, có bán nhiều sản phẩm bảo hiểm khác nhau cho khách hàng ở toàn quốc (Việt Nam). Do nhiều lý do khác nhau, dữ liệu hệ thống cần được lưu trữ phân tán trên một hệ thống sharding gồm 3 shard node. Theo kết quả phân tích workload, cho thấy người dùng hệ thống thường xuyên truy vấn các hồ sơ bảo hiểm của khách hàng thành phố mà khách hàng trực thuộc.

Bạn hãy đề xuất một kế hoạch cho việc dựng một sharding sao cho tăng được hiệu năng của hệ thống trong khi truy vấn dữ liệu.

Nội dung 3: một số câu hỏi lý thuyết gợi ý

* Những yếu tố nào ảnh hưởng đến hiệu năng của hệ thống, cho ví dụ và giải thích lý do tại sao
* Để nâng cao hiệu năng của hệ thống, một trong những giải pháp là sử dụng RAM hiệu quả. Bạn đề xuất các giải pháp có thể có trong khi thiết kế cấu trúc lưu trữ dữ liệu mà giúp ta dùng RAM hiệu quả, giải thích tại sao?
* Hệ thống có sử dụng replication, hãy cho ví dụ 1 trường hợp là chúng ta nên đọc (Read) dữ liệu từ secondary tốt hơn là từ Primary, giải thích lý do
* Những yếu tố nào ảnh hưởng đến tốc độ của câu truy vấn (Find) dữ liệu trong mongoDB, cho ví vụ và nêu lý do tại sao?
* Đứng dước gốc độ người optimize hiệu năng của hệ thống hãy cho biết ưu điểm và nhược điểm lưu trữ dữ liệu phân tán trên hệ thống sharding.