

BÀI TẬP CUỐI KHÓA CAO HỌC K35-36 – CNTT

(Thực hiện theo nhóm: 3 học viên, 30 giờ)

Hạn nộp bài 05/01/2026. Nộp trên hệ thống <http://lms.hnue.edu.vn>

TS. Nguyễn Duy Hải

Trường Đại học Sư phạm Hà Nội

Yêu cầu cần đạt:

- ✓ Hiểu sâu các mô hình cơ sở dữ liệu hiện đại: *NoSQL, Distributed Database Systems.*
- ✓ Vận dụng kiến thức về phân mảnh dữ liệu, sao chép (*replication*), cân bằng tải (*load balancing*) trong thiết kế hệ thống phân tán.
- ✓ Thiết kế mô hình dữ liệu đa chi nhánh (*multi-site*) đảm bảo tính nhất quán, sẵn sàng, chịu lỗi theo định lý CAP.
- ✓ Cài đặt thử nghiệm hệ thống cơ sở dữ liệu *NoSQL (MongoDB)* với API Web để thao tác dữ liệu phân tán.
- ✓ Trình bày mô hình và đánh giá hiệu năng, độ tin cậy và tính mở rộng (*scalability*) của hệ thống.

Bối cảnh

Trường Đại học Sư phạm Hà Nội triển khai hệ thống “**Học liệu mở – SmartLearn**”, hỗ trợ giảng viên và học viên truy cập, lưu trữ và chia sẻ học liệu số tại nhiều cơ sở đào tạo khác nhau (Hà Nội, Đà Nẵng, Tp.HCM). (Học viên/Nhóm có thể lựa chọn các bài toán khác trong danh sách ở **Phụ lục I** và phân tích các bảng/thực thi dữ liệu cần phân tán)

Hệ thống cần đảm bảo:

- Truy cập nhanh từ các địa điểm khác nhau.
- Dữ liệu đồng bộ giữa các node khu vực.
- Khi một node gặp sự cố, hệ thống vẫn hoạt động bình thường (high availability).
- Hỗ trợ thông kê học liệu, người dùng và hoạt động học tập theo thời gian thực.

Yêu cầu bài tập

1. Thiết kế cơ sở dữ liệu NoSQL

- Chọn MongoDB để xây dựng mô hình dữ liệu học liệu:
 - users (người dùng: giảng viên, học viên)
 - courses (khóa học/môn học)
 - materials (học liệu / tài nguyên số)

- o activities (lịch sử truy cập, lượt tải, lượt xem)
- Phân tích mối quan hệ giữa các collection/table.
- Đè xuất mô hình truy vấn tối ưu (indexing, sharding key, partition key).
- Tạo dataset mẫu (~500–1000 bản ghi).

2. Xây dựng mô hình phân tán

- Triển khai tối thiểu 3 node MongoDB (2 node chi nhánh + 1 node trung tâm) trên Docker hoặc máy ảo.
- Thực hiện:
 - o Replication: đảm bảo sao chép dữ liệu tự động.
 - o Sharding: phân chia dữ liệu học liệu theo Khoa hoặc Cơ sở đào tạo.
 - o Failover test: kiểm tra khi 1 node ngắt kết nối, hệ thống vẫn phục vụ được truy vấn.
- Minh họa bằng sơ đồ kiến trúc mạng phân tán (Network Topology Diagram).

3. Phát triển hệ thống Web/API

- Xây dựng API hoặc ứng dụng web (Node.js, PHP, hoặc Python Flask/FastAPI) cho các chức năng:
 - o Đăng nhập và phân quyền (admin/giảng viên/học viên).
 - o Quản lý học liệu: thêm, sửa, xóa, tìm kiếm, lọc theo khóa học.
 - o Ghi log hoạt động vào collection activities.
 - o Thống kê số lượt tải/tương tác theo ngày hoặc người dùng.
- API thao tác trực tiếp với các node NoSQL qua driver hoặc ORM tương ứng.
- Triển khai giao diện báo cáo (Dashboard) hiển thị thông kê bằng Chart.js hoặc tương tự.
- Mã hóa mật khẩu người dùng (bcrypt hoặc password_hash).
- Sử dụng JWT (JSON Web Token) cho xác thực người dùng.

5. Báo cáo kết quả

Báo cáo gồm các phần:

1. Giới thiệu vấn đề và mục tiêu thiết kế.
2. Phân tích mô hình dữ liệu NoSQL.
3. Kiến trúc hệ thống phân tán (sơ đồ + mô tả).
4. Các truy vấn mẫu (insert, update, aggregate, map-reduce, API calls).

5. Kết quả kiểm thử & đánh giá hiệu năng.
6. Kết luận và hướng phát triển.

Tệp nộp gồm:

1. Mã nguồn ứng dụng (Web/API + cấu hình CSDL).
2. File dataset mẫu (.json hoặc script tạo dữ liệu).
3. Báo cáo PDF (15–20 trang).
4. Ảnh chụp màn hình demo.
5. Link video demo (tùy chọn).

6. Tiêu chí đánh giá kết quả xem tại Phụ lục II.

Phụ lục I. DANH SÁCH CÁC ĐỀ TÀI KHÁC

ĐỀ 1 — Hệ thống quản lý Học liệu số

- Lưu trữ hàng chục nghìn học liệu theo khoa, môn, giảng viên.
- Hệ thống NoSQL shard theo Khoa hoặc CourseID.
- Phân tích thống kê lượt truy cập theo thời gian thực.
- Chống trùng lặp học liệu bằng hàm hash.
- Yêu cầu kỹ thuật: MongoDB + API FastAPI.

ĐỀ 2 — Hệ thống e-Library phân tán nhiều cơ sở

- Quản lý sách, người mượn, lượt mượn, lịch sử giao dịch.
- Thiết kế hệ thống phân tán 3 node (Hà Nội – Hải Phòng – Đà Nẵng).
- Tối ưu truy vấn tìm kiếm bằng index + text search NoSQL.
- Đồng bộ dữ liệu tự động khi có giao dịch mượn/trả.
- Yêu cầu: MongoDB Replica Set + Full-text Search.

ĐỀ 3 — Hệ thống phân tích hành vi người học trên LMS (Learning Analytics)

- Thu thập log học tập (click, xem video, thời gian hoạt động...).
- Lưu log vào hệ thống NoSQL dung lượng lớn.
- Thực hiện thống kê: đường cong học tập, mức độ tương tác.
- Mô hình dữ liệu dạng time-series.
- Yêu cầu: MongoDB + Dashboard.

ĐỀ 4 — Hệ thống quản lý bệnh nhân phân tán cho bệnh viện

- Lưu trữ hồ sơ y tế đa chi nhánh (Clinic A, B, C).
- Hỗ trợ truy cập nhanh tại chỗ, đồng bộ theo thời gian thực.
- Lưu lịch sử khám, đơn thuốc, xét nghiệm.
- Bảo mật cao + phân quyền nghiêm ngặt.
- Yêu cầu: Replication + Sharding.

ĐỀ 5 — Hệ thống bán lẻ & phân tích hóa đơn thời gian thực

- Lưu trữ hóa đơn, sản phẩm, khách hàng, lịch sử giao dịch.
- Phân tích số liệu bán hàng theo ngày – tuần – tháng.
- Hệ thống phân tán đảm bảo chịu tải cao vào giờ cao điểm.

- Tối ưu write-heavy workload.
- Yêu cầu: MongoDB

ĐỀ 6 — Hệ thống gợi ý sản phẩm (Recommendation System)

- Thu thập hành vi người dùng (click, view, rating).
- Lưu vào NoSQL và phân tích bằng các truy vấn aggregation pipeline.
- Xây dựng API gợi ý sản phẩm dựa trên dữ liệu phân tán.
- Tối ưu tốc độ truy cập.
- Yêu cầu: MongoDB.

Phụ lục II. RUBRIC ĐÁNH GIÁ BÀI TẬP CUỐI KỲ

Tiêu chí đánh giá	Mô tả yêu cầu	Điểm tối đa
1. Thiết kế mô hình CSDL NoSQL	<ul style="list-style-type: none"> - Mô hình dữ liệu logic + physical đầy đủ.- Lựa chọn key, partition key, shard key hợp lý. - Thể hiện mối quan hệ và chiến lược truy vấn tối ưu. - Dataset mẫu đa dạng (≥ 500 bản ghi). 	20 điểm
2. Triển khai hệ thống CSDL phân tán	<ul style="list-style-type: none"> - Cấu hình ≥ 3 node (2 chi nhánh + 1 trung tâm). - Triển khai replication/sharding đúng chuẩn. - Test failover và mô tả kết quả. - Sơ đồ kiến trúc phân tán rõ ràng. 	20 điểm
3. Xây dựng API/Web kết nối NoSQL	<ul style="list-style-type: none"> - Tối thiểu 4 nhóm chức năng CRUD hoàn chỉnh. - API chạy ổn định, trả kết quả đúng. - Có thống kê/aggregation pipeline. - Giao diện thân thiện (nếu có). 	15 điểm
4. Xử lý truy vấn và tính toán nâng cao	<ul style="list-style-type: none"> - Sử dụng đúng và hiệu quả các truy vấn: aggregation, index, map-reduce hoặc equivalent. - Minh chứng tối ưu hóa truy vấn. - So sánh hiệu năng khi sharding. 	15 điểm

Tiêu chí đánh giá	Mô tả yêu cầu	Điểm tối đa
5. Bảo mật và phân quyền	<ul style="list-style-type: none"> - Áp dụng session/JWT. - Hash mật khẩu đúng chuẩn. - Kiểm soát truy cập theo vai trò (RBAC). - Xử lý lỗi bảo mật cơ bản (SQL/NoSQL injection, brute-force). 	10 điểm
6. Hiệu năng & đánh giá hệ thống	<ul style="list-style-type: none"> - Chạy thử nghiệm thực tế với dataset lớn. - Báo cáo latency, throughput, replication lag. - Phân tích ưu/nhược điểm mô hình. 	10 điểm
7. Báo cáo cuối kỳ (PDF)	<ul style="list-style-type: none"> - Trình bày mạch lạc, khoa học. - Có đầy đủ: mô hình CSDL, kiến trúc phân tán, truy vấn, API, ảnh demo. - Có phân tích, tự đánh giá, hướng phát triển. 	5 điểm
8. Demo & trả lời vấn đáp	<ul style="list-style-type: none"> - Demo mượt, đầy đủ chức năng. - Trả lời rõ ràng các câu hỏi về thiết kế, hiệu năng, mô hình phân tán. 	5 điểm
Tổng điểm		100 điểm

