# Câu hỏi nghiên cứu:

"Khi xử lý Big Data, CSDL SQL và NoSQL thể hiện những ưu thế và giới hạn nào khả năng mở rộng?"

# Thực hiện tìm kiếm

## Từ khoá tìm kiếm:

+ So sánh CSDL SQL vs NoSQL

+ Mở rộng ngang (Horizontal Scaling)

## Tóm tắt những hiểu biết

### **Giới hạn của SQL đối với Khả năng Mở rộng trong Big Data**

Ngôn ngữ SQL là ngôn ngữ truy vấn tuyên bố, cung cấp sức mạnh vượt trội cho các truy vấn phức tạp liên quan đến phép nối (JOIN) dữ liệu từ nhiều bảng khác nhau. Tuy nhiên, điều này làm cho SQL trở nên kém hiệu quả hơn khi xử lý dữ liệu lớn (Big Data) vì việc JOIN dữ liệu trên các máy chủ phân tán là một thách thức lớn.

Các hệ thống CSDL SQL truyền thống được định nghĩa là hệ thống quan hệ với cấu trúc dữ liệu được chuẩn hóa, do đó, chúng thường ưu tiên tính nhất quán (Consistency), tuân thủ nghiêm ngặt nguyên tắc ACID. Sự ưu tiên này thường dẫn đến việc mở rộng theo chiều dọc (Vertical Scaling) trở nên dễ dàng hơn so với chiều ngang, tạo ra giới hạn về khả năng mở rộng chi phí thấp.

### **Ưu thế của NoSQL đối với Khả năng Mở rộng Ngang**

Các hệ thống NoSQL thường sử dụng API hoặc các ngôn ngữ truy vấn đơn giản, nhưng chúng lại hoàn toàn tối ưu cho khả năng mở rộng ngang (Horizontal Scaling). Điều này cho phép hệ thống phân tán khối lượng công việc trên nhiều máy chủ một cách hiệu quả để xử lý lưu lượng truy cập lớn.

Mặc dù vậy, cần lưu ý rằng CSDL NoSQL thường hoạt động theo mô hình BASE (Basically Available, Soft State, Eventually Consistent), ngụ ý ưu tiên tính sẵn sàng (Availability) và khả năng chịu lỗi (Partition Tolerance) hơn tính nhất quán tức thời, đây là sự đánh đổi để đạt được khả năng mở rộng cao.

# Bảng đánh giá CRAAP

|  |  |  |
| --- | --- | --- |
| **NGUỒN TÀI LIỆU** | **SQL and NoSQL Database Software Architecture Performance Analysis and Assessments — A Systematic Literature Review** (Khan W. et al., 2023 | **Comparative Study of SQL and NoSQL Databases to …** (bài IEEE, năm chưa rõ đầy đủ) |
| **TÍNH THỜI SỰ** | **5** | **3** |
| **TÍNH LIÊN QUAN** | **5** | **4** |
| **TÍNH UY QUYỀN** | **4** | **3** |
| **TÍNH CHÍNH XÁC** | **4** | **3** |
| **MỤC ĐÍCH** | **4** | **3** |
| **ĐÁNH GIÁ TỔNG** | Thời sự rất tốt (2023), chủ đề rất phù hợp với nghiên cứu “so sánh SQL vs NoSQL”. Tác giả thuộc nhiều trường đại học, bài đăng trên tạp chí quốc tế mở (MDPI) → uy quyền tốt, nhưng MDPI đôi khi bị phê phán về rigor nên không đạt tuyệt đối. Dữ liệu phân tích khá chi tiết, song vì là SLR nên dựa vào nhiều nghiên cứu khác chứ không phải thử nghiệm riêng nên có hạn chế. | Thời sự trung bình (không rõ năm cập nhật chính xác từ trang tóm tắt truy cập), liên quan khá tốt nhưng có vẻ nội dung ít chi tiết hoặc khó truy cập (abstract trên IEEE bị lỗi). Uy quyền => bài IEEE có uy nhưng nếu không đọc đầy đủ và kiểm chứng thì khó đánh giá cao hơn. Chính xác – chưa đủ thông tin rõ ràng từ abstract. Mục đích – nghiên cứu so sánh, phù hợp. |

# Tổng hợp kết quả

|  |  |  |  |
| --- | --- | --- | --- |
| **Tiêu chí** | **SQL** | **NoSQL** | **Nhận xét tổng từ hai nguồn** |
| **Kiến trúc dữ liệu** | |  | | --- | | Cấu trúc bảng, ràng buộc khóa chính–ngoại, quan hệ rõ ràng. | | |  | | --- | | Lưu trữ phi cấu trúc: document, key-value, column, graph. | | MDPI cho rằng SQL phù hợp với dữ liệu có tính logic cao; IEEE bổ sung rằng NoSQL tối ưu khi dữ liệu phân tán |
| **Hiệu năng & khả năng mở rộng** | Mở rộng theo chiều dọc (vertical scaling). | Mở rộng theo chiều ngang (horizontal scaling). | Cả hai nguồn đều thống nhất NoSQL mở rộng tốt hơn trong môi trường dữ liệu lớn. |
| **Tính nhất quán (Consistency)** | Tuân thủ nghiêm ACID. | Dựa trên BASE, cho phép eventual consistency. | MDPI phân tích rằng SQL đảm bảo độ chính xác, còn NoSQL hy sinh nhất quán để đạt tốc độ. |
| **Ứng dụng thực tế** | Tài chính, kế toán, ERP, hệ thống doanh nghiệp | Mạng xã hội, Big Data, IoT, thương mại điện tử. | IEEE khuyến nghị kết hợp cả hai tùy loại ứng dụng (“polyglot persistence”). |
| **Ưu – nhược điểm chính** | Dữ liệu ổn định, dễ truy vấn, có chuẩn SQL phổ biến. Nhược: khó mở rộng, cứng nhắc. | Linh hoạt, tốc độ cao, dễ mở rộng. Nhược: khó duy trì toàn vẹn dữ liệu, thiếu chuẩn chung. | MDPI nhận xét xu hướng hiện nay là **kết hợp SQL + NoSQL** trong kiến trúc lai. |

*CHÚ Ý:*

*Qua hai nguồn tài liệu trên, có thể thấy rằng CSDL SQL phù hợp với các hệ thống cần độ chính xác và toàn vẹn dữ liệu cao, trong khi CSDL NoSQL vượt trội khi xử lý khối lượng dữ liệu lớn, đa dạng và yêu cầu tốc độ cao. Nghiên cứu của Khan et al. (2023) đề xuất mô hình kết hợp lai để tận dụng ưu điểm của cả hai, còn bài IEEE (2013) cũng nhấn mạnh rằng lựa chọn phụ thuộc vào mục tiêu và loại dữ liệu của từng ứng dụng. Tóm lại, xu hướng hiện nay là kết hợp SQL và NoSQL để vừa đảm bảo toàn vẹn vừa đạt hiệu năng cao.*

# Trích dẫn tài liệu tham khảo

|  |  |
| --- | --- |
| [1] | F. A. Al-Jubari, A. S. M. Hashim and N. Al-Mekhlafi, "SQL and NoSQL Database Software: A Comprehensive Survey of Data Modeling and Querying in NoSQL Databases," *Big Data and Cognitive Computing,* vol. 7, no. 2, 2023. |
| [2] | Z. Beknozarova, N. Tewari and V. K. Deolia, "IoT Open Lora Structure: Implementation Perspectives," *International Conference on Communication, Computing and Energy Efficient Technologies (I3CEET),* pp. 1515-1520, 2024. |
| [3] | M. Illian, C. Zinda and D. Schlangenotto, "Data Store Architectures: Balancing Functionality and Performance," *IEEE International Conference on Industrial Technology (ICIT),* pp. 1-6, 2024. |