

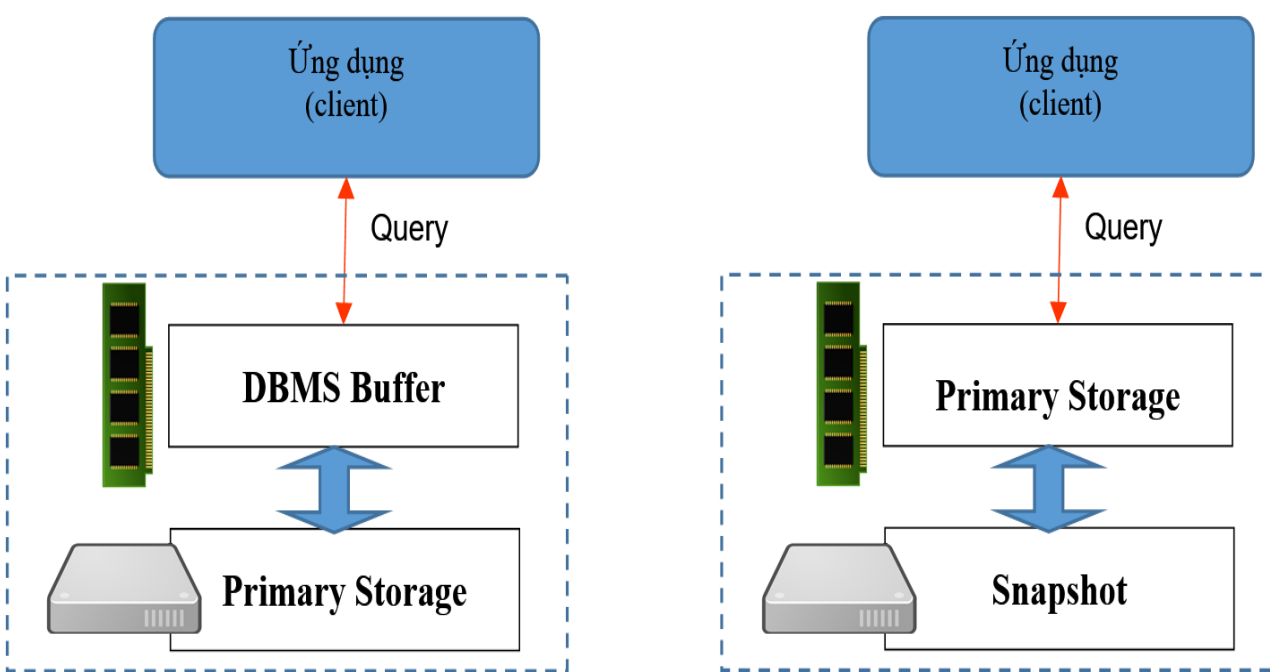
Nâng cao hiệu năng xử lý truy vấn với mô hình dữ liệu trong bộ nhớ

Dư Phương Hạnh, Nguyễn Hải Châu, Nguyễn Kim Khoa



Giới thiệu

- Các hệ cơ sở dữ liệu (CSDL) truyền thống, dữ liệu được lưu trữ trên thiết bị từ tính, phổ biến vẫn là đĩa cứng. Khi thi hành truy vấn, các phương pháp xử lý cơ bản vẫn phải dựa trên thao tác với đĩa cứng. Chính vì thế mà hiệu năng xử lý không những phụ thuộc các thao tác xử lý dữ liệu mà còn phụ thuộc rất lớn vào các thao tác đọc/ghi trên đĩa. Vì thế nếu giảm thiểu hay loại bỏ được các thao tác truy xuất đĩa sẽ có ý nghĩa với hiệu năng xử lý của hệ thống

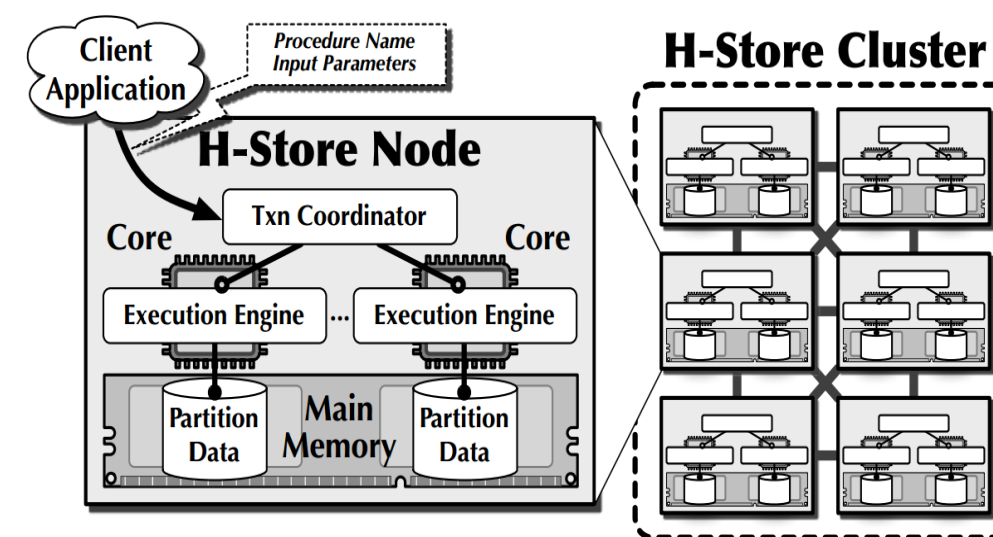


a. Hệ quản trị CSDL truyền thống

b. Hệ quản trị CSDL trong bộ nhớ

- Khi bộ nhớ bán dẫn ngày càng rẻ và dung lượng càng lớn, cách tiếp cận đưa tất cả dữ liệu vào bộ nhớ chính trở nên khả thi hơn. Phương án này hình thành cách tiếp cận xây dựng các hệ quản trị CSDL trong bộ nhớ (In Memory Database - IMDB). Với mô hình IMDB, việc tối ưu hoá truy vấn sẽ không cần phải quan tâm nhiều đến chi phí cho phần vào/ra trên thiết bị lưu trữ, một trong những yếu tố then chốt ảnh hưởng đến hiệu năng xử lý giao tác [1]. Ưu điểm này đến từ bản chất của bộ nhớ chính: dữ liệu được truy xuất theo kiểu RAM (Random Access Memory) thay vì kiểu truy cập trực tiếp như ổ đĩa cứng. Chính vì thế, việc thi hành các giao tác của người dùng trên các hệ quản trị CSDL trong bộ nhớ có hiệu năng vượt trội so với những hệ quản trị truyền thống.

- Các hệ CSDL truyền thống cũng thường được xây dựng phương pháp xử lý trên mô hình hệ thống tính toán đơn nhân CPU. Chính vì thế chưa khai thác hết năng lực tính toán của những hệ thống máy tính có nhiều CPU, mỗi CPU lại cho nhiều lõi khác nhau [2].



c. Kiến trúc xử lý nhóm đa lõi H-Store

- Ngoài ra, với công nghệ tính toán đám mây hiện nay, chúng ta hoàn toàn có thể khai thác những điểm mạnh của công nghệ này để cấp phát và quản lý linh động cả tài nguyên tính toán CPU lẫn bộ nhớ để có thể nâng cao hiệu năng xử lý các truy vấn đồng thời [3-4].

Mục tiêu

- Nghiên cứu, đánh giá mô hình dữ liệu trong bộ nhớ và những phương pháp xử lý truy vấn đồng thời
- Đề xuất giải pháp xử lý các truy vấn không gian theo cách tiếp cận khai thác tối đa không gian bộ nhớ chính
- Xây dựng giải pháp xử lý đồng thời nhiều truy vấn trên các hệ thống tính toán có nhiều lõi, nhiều CPU
- Đề xuất phương pháp cấp phát và quản lý tài nguyên bộ nhớ và CPU động khi xử lý truy vấn trong môi trường tính toán đám mây.
- Thử nghiệm với những bài toán kinh điển như tìm lộ trình tối ưu giữa hai vị trí địa lý, các truy vấn đơn giản (thêm/xoá/tìm kiếm) nhưng quy mô lớn trên mạng xã hội, mạng IoT,

Kết quả dự kiến

- Phương pháp và kỹ thuật xử lý các truy vấn đối với dữ liệu không gian khai thác tối đa được không gian bộ nhớ chính
- Một số kỹ thuật xử lý đồng thời nhiều truy vấn trên các hệ thống tính toán có nhiều lõi, nhiều CPU
- Phương pháp cấp phát và quản lý tài nguyên bộ nhớ và CPU động khi xử lý truy vấn trong môi trường tính toán đám mây.
- Công bố từ 3-4 công trình được đăng trong kỷ yếu hội nghị quốc tế

Tài liệu tham khảo

- [1] J. DeBrabant, A. Pavlo, S. Tu, M. Stonebraker, and S. Zdonik, "Anti-Caching: A New Approach to Database Management System Architecture," Proc. VLDB Endow., vol. 6, pp. 1942-1953, 2013.
- [2] M. Serafini, et al., *Accordion: Elastic Scalability for Database Systems Supporting Distributed Transactions*. Proceedings of the VLDB Endowment (PVLDB), volume 7, number 12, pages 1035-1046, August 2014.
- [3] Goetz Graefe, Haris Volos, Hideaki Kimura, Harumi Kuno, Joseph Tucek, Mark Lillibridge, and Alistair Veitch. 2014. In-memory performance for big data. Proc. VLDB Endow. 8, 1 (September 2014)
- [4] Darius Šidlauskas and Christian S. Jensen. 2014. Spatial joins in main memory: implementation matters!. Proc. VLDB Endow. 8, 1 (September 2014), 97-100.
- [5] K. Molka, G. Casale, T. Molka and L. Moore, "Memory-aware sizing for in-memory databases," *Network Operations and Management Symposium (NOMS), 2014 IEEE, Krakow, 2014*, pp. 1-9.