

PHÂN MẢNH DỮ LIỆU TRONG THIẾT KẾ CƠ SỞ DỮ LIỆU PHÂN TÁN DỰA VÀO KỸ THUẬT PHÂN CỤM HƯỚNG TRI THỨC

FRAGMENTATION IN DISTRIBUTED DATABASE DESIGN BASED ON KNOWLEDGE-ORIENTED CLUSTERING TECHNIQUE

Tác giả: [Lê Văn Sơn](#), [Lương Văn Nghĩa](#)

Tóm tắt bằng tiếng Việt:

Bài toán tối ưu hóa cơ sở dữ liệu phân tán bao gồm các bài toán: phân mảnh và định vị dữ liệu. Có nhiều phương pháp tiếp cận khác nhau và nhiều thuật toán được đề xuất để giải quyết các bài toán này. Tuy nhiên, độ phức tạp của thuật toán vẫn còn là một thách thức. Trong bài báo này, chúng tôi sử dụng kỹ thuật phân cụm hướng tri thức cho cả hai bài toán phân mảnh ngang và phân mảnh dọc dữ liệu. Độ đo tương tự sử dụng trong hai thuật toán là các độ đo được phát triển từ các độ đo cổ điển. Kết quả thử nghiệm trên các tập dữ liệu nhỏ hoàn toàn trùng khớp với kết quả phân mảnh dựa vào các thuật toán cổ điển. Thời gian thực hiện phân mảnh dữ liệu cũng được giảm đáng kể (mặc dù độ phức tạp thuật toán trong trường hợp tổng quát vẫn chưa thay đổi).

Từ khóa: Cơ sở dữ liệu phân tán; Phân mảnh; Định vị; Độ đo tương tự; Phân cụm; kỹ thuật phân cụm hướng tri thức.

Tóm tắt bằng tiếng Anh:

The optimization problem of data fragmentation is requiring to several interrelated problems including: Data fragmentation and Data allocation. Although we had many different algorithms to approach solving problems, the complexity of algorithm is always a big challenge to solve. In this paper, we presented a knowledge-oriented clustering technique that is applying both of vertical fragmentation and horizontal fragmentation problems. Similarity measures are used in both of algorithms which were built in the traditional measures. The experimental result of small data files and the fragmentation result based-on traditional algorithm are similar. The execution time of fragmented data is significantly reduced. (Although, the complexity of algorithm in the general case is still unchanged).

Key words: Distributed database; Fragmentation; Allocation; Similarity measures; Clustering; Knowledge-oriented clustering technique.