Tạp chí Tin học và Điều khiển học, T. 18, S. 2 (2002), 149–154

ĐẠI SỐ KHỐI TRÊN CÁC CƠ SỞ DỮ LIỆU ĐA CHIỀU

NGUYỄN KIM ANH

Abstract. On-Line Analytical Processing (OLAP) is a new technology that refers to the techniques of performing complex analysis over a very large repository of historical data. The complexity of queries required to support OLAP applications makes it difficult to implement using standard relational database technology. In this paper, we present a model for multidimensional data cubes and an algebra to support OLAP operations on this cubes. The model which presented is independent of specific implementation and the cube algebra provides a mean to concisely express the analytical queries typical of OLAP environment.

Tóm tắt. Xử lý phân tích trực tuyến (OLAP) là một công nghệ mới với các kỹ thuật thực hiện các phân tích phức tạp trên một kho dữ liệu lịch sử rất lớn. Tính phức tạp của các câu hỏi được yêu cầu để hỗ trợ các ứng dụng OLAP làm cho nó khó cài đặt với việc sử dụng công nghệ cơ sở dữ liệu quan hệ. Trong bài báo này, chúng tôi trình bày một mô hình dữ liệu đa chiều và một đại số hỗ trợ các phép toán OLAP trên các khối đó. Mô hình dữ liệu mà chúng tôi trình bày là độc lập đối với một sự cài đặt cụ thể và đại số khối cung cấp một công cụ để biểu diễn một cách chính xác các câu hỏi phân tích điển hình của môi trường OLAP.

1. GIỚI THIỆU

Cơ sở dữ liệu đa chiều (CSDLĐC) là một bộ sưu tập các dữ liệu lớn, mang tính lịch sử, được sử dụng với mục đích phân tích dữ liệu để hỗ trợ việc ra các quyết định. Công việc này được hỗ trợ bởi một công nghệ phần mềm với tên gọi OLAP "Xử lý phân tích trực tuyến" và nhiều công cụ thương mại với OLAP đã được đưa ra thị trường. Hệ thống OLAP hộ trợ phân tích dữ liệu nhờ mô hình dữ liệu đa chiều (MHDLĐC). Việc xây dựng MHDLĐC cùng với các phép toán hỗ trợ cho việc phân tích đa chiều là không đơn giản. Cho đến nay, chưa có một MHDLĐC chuẩn và có nhiều cách tiếp cận để xây dựng MHDLĐC cùng các phép toán trên mô hình đó. Một trong các cách tiếp cận xây dựng MHDLĐC là xây dựng MHDLĐC dựa trên cơ sở mô hình quan hệ, do vậy, các phép toán đa chiều trong mô hình đó có thể được kế thừa và mở rộng từ các phép toán của đại số quan hệ truyền thống. Tuy nhiên, các cách tiếp cận mô hình hóa trực tiếp và tự nhiên hơn các CSDLĐC là các cách tiếp cận hướng khối. Trong bài báo này, chúng tôi trình bày một cách tiếp cận hướng khối mô hình hóa CSDLĐC và một đại số khối bao gồm các phép toán trên các khối của CSDLĐC. Chúng tôi cũng đưa ra các đánh giá về khả năng biểu diễn của đại số khối và so sánh cách tiếp cận của chúng tôi với các cách tiếp cận khác. Nội dung của bài báo được trình bày như sau. Phần 2 trình bày một định nghĩa hình thức về MHDLĐC với các khái niệm cơ bản như chiều, khối và CSDLĐC. Phần 3 giới thiệu một đại số khối của CSDLĐC và các ví dụ biểu diễn các câu hỏi phân tích điển hình của môi trường OLAP. Cuối cùng, để kết luận, trong Phần 4 chúng tôi đưa ra các đánh giá về khả năng biểu diễn và các so sánh.

2. MÔ HÌNH DỮ LIỆU ĐA CHIỀU

2.1. Chiều và khối dữ liệu

Giả sử chúng ta có các tập sau: $\underline{\mathbf{L}}$: tập mức (level) trong đó với mỗi mức l thuộc $\underline{\mathbf{L}}$ sẽ có miền trị tương ứng dom(l); $\underline{\mathbf{D}}$: tập chiều (demension); $\underline{\mathbf{C}}$: tập các khối.

Định nghĩa 1: Sơ đồ chiều và thể hiện

Sơ đồ chiều là bộ (dname, $L, \leq$), trong đó: dname $\in \underline{\mathbf{D}}$ là tên chiều; $L \in \underline{\mathbf{L}}$ là tập hữu hạn các mức trong đó có chứa một mức đặc biệt gọi là All với dom(All) = {all}; $\leq$ là quan hệ có thứ tự bộ phận trên các mức. $\leq^*$ là bao đóng truyền ứng và phản xạ của quan hệ $\leq$, với duy nhất một mức $l_{\inf}$ gọi là mức thấp nhất, duy nhất một mức cao nhất gọi là All và với mỗi mức $l \in L$ ta có $l_{\inf} \leq^* l$ và $l \leq^*$ All. Hơn nữa nếu l_a, l_b là các mức trong L và $l_a \leq^* l_b$ thì không tồn tại $l_a \leq l_b$. Nếu có quan hệ $l_1 \leq l_2$ chúng ta nói rằng l_1 cuốn trên l_2 hoặc l_2 khoan xuống l_1.

Thể hiện chiều là bộ (D, RUP) trong đó D là sơ đồ chiều và RUP là tập các hàm sao cho:
(i) Với mỗi cặp mức l_1, l_2 thỏa mãn $l_1 \leq l_2$ thì tồn tại một hàm roll-up $\mathrm{RUP}_{l_1}^{l_2} : \mathrm{dom}(l_1) \to \mathrm{dom}(l_2)$. Nếu $\mathrm{RUP}_{l_1}^{l_2}(o_1) = o_2$ chúng ta nói rằng o_1 cuốn trên o_2 hoặc o_2 khoan xuống o_1.

(ii) Chúng ta xây dựng đồ thị vô hướng ứng với quan hệ $\leq$ và tập L như sau:
Đỉnh là mức thuộc L, $l_i l_j$ là cạnh nếu $l_i, l_j \in L$ và $l_i \leq l_j$.

Cho $\Pi_1 = \langle l_1, l_2, \ldots, l_{n-1}, l_n\rangle$ là đường đi từ l_1 tới l_n và $\Pi_2 = \langle l_1, l'_2, \ldots, l'_{m-1}, l_m\rangle$ là đường từ l_1 tới l_m, thì chúng ta sẽ có:

$$\mathrm{RUP}_{l_{n-1}}^{l_n} \circ \mathrm{RUP}_{l_{n-2}}^{l_{n-1}} \circ \cdots \circ \mathrm{RUP}_{l_1}^{l_2} = \mathrm{RUP}_{l'_{m-1}}^{l_m} \circ \mathrm{RUP}_{l'_{m-2}}^{l'_{m-1}} \circ \cdots \circ \mathrm{RUP}_{l_1}^{l'_2},$$

(iii) Với ba mức l_1, l_2, l_3 sao cho $l_1 \leq l_2$ và $l_2 \leq l_3$ ta có $\mathrm{ran}(\mathrm{RUP}_{l_1}^{l_2}) \subseteq \mathrm{dom}(\mathrm{RUP}_{l_2}^{l_3})$.

Để minh họa cho định nghĩa trên chúng ta xét ví dụ sau:

Ví dụ 1. Cho sơ đồ chiều (dname, L, $\leq$), với dname = Product,
L = {ItemId, Brand, Company, Category, Corporation, Product, All}.
Quan hệ $\leq$= {ItemId $\leq$ Brand, Brand $\leq$ Company, Company $\leq$ Corporation, Corporation $\leq$ All, ItemId $\leq$ Category, Category $\leq$ Corporation}. Hình vẽ sau cho chúng ta thấy sơ đồ chiều Product, và thể hiện chiều Product.

Sơ đồ chiều Product Thể hiện chiều Product

Chú ý: Ngoại trừ các trường hợp đặc biệt, từ nay về sau chúng ta sẽ gọi chiều và thể hiện chiều là như nhau.

Chúng ta kí hiệu RUP* là tập các hàm cuốn gián tiếp và RUP là tập các hàm cuốn trực tiếp. RUP* là tập các hàm cuốn thỏa mãn điều kiện: với mỗi cặp mức $l_m, l_n \in L$, $l_m \leq^* l_n$, thì

- nếu $l_m = l_n$ thì $\mathrm{RUP^*}_{l_m}^{l_n} = \mathrm{identity}$,
- nếu $\langle l_m \ldots l_n\rangle$ là một đường đi từ l_m đến l_n trong đồ thị nêu trên thì $\mathrm{RUP^*}_{l_m}^{l_n} = \mathrm{RUP}_{l_{n-1}}^{l_n} \circ \cdots \circ \mathrm{RUP}_{l_m}^{l_{m+1}}$.

Cũng cần phải nói thêm rằng: Các hàm cuốn được định nghĩa trên cùng một mức phải có cùng miền xác định. Cụ thể nếu cho sự thể hiện chiều, với l, l', l'' sao cho $l \leq l'$ và $l \leq l''$ thì $\mathrm{dom}(\mathrm{RUP}_l^{l'}) = \mathrm{dom}(\mathrm{RUP}_l^{l''})$.

Cho một chiều với mỗi cặp mức l, l' sao cho $l \leq l'$. Ta gọi tập thể hiện của mức l là $\mathbf{instset}(l) = \mathrm{dom}(\mathrm{RUP}_l^{l'})$. Trong Ví dụ 1 với chiều Product tập thể hiện của mức ItemId là $\{i_1, i_2, i_3, i_4\}$.

Định nghĩa 2: Sơ đồ khối và thể hiện

Một sơ đồ khối (cube) là bộ $(C, Lset, m)$, trong đó $C \in \underline{\mathbf{C}}$ là tên khối, $Lset$ là tập các mức, m là độ đo của khối. Cho sơ đồ khối C với $Lset = \{l_1, l_2, \dots, l_n\} \in \underline{\mathbf{L}}$ được gọi là khối n chiều, tọa độ của C là bộ $O = \{o_1, o_2, \dots, o_n\}$ với $o_i \in \mathrm{dom}(l_i)$, $i = 1, \dots, n$. Thể hiện của khối $(C, Lset, m)$ là một hàm bộ phận, hàm này ánh xạ từ các tọa độ thuộc C tới các phần tử thuộc $\mathrm{dom}(m)$. f_C : {tập các tọa độ của C} $\rightarrow \mathrm{dom}(m)$. Ta gọi O là một điểm của C nếu f_C được xác định. Do vậy, chúng ta có thể kí hiệu thể hiện của khối C là $C = \{O, f_C(O)) \,|\, O$ điểm của $C\}$.

2.2. Cơ sở dữ liệu đa chiều

Định nghĩa 3: Sơ đồ cơ sở dữ liệu đa chiều và thể hiện

Sơ đồ cơ sở dữ liệu đa chiều là một cặp $MS = (\mathbf{DC}, \mathbf{CS})$, trong đó **DS** là tập các sơ đồ **chiều** và **CS** là tập các sơ đồ **khối**. Thể hiện của cơ sở dữ liệu đa chiều là bộ $MD = (\mathbf{D}, \mathbf{C})$, trong đó **D** là tập các **thể hiện chiều** và **C** là tập các **thể hiện khối**.

3. ĐẠI SỐ KHỐI

Trước khi định nghĩa các phép hợp, phép trừ, phép giao chúng ta đưa ra khái niệm hai khối khả hợp.

Định nghĩa 4. Hai khối $C_1(\mathrm{Lset}_1, m_1)$ và $C_2(\mathrm{Lset}_2, m_2)$ được gọi là khả hợp nếu $\mathrm{Lset}_1 \equiv \mathrm{Lset}_2$ và $m_1 \equiv m_2$.

3.1. Phép hợp

Cho hai khối $C_1(\mathrm{Lset}_1, m_1)$ và $C_2(\mathrm{Lset}_2, m_2)$ là khả hợp. Phép hợp của hai khối C_1 và C_2, kí hiệu là $C = \mathrm{UN}(C_1, C_2, f_U)$ với $C(\mathrm{Lset}, m)$, trong đó $\mathrm{Lset} \equiv \mathrm{Lset}_1 \equiv \mathrm{Lset}_2$, $m \equiv m_1 \equiv m_2$ và $f_U : 2^m \rightarrow m$ là hàm tập hợp được định nghĩa như sau:

$$C = \{(O, f_C(O))\} = \begin{cases} (O, f_{C_1}(O)) \text{ nếu } O \in C_1 \text{ và } O \notin C_2 \\ (O, f_{C_2}(O)) \text{ nếu } O \in C_2 \text{ và } O \notin C_1 \\ (O, f_U(f_{C_1}(O), f_{C_2}(O)) \text{ nếu } O \in C_1 \text{ và } O \in C_2 \end{cases}$$

3.2. Phép trừ

Cho hai khối $C_1(\mathrm{Lset}_1, m_1)$ và $C_2(\mathrm{Lset}_2, m_2)$ là khả hợp. Phép hiệu của hai khối C_1 và C_2 kí hiệu là $C = \mathrm{DF}(C_1, C_2, f_D)$ với $C(\mathrm{Lset}, m)$ trong đó $\mathrm{Lset} \equiv \mathrm{Lset}_1 \equiv \mathrm{Lset}_2$, $m \equiv m_1 \equiv m_2$ và $f_D : 2^m \rightarrow m$ là hàm tập hợp được định nghĩa như sau:

$$C = \{(O, f_C(O))\} = \begin{cases} (O, f_{C_1}(O)) \text{ nếu } O \in C_1 \text{ và } O \notin C_2 \\ (O, f_D(f_{C_1}(O), f_{C_2}(O)) \text{ nếu } O \in C_1 \text{ và } O \in C_2 \\ f_C(O) \text{ không xác định trong các trường hợp khác} \end{cases}$$

3.3. Phép giao

Cho hai khối $C_1(\mathrm{Lset}_1, m_1)$ và $C_2(\mathrm{Lset}_2, m_2)$ là khả hợp. Phép giao của hai khối C_1 và C_2 kí hiệu là $C = \mathrm{IN}(C_1, C_2, f_I)$ với $C(\mathrm{Lset}, m)$, trong đó $\mathrm{Lset} \equiv \mathrm{Lset}_1 \equiv \mathrm{Lset}_2$, $m \equiv m_1 \equiv m_2$ và $f_I : 2^m \rightarrow m$ là hàm tập hợp được định nghĩa như sau:

$C = \{O, f_C(O))\} = \{(O, f_I(f_{C_1}(O_1), f_{C_2}(O_2)) \,|\, O \in C_1,\, O \in C_2\}$.

Nhận xét: Do khối có đặc trưng là độ đo, vì vậy, trong các phép toán hợp, trừ, giao chúng ta phải đưa vào một hàm tập hợp như là một tham số của các phép toán. Do vậy, một cách tổng quát, một số tính chất của các phép toán tập hợp truyền thống có thể không còn nữa. Tuy nhiên chúng ta có thể giữ nguyên được các tính chất này nếu chúng ta chọn các hàm tập hợp một cách thích hợp.

3.4. Phép thu gọn khối

Cho khối $C(\mathrm{Lset}, m)$. Phép thu gọn khối của khối C kí hiệu $C'(\mathrm{Lset}', m) = \mathrm{RED}(C(\mathrm{Lset}, m))$, trong đó $\mathrm{Lset}' = \mathrm{Lset} \setminus \{l \in \mathrm{Lset} \mid \exists l' \in \mathrm{Lset}$ mà $l' \leq^* l$ và $l \neq l'$ hoặc $l = \mathrm{All}\}$ được xác định như sau:

$C' = \{(O', f_{C'}(O')) \mid \exists O$ điểm của C sao cho $O' = O.\mathrm{Lset}'$, $f_{C'}(O') = f_C(O)\}$ với $O.\mathrm{Lset}'$ - kí hiệu tọa độ của O trên tập các mức Lset'.

3.5. Phép chọn

Cho khối $C(\mathrm{Lset}, m)$. Phép chọn trên khối C sẽ cho chúng ta một khối mới có cùng cấu trúc với khối đã cho và thỏa mãn điều kiện chọn F, trong đó biểu thức chọn là tổ hợp logic của các toán hạng, mỗi toán hạng là một phép so sánh có dạng sau: $l\theta c$ với $c \in \mathrm{dom}(l)$ hoặc $\mathrm{RUP}_l^{l'}(l)\theta c'$ với $c' \in \mathrm{dom}(l')$ hoặc $l\theta l'$ với $\mathrm{dom}(l) = \mathrm{dom}(l')$. $\theta \in \{<, >, <=, >=, \neq, =\}$. Phép chọn trên khối C với điều kiện chọn F kí hiệu $C' = \mathrm{SL}_F(C)$ được định nghĩa như sau:

$C' = \{(O, f_C(O)) \mid F(O)$ đúng$\}$.

3.6. Phép cuốn (ROLL-UP)

Cho khối $C(\mathrm{Lset}, m)$ trong đó $\mathrm{Lset} = \{l_1, l_2, \ldots, l_n\}$, đặt $\mathrm{Lset}' = \{l'_{11}, \ldots, l'_{1m_1}, \ldots, l'_{n1}, \ldots, l'_{nm_n}\}$ với $l_i \leq^* l'_{ij}$, $i = 1, \ldots, n$, $j = 1, \ldots, m_i$.
$O' = \{o'_{11}, \ldots, o'_{1m_1}, \ldots, o'_{n1}, \ldots, o'_{nm_n}\}$ là một điểm của $C'(\mathrm{Lset}', m')$ nếu tồn tại: $O = (o_1, o_2, \ldots, o_n)$ là một điểm của C sao cho: $o'_{ij} = \mathrm{RUP}{*}_{l_i}^{l'_{ij}}(o_i)$ với $i = 1, \ldots, n,\ j = 1, \ldots, m_i$.

Chúng ta kí hiệu $O' = \mathrm{RUP}{*}_{\mathrm{Lset}}^{\mathrm{Lset}'}(O)$ và định nghĩa $f_M : 2^m \to m'$ là hàm tập hợp. Phép cuốn của khối C từ các tập mức Lset lên Lset' kí hiệu $C' = \mathrm{RUP}_{\mathrm{Lset}}^{\mathrm{Lset}'}(C, f_M)$ được định nghĩa như sau:

$C' = \mathrm{RUP}{*}_{\mathrm{Lset}}^{\mathrm{Lset}'}(C, f_M)$
$= \{(O', f_{C'}(O') \mid O'$ là một điểm của C' và $f_{C'}(O') = f_M(\{f_C(O) \mid O' = \mathrm{RUP}{*}_{\mathrm{Lset}}^{\mathrm{Lset}'}(O)\})\}$.

Cuối cùng chúng ta thực hiện phép thu gọn với khối $C' = \mathrm{RED}(C')$.

3.7. Phép kết nối

Định nghĩa 5. Hai mức l_a, l_b là khả tương thích nếu $l_a \leq^* l_b$ hoặc $l_b \leq^* l_a$.

Cho hai khối $C_1(\mathrm{Lset}_1, m_1)$ và $C_2(\mathrm{Lset}_2, m_2)$. Điều kiện kết nối F của hai khối là phép hội của các toán hạng, trong đó mỗi toán hạng là một phép so sánh giữa hai mức khả tương thích của hai khối có dạng

$l_a\theta l_b$, $l_a \in \mathrm{Lset}_1$ của C_1, $l_b \in \mathrm{Lset}_2$ của C_2 và $\theta \in \{<, >, <=, >=, \neq, =\}$.

Phép kết nối của hai khối C_1 và C_2 với điều kiện kết nối F kí hiệu

$C_3(\mathrm{Lset}_3, m_3) = \mathrm{JN}_F(C_1(\mathrm{Lset}_1, m_1), C_2(\mathrm{Lset}_2, m_2), f)$ trong đó $\mathrm{Lset}_3 = \mathrm{Lset}_1 \cup \mathrm{Lset}_2$,

$f : m_1 \times m_2 \to m_3$ được định nghĩa như sau:

$C_3 = \{(O_3, f_{C_3}(O_3)) \mid O_3 = (O_1, O_2)$, O_1 - điểm của C_1, O_2 - điểm của C_2, $F(O_3)$ đúng và $f_{C_3}(O_3) = f(f_{C_1}(O_1), f_{C_2}(O_2))\}$.

Cuối cùng chúng ta thực hiện phép thu gọn khối: $C_3 = \mathrm{RED}(C_3)$.

3.8. Biểu diễn các yêu cầu

Để minh họa cách biểu diễn các câu hỏi phân tích điển hình của OLAP bằng đại số khối, trong phần này để trực quan hơn chúng ta xem xét một ví dụ khác. Giả sử cho một CSDLĐC gồm hai khối sau:

Sales($\{\mathrm{Day}, \mathrm{Item}, \mathrm{Store}\}$, Numeric) và Price_list($\{\mathrm{Item}, \mathrm{Month}\}$, Numeric)

với các chiều **Time**, **Store**, **Product**, và có sơ đồ như sau:

Sơ đồ chiều Time Sơ đồ chiều Store Sơ đồ chiều Product

Câu hỏi 1: Biểu diễn các thông tin về doanh thu với các kho ở Rome

$$\mathrm{SL}\;\mathrm{RUP}^{\mathrm{City}}_{\mathrm{Store}}\ _{(\mathrm{Store})=\mathrm{Rome}}(\mathrm{Sales})$$

Câu hỏi 2: Biểu diễn doanh thu theo ngày đối với mỗi kho và mỗi mặt hàng

$$\mathrm{TG{=}JN}\;\mathrm{RUP}^{\mathrm{Month}}_{\mathrm{Day}}\ _{(\mathrm{Day})=\mathrm{Month}\ \text{và}\ \mathrm{Sales\_Item}=\mathrm{Price\_list\_Item}}(\mathrm{Sales}, \mathrm{Price}, \mathrm{List},\ f)$$

Hàm tập hợp ở đây là hàm tích

KQ=RED(TG).

Câu hỏi 3: Biểu diễn doanh thu theo tháng đối với mỗi mặt hàng và mỗi kho

$$\mathrm{RUP}^{\mathrm{Month},\,\mathrm{Item},\,\mathrm{Store}}_{\mathrm{Day},\,\mathrm{Item},\,\mathrm{Store}}(\mathrm{Sales}, \mathrm{Sum})$$

4. SO SÁNH VÀ ĐÁNH GIÁ

4.1. Đánh giá về khả năng biểu diễn

Chúng ta có thể thấy tính tương tự của các phép toán được đề nghị đối với đại số khối với các phép toán của đại số quan hệ. Trong quá trình phát triển, các phép toán này có thể được dịch sang SQL và chúng tôi hy vọng có thể tạo được một đường dẫn nhanh chóng với việc cung cấp các khả năng của OLAP trên đỉnh các hệ CSDL quan hệ. Hơn nữa nó cũng chỉ ra được định hướng trong vấn đề tối ưu hóa và tổ chức lưu trữ đối với CSDL đa chiều.

Cuối cùng chúng ta cũng có thể nhận thấy sự vắng mặt của số phép toán trong đại số quan hệ như phép chiếu, phép tích đề các.

Phép chiếu: Phép chiếu của khối $C(\mathrm{Lset}, m)$ trên tập các mức $\mathrm{Lset}' \subseteq \mathrm{Lset}$ được thực hiện thông qua phép cuốn với Lset' là tập các mức với mức All được gán cho các mức $l \in \mathrm{Lset}$ mà $l \notin \mathrm{Lset}'$. Sau đó áp dụng phép thu gọn khối để xóa đi các mức All.

Phép tích Đêcac: Phép tích Đêcac là trường hợp đặc biệt của phép kết nối khi hai khối không có hai mức nào là khả tương thích, khi đó điều kiện kết nối $F = \emptyset$.

Với những nhận xét trên dễ dàng nhận thấy rằng đại số khối của chúng tôi ít nhất là mạnh bằng đại số quan hệ.

4.2. So sánh với các cách tiếp cận khác

Trong [5], một MHDLĐC và một đại số trên các khối được trình bày. Ở đây, CSDLĐC được biểu diễn trong một hoặc nhiều khối, giá trị các ô trong khối có thể là một n-bộ hoặc một giá trị thuộc tập chứa 0 và 1. Các chiều không có cấu trúc phân cấp một cách tường minh và không phân biệt giữa cấu trúc và nội dung rõ ràng.

Trong [3], một MHDLĐC và một ngôn ngữ hỏi khai báo dựa trên các phép tính toán tân từ được đề nghị. Trong cách tiếp cận này, các độ đo phức tạp không được mô hình hóa trực tiếp và thường đòi hỏi các cách xử lý khác nhau đối với mỗi độ đo phức tạp cụ thể.

Trong [2] và [4], một MHDLĐC được đưa ra với các ứng dụng OLAP. Do vậy, cả hai cách tiếp cận này đều phụ thuộc vào những chi tiết về lưu trữ dữ liệu đa chiều theo mô hình quan hệ dưới dạng các bảng. Việc biểu diễn các yêu cầu khá phức tạp và khó khăn với quá nhiều phép toán.

Tóm lại, so sánh với các cách tiếp cận khác, MHDLĐC của chúng tôi là độc lập đối với một sự cài đặt đặc biệt nào đó, cho phép đối xử một cách đối xứng với các chiều và các độ đo, cho phép phân biệt giữa cấu trúc và nội dung và hỗ trợ sự phân cấp phức tạp trên các chiều. Đại số khối của chúng tôi có khả năng biểu diễn tương đương với các đại số và ngôn ngữ hỏi được đề nghị trên các MHDLĐC đã được xây dựng. Thêm vào đó, kết quả của chúng tôi có thể mở ra một số vấn đề cần nghiên cứu trong tương lai như tính chất của các phép toán của đại số khối và các kỹ thuật cài đặt đại số khối và đại số khối có thể cài đặt trên đỉnh của một hệ quản trị CSDL quan hệ như thế nào... Cuối cùng, vấn đề tối ưu hóa câu hỏi có thể được thực hiện một cách hiệu quả bởi việc sử dụng một ngôn ngữ hỏi đại số thủ tục với các nghiên cứu về các phép biến đổi tương đương của đại số đó. Do vậy, chúng tôi hy vọng có thể tiếp tục phát triển các kết quả đạt được với việc nghiên cứu vấn đề tối ưu hóa và các kỹ thuật tổ chức lưu trữ CSDLĐC.

TÀI LIỆU THAM KHẢO

[1] A. Data, H. Thomas, *A Conceptual Model and an Algebra for On-Line Analytical Processing in Data Warehouses*, WITS, 1997.

[2] C. Li and X. S. Wang, A data model for On-line analytical processing, *Pro. Conf. on Information and Knowledge Management*, Novembre 1996.

[3] L. Cabibbo, R. Torlone, Querying multidimensional databases, *Pro. of the 6^{th} DBPL*, 1997.

[4] M. Gyssens, L. V. S. Lakshmann, A foundation for multidimensional databases, *Pro. of the VLDB*, Athens, Greece, 1997.

[5] R. Agraval, A. Gupta, S. Sarawagi, Modelling multidimensional databases, *Pro. of the ICDE*, 1997.

Nhận bài ngày 28 - 10 - 2001

Khoa Công nghệ thông tin
Trường Đại học Bách khoa Hà Nội