

1) Giả sử một kho dữ liệu bao gồm 3 chiều: thời gian, bác sỹ và bệnh nhân, và hai độ đo là số đếm và chi phí với chi phí là tổng số tiền mà một bệnh nhân phải trả cho một bác sỹ trong một lần khám bệnh.

(a) Hãy liệt kê ba loại lược đồ dữ liệu có thể được sử dụng để mô hình hóa kho dữ liệu và sử dụng một trong số chúng để vẽ lược đồ dữ liệu cho kho dữ liệu mô tả ở trên.

(b) Xuất phát từ khối cơ bản [ngày, bác sỹ, bệnh nhân], những thao tác OLAP nào cần được thực hiện để liệt kê tổng số tiền kiểm được của mỗi bác sỹ trong năm 2012?

(c) Viết các câu lệnh DMQL để định nghĩa lược đồ dữ liệu trong câu (a) và câu lệnh SQL để tính tổng tiền trong câu (b)

2) Một cơ sở dữ liệu có 5 giao dịch như dưới đây. Giả sử min sup = 60% và min conf = 80%.

Mã giao dịch	Mặt hàng mua
--------------	--------------

T100	A, B, C, D, E
------	---------------

T200	B, E, A, C, H
------	---------------

T300	P, A, G, E
------	------------

T400	U, R, B, A, N, A
------	------------------

T500	G, A, U, D, I
------	---------------

a) Hãy tìm tất cả những tập mặt hàng thường xuyên sử dụng thuật toán Apriori và FP-growth. So sánh tính hiệu quả của hai phương pháp khai phá trên.

b) Hãy liệt kê tất cả những luật kết hợp (với độ hỗ trợ s và độ tin cậy c) tuân thủ theo siêu luật được mô tả dưới đây, trong đó X là một biến đại diện cho khách hàng và mh_i thể hiện các biến đại diện cho các mặt hàng (ví dụ "A", "B", v.v...): cho tất cả x trong giao dịch; mua(X, mh1) mua(X, mh2) [s, c]