Cho quan hệ PROJ(PNo, PName, Budget, Loc). Giả sử rằng các ứng dụng sử dụng tập câu truy vấn Q(q1, q2, q3, q4) truy cập đến quan hệ PROJ như sau:

           q1: Tìm ngân sách của dự án theo mã số dự án.

           q2: Tìm tên và ngân sách của tất cả các dự án.

           q3: Tìm tên của dự án theo vị trí.

           q4: Tìm tổng ngân sách dự án tại mỗi vị trí.

Dựa vào q1, q2, q3, q4, chúng ta có thể định nghĩa được giá trị sử dụng thuộc tính. Gọi A1=PNo, A2=PName, A3=Budget, A4=Loc và cho hình ảnh ma trận Use và ma trận tần số truy cập Acc bên.

Yêu cầu:

       a)Tính matrix Use.

       b)Giả sử cho số đo tần số truy cập ứng dụng bên dưới. Hãy xây dựng matrix tần số truy cập accl(qk) cho 4 câu truy vấn và 3 site như sau:

|  |  |  |
| --- | --- | --- |
| acc1(q1) = 15 | acc2(q1) = 20 | acc3(q1) = 10 |
| acc1(q2) = 5 | acc2(q2) = 0 | acc3(q2) = 0 |
| acc1(q3) = 25 | acc2(q3) = 25 | acc3(q3) = 25 |
| acc1(q4) = 3 | acc2(q4) = 0 | acc3(q4) = 0 |

a)Tính ma trận AA

b)Tính ma trận CA

c) Tìm hai mảnh dọc tối ưu

***Lưu ý: Anh/Chị có thể làm đáp án bằng file word, sau đó tải file word lên.***

Bài làm

a)

Matrix Use:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | A1 | A2 | A3 | A4 |
| Q1 | 1 | 0 | 1 | 0 |
| Q2 | 0 | 1 | 1 | 0 |
| Q3 | 0 | 1 | 0 | 1 |
| Q4 | 0 | 0 | 1 | 1 |

Matric Acc

|  |  |  |  |
| --- | --- | --- | --- |
|  | S1 | S2 | S3 |
| Q1 | 15 | 20 | 10 |
| Q2 | 5 | 0 | 0 |
| Q3 | 25 | 25 | 25 |
| Q4 | 3 | 0 | 0 |

Tổng tần số dựa theo câu truy vấn:

|  |
| --- |
| SUM |
| 45 |
| 5 |
| 75 |
| 3 |

Ma trận AA:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | A1 | A2 | A3 | A4 |
| A1 | 45 | 0 | 45 | 0 |
| A2 | 0 | 80 | 5 | 75 |
| A3 | 45 | 5 | 53 | 3 |
| A4 | 0 | 75 | 3 | 78 |

b)

Từ Matrix AA. Chọn và thêm trước 2 cột A1 và A2 vào ma trận CA. Ta được CA hiện tại như sau:

Đặt A3 : có 3 cách đặt

Đặt A3 trước A1 :

cont(\_, A3 , A1 ) = 2bond(\_, A3 ) + 2bond(A3 , A1 ) - 2bond(\_, A1 ) = 2\*0+2\*4410-2\*0 = 8820 bond(\_, A3 ) = 0 bond(A3 , A1 ) = 45\*45 + 5\*0 + 53\*45 + 3\*0 = 4410

bond(\_, A1 ) = 0

bond(A3 , A1 ) = 45\*45 + 5\*0 + 53\*45 + 3\*0 = 4410

bond(\_, A1 ) = 0

Đặt A3 giữa A1 và A2 :

cont(A1 , A3 , A2 ) = 2bond(A1 , A3 ) + 2bond(A3 , A2 ) - 2bond(A1 , A2 ) = 2\*4410+2\*890-2\*225 = 10150 bond(A1 , A3 ) = 45\*45 + 0\*5 + 45\*53 + 0\*3 = 4410

bond(A3 , A2 ) = 45\*0 + 5\*80 + 53\*5 + 3\*75 = 890

bond(A1 , A2 ) = 45\*0 + 0\*80 + 45\*5 + 0\*75 = 225

Đặt A3 sau A2 :

cont(A2 , A3 , \_) = 2bond(A2 , A3 ) + 2bond(A3 , \_) - 2bond(A2 , \_) = 2\*890+2\*0-2\*0 = 1780

bond(A2 , A3 ) = 0\*45 + 80\*5 + 5\*53 + 75\*3 = 890

bond(A3 , \_) = 0

bond(A2 , \_) = 0

Vì cách đặt thứ 2 có contribution lớn nhất (10150) nên ta chọn đặt "A3" theo cách này. Sau khi đặt "A3" vào CA ta được:

|  |  |  |  |
| --- | --- | --- | --- |
| CA | A1 | A3 | A2 |
| A1 |  |  |  |
| A3 |  |  |  |
| A2 |  |  |  |

Đặt A4 : có 4 cách đặt

Đặt A4 trước A1 :

cont(\_, A4 , A1 ) = 2bond(\_, A4 ) + 2bond(A4 , A1 ) - 2bond(\_, A1 ) = 2\*0+2\*135-2\*0 = 270

bond(\_, A4 ) = 0

bond(A4 , A1 ) = 45\*0 + 0\*75+45\*3+0\*78 = 135

bond(\_, A1 ) = 0

Đặt A4 giữa A1 và A3 :

cont(A1 ,A4 ,A3 ) = 2\*bond(A1 , A4 ) + 2\*bond(A4 , A3 ) - 2\*bond(A1 , A3 ) = 2\*135+2\*768-2\*4410=-7014

bond(A1 , A4 ) = 45\*0 + 0\*75 + 45\*3 + 0\*78 = 135

bond(A4 , A3 ) = 0\*45 + 75\*5 + 3\*53 + 78\*3 = 768

bond(A1 , A3 ) = 45\*45 + 0\*5 + 45\*53 + 0\*3 = 4410

Đặt A4 giữa A3 và A2 :

cont(A3 ,A4 ,A2 ) = 2\*bond(A3 , A4 ) + 2\*bond(A4 , A2 ) - 2\*bond(A3 , A2 ) = 2\* 768+2\* 11865-2\*890= 23486

bond(A3 , A4 ) = 768

bond(A4 , A2 ) = 11865

bond(A3 , A2 ) = 890

Đặt A4 sau A2 :

cont(A2 ,A4 ,\_) = 2\*bond(A2 , A4 ) + 2\*bond(A4 , \_) - 2\*bond(A2 , \_) = 2\*11865 + 2\*0 - 2\*0= 23730

bond(A4 , A2 ) = 11865

bond(A4 , \_) = 0

bond(A2 , \_) = 0

Vì cách đặt thứ 4 có contribution lớn nhất (23730) nên ta chọn đặt "A4" theo cách này. Sau khi đặt "A4" vào CA ta được:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| CA | A1 | A3 | A2 | A4 |
| A1 | 45 | 45 | 0 | 0 |
| A3 | 45 | 53 | 5 | 3 |
| A2 | 0 | 5 | 80 | 75 |
| A4 | 0 | 3 | 75 | 78 |

c)

Chọn điểm cắt tại bên trái và phía trên của dòng "A2" và cột "A2" và ta tiến hành tính z:

TA = { A1 , A3 , A2 }

BA = { A4 }

TQ = { q1 , q2 }

BQ = { φ }

OQ = { q3 , q4 }

CTQ = acc1 (q1 ) + acc2 (q1 ) + acc3 (q1 ) + acc1 (q2 ) + acc2 (q2 ) + acc3 (q2 ) = 15 + 20 + 10 + 5+ 0 + 0 = 50

CBQ = 0

OQ = acc1 (q3 ) + acc2 (q3 ) + acc3 (q3 ) + acc1 (q4 ) + acc2 (q4 ) + acc3 (q4 ) = 25 + 25 + 25 + 3 + 0 + 0 = 78

z = (CTQ \* CBQ) - COQ2 = (50 \* 0) - 782 = -6084

Chọn điểm cắt tại bên trái và phía dưới của cột "A3" và dòng "A3" và ta tiến hành tính z:

TA = { A1 , A3 }

BA = { A2 , A4 }

TQ = { q1 } BQ = { q3 }

OQ = { q2 , q4 }

CTQ = acc1 (q1 ) + acc2 (q1 ) + acc3 (q1 ) = 15 + 20 + 10 =45

CBQ = acc1 (q3 ) + acc2 (q3 ) + acc3 (q3 ) = 25 + 25 + 25 = 75

OQ = acc1 (q2 ) + acc2 (q2 ) + acc3 (q2 ) + acc1 (q4 ) + acc2 (q4 ) + acc3 (q4 ) = 5 + 0 + 0 + 3 + 0 + 0 = 8

z = (CTQ \* CBQ) - COQ2 = (45 \* 75) - 82 = 3311

Chọn điểm cắt tại bên trái và phía dưới của cột "A1" và dòng "A1" và ta tiến hành tính z:

TA = { A1 }

BA = { A3 , A2 , A4 }

TQ = {φ}

BQ = { q2 , q3 , q4 }

OQ = { q1 }

CTQ = 0

CBQ = acc1 (q2 ) + acc2 (q2 ) + acc3 (q2 ) + acc1 (q3 ) + acc2 (q3 ) + acc3 (q3 ) + acc1 (q4 ) + acc2 (q4 ) + acc3 (q4 ) = 5 + 0 + 0 + 25 + 25 +25 + 3 + 0 + 0 = 83

OQ = acc1 (q1 ) + acc2 (q1 ) + acc3 (q1 ) = 15 + 20 + 10 = 45

z = (CTQ \* CBQ) - COQ2 = (0 \* 83) - 452 = -2025

Ta chọn điểm cắt z có chi phí lớn nhất là: 3311.

Từ điểm cắt này ta tiến hành phân mảnh. Kết quả cuối cùng ta được 2 mảnh dọc tối ưu đã được bổ sung thêm thuộc tính khóa "A1" như sau:

PROJ1 = { A1 , A3 } - PROJ1 (PNo, Budget)

PROJ2 = { A1 , A2 , A4 } - PROJ2(PNo, PName, Loc)