

Cơ Sở Dữ Liệu Phân Tán

Bài Tập Tối Ưu Hóa Truy Vấn Trong Phân Tán - IS211.M21

Nguyễn Hồ Duy Tri, Nguyễn Thị Kim Yến
Sinh viên: Phạm Đức Thế - 19522253

Thứ 2, ngày 20 tháng 06 năm 2022

Bài Tập Chương 5: Tối Ưu Hóa Truy Vấn Trong Phân Tán

Bài Tập 1

Marie Stopes International tại Việt Nam (MSIVN) hiện đang vận hành 4 phòng khám đa khoa tại các thành phố: Hồ Chí Minh, Hà Nội, Hải Phòng và Đà Nẵng. Cho biết lược đồ CSDL quan hệ toàn cục của phòng khám MSIVN như sau:

PhongKham(MaPK, TenPK, ThanhPho, DienThoai)
BenhNhan(MaBN, TenBN, NgaySinh, DiaChi, DienThoai, GioiTinh)
BacSy(MaBS, TenBS, TenDangNhap, MatKhau)
KhamBenh(MaKB, MaBN, YeuCauKham, NgayKham, KetLuan, MaBS, MaPK)
ThuPhi(MaTP, MaKB, MaDV, SoLuong, ThanhTien)
DichVu(MaDV, TenDV, DonGia)

Quan hệ **PhongKham** phân mảnh ngang chính theo thành phố:

PhongKham1 = $\sigma_{\text{ThanhPho} = \text{"HCM"}} \text{PhongKham}$
PhongKham2 = $\sigma_{\text{ThanhPho} = \text{"Hà Nội"}} \text{PhongKham}$
PhongKham3 = $\sigma_{\text{ThanhPho} = \text{"Hải Phòng"}} \text{PhongKham}$
PhongKham4 = $\sigma_{\text{ThanhPho} = \text{"Đà Nẵng"}} \text{PhongKham}$

Quan hệ **KhamBenh**, **ThuPhi** phân mảnh ngang dẫn xuất như sau:

KhamBenh1 = KhamBenh \bowtie_{MaPK} PhongKham1
KhamBenh2 = KhamBenh \bowtie_{MaPK} PhongKham2
KhamBenh3 = KhamBenh \bowtie_{MaPK} PhongKham3
KhamBenh4 = KhamBenh \bowtie_{MaPK} PhongKham4
ThuPhi1 = ThuPhi \bowtie_{MaKB} KhamBenh1
ThuPhi2 = ThuPhi \bowtie_{MaKB} KhamBenh2
ThuPhi3 = ThuPhi \bowtie_{MaKB} KhamBenh3
ThuPhi4 = ThuPhi \bowtie_{MaKB} KhamBenh4

Quan hệ **BenhNhan** phân mảnh hỗn hợp như sau:

BenhNhan1_A = $\Pi_{\text{MaBN}, \text{DienThoai}}(\text{BenhNhan} \bowtie_{\text{MaBN}} \text{KhamBenh1})$
BenhNhan1_B = $\Pi_{\text{MaBN}, \text{TenBN}, \text{NgaySinh}, \text{DiaChi}, \text{GioiTinh}}(\text{BenhNhan} \bowtie_{\text{MaBN}} \text{KhamBenh1})$
BenhNhan2_A = $\Pi_{\text{MaBN}, \text{DienThoai}}(\text{BenhNhan} \bowtie_{\text{MaBN}} \text{KhamBenh2})$
BenhNhan2_B = $\Pi_{\text{MaBN}, \text{TenBN}, \text{NgaySinh}, \text{DiaChi}, \text{GioiTinh}}(\text{BenhNhan} \bowtie_{\text{MaBN}} \text{KhamBenh2})$
BenhNhan3 = BenhNhan \bowtie_{MaBN} KhamBenh3
BenhNhan4 = BenhNhan \bowtie_{MaBN} KhamBenh4

Quan hệ **BacSy**, **DichVu** được nhân bản ở tất cả các phòng khám.

Yêu cầu:

1. Hãy đơn giản hóa câu truy vấn **Q1** bằng cách sử dụng qui tắc lũy đẳng:

```
Q1: SELECT MaBN, TenBN, KetLuan
FROM BenhNhan BN, KhamBenh KB
WHERE BN.MaBN = KB.MaBN
      AND (MaBN = 1 AND (NgayKham = '01/12/2015' OR MaBN = 2))
      OR NgayKham = '01/12/2015'
```

Lời giải.

Đặt:

- p1: BN.MaBN = KB.MaBN
- p2: MaBN = 1
- p3: NgayKham = '01/12/2015'
- p4: MaBN = 2

Viết lại mệnh đề WHERE trong câu truy vấn **Q1**, ta có:

$p1 \wedge (p2 \wedge (p3 \vee p4)) \vee p3$
 $\Leftrightarrow p1 \wedge ((p2 \wedge p3) \vee (p2 \wedge p4)) \vee p3$ (áp dụng luật phân phối)
 $\Leftrightarrow p1 \wedge ((p2 \wedge p3) \vee False) \vee p3$
 $\Leftrightarrow p1 \wedge (p2 \wedge p3) \vee p3$ (áp dụng luật đồng nhất)
 $\Leftrightarrow (p1 \vee p3) \wedge ((p2 \vee p3) \wedge p3)$ (áp dụng luật phân phối)
 $\Leftrightarrow (p1 \vee p3) \wedge p3$ (áp dụng luật hấp thụ)
 $\Leftrightarrow p3$ (áp dụng luật hấp thụ)

Vậy câu truy vấn **Q1** được biến đổi thành:

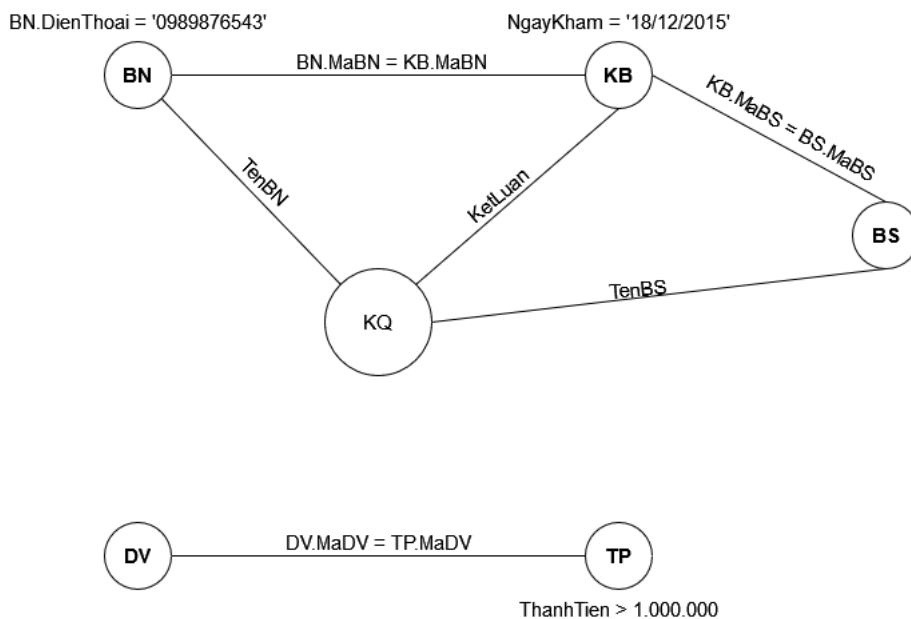
```
Q1: SELECT MaBN, TenBN, KetLuan
FROM BenhNhan BN, KhamBenh KB
WHERE NgayKham = '01/12/2015'
```

2. Vẽ đồ thị truy vấn (query graph) cho câu truy vấn **Q2**. Dựa vào đồ thị, hãy cho biết câu truy vấn này có sai ngữ nghĩa không? Nếu sai, hãy viết lại câu truy vấn **Q2** cho đúng ngữ nghĩa.

```
Q2: SELECT TenBN, TenBS, KetLuan
FROM BenhNhan BN, KhamBenh KB, DichVu DV, BacSy BS, ThuPhi TP
WHERE BN.MaBN = KB.MaBN
      AND DV.MaDV = TP.MaDV
      AND KB.MaBS = BS.MaBS
      AND ThanhTien > 1.000.000
      AND NgayKham = '18/12/2015'
      AND BN.DienThoai = '0989876543'
```

Lời giải.

Đồ thị truy vấn của câu truy vấn **Q2**:

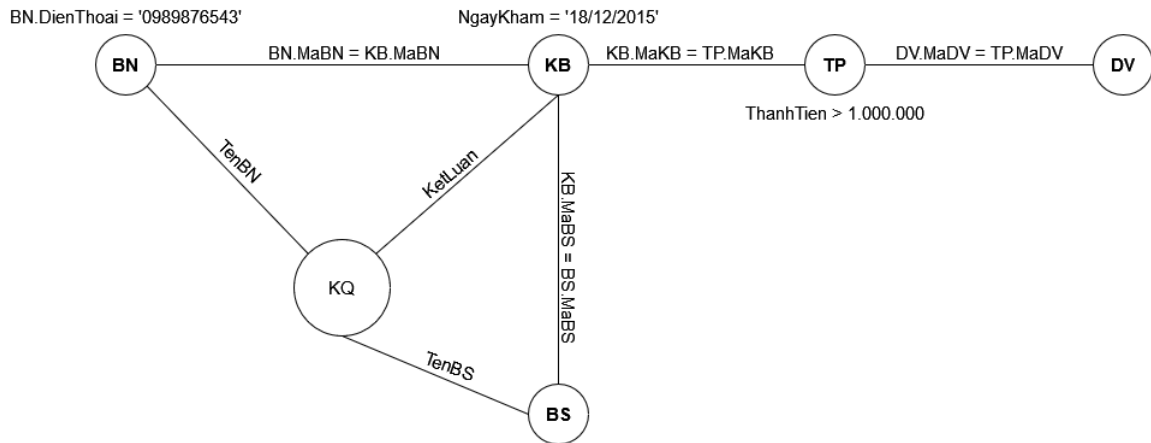


Ta thấy đồ thị truy vấn không liên thông.

\Rightarrow Câu truy vấn **Q2** sai ngữ nghĩa.

Viết lại câu truy vấn **Q2**:

```
Q2: SELECT TenBN, TenBS, KetLuan
FROM BenhNhan BN, KhamBenh KB, DichVu DV, BacSy BS, ThuPhi TP
WHERE BN.MaBN = KB.MaBN
      AND DV.MaDV = TP.MaDV
      AND KB.MaBS = BS.MaBS
      AND KB.MaKB = TP.MaKB
      AND ThanhTien > 1.000.000
      AND NgayKham = '18/12/2015'
      AND BN.DienThoai = '0989876543'
```



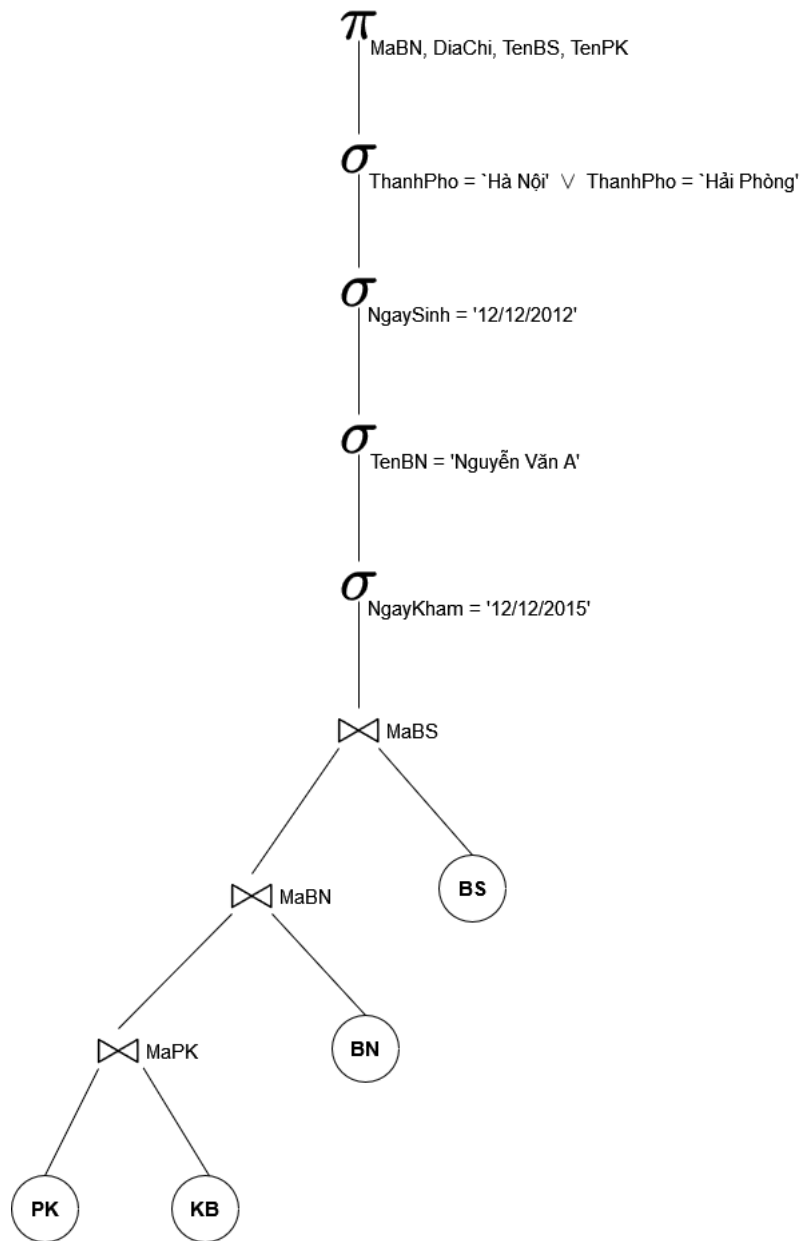
3. Cho câu truy vấn **Q3** như sau:

```
Q3: SELECT MaBN, DiaChi, TenBS, TenPK
FROM PhongKham PK, BenhNhan BN, KhamBenh KB, BacSy BS
WHERE PK.MaPK = KB.MaPK
      AND BN.MaBN = KB.MaBN
      AND KB.MaBS = BS.MaBS
      AND NgayKham = '12/12/2015'
      AND TenBN = 'Nguyễn Văn A'
      AND NgaySinh = '12/12/2012'
      AND (ThanhPho = 'Hà Nội' OR ThanhPho = 'Hải Phòng')
```

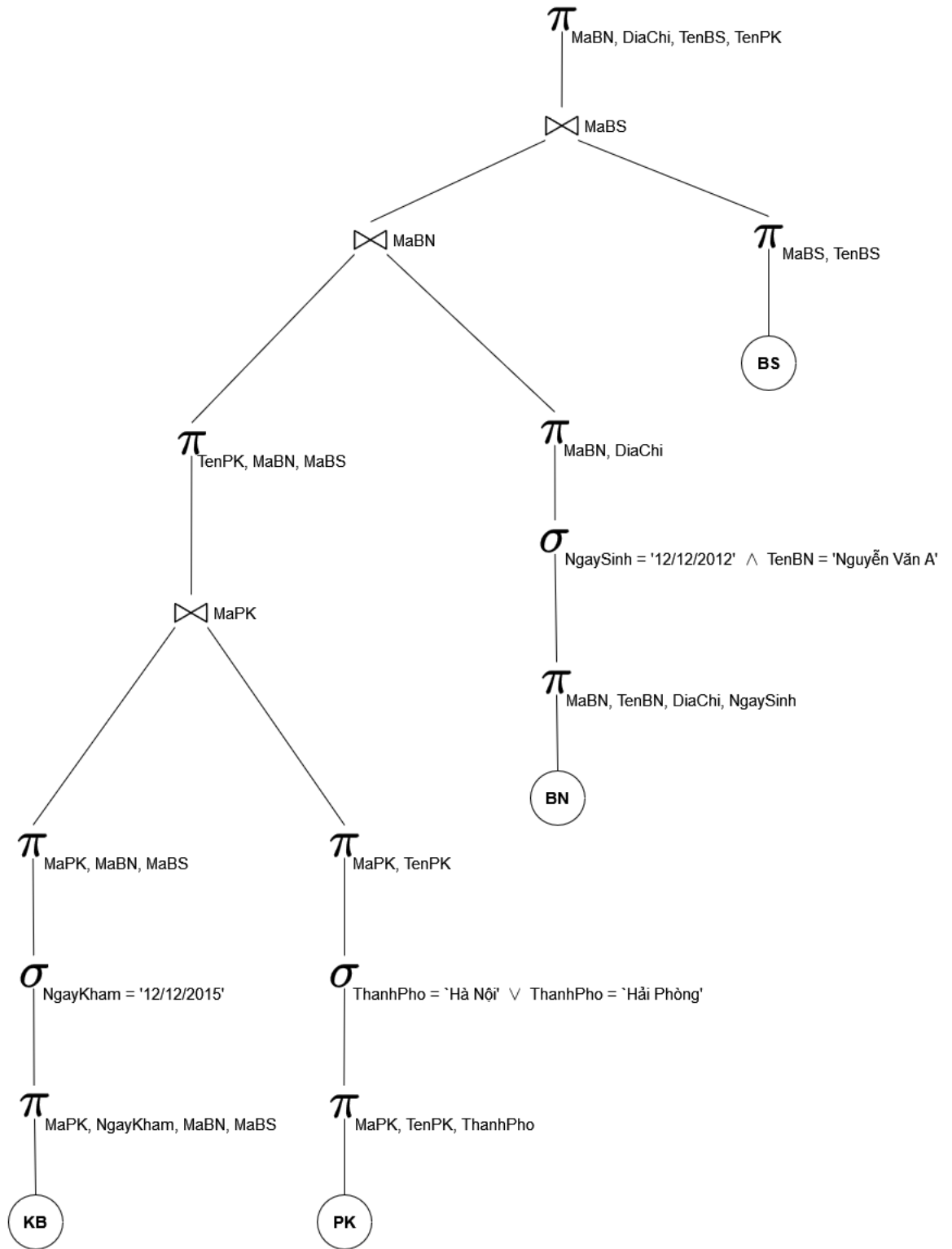
(a) Phân rã truy vấn để tối ưu hóa toàn cục câu truy vấn Q3.

Lời giải.

Cây truy vấn đại số quan hệ ban đầu của **Q3** là:



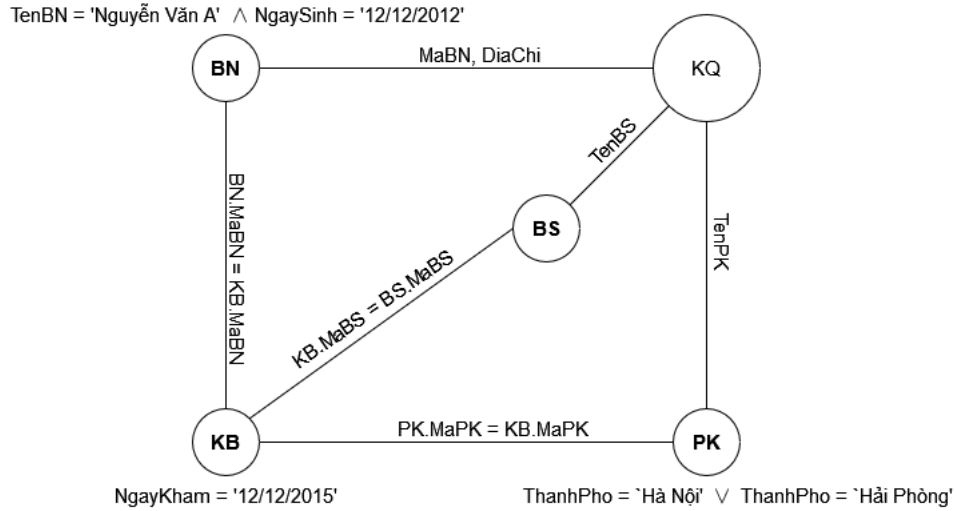
Cây truy vấn đại số quan hệ sau khi tối ưu hóa toàn cục là:



- (b) Vẽ đồ thị truy vấn (query graph) cho câu truy vấn **Q3**. Dựa vào đồ thị, hãy cho biết câu truy vấn này có sai ngữ nghĩa không?

Lời giải.

Đồ thị truy vấn của câu truy vấn **Q3**:



Ta thấy đồ thị truy vấn liên thông.

\Rightarrow Câu truy vấn **Q3** đúng ngữ nghĩa.

- (c) Biến đổi câu truy vấn a thành một câu truy vấn rút gọn trên các mảnh. (**Làm sau**)

Lời giải.

Để ngắn gọn, trong bài làm sẽ sử dụng các từ viết tắt như sau:

Từ viết tắt	Từ đầy đủ
PK	PHONGKHAM
BN	BENHNHAN
BS	BACSY
KB	KHAMBENH
TP	THUPHI
DV	DICHVU

Dựa vào lược đồ phân mảnh ta có chương trình định vị cho các quan hệ như sau:

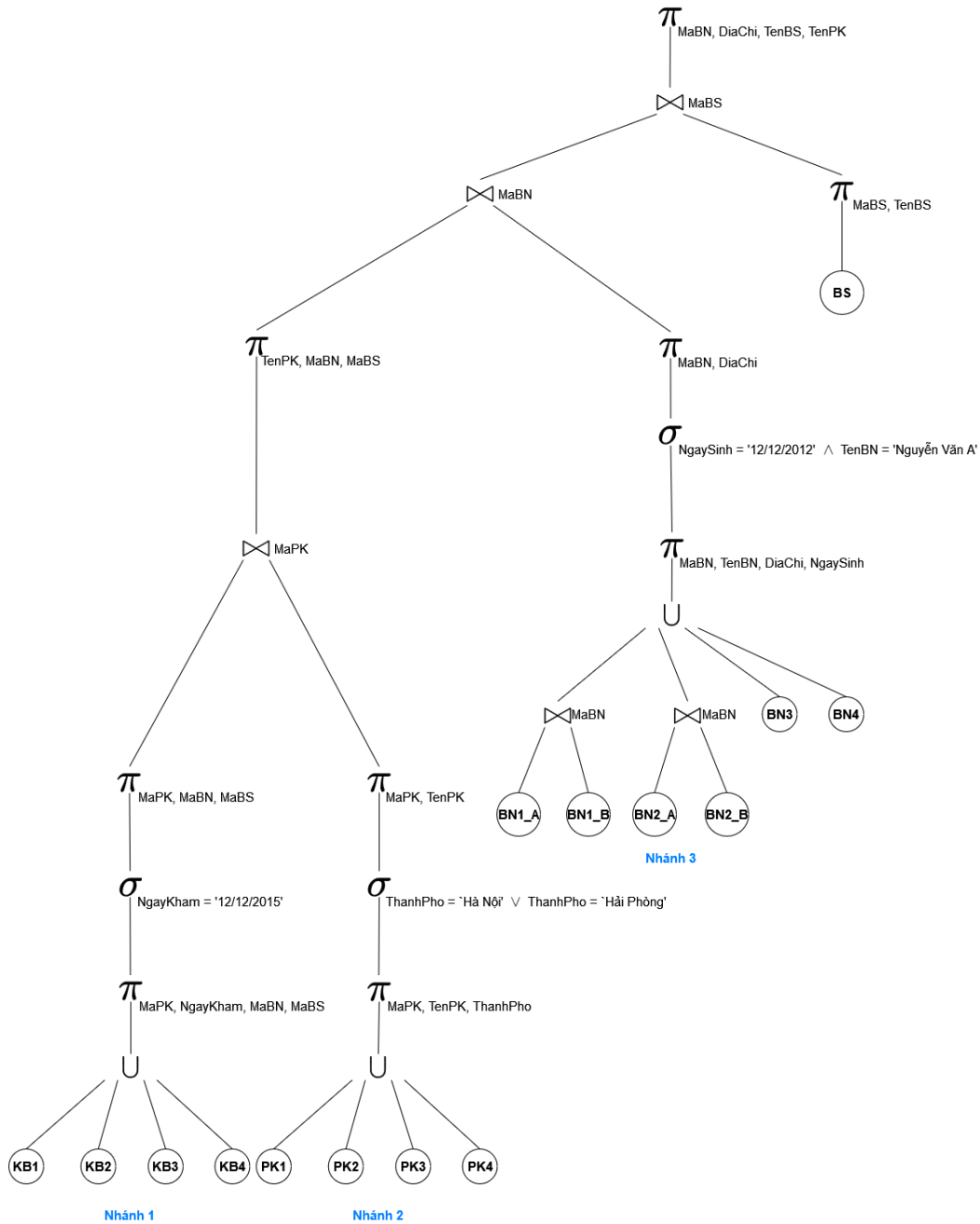
$PK = PK1 \cup PK2 \cup PK3 \cup PK4$

$KB = KB1 \cup KB2 \cup KB3 \cup KB4$

$BN = (BN1_A \bowtie BN1_B) \cup (BN2_A \bowtie BN2_B) \cup BN3 \cup BN4$

Quan hệ BS được nhân bản.

Thay thế chương trình định vị vào cây truy vấn đại số quan hệ tối ưu ở câu a, ta được cây truy vấn đại số quan hệ trên các mảnh:



- Xét nhánh 2 ta có:

$$\begin{aligned}
 & \sigma_{\text{ThanhPho}} = \text{"Hà Nội"} \vee \text{ThanhPho} = \text{"Hải Phòng"} \text{ (PK1} \cup \text{PK2} \cup \text{PK3} \cup \text{PK4)} \\
 & \Leftrightarrow \sigma_{\text{ThanhPho}} = \text{"Hà Nội"} \vee \text{ThanhPho} = \text{"Hải Phòng"} \text{ (PK1)} \\
 & \cup \sigma_{\text{ThanhPho}} = \text{"Hà Nội"} \vee \text{ThanhPho} = \text{"Hải Phòng"} \text{ (PK2)} \\
 & \cup \sigma_{\text{ThanhPho}} = \text{"Hà Nội"} \vee \text{ThanhPho} = \text{"Hải Phòng"} \text{ (PK3)} \\
 & \cup \sigma_{\text{ThanhPho}} = \text{"Hà Nội"} \vee \text{ThanhPho} = \text{"Hải Phòng"} \text{ (PK4)} \\
 & \Leftrightarrow \emptyset \cup \sigma_{\text{ThanhPho}} = \text{"Hà Nội"} \vee \text{ThanhPho} = \text{"Hải Phòng"} \text{ (PK2)} \\
 & \cup \sigma_{\text{ThanhPho}} = \text{"Hà Nội"} \vee \text{ThanhPho} = \text{"Hải Phòng"} \text{ (PK3)} \cup \emptyset \\
 & \Leftrightarrow \sigma_{\text{ThanhPho}} = \text{"Hà Nội"} \vee \text{ThanhPho} = \text{"Hải Phòng"} \text{ (PK2)} \\
 & \cup \sigma_{\text{ThanhPho}} = \text{"Hà Nội"} \vee \text{ThanhPho} = \text{"Hải Phòng"} \text{ (PK3)} \\
 & \Leftrightarrow \text{PK2} \cup \text{PK3} \text{ (1')}
 \end{aligned}$$
- Xét phép kết giữa quan hệ ở (1') và nhánh 1, ta có:

$$\begin{aligned}
 & (\text{PK2} \cup \text{PK3}) \bowtie (\text{KB1} \cup \text{KB2} \cup \text{KB3} \cup \text{KB4}) \\
 & \Leftrightarrow (\text{PK2} \bowtie \text{KB1}) \cup (\text{PK2} \bowtie \text{KB2}) \cup (\text{PK2} \bowtie \text{KB3}) \cup (\text{PK2} \bowtie \text{KB4}) \\
 & \cup (\text{PK3} \bowtie \text{KB1}) \cup (\text{PK3} \bowtie \text{KB2}) \cup (\text{PK3} \bowtie \text{KB3}) \cup (\text{PK3} \bowtie \text{KB4}) \\
 & \Leftrightarrow \emptyset \cup (\text{PK2} \bowtie \text{KB2}) \cup \emptyset \cup \emptyset \cup \emptyset \cup \emptyset \cup (\text{PK3} \bowtie \text{KB3}) \cup \emptyset
 \end{aligned}$$

$$\Leftrightarrow (PK2 \bowtie KB2) \cup (PK3 \bowtie KB3) \text{ (2')}$$

Nhờ phép biến đổi này, ta loại bỏ được các mảnh quan hệ không cần thiết là KB1 và KB4, chỉ giữ lại mảnh KB2 và KB3.

- Xét phép kết giữa quan hệ ở (2') và nhánh 3, ta có:

$$[(PK2 \bowtie KB2) \cup (PK3 \bowtie KB3)] \bowtie [(BN1_A \bowtie BN1_B) \cup (BN2_A \bowtie BN2_B) \cup BN3 \cup BN4]$$

$$\Leftrightarrow [(PK2 \bowtie KB2) \bowtie (BN1_A \bowtie BN1_B)] \cup [(PK2 \bowtie KB2) \bowtie (BN2_A \bowtie BN2_B)] \cup [(PK2 \bowtie KB2) \bowtie BN3] \cup [(PK2 \bowtie KB2) \bowtie BN4]$$

$$\cup [(PK3 \bowtie KB3) \bowtie (BN1_A \bowtie BN1_B)] \cup [(PK3 \bowtie KB3) \bowtie (BN2_A \bowtie BN2_B)] \cup [(PK3 \bowtie KB3) \bowtie BN3] \cup [(PK3 \bowtie KB3) \bowtie BN4]$$

$$\Leftrightarrow \emptyset \cup [(PK2 \bowtie KB2) \bowtie (BN2_A \bowtie BN2_B)] \cup \emptyset \cup \emptyset$$

$$\cup \emptyset \cup \emptyset \cup [(PK3 \bowtie KB3) \bowtie BN3] \cup \emptyset$$

$$\Leftrightarrow (PK2 \bowtie KB2 \bowtie BN2_A \bowtie BN2_B) \cup (PK3 \bowtie KB3 \bowtie BN3) \text{ (3')}$$

Vì $(TenBN, DiaChi, NgaySinh) \cap (DienThoai) = \emptyset$, nên phép chiếu trên BN2_A là vô ích, nên phép chiếu chỉ cần gán vào BN2_B.

$$(3') \Leftrightarrow (PK2 \bowtie KB2 \bowtie BN2_B) \cup (PK3 \bowtie KB3 \bowtie BN3) \text{ (3'')}$$

Áp dụng các phép biến đổi trên, ta được cây truy vấn đại số quan hệ rút gọn trên các mảnh dưới đây:

