

CƠ SỞ DỮ LIỆU



TỐI ƯU HÓA TRUY VẤN

Trần Ngọc Bảo
Email: tnbao.dhsp@gmail.com

TỐI ƯU HÓA TRUY VẤN



- **Giới thiệu về tối ưu hóa truy vấn**
- **Các qui tắc biến đổi tương đương**
- **Giải thuật Heuristic cho tối ưu hóa truy vấn**

TỐI ƯU HÓA TRUY VẤN



- **Giới thiệu về tối ưu hóa truy vấn**
- Các qui tắc biến đổi tương đương
- Giải thuật Heuristic cho tối ưu hóa truy vấn

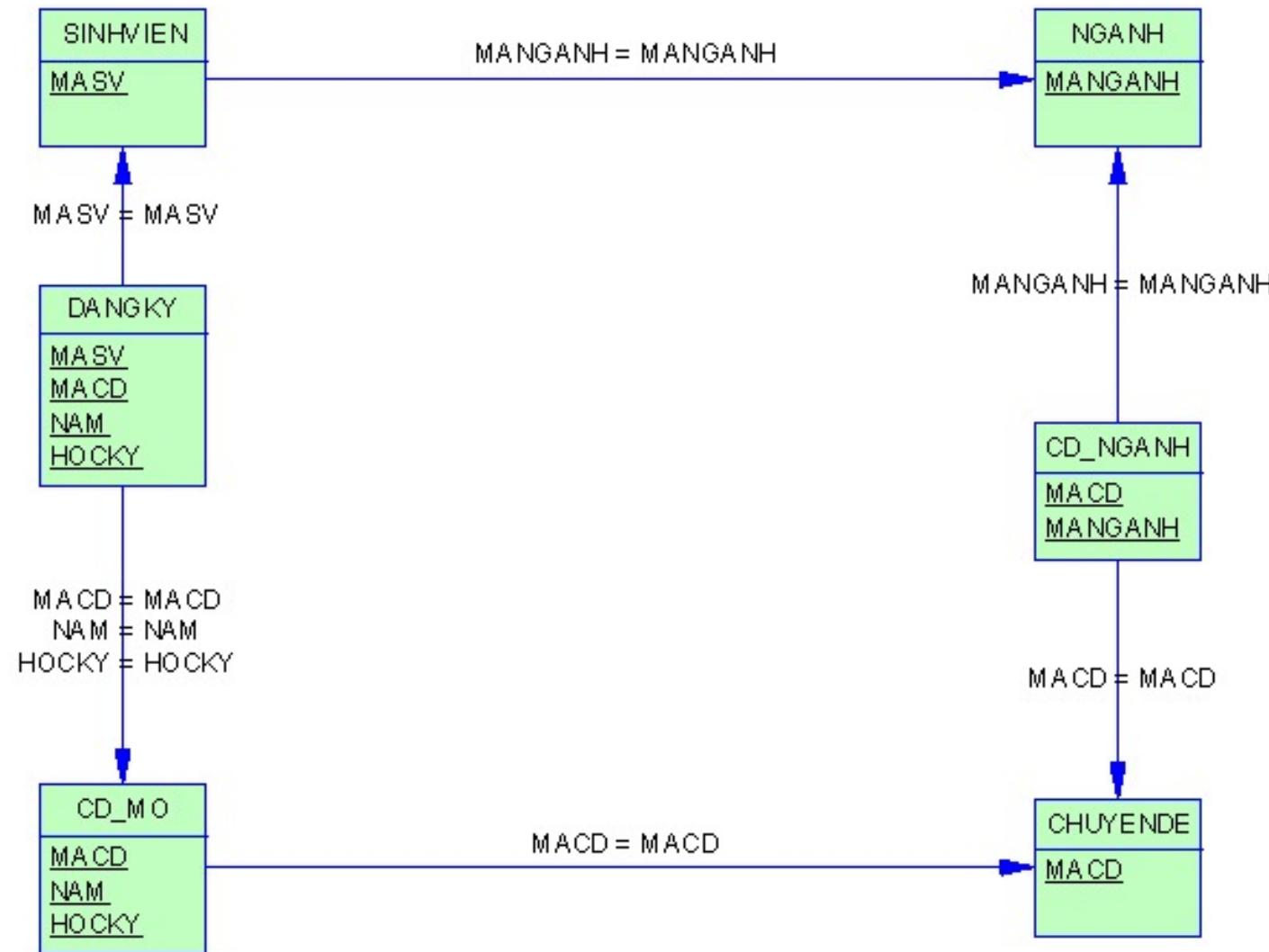
Giới thiệu về tối ưu hóa truy vấn



Phòng giáo vụ tại một trường Đại học muốn tin học hóa việc quản lý đăng ký học chuyên đề của sinh viên. Kết quả phân tích thiết kế được mô hình CSDL quan hệ như sau:

1. SINHVIEN (MASV, HOTEN, PHAI, NGAYSINH, MANGANH)
2. NGANH (MANGANH, TENNGANH, SOCD, TSSV)
3. CHUYENDE (MACD, TENCD, SOSVTD)
4. CD_NGANH (MACD, MANGANH)
5. CD_MO(MACD, NAM, HOCKY)
6. DANGKY(MASV, MACD, NAM, HOCKY)

Giới thiệu về tối ưu hóa truy vấn



Lược đồ CSDL ứng dụng quản lý chuyên đề

Giới thiệu về tối ưu hóa truy vấn



Yêu cầu truy vấn: **Liệt kê danh sách (MaSV, HoTen, NgaySinh)**
sinh viên Nam thuộc ngành “Sư phạm Tin học”

1. SINHVIEN (MASV, HOTEN, PHAI, NGAYSINH, MANGANH)

MASV	HOTEN	PHAI	NGAYSINH	MANGANH
K29.101.001	Nguyen Cong Phu	Nam	10/10/1979	SP101
K29.101.002	Phan Anh Khanh	Nu	1/10/1979	SP101
K29.103.001	Han Quoc Viet	Nam	3/2/1979	SP103
K29.103.002	Pham Khanh Nhu	Nu	10/12/1979	SP103
K29.201.001	Ly Thanh	Nam	3/8/1979	SP201
K29.201.002	Tran Ngoc Dung	Nu	4/5/1979	SP201
K29.104.001	Ho Anh Thu	Nu	1/4/1979	SP104

Giới thiệu về tối ưu hóa truy vấn



Yêu cầu truy vấn: **Liệt kê danh sách (MaSV, HoTen, NgaySinh)**
sinh viên Nam thuộc ngành “Sư phạm Tin học”

2. NGANH (MANGANH, TENNGANH, SOCD, TSSV)

MANGANH	TENNGANH	SOCD	TSSV
SP103	Sư Phạm Tin học	8	800
SP101	Sư Phạm Toán	5	1000
SP201	Sư Phạm Hóa	3	1500
SP102	Sư Phạm Lý	4	1200
CNTT	Khoa học máy tính	8	160
SP104	Sư Phạm Lý	6	1000
SP105	Sư phạm Anh	2	1100

Giới thiệu về tối ưu hóa truy vấn

BÀI GIẢNG HỌC PHẦN CƠ SỞ DỮ LIỆU TỐI ƯU HÓA TRUY VẤN



Yêu cầu truy vấn: **Liệt kê danh sách (MaSV, HoTen, NgaySinh)**
sinh viên Nam thuộc ngành “Sư phạm Tin học”

MASV	HOTEN	PHAI	NGAYSINH	MANGANH
K29.101.001	Nguyen Cong Phu	Nam	10/10/1979	SP101
K29.101.002	Phan Anh Khanh	Nu	1/10/1979	SP101
K29.103.001	Han Quoc Viet	Nam	3/2/1979	SP103
K29.103.002	Pham Khanh Nhu	Nu	10/12/1979	SP103
K29.201.001	Lý Thanh	Nam	3/8/1979	SP201
K29.201.002	Tran Ngoc Dung	Nu	4/5/1979	SP201
K29.104.001	Ho Anh Thu	Nu	1/4/1979	SP104

MANGANH	TENNGANH	SOCD	TSSV
SP103	Sư Phạm Tin học	8	800
SP101	Sư Phạm Toán	5	1000
SP201	Sư Phạm Hóa	3	1500
SP102	Sư Phạm Lý	4	1200
CNTT	Khoa học máy tính	8	160
SP104	Sư Phạm Lý	6	1000
SP105	Sư phạm Anh	2	1100

Kết quả truy vấn: 1 dòng

MASV	HOTEN	NGAYSINH
K29.103.001	Han Quoc Viet	3/2/1979

Giới thiệu về tối ưu hóa truy vấn



Yêu cầu truy vấn: **Liệt kê danh sách (MaSV, HoTen, NgaySinh)**
sinh viên Nam thuộc ngành “Sư phạm Tin học”

1. Cách 1:

```
SELECT MASV, HOTEN, NGAYSINH  
FROM SINHVIEN SV, NGANH N  
WHERE SV.MANGANH = N.MANGANH AND PHAI = 'NAM'  
AND TENNGANH = 'Sư phạm Tin học'
```

2. Cách 2:

```
SELECT MASV, HOTEN, NGAYSINH  
FROM SINHVIEN SV JOIN NGANH N ON SV.MANGANH =  
N.MANGANH  
WHERE PHAI = 'NAM' AND TENNGANH = 'Sư phạm Tin  
học'
```

Giới thiệu về tối ưu hóa truy vấn



Yêu cầu truy vấn: **Liệt kê danh sách (MaSV, HoTen, NgaySinh)**
sinh viên Nam thuộc ngành “Sư phạm Tin học”

3. Cách 3:

```
SELECT MASV, HOTEN, NGAYSINH  
FROM  
(SELECT MASV, TENSV, NGAYSINH, MANGANH  
FROM SINHVIEN  
WHERE PHAI = 'NAM') AS SV JOIN  
(SELECT MANGANH  
FROM NGANH  
WHERE TENNGANH = 'Sư phạm Tin học') AS N  
ON SV.MANGANH = N.MANGANH
```

Trong 3 cách biểu diễn như trên cách nào biểu diễn
nào chạy chậm nhất (xấu nhất) ?, cách biểu diễn
nào chạy nhanh nhất (tối ưu nhất) ?

Giới thiệu về tối ưu hóa truy vấn



Yêu cầu truy vấn: **Liệt kê danh sách (MaSV, HoTen, NgaySinh)**
sinh viên Nam thuộc ngành “Sư phạm Tin học”

1. Cách 1:

```
SELECT MASV, HOTEN, NGAYSINH  
FROM SINHVIEN SV, NGANH N  
WHERE SV.MANGANH = N.MANGANH AND PHAI = 'NAM'  
AND TENNGANH = 'Sư phạm Tin học'
```

Giới thiệu về tối ưu hóa truy vấn



Yêu cầu truy vấn: **Liệt kê danh sách (MaSV, HoTen, NgaySinh)**
sinh viên Nam thuộc ngành “Sư phạm Tin học”

1. Cách 1: SinhVien x Nganh → **7 x 7 = 49 dòng**

MASV	HOTEN	PHAI	NGAYSINH	MANGANH
K29.101.001	Nguyen Cong Phu	Nam	10/10/1979	SP101
K29.101.002	Phan Anh Khanh	Nu	1/10/1979	SP101
K29.103.001	Han Quoc Viet	Nam	3/2/1979	SP103
K29.103.002	Pham Khanh Nhu	Nu	10/12/1979	SP103
K29.201.001	Lý Thanh	Nam	3/8/1979	SP201
K29.201.002	Tran Ngoc Dung	Nu	4/5/1979	SP201
K29.104.001	Ho Anh Thu	Nu	1/4/1979	SP104

MANGANH	TENNGANH	SOCD	TSSV
SP103	Sư Phạm Tin học	8	800
SP101	Sư Phạm Toán	5	1000
SP201	Sư Phạm Hóa	3	1500
SP102	Sư Phạm Lý	4	1200
CNTT	Khoa học máy tính	8	160
SP104	Sư Phạm Lý	6	1000
SP105	Sư phạm Anh	2	1100

Giới thiệu về tối ưu hóa truy vấn



Yêu cầu truy vấn: **Liệt kê danh sách (MaSV, HoTen, NgaySinh)**
sinh viên Nam thuộc ngành “Sư phạm Tin học”

1. Cách 1: SinhVien x Nganh **7 x 7 = 49 dòng**

MASV	HOTEN	PHAI	NGAYSINH	MANGANH	MANGANH	TENNGANH	SOCD	TSSV
K29.101.001	Nguyen Cong Phu	Nam	10/10/1979	SP101	SP103	Sư Phạm Tin học	8	800
K29.101.001	Nguyen Cong Phu	Nam	10/10/1979	SP101	SP101	Sư Phạm Toán	5	1000
K29.101.001	Nguyen Cong Phu	Nam	10/10/1979	SP101	SP201	Sư Phạm Hóa	3	1500
K29.101.001	Nguyen Cong Phu	Nam	10/10/1979	SP101	SP102	Sư Phạm Lý	4	1200
K29.101.001	Nguyen Cong Phu	Nam	10/10/1979	SP101	CNTT	Khoa học máy tính	8	160
K29.101.001	Nguyen Cong Phu	Nam	10/10/1979	SP101	SP104	Sư Phạm Lý	6	1000
K29.101.001	Nguyen Cong Phu	Nam	10/10/1979	SP101	SP105	Sư phạm Anh	2	1100
K29.101.002	Phan Anh Khanh	Nu	1/10/1979	SP101	SP103	Sư Phạm Tin học	8	800
K29.101.002	Phan Anh Khanh	Nu	1/10/1979	SP101	SP101	Sư Phạm Toán	5	1000
K29.101.002	Phan Anh Khanh	Nu	1/10/1979	SP101	SP201	Sư Phạm Hóa	3	1500
.....								

Giới thiệu về tối ưu hóa truy vấn



Yêu cầu truy vấn: **Liệt kê danh sách (MaSV, HoTen, NgaySinh)**
sinh viên Nam thuộc ngành “Sư phạm Tin học”

1. Cách 1: SinhVien x Nganh → **7 x 7 = 49 dòng**

MASV	HOTEN	PHAI	NGAYSINH	MANGANH	MANGANH	TENNGANH	SOCD	TSSV
K29.103.001	Han Quoc Viet	Nam	3/2/1979	SP103	SP103	Sư Phạm Tin học	8	800
K29.103.001	Han Quoc Viet	Nam	3/2/1979	SP103	SP101	Sư Phạm Toán	5	1000
K29.103.001	Han Quoc Viet	Nam	3/2/1979	SP103	SP201	Sư Phạm Hóa	3	1500
K29.103.001	Han Quoc Viet	Nam	3/2/1979	SP103	SP102	Sư Phạm Lý	4	1200
K29.103.001	Han Quoc Viet	Nam	3/2/1979	SP103	CNTT	Khoa học máy tính	8	160
K29.103.001	Han Quoc Viet	Nam	3/2/1979	SP103	SP104	Sư Phạm Lý	6	1000
K29.103.001	Han Quoc Viet	Nam	3/2/1979	SP103	SP105	Sư phạm Anh	2	1100
K29.103.002	Pham Khanh Nhu	Nu	10/12/1979	SP103	SP103	Sư Phạm Tin học	8	800
K29.103.002	Pham Khanh Nhu	Nu	10/12/1979	SP103	SP101	Sư Phạm Toán	5	1000
K29.103.002	Pham Khanh Nhu	Nu	10/12/1979	SP103	SP201	Sư Phạm Hóa	3	1500
.....								

Giới thiệu về tối ưu hóa truy vấn



Yêu cầu truy vấn: **Liệt kê danh sách (MaSV, HoTen, NgaySinh)**
sinh viên Nam thuộc ngành “Sư phạm Tin học”

1. Cách 1: SinhVien x Nganh → **7 x 7 = 49 dòng**

Kết quả truy vấn: 1 dòng

MASV	HOTEN	PHAI	NGAYSINH	MANGANH	MANGANH	TENNGANH	SOCD	TSSV
K29.103.001	Han Quoc Viet	Nam	3/2/1979	SP103	SP103	Sư Phạm Tin học	8	800
K29.103.001	Han Quoc Viet	Nam	3/2/1979	SP103	SP101	Sư Phạm Toán	5	1000
K29.103.001	Han Quoc Viet	Nam	3/2/1979	SP103	SP201	Sư Phạm Hóa	3	1500
K29.103.001	Han Quoc Viet	Nam	3/2/1979	SP103	SP102	Sư Phạm Lý	4	1200
K29.103.001	Han Quoc Viet	Nam	3/2/1979	SP103	CNTT	Khoa học máy tính	8	160
K29.103.001	Han Quoc Viet	Nam	3/2/1979	SP103	SP104	Sư Phạm Lý	6	1000
K29.103.001	Han Quoc Viet	Nam	3/2/1979	SP103	SP105	Sư phạm Anh	2	1100
K29.103.002	Pham Khanh Nhu	Nu	10/12/1979	SP103	SP103	Sư Phạm Tin học	8	800

Giới thiệu về tối ưu hóa truy vấn



Yêu cầu truy vấn: **Liệt kê danh sách (MaSV, HoTen, NgaySinh)**
sinh viên Nam thuộc ngành “Sư phạm Tin học”

2. Cách 2:

```
SELECT MASV, HOTEN, NGAYSINH  
FROM SINHVIEN SV JOIN NGANH N ON SV.MANGANH =  
N.MANGANH  
WHERE PHAI = 'NAM' AND TENNGANH = 'Sư phạm Tin  
học'
```

Giới thiệu về tối ưu hóa truy vấn



Yêu cầu truy vấn: **Liệt kê danh sách (MaSV, HoTen, NgaySinh)**
sinh viên Nam thuộc ngành “Sư phạm Tin học”

2. Cách 2: SinhVien \bowtie Nganh **7 x 1 = 7 dòng**

MASV	HOTEN	PHAI	NGAYSINH	MANGANH
K29.101.001	Nguyen Cong Phu	Nam	10/10/1979	SP101
K29.101.002	Phan Anh Khanh	Nu	1/10/1979	SP101
K29.103.001	Han Quoc Viet	Nam	3/2/1979	SP103
K29.103.002	Pham Khanh Nhu	Nu	10/12/1979	SP103
K29.201.001	Ly Thanh	Nam	3/8/1979	SP201
K29.201.002	Tran Ngoc Dung	Nu	4/5/1979	SP201
K29.104.001	Ho Anh Thu	Nu	1/4/1979	SP104

MANGANH	TENNGANH	SOCD	TSSV
SP103	Sư Phạm Tin học	8	800
SP101	Sư Phạm Toán	5	1000
SP201	Sư Phạm Hóa	3	1500
SP102	Sư Phạm Lý	4	1200
CNTT	Khoa học máy tính	8	160
SP104	Sư Phạm Lý	6	1000
SP105	Sư phạm Anh	2	1100

Giới thiệu về tối ưu hóa truy vấn



Yêu cầu truy vấn: **Liệt kê danh sách (MaSV, HoTen, NgaySinh)**
sinh viên Nam thuộc ngành “Sư phạm Tin học”

2. Cách 2: SinhVien \bowtie Nganh **7 x 1 = 7 dòng**

MASV	HOTEN	PHAI	NGAYSINH	MANGANH
K29.101.001	Nguyen Cong Phu	Nam	10/10/1979	SP101
K29.101.002	Phan Anh Khanh	Nu	1/10/1979	SP101
K29.103.001	Han Quoc Viet	Nam	3/2/1979	SP103
K29.103.002	Pham Khanh Nhu	Nu	10/12/1979	SP103
K29.201.001	Lý Thanh	Nam	3/8/1979	SP201
K29.201.002	Tran Ngoc Dung	Nu	4/5/1979	SP201
K29.104.001	Ho Anh Thủ	Nu	1/4/1979	SP104

MANGANH	TENNGANH	SOCD	TSSV
SP103	Sư Phạm Tin học	8	800
SP101	Sư Phạm Toán	5	1000
SP201	Sư Phạm Hóa	3	1500
SP102	Sư Phạm Lý	4	1200
CNTT	Khoa học máy tính	8	160
SP104	Sư Phạm Lý	6	1000
SP105	Sư phạm Anh	2	1100

Giới thiệu về tối ưu hóa truy vấn



Yêu cầu truy vấn: **Liệt kê danh sách (MaSV, HoTen, NgaySinh)**
sinh viên Nam thuộc ngành “Sư phạm Tin học”

2. Cách 2: SinhVien \bowtie Nganh **$7 \times 1 = 7$ dòng**

Kết quả truy vấn: 1 dòng

MASV	HOTEN	PHAI	NGAYSINH	MANGANH	MANGANH	TENNGANH	SOCD	TSSV
K29.103.001	Han Quoc Viet	Nam	3/2/1979	SP103	SP103	Sư Phạm Tin học	8	800

Giới thiệu về tối ưu hóa truy vấn



Yêu cầu truy vấn: **Liệt kê danh sách (MaSV, HoTen, NgaySinh)**
sinh viên Nam thuộc ngành “Sư phạm Tin học”

3. Cách 3:

```
SELECT MASV, HOTEN, NGAYSINH  
FROM  
(SELECT MASV,HOTEN, NGAYSINH, MANGANH  
FROM SINHVIEN  
WHERE PHAI = 'NAM') AS SV JOIN  
(SELECT MANGANH  
FROM NGANH  
WHERE TENNGANH = 'Sư phạm Tin học') AS N  
ON SV.MANGANH = N.MANGANH
```

Giới thiệu về tối ưu hóa truy vấn

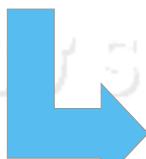


Yêu cầu truy vấn: **Liệt kê danh sách (MaSV, HoTen, NgaySinh)**
sinh viên Nam thuộc ngành “Sư phạm Tin học”

3. Cách 3: Bước 1

SELECT MASV, HOTEN, NGAYSINH
FROM

(SELECT MASV, HOTEN, NGAYSINH, MANGANH
FROM SINHVIEN
WHERE PHAI = 'NAM') AS SV JOIN
(SELECT MANGANH
FROM NGANH
WHERE TENNGANH = 'Sư phạm Tin học') AS N
ON SV.MANGANH = N.MANGANH



Chọn các dòng [MASV, HOTEN, NGAYSINH, MANGANH]
có thuộc tính PHAI = 'NAM' trong quan hệ SINHVIEN

Giới thiệu về tối ưu hóa truy vấn



Yêu cầu truy vấn: **Liệt kê danh sách (MaSV, HoTen, NgaySinh)**
sinh viên Nam thuộc ngành “Sư phạm Tin học”

3. Cách 3: Bước 1

```
(SELECT MASV, TENSV, NGAYSINH,  
MANGANH FROM SINHVIEN  
WHERE PHAI = 'NAM')
```

MASV	HOTEN	PHAI	NGAYSINH	MANGANH
K29.101.001	Nguyen Cong Phu	Nam	10/10/1979	SP101
K29.101.002	Phan Anh Khanh	Nu	1/10/1979	SP101
K29.103.001	Han Quoc Viet	Nam	3/2/1979	SP103
K29.103.002	Pham Khanh Nhu	Nu	10/12/1979	SP103
K29.201.001	Ly Thanh	Nam	3/8/1979	SP201
K29.201.002	Tran Ngoc Dung	Nu	4/5/1979	SP201
K29.104.001	Ho Anh Thu	Nu	1/4/1979	SP104

Kết quả truy vấn: 3 dòng

MASV	HOTEN	NGAYSINH	MANGANH
K29.101.001	Nguyen Cong Phu	10/10/1979	SP101
K29.103.001	Han Quoc Viet	3/2/1979	SP103
K29.201.001	Ly Thanh	3/8/1979	SP201

Giới thiệu về tối ưu hóa truy vấn



Yêu cầu truy vấn: **Liệt kê danh sách (MaSV, HoTen, NgaySinh)**
sinh viên Nam thuộc ngành “Sư phạm Tin học”

3. Cách 3: Bước 2

```
SELECT MASV, HOTEN, NGAYSINH  
FROM  
(SELECT MASV,HOTEN, NGAYSINH, MANGANH  
FROM SINHVIEN  
WHERE PHAI = 'NAM') AS SV JOIN  
(SELECT MANGANH  
FROM NGANH  
WHERE TENNGANH = 'Sư phạm Tin học') AS N  
ON SV.MANGANH = N.MANGANH
```



**Chọn các dòng [MANGANH] có thuộc tính TENNGANH =
'Sư phạm Tin học' trong quan hệ NGANH**

Giới thiệu về tối ưu hóa truy vấn



Yêu cầu truy vấn: **Liệt kê danh sách (MaSV, HoTen, NgaySinh)**
sinh viên Nam thuộc ngành “Sư phạm Tin học”

3. Cách 3: Bước 2

**(SELECT MANGANH
FROM NGANH
WHERE TENNGANH = ‘Sư
phạm Tin học’)**

Kết quả truy vấn: 1 dòng

MANGANH
SP103

MANGANH	TENNGANH	SOCĐ	TSSV
SP103	Sư Phạm Tin học	8	800
SP101	Sư Phạm Toán	5	1000
SP201	Sư Phạm Hóa	3	1500
SP102	Sư Phạm Lý	4	1200
CNTT	Khoa học máy tính	8	160
SP104	Sư Phạm Lý	6	1000
SP105	Sư phạm Anh	2	1100

Giới thiệu về tối ưu hóa truy vấn



Yêu cầu truy vấn: **Liệt kê danh sách (MaSV, HoTen, NgaySinh)**
sinh viên Nam thuộc ngành “Sư phạm Tin học”

3. Cách 3: Bước 3

```
SELECT MASV, HOTEN, NGAYSINH  
FROM  
(SELECT MASV,HOTEN, NGAYSINH, MANGANH  
FROM SINHVIEN  
WHERE PHAI = 'NAM') AS SV JOIN  
(SELECT MANGANH  
FROM NGANH  
WHERE TENNGANH = 'Sư phạm Tin học') AS N  
ON SV.MANGANH = N.MANGANH
```

Giới thiệu về tối ưu hóa truy vấn



Yêu cầu truy vấn: **Liệt kê danh sách (MaSV, HoTen, NgaySinh)**
sinh viên Nam thuộc ngành “Sư phạm Tin học”

3. Cách 3: Bước 3 [Bước 1] AS SV JOIN

[Bước 2] AS N

ON SV.MANGANH = N.MANGANH

Kết quả truy vấn bước 1 (SV) \bowtie Kết quả truy vấn bước 2 (N)

MASV	HOTEN	NGAYSINH	MANGANH
K29.101.001	Nguyen Cong Phu	10/10/1979	SP101
K29.103.001	Han Quoc Viet	3/2/1979	SP103
K29.201.001	Ly Thanh	3/8/1979	SP201

\bowtie

MANGANH
SP103



MASV	HOTEN	NGAYSINH	MANGANH	MANGANH
K29.103.001	Han Quoc Viet	3/2/1979	SP103	SP103

Giới thiệu về tối ưu hóa truy vấn



Yêu cầu truy vấn: **Liệt kê danh sách (MaSV, HoTen, NgaySinh)**
sinh viên Nam thuộc ngành “Sư phạm Tin học”

3. Cách 3: Bước 4

**SELECT MASV, HOTEN, NGAYSINH
FROM**

(SELECT MASV, TENSV, NGAYSINH, MANGANH
FROM SINHVIEN
WHERE PHAI = ‘NAM’) AS SV JOIN
(SELECT MANGANH
FROM NGANH
WHERE TENNGANH = ‘Sư phạm Tin học’) AS N
ON SV.MANGANH = N.MANGANH



Chiếu trên thuộc tính **[MASV, HOTEN, NGAYSINH]** trong
kết quả của **Bước 3**

Giới thiệu về tối ưu hóa truy vấn



Yêu cầu truy vấn: **Liệt kê danh sách (MaSV, HoTen, NgaySinh)**
sinh viên Nam thuộc ngành “Sư phạm Tin học”

3. Cách 3: Bước 4

**SELECT MASV, HOTEN,
NGAYSINH
FROM [KQ BƯỚC 3]**

Kết quả truy vấn bước 3

MASV	HOTEN	NGAYSINH	MANGANH	MANGANH
K29.103.001	Han Quoc Viet	3/2/1979	SP103	SP103



MASV	HOTEN	NGAYSINH
K29.103.001	Han Quoc Viet	3/2/1979

Giới thiệu về tối ưu hóa truy vấn



Trong 3 cách biểu diễn như trên cách nào biểu diễn nào chạy chậm nhất (xấu nhất) ?, cách biểu diễn nào chạy nhanh nhất (tối ưu nhất) ?

1. **Cách 1: Xấu nhất**



Đơn giản nhất

2. **Cách 2: Thường sử dụng nhất**



Trung bình

3. **Cách 3: Tốt nhất**



Phức tạp nhất

Giới thiệu về tối ưu hóa truy vấn



Yêu cầu truy vấn: **Liệt kê danh sách (MaSV, HoTen, NgaySinh)**
sinh viên Nam thuộc ngành “Sư phạm Tin học”

1. Cách 1:

Ngôn ngữ SQL

```
SELECT MASV, HOTEN, NGAYSINH  
FROM SINHVIEN SV, NGANH N  
WHERE SV.MANGANH = N.MANGANH AND PHAI = 'NAM'  
AND TENNGANH = 'Sư phạm Tin học'
```

Ngôn ngữ ĐSQL

$$\prod_{\text{MASV, HOTEN, NGAYSINH}} \sigma_{\text{TENNGANH} = \text{'Sư phạm Tin học'}} \\ \wedge \text{PHAI} = \text{'NAM'} \wedge \text{SINHVIEN.MANGANH} = \text{NGANH.MANGANH} \\ (\text{SINHVIEN} \times \text{NGANH})$$

Giới thiệu về tối ưu hóa truy vấn



Yêu cầu truy vấn: **Liệt kê danh sách (MaSV, HoTen, NgaySinh)**
sinh viên Nam thuộc ngành “Sư phạm Tin học”

2. Cách 2:

Ngôn ngữ SQL

```
SELECT MASV, HOTEN, NGAYSINH  
FROM SINHVIEN SV JOIN NGANH N ON SV.MANGANH =  
N.MANGANH  
WHERE PHAI = 'NAM' AND TENNGANH = 'Sư phạm Tin  
học'
```

Ngôn ngữ ĐSQL

$$\prod_{\begin{array}{l} \text{MASV, HOTEN, NGAYSINH} \\ \wedge \text{PHAI} = 'NAM' \end{array}} \sigma_{\begin{array}{l} \text{TENNGANH = 'Sư phạm Tin học'} \\ \text{SINHVIEN.MANGANH = NGANH.MANGANH} \end{array}}$$

Giới thiệu về tối ưu hóa truy vấn



Yêu cầu truy vấn: **Liệt kê danh sách (MaSV, HoTen, NgaySinh)**
sinh viên Nam thuộc ngành “Sư phạm Tin học”

3. Cách 3:

Ngôn ngữ SQL

```
SELECT MASV, HOTEN, NGAYSINH
FROM
    (SELECT MASV,HOTEN, NGAYSINH, MANGANH   FROM SINHVIEN
     WHERE PHAI = 'NAM') AS SV JOIN
    (SELECT MANGANH
     FROM NGANH
     WHERE TENNGANH = 'Sư phạm Tin học') AS N
     ON SV.MANGANH = N.MANGANH
```

Ngôn ngữ ĐSQL

$$\Pi_{MASV, HOTEN, NGAYSINH} ((\pi_{MASV, HOTEN, NGAYSINH, MANGANH} \sigma_{PHAI = 'NAM'}(SINHVIEN)) \bowtie (\pi_{MANGANH} \sigma_{NGANH = 'Sư phạm Tin học'}(NGANH)))$$
$$SINHVIEN.MANGANH = NGANH.MANGANH$$

Giới thiệu về tối ưu hóa truy vấn



**Tối ưu hóa truy vấn là một giải thuật
Heuristics tìm cách biểu diễn các truy vấn
cho kết quả nhanh nhất (tối ưu nhất) áp
dụng một số qui tắc biến đổi tương đương**

1. Cách 1: Xấu nhất



3. Cách 3: Tốt nhất

TỐI ƯU HÓA TRUY VẤN



- Giới thiệu về tối ưu hóa truy vấn
- **Các qui tắc biến đổi tương đương**
- Giải thuật Heuristic cho tối ưu hóa truy vấn

Các qui tắc biến đổi tương đương



1. Qui tắc 1: biến đổi dãy các phép chọn

$$\sigma_{c1 \text{ AND } c2 \dots \text{ AND } cn}(R) \equiv \sigma_{c1}(\sigma_{c2}(\dots \sigma_{cn}(R))\dots)$$

Ví dụ

1. **SINHVIEN (MASV, HOTEN, PHAI, NGAYSINH, MANGANH)**

$\sigma_{MANGANH = 'SPS101' \text{ AND } PHAI = 'NAM'}(SINHVIEN)$

=

$\sigma_{MANGANH = 'SPS101'}(\sigma_{PHAI = 'NAM'}(SINHVIEN))$

Các qui tắc biến đổi tương đương



2. Qui tắc 2: Tính giao hoán của phép chọn

$$\sigma_{c1}(\sigma_{c2}(R)) = \sigma_{c2}(\sigma_{c1}(R))$$

Ví dụ

1. SINHVIEN (MASV, HOTEN, PHAI, NGAYSINH, MANGANH)

$$\sigma_{MANGANH} = 'SPS101' (\sigma_{PHAI} = 'NAM' (SINHVIEN)) \\ =$$

$$\sigma_{PHAI} = 'NAM' (\sigma_{MANGANH} = 'SPS01' (SINHVIEN))$$

Các qui tắc biến đổi tương đương



3. Qui tắc 3: Biến đổi dãy các phép chiếu

$$\pi_{\langle DS1 \rangle} (\pi_{\langle DS2 \rangle} (\dots \pi_{\langle DS_n \rangle} (R)) \dots) = \pi_{\langle DS1 \rangle} (R)$$

Ví dụ

1. SINHVIEN (MASV, HOTEN, PHAI, NGAYSINH, MANGANH)

π MASV, HOTEN (π MASV, HOTEN, PHAI, NGAYSINH (SINHVIEN))

=

π MASV, HOTEN (SINHVIEN))

Các qui tắc biến đổi tương đương



4. Qui tắc 4: Tính giao hoán của phép chọn với phép chiếu

$$\pi_{A1, A2, \dots, An}(\sigma_c(R)) \equiv \sigma_c(\pi_{A1, A2, \dots, An}(R))$$

Ví dụ

1. SINHVIEN (MASV, HOTEN, PHAI, NGAYSINH, MANGANH)

π MASV, HOTEN, PHAI (σ PHAI = 'NAM' (SINHVIEN))

=

σ PHAI = 'NAM' (π MASV, HOTEN (SINHVIEN))

Các qui tắc biến đổi tương đương



5. Qui tắc 5: Tính giao hoán của phép tích, phép kết

$$(R \bowtie_C S) = (S \bowtie_C R)$$

Ví dụ

$$(R \times S) = (S \times R)$$

1. **SINHVIEN (MASV, HOTEN, PHAI, NGAYSINH, MANGANH)**
2. **NGANH (MANGANH, TENNGANH)**

$$(\text{SINHVIEN} \bowtie \text{NGANH})$$

$$\text{SINHVIEN.MANGANH} = \text{NGANH.MANGANH}$$

$$= (\text{NGANH} \bowtie (\text{SINHVIEN}))$$

$$\text{SINHVIEN.MANGANH} = \text{NGANH.MANGANH}$$

Các qui tắc biến đổi tương đương



6. Qui tắc 6:

Nếu tất cả các thuộc tính trong c đều là thuộc tính của R , thì

$$\sigma_c(R \bowtie S) \equiv (\sigma_c(R)) \bowtie S$$

Ví dụ

1. SINHVIEN (MASV, HOTEN, PHAI, NGAYSINH, MANGANH)
2. NGANH (MANGANH, TENNGANH)

$\sigma_{\text{PHAI} = \text{'NAM'}}(\text{SINHVIEN} \bowtie \text{NGANH})$

=

$(\sigma_{\text{PHAI} = \text{'NAM'}}(\text{SINHVIEN})) \bowtie \text{NGANH}$

Các qui tắc biến đổi tương đương



6. Qui tắc 6:

Nếu $c=c_1 \text{ and } c_2$, trong đó c_1 là một thuộc tính của R , c_2 là một thuộc tính của S , thì:

$$\sigma_c(R \bowtie S) \equiv (\sigma_{c_1}(R)) \bowtie (\sigma_{c_2}(S))$$

Ví dụ

1. SINHVIEN (MASV, HOTEN, PHAI, NGAYSINH, MANGANH)
2. NGANH (MANGANH, TENNGANH)

$\sigma_{\text{PHAI} = \text{'NAM'} \text{ AND } \text{TENNGANH} = \text{'SP TIN'}}(\text{SINHVIEN} \bowtie \text{NGANH})$

=

$(\sigma_{\text{PHAI} = \text{'NAM'}}(\text{SINHVIEN})) \bowtie (\sigma_{\text{TENNGANH} = \text{'SP TIN'}}(\text{NGANH}))$

Các qui tắc biến đổi tương đương



6. Qui tắc 6:

Nếu tất cả các thuộc tính trong **c** đều là thuộc tính của **R**, thì

$$\sigma_c(R \times S) \equiv (\sigma_c(R)) \times S$$

Ví dụ

1. **SINHVIEN** (MASV, HOTEN, PHAI, NGAYSINH, MANGANH)
2. **NGANH** (MANGANH, TENNGANH)

$$\sigma_{\text{PHAI} = \text{'NAM'}}(\text{SINHVIEN} \times \text{NGANH})$$

=

$$(\sigma_{\text{PHAI} = \text{'NAM'}}(\text{SINHVIEN})) \times \text{NGANH}$$



Các qui tắc biến đổi tương đương

6. Qui tắc 6:

Nếu $c=c_1$ and c_2 , trong đó c_1 là một thuộc tính của R , c_2 là một thuộc tính của S , thì:

$$\sigma_c(R \times S) \equiv (\sigma_{c_1}(R)) \times (\sigma_{c_2}(S))$$

1. SINHVIEN (MASV, HOTEN, PHAI, NGAYSINH, MANGANH)
2. NGANH (MANGANH, TENNGANH)

$\sigma_{\text{PHAI} = \text{'NAM'}} \text{ AND } \text{TENNGANH} = \text{'SP TIN'}$ ($\text{SINHVIEN} \times \text{NGANH}$)

=

($\sigma_{\text{PHAI} = \text{'NAM'}} (\text{SINHVIEN})$) \times ($\sigma_{\text{TENNGANH} = \text{'SP TIN'}} (\text{NGANH})$)



Các qui tắc biến đổi tương đương

7. Qui tắc 7:

Giả sử $L = \{A_1, \dots, A_N, B_1, \dots, B_M\}$, trong đó A_1, \dots, A_N là thuộc tính của R ; B_1, \dots, B_M là thuộc tính của S

Nếu điều kiện kết nối c chỉ chứa các thuộc tính trong L thì

$$\prod_L (R \bowtie_C S) \equiv (\prod_{A_1, A_2, A_3, \dots, A_N} (R)) \bowtie_C (\prod_{B_1, B_2, B_3, \dots, B_M} (S))$$

1. **SINHVIEN (MASV, HOTEN, PHAI, NGAYSINH, MANGANH)**
2. **NGANH (MANGANH, TENNGANH)**

$\pi_{MASV, HOTEN, MANGANH, TENNGANH} (\text{SINHVIEN} \bowtie \text{NGANH})$



SINHVIEN.MANGANH = NGANH.MANGANH

$= (\pi_{MASV, HOTEN, MANGANH} (\text{SINHVIEN})) \bowtie (\pi_{TENNGANH, MANGANH} (\text{NGANH}))$



SINHVIEN.MANGANH = NGANH.MANGANH

$(\pi_{TENNGANH, MANGANH} (\text{NGANH}))$



Các qui tắc biến đổi tương đương

7. Qui tắc 7:

Nếu điều kiện kết nối c có chứa các thuộc tính không ở trong L , giả sử là A_{N+1}, \dots, A_{N+K} của R và B_{M+1}, \dots, B_{M+P} của S thì

$$\prod_L (R \bowtie_c S) \equiv (\prod_{A_1, A_2, A_3, \dots, A_N, A_{N+1}A_{N+2}\dots A_{N+K}} (R)) \bowtie_c (\prod_{B_1, B_2, B_3, \dots, B_M B_{M+1}B_{M+2}\dots B_{M+P}} (S))$$

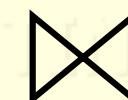
1. **SINHVIEN (MASV, HOTEN, PHAI, NGAYSINH, MANGANH)**
2. **NGANH (MANGANH, TENNGANH)**

$\pi_{MASV, HOTEN, TENNGANH} (\text{SINHVIEN} \bowtie \text{NGANH})$



SINHVIEN.MANGANH = NGANH.MANGANH

= $(\pi_{MASV, HOTEN, MANGANH} (\text{SINHVIEN})) \bowtie (\pi_{TENNGANH, MANGANH} (\text{NGANH}))$



SINHVIEN.MANGANH = NGANH.MANGANH

$(\pi_{TENNGANH, MANGANH} (\text{NGANH}))$

Các qui tắc biến đổi tương đương



7. Qui tắc 7:

Giả sử $L = \{A_1, \dots, A_N, B_1, \dots, B_M\}$, trong đó A_1, \dots, A_N là thuộc tính của R ; B_1, \dots, B_M là thuộc tính của S

$$\prod_L (R \times S) \equiv (\prod_{A_1, A_2, A_3, \dots, A_N} (R)) \times (\prod_{B_1, B_2, B_3, \dots, B_M} (S))$$

1. **SINHVIEN (MASV, HOTEN, PHAI, NGAYSINH, MANGANH)**
2. **NGANH (MANGANH, TENNGANH)**

$\pi_{MASV, HOTEN, MANGANH, TENNGANH} (\text{SINHVIEN} \times \text{NGANH})$

SINHVIEN.MANGANH = NGANH.MANGANH

= $(\pi_{MASV, HOTEN, MANGANH} (\text{SINHVIEN})) \times$

SINHVIEN.MANGANH = NGANH.MANGANH

$(\pi_{TENNGANH, MANGANH} (\text{NGANH}))$

Các qui tắc biến đổi tương đương



8. Qui tắc 8: Tính giao hoán của phép hội, giao

$$R \cup S \equiv S \cup R, R \cap S \equiv S \cap R$$

Các qui tắc biến đổi tương đương



9. Qui tắc 9: Tính kết hợp

$$(R \theta S) \theta T = R \theta (S \theta T)$$

Trong đó θ là 1 trong các phép toán $\bowtie, \times, \cap, \cup$

Các qui tắc biến đổi tương đương



10. Qui tắc 10: Tính phân phối của phép chọn σ đối với các phép toán Giao (\cap), Hội (\cup), Trừ (-)

$$\sigma_c(R \cap S) = (\sigma_c(R)) \cap (\sigma_c(S))$$

Các qui tắc biến đổi tương đương



11. Qui tắc 11: Tính phân phối của phép chiếu π đối với các phép toán Giao (\cap), Hội (\cup), Trừ (-)

$$\Pi_L(R \cap S) = (\Pi_L(R)) \cap (\Pi_L(S))$$

Các qui tắc biến đổi tương đương



12. Qui tắc 12: Các biến đổi khác

Chuyển các phép (σ , \times) thành phép \bowtie

$$\sigma_c(R \times S) = R \bowtie_c S$$

Hoặc

$$c \equiv \text{NOT}(c_1 \text{ AND } c_2) \equiv \text{NOT}(c_1) \text{ OR } \text{NOT}(c_2)$$

$$c \equiv \text{NOT}(c_1 \text{ OR } c_2) \equiv \text{NOT}(c_1) \text{ AND } \text{NOT}(c_2)$$

TỐI ƯU HÓA TRUY VẤN



- Giới thiệu về tối ưu hóa truy vấn
- Các qui tắc biến đổi tương đương
- **Giải thuật Heuristic cho tối ưu hóa truy vấn**



1. Dùng **qui tắc 1**, tách các phép chọn liên kiện thành 1 dây các phép chọn.
2. Dùng **qui tắc 2,4,6 và 10**, liên quan đến tính giao hoán của phép chọn và các phép khác để di chuyển phép chọn xuống nhánh của cây.
3. Dùng **qui tắc 5 và 9**, liên quan đến tính kết hợp của các phép 2 ngôi để sắp xếp lại các nút lá của cây để các phép chọn được ưu tiên thực hiện trước.
4. Dùng **qui tắc 12**, tổ hợp các tích Descartes với phép chọn thành các phép kết.
5. Sử dụng các **qui tắc 3, 4, 7 và 11** để tách và đẩy các phép chiếu xuống các nhánh.
6. Nhận biết từng nhánh con biểu diễn cho một nhóm các thao tác có thể được thi hành bằng một chiến lược thực hiện đơn.



Yêu cầu truy vấn: **Liệt kê danh sách (MaSV, HoTen, NgaySinh)**
sinh viên Nam thuộc ngành “Sư phạm Tin học”

1. Cách 1:

Ngôn ngữ SQL

```
SELECT MASV, HOTEN, NGAYSINH  
FROM SINHVIEN SV, NGANH N  
WHERE SV.MANGANH = N.MANGANH AND PHAI = 'NAM'  
AND TENNGANH = 'Sư phạm Tin học'
```

Ngôn ngữ ĐSQL

$$\prod_{\text{MASV}, \text{HOTEN}, \text{NGAYSINH}} (\sigma_{\text{TENNGANH} = \text{'Sư phạm Tin học'}} \wedge \text{PHAI} = \text{'NAM'} \wedge \text{SINHVIEN.MANGANH} = \text{NGANH.MANGANH} \\ (\text{SINHVIEN} \times \text{NGANH}))$$



Π

MASV, HOTEN, NGAYSINH σ TENNGANH = 'Sư phạm Tin học'

\wedge PHAI = 'NAM' \wedge SINHVIEN.MANGANH = NGANH.MANGANH

(SINHVIEN x NGANH)



Phép chọn



Phép chiếu



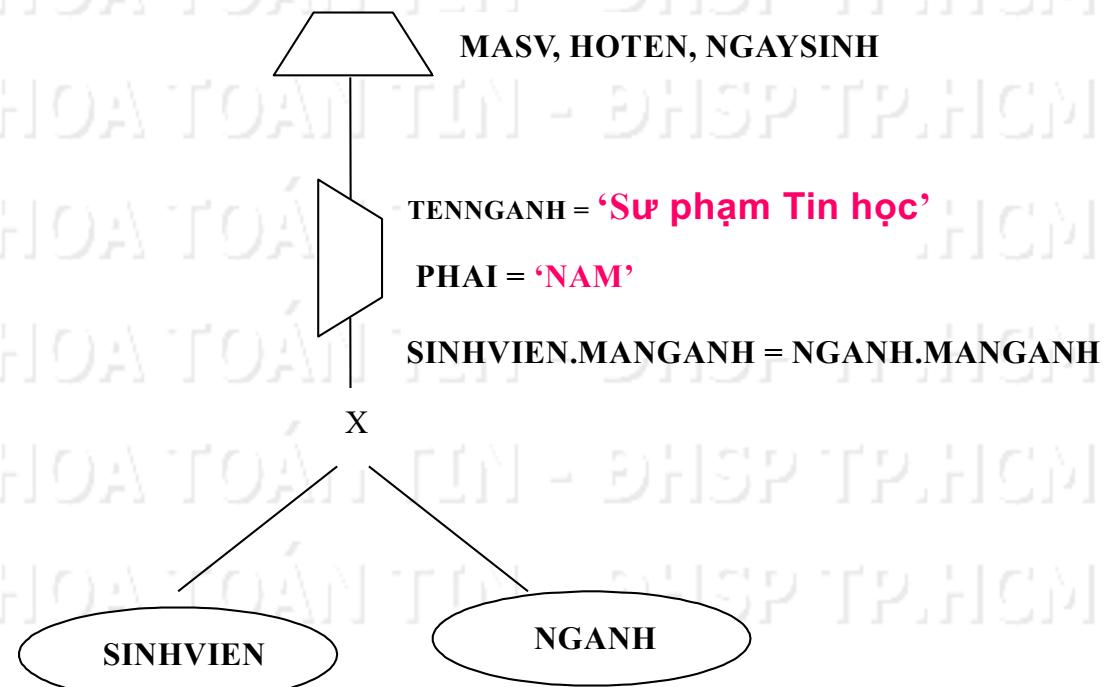
Quan hệ

X

Phép tích

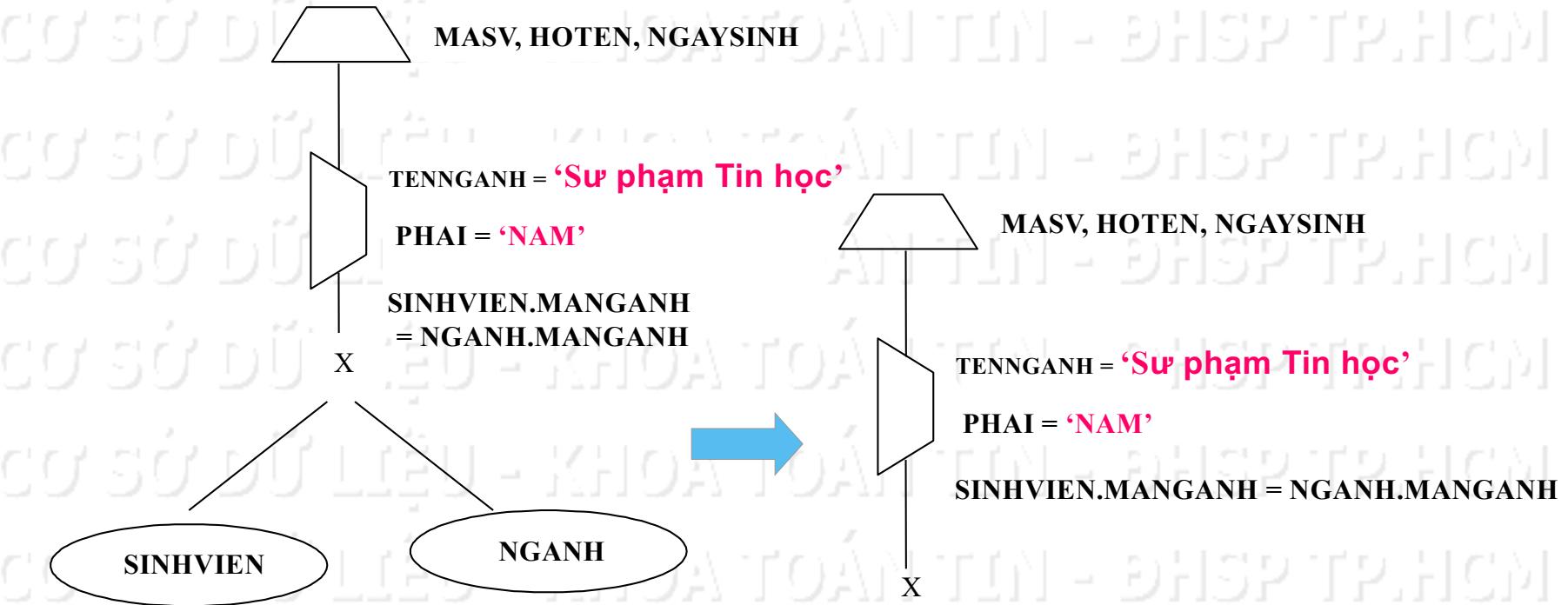


Phép kết

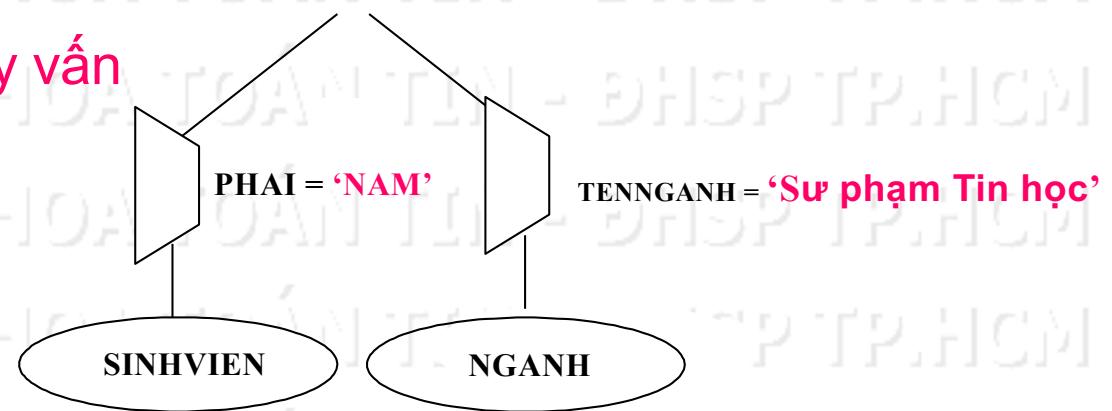


Cây biểu diễn biểu thức truy vấn

Giải thuật heuristic cho tối ưu hóa biểu thức ĐSQH

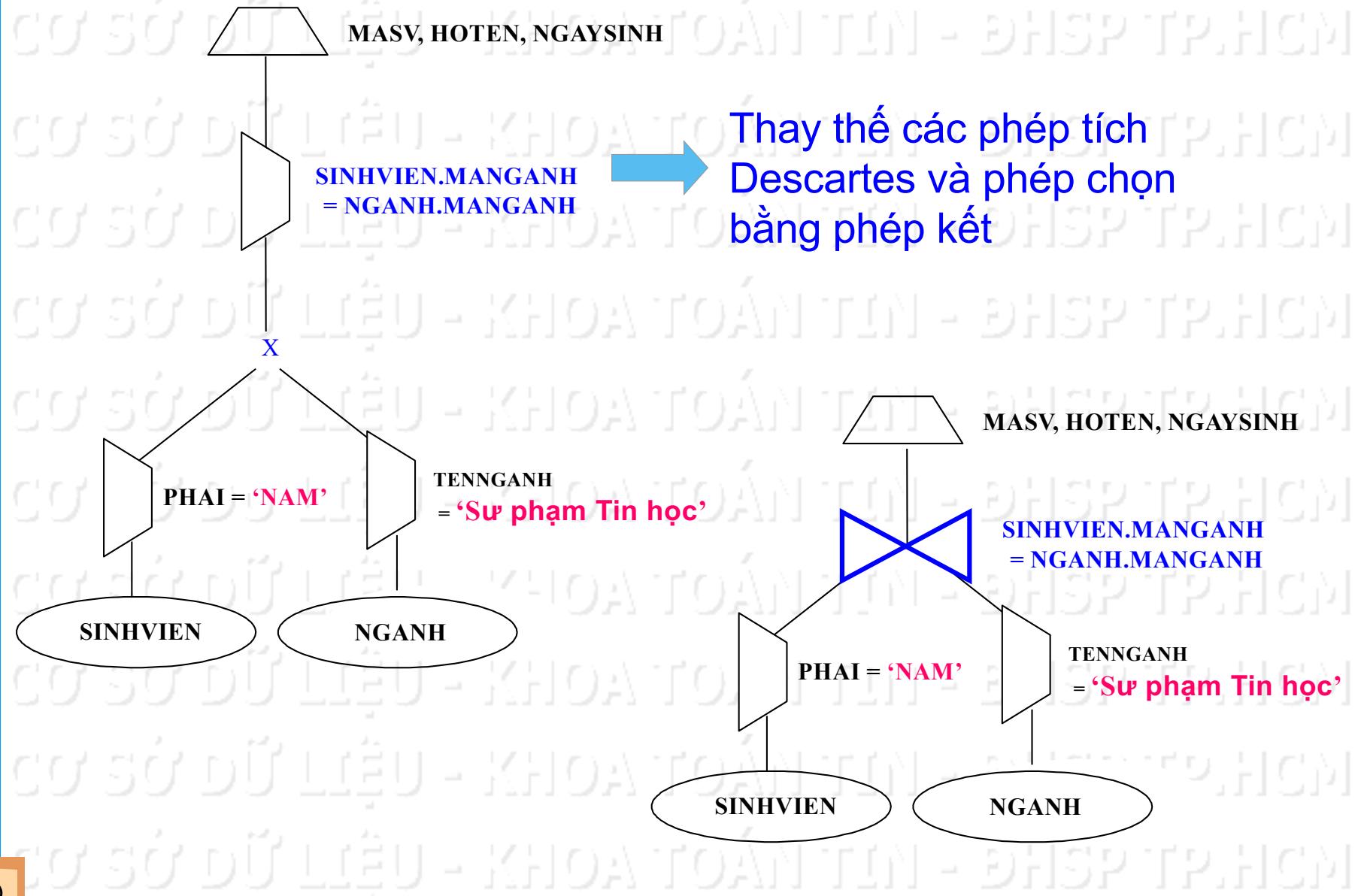


Cây biểu diễn biểu thức truy vấn

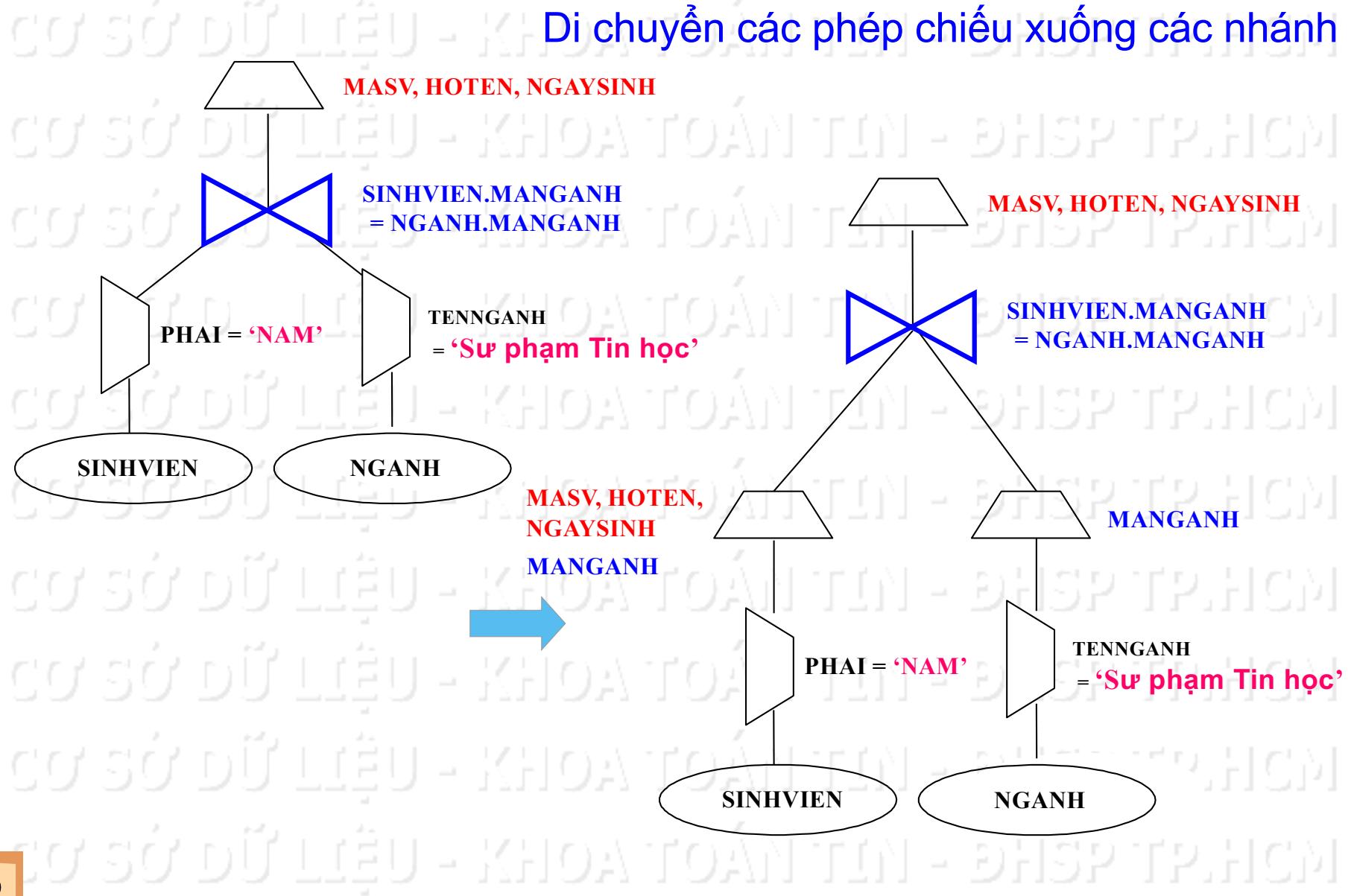


Đưa phép chọn xuống sâu các nhánh

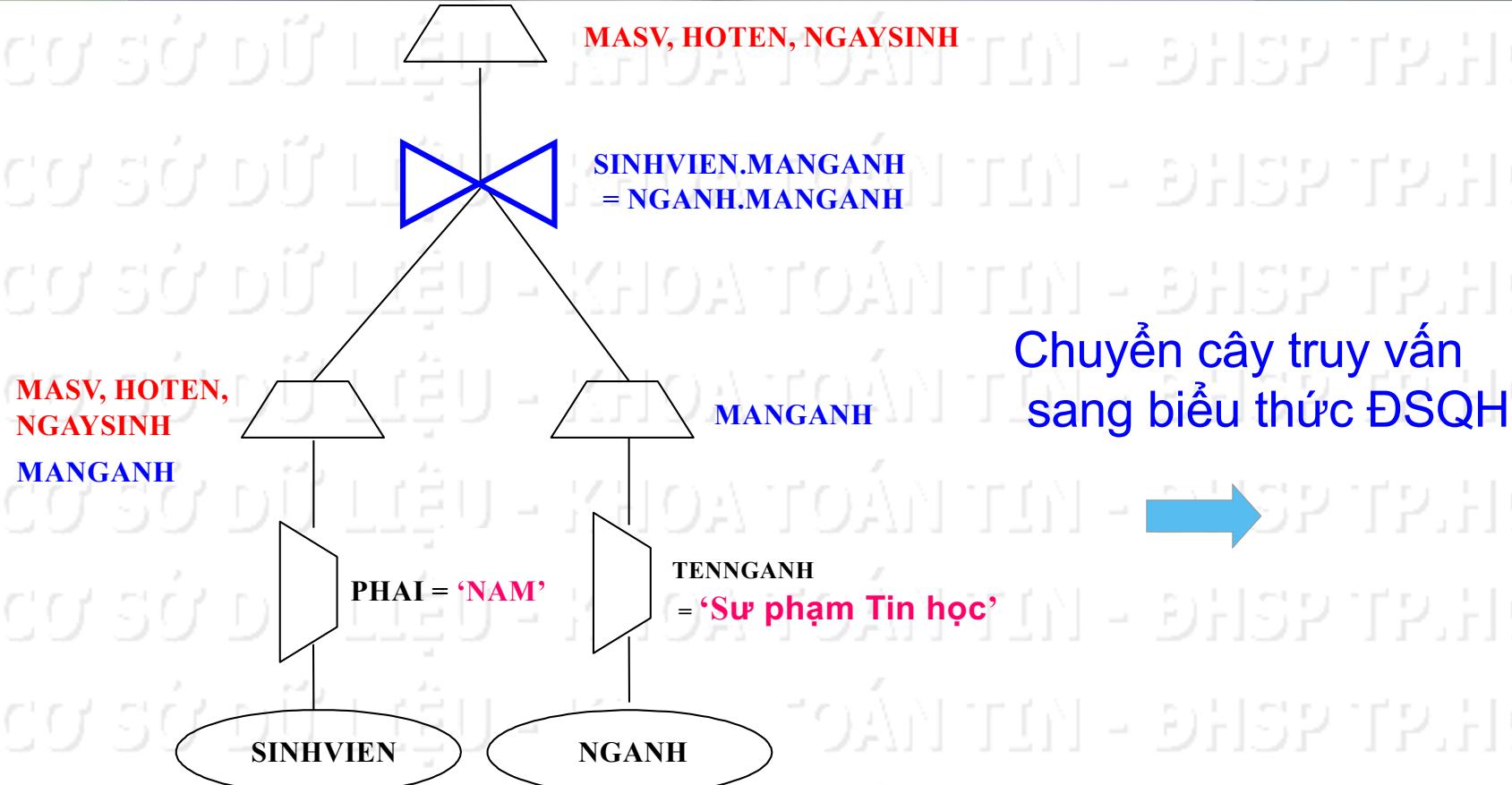
Giải thuật heuristic cho tối ưu hóa biểu thức ĐSQH



Giải thuật heuristic cho tối ưu hóa biểu thức ĐSQH



Giải thuật heuristic cho tối ưu hóa biểu thức ĐSQH



PHAI = 'NAM' (SINHVIEN) \bowtie (π_{MANGANH} ($\sigma_{\text{NGANH} = \text{'Sư phạm Tin học'}}$

$(\text{NGANH}))$)

$\text{SINHVIEN.MANGANH} = \text{NGANH.MANGANH}$



Yêu cầu truy vấn: **Liệt kê danh sách (MaSV, HoTen, NgaySinh)**
sinh viên Nam thuộc ngành “Sư phạm Tin học”

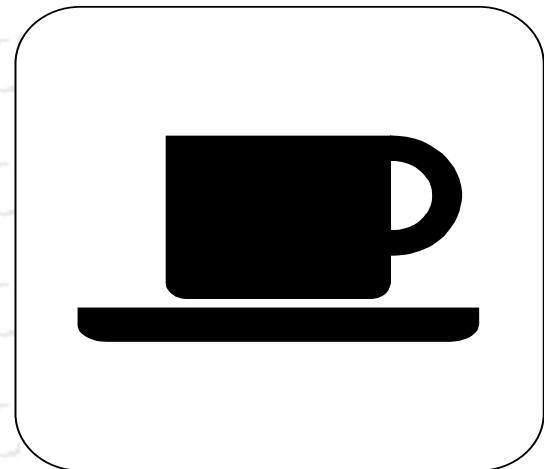
❑ Ngôn ngữ ĐSQL

$$\Pi_{MASV, HOTEN, NGAYSINH}((\pi_{MASV,HOTEN,NGAYSINH,MANGANH}(\sigma_{PHAI = 'NAM'}(SINHVIEN))) \bowtie (\pi_{MANGANH}(\sigma_{NGANH = 'Sư phạm Tin học'}(NGANH))))$$
$$SINHVIEN.MANGANH = NGANH.MANGANH$$

❑ Ngôn ngữ SQL

```
SELECT MASV, HOTEN, NGAYSINH
FROM
  (SELECT MASV, HOTEN, NGAYSINH, MANGANH
   FROM SINHVIEN
   WHERE PHAI = 'NAM') AS SV JOIN
  (SELECT MANGANH
   FROM NGANH
   WHERE TENNGANH = 'Sư phạm Tin học') AS N
  ON SV.MANGANH = N.MANGANH
```

BÀI GIẢNG HỌC PHẦN CƠ SỞ DỮ LIỆU
TỐI ƯU HÓA TRUY Y VĂN





❑ Ví dụ

```
SELECT NV.TENNV
FROM NHANVIEN NV, THUCHIEN TH, DUAN DA
WHERE (DA.TEN = "A") AND (DA.MADA = TH.MADA) AND
(NV.MANV = TH.MANV) AND (NS > "12/31/1978")
```

❑ Ngôn ngữ ĐSQL

$$\Pi_{\text{NV.TENNV}} (\sigma_{\text{DA.TTEN} = 'A' \text{ AND } \text{DA.MADA} = \text{TH.MADA} \text{ AND } \text{NV.MANV} = \text{TH.MANV} \text{ AND } \text{NS} > '12/31/1978'} (\text{NHANVIEN} \times \text{THUCHIEN} \times \text{DUAN}))$$



Π

NV.TENNV Π DA.TTEN = 'A' AND
DA.MADA = TH.MADA AND NV.MANV = TH.MANV
AND NS > '12/31/1978'

(NHANVIEN \times THUCHIEN \times DUAN))



Phép chọn



Phép chiếu



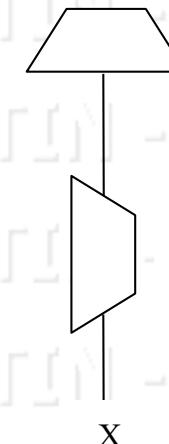
Quan hệ

\times

Phép tích

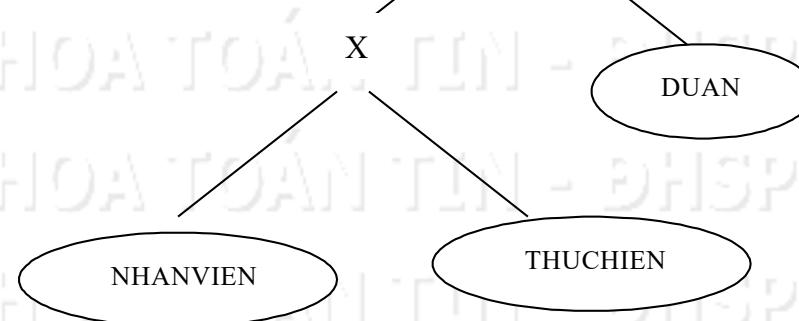


Phép kết

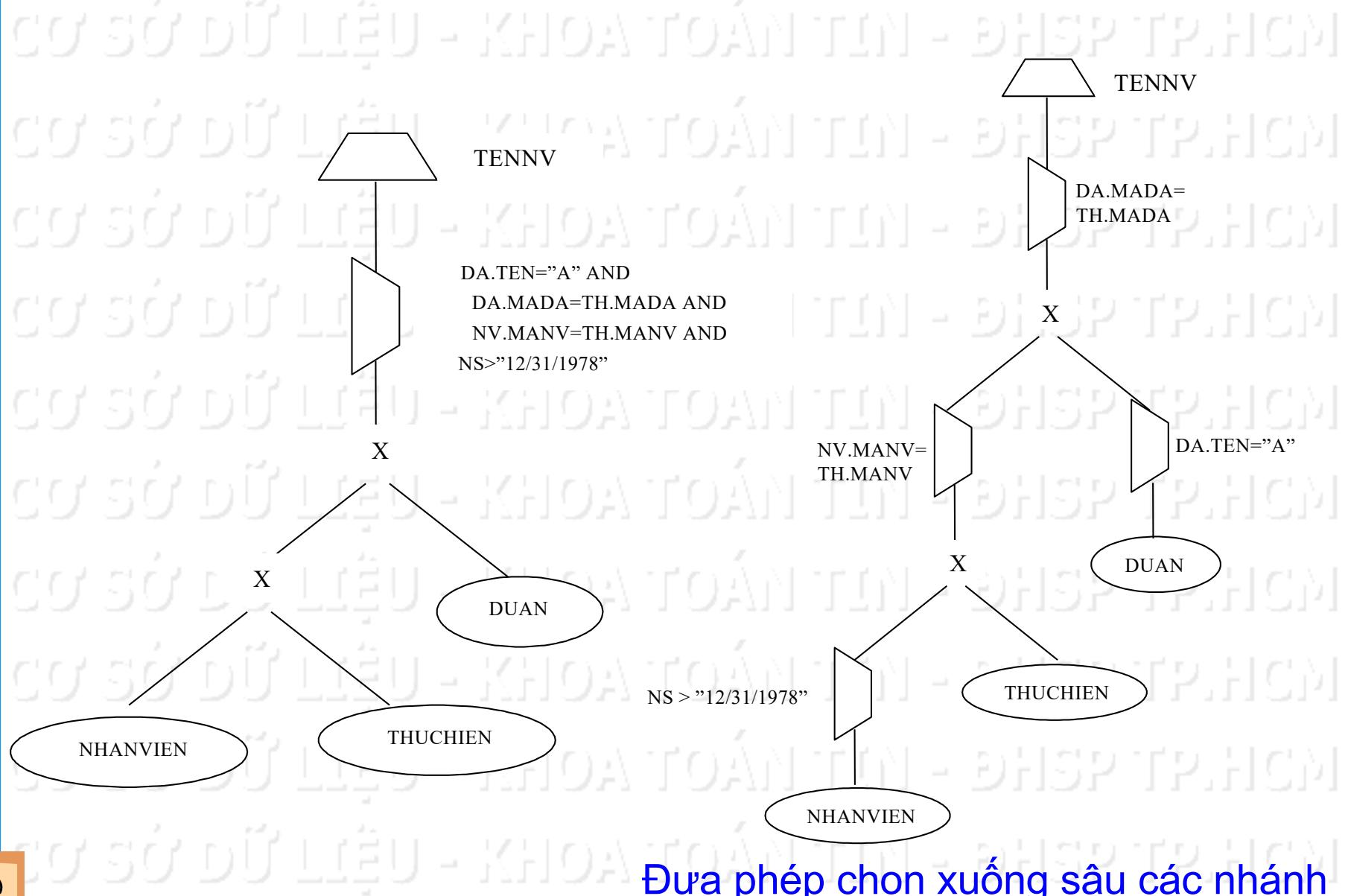


TENNV

DA.TEN="A" AND
DA.MADA=TH.MADA AND
NV.MANV=TH.MANV AND
NS>"12/31/1978"

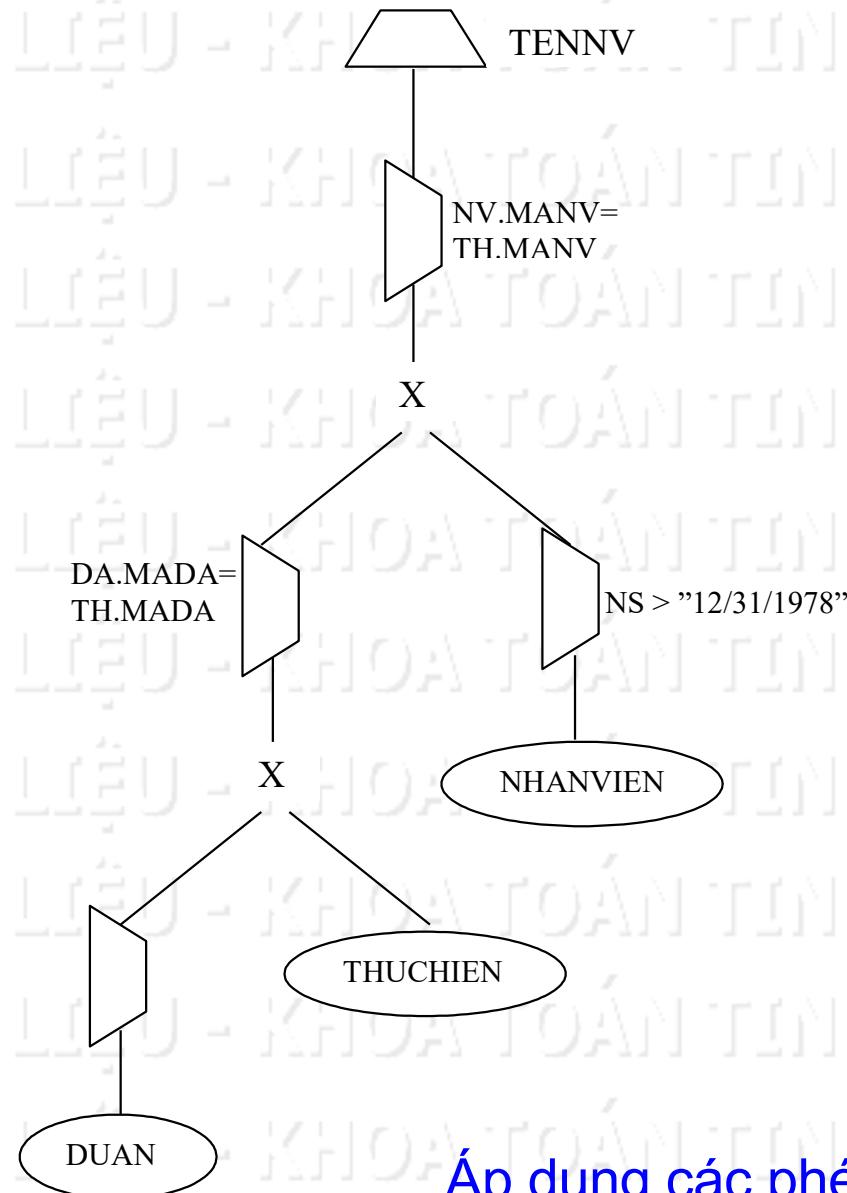


Cây biểu diễn biểu thức truy vấn

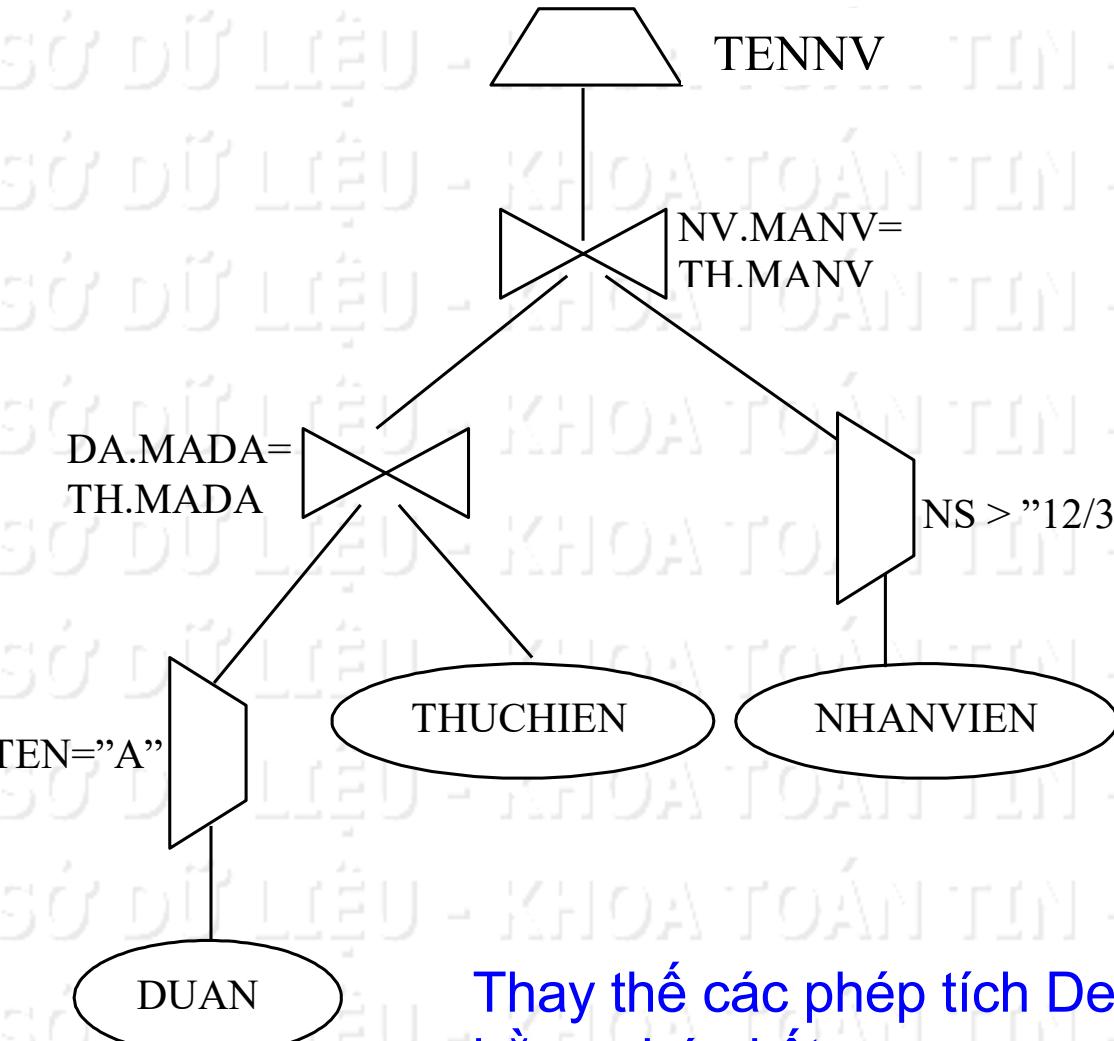


Đưa phép chọn xuống sâu các nhánh

Giải thuật heuristic cho tối ưu hóa biểu thức ĐSQH

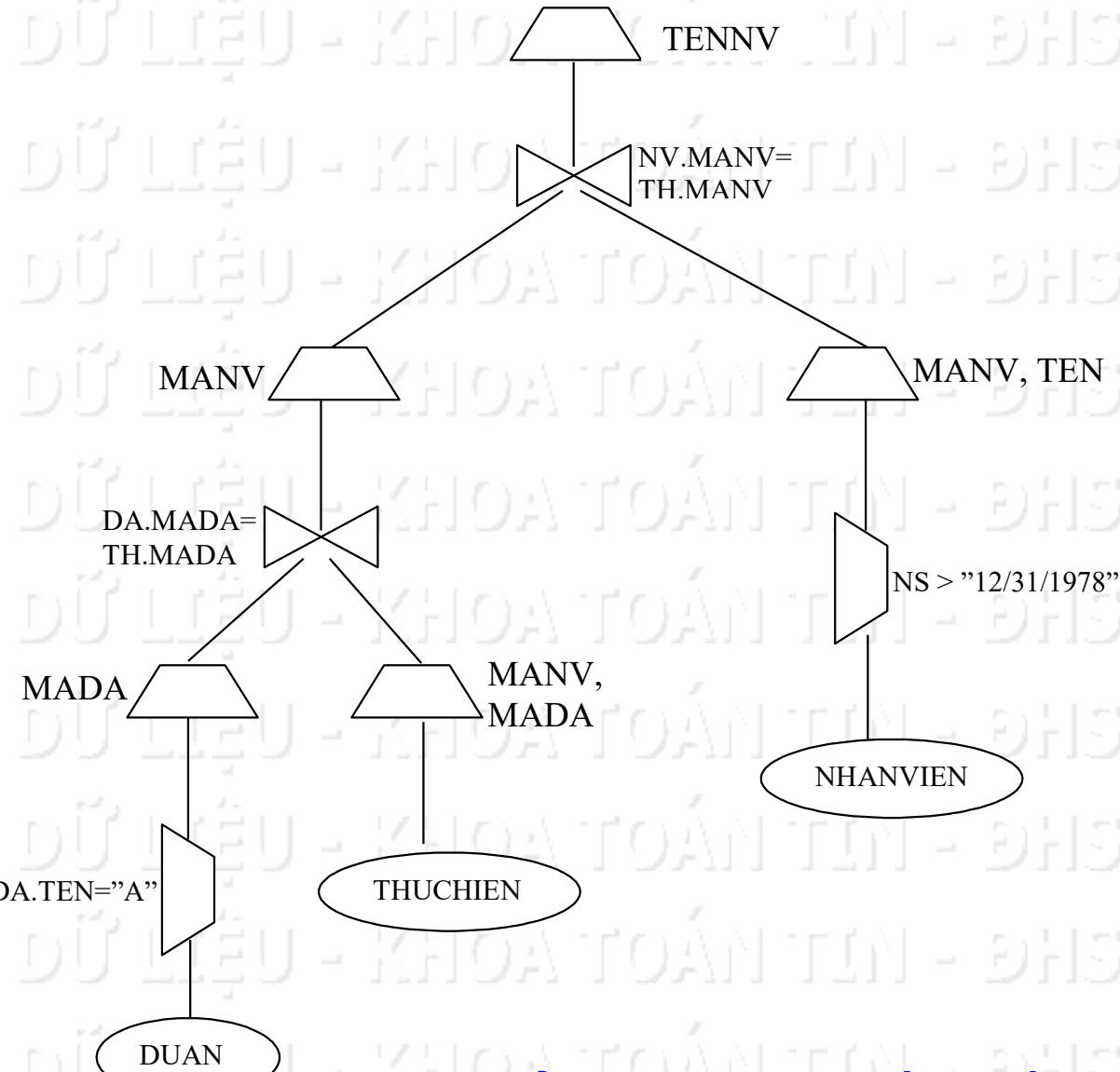


Áp dụng các phép chọn chặt chẽ trước



Thay thế các phép tích Descartes và phép chọn bằng phép kết

Giải thuật heuristic cho tối ưu hóa biểu thức ĐSQH



Di chuyển các phép chiếu xuống các nhánh



☐ Ngôn ngữ ĐSQL

$$\Pi_{NV.TENNV} ((\Pi_{MANV, TENNV} (\sigma_{NS > '12/31/1978'} (NHANVIEN)))$$


NHANVIEN,MANV = THUCHIEN.MANV

$$(\Pi_{MANV} (\Pi_{MADA} (\sigma_{TENDA='A'} (DUAN))) \bowtie (\Pi_{MADA, MANV} (THUCHIEN)))$$

DUAN.MADA = THUCHIEN.MADA