



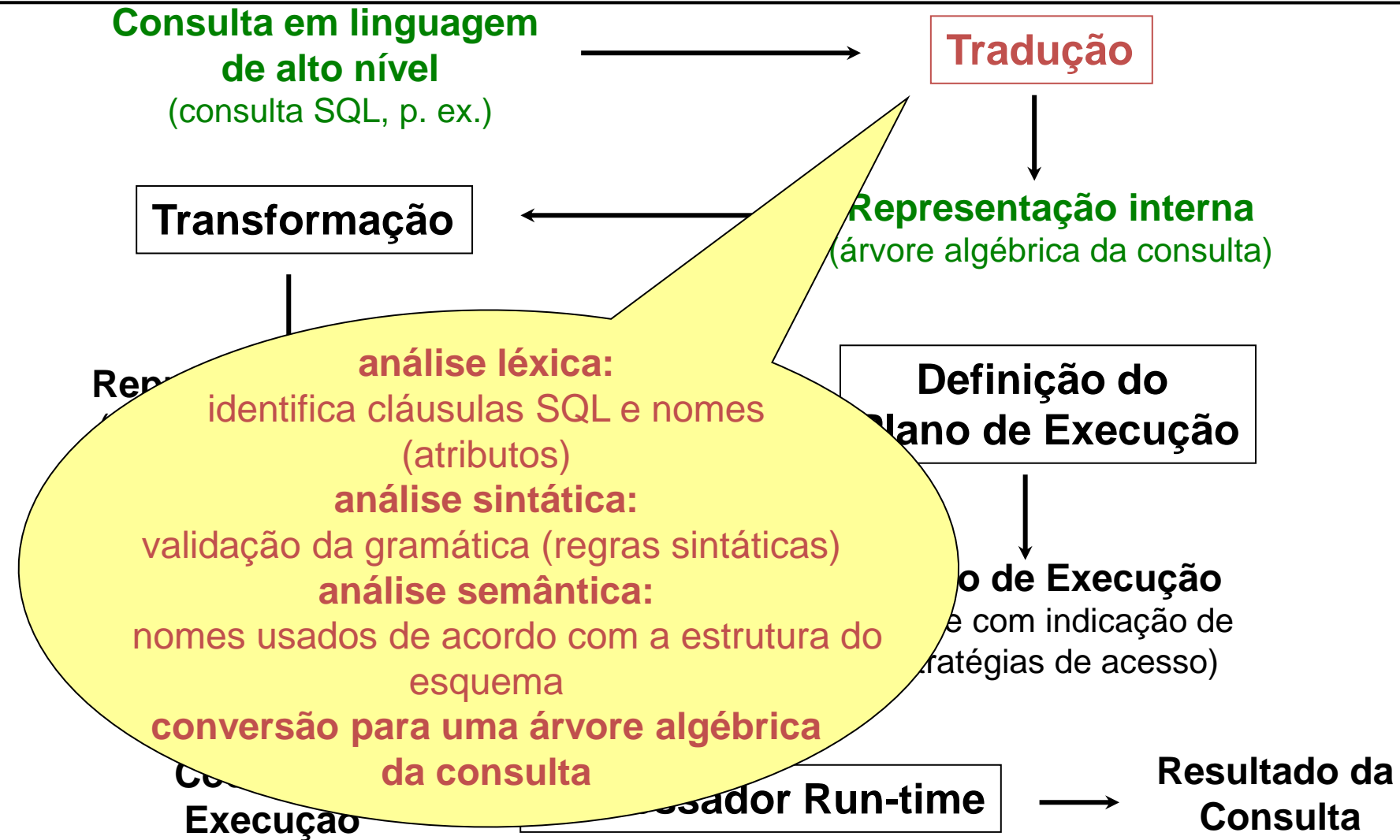
PUCRS
Pontifícia Universidade Católica
do Rio Grande do Sul

**ESCOLA
POLITÉCNICA**

Otimização Algébrica

98902-02 - Banco de Dados II

Prof. Msc. Eduardo Arruda
eduardo.arruda@pucrs.br



- Estrutura que representa o mapeamento da consulta para a álgebra relacional
 - uma expressão da álgebra relacional “estendida”
 - pode indicar alguma computação (função agregação, atributo calculado, ...)

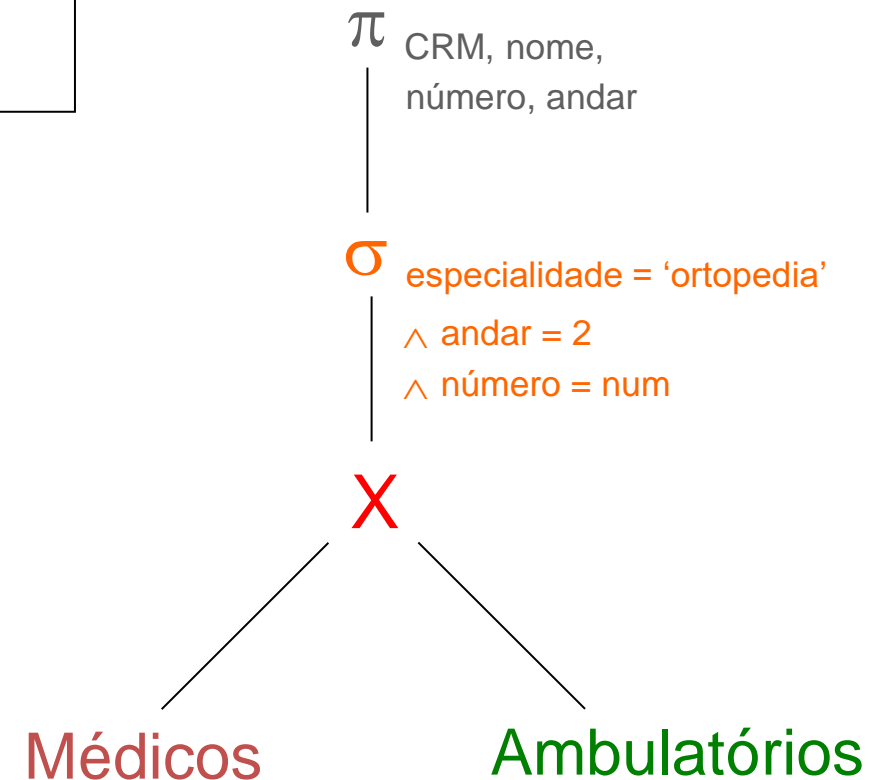
```
select CRM, nome, número, andar  
from Médicos, Ambulatórios  
where especialidade = 'ortopedia'  
and andar = 2  
and número = num
```

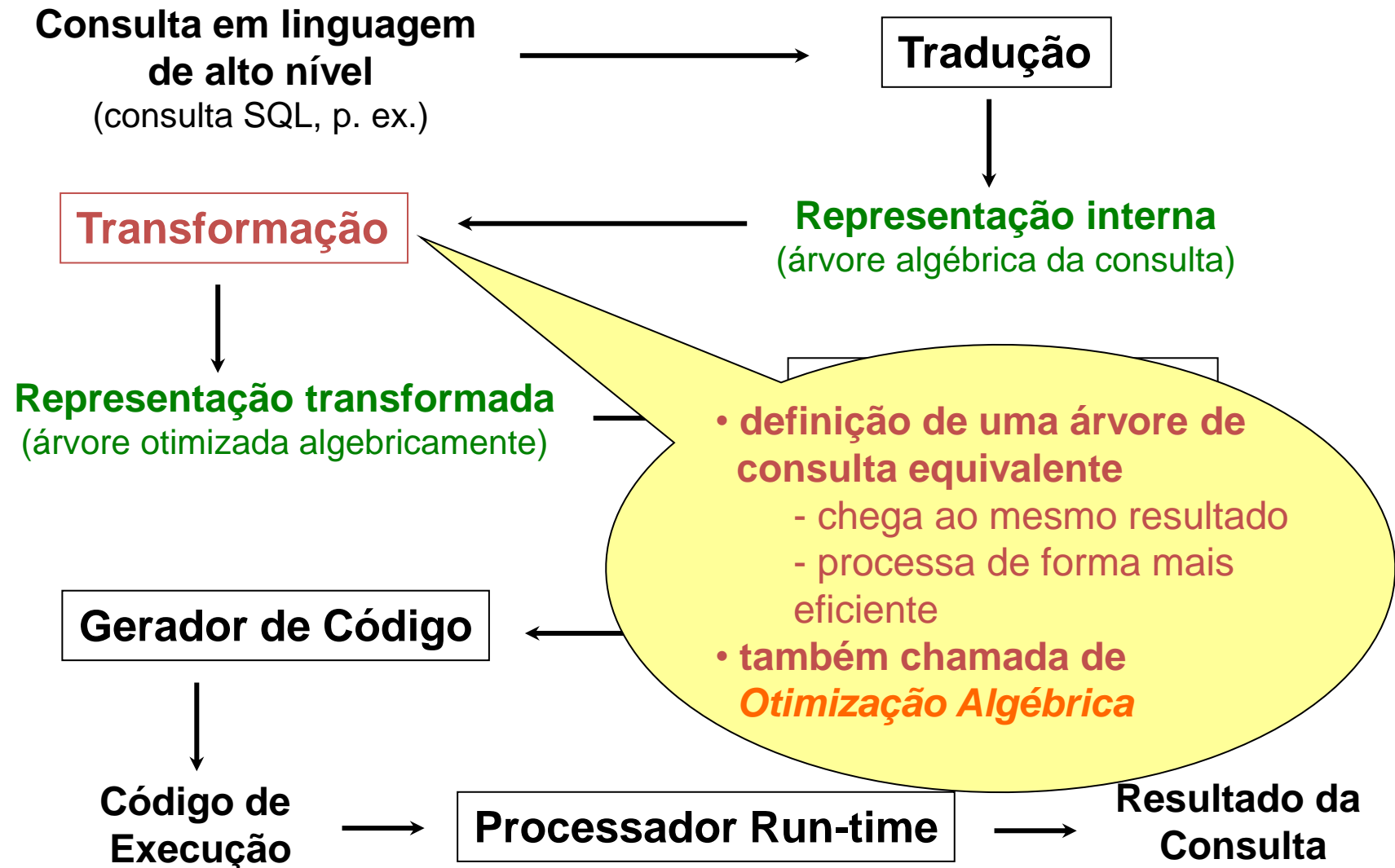
π CRM, nome, numero, andar (σ especialidade="ortopedia" and andar=2 and
numero=num (Medicos X Ambulatorios))

- Árvore da consulta:
 - Expressão em álgebra \rightarrow árvore da consulta
 - nodos folha: relações (do BD ou resultados intermediários)
 - nodos internos: operações da álgebra
- Processamento da árvore
 - nodos internos são executados quando seus operandos estão disponíveis
 - são substituídos pela relação resultante
 - a execução termina quando o nodo raiz é executado

```
select CRM, nome, número, andar  
from Médicos, Ambulatórios  
where especialidade = 'ortopedia'  
and andar = 2  
and número = num
```

π CRM, nome, numero, andar
(σ especialidade="ortopedia"
and andar=2 and numero=num
(Medicos X Ambulatorios))

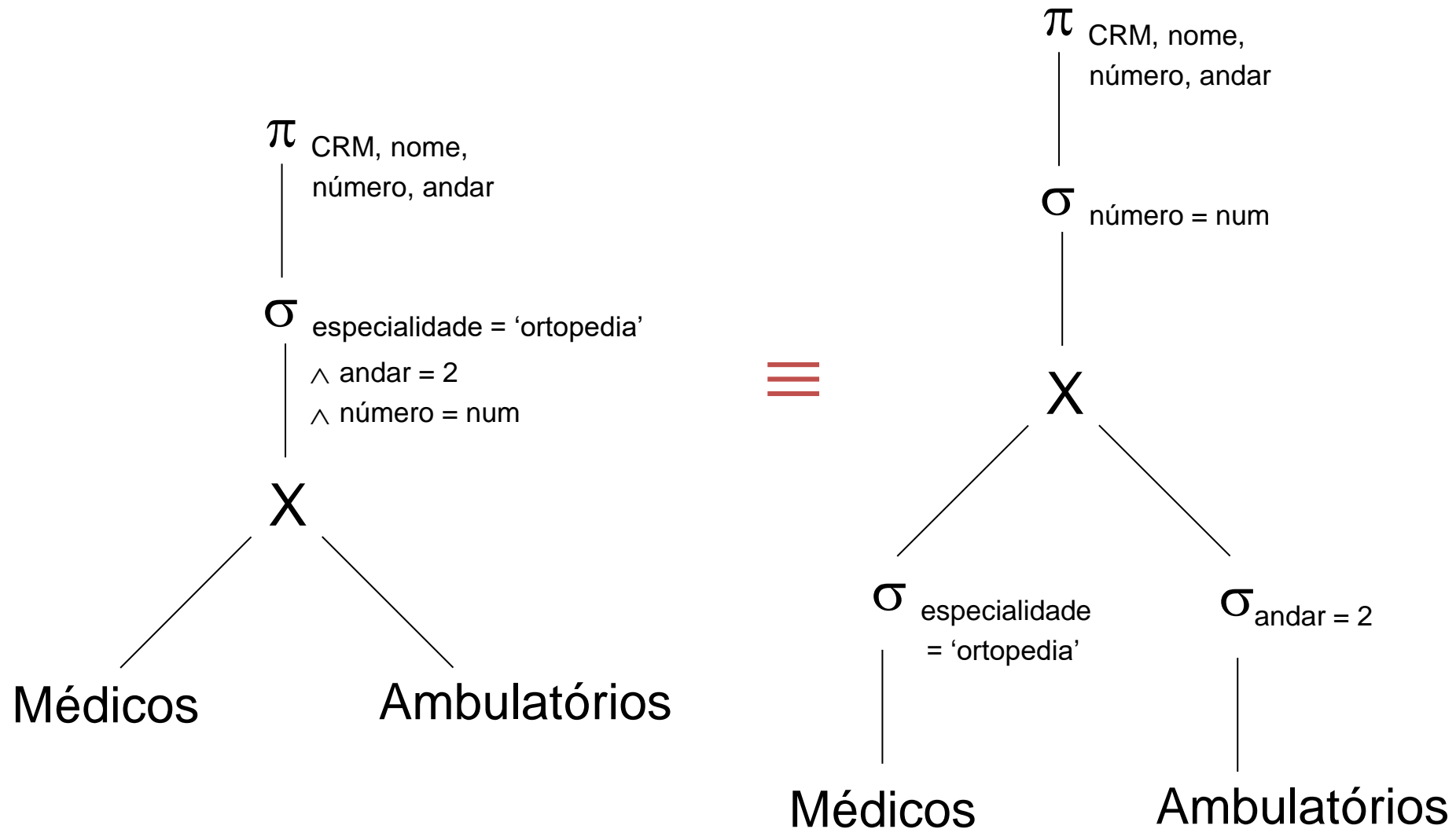


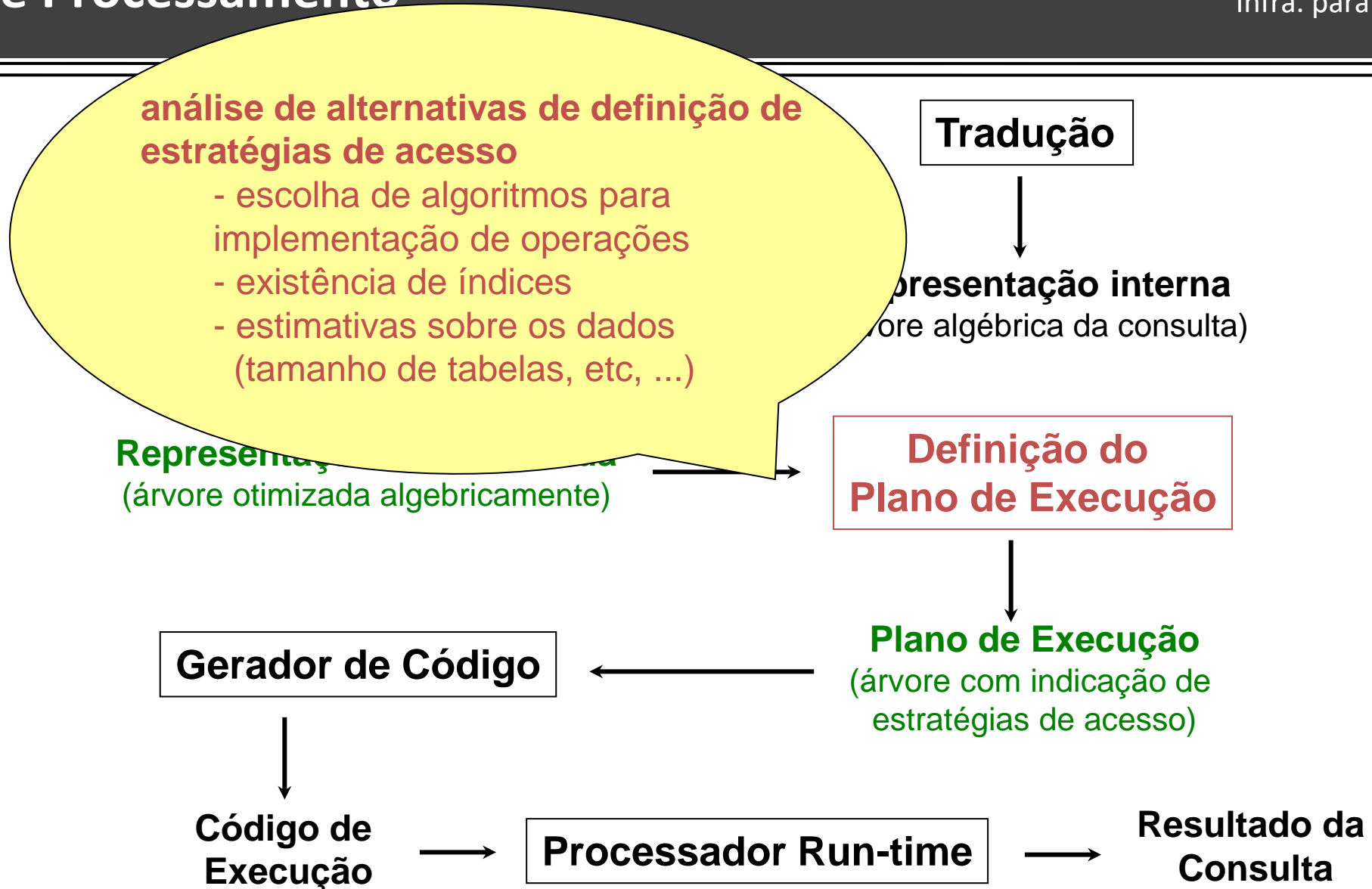


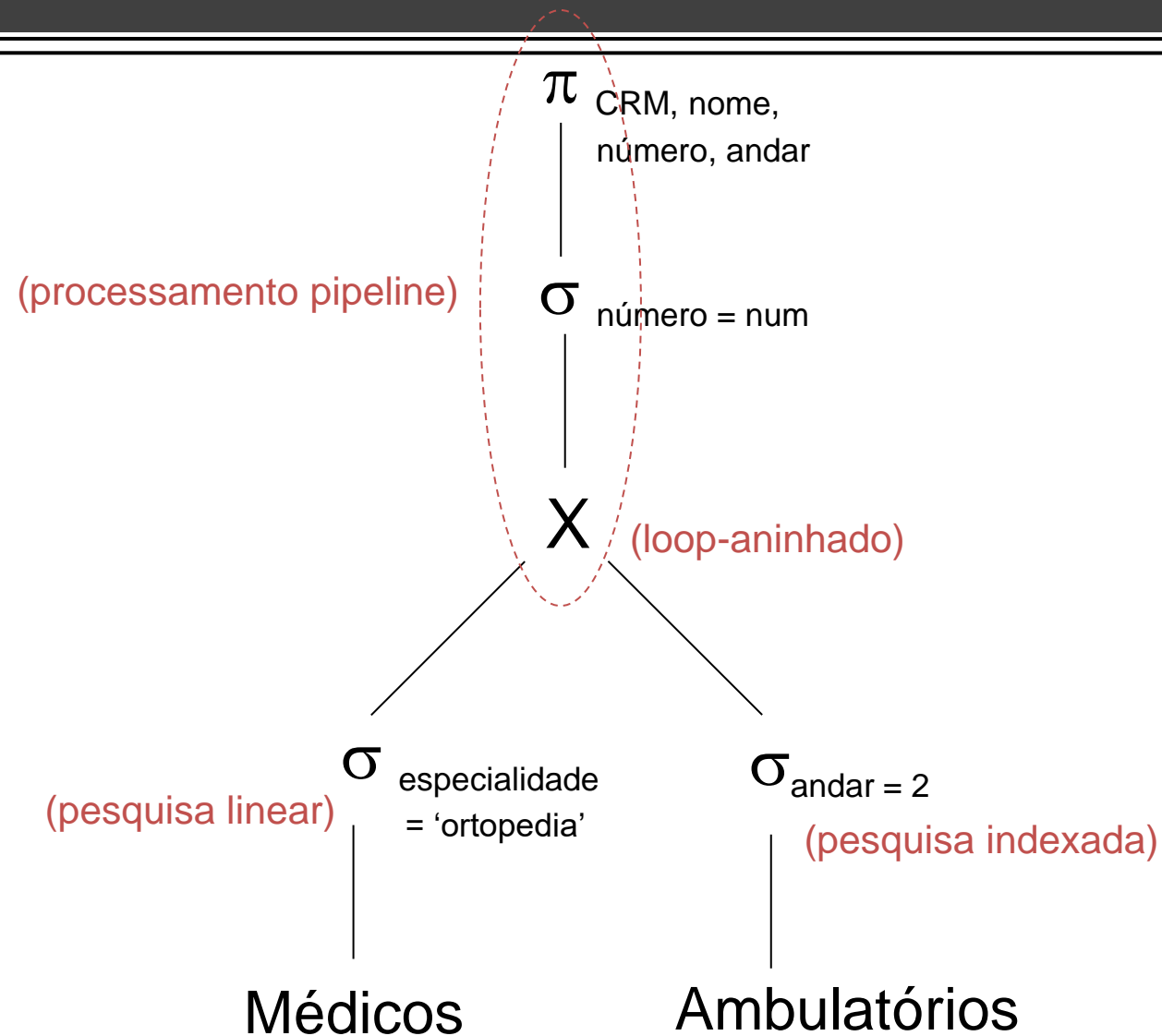
Exemplo de Árvore Equivalente

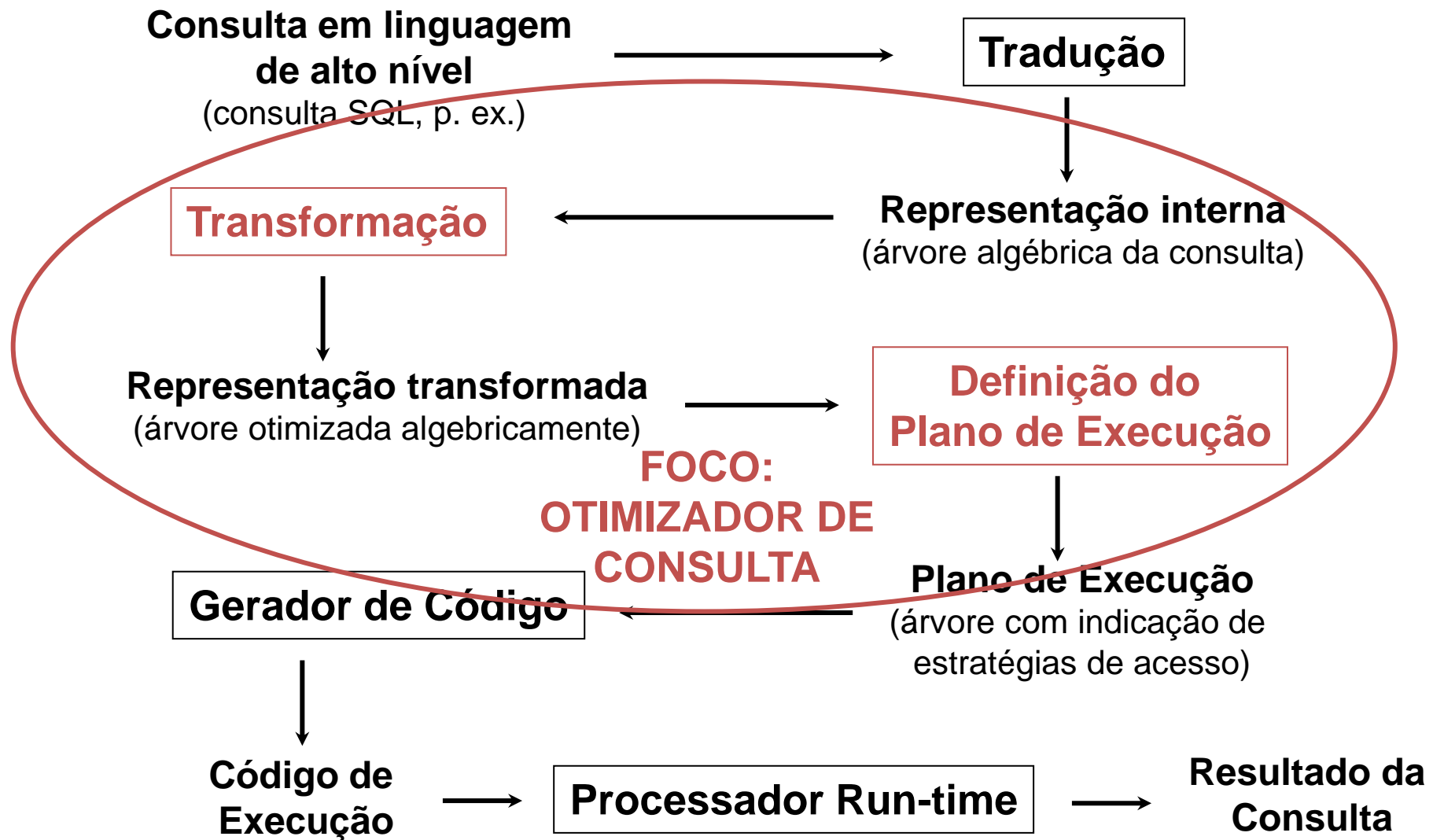
98H00-04

Infra. para Gestão de Dados







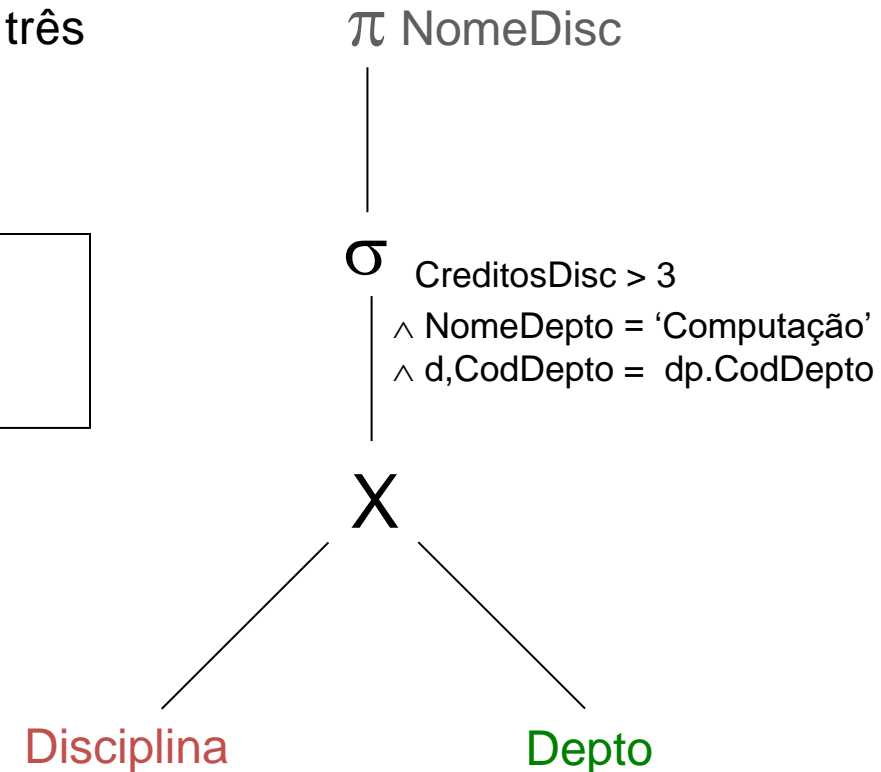


- Objetivo do passo de Transformação
 - entrada: árvore da consulta inicial
 - saída: árvore da consulta otimizada (pode manter a mesma árvore)
- Base
 - regras de equivalência algébrica
 - devem ser conhecidas pelo otimizador para que possam ser geradas transformações válidas
 - algoritmo de otimização algébrica
 - indica a ordem de aplicação das regras e de outros processamentos de otimização

- Estrutura que representa o mapeamento da consulta para a álgebra relacional
 - uma expressão da álgebra relacional “estendida”
 - pode indicar alguma computação (função agregação, atributo calculado, ...)
 - nodos folha: relações (do BD ou resultados intermediários)
 - nodos internos: operações da álgebra
- Processamento da árvore
 - nodos internos são executados quando seus operandos estão disponíveis
 - são substituídos pela relação resultante
 - a execução termina quando o nodo raiz é executado

Obtenha os nomes das disciplinas do depto de computação que possuem mais de três créditos.

π NomeDisc (σ (CreditosDisc > 3 and NomeDepto = 'Computação' and dp.CodDepto = d.CodDepto) (Disciplina d X Depto dp))



- Desenhar a árvore de consulta inicial para as expressões algébricas abaixo:
 - $\pi_{\text{CodProf}} (\sigma_{(\text{AnoSsem} = '2010/1')} (\text{ProfTurma}))$
 - $\pi_{\text{NomeDisc}, \text{NomeDepto}} (\sigma_{(\text{Disciplina.CodDepto} = \text{Depto.CodDepto})} (\text{Disciplina X Depto}))$
 - $\pi_{\text{NomeProf}} (\sigma_{(\text{NomeTit} = 'Dr.')} (\text{Titulacao} \bowtie \text{Professor}))$
 - $\pi_{\text{NomeDisc}} (\sigma_{(\text{CreditosDisc} > 3 \text{ and } \text{NomeDepto} = 'Computação')} (\text{Disciplina} \bowtie \text{Depto}))$
 - $\pi_{\text{NomeProf}, \text{NomeTit}} (\sigma_{(\text{AnoSsem} = '2009/2')} (\sigma_{(\text{Titulacao.CodTit} = \text{Professor.CodTit})} (\text{Titulacao X } (\sigma_{(\text{Professor.CodProf} = \text{ProfTurma.CodProf})} (\text{Professor X ProfTurma)))))))$

- Cascata de Seleções: uma seleção conjuntiva pode ser quebrada em uma cascata (seqüência) de operações de seleção individuais

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

Exemplo:

$$\begin{aligned} &\sigma_{\text{salario} > 1000 \wedge \text{idade} < 40}(\text{Professor}) \\ &\quad \equiv \\ &\sigma_{\text{salario} > 1000}(\sigma_{\text{idade} < 40}(\text{Professor})) \end{aligned}$$

- Comutatividade de Seleções

$$\sigma_{c_1}(\sigma_{c_2}(R)) \equiv \sigma_{c_2}(\sigma_{c_1}(R))$$

Exemplo:

$$\begin{aligned} \sigma_{\text{salario} > 1000}(\sigma_{\text{idade} < 40}(\text{Professor})) \\ \equiv \\ \sigma_{\text{idade} < 40}(\sigma_{\text{salario} > 1000}(\text{Professor})) \end{aligned}$$

- Cascata de Projeções: em uma sequência de projeções, todas, exceto a última, podem ser ignoradas

$$\pi_{\text{listaAtributos1}}(R) \equiv \pi_{\text{listaAtributos1}}(\pi_{\text{listaAtributos2}}(\dots(\pi_{\text{listaAtributosN}}(R))))$$

Exemplo:

$$\begin{aligned} &\pi_{\text{nome}}(\text{Professor}) \\ &\quad \equiv \\ &\pi_{\text{nome}}(\pi_{\text{nome, cidade}}(\pi_{\text{nome, cidade, salario}}(\text{Professor}))) \end{aligned}$$

- Comutatividade de Seleções e Projeções: se a condição de seleção c envolver apenas os atributos a_1, a_2, \dots, a_n da lista de projeções, as duas operações podem ser comutadas.



$$\pi_{a_1, a_2, \dots, a_n}(\sigma_c(R)) \equiv \sigma_c(\pi_{a_1, a_2, \dots, a_n}(R))$$

Exemplo:

$$\begin{aligned} \pi_{\text{nome}, \text{salario}}(\sigma_{\text{nome}=\text{"Ana"} \wedge \text{salario} > 1000}(\text{Professor})) \\ \equiv \\ \sigma_{\text{nome}=\text{"Ana"} \wedge \text{salario} > 1000}(\pi_{\text{nome}, \text{salario}}(\text{Professor})) \end{aligned}$$

- Comutatividade de Operações Produtórias (“X”)

$$R \text{ “X” } S \equiv S \text{ “X” } R$$

- Por “X” entenda-se: **X** ou  ou _c
- Embora a ordem dos atributos possa não ser a mesma o “significado” é o mesmo

Exemplo:

Professor X Depto

≡

Depto X Professor

- Comutatividade de Seleções e Operações Produtórias: se todos os atributos da seleção c envolvem apenas os atributos de uma das relações participantes do produto (R), as duas operações podem ser comutadas como segue:

$$\sigma_c (R \text{ "X" } S) \equiv (\sigma_c (R)) \text{ "X" } S$$

Exemplo:

$$\begin{aligned} &\sigma_{\text{salario} > 1000 \wedge \text{cidade} = \text{"POA"}} (\text{Professor X Depto}) \\ &\quad \equiv \\ &(\sigma_{\text{salario} > 1000 \wedge \text{cidade} = \text{"POA"}} (\text{Professor})) \text{ X Depto} \end{aligned}$$

- Associatividade de Operações Produtórias e de Conjunto (“X”)

$$(R \text{ “X” } S) \text{ “X” } T \equiv R \text{ “X” } (S \text{ “X” } T)$$

- por “X” entenda-se: \times ou \bowtie_c ou \bowtie ou \cup ou \cap

Exemplo:

$$\begin{aligned} &(\text{Professor X Depto}) \text{ X Disciplina} \\ &\equiv \\ &\text{Professor X (Depto X Disciplina)} \end{aligned}$$

- Cuidado com a associatividade da operação de Junção:
 - A associatividade só ocorre entre tabelas que podem sofrer a junção

Exemplo:

$$\begin{aligned} & (Peca \bowtie Embarq) \bowtie Fornecedor \\ & \quad \equiv \\ & Peca \bowtie (Embarq \bowtie Fornecedor) \\ & \quad \not\equiv \\ & (Peca \bowtie Fornecedor) \bowtie Embarq \end{aligned}$$

Não existe como fazer junção entre Peca e Fornecedor

- Comutatividade de Seleção e Operações de Conjunto (“o”)

$$\sigma_c(R \text{ “o” } S) \equiv (\sigma_c(R)) \text{ “o” } (\sigma_c(S))$$

- por “o” entenda-se: \cup ou \cap ou -

Exemplo:

$$\begin{aligned} &\sigma_{\text{salario} > 2000} (\text{Professor} \cup \text{Aluno}) \\ &\quad \equiv \\ &(\sigma_{\text{salario} > 2000} (\text{Professor})) \cup (\sigma_{\text{salario} > 2000} (\text{Aluno})) \end{aligned}$$

- Comutatividade de Projeção e União

$$\pi_{\text{listaAtributos}}(R \cup S) \equiv (\pi_{\text{listaAtributos}}(R)) \cup (\pi_{\text{listaAtributos}}(S))$$

supõe-se atributos de R e S renomeados para manter os mesmos nomes

Exemplo:

$$\begin{aligned} &\pi_{\text{nome, salario}}(\text{Professor} \cup \text{Aluno}) \\ &\quad \equiv \\ &(\pi_{\text{nome, salario}}(\text{Professor})) \cup (\pi_{\text{nome, salario}}(\text{Aluno})) \end{aligned}$$

- Conversão de uma seqüência (σ, X) em junção

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

Exemplo:

$$\sigma_{\text{codDeppto}=\text{numDeppto}}(\text{Professor} \times \text{Deppto}) \equiv \text{Professor} \bowtie_{\text{codDeppto}=\text{numDeppto}} \text{Deppto}$$

1. Cascata de Seleções

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

2. Comutatividade de Seleções

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

3. Cascata de Projeções

$$\pi_{\text{listaAtributos1}} (R) \equiv \pi_{\text{listaAtributos1}} (\pi_{\text{listaAtributos2}} (\dots (\pi_{\text{listaAtributosN}} (R))))$$

4. Comutatividade de Seleções e Projeções

$$\pi_{a1, a2, \dots, an} (\sigma_c (R)) \equiv \sigma_c (\pi_{a1, a2, \dots, an} (R))$$

5. Comutatividade de Operações Produtórias

$$R \text{ "X" } S \equiv S \text{ "X" } R$$

6. Comutatividade de Seleções e Operações Produtórias

$$\sigma_c (R \text{ "X" } S) \equiv (\sigma_c (R)) \text{ "X" } S$$

7. Associatividade de Operações Produtórias e de Conjunto

$$(R \text{ "X" } S) \text{ "X" } T \equiv R \text{ "X" } (S \text{ "X" } T)$$

8. Comutatividade de Seleção e Operações de Conjunto

$$\sigma_c (R \text{ "o" } S) \equiv (\sigma_c (R)) \text{ "o" } (\sigma_c (S))$$

9. Comutatividade de Projeção e União

$$\pi_{\text{listaAtributos}} (R \cup S) \equiv (\pi_{\text{listaAtributos}} (R)) \cup (\pi_{\text{listaAtributos}} (S))$$

10. Conversão de uma seqüência (σ, X) em junção

$$\sigma_c (R \text{ X } S) \equiv R \bowtie S$$

1. Separar seleções conjuntivas em operações de seleção isoladas
2. Mover as operações de seleção para baixo, para serem executadas o mais cedo possível
3. Reorganizar a árvore, de forma que as relações folhas com seleções mais restritivas sejam executadas primeiro
4. Substitua produtos cartesianos seguidos de seleção por junção
5. Separe e mova para baixo os operadores de projeção, para que sejam executados o mais cedo possível

- Composto de 5 grandes passos
- Passo 1
 - desmembrar operações de seleção
 - maior flexibilidade para mover seleções
- Passo 2
 - mover seleções para níveis inferiores da árvore o máximo possível

- Passo 3
 - mudar de posição sub-árvores envolvidas em operações produtórias
 - objetivos
 - combinar prioritariamente sub-árvores com menor número de dados
 - investigar sub-árvores com seleções mais restritivas
 - como saber quais as seleções mais restritivas?
 - análise do grau de seletividade de um predicado
 - estatística geralmente mantida no Dicionário de Dados

- Definido pela seguinte razão
 - $GS_{ai}(R) = tp(R) / |R|$
 - $tp(R)$ é o número de tuplas que satisfazem o predicado aplicado sobre um atributo ai em uma relação R
 - $|R|$ é o número de tuplas em R ($GS \in [0,1]$)
- $GS_{ai}(R)$ pequeno (≈ 0) \Rightarrow seleção mais restritiva

- Passo 4
 - otimizar operações produtórias (conversão de σ, X) em junção)
- Passo 5
 - desmembrar e mover projeções para níveis inferiores da árvore, tanto quanto possível, definindo novas projeções conforme se faça necessário

- Consulta sobre um Banco de Dados de empregados

Empregado = (NumEmp, NomeEmp, Sobrenome, Salário, DataNasc, DeptoNum)

Departamento = (DeptoNum, NomeDepto, Orçamento)

Projeto = (ProjNum, NomeProj, DeptoNum)

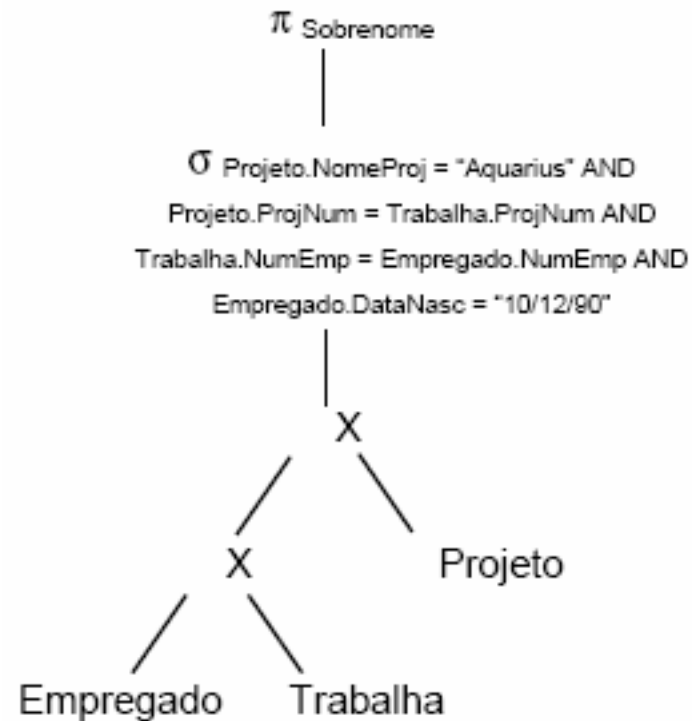
Trabalha = (NumEmp, ProjNum)

- Obter os sobrenomes dos empregados que trabalharam no projeto Aquarius e que nasceram em 10/12/1990

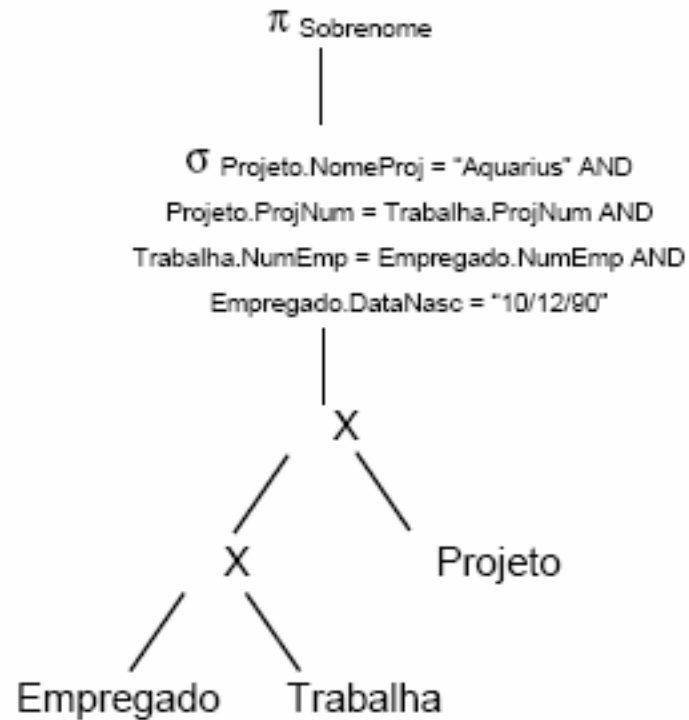

```
SELECT Sobrenome
FROM
    Empregado
    INNER JOIN Trabalha
        ON Empregado.NumEmp = Trabalha.NumEmp
    INNER JOIN Projeto
        ON Trabalha.ProjNum = Projeto.ProjNum
WHERE
    Projeto.NomeProj = 'Aquarius' AND
    Empregado.DataNasc = '10/12/1990'
```

$\pi_{\text{Empregado.Sobrenome}} ($
 $\sigma_{\text{(Projeto.NomeProj = "Aquarius" AND$
 Projeto.ProjNum = Trabalha.ProjNum AND
 Trabalha.NumEmp = Empregado.NumEmp AND
 Empregado.DataNasc = "10/12/90")
 (Empregado X Trabalha X Projeto)

1 Árvore de Consulta Algébrica



1 Árvore de Consulta Algébrica

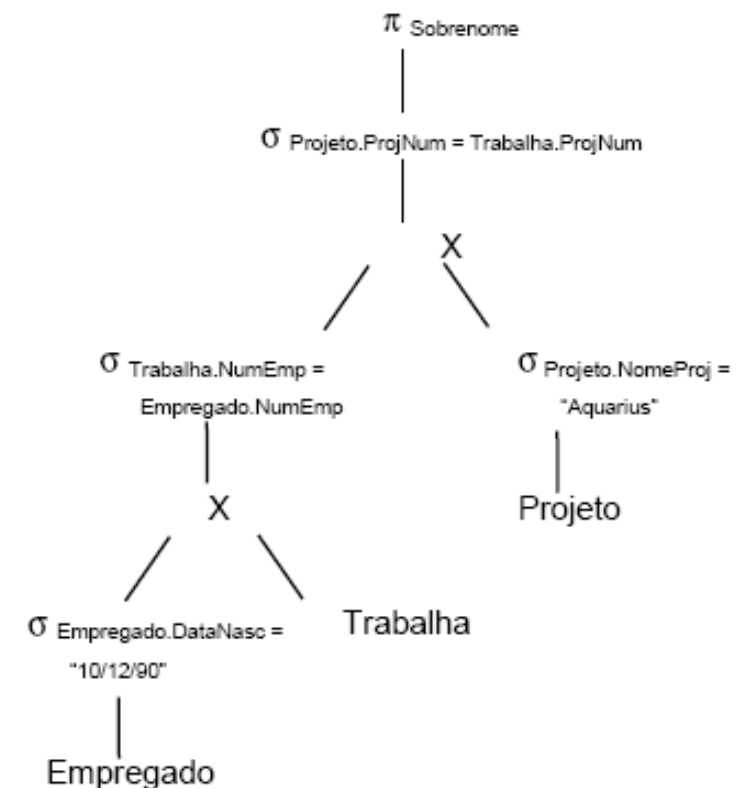


Passo 1: Desmembrar operações de seleção
Maior flexibilidade para mover seleções

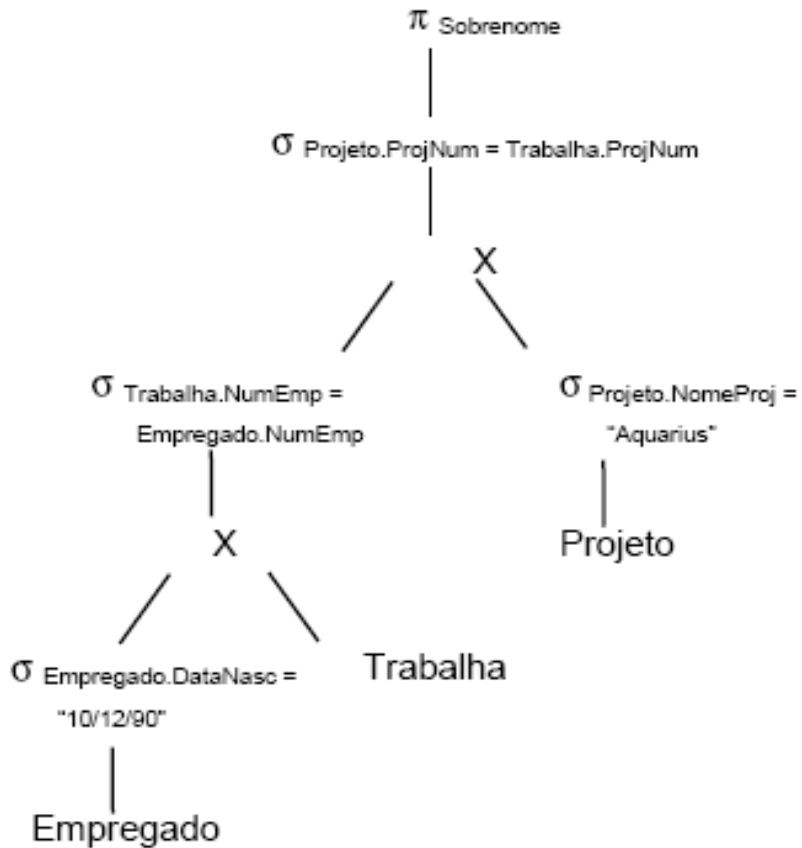


Passo 2: Mover seleções para níveis inferiores da árvore o máximo possível
Diminuir resultados intermediários

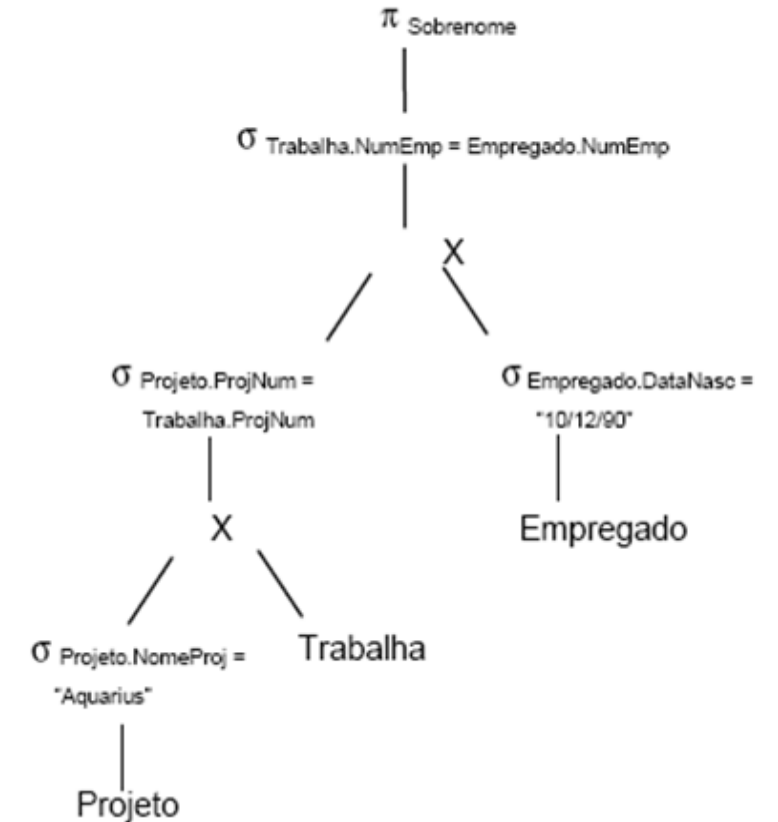
2 Árvore de Consulta Transformada Passo 1 e Passo 2



2 Árvore de Consulta Transormado Passo 1 e Passo 2



3 Árvore de Consulta após Passo 3



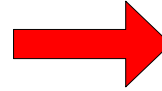
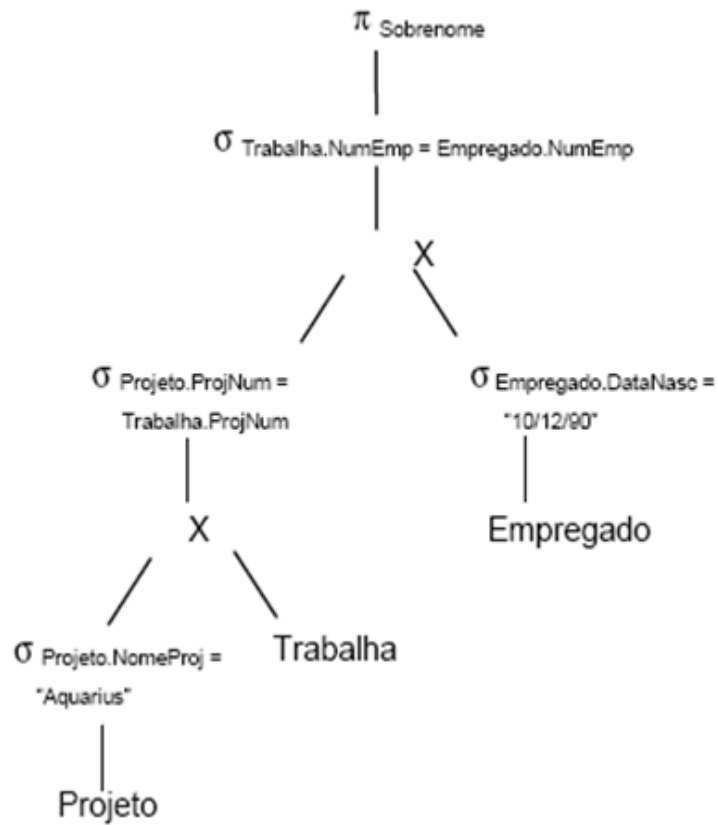
Passo 3: Mudar de posição sub-árvores envolvidas em operações produtórias.

Combinar prioritariamente sub-árvores com menor número de dados

Normalmente utiliza-se primeiro sub-árvores com seleções mais restritivas

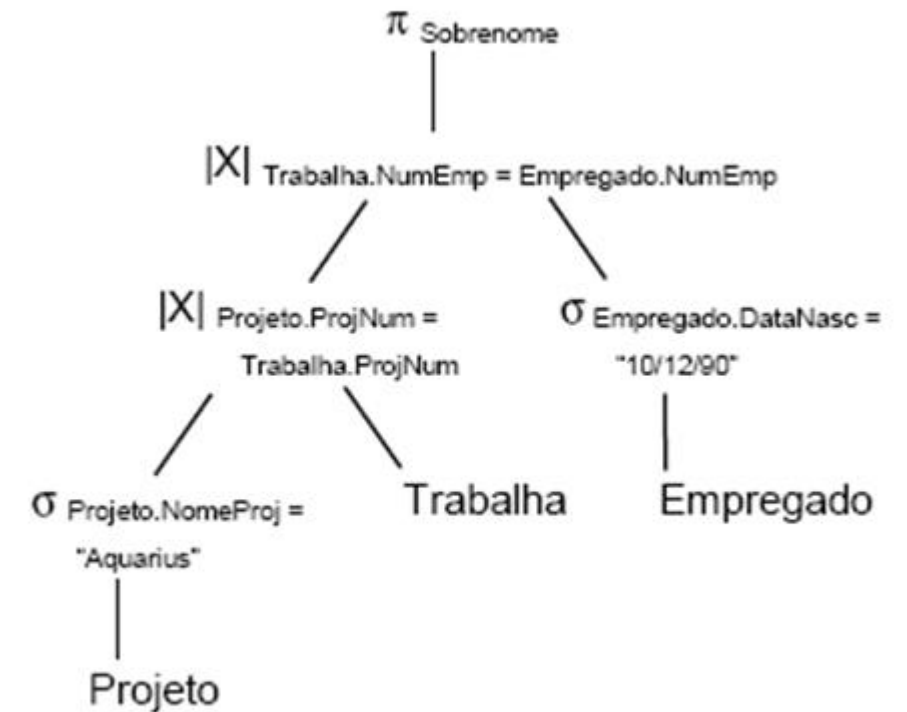
3

Árvore de Consulta após Passo 3



4

Árvore de Consulta após Passo 4



Passo 4: Otimizar operações produtórias (converter de produto cartesiano para junção)

4 Árvore de Consulta após Passo 4

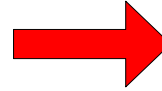
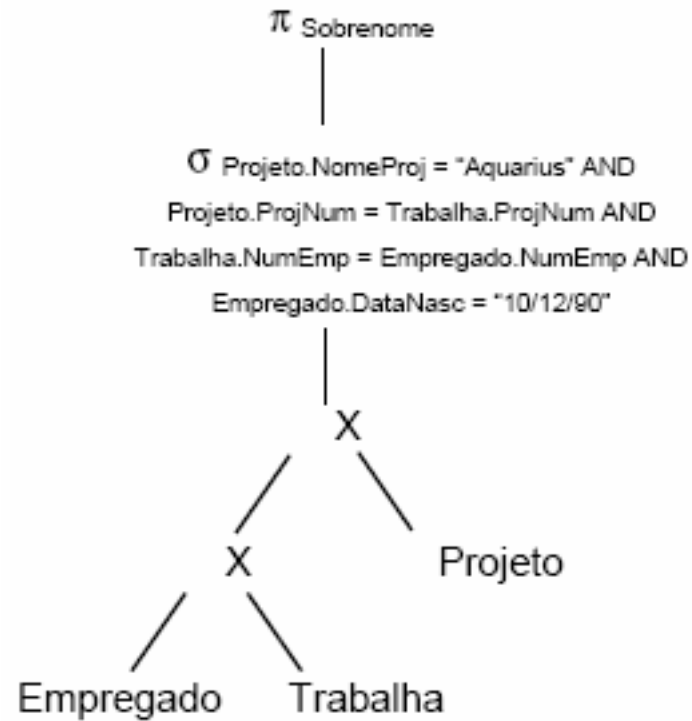


Passo 5: Desmembrar e mover projeções para níveis inferiores da árvore, tanto quanto possível, definindo novas projeções conforme se faça necessário.

5 Árvore de Consulta após Passo 5



1 Árvore de Consulta Algébrica



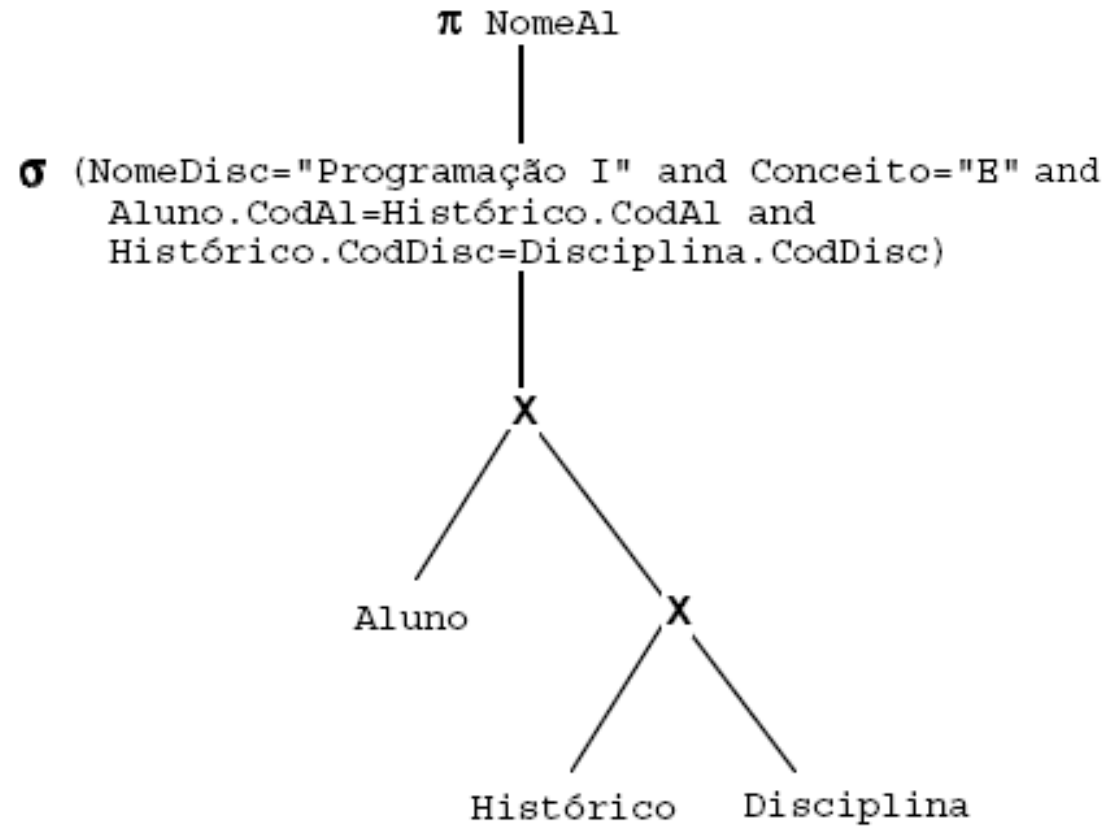
5 Árvore de Consulta após Passo 5



- Consulta sobre um Banco de Dados acadêmico
Aluno (codAl, nomeAl,)
Disciplina (codDisc, nomeDisc, ...)
Historico (codAl, codDisc, xxxx, conceito...)
- Obter os nomes dos alunos que obtiveram conceito "E" em disciplina denominada "Programação I"

```
SELECT Aluno.NomeAl
FROM
    Aluno, Historico, Disciplina
WHERE
    Disciplina.NomeDisc = 'Programação I' AND Historico.Conceito = 'E' AND
    Aluno.CodAl = Historico.CodAl AND
    Historico.CodDisc = Disciplina.CodDisc
```

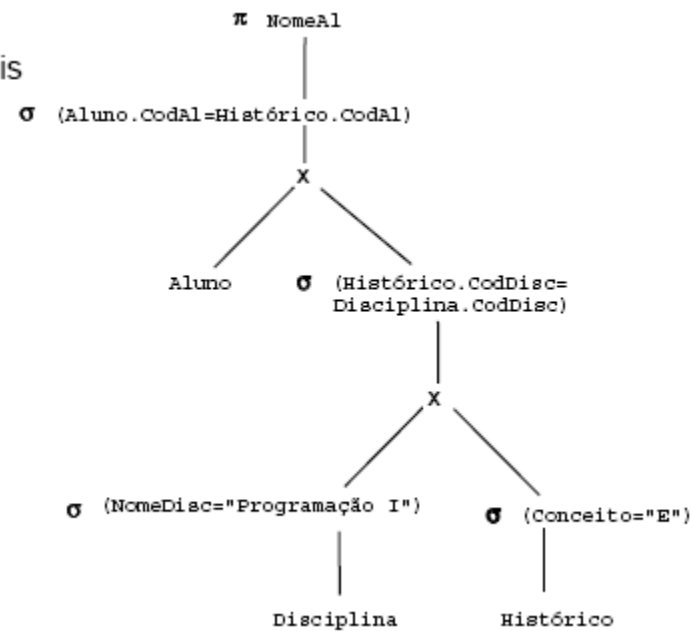
```
 $\pi$  NomeAl  
( $\sigma$  (NomeDisc="Programação I" and Conceito="E"  
    and  
    Aluno.CodAl=Histórico.CodAl  
    and  
    Histórico.CodDisc=Disciplina.CodDisc)  
  (Aluno  $\bowtie$   
    (Histórico  $\bowtie$   
      (Disciplina))) )
```



Passo 1: Desmembrar operações de seleção
Maior flexibilidade para mover seleções

Passo 2: Mover seleções para níveis inferiores da árvore o máximo possível
Diminuir resultados intermediários

Seleções são executadas mais cedo

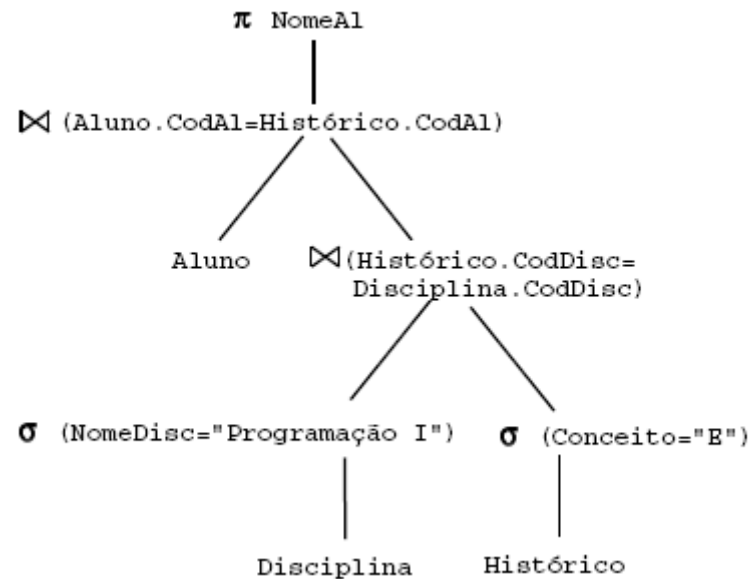


Passo 3: Mudar de posição sub-árvores envolvidas em operações produtórias.

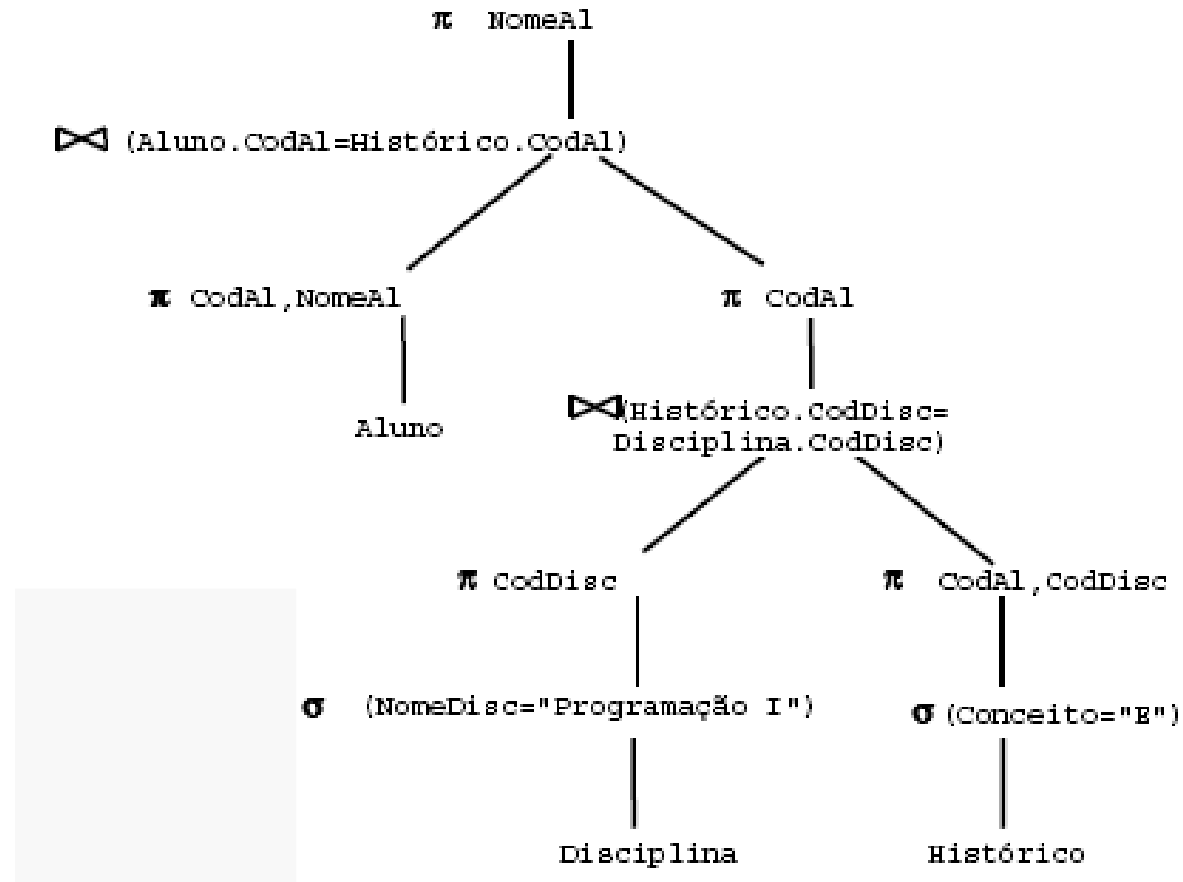
Combinar prioritariamente sub-árvores com menor número de dados

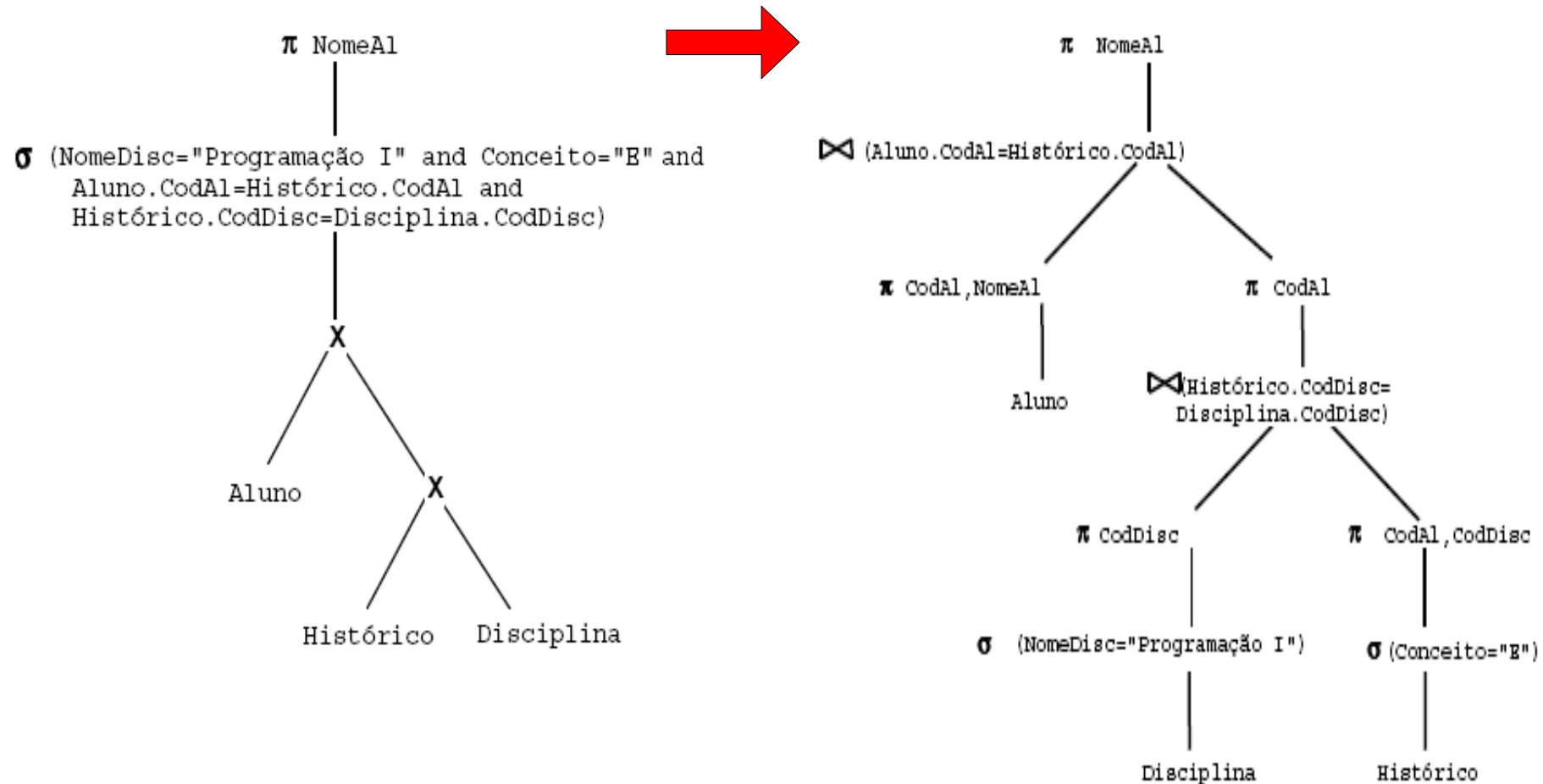
Normalmente utiliza-se primeiro sub-árvores com seleções mais restritivas

Passo 4: Otimizar operações produtórias (converter de produto cartesiano para junção)



Passo 5: Desmembrar e mover projeções para níveis inferiores da árvore, tanto quanto possível, definindo novas projeções conforme se faça necessário.





- A otimização heurística não leva em conta volumes de dados nem tempos de acesso
 - Com isso, pode gerar árvores de consulta menos eficientes
- Otimização baseada em estimativas de custo
 - Para diferentes árvores de consulta, equivalentes em performance do ponto de vista das regras heurísticas, é escolhida uma que se estima ser de menor custo de implementação
 - Custo = volume de acesso a disco
 - Como é oneroso verificar volumes a cada transação, SGBD normalmente coleta dados estatísticos sobre as tabelas (número de linhas, número de valores diferentes por coluna, ...)

- Para entender a otimização baseada em custos, é preciso entender os algoritmos possíveis para cada uma das operações da árvore algébrica
- Os algoritmos possíveis de utilizar e os custos estimados para cada um deles é que irão determinar as estratégias de acesso, que irão compor o plano de execução da consulta

- Considerando o modelo a seguir:

Depto(CodDepto, NomeDepto)

Disciplina(CodDepto, NumDisc, NomeDisc, CreditosDisc)
CodDepto referencia Depto

PreReq(CodDepto, NumDisc, CodDeptoPreReq, NumDiscPreReq)
(CodDepto, NumDisc) referencia Disciplina
(CodDeptoPreReq, NumDiscPreReq) referencia Disciplina

Turma(AnoSem, CodDepto, NumDisc, SiglaTur, CapacOfere) (CodDepto, NumDisc) referencia Disciplina

Horario(AnoSem, CodDepto, NumDisc, SiglaTur, DiaSem, HoraInicio, NumHoras, CodPred, NumSala)
(AnoSem, CodDepto, NumDisc, SiglaTur) referencia Turma
(CodPred, NumSala) referencia Sala

Predio(CodPred, NomePred)

Sala(CodPred, NumSala, CapacSala)
CodPred referencia Predio

Professor(CodProf, NomeProf, CodTit, CodDepto)
CodTit referencia Titulação
CodDepto referencia Departamento

ProfTurma(AnoSem, CodDepto, NumDisc, SiglaTur, CodProf)
(AnoSem, CodDepto, NumDisc, SiglaTur) referencia Turma
CodProf referencia Professor

Titulação(CodTit, NomeTit)

- Gere a árvore algébrica otimizada para as seguintes consultas, mostrando passo a passo o processo de otimização:
 - Obtenha os nomes das disciplinas do depto de computação com mais de 3 créditos
 - Obtenha os nomes das disciplinas que foram ministradas em 2000/1 pelo professor Mário
 - Obtenha a titulação dos professores que ministraram turmas em 1999/2

