

ACH2025

Laboratório de Bases de Dados

Aula 13

Processamento de Consultas - Parte 1

Professora:

➤ **Fátima L. S. Nunes**



Até agora, já sabemos...

- ✓ Modelagem de BD
- ✓ Álgebra Relacional
- ✓ SQL
- ✓ Como dados são armazenados fisicamente
- ✓ Como construir índices
- ✓ Como armazenar dados fisicamente e construir índices usando hashing

Processamento de consultas

✓ O que veremos:

- Como SGBD processa consultas
- Fatores envolvidos no processamento de consultas
- Custos de consulta

Processamento de consultas

✓ Finalidade???

Processamento de consultas

✓ Finalidade???

- Suportar atividades envolvidas na ação de extrair dados de um BD

✓ Que atividades são essas?

Processamento de consultas

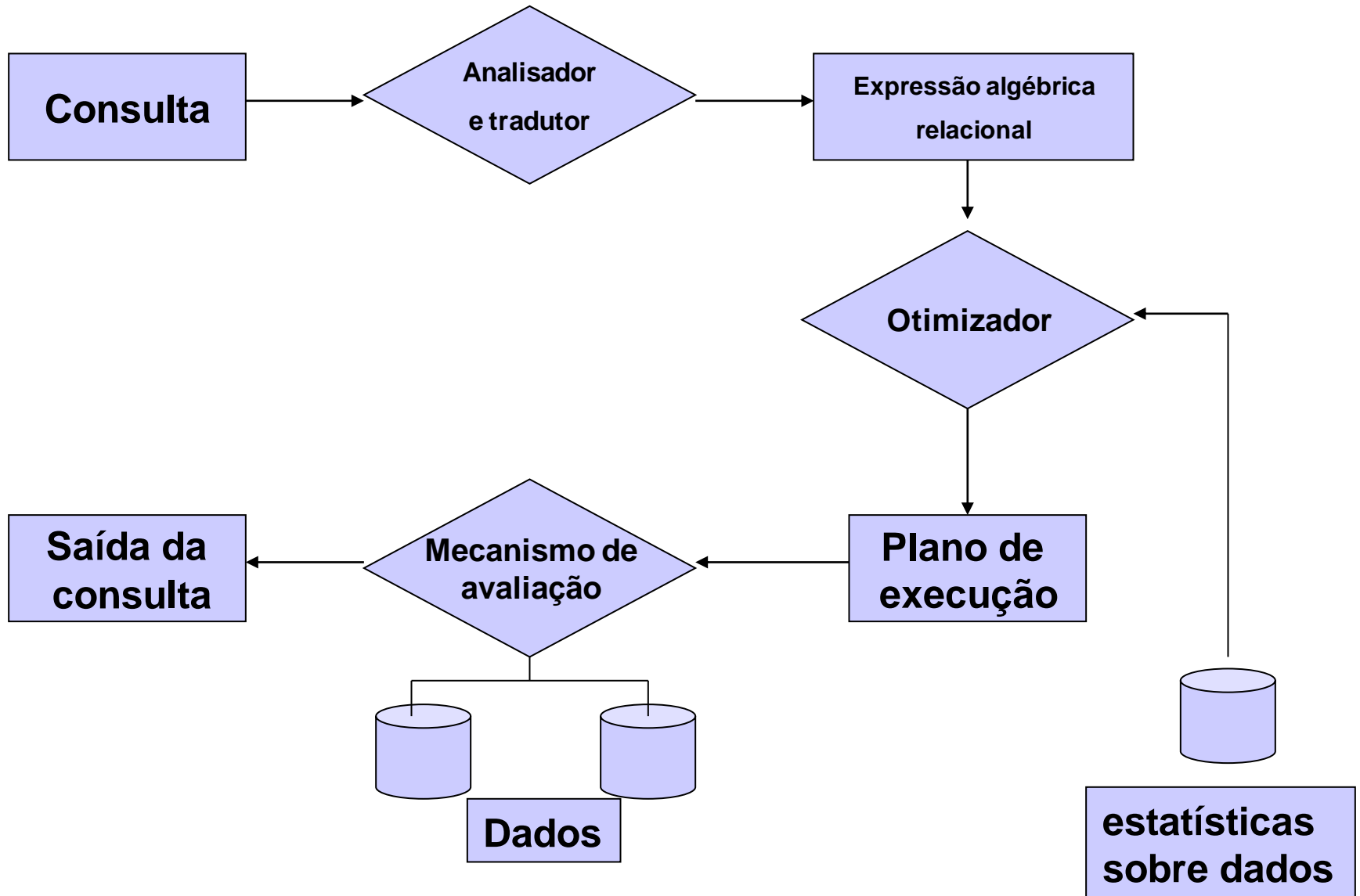
✓ Finalidade???

- Suportar atividades envolvidas na ação de extrair dados de um BD

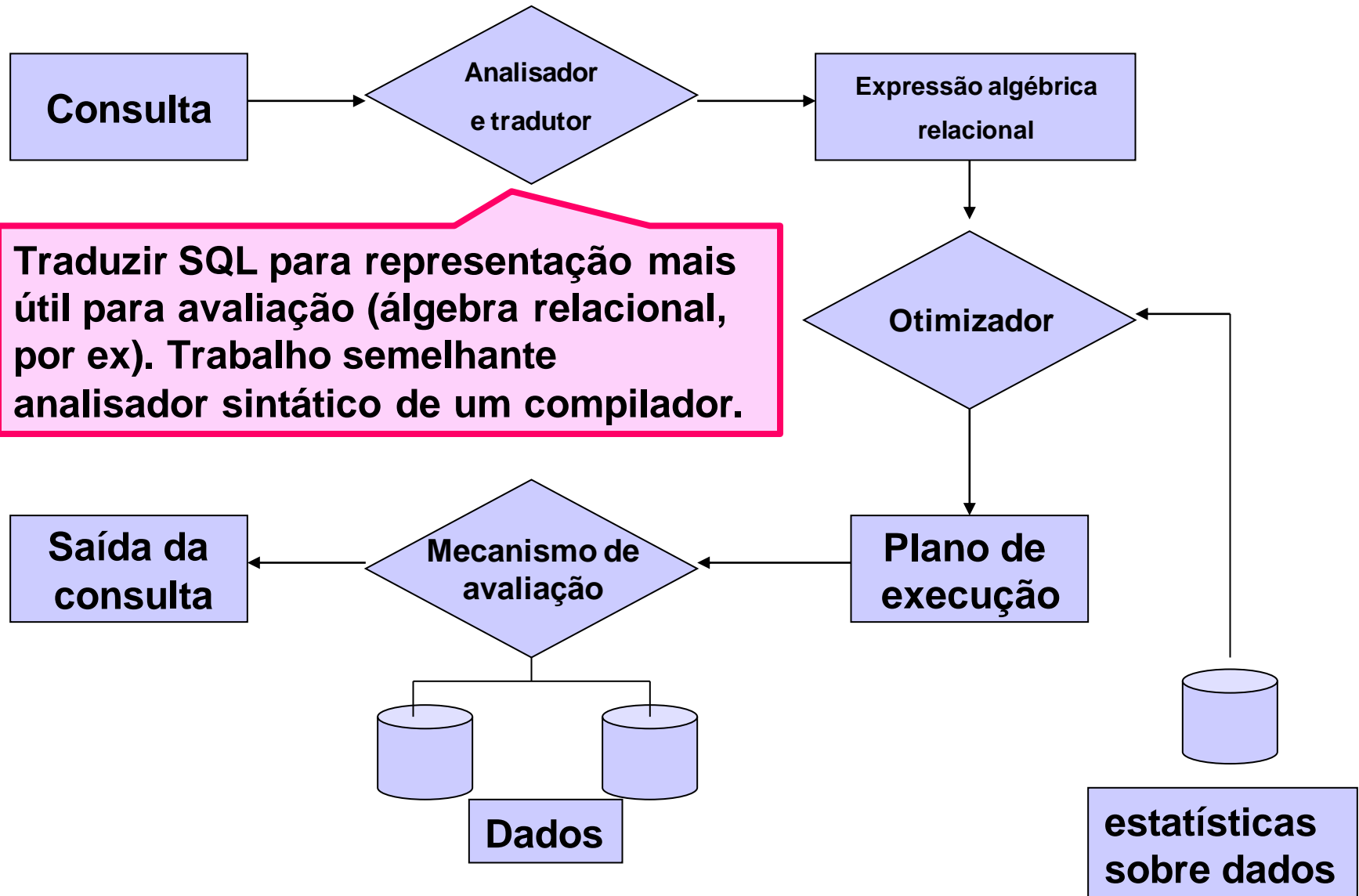
✓ Que atividades são essas?

- Tradução alto nível → expressões implementadas em nível físico
- Otimizações
- Avaliação

Processamento de consultas



Processamento de consultas



Processamento de consultas

✓ Exemplo de consulta em SQL:

```
select saldo
```

```
from conta
```

```
where saldo < 2500
```

✓ Como fica em Álgebra Relacional?

Processamento de consultas

✓ Exemplo de consulta em SQL:

```
select saldo  
from conta  
where saldo < 2500
```

✓ Como fica em Álgebra Relacional?

- $\sigma_{\text{saldo} < 2500} (\pi_{\text{saldo}} (\text{conta}))$
- $\pi_{\text{saldo}} (\sigma_{\text{saldo} < 2500} (\text{conta}))$

Processamento de consultas

✓ Exemplo de consulta em SQL:

```
select saldo
```

```
from conta
```

```
where saldo < 2500
```

✓ Como fica em Álgebra Relacional?

- $\sigma_{\text{saldo} < 2500} (\pi_{\text{saldo}} (\text{conta}))$

- $\pi_{\text{saldo}} (\sigma_{\text{saldo} < 2500} (\text{conta}))$

Além de poder ter mais de uma representação algébrica, há vários algoritmos para cada operação algébrica relacional.

Processamento de consultas

✓ Exemplo de consulta em SQL:

```
select saldo  
from conta  
where saldo < 2500
```

✓ Como fica em Álgebra Relacional?

- $\sigma_{\text{saldo} < 2500} (\pi_{\text{saldo}} (\text{conta}))$
- $\pi_{\text{saldo}} (\sigma_{\text{saldo} < 2500} (\text{conta}))$

Além de poder ter mais de uma representação algébrica, há vários algoritmos para cada operação algébrica relacional.

Usando o que aprendemos com índices, como pode ser feita a seleção?

Processamento de consultas

✓ Exemplo de consulta em SQL:

```
select saldo
from conta
where saldo < 2500
```

✓ Como fica em Álgebra Relacional?

- $\sigma_{\text{saldo} < 2500} (\pi_{\text{saldo}} (\text{conta}))$
- $\pi_{\text{saldo}} (\sigma_{\text{saldo} < 2500} (\text{conta}))$

Para avaliar uma consulta, não basta expressão algébrica relacional. É necessário:

- anotações de instruções que especificam como avaliar cada operação;
- **expressão algébrica relacional anotada** com instruções sobre como ser avaliada é chamada avaliação primitiva (AP);
- várias APs podem ser agrupadas em *pipeline*, permitindo execuções em paralelo;
- sequência de APs é chamada de plano de execução de consulta.

Processamento de consultas

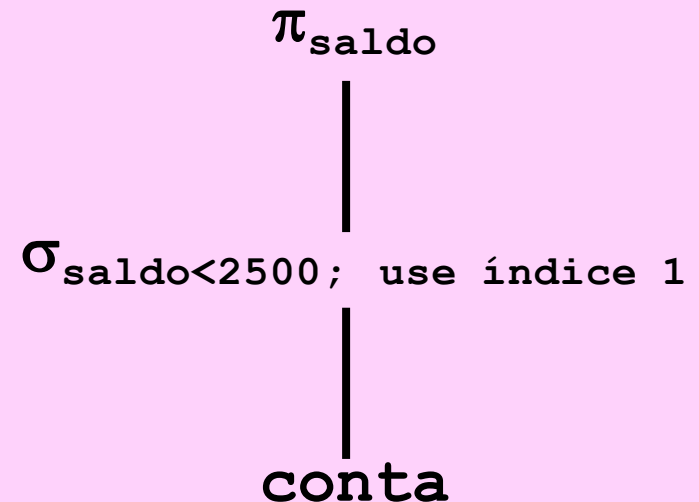
✓ Exemplo de consulta em SQL:

```
select saldo
from conta
where saldo < 2500
```

✓ Como fica em Álgebra Relacional?

- $\sigma_{\text{saldo} < 2500} (\pi_{\text{saldo}} (\text{conta}))$
- $\pi_{\text{saldo}} (\sigma_{\text{saldo} < 2500} (\text{conta}))$

- Sequência de APs é chamada de plano de execução de consulta.
- Exemplo:



Processamento de consultas

✓ Exemplo de consulta em SQL:

```
select saldo
```

```
from conta
```

```
where saldo < 2500
```

✓ Como fica em Álgebra Relacional?

▪ $\sigma_{\text{saldo} < 2500} (\pi_{\text{saldo}} (\text{conta}))$

▪ $\pi_{\text{saldo}} (\sigma_{\text{saldo} < 2500} (\text{conta}))$

- Sequência de APs é chamada de plano de execução de consulta.

- Diferentes planos podem ter diferentes custos.

- É dever do SGBD construir um plano que minimize custos.

- **Medida de desempenho mais relevante???**

Processamento de consultas

✓ Exemplo de consulta em SQL:

```
select saldo
```

```
from conta
```

```
where saldo < 2500
```

✓ Como fica em Álgebra Relacional?

▪ $\sigma_{\text{saldo} < 2500} (\pi_{\text{saldo}} (\text{conta}))$

▪ $\pi_{\text{saldo}} (\sigma_{\text{saldo} < 2500} (\text{conta}))$

- Sequência de APs é chamada de plano de execução de consulta (PEC).

- Diferentes PECs podem ter diferentes custos.

- É dever do SGBD construir um plano que minimize custos.

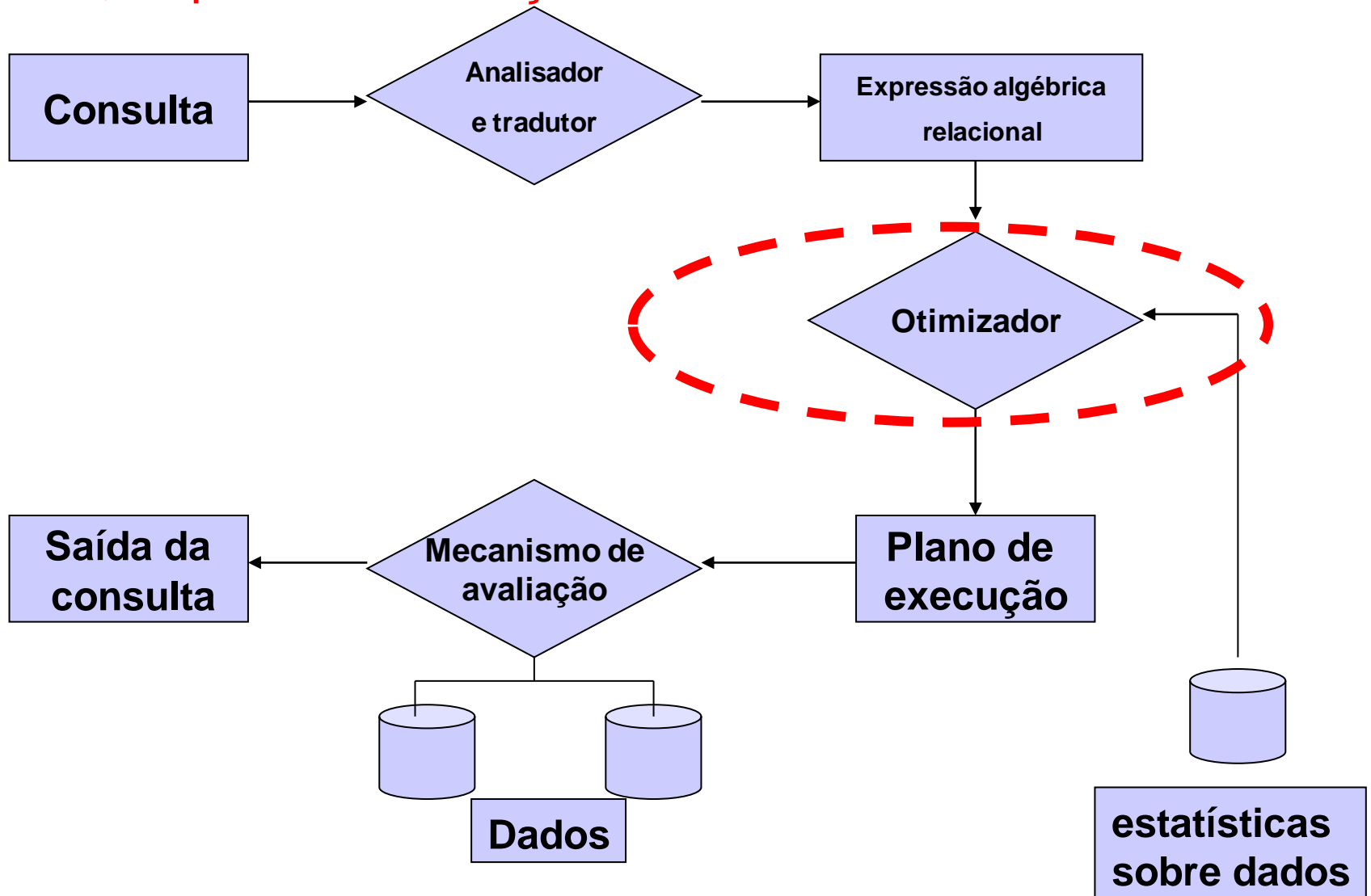
- **Medida de desempenho mais relevante??? NÚMERO DE ACESSOS AO DISCO**

Processamento de consultas

✓ Então, o que é a otimização de consultas?

Processamento de consultas

✓ Então, o que é a otimização de consultas?



Processamento de consultas

✓ Então, o que é a otimização de consultas?

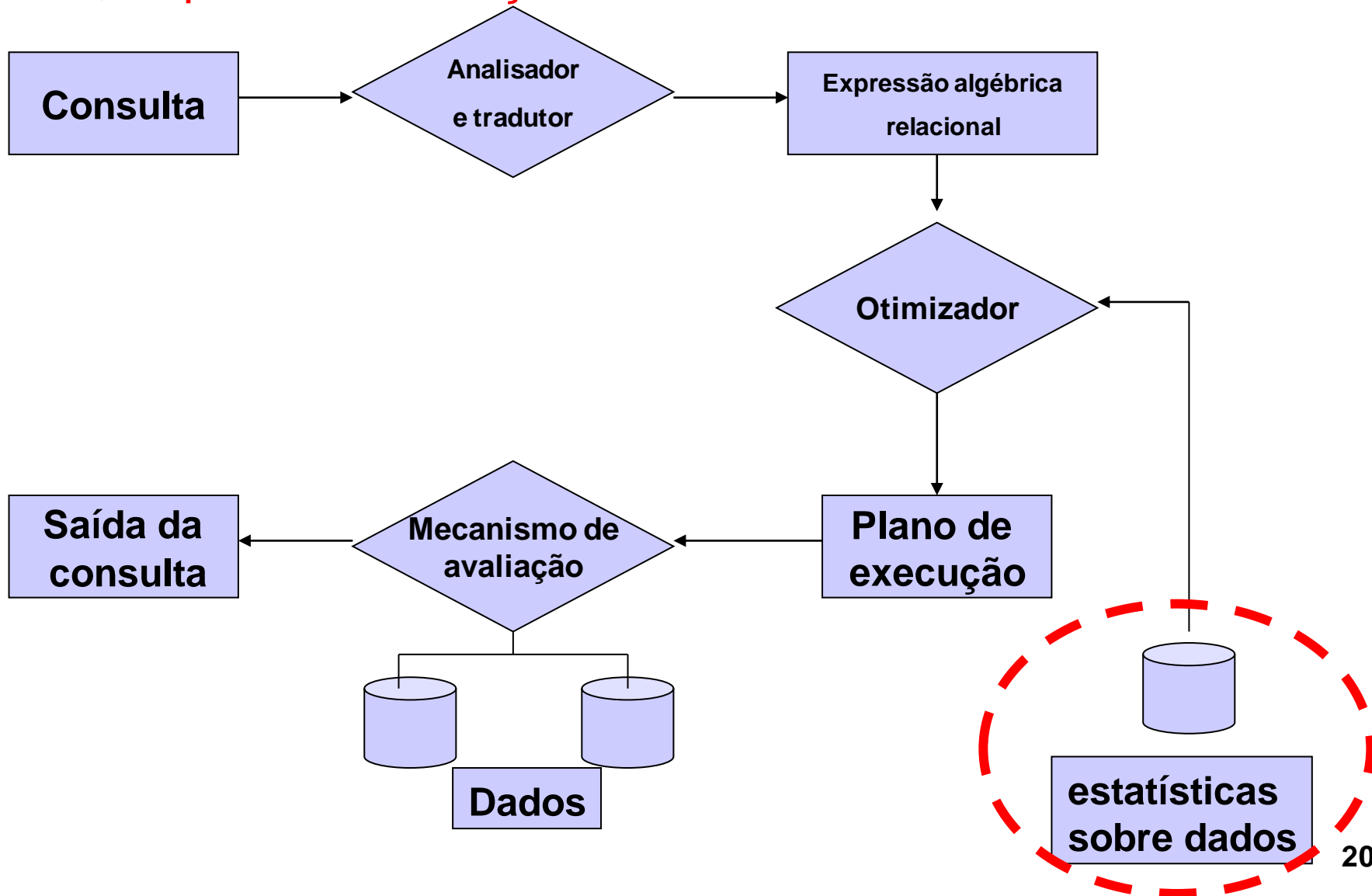
- processo de selecionar o plano de avaliação de consulta mais eficiente para uma consulta.
- possibilidades:
 - achar expressão equivalente em álgebra relacional que seja mais eficiente (veremos mais adiante);
 - estratégia adequada para processar consulta (algoritmos, índices).

✓ Como escolher melhor plano de avaliação de consulta?

- não é possível calcular custo exato sem executar plano;
- solução: uso de estatísticas.

Processamento de consultas

✓ Então, o que é a otimização de consultas?



Como calcular o custo de um PEC

✓ Informações relevantes ???

Como calcular o custo de um PEC

✓ Informações relevantes sobre as relações:

- n_r – número de tuplas na relação r
- b_r – número de blocos que contêm tuplas da relação r
- s_r – tamanho em bytes de uma tupla da relação r
- f_r – fator de bloco da relação r : número de tuplas da relação r que cabe em um bloco
- $V(A, r)$ – número de valores distintos para o atributo A na relação r
- $SC(A, r)$ – cardinalidade da seleção do atributo A da relação r (número médio de registros que satisfazem uma condição de igualdade no atributo A)

Como calcular o custo de um PEC

✓ Informações relevantes sobre os índices:

- f_i – *fan-out* médio dos nós internos do índice i para índices estruturados em árvore
- HT_i – número de níveis no índice i (altura do índice). Se árvore balanceada, $HT_i = \lceil \log_{f_i}(V(A, r)) \rceil$
- LB_i – número de blocos de índice de nível mais baixo no índice i (nível de folha)

Como calcular o custo de um PEC

✓ Problema:

- manter estatísticas atualizadas. Por quê?

Como calcular o custo de um PEC

✓ Problema:

- manter estatísticas atualizadas. **Por quê?**
- *overhead* devido à atualização dos dados estatísticos toda vez que houver alteração na relação ou no índice
- solução: atualização periódica.

Como calcular o custo de um PEC

✓ Agora que sabemos as variáveis, o que falta?

Como calcular o custo de um PEC

✓ Agora que sabemos as variáveis, o que falta?

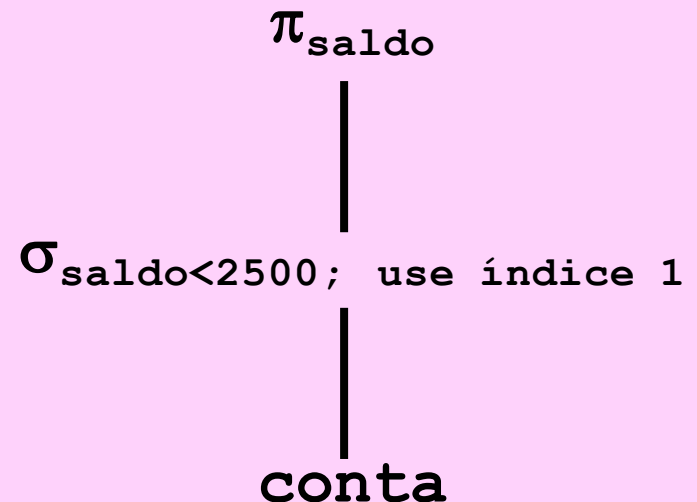
– Saber os algoritmos para as operações envolvidas

▪ $\sigma_{\text{saldo} < 2500} (\pi_{\text{saldo}} (\text{conta}))$

▪ $\pi_{\text{saldo}} (\sigma_{\text{saldo} < 2500} (\text{conta}))$

• Sequência de APs é chamada de plano de execução de consulta.

• Exemplo:



Como calcular o custo de um PEC

- ✓ Chamamos de E_{A_i} o custo do algoritmo A_i
- ✓ Operação de Seleção
 - Dois algoritmos de varredura para implementá-la
 - A1: busca linear
 - A2: busca binária

Como calcular o custo de um PEC

- ✓ Chamamos de E_{Ai} o custo do algoritmo A_i
- ✓ Operação de Seleção
 - A1 (busca linear)
 - varre cada bloco do arquivo e todos registros são testados para verificar a condição de seleção
 - todos blocos precisam ser lidos. Então, custo:

n_r – número de tuplas na relação r
 b_r – número de blocos que contêm tuplas de r
 s_r – tamanho em bytes de uma tupla r
 f_r – número de tuplas que cabe em um bloco
 $V(A, r)$ – número de valores distintos do atributo A em r .
 $SC(A, r)$ – número médio de registros que satisfazem uma condição de igualdade no atributo A



Como calcular o custo de um PEC

- ✓ Chamamos de E_{A_i} o custo do algoritmo A_i
- ✓ Operação de Seleção
 - A1 (busca linear)
 - varre cada bloco do arquivo e todos registros são testados para verificar a condição de seleção
 - todos blocos precisam ser lidos. Então, custo: $E_{A1} = b_r$

n_r – número de tuplas na relação r
 b_r – número de blocos que contêm tuplas de r
 s_r – tamanho em bytes de uma tupla r
 f_r – número de tuplas que cabe em um bloco
 $V(A, r)$ – número de valores distintos do atributo A em r .
 $SC(A, r)$ – número médio de registros que satisfazem uma condição de igualdade no atributo A



Como calcular o custo de um PEC

- ✓ Chamamos de E_{Ai} o custo do algoritmo A_i
- ✓ Operação de Seleção
 - A1 (busca linear)
 - varre cada bloco do arquivo e todos registros são testados para verificar a condição de seleção
 - todos blocos precisam ser lidos. Então, custo: $E_{A1} = b_r$
 - se seleção for em atributo chave (de organização física), supõe-se que metade dos blocos é varrida → Qual é o custo?

n_r – número de tuplas na relação r
 b_r – número de blocos que contêm tuplas de r
 s_r – tamanho em bytes de uma tupla r
 f_r – número de tuplas que cabe em um bloco
 $V(A, r)$ – número de valores distintos do atributo A em r .
 $SC(A, r)$ – número médio de registros que satisfazem uma condição de igualdade no atributo A



Como calcular o custo de um PEC

✓ Chamamos de E_{Ai} o custo do algoritmo A_i

✓ Operação de Seleção

– A_1 (busca linear)

- varre cada bloco do arquivo e todos registros são testados para verificar a condição de seleção
- todos blocos precisam ser lidos. Então, custo: $E_{A1} = b_r$
- se seleção for em atributo chave (de organização física), supõe-se que metade dos blocos é varrida → Qual é o custo?

– custo médio = $E_{A1} = \frac{b_r}{2}$

Como calcular o custo de um PEC

✓ Operação de Seleção

- A2 (busca binária) – considerando arquivo ordenado fisicamente pelo atributo usado na seleção:

$$E_{A2} = \lceil \log_2(b_r) \rceil + \left\lceil \frac{SC(A, r)}{f_r} \right\rceil - 1$$

n_r – número de tuplas na relação r

b_r – número de blocos que contêm tuplas de r

s_r – tamanho em bytes de uma tupla r

f_r – número de tuplas que cabe em um bloco

$V(A, r)$ – número de valores distintos do atributo A em r .

$SC(A, r)$ – número médio de registros que satisfazem uma condição de igualdade no atributo A



Como calcular o custo de um PEC

✓ Operação de Seleção

- A2 (busca binária) – considerando arquivo ordenado pelo atributo usado na seleção:

$$E_{A2} = \left\lceil \log_2(b_r) \right\rceil + \left\lceil \frac{SC(A, r)}{f_r} \right\rceil - 1$$

Custo para localizar primeira tupla

1 registro já havia sido recuperado

Quantidade total de registros que satisfazem a condição dividido pela quantidade total de tuplas que cabem no bloco

n_r – número de tuplas na relação r

b_r – número de blocos que contêm tuplas de r

s_r – tamanho em bytes de uma tupla r

f_r – número de tuplas que cabe em um bloco

$V(A, r)$ – número de valores distintos do atributo A em r .

$SC(A, r)$ – número médio de registros que satisfazem uma condição de igualdade no atributo A



Como calcular o custo de um PEC

✓ Operação de Seleção

- A2 (busca binária) – considerando arquivo ordenado pelo atributo usado na seleção:

$$E_{A2} = \lceil \log_2(b_r) \rceil + \left\lceil \frac{SC(A, r)}{f_r} \right\rceil - 1$$

Se condição for em atributo
que é chave primária,
 $SC(A, r) = ???$

n_r – número de tuplas na relação r
 b_r – número de blocos que contêm tuplas de r
 s_r – tamanho em bytes de uma tupla r
 f_r – número de tuplas que cabe em um bloco
 $V(A, r)$ – número de valores distintos do atributo A em r .
 $SC(A, r)$ – número médio de registros que satisfazem uma condição de igualdade no atributo A



Como calcular o custo de um PEC

✓ Operação de Seleção

- A2 (busca binária) – considerando arquivo ordenado pelo atributo usado na seleção:

$$E_{A2} = \lceil \log_2(b_r) \rceil + \left\lceil \frac{SC(A, r)}{f_r} \right\rceil - 1$$

Se condição for em atributo que é chave primária, $SC(A, r) = 1$. Então tem-se:

$$E_{A2} = \lceil \log_2(b_r) \rceil$$

n_r – número de tuplas na relação r

b_r – número de blocos que contêm tuplas de r

s_r – tamanho em bytes de uma tupla r

f_r – número de tuplas que cabe em um bloco

$V(A, r)$ – número de valores distintos do atributo A em r .

$SC(A, r)$ – número médio de registros que satisfazem uma condição de igualdade no atributo A



Como calcular o custo de um PEC

✓ Operação de Seleção

- A2 (busca binária) – considerando arquivo ordenado pelo atributo usado na seleção:

$$E_{A2} = \lceil \log_2(b_r) \rceil + \left\lceil \frac{SC(A, r)}{f_r} \right\rceil - 1$$

n_r – número de tuplas na relação r

b_r – número de blocos que contêm tuplas de r

s_r – tamanho em bytes de uma tupla r

f_r – número de tuplas que cabe em um bloco

$V(A, r)$ – número de valores distintos do atributo A em r .

$SC(A, r)$ – número médio de registros que satisfazem uma condição de igualdade no atributo A

- Estimativa considera blocos da relação armazenados de forma contígua. Caso contrário, custo da procura nas estruturas de acesso a arquivos deve ser acrescentado.



Como calcular o custo de um PEC

✓ Operação de Seleção

- Deve ser considerado ainda tamanho do resultado da seleção.
- Supondo distribuição uniforme de valores, a consulta $S_{A=a}(r)$ tem:

$$SC(A, r) = \left[\frac{n_r}{V(A, r)} \right]$$

n_r – número de tuplas na relação r

b_r – número de blocos que contêm tuplas de r

s_r – tamanho em bytes de uma tupla r

f_r – número de tuplas que cabe em um bloco

$V(A, r)$ – número de valores distintos do atributo A em r .

$SC(A, r)$ – número médio de registros que satisfazem uma condição de igualdade no atributo A



Como calcular o custo de um PEC

✓ Operação de Seleção - Exemplo

$$\sigma_{\text{nome_agencia}=\text{"Centro"}}(\text{conta})$$

- $f_{\text{conta}}=20$
- $V(\text{nome_agencia}, \text{conta}) = 50$
- $V(\text{saldo}, \text{conta}) = 500$
- $n_{\text{conta}}=10000$

Qual o custo para:

- varredura de arquivo simples?
- busca binária se relação conta estiver ordenada fisicamente por nome_agencia?

n_r – número de tuplas na relação r
 b_r – número de blocos que contêm tuplas de r
 s_r – tamanho em bytes de uma tupla r
 f_r – número de tuplas que cabe em um bloco
 $V(A, r)$ – número de valores distintos do atributo A em r .
 $SC(A, r)$ – número médio de registros que satisfazem uma condição de igualdade no atributo A



Como calcular o custo de um PEC

✓ Operação de Seleção - Exemplo

$\sigma_{\text{nome_agencia}=\text{"Centro"}}(\text{conta})$

- $f_{\text{conta}}=20$
- $V(\text{nome_agencia}, \text{conta}) = 50$
- $V(\text{saldo}, \text{conta}) = 500$
- $n_{\text{conta}}=10000$

Qual o custo para:

- varredura de arquivo simples?

$$E_{A1} = b_r$$

- busca binária se relação conta estiver ordenada fisicamente por nome_agencia?

$$E_{A2} = \lceil \log_2(b_r) \rceil + \left\lceil \frac{SC(A, r)}{f_r} \right\rceil - 1$$

n_r – número de tuplas na relação r

b_r – número de blocos que contêm tuplas de r

s_r – tamanho em bytes de uma tupla r

f_r – número de tuplas que cabe em um bloco

$V(A, r)$ – número de valores distintos do atributo A em r .

$SC(A, r)$ – número médio de registros que satisfazem uma condição de igualdade no atributo A



Como calcular o custo de um PEC

✓ Operação de Seleção - Exemplo

$\sigma_{\text{nome_agencia}=\text{"Centro"}}(\text{conta})$

- $f_{\text{conta}}=20$
- $V(\text{nome_agencia}, \text{conta}) = 50$
- $V(\text{saldo}, \text{conta}) = 500$
- $n_{\text{conta}}=10000$

$b_{\text{conta}} = 10000/20 = 500$

Qual o custo para:

- varredura de arquivo simples?

500 acessos de bloco

- busca binária se relação conta estiver ordenada por nome_agencia?

$V(\text{nome_agencia}, \text{conta}) = 50$

Esperamos que $10000/50$ tuplas pertençam à agência "Centro"

$200/20 = 10$ blocos

Encontrar primeiro bloco = $\log_2(500) = 9$

Total = $9 + 10 - 1 = 18$ acessos de bloco

n_r – número de tuplas na relação r

b_r – número de blocos que contêm tuplas de r

s_r – tamanho em bytes de uma tupla r

f_r – número de tuplas que cabe em um bloco

$V(A, r)$ – número de valores distintos do atributo A em r .

$SC(A, r)$ – número médio de registros que satisfazem uma condição de igualdade no atributo A

Como calcular o custo de um PEC

✓ Operação de Seleção usando índices

– A3 (índice **primário**, igualdade de chave única)

- usa índice para recuperar um único registro que satisfaz condição;
- para recuperar 1 registro: recuperar um bloco a mais que o número de níveis do índice:

$$E_{A3} = HT_i + 1$$

Exemplo: $\sigma_{\text{num_agencia}=1267}$ (**agencia**)
(*arquivo organizado por num_agencia*)

n_r – número de tuplas na relação r

b_r – número de blocos que contêm tuplas de r

s_r – tamanho em bytes de uma tupla r

f_r – número de tuplas que cabe em um bloco

$V(A, r)$ – número de valores distintos do atributo A

$SC(A, r)$ – núm médio registros satisfazem condição

f_i – fan-out médio dos nós internos do índice i

HT_i – número de níveis no índice i (altura do índice).

LB_i – número de blocos de índice de nível mais baixo no índice i

Como calcular o custo de um PEC

✓ Operação de Seleção usando índices

- A3 (índice **primário**, igualdade de chave única)

$$E_{A3} = HT_i + 1$$

Exemplo: $\sigma_{\text{num_agencia}=1267}$ (**agencia**)
(arquivo organizado por num_agencia)

- $f_{\text{conta}}=20$
- $V(\text{nome_agencia}, \text{conta}) = 50$
- $V(\text{saldo}, \text{conta}) = 500$
- $n_{\text{conta}}=10000$
- $f_i = 40$ (árvore B+)

n_r – número de tuplas na relação r
 b_r – número de blocos que contêm tuplas de r
 s_r – tamanho em bytes de uma tupla r
 f_r – número de tuplas que cabe em um bloco
 $V(A, r)$ – número de valores distintos do atributo A
 $SC(A, r)$ – núm médio registros satisfazem condição

f_i – fan-out médio dos nós internos do índice i
 HT_i – número de níveis no índice i (altura do índice).
 LB_i – número de blocos de índice de nível mais baixo no índice i

Como calcular o custo de um PEC

✓ Operação de Seleção usando índices

- A3 (índice **primário**, igualdade de chave única)

$$E_{A3} = HT_i + 1$$

Exemplo: $\sigma_{\text{num_agencia}=1267}$ (**agencia**)
(*arquivo organizado por num_agencia*)

- $f_{\text{conta}}=20$
- $V(\text{nome_agencia}, \text{conta}) = 50$
- $V(\text{saldo}, \text{conta}) = 500$
- $n_{\text{conta}}=10000$
- $f_i = 40$ (árvore B+)

$$HT_i = \log_{f_i/2}(10000) = 4$$

$$E_{A3} = 5$$

n_r – número de tuplas na relação r
 b_r – número de blocos que contêm tuplas de r
 s_r – tamanho em bytes de uma tupla r
 f_r – número de tuplas que cabe em um bloco
 $V(A, r)$ – número de valores distintos do atributo A
 $SC(A, r)$ – núm médio registros satisfazem condição

f_i – fan-out médio dos nós internos do índice i
 HT_i – número de níveis no índice i (altura do índice).
 LB_i – número de blocos de índice de nível mais baixo no índice i

Como calcular o custo de um PEC

✓ Operação de Seleção usando índices

- A4 (índice **primário**, igualdade em atributo não chave única)
 - $SC(A, r)$ registros satisfarão a condição
 - $\lceil SC(A, r) / f_r \rceil$ blocos serão acessados
 - Registros armazenados em blocos contíguos.

$$E_{A4} = HT_i + \left\lceil \frac{SC(A, r)}{f_r} \right\rceil$$

Exemplo: $\sigma_{\text{cidade_agencia}=\text{"Santos"}}$ (**agencia**)
(*arquivo organizado por cidade_agencia*)

n_r – número de tuplas na relação r

b_r – número de blocos que contêm tuplas de r

s_r – tamanho em bytes de uma tupla r

f_r – número de tuplas que cabe em um bloco

$V(A, r)$ – número de valores distintos do atributo A

$SC(A, r)$ – número médio registros satisfazem condição

f_i – fan-out médio dos nós internos do índice i

HT_i – número de níveis no índice i (altura do índice).

LB_i – número de blocos de índice de nível mais baixo no índice i

Como calcular o custo de um PEC

✓ Operação de Seleção usando índices

- A4 (índice **primário**, igualdade em atributo não chave única)

$$E_{A4} = HT_i + \left\lceil \frac{SC(A, r)}{f_r} \right\rceil$$

Exemplo:

$\sigma_{\text{cidade_agencia}=\text{"Santos"}}$ (**agencia**)
(*arquivo organizado por cidade_agencia*)

$f_{\text{conta}}=20$

$V(\text{cidade_agencia}, \text{conta}) = 50$

$V(\text{saldo}, \text{conta}) = 500$

$n_{\text{conta}}=10000$

$f_i = 40$ (árvore B+)

n_r – número de tuplas na relação r

b_r – número de blocos que contêm tuplas de r

s_r – tamanho em bytes de uma tupla r

f_r – número de tuplas que cabe em um bloco

$V(A, r)$ – número de valores distintos do atributo A

$SC(A, r)$ – número médio registros satisfazem condição

f_i – fan-out médio dos nós internos do índice i

HT_i – número de níveis no índice i (altura do índice).

LB_i – número de blocos de índice de nível i mais baixo no índice i 46

Como calcular o custo de um PEC

✓ Operação de Seleção usando índices

- A4 (índice **primário**, igualdade em atributo não chave única)

$$E_{A4} = HT_i + \left\lceil \frac{SC(A, r)}{f_r} \right\rceil$$

Exemplo:

$\sigma_{\text{cidade_agencia}=\text{"Santos"}}$ (**agencia**)
(arquivo organizado por cidade_agencia)

$f_{\text{conta}}=20$

$V(\text{cidade_agencia}, \text{conta}) = 50$

$V(\text{saldo}, \text{conta}) = 500$

$n_{\text{conta}}=10000$

$f_i = 40$ (árvore B+)

$SC(\text{cidade_agencia}, \text{conta}) = 10000/50 = 200$

Número de chaves esperadas no índice = 50

$HT_i = \log_{f_i/2}(50) = 2$

$E_{A4} = 2 + 200/20 = 12$

n_r – número de tuplas na relação r

b_r – número de blocos que contêm tuplas de r

s_r – tamanho em bytes de uma tupla r

f_r – número de tuplas que cabe em um bloco

$V(A, r)$ – número de valores distintos do atributo A

$SC(A, r)$ – número médio registros satisfazem condição

f_i – fan-out médio dos nós internos do índice i

HT_i – número de níveis no índice i (altura do índice).

LB_i – número de blocos de índice de nível mais baixo no índice i

Como calcular o custo de um PEC

✓ Operação de Seleção usando índices

– A5 (índice **secundário**, igualdade)

- Recupera 1 registro se campo indexação for uma chave única

$$E_{A5} = HT_i + 1$$

Exemplo: $\sigma_{\text{num_agencia}=1267}$ (**agencia**)

(arquivo organizado por atributo diferente de num_agencia)

n_r – número de tuplas na relação r

b_r – número de blocos que contêm tuplas de r

s_r – tamanho em bytes de uma tupla r

f_r – número de tuplas que cabe em um bloco

$V(A, r)$ – número de valores distintos do atributo A

$SC(A, r)$ – núm médio registros satisfazem condição

f_i – fan-out médio dos nós internos do índice i

HT_i – número de níveis no índice i (altura do índice).

LB_i – número de blocos de índice de nível mais baixo no índice i 48

Como calcular o custo de um PEC

✓ Operação de Seleção usando índices

– A5 (índice **secundário**, igualdade)

- Recupera vários registros se campo indexação não for chave única
- $SC(A, r)$ registros satisfarão a condição
- Pior caso: cada registro que satisfaz condição em 1 bloco diferente.

$$E_{A5} = HT_i + SC(A, r)$$

Exemplo: $\sigma_{\text{nome}=\text{"Maria"}}$ (cliente)
(arquivo organizado por atributo diferente de nome)

n_r – número de tuplas na relação r

b_r – número de blocos que contêm tuplas de r

s_r – tamanho em bytes de uma tupla r

f_r – número de tuplas que cabe em um bloco

$V(A, r)$ – número de valores distintos do atributo A

$SC(A, r)$ –número médio registros satisfazem condição

f_i – fan-out médio dos nós internos do índice i

HT_i – número de níveis no índice i (altura do índice).

LB_i – número de blocos de índice de nível mais baixo no índice i

Como calcular o custo de um PEC

✓ Operação de Seleção usando índices

- A5 (índice **secundário**, igualdade)

$$E_{A5} = HT_i + SC(A, r)$$

Exemplo:

$\sigma_{\text{cidade_agencia}=\text{"Santos"}}$ (**agencia**)
(arquivo *NÃO* organizado por cidade_agencia)

$f_{\text{conta}}=20$

$V(\text{cidade_agencia}, \text{conta}) = 50$

$V(\text{saldo}, \text{conta}) = 500$

$n_{\text{conta}}=10000$

$f_i = 40$ (árvore B+)

$SC(\text{cidade_agencia}, \text{conta}) = 10000/50 = 200$

Número de chaves esperadas no índice = 50

$HT_i = \log_{f_i/2}(50) = 2$

$E_{A4} = 2 + 200 = 202$

n_r – número de tuplas na relação r

b_r – número de blocos que contêm tuplas de r

s_r – tamanho em bytes de uma tupla r

f_r – número de tuplas que cabe em um bloco

$V(A, r)$ – número de valores distintos do atributo A

$SC(A, r)$ – número médio registros satisfazem condição

f_i – fan-out médio dos nós internos do índice i

HT_i – número de níveis no índice i (altura do índice).

LB_i – número de blocos de índice de nível 50 mais baixo no índice i

Como calcular o custo de um PEC

✓ Operação de Seleção usando índices – Exemplo

$\sigma_{\text{nome_cliente}=\text{"Maria"}}(\text{cliente})$

- $f_{\text{cliente}}=10$
- $V(\text{nome_cliente}, \text{cliente}) = 5000$
- $n_{\text{conta}}=20000$
- índice árvore-B+ primário para atributo *nome_cliente*
- índice armazena 10 ponteiros por nós

n_r – número de tuplas na relação r

b_r – número de blocos que contêm tuplas de r

s_r – tamanho em bytes de uma tupla r

f_r – número de tuplas que cabe em um bloco

$V(A, r)$ – número de valores distintos do atributo A

$SC(A, r)$ – número médio registros satisfazem condição

f_i – fan-out médio dos nós internos do índice i

HT_i – número de níveis no índice i (altura do índice).

LB_i – número de blocos de índice de nível i mais baixo no índice i

Como calcular o custo de um PEC

✓ Operação de Seleção usando índices – Exemplo

$$\sigma_{\text{nome_cliente}=\text{"Maria"}}(\text{cliente})$$

- $f_{\text{cliente}}=10$
- $V(\text{nome_cliente}, \text{cliente}) = 5000$
- $n_{\text{conta}}=20000$
- índice árvore-B+ primário para atributo *nome_cliente*
- índice armazena 10 ponteiros por nós

Qual o custo para busca com índice primário?

n_r – número de tuplas na relação r

b_r – número de blocos que contêm tuplas de r

s_r – tamanho em bytes de uma tupla r

f_r – número de tuplas que cabe em um bloco

$V(A, r)$ – número de valores distintos do atributo A

$SC(A, r)$ – número médio registros satisfazem condição

f_i – fan-out médio dos nós internos do índice i

HT_i – número de níveis no índice i (altura do índice).

LB_i – número de blocos de índice de nível mais baixo no índice i

Como calcular o custo de um PEC

✓ Operação de Seleção usando índices – Exemplo

$\sigma_{\text{nome_cliente}=\text{"Maria"}}(\text{cliente})$

- $f_{\text{cliente}}=10$
- $V(\text{nome_cliente}, \text{cliente}) = 5000$
- $n_{\text{conta}}=20000$
- índice árvore-B+ primário para atributo *nome_cliente*
- índice armazena 10 ponteiros por nós

Qual o custo para busca com índice primário?

Qual dos algoritmos anteriores usar?

n_r – número de tuplas na relação r

b_r – número de blocos que contêm tuplas de r

s_r – tamanho em bytes de uma tupla r

f_r – número de tuplas que cabe em um bloco

$V(A, r)$ – número de valores distintos do atributo A

$SC(A, r)$ – número médio registros satisfazem condição

f_i – fan-out médio dos nós internos do índice i

HT_i – número de níveis no índice i (altura do índice).

LB_i – número de blocos de índice de nível mais baixo no índice i

Como calcular o custo de um PEC

✓ Operação de Seleção usando índices – Exemplo

$\sigma_{\text{nome_cliente}=\text{"Maria"}}(\text{cliente})$

- $f_{\text{cliente}}=10$
- $V(\text{nome_cliente}, \text{cliente}) = 5000$
- $n_{\text{conta}}=20000$
- índice árvore-B+ primário para atributo *nome_cliente*
- índice armazena 10 ponteiros por nós

Qual o custo para busca com índice primário?

$$E_{A4} = HT_i + \left[\frac{SC(A, r)}{f_r} \right]$$

n_r – número de tuplas na relação r
 b_r – número de blocos que contêm tuplas de r
 s_r – tamanho em bytes de uma tupla r
 f_r – número de tuplas que cabe em um bloco
 $V(A, r)$ – número de valores distintos do atributo A
 $SC(A, r)$ – número médio registros satisfazem condição

f_i – fan-out médio dos nós internos do índice i
 HT_i – número de níveis no índice i (altura do índice).
 LB_i – número de blocos de índice de nível mais baixo no índice i

Como calcular o custo de um PEC

✓ Operação de Seleção usando índices – Exemplo

$\sigma_{\text{nome_cliente}=\text{"Maria"}}(\text{cliente})$

- $f_{\text{cliente}}=10$
- $V(\text{nome_cliente}, \text{cliente}) = 5000$
- $n_{\text{conta}}=20000$
- índice árvore-B+ primário para atributo *nome_cliente*
- índice armazena 10 ponteiros por nós

Qual o custo para busca com índice primário?

$$E_{A4} = HT_i + \left\lceil \frac{SC(A, r)}{f_r} \right\rceil$$

- $SC(\text{nome_cliente}, \text{cliente}) = 20000/5000 = 4 \rightarrow$ leitura de 1 bloco de dados (4/10)
- 5000 nomes diferentes de cliente \rightarrow profundidade da árvore $= \log_{10/2} 5000 \rightarrow$ leitura de 6 blocos de índice
- Então: estratégia requer $6+1=7$ leituras de bloco

n_r – número de tuplas na relação r
 b_r – número de blocos que contêm tuplas de r
 s_r – tamanho em bytes de uma tupla r
 f_r – número de tuplas que cabe em um bloco
 $V(A, r)$ – número de valores distintos do atributo A
 $SC(A, r)$ – número médio registros satisfazem condição

HT_i – número de níveis no índice i (altura do índice).

LB_i – número de blocos de índice de nível mais baixo no índice i 55

Como calcular o custo de um PEC

✓ Operação de Seleção usando índices – Exemplo

$\sigma_{\text{nome_cliente}=\text{"Maria"}}(\text{cliente})$

- $f_{\text{cliente}}=10$
- $V(\text{nome_cliente}, \text{cliente}) = 5000$
- $n_{\text{conta}}=20000$
- índice árvore-B+ **secundário** para atributo *nome_cliente*
- índice armazena 10 ponteiros por nós

Qual o custo para busca com índice secundário ?

n_r – número de tuplas na relação r

b_r – número de blocos que contêm tuplas de r

s_r – tamanho em bytes de uma tupla r

f_r – número de tuplas que cabe em um bloco

$V(A, r)$ – número de valores distintos do atributo A

$SC(A, r)$ – número médio registros satisfazem condição

f_i – fan-out médio dos nós internos do índice i

HT_i – número de níveis no índice i (altura do índice).

LB_i – número de blocos de índice de nível mais baixo no índice i

Como calcular o custo de um PEC

✓ Operação de Seleção usando índices – Exemplo

$\sigma_{\text{nome_cliente}=\text{"Maria"}}(\text{cliente})$

- $f_{\text{cliente}}=10$
- $V(\text{nome_cliente}, \text{cliente}) = 5000$
- $n_{\text{conta}}=20000$
- índice árvore-B+ secundário para atributo *nome_cliente*
- índice armazena 10 ponteiros por nós

Qual o custo para busca com índice secundário?

$$E_{A5} = HT_i + SC(A, r)$$

n_r – número de tuplas na relação r
 b_r – número de blocos que contêm tuplas de r
 s_r – tamanho em bytes de uma tupla r
 f_r – número de tuplas que cabe em um bloco
 $V(A, r)$ – número de valores distintos do atributo A
 $SC(A, r)$ – número médio registros satisfazem condição

f_i – fan-out médio dos nós internos do índice i
 HT_i – número de níveis no índice i (altura do índice).
 LB_i – número de blocos de índice de nível mais baixo no índice i

Como calcular o custo de um PEC

✓ Operação de Seleção usando índices – Exemplo

$\sigma_{\text{nome_cliente}=\text{"Maria"}}(\text{cliente})$

- $f_{\text{cliente}}=10$
- $V(\text{nome_cliente}, \text{cliente}) = 5000$
- $n_{\text{conta}}=20000$
- índice árvore-B+ primário para atributo *nome_cliente*
- índice armazena 10 ponteiros por nós

Qual o custo para busca com índice primário?

$$E_{A5} = HT_i + SC(A, r)$$

- $SC(\text{nome_cliente}, \text{cliente}) = 20000/5000 = 4$
- 5000 nomes diferentes de cliente → profundidade da árvore = $\log_{10/2} 5000$ → leitura de 6 blocos de índice
- Então: estratégia requer 6+4=10 leituras de bloco

n_r – número de tuplas na relação r
 b_r – número de blocos que contêm tuplas de r
 s_r – tamanho em bytes de uma tupla r
 f_r – número de tuplas que cabe em um bloco
 $V(A, r)$ – número de valores distintos do atributo A
 $SC(A, r)$ – número médio registros satisfazem condição

T_i – fan-out médio dos nós internos do índice i

HT_i – número de níveis no índice i (altura do índice).

LB_i – número de blocos de índice de nível mais baixo no índice i 58

Como calcular o custo de um PEC

✓ Operação de Seleção com comparações

– A6 (índice **primário**, comparação)

$$\sigma_A \leq_v (r)$$

- suposição: metade dos registros satisfaz a condição → resultado tem $n_r/2$ tuplas
- condição do tipo $A \geq v$: ???

Como calcular o custo de um PEC

✓ Operação de Seleção com comparações

– A6 (índice **primário**, comparação)

$$\sigma_A \leq_v (r)$$

- suposição: metade dos registros satisfaz a condição → resultado tem $n_r/2$ tuplas
- condição do tipo $A \geq v$: encontra primeiro registro que atende a condição e varre arquivo até final.

Como calcular o custo de um PEC

✓ Operação de Seleção com comparações

$$\sigma_A \leq v(r)$$

– A6 (índice primário, comparação)

- suposição: metade dos registros satisfaz a condição → resultado tem $n_r/2$ tuplas
- condição do tipo $A \geq v$: encontra primeiro registro que atende a condição e varre arquivo até final.

$$E_{A6} = HT_i + \frac{b_r}{2}$$

ou

$$E_{A6} = HT_i + \frac{c}{f_r}$$

n_r – número de tuplas na relação r

b_r – número de blocos que contêm tuplas de r

s_r – tamanho em bytes de uma tupla r

f_r – número de tuplas que cabe em um bloco

$V(A, r)$ – número de valores distintos do atributo A

$SC(A, r)$ – núm médio registros satisfazem condição

f_i – fan-out médio dos nós internos do índice i

HT_i – número de níveis no índice i (altura do índice).

LB_i – número de blocos de índice de nível mais baixo no índice i

Como calcular o custo de um PEC

✓ Operação de Seleção com comparações

$$\sigma_A \leq v(r)$$

– A6 (índice primário, comparação)

- suposição: metade dos registros satisfaz a condição → resultado tem $n_r/2$ tuplas
- condição do tipo $A \geq v$: encontra primeiro registro que atende a condição e varre arquivo até final.

$$E_{A6} = HT_i + \frac{b_r}{2}$$

ou

$$E_{A6} = HT_i + \frac{c}{f_r}$$

n_r – número de tuplas na relação r

b_r – número de blocos que contêm tuplas de r

s_r – tamanho em bytes de uma tupla r

f_r – número de tuplas que cabe em um bloco

$V(A, r)$ – número de valores distintos do atributo A

$SC(A, r)$ – núm médio registros satisfazem condição

- se valor real usado na comparação (v) estiver disponível, estimativa pode ser mais precisa
→ c = número de valores que satisfazem condição

Como calcular o custo de um PEC

✓ Operação de Seleção com comparações

$$\sigma_A \leq v (r)$$

– A6 (índice primário, comparação)

- suposição: metade dos registros satisfaz a condição → resultado tem $n_r/2$ tuplas
- condição do tipo $A \geq v$: encontra primeiro registro que atende a condição e varre arquivo até final.

$$E_{A6} = HT_i + \frac{b_r}{2} \quad \text{ou}$$

$$E_{A6} = HT_i + \frac{c}{f_r}$$

- condição do tipo $A \leq v$: ???

n_r – número de tuplas na relação r

b_r – número de blocos que contêm tuplas de r

s_r – tamanho em bytes de uma tupla r

f_r – número de tuplas que cabe em um bloco

$V(A, r)$ – número de valores distintos do atributo A

$SC(A, r)$ – núm médio registros satisfazem condição

f_i – fan-out médio dos nós internos do índice i

HT_i – número de níveis no índice i (altura do índice).

LB_i – número de blocos de índice de nível mais baixo no índice i

Como calcular o custo de um PEC

✓ Operação de Seleção com comparações

$$\sigma_A \leq v (r)$$

– A6 (índice primário, comparação)

- suposição: metade dos registros satisfaz a condição → resultado tem $n_r/2$ tuplas
- condição do tipo $A \geq v$: encontra primeiro registro que atende a condição e varre arquivo até final.

$$E_{A6} = HT_i + \frac{b_r}{2}$$

ou

$$E_{A6} = HT_i + \frac{c}{f_r}$$

- condição do tipo $A \leq v$: varre do início do arquivo até encontrar tupla com valor $\geq v$. Índice não é necessário.

n_r – número de tuplas na relação r

b_r – número de blocos que contêm tuplas de r

s_r – tamanho em bytes de uma tupla r

f_r – número de tuplas que cabe em um bloco

$V(A, r)$ – número de valores distintos do atributo A

$SC(A, r)$ – núm médio registros satisfazem condição

f_i – fan-out médio dos nós internos do índice i

HT_i – número de níveis no índice i (altura do índice).

LB_i – número de blocos de índice de nível mais baixo no índice i

Como calcular o custo de um PEC

✓ Operação de Seleção com comparações - Exemplo

$\sigma_{\text{nome_agencia} \leq \text{"Centro"}} (\text{conta})$

$$E_{A6} = \frac{b_r}{2}$$

- $f_{\text{conta}} = 20$
- $V(\text{nome_agencia}, \text{conta}) = 50$
- $V(\text{saldo}, \text{conta}) = 500$
- $n_{\text{conta}} = 10000$
- índice árvore-B+ primário para atributo *nome_agencia*
- índice armazena 20 ponteiros por nós

• Esperamos metade das tuplas com valores \leq "Centro" → 5000 tuplas
→ 5000/20 blocos → 250 blocos

$$E_{A6} = 250$$

n_r – número de tuplas na relação r
 b_r – número de blocos que contêm tuplas de r
 s_r – tamanho em bytes de uma tupla r
 f_r – número de tuplas que cabe em um bloco
 $V(A, r)$ – número de valores distintos do atributo A
 $SC(A, r)$ – núm médio registros satisfazem condição

f_i – fan-out médio dos nós internos do índice i
 HT_i – número de níveis no índice i (altura do índice).
 LB_i – número de blocos de índice de nível mais baixo no índice i

Como calcular o custo de um PEC

✓ Operação de Seleção com comparações - Exemplo

$\sigma_{\text{nome_agencia} \geq \text{"Centro"}} (\text{conta})$

- $f_{\text{conta}} = 20$
- $V(\text{nome_agencia}, \text{conta}) = 50$
- $V(\text{saldo}, \text{conta}) = 500$
- $n_{\text{conta}} = 10000$
- índice árvore-B+ primário para atributo *nome_agencia*
- índice armazena 20 ponteiros por nós

$$E_{A6} = HT_i + \frac{b_r}{2}$$

- Esperamos metade das tuplas com valores \leq "Centro" → 5000 tuplas → 5000/20 blocos → 250 blocos

$$HT_i = \log_{f_i/2}(50) = 2$$

$$E_{A6} = 2 + 250 = 252 \text{ blocos}$$

n_r – número de tuplas na relação r
 b_r – número de blocos que contêm tuplas de r
 s_r – tamanho em bytes de uma tupla r
 f_r – número de tuplas que cabe em um bloco
 $V(A, r)$ – número de valores distintos do atributo A
 $SC(A, r)$ – núm médio registros satisfazem condição

f_i – fan-out médio dos nós internos do índice i
 HT_i – número de níveis no índice i (altura do índice).
 LB_i – número de blocos de índice de nível mais baixo no índice i

Como calcular o custo de um PEC

✓ Operação de Seleção com comparações

$$\sigma_A \leq_v (r)$$

– A7 (índice **secundário**, comparação)

- **blocos de índice de nível mais baixo** são varridos a partir do maior (\geq) ou menor valor (\leq) até v .
- suposição: metade dos registros satisfaz a condição → **quantos blocos de índice de nível mais baixo serão acessados ???**

n_r – número de tuplas na relação r

b_r – número de blocos que contêm tuplas de r

s_r – tamanho em bytes de uma tupla r

f_r – número de tuplas que cabe em um bloco

$V(A, r)$ – número de valores distintos do atributo A

$SC(A, r)$ – núm médio registros satisfazem condição

f_i – fan-out médio dos nós internos do índice i

HT_i – número de níveis no índice i (altura do índice).

LB_i – número de blocos de índice de nível mais baixo no índice i

Como calcular o custo de um PEC

✓ Operação de Seleção com comparações

$$\sigma_A \leq v (r)$$

– A7 (índice **secundário**, comparação)

- **blocos de índice de nível mais baixo** são varridos a partir do maior (\geq) ou menor valor (\leq) até v .
- suposição: metade dos registros satisfaz a condição → **quantos blocos de índice de nível mais baixo nível serão acessados ???**
METADE

n_r – número de tuplas na relação r

b_r – número de blocos que contêm tuplas de r

s_r – tamanho em bytes de uma tupla r

f_r – número de tuplas que cabe em um bloco

$V(A, r)$ – número de valores distintos do atributo A

$SC(A, r)$ – núm médio registros satisfazem condição

f_i – fan-out médio dos nós internos do índice i

HT_i – número de níveis no índice i (altura do índice).

LB_i – número de blocos de índice de nível mais baixo no índice i

Como calcular o custo de um PEC

✓ Operação de Seleção com comparações

$$\sigma_A \leq v (r)$$

– A7 (índice **secundário**, comparação)

- **blocos de índice de nível mais baixo** são varridos a partir do maior (\geq) ou menor valor (\leq) até v .
- suposição: metade dos registros satisfaz a condição → **quantos blocos de índice de nível mais baixo nível serão acessados ???**
METADE
- E ainda: caminho percorrido no índice a partir da raiz do bloco até primeiro bloco folha ser usado.

n_r – número de tuplas na relação r

b_r – número de blocos que contêm tuplas de r

s_r – tamanho em bytes de uma tupla r

f_r – número de tuplas que cabe em um bloco

$V(A, r)$ – número de valores distintos do atributo A

$SC(A, r)$ – núm médio registros satisfazem condição

f_i – fan-out médio dos nós internos do índice i

HT_i – número de níveis no índice i (altura do índice).

LB_i – número de blocos de índice de nível mais baixo no índice i

Como calcular o custo de um PEC

✓ Operação de Seleção com comparações

$$\sigma_A \leq v (r)$$

– A7 (índice **secundário**, comparação)

- **blocos de índice de nível mais baixo** são varridos a partir do maior (\geq) ou menor valor (\leq) até v .
- suposição: metade dos registros satisfaz a condição → **quantos blocos de índice de nível mais baixo nível serão acessados ??? METADE**
- E ainda: caminho percorrido no índice a partir da raiz do bloco até primeiro bloco folha ser usado.

$$E_{A7} = HT_i + \frac{LB_i}{2} + \frac{n_r}{2}$$

n_r – número de tuplas na relação r

b_r – número de blocos que contêm tuplas de r

s_r – tamanho em bytes de uma tupla r

f_r – número de tuplas que cabe em um bloco

$V(A, r)$ – número de valores distintos do atributo A

$SC(A, r)$ – núm médio registros satisfazem condição

f_i – fan-out médio dos nós internos do índice i

HT_i – número de níveis no índice i (altura do índice).

LB_i – número de blocos de índice de nível mais baixo no índice i 70

Como calcular o custo de um PEC

✓ Operação de Seleção com comparações – Exemplo

$$\sigma_{\text{saldo} < 1500} (\text{conta})$$

- $f_{\text{conta}}=20$
- $V(\text{nome_agencia}, \text{conta}) = 50$
- $V(\text{saldo}, \text{conta}) = 500$
- $n_{\text{conta}}=10000$
- índice secundário árvore-B+ para *saldo*
- índice armazena 20 ponteiros por nó

n_r – número de tuplas na relação r

b_r – número de blocos que contêm tuplas de r

s_r – tamanho em bytes de uma tupla r

f_r – número de tuplas que cabe em um bloco

$V(A, r)$ – número de valores distintos do atributo A

$SC(A, r)$ – núm médio registros satisfazem condição

f_i – fan-out médio dos nós internos do índice i

HT_i – número de níveis no índice i (altura do índice).

LB_i – número de blocos de índice de nível mais baixo no índice i

Como calcular o custo de um PEC

✓ Operação de Seleção com comparações – Exemplo

$$\sigma_{\text{saldo} < 1500}(\text{conta})$$

- $f_{\text{conta}}=20$
- $V(\text{nome_agencia}, \text{conta}) = 50$
- $V(\text{saldo}, \text{conta}) = 500$
- $n_{\text{conta}}=10000$

Qual o custo para busca com índice secundário?

$$E_{A7} = HT_i + \frac{LB_i}{2} + \frac{n_r}{2}$$

- índice secundário árvore-B+ para *saldo*
- índice armazena 20 ponteiros por nó

n_r – número de tuplas na relação r

b_r – número de blocos que contêm tuplas de r

s_r – tamanho em bytes de uma tupla r

f_r – número de tuplas que cabe em um bloco

$V(A, r)$ – número de valores distintos do atributo A

$SC(A, r)$ – núm médio registros satisfazem condição

f_i – fan-out médio dos nós internos do índice i

HT_i – número de níveis no índice i (altura do índice).

LB_i – número de blocos de índice de nível mais baixo no índice i

Como calcular o custo de um PEC

✓ Operação de Seleção com comparações – Exemplo

$$\sigma_{\text{saldo}} < 1500 \quad (\text{conta})$$

- $f_{\text{conta}}=20$
- $V(\text{nome_agencia}, \text{conta}) = 50$
- $V(\text{saldo}, \text{conta}) = 500$
- $n_{\text{conta}}=10000$
- índice secundário árvore-B+ para *saldo*
- índice armazena 20 ponteiros por nó

$$E_{A7} = HT_i + \frac{LB_i}{2} + \frac{n_r}{2}$$

n_r – número de tuplas na relação r
 b_r – número de blocos que contêm tuplas de r
 s_r – tamanho em bytes de uma tupla r
 f_r – número de tuplas que cabe em um bloco
 $V(A, r)$ – número de valores distintos do atributo A
 $SC(A, r)$ – número médio registros satisfazem condição

Qual o custo para busca com índice secundário?

- 500 valores diferentes de saldo e cada nó folha armazenado deve preencher pelo menos a metade das chaves → árvore tem entre 25 e 50 nós folha
- Profundidade da árvore = $\log_{fi/2}(500) = \log_{10} 500$ → leitura de 3 blocos de índice
- Pior caso: 50 nós folhas e tem que acessar metade → 25 leituras de blocos
 - Para cada tupla → recuperá-la da relação → 5000 tuplas satisfazem a condição → 5000 leituras de blocos (pior caso)
- Então: estratégia requer 5028 (3 + 25 + 5000) acessos a blocos

HT_i – número de níveis no índice i (altura do índice).

LB_i – número de blocos de índice de nível mais baixo no índice i

Como calcular o custo de um PEC

✓ Operação de Seleção com comparações – Exemplo

$$\sigma_{\text{saldo} < 1500} (\text{conta})$$

- $f_{\text{conta}}=20$
- $V(\text{nome_agencia}, \text{conta}) = 50$
- $V(\text{saldo}, \text{conta}) = 500$
- $n_{\text{conta}}=10000$
- índice secundário árvore-B+ para *saldo*
- índice armazena 21 ponteiros por nó

Qual o custo dessa busca considerando varredura simples do arquivo, sem índice?

$$E_{A7} = HT_i + \frac{LB_i}{2} + \frac{n_r}{2}$$

n_r – número de tuplas na relação *r*

b_r – número de blocos que contêm tuplas de *r*

s_r – tamanho em bytes de uma tupla *r*

f_r – número de tuplas que cabe em um bloco

$V(A, r)$ – número de valores distintos do atributo *A*

$SC(A, r)$ – número médio registros satisfazem condição

f_i – fan-out médio dos nós internos do índice *i*

HT_i – número de níveis no índice *i* (altura do índice).

LB_i – número de blocos de índice de nível mais baixo no índice *i*

Como calcular o custo de um PEC

✓ Operação de Seleção com comparações – Exemplo

$$\sigma_{\text{saldo} < 1500} \text{ (conta)}$$

- $f_{\text{conta}}=20$
- $V(\text{nome_agencia}, \text{conta}) = 50$
- $V(\text{saldo}, \text{conta}) = 500$
- $n_{\text{conta}}=10000$
- índice secundário árvore-B+ para *saldo*
- índice armazena 21 ponteiros por nós

$$E_{A7} = HT_i + \frac{LB_i}{2} + \frac{n_r}{2}$$

Qual o custo dessa busca considerando varredura simples do arquivo, sem índice?

- $10000/20 = 500$ acessos de bloco !!!

n_r – número de tuplas na relação r
 b_r – número de blocos que contêm tuplas de r
 s_r – tamanho em bytes de uma tupla r
 f_r – número de tuplas que cabe em um bloco
 $V(A, r)$ – número de valores distintos do atributo A
 $SC(A, r)$ – número médio registros satisfazem condição

f_i – fan-out médio dos nós internos do índice i
 HT_i – número de níveis no índice i (altura do índice).
 LB_i – número de blocos de índice de nível mais baixo no índice i

Como calcular o custo de um PEC

✓ Seleções Complexas (Conjunção, Disjunção, Negação)

- Deve se calcular o custo de cada seleção
- Mas, por onde começamos?
- Exemplo:

$\sigma_{\text{nome_agencia}=\text{"Centro"} \wedge \text{saldo} = 1500 \wedge \text{estado} = \text{"SP"} \text{ (conta)}}$

n_r – número de tuplas na relação r

b_r – número de blocos que contêm tuplas de r

s_r – tamanho em bytes de uma tupla r

f_r – número de tuplas que cabe em um bloco

$V(A, r)$ – número de valores distintos do atributo A

$SC(A, r)$ – núm médio registros satisfazem condição

f_i – fan-out médio dos nós internos do índice i

HT_i – número de níveis no índice i (altura do índice).

LB_i – número de blocos de índice de nível mais baixo no índice i

Como calcular o custo de um PEC

✓ Seleções Complexas (Conjunção, Disjunção, Negação)

- Deve se calcular o custo de cada seleção
- Mas, por onde começamos?
- Exemplo:

$\sigma_{\text{nome_agencia}=\text{"Centro"} \wedge \text{saldo} = 1500 \wedge \text{estado} = \text{"SP"} (\text{conta})$

- Precisamos conhecer a **seletividade** de cada predicado

n_r – número de tuplas na relação r

b_r – número de blocos que contêm tuplas de r

s_r – tamanho em bytes de uma tupla r

f_r – número de tuplas que cabe em um bloco

$V(A, r)$ – número de valores distintos do atributo A

$SC(A, r)$ – núm médio registros satisfazem condição

f_i – fan-out médio dos nós internos do índice i

HT_i – número de níveis no índice i (altura do índice).

LB_i – número de blocos de índice de nível mais baixo no índice i

Como calcular o custo de um PEC

✓ Seleções Complexas (Conjunção, Disjunção, Negação)

✓ Conjunção

$$\sigma_{\theta_1 \wedge \theta_2 \wedge \dots \wedge \theta_n}(r)$$

- estimar o tamanho da seleção $\sigma_{\theta_i}(r)$, denominado s_i (qtde tuplas)
- probabilidade de uma tupla satisfazer condição $\sigma_{\theta_i} \rightarrow s_i/n_r \rightarrow$ esta probabilidade é chamada de **seletividade**
- se condições forem **independentes** entre si: probabilidade de uma tupla satisfazer todas as condições é o produto de todas as probabilidades
- Então, tamanho da seleção completa =
$$n_r * \frac{s_1 * s_2 * \dots * s_n}{n_r^n}$$

n_r – número de tuplas na relação r

b_r – número de blocos que contêm tuplas de r

s_r – tamanho em bytes de uma tupla r

f_r – número de tuplas que cabe em um bloco

$V(A, r)$ – número de valores distintos do atributo A

$SC(A, r)$ – núm médio registros satisfazem condição

f_i – fan-out médio dos nós internos do índice i

HT_i – número de níveis no índice i (altura do índice).

LB_i – número de blocos de índice de nível mais baixo no índice i

Como calcular o custo de um PEC

✓ Seleções Complexas (Conjunção, Disjunção, Negação)

✓ Disjunção

$$\sigma_{\theta_1 \vee \theta_2 \dots \vee \theta_n}(r)$$

- estimar o tamanho da seleção $\sigma_{\theta_i}(r)$, denominado s_i (qtde tuplas)
- probabilidade de uma tupla satisfazer condição $\sigma_{\theta_i} \rightarrow s_i / n_r$
probabilidade de uma tupla satisfazer disjunção $\rightarrow 1 - \text{probabilidade de que ela não satisfaça nenhuma das condições:}$
$$1 - \left(1 - \frac{s_1}{n_r}\right) * \left(1 - \frac{s_2}{n_r}\right) * \dots * \left(1 - \frac{s_n}{n_r}\right)$$
- tamanho da seleção:
$$n_r * \left[1 - \left(1 - \frac{s_1}{n_r}\right) * \left(1 - \frac{s_2}{n_r}\right) * \dots * \left(1 - \frac{s_n}{n_r}\right)\right]$$

n_r – número de tuplas na relação r

b_r – número de blocos que contêm tuplas de r

s_r – tamanho em bytes de uma tupla r

f_r – número de tuplas que cabe em um bloco

$V(A, r)$ – número de valores distintos do atributo A

$SC(A, r)$ – núm médio registros satisfazem condição

f_i – fan-out médio dos nós internos do índice i

HT_i – número de níveis no índice i (altura do índice).

LB_i – número de blocos de índice de nível mais baixo no índice i

Como calcular o custo de um PEC

- ✓ Seleções Complexas (Conjunção, Disjunção, Negação)
- ✓ Negação

$$\sigma_{\neg_{\theta}}(r)$$

- Resultado: tuplas de r que não estão em $\sigma_{\theta}(r)$
- tamanho:

$$\text{tamanho}(r) - \text{tamanho}(\sigma_{\theta}(r))$$

n_r – número de tuplas na relação r

b_r – número de blocos que contêm tuplas de r

s_r – tamanho em bytes de uma tupla r

f_r – número de tuplas que cabe em um bloco

$V(A, r)$ – número de valores distintos do atributo A

$SC(A, r)$ – núm médio registros satisfazem condição

f_i – fan-out médio dos nós internos do índice i

HT_i – número de níveis no índice i (altura do índice).

LB_i – número de blocos de índice de nível mais baixo no índice i

Como calcular o custo de um PEC

✓ Seleções Complexas – Usam algoritmos anteriores

- A1 (busca linear)
- A2 (busca binária)
- A3 (índice primário, igualdade de chave)
- A4 (índice primário, igualdade em atributo não chave)
- A5 (índice secundário, igualdade)
- A6 (índice primário, comparação)
- A7 (índice secundário, comparação)

n_r – número de tuplas na relação r

b_r – número de blocos que contêm tuplas de r

s_r – tamanho em bytes de uma tupla r

f_r – número de tuplas que cabe em um bloco

$V(A, r)$ – número de valores distintos do atributo A

$SC(A, r)$ – núm médio registros satisfazem condição

f_i – fan-out médio dos nós internos do índice i

HT_i – número de níveis no índice i (altura do índice).

LB_i – número de blocos de índice de nível mais baixo no índice i

Como calcular o custo de um PEC

✓ Seleções Complexas – Algoritmos (Conjunção, Disjunção)

– A8 (seleção de conjunção usando um índice)

- determinar se há um caminho de acesso para um dos atributos em uma das condições simples → escolher um dos algoritmos (A2 a A7) para recuperar registros daquela condição
- completar operação testando se cada registro no buffer de memória satisfaz as demais condições simples
- Para determinar ordem: primeira varredura deveria ser com condição mais seletiva (menor **seletividade**).

A1 (busca linear)
A2 (busca binária)
A3 (índice primário, igualdade de chave)

A4 (índice primário, igualdade em atributo não chave)
A5 (índice secundário, igualdade)
A6 (índice primário, comparação)
A7 (índice secundário, comparação)

n_r – número de tuplas na relação r
 b_r – número de blocos que contêm tuplas de r
 s_r – tamanho em bytes de uma tupla r
 f_r – número de tuplas que cabe em um bloco
 $V(A, r)$ – número de valores distintos do atributo A
 $SC(A, r)$ – número médio registros satisfazem condição

f_i – fan-out médio dos nós internos do índice i
 HT_i – número de níveis no índice i (altura do índice).
 LB_i – número de blocos de índice de nível mais baixo no índice i

Como calcular o custo de um PEC

✓ Seleções Complexas – Algoritmos (Conjunção, Disjunção)

- A9 (seleção de conjunção usando **índice composto**)
 - índice composto → índice que usa os atributos combinados
 - usar cálculo de probabilidade para achar menor seletividade
 - procurar diretamente no índice
 - usar melhor algoritmo (A3, A4 ou A5)

| | |
|---|--|
| A1 (busca linear) A2 (busca binária) A3 (índice primário, igualdade de chave) | A4 (índice primário, igualdade em atributo não chave) A5 (índice secundário, igualdade) A6 (índice primário, comparação) A7 (índice secundário, comparação) |
| n_r – número de tuplas na relação r b_r – número de blocos que contêm tuplas de r s_r – tamanho em bytes de uma tupla r f_r – número de tuplas que cabe em um bloco $V(A, r)$ – número de valores distintos do atributo A $SC(A, r)$ – núm médio registros satisfazem condição | f_i – fan-out médio dos nós internos do índice i HT_i – número de níveis no índice i (altura do índice). LB_i – número de blocos de índice de nível mais baixo no índice i |

Como calcular o custo de um PEC

✓ Seleções Complexas – Algoritmos (Conjunção, Disjunção)

- A10 (seleção de conjunção usando intersecção de identificadores)
 - identificadores → ponteiros de registros
 - requer índices com registros de ponteiros para atributos das condições individuais
 - cada índice é varrido buscando-se os ponteiros que satisfazem condição
 - intersecção entre ponteiros recuperados fornece resultado.
 - se não há índices disponíveis para todas condições ???

A1 (busca linear)
A2 (busca binária)
A3 (índice primário, igualdade de chave)

A4 (índice primário, igualdade em atributo não chave)
A5 (índice secundário, igualdade)
A6 (índice primário, comparação)
A7 (índice secundário, comparação)

n_r – número de tuplas na relação r
 b_r – número de blocos que contêm tuplas de r
 s_r – tamanho em bytes de uma tupla r
 f_r – número de tuplas que cabe em um bloco
 $V(A, r)$ – número de valores distintos do atributo A
 $SC(A, r)$ – número médio registros satisfazem condição

f_i – fan-out médio dos nós internos do índice i
 HT_i – número de níveis no índice i (altura do índice).
 LB_i – número de blocos de índice de nível mais baixo no índice i

Como calcular o custo de um PEC

✓ Seleções Complexas – Algoritmos (Conjunção, Disjunção)

- A10 (seleção de conjunção usando **intersecção de identificadores**)
 - identificadores → ponteiros de registros
 - requer índices com registros de ponteiros para atributos das condições individuais
 - cada índice é varrido buscando-se os ponteiros que satisfazem a condição
 - intersecção entre ponteiros recuperados fornece resultado.
 - **se não há índices disponíveis para todas condições ???** registros recuperados são testados para condições restantes (sem índices)

| | |
|---|---|
| A1 (busca linear) | A4 (índice primário, igualdade em atributo não chave) |
| A2 (busca binária) | A5 (índice secundário, igualdade) |
| A3 (índice primário, igualdade de chave) | A6 (índice primário, comparação) |
| | A7 (índice secundário, comparação) |
| n_r – número de tuplas na relação r | f_i – fan-out médio dos nós internos do índice i |
| b_r – número de blocos que contêm tuplas de r | HT_i – número de níveis no índice i (altura do índice). |
| s_r – tamanho em bytes de uma tupla r | LB_i – número de blocos de índice de nível mais baixo no índice i |
| f_r – número de tuplas que cabe em um bloco | |
| $V(A, r)$ – número de valores distintos do atributo A | |
| $SC(A, r)$ – núm médio registros satisfazem condição | |

Como calcular o custo de um PEC

- ✓ Seleções Complexas – Algoritmos (Conjunção, Disjunção)
 - A11 (seleção de disjunção usando **união de identificadores**)
 - requer índices com registros de ponteiros para atributos das condições individuais
 - cada índice é varrido buscando-se os ponteiros que satisfazem a condição
 - união entre ponteiros recuperados fornece resultado.
 - se não há índices disponíveis para uma condição: varredura linear da relação para encontrar tuplas que satisfazem condição.

| | | | |
|---|--|--|--|
| A1 (busca linear) A2 (busca binária) A3 (índice primário, igualdade de chave) | | A4 (índice primário, igualdade em atributo não chave) A5 (índice secundário, igualdade) A6 (índice primário, comparação) A7 (índice secundário, comparação) | |
| n_r – número de tuplas na relação r | | f_i – fan-out médio dos nós internos do índice i | |
| b_r – número de blocos que contêm tuplas de r | | HT_i – número de níveis no índice i (altura do índice). | |
| s_r – tamanho em bytes de uma tupla r | | LB_i – número de blocos de índice de nível mais baixo no índice i | |
| $V(A, r)$ – número de valores distintos do atributo A | | | |
| $SC(A, r)$ –número médio registros satisfazem condição | | | |

Como calcular o custo de um PEC

✓ Seleções Complexas – Exemplo

$\sigma_{\text{nome_agencia}=\text{"Centro"} \wedge \text{saldo} = 1500}(\text{conta})$

- $f_{\text{conta}}=20$
- $V(\text{nome_agencia}, \text{conta}) = 50$
- $V(\text{saldo}, \text{conta}) = 500$
- $n_{\text{conta}}=10000$
- índice primário para *nome_agencia*
- índice secundário árvore-B+ para *saldo*
- índices armazenam 20 ponteiros por nós

Qual o custo dessa busca ?

A1 (busca linear)
A2 (busca binária)
A3 (índice primário, igualdade de chave)

A4 (índice primário, igualdade em atributo não chave)
A5 (índice secundário, igualdade)
A6 (índice primário, comparação)
A7 (índice secundário, comparação)

n_r – número de tuplas na relação *r*
 b_r – número de blocos que contêm tuplas de *r*
 s_r – tamanho em bytes de uma tupla *r*
 f_r – número de tuplas que cabe em um bloco
 $V(A, r)$ – número de valores distintos do atributo *A*
 $SC(A, r)$ – número médio registros satisfazem condição

f_i – fan-out médio dos nós internos do índice *i*
 HT_i – número de níveis no índice *i* (altura do índice).
 LB_i – número de blocos de índice de nível mais baixo no índice *i*

Como calcular o custo de um PEC

✓ Seleções Complexas – Exemplo

$\sigma_{\text{nome_agencia}=\text{"Centro"} \wedge \text{saldo} = 1500}(\text{conta})$

- $f_{\text{conta}}=20$
- $V(\text{nome_agencia}, \text{conta}) = 50$
- $V(\text{saldo}, \text{conta}) = 500$
- $n_{\text{conta}}=10000$
- índice primário para *nome_agencia*
- índice secundário árvore-B+ para *saldo*
- índices armazenam 20 ponteiros por nós

Que PEC são possíveis ?

A1 (busca linear)
A2 (busca binária)
A3 (índice primário, igualdade de chave)

A4 (índice primário, igualdade em atributo não chave)
A5 (índice secundário, igualdade)
A6 (índice primário, comparação)
A7 (índice secundário, comparação)

n_r – número de tuplas na relação *r*
 b_r – número de blocos que contêm tuplas de *r*
 s_r – tamanho em bytes de uma tupla *r*
 f_r – número de tuplas que cabe em um bloco
 $V(A, r)$ – número de valores distintos do atributo *A*
 $SC(A, r)$ – número médio registros satisfazem condição

f_i – fan-out médio dos nós internos do índice *i*
 HT_i – número de níveis no índice *i* (altura do índice).
 LB_i – número de blocos de índice de nível mais baixo no índice *i*

Como calcular o custo de um PEC

✓ Seleções Complexas – Exemplo

$\sigma_{\text{nome_agencia}=\text{"Centro"} \wedge \text{saldo} = 1500}(\text{conta})$

- $f_{\text{conta}}=20$
- $V(\text{nome_agencia}, \text{conta}) = 50$
- $V(\text{saldo}, \text{conta}) = 500$
- $n_{\text{conta}}=10000$
- índice primário para *nome_agencia*
- índice secundário árvore-B+ para *saldo*
- índices armazenam 20 ponteiros por nós

Qual o custo dessa busca ?

1) usando índice para *nome_agencia*
= 12 leituras de bloco (já calculado antes)

A1 (busca linear)
A2 (busca binária)
A3 (índice primário, igualdade de chave)

A4 (índice primário, igualdade em atributo não chave)
A5 (índice secundário, igualdade)
A6 (índice primário, comparação)
A7 (índice secundário, comparação)

n_r – número de tuplas na relação *r*
 b_r – número de blocos que contêm tuplas de *r*
 s_r – tamanho em bytes de uma tupla *r*
 f_r – número de tuplas que cabe em um bloco
 $V(A, r)$ – número de valores distintos do atributo *A*
 $SC(A, r)$ – número médio registros satisfazem condição

f_i – fan-out médio dos nós internos do índice *i*
 HT_i – número de níveis no índice *i* (altura do índice).
 LB_i – número de blocos de índice de nível mais baixo no índice *i*

Como calcular o custo de um PEC

✓ Seleções Complexas – Exemplo

$\sigma_{\text{nome_agencia}=\text{"Centro"} \wedge \text{saldo} = 1500}(\text{conta})$

- $f_{\text{conta}}=20$
- $V(\text{nome_agencia}, \text{conta}) = 50$
- $V(\text{saldo}, \text{conta}) = 500$
- $n_{\text{conta}}=10000$
- índice primário para *nome_agencia*
- índice secundário árvore-B+ para *saldo*
- índices armazenam 20 ponteiros por nós

Qual o custo dessa busca ?

- 1) usando índice para *nome_agencia* = 12 leituras de bloco
- 2) Usando índice para *saldo* ???

A1 (busca linear)
A2 (busca binária)
A3 (índice primário, igualdade de chave)

A4 (índice primário, igualdade em atributo não chave)
A5 (índice secundário, igualdade)
A6 (índice primário, comparação)
A7 (índice secundário, comparação)

n_r – número de tuplas na relação *r*
 b_r – número de blocos que contêm tuplas de *r*
 s_r – tamanho em bytes de uma tupla *r*
 f_r – número de tuplas que cabe em um bloco
 $V(A, r)$ – número de valores distintos do atributo *A*
 $SC(A, r)$ – número médio registros satisfazem condição

f_i – fan-out médio dos nós internos do índice *i*
 HT_i – número de níveis no índice *i* (altura do índice).
 LB_i – número de blocos de índice de nível mais baixo no índice *i*

Como calcular o custo de um PEC

✓ Seleções Complexas – Exemplo

$\sigma_{\text{nome_agencia}=\text{"Centro"} \wedge \text{saldo} = 1500}(\text{conta})$

- $f_{\text{conta}}=20$
- $V(\text{nome_agencia}, \text{conta}) = 50$
- $V(\text{saldo}, \text{conta}) = 500$
- $n_{\text{conta}}=10000$
- índice primário para *nome_agencia*
- índice secundário árvore-B+ para *saldo*
- índices armazenam 20 ponteiros por nós

Qual o custo dessa busca ?

1) usando índice para *nome_agencia* =
12 leituras de bloco

2) Usando índice para *saldo* ???

$$E_{A5} = HT_i + SC(A, r)$$

A1 (busca linear)
A2 (busca binária)
A3 (índice primário, igualdade de chave)

A4 (índice primário, igualdade em atributo não chave)
A5 (índice secundário, igualdade)
A6 (índice primário, comparação)
A7 (índice secundário, comparação)

n_r – número de tuplas na relação *r*
 b_r – número de blocos que contêm tuplas de *r*
 s_r – tamanho em bytes de uma tupla *r*
 f_r – número de tuplas que cabe em um bloco
 $V(A, r)$ – número de valores distintos do atributo *A*
 $SC(A, r)$ – número médio registros satisfazem condição

f_i – fan-out médio dos nós internos do índice *i*
 HT_i – número de níveis no índice *i* (altura do índice).
 LB_i – número de blocos de índice de nível mais baixo no índice *i*

Como calcular o custo de um PEC

✓ Seleções Complexas – Exemplo

$\sigma_{\text{nome_agencia}=\text{"Centro"} \wedge \text{saldo} = 1500}(\text{conta})$

- $f_{\text{conta}}=20$
- $V(\text{nome_agencia}, \text{conta}) = 50$
- $V(\text{saldo}, \text{conta}) = 500$
- $n_{\text{conta}}=10000$
- índice primário para *nome_agencia*
- índice secundário árvore-B+ para *saldo*
- índices armazenam 20 ponteiros por nós

Qual o custo dessa busca ?

1) usando índice para *nome_agencia* =
12 leituras de bloco

2) Usando índice para *saldo*:

- $V(\text{saldo}, \text{conta}) = 500 \rightarrow$ espera-se que $10000/500=20$ tuplas tenham saldo 1200
- índice para saldo não é primário: uma leitura de bloco para cada tupla (pior caso) \rightarrow 20 leituras de bloco
- Profundidade do índice: $\log_{\lceil f_i/2 \rceil} 500 = 3 \rightarrow$ 3 leituras de bloco
- Então: $20 + 3=23$ leituras de blocos

Conclusão: melhor pelo índice por *nome_agencia*

n_r – número de tuplas na relação *r*
 b_r – número de blocos que contêm tuplas de *r*
 s_r – tamanho em bytes de uma tupla *r*
 f_r – número de tuplas que cabe em um bloco
 $V(A, r)$ – número de valores distintos do atributo *A*
 $SC(A, r)$ –número médio registros satisfazem condição

mais baixo no índice *i*

Como calcular o custo de um PEC

✓ Seleções Complexas – Exemplo

$\sigma_{\text{nome_agencia}=\text{"Centro"} \wedge \text{saldo} = 1500}(\text{conta})$

- $f_{\text{conta}}=20$
- $V(\text{nome_agencia}, \text{conta}) = 50$
- $V(\text{saldo}, \text{conta}) = 500$
- $n_{\text{conta}}=10000$
- índice primário para *nome_agencia*
- índice secundário árvore-B+ para *saldo*
- índices armazenam 20 ponteiros por nós

Qual o custo dessa busca se índice por *nome_agencia* fosse secundário?

n_r – número de tuplas na relação r

b_r – número de blocos que contêm tuplas de r

s_r – tamanho em bytes de uma tupla r

f_r – número de tuplas que cabe em um bloco

$V(A, r)$ – número de valores distintos do atributo A

$SC(A, r)$ – número médio registros satisfazem condição

f_i – fan-out médio dos nós internos do índice i

HT_i – número de níveis no índice i (altura do índice).

LB_i – número de blocos de índice de nível mais baixo no índice i

Como calcular o custo de um PEC

✓ Seleções Complexas – Exemplo

$\sigma_{\text{nome_agencia}=\text{"Centro"} \wedge \text{saldo} = 1500}(\text{conta})$

- $f_{\text{conta}}=20$
- $V(\text{nome_agencia}, \text{conta}) = 50$
- $V(\text{saldo}, \text{conta}) = 500$
- $n_{\text{conta}}=10000$
- índice primário para *nome_agencia*
- índice secundário árvore-B+ para *saldo*
- índices armazenam 20 ponteiros por nós

Qual o custo dessa busca se índice por *nome_agencia* fosse secundário?

- $V(\text{saldo}, \text{conta}) = 500 \rightarrow$ espera-se que $10000/500=20$ tuplas tenham saldo = 1200 **$\rightarrow 20$ leituras de bloco + 3 leituras de blocos de índice = 23**
- $V(\text{nome_agencia}, \text{conta}) = 50 \rightarrow$ espera-se que $10000/50=200$ tuplas **$\rightarrow 200$ leituras de bloco + 2 leituras de blocos de índice ($\log_{\lceil f_i/2 \rceil} 50$) = 202**

Conclusão: melhor pelo índice por *saldo*

n_r – número de tuplas na relação *r*

b_r – número de blocos que contêm tuplas de *r*

s_r – tamanho em bytes de uma tupla *r*

f_r – número de tuplas que cabe em um bloco

$V(A, r)$ – número de valores distintos do atributo *A*

$SC(A, r)$ – núm médio registros satisfazem condição

f_i

HT_i – número de níveis no índice *i* (altura do índice).

LB_i – número de blocos de índice de nível mais baixo no índice *i*

Como calcular o custo de um PEC

✓ Classificação

- Por que é necessária?

Como calcular o custo de um PEC

✓ Classificação

- Por que é necessária?
 - Consultas com saídas ordenadas
 - Junções mais eficazes
- Se relação cabe inteira na memória: técnicas clássicas (quicksort, por exemplo)
- Qual é o problema daquelas que não cabem na memória?

Como calcular o custo de um PEC

✓ Classificação

- Por que é necessária?
 - Consultas com saídas ordenadas
 - Junções mais eficazes
- Se relação cabe inteira na memória: técnicas clássicas (quicksort, por exemplo)
- Qual é o problema daquelas que não cabem na memória?
 - Novamente a questão do acesso a disco!!!
 - Para esses casos: algoritmo de ordenação externa: **sort-merge externo**

Como calcular o custo de um PEC

✓ Classificação

- Algoritmo **sort-merge externo**
- M = quantidade de blocos que cabem na memória principal

Fase 1

$i = 0$

repita

leia M blocos da relação (ou relação até final se for menor que M)

ordene a parte que está na memória

escreva dados ordenados no arquivo temporário R_i

$i = i + 1$

até fim da relação

Como calcular o custo de um PEC

✓ Classificação

- Algoritmo **sort-merge externo**
- **M** = quantidade de blocos que cabem na memória principal

E a Fase 2 ???

Como calcular o custo de um PEC

✓ Classificação

- Algoritmo **sort-merge externo**
- M = quantidade de blocos que cabem na memória principal

Fase 2 (merge dos arquivos temporários)

leia um bloco de cada um dos N arquivos R_i para uma página de buffer na memória (*considerando $N < M$*)

repita

escolha a primeira tupla entre todas as páginas de buffer
escreva tupla no resultado e apague-a da página do buffer
se página de buffer de qualquer R_i for vazia e não for fim de R_i

leia próximo bloco de R_i na página do buffer
até todas páginas de buffer estiverem vazias

Como calcular o custo de um PEC

✓ Classificação

- sort-merge externo supondo que somente 1 tupla cabe em um bloco e que a memória mantém no máximo 3 *frames*. Durante o merge são usados 2 frames e reservado 1 para guardar o resultado

| | |
|---|----|
| g | 24 |
| a | 19 |
| d | 31 |
| c | 33 |
| b | 14 |
| e | 16 |
| r | 16 |
| d | 21 |
| m | 3 |
| p | 2 |
| d | 7 |
| a | 14 |

**Relação
inicial**

| | |
|---|----|
| a | 19 |
| d | 31 |
| g | 24 |

| | |
|---|----|
| b | 14 |
| c | 33 |
| e | 16 |

| | |
|---|----|
| d | 21 |
| m | 3 |
| r | 16 |

| | |
|---|----|
| a | 14 |
| d | 7 |
| p | 2 |

Temporários



Criação dos arquivos temporários

Como calcular o custo de um PEC

✓ Classificação

- sort-merge externo supondo que somente 1 tupla cabe em um bloco e que a memória mantém no máximo 3 *frames*. Durante o merge são usados 2 frames e reservado 1 para guardar o resultado

| | |
|---|----|
| g | 24 |
| a | 19 |
| d | 31 |
| c | 33 |
| b | 14 |
| e | 16 |
| r | 16 |
| d | 21 |
| m | 3 |
| p | 2 |
| d | 7 |
| a | 14 |

**Relação
inicial**

| | |
|---|----|
| a | 19 |
| d | 31 |
| g | 24 |

| | |
|---|----|
| b | 14 |
| c | 33 |
| e | 16 |

| | |
|---|----|
| d | 21 |
| m | 3 |
| r | 16 |

| | |
|---|----|
| a | 14 |
| d | 7 |
| p | 2 |

Temporários

Como calcular o custo de um PEC

✓ Classificação

- sort-merge externo supondo que somente 1 tupla cabe em um bloco e que a memória mantém no máximo 3 *frames*. Durante o merge são usados 2 frames e reservado 1 para guardar o resultado

| | |
|---|----|
| g | 24 |
| a | 19 |
| d | 31 |
| c | 33 |
| b | 14 |
| e | 16 |
| r | 16 |
| d | 21 |
| m | 3 |
| p | 2 |
| d | 7 |
| a | 14 |

**Relação
inicial**

| | |
|---|----|
| a | 19 |
| d | 31 |
| g | 24 |

| | |
|---|----|
| b | 14 |
| c | 33 |
| e | 16 |

| | |
|---|----|
| d | 21 |
| m | 3 |
| r | 16 |

| | |
|---|----|
| a | 14 |
| d | 7 |
| p | 2 |

Temporários

| | |
|---|----|
| a | 19 |
|---|----|

Como calcular o custo de um PEC

✓ Classificação

- sort-merge externo supondo que somente 1 tupla cabe em um bloco e que a memória mantém no máximo 3 *frames*. Durante o merge são usados 2 frames e reservado 1 para guardar o resultado

| | |
|---|----|
| g | 24 |
| a | 19 |
| d | 31 |
| c | 33 |
| b | 14 |
| e | 16 |
| r | 16 |
| d | 21 |
| m | 3 |
| p | 2 |
| d | 7 |
| a | 14 |

**Relação
inicial**

| | |
|---|----|
| | |
| d | 31 |
| g | 24 |

| | |
|---|----|
| b | 14 |
| c | 33 |
| e | 16 |

| | |
|---|----|
| d | 21 |
| m | 3 |
| r | 16 |

| | |
|---|----|
| a | 14 |
| d | 7 |
| p | 2 |

Temporários

| | |
|---|----|
| a | 19 |
|---|----|

Como calcular o custo de um PEC

✓ Classificação

- sort-merge externo supondo que somente 1 tupla cabe em um bloco e que a memória mantém no máximo 3 *frames*. Durante o merge são usados 2 frames e reservado 1 para guardar o resultado

| | |
|---|----|
| g | 24 |
| a | 19 |
| d | 31 |
| c | 33 |
| b | 14 |
| e | 16 |
| r | 16 |
| d | 21 |
| m | 3 |
| p | 2 |
| d | 7 |
| a | 14 |

**Relação
inicial**

| | |
|---|----|
| | |
| d | 31 |
| g | 24 |

| | |
|---|----|
| | |
| c | 33 |
| e | 16 |

| | |
|---|----|
| d | 21 |
| m | 3 |
| r | 16 |

| | |
|---|----|
| a | 14 |
| d | 7 |
| p | 2 |

Temporários

| | |
|---|----|
| a | 19 |
| b | 14 |

Como calcular o custo de um PEC

✓ Classificação

- sort-merge externo supondo que somente 1 tupla cabe em um bloco e que a memória mantém no máximo 3 *frames*. Durante o merge são usados 2 frames e reservado 1 para guardar o resultado

| | |
|---|----|
| g | 24 |
| a | 19 |
| d | 31 |
| c | 33 |
| b | 14 |
| e | 16 |
| r | 16 |
| d | 21 |
| m | 3 |
| p | 2 |
| d | 7 |
| a | 14 |

**Relação
inicial**

| | |
|---|----|
| | |
| d | 31 |
| g | 24 |

| | |
|---|----|
| | |
| | |
| e | 16 |

| | |
|---|----|
| d | 21 |
| m | 3 |
| r | 16 |

| | |
|---|----|
| a | 14 |
| d | 7 |
| p | 2 |

Temporários

| | |
|---|----|
| a | 19 |
| b | 14 |
| c | 33 |

Como calcular o custo de um PEC

✓ Classificação

- sort-merge externo supondo que somente 1 tupla cabe em um bloco e que a memória mantém no máximo 3 *frames*. Durante o merge são usados 2 frames e reservado 1 para guardar o resultado

| | |
|---|----|
| g | 24 |
| a | 19 |
| d | 31 |
| c | 33 |
| b | 14 |
| e | 16 |
| r | 16 |
| d | 21 |
| m | 3 |
| p | 2 |
| d | 7 |
| a | 14 |

**Relação
inicial**

| | |
|---|----|
| | |
| | |
| g | 24 |

| | |
|---|----|
| | |
| | |
| e | 16 |

| | |
|---|----|
| d | 21 |
| m | 3 |
| r | 16 |

| | |
|---|----|
| a | 14 |
| d | 7 |
| p | 2 |

Temporários

| | |
|---|----|
| a | 19 |
| b | 14 |
| c | 33 |
| d | 31 |

Como calcular o custo de um PEC

✓ Classificação

- sort-merge externo supondo que somente 1 tupla cabe em um bloco e que a memória mantém no máximo 3 *frames*. Durante o merge são usados 2 frames e reservado 1 para guardar o resultado

| | |
|---|----|
| g | 24 |
| a | 19 |
| d | 31 |
| c | 33 |
| b | 14 |
| e | 16 |
| r | 16 |
| d | 21 |
| m | 3 |
| p | 2 |
| d | 7 |
| a | 14 |

**Relação
inicial**

| | |
|---|----|
| | |
| | |
| g | 24 |

| | |
|--|--|
| | |
| | |
| | |

| | |
|---|----|
| d | 21 |
| m | 3 |
| r | 16 |

| | |
|---|----|
| a | 14 |
| d | 7 |
| p | 2 |

Temporários

| | |
|---|----|
| a | 19 |
| b | 14 |
| c | 33 |
| d | 31 |
| e | 16 |

Como calcular o custo de um PEC

✓ Classificação

- sort-merge externo supondo que somente 1 tupla cabe em um bloco e que a memória mantém no máximo 3 *frames*. Durante o merge são usados 2 frames e reservado 1 para guardar o resultado

| | |
|---|----|
| g | 24 |
| a | 19 |
| d | 31 |
| c | 33 |
| b | 14 |
| e | 16 |
| r | 16 |
| d | 21 |
| m | 3 |
| p | 2 |
| d | 7 |
| a | 14 |

**Relação
inicial**

| | |
|--|--|
| | |
| | |
| | |

| | |
|--|--|
| | |
| | |
| | |

| | |
|---|----|
| d | 21 |
| m | 3 |
| r | 16 |

| | |
|---|----|
| a | 14 |
| d | 7 |
| p | 2 |

Temporários

| | |
|---|----|
| a | 19 |
| b | 14 |
| c | 33 |
| d | 31 |
| e | 16 |
| g | 24 |

Como calcular o custo de um PEC

✓ Classificação

- sort-merge externo supondo que somente 1 tupla cabe em um bloco e que a memória mantém no máximo 3 *frames*. Durante o merge são usados 2 frames e reservado 1 para guardar o resultado

| | |
|---|----|
| g | 24 |
| a | 19 |
| d | 31 |
| c | 33 |
| b | 14 |
| e | 16 |
| r | 16 |
| d | 21 |
| m | 3 |
| p | 2 |
| d | 7 |
| a | 14 |

**Relação
inicial**

| | |
|---|----|
| a | 19 |
| d | 31 |
| g | 24 |

| | |
|---|----|
| b | 14 |
| c | 33 |
| e | 16 |

| | |
|---|----|
| d | 21 |
| m | 3 |
| r | 16 |

| | |
|---|----|
| a | 14 |
| d | 7 |
| p | 2 |

Temporários

| | |
|---|----|
| a | 19 |
| b | 14 |
| c | 33 |
| d | 31 |
| e | 16 |
| g | 24 |

| | |
|---|----|
| a | 14 |
| d | 7 |
| d | 21 |
| m | 3 |
| p | 2 |
| r | 16 |

Temporários

Passo 1 do merge

Como calcular o custo de um PEC

✓ Classificação

- sort-merge externo supondo que somente 1 tupla cabe em um bloco e que a memória mantém no máximo 3 *frames*. Durante o merge são usados 2 frames e reservado 1 para guardar o resultado

| | |
|---|----|
| g | 24 |
| a | 19 |
| d | 31 |
| c | 33 |
| b | 14 |
| e | 16 |
| r | 16 |
| d | 21 |
| m | 3 |
| p | 2 |
| d | 7 |
| a | 14 |

**Relação
inicial**

| | |
|---|----|
| a | 19 |
| d | 31 |
| g | 24 |

| | |
|---|----|
| b | 14 |
| c | 33 |
| e | 16 |

| | |
|---|----|
| d | 21 |
| m | 3 |
| r | 16 |

| | |
|---|----|
| a | 14 |
| d | 7 |
| p | 2 |

Temporários

| | |
|---|----|
| a | 19 |
| b | 14 |
| c | 33 |
| d | 31 |
| e | 16 |
| g | 24 |

| | |
|---|----|
| a | 14 |
| d | 7 |
| d | 21 |
| m | 3 |
| p | 2 |
| r | 16 |

Temporários

| | |
|---|----|
| a | 14 |
| a | 19 |
| b | 14 |
| c | 33 |
| d | 7 |
| d | 21 |
| d | 31 |
| e | 16 |
| g | 24 |
| m | 3 |
| p | 2 |
| r | 16 |

**Resultado
classificado**

**Passo 2
do
merge**

Como calcular o custo de um PEC

✓ Classificação

- Algoritmo **sort-merge externo**
- **M** = quantidade de blocos que cabem na memória principal
- Se **$i > M$** ???

Como calcular o custo de um PEC

✓ Classificação

- Algoritmo **sort-merge externo**
- M = quantidade de blocos que cabem na memória principal
- Se $i > M$???
 - merge em múltiplos passos
 - cada merge terá $M - 1$ arquivos temporários de entrada
- Como fazer:
 - merge nos primeiros $M - 1$ temporários, obtendo um único temporário para o próximo passo;
 - merge nos próximos $M - 1$ de forma semelhante;
 - repete até que todos temporários iniciais tenham sido processados;
 - cada passo reduz o número de temporários por um fator $M - 1$.

Como calcular o custo de um PEC

✓ Classificação

- Quantas transferências de bloco são necessárias ?
 - Primeira fase:
 - todo bloco da relação é lido e escrito novamente: ???



Como calcular o custo de um PEC

✓ Classificação

- Quantas transferências de bloco são necessárias ?
 - Primeira fase:
 - todo bloco da relação é lido e escrito novamente: $2 b_r$
 - número inicial de temporários: ???



Como calcular o custo de um PEC

✓ Classificação

- Quantas transferências de bloco são necessárias ?
 - Primeira fase:
 - todo bloco da relação é lido e escrito novamente: $2 b_r$
 - número inicial de temporários: $[b_r / M]$
 - número de temporários diminui por um fator de $M - 1$ em cada passo de merge → número total de passos de merge necessários: ???



Como calcular o custo de um PEC

✓ Classificação

- Quantas transferências de bloco são necessárias ?
 - Primeira fase:
 - todo bloco da relação é lido e escrito novamente: $2 b_r$
 - número inicial de temporários: $[b_r / M]$
 - número de temporários diminui por um fator de $M - 1$ em cada passo de merge → número total de passos de merge necessários: $[\log_{M-1} (b_r / M)]$
 - cada um desses passos lê cada bloco da relação uma vez e os escreve uma vez, com 2 exceções:
 - » passo final pode produzir resultado ordenado sem escrevê-lo no disco;
 - » pode haver temporários não lidos ou não escritos durante um passo (se houver M temporários, o merge será feito sobre $M-1$).



Como calcular o custo de um PEC

✓ Classificação

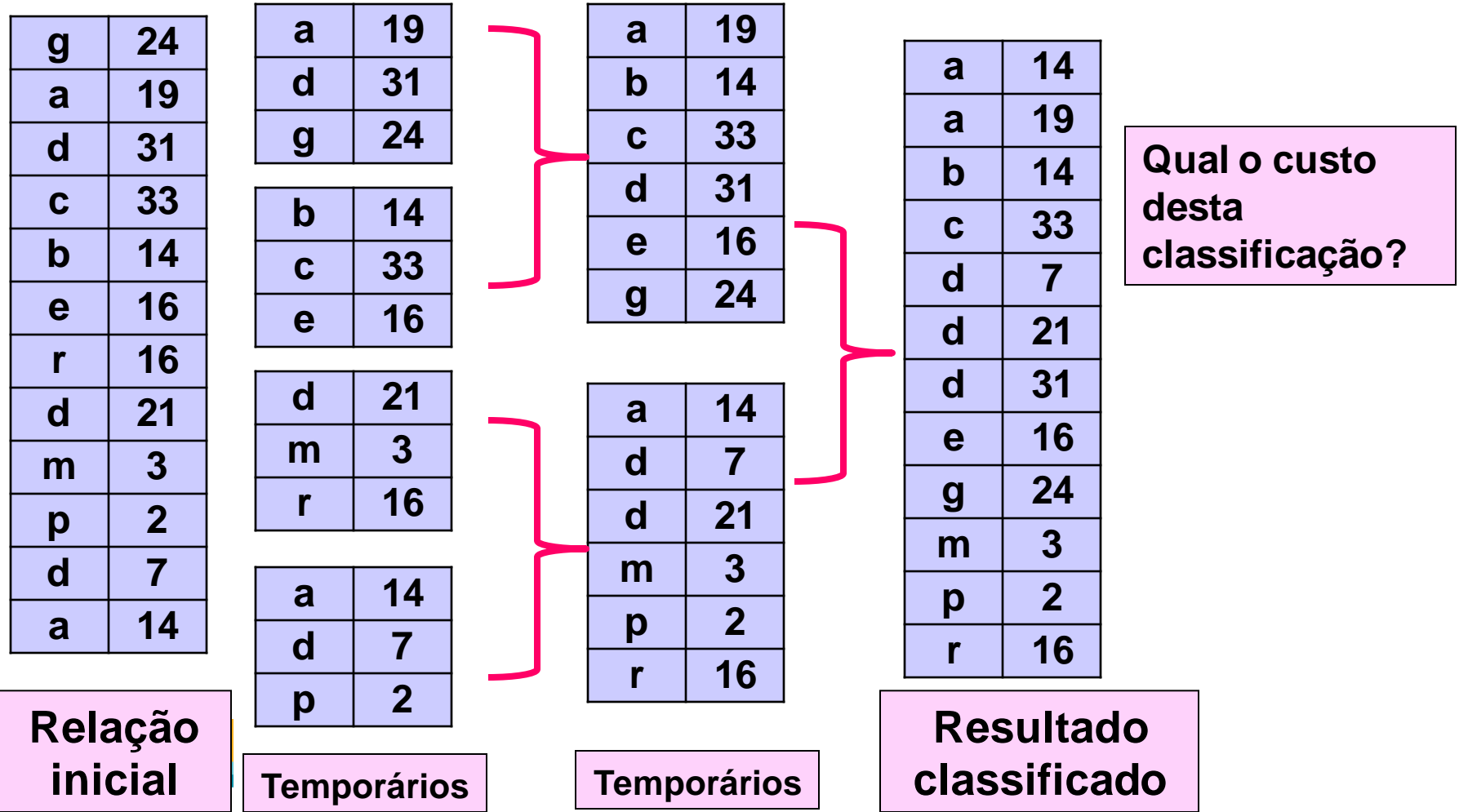
- Quantas transferências de bloco são necessárias ?
 - Primeira fase: todo bloco da relação é lido e escrito novamente: $2 b_r$
 - número inicial de temporários: $\lceil b_r / M \rceil$
 - número de temporários diminui por um favor de $M - 1$ em cada passo de merge → número total de passos de merge necessários: $\lceil \log_{M-1} (b_r / M) \rceil$.
 - cada uma dessas passadas lê cada bloco da relação uma vez e os escreve uma vez, com 2 exceções:
 - » passo final pode produzir resultado ordenado sem escrevê-lo no disco;
 - » pode haver temporários não lidos ou não escritos durante um passo.
 - Ignorando essas situações de exceção, custo total será:

$$b_r \left(2 \lceil \log_{M-1} (b_r / M) \rceil + 1 \right)$$

Como calcular o custo de um PEC

✓ Classificação – Exemplo

$$b_r \left(2 \left[\log_{M-1} (b_r / M) \right] + 1 \right)$$



Como calcular o custo de um PEC

✓ Classificação – Exemplo

$$b_r \left(2 \left[\log_{M-1} (b_r / M) \right] + 1 \right)$$

| | |
|---|----|
| g | 24 |
| a | 19 |
| d | 31 |
| c | 33 |
| b | 14 |
| e | 16 |
| r | 16 |
| d | 21 |
| m | 3 |
| p | 2 |
| d | 7 |
| a | 14 |

**Relação
inicial**

| | |
|---|----|
| a | 19 |
| d | 31 |
| g | 24 |

| | |
|---|----|
| b | 14 |
| c | 33 |
| e | 16 |

| | |
|---|----|
| d | 21 |
| m | 3 |
| r | 16 |

| | |
|---|----|
| a | 14 |
| d | 7 |
| p | 2 |

Temporários

| | |
|---|----|
| a | 19 |
| b | 14 |
| c | 33 |
| d | 31 |
| e | 16 |
| g | 24 |

| | |
|---|----|
| a | 14 |
| d | 7 |
| d | 21 |
| m | 3 |
| p | 2 |
| r | 16 |

Temporários

| | |
|---|----|
| a | 14 |
| a | 19 |
| b | 14 |
| c | 33 |
| d | 7 |
| d | 21 |
| d | 31 |
| e | 16 |
| g | 24 |
| m | 3 |
| p | 2 |
| r | 16 |

**Resultado
classificado**

Qual o custo desta
classificação?

1 tupla = 1 bloco
 $M = 3$

$$12(2[\log_2(12/3)]+1) =$$

$$12 \cdot (4+1) = 60$$

Como calcular o custo de um PEC

✓ Operação de Junção

- vamos usar o termo *equi-join* para indicar $r \bowtie_{A=B} s$, onde A e B são conjuntos de atributos das relações r e s , respectivamente.
- O que precisamos saber:
 - estimar tamanho das junções;
 - algoritmos e seus custos para executar junções.

Como calcular o custo de um PEC

✓ Operação de Junção - tamanho das junções - exemplo

depositante |X| cliente

- $n_{\text{cliente}} = 10.000$
- $f_{\text{cliente}} = 25$
- $b_{\text{cliente}} = 10.000/25 = 400$
- $n_{\text{depositante}} = 5.000$
- $f_{\text{depositante}} = 50$
- $b_{\text{depositante}} = 5.000/50 = 100$
- $V(\text{nome_cliente}, \text{depositante}) = 2500 \rightarrow$
em média, cada cliente tem duas contas
- $V(\text{nome_cliente}, \text{cliente}) = 10000$

Qual a estimativa de tamanho para *depositante |X| cliente* sem utilizar informações sobre chaves estrangeiras?

n_r – número de tuplas na relação r
 b_r – número de blocos que contêm tuplas de r
 s_r – tamanho em bytes de uma tupla r
 f_r – número de tuplas que cabe em um bloco
 $V(A, r)$ – número de valores distintos do atributo A
 $SC(A, r)$ – número médio registros satisfazem condição

f_i – fan-out médio dos nós internos do índice i
 HT_i – número de níveis no índice i (altura do índice).
 LB_i – número de blocos de índice de nível mais baixo no índice i

Como calcular o custo de um PEC

✓ Operação de Junção - tamanho das junções - exemplo

depositante |X| cliente

- $n_{\text{cliente}} = 10.000$
- $f_{\text{cliente}} = 25$
- $b_{\text{cliente}} = 10.000/25 = 400$
- $n_{\text{depositante}} = 5.000$
- $f_{\text{depositante}} = 50$
- $b_{\text{depositante}} = 50.000/50 = 100$
- $V(\text{nome_cliente}, \text{depositante}) = 2500$
→ em média, cada cliente tem duas contas
- $V(\text{nome_cliente}, \text{cliente}) = 10000$

Qual a estimativa de tamanho para *depositante |X| cliente* sem utilizar informações sobre chaves estrangeiras?

• $V(\text{nome_cliente}, \text{depositante}) = 2500$ e $V(\text{nome_cliente}, \text{cliente}) = 10000$:

1) $5000 * 10000/2500 = 20000$

2) $5000 * 10000/10000 = 5000$

Escolhemos a menor estimativa: 5000.

n_r – número de tuplas na relação r

b_r – número de blocos que contêm tuplas de r

s_r – tamanho em bytes de uma tupla r

f_r – número de tuplas que cabe em um bloco

$V(A, r)$ – número de valores distintos do atributo A

$SC(A, r)$ – número médio registros satisfazem condição

f_i –

HT_i – número de níveis no índice i (altura do índice).

LB_i – número de blocos de índice de nível mais baixo no índice i

Como calcular o custo de um PEC

✓ Operação de Junção - tamanho das junções - exemplo

depositante |X| cliente

- $n_{\text{cliente}} = 10.000$
- $f_{\text{cliente}} = 25$
- $b_{\text{cliente}} = 10.000/25 = 400$
- $n_{\text{depositante}} = 5.000$
- $f_{\text{depositante}} = 50$
- $b_{\text{depositante}} = 50.000/50 = 100$
- $V(\text{nome_cliente}, \text{depositante}) = 2500$
→ em média, cada cliente tem duas contas
- $V(\text{nome_cliente}, \text{cliente}) = 10000$

n_r – número de tuplas na relação r

b_r – número de blocos que contêm tuplas de r

s_r – tamanho em bytes de uma tupla r

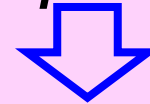
f_r – número de tuplas que cabe em um bloco

$V(A, r)$ – número de valores distintos do atributo A

$SC(A, r)$ – número médio registros satisfazem condição

Qual a estimativa de tamanho para *depositante |X| cliente* utilizando informações sobre chaves estrangeiras?

- se $\text{Depositante} \cap \text{Cliente}$ é uma chave estrangeira para **Cliente** → número de tuplas de *Depositante |X| Cliente* é exatamente igual ao número de tuplas de *Depositante* = 5000



mesmo número encontrado anteriormente!

Como calcular o custo de um PEC

- ✓ Operação de Junção – Estimativas de custo
 - Vários algoritmos
 - Junção de laço aninhado
 - Junção de laço aninhado de blocos
 - Junção de laço aninhado indexada
 - Merge-junção
 - Hash-Junção



Como calcular o custo de um PEC

- ✓ Operação de Junção – Estimativas de custo
- ✓ Junção de laço aninhado ($r \bowtie_{\theta} s$)

```
para cada tupla  $t_r$ , em  $r$  faça      //  $r$  é relação externa
  para cada tupla  $t_s$  em  $s$  faça      //  $s$  é relação interna
    se par  $(t_r, t_s)$  satisfaz condição  $\theta$ 
      adicione  $t_r, t_s$  ao resultado
    fim para
  fim para
fim para
```

- custo semelhante varredura linear de arquivo para seleção
 - extensão do algoritmo de seleção: junção θ seguida de projeção para eliminar atributos repetidos para tuplas t_r, t_s antes de incluir no resultado
 - número de pares de tupla a ser considerado: ???

Como calcular o custo de um PEC

- ✓ Operação de Junção – Estimativas de custo
- ✓ Junção de laço aninhado ($r \bowtie_{\theta} s$)

```
para cada tupla  $t_r$ , em  $r$  faça      //  $r$  é relação externa
  para cada tupla  $t_s$  em  $s$  faça      //  $s$  é relação interna
    se par  $(t_r, t_s)$  satisfaz condição  $\theta$ 
      adicione  $t_r, t_s$  ao resultado
    fim para
  fim para
fim para
```

- custo semelhante varredura linear de arquivo para seleção
 - extensão do algoritmo de seleção: junção θ seguida de projeção para eliminar atributos repetidos para tuplas t_r, t_s antes de incluir no resultado
 - número de pares de tupla a ser considerado: $n_r * n_s$
 - pior caso: buffer pode manter somente um bloco de cada relação →
 $n_r * b_s + b_r$ acessos de blocos necessários
 - melhor caso: buffer consegue manter relações inteiras na memória →
 $b_r + b_s$ acessos de blocos



Como calcular o custo de um PEC

- ✓ Operação de Junção – Estimativas de custo
- ✓ Junção de laço aninhado ($r \bowtie_{\theta} s$)

```
para cada tupla  $t_r$ , em  $r$  faça      //  $r$  é relação externa
  para cada tupla  $t_s$  em  $s$  faça      //  $s$  é relação interna
    se par  $(t_r, t_s)$  satisfaz condição  $\theta$ 
      adicione  $t_r, t_s$  ao resultado
  fim para
fim para
```

- custo ser “custo de acesso linear de arquivo para seleção
• extensão da relação s para a junção θ seguida de projeção para eliminar
atributos não necessários antes de incluir no resultado
• número de blocos de s considerado: $n_r * n_s$
• pior caso: $n_r * b_s + b_r$ antes um bloco de cada relação →
 $n_r * b_s + b_r$ acessos de blocos necessários
• melhor caso: buffer consegue manter relações inteiras na memória →
 $b_r + b_s$ acessos de blocos

**quantidade de
blocos de s : $n_r * b_s$**

**quantidade de
blocos de r : b_r**

Como calcular o custo de um PEC

- ✓ Operação de Junção – exemplo
- ✓ Junção de laço aninhado ($r \bowtie s$)

depositante |X| cliente

- $n_{\text{cliente}} = 10.000$
- $f_{\text{cliente}} = 25$
- $b_{\text{cliente}} = 10.000/25 = 400$
- $n_{\text{depositante}} = 5.000$
- $f_{\text{depositante}} = 50$
- $b_{\text{depositante}} = 50.000/50 = 100$
- $V(\text{nome_cliente}, \text{depositante}) = 2500$
- depositante = relação externa
- cliente = relação interna

Qual o custo usando junção de laço aninhado?

- 1) considerando *depositante* como relação externa
- 2) considerando *cliente* como relação externa

- pior caso: $n_r * b_s + b_r$
- melhor caso: $b_r + b_s$

Como calcular o custo de um PEC

- ✓ Operação de Junção – exemplo
- ✓ Junção de laço aninhado ($r \bowtie_{\theta} s$)

depositante \bowtie_X cliente

- $n_{\text{cliente}} = 10.000$
- $f_{\text{cliente}} = 25$
- $b_{\text{cliente}} = 10.000/25 = 400$
- $n_{\text{depositante}} = 5.000$
- $f_{\text{depositante}} = 50$
- $b_{\text{depositante}} = 50.000/50 = 100$
- $V(\text{nome_cliente}, \text{depositante}) = 2500$
- depositante = relação externa
- cliente = relação interna

Qual o custo usando junção de laço aninhado?

1) **depositante como relação externa**

- temos que examinar $5000 * 10000 = 50 * 10^6$ pares de tuplas
- pior caso $\rightarrow 5000 * 400 + 100 = 2.000.100$ acessos de bloco
- melhor caso \rightarrow ler ambas relações de uma só vez $\rightarrow 100 + 400 = 500$ acessos a blocos

2) **cliente como relação externa**

- pior caso: $10000 * 100 + 400 = 1.000.400$ acessos de bloco
- (melhor caso igual 1)



Como calcular o custo de um PEC

- ✓ Operação de Junção – exemplo
- ✓ Junção de laço aninhado ($r \bowtie_{\theta} s$)

depositante \bowtie cliente

- $n_{\text{cliente}} = 10.000$
- $f_{\text{cliente}} = 25$
- $b_{\text{cliente}} = 10.000/25 = 400$
- $n_{\text{depositante}} = 5.000$
- $f_{\text{depositante}} = 50$

Custo muito alto!!!

Como melhorar???

Qual o custo usando junção de laço aninhado?

1) **depositante como relação externa**

- temos que examinar $5000 * 10000 = 50 * 10^6$ pares de tuplas
- pior caso $\rightarrow 5000 * 400 + 100 = 2.000.100$ acessos de bloco
- melhor caso \rightarrow ler ambas relações de uma só vez $\rightarrow 100 + 400 = 500$ acessos a blocos

2) **cliente como relação externa**

- pior caso: $10000 * 100 + 400 = 1.000.400$ acessos de bloco
- (melhor caso igual 1)



Como calcular o custo de um PEC

- ✓ Operação de Junção – exemplo
- ✓ Junção de laço aninhado de blocos
 - alternativa quando buffer for muito pequeno para manter as 2 relações na memória;
 - cada bloco da relação interna é emparelhado com cada bloco da relação externa;
 - dentro de cada par de blocos, as tuplas são emparelhadas.

```
para cada bloco  $B_r$  de  $r$  faça //  $r$  é relação externa
  para cada bloco  $B_s$  de  $s$  faça //  $s$  é relação interna
    para cada tupla  $t_r$  em  $B_r$  faça
      para cada tupla  $t_s$  em  $B_s$  faça
        se par  $(t_r, t_s)$  satisfaz condição  $\theta$ 
          adicione  $t_r, t_s$  ao resultado
      fim para
    fim para
  fim para
fim para
```

Como calcular o custo de um PEC

- ✓ Operação de Junção – exemplo
- ✓ Junção de laço aninhado de blocos
 - pior caso: cada bloco da relação interna s é lido apenas uma vez para cada bloco da relação externa $r \rightarrow b_r * b_s + b_r$ acessos
 - melhor caso: $b_r + b_s$ acessos de blocos

```
para cada bloco  $B_r$  de  $r$  faça //  $r$  é relação externa
  para cada bloco  $B_s$  de  $s$  faça //  $s$  é relação interna
    para cada tupla  $t_r$  em  $B_r$  faça
      para cada tupla  $t_s$  em  $B_s$  faça
        se par  $(t_r, t_s)$  satisfaz condição  $\theta$ 
          adicione  $t_r, t_s$  ao resultado
      fim para
    fim para
  fim para
fim para
```

Como calcular o custo de um PEC

- ✓ Operação de Junção – exemplo
- ✓ Junção de laço aninhado de blocos

depositante |X| cliente

- $n_{\text{cliente}} = 10.000$
- $f_{\text{cliente}} = 25$
- $b_{\text{cliente}} = 10.000/25 = 400$
- $n_{\text{depositante}} = 5.000$
- $f_{\text{depositante}} = 50$
- $b_{\text{depositante}} = 50.000/50 = 100$
- $V(\text{nome_cliente}, \text{depositante}) = 2500$
- **depositante** = relação externa
- **cliente** = relação interna

Qual o custo usando junção de laço aninhado de blocos?

- 1) considerando **depositante** como relação externa
 - 2) considerando **cliente** como relação externa
- pior caso: $b_r * b_s + b_r$
 - melhor caso: $b_r + b_s$

Como calcular o custo de um PEC

- ✓ Operação de Junção – exemplo
- ✓ Junção de laço aninhado de blocos

depositante |X| cliente

- $n_{\text{cliente}} = 10.000$
- $f_{\text{cliente}} = 25$
- $b_{\text{cliente}} = 10.000/25 = 400$
- $n_{\text{depositante}} = 5.000$
- $f_{\text{depositante}} = 50$
- $b_{\text{depositante}} = 50.000/50 = 100$
- $V(\text{nome_cliente}, \text{depositante}) = 2500$
- depositante = relação externa
- cliente = relação interna

Qual o custo usando junção de laço aninhado de blocos

(*depositante* relação externa)?

- pior caso: $b_r * b_s + b_r$
- melhor caso: $b_r + b_s$
- pior caso → $100 * 400 + 100 = 40.100$ acessos de bloco
- melhor caso → ler ambas relações de uma só vez → $100 + 400 = 500$ acessos a blocos

Como calcular o custo de um PEC

- ✓ Operação de Junção – exemplo
- ✓ Junção de laço aninhado de blocos

depositante |X| cliente

- $n_{\text{cliente}} = 10.000$
- $f_{\text{cliente}} = 25$
- $b_{\text{cliente}} = 10.000/25 = 400$
- $n_{\text{depositante}} = 5.000$
- $f_{\text{depositante}} = 50$
- $b_{\text{depositante}} = 50.000/50 = 100$
- $V(\text{nome_cliente}, \text{depositante}) = 2500$
- depositante = relação externa
- cliente = relação interna

Qual o custo usando junção de laço aninhado de blocos (**cliente relação externa**)?

- pior caso: $b_r * b_s + b_r$
- melhor caso: $b_r + b_s$
- pior caso → $400 * 100 + 400 = 40.400$ acessos de bloco
- melhor caso → ler ambas relações de uma só vez → $100 + 400 = 500$ acessos a blocos

Como calcular o custo de um PEC

- ✓ Operação de Junção
- ✓ Como melhorar desempenho dos 2 algoritmos vistos ???



Como calcular o custo de um PEC

- ✓ Operação de Junção
- ✓ Como melhorar desempenho dos 2 algoritmos vistos ???
 - se atributos formarem chave na relação interna: laço interno pode parar quando a primeira correspondência for encontrada;
 - junção aninhada de blocos: em vez de usar blocos de disco como a unidade de bloco para a relação externa → usar maior tamanho que cabe na memória;
 - varrer laço interno alternadamente para frente e para trás → ordena varredura de forma que dados da varredura anterior que permanecem no buffer podem ser usados novamente reduzindo acesso a disco;
 - usar índice no atributo de junção do laço interno (**próximo algoritmo**).



Como calcular o custo de um PEC

✓ Operação de Junção

✓ Junção de laço aninhado indexada

- alternativa quando há índice para atributo de junção do laço interno
- índices podem ser permanentes ou temporários

```
para cada tupla  $t_r$ , em  $r$  faça      //  $r$  é relação externa
    usar índice para procurar tuplas  $t_s$  em  $s$ 
    que satisfaz condição  $\theta$ 
    adicione  $t_r, t_s$  ao resultado
fim para
```

- procura de tuplas é uma seleção em s
- pior caso → espaço no buffer somente para uma página de r e uma página do índice → $b_r + n_r * c$
- c é o custo de seleção única em s usando condição de junção → já visto em custo de seleção utilizando diversos tipos de índice.



Como calcular o custo de um PEC

- ✓ Operação de Junção – exemplo
- ✓ Junção de laço aninhado indexada

depositante |X| cliente

- $n_{\text{cliente}} = 10.000$
- $f_{\text{cliente}} = 25$
- $b_{\text{cliente}} = 10.000/25 = 400$
- $n_{\text{depositante}} = 5.000$
- $f_{\text{depositante}} = 50$
- $b_{\text{depositante}} = 50.000/50 = 100$
- $V(\text{nome_cliente}, \text{depositante}) = 2500$
- $\text{depositante} = \text{relação externa}$
- $\text{cliente} = \text{relação interna}$
- índice árvore-B+ primário no atributo de junção *nome_cliente*, com 20 entradas em cada nó do índice

Qual o custo usando junção de laço aninhado indexada (depositante como relação externa)?

$$b_r + n_r * c$$

- cliente tem 10000 tuplas →
 $HT_i + SC(\text{nome_cliente}, \text{cliente}) = \log_{10} 2500 + (10000/2500)/25 = 5$
- $\therefore c = 5$
- $n_{\text{depositante}} = 5000 \rightarrow$
 $\text{custo} = 100 + 5000 * 5 = 25100$
acessos de disco.

Como calcular o custo de um PEC

✓ Operação de Junção

✓ Merge-Junção

- usado para calcular junção natural ou *equi-join*
- $r(R)$ e $s(S)$: relações sobre as quais será calculada junção
- $R \cap S$: atributos em comum
- r e s estão classificadas nos atributos $R \cap S$
- Algoritmo:
 - associa ponteiro a cada relação – apontam para 1ª tupla
 - ponteiros se movem ao longo da relação
 - grupo de tuplas de uma relação com mesmo valor para os atributos de junção é lido em S_s
 - tuplas correspondentes da outra relação são lidas e armazenadas

*Algoritmo completo: Silberschatz, A.; Korth, H.F.; Sudarshan, S. “Sistema de Banco de Dados”, 3a. edição, Makron Books, 1999 (capítulo 12)

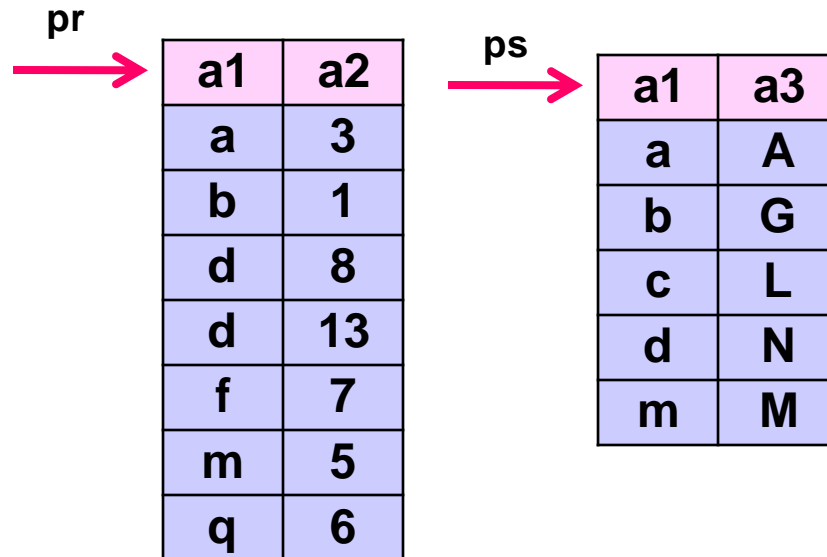


Como calcular o custo de um PEC

✓ Operação de Junção

✓ Merge-Junção

– cada tupla é lida somente uma vez → consequentemente cada bloco é lido somente uma vez



Relações classificadas pelo atributo *a1*.

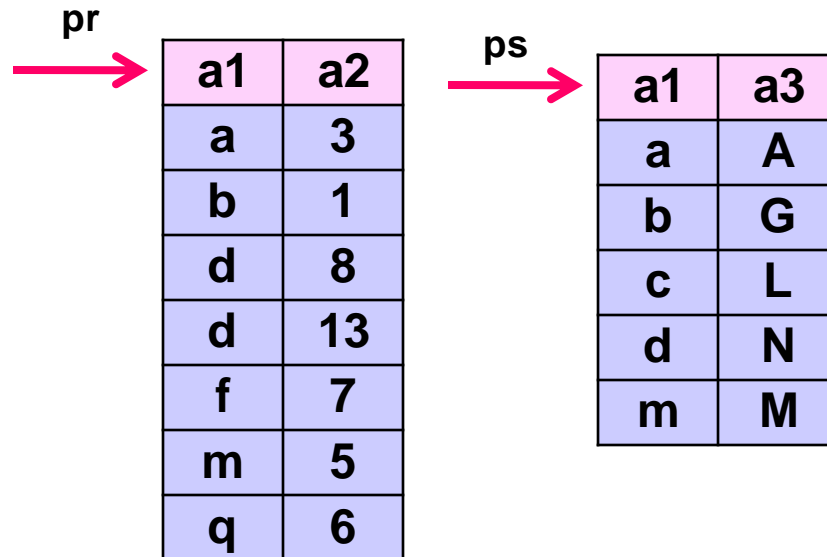
Como calcular o custo de um PEC

✓ Operação de Junção

✓ Merge-Junção

– cada tupla é lida somente uma vez → consequentemente cada bloco é lido somente uma vez

– Custo???



Relações classificadas pelo atributo a1.

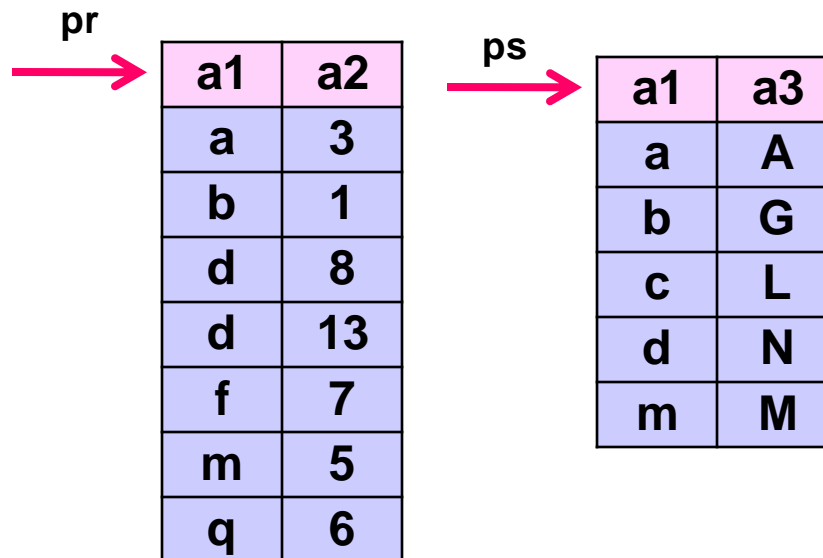
Como calcular o custo de um PEC

✓ Operação de Junção

✓ Merge-Junção

– cada tupla é lida somente uma vez → consequentemente cada bloco é lido somente uma vez

– Custo: $b_r + b_s$



Relações classificadas pelo atributo a_1 .

Como calcular o custo de um PEC

- ✓ Operação de Junção – exemplo
- ✓ Merge-Junção

depositante |X| cliente

- $n_{\text{cliente}} = 10.000$
- $f_{\text{cliente}} = 25$
- $b_{\text{cliente}} = 10.000/25 = 400$
- $n_{\text{depositante}} = 5.000$
- $f_{\text{depositante}} = 50$
- $b_{\text{depositante}} = 50.000/50 = 100$
- $V(\text{nome_cliente}, \text{depositante}) = 2500$
- depositante = relação externa
- cliente = relação interna
- relações estão ordenadas no atributo de junção *nome_cliente*

Qual o custo usando Merge-Junção?

$b_r + b_s$

Como calcular o custo de um PEC

- ✓ Operação de Junção – exemplo
- ✓ Merge-Junção

depositante |X| cliente

- $n_{\text{cliente}} = 10.000$
- $f_{\text{cliente}} = 25$
- $b_{\text{cliente}} = 10.000/25 = 400$
- $n_{\text{depositante}} = 5.000$
- $f_{\text{depositante}} = 50$
- $b_{\text{depositante}} = 50.000/50 = 100$
- $V(\text{nome_cliente}, \text{depositante}) = 2500$
- depositante = relação externa
- cliente = relação interna
- relações estão ordenadas no atributo de junção *nome_cliente*

Qual o custo usando Merge-Junção?

$$\text{custo} = 400 + 100 = 500$$

Como calcular o custo de um PEC

- ✓ Operação de Junção – exemplo
- ✓ Merge-Junção

depositante |X| cliente

- $n_{\text{cliente}} = 10.000$
- $f_{\text{cliente}} = 25$
- $b_{\text{cliente}} = 10.000/25 = 400$
- $n_{\text{depositante}} = 5.000$
- $f_{\text{depositante}} = 50$
- $b_{\text{depositante}} = 50.000/50 = 100$
- $V(\text{nome_cliente}, \text{depositante}) = 2500$
- depositante = relação externa
- cliente = relação interna
- relações estão ordenadas no atributo de junção *nome_cliente*

E se relações não estão ordenadas?

Como calcular o custo de um PEC

- ✓ Operação de Junção – exemplo
- ✓ Merge-Junção

depositante |X| cliente

- $n_{\text{cliente}} = 10.000$
- $f_{\text{cliente}} = 25$
- $b_{\text{cliente}} = 10.000/25 = 400$
- $n_{\text{depositante}} = 5.000$
- $f_{\text{depositante}} = 50$
- $b_{\text{depositante}} = 50.000/50 = 100$
- $V(\text{nome_cliente}, \text{depositante}) = 2500$
- depositante = relação externa
- cliente = relação interna
- relações estão ordenadas no atributo de junção *nome_cliente*

E se relações não estão ordenadas?

Calcular custo para classificação (já visto)

Como calcular o custo de um PEC

- ✓ Operação de Junção – exemplo
- ✓ Merge-Junção

depositante |X| cliente

- $n_{\text{cliente}} = 10.000$
- $f_{\text{cliente}} = 25$
- $b_{\text{cliente}} = 10.000/25 = 400$
- $n_{\text{depositante}} = 5.000$
- $f_{\text{depositante}} = 50$
- $b_{\text{depositante}} = 50.000/50 = 100$
- $V(\text{nome_cliente}, \text{depositante}) = 2500$
- *depositante* = relação externa
- *cliente* = relação interna
- relações estão ordenadas no atributo de junção *nome_cliente*

E se relações não estão ordenadas?

Considerando tamanho de memória no pior caso = 3 blocos:

- custo para classificar *cliente* = $400 * (2[\log_2(400/3)] + 1) = 6800$ e + 400 transferências para escrever resultado
- custo para classificar *depositante* = $100 * (2[\log_2(100/3)] + 1) = 1300$ e 100 transferências para gravar resultados
- custo total = $7200 + 1400 + 100 + 400 = 9100$ acessos



Como calcular o custo de um PEC

✓ Operação de Junção

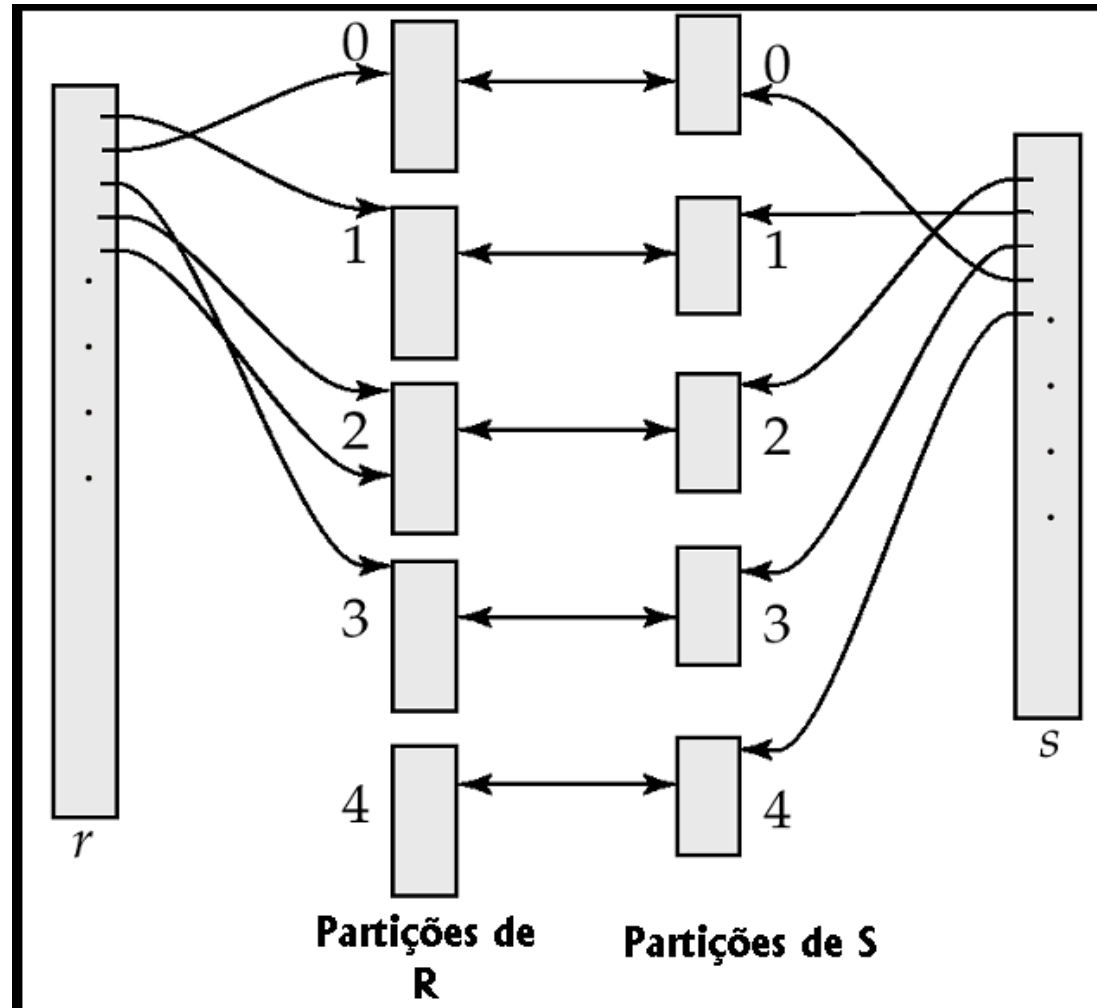
✓ Hash-Junção

- usado para calcular junção natural ou *equi-join*
- $R \cap S$: atributos em comum (atributos de junção)
- função *hash* é usada para particionar as tuplas de ambas as relações em conjuntos que têm o mesmo valor nos atributos de junção
- h : função *hash* que faz o mapeamento de $R \cap S$
- $H_{r_0}, H_{r_1}, \dots, H_{r_{max}}$: partições das tuplas de r
- $H_{s_0}, H_{s_1}, \dots, H_{s_{max}}$: partições das tuplas de s



Como calcular o custo de um PEC

- ✓ Operação de Junção
- ✓ Hash-Junção



Como calcular o custo de um PEC

✓ Operação de Junção

✓ Hash-Junção - ideia do algoritmo:

- se uma tupla de r e uma tupla de s satisfazem condição de junção → têm mesmo valor para atributos de junção
- tupla de r está em H_{r_i} → i é o valor retornado da função *hash*
- tupla de s está em H_{s_j} → i é o valor retornado da função *hash*
- portanto: tuplas de r em H_{r_i} precisam ser comparadas apenas com tuplas de s em H_{s_j}
- após particionar relações, executa-se junção de laço aninhado indexada em cada um dos pares de partição i , para $i=0, \dots, \max$
 - constrói índice *hash* em cada H_{s_j} (relação de construção)
 - testa (procura em H_{s_j}) com as tuplas de H_{r_i} (relação de teste)

* Algoritmo completo: Silberschatz, A.; Korth, H.F.; Sudarshan, S. "Sistema de Banco de Dados", 3a. edição, Makron Books, 1999 (capítulo 12)

Como calcular o custo de um PEC

✓ Operação de Junção

✓ Hash-Junção - custo

- índice *hash* é construído na memória → sem custo
- fases de construção e de teste requerem somente um passo
- se *max* é maior ou igual ao número de frames da memória → não dá para fazer particionamento em em único passo → nesse caso, faz-se ***particionamento recursivo***



Como calcular o custo de um PEC

✓ Operação de Junção

✓ Hash-Junção – custo

–considerando que não há overflow da tabela *hash* e não é necessário particionamento recursivo:

- particionamento de r e s : leitura e escrita de todas as tuplas $2(b_r + b_s)$
- fases de construção e teste: ler uma vez cada uma das partições $\rightarrow b_r + b_s$
- número de blocos ocupados pelas partições pode ser maior que $b_r + b_s$ (blocos parcialmente cheios) \rightarrow overhead de no máximo $2 * \max$
- *custo total* = $3(b_r + b_s) + 2 * \max$



Como calcular o custo de um PEC

✓ Operação de Junção

✓ Hash-Junção – custo

–considerando necessário particionamento recursivo:

- cada passo reduz o tamanho do particionamento por um fator $M - 1$;
- redução repetida até que cada partição tenha no máximo M blocos;
- número de passos para particionar s : $\lceil \log_{M-1}(b_s) \rceil - 1$;
- em cada passo todos os blocos de s são lidos e escritos: $2b_s \lceil \log_{M-1}(b_s) \rceil$;
- número de passos para particionar $r =$ igual anterior;
- *custo total* = $2(b_r + b_s) + \lceil \log_{M-1}(b_s) \rceil - 1 + b_r + b_s$



Como calcular o custo de um PEC

✓ Outras operações

- Eliminação de duplicidade
- Projeção
- Operações de Conjunto
- Junção Externa
- Agregação



Como calcular o custo de um PEC

✓ Eliminação de duplicidade

- Fácil de implementar usando classificação:
 - sort-merge externo: tuplas duplicadas podem ser eliminadas antes do arquivo temporário ser escrito em disco, reduzindo número de acessos;
 - Estimativa de custo para pior caso: ???



Como calcular o custo de um PEC

✓ Eliminação de duplicidade

- Implementação usando classificação:
 - sort-merge externo: tuplas duplicadas podem ser eliminadas antes do arquivo temporário ser escrito em disco, reduzindo número de acessos;
 - Estimativa de custo para pior caso: **mesmo da classificação da relação.**
- Implementação usando hashing:
 - relação particionada com base em função *hash*;
 - cada partição é lida e índice *hash* é construído na memória;
 - enquanto constrói índice *hash*: só insere tupla se ainda não estiver lá
 - após todas tuplas da partição serem processadas → tuplas do índice são escritas no resultado;
 - Estimativa de custo: **igual ao custo do processamento (particionamento e leitura de cada partição) da relação de construção em uma hash-junção.**



Como calcular o custo de um PEC

✓ Projeção

- Projeção em cada tupla, seguida de eliminação de duplicidade (se necessário);
- se atributos projetados incluem uma chave da relação → não existirá duplicatas;
- tamanho da projeção na forma $\pi_A(r)$ é calculado como $V(A,r)$ considerando que projeção elimina duplicatas.
- Estimativa de custo: mesmo da eliminação de duplicidade.



Como calcular o custo de um PEC

✓ Operações de Conjunto (União, Intersecção e Diferença)

- Usando classificação:
 - classificação de ambas as relações;
 - varredura de cada relação (uma vez cada uma) para obter resultado.
 - Estimativa de custo: $b_r + b_s + \text{custo de classificação (se necessário)}$
- Usando hashing:
 - particionar as duas relações com base na mesma função *hash*, criando partições $H_{r_0}, H_{r_1}, \dots, H_{r_{max}}$ e $H_{s_0}, H_{s_1}, \dots, H_{s_{max}}$
 - Executar para cada partição $i, i=0, \dots, max$:
 - construir índice hash na memória em H_{r_i}



Como calcular o custo de um PEC

✓ Operações de Conjunto (União, Intersecção e Diferença)

– Usando hashing:

- particionar as duas relações com base na mesma função *hash*, criando partições $H_{r_0}, H_{r_1}, \dots, H_{r_{max}}$ e $H_{s_0}, H_{s_1}, \dots, H_{s_{max}}$
- Executar para cada partição $i, i=0, \dots, max$:
 - construir índice hash na memória em H_{r_i}

$r \cup s$:

- » adicionar as tuplas de H_{s_i} ao índice hash somente se não estiverem presentes
- » adicionar as tuplas do índice hash ao resultado

$r \cap s$:

- » para cada tupla de H_{s_i} testar o índice hash e adicionar a tupla ao resultado somente se ela estiver presente no índice hash

$r - s$:

- » para cada tupla de H_{s_i} testar o índice hash e, se a tupla estiver presente no índice hash, removê-la do índice hash
- » adicionar as tuplas restantes do índice hash ao resultado



Bibliografia

- ✓ Silberschatz, A.; Korth, H.F.; Sudarshan, S. “Sistema de Banco de Dados”, 3a. edição, Makron Books, 1999 (capítulo 12)
- ✓ Elmasri, R.; Navathe, S. “Sistemas de Banco de Dados”, 4ª edição, Addison-Wesley, São Paulo, 2005 (capítulo 15)

Exercícios

- ✓ Silberschatz, A.; Korth, H.F.; Sudarshan, S. “Sistema de Banco de Dados”, 3a. edição, Makron Books, 1999 – página 436 – exercícios 1 a 18
- ✓ Elmasri, R.; Navathe, S. “Sistemas de Banco de Dados”, 4ª edição, Addison-Wesley, São Paulo, 2005 – página 384 – exercícios 1 a 7, 9 a 11

ACH2025

Laboratório de Bases de Dados

Aula 13

Processamento de Consultas - Parte 1

Professora:

➤ Fátima L. S. Nunes

