

Projeto de Bancos de Dados Distribuídos (Parte 01)

IN1128/IF694 – Bancos de Dados Distribuídos e Móveis
 Ana Carolina Salgado – acs@cin.ufpe.br
 Bernadette Farias Lôscio – bfl@cin.ufpe.br



Cin.ufpe.br

Tipos de fragmentação

- Fragmentação horizontal
- Fragmentação vertical



Grau de fragmentação

- É preciso encontrar um **nível adequado de fragmentação** que mantenha um compromisso entre os dois extremos
 - Nenhuma fragmentação
 - Fragmentação até o nível de tuplas ou de atributos
- Este equilíbrio pode ser **determinado de acordo com as aplicações** que farão uso do banco de dados e com um conjunto de parâmetros que caracterizam essas aplicações



Correção das regras de fragmentação

- **Fragmentação vertical e normalização** são processos semelhantes
- São necessárias regras para assegurar que o BD não sofrerá **nenhuma mudança semântica** durante a fragmentação
 - Completeza
 - Reconstrução
 - Disjunção



Completeza

- Se uma instância da relação R é **decomposta em fragmentos** R_1, R_2, \dots, R_n , cada item de dados que pode ser encontrado em R também pode ser encontrado em um ou mais fragmentos R_i
- Propriedade idêntica à **decomposição sem perdas** da normalização
- Os dados em uma relação global serão mapeados em fragmentos sem perdas
- **Fragmentação horizontal**
 - um item = tupla
- **Fragmentação vertical**
 - um item = atributo



Reconstrução

- Se uma relação R é decomposta em fragmentos R_1, R_2, \dots, R_n , deve ser possível definir um **operador relacional** ∇ tal que:
 - $R = \nabla R_i, \forall R_i \in F_R$
- O operador dependerá do tipo de fragmentação
- Esta regra garante que as restrições definidas sobre os dados na forma de dependências serão preservadas



Disjunção

- Se uma relação R é **decomposta horizontalmente em fragmentos** R_1, R_2, R_n , e o item de dados d_1 está em R_j , ele não está em qualquer outro fragmento $R_k (k \neq j)$
- Os fragmentos horizontais são disjuntos
- No caso da fragmentação vertical, a disjunção só é definida para os atributos que não fazem parte da chave primária
 - Se a relação é decomposta verticalmente, os atributos que compõem a chave primária se repetem em todos os fragmentos



Alternativas de alocação

- Quando os dados são **alocados**, eles podem ser replicados ou mantidos como uma única cópia
- As **razões para a replicação** são a confiabilidade e a eficiência de consultas somente de leitura
- A decisão sobre replicação é um **compromisso** que depende da proporção **entre consultas de leitura e consultas de atualização**



Alocação

- Banco de dados não replicado é chamado de **BD particionado**
 - Fragmentos alocados a sites e só existe uma cópia de cada fragmento na rede
- O BD existe **por inteiro em cada site** (totalmente replicado) ou os **fragmentos estão distribuídos pelos sites** de maneira que cópias de um fragmento podem residir em mais de um site (parcialmente replicado)
- No caso de ser **parcialmente replicado**, o número de cópias de cada fragmento pode ser uma entrada para um algoritmo de alocação



Requisitos de informações

- As informações necessárias para o projeto de distribuição podem ser divididas em quatro categorias:
 - Informações do BD
 - Informações de aplicativos
 - **Informações da rede de comunicações**
 - **Informações de sistemas computacionais**
 - Os dois últimos tipos são de natureza quantitativa, sendo utilizadas apenas para a alocação e não em algoritmos de fragmentação



Fragmentação horizontal

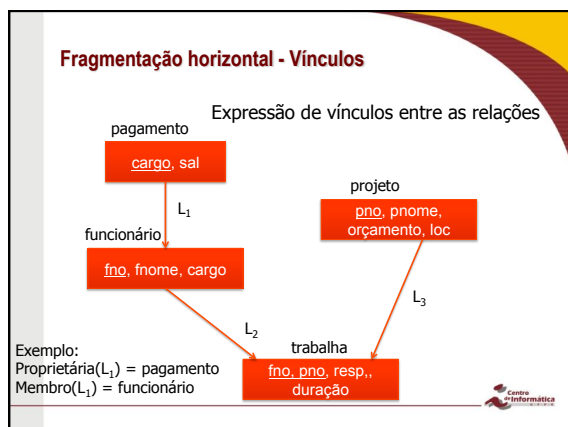
- Versões de fragmentação horizontal: primária e derivada
- **Fragmentação horizontal primária** de uma relação é executada com o uso de predicados definidos sobre esta relação
- **Fragmentação horizontal derivada** é o particionamento de uma relação que resulta da definição de predicados sobre outra relação



Fragmentação horizontal - Informações do BD

- **Esquema conceitual global**
 - Como as relações do BD estão conectadas entre si, em especial as junções
- É preciso identificar a origem e o destino de um vínculo (isto é, **ligação, relacionamento, relação**)
- A informação quantitativa exigida sobre o BD é a cardinalidade de cada relação R, denominada **card (R)**





- ### Fragmentação horizontal - Informações de aplicativos
- Informações qualitativas e quantitativas
 - A **qualitativa** orienta a atividade de fragmentação (ex: predicados usados nas consultas de usuários)
 - 20% das consultas de usuários mais ativos respondem a 80% do total de acessos aos dados
 - Regra 80/20 pode ser usada como diretriz para a execução dessa análise
 - A informação qualitativa é incorporada nos modelos de alocação

- ### Fragmentação horizontal
- Predicados podem ser **simples**, envolvendo apenas operadores relacionais, por exemplo
 - Predicados podem ser mais elaborados, os quais são **combinações booleanas de predicados simples**
 - Um predicado **minterm** é a conjunção de predicados simples
 - Tendo em vista que sempre é possível converter uma expressão booleana para a forma normal conjuntiva, o uso de predicados *minterm* nos algoritmos de projeto não causa qualquer perda de generalidade

Fragmentação horizontal

- Na lógica booleana, uma fórmula está na forma normal conjuntiva (FNC) se é uma conjunção de cláusulas, onde uma cláusula é uma disjunção de literais.

Funcionário			Trabalha			
FNO	FNome	Cargo	FNO	PNO	Respons.	Dur.
F1	João	Engenheiro	F1	P1	Gerente	12
F2	Paulo	Analista	F2	P1	Analista	24
F3	Maria	Professor	F2	P2	Analista	6
F4	Pedro	Programador	F3	P3	Consultor	10
F5	Luiza	Analista	F3	P4	Engenheiro	48
F6	Luis	Engenheiro	F4	P2	Programador	18
F7	Beatriz	Professor	F5	P2	Gerente	24
F8	Paula	Analista	F6	P4	Gerente	48
			F7	P3	Engenheiro	36
			F8	P3	Gerente	40

Projeto				Pagamento	
PNO	PNome	Orçamento	Loc	Cargo	Salário
P01	BI	150000	Montreal	Engenheiro	40000
P02	Des. de BD	135000	Nova York	Analista	34000
P03	CADI/CAM	250000	Nova York	Professor	27000
P04	Manutenção	310000	Paris	Programador	24000

- ### Fragmentação horizontal
- **Predicado simples**
 - p₁: cargo = "Professor"
 - p₂: cargo = "Engenheiro"
 - p₃: cargo = "Programador"
 - p₄: cargo = "Analista"
 - p₅: sal ≤ 30000
 - p₆: sal > 30000

Fragmentação horizontal

- **Alguns predicados minterm**
- m_1 : cargo = "Engenheiro" \wedge sal \leq 30000
- m_2 : cargo = "Engenheiro" \wedge sal $>$ 30000
- m_3 : not (cargo = "Engenheiro") \wedge sal \leq 30000
- m_4 : not (cargo = "Engenheiro") \wedge sal $>$ 30000
- m_5 : cargo = "Programador" \wedge sal \leq 30000
- m_6 : cargo = "Programador" \wedge sal $>$ 30000



Fragmentação horizontal

- **Algumas observações:**
 - Esses não são todos os predicados minterm que podem ser definidos; esta é apenas uma amostra representativa
 - Alguns predicados podem não ter significado, dada a semântica da relação *pagamento*



Fragmentação horizontal

- **Informações quantitativas sobre aplicativos incluem:**
 - **Seletividade de minterm:** a quantidade de tuplas da relação que seria acessada por uma consulta do usuário especificada de acordo com um predicado minterm. Por exemplo, a seletividade de m_1 é 0 porque não existem tuplas que satisfazem esse predicado.



Fragmentação horizontal

- **Informações quantitativas sobre aplicativos incluem:**
 - **Frequência de acesso:** é a frequência com que os aplicativos do usuário acessam os dados. Se $Q = \{q_1, q_2, \dots, q_n\}$ é um conj. de consultas do usuário, $acc(q_i)$ indica a frequência de acesso da consulta q_i em determinado período.
 - As frequências de acesso minterm podem ser determinadas a partir das frequências de consultas. A frequência de acesso de um minterm m_i é denotada por $acc(m_i)$



Fragmentação horizontal primária

- É definida por uma operação de seleção sobre as relações proprietárias de um esquema do banco de dados
- Dada a relação R, seus fragmentos horizontais são determinados por:
- $R_i = \sigma_{F_i}(R)$, $1 \leq i \leq w$, onde w é o número de fragmentos definidos sobre R
- Se F, está em forma normal conjuntiva então é um predicado minterm(m_i)
 - os algoritmos de fragmentação horizontal geralmente consideram F_i como sendo um predicado minterm



Fragmentação horizontal primária

- **Exemplo:**
 - Decomposição da relação Projeto em Projeto1 e Projeto2
 - Projeto1 = $\sigma_{\text{orçamento} \leq 200000}(\text{Projeto})$
 - Projeto2 = $\sigma_{\text{orçamento} > 200000}(\text{Projeto})$



Fragmentação horizontal

- Um fragmento horizontal R_i da relação R consiste de todas as tuplas de R que satisfazem o predicado *minterm* m_i
- Dado um conjunto de predicados *minterm* M , existem tantos fragmentos horizontais da relação R quantos forem os predicados *minterm*
- Conjunto de fragmentos horizontais também é referido como conjunto de *fragmentos minterm*



Fragmentação horizontal

- A definição de fragmentos horizontais depende dos predicados *minterm*
- O primeiro passo para qualquer algoritmo de fragmentação é definir um conjunto de predicados simples que formarão os predicados *minterm*



Fragmentação horizontal primária - algoritmo

- Dados: Uma relação R , o conjunto de predicados simples Pr
- Saída: O conjunto de fragmentos de $R = \{R_1, R_2, \dots, R_w\}$ que obedecem as regras de correção da fragmentação
- Sendo que :
 - Pr deve ser *completo*
 - Pr deve ser *minimo*



Fragmentação horizontal – Completeza de predicados simples

- Um conjunto de predicados simples Pr é considerado *completo* sse existe uma probabilidade igual de acesso de cada aplicativo a qualquer tupla pertencente a qualquer fragmento *minterm* definido de acordo com Pr
- Cada tupla tem igual probabilidade de ser acessada dentro de um fragmento



Fragmentação horizontal – Completeza de predicados simples

- Exemplo:
- Considere que a relação Projeto(PNO, PNome, Orçamento, Loc) tem 2 aplicações definidas sobre ela:
 - (1) Encontre os orçamentos dos projetos de cada localidade
 - (2) Encontre os projetos com orçamento < 200000



Fragmentação horizontal – Completeza de predicados simples

- De acordo com (1),
 - $Pr = \{Loc = "Montreal", Loc = "New York", Loc = "Paris"\}$
- Pr não é completo com respeito a (2), pois dentro de um mesmo fragmento uma tupla terá maior probabilidade de ser acessada do que outra

Projeto

PNO	PNome	Orçamento	Loc
P01	BI	150000	Montreal
P02	Des. de BD	135000	Nova York
P03	CAD/CAM	250000	Nova York
P04	Manutenção	310000	Paris



Fragmentação horizontal – Completeza de predicados simples

- **Modifique**
 - $Pr = \{Loc = "Montreal", Loc = "New York", Loc = "Paris", Orçamento \leq 200000, Orçamento > 200000\}$
- **Agora Pr é completo, pois todas as tuplas terão a mesma probabilidade de serem acessadas em cada um dos fragmentos *minterm***



Fragmentação horizontal – Completeza de predicados simples

- **Os fragmentos obtidos de acordo com um conjunto completo de predicados são logicamente uniformes**
- **São estatisticamente homogêneos na forma como os aplicativos têm acesso a eles**
- **Um conjunto completo de predicados será usado como base da fragmentação horizontal primária**



Fragmentação horizontal – Minimalidade de predicados simples

- **Se um predicado influencia o modo como a fragmentação é efetuada (isto é, faz com que um fragmento f seja ainda mais fragmentado em, digamos, f_i e f_j) deve haver pelo menos um aplicativo que acesse f_i e f_j de maneira diferenciada**
- **Cada predicado simples deve ser “relevante” na determinação de uma fragmentação**
- **Se todos os predicados de um conjunto Pr são relevantes, Pr é mínimo**



Fragmentação horizontal – Minimalidade de predicados simples

- **O conjunto $Pr = \{Loc = "Montreal", Loc = "New York", Loc = "Paris", Orçamento \leq 200000, Orçamento > 200000\}$ é mínimo**
- **Se adicionássemos $P_{nome} = "BI"$ ao conjunto Pr ele deixaria de ser mínimo, porque o novo predicado não é relevante com relação a Pr**
- **Este novo predicado não afetaria os fragmentos que já foram criados**



Fragmentação horizontal

- **Um algoritmo para determinar um conjunto completo e mínimo de predicados Pr' dado um conjunto de predicados simples Pr deve adotar a seguinte regra:**
 - Regra 1: a regra fundamental de completeza e minimalidade, que estabelece que uma relação ou fragmento é particionado **“em pelo menos duas partes que são acessadas de maneira diferente por pelo menos um aplicativo”**
- **No algoritmo de fragmentação horizontal os fragmentos são determinados de acordo com um predicado *minterm* definido sobre os predicados de Pr**



Fragmentação horizontal – Exemplo 1

- **Fragmentação da relação Pagamento**
 - Aplicativo: Verifica as informações sobre salário e determina um aumento adequado
 - Registros de funcionários são administrados em 2 nós (aplicativo executado em 2 nós)
 - Predicados simples
 - $p1$: Salário \leq 30000
 - $p2$: Salário $>$ 30000
 - $Pr = \{p1\}$, que é completo e mínimo ($p2$ não particionaria $f1$, que é o fragmento *minterm* formado com relação a $p1$)
 - Predicados *minterm*
 - $m1$: (Salário \leq 30000)
 - $m2$: NOT(Salário \leq 30000) = (Salário $>$ 30000)



Fragmentação horizontal – Exemplo 1

- Fragmentação horizontal da relação Pagamento

Pagamento	
Cargo	Salário
Engenheiro	40000
Analista	34000
Professor	27000
Programador	24000

Salário ≤ 30000 Salário > 30000

Pagamento1		Pagamento2	
Cargo	Salário	Cargo	Salário
Professor	27000	Engenheiro	40000
Programador	24000	Analista	34000

Fragmentação horizontal – Exemplo 2

- Fragmentação da relação Projeto
 - Suponha que existam dois aplicativos
 - O primeiro aplicativo é executado em 3 sites e procura nomes e orçamentos de projetos dadas suas localizações
 - Select Pnome, Orçamento
 - From projeto
 - Where Loc = valor
 - Predicados simples (aplicativo 1)
 - p1 : Loc = "Montreal"
 - p2 : Loc = "New York"
 - p3 : Loc = "Paris"

Fragmentação horizontal – Exemplo 2

- Fragmentação da relação Projeto (cont.)
 - O segundo aplicativo é emitido em dois sites e está relacionado com a administração de projetos.
 - Os sites administram projetos de acordo com o orçamento
 - Um nó administra projetos com orçamento ≤ 200000 e outro administra projetos com orçamento > 200000
 - Predicados simples (Aplicativo 2)
 - p4 : Orçamento ≤ 200000
 - p5 : Orçamento > 200000

Fragmentação horizontal – Exemplo 2

- Fragmentação da relação Projeto (cont.)
 - O conjunto Pr = {p1, p2, p3, p4, p5} é completo e mínimo
 - Com base em Pr, os seis predicados *minterm* a seguir podem ser definidos (predicados *minterm* obtidos após a eliminação dos contraditórios)
 - m1 : (Loc = "Montreal") e (Orçamento ≤ 200000)
 - m2 : (Loc = "Montreal") e (Orçamento > 200000)
 - m3 : (Loc = "New York") e (Orçamento ≤ 200000)
 - m4 : (Loc = "New York") e (Orçamento > 200000)
 - m5 : (Loc = "Paris") e (Orçamento ≤ 200000)
 - m6 : (Loc = "Paris") e (Orçamento > 200000)

Fragmentação horizontal - Exemplo 2

Projeto			
PNO	PNome	Orçamento	Loc
P01	BI	150000	Montreal
P02	Des. de BD	135000	Nova York
P03	CAD/CAM	250000	Nova York
P04	Manutenção	310000	Paris

Projeto1 Projeto3

PNO	PNome	Orçamento	Loc
P01	BI	150000	Montreal

PNO	PNome	Orçamento	Loc
P02	Des. de BD	135000	Nova York

Projeto4 Projeto6

PNO	PNome	Orçamento	Loc
P03	CAD/CAM	250000	Nova York

PNO	PNome	Orçamento	Loc
P04	Manutenção	310000	Paris

Fragmentação horizontal derivada

- Uma fragmentação horizontal derivada é definida sobre uma relação membro de um vínculo de acordo com uma operação de seleção especificada sobre sua proprietária
- Exemplo:
 - Diagrama de dependência:
 - funcionário (fno, fnome, cargo) -- L1 --> pagamento (cargo, sal)
 - Proprietária(L1) = pagamento
 - Membro(L1) = funcionário
 - Na fragmentação horizontal derivada, um fragmento de funcionário seria definido de acordo com uma seleção em pagamento

Fragmentação horizontal derivada

- Em uma fragmentação horizontal derivada, os fragmentos são definidos considerando operações de semijunção
 - Deseja-se fazer o particionamento de uma relação membro de acordo com uma relação proprietária, porém o resultado da fragmentação deve ser definido somente sobre os atributos da relação membro
 - Um semijunção da relação R definida sobre o conjunto de atributos A pela relação S definida sobre o conjunto de atributos B é o subconjunto das tuplas de R que participam da junção de R com S



Fragmentação horizontal derivada

- Dada uma ligação L onde $proprietária(L)=S$ e $membro(L)=R$, Os fragmentos horizontais derivados de R são definidos como:
 - $R_i = R \text{ semijoin } S_i, 1 \leq i \leq w$, onde w é o número de fragmentos definidos sobre R e
 - $S_i = \sigma_{F_i}(S)$, onde F_i é a fórmula segundo a qual o fragmento horizontal primário S_i é definido



Fragmentação horizontal derivada - exemplo

- Dada a ligação L_1 onde:
 - Proprietária(L_1) = Pagamento
 - Membro(L_1) = Funcionário
- A relação funcionário pode ser particionada da seguinte forma:
 - Funcionário1 = Funcionário *semijoin* Pagamento1
 - Funcionário2 = Funcionário *semijoin* Pagamento2
 - onde:
 - Pagamento 1 = $\sigma_{salário \leq 30000}$ (Pagamento)
 - Pagamento 2 = $\sigma_{salário > 30000}$ (Pagamento)



Fragmentação horizontal

- É importante notar que é possível ter mais de uma fragmentação candidata para uma relação
- A escolha final do esquema de fragmentação pode ser um problema de decisão solucionado durante a alocação

