

# Proc. Consulta e Algebra Relacional: Uma introdução gentil

João Marcelo Borovina Josko

`marcelo.josko@ufabc.edu.br`

June 29, 2018

# Tópicos

- 1 Álgebra Relacional
  - Fundamentos
  - Operadores Clássicos
  - Operadores Estendidos
- 2 Multiset ou BAG
- 3 Processamento de Consulta
  - Catálogo
  - Parsing
  - Otimização
  - Plano de Acesso

# Tópicos

## 1 Álgebra Relacional

- Fundamentos
- Operadores Clássicos
- Operadores Estendidos

## 2 Multiset ou BAG

## 3 Processamento de Consulta

- Catálogo
- Parsing
- Otimização
- Plano de Acesso

# Álgebra Relacional

- Linguagem procedural restrita em operações sobre um SGBD Relacional [Date 2003]
- Origem da perspectiva [Koppelaars *et al.* 2007]
  - Restrição facilita a otimização em operações escritas em linguagens de alto nível (SQL)
- Cada operação algébrica atua sobre uma ou mais relações e leva a uma nova relação [Date 2003]
- Grupo de operações denotam um Expressão Algébrica relacional

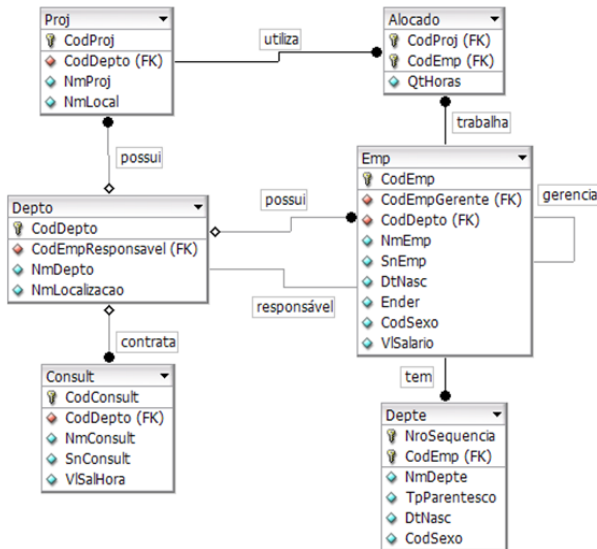
# Álgebra Relacional (Cont.)

- A ideia central de um SGBD Relacional é otimização algébrica de operações
- Exemplo no mundo dos inteiros ( $\mathbb{Z}$ )
  - A expressão  $((a * 3) + ((a * 5) + 0))/1$  é otimizada para  $(3 + 5) * a$
- Três possibilidades. Qual a melhor?
  - $\sigma_{p=a}(R) \bowtie_{o=s} (\sigma_{p=b}(R) \bowtie_{o=s} \sigma_{p=b}(R))$
  - $(\sigma_{p=a}(R) \bowtie_{o=s} \sigma_{p=b}(R)) \bowtie_{o=s} \sigma_{p=b}(R)$
  - $(\sigma_{p=a \wedge o=s \wedge p=b}(R) \bowtie_{o=s} (R \times R \times R))$

# Lista de Operadores

- Operadores Clássicos [Elmasri 2008]
  - Projeção, Seleção e Renomeação
  - Junção (Natural e Theta)
  - Divisão, União, Intersecção, Diferença
  - Produto Cartesiano

# Banco de Dados de Referência



# Projeção ( $\pi$ )

- Extrai um subconjunto de atributos de uma relação qualquer [Elmasri 2008, Date 2003]
- Sintaxe:
  - $\pi_{a_1, a_2, \dots, a_n}(R)$ , onde  $a_i, i = [1, n], a_i \in R$

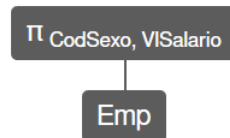


# Projeção (Cont.)

$\pi_{\text{CodSexo, VISalario}}(\text{Emp})$

<b>Emp.CodSexo</b>	<b>Emp.VISalario</b>
M	21000
F	4000
F	7500
M	8000
F	2600
F	4500
M	5000

```
select CodSexo, VISalario  
from Emp
```



Fonte: Elaborado por BOROVINA JOSKO, J.M.

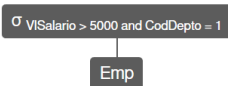
# Seleção (Sigma)

- Extrai um subconjunto de tuplas de uma relação  $R$
- Tuplas extraídas satisfazem um conjunto de predicados  $P$  envolvendo atributos da relação  $R$ 
  - $P = \{p_1, p_2, \dots, p_k\}$ , onde  $p_j, j = [1, k]$ ,  
 $p_j = \langle \text{atributo} \rangle \langle \text{operador} \rangle \langle \text{atributo ou constante} \rangle$
  - Predicados são conectados por operadores lógicos  $\Phi$ , formando  
 $p_1 \Phi p_2 \Phi \dots p_{k-1} \Phi p_k$
- Sintaxe:
  - $\sigma_P(R)$

# Seleção (Cont.)

$\sigma_{\text{VISalario} > 5000 \text{ and } \text{CodDepto} = 1}(\text{Emp})$

```
select * from Emp
where VISalario > 5000
and CodDepto = 1
```



Emp.CodEmp	Emp.CodEmpGerente	Emp.CodDepto	Emp.NmEmp	Emp.SnEmp	Emp.DtNasc	Emp.Ender	Emp.CodSexo	Emp.VISalario
13	55	1	Gilberto	Vurg	22-JAN-1960	R. Prof. Faustino, 111	M	8000

```
select * from Consult
where VISalHora is not null
```

$\sigma_{\text{VISalHora} \neq \text{null}}(\text{Consult})$



Consult.CodConsult	Consult.CodDepto	Consult.NmConsult	Consult.SnConsult	Consult.VISalHora
1	2	Carlos	Figueroa	32
3	2	Cristina	Fernandes	28
4	3	Humberto	Osvaldo	31
6	3	Paula	Gilt	41
8	2	Francisca	da Silva	33

Fonte: Elaborado por BOROVINA JOSKO, J.M.

# Expressão em Álgebra Relacional

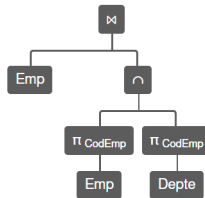
- Operadores algébricos  $O_1, O_2, \dots, O_n$  podem ser combinados em um expressão  $E$
- O resultado da operação interna é utilizado pela operação externa subsequente
- Sintaxe em *pipeline*:
  - $O_1(O_2(O_n))$
- Sintaxe procedural:
  - $T \leftarrow O_n$
  - $G \leftarrow O_2(T)$
  - $O_1(G)$

# Expressão em Álgebra Relacional

```

6 |-- X =  $\pi$  CodEmp (Emp)
7 -- Y =  $\pi$  CodEmp (Depte)
8 -- Z = X  $\cap$  Y
9 -- Emp join Z
10
11
12 Emp join (( $\pi$  CodEmp (Emp))  $\cap$  ( $\pi$  CodEmp (Depte)))

```



$$\text{Emp} \bowtie ((\pi_{\text{CodEmp}}(\text{Emp})) \cap (\pi_{\text{CodEmp}}(\text{Depte})))$$

I

Emp.CodEmp	Emp.CodEmpGerente	Emp.CodDepto	Emp.NmEmp	Emp.SnEmp	Emp.DtNasc	Emp.Ender	Emp.CodSexo	Emp.VISalario
29	55	3	Tereza	Kohn	22-JUN-1963	R. Prof. Mario Sá, 198	F	7500
13	55	1	Gilberto	Vurg	22-JAN-1960	R. Prof. Faustino, 111	M	8000
1	13	1	Jose	Santos	06-OUT-1961	R. Prof. Sebastiao, 15	M	5000

# Renomear ( $\rho$ )

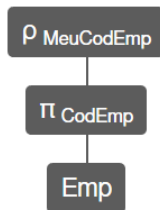
- Renomeia (atribui um alias) a atributos de uma relação  $R$
- Sintaxes:
  - $\rho \text{ Alias } R(R)$
  - $\rho \text{ Alias } a_1, a_2, \dots, a_n(R)$
  - $\rho \text{ Alias } R(a_1, a_2, \dots, a_n)(R)$

# Renomear

$\rho_{\text{MeuCodEmp}} (\pi_{\text{CodEmp}} (\text{Emp}))$

select CodEmp as MeuCodEmp  
from Emp

MeuCodEmp.CodEmp
55
39
29
13
17
16
2
1



Fonte: Elaborado por BOROVINA JOSKO, J.M.

# Produto Cartesiano

- Realiza a operação de combinação das tuplas das relações  $R_1$  e  $R_2$  e gera uma nova relação  $T$
- Seja  $\langle Relação \rangle$  o **número de atributos de uma relação**, então  $\langle T \rangle = \langle R_1 \rangle + \langle R_2 \rangle$
- Seja  $|Relação|$  o **número de tuplas de um relação**, então  $|T| = |R_1| * |R_2|$
- Sintaxe:
  - $R_1 \times R_2$

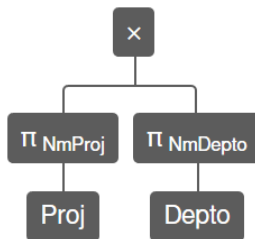


# Produto Cartesiano (Cont.)

$$\pi_{NmProj}(Proj) \times \pi_{NmDepto}(Depto)$$

```
select NmProj, NmDepto
from Proj, Depto
```

Proj.NmProj	Depto.NmDepto
reengenharia	pesquisa
reengenharia	administracao
reengenharia	financeiro
produto X	pesquisa
produto X	administracao
produto X	financeiro
produto Y	pesquisa
produto Y	administracao
produto Y	financeiro
automatizacao	pesquisa
automatizacao	administracao
automatizacao	financeiro



Fonte: Elaborado por BOROVINA JOSKO, J.M.

# Junção Natural

- Extrai tuplas das relações  $R$  e  $T$  que possuem em comum um ou mais atributos
- Combinação de *Produto Cartesiano* e *Seleção*
- Pode apresentar um conjunto de predicados (similar a seleção)
- Seja  $R(A)$  e  $T(A)$  os atributos das relações  $R$  e  $T$ , então  $R(A) \cap T(A) \neq \emptyset$
- Sintaxe:
  - $R \bowtie T$
  - $R \bowtie_P T$

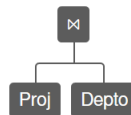
# Junção Natural (Cont.)

$(Proj) \bowtie (Depto)$

select \* from  
Proj Natural Join Depto

ou

select \* from  
Proj Join Depto using (CodDepto)



Proj.CodProj	Proj.CodDepto	Proj.NmProj	Proj.NmLocal	Depto.CodEmpResponsavel	Depto.NmDepto	Depto.NmLocalizacao
1	2	reengenharia	SP	null	administracao	campinas
2	1	produto X	SP	null	pesquisa	sao paulo
3	2	reengenharia	MG	null	administracao	campinas
4	1	produto Y	SP	null	pesquisa	sao paulo
5	3	automatizacao	SP	null	financeiro	campinas

Fonte: Elaborado por BOROVIINA JOSKO, J.M.

# Junção Theta

- Estende a junção natural ao permitir indicar predicados sobre atributos das relações envolvidas
- Sintaxe:
  - $R \bowtie_P T$

# Junção Theta (Cont.)

Alocado  $\bowtie$  ( QtHoras < 100 and Alocado.CodProj = Proj.CodProj ) Proj

$\bowtie$  ( QtHoras < 100 and Alocado.CodProj = Proj.CodProj )

```
select * from Alocado Join Proj using (CodProj)
where QtHoras < 100
```



Alocado.CodProj	Alocado.CodEmp	Alocado.QtHoras	Proj.CodProj	Proj.CodDepto	Proj.NmProj	Proj.NmLocal
2	13	80	2	1	produto X	SP
4	13	60	4	1	produto Y	SP
4	1	50	4	1	produto Y	SP
1	55	40	1	2	reengenharia	SP
1	13	25	1	2	reengenharia	SP
3	55	30	3	2	reengenharia	MG
4	29	15	4	1	produto Y	SP
5	29	70	5	3	automatizacao	SP
3	17	40	3	2	reengenharia	MG

Fonte: Elaborado por BOROVINA JOSKO, J.M.

# Lista de Operadores

- Operadores Relacionais Estendidos [Elmasri 2008]
  - Ordenação
  - Projeção Estendida
  - Agrupamento
  - Funções de Agregação
  - Junção Externa

# Ordenação (Tau)

- Realiza a ordenação por ordem ascendente ou descendente de atributos de um relação  $R$
- Sintaxe:
  - $\mathcal{T}_{atributo_1[asc,desc], \dots, atributo_K[asc,desc]}R$

# Ordenação (Cont.)

$\pi_{\text{Emp.NmEmp}, \text{Emp.VISalario}} (\tau_{\text{Emp.VISalario desc}, \text{Emp.NmEmp asc}} (\text{Emp}))$

<b>Emp.NmEmp</b>	<b>Emp.VISalario</b>
PAulo	21000
Gilberto	8000
Tereza	7500
Jose	5000
Fabiana	4500
Mariana	4500
Ana	4000
Daniela	2600

```
select NmEmp, VISalario
from Emp
Order by VISalario desc, NmEmp Asc
```



# Projeção Estendida

- Permite aplicar expressões aritméticas sobre atributos de um relação  $R$
- Sintaxe:
  - $\pi_{\text{atributo}_1 \text{ operação atributo}_K} R$
- Exemplo:
  - $\tau_{nmemp, vlsalario * 12} Emp$
  - SQL: *Select nmemp, vlsalario \* 12 from Emp*

# Agrupamento (Gamma)

- Permite agrupar tuplas da relação  $R$  de acordo com uma lista de atributos
- Sintaxe:
  - $\gamma_{\text{atributo}_1, \text{atributo}_k; \text{funçãoAgregação}} \text{agregação} \rightarrow \text{alias} \quad R$

# Agrupamento (Cont.)

- Funções de agregação não são operadores da Álgebra relacional
- Consomem um conjunto de tuplas de uma relação para produzir um valor único
- Principais funções:
  - Count
  - Sum
  - Max
  - Min
  - Avg

# Agrupamento (Cont.)

Y CodDepto, CodSexo; COUNT(CodSexo)→Total, SUM(VISalario)→Soma (Emp)

Emp.CodDepto	Emp.CodSexo	Total	Soma
2	M	1	21000
2	F	1	4000
3	F	3	14600
1	M	2	13000
1	F	1	4500

```
select CodDepto, CodSexo, count(Codsexo) Total, sum(VISalario) Soma
from Emp
Group by CodDepto, CodSexo
```

Y CodDepto, CodSexo; COUNT(CodSexo)→Total, SUM(VISalario)→Soma

Emp

# Junção Externa

- Extrai tuplas da relação  $R$  e  $T$  que podem ou não estar relacionadas
- Possibilidades:
  - Junção externa a esquerda  $\rightarrow$  Tuplas de  $R$  que não estão em  $T$
  - Junção externa a direita  $\rightarrow$  Tuplas de  $T$  que não estão em  $R$
  - Junção externa cheia  $\rightarrow$  Tuplas de  $R$  que não estão em  $T$  e vice-versa
- As respectivas syntaxes são:
  - $R \bowtie_P T$
  - $R \bowtie_{\sqsubset_P} T$
  - $R \bowtie_{\sqsupset_P} T$

# Exercícios

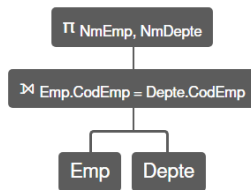
- Selecionar a quantidade de projetos por nome de departamento. Ordene o resultado por nome do departamento em ordem decendente
- Selecionar o nome do consultor e o local do departamento que está associado. Deseja-se todos os locais, independente de terem consultores associados

# Agrupamento (Cont.)

$\pi_{NmEmp, NmDepte} (Emp \bowtie_{Emp.CodEmp = Depte.CodEmp} Depte)$

Emp.NmEmp	Depte.NmDepte
PAulo	<i>null</i>
Ana	<i>null</i>
Tereza	Gabriela Kohn
Tereza	Ricardo Kohn
Gilberto	Carla Vurg
Daniela	<i>null</i>
Fabiana	<i>null</i>
Mariana	<i>null</i>
Jose	Ramiro Santos
Jose	Samuel Santos

select NmEmp, NmDepte  
from Emp RIGHT OUTER JOIN Depte  
on Emp.CodDepte = Depte.CodDepte



# Tópicos

- 1 Álgebra Relacional
  - Fundamentos
  - Operadores Clássicos
  - Operadores Estendidos
- 2 Multiset ou BAG
- 3 Processamento de Consulta
  - Catálogo
  - Parsing
  - Otimização
  - Plano de Acesso



# Um pouco de história

- Da proposta inicial do Modelo Relacional.... [Elmasri 2008]
  - Baseada na álgebra de Conjuntos de Tuplas (**Set**)
  - Restrito às operações de União, Produto Cartesiano, Junção
- ...até a sua implementação em produtos SGBD
  - Baseado no conceito de **Multiset** ou **Bag**
  - Grupo estendido de operações

# Implementação

- SGBDR comerciais raramente utilizam o conceito de **Set**
  - Não permite duplicações de tuplas
- O conceito utilizado é o de **Multiset** ou **Bags**
  - Não é necessário checar por duplicados a cada operação
  - Não desconsidera valores iguais durante operações (e.g., agrupamento)
- Implicações...
  - Aumenta a velocidade dos operadores
  - Todos operadores algébricos em **Set** funcionam de modo distinto em **Multiset**
  - Operador algébrico de conversão de **Multiset** para **Set**

# Valores Distintos (Delta)

- Permite eliminar tuplas de uma operação sobre a relação  $R$  com conteúdo idêntico para uma lista de atributos
- Sintaxe:
  - $\Delta_{atributo_1, atributo_k} R$

# Valores Distintos (Cont.)



Operação em SET

2 `select nmlocalizacao from depto`

	nmlocalizacao character varying (45)
1	sao paulo
2	campinas
3	campinas

Operação em Multiset

2 `select distinct nmlocalizacao from depto`

	nmlocalizacao character varying (45)
1	sao paulo
2	campinas

Operação conversão Multiset para SET

Figure: Distinção entre Set e Multiset

# Tópicos

- 1 Álgebra Relacional
  - Fundamentos
  - Operadores Clássicos
  - Operadores Estendidos
- 2 Multiset ou BAG
- 3 Processamento de Consulta
  - Catálogo
  - Parsing
  - Otimização
  - Plano de Acesso

# Do you remember me?

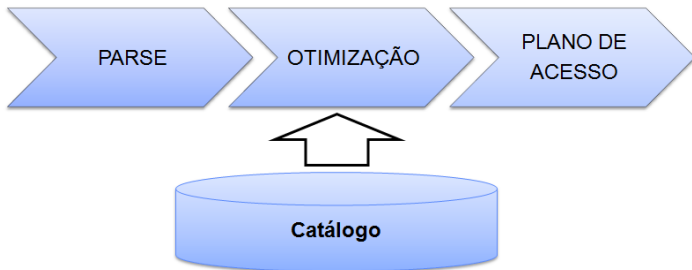


Figure: Visão Resumida dos passos do Processamento de Consultas

# Vamos pensar ... (Cont.)

```
5 select nmlocalizacao, count(*) from depto group by nmlocalizacao
```

Data Output Explain Messages Notifications Query History

	nmlocalizacao character varying (45)	count bigint
1	sao paulo	1
2	campinas	2

Y NmLocalizacao; COUNT(\*)→Total

Depto

Y NmLocalizacao; COUNT(\*)→Total (Depto)

Depto.NmLocalizacao	Total
sao paulo	1
campinas	2

Figure: What!!!!!! São iguais?! Why?

# Vamos pensar ... (Cont.)

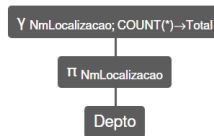
```

5 select a.nmlocalizacao, count(*)
6 from (select nmlocalizacao from depto) A
7 group by a.nmlocalizacao

```

Data Output Explain Messages Notifications Query History

	nmlocalizacao character varying (45)	count bigint
1	sao paulo	1
2	campinas	2



Y NmLocalizacao; COUNT(\*)→Total ( π NmLocalizacao (Depto))

Depto.NmLocalizacao	Total
sao paulo	1
campinas	1

Figure: Agora são diferentes? Que Louco!



# O que é?

- Representa um conjunto de **relações do SGBD**
- Descreve uma série de características físicas e de conteúdo de **relações do usuário**
  - **Esquema** (Relação, seus atributos e tipos, regras de integridade e unicidade, etc.)
  - **Estatísticas** (Volume e tamanho de relações e índices)
  - **Autorização** (Logins, senhas e permissão de acesso e operações)
- Recurso elegante
  - Relações do catálogo são tratadas do mesmo modo que relações do usuário

# Catálogo no PostgreSQL

Catalog Name	Purpose
<a href="#"><code>pg_aggregate</code></a>	aggregate functions
<a href="#"><code>pg_am</code></a>	index access methods
<a href="#"><code>pg_amop</code></a>	access method operators
<a href="#"><code>pg_amproc</code></a>	access method support procedures
<a href="#"><code>pg_attrdef</code></a>	column default values
<a href="#"><code>pg_attribute</code></a>	table columns ("attributes")
<a href="#"><code>pg_authid</code></a>	authorization identifiers (roles)
<a href="#"><code>pg_auth_members</code></a>	authorization identifier membership relationships
<a href="#"><code>pg_cast</code></a>	casts (data type conversions)
<a href="#"><code>pg_class</code></a>	tables, indexes, sequences, views ("relations")
<a href="#"><code>pg_collation</code></a>	collations (locale information)
<a href="#"><code>pg_constraint</code></a>	check constraints, unique constraints, primary key constraints, foreign key constraints
<a href="#"><code>pg_conversion</code></a>	encoding conversion information
<a href="#"><code>pg_database</code></a>	databases within this database cluster
<a href="#"><code>pg_db_role_setting</code></a>	per-role and per-database settings
<a href="#"><code>pg_default_acl</code></a>	default privileges for object types
<a href="#"><code>pg_depend</code></a>	dependencies between database objects
<a href="#"><code>pg_description</code></a>	descriptions or comments on database objects
<a href="#"><code>pg_enum</code></a>	enum label and value definitions
<a href="#"><code>pg_event_trigger</code></a>	event triggers
<a href="#"><code>pg_extension</code></a>	installed extensions
<a href="#"><code>pg_foreign_data_wrapper</code></a>	foreign-data wrapper definitions
<a href="#"><code>pg_foreign_server</code></a>	foreign server definitions
<a href="#"><code>pg_foreign_table</code></a>	additional foreign table information
<a href="#"><code>pg_index</code></a>	additional index information
<a href="#"><code>pg_inherits</code></a>	table inheritance hierarchy

Figure: Trecho das relações que compõem o catálogo do PostgreSQL

# Catálogo no PostgreSQL (Cont.)

```
select relname, relpages, reltuples from pg_class where relname not like 'pg%'
```

relname name	relpages integer	reltuples real
pg_statistic	15	384
pg_type	9	359
pg_toast_2604	0	0
pg_toast_2604_index	1	0
pg_toast_2606	0	0
pg_toast_2606_index	1	0
pg_toast_2609	0	0
pg_toast_2609_index	1	0
pg_toast_1255	0	0
pg_toast_1255_index	1	0
pg_toast_2618	47	205
pg_toast_2618_index	2	205
pg_toast_3596	0	0
pg_toast_3596_index	1	0
pg_toast_2619	3	13
pg_toast_2619_index	2	13
pg_toast_2620	0	0

Figure: Objetos gerenciados pela relação de catálogo *pg\_class*

# Funcionamento Geral

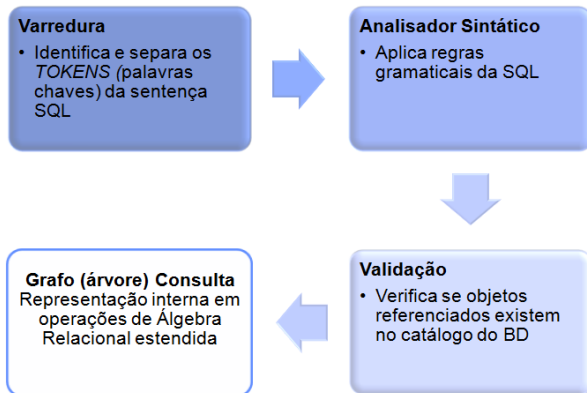


Figure: Principais passos da etapa de *parsing*

# Funcionamento Geral

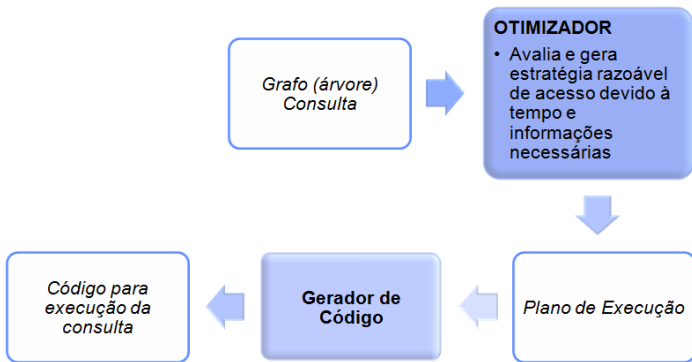


Figure: Principais passos na etapa de *otimização*

# Query Tree

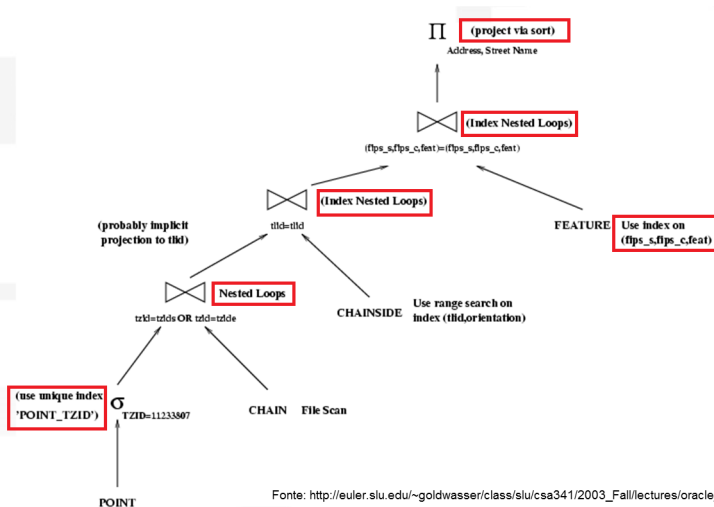


Figure: Escolhas do otimizador - algoritmos e índices

# Qual é a base de escolhas do Otimizador

- Existem diversos algoritmos para um mesmo operador algébrico
- Escolhas dependem das estatísticas correntes do BD → determinar **custo**
- Principais técnicas...
  - Índices seletivos
  - Iterações (e.g., sort-merge para operações de junção)
  - Particionamento dos dados

# Referências I

**Date (2003)** C. J. Date. *An introduction to database systems*. Pearson, 8th. ed.

**Elmasri (2008)** Ramez Elmasri. *Fundamentals of database systems*. Pearson Education India.

**Koppelaars et al. (2007)** Toon Koppelaars et al. *Applied mathematics for database professionals*. Apress.