

Análise de Dados com Base em Processamento Massivo em Paralelo

Índices

Guilherme Muzzi da Rocha



Agenda

- Motivação
- Tipos de índices
 - Árvores
 - Hash
 - Bitmap
- Índice para DW
- Índice para similaridade
- Conclusão

Agenda

- Motivação
- Tipos de índices
 - Árvores
 - Hash
 - Bitmap
- Índice para DW
- Índice para similaridade
- Conclusão

Motivação

- Índice
 - Velocidade de busca

Conteúdo

1 Business Intelligence	3
2 Data Warehousing	4
3 Diferenças entre os Ambientes Operacional e Informacional	6
3.1 Aspectos Gerais	7
3.2 Aspectos Relacionados aos Usuários	8
3.3 Aspectos Relacionados às Operações	9
3.4 Aspectos Relacionados aos Dados	10
3.5 Exemplos de Aplicação	11
4 Conclusão	11

Agenda

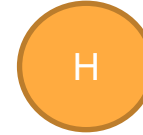
- Motivação
- Tipos de índices
 - Árvores
 - Hash
 - Bitmap
- Índice para DW
- Índice para similaridade
- Conclusão

Tipos de índices - Árvores

- Estrutura específica
 - Esquerda
 - Elementos menores
 - Direita
 - Elementos maiores

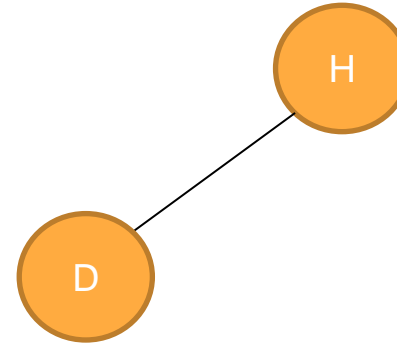
Tipos de índices - Árvores

- Estrutura específica
 - Esquerda
 - Elementos menores
 - Direita
 - Elementos maiores
- Inserção
 - H



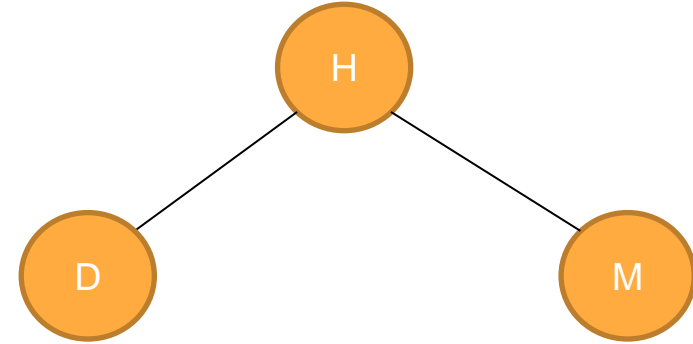
Tipos de índices - Árvores

- Estrutura específica
 - Esquerda
 - Elementos menores
 - Direita
 - Elementos maiores
- Inserção
 - H – D



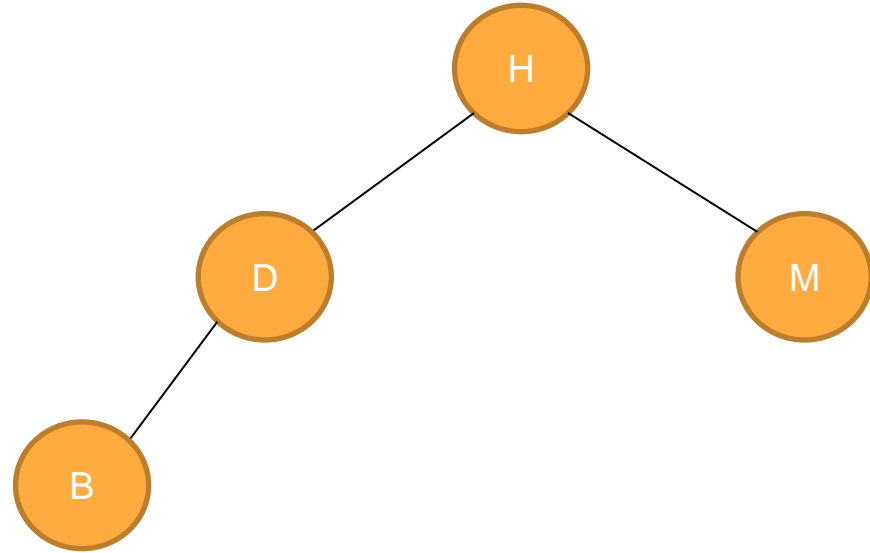
Tipos de índices - Árvores

- Estrutura específica
 - Esquerda
 - Elementos menores
 - Direita
 - Elementos maiores
- Inserção
 - H – D – M



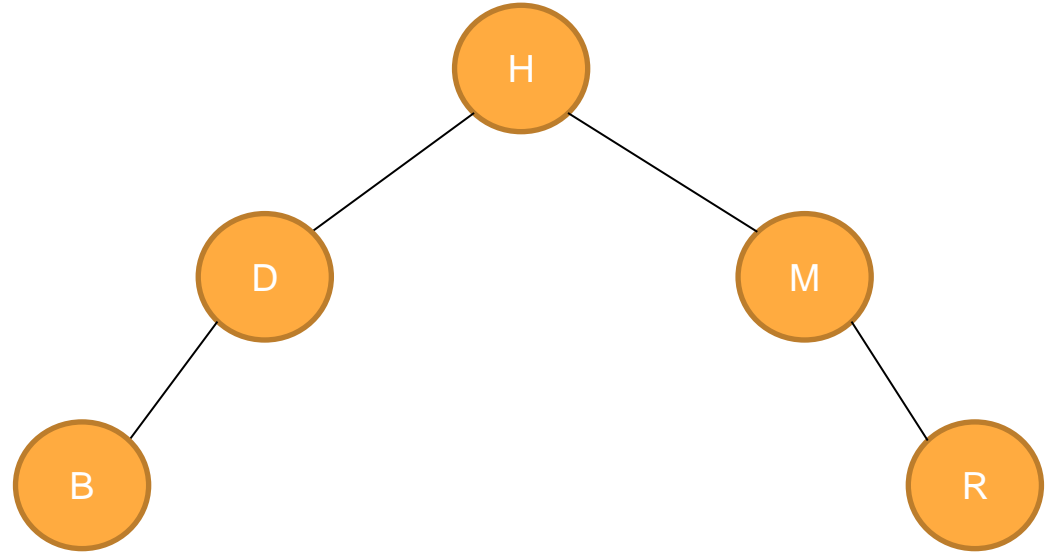
Tipos de índices - Árvores

- Estrutura específica
 - Esquerda
 - Elementos menores
 - Direita
 - Elementos maiores
- Inserção
 - H – D – M – B



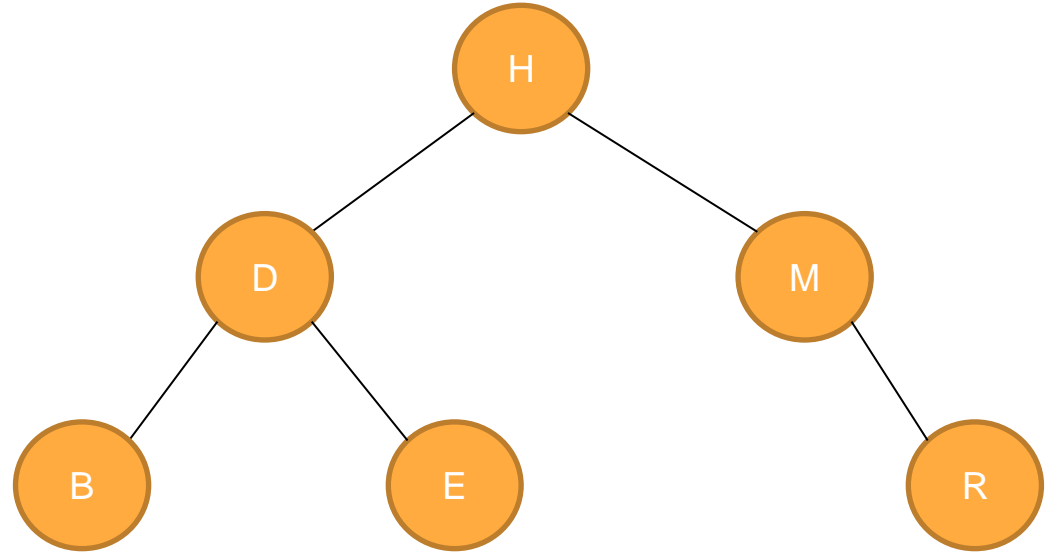
Tipos de índices - Árvores

- Estrutura específica
 - Esquerda
 - Elementos menores
 - Direita
 - Elementos maiores
- Inserção
 - H – D – M – B – R



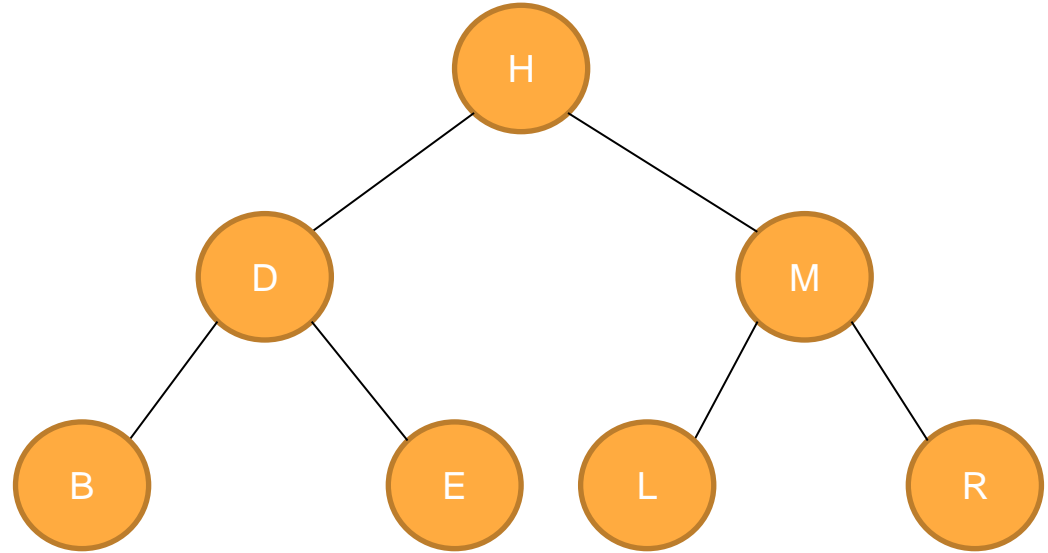
Tipos de índices - Árvores

- Estrutura específica
 - Esquerda
 - Elementos menores
 - Direita
 - Elementos maiores
- Inserção
 - H - D - M - B - R - E



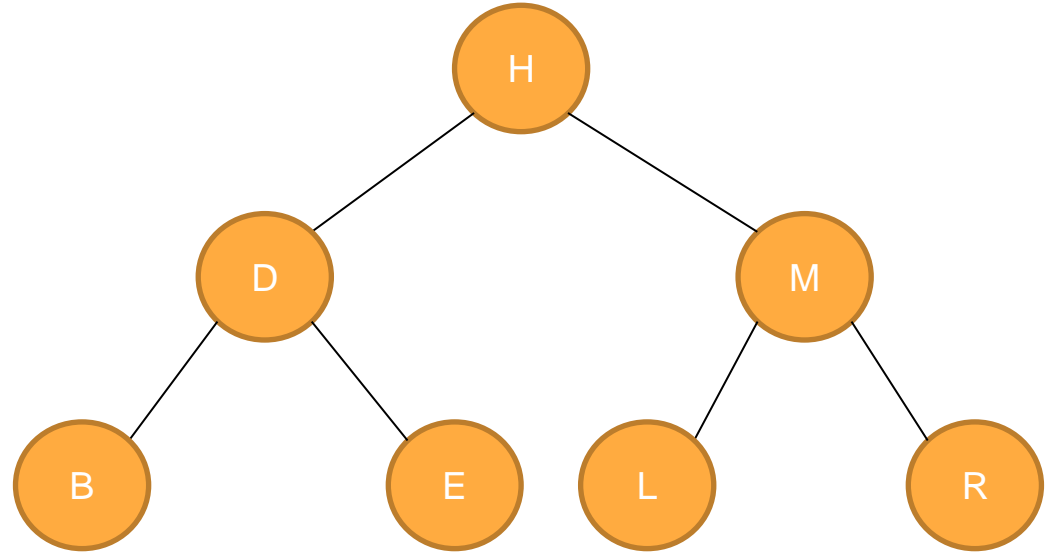
Tipos de índices - Árvores

- Estrutura específica
 - Esquerda
 - Elementos menores
 - Direita
 - Elementos maiores
- Inserção
 - H - D - M - B - R - E - L



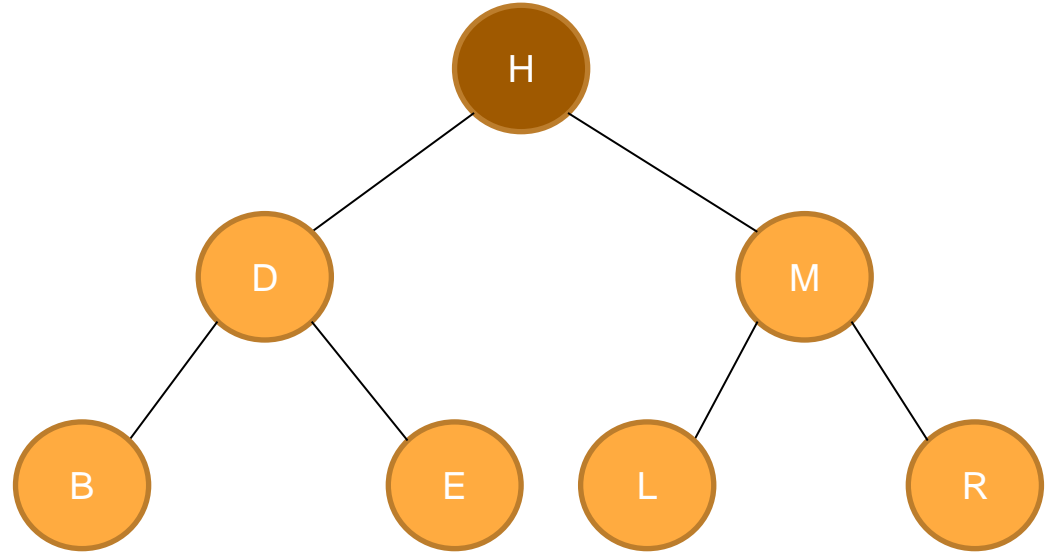
Tipos de índices - Árvores

- Estrutura específica
 - Esquerda
 - Elementos menores
 - Direita
 - Elementos maiores
- Inserção
 - H - D - M - B - R - E - L
- Busca
 - H



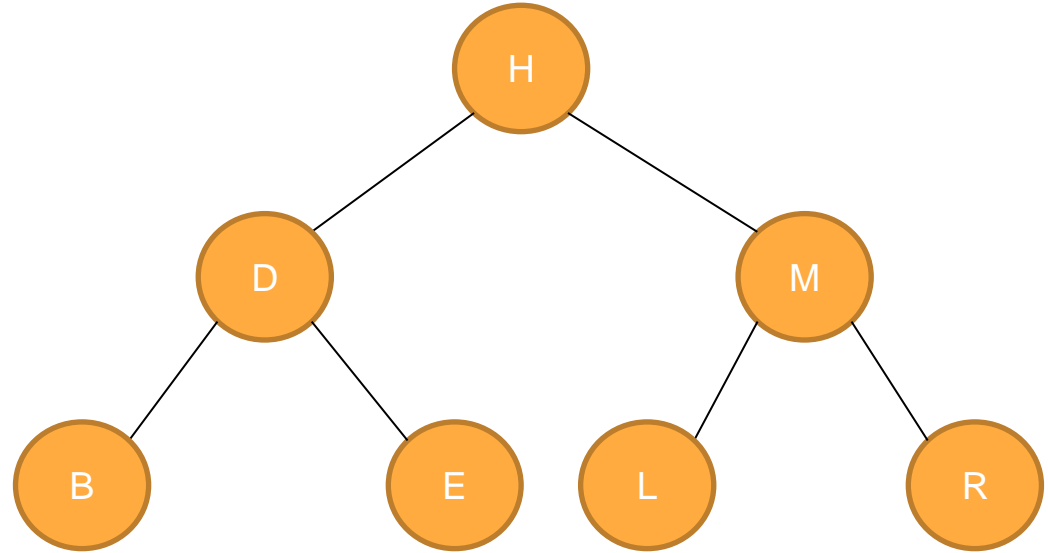
Tipos de índices - Árvores

- Estrutura específica
 - Esquerda
 - Elementos menores
 - Direita
 - Elementos maiores
- Inserção
 - H - D - M - B - R - E - L
- Busca
 - H - 1 acesso



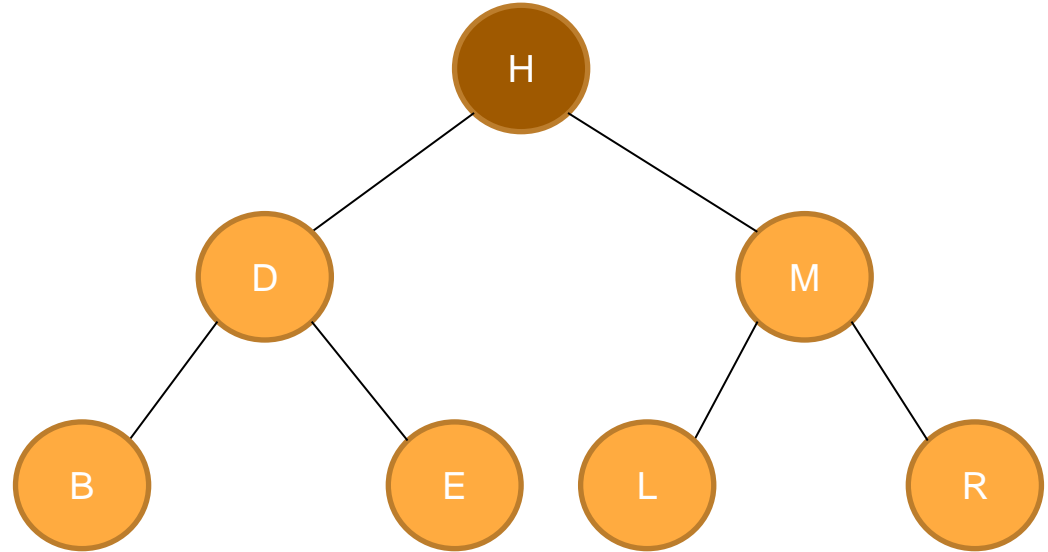
Tipos de índices - Árvores

- Estrutura específica
 - Esquerda
 - Elementos menores
 - Direita
 - Elementos maiores
- Inserção
 - H - D - M - B - R - E - L
- Busca
 - H - 1 acesso
 - M



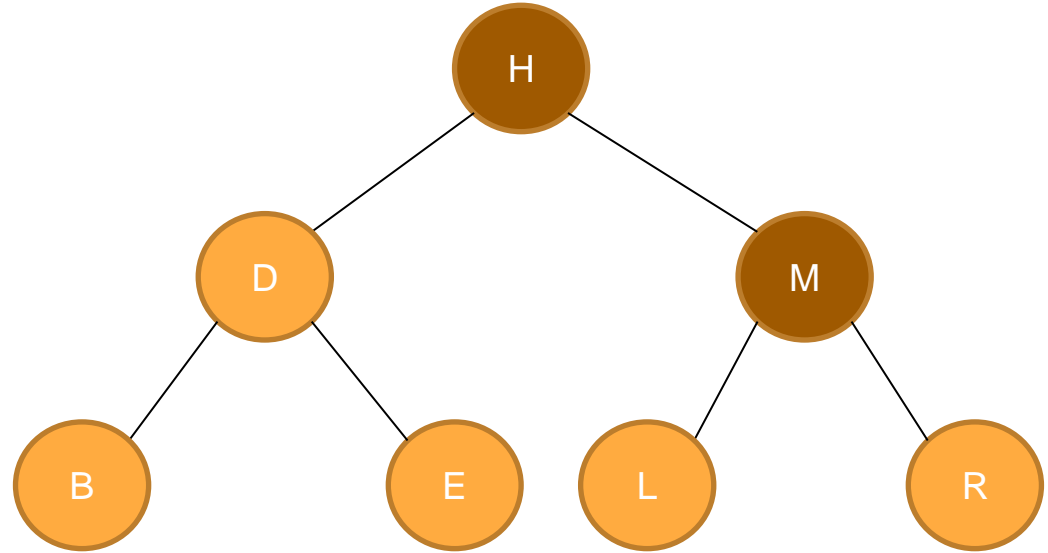
Tipos de índices - Árvores

- Estrutura específica
 - Esquerda
 - Elementos menores
 - Direita
 - Elementos maiores
- Inserção
 - H - D - M - B - R - E - L
- Busca
 - H - 1 acesso
 - M



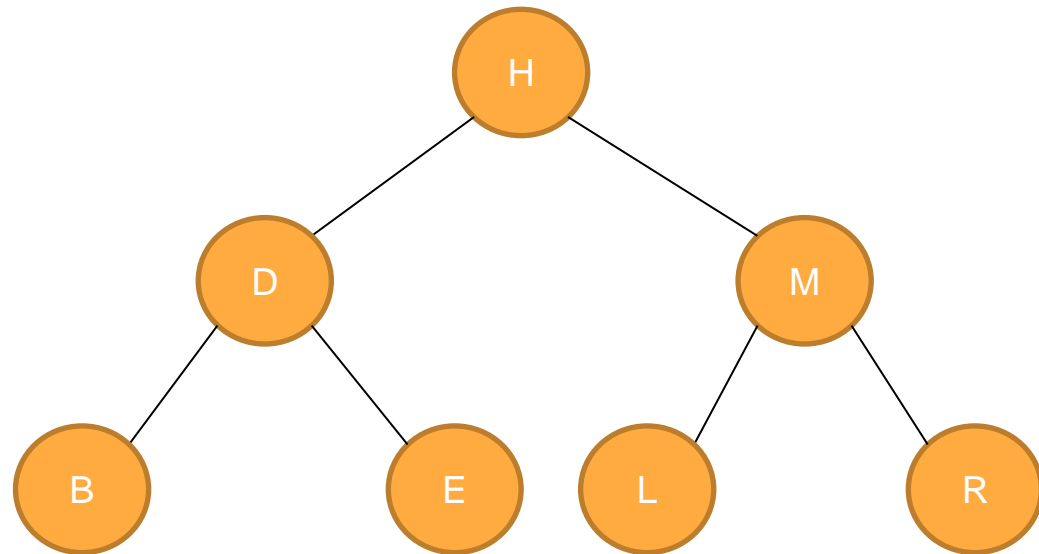
Tipos de índices - Árvores

- Estrutura específica
 - Esquerda
 - Elementos menores
 - Direita
 - Elementos maiores
- Inserção
 - H – D – M – B – R – E – L
- Busca
 - H – 1 acesso
 - M – 2 acessos



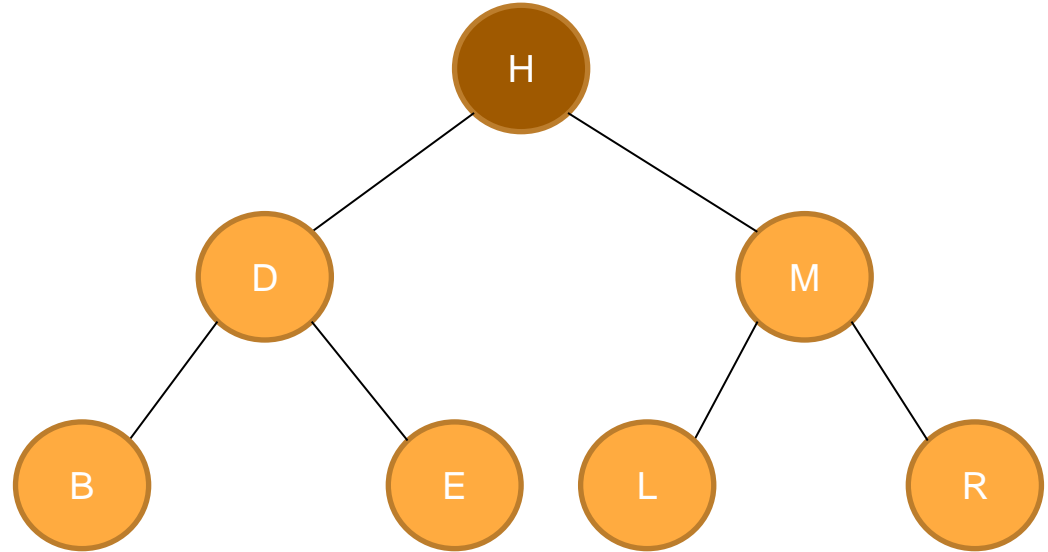
Tipos de índices - Árvores

- Estrutura específica
 - Esquerda
 - Elementos menores
 - Direita
 - Elementos maiores
- Inserção
 - H - D - M - B - R - E - L
- Busca
 - H - 1 acesso
 - M - 2 acessos
 - E



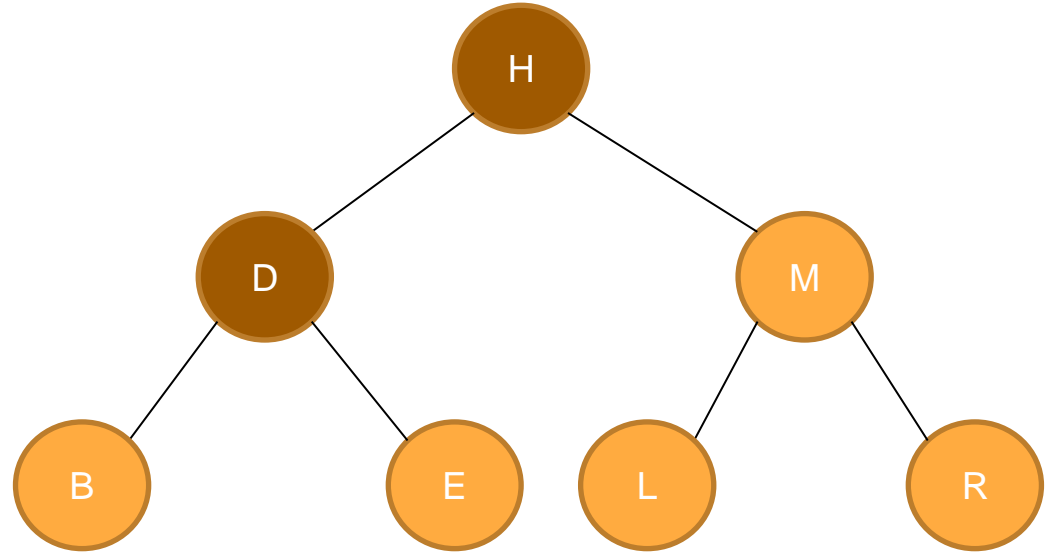
Tipos de índices - Árvores

- Estrutura específica
 - Esquerda
 - Elementos menores
 - Direita
 - Elementos maiores
- Inserção
 - H - D - M - B - R - E - L
- Busca
 - H - 1 acesso
 - M - 2 acessos
 - E



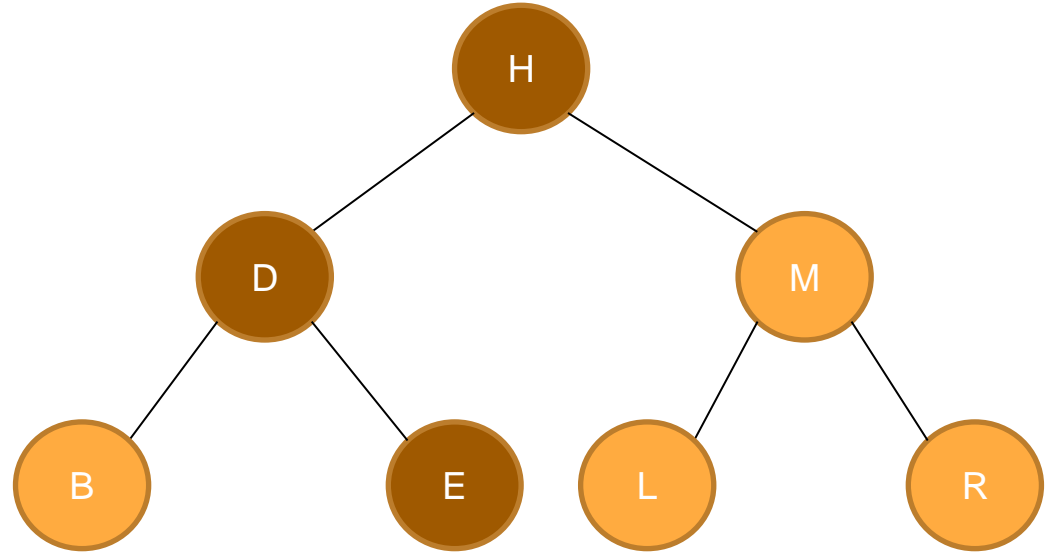
Tipos de índices - Árvores

- Estrutura específica
 - Esquerda
 - Elementos menores
 - Direita
 - Elementos maiores
- Inserção
 - H - D - M - B - R - E - L
- Busca
 - H - 1 acesso
 - M - 2 acessos
 - E



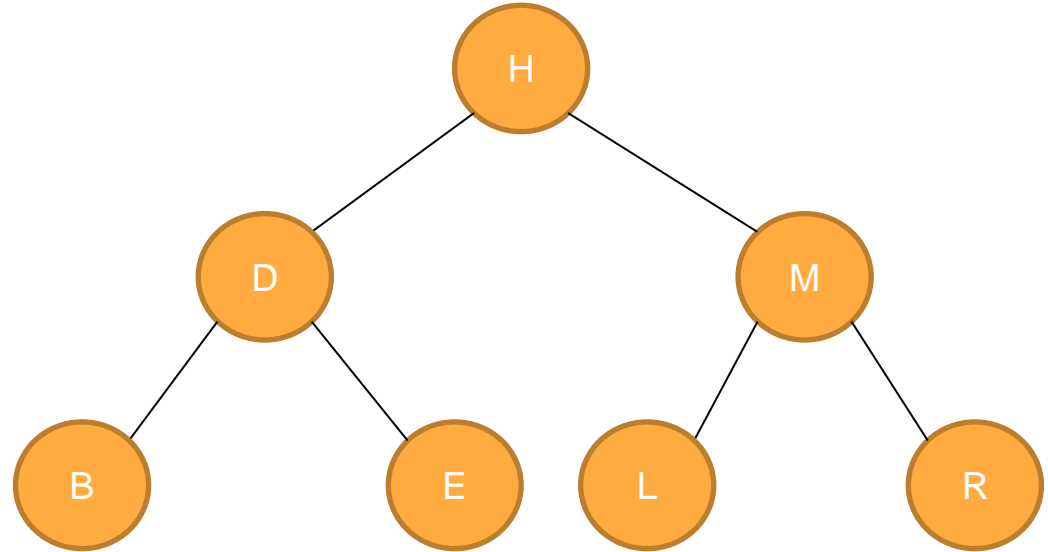
Tipos de índices - Árvores

- Estrutura específica
 - Esquerda
 - Elementos menores
 - Direita
 - Elementos maiores
- Inserção
 - H - D - M - B - R - E - L
- Busca
 - H - 1 acesso
 - M - 2 acessos
 - E - 3 acessos



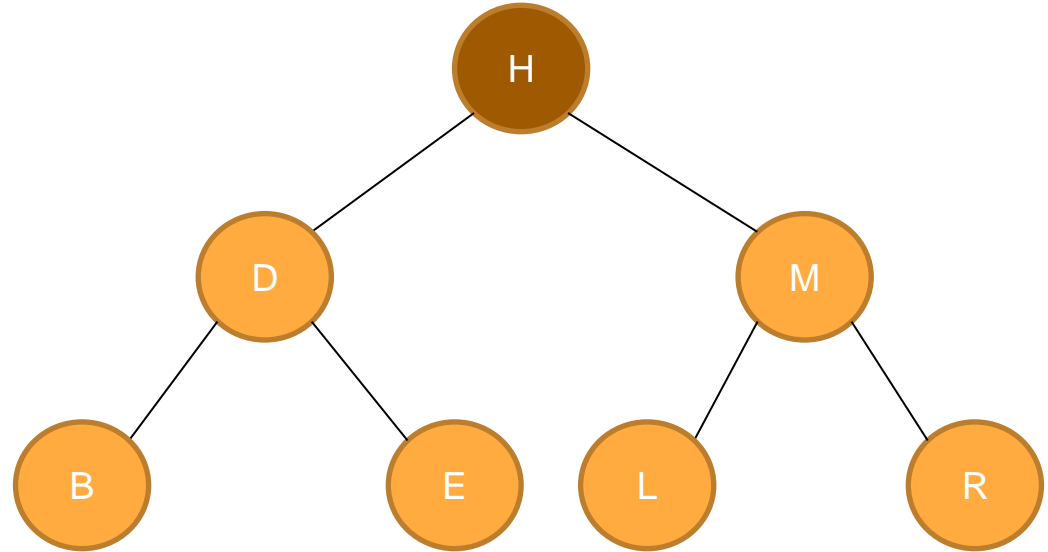
Tipos de índices - Árvores

- Estrutura específica
 - Esquerda
 - Elementos menores
 - Direita
 - Elementos maiores
- Inserção
 - H - D - M - B - R - E - L
- Busca
 - H - 1 acesso
 - M - 2 acessos
 - E - 3 acessos
 - A



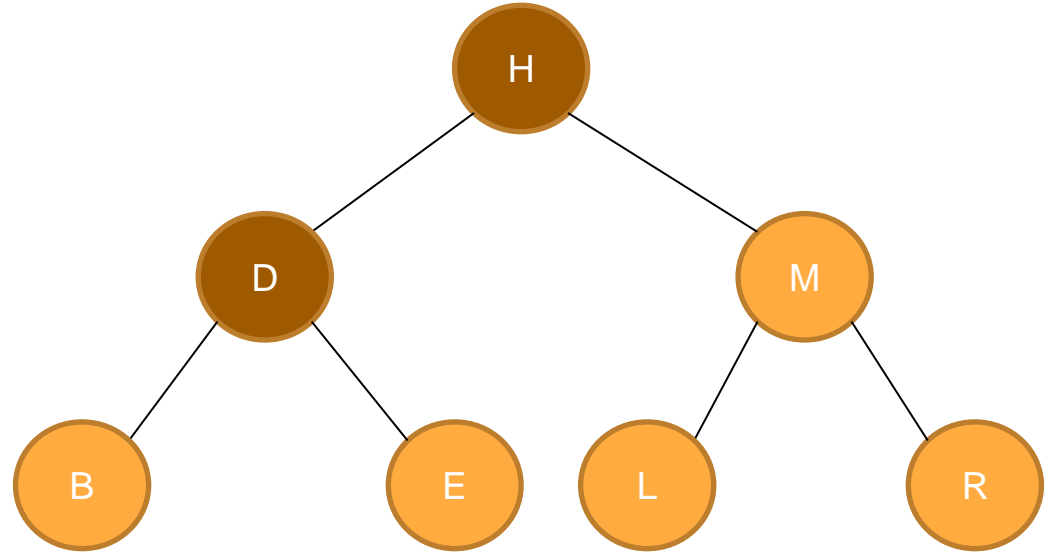
Tipos de índices - Árvores

- Estrutura específica
 - Esquerda
 - Elementos menores
 - Direita
 - Elementos maiores
- Inserção
 - H - D - M - B - R - E - L
- Busca
 - H - 1 acesso
 - M - 2 acessos
 - E - 3 acessos
 - A



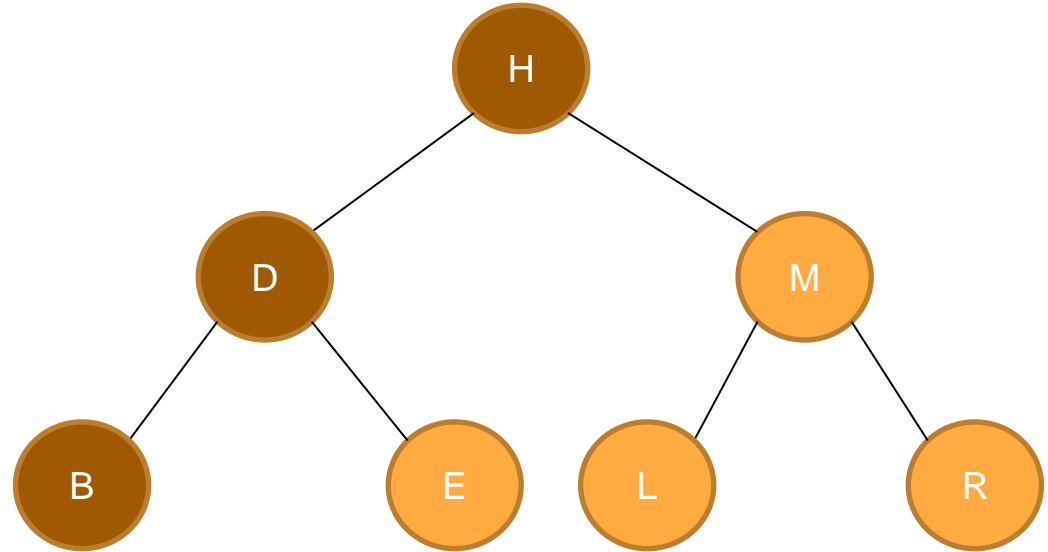
Tipos de índices - Árvores

- Estrutura específica
 - Esquerda
 - Elementos menores
 - Direita
 - Elementos maiores
- Inserção
 - H - D - M - B - R - E - L
- Busca
 - H - 1 acesso
 - M - 2 acessos
 - E - 3 acessos
 - A



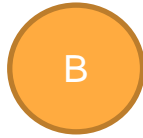
Tipos de índices - Árvores

- Estrutura específica
 - Esquerda
 - Elementos menores
 - Direita
 - Elementos maiores
- Inserção
 - H – D – M – B – R – E – L
- Busca
 - H – 1 acesso
 - M – 2 acessos
 - E – 3 acessos
 - A – Não encontrado



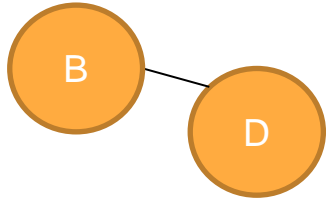
Tipos de índices - Árvores

- Problema
 - Árvore desbalanceada
- Inserção
 - B



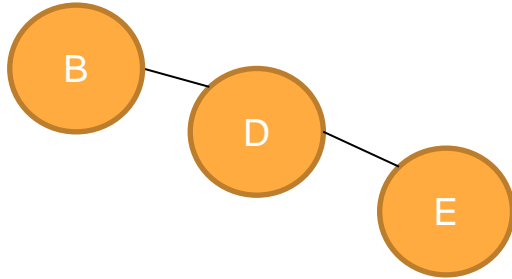
Tipos de índices - Árvores

- Problema
 - Árvore desbalanceada
- Inserção
 - B – D



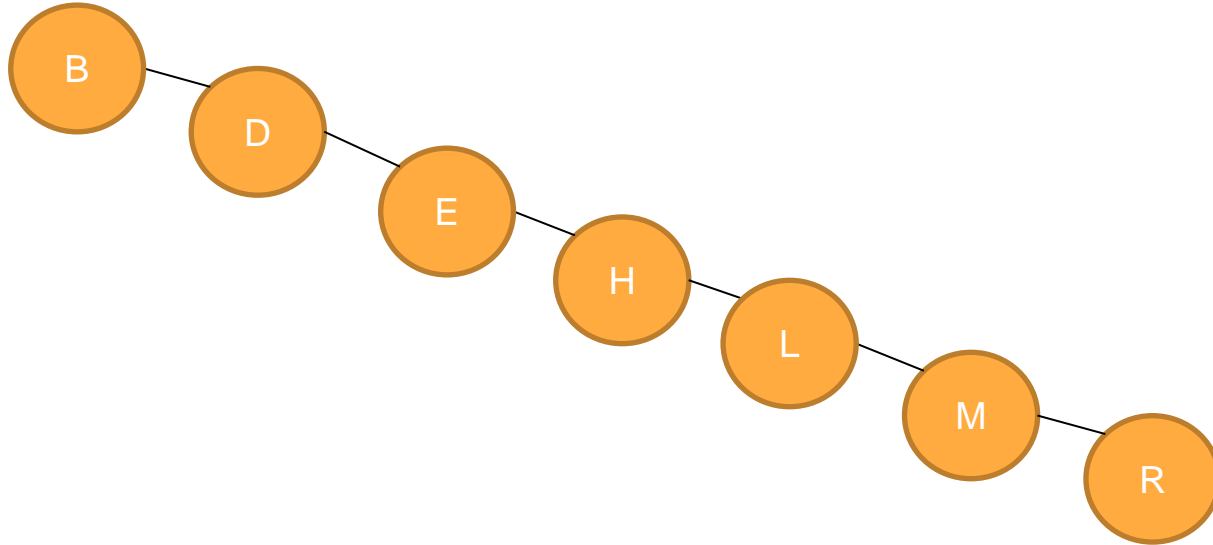
Tipos de índices - Árvores

- Problema
 - Árvore desbalanceada
- Inserção
 - B – D – E



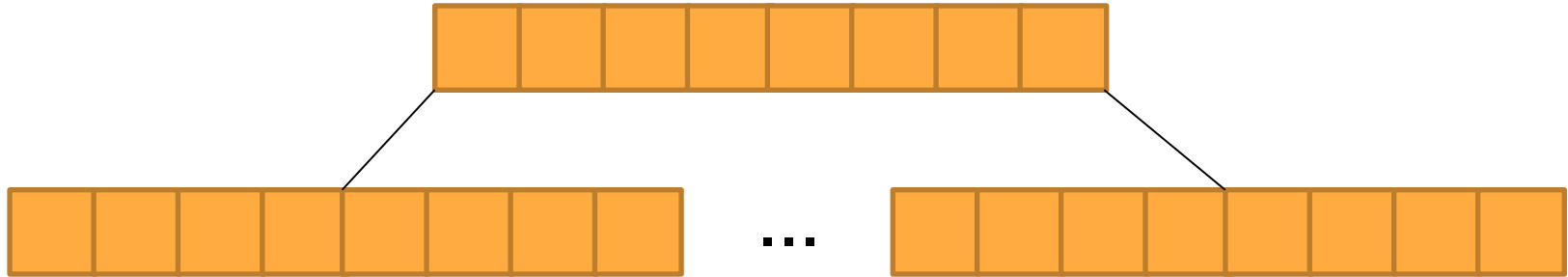
Tipos de índices - Árvores

- Problema
 - Árvore desbalanceada
- Inserção
 - B-D-E-H-L-M-R



Tipos de índices - Árvores

- Árvore B
 - Página de Disco



Tipos de índices - Árvores

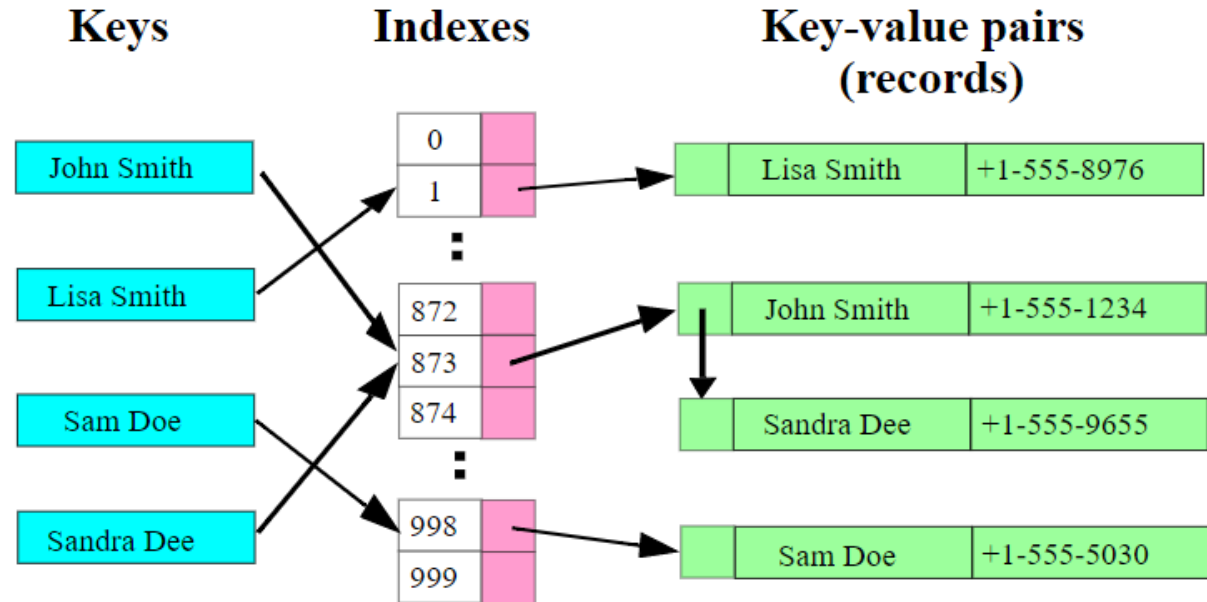
- Árvore B
 - Indicado
 - Intervalos – alta cardinalidade
 - Não indicado
 - Poucos valores – baixa cardinalidade

Agenda

- Motivação
- Tipos de índices
 - Árvores
 - Hash
 - Bitmap
- Índice para DW
- Índice para similaridade
- Conclusão

Tipos de índices - Hash

- Função Hash
 - Mapeamento direto



Tipos de índices - Hash

- Hash
 - Indicado
 - Acesso a linhas específicas
 - Não indicado
 - Intervalos

Agenda

- Motivação
- Tipos de índices
 - Árvores
 - Hash
 - Bitmap
- Índice para DW
- Índice para similaridade
- Conclusão

Tipos de índices - Bitmap

- Estruturas auxiliares
 - Binárias
 - Uma para cada valor possível

	LINHA																		
VALOR	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
Masculino	0	0	1	1	1	0	1	0	0	0	0	1	0	1	1	1	0	1	0
Feminino	1	1	0	0	0	1	0	1	1	1	1	0	1	0	0	0	1	0	1

Tipos de índices - Bitmap

- Bitmap
 - Indicado
 - Conjuntos de valores e baixa cardinalidade
 - Não indicado
 - Conjuntos grandes e atualização

Agenda

- Motivação
- Tipos de índices
 - Árvores
 - Hash
 - Bitmap
- Índice para DW
- Índice para similaridade
- Conclusão

Índice para DW

- Índice bitmap

filial

chaveFilial	nomeFilial	cidade	estado	regiao	pais
1	Filial 1	Sao Carlos	SP	SE	Brasil
2	Filial 2	Araraquara	SP	SE	Brasil
3	Filial 3	Recife	PE	NE	Brasil
4	Filial 4	Ribeirao Preto	SP	SE	Brasil
5	Filial 5	Jaboatao	PE	NE	Brasil

Araraquara	Recife	Ribeirao Preto	Sao Carlos	Jaboatao
0	0	0	1	0
1	0	0	0	0
0	1	0	0	0
0	0	1	0	0
0	0	0	0	1

índice bitmap para o atributo **cidade**

NE	SE
0	1
0	1
1	0
0	1
1	0

índice bitmap para o atributo **região**

Índice para DW

- Índice bitmap de junção

tabela de fatos: vendas

chave Filial	chave Produto	chave Tempo	lucro Dólar	unidades Vendidas
1	1	1	1	2
2	1	1	2	4
3	1	1	3	6
1	1	2	3.5	7
3	1	2	1	2
4	1	2	2	4
2	1	3	1.5	3
3	1	3	4	8
5	1	3	4.5	9
1	1	4	5	10
4	1	4	1	2
5	1	4	7.5	15

NE	SE
0	1
0	1
1	0
0	1
1	0
0	1
0	1
1	0
1	0
0	1
0	1
0	1
1	0

2014	2015
1	0
1	0
1	0
1	0
1	0
0	1
0	1
0	1
0	1
0	1
0	1
0	1
0	1

índice bitmap para
o atributo **região**

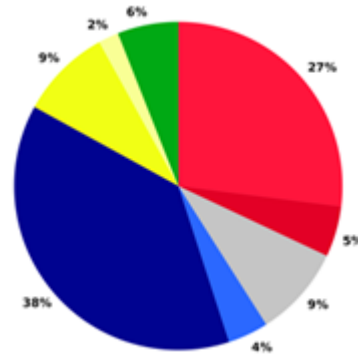
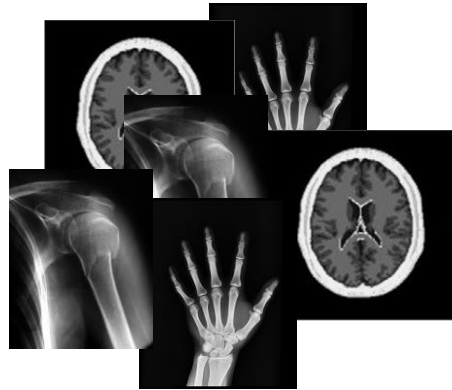
índice bitmap para
o atributo **ano**

Agenda

- Motivação
- Tipos de índices
 - Árvores
 - Hash
 - Bitmap
- Índice para DW
- Índice para similaridade
- Conclusão

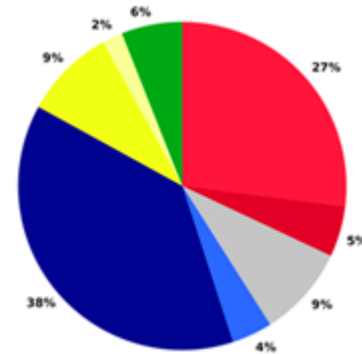
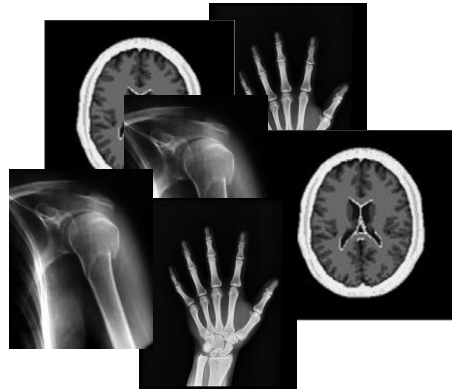
Similaridade

- Objetos complexos



Similaridade

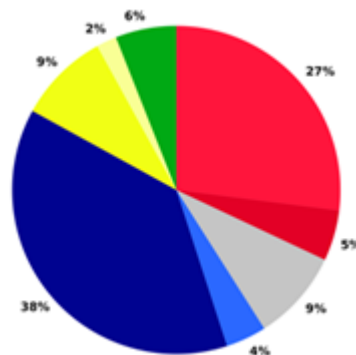
- Objetos complexos



Cor, textura, forma

Similaridade

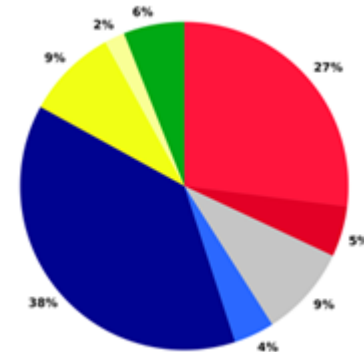
- Objetos complexos



Lat, Long

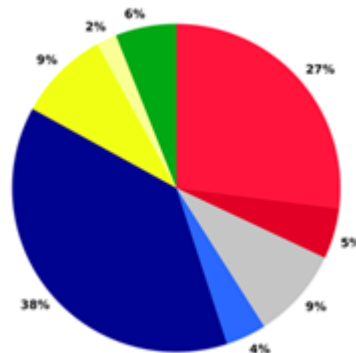
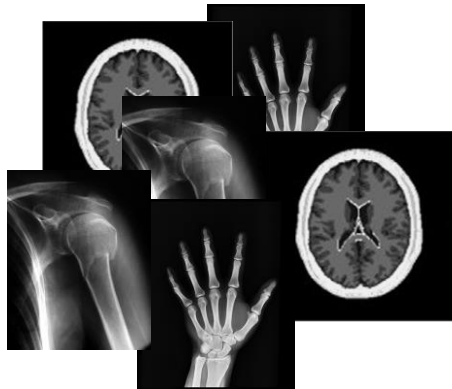
Similaridade

- Objetos complexos



Similaridade

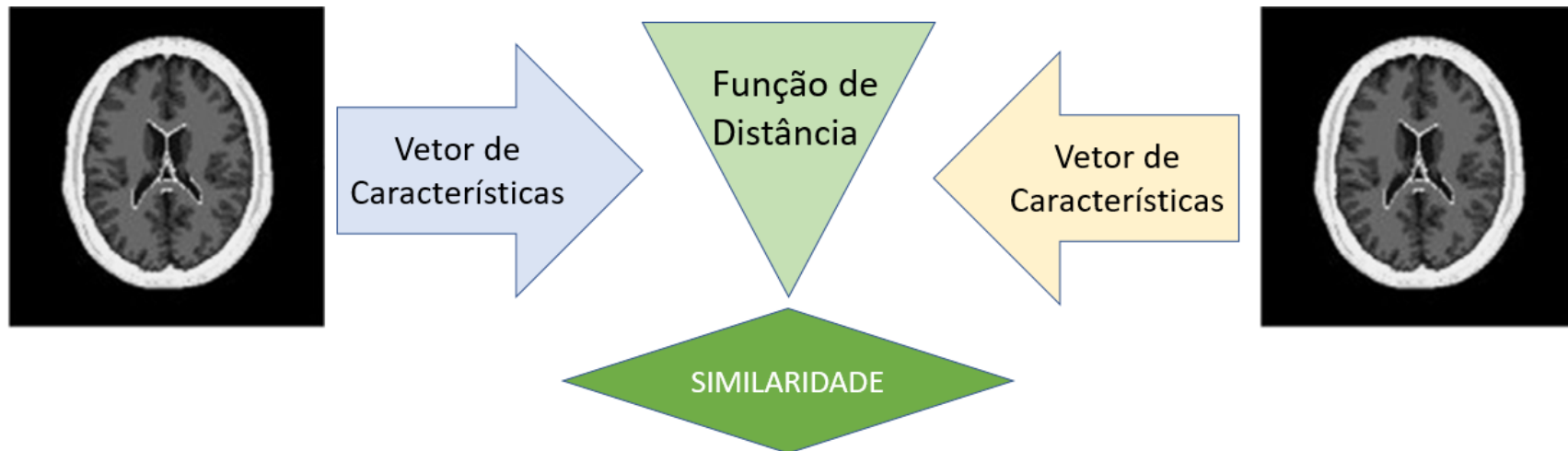
- Objetos complexos



Energia, valência

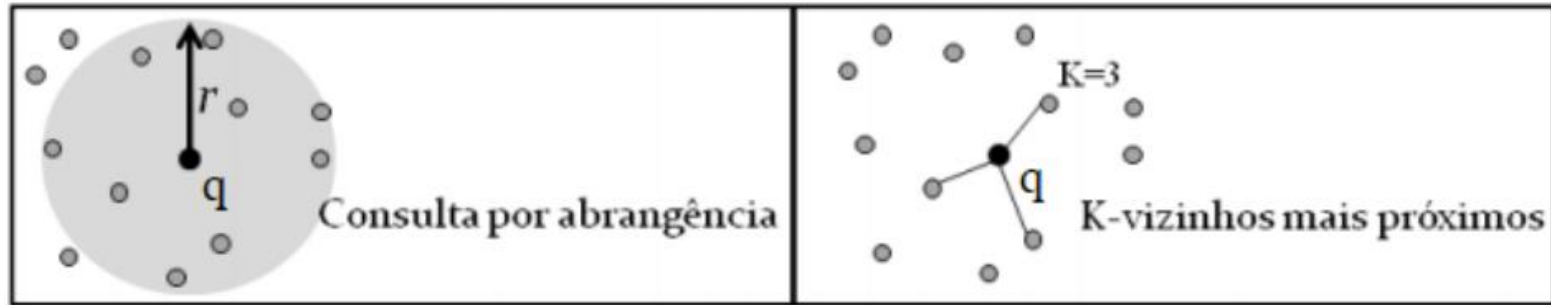
Similaridade

- Similaridade entre dois objetos complexos



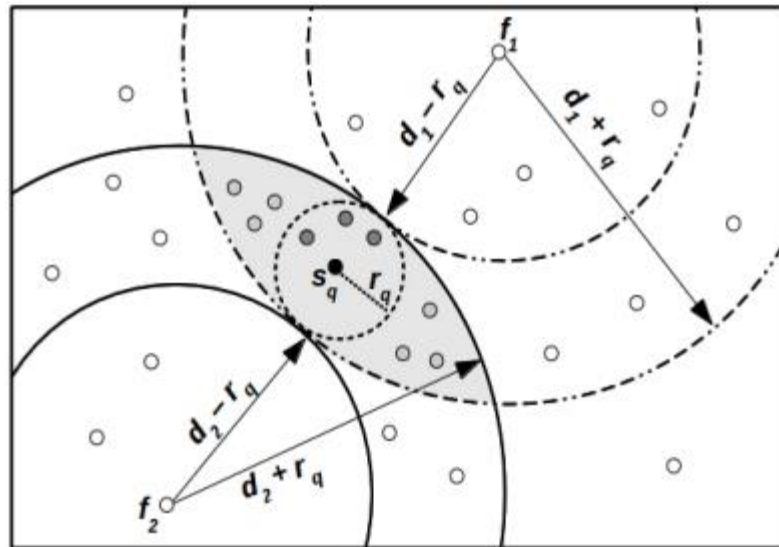
Similaridade

- Operações de similaridade



Método de Acesso Métrico

- Técnica Omni
 - Elementos representativos
 - Região *mbOr*
 - Filtragem
 - Refinamento



Agenda

- Motivação
- Tipos de índices
 - Árvores
 - Hash
 - Bitmap
- Índice para DW
- Índice para similaridade
- Conclusão

Conclusão

- Índices
 - Otimização de consultas
 - Estruturas auxiliares
 - Diferentes tipos
 - Custo em manutenção

Agenda

- Motivação
- Tipos de índices
 - Árvores
 - Hash
 - Bitmap
- Índice para DW
- Índice para similaridade
- Conclusão