

Programação e Otimização em SQL (CET 0604)



Apresentação

Manoel Augusto, Engenheiro
Eletricista com Ênfase em Computação
e Mestre em Ciências em Engenharia
de Computação pela Escola Politécnica
da Universidade de São Paulo.

Experiência de 20 anos em
desenvolvimento de softwares com
grande volume de dados e acessos
concorrentes.



Conhecendo a Turma

- Diferentes perfis: devs, dbas, gestores (coordenadores / supervisores), estatísticos, analistas de sistemas, suporte, engenheiro / cientistas de dados
- Diferentes experiências
- Diversas indústrias



Objetivos da Disciplina

- Otimizações serão focadas em bancos de dados SQL
- Entender em detalhes o funcionamento do sistema de banco de dados
- Compreender os fatores técnicos que influenciam a performance de um banco de dados
- Utilizar as ferramentas de análise de plano de execução de consultas
- Conhecer as alternativas possíveis para modificar o plano de execução de consultas
- Monitorar consistentemente a performance do banco de dados

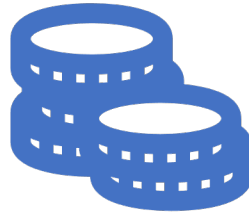
Plano de Aulas

- Distribuição dos conteúdos
- Trabalho Individual, Projeto Prático e Avaliação Teórica

Objetivos da Otimização



Melhorar a Experiência
do Cliente



Economia



Reputação

Importância da Métrica

“Aquilo que não se pode medir, não se pode melhorar”.

William Thomson – Lord Kelvin



Métricas de Otimização

Respostas mais rápidas

Respostas médias

Pior resposta

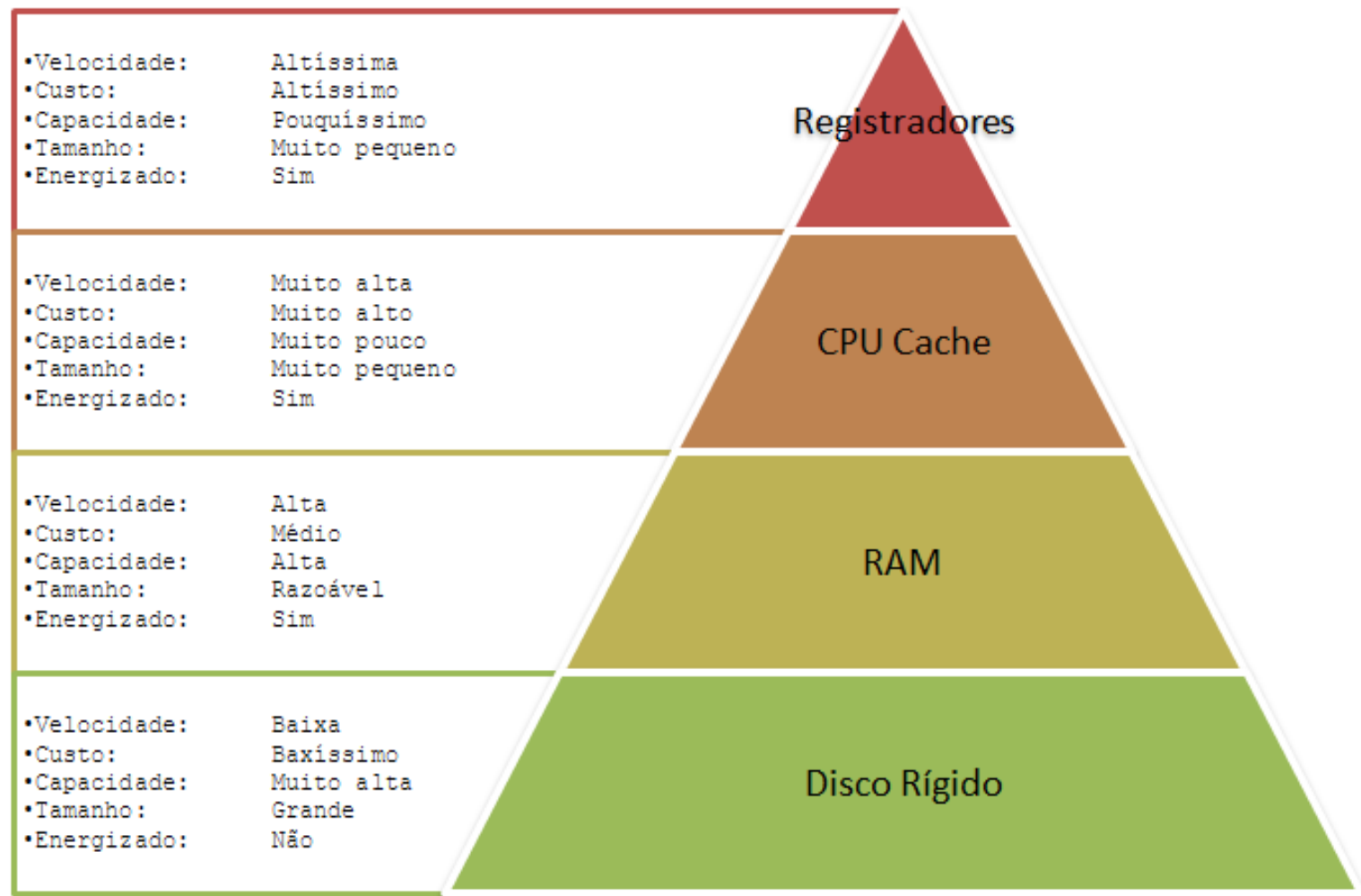
Querys específicas

Picos de acesso

Menor consumo de hardware

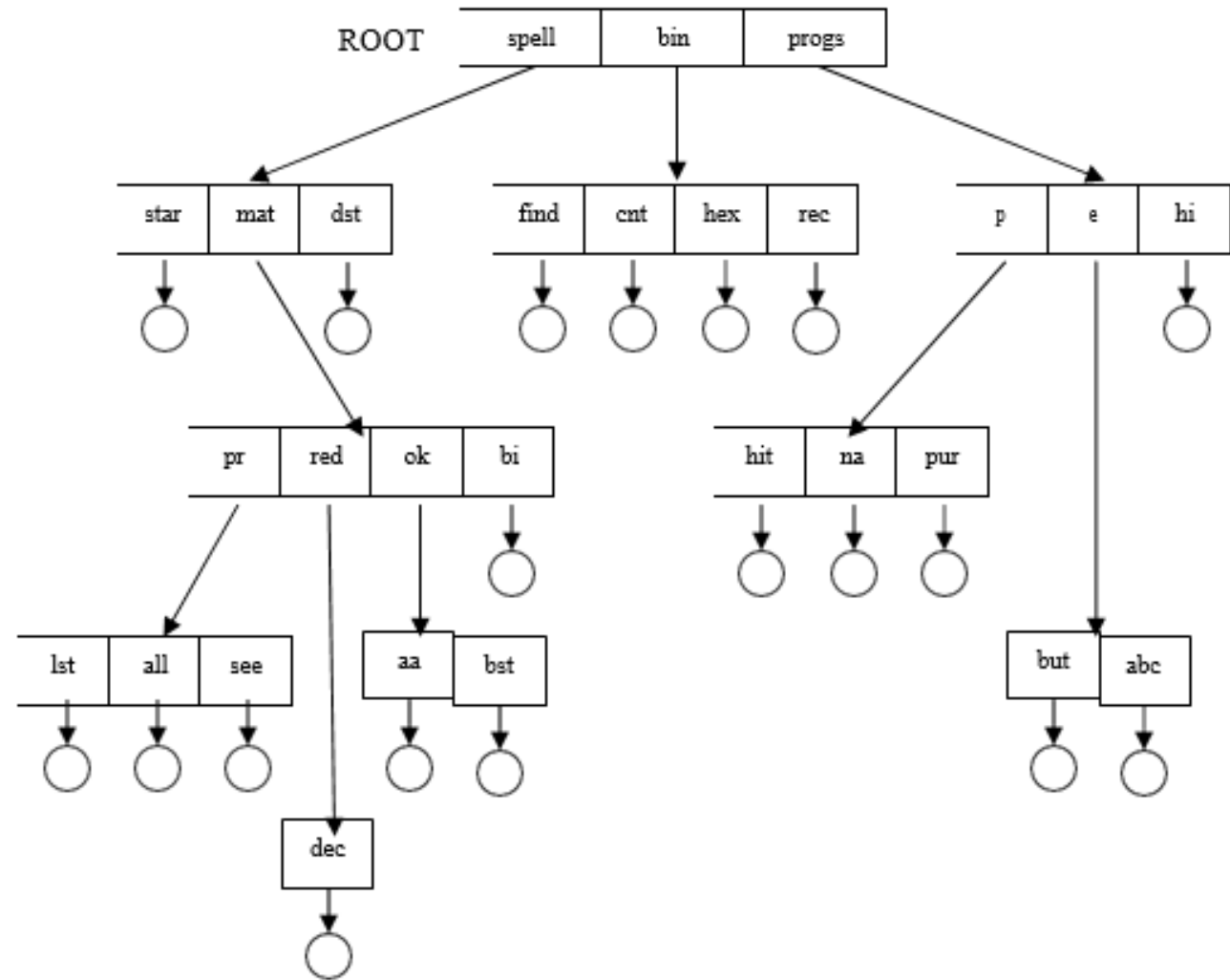
Menor custo financeiro

Hierarquia Memória



Sistema Operacional

Gerenciamento de Arquivos



Tree-Structured Directory Structure

Oportunidades de Otimização

- Modelagem
- Arquitetura de servidores
- Logs de queries
- Arquivamento de dados
- Views
- Reorganização de horários de scripts
- Cache
- Poll de conexões



Modelagem

- Normalização das tabelas
- Modelagens especiais
- Escolha correta dos tipos de dados
- Chaves primárias e estrangeiras
- Criação de índices



Normalização das tabelas

- Normalização é o processo de modelar o banco de dados projetando a forma como as informações serão armazenadas com o objetivo de eliminar a redundância.
- Forma Não Normalizada

CodProj	Tipo	Descr	Emp					
			CodEmp	Nome	Cat	Sal	DataIni	TempAl
LSC001	Novo Desenv.	Sistema de Estoque	2146	João	A1	4	1/11/91	24
			3145	Silvio	A2	4	2/10/91	24
			6126	José	B1	9	3/10/92	18
			1214	Carlos	A2	4	4/10/92	18
			8191	Mário	A1	4	1/11/92	12
PMG02	Manutenção	Sistema de RH	8191	Mário	A1	4	1/05/93	12
			4112	João	A2	4	4/01/91	24
			6126	José	B1	9	1/11/92	12

1a Forma Normal

- Eliminar repetições
- Cada célula deve conter um único valor
- Cada registro deve ser único

1a Forma Normal

Exemplo de 1FN: Livros

Tabela de Livros: Estrutura original

<u>IdLivro</u>	Título	Assunto	Autor1	Autor2	Autor3
21237	Os Sertões	Ficção	E. Cunha		
33455	Eletricidade básica	Física	A. Silva	B. Santos	
12312	Atlas do Brasil	Geografia	IBGE		

Estrutura normalizada na 1FN:

Tabela de Livros

<u>IdLivro</u>	Título	Assunto
21237	Os Sertões	Ficção
33455	Eletricidade básica	Física
12312	Atlas do Brasil	Geografia

Tabela Autores_Livros

<u>IdLivro</u>	<u>Autor</u>
21237	E. Cunha
33455	A. Silva
33455	B. Santos
12312	IBGE

1a Forma Normal

Exemplo de 1FN: Clientes

Tabela de Clientes: Estrutura original

<u>Codigo</u>	Nome	Telefone	Tipo_tel	Rua	No	Cidade
00001	Maria	3441 8566	Residencial	Contorno	2316	Belo Horizonte
00001	Maria	3215 8751	Serviço	Contorno	2316	Belo Horizonte
00001	Maria	9158 3239	Celular	Contorno	2316	Belo Horizonte
00002	Antônio	8874 5698	Celular	Afonso Pena	5693	Belo Horizonte

Estrutura normalizada na 1FN:

Tabela: Clientes

<u>Codigo</u>	Nome	Rua	No	Cidade
00001	Maria	Contorno	2316	Belo Horizonte
00002	Antônio	Afonso Pena	5693	Belo Horizonte

Tabela: Telefone_Clientes

<u>Codigo</u>	<u>Telefone</u>	<u>Tipo_tel</u>
00001	3441 8566	Residencial
00001	3215 8751	Serviço
00001	9158 3239	Celular
00002	8874 5698	Celular

1a Forma Normal

Código	Nome	Endereço	Telefone
1001	Diego Machado	Rua Tal 321 Porto	5312345678 5398765432
1002	Fulano de Tal	Avenida Tal 71 Centro	5187654321 5143215678

Código	Nome	Endereço	Bairro
1001	Diego Machado	Rua Tal 321	Porto
1002	Fulano de Tal	Avenida Tal 71	Centro

Código	Telefone
1001	5312345678
1001	5398765432
1002	5112345678
1002	5187654321

2a Forma Normal

- Satisfazer as regras da Primeira Forma Normal
- Todos os atributos de uma tabela devem depender da chave primária

2a Forma Normal

Tabela de Empregado_Projeto: Estrutura original

<u>Num_emp</u>	<u>Num_proj</u>	Horas	Nome_emp	Nome_proj	Local_proj
00001	001	8	Maria	Versão Evolutiva 3.22	João Monlevade
00002	001	18	José	Versão Evolutiva 3.22	João Monlevade
00003	002	12	Samara	Versão Corretiva 3.21	Belo Horizonte

Estrutura normalizada - 2FN

Tabela: projetos

<u>Num_proj</u>	Nome_proj	Local_proj
001	Versão Evolutiva 3.22	João Monlevade
002	Versão Corretiva 3.21	Belo Horizonte

Tabela: Empregado_projeto

<u>Num_emp</u>	<u>Num_proj</u>	Horas
00001	001	8
00002	001	18
00003	002	12

Tabela: Empregado

<u>Num_emp</u>	Nome_emp
00001	Maria
00002	José
00003	Samara

2a Forma Normal

cd_locacao	cd_filme	titulo_filme	devolucao	cd_cliente
1010	201	The Matrix	2011-10-12	743
1011	302	O Grito	2011-12-10	549
1012	201	The Matrix	2011-12-30	362

cd_filme	titulo_filme
201	The Matrix
302	O Grito

cd_locacao	cd_filme	devolucao	cd_cliente
1010	201	2011-10-12	743
1011	302	2011-12-10	549
1012	201	2011-12-30	362

3a Forma Normal

- Satisfazer as regras da Segunda Forma Normal
- Não pode ter atributos não-chave se referindo a outros atributos não-chave

3a Forma Normal

Exemplo de 3FN: Empregados trabalhando em departamentos

Tabela de Empregado_Depto: Estrutura original

<u>Num_emp</u>	Nome	Data_nasc	Num_Depto	Nome_Depto	Emp_Ger_Depto
00001	Maria	06/03/1977	001	Homologação	018
00002	José	27/05/1973	002	Homologação	018
00003	Samara	24/08/1984	003	Desenvolvimento	005

Tabela: Empregado_Depto

<u>Num_emp</u>	Nome	Data_nasc	Num_Depto
00001	Maria	06/03/1977	001
00002	José	27/05/1973	002
00003	Samara	24/08/1984	003

Tabela: Departamento

<u>Num_Depto</u>	Nome_Depto	Emp_Ger_Depto
001	Homologação	018
002	Homologação	018
003	Desenvolvimento	005

3a Forma Normal

placa	modelo	qtd_kmetro	cod_fab	nome_fab
qwe1234	Modelo1	867	3004	fabricante1
asd456	Modelo2	928	3005	fabricante2

placa	modelo	qtd_kmetro	cod_fab
qwe1234	Modelo1	867	3004
asd456	Modelo2	928	3005

cod_fab	nome_fab
3004	fabricante1
3005	fabricante2

Sharding

Original Table

CUSTOMER ID	FIRST NAME	LAST NAME	CITY
1	Alice	Anderson	Austin
2	Bob	Best	Boston
3	Carrie	Conway	Chicago
4	David	Doe	Denver

Vertical Shards

VS1

CUSTOMER ID	FIRST NAME	LAST NAME
1	Alice	Anderson
2	Bob	Best
3	Carrie	Conway
4	David	Doe

VS2

CUSTOMER ID	CITY
1	Austin
2	Boston
3	Chicago
4	Denver

Horizontal Shards

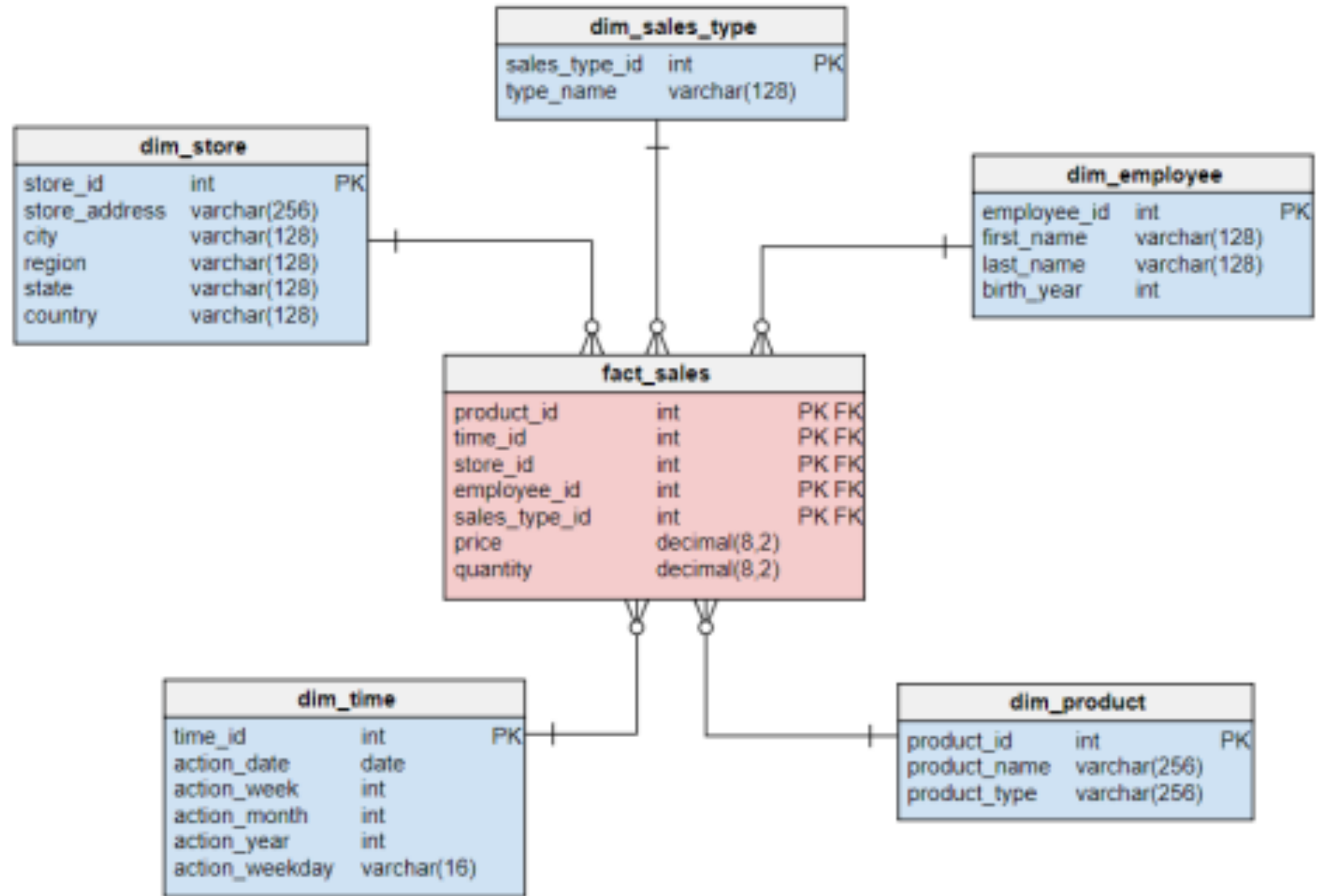
HS1

CUSTOMER ID	FIRST NAME	LAST NAME	CITY
1	Alice	Anderson	Austin
2	Bob	Best	Boston

HS2

CUSTOMER ID	FIRST NAME	LAST NAME	CITY
3	Carrie	Conway	Chicago

Fato Dimensão (Star Schema)



A large orange circle is positioned on the left side of the slide, partially cut off by the edge.

Fato
Dimensão

“Entenda o esquema em estrela e a importância para o Power BI”

<https://docs.microsoft.com/pt-br/power-bi/guidance/star-schema>

Estudem o material para a próxima aula

A yellow dashed line is located in the bottom right corner of the slide, consisting of several short, curved segments.

Arquitetura de servidores

- Melhor formato para escalabilidade do sistema
 - Cluster Público, Privado ou Híbrido
 - Múltipla escrita
 - Réplicas de consulta
- Hardware adequado
 - Processador
 - Memória
 - Disco
 - Rede





Logs

Análise do tipo de consultas recebidas

Análise da frequência

Repetições

Picos de acesso

Consome recursos do banco de dados,
gravação em disco





Arquivamento de Dados

- Validade de uma informação
- Shards Horizontais para separar dados com validades diferentes
- Versionamento de informações
- Tabelas quentes x tabelas frias
- Otimização de velocidade
- Redução de custos

Views

- Restringir os acessos a dados escolhidos e delimitados
- Segurança
- Performance

Reorganização de Horários de Scripts

- Agendar as tarefas possíveis fora dos horários de pico de acesso
- Não agrupar diversas tarefas com um curto espaço de tempo
- Monitorar a possível degradação do tempo de execução dos scripts



Cache

$$\frac{\text{Number of cache hits}}{(\text{Number of cache hits} + \text{Number of cache misses})} = \text{Cache hit ratio}$$

- Evitar o processamento completo ou parcial das consultas
- Habilitar no banco de dados
- Implementar na camada de aplicação
- Possível utilizar ambas estratégias para aumentar a performance
- Monitorar o cache hit ratio

+

•

○

Índices

Inspirado na organização de bibliotecas (número decimal de Dewey)

500 Natural sciences and mathematics

510 Mathematics

516 Geometry

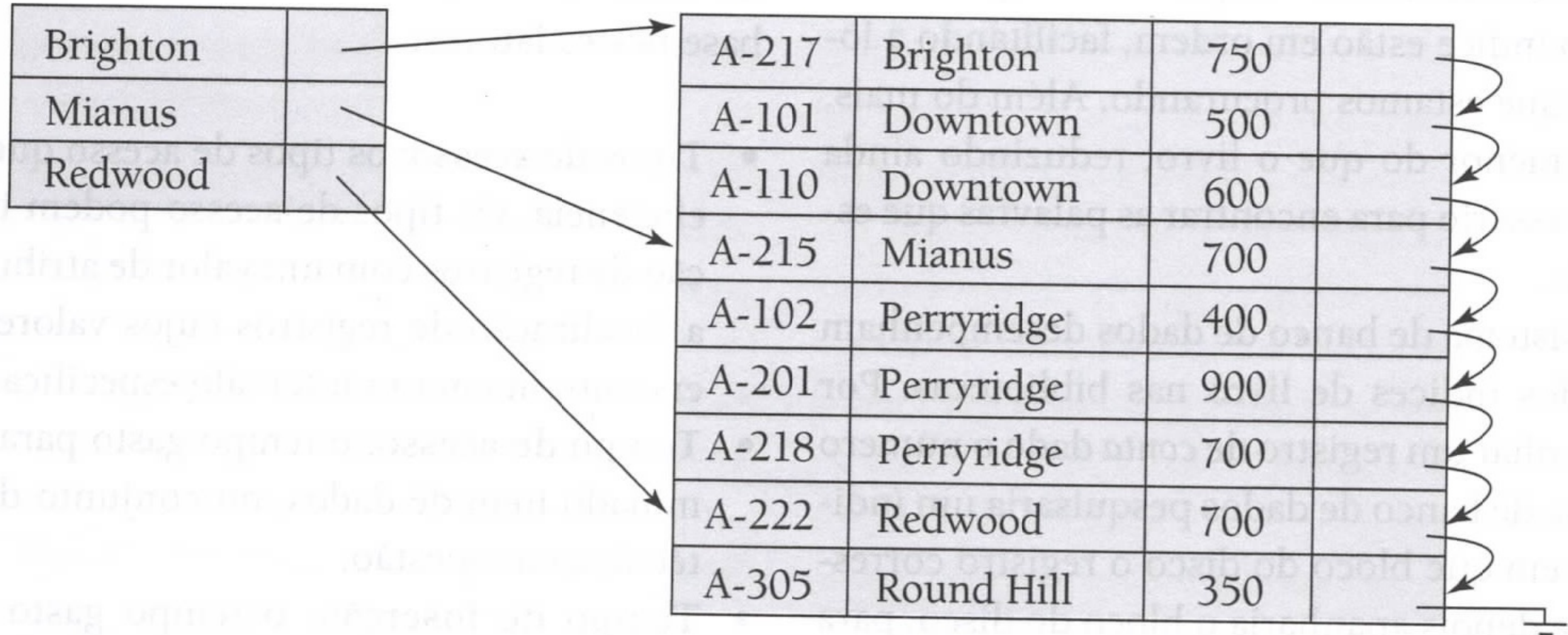
516.3 Analytic geometries

516.37 Metric differential geometries

516.375 Finsler geometry

- Diferentes tipos
- Servem para acelerar os resultados das queries

Registros de Datos



Índice Denso

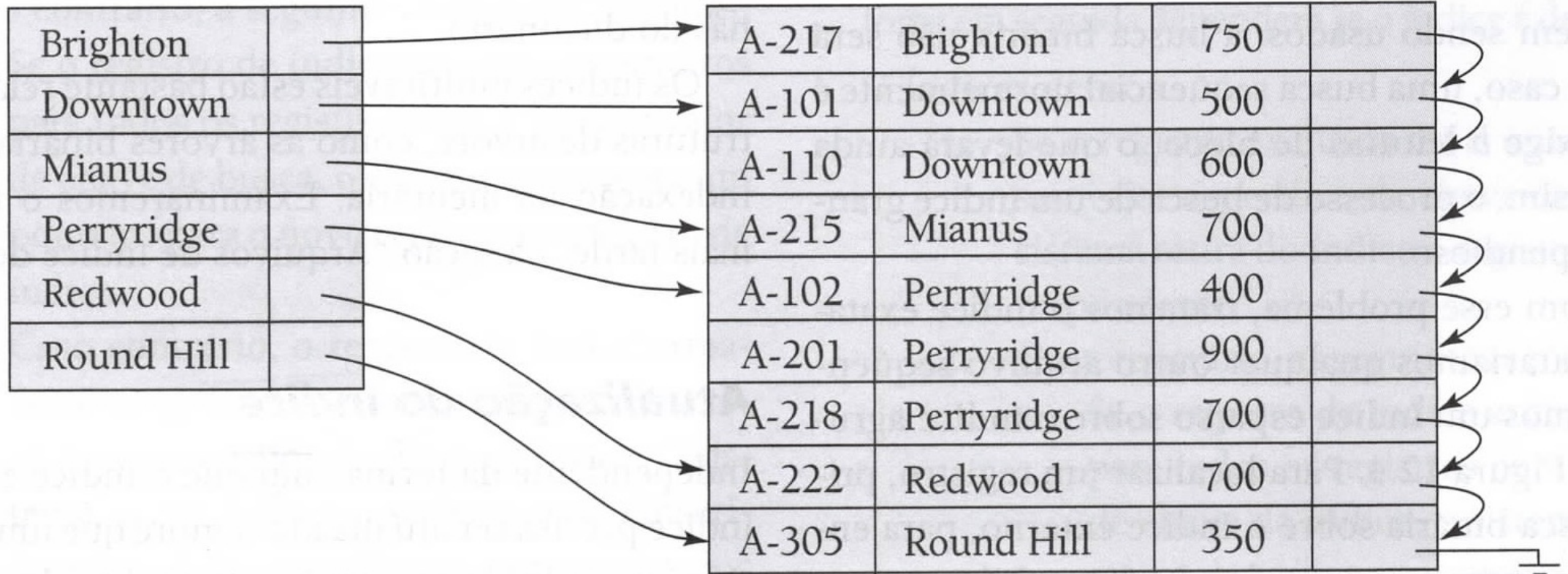


Figura 12.2 Índice denso.

Índice Esparsos

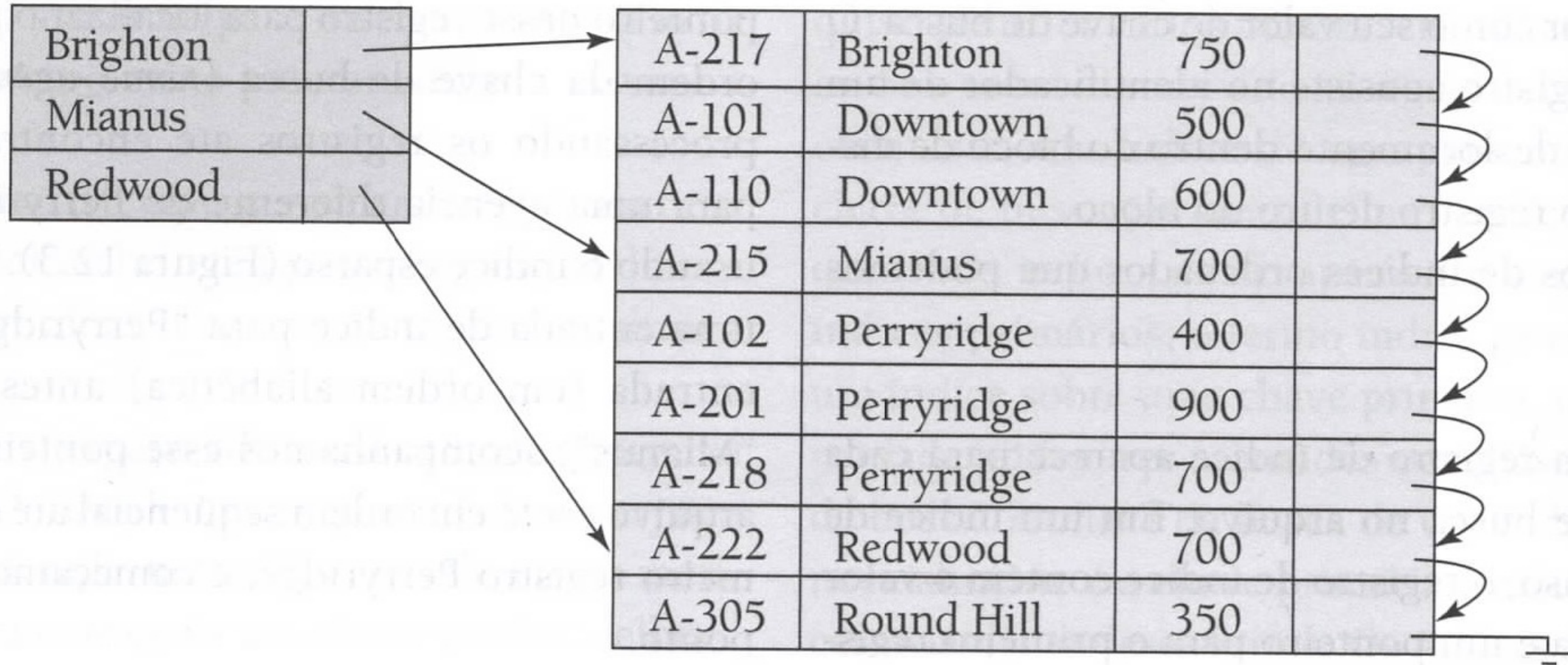


Figura 12.3 Índice esparso.

Índice Multi-nível

- Ex. Dicionário

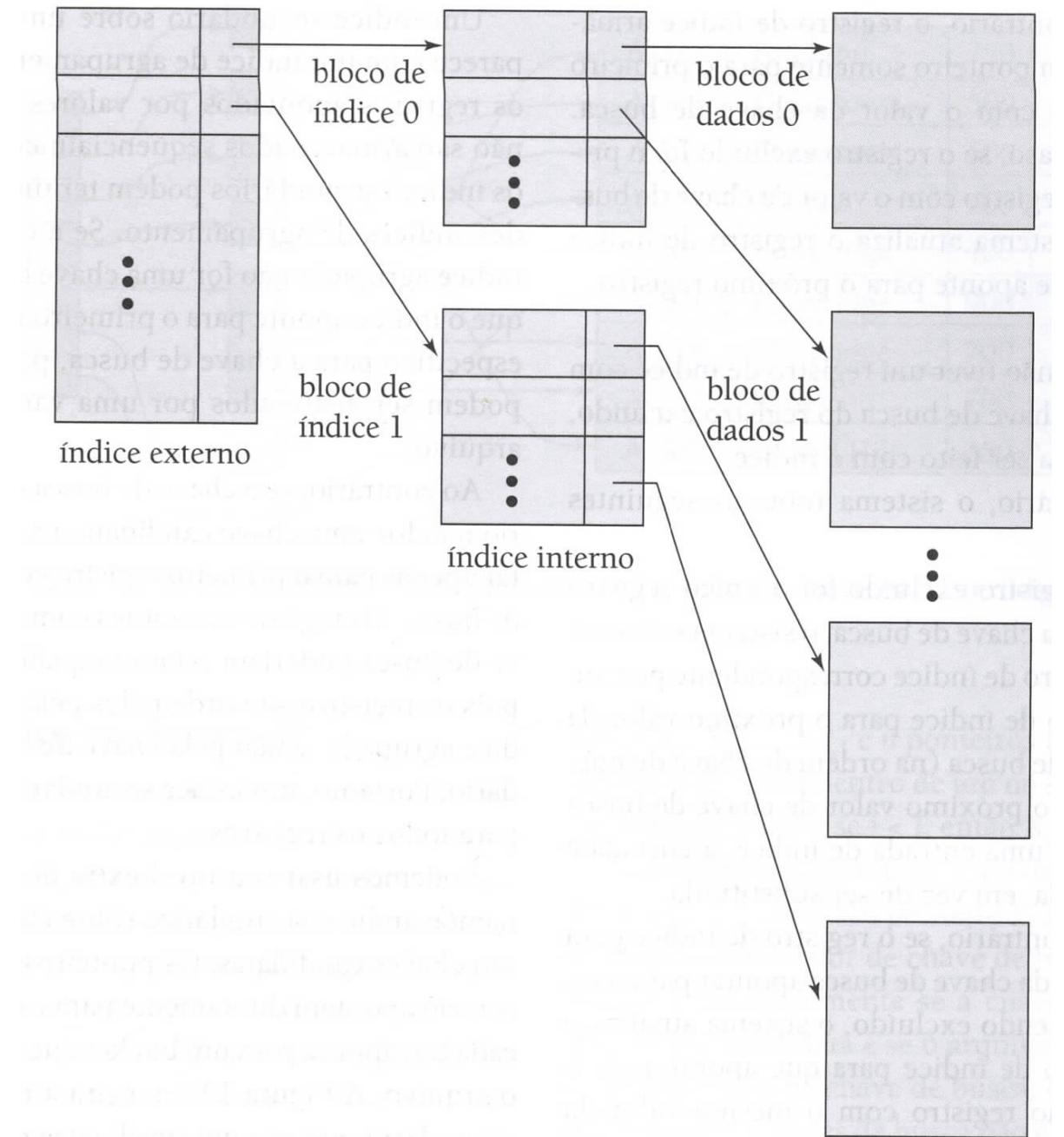


Figura 12.4 Índice esparso de dois níveis.

Índice Secundário -> Densos

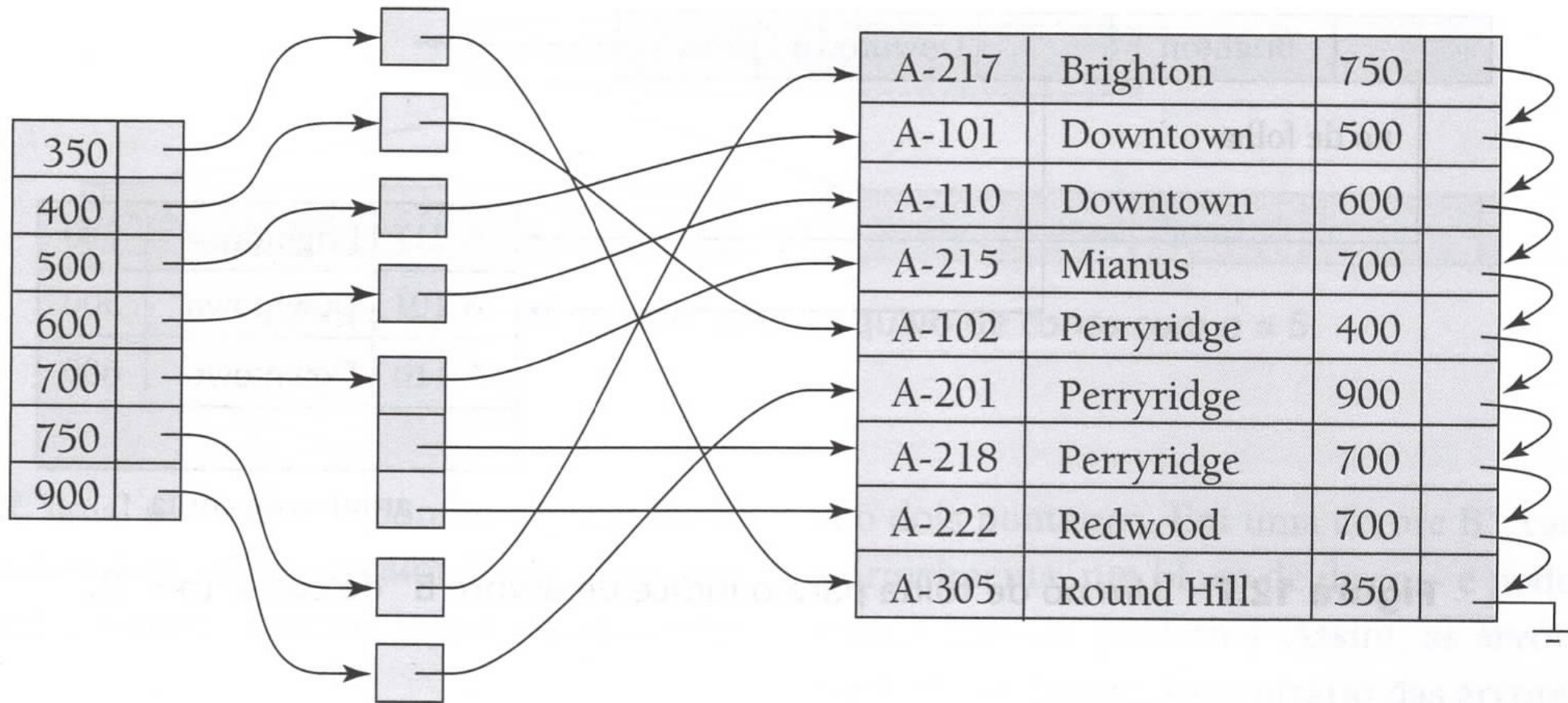



Figura 12.5 Índice secundário no arquivo de *conta*, sobre a chave não candidata *saldo*.

Problema resolvido?

Seria somente criar índices para todas as queries possíveis?



Lado Ruim:
Quanto mais
índices

- Maior a lentidão na escrita de dados
 - INSERT
 - UPDATE
 - DELETE
- Maior consumo de espaço
- Maior utilização da memória RAM do servidor

A large orange circle occupies the left side of the slide, partially cut off by the edge.

Discussão
Artigo

TUNING - Técnicas de Otimização de Banco de Dados - Um Estudo Comparativo: Mysql e Postgresql



A large orange circle occupies the left side of the slide, partially cut off by the edge.

Artigo
Leitura

Performance Impact of Optimization Methods on MySQL Document-Based and Relational Databases





Trabalhos Avaliativos

Análise crítica dos 3 artigos científicos:

- Resumo
- Como seria possível avançar mais no tema?

Avaliação: $(TI * 0,2) + (PP * 0,4) + (AV * 0,4)$

- TI – Trabalho Individual
 - PP – Projeto Prático
 - AV – Avaliação Teórica
-
- Prazo: 19/04/2024 23:59

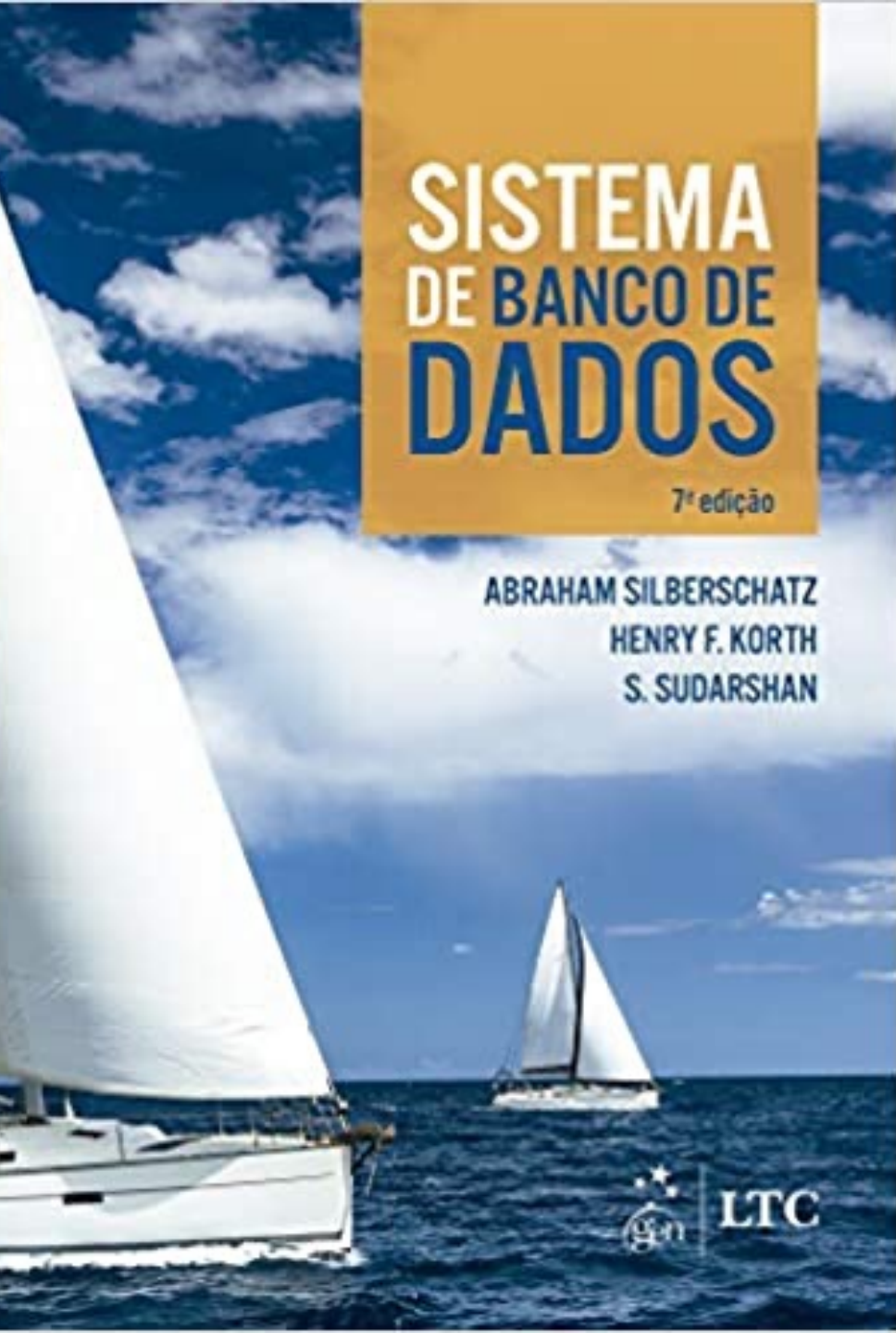
Instalação Mysql

Conseguiram?

<https://dev.mysql.com/doc/mysql-getting-started/en/#mysql-getting-started-installing>

Instalem também o Mysql Workbench:

<https://www.mysql.com/products/workbench/>



Aprofundar mais

Sistema de Banco de Dados, Abraham Silberschatz, Henry Korth e S. Sudarshan

Capítulo – INDEXAÇÃO E HASHING