



## Primeira lista de exercícios de SCC-243 – Arquitetura de SGBD

Primeiro semestre de 2013

Prof. Dr. Caetano Traina Júnior PAE: Gabriel Vicente De Pierro

Para os exercícios a seguir considere a base de dados real DBLP<sup>1</sup> que armazena informações sobre as publicações da área de Ciências de Computação, registrando autores e suas respectivas publicações em conferências, periódicos científicos e livros. A base será composta das seguintes tabelas e índices:

```
CREATE TABLE Authors (  
  IDAuthors number PRIMARY KEY,  
  AuthorName varchar2(100)  
);  
--IDAuthors - ID sequencial,  
--AuthorName - Nomes válidos para os autores.  
  
CREATE TABLE Publications (  
  IDPublication number PRIMARY KEY,  
  CodePublication varchar2(120),  
  Publisher varchar2(250),  
  Title varchar2(500),  
  VolumePub varchar2(50),  
  Year number,  
  TypePub varchar2(50)  
);  
--IDPublication - ID sequencial para junções  
--CodePublication - Chave para as publicações armazenadas (diretório).  
--Publisher - Nomes validos para as editoras.  
--Title - Títulos válidos para os livros publicados.  
--VolumePub - Volume válido para os livros publicados.  
--Year - Ano de Publicação.  
--TypePub - Tipos de publicação - Restringido ao valor - Book.  
  
CREATE TABLE Publish (  
  IDPub number,  
  IDAuth number  
  FOREIGN KEY (IDPub) REFERENCES Publications(IDPublication)  
  FOREIGN KEY (IDAuth) REFERENCES Author(IDAuthors)  
);  
--IDPub - Referência para tabela Publications  
--IDAuth - Referência para tabela Authors  
  
CREATE INDEX IdxPub ON Publications(Publisher);  
CREATE INDEX IdxType ON Publications(Typepub);  
CREATE INDEX IdxYear ON Publications(Year);  
CREATE INDEX IdxPubType ON Publications(Publisher, Typepub);  
CREATE BITMAP INDEX BtTPub ON Publications (Typepub);
```

---

<sup>1</sup><http://www.informatik.uni-trier.de/ley/db/>

Métricas das tabelas:

Tabela	Atrib	ColSIZE	ColCard	Distr	Low2Key	High2Key
Publications	IDPublication	2	2200000	2	2	2199999
Publications	CodePublication	120	2200000	2	-	-
Publications	Publisher	250	90	2	'ACM'	'World Scientific'
Publications	Title	500	2200000	2	'1.2 GS/s Hadamard Transform...'	'Zero-Knowledge Using Garbled Circuits...'
Publications	VolumePub	50	100	2	2	99
Publications	Year	4	76	2	1937	2012
Publications	TypePub	50	8	2	'inproceedings'	'mastersthesis'
Authors	IDAAuthors	75000	4	2	1	74999
Authors	AuthorName	100	70000	2	'Aaron'	'Zhao'
Publish	IDPub	4	2200000	2	2	2199999
Publish	IDAAuth	4	75000	2	2	74999

Exercício 1) Para cada uma das consultas abaixo, efetue as seguintes tarefas:

- Escreva a consulta em SQL;
- Crie a árvore canônica que expressa a consulta;
- Apresente uma árvore de comando alternativa para a consulta;
- Indique quais são os predicados indexáveis para quais índices dessa consulta;
- Indique quais são os predicados de screening dessa consulta (se existirem).
- Indique a seletividade de cada predicado.

Q1 - Liste o título de todos os trabalhos publicados pela Elsevier.

Q2 - Liste os títulos e o nome dos autores de artigos publicados em periódicos (journal) a partir do ano 2010 contendo *big data* no final do título.

Q3 - Liste todos os títulos de livros (book) publicados pela ACM.

Q4 - Liste o nome dos autores com mais de 10 trabalhos publicados desde 2005.

Exercício 2) A DBLP armazena cerca de 2.200.000 publicações na sua tablespace. Suponha que ela utiliza uma unidade de disco que tem as seguintes características:

Interface: SAS 6-Gb/s  
 Cache: 16MB  
 Capacity: 600GB  
 Area density (avg): 250Gb/in<sup>2</sup>  
 Guaranteed Sectors: 1.175.123.456  
 Spindle Speed: 10.000 RPM  
 Sustained data transfer rate: 120Mb/s  
 Average latency: 3,0ms  
 Random read seek time: 4,0ms  
 Random write seek time: 4,0ms  
 I/O data transfer rate: 600MB/s  
 MTBF: 1.000.000 hours

Sabe-se que a base de dados usa Oracle e é estruturada em páginas de 4KB, e que em média são usados 60 caracteres para cada código de publicação, 120 caracteres para cada editora, 200 caracteres para cada título, 10 caracteres para cada volume e 10 para tipo de publicação.

a) Se a relação Publications fosse criada da seguinte forma:

```
CREATE TABLE Publications (  
IDPublication number PRIMARY KEY,  
CodePublication varchar2(120),  
Publisher varchar2(250),  
Title varchar2(500),  
VolumePub varchar2(50),  
Year number,  
TypePub varchar2(50)  
STORAGE (INITIAL 30K NEXT 32k)  
PCTFREE 10 PCTUSED 70  
TABLESPACE DBLP;  
);
```

quanto tempo leva para ser feita a leitura de toda a tabela em uma busca sequencial?

- b) Quanto tempo levaria essa busca se no `CREATE TABLE` não fosse usada a cláusula `STORAGE`?
- c) Como esse comando seria re-escrito se a base usada fosse Postgres?

Exercício 3) Explique sucintamente:

- a) Quais são as diferenças entre predicado indexável, predicado de restrição (Screening) e predicado de busca (Matching)?
- b) Por que é vantajoso acessar o maior número possível de dados em cada operação de acesso ao disco ?
- c) Como a coleta de estatísticas sobre os dados de uma base auxiliam no processamento de consultas ?

Exercício 4) Para cada uma das consultas do exercício 1 responda:

- a) - Quais métodos de acesso podem ser usados por cada consulta ?
- b) - Faça uma previsão de custo para cada método de acesso que pode ser utilizado.

Exercício 5) Para os 5 índices criados na base responda (assumindo que a base foi criada em oracle com valores default para `TABLESPACE`):

- a) - Qual o espaço em disco ocupado por cada índice (em KBytes e número de páginas) ?
- b) - Qual a altura e número de folhas de cada índice ?