



Instituto de Ciências Matemáticas e de Computação – USP
Programa de Ciências de Pós-graduação em Computação e Matemática Computacional

Lista de exercícios: Seletividade de Condições — Histogramas

Prof. Dr. Caetano Traina Júnior

Exercício 1) Responda:

1. Além do uso para a previsão de seletividade, descreva mais três aplicações para histogramas em SGBDRs.
2. Porque o algoritmo usado para preparar os histogramas em SGBDR não detalham atributos de baixa frequência do domínio? Porque o custo de prepará-lo aumenta?
3. Criar histogramas reduz o desempenho do comando `SELECT`? e do comando `UPDATE`? E quanto a acrescentar índices?
4. Porque deve-se dar preferência a criar histogramas sobre atributos que formam a cabeça da chave de um índice?
5. Quais as diferenças entre os histogramas que os SGBD Oracle e Postgres podem gerar?
6. Qual a diferença entre as *metaviews* `db_stats` e `db_statistics` em Postgres?
7. Qual a diferença entre usar `ANALYZE STATISTICS` e `DBM_STATS` e `PDATEm` Oracle?
8. Quais são os atributos mais importantes para ter Histogramas? Quais não justificam ter histogramas?

Exercício 2) Responda às seguintes questões explicando os conceitos envolvidos que você usou para sustentar sua resposta:

1. Histogramas auxiliam a otimização lógica ou física das expressões de consulta?
2. Para obter uma boa atuação do otimizador físico, as estatísticas da base de dados devem ter sido coletadas recentemente?
3. É possível configurar um histograma para que o SGBD tenha um desempenho equivalente à operação em modo uniforme?

Exercício 3) Uma aplicação registra em uma tabela as vendas realizadas em cada dia da semana, usando o esquema:

Venda={Código,DiaDaSemana,Dia}.

Assuma que, durante um mês de Abril que teve início numa segunda-feira, foram realizadas vendas em cada dia exatamente igual a cada dia do mês – quer dizer, uma venda dia primeiro, duas no dia dois, etc. Nessa especificação, responda às seguintes questões:

- Crie um histograma para o atributo `DiaDaSemana` com as seguintes características:
 1. um histograma trivial;
 2. um histograma completo;

3. um histograma Equi-faixa com $\beta = 5$;
 4. um histograma Equi-frequência com $\beta = 5$;
 5. um histograma Equi-área com $\beta = 6$;
 6. um histograma Seriado decrescente com $\beta = 4$;
 7. um histograma Seriado decrescente com extremo superior detalhado e $\beta = 5$;
 8. um histograma *v-optimal* com $\beta = 4$;
 9. um histograma **max-diff** com $\beta = 5$;
 10. um histograma **compressed** com $\beta = 5$ e $R = 2$;
 11. um histograma Seriado decrescente com $\beta = 5$ e aproximação $P = 2$ de frequência.
- Dos histogramas construídos, e desconsiderando o histograma completo, qual é o que melhor aproxima a distribuição real dos dados neste caso em particular? E no caso geral? Porque?
 - Mostre os histogramas que são gerados pelo Postgres com a seguinte sequência de comandos:
`ALTER TABLE Venda ALTER Dia SET STATISTICS;`
`ALTER TABLE Venda ALTER DiaDaSemana SET STATISTICS 4;`
`ALTER TABLE Venda ALTER DiaDaSemana (N_DISTINCT=2);`
`ANALYZE Venda;`
 - Escreva um comando para Postgres (ibidem Oracle) que mostra quais histogramas foram criados e quantas faixas cada um tem, para a tabela `Venda`.
 - Dos histogramas construídos, e desconsiderando o histograma completo, qual é o que melhor aproxima a distribuição real dos dados neste caso em particular? E no caso geral? Porque?
 - Suponha que o seguinte comando é executado:
`DELETE FROM Venda WHERE Dia BETWEEN 10 AND 16;`
 O que muda em cada um dos histogramas?

Exercício 4) Considere os dois seguintes comandos que podem ser emitidos em Oracle:

```
ANALYZE TABLE Tabela
ESTIMATE SYSTEM STATISTICS SAMPLE 10 PERCENT
FOR TABLE SIZE 20;
```

```
ANALYZE TABLE Tabela
ESTIMATE STATISTICS SAMPLE 10000 ROWS
ALL INDEXED COLUMNS SIZE 25
INTO Tabela-Estats
```

- Descreva o que cada um faz, ressaltando as diferenças entre eles.
- Assumindo que a cardinalidade de `Tabela` seja 150K, `PAGESIZE=2K` e `ROWSIZE=150`, qual dos dois gera estatísticas mais precisas? Qual é executado mais rapidamente? Qual gera mais acessos a disco?

Exercício 5) Quais são as tabelas/visões que guardam informações sobre estatísticas em:

- Oracle

- Postgres

Como são guardados os histogramas em cada uma dessas estruturas?

— 3 de maio de 2013—