

OTIMIZAÇÃO DE INTERROGAÇÕES - CANDIDATOS

OBJETIVOS DO TRABALHO

Analisar os planos de execução de diferentes interrogações SQL a uma BD de teste. Avaliar o impacto da existência de índices e de estatísticas e do recurso a diferentes estratégias de estruturação das perguntas.

GRUPOS DE TRABALHO

O trabalho deve ser realizado por grupos de um a três elementos.

TEMA

A BD a utilizar neste trabalho refere-se a dados sobre candidaturas e inscrições de alunos em licenciaturas, ao longo dos anos. As tabelas são: a tabela **lics**, que contém a lista de cursos de licenciatura; a tabela **cands** que regista as candidaturas; a tabela **alus** que contém as inscrições dos alunos e a tabela **anos** que contém a lista dos anos em que há matrículas. Os dados referem-se a todos os alunos que estavam a frequentar em 1992 ou se matricularam posteriormente, tendo sido obtidos no final de 2002, já com as matrículas de 2002/2003 feitas.

Antes de um aluno ser matriculado, é registada a sua candidatura a um determinado curso, num dado ano lectivo. O aluno pode candidatar-se a mais do que um curso. Nessa fase, é identificado pelo BI, uma vez que só adquire número de aluno ao ser matriculado. Relativamente a cada candidatura, sabe-se o curso pretendido, o resultado e, nalguns casos, a média de candidatura. Nem todos os candidatos aceites acabam por se matricular. Uma vez matriculado, o aluno tem um estado que indica se está a frequentar ou se terminou o curso.

O esquema da base de dados é o seguinte:

- A tabela **lics** contém a lista de cursos de licenciatura, com as seguintes colunas:
 - **codigo** NUMBER(4) NOT NULL primary key,
 - **sigla** VARCHAR2(5) NOT NULL,
 - **nome** VARCHAR2(80) NOT NULL
- A tabela **cands** regista as candidaturas, com as seguintes colunas:
 - **bi** VARCHAR2(12) NOT NULL,
 - **curso** NUMBER(4) NOT NULL references lics,
 - **ano_lectivo** NUMBER(4) NOT NULL,
 - **resultado** VARCHAR2(1), -- C – aceite; E – excluído; S - suplente
 - **media** NUMBER(5,2),
 - **PRIMARY KEY** (bi, curso, ano_lectivo)
- A tabela **alus** tem os alunos que se matricularam e informação sobre a evolução do curso. Note que, se uma pessoa mudar de curso, vai ter dois números de aluno diferentes, embora o mesmo BI. Os anos lectivos são representados pelo número do ano civil inicial, isto é 1999 significa “1999/2000”. As colunas são:
 - **numero** VARCHAR2(9) primary key,
 - **bi** VARCHAR2(12) NOT NULL,
 - **curso** NUMBER(4) references lics,
 - **a_lect_matricula** NUMBER(4) NOT NULL,
 - **estado** VARCHAR2(2), -- F – frequentar; C – concluído
 - **a_lect_conclusao** NUMBER(4),
 - **med_final** NUMBER(4,2)
 - **FOREIGN KEY** (bi, curso, a_lect_matricula) REFERENCES cands
- A tabela **anos** tem a lista de anos em que há matrículas
 - **ano** NUMBER(4).

DADOS

As tabelas encontram-se no utilizador GTD2, de onde devem ser copiadas, a triplicar, para as tabelas xalus, xlics, ..., yalus, ylics, ..., zalus, zlics, O objetivo é criar três ambientes de experimentação:

X - Sem índices nem restrições de integridade.

Y - Com as restrições de integridade standard (chave primária e chave externa).

Z - Com as restrições de integridade standard e os índices extra que achar convenientes, justificando a razão de inclusão de cada um deles.

MÉTODO

0) Comece com a justificação das restrições (Y) e dos índices extra (Z).

Relativamente a cada questão abaixo, deve escrever:

- 1) A pergunta SQL;
- 2) A resposta;
- 3) Uma análise dos três planos de execução nos três ambientes e do esforço estimado correspondente;
- 4) Se o tempo de execução for mensurável, uma comparação dos tempos de execução nos três ambientes.

QUESTÕES A ANALISAR

1. Seleção e junção.
Qual a lista dos números dos alunos especiais, que terminaram o curso com sigla EIC em menos de cinco anos, e quantos anos demoraram.
2. Agregação.
Qual a média mínima de candidatura em cada curso, em cada ano, dos alunos matriculados? Nem todas as candidaturas têm a média preenchida.
 - a. Quais os cursos que têm alunos excluídos em todos os anos de 1995 a 2002? Indique o código e o nome do curso.
3. Considere a questão de saber quantos candidatos aceites não se matricularam nesse ano lectivo. Compare uma formulação que use uma subpergunta constante com a equivalente que use uma subpergunta variável (sugestão: usar EXISTS).
4. Estude as tentativas de resposta à questão “Qual o curso com a melhor média de conclusão em cada ano lectivo” apresentadas abaixo. Comente-as.
5. Compare os planos de execução da pesquisa “Quantos candidatos tiveram como resultado algo diferente de “C” ou “E”, usando, no contexto Z,
 - a. Com índice árvore-B em Resultado;
 - b. Com índice bitmap em Resultado.
6. A pergunta “Há, em algum ano algum curso (ano_lectivo, sigla e nome) que tenha todas as candidaturas aceites transformadas em matrículas, nesse mesmo ano?” é de natureza universal. Compare do ponto de vista temporal e de plano de execução as estratégias da dupla negação e da contagem.