

Tecnologia de Bases de Dados

Trabalho Prático nº 1

OTIMIZAÇÃO DE INTERROGAÇÕES

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. O relatório a produzir deve, para cada situação, indicar a formulação da pergunta em SQL, a resposta obtida, as condições de execução do teste incluindo duração se relevante, o plano de execução e as conclusões a que chegou.

GRUPOS DE TRABALHO

O trabalho deve ser realizado por grupos de um ou dois elementos.

TEMA

A base de dados do exemplo contém as respostas dos inquéritos pedagógicos até 2008/2009. As tabelas relevantes são:

- ipdw_respostas: tabela com as respostas individuais a cada pergunta para cada disciplina em cada questionário, usando valores de 1 a 5;
- ipdw_semestre: tabela com um registo para cada inquérito, num dado ano letivo e semestre;
- ipdw_pergunta: tabela com as perguntas utilizadas em cada inquérito, as quais variam um pouco de uns anos para os outros (nos últimos anos utilizaram-se as do tipo_de_inquerito 'Inq2006-31');
- ipdw_disciplina: tabela com todas as disciplinas.

DADOS

As tabelas encontram-se no utilizador GTD2, de onde devem ser copiadas, a triplicar, para as tabelas xrespostas, xsemestre, xpergunta, xdisciplina, yrespostas, ysemestre, ypergunta, ydisciplina, zrespostas, zsemestre, zpergunta e zdisciplina. 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

Pretende-se, relativamente às perguntas abaixo, quais as formulações SQL e as respetivas respostas, uma análise dos planos de execução dessas interrogações e qual o esforço estimado. As perguntas devem ser executadas nos três contextos.

QUESTÕES A ANALISAR

1. Seleção.

- Quantas são as respostas do inquérito de id 21 (2008/2009, 1S)?
- Quantas são as respostas relativas à disciplina de id 1237 (FPRO)?
- Quantas são as respostas relativas ao inquérito de 2008/2009, 1S?

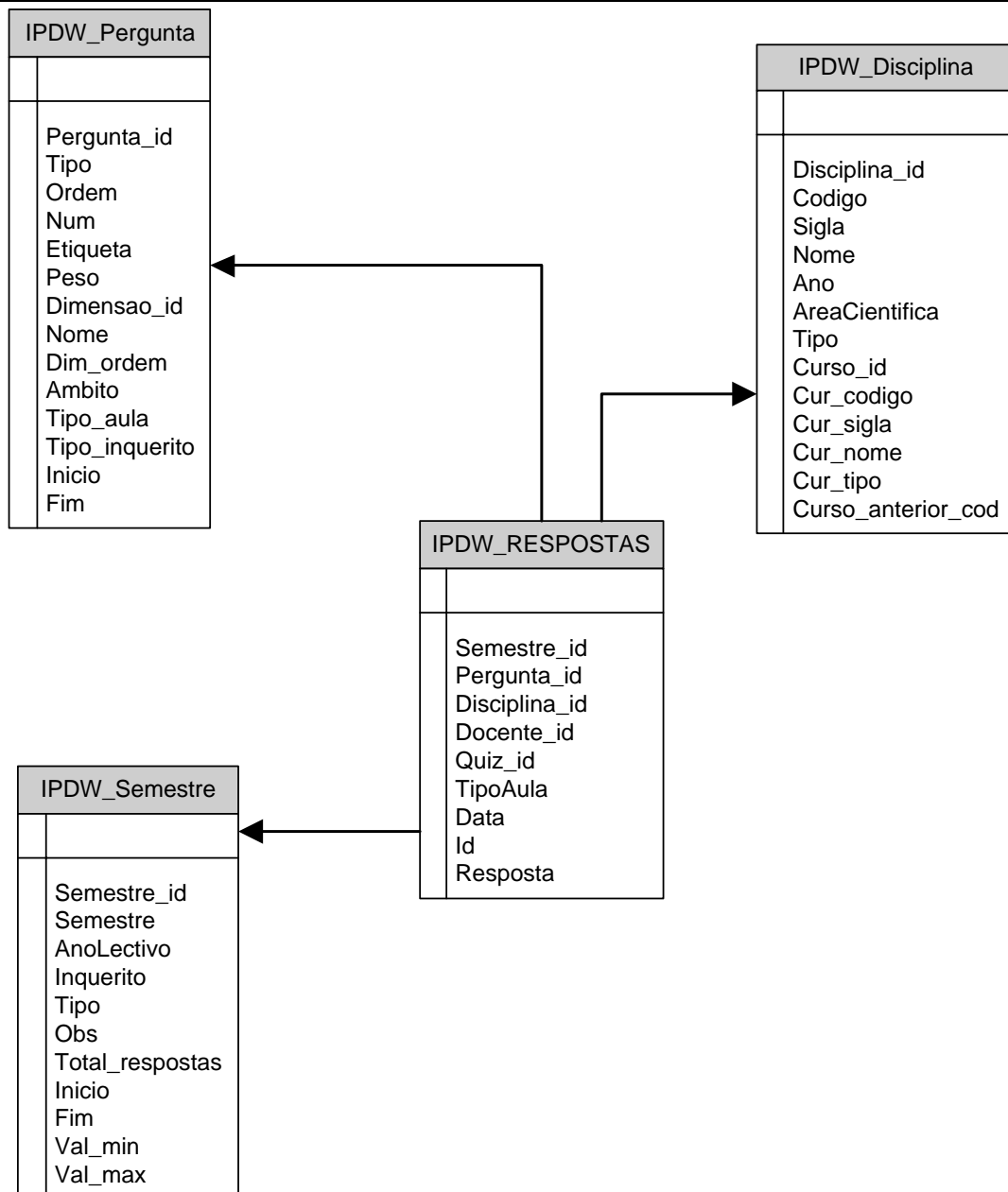


Figura 1: Esquema relacional.

2. **Agregação.** Faça uma análise semelhante à anterior, relativamente a perguntas que envolvem agregações ou cláusula having.
 - a. Qual a média dos resultados dos vários inquéritos à disciplina FPRO?
 - b. Qual o id e a sigla da disciplina que, no inquérito de id 21, teve o melhor resultado, das que tiveram mais do que 300 respostas?
 - c. Qual o id e a sigla das disciplinas que têm uma média da pergunta ‘Apreciação Global’ superior em mais do que 0.1 à média de todas as suas respostas, no inquérito de id 21 e entre as que tiveram mais de 300 respostas?
3. **Negação.** Analise de igual forma a questão “Quais as disciplinas que não têm qualquer resposta registada?” Compare duas soluções equivalentes, uma baseada em “not in” e outra baseada em junção externa e filtragem para nulo. Verifique a influência dos modos do optimizador (CHOOSE e RULE) nos resultados obtidos.
4. **Quantificação universal.** Analise a pergunta “Quais os id e as siglas das disciplinas que tiveram pelo menos um 5 a cada uma das perguntas do inquérito de id 21?” Compare do ponto de vista de esforço e de plano de execução as estratégias da dupla negação e da contagem.