

Administração de Bases de Dados

Engenharia Informática – Universidade do Minho

Trabalho Prático – 2020/2021

O resultado do trabalho é um relatório escrito. O relatório deve omitir considerações genéricas sobre as ferramentas utilizadas, focando a apresentação e justificação dos objetivos atingidos. O relatório deve ser entregue por apenas um dos elementos do grupo, identificando também o nome e número de todos os outros na capa. A entrega do relatório é feita na área da Unidade Curricular no *e-Learning*. A data limite é 15 de janeiro de 2020.

1 Contexto

O trabalho prático consiste na configuração, otimização e avaliação do *benchmark* TPC-C com alguns dados e interrogações adicionais. O TPC-C simula um sistema de bases de dados de uma cadeia de lojas, suportando a operação diária de gestão de vendas e stocks. As interrogações analíticas adicionais, baseadas na adaptação do TPC-H, poderão ser corridas ocasionalmente. A descrição dos *benchmarks* e da respetiva implementação está disponível no Anexo A.

2 Objetivos

1. Instalar e configurar o benchmark *TPC-C* usando um máquina virtual "n1-standard-2" para o servidor PostgreSQL.
2. Propôr uma configuração de referência em termos de número de *warehouses* e número de clientes.
3. Escolher e adaptar n interrogações analíticas (1 por cada elemento do grupo) e otimizar o seu desempenho tendo em conta, principalmente, os respetivos planos e os mecanismos de redundância que estão a ser usados. <https://db.in.tum.de/research/projects/CHbenCHmark/index.shtml?lang=en>, usar datas recentes
4. Otimizar o desempenho da carga transacional tendo em conta, principalmente, os parâmetros de configuração do PostgreSQL.

3 Notas

- Como medidas de desempenho da carga operacional (TPC-C) deve considerar-se principalmente o **debito máximo** atingível. Nas operações analíticas, deve considerar-se o **tempo de resposta**.
- Poderão modificar as interrogações SQL e o código Java para atingir os objetivos 2 e 3. Devem explicar no relatório em que medida essas alterações preservam o funcionamento da aplicação original.
- Não poderão considerar todas as otimizações possíveis nas suas várias combinações... Devem focar-se nas que consideram mais prometedoras e que mais vos interessam, justificando no relatório essas opções.

2) Configurar o numero de warehouses e clientes que usam, bd nem de 100mb nem 6gb. 3/4gb de BD
Numero de clientes: threads por transacao, nem 1% de cpu nem 90%. Carga razoavel 30/40/50 e dimensional em termos de disco

- É aceitável que não consigam uma melhoria de desempenho em relação à configuração base. Nesse caso, devem explicar porque é que a configuração de referência já era ótima. Por outro lado, a simples apresentação de uma melhoria, não justificada, não é muito interessante.
- A utilização de ferramentas de monitorização e diagnóstico do PostgreSQL e do sistema operativo (e.g., pgbadger, iostat) valoriza o trabalho.
- A automatização da instalação e execução do *benchmark* permitirá obter resultados em maior quantidade, permitindo uma análise mais profunda e valorizando o trabalho.
- Devem também procurar estratégias para poupar recursos na *Google Cloud*, por exemplo, armazenando os dados em *Cloud Storage* e reutilizando-os, em vez de re-executar o `load.sh`. Tenham o cuidado de escolher as configurações mais baratas na *Google Cloud* (i.e., regiões nos EUA, *preemptible*, ...) e de não deixar máquinas virtuais ativas a consumir recursos desnecessariamente!

A Referências

1. Código fonte e instruções de instalação:
<https://github.com/jopereira/EscadaTPC-C>
<https://gist.github.com/jopereira/4086237>
2. Interrogações analíticas:
<https://db.in.tum.de/research/projects/CHbenCHmark/index.shtml?lang=en>
3. Documentação adicional:
<http://www.tpc.org/tpcc/>
<http://www.tpc.org/tpch/>