

Administração de Bases de Dados

Engenharia Informática – Universidade do Minho

Trabalho Prático – 2018/2019

Os grupos de trabalho devem ser constituídos por 3 (três) elementos, todos inscritos na Unidade Curricular. Será disponibilizado saldo na *Google Cloud Platform* para realizar este trabalho a cada um dos grupos. 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 30 de dezembro de 2018.

1 Contexto

O trabalho prático consiste na configuração, otimização e avaliação do *benchmark* TPC-C. 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. Devem também ser consideradas algumas interrogações analíticas, baseadas na adaptação do TPC-H e listadas no Anexo A, que poderão ser corridas ocasionalmente. A descrição dos *benchmarks* e da respetiva implementação está disponível no Anexo B.

2 Objetivos

1. Instalar e configurar o benchmark *TPC-C* obtendo uma **configuração de referência** em termos de: número de *warehouses*, número de clientes e *hardware*.
2. Usando a configuração de referência, otimizar ou justificar o desempenho tendo em conta as interrogações, os respectivos planos, os mecanismos de redundância que estão a ser usados e os parâmetros de configuração do PostgreSQL.
3. Propôr e testar uma configuração usando replicação, *sharding* e/ou processamento distribuído com qualquer uma das ferramentas estudadas. Esta configuração deve atingir uma escala superior à da configuração de referência, com a qual deve ser comparada.

3 Notas

- Como medidas de desempenho da carga operacional (TPC-C) deve considerar-se principalmente o débito máximo atingível. Nas operações analíticas, deve considerar-se o tempo de resposta.
- Ao escolher a configuração de referência, devem ter em conta o saldo disponível na *Google Cloud* para efetuar todos os testes e o tempo que cada teste demora a correr. Se escolherem uma configuração demasiado grande, não conseguirão fazer todos os testes necessários. Por outro lado, se escolherem uma configuração demasiado pequena, vão ter poucas oportunidades de otimização.

- 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.
- Nos objetivos 2 e 3 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.
- Nos objetivos 2 e 3 é aceitável que não consigam uma melhoria de desempenho. Nesse caso, devem explicar porque é que a configuração de referência já era ótima. Por outro lado, a simples apresentação de uma melhoria de desempenho, 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 Interrogações analíticas

A.1

```
select  sum(ol_amount) / 2.0 as avg_yearly
from    order_line, (select  i_id, avg(ol_quantity) as a
                    from    item, order_line
                    where    i_data like 'b%'
                    and ol_i_id = i_id
                    group by i_id) t
where   ol_i_id = t.i_id
        and ol_quantity < t.a;
```

A.2

```
select  c_count, count(*) as custdist
from    (select c_id, count(o_id)
        from customer left outer join orders on (
            c_w_id = o_w_id
            and c_d_id = o_d_id
            and c_id = o_c_id
            and o_carrier_id > 8)
        group by c_id) as c_orders (c_id, c_count)
group by c_count
order by custdist desc, c_count desc;
```

A.3

```
select  sum(ol_amount) as revenue
from    order_line, item
where   (
        ol_i_id = i_id
        and i_data like 'a%'
        and ol_quantity >= 1
        and ol_quantity <= 10
        and i_price between 1 and 400000
        and ol_w_id in (1,2,3)
    ) or (
        ol_i_id = i_id
        and i_data like 'b%'
        and ol_quantity >= 1
        and ol_quantity <= 10
        and i_price between 1 and 400000
        and ol_w_id in (1,2,4)
    ) or (
        ol_i_id = i_id
        and i_data like 'c%'
        and ol_quantity >= 1
        and ol_quantity <= 10
        and i_price between 1 and 400000
        and ol_w_id in (1,5,3)
    );
```

A.4

```
select  substr(c_state,1,1) as country,
        count(*) as numcust,
        sum(c_balance) as totacctbal
from    customer
where   substr(c_phone,1,1) in ('1','2','3','4','5','6','7')
        and c_balance > (select avg(c_BALANCE)
                          from  customer
                          where  c_balance > 0.00
                          and substr(c_phone,1,1) in ('1','2','3','4','5','6','7'))
        and exists (select *
                    from  orders
                    where  o_c_id = c_id
                          and o_w_id = c_w_id
                          and o_d_id = c_d_id)
group by substr(c_state,1,1)
order by substr(c_state,1,1);
```

B Referências

1. Código fonte e instruções de instalação:

<https://github.com/jopereira/EscadaTPC-C>

<https://gist.github.com/jopereira/4086237>

2. Documentação:

<http://www.tpc.org/tpcc/>

<https://db.in.tum.de/research/projects/CHbenCHmark/index.shtml?lang=en>