

Engenharia de Aplicações

Frequência¹

14 de Fevereiro de 2011

Duração: 2h00min

I

- 1 Diga o que é um *clustered index*. Dê exemplos de interrogações em que a sua existência seja/não seja vantajosa.
- 2 Descreva o método de *RE-DO logging*. Porque é que normalmente um SGBD não usa **apenas** este método?
- 3 Quais a vantagem/desvantagem principal de um SGBD multi-versão vs. um SGBD que faça uso de *2 phase locking*? Justifique e dê exemplos.
- 4 Identifique o método que pode ser usado para gerir uma quantidade de dados de grandes dimensões. Quais os principais obstáculos à aplicação deste técnica?

II

- 1 Normalmente no desenho de uma arquitectura computacional de elevada disponibilidae, existe um conjunto de serviços por exemplo DHCP, DNS, NTP, LDAP, que não são explicitamente referenciados mas que são muitas vezes vitais para o correcto funcionamento da infra-estrutura. Descreva os serviços e discuta a veracidade da afirmação.
- 2 Tendo em atenção o enunciado do grupo III, proponha uma solução adequada de storage (volumes de dados, tipo de redundância e sistema de ficheiros). Tenha em consideração as questões de disponibilidade da base de dados e que os ficheiros de imagens podem ser acedidos simultaneamente por vários servidores.
- 3 Com a evolução da aplicação, foi decidido utilizar aplicações de *transcoding* que geram para cada imagem várias versões em vários tamanhos e que são apenas guardadas em disco numa cache (i.e. estas imagens podem ser apagadas se for necessário espaço). Apesar da utilização deste sistema de cache e devido ao elevado número de acessos os tempos de resposta são muito elevados. Proponha reformulações a nível da storage que permitam minimizar este problema, isto é, permitam aumentar o desempenho da storage, mantendo o número de discos existentes. A solução que propõe pode ser implementada sem obrigar a interrupção de serviço?

¹Cotação — (7) + (8) + (3 + 2)

III

Considere a base de dados de uma loja online, que inclui uma tabela ITEMS com os produtos disponiveis e uma linha por cada item vendido na tabela VENDAS.

Considere ainda que a cada item estão associadas imagens com diferentes resoluções diferentes para utilização em diferentes tipos de média, e que a grandeza do número de imagens é da ordem das dezenas de milhões.

1 Considere a seguinte interrogação usada para calcular o top 10 de vendas em diferentes períodos:²

```
select i.nome from VENDAS v inner join ITEMS i on i.id=v.id where  
v.data > ... group by v.id , i.nome order by sum(v.quant) desc limit 10;
```

1.1 Diga qual o plano de execução que espera que seja utilizado. Diga quais os pressupostos de que partiu.

1.2 Admita que esta interrogação está muito lenta ao calcular o top 10 relativo ao dia de hoje. O que recomenda?

1.3 Admita que esta interrogação está muito lenta ao calcular o top 10 de vendas do mês actual. O que recomenda?

2 Proponha uma arquitectura de elevada disponibilidade e escalável para o serviço em causa.

²Responda em alternativa à realização do trabalho prático sobre TPC-W.