GBC053–Gerenciamento de Banco de Dados

Armazenamento de Dados

RAID e Gerência de Espaço em Disco

Ilmério Reis da Silva
ilmerio arroba ufu.br
MS Teams: GBC053.2021.2
UFU/FACOM

Armazenamento de Dados - Desempenho de Discos – RAID Motivação

Acesso paralelo às trilhas de um cilindro é de difícil sincronização, uma solução para prover paralelismo e maior confiabilidade em discos é a Tecnologia RAID.

RAID - Redundant Arrays of Independent Disks ou Conjunto Redundante de Discos Independentes

Uma tecnologia para acesso a múltiplos discos!

Armazenamento de Dados - Desempenho de Discos – RAID Espelhamento e Espalhamento

- ***Melhoria da confiabilidade por meio da redundância (Espelhamento-Mirroring).***
- ***Melhoria do desempenho por meio do paralelismo (Espalhamento-Striping)***

Armazenamento de Dados - Desempenho de Discos – Espalhamento em RAID

- ***Espalhamento melhora desempenho***
  - partições de mesmo tamanho distribuídos em discos
  - Para D discos a partição i é escrita no disco (i mod D)
  - Permite leitura em paralelo
  - Partição pode ser por bit ou bloco

Armazenamento de Dados - Desempenho de Discos – Espelhamento/Bit de paridade RAID

- ***Redundância melhora a confiabilidade***
  - Espelhamento ou
- ***Discos de dados com espalhamento + disco de verificação com bit de paridade:***
  - permite reconstrução de discos com falha, por exemplo:
    - ✓Paridade 1 sse número de 1's é ímpar
    - ✓bit do disco que falhou é inferido pelo valor do bit de paridade

Armazenamento de Dados - Desempenho de Discos – NÍVEIS em RAID

- ***RAID nível 0 espalhamento nível de bloco, sem qualquer redundância => melhora write; diminui confiabilidade***
- ***RAID nível 1 espelhamento => melhora confiabilidade***
- ***RAID nível 0 + 1 espelhamento e espalhamento (RAID10)***
- ***RAID nível 2 espalhamento com bits de paridade => melhora confiabilidade***
- ***RAID nível 3 espalhamento por bit com bits de paridade para correção de erro de uma forma otimizada => indentifica disco que falhou***
- ***RAID nível 4 espalhamento por bloco com bits de paridade de uma forma otimizada => explora melhor paralelismo***
- ***RAID nível 5 espalhamento por bloco combinado com bits de paridade distribuídos => elimina gargalo***
- ***RAID nível 6 semelhante ao Raid nível 5, mas armazena informações redundantes extras para proteger contra múltiplas falhas de disco.***

Armazenamento de Dados - Desempenho de Discos – NÍVEIS em RAID - ILUSTRAÇÃO

C = Cópia de Dados

P = Bits de Paridade

Armazenamento de Dados - Desempenho de Discos – NÍVEIS em RAID e suas Indicações

- ***RAID nível 0 caso não haja problemas com perdas***
- ***RAID nível 0 + 1 para pequeno volume de dados e muita gravação (write)***
- ***RAID nível 2 e 4 não são utilizados, pois 3 e 5 substituem***
- ***RAID nível 3 grandes transferências de blocos contíguos***
- ***RAID nível 5 genérico com bom desempenho médio***
- ***RAID nível 6 sistemas que necessitam alta confiabilidade***

Armazenamento de Dados - Desempenho de Discos - MTTF

MTTF (mean-time-to-failure)

- Exemplo de MTTF em um disco: 50000 horas (5,7 anos)
- Em 100 desses discos : 50000/100 horas (21 dias)
- Usando 10 discos de verificação podemos melhorar a MTTF do sistema:

(100 discos de dados + 10 de verificação) > 250anos,

pois o sistema falha se houver falha simultanea de um disco de dados e de um disco de verificação:

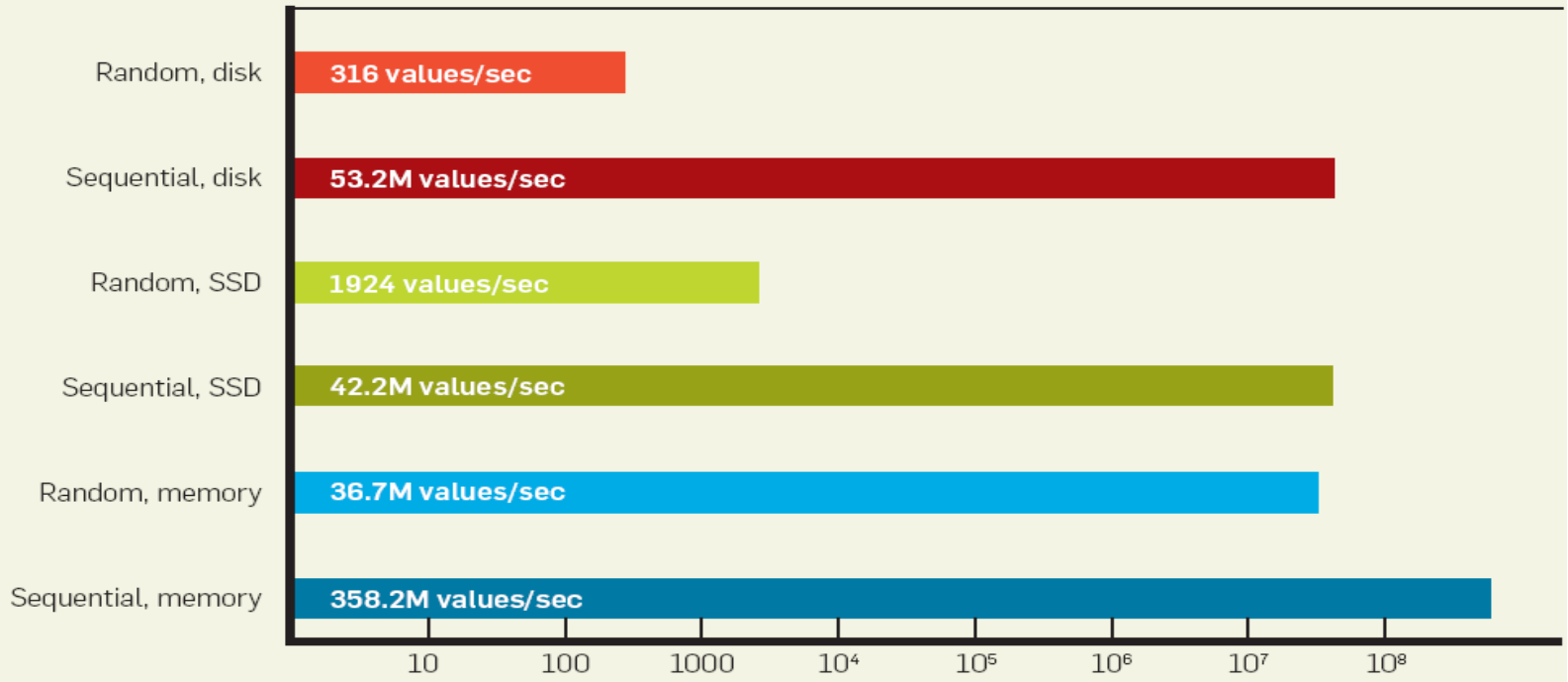
(50000/100) * (50000/10)=2.500.000 horas

Armazenamento de Dados - Desempenho de Discos - MTTR

- ***A tecnologia RAID 5 recupera falha de um disco***
- ***A tecnlogoia RAID 6 pode recuperar falha de mais de um disco***
- ***Entretanto, para calcular precisamente a nova confiabilidade precisamos definir o tempo de reparo do disco (MTTR-Mean Time To Repair)***

Armazenamento de Dados - Desempenho de Discos, Exemplo HD – SSD - RAM

Figure 3. Comparing random and sequential access in disk and memory.



* Disk tests were carried out on a freshly booted machine (a Windows 2003 server with 64GB RAM and eight 15,000RPM SAS disks in RAID5 configuration) to eliminate the effect of operating-system disk caching. SSD test used a latest generation Intel high-performance SATA SSD.

*** Jacobs, A. ,"The Patologies of Big Data", CACM, V.52, N.8, August, 2009**

Gerência de espaço em disco

Armazenamento de Dados – Gerência de Espaço em Disco

- ***Página ou bloco é a unidade de acesso definida pelo software, no caso o SGBD***
- ***Otimização de acesso sequencial é feita por meio de alocação de blocos contíguos (mesma trilha, mesmo cilindro, cilindros adjacentes)***
- ***Modificações podem criar espaços livres***
- ***Gerência de espaços livres pode ser por lista de blocos livres ou bitmap***

Armazenamento de Dados – Gerência de Espaço em Disco

- ***Quem gerencia o espaço?***
  - Sistema operacional ou sistema de arquivos; ou
  - Camada de baixo nível do SGBD
    - ✓ dá maior portabilidade ao sistema e melhora gerência de *buffer pool* (próxima seção)
  - Gerência compartilhada (SO + SGBD)
  - Deixando a alocação física de páginas para camadas de baixo nível, podemos trabalhar com a seguinte abstração:
    - ✓ Arquivo: array de bytes (ou de páginas)
    - ✓ Solicitação: acesso byte *i* (ou página *i*) do arquivo *f*
    - ✓ Execução pelas camadas de baixo nível: acesso ao bloco *m* da trilha *t* do cilindro *c* no disco *d*

FIM - RAID e Gerência de Espaço em Disco