

Segurança Informática e nas Organizações

Resumos
2016/2017

João Alegria | 68661

Capítulo 10

Armazenamento da Informação

Problemas

- Os discos avariam
 - E cada vez há mais informação digital vital
 - É preciso minimizar a falta de discos ou a perda de informação
- O acesso mecânico à informação é lento
 - $\text{tempo} = \text{tempo de translação} + \text{tempo de rotação}$

↓ ↓
relativo à agulha relativo ao disco

- Mais informação -> maior estrangulamento
- São precisos discos mais eficientes
- Alternativas de solução
 - cópias de segurança (*backups*) - locais ou remotas
 - RAID

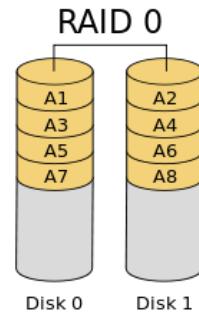
Objetivos

- Garantir a sobrevivência da informação
 - Os dados só se perdem se falharem mais do que X discos do RAID, onde o valor de X depende do tipo de RAID
- Solução de baixo custo e eficiente
 - Permite usar hardware barato e fálgivel
 - Acelerar o desempenho nas leituras e escritas em discos
- O RAID não substitui o backup!
 - Não tolera falhas catastróficas em mais do que X discos em N
 - Não tolera erros dos utentes ou do sistema
- O RAID pode aumentar a probabilidade de falha do sistema!
 - Se o objetivo for apenas acelerar o mesmo

RAID 0 (*striping*)

Se um disco falhar, falha tudo!

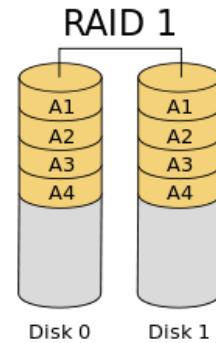
- Objetivo: Acelerar o acesso à informação em disco
- Aproximação: Acesso a discos em paralelo
 - Stripping:
 - A informação lógica de um volume é subdividido em fatias (*stripes*)
 - As fatias são intercaladas nos discos
 - Não há desperdício de informação (não aumenta em razão direta)
- Prós: Aceleração dos acessos aos discos até N vezes
 - Não é necessário percorrer o disco todo, basta aceder a partes dos N discos
- Contras: Maior probabilidade de perda de informação
 - Se P_F for a probabilidade de falha num disco, a probabilidade de perder informação com N discos é $1 - (1 - P_F)^N$
 - Aumento do número de dispositivos, pelo menos para o dobro



RAID 1 (*mirroring*)

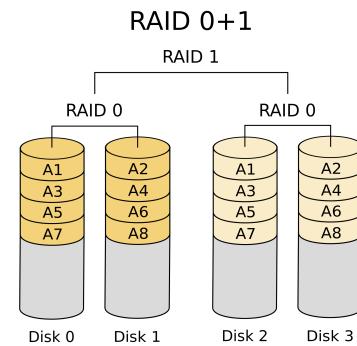
Se um disco falhar, os dados podem ser recuperados porque existem cópias

- Objetivo: Tolerar a falha de discos
- Aproximação: Acesso a discos em paralelo
 - Duplicação da informação (*mirroring*)
 - A mesma informação nos N discos, escrita sincronizada
 - Na leitura é feita uma verificação de coerência (igualdade)
 - O sistema RAID1 deve funcionar aos pares de forma a que haja um clone
 - Aumento do desperdício, tudo o que é colocado é redundância (aumenta na razão direta)
 - Prós: Diminuição da probabilidade de perda de informação
 - Se P_F for a probabilidade de falha de um disco, a probabilidade de perder informação com N discos é $(P_F)^N$
 - Contra: Desperdício da capacidade de armazenamento, mais de 50% de capacidade total de armazenamento $\left(\frac{N-1}{N}\right)$
 - Sistema que apresenta maior desperdício de espaço de armazenamento
 - Aumento do número de dispositivos, no mínimo para o dobro



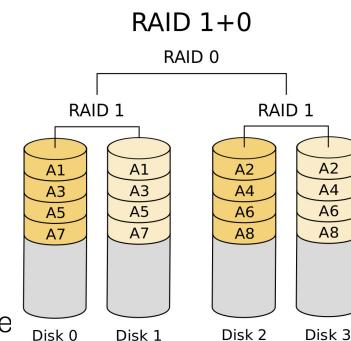
RAID 0+1

- Objetivo: Benefícios do RAID 0 (desempenho)
Benefícios do RAID 1 (tolerância a falhas)
- Aproximação: RAID 0 num primeiro nível
RAID 1 num segundo nível
 - *Mirroring de volumes striped*
 - É um sistema híbrido!
 - Combina RAID 0 com RAID 1
 - O sistema precisa de ter pelo menos 4 unidades de armazenamento, duas para cada nível
 - RAID 0+1 considera o aspeto do desempenho e redundância
- Vantagens: Maior desempenho graças ao paralelismo
Proteção contra perdas graças ao RAID 1
- Desvantagens: Desperdício de capacidade de armazenamento, mais de 50% da capacidade total de armazenamento $\left(\frac{N-1}{N}\right)$
 - Aumento do número de dispositivos, no mínimo para o dobro



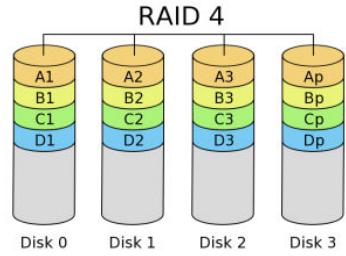
RAID 1+0

- Objetivo: Semelhante ao 0+1
Inverte *stripping* e *mirroring*
- Aproximação: RAID 1 num primeiro nível
RAID 0 num segundo nível
 - *Stripping de volumes sobre volumes mirrored*
- Em caso de falha
 - RAID 0+1 - O sistema transforma-se em RAID 0
 - RAID 1+0 - O sistema assume o nível RAID 1
- Desvantagem: Ligeiramente pior fiabilidade prática, se perder-se todo um subsistema com RAID 1



RAID 4

- Objetivo: Ter proteção do RAID 1, tendo um desempenho e um aproveitamento de espaço próximo do RAID 0
- Aproximação: Armazenamento de dados em N-1 discos
 - Há um disco reservado para a paridade
 - Desperdício de espaço é igual à capacidade de cada disco
 - Os dados de quaisquer N-1 discos geram um outro
- Problemas: Necessita de 3 ou mais discos
 - A atualização da paridade é complexa e demorada
 - A recuperação é mais demorada do que com o RAID 1
 - Se falhar dois ou mais discos, o sistema falha e perde informação $1 - (1 - P_F)^N$
 - Se falhar um disco, o sistema perde informação e recupera-a $1 - (1 - P_F)^N - N(P_F(1 - P_F)^{N-1})$



RAID 5

- Objetivo: Semelhante ao RAID 4, mas mais eficiente nas escritas
- Aproximações: Blocos de paridade dispersos por todos os discos
 - O desperdício de espaço é igual ao do RAID 4 (igual à capacidade de cada disco)
 - A concorrência nas escritas é melhorada
 - Em vez de existir uma unidade de armazenamento inteira como réplica, os próprios discos servem de proteção
 - Os dados são divididos em pequenos blocos, cada um deles recebe um bit de paridade
 - As informações de paridade, assim como os próprios dados, são distribuídos entre todos os discos do sistema
 - O espaço destinado à paridade é equivalente ao tamanho de um dos discos
- Problemas: Mais complexo que o RAID 4
 - Necessita de 3 ou mais discos
 - Se falhar dois ou mais discos, o sistema falha e perde informação $1 - (1 - P_F)^N$
 - Se falhar um disco, o sistema perde informação e recupera-a $1 - (1 - P_F)^N - N(P_F(1 - P_F)^{N-1})$

