

Arquitetura e Organização de Computadores II

Sistemas de Armazenamento

Prof. Nilton Luiz Queiroz Jr.

Sistemas de armazenamento

- Com a popularidade de serviços para internet a necessidade de melhorar o desempenho de I/O se tornou muito importante;
 - O foco em armazenamento de informação acabou crescendo se comparado ao foco em computar dados;
- Questões como confiabilidade, escalabilidade e custo-performance foram enfatizadas;
- Em geral, a dependência causada pelos sistemas de armazenamento é a maior entre o usuário e o sistema computacional;
 - Perder os dados armazenados no computador sem dúvida impacta mais um usuário do que um *crash* no sistema

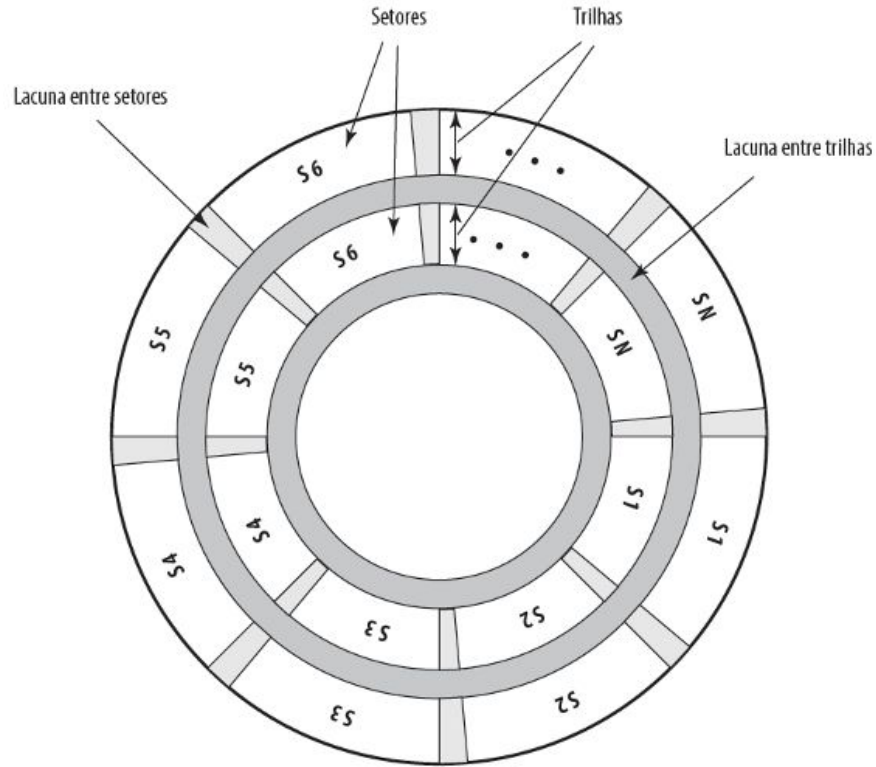


Discos magnéticos

- Discos magnéticos são a tecnologia mais comum no armazenamento secundário;
- São pratos circulares construídos de material não magnético, cobertos com um material magnetizável;
 - São divididos em trilhas e setores;
- Alguns discos possuem mais de um prato;
 - Diversos pratos que armazenam dados próximos na mesma trilha de cada prato;
 - Cada conjunto de setores que estão posicionados sob o braço de leitura e escrita ao mesmo instante é chamado de cilindro;



Discos magnéticos



Discos magnéticos

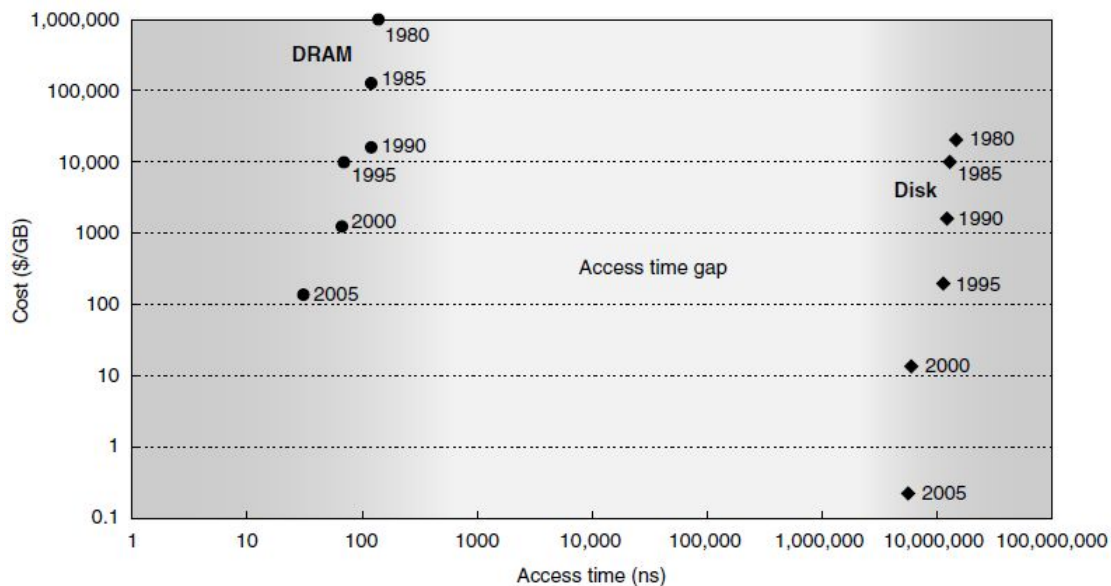
- Historicamente o desafio em produção de discos é incrementar a capacidade de armazenamento por área;
 - Quanto maior a densidade por área maior a capacidade de armazenamento;
 - A densidade está ligada ao tamanho da trilha;

$$\text{Areal density} = \frac{\text{Tracks}}{\text{Inch}} \text{ on a disk surface} \times \frac{\text{Bits}}{\text{Inch}} \text{ on a track}$$



Discos magnéticos

- O tempo de acesso de um disco magnético é muito maior que o tempo de acesso da memória ram;



Discos magnéticos

- Outro problema para discos é a largura de banda:
 - Discos rápidos, em 2011, de aproximadamente 600GB transferiram 200MB/s;
 - Memórias DRAM de 4GB transferiam 16000MB/s
- Assim existem dois grandes gaps entre memórias do tipo DRAM e discos;
 - Largura de banda;
 - Tempo de acesso;
- Alguns projetos tentaram preencher esse gap;
 - Memória flash chegou perto da questão de tempo de acesso;
 - Cem até mil vezes mais rápido em tempo de acesso;
 - Largura de banda é a mesma;
 - Porém é limitada na quantidade de escrita;
 - Em geral limitadas a 1 milhão de escritas;



Discos magnéticos

- O modelo setor-trilha-cilindro tende a desaparecer;
 - Inicialmente os discos começaram a oferecer interfaces inteligentes com um microprocessador dentro do disco;
 - Organização do disco com arquivos de maneira sequencial e não distribuídos aleatoriamente;
 - Depois dos microprocessadores nos discos começaram a surgir buffers e caches;
 - Buffers para manter dados até que o computador estivesse pronto para recebê-los;
 - Cache para evitar acessos a leitura;
 - Além disso ainda existia uma fila de comandos para que o disco pudesse decidir a melhor ordem para realizar operações para maximizar o desempenho sem alterar o comportamento das escritas e leituras;
 - O número de pratos reduziu de 12 para 4 ou até 1 em discos atuais;



Disk arrays

- Uma inovação que melhora tanto a questão de confiabilidade quanto a questão de escalabilidade em discos é o array de discos;
 - Também chamado de:
 - Redundant Array of Independent Disks - RAID;
- Dessa maneira mais de um disco é colocado para operar independentemente e em paralelo com outros;
 - Como existem vários discos, existem também várias maneiras de se organizar os dados;



RAID

- Existem vários níveis de RAID;
- Não são hierárquicos;
- Designam diferentes arquiteturas de projetos que compartilham características em comum:
 - São vistos como uma única unidade lógica pelo sistema operacional;
 - Dados distribuídos pelos discos em um esquema de interpolação de dados;
 - Pode-se armazenar informações de paridade;
 - Nem todo nível de raid aceita essa característica;



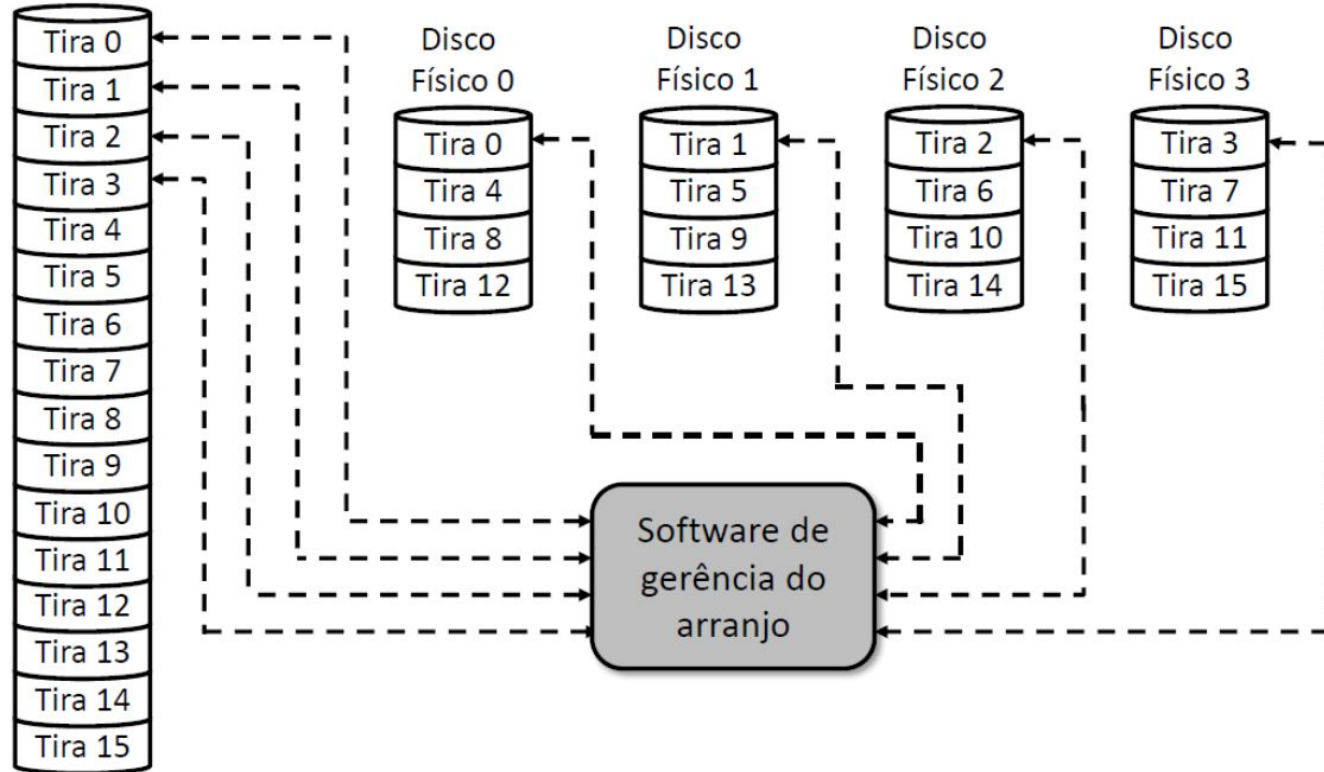
RAID 0

- Raid nível 0 não é um membro verdadeiro da família RAID;
 - Não dá suporte a redundância;
 - Apelidado de JBOD (Just a Bunch of Disks);
- Os dados do usuário são espalhados pelo sistema de armazenamento todo;
 - Em alguns casos onde mais de uma solicitação ao disco é feita, se todas estiverem em locais diferentes, ambas podem ser atendidas em paralelo;
- Os dados são armazenados de maneira intercalada entre os discos disponíveis (técnica também chamada de striping);
 - Todos dados são vistos em um único disco lógico dividido em tiras;
 - Desse modo, a tira i vai para o disco $i\%n$, onde n é o número de discos;



RAID 0

Disco lógico



RAID 0

- Vantagem:

- Maior velocidade;
 - Busca de dados em paralelo;
 - Um mesmo arquivo provavelmente será espalhado por diversos discos;
 - Múltiplas solicitações de dados provavelmente serão feitas em discos diferentes;

- Desvantagem:

- Sem proteção aos dados;

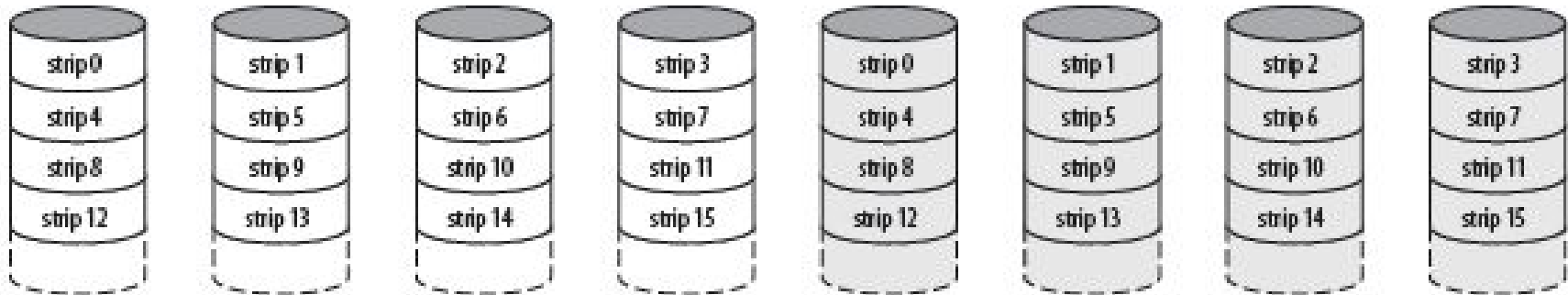


RAID 1

- Raid nível 1 faz cópias de todos os dados de maneira completa;
 - Também chamada de espelhamento;
- É o esquema de redundância de discos mais antigo;
- Tem o maior custo dentre todos os esquemas de RAID;
- É comum se usar striping de dados;
- Alguns controladores podem otimizar as leituras;
 - Os discos irão ler independentemente quando houver duas solicitações para “o mesmo disco”;



RAID 1



RAID 1

- Vantagens:

- Leituras podem ser atendidas por dois discos;
- Recuperação de falhas é simples;
- Não existe necessidade de cálculos de bit de paridade para gravação;

- Desvantagem:

- Muito espaço;
 - Dados todos duplicados;

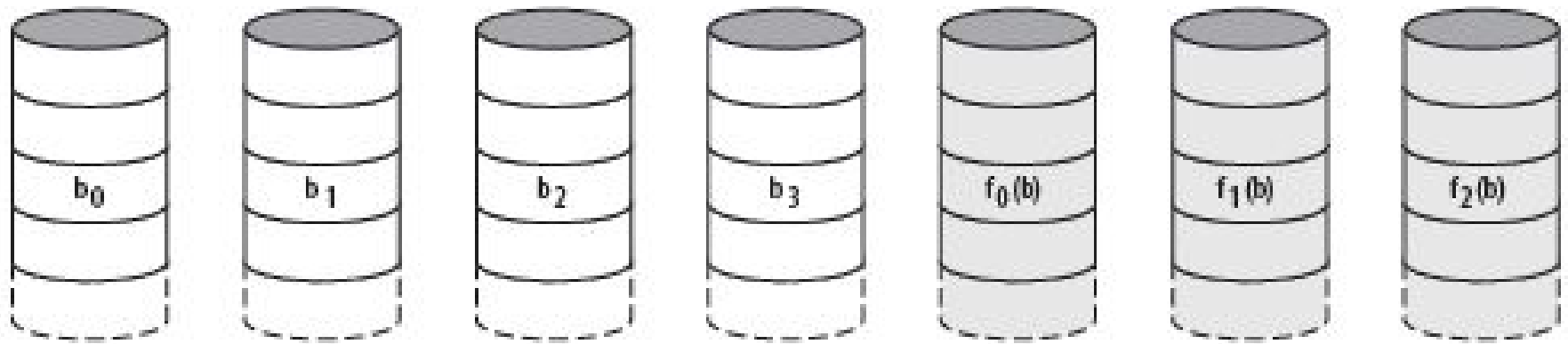


RAID 2

- Utiliza técnica de acesso paralelo;
- Todos discos membros participam na execução de cada solicitação E/S;
- Faz uso de stripping de dados;
 - Stripping pequenos (um byte ou uma palavra);
- Um código de correção de erro é calculado para os bits correspondentes em cada disco de dados;
 - Esses códigos são armazenados nos discos de paridade, cada bit em um disco;
 - Normalmente usa-se o código de HAMming;
- Geralmente não implementado atualmente;



RAID 2



RAID 2

- Vantagem:
 - Tem verificação própria de danos no disco;
- Desvantagem;
 - Por ser apenas a verificação de erros gasta muito espaço;

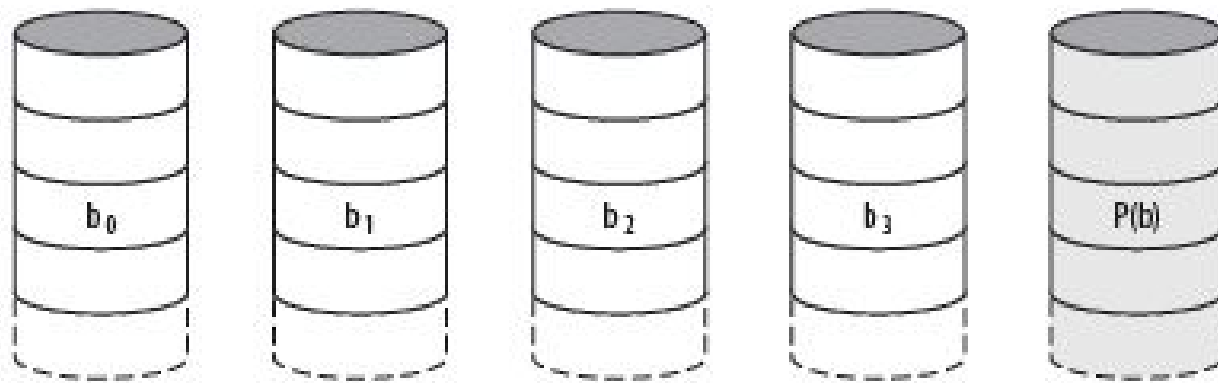


RAID 3

- O RAID 3 é uma variação do RAID 2, porém usando apenas um único disco de redundância;
- Um único bit de paridade é calculado para cada posição de todos os discos;
- Somente uma solicitação pode ser executada por vez;
- Alta velocidade em taxas de transferência;
- Pode recuperar dados de um disco que falou;
 - Quando um disco falha basta “subtrair” os bits dos outros discos;
- Assume que os dados estão espalhado entre todos discos;



RAID 3



RAID 3

- Vantagem:
 - Pouco overhead de verificação de bits;
 - Largura de banda alta para arquivos grandes;
- Desvantagem:
 - Não tem suporte para acessos:
 - Pequenos;
 - Aleatórios;

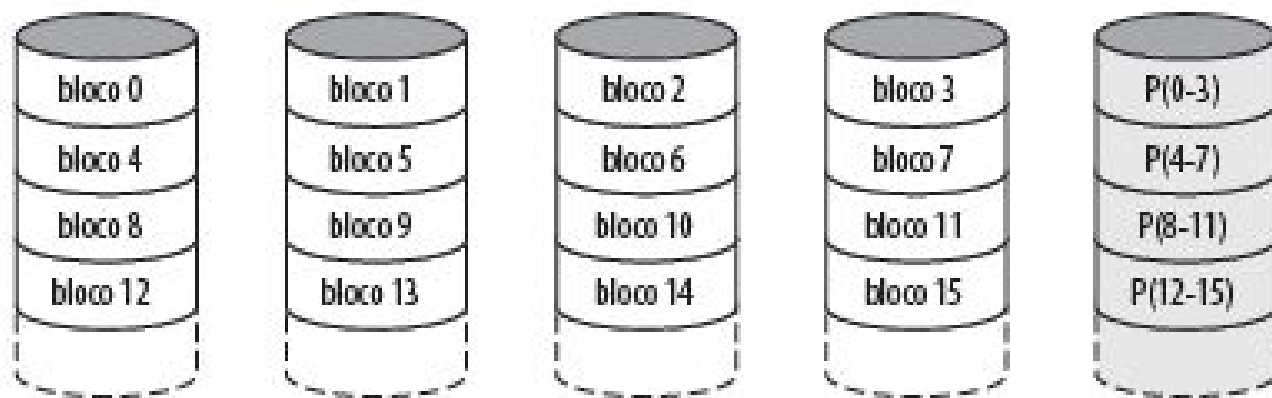


RAID 4

- Usa técnica de acesso independente, permitido solicitações de E/S paralelas;
- Utiliza stripping de dados;
- Um bit de paridade para cada strip;
- Armazenam todos os strips de paridade em um único disco, porém um bit para cada strip;
- Todas alterações envolvem o disco de paridade;
 - Gravação lenta



RAID 4



RAID 4

- Vantagem:
 - Pouco overhead de verificação de bits;
 - Mais largura de banda para pequenos acessos de leitura;
- Desvantagem:
 - Disco de paridade se torna um gargalo;
 - Escrita lenta;

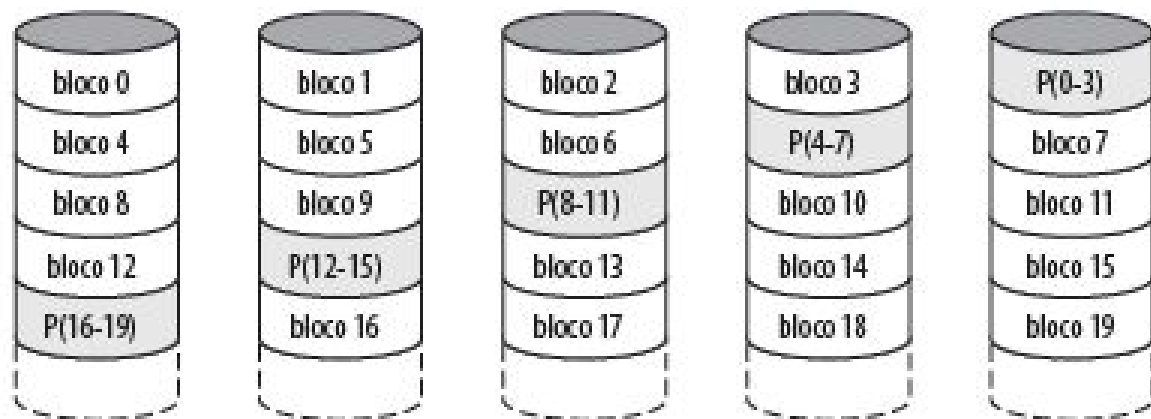


RAID 5

- O esquema de RAID 5 é semelhante ao RAID 4;
 - Resolve o problema do gargalo de escrita no RAID 4 distribuindo os strips de paridade;
- A principal diferença é que distribui o disco de paridade entre os demais existentes;
 - Em um esquema com n discos:
 - O strip de paridade dos discos 1 até $n-1$ está armazenado no disco n ;
 - O strip de paridade dos discos 1 até $n-2$ e do disco n está armazenado no disco $n-1$;
 - Assim por diante;



RAID 5



RAID 5

- Vantagens:
 - Pouco overhead de verificação de bits;
 - Mais largura de banda para pequenos acessos de leitura e escrita;
- Desvantagens:
 - Apesar de alta largura de banda, escritas pequenas acessam mais de um disco;

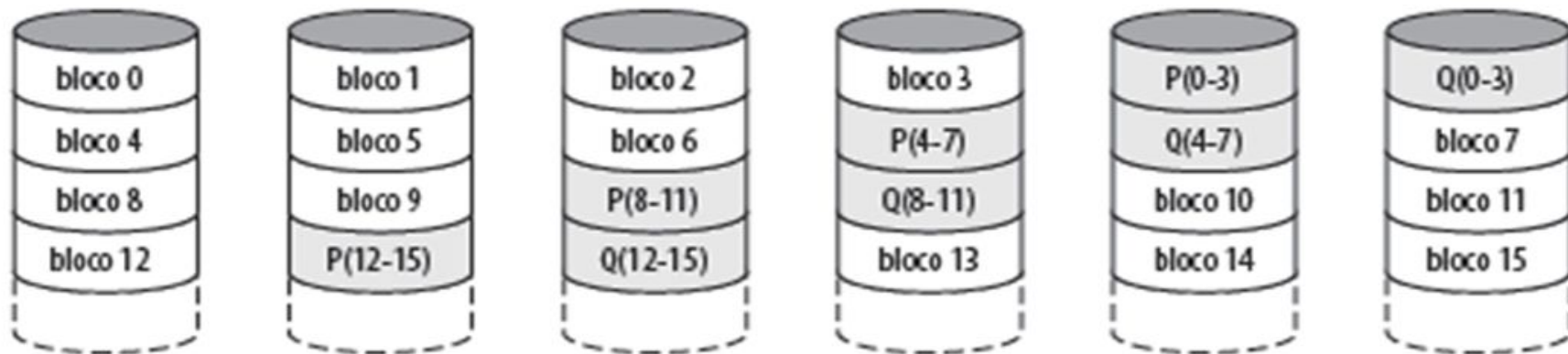


RAID 6

- Dois cálculos de paridade;
 - Necessários dois discos a mais;
- Cada cálculo de paridade para um strip é armazenado em um disco diferente;
 - Um strip de paridade de cada um dos discos 1 até $n-1$ está armazenado no disco n e outro no disco $n+1$;
 - Um strip de paridade de cada um dos discos 1 até $n-2$ e do disco $n+1$ está armazenado no disco $n-1$ e no disco n ;
 - Um strip de paridade de cada um dos discos 1 até $n-3$ e dos discos n e $n+1$ está armazenado no disco $n-2$ e outro no disco $n-1$;
 - e assim por diante;
- Tolerância a falha de até dois discos;
 - Os cálculos de paridade são diferentes;



RAID 6



RAID 6

- Vantagens:
 - Protege até duas falhas de disco;
- Desvantagem:
 - Verificação de paridade consome o dobro do espaço dos RAID 3, 4 e 5;
 - Escrita de dados pequenos em disco requerem mais acessos a disco que em RAID 3,4 e 5;

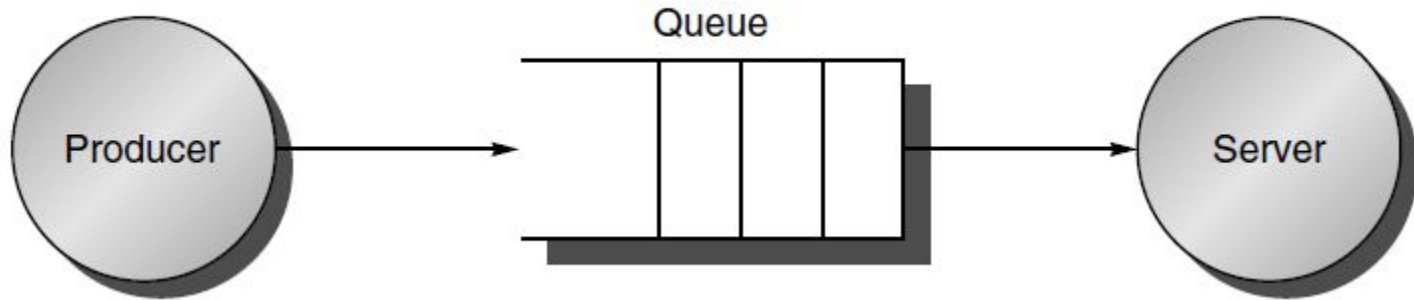


Desempenho de E/S

- Medidas tradicionais também podem ser usada em dispositivos de E/S, como discos magnéticos;
 - Throughput, que também é chamada de largura de banda;
 - Tempo de resposta, que também é chamado de latência;
- Num modelo produtor servidor, o produtor cria as tarefas e as coloca num buffer enquanto o servidor pega a primeira do buffer e a executa;



Desempenho de E/S



Desempenho de E/S

- Tempo de resposta é definido como o tempo total gasto entre o momento que a tarefa é emitida para o buffer e o fim de sua execução;
- Throughput é a média de tarefas terminadas por um servidor num determinado período de tempo;



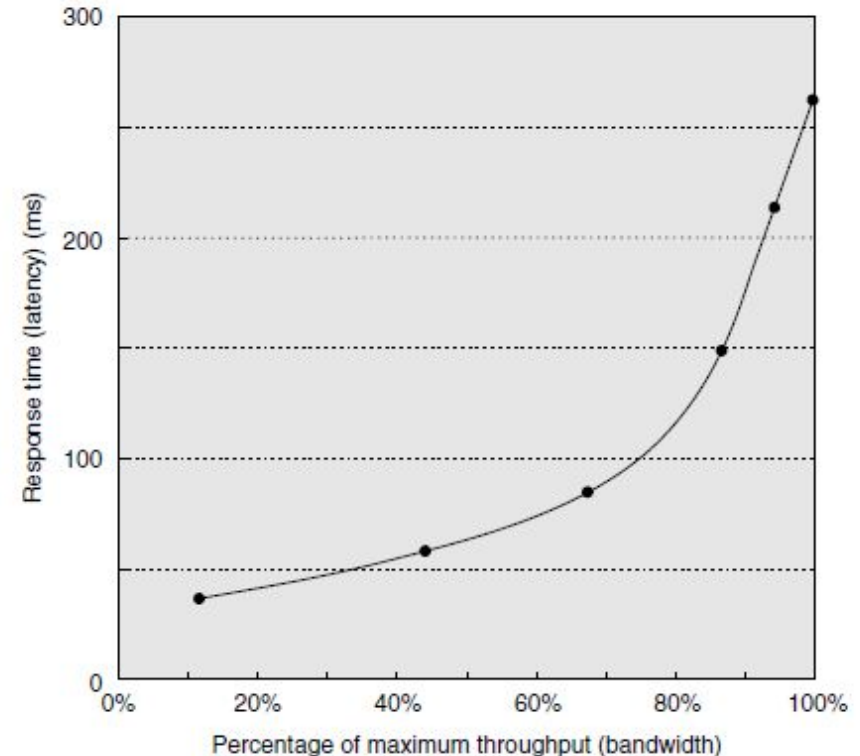
Desempenho de E/S

- Maximizar o throughput requer que o servidor nunca fique ocioso;
- Por outro lado, para tempo de resposta o tempo no buffer conta, então para um tempo de resposta “perfeito” o buffer deve ficar sempre vazio;



Throughput x tempo de resposta

- Existe um tradeoff entre throughput e tempo de resposta:
 - Aumentar o throughput aumenta o tempo de resposta;
 - Diminuir o tempo de resposta diminuir o throughput;
- Como balancear?



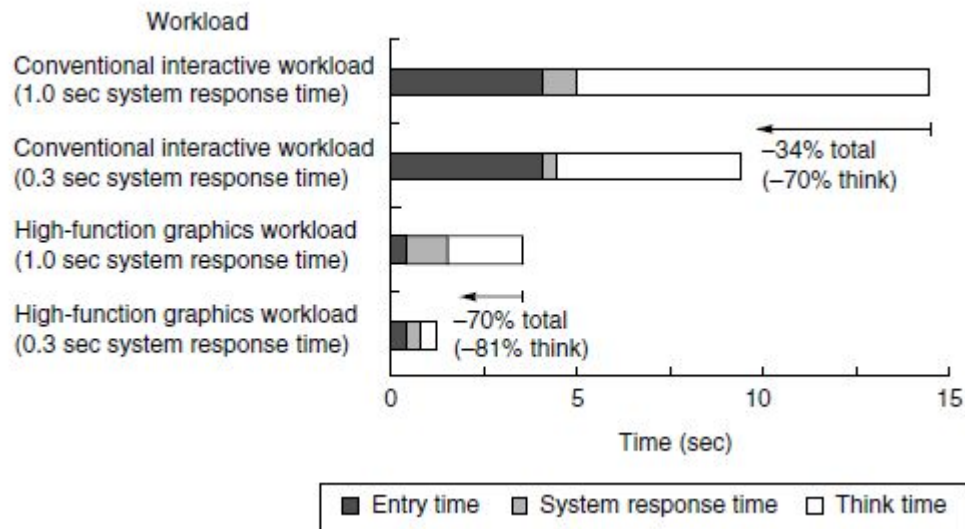
Throughput x tempo de resposta

- Computadores que interagem com seres humanos podem dividir a interação em 3 partes:
 - Tempo de entrada;
 - Tempo que o usuário gasta para fornecer a entrada;
 - Tempo de resposta do sistema;
 - Tempo entre o usuário entrar com os dados (ou comandos) e o sistema gerar a saída;
 - Tempo para pensar/
 - Tempo para a percepção da resposta até a entrada de um novo comando;



Throughput x tempo de resposta

- Estudos mostram que quanto maior o tempo de resposta mais o tempo para que uma nova interação ocorra;
 - O usuário se torna mais produtivo com um menor tempo de resposta;



Desempenho de E/S

- O throughput e o tempo de resposta é possível usar uma ferramenta matemática para auxiliar nos cálculos de tempo de resposta e throughput;
 - Teoria das filas;
- Usando a teoria das filas, fazendo uma adaptação da lei de Little e assumindo sistemas em equilíbrio (taxa de entrada é igual a taxa de saída) podemos chegar a equações para encontrar:
 - A média de tarefas no sistema:
 - $\text{número médio de tarefas no sistema} = \text{média de chegada de requisições} \times \text{tempo médio de resposta}$
 - A utilização do servidor:
 - $\text{utilização do servidor} = \text{média de chegada de requisições} \times \text{tempo no servidor}$;
 - Qual o tempo total de resposta (ou tempo total no sistema):
 - $\text{tempo no sistema} = \text{tempo na fila} + \text{tempo no servidor}$

Referências

STALLINGS, W. Arquitetura e organização de computadores. 8. ed. Prentice Hall, 2009.

HENNESSY, John L.; PATTERSON, David A. Computer architecture: a quantitative approach. Elsevier, 2011.

