Administração de Sistemas Informáticos
2005 / 2006

Dispositivos de Armazenamento de dados (HDD)

1000002 – Fernando Leal
1020113 – Hélder Ferreira
1000313 – Paulo Santos
1000941 – Ana Silva

Departamento de Engenharia Informática
ISEP – DEI

Índice

1.0 – Introdução

O objectivo principal deste trabalho consistiu na recolha e análise de toda a informação necessária para elaborar um relatório sobre discos rígidos de computadores. Tema escolhido pelos alunos, devido à sua importância e sua evolução ao longo do tempo.
No desenvolver deste empreendimento, são de realçar vários contributos ao nível académico, salientando a aprendizagem de novas descobertas relativamente a promenores históricos desconhecidos até então. Assim como a consolidação de conhecimentos préviamente adquiridos em cadeiras anteriores.
Desde o uso original do disco rígido num único computador, as técnicas de prevenção contra falhas dos discos rígidos foram desenvolvidas, assim como a RAID ("Redundant Array of Independent disks"). Também se encontram discos rígidos na NAS ("Network Attached Storage"), mas para grandes volumes de informação a SAN ("Storage Área Network").
A título de curiosidade, é interessante mencionar que as aplicações para os discos rígidos expandiram de forma a incluírem câmaras digitais e agendas electrónicas. Em 2005 foram introduzidos pela Nokia e Samsung, os primeiros telemóveis que incluíam discos rígidos.
O relatório encontra-se organizado por capítulos e respectivos subcapítulos. O primeiro capítulo diz respeito à *Introdução*, onde é apresentado o tema escolhido e seus objectivos, a estrutura do relatório. O segundo capítulo apresenta um resumo histórico, onde menciona a criação do primeiro disco rígido e suas eventuais evoluções/adaptações até aos discos usados actualmente. No terceiro capítulo consta a definição de um disco rígido e respectivo funcionamento mecânico. O quarto refere as diferentes conexões e modos de endereçamento de um disco.

2.0 - História

Os discos rígidos tal como os vários componentes do computador foram evoluindo ao longo do tempo.
As evoluções (alterações) nos discos rígidos foram a nível do seu tamanho, capacidade, velocidade de funcionamento e velocidade de comunicação.
O primeiro disco rígido foi o produzido pela IBM em 1955 / 1957, o IBM 305 RAMAC (abreviatura de Random Access Method of Accounting and Control) com 50 pratos (discos) de 24 polegadas (60.96 CM) de diâmetro, com uma capacidade de 5 Megabytes.
Em 1973 a IBM lançou o modelo 3340 "Winchester", com dois pratos de 30 Megabytes e tempo de acesso de 30 milisegundos, o primeiro a usar uma caixa exterior protectora, estas características levaram a que ficasse conhecido como "Winchester 30/30 rifle".
Durante muitos anos, os discos duros eram dispositivos grandes e incómodos, mais indicados para serem usados em ambientes protegidos como por exemplo num centro de dados ou num grande escritório e não noutros tipos de ambientes, como por exemplo, num ambiente industrial, escritório ou em casa (devido ao seu tamanho e consumo de energia). Antes de 1980, a maioria dos discos duros tinham pratos de 8-inch (20,32 cm) ou 14-inch (35.56 cm), requeriam equipamento extra ou uma grande quantidade de espaço (especialmente os removíveis, referidos frequentemente como "máquinas de lavar"), e em muitos casos necessitavam de fontes de alimentação devido aos grandes motores que usavam.
No ano de 1980 foi apresentado ao mundo primeiro disco rígido de formato pequeno pela "Seagate

Technology" que introduziu o ST-506, o primeiro disco duro de 5.25-inch (13,335 cm), com uma capacidade de 5 megabytes.
Em 1992 os discos rígidos ficaram pequenos em tamanho físico, possibilitando o seu uso em notebooks. Em 1997 a IBM impressiona e cria uma nova tecnologia chamada Giant Magnetoresistive (GMR), que abriu precedentes na indústria do armazenamento de dados, chegando a quase 17 Gigabytes de capacidade. Um ano depois a própria IBM lançava sua primeira unidade com 25GB, e em 1999, a mesma lança o disco rígido recordista de capacidade, com 73GB, trabalhando a 10.000 RPM. No ano 2000 foi lançado pela IBM o primeiro microdrive o menor disco rígido do mundo, com capacidade de 1GB.
No que respeita à história as principais famílias de discos são MFM, RLL, ESDI, SCSI, IDE, EIDE e agora SATA.
Os MFM necessitavam que os componentes electrónicos do "controlador do disco" fossem compatíveis com os da placa controladora.
RLL (Run Length Limited) consistiu num modo de codificar os bits nos discos para uma melhor densidade.
ESDI foi uma interface desenvolvida pela Maxtor, que permitia uma comunicação mais rápida entre o PC e o disco.
SCSI (Small Computer System Interface) foi um concorrente inicial do ESDI, quando os preços dos componentes electrónicos baixaram, os componentes electrónicos que vinham sido armazenados na controladora foram incorporados no próprio disco, este avanço ficou conhecido como "Integrated Drive Electronics" ou IDE.
Actualmente é difícil comprar um computador (Desktop) com um disco rígido menor que 40 GB e com uma velocidades de funcionamento inferiores a 7200RPM.
A interface de comunicação actualmente mais comum é a IDE/ATA Ultra DMA100 e Ultra DMA133, medidos em Megabytes por segundo. Estando a imergir um novo padrão o SATA, ou Serial ATA, que já começou a 150 Megabytes por segundo e promete chegar a 333 Megabytes por segundo, já que o padrão ATA está estagnado a 133 Megabytes por segundo. Espera-se que o SATA seja o padrão num futuro muito próximo.

3.0 - O que é um HDD (Hard Disk Drive)

Um disco duro é de facto um conjunto de discos todos eles revestidos por uma camada superficial magnetizável. A informação (os bits) é gravada nos discos magnetizando localmente um ponto sobre o disco.
Um disco duro é um dispositivo de armazenamento de dados não volátil, ou seja, após todo o sistema ser desligado a informação permanece literalmente na superfície do disco e pode ser recuperada quando todo o sistema for ligado novamente.

Foto 1- Vários discos duros

3.1 - Funcionamento mecânico básico de um HDD

Existem materiais que quando estão sob a acção de um campo eléctrico têm tendência para se orientarem segundo este. Este facto é facilmente verificado através de uma simples experiência usando uma bússola e um condutor eléctrico ligado a uma pilha.
Essencialmente, um disco duro usa este princípio para guardar informação. Como é que isto é

conseguido? Para que isto seja compreendido é essencial conhecer as partes essenciais de um disco duro.

Um disco duro é constituido por um grupo de discos de metal ou plástico do mesmo diâmetro revestidos por um material ferromagnético (por exemplo: óxido de ferro). Estes disco encontram-se juntos no mesmo eixo. Por sua vez, este eixo está ligado a um motor passo-a-passo.

Existem duas cabeças de escrita e leitura para cada disco, uma para cada face. As cabeças encontram-se muito próximas da superfície do disco. Todas as cabeças são montadas em cima de um braço, em forma de pente, que descreve um movimento radial, aquando de qualquer acesso ao disco. A posição onde a informação é gravada, é referenciada através da pista, sector e cabeça.

Foto 2- Vista interna de um disco duro

Quando em funcionamento os discos estão a girar a uma velocidade constante e as cabeças pairam apenas a alguns nanómetros de superfície do disco, transmitindo as variações do campo magnético quando estão a ler, ou magnetizando a superfície quando estão a escrever.

Um pormenor interessante é a forma como as cabeças se deslocam sobre os discos e de que forma isso afecta o design de novos discos duros:

- Como as distâncias entre as cabeças e os discos são mínimas todo o ambiente no interior do HDD não pode conter poeiras uma vez que estas poderiam originar uma colisão entre a cabeça e a superfície do disco, dando origem a danos físicos na camada magnetizada do disco arranhando-a, por exemplo.
- Por outro lado se o disco estivesse exposto às variações da humidade ambiente, isto poderia dar origem a que após algum tempo parado, a cabeça colasse à superfície do disco. Este contacto aceleraria o desgaste tanto da cabeça como da superfície.
- Não existe vácuo no interior das caixas seladas dos discos duros. A baixa pressão dificultaria o funcionamento tão próximo da superfície e aumentaria a probabilidade de colisão com a superfície. Para que a cabeça flutue à superfície é necessário que o ar esteja a passar por entre a superfície do disco e a cabeça. Análogamente a cabeça é como a asa de um avião, o vento a passar debaixo da asa é que fornece força de sustentação.

Mesmo assim o design das cabeças de leitura actuais é tão robusto que permitem que estas resistam até 50 000 impactos com a superfície antes de a probabilidade de ocorrer danos seja superior a 50%

4.0 - Diferentes modos de endereçamento e conexão de HDD's

Um disco rígido é geralmente acedido através de um dos inúmeros tipos de protócolos de comunicação, entre os quais ATA (*"AT Attachment Interface"*) (IDE, EIDE), SCSI, FireWire/IEEE 1394, USB, e Fibre Channel.

Na altura da interface ST-506, o método de codificação de dados também era importante. Os primeiros discos ST-506 usavam *"Modified Frequencie Modulation"* (MFM), actualmente usado nas disquetes, e obtinham velocidades na ordem dos 5 megabits por segundo (Mbps), posteriormente chegando até aos 7.5 Mbps, aumentando em 50% a sua capacidade.

Diversos discos de interface ST-506 não eram certificados para executar às velocidades máximas mas a um terço da velocidade e capacidade, dado que eram muito dispendiosos.

ESDI também suporta múltiplos modos de velocidade, sendo esta automáticamente seleccionada pela controladora, e traziam *conectores* para permitir a configuração do disco.

4.1 - IDE

A interface IDE/ATA (Integrated Drive Eletronics - Electrónica de drives integrada/Advanced Technology Attachment - ligação de tecnologia avançada) para HDs é a mais popular há cerca de uma década.
O padrão IDE sofreu bastantes alterações desde sua concepção original até à versão mais recente a Ultra DMA/133 e que permite as mais elevadas taxas de transferência, taxas de até 133 MB/s (megabytes por segundo).
Um aspecto que contribui para a eficiência dos HDs é a alta taxa de rotação. Desde que os dados consigam ser extraídos ou inseridos em altas rotações, fica claro que se os discos girarem mais rápido mais dados passam pelas cabeças de leitura e escrita durante um mesmo intervalo de tempo. Os HDs mais comuns actualmente atingem taxas de 10000 rpm (rotações por minuto). Além disso o tempo de acesso médio tem caído bastante e hoje em dia encontra-se HDs que oferecem apenas 9ms em comparação a 20ms de modelos obsoletos. O tempo de acesso médio é uma composição entre diversas situações de acessos, como acesso entre pistas consecutivas e entre pistas bem distantes. É interessante ressaltar que o modelo de DMA (Direct Memory Access - acesso directo à memória) utilizado por esses HDs não é o mesmo que o utilizado por placas de som, placas de vídeo ou outros dispositivos. Estes outros dispositivos precisam da ajuda do chipset da placa-mãe para poder utilizar a memória RAM sem intervenção do processador - isto é DMA. Os HDs compatíveis com o protocolo DMA possuem um sistema próprio que liga directamente à memória sem ajuda do processador ou dos recursos da placa-mãe. E porque é importante que o HD comunique rapidamente com a RAM? Simplesmente porque o processador opera eficientemente com a RAM e os principais dados dos programas estão armazenados no HD, por isso o HD precisa transferir dados para a memória da maneira mais rápida possível.
Pouco antes da introdução dos modos DMA, o topo em termos de performance era o protocolo PIO (Programmed Input/Output). O problema do PIO é que obrigava o processador a gerir as transações entre o HD e a memória. Como o tempo do processador é precioso, fica evidente que os modos DMA são muito mais eficientes.
Qual a diferença entre um dispositivo IDE ajustado para *master e* outro para slave? Ao contrário do que sugere a nomenclatura, não há privilégios extras para o *master* na maioria das situações e nem é o *slave* um dispositivo auxiliar do *master* ou que tem menor prioridade. A eles são dados iguais privilégios. Na verdade a convenção foi criada com objectivo de definir as denominações lógicas (letras dos drives). Por isso, a partição activa do *master* do primeiro canal sempre recebe a denominação c, além disso ela também será o dispositivo de arranque (boot) preferencial. As partições lógicas secundárias sempre recebem denominações inferiores a mais baixa partição activa de um canal *master*, incluindo dispositivos SCSI.

4.2 - SCSI

A interface SCSI (*Small Computer Systems Interface* - interface de sistemas para computadores de pequeno porte) não é das mais populares entre os utilizadores domésticos, sendo mais usada em ambientes corporativos e pequenos negócios. Isto deve-se a factores que tornam a interface SCSI mais bem moldada para situações comuns nestes meios. O custo de implantação é maior do que a concorrente IDE, por isso a sua popularidade não é tão grande já que a interface IDE pode cumprir

os objectivos da maioria dos utilizadores. Apesar de todos os avanços da IDE/ATA, a SCSI (lê-se scãzi), que também está em constante avanço, ainda é um pouco mais eficiente e mais robusta. Além disso, a SCSI pode ligar outros dispositivos internos ou externos, enquanto que a IDE está limitada a dispositivos internos e de uso exclusivo para armazenamento. Uma das grandes vantagens da SCSI é a imbatível capacidade para gerir os diversos dispositivos que podem estar ligados no mesmo barramento, principalmente em sistemas multitarefa e de rede nos quais mais de um programa ou utilizador podem querer aceder ao mesmo periférico simultaneamente, nomeadamente em servidores de rede. Para isso há um elaborado esquema comum referenciado como *queue* (fila de tarefas usualmente ordenada por ordem temporal de requisição).
Os HDs SCSI são mais caros que os IDE de mesma capacidade e além disso necessitam de uma placa controladora. Estas controladoras podem acomodar entre 7 e 15 periféricos no caso de controladoras com suporte para wide SCSI - 16bits de banda de transferência. Os periféricos podem ser scanners, dispositivos de armazenamento removível, gravadores de CD-R/RW entre outros.
De certa forma o barramento SCSI é como um canal IDE. Assim, também é necessário que cada dispositivo ligado a ele possua um identificador como *o Master / Slave* do IDE, mas de maneira mais abrangente. Como a controladora também ocupa uma posição no barramento é possível escolher uma identificação (ID) entre O e 7 ou O e 15 para cada periférico. O barramento SCSI também precisa definir a posição de terminadores. Terminadores são elementos eléctricos utilizados para reduzir interferências e reflexões electromagnéticas dos próprios sinais do barramento nos cabos que ligam os diversos periféricos internos e externos. Deve haver um terminador em cada periférico ligado a um extremo físico do cabo SCSI.
Um recurso bem interessante de HDs SCSI é o RAID (Redundam Array of Independent Disks - conjunto redundante de discos independentes; algumas vezes encontra-se o termo *Inexpensive* - barato - no lugar *de Independente.* Há diversos modelos RAID que recebem um algarismo de identificação. Por exemplo, o RAID nível l consiste de um conjunto de dois ou mais discos que armazenam exactamente as mesmas informações (disk mirroring - espelhamento de discos). Quando o disco principal falha, o disco espelho é utilizado até que o primário seja substituído. Alguns sistemas RAID permitem que o disco defeituoso seja substituído mesmo com os equipamentos ligados (hot swap) como se nada estivesse acontecendo. Para utilizar-se de recursos RAID é necessária a utilização de placas especiais e frequentemente gabinetes extras para acomodar os sistemas mais refinados. Desde o princípio os HDs SCSI utilizam o modelo LBA para localizar os dados nas superfícies dos discos, o mesmo modo que os HDs IDE começaram a adoptar posteriormente.
O modelo LBA (Logical Block Addressing - endereçamento por blocos lógicos) é o mais refinado no momento e também bastante genérico, tanto é que é aplicado também em HDs SCSI. Um HD com organização LBA interpreta dados requisitados num formato linear de indexação e converte-os para valores reais da geometria (cilindro, sector e cabeça). Em geral os novos HDs utilizam uma quantidade de sectores dependente da posição ocupada pela pista no disco, por isso, o HD é encarregado de mapear posições lógicas para as posições físicas correspondentes. Com isso juntamente a outros processos de manufactura possibilita-se um melhor aproveitamento do espaço disponível nos cilindros para armazenamento.

4.3.1 - Interface ATA.

Os discos ATA tipicamente não tinham problemas nem com velocidade nem com a *interleave*[1].

1 ***Interleave***: . Distância entre sectores comparáveis em pistas adjacentes. É necessário devido ao tempo que demora a cabeça a deslocar-se entre pistas. Em sistemas antigos podia ser definido pelo utilizador.

Devido ao design da sua controladora, mas muitos dos primeiros discos ATA eram mutuamente incompatíveis entre si, e não funcionavam no modo escravo/mestre. Por volta de 1990, o problema resolveu-se com a padronização e clarificação de detalhes ocorreu na norma ATA, mas ocasionalmente ainda ocorrem alguns problemas quando se juntam periféricos ultra DMA ("Direct Memmory Acess") e não ultra DMA.
Alguns dos modos mais comuns de funcionamento dos discos ATA são ATA33 (UDMA33), ATA66 (UDMA66), ATA100 (UDMA100), ATA133 (UDMA133), sendo que a partir do ATA66, inclusivé, é necessário um cabo especial de 80 pinos. No entanto nenhum dos discos existentes dificilmente consegue atingir mais do que 33MBps, o que é muito inferior ao da norma máxima de 133MBps do ATA133.

4.3.1.1 - Endereçamento da interface ATA.

Existem dois tipos de endereçamento para discos ATA, o CHS (*Cilinder-Head-Sector*) e o LBA (*Logical Block Addressing*), sendo que os discos mais antigos (até 528 MBps) usavam o CHS e depois com a evolução da capacidade passaram a utilizar o LBA de forma a permitir capacidades até 2^28 bits.

4.3.2 - Descrição do Serial ATA (SATA).

No fim de 2002 foi introduzido Serial ATA, que aboliu a norma de funcionamento escravo/mestre, garantindo a cada disco um canal exclusivo de comunicação. Esta norma usa dois fios, um envia e o outro recebe dados para e do disco respectivamente, a 1.5 GBps ou mais.

4.3.3 - Fibre Channel

Os discos de interface Fibre Channel são do tipo interface *Ethernet* e *InfiniBand*[1] já que não se destinam exclusivamente a discos mas também a periféricos. Frequentemente usado para ligar um disco SCSI a uma RAID ou RAID a uma rede, com velocidades entre os 106 MBps e os 1060MBps.

4.3.4 - FireWire

IEEE 1394 (FireWire), pode também ser usado como no caso anterior para ligar discos SCSI ou IDE, permitindo o *hotplug* desses discos permitindo velocidades de 100 a 400 Mbps.

4.3.5 - USB

USB (*Universal Serial Bus*) essencialmente funciona como o FireWire mas na versão 1.1 permite velocidades até um máximo de 12Mbps, enquanto que na versão 2.0 permite velocidades na ordem de 480 Mbps.

5.0 - Dispositivos RAID

Um sistema RAID consiste na utilização de múltiplos discos rígidos para entre eles partilhar ou duplicar informação. O benefício da utilização de um sistema RAID verifica-se com o aumento da integridade da informação, da tolerância a falhas, ou da taxa de transmissão da informação, consoante a versão do sistema escolhida.
Na sua versão original, patenteada pela IBM[2] em 1978, RAID significava *Redundant Array of*

1 **Infin*i*Band :**É uma Arquitetura de I/O e também uma especificação para a transmissão de dados entre processadores e dispositivos de I/O que têm vindo a substituir gradualmente o *bus* do PCI em servidores e *PCs*. Em vez de enviar dados na paralelamente, que é o que o PCI faz, o *InfiniBand* envia dados em série e pode levar os dados por múltiplos canais ao mesmo tempo num sinal *multiplexing* (envio de múltiplos sinais ou conjunto de informação de uma só vez e ao mesmo tempo, recuperando mais tarde os sinais separados na extremidade/no fim).

2 ***I**nternational **B**usiness **M**achines*, E.U.A.

***Inexpensive** Disks*. Um dos benefícios do sistema RAID era a utilização de discos "baratos" do tipo IDE. Actualmente, usam-se frequentemente discos SCSI, que são agora muito mais económicos que outrora e que aqueles que o RAID original pretendia substituir.
Na sua versão mais simples, RAID combina múltiplas discos numa única unidade lógica, de tal forma que os sistema operativo "vê" apenas um disco. RAID é tipicamente usado em servidores, usando discos de características idênticas.
A especificação original sugeria um número de "níveis" para um sistema RAID, correspondentes a diferentes tipos de implementação do sistema. Cada um destes níveis trazia vantagens e desvantagens.

5.1 - Hardware vs. Software

Um sistema RAID pode ser implementado directamente em *hardware* (ao nível do controlador de discos) ou em *software*. Há ainda soluções híbridas.
A implementação em *software* usa um controlador normal (IDE, SCSI...). Com as actuais velocidades dos processadores, esta implementação pode ser a mais rápida, mas tem a desvantagem de usar ciclos de CPU que podem eventualmente ser melhor empregues noutras tarefas. Outra desvantagem da implementação em *software* é que, em caso de falha de um disco, e em função dos parâmetros de arranque do sistema, este pode não conseguir arrancar até que o problema seja solucionado.
A implementação em *hardware* requer (no mínimo) um controlador de discos RAID. Pode ser uma placa de expansão PCI, ou estar construído na própria *motherboard*, ou ainda estar alojado num dispositivo externo. Os discos podem ser IDE, ATA, SATA, SCSI, etc., ou qualquer combinação destes. O controlador faz a gestão dos discos e trata dos cálculos de paridade (necessários para alguns dos níveis RAID).
Esta implementação tende a ter melhor performance, e é a mais simples do ponto de vista da implementação e da manutenção do sistema informático. Tem ainda a vantagem de, frequentemente, estes controladores suportarem *hot swapping* dos discos, permitindo a troca de um (ou mais) discos eventualmente danificados com o sistema em funcionamento.
Os sistemas de RAID híbrido tornaram-se muito populares com o aparecimento no mercado de controladores RAID muito baratos. Na realidade, o *hardware* é um controlador normal sem quaisquer funções RAID, mas há uma rotina no arranque da máquina que permite ao administrador do sistema definir conjuntos RAID que são controlados pela BIOS. É necessário depois utilizar um *driver* especial no sistema operativo para tirar partido destas funcionalidades.

5.2 - Níveis RAID Standard

Nível 0

Um sistema RAID 0 divide a informação de forma igual entre dois (ou mais) discos, sem cálculo de paridades para efeitos de redundância. Os blocos de informação são guardados alternadamente nos vários discos que compõem o sistema – no caso de dois discos, os blocos pares são guardados num e os ímpares noutro. Desta forma, pretende-se maximizar a velocidade de leitura de informação, sendo que a capacidade de transmissão no barramento é normalmente muito superior à de leitura dos discos. O sistema pode também ser usado para obter unidades lógicas de maior capacidade que cada uma das

RAID 0	
A1	A2
A3	A4
A5	A6
A7	A8

unidades físicas quando consideradas separadamente.
Discos de capacidades diferentes podem ser usados num sistema RAID 0, mas o tamanho combinado será limitado pelo disco de menor capacidade (se, por exemplo, foram usados dois discos, um de 100MB e outro de 150 MB, o tamanho combinado será de 200MB – os 50MB extra do segundo disco em relação ao primeiro não serão usados).
É possível usar RAID 0 com mais que dois discos, mas quanto maior for o número de discos, mais provável se torna que um deles possa falhar. A fiabilidade de um sistema RAID 0 corresponde à fiabilidade média de cada um dos discos dividida pelo número de discos do conjunto. Assim, a fiabilidade (medida como o tempo médio entre falhas) de um sistema RAID 0 com dois discos é, grosso modo, metade da fiabilidade de cada um dos discos quando considerados separadamente; com quatro discos, será um quarto, e assim sucessivamente.
A maior vantagem de um sistema RAID 0 é a performance em termos de velocidade de leitura. A visão de várias unidades físicas (vários discos) como uma única unidade lógica pode também ser vantajosa. A grande desvantagem é a possibilidade de falha, que aumenta proporcionalmente ao número de discos. Como não é guardada informação de paridade, se um dos discos falhar todo o conjunto estará comprometido – nenhuma informação será recuperável, porque ficar-se-á apenas com os sectores ímpares (ou pares) de todo a conjunto (caso de dois discos).

RAID 1

Um sistema RAID 1 cria uma cópia exacta (*mirror*) de um ou mais discos. Esta hipótese é importante para sistemas informáticos em que a redundância é mais importante que a máxima capacidade de armazenamento dos discos. O conjunto terá apenas a capacidade do disco mais pequeno.
Numa configuração típica são usados dois discos. Ao contrário do RAID 0, aqui a fiabilidade do sistema é multiplicada por dois, relativamente a ter um único disco. É no entanto possível ter mais que uma cópia, variando a fiabilidade de forma linear com o número de discos.
Num sistema RAID 1 a velocidade de leitura, tal como no caso do RAID 0, é superior à de um disco isolado pelo factor do número de discos usado (até ao limite de largura de banda do barramento de transmissão). Apesar de cada disco conter a mesma informação, é possível dividir cada pedido de informação em blocos e ter cada disco a ler diferentes blocos em simultâneo. Ao escrever, a velocidade do conjunto é a mesma de um único disco (a informação é escrita em cada um dos discos).

RAID 2

Um sistema RAID 2 separa informação com base no bit (e não no bloco, como o RAID 0). Este é o único dos níveis originais do RAID que já não é usado.

RAID 3

Um sistema RAID 3 faz a separação da informação com base no byte, e usa um disco dedicado para guardar informação de paridade. Este nível é só muito raramente usado na prática. Um dos inconvenientes é que qualquer operação de I/O implica actividade em cada um dos vários discos do conjunto, porque cada bloco de informação, por definição, está dividido por todos os discos. Desta forma, apenas um pedido pode ser atendido de cada vez.

RAID 4

RAID 3			
A1	A2	A3	Ap(1-3)
A4	A5	A6	Ap(4-6)
A7	A8	A9	Ap(7-9)
B1	B2	B3	Bp(1-3)

O RAID 4 é idêntico ao RAID 3, mas divide a informação em blocos e não em bytes. Isto significa que cada disco pode ser usado independentemente, e vários pedidos por blocos distintos, desde que gravados em discos distintos, podem ser atendidos em simultâneo.

RAID 5

Um sistema RAID 5 usa a divisão em blocos pelos vários discos, mas guarda a informação de paridade em cada disco. Este é um dos níveis RAID mais populares.

RAID 5			
A1	A2	A3	Ap
B1	B2	Bp	B3
C1	Cp	C2	C3
Dp	D1	D2	D3

De cada vez que um bloco de informação é escrito num dos discos do conjunto, é criado (ou actualizado) o respectivo bloco de paridade. O "disco" usado para armazenamento dos blocos de paridade passa a estar distribuído pelos vários discos do conjunto RAID. Os blocos de paridade não são lidos a cada leitura dos blocos de informação respectivos – isso implicaria uma diminuição significativa da performance -, mas apenas quando é detectado um erro de CRC[1]. Neste caso, informação de blocos contíguos na mesma faixa e a informação do bloco de paridade são usadas para reconstruir o bloco errante. Da mesma forma, se um disco falha no conjunto RAID, a informação dos outros discos serve de base a um cálculo matemático, apoiado nos blocos de paridade, que permite reconstruir a informação do disco avariado em tempo-real. Desta forma, o sistema operativo "sabe" que um disco falhou e pode notificar o administrador, mas os programas a serem executados não se apercebem da falha.

Num sistema RAID 5 é assim possível suportar a falha de um dos discos do conjunto. Note-se, contudo, que a falha simultânea de um segundo disco resulta na perda total da informação.

O número de discos num sistema RAID 5 é teoricamente ilimitado, mas não é normal que ultrapasse as quatorze unidades. Isto porque a probabilidade de acontecerem duas falhas simultâneas em discos diferentes torna-se maior quanto maior for o número de discos.

Outro efeito para o qual convém estar alerta é o tempo médio de vida útil dos discos. Se todos os discos usados forem idênticos e fabricados sensivelmente na mesma altura, é provável que atinjam o seu limite de vida útil sensivelmente ao mesmo tempo. Por este motivo, é razoável utilizar no mesmo conjunto discos em diferentes períodos do seu tempo útil de vida.

1 ***C**yclic **R**edundancy **C**heck*

RAID 6

RAID 6 é uma extensão do RAID 5, e não estava incluído nos níveis RAID originais. A diferença para o RAID 5 é que usa um bloco de paridade adicional, distribuído por todos os discos membros do conjunto. A vantagem desta modalidade é a tolerância a falhas simultâneas em dois discos. A desvantagem é o *overhead* associado ao cálculo da informação extra de paridade de cada vez um bloco é escrito. O facto de usar mais espaço que o RAID 5 para guardar a informação adicional não é significativo em conjuntos com um número elevado de discos, mas é obviamente penalizador se o número de discos for reduzido.

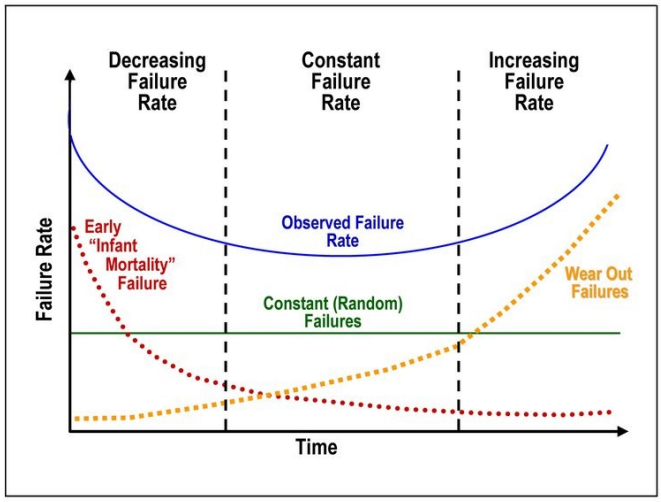

Figura 1: Probabilidade de falha de um dispositivo ao longo do seu tempo de vida

5.3 - Níveis RAID Mistos

Alguns controladores permitem vários níveis de RAID em simultâneo, isto é, usam um nível RAID como a unidade para o seguinte, em vez de discos físicos. Constroem-se assim "camadas" de RAID umas sobre as outras, com discos apenas ao nível mais baixo da "pilha".
Desta forma procura-se combinar as vantagens (e contornar as desvantagens) de cada um dos níveis RAID quando considerados isoladamente. A título de exemplo, é possível obter RAID 0+1 (ou RAID 01) "espelhando" um sistema RAID 0, ou RAID 10 (RAID 1+0).
O *kernel* Linux implementa em *software* alguns níveis de RAID, nomeadamente RAID 10.