Se algo de errado acontecer com seus dados, seu chefe irá perguntar:

E aí, tem becape?

Dados em computadores sempre tendem a se corromper exatamente quando mais precisamos, mas uma estratégia bem pensada de backup pode ajudá-lo a colocar aqueles arquivos perdidos de volta em seus lugares num piscar de olhos.

POR MARC ANDRÉ SELIG

Há inúmeras razões para a perda de dados. Por exemplo, um administrador distraído pode digitar um fatídico `rm -rf *` no diretório errado, apagando centenas de arquivos importantíssimos. Há também os casos de perda total de dados, quando – por exemplo – o disco rígido decide partir desta para uma melhor. E há ainda os casos em que não é possível, ou sensato, confiar na integridade dos dados – como logo após uma invasão, por exemplo. Nenhuma dessas situações deixa outra alternativa ao administrador de sistemas a não ser restaurar uma cópia de segurança (e é bom que ela exista e seja confiável), o popular "becape", que esteja comprovadamente livre de qualquer suspeita.

Como as causas e necessidades envolvendo perda de dados podem ser muito diferentes, diversas formas de lidar com o problema vieram à tona ao longo dos anos. Todas elas possuem benefícios e desvantagens. Este mês, descreveremos algumas das ferramentas de backup disponíveis e as técnicas com as quais devem ser usadas.

Tabela 1: Escolha sua mídia

	Vida útil	Confiabilidade	Armazenamento*	Velocidade	Disponibilidade**
Fita magnética	++	++	++	-	Após restauração
CD/DVD	+	+	++	+	Após restauração
MO	++	+	++	+	Após restauração
Disco rígido interno	-	-	-	++	Imediatamente ("*a quente*")
Disco rígido externo/removível	-	-	++	o	Imediatamente após a conexão

++: Ponto forte, **+:** Aplicável, **o:** Parcialmente aplicável, dependendo da mídia, **-:** Não aplicável

* *Armazenamento* indica a possibilidade de armazenar as mídias em outro local geograficamente distante de seus servidores, dando mais segurança contra acidentes, desastres naturais e roubo.

** *Disponibilidade* indica quando os dados estarão disponíveis caso se precise da cópia de segurança.

Escolhendo a mídia

A **tabela 1** mostra uma breve comparação entre os tipos de meios de armazenamento disponíveis. As fitas magnéticas já foram as estrelas solitárias do armazenamento de dados, desbancando os cartões perfurados na década de 60 e seguintes. Em tempos mais recentes, elas ainda brilhavam para as cópias de segurança dos frágeis e ineficientes discos rígidos e, ainda hoje, são as vedetes dos CPDs e *Data Centers* que precisam trabalhar com quantidades monstruosas de dados. Fitas magnéticas são muito baratas, mesmo considerando a quantidade fabulosa de dados que podem guardar, mas têm uma desvantagem gritante: a velocidade de acesso é bastante baixa. Não seria tão mau se a outra desvantagem não fosse proibitiva: os *drives* de fita – mesmo os mais baratos – são esmagadoramente caros para a maioria das pequenas e médias empresas em qualquer parte do mundo. Ainda assim, uma (ou várias) unidades de fita conectadas com uma *jukebox* robotizada é o que se tem de melhor ainda hoje para backups automatizados de alta capacidade.

Para nós, reles mortais, os CDs e DVDs graváveis, as memórias *flash* e os discos rígidos (internos ou externos, removíveis ou não) especiais para backup são as soluções mais comuns e acessíveis. Em empresas maiores, os administradores podem contar com um NAS (*Network Attached Storage* ou sistema de armazenamento ligado em rede) para aumentar a capacidade dos discos rígidos centrais.

Assim como há diferentes mídias para backup, há também estratégias diferentes. Na maioria dos casos, os administradores optam pelo conhecido método de *backups incrementais*, que armazenam apenas as mudanças que ocorreram desde o último backup. Essa estratégia poupa bastante espaço nas mídias de segurança, o que melhora a relação custo/benefício do sistema de backups como um todo. A maioria das ferramentas de backup existentes podem trabalhar com backups incrementais.

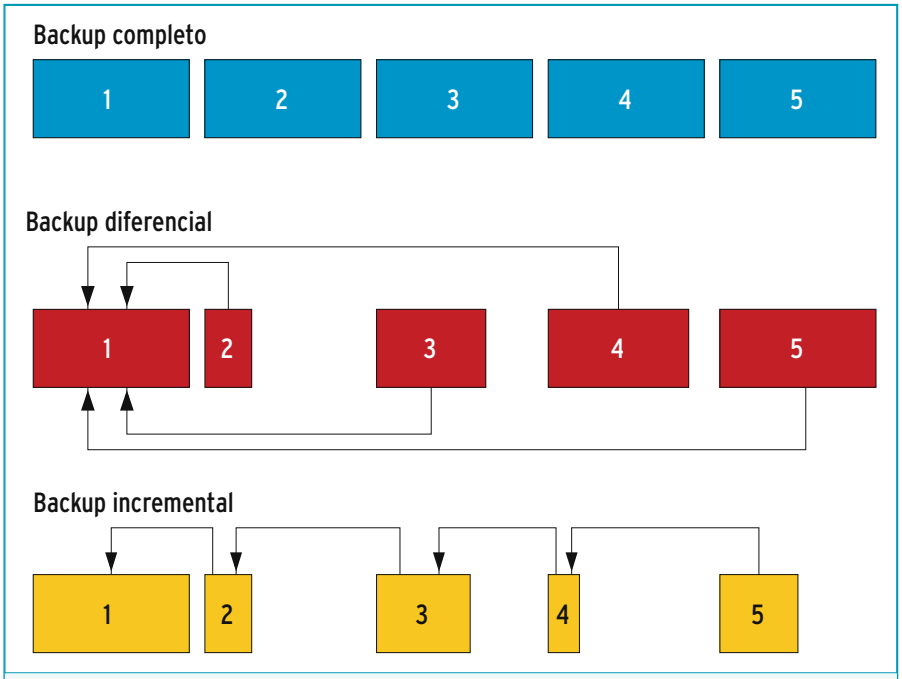

Figura 1: Um backup completo guarda todos os arquivos. Um backup diferencial armazena as mudanças ocorridas desde o último backup completo. Um backup incremental, por sua vez, armazena arquivos modificados desde o último backup, seja ele completo ou diferencial.

A grande desvantagem dos backups incrementais é que restaurar os dados perdidos dá muito mais trabalho e toma muito mais tempo do que restaurar um backup completo. Os backups incrementais armazenam as diferenças desde o último backup, mesmo que ele seja outro backup incremental. Além disso, os administradores têm que ficar trocando as mídias de backup se não possuírem orçamento para uma *jukebox* robotizada. Há uma terceira modalidade chamada *backup diferencial*, que sempre armazena as mudanças em relação ao último backup completo, diminuindo um pouco o problema. A **figura 1** ilustra os três métodos.

Uma palavra sobre disponibilidade

A escolha do método de backup depende das circunstâncias em que os dados devem ser guardados. Se o arquivo de que o usuário desesperadamente precisa está numa fita jogada numa prateleira, o acesso ao arquivo vai precisar de intervenção humana. Isso pode ser uma vantagem: afinal, um invasor não conseguirá, por mais que tente, comprometer uma fita que não está no drive. Entretanto, o ato de restaurar o arquivo precisará de tempo e de alguém que o faça.

Na outra extremidade do problema, há soluções imediatas de backup em que os dados estão em uma mídia disponível 24x7. Esse método poupa tempo – e possivelmente dinheiro – mas é passível de outros tipos de falha. Uma delas é que essa mídia pode ser comprometida por nossos amigos *crackers*.

Há sistemas que podem criar os chamados *backups instantâneos* ou *backups a quente* (*hot backups*) a intervalos regulares ou mesmo de forma contínua. Entretanto, esse tipo de cópia de segurança protege apenas contra falhas no hardware. Não há proteção contra erros dos usuários ou do administrador, que serão propagados para o backup no exato instante em que foram cometidos. Por essa razão,

muitos administradores não confiam nas técnicas de backup instantâneo e não querem nem ouvir falar de substituir as técnicas tradicionais.

Formatos

Os administradores discordam entre si sobre os prós e os contras de se gravar cada arquivo e cada diretório individualmente no backup – basta simplesmente jogar os diretórios e arquivos na mídia e pronto. Alguns acham que esse é o caminho, mas outros preferem criar um pacotão com estruturas mais complexas de controle, com dados informativos (metadados) e numeros de verificação de integridade (*checksum*).

Backups em que cada arquivo individual é gravado na mídia tendem a ser mais rápidos para guardar e restaurar. Além disso, se a mídia tiver um pequeno defeito ("deu fungo no CD" ou "o cachorro comeu a ponta da fita" são bastante comuns...) apenas um punhado de arquivos – se muito – são afetados. Se o mesmo acontecer com um pacote de backup (ou seja, todos os arquivos empacotados e, possivelmente, compactados dentro de um arquivão) uma quantidade bem maior de dados vai para o espaço. Dependendo da gravidade, o backup todo fica comprometido.

Entretanto, os pacotes ou contêineres de arquivos oferecem benefícios que o método de armazenar arquivos um a um não pode trazer. Por exemplo, é possível armazenar informações como o proprietário e o grupo dos dados, os privilégios de acesso e as datas de criação e alteração dos arquivos individuais. É possível inclusive fazer cópias de segurança de dispositivos inteiros do diretório `/dev`. Além disso, as fitas magnéticas não são exatamente o meio ideal para armazenar uma multidão de pequenos arquivos. Muito pelo contrário, esses dinossauros da tecnologia são campeões no armazenamento de um só arquivo bem grande.

Muitos programas, incluindo o *tar* e o *cpio*, tentam encontrar o equilíbrio perfeito. Se um arquivo cpio estiver corrompido, o dano fica restrito aos arquivos armazenados no local da mídia em que a falha ocorreu. O programa resincroniza seus contadores internos com o próximo marcador de fim de arquivo após a falha, o que minimiza o prejuízo. Dessa forma, os arquivos que estiverem *depois* do local danificado podem ser restaurados.

Se vamos entrar na discussão "pacotão único contra arquivos individuais", também precisamos trazer à baila os problemas da compactação e da criptografia dos dados. O esquema de ressincronização do cpio funciona *apenas* para backups que não foram compactados. Se um erro de leitura impede que o arquivo seja descompactado, o cpio lhe será de pouca utilidade.

O popular `gzip` simplesmente aborta a descompactação quando encontra o primeiro erro no arquivo compactado. ***Evite o gzip para backups como o diabo foge da cruz:*** há inúmeros relatos de administradores de sistema em desespero porque perderam vários gigabytes de dados não-corrompidos porque o gzip simplesmente se recusa a continuar a descompactação a partir de um mísero bit defeituoso – o *zcat* pode, ao menos, recuperar parte dos dados até o ponto onde o erro ocorreu, mas nada além disso. O formato alternativo, *bzip2*, compacta e descompacta os arquivos em blocos de 900 KBytes no máximo. Se um erro de leitura ocorrer, perde-se um pequeno bloco de dados mas os blocos seguintes a ele não são afetados. Novamente avisamos: **não use o *gzip***.

Os administradores são confrontados com um dilema semelhante quando precisam criptografar os dados. Muitos algoritmos de criptografia usados pelos programas de backup são tão bons que os dados ficam inacessíveis em caso de falha na mídia. Uma possível forma de contornar isso poderia ser comprimir cada arquivo sozinho antes de gravá-lo no pacote. A ferramenta *afio* [1] é um candidato a substituir o cpio nessa tarefa, pois tem sintaxe de opções semelhante e pode criptografar individualmente cada um dos arquivos.

Quadro 1: Backup em fita

As fitas são muito populares e conhecidas. Raramente são vítimas de erros isolados de leitura e mesmo essas raras falhas podem ser evitadas com ferramentas de software mais sofisticadas. O que torna as coisas piores é o fato de muitos *drivers* para o kernel precisarem de blocos pré-formatados para dispositivos de fita. Em outras palavras, nem todas as unidades de fita do mercado são boas para se usar como dispositivo-alvo no comando `tar cpf`.

A maneira mais fácil de usar unidades de fita é empregar um software do tipo pronto-para-usar como o *Amanda* [2], que pode coletar dados de um número praticamente ilimitado de máquinas na rede e gravá-los em uma fita. O Amanda funciona com uma grande variedade de sistemas Unix e possui inclusive clientes para o Microsoft Windows® [3].

O sistema é baseado no modelo cliente/servidor. É preciso instalar um programa cliente do Amanda em cada máquina que deve ter seus dados guardados em uma cópia de segurança. Como é óbvio e ululante, o cliente precisa ter acesso de leitura para qualquer dado que deva ser guardado no servidor Amanda. O servidor envia requisições periódicas aos clientes pelo protocolo UDP e eles respondem com os dados a serem guardados transportados via TCP. O Amanda pode usar tanto o comando `dump` como o `tar` para criar os pacotes de arquivos.

O Amanda possui um sistema sofisticado de agendamento de backups. O programa servidor consulta as unidades de fita para ver quais estão livres e verifica na tabela de backups agendados quem está na vez e que tipo de cópia deve ser feita: completa ou incremental. Ou seja: cada máquina na rede tem seu backup feito sempre que possível e, pelo menos, no intervalo configurado. O Amanda também sabe quanto espaço há nas fitas já usadas e guarda nelas os backups incrementais.

Listagem 1: Script simples de backup

```
01 #!/bin/sh
02
03 [ `id -u` -eq 0 ] || ( echo 'É preciso ser root para gravar os dados no CD/DVD!' && exit )
04
05 TODAY=`date +%Y%m%d.%H%M`
06 MYKEY='0x598342d9'
07
08 umask 022
09 mkdir -p /tmp/root/backup-$TODAY
10
11 cd /
12 tar cf - etc home usr/local | \
13  gpg -v --homedir $HOME/.gnupg -e -r $MYKEY | \
14  tee /tmp/root/backup-$TODAY/backup-$TODAY.tar.gpg | \
15  md5sum -b >/tmp/root/backup-$TODAY/backup-$TODAY.tar.gpg.md5
16
17 cd /tmp/root
18 mkisofs -r -pad -o backup.iso backup-$TODAY
19 cdrecord -v -eject -multi dev=0,0,0 -driveropts=burnproof -speed=24 -pad backup.iso
20
21 rm -rf backup-$TODAY backup.iso
```

Backup em CD

O backup em fita, especialmente se controlado pelo Amanda (ver **quadro 1**), pode ser usado em ambientes mais modestos, mas sente-se bem mais à vontade em grandes corporações. Usuários domésticos e pequenas empresas talvez fiquem mais à vontade com um sistema de backup mais simples, baseado em CDs ou DVDs. Em comparação com as fitas magnéticas, e considerando um volume pequeno de dados, os CDs e DVDs são extremamente baratos e possuem um ciclo de vida maior.

A **listagem 1** mostra um script de backup bem simples, que chama o utilitário `gpg` para criptografar os dados e gera um arquivo MD5 para verificação de integridade. Se um CD se perder (e isso acontece com frequência quando não se é cuidadoso), você não precisa, pelo menos, se preocupar com acesso não-autorizado a seus dados. Modifique nosso script a seu gosto, fazendo-o usar cartões *flash* ou discos rígidos externos. Se sua distribuição Linux não usa emulação SCSI para gravação de CDs (como acontece com muitas das distribuições recentes), consulte a documentação do *cdrecord* para saber como fazer e altere a penúltima linha do script.

Faça a coisa certa

Um sistema de backups é tão bom quanto os dados gravados na mídia – e nem sempre esses dados são o que o programa de backup teve a intenção de gravar. Portanto, a melhor prática é verificar periodicamente suas cópias de segurança para garantir que tudo pode ser lido, que o lido é exatamente o que foi gravado, e que você gravou os dados certos, pra começo de conversa.

Além disso, é preciso criar mecanismos que permitam que várias pessoas ou mesmo o próprio usuário possam restaurar facilmente os dados em caso de emergência. Não há nada mais desesperador do que ter que restaurar um backup bem antigo e não ser capaz de fazê-lo porque o administrador que o gerou não está mais disponível – e o pior, levou com ele a senha…

Para os casos de perda total, há outras coisas a considerar. Como em muitos casos o sistema operacional (ou o hardware) do sistema principal podem estar bastante danificados, é obrigatório ter um computador reservado exclusivamente para a restauração dos backups. Esse sistema de socorro deve iniciar a partir de um CD ou disco rígido externo e permitir que o administrador restaure completamente os dados – em outras palavras, deve haver um servidor completo guardado no armário para colocar no lugar do que "morreu". É claro, esse tipo de solução requer planejamento, traquejo e investimento. ■

Informações

[1] Página oficial do *Afio*: directory.fsf.org/sysadmin/backup/afio.html

[2] Página oficial do Amanda: www.amanda.org

[3] Cliente do Amanda para Windows®: sourceforge.net/projects/amanda-win32