Implementando o UFS Journaling em um Desktop PC

Resumo

Um sistema de arquivos com journaling utiliza um log para registrar todas as transações que ocorrem no sistema de arquivos e preserva sua integridade em caso de falha do sistema ou queda de energia. Embora ainda seja possível perder alterações não salvas em arquivos, o journaling quase elimina completamente a possibilidade de corrupção do sistema de arquivos causada por uma desligamento incorreto. Ele também reduz ao mínimo o tempo necessário para a verificação do sistema de arquivos após uma falha. Embora o sistema de arquivos UFS utilizado pelo FreeBSD não implemente o journaling por si só, a nova classe de journaling do framework GEOM no FreeBSD 7.*X* pode ser usada para fornecer o journaling independente do sistema de arquivos. Este artigo explica como implementar o journaling do UFS em um cenário típico de um PC desktop.

Índice

1. Introdução

Enquanto servidores profissionais geralmente estão bem protegidos contra desligamentos imprevistos, os desktops típicos estão sujeitos a quedas de energia, reinicializações acidentais e outros incidentes relacionados ao usuário que podem levar a desligamentos incorretos. As Soft Updates geralmente protegem o sistema de arquivos de forma eficiente nesses casos, embora na maioria das vezes seja necessária uma verificação em segundo plano demorada. Em raras ocasiões, a corrupção do sistema de arquivos atinge um ponto em que é necessária a intervenção do usuário e dados podem ser perdidos.

O novo recurso de journaling fornecido pela GEOM pode ajudar bastante nesses cenários, praticamente eliminando o tempo necessário para a verificação do sistema de arquivos e garantindo que o sistema de arquivos seja rapidamente restaurado para um estado consistente.

Este artigo descreve um procedimento para implementar o journaling do UFS em um cenário típico de um PC desktop (um disco rígido usado tanto para o sistema operacional quanto para os dados). Ele deve ser seguido durante uma nova instalação do FreeBSD. Os passos são simples o suficiente e

não exigem interações excessivamente complexas com a linha de comando.

Depois de ler este artigo, você saberá:

- Como reservar espaço para o journaling durante uma nova instalação do FreeBSD.
- Como carregar e ativar o módulo `geom_journal` (ou compilar o suporte para ele em seu kernel personalizado).
- Como converter seus sistemas de arquivos existentes para utilizar journaling e quais opções usar no arquivo /etc/fstab para montá-los.
- Como implementar o journaling em novas partições (vazias).
- Como solucionar problemas comuns associados ao journaling.

Antes de ler este artigo, você deve ser capaz de:

- Entender os conceitos básicos do UNIX® e do FreeBSD.
- Estar familiarizado com o procedimento de instalação do FreeBSD e com a ferramenta sysinstall.

A procedimento descrito aqui destina-se a preparar uma nova instalação na qual ainda não existe nenhum dado do usuário armazenado no disco. Embora seja possível modificar e estender este procedimento para sistemas já em produção, você deve fazer um *backup* de todos os dados importantes antes de fazê-lo. Mexer com discos e partições em um nível baixo pode levar a erros fatais e a perda de dados.

2. Compreendendo o journaling no FreeBSD

O journaling fornecido pelo GEOM no FreeBSD 7.*X* não é específico do sistema de arquivos (ao contrário do sistema de arquivos ext3 no Linux®), mas funciona no nível de bloco. Embora isso signifique que ele possa ser aplicado a diferentes sistemas de arquivos, no FreeBSD 7.0-RELEASE, ele só pode ser usado no UFS2.

Essa funcionalidade é fornecida carregando o módulo geom_journal.ko no kernel (ou compilando-o em um kernel personalizado) e usando o comando `gjournal` para configurar os sistemas de arquivos. Em geral, você gostaria de registrar grandes sistemas de arquivos, como o /usr. No entanto, você precisará reservar algum espaço livre em disco (consulte a próxima seção).

Quando um sistema de arquivos é jornalizado, é necessário um espaço em disco para armazenar o próprio journal. O espaço em disco que contém os dados reais é chamado de *provedor de dados* (data provider), enquanto o que contém o journal é chamado de *provedor de journal* (journal provider). Os provedores de dados e de journal precisam estar em partições diferentes ao jornalizar uma partição existente (não vazia). Ao jornalizar uma nova partição, você tem a opção de usar um único provedor para ambos os dados e o journal. Em qualquer caso, o comando `gjournal` combina ambos os provedores para criar o sistema de arquivos final com journaling. Por exemplo:

- Você deseja fazer o journaling do sistema de arquivos /usr, armazenado em /dev/ad0s1f (que já

contém dados).

- Você reservou um espaço livre em disco em uma partição em [/dev/ad0s1g].
- Usando `gjournal`, um novo dispositivo /dev/ad0s1f.journal é criado, onde /dev/ad0s1f é o provedor de dados e /dev/ad0s1g é o provedor de journal. Esse novo dispositivo é então usado para todas as operações de arquivo subsequentes.

A quantidade de espaço em disco que você precisa reservar para o provedor de journal depende da carga de uso do sistema de arquivos e não do tamanho do provedor de dados. Por exemplo, em um desktop de escritório típico, um provedor de journal de 1 GB para o sistema de arquivos /usr será suficiente, enquanto uma máquina que lida com intensas operações E/S de disco (por exemplo, edição de vídeo) pode precisar de mais espaço. Ocorrerá um kernel panic se o espaço do journal for esgotado antes que ele tenha a chance de ser confirmado (committed).

Os tamanhos de journal sugeridos aqui são altamente improváveis de causar problemas em uso típico de desktop, como navegação na web, processamento de texto e reprodução de arquivos de mídia. Se a sua carga de trabalho inclui atividade intensa de disco, utilize a seguinte regra para obter máxima confiabilidade: o tamanho da sua memória RAM deve caber em 30% do espaço do provedor de journal. Por exemplo, se o seu sistema possui 1 GB de RAM, crie um provedor de journal de aproximadamente 3,3 GB (multiplique o tamanho da sua RAM por 3,3 para obter o tamanho do journal).

Para obter mais informações sobre o journaling, por favor, leia a página do manual gjournal(8).

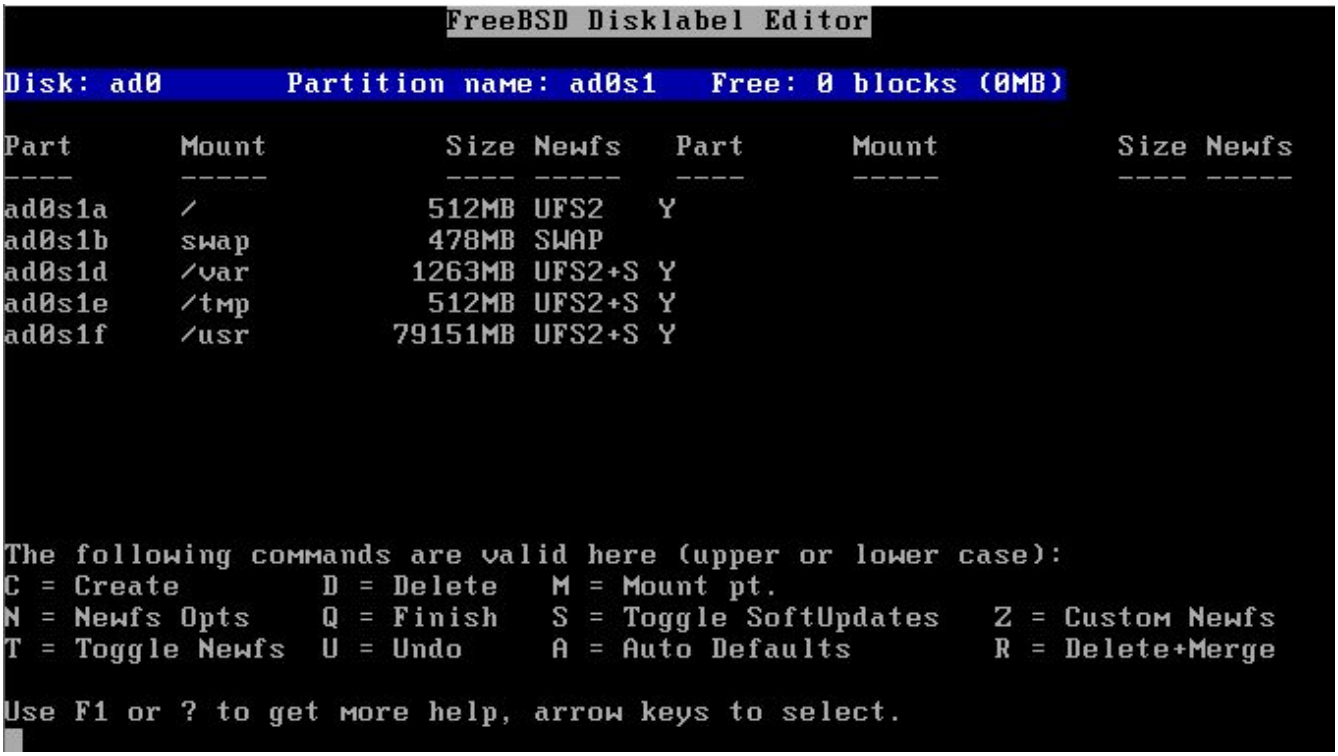
3. Etapas durante a instalação do FreeBSD

3.1. Reservando espaço para o journaling

Uma máquina desktop típica geralmente possui um único disco rígido que armazena tanto o sistema operacional quanto os dados do usuário. Argumentavelmente, o esquema de particionamento padrão selecionado pelo sysinstall é mais ou menos adequado: uma máquina desktop não precisa de uma partição /var grande, enquanto a partição /usr é alocada para a maior parte do espaço em disco, uma vez que os dados do usuário e muitos pacotes são instalados em seus subdiretórios.

A partição padrão (aquela obtida ao pressionar `A` no editor de partição do FreeBSD, chamado Disklabel) não deixa nenhum espaço não alocado. Cada partição que será jornalizada requer outra partição para o journal. Como a partição /usr é a maior, faz sentido reduzir levemente essa partição para obter o espaço necessário para o journaling.

No nosso exemplo, um disco de 80 GB está sendo utilizado. A captura de tela a seguir mostra as partições padrões criadas pelo Disklabel durante a instalação:



Se isso é mais ou menos o que você precisa, é muito fácil ajustar para o journaling. Basta usar as teclas de seta para mover o destaque para a partição /usr e pressionar `D` para excluí-la.

Agora, mova o destaque para o nome do disco no topo da tela e pressione `C` para criar uma nova partição para /usr. Essa nova partição deve ser menor em 1 GB (se você pretende jornalizar apenas /usr) ou 2 GB (se você pretende jornalizar tanto /usr quanto /var). No pop-up que aparece, opte por criar um sistema de arquivos e digite /usr como ponto de montagem.

Você deve jornalizar a partição /var? Normalmente, o journaling faz sentido em partições bastante grandes. Você pode optar por não jornalizar /var, embora fazê-lo em um desktop típico não cause problemas. Se o sistema de arquivos tiver um uso leve (o que é bastante provável para um desktop), você pode desejar alocar menos espaço em disco para o seu journal.

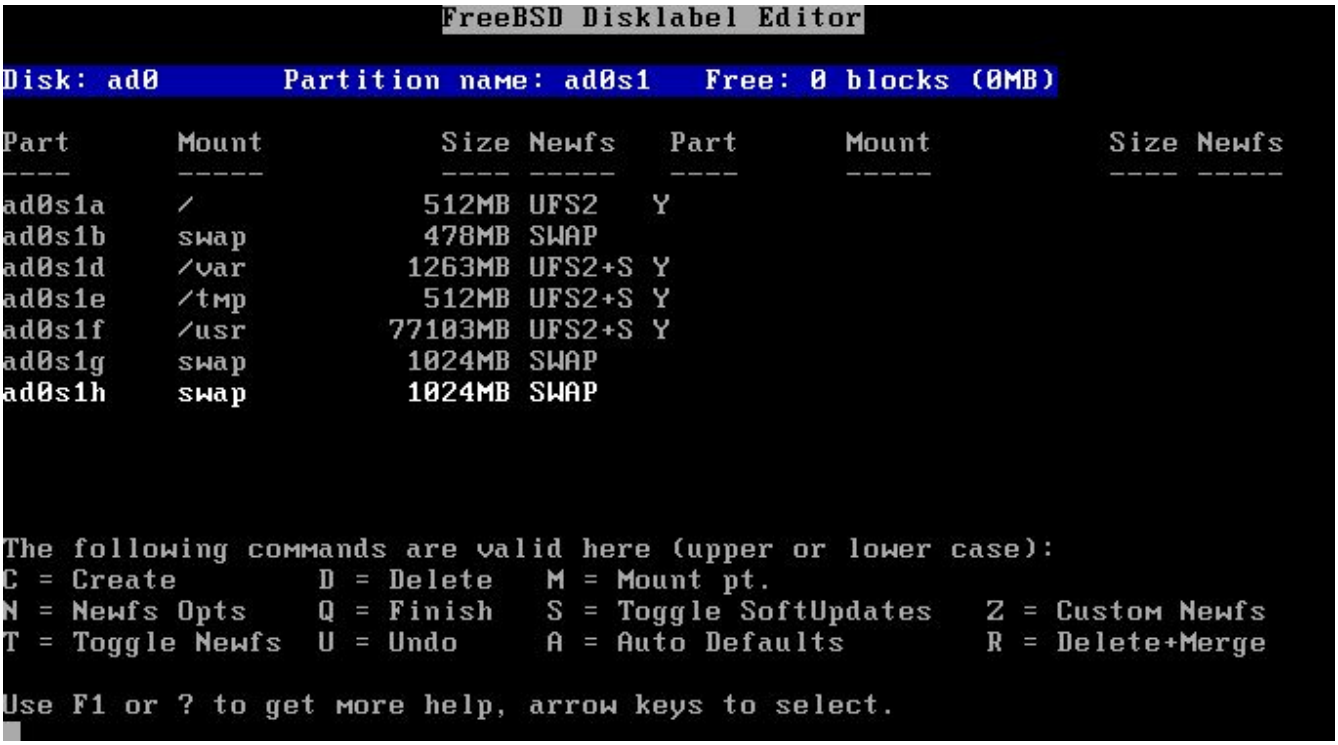
No nosso exemplo, nós aplicamos o journaling nas partições /usr e /var. Você pode, é claro, ajustar o procedimento de acordo com suas próprias necessidades.

Para facilitar o processo o máximo possível, vamos usar o sysinstall para criar as partições necessárias para o journaling. No entanto, durante a instalação, o sysinstall insiste em solicitar um ponto de montagem para cada partição que você cria. Neste momento, você não possui nenhum ponto de montagem para as partições que irão armazenar os journals e, na realidade, você nem mesmo precisa deles. Essas não são partições que serão montadas em algum lugar.

Para evitar esses problemas com o sysinstall, vamos criar as partições de journal como espaço de swap. O swap nunca é montado, e o sysinstall não tem problemas em criar quantas partições de swap forem necessárias. Após o primeiro reinício, será necessário editar o arquivo /etc/fstab e remover as entradas de espaço de swap adicionais.

Para criar a partição de swap, novamente use as teclas de seta para mover o destaque para o topo da tela do Disklabel, de modo que o próprio nome do disco seja destacado. Em seguida, pressione `N`,

insira o tamanho desejado (*1024M*) e selecione "swap space" no menu pop-up que aparece. Repita esse processo para cada journal que você deseja criar. No nosso exemplo, criaremos duas partições para fornecer os journals de /usr e /var. O resultado final é mostrado na captura de tela a seguir:



Quando você tiver concluído a criação das partições, sugerimos que anote os nomes das partições e os pontos de montagem para que você possa se referir facilmente a essas informações durante a fase de configuração. Isso ajudará a evitar erros que possam danificar sua instalação. A tabela a seguir mostra nossas anotações para a configuração de exemplo:

Tabela 1. Partições e Journals

Partição	**Ponto de Montagem**	**Journal**
ad0s1d	/var	ad0s1h
ad0s1f	/usr	ad0s1g

Continue a instalação como você normalmente faria. No entanto, sugerimos que você adie a instalação de softwares de terceiros (pacotes) até ter configurado completamente o journaling.

3.2. Inicializando pela primeira vez

Seu sistema inicializará normalmente, mas você precisará editar o arquivo /etc/fstab e remover as partições de swap extras que você criou para os journals. Normalmente, a partição de swap que você realmente usará é aquela com o sufixo "b" (por exemplo, ad0s1b em nosso exemplo). Remova todas as outras entradas de espaço de swap e reinicie para que o FreeBSD deixe de usá-las.

Quando o sistema voltar a funcionar, estaremos prontos para configurar o journaling.

4. Configurando o journaling

4.1. Executando o comando `gjournal`

Depois de preparar todas as partições necessárias, é bastante fácil configurar o journaling. Será necessário mudar para o modo de usuário único (single user mode). Para isso, faça o login como `root` e digite o seguinte comando:

```
# shutdown now
```

Pressione `Enter` para obter o shell padrão. Agora, você precisará desmontar as partições que serão jornalizadas, no nosso exemplo /usr e /var:

```
# umount /usr /var
```

Carregue o módulo necessário para o journaling:

```
# gjournal load
```

Agora, use suas anotações para determinar qual partição será usada para cada journal. No nosso exemplo, /usr é ad0s1f e seu journal será ad0s1g, enquanto /var é ad0s1d e será jornalizada em ad0s1h. Os seguintes comandos são necessários:

```
# gjournal label ad0s1f ad0s1g
GEOM_JOURNAL: Journal 2948326772: ad0s1f contains data.
GEOM_JOURNAL: Journal 2948326772: ad0s1g contains journal.

# gjournal label ad0s1d ad0s1h
GEOM_JOURNAL: Journal 3193218002: ad0s1d contains data.
GEOM_JOURNAL: Journal 3193218002: ad0s1h contains journal.
```

Se o último setor de qualquer uma das partições estiver em uso, o `gjournal` retornará um erro. Nesse caso, você precisará executar o comando usando a opção `-f` para forçar a sobrescrita. Por exemplo:

```
# gjournal label -f ad0s1d ad0s1h
```

Como esta é uma nova instalação, é altamente improvável que qualquer coisa seja realmente sobrescrita.

Neste ponto, dois novos dispositivos são criados, chamados ad0s1d.journal e ad0s1f.journal. Eles representam as partições /var e /usr que devemos montar. No entanto, antes de montá-los,

precisamos definir a flag de journaling neles e desativar a flag de Soft Updates:

```
# tunefs -J enable -n disable ad0s1d.journal
tunefs: gjournal set
tunefs: soft updates cleared

# tunefs -J enable -n disable ad0s1f.journal
tunefs: gjournal set
tunefs: soft updates cleared
```

Agora, monte manualmente os novos dispositivos em seus respectivos locais (observe que agora podemos usar a opção de montagem `async`):

```
# mount -o async /dev/ad0s1d.journal /var
# mount -o async /dev/ad0s1f.journal /usr
```

Edite o arquivo /etc/fstab e atualize as entradas para /usr e /var:

```
/dev/ad0s1f.journal     /usr            ufs     rw,async      2       2
/dev/ad0s1d.journal     /var            ufs     rw,async      2       2
```

Certifique-se de que as entradas acima estão corretas ou você terá problemas para inicializar normalmente após o reboot!

Por fim, edite o arquivo /boot/loader.conf e adicione a seguinte linha para carregar o módulo gjournal(8) em cada inicialização do sistema:

```
geom_journal_load="YES"
```

Parabéns! Seu sistema está agora configurado para o journaling. Você pode digitar `exit` para retornar ao modo multiusuário ou reiniciar para testar sua configuração (recomendado). Durante a inicialização, você verá mensagens como as seguintes:

```
ad0: 76293MB XEC XE800JD-00HBC0 08.02D08 at ata0-master SATA150
GEOM_JOURNAL: Journal 2948326772: ad0s1g contains journal.
GEOM_JOURNAL: Journal 3193218002: ad0s1h contains journal.
GEOM_JOURNAL: Journal 3193218002: ad0s1d contains data.
GEOM_JOURNAL: Journal ad0s1d clean.
GEOM_JOURNAL: Journal 2948326772: ad0s1f contains data.
GEOM_JOURNAL: Journal ad0s1f clean.
```

Após um encerramento não limpo, as mensagens variam ligeiramente, ou seja:

```
GEOM_JOURNAL: Journal ad0s1d consistent.
```

Isso geralmente significa que o gjournal(8) utilizou as informações no journal provider para restaurar o sistema de arquivos a um estado consistente.

4.2. Fazendo journaling de partições recém-criadas

Enquanto o procedimento acima é necessário para jornalizar partições que já contêm dados, jornalizar uma partição vazia é um pouco mais fácil, pois tanto o provedor de dados quanto o provedor de journal podem ser armazenados na mesma partição. Por exemplo, suponha que um novo disco tenha sido instalado e uma nova partição /dev/ad1s1d tenha sido criada. Criar o journal seria tão simples quanto:

```
# gjournal label ad1s1d
```

O tamanho do journal será de 1 GB por padrão. Você pode ajustá-lo usando a opção `-s`. O valor pode ser dado em bytes, ou pode ser seguido por `K`, `M` ou `G` para representar Kilobytes, Megabytes ou Gigabytes, respectivamente. Note que o `gjournal` não permitirá que você crie tamanhos de journal excessivamente pequenos e inadequados.

Por exemplo, para criar um journal de 2 GB, você poderia usar o seguinte comando:

```
# gjournal label -s 2G ad1s1d
```

Em seguida, você pode criar um sistema de arquivos na sua nova partição e habilitar o journaling usando a opção `-J`:

```
# newfs -J /dev/ad1s1d.journal
```

4.3. Adicionando suporte ao journaling no seu kernel personalizado

Se você não deseja carregar `geom_journal` como um módulo, você pode incorporar suas funções diretamente no seu kernel. Edite o arquivo de configuração do seu kernel personalizado e verifique se ele inclui as seguintes linhas:

```
options UFS_GJOURNAL # Note: This is already in GENERIC

options GEOM_JOURNAL # You will have to add this one
```

Recompile e reinstale o seu kernel seguindo as instruções relevantes no Handbook do FreeBSD

Não se esqueça de remover a entrada relevante de "load" do arquivo /boot/loader.conf se você a tiver usado anteriormente.

5. Solução de problemas com journaling

A seção a seguir aborda as perguntas mais frequentes relacionadas a problemas relacionados ao journaling.

5.1. Estou recebendo um kernel panic durante períodos de alta atividade de disco. Como isso está relacionado ao journaling?

O journal provavelmente fica cheio antes de ter a chance de ser confirmado (gravado) no disco. Lembre-se de que o tamanho do journal depende da carga de uso e não do tamanho do provedor de dados. Se a atividade do disco for intensa, será necessário uma partição maior para o journal. Consulte a nota na seção Compreendendo o Journaling para mais informações.

5.2. Eu cometi algum erro durante a configuração e não consigo inicializar normalmente agora. Isso pode ser resolvido de alguma forma?

Parece que você esqueceu (ou digitou incorretamente) a entrada no arquivo /boot/loader.conf ou existem erros no arquivo /etc/fstab. Esses erros geralmente são fáceis de corrigir. Pressione `Enter` para acessar o shell do sistema no modo de usuário unico. Em seguida, localize a raiz do problema:

```
# cat /boot/loader.conf
```

Se a entrada `geom_journal_load` estiver ausente ou digitada incorretamente, os dispositivos com journaling não serão criados. Carregue o módulo manualmente, monte todas as partições e continue com a inicialização em modo multiusuário:

```
# gjournal load

GEOM_JOURNAL: Journal 2948326772: ad0s1g contains journal.
GEOM_JOURNAL: Journal 3193218002: ad0s1h contains journal.
GEOM_JOURNAL: Journal 3193218002: ad0s1d contains data.
GEOM_JOURNAL: Journal ad0s1d clean.
GEOM_JOURNAL: Journal 2948326772: ad0s1f contains data.
GEOM_JOURNAL: Journal ad0s1f clean.

# mount -a
# exit
(boot continues)
```

Por outro lado, se essa entrada estiver correta, verifique o arquivo /etc/fstab. Provavelmente você encontrará uma entrada ausente ou digitada incorretamente. Nesse caso, monte todas as partições restantes manualmente e continue com a inicialização em modo multiusuário.

5.3. Posso remover o registro no journal e retornar ao meu sistema de arquivos padrão com o Soft Updates?

Claro. Use o seguinte procedimento, que reverte as alterações. As partições que você criou para os provedores de journal podem ser usadas para outros fins, se desejar.

Faça login como `root` e altere para o modo de usuário único:

```
# shutdown now
```

Desmonte as partições journaled:

```
# umount /usr /var
```

Sincronize os journals:

```
# gjournal sync
```

Pare os provedores de journaling:

```
# gjournal stop ad0s1d.journal
# gjournal stop ad0s1f.journal
```

Limpe os metadados de journaling de todos os dispositivos usados:

```
# gjournal clear ad0s1d
# gjournal clear ad0s1f
# gjournal clear ad0s1g
# gjournal clear ad0s1h
```

Limpe o sinalizador de journaling do sistema de arquivos e restaure a flag do Soft Updates:

```
# tunefs -J disable -n enable ad0s1d
tunefs: gjournal cleared
tunefs: soft updates set

# tunefs -J disable -n enable ad0s1f
tunefs: gjournal cleared
```

```
tunefs: soft updates set
```

Remonte os dispositivos antigos à mão:

```
# mount -o rw /dev/ad0s1d /var
# mount -o rw /dev/ad0s1f /usr
```

Edite o arquivo /etc/fstab e restaure-o para seu estado original:

```
/dev/ad0s1f     /usr            ufs     rw      2       2
/dev/ad0s1d     /var            ufs     rw      2       2
```

Finalmente, edite o arquivo /boot/loader.conf, remova a entrada que carrega o módulo `geom_journal` e reinicie o sistema.

6. Leitura Adicional

O journaling é um recurso relativamente novo no FreeBSD e, portanto, ainda não está muito bem documentado. No entanto, você pode encontrar as seguintes referências adicionais úteis:

- A nova seção sobre journaling agora faz parte do FreeBSD Handbook.
- Esta mensagem na lista de discussão do FreeBSD-CURRENT enviada por um desenvolvedor do gjournal(8)'s, `Paweł Jakub Dawidek <pjd@FreeBSD.org>`.
- Esta mensagem na lista de discussão para perguntas gerais sobre o FreeBSD enviada por `Ivan Voras <ivoras@FreeBSD.org>`.
- As páginas de manual gjournal(8) e geom(8).