

Implementando o UFS Journaling em um Desktop PC

Resumo

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

Índice

1. Introdução	1
2. Compreendendo o journaling no FreeBSD	2
3. Etapas durante a instalação do FreeBSD	3
4. Configurando o journaling	6
5. Solução de problemas com journaling	9
6. Leitura Adicional	11

1. Introdução

Embora os servidores profissionais estejam geralmente bem protegidos contra desligamentos imprevistos, um desktop típico fica à mercê de falhas de energia, reinicializações acidentais e outros incidentes relacionados ao usuário que podem levar a paradas abruptas. Os soft updates costumam proteger o sistema de arquivos de maneira eficiente nestes casos, embora na maioria das vezes seja necessária uma longa verificação em background. Em raras ocasiões, a corrupção do sistema de arquivos atinge um ponto em que a intervenção do usuário é necessária e os 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 PC de mesa (um único disco rígido usado para o sistema operacional e para os dados). Deve ser seguido durante uma nova instalação do FreeBSD. As etapas são simples o suficiente e não

requerem interação excessivamente complexa 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 como compilar o suporte para ele em seu kernel customizado).
- Como converter seus sistemas de arquivos existentes para utilizar o journaling e quais opções usar em `/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 o utilitário `sysinstall`.



O procedimento descrito aqui é destinado a preparar uma nova instalação na qual ainda não temos nenhum dado real do usuário é armazenado no disco. Embora seja possível modificar e estender este procedimento para sistemas já em produção, você deve efetuar o *backup* de todos os dados importantes antes de fazer isso. Mexer com discos e partições em um baixo nível 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 (diferentemente do sistema de arquivos `ext3` no Linux®), funcionando a nível de bloco. Embora isso signifique que ele possa ser aplicado a sistemas de arquivos diferentes, no FreeBSD 7.0-RELEASE, ele só pode ser usado com o UFS2.

Esta funcionalidade é fornecida pelo carregamento do módulo `geom_journal.ko` no kernel (ou através da compilação de um kernel personalizado) e pelo uso do comando `gjournal` para configurar os sistemas de arquivos. Em geral, você gostaria de utilizar o journal em grandes sistemas de arquivos, como o `/usr`. Você precisará no entanto (veja a seção seguinte) reservar algum espaço livre em disco para isso.

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

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

já contém dados).

- Você reservou algum espaço livre no disco, na partição `/dev/ad0s1g`.
- Usando o comando `gjournal`, um novo dispositivo `/dev/ad0s1f.journal` é criado no qual o `/dev/ad0s1f` é o data provider, e o `/dev/ad0s1g` é o journal provider. Este novo dispositivo é então usado para todas as operações de arquivo posteriores.

A quantidade de espaço em disco que você precisa reservar para o journal provider depende da carga de uso do sistema de arquivos e não do tamanho do data provider. Por exemplo, em um desktop típico de escritório, um journal provider de 1 GB para o sistema de arquivos `/usr` será suficiente, enquanto uma máquina que lida com I/O de disco pesado (por exemplo, edição de vídeo) pode precisar de mais. Um kernel panic ocorrerá se o espaço do journal estiver esgotado antes de ter a chance de ser committed.



É improvável que os tamanhos de journal sugeridos aqui causem problemas no uso típico de um desktop (como navegação na Web, processamento de texto e reprodução de arquivos de mídia). Se sua carga de trabalho incluir intensa atividade de disco, use a regra a seguir para obter a confiabilidade máxima: o tamanho da RAM deve caber em 30% do espaço do journal provider. Por exemplo, se o seu sistema tiver 1 GB de RAM, crie um journal provider de aproximadamente 3,3 GB. (Multiplique o tamanho total da sua RAM por 3.3 para obter o tamanho do journal).

Para mais informações sobre journaling, leia a página de manual do [gjournal\(8\)](#).

3. Etapas durante a instalação do FreeBSD

3.1. Reservando espaço para o journaling

Normalmente, um desktop típico tem um disco rígido que armazena o sistema operacional e os dados do usuário. Indiscutivelmente, o esquema de particionamento padrão selecionado pelo `sysinstall` é mais ou menos adequado: Um desktop não precisa de uma grande partição `/var`, enquanto o `/usr` é alocado com a maior parte do espaço em disco, já que os dados do usuário e muitos pacotes são instalados em seus subdiretórios.

O particionamento padrão (aquele obtido pressionando **A** no editor de partições do FreeBSD, chamado `Disklabel`) não deixa nenhum espaço não alocado. Cada partição que será `journal`ed, requer outra partição para `journal`. Como a partição `/usr` é a maior, faz sentido reduzir ligeiramente essa partição, para obter o espaço necessário para o `journaling`.

No nosso exemplo, um disco de 80 GB é usado. A captura de tela a seguir mostra as partições padrões criadas por `Disklabel` durante a instalação:

```

FreeBSD Disklabel Editor

Disk: ad0      Partition name: ad0s1      Free: 0 blocks (0MB)

Part      Mount      Size Newfs      Part      Mount      Size Newfs
----      -
ad0s1a    /           512MB UFS2      Y
ad0s1b    swap        478MB SWAP
ad0s1d    /var        1263MB UFS2+S   Y
ad0s1e    /tmp        512MB UFS2+S   Y
ad0s1f    /usr        79151MB UFS2+S   Y

The following commands are valid here (upper or lower case):
C = Create      D = Delete      M = Mount pt.
N = Newfs Opts  Q = Finish      S = Toggle SoftUpdates  Z = Custom Newfs
T = Toggle Newfs  U = Undo        A = Auto Defaults      R = Delete+Merge

Use F1 or ? to get more help, arrow keys to select.

```

Se isso é mais ou menos o que você precisa, é muito fácil se ajustar ao journaling. Simplesmente use as teclas de seta para mover o realce para a partição /usr e pressione **D** para excluí-la.

Agora, mova o realce para o nome do disco na parte superior da tela e pressione **C** para criar uma nova partição para /usr. Esta nova partição deve ser menor em 1 GB (se você pretende registrar apenas /usr), ou 2 GB (se você pretende registrar ambos /usr e /var). No pop-up exibido, opte por criar um sistema de arquivos e digite /usr como o ponto de montagem.



Você deve fazer o journal da partição /var? Normalmente, o journaling faz sentido em partições grandes. Você pode decidir não fazer o journal do /var, embora fazê-lo em um desktop típico não cause nenhum dano. Se o sistema de arquivos é usado levemente (bastante provável para um desktop) você pode querer alocar menos espaço em disco para o seu journal.

Em nosso exemplo, nós fizemos o journal em ambos /usr e /var. Você pode, naturalmente, ajustar o procedimento às suas próprias necessidades.

Para manter as coisas o mais fáceis o possível, vamos usar o sysinstall para criar as partições necessárias para o journaling. No entanto, durante a instalação, o sysinstall insiste em pedir um ponto de montagem para cada partição criada. Neste ponto, você não tem nenhum ponto de montagem para as partições que irão manter os journals, e na realidade você *nem precisa deles*. Estas não são partições que iremos montar em algum lugar.

Para evitar esses problemas com o sysinstall, vamos criar as partições de journal como espaço de troca. O swap nunca é montado, e o sysinstall não tem problemas para criar tantas partições de troca quantas forem necessárias. Após a primeira reinicialização, o /etc/fstab terá que ser editado, e as entradas extras do espaço de troca serão removidas.

Para criar o swap, use novamente as teclas de seta para mover o realce para a parte superior da tela do Disklabel, para que o nome do disco seja realçado. Em seguida, pressione **N**, insira o

tamanho desejado (1024M) e selecione "swap space" no menu pop-up exibido. Repita para cada journal que você deseja criar. Em nosso exemplo, criamos duas partições para fornecer os diários de /usr e /var. O resultado final é mostrado na seguinte captura de tela:

```

FreeBSD Disklabel Editor
Disk: ad0 Partition name: ad0s1 Free: 0 blocks (0MB)
Part      Mount      Size Newfs  Part      Mount      Size Newfs
----      -
ad0s1a    /           512MB UFS2   Y
ad0s1b    swap        478MB SWAP
ad0s1d    /var        1263MB UFS2+S Y
ad0s1e    /tmp        512MB UFS2+S Y
ad0s1f    /usr        77103MB UFS2+S Y
ad0s1g    swap        1024MB SWAP
ad0s1h    swap        1024MB SWAP

The following commands are valid here (upper or lower case):
C = Create      D = Delete      M = Mount pt.
N = Newfs Opts  Q = Finish      S = Toggle SoftUpdates  Z = Custom Newfs
T = Toggle Newfs U = Undo        A = Auto Defaults      R = Delete+Merge

Use F1 or ? to get more help, arrow keys to select.

```

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

Tabela 1. Partições e Journals

Partições	Ponto de montagem	Journal
ad0s1d	/var	ad0s1h
ad0s1f	/usr	ad0s1g

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

3.2. Inicializando pela primeira vez

Seu sistema irá iniciar normalmente, mas você precisará editar o /etc/fstab para remover as partições extras de swap que você criou para os journals. Normalmente, a partição swap que você irá usar é aquela com o sufixo "b" (por exemplo, ad0s1b no nosso exemplo). Remova todas as outras entradas de espaço swap e reinicialize para que o FreeBSD pare de usá-las.

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

4. Configurando o journaling

4.1. Executando o **gjournal**

Tendo preparado todas as partições requeridas, é bastante fácil configurar o journaling. Nós precisaremos mudar para o modo de single user, então entre como **root** e digite:

```
# shutdown now
```

Pressione **Enter** para obter o shell padrão. Nós precisaremos desmontar as partições que serão registradas no diário, 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 diário. Em nosso exemplo, /usr é ad0s1f e seu journal será ad0s1g, enquanto /var é ad0s1d e será journaled para 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 partição for usado, o **gjournal** retornará um erro. Você terá que executar o comando usando o sinalizador **-f** para forçar uma substituição, ou seja:



```
# 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, a saber ad0s1d.journal e ad0s1f.journal. Os quais representam as partições /var e /usr que temos que montar. Antes de montar, devemos definir o flag de Journal e limpar o flag de Soft Updates:

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

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

Agora, monte os novos dispositivos manualmente em seus respectivos locais (note 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 `/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!

Finalmente, edite o `/boot/loader.conf` e adicione a seguinte linha para que o módulo `gjournal(8)` seja carregado em cada boot:

```
geom_journal_load="YES"
```

Parabéns! Seu sistema está agora configurado para journaling. Você pode digitar `exit` para retornar ao modo multiusuário ou reinicializar 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)` usou as informações no journal provider para retornar o sistema de arquivos a um estado consistente.

4.2. Fazendo journaling de partições recém-criadas

Embora o procedimento acima seja necessário para partições que fazem uso de journaling e que já contêm dados, o journaling de uma partição vazia é um pouco mais fácil, uma vez que os dados e o journal provider 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 1 GB por padrão. Você pode ajustá-lo usando a opção `-s`. O valor pode ser dado em bytes, ou acrescentado por `K`, `M` ou `G` para indicar Kilobytes, Megabytes ou Gigabytes, respectivamente. Note que o comando `gjournal` não permitirá que você crie journals de tamanhos pequenos e inadequados.

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

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

Você pode criar um sistema de arquivos em sua nova partição e ativar 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 o `geom_journal` como um módulo, você pode construir suas funções diretamente em seu kernel. Edite seu arquivo de configuração do kernel personalizado e verifique se ele inclui estas duas linhas:

```
options UFS_GJOURNAL # Note: This is already in GENERIC
options GEOM_JOURNAL # You will have to add this one
```

Recompile e reinstale seu kernel seguindo as instruções [relevantes no Handbook do FreeBSD](#).

Não se esqueça de remover a entrada relevante "load" do `/boot/loader.conf` se você a usou 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 se enche antes que ele tenha a chance de ser enviado (descarregado) para o 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 alta, você precisará de uma partição maior para o journal. Veja a nota na seção [Compreendendo o journaling no FreeBSD](#).

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

Você esqueceu (ou escreveu incorretamente) a entrada em `/boot/loader.conf`, ou existem erros no seu arquivo `/etc/fstab`. Estes erros geralmente são fáceis de corrigir. Pressione `Enter` para acessar o shell padrão do modo single user. Em seguida, localize a raiz do problema:

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

Se a entrada `geom_journal_load` estiver ausente ou incorreta, os dispositivos registrados nunca serão criados. Carregue o módulo manualmente, monte todas as partições e continue com a inicialização do modo multi usuário:

```
# 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)
```

Se, por outro lado, esta entrada estiver correta, dê uma olhada em `/etc/fstab`. Você provavelmente encontrará uma entrada incorreta ou faltando. Nesse caso, monte todas as partições restantes manualmente e continue com o boot em modo multi-usuários.

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

Certo. Use o procedimento a seguir, que inverte as alterações. As partições que você criou para os provedores de journal podem ser usadas para outros propósitos, se você desejar.

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

```
# shutdown now
```

Desmonte as partições journalled:

```
# 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:

```
# tune2fs -J disable -n enable ad0s1d  
tune2fs: gjournal cleared  
tune2fs: soft updates set  
  
# tune2fs -J disable -n enable ad0s1f  
tune2fs: gjournal cleared  
tune2fs: soft updates set
```

Remonte os dispositivos antigos à mão:

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

Edite o `/etc/fstab` e restaure-o ao seu estado original:

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

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

6. Leitura Adicional

Journaling é um recurso relativamente novo do FreeBSD e, como tal, ainda não está muito bem documentado. Você pode, no entanto, encontrar as seguintes referências adicionais úteis:

- Uma [nova seção sobre journaling](#) agora faz parte do Handbook do FreeBSD.
- [Este post](#) em [freebsd-current](#) pelo desenvolvedor do [gjournal\(8\)](#), Paweł Jakub Dawidek pjd@FreeBSD.org.
- [Este post](#) em [freebsd-questions](#) por Ivan Voras ivoras@FreeBSD.org.
- As páginas de manual do [gjournal\(8\)](#) e [geom\(8\)](#).