Version: 2.0.0
Last Updated: 05 Sept 1996
Author: _Timothy J. Luoma_
Converted to HTML: 5 Sept 1996

_

Q: What is this?
A: The first-known attempt to create a single resource of information concerning the concept of a "swapdisk". That is, a separate space set aside for swapping.

Q: Is this guaranteed? Is it exhaustive? Will it answer all of my questions?
A: NO! No. Probably not. It is an attempt to answer some of the questions which have been asked and answered REPEATEDLY in the past.

This is an ongoing process, and I'll probably be expanding it as new information comes in. I'm not an expert at this. Everything I know is contained in this document.

_

References:

> *The author of this FAQ assumes that you have read the following reference materials BEFORE reading this document. In fact, you may very well find out that your question and/or problem is solved just by reading the NeXTAnswers on the subject. If you find a term which you don't understand, it probably means you didn't read something from the list below. I have tried to write this in as "plain English" as possible, but at some point one must use the jargon in order to keep from having to explain each detail fully with long explanitary notes.... like this one...*

Optimizing Virtual Memory with swaptab (NeXTAnswer #1020) ***A MUST READ!!!!!!!!!!!***

Swabtab High Water Mark (hiwat) Ignored (NeXTAnswer #1386)
Swapfile size issues (NeXTAnswer #1388)
Swapdisk no longer functions (NeXTAnswer #1387)
man fstab (NeXTAnswer #2050)
man swaptab
man mach_swapon

There was also a very long discussion about swapdisks in comp.sys.next.sysadmin from July-November 1995. You can find archives of the on several of the NeXT archives.

Well, if you're all ready, go to the _Question and Answer_ section. (28k)

Version: 2.0
Last Updated: 05 Sept 1996
Author: _Timothy J. Luoma_
Converted to HTML: 5 Sept 1996

_

The Swapfile and Swapdisk FAQ Q&A

SWAPFILE

SWAPDISK

- _Q: What is a swapdisk?_
- _Q: Why would I want a swapdisk?_
- _Q: What are the problems with the 'swapdisk' method?_
- _Q: Any reason I shouldn't use the free space on my swapdisk?_
- _Q: What kinds of drives would be good for use as swapdisks?_
- _Q: Is a separate swapdrive worth the cost?_
- _Q: How can I fine-tune my swapdisk's performance?_
- _Q: How do I set aside a partition for swapspace?_

_

What is the swapfile?

> The swapfile is the virtual memory for the NeXT. It usually is mounted on /private/vm/swapfile. When processes need memory, it pages in and out to the swapfile.

_

What can I do to reclaim the swapfile space?

> Reboot your computer. There is no other way to reclaim swap space. If your computer needs to be rebooted often, you may want to consider a cron job using /usr/etc/shutdown -r.
>
> Some have suggested that logging out completely and then logging in as 'exit' (which restarts the WindowServer) might help.

_

Can't I just delete the swapfile?

> NO! This is among the worst of the "Very Bad Ideas" which one might have, ranking up there with putting magnets on your hard drive, eating potato chips over your motherboard, and dousing yourself with gasoline before smoking.

_

Why doesn’t the swapfile ever shrink?

Theoretically, it should. However, here is how it seems to work *(if anyone has a better way of explaining this, please _email me_)*

You start some app (let’s say Edit.app). Edit.app requests some part of the swapfile, and is given some. Then you start Webster.app to check a word, and Webster.app requests a part of the swapfile. Basically now the Edit.app part of the swapfile is tra pped underneath the Webster.app part. In a perfect world, one would think that by quitting Webster.app and then Edit.app, one would return to the same size swapfile as one had before starting them. Unfortunately it doesn’t work that way, and so the swap file grows. It’s a poor explanation, but you get the idea....

_

What can I do to slow the growth of the swapfile?

Decrease the memory demand on your computer. The best way is to add RAM which will help, but not solve, the problem.

_

What are the lowat and hiwat?

LOWAT is the "low water mark" and HIWAT is the "high water mark". The LOWAT is the LOWEST the swapfile will ever grow, and the HIWAT is the highest the swapfile will ever grow.

_

Should I make the lowat lower than the default?

NO. The LOWAT should not be set lower than the default size of 16777216. Lower settings can cause excessive fragmentation.

_

Should I make the lowat higher than the default?

Some people choose to set a higher LOWAT because a contiguous swap zone gives better performance that a fragmented swap zone, since it cuts done the amount of seeking to non-adjacent cylinders.

_

How do I calculate a hiwat or lowat size?

Multiply the number of megabytes you want by 1048576. For example, if you want a 50 megabyte LOWAT, set LOWAT to 52428800 (which is 50 * 1048576)

HIWAT NOTE: If you use compression on the swapfile, the HIWAT is checked against the size of the 'swapfile.front' file, which is not really a file but a virutal file, so some guessing has to be done as to how high the HIWAT can/should be. (The HIWAT behavior regarding compressed swapfiles is considered a bug, at least by the author.)

_

I set a new, higher lowat, but when I rebooted, the swapfile was SMALLER than the lowat, what happened?

The LOWAT is only the LOWEST that the swapfile will be trimmed down TO. Rebooting will never make the swapfile LARGER.

See also: How do I create a swapfile larger than the default?

_

How do I create a swapfile larger than the default?

The best way is to boot into single-user mode, delete the old swapfile and use mkfile (8). Then reboot.

_

Why is /dev/sd0a identical to swapfile?

While I did 'df' on my NeXTstation which I loaded very recently with NS 3.3, I got the following:

```
Filesystem...........kbytes....used...avail.capacity..Mounted.on.
/private/vm/swapfile..402253..294450...67577....81%..../private/vm/swapfile.front
/dev/sd0a............402253..294450...67577....81%..../
```

You need not worry. This is normal. It is telling you that your swapfile resides on your primary SCSI disk. If you have a secondary hard drive, you can make it a swapdisk or put the swapfile on it. Depending on the speed of the secondary drive, it should improve performance and provide a small degree of protection should your disk ever fill up.

_

Should I set a hiwat?

Opinions vary. Some people say that you should, because it will keep your system from filling up, which can be bad (causing system panics and fragmentation). Others say that you shouldn't, because if the swapfile grows to the HIWAT the system will stop until space becomes available.

_

What happens when my swapfile hits the hiwat?

From: _Chuck Swiger_

Your system won't be able to do page outs. You'll get a ton of warning messages via syslogd, process forking will start to fail, requests to malloc will start to fail, and so forth. The WindowServer is pretty likely to crash under these circumstances, and it's possible for the kernel itself to panic.

But you do have some flexibility since the system can deal with these circumstances to some extent, and you can sometimes kill a large process to recover.

From: _Charles Fu_

Unless you have another swapfile (on another disk). IMHO, the high water mark is mostly useful for limiting the amount of space used on particular disks when you have multiple swapfiles. Unfortunately, the high water mark is not as useful if you are using a compressed swapfile (since it regulates the size _prior_ to compression--plus the "compressed" swapfile can be larger than the uncompressed version in some not so uncommon circumstances).

_

What is the swapfile.front?

swapfile.front is the file which swapfile uses when you have enabled "compression" on your swapfile. The swapfile.front file does not actually exist on the system, it is a virtual file.

Note: only ONE swapfile can be compressed

See also: man swaptab

_

How can I change the size of swapfile.front?

The size of 'swapfile.front' completely does not matter. It is not really a file which exists on your computer, it is a virtual file. Only the size of 'swapfile' matters.

_

How can I force swapfile compression?

Q: Is there any way to force swapfile compression to happen? I don't appear to have swapfile compression happening (there is no swapfile.front), and I want to enable it. I don't have it specifically turned off in /etc/swaptab or /etc/rc, so why isn't it compressing?

A: It should be automatic. However, only ONE swapfile can be compressed.

_

Why won't my secondary swapfile compress?

Only ONE swapfile can be compressed. Feel free to consider this a bug.

_

NOTE: hiwat and compression

If there is a HIWAT set on a swapfile which is being compressed, the HIWAT will be reached with the COMPRESSED swapfile (ie swapfile.front) reaches the HIWAT level.

Since the 'swapfile.front' is not really a file, but rather a virtual file, this behavior seems a little odd, and one is left to wonder why it acts this way.

_

What is a swapdisk?

A swapdisk is a separate disk used solely for swapping.

In the early days, NeXT used to ship a 40meg disk with their hardware, to be used for swapping purposes only. For ease of operation and installation, the 'rc' files (in /etc) are setup to recognize a secondary disk labelled "swapdisk" as a disk set aside for that purpose.

A disk which is NOT labelled 'swapdisk' can be used, if one wants to setup the appropriate parameters and configure the appropriate files (see below for an outline of how this is done). In the author's opinion, however, it is easier/better/safer to use a disk named 'swapdisk' for the express purposes of swapping alone. IF you take the time to read through this document and the files it references, you can configure your 'swapdisk' for the size drive/parition you have, and allow the rc script to automatically handle the mounting, etc.

If you do not read and understand what is going on, you will very likely NOT get the best performance out of your swapdisk. In other words, read this, I spent a good deal of time on it to make it as clear as possible.

For the purposes of this document, the word "swapdisk" will be used to mean any disk or partition set aside for the purposes of swapping, unless specifically stated otherwise.

NOTE: If the disk is labelled 'swapdisk' then 'rc.swap' and NOT /etc/swaptab controls the setup of the swapfile (lowat,hiwat,nocompress,etc,etc).

Why would I want a swapdisk?

The theory behind the swapdisk concept is that a separate disk set aside for swapping increases performance, because while reading/writing/etc is being done to the primary hard drive, all of the swapping is being handled by a second drive. This is commonly referred to as the "double spindle" advantage, because with a swapdisk you have two hard drives spinning at once, splitting up the work. (Note: setting aside a partition on your primary drive solely for swapping will NOT give you the double-spindle advantage, which is hopefully obvious, but I thought warranted mentioning.)

Having a swapdisk also means that any fragmentation which is caused by the swapfile is contained to a single area, which can be advantageous should the swapfile grow too large and cause pageout errors (which occurs when some process tries to request space in the swapfile but there are none left to give out).

What are the problems with the 'swapdisk' method?

[Note: in this section, when I write "swapdisk" I mean specifically a disk with the label "swapdisk"]

One man's preference is another man's bug when it comes to the 'swapdisk' setup. For example, by default the swapdisk's swapfile has a HIWAT of approximately 30 megabytes. It has been correctly pointed out that this may be foolish today when some people are buying GIG drives to use for swapdisks.

Also, when the HIWAT is reached on the swapdisk's swapfile, swapping will begin on /private/vm/swapfile. Again, this could be seen as a bug or a feature, depending on your point of view. [Note: this can be prevented by editing out the /etc/swaptab entry for /private/vm/swapfile]

One of the more serious issues is the location of the /tmp. By default, rc.swap sets the /tmp to be located on the 'swapdisk' rather than on the primary drive. Realize that any space used by /tmp is not available for swapping, and MANY applications use /tmp to dump scratch files. There's an advantage to having /tmp be on another drive if you write a lot of scratch files and want to get the "double spindle" effect on that also. [Note: This can be changed by editing /etc/rc.swap and changing all the references to /private/swapdisk/tmp to /private/tmp]

Chuck Swiger and _Garance A Drosehn_ have differing opinions on the matter, each with certain "good points" which I think are note worthy. I have included a post from Chuck (which has a quotation from an earlier posting by Garance) and then a followup by Garance after Chuck's post.

> While naming the disk "swapdisk" does trigger automatic processing, I (for one) do not recommend doing that. The automatic processing that's triggered by the swapdisk name was designed for a different era. It thinks you've got a 40-meg hard disk for swapping and /tmp space, from the days when you were running the system off an optical cartridge.

That's true, although it's not a problem, either. If you don't want the high watermark on the second drive to be the default value of 40 MB, edit the mach_swapon line in /etc/rc.swap to something like:

/usr/etc/mach_swapon -v -o prefer,lowat=33554432,hiwat=131457280,nocompress
$NEWSWAPFILE >/dev/console 2>&1

> I recommend that people figure out swaptab and fstab instead of tripping upon that older swapdisk logic.

Well, the problem with that is an error made in /etc/fstab will result in an unbootable system. I agree with you that people should understand how to deal with these files, so if you're motivated why don't you write up a step-by-step guide?

Using the swapdisk method takes one step, two if you wish to adjust the swapfile size:

1) /usr/etc/disk -L swapdisk /dev/rsd1a (or whatever device it is)
2) [optional] edit /etc/rc.swap as described above to change the hiwat.

Here is Garance's response

Editing rc.swap is no less dangerous than editting /etc/fstab. In either case, you can work your way around any errors by booting in single-user mode. And, btw, the default high-water mark is 30-meg (which is picked because the code thinks you have a 40-meg drive, and it wants to save 10-meg for /tmp).

I've received a few messages via email asking me why I offer such a strange recommendation. Here's some comments I wrote up for the NS/Intel homebrew mailing list, including a few extra points that I have learned since writing those comments.

For what (little) it's worth, I avoid the special /swapdisk processing. If I have disk or partition I want to use as swap, I'd rather set it up by hand (using /etc/swaptab) and make sure I avoid the /etc/rc.swap magic.

For one thing, if there is no swapfile on that /swapdisk, then /etc/rc.swap just *touches* the file to create it. Oh joy. I'd much rather have something do a mkfile at a reasonable size, instead of starting my swapfile at zero bytes. If the swapfile was really going to stay at zero bytes, then I wouldn't need a whole new disk to hold it...

For two, it also starts up swapping with: /usr/etc/mach_swapon -v -o prefer,lowat=16777216,hiwat=31457280

So, you have a grand total of 30meg of swapspace. This made sense for the original expectation of /swapdisk's, but now that I'm buying 1-gig disks it seems pretty silly to pretend my swapfile has to stop growing at 30meg. The *smallest* NeXTSTEP machine I run has 20meg of RAM, and all the rest have 32meg or more of RAM. Chances are that 32-meg of swapspace (as a high-water mark!) is going to fall far short of my needs.

Note that this 30-meg highwater mark is on a *preferred* swap file. You'll still be using whatever swapfiles are listed in /etc/swaptab in *addition* to this 30meg swapfile. So, what happens when you start paging to megabyte 31? All swapspace after megabyte 30 is going to be using up space back on your root hard disk. If you're bothering with a swapdisk at all, then you're probably doing it because you don't have *room* on your root partition. I've had people create swapdisks like this, and then still have system crashes because their root partition runs out of disk space.

For three, it also creates the /tmp directory on the /swapdisk. I generally want the /tmp space on the same partition as my home directory (so I can 'mv' files back and forth between /tmp and /Users, instead of 'cp'-ing them), and I certainly don't want my /tmp competing for space on the same partition as my swapfile. I had that once, and some user who was telnetted into my machine filled up /tmp and froze the entire system on me (once it needed to expand the swapfile). Note that this is just a personal preference of mine, in that I simply do not like having tmp space and swapspace on the same partition. There isn't really anything *wrong* with it, but I don't like it.

So, if you do want to use /etc/rc.swap magic processing, you might want to look thru it and see if there's places you want it to work differently than the current logic works. The alternative (which I do) is to really add the disk in /etc/fstab, and then change the entries in /etc/swaptab so *that* is doing what you want it to. This is a little riskier, as it means you can't just attach and detach that swapdisk with reckless abandon. The /swapdisk code made sense for the time when it was written, but I think it's current coding doesn't make a lot of sense with the disks and machines that we're now running NeXTSTEP on. If you're buying an extra disk for swapping these days, it's probably "a little bit" larger than 40-megabytes!

The /swapdisk model has two advantages. 1) It's easy. 2) It works automatically when booting off a CD-ROM disc from NeXT (should you happen to want to do that). In my opinion though, being easy isn't much of an advantage if the processing is not doing what you want it to do. The logic in /etc/rc.swap, which was fine for the early days of NeXT hardware, simply does not do what I want it to do with the hardware configurations I'm running with these days.

Any reason I shouldn't use the free space on my swapdisk?

Reply from _William Herndon_

No hard and fast evidence, but I partioned my swapdisk and used the other half as a netboot partition for an additional NeXT. I ran with this configuration for about a year, and all during that time I was getting mysterious system crashes on my netboot server (the system with the partioned swapdisk). It may simply have been the extra overhead of "non-swap" traffic to the swapdisk; or it may have been something else, but it was a real pain. My recommendation is don't do it. Leave the swapdisk to be just a swapdisk.

What kinds of drives would be good for use as swapdisks?

From: _David Finton_
Date: 20 Jan 1995 22:53:23 GMT

I've been looking for a swapdrive for my slab: something like 100 MB, <= 10 ms, SCSI, cheap. Although I found several <= 170 MB drives advertized, none were still available. Here's a summary of the best drives I've found so far, after scouring the current Computer Shopper.

1. Used drives. These are usually small and < $100. Some are refurbs, some are just lightly used (you hope).

2. IBM 270 MB, 12 ms, 1" high, 2-year warranty, $172 from Optional Systems Resources, Inc.

3. Quantum 365 MB, 11 ms, 1" high, 2-year warranty, $209 from Insight/HDI

4. IBM 540 MB, < 9 ms, 1" high, can't remember if warranty is 2 years or (probably) longer, $304 from DATA NET.

[NOTE: pricing and availability information is bound to change over time. Basically you want to find a relatively fast drive that you can set aside and leave to swap]

Is a separate swapdrive worth the cost?

Most people I know of who have gotten seperate swapdrives have felt it was a noticable improvement in terms of system responsiveness. It's pretty helpful when doing things that create scratch files in /tmp as well.

> Is this true even when the swap drive is significantly slower than the current main drive? Or how much slower can the swap drive be, and still help to improve overall performance... Or must the swap drive be at least as fast as the current drive to make a difference?

Chuck Swiger *responded:*

That's a complicated question because it depends on what you're doing with the machine. Most of the time, large data sets which involve a lot of swapping also involve a lot of file I/O (in order to generate, manipulate, or save that data, usually), so having a seperate swapdrive helps in distributing the load between different drives.

However, you can come up with tasks where this would not be the case, and using a faster "primary" drive for swapping would be faster than using a slower swapdrive.

Anyway, for most practical purposes: unless your swapdrive is much slower as your other drive(s), using a swapdrive will probably be a performance win. It certainly is in the case of my system with a 14 ms swapdrive versus my primary drives, which are ar ound 8-10 ms.

There isn't a definitive answer to such a general question. You can get a definitive answers for specific system configurations by benchmarking it, though.

How can I fine-tune my swapdisk's performance?

*With many thanks to _Tim Scanlon_*

I wanted to pass on a technique that I've found to increase the efficiency & speed of a (my) swapdisk. It's somewhat technical, but I found it to be worth the effort I've gone to to do it.

It is a very simple solution. And involves one of the more traditionaly overlooked switches in a tradtionaly overlooked program, "tunefs". I am sure there'sa good chunk of you out there who have messed with the -m & -o prefrences, especialy if you are on black h/w & are using the old 105 to swap to... if you're not, and you have the 105 you should be

Basicly -m does "minfree", the 105 get's set at 5% by default, which is VERY low. it should be at least 15%, so you do -m 15 to fix it. The -o switch is either "time" or "space", and "space" is the default on 105's.

Another ~very~ important factor to recall is that the machine, (and this applies to ALL disks on ALL platforms) will automaticly get changed from "time" to "space" if you fill the disk, so an occasional 'tunefs' is called for as part of normal maintinance. it doesn't automaticly change back...

Overlooked however, and I say this because I have NEVER seen anyone speak of it, refer to it, or otherwise make any note of it in the NeXT community in the past 6 or so years is the "-e" switch, which does the max contigious blocks in a cylinder group.

The man page describes greater values as being better for large files (like a swapfile), and damned if it doesn't make a big difference to change it.

The default on a 105 meg disk is 256 bytes per cylinder group, which if the setup is "typical" according to the man page is 1/4 of the group. not too efficent for large files at all, cause it goes read 1/4, seek 3/4, read 1/4, etc.

I decided to change this, basicly to be more efficient, and here's what I did:
- Booted the machine without having the swapdisk mount.
- removed the "old" swapfile,
- unmounted the drive
- "tunefs -e 1024 -m 15 -o time /dev/rsdXa" where X is my drive #.
- mounted the drive.
- created a new swapfile "mkfile -v 50m swapfile"
- rebooted

I belive you can get away with doing it on a mounted filesystem with no negative impact too. Apperently tunefs is designed to work on mounted devices ok.

To explain, the default was 256, changing it at all returns a "was" & "is now" reading from it. So you can change it back. I derived 1024 becasue it was 1/4 of the default and I knew that 1024 was the frag size, so I figured it was a good value. Going higher than frag size seems like it'd be a Bad Thing.

The result I got was a perceptible increase in my machine's speed. The swap i/o seems to be much faster, and there's the aural factor. All of a sudden my disk is way the hell quieter, and is making far less "seeking" noises. I use a 50 meg swapfile by default, and grow it as needed up to 80 megs, & then kick over to a non-prefered file on a different drive.

I highly reccomend that if you are swaping on a separate disk, that you investigate this as a solution. YMMV but it worked well for me.

_

How do I set aside a partition for swapspace?

Configure the swapping via /etc/swaptab. Of course, remember to put an entry in /etc/fstab to mount the extra drive at boot. These suggestions are only educated speculation, for I've only slightly experimented using my own suggestions.

_

_

End of File: http://www.nerc.com/~luomat/next/swapfaq-qa.html

My Home Page
My NeXT Page
Intro to Swapfile/Swapdisk FAQ page