a foolproof guide to backups

by Andreas Sautter

Backing up the data on your computers is an integral part of day-to-day computer operations. Even though backups are commonly referred to as a necessary evil, they're the only way to be prepared for data loss. Data can be lost in even the most advanced environments. Whether you need to restore an accidentally removed file or rebuild an entire system, doing regular backups helps assure the security of your data and, if necessary, its efficient restoration.

This article tells you what data needs to be backed up and describes tried-and-true methods of performing various backup operations.

what data needs to be backed up

First, you need to differentiate between two types of dataÐoperating system data and user data. These data types require different backup strategies and media, depending on the size of the network and method of data collection.

operating system data

In contrast to configuration files,   such as the NetInfo database, most of the operating system doesn't often change. Operating system data usually lives on the computer's internal disk, also referred to as the boot device.

Unless you run a highly customized operating system environment, the core of the operating system doesn't need regular backups. You can rebuild vanilla systems fairly easily using standard NeXT software release products. The Release 2.1 Installation Set and, if appropriate, the Release 2.1 Extensions Installation Set come on floppy disks and should be used to rebuild 2.1 systems. If your system is equipped with an optical disk drive, you can use a bootable optical disk with the Release 2.1 software to rebuild your system. Release 2.2 is available on CD-ROM as a special Disk Recovery Tool   Set, and Release 3 will be shipped on CD-ROM with every new NeXT system (see table 1). For more information, see ªNeXT's strategy for software distributionº in this issue.

user data

Home directory files tend to change frequently, as e-mail is received and documents are created and changed almost daily. On NeXT machines, home directories are usually grouped within the same folder, typically named Users. In a small environment, home directories may live on the computer's internal disk. In large network installations with many users, the home directories are often grouped by business function or department and are likely to be located on server machines' external disks because of size limitations on the boot device.

table 1: NeXT software releases

software release	**medium**	**what to use to rebuild**
2 and 2.1	floppy disks	Release 2.1 Installation Set or, if appropriate, the Release 2.1 Extensions Installation Set.
2.2	CD-ROM	Release 2.2 CD-ROM, available at most NeXT Authorized Support Centers or directly through NeXTedge.
3	CD-ROM	Release 3 CD-ROM shipped with every system will provide an automatic system software backup, unless the operating system is customized.

backup media and devices

A number of devices are available for backups on NeXT computers. Your choice should be guided by how much money you can spend on backup media and devices and how much data you need to back up. A good backup medium should be:

- *Reliable*. The media should not deteriorate over time and must be resistant to physical hazards, such as magnetism and extreme temperatures.
- *Removable*. Consider it mandatory to store your backups in a different place from the original data. In case of fire or theft, you could be faced with complete loss of files if data is not stored elsewhere.
- *Portable*. Users who want to travel with their data must have a portable medium.
- *Fast*. Time is always a critical factor when a computing environment needs to be restored. Quick access to the right set of data is essential.
- *Cost-effective*. It's important to be able to store data as economically as possible. The more data you have, the more critical the cost per megabyte becomes.

Commonly used media for NeXT computers include floppy disks, optical disks, 8-mm videotapes, and DAT (digital audio tape) tapes. NeXT's Software and Peripherals listing provides more information about specific third-party vendors that offer backup devices for NeXT machines.

A05.floppy.tiff ¬

floppy disks

Floppy disks are nearly ideal for backing up small amounts of data. They're portable, widely available, and easy to use. If you use MS-DOS formatting, the floppy disk can be read on other computers, such as DOS personal computers and Macintosh computers.

Most NeXT systems have an internal floppy disk drive. A number of third-party devices can also be hooked up to the SCSI port. To back up data, simply copy files in the Workspace Manager™ and, if necessary, compress files with a third-party compression tool such as RunningBack or Squash.

A03.optical.tiff ¬

optical disks

You can store the whole operating system on a bootable NeXT optical disk, which makes it a great medium for restoring whole systems. Although several external optical disk systems are available from third-party vendors, the format used in those drives is not compatible with the format of NeXT optical disks. For backup programs, use dump, tar, and cp.

A15.8mm.icon.tiff ¬

8-mm videotapes

These tapes are used mainly for middle-sized to large network installations. 8-mm videotape drives, available from a variety of third-party vendors (see the sidebar ªthird-party drives for NeXT computersº), can be connected to a NeXT machine through the SCSI port. For backup programs, use dump and tar.

A16.dat.icon.tiff ¬

DAT tapes

Because DAT tapes have only recently become available, the DAT format is not as widely used for backups as 8-mm videotapes. DAT drives from third-party vendors can be connected to a NeXT machine through the SCSI port (see the sidebar ªthird-party drives for NeXT computersº).

less frequently used media

A number of other media are available: .25-inch tapes, 9-track tapes, write-once read-many (WORM) disks, and hard disks. The .25-inch tape drive, a common device on UNIX machines during the 1980s, has been largely replaced by 8-mm video and DAT drives.

Nine-track tapes predate the .25-inch tape. They are mostly used on large installations that have had such a device installed and keep them for compatibility reasons.

WORM disks are a special form of optical disk. As the name indicates, the medium is written to once but can be read as many times as you like. This feature is important when you must guarantee that the data, once written, cannot be altered under any circumstances.

Hard disks can be used to provide a fast on-line backup system.

Table 2 gives an overview of the advantages and disadvantages of the most common media.

table 2: advantages and disadvantages of common backup media

medium	**advantages**	**disadvantages**
floppy disks	portable	access is slow
	widely available	file compression necessary for more than 2.8 MB of data
	usable on variety of computer types	
	inexpensive for small amounts of data	
optical disks	portable	expensive relative to other backup media
	holds up to 256 MB of data (the complete operating system fits)	
	can't be corrupted by magnetic influence	
	can boot computer	
8-mm videotapes	de facto standard for large-capacity media	no random access
	storage capacity up to 5 GB	frequent cleaning necessary
	standard 8-mm videotape format	data integrity requires expensive high-quality tapes

		drives are expensive
DAT tapes (4 mm)	storage capacity up to 5 GB	no random access
	standard DAT tape format	drives are expensive

commands and utilities for backing up and restoring

This section presents an overview of commands and utilities available on your NeXT system to back up and restore data. Most of the commands are used in similar ways but for different tasks. For a complete list of the command-specific parameters, see the respective UNIX manual pages listed in the references at the end of this article.

dump/rdump and restore/rrestore

Some data on a system never or only rarely changes. The dump command provides you with different levels so you can save only the data that has been changed. The restore command is used to retrieve data that has been stored with the dump command. The data format of dump.old/restore.old is compatible with other BSD implementations, making tapes interchangeable with other systems.

tar

The tar command is used to back up specific directories and files. A more extensive implementation of the tar command, GNU tar version 1.10, is available for the NeXT machine, too. It has been posted in both source and binary form (compiled on System 2.1) on the following anonymous File Transfer Protocol (FTP) sites:

site	**directory**
sonata.cc.purdue.edu	pub/next/2.0-release/source/tar-1.10.tar.Z
cs.orst.edu	pub/next/sources/tools/tar-1.10.tar.Z.

For NeXT-specific information about GNU tar, consult the README file in the subdirectory tar-1.10/next. The GNU tar command is not part of the NeXT system software and is not supported by NeXT.

third-party drives for NeXT computers

The following is a list, by no means complete, of third-party drives that work with NeXT computers. NeXT does not specifically endorse any of these vendors.

8-mm videotape drives include:

- Jetstream
- Exabyte 8200ST
- Mountain Series 2100 1.4 GB

DAT drives include:

- 2000e ArchiveST 2.0GB External 4-mm DAT Drive
- Archive Python DAT 1.2 GB Tape

Note: Some DAT and 1/4-inch tape drives require the SCSI device driver to be set to recognize fixed-size transfer blocks. For detailed setup information, see NeXTanswer hardware.762.

compress and split

The compress command can be used along with dump/restore and tar to not only store but also compress the data.

The split command can partition a big file into multiple pieces, which can be useful if a file doesn't fit on a particular medium, especially on a floppy disk.

cp, rcp, rdist, and mt

The cp, rcp, and rdist commands can also be used for backups. They're used in special cases, such as keeping system files consistent in a network. For more information on these commands, see their respective UNIX manual pages.

The mt command is used to position the tape to a particular backup file on a tape and to rewind a tape.

table 3. advantages and disadvantages of commands and utilities

commands and utilities	**advantages**	**disadvantages**
dump and restore, rdump and rrestore	very effective, straightforward way to retrieve a large amount of data using restore's r option	dump can only back up entire file
	enables fast, unattended backups	
	back-to-back capability (to simultaneously back up and restore)	
	restore's i option makes interactive restoration possible	
tar	backs up only what is needed	100-character path length limit
	back-to-back capability	doesn't cross media boundaries
		no incremental backup possible
compress	requires less space (and therefore backup media) for the same	slows down the store/ retrieve

amount of data

backup strategies

Strategies for backing up your data vary greatly, depending on the types of data you need to back up and the kind of devices you have at your site. The strategies outlined here are not recipes but rather starting points for your personal backup approach.

backing up user data

Mere mortals can easily use the Workspace Manager to copy their most important files onto floppy or optical disks. No operator involvement is necessary. This kind of backup is usually not very reliable because the process can hardly be automated and depends on how conscientious the user is about keeping backups up-to-date. It is, however, practical if a user needs to have files handy on removable media.

All user data should be backed up systematically by a system administrator. The dump command provides the most effective and secure way to back up and restore user data. dump can be used to use back up only the data that has been changed since the last backup on a particular file system, which has the advantage of not using up valuable storage space.

You might, however, need to restore multiple levels before reaching the most current version of a particular file. A good way to use backup levels is to run a full level-0 backup once a week, complemented by an incremental level-1 backup every day. If you have enough space on your backup device, you might just want to run a level-0 dump every day. If backup speed and space are critical, you might do a level-0 backup once a month, a level-1 every week, and a level-2 the rest of the days (see figure 1 (page 21) in the hard copy *support bulletin*). Again, use your judgment and experiment with different backup levels to find the ideal solution for your site.A17 ¬A17 ¬A17 ¬A17 ¬

backing up system data

It's not usually necessary to have more than one backup of the whole operating system. Be sure, however, that all the configuration files, applications, and administrative data are being backed up at least once a week. Those files usually live in the following directories:

- /etc
- /usr/local
- /LocalLibrary
- /LocalApps
- /usr/spool

Note that /etc is actually a symbolic link to /private/etc.

You can use tar to back up these directories or selective files within these directories. Always be aware of the tar command's path limit.

In case of a complete system data loss, such as with a disk crash, you can rebuild a system to a vanilla version using the BuildDisk application. Then reinstall the custom files using tar.

tips and techniques

Before you start doing backups at your site, prepare a backup plan with a list of all the data that needs to be backed up. Then decide on the media and commands that you'll use.

doing unattended backups

You'll have to decide if the backup can be done unattended. Unattended backups are convenient since they don't require human intervention and can be run at night or on weekends, when few people are disturbed by the backup activity. If the backup fails, however, the error will be detected only later, and some of the data might not be backed up at all.

keeping your most important data safe

Back up your most important data first. If something goes wrong during the backup, or if your backup medium runs out of space, you're still sure that the most important data is safe. Usually, you should back up user data before system data.

working in a single-user mode

Ideally, backups are done in single-user mode. The restrictions of a working environment make this tricky. Do your backups when nobody else is likely to be using the system, for example, at night.

keeping records

Label your media properly. Make sure that both the medium and its container are labeled. If you make multiple backups onto one medium, most information can be retrieved from the table of contents (TOC) on the medium. Recording as much information as possible on the container of the medium not only clearly identifies the backups but also can also save you time when you need to retrieve data. (For example, see the sample backup script in the next section.)

A label on a backup medium should contain the following:

- Date and time of backup
- Type of backup (commands used)
- What has been backed up

The TOC should be the first file you write to the backup medium. It should contain the following information:

- Miscellaneous location and version information
- Dates of current and previous backup
- Number of files on the medium
- For each backup on the tape:
  - backup contents

backup command and arguments
restore agent

A typical TOC of a backup tape might look like this:

```
==== TABLE OF CONTENTS  (TOC)                                   ====
==== AcME HQ Backup Version 1.0                                 ====
==== Site: AcME HQ                                              ====
==== Backup Date: Tue Sep  3 16:25:49 BST 1991                  ====
==== Last Backup Date: Tue Aug  26 16:25:51 BST 1991            ====

====================================================================
File #  Host        Backup  Restore Description
====================================================================
1       -           tar         tar Table of Contents (this file)
2       server1 tar         tar /private/etc /usr/local
3       server1 dump-0  restore /Demos
4       server2 tar         tar /private/etc /usr/local
6       server2 dump-4      rrestore    /Users
```

After creating your backup, set the write protection on each medium to ensure that no accidental overwrites occur.

On each file system that you back up, create a dummy file that is touched every day and that is in itself the name of the machine and file system. This kind of file immediately identifies the file system and machine on a backup.

```
    mymachine> ls -lag /mymachine-ROOT_sd0a
    -rw-r--r--  1 root     wheel          0 Aug 24 00:25 mymachine-ROOT_sd0a
```

As an example, use the following crontab entry to create the dummy file:

```
    0 0 * * * * root /usr/bin/touch /mymachine-ROOT_sd0a
```

Use the revision control system (RCS) to guarantee that your scripts are properly maintained throughout multiple revisions and edits.

Document the exact location of the backup media within the building and how to gain access to themÐfor example, top drawer of low file cabinet in cubicle 127, building B. In a shared environment, backups must be accessed quickly to minimize system downtime.

managing the process

In a networked environment with a number of users, consider putting all the users' home directories on a separate disk or partition.   Doing this enables you to use the dump command on all the users' data.

Automating the backup procedure as much as possible will save you time and will assure that all the backups are done consistently.

Add the accounts of those who do the backups to the operator group. This allows them to read disks without becoming root and minimizes the risk of data loss from your disks.

Use the rdump and rrestore commands for backups across the network to a centralized backup device, for example, to an 8-mm video or DAT tape drive.

If tar or level-0 dumps are used to make a backup, you can predict exactly how much space it will take on the backup medium. Thus, you can find out how much of the backup will fit.

Keep at least five sets of your tapes around, one for every week of the month. Do not overwrite the most recent backup set since it contains the most up-to-date data. This allows you to restore files that are up to one month old.

Store all backup media in a safe place. Your backups are vital to your operation. Make sure that only authorized people have access to them. Your primary backup should be in a fireproof location on-site for efficient restoration of the data. A very safe way of storing your checkpoint backup media is off-site, which prevents complete data destruction resulting from a disaster at your site.

a sample backup script

The following is a very basic template script that you can use as a basis for your site:

```
#! /bin/sh-u
#
# very basic backup script
#
# possible improvements:
#  - error checking and warning, authorized user
#  checking, mail results of backups, check tape in
#  drive, clean up after interrupts
#

#
# doing inits first
#
PATH="/usr/local/bin:/usr/ucb:/bin:/usr/bin:/etc:/usr/etc"
export PATH
#
name=Ábasename $0Á
vers=1.0
site="AcME_HQ"
hostname=ÁhostnameÁ
usage="usage:  ${name} mode"
#
mode=$1
if [ "${mode=foo}" = "weekly" ]; then
```

```
    level=0
elif [ "${mode=foo}" = "daily" ]; then
    level=1
else
    echo ${usage}
    exit 1
fi
backup_dev=/dev/nrxt0
date=ÁdateÁ
#
# setting targets
#
targets="\
tar:myhost:/usr/local/bin \
tar:myhost:/etc/netinfo \
dump:myhost:/"
tmp_dir=/tmp
toc_file=${tmp_dir}/toc.$$
if [ -s ${toc_file} ]; then
    rm ${toc_file}
fi
#
# writing the TOC file
#
echo "#${name} Version: ${vers}" >> ${toc_file}
echo "Table of Contents (TOC)" >> ${toc_file}
echo "Site: ${site}" >> ${toc_file}
echo "Backup Date: ${date}" >> ${toc_file}
echo ${targets} >> ${toc_file}
#
# writing TOC to backup media
#
dd "if=${toc_file}" "of=${backup_dev}" 2>&1
stat=$?
if [ $stat -ne 0 ]; then
```

```
    echo "couldn©t write TOC to backup_media, aborting..."
    exit 1
fi
#
# starting to do the actual backups
#
for a_target in ${targets}; do
    command=Áecho "${a_target}"|awk -F: ©{print $1}©Á
    host=Áecho "${a_target}"|awk -F: ©{print $2}©Á
    dir=Áecho "${a_target}"|awk -F: ©{print $3}©Á
    if [ "${command=foo}" = "dump" ]; then
        options="${level}suf 1200000"
    elif [ "${command=foo}" = "tar" ]; then
        options="-cvf"
    else
        echo "unknown command ${command}, skipping backup for \
            ${host}:${dir}..."
        break
    fi
    remote=no
    if [ "${host=foo}" != ${hostname} ]; then
        remote=yes
    fi
    if [ ${remote} = "no" ]; then
        ${command} ${options} ${backup_dev} ${dir} 2>&1
    else
        if [ ${command} = "tar" ]; then
            rsh ${host} "cd ${dir}/..; /bin/tar ${options} - ${dir}" | \
                dd of=${backup_dev}
            stat=$?
            if [ $stat -ne 0 ]; then
        echo "couldn©t ${command} ${host}:${dir}, aborting..."
        exit 1
    fi
        else
```

```
            rsh ${host} "/etc/rdump ${options} ${hostname}: \
                ${backup_dev} ${dir}"
            stat=$?
            if [ $stat -ne 0 ]; then
        echo "couldn©t ${command} ${host}:${dir}, aborting..."
        exit 1
    fi

        fi
    fi
    stat=$?
    if [ $stat -ne 0 ]; then
        echo "couldn©t ${command} ${host}:${dir}, aborting..."
        exit 1
    fi
done
#
# let©s clean up
#
rm ${toc_file}
exit 0
```

in conclusion

With a little planning, backups aren't that difficult. Before you start, take another look at the sections describing what data and how much needs to be backed up. You can then determine the most appropriate backup device and utilities. Using the tips and techniques described in this article, you can tailor the backup procedure to your needs. The efforts you make will be well worth it should you ever have a data loss.

references

The following UNIX manual pages provide an in-depth explanation of the various commands along with useful examples:

dump(5)	incremental dump format
dump(8)	incremental file system dump
restore(8)	incremental file system restore
rdump(8c)	file system dump across the network
rrestore(8c)	restore a file system dump across the network
tar(5)	tape archive file format
tar(1)	tape archiver
cp(1)	copy
rcp(1c)	remote file copy
rdist(1)	remote file distribution program
rmt(8c)	remote magtape protocol module
st(4)	SCSI Tape Device Driver
compress(1)	compresses and expands data
split(1)	splits a file into pieces
dd(1)	converts and copies a file
mt(1)	magnetic tape manipulating program

The following NeXTanswers provide some more information and hints on the backup topic:

- sysadmin.539
- sysadmin.416
- sysadmin.379
- hardware.762
- hardware.340

Chapter 12 of the *NeXT Network and System Administration* manual and on-line documentation introduces backup strategies on the NeXT machine.

In the *UNIX System Administration Handbook* (Nemeth, Evi, Garth Snyder, Scott Seebass; Englewood Cliffs, NJ: Prentice-Hall,. 1989), you'll find a very thorough and complete description of the different backup procedures on UNIX systems.

Last but not least, consult the reference manual for your third-party devices. Use that documentation for device-specific setup instructions.