

Net Integration Technologies, Inc.



<http://www.net-itech.com>

Net Integrator Backup System

Technical Overview

Version 1.00

TABLE OF CONTENTS

1 Introduction	1
2 How Backups Work	2
2.1 Which data is backed up?	2
2.2 Where is the backup stored?	2
2.3 How often do backups occur?	3
3 Backups on Tape	4
3.1 Advantages of the tape system	4
3.1.1 One-touch backup system	4
3.1.2 Removable storage media	4
3.1.3 Intelligent tape spanning	4
3.2 Details on Net Integrator tape backup features	5
3.2.1 Faster, smaller backups: Compression	5
3.2.2 Faster recovery using logical divisions	5
4 Intelligent Disk Backup	7
4.1 Advantages of idb	7
4.1.1 Backups are fast	7
4.1.2 Large storage capacity	7
4.1.3 Removable storage media	7
4.1.4 idb is completely automated	8
4.2 Details on Net Integrator idb features	8
4.3 Incremental backups	8
4.3.1 Micro-backups	9
5 About Net Integration Technologies	10

1 INTRODUCTION

No system's data is completely safe.

Hard disks can fail at any time, and indeed all will fail eventually as their components wear out. Lightning strikes can produce power surges that can destroy storage devices. Humans can destroy data: they can accidentally delete important files or knock servers over, smashing the contents. Sometimes, the data is recoverable. Often, some, most, or all of the damage is irreparable.

Preventative measures, such as better power regulation or improved mechanical quality can reduce the risk of catastrophic failures. But these things can never reduce the risk to zero—there is always a chance that a system's data can be irreversibly lost.

But there is hope. If the data is regularly copied onto other media, a recent copy can be restored in the event that the data is lost. These *backup* copies provide extra safety in an emergency.

Modern systems, especially those that contain the data of many users, almost always have backup systems. But these systems can often be tedious: they can be complex, slow, and sometimes not even complete.

The Net Integrator's backup system remedies these problems with its tape backup and Intelligent Disk Backup (idb) systems, which are simple to set up, fast, and always complete.

2 HOW BACKUPS WORK

A backup refers to a secondary copy of data that is used elsewhere. Users never work directly on a backup; instead, they work on the original data, which is periodically copied to the backup device. For example, a home computer user might make a copy of all the data on the computer's hard drive by creating a CD-ROM every month. In office settings, important files are often copied to magnetic tapes on a daily basis.

2.1 Which data is backed up?

The data to be backed up on a server generally includes all users' files—such as reports, programming source code, or World Wide Web homepages—plus the system files—applications, data files, and so on.

Due to the nature of the Net Integrator, a backup only needs to include user files plus one system file. All other system files are stored on a flash disk that can easily be replaced in the event of a failure. For a given release, say 3.50b, these files do not change, so there is no need to back them up; the desired version can be retrieved from the SoftUpdate servers after a new flash disk is installed.

The sole system file that is included in backups is the system configuration file. This contains all of the individual settings of the Net Integrator, such as user names, the hostname, and server configurations. This file is stored both on the hard disk and on the flash disk, so if one device fails, the configuration can be retrieved from the other. It is backed up because it might be desirable to restore an earlier configuration, or both the flash and hard disks might fail (which, however, is extremely unlikely).

In addition to users' regular files, the email stored on the server is also backed up. When restored, it immediately appears in the users' folders, whether IMAP, POP3, or WebMail.

2.2 Where is the backup stored?

The Net Integrator Mark I and Mark II contain either a magnetic tape drive or a secondary, removable hard drive for backup purposes.

Tape drives are a common form of backup device. The tape drives used in Net Integrators are fast, extremely reliable, and have a large storage capability. The tapes themselves are relatively inexpensive.

Use of a secondary hard disk drive for backup purposes is a feature unique to Net Integrators. Using a system called Intelligent Disk Backup, or *idb*, backups can be performed quickly and reliably without user intervention. Furthermore, backups on *idb* systems can be performed as often as every fifteen minutes, so the backups are always nearly up-to-date.

2.3 How often do backups occur?

User files change often as the user updates web pages, writes more source code, or revises reports. Thus, backups must be performed frequently. If a hard disk drive crashes and loses its data, a backup made three months before is probably not useful on a system that is regularly accessed by users. Most administrators perform tape backups on a daily basis; *idb* systems automatically perform backups as often as every fifteen minutes.

Unlike many systems, the Net Integrator does not require any services to be disabled while performing a backup. All normal functions—file serving, web serving, email, and so on—continue running during the backup process.

3 BACKUPS ON TAPE

Tape drives are similar to the audio tape recorder found in stereo systems. In such tape recorders, playing music found in the middle of the tape requires rewinding or fast-forwarding to position the tape head. To find the beginning of a song, one must frequently play the tape to determine how much farther the tape needs to be fast-forwarded or rewound. Once at the appropriate place, the tape recorder can play or record sound until it reaches the end of the tape.

This is similar to how a computer's tape drive works, except that searching for files occurs only in the forward direction. The tape is rewind, then searched in order until the proper spot is found. If the data needed is located near the end of the tape, it can take a long time to find it. The time to find this place, the *seek time*, is much higher on tape drives than on most other storage media, such as hard drives.

3.1 Advantages of the tape system

3.1.1 One-touch backup system

In addition to the WebConfig backup pages, backups can be started by simply pressing the "Backup" button on the front panel, in which case the backup uses the settings from the last backup initiated through WebConfig. A complete restore is also available through the "Restore" button.

3.1.2 Removable storage media

Since individual tapes are inexpensive and removable, most organizations store one or more backups offsite. In the event of a catastrophe at the workplace, the offsite tapes are safe and can be used if the others are destroyed. Tape backups can also be used in other systems. This allows a complete restore of the data on one machine to be performed on another, which would be useful in the case of a total system failure. It also provides a simple way to transfer large amounts of data from one Net Integrator to another.

3.1.3 Intelligent tape spanning

All data is compressed as it is being backed up, so backups usually fit onto a single tape. However, certain types of data do not compress well. Digital multimedia, for example, is already compressed. Re-compressing this does little if anything to reduce the file size.

If the files on a Net Integrator are mostly compressed data like digital video, there is a chance that an entire backup will not fit on one tape. In this case, the Net Integrator's backup report, which is generated and emailed to the administrator or another designated person, indicates which files have been backed up. The original backup can be continued onto a new tape simply by activating the backup a second time, after a new tape has been inserted into the drive. In addition, the Net Integrator ensure that a user's files are always on one tape and never span tapes; in other words, if a tape fills up while backing up files belonging to a particular user, the backup of that user will be shifted entirely to the second tape. This ensures that, when restoring one user, the administrator never has to wait for more than one tape.

3.2 Details on Net Integrator tape backup features

Net Integrators use a custom approach to backing up data onto magnetic tapes. The data is written in compressed blocks of about 256 kilobytes, with logical divisions every 2 gigabytes, approximately.

3.2.1 Faster, smaller backups: Compression

Because the tape drives on Net Integrators have a smaller storage capacity than the hard drives, the data is compressed to improve the probability that a complete backup will fit onto one tape. Compression is a way to decrease the size of data files without losing any data. Most forms of data, such as text files or other office files, contain a certain amount of repetition that can be eliminated by compressing them. To use the files, they must be decompressed, which occurs automatically when a backup is restored.

The tape backup system uses the zlib library, which is used in the common UNIX and Linux programs `gzip` and `gunzip`. The backup is broken up into chunks of about 256 kilobytes, which are individually compressed. This was determined by internal testing to be the optimal size, as smaller blocks result in poorer compression, and large blocks take significantly longer to compress while resulting in only marginally better compression.

3.2.2 Faster recovery using logical divisions

As explained above, tape drives take a relatively long time to find a specific spot on a tape. Finding a file on the tape can take a very long time, as the backup system needs to go through each file in the order that they were written to the tape to find the desired one. During a full restore, finding a specific file is not needed. But if the administrator needs

to restore only a few files, which is called a partial restore, the time required can take almost as long as a full restore, especially if the desired files are near the end of the tape.

Luckily, the Net Integrator's tape drives have a feature called *file marks*. These are used to mark spots on the tape so that the drive can quickly find a particular location. A file mark is similar to a bookmark. Without a bookmark, a reader would have to skim through the book, page by page, to find a specific passage. But with a bookmark, the reader need only open the book to the correct spot.

But writing too many file marks can increase the amount of time required to perform a backup. After some internal testing, it was determined that 2 gigabytes was the optimal space in between file marks. This represented a trade off between the time to perform a backup and the time to perform a partial restore. An index of files and file marks is stored at the end of the tape, and, when a particular file is required from the tape, the backup system reads the index and finds the file mark that precedes the desired file. It then rewinds the tape, fast-forwards to that file mark, and reads the tape until the appropriate file is located.

4 INTELLIGENT DISK BACKUP

The other backup system used in some Net Integrators is Intelligent Disk Backup, or idb. idb is a unique approach to system backup.

Hard drive technology is constantly evolving. It is quite easy to replace a week's worth of backup tapes with a single removable hard drive. Hard drives do not suffer the same disadvantages as tape drives do when it comes to seek time and therefore are much faster. Net Integration Technologies' idb solution replaces the tape drive with a secondary hard drive.

Much of the idb system is similar to the tape backup system. A full backup consists of copying all files to the idb disk. Hard drives, like most storage media except tapes, use something called *addressing* to keep track of the locations of files. Every file has a numeric address which points to the physical location of the file on the drive. Finding a file is as simple as looking up its address; hence, file marks are unnecessary.

4.1 Advantages of idb

4.1.1 Backups are fast

idb backups and restores take only a fraction of the time required to perform tape backups and restores. This is due to a combination of the raw speed of hard drives, the addressing system described above, and the use of incremental backups, which are described in detail below.

4.1.2 Large storage capacity

idb drives have vast storage capabilities. This means that every drive can store many backups, even if the Net Integrator has a large user base who all have large amounts of data. The storage ability of idb drives comes partly from the fact that hard drives can hold tens of gigabytes, but also from the incremental backup system.

4.1.3 Removable storage media

Like tapes, idb disks can easily be replaced. This does not confuse the incremental backup system, regardless of when the disks are switched. However, the large storage capacity of idb disks means that idb disks do not have to be switched frequently. Like

tapes, idb disks can be stored offsite, and idb allows a particular Net Integrator's data to be restored to another.

4.1.4 idb is completely automated

Backups occur automatically on a monthly, weekly, and daily basis and can be set to occur as often as every fifteen minutes. Again, due to the large storage capability of idb drives, many backups can be stored on one disk before it must be replaced. The administrator does not have to bother with replacing tapes nightly. And, if a backup is needed immediately, it can be performed simply by pressing the "Backup" button on the front panel.

4.2 Details on Net Integrator idb features

A key feature of the idb system is its ability to perform incremental backups. As mentioned above, this both speeds up the backup process as well as increasing the number of backups that can be saved to one idb disk.

4.3 Incremental backups

Due to the flexibility of hard drives, it has been possible to expand the backup system beyond simply copying every file on the main hard drive on a daily basis. idb uses a technique called *incremental backups* to fit many days' worth of backups onto one idb disk.

The Net Integrator idb system only backs up all files once per month. Otherwise, only changed and new files are backed up. These backups are based on the full backups: a weekly backup copies only the files that have been changed or added since the monthly backup, and a daily backup copies only changed and new files since the weekly backup. Using this technique, it is possible to fit many more backups on one idb disk than if full backups are performed every day.

An example will help illustrate this technique. Let's say that the system "metamorphosis" performs a full backup on the 1st day of every month. On September 1st, which happens to be a Saturday, every user file, plus the system configuration file, is copied to the idb disk. That day, josefk, being a hard worker, saves a report (derprozess.pdf) to his user directory. On Sunday, a weekly backup is performed. The idb system looks at all of the user files on metamorphosis and sees that only one,

derprozess.pdf, has been changed or added since the monthly backup. It copies only that file as the weekly backup. So there is now a complete backup on the idb disk, which consists of one monthly and one weekly backup. A similar process is repeated for each day after the weekly backup: once per day, the idb system examines the user files and copies changed and new files since the last weekly backup as its daily backup.

It is important to note that the idb system does all the decision-making. Administrators never need to decide whether to perform a full or an incremental backup; the Net Integrator does this for them, based on the criteria described above and the existing contents of the idb disk.

Restoring a backup is similar to the reverse of the backup process. The Net Integrator restores all the files in the selected backup first. Then, all files in the parent backup (upon which the first backup is based) are restored. The parent of that backup is then restored, and this continues until there are no more parents, which is generally when the most recent full backup is restored. For instance, if an idb disk had one monthly, one weekly, and one daily backup, the daily would be restored first, followed by the weekly, followed by the monthly.

4.3.1 Micro-backups

Suppose that daily backups on metamorphosis are performed at 2:00 am each night. If a hard drive crash occurs at 5:00 pm one work day, all the data done that day is lost, since the last backup was at 2:00 am that morning. Micro-backups solve this problem by performing small incremental backups throughout the day.

Micro-backups are always based on the most recent backup. The first micro-backup of the day is based on the last daily backup, and subsequent micro-backups are based on the previous micro-backup. The default interval is 15 minutes; that is, every 15 minutes all changed and new files so far that day are backed up. Since it is likely that only a small amount of data will have been changed or added, these backups are typically very small and do not impact the performance of the Net Integrator.

5 ABOUT NET INTEGRATION TECHNOLOGIES

Net Integration Technologies Inc. develops and manufactures affordable, high-performance all-in-one network server appliances designed to fulfill the IT needs of small to medium-sized organizations. The Net Integrator family of products leverages Net Integration's unique technologies to deliver an easy-to-use, reliable, powerful and cost-effective IT solution.

The Net Integration product family includes Net Integrator Lite, Net Integrator Mark I, Net Integrator Mark II, Net Integrator Mark I-idb and Net Integrator Mark II-idb. This family of products is flexible, scalable and does not require specialized IT skills to set up, maintain or manage. By eliminating the need for expensive hardware, software and highly skilled IT professionals, Net Integrator gives small to medium-sized businesses access to the IT services needed to succeed in today's business environment.

Net Integrator products are delivered through a global network of value-added resellers (VARs) and directly from Net Integration Technologies, Inc. Net Integration Technologies is a privately owned, venture-backed company headquartered in Toronto, Canada.

For more information, please visit the Net Integration web site at www.net-itech.com