zmanda

Amanda Enterprise **White Paper**

Advantages of Amanda Over Proprietary Backup Software

465 S. Mathilda Ave., Suite 300 • Sunnyvale, CA 94086 • t: (888) 496-2632 • zsales@zmanda.com • www.zmanda.com

Abstract

This white paper discusses how Amanda compares to proprietary backup software. It will help you understand some key Amanda advantages, and will help you evaluate Amanda for protecting your data.

Table of Contents

Introduction

What is Amanda

Today there are a number of choices when choosing a backup and recovery solution. Perhaps you are already using one of the old commercial packages, such as NetBackup, NetWorker, or TSM. So why consider an alternative? Because with these products you face many challenges – they are high in cost, complicated, and you are forced into vendor lock-in.

This paper will show you how Amanda compares with commercial software packages in the following ways:

- Amanda is **an open source** product, while the others are closed-source.
- Amanda **determines the backup level** for a given scheduled backup run, rather than you having to determine this ahead of time.
- Amanda uses **standard backup tools and data layouts** *(gnutar*, *dump, Schily tar and zip*), instead of proprietary ones.
- **Backup to disk is integral** to Amanda's operation..
- A **holding disk** is used to temporarily store backup data on the Amanda server.
- Amanda supports open-source and user configurable **encryption and data compression**, while other solutions force you to use their encryption and compression routines.
- Amanda provides an option to backup your data to Amazon S3 storage cloud.

Benefits of an Open Source Product

The primary benefit of Amanda being open source is that the source code is available to all for review and modification. A community of developers, from various companies and industries, works to maintain and improve Amanda. This is in sharp contrast to closed-source backup software where only the vendor can change the product. With a closed-source backup product, you must submit your enhancement request to a single company, and hope that there is enough of a business case for them to add the enhancement. With Amanda's open-source approach, you have an option to add the enhancement yourself, or rally other users who are interested to do so.

This can greatly reduce the time it takes to get changes into the product. For example, in the recent past the following features have been added to Amanda:

- Tape spanning – which allows Amanda dumps to span from one media volume to another. This enables you to backup data sets of virtually any size using Amanda.
- Encryption and compression of the backup data, either on the client or on the server, using industry standard algorithms and tools.
- Enhanced authentication methods between the Amanda server and client.
- Support for IPv6.
- Support for Volume Shadow Copy Services (VSS) for backup of open files and hot backup of Exchange, MS SQL, Oracle and other applications on Windows.
- Backup of Oracle on Linux and Solaris via Oracle Recover Manager (RMAN).
- Backup over the Internet to Amazon Simple Secure Storage (S3) Service

Advantages of Amanda Over Proprietary Backup Software

This is just a partial list, highlighting some of the more significant features added to Amanda. In a closed-source product, these changes would have taken much longer. Indeed, most commercial backup software packages have not yet released support for IPv6 and do not provide backup to a storage cloud.

Despite this clear advantage, some may hesitate to venture into open-source products, fearing that they will have to write, compile, or otherwise manipulate the source code, using complicated and arcane procedures. However, this is not true in Amanda's case because you can obtain ready-made, tested and validated installation packages of Amanda Enterprise from Zmanda, Inc.

Another important benefit of Amanda's open-source model is external auditing of code quality. With a closed-source product, how do you measure intrinsic code-quality? You might keep track of the number of bugs you report, or the number of bugs fixed in a given release (if the vendor publishes such a list). Other than this, it is difficult, if not impossible, to measure code quality. Closed-source backup products keep hidden their code quality issues, including critical security problems.

Amanda's source code is available for inspection and has been analyzed for code quality. This transparency significantly increases confidence in the product. In fact, two organizations have examined and measured code quality for Amanda. One is Klokwork, the other is Coverity, which collaborates with Stanford University. Coverity applied a source code analysis tool named "Scan" to Amanda, and all defects found to date have been corrected. Because of code transparency, Amanda has been certified by the US Department of Homeland Security and is the only backup product to achieve such a distinction.

How Backup Levels Are Determined

In a backup product like NetWorker from EMC, you must specify which backup levels are to be performed, and on which specific days to perform these backup levels. If you have a large number of backup clients to configure, this can be tedious and complicated. Amanda was designed to determine the backup levels for you, relieving you of this burden.

Another consideration is that in NetWorker, NetBackup or BackupExec you tend to do all of your full backups on one day, and perform incremental backups the rest of the days. For example, you may decide to perform a full backup on Sunday, and incremental backups on the rest of the days of the week:

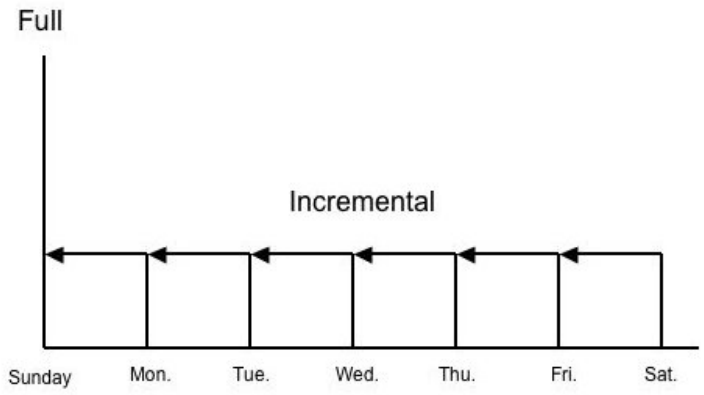


Figure 1: Example NetWorker Backup Schedule

All backup clients that use this schedule will perform their full backup on Sunday. This dramatically increases the amount of time for Sunday's backup relative to the rest of the week. You no longer have a consistent amount of time for your backups each day. This is of course because dramatically different amounts of data will be backed up each day. In addition, much more backup media will be consumed on Sunday meaning you have to verify if enough media is available to finish the full backup.

Amanda, on the other hand, has the following design goals related to dump levels:

- Dump a consistent amount of data each dump run.
- Thus consume a consistent amount of backup media each dump run.
- Run each dump run within a consistent amount of time (backup window).
- Recalculate and re-balance the schedule when you add a new client to the network or an existing client is not available.

To accomplish these goals Amanda has an estimation or planning phase at the beginning of a dump run. A dump run is a backup of multiple file systems, directories and databases, often from multiple backup clients. The goal of the planning phase is to determine the backup levels for the individual dumps in the dump run. Each dump run results in a different mixture of full and incremental dumps.

One determining factor for the individual dump levels to be run is how much data will that level produce. Amanda attempts to keep the total amount of data dumped for each dump run the same, which results in a consistent backup window.

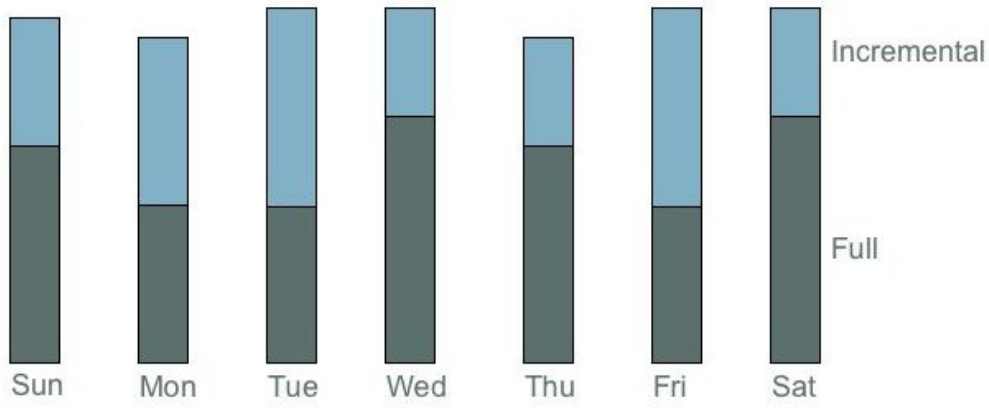


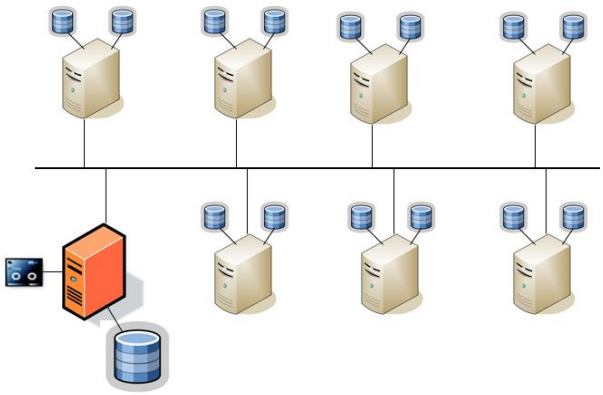
Figure 2: Example Zmanda Backup Schedule

Amanda's backup level algorithm also manages spikes in data changes. For example, suppose that on Tuesday an unusually large amount of data is changed. This will cause an unusually large amount of data to be backed up for the incremental dumps being run that day. Amanda's planning algorithm can balance the large amount of incremental data by promoting some backup jobs from to full dumps.

How to Control Amanda Dump Levels

The primary mechanism is the *dumpcycle* parameter (*Backup Cycle* in Zmanda Management Console available in Amanda Enterprise). The *dumpcycle* is the number of days over which you would like individual dumps in the dump run to have at least one full dump.

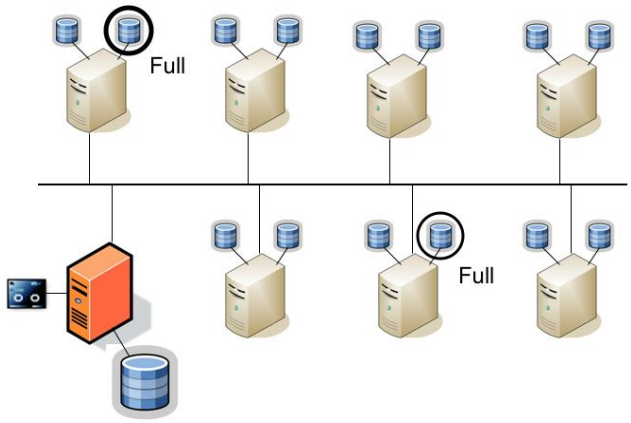
For example suppose you have seven client systems, each with two file systems (we'll make this a simple example for illustration purposes).



Amanda Server with 7 Clients
Each Client has 2 file systems

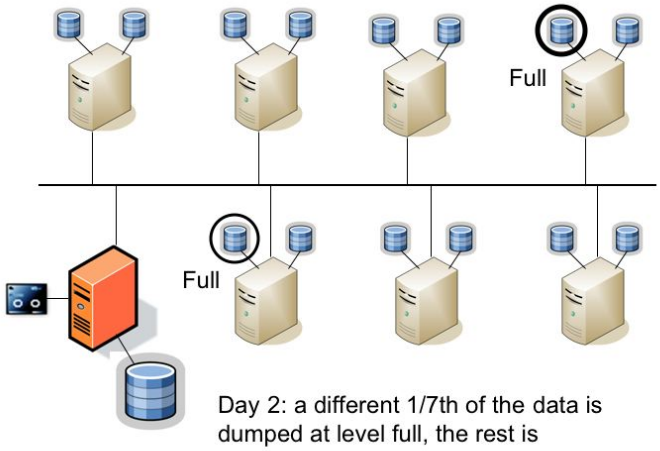
It is not uncommon for full backups to be done weekly. To do this with Amanda you would specify a *dumpcycle* of 7 days. However, unlike other backup products this does not mean to perform full dump every seventh day. Rather, it means that within seven days each of the file systems in our example will be backed up at level full, but not on the same days.

What Amanda's algorithm tries to do is smooth out the amount of data backed up each day. What this approximates to in our example is to perform a full dump of 1/7th of all file systems each day, and perform incremental dumps on the rest of the 6/7th of file systems.



Day 1: 1/7th of the data is dumped at level full, the rest is incremental

The next day a different 1/7th of the file systems will be dumped at level full, and the remainder is dumped as incremental.



Day 2: a different 1/7th of the data is dumped at level full, the rest is incremental

Amanda continues this process across the 7-day *dumpcycle*, so that by the end of those 7 days all file systems have been backed up at level full.

Amanda has more controls that you can apply to backup levels. For example, you can control when Amanda uses various levels of incremental dumps (for more details see http://wiki.zmanda.com/index.php/FAQ and read the section on "bumping"). These controls are lacking in other backup products like NetWorker.

Use of Standard Backup Tools and Data Layouts

Disadvantages of Proprietary Tools

Commercial products such as IBM TSM and Symantec NetBackup use their own proprietary backup algorithms, tools, and data layouts. The use of proprietary tools locks you into their product, since it provides the only way to recover your data. Your backup data is in essence held hostage by the vendor. You may use their product for several years, creating a large investment in backup data. The vendor knows that it would be difficult for you to switch backup solutions, because you would no longer be able to recover the older backup data.

Advantages of Amanda's Approach

Amanda, on the other hand, does not use proprietary tools and data layouts when backing up your data. Amanda uses *tar*, *dump*, or *Schily tar* as backup tools on Linux, UNIX and Mac OS X platforms, and Microsoft Volume Shadow Copy Services (VSS) on Windows. These are readily available, industry-standard tools. Their specifications and data layouts have been stable for many years, with the promise that this will continue to be stable and accessible for years to come.

Amanda adds value to these tools by providing several features:

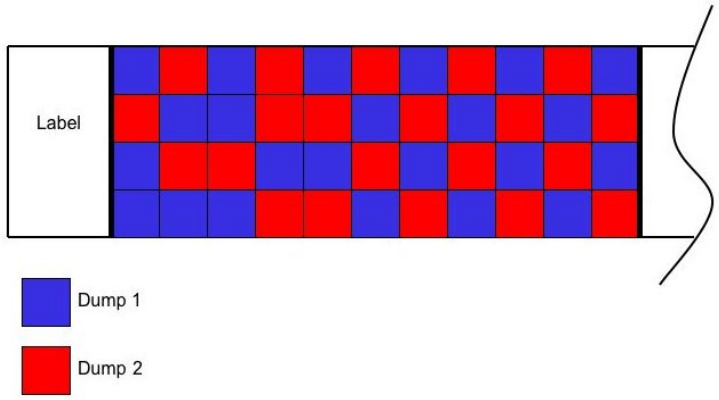
- Amanda keeps an index of files for ease of recovery
- Amanda provides its own toolset for recovering files. These tools provide ease of use and index navigation.
- It automates the running of the standard tools.
- Amanda can send the output of the standard tools across the network to a centralized backup server.
- The output of several backup clients, using the standard tools, can be written to tape on the backup server.
- Additionally, data from Amanda dumps can be restored without the use of any Amanda commands.
- Windows client that uses VSS to backup NTFS file systems even when some files are open. Backup data is in a standard zip format used by "Compressed Folders" feature in Windows XP and later.

Amanda Enterprise adds:

- Zmanda Management Console that simplifies all day-to-day activities of backup administrators.
- Backup of System State for Windows clients
- Hot backup of MS Exchange, MS SQL and Oracle databases.
- Option to backup data over the Internet to Amazon Simple Secure Storage (S3) Service.
- New functionality and security updates and 24x7 support from Zmanda.

Proprietary Media Layout

Many backup products use proprietary tape layouts where the data from different dumps are intermixed on the tape:



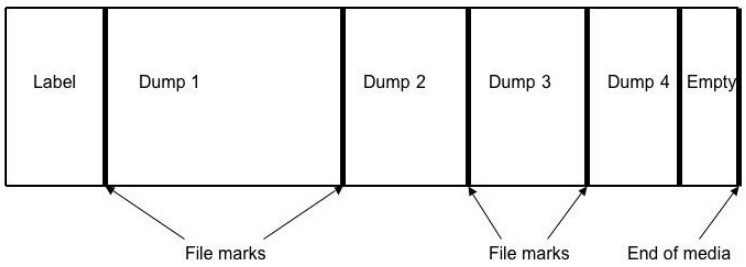
Proprietary tape layout

Notice that the data from dumps 1 and 2 are intermixed. One side effect is that you cannot decode the contents of the tape using any sort of standardized tool. In order to recover either dump you *must* use the proprietary product.

Speed and ease of recovery also becomes an issue. Since the data from the two dumps are intermixed on the tape, to recover just one of the dumps requires skipping blocks that are not related to the recovery. For example, to recover from dump 2 above, all of the blocks associated with dump 1 must be ignored. However, those ignored blocks take up tape space. This increases the amount of tape that must be traversed in order to recover dump 2 and all the negative effects of shoe-shining for a drive that is reading the tape. This in turn increases the amount of time required for the recovery. The more dumps are intermixed on tape (or other media) the longer it takes to recover the data.

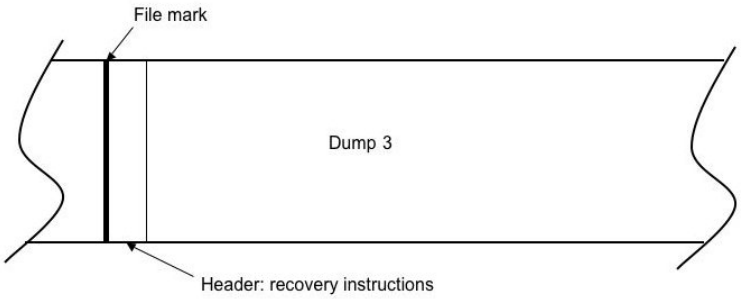
Amanda's Media Layout

Amanda allows for several backups from several sources to reside on the same piece of media. Amanda separates each backup on tape using a tape file mark.



Amanda tape layout

When Amanda writes the output of a backup to tape it does not alter the output in any significant way (see section below on encryption and compression). It does precede each *tar* or *dump* output with a header of a given size (typically 32K). This header documents the exact command sequence needed to restore the data with standard operating system utilities (e.g. *tar* or *dump*).



Amanda dump header

The instructions in this header are optional – you only need to use these instructions if you are trying to recover data from the dump without the use of Amanda recovery commands. Below are some typical contents of a header:

AMANDA: FILE 20070604161605 tm-amserver01 /testdata01 lev 0 comp N program /bin/tar

To restore, position tape at start of file and run:

dd if=<tape> bs=32k skip=1 | /bin/tar -xpGf - ...

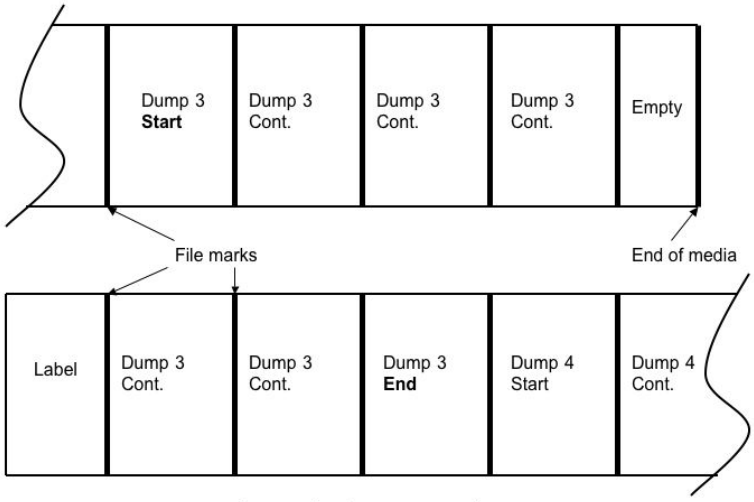
The first line of output describes the host name (tm-amserver01), the file system or directory that was dumped (/testdata01), the backup level (0), whether or not compression was used, and the program used to create the dump. The rest of the header gives detailed instructions on how to use non-Amanda commands to restore the

data.
Amanda uses the very same data layout for any supported media. So unlike proprietary solution Amanda does not hold your data hostage. If Amanda is unavailable for any reason you can still
recover the data.

Amanda Tape Spanning

One recent addition to Amanda is the ability to split a single dump across tape media. Amanda accomplishes this by dividing the dump into smaller pieces and writing the smaller pieces to tape. The size of the pieces written to tape is controlled by the *tape_splitsize* and the *fallback_splitsize* parameters in the amanda.conf file. The *fallback_splitsize* is used when there is no holding disk being used for the dump (see "Holding Disk" below).

Each piece of the dump is separated with tape file marks. When it reaches the end of tape Amanda uses various methods to detect that the piece being written did not fit on the tape. This piece is written on the next tape, immediately after the label.



Amanda tape spanning

Use of Standard Encryption and Compression Algorithms

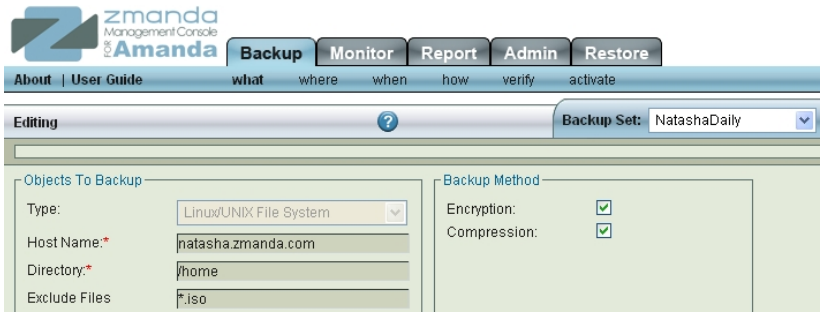
During the past several years there have been articles in the press about various companies who have lost their backup tapes. Sometimes these tapes are misplaced by courier services; sometimes they are stolen; sometimes they just go missing. If these tapes have sensitive data, such as account numbers, the loss represents a significant financial risk to the company and their customers. It also places the company loosing the data in a bad light.

To combat this risk that backup data could be misused, most commercial backup products have software encryption algorithms available. However, you cannot examine the algorithms, nor can you choose to use a different algorithm. If you need a better encryption algorithm than the one provided, you are just out of luck.

Amanda allows you to choose the encryption command, which puts you in the driver's seat when choosing the appropriate level of encryption for your security needs. Any encryption command that reads from standard input and writes to standard output can be used. This flexibility allows Amanda to benefit from rapid advances in encryption technologies. There is no need to wait for changes from a proprietary software vendor to increase your level of security.

Unlike some commercial software packages, Amanda gives you the flexibility to encrypt the data either on the backup client or on the Amanda server. Other packages, such as EMC's NetWorker, only allows for encryption on the backup client. Since Amanda gives you the choice, you can encrypt the data on the most appropriate host. For example, if you were concerned about the backup traffic being intercepted as it flows over TCP/IP, encrypting the data on the client would be most appropriate. In other cases, network security is provided by other means. Since encryption can take significant resources you may choose to encrypt the backup data on the Amanda server.

Amanda, including Amanda Enterprise, comes with interfaces to standard open-source encryption tools such as *openssl* and *GPG*. Zmanda Management Console (ZMC) users can easily turn on server-side encryption. Simply click a check box labeled "Encryption" when defining the directory to be backed up (shown below).



Use of Standard Compression Algorithms

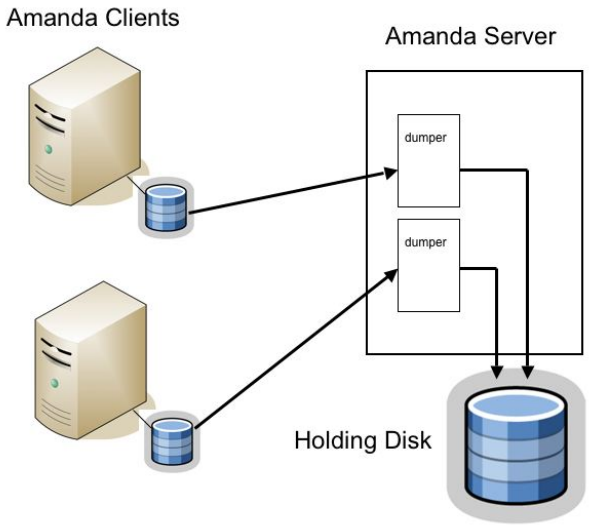
Amanda also implements support for standard compression utilities, such as ***gzip***. You may choose any suitable program that compresses and decompresses data. This allows you to choose a compression algorithm tailored to your data and the resources available to you (CPU, memory, etc.). Also, like Amanda encryption, you may choose to run the compression on the client or the server.

Proprietary backup solutions do not give you these options – you are locked in to their compression routines.

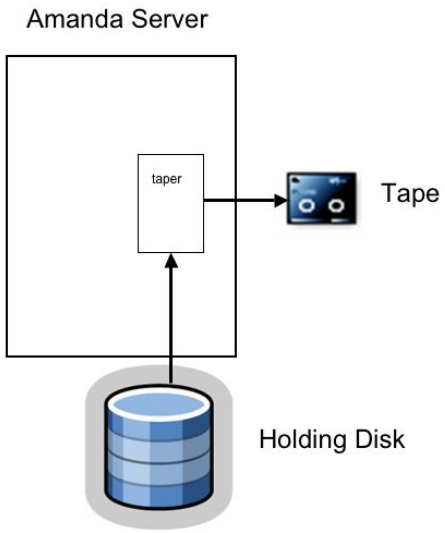
Using a Holding Disk

Backing up to disk first and then moving the backup to tape, is a common function in commercial backup software. Amanda can also do this with holding disks, which are configured on the Amanda server. The purpose of these holding disks is to temporarily store backup data on the Amanda server. The data is stored on the holding disks as the dumps are processed, and are later transferred to tape
or other media.

One reason to use holding disk is that the Amanda server there can be multiple *dumper* processes. Each *dumper* can handle one dump, and write the contents of that dump to the holding disk(s) in parallel with other dumper processes.



Another advantage of backup-to-disk is that it is often faster than backing up to tape. When holding disks are used the client dump is effectively complete once the dump is transferred to the holding disk. After the transfer is complete no additional resources from the client are required. The Amanda backup server can then transfer the backup from the holding disk to tape or other media at its leisure. The *taper* process on the Amanda server performs this transfer.



Another issue that this design addresses is the tendency of tape drives to stop and start if enough data is not fed to the tape drive, often called “shoe-shining”. The ideal situation is the opposite – to feed the tape drive enough information so that it moves forward in a continuous flow, often called streaming. For Amanda backups across the network, it may not be possible to transfer enough data quickly enough to keep the tape drives streaming. By having dumps cached on the holding disk, Amanda significantly streamlines tape drive usage.

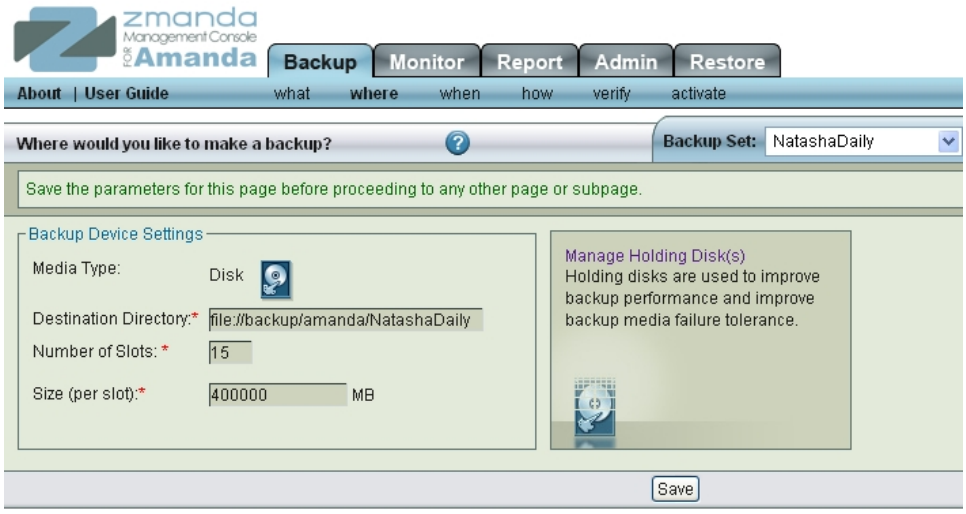
A final issue that is addressed by Amanda’s design is fault tolerance. If you backup straight to tape, and your tape subsystem fails, your backups will fail as well. Holding disks solve this problem, since the backups are cached there. While you fix your tape subsystem Amanda will continue to send backups to the holding disk. And should you need to recover data during this time, no problem – data on the holding disk is recoverable using the same tools and techniques that are used for tape backups.

Backup to Disk

More and more system administrators are using backup to disk. Backup to disk has several advantages over backup to tape, one being the speed of recovery. Most commercial backup packages include some level of support for backup-to-disk, and Amanda does as well. Amanda’s backup to disk is as easy to configure as any other product.

Some commercial products (e.g. NetWorker) can only emulate a single tape drive with a single piece of media. Amanda on the other hand emulates an entire tape library. This allows you as the system administrator to configure both backup to disk and backup to tape in the same way – there is no need to learn two different ways.

In Zmanda Management Console you start your virtual tape configuration by basing your configuration on the hard-disk template. You then define a few parameters on the *Backup -> Where* page:



If you selected backup to disk template while creating a new backup set, Disk will be the only choice for The *Media Type*. The *Size* is the maximum size of each virtual tape. This space is not pre-allocated, so Amanda will

only consume the amount of space required for each virtual tape.

The *Destination Directory* specifies which directory will hold the virtual tape library. ZMC provides a default location (under the *amandabackup* user's home directory), but you can change it to a more suitable location. A best practice is to use a separate disk for your virtual tape system.

The number of slots defines the number of virtual library slots and virtual tape media to create. This creates a virtual tape library structure on disk as shown below:

```
.../vtapes/TestData
        data -> /var/lib/amanda/vtapes/TestData/slot1
        info
        slot1
        . . .
        slotn
```

The symbolic link data is used to point to the slot currently loaded into the virtual tape drive, in this case slot1. The info file is used to keep track of positioning data for the current virtual tape. The directories slot1, slot2, etc. contain the actual backup data.

You may want to archive your virtual tapes to actual tapes for off-site storage, which is very easy to accomplish with Amanda. In the slot directories there are a series of files. By copying the correct files to actual tape (using the *dd* command) you can create a physical copy of the virtual tape.

Backup to Amazon S3

In addition to traditional backup to disk, tape and optical devices, Amanda Enterprise can now use Amazon Simple Secure Storage (S3) Service to backup, archive and retrieve any amount of data, at any time, from anywhere on the Internet. You are getting access to the same scalable, reliable, fast, inexpensive data storage infrastructure that Amazon uses to run its own global network of web sites. Storing a copy of your data on S3 gives you many benefits:

- Streamlined offsite backup storage & recovery process
- Ability to retrieve data anytime, anywhere.
- Restoration of data from S3 is much faster than from offsite tape storage
- Cost effective "pay as you go" ($0.20 per GB)
- Online storage service backed by Amazon's 99.9% SLA

When you configure a new backup set you simply select Amazon S3 as a backup destination (below). After initial configuration, Amanda Enterprise securely sends a copy of your data for safe and secure storage on Amazon S3.

In comparison to traditional storing of tapes off-site for disaster recovery, the Amazon S3 provides streamlined recovery process with minimal downtime for your business.

Recovery from off-site tape	Recovery from Amazon S3
Identify location of the required tapes	Find files or directories using intuitive ZMC and click Restore
Request your off-site tape storage vendor, e.g. Iron Mountain, to deliver the tapes	
Mount and read the tape until required data is found and recovered	
Unload the tape and repeat with another tape if required	
Total downtime for best case scenario is hours, but usually takes several days	**Total downtime is a few minutes**

Conclusion

Amanda is a modern, robust backup system. It differentiates itself from proprietary solutions in several key areas:

1. Amanda is an open-source product, with all the attendant benefits of an open-source product.
2. A unique approach to backup level scheduling is used, which leads to consistency in backup window, media use and amount of data dumped.
3. Amanda uses industry standard dump tools and data layouts. This ensures that you can recover data, even if Amanda tools are unavailable at recovery time.
4. Amanda uses industry standard encryption and compression tools.
5. Amanda Enterprise provides an intuitive user interface making it easy to install, configure and manage all day-to-day backup and restore operations
6. Amanda Enterprise provides a cost-effective alternative to proprietary software with an average cost savings of 60 percent or more over proprietary software.

For further information contact zsales@zmanda.com