



zmanda
Open Source Backup

Amanda Enterprise vTape Planning Guide

Abstract

This guide will help you install and set-up backup to disk utilizing vTapes in Amanda Enterprise. It outlines considerations that need addressing and provides a methodology for determining optimal vTape values. It also addresses some of the questions that arise when deploying vTape backups.



Introduction

When backing-up to disk, Amanda uses a built-in virtual tape (vTape) functionality to store the backup data. When using vTapes, often it is not clear how to estimate the size of the vTape slots, or how many vTape slots to allocate. This guide provides a methodology for determining vTape values. Whether you are using the Amanda community release or Amanda Enterprise, it is a good starting point to show some of the considerations and address some of the questions that arise when deploying vTape backups.



Estimating Your Storage Capacity Requirements

In order to accurately estimate your storage capacity needs for disk-based backup you will need to have a good idea about several parameters.

The easiest one is often the *Full Data Set*, or the size of data you plan to back up. This may be as easy as the total amount of disk space on some servers and/or clients. This is usually not very difficult to at least estimate accurately.

The next one, is usually much more difficult, is the *Incremental Data Set* which is basically the rate of change of your data, or how much data will change from one backup run to the next. An easy way to think of this is the combination of how much new data will be generated and how much old data will be removed:

$$\text{Incremental Data Set} = \text{Data Added} - \text{Data Removed}$$

Often, we have no idea how much of our data changes from day to day. If there is no easy way to estimate or calculate this, make an educated guess, but be conservative. Don't worry! As long as you are able to add disk space¹, you can always adjust this in the future.

Next is the *Backup Cycle*. This is defined simply as the “maximum amount of time that may elapse between full (Level 0) backups”. As the planning and scheduling of backups in Amanda Enterprise differs somewhat from other backup programs, it is important to note that this is, truly, the **maximum** amount of time allowed, and not just the interval between full backups.

Finally, *Data Retention Policy* is how long you want to keep backups around. This is often a matter of practicality with vTape backups, as they usually implemented for ease of restore, rather than archival purposes. Your *Data Retention Policy* may be determined by corporate data archiving policy or legal and regulatory requirements. You can also implement multiple policies using multiple Backup Sets²,

These parameters can be summarized as:

- The size of your *Full Data Set* of data being backed-up (GB)
- The size of your *Incremental Data Set* of data being backed-up (GB)
- Maximum allowable time between full backups, or *Backup Cycle* (Days)
- Desired length of time for *Data Retention Policy* (Days)

¹The best way to do this is to store your vTapes on some kind of volume-managed file system, such as LVM on Linux or ZFS on Solaris.

²A Backup Set is how Amanda segregates work. For more information, consult the Amanda Wiki at <http://wiki.zmanda.com> or check out the “introduction to Amanda” white paper on the Zmanda Network at <http://network.zmanda.com>



These variables all combine to determine the required backup space through the following formula:

$$\text{Storage Capacity Required} = \frac{\text{Data Retention Policy}}{\text{Backup Cycle}} \times (\text{Full Data Set} + ((\text{Backup Cycle} - 1) \times \text{Incremental Data Set}))$$

This looks a little complex, but we can simplify it, by using some understanding about how the planning mechanism inside of Amanda Enterprise works.

By default, Amanda Enterprise creates its plans to balance the data stored for each backup run. This means that full backups are normally spread relatively evenly around the *Backup Cycle*, and we can use this fact in our planning. We can assume that for each backup cycle, we will have the equivalent of one Full Backup of each client (the *Full Data Set*), and the rest are Incremental Backups (*Incremental Data Set*.) This is where this part of the equation comes from:

$$\text{Backup Data} = \text{Full Data Set} + ((\text{Backup Cycle} - 1) \times \text{Incremental Data Set})$$

We can then just multiply this value times the number of times our *Backup Cycle* fits into our *Data Retention* policy. So, a simpler way of writing the equation becomes:

$$\text{Storage Capacity Required} = \frac{\text{Data Retention}}{\text{Backup Cycle}} \times \text{Backup Data}$$

An Example

While the formula may seem complex, it really becomes simple if broken down into component parts. Let's use an example to illustrate it. Let's use the following values for our example:

- *Full Data Set Size*: 100GB
- *Incremental Data Set*: 10GB
- *Backup Cycle*: 7 Days
- *Data Retention*: 14 Days

This makes the equation look like this:

$$\text{Storage Capacity Required} = \frac{14 \text{ Days}}{7 \text{ Days}} \times (100\text{GB} + ((7 \text{ Days} - 1 \text{ Day}) \times 10\text{GB})) = 320\text{GB}$$

Since our *Full Data Set* size is 100GB, a full (Level 0) backup will create a vTape archive of this size³. Because our *Incremental Data Set* size is 10GB (our Rate of Change of data is therefore around 10%), our incremental (Level 1) backups will be 10GB.

There will be at least one full backup during the amount of time specified for the *Backup Cycle*, so if our cycle is 7 days long, we can assume that there will be at least one full backup during that period, and the rest are likely to be incremental backups.

This means we have one full (100GB) backup, and six incremental (10GB) backups for the remainder of our cycle (7 days - 1 day.) As we are storing 14 days worth of data, or 2 cycles, we need a total of 320GB of space for this backup set.

Where things get interesting is how these values translate to Amanda Enterprise options. In order to satisfy our *Data Retention Period* requirement, we will need to have 14 vTape slots. In order to satisfy the *Full Data Size* requirement, each slot will need to be able to hold a full backup. This would seem to indicate that we should need:

$$100\text{GB} \times 14 \text{ vTapes} = 1.4\text{TB}$$

1.4TB is a lot larger than our calculated 320GB! However, vTapes only utilize the space on disk that is needed at each time. While each vTape needs to be able to hold the full *Data Size*, as we pointed-out earlier, we are only going to store a single full backup for each *Backup Cycle*, and we are allocating a *Data Retention* policy of storing two backup cycles. The rest will be incremental backups, and will be much smaller:

$$(100\text{GB} \times 2 \text{ vTapes}) + (10\text{GB} \times 12 \text{ vTapes}) = 200\text{GB} + 120\text{GB} = 320\text{GB}$$

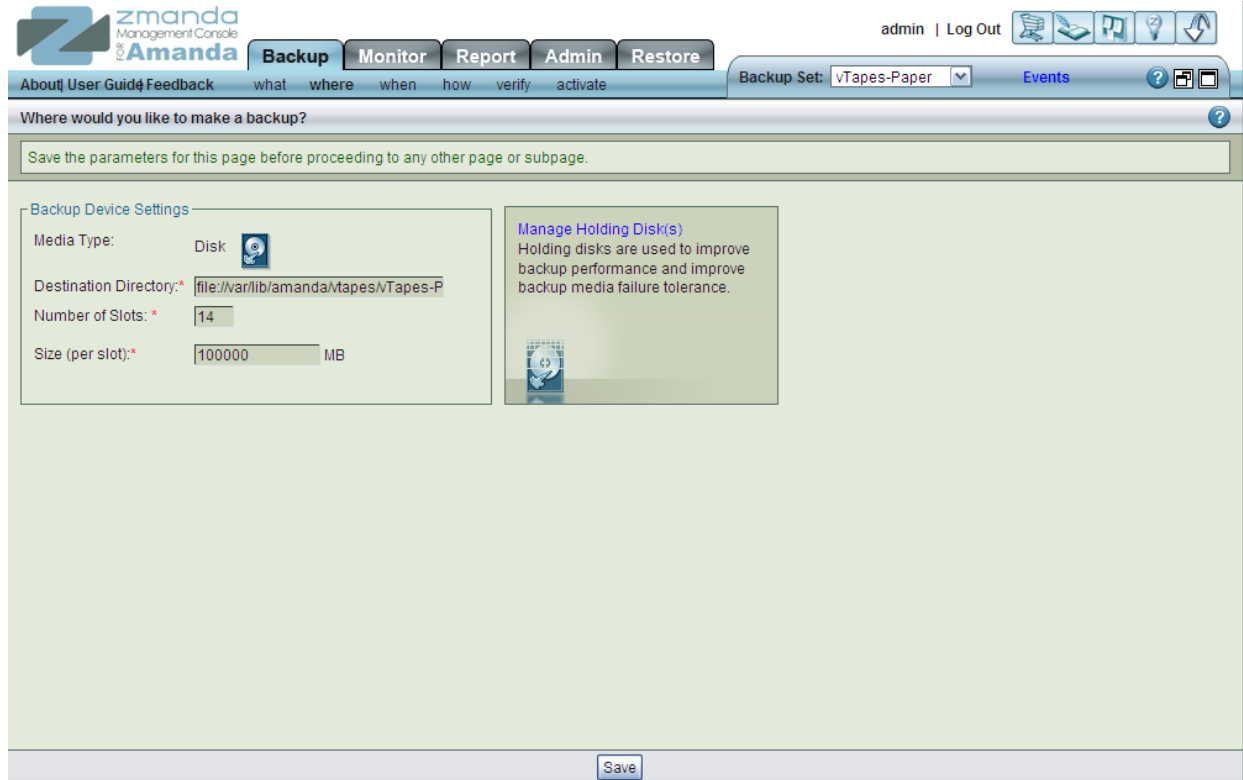
³ Let's assume, to simplify things, that we are not using any compression for now.



That's more like it!

Using the Values in the Zmanda Management Console (ZMC)

Now that you’ve calculated some values, where do they go? In the Zmanda Management Console (ZMC) for Amanda Enterprise 2.6.4, the values we calculated are used to populate certain fields in the Backup Set.



First, on the “Backup Where” tab, the “Number of Slots” correlates to the maximum number of vTapes that you plan to use. This will be equal to the number of days you planned for *Data Retention Policy*. (In our example, this would be 14.) The vTape size needs to be the maximum amount of data that we need to back up for a full (Level 0) backup, which is our *Full Data Set Size* which is 100GB or 100000MB.



admin | Log Out

Backup Set: vTapes-Paper

When would you like to do a backup?

Specify here the backup schedule for this backup set.
NOTE: After configuration and verification, go to the [activate](#) page to activate this backup set.

Backup Cycle

Time Range:* Days

Backup Cycle:* 7 Days

Backup Time:* 2 Hr 0 Mins

Volumes in Rotation:* 14

Backup Week

Mo: Tu: We: Th: Fr: Sa: Su:

Specify which days to perform backups.

Save

Next, on the “Backup When” page, the “Backup Cycle” entry is the same as our *Backup Cycle (7)*. The “Volumes in Rotation” should be the same as our *Data Retention Policy (14)* and also the same as the “Number of Slots” from the “Backup Where” page (14). Using a value less than the “Number of Slots” wastes space, as only one vTape is used per Backup Run.



Planning Considerations

The methodology above provides a good set of starting calculations for determining vTape sizes; however, every installation will be unique. In order to help in customizing the settings and make them more accurate, consider the following when making your initial assumptions for values:

- When planning your *Full Data Set Size*, it will probably grow over time. Unless you are backing up a finite set of disks that will not change or be added to, build-in some room for growth.
- Offsetting what was just said is data compression. For our example, we only used raw data sizes. In reality, your data may compress very well, and you may be able to get by with much less disk space allocated. Amanda Enterprise utilizes standard compression formats (gzip and IZip, so you can utilize those tools to see how “compressible” your data is, and add that to your calculated *Full* and *Incremental Data Set Sizes*.
- Unless you have some sort of highly predictable data generation process, the *Incremental Data Set Size* is most likely an estimate. You will probably want to build-in a “fudge factor” just to be sure. Over time, the rate of change of your data can be found by monitoring how much data gets backed-up for *Incremental Data Set Sizes*. An easy way to do this is to create a report in the “Custom Reports” tab in the Zmanda Management Console, and to periodically examine how much data you're backing-up.
- You know that “only one full backup per *Backup Cycle*” assumption? It may be wrong – especially if your *Rate of Change* of data is really high or low. Amanda often attempts to promote clients to a full (Level 0) backup, especially if they appear and disappear from the network during backup runs, or have a high rate of change of data. This can affect how fast your vTapes fill-up.
- Remember that this has to be done for each Backup Set in Amanda Enterprise. You will probably have several of these calculations to run, and will need to have enough room for the aggregate storage requirements.
- All of these parameters are not set in stone! If you plan your vTape storage on an expandable platform (RAID, NAS, LVM, ZFS, etc.) then you can always grow it over time, and expand the size (or number of vTapes!) you use for storage.



Conclusion

This guide should be a good starting point to assist in your planning of vTape sizes for disk-based backup in Amanda Enterprise. While there is no “one true way” for planning vTapes, the methods presented will allow you to start from information you already know, and to determine or estimate values you do not have.

Additionally, while the bulk of configuration work is done when you first deploy Amanda, you will need to monitor and observe your backup runs over time. Remember to read your reports! Compression ratios, backup run sizes (which tell you the rate of change of your data), vTape usage, and other important clues are there. This will help you fine-tune the parameters of your Backup Set, and prevent backups from failing because you run out of space.