



# *Key Issues in Storage Resource Reporting*

*An NTP Software  
White Paper*

## **Abstract**

*The storage resources on your network are one of its most costly components and a critical service for your clients. Understanding how these resources are being used and their contents is critical to the success of your storage strategy. In spite of this fact, many institutions do little more to understand their investment than check available space on their drives.*

*The general perception that storage costs little is false. A reporting strategy of only looking at storage when problems arise is unnecessarily risky and not cost effective. This paper discusses the key issues that underlie Storage Resource Reporting and provides a framework for addressing these issues.*

Rev 1.0, October 2005

The information contained in this document is believed to be accurate as of the date of publication. Because NTP Software must constantly respond to changing market conditions, what is here should not be interpreted as a commitment on the part of NTP Software, and NTP Software cannot guarantee the accuracy of any information presented after the date of publication.

This Technical Note is for informational purposes only. NTP SOFTWARE MAKES NO WARRANTIES, EXPRESS OR IMPLIED, IN THIS DOCUMENT.

Product and company names mentioned herein may be the trademarks of their respective owners.

NTP Software, 15 Charron Ave., Nashua, NH 03063, USA

## Introduction

The storage resources on your network are one of its most costly components and a critical service for your clients. Understanding how these resources are being used and their contents is critical to the success of your storage strategy. In spite of this fact, many institutions do little more to understand their investment than check available space on their drives.

The general perception that storage costs little is false. A reporting strategy of only looking at storage when problems arise is unnecessarily risky and not cost effective. This paper discusses the key issues that underlie Storage Resource Reporting and provides a framework for addressing these issues.

The purpose of this paper is to help you understand the issues surrounding Storage Resource Reporting. Specifically, we will talk about the issues relating to the many approaches to understanding your existing storage as well as the alternatives in using this information to drive efficiencies in your enterprise.

After reading this paper, you should better appreciate the alternatives open to you and understand how to mitigate risks by implementing best practices based on solid reporting. You should also be able to identify practical and effective solutions to the Storage Resource Reporting problem.

## What is Storage Resource Reporting?

Networks (whether Windows based or otherwise) must provide large amounts of storage resources that are accessed by your user community which often numbers in the tens of thousands. Unlike stereo equipment, networks have many more moving parts and require provisioning on an ongoing basis, changes to storage access, upgrades to existing hardware and software. In these days of ever increasing storage consumption the demand for more compliance coverage and larger items of storage, provisioning of storage resources has become a constant activity that keeps administrative staff busy. Ultimately it would be nice if the owner of the storage infrastructure could spend five minutes a week to keep their company's network humming along, but this only works for organizations where the storage resources are covered by a reliable and comprehensive reporting system.

The trick is to ensure that you avoid getting buried in the data. With Terabytes of storage in use it does no good to spit out scattered facts or worse yet, reams of meaningless numbers. The standards of practice for storage resource reporting will decide if it is actually a planned growth network or a jungle of data sumps. There are true horror stories of rogue users and administrators stealing and hoarding company storage resources for uses that are not only off mission, but dangerous from a liability standpoint. These threats along with the potential fact that company data systems are a valuable resource whose use is worth understanding are the real reason that the practices outlined in this document must be adopted.

## The Various Reporting Models

Once you understand the task at hand it is obvious that a good network requires good storage resource management, but before you can get to the task you first need eyes into the current situation and a way to measure that progress is being made once steps to improve are taken. On a network this information has historically been associated with virtually random observations by administrators that happen upon a cache of inappropriate files or who notice that disk space is rapidly diminishing on a specific drive. In the days of thousands of files and hundreds of megabytes in network storage resources this wasn't such a bad model. We will call this the Spot Check Storage Reporting Model. It represents the assumption that if you regularly browse around you will find the problems with a little experience and vigilance. Makes sense on the surface, but as you will quickly realize, in all but the simplest environments this is an unworkable mechanism.

## Problems with the Spot Check Storage Reporting Model

In a small single file server environment especially where the administrator is the owner of the business then the Spot Check Storage Reporting Model will often suffice. While we will still argue against some of its component parts even in the smallest of organizations, it doesn't pose the risks we see it inflicting on larger networks. The problems with the Spot Check Storage Reporting Model include that it depends too much on the observational skills of the network staff, it limits the ability to see trends and be predictive, and it requires a Herculean effort to actually survey even a medium sized network's storage resources. Any one of these may not be enough to convince a company to move onto a more advanced reporting model; however, when taken together it is hard to accept the cumulative failings.

Let's look at how each of these problems we have listed manifest themselves in the Spot Check Storage Reporting Model:

### Spot Checks

Administrative staff is busy and too distracted to be relied on to catch every nuance of your storage usage as they carry out their regular duties. Often they are not even looking at the contents of the storage devices, but instead are running utilities and diagnostics. These tools shield them from the raw data.

Looking for trouble and torn loyalties also play a part in undermining the administrative spot check mechanism as personal relations and grudges are likely to influence who gets policed and who gets a pass on what they store on your network.

### Inability to Trend and Predict

If you are told that your NAS has X files along with the distribution by file type then this is an interesting bit of information, but it fails to describe where things are going. Without a track record and an easy way to visualize how the data hangs together you are left to guess about your best next move. Trending allows a quality Storage Reporting package to predict when your critical devices will likely run out of space based on actual experience in your network rather than relying on statistics that predict storage growth on abstract industry averages. If marketing is consuming 20 GB per month for the last three months that may be a disaster waiting to happen if you find out too late. Even the users

are often surprised to see what space they have consumed once the reports bring it to light.

## **Volume of Data to Survey**

The concept of a trillion plus files is no longer speculation, but no human or even reasonably sized team of humans can hope to browse and spot irregularities in such a massive repository. Expecting this trend of larger and larger data repositories to continue has become an accepted fact. As our ability to create and retrieve data increases with the introduction of powerful software algorithms we have to rely on equally powerful automated systems to help us see the forest for the trees.

## **Reporting Solution Architectures**

Once we discard the somewhat luck based Spot Check Storage Reporting Model the big choice becomes between real time and interval based reporting solutions. While real time reporting is attractive on the surface, it is fraught with problems. If you implement a real time solution on a network of any material size you will not be able to duplicate information. The hallmark of a usable reporting system for business decision making and planning purposes is to not only have stable points of information, but to also have the ability to look back in time. Imagine trying to do calculations using the real time national debt. Your results would vary with the second that the calculation was performed. You would also be beset by the storage needs of keeping history information to the extent that the history would quickly eclipse in size the original storage reported against. Finally in very large storage scenarios you are likely to damage the overall performance of the system in your attempts to keep up with the changes as they happen.

For these reasons we look to a periodic model. In a periodic reporting model such as that used by NTP Software Storage Modeling & Analysis (M&A) agents run on the individual storage hosts and report back to a centralized data store. In the case of Storage M&A this centralized reporting database is an ODBC compatible SQL Server. This system provides a web based front end to view the results and even provides ad hoc query features. Given three or more data points accumulated in the database, the system can project forward the increasing use of storage and predict when the device will exceed capabilities.

## **A Better Model of Storage Resource Reporting**

As described above, NTP Software's Storage M&A product exemplifies the best in a Periodic Storage Reporting Model. Using this product as an example we will discuss how this model solves the hardest problems presented by reporting on massive storage infrastructures with minimal pain and inconvenience.

### **Ability to forecast needs accurately**

How much storage will you need tomorrow, next month, next year, and five years out? These are things that anyone who has to budget for storage has struggled with. M&A gives you proof through statistics based on historical fact. Understanding your company's growth rate based on actual numbers gathered on your network allows you to know how much storage to plan for and rather than having to rely on a guess.

### **Easy to Find the Answer**

When you have a periodic snapshot based system the organization can rely on the reports as it does other periodic information. Reliable reporting is repeatable reporting whereas a real-time system changes as you look at it making it impossible to get a grip on what the data means. It would be a disaster to represent the situation and then not be able to reproduce the reports because the data changed unexpectedly (i.e. covering tracks). Ease of use makes this kind of system much preferred over other variants. Sometimes the answers are to questions that even seasoned network professionals and managers do not even realize need answering. By having a reporting system you get the reports most asked for by the community based on experience of what metrics actually matter to the health of your storage infrastructure. Breaking storage down by extension for instance is a very useful way to clear out the noise, but understand exactly how storage is utilized. This kind of report often identifies files that should not be stored on servers given company policy or just common sense.

## **Accountability**

While blame is never the end goal, it is important to understand who is causing problems and correct their behaviors via education, punishment or simple exposure of the error. A solid reporting system will have to delineate files by owner in a clear report that helps pinpoint action points such as users with huge home directories. A typical organization will see the 80/20 rule even in this area with a small number of users having home directories in the Gigabytes, while the average size could well fall below 10 Megabytes. The leverage of addressing just the small number of “storage abusers” often regains vast amounts of storage in a short period of time. Also knowing that the store is being minded will influence the user community to be more responsible in what they put on company storage devices.

## **Chargeback**

Since the storage infrastructure is not a profit driving part of most businesses it is often highly desirable to be able to chargeback or bill departments for their usage. Manual audits for this information are time consuming and error prone. With a source of data that gives you periodic breakdowns of actual usage you have a situation tailor made for implementing a fair and highly reliable chargeback system that will help justify proper provisioning of your companies often under appreciated storage infrastructure.

## **Centralized Configuration**

NTP Software Modeling and Analysis agents are centrally configured to get all the information that’s important to report on your environment. This configuration is done with just a few clicks on a secure website. Each agent will then check its configuration and run as instructed. This central administration makes it easy to configure 1 or 1,000 agents to run and gather information on all the storage being consumed on your network.

## **RIAA and Other Outside Reasons to Change**

The Recording Industry Association of America (RIAA) and other media associations have recently gone on a fairly public crusade to punish companies and individuals (but mostly companies) that have pirated music and other media files. Not only are MP3s and other media files huge, they can also cost your company dearly in liability due to lawsuits such as these. Even the primary law firm that is prosecuting the lawsuits for the RIAA has searched for a mechanism to ensure that they are not in breach of the same rules they are enforcing.

Even in cases where you do not succeed 100% in keeping out violating materials, the fact that your company has gone through the efforts of implementing a solution and setup a regimen on your primary storage platforms of regular reporting with a product such as NTP Software Storage M&A can mitigate your vulnerability. By configuring the agents to gather data on media files you will know when a user is putting your organization in danger of violating the rules.

## **Conclusion**

The best practices outlined in this document are typically only seen in organizations that have made a strong commitment to knowing and managing their storage infrastructure. As storage increases geometrically year after year the old practices are not only inadequate, but they are indefensible. In this escalating environment we feel that having a comprehensive Storage Resource Reporting System in place is no longer optional!

Virtually all organizations will benefit from knowing how their storage is being used and or abused. In this paper we have sought to point out where your organization may look to change or validate existing practices in the areas of storage resource reporting.

NTP Software Storage M&A is a powerful weapon in your battle to manage your organizational storage resources. It is the best of breed storage resource reporting product on the market and will help you get the handle you need on one of the most valuable assets in your company.

## **Related Documents**

### **White Papers**

- Do You Need a Storage Management Policy?
- NTP Software QFS Best Practices
- Understanding EASE

### **Technical Notes**

- NTP Software QFS vs. Windows Native Quotas
- Active Directory Storage Quotas