

Lowering the Cost of Deleting Data

© 2014, SolarWinds Worldwide, LLC. All rights reserved.

Follow SolarWinds:   

One analogy that's common in the storage industry is that going through old file systems is like cleaning out the garage. It's a big job that needs to be done, but nobody wants to do it. Ironically, since the cost of storage capacity keeps dropping it can sometimes seem more cost-effective for companies to add storage (build a bigger garage), than to go through their existing data. So historically, most companies haven't regularly cleaned out obsolete files, a situation that's prompted IT Managers to ask "How do I remove old data from my storage systems?"

Cost of file deletion

Part of the problem is the cost of determining which files should be deleted; this could be called the cost of "file deletion". Since most of the data that's accumulating on corporate file servers is owned by people and departments outside of IT, cleaning it out requires getting other people to spend time going through old file stores. And when they finally do, odds are they'll just say to keep much of that data anyway. Users often don't remember what exactly is in each file, especially older files, and therefore, don't feel comfortable parting with it. This results in even less data being evicted, increasing the cost of file deletion.

Cost of file storage

The real cost of file storage is more than simply buying storage media, it also includes the cost of overhead and related infrastructure, such as storage arrays and networking. The cost of backup must be included as well since each GB of primary data can consume up to 5GB of capacity in the backup system. While the raw cost per GB of storage keeps dropping, new technologies that can handle large data stores, like object storage are implemented, but it's still not enough to keep up with data growth in most organizations.

Cost of change

There's a principle of behavior that states people won't change until there's enough pain with what they're currently doing. In business, cost is substituted for pain, but the principle is the same. Businesses keep doing things the same way until it's less expensive to change the process. In this scenario, the cost of file deletion represents change and file storage represents the cost of staying the same. By this equation, until the cost of deleting data comes down, many companies won't be compelled to make any changes, they'll just keep buying more storage.

But what if the process of going through existing data sets was easier and less expensive? It could indeed be cost effective to stop building more storage infrastructure and instead start deleting old files. This could have some real benefits down the road when the long term costs of storage are factored into the equation, or when other aspects besides cost are considered.

Cost isn't the only issue

In addition to cost containment, compliance regulations can force a company to get their file storage 'house in order'. For example, if personal information must be stored on certain systems in order to meet data protection or security

requirements, companies need to make sure these files are not scattered across the environment. Similarly, an eDiscovery motion - or just the threat of one - can make an IT Manager wish their file storage was better organized.

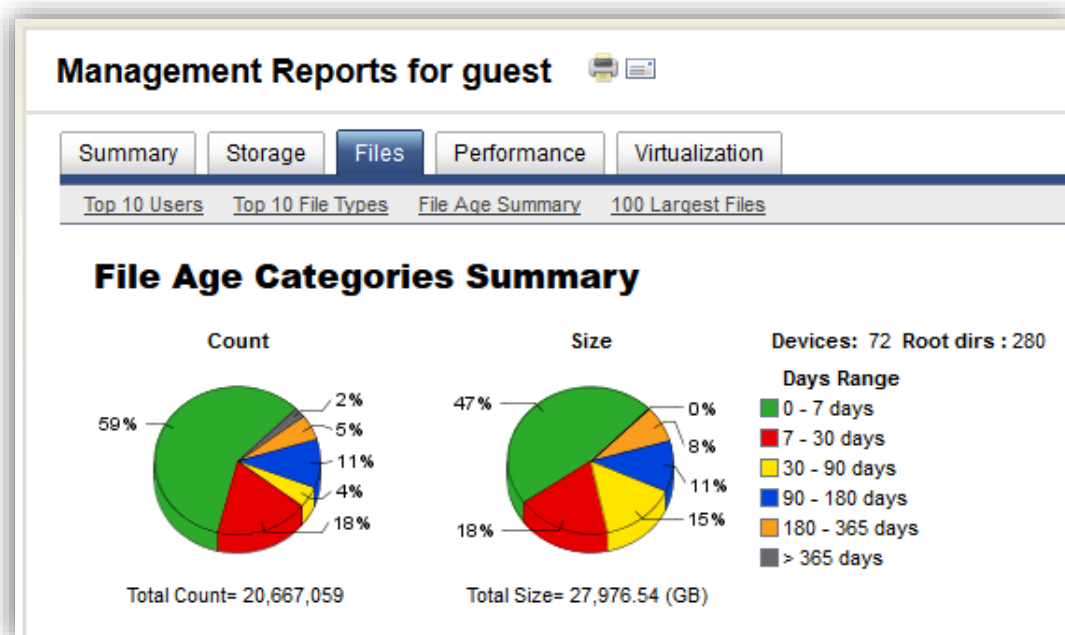
It's clear that reducing the cost of file data management and organization is a valuable thing to have and increasing the effectiveness of that system is also valuable. But what would such a system do?

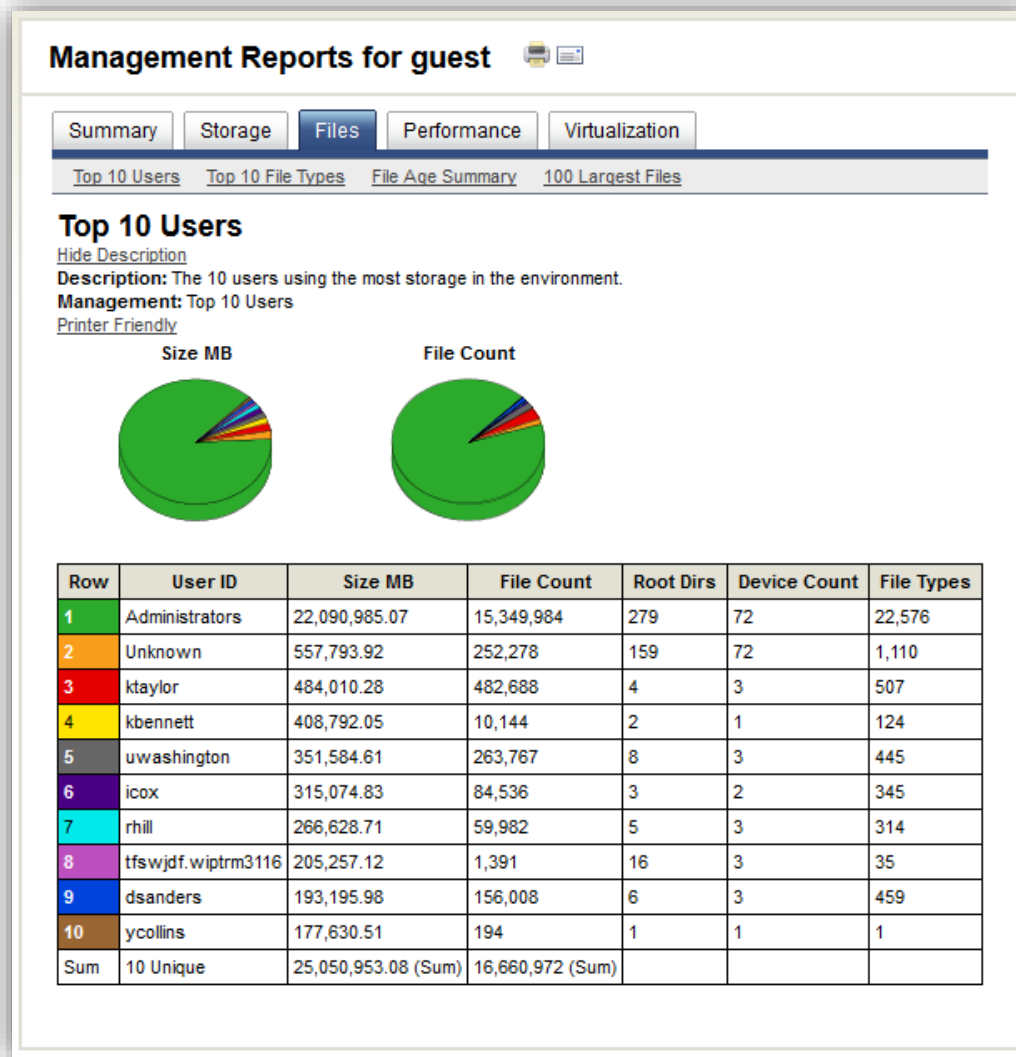
File Analysis

To start with, this system should perform a detailed file analysis, so that files can be accurately identified and their value determined, and so that file stores can be organized in meaningful way. After all, these files will need to be accessed at some point as well. This file analysis should capture data about each file, such as the file type, its age, the owner(s), where it's stored, etc. This data will become the foundation of a more sophisticated file management process.

Create rules, policies and reports

File characteristics should be available to create rules and policies. These can then be applied to data stores automatically based on specific requirements of the company and the users that are examining file data. As an example, [SolarWinds® Storage Manager](#) can combine file age and size characteristics to create a rule that lets administrators know how many files greater than 1MB exist and that are older than one year. Or they can combine file owner and file type to identify which users are the worst offenders of the company's policy against storing MP3 files.





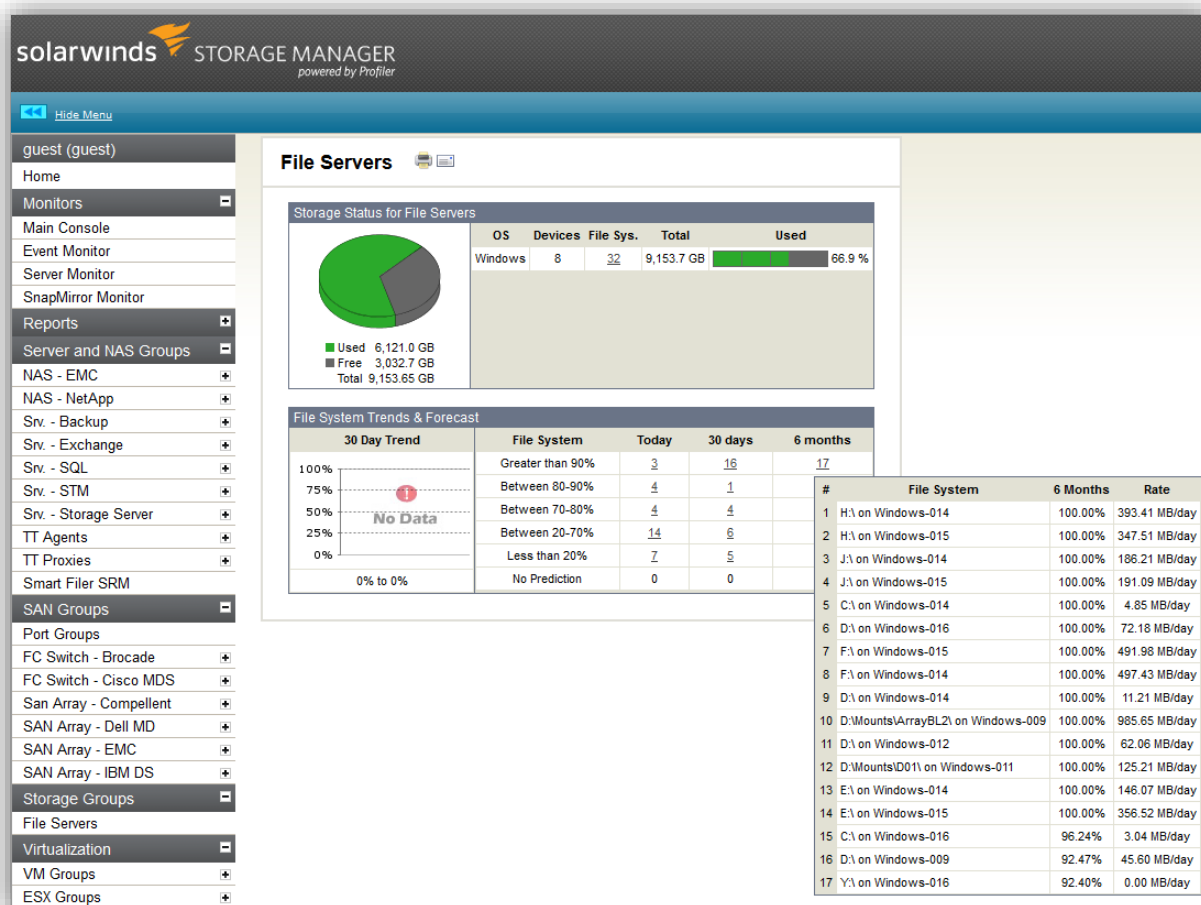
This rule can be applied automatically by specific policies that run it every month to generate a report that can help users identify which files are good candidates for deletion. These reports can also provide IT with some insight into their file organization problem. Aside from the immediate relief of data reduction and identifying specific files, this is the reason to conduct this work, so that the problem can be proactively managed. But in addition to file analysis, rules and policies, what other characteristics should this system have?

Flexibility

The system should allow users of different abilities and backgrounds to customize data captured from the system. Users will learn more about their environments as they start conducting searches and require the ability to drill down on specific files and characteristics. Returning to the 'garage full of junk' analogy, from the outset you're not sure if you'll need shelves, hooks, or several large bins to organize the garage's contents. File owners and IT administrators need the flexibility to deploy a number of different tools as well.

Comprehensive

Files can live anywhere, on traditional file servers, on NAS appliances, or on storage that's directly attached to application servers, physical servers, or VMs. The file analysis system must be able to see files wherever they live and consolidate their management at the device or share level, as well as by the different characteristics listed above. Running different utilities to manage each different storage system is just not an option. SolarWinds Storage Manager allows IT to look at their entire formatted space, whether on file servers, NAS devices or application servers - physical or virtual.



Simple to use

The file analysis system should have a strong set of default tools. In turn, the user can apply them and easily get value from their time invested. If data owners are required to conduct this work, the system should be easy for users to get started and see progress quickly. This means that default configuration settings should be available, along with a complete set of pre-defined reports.

Strong filtering

Historically, companies have been letting this pile of data in their file shares grow into mountains. The first steps in tackling a file organization problem is running different filters to cut the job down to size, since users can easily get overwhelmed when faced with a mountain of data.

Conclusion

Exploding file stores are the 'elephant in the room' for many companies. While the cost of raw data storage keeps dropping, the cost of deleting obsolete or inappropriate data can be prohibitive. This results in the unspoken accumulation of a proverbial mountain of data that has both cost and compliance ramifications for the organization. The reason for this is the difficulty in identifying which files are good candidates for deletion, a decision process that usually requires effort by the data owner, not just IT. By applying file analysis techniques in a flexible, easy-to-use format, comprehensive file organization solutions can reverse this trend and enable companies to get their collective arms around their file storage problems.

SolarWinds Storage Manager

- Monitoring storage performance & isolating hotspots in your multi-vendor SAN fabric
- Maps virtual machines to physical storage
- Automates storage capacity planning & reporting
- Simplifies analysis of storage usage & reclamation of storage space
- Monitor storage arrays from EMC®, NetApp®, HP®, IBM®, Dell®, Hitachi Data Systems® (HDS), and more
- Integration with Virtualization Manager and Server and Application Monitor for end-to-end visibility from application to virtualization to the storage

[Q LEARN MORE »](#)[↓ DOWNLOAD FREE TRIAL](#)

About SolarWinds

[SolarWinds](#) (NYSE: SWI) provides powerful and affordable IT management software to customers worldwide. Focused exclusively on IT Pros, we strive to eliminate the complexity in IT management software that many have been forced to accept from traditional enterprise software vendors. SolarWinds delivers on this commitment with unexpected simplicity through products that are easy to find, buy, use, and maintain, while providing the power to address any IT management problem on any scale. Our solutions are rooted in our deep connection to our user base, which interacts in our online community, [thwack®](#), to solve problems, share technology and best practices, and directly participate in our product development process. Learn more at <http://www.solarwinds.com>.