TechTarget

Special Report

Data Archiving 101

Every organization needs to have some form of archiving strategy in place. This quick guide features tips and best practices that can help you get started with a new data archiving plan, or help refine one that you already have running.

Data Archiving Best Practices: The Difference Between Backups and Archives

One of the most common questions that come up again and again in data backup and recovery is "What's the different between data backups and archives? Data archives are often confused with data backups. Data backups are used to restore data in case it is lost, corrupted or destroyed. In contrast, data archives protect older information that is not needed for everyday business operations but may occasionally need to be accessed. It's crucial to learn data archiving best practices -- an effective data archiving strategy is a necessary part of every IT organization.

Data archiving is the practice of moving data that's no longer being used to a separate storage device. Data backup expert and a senior consultant with Long View Systems Inc. Pierre Dorion defines data archiving as "a single or a collection of historical records specifically selected for long-term retention and future reference." In addition, data archives consist of older data that is still important and necessary for future reference, as well as data that must be retained for regulatory compliance. Data archives are also indexed and have search capabilities so that files and parts of files can be easily located and retrieved.

To help you stay up to date on the latest information on data archiving best practices and data backup, we've collected our top five tips on data backup vs. archiving. Learn why you shouldn't use your backups as archives; whether you use should tape, disk, or the cloud for archiving; and if you go with cloud archiving, how to choose the best cloud archiving service. In addition, learn about how archiving your data prior to backup can increase your data reduction ratios.

DATA ARCHIVING BEST PRACTICE #1

Data backup vs. data archiving

The backup vs. archive debate has been going on for years, and backup and recovery pundits are constantly saying "backups aren't archives." But data backup and recovery software vendors have started to integrate different functionality into their software, so is this still true? Some examples include data deduplication and data lifecycle management with storage tiering. Is backup catching up with archive functionality, or are the two so fundamentally different that there will always be a void that backup will never be able to bridge? Learn about data backup vs. data archive in this expert tip.

DATA ARCHIVING BEST PRACTICE #2

Don't use your archiving storage as backup storage

Another important distinction between backups and archives, says W. Curtis Preston, independent backup and recovery expert, is backups are for disaster recovery and data archives are for discovery. A data backup is for restoring lost or corrupted files. So if you accidentally deleted some files, and you need to restore things to the way they previously looked, if you have your backup, you're still in business. In addition, traditional backup software isn't going to help you with archiving so it's important to have separate data archiving software. In this podcast with Curtis, learn about the do's and dont's of using backup storage as archiving storage.

DATA ARCHIVING BEST PRACTICE #3

Data archiving with tape, disk or the cloud

The Storage Networking Industry Association (SNIA) defines an archive as "A collection of data objects, perhaps with associated metadata, in a storage system whose primary purpose is the long-term preservation and retention of that data." In addition, data that is archived is not usually expected to be readily searchable. This definition sounds simple, but presents many problems for administrators. For example, the type of media the data is stored on will affect the speed and ease with which it's restored. Your three basic choices for archiving are tape, disk and the cloud. But how do you choose what's best for your organization? Which is the most expensive? In this article, get the answers to these questions, and learn about data archiving and tape, disk, and cloud storage.

DATA ARCHIVING BEST PRACTICE #4

Cloud archiving services

As mentioned earlier, a data archiving strategy that's being implemented by more and more organizations is to use a cloud archiving service. Even though there are a plethora of cloud data storage providers, there are only two online storage services that can really be considered viable for enterprises interested in cloud archiving. In the author of this article's opinion, these two companies are Autonomy Zantaz and Iron Mountain Inc. In this article, learn about the pros and cons of these two different cloud archiving services, and how to choose the best cloud service provider for archiving.

DATA ARCHIVING BEST PRACTICE #5

Data archiving increasing data reduction when done prior to data deduplication

In enterprise data storage, the theme for the past year has been to "do more with less," and some users are controlling data growth by archiving their inactive data before it

ever enters the data backup cycle. While this archiving requires some work, storage managers at organizations with data archiving in place have found additional benefits, particularly in data backup, another place where data growth has challenged budgets and infrastructures this year. When this process is done before data deduplication, some users are able to reduce even more data. In this article, learn about data archiving and data deduplication and how using both of these technologies can reduce your backup data.

Archive vs. Backup and Why You Need to Know The Differences

Backups are primarily used for operational recoveries, to quickly recover an overwritten file or corrupted database. The focus is on speed, both to back up and recover, and on data integrity. Archives, on the other hand, typically store a version of a file that's no longer changing, or shouldn't be changing.

Speed is less important in archives; even if the event is a legal action, you typically only have a few days to respond. Searchabilty is more critical in archives. In addition, importance is placed on the ability to scale data integrity and data retention over a long period of time, possibly decades. An archive is no longer limited to traditional files and images; most database applications have specific archive capabilities to allow the primary database to stay lean and fast while the archive is retained for research and compliance.

Email archiving applications are often the catalyst for establishing a separate archive process. It's important to realize that you are legally responsible to do more than just capture email.

When considering combining archive and backup onto a single platform, the decision will depend on the specific platform, what the organization's retention requirements are, and the expected goals of the backup and archive process.

Can tapes be used for archives?

While the vast majority of organizations consider tape for their long-term archives, and companies like Index Engines Inc. provide the ability to more effectively search for data on tape, there's a risk in counting on tape for the storage of archive data.

Just as disk has become a popular addition to the backup process because of the concerns about recovery from tape, data that's archived to tape should be considered

just as vulnerable. It's difficult to develop an ongoing process to verify the integrity of tape, leading to greater concerns the longer the media sits on a shelf. There's also a simple technology issue. Even if your retention requirements are only seven years, think back seven years ago -- LTO-1 or LTO-2 was becoming the standard, DLT in the Super DLT form was still considered competition. What's the likelihood that the LTO-1 tape that's been sitting on the shelf can be read and successfully restored from in the new LTO-4 drives in your data center? Anecdotal stories of sub-50% success rates aren't uncommon.

Even if the hardware works, how are you going to find a piece of data that's seven years old from hundreds -- and possibly thousands -- of tapes? Most backup applications don't keep their metadata (the data about the data being backed up) very long. In fact, the average length of time is approximately 90 days to 120 days. After that, it's up to your records-retention skills or the person who had your job before you, or even the person before them. Recovery of data this old is most likely going to require guesswork, plenty of manual scans and lots of time.

Should you use disk for archives?

The thought of keeping all archives on disk may seem impossible and costly, but companies like EMC Corp., Hewlett-Packard Co. and Permabit Technology Corp. are delivering technology today to make a disk archive that can last for 25, 50 or 100 years a reality. But, the disk drive you start with today won't be the same disk drive that you use 100 years from now (if we still even use disk drives 100 years from now).

While disk also makes combining the process of backup and recovery into a single platform more realistic than tape, a best practice is to have a specific system for archives. Archives have different retention requirements, different recovery needs and different searchability requirements than backups.

Most disk-based archive systems present themselves as a network mount point, which makes access over time realistic. Unlike a seven-year-old tape drive, you access a CIFS or NFS mount in an almost identical fashion from seven years ago as you do today.

The Importance of Lifecycle Planning for Exchange Email Archives

In many organizations, message archiving is thought of as something that has to be done just to keep the lawyers happy. Typically, a lot of planning goes into implementing

a message archiving system in Exchange Server organizations. Most of it, however, focuses on controlling costs and ensuring that the proposed archiving system complies with regulations that apply to that organization's particular industry.

Message lifecycle planning is the process of developing an email message retention policy. Part of the planning process involves making up-front decisions about storing and retaining email messages, such as where messages will be archived, how those archives will be backed up and for what length of time messages should be retained. When it comes to creating a messaging archival strategy for your organization, less can be more.

Unless there is a critical reason behind retaining messages indefinitely, it's often better to store messages for no longer than what is required by the law or by your business needs. Keep in mind that your message archives typically contain an all-encompassing picture of how your organization operates.

Unless explicitly forbidden, virtually every aspect of a company's day-to-day operations is discussed through email. For instance, companies typically use email to communicate with customers, negotiate contracts, plan meetings and discuss marketing strategies on new products.

It's important to realize that your message archives can be a double-edged sword. On one hand, they contain valuable information related to the organization's business. On the other hand, message archives contain references to some of your company's dirtiest little secrets.

Complying with message-retention policies

Planning a message retention policy is something of an art form. You have to retain messages long enough to comply with any applicable government regulations. Beyond that, though, you should consider how long the messages are going to be of value to your organization.

You should also consider the risks involved with storing older messages that have exceeded the required retention period. Remember: All regulations that require email messages to be archived and retained for a specific period of time do not exist to benefit the company specifically. Email archives exist to allow lawyers to search for evidence of wrongdoing in the event of a company lawsuit. These archives can actually be used against the company. Some companies retain all email messages indefinitely to ensure that no one can accuse the company of not being in compliance.

But imagine if a company was involved in a lawsuit in which message archives were subpoenaed. Lawyers won't simply ignore older messages just because those messages

no longer need to be retained. Along these lines, consider how long the message format you're using for your archives will be valid.

Many current message archiving products use .PST files as a repository for archived messages. But what would happen if Microsoft stopped supporting .PST files 10 years from now? How would you retrieve those records from the archives? Storing messages for too long and storing them in an unsupported -- or soon-to-be unsupported -- format can cause several issues down the road.

How long should I retain messages?

When determining how long to retain company messages, no clear-cut recommendations exist. There are various legislative regulations that include email archiving requirements. For instance, all publicly traded companies are subject to the Sarbanes-Oxley (SOX) Act and the Gramm-Leach-Bliley (GLB) Act, both of which define email retention requirements, among other things related to data storage and security.

Even so, there are other more restrictive regulations that apply to specific industries, including financial services, healthcare and government. The entire healthcare industry is subject to the Health Insurance Portability and Accountability Act (HIPPA).

Financial services companies may be subject to Securities and Exchange Commission (SEC) regulations and regulations related to the National Association of Securities Dealers (NASD) or the New York Stock Exchange (NYSE). The General Records Schedules from the National Archives and Records Administration mandate archival requirements for government agencies.

The key to determining how long to retain a company's archives is to understand which regulations apply to your particular industry and which retention rules apply to that regulation. Smaller, privately owned companies may not be required to retain message archives at all. In any case, you also need to determine what your business needs are and balance that with any applicable regulations.

For example, if you owned a Web-based store that had a 60-day return policy on items, it would be good practice to retain email messages for at least 60 days after the sale -- even if it's not required. In other cases, you need to retain email messages for three years for IRS purposes as well.

Storage considerations

In addition to legal ramifications, you must consider backup and storage in your long-term message archival plan. Most archiving products on the market use compression

and single-instance storage to minimize the amount of disk space that the archives consume. Even so, long-term storage of messaging data will consume disk space at an ever increasing rate if your organization is storing messages indefinitely.

Generally, message archives are not stored on an Exchange server; therefore, they aren't typically included in the normal Exchange Server backup process. Although some administrators tend to think of email archives as a type of backup, it's also imperative to regularly back up your message archives.

For example, it would be difficult to explain to the courts that your company doesn't have the archived material that was required by law because a hard drive on your archive server failed and you didn't make a backup.

Some organizations avoid the long-term storage issue by outsourcing their archives. This means that the data is stored off-site, and the archival company it hires deals with the headaches of long-term storage and all necessary backups. If your organization decides to outsource storage of its message archives, be sure to read the service provider's contract carefully.

Be certain that you're protected against data loss and service interruptions and make sure that you retain possession of your data. Some unscrupulous archival companies try to retain customers indefinitely by claiming ownership of their data. If a customer tries to cancel an account, the archival company threatens to delete the data. Although most archival companies don't operate like that, read the fine print in the service contract to be sure your company is protected from such practices.

The risks of long-term message retention

Many organizations are required by federal law to retain all email messages for a specific length of time. Although you can store messages for longer than is required by law, there are significant legal risks in doing so. When the government requires you to archive old messages, it's for the government's benefit, not yours.

The underlying assumption is that if a company's business practices are ever called into question, the courts can subpoena message archives and search those archives for incriminating messages. Storing messages for longer than is required means there's more potential evidence. Remember: A message isn't exempt from being used as evidence just because it has exceeded the required retention period.

Sometimes there are legitimate business requirements for retaining an email message longer than is required. It's important, however, to strike a balance between the business's needs and the legal risks associated with long-term message retention.

Free resources for technology professionals

TechTarget publishes targeted technology media that address your need for information and resources for researching products, developing strategy and making cost-effective purchase decisions. Our network of technology-specific Web sites gives you access to industry experts, independent content and analysis and the Web's largest library of vendor-provided white papers, webcasts, podcasts, videos, virtual trade shows, research reports and more —drawing on the rich R&D resources of technology providers to address market trends, challenges and solutions. Our live events and virtual seminars give you access to vendor neutral, expert commentary and advice on the issues and challenges you face daily. Our social community IT Knowledge Exchange allows you to share real world information in real time with peers and experts.

What makes TechTarget unique?

TechTarget is squarely focused on the enterprise IT space. Our team of editors and network of industry experts provide the richest, most relevant content to IT professionals and management. We leverage the immediacy of the Web, the networking and face-to-face opportunities of events and virtual events, and the ability to interact with peers—all to create compelling and actionable information for enterprise IT professionals across all industries and markets.

Related TechTarget Websites

- SearchDataBackup
- SearchCloudStorage
- SearchDisasterRecovery
- SearchVirtualStorage
- SearchSMBStorage
- SearchStorage