Data Archiving Best Practices

Many organizations today ask if they should invest in a data archiving solution—these organizations should keep multiple definitions of “archiving” in mind. Archiving doesn’t always mean moving data entirely offline. Sometimes it simply means moving data to an online, but lower-cost media.

Archiving is largely about control—knowing precisely what data you have, where you have it and having it at the right place, at the right time. Here are five tips to help you trim down your essentials and archive your rarely used content.

1. Leverage Storage Resource Management Tools

Improve budgeting and long-term planning

When you're designing and implementing a data archiving solution, a good first step is to establish a data management baseline of your current environment. By identifying and classifying the various data pools within your organization, you can better understand your archiving requirements. Storage resource management (SRM) tools present the best approach for capturing this baseline. SRM tools can greatly improve budgeting and long-term planning, which in turn can offer significant ROI.

By implementing and using a storage resource management (SRM) solution, you’ll have a more accurate perspective of your storage infrastructure. SRM gives you accurate proactive capacity planning and more accurate growth forecasting. And with the storage performance trending and near real-time statistics, you have much better insight into your day-to-day data usage patterns. This provides early warning signs to potential performance and capacity-related problems and downtime.

2. Implement an Email Archiving Strategy

Save on backup costs and avoid legal trouble

With an effective email archiving strategy, companies can often save as much as 50% on backup costs. The average employee sends and receives more than 140 emails each day, according to a recent Osterman Research report¹, and according to courts and regulators, emails are official business documents. Motivated in part by the Federal Rules of Civil Procedure, as well as case law emerging around email discovery, companies now have new and increased responsibilities to control email in response to potential litigation.

In order to address this issue, organizations often impose strict mailbox quotas, limiting the amount of email any single user can save. Unfortunately, limiting email storage only encourages users to store email messages on their own desktop or laptop in PST files. This strategy presents complications if companies need to provide litigation holds on this data. Additionally, storing email messages on the corporate server negatively affects its performance over time, so it takes longer to recover in the event of a disaster, and is more likely to need additional capacity and other upgrades to accommodate the increased load.

Top 5 Archiving Best Practices:

1. Leverage storage resource management tools
2. Implement an email archiving strategy
3. Implement information lifecycle management
4. Ensure data has the proper privacy controls
5. Establish a data retention policy

3 Implement Information Lifecycle Management

Prioritize and store data according to its value

When you implement information lifecycle management (ILM), you effectively store your information in a manner consistent with the value of the data. All data maintained on storage networks has a defined lifecycle. This lifecycle identifies the way information travels through an organization from its creation to its archival and removal. The exact steps in data lifecycles largely depend on organizational policy, though data generally travels through three stages:

Stage 1: Creation/Acquisition of Data. During the creation of data, both data availability and data value are extremely high.

Stage 2: Publication. The value and availability requirements of published data, whether printed or accessed through other means, often depend on the content of that data.

Stage 3: Retention and Data Disposal. The length of time an organization archives and retains information depends on the nature of the data. However, increasing federal regulations, standards, and compliance measures often govern how long organizations must keep certain types of data.

The changing importance of data and the requirement for data availability create problems—it is costly to store all information on expensive, high-availability Tier 1 storage systems. At some point, organizations must shift at least portions of corporate data to less expensive storage media. However, that merely raises the questions, “What data do we move to cheaper media? And when do we move it?”

ILM provides a strategy for data management throughout the information lifecycle. It identifies the processes and technologies that determine how data flows through an environment. Information path management is another consideration with ILM—organizations are unlikely to offload all rarely used data to cheaper storage unless they can still access the data reliably, if needed.

4 Ensure Data Has the Proper Privacy Controls

Protect your data and your privacy

Data protection and privacy continue to be a tremendous focus and risk for IT communities. While companies make great strides protecting data privacy in production application environments, they often overlook implementing similar strategies in non-production environments such as testing, development, and training.

Reliable, safe, and effective mechanisms for securing data at rest require the adoption and rigorous execution of well-defined processes for handling keys used to encrypt data and keys used to safeguard the data encryption keys. Key management is a comprehensive term that covers these controls—including the creation, distribution, deployment storage, transmission, and destruction of keys used to encipher data.

5 Establish a Data Retention Policy

Categorize archived data and set a retention period for each category

Data retention policies are a critical component of data archiving. Once you have a data retention policy, enforce it for all the information you have on your network. You should also record the retention periods, both for distribution to users and as part of the legal
defensibility record. The retention schedule does not need to be elaborate, but it should include specifics about the various categories and associated retention periods. When creating a written schedule, organizations should focus on two main criteria: the maximum retention periods and data distribution.

As part of a data retention policy, you should set minimum retention periods and maximum retention periods, in order to avoid over-retention. You can either include specific maximum retention periods for each category, or you can include a general clause that when the minimum retention period expires, the company no longer retains the data. This maximum retention period should also address the disposal of any currently retained data, including information archived under an interim infinite-retention period.

Once you establish your archival retention periods, distribute the information to all users, both to communicate policy and to address the storage habits of individual users. In cases where individual employees use a variety of different email programs or in cases where it is common for users to archive e-mail messages, the retention schedule should make clear that this is not “retention” for purposes of the organization.

Organizations can meet legal preservation obligations with interim archiving, regardless of whether a complete retention policy is in place. Litigation requires that a company be immediately able to preserve and retain information (also known as instituting a legal hold) as soon as it reasonably anticipates litigation or governmental action.

Simply moving information into an archive does not constitute effective information management. In order to gain the highest benefit from archiving, companies should categorize all archived information, setting retention periods for each category. Well-established retention periods with automated, effective, and easy-to-use archiving systems create streamlined, customized, and uncomplicated data preservation systems.

Trim down your essentials

These data archiving tips can help you gain the control you need to effectively trim down your essentials and archive your rarely used content. Once you accomplish that, you can help your company trim costs by reserving the more expensive storage for more frequently accessed data.

About the Author

Mark Teter is the Chief Technology Officer at Advanced Systems Group and an internationally recognized authority on information technology who regularly advises IT organizations, vendors, and government agencies on a broad range of information management issues. Each year, Mark conducts dozens of seminars and training programs for corporate and government institutions. He has recently published Paradigm Shift: Seven Keys of Highly Successful Linux and Open Source Adoptions.

About Advanced Systems Group

Advanced Systems Group is a Denver-based IT consulting, integration, and project management firm. As a consistent member of the CRN Tech Elite 250, ASG pursues active involvement in the industry, maintaining the highest level of engineering certifications with partners and the vendor community.