White Paper

Optimizing Data Backup with Intelligent File Virtualization

Despite the advent of snapshot and other disk-based backup technologies, tape backup remains an integral component of data protection for enterprises. Intelligent file virtualization integrates with existing backup solutions to optimize data protection in today's enterprise.

by Renny Shen
Product Marketing Manager

Contents

Introduction

Rapid data growth is challenging enterprises' ability to maintain adequate data protection. Not only is the amount of managed data growing, but file storage environments are becoming increasingly complex as file systems proliferate and increase in size. These factors increase the amount of time required to back up data. For many organizations, backup times now exceed their available window, which results in data protection requirements negatively affecting business operations.

Organizations must maintain user access to data; however this limits their flexibility to manage the influx of data. Moving data or changing file systems can disrupt user access and cause downtime. To address the challenges around data backup, organizations must be able to separate how file data is accessed from how it is backed up.

F5® ARX® intelligent file virtualization helps organizations optimize file-based storage. ARX devices provide data mobility in the file storage infrastructure by enabling files to be moved without disrupting user or application access. ARX provides automated data management policies that intelligently move or place files in the optimal location to achieve business and IT goals.

ARX integrates with organizations' existing backup technologies and processes to provide key benefits in the backup environment:

- **Reduced backup times.** ARX can dramatically reduce the amount of data being redundantly backed up on a frequent basis, thereby reducing the amount of time needed to perform weekly full backups. In addition, ARX allows organizations to back up their largest file systems as multiple smaller ones, each requiring significantly less time to back up than a large system as a whole.
- **Reduced costs.** Reducing the amount of data backed up on a weekly basis also reduces the amount of tape media required per backup. This savings is further multiplied by the number of backup copies an organization retains.
- **Improved flexibility.** ARX provides the flexibility to apply different backup policies to different types of data. For example, ARX can automatically place all PST files on a separate physical file system, which can then be backed up more or less frequently according to business requirements.

The Foundation: Data Backup

Data backup is a fundamental component of an enterprise data protection strategy, the primary purpose of which is to ensure the availability of data in the event of loss. Organizations can accomplish this by creating and retaining backup copies with which IT can restore an organization's data.

How It Works

A backup is a copy of a file system as it existed at a single point in time. This copy includes everything in the file system that was written to disk at the time that it was created, including the directory tree structure, files, and file properties. Transactions that were not written to disk, such as any data cached on the client, are not included in the backup. There are three primary methods used to back up a NAS (network-attached storage) device: mount-based, NDMP (Network Data Management Protocol), and snapshots.

Mount-based backup

Mount-based backup (also called "network backup") involves a backup server mapping to a Common Internet File System (CIFS) share or mounting a Network File System (NFS) export. Administrators perform the backup by copying all the files in the share or export over the LAN through the backup server to an attached tape device.

When restoring, administrators must locate the data using catalogs maintained by the backup server and subsequently copy it to the target destination. Recovered data is often restored to a staging area so that it can be verified before being copied back into the live file system.

NDMP backup

NDMP enables centralized management and control of backup and recovery across heterogeneous file servers. NDMP allows a backup application running an NDMP-compliant implementation to back up different servers from multiple vendors.

In a typical configuration, a backup server will send backup and restore directions using NDMP to a NAS device. The NDMP agent on the NAS device may back up or restore data in one of several configurations: locally to a direct attached tape device; over the network to another NAS system with a locally attached tape device; or via

the network to a backup server with a locally attached tape device. The backup and restore workflow is the same for each configuration.

Snapshot-based backup

A snapshot is a copy of a file system as it existed at a single point in time. As opposed to mount-based and NDMP backup solutions, snapshots are implemented by the file storage device and are stored on disk rather than tape media. Snapshots can be created instantaneously by the file storage device and store only the changes that have occurred since the previous snapshot.

Restoring from a snapshot is relatively straightforward. The file storage device can present the snapshot image as a file system that an administrator or user can navigate to find and restore individual files. Compared to tape-based backups, snapshots are typically used more for short-term data protection due to the higher cost of the disk capacity consumed by the snapshot copies.

Backup Policy

Because data is constantly changing, many of the files preserved in a backup copy quickly become outdated. To provide data protection over time, organizations create a backup policy, which designates backup intervals and defines the periods of time it must retain those backup copies.

A typical backup policy entails performing a combination of weekly, monthly, quarterly, and yearly full backups. A multi-layered policy ensures the availability of not only the latest version of data, but also copies of data as it existed at previous points in time. Each backup copy is retained for a period of time relative to the interval at which it was taken, often using an N+1 system. Figure 1 is an example of a typical backup policy, where weekly backups are retained for five weeks (four weeks per month), monthly backups for four months (three months per quarter), and quarterly backups for five quarters (four quarters per year).

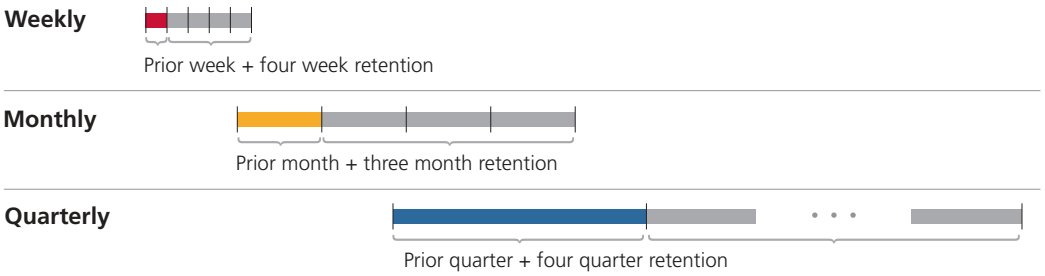

Figure 1: Example of a typical enterprise backup policy

While this backup policy allows an organization to cost-effectively retain copies of their data over long periods of time, it places a higher value on recent data changes by providing more granular recovery point objectives (RPOs). As the backup interval increases from weekly to quarterly, the available RPO becomes less granular. Organizations make this tradeoff to control the costs of backup media. For example, retaining a quarter's worth of tape-based backup copies performed at a weekly interval would require twelve copies, compared to a single copy for a quarterly interval. Similarly, snapshots are often performed at frequent intervals, such as on a nightly basis, but only retained for a short period of time to minimize the consumption of higher-cost disk capacity.

Backup Times and Windows

Understanding backup times

The amount of time required to perform a full backup depends on a number of factors:

- **Performance.** Any backup infrastructure consists of a number of discrete components, such as the backup software, the server on which the backup software is installed, the storage device on which the live data resides, the network, and the tape library. Each of these components will have performance constraints that may affect the speed of the backup.
- **Amount of data.** A full backup essentially reads all of the data in the source file system and writes it to tape. As managed data continues to grow, full backups will require an increasing amount of time to complete.
- **Number of files.** The number of files in a file system can also affect backup times, as the backup server must read each file before writing it to tape. Even a small amount of data can take a long time to back up if it is stored in a large number of smaller files.

- **File system size.** Because backups are performed against a file system, the size of the file system directly affects the amount of time required to perform the backup. Larger file systems can hold more data and more files that must be read and written to tape media, and therefore are more time-consuming.

What is a backup window?

Unlike backup times, the backup window is a business constraint imposed on the backup process by the needs of the business. When performing a backup, backup software can impose a significant performance load on file systems and file storage devices. This can adversely affect users and applications when accessing files for business purposes. As a result, organizations typically restrict backup operations to specific windows. These are negotiated with the organization's business groups to occur during periods of lower activity, such as over the weekend.

Because of recent growth in both the amount of data and the size of modern file systems, backup operations are often exceeding the traditional weekend backup window. This requires organizations to make a choice: cancel the backup or adversely affect users and applications.

Optimizing Backups with ARX

ARX helps organizations optimize their file storage environments for a variety of purposes, including to reduce backup times and costs. ARX provides two essential capabilities. First, it virtualizes the existing file storage environment with a presentation layer. Often referred to as a Global Namespace, this layer decouples logical access to files from their physical location on storage. This allows individual files to be moved without disrupting application or user access. Next, ARX devices can classify data based on characteristics such as age, type, and size; then they can perform various actions based on these classifications through a suite of intelligent data management policies to achieve a variety of business and IT goals.

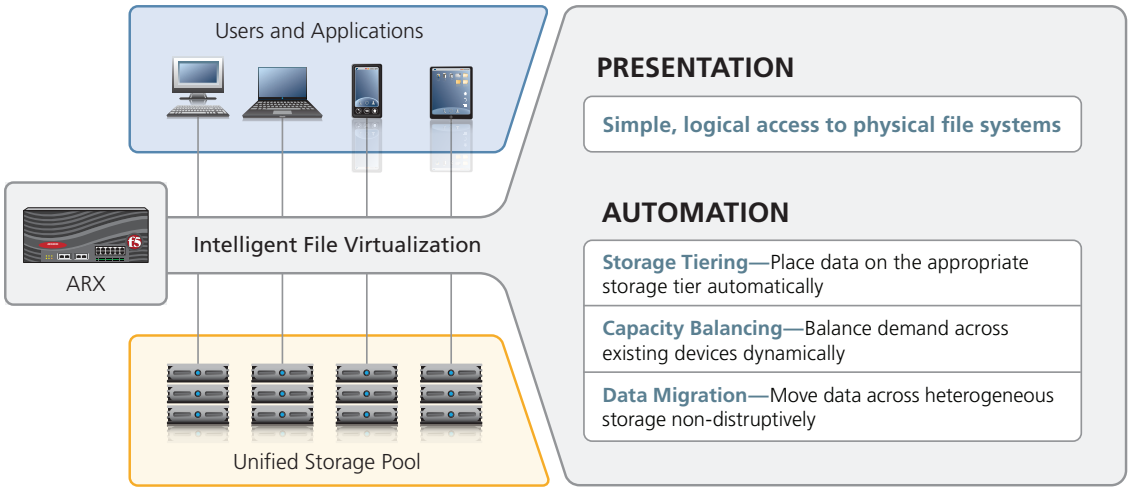

Figure 2: ARX intelligent file virtualization

Designing a Global Namespace

File virtualization offers a unique opportunity to address a key backup challenge. Using ARX, organizations can construct a Global Namespace that decouples the file serving structure between client systems and physical storage devices. As shown in Figure 3, the Global Namespace is a virtual construct created by an ARX device. ARX presents a collection of virtual file systems (i.e., CIFS shares or NFS exports) to users and applications. The virtual file systems are not tied to specific hardware devices, but instead can federate capacity from physical file systems presented by multiple storage devices. The ARX device then proxies any file access from a client system to the proper physical file system where the file is actually stored.

The design of the Global Namespace can have a significant effect on backup times in that it reduces the size of the physical file systems being backed up. In Figure 3, ARX is presenting a 16 TB virtual file system to users and applications. The virtual file system comprises and federates the capacity of 16 physical file systems, each 1 TB in size. As clients create new files and store them in the virtual file system, ARX automatically distributes the new files evenly across all of the physical file systems.

In a virtualized file storage environment, administrators can perform full backups at either the virtual or physical layer. With the former, the backup server can be pointed at the virtual file system presented by the ARX device. As with user and application clients, the backup server will only be aware of the single virtual file system. Note that a virtual backup does not provide any reduction in backup times as each backup image must still traverse and capture all of the data within a 16 TB

file system. And, depending on the size of the virtual file system required, some organizations may not have sufficient time within their backup window to actually perform the backup.

Performing full backups at the physical layer can provide significant benefits. As shown in Figure 3, the backup server can also be pointed individually at each of the smaller physical file systems. While this does entail performing a greater number of backup operations, each operation will require significantly less time to back up. In addition, multiple file systems can be backed up in parallel, further reducing the amount of time required. This can allow organizations to back up even their largest file systems, supporting application requirements while maintaining the required level of data protection. By decoupling the file serving structure between client systems and physical storage devices, organizations can design each independently to best meet their application and backup requirements.

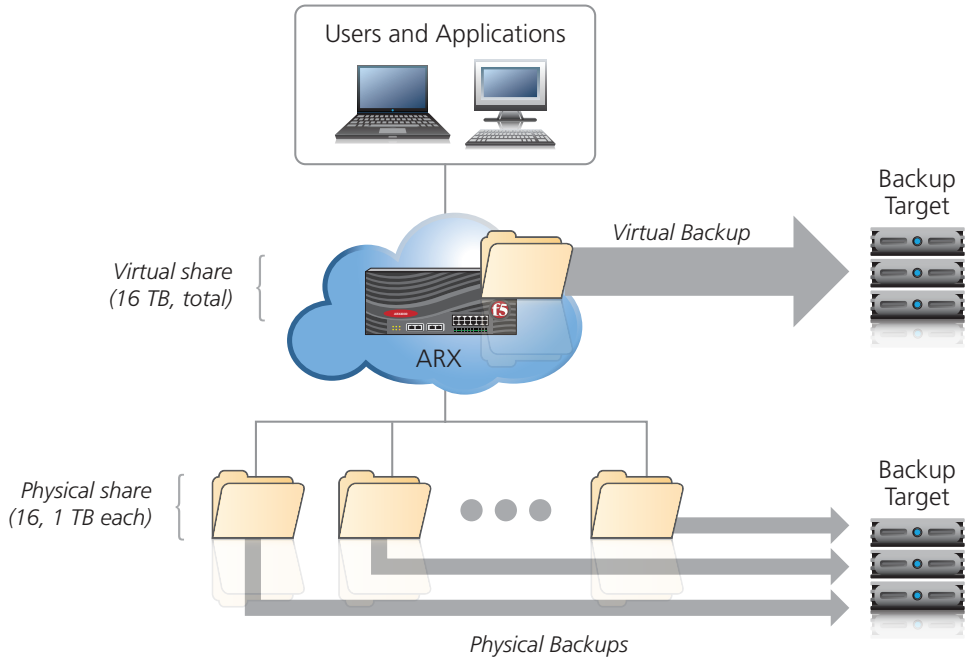

Figure 3: Benefits of physical versus virtual backups with ARX

Note that ARX does not affect the process for incremental backups—organizations can continue making nightly incremental backups.

Leveraging Data Management Policies

In addition to the Global Namespace, ARX provides a suite of data management policies that can intelligently move or place files to achieve business and IT goals. ARX policies use various file attributes, such as file age (last modified, last accessed), type, and size, to determine the optimal location to store those files. For backup purposes, two characteristics are particularly useful: age and type.

Optimizing backup by file age

As the total amount of file data under management continues to grow, the majority of that data is inactive or has not changed in some time. Organizations that perform frequent full backups may find that much of the time and backup media required is spent backing up files that have not changed from previous backups. The operational challenge is that active and inactive data cohabitate many of their file systems, making it difficult to distinguish between the two, or to separate them.

Using ARX, organizations can create automated storage tiering policies that automatically migrate files from primary to secondary storage based on their age. As shown in Figure 4, this configuration builds on the Global Namespace to physically separate active from inactive data in different physical file systems. Users and applications access and store all of their files through the virtual file system. When a file is first created, ARX places it on Tier 1. As a file ages, ARX automatically migrates it to Tier 2 according to a defined policy.

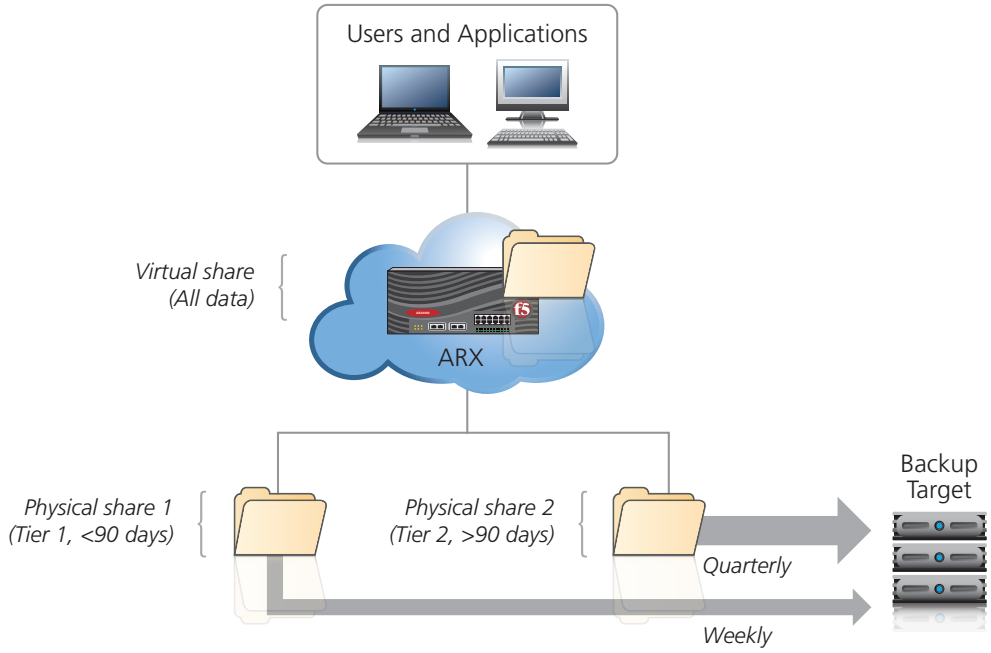

Figure 4: Reduced weekly backup times and costs with age-based storage tiering policies

There are several things to consider when creating an age-based tiering policy with ARX for backup optimization purposes:

- **Schedule.** ARX devices provide both real-time and scheduled data management policies. Storage tiering policies involving file age use scheduled policies to minimize the amount of file movement between tiers. Scheduled policies are ideal for backup optimization because all file movement occurs systematically at known points in time. Administrators can configure the policy schedule to align with their existing backup schedule (weekly, monthly, quarterly).
- **Criteria.** ARX can use both the last modified and last accessed dates as criteria for its storage tiering policies. Using the last modified date is preferred for backup purposes, as the goal is to reduce the amount of redundant backups of unchanged data.
- **Number of tiers.** Although Figure 4 only shows a two-tiered environment, organizations can use ARX to create up to 64 tiers to meet their storage and backup requirements.
- **Number of file systems per tier.** Figure 4 depicts a single physical file system for Tier 1 and Tier 2, respectively; but ARX can use multiple physical file systems for each tier and will automatically distribute files across each file

system in the appropriate tier. This further reduces the amount of time required to back up each tier.

Figure 4 provides a good example of how ARX automated storage tiering policies can align with existing backup processes to optimize full backups. Here, ARX has been configured with a storage tiering policy that automatically migrates files that have not been modified in the last 90 days to Tier 2. The storage tiering policy is configured to run every three months. With the ARX storage tiering policy, every file on Tier 2 is known to not have changed in the last 90 days. Therefore, this organization can perform full backups of their Tier 2 data on a monthly or quarterly basis. ARX dramatically reduces the amount of data in the weekly backup data set, correspondingly reducing the amount of time required for performing weekly full backups and the amount of backup media consumed.

Inline recall is the final backup consideration when creating an age-based tiering policy. When a file that has been moved to Tier 2 is modified, ARX can immediately recall the file and place it on Tier 1. This allows all recent modifications to be captured by the weekly full and nightly incremental backups, regardless of the previous location of the file.

Optimizing backups by file type

Many organizations have special backup requirements for different types of files—they may prefer to back up specific data sets either more or less often. A common example is with Personal Storage Table (PST) files. PST files are created by users as an offline archive for their Microsoft Exchange data. In many cases, these files are scattered across an organization's user home directories, making it difficult to apply a different backup policy specific to these files.

Using ARX, organizations can create storage tiering policies that automatically place newly created files on specific physical file systems based on their type. As with an age-based tiering policy, users and applications continue to access and store all of their files through the virtual file system. However, a type-based tiering policy is not a scheduled policy, but rather a real-time policy. With a type-based policy, ARX ensures that specified files are placed in the designated physical file system at the time of creation.

In Figure 5, users access and store PST files through a virtual home directory file system presented by an ARX device. The ARX device has been configured with a storage tiering policy that automatically places PST files separate from other files on file system 2. As with age-based storage tiering, the organization can back up this

file system independently from others, including applying a different backup schedule appropriate to the business value of the contents.

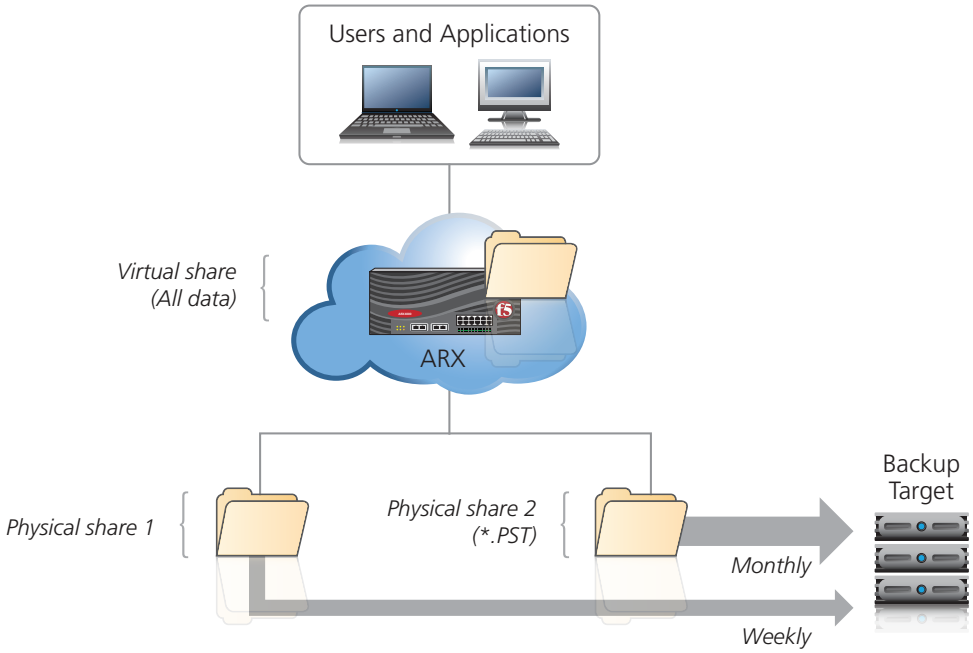

Figure 5: Customizing backup policies with type-based storage tiering policies

Virtualizing snapshots

ARX can simplify the management of snapshots in heterogeneous storage environments. Rather than administrators having to individually manage snapshots for every file system, ARX can coordinate the physical snapshot creation operation and aggregate them into a virtual snapshot presentation. The virtual snapshot can include physical snapshots from multiple file systems, storage devices, or different types of devices. In addition, it will have the same presentation as the Global Namespace, simplifying the file restoration process. Administrators and users can navigate the virtual snapshot in the same manner as they do live file systems, and easily identify files they need to restore.

Advantages of the F5 Approach

Reduced Backup Times and Costs

ARX can significantly reduce the amount of time required to perform full backups. With the Global Namespace, organizations can design large virtual file systems that are composed of multiple smaller physical file systems. Users and applications access files through the virtual file system, while backup servers can target the smaller physical file systems for shorter backup times. Organizations should use a physical file system size that is appropriate for the length of their backup window.

In addition, the majority of most organizations' data is inactive and has not been modified for some time. Figure 6 shows part of a sample report generated by F5 Data Manager™. Data Manager is a storage resource management and reporting tool that helps organizations better understand their file data.

In this example, almost 75 percent of every weekly full backup includes data that has not changed from the week or month before. ARX includes automated storage tiering policies that physically separate active and inactive data into separate file systems, without affecting user and application access to those files. This reduces the size of the backup data set and significantly reduces the amount of time required to perform weekly full backups.

Beyond reducing the amount of time required, ARX can also reduce costs associated with backup media. As shown in Figure 6, an appropriate storage tiering policy can reduce the size of the backup data set by 75 percent. With a typical retention policy of five copies, this can dramatically reduce the amount of tape capacity required with no degradation in data protection.

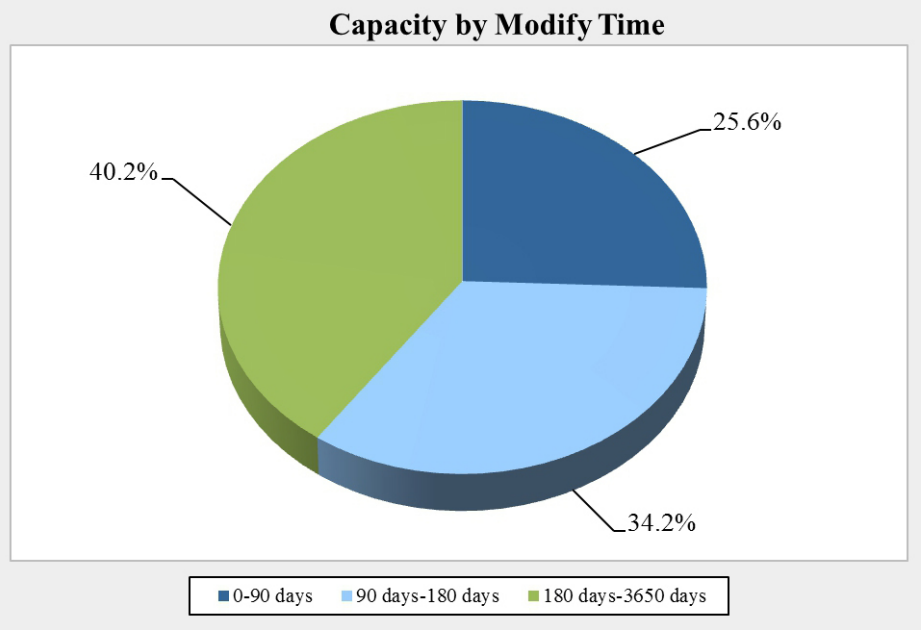

Figure 6: Sample characterization of file data by last modified time

Other Advantages

Improved backup flexibility

With a Global Namespace and intelligent data management policies, ARX offers organizations significant flexibility in how full backups are performed. Organizations can independently design the virtual environment to improve user and application access to file data and the physical environment to improve data protection and reduce costs. By enabling the physical separation of different types of data, ARX can help organizations optimize their backup processes in a variety of ways that make sense to their business, using file characteristics such as age and type.

Integration with existing processes

A key advantage of ARX is its seamless integration with existing backup operations and processes. Because the ARX employs a proxy architecture to virtualize the environment, all data resides on physical file systems presented by a file storage device, similar to unvirtualized environments. This means that backups can continue to operate in the same manner as before, interfacing with physical file systems or devices using existing processes.

Simplified snapshot management

Virtual snapshots minimize the complexity of managing the snapshot process across multiple and potentially heterogeneous file servers. The ARX device coordinates snapshots from all of the physical shares within each virtual share to reduce the administrative overhead. In addition, it abstracts the interfaces to proprietary snapshot implementations, allowing administrators to apply a consistent snapshot policy across the virtualized environment. Finally, ARX presents the virtual snapshot image to users and administrators, simplifying the process of restoring individual files.

Conclusion

Ensuring data protection in the face of rapid data growth is a top priority for every IT organization. But this challenge is made even more difficult when backup times exceed the maximum available backup window. As a result, backup processes start to impede business operations with users and applications experiencing poor performance. ARX provides a flexible approach that helps organizations achieve their backup objectives while decreasing the burden on IT resources.

ARX devices use virtualization technology to give organizations the flexibility to optimize their physical file storage environment for backup purposes, without affecting users and application access. Using a Global Namespace and intelligent data management policies, ARX allows large file systems to be backed up in smaller increments and reduces the amount of data that needs to be backed up on a weekly basis. This leads to dramatically shorter backup times and reduced backup media consumption so organizations can ensure that their data remains adequately protected.

F5 Networks, Inc. 401 Elliott Avenue West, Seattle, WA 98119 888-882-4447 www.f5.com

F5 Networks, Inc.
Corporate Headquarters
info@f5.com

F5 Networks
Asia-Pacific
apacinfo@f5.com

F5 Networks Ltd.
Europe/Middle-East/Africa
emeainfo@f5.com

F5 Networks
Japan K.K.
f5j-info@f5.com