



Enabling Flexibility with Intelligent File Virtualization

Enabling Flexibility with Intelligent File Virtualization

White Paper



WHITE PAPER

Enabling Flexibility with Intelligent File Virtualization

Introduction

Enterprises today face the challenge of managing and storing an ever-increasing amount of data. Not only is data being created at an unprecedented rate, but regulatory and business needs are requiring data to be preserved for longer periods. At the same time, IT organizations must do more with less-capital and operational budgets have not kept pace with the rapid growth in storage capacity requirements.

Virtualization provides an elegant solution to efficiently managing dynamic and growing file storage environments. A common characteristic of these environments is the unpredictability of where and how fast data will grow. Much of the management overhead can be attributed to continually optimizing file storage resources to accommodate the changing nature of the data they contain. Optimizing for higher efficiency or lower costs often requires moving data, a task that is operationally disruptive for users and applications.

The F5 ARX intelligent file virtualization solution introduces the ability to move file data without disrupting user or application access. Through file virtualization, ARX preserves the logical access to files regardless of their current location on storage. From an operational perspective, this reduces the cost of moving data by reducing the time and IT overhead required as well as minimizing downtime. From a strategic perspective, data mobility provides flexibility in the storage infrastructure, which is necessary for organizations to respond to data growth.

ARX builds on that flexibility with intelligent data management policies that allow IT to rapidly respond to developing changes in the environment. These customizable policies automatically place or move individual files to optimize file storage environments to meet business goals, such as greater efficiency, lower storage costs, and shortened backup times.

Managing Change: The Challenge

Constraints in Responding to Constant Change

The high rate of data growth is generating constant change in today's file storage environments. However, two constraints inherent in current file storage infrastructures hamper the ability of IT organizations to manage that change effectively:

- **Complexity.** Today's typical enterprise storage infrastructure consists of a complex collection of storage platforms, file systems, and operating systems. As data continues to grow, organizations add new storage devices, potentially from different vendors. Unfortunately, these devices often do not work



WHITE PAPER

Enabling Flexibility with Intelligent File Virtualization

seamlessly together. This is especially true as organizations begin to embrace cloud storage, which typically requires standalone gateway devices to interface with the cloud API.

- **Inflexibility.** Users and applications are statically mapped to the physical file storage resources that contain the data they need to access, with many environments containing hundreds or thousands of individual mappings (or mount points). As file systems grow, files move, or devices change, these static mappings are disrupted. Updating them to account for environmental change requires manual configuration and system downtime.

Consequences of Complexity and Inflexibility

A changing environment creates imbalances between how files should be optimally stored and how they are actually stored. The complexity and inflexibility of the file storage infrastructure then makes it difficult to move files to the best storage location, with the following consequences:

- **Inefficiency.** Unpredictable growth leads to uneven utilization among different file storage resources. While there may be specific file servers or network-attached storage (NAS) devices that are highly utilized, in aggregate, file storage devices tend to be underutilized, averaging just 40 to 50 percent capacity utilization in most enterprise environments. This is typically not the result of underuse, but rather of over-provisioning because there has not been a simple, non-disruptive way to balance demand or add capacity to file systems without incurring an outage.
- **Higher capital costs.** Even as the total amount of data continues to grow, the majority of it is older or inactive. Inactive data remains on high-cost storage long after its business value has diminished, because of the disruption and IT overhead required to identify, classify, and move it to a more appropriate location. As a result, organizations remain locked into a higher cost for storage capacity and find it difficult to take advantage of lower-cost storage options, such as deduplicated storage systems or the cloud.
- **Higher operational costs for ongoing management.** Every new device added to meet capacity demands increases environmental complexity and escalates operational costs. Storage environments become increasingly siloed over time, with each silo needing to be managed as a discrete "island" of storage. In addition, the proliferation of different storage platforms places an extra burden on IT staff, who must be trained to manage each of them.
- **Higher operational costs for migrating data.** In traditional environments, these operations incur an overhead cost beyond the cost of manually moving the data. A typical migration project also involves a significant amount of time for planning prior to the migration, and error-fixing and reconfiguring affecting client systems afterward. In addition, many IT organizations support businesses that operate 24x7 and recognize the cost of downtime.



Flexibility to Respond to Change

The key to breaking free from the constraints outlined above lies in the ability to eliminate the static mapping between client and storage resources, which allows the composition of storage resources to change and data to move freely between resources, without affecting client access to data. Intelligent file virtualization provides a layer of virtualization in the network that decouples the logical access to files from the physical location of those files. With this layer in place, data is free to move and storage resources are free to change, without the disruption previously associated with these actions.

The next step in creating a dynamic storage infrastructure is to introduce intelligence that can respond to ongoing change. The file virtualization layer provides an ideal location for such intelligence because it is aware of any changes to files as they occur. Intelligent data management policies can monitor files as they are created or change over time, and then take appropriate action based on the environmental conditions. For example, a policy can move individual files to lower cost storage as they age. Intelligent policy enforcement is a valuable tool for organizations that can reduce the IT overhead involved in responding to changes in the environment.



Key Benefits

This intelligent file virtualization layer brings several key benefits to the environment:

- **Simplified file access.** Hundreds or even thousands of physical client-to-resource mappings can be consolidated down to a much more manageable number of logical mappings, as shown in Figures 1a and 1b. Even more important, these logical mappings need never change—they are persistent. IT staff can perform storage management tasks such as provisioning, consolidation, and migration without having to reconfigure client systems.
- **Increased operational flexibility.** File virtualization decouples the logical access to files from their physical location on storage. Data is no longer bound to physical storage resources and can be moved at will, without affecting client access to that data or incurring downtime. This gives organizations more flexibility to better respond to data growth or ongoing change in their file storage environment.
- **Increased architectural flexibility.** By decoupling logical file access, virtualization abstracts the physical file storage infrastructure from users and applications. This provides the flexibility to utilize different file storage technologies, platforms, or devices, as well as change vendors to meet business and IT requirements over time. Removing the inherent barriers to infrastructure change helps organizations take advantage of new technologies, such as solid state drives (SSDs), data deduplication, and cloud storage.
- **Ongoing optimization of storage resources.** Intelligent file virtualization automatically optimizes the storage of file data over time based on business goals. Data management policies monitor both existing and newly created files and automatically match them with the appropriate type of storage according to their business value. This is determined using flexible criteria, such as file age, type, and size.

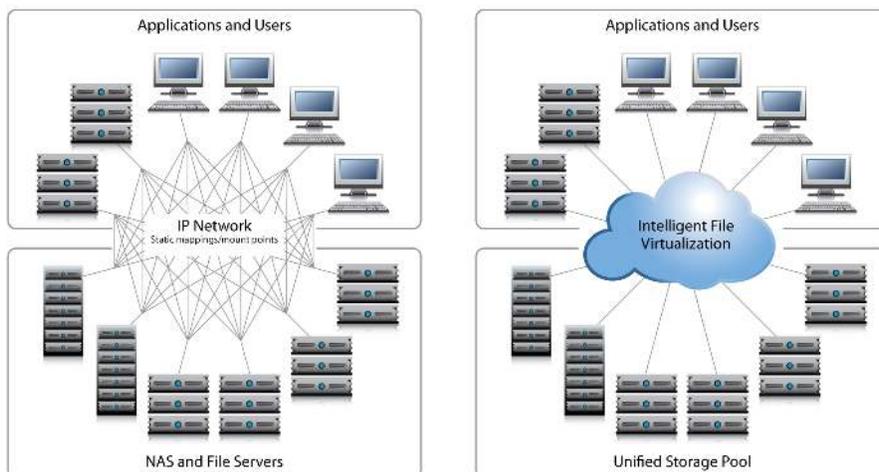


Figure 1a: Current file storage infrastructure, Figure 1b: Intelligent file virtualization



WHITE PAPER

Enabling Flexibility with Intelligent File Virtualization

Attributes of an Intelligent File Virtualization Solution

An intelligent file virtualization solution has several key elements:

- **Heterogeneous and multi-vendor.** The steady introduction of new storage technologies (such as the cloud) highlights the need for intelligent file virtualization to support a multi-vendor, heterogeneous file storage infrastructure. To achieve this, it must be able to present a logical abstraction of all types of physical devices being virtualized, irrespective of file system, platform, vendor, or protocol.
- **Automation and real-time policy.** Intelligent file virtualization maximizes value not only by simplifying file access through presentation, but also by simplifying storage management through automation. To automate storage management operations, a solution must be able to react in real time to dynamic environmental conditions. A policy that only reacts after the condition it sought to avert occurs is of no value. Similarly, a policy that informs about a condition but is powerless to act upon it is of little value. To realize the benefits of automation, file virtualization solutions need to enforce management policies in real time.
- **Performance and scale.** Intelligent file virtualization cannot introduce new bottlenecks into the file storage infrastructure. This means that the intelligent file virtualization layer must be faster than the aggregate of the physical devices it virtualizes. It must also be able to scale to meet future capacity demands, in order to help organizations better manage their rapid data growth.
- **Data integrity and availability.** Intelligent file virtualization cannot introduce any single points of failure into the file storage environment, nor can it compromise data integrity in any way. The intelligent file virtualization layer must meet or exceed the availability characteristics of the most highly available systems it virtualizes.

ARX Intelligent File Virtualization

F5 ARX helps organizations better respond to and manage change in their environments. ARX devices employ a unique network-based architecture to provide flexibility in the file storage infrastructure. They interface directly with the IP/Ethernet network fabric to provide an additional layer of intelligence—a "file awareness"—to the network.

ARX devices use industry-standard file access protocols to communicate with clients and servers—CIFS for Windows devices and NFS for UNIX or Linux devices. The ARX device does not introduce a new file system; rather it acts as a proxy to the file systems that are already there. Enterprises are not required to forklift upgrade hardware, replace existing file systems, or load software agents across the enterprise to gain the benefits of virtualization.



WHITE PAPER

Enabling Flexibility with Intelligent File Virtualization

The scale and performance requirements of an enterprise-level intelligent file virtualization solution in turn require a purpose-built architecture that can scale to billions of files and handle gigabytes of throughput. The ARX architecture is the only file virtualization solution proven to scale to these levels, and it is the reason that ARX devices are deployed in many of the world's leading large enterprises today.

When it comes to availability, ARX devices provide equivalent or better availability than the leading high-end clustered NAS devices in the market today. Services transparently fail over between ARX devices in a cluster upon failure, ensuring data integrity throughout the entire process.

ARX devices are unique in their ability to monitor client demand, resource capacity, and network conditions, and to adapt in real time to respond to these changing dynamics. This enables ARX to perform several unique functions, including dynamic load balancing and placing data on appropriate storage in real time. It also eliminates much of the overhead associated with searching entire file systems to determine policy actions, enabling high-performance, low-latency, real-time policy enforcement in ARX.

There are two major components to the ARX solution: presentation and automation.

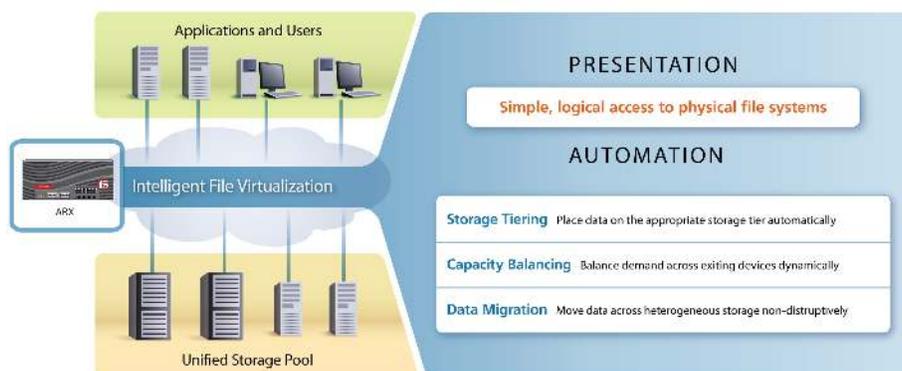


Figure 2: ARX intelligent file virtualization

Presentation: Simplifying File Access

Think of the presentation layer as the client-facing side of intelligent file virtualization. It is the logical abstraction of the physical environment that the client sees. The presentation layer enables simple, logical access to physical file systems and hides storage changes from clients.



How does the presentation layer work?

Every storage device presents a namespace—a collection of shared file systems, such as CIFS shares and NFS exports. Because each namespace is tied to a specific device, storing and accessing files through the namespace is limited to that device. For example, file systems in each namespace can only be provisioned storage capacity from the presenting storage device. This physical relationship is a principal reason for the inflexibility attributed to traditional file storage environments.

The ARX presentation layer federates the individual namespaces of multiple heterogeneous storage devices. Figure 3 shows how a presentation layer works. By creating what is called a Global Namespace, ARX presents a collection of virtual CIFS shares and NFS exports that can comprise capacity from any of the physical file systems behind it. The physical storage devices can be of different types, based on different platforms, or even from different vendors.

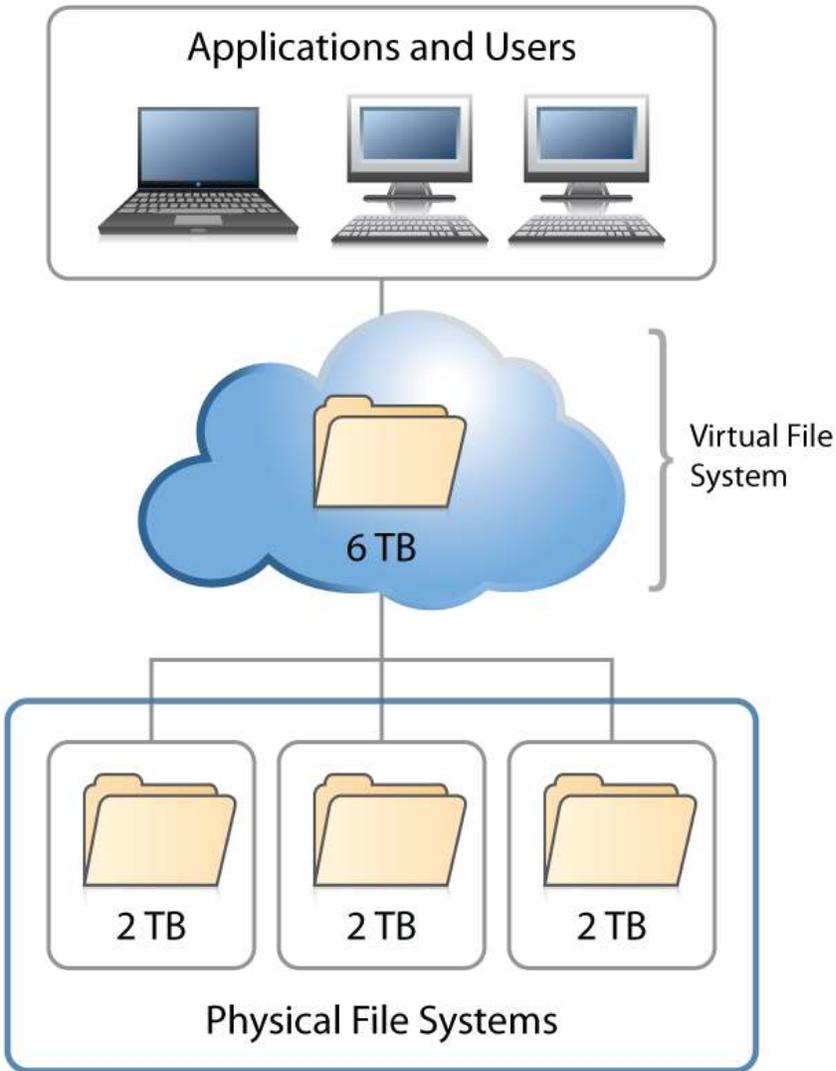


Figure 3: How the presentation layer works

The presentation layer enables administrators to decouple logical access to files from their physical location on storage. Application and user clients logically access file data through the presentation layer; that is to say, they access and store their files through a virtual file system presented by ARX. The files may be located on any of the physical file systems or storage devices behind it. The ARX device keeps track of the current physical location for every file and proxies any file access to that location. If a file is physically moved, the ARX will proxy access to the new location with no effect on application or user client systems.



WHITE PAPER

Enabling Flexibility with Intelligent File Virtualization

Benefits for users

The presentation layer affects storage managers, users, and application managers. From a user or application manager perspective, there are several key benefits to the presentation layer.

- In the past, if files were moved or storage was reconfigured, access was interrupted as storage administrators had to reconfigure login scripts and drive mappings to access the new location. With ARX, these changes become invisible to users and application managers. Consequently, companies are free to leverage the latest technology, or seamlessly expand capacity, without experiencing downtime.
- In the past, users and application managers needed to know where files were located to access them successfully. Often this required complex mappings, as the data could be spread out across many different file systems. ARX dramatically simplifies this situation by allowing clients to view all file storage as one unified pool, making it much easier to find and access data.

Benefits for IT

The presentation layer also brings some significant benefits to the storage manager.

- Storage management tasks such as adding or decommissioning file servers or moving files used to mean outages. By necessity, these outages occurred at inconvenient times—over weekends, at night, and during holidays. However, with ARX, outages due to storage management tasks can be eliminated.
- Previously, many storage management tasks required storage managers to reconfigure file systems and client machines. ARX eliminates this reconfiguration overhead because all drive mappings and mount points are now persistent, and thus don't require reconfiguration when the data location or infrastructure changes.
- Until now, responding to data growth has often been disruptive. Re-provisioning file systems that had run out of capacity required outages that could adversely affect the business. With ARX, new capacity can be easily provisioned into existing virtual file systems, allowing the environment to seamlessly scale with the growth in data.
- In the past, the file storage environment was restricted by the hardware devices it contained. For example, the maximum file system sizes supported by storage devices are often smaller than what is required to support an application or user groups. With ARX, organizations can scale their environment beyond the limitations of individual devices, allowing them to better support their IT requirements.

Automation: Simplifying Storage Management

The automation layer automates storage management tasks that were previously manual. It is invisible to users and application managers, and it holds significant benefits for the storage manager.



WHITE PAPER

Enabling Flexibility with Intelligent File Virtualization

There are several different automation components, including:

- Data migrations move data, potentially across platforms from different vendors, without disrupting access to that data.
- Storage tiering, or automated Information Lifecycle Management (ILM) policy, places data on appropriate classes (tiers) of storage according to business need, both automatically and non-disruptively.
- Capacity balancing dynamically distributes demand across existing devices to optimize application performance.
- Backup optimization reduces the amount of data being redundantly backed up on a frequent basis, and breaks large file systems into smaller ones for faster backups.

Non-disruptive data migration

Data migrations are a common occurrence in enterprises. There are many reasons IT organizations move data on a daily basis, from one-time events such as server consolidations, platform upgrades, and vendor transitions, to everyday occurrences such as capacity balancing or re-provisioning events. Whatever the reason, ARX enables organizations to move data without affecting user access.

In addition, ARX provides powerful policies that simplify a range of data migration tasks—from moving entire file systems to just individual files. Data migration can take place between heterogeneous storage devices, for both CIFS- and NFS-oriented data, and administrators can schedule migrations not to coincide with peak traffic times or backup windows. Furthermore, ARX is uniquely able to handle complex and large-scale data migrations. Because it decouples logical access from physical location, organizations can restructure the file system layout to simplify ongoing storage management tasks without affecting client access.

WHITE PAPER

Enabling Flexibility with Intelligent File Virtualization

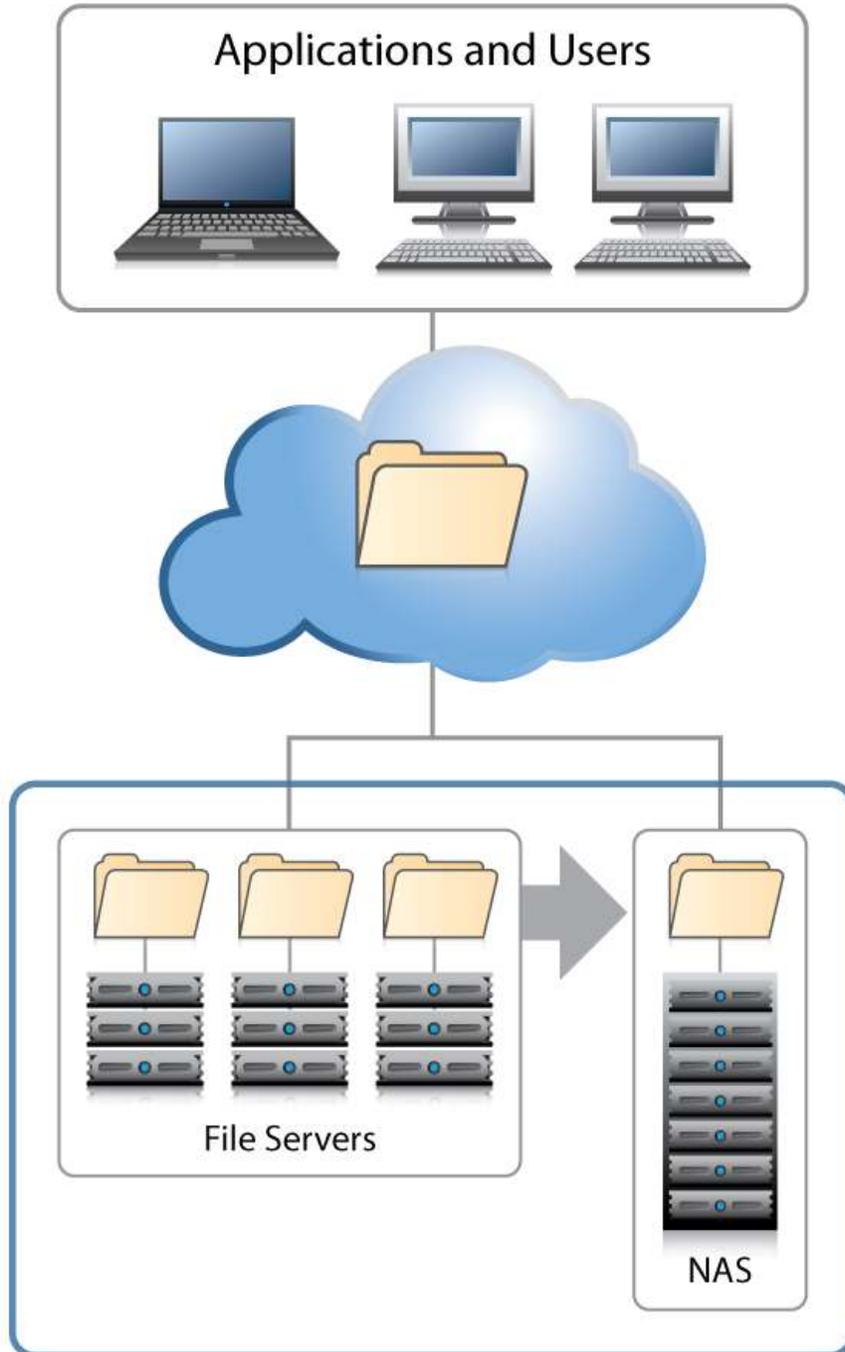


Figure 4: Performing a data migration without disrupting applications and users



WHITE PAPER

Enabling Flexibility with Intelligent File Virtualization

With ARX, the outages and business disruption previously associated with data migrations are a thing of the past. Administrators have less overhead as they no longer have to reconfigure applications or client machines, and much of the risk associated with these operations is mitigated by eliminating operator error through automation.

Automatic storage tiering

Today many enterprises are seeking to reduce their storage costs by implementing a tiered storage strategy. These organizations are interested in augmenting their high-end, high-cost storage with more cost-effective storage technologies like SATA, data deduplication, or the cloud. With a tiered storage strategy, organizations can realize significant capital and operational savings by moving non-critical business data off expensive storage resources to lower-cost alternatives.

ARX automates the placement and movement of data between different tiers (or classes) of storage, with each tier potentially comprising devices from multiple vendors. ARX storage tiering policies operate at the file level because most enterprises need this flexibility—typically organizations want to move files or projects of a certain age or of a certain type, rather than entire file systems. It is important to note that when ARX devices move files between tiers, they do not leave behind stubs or pointers, which pose availability risks and complicate backup and recovery procedures.

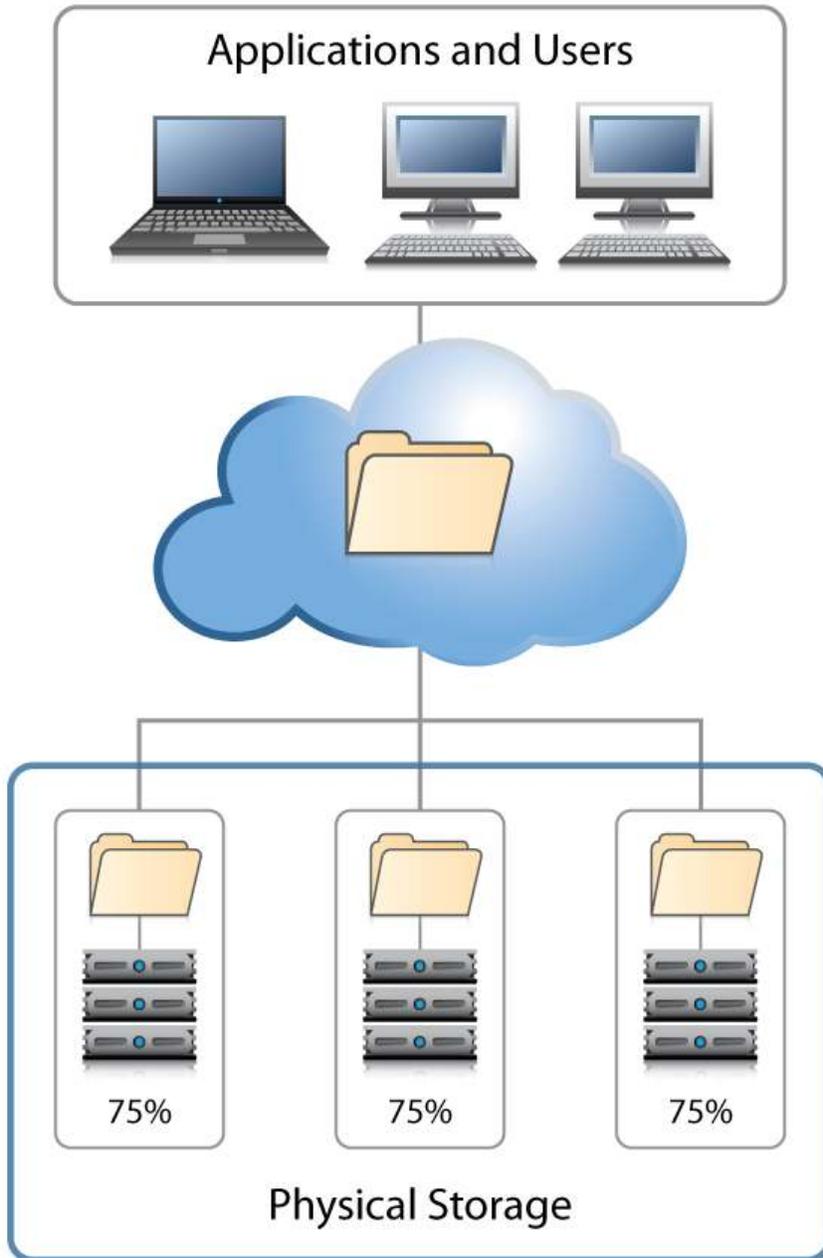


Figure 5: Tiering to the cloud with ARX



WHITE PAPER

Enabling Flexibility with Intelligent File Virtualization

ARX is unique in that it can enforce policies in real time. For example, if a policy on an ARX device dictates that a file of a certain type be placed on a specific storage device, it happens automatically, in real time, without having to be moved later on when an out-of-band policy engine has queried the entire file system. This real-time capability means ARX devices can more efficiently manage the lifecycle of data over time. For example, if a particular file has been moved from tier 1 to tier 2 because it has not changed in a given period, ARX can automatically move this file back to tier 1 if it is changed in the future.

With ARX, enterprises can better utilize their most expensive storage tier and reduce costs by augmenting this tier with lower-cost storage. They can also easily take advantage of new technologies, seamlessly integrate them into existing file storage environments, and maximize the benefits realized.

Dynamic capacity balancing

The rapid overall rate of data growth often obscures another important characteristic—the uneven and unpredictable manner in which it grows. Within an organization, different applications and users will generate new data at different rates and in different quantities. This leads to an environment where some storage resources are heavily utilized, while others are barely utilized at all. As a result, organizations must often purchase additional storage capacity despite having free or stranded capacity elsewhere in their environment.

ARX is unique in its ability to support real-time capacity balancing policies that can increase storage efficiency in heterogeneous storage environments. ARX policies can automatically distribute files across multiple physical file systems, resulting in even utilization of all storage resources. ARX can also help scale the file storage environment along with data growth. When administrators provision new capacity into an existing virtual file system, the ARX device will balance utilization with that of existing file systems to maintain an even level of utilization throughout the environment.

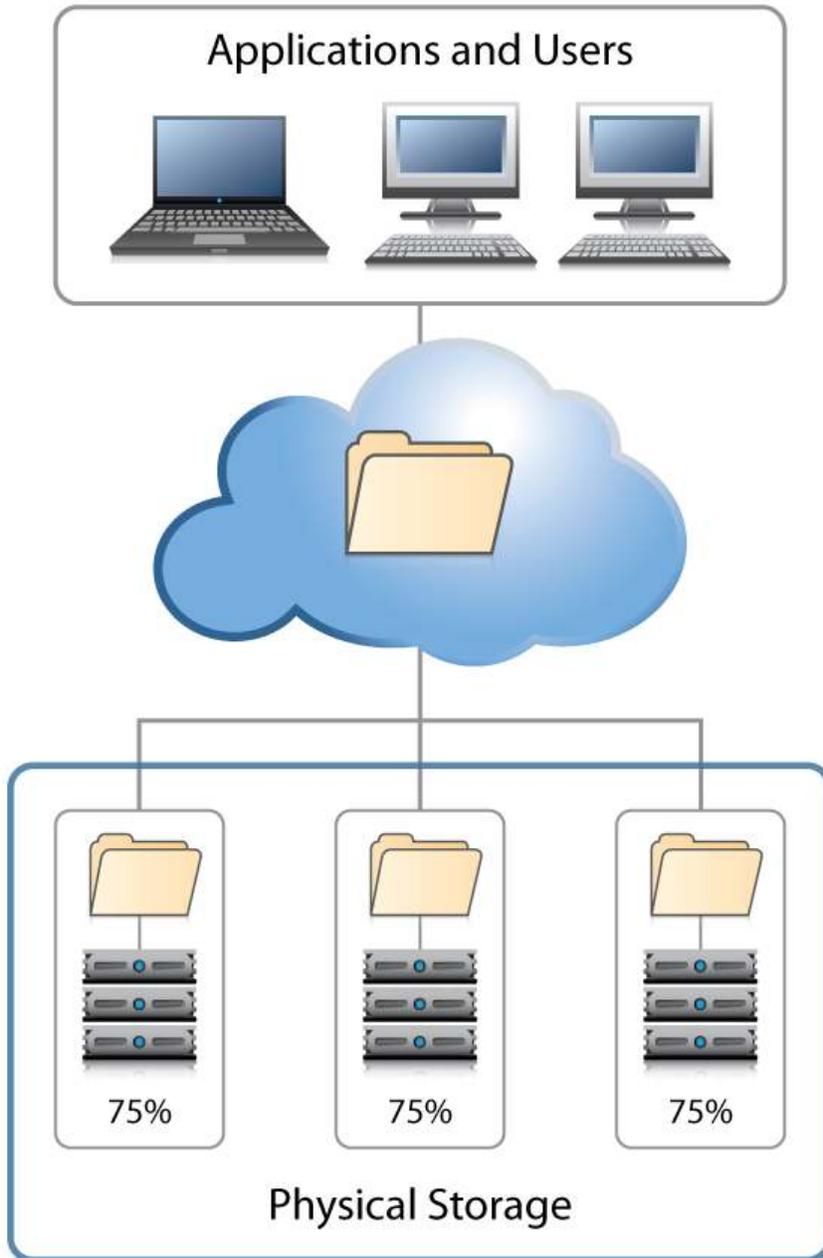


Figure 6: Balancing capacity utilization with ARX



WHITE PAPER

Enabling Flexibility with Intelligent File Virtualization

ARX helps smooth out the effects of uneven and unpredictable data growth. Without having to worry about sudden spikes in utilization, organizations can target higher levels of aggregate utilization in their storage environment. This increases storage efficiency and reduces the total cost of storing data. In addition, dynamic capacity balancing policies can help organizations easily scale their file storage environments on-demand and without disrupting applications or users.

Backup optimization

In addition to reducing storage costs, ARX also provides an important secondary benefit in optimizing data backups. Backup times have grown along with the amount of data under management, and many organizations now face challenges in meeting their backup windows. There are two primary factors contributing to this:

- Organizations have more data that needs to be backed up on a weekly, monthly, and quarterly basis.
- Larger file systems require more time for backup software to traverse. It can often take more time to back up a large file system than is available in the backup window.

ARX can help dramatically reduce the amount of time required to perform a full backup of data. For most organizations, the majority of their data is inactive and not changing. Backing this data up on a weekly basis increases backup times unnecessarily, without improving the level of data protection. Automated storage tiering can separate active and inactive data between different physical locations, without disrupting logical access to those files. For example, ARX can be configured to automatically move files that haven't been modified in over 90 days to tier 2. Organizations can continue backing up active data weekly, but reduce backups of inactive data to a monthly or quarterly basis.

ARX can also present a large virtual file system that's composed of smaller physical file systems to meet scale requirements and for ease of backup. For example, an organization can create a 16 TB virtual file system comprising 32 500 GB physical file systems. Each physical file system can be backed up in much less time, and multiple file systems can be backed up in parallel.

Conclusion

ARX eliminates the complexity and inflexibility of traditional file storage infrastructures. It provides enterprises with flexibility to better manage growing and constantly changing file storage environments. By decoupling the logical access to files from their physical location, ARX offer significant benefits, including:

- Reduced capital and operational expenditures.
- Improved storage efficiency.



WHITE PAPER

Enabling Flexibility with Intelligent File Virtualization

- Reduced backup times and costs.
- Minimized downtime and business disruption.
- Freedom to choose the technology most appropriate for an organization's particular business need.

ARX devices automate what are currently manual storage management tasks and eliminate the downtime associated with these tasks. These capabilities enable organizations to realize significant cost savings, extend the value of their existing file storage investment, and enhance business workflow.

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

Americas
info@f5.com

Asia-Pacific
apacinfo@f5.com

Europe/Middle-East/Africa
emeainfo@f5.com

Japan
f5j-info@f5.com