



F5 White Paper

Building a Cloud-Enabled File Storage Infrastructure

A cloud-enabled infrastructure can help your organization seamlessly integrate cloud storage and maximize cost savings, while also offering significant benefits to your traditional file storage environments.

by Renny Shen

Product Marketing Manager



Contents

Introduction	3
<hr/>	
What Makes a Cloud?	3
Types of Cloud Storage	4
<hr/>	
What Makes Cloud Storage Different?	4
Accessing Files Remotely over the Network	5
Accessing Files on Object-Based Storage	5
Unique Cost Structure	6
<hr/>	
Where Clouds Make Sense	7
Fitting the Cloud into a Tiered Storage Framework	7
Expanding the Parameters for Tiering with the Cloud	8
<hr/>	
Defining Cloud-Enabled	9
Integrating Different Types of Storage	10
Non-Disruptive File Migration	11
Automated Storage Tiering	11
<hr/>	
Benefits of a Cloud-Enabled Infrastructure	12
Reduced Storage Costs	12
Reduced Backup Times and Costs	13
Reduced Operational Costs	13
<hr/>	
The F5 Cloud Storage Model	13
Creating a Private Cloud	15
<hr/>	
Conclusion	18



Introduction

Cloud storage offers enterprise organizations the opportunity to bring constantly rising file storage costs and management burden under control. By moving appropriate types of files to the cloud, organizations can reduce not only the amount of storage capacity that they need to purchase, but also the operational overhead involved in managing it. In addition, the cloud enables storage capacity to be increased on demand, while charging organizations only for the amount of storage that is actually utilized.

Cloud storage will bring many changes to the way enterprises manage storage. As with any disruptive technology, the rate of adoption among organizations will vary. Early adopters might be ready to tier the majority of their business data to the cloud today. Others might prefer to wait or experiment with the cloud in a test environment for an extended time.

Regardless of where your organization is in the adoption curve for cloud storage, there are compelling reasons to prepare the file storage infrastructure for that transition. The same capabilities that are required to integrate cloud storage into existing environments can also offer benefits in traditional environments. These benefits include:

- Reduced storage costs.
- Optimized backup infrastructure.
- Increased operational flexibility with lower overhead.
- Flexibility to easily integrate cloud and other new technologies when and where appropriate.

A cloud-enabled infrastructure can help your organization maximize the capital and operational cost savings from cloud storage. In addition, it enables the flexibility to seamlessly integrate the cloud when ready.

What Makes a Cloud?

As a new and evolving technology, there are often varying opinions about the precise definition of cloud storage. However, the common factor is that of providing storage capacity as a service, typically from a remote location.



Types of Cloud Storage

There are several broadly defined types of cloud storage:

Public cloud

Public cloud storage is the model of delivering storage as a service that is the easiest to understand, as well as the one most frequently associated with the cloud. With a public cloud, organizations utilize storage capacity provided by a third-party entity, located off-premises in a cloud data center, and accessed remotely, over a public network, such as a wide area network (WAN).

Hybrid cloud

A hybrid cloud blends aspects of local and cloud storage. A common example of a hybrid cloud combines storage capacity from a public cloud-storage provider with a local device known as a cloud storage gateway. The gateway makes cloud-based storage capacity appear as a local storage device and performs any necessary protocol translation (see “Accessing Files on Object-Based Storage,” below).

Private cloud

With a private cloud, a central IT group manages storage capacity for the rest of the company and offers it as a service to individual business groups, users, or applications. For smaller organizations, a private cloud can be located in the same data center as its users, whereas larger organizations might support multiple remote facilities from a central data center. Depending on the distance between the remote facilities and cloud data center, some form of WAN optimization might be needed.

What Makes Cloud Storage Different?

Cloud storage differs from traditional storage infrastructures in regard to three key aspects: accessing files remotely over the network, accessing files on object-based storage, and the unique cost structure.



Accessing Files Remotely over the Network

Cloud storage provides geographically dispersed users with storage capacity managed from a central location. By definition, this entails storing data at a location different from where it was created or used. Users typically must access data stored remotely in the cloud. This raises two important considerations:

- **Performance.** Data access over long distances might be impacted by unpredictable network conditions. Depending on the distance between the local facility and the cloud data center, users might experience significant latency when accessing data stored in the cloud. This can result in a poor user experience or unacceptable application performance.
- **Data security.** For public or hybrid clouds, storing business data outside your organization's control creates a new requirement to encrypt stored data. However, even private clouds might involve transmitting files over the public networks. To fully secure business data, it is important to take steps to encrypt data not just once it is in the cloud, but also before it leaves the data center.

Accessing Files on Object-Based Storage

Many cloud storage offerings (either cloud storage services or systems explicitly marketed as cloud) are built on an object-based storage platform. These platforms offer high levels of scalability (in terms of capacity and performance) as well as easy data access over the network via Hypertext Transfer Protocol (HTTP). But, they create an impediment for enterprise organizations looking to integrate the cloud with their existing file storage infrastructure.

Data access with object-based storage is performed through a web services application programming interface (API), based on either the Simple Object Access Protocol (SOAP) or Representational State Transfer (REST) protocol. However, enterprise organizations access their file data through industry-standard Common Internet File System (CIFS) or Network File System (NFS) protocols.

In order to deploy cloud storage in the least disruptive manner, enterprises will need to use a cloud storage gateway. As shown in Figure 1, a cloud storage gateway provides a local file system interface for the remote object-based storage platform. Users and applications access files using standard CIFS or NFS protocols. The gateway translates file access to the appropriate web services API, retrieves the file from the cloud, and places it in the local file system for users to access.

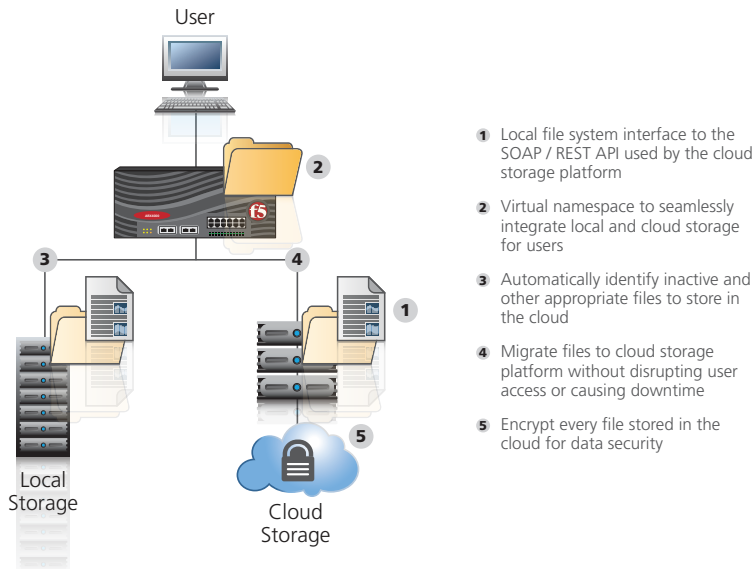


Figure 1: The role of a cloud storage gateway

Unique Cost Structure

Public and hybrid clouds have a different cost structure from traditional storage, due to two key differences: being a managed service and the nature of accessing files on remote storage. Depending on its configuration, the cost of a private cloud might be closer to that of traditional storage.

Cost structure of a managed service

With a public or hybrid cloud, the cloud storage provider rents storage capacity to organizations, charging a monthly fee for the amount of capacity utilized plus network bandwidth costs (see “Added costs of remote storage,” below). Because of the different cost structure, the economics of some cloud offerings initially might not appear to compare favorably to that of traditional storage. However, a comprehensive comparison must consider several elements:

- **Initial investment.** With cloud storage, you can avoid the upfront costs required to purchase traditional storage systems and capacity.
- **Efficiency.** Unlike traditional storage, cloud storage is effectively 100 percent utilized. You only pay for the amount of capacity that is actually utilized for storing data.
- **Data center costs.** Physical storage devices consume power, require cooling, and take up space in your data center. Public and hybrid clouds offload these costs to the cloud storage provider.



- **Operational overhead.** Public and hybrid clouds offload the operational costs of managing storage over time to the cloud storage provider.
- **Managing growth.** Public and hybrid clouds enable more capacity to be added at any time, seamlessly scaling the storage environment as data grows.
- **Support.** With cloud storage, you can avoid the annual expense of vendor support and maintenance for a purchased storage system.

Added costs of remote storage

Because cloud storage entails storing and accessing files over the public networks, organizations must consider the additional costs of network bandwidth. Organizations are responsible for the cost of network bandwidth out from their own data centers. In addition, the public cloud storage provider might levy a fee for the bandwidth utilized in to put files into and retrieve files from their cloud data center.

Private clouds

From the perspective of IT, a private cloud is not a managed service. Since storage capacity must still be purchased and managed, the cost structure is similar to a non-cloud environment. However, the centrally managed private cloud does offer greater economies of scale and operational efficiencies than an operationally dispersed storage infrastructure with many individual points of management. It also offers a much greater degree of control over the storage of business data, including the ability to leverage different storage technologies where appropriate, as well as a lower risk for data security.

With a private cloud, an organization owns both the local and the cloud data center and is responsible for the cost of network bandwidth between them. This provides an opportunity for symmetric WAN optimization to reduce bandwidth consumption.

Where Clouds Make Sense

Fitting the Cloud into a Tiered Storage Framework

Within a tiered storage framework (see Figure 2), the most appropriate type of storage for any type of data is the one that provides the required performance and availability at the lowest cost. A traditional tiered storage environment maximizes user productivity and application performance by placing files with the greatest



business value and/or the highest demand on the storage system with the best performance. As files decrease in value or demand, they are subsequently tiered to lower-cost types of storage.

A geographically dispersed storage environment adds another element of adjusting performance: locality. A shorter distance between a user or application and the file being accessed lowers the potential amount of network latency. Maximizing performance requires storing files that are in high demand on a system that has high performance and that is located close to where those files will be used. Conversely, files that are no longer in demand can be stored more remotely.

Given the additional cost and inherent performance impact of accessing files remotely over the network, cloud storage offerings primarily focus on inactive files—files that are unlikely to be accessed but must be retained. Users and applications are unlikely to access these files and are more likely to tolerate additional latency if they do. This minimizes the risk and impact of any potential performance degradation.

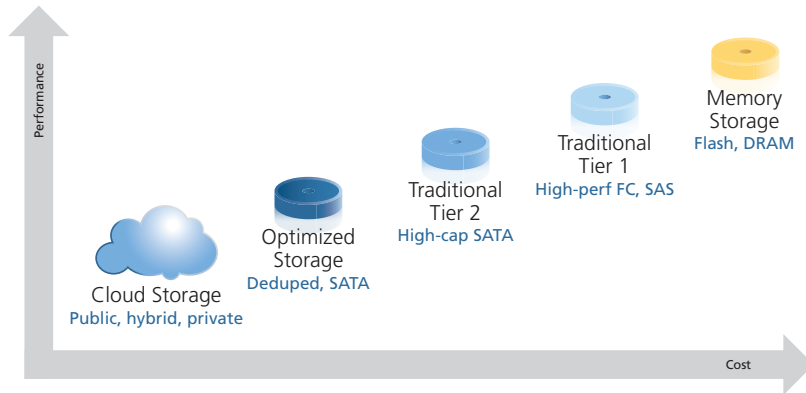


Figure 2: The cloud is a natural extension of the tiered storage framework

Expanding the Parameters for Tiering with the Cloud

Cloud storage goes beyond the traditional parameters involved in characterizing storage tiers. Traditional tiered storage incorporates different storage technologies within a single location and managed by the same staff. By doing so, it isolates the differences between various types of storage to the parameters of performance, availability, and cost per gigabyte of capacity.

While a private cloud can fit into this framework, public or hybrid clouds are remotely managed services. This presents additional parameters to be considered when determining the suitability of a public or hybrid cloud for a particular storage tier:



- **Network cost.** In the context of tiered storage, higher cost is typically associated with a higher level of performance. Cloud storage differs significantly in that organizations incur additional network costs to access a type of storage with lower performance. In addition, these costs apply to continued file access over time, as opposed to an upfront investment.
- **Operational cost savings.** Before the cloud, storage tiering did not consider the operational costs of different tiers. Storage options for all tiers required a similar management overhead proportional to the amount of data stored. However, as a managed service, a public or hybrid cloud offloads the cost of managing storage for any data that is tiered to the cloud.
- **Data security.** Similar to operational costs, data security is not generally as heavily emphasized when deciding the appropriateness of a particular storage tier in traditional storage environments. Typically, every tier is located within the same data center, has the same scheme for access control, and is managed by the same IT staff. However, a public or hybrid cloud provides capacity managed by a third-party entity and located in a cloud data center. As such, you need to consider which types of files can be safely outsourced to a third-party and which types must be kept within your organization's direct control.

Defining Cloud-Enabled

Cloud storage has dramatically different characteristics from on-premises storage systems being deployed today. Organizations looking to maximize the cost savings that cloud storage can offer must be able to effectively leverage storage capacity that is both geographically dispersed and potentially managed by a third-party. This requires certain capabilities within the storage infrastructure to be enabled, including:

- A method of integrating local and cloud storage that is transparent to users and applications.
- The ability to migrate files among different tiers while preserving access by users and applications.
- The ability to automatically classify and migrate files to the most appropriate storage tier.

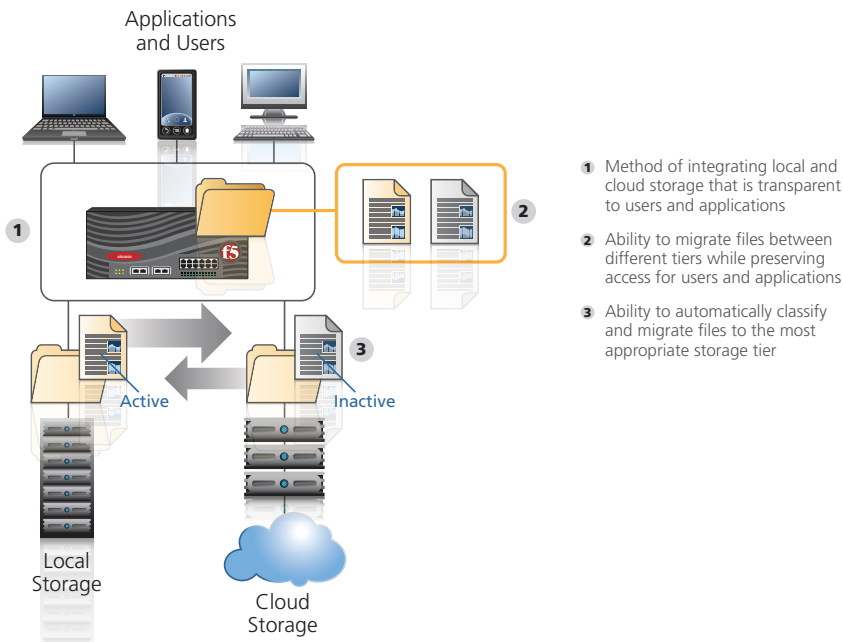


Figure 3: Storage infrastructure capabilities needed to be cloud-enabled

Integrating Different Types of Storage

Most organizations plan to utilize cloud storage to augment their existing file storage infrastructure. Files in active use will be stored on local systems, while files that are unlikely to be accessed will be stored in the cloud. However, users and applications must be able to access files stored on local systems and in the cloud.

The most elegant method of integrating multiple storage systems or different storage types is through a global namespace. A global namespace federates multiple physical file systems and presents them as a single virtual one. Just as a physical file system hides the characteristics of the underlying disk from client systems, a virtual file system hides the characteristics of the underlying storage platforms or systems.

The ability to incorporate storage systems from different vendors or platforms is particularly important with the cloud. The cloud storage market continues to be fragmented, with a multitude of offerings from leading traditional storage vendors as well as startup companies that focus on specific components of the overall cloud storage solution (such as providing a cloud storage gateway).



Non-Disruptive File Migration

The changing nature of business data means that its value to the business will also change. Organizations need the ability to move appropriate types of data (such as inactive files) to the cloud over time. However, users and applications must be able to access files regardless of their current location on local storage or in the cloud, and preferably without IT intervention. This necessitates preserving user access to files over the entirety of their lifecycle with minimal operational overhead.

The global namespace works by decoupling logical access to files from their physical location, as shown in Figure 3. Users and applications access files through persistent logical mappings presented by the virtual file system. The global namespace will proxy a logical file access to the physical file system where the file is stored. This proxy architecture enables the global namespace to mask the physical location of individual files from client systems. Equally important, it preserves logical access to files as they move among different storage systems.

Automated Storage Tiering

The global namespace provides a foundation that enables the movement of files among different storage systems without disrupting user access. Building on that foundation, the storage infrastructure needs the capability to automatically identify the business value of individual files and move them to the most appropriate storage system. This automated storage tiering solution must have the following characteristics:

- **File-level granularity.** An effective storage tiering policy can recognize the unique business value of individual files and make the most appropriate decision for each.
- **Multiple criteria.** Because of the unique characteristics of the cloud, the appropriateness of storing a file in the cloud must consider multiple criteria, such as the file's age and its degree of business sensitivity.
- **Ongoing.** Because the business value of files changes over time, an effective tiering policy must monitor these changes and make adjustments to the placement of files on an ongoing basis.
- **Automated.** Most organizations do not have the resources or time to manually classify individual files on an ongoing basis. In addition, users typically do not have the discipline to properly classify their own files. In order to properly utilize multiple types of storage over time, an effective policy must be able to automatically tier files without IT intervention.



Benefits of a Cloud-Enabled Infrastructure

When you are ready to adopt cloud storage, a cloud-enabled infrastructure provides the ability to integrate cloud storage in the most effective and efficient manner. It also provides significant immediate benefits that can be realized today with existing storage environments.

Reduced Storage Costs

A cloud-enabled infrastructure can help you reduce storage costs even in non-cloud environments. The operational flexibility and intelligent policies required to maximize the benefits of the cloud also apply to traditional storage technologies and systems.

Automated storage tiering

The ability to automatically move files to the most appropriate type of storage is not limited to the cloud. Cloud storage is the latest of many technologies that can be effectively deployed to meet a specific storage requirement, such as Solid State Drive (SSD), Fibre Channel (FC), Serial Advanced Technology Attachment (SATA), and data deduplication. For example, storage technologies that provide high performance have typically come with a high cost per gigabyte of capacity. In this case, storage tiering can ensure that only files that need high performance are placed on the highest tier, maximizing the performance benefit and minimizing the capacity cost per input/output (I/O). On the other hand, technologies that focus on lower cost often come at the expense of lower performance. In this case, storage tiering can ensure that only inactive files are placed on lower tiers. Since inactive data typically represents 70 to 90 percent of all data under management, automated storage tiering maximizes the amount of data placed in lower-capacity storage, providing cost benefits.

Increased utilization

A cloud-enabled storage infrastructure also offers an opportunity to significantly improve operational efficiency. Many storage environments today operate at persistently low levels of aggregate utilization despite high levels of data growth. The disruption and operational overhead of reprovisioning events encourage organizations to overprovision storage capacity up front, resulting in low utilization.



A global namespace eliminates that disruption by enabling additional capacity to be easily provisioned into existing file systems when needed. This enables you to operate at higher levels of target utilization and to better utilize existing capacity rather than purchasing more.

Reduced Backup Times and Costs

In addition to reducing capacity costs, a cloud-enabled storage infrastructure can also optimize backup processes, dramatically reducing backup times and media consumption. Storage tiering effectively separates various ages or types of files among different physical file systems. This enables the backup process for each tier to be customized by the characteristics of files on that tier. For example, consider a tiering policy that moves files that have not been modified in the last 30 days to a separate tier. You can continue performing weekly full backups on active data, while reducing full backups of inactive data to once a month.

In addition, the multiple physical file systems that compose the global namespace can be backed up independently. Not only does each physical file system contain less data and require less time to back up, but multiple physical file systems can be backed up in parallel, further reducing backup times.

Reduced Operational Costs

A cloud-enabled storage infrastructure eliminates the disruption caused by data movement by providing the operational flexibility to perform many storage management tasks in less time and with lower overhead. In traditional environments, moving files can be a highly disruptive operation, requiring significant planning, scheduled downtime, and client reconfiguration. The global namespace decouples the logical access to files from their physical location, preserving access to files regardless of their current locations. This gives you the freedom to move files whenever and wherever you want, and with less operational overhead.

The F5 Cloud Storage Model

When your organization is ready to move to the cloud, F5 offers solutions that help build the right cloud to meet any need—in the most flexible, effective, and economical manner.

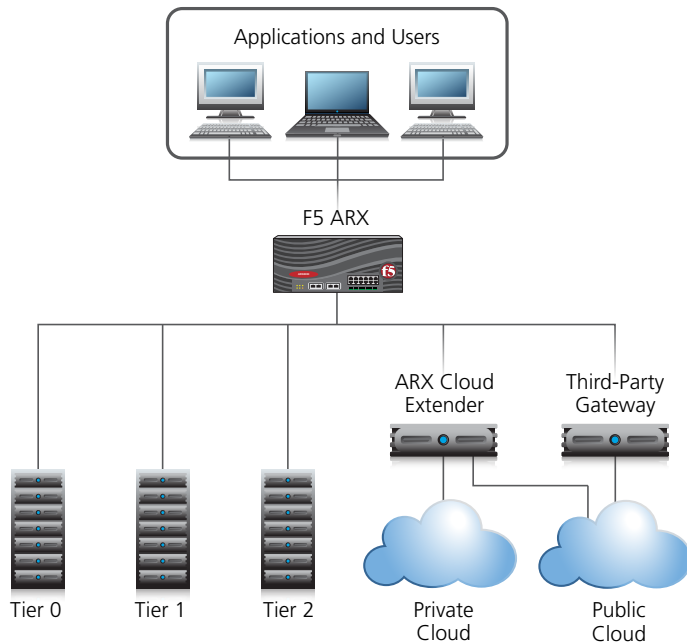


Figure 4: F5 has a flexible and extensible model for integrating and building public and private clouds

Tiering to the Cloud with F5

As shown in Figure 4, F5 offers a flexible and extensible cloud storage model that can help you integrate a variety of cloud storage offerings.

Implementing storage tiering

The F5 ARX® intelligent file virtualization solution establishes the foundation for a flexible cloud storage model by providing a global namespace to virtualize the file storage environment and integrate storage capacity from a variety of different systems, platforms, and vendors. The global namespace decouples logical access to files from their physical locations and enables the movement of files among storage tiers without impacting users and applications.

ARX solutions provide a suite of data management policies that automate the placement and movement of individual files among storage tiers based on their business value. Today, organizations use ARX to create a tiered storage infrastructure within their data centers, whether with existing storage systems or systems that are augmented by high-performance or low-cost storage technologies.



Adding a cloud storage tier

The cloud is a natural extension to a tiered storage infrastructure, augmenting the existing storage tiers within the data center with storage capacity in the cloud. ARX solutions integrate the cloud into a tiered storage infrastructure by incorporating a cloud storage gateway into the global namespace and automatically migrating files to the gateway using a storage tiering policy. Depending on the characteristics of the cloud storage offering, a policy could involve several criteria, including file age, type, size, name, or location.

ARX supports two options for the cloud storage gateway:

- **F5 ARX® Cloud Extender™** A software solution installed on a Windows-based file server, ARX Cloud Extender provides native CIFS access to files stored on a range of public or private cloud storage options. It translates file access to the appropriate web services API and provides additional services such as data encryption and metadata caching.
- **Third-party gateway.** Cloud storage offerings built on object-based storage platforms often provide their own cloud storage gateway to facilitate the migration of files to the cloud storage platform.

Securing your data in the cloud

F5 helps protect your data when tiering to the cloud in two important ways. First, ARX can help ensure that only files that can be appropriately stored at a third-party site are tiered to the cloud. The ARX device's tiering policies are highly customizable and can be configured to exclude sensitive business file types from being tiered to the cloud storage gateway. Second, the ARX Cloud Extender independently encrypts every file using Advanced Encryption Standard (AES) algorithms before they are migrated to the cloud. This ensures that your data is secure when being sent to, stored with, and retrieved from the cloud storage provider.

Creating a Private Cloud

Some organizations might have data storage requirements that cannot be met by existing cloud storage offerings. F5 can help you build a private cloud that meets your needs.



Internal cloud

Organizations with more stringent data security requirements might not feel comfortable storing sensitive business data off-premises and under the control of a third-party entity. With ARX, they have the option of building an internal cloud using either a private cloud storage platform (see “Tiering to the Cloud with F5,” above) or traditional file storage devices. The ARX solution’s global namespace, non-disruptive file movement, and automated tiering policies provide the flexibility to create different types of storage services to meet different requirements.

Every organization has multiple constituencies—including independent business units, project groups, applications, and individual users—with different storage requirements. ARX makes it easy to customize services to meet the requirements of each constituency. A defined service offers a different combination of multiple options, including:

- **Types of storage.** ARX can integrate multiple types of storage, such as SSD, SATA, and deduplicated storage, within its global namespace. It also provides significant flexibility in terms of where and how much of each type of storage is offered. Services for different constituencies can offer various combinations of storage types to meet specific needs.
- **Tiering policy.** Within an organization, each constituency creates and uses different types of data, and in varied ways. You can customize the tiering policy to recognize the criteria for business value that each constituency has.
- **Data protection.** Different constituencies will have varying requirements for data protection, including recovery point objective (RPO) and recovery time objective (RTO). Depending on the requirements, a service can offer disk-based (for example, snapshot) and/or tape-based backup with appropriate full backup and retention policies.
- **Cost.** The cost of a service will be dictated by the types of storage selected, the policy used to tier files between them, and the cost of supporting infrastructure and operations. However, creating a service for budget-sensitive constituencies can start from a target cost per gigabyte and then consider options that help achieve that target.

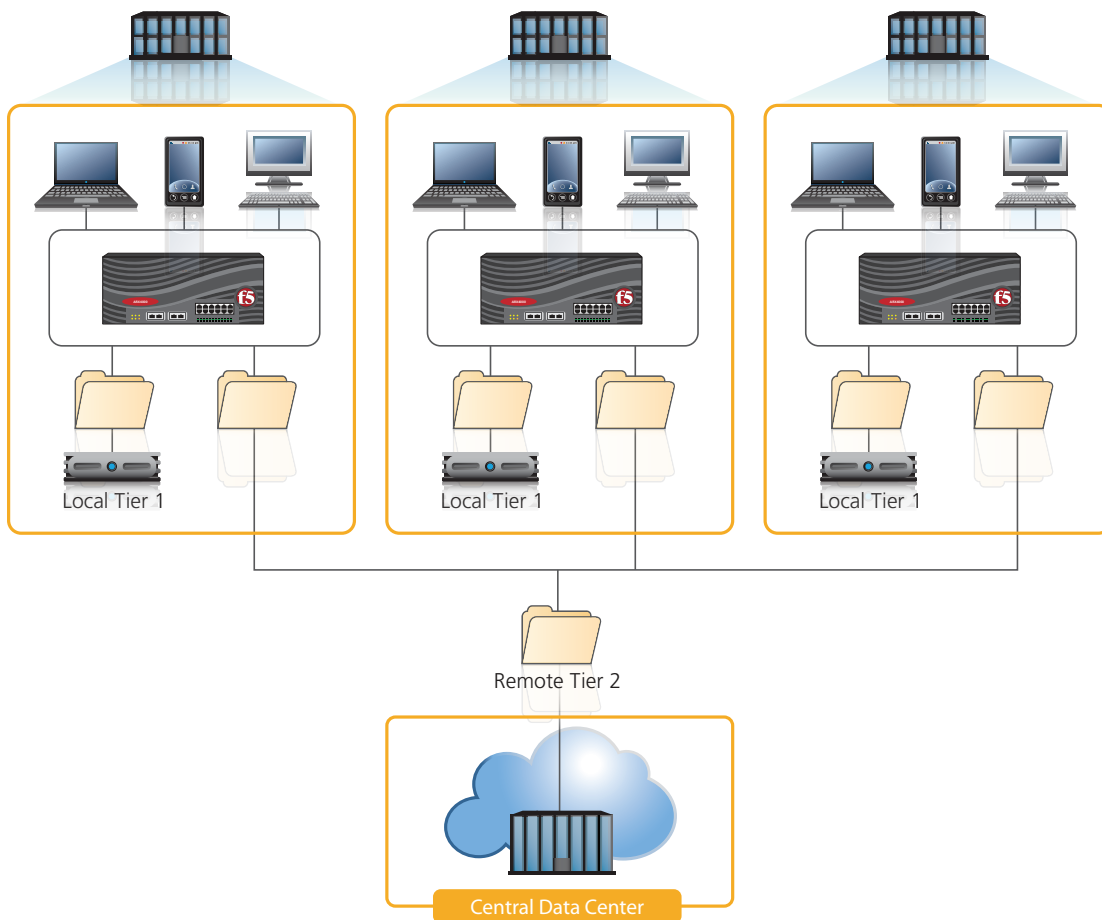


Figure 5: Example of tiering over the WAN

Tiering over the WAN

Many organizations operate multiple facilities, each with its own users, applications, and storage infrastructure. Management tasks, such as purchasing new capacity, managing capacity utilization, performing data migrations, and backing up data are replicated at each location, with the corresponding capital and operational costs. And because data is dispersed across multiple locations, it can be difficult to find and apply economies of scale or operational efficiencies to the management of data and storage.

Tiering over the WAN enables you to offer storage capacity to distributed facilities for inactive data that is remotely managed at the central data center. As shown in Figure 5, organizations can deploy an ARX device in each of their



facilities. ARX federates a small amount of local storage capacity for active data, augmented by additional capacity in the remote data center for inactive data. The ARX device then employs a tiering policy to determine which files must be retained locally and which can be safely tiered to the central data center. By consolidating inactive data at the central data center, you can apply greater economies of scale and operational efficiencies to the majority of data. For example, you can simplify the backup of inactive data by managing it from a central location, rather than across multiple locations. Consolidation also reduces the storage footprint at distributed locations, along with the corresponding capital and operational costs.

Role of WAN optimization

With a private cloud, an organization manages both distributed and central data centers, as well as the network connection between them. This provides an opportunity to improve the performance of accessing or tiering files over the WAN using symmetric WAN optimization:

- **Faster access.** WAN optimization accelerates network traffic over the WAN to reduce the impact of network latency on file access performance.
- **Reduced bandwidth utilization.** WAN optimization can apply compression and/or deduplication algorithms to any files being moved over the WAN to reduce the total amount of bandwidth utilized.

Organizations can easily and cost-effectively enable WAN optimization in their network infrastructure with the F5 BIG-IP® WAN Optimization Module!™

Conclusion

In the face of rapid data growth, cloud storage offers enterprises the opportunity to reduce not only the capital costs of purchasing storage but also the operational costs of managing it. However, the cloud remains a new and evolving technology with significant differences from traditional storage systems. In addition, the implementation of cloud storage can come in many forms, broadly categorized into public, hybrid, and private clouds. Each implementation presents a different set of benefits and limitations. Organizations must consider these attributes in order to determine how and where to best leverage the cloud to meet their needs.

White Paper

Building a Cloud-Enabled File Storage Infrastructure

F5 can help you build a cloud-enabled file storage infrastructure by enabling three essential capabilities:

- **A global namespace** to seamlessly integrate cloud-based storage capacity by presenting it to users and applications alongside existing storage systems
- **Non-disruptive file migration** to preserve user access to files regardless of their current location on local systems or in the cloud
- **Automated storage tiering policies** to maximize the benefit realized from cloud storage by moving only appropriate files into the cloud and with minimal operational overhead

These capabilities can provide immediate, significant benefits to organizations in traditional environments, and ease the eventual transition to the cloud. F5 offers a flexible and extensible cloud storage model to help you incorporate the right cloud into your file storage infrastructure. Using F5 solutions, organizations can leverage public, hybrid, or private cloud storage offerings, or create their own private cloud to build a storage infrastructure that best meets their unique needs.

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

