

WAFS Solutions

Separating the Have and Have-Nots

WAFS Solutions – Separating the Have and Have-Nots

Summary

When is a WAFS offering a true WAFS? It turns out most offerings are not really a WAFS, but merely protocol optimizers that do not address critical issues of data integrity and security.

Introduction

For most people, the much hyped and abused term “WAFS” or Wide Area File Services inherently implies something users have wanted badly for decades: Global File Access – not just in a data center; not just in a metropolitan area, but access to any file, at any time, from anywhere in the world, by any authorized user(s) with guaranteed data integrity and performance approaching LAN access. A global namespace and performance QOS are also desirable as is global sharing and protection of laptop data. Such a solution must also provide moderate to massive scalability and be easy to centrally manage.

This, of course, was the promise of Distributed File Systems (DFSs). However, although many DFSs have emerged over the years, none have had much success in the market. They worked well in small environments, but ultimately they were difficult to scale and manage, suffered performance issues, had security and integrity issues, and were technically difficult to develop, deploy and maintain. Thus, they were mostly used in University environments where labor is cheap. Nonetheless, some Global 500 companies deployed DFSs such as the Andrew File System (AFS) despite the cons. However, AFS has since been withdrawn from the market. And, while users are being promised alternatives they are not happening quickly. Even most WAFS-like offerings do not provide trustable Global File Access.

Note that historically DFSs were designed so that only a single copy of a particular file existed and that file could be anywhere. To improve performance, some DFSs added caching and locking, but these systems suffered from cache coherency and lock management issues.

With WAFS, a single file instance, known as the authoritative copy, is kept in one centralized location. Some of today’s WAFS solutions keep “scratch” copies of files to improve performance, but the authoritative copy always resides in one centralized location.

Interestingly, this policy makes continued write access from a remote location impossible if the WAN goes down. We expect some vendors to introduce recovery

features that temporarily shift the location of the authoritative copy until the WAN is up again.

So Why the WAFS Hype?

The reason WAFS-like offerings are getting so much attention lately is that they address another set of key user needs: WAFS addresses data centralization and consolidation, and no more backing up over a WAN (which never worked very well). With over 60% of corporate data residing outside the data center, storage management of data in branch offices and mobile workers' PCs has been a nightmare. The promise of WAFS-like offerings is that storage can be consolidated and managed in central locations while users still access their data remotely and do not need to even know where the authoritative copies reside and are managed. Meanwhile, by keeping a copy of this remote data, central locations can employ more robust, centralized and proven storage management infrastructure processing for the remote data – the same thing they do for local data.

As such, there is a growing spectrum of offerings billed as WAFS. However, each one is different. They all optimize at least one protocol, but many only address a few layers of the stack. In addition each provides varying levels of optimization, availability, resiliency, cost and manageability.

Protocol Optimization is not Really WAFS

While much has been written on these WAFS-like offerings, little analysis has been done on how well these offerings do or do not address the Global File Access issue and whether they provide true data integrity (see insert). In fact, most of the offerings do little more than protocol optimization including simple data compression, using all of the space available in a network packet, and reducing the number of packets and network turn-arounds common in “chatty” protocols including TCP/IP, CIFS and NFS.

Vendors that provide network file system protocol optimization:

The Risk of File Protocol Optimization - Silent Data Loss with Buffer Flush on Close

- Application opens a file.
- Application writes to the file and the WAFS appliance acknowledges the write operations.
- Application initiates a close on the file and waits since the WAFS appliance is still sending the buffered writes to the central file server.
- WAFS appliance encounters an error (due to WAN disruption, quota overuse, media errors, CIFS timeouts, etc) while sending pending writes to the file server. *At this point data is lost and the file write operation to the authoritative copy is incomplete.*
- WAFS appliance attempts to propagate the error back to the application by sending an error in the file-close response.
- The desktop OS strips the error and propagates a successful close-response to the application.
- Application believes all writes successfully made it to the file server even though some of the data is lost and the file has become corrupt.
- Note that even on UNIX derivatives that correctly propagate errors on close responses almost every application still ignores errors seen in the close response.

- Citrix Systems Inc (acquired Orbital Data Corp.) -- www.orbitaldata.com or www.citrix.com
- F5 Networks Inc. (acquired Swan Labs) -- www.f5.com
- Riverbed Technology Inc. -- www.riverbed.com
- Juniper Networks (acquired Peribit) -- www.juniper.net
- Silver Peak -- www.silver-peak.com

Many of the above listed vendors offer various features in addition to protocol optimization such as compression and network quality of service. But they do not really address the Global File Access issue. Other vendors that are often incorrectly associated with the WAFS market, but actually focus on replication and continuous data protection for remote offices include Avail, Inc., Signiant Inc. and Xosoft (acquired by Computer Associates).

What is a Real WAFS?

A real WAFS solution is a marriage of distributed file system techniques, protocol optimization and a keen understanding of native file system and operating system vagaries. It must survive WAN disruptions without corrupting data. It must also survive appliance, server and software failures. It provides for disconnected/offline operations. It maintains a persistent connection to prevent timeouts. It supports full storage security. It overcomes file protocol optimization risks such as silent data loss (see insert). To aid performance, it understands file structures and, therefore, can offer file differencing. Ideally, it can synchronize laptop data with the WAFS appliance and then synchronize it with the data center. It must also provide easy centralized management including comprehensive lock/lease management.

Unique Features Needed to Implement a Real WAFS

The primary feature of a real WAFS is persistent and sophisticated caching – not just simple file caching, but rather acutely file-aware persistent caching that caches not only the file data, but also the file metadata, user credentials and the file identity in addition to file handle. If the file identity is not cached, recovery cannot be assured. This feature must, of course, guarantee cache coherency and provide scalable file lock management.

Secondly, a true WAFS must possess acute awareness of file access protocols that requires a deep understanding of esoteric application vagaries and shortcomings. Without this awareness, silent data loss is almost guaranteed.

Thirdly, in order to recover from WAFS appliance and WAN failures a true WAFS must employ sophisticated logging and journaling combined with advanced replication – similar to replication systems used in today's modern block serving controllers which faithfully replicate blocks even while enduring WAN and site outages – albeit eventually.

Fourthly, storage security must include Kerberos, SMB signing, and encryption such as SSL and IPsec. Support for ActiveDirectory and LDAP is also a must.

Fifth, a well-designed WAFS incorporates the local native file system protocol stack, thus avoiding reverse engineering and allowing it to incorporate new features more easily. This is particularly important in Microsoft environments.

WAFS vendors that provide products with some of these required features:

- Packeteer (acquired Tacit Networks Inc.) www.tacitnetworks.com or www.packeteer.com
- Expand Networks (acquired Disksites) -- www.expand.com
- Cisco Systems Inc. (acquired Actona) -- www.cisco.com

Some of the vendors address the most important of these five requirements, especially for Microsoft environments that dictate a thorough treatment of SMB, application vagaries and security issues. Some do not and, thus, can suffer silent data loss.

Actionable Advice

If file traffic is required, mere protocol optimization is not sufficient as there will be data integrity and security issues. Ask the vendor how its product protects against silent data loss. Reject products that do not cache files or where the caching techniques are unsophisticated.

Benchmark your own applications – there is no industry benchmark.

Test it to exhaustion and then test it again with even larger files.

Carefully evaluate performance, recovery behavior, manageability, ease-of-deployment and scalability.

See our website for questions to ask the WAFS vendors to better understand application vagaries and how well a WAFS offering addresses these issues.

True WAFS plus WAN Optimization - the Best of All Worlds

One of the reasons that protocol optimization approaches sometimes seem worthwhile is that other, non-file based optimization products can also reside on the same equipment to improve other types of applications that operate over a distributed network. WAN optimization techniques like data compression, protocol acceleration and various caching technologies can broaden the appeal of a new piece of equipment, and extend the value to other applications. That value, however, should not trump the issues we've outlined here, especially when some vendors have found a way to provide true WAFS capabilities, yet also add general WAN optimization technologies.