



Globally Distributed Enterprise File Sharing with Azure NetApp Files and NetApp Global File Cache

Modernizing distributed storage in Microsoft Azure

Streamline and simplify IT storage and infrastructure by centralizing unstructured data into Microsoft Azure using Azure NetApp Files. Provide fast local and geographically distributed access with NetApp Global File Cache.

Contents

1. Executive Summary	3
2. Solution Overview	4
3. About NetApp Global File Cache Intelligent File Caching	6
4. Deployment Methodologies	6
5. User Experience	11
6. Summary	12

1 Executive Summary

One of the biggest challenges that distributed enterprises face is the evolution of consolidation and virtualized infrastructure. However, staying responsive to users' business needs must also remain a priority. Many solutions have tried to provide efficient data management, global enterprise file sharing, and collaboration. But they haven't been able to guarantee a high-quality end-user experience, maintain or improve the performance of enterprise (application) file access, and simplify and improve day-to-day data management tasks.

By overcoming these challenges, you can secure and optimize the delivery of business services and data for all users across your distributed enterprise—including users in or near cloud regions, internet breakouts, branch offices, data centers, and individual endpoints. Solutions need to deliver increasingly high-performance and consistent end-user experiences regardless of geographical location. Furthermore, you need to enable full-scale branch office consolidation as you eliminate performance bottlenecks.

To achieve seamless integration without disrupting user workflow, user experience, and performance, intelligent storage solutions should transparently support the integral part of the Windows file-sharing framework: the SMB protocol. Optimization is key to increasing productivity for file-based applications. With optimization user performance improves, and costs are reduced because you can use your own server infrastructure.

NetApp provides an intelligent, edge-based global file sharing and collaboration environment that is application agnostic, uses end-to-end distributed file locking, and delivers file integrity between users in globally distributed environments—on-premises, in a public cloud, or both. This locking/leasing mechanism allows you to fully benefit from industry-standard application locking principles, which improve productivity and eliminate data loss and file duplicates.

In today's distributed enterprises, over 70% of data currently resides at the edge of a network, and organizations need to secure and manage those file sets more effectively, and preferably centrally. Organizations benefit from the technical and business architecture provided by NetApp® Global File Cache in combination with Microsoft's Azure NetApp Files. This approach transforms a global enterprise's multisite collaboration by improving productivity and business data protection.

NetApp Global File Cache (GFC) and Azure NetApp Files help unify the distribution and presentation of globally accessible content for distributed storage and business applications. The solution:

- Allows for a single set of centralized data that is transparently available throughout the distributed enterprise, at the data center, in branch offices, in remote locations, and within the Microsoft Azure intelligent cloud
- Centralizes and consolidates petabytes of unstructured data on the enterprise-grade Azure NetApp Files file service, hosted in the Azure intelligent cloud
- Provides a software-based solution that delivers global file sharing and collaboration with central file locking, which facilitates flexibility, scalability, ease of deployment, and simplified management
- Enhances RTO/RPO for on-site infrastructure outage from days to minutes; you can continue working from your central dataset while IT is restoring caching functionality

- Further enhances RTO/RPO (down to minutes) by using enterprise-grade time and space-efficient snapshots; enables backup and recovery of entire datasets (up to 100TiB) in a matter of minutes
- Delivers a unified approach to global file sharing that eliminates locally managed storage and backups
- Provides a high-performance end-user experience for your critical business applications on-premises and in the cloud
- Delivers a unified namespace by aggregating datasets into a single pane of glass for global teams
- Provides a cost-effective solution that simplifies IT service management for the enterprise

2 Solution Overview

NetApp and Microsoft—using NetApp ONTAP® data management software—help organizations modernize their distributed storage and their journeys into on-premises, hybrid, or public cloud infrastructures. This proven solution strategy allows enterprises to centralize and consolidate unstructured data while harnessing a software fabric that caches “active datasets” in globally distributed offices. As a result, business users get transparent data access and optimal performance on a global scale.

NetApp and Azure NetApp Files: a major step in unstructured data management for the distributed enterprise

Eighty-five percent of companies are in the process of adopting a cloud transformation strategy. They must combine on-premises, hybrid, and public cloud services and associated storage technologies, such as file-based, block-based, and object storage (for example, Azure Blob), to host both structured and unstructured data.

NetApp and Microsoft recognize the impact on your organization, end users, distributed IT strategy, data center, and cloud operations. The NetApp and Microsoft joint-solution approach allows a scalable, flexible, and cost-effective strategy by addressing all layers of the enterprise, from end users to branch offices, the data center, and cloud infrastructure.

With NetApp Global File Cache intelligent file caching software and Azure NetApp Files, along with Microsoft Azure, your enterprise can do more than just control its data. These products can revolutionize the way you manage unstructured data, both in your daily operations and in user access of that data globally and in the cloud.

Drastically reduce the storage footprint

Consolidation with Azure NetApp Files, enhanced with Global File Cache, gives your distributed branch offices total access to the entire directory structure. Your users have immediate access to all centralized data, which could be hundreds of terabytes or even petabytes of unstructured data. They get streamlined access to all company data, and only active datasets that are relevant to the users in that site are cached locally in the Global File Cache edge cache. Also, as the active dataset ages over time, the Global File Cache intelligent file caching purging mechanism clears the least recently used (LRU) cached files from the local storage cache volumes.

Streamline and simplify distributed IT

As your organization aims to centralize and consolidate its branch office IT storage assets, you can significantly save money by eliminating complexity and backups. NetApp Global File Cache deploys transparently on a (virtual) Microsoft Windows Server instance, on traditional servers or on virtualization platforms like Microsoft Hyper-V or VMware vSphere. Enterprises can consolidate local storage and embed services like Microsoft Active Directory, DNS/Dynamic Host Configuration Protocol (DHCP), Distributed File System (DFS) Namespaces, and software distribution in their streamlined and standardized branch office IT image. Because the actual data is consolidated and stored safely and securely in Azure, you can also centrally handle various data-management tasks like backup and restore, disaster recovery, and archiving, applying powerful mechanisms such as time- and space-efficient snapshots and replication within the Azure intelligent cloud. This approach not only simplifies operations (by taking away these tasks at the edge) but also lets you achieve more aggressive SLAs, further enhancing the business and decreasing business risks.

The Global File Cache fabric, transparently integrated with Azure NetApp Files

Azure NetApp Files stores unstructured data on centralized SMB/file shares in Microsoft Azure using the highest-performing cloud storage technology available. In order to leverage that consolidated data, you can make Azure NetApp Files file shares globally accessible by implementing NetApp Global File Cache software. Global File Cache presents a virtual file share to your network, with an intelligent file cache at each location. This fabric fully integrates with security principles like those employed by Active Directory, access control lists (ACLs), NTFS permissions, and DFS Namespaces to ensure enterprise-grade security and authorization worldwide.

Enable an optimal user experience

With Global File Cache, your distributed users get an optimal experience. By accessing and collaborating on centrally located data in real time—transparent to all client platforms—you feel as if you're all working in the same office, anywhere in the world, regardless of bandwidth, latency, and distance. Imagine the power of each user being able to tap into the sum total of the company's data assets without being limited to the size of the local file server.

Unleash the power of the Azure intelligent cloud

The combination of Azure NetApp Files and the Global File Cache service allows direct access to the data in the cloud by cloud-based applications and users (for example, a WVD workspace). Because advanced (infrastructure as a service) applications and various platform services can access the data directly in Azure NetApp Files, your business can take advantage of powerful, fast, and scalable capabilities to process your data. These capabilities might entail file indexing, searchability, high-performance computing (HPC), and analytics features that are available in Azure or third-party offerings. With Global File Cache, that same data can then be available to authorized users anywhere in the world, in real time.

3 About NetApp Global File Cache Intelligent File Caching

Organizations are facing excessive growth of unstructured data, and they need to be able to centralize and manage those datasets efficiently. The largest Global 2000 companies in architecture, engineering, construction, manufacturing, energy, offshore, healthcare, financial services, and legal have partnered with NetApp to centralize and consolidate unstructured, file-level data more effectively by using enterprise-grade cloud storage while enabling real-time global file sharing and collaboration for the distributed enterprise.

By using its unique approach to a single set of data and applying the intelligent file cache and virtual file share, NetApp lets you access your centralized datasets in the cloud while working on your active dataset (cached locally), with no change in workflow, applications, or user experience.

Global File Cache endpoints are installed by deploying small-footprint Windows Server (2012 R2, 2016, or 2019) virtual appliances in Azure (core) and at each distributed branch office (edges) where data access will be granted. Remote office users connect to their local edge instance, which communicates with the core instance to provide seamless data access with central global file locking. Upon user request, the file is open in the cloud and locked to prevent conflict; it is then either served locally out of cache or compressed and streamed through the edge's intelligent file cache to be served to the user (with subsequent accesses via cache). With the delta differencing mechanism, only the changes to files are sent back to the authoritative source in Azure. Because the data is cached as closely as possible to the users, you experience highly optimized data access in your daily workflow.

Users outside the Global File Cache fabric—for example, those working from home or on-the-road locations—can still connect to the central dataset directly through VPN if that's closer for them. This gives ultimate flexibility, with the ability to access and collaborate on the centralized datasets. Going directly to the data source in Azure mitigates the traversal time and latency in some scenarios.

4 Deployment Methodologies

Global File Cache intelligent file-caching software can be deployed in various ways, either on physical hardware or on virtualization platforms such as Microsoft Hyper-V and VMware. Depending on your needs, the software fabric can be architected as a hub-and-spoke to a main central data store, or via symmetric or hybrid deployment to more than one storage repository in different locations.

Operating environment summary

The most common topology is a hub-and-spoke deployment, because it is typically used for data centralization and consolidation of storage from distributed branch offices into a data center. The key points of the reference architecture are as follows:

- Centralized datastore: Azure NetApp Files service in the Azure intelligent cloud
- Data visibility/access/optimization layer: NetApp Global File Cache fabric for extension of your central datastore to the distributed locations:
 - Global File Cache core virtual machine (VM) instances (for example a DS4_v3 instance) mounting Azure NetApp Files file shares (SMB) in Microsoft Azure
 - Global File Cache edge instances running in each distributed location
 - Presents a virtual file share that provides access to central data in real time
 - Hosts the intelligent file cache on a custom-sized NTFS volume (D:\)
 - Caches active data on-demand or by using scheduled pre-population jobs

- Secure network connection
 - Microsoft Azure ExpressRoute
 - MPLS
 - VPN connectivity (site-to-site or point-to-site)
 - Secure Sockets Layer (SSL) connection

- Authentication/control: integration with customer's Active Directory domain
 - The Active Directory implementation should be a standard Active Directory configuration with on-premises domain controllers; Azure Active Directory or Active Directory Domain Services is not currently supported
 - Only a single domain infrastructure is supported currently (for multiple domain configurations using trust relationships, contact NetApp)
 - The service account should be made a part of the Domain Admins group in the Active Directory domain that contains the Azure NetApp Files resource

- Namespace presentation: DFS Namespaces for the use of a global namespace is recommended
 - Create a unified namespace so that users can access data consistently, always using the local cache in the (nearest) branch office or directly in Microsoft Azure
 - Take advantage of Active Directory Sites and Services to enable location awareness and failover or fallback redundancy if a local branch office outage occurs
 - Aggregate centralized volumes, file shares, and services into a single namespace to allow multi-petabyte scale

Centralized datastore with Azure NetApp Files

The main repository for the unstructured data consists of shares configured on your Azure NetApp Files storage account, hosted in Microsoft Azure, providing direct SMB access to both cloud-resident applications and users (such as WVD workspaces). This centralized approach to storage management enables your organization to apply best-practices for unstructured data in the enterprise, and to scale storage on demand, using proven solutions and processes for file capacity expansion, off-site storage, and data archiving.

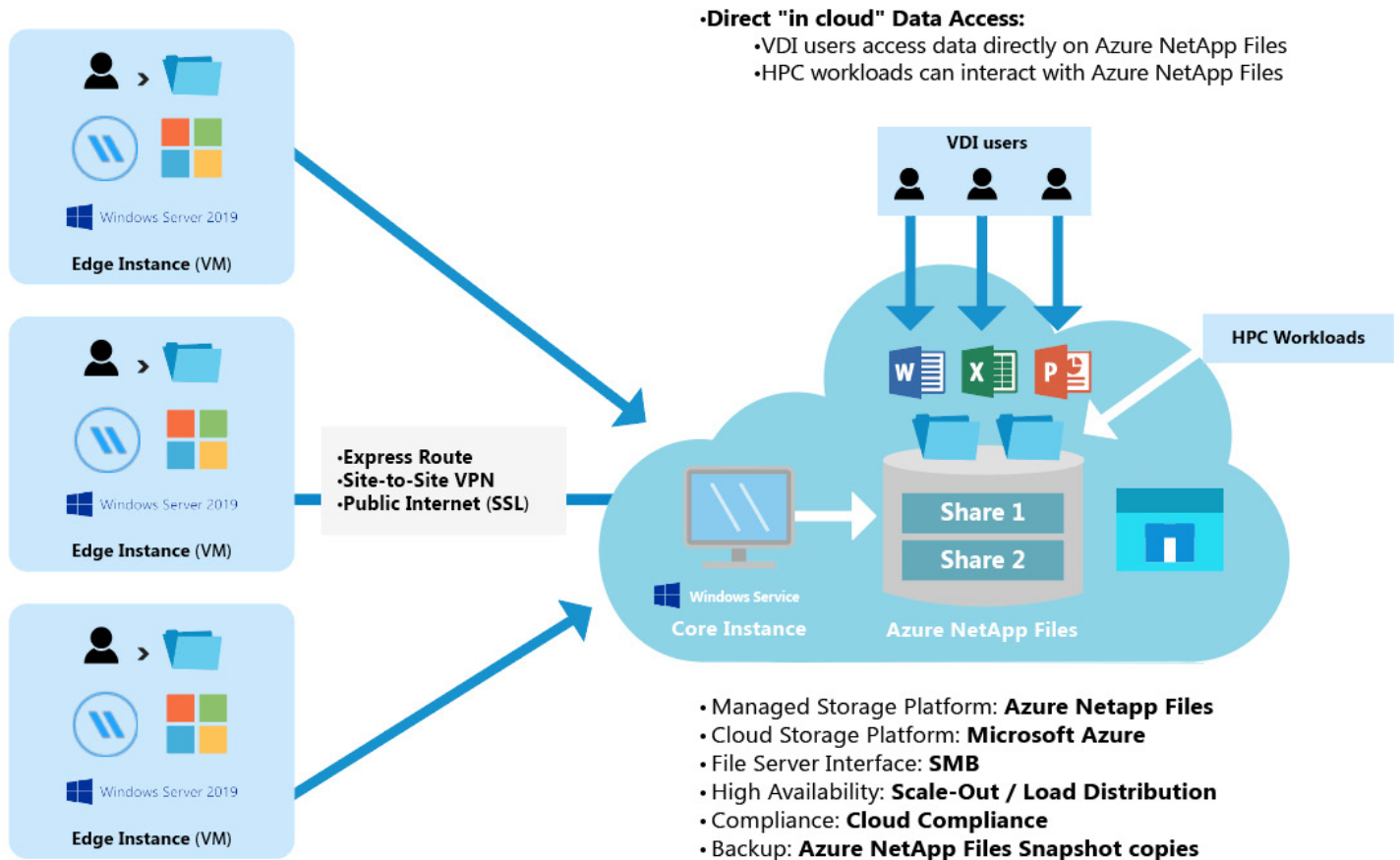


Figure 1) Centralized storage in Azure NetApp Files

Azure NetApp Files is a powerful, cost-effective, and easy-to-use data management solution for your enterprise cloud NAS storage workloads. With Azure NetApp Files, organizations can build a public cloud “data center” that is the foundation of a data fabric that serves the needs of users worldwide. IT teams can then seamlessly manage data as it flows to wherever it’s needed most, using a centrally managed approach to all data, with the benefits and scale of public cloud.

Azure NetApp Files addresses challenges that the digital enterprise faces by:

- Consolidating distributed file servers into Microsoft Azure
- Managing distributed storage and IT infrastructure
- Simplifying deployment and data management
- Flexibly supporting your changing business needs
- Providing scalable, on-demand capacity and performance
- Powering your enterprise applications, both on-premises and in the cloud

Your enterprise can simplify its storage environment with proven operational efficiency by employing economies of scale and the resiliency of the Microsoft Azure public cloud infrastructure and services. Data is secured with built-in encryption, and a consolidated approach with local caching provides a global view of storage with a single management console and a single physical footprint. With this approach, your enterprise can take advantage of:

- Seamless workload migration
- Low-cost disaster recovery using the public cloud
- A reduced storage footprint
- Grow-as-you-go file shares
- Automated DevOps environments
- Azure NetApp Files managed encryption at rest
- Cost-effective data protection

Furthermore, Azure NetApp Files allows your enterprise to simplify operations by unifying data management across a consolidated data footprint. This benefit is further expanded when all enterprise data—even that used primarily by users outside the locale—uses the centralized and consolidated infrastructure, hosted in the public cloud. IT staff can become more efficient, easily moving data and scaling storage resources when and where they're needed most. As a foundation for a global data fabric, Azure NetApp Files provides increased flexibility, control, and security across the public cloud storage environment.

NetApp Global File Cache fabric

When an enterprise introduces NetApp Global File Cache and integrates with existing Azure NetApp Files storage in the cloud, all distributed locations can use the centralized Azure NetApp Files file storage resources as if they were local. The result is a single, centralized storage footprint, versus a distributed storage architecture that requires local data management, backup, security management, storage, infrastructure footprint, and so on, in each location.

Global File Cache edge instances transparently integrate with the Global File Cache fabric in your Azure subscription:

- Distributed locations connect to the Azure instances through the Global File Cache fabric
- Global File Cache presents a virtual file share, with user access optimized via intelligent file cache at each location
- High-performance global file sharing with real-time distributed file locking is enabled

The software overlays the Microsoft Windows file-sharing mechanism, fully integrating with Microsoft security principles like Active Directory, ACLs, and NTFS permissions. This integration allows the software to perform at a global scale, even in locations that are challenged with poor connectivity (low bandwidth or high latency). Global File Cache edge instances offer the following benefits:

- Flexible: storage agnostic; works with any SMB infrastructure/version presented by Azure NetApp Files
- Intelligent: caches only what's needed at the edge (active dataset)
- Zero-touch: automatically purges "stale" cached files over time (LRU) to retain a consistent performance profile
- High-performing: compresses, streams, and reduces data in the event that WAN transmission is needed
- Consistent: distributed file locking for enterprise applications

Global File Cache core instances

Sitting in front of the Azure NetApp Files storage service is one or more Global File Cache core instances, each configured on a Windows VM in co-resident in the same subnet as the Azure NetApp Files resource. The Global File Cache core extends your centralized file shares to your organization's distributed locations, acting as a "traffic cop" between users in the distributed locations and the actual file storage in the Azure NetApp Files volumes. With Global File Cache core instances, ACLs and NTFS file semantics are fully supported to maintain data coherency and integrity. Distributed file locking is also assured regardless of where users are located when they open and work on files. A Global File Cache core instance:

- Mounts corporate file shares, hosted on Azure NetApp Files
- Presents the namespace for the Azure NetApp Files resident shares to the edge locations
- Provides data-in-transit encryption, streaming, and compression between Microsoft Azure and the network edge
- Manages distributed file locks and leases, directly handled by the Azure NetApp Files service

Global File Cache edge instances

Each distributed office runs an instance of the Global File Cache software, configured as an edge instance. The edge VMs provide critical performance-enhancing functions such as file caching, file-level differencing, and local service to users. Each edge instance gives you in that location a full view of all the cloud-resident shares to which you are authorized, essentially allowing even small locations to take advantage of all enterprise information assets without having that data housed on-site.

Edge instances can run on Windows Server 2012 R2 and later, either on physical hardware or virtual infrastructure in the remote location. They employ an intelligent file-caching partition that uses an array of algorithms to retain the most active dataset for that location, providing optimal performance to users located there.

Global File Cache edge instances:

- Have a software installation package or virtual appliance running on Windows Server 2012 R2 and later
- Create a virtual file share; for example, `\\Edge\FASTData[datacenter][fileserver][share][folder]`
- Contain the Global File Cache intelligent file cache

Network connectivity

Connectivity is provided by the customer's existing network infrastructure, either Microsoft Azure ExpressRoute or a secure site-to-site or point-to-site VPN connection from each location needing access to the centralized data in Azure. A VPN connection must be able to carry bi-directional traffic on TCP ports 6618–6630 between the Global File Cache edge and the corresponding Global File Cache core instances. If no direct connection or VPN is feasible, an SSL connection can be enabled between the core and edge instances.

Configuration guidelines

For optimal deployment and performance, see the Global File Cache hardware and software requirements and application best-practices guides at www.cloud.netapp.com/global-file-cache.

5 User Experience

When Global File Cache extends Azure NetApp Files centralized file shares to globally distributed users, users have access to the shares through either a Uniform Naming Convention (UNC) path or a DFS Namespace.

When the system is configured properly, you should feel as if you're using a local file server. For example, users or applications can navigate to a directory structure, select shares and folders, and work with files. You get the complete range of file operations (open, save, copy, paste, and so on) and don't have to change your workflow, and there are no special client software or agents required. The workflow is as follows:

1. When a user requests use of a file in the central repository, Active Directory authenticates that user's access rights.
2. After a successful authentication, the file is opened centrally in Azure NetApp Files by the Global File Cache core, and a lock is applied (centrally by Azure NetApp Files) to that file.
3. If the file has never been used by a user in that location, the file is served from Azure NetApp Files through the Global File Cache core to the Global File Cache edge; Global File Cache proprietary streaming and compression technologies are used to improve performance.
 - If, however, the file has been used by a user in that location before, it is likely to already be present in the local Global File Cache edge cache. If so, the file is served out of the local cache without incurring network transfer operations, providing a high-performance experience, albeit with centrally managed files.
 - If the file in the local cache is less recent than the version in the authoritative back-end file server repository in Azure NetApp Files, any differences (and only the differences) are sent to the local cache and merged with the cached version of the file as it is being opened. This approach maximizes performance, minimizes network resource use, and is totally transparent to the user.

Note: The file remains locked at the central repository in Azure NetApp Files, and is served only after authentication and lock are performed.

4. User operations continue normally, and any updates, changes, and writes are cached locally using the write-back caching function of the Global File Cache edge.
5. Upon save/exit, any changes to the file are differenced back to the authoritative central copy.
6. Upon exit, after saves are completed centrally, the file is closed, and the lock is released and available to other users

6 Summary

This combination of Azure NetApp Files and NetApp Global File Cache software allows your enterprise to consolidate your unstructured data to a centralized single set of data. Your enterprise can take advantage of the flexibility, availability, and economics of a centralized storage model in the public cloud for one of your largest use cases—unstructured data—while maintaining a reduced storage footprint at the data center and distributed branch locations.

The effect on your business is significant:

- A consolidated file storage environment that gains the benefits of the public cloud and the scalability and performance of an enterprise storage service
- A reduced storage-infrastructure footprint at distributed locations through intelligent file caching
- Significant cost savings in the hardware and management aspects of providing file services to distributed users
- The ability to use the enterprise-class scale, flexibility, and security of Azure NetApp Files—resident data across all users, regardless of location, both on-premises and in the cloud
- Increased flexibility and agility through enhanced global collaboration
- Enhanced security and compliance by removing the risk of data loss or leak at distributed locations through error, disaster, and intrusion

About NetApp

NetApp is the leader in cloud data services, empowering global organizations to change their world with data. Together with our partners, we are the only ones who can help you build your unique data fabric. Simplify hybrid multi-cloud and securely deliver the right data, services and applications to the right people at the right time. Learn more at www.netapp.com.

Refer to the Interoperability Matrix Tool (IMT) on the NetApp Support site to validate that the exact product and feature versions described in this document are supported for your specific environment. The NetApp IMT defines the product components and versions that can be used to construct configurations supported by NetApp. Specific results depend on each customer's installation in accordance with published specifications.

Copyright Information

Copyright© 2020 NetApp, Inc. All Rights Reserved. Printed in the U.S. No part of this document covered by copyright may be reproduced in any form or by any means—graphic, electronic, or mechanical, including photocopying, recording, taping, or storage in an electronic retrieval system—without prior written permission of the copyright owner.

Software derived from copyrighted NetApp material is subject to the following license and disclaimer:

THIS SOFTWARE IS PROVIDED BY NETAPP "AS IS" AND WITHOUT ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE, WHICH ARE HEREBY DISCLAIMED. IN NO EVENT SHALL NETAPP BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

NetApp reserves the right to change any products described herein at any time, and without notice. NetApp assumes no responsibility or liability arising from the use of products described herein, except as expressly agreed to in writing by NetApp. The use or purchase of this product does not convey a license under any patent rights, trademark rights, or any other intellectual property rights of NetApp.

The product described in this manual may be protected by one or more U.S. patents, foreign patents, or pending applications.

Data contained herein pertains to a commercial item (as defined in FAR 2.101) and is proprietary to NetApp, Inc. The U.S. Government has a non-exclusive, non-transferrable, non-sublicensable, worldwide, limited irrevocable license to use the Data only in connection with and in support of the U.S. Government contract under which the Data was delivered. Except as provided herein, the Data may not be used, disclosed, reproduced, modified, performed, or displayed without the prior written approval of NetApp, Inc. United States Government license rights for the Department of Defense are limited to those rights identified in DFARS clause 252.227-7015(b).

Trademark Information



NETAPP, the NETAPP logo, and the marks listed at <http://www.netapp.com/TM> are trademarks of NetApp, Inc. Other company and product names may be trademarks of their respective owners.