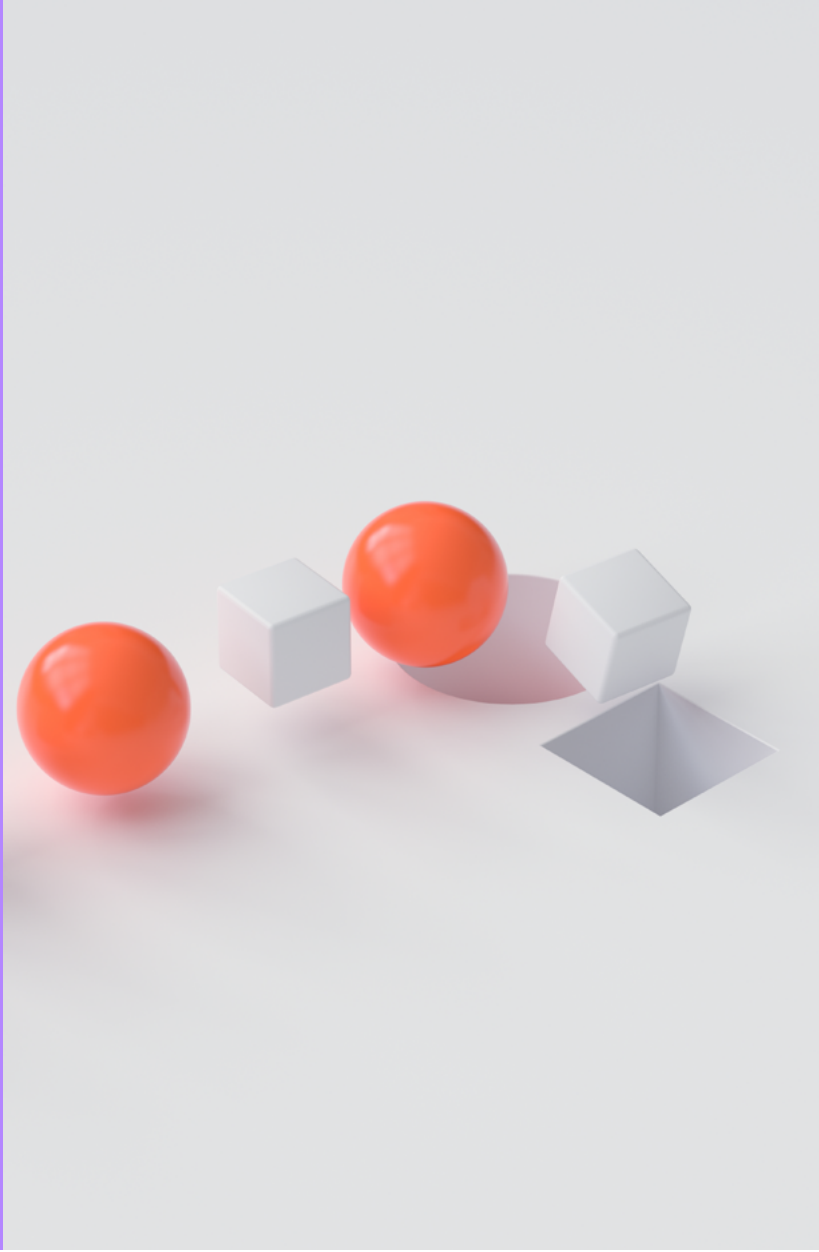
WHITEPAPER

Global File Cache Reference Architecture

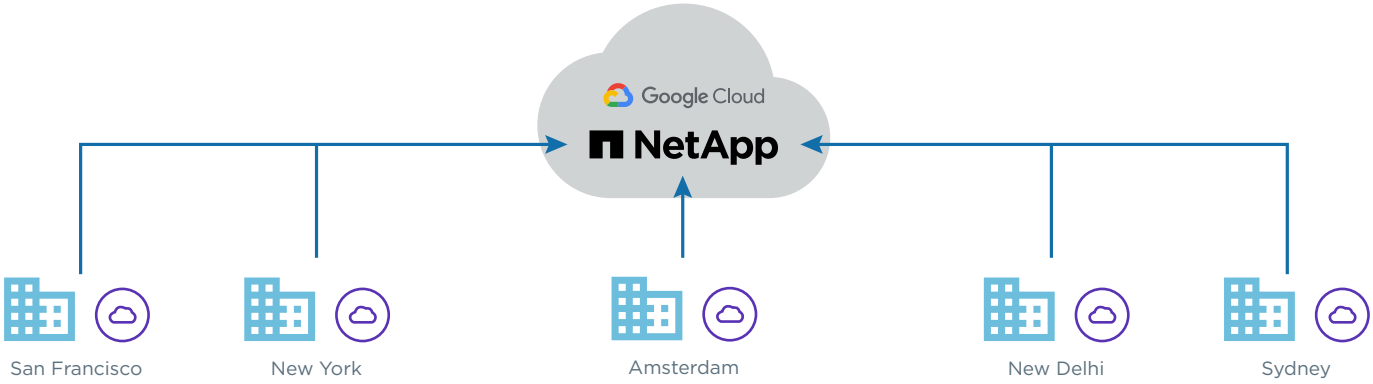
Extending NetApp Cloud Volumes for Google Cloud



NetApp | Google Cloud

Table of contents

Introduction	**3**
Deployment methodologies	**3**
Centralized data store with NetApp Cloud Volumes for Google Cloud	**4**
NetApp Cloud Volumes for Google Cloud	**4**
Features of Cloud Volumes for Google Cloud	**5**
Joint-Solution: NetApp Cloud Volumes Services and Global File Cache Intelligent File Caching software	**5**
Global File Cache Fabric	**5**
Global File Cache Core Instance	**6**
Global File Cache Edge Instance	**6**
Network connectivity	**6**
Configuration guidelines	**7**
User experience	**7**
Summary	**7**



Introduction

NetApp® Global File Cache and NetApp Cloud Volumes for Google Cloud allow businesses to centralize data into a single cohesive footprint by consolidating distributed storage and IT assets into the Google public cloud platform. This software-defined solution strategy enables enterprises to transparently extend Cloud Volumes for Google Cloud managed data to users globally, with real-time global file sharing and collaboration, without changing their workflow or experience.

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 or VMware. Depending on the client's needs, the software can be architected as a hub-and-spoke, symmetric, or hybrid deployment.

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 cloud instance. This paper describes the integration of Global File Cache with NetApp Cloud Volumes for Google Cloud storage in Google Cloud. This integration enables consolidation of all enterprise data onto the secure and scalable platform for access by all users.

Operating Environment Summary

The topology described in this paper is a hub-and-spoke model. In this model, the distributed offices or locations are all accessing one common set of data in your Google Cloud environment. The key points of the reference architecture are as follows:

- **A centralized datastore** consisting of NetApp Cloud Volumes for Google Cloud
- **A Global File Cache fabric**, extending the central datastore to the distributed locations:
  - A Global File Cache core virtual machine (VM) instance, GCP n1-standard-8, mounting to Cloud Volumes shares (CIFS/SMB) in public Google Cloud
  - A Global File Cache edge instance, running as a small Windows VM in each distributed location
- **A virtual file share**, which provides access to centralized data in real time:
  - Hosts the intelligent file cache on a custom-sized NTFS volume (D:\) on the edge instance in each location
  - Caches "active" data dynamically or by using scheduled pre-population jobs
- **Network configuration** support includes:
  - Google Cloud Interconnect/Virtual Private Cloud (VPC)
  - VPN connectivity
  - Secure Sockets Layer (SSL) connection
- **Integration with the customer's Active Directory domain**
- **Microsoft Distributed File System (DFS) Namespaces** for the use of a global namespace (recommended)

Centralized Datastore with NetApp Cloud Volumes for Google Cloud Storage

The main repository for the unstructured data is a share (or several shares) configured on the customer's NetApp Cloud Volumes for Google Cloud instance hosted in Google's public cloud, providing SMB/CIFS integration directly. This environment provides storage volumes and their associated file shares. See Figure 1.

This centralized approach to storage management enables organizations to apply their public cloud and data center best practices to unstructured data across the enterprise. It also enables them to scale storage on demand by using their proven solutions and processes for file capacity expansion, off-site storage, and data archiving.

NetApp Cloud Volumes for Google Cloud

NetApp Cloud Volumes for Google Cloud is a powerful, cost-effective, and easy-to-use data management solution for your enterprise cloud storage workloads.

With NetApp Cloud Volumes, organizations can build a public cloud data center that is the foundation of a data fabric that serves users worldwide. IT teams can then seamlessly manage data as it flows wherever it is needed most, by using a centrally managed approach to all data, with the benefits and scale of public cloud. The familiar and reliable NetApp ONTAP® data management software provides industry-leading data protection and storage efficiency features.

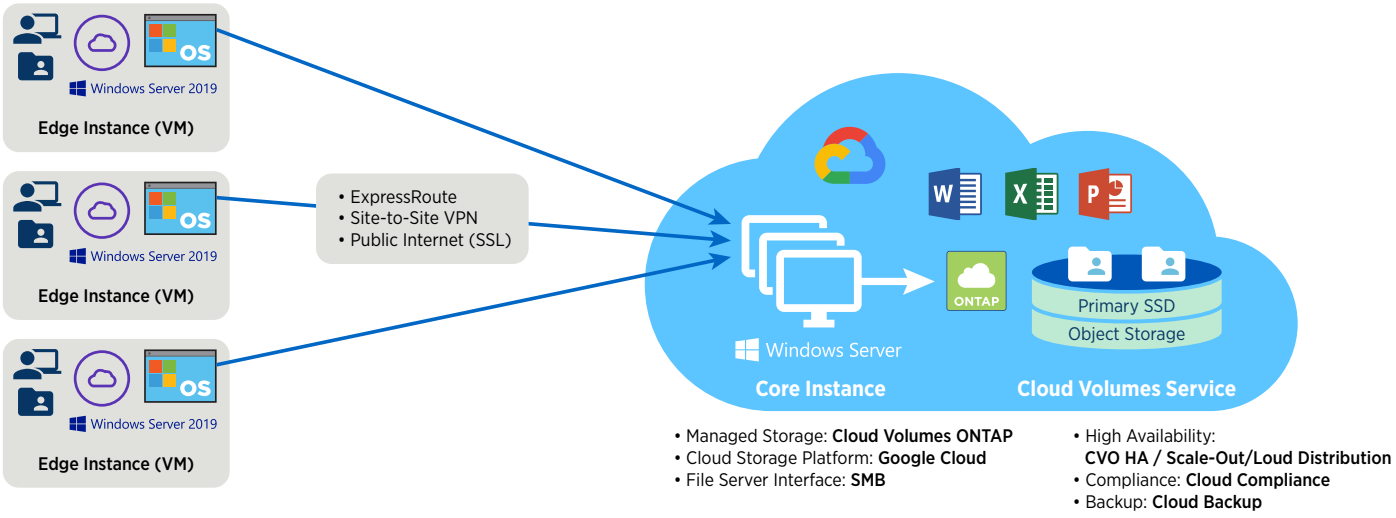


Figure 1) File shares hosted on NetApp Cloud Volumes for Google Cloud.

Cloud Volumes for Google Cloud combined with the Global File Cache service addresses the challenges the digital enterprise faces by:

- Consolidating distributed file servers into Google Cloud
- Managing distributed storage and IT infrastructure
- Simplifying deployment and data management
- Flexibly supporting your changing business needs
- Powering your enterprise applications

Enterprises can simplify their storage environments with proven operational efficiency by combining economies of scale and the resiliency of Google's public cloud infrastructure. Data can be secured with built-in encryption, and a consolidated approach gives a global view of storage with a single management console and a single physical footprint.

Features of Cloud Volumes for Google Cloud

- Nondisruptive operations
- Seamless workload migration
- Low-cost disaster recovery using a public cloud
- Reduced storage footprint
- Grow-as-you-go file shares
- Automated DevOps environments
- Encryption at rest managed by Cloud Volumes for Google Cloud
- Cost-effective data protection
- Integration into Google Cloud UI; billing and support from Google Cloud

Unified Data Management Using Public Cloud
Cloud Volumes for Google Cloud allows the enterprise to simplify operations by unifying data management across a consolidated data footprint. This benefit increases when all the enterprise data—even the data used primarily by staff outside the data center locale—uses the centralized and consolidated infrastructure, hosted in the public cloud. This approach can increase the efficiency of IT staff, because they can easily manage data and scale storage resources when and where they're needed most. As a foundation for a global data fabric, Cloud Volumes for Google Cloud provides increased flexibility, control, and security across the public cloud storage environment.

Extending the Cloud: NetApp Global File Cache Service

Global File Cache extends the cloud by offering the following benefits:

- Centralizes and consolidates unstructured data workloads into public cloud storage platforms that run NetApp Cloud Volumes for Google Cloud
- Provides a transparent SMB/CIFS interface through the Global File Cache fabric, presented by Cloud Volumes for Google Cloud
- Provides native service and software-defined storage and data management (LUNs, volumes, ACLs, NTFS permissions)
- Takes advantage of data deduplication, compression, and volume encryption
- Integrates with Microsoft Active Directory authentication and authorizationIntegrates with NetApp Cloud Volumes for Google Cloud business continuity solutions (recovery time objective [RTO] and recovery point objective [RPO]) for NetApp Snapshot™ copies, backup/restore, and archiving
- Strengthens disaster recovery by using Snapshot copies and Global File Cache fabric clustering
- Transparently extends central file storage to distributed locations, integrating the Global File Cache intelligent file cache at the edge through Microsoft Hyper-V and VMware vSphere

Global File Cache Fabric

When an enterprise introduces Global File Cache software, all distributed locations can use the centralized Cloud Volumes file storage resources as if they were local. The result is a single, centralized storage footprint, versus a distributed storage architecture that requires data management, backup, security management, storage, and infrastructure footprint in each location.

The Global File Cache edge instances transparently integrate with the Global File Cache fabric hosted in Google Cloud:

- Distributed locations connect to Cloud Volumes through the Global File Cache fabric.
- The software provides a virtual file share and intelligent file cache at each location.
- The software enables high-performance global file sharing with real-time distributed file locking.

The Global File Cache edge instance overlays the Microsoft Windows file-sharing mechanism, fully integrating with Microsoft security principles such as those used by Active Directory, access control lists (ACLs), and NTFS permissions. The result is performance at a global scale, even in locations that are challenged with poor connectivity (low bandwidth or high latency).

The Global File Cache fabric offers the following benefits:

- **Flexible**. Various Cloud Volumes offerings are supported, including self-managed NetApp Cloud Volumes ONTAP and fully managed Cloud Volumes Service for Google Cloud, a native Google Cloud service. The offerings all use the industry-standard SMB/CIFS interface, so your choice depends on your requirements for balancing cost and management ease.
- **Intelligent**. The service caches only what's needed at the branch (the active dataset).
- **Zero-touch**. Global File Cache automatically purges "stale" cached files over time (least recently used, or LRU).
- **High performing**. The solution compresses, streams, and reduces data.
- **Consistent**. The solution provides distributed file locking for enterprise applications.

Global File Cache Core Instance

Sitting in front of the Cloud Volumes for Google Cloud storage infrastructure is a Global File Cache core instance, configured on a VM instance. The Global File Cache core extends the centralized file shares to the organization's distributed locations. It manages traffic among users in the distributed locations and the actual file storage in the public cloud data center.

With the Global File Cache core instance, ACLs and NTFS file semantics are fully supported to maintain data coherency and integrity. In addition, distributed file locking is implemented regardless of where users are located geographically when they open and work on files.

The Global File Cache core instance does the following:

- Mounts corporate file shares, hosted on Cloud Volumes for Google Cloud storage in the public cloud
- Provides streaming and compression between the data center and the edge
- Manages distributed file locks and leases

Global File Cache Edge Instance

Each distributed office runs an instance of the Global File Cache software, configured as an edge instance. The edge VMs provide the critical performance-enhancing functions such as file caching, file-level differencing, and local service to users. Each edge instance lets the users in that location view the Google Cloud-resident shares to which they are authorized. In this way, even smaller locations can take advantage of enterprise information assets; that data doesn't need to be physically housed on site.

The edge instances can run on Windows Server 2012 R2 and later, either on physical hardware or virtual infrastructure in the distributed location. An edge instance employs an intelligent file caching partition that uses an array of algorithms to retain the most active dataset for that location, resulting in optimal performance for users located there.

The Global File Cache Edge instance:

- Is a software installation package or virtual appliance running on Windows Server 2012 R2 and later
- Creates a virtual file share: \\edge\Global File CacheData\[datacenter]\[fileserver]\[share]\[folder]\
- Contains the Global File Cache intelligent file cache

Network Connectivity

Connectivity is provided by the customer's existing network infrastructure. That infrastructure can be either of the following:

- Google Cloud Interconnect/VPC
- A secure site-to-site or point-to-site VPN connection between the centralized data center and each location needing access to the data center

If no direct connection or VPN is available or feasible, there is also the choice of enabling an SSL connection between the core and edge instances. When a VPN is used, it must be able to carry bidirectional traffic on TCP ports 6618–6630 between the Global File Cache edge and the corresponding Global File Cache Core instances.

Configuration Guidelines

For optimal deployment and performance, review the Global File Cache Hardware and Software requirements and Application Best practices guides at www.cloud.netapp.com/googlecloud.

User Experience

By using the Global File Cache solution with Cloud Volumes for Google Cloud, globally distributed users have access to file shares through either a Uniform Naming Convention (UNC) path or a DFS Namespace.

For properly configured systems, using the solution is like using a local file server: users or applications can navigate to a directory structure, select shares or folders, and work with files. The complete range of file operations (open, save, copy, paste, and so on) is available, requiring no change in user workflow.

- When a user requests use of a file in the central repository, Active Directory authenticates that user's access rights.
- After a successful authentication, the file is opened centrally in the Cloud Volumes instance by the Global File Cache core, and a lock is applied (centrally) to that file.
- If the file has never been used by a user in that location, the file is served through the proprietary streaming and compression technologies of the Global File Cache solution to improve performance.
  - If, however, the file has previously been used by any user in that location, it is probably already present in the local cache. If it is present, the file is served out of the local cache without incurring network transfer operations; this approach provides a high-performance experience.
  - If the file in the local cache is not the latest version that is in the authoritative back-end file server repository, any differences (and only the differences) are sent to the local cache and merged with the cached version of the file when the file is opened. This approach maximizes performance and minimizes network resource use. Note: The file remains locked at the central repository, and is served only after authentication and lock are performed.
  - User operations continue as normal, and any updates, changes, or writes are cached locally.
  - Upon save or exit, any changes to the file are differenced back to the authoritative central copy.
  - Upon exit, after the saves are completed centrally, the file is closed and the lock is released and available to other users.

Summary

The combination of public cloud storage technologies in Cloud Volumes for Google Cloud with the Global File Cache service allows enterprises to consolidate their unstructured data to a centralized "single set of data." Enterprises can take advantage of the flexibility, availability, and economics of a centralized storage model in the public cloud for one of their largest use cases—unstructured data. At the same time, they can maintain a reduced storage footprint throughout all of their locations. The resulting effects on the business are substantial:

- A consolidated file storage environment that exploits the benefits of the public cloud and the scalability and performance from an enterprise storage system
- 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 apply the enterprise-class scale, flexibility, and security of Cloud Volumes resident data across all users, regardless of location
- 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 For more information, visit: www.cloud.netapp.com/googlecloud.

About NetApp

In a world full of generalists, NetApp is a specialist. We're focused on one thing, helping your business get the most out of your data. NetApp brings the enterprise-grade data services you rely on into the cloud, and the simple flexibility of cloud into the data center. Our industry-leading solutions work across diverse customer environments and the world's biggest public clouds.

As a cloud-led, data-centric software company, only NetApp can help build your unique data fabric, simplify and connect your cloud, and securely deliver the right data, services and applications to the right people—anytime, anywhere. www.netapp.com

© 2021 NetApp, Inc. All Rights Reserved. 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. NA-428-0121

Copyright Information

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 that are supported by NetApp. Specific results depend on each customer's installation in accordance with published specifications.

Copyright © 2021 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.