The NetApp Guide to Storage Tiering: Bringing the Cloud to the Data Center
Introduction

Storage tiering hasn’t always been as simple as using the Cloud Tiering service for NetApp AFF and SSD-backed FAS systems. Shifting data between storage systems with different performance characteristics has been a long-standing practice to optimize the data lifecycle, and some have been more practical than others. The same is true with the latest development in tiering technology: moving data from data centers to the cloud.

What did it take to get from there to here? In this guidebook, we will outline the beginnings of storage tiering, the current uses for the technology, and the cloud evolution that is happening now in tiering physical storage to cloud-based storage, including NetApp’s Cloud Tiering service.
# Table of Contents

- Introduction ............................................................................................................................... 2
- Table of Content ......................................................................................................................... 3
- What Is Storage Tiering? ............................................................................................................ 4
- Different Forms of Tiering .......................................................................................................... 5
- A Look Back at Tiering History .................................................................................................. 6
- How Tiering Should Be Done ...................................................................................................... 7
- Meet the Cloud-Based Object Storage Offerings ....................................................................... 8
  - Amazon Web Services S3 storage ............................................................................................ 8
  - Azure Blob Storage .................................................................................................................. 9
  - Google Cloud Storage .............................................................................................................. 9
- How Different Storage Makers Tier Data to the Cloud .............................................................. 10
  - Dell EMC ............................................................................................................................... 10
  - HPE ........................................................................................................................................ 11
  - Hitachi Vantara ...................................................................................................................... 11
  - IBM ....................................................................................................................................... 12
  - NetApp ................................................................................................................................. 12
- More on Cloud Tiering ................................................................................................................. 13
- Major Use Cases .......................................................................................................................... 14
- All Tiering Isn’t on the Same Level ............................................................................................ 15
What Is Storage Tiering?

Storage tiering (also called data tiering) describes the ability to use at least two storage technologies with different performance levels but presenting them as a single pool of storage. In this configuration, the different storage technologies are considered tiers, and data can be moved between these tiers for optimal usage and cost. Users accessing the data see a single, unified view.

All storage technologies have different performance levels and are normally grouped by data transfer speed and the type of storage used, whether it’s tape, SAS Disks, SATA Disks, or SAS-based and NVMe-based SSDs. Now, with highly durable cloud object storage technologies such as Amazon S3, Azure Storage Blob, and Google Cloud Storage, cloud storage can also be used as storage tiers.

The higher-performance tier generally runs on more expensive storage, such as SAS-based or NVMe-based SSD, though it will often have less capacity than the colder tiers. When it comes to on-prem machines, the performance tier’s ability to scale capacity is normally limited by the enormous costs of CAPEX spending, physical maintenance and the machine’s limitations. This performance tier stores the most frequently accessed data, which requires low latency (also known as hot data).

The slower tier will use less expensive storage, and normally has a higher capacity due to the lower cost, and stores the infrequently-accessed data (also known as cold data), and/or data where performance is not a priority.

There may be a middle-tier which stores warm data. This tier may be part of the initial design, but it may also be a former performance tier that has been demoted. That situation can happen when a higher performance level requires a new machine, but the machine that was previously being used as the hot tier hasn’t reached the end of its life yet.

For a simple and efficient data tiering operation, the storage system itself must manage the movement of data blocks between the tiers and make it appear seamless. However, data blocks can only reside on one tier at any time; thus, it should not be confused with backups or cache where data is replicated. Overall, storage tiering can provide capacity and performance at the right budget.
Different Forms of Tiering

In two-tier systems, the higher performance tier—where data that is needed frequently is located—is called the hot tier. The lower performance tier that houses data that isn’t in frequent use is called the cold tier. Data is moved between these two tiers based on an algorithm written by the storage vendor which makes the movement completely transparent to users. Three-tier systems may have a hot tier for performance, a warm tier for capacity, and the cold tier for long term storage.

Tier Types:

<table>
<thead>
<tr>
<th>Cold Tier</th>
<th>Hot Tier</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Data not in frequent access</td>
<td>- Data needed on a constant basis</td>
</tr>
<tr>
<td>- Higher latency</td>
<td>- Low latency</td>
</tr>
<tr>
<td>- Lower cost</td>
<td>- High cost</td>
</tr>
</tbody>
</table>

In storage tiering, data is either whole files or blocks of which files are composed, but each piece of data can only exist on a single tier. As opposed to caching and backup data, where data is copied between tiers, storage tiering moves a single copy of the data from one tier to another.

The algorithm has to manage the placement of data on the best tier to provide an overall performance or capacity benefit. A common approach is to move the most frequently accessed data to the hot tier, some algorithms will move data older than a defined age to the cold tier, and if a backup facility is integrated into the storage operating system, the backups could be moved to the cold tier immediately. Data moved to archive storage is still effectively online and will be retrieved by the system when accessed.

The following technologies can be used as storage tiers:

<table>
<thead>
<tr>
<th>Technology</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>NVMe-based SSD</td>
<td>Fastest but most expensive.</td>
</tr>
<tr>
<td>SAS-based SSD</td>
<td>Very fast and quite expensive.</td>
</tr>
<tr>
<td>SAS Disk Drives</td>
<td>More expensive than SATA and uses more power.</td>
</tr>
<tr>
<td>SATA Disk Drives</td>
<td>Considered slower than SAS and has lower MTBF.</td>
</tr>
<tr>
<td>Tape Drives</td>
<td>Inexpensive, but high latency. See below.</td>
</tr>
<tr>
<td>Cloud Storage</td>
<td>Most inexpensive. More below.</td>
</tr>
</tbody>
</table>

There are large variations in the performance and cost of tape and cloud. Tape drives have high sustained transfer rates, but latency can be high due to the time it takes for tape robot loading and unloading. Cloud storage, while extremely inexpensive due to its total lack of physical storage requirements and costs, depends on your cloud connectivity throughput, as well as the cloud provider storage choice.
A Look Back at Tiering History

There’s a long tradition of moving data from one storage system to another to make the best use of the data lifecycle. Some of these were extremely labor intensive as well as hugely expensive. In this section, we’ll take a look at the beginnings of storage tiering, and its evolution into the cloud.

Earlier storage tiering used tape as the capacity tier and disk as the performance tier, with a robot moving tapes between shelves of tape storage and the tape drive(s) when required. Advances in technology have increased performance but tape storage has been kept around as a useful cold tier for a very long time, and is still in use to this day.

Another common example of tiering was using SAS Disks for the hot tier and SATA disks for the cold tier, but advances in SATA technology made performance similar to SAS when used as part of storage infrastructure for sequential workloads in nature. In recent years, SSDs became cost-effective as the hot tier with an extreme level of performance; in this case using SATA as a cold tier became a very cost-effective solution. Eventually, increases in SSD capacity, reliability, and storage efficiency features made it possible to use it as just a single tier. However, the exponential increase in data being generated means the search continues for more ways to optimize storage infrastructure and reduce its associated costs. This is mostly being done by leveraging emerging technologies.

In the next section we’ll dive into where tiering is going now and how NetApp is going to take you there with Cloud Tiering.

Tiering Models

- **Disk to Tape**
  - Slow, high durability, expensive

- **SAS Disks to SATA**
  - Originally ideal, but SATA became as performant as SAS, negating benefits

- **SSD to SATA**
  - Cost effective, but still required on-prem storage

- **SSD to the Cloud**
  - Most cost effective tiering model
How Tiering Should Be Done

With budgets prioritizing Operational Expenditure (OPEX) over Capital Expenditure (CAPEX), and cloud deployment becoming the dominant IT strategy across all industries, the deciding factors in storage architecture design have changed. The data center cold tier’s shelves of spinning disks or the robot that manages the tapes are now simply power, cooling, and floor tile rent. When you add maintenance and media replacement, a newer storage tiering architecture starts to lead—tiering data to the cloud.

Examples of use cases for cloud tiered data include:

Virtualized environments are where backups of virtual machines sit on high-performance storage or are backed up to existing capacity tier; however, its capacity limits retention time for backups and therefore limits resilience. Cloud tiering would automatically move backups to cloud-based storage for unlimited time, increasing resilience, and during restore, data would be moved back to the performance tier for faster recovery.

Data retention for compliance requires planning and purchase of storage which may sit empty until required, which could be a year or longer. Cloud data tiering technology would move the cold data to cloud storage, which is effectively whatever size you need, as more cloud-based storage can simply be added when required, and thus you will only be paying for the amount of storage that you are using.
Meet the Cloud-Based Object Storage Offerings

In this section we’ll look at the most common compatible storage offerings from the cloud vendors. Listed here are the cost of storage and retrieval rate to show the differences in the most basic costs. Note that this table doesn’t include any of the many small costs associated with cloud-based storage, such as network egress and operations/API costs.

Amazon Web Services S3 storage

Amazon Simple Storage Service (Amazon S3) was the initial public cloud storage offering. As such, it is the most widely used low-cost cloud-based object storage. There are four relevant storage classes that data can be tiered to that vary by cost and availability. (Note: The below details are for storage within the US West 2 region - Oregon)

<table>
<thead>
<tr>
<th>Class</th>
<th>Storage $ per GB</th>
<th>Retrieval $ per GB</th>
<th>Minimum Days Charged</th>
</tr>
</thead>
<tbody>
<tr>
<td>S3 Standard</td>
<td>0.023</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>S3 Standard - Infrequent Access</td>
<td>0.0125</td>
<td>0.01</td>
<td>30</td>
</tr>
<tr>
<td>S3 Glacier</td>
<td>0.004</td>
<td>0.01</td>
<td>90</td>
</tr>
<tr>
<td>S3 Glacier Deep Archive</td>
<td>0.00099</td>
<td>0.02</td>
<td>180</td>
</tr>
</tbody>
</table>
Azure Blob Storage

Azure Block Blob storage offers three classes of object storage that vary by cost and the minimum number of days charged for usage (even if deleted).

(Note: The below details are for storage within the US West 2 region - Washington)

<table>
<thead>
<tr>
<th>Class</th>
<th>Storage $ per GB</th>
<th>Retrieval $ per GB</th>
<th>Minimum Days Charged</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hot</td>
<td>0.0184</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>Cool</td>
<td>0.01</td>
<td>0.01</td>
<td>30</td>
</tr>
<tr>
<td>Archive</td>
<td>0.00099</td>
<td>0.02</td>
<td>180</td>
</tr>
</tbody>
</table>

Google Cloud Storage

On Google Cloud, object storage is offered in the form of Google Cloud Storage. There are three classes of Cloud Storage that vary by cost, the minimum number of days charged for usage (even if deleted), and a slight decrease in potential availability.

(Note: The below details are for storage within the US West 1 region - Oregon)

<table>
<thead>
<tr>
<th>Class</th>
<th>Availability</th>
<th>Storage $ per GB</th>
<th>Retrieval $ per GB</th>
<th>Minimum Days Charged</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard</td>
<td>99.9%</td>
<td>0.02</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>Nearline</td>
<td>99.0%</td>
<td>0.01</td>
<td>0.01</td>
<td>30</td>
</tr>
<tr>
<td>Coldline</td>
<td>99.0%</td>
<td>0.004</td>
<td>0.02</td>
<td>90</td>
</tr>
</tbody>
</table>
How Different Storage Makers Tier Data to the Cloud

Storage tiering makes it possible to match data with the most optimal type of storage, based on the usage of that data. Tiering infrequently used data to the cloud is the latest development in storage tiering, and as such many of today’s largest storage vendors offer services for it.

The following is a summary of how the five top storage array vendors tier data to the cloud.

Dell EMC

When Dell bought EMC in 2016, they gained storage hardware and a cloud tiering product along with their existing storage hardware and cloud tiering products. Dell EMC has the following three cloud tiering capable systems:

Isilon Storage Array

- Hardware storage appliance providing NAS storage, including data tiering between Isilon Cluster nodes, called SmartPools.
- CloudPools (A SmartPools extension) allows files to be tiered to cloud storage, leaving 8KB stub files behind.

PowerProtect DD Appliances

- Backup appliances that use the Data Domain Filesystem, which has a two-tier cloud tiering mechanism built in.
- Supports S3 compatible providers of cloud storage such as Virtustream Storage Cloud, Amazon S3, or Dell EMC ECS S3.
- Can also present an iSCSI target of a backup device, allowing tiered cloud storage to function as a virtual tape library.

Unity Storage Arrays

- Unity Arrays use the Cloud Tiering Appliance (CTA) software to tier data to the cloud installed on a physical server or deployed as a virtual machine.
- Performs file-level tiering, leaving 8KB stub files behind.
- Supported cold tiers in CTA are Amazon S3, Azure Blob, IBM Cloud Object Storage, Dell EMC Virtustream and ECS.
- Block tiering (known as block archiving in Dell EMC terminology) can be performed to archive full copy of LUN, Consistency Group or Thin Clone and subsequent snapshots.
HPE has two products that can have their data tiered to cloud storage, together they cover most of the HPE storage hardware range.

**HPE 3PAR and Nimble Storage Systems**

- Secondary data copies can be tiered to the cloud using HPE’s StoreOnce Data Protection Backup Appliances together with HPE Cloud Bank Storage.
- HPE Cloud Bank Storage can tier backup/DR data to Amazon S3, Microsoft Azure Blob storage, and to the Scality object storage solution.
- To optimize the usage of cloud storage it includes change block tracking and data deduplication to reduce the amount of data migrated to and from cloud storage.

**Hitachi Vantara**

**Hitachi Content Platform**

- The Hitachi Content Platform (HCP) uses adaptive cloud tiering (ACT) functionality, allowing hybrid storage pools to leverage storage from third-party sources on-site or off-site.
- Each HCP node will contain some internal storage and can be connected to off-site public cloud resources and tier data into Amazon S3, Azure Blob, Google Cloud Storage, or other S3 compatible storage.
- For remote offices that require tiering data to cloud storage, the Hitachi Anywhere Edge Device can be used to tier data to the HCP cluster, which can then be tiered to cloud storage.

**Hitachi NAS Platform (HNAS) and VSP N series**

- These systems come with an intelligent file tiering feature allowing migration of data to private cloud object storage based on HCP or public cloud object storage.
- Support for Amazon S3, Azure Blob, and IBM Cloud Object Storage.
- Tiering to the cloud is done at the file level and it is a policy-based process that runs on a predefined schedule.
• IBM Spectrum Virtualize pools storage from multiple storage systems, even from multiple vendors, allowing data tiering across heterogeneous storage environments.
• Tiering is done using the IBM Easy Tier function that supports up to three tiers and moves data extents (blocks) to the appropriate tier automatically.
• However, Easy Tier isn’t a true cloud-based solution, since it doesn’t tier to cloud storage directly. It replicates to IBM Spectrum Virtualize for Public Cloud instances running in AWS or IBM Cloud.
• IBM Spectrum Protect can tier to cloud storage on AWS or Azure to back up many different types of enterprise environments, from applications to virtual hosts, to databases.

NetApp

NetApp All Flash FAS (AFF) and FAS Hybrid Arrays running NetApp’s ONTAP data management software can tier data to a cloud-based object storage via the Cloud Tiering service which is based on NetApp FabricPool technology.

Tiering is completely automated and offers users three policies for tiering data to the cloud:

• Auto Tiering Policy: Blocks that have been cold for a predefined period are moved to the cloud tier.
• Snapshot-Only Tiering Policy: Snapshot blocks are moved to the cloud tier after a default or specified cooling period.
• All Tiering Policy: Moves all volume data blocks to the cloud, normally on volumes that contain finished projects, historical data, backups or archive data.

When blocks that have been tiered are accessed, only those 4KB blocks are returned, not the whole 4MB object. This ensures both latency and costs are minimized. Cloud Tiering supports tiering cold data to Amazon S3, Azure Blob storage, and Google Cloud Storage.

There are a few major advantages to Cloud Tiering with NetApp than the solutions mentioned above. Data tiering is performed at the block level, without any “stubs” or “smart links” left behind. That means instead of entire files being migrated to an external cloud tier, only the blocks marked cold are moved, cutting down on the amount of cloud storage that is used. All of this is done while preserving the existing namespace, making it transparent to applications when accessing their data.

Cloud Tiering is also designed to handle both primary and secondary copies of data, so ONTAP-based systems can be utilized more efficiently, allowing you to reclaim performant storage space to be used by performance-sensitive applications and have a cost-efficient secondary storage. Additionally, Cloud Tiering allows the on-prem machines to tier data directly to the cloud, without the data having to be moved to another level before it ever gets to the cloud.

Cloud Tiering is an excellent way to manage the data lifecycle, and lower the costs of hosting infrequently used data in the data center.
More on Cloud Tiering

NetApp Cloud Tiering service is an almost hidden process, with no manual work once the initial configuration has been completed. Data is automatically moved out to the cloud when it is no longer actively being used and moved back when it is required. Data tiered to the cloud using NetApp Cloud Tiering service is encrypted in transit and at rest and thus your data is secured end-to-end.

With more space freed up on your AFF system, that means there is less need to roll in a new machine. Using Cloud Tiering is considered part of a hybrid-cloud strategy even though you haven’t decommissioned any part of your data center or had to make any changes to your applications. Cloud tiering can really improve your data lifecycle management.
Major Use Cases

Delay CAPEX Spending for New Systems
Put off making data center refresh purchases by optimizing your current data usage and keeping only frequently-used data on-prem.

Avoid Adding Disk Shelves to Existing Systems
Bypass issues of compatibility, interoperability, capacity sizing, and added costs that come with expanding existing storage systems.

An Existing AFF Running Out of Space
Leverage the cloud to make room on systems that are too new for a refresh or reached their physical limit without limiting performance.

Reducing Data Center Costs
Sunset older machines that may be dedicated solely for storing cold data such as snapshot versions and completed projects.

Avoid the Costs and Time Constraints of Using Tape or SATA
Reduce the amount of time it takes to manage data at the end of its lifecycle.

Take a First Step into the Cloud
Enable a hybrid deployment with the click of a button—without making any changes to your existing applications or processes.

Lower DR or Backup Site Costs
Tier entire volumes to the cloud instead of maintaining costly off-site locations for cold data storage such as DR and backup.

Snapshot Copies and Backups
Increase storage utilization by offloading important snapshot data to the cloud.

Prevent Cloud Vendor Lock-In
Enable storage tiering in more than one cloud to avoid vendor lock-in, a major benefit to a multicloud strategy.

Lower Compliance Costs
Spend less housing data that is required to be stored under regulations such as HIPAA and GDPR.
All Tiering Isn’t on the Same Level

Storage or data tiering is a way to intelligently manage data growth and be cost-effective, as you only move cold data to a less performant, less expensive storage tier. However, the first thing to notice is that there are different interpretations of tiering by different vendors.

NetApp Cloud Tiering service is a key part of NetApp’s Data Fabric solutions, that allows customers benefit from the high performant on-premises storage platforms such as NetApp AFF and SSD-backed FAS systems, while leveraging the low costs and high durability of cloud-based object storage such as Azure Blob, Amazon S3, and Google Cloud Storage.

NetApp Cloud Tiering optimizes storage tiering by only moving infrequently-accessed data blocks directly to the cloud tier (and back to the performance tier when requested) for both primary and secondary workloads providing a single fixed access point for all applications, so unneeded data is always stored cost efficiently.

For organizations that are just beginning to work on a digital transformation, Cloud Tiering can offer an easy, low-cost first step into the cloud.

Get a higher level of tiering

Sign up for a free trial of the Cloud Tiering service
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 Information
Copyright © 1994–2020 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.

RESTRICTED RIGHTS LEGEND: Use, duplication, or disclosure by the government is subject to restrictions as set forth in subparagraph (c)(1)(ii) of the Rights in Technical Data and Computer Software clause at DFARS 252.277-7103 (October 1988) and FAR 52.227-19 (June 1987).

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.

NA-287-0218