



## **Six Ways to Implement StorCycle in a Modern Storage Workflow**

## A Modern Storage Paradigm for Managing and Storing Data

We live in a new world, and that is largely due to our ability to use data to enhance our world. It's interesting to look at data growth numbers for social media, but that's *not* at the heart of data growth for most organizations. Yet we are all finding new ways to use the data we collect or create to further organizational goals. In fact, recent studies show that 97.2% of organizations are investing in big data and artificial intelligence.<sup>1</sup>

Data growth is not the problem, it's the opportunity. That is of course provided your organization is prepared to deal with this growth. Storage of information is no longer an afterthought. If it is, chances are that you're falling into a statistic of which you don't want to be a part of – as much as 80 percent of enterprise data may qualify as “inactive,” yet it resides on primary storage.<sup>2</sup>

Traditional storage paradigms have focused too much on the storage “tier” without offering insight into the *data* that is being stored on any given tier. If inactive data is being stored on the wrong tier of storage, it could end up costing organizations millions of dollars a year as well as limiting their ability to monetize and share this data properly. Ideally, the “knowledge” we need starts with knowledge of the data we already have. That's the missing piece in traditional storage paradigms. Spectra solves that problem with the introduction of StorCycle.

StorCycle, Spectra's groundbreaking storage lifecycle management software, enables a modern storage paradigm based on a two-tiered storage model. Rather than focusing exclusively on the storage medium, this model is based on the data or digital content that is actually being stored. We start by classifying data into two categories – “Active,” meaning it's being edited, processed or changed in some way, and “Inactive” which quite simply refers to everything else. This results in a *Primary Tier* for the active data and a *Perpetual Tier* for inactive data.

The *Primary Tier* holds all active data and is most commonly composed of flash, NVMe and high-performance disk drives. By moving inactive data out of the Primary Tier and into the Perpetual Tier, organizations can significantly decrease the size of the Primary Tier. This allows administrators to better configure this tier using a combination of high-speed storage mediums in order to achieve the performance required for workflows associated with highly active data.

The *Perpetual Tier* is dedicated to inactive data and is designed to keep multiple copies of data on multiple storage mediums including NAS, object storage disk, cloud and tape. While the data is not considered “active” on the Perpetual Tier, there is quite a bit happening at this level. The Perpetual Tier is used for secondary storage; distribution; multiple copies (a responsive copy and DR copy); backup; archive; project archive; and traditional disaster recovery.



*Two-Tier Storage Model*



StorCycle allows organizations to configure the more economical Perpetual Tier to be as responsive as their workflows demand. Administrators can create “responsive copies” on low-cost NAS or another local format for data requiring quick but infrequent access. This copy can be used for data accessed directly by machine or application or may be used as a distribution copy. Simultaneously, StorCycle can create an identical copy on tape, for true air-gap protection from ransomware; a copy to cloud, for offsite DR; an additional copy to tape to avoid cloud restoration charges; or any combination thereof.

## The Promise of a Perpetual Tier

Enabling a Perpetual Tier of storage allows organizations to address storage problems of the past and present as well as create an environment open to future growth, development and change.

A properly implemented Perpetual Tier has significant benefits for the Primary Tier as well. A smaller Primary Tier reduces primary storage costs in hardware, software, and storage licensing by reducing storage requirements from primary storage systems. A reduction in expensive primary storage will also lead to less administrative and maintenance costs for the primary storage support, and allow IT administrators to be more productive in organizing and managing their infrastructure.

The following six use cases give more granular insight into how the “Perpetual Tier” approach reduces costs associated with the Primary Tier and ways to implement StorCycle in a modern storage workflow.

### Midsize Corporate IT Department Upgrading Current Primary Tier Storage to an All Flash Array

One general IT shop we’re working with is in the process of upgrading the performance of their primary storage tier. They have targeted the *Nimble All Flash Array A300* to replace their current HDD infrastructure. They have roughly 500TB of primary storage. They originally decided to forego the upgrade due to cost. The 500TB solution they selected had a list price of \$3,846,000. With significant discounts, this solution would be roughly \$1.1M or \$2,200/TB.

They estimate 60 percent of their data currently stored on the Primary Tier is inactive and can, therefore, be moved to the Perpetual Tier. The majority of that migrated data (300TB) will reside on low-cost NAS, roughly 7¢/GB or \$70/TB. Cost for the Perpetual Tier NAS storage will be roughly \$21,000. They now estimate the cost of creating a 200TB flash-only, Primary Tier platform for roughly \$440,000. The total savings from this simple application of a Perpetual Tier is close to \$600,000 including the cost of StorCycle. But more importantly, they are able to significantly increase the performance of their data center and work with state-of-art technology, and users have seamless access to migrated data on the NAS tier of perpetual storage.

StorCycle is causing organizations to rethink the most effective way to store their data. By removing inactive and “low transaction” data from the Primary Tier, an organization can now more affordably upgrade to a high-performance Primary Tier that uses SSD, NVMe flash or other cutting-edge high-performance technologies.



## Midsized Corporate IT Department with Exponential Data Growth on Primary Tier Storage

This group is similar in size to the above IT department, but their challenges and goals are different. They currently run 600TB on their primary storage tier, but are increasing Primary Tier storage by roughly 40 percent per year. They use high-performance HDD in the Primary Tier. The performance is sufficient, but adding 240TB per year (at a discounted price of roughly \$850,000) doesn't fit their budget. They are especially concerned by the compounded growth rate for following years. They saw their only option as implementing severe storage quotas to departments, which is not in line with other business growth mandates.

With StorCycle Storage Lifecycle Management Software, they will be able to move an estimated 70 percent of data deemed "cold data" to the Perpetual Tier, consisting of a combination of NAS and tape. Data not accessed in at least 30 days will go to NAS, and any data not accessed in over six months will be moved to tape. Total cost of implementation will run well under \$200,000.

The upside is significant in this case. Instead of adding an additional 240TB of high-speed HDD, they now project being able to hold their primary storage growth level for 18 to 24 months given the amount of current storage being moved to the Perpetual Tier. Their users are also seeing an upside – they have actually *increased* storage quotas for NAS and are now able to completely eliminate storage quotas for anything moved to tape – where users will still have seamless access to that data – for roughly 2¢/GB.

As data continues to grow, the Primary Tier of storage can be seamlessly extended to other tiers allowing a user to easily restore any file (or directory) that exists on low-cost storage from their local work machine. This means less future growth of the expensive Primary Tier.

Without StorCycle, both of the above projects would involve adding headcount and management. StorCycle, in effect, acts as an additional storage administrator to an IT Team.

## Small Corporate IT Department Plans on Moving Data to the Cloud

This use case doesn't involve new hardware. This IT group has roughly 12TB of data on Primary Tier and is growing at 25 percent per year. Given the relatively small amount of data, and the limited resources they have available, they are moving to the cloud as a backup target.

Their backup software licensing is capacity-based with identical charges for any given storage target – NAS, tape or cloud. Their original software licensing to back up the 12TB to NAS ran around \$30,000 with a 20 percent per year support charge. They pay an additional \$2,000/TB for any additional data backup over the original 12TB license.

In addition to current yearly support charges, at a data growth rate of 25 percent per year, they expect to pay an additional \$6,000 in software licenses in year 2 for new data to be backed up as well as \$1,200 in additional support.

They have conservatively estimated that 50 percent of their Primary Tier data (6TB) is inactive and can be moved to the cloud via StorCycle. The *Administrator License* of StorCycle licensing model allows for up to 25TB to be moved to the cloud for no additional charge, so they will



have no capacity charges to worry about. The responsive tier of their StorCycle archive will be kept in AWS Infrequent Access storage for recall (6TB @ \$900/yr), and the DR tier in AWS Deep Glacier for (6TB @ \$71/yr). This now provides two copies for redundancy at a cloud storage cost of roughly \$1000/year. The remaining 6TB of active data will still be backed up on a daily basis. Backup windows will be cut by half.

By deploying StorCycle to move inactive data to the cloud, they will eliminate the need to increase backup licenses year over year, saving the above mentioned \$7,200 in the first year. They will have to invest \$12,000 in a StorCycle license, \$1,800 in StorCycle support and \$1,000 for the AWS cloud storage. By end of year two, as the data continues to grow, no additional StorCycle licensing and fixed cost on support means there will be an actual cost savings of \$7,400. By end of year three, they will have an additional cost savings of \$11,052. This represents a total cost savings over the three-year period of \$10,852 – in addition to having recovered all costs for StorCycle. Extra savings will also be created by the reduction in new data residing on Primary Tier storage as StorCycle is continuously used to move inactive data to the Perpetual Tier of storage. The organization will have a more manageable Primary Tier storage infrastructure and a more fully implemented cloud archive strategy.

As the size of an organization's primary storage is reduced, daily, weekly and monthly backups or replication snapshots are also smaller, leading to shorter backup windows and reduced backup storage costs. Likewise, new cloud strategies are both easier and less costly to implement.

## **Project Archive within the Perpetual Storage Tier**

As we continue to look at use case examples, it is important to keep in mind that the Perpetual Tier of storage is not limited to *older data*. The above use cases focus on moving data into the Perpetual Tier based on age and last access date. The Perpetual Tier should also serve as an archive tier for large data sets that may be moved immediately after creation or collection. By allowing users to archive recent or even older "project based" files or directories, historically critical data can be maintained and protected indefinitely.

Any of the above use cases could also choose to use StorCycle's unique Project Archive feature, but we tend to think of Project Archive more along the lines of preserving large data sets generated by data-intensive organizations that generate and manage project data as part of their mission. Universities, government agencies, genomics, research labs, media & entertainment organizations and associated post-production houses, weather research and forecasting are just a small sampling of the many organizations creating enormous amounts of data on an ongoing basis that needs to be managed outside the confines of high-performance, Primary Tier storage.

The following use cases are as varied as the organizations deploying project archives, but there are commonalities which allow any organization to gain insight into better ways of managing and archiving large data sets be they application, human or machine generated.



## A Large University Supporting Multiple Research Projects

One of the data centers we work with is part of a large university that supports the research efforts of over 50 different groups within the university. They offer various service level agreements (SLAs) based on the performance of the storage. Each research group is billed for the amount of storage they use based on the SLA they select.

The university has standardized on three storage performance levels: an SSD-based tier; high-speed, disk-based tier; and a NAS-based tier targeted for archive of projects. There are multiple challenges for both the university and the individual researchers that the university would like to overcome.

While NAS is the lowest cost repository in the current storage model, archiving large amounts of fixed content becomes extremely expensive over time. Researchers have asked for a lower cost solution for archiving. The university would like to introduce tape storage as part of the Perpetual Tier for archiving, but they have no way to introduce “rule-based” file movement across the storage infrastructure to a Perpetual Tier of tape.

The university actively encourages the researchers to move data off of the high-speed Primary Tier. The university actively encourages the researchers to move data off of the high-speed Primary Tier with a bill-back system that offsets their costs. However, when the university runs out of high-speed storage, they don't always have the funds to acquire more before the offsets come in, which can hamper research efforts.

The researchers would like to move their data to a low-cost storage tier, but they have a challenge to accomplish this. The data can be human generated, application generated or machine generated, and it has been accumulating for years. The researchers have access to the data, but they have no way to identify what is actively used, what needs to be archived, and what data is orphaned. And if the data is manually moved, how can it be accessed after the researcher leaves the university? Grants for research often require storage of the data for periods longer than the researchers work for the lab or university involved.

StorCycle is a perfect solution for both parties. StorCycle offers a seamless view across all of the storage it manages. Both the university and the researchers can have access across all data on the Perpetual Tier.

Researchers can use the scanning capabilities of StorCycle to identify and target inactive data sets for archive. StorCycle's Project Archive will assure this situation doesn't reoccur. Project Archive allows users to identify any files or directories associated with a project and archive them as a group. This can be done immediately after a large project is completed.

Archived data sets can be tagged with additional information to identify anything of importance to the project, be it grants associated with the project, researchers involved, project names, etc. This metadata can easily be searched at any point in the future. Likewise, StorCycle produces a manifest for each project archive which can be accessed as a digital file. The manifest shows exactly what was moved, where the data originated, where it was moved to, and when it was moved. It can be digitally displayed by clicking on the finished project archive and stored with other files in the project. It does not require a query into the database, and can be worked into existing workflows.



The university will now be able to deploy an affordable tape storage specifically for archiving. The cost to the university for automated tape will be around 2.5 cents/GB (US Dollars) and additional copies on tape will be around 1.2 cents/GB. Researchers will be able to identify data and move it to the Perpetual Tier, and that migrated data will be protected and easily accessible into perpetuity. All of this relieves pressure on the most expensive Primary Tier storage, which the university is responsible for purchasing and maintaining.

## **A Post-Production Studio in Media & Entertainment Supporting Advertising Agencies**

Movies, television series, reality TV and commercials – They seem relatively straightforward when we watch them, but what goes on behind the scenes is tremendously complex. A lot goes into creating a simple commercial – film ingest, QC, logging, audio synch, creating a proxy or mezzanine, editing, rendering, adding special effects, dubbing – it’s all done on high-speed, high-cost disk storage.

In this manner, a single commercial is typically 50 to 100TB. That’s the output; all of the input (think 1200:1 ratio) is kept as well. On average 10 to 15 final versions of the commercial will be presented for review and selection by the ad agency. Here’s where the numbers really become staggering – A single post-production house will typically support multiple advertising agencies and a single advertising agencies will support multiple clients. A single client can easily run four to eight campaigns a year, each requiring multiple commercials, which easily creates 500TB to 1PB per client per year.

Unlike large movie or television studios, many post-production studios don’t use Media Asset or Production Asset Management software for archiving content. Those solutions are often cost prohibitive. Spectra works with many post-production studios.

This particular post-production studio will use StorCycle’s Project Archive feature to solve this problem. All digital assets will be stored in directories associated with a given client or commercial. As soon as the commercial is completed, all associated directories and files will be archived off of their Primary Tier of storage to their Perpetual Tier of storage via StorCycle Project Archive. Additional metadata will be added for more granular searches in the future such as type of commercial (seasonal/geography/length/cost/end user client/ad agency client, etc.).

Due to the volume of data created in each case, the content will all be archived to tape via StorCycle and the fully integrated Spectra BlackPearl Converged Storage solution. This allows them to manage the content on their newly created Perpetual Tier with the advantages of object storage even when written to tape. Future migration is seamless, floor-life of a tape library is 10 to 12 years vs. three to five years for disk, and additional copies for vault storage or sending back to the ad agency will run under 2¢/GB.

Instead of spending roughly 50¢/GB for high performance disk, the content can be archived to tape for roughly 3.8¢/GB (including cost for BlackPearl). Given those numbers, that takes the cost of storage from \$500,000 per PB to \$38,000 per PB. The *solution* is now three-fold: They can afford to keep all data associated with an ad agency or client indefinitely; they can seamlessly move it to a Perpetual Tier of storage and easily bring it back when needed; and StorCycle’s Project Archive feature assures all content will be available and searchable as a single project.



## U.S. Governmental Research Facility Utilizing Project Archive

Virtually any governmental agency, in any country across the globe, deals with large amounts of data. This particular agency creates, collects and distributes scientific information used by other governmental offices, both U.S. and internationally; non-governmental agencies; other researchers; and even individual citizens.

Data gathered can be generated by application output, field sensors, machines, cameras, individuals, or other data creation methods. After data is gathered, it is further analyzed, categorized or simply stored for possible future use or reference. No data gathered is considered “disposable”. As technology, science and even the earth itself evolve, new exploration often draws on historical data – weather patterns, ocean currents, agricultural yields, and mineral exploration, to name a few.

After extensive search for a storage lifecycle management software application, this particular agency was drawn to StorCycle specifically for its ability to archive data based on the data’s association with a given project. How else could multiple forms of data from multiple types of data generators be collected and archived if not through some form of “project identification”?

But they still had a challenge. Machine-generated data accounts for much of the data they collect. Examples of this are data from sensors that may detect physical phenomena such as light and sound and turn it into a data stream, or calculations from algorithms predicting risk of earth movement based on other seismic data sets. The output may not be analyzed immediately, or the researcher may deem it unnecessary for a current project, but they wish to keep it for future reference. Most of the machine-generated data they receive initially requires high-speed disk as a landing zone. The researchers had no way to move it to lower cost storage and bring it back when needed. Therefore, it stayed on the Primary Tier of storage – at great expense – even if it was never accessed.

A combination of StorCycle, Spectra’s BlackPearl Converged Storage System, and a Spectra T950 Tape Library make perfect sense. By setting up an Archive Directory with StorCycle, even machine data can be immediately archived as it comes in – using high-speed disk storage for ingest, but immediately moving it to a lower performance storage tier. It’s then deleted from the primary storage.

As the output storage target for StorCycle, a flash-based BlackPearl Converged Storage System can not only ingest the archived data at great speeds, but it can also direct it to the tape library at great speeds – easily streaming 12 or more LTO tape drives simultaneously. The final archive tier will be the Spectra T950, which can hold over 11PB of uncompressed data in the footprint of a single rack and expand to hold over 120PB of uncompressed data via expansion frames. It’s as if the high speed Primary Tier of storage is extended indefinitely at pennies per gigabyte.

Most importantly, the Archive Directory can be associated with the project which created the data so that the data captured can be seamlessly tracked as part of the larger project. Individual researchers can designate the storage layer for data or content, but the ability to query the StorCycle database means other individuals, not originally associated with the research, can find it throughout its lifecycle.

Other data sets within the agency are posted to the cloud for distribution or sharing. StorCycle’s ability to migrate to multiple targets allows the agency to direct data to the same long-term, tape storage tier –



which is not externally accessible – as well as direct a copy to the cloud for distribution or sharing. The cloud copy can expire in a year or two while the DR copy will remain on tape into perpetuity.

All of the above examples for Project Archive require a Perpetual Tier that is easily accessed, easily searchable by project name or other tagged metadata, and built for “forever” retention.

StorCycle is the only data management software package that offers Project Archive, making it an ideal solution for preserving large data sets generated by data-intensive organizations that generate and manage project data as part of their mission.

These are but a few of the ways in which StorCycle is making a simple, two-tiered storage paradigm a reality for organizations working with a few terabytes up to those working with hundreds of petabytes and beyond.

## Summary:

If we inherently knew that all data resided on the appropriate level of storage, storage quotas could be a thing of the past, storage budgets could be more accurately forecasted and managed, and the benefits of any new storage medium could be easily implemented without an overhaul to the existing storage infrastructure. In this way, efficiencies could be maximized, data could be utilized to create further value and storage costs could actually go down instead of up... even when adding new technology.

Arguments over “end point” storage solutions – disk vs. tape, public cloud vs. private cloud, file vs. object – have consumed too much of the storage conversation and have deterred organizations from being able to focus on the real point behind storage – meeting the desired organizational goals that the information/data/content is used for.

Drawing from our 40 years of experience in the storage industry, Spectra is excited to introduce StorCycle storage lifecycle management software, and a new, two-tiered paradigm for storage that enables data to reside on the appropriate level of storage. This means more storage, lower costs, greater access, enhanced protection and fewer silos.

## Footnotes:

1. <https://techjury.net/stats-about/big-data-statistics/>
2. <https://storageswiss.com/2019/04/12/inactive-data-how-much-do-you-really-have/>



## About Spectra Logic Corporation

Spectra Logic develops data storage and data management solutions that solve the problem of long-term digital preservation for organizations dealing with exponential data growth. Dedicated solely to storage innovation for over 40 years, Spectra Logic's uncompromising product and customer focus is proven by the adoption of its solutions by leaders in multiple industries globally. Spectra enables affordable, multi-decade data storage and access by creating new methods of managing information in all forms of storage — including archive, backup, cold storage, private cloud and public cloud.

To learn more, visit [www.SpectraLogic.com](http://www.SpectraLogic.com) or contact our sales staff at [sales@spectralogic.com](mailto:sales@spectralogic.com)

Copyright ©2020 Spectra Logic Corporation. All rights reserved worldwide. Spectra and Spectra Logic are registered trademarks of Spectra Logic. All other trademarks and registered trademarks are property of their respective owners. This document is provided for information purposes only, and the contents hereof are subject to change without notice. This document is not warranted to be error-free, nor subject to any other warranties or conditions, whether expressed orally or implied in law. We specifically disclaim any liability with respect to this document, and no contractual obligations are formed either directly or indirectly by this document. All opinions in this white paper are those of Spectra Logic and are based on information from various industry reports, news reports and customer interviews.

303-449-6400 • 800-833-1132 • 6285 Lookout Road • Boulder, CO 80301 USA • [spectralogic.com](http://spectralogic.com)