The Modern Storage Model

Perpetual Tier (Multiple copies, secondary storage distribution, backup, archive, DR)

OBJECT STORE • DISK • TAPE • CLOUD

Primary Tier
Solid State/Flash Memory
(Creation, editing, production and processing)

SPECTRA

StorCycle™

Breaking Through Data Lifecycle Management Barriers

HORISON Information Strategies

Horison Information Strategies
Fred Moore, President
www.horison.com

Introduction

If data is the new currency, then storage is the new bank. IT staffs are under increasing pressure as the volume and complexity of managing daily workloads defy the traditional approach of simply adding costlier flash and disk drives when capacity is maxed out. Too often most of this data is stored on the most expensive storage tiers. Lower cost storage options are available, such as archival disk, high-capacity tape, and the cloud, which are the optimal solutions for infrequently accessed and cold data. Traditional storage management techniques have left data centers struggling with as much as 80% of their data being stored on the *wrong tier* of storage costing organizations millions of dollars a year.

Storage TCO analysis are readily available and reveal the expected cost savings if the data is stored on the *right tier*. For example, a [recent TCO study](#) indicated disk storage TCO was ~2.45 times higher than an all disk cloud and ~7.16 times higher than LTO-8 tape for archiving and is a revealing representation of how expensive storing the wrong data in the wrong place can be. From 2011-2020, data was expected to grow ~50 times while the number of trained IT professionals was projected to grow just ~1.5 times. The storage management challenge is undeniable and a new management architecture, breaking from traditional methods, is urgently needed.

Importance of Digital Preservation – The Value of Data

Digital Preservation is the management, maintenance and safeguarding of data (files and objects) so they can be accessed by future users. Organizations routinely plan to preserve a large amount of data in some digital format for 50 years or more and a growing amount of this archival data will never be modified or deleted. The world now produces ~2.5 exabytes ($1x10^{18}$) of data a day and much of it will be preserved with hopes of someday capitalizing on its potential value. Corporations have been the greatest
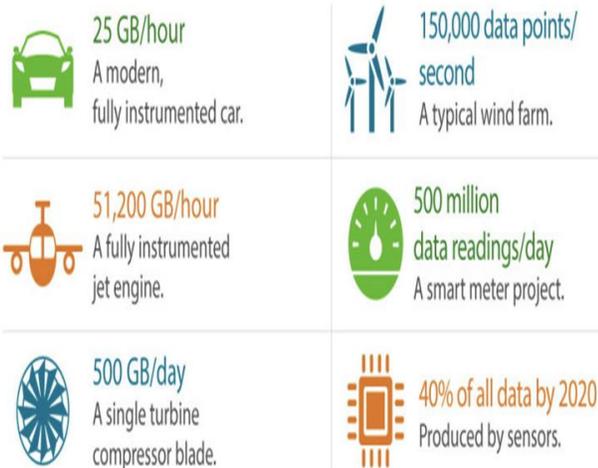
beneficiaries of this data revolution. In 2006, oil and energy companies dominated the list of the top six most valuable firms in the world, but today the list is dominated by firms based on digital content like Alphabet (Google), Apple, Amazon, Facebook and Microsoft. It can be said that digital data is becoming the newest type of intellectual property (IP) and must be protected as such. There is no one rule to determine the value of data, but the Equifax data breach gives us some indication of how data is valued. Equifax, a U.S.- based consumer credit reporting agency that collects and aggregates information on over 800 million individual consumers and more than 88 million businesses worldwide, suffered a data breach in 2017 of 143 million users. As a result, they're facing a class-action lawsuit of up to $70 billion, representing the perceived value of the data at risk.

The Archive Datasphere Is Getting Much Larger – IOT and Sensor Data Take Off
Even as data centers struggle with the enormous growth of disk farms, which are devouring IT budgets and overcrowding data centers, many continue to maintain expensive disks -- often half full of data which has not been accessed or used for years. Archival data is the largest and fastest growing data classification segment. Most data typically reach archival status in 90 days or less after its creation, and archival data is accumulating at over 50% compounded annually as the total volume of archival data has surpassed the amount of backup data. Few data centers can afford to sustain this degree of inefficiency and continue to use traditional storage management software which is very expensive and has become a cost multiplier vs. a cost reducer because it hasn't achieved the goal for optimizing data center storage.

Applications driving archival storage requirements include compliance data, GDPR, medical records, photos and images, email history, unstructured file data, scientific, video, movies, audio, documents, surveillance, collaboration, social media history, archive cloud applications, security system history and archives, media and entertainment, and business continuity and disaster recovery (DR). The IoT, fueled with the arrival of 5G networks that are ~20x faster than 4G, promises to generate massive floods of data from billions of sensors. Many companies have spent years collecting data that piles up in archives and they still haven't figured out what to do with it. These barriers will only become greater without major data lifecycle management breakthroughs.

Data Produced by IoT Devices

25 GB/hour — A modern, fully instrumented car.

150,000 data points/second — A typical wind farm.

51,200 GB/hour — A fully instrumented jet engine.

500 million data readings/day — A smart meter project.

500 GB/day — A single turbine compressor blade.

40% of all data by 2020 — Produced by sensors.

Source: Simafore, RTInsights, Cisco

Mordor Intelligence

By 2025 IDC predicts that 49% of the world's stored data will reside in public cloud environments mandating that businesses understand cloud storage parameters carefully. Object storage is the basis for cloud-based protocols and designed for scale, unstructured data, and potentially good for analytics having rich metadata; however, it lacks high-end performance and robust data protection. In addition,

unstructured (hard to search) data will account for approximately 80% of all new data created. Unfortunately, much of this data also resides in the *wrong place*.

Clean Up the Backup Profile

Backup (make a copy) and archive (move the data) are not the same. Including archive data in the backup set profile is a costly process and becomes more so as the amount of backup data increases. There's no point in repeatedly backing up unchanged and archival data as this lengthens the time and the amount of data to manage in the backup cycle. Archiving removes unchanged data from the backup set to speed up the backup (and restore) process and free up costly disk capacity in the process. Though disk backup processes using compression or deduplication can help, the growing length of backup windows remains under constant pressure as data growth appears to be endless. It's time to clean it up.

Current Challenges for Primary Storage

- IT Budgets are not scaling as fast as data generation and data creation
- Tier 0 flash and tier 1 HDD capacity can become extremely expensive (use it wisely)
- Degraded performance and inefficiency on primary storage
- Aging, inactive, and unstructured data clogging up tier 0 and tier 1
- Backing up all data is not cost effective and is time consuming

Traditional Technology Tiers

The traditional tiered storage model presents a technology-centric view and is delineated by differences in four primary attributes: price, performance, capacity and functionality. As storage pools grow, an effective automated tiered storage becomes the optimal storage architecture since 1) manual data movement is very time consuming; 2) the amount of digital data is continually increasing; and 3) limited staff resources leave storage administrators stretched too thin.

Traditional Technology Hierarchy and Tiers

Percent of WW Digital Data if Optimally Stored by Tier In 2023.
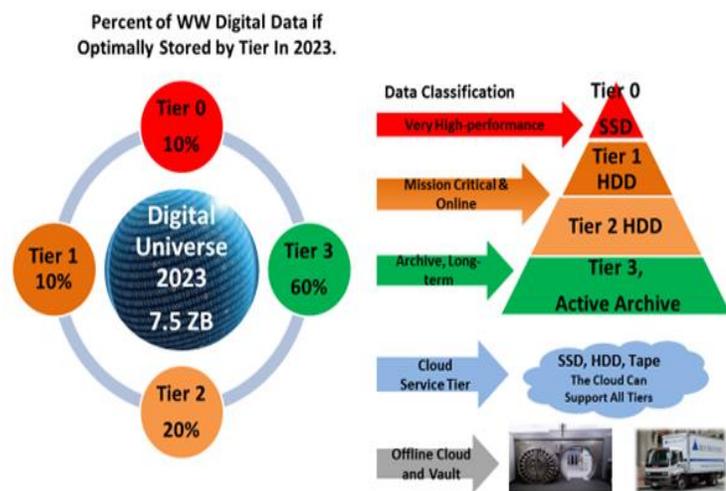
Source: Horison, Inc. The Digital Universe IDC

The optimally stored allocations shown by each tier above are rarely achieved. Many management packages are available to automate the traditional technology tiers; however, many data centers are left with as much as 80% of the data remaining on costly tier 0 and tier 1 storage -- the wrong place. Additionally, the escalating cost and licensing models of storage management software often exceed the cost of the storage being managed. Clearly a new approach is long overdue.

StorCycle – Introducing the Next Storage Architecture

With as much as 80% of data being stored on the wrong tier of storage, the storage industry hasn't delivered tools that can optimally manage data -- until now. Developed by Spectra Logic, StorCycle™ has answered the call to deliver a new, modern storage model to address data lifecycle workflow that reduces overall storage costs by up to 70% while *finally* enabling data centers to get the right data in the right place.

Key Aspects of StorCycle Include:

- Low cost and simple software application to offload less-active files from tier 0 and tier 1
- Find, manage, organize, and access all unstructured data with a non-proprietary data format
- Identify and migrate data to lower cost tiers; and easily store and retrieve migrated files
- Straightforward data analytics for the data that has been migrated.
- Complete integration with email and ADFS
- Priced by the number of users
- One-time permanent or annual software subscription options

Spectra's StorCycle is a storage management software application that looks at the usage of data rather than the underlying hardware technology to manage data through its lifetime. Unlike the traditional four-tier technology hierarchy, this new model includes a two-tier model, consisting of a Primary Tier and a Perpetual Tier, a new tier as shown in the adjacent diagram.

Spectra StorCycle scans data on primary and high-cost storage tiers and identifies infrequently accessed data which typically consumes a significant amount of capacity. StorCycle then automatically migrates data based on file parameters, such as file size, file type, last accessed date, last modified date, and location, to lower-cost, long-term storage in the Perpetual Storage Tier, while enabling the migrated data to be easily accessible to users.

Modern Storage Model

The **Primary Tier** holds all active data and is comprised of flash and high-performance disk drives to achieve the performance required for workflows around highly active or response time critical data. By removing inactive data from the primary storage tier, an organization can affordably upgrade to a high-performance primary storage tier that uses SSD, NVMe, or new memory-based high-performance technologies as they arrive.

The **Perpetual Tier** is dedicated to inactive data and is designed to keep numerous copies of data on less costly storage mediums including cloud, NAS disk, object storage disk, and tape. The Perpetual Tier is used for secondary storage, distribution, backup, archive and disaster recovery. By moving data, including

finished projects, off the Primary Tier, which is dedicated to high performance, astute organizations can easily configure the more economical Perpetual Tier to be as responsive as their workflows demand while enjoying significant cost savings. With StorCycle, there are two basic methods for migrating data – Project Archive and Auto Migrate. The Project Archive method is for completed data sets such as research data, raw video, semiconductor designs, completed digital projects, etc. The Auto Migrate method identifies inactive files based on age and size policies and moves those files to the Perpetual Tier. Migrating infrequently accessed data to the Perpetual Tier yields the greatest benefits, while enabling the data to be accessible between the two tiers as required.

Optimizing the Perpetual Tier

The creation, storage, and protection of digital data has been ongoing for decades; however, the idea of retaining data for indefinite periods in anticipation of discovering unknown value is a relatively new concept. Therefore, economic and archival value of optimizing the Perpetual Tier storage resources cannot be underestimated. The Perpetual Tier provides a landing zone for less-active data and inactive data in the form of backup, disaster recovery and archives and can use the cloud, archive disk and tape. A combination of cloud and tape can help to achieve best practices in digital preservation for the long-term by providing IT managers with three copies of data, on two different media types and one offsite.

StorCycle is Built Around Four Main Pillars of Functionality:

1) *Identify* and Scan primary storage and identify inactive files

2) *Migrate* and store data to one or more of the following targets:
   - Spectra BlackPearl® – An object storage disk solution that tiers to cloud and tape
   - NAS – Spectra BlackPearl® NAS, any other vendor's NAS, or any addressable file path
   - Public Cloud – Including Amazon S3 storage services

3) *Protect long-term* – Ensure files remain safe on the Perpetual Storage Tier indefinitely

4) *Access and Retrieve* – Allow users to easily access and retrieve migrated files from the Perpetual Tier
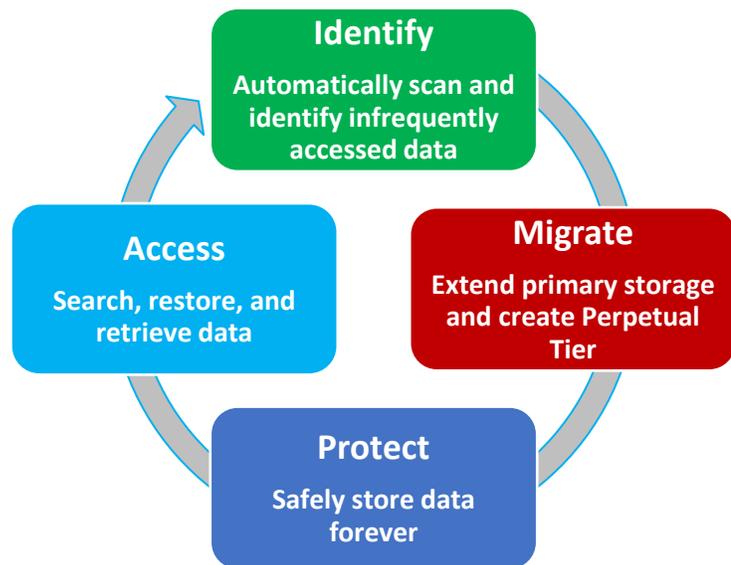
StorCycle Resolves Storage Challenges - Today *and* Tomorrow

IT Budgets are not scaling as fast as data is being generated creating added pressure on the overall IT organization to optimize resources. Primary storage solutions are the most expensive to acquire and operate and need to be used efficiently and wisely. Leaving only the mission-critical and important data on primary storage systems will lead to higher performance of the most active data. Otherwise, degraded performance and reduced efficiency will negate many of the benefits of primary storage. As StorCycle scans data, migrates data, and frees primary storage space, backup replication and snapshot copies are

also smaller, leading to shorter backup windows and reduced backup costs. Alleviating storage constraints on the primary storage tier can also lead to an organization's ability to reduce the overall software licensing fees for the primary storage systems.

Data centers can choose from a variety of storage management software packages, but the cost of these packages can become extremely expensive and often exceeds the entire cost of the secondary storage solution. To directly address mounting software costs, StorCycle provides two licensing models - a permanent license model and an annual subscription model. These models provide greater flexibility for organizations to choose whether they want to use StorCycle as a capital or operational expenditure and offer significant savings over standard software licensing models.

With the StorCycle policies configured, a user can select the storage destination(s) for the managed data and assign metadata tags for easier search and retrieval at a later time. Multiple storage targets can be assigned to each job, allowing for multiple copies of data to be created on premise, on tape, and/or in the cloud. StorCycle will manage multiple copies of data and can ensure that at least one copy is stored in a location that best addresses the users' needs.

Conclusion

For most organizations, facing the challenge of managing terabytes, petabytes and even exabytes of archive data for the first time can force the redesign of their entire storage strategy and infrastructure. As businesses, governments, societies, and individuals worldwide increase their dependence on data, archiving and data preservation have become a critical practice for the 21st century data center.

Spectra has identified the long-standing need for a solution that reduces the costly primary storage costs for hardware acquisition, TCO and storage software licensing by reducing the data storage load from primary storage systems. It's truly amazing that data centers have lived with so many costly storage inefficiencies for decades without demanding significant improvement. Spectra's new StorCycle solution, which maximizes the Perpetual Tier, is designed to reduce the overall strain on the Primary Tier and to empower organizations to fully leverage and realize the benefits of their primary storage investments.

Spectra logic is committed to delivering a high value solution with StorCycle and has invested considerable resources to develop an exciting roadmap to ensure the product meets the customers' changing and growing requirements. With the amount of storage and the related inefficiencies growing unabated, sooner or later organizations will need to implement a highly effective storage architecture to break through data lifecycle management barriers before they become insurmountable.