

OPTIMIZING PRIMARY STORAGE

WHITE PAPER

FILE ARCHIVING SOLUTIONS FROM
QSTAR AND CLOUDIAN



CONTENTS

EXECUTIVE SUMMARY	1
The Challenges of Data Growth	1
SOLUTION OVERVIEW	3
SOLUTION COMPONENTS	4
Cloudian HyperStore® Software	4
QStar Archiving Software	4
TECHNICAL DETAILS	5
Setting Archiving Policies	6
How much data can be reclaimed?	7
CONCLUSION	8



EXECUTIVE SUMMARY

Deployed together, QStar and Cloudian provide a robust file archive solution that allows organizations to easily migrate static data from primary storage systems such as NetApp filers, onto a cost efficient, highly scalable, object-based storage platform running Cloudian HyperStore® software. The components that make up this solution include the following:

- **QStar Archive Manager:** Acts as a local cache into backend Cloudian object store. A CIFS or NFS gateway is presented out and used as a target for QStar Network Migrator. Archiving to object storage happens automatically when certain thresholds and conditions are met (i.e. Cache reaches 80% capacity).
- **QStar Network Migrator:** This component is responsible for the actual data movement from primary storage to the archive point and subsequently the Cloudian HyperStore platform. Network Migrator is also responsible for file stubbing and supports NetApp filers using FPolicy (CIFS).
- **Cloudian HyperStore:** A feature rich and highly scalable software-defined, object based storage platform that is 100 percent S3 compatible. In this architecture, Cloudian acts as the central repository for all archive data. Cloudian is a scale out, geo cluster solution that supports various features and technologies including replication, erasure coding, multi-tenancy and QoS, as well as the ability to tier-out into any other S3 compatible platform such as Amazon S3, Amazon Glacier or another Cloudian system.

THE CHALLENGES OF DATA GROWTH

In an ever-evolving landscape, storage administrators and IT departments are often challenged with increasing demands for additional storage capacity to introduce new services, drive productivity, streamline business processes and accommodate the 'natural growth' of data. In addition to maintaining service levels for current platforms they are often expected to accommodate these requests while also reducing cost.

This can prove challenging with traditional NAS and SAN storage systems due to architectural limitations of scalability and performance, as well increased operational and support costs as systems fill up with data and reach the end of their expected duty cycles. Many frame-based storage arrays can only support a set number of drives before customers are forced to buy larger controllers, or 'rip and replace' entire systems also known as forklift upgrades. These upgrades are typically costly, can be disruptive and also have to be repeated periodically as older systems are phased out by storage vendors.

Traditional approaches to data storage and management are changing with utility compute models now becoming mainstream and many organization's looking to cloud and SaaS based solutions to reduce primary storage spend and facilitate the introduction of new services.

The traditional NAS and SAN foothold on IT storage is in decline and sales are being disrupted by customers who now opt for flash based (hybrid type) systems to deliver high performance at reasonably low cost for Tier 1 data sets and applications. In addition to SSD caching, other technologies such as compression and deduplication are often integrated within these new systems and as such, the traditional storage vendors of old with their monolithic offerings are struggling to compete. This decline is also influenced by the fact that although transactional data sets are growing year on year, these workloads are the slowest moving trend of predicted data growth in the coming years.

At the other end of the spectrum, where performance is less important, cloud based services are now becoming the de-facto standard to deliver high capacity storage repositories and data archives. Many of these services can be consumed in the public cloud (e.g. Amazon S3) or privately, on site, behind the customer's firewall (e.g. Clouidian HyperStore). Whether an organization chooses to deploy cloud services privately, or use a public cloud service, or a combination of the two, is often a matter for internal politics and debate but major software vendors are increasingly adopting these standards. For example, the S3 ecosystem has over 500 independent software vendors who now support the S3 standard, and this trend is set to rise. Unlike the transactional data sets described above, unstructured data sets are growing at an exponential rate.

Gartner, IDC and almost all other analysts suggest that the rise of unstructured data will account for up-to 80% of all data created in the next decade. There are multiple drivers for this extreme growth including a massive increase in the amount of 'human data' that is being generated. This in itself is being driven by various market trends and technologies, including the consumerization of IT, an influx of mobility solutions, social platforms, online media, file sync and share tools and so on. Other major contributors to this growth include increasing amounts of 'Interaction' and 'Log generated' data, not to mention large data sets created by 'Big Data' & 'Analytics' as well as other machine-generated data from the 'Internet of Things'. (Figure 1)

With the exception of certain workloads, this unstructured data typically does not have the same performance characteristics of traditional Tier 1 data and can therefore reside in a more cost-effective storage medium such as, Clouidian HyperStore, a highly scalable object storage platform.

DATA GROWTH - INDUSTRY DRIVERS

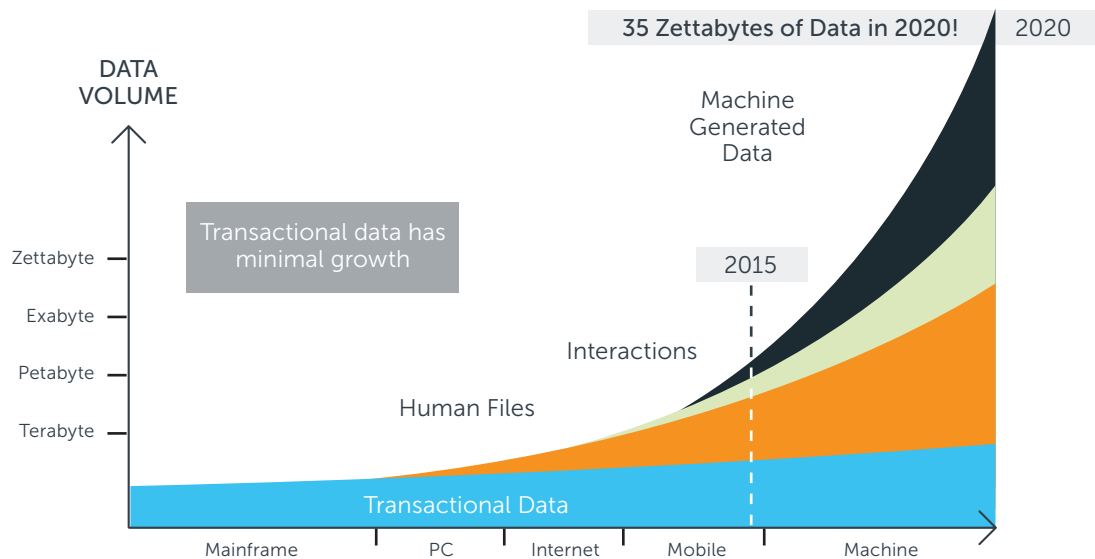


Figure 1 - Growth Trends

Very often organization's will introduce new, capacity hungry services and applications using local primary storage systems, especially if these systems have been over provisioned and capacity is available. This can present challenges for IT and storage admins as capacity is quickly consumed, creating 'bloated' storage arrays that contain mostly aged, static and dormant data sets.

As a result, management of these platforms becomes more complex and challenging, support costs skyrocket, performance suffers and problems arise when Tier 1 applications require additional space, which has been consumed by data considered less business critical. Adding additional capacity to these systems is only a short-term fix and forklift upgrades will only take you so far before you need to repeat and restart the process all over again.

This is where QStar and Cloudian can provide a solution.

SOLUTION OVERVIEW

Used together, QStar and Cloudian provide a robust mechanism for organizations to easily identify and migrate static data sets from existing primary storage systems to a highly scalable, feature rich object based storage platform. Using this approach, customers can migrate data to the appropriate storage tier based on user defined policies such as file type, file modification time or file size. Once data has been migrated to Cloudian, QStar then creates a stub file to the data and users & applications continue to access the data transparently in the normal way.

SOLUTION BENEFITS:

- Lower storage total-cost-of-ownership (TCO)
- Extend the lifecycle of Tier 1 storage systems (NetApp, HDS HNAS, IBM GPFS, Windows, Linux, Mac, UNIX)
- Cap primary storage
- Reduce backup times
- Improve data governance
- Improve performance of Tier 1 systems (NetApp specifically)
- Achieve better storage efficiency by storing data in the appropriate tier based on performance characteristics

Take advantage of a rich S3 ecosystem, use cases include the following:

- Backup and Archive
 - EFSS & File Collaboration services
 - Web Content Storage
 - Big Data Analytics
 - Storage as a Service
 - Hybrid Cloud with AWS
- And many more...



SOLUTION COMPONENTS

CLOUDIAN HYPERSTORE® SOFTWARE

Cloudian HyperStore software delivers a fully S3 API compliant, multi-tenant, and multi-datacenter hybrid cloud storage solution. Cloud service providers use Cloudian HyperStore software to deploy public clouds and managed private clouds. Enterprises use Cloudian HyperStore software to deploy private and hybrid clouds.

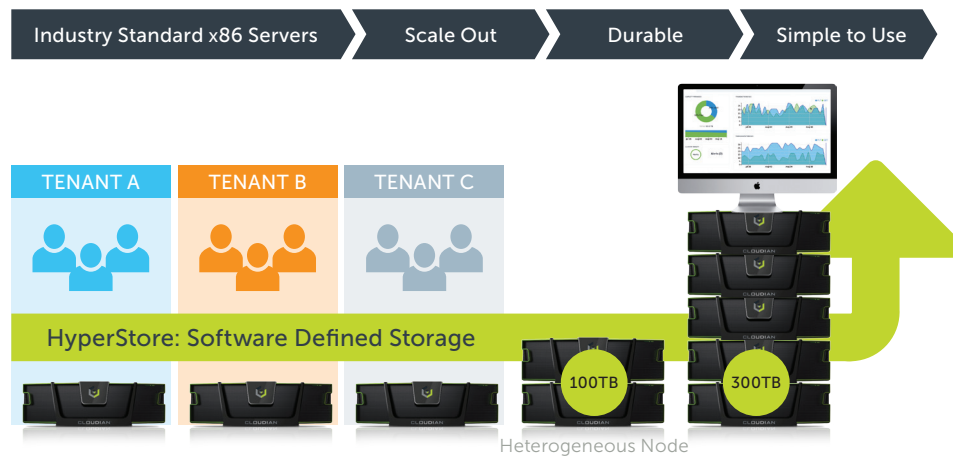


Figure 2 - Cloudian HyperStore Software

Cloudian HyperStore software employs a fully distributed and replicated peer-to-peer architecture with no single point of failure. It easily scales horizontally using industry standard x86 hardware so deployments can start with a few servers in a single datacenter and then scale out as usage increases to thousands of servers distributed across multiple datacenters managing hundreds of petabytes of data. Its distributed architecture with automatic replication and recovery services makes it highly resilient to network and node failures without data loss. Similarly, when scaling the storage cluster or performing maintenance, changes in node availability are automatically detected without service interruption.

Features like hybrid cloud streaming, virtual nodes, configurable erasure coding, data compression, and encryption provide highly efficient storage and data management that lets users store and access their data where they want it, when they want it.

QSTAR ARCHIVING SOFTWARE

QStar offers performance and cost flexibility with unlimited scalability. QStar Active Manager Software can be easily integrated with an organization's network environment and can be installed on physical or virtual servers using Windows or Linux operating systems. It integrates seamlessly with popular digital asset management and media asset management systems, providing simple access to Cloudian HyperStore, without API support.

QStar's Archive Manager Software creates an Active Archive gateway for Cloudian HyperStore to provide a quick and easy method of archiving any file-based archive content. The software presents the archive as a network share or mount point. Using standard network protocols such as CIFS or NFS, creative users, editors and administrators can easily store, search and retrieve data within the archive. In addition, completed digital content from production, post-production, mastering, transcoding or distribution can be archived, freeing up capacity on primary storage for new content. QStar software then uses the Cloudian S3 API to move content across the LAN or WAN into the Cloudian HyperStore object-based storage solution – all while being transparent to applications.

QStar software allows retention periods to be set, converting data into a secure read-only format for a set period of time. Data can be automatically removed at the end of this period, allowing the re-use of this capacity for new content. Multiple retention periods can be created to support different data sets to meet varying business needs. Additionally, QStar can independently replicate data to multiple sites and to other archive technologies, such as LTFS tape for example.

For digital content already stored on SAN or NAS primary disk systems, organizations can create policies using QStar Network Migrator software, which automatically migrates content to QStar Archive Manager, and then to the Cloudian HyperStore object-based dispersed storage system.

TECHNICAL DETAILS

QStar software easily integrates within an organization's network environment and can be installed on physical or virtual servers using Windows or Linux operating systems. QStar Archive Manager creates an "Active Archive gateway" for Cloudian HyperStore to provide a quick and easy method of archiving any file-based archive content. The software presents the archive as a network share or local mount point. Using standard network protocols such as CIFS or NFS, users, applications and administrators can easily store, search and retrieve data within the archive.

The ASM server, or server cluster is responsible for data archiving from the local cache to the Cloudian backend storage platform. High and low capacity thresholds can be configured to govern when actual data archiving to the object store happens, and it is the ASM server (or cluster) that is responsible for maintaining all indexes and catalogues pertaining to media management and the QStar Structured Storage Device (QSSD), which in this case is the Cloudian HyperStore storage platform.

Aged and unstructured primary storage data sets can be easily archived, freeing up capacity for new content. The software then uses the Cloudian S3 API to move content across the LAN or WAN into the Cloudian HyperStore object-based storage solution all completely transparent to users and applications.



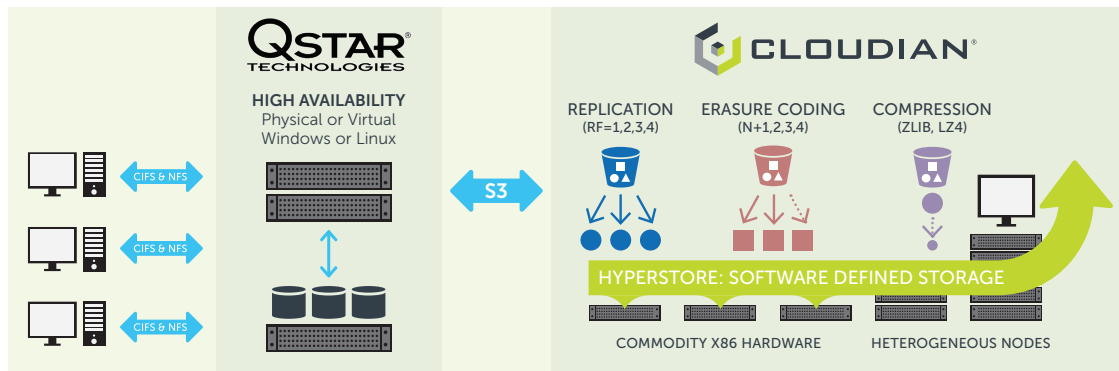


Figure 3 - ASM Solution Overview

The introduction of a secondary storage platform can help to reduce the overall total cost of ownership (TCO) for storage as well as future proof storage investments. By taking advantage of Cloudian object storage and QStar software, customers can easily move data into the appropriate storage platform, based on custom defined criteria automatically and do this completely transparently. Traditional approaches to data storage are inefficient; RAID overhead, replication, over provisioning of volumes all lead to very poor storage utilization rates, object storage overcomes these challenges and provides a robust scalable platform for the future.

SETTING ARCHIVING POLICIES

In order to migrate data already stored on SAN or NAS primary disk systems, organizations can create policies using QStar Network Migrator software, which automatically migrates content to the QStar Archive Manager access point, and then ultimately to the Cloudian object-based dispersed storage system when the cache reaches configurable thresholds.

Using the Network Migrator component, administrators are able to define specific policies for data movement based on standard file metadata. These policies can be custom tuned and are based on various parameters including file type, file age, last access time, size and many more. Once a policy has been defined and is executed, QStar will copy, move or migrate the data to the Cloudian HyperStore cluster via the 'Archive Manager Software' (ASM) solution, which provides a CIFS/NFS gateway with S3 support into Cloudian.

Once data has been migrated to the object store, QStar will leave a stub file on the file system. This stub file, or reparse point, is used to call the data from the S3 object store. When a user or application attempts to open the file, the object is pulled from the Cloudian backend storage repository or if the file has already been retrieved, from the cache. An example stub file is shown below, in Figure 4.

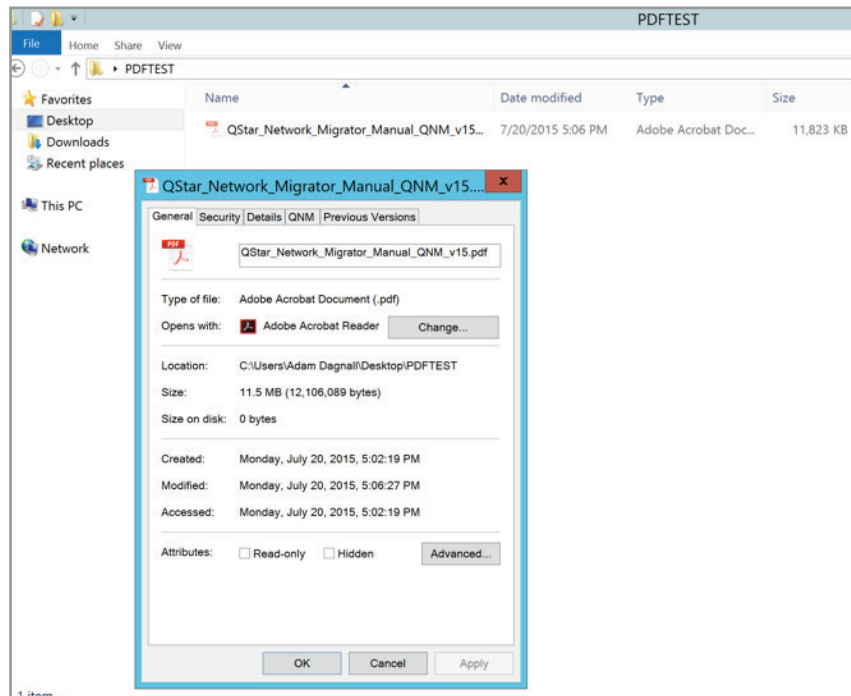


Figure 4 – Stub File Properties

As the QStar Network Migrator policy server is not in the data path, it is therefore not a single point of failure. Operations will continue even if this server is offline. All information pertaining to the location of the file is held within the stub file. A SQL database is maintained on the Network Migrator server which provides reports on files moved and helps with simulating policies to view their effectiveness before being applied. The database also allows QStar to recreate the stubs on the primary storage if accidentally deleted.

HOW MUCH DATA CAN BE RECLAIMED?

As well as actually being able to copy, move or migrate data from primary to secondary storage systems, QStar's Network Migrator product also has another useful feature in terms of its 'Storage Reporter' capabilities. This tool (which can be run for free and does not require installation on primary systems or shares) can be easily used to determine how much data would be archived to the secondary system based on user defined characteristics. This gives organizations the ability to easily identify how much capacity can be reclaimed without impact to services. Depending on the data type and use case, up to 80% of inactive file data can be migrated to Cloudian via QStar software.

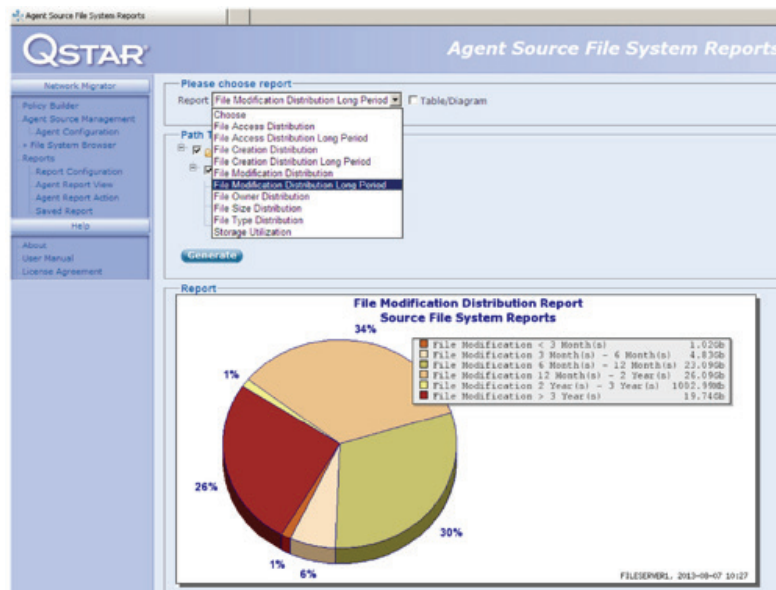


Figure 5 - Storage Reporter Output

CONCLUSION

Deployed together, QStar and Cloudian can be used to help to extend the life of existing primary storage systems by optimizing storage placement of file data based on the performance characteristics of the data itself. Once the data has been moved to the appropriate platform, administrators should see capacity and performance gains on their primary storage systems. For example, with NetApp, 'WAFL' file system performance significantly degrades over time as system capacity increases with data and snapshots. Figure 6. More information can be found in this post http://wikibon.org/wiki/v/WAFL_Performance

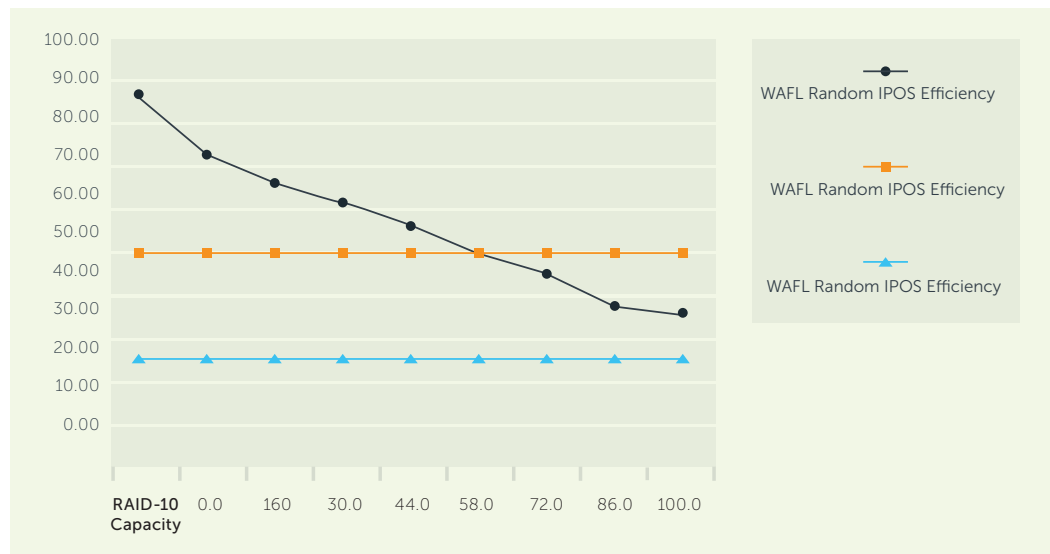
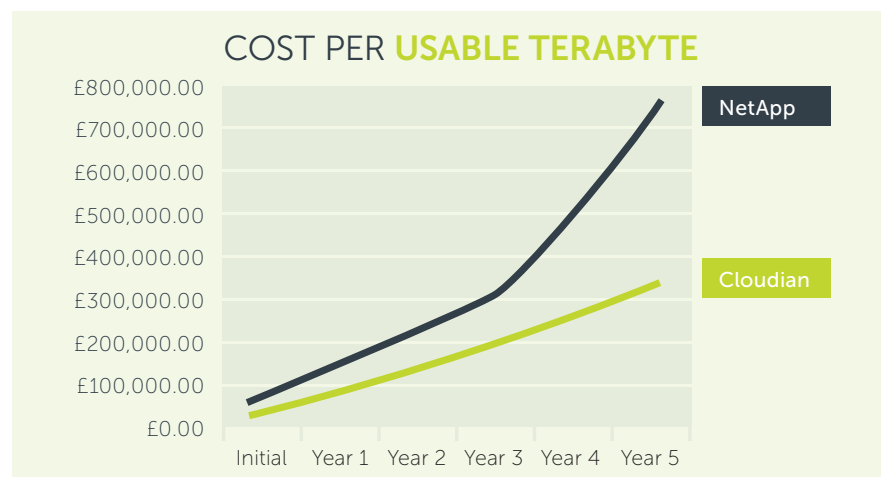


Figure 6 - NetApp - WAFL Performance Degradation on Loaded Systems

By moving inactive data sets to Cloudian and freeing up capacity performance and response times will improve for Tier 1 applications and workloads.

Figure 7 highlights the delta between primary and secondary storage system costs over time, although to calculate costs for your own environment a TCO calculator can be located here. For accurate QStar and HyperStore costs, please work with your local Cloudian sales reps.

Click the paper clip icon to launch the TCO Calculator



The introduction of a secondary storage platform can help to reduce the overall total cost of ownership (TCO) for storage as well as future proof storage investments. By taking advantage of Cloudian object storage and QStar software, customers can easily move data into the appropriate storage platform, based on custom defined criteria automatically and do this completely transparently. Traditional approaches to data storage are inefficient; RAID overhead, replication, over provisioning of volumes all lead to very poor storage utilization rates, object storage overcomes these challenges and provides a robust scalable platform for the future.

Get started today and receive 10TB for free with our Community Edition:
<http://www.cloudian.com/free-trial/>

ABOUT QSTAR

QStar's archive philosophy is reflected in the architecture of our software. Designed to be operating system and storage hardware independent, QStar customers are not locked into vendor specific server and storage hardware. In addition they have the choice of using QStar's optimized proprietary file system (TDO) or industry standard files systems, such as LTFS for tape, or UDF for optical. The modular platform supports incremental capacity expansion from terabytes to petabytes, and offers advanced features such as replication and real-time mirroring. This unique approach gives QStar customers a long-term data archive strategy with the agility they need to evolve in changing market and financial conditions.

ABOUT CLOUDIAN

Cloudian is a Silicon Valley-based software company specializing in enterprise-grade storage. Its flagship product, Cloudian HyperStore®, is an S3-compatible storage platform that enables service providers and enterprises to build reliable, affordable and scalable hybrid cloud storage solutions.

Follow us on Twitter @CloudianStorage



Cloudian, Inc. | 177 Bovet Road, Suite 450 | San Mateo, CA 94402
Tel: 1.650.227.2380 | **Email:** info@cloudian.com | www.cloudian.com

©2015 Cloudian, Inc. Cloudian, the Cloudian logo, and HyperStore are registered trademarks or trademarks of Cloudian, Inc. All other trademarks are property of their respective holders.

CLO-WP-13001-EN-01