REFERENCE ARCHITECTURE

A Reference Architecture for Building a Searchable Content Archive Using CatDV and ActiveScale

Abstract

This document defines a reference architecture for a combined Quantum CatDV and ActiveScale workflow. The storage infrastructure is implemented with Quantum software and appliances.

Quantum

Table of Contents

Executive Summary

The objective of this paper is to present a detailed, repeatable reference architecture for deploying asset management and archiving solutions using the Quantum CatDV media management and orchestration platform integrated with Quantum ActiveScale object storage. The architecture demonstrated can scale into the Petabyte range and beyond, protects assets with a very high degree of resiliency, offers retrievability of assets with excellent performance, all while making the entire library available to and tightly integrated with the Quantum CatDV asset management platform.

By implementing the MAM as the creative user's primary interface into their workflow and archive, customers will be able to index and ingest master files, automatically generate links back to all the raw sources of those masters, which are then automatically retrieved and presented to the user or solution when needed. The customer can organize their assets, monitor asset locations and levels of protection, and have a searchable interface to quickly find content across their entire deployment. By configuring powerful workflow automation engines users can order and speed their workflow steps including managing ingest of raw or source files, production management and planning, craft editing and finishing, review and approval, asset enhancement and monetization, and archive workflow steps.

At a very high level, Figure 1 below provides a view of the components and options that go into a Quantum Solution:

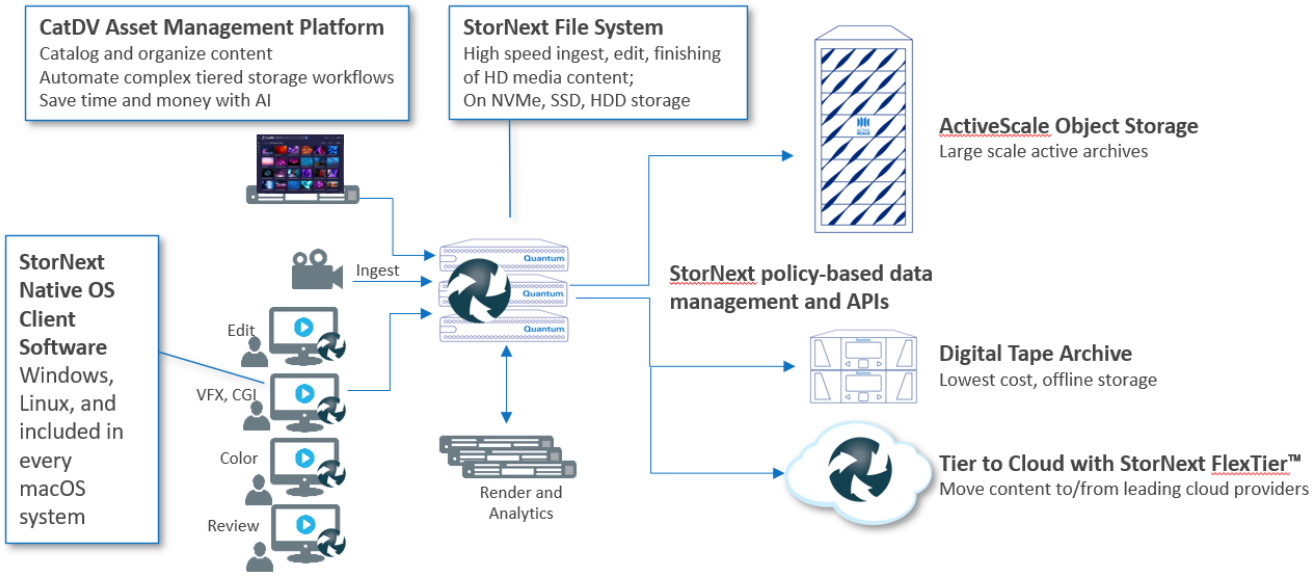


Figure 1 – Quantum Solutions for Media Workflows

The Need for This Solution

With rapidly increasing media resolutions, delivery formats, and the need to consolidate ever larger content libraries to speed content production and support extended monetization efforts, there is an urgent need for content library storage solutions that can seamlessly support asset management workflows. The challenge facing architects today is to design a solution that can track, store, and retrieve these assets even while the asset library and creation workflows are growing dynamically and beyond traditional storage architectures. The number of collected assets, multiple ongoing projects, geo-dispersed sites and remote operators all make it extraordinarily challenging for organizations struggling to maintain order in media management and prevent the loss of irreplaceable content. Over the last year, facilities have been required to deliver highly reliable workflows despite a near-total remote work force. With the surge of new content and multiple content staging areas such as tape, nearline, and other disk storage, tracking all this data and managing their multiple locations manually is both error-prone and inefficient.

By implementing CatDV as the Media Asset Management (MAM) solution that presents a unified view of asset inventory across local and object storage that automates file movement and proxy generation, a single canonical view of content inventory is achieved while also allowing creatives and asset managers to be more efficient and work with content as an intrinsic part of their workflow rather than a separate effort to manage content, or storage, directly.

Implementing a private cloud object storage as the primary content storage system presents a more reliable and faster way to store these assets for both short or long term at any scale. Erasure coding techniques within ActiveScale ensure the integrity of the data, and delivers resiliency of that data by spreading the data over different locations, while asset locations and cloud-delivered tools for asset enhancement are all orchestrated within a centralized CatDV server, database, and worker automation routines. Together, the integrated solution addresses the key concerns of collaborative asset management and project workflows, makes assets available wherever needed, yet also protects the original assets with a high degree of resiliency and near-instant file retrieval performance.

When implemented in this way, content creation and archival workflows become a virtuous circle of media management and production. In the diagram below, we observe the workflow steps of asset and content creation through archive becomes a cyclical pattern, completing the circle with the ability to retrieve archived assets that can in-turn be used to create more assets or select for the next project combined with newly created media.

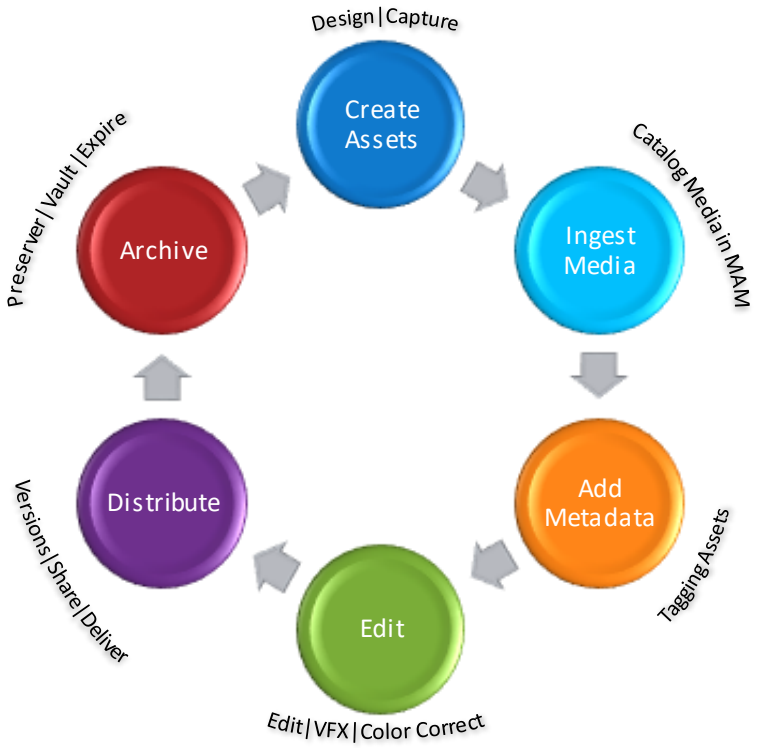


Figure 2 – Media Asset Life Cycle

User Workflow Examples

Following are a few examples of workflows that highlight the applications of the components of this Reference Architecture.

User Workflow Example 1: Ingest Master Files After Project Completion

In this workflow, editors or producers might have used CatDV primarily for work in progress while a project was active, and now upon completion and delivery of the finished work to the client now want to ensure that all potentially valuable source files are protected and associated with the project. This might occur if the project collaboration occurred on shared storage such as SAN/NAS, or when a finished project is copied over from an external hard drive or location. To proceed, *create an "Archive Project Drop"* watch folder to receive all project files for archiving.

The user will then create a worker action within the file orchestration worker automation environment. This automatic workflow step will perform the archiving steps, moving the files from the local, work in progress storage from the watch folder created to archive. This workflow step should also create a production catalog manually or an automatic step of the worker action, create any needed proxy or reference versions of media, especially to ease web-based content curation, and in working with high resolution files. Next, the workflow step will automatically archive content to Active Scale object storage, and optionally to other tape, private cloud, or public cloud destination. Once the archiving step is complete, the content record remains in CatDV, and are easily inspectable to confirm that content is stored in archive and no longer on work in progress shared storage. At any time, the user can restore and validate the original files to work with again.

User Workflow Example 2: Archive Master Files on Ingest

In this workflow, users want to get master files such as those captured during production and saved to camera carts protected as soon as possible, while also making these files immediately available to the entire distributed team for grading or tagging. Users ingest raw camera media to CatDV which then creates asset records and immediately archives the media on ActiveScale. CatDV will automatically create mezzanine or proxy files for editorial and work in progress. Once the project is ready to deliver, the archived camera native files will be restored for conform steps. After the master is delivered, camera native can be safely deleted from primary storage, and the master and the project will be archived to ActiveScale.

User Workflow Example 3: Public Cloud Disaster Recovery

In this workflow, a user has assets in a public cloud store such as Amazon S3. The user will use Quantum CatDV to index all content stored in the cloud bucket. Once indexed, the customer can then assess the cloud provider and category of storage, reduce egress fees by performing work on the proxy or reference file, or shift their primary asset store to an object store that delivers faster performance, predictability, and retrieval time with a solution incorporating CatDV and ActiveScale. With a single content reference in CatDV, customers can retain a copy of content in their public cloud provider, and a reference to the secondary asset location with ActiveScale as the primary archive.

User Workflow Example 4: Legacy Archive Ingest from External or S3-Compatible Storage

In this workflow, customers wish to simultaneously organize asset collections on multiple locations, including S3-compatible storage. All content will be indexed by CatDV, regardless of initial location, to inventory and index all content while extracting file metadata, making it simple to search and retrieve from a single interface and support ongoing content consolidation operations. No matter the location of the indexed file, users can restore media as needed, generate any needed proxies and archive a protected copy of the file on ActiveScale.

Solution Overview

With the ever-increasing prevalence and demand for digital assets and rich media content production in all forms of communication, every industry is realizing the need for resilient, highly scalable, and reliable solutions to properly track, store, and retrieve these assets, even while the asset library and creation workflows are growing dynamically. The number of collected assets, multiple ongoing projects, global diversity of sites and locations and more remote operators and producers all contribute to the challenges of organizations struggling to maintain order in media management. Facilities have been tasked with providing highly reliable workflows despite a near-total remote work force. As companies have seen, a focused content and data-centric lifecycle with shared accessibility across worksites and remote workers has proven highly effective and offers a model for new levels of reliability and capability. Businesses that have adopted this model can transform their operations and capabilities by rising to evolving challenges of data and content collection, production, archiving, and retrieving of assets in a timely manner.

With all this content and the proliferation of multiple staging areas such as tape, nearline and other storage volumes, tracking all this data and its location with a spreadsheet or notes catalog is not efficient, or safe. Implementing a private cloud object storage to organize all content presents a more reliable and faster way to store these assets short or long term. By implementing CatDV as the organization's Media Asset Management (MAM) system presents a single, canonical view of all content in a single place. This highly valuable content is protected with erasure coding technology and other data protection features of ActiveScale to assure the integrity of all data. Further, the archive system can be divided across different locations, or have multiple copies in multiple locations for resilience or workflow need, and CatDV can then provide a unified view of all content across all locations. With this solution, companies can make work in progress shared storage highly efficient and sized for the needs of active projects only, with the ability to scale to any size needed on massively scalable object storage, and track all assets from first capture to long term archive.

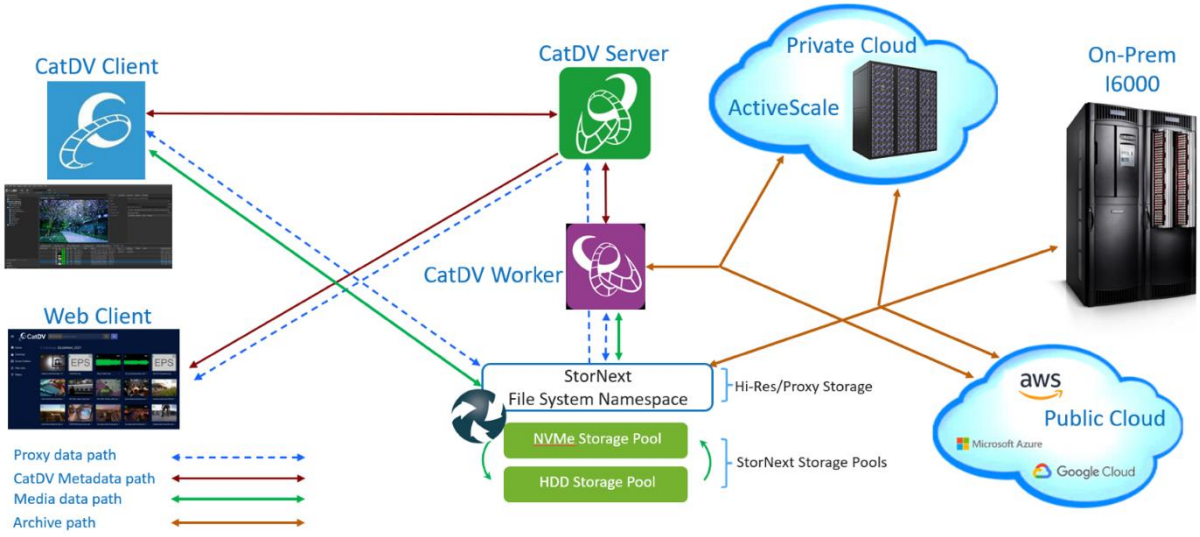


Figure 3 – High-Level CatDV and ActiveScale Architecture

Technology Summary

The table below lists the technology components that make up the Quantum CatDV and ActiveScale reference architecture outlined in this document. The paragraphs that follow the table provide more detail on the function of these components in the solution.

Solution Components	Item	Notes	
ActiveScale	S3 Credentials	The access and secret keys were created to facilitate the S3 connection to the endpoint bucket on the ActiveScale System	
CatDV	CatDV Credentials	Ability to login to CatDV, user or Administrator	
ActiveScale Protocol	**Version (s)**	**Inbound/Outbound**	**Notes**
Amazon S3 Signature	2	Inbound	Provides S3 connectivity to CatDV
CatDV Protocol	Version (s)	Inbound/Outbound	Notes
Amazon S3 Signature	2	Inbound/Outbound	Provide S3 connectivity
REST API			HTTP-based API
HTML 5/CSS3/JavaScript			
Software	**Version (s)**		**Notes**
StorNext	7.0.1		CentOS Linux release 7.7.1908
ActiveScale	5.7.1		
CatDV Server	9.2.0		Windows 10 Server
CatDV Worker	8.1.9.19249		Windows 10 Worker
CatDV Pegasus Client	13.0.14		macOS Big Sur Pegasus Client

TABLE-1 – TECHNOLOGY SUMMARY

Quantum CatDV Media Asset Management Platform

The figure-7 below lists the server versions available and the specific features with each.

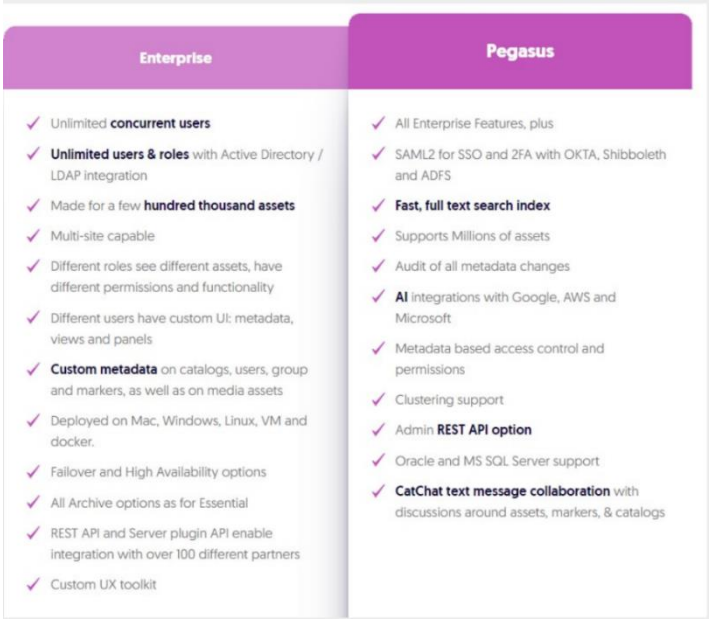


Figure 4 – Quantum CatDV Server Versions

Quantum CatDV is an agile asset management and workflow orchestration platform that provides powerful asset management, automation, and collaboration tools for organizations managing large volumes of digital media. The platform delivers a wide variety of media-centric capabilities, including traditional PAM, MAM, and DAM, sophisticated workflow automation, and fully customized applications. CatDV deployments create transparency and visibility, ensure that best practices are followed, and increase team productivity. CatDV unlocks value buried in large media libraries, saving time, money, and stress. Whether deployed on-premises or in the cloud, across traditional and object storage tiers, customer benefits include:

- Communicating and collaborating more effectively
- Automating workflows across tiered storage products
- Unlocking value in large content archives
- Saving time and money by leveraging Artificial Intelligence

CatDV is used today in post-production, corporate video, sports, government, and education markets, and has potential to expand to other markets using specifically designed plug-ins for expanded use cases such as genomics research, autonomous vehicle design, geospatial exploration, and any use case dealing with large unstructured data. For a closer look at what CatDV with Quantum brings follow the link here to find latest in what Asset Management from Quantum brings you.

Quantum ActiveScale Object Storage

Outstanding data availability and data integrity are essential for a world-class infrastructure. ActiveScale supports configurations where data is dispersed across three locations for extreme availability with ActiveScale and dynamic data placement technology (DDP), an advanced erasure code alternative to RAID. Even with a full data center outage, the 3-site configuration delivers continuous data availability and operations. In addition, ActiveScale proactive data repair provides high data integrity with data verification performed in the background.

All security threats may not be addressed with a single tool; ActiveScale supports keyless encryption of data at rest to prevent access via unauthorized applications. Object Lock makes objects immutable; once set, immutability can't be modified even by an administrator, making data safe from ransomware attacks. The system may be partitioned into buckets whereby different protection policies may be applied to each bucket to align with the security requirements.

ActiveScale systems are built with industry-standard, commodity-based server platforms. Each server is referred to as a node. All nodes execute the same embedded software. ActiveScale object storage systems are built with a predefined configuration using dedicated 10GigE or 25GigE switched networks. A single namespace or storage pool is maintained across the multiple nodes.

ActiveScale software enables your environment to scale from terabytes to exabytes in a scale-out configuration. Adding more capacity is seamless; as new clusters of nodes are added to the system, capacity is added to the pool and the dynamic data placement algorithm balances data placement and makes capacity available to applications. ActiveScale presents a standard S3-compliant object interface.

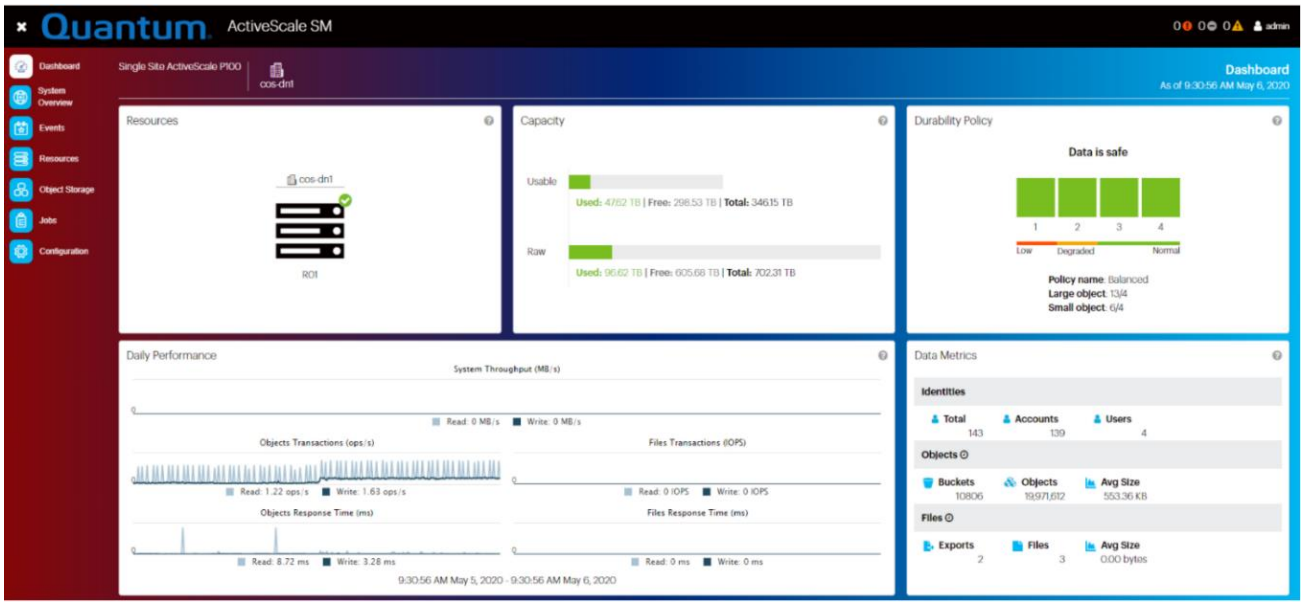


Figure 5 –ActiveScale Dashboard

Reference Architecture

Quantum CatDV Server and Worker nodes are server-side components typically reside in the datacenter, deployed on bare metal computers or VMs. Both applications can run on Mac, Windows, or Linux system, however, Windows and Linux are the predominant choices. It is recommended that the Server and Worker node are connected via 10Gig Ethernet to the file system and to the ActiveScale systems. Both can also be connected via block level access over Fibre Channel or High-Performance Ethernet. Whichever deployment and connection selected, the worker node must have read/write access to the file system.

The following diagram represents the lab environment that was configured to show a typical implementation of the Reference Architecture and a combination of components used:

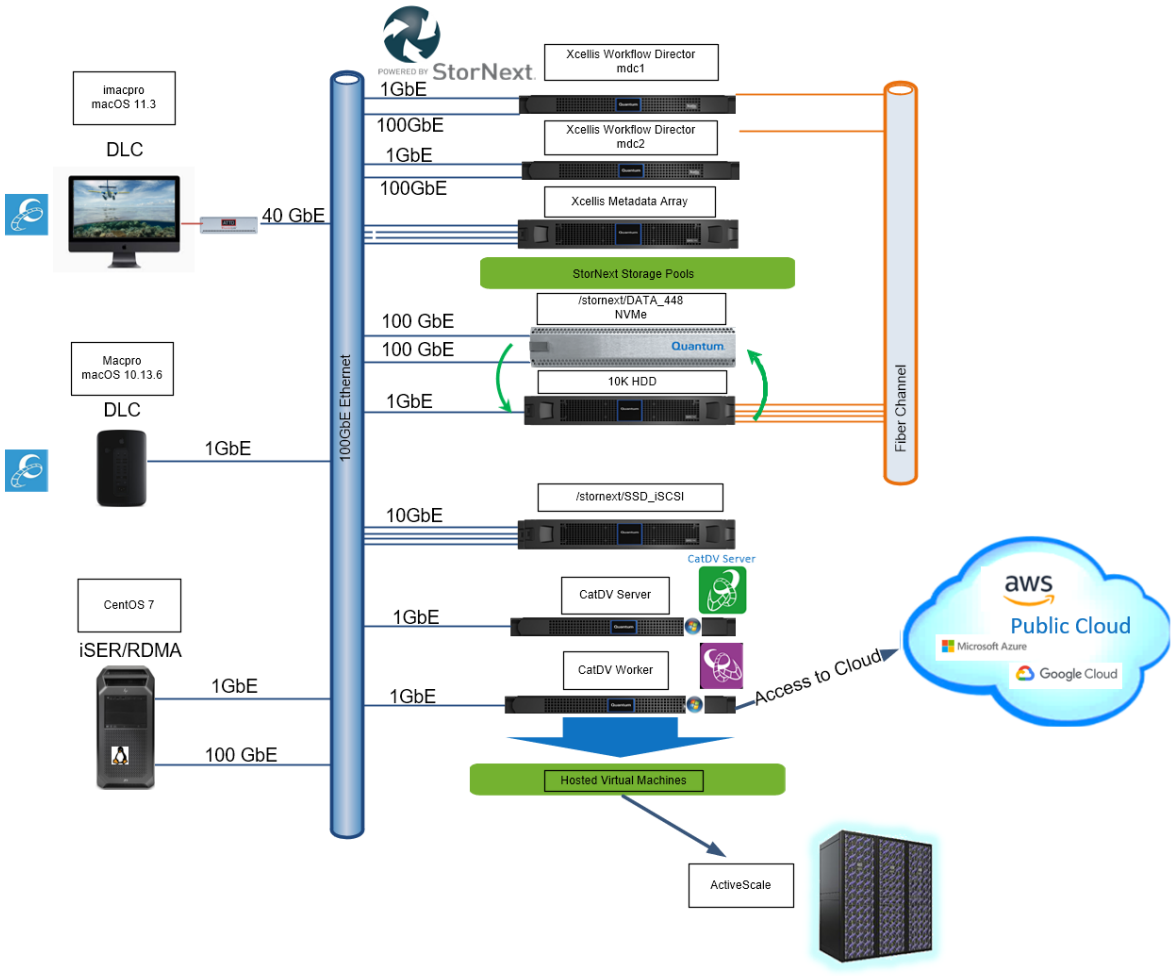


Figure 6 – CatDV and ActiveScale Reference Architecture

Requirements

File System

The file system needs to have the ability to playback footage at real-time speeds for various workloads up to 8K. Depending on the workflow, multiple streams of high resolution and high frame rate files will need to be supported. If the workflow is designed around proxy workflows the file system needs to have appropriate capacity to handle additional assets. Further, the file system should also be highly available for uninterrupted access to assets. The file system should be presented to the client's workstations NFS and SMB NAS protocols, in addition to direct access and seen as block storage devices with fibre channel and iSER. Distributed LAN Client (DLC) is another ethernet connectivity option but presented with a proprietary LAN client software. As the underlying file system, StorNext can provide these options in addition to multiple tiering options including S3 targets and single namespace tiering called pools.

Networking

The network requirements can vary depending on the facility and the specific workflow. For this Reference Architecture, we selected 100GbE networking from storage to specific Linux clients. There is also a limited amount of 32Gbit fibre channel storage connected to the file system. ActiveScale S3 targets are connected via 10GbE ethernet. CatDV servers and workers are connected via 1 or 10GbE ethernet. Clients in this RA are connected via 1, 10, 100GbE ethernet with limited 32Gb fibre channel connections.

S3 and Public Cloud Targets

ActiveScale will need appropriate credentials as part of the CatDV connections for private cloud locations. Multiple locations can be freely added to support advanced, hybrid environment, and multi-geo configurations.

Database

A database will be required for any CatDV solution, while a light-weight built-in database is provided during installation, performance is highly tuned for MySQL and Microsoft SQL.

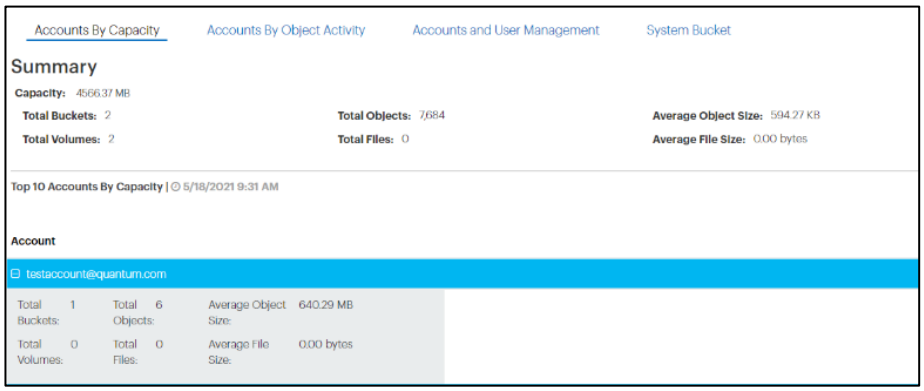
Configuration

The following sections detail the configuration required for CatDV and ActiveScale for this reference architecture.

The objective of this section is to provide enough setup details that the reader can get the products working together and create a basic workflow.

Quantum ActiveScale Product Configuration Details

1. Create a user account on the ActiveScale System known ActiveScale system in place
2. Create the S3 access and secret key credentials on the ActiveScale system
3. Create a bucket on the ActiveScale System (e.g. bucket-1)
4. Document the domain name of the ActiveScale System (e.g. s3p001.hgstcloud.com)
   Exp. https://Private.ActiveScale/operate/s3

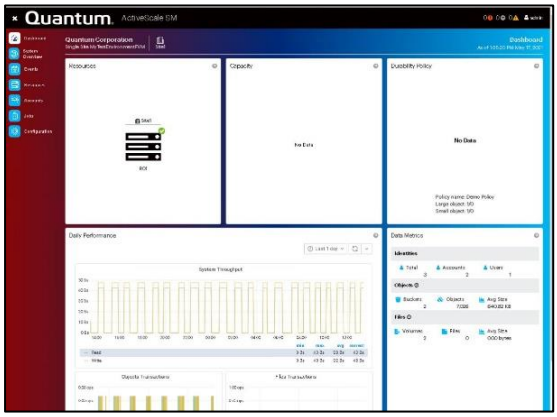


Quantum

Quantum CatDV Product Configuration Details

1. **Server** - Download installer for specific operating system here
    a. Post install – Configure through Server Control Panel 5 tabs, instructions follow link above
        1. Installation – location of installation
        2. Licensing – install registration code
        3. Database, build-in DB or MySQL or Microsoft SQL, if choosing third party, that will need to be setup and licensed, but probably a better choice
        4. Web Server, CatDV has its own built in one
        5. CatDV Server set server IP address and port number

    b. CatDV Enterprise or Pegasus Server Software - minimum requirements

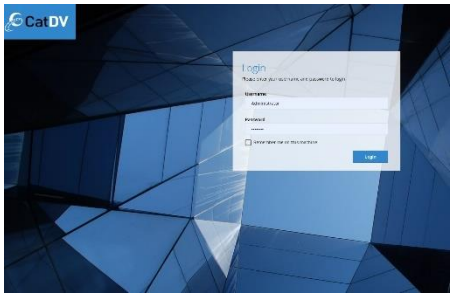


- Windows or Linux Servers (Can be virtualized)
- 4 Cores
- 16GB RAM (32 Recommended)
- SSD (NVMe recommended) 40GB (More for larger databases)
- 10GbE network connection

2. **Worker** - Download installer for specific operating system from here
    a. Install client per instructions, follow link above
        1. Install License
        2. Enter Server hostname or IP under the CatDV tab
        3. Proceed to "watch actions, click + to add new action. This part takes some investigation and planning. Review "Configure the Worker Node" in the manual
        4. Another good section "Hints and tips"
    b. CatDV Enterprise or Pegasus Worker node Software – minimum requirements
        - Windows or Linux Servers (Can be virtualized)
        - 12 Cores
        - 24GB RAM
        - SSD 40GB
        - 10GbE network connection

This example was designed with 4 simultaneous Worker node threads running at any time. More may be needed depending on the customer environment

3. **Web-Client**
    b. Most Web browsers compatible, interface uses HTML 5/CSS3/JavaScript
    c. Licenses should be obtained along with Server license.
    d. Can be deployed in a DMZ with SSL certificate
    e. Administrator or created username, then opens Web3 interface



4. **Desktop Client -** Download installer for specific operating system from here
    a. Install the client and following instructions
    b. License the application with registration code
    c. Connect to server with proper credentials
    d. Overview of main features
        1. Import clips
        2. View clips
        3. Outputting clips
5. **CatDV Active Scale Archive Plugin**
    a. Installed on CatDV Server
    b. Can be deployed on 1 or more dedicated Archive servers for load balancing and redundancy
6. **Plugins**
    a. Server-side
        1. CatDV Project Template Plugin
            - Allows definition of one more Project templates
        2. Some licenses for worker plugins will be licensed on the server

    b. Worker-side
        1. Download worker plugins from here
        2. Install plugins per instructions, follow link above, each plugin has different instructions, but the plugin usually resides in:
            - Windows - C:\ProgramData\Square Box\Extensions
            - Linux: /usr/local/catdvWorker/extensions
            - Mac: /Library/Application Support/Square Box/Extensions

7. **Optional CatDV Tools**
    a. CatDV Enterprise or Pegasus Client

**Pegasus Client is required for complex cinematic media types to playback natively in the application examples - RED, R3D, Arri, etc...*

8. **CatDV Adobe Panels and FCPX Extensions**
    a. Installation instructions here
9. **Dedicated CatDV Archive Server**
    a. Final option, have a dedicated server to manage archive tasks

Validation Activities

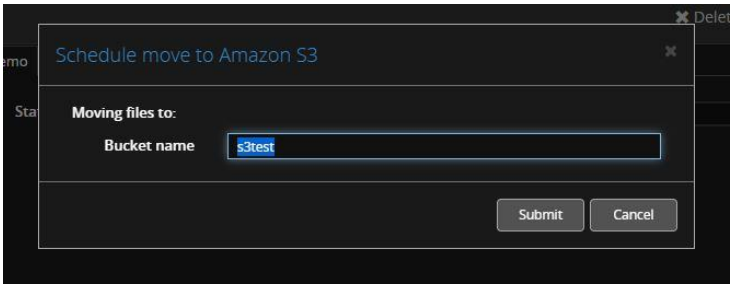
The following are procedures used to validate the solution's operation. These steps can be summarized as ingest, verify, archive to an archive target, verify, search for an asset, restore that asset, and verify again.

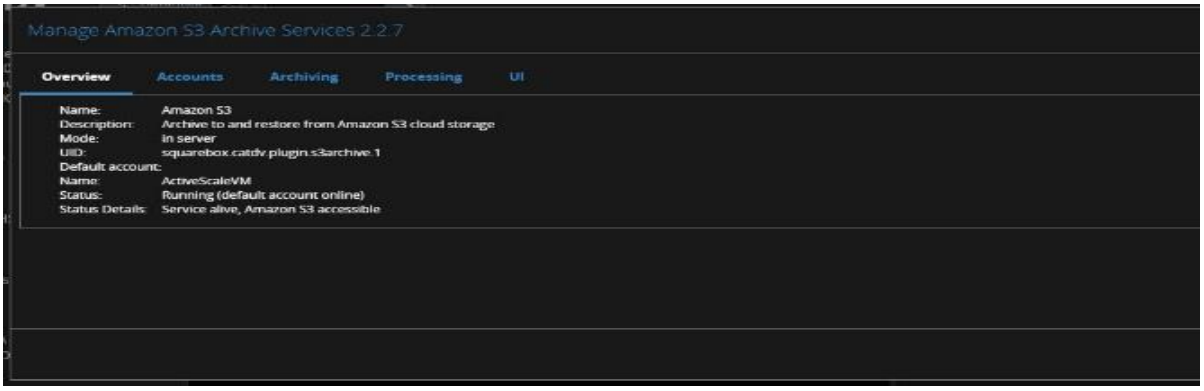
1. Ingest/Catalog assets into CatDV, then verify
2. Backup from CatDV to ActiveScale
3. CatDV shows that assets are tracked into ActiveScale
4. Search archived data when only in ActiveScale
5. Restore from ActiveScale to CatDV

ActiveScale and CatDV Workflow Testing

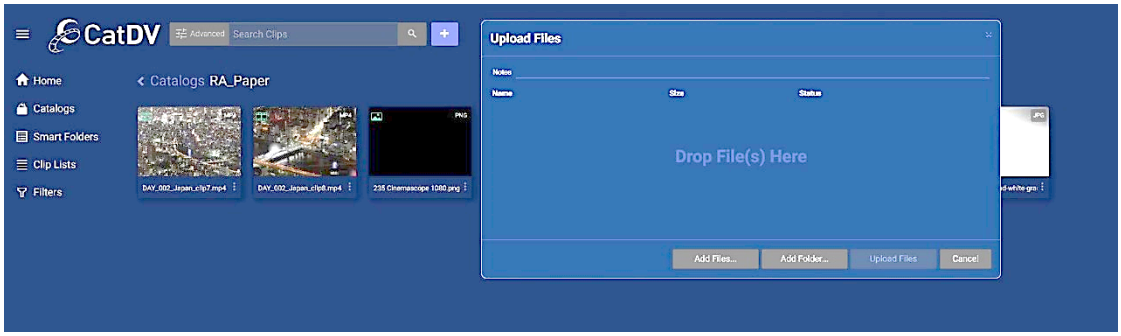
1. Define a cloud storage bucket to serve as the target location within CatDV

    a. Enter the bucket name

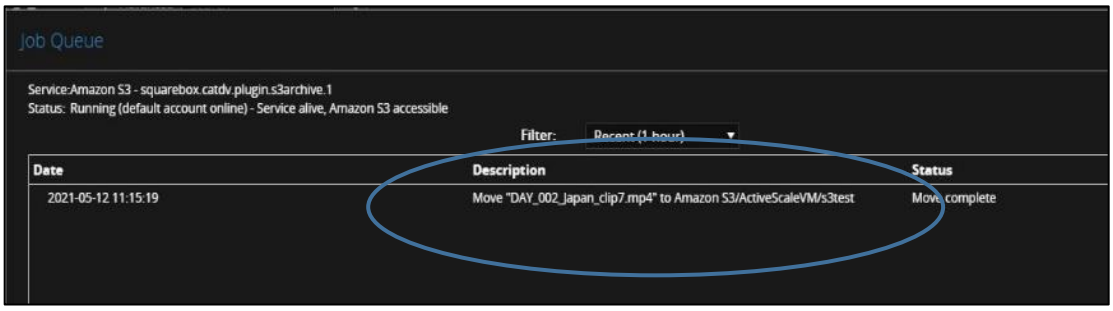




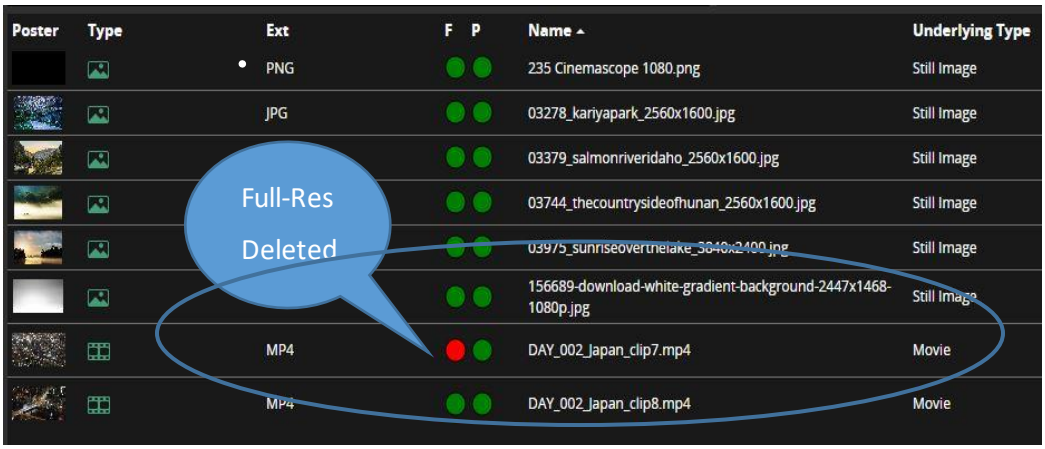
2. Confirm that assets are properly ingested into CatDV



3. Observe the content moving to the ActiveScale archive
    a. In client or Web Client select Tools > Schedule move to Amazon S3(compatible S3) > select "preconfigured" Bucket > select ok

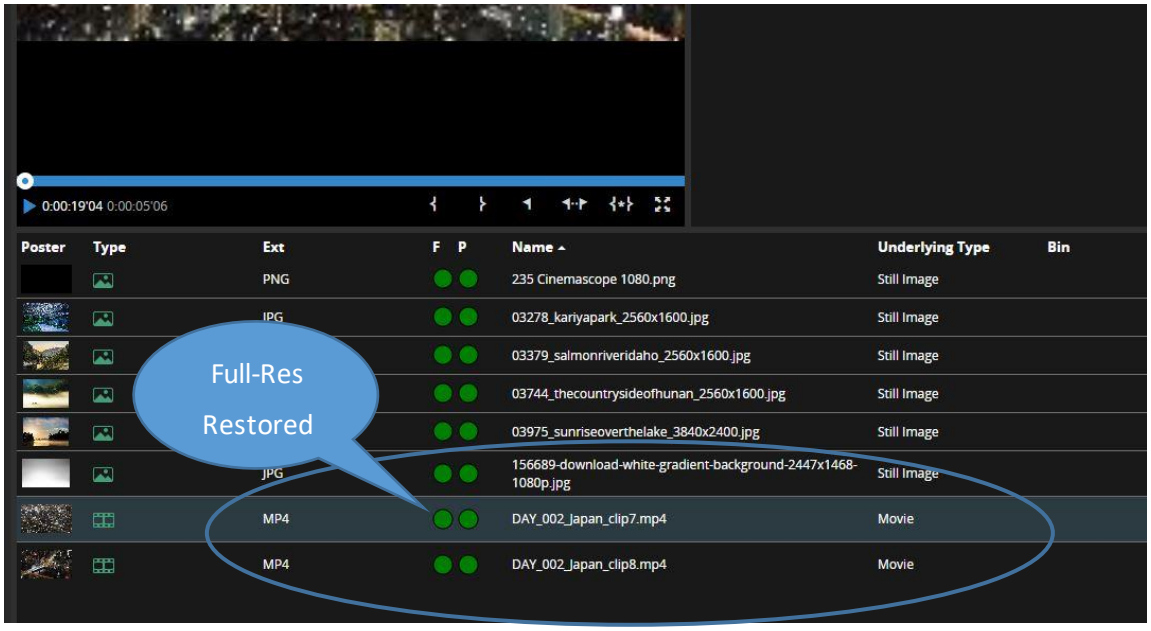


4. Confirm that local file versions on work in progress storage are removed and marked 'off-line.'



5. Restore the file from Archive within CatDV
    a. In client or Web Client select Tools > Schedule restore from Amazon S3 (compatible S3) > select “preconfigured” Bucket > select ok

6. Verify the Asset’s health in CatDV



Considerations

When considering implementation of this workflow by sales, presales, and solution architects, it is recommended to conduct a thorough inventory the customer's unique workflow desired, current and future needs, and the existing systems and network in place. What does the capacity and speed of the current installation look like? What is the growth rate of content, project demands? For new installations, be prepared to architect a solution for immediate concerns but know the growth and sizing steps likely to be needed up to 6 years of projected future demand. What does the landscape look like in terms of expandability of storage capacity, performance requirements, asset management growth? Will the production facilities of tomorrow look different – perhaps emphasizing more dispersed, local facilities in different regions rather than very large, consolidated production centers? Will there be a higher emphasis on NVMe storage locally, and object storage everywhere else? Will cloud delivered content analysis solutions and API-level reporting or solution integration be needed? Look for future opportunities to expand capacity and performance as the operation grows, and seek to build the infrastructure capability for the future today, with a flexible mindset, and platforms that can help you respond to ever-changing needs.

Summary

A solution based on Quantum StorNext, CatDV and ActiveScale provides a wide range of options and a highly flexible framework to meet a customer's asset management and protection needs for the tremendous amount of assets being gathered today. With ActiveScale as a target, you can easily deliver an on-premises private cloud, a highly geo-distributed or 3-site configuration to provide extreme levels of data availability, resiliency and can be further integrated with the Quantum StorNext shared storage platform for further content creation and production workflow tuning.

References

The documents below are references to configure the software and the systems for functional testing

Document Title	Download URL
ActiveScale Documentation Center	Link Here
CatDV Server installation	Link Here
CatDV Worker Manual	
CatDV Client Manual	

Version History

Version	Notes	Date
1.0	Initial Release CatDV and ActiveScale Reference Architecture	August 2021

Quantum