Technical White Paper

Nodeum, Active Archive

Manage data archiving and reduce complexity to access unstructured data

Nodeum is a software which manage data mobility and data organization when data volumes and capacity are exploding. It unifies the management of all data movement workflows. With logs and reports for each task, the solution provides full visibility and control to ensure the security and integrity of data.

In addition, Nodeum provide a large-scale posix compliant and scalable file system integrated with Data Mover and VFS technology to manage exabytes of unstructured data.

Built around a micro-services architecture that scales to an unlimited number of nodes, the solution is an answer to data-intensive computing environments. It scales data movement, simplifies long-term data operation and management, and improves accuracy levels for data movement operations.

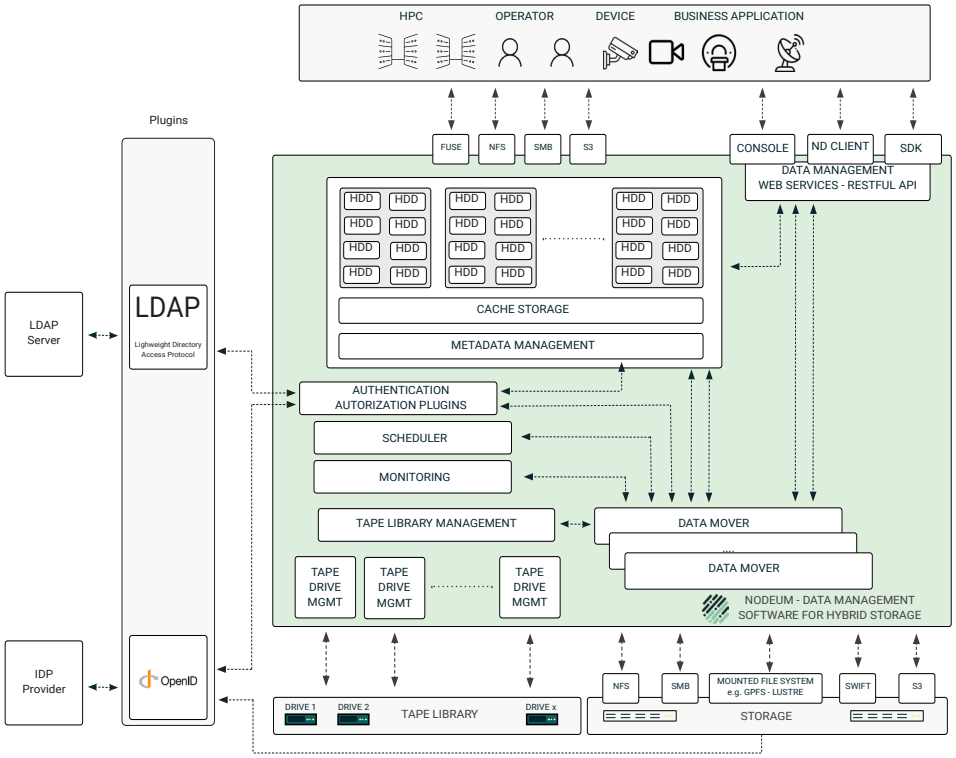
Nodeum combines leading technologies to create a powerful data management solution that openly integrates different types and classes of storage such as mounted file systems (GPFS, Lustre,...), NFS and SMB enterprise storage, S3 -Swift Object Storage cloud storage and finally Tape Storage.

Unstructured Data Archiving

Nodeum's Active Archive feature offers a multi-protocol storage access solution for long-term data archiving. It effectively addresses to the growing demand for a cost-effective storage tier for infrequently accessed cold data, which is typically stored on flash or disk-based platforms. By utilizing Active Archive, enterprises can reduce their on-premises storage costs.

Additionally, this feature serves as an alternative to cloud storage, enabling companies to decrease recurring storage fees and expensive egress fees. It provides an on-premises cold data repository that functions similarly to the cloud.

The Active Archive module, when combined with the overall Nodeum solution, offers a unified data management platform for organizing data flows between primary and secondary storage. It supports multi-protocol interfaces for seamless data ingestion and retrieval. This capability addresses the primary objective of managing data archiving, regardless of whether the original files are stored on NAS, S3 Buckets, or even Tapes as Secondary Storage



Nodeum software is an open platform which enables our customers to focus on their business values. The storage abstraction and orchestration layer design a hybrid data services on top of different hardware and silo technologies such as flash, spinning disk, tapes, cloud. All behind a unique file repository. This allows to manage, automate, and facilitate the archiving of massive oldest data storage created by business data workflows.

Workflow Data Management

Nodeum empowers to establish seamless routines and workflows, streamlining the efficient transfer of data across various classes of archive storage. The definition of retentions and multiple storage classes is essential to your data archiving operations. allows organizations to better control their data. Our categorization of archives encompasses three key segments:

1. Immediate Access Storage Pool:
   Archives demanding instant access and peak performance.

2. Infrequently Accessed Data Storage Pool:
   Archives not requiring immediate access but necessitating flexibility for retrieving expansive datasets. Take advantage of the Flexible Retrieval feature, enabling retrieval within minutes or bulk retrievals within hours.

3. Long-Term Archive Storage Pool:
   Ideal for compliance archives and digital media preservation. Equivalent to Deep Archive, this option is the epitome of cost-effective storage, offering data retrieval within a day.

GETTING ON BOARD A USER JOURNEY WITH NODEUM INVOLVES THREE STRAIGHTFORWARDS STEPS

- Specify a required storage pools for access.
- Obtain your access credentials.
- Define your desired data retention period.

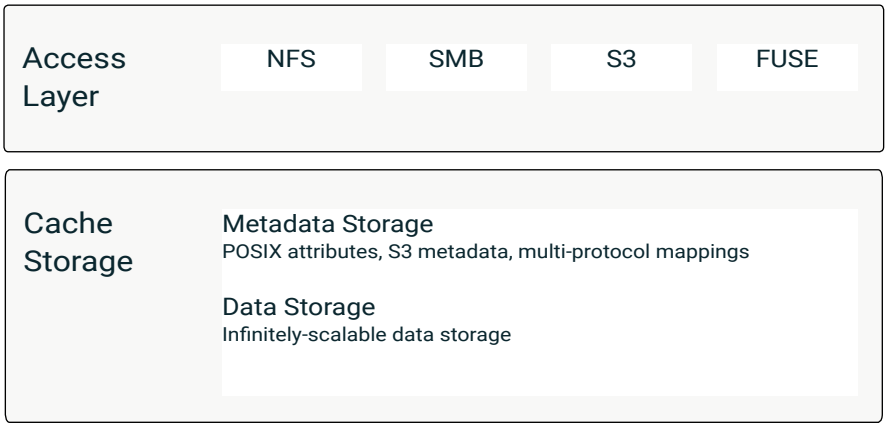
By defining data retention periods, automatic data movement between diverse storage pools is ensured, accompanied by variable retrieval speeds based on file types and the chosen storage pool. Automation orchestrates data flows across disparate storage pools while upholding data accessibility for users. Additional advantages encompass structured data organization by file type, accessibility of files, consideration of file creation dates, and more.

Regardless of the storage medium employed — be it NAS Storage, S3 Object Storage, or Tape Storage — user archives remain perpetually accessible.

> Elevate your archiving experience with Nodeum where data movement becomes effortless and accessibility paramount.

Benefits of using Multi-Protocols Storage Access

Nodeum does not store data in a traditional file system or object store but uses an abstraction layer that serves to present data to traditional file storage protocols (NFS, SMB) as well as to present data, which may be the same, to the S3 object storage protocol.



SMB

Nodeum supports SMB storage protocol and provide a reliable and efficient way for users to access and share files stored on a repository seen as a Network Shared Folder. Its versatility and ease of use make it an ideal solution for collaborative work environments. Its advanced security features help ensure that data is accessed securely. Overall, Nodeum's SMB support is a valuable feature that enhances the platform's capabilities and provides users with an efficient and secure file sharing solution.

One of the primary advantages of using SMB is its ease of use. It is supported by most operating sys-tems and devices, including Windows, macOS, Linux.

NFS

Nodeum supports NFSv3 and NFSv4 protocols. It allows for easy and efficient file sharing across a net-work. With its support for both versions, users can choose the protocol that best suits their needs, whether they require an established and reliable protocol like NFSv3 or the advanced se-curity and per-formance features of NFSv4.

Overall, Nodeum's NFS support provides an ideal solution for organizations that need to share files across multiple devices while maintaining security and efficiency.

S3

Nodeum supports the widely recognized S3 protocol format for seamless communication with object storage. It uses S3 compatible APIs for data operations to enable easy and seamless integration with existing object storage platforms and applications that can interact with S3 compatible storage.

In addition, S3 Data Lifecycle, Versioning, SSE encryption, S3 Glacier extension extend the capabilities of solution to manage and archive contents. Storing data to tape could be done in defining the storage class as S3 Glacier storage class which will be directly interpret that the data must be stored into a “Cold” Storage tier category.

S3 Glacier Storage Class will also allow the option to restore the contents from this tier prior to access it. This mechanism avoids timeout issues and unnecessary overhead on applications in case there is latency associated with restoring data from tapes.

Leveraged any type of secondary storage to offload data behind POSIX compliant integrated Virtual File System

Mounted local file system

Nodeum extends its support for storage mounted by their proper client in addition to the storage types it already supports. This feature enables Nodeum to perform data archiving operations on a wider range of storage types, providing greater flexibility and versatility in managing data.

It can support storage devices that are directly connected to a client and mounted on a local directory. By supporting storage mounted by their proper client, Nodeum can provide more comprehensive data management capabilities and increase the efficiency of data transfer operations.

SMB file storage

Nodeum's support for SMB storage protocol provides a reliable and efficient way for users to archive and stored data on network. Overall, Nodeum's SMB support is a valuable feature that enhances the platform's capabilities and provides users with an efficient and secure file sharing solution.

One of the primary advantages of using SMB is its ease of use. It is supported by most operating sys-tems and devices, including Windows, macOS, Linux. With Nodeum's support for SMB, users can easily access files remotely from any NAS or SMB compatible device.

NFS file storage

Nodeum's support for NFSv3 and NFSv4 protocols allows for easy and efficient file sharing across a network. With its support for both versions, users can choose the protocol that best suits their needs, whether they require an established and reliable protocol like NFSv3 or the advanced security and performance features of NFSv4.

S3 object storage

Nodeum supports S3 protocol for seamless communication with object storage. It ensures a reliable and secure connection with a variety of object storage providers, whether public or private. Nodeum boasts an impressive compatibility matrix that covers a wide range of technologies and vendors, ena-bling users to effortlessly integrate Nodeum into their existing storage infrastructure.

The compatibility matrix highlights the versatility of Nodeum and its ability to communicate with vari-ous storage providers such as AWS S3, Microsoft Azure Blob Storage, Google Cloud Storage, and many more. With Nodeum, users can confidently manage their data arching workflows, knowing that their storage infrastructure is fully supported.

Users have the freedom to choose the cloud provider that best suits their needs, and can easily switch to another provider without the worry of being locked in. The solution provides a complete control over the frequency of data movement, allowing them to schedule it following their requirements.

Data encryption capabilities provide an extra layer of security for data in transit. Overall, Nodeum provides a highly flexible and secure solution for customers to manage their data across different cloud providers, without any restrictions or limitations.

Swift object storage

In addition to the support of S3 protocol, Nodeum supports Swift compatible including its integration with the OpenStack Swift object storage platform. Nodeum Swift storage configuration include the full support of the different version of Keystone, the authentication extension of Swift Storages.

Nodeum's integration with OpenStack Swift allows users to leverage the scalability and durability of Swift storage. Users can easily create and manage workflow that copy or move data from and to Swift Object Storage containers and objects using Nodeum's interface or API.

Tape storage

Nodeum includes a service to manage natively all communications with Tape Library. This includes native tape automation including tape movement across slots, drives, i/o module. Tapes and Drives statistics information are retrieved to enhance the monitoring level and the reliability of the Tape usage.

This service is responsible to manage the drive allocations, monitor Tape Library and reports status to required Nodeum services.

Nodeum uses the LTFS open format to write data on magnetic tapes.
The LTFS format (Linear Tape File System) is a format supported by the drive and tape library manufacturers which are the founders of the LTO Ultrium consortium. This format has existed from the LTO-5 generation. It offers the unique feature to store both data and index data on tape.

A tape written with the LTFS format can be read independently of any storage system and offers a file system to access the data written on that tape. The LTFS drivers for any kind of operating systems can read the index and then execute functions like copy, update or delete. It is also possible to use a graphical interface for features like a simple drag and drop.

Unlike tapes generated by backup software, LTFS tapes can be removed from the appliance, shared, and stored. And, because they are stored following an open format under a common standard, other divisions, customers, partners, or workgroups can easily access data without having to purchase specialized hardware or additional software.

Nodeum is listed as one of the LTFS Partners. Moreover, it is important to know that Nodeum has been certified as a LTFS certified appliance by the LTO consortium.

Key Features

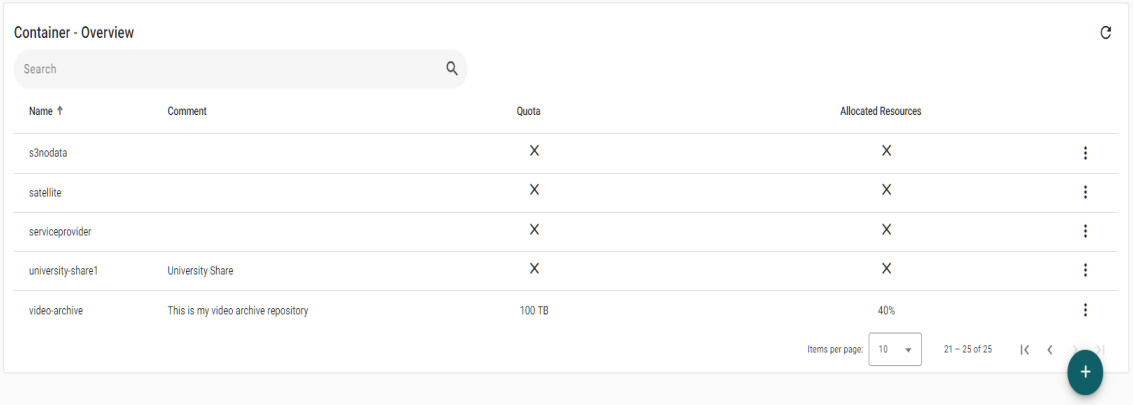
Container

Containers are storage space where any client can access to read, write, or modify any files. These spaces are accessible through three different protocols: NFS/SMB/S3.

The solution uses disk caching which allows fast access for all data transactions. This caching definition is flexible to fit into different use case.

Depending on the protocol, for SMB, the container is seen as network folder, for NFS, it is seen as target, and for S3, it is seen as a bucket. Set of access rights can defined for one or more users and/or groups of users. The authentication and authorization can be managed through different authentica-tion method, locally, with Active Directory/LDAP, with pair of keys.

Quota and disk allocation are settings to limit caching usage per container and global capacity. Nodeum allows creation of multiple different “Container” and multi-tenant.



Versioning

Nodeum supports versioning features of content stored in Container. The number of versions is defined per Container configuration and can be overwrite through API call or through S3 query. Write operations which would normally overwrite an existing file instead result in the creation of a new version of the file.

This allows protection from unintended overwrites and deletions while providing support for “un-doing” a write operation. Clients can then explicitly choose to list, retrieve, or remove a specific file version.

Data Lifecycle Management

Nodeum as a data management platform provides lifecycle management capabilities of any data managed by the solution. Data lifecycle refers to the different stages that data goes through during its existence, from creation to deletion or archival.

Available Parameters

- Storage Source / Storage Destination
- Action Types: Copy / Move
- File / Folder selector with inclusion / exclusion of files
- Integrity Verification (MD5 – XXHASH64)
- Filtering: On age / date accessibility of the files / On file size
- Task Priority
- Schedule: Once, immediately , Automatically (several standard time intervals)
- Planned according to a schedule

Nodeum provides different way to handle configuration of the data lifecycle policies:

NFS/SMB: The data lifecycle rules are designed through the Nodeum's task management. Per container, different rules can be defined and will be executed accordingly as 'task'. The policy configuration can be determined by either a configuration through the Nodeum GUI, or through an API call in using the Nodeum REST API.

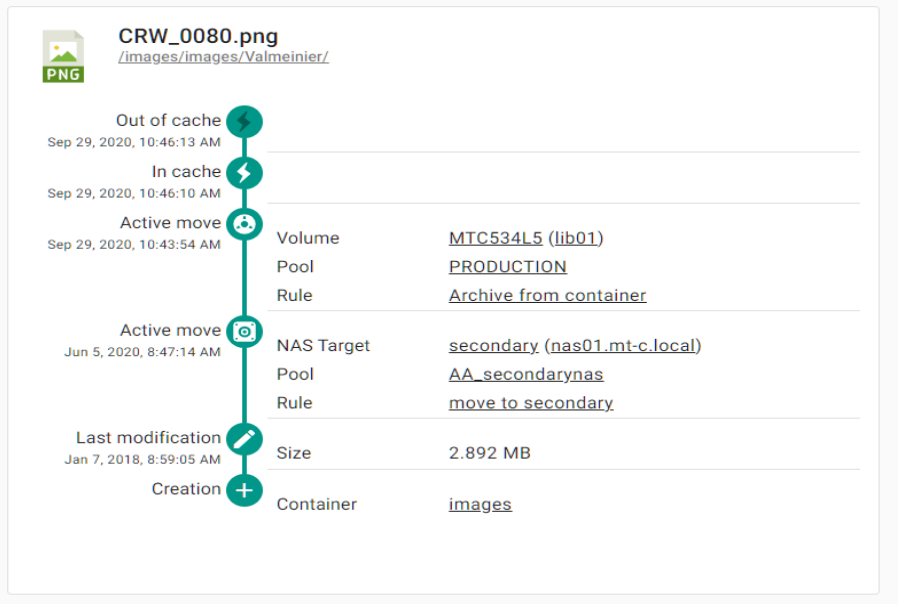
S3: the protocol includes extension to configure proper S3 Lifecycle. This is about a set of rules that define actions that Amazon S3 applies to a group of objects.

There are two types of actions:

- Transition actions: These define if the content has to be stored in "standard storage class" or to another storage class such as S3 Glacier Flexible Retrieval storage class
- Expiration actions – These actions define when the content expire, which means that the deletion date time can be scheduled on your behalf.

Nodeum's seamlessly integrates with workflows, offering the different S3 Storage Class It supports standard PUT, RESTORE, and GET primitives for easy management of data. With lifecycle policies, content can be automatically transitioned from Cache Storage to any compatible secondary Storage.

It integrates the capability of Nodeum to present Lifecyle of any data. Here's a general overview of the data lifecycle in the context of Nodeum:



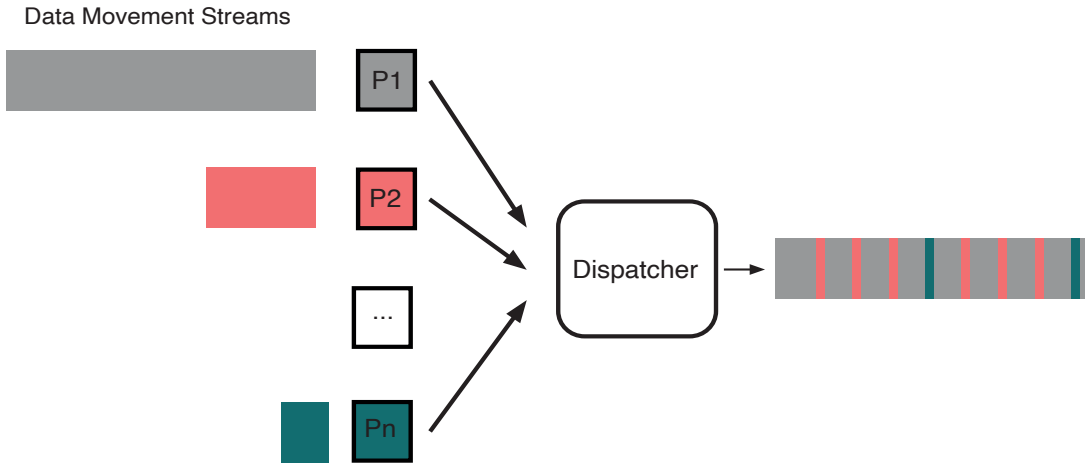
Priority Management

A priority management is provided into the workflow management to enable efficient and effective workflow management by allowing prioritization of data movement workflows.

This feature ensures that higher priority tasks do not consume all the available resources, leaving other workflows waiting indefinitely.

To achieve this, Nodeum uses Quality of Service (QoS) techniques to prioritize data movement operations based on their priority level. This approach ensures that critical demands receive the necessary resources, while still providing fair access to non-critical demands. By assigning different priority levels to different tasks, Nodeum ensures that resources are allocated based on their priority level.

Additionally, Nodeum uses Fair Queuing techniques to allocate resources to different requests,



ensuring that each remaining request gets an equal share of the remaining resources. This method helps prevent one data movement workflow from taking up all the available resources, ensuring that all remaining workflows get fair access.

Overall, the priority management and QoS techniques implemented in Nodeum ensure efficient resource allocation and workflow management, preventing resource contention and ensuring that all workflows receive fair access to resources.

Metadata Management

The Metadata Management implementation doesn't require any database. This is a design choice made as a major factor to ensure the Nodeum' s ability to scale in a fault tolerant manner across thousands of servers. Instead of a database, Nodeum uses consistent hashing and the file system to store all Metadata information.

Metadata are stored into the file system through a serialization format, MessagePack format which keeps the extensibility of JSON, by allowing keys to be added/removed.

For each file, a list of metadata is automatically managed:

- File size,
- File permissions,
- Access time,
- Version,
- Data lifecycle.

Metadata Management Benefit

There is no direct operations on the secondary storage such as tape or s3 when the metadata of the file system are changed. There is background process which will update the secondary storage (cartridge(s) or bucket) in a second step. Then, there is no operation done when end-user move – rename files.

Multi-tenancy

Nodeum supports multi-tenancy at the user level, simply by creating multiple accounts in the system. Each account will be allowed to can issue its own set of credentials including username - password or pair of keys.

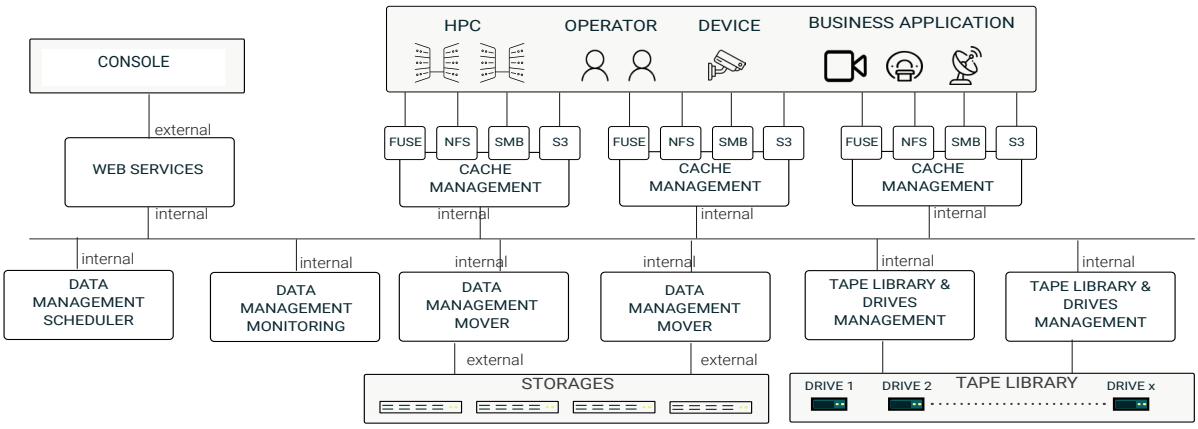
Credentials allow granular privileges to access the defined storage and execute appropriated actions. This depends on the protocol types. With the appropriated roles and permission, S3 includes creation of buckets and data access to a user. A bucket created by a user with a set of access keys issued by one account will not be visible to another user accessing the system with a different set of access keys that were issued by a different account.

By utilizing different accounts, the solution can be used in a multi-tenant environment, for data segregation between applications, departments, internal and external customers.

Pool Management including HSM Management to integrate multiple backend storage into a single name space

Nodeum manages its storage abstraction behind pool manager. A pool is a logical mapping and grouping of storage resources, such as NAS, tape libraries, or cloud storage.

Each pool has a type which define the role of storage. This role includes different options such as: Primary Storage, Active Archive. Each type determines what the storage will be used for. The pool also includes the mapping with the final storage resources which differs depending on the type of the storage and the protocol used to access it.



In this context, any executed Data Movement workflow always move data between Pools. In this context, pool will be used to defined where the data written to the defined Container, will be to store data.

NAS: It will allow mapping between the pool and a Network Shared Folder if the protocol set is SMB; and this will allow mapping between the pool and a NFS Target if the protocol set is NFS.

Cloud: The mapping allows the link between the pool and a bucket.

Tape: Manage multiple tapes behind a same pool. In this type of configuration, during the execution of a task workflow, once a tape is full, the system will automatically unload/reload a new tape to continue its writing process.

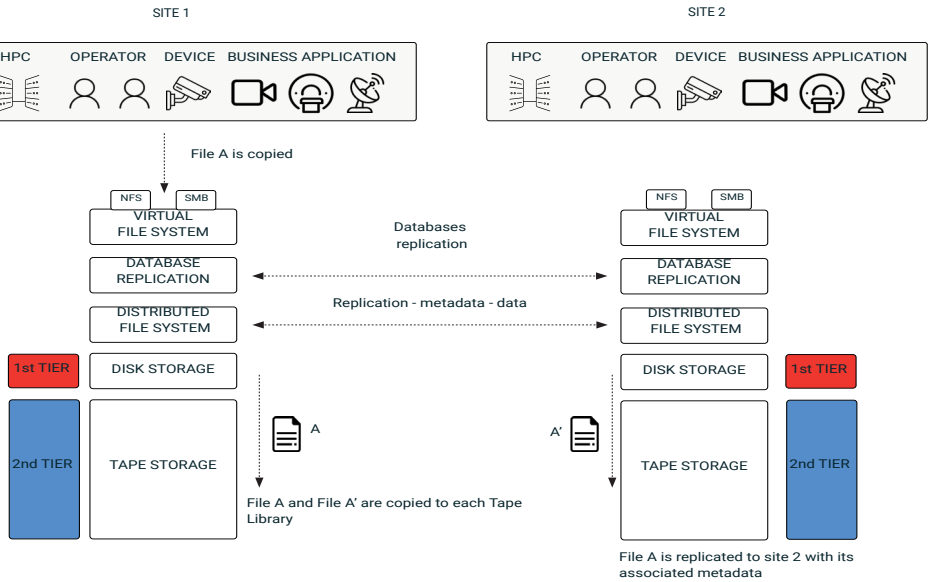
By using this feature, multi-tenant data management can be further enhanced in case physical data segregation by storage pool is desired for different accounts, or by Container.

Tape Management - Mutliple Copies Automation

Nodeum provides the capability to handle multiple copies of data automatically and on separated storage. Different pool type can be used for allowing the different copies. For Tape Storage, it allows the definition of multiple sets of tapes.

The Archiving Workflow policy can be allowed between 1, 2, or 3 copies, which can be set up across multiple pools. Copies are created simultaneously on separate storage. As a best practice for data protection, applying the 2-copy or 3-copy policy is highly recommended. In case of a copy is lost or becomes unusable for any reason, a new copy can be created by copying objects from another storage copy, to recover the defined policy. The 1-copy policy still remains an option.

Example: Multiple tape library pool in multiple sites.



Tape Rehydratation

The tape rehydration process is design to optimize the space on a tape. When files are deleted from a tape, they are not completely erased. To completely erase the files, a rehydration task is required.

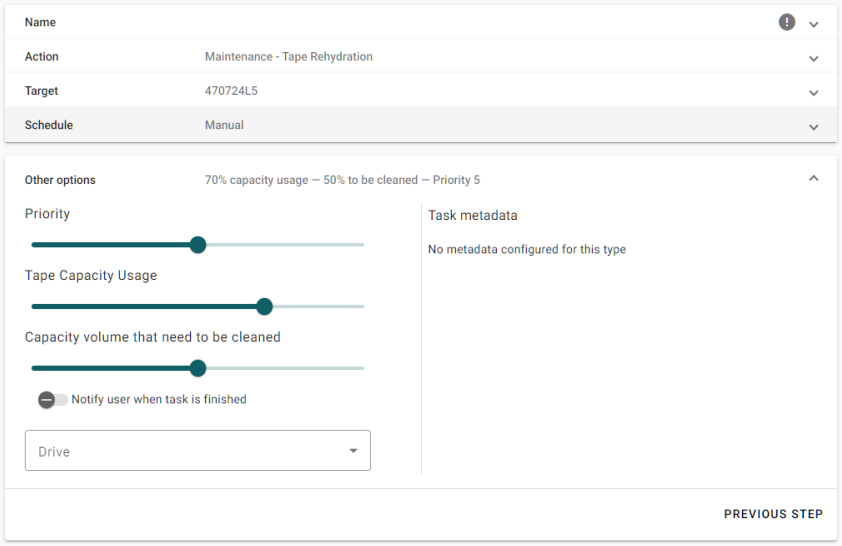
This process enhances performance and improve the end-user experience.
In these situations, Nodeum does not delete or modify the files stored on the tapes. For deletions, the entry in the catalog is removed, and for updates, the new version is written while the index pointer is changed. The files are never physically deleted from the tape.

Therefore, when a certain number of file deletions or updates occur, the ratio between the available free space and the actual consumed space on these tapes becomes imbalanced.

This process ensures the physical deletion of previously deleted data on tapes and can be used for regulatory compliance such as GDPR.

Tape rehydration can be executed by selection of individual tapes or by selection of a group of tapes. Alternatively, tape rehydration can be set to be executed automatically, by setting threshold on the percentage of tape capacity totally or on the percentage of used deleted capacity of a tape for the rehydration process to be initiated.

The process is done by reading all files back to Nodeum server cache storage, formatting the tapes, then writing back the remaining files to a new tape. This process takes place in the background when the system is idle and will be paused as necessary not to interfere with normal write, retrieval opera-tions.



Automatic Tape Generation Migration

Tape Migration from a generation to another can be a lengthy process that requires attention, time, and resources. At Nodeum, a feature exists to automate all this migration effort in the background.

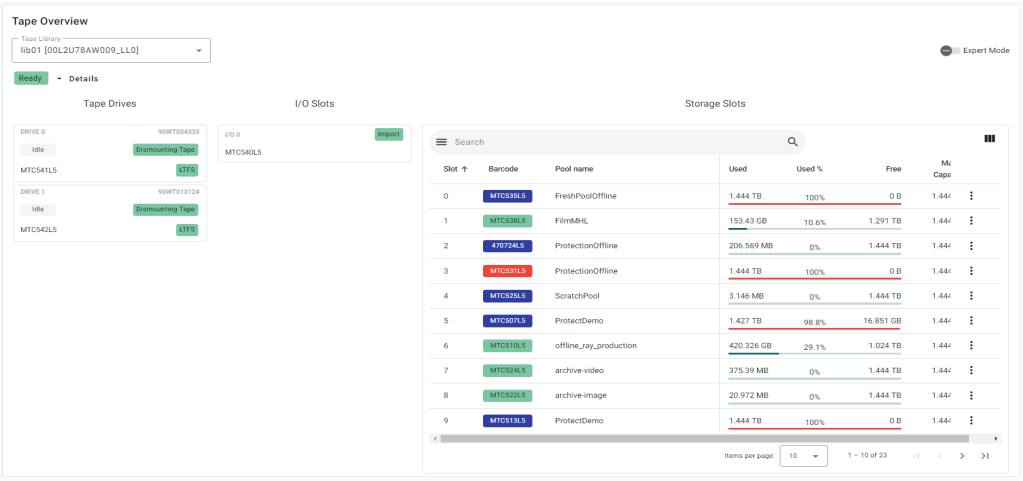
As a prerequisite, the tape library (or libraries) managed by Nodeum will need to be equipped with a set of tape drives that can read the old generation tapes and another set of drives that can write to the new generation tapes. It is recommended to ensure an ample number of tape drives are available to support the data migration procedure as well as normal operations.

As an option, the old generation tapes that have completed the migration process can be logically formatted for disposal. Same as the tape rehydration feature, the tape generation migration process takes place in the background when the system is idle and will be paused as necessary not to interfere with normal write, retrieval operations.

Tape Library Monitoring and Management

Nodeum is natively designed to efficiently manage tape library, it supports popular tape library brands and models, ensuring compatibility and interoperability across different version and generation.

The solution includes a full set of Tape Lifecycle Management, this facilitates comprehensive tape lifecycle management, it enables tracking and monitoring of tape usage, including information on tapes' health and utilization. This ensures optimal utilization of tapes, effective rotation, and timely replacement when necessary.



Nodeum provides a centralized management interface for overseeing and controlling tape library operations. Administrators can monitor Tape Drives, Tape Library Slots and I/O module utilization, track performance metrics when tapes are doing read-write operations in tape drives. This centralized approach simplifies management, ensures a door-to-door control of tape usage operations.

User Interfaces

ND Client

ND Client is a command-line interface (CLI) provided by Nodeum for managing the system and performing various tasks. It offers several benefits to users, such as:

FLEXIBILITY

ND Client provides a flexible way to interact with Nodeum services, allowing user to execute commands, scripts, and automate workflows. It enables users to perform various tasks in a customizable way, depending on their specific needs.

EFFICIENCY

With ND Client, user performs multiple tasks simultaneously or in rapid succession, which increase efficiency and productivity.

CONTROL

It provides a high degree of control over the system, allowing user to manage and monitor various aspects of Nodeum, including data movement operations.

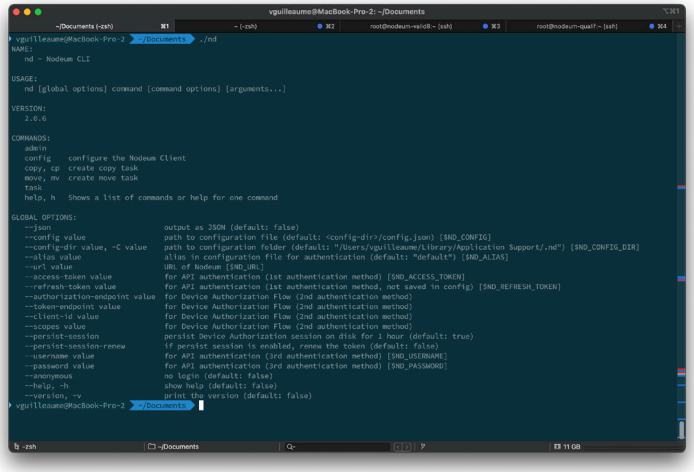
SECURITY

It runs on a secure and encrypted channel, which ensures that all communications between the client and server are encrypted and secure. This feature protects sensitive data and prevents unauthorized access to the system.

COMPATIBILITY

ND Client is compatible with different operating systems, including Linux, macOS, and Windows. This compatibility enables users to access and manage Nodeum from various devices and platforms.

In summary, the ND Client provides a flexible, efficient, and secure way for users to manage and interact with Nodeum, allowing them to automate workflows, manage storage, and monitor the system with ease.



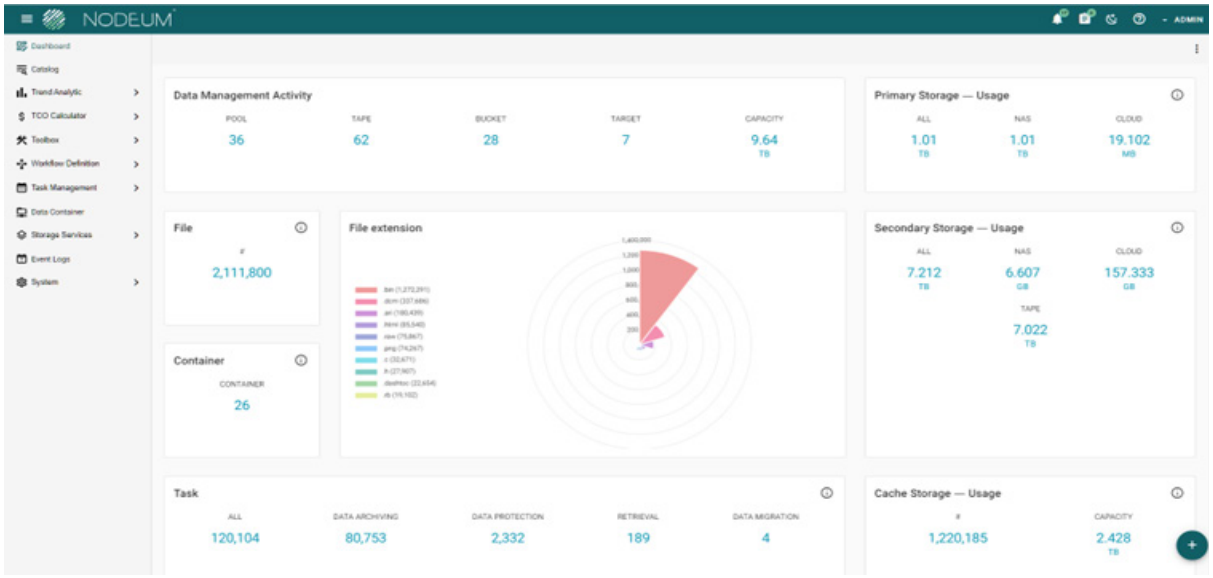
Console

A Web console, HTML 5 is available to get access for User to manage their data from anywhere. It sim-plifies the data management operations and making it easy.

The Console provides a global centralized view of all data archives, regardless of where they are stored. It is easy to navigate through their data archives using a familiar file structure and can also search for specific files using metadata tags or file names.

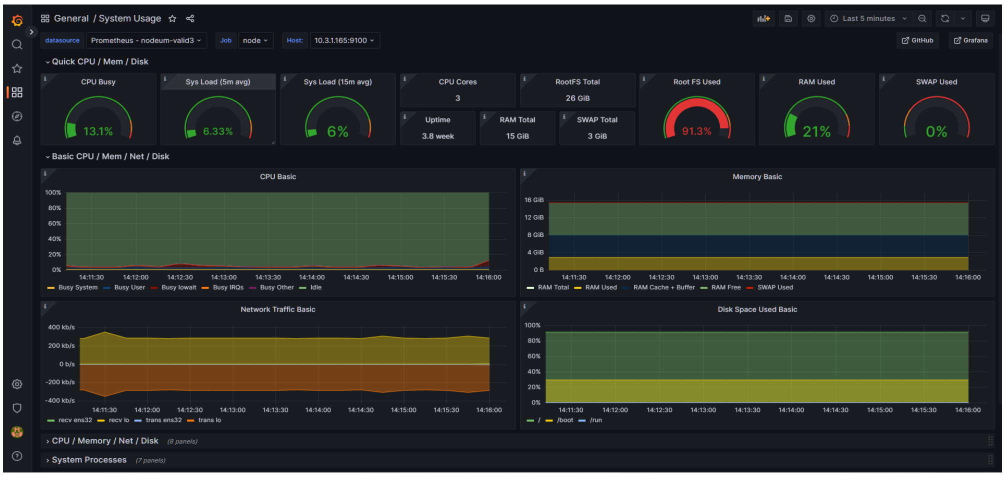
For example, users can set up data movement tasks, view task status and history, and monitor data usage and performance.

Overall, the Nodeum Console is a powerful tool that provides users with a centralized and user-friendly interface for managing their data archiving. Its intuitive interface, robust manage-ment tools, and customization options make it an essential component of any data archiving strategy.



Alert & Monitoring

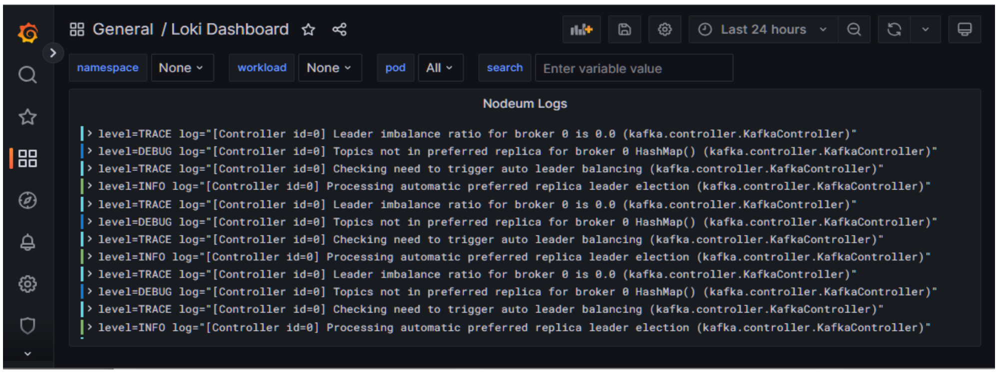
Nodeum measures the status of all cluster nodes, including a set of metrics for system resource utilization. This data is stored in a local Prometheus database, guaranteeing long-term retention. These metrics can then be exported to Grafana visualization tools.



As part of the alert & monitoring capabilities, Nodeum services generate continuous logs. All of these logs are centralized in a dedicated service.

Using this feature, the user gains insight into all operations executed within the solution. Each log is associated with a tag that refers to the process and with a level. This helps the user to sort and filter each event and find out information to answer a specific case.

In addition, the solution allows integration with Grafana/Loki to export the generated logs to this centralized log management platform.



Discover right now Nodeum.

Follow this link and start the adventure with Nodeum !

www.nodeum.io

Nodeum's Key Features

- Policy-Based Workflow Manager
- Hybrid Storage Management (NAS, Object, Cloud)
- Software-Defined
- Microservices Architecture
- Highly Scalable
- RESTful API
- Console
- ND Client
- Monitoring

Summary

Nodeum Active Archive offers a solution for creating archives that help optimize storage utilization and cost while keeping all your organization's data accessible online. By migrating infrequently accessed cold data to tape, you can free up valuable space on expensive primary storage or object storage, ensuring a balance between data access speed and storage costs.

Identifying and categorizing cold data may initially seem challenging, but there are automated approaches available. Data management software, often referred to as a data mover, can assist in identifying cold data based on criteria such as data age or time since last access. It enables the automatic movement of such data to the archive.

Alternatively, many existing products have built-in features that allow you to define policies for archiving data to external storage. These features enable automated archival based on preset rules and streamline the process of moving data to the archive.

Nodeum Active Archive provides flexibility and automation in managing cold data, allowing organizations to optimize their storage resources effectively while keeping data readily accessible.

NODEUM®

About Nodeum

Nodeum is data management software, agentless, running on commodity hardware. The solution manages and abstracts storage tiers: NAS, Cloud, Object & Tape. The objective is to automate the data management process, integrate it into customer business logic and connect end-user pipelines with any storage system for data movement, archiving, and protection. It achieve high-performance workflows to gain control of exponential data growth. And save time with a single, intuitive storage and process management software that works across multiple storage silos.

Do you have any questions?

Get in touch!

Further information on **www.nodeum.io** & **sales@nodeum.io**

H.Q. Office

+32 4 264 23 74