commvault
solving forward™

A CommVault White Paper:

Data Archiver Modules

Active Archiving for Storage Management, Compliance, and Discovery

CommVault Corporate Headquarters
2 Crescent Place
Oceanport, New Jersey 07757-0900 USA
Telephone: 888.746.3849 or 732.870.4000

CommVault® Data Archiver: Active Archiving for Storage Management, Compliance and Discovery

Table of Contents

Data Archiver

Data Archiver™ (DA) is a hierarchical storage management (HSM) solution that allows an organization to reclaim critical primary storage space by moving and automatically recalling data from secondary storage for active use. By applying policies to select, move and retain fixed content[1], organizations gain significant economic and productivity benefits. It allows companies to address aggressive data growth by redistributing older data across a tiered-storage architecture. CommVault® Data Archiver's transparent, in-place recall mechanism maximizes data accessibility for users, without imposing the need for extensive training, physical staging or catalog maintenance over the data's lifecycle. With this, administrators maintain more efficient primary storage architectures. They reduce the costs of retaining and accessing fixed data by aligning it with less expensive forms of tiered storage.

As part of the CommVault® Singular Information Management™ (SIM) platform, Data Archiver fully complements data protection strategies for Exchange, Lotus Domino, SharePoint, Windows file systems, Unix file systems or Network Attached File Share environments. The impact from migrating data is immediate: a substantial reduction in the production environment's physical storage space directly translates into reclaimed disk capacity and improvements in the speed and size of backup and recovery operations.

Longer-term benefits include reduced problems related to user quota management (such as out-of-disk or the excessive usage of personal archiving PST file or .nsf alternatives) and the ability to archive data for retention or reference purposes[2]. Although Data Archiver is fully deployable in a standalone mode (for usage in non-CommVault® Galaxy™ Backup environments), it leverages the CommVault Common Technology Engine to deliver very cost-effective and automated data archiving services. Please refer to product documentation for agent support along with specific configuration and deployment parameters.

[1] Fixed Content represents data objects that have both reached final form (no further versioning) and aged beyond a requirement for immediate availability on fast access primary disk. At that point they become candidates to better manage when stored on cheaper media outside the primary system, while retaining the ability to access any item within a reasonable response time (correlated to media formats and system layout).

[2] DA rules can be set to provide a fundamental archiving role of select data files. This is used with the "no-stubs" option and data will be moved off the primary store and deposited in a retention-based storage policy. Access to the "archived" data is provided through the browse & search application in the CommCell Browser. Note that compliance or regulatory Email Retention is not the role of DA; rather it is the special role of Data Archiver Compliance Option.

Strategic Value

Data Archiver helps deliver on the following strategic objectives**.**

Information Lifecycle Management (ILM) – Data Archiver is a critical lynchpin within CommVault's ILM strategy. It provides an automated ability to seamlessly move data across tiered storage with independent retention policies and automatic recall. Independent media retention cycles with scheduled pruning allow you to align data against a lifecycle model (balancing access vs. cost). Data is available for automatic recall from any active retention copy over mixed media. This eliminates the need to restage that data prior to recall triggered by a pull (stubs) or push (browse/recover) mechanism. Data Archiver makes ILM a reality because it manages and moves data as it ages and becomes fixed in nature without sacrificing direct access for your users or applications.

Application Availability – Data Archiver dramatically reduces application recovery time in the case of data corruption or loss. It reduces the actual size of the data store by moving/stubbing aged, fixed content to secondary storage. Data Archiver ensures that users retain continuous access to data in the original file or message location.

Complete Data Management Solution – Customers understand the need for integrated software solutions that offer advantages like reduced total cost of ownership, simplified administration, lower acquisition price and faster deployment due to its less complicated architecture and single software stack.

Enterprise Wide e-Discovery - With the recent amendments to the Federal Rules of Civil Procedures (FRCP), organizations are realizing the Electronically Stored Information (ESI) no longer means just email. Data Archiver allows for retention policies to be defined in one application and can span all data types and repositories. This ensures proper data retention and disposition is managed but more importantly, in the event of litigation, allows for a legal retention hold to be placed across all relevant data all from within one user interface. Built-in audit reporting provides the proof needed to ensure proper chain-of-custody.

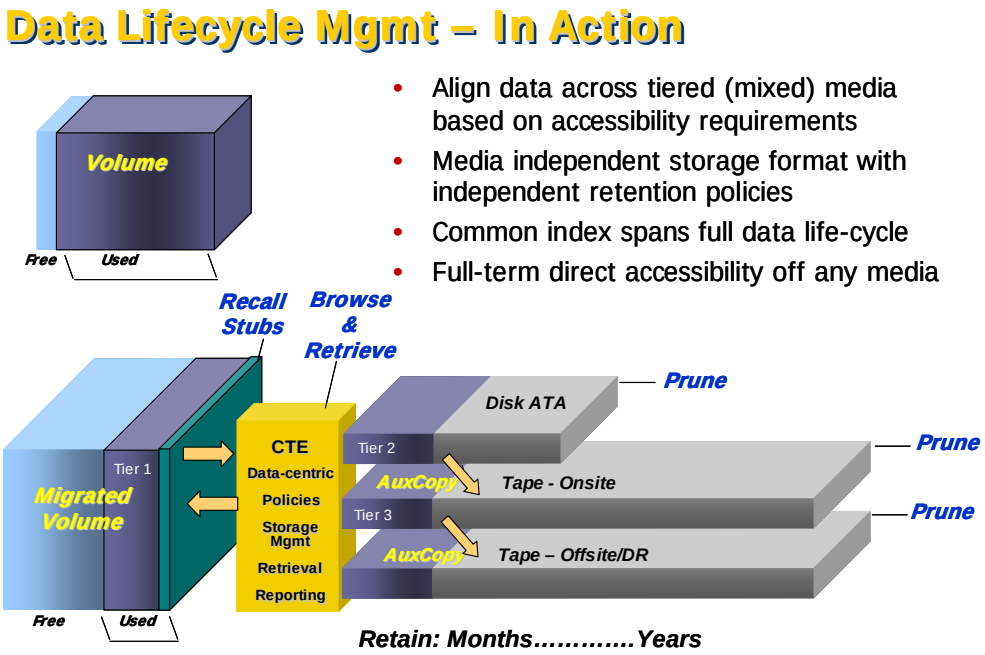
Challenges Solved by Data Archiver

Multi-Tiered Data Lifecycle Management

A Data Archiver scenario is constructed around two primary sets of adjoining policies; sub-client and storage policies. A set of sub-client policies is used to assign scanning ownership of content sources, find candidate data for migration based on aging, sizing and filtering rules and compress/encrypt/migrate[3] the data to an assigned storage policy. Sub-client rule classes are based on the supported Data Archiver agents. The storage policy is a management component of the Common Technology Engine (CTE) which underpins Data Archiver. Storage policies assign media, retention policies, auxiliary copy hierarchies and data verification policies. Storage policies afford organizations with unmatched flexibility in easily configuring multi-tiered backend storage environments behind Data Archiver. This capability guarantees that the organization can maximize the benefits of broader archive scenarios by managing the complete lifecycle of the archived data as it moves across a heterogeneous, multi-tiered secondary storage environment.

Although any of the CTE-supported media formats may be employed to store the archived content (i.e., disk-DAS/SAN/NAS/CAS, optical, tape, WORM), we generally recommend employing magnetic media as the initial storage location (tier-two) to ensure the greatest degree of recall transparency for end-users. A subsequent storage tier (tier-three) may exist on tape with an extended retention period relative to the disk copy. The advanced indexing feature within the CTE ensures that all stub-recalls may be serviced from data residing on any active tier in the backend. This provides critical data retrieval integrity to the end-user scenario by infusing redundancy and high availability of the data across all the tiers of storage (i.e. if disk storage on tier-two is lost, direct retrievals are automatically serviced off tier-three/tape). The CTE data aging (a.k.a. pruning) policy completes the data lifecycle management process by automatically deleting/expiring data after it completes its retention cycle on each of the distinct tiers. When the last instance of the migrated data is aged off the CTE, Data Archiver will automatically clean out the associated stub from the source system on the next job.

[3] Migrate: The migration process utilizes an indexing and archiving phase where the candidate data is moved out of the primary store, followed by the stubbing phase which replaces the original message/file objects with the recall stub objects. If the no-stub rule is used then an archiving-only job is implied. This requires all data access for restore to be conducted through the CV Browse, Search & Recovery application.



Archiving Email Data

Email data is steadily increasing in size and volume, driving administrators to turn to solutions that seamlessly employ policies to archive older messages and/or attachments to secondary media. Data Archiver provides advanced archiving rules that can be tuned to individual groups of mailboxes to align with specialized SLA requirements (i.e. mobile executives may use a 180 day age rule versus the 30 day default applied to the bulk of the organization). Extended across migration policies, single instancing of shared message and attachment data ensures efficient storage of archived data on secondary media.

When messages are archived, they are replaced with powerful, self-contained stub message forms that retain the key properties of the original messages. They consume a tiny footprint within the email information store (typically <1Kb[4]) and embed self-contained triggers for direct, in-place recall. Recall can be configured to automatically replace the stub with the original complete message by simply double-clicking; this compares with more difficult traditional email archiving solutions which require searching, selecting and accessing data outside the messaging environment. Data Archiver lets users organize and search a stub just like a regular message within their own mailbox folders.

[4] Attachment annotations and optional body text added to the stub can result in the actual size ranging between 3-6Kb.

After reviewing a new batch of messages with large attachments, end-users can easily tag select messages to add them to the next archiving job outside the normal aging/sizing candidate rule that are applied to the mailbox. Regular archive strategies dramatically impact the amount of active storage required per user. This enables Exchange users to retain longer access to an expanded collection of data items directly from their Outlook or OWA client. For Lotus Notes users environments, users can directly access from the Domino Journal mailboxes. A very valuable by-product of an effective email archivingstrategy is the reduced demands placed on users and administrators to conduct frequent mailbox management within the bounds of quota management. Aside from the productivity and welfare benefits this provides to overburdened users, active archiving helps contain and constrict the production of large collections of PST or .nsf container files by users.

Archiving SharePoint Data

In addition to messaging, proliferation of collaboration system data is also causing storage management burdens. Data Archiver for Microsoft SharePoint migrates infrequently accessed or older SharePoint documents and other objects from higher-cost primary storage to less-expensive secondary or tertiary storage, further improving bottom-line efficiencies. Keeping costs down is especially important when managing multiple versions of files within SharePoint (each file version is an independent copy in the SharePoint database). Data Archiver keeps users connected to their vital corporate knowledge assets within SharePoint with size or aged-based migration policies and keeps it available where it can be searched and accessed automatically.

Archiving File System Data

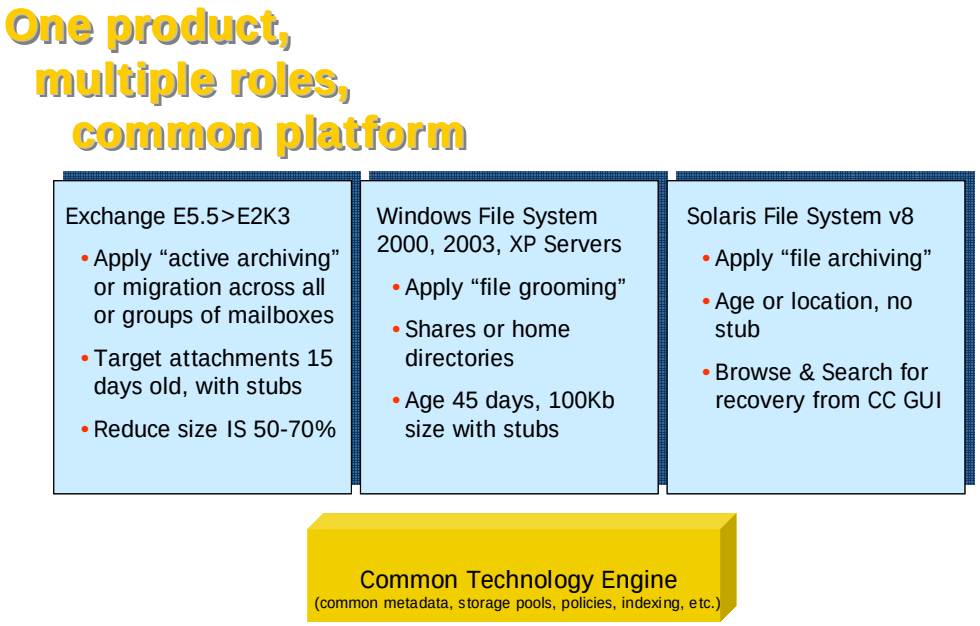
File shares are under constant assault as relentless data growth is driving an insatiable appetite for additionally expensive storage and heavier burdens in ensuring levels of application availability. With applications driving larger and larger file formats (i.e. Word documents now retain a full collaborative revision history) and enterprises undergoing extensive reorganization, consolidation and user turnover, open file systems represent vast stores of unmanaged, stale data effectively collecting dust on expensive spinning disk. While incremental backup strategies can be used to skip these fixed files in the nightly backup jobs, a hefty price is still paid in recovery time and overall backup media requirements. Broader server consolidation projects can further exacerbate these problems as larger stores of data are created – in many cases defeating the recoverability service level agreements for those servers due to the overall size of the data store. This condition is easily relieved with a strategy to select and move older data to secondary storage while retaining in-place access in the primary volume.

Data Archiver provides enterprises with a robust HSM solution to redistribute their file stores across a tiered-storage architecture. Data Archiver policies are unique compared to traditional HSM solutions, in that they work inside the existing file system directory structures. This means that archiving can be applied quickly to existing data locations, by archiving data where it sits, rather than requiring it to be moved to media staging folders[5]. This provides a more transparent experience for end-users and applications because the access point remains where they left the data and the stub retains the existing ACL security properties of the original file.

Data Archiver provides organizations with a powerful capability to groom their file shares and move older or orphaned data out to secondary media while preserving in-place (stubbed) access and delivering cataloged browse/search index restore across the retention period. The Data Archiver for File Systems agent can open a wider variety of deployment scenarios in providing ease and flexibility in moving data files across various media formats. Examples:

- Standard File Archiving: Archiving files based on age, use a tiered 12 month disk retention & 3 year tape retention = typical tiered model for longer term retention employing AuxCopy
- Groom Prohibited File Types: Use filter to exclusively include file type="*.mp3", archive to storage policy with 3-day retention policy; short retention policy washes out files and stubs after expiration
- File Archiving Role: Employ "no-stub" rule, target special locations and set archiving to capture all existing content, use browse & search features in the CTE to select and recover data to alternate location for review/use

[5] DA can easily be used in a traditional staged media folder role where special folder locations are configured on the share to collect files for movement to secondary media (typically users will be required to move files to these folders for migration). This scenario also aligns to an "archiving files" approach, where the no-stub rule will be used and access is only permitted in the authorized browse & search function in the CommCell.



Improving Utilization and Controlling Storage Costs

In many cases, organizations have already made sufficient investments in secondary storage architectures (i.e. DAS, SAN, NAS, Tape, and Optical) that can be shared with Data Archiver to reclaim vital space in primary storage arrays. This secondary storage is generally underutilized because of the inability to effect data management over the primary store by monitoring, selecting and moving data onto this secondary capacity. When Data Archiver is successfully deployed, it can help offset future hardware investments by better utilizing the existing storage architectures and employing libraries to support data access during normal working hours.

Business Continuity and Disaster Recovery

Today's business critical application must be available 24 x 7, 365 days per year. Users demand continuous access to data. This places a greater burden on the shoulders of the IT group as data growth continues to challenge their ability to successfully recover within a fixed window. HSM provides relief in ensuring that primary data can be recovered more quickly, since the actual amount of data in the store is reduced by a significant factor. How? In many cases customers report a 50% reduction in physical size of existing email stores from the simple act of executing a policy to archive aged attachments only. This translates directly to faster backups and recoveries. Employing the CTE and the common data storage format, Data Archiver enables organizations to easily protect archived data with advanced AuxCopy capabilities – mixing retention times and media formats for both on-site and off-site data protection. This

ensures the integrity and availability of the archived data managed within storage policies. Coupled with the inherent data protection features of the CTE architecture, Data Archiver provides a complete solution for customers compared to competitive archiving products that force IT organizations to make additional investments and integrate 3rd party bolt-on data protection or HSM tools.

Key Features and Customer Benefits

Data Archiver provides policy-based, data lifecycle management for email, collaboration and open file system data to reclaim space, reduce production store size, address quota issues and enable file Hierarchical Storage Management (HSM) and archiving services. The following tables outline the key Data Archiver features that help resolve space management challenges facing many organizations today. The following is NOT an exhaustive list of every product feature. Rather, it provides a comprehensive view of the key features that differentiate the product and deliver valuable benefits to the organization.

Data Archiver Key Features and Customer Benefits

Key Features	Customer Benefits
1. Advanced hierarchical data archiving with direct recall	• Provides automated movement of older data off primary store while retaining in-place recall • Target candidate data with rich, granular migration rules and scheduled jobs • Ensure transparent recall process for end-users and applications • Reduce cost of storing data and ensuring longer term accessibility for end-users • Reduce overall size of primary store – to improve backup/recovery and regular operations – without limiting user access to that data
2. Rich centralized migration and storage policy management	• Policies can be broadly applied or focused on select groups (Mailboxes or Directories). • Powerful multi-dimensional rules are easily configured and tuned • Centrally managed across all servers from the CommCell Console • Conduct “what-if” analysis with the DA Predictor tools

Key Features	Customer Benefits
3. Streamlined, in-place recall	• Transparent process for end-users and applications (low training, easier deployment) • Fast and direct recall of original data (no need to navigate to the 'archive' for retrieval) • Secure and authorized data access (retains original security model)
4. Consolidated data storage platform	• Cross platform DA agents can use shared media to reduce TCO • Reduces administration and operational management costs • Policy-based-simple-to-administer-low burden/high return for the organization
5. Open systems software architecture seamlessly integrates with existing or new heterogeneous environments	• Easily implement ILM – multi-tiered storage with full life-cycle/retention management • Minimal deployment risk or chance of obsolescence across storage architecture • Excellent TCO and ROI compared to individual point solutions. • Fits seamlessly into existing networked storage environments

Data Archiver Feature Highlights

A) Product Interoperability
B) User Experience
C) Archive Policies/Operations
D) Efficient Storage Management
E) Multiple Platform Support

Feature	Description
A.) Product Interoperability	
Common CTE Platform for all agents	The CTE platform provides a common layer of services that are applied to all DA agents. This includes a common integrated interface (CommCell Console) accessing the core services of the CommServe engines: • Common user experience ensuring ease of use -- Intuitive and application data-centric policies, centralized administration and transparent end-user recall. • Common policy-based data management and movement -- Streamlined and responsive centralized control, ensuring the highest-degree of data integrity for archived data. • Common content indexing for full data life-cycle management -- Distributed model to ensure scalability and accessibility • Common media management layer -- Multi-tiered, media independent, retention-based data storage format. Enables sharable media across all Data Mgmt products. • Data portability addresses technology obsolescence.

	• Lower Cost of Ownership -- Modular products integrated into common platform, delivers high reliability, quick to implement, easy to learn, provides efficient storage management
Exchange DA Agent	The agent is deployed on a server basis and sub-client policies are defined within the CommCell Console to target groups of mailboxes and apply candidate selection rules. All sub-client policies are maintained in the CommServe and processed by the iDA during archive jobs. No scripting is required. No rules are embedded in Exchange folders or directories. • Source content: Individual mailboxes are assigned to sub-client policies – auto-discovery is used to quickly assign accounts to sub-clients and it may be configured to automatically monitor MailStores and associate new accounts to a specific sub-client. • Migration scan rules are defined to trigger which mailboxes qualify for migration – this may be based on quota watermarks or overall disk usage on the Exchange volume. • Migration candidate rules are defined with aging and/or sizing rules. These are applied to all mailbox objects (except sticky notes and personal journal entries). Special rule combinations are available (i.e. attachments only, no stubs…). • Filters are provided to target inclusion or exclusion rules on attachment types, folder or mailbox users (for instance this can exclude Calendar and Contacts folders from migration scans). • Stub rules define the lifecycle of stubs within the mailbox – they may be retained until all data is pruned, after which DA will prune the appropriate stubs from the mailbox, or stubs may be assigned to a fixed retention period within the mailbox independent of the retention policy of the migrated data (i.e. Stubs are pruned at 180 days post migration, migrated data is pruned at 365 days). • PST migration is now supported to include existing PST file message content and re-associate it back to the owner's mailbox. • Single instance storage is preserved directly from Exchange on all sub-client data. No additional servers or specialized infrastructure (SQL Db) is required for this capability. • Migrated data is assigned to Storage Policies which manage retention policies and media management. See Efficient Storage Mgmt Features below. • Data compression and data encryption are fully configurable at different stages of the data migration process. • A predictive reporting tool is provided to run predictive analysis of rule combinations against the source environment. This provides rich output that can be used to evaluate and configure optimal migration strategies for the individual environment. Refer to Exchange Predictor Tool. • Exchange public folders are supported.
Lotus Domino DA Agent	The agent is deployed on a server basis and sub-client policies are defined within the CommCell Console to targets groups of mailboxes and apply candidate selection rules. All sub-clients policies are maintained in the CommServe and processed by the iDA during migration jobs. No scripting is required. No rules are embedded in Lotus Domino mail directories. • Source content: Individual mailboxes are assigned to sub-client policies

	- Auto-discovery is used to quickly assign accounts to sub-clients and it may be configured to automatically monitor Mail Directories and associate new accounts to specific sub-client. • Migration scan rules are defined to trigger which mailboxes qualify for migration • Migration candidate rules are defined with aging and/or sizing rules. These are applied to all mailbox objects. Special rule combinations are available (i.e. read, unread) • Filters are provided to target inclusion or exclusion rules on attachment types or mailboxes users (for instance this can exclude Calendar and address book items from migration scans). • Stub rules define the lifecycle of stubs within the mailbox - they may be retained until all data is pruned, after which DA will prune the appropriate stubs from the mailbox, or stubs may be assigned to a fixed retention period within the mailbox independent of the retention policy of the migrated data (i.e. Stubs are pruned at 180 days post migration, migrated data is pruned at 365 days). • Single instance storage is preserved directly from Lotus Domino on all sub-client data. No additional servers or specialized infrastructure (SQL Db) is required for this capability. • Migrated data is assigned to Storage Policies which manage retention policies and media management. See Efficient Storage Management Features below. • Data compression and data encryption are fully configurable at different stages of the data migration process.
File System DA Agent (AIX, Linux, Netware, Windows, Solaris, NAS / CIFS Shares)	The agent is deployed on a server basis and sub-client policies are defined within the CommCell Console to target directory locations and apply candidate selection rules. All sub-client policies are maintained in the CommServe and processed by the iDA during migration jobs. No scripting is required. No rules are embedded in the volume folders or directories. • Source content: Directory locations are assigned to sub-client(SC) policies – browse and select is used to easily assign paths to SC properties. The SC will monitor the contents of the tree from that location downwards – unless another SC has ownership of a lower level area. Content ownership is exclusive to a sub-client. • Migration scan rules are defined to trigger migration on the volume based on a start/stop % free space watermarks. If a volume has % free below the Start threshold – sub-client migration scanning will proceed. That job will run until all the candidates are selected or until the Stop threshold point is reached on freed space. • Migration candidate rules are defined with aging and/or sizing rules. These are applied to all NTFS file objects (WFS). • Filters are provided to target inclusion or exclusion rules on file types or folder pattern matches. • Stubs are retained until all data is pruned, after which DA will prune the appropriate stubs from the volume after the archived data has been pruned from the Storage Policy. This automatically cleans out the expired stubs on the new DA run. • Archived data is assigned to Storage Policies which manage retention policies and media management. See Efficient Storage Mgmt Features below. • Data compression and data encryption are fully configurable at different stages of the data archiving process.

	• A predictive reporting tool is provided to run predictive analysis of rule combinations against the source environment. This provides rich output that can be used to evaluate and configure optimal migration strategies for the individual environment. Refer to Windows FS Predictor Tool.

B.) User Experience	
Transparent and direct data recall by end-user or production application	Minimizes user training or application customization required to execute direct recall of archived data. • User controlled direct recall triggered on the 'Read' command of a stubbed object. No admin interaction required. Maintains existing security/access model (mailbox or ACLs). Single action recall is permitted for end-user.
Search and retrieval	Extend end-user access to data after stubs expire by enabling rich search and retrieval options. • Outlook users are offered new capabilities to engage searches of migrated message/attachment data directly from their Outlook client.
Stubs preserve attributes of original files	Minimizes impact on production system – no renaming, full attributes retained where possible – enables users to easily identify key attributes of original messages or files directly from stub object. • WFS: Re-parse points/sparse files preserve key attributes of original file (size, dates, and name), annotated with 'Offline' attribute and Windows supported offline icons; Exchange: Custom forms preserve header properties; can be recognized with Simpana icon (customizable); SFS: Retains full attributes.
End-user or external forced migration flags	Enable users or external applications to select items for archiving – in addition to archive policies. • Outlook users can select items for 'forced' archiving – independent of the aging or filtering rules based on the sub-client. The next migration run on the mailbox will process these flagged items. • WFS: The OnDemand feature allows external collect-lists of path/filenames to be submitted for a DA job. DA will migrate these items – independent of the candidate rules of the sub-client
Stub portability, within and across volume	Ensures maximum data portability and access to the user. Allows user to easily reorganize stubbed objects within or across folders or directories without triggering unwanted recalls. • Stubs are self-contained and can be easily be moved by the user within the source volume by drag and drop or using the "Move To" command. Stub recall is linked to "Read" actions triggered by opens, edits or copy commands. For Exchange Stubs, support full mailbox move operations without triggering recalls or requiring additional synchronization. • A unique capability exists in employing the FS iDA within the std DA stack to copy out full directories with migrated data and move them across volumes with backup/recovery process. Use case: move a home directory - backup the folder and recover it to the new volume; this will not trigger stub recalls. As long as that volume is also monitored by DA it will work fine. A command line utility is also provided for copying stubs across volumes and monitoring new locations.

C.) Archive Policies/Operations	
Streamlined archive process	The archive process is applied directly to existing content sources/structures and produces immediate capacity improvements in those targets. ● Sub-client policies contain rules to sweep content sources (folders/directories or mailboxes) and look for for candidate files based age, size or pattern rules. ● We implement a single stage archive process, which is cleaner and easier to manage and requires less overall storage since the file or message data is moved and stubbed within the same process. Alternative approaches, using a delayed two stage process, consume more active storage (concurrent copies existing in primary, secondary, backupsets) and pose more challenges in data recovery scenarios where copies can easily get out of sync. ● Our single-pass process provides immediate results in reclaiming space (vs. two stage model). In the case of Exchange – white-space is created within the Information Store database and can be physically reclaimed with the ESE defrag utility (reducing the overall size of the *.edb database). In the case of Lotus Domino, white-space is created within the mail directory and can be physically reclaimed with the Compact command (reducing the overall size of the *.nsf files).
Centralized policy management	Policies are easily administrated and centrally managed. Migration policies are simple and quick to configure and provide a broad or very granular application. All, groups or a single folder or mailbox may be targeted with unique rules. ● DA policies are managed within the CommCell GUI for all servers and agents within the CommCell – providing centralized management with remote administration across all servers. This maximizes the reach of the administrators while reducing the amount of effort and training required managing consolidated data management operations across backup, migration and archiving operations. All the storage policies and operations are centrally managed from the same interface – this greatly reduces the administrative burden required to configure and manage the system over time; translating to a lower TCO. ● The policies are protected within the CTE backupset – ensuring full data protection and reducing the burden of embedding external rules within the source environments. No scripting is required and granularity can range from a server down to a specific mailbox (sub-client content sources).
Auto discovery of content sources	Minimizes the administration tasks required to target/update archive policies with new users or changes in file system structure. ● Auto-discovery of mailboxes permits administrator to quickly assign mailboxes to migration sub-clients. Auto-discovery can also be assigned on a Storage Group/Mailbox Store basis or name pattern matches to automatically add new accounts to a primary sub-client (i.e. Default). ● File system sub-client provides integrated browse and select capability to select content sources for a sub-client. No need to manually enter paths or accounts – simply browse to the target and add it to the source list.

Scheduled archive jobs	Archive jobs are schedulable with full job history reporting and alerts. Jobs may also be run in on-demand and scripted mode. • DA jobs are scheduled in standard fashion with all Simpana data mgmt operations. A job is created for each scheduled sub-client. Follows standard prioritization, job scheduling and CTE features of the Singular Information Management. Jobs may also be run in immediate mode or a script may be produced to trigger the job per a pre- or post-command or external job manager.

Highly scalable index/catalog for search and retrieval	Indexing/catalogue model ensures long-term accessibility and search without constraining the scalability of the system. • We do not implement a traditional physical central catalog; rather our advanced distributed indexing architecture provides data protection (indexes follow data) and powerful browsing and search abilities from CommCell Console across all Storage Policy copies. We provide the benefit of the central catalog without the risk of the single point-of-failure catalog that is common in archiving products. • Authorized users can be enabled in the CommCell security model to browse (hierarchical navigation view/select of folder trees (FS) or mailbox folders or to search migrated data for retrieval purposes. Messaging data is recovered back to the originating mailbox; cross-server recovery is enabled to support mailbox movement. File system data can be recovered to new folder locations on the original or alternative servers. • Robust search engine provides search-ability by dates, directories, filenames, folder, mail header properties etc. The advanced content indexing options have extended to migration data to include full text search. • Archive Migration jobs run with the ”no stub” rule will execute archiving roles – where the migrated data is completely removed from the content sources and access is enabled through the CommCell browse and search capabilities. • CTE distributed indexing/archive model imposes no limitations on index size restrictions and does not force the administrator to partition the backend as more data is archived.
Minimize file or folder management requirements	Archive policies are flexibly applied to existing file structures or folders – compared to forcing the organization to physically move data to a media folder staging area for migration purposes. • We support a very flexible model that is not intrusive and can apply rules directly to existing file structures or mailboxes. This does not require users to stage files to areas for archiving, although that model can be implemented if required. The use of filters on the sub-client permit inclusion and exclusion rules to be applied to content • WFS: Repeated migrations of changed files with same name supported by versioning in storage policies. User can recover past active versions from CommCell Browse • Exchange: No folder modification or hidden rules need to be used. User fully enabled with forced migration to select specific content for migration. Filters used to exclude folders.

D.) Efficient Storage Management

Extend data movement and access across multi-tiered or classed media (ILM)	Provides multi-tiered data movement for all migrated data. • DA manages migrated data across the first tier by writing data into a Storage Policy primary copy in the CTE. • AuxCopy is a very powerful mechanism to create synchronous copies with individual retention on different media-enabling multi-tiered storage of the migrated data. This lets an administrator implement an ILM, multi-tiered HSM strategy very easily.
Provide data portability across media types; move data to new technology (i.e. tape refresh or upgrade DLT > LTO)	Offsets technology obsolescence risks in storage architecture by allowing archived data to be easily moved across media formats without requiring full de-archiving of the data. • The CTE allows active archived data to be easily moved across media formats and retention times with Cascading Copies using the AuxCopy process. No need to fully de-archive data to change backend media. Imposes no impact on production volumes.
Enable data health check across full life cycle	Ensures data availability and accessibility for the entire managed lifecycle. • The Data Verification Option provides an added layer of protection in scheduling regular tests of that data to ensure it is fully accessible for recall. This complements the standard CRC checks on storage devices.
Data-centric management with retention/pruning; preserve data for retention period	Data is managed with independent retention policy per media tier. Upon browsing data – user is presented with a complete view of the migrated data set based on the time range, abstracting a composite data view independent of actual media copies. • Fully leverages CTE data management model to apply individual retention policy to all storage policy copies. • Provides full lifecycle management ability applicable over tiered media; pruning or aging process completes the final stage of lifecycle.
Fully integrated reporting and alerts management	Enables rich automated reporting of archive, storage and recovery operations. Job results, logs and user alerts are fully supported for DA and CTE components. • Reports can be scheduled, emailed on regular basis and posted as web pages. Data export to excel or publication to PDF formats fully supported. • Enhanced capabilities are available with the use of the CommNet and Storage Manager management system. This provides advanced reporting and analysis including capabilities such as primary/secondary costing, usage and predictive analysis of archived data sets.

Provide full data protection for the entire solution	All components of the CommVault Archive solution can be fully protected without the need for additional agents or 3rd party products. All DM requires is addition media to host the backup copies. • Data and indexes are naturally protected with AuxCopy process which creates active redundant copies for recovery use. No need to promote copy to access and recover data. AuxCopy data sets are asynchronous copies readily available for direct access for stub-recalls or GUI browsing & recovery. • No restrictions on media copies. We provide ability to mix – primary migration data may be on NAS, AuxCopy can create synchronous copy on Tape. This applies to EMC Centera – our process is unique in the case of backing up migration data from a Centera to an alternate media form (tape). • Stubs are backed up normally by existing data protection processes. Stubs are recovered from backup sets after migration of original file. Stand-alone DA has been tested successfully with third-party backup/recovery products. • CTE includes inherent backup and recovery for CommServe (ER backup sets). Protects all rules, policies, indexes and configurations.

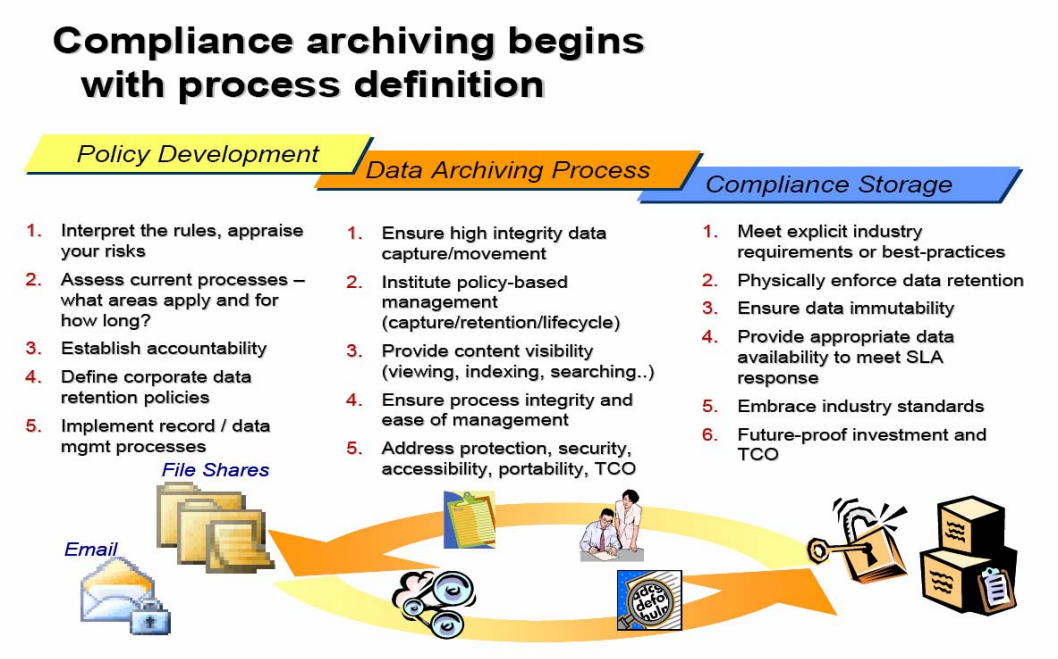
E.) Multiple Platform Support	
	• CommVault® Archive solutions enable enterprise-wide storage management and e-discovery with solutions for email, collaboration, file system and Network Attached Storage (NAS) data. • With built-in tiered storage and multi-platform support including: Microsoft Exchange, Lotus Notes, and Microsoft SharePoint data, comprehensive archive management is simplified.

Data Archiver Compliance Option

CommVault Data Archiver Compliance Option extends the power of Singular Information Management™ to address the specialized archiving and preservation needs of organizations that face regulatory retention or legal discovery requirements. DACO is built on the proven data movement, management and preservation capabilities of the Common Technology Engine (CTE). This modular product offers features directly related to regulated data retention and audit requirements. When combined with the platform features of the CTE and WORM capable media, it provides organizations with the necessary components to implement compliance data retention and preservation processes that support a wide array of industry requirements for retention, audit and discovery of Electronically Stored Information (ESI).

Strategic Value:

Data Archiver Compliance Option has been designed to support organizations facing increasing demands for long-term email retention and accessibility beyond the capabilities of traditional data protection strategies. For organizations who have begun to assess and align new data retention requirements against the backdrop of industry regulation and corporate governance policies, the product helps IT implement high-integrity data archiving processes. DACO can satisfy retention requirements and reduce the over cost and complexities of managing the entire lifecycle of that data.



Challenges Solved by Data Archiver Compliance Option

1. Establish a long-term email archive strategy. Using the Exchange or Lotus Domino Journaling service as the primary data capture mechanism, the service is configured to deposit a duplicate copy of all the mail messages sent or received defined by the Storage Policy. This operation happens independently of user interaction to ensure that a complete email message collection is created and deposited in the proper mailbox for archiving purposes. DACO policies are used to monitor the collection mailboxes for content to be archived.

2. Backfill older email messages into the archive. The default mode (using Journaling as data source) creates the collection beginning at a point-in-time (today) and continues forward. Older email messages existing in either the Information Store (Exchange) or individual mail files (Lotus Notes) may be backfilled into the archive by either importing older PST files (Ex) or .nsf files (Lotus Notes) or configuring DACO to

archive specific production mailboxes without deleting the actual mail items. The process of archiving the existing mailboxes is point-in-time based on the job and will not resurrect deleted items.

3. Create a corporate email archive where messages cannot be deleted or archive to comply with regulations. DACO archives, indexes and then deletes each message from the journal mailbox. DACO has no material impact on individual user mailboxes.

4. Archive email content for long periods of time. The longer the retention requirements, the deeper the archive store lifecycle management requirements become relating to issues such as technology obsolescence (media refresh), onsite/offsite management, data heath checks and multi-tiered storage management.

5. Verify data, generate reports and manage media. DACO extends many of the CTE management and reporting features for compliance-based archiving.

6. Audit content search and retrieval. Audit ability is applied not simply to the content, but also to the authorized users of the review and retrieval process.

7. Search email body text and attachment content. DACO is enabled to use the CommVault Content Indexing Option as part of the archiving process; the CI Option is used to create and preserve a full-text index of the content that is readily available for the full data lifecycle of the archive storage policy.

8. Enable authorized business users to quickly and transparently access, content-search and retrieve archive email in a self-sufficient mode.

9. Meet organizational demands for historical email retention. A single email message restoration process from data protection backups can consume a range of ten to well over one hundred man-hours per event. By comparison, once the email has been archived with DACO, legal or business users can quickly retrieve required message sets within minutes without engaging any IT resources.

10.Use information as corporate asset and competitive advantage when responding to regulatory or legal challenges.

Data Archiver Compliance Option Feature Table

• Fully integrated with Microsoft® Exchange and Lotus Domino mailboxes that are targets of the Journaling service. DACO can also target Exchange mailboxes that are SMTP targets of third party Instant Messaging archiving solutions[6]. Envelope journaling is fully supported.
• Acts as an iDataAgent, under the policy configuration and job scheduling control of the consolidated CommCell GUI command and control management interface.
• The archiving process scans the Journal, applies single instancing to both messages and attachments, creates a property-data and full-text index and extracts all available messages for archival storage within a Storage Policy.
• On successful completion of the archiving process, the candidate messages are deleted from the Journal mailbox(s) to complete the data capture process.
• The original email records, attachments and associated indexes are stored within protected archived files on the secondary media (in the SEC case, WORM media). This affords stronger safeguards to ensure that archived data is never altered or modified aside from physical destruction of the media.
• Job histories and logs are maintained. They detail the archiving and retrieval jobs for audit purposes with date/time stamps.
• The data collected within an archive job is actively managed by the assigned QiNetix Storage Policy that associates physical media and retention rules.
• The archived data and indexes are managed in a controlled, read-only state for the length of time set by the retention rules. During this period the data is readily available for access and retrieval by authorized and audited users. By storing the index data along with the email data, our distributed indexing architecture enables organizations to easily layer in data protection with redundant offsite copies and index preservation in accordance with regulations.
• The Auxiliary Copy process provides an easy way to quickly create redundant data copies (across different media forms) to ensure data protection and disaster recovery within your archival system. The CTE affords customers with built-in data protection features that eliminate the need for third party data protection or HSM bolt-on extensions.
• After the required retention time has transpired, the aged data is deemed ready for expiration and applicable to a scheduled pruning-maintenance job. This job purges the data and index entries from the active retrieval system. In the case of WORM media once the last data-set on the media has been logically “pruned” - the platter is marked as expired media and is then available for physical destruction. In the case of recycle-able media, a pruning process will make that media available to the storage pool for reuse.

[6] Third-party Instant Messaging archive tools can be configured to send transcripts of IM conversations in the form of email messages to SMTP addresses. These messages can be captured and archived from that mailbox by DataArchiver Compliance Option.

• Active archive data managed within a Storage Policy is readily available for review, search and retrieval by authorized users. We provide two interactive models to browse and select email data for retrieval from the maintained indexes.
• Navigation tree/folder/sub-folder representation.
• Boolean search set representation (full-text search enabled).
• A sampling mechanism within the search application enables the user to reduce the size selection set from a larger population of search results, in order to optimize the scope of records submitted for compliance review tasks.
• All the email data, attachments and indexes are always protected from any edit or alteration from all users within the context of the archival system. Users are restricted to retrieving duplicate copies of the system of record original instance that remains under the management of the Storage Policy. The duplicate data is retrieved to review folders in pre-assigned Exchange mailbox(s). For Lotus Notes environments, the duplication data is retrieved out-of-place (i.e., to a secure mailbox created exclusively for compliance purposes). The Retrieval function creates a new review folder, annotates and time-stamps it with the Retrieval job reference and proceeds to deposit all the selected contents of the retrieval set within that folder.
• The CommVault SIM architecture employs multiple layers of safe-guards and specialized auditing features across the entire data capture, movement, management and preservation processes to ensure the integrity and validity of the archive data under management. This is achieved with a solution that delivers the highest levels of manageability and usability balanced by the lowest TCO for long-term data preservation.

A Powerful, Simple and Safe Approach

Data Achiver Compliance Option provides a powerful, yet simple approach that allows organizations to quickly address the mounting requirements for long-term email retention and compliance management. By aligning archiving policies to implement corporate record management strategies for email data, DACO reduces the risk of inadvertent or deliberate data losses that may lead to significant fines or legal action. Enabling the organization with extensive content visibility across the email archive provides easier ways to proactively engage surveillance or discovery activities to further mitigate business risk.

Although DACO can be deployed either a standalone or integrated role with other offerings in the Simpana™ product suite, it brings forth the power of SIM architecture to ensure designed in data protection, ease-in-administration, enterprise scalability and seamless integration with special tamper-proof compliance media formats. Together these powerful features drive down the fill lifecycle costs in managing and provisioning a long-term data archive.

Finally, for those customers who have already made investments in the CommVault product suite, DACO affords fast deployment and leverages many of the skills and resources already existing in the organization. As evidenced by the shifts in the broader marketplace, point-solutions drive up costs and complexities as they grow, increasing the burden placed on the organization. DACO provides leveraged investment, infrastructure and training advantages with the context of existing operations.

Conclusion:

Used separately or in combination, CommVault Archive Modules help control data growth, reduce costs, and manage risks associated with non-compliance. CommVault Archive Software serves mailbox end-users, administrators, legal discovery and compliance teams and maintains high-speed performance in the face of increasing data volumes.

For Additional Information

For additional information relating to the Data Archiver product and Common Technology Engine please consult documentation available on the CommVault website at http://www.commvault.com.

CommVault
Corporate Headquarters
2 Crescent Place
Oceanport, NJ 07757
Phone: (888) 746-3849
Phone: (732) 870-4000
Fax: (732) 870-4525
www.commvault.com