



Intelligent Data Management Required When Implementing a Modern Storage Infrastructure

Intelligent Data Management Required When Implementing a Modern Storage Infrastructure White Paper

Applying intelligence and business understanding to corporate data is the key to deriving immense financial, strategic, and ROI benefits from a company's data assets. It is not difficult to apply this intelligence to data but it requires tools that manage the data from a more disciplined perspective utilizing metadata. Rocket® Arkivio software has been helping the largest companies in the world to better understand and manage their data for over a decade. After thousands of data discovery and analysis projects, the Rocket Arkivio team has simplified the process of intelligent data management to "Gather, Group, Act"—and provides the software necessary to quickly implement a solution. Rocket Arkivio data management software enables IT departments to intelligently deliver ROI results to the business by optimizing backup and recovery, reclaiming primary storage, and providing regulatory compliance and retention management. This white paper describes the benefits, functionality, and technical architecture of the Rocket Arkivio Autostor intelligent data management solution.

Revised September 2014



Table of Contents

Overview.....	3
Introducing Rocket Arkivio Autostor	5
Rocket Arkivio Autostor in Detail.....	7
Product Features	7
Highlights.....	7
Benefits	7
Gather.....	8
Data Collection.....	8
Group	8
Resource Grouping and Classification.....	8
Volume Groups.....	8
File Groups	10
File Location Groups.....	11
Reporting	12
Resource Usage Reports.....	12
Data Usage Reports	12
Policy Monitoring Reports	12
Custom Reports	12
Act.....	13
Policy-based Data Management	14
Scheduling.....	15
Simulation.....	15
Policy Actions	15
Copy Policy.....	16
Migrate Policy.....	16
Move Policy	17
Delete Policy	17
Retention Policy Actions.....	17
Restore.....	17
Delete.....	18
Storage Device API Integration	18
EMC Centera API Integration.....	18
EMC VNX/Celerra API Integration	18
Hitachi HCP API Integration.....	19
Network Appliance API Integration	19



Quantum Lattus API Integration	20
Tape/MO Support	20
Rocket Arkivio Autostor Architecture.....	20
Overview.....	20
Scalability	22
Agentless Discovery and Data Movement.....	22
Hardware and File System Independent	22
Content-based Global Policies	22
Modular Building Blocks	22
Capacity-based Alerts	23
Out-of-band.....	23
Anti-virus and Backup Software Compatibility.....	23
Disaster Recovery.....	23
Disaster Recovery for ACS.....	24
Disaster Recovery for RSA.....	24
Self Description	24
Metadata Accessibility.....	25
Tag/Link Reconstruction	25
Conclusion	25



Intelligent Data Management Required When Implementing a Modern Storage Infrastructure



Overview

What value does data have for your business? Modern processing power and software technology such as Hadoop allow companies of all sizes to find revenue opportunities by accessing and analyzing company generated and managed data, as well as third party data. However, in order to utilize data as an asset and use it to generate revenue, a new business approach is required that moves beyond raw processing power. Even though 94% of companies surveyed recently said that implementing a Big Data system was a top priority over the next few years, most Big Data implementations have failed in spite of millions of dollars of investment and the allocation of significant resources. The reason is simple – "boiling the ocean" of data is a near impossible task except for the largest, best funded organizations. Data discovery and categorization platforms are a key missing component of an implementation to optimize value creation from a company's data assets.

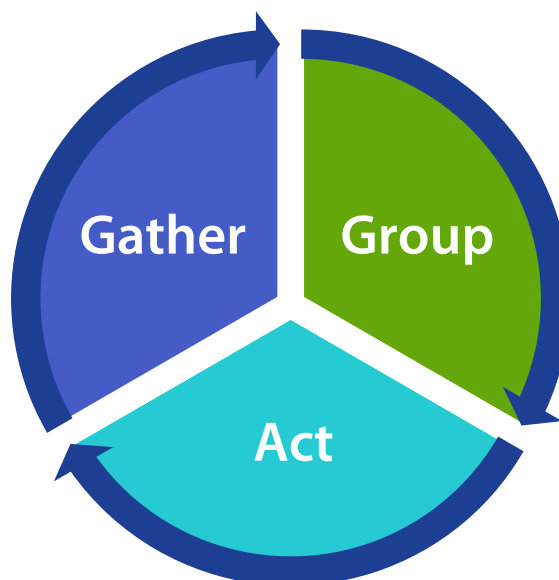


Cloud and virtualization implementations have the same challenge. Do we move all our data to the cloud? Do we virtualize all our data? Is implementing this technology across the entire company a good use of resources? A data discovery and categorization system is the ideal platform to segregate which data should be included in the implementation to generate the best value for the organization.

Most companies treat their data as a cost center. Existing data is ignored as "stale" in favor of projects that purchase more capacity or process data faster. They see it as a nuisance that is constantly nagging for more attention, increased capacity, and to be backed up and archived. However, mining stale data can not only free up primary storage but it can also decrease backup time. This data also contains the life blood of the organization and is a source of revenue; intellectual property, contacts, accounts, projects, transactions, financials, and personnel files, to name a few. Yet data management is seldom or poorly practiced at most companies because of the seeming enormity of the task to manage data growing on average 40% a year and in most cases, much faster than the IT budget.

Data management and the use of classification tools can change that outcome. The difference can be illustrated simply. If you want to find all the chemical engineers in the world—does it make sense to create an index and profile of every one of the 7 billion people on earth first, and then analyze that information to find the chemical engineers? Of course not, it would be much easier to classify people into broad groups and only index and research people that have graduated from an engineering academic institution. This insight can be extended to a practical business example. When looking for intellectual property at a company, do you first index all the data at a company then search for key words that might expose which files contain intellectual property or do you first group those that work in technology positions or departments and then index their data?

Applying intelligence and business understanding to corporate data is the key to deriving immense financial, strategic, and ROI benefits from a company's data assets. It is not difficult to apply this intelligence to data but it requires tools that manage the data from a more disciplined perspective utilizing metadata. The Rocket Arkivio® team has been helping the largest companies in the world to better understand and manage their data for over a decade. After thousands of data discovery and analysis projects, the Rocket Arkivio team has simplified the process of intelligent data management to "Gather, Group, Act"—and provides the software necessary to quickly implement a solution.





IT management decisions become straightforward after implementing a Rocket Arkivio system. The software allows users to readily implement a tiered and multi-vendor storage architecture, apply data retention policies, implement requirements from the compliance officer, improve the backup and disaster recovery processes, know how much data is being consumed by which department, application, and users, and project future capacity and infrastructure requirements.

Rocket Arkivio software empowers an IT department to traverse the changes in data types and volumes as well as storage technology to derive as much value as possible. Continuing advances in the Rocket Arkivio technology over the past ten plus years brings a level of experience unmatched in the industry to help you deliver the ROI needed to justify any expenditure while enabling new virtualization, Big Data, and Cloud implementations. This intelligence applied to data management has the added benefit of reducing storage capacity, backup, and archiving costs, as well as helping to ensure a company meets regulatory compliance rules.

Introducing Rocket® Arkivio Autostor

With file system data growing at an exponential rate (a McKinsey study projected global data to grow 40% annually while global IT spending will only grow at 5%), where does an IT department start in order to get a handle on the data they already have, as well as what is being created every day? Vendors tout offerings such as Software Defined Storage Infrastructure, SSD Storage, Hybrid Storage, Virtualized Storage, Object-based storage, VDI storage infrastructure, clustered NAS, Scale-out NAS, Cloud Gateway, Hybrid Cloud, or Private Cloud to solve this data storage problem. But getting control of the growing unstructured data in a company is a data management problem rather than a storage infrastructure problem. These "boil the ocean" solutions ignore one simple fact; no company can afford to transition 100% of data to a new infrastructure. A company needs to first understand what data will be included in the new implementation and what data stays in the existing infrastructure.

Companies can also derive tremendous ROI intelligently managing existing data in the IT infrastructure. Storage audits at thousands of large companies show that the typical organization has not accessed 60-80% of its unstructured data residing on primary storage in over a year. Rocket® Arkivio data management software enables IT departments to intelligently manage this situation by optimizing backup and recovery, reclaiming primary storage, and providing regulatory compliance and retention management.

Clearly, just adding more storage locally or in the cloud is not a sufficient solution to this challenge. Rocket Arkivio Autostor simplifies intelligent data management through its three step process—Gather, Group, Act. At the heart of Rocket Arkivio Autostor is an agentless meta-data collection system that automatically



gathers the most critical information about the data. The collected metadata is presented through standard and custom reports that identify obvious opportunities for ROI and quickly enables grouping or classifying data into logical groups based on their importance to the business. From these reports, the IT department and business can take decisive, consistent action on the data in line with its objectives.

Rocket Arkivio Autostor automates and integrates these three key components into one seamless, software system: metadata collection (Gather), data classification (Group), and data management (Act). Competitive products typically require the purchase of three or four products that must be manually integrated and deployed to equal what Rocket Arkivio Autostor can do by itself. In addition, Rocket Arkivio Autostor provides the only solution that has a classification system that enables customers to easily act on hundreds of millions to billions of files or objects through the use of intelligent grouping. These logical groups simplify automating global policies and actions on the data. Rocket Arkivio Autostor is easy to drop into an existing infrastructure since it requires no server agents and the solution stays out of the data path by design.

The remainder of this paper describes in detail the technology behind Rocket Arkivio Autostor proven implementation process—Gather, Group, Act.

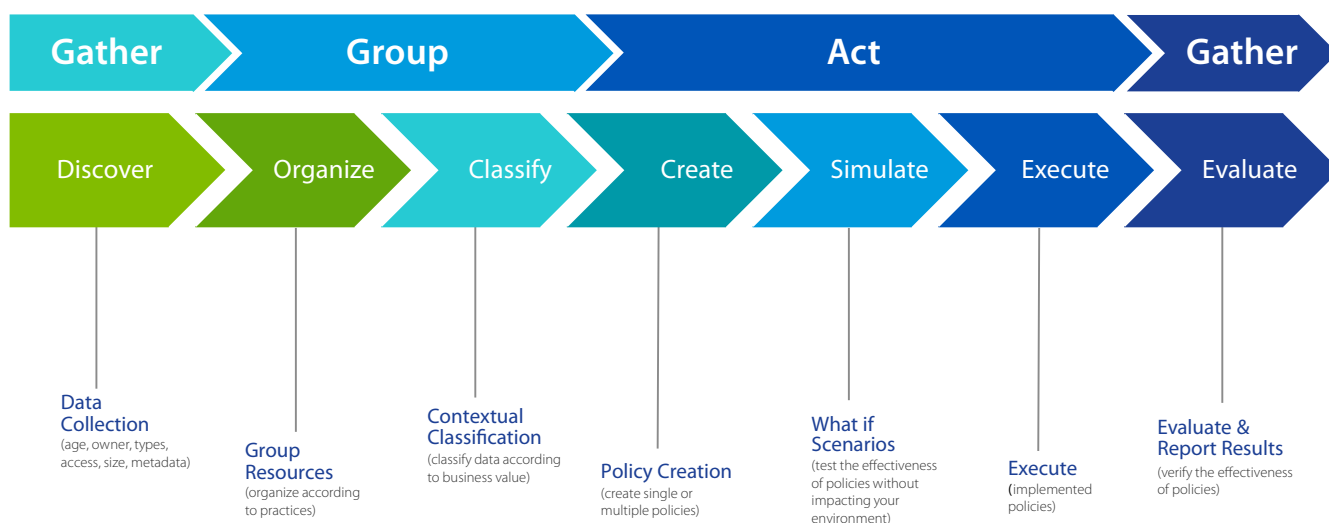


Figure 1: Intelligent Data Management Implementation Process



Rocket Arkivio Autostor in Detail

Product Features

Rocket Arkivio Autostor is built around a process model (see Figure 1) consisting of the following components: Gather: data collection; Group: organization, classification; Act: policy creation, policy simulation, policy execution, and self-evaluation of automated policy.

Highlights

- ❖ Integrated System - Combines discovery, classification, reporting, automated data movement and continuous monitoring across multiple operating environments (NAS, Windows and UNIX/Linux)
- ❖ Contextual Classification - Provides flexible grouping and prioritization of data based on intelligence collected from files, user groups, directory structure, and the business relevance assigned to each group
- ❖ Policy-Based Management - Automates data management actions via a multi-threaded, event-driven policy engine
- ❖ Transparent Data Access - Automatically and non-disruptively redirects end-user and application requests for access to migrated data to the repository
- ❖ Enterprise Scalability - Scans and manages over six billion files spread across hundreds of servers, adapting to variations in network, OS, protocols, and domain architectures
- ❖ Multi-threaded Data Collectors - High-performance, agentless data collectors run independently with extraordinary speed and scale in complex heterogeneous enterprise environments
- ❖ Extensible Reporting - Aggregates and analyzes data across entire system to provide customizable views of point-in-time and trending information
- ❖ Global Deployment - Unicode-compliant software ready for deployment in international locations

Benefits

- ❖ Discover and collect statistics on storage utilization and data usage patterns across heterogeneous systems without deploying server agents
- ❖ Classify data and storage resources into logical groups based on business criteria for efficient data retention management
- ❖ Automate data management actions such as copy, migrate, move, retain, and delete based on changing data retention and availability requirements
- ❖ Deploy policies to automate migration of regulatory data from primary storage to WORM devices for compliance
- ❖ Simplify tiered-storage management by automating transparent migration of data from primary storage to secondary or tertiary storage (e.g., ATA, NAS, Tape, Optical)



- ❖ Continuously monitor and alert on storage utilization levels, data usage patterns, and policy effectiveness to proactively resolve capacity management issues
- ❖ Allow end user and management access to the reporting system to proactively understand data and data utilization business opportunities

Gather

Data collection

Rocket Arkivio Autostor uses an agentless approach for data collection which eliminates the need to install and maintain server agents. The Rocket Arkivio Autostor software is hosted on a physical or virtual server running Windows 2008 R2 Server (64-bit, Standard or Enterprise). The administrator can then configure the software to scan selected volumes on file servers and NAS devices in the network utilizing industry standard protocols (CIFS, NFS). The data collection process can be scheduled or manually launched by the administrator on a select group of servers.

Rocket Arkivio Autostor collects information on total available and used capacity by volume across the defined servers. The system also collects information on files such as size (physical and logical), file type, owner, creation date, last modified and last accessed dates, directory path, and owner information (if a directory service such as Active Directory or NIS is used). While the most common usage is the management of Volumes, Rocket Arkivio software also supports managing the logical storage of a DFS namespace.

Group

Resource Grouping and Classification

Rocket Arkivio Autostor provides an intuitive mechanism for organizing data and storage resources based on an organization's specific management needs whether it is the IT organization itself or the user, owners, compliance officers or HR and Legal teams. There are three major types of logical resource groups: Volume Groups, File Groups, and File Location Groups.

Volume Groups

A volume group consists of any combination of unique volumes that the administrator logically organizes for monitoring, reporting, and policy-based management. Volume groups can be defined based on the file protocol type (CIFS or NFS), organization, or location. The Rocket Arkivio Autostor solution supports two different categories of volume groups: Host-Based and List-Based.

Host-based volume groups are created automatically by the Rocket Arkivio Autostor software as servers are added to the systems under management. The host-based volume groups contain all volumes



presented to the network by a host server or NAS host. List-based volume groups are defined based on an explicit selection of volumes by the administrator. Because the administrator can select membership based on individual volumes, these groups can include volumes across heterogeneous storage devices and sub-sets of storage devices. Figure 2 below provides several examples of custom volume groups created with the Rocket Arkivio Autostor software solution.

When a new volume is detected, Rocket Arkivio Autostor automatically alerts the administrator of the new addition and automatically associates that volume with its existing Host-Based volume group. The new volume can also be easily added to additional List-Based volume groups, and once an Admin selects it for management, it will be set to observe all the existing policy and reporting actions of those volume groups.

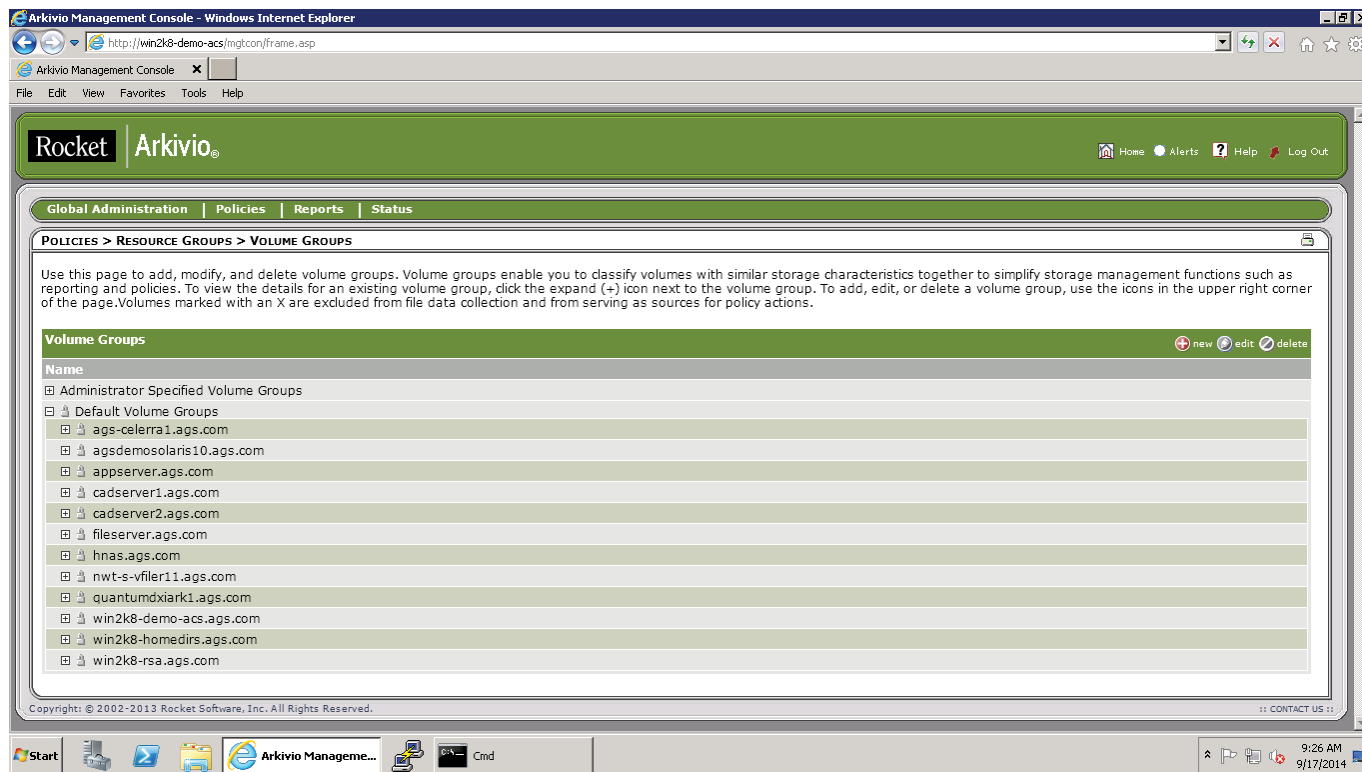


Figure 2: Volume groups

Volume groups ease the storage management burden by enabling administrators to logically manage large distributed storage environments and facilitate group-based reporting in order to identify over-utilized servers by department or by application. Administrators are also able to expedite decision making and justify future storage acquisitions based on utilization trends.



File Groups

A file group is a logical collection of files that are stored on distributed servers and storage volumes across the network. Rocket Arkivio Autostor automatically groups common file types, for example, Microsoft Office documents and image files. In addition, administrators can create custom file groups using one or more of the following attributes: file type (extension), size, owner, age (last modified or last accessed), and directory paths which can span different hardware and different volumes. Figure 3 below provides several examples of custom file groups created with the Rocket Arkivio Autostor software solution.

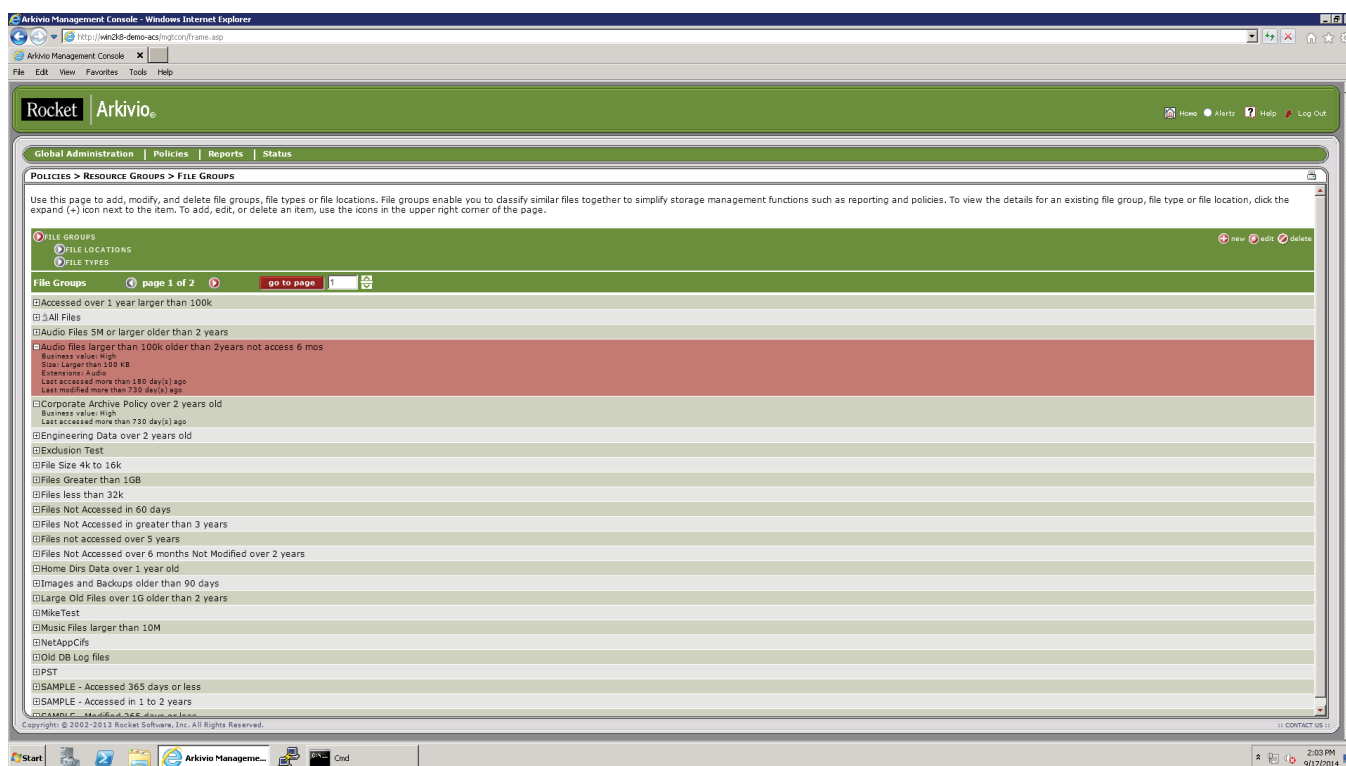


Figure 3: File Groups

File groups enable administrators to identify what data on the network can be managed similarly based on the data's lifecycle or regulatory requirements. For instance, administrators might group all high-value engineering files associated with a new product design or they may group all medical image files that must be retained to comply with HIPAA guidelines.

Custom reports can be generated to monitor how different groups of file data are being utilized or to uncover groups of data that are occupying the most space in volume groups. For example, old or "stale" data that has been identified by the metadata collector can be quickly grouped by creation, last modified,



or last access date of the files. The report can be viewed or emailed to end users for remediation or acted on by creating a policy that deletes or archives those files that have passed corporate retention guidelines or after a capacity threshold is reached. Groups could also be created to view the number and location of TMP or PST files making them available for removal or other action. Since a file group is a logical view of metadata attributes, a file can be a member of multiple groups and no data is moved when creating a group.

A major benefit of file groups is that they give administrators the ability to organize and prioritize data according to the data's business value without having to worry about where on the network or on what type of storage device the data resides. Rocket Arkivio Autostor enables this through a simple and straightforward classification scheme of high, medium, or low for file groups. For instance, the administrator might rank a file group as "high priority" when the data is extremely important to the business, such as financial data summarizing the current fiscal quarter results. In contrast, a "low-priority" file group might include data that is inconsequential to current business operations, such as temporary or log files. When multiple file groups are included in a policy, Rocket Arkivio Autostor will select files based on their priority levels, ensuring proper disposition of data based on its value and the company's business priorities.

File Location Groups

A File Location Group (FLG) is unique to Rocket Arkivio Autostor and has the power of file and volume groups discussed in detail above. The FLG allows administrators to customize the software's reporting and actions based on how the businesses they support are organized or how data is stored across different business units, servers, applications, or storage devices. Specifically, a FLG is a group of selected directory trees from any number of volumes across one or multiple servers. These groups enable a more granular level of policy actions than file or volume groups by acting upon files in just those specific directories.

The directories specified in a FLG also increase the granularity of the standard or custom reports generated by the system. This type of grouping allows you to report on or act on sets of files that span an entire enterprise but with the ability to include only the specific locations that have meaning to the report or policy. For example, in a multiple division organization with HR departments in each organization, directories supporting salary and performance administration located across multiple servers and locations can be placed in one file location group for reporting and compliance retention actions.

The power of this unique grouping structure comes from its flexibility. Any number of FLGs can be defined, with any number of directory trees, across any set of volumes and servers. This gives both the ability to intelligently pinpoint policy actions to specific directory trees that are logically associated with each other and to report on each of these directories and on the group of directory trees as a whole.



Reporting

Rocket Arkivio Autostor gives users the ability to gain detailed insight and intelligence on their existing data using a comprehensive set of standard and customizable reports built from the storage and data information gathering process. These reports enable the organization to uncover detailed information about their data and storage resources, such as resource inventory, resource usage, data usage, capacity planning, and policy monitoring. These reports can be viewed through an onscreen dashboard or exported in a variety of formats. In addition, the system administrators can set up an automated email of selected reports to be delivered directly to end user or management or grant password protected access to the reporting system so managers, end users, and compliance officers can directly access the system and create, run, and/or view needed reports on demand.

Below is a detailed description of the different report categories offered by Rocket Arkivio Autostor:

Resource Usage Reports

These reports display current usage information of the storage resources managed by Rocket Arkivio Autostor. Administrators can view space consumption by top N files and directories, current space consumption of each storage device, and space consumption trends for the last N months or years. Further drill-down to view information pertaining to volume or file groups is also provided for example cumulative size of files distributed by "last access time" for a given system.

Data Usage Reports

The data usage reports display aggregated current space usage information at a file level, based on creation, last modified, and last accessed time. For example, administrators can see the number of files that have not been accessed for a given period on each storage device. Users can also choose to further break down the information based on volume or file groups. Figure 4 shows same sample reports for intelligent data management of storage resources.

Policy Monitoring Reports

Rocket Arkivio Autostor offers a comprehensive set of policy monitoring reports that provide continuous feedback on the success of implemented policies. These reports show the net results of different policy actions based on high-level statistics, such as how many or which files were acted upon by a policy from a source volume group. The reports also summarize how much capacity has been freed at the source volume and how much capacity is utilized at the target volume.



Custom reports:

The Rocket Arkivio Report Builder allows for the defining of custom reports that can be distributed through email to end users, managers, or others on a scheduled basis or viewed in the Rocket Arkivio reporting dashboard. These emails can include embedded tables and graphs and may contain information about their data, the potential for savings, the storage growth, or abusive usage. While Report Builder provides custom reporting it also has a variety of built-in sample reports to enable quick customization or email distribution of important resource utilization information.

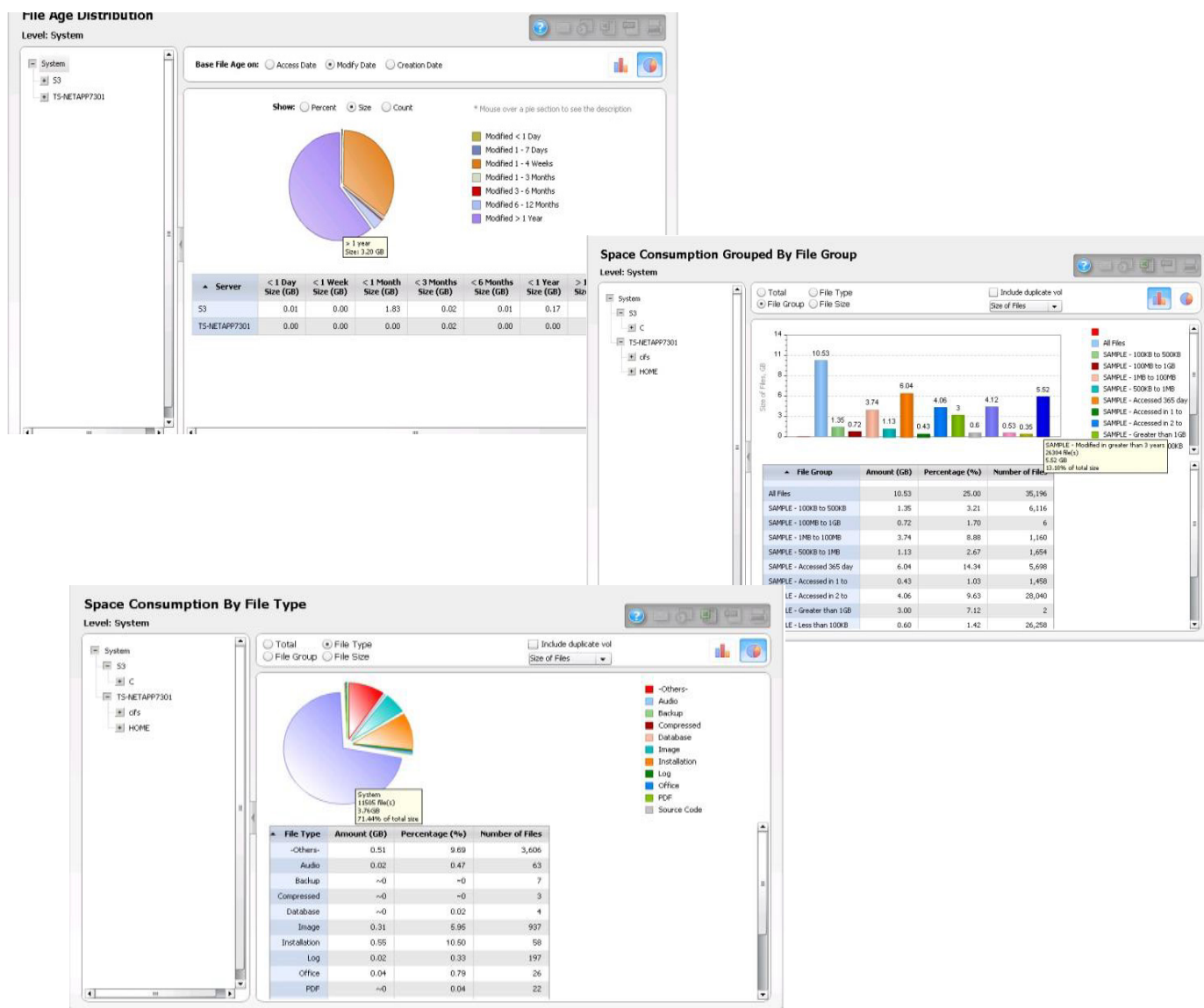


Figure 4: Reports for Intelligent Data Management

The Rocket Arkivio Autostor Policy Automation Engine is responsible for executing the data management actions. It is unlike any other policy-based solution available today because it uses a patented algorithm to select files for movement based on detailed parameters such as age, size, last modified, last accessed date, etc., combined with the business value or priority that was assigned for the different file groups.

Global policies allow administrators to manage very large distributed, multi-vendor storage environments that are deployed across logical resource groups, rather than having to create and manage individual policies across multiple physical servers and volumes. Using file groups, for example, the administrator could configure a policy to copy high-value data from many servers in a primary site to a remote site, the cloud, or object storage for specific files that match the specified criteria, as illustrated in Figure 5 below. When multiple file groups are included in a policy action, the file groups are acted upon according to the priority assigned to each group.





Scheduling

Acting on data through policies is simplified by the system's built in scheduling module. Policies can be scheduled to act on data just once, run automatically on a periodic basis, e.g. daily, weekly, or within a certain time period, or when an event occurs such as a capacity threshold being reached on a primary storage system or volume.

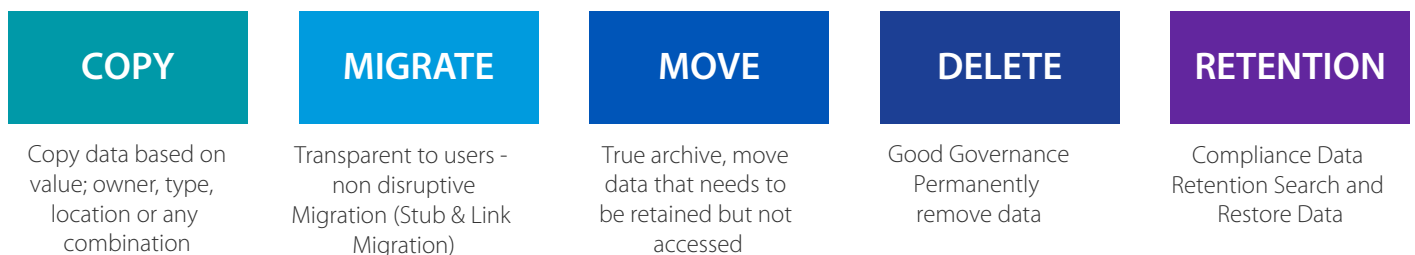
Simulation

Once data management policies have been defined, a powerful feature of Rocket Arkivio Autostor is its ability to simulate the different policies and to test their expected results before actually executing them. Various what-if scenarios can be quickly run to determine the outcome and impact of the planned policies, including the ability to drill down to see the individual files that will be moved. Simulation of data management policies before actual execution provides a better understanding of policy scope and effectiveness. For example, the administrator can identify the capacity to be freed by a data migration policy, as well as see a list of files that would be affected. Additionally, the same detailed set of reports which are available for implemented policies are available for simulated policies.

Policy Actions

Rocket Arkivio Autostor policy automation engine enables managing the lifecycle of a file from "cradle to grave" (Figure 6). While individually managing billions of files is an impossible task, the software's gathering of file information and grouping capability along with its integrated policy action simplifies intelligent data management across the enterprise.

Figure 6: Cradle to Grave Data Management





Copy Policy

The automated Copy action is typically used to provide a compliance backup copy of data to ensure data accessibility. The ability to Copy only valuable or compliance data ensures availability of an immutable version of the data.

Migrate Policy

Migrate actions are implemented to maintain transparent user and application access to data after moving the data to a different storage location. Rocket Arkivio Autostor supports two different types of migration: Tag Migration and Link Migration. When a Tag Migrated file is accessed, the server can be configured to automatically restore data from the migration target to its original location or files can be read directly from the target without recall of the files. Link Migration moves file content to a different storage volume, while at the same time providing direct, transparent read-access to the data from the original location.

Using Tag Migration to move data from a source volume to a target volume leaves behind a stub file that includes all the file property information and takes up only a small amount of space (4KB or 8KB, depending on the storage system block size). The stub file is encoded with the location of the migrated file, allowing access to the migrated file independent of the Rocket Arkivio database. To end users and applications, the tag file appears as if it is the original file. They have access to the data transparently through the tag, even though the content of the file has been moved to a target storage repository that can be located on any server or storage device available to the Rocket Arkivio Autostor software.

When files are migrated using a link or shortcut, the content is moved to a target volume and directory specified by the administrator during the configuration of the policy. The Link Migration action involves permanently migrating files from the source to the target volume, while leaving behind a small¹ symbolic link (NFS) or shortcut (CIFS) at the source location that is encoded with the new location of the data. Accessing the migrated files through a symbolic link or shortcut will not cause the migrated files to be recalled. Instead, files are read directly from the target location. Since the link or shortcut file includes the location of the target file, Rocket Arkivio Autostor is out of the data path unless required for accessing a retention storage device such as HCP or Centera. This makes it possible to access the target file even if Rocket Arkivio Autostor is no longer available or is removed from the environment.

¹Size of shortcuts is 4 KB. Size of symbolic links depends on the target path.



Move Policy

A Move policy is typically used to automate server and storage consolidation, relocate data for Cloud and Big Data applications, or move little used data to an active archive or other long term storage. It may also be used to comply with federal regulations by moving data to storage targets that support retention periods, such as WORM devices. A Move action permanently relocates files to a target directory location and, unlike migration, does not leave behind a tag file or link at the original location. A Move policy can be configured to move all the specified files in the selected file groups, or only enough files to relieve an over-capacity condition on a volume.

When a Move policy encounters a file that has been previously tag-migrated, the policy can optionally move the tag file to the new location, even if the target is of a different storage type from the source storage system (e.g.: NetApp to VNX or vice versa).

Delete Policy

The Delete action allows unwanted data, such as temporary files, log files, or data that has passed its retention limit, to be automatically deleted on a scheduled basis or when the utilization of a volume reaches a certain threshold. The file deletion policy is specified to delete all files matching a file group's criteria or only enough files matching the criteria to relieve an over-capacity condition. The Delete policy can be configured to perform the clean-up automatically - or to do so only after the administrator has verified the list of files that will be removed. When tag or link-migrated files are deleted, the policy can delete the tag or link as well as the migrated file.

Retention Policy Actions

Retention actions assist organizations in achieving regulatory compliance or maintaining good corporate governance by adding retention time periods to the data management policies (Migrate, Move, Copy, or Delete) through tight integration with Object Store or Retention devices. The Retention actions operate the same way as regular data movement actions. They have the added benefit of allowing the administrator to set a specific retention period for each data movement policy. Retention periods are enforced through the ability to use the native Retention device API to set the amount of time the files must be retained.

Restore

For the situation in which files need to be brought back from a compliance device for audit, legal discovery, or recovery purposes, Rocket Arkivio Autostor provides an automated feature that restores the files to their original location on the file system or an alternate location. This comprehensive Restore feature allows administrators to pick the exact files they want to restore using the



intelligence stored in the database. Administrators can view migrated files on a per-server and per-volume basis, within a specified date range. In addition, they also have the freedom to locate files based on modification date, version, migration type, directory name, file name, and file extension. By applying sophisticated filter criteria, administrators can quickly locate the files they want to restore.

Delete

Rocket Arkivio Autostor offers administrators a Delete task under the circumstance in which files previously migrated, moved, or copied to a compliance device need to be removed.

Administrators can specify which files to delete by selecting from a list that includes all eligible files. Eligible files are defined as those that have expired or have no specified retention period. The Delete task has the same filtering capability as the restore task, allowing administrators to locate files based on source server, source volume, modification date, version, migration type, directory name, file name, and file extension.

Storage Device API Integration

EMC Centera API Integration

Rocket Arkivio Autostor is fully integrated with the EMC Centera API to perform all data management actions (migrate, move, copy, or delete), including retention for compliance and data de-duplication. This integration facilitates using the EMC Centera as a target for data actions, as well as setting Retention policies on data and performing all compliance device functions. Further, Rocket Arkivio Autostor supports seamless fail-over when multiple, replicated Centera clusters are deployed. This ensures that large or complex environments maintain maximum availability of the archived data with no additional administrative effort.

EMC VNX/Celerra API Integration

Rocket Arkivio Autostor is fully integrated with EMC's DART CFM/VFM service to effectively manage data residing on EMC file servers. All data management actions such as migrate, move, copy, and delete with and without retention are supported for moving data between file servers, as well as to other storage devices.

Files that are tag-migrated from a VNX/Celerra volume will be replaced by an EMC-specific stub file from which end-users can access migrated files transparently. Files migrated through this mechanism will be accessed and read by the operating system, with no interaction or intervention required by the Rocket Arkivio Autostor software or its driver. The CFM API enables Rocket Arkivio Autostor to create tags that have different data access features including recall of data to the primary storage, direct read of the data



from the repository (pass-thru read), and partial read. Additionally, Rocket Arkivio Autostor encodes the file location in the tag file which enables the VNX/Celerra operating system to access the data without the need to communicate with the Rocket Arkivio server. End users still have access to their tag or link migrated data in disaster situations where the Arkivio Autostor software is unavailable.

When files are tag-migrated from the VNX/Celerra volume, they are stored in a migration target repository on a volume managed by the Rocket Arkivio Autostor software. If a previously migrated file is recalled and then becomes eligible for migration again, Rocket Arkivio Autostor will detect that the size and contents of the file are not changed and simply re-tag the file without performing another migration operation. This ensures that only one copy of the file's content is maintained in the repository at all times.

Hitachi HCP API Integration

Rocket Arkivio Autostor is fully integrated with the Hitachi Content Platform (HCP) API to perform all data management actions (migrate, move, copy, or delete), including retention for compliance. This integration facilitates using the Hitachi HCP as a target for data actions, as well as setting retention policies on data. REST protocol integration facilitates seamless fail-over in multi-HCP scenarios to ensure maximum availability with no additional administrative effort, and support for multiple HCP Namespaces and Tenants is included.

Network Appliance API Integration

Rocket Arkivio Autostor is fully integrated with the NetApp 7- and C-mode FPolicy APIs to effectively manage data residing on the NetApp family of primary and nearline storage. The integration with the NetApp APIs enables Rocket Arkivio Autostor to work directly with the Data ONTAP operating system to migrate data from primary storage while providing transparent access to the archived files for both CIFS and NFS users.

All data management actions such as migration, move, copy, and delete, with and without retention, are supported for moving data between NetApp filers, as well as to other storage devices. In the case of tag migration, the file content is replaced by a NetApp-specific stub file. When a tag file is accessed by the end-user or an application, the Rocket Arkivio Autostor software interacts with the Data ONTAP operating system to either automatically restore the file to its original location on the primary storage or stream just the file contents to the requesting application (Pass through) without causing a recall. The end-user experience is transparent. All access rights to the data are preserved through the Rocket Arkivio Autostor migration policy and continue to be enforced through the Data ONTAP operating system.

In order to ensure that archived data is always available, even if the Rocket Arkivio Autostor system responsible for managing the data is not available, Rocket Arkivio Autostor implements $N + 1$



redundancy. A second Arkivio system can be registered with the Data ONTAP operating system as a secondary recall server for recalling data. If the primary system is unavailable, Data ONTAP will automatically request the data from the secondary system. There is no need for administrator intervention in order for the secondary system to automatically take over because of the intelligence Rocket Arkivio Autostor incorporates in the tag file itself.

Quantum Lattus API Integration

Rocket Arkivio Autostor provides built-in support for the REST API for storing objects (files) on the Quantum Lattus object storage platform which delivers the flexibility and scalability of a private cloud. Customers have the ability to create namespaces within the Lattus system as targets for files managed by the Rocket Arkivio software. Quantum's Fountain Erasure Coding technology enables data to be spread to multiple Lattus nodes which delivers the highest level of data reliability available today. The Lattus' ability to scale from 100's of terabytes to 100's of petabytes, without the need to rebuild the data when new storage is added, combined with the erasure code technology enables the Lattus storage platform to provide quick access to archived files across multiple, geographically distributed storage nodes. By implementing all the advantages of intelligent data management available in Rocket Arkivio Autostor in conjunction with the Quantum Lattus storage system allows clients to easily understand the data they have, determine how best to manage it, and direct it to an online archive storage platform that will support them for decades.

Tape/MO support

Rocket Arkivio Autostor supports a variety of tape and optical storage repositories, such as those from Quantum and Oracle (Sun/StorageTek) as well as major VTL and de-duplication solutions, such as EMC (Data Domain). These devices can be used as targets for data migration or true archive targets for long-term storage of data not required online.

Rocket Arkivio Autostor Architecture Overview

Rocket Arkivio Autostor is composed of two components: the Rocket Arkivio Central Server (ACS) and the Remote Server Assistant (RSA).

The ACS software resides on a dedicated Windows 2008 R2 Server and is the control center for all data collection, monitoring, and policy-based management. It consists of the Policy Automation Engine (PAE); a metadata repository; intelligent data movers; and monitoring, reporting, and web-based management interface components.



The PAE manages all automation actions performed by either the ACS or the RSA. The ACS is capable of monitoring and managing storage resources itself, and is able to manage resources via Remote Server Assistants.

The Remote Server Assistant (RSA) software resides on a dedicated Windows 2008 R2 Server (64-bit, Standard or Enterprise), so that additional server and storage resources can be managed without the use of agents. In addition to providing agentless management, the RSA provides scalability for managing resources within large enterprise environments by off-loading processing from the ACS. RSAs are deployed using automated push-install technology from the ACS central management console.

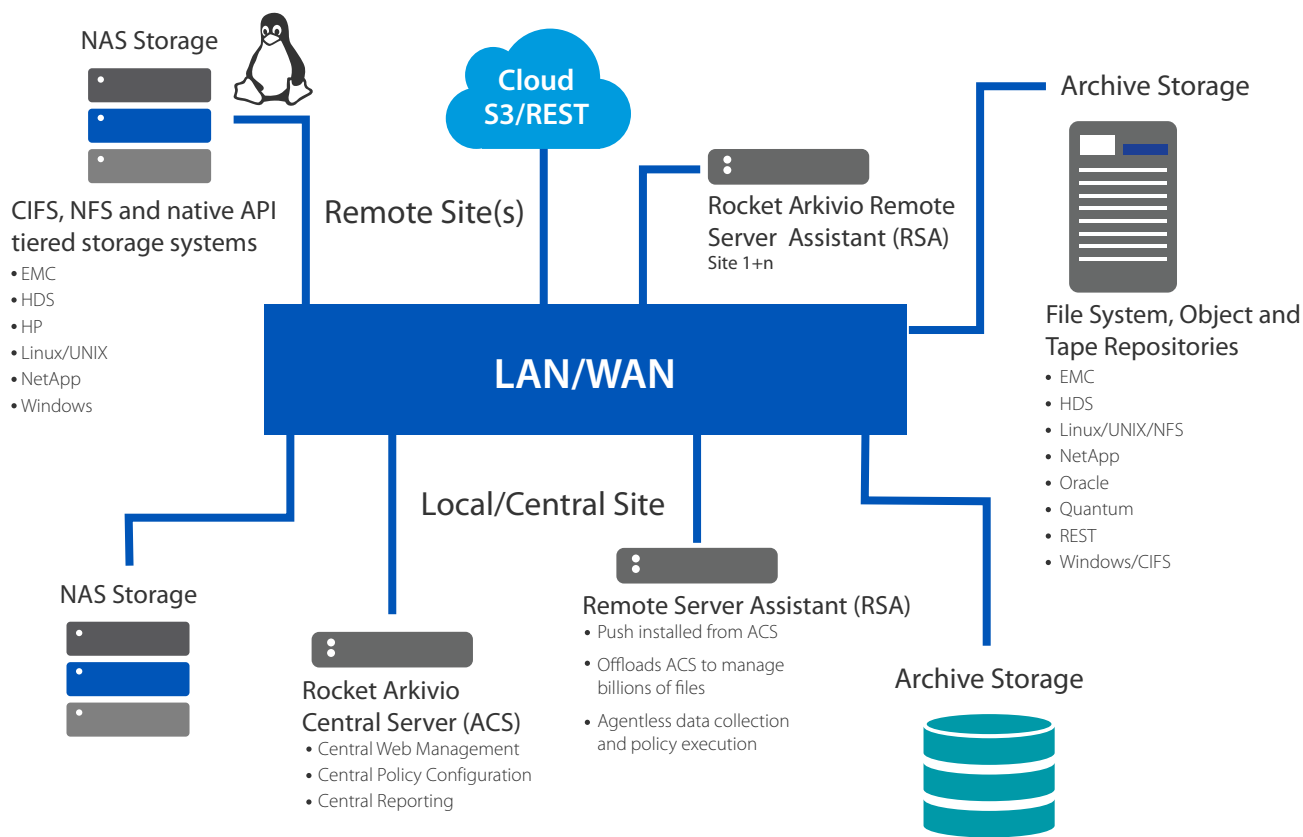


Figure 7: Rocket Arkivio deployment architecture

Figure 7 provides an illustration of the Rocket Arkivio Autostor deployment architecture when RSAs are deployed.



Scalability

Enterprise scalability can be achieved by adding RSAs to the system, which off-loads storage and data resource management from the ACS. A single ACS with a network of RSAs is capable of supporting billions of files and objects, or petabytes of data under management. Administrators can strategically assign different storage units to be managed by an RSA based on location, access load, and type of storage. For even larger environments, multiple Arkivio instances can be deployed.

Agentless Discovery and Data Movement

Rocket Arkivio Autostor is designed to function across heterogeneous file servers without deploying server agents. Unlike agent-based solutions which require proprietary software installation, maintenance, and labor cost, Rocket Arkivio Autostor reduces the total cost of ownership by:

- ❖ Allowing transparent deployment, without server or network modifications
- ❖ Eliminating the need for extra memory or CPU power on production systems
- ❖ Eliminating the need to take a production system offline for agent maintenance

Hardware and File System Independent

Rocket Arkivio Autostor supports CIFS, NFS, and REST protocols on major operating systems, storage servers, NAS, object storage, LTFS, and tape libraries. This means most servers, applications, and storage devices from major manufacturers can be supported in an intelligent data management environment including Windows, UNIX, Linux, Network Appliance, EMC (Atmos, Celerra, Centera, Data Domain, VNX), HDS, Quantum (Lattus, StorNext), HP, Cleversafe, Amazon S3, and many more (contact the Rocket Arkivio team to check if your specific storage platform is supported). Seamless cross-platform data movement is performed by one product, without the need for integration with other products. In addition, Rocket Arkivio Autostor provides the intelligence to select file system data that should be moved to Retention storage devices for compliance retention and archiving.

Content-based Global Policies

It is critical to base Intelligent Data Management solutions on the business value of the data; therefore, Rocket Arkivio Autostor is designed to take action only on the subset of data that is relevant to the business or data owners. Policies that act on these data groups can be set up and run on schedules that best fit the environment and the rate of change or growth that needs to be managed.

Modular Building Blocks

The Rocket Arkivio Autostor solution adopts a truly modular deployment methodology giving administrators the flexibility to define small building blocks that are reusable for policy definition and



report generation. For example, a data group can be defined based on attributes such as the type of file, combined with a path location. A group only needs to be defined once, and multiple policies can then make use of this component, simplifying the process when change is required. Additionally, Rocket Arkivio Autostor does not dictate the order in which components need to be defined. Instead, it lets users define them as the need arises. This modular approach to defining building blocks is the key to converting from a physical view of managing storage resources to a logical or business-value approach.

Capacity-based Alerts

Data management policies are often designed to help avoid critical thresholds, but there may be instances when policies alone cannot mitigate the situation. Rocket Arkivio Autostor generates capacity threshold alerts when a volume's capacity reaches an administrator-defined limit. In addition to the alert, policies can also be setup to automatically act on such volumes to relieve the constraint, independent of the normally scheduled policies. This helps companies avoid downtime due to full volumes or unexpected capacity shortages, and enables IT departments to better plan for maintenance and storage expansion.

Out-of-band

Rocket Arkivio Autostor operates out of the primary data path, requiring no changes to the way that users access the existing primary storage. Instead, all user behavior continues as before, with the high-availability and performance of primary storage systems preserved. Rocket Arkivio Autostor is only involved in the migration of data and, in some cases, when required by the object storage platform APIs, in the retrieval of that data back to the primary storage.

Anti-virus and Backup Software Compatibility

Anti-virus software and backup software are notorious for causing problems for system management applications. Agent-based software is difficult to upgrade and tag files can be unnecessarily recalled by anti-virus software and backup software. Rocket Arkivio Autostor uses no server agents and has been tested with major anti-virus software including McAfee, Norton, CA, and TrendMicro to ensure proper interoperability with these applications. Rocket Arkivio Autostor has also been tested with major backup software including Symantec, EMC, and CA to smoothly integrate with an organization's existing backup processes. Full-file recalls will not be initiated by the backup process, unless desired by the organization.

Disaster Recovery

The Rocket Arkivio Autostor solution is designed to handle disaster situations gracefully, allowing end-users to access their files without the immediate need for administrator action.



Disaster Recovery for ACS

The metadata repository on the Rocket Arkivio Central Server (ACS) can be easily rebuilt via an automated backup mechanism which takes snapshots of the system and stores them in a backup directory, which can then be protected by corporate backup policies. In the event of an ACS failure, the system can be quickly restored by redeploying the software on a fresh server and applying the latest snapshot. During the time that the ACS is offline, users can still access migrated files through the local RSA server, including access to files stored on a Retention Target storage system. This is possible because the Rocket Arkivio Autostor architecture is designed such that the ACS is not in the data path for accessing migrated data.

Disaster Recovery for RSA

The metadata repository on the Rocket Arkivio Remote Server Assistant (RSA) is designed to store only information collected on volumes and storage devices managed by the RSA. In the event of an RSA failure, the system can be rebuilt by simply push deploying the Assistant software onto a rebuilt server. While an RSA is down, the only lost functionality is the ability to execute scheduled policies for storage resources managed by the RSA. Read-access to link-migrated data still functions normally for CIFS and NFS target volumes. For tag-migrated data which utilizes the NetApp Open API, data can be made continuously available by designating an RSA as the secondary (N+1) server for recall. This designated “secondary” RSA is a fully functional member of the distributed Rocket Arkivio Autostor architecture, with its own storage management capabilities. For tag-migrated data which utilizes the EMC Celerra API or from Windows 2008 servers, the RSA does not have to be available for recall.

Self Description

All tag or link files created by Rocket Arkivio Autostor contain information that describes the target location of the associated data. This technology, called “Self Description”, ensures that data remains accessible independent of the preservation of any component of the Rocket Arkivio Autostor solution. Since the tag and link files are self-describing, the ACS and RSA servers do not require any prior knowledge of data migration tasks, nor do they perform bandwidth-intensive database replication to stay in sync. Instead, any data request requiring the interaction of the Rocket Arkivio Autostor solution is processed based on the information contained in the tag or link, resulting in maximum efficiency of data access. Further, the Self Description process ensures that tag and link files can be moved amongst managed systems just like any other files in the environment – allowing for traditional methods of storage allocation and management to proceed without any process change.



Metadata Accessibility

Each ACS and RSA maintains its own metadata repository that contains file and storage level information collected during the data collection process. In the event of a disaster, administrators will not have access to the metadata maintained by the specific ACS/RSA. Once the server comes back online, the repository is easily rebuilt by rerunning the data collection process, and the information will then be available for aggregated viewing and reporting.

Tag/Link Reconstruction

Rocket Arkivio Autostor is capable of dynamically reproducing tag files which were inadvertently deleted, moved, or lost as a result of a disaster. This provides the ability to quickly restore lost tag files without resorting to traditional backup and restore mechanisms. Tag files can even be eliminated from backup entirely, allowing for further optimizations of the backup infrastructure, while relying on the Rocket Arkivio Autostor solution to restore these if necessary.

Conclusion

IT departments can no longer afford to look to the storage industry to solve their data management problems by just providing faster and faster data storage and networking products. They need to demand solutions that track and intelligently manage file system data over their entire lifecycle. Rocket Arkivio Autostor provides an enterprise-class solution for gathering information on data and storage resources, grouping and classifying data with the ability to assigning business value, and automating data management actions to support the business requirements of the IT department. As a result, IT managers are able to demonstrate their contribution to the business objectives of the organization.

Rocket Arkivio Autostor Intelligent Data Management software has redefined storage management. It enables IT departments to understand the criticality of their data within their organization and automate its discovery, classification and placement on the most appropriate storage resource (Gather, Group, and Act) across heterogeneous DAS/NAS/SAN/Cloud environments without deploying agents – driving the cost per managed terabyte of storage to its lowest possible level, while providing significant ROI.

Call the Rocket Arkivio sales hotline at +1-650-237-6246 or visit the website at:

<http://www.rocketsoftware.com/brand/rocket-arkivio> if you would like to know more about Rocket Arkivio solutions and how you can implement a cost-effective, enterprise-class, intelligent data management system.

