DataFrameworks

ClarityNow! Overview

Table of Contents

Overview

Content generated by enterprises continues to proliferate. This growth strains enterprise storage infrastructure, business users, IT administrators and corporate IT budgets. While the price of storage continues to fall, the cost of managing complex data and storage infrastructure is increasing at a more accelerated pace. ClarityNow! uniquely responds to this growing problem.

ClarityNow! addresses management of unstructured data across complex heterogeneous storage networks. Our customers have found continued savings in OpEx and CapEx. Further, ClarityNow! is a flexible platform that provides structure for end users working on file based workflows, while maintaining an agnostic view of storage infrastructure.

Current file managers like Explorer, Finder, and Nautilus do not scale to manage multi-Petabyte environments. File system management often consists of data managers building internal scripts and home grown tools to index the collection of file systems. Spreadsheets are leveraged to present data and manually attach business tags and metadata in order to correlate to lines of business. The resulting data is often out dated upon delivery due to poor indexing capabilities. The entire process is rinsed and repeated as necessary to produce static report.

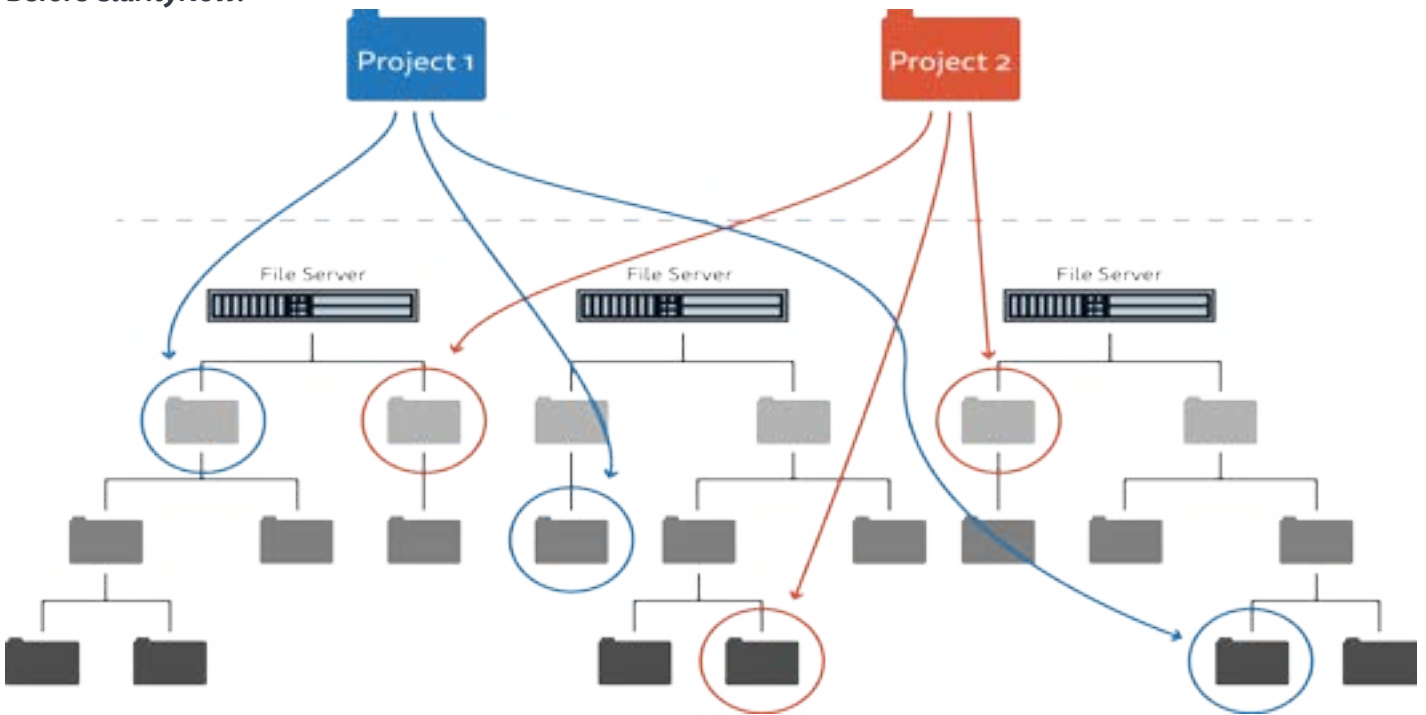# Before ClarityNow!

Before ClarityNow!

Fig. 1 - Before ClarityNow!

Before ClarityNow!

As file-based workflows continue to evolve, a variety of problems regarding the efficiency, effectiveness, and productivity of production-line data systems have become apparent. Because analyses of these production systems are often limited only to cleanup and crises, the infrastructure is often considered generally inefficient and unproductive, but without specific characterization or full root-cause analysis that would put a dollar figure on the issue and point the way to a solution.

An analysis of current file management is required to gain and understand cost and efficiency. Full characterization is akin to taking a wall to wall inventory, except the costs are even greater. Due to the sheer size (often hundreds of millions of files or more), the frequent churn, and the dispersed, heterogeneous nature of storage resources, enterprises almost always have great difficulty determining the status of their data production.

Superior File System Management

ClarityNow! provides functionality analogous to a file manager interface. Operating system file managers lack components required to manage large scale data and storage environments. For file system management in complex environments, best practice expertise is incorporated into ClarityNow! to address all large scale file based workflows.
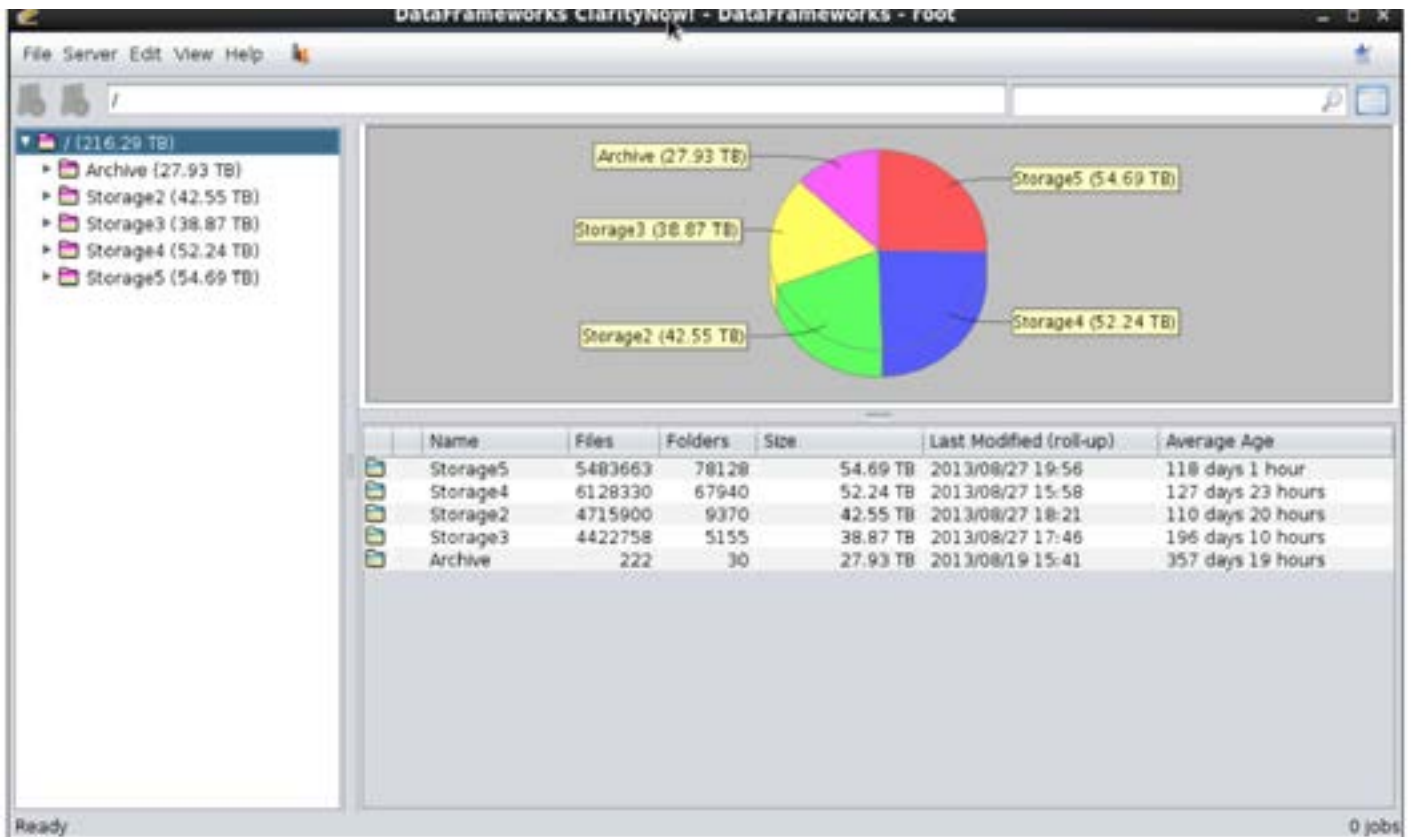

Fig. 2 — Storage resources

Unified View of Data and High Speed Search Across Storage Tiers

ClarityNow! benefits infrastructure workflow environments where data:

- Is moved between production and archive storage, end users need a mechanism to locate where data resides: a) production storage, or b) archive storage.

- Is distributed among various locations (i.e. different file systems, user directories, etc.)

ClarityNow! offers a unified view of all data with integrated high speed search. Below is a snapshot demonstrating the unified view (/archive, /aspera, /vol1, /vol2, /vol3). The search window shows all results for 'avatar' data.
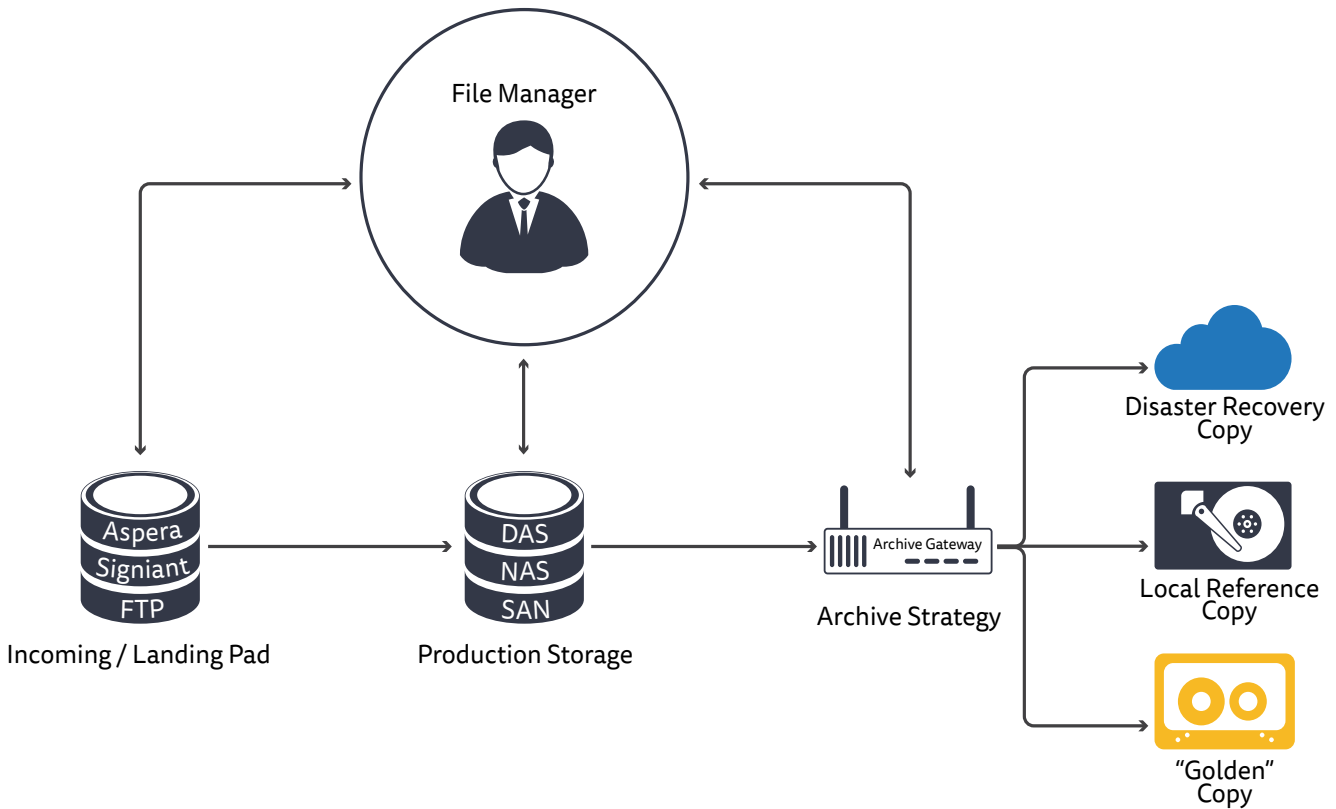

Fig. 3 — Logical representation of multiple copies of data cascading through the environment.

Presentation of Data in Actionable Format

Many causes of storage expansion are unavoidable. Increased service demands, higher resolution images, etc. However, a significant amount of capacity growth results from avoidable factors such as underutilization of existing capacity, storing duplicate copies of files, and retaining files on the data production line that are no longer required. The capacity growth drives costs higher despite the declining cost of per-unit storage.

DataFrameworks

Fig. 4 — Logical representation of multiple copies of data cascading through the environment.

Leveraging Intelligence for the "What to and when to archive" Quandary

ClarityNow! provides a flexible, industry standard file system based approach to higher level archive intelligence. Businesses can capture this value today with ClarityNow!

- Group and relate data to a project, product, job, business purpose, think of asset preservation. This data is what companies want to manage and archive for later usages, including monetization.

- Apply protection policies at the asset level, however some files within the asset may be tagged to always stay on disk in the HSM tier. Additionally, unnecessary files can be parsed out prior to archive.

- Unified views of all the data: incoming, production and archive by utilitizing unique scanning architecture.

- Provide high speed search across data silos (i.e. where is my data?, has it been archived? how do I find it?)

- Enable workflow automation with tagging and the ability to couple the file systems to key applications. One powerful ClarityNow! approach is to tag one attribute and use that attribute to query metadata from other applications. The business context is then visible and actionable within ClarityNow!.

- Report cost information across all file systems, projects, clients, archive, etc.

Chargeback, Showback

The policy of allocating resource expense back to the original or initial user or group allows companies to gain insight into workflow costs. The insight of showback as part of a chargeback policy enables the accountability of the resources to be viewed in a quantifiable way. Chargeback and showback promote awareness of resource expense within workflows.

Extensible to Incorporate into Workflow

ClarityNow! is designed to standardize, simplify, and enable seamless scaling of data and storage infrastructure through automation. ClarityNow!'s zero-footprint solution layers on top of existing file systems enabling business logic, policies and budgets to be incorporated into the file systems. ClarityNow! provides an essential abstraction layer within enterprise storage that powerfully connects business users with data assets, enabling:

- A collaborative business and file centric view for intuitive file system navigation;

- Enable naming convention compliance through audit and actionable visibility

DataFrameworks

- Incorporates business attributes into the file system enabling visualization of business logic, policies, and budgets

- Establishes normal use baselines and actively monitors file system for abnormal use; and

- Provides business facing reporting functionality delivering actionable information directly to business users.

Plug-In Ecosystem

The ClarityNow! Plug-In ecosystem has been architected to support multiple different types of plug-in functions to provide data creators and consumers a self-service capability, whereby users can extend the management characteristics of their ClarityNow! environment.

Plug-ins offer advantages such as the ability to send data to an Amazon S3 Bucket, or to support automatic transfer of meta data from ClarityNow! to delete data from predetermined areas on file systems without actually granting users file system access. The duplicate plugin for ClarityNow! is designed to leverage the fast, lightweight scanning mechanism within ClarityNow! to identify potential duplicates across all file systems, or subset thereof. The plugin will perform and return a high level identification based on filename, size, and last modified. Additional duplicate verification (i.e. MD5 checksums) can then be invoked upon user request for items within the high level results list. A plugin ecosystem enables the customer to take more of a vendor agnostic approach to any particular data silo and use the collection of underlying file systems as the silo maximizing the OpEx and CapEx of ClarityNow!. A sample list of plug-ins is below.

Amazon S3	Archive	ASG/Atempo	Audited Deleted
DMF Copy	SGI DMF	Duplicated	Filtered Archive
Media Info	Object Interface	Pixspan Compression	Restore

Fig. 5 — Sample listing of the plug-in types in the ClarityNow! Ecosystem

Security

ClarityNow! is used in highly-secure environments, including those that pass MPAA and Marvel audits. ClarityNow! can be used to separate data knowledge from data access, allowing certain users to view file systems without access to the file system data. Further, these non-file-system users can still perform specific, auditable operations via ClarityNow! plug-ins.

After ClarityNow!

ClarityNow! provides near-real-time, actionable views of your storage, via both raw file system views and business-aware views. ClarityNow!'s zero-footprint solution layers on top of existing file systems, so business logic, policies and budget information is at the user's fingertips while viewing storage. Advantages:

- Instantaneous search of file data across entire storage infrastructure allows fast answers to critical questions
- Tagging and instant rollups expose redundant and dormant data, leaving only profitable files on storage
- Automatic file and folder tagging enforces naming policies and enables reporting and automation
- Cost-based reports enable executive overview, with drillable answers to "why"
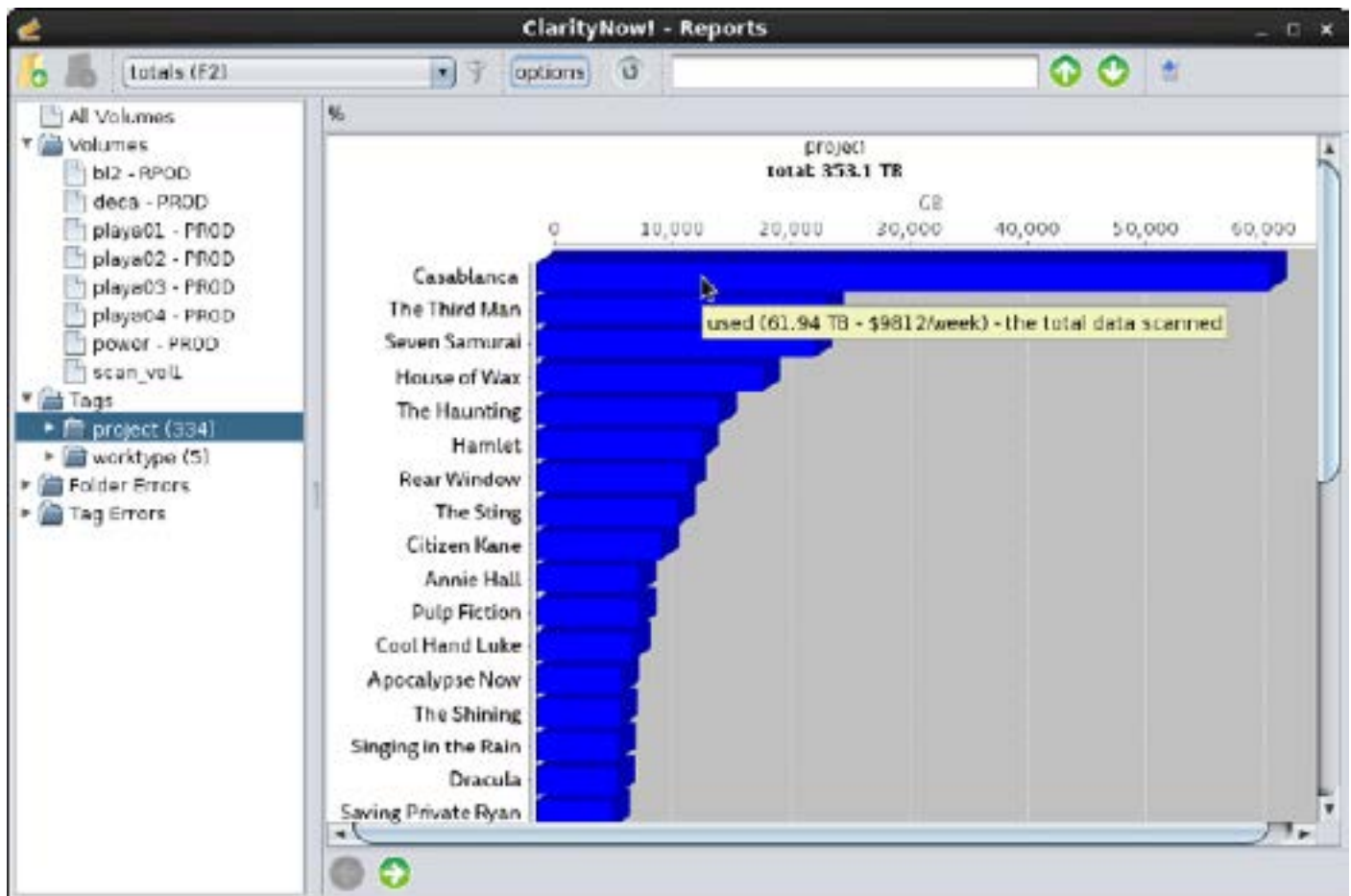- Workflow integration allows safe, auditable data movement (archive, retrieve, delete)

Fig. 6 — After ClarityNow!

Summary

As with most anything, the benefit of gaining an overall perspective is well understood and sought after. But such perspective in the increasingly complex data infrastructure topology of today's data centers has been more and more difficult to attain with its data silos, variants of unstructured data, disparate files systems and sheer data growth.

Yet the unified view, high speed search, segmentation, and access authorization capabilities of ClarityNow! offer IT managers an overall perspective that results in significant cost savings gained from the improvement in the overall infrastructure manageability it provides. Tool and utilities can be eliminated, while time efficiencies (in necessary human processes and analysis) are gained. These are real savings that go right to the organization's bottom line.

ClarityNow! is undoubtedly the industry standard choice for providing visibility and perspective to heterogeneous and complex data infrastructures, improving manageability, as well as significantly decreasing operational cost requirements.

DataFrameworks