Metadata Analytics for Infrastructure Assessment.

DATA
DYNAMICS

Table of Contents

BUSINESS CHALLENGE

Most organizations have some kind of Infrastructure monitoring. This can include not only availability, but also capacity and throughput. These are mission critical and cannot be overlooked. However, these systems cannot give organizations a picture of what data is where, how old, or how active the data is. There seems to be no shortage of studies that show unstructured data is growing at an alarming rate and managing that data can be another critical tool to organizations. This growth affects everything from budget, to risk, and becomes another for data governance data categorization.

This whitepaper is a high level overview of how to use StorageX Metadata Analytics to assess your NAS, unstructured data environment.

METADATA

Enterprises across all industries and geographies are moving to the Cloud to reap financial, operational, and technology benefits. Cloud strategy for each enterprise could vary according to their business goals and digital strategy. Most of them adopt the Cloud for improving efficiency, performance, and agility in the system. At the same time, others are focused on innovation and technological expansions. This topic could fill volumes, but we can summarize it with this SWOT analysis approach:

Table 1-Typical Metadata

|	Share or export	Directory (Path)	Filename	Other Metata
Examples	\\server1\share2 Linux1:/export2	\finance	invoice.pdf	File creation date File modified date File access date Owner (represented through security) Security bits (archive etc) Secuirty acls

StorageX store's all of the typical metadata but also adds several other pieces of information. The SID strings behind the actual security owner, the sid string of the machine, and the authority name are a few of these.

These can be used in combination with our discovery tagging to create queries to locate, categorize, and quantify the unstructured NAS infrastructure maintained in your organization.
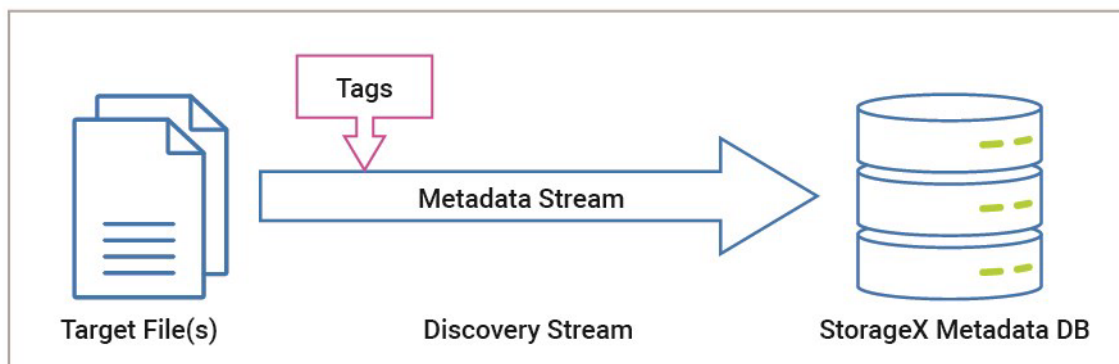
DISCOVERY

When you implement Metadata Analysis with StorageX, you are creating a process to gather detailed file information into a powerful database and build a set of queries to give you information about the unstructured data that's being stored. The data discovery setup is easy to implement but requires planning. Typically, Data Governance, Security, and IT would be involved in the planning. The idea is to create a set of standards by which you would search and categorize data by.

Figure 1- Overall Process
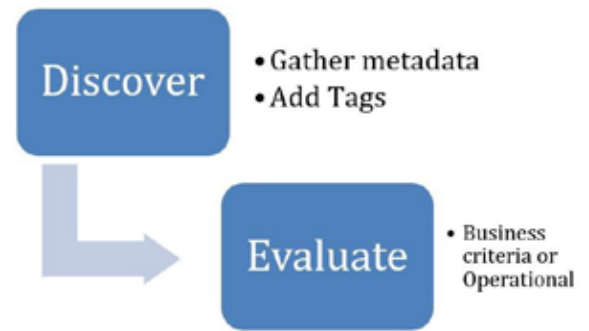
These standards should include:

˜ Data Tagging

   ▪ Creation of tags allows you to group data in a manner that matches your organization.

   ▪ E.g. Department=Finance, Sub_Department=AP

   ▪ The above would let you group all of finance but separate out the 'AP' department in queries.

˜ Standard Queries for Categorization

   ▪ These might be different for different organizations. Finance might have a 10 year hold on data, where as Engineering might have a 5 year hold on data.

   ▪ Because StorageX can do combined queries, it's possible to take multiple factors into consideration when assessing the infrastructure. Examples of this are:

      ˜ Files that were created 5 years ago, that haven't been accessed for 3 years, and were created by employees who have left the company.

      ˜ Files that were create by an employee who was let go for legal reasons

      ˜ Files with file types no longer used. (*.nsf)

      ˜ Files with particular bit settings (e.g. Archive bit)

˜ The planned combination of the above gives you an infinite possibility to evaluate the data.

˜ Each discovery stream can consist of multiple shares or exports.

By planning use of tags and queries the evaluation can be consistent with corporate or business unit goals.

DATA DYNAMICS

EXECUTION AND USE CASES

Executing on the overall plan, starts with discovery. When you create discovery sets, you can create datasets by either using the scan name or by combining tags and scan names.   Your next step is to evaluate. You can, of course, look at the entire dataset you've created. StorageX's ability to apply queries allows you to use the criteria created during the planning stage. Because you're using the DB in the background, you can try multiple queries to create the report that is relevant without 're-doing' the discovery.

Target Data	Use Case Description
Illegal Files	.In some organizations, certain files types could be banned. Examples of these include movie files, audio files or certain databases. By searching for the file type you can evaluate your risk. e.g. File extension equals ".mp4"
Outdated or Expired data	Metadata contains three dates to evaluate. These are created, modified, and accessed. You can use a combination or just one to evaluate by date.
Ex-Employees	1) In the CIFS protocol, when a user is removed from Active Directory, there is no longer a name associated with the Security Identifier. Many companies remove users names from Active directory. This causes 'Orphaned SIDS'. Named because the SID no longer has a name associated. A StorageX query that includes "Owner contains unknown" will return the list of files that are no longer associated with an employee. 2) In some companies, they rename the user with employees leave. An example of this is turning user 'Smith' into 'Smith_NLE'. This way they can track files and folders that were owned. In StorageX we can find these files with the query "Owner contains NLE".

These are basic examples of where you can start. The query language allows for 'grouping' and using "AND" and "OR" statements. An example of this could be:

Dataset=finance_scan AND

{Owner contains "unknown"

OR

Owner contains "_NLE"

OR

Owner contains "Fred"

}

AND Modified Date is older than 5 years

DATA DYNAMICS

Summary

StorageX's analytics capabilities give organization's the ability to perform a deep assessment of the infrastructure. There can be significant risk in not understanding depth and breadth of the file systems maintained in an organization. Data Dynamics has been used enterprise companies to assess the file infrastructure.   Some examples:

~ On the largest share: 78% of the file were owned by Orphaned SIDs, of that 90% hadn't been accessed in 3 years. (waste of resource)

~ On NFS exports, 9 Million files had not been accessed in more than 5 years

~ On another, 2.5M files where older than 10 years

~ On one share 6000+ files of old .BCK and .log files where more than 5 years old and had never been accessed. 10,000 were 3-5 years old. None have been created in the last 90 days (application retired)

Assessment is the first step, but taking action is the next. StorageX has a full complement of data movement capabilities. These include:

~ Standard Migration

~ Migration by assessment of individual folders

~ Archive to Object store

~ Replication to Object store

~ Workload tiering: Assessing the complete share and creating policies to move to an appropriate tier

About Data Dynamics

Data Dynamics is a leader in intelligent file management solutions that empower enterprises to seamlessly analyze, move, manage and modernize critical data across hybrid, cloud and object-based storage infrastructures for true business transformation.

Its award-winning StorageX platform eliminates vendor lock-in and provides a policy-based, storage management platform to provide the insight, agility, and operational efficiency to transform your data assets into competitive advantage. Used today by 26 of the top Fortune 100 companies, StorageX has optimized more than 350 PB of storage, saving more than 170 years in project time and $250 million in total storage costs. For more information, please visit: www.datadynamicsinc.com.

Data Dynamics, a global leader in enterprise data management, stands at the forefront of the industry-wide shift towards Digital Trust & Data Democracy. Trusted by 300+ organizations, including 25% of the Fortune 20, the company is recognized for its commitment to creating a transparent, unified, and empowered data ecosystem. Whether addressing data risk, privacy, sovereignty, optimization, sustainability, or facilitating seamless, policy-driven data migration across hybrid and multi-cloud environments, the company is ushering in a new era where data ownership, control, & actionability reside with the data owners.

DATA DYNAMICS®

Contact Sales   Book a Demo