Storage Optimization

> Setting up StorageX for Cost Reduction and Efficiency

DATA DYNAMICS

Table of Contents

DATA
DYNAMICS

Preface

Implementing Data Dynamics' StorageX software in your NAS storage environment will help you save money and create a more streamlined and efficient storage environment.

This document describes the various methods and use case outcomes using StorageX for dynamic file management.

Please refer to the latest admin guides for specifics and details of the hardware/software configuration.

Overview

Successful implementation of any strategy requires planning. You should gather requirements from your company's governance, business units, and IT groups.

The requirements gathered will be used to create policies, search criteria, and archive or workload tiering projects.
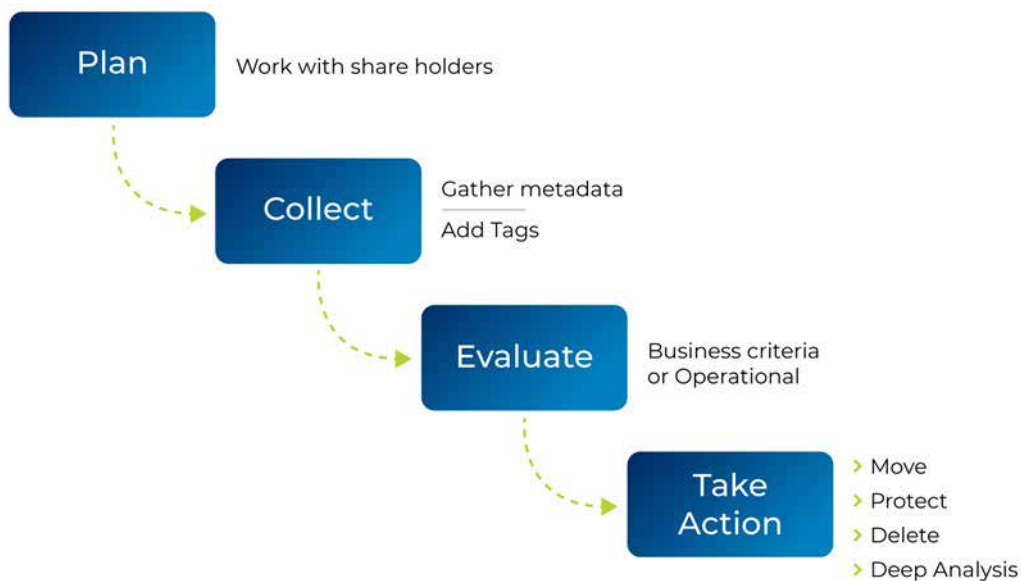

Figure 1: Process for Storage Optimization

Understand StorageX Modules for Storage Optimization
StorageX has several modules that can be used to achieve your goals.

File Replication	File Migration	Metadata Analytics
NAS to NAS file replication. Capable of one to many.	NAS to NAS file migration includes the capability to perform cutovers	Gathers all file metadata for users to evaluate their unstructured data and storage allocation
Uses: 1. Heterogenous D/R. 2. Replication for distribution	Uses: 1. Migration to new storage 2. Access to migration folders 3. Cloud Migration	Uses: 1. Discover ROT 2. Identify Duplicate Files 3. Evaluate files for Cloud Migration
Saves money by replicating to a cheaper storage	Capable of moving massive amounts of data efficiently	Deep metadata knowledge targets cost savings

Intelligent Archival	File to Object Transformation	Duplicate File Report
Using Metadata Analytics, users can target data sets that can be moved to object storage	Replicates a file share or export directly to the selected object storage	Identify duplicate files across shares or exports and provide a downloadable report
Uses: 1. Compliance 2. Cost savings/ rationalization	Uses: 1. Copy to S3 for sharing 2. Low cost backup	Uses: 1. Duplicate files can be deleted to regain storage space
Moving data to object storage can reduce overall storage costs	Object storage could be in the cloud or onsite	Storage savings.

Planning

Developing the plan starts with gathering the inputs from the stakeholders. These inputs will support the storage optimization plan. There also needs to be a plan for data placement. What systems are involved? What is their capacity and cost? Where are they?

Example:

Location	System	Tech/Protocol	Capacity	Cost	Owner
NY Datacenter	Systema-001	NFS/CIFS	100/140TB	A	IT
AZ Datacenter	Objecta-001	S3	10/1000TB	D	CISO
CH Datacenter	Win-123	CIFS	99/110TB	B	IT/Marketing

Understanding the physical infrastructure will help identify where you need to optimize and where to take advantage of additional storage space.

Include a column for "cost" and since cost can vary, we use a 'rating' of A to D. A is high cost (Tier 1 storage) and D is the lowest (object storage). IT can provide the total cost of ownership with exact numbers, but a range is a good place to start. In the example above, Win-123 is nearly full and is one tier below the highest cost. This could be a good place to start your assessment.

The methods to assess storage resources and orchestrate data using StorageX will also help you gather requirements for your storage optimization plan.

Method	Granularity	Data Orchestration	Module/License
Metadata Analysis	File	Individual files can be archived and removed.	Analytics /Archive
Meta Data Analysis	Share	Entire shares can be identified via Metadata and migrated to the target NAS	Workload Tiering
Migration Archive	Folder	Automatically creates a migration policy targeted the folders identified	Migration
Duplicate File Report	File	Works across volumes within the same dataset	Analytics

DATA DYNAMICS

However, there are times when you don't need to do an assessment.
1. If you buy new storage systems, use migration for quick adoption
   - Note: Using metadata analysis and archiving the result can significantly reduce the amount of data migrated.
2. If you need a quick backup, use Replication of File to Object.
   - This puts you in control of individual file restore

Planning the Discovery and Analysis

When StorageX is used to analyze metadata, it gathers the data into a database. StorageX also gives you the capability to add custom tags. These can help you categorize data later in analysis. They also can follow the file as http headers when it's transformed into an object.

Once the file metadata is in in the database, you can perform a limitless number of analysis based on data that matters to your business goals.

Establishing dataset criteria and standard set of queries is the next part of the plan. The dataset is created when create the Analysis. This can be a combination of:

- Discovery Dataset Name
- Discovery Dataset Description
- Custom Tags

Plan a standard for these items. Here are some examples.

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 infinite possibilities to evaluate the data.
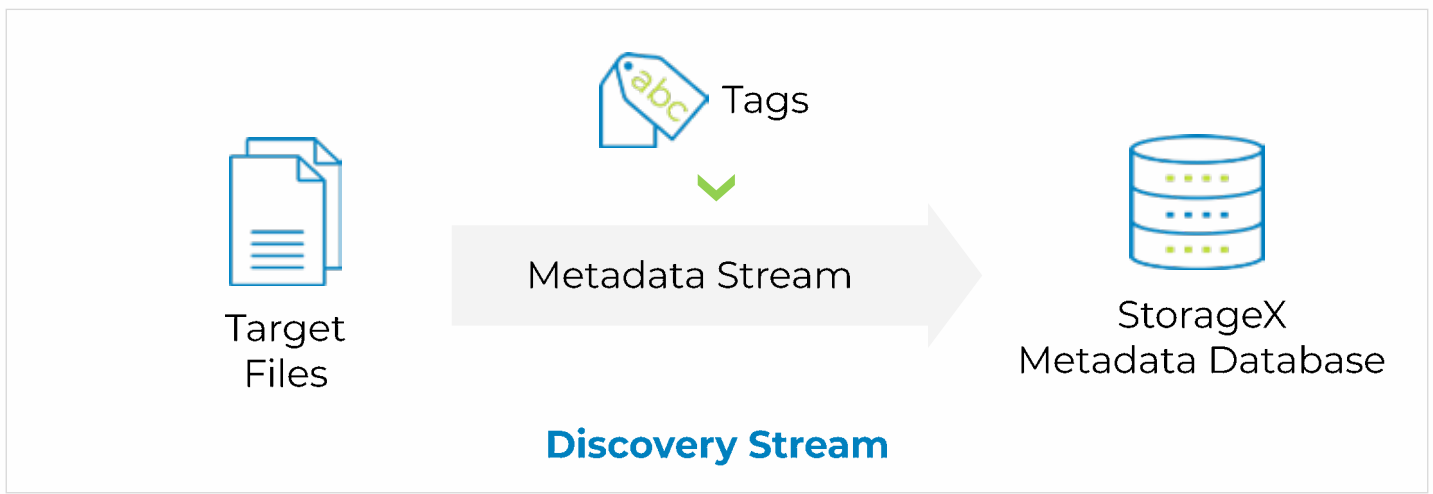Each discovery stream can consist of multiple shares or exports.

DATA DYNAMICS

Figure 2- Discovery stream

An example of tags:

Data Center	Share(s)	Tags
NYC	\\xntap32\finance \\xntap12\ap_share	DataCenter=NYC Org=Finance Sub-Org=AP
WDC	\\Isilon02\finusers \\Isilon02\ap_dump	DataCenter=WDC Org=Finance Sub-Org=AP
WDC	\\Isilon02\ar_dump	DataCenter=WDC Org=Finance Sub-Org=AR
Dallas	\\server4\Payroll	DataCenter=Dallas Org=Finance Sub-Org=Payroll

Keeping a coherent list of tags can help in planning. Note that StorageX uses ASCII characters, so capitalization matters. e.g. Finance != finance

Next, when planning discovery scans you also have choices. Examples of this are:

What shares or exports go into a discovery. StorageX allows 100's of shares to be grouped in a discovery. The tags you choose are associated with the file information brought in.
- Consider grouping shares/exports by department or project.
- Since the same tag can be applied to multiple discoveries, for later grouping.

Names and descriptions should mean something to the project/business or project management team. Consider you might have thousands and need to sort, organize and find discoveries.
- Examples might be:

Name	Descriptions	# of shares/exports
Project_202011_123	DataScience project for ROT in 2020	154
Finance_AP_exports	Scan Exports for Finance Compliance project	23
AP_share23	AP Share23 on server AP43-NYC1	1

- You can also consider putting the list of shares or export in your planning document.

Analysis Query Plan

The query plan can be to just find data for knowledge, but the key in creating a budget saving strategy is find data that can be moved or deleted.

StorageX applies queries so you can use the criteria created during the planning stage. Because you're using the database in the background, you can try multiple queries to create the report that is relevant without 're-doing' the discovery.

The possibilities have a wide range of capabilities and applying your business problem/solution is key. Some examples of 'what to look for' are shown below.

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 3 dates to evaluate. These are created, modified, and accessed. You can use a combination or just one to evaluate by date. Outdated files could also include application files from expired or unused applications. E.g, .nfs is used in lotus notes, so if that is not supported, it would be an extension to look for.
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".

You can combine queries to isolate the exact data you want to find.

DATA DYNAMICS

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:

Notes:
- StorageX stores Orphaned SIDs as "unknown".
- _NLE could stand for "no longer employed"
- Backup might be an application created file

Dataset=finance_scan AND
    {Owner contains "unknown"
    OR
    Owner contains "_NLE"
    OR
    Owner contains "Backup"
    }
AND Modified Date is older than 5 years

Planning Summary

Now, you have all the information you need to start your data orchestration process and optimize your unstructured data environment. This includes
    A list of target systems for discovery and the targets for data movement
    Discovery datasets with names and descriptions
    Custom Tags to help identify and group analysis
    Standard queries
Finally, your team should have contacted Data Dynamics support to understand policies and procedures so you can make the most of deep knowledge.

Support@datdyn.com

Execution of the Storage Optimization Process

Actual execution can will be somewhat fluid because as you perform analysis, you may, or may not find data to take action on. There should be a constant feedback loop to the stake holders on findings and resulting changes. This can help in both on-going processes and in managing business changes.

What follows is a discussion of the mobility features and their process. Note that these can be implemented via the Management console or by StorageX API's where noted.

File to File NAS Replication

There is no analysis other than business need for this replication. Business needs can be as simple as lower cost storage. You also might have an application need.

The replication can be setup by minute, hour, or daily. Remember that a file is moved based on being modified and won't be replicated until it's closed. This makes it great for general shares and exports and possibly application files. Things like database files will not work as well unless you can coordinate file closures. F2F replication can be integrated with mail for statuses.

Possible Use Cases:

- Quick replication of files as a quickly accessible backup for testing
- D/R in a heterogeneous NAS environment.
    E.g., Netapp Ontap share to Windows Server.
- Data distribution to a set of heterogeneous NAS servers
    E.g., Windows NFS export
      Isilon export
      Centos export
      NetApp export
    This might give several users in different datacenter the ability to test
    with a common dataset.

DATA DYNAMICS

Use Reporting to Identify Targets	Create Policies	Execute and Monitor	Establish Recovery Criteria
Share or Export Lists	Work with Busines to identify priorities	Evaluate the incremental time	Gain Agreement from Bus. to facilitate

Figure 3 - F2F Replication

Migration Workflow

The business need behind migrating data can be as simple as a new hardware platform. Using StorageX mitigates risk and also reduces cost through efficiency automation by having a centralize platform. StorageX Migration is capable of facilitating mergers, acquisitions, and even divestitures with its advanced Security Identifier management capabilities. The ability to do one to many or many to one type of migration gives the business units the capability to re-org their data as they move to a new structure.

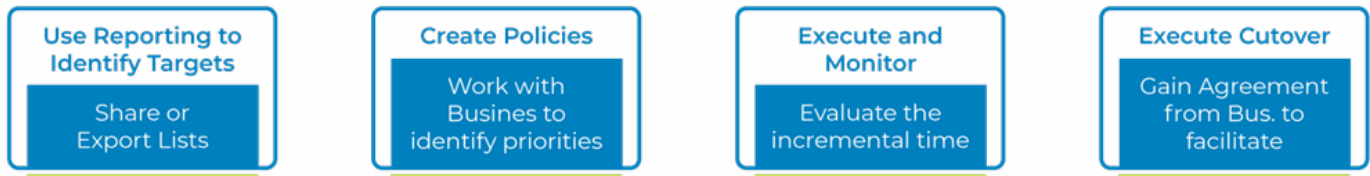| Use Reporting to Identify Targets	Create Policies	Execute and Monitor	Execute Cutover
Share or Export Lists	Work with Busines to identify priorities	Evaluate the incremental time	Gain Agreement from Bus. to facilitate

Figure 4 - Migration Policy, Business Need

With StorageX, the migration policies can be created via console or API. In the console, either targeting one by one, or importing them, 50 at a time via CSV file. You can also use StorageX API's create and manage migration policies.

The StorageX template configuration can be key in your planning and execution. The template for migration has over 60 different items that can be customized to make your migration successful. By using a template, you can faithfully create migration policies that will do exactly what you need. Here are some key features that should be part of your consideration.

- Scheduling. Keep in mind StorageX is moving files. Typically, customers schedule overnight once a day.
- Preserve last access time on source
- Delete orphaned files and folders. This deletes files on the target that were deleted on the source.

There are many options, and you should refer to the Migration administration guide for a full explanation.

> Note: If you are unsure of how a feature will work in your environment, it is advised that you a) test and b) work with Data Dynamics Support.

Some other items to consider.
- **Infrastructure**
    UDE placement, keep it close to the target data
    Overall sizing and configuration. Follow the admin guide, and consider working with your Systems Engineer (or support) to validate
    Multi Country? Some customers deploy one infrastructure per Continent.
- **Manpower**
    Take advantage of StorageX Training online
    Plan on connecting your StorageX instance to email so status's can be automatically delivered
    Plan on using RBAC to give the proper permissions.

Archive Workflow
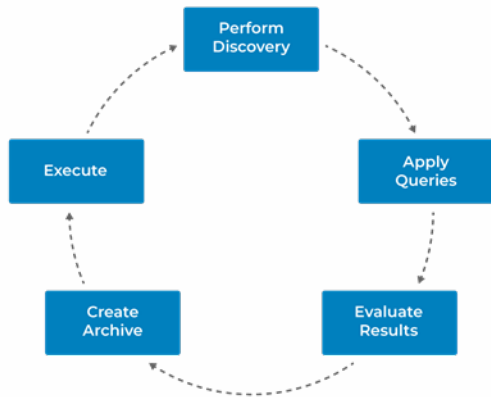
Note: Queries are directly applicable to archive policies. Choosing file modified date older than 5 years

Figure 5 - Archive Workflow

As each cycle finishes in the Archive workflow, the files that identified, are moved to the target object storage, and removed from the source system. This allows IT to reclaim space, lessen backup time, and reduce costs. Each archived file is written with a unique ID. This guarantees there will no two objects with the same name. Additionally, StorageX writes a companion object. This is specially created at a JSON formatted mini database of every piece of metadata. At the very least, this companion object can tell, without a doubt, exactly where and when the primary object was created, as well as who did it.

Possible use case for the companion object include applying Machine Learning, AI, and custom queries. With these options finding long term patterns for any piece of meta data is possible. This can lead to predictive analysis and proactive remediation.

Migration Policy, Archive Workflow

This feature allows you to evaluation down to the folder level. These are based one or more restricted list shown below.

Note that in Migration Archive, there some choices that are not in the traditional meta data. Folder is more than x old, size, as well as number of files are part of this analysis as an example

The archive template in Migration, has the ability to generate policies that rationalize folders against a given set of meta data. While you can choose to automatically generate the migration policy, the workflow below breaks it out to assessment and the action of performing the migration.
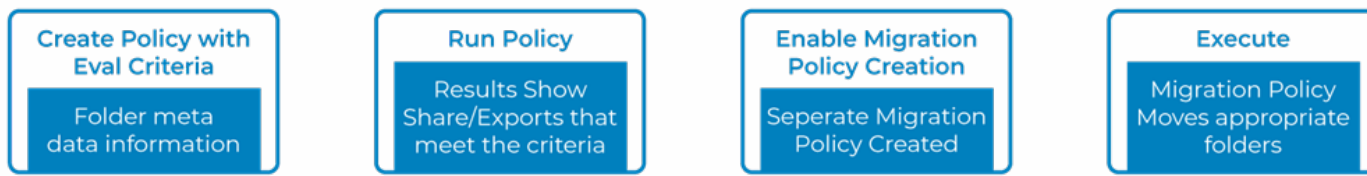

Figure 6 - Migration Policy, Archive Workflow

The output will specific folders and will include sub-folders. In Archive-Migration the policy is a migration policy.   If you want to delete the source data, you'll need to either delete it manually or create a script that automatically deletes the source. This can be achieved easily and generically, as StorageX passes the source and target into scripts being executed.

Workload Tiering

In workload tiering, you evaluate entire shares based on the metadata criteria selected. An example of this might be:

In this example 90% of the query must meet the criteria. In this case the owner contains orphaned SIDS and they have not been modified in more than 5 years.

> Dataset=finance_scan AND
>
> 90% or more
>
> > {Owner contains "unknown"
> >
> > }
>
> AND Modified Date is older than 5 years

The start of this in the discovery. The goal is to find shares that you want to move to a different tier. The focus on this is to move from an expensive platform (ssd?) to a cheaper, NAS based storage (Sata?). You and also evaluate based on the opposite.

Either way, you once you've decided, the output of your evaluation will be migration policies that will move the data to the targeted storage.

Some things to consider:
- Choose a migration template with the proper setup for your environment.
  - Refer to the storagex migration admin manual for more details
- In order to get back your storage space, consider automatically deleting the source after the cutover is done. This can be scripted and generic fashion as StorageX will pass the source and target as %1 and %2 in batch files.
- Before cutovers of the migration, create a ticket with DD Support.
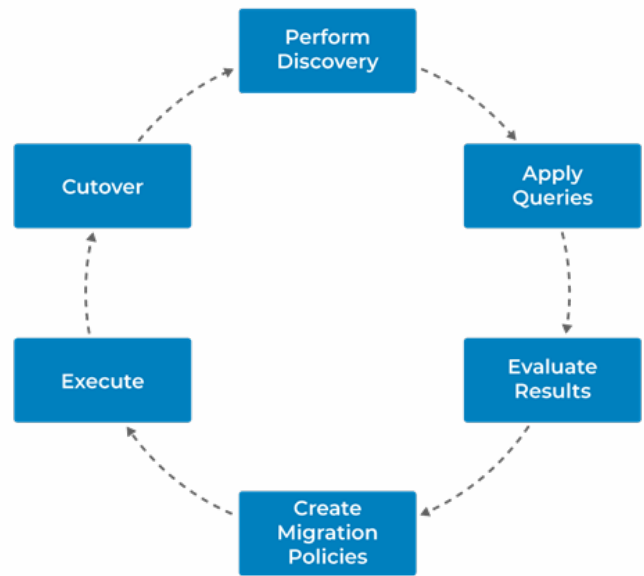
Figure 7 - Workload Tiering

Duplicate Files

Duplicate files can take up a major amount of space, with little or no return to the business. These are generated from a variety of methods. StorageX can create lists of these files. These lists can include:

- File Name
- File Type
- Creation Time
- Last Access Time
- Last Modified Time
- Size in Bytes
- Resource – Server name
- Platform – e.g. NetApp 7 mode
- CIFS Access Path

- CIFS File Owner
- Export Path
- ACL
- Share Name / Export Name
- NFS User ID
- NFS Group ID
- Sticky Bit
- Offline
- Mode bits

To start you need to run a discovery with an md5 hash. The discovery can be on multiple shares and multiple source servers. When you run the duplicate file report, it delivers all the files that have matching names and MD5 hashes. Because there are so many possibilities, StorageX leaves the remediation to the customer. However, you can output a report for a custom script to take action
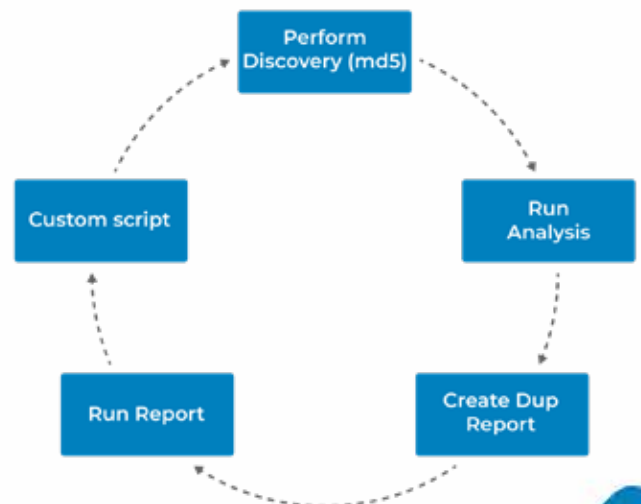
Figure 8 - Duplicate Files

DATA DYNAMICS

Appendix

What follows are items that could be of interest or used in planning. Note that StorageX has infastructure outputs for things like lists of shares or exports.

NAS Infrastructure sheet

LOCATION	SYSTEM	TECH/ PROTOCOL	CAPACITY	COST	OWNER
NY DATACENTER	SYSTEMA-001	NFS/CIFS	100/140TB	A	IT
AZ DATACENTER	OBJECTA-001	S3	10/1000TB	D	CISO
CH DATACENTER	WIN-123	CIFS	99/110TB	B	IT/MARKETING

NAS export/share list

SYSTEM	SHARE/ EXPORT	PRIMARY OWNER/ STAKEHOLDER	MAX USERS	USED STORAGE	TOTAL STORAGE
\\SYSD-ABC	Share1	Finance	2029	15TB	17TB

S3 Target list

VENDOR	URL	ACCOUNT NAME	ACCESS KEY ID	SECRET KEY	BUCKET
SG	Abc:8082	Sx1	Nnn-xxxx-nnn-xxxss-xxxx	sdfsdafasf	Finance1

DATA DYNAMICS

Companion Object

The out of the archive policy is not only the file transformed into an object but StorageX also creates a companion object. This companion object is like a fingerprint of the file. This can be used for compliance or for AI/ML applications as time goes by.

Example Contents:

```
{
  "CifsMetadata": {
    "AllocationSize": 60325888,
    "ExtendedAttributesSize": 0,
    "FileAttributes": 32,
    "FileSecurity": {
      "ControlFlags": 32788,
      "Dacl": {
        "Aces": [
          {
            "AccessMask": 2032127,
            "AceFlags": 16,
            "AceType": 0,
            "MachineName": "ntap-81-7m.training.local",
            "Trustee": {
              "AccountName": "storagex",
              "AuthorityName": "TRAINING",
              "MachineName": "",
              "SidString": "S-1-5-21-801905265-1630774311-3690633751-1606"
            }
          },
          {
            "AccessMask": 2032063,
            "AceFlags": 16,
            "AceType": 0,
            "MachineName": "ntap-81-7m.training.local",
            "Trustee": {
              "AccountName": "Everyone",
              "AuthorityName": "",
              "MachineName": "",
              "SidString": "S-1-1-0"
            }
          },
          {
            "AccessMask": 2032127,
            "AceFlags": 16,
            "AceType": 0,
            "MachineName": "ntap-81-7m.training.local",
            "Trustee": {
              "AccountName": "SYSTEM",
              "AuthorityName": "NT AUTHORITY",
              "MachineName": "",
              "SidString": "S-1-5-18"
            }
          },
          {
            "AccessMask": 2032127,
            "AceFlags": 16,
```

```
"AceType": 0,
        "MachineName": "ntap-81-7m.training.local",
        "Trustee": {
         "AccountName": "Administrators",
         "AuthorityName": "BUILTIN",
         "MachineName": "",
         "SidString": "S-1-5-32-544"
        }
      }
     ],
     "MachineName": "ntap-81-7m.training.local"
    },
    "Group": {
     "AccountName": "Domain Users",
     "AuthorityName": "TRAINING",
     "MachineName": "",
     "SidString": "S-1-5-21-801905265-1630774311-3690633751-513"
    },
    "MachineName": "ntap-81-7m.training.local",
    "Owner": {
     "AccountName": "bowie",
     "AuthorityName": "TRAINING",
     "MachineName": "",
     "SidString": "S-1-5-21-801905265-1630774311-3690633751-19111"
    },
    "Revision": 1,
    "Sacl": {
     "Aces": [],
     "MachineName": "ntap-81-7m.training.local"
    }
   },
   "Group": "TRAINING\\Domain Users",
   "IsNormal": false,
   "IsSystem": false,
   "IsTemporary": false,
   "Owner": "TRAINING\\bowie"
  },
  "Data": {
   "FileStreams": [
    {
     "BytesToEndOfFile": 60322320,
     "IsCifsFileStream": true,
     "IsNfs3FileStream": false,
     "Name": "",
     "ObjectTransferRecords": [
      {
       "ObjectRecords": [
        {
         "AccessId": "P8BWZRPDUX4EYQH8XW14",
         "AuthRegion": "",
         "EndpointUrl": "https://webscaledemo.netapp.com:8082/pfeiff",
         "Index": 0,
         "ObjectKey": "5cc462d6-5439-4d79-ac2b-84aeaab51658",
         "StorageType": 0
        }
       ],
       "TransferTime": 132042293278204980
      }
     ],
```

```
      "Size": 60322320
    }
  ],
  "IsCifsFileData": true,
  "IsNfs3FileData": false,
  "NumberOfFileStreams": 1,
  "TotalSizeOfFileStreams": 60322320
},
"GeneralMetadata": {
  "AccessPath": "\\\\ntap-81-7m.training.local\\TestData1\\Movies\\The Goonies.mov",
  "ChangeTime": 131962702716465900,
  "ConnectionPoint": "TestData1",
  "CreationTime": 129829827530000000,
  "FilenameExtension": ".mov",
  "FilesystemType": "NTFS",
  "IsArchive": true,
  "IsCompressed": false,
  "IsDirectoryItem": false,
  "IsEncrypted": false,
  "IsFileItem": true,
  "IsHidden": false,
  "IsOffline": false,
  "IsReadOnly": false,
  "IsReparsePoint": false,
  "IsSparse": false,
  "IsSymlink": false,
  "LastAccessTime": 130543805110000000,
  "LastModifiedTime": 130540764460000000,
  "LocalPath": "/vol/TestData1",
  "MachineName": {
    "AccessName": "ntap-81-7m.training.local",
    "DomainFullyQualifiedName": "ntap-81-7m.training.local",
    "DomainName": "training.local",
    "KnownIpAddresses": [
      "10.0.0.20"
    ],
    "KnownNames": [
      "ntap-81-7m.training.local"
    ]
  },
  "MimeType": "video/quicktime",
  "Name": "The Goonies.mov",
  "PlatformDescription": "Data ONTAP 8.1.4 7-Mode",
  "ShortName": "THEGOO~2.MOV",
  "Size": 60322320
},
"NetAppMetadata": {
  "ContainerName": "vol0",
  "HostingResourceName": "ntap-81-7m.training.local",
  "Is7mode": true,
  "IsCmode": false,
  "IsFlexGroup": false,
  "IsFlexVol": true,
  "IsInfiniteVServer": false,
  "IsTraditionalVol": false,
  "IsVirtualResource": false,
  "LanguageCode": "en_US",
  "ONTAPGeneration": 8,
  "ONTAPMajorVersion": 1,
```

    "ONTAPMinorVersion": 4,
    "ONTAPVersion": "NetApp Release 8.1.4 7-Mode: Wed Nov 20 16:16:17 PST 2013",
    "Oplock": "enabled",
    "QtreeName": "",
    "SecurityStyle": "ntfs",
    "VirtualResourceName": "vfiler0"
  },
  "ObjectTransferRecords": [
    {
      "ObjectRecords": [
        {
          "AccessId": "P8BWZRPDUX4EYQH8XW14",
          "AuthRegion": "",
          "EndpointUrl": "https://webscaledemo.netapp.com:8082/pfeiff",
          "Index": 0,
          "ObjectKey": "7e65b4aa-330d-435f-8e07-c72e70bc56c8",
          "StorageType": 0
        }
      ],
      "TransferTime": 132042293278204980
    }
  ],
  "Tags": [
    {
      "Name": "isx-job-run-id",
      "Value": "2cb410ac-6ce2-42e1-b993-e48186383cfd"
    },
    {
      "Name": "isx-archive-policy-name",
      "Value": "Final_Archive"
    },
    {
      "Name": "Datacenter",
      "Value": "Skytap"
    },
    {
      "Name": "Location",
      "Value": "New_Jersey"
    },
    {
      "Name": "Organization",
      "Value": "Marketing_All"
    },
    {
      "Name": "Project_Name",
      "Value": "Alan_Parsons"
    },

```
    {
        "Name": "Project_Number",
        "Value": "867-5309"
    },
    {
        "Name": "isx-scan-description",
        "Value": "Scan_for_Rock_Stars"
    },
    {
        "Name": "isx-scan-name",
        "Value": "Rock_Star_Scan"
    }
  ]
}
```

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