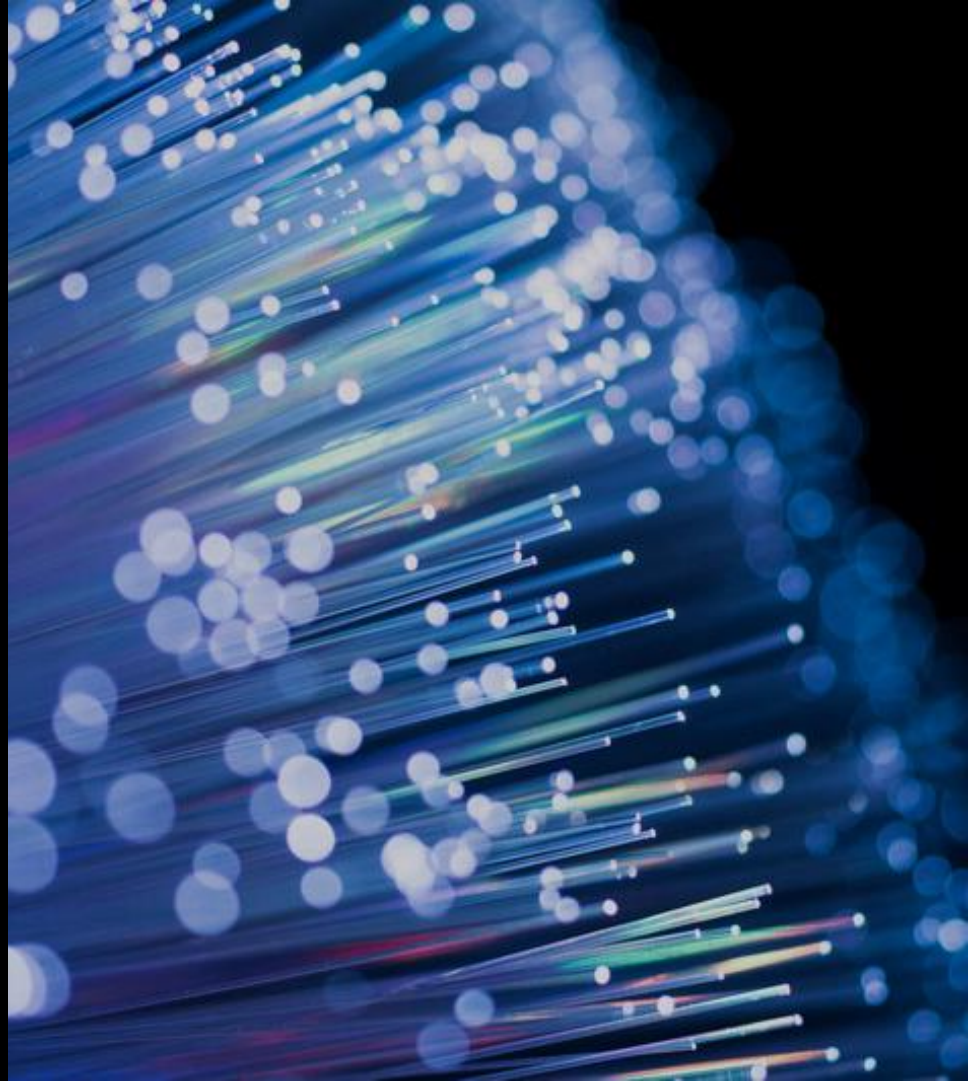
Storage Scale and Storage Protect/Archive



Disclaimer

IBM's statements regarding its plans, directions, and intent are subject to change or withdrawal without notice at IBM's sole discretion. Information regarding potential future products is intended to outline our general product direction and it should not be relied on in making a purchasing decision. The information mentioned regarding potential future products is not a commitment, promise, or legal obligation to deliver any material, code, or functionality. The development, release, and timing of any future features or functionality described for our products remains at our sole discretion.

IBM reserves the right to change product specifications and offerings at any time without notice. This publication could include technical inaccuracies or typographical errors. References herein to IBM products and services do not imply that IBM intends to make them available in all countries.

Scale and Protect/Archive - should we talk about HPSS?

Data exchange

➔Leverage import / export function, copy function and standardized format

Exchange large volumes of data

Provide access via global name space

Tiered storage

➔Leverage simplicity and tape TCO

Policy based placement and migration

Use tape for infrequently accessed files

Archiving

➔Leverage simplicity, tape TCO

Archive large volumes of data / files

Use policy based migration to tape

Retain data for long periods of time

Backup

➔Leverage simplicity and tape TCO

Backup data to managed file system

Immediate migration to tape

Easy data access and restore

IBM Storage Protect/Archive & Fusion Data Cataloging Service (Discover)

IBM Fusion Data Cataloging Service (Discover)

- Understand the state of files in the file system
- Build and deliver reports that include the active and archive parts of the filesystem
- Create a full view of the unified file system including all storage pools both Internal and External.
- Understand temperature / access times of files and data to be able to adjust policies accordingly and understand the usage pattern of the data
- ScaleILM application to move data to different tiers (pools) that are configured on the IBM Storage Scale connection including IBM Storage Archive

Consistent View File Status

IBM

Storage Scale backup and restore

IBM

Storage Scale with Storage Protect for Space Management

IBM

Scale Out Backup and Restore (SOBAR)

IBM

Storage Archive

IBM

Safe Guarded Copy with Tape

Storage Scale backup and restore

IBM Storage Scale Backup

New!

Simple licensing option

For IBM Storage Scale and IBM Storage Scale System solutions

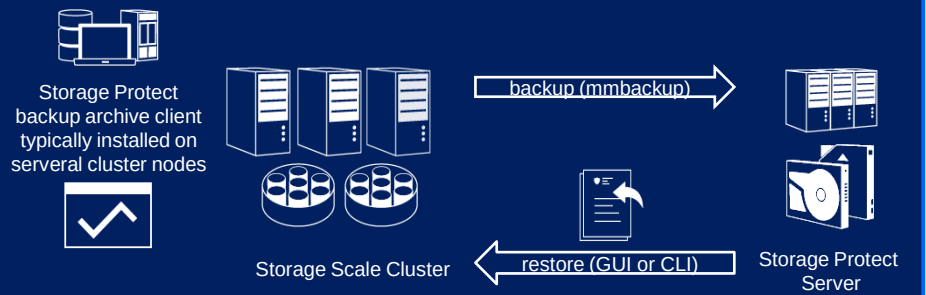
Available only to new and existing Storage Scale and Storage Scale System Customers

- Backup
- Space Management for Storage Scale (HSM)
- Fast Disaster Recovery

IBM Storage Scale Backup is a seamless, integrated, backup and archiving solution for data protection and cyber security requirements

IBM Storage Scale Backup combines

- *IBM Storage Protect Extended Edition (5725-W99)*
- *IBM Storage Protect for Space Management (5725-X04)*
- *Offering. that is priced per terabyte, in alignment with IBM Storage Scale capacity editions**



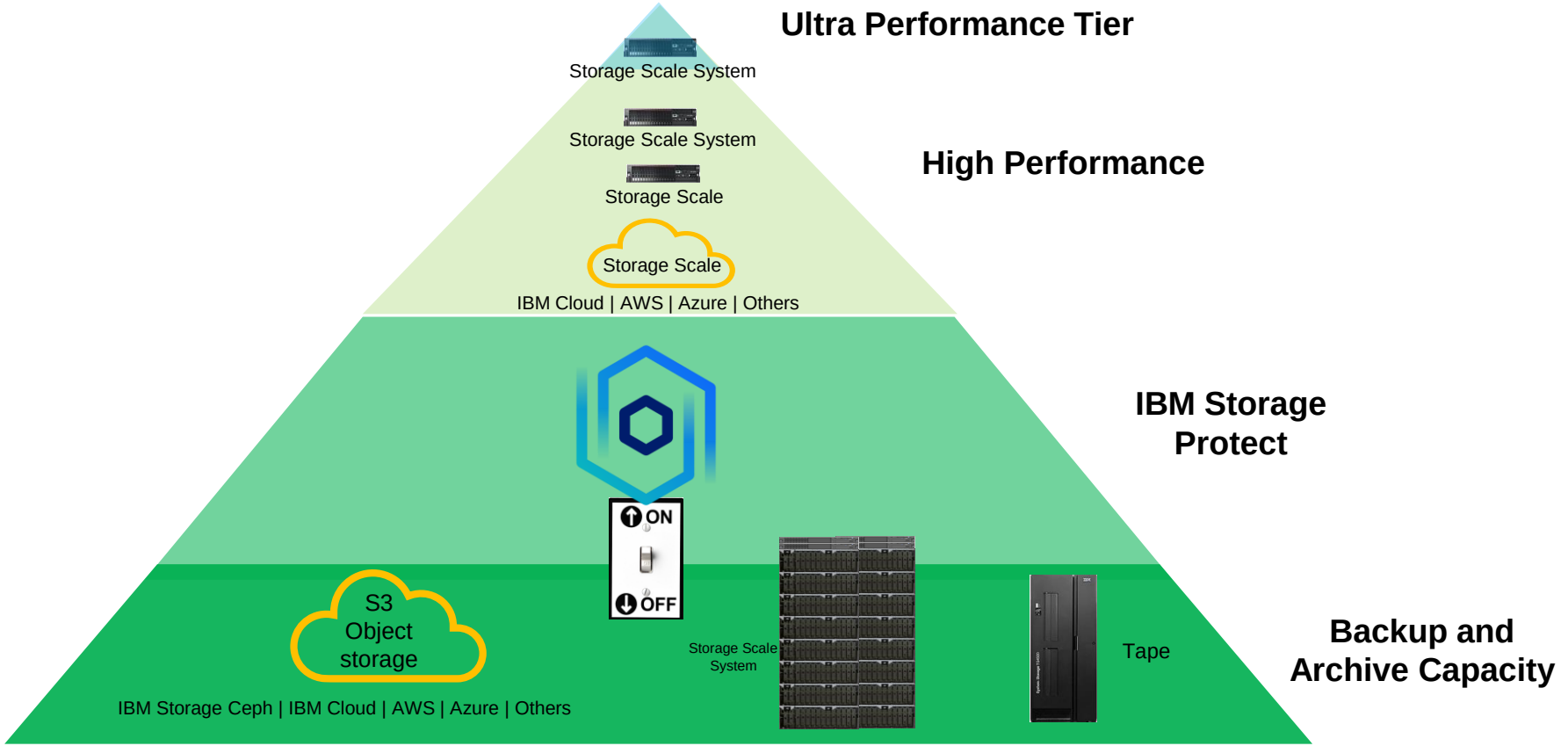
*Not available for socket editions of Storage Scale
**Not Available in EMEA in initial offering

Use Cases

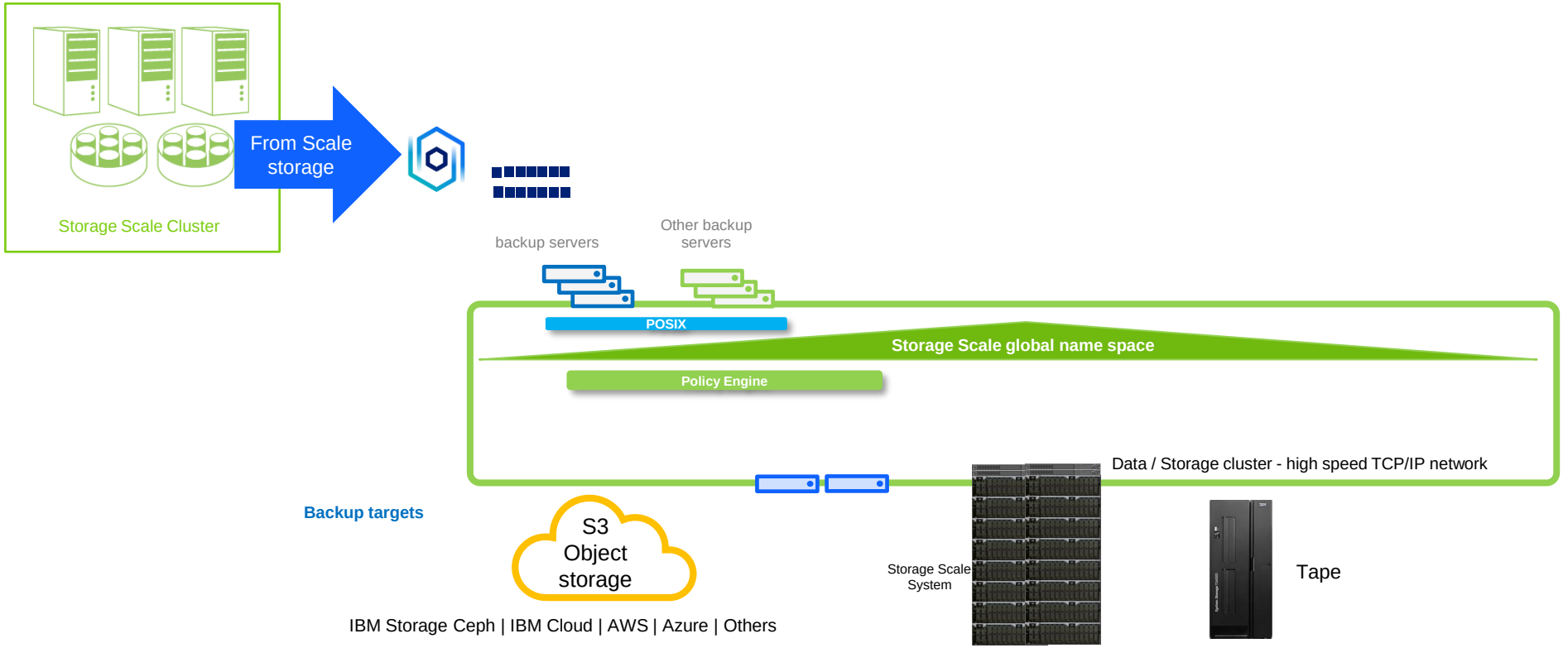
Storage Scale Backup Use Case	Storage Scale Backup Entitlement
Backup Storage Scale or Storage Scale System Data to Storage Protect	• *Storage Scale works with Storage Protect as a Backup target* • *Storage Protect is member of Scale Cluster directly accessing file system*
Utilize Storage Scale or Storage Scale System to contain Storage Protect Storage Pools	• *Storage Protect also provides Space Management functions for Storage Scale (HSM)* • *Storage Protect is member of Scale Cluster directly accessing file system*
Fast Disaster Recovery (DR) **Utilize Storage Scale or Storage Scale System for DR – Scale-out Backup and Recovery (SOBAR)**	
Storage Scale and Storage Scale System is used as a High-Performance Tier as front end to Storage Protect backup solution	• *Using AFM, Storage Scale and Storage Scale System is deployed as a caching tier in front of a Storage protect backup Solution*

Management Services Transparently Optimize Capacity

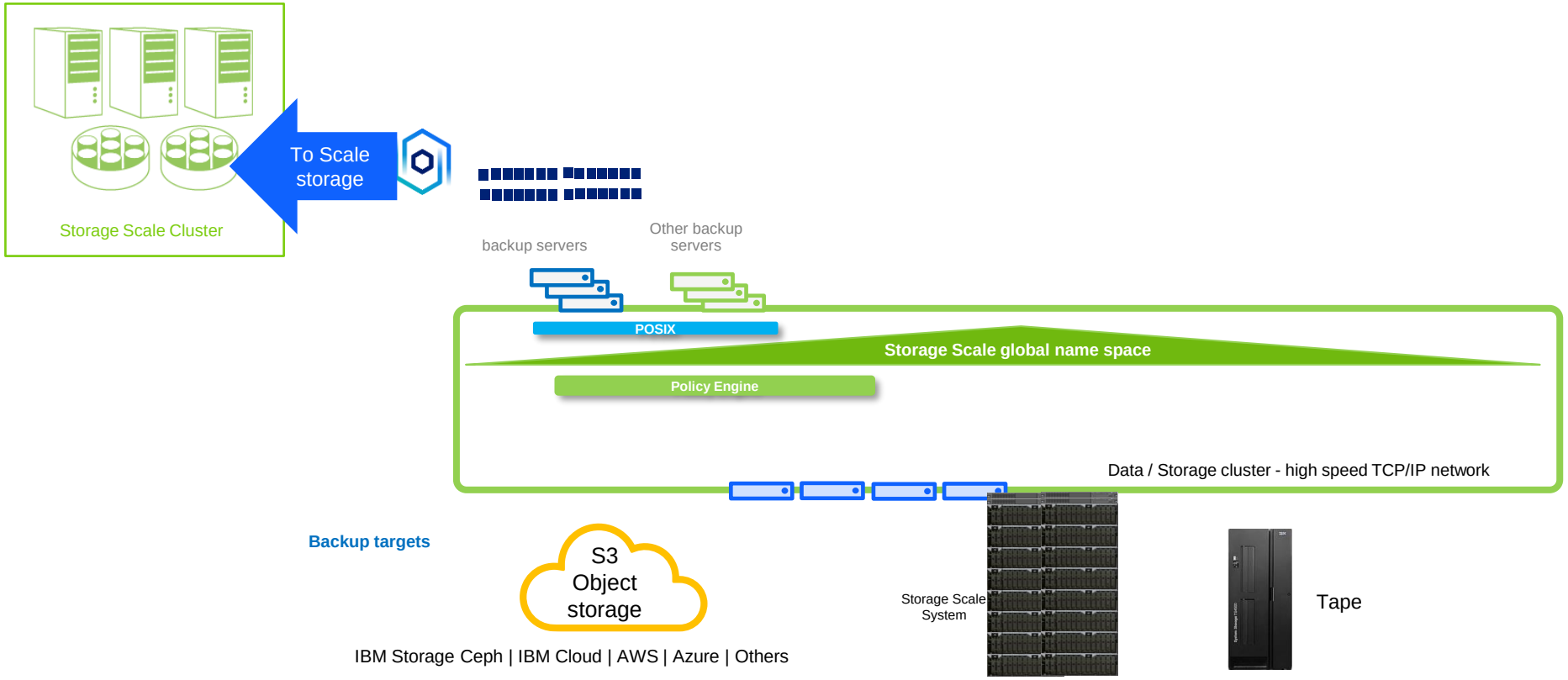
Lowers Energy Consumption and Costs with Data Lifecycle Management



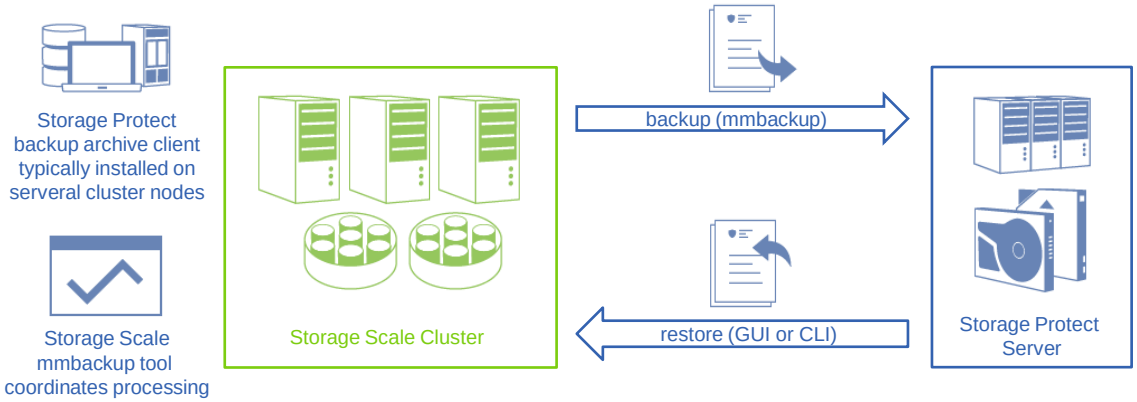
Backup Scale



Restore to Scale



Scalable Backup And Restore



- Massive parallel filesystem backup processing
- Storage Scale mmbackup creates local shadow of Storage Protect DB and uses policy engine to identify files for backup
- Storage Protect backup archive client is used under the hood to backup files to Storage Protect Server
- Storage Protect restore (CLI or GUI) can be used to restore files

Backup recommendations

- Do not mix platforms, run mmbackup on either AIX, xLinux, pLinux, zLinux nodes

- Consider additional workload generated by backup runs
  - mmbackup generates workload on nodes and storage
  - When backing up multiple file systems / file sets then serialize backups

- Backup from snapshots prevents skipped files

- Rename of files or directories and ACL changes cause backup of these files
  - When HSM is used this may require recall
  - **Scale/Protect teams have scoped this work to resolve. Just a matter of an RFE/IDEA, resources and time.**

Align backup parameter among the components

Define number of bytes to be used in a single backup transaction

`--max-backup-size` ——— `TXNBYTELIMIT`

Define number of files to be used in a single backup transaction

`--max-backup-count`
`--max-expire-count`
——— `* N (number of nodes)` ——— `TXNGROUPMAX`

Define number of parallel backup processes

`--expire-threads`
`--backup-threads`
——— `* N (number of nodes)` ——— `RESOURCEUTILIZATION` ——— `MAXSESSIONS`
`MAXNUMMP`

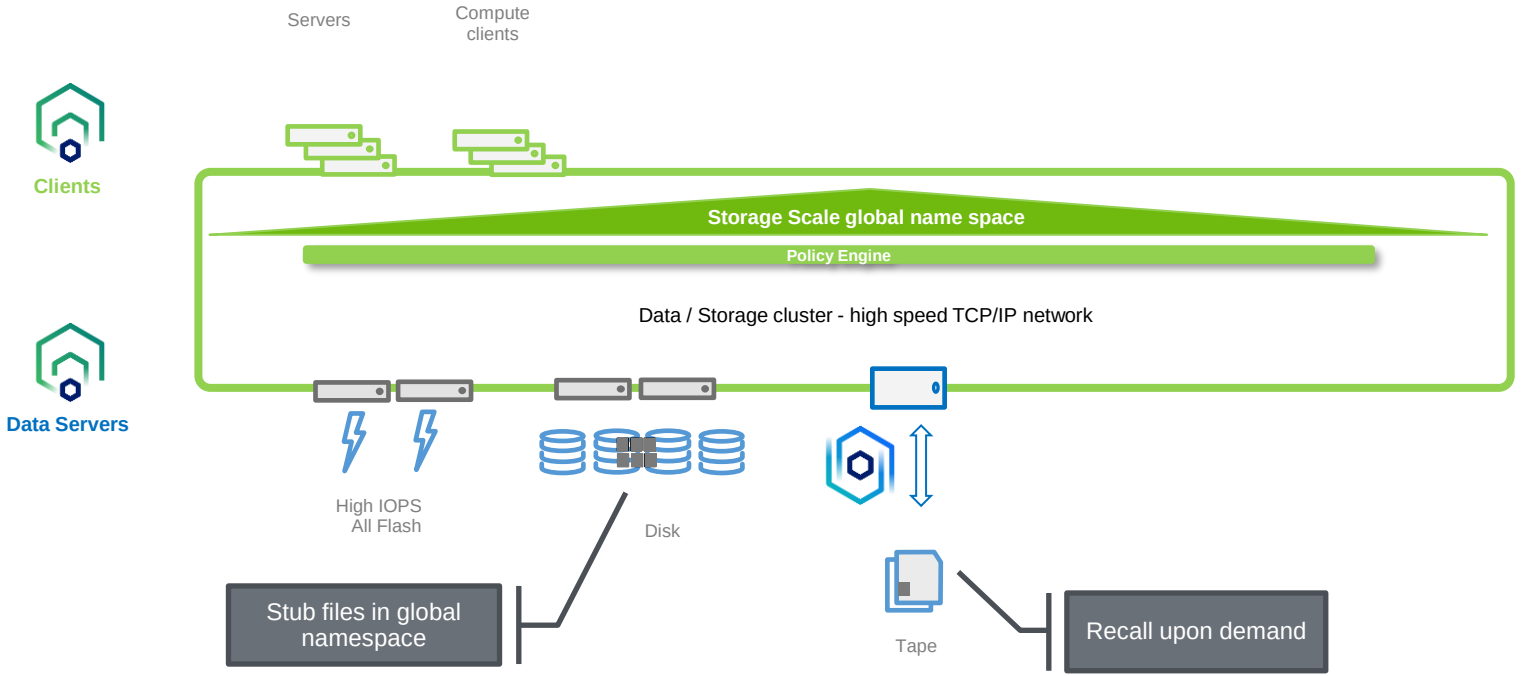
IBM Storage Scale mmbackup command line options

IBM Storage Protect backup archive client processing options

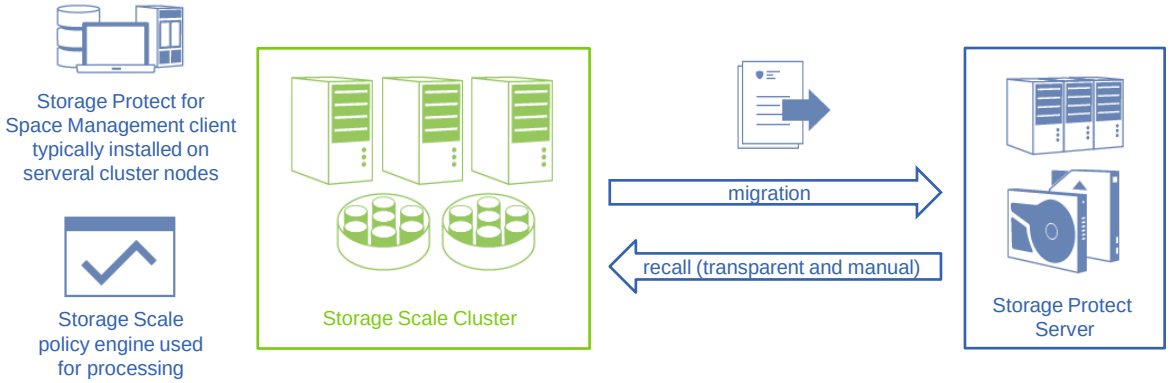
IBM Storage Protect Server processing options

Storage Scale with Storage Protect for Space Management

Utilize Storage Hierarchical Storage Management (HSM) with Storage Scale and Protect Space Management

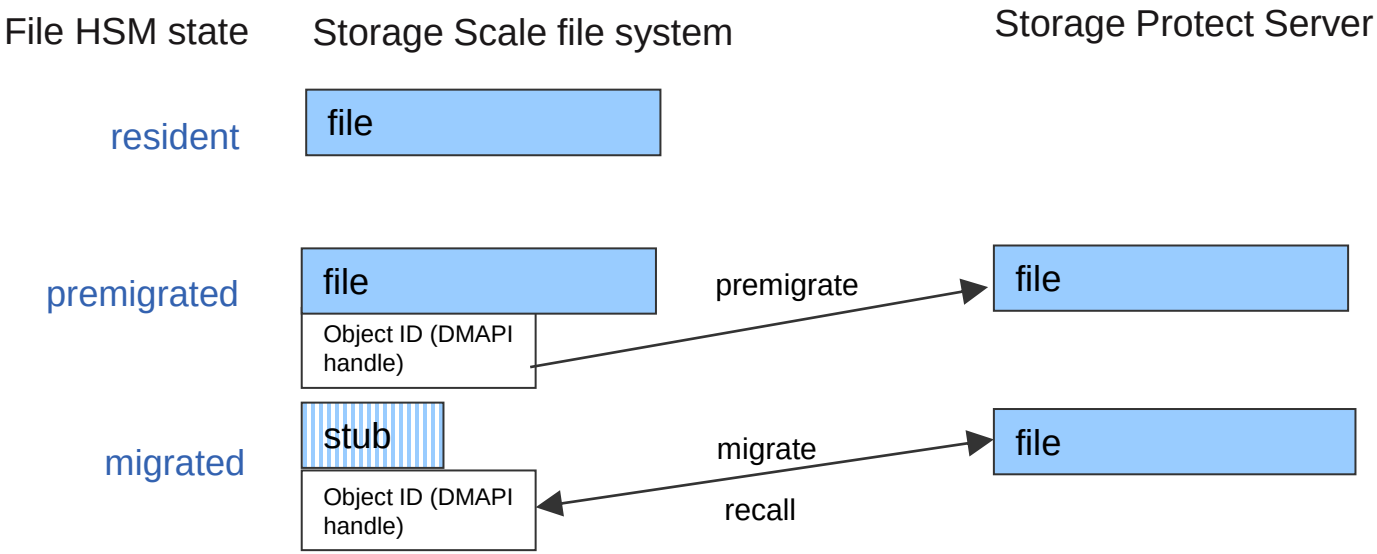


Hierarchical Storage Management



- Massive parallel migration processing to efficiently reduce filesystem occupation
- Cluster wide distributed parallel recall of files in file specific mode
- Tape optimized recall for mass recalls in batch mode
- Close integration with Storage Protect Backup Archive Client
- Recovery of migrated files in stub format

HSM file states



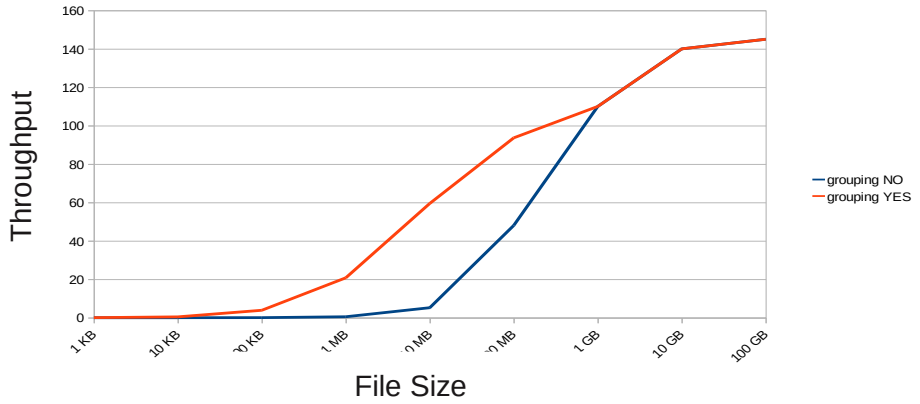
Functional Overview - Migration

- Supports pre/migration to address different use cases
- Stub size can be defined
- Fully integrated with Storage Scale cluster and file system capabilities
  - Node wide and cluster wide parallelism
- Supports multiple Storage Protect servers for a single file system
- Close integration with Storage Protect Backup Archive Client
  - Migration requires backup
  - Inline copy during backup if file is migrated up already
- The migration of files in tape pools of the SP server is optimized for performance
- Set the quota for migrated file capacity !!
  - Initially set to file system capacity, but you might want to migrate more

```
# dsmmigfs q –d fspath
```

Tape Optimized Migration

- Implemented to improve the migration performance directly to SP server tape pools
  - Performance is improved by reducing the number of tape commits required
  - Accelerates migration especially for small files
- Option HSMGROUPEDMIGRATE YES must be set to enable the optimization
  - HSM migration process will group several files in one TSM server transaction
- Multiple file transactions limited by
  - option TXNBYTELIMIT and TXNGROUPMAX
  - alternation of management class in file list
  - alternation of file system in file list
- No influence to recall performance



Functional Overview - Recall

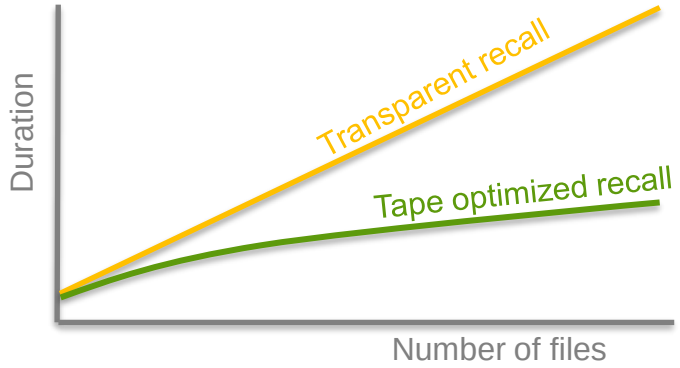
- Supports three different recall modes:
  - Normal (Application can read the file after recall has finished)
  - Streaming (Application can read the file after a administrator defined portion of the file was recalled)
  - Partial (Application can read the sub portion of the file that was requested after it was recalled)
- Recall options can be defined in detail with a common interface
- Recalls are distributed to different cluster nodes in Storage Scale clusters automatically
- Close integration with Storage Scale mmbackup
  - Migrated files are skipped and their names are stored in a list for subsequent processing
- A special manual recall method is available for tape optimized recalls

Tape Optimized Recall

- Tape optimized recall can be used to recall a huge amount of files on demand
- The procssing requires a list of files that has to be recalled

```
# dsmrecall [-preview] –filelist=list fspath
```

- Processing is divided into two steps:
  - Preparation (query the Storage Protect server for media and ordering information)
  - Processing (perform parallel recall of files)
- Option HSMMAXRECALLTAPEDRIVES (1-10, 5) defines the number of parallel recall threads.
  - Should be equal or less then MAXNUMMP.
- Tape optimized recall has higher priority than transparent recall
- Tape optimized recall is not distributed to other cluster nodes automatically



Scale Out Backup and Restore (SOBAR)

Level of file and file system recovery using IBM Storage Protect

Restore Type	Restored Content	Comments
Full Restore	• directories • file data • file metadata (POSIX/ACL/EA)	• full data transfer over network
Stub Restore	• directories • file metadata (POSIX)	• quick • next backup requires recall of files • reset of ACL/EA required
SOBAR Restore	• directories • file metadata (POSIX/ACL/EA)	• fast • scalable

Level of file and file system recovery using IBM Storage Protect

Small number of files (Full Restore):

- dsmc restore –restormigstate=no

Huge number of files or full directory trees (Stub Restore):

- dsmc restore –restoremigstate=yes

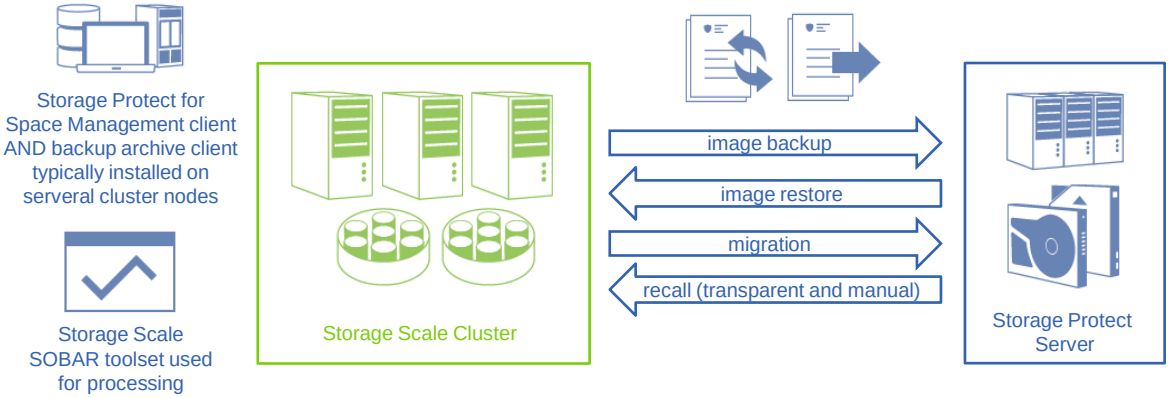
or

- dsmc restore –dirsonly
- dsmmigundelete (does not bring back the ACL)

Full filesystem recovery (SOBAR Restore):

- SOBAR

Fast Disaster Recovery



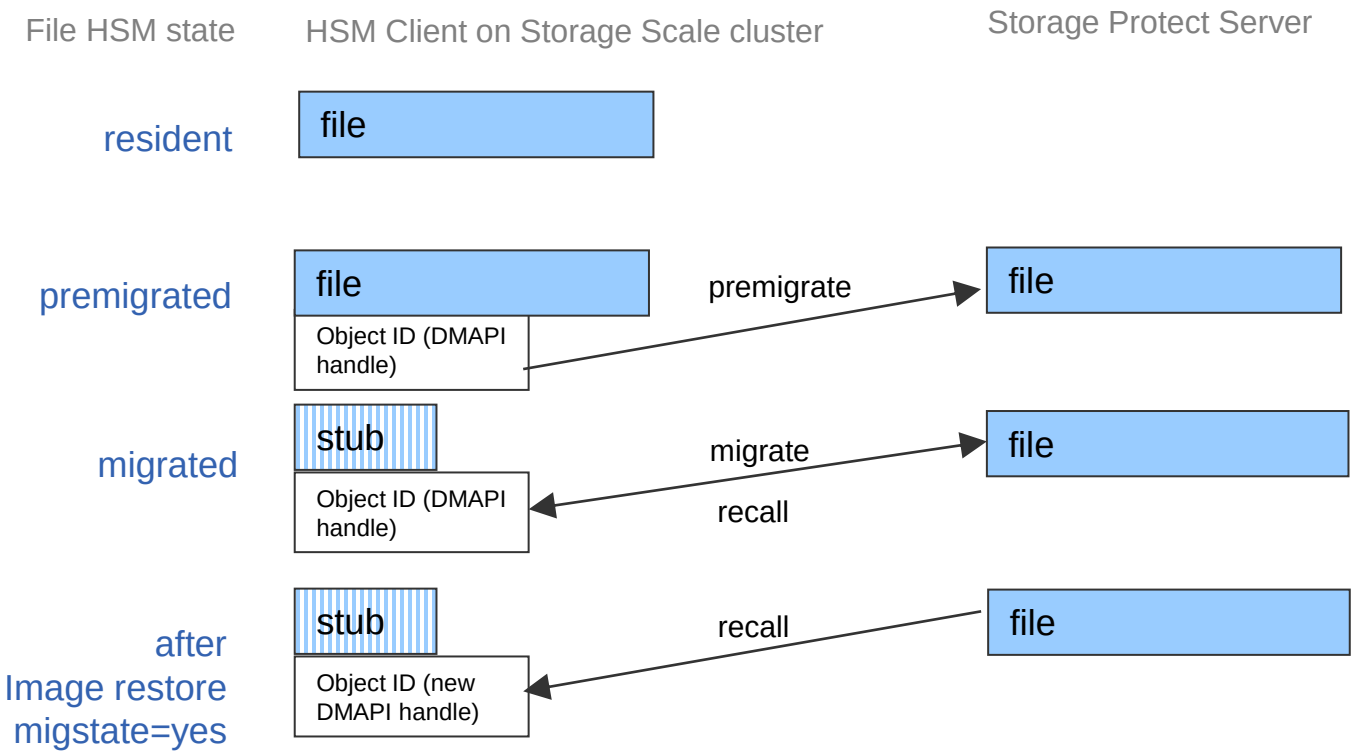
Function Backup

- Storage Protect HSM used to premigrate files
- SOBAR toolset used to generate filesystem metadata image
- Storage Protect backup archive client used to backup image files

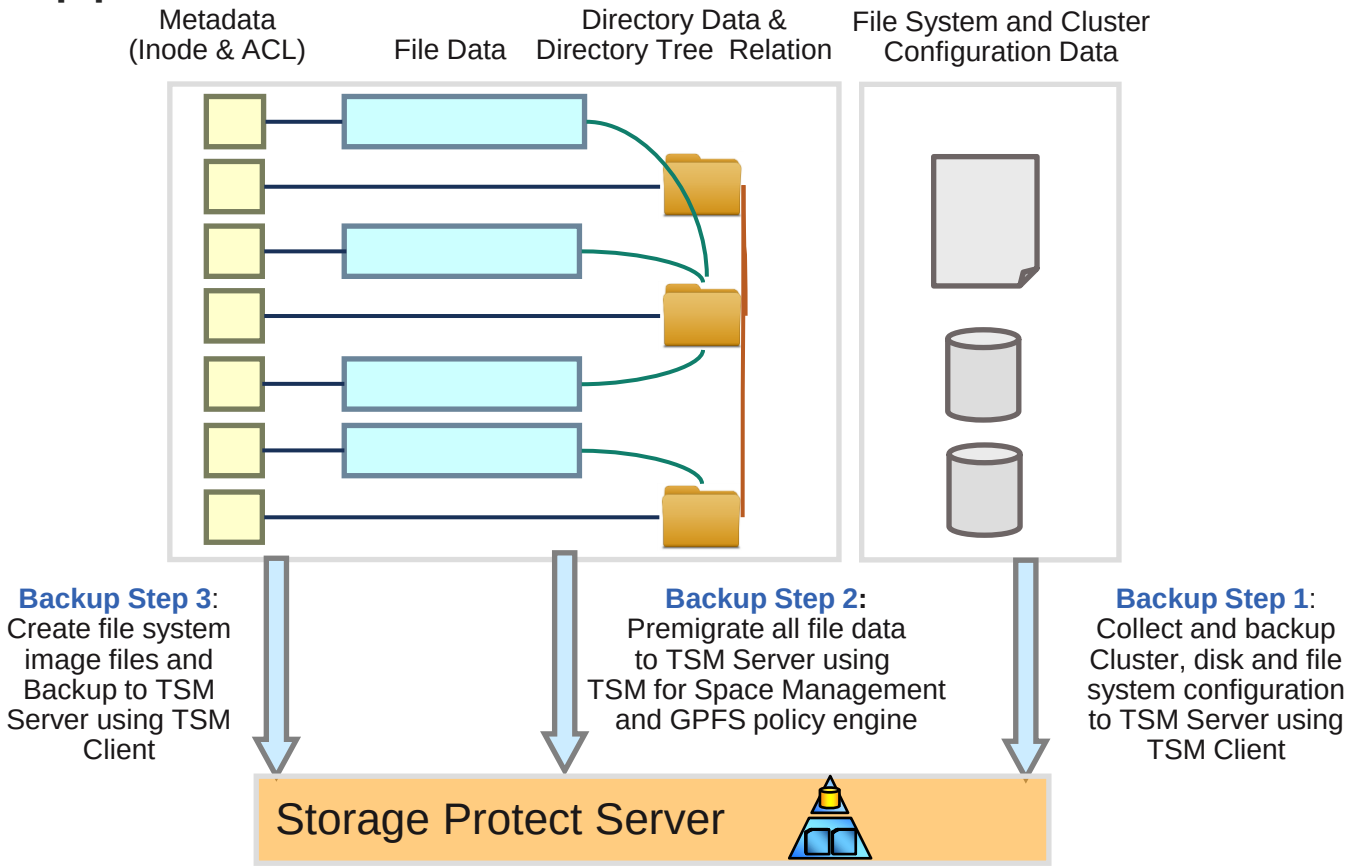
Function Restore

- Storage Protect backup archive client used to restore image files
- SOBAR toolset used to recreate file system structure
- Storage Protect HSM used to pre-fetch files and allow direct access by applying transparent recall

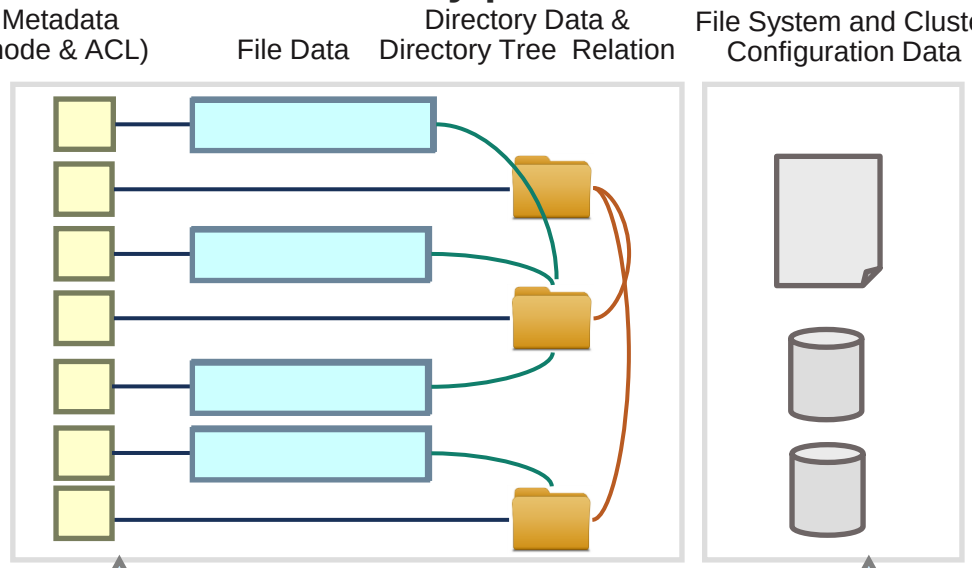
HSM file states



SOBAR Backup process overview



Scale Out Backup And Restore – Recovery process



Storage Scale SOBAR Characteristics

- Recovery of the Storage Scale file system includes all directories and files in stub format
  - Including filesets

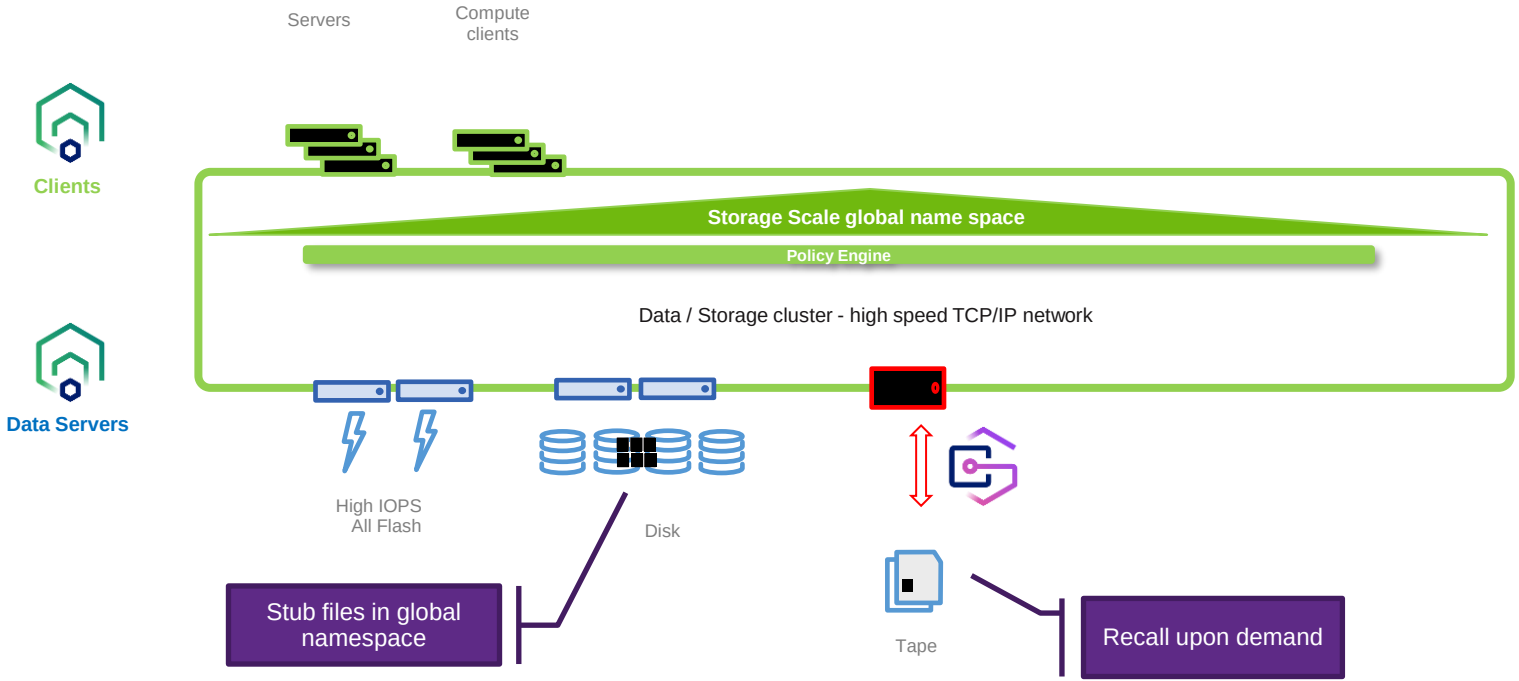
- High scalable “backup”:
  - Leverages Storage Scale policy engine for fast file identification
  - Files are “backed“ up incrementally forever through pre-migration or migration
- High restore performance
  - Only file metadata is applied without transferring file data
  - File data resides on the Storage Protect Server and recall happens on demand

- Subsequent backup (mmbackup) will not backup any files

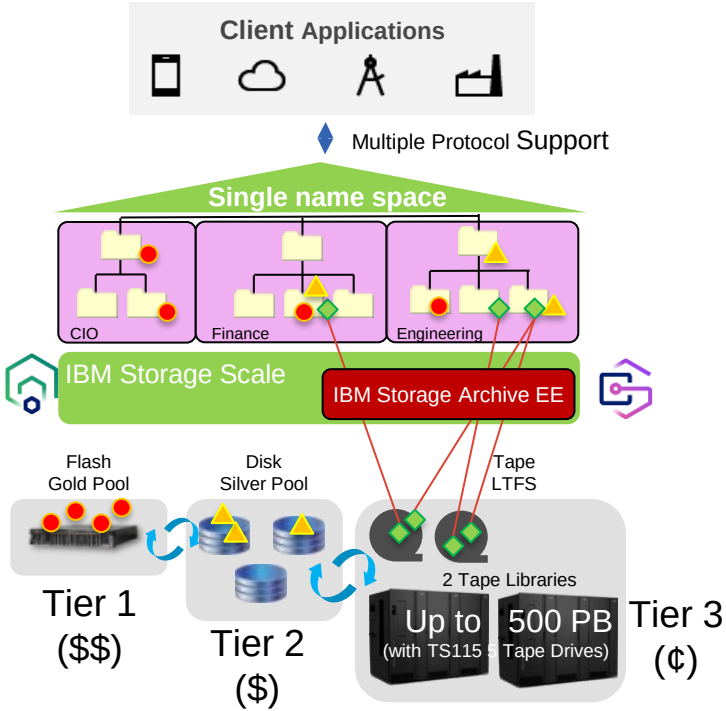
- Can be used to migrate data from one cluster to another
  - Requires the file system being migrated to be space managed

Storage Archive

Archive with Storage Archive



IBM Storage Archive Enterprise Edition (EE)



Persistent view of the data - tape storage under the single namespace

- Policy-based data placement for cold/idle data
- Recall data from tape on demand

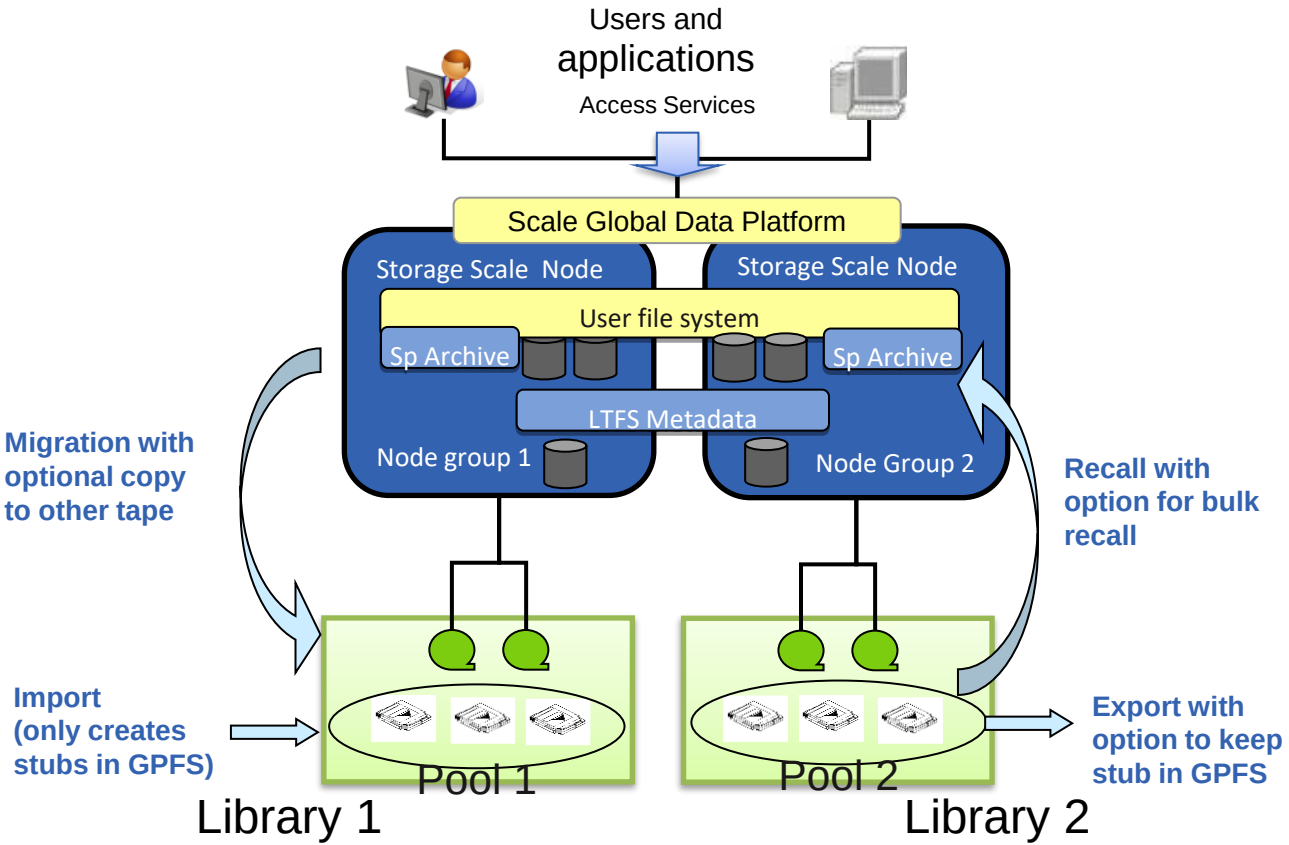
Integrated Tape Tier

- » Up to 3 data replicas
- » Data Encryption with IBM GKLM server (LME)
- » WORM tape for anti-tampering
- » Offline tapes to store the media in an isolated environment – “air gap” for greater protection of sensitive corporate data, or extend the storage capacity beyond the library limit
- » Automated Tape Validation available with IBM TS4500

Export the LTFS tapes for data exchange

- » Remove data from Scale namespace, and export tapes for the use in other application

Storage Archive functional overview

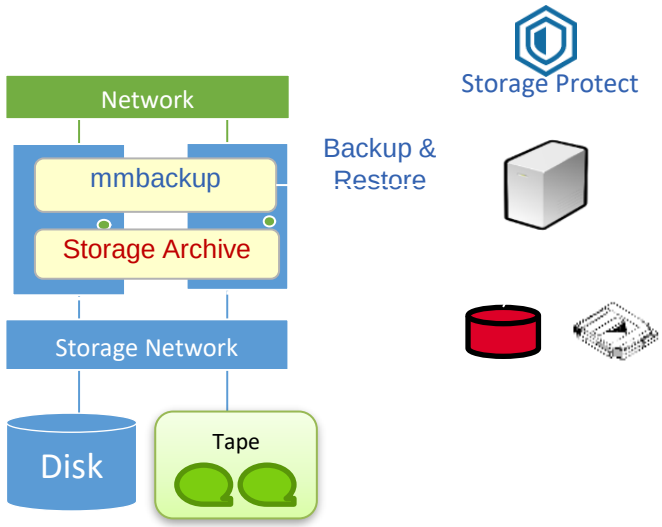


Integration with Storage Protect backup

- Backup of Storage Scale file system can be done using mmbackup
  - Integrates with IBM Storage Protect B/A client

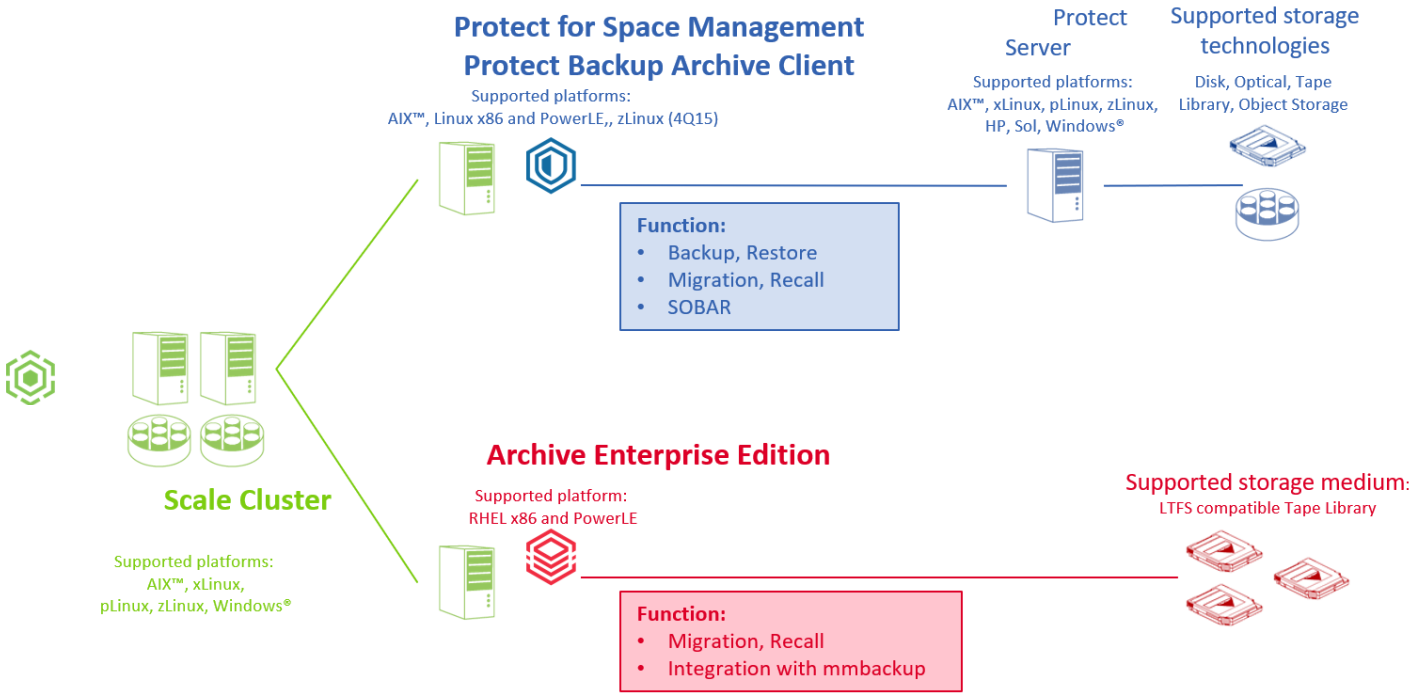
- Storage Archive EE can be configured to prevent migration of files that have not been backed up [Link]
  - Uses the mmbackup shadow DB to determine if a file has been backed up
  - Files that are not backed up are not migrated
  - Use this option `-mmbackup /path-to-shadowDB` with the command `eeadm migrate`
  - Or with EXTERNAL POOL rule:

```
 RULE EXTERNAL POOL 'ltfs' EXEC
'/opt/ibm/ltfsee/bin/eeadm'
 OPTS '-p pool1@lib1 -mmbackup /path-to-shadowDB'
```



- If migrated file is deleted, then it can be restored from the backup in Storage Protect
  - Using the B/A client (dsmc restore)

Comparison with IBM Storage Protect for Space Management



Consideration: Backup with Storage Archive

Objective of backup is to be able to restore files whatsoever

Storage Archive does not provide restore capabilities

After pre-migration file is dual resident: in GPFS and on LTFS tape

When file is deleted in GPFS it cannot easily be found on tape

ACLs are not stored on LTFS tape

Storage Archive integrates with mmbackup to Storage Protect

EE version 1.3 allows to exclude files from migration that have not been backed up

Storage Scale mmbackup allows to skip migrated files

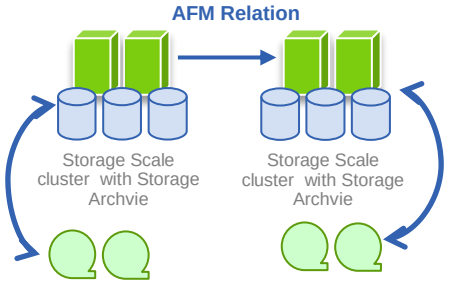
Bottom Line: Storage Archive is made for archiving, not for backup

When backup is required consider mmbackup with Storage Protect

Considerations: AFM with Storage Archive

➢ Storage Archive is supported with AFM IW mode (link)!!

- Files changed on the primary are asynchronously copied
  - Intermediate migration of dirty files will generate errors
- Recall storms on the primary site when many migrated files are accessed



- Recall storms for migrated files on secondary site
  - When many files are replicated from primary
  - When many files are accessed after failover
  - When many files are replicated to old primary during failback
- Peer snapshots cause recalls when migrated files are deleted
  - Deletion of files on primary can cause recall storms on both sites
  - Failover and failback will restore peer snapshots and may cause recalls

Safe Guarded Copy with Tape

Air Gap with SafeGuarded Tape

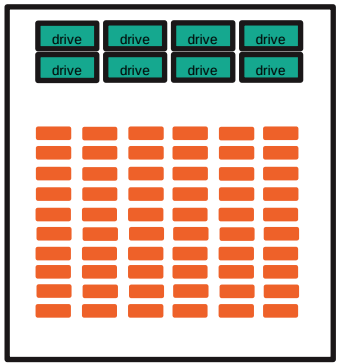
Orchestrated process creating an air-gap, without physical media handling.

1. Admin creates disconnected drive in the logical library
2. Admin determines air-gap criteria
3. Air-gap qualified media is logically assigned to Safeguarded logical library
4. All criteria assessment and assignment via CLI
   - Dark handling
   - No physical media movement
   - Robotic Process Automation to meet business need
5. “Gap” protection
   - Air-gap –Application cannot see the safeguarded media
   - Time-gap – Internal actor must move identify and move media in order to attack
   - Separation of Duty
     - ISV (ie IBM Storage Archive) admin cannot see media in the logical library
     - Storage admin cannot see data on tape

Air Gap with SafeGuarded Tape

Physical Library

- 60 Cartridges
- 12 tape drives
- All visible to ISV (ie IBM Storage Archive)



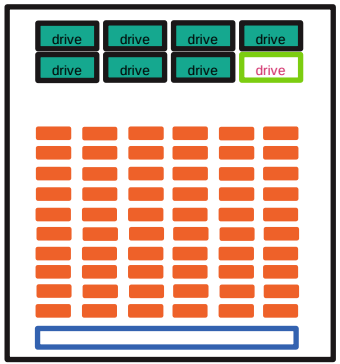
Physical Library with Safeguarded tape

60 Cartridges

11 ISV tape drives

All visible to ISV

Safeguarded LL 1 drive*



Tape drive assigned is not connected to SAN
Library is a logical assignment with no media

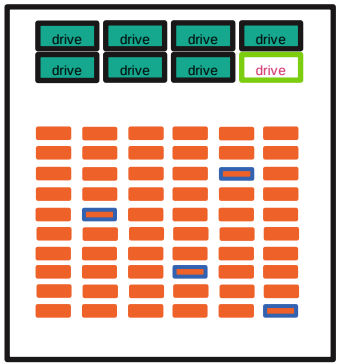
Safeguard Orchestration

ISV issues export from library

Automation/Orchestration

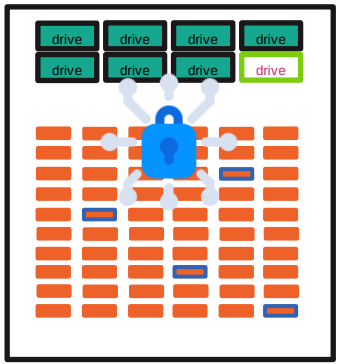
- CLI query export
- CLI import to Safeguarded LL

No physical tape handling



Safeguarded Tape Outcome

- ISV has no visibility to ST media
- StorAdmin has no data access
- Media Verification schedule
- Reduced manual handling risk




IBM