Nils Haustein | Senior Technical Staff Member | EMEA Storage Competence Center

IBM Spectrum Scale | Information Lifecycle Management

© 2019 IBM Corporation

Agenda

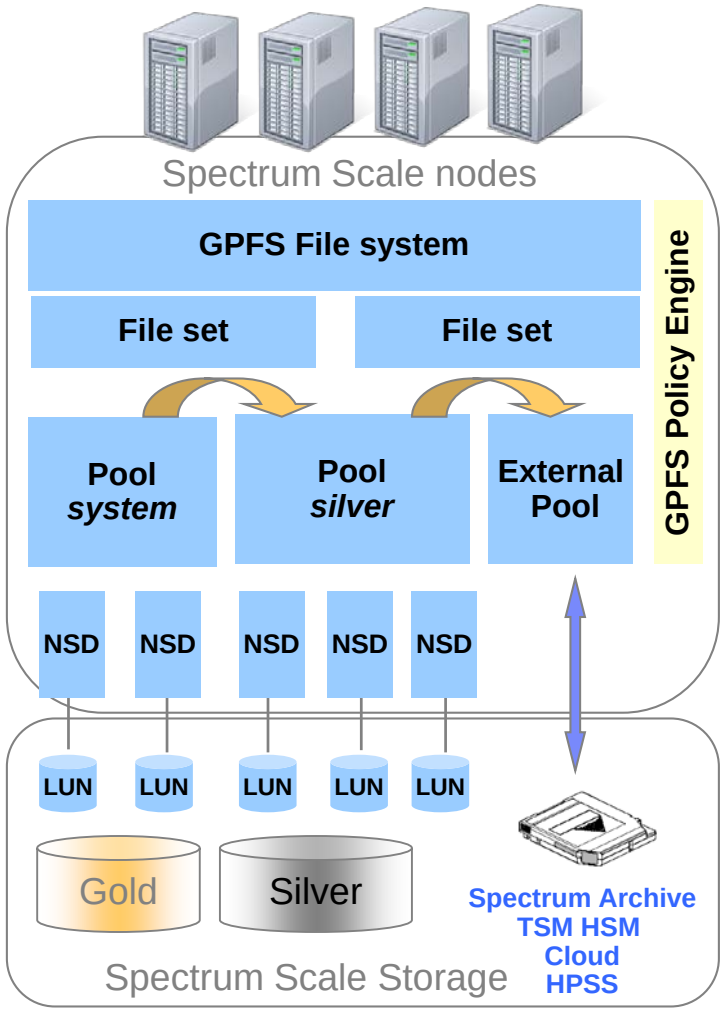
➔**Spectrum Scale ILM**

Hints & Tips

© 2019 IBM Corporation

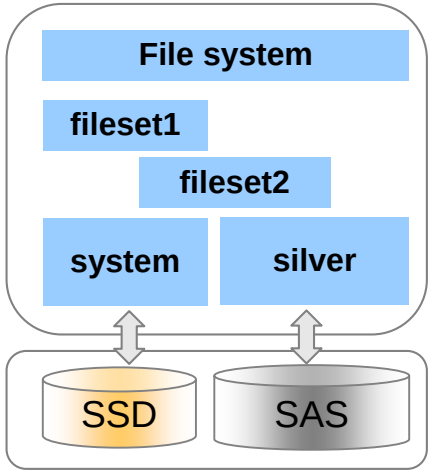
Spectrum Scale ILM tools

- Files stored in file system and underlying pools
  - Storage pool is a collection of disks of the same type
- **Placement policy** controls where files are placed
  - Files are placed in pools
  - Placement policy is applied during file creation
- **Migration policies** control transparent migration of files from one pool to another
  - Applied during life cycle
  - Migrated files can be transparently accessed
  - Files can also be migrated to tape
    - Transparent access through file system
- **Policy engine** applies and executes policies



Spectrum Scale User Group Internal use
© 2019 IBM Corporation

File system and filesets



- User stores data in file system (global name space)
- **Fileset** is a partition of the file system
  - For the user it is a directory
  - For the admin it allows more granular operations such as quota, snapshots*, AFM caching*, immutability, backup and tiering
- Two kinds of filesets: independent and dependent
  - Independent file sets have own inode space and allow snapshot & AFM
- File system is comprised of storage **pools**
  - There is at least one storage pool per file system (system pool)
  - File system capacity is represented by all pools
  - Pools are used for placement and migration
  - Data in a file set can be aligned to a pool using placement policies

* Independent fileset only

Spectrum Scale User Group Internal use
© 2019 IBM Corporation

File system storage pools

- **Internal pool** – allows placement and migration
  - A collection of disks or arrays with similar properties that are managed together
  - Every file system has at least a “system” pool
    - GPFS file system metadata is only stored in “system” pool
    - Maximum of 8 storage pools are possible per file system
- **External pool** – allows migration
  - External pool is an interface (script) to an external application
  - Scripts supported for Spectrum Archive, TSM HSM, Cloud Object Storage or HPSS
  - External application is invoked by policy engine or by command
- Pools are designated by names
  - Pool name *system* is reserved for the standard pool

© 2019 IBM Corporation

Spectrum Scale policies and rules

- A **policy** is a set of rules that describes the life cycle of files
- Each **rule** defines selection criteria for files and operation to be applied
  - File selection based on file attributes (name, size, age, ...)
  - Selected files are processed according to rule type and definition
- Rule types are:
  - **File placement** OR
  - **File migration (threshold / scheduled)** OR
  - File deletion OR
  - File encryption etc.
- A policy can include a mix of different rules types

Spectrum Scale User Group Internal use
© 2019 IBM Corporation

Active and scheduled policies

- **One active policy** can be configured for a file system
  - Is automatically invoked based on rule type and definitions
  - Active ILM policy for one file system can include:
    - File placement rules
    - Threshold migration rules (space triggers)

- **One or more scheduled policies** can be run for a file system
  - Are invoked manually or via scheduler
  - Scheduled policies include migration rules
    - Based on file attribute (size, age)

- Active migration policies must be threshold based and simple
- Scheduled migration policies do not require threshold and can be more advanced

Spectrum Scale User Group Internal use
© 2019 IBM Corporation

Active migration policy (threshold based)

- Active migration policy triggers automated migration based on capacity threshold

```
RULE 'rule-name' MIGRATE FROM POOL 's-pool'
THRESHOLD(%high,%low) [WEIGHT(expression)] TO POOL 'd-pool'
[FOR FILESET ('fileset-name') WHERE (conditions)]
```

- Threshold statement defines threshold at which this rule is triggered
  - %high: Threshold triggers migration
  - %low: continue migration until this threshold is reached

- Activated for entire file system with command: `mmchpolicy`
  - When threshold is reached migration rules are applied automatically
  - Invokes GPFS callback (must be configured)
  - Callback invokes `mmapplypolicy` for this policy

Spectrum Scale User Group Internal use
© 2019 IBM Corporation

Migration Callback

- To invoke the policy engine when threshold is met a callback must be configured
  - For internal and external pool migrations
- Callback is invoked upon GPFS events
  - Relevant migration events are: `lowDiskSpace,noDiskSpace`
- Callback executes a customizable script with selective parameters
  - Parameters are typically: `'%eventName %fsName' (%storagePool)`
  - Default script is: `/usr/lpp/mmfs/bin/mmstartpolicy` - invokes `mmapplypolicy`
- Example for callback setup:

```
mmaddcallback MIGRATION --command /usr/lpp/mmfs/bin/mmstartpolicy
--event lowDiskSpace,noDiskSpace --parms '%eventName %fsName'
```


Spectrum Scale User Group Internal use
© 2019 IBM Corporation

Manual migration (scheduled)

- Manual migration is started on-demand

```
RULE 'rule-name' MIGRATE FROM POOL 's-pool'
[WEIGHT(expression)] TO POOL 'd-pool'
[FOR FILESET ('fileset-name')] WHERE (conditions)
```

- Conditions describe files to be selected for migration
  - Can be based on file attributes (names, size, age, owner)
  - Use of threshold is optional

- Start the policy engine for the entire file system:

```
mmapplypolicy <fsname> -P policyfile
[–s workdir -g workdir -I test|yes]
```


Spectrum Scale User Group Internal use
© 2019 IBM Corporation

Spectrum Scale policy engine

- Policy engine evaluates and applies policy rules by:
  - Selecting files based on criteria defined in rules
  - Initiating an action defined by the rules (placement, migration)

- Policy engine is invoked according to policy type
  - Active placement policy: when file is created (automatic)
  - Active migration policy: when threshold is reached (automatic)
  - Manual migration policy: when policy engine is invoked

- Policy engine runs in three phases on multiple Spectrum Scale nodes
  1. Metadata scan: reading directory file metadata from inodes
  2. Rule evaluation: matching file metadata against rules (top down) and selecting file
  3. File operation: performing the action (placement, migration, etc.)

Spectrum Scale User Group Internal use
© 2019 IBM Corporation

Starting the policy engine

- Active policy for a file system is configured with command: `mmchpolicy`
  - Starts policy engine automatically for **placement and threshold based migration**

```
mmchpolicy filesystem policyfile [–I yes|test]
```

  - When threshold is reached `mmapplypolicy` is invoked via callback

- Scheduled policy for a file system is invoked with command: `mmapplypolicy`
  - Starts policy engi**ne manually for migration**

```
mmapplypolicy filesystem –P policyfile [–I yes|test|defer|prepare]
```


Spectrum Scale User Group Internal use
© 2019 IBM Corporation

Migration to external pool

- External pool must be defined with an additional external pool rule

```
RULE EXTERNAL POOL 'ext-pool' EXEC 'program' OPTS option

RULE 'rule-name' MIGRATE FROM POOL 'int-pool' TO POOL 'ext-pool'
WHERE (conditions)
```

- Example for migration to Spectrum Archive EE:

```
RULE EXTERNAL POOL 'ltfs' EXEC '/opt/ibm/ltfsee/bin/ltfsee'
OPTS ‘-p Tapepool@Library'
RULE 'autoMigSilver' MIGRATE FROM POOL 'silver' THRESHOLD(90,70)
WEIGHT(KB_ALLOCATED) TO POOL 'ltfs’
```

  – To start the policy engine:

```
# mmapplypolicy <file-system> -P policyfile
-m 3 –B 1000 -N ltfsnodes -n 1 --single-instance [-I test|yes]
```


© 2019 IBM Corporation

Consideration for mmapplypolicy parameters for external pools

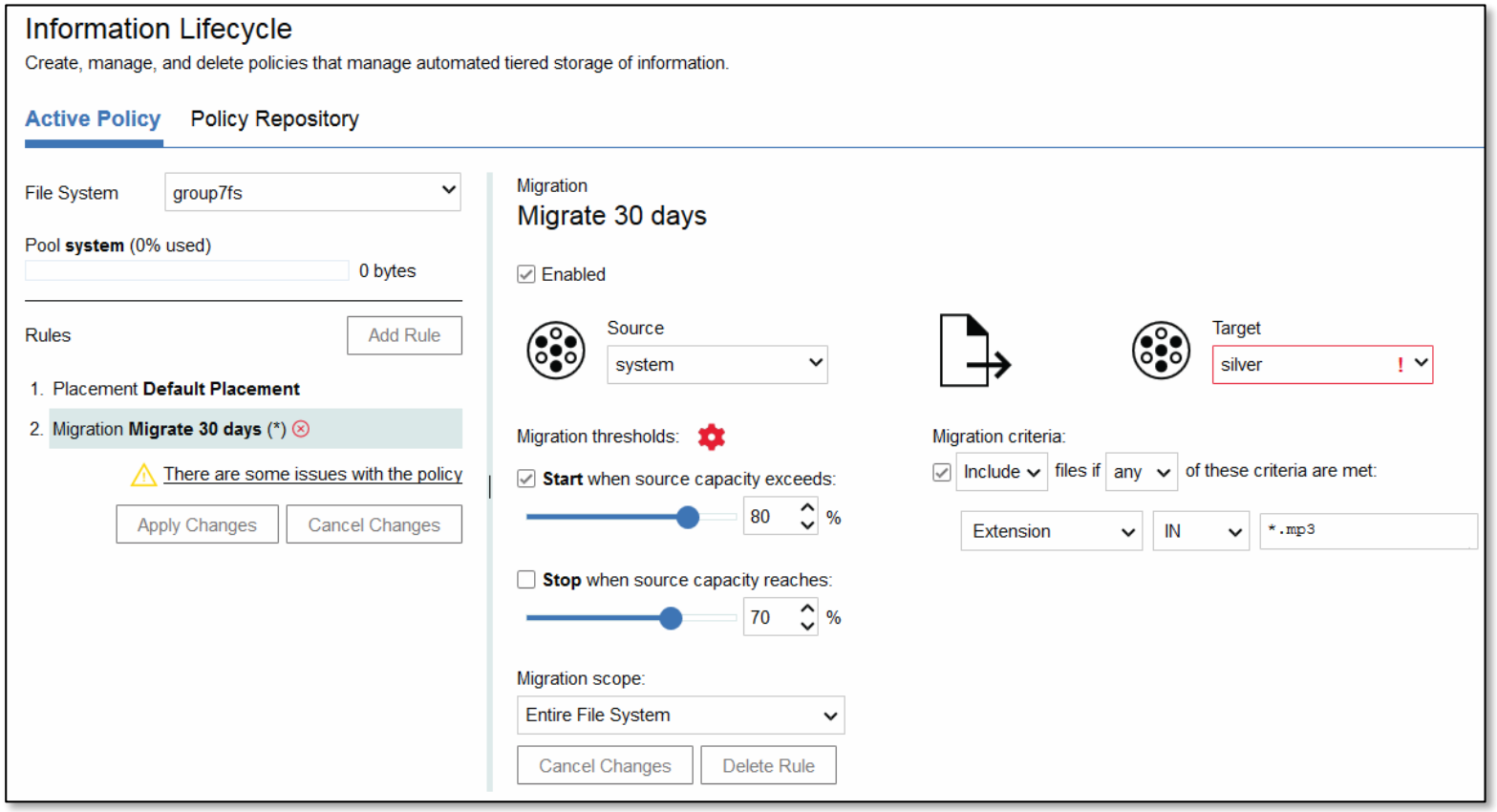
Option	Description	Consideration
-m	Number of threads dispatched for migration, each thread processes a bucket	Depend on number of nodes, bucket size and size of files
-B	Bucket size: number of files per bucket (per file list)	1000 – 10.000, depends on number and size of files
-n	Number of directory scan threads	1
-N	Node names to be involved in the policy execution	Nodes where Spectrum Archive is installed and running
--single-instance	Only one instance of mmapplypolicy can run for this file system	Should be used
-I	Run mode of policy (run, test or defer)	For testing
-s <directory>	Local work directory, default /tmp	Set to shared file system with sufficient capacity
-g <directory>	Global work directory, default sharedTmpDir	
-M	Allows to substitute strings in the policy file	Special use cases
-L <num>	Run in debug mode	For testing, shows the name of the selected files

- Parameters –m and –B can be used to control the max. numbers of files being migrated in parallel (m * B * N)

Spectrum Scale User Group Internal use
© 2019 IBM Corporation

GUI integration of ILM

- GUI intuitively allows to manage placement and migration policies
  - Including syntax and prerequisite check, displays rules



© 2019 IBM Corporation

Agenda

Spectrum Scale ILM concept

➔**Hints & Tips**

© 2019 IBM Corporation

General guidance for Spectrum Scale ILM

- Use scheduled policies to control pool utilization
  - Use Threshold based migration as last line of defense
    - Kicks in unexpected and generates unexpected workloads
    - Triggers every two minutes
- Policies for Threshold based migration should be simple and effective
  - Scheduled policies can be more sophisticated
- Migration between pools causes additional I/O – plan for this
- Use `–single-instance` when migrating to external pools
- Use parameter `–s` and `–g` to set the directories for temporary files
  - Prior to Spectrum Scale 5.0 default location is /tmp
  - With Spectrum Scale version 5.0 default location is /fs/.mmSharedTmpDir
    - Can be set with global configuration parameter: `sharedTmpDir`
- Exclude certain directories when migrating to external pools
  - Such as .SpaceMan, .snapshot, mmSharedTmpDir, CES shared root and .ltfsee
  - Consider EXCLUDE rules

© 2019 IBM Corporation

Macros

- Macros allow pre-defining certain conditions to make rules less complex to read

```
define( macro-name, (conditions & expressions) )
```

- Example: Migrate files older than 30 days

```
define( access_age,(DAYS(CURRENT_TIMESTAMP) - DAYS(ACCESS_TIME)) )
.
RULE 'MigToTape' MIGRATE FROM POOL ‘system' TO POOL 'ltfs'
WHERE (access_age > 30)
```

- More examples: define *migrated* state:
  `define( is_migrated,(MISC_ATTRIBUTES LIKE '%V%') )`

Spectrum Scale User Group Internal use
© 2019 IBM Corporation

Exclude rules

- Allow to exclude files and directories from migration (not for placement)

```
RULE ['RuleName'] EXCLUDE [DIRECTORIES_PLUS]
[FOR FILESET ('FilesetName'[,'FilesetName']...)]
[WHERE SqlExpression]
```

- Exclude rules are evaluated during inode scan phase
  - Files are excluded prior to matching the subsequent migration rules
  - Exclude rules are more efficient and faster than exclusion in WHERE clause

- To exclude directories from migration:

```
RULE 'exclude' EXCLUDE DIRECTORIES_PLUS WHERE
(PATH_NAME LIKE '%/.SpaceMan/%' OR
PATH_NAME LIKE  '%/.snapshots/%' OR
PATH_NAME LIKE  ‘%/.mmSharedTmpDir/%' OR
PATH_NAME LIKE  '%/.mmbackup/%')
```


Spectrum Scale User Group Internal use
© 2019 IBM Corporation

Weight statement

- Weight statement can be used in action rule to sort the files provided in file lists
  - Default weight is `KB_ALLOCATED` (largest files first)
- For age based migration access age my be used as weight
  - `(DAYS(CURRENT_TIMESTAMP) - DAYS(ACCESS_TIME)`
- Example: to prefer certain directories for migration:

```
define(dir_weight,
  (CASE
      /*=== dir1 has high priority ===*/
    WHEN PATH_NAME like '%/dir1/%'  THEN 4
      /*=== dir2 has medium priority ===*/
    WHEN PATH_NAME like '%/dir2/%'  THEN 3
  END)
)

RULE 'MigToTape' MIGRATE FROM POOL ‘system’ WEIGHT(dir_weight)
TO POOL 'ltfs' WHERE …
```


Spectrum Scale User Group Internal use
© 2019 IBM Corporation

Premigration

- Premigration copies files to the other pool, files are dual resident
  - Used with external pools
- Premigration can be configured with threshold rules

```
RULE 'rule-name' MIGRATE FROM POOL 's-pool'
THRESHOLD(%high,%low,%premig) [WEIGHT(expression)]
TO POOL 'd-pool' [FOR FILESET ('fileset-name') WHERE (conditions)]
```

- Thresholds control premigration and migration
  - %high: Threshold triggers migration rule
  - %low: Threshold stops pre-migration
  - %premig: Threshold triggers premigration. Must be between 0 and the %low value.
- If occupancy is above %low then files will be migrated until %low is met
  - Capacity between %low and %premig is premigrated

Spectrum Scale User Group Internal use
© 2019 IBM Corporation

Premigration example (schedule policies only)

- Always premigrate:

```
RULE ‘premig' MIGRATE FROM POOL ‘system’ THRESHOLD(0,100,0)
TO POOL ‘ltfs’ WHERE KB_ALLOCATED > 0
```

  - %high = 0: premigration rule is always started
  - %low = 100: always premigrate
  - %premig = 0: premigrate everything between 0 – 100%
  - Invoke this rule only as scheduled policy because %high is 0 !!!

- Premigrate if pool is below 70 %, otherwise migrate

```
RULE ‘premig' MIGRATE FROM POOL ‘system’ THRESHOLD(0,70,0)
TO POOL ‘ltfs’ WHERE KB_ALLOCATED > 0
```

  - Below 70% will premigrate, above 70% will migrate until 70% is reached and than premigrate all files that are resident
  - Invoke this rule only as scheduled policy because %high is 0.

Spectrum Scale User Group Internal use
© 2019 IBM Corporation

Premigration example with active policy

- Challenge is that active policy is triggered with %high
  - %high should have a non-zero value, otherwise the policy will be constantly triggered

- Premigrate between 10 – 70 %, migrated above 70%

```
RULE ‘premig' MIGRATE FROM POOL ‘system’ THRESHOLD(80,70,10)
TO POOL ‘ltfs’ WHERE KB_ALLOCATED > 0
```

  - This active policy kicks in if pool has reached 80%
  - Note, the active policy will kick in un-controlled.

© 2019 IBM Corporation

List policies

- List policies allow identifying files based on file attributes using the policy engine

```
RULE EXTERNAL LIST 'list-name' EXEC 'program' OPTS option
RULE 'rule-name' LIST 'list-name' WHERE (conditions)
```

- Example of policy to list migrated files:

```
RULE EXTERNAL LIST 'migrated' EXEC ''
RULE 'm_files' LIST 'migrated' WHERE (MISC_ATTRIBUTES LIKE '%V%')
```

  – Use mmapplypolicy to run list policy and create file list:

```
# mmapplypolicy /filesystem -P policyfile -f ./fileprefix -I defer
```

  – Output file is stored in file **./fileprefix.list.migrated**

  – Format of the output (inodenumber, inodegeneration, snapid – path and filename)

```
27648 1566354680 0  -- /mnt/user1/test/file.doc
```


Spectrum Scale User Group Internal use
© 2019 IBM Corporation

List policy example

- Create a list of non-resident files including the tape ID (Spectrum Archive)

```
/* macro defining resident state */
define( is_resident,(MISC_ATTRIBUTES NOT LIKE '%M%') )

/* External list rule showing the tape ID */
RULE EXTERNAL LIST 'tape-id' EXEC ''
RULE 'tape-id' LIST 'tape-id' SHOW(xattr('dmapi.IBMTPS'))
where NOT (is_resident)
```

  - Run the policy

```
#  mmapplypolicy fsname -P policyfile -f /tmp/mia -I defer
```

  - Output file /tmp/mia.list.tape-id

```
27648 1566354680 0  1 SLE033L6@33490ef5-be30-4f23-b328-
527b5e3109a4@0000013100730409 -- /mnt/user1/test/file.doc
```


Spectrum Scale User Group Internal use
© 2019 IBM Corporation

List policy with external scripts

- List policies can invoke an external script for processing the identified files
  - Policy engine identifies files according to certain criteria
  - Policy engine passes lists of files to the external script
  - External script processes the file according to the needs

- Advantages:
  - Leverage speed of policy engine for file identification
  - Leverage parallelism of policy engine for processing the file lists

- EXTERNAL LIST rule allows to specify the name of an external script

```
RULE EXTERNAL LIST 'list-name' EXEC ‘external script' OPTS option
RULE 'rule-name' LIST 'list-name' WHERE (conditions)
```


Spectrum Scale User Group Internal use
© 2019 IBM Corporation

List Policy - External script

- External script is invoked by the policy engine with the following parameters:
  1. Parameter: string describing the operation (TEST and LIST)
     - TEST is invoked first and allows the script to perform some checking
     - LIST will pass the file lists in the second parameter
  2. Parameter: depends on 1. Parameter.
     - If the first parameter is LIST then the second parameter is the name of the file list
     - If the first parameter is TEST then the second parameter is the name of the file system.
  3. Parameter (optional): OPTS encoded in the EXTERNAL LIST rule

- Each file lists includes (-B bucket size) entries in the following format

```
27648 1566354680 0  -- /mnt/user1/test/file.doc
```

- External script has to process the parameters and perform its tasks:

```
case $1 in
  (TEST)     check if file system in $2 is mounted & exit;;
  (LIST)     convert and process file list;;
esac
```


Spectrum Scale User Group Internal use
© 2019 IBM Corporation

List policy with external script - implementation

- Create a EXTERNAL LIST policy that identifies files not immutable files

```
RULE EXTERNAL LIST ‘setWorm' EXEC ‘/usr/local/bin/setWorm.sh’
RULE ‘notimm' LIST ‘setWorm' WHERE NOT (MISC_ATTRIBUTES LIKE '%X%’)
```

- Create an interface script that sets files to immutable:
  - Obtains file list ($2) with file names identified by policy engine
  - Extracts file names
  - Set retention time (mmchattr –E) and immutability flag (mmchattr –i yes)

- Run the policy using

```
# mmapplypolicy fsname -P policyfile –N nodes –m 1 –B 1000
```

  - N specifies nodes running the external scripts – create a node class for this
  - m specifies the number of concurrent processes per node
  - B specifies the number of file names in one file list

© 2019 IBM Corporation

List policy with external script – using OPTS

- EXTERNAL LIST rule allows pass options to the external script:
  - Set different retention times for different file types:

```
define( immutable, MISC_ATTRIBUTES LIKE '%X%')

RULE EXTERNAL LIST 'mp3' EXEC ‘/…/setWorm.sh' OPTS ‘365'
RULE 'mp3' LIST 'mp3' WHERE NOT (immutable) and (NAME LIKE '%.mp3')

RULE EXTERNAL LIST 'pdf' EXEC ‘/…/setWorm.sh' OPTS ‘730'
RULE 'pdf' LIST 'pdf' WHERE NOT (immutable) and (NAME LIKE '%.pdf’)
```

- External script obtains the OPTS parameter together with the file list

```
filelist=$2; retime=$3
for each line in $filelist
  extract filename
  set WORM (mmchattr –E $(current + $retime) –i yes filename)
done
```


© 2019 IBM Corporation

Extracting the file names from the file lists

- Format of the file lists (inodenumber, inodegeneration, snapid – path and filename)

```
27648 1566354680 0  -- /mnt/user1/test/file_9.doc
```

  - Extracting path and file names can be challenging if they include spaces

- Use shift to isolate the file name from file list ($fileList):

```
cat $fileList | while read line
do
  set $line; shift 4
  fName="$*“
  … process file name …
done
```

- Use awk:

```
awk -F '[ ]' '{ for(i=7; i<=NF; i++) \
printf "%s",$i (i==NF?ORS:OFS) }’ $fileList
```


Spectrum Scale User Group Internal use
© 2019 IBM Corporation

String substitution in policies

- Policy engine (mmapplypolicy) allows to substitute strings in policy rules
  - Policy rule allowing to substitute the fileset name

```
RULE 'MigToTape' MIGRATE FROM POOL ‘system’ TO POOL 'ltfs’
FOR FILESET ('FILESETNAME’) WHERE (access_age > 30)
```

  - Substitute the string (fileset name = test) when running the policy

```
# mmapplypolicy fsname -P policyfile -M "FILESETNAME=test“ …
```

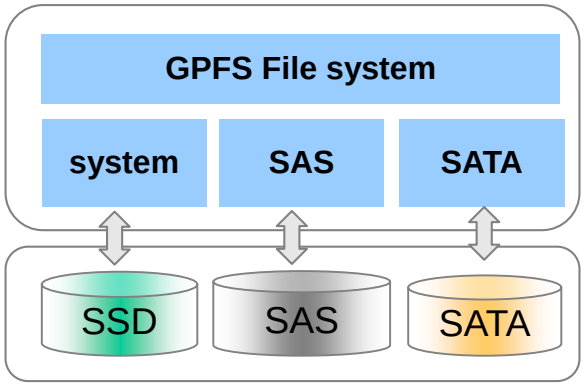
- Substitute fileset name and access age:

```
RULE 'MigToTape' MIGRATE FROM POOL ‘system’ TO POOL 'ltfs’
FOR FILESET ('FILESETNAME’) WHERE (access_age > ‘AGE’)

# mmapplypolicy fsname -P policyfile
  -M "FILESETNAME=test“ -M “AGE=30” …
```


Spectrum Scale User Group Internal use
© 2019 IBM Corporation

Using *Group Pools*



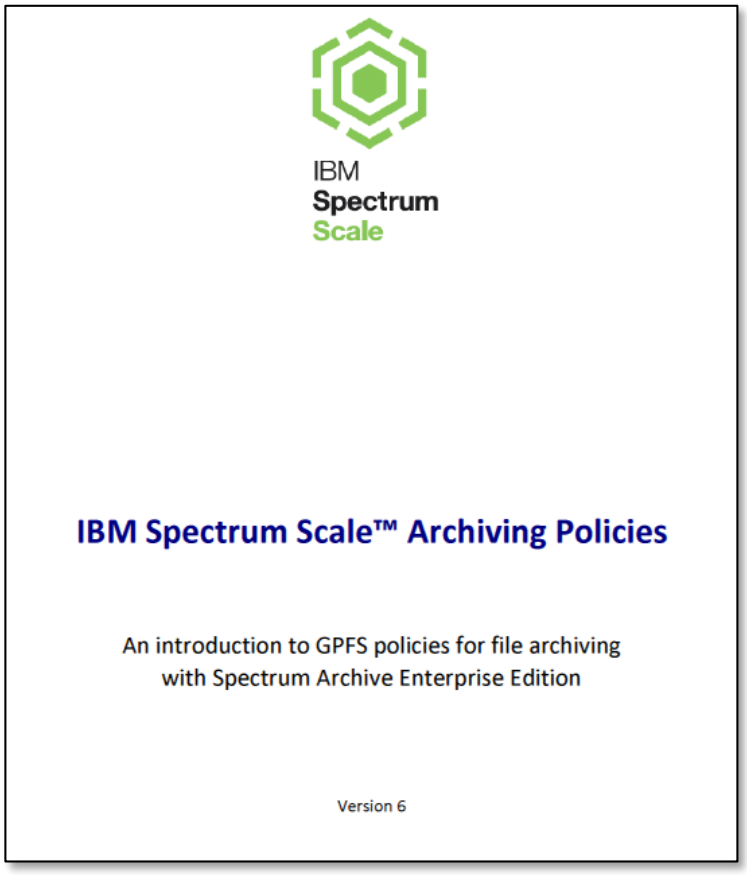
- Group pool combine the content of multiple pools into one virtual pool
  - Allows for redistribute files across the pools in the group
  - Example for redistribution based on file heat:

```
RULE 'HeatTiering' GROUP POOL 'TIERS' IS
     'system' LIMIT(80)
     THEN 'sas' LIMIT(90)
     THEN 'sata'
RULE 'Rebalance' MIGRATE FROM POOL 'TIERS' TO POOL 'TIERS' WEIGHT(FILE_HEAT)
```

  - Puts hottest files in pool “system” up to the utilization limit (80%)
  - Processes the other pools in the order they are listed in the GROUP POOL rule

© 2019 IBM Corporation

Recommended reading



Spectrum Scale ILM Policy Guide

Automation of ILM policies

Spectrum Scale User Group Internal use
© 2019 IBM Corporation

© 2019 IBM Corporation

Links

- GPFS ILM Policy Guide:
  http://www.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/WP102518
- Whitepaper: Introduction to Elastic Storage:
  http://public.dhe.ibm.com/common/ssi/ecm/en/dcw03057usen/DCW03057USEN.PDF
- Whitepaper: Elastic Storage Archiving with IBM LTFS Enterprise Edition
  http://www.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/WP102504
- Whitepaper: Spectrum Protect HSM with Spectrum Scale AFM:
  https://www.ibm.com/developerworks/community/wikis/home?lang=en#!/wiki/Tivoli%20Storage%20Manager/page/Configuring%20IBM%20Spectrum%20Scale%20Active%20File%20Management
- GPFS Home Page
  http://www.ibm.com/systems/gpfs
- GPFS Information Center:
  http://www-01.ibm.com/support/knowledgecenter/SSFKCN/gpfs_welcome.html
- GPFS Wiki
  http://www.ibm.com/developerworks/wikis/display/hpccentral/General+Parallel+File+System+(GPFS)
- GPFS FAQ
  http://publib.boulder.ibm.com/infocenter/clresctr/vxrx/topic/com.ibm.cluster.gpfs.doc/gpfs_faqs/gpfsclustersfaq.pdf

© 2019 IBM Corporation

Disclaimer

- This information is classified for **IBM Spectrum Scale User Group Internal use** and shall not be used or published outside this scope.
- This information is provided on an "AS IS" basis without warranty of any kind, express or implied, including, but not limited to, the implied warranties of merchantability and fitness for a particular purpose. Some jurisdictions do not allow disclaimers of express or implied warranties in certain transactions; therefore, this statement may not apply to you.

Important notes:

- 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.
- IBM makes no warranties, express or implied, regarding non-IBM products and services, and any implied warranties of merchantability and fitness for a particular purpose. IBM makes no representations or warranties with respect to non-IBM products. Warranty, service and support for non-IBM products is provided directly to you by the third party, not IBM.
- All part numbers referenced in this publication are product part numbers and not service part numbers. Other part numbers in addition to those listed in this document may be required to support a specific device or function.
- When referring to storage capacity, GB stands for one billion bytes; accessible capacity may be less. Maximum internal hard disk drive capacities assume the replacement of any standard hard disk drives and the population of all hard disk drive bays with the largest currently supported drives available from IBM.

IBM Information and Trademarks

- The following terms are trademarks or registered trademarks of the IBM Corporation in the United States and / or other countries: IBM, IBM Spectrum Storage, IBM Spectrum Protect, IBM Spectrum Scale, IBM Spectrum Accelerate, IBM Spectrum Virtualize, IBM Spectrum Control, Tivoli, IBM Elastic Storage
- Linux is a registered trademark of Linus Torvalds in the United States, other countries, or both.
- UNIX is a registered trademark of The Open Group in the United States and other countries.
- Other company, product, and service names may be trademarks or service marks of others.

© 2019 IBM Corporation