

IBM Spectrum Scale a Good Place For Stuff

What's new in 4.1.1

Scott Fadden,
IBM Spectrum Scale Development

#IBMSpectrum



Exploring IBM Software Defined Storage Capabilities

IBM Spectrum Storage Family

Storage and Data Control

Storage Management Policy Automation Analytics & Optimization Snapshot & Replication Management Integration & API Services Self Service Storage

Spectrum Control
Virtual Storage Center

Data Protection

Spectrum Protect
Tivoli Storage Manager

Data Access

Traditional Applications

New Generation Applications

Virtualized SAN Block

Spectrum Virtualize
SAN Volume Controller

Hyperscale Block

Spectrum Accelerate
Based on XIV Software

Global File & Object

Spectrum Scale
GPFS/Elastic Storage

Active Data Retention

Spectrum Archive
LTFS

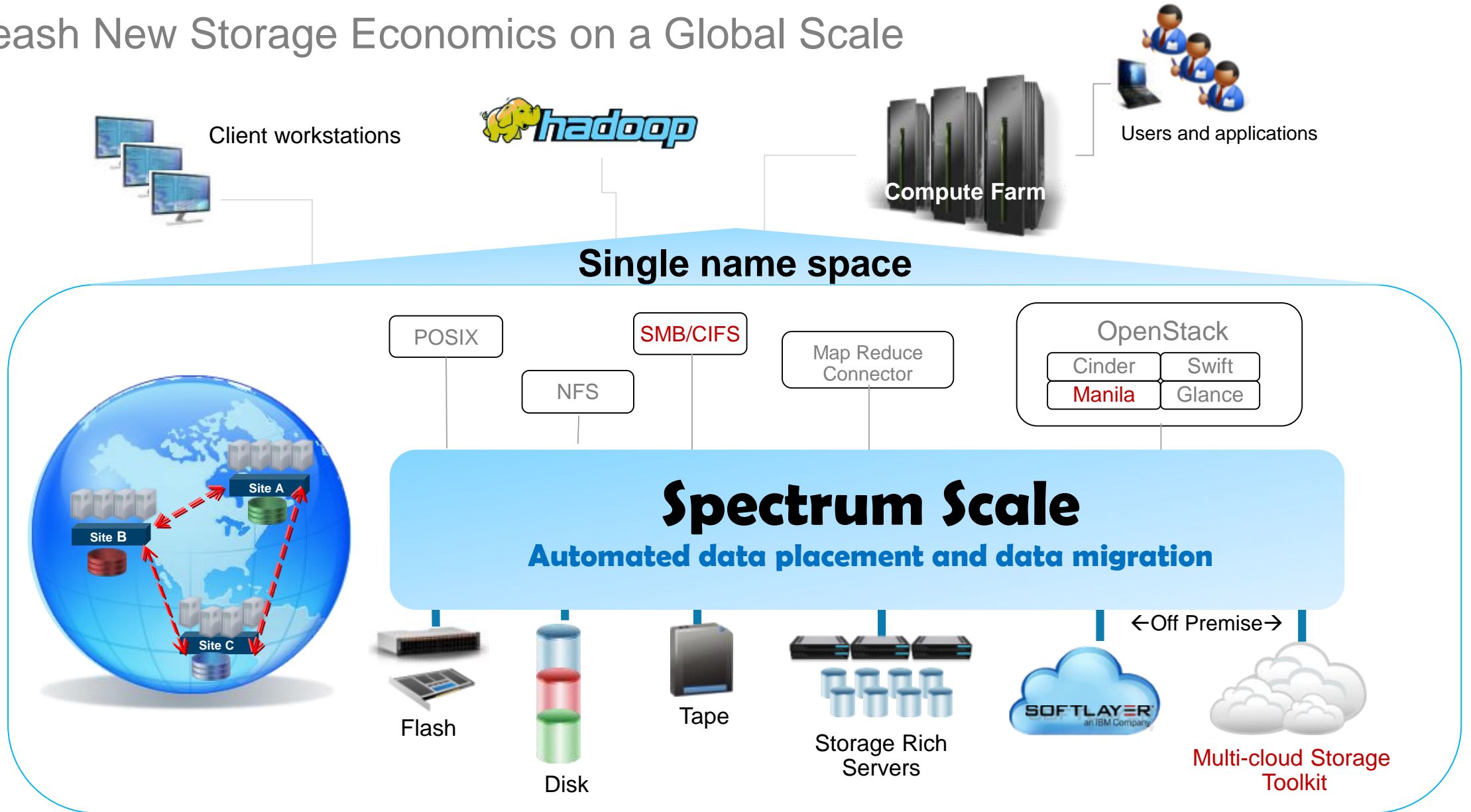


Spectrum Scale simplifies data management at scale.

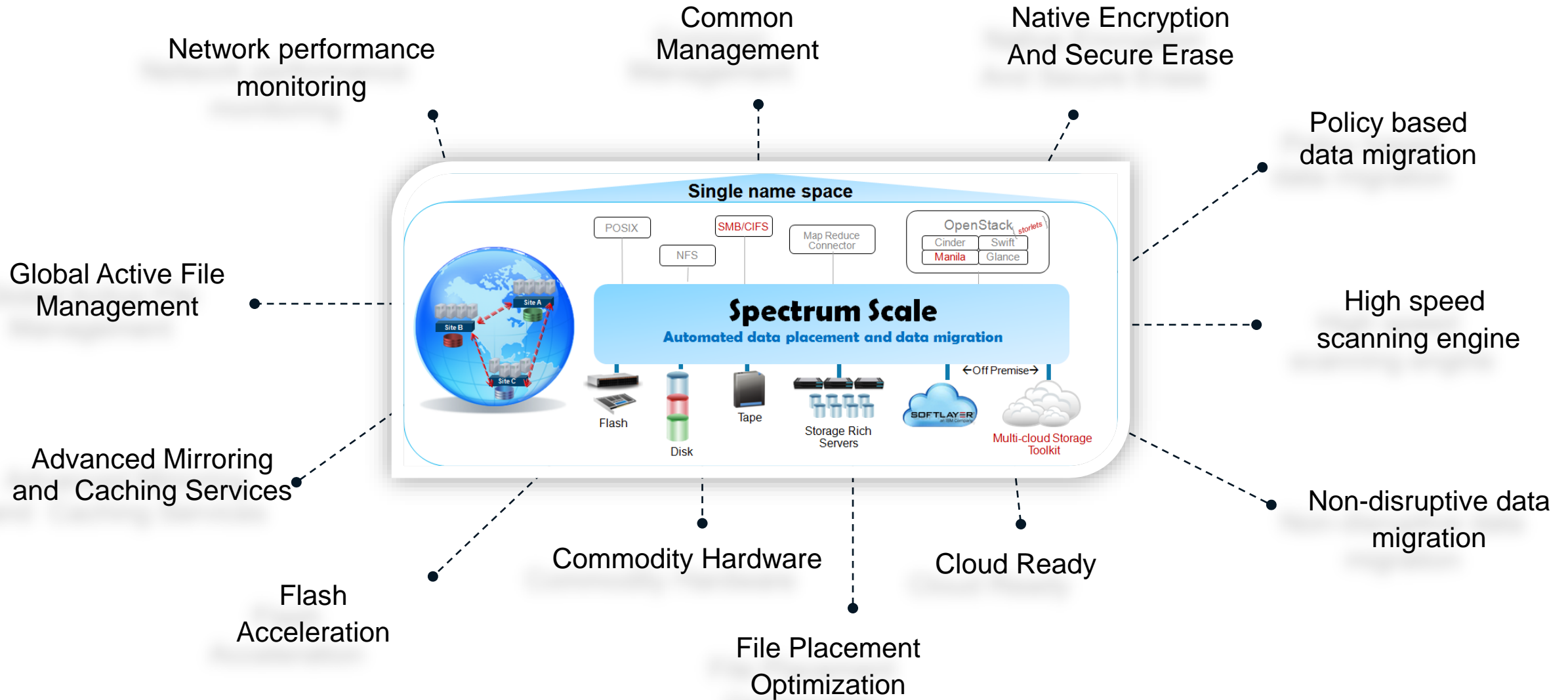
For enterprises that are swamped by unstructured data IBM Spectrum Scale **software** lets you share the storage infrastructure while it *automatically moves file and object data to the optimal storage tier as quickly as possible.*



Unleash New Storage Economics on a Global Scale



Spectrum Scale services



Different ways to view the storage environment

Vertical View (select one)

I/O Acceleration for Tech Computing	<input type="radio"/>
Analytics for FSS	<input type="radio"/>
Global Collaboration Electronic Design Automation (EDA) for Auto	<input type="radio"/>
Video surveillance for Government	<input type="radio"/>
Seismic data ingest and analytics for Petro	<input type="radio"/>
Video Streaming and production workflow for Entertainment	<input type="radio"/>
Synch and Share global collaboration (multiple verticals)	<input type="radio"/>
Medical image archive for Healthcare	<input type="radio"/>
Other ...	<input type="radio"/>

Common Data Management Tasks (select all that apply)

Storage Pools	<input type="radio"/>
Quotas	<input type="radio"/>
Enterprise Archiving	<input type="radio"/>
Backup / Restore	<input type="radio"/>
Disaster Recovery	<input type="radio"/>
Snapshots	<input type="radio"/>
Clone	<input type="radio"/>
Metering and usage	<input type="radio"/>

How do I drop in/out of storage? (select all that apply)

POSIX	<input type="radio"/>
NFS	<input type="radio"/>
SMB/CIFS	<input type="radio"/>
Hadoop	<input type="radio"/>
Open Stack Cinder	<input type="radio"/>
Open Stack Swift	<input type="radio"/>
Https REST interface	<input type="radio"/>
S3	<input type="radio"/>
Other ...	<input type="radio"/>

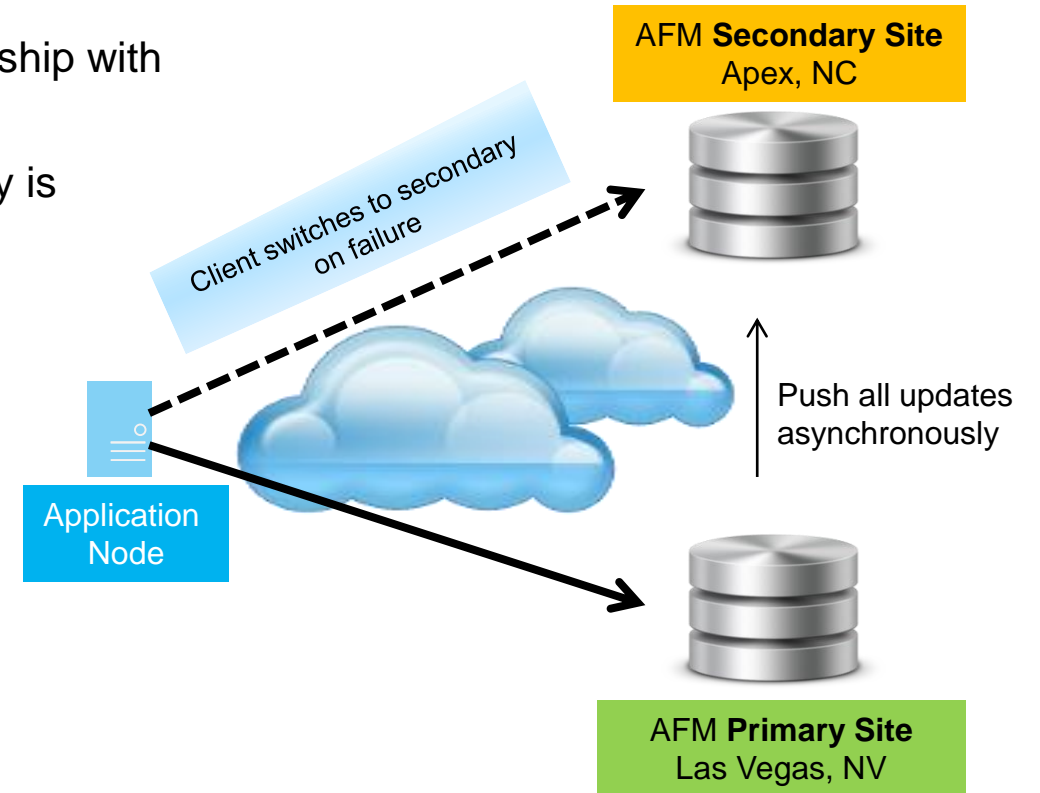
Access
protocols

Spectrum Scale Data Layer

Storage

Asynchronous Disaster Recovery

- Asynchronously Replicate data from primary to secondary site using AFM
 - Failover to secondary when Primary fails
 - Failback when primary comes back
- Allow primary to operate actively with no interruption when the relationship with secondary fails
- Active / Passive model – Primary is Active (Read/Write) and secondary is passive (Read Only)
- Supports Recovery Point Objective (RPO)
- Recovery Time Objective (RTO) determined by network bandwidth, amount of data changed, number of files
- Usability: simpler and fewer commands to do multisite management



Introducing – “The Protocols”

NFS
SMB/CIFS
Object

- Simple installation process
- Tools to manage
- Cluster Export Services (CES)
- New performance tools and metrics
- Protocol Node
- Redhat 7

NFS

Ganesha 2.2

NFSv3 and NFSv4.0

NFS	
# of Protocol Nodes	32
# of Exports	1,000
# of Connections	4,000-5,000 per node
# of Files per File Systems	9 billion
# of Groups	16 group limit needs to be addressed
Cross Protocol Locking	Advisory
Rolling Upgrade	Yes

Object

Openstack Swift (K Release)**Data access using REST protocol**

- PUT (upload)
- POST (update metadata)
- GET (download)
- DELETE

Amazon S3 Protocol Emulation support

Object	
# of Protocol Nodes	16
# of Exports	4M containers (1 M containers per account x 4 accounts)
# of Connections	TBD
# of Files per File Systems	1 billion objects
# of Groups	n/a
Cross Protocol Locking	No
Rolling Upgrade	Yes

SMB

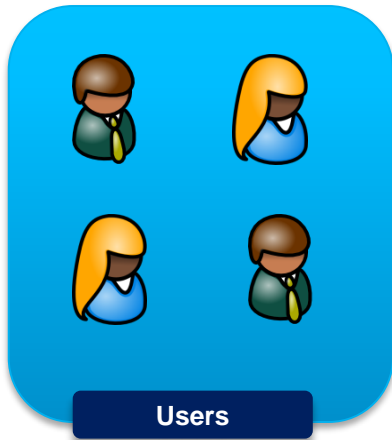
Based on Samba 4.2

- SMB2, SMB2.1
- SMB3 Support includes mandatory features + SMB encryption

SMB	
# of Protocol Nodes	16
# of Exports	1,000
# of Connections	3,000 per node 20,000 per cluster
# of Files per File Systems	9 billion
# of Groups	1,000 per user
Cross Protocol Locking	Mandatory
Rolling Upgrade	No – 2 phase upgrade

Roadmap Details ✨

Client Experience



- Streamline Linux installs further by automating kernel module creation and installation.
- New industry standard methods to protect from system administrator errors, e.g. new partition type
- Real-time monitoring of network performance speeds detection of network issues.
- Monitoring of storage and application performance can mean more consistent application performance.

Immutable Filesets

- Was put in years ago for Integrated Archive (IBM Product)
- New mmchfileset –m option - modes
 - **Regular**
 - **Advisory:** Hardlink is not allowed, Directory can NOT be renamed or deleted unless empty, Immutable files can not be changed to mutable. Retention rule is NOT enforced, Expiration time can be set backwards.
 - **Non-Compliant:** + Retention rule is enforced, expiration time can be set backwards
 - **Compliant:** + expiration time can NOT be set backwards.

Other Stuff

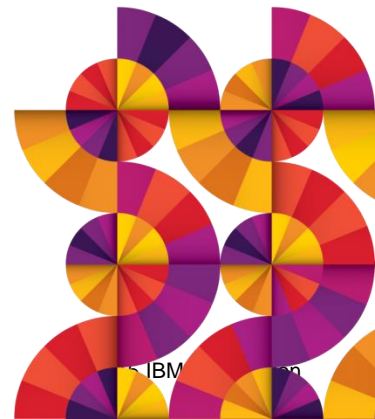
- Fileset level backups with mmbackup
- New preferred read option – read fastest
- Remote command execution (Some beta scripts available – email gpfs@us.ibm.com)
- Speed-up inode expansion
- Allocate token manager memory on the fly
- Maintain disk descriptors, not just on start dynamically fixing bad ones

Other Stuff

- Specify your own sort command << Requested by GPFS UG !!
 - Mmapplypolicy --sort-command option
- Placement Policy now defaults to first data pool
- Faster more informative delete of bad/lost disks
 - Add “empty” option does not scan to drain data
 - Collect information on what files were affected 2 flags
[--inode-criteria CriterialFile] [-o InodeResultFile]

Print interesting inodes BROKEN, illReplicated, unbalanced, dataupdatemiss, metaupdatemiss, exposed, illplaced

PROBLEM DETERMINATION



Blame the network...and use nsdperf to do it

- Works with TCP or RDMA
- Many to many
- Tests many parameters easily
- Does not require GPFS to be installed

Looking into AFM

- Make sure your connection method works NFS or GPFS
- See the state

```
# mmadmctl fs1 getstate
Fileset Name      Fileset Target      Cache State      Gateway Node      Queue Length      Queue numExec
-----
iw_test1          gpfs:///gpfs/homefs1/homefs1  Active           hs22n67           0                 4
iw_testnfs        nfs://hs22n68/gpfs/fs1/nfshome  Active           hs22n67           0                 1
```

- To see active operations

```
mmfsadm dump afm
```

Monitor AFM

- Monitor

- `DSTAT_GPFS_WHAT=vfs,afm dstat -gpfops`

```
-----gpfs-vfs-ops-----#-----gpfs-afm-----
cr  del  op/cl  rd    wr  trunc fsync looku gattr sattr other creat delet  stat  read write sattr rdio  wrio wpend
0   0   0     0    0   0     0   0     0     0     0     0     0     0     0     0     0     0     0     0
```

- Or using `mmpmon`

```
# echo afm_s | mmpmon -s -r 0 -d 2000
mmpmon node 192.168.116.127 name hs22n67 afm_s s OK
Name          Queued    Inflight  Completed  Errors    Filtered  ENOENT
lookup        0         0         4          0         0         0
getattr       0         0         1          0         0         0
readdir       0         0         1          0         0         0
Async Msgs (expire:1 force:0 sync:6 revoke:0)
NumMsgExecuted = 7
```