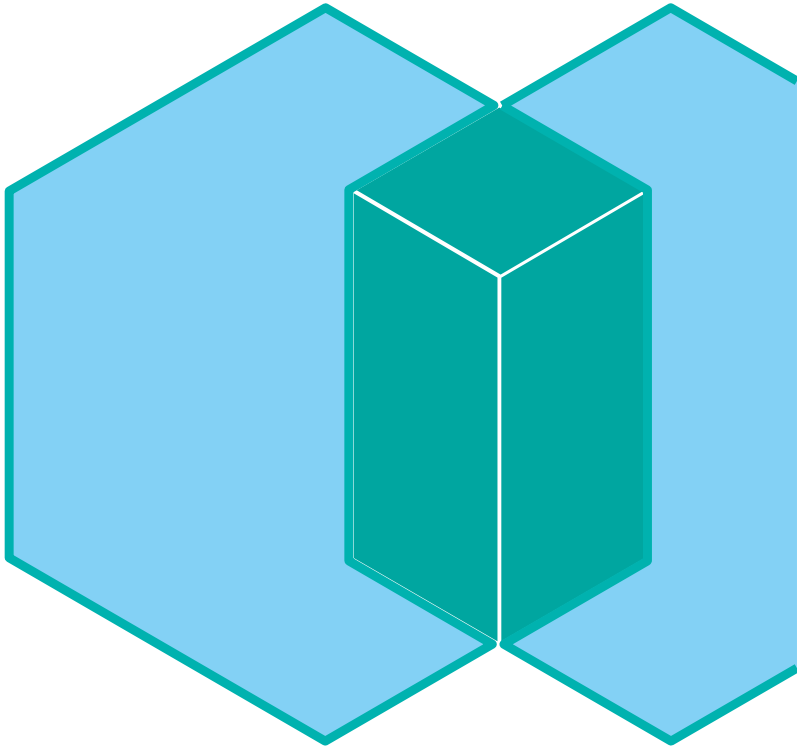


IBM Spectrum Scale

– Recent Updates and Outlook –

Meet the Devs – Oxford – Feb 24, 2016 – Ulf Troppens

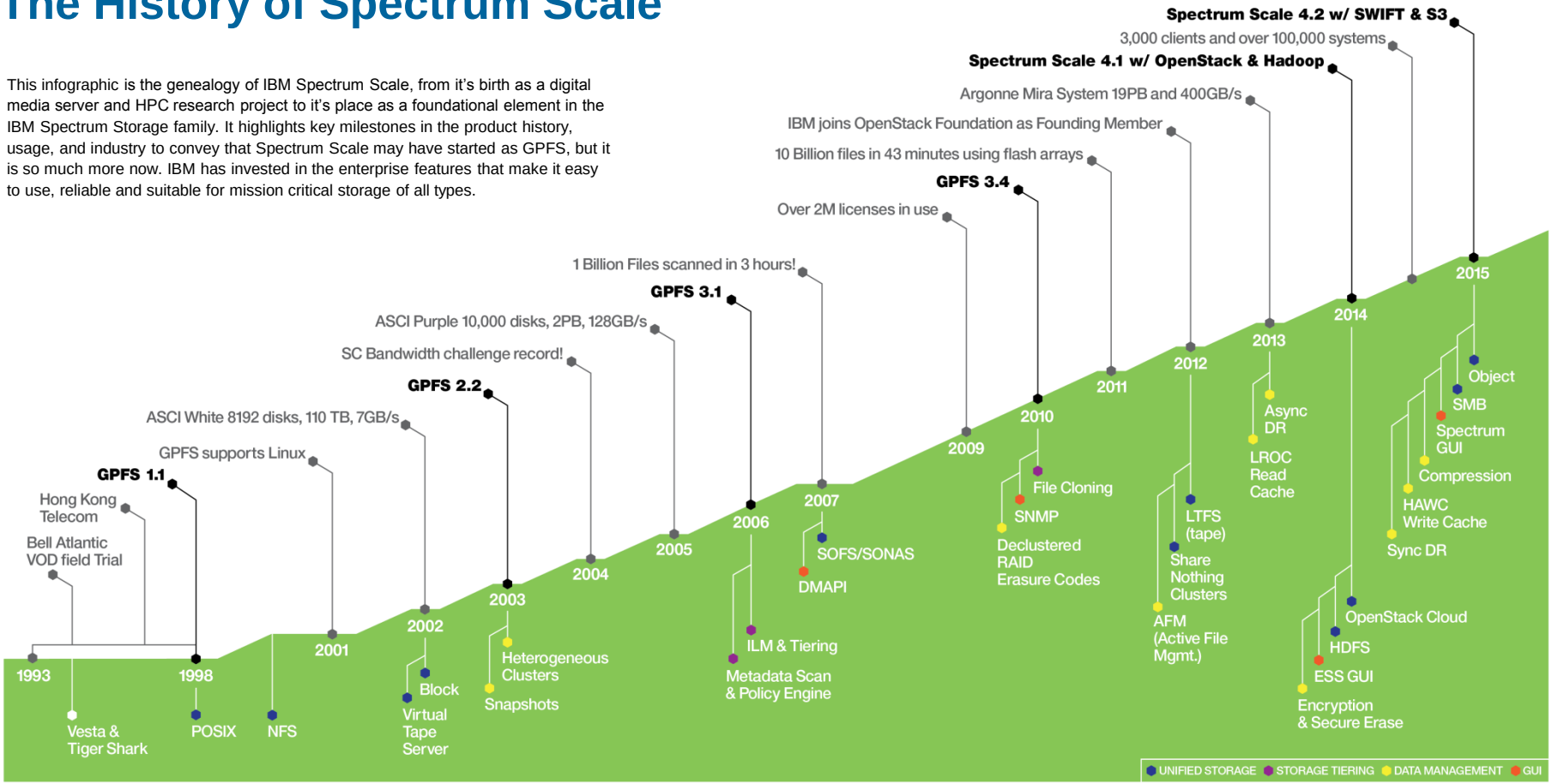
New in Spectrum Scale 4.2
Priorities 2016
Security
Hadoop Integration
Problem Determination

Outline

New in Spectrum Scale 4.2

The History of Spectrum Scale

This infographic is the genealogy of IBM Spectrum Scale, from it's birth as a digital media server and HPC research project to it's place as a foundational element in the IBM Spectrum Storage family. It highlights key milestones in the product history, usage, and industry to convey that Spectrum Scale may have started as GPFS, but it is so much more now. IBM has invested in the enterprise features that make it easy to use, reliable and suitable for mission critical storage of all types.



Store everywhere. Run anywhere.

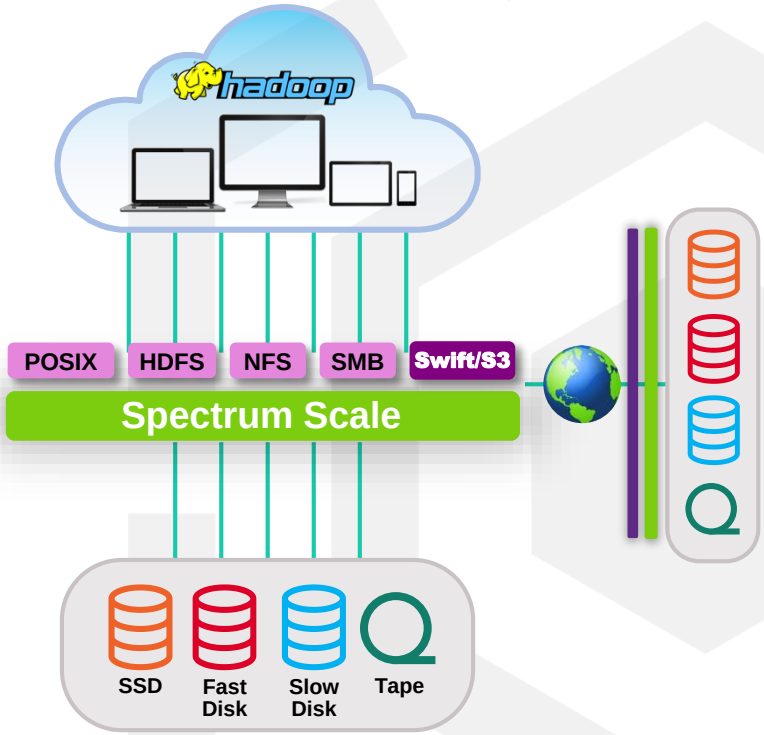
Remove data-related bottlenecks

Challenge

- Managing data growth
  - Lowering data costs
  - Managing data retrieval & app support
  - Protecting business data

Unified Scale-out Data Lake

- File In/Out, Object In/Out; Analytics on demand.
- High-performance native protocols
- Single Management Plane
- Cluster replication & global namespace
- Enterprise storage features across file, object & HDFS



Store everywhere. Run anywhere.

Content Repositories

Challenge

Object storage for static data

- Seamless scaling
- RESTful data access
- Object metadata replaces hierarchy
- Storage efficiency

Spectrum Scale Swift & S3

- High-performance for object
- Native OpenStack Swift support w/ S3
- File or object in; Object or file out
- Enterprise data protection
- Spectrum Scale RAID (ESS)
- Transparent ILM
- Encryption of data at rest and Secure Erase

Store everywhere. Run anywhere.

Analytics without complexity

Challenge

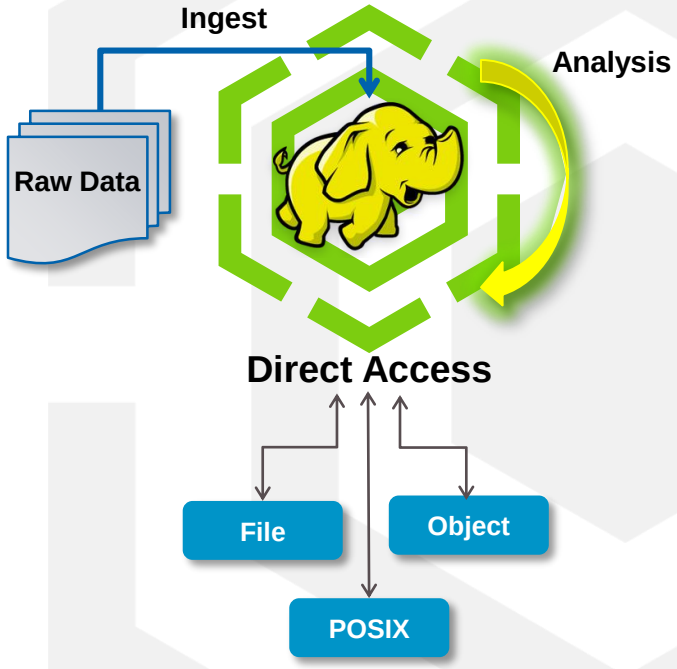
Separate storage systems for ingest, analysis, results

- HDFS requires locality aware storage (namenode)
- Data transfer slows time to results
- Different frameworks & analytics tools use data differently

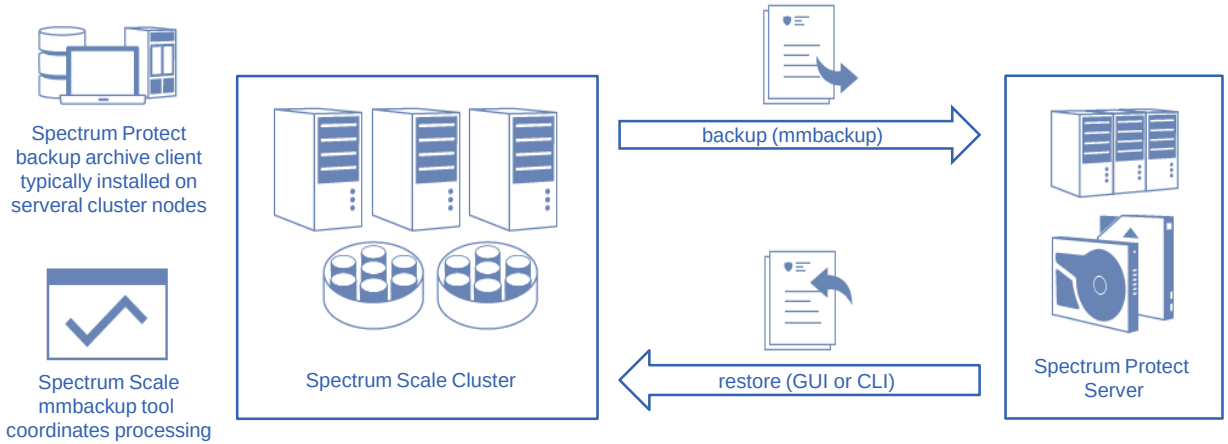
HDFS Transparency

- Map/Reduce on shared, or shared nothing storage
- No waiting for data transfer between storage systems
- Immediately share results
- Single ‘Data Lake’ for all applications
- Enterprise data management
- Archive and Analysis in-place

➔ Analyze object and file data without copying into HDFS



Backup Of Large Spectrum Scale File Systems



Function

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

➔ Use any backup program to backup file, object and Hadoop data
➔ Use Spectrum Protect to benefit from mmbackup and SOBAR to backup and restore huge amounts of data

New in Spectrum Scale 4.2

	New Feature	Benefit
Client Experience Focus	• Common interface across Spectrum Portfolio • GUI Phase 1	Easy to learn UI and integration across Spectrum Storage portfolio Simplify common management functions, including • Enabling protocols • Policy driven placement and ILM • Monitoring • Troubleshooting
Object Storage	• Unified File and Object • Extended S3 API support	Single view of data with wither file or object read and write Enable applications originally written for AWS
Big Data & Analytics	• Native Hadoop Support • Ambari Integration	Higher performance and broader integration with HDFS applications to go beyond Hadoop and embrace Map/Reduce ecosystem
Storage efficiency	• Compression of Cold data for File & Object	• Improve Storage utilization & efficiency for Cold data • Efficienciently reduce data size using compression policies
General	• Quality of Service for File • z Linux support • Sudo wrappers	Expanding functionality in Spectrum Scale data aware policy engine: • Performance reservations to meet SLAs – even by time of day • Extending multi-site resiliency features to z-Linux

Speed and simplicity: Graphical user interface

Reduce administration overhead

- Graphical User Interface for common tasks

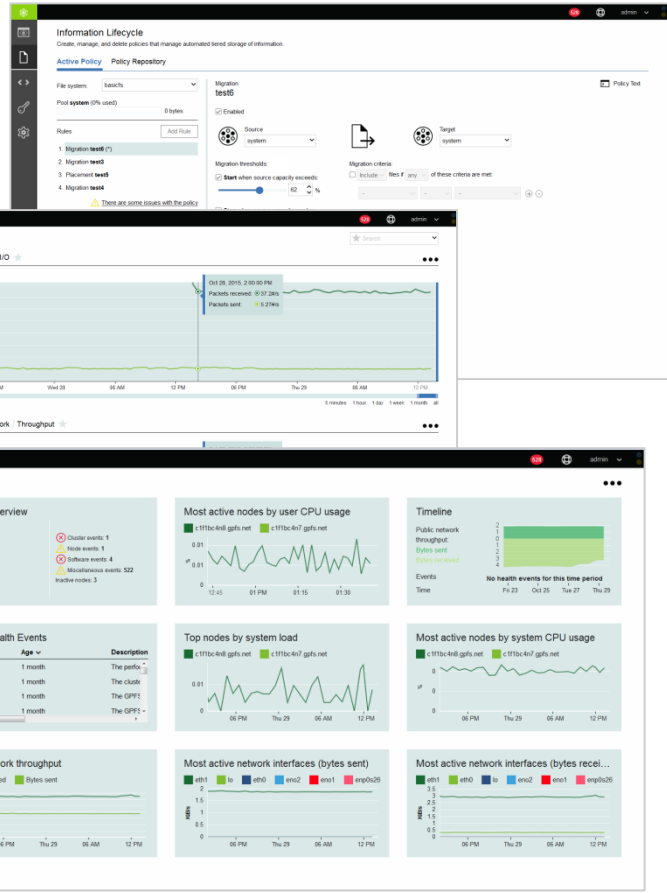
Easy to adopt

- Base interface on common IBM Storage Framework

Integrated into Spectrum Control

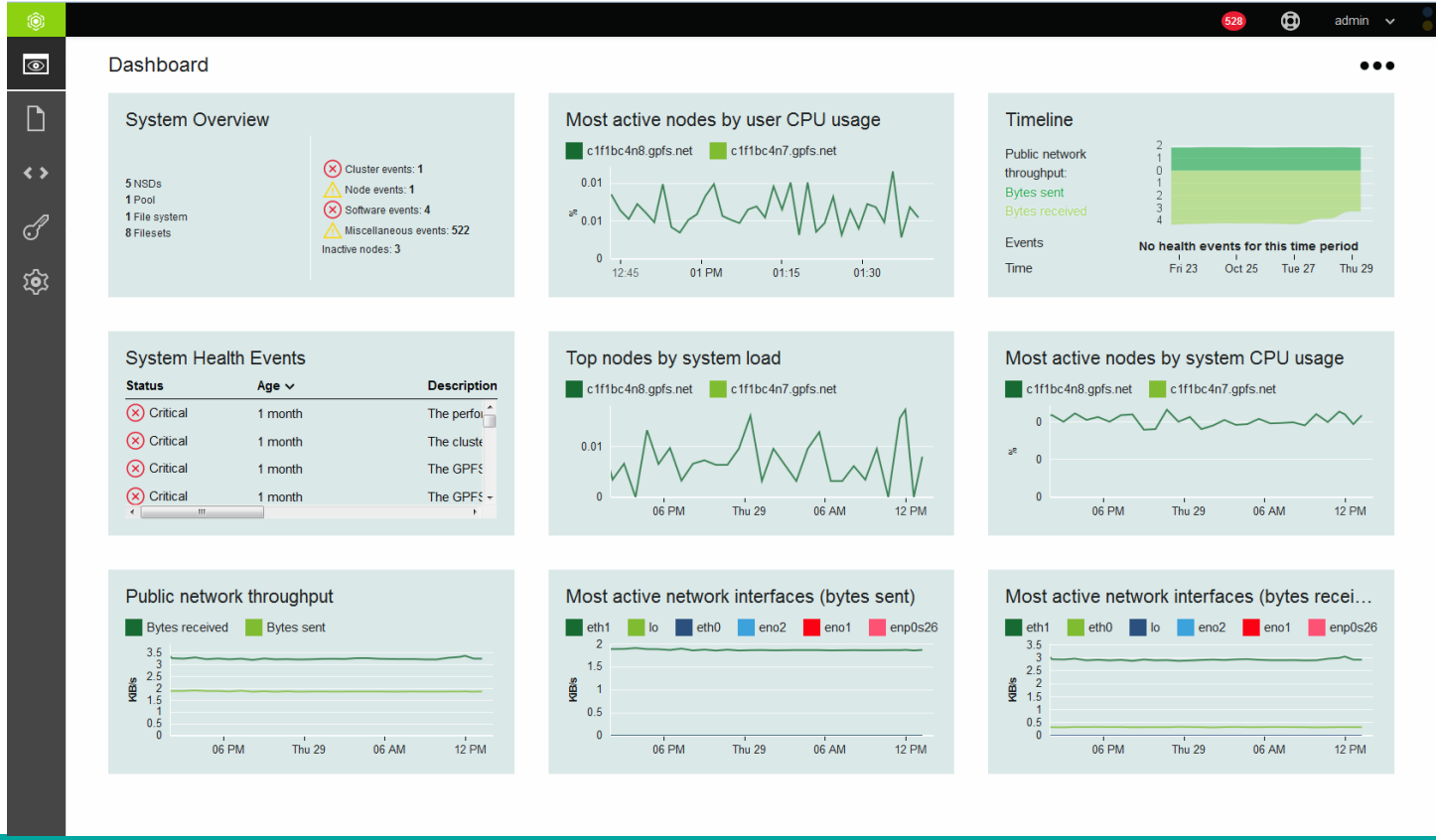
- Storage portfolio visibility
  - Consolidated management
  - Multiple clusters

IBM Spectrum Control



Speed and simplicity: Performance monitoring highlights

System health
Node performance
Network traffic
Historical trends



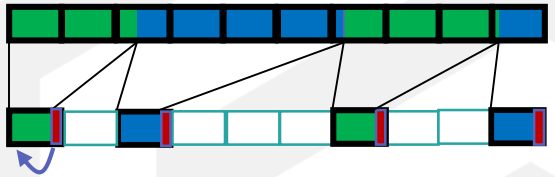
Reduce costs: Compression

Improved storage efficiency
- Typically 2x improvement in storage efficiency

Improved i/o bandwidth
- Read/write compressed data reduces load on storage backend

Improved client side caching
- Caching compressed data increases apparent cache size

Compression is controlled per file
- By administrator defined policy rules



Vision

Which files to compress

When to compress the file data

How to compress the file data

Native Encryption and Secure Erase

Encryption of data at rest

Files are encrypted before they are stored on disk

Keys are never written to disk

No data leakage in case disks are stolen or improperly decommissioned

Secure deletion

Ability to destroy arbitrarily large subsets of a filesystem

No “digital shredding”, no overwriting: secure deletion is a cryptographic operation

Spectrum Scale Virtual Machine

Turn-key Spectrum Scale VM available for download

- Try the latest Spectrum Scale enhancements
- Full functionality on laptop, desktop or server
- Incorporate external storage

Use for live demonstrations, proof of concepts, education, validate application interoperability

- Scripted demonstrations

Limitations

- VirtualBox hypervisor only
- Type-2 Hypervisor limits performance
- Not supported for production workloads
- *Can not be migrated to bare metal*

Priorities 2016

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. Information about potential future products may not be incorporated into any contract. The development, release, and timing of any future features or functionality described for our products remains at our sole discretion.

Performance is based on measurements and projections using standard IBM benchmarks in a controlled environment. The actual throughput or performance that any user will experience will vary depending upon many factors, including considerations such as the amount of multiprogramming in the user’s job stream, the I/O configuration, the storage configuration, and the workload processed. Therefore, no assurance can be given that an individual user will achieve results similar to those stated here

2016 Development Priorities

Every year we define a set of goals

- Based mainly on client feedback and market opportunity
- Target is to achieve them within the year

Focus areas

- Problem determination
- Documentation
- Security
- Defect backlog

Functional enhancements

- Improvements for Big Data
- More flexibility for GNR

Hills – Problem Determination

1. An IT administrator who monitors Spectrum Scale can be made aware of the health of his Spectrum Scale components in one cluster, from a single place.

2. An IT Administrator, can perform self-service problem determination by utilizing provided guidance or automated solutions to problems, without contacting IBM Support.

3. An IT Administrator, can pre-check/check Spectrum Scale and its operating environment to avoid potential problems after initial installation or when changes are made, from a single tool.

Security

Security Work 2016

Subject to change.
Details are under investigation.

Sudo wrapper / no root ssh

- Make GUI functional

File encryption (on rest)

- Consumability improvements in the configuration of SKLM
- Support for the Vormetric key server
- File encryption performance (whitepaper)

Authentication

- GUI admin user can authenticate via external AD or LDAP server (delivered with 4.2.0-1)
- External Keystone SSL support for object

Miscellaneous

- Spectrum Scale security best practices (whitepaper)
- Multi-region object deployment with a highly available keystone service (whitepaper)

Hadoop Integration

A Tale of Two Connectors

GPFS Hadoop Connector

- Henceforth known as the “old” connector
- Emulates a Hadoop compatible filesystem – i.e. replaces HDFS
- Stateless
- Free download – link
- Supports Spectrum Scale 4.1.x, 4.1.1.x and 4.2
- Currently supported with IOP 4.0.x and 4.1.x
- Integrated with Ambari (IOP 4.1.x)

Spectrum Scale HDFS Transparency Connector

- Henceforth known as the “new” connector
- Integrates with HDFS – reuses HDFS client and implements NameNode and DataNode RPCs
- Stateless
- Free download – link
- Supports Spectrum Scale 4.1.x, 4.1.1.x and 4.2
- Planned for IOP 4.2 (April timeframe)
- Ambari integration being developed

Old GPFS Hadoop Connector Approach

How can we be sure we're compatible?
Hadoop File System API intended to be open.

public abstract class
org.apache.hadoop.fs.FileSystem

Source: *hadoop.apache.org*

"All user code that may potentially use the Hadoop Distributed File System should be written to use a FileSystem object."

Latest File System APIs are described here:
https://hadoop.apache.org/docs/current/api/org/apache/hadoop/fs/FileSystem.html

Old GPFS Hadoop Connector Approach

All based on org.apache.hadoop.fs.FileSystem API

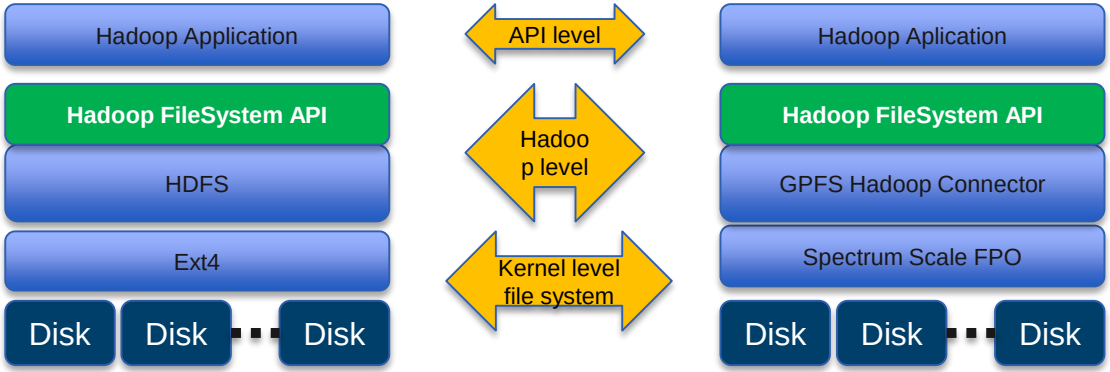
	Optimized for
HDFS	General Hadooop
GlusterFS	file-based scale-out NAS
OrangeFS	high end computing (HEC) systems
SwiftFS	write directly to containers in an OpenStack Swift object store
GridGain	In-Memory Data Fabric
Lustre	
MapR FileSystem	
Quantcast File System	
▪etc...	

Spectrum Scale (GPFS) is no different

Source: https://wiki.apache.org/hadoop/HCFS

Old GPFS Hadoop Connector Approach

Applications communicate with Hadoop using FileSystem API. Therefore, transparency is preserved.



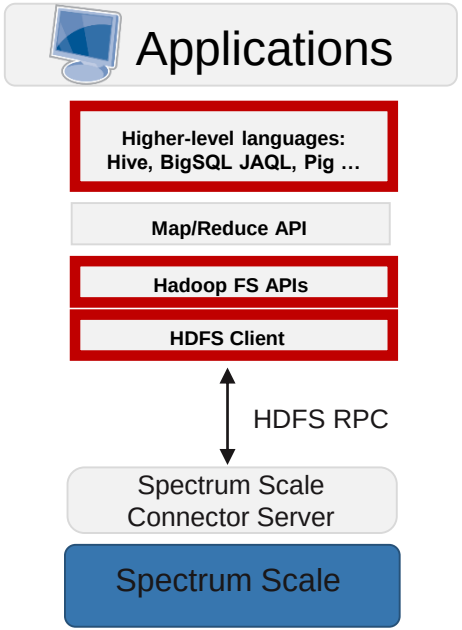
*"All user code that may potentially use the Hadoop Distributed File System should be written to use a **FileSystem** object."*

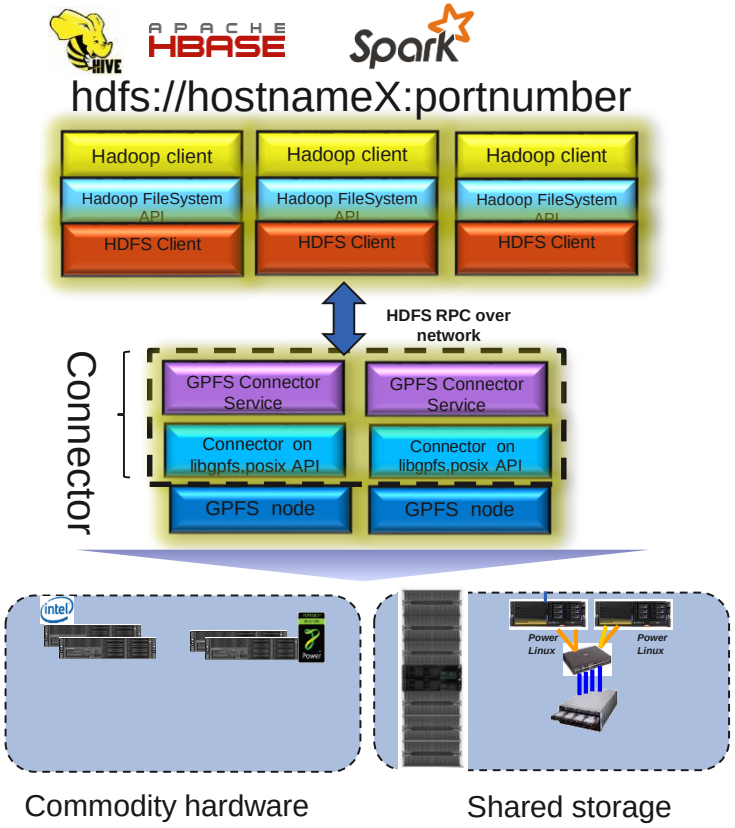
Source: *hadoop.apache.org*

New Spectrum Scale HDFS Transparency Design

- Issues with old Hadoop Connector
  - Some applications and many tools do not use org.apache.hadoop.fs.FileSystem
  - Those applications and tools fail with HDFS Connector
- Key Advantages of new HDFS Transparency Connector
  - Support workloads that have hard coded HDFS dependencies
  - Simpler integration for currently compatible workloads & components
  - Leverage HDFS Client cache for better performance
  - No need to install Spectrum Scale clients on all nodes
  - Full Kerberos support for Hadoop ecosystem

New Spectrum Scale HDFS Transparency Design

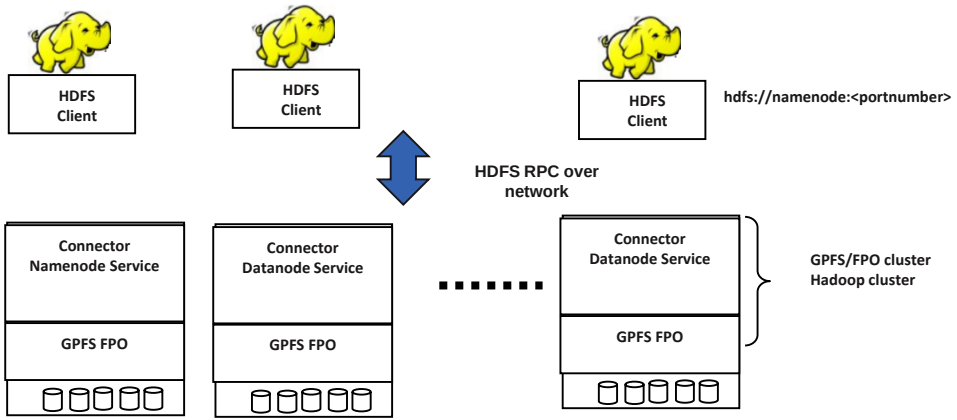




Supported Hadoop versions: 2.7.1

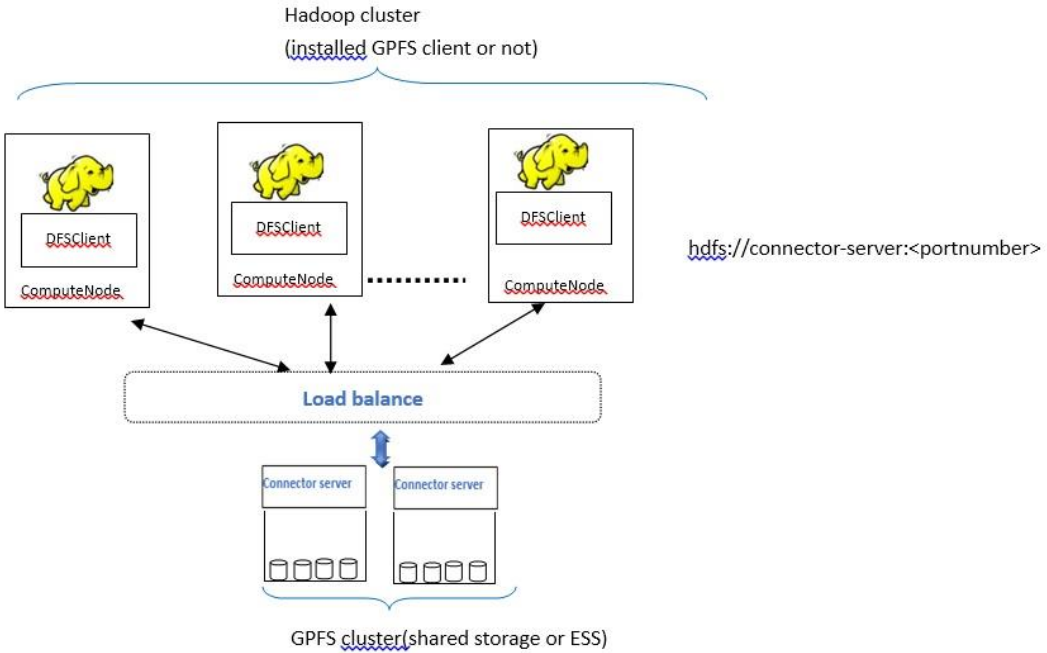
New Spectrum Scale HDFS Transparency Design

- Each node will be installed with connector datanode server
- Only one node will be installed with connector namenode server
- Connector namenode server will be configured with HA, just similar as HDFS
- GA'ed 2015/11/20



New Spectrum Scale HDFS Transparency Design

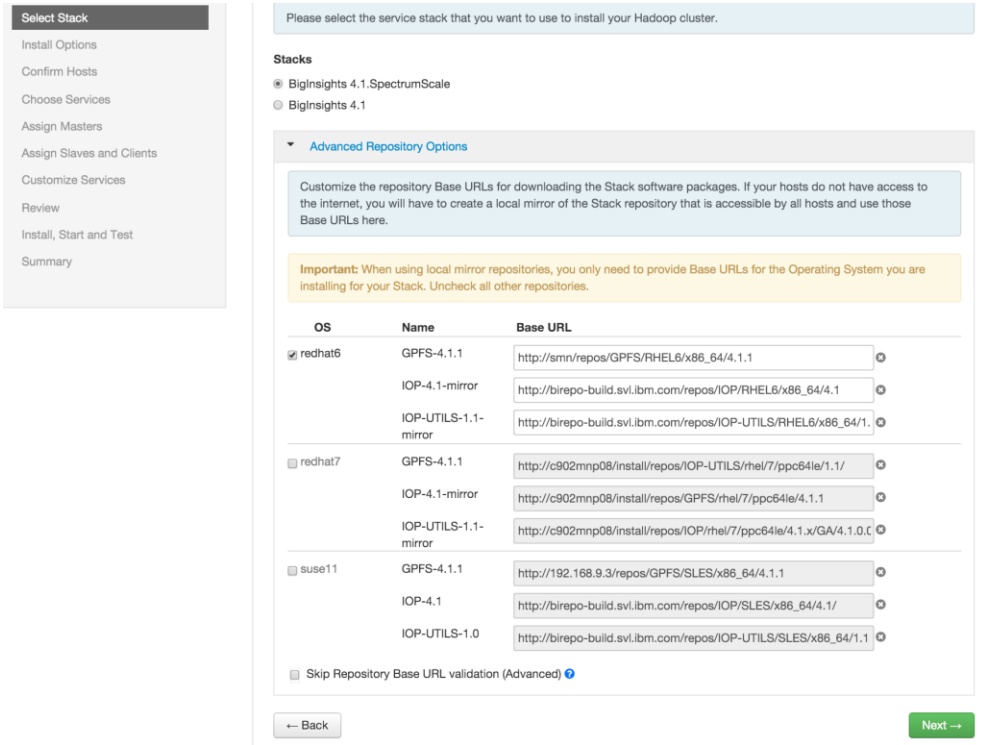
- Connector servers are installed over limited nodes (ex. GPFS NSD servers)
- GPFS client is not needed over the Hadoop computing nodes
- DNS rotation or CES can be used to load balance for HDFS Client
- GA'ed for 2016/1/22



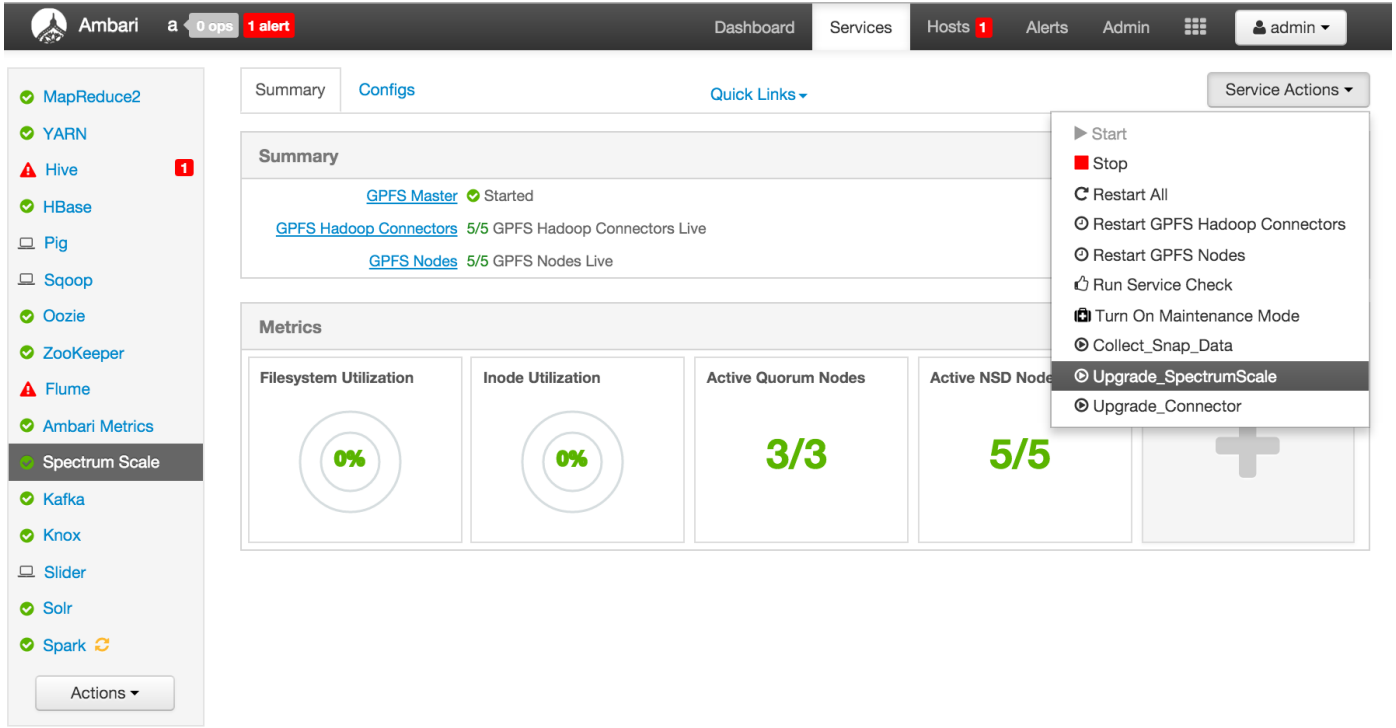
Current Ambari Integration

- New BigInsights 4.1.SpectrumScale stack
- Inherits from BigInsights 4.1 stack
- Removes HDFS, add Spectrum Scale, change all dependencies

- Can install IOP + Spectrum Scale (either new GPFS filesystem or integrate with existing filesystem)
- Value Add integration
- Basic Spectrum Scale monitoring (AMS)
- Support separate connector control
- Support GPFS and connector upgrades
- Collect GPFS snap
- Change GPFS parameters
- Add new nodes
- Remove nodes
- Provide quick link to Spectrum Scale GUI for full management and monitoring



Current Ambari Integration



Ambari Integration with HDFS Transparency

- Biggest change is that there is no new stack
- Spectrum Scale is added as a new service after full IOP install with HDFS (use dummy directory / mount point for HDFS)
- Spectrum Scale service “integrates” with HDFS
- Will support “un-integrate” capability
  - Flip back and forth between HDFS & GPFS
  - Will not move data back and forth between HDFS & GPFS
- Will simplify future upgrades

Subject to change.
Details are under investigation.

Outlook

- Coming soon
  - BigInsights 4.2 support (additional components)
  - HDFS + Spectrum Scale Federation
  - Federate multiple Spectrum Scale clusters
  - Isolate multiple Hadoop clusters on the same filesystem (restrict to sub-directory)

Problem Determination – Health Status (Hill 1)

1

An IT administrator who monitors Spectrum Scale can be made aware of the health of his Spectrum Scale components in one cluster, from a single place.

A user will be able to:

- Issue a single command and see status for all components
- Create thresholds for any ZIMON metric & be notified if it is hit
- Identify the top processes by CPU, memory, network

Today

"There is really no clear way to understand what a healthy cluster looks like. If there is someone who knows, I'd love to talk to them."

Users rely on a wide variety of commands to monitor their Spectrum Scale cluster. This requires them to understand:

- Which components are important to monitor?
- Which commands should I use to monitor each component type?
- How do I interpret the results of all of the commands?
- How to assemble some sort of monitoring framework to piece everything together.

Subject to change.
Details are under investigation.

Outcome

Mockup!!

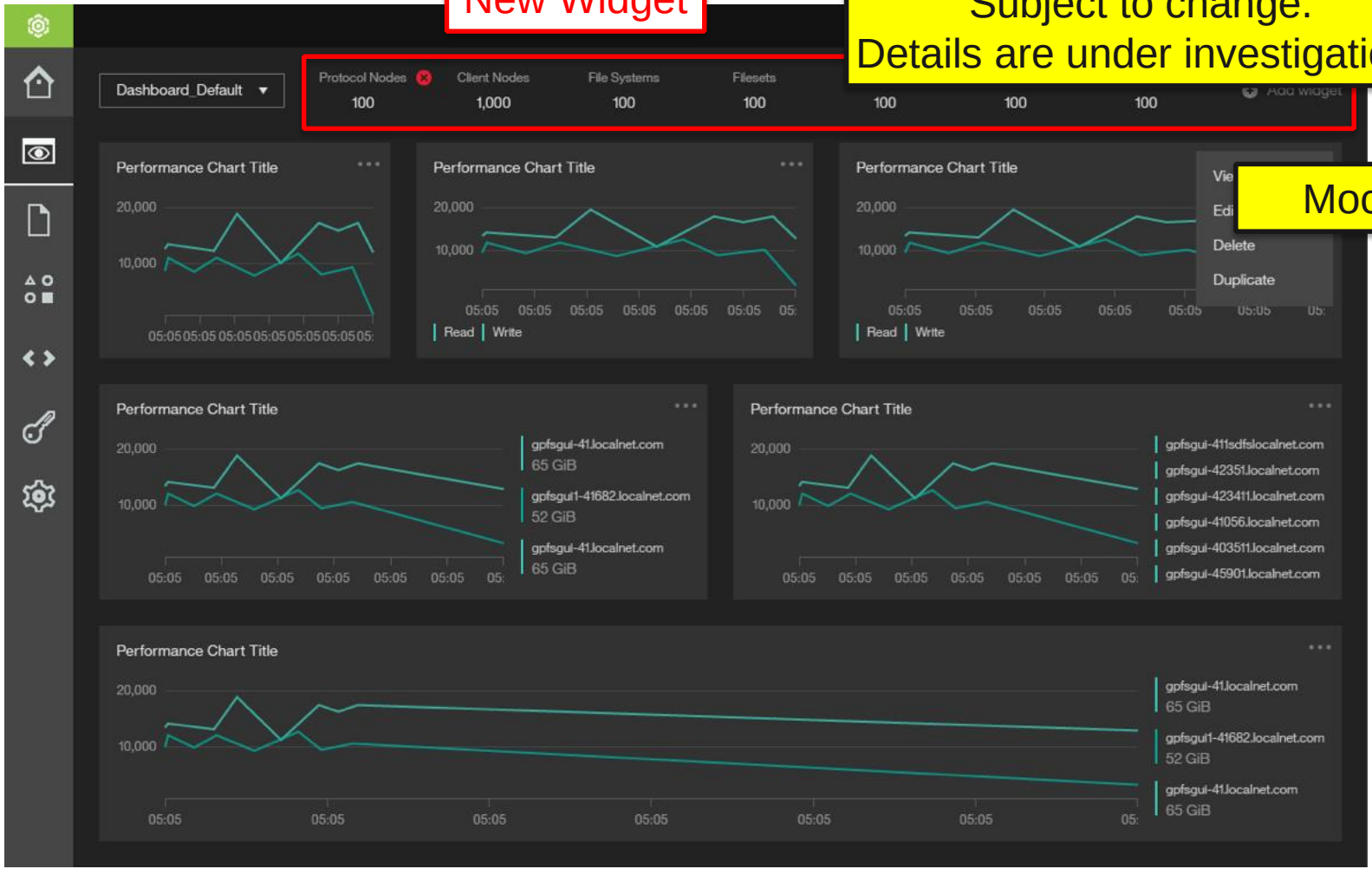
A single CLI command that provides a health overview of all key components in the entire cluster.

```
$ mmhealth node show

Node name:      test_node
Node status:    degraded

Component       Status          Reasons
-----------------------------------------
GPFSDaemon      healthy         -
CES             failed          smbd_down
ZIMon           healthy         -
Network         healthy         -
LocalDisk       healthy         -
FSMount         healthy         -
```

```
$ mmhealth node show -v

Node name:      test_node
Node status:    degraded

Component       Status          Reasons
-----------------------------------------------
GPFSDaemon      healthy         -

CES             failed          smbd_down
    Auth            healthy       -
    OBJ_Auth        healthy       -
    NFS             healthy       -
    OBJ             healthy       -
    SMB             failed        smbd_down

ZIMon           healthy         -

Network         healthy         -

LocalDisk       healthy         -
    DiskA           healthy       -
    DiskB           healthy       -
    DiskC           healthy       -
    DiskD           healthy       -
    DiskE           healthy       -

FSMount         healthy         -
    FSI             healthy       -
    FSII            healthy       -
```

New Widget

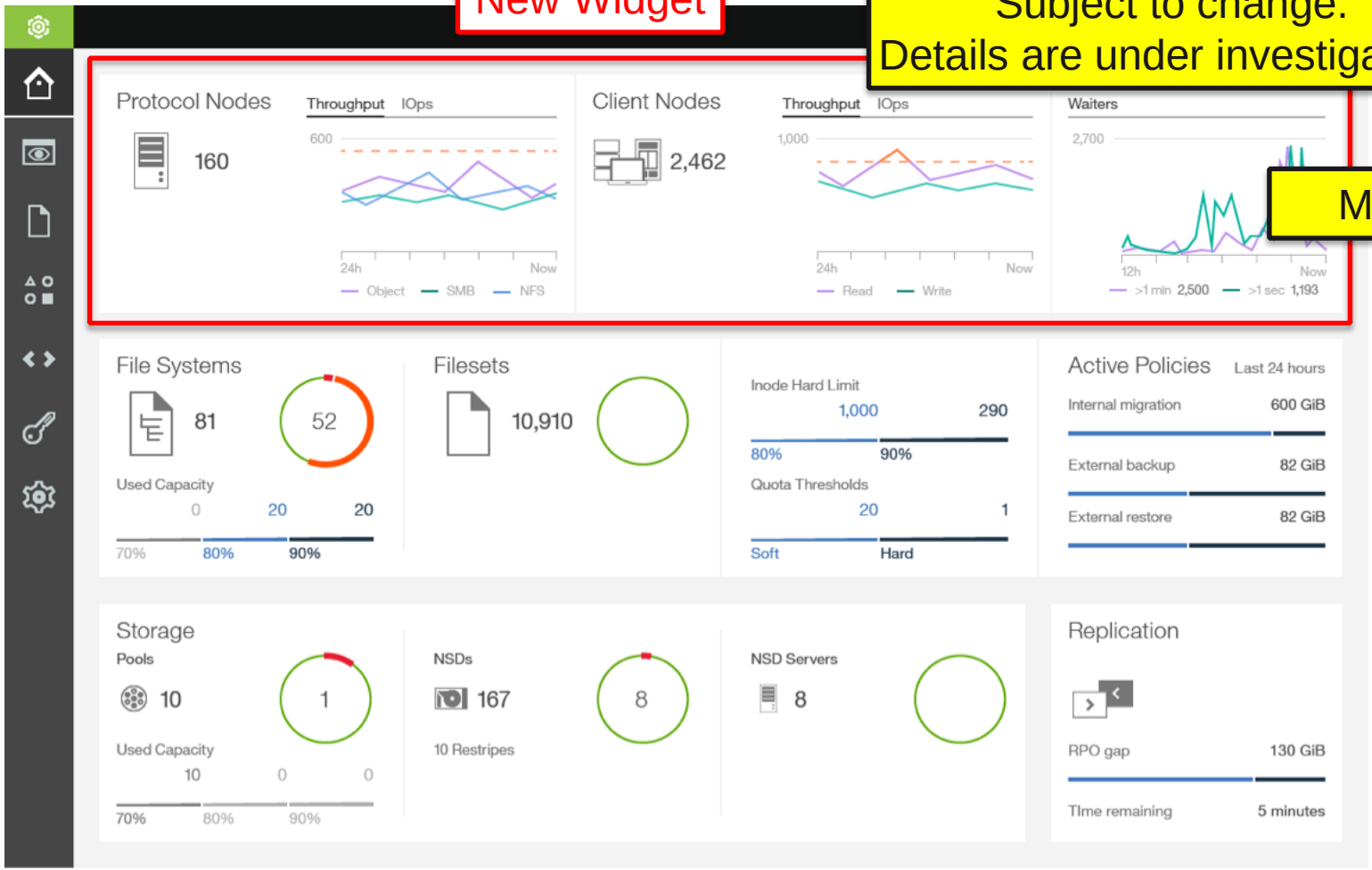
Subject to change.
Details are under investigation.



Mockup!!

New Widget

Subject to change.
Details are under investigation.



Mockup!!

Today

Operations Team Members don't know if a value is good or bad.

Administrator and architect level users want the ability to set thresholds so lower level operations teams can assess if a value is a problem or not.

Subject to change.
Details are under investigation.

Outcome

A user can create thresholds for any Zimon metric and be notified if the threshold is hit.

Problem Determination – Self-Service Problem Determination (Hill 2)

2

An IT Administrator, can perform self-service problem determination by utilizing provided guidance or automated solutions to problems, without contacting IBM Support.

A user will be able to:

- Receive guidance for key problems
- Read documentation on best practices and troubleshooting

Today

"I can look at each individual file system in the cluster and see the reads, the writes, the opens, the closes. I use that frequently to look into performance related issues where I see a large amount of traffic in the cluster. If I can isolate traffic to a particular file system, I can figure out which group is doing the traffic. I can look at individual nodes and determine if they are doing a large amount of traffic on the file system. This allows me to trace back to the job that is running on the node at the time of the issue.”

"First I look for the most active file system and then figure out who has the most jobs and activity going."

Outcome

A user can troubleshoot the performance of a file system

- Overall client workload
- Top clients by workload
- Storage workload and latency
- Waiters
- Protocol workload
- ILM policy

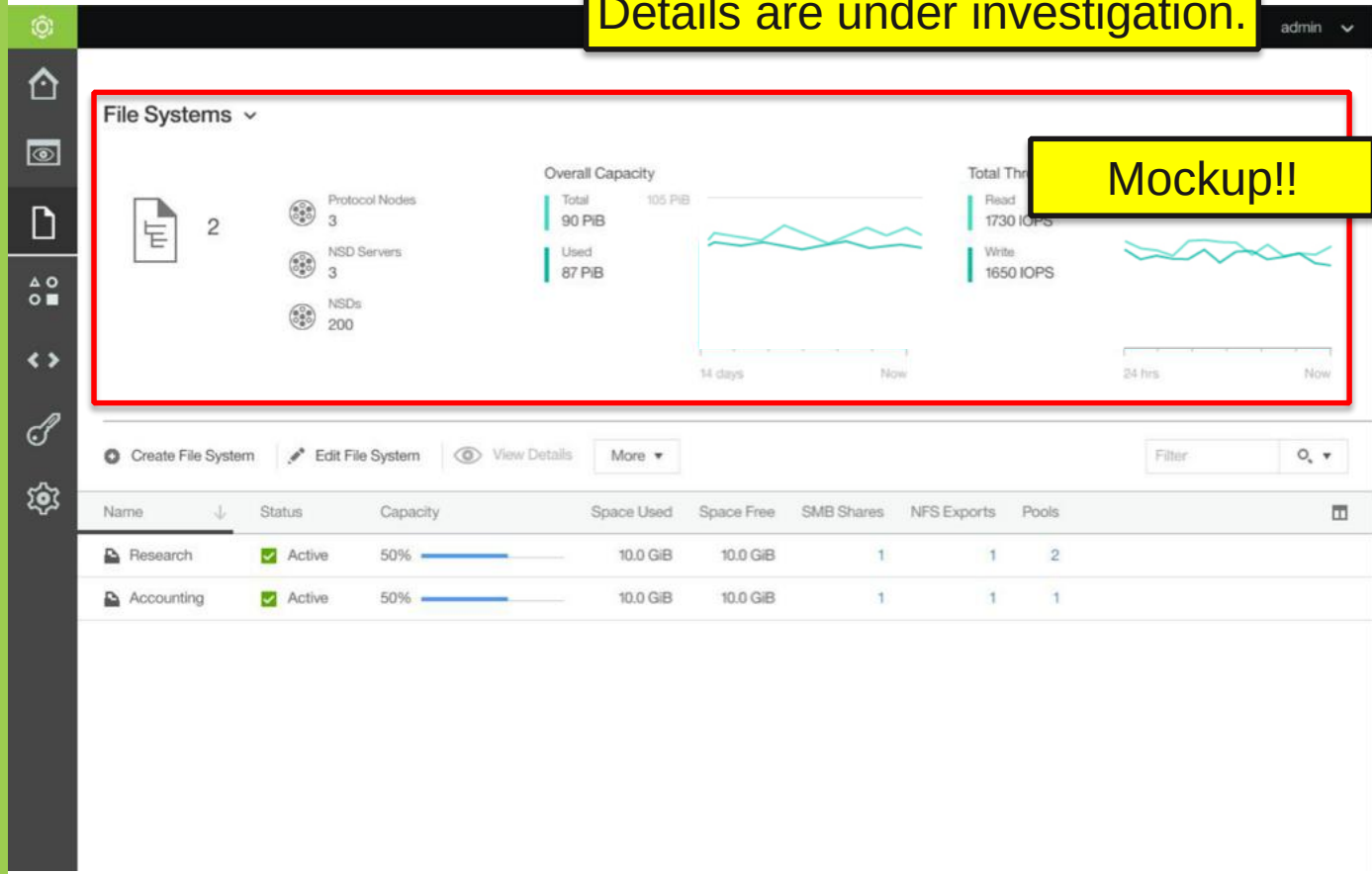
Scenario 1:

A Spectrum Scale admin has been informed that jobs are taking a long time to run on a file system. They are able to view:

- The health of a file system
- Events that have impacted the health
- Average response time for the file system to understand if it is unusually high.
- Storage that the file system is built off of and determine which NSDs have the highest latency.
- Overall workload running against the file system so they can determine if it is unusually high.

New Widget

Subject to change.
Details are under investigation.



Concept, not final design

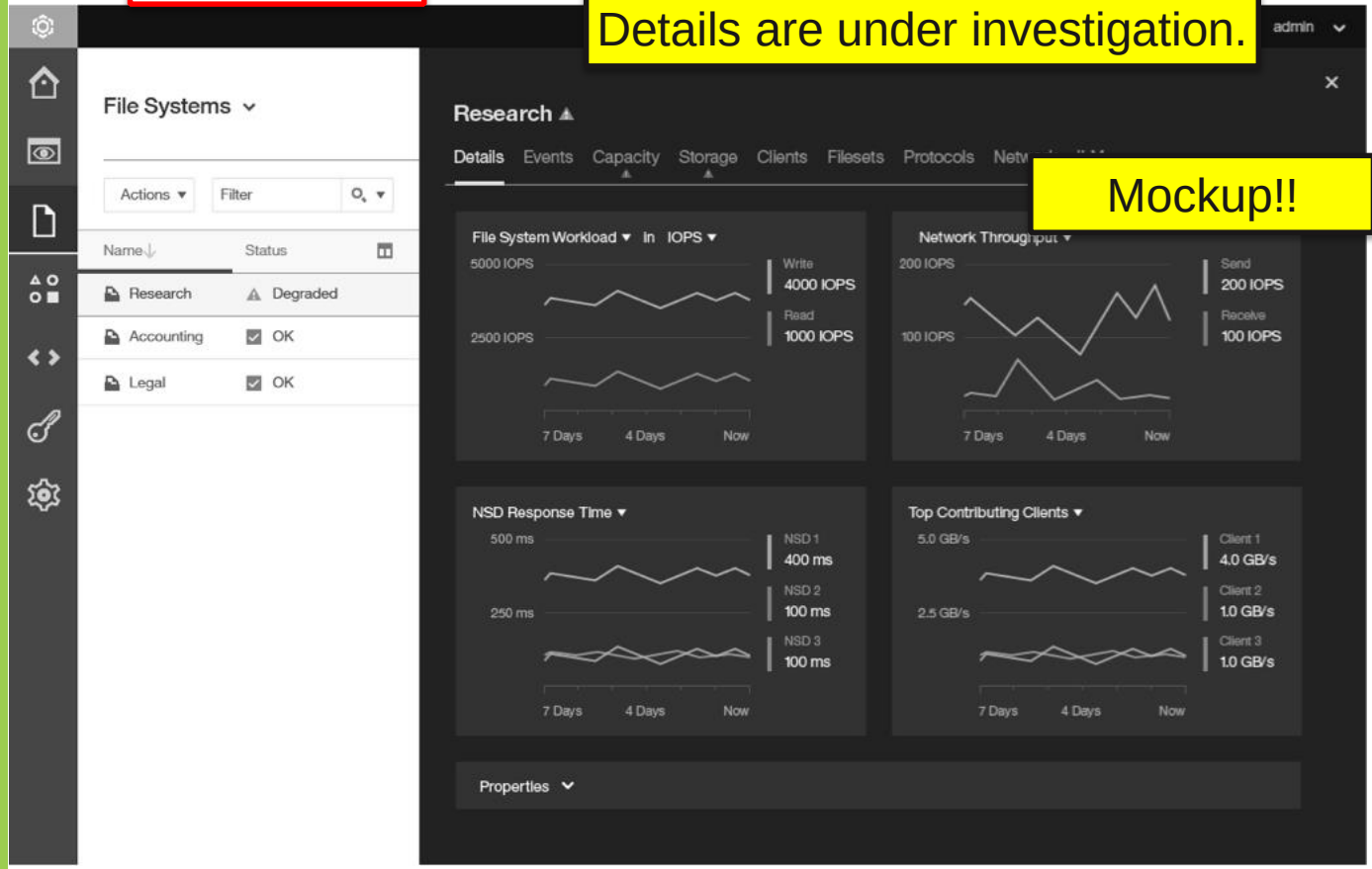
Scenario 1:

A Spectrum Scale admin has been informed that jobs are taking a long time to run on a file system. They are able to view:

- The health of a file system
- Events that have impacted the health
- Average response time for the file system to understand if it is unusually high.
- Storage that the file system is built off of and determine which NSDs have the highest latency.
- Overall workload running against the file system so they can determine if it is unusually high.

New Panel

Subject to change.
Details are under investigation.



Mockup!!

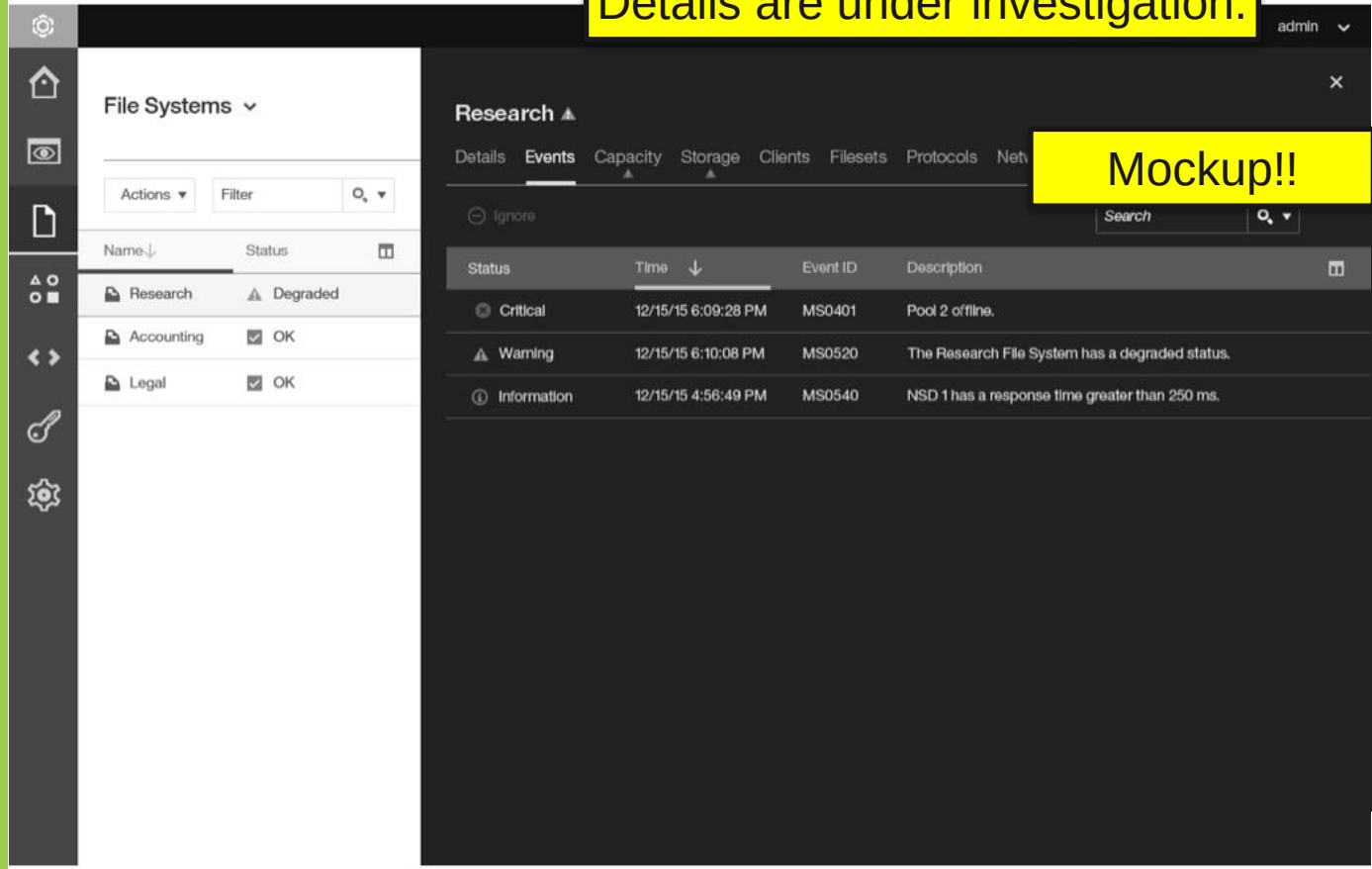
Scenario 1:

A Spectrum Scale admin has been informed that jobs are taking a long time to run on a file system. They are able to view:

- The health of a file system
- Events that have impacted the health
- Average response time for the file system to understand if it is unusually high.
- Storage that the file system is built off of and determine which NSDs have the highest latency.
- Overall workload running against the file system so they can determine if it is unusually high.

New Panel

Subject to change.
Details are under investigation.



Mockup!!

Concept, not final design

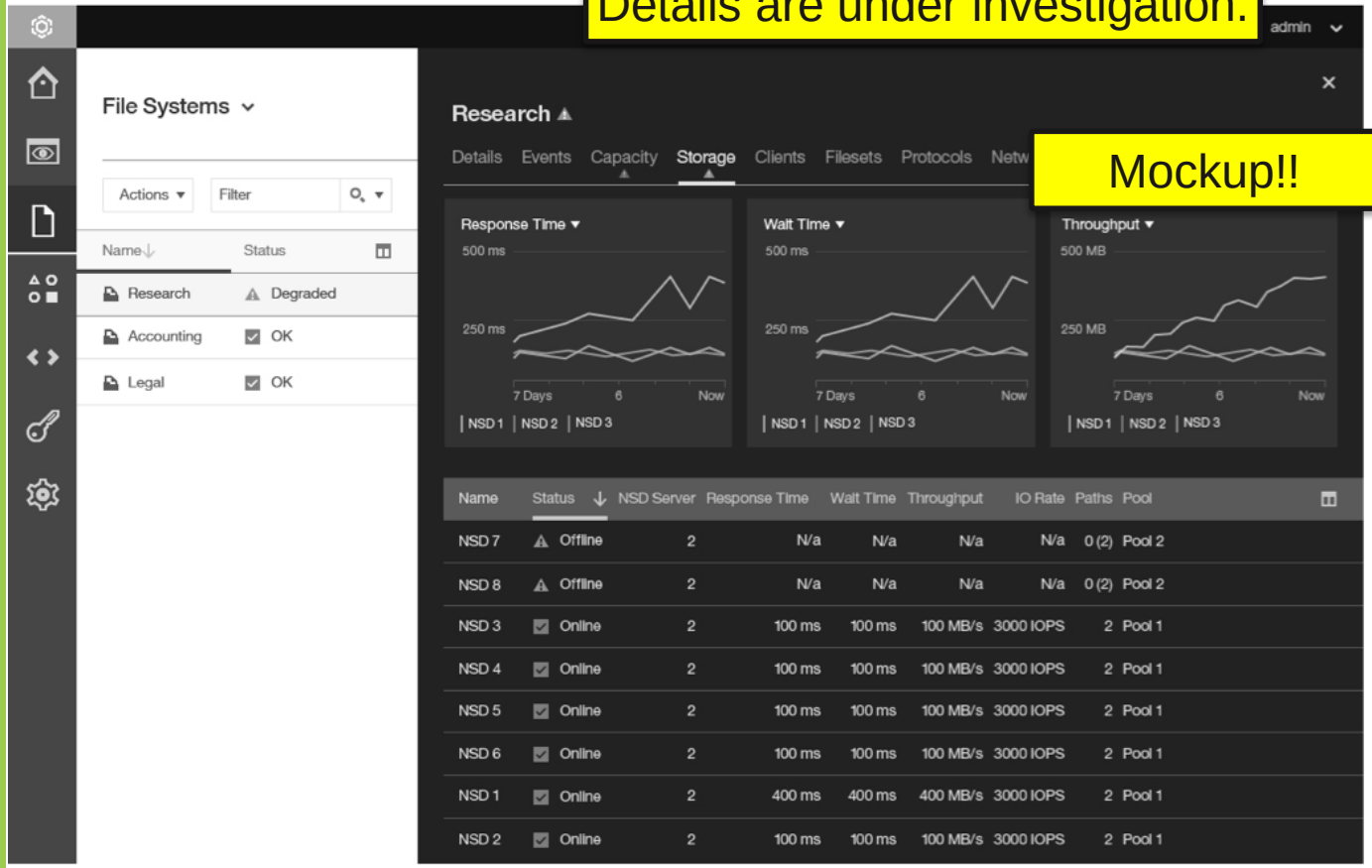
Scenario 1:

A Spectrum Scale admin has been informed that jobs are taking a long time to run on a file system. They are able to view:

- The health of a file system
- Events that have impacted the health
- Average response time for the file system to understand if it is unusually high.
- Storage that the file system is built off of and determine which NSDs have the highest latency.
- Overall workload running against the file system so they can determine if it is unusually high.

New Panel

Subject to change.
Details are under investigation.



Concept, not final design

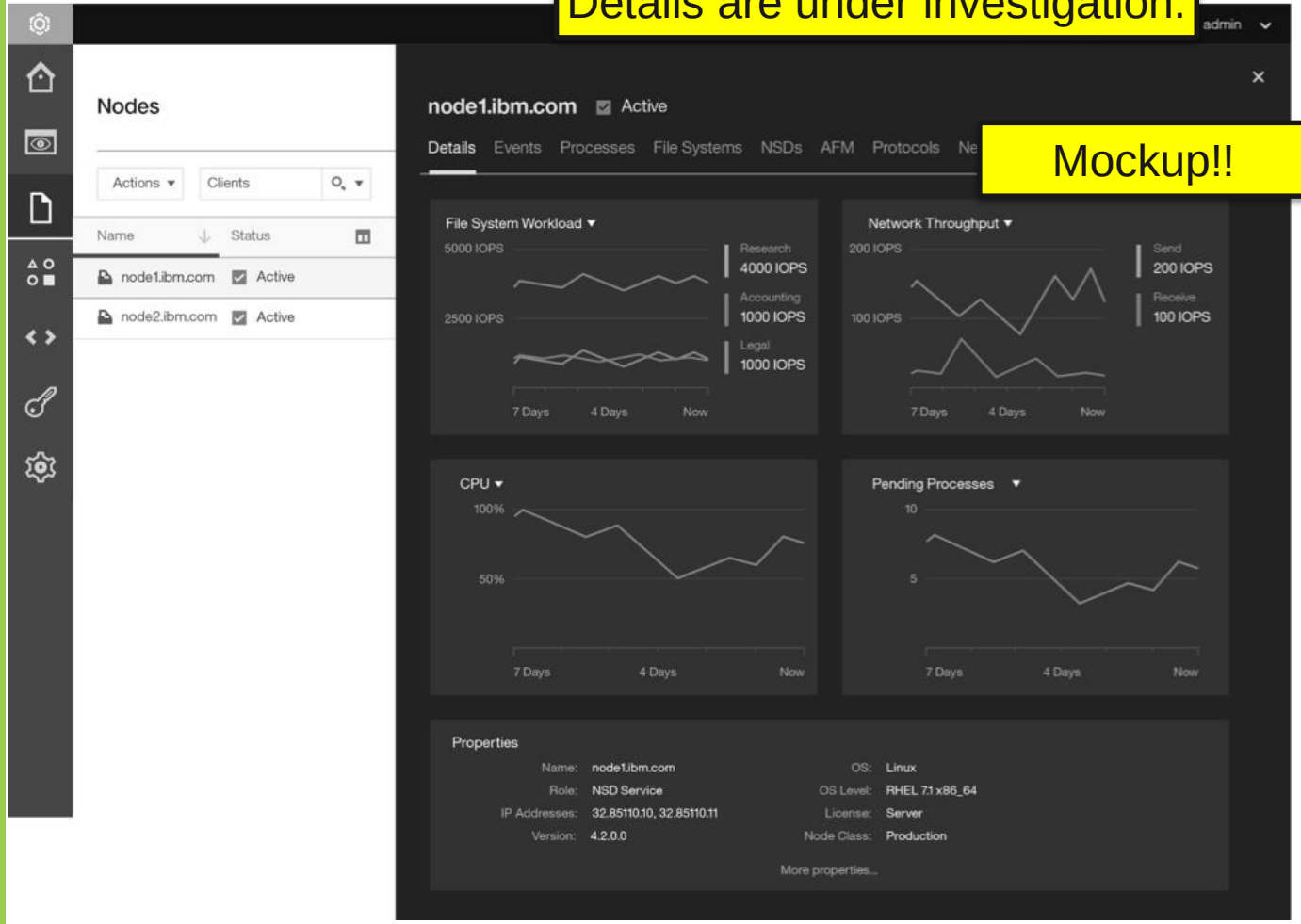
Scenario 2:

The Spectrum Scale admin verifies that the file system's workload is unusually high. They are able to:

- Identify the client that is pushing the highest workload against the file system
- Determine the top processes running on the client
- Transaction size for the client workload and how it has changed over time
- Network workload for the node

New Panel

Subject to change.
Details are under investigation.



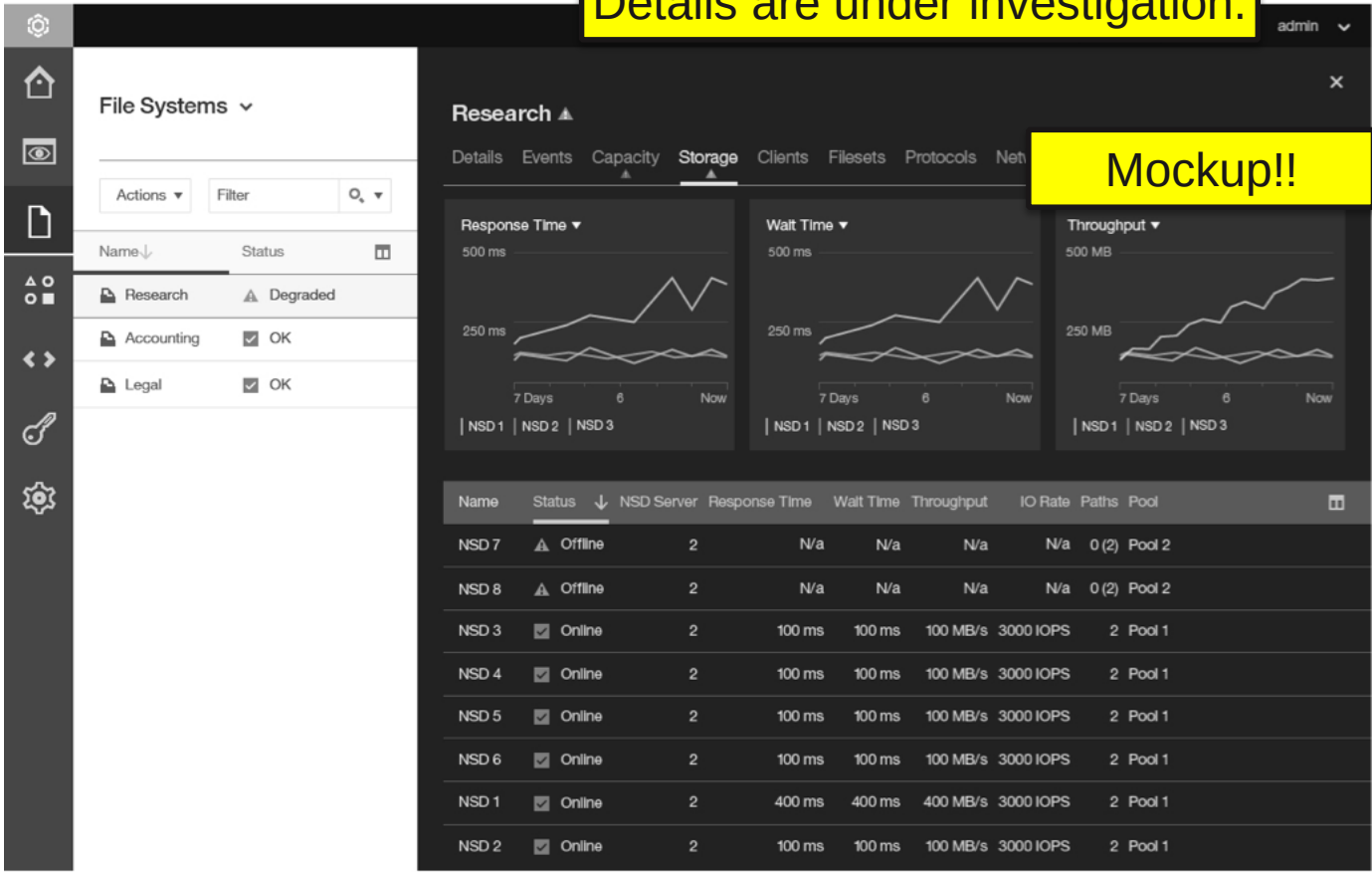
Scenario 3:

A Spectrum Scale admin has realized that file system latency is due to a storage latency. They are able to understand:

- The performance of the NSDs that the file system is built off of
- The performance of the NSD servers that are providing access to the NSDs
- Whether reduced paths for an NSD is causing a particular NSD server to be a bottleneck

New Panel

Subject to change.
Details are under investigation.



Mockup!!

Concept, not final design

Today

"Utilizing the GPFS waiter information, it becomes obvious that all of the waiters on one server."

"Understanding and analyzing this is key to getting to the bottom of many problems"

"Looking at waiters tells you what's backed up, so checking that for a pattern can reveal bad applications which are beating up the file system."

New Panel

Subject to change.
Details are under investigation.

Outcome

Mockup!!

Considered by customers to be the key metric that they monitor and use for problem determination

Waiter metrics will be added to existing performance charts

Problem Determination – Check Spectrum Scale and its environment (Hill 3)

3

An IT Administrator, can pre-check/check Spectrum Scale and its operating environment to avoid potential problems after initial installation or when changes are made, from a single tool.

A user will be able to use a:

- Network verification tool to understand if there are network problems
- Active directory monitoring tool to prevent issues

Today

"When we have issues and we're pretty sure it is the network, we still have to spend however many hours to write a test case that doesn't involve GPFS to prove that it is exclusively the network...A network verification tool would be a big help."

Subject to change.
Details are under investigation.

Outcome

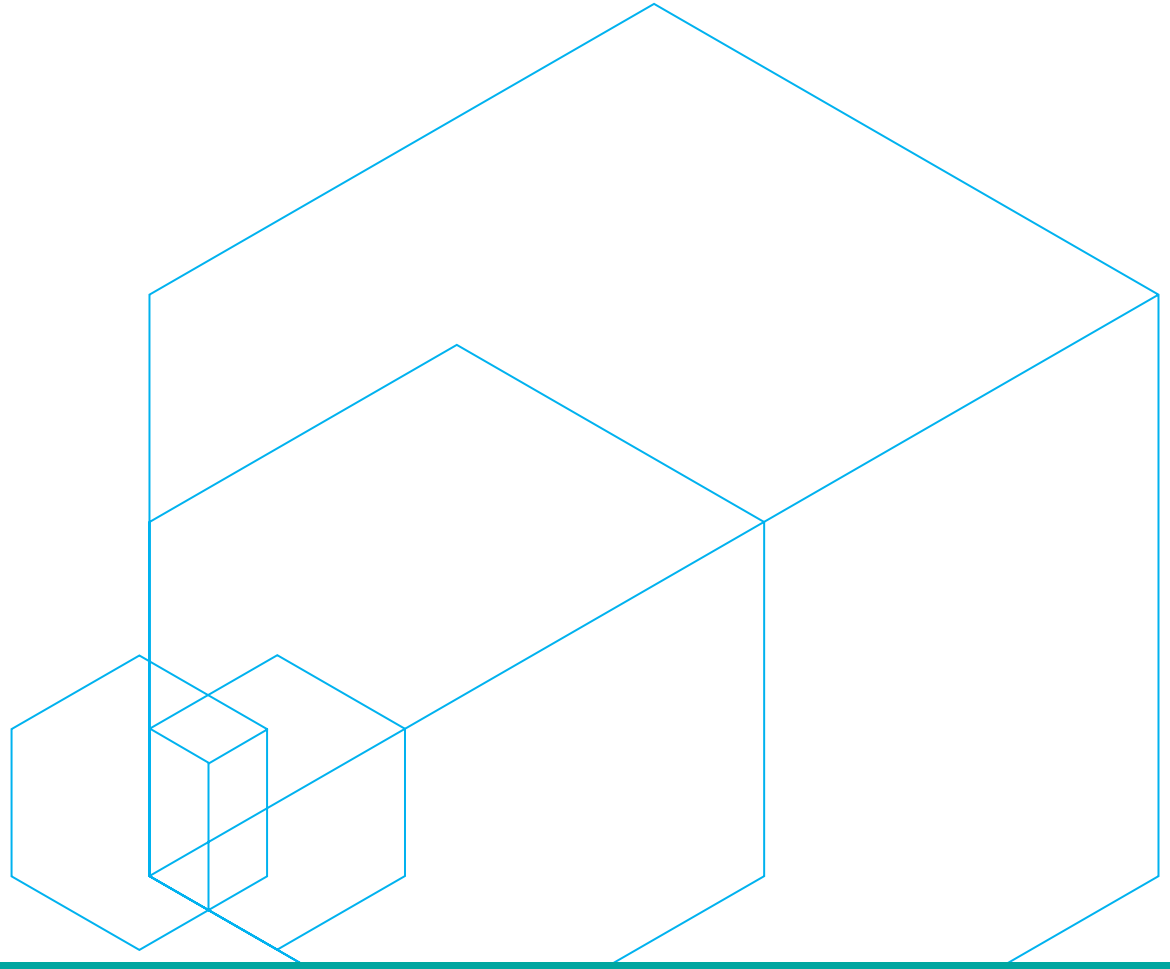
Users can verify node to node connectivity to detect to common network issues and point customers to the root cause

- General connectivity issues
e.g. broken IP/Routing/Switch config, Infiniband connectivity

- Firewall configuration issues
GPFS User Group feedback:
“...had misconfigured firewalls, such that they could reach our home cluster nodes on port 1191, but our home cluster nodes could not reach them on 1191 or on any of the ephemeral ports.”

- Network Performance issues
Low throughput / High Latency
Run reference workload (nsdperf) and measure performance

Thank you.



ibm.com/systems

IBM Spectrum Scale Value

Storage management at scale	Store everywhere. Run anywhere.	Improve data economics	Software Defined Open Platform
New GUI & health monitoring	Advanced routing with latency awareness	Tier seamlessly	Heterogeneous commodity storage: flash, disk, & tape
Unified File, Object & HDFS	Read or Write Caching	Incorporate and share flash	Software, appliance or Cloud
Distributed metadata & high-speed scanning	Active File Management for WAN deployments	Policy driven compression	Data driven migration to practically any target
QoS management	File Placement Optimization	Data protection with erasure code and replication	File/Object In/Out with OpenStack SWIFT & S3
1 Billion Files & yottabytes of data	End-to-end data integrity	Native Encryption and Secure Erase compliance	Transparent native HDFS
Multi-cluster management with Spectrum Control	Snapshots	Target object store and cloud	Integration with cloud
	Sync or Async DR	Leading performance for Backup and Archive	

Introducing IBM Spectrum Scale

Highly scalable high-performance unified storage
for files and objects with integrated analytics

Remove data-related bottlenecks
Demonstrated 400 GB/s throughput

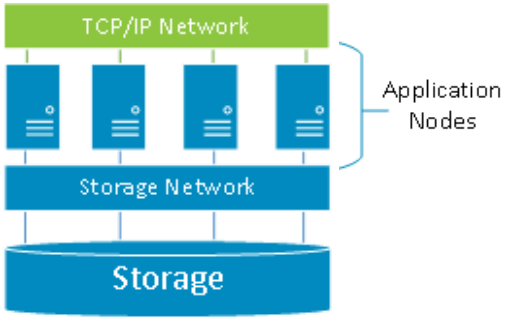
Enable global collaboration
Data Lake serving HDFS, files & object across sites

Optimize cost and performance
Up to 90% cost savings & 6x flash acceleration

Ensure data availability, integrity and security
End-to-end checksum, Spectrum Scale RAID, NIST/FIPS certification

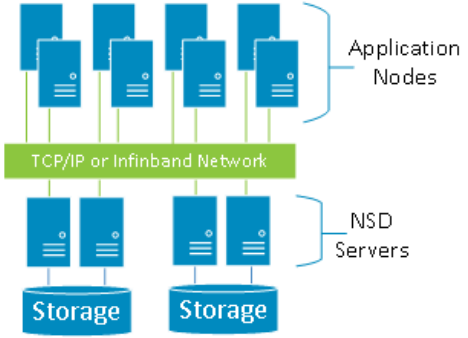
Spectrum Scale deployment models

Enterprise Integrated Model



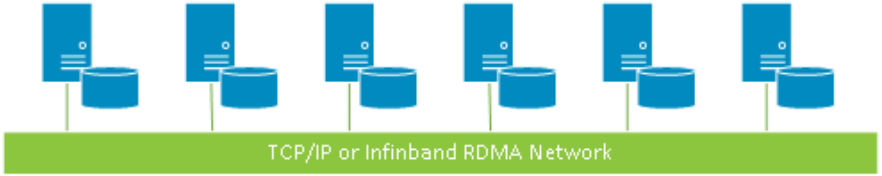
Unify and parallelize storage silos

Network Shared Disk (NSD) Model



Modular High-Performance Scaling

Shared Nothing Cluster (SNC) Model



Span storage rich servers for converged architecture or HDFS deployment

Spectrum Scale Parallel Architecture

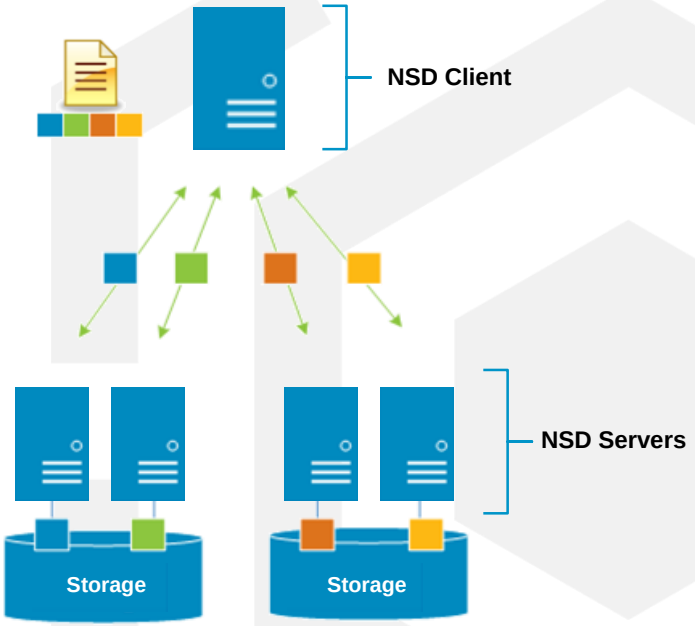
No Hot Spots

All NSD servers export to all clients in active-active mode

Spectrum Scale stripes files across NSD servers and NSDs in units of file-system block-size

File-system load spread evenly

Easy to scale file-system capacity and performance while keeping the architecture balanced



NSD Client does real-time parallel I/O
to all the NSD servers and storage volumes/NSDs

IBM Spectrum Scale performance features

Quality of Service

- Throttle background functions such as rebuild or async replication
- Set by flexible policy, such as day-of-week and time-of-day

Highly Available Write Cache (HAWC)

- Improves performance of small synchronous writes
- Small synch writes are written to the log. As log fills, rewrite to home.

Local Read Only Cache (LROC)

- Extend the page pool memory to include local DAS/SSD for read caching

Policy driven compression

- Compress only what makes sense & extends to cache

Distributed and flash accelerated metadata

- Metadata includes directories, inodes, indirect blocks

Lift data to the highest tiers based on the file's "heat"

Store everywhere. Run anywhere.

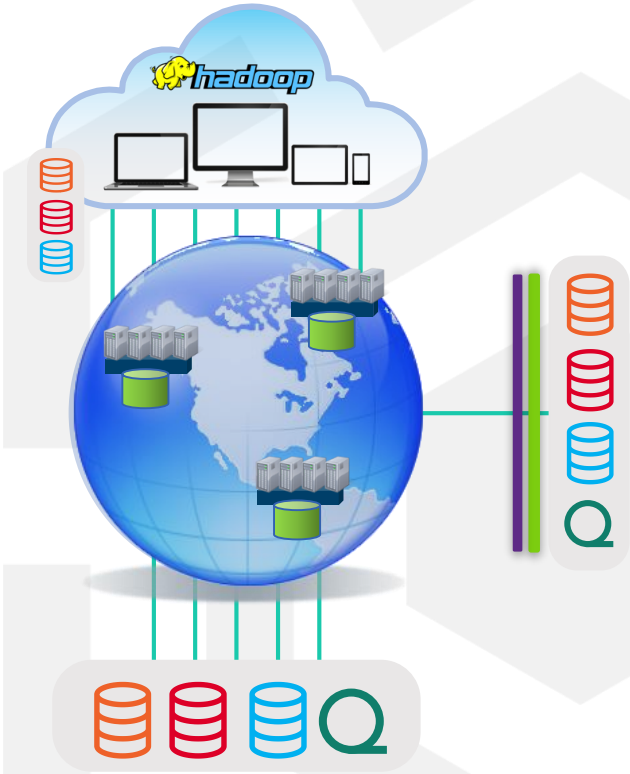
Enable Global Collaboration

Challenge

- Multiple sites working on same data
  - Remote access is slower than local
  - Consistent metadata & data locking
  - Support for mission critical transactional replication
  - Manage unreliable, remote sites

Advanced File Management, Routing & Caching

- Global namespace with fast, consistent metadata
- Latency aware
- Multi-writer and multi-reader
- Automatic failover and seamless file-system recovery



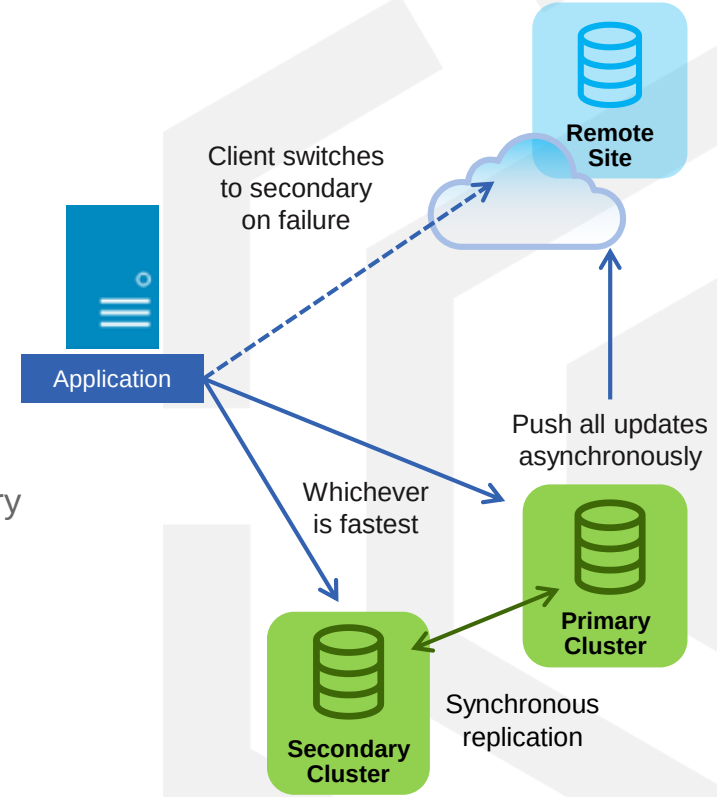
Global collaboration options

Single global namespace enables:

Remote Mount
- Single copy of data
- Use caching to speed local access

Synchronous replication
- Active/Active data access
- Simultaneous write is sensitive to network latency
- Read from fastest source
- DR with automatic failover and seamless file-system recovery

Asynchronous replication
- Active/Passive data access
- Write now, copy later across network
- Write to Active, Read from fastest
- Any storage target, including cloud



Spectrum Scale Advanced File Management (AFM)

Spans geographic distance and unreliable networks

- Caches local ‘copies’ of data distributed to one or more Spectrum Scale clusters
- Low latency ‘local’ read and write performance
- As data is written or modified at one location, all other locations see that same data
- Efficient data transfers over wide area network (WAN)

Speeds data access to collaborators and resources around the world

- Unifies heterogeneous remote storage

Asynchronous DR is a special case of AFM

- Bidirectional awareness for Fail-over & Fail-back with data integrity
- Recovery Point Objectives for volume & application consistency

Store everywhere. Run anywhere.

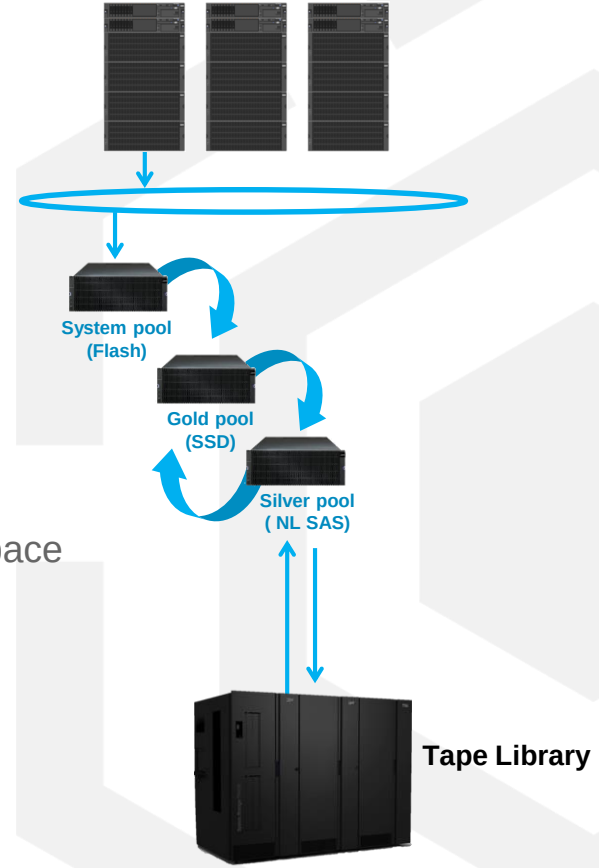
Optimize Cost and Performance

Challenge

- Data growth is outpacing budget
  - Low-cost archive is another storage silo
  - Flash is under utilized because it isn't shared
  - Locally attached disk can't be used with centralized storage
  - Migration overhead is preventing storage upgrades

Automated data placement

- Span entire storage portfolio, including DAS, with a single namespace
- Policy driven data placement & data migration
- Share storage, even low-latency flash
- Automatic failover and seamless file-system recovery
- Lower TCO



Data aware cost optimization

Powerful policy engine

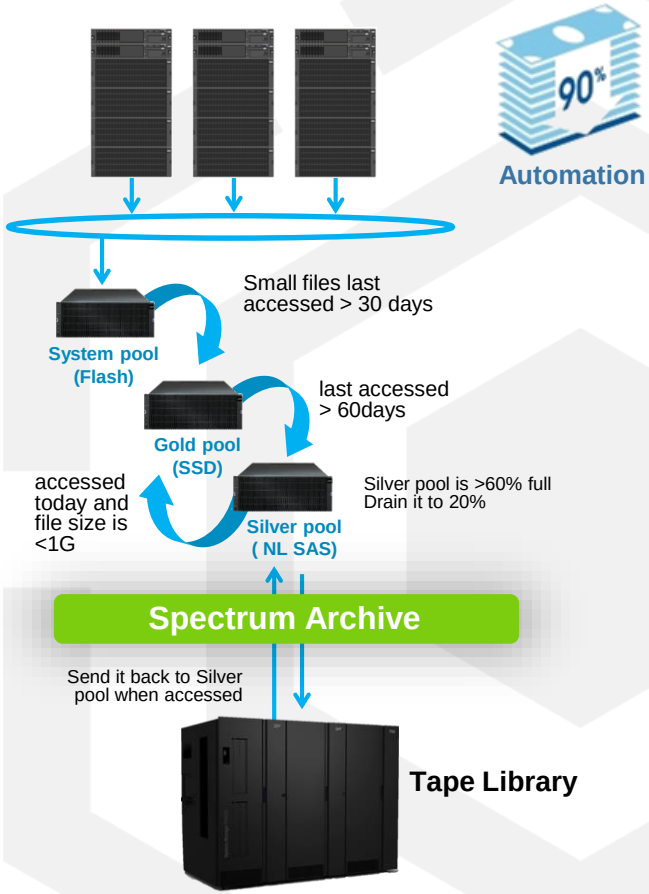
- Information Lifecycle Management
- Fast metadata ‘scanning’ and data movement
- Automated data migration to based on threshold

Users not affected by data migration

- Single namespace

Example: Online storage reaches 90% full then move all 1GB or larger files that are 60 days old to offline to free up space

Integrated with Spectrum Archive

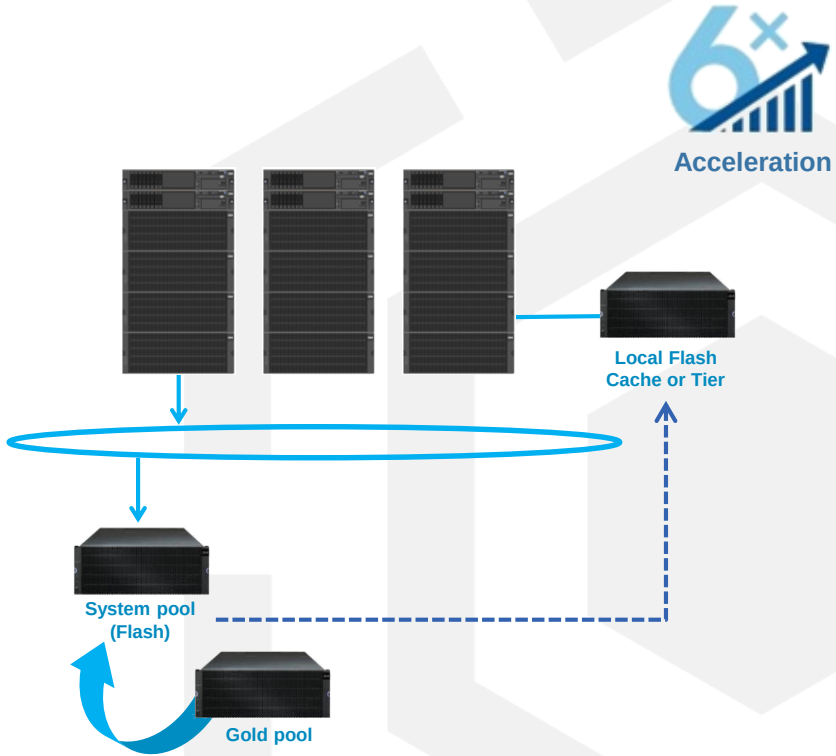


Data aware performance optimization

Alternative to explicit policies
- Respond to changing workload

Data identified as “Hot” data
- High-speed metadata
- Access pattern analysis
- Migrate closer to client

Flash can be added anywhere
- Read from “Fastest”
- Latency & cache aware



Store everywhere. Run anywhere.

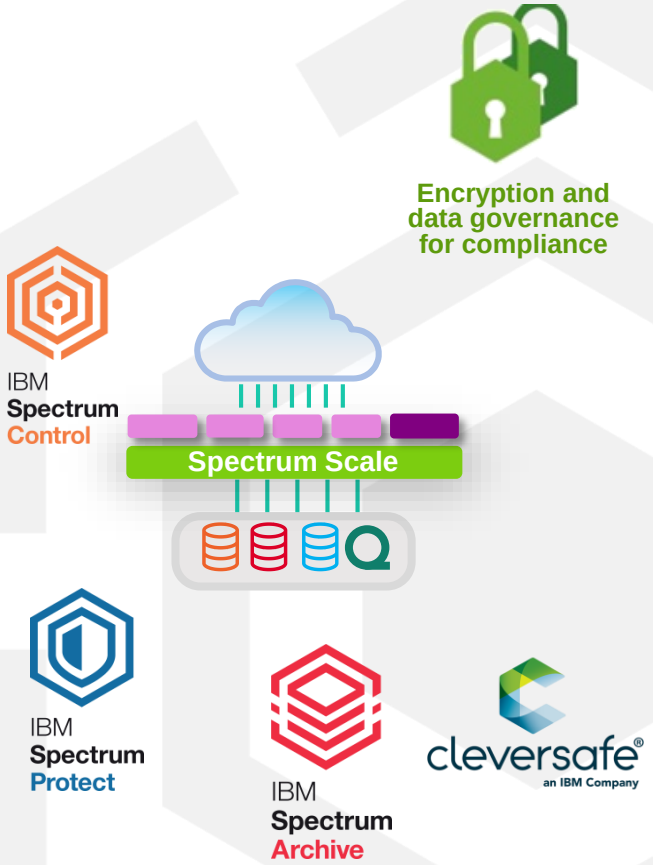
Ensure data availability, integrity and security

Challenge

- Business data is going on new storage types
  - HDFS replication scheme lacks data integrity
  - Object storage lacks features, including backup
  - Authentication across data center should be the same

Enterprise Features

- Universal data access
- A single authentication scheme
- Data dispersal and erasure code for faster rebuild times
- End-to-end checksum to catch errors
- Data protection through Snapshots, Replication, Backup, and/or Disaster Recovery
- Data encryption and cryptographically secure erase
- Integration to Spectrum Family



Native Encryption and Secure Erase

Native: Encryption is built into the **“Advanced”** product

Protects data from security breaches, unauthorized access, and being lost, stolen or improperly discarded

Cryptographic erase for fast, simple and secure file deletion

Complies with **NIST SP 800-131A** and is **FIPS 140-2** certified

Supports HIPAA, Sarbanes-Oxley, EU and national data privacy law compliance

Get it your way

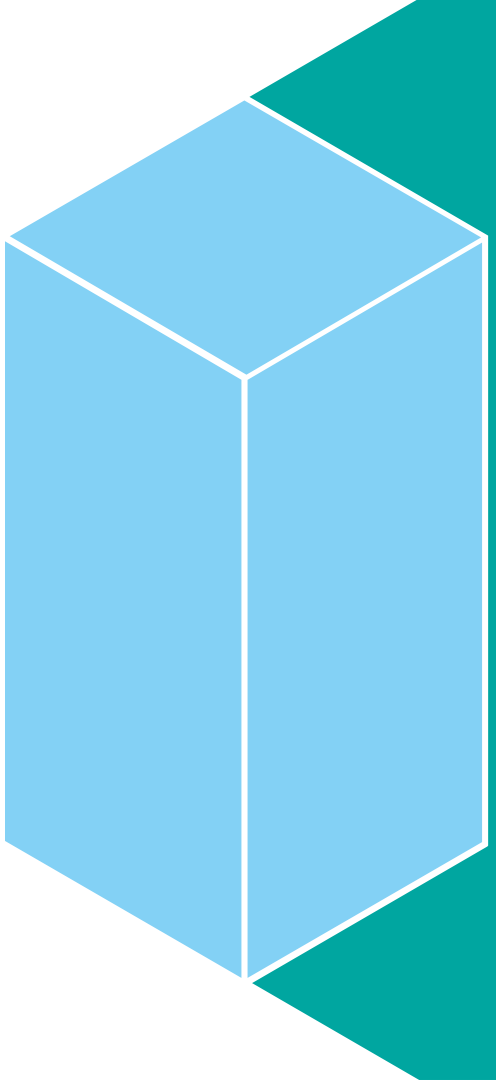
Software

Appliance

Cloud service

Elastic Storage Server

Appliance



IBM Elastic Storage Server (ESS)

Integrated scale out data management for file and object data

Optimal building block for high-performance, scalable, reliable enterprise storage

- Faster data access with choice to scale-up or out
- Easy to deploy clusters with unified system GUI
- Simplified storage administration with IBM Spectrum Control integration

One solution for all your data needs

- Single repository of data with unified file and object support
- Anywhere access with multi-protocol support: NFS 4.0, SMB, OpenStack Swift, Cinder, and Manila
- Ideal for Big Data Analytics with full Hadoop transparency with 4.2

Ready for business critical data

- Disaster recovery with synchronous or asynchronous replication
- Ensure reliability and fast rebuild times using Spectrum Scale RAID's dispersed data and erasure code

Advantages of Spectrum Scale RAID

Use of standard and inexpensive disk drives

- Erasure Code software implemented in Spectrum Scale

Faster rebuild times

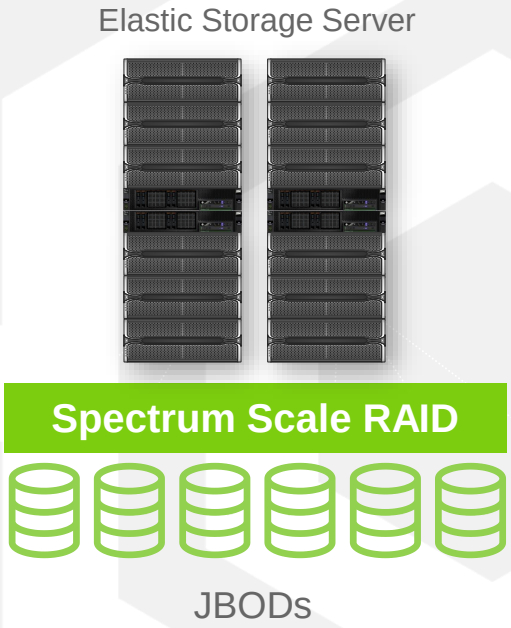
- More disks are involved during rebuild
- Approx. 3.5 times faster than RAID-5

Minimal impact of rebuild on system performance

- Rebuild is done by many disks
- Rebuilds can be deferred with sufficient protection

Better fault tolerance

- End to end checksum
- Much higher mean-time-to-data-loss (MTTDL)
  - 8+2P: ~ 200 Years
  - 8+3P: ~ 200 Million Years



Getting started

Do something today

Schedule remote Proof of Technology Lab
– *Three global labs with deep expertise*

Experience virtual machine demonstration
– *Download & run on your systems for POC*

Spectrum Scale to the Rescue!

Add management, performance and scalability to existing storage

Start Smart!

Anticipate data growth and flexibility

- *HDFS & Big Data Analytics*
- *Private Cloud*
- *Object Storage*

ibm.com/systems/storage/**spectrum**/**scale**/

Store **Everywhere.** Run **Anywhere.**

Legal notices

Product data has been reviewed for accuracy as of the date of initial publication. Product data is subject to change without notice. This document could include technical inaccuracies or typographical errors. IBM may make improvements and/or changes in the product(s) and/or program(s) described herein at any time without notice. Any statements regarding IBM's future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only. References in this document to IBM products, programs, or services does not imply that IBM intends to make such products, programs or services available in all countries in which IBM operates or does business. Any reference to an IBM Program Product in this document is not intended to state or imply that only that program product may be used. Any functionally equivalent program, that does not infringe IBM's intellectually property rights, may be used instead.

THE INFORMATION PROVIDED IN THIS DOCUMENT IS DISTRIBUTED "AS IS" WITHOUT ANY WARRANTY, EITHER OR IMPLIED. IBM LY DISCLAIMS ANY WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE OR NONINFRINGEMENT. IBM shall have no responsibility to update this information. IBM products are warranted, if at all, according to the terms and conditions of the agreements (e.g., IBM Customer Agreement, Statement of Limited Warranty, International Program License Agreement, etc.) under which they are provided. Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products in connection with this publication and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. IBM makes no representations or warranties, ed or implied, regarding non-IBM products and services.

The provision of the information contained herein is not intended to, and does not, grant any right or license under any IBM patents or copyrights. Inquiries regarding patent or copyright licenses should be made, in writing, to:

IBM Director of Licensing
IBM Corporation
North Castle Drive
Armonk, NY 1 0504- 785
U.S.A.

Information and trademarks

Notes:

Performance is in Internal Throughput Rate (ITR) ratio based on measurements and projections using standard IBM benchmarks in a controlled environment. The actual throughput that any user will experience will vary depending upon considerations such as the amount of multiprogramming in the user's job stream, the I/O configuration, the storage configuration, and the workload processed. Therefore, no assurance can be given that an individual user will achieve throughput improvements equivalent to the performance ratios stated here.

All customer examples cited or described in this presentation are presented as illustrations of the manner in which some customers have used IBM products and the results they may have achieved. Actual environmental costs and performance characteristics will vary depending on individual customer configurations and conditions.

This publication was produced in the United States. IBM may not offer the products, services or features discussed in this document in other countries, and the information may be subject to change without notice. Consult your local IBM business contact for information on the product or services available in your area.

All statements regarding IBM's future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only.

Information about non-IBM products is obtained from the manufacturers of those products or their published announcements. IBM has not tested those products and cannot confirm the performance, compatibility, or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

Prices subject to change without notice. Contact your IBM representative or Business Partner for the most current pricing in your geography.

This presentation and the claims outlined in it were reviewed for compliance with US law. Adaptations of these claims for use in other geographies must be reviewed by the local country counsel for compliance with local laws.

Special notices

This document was developed for IBM offerings in the United States as of the date of publication. IBM may not make these offerings available in other countries, and the information is subject to change without notice. Consult your local IBM business contact for information on the IBM offerings available in your area.

Information in this document concerning non-IBM products was obtained from the suppliers of these products or other public sources. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

IBM may have patents or pending patent applications covering subject matter in this document. The furnishing of this document does not give you any license to these patents. Send license inquires, in writing, to IBM Director of Licensing, IBM Corporation, New Castle Drive, Armonk, NY 10504-1785 USA.

All statements regarding IBM future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only.

The information contained in this document has not been submitted to any formal IBM test and is provided "AS IS" with no warranties or guarantees either expressed or implied.

All examples cited or described in this document are presented as illustrations of the manner in which some IBM products can be used and the results that may be achieved. Actual environmental costs and performance characteristics will vary depending on individual client configurations and conditions.

IBM Global Financing offerings are provided through IBM Credit Corporation in the United States and other IBM subsidiaries and divisions worldwide to qualified commercial and government clients. Rates are based on a client's credit rating, financing terms, offering type, equipment type and options, and may vary by country. Other restrictions may apply. Rates and offerings are subject to change, extension or withdrawal without notice.

IBM is not responsible for printing errors in this document that result in pricing or information inaccuracies.

All prices shown are IBM's United States suggested list prices and are subject to change without notice; reseller prices may vary.

IBM hardware products are manufactured from new parts, or new and serviceable used parts. Regardless, our warranty terms apply.

Any performance data contained in this document was determined in a controlled environment. Actual results may vary significantly and are dependent on many factors including system hardware configuration and software design and configuration. Some measurements quoted in this document may have been made on development-level systems. There is no guarantee these measurements will be the same on generally-available systems. Some measurements quoted in this document may have been estimated through extrapolation. Users of this document should verify the applicable data for their specific environment.

DCP03213USEN-06