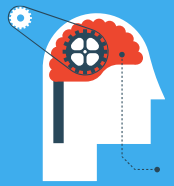
Spectrum Scale Expert Talks

Episode 16:

What is new in Spectrum Scale 5.1.2?

Show notes:
www.spectrumscaleug.org/experttalks

Join our conversation:
www.spectrumscaleug.org/join

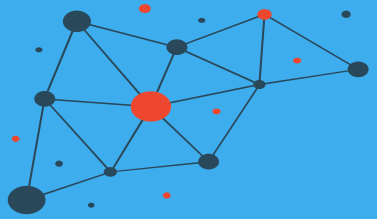


SSUG::Digital

Welcome to digital events!

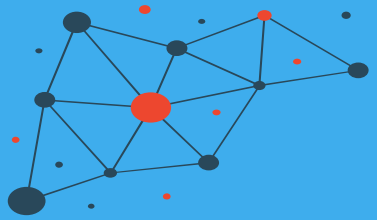
Show notes:
www.spectrumscaleug.org/experttalks

Join our conversation:
www.spectrumscaleug.org/join



About the user group

- Independent, work with IBM to develop events
- Not a replacement for PMR!
- Email and Slack community
- https://www.spectrumscaleug.org/join



We are …

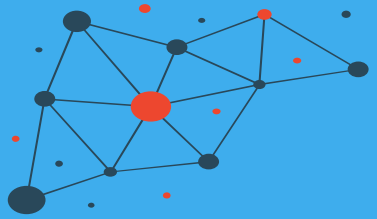
Current User Group Leads

- Paul Tomlinson (UK)
- Kristy Kallback-Rose (USA)
- Bob Oesterlin (USA)

Former User Group Leads

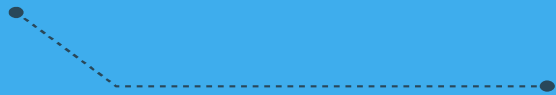
- Simon Thompson (UK)
- Bill Anderson (USA)
- Chris Schlipalius (Australia)

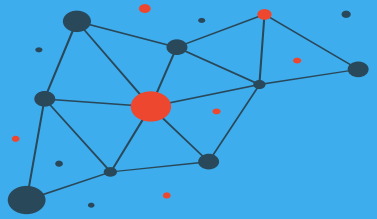
IBM CHAMPION



Check https://www.spectrumscaleug.org/experttalks for charts, show notes and upcoming talks

- Past talks:
  - 001: What is new in Spectrum Scale 5.0.5?
  - 002: Best practices for building a stretched cluster
  - 003: Strategy update
  - 004: Update on performance enhancements in Spectrum Scale (file create, MMAP, direct IO, ESS 5000)
  - 005: Update on functional enhancements in Spectrum Scale (inode management, vCPU scaling, NUMA considerations)
  - 006: Persistent Storage for Kubernetes and OpenShift environments
  - 007: Manage the lifecycle of your files using the policy engine
  - 008: Multi-node scaling of AI workloads using Nvidia DGX, OpenShift and Spectrum Scale
  - 009: Continental: Deep Thought – An AI Project for Autonomous Driving Development
  - 010: Data Accelerator for Analytics and AI (DAAA)
  - 011: What is new in Spectrum Scale 5.1.0?
  - 012: Lenovo - Spectrum Scale and NVMe Storage
  - 013:Event driven data management and security using Spectrum Scale Clustered Watch Folder and File Audit Logging
  - 014: What is new in Spectrum Scale 5.1.1?
  - 015: IBM Spectrum Scale Container Native Storage Access





Speakers

- Chris Maestas (IBM)
- Loads of thanks to Nikhil Khandelwal!!!! (IBM)

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.

Featured Updates

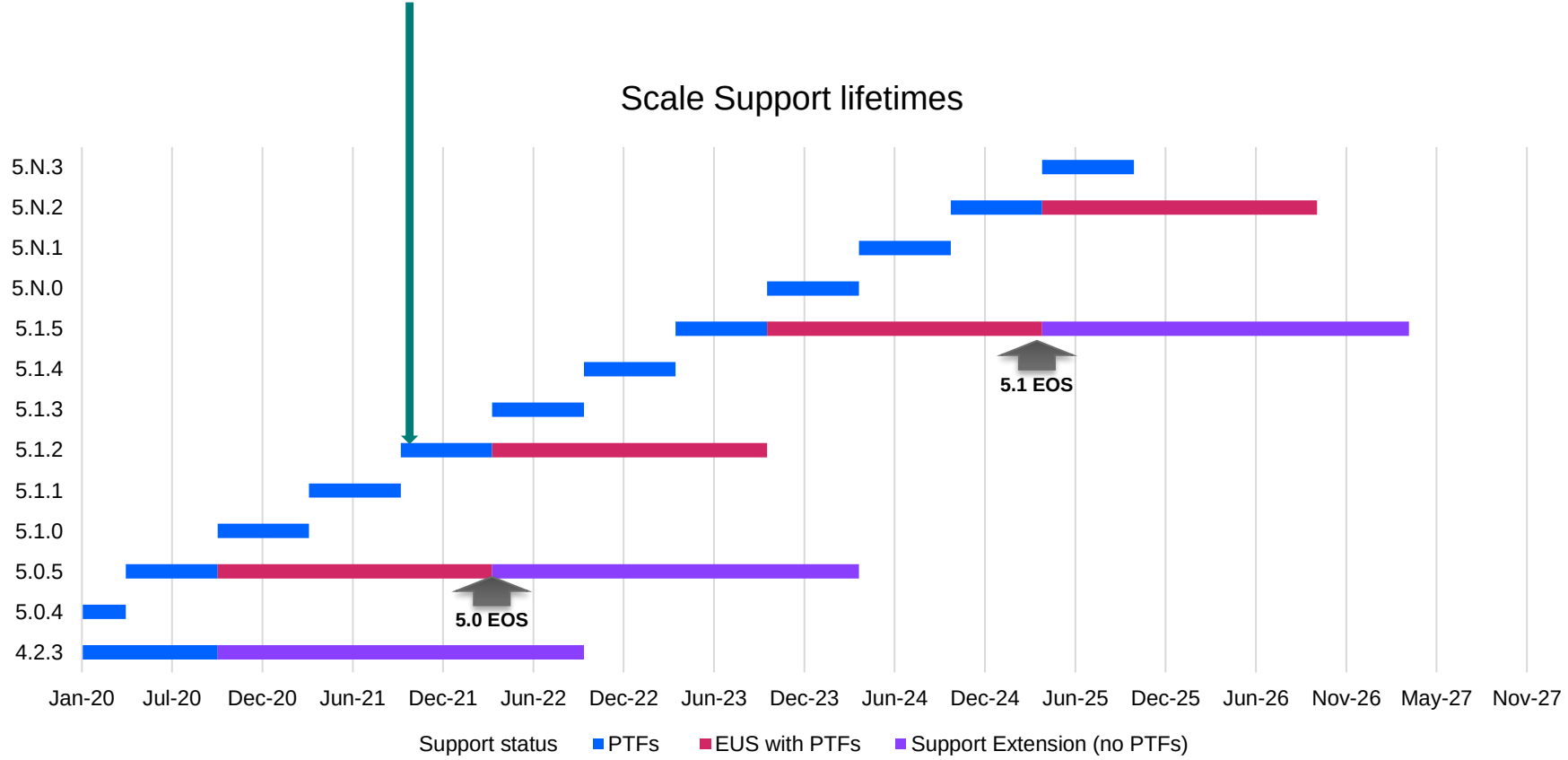
Continued enhancement to integration and deployment features with containers that now allows containerized storage deployment within Spectrum Fusion technology.

High Performance Object (HPO) support as part of next generation protocol services stack.

GPU Direct Storage (GDS) general availability

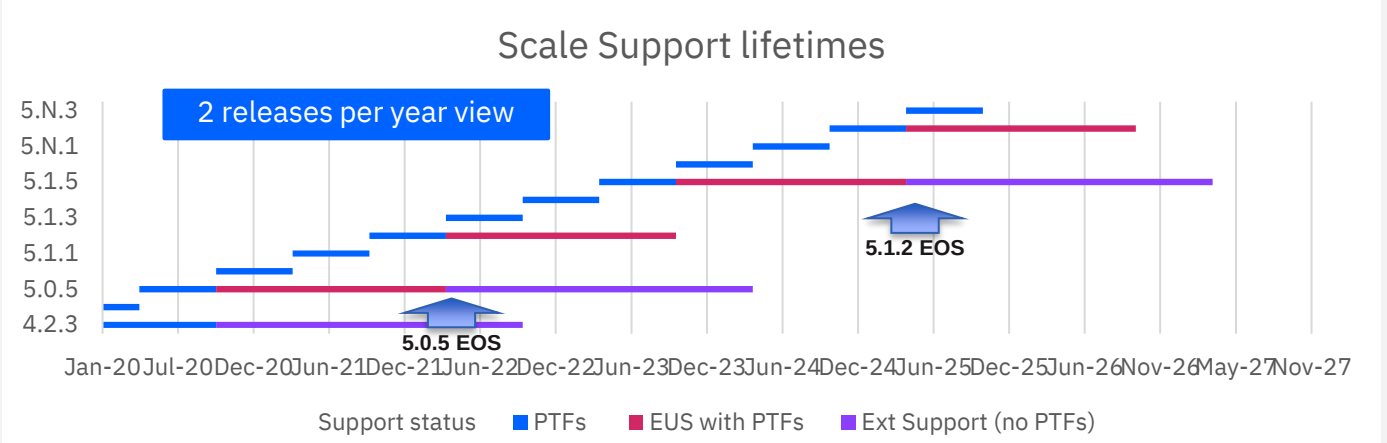
Enhanced reliability, monitoring and security with Active File Management use cases in Object and Disaster Recovery scenarios.

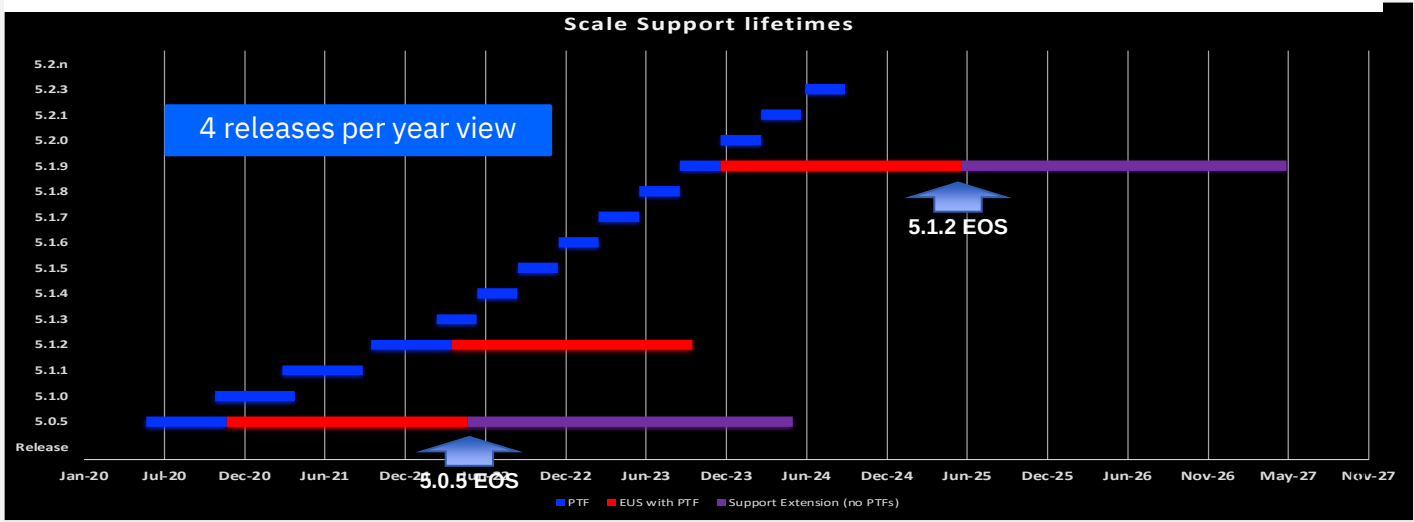
Spectrum Scale: Release Plan



Shift to Quarterly Release Cadence

- **Why?**
  - To address request for quarterly updates to bring new features out sooner.
- **Maintain Extended Update Support concept**
  - EUS with PTFs every 18 months
  - Extended support on last EUS within a release (example: V.R.x, 4.2.3, 5.1.2, 5.1.last)
  - Increase the number of Modification levels with new function





Container Native Storage Access

Spectrum Scale

Deploy Spectrum Scale on containers.

IBM Spectrum Scale container native storage access (CNSA) supports the Red Hat OpenShift Container Platform with a fully containerized deployment.

This allows deployment of IBM Spectrum Scale on Red Hat CoreOS worker nodes where classic packages cannot be installed.

Container Native Storage Access

Spectrum Scale

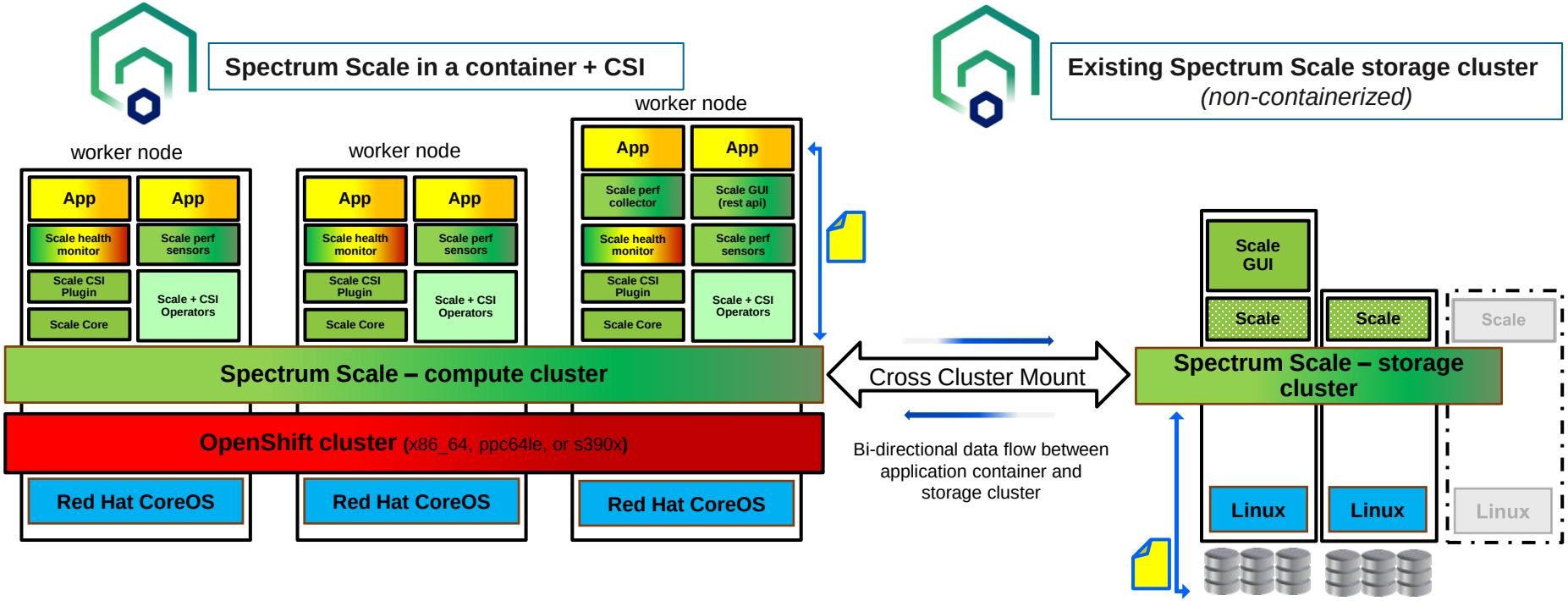
Improvements introduced in CNSA 5.1.2.1

Wider support to use the latest CNSA functionality.

- Support for upgrading IBM Spectrum Scale Container Native Storage Access (CNSA) from v5.1.1.4 to 5.1.2.1
- Support for RedHat OpenShift Container Platform 4.[7,8,9]
- CNSA images now hosted on the entitled IBM Cloud Container Registry.
- Automated deployment of the CSI driver
- Support for storage cluster encryption
- Rolling upgrade of IBM Spectrum Scale is supported
- Support for a limited set of IBM Spectrum Scale configuration settings to be set directly
- Grafana support
- Support for X86, Power and Z.
- Direct storage attachment on x86 and power servers.
- Automatic quorum selection is Kubernetes topology aware.

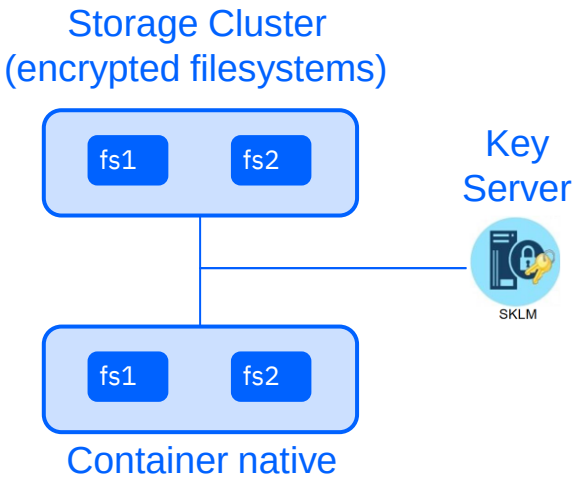
IBM Spectrum Scale Container Native Storage Access (CNSA)

Cluster Overview



CNSA: Support for storage cluster encryption

- To **access encrypted data from a remote mounted filesystem of a storage cluster**, the encryption key must be present in the container native cluster.
- To access the encryption key from the key-server, the key-client must be configured in the cluster.
- To configure **key-server** and **key-client**, one needs to create an **EncryptionConfig custom resource.**
- Key servers with CA signed server certificates are supported.
- The encryption controller applies the configuration of the encryption CR.
- Using multiple key servers is supported (create one CR per key server).



AFM, Compression, ILM for Containerized environments

- AFM for Caching or DR
  - Setup AFM gateway nodes in the external storage cluster for Caching or DR use case
  - No configuration required on the Scale compute cluster
- Compression
  - Run compression policies in the external storage cluster
  - Scale compute cluster will read compressed data by uncompressing it on the clients
- Storage Tiering
  - Define Storage pools, apply placement, migration policies in external cluster
  - No way to define placement policies from Scale compute cluster

Container Storage Interface

Spectrum Scale

Improvements introduced in CSI ~~2.3.1~~ 2.4.0

Upgrades for OpenShift, Kubernetes and Ansible as well as improved functionality that support simpler administration and configuration.

- Support for Red Hat OpenShift 4.[7,8,9] and Kubernetes 1.20, 1.21, 1.22.

- CSI Volume Cloning which provides ability to clone a persistent volume.
- CSI Volume Expansion which provides support for volume expansion after volume creation.
  - requires IBM Spectrum Scale v5.1.1.2 or higher.

- Support for Kubernetes subpath using the new parameter “permissions”

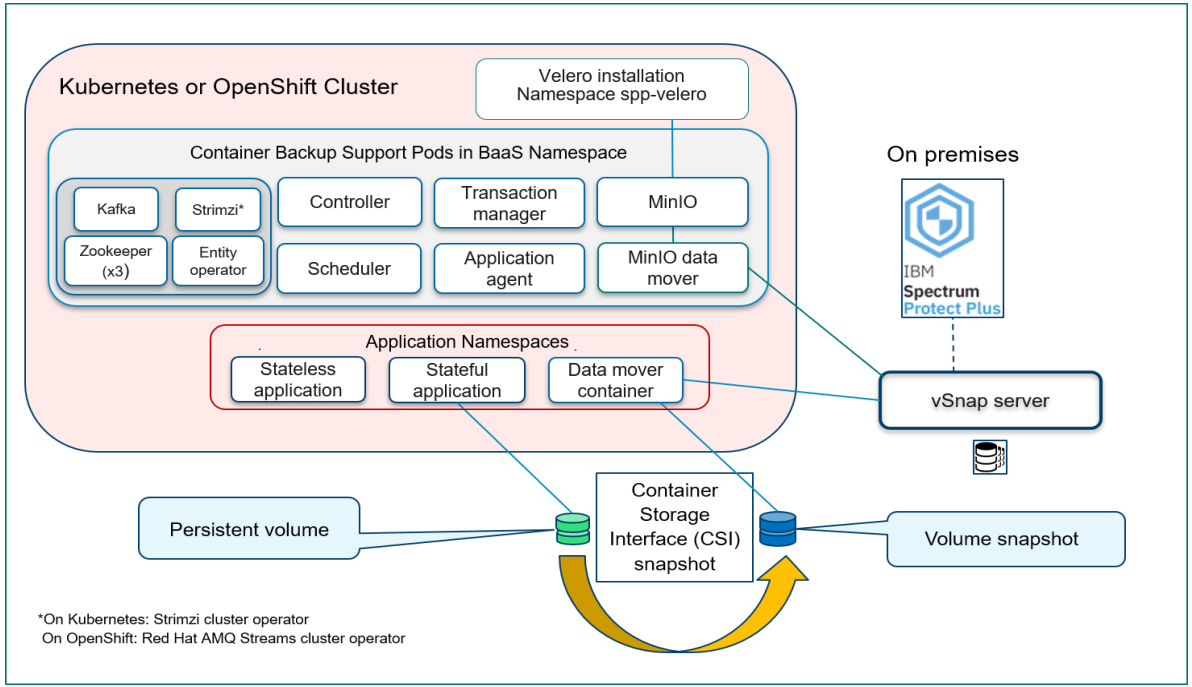
- Usability improvements for small volume and inode management
  - inode calculation updates for independent fileset based volumes
    - 100K inodes for <= 10 GB and 200K for > 10 GB volume size

- Support for liveliness probe in CSI driver pods

- Upgraded the ansible-operator SDK version from 1.5.2 to 1.13.0 to get the latest security and stability updates

Container Backup and Restore Support

- To protect volume data, IBM Spectrum Protect Plus (SPP) provides Container Backup Support for IBM Spectrum Scale using CSI.
- For backups, IBM Spectrum Protect Plus uses CSI for taking snapshots to create a backup.
- For restores, IBM Spectrum Protect Plus uses CSI for creating a PVC where restore is done.
- IBM SPP Container backup support added support for IBM Spectrum Scale CSI driver from IBM Spectrum Protect Plus V10.1.8 onwards.



IBM SPP Container Backup Support deployment diagram

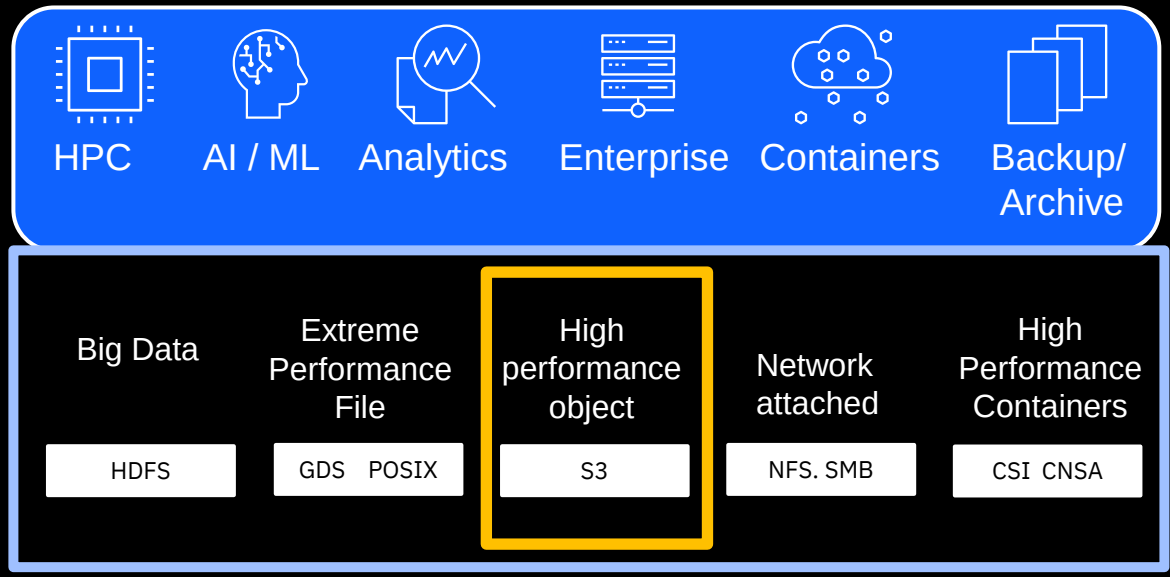
Link to blog on [IBM Spectrum Scale CSI Volume Snapshots and its usage by IBM Spectrum Protect Plus]

Ingest or access data with high performance S3 interface

- fast AI results for S3 cloud native applications
- scalable solution for ingesting high performance S3 object data from remote locations
- scale performance and capacity as needed
- container native deployment for easy OpenShift integration
- applications can now optimize with the interface they need to access all the data they require
  (example: ingest S3 and access via file)*

*Concurrent access supported 1H 2022

GB/s to TB/s performance for S3 object data



Data Access Services – S3 object access – Tech Preview

Containerized S3 object access integrated within Spectrum Scale delivering high performance object for AI and analytics workloads

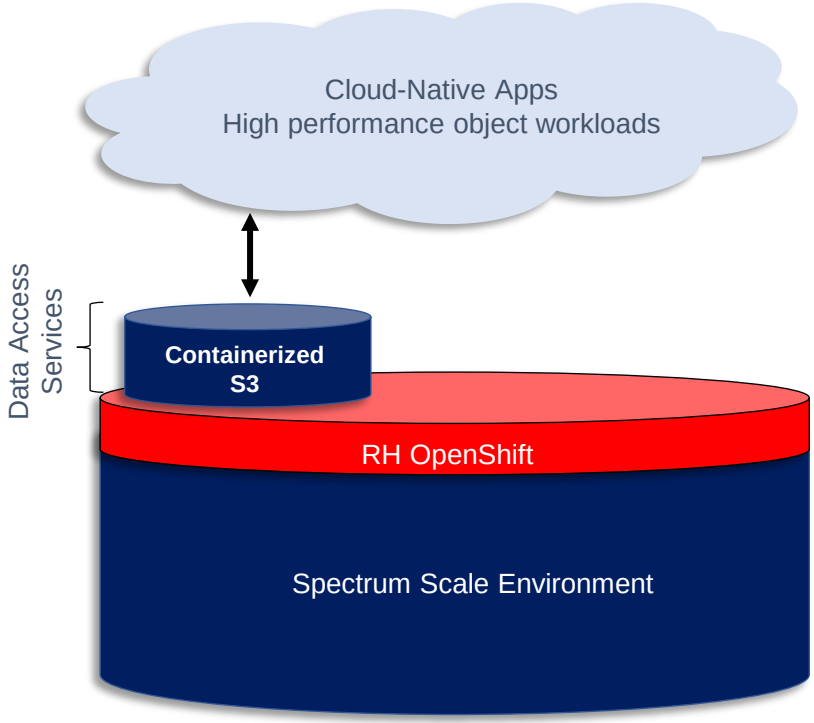
Customer Requirements & DAS Dependencies:

- Spectrum Scale 5.1.2.1: DAE, DME, ESS for DAE, ESS for DME, ECE (future)
- OpenShift 4.9.8
- CNSA 5.1.2.1 / CSI 2.4.0
- ESS models at GA followed by any storage (supported by CNSA)

Performance: MVP baseline 40 GB/s w/ 3 DAN (Data Access) nodes on vanilla ethernet, scales linearly, increased performance with each release

Advanced S3 Capabilities on roadmap

- S3 Encryption Headers
- S3 Select
- S3 Versioning
- S3 Object Lock



GPU Direct Storage (GDS)

Spectrum Scale

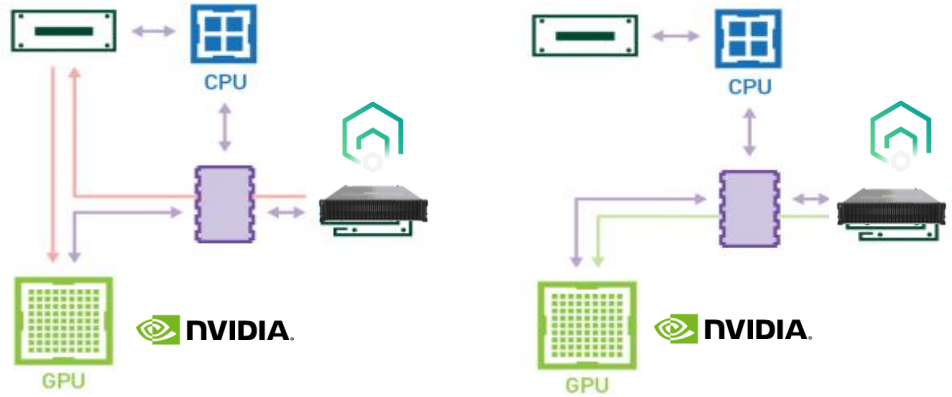
Scale with NVIDIA

Greatly reduce CPU overhead when talking to IBM Spectrum Scale from the GPU

IBM Spectrum Scale 5.1.2

GPUDirect Storage (GDS) enables an NVIDIA developer to:

- Make a direct memory access (DMA) between GPU memory and storage
- Bypass the CPU and system memory
- Reduce latency, Increase bandwidth, and lower CPU utilization for a specific read from storage



* Requires CUDA, provided by Nvidia, to be installed on the Spectrum Scale client

* GDS Read Operations are fully supported

* Writes (cuFileWrite) are supported in compatibility mode for good path scenarios only. Failure scenarios for the writes in compatibility mode are available as a Technical Preview feature (to be lifted in an upcoming PTF).

GPU Direct Storage (GDS)

Spectrum Scale

Scale with NVIDIA

Understand how to get GDS and the requirements.

Spectrum Scale Knowledge Center:
https://www.ibm.com/docs/en/spectrum-scale/5.1.2?topic=summary-changes
https://www.ibm.com/docs/en/spectrum-scale/5.1.2?topic=architecture-gpudirect-storage-support-spectrum-scale

Nvidia GDS Documentation:
https://docs.nvidia.com/gpudirect-storage/index.html
https://developer.nvidia.com/gpudirect-storage

For help getting started: scale@us.ibm.com

Which GDS Release?

- CUDA 11.4 or later

Supported Storage

- Spectrum Scale 5.1.2
- ESS or any NSD client-server storage model

Supported Network

- Infiniband (RDMA)
- MOFED 5.4-1.0.3.0

GPUs and OS

- NVIDIA Ampere (e.g. NVIDIA A100)
- RHEL 8.3, Ubuntu 20.04

Big Data & Analytics

Spectrum Scale

Extended support:

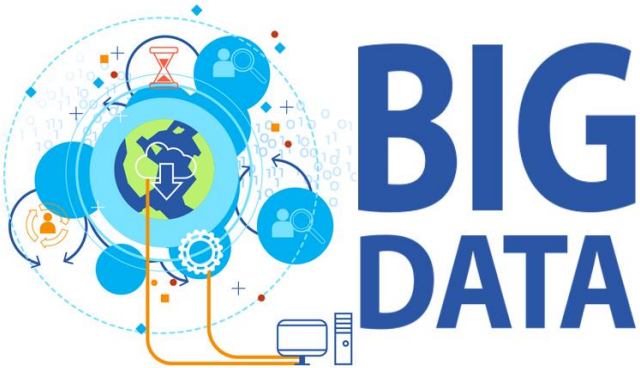
- Cloudera Data Platform (CDP) Private Cloud Base is certified with IBM Spectrum Scale on x86_64 and ppc64le
- CDP Private Cloud Base was 1st certified on IBM Spectrum Scale 5.1.0.1 with HDFS Transparency 3.1.1-3 in December 2020

Simplified automated deployment:

- Ansible toolkit deployment for CES HDFS in Scale 5.1.2

Improved performance:

- HDFS Transparency 3.1.1-6 implements performance enhancements in metadata for the NameNode.
- Optimized parallelism for DataNode request processing.
  - See the KnowledgeCenter for DataNode performance Options.



Big Data & Analytics

Spectrum Scale

Enhanced security and reliability for HDFS Transparency:

- Version 3.1.1-6 in IBM Spectrum Scale 5.1.2 contains RPC enhancements and contains fixes for metadata handling when listing files would not show the correct creation time

Enhanced support and management with Mpack 2.7.0.10

- Service can now be deployed or upgraded in a single or multiple HDFS namespace configuration. Includes adding DataNode using Ambari in multiple HDFS namespaces.
- Decommissioning DataNodes using the Ambari HDFS service is now supported.
- Service can now be deployed in Ambari in remote cluster mount configuration for non-root Ambari and IBM Spectrum Scale environment.
- The MoveNameNodeTransparency.py script now supports moving the HDFS Transparency NameNode when Kerberos is enabled.

AFM Performance Updates

Spectrum Scale

AFM resync version 2 improves replication on heavily stressed systems

- Updates to message queuing to improve AFM resync and recovery
- Lower memory usage on gateway nodes with faster replication
- Faster recovery and resync after gateway node failures
- Improved role reversal in AFM-DR
- See ‘AFM resync version 2’ in the knowledge center, and the afmResyncVer2 parameter in mmchfileset to activate this feature

NFSv4 support for AFM replication

- Allows NFSv4 to be used as the underlying AFM protocol for data transfer
- Supports NFSv4 ACLs on third-party storage to support data migration

New queries and statistics available in mmafmctl

- Query uncached and dirty files, and new read and write statistics
- See mmafmctl man page for more details

Spectrum Scale Core Improvements

Spectrum Scale

Administration and reliability

Simpler and more flexible administration that allows you better control.

- **mmlsnsd** update for LROC size reporting
- mmcrfs –metadata-block-size is being deprecated
  - No longer required for performance with file system format 5.0.0.
- **mmgetstate --fqdn** to display the fully qualified domain name of nodes
- Clear QoS configuration automatically when filesystem objects (filesets, nodes, node classes, remote clusters, and pools) are deleted

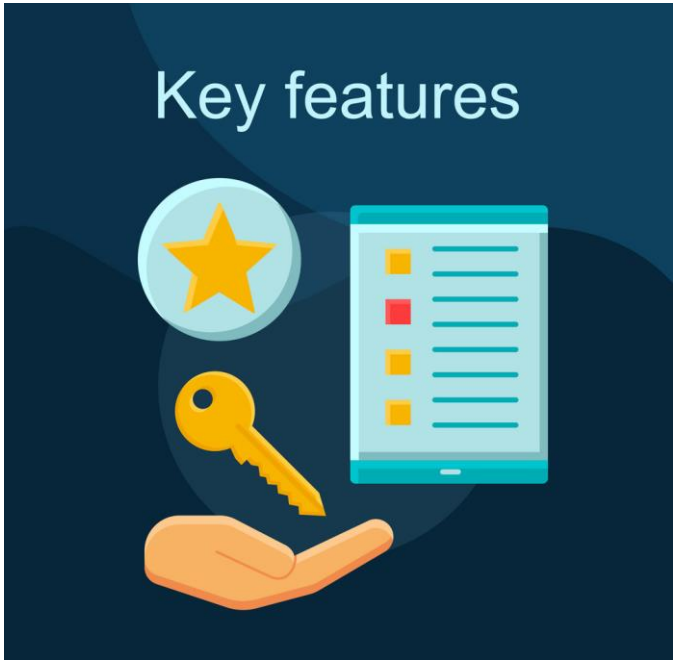
Spectrum Scale Core Improvements

Spectrum Scale

Features

Features that allow you to improve your resource utilization and problem determination.

- New command: **mmxcp** performs parallel copies of files within a cluster. Copies are distributed across multiple nodes, and can be used to copy data between 2 file systems, within a file system, or from a snapshot to a file system *[including ACLs and file attributes!]*
- New *diskDown* callback triggered when a disk goes down. This can be used to monitor disks and provides the reason a disk has changed its status.
- *Windows* support for *4k disk sectors*. This removes the restriction that existed in previous versions of Spectrum Scale for Windows
- IBM Spectrum Scale now supports the use of *O_TMPFILE* on opens with Linux kernel 3.15 or greater



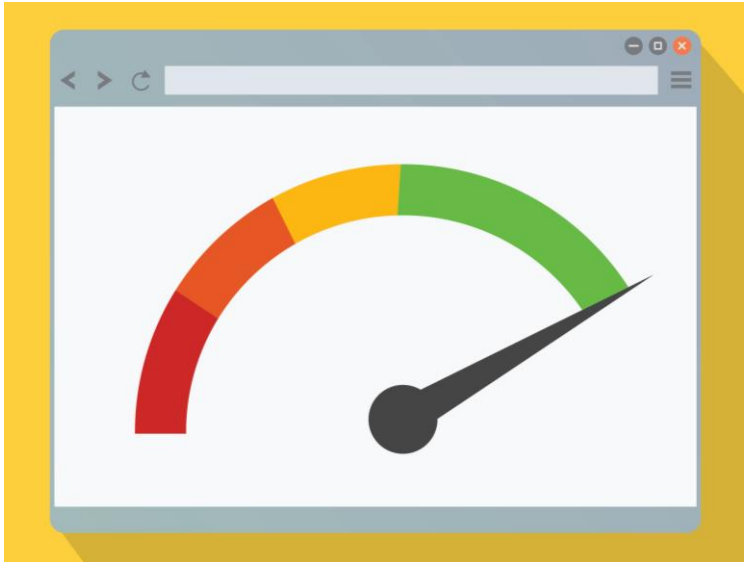
Spectrum Scale Core Improvements

Spectrum Scale

Performance Improvements

File system utilities and network enhancements that allow you to maximize your system utilization.

- *NUMA* aware *RDMA adapter* selection and thread affinity. The new **verbsNumaAffinity** parameter enables NSD servers to select an adapter with NUMA awareness
- Faster file creations by reducing context switching during inode opens.
- Direct IO vectored read improvements with more efficient locking
- Concurrent hard link creation improvements by enabling fine-grained directory locking for hard link creation
- **mmap** performance improvements when pre-fetching file data blocks



Spectrum Scale Core Improvements

Spectrum Scale

Scale on Z Systems

Better resource allocation and support for NFS clients

- Spectrum Scale ECE support for NVMe drives on the IBM z15 server
  - Supports up to 4 storage nodes in native LPARs, each with up to 3 NVMe drives
  - HiperSocket network support
- Thin provisioning support on DS8000
- Spectrum Scale for Openshift supports direct attach of FCP and ECKD storage to workers with cross-cluster mounts

Security Improvements

Spectrum Scale

- Support for IBM Security Guardium Key Lifecycle Manager (GKLM, formerly SKLM) 4.1.0.1
- Performance Monitor tool improvements to limit connections to known cluster node IPs
  - Can be enabled by modifying the *ipfixiplist* and *ipfixiplistrefresh* attributes in the */opt/IBM/zimon/ZIMonCollector.cfg* configuration file.
- Changing file system name is restricted if file audit logging or clustered watch folder is enabled not use the **mmchfs** command to change the file system name if file audit logging or clustered watch folder is enabled for the file system

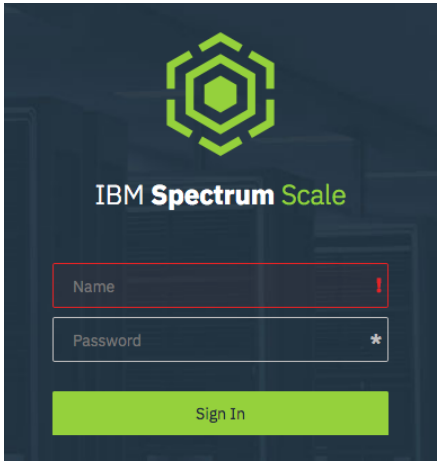
Management GUI/API Changes

Spectrum Scale

Administration and reliability

Simpler management.

- Enhanced AFM performance metrics
  - Total requests written to and read from remote systems can be viewed and sorted in the GUI performance metric table on the Active File Management page.
- Enhanced Management API encryption endpoints. Better management of tenants and encryption keys through the management API



Monitoring, Availability & Proactive Services (MAPS) Updates

Spectrum Scale

System Health & Monitoring

Enhanced awareness on the status of your system components

•New components and performance counters to support GPUDirect Storage

•New proactive system health HEALTHCHECK alerts to provide direct feedback from IBM support.

- •Scheduled call home data is monitored for security or configuration issues, and feedback is sent directly to mmhealth

•New API sub-commands in mmperfmon delete to delete expired API keys for ease of management

•New commands added to scheduled data upload list for call home. See knowledge center for details

Spectrum Scale Erasure Code Edition Changes

Spectrum Scale

Simpler installation and enhanced mechanisms for cluster management.

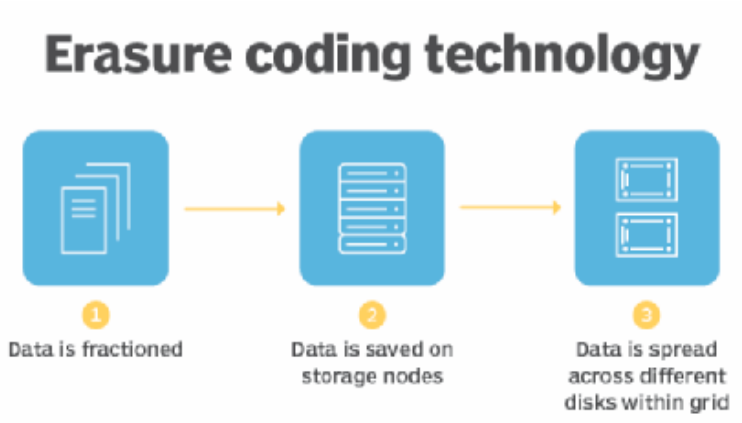
GPFS Native RAID (GNR) metadata flushing performance enhancements. Enhanced monitoring to find bottlenecks, better thread management for increased performance

Larger log home *vdisks* (default from 2 GiB to 32 Gib) to allow more time to flush GNR metadata.

mmvdisk suspend node enhancements to avoid I/O pauses while suspending ECE nodes

Additional VIO metrics in performance monitor for enhanced trouble shooting and management

Enablement for larger recovery groups and drive configurations



Discontinued Features

https://www.ibm.com/docs/en/STXKQY/pdf/scale_deprecated_features.pdf

Category	Discontinued functionality	Recommended Action
Security	The use of TLS 1.0 and 1.1 for authorization within and between IBM Spectrum Scale clusters.	Upgrade to TLS 1.2 or later.
GUI/REST API	The use of TLS 1.0 and 1.1 for authorization with the GUI/REST API server	Upgrade to TLS 1.2 or later.
Platforms	Encryption acceleration library for Power7 (CLIC)	If encryption performance is critical, migrate to newer generations of Power Systems .
	Big Endian Power servers	Upgrade to newer generations of Power systems or remain on IBM Spectrum Scale 5.1.2.
	Linux support for IBM Power7 systems	Plan to migrate to newer generations of Power systems.
Protocols	iSCSI as a target for remote boot	Use some other block services provider.
Containers	Storage Enabler for Containers (SEC)	Migrate to Container Storage Interface (CSI).

Deprecated Features

https://www.ibm.com/docs/en/STXKQY/pdf/scale_deprecated_features.pdf

Category	Deprecated functionality	Recommended Action
Platforms	AIX support for IBM Power7 systems	Plan to migrate to newer generations of Power systems.
Watch folders	Kafka queue for watch folders	If you are using Kafka for other purposes, install it separately from IBM Spectrum Scale. It is no longer installed as part of the IBM Spectrum Scale installation.
Security	Support for Vormetric DSM V5	Upgrade to Vormetric DSM V6.2 or later
Protocols	mmcesdr command (Protocols cluster disaster recovery)	Use AFM and AFM DR to set up your own replication strategies between clusters.
Cluster configuration	The primary and secondary configuration server functionality. Instead of this, clusters must use CCR.	The default configuration service is CCR, and new clusters are created using CCR. If not yet operating with CCR, change to that mode with mmchcluster --ccr-enable.
Block size	The --metadata-block-size option of mmcrfs command is deprecated. This option is used for defining metadata blocks to a different size than the data blocks.	Only a single definition for the number of subblocks per block exists per file system. Selecting a smaller metadata block size has the unintended side effect of increasing the subblock size for data blocks. Although it is supported to set metadata blocks to a different size than data blocks by using the --metadatablock-size parameter, it is not recommended to use that option. This

Stabilized Features

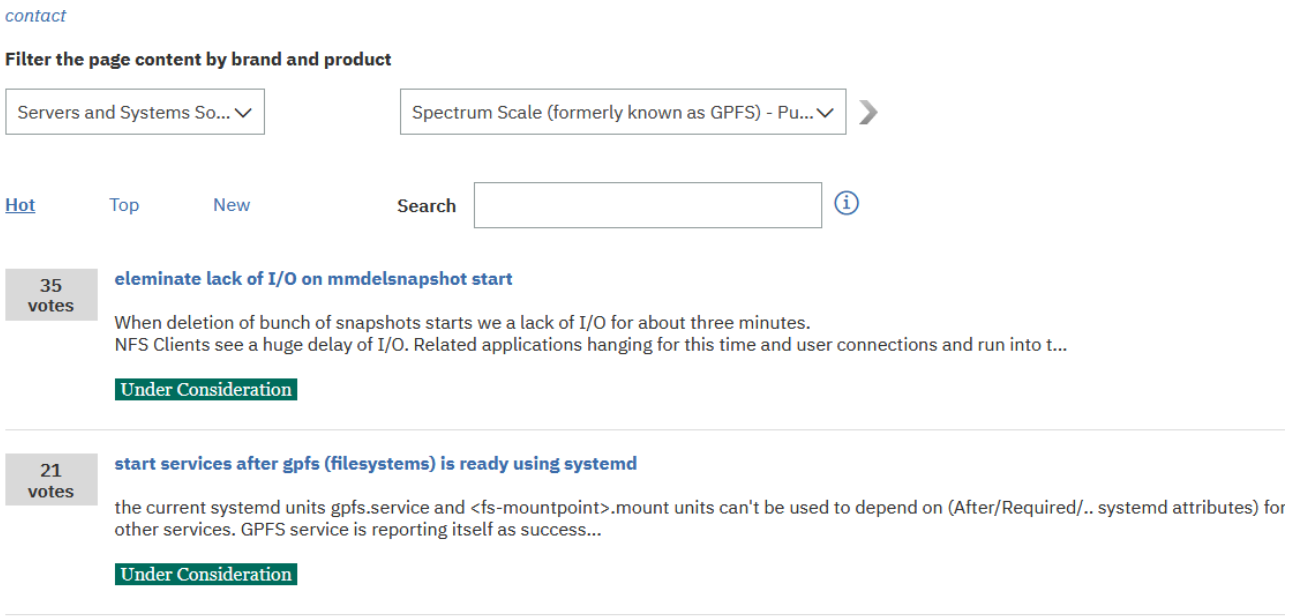
https://www.ibm.com/docs/en/STXKQY/pdf/scale_deprecated_features.pdf

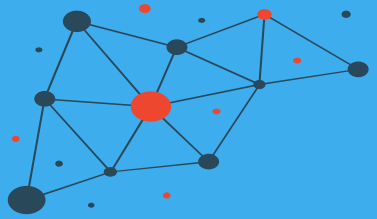
Category	Stabilized functionality	Recommended Action
cNFS	The use of TLS 1.0 and 1.1 for authorization within and between IBM Spectrum Scale clusters.	IBM®'s strategic path is to invest in User Space solutions for NFS support of Scale workloads. Once User Space performance and function are considered to be sufficient to replace cNFS, anticipate that the support for cNFS is deprecated.

Log your RFE!

https://www.ibm.com/developerworks/rfe/execute?use_case=productsList

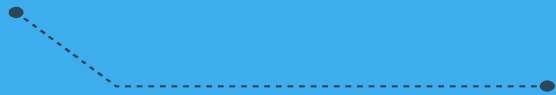
- Spectrum Scale (formerly known as GPFS) - Private RFEs
- Spectrum Scale (formerly known as GPFS) - Public RFEs





Check https://www.spectrumscaleug.org/experttalks for charts, show notes and upcoming talks

- Past talks:
  - 001: What is new in Spectrum Scale 5.0.5?
  - 002: Best practices for building a stretched cluster
  - 003: Strategy update
  - 004: Update on performance enhancements in Spectrum Scale (file create, MMAP, direct IO, ESS 5000)
  - 005: Update on functional enhancements in Spectrum Scale (inode management, vCPU scaling, NUMA considerations)
  - 006: Persistent Storage for Kubernetes and OpenShift environments
  - 007: Manage the lifecycle of your files using the policy engine
  - 008: Multi-node scaling of AI workloads using Nvidia DGX, OpenShift and Spectrum Scale
  - 009: Continental: Deep Thought – An AI Project for Autonomous Driving Development
  - 010: Data Accelerator for Analytics and AI (DAAA)
  - 011: What is new in Spectrum Scale 5.1.0?
  - 012: Lenovo - Spectrum Scale and NVMe Storage
  - 013:Event driven data management and security using Spectrum Scale Clustered Watch Folder and File Audit Logging
  - 014: What is new in Spectrum Scale 5.1.1?
  - 015: IBM Spectrum Scale Container Native Storage Access



Thank you!

Please help us to improve Spectrum Scale with your feedback

- If you get a survey in email or a popup from the GUI, please respond
- We read every single reply

Spectrum Scale User Group

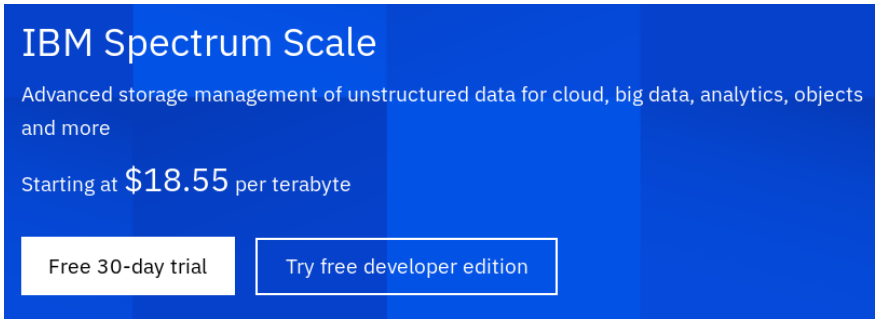
The Spectrum Scale (GPFS) User Group is free to join and open to all using, interested in using or integrating IBM Spectrum Scale.

The format of the group is as a web community with events held during the year, hosted by our members or by IBM.

See our web page for upcoming events and presentations of past events. Join our conversation via mail and Slack.

www.spectrumscaleug.org

Spectrum Scale Developer Edition!



Fully functional!

- Based on first PTF of a release
- Derived from **Data Management Edition (DME)**
- Limited to 12 TBs:
  enough for a small test cluster
- Available from the Scale ”try and buy” page on ibm.com

Free for non-production use, e.g. test, learning, upgrade prep…

- If you have to ask, it’s probably not permitted

Not formally supported

Spectrum Scale Early Programs

Spectrum Scale

Types of Programs:

Alpha

Influence the development of new technology by gaining before market access to product code. Alpha programs are typically confidential and the first opportunity for you to interact with a feature or function.

Beta

Try out a new offering with the team who owns the product and influence its usability and design. A Beta program gives you the ability to evaluate and provide feedback on IBM products before the products general availability. Beta programs are typically confidential and run prior to GA.

Early Support Program (ESP)

Be one of the few selected participants to validate new Software or Hardware and potentially give your enterprise an edge over the competition. The IBM early support programs give you and IBM the opportunity to develop, evaluate, and gain experience with a product or a set of products in your enterprise environment.

Customer Success

- ❑ Evaluate new IBM HW or SW in your environment.
- ❑ Validate procedures and interoperability with other products in your enterprise.
- ❑ Opportunity to Influence Product Design
- ❑ Early Enablement and education
- ❑ Strengthen Partnership with IBM

Talk to your IBM contact or Partner to be nominated!

Spectrum Scale on GitHub!
https://github.com/IBM/SpectrumScaleTools

- IBM Spectrum Scale Bridge for Grafana
- IBM Spectrum Scale cloud install
- IBM Spectrum Scale Container Storage Interface driver
- IBM Spectrum Scale install infra
- IBM Spectrum Scale Security Posture
- Oracle Cloud Infrastructure IBM Spectrum Scale terraform template
- SpectrumScale_ECE_CAPACITY_ESTIMATOR
- SpectrumScale_ECE_OS_OVERVIEW
- SpectrumScale_ECE_OS_READINESS
- SpectrumScale_ECE_STORAGE_READINESS
- SpectrumScale_ECE_tuned_profile
- SpectrumScale_NETWORK_READINESS

Find open source tools that are related with IBM Spectrum Scale.

Unless stated otherwise, the tools compiled in this list come with no warranty of any kind from IBM.

Check out the FAQ!

https://www.ibm.com/support/knowledgecenter/en/STXKQY/gpfsclustersfaq.html
https://www.ibm.com/support/knowledgecenter/STXKQY/gpfsclustersfaq.pdf?view=kc
https://www.ibm.com/support/knowledgecenter/SSYSP8/gnrfaq.html

HTML or PDF

Spectrum Scale version compatibility with OS or kernels

Updated regularly!