

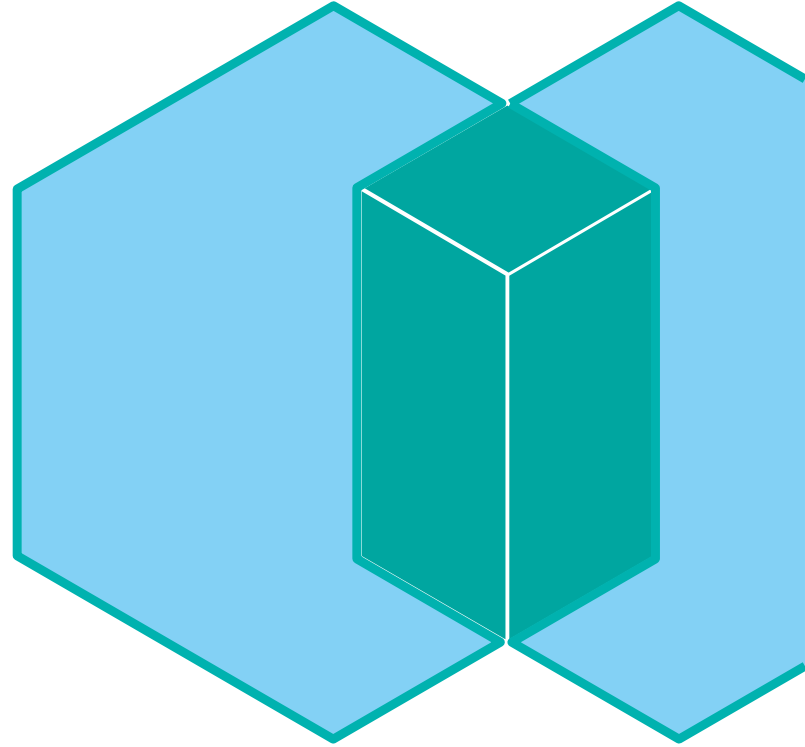


Spectrum Scale Update

- 5.01 / 5.0.2 updates
- new: “SpectrumAI”

Spectrum Scale User Group, CIUK, Manchester
12th December 2018

Dr. Daniel Kidger, IBM





Spectrum Scale User Group



SSUG Manchester 2017

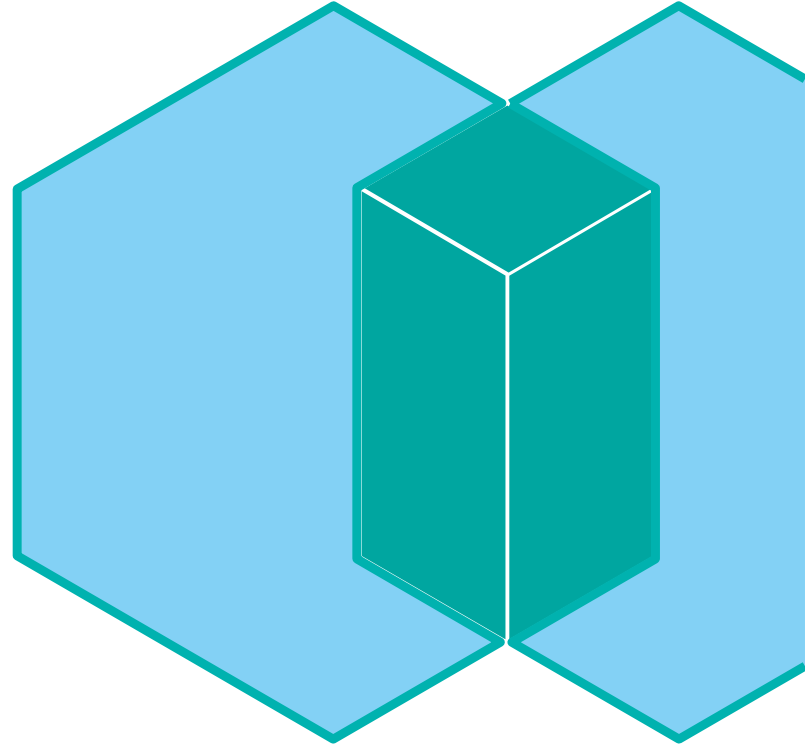




New! Spectrum AI

Integrated Spectrum Scale with
NVIDIA DGX solution

Announced: 11th Dec 2018



SpectrumAI = Spectrum Scale + Nvidia DGX

- Integrated DL / ML solution.
- Nvidia DGX-1 POD
 - 8 Tesla V100 GPGPUs
 - interconnected with NVLink
- Mellanox interconnect
 - 4 RDMA IB EDR cards per DGX
- IBM prototype all-flash 2U storage appliance*
 - 2nd generation NVMe from TMS acquisition
- Spectrum Scale
 - Embedded on the two dual-socket x86 CPUs in the NVMe array



Demonstrated to deliver industry's best performance for running modern ML/DL workloads with software-defined scalability and 4.5x more data throughput than alternatives with scalability to 120GB/s*

Readily plugs into your overall AI data pipeline with joint support, optimized container lifecycle, in-place analytics, and smart policy engine for data life cycle management.

Highlights:

- **Converged Solution:** Ready to deploy for your most demanding ML/DL projects
- **IBM Spectrum Scale v5:** Software-defined file, object and HDFS to streamline data movement through your AI data pipeline from ingest through training, inference, and archive
- **NVIDIA DGX:** Optimized to accelerate data science workflows
- **Performance:** Demonstrated 120GB/s random read throughput in 6U to support up to 9 DGX-1 POD
- **Foundational:** A shared data plane for your containerized AI workloads that extends with IBM Storage

The AI Workflow

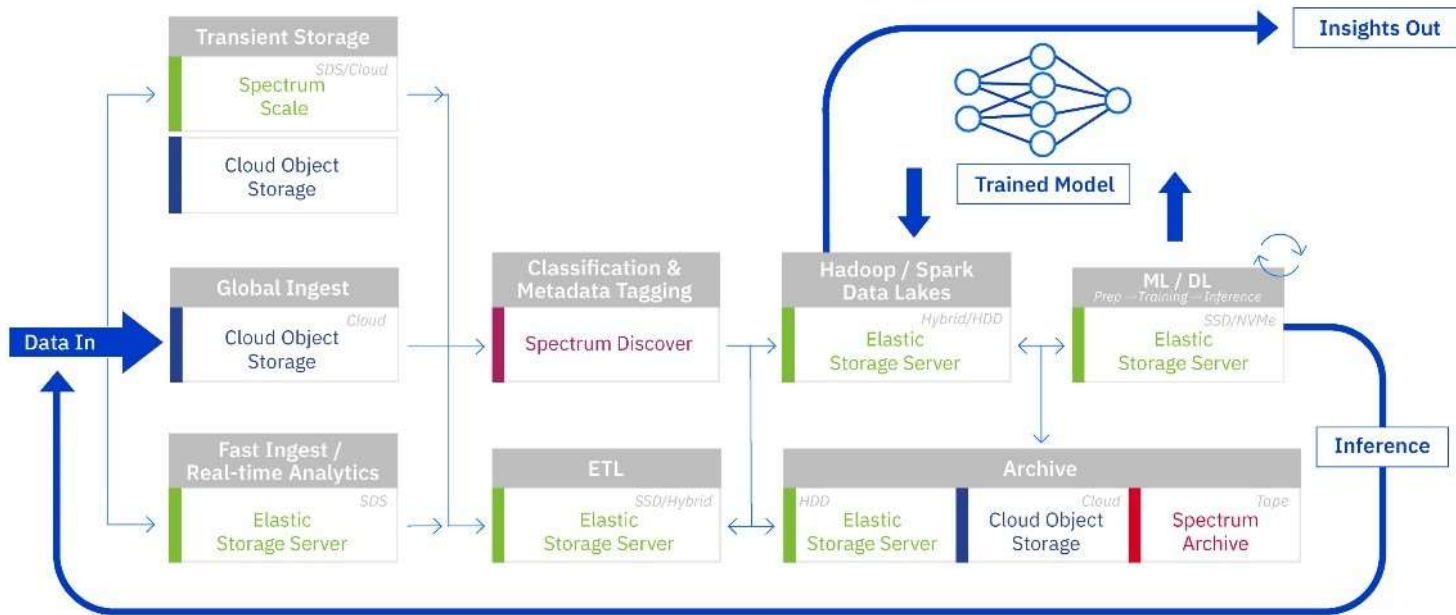
EDGE

INGEST

CLASSIFY / TRANSFORM

ANALYZE/TRAIN

INSIGHTS





40 Gbytes/s
per shelf

Spectrum Scale NVMe all-flash appliance

New IBM 19.2TB NVMe
Flashcore Module



Original Flashcore form factor

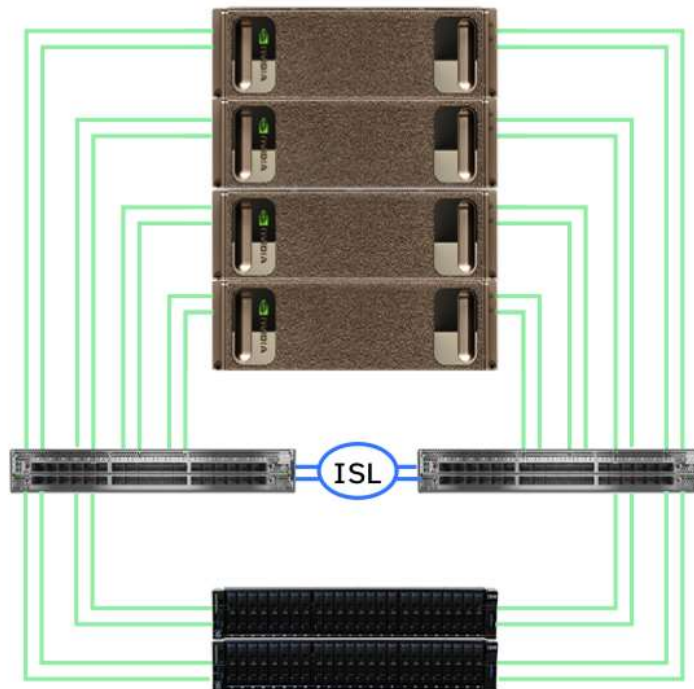
Topology

1. Multiple DGX PODs
2. Dual EDR Infiniband
3. IBM NVMe + embedded Spectrum Scale appliances

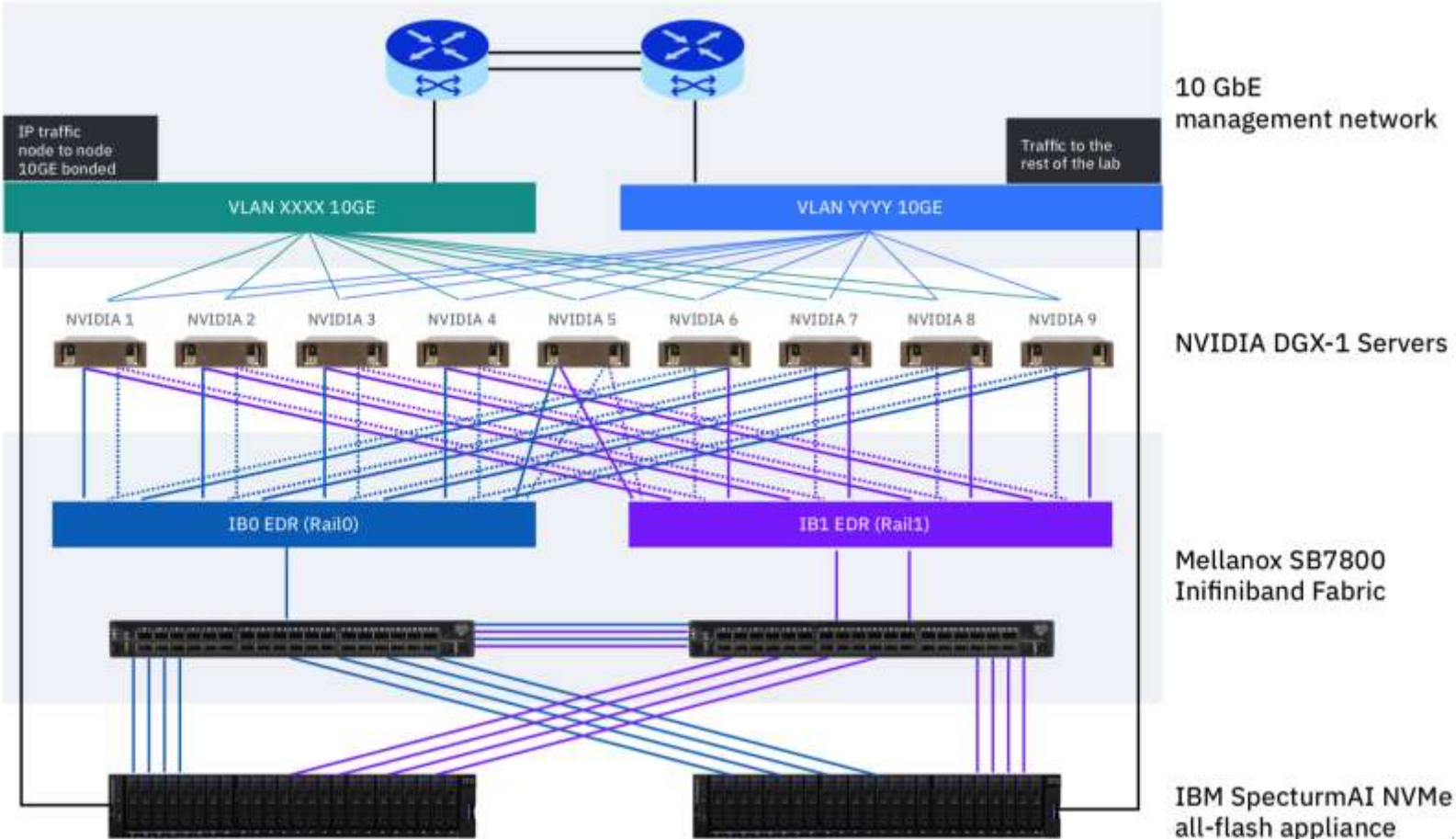
1

2

3



Reference Topology used in Benchmark Tests



10 GbE management network

NVIDIA DGX-1 Servers

Mellanox SB7800 Infiniband Fabric

IBM SpecturmAI NVMe all-flash appliance

Potential configurations



3:1 Configuration



6:2 Configuration



9:3 Configuration

Throughput performance

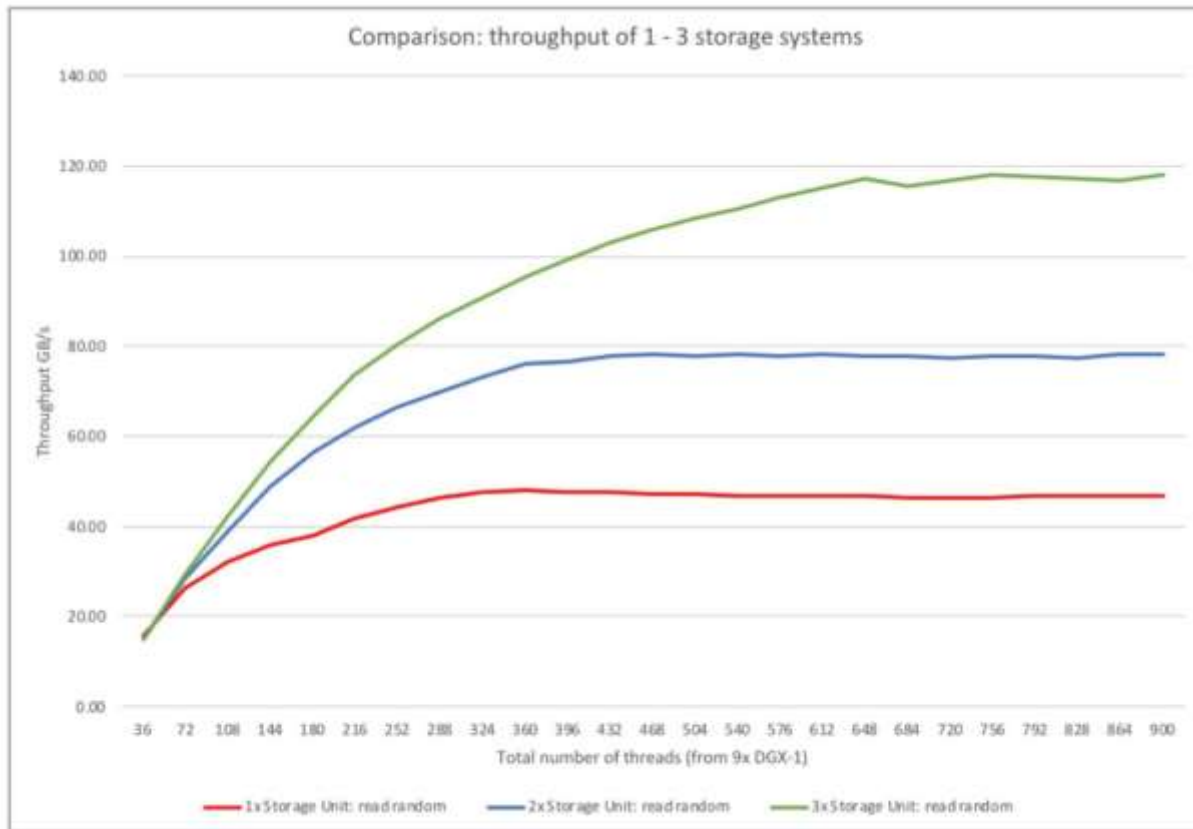


Figure 13: System scalable throughput using fio benchmark

Spectrum AI : more information

- Spectrum AI Introduction (16 pages)
https://public.dhe.ibm.com/common/ssi/ecm/81/en/81022381usen/ibm-spectrumai-ref-arch-dec10-v6_81022381USEN.pdf
- IBM Flashcore technology
<https://developer.ibm.com/storage/2018/08/06/new-form-factor-ibm-flashcore/>
- Blog (Eric Herzog):
<https://www.ibm.com/blogs/systems/introducing-spectrumai-with-nvidia-dgx>

Please Note

IBM's statements regarding its plans, directions, and intent are subject to change or withdrawal without notice at IBM's sole discretion.

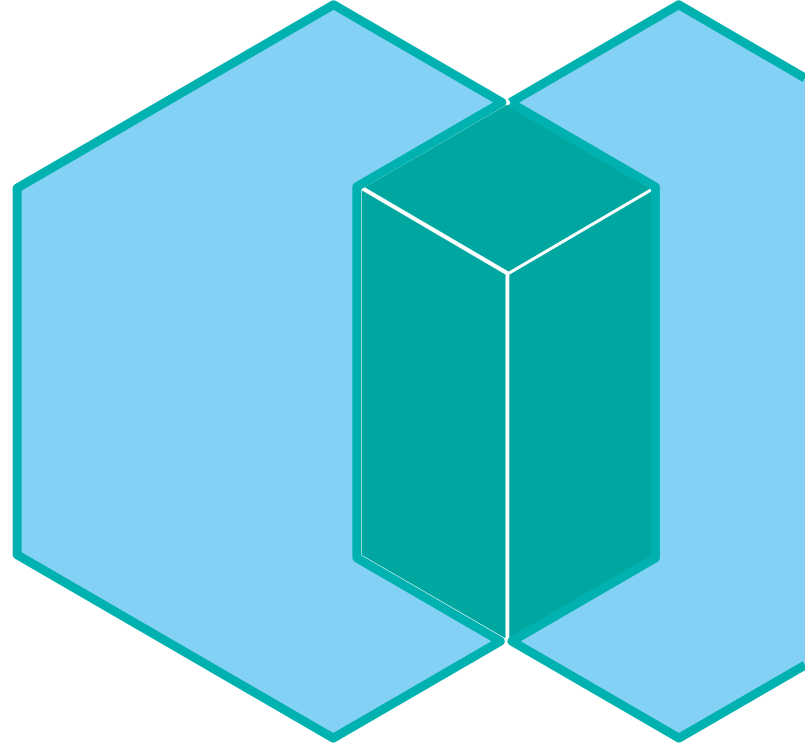
Information regarding potential future products is intended to outline our general product direction and it should not be relied on in making a purchasing decision.

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.

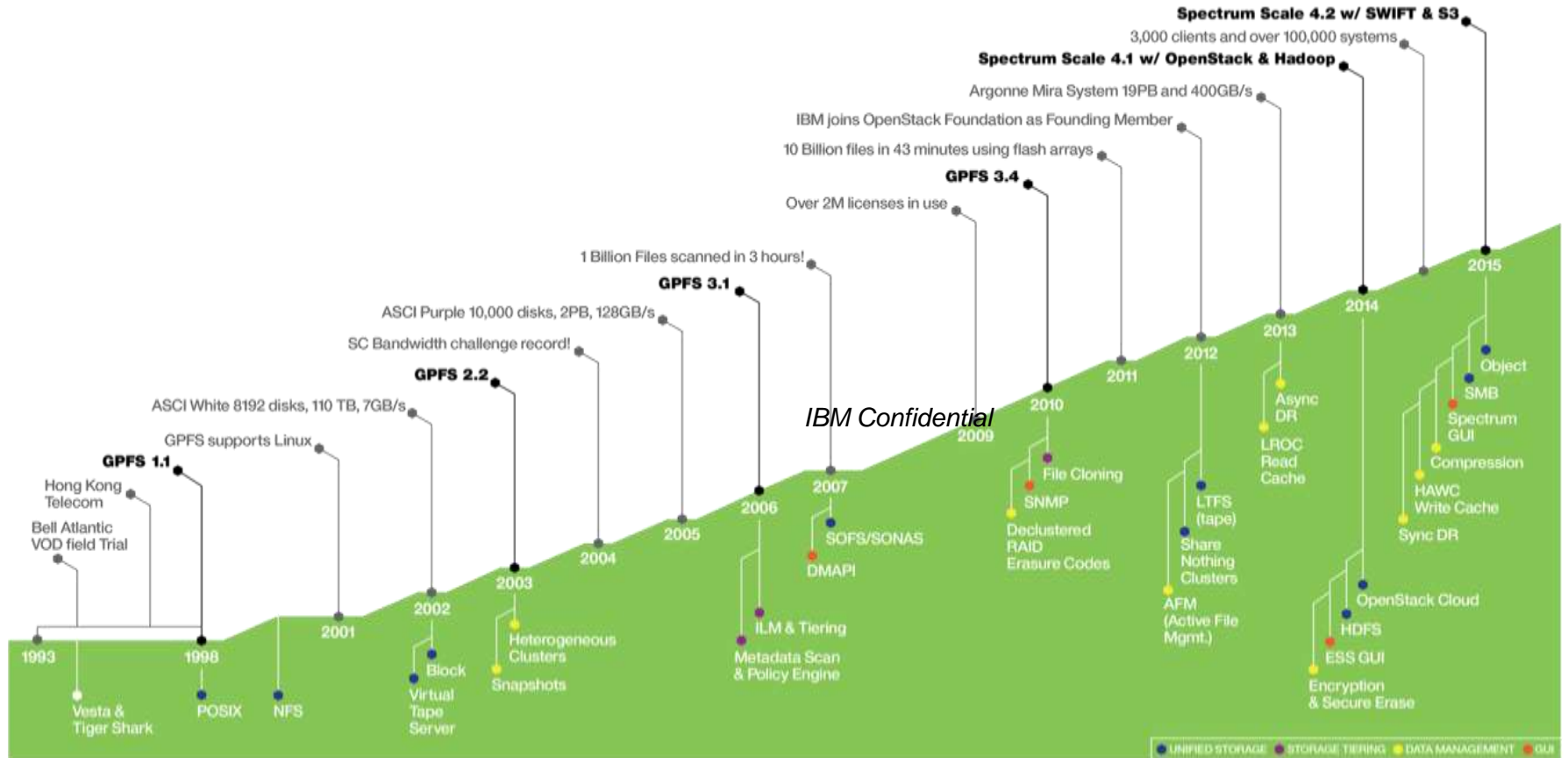


IBM Spectrum Scale

What's new in 5.0.1



The History of Spectrum Scale



Refresh of why IBM went to version 5

1. Separation of the sub-block size from the blocksize
 - New default blocksize is 4MB with (512of) 8KB sub-blocks (was 32of 128kB sub-block)
2. ESS: No longer needs to pre-allocate say 5% for Metadata
3. GUI extended to almost all features including AFM, TCT, etc.
 - So single 'pane-of-glass' like several competitors have.
4. TCT enhancements eg multi-cloud, support for Multi-cluster etc.
5. Support for TCT with direct access to the Object Store. (Cloud Sharing)
6. Lots of performance improvements, largely driven by CORAL eg 3-5x on metadata
7. Logging of every file access (with DME edition)
 - Audit logging, anti-virus. Uses LWE like TCT does.
8. Now available on Amazon AWS
 - As well as IBM Cloud (aka Softlayer)

5.0.1 Updates

Filesystem Enhancements

- Improved fsck validation of directory blocks
- Improved ability to add disks to a full filesystem
- Improved integration with systemd
- Support for end-to-end data checksums in non-ESS environments with unreliable network connectivity

Currency Updates

- Power 9 processor support (Linux kernel 4.11 and 4.14 support)
- SKLM 3.0 support
- Openstack Pike support

5.0.1 Updates (continued)

GUI Updates

- New Services Panel
 - Manage (start, stop) and monitor services like SMB, NFS, CES, GPFS, Zimon
- Display if cluster is configured using tiebreaker disks
- Panel for Audit Events for the cluster
- TCT Monitoring: list accounts and details, show container pairs
- Display performance data from remote clusters
- Improved usability for SMB share/NFS export creation

File Audit Logging Updates

- Support rolling upgrade of Kafka
- Improved error messages for easier problem determination

5.0.1 Updates (continued)

Deployment Toolkit Updates

- Improved installation and configuration for performance monitoring (zimon) toolkit
- Support rolling upgrade of zimon
- Support configuration population functionality for Call Home and File Audit Logging
- Improved shutdown and unmount orchestration during upgrades
- Simplify setup and upgrade of LDAP/AD configuration for GUI authentication

Transparent Cloud Tiering Updates

- Scaling improvements (tested to 1 Billion files)
- Integration with SOBAR (Scale out backup and restore) for service restoration
- Performance tuning
- Automation of common maintenance actions

5.0.1 Updates (continued)

Health Monitoring Updates

- Improved configuration options for performance monitoring tools
- Enhanced checking of Infiniband port state: monitoring of port speed and width
- Addition of performance data to callhome package
- CES improvements:
 - Network monitoring for CES IPs when using node affinity
 - Identification of unassigned CES IPs
 - Improve error messages for easier problem determination
 - Enhanced reporting for CES IP moves and rebalancing

Big Data and Analytics Updates

- HDFS Transparency rolling code updates
- Hortonworks Data Platform (HDP) 2.6.4 support
- Enhanced Snapshot support (fileset and remote-filesystem snapshots supported)

5.0.1 Updates (continued)

NFS Enhancements

- Renamed NFS daemon for improved compatibility with SELinux
- Support for netgroup names with a '.' character
- Automatic backtrace generation for better problem determination
- Logging improvements and cleanup of unnecessary messages
- Automatic restart of nfs-ganesha for better error recovery

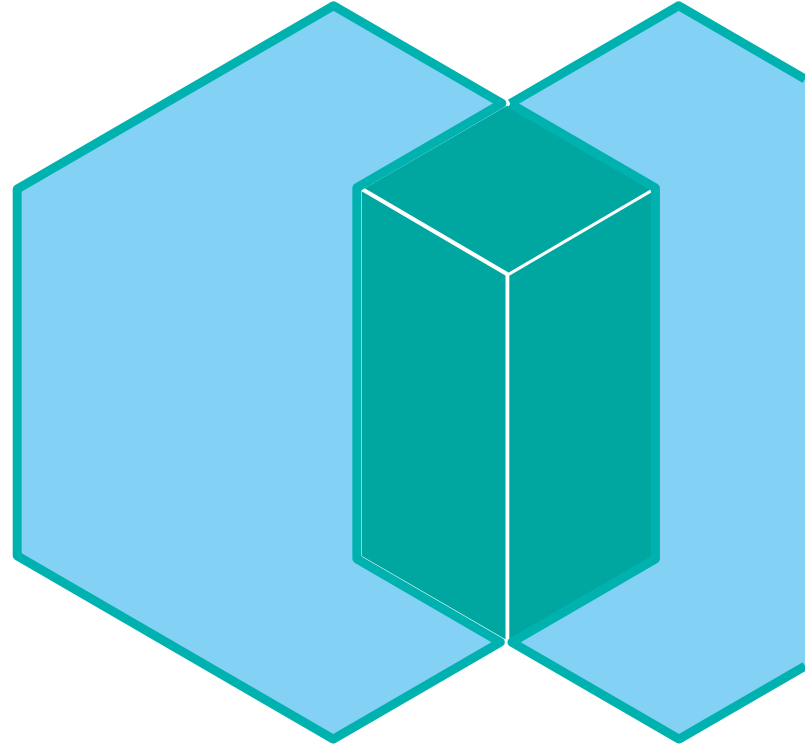
AFM and AFM-DR Enhancements

- When using AFM-DR, a secondary site can be converted into the primary site if the primary site becomes unavailable
- Secure NFS communication for enhanced security of data in-flight between clusters



IBM Spectrum Scale

What's new in 5.0.2



5.0.2 New Feature - Watch Folders

Watch Folders

- Flexible API that allows programmatic actions to be taken based on filesystem events
 - Can be run against directories, filesets, and inode spaces
- Modeled after Linux inotify, but works with clustered filesystems and supports recursive watches for filesets and inodespaces
- 2 primary components:
 - Spectrum Scale C API
 - mmwatch command for monitoring watches running in the cluster
- A watch folder uses the API to run as an executable C program on a node within the cluster
 - Utilizes message queues to receive events from multiple cluster nodes and consume events on the node running the application
 - Lightweight events are generated by eligible nodes within the cluster, and from remote clusters
- Integration into callhome, GPFS snaps, and GPFS tracing
- Contact IBM support for additional details prior to use

5.0.2 Updates

Filesystem

- Updates to stat cache on Linux systems
 - Stat cache is now usable on Linux systems
 - See updated performance guidelines regarding stat cache for more information
- Automatic rebuild of GPFS portability layer on Linux systems after kernel upgrades
 - Configurable via the 'autoBuildGPL' configuration option
- Addition of new filesystem maintenance mode
 - Disables certain filesystem operations, such as mounts, write I/O, disk down scenarios, and others if a filesystem is placed in maintenance mode
- mmnetverify enhancements, including remote cluster testing
- mmchconfig performance improvements
 - Reduce number of file operations – most noticeable in diskless or slow-disk environments
- LTFS EE and mmbackup integration improvements
 - LTFS EE awareness of backed up files to reduce the chance of recall storms during backups
 - Requires updates to LTFS EE, planned for future releases

5.0.2 Updates (continued)

GUI

- Display of quota information and capacity data for remote clusters
- Node classes can now be created, displayed, and modified
- Ability to enable and disable File Audit Logging
- Health events show up faster, especially on large clusters
- The cluster name is now displayed in the banner (usability)
- Users are logged off immediately if a password expires, a user role is changed, or user is deleted

Health and monitoring

- CES IP address lists show node affinity
- Improved CES node suspend for easier service and unmounts of filesystems
- Additional File Audit Logging events
- Ability to remove performance data for cluster components that have been removed
- Grafana bridge updates

5.0.2 Updates (continued)

REST API

- SMB ACL's can now be retrieved and modified
- File Audit Logging can be enabled and disabled and configuration information can be retrieved

File Audit Logging

- Multi-cluster/remote mount support for file audit logging
- Improved monitoring of Kafka producers
- Support for Linux on Z systems

NFS and protocol

- NFS CLI performance improvements
- Package names changed to avoid conflicts with distribution packages
- NFSv4 pseudo path support

5.0.2 Updates (continued)

Deployment Toolkit

- Linux on Z systems support
- Ubuntu 18.04/18.04.1 support for install, deployment, and upgrade
- Additional support for file audit logging and watch folders

Upgrade

- Documentation and flowchart improvements for easier upgrade planning
- Deployment toolkit tolerance of unhealthy nodes prior to an upgrade
- Improved ability to restart a failed upgrade in the deployment toolkit
- Improved filesystem unmount experience on CES-enabled nodes
- Documentation updates and performance tuning guide available

AFM

- Allow users to modify AFM gateway nodes used for a fileset
- Additional prefetch options
- Read-only NFS exports supported for read-only AFM relationships

5.0.2 Updates (continued)

Big Data & Analytics

- Support for Hortonworks Data Platform 3.0 and Management Pack 2.7.0.0
- Support for Apache Hadoop 3.0.x
- Support for native HDFS encryption
- Changes to configuration and logging paths for greater consistency
- FPO improvements to scanning for inconstant replicas and automatic recovery from certain node/disk failures

Currency

- Ubuntu 18.04.1 kernel support
- Windows 10 Enterprise Edition client support

LEARN MORE

Client Enablement Material

New videos, blogs and presentations:

- [https://www.ibm.com/developerworks/community/wikis/home?lang=en#!/wiki/General%20Parallel%20File%20System%20\(GPFS\)/page/White%20Papers%20%26%20Media](https://www.ibm.com/developerworks/community/wikis/home?lang=en#!/wiki/General%20Parallel%20File%20System%20(GPFS)/page/White%20Papers%20%26%20Media)

Watch for new blogs at:

- <https://developer.ibm.com/storage/blog/>

Turn-key Spectrum Scale VM available for download

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

Spectrum Scale Blueprints for Genomic Medicine Workloads

- Enablement and Redpapers available for these workloads

Thank You.

IBM Storage & SDI

