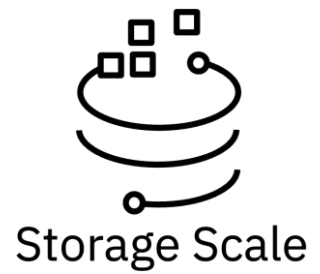# What's new in

IBM Storage Scale

the Global Data Platform
for all data lakes

Chris Maestas
IBM Chief Executive Architect
Storage Solutions for Data and AI
cdmaestas@us.ibm.com

IBM.

Disclaimer

Storage Scale

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.

Data is everywhere in a hybrid and multi-cloud world
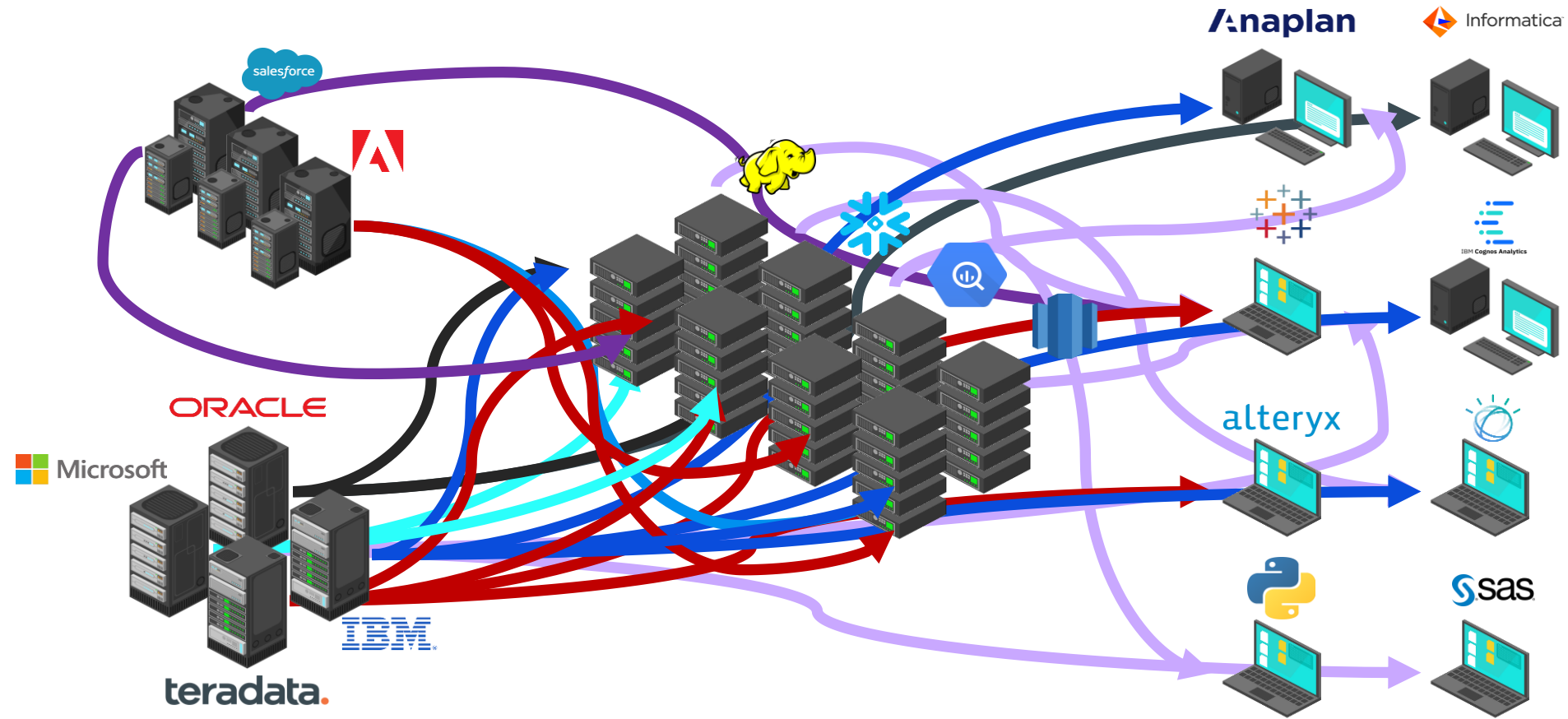
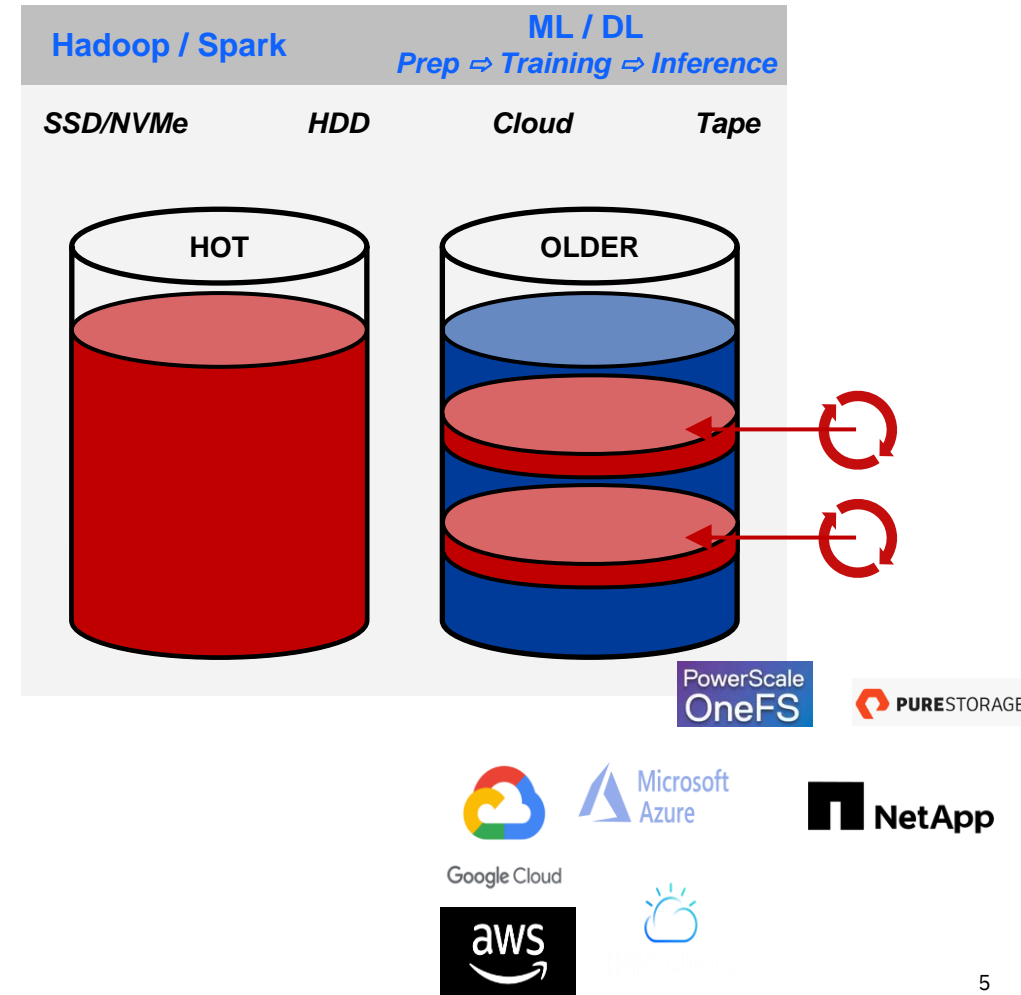# The Other Data Problem – let's look at your Archive!

We see:

Customers across all verticals are creating large PB to EB data stores.

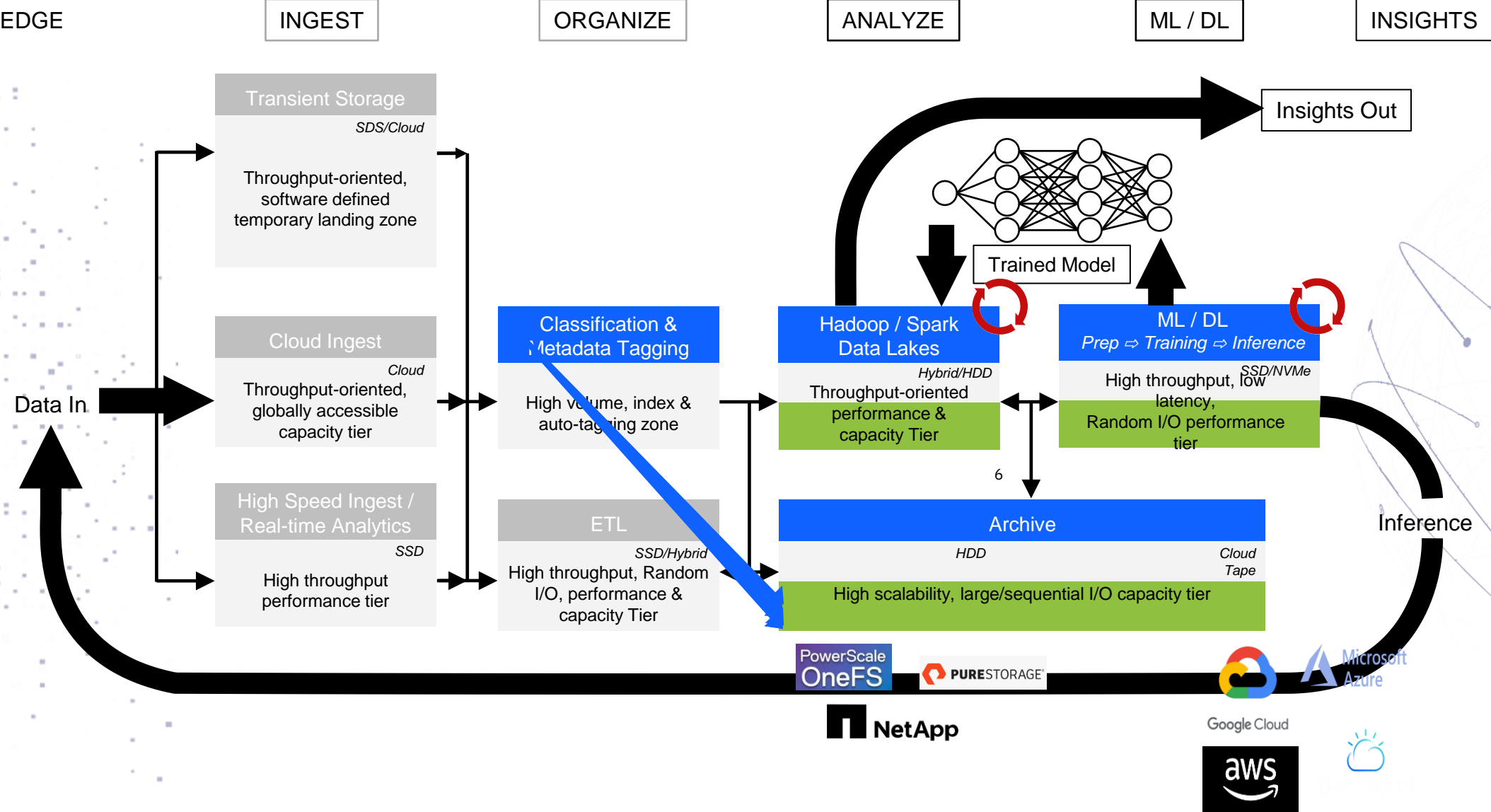Other just doesn't mean cloud or tape, it could be existing storage

Vast majority of data is relatively (c)old, but still required for periodic trend analysis.

AI / Analytics require high performance, low latency storage to keep expensive CPU / GPU / TPU / FPGA busy.

AI and Analytics Data – Archive Ingest
Data that resides in your existing storage!

EDGE **INGEST** **ORGANIZE** **ANALYZE** **ML / DL** **INSIGHTS**

Insights Out

Trained Model

Transient Storage
SDS/Cloud

Throughput-oriented, software defined temporary landing zone

Cloud Ingest
Cloud

Throughput-oriented, globally accessible capacity tier

Data In

High Speed Ingest / Real-time Analytics
SSD

High throughput performance tier

Classification & Metadata Tagging

High volume, index & auto-tagging zone

ETL
SSD/Hybrid

High throughput, Random I/O, performance & capacity Tier

Hadoop / Spark Data Lakes
Hybrid/HDD

Throughput-oriented performance & capacity Tier

ML / DL
Prep ⇨ Training ⇨ Inference
SSD/NVMe

High throughput, low latency, Random I/O performance tier

6

Archive
HDD   *Cloud Tape*

High scalability, large/sequential I/O capacity tier

Inference

PowerScale OneFS   PURESTORAGE

NetApp

Google Cloud   Microsoft Azure   aws

IBM Storage Scale
Delivering a High-Performance Global Data Platform for Unstructured Data

IBM Storage Scale Developer Edition

https://www.ibm.com/products/storage-scale
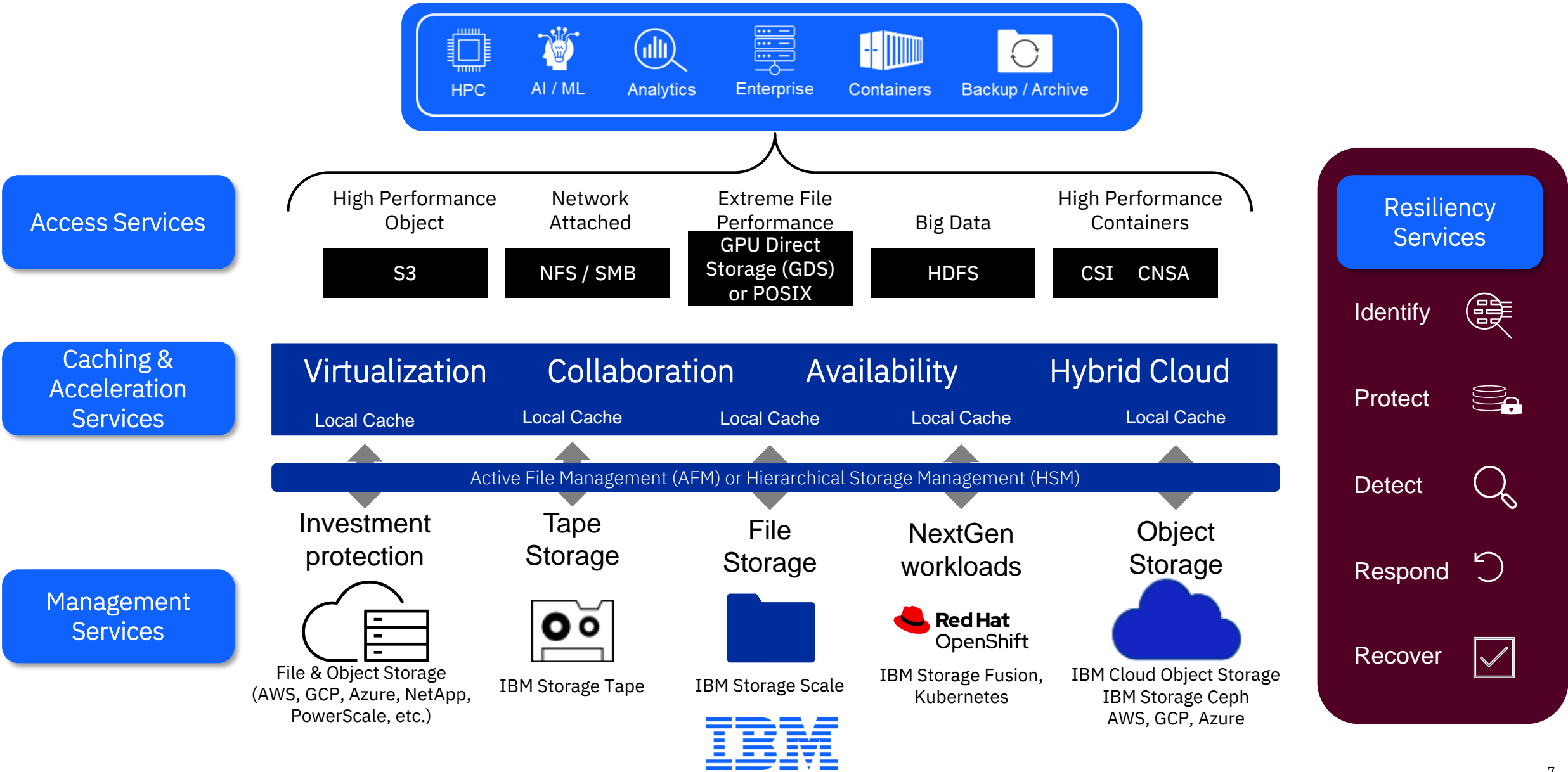
IBM Storage Scale

Fully functional!

Derived from the **most advanced Scale Technology**

Download it or schedule a demonstration!

Accelerate AI and unlock value from your data

★★★★⯪ 17 Reviews - G2 Crowd

Try the free developer edition →   Schedule a free demo →

Storage Scale User Group

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

Techzone – https://techzone.ibm.com/collection/ibm-storage-scale-developer-edition-labs

IBM Storage Resources

★★★★★ (1) Rat

Apr 16, 2023
IBM Storage Scale Developer Edition - Installation Lab

Ibmcloud 2: us-east, us-south, ca-tor, eu-gb, eu-de, jp-tok, jp-osa

IBM Storage Scale Developer Edition - Installation Lab

Visibility
IBMers, Business Partners

Reserve

Apr 19, 2023
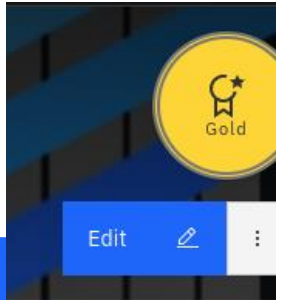**IBM Storage Scale Developer Edition Lab**

Ibmcloud 2: us-south, us-east, ca-tor, eu-de, eu-gb, jp-tok, jp-osa

IBM Storage Scale Developer Edition Installed on a 5 node system consisting of a GUI, 2 clients and 2 storage servers.

Visibility
IBMers, Business Partners

Reserve

Edit ✏ ⋮

Gold

Mar 14, 2023
Login - to Storage Scale Developer Edition Labs via SSH key method

Powerpoint step by step - login to Techzone Storage Scale Developer Edition Labs via SSH key method.

Visibility

IBMers, Business Partners

📄

Mar 14, 2023
Lab Guide - Installation Toolkit Lab

README.md file step by step instructions, commands to do Storage Scale install using Storage Scale Install Toolkit.

Visibility

IBMers, Business Partners

🔗

Mar 14, 2023
Lab Guide - Storage Scale guided exercises for GUI and command line

README.md file step by step instructions to perform various Storage Scale operations on GUI and command line.

Visibility

IBMers, Business Partners

🔗

Mar 14, 2023
Video - Storage Scale GUI Overview

Video showing extensive step by step examples of using Storage Scale GUI.

Visibility

IBMers, Business Partners

📹

9

Mar 14, 2023
Video - Storage Scale Data Management Services

Video showing extensive step by step examples of using Storage Scale Integrated Life Cycle Management (ILM) - place data, move and manage data.

Mar 14, 2023
Video - Storage Scale Data Caching Service

Video showing setup Storage Scale Active File Management (AFM) connections to NFS or S3 target.

Mar 28, 2023
IBM Storage Scale New User Group

This is intended for a New Scale User Group Session. The agenda is as follows:

- 5 minutes - Welcome
- 25 minute - Planning for Data-Intensive Science

Mar 28, 2023
2-3 day - IBM Storage Scale and Scale System Client Workshop

Day 1

- 15 minute - Welcome and introductions
- 45 min - Overview of the IBM Global Data Platform powered by Storage Scale

Storage Scale Editions and Licensing

Editions have various function levels:
- Data Access Edition (DAE) – standard level often used for HPC
- Data Management Edition (DME) - adds advanced functions, valuable in commercial environments
  - Free Developer Edition (DE)
- Erasure Code Edition (ECE) - aimed at hyperscale, web-scale service providers

Capacity licensing: built for simplicity
- Easy to purchase, expand, budget, renew
- Entitled to unlimited number of IBM Storage Scale client and server licenses

Feature	Data Access Edition	Data Management or Developer Edition	Erasure Code Edition
Multi-protocol scalable file service with simultaneous access to a common set of data	Yes	Yes	Yes
Facilitate data access with a global namespace, massively scalable file system, quotas and snapshots, data integrity and availability and filesets	Yes	Yes	Yes
Simplify management with GUI	Yes	Yes	Yes
Improved efficiency with QoS and compression	Yes	Yes	Yes
Create optimized tiered storage pools based on performance, locality, or cost	Yes	Yes	Yes
Simplify data management with Information Lifecycle Management (ILM) tools that include policy-based data placement and migration	Yes	Yes	Yes
Enable worldwide data access using AFM asynchronous replication	Yes	Yes	Yes
Immutability (WORM / Write Once Read Many)	Yes	Yes	Yes
Container Native Storage Access (CNSA)	Yes	Yes	Yes
Storage Scale Back-up Leverage	Yes	Yes	Yes
Asynchronous multi-site Disaster Recovery		Yes	Yes
Protect data with native software Encryption and secure erase, NIST compliant and FIPS certified		Yes	Yes
File audit logging		Yes	Yes
Watch folder		Yes	Yes
Fusion Data Catalog Entitlement (Discover)		Yes	Yes
Erasure coding	Scale System only	Scale System only	Yes

	1H23	4Q23	1H24	2H24
Access	• HPO: Multi-protocol data access • SMB 4.17	• Scale Client on ARM (Tech Preview) • Ganesha 4.3		
Core and Caching	• Dynamic Pagepool (Tech Preview) • GNR Checksum improvement +30% throughput	• Dynamic Pagepool (GA) • migrate from TCT to AFM-MU • Native Azure blob / Ceph S3 support • Test and validate your S3! • Scale System: NVMeoF high IOPs with data integrity (Ustore)		

Management	• Scale DME/ECE includes Fusion Data catalog
	• Cloudkit: Automation to provision and deploy Scale on Cloud (AWS)
		• Cloudkit
		  • AWS Upgrade support
		  • GCP
		• mmxcp support hardlinks and WORM attributes
		• mmapplypolicy silent on delete
		• Data Orchestrator Sponsor User		

Resiliency	• Storage Scale Backup • GKLM for Scale System (per install)	• Scale control plane and security architecture modernization - RBAC Tech Preview 2H23 • Support signed kernel modules for UEFI Secure boot		

Access Services
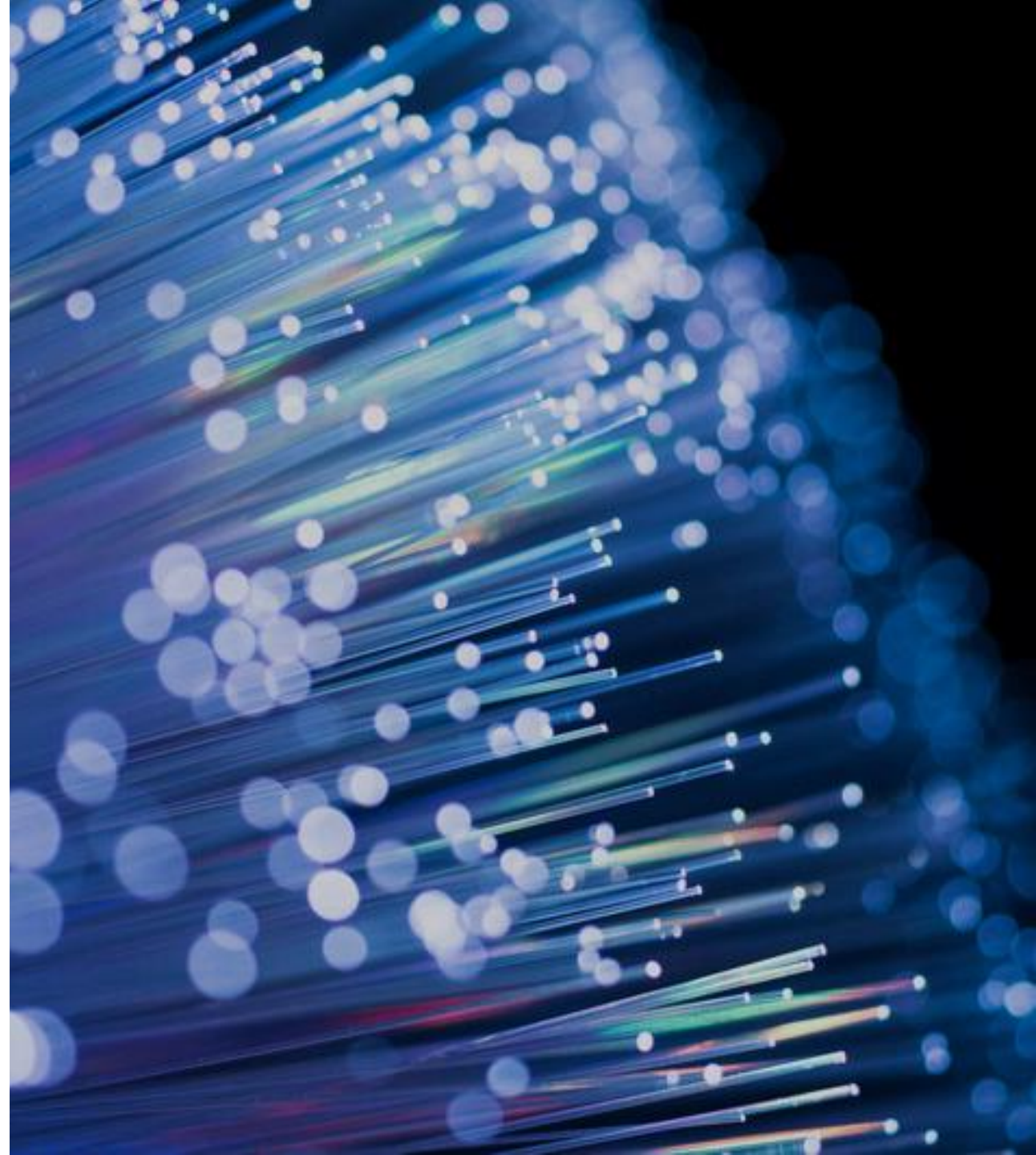
Access Services – ARM

Wider support to use ARM functionality Data Processing Units (DPU)

QuantaGrid S74G-2U

supported on the ARM processor as a technology preview (==nonproduction environments==), starting with IBM Storage Scale 5.1.9.
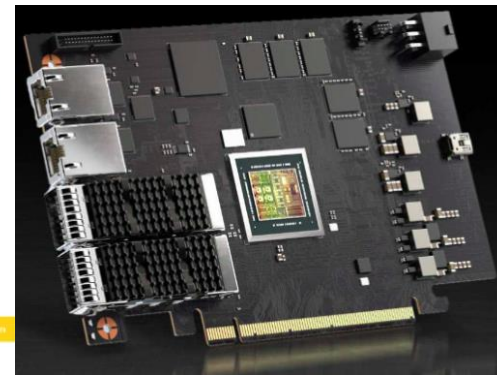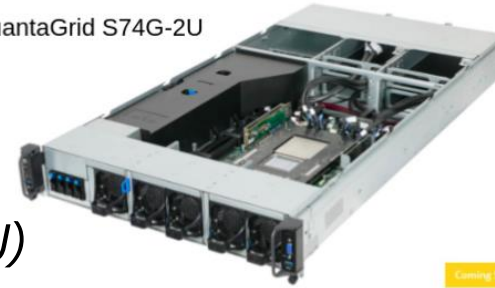
developed for the ARM processor with an instruction set of at least version 8.2-A

      Raspberry Pi 4 and 5 not supported!
      Missing Privileged Access Never (PAN)
          support or not supported

The official Architecture name is aaarch64

- Target: IBM Storage Scale client for HPC compute node
- **Included**
  - rpm based install
  - NSD client
  - Scale base functionality (IO, policies, remote mounts, snapshots, quotas, etc.)
  - Manager roles: file system manager / token manager / cluster manager
  - RDMA (IB or RoCE)
  - Health Monitoring
  - Target OS: RHEL 9 (4K page size)   *[RHEL 8 and Ubuntu packages are available, if needed]*
- **Excluded, but planned for future releases**
  - **Install toolkit**
  - File audit logging, watch folders folders (GA)
  - NSD servers (traditional, ECE, SNC)   *[NSD server has paretially been tested already, working]*
  - GDS   *[work in progress, issues found]*
  - GUI
  - Call home
  - Compression
- **Excluded**
  - Protocols
  - BDA / HDFS
  - CNSA
  - GNR *[plans to get ECE on ARM are emerging]*
  - TCT
  - HSM

Access Services – Container Native Storage Access (CNSA)
Improvements introduced in CNSA 5.1.7

Wider support to use the latest CNSA functionality.

Support for upgrading IBM Storage Scale Container Native Storage Access (CNSA) from v5.1.6.x to 5.1.7.0

Support for RedHat OpenShift Container Platform 4.12

Support of IBM Storage Scale Data Access Services (DAS) 5.1.7

AWS ROSA with IBM Storage Scale container-native storage access configuration

- Support for X86, Power and Z.
- Support of custom tolerations for IBM Storage Scale container native core pods
- Reduced memory footprint of CNSA operator with additional memory and cpu profiling

Access Services – Container Storage Interface

Improvements introduced in CSI 2.9

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

- Planned support for Red Hat OpenShift 4.12 and Kubernetes 1.26

- FSGroup support – security in k8s that is a supplemental group for containers in a pod
  - Allow multiple pods to access common files

- Volume stat support – when the volume is attached to the running pod, can show the used size under the used column (openshift today).

- Clone and restore volumes across different storage classes

- Kubernetes CSI sidecar containers upgrade

- Resource requests and limits support

Access Services – HDFS, NFS, SMB

Support and Currency:

- Cloudera Data Platform (CDP) Private Cloud Base is certified with IBM Storage Scale on x86_64 and ppc64le since December 2020.

- Opensource Hadoop 3.1.3, 3.2.2, 3.3.0

- Includes HDFS Transparency 3.1.1-15, HDFS Transparency 3.2.2-6

- Changed HDFS install process to use self-provided JAR files

- Support **mmhdfs config dump**

- NFS-Ganesha support for ~~3.5~~ 4.3 code base

- Continued partnership with Tuxera for high-performance SMB

Improved performance:

- HDFS - Buffered logging and log filtering for increased HDFS throughput

- HDFS - Increase recursive deletions of snapshot-enabled directories, by avoiding mmlssnapshot dependency

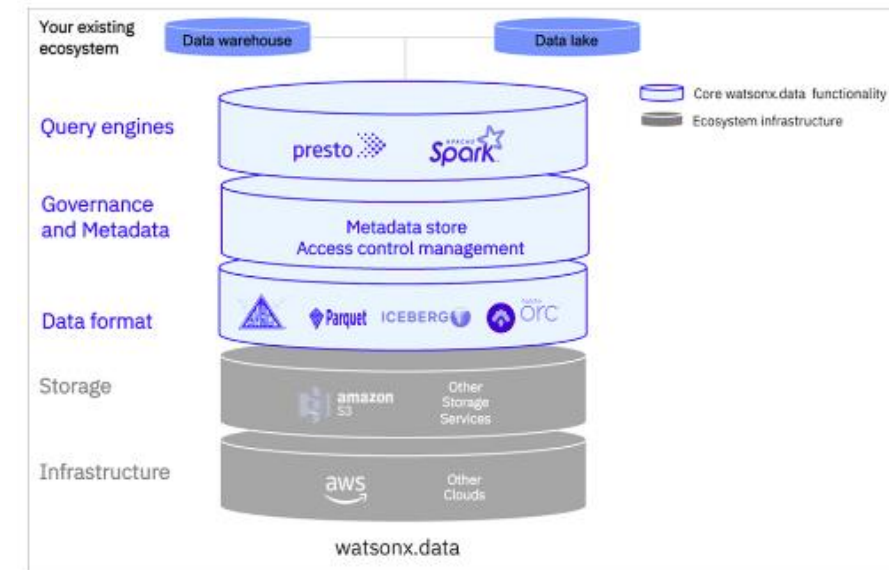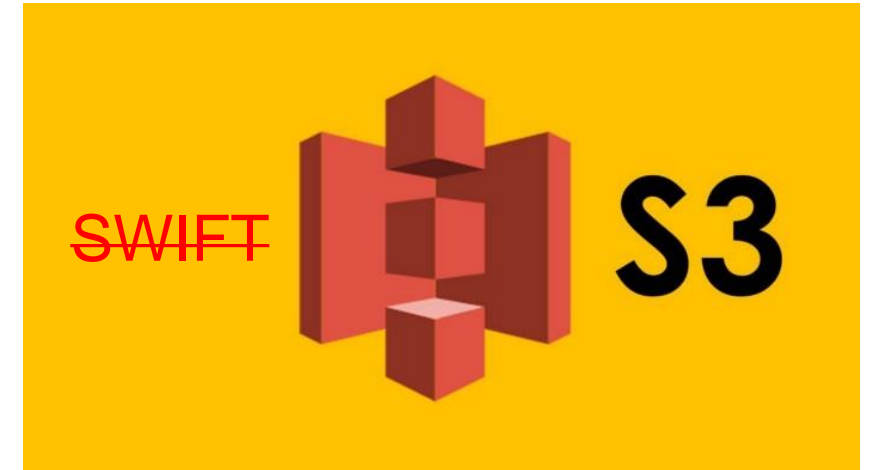- Reduced latency for Ganesha NFs v3/4/4.1 clients
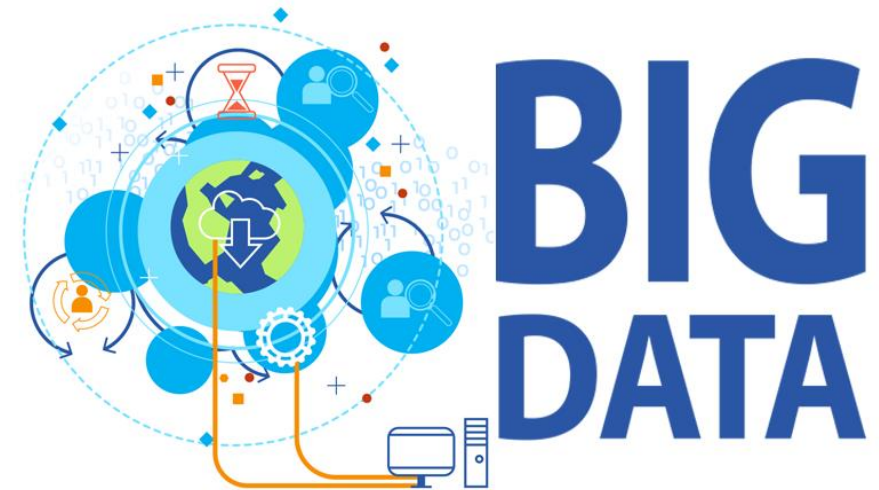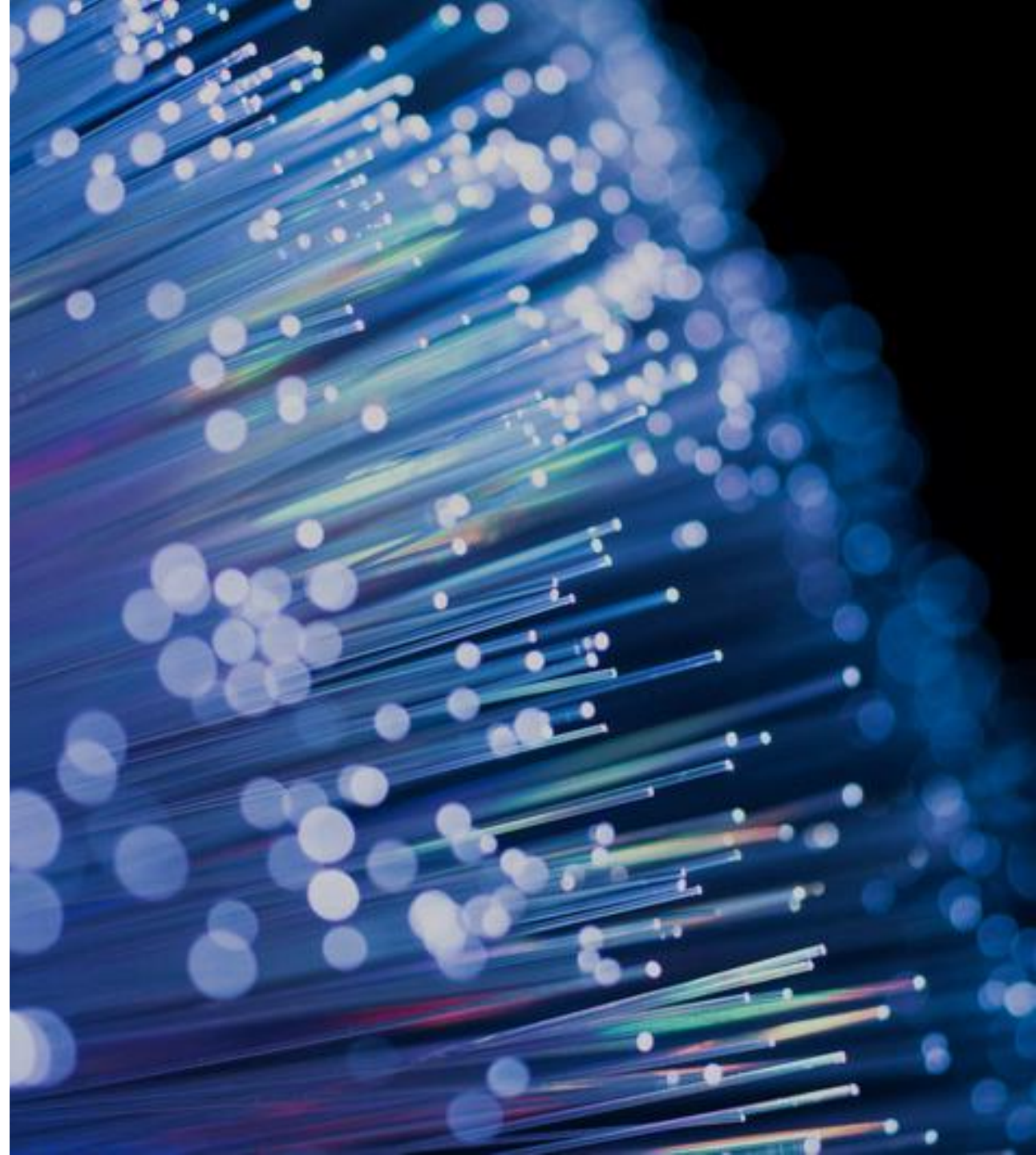
Access Services – Object

Support and Currency:

- Swift is being Discontinued
  - You can use 5.1.8 Swift code in CES of 5.1.9
- NSFS (Standalone NooBaa) is coming!
  - https://community.ibm.com/community/user/storage/blogs/pratibha-joshi/2023/09/29/using-nsfs-standalone-noobaa-with-ibm-storage-scal?CommunityKey=1142f81e-95e4-4381-95d0-7977f20d53fa
  - Tested at "The University of Queensland" and Americas POK environment

Improved performance:

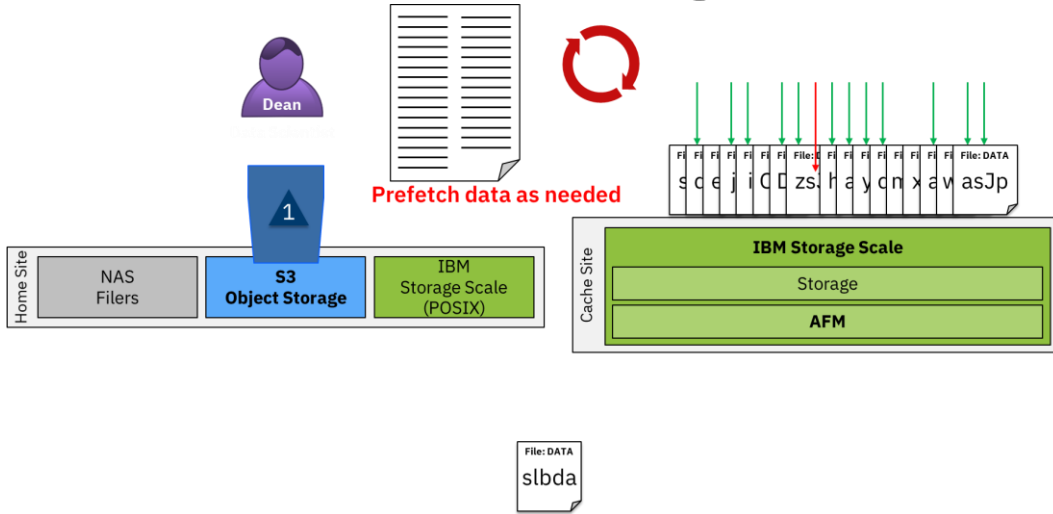- NSFS Standalone NooBaa testing is currently at about
  - 10.5 GB/s of read from single NSFS server
  - 5 GB/s of write
  - Cosbench testing underway
    - https://community.ibm.com/community/user/storage/blogs/silvana-de-gyves-avila1/2022/05/20/ibm-data-access-services-performance-evaluation

Acceleration (Core) and Caching Services

Caching and Acceleration Services – Active File Management (AFM)

- Support Ceph S3 as object storage backend

- Support of Azure Blob native API for AFM to Cloud Object Storage filesets.

- Validation Tool using REST API and Certification Tool using: mmafmtransfer

- Migrate TCT enabled fileset to AFM-S3 MU (Tiering only)

- Out-of-band metadata pre-population with AFM-POSIX
  - Avoid queueing any messages at gateway node

- Support of creation/deletion of dependent fileset in AFM-DR without unlinking.
- Disable AFM LU and RO mode fileset without unlink

- Return ENOTREADY to the application instead of EIO

- Enable tiering at the filesystem level
  - (AFM COS MU mode)

- Skip directories using skip-list
- Empty ptrash while prefetching

- Tested AFM DR RPO to 15 mins and lookup performance improvements to be GA in the next release
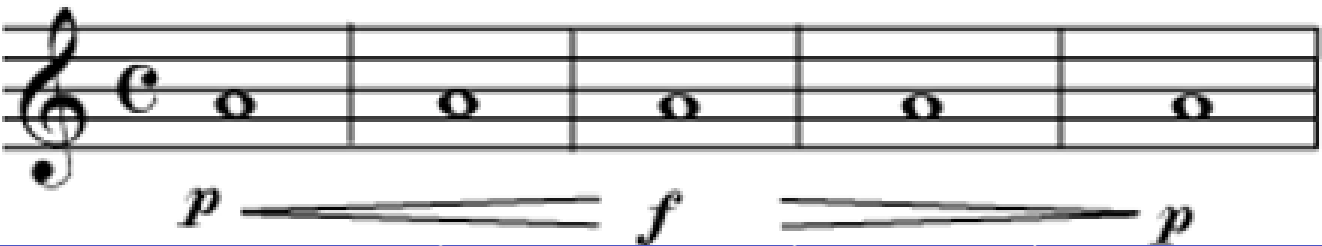
Caching and Acceleration Services – Dynamic Page Pool

Dynamic workload management!

Scale detects a shortage of the pagepool memory, then attempts to increase the pagepool size.

When the Linux kernel detects the memory pressure, it requests Scale to shrink the size of the pagepool.

Configuration:

mmchconfig dynamicPagepoolEnabled=yes –N node1

mmchconfig pagepool=default –N node1

mmshutdown –N node1

mmstartup –N node1

mmdiag –pagepool

GPFSBufMgr monitor pagepool size via zimon

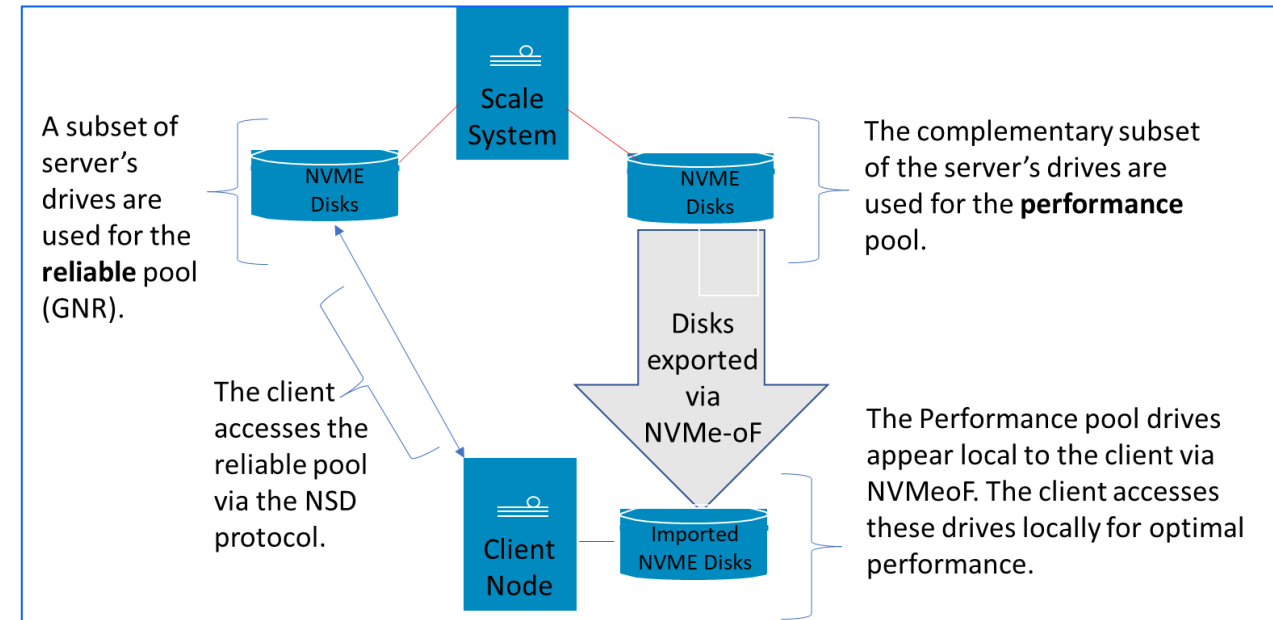| Config parameter	Allowed values	Default	Description
dynamicPagepoolEnabled	yes/no	no	Enable dynamic pagepool vs. static pagepool
pagepoolMinPhysMemPct	1-50	5	Minimum size of dynamic pagepool as percentage of physical memory.
PagepoolMaxPhysMemPct	10-90	75	Maximum size of dynamic pagepool as percentage of physical memory.
pagepoolChangeGracePeriod	1-86400	10	The grace period for growing the dynamic pagepool, in seconds. The dynamic pagepool grows only once every grace period

Caching and Acceleration Services – Ustore
Accelerate AI and Analytics by storing the data as close to the compute as possible

Leverage both shared storage (e.g. NVMeoF) and storage inside of the compute node

Support **Asymmetric Replication** with one erasure-encoded copy of the data and one performance copy for high-speed access

Creates a **Shared Co-operative Cache** across all compute nodes. Any node can access all cached data, regardless of physical location.

The first release, writes update all copies. In a follow-on release, allow writes to performance copy only with **Eventual Reliability** (e.g. Burst Buffer)


A subset of server's drives are used for the **reliable** pool (GNR).

The complementary subset of the server's drives are used for the **performance** pool.

The client accesses the reliable pool via the NSD protocol.

The Performance pool drives appear local to the client via NVMeoF. The client accesses these drives locally for optimal performance.

Ustore has 2 advantages over LROC:

– LROC cache is potentially redundant because the same data can be cached on multiple nodes, but only a single copy of the data will be contained in the Ustore performance pool.

– LROC is not a persistent cache because a reload is needed after a failed node is back up. With Ustore, the performance copy is persistent except when a drive goes down (the reliable copy is used until the performance copy is available and any missing updates are applied)

Management Services
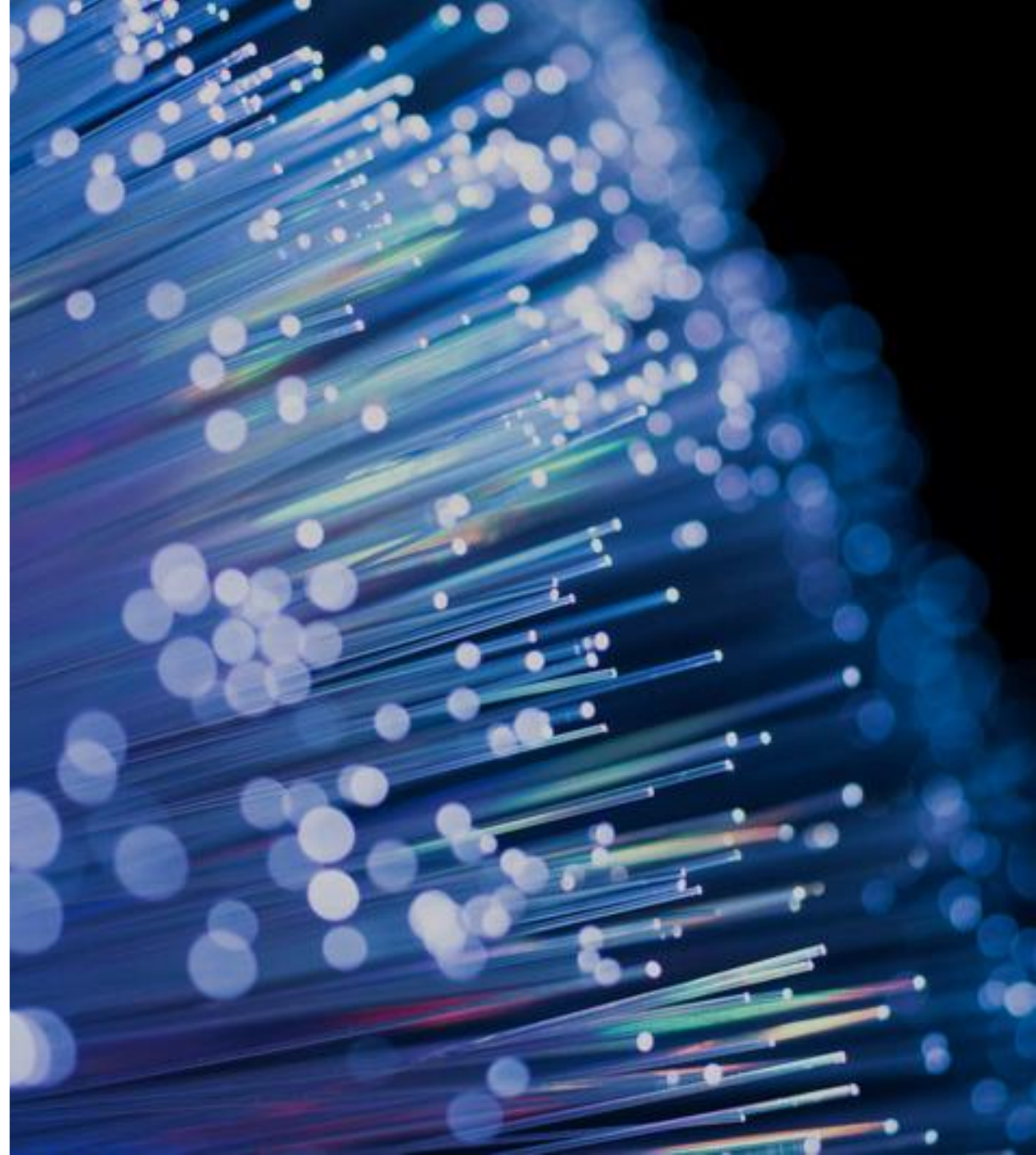
IBM Storage Fusion Data Cataloging Services entitlement with IBM Storage Scale

IBM Storage Scale announcement May 2023

https://www.ibm.com/docs/en/announcements/storage-scale-518-dual-pipe?region=US

- IBM Storage Scale 5.1.8 extends the management services of the global data platform. IBM Storage Scale **Data Management Edition** and IBM Storage Scale **Erasure Code Edition** entitle clients to use IBM Storage Fusion Data Cataloging Services
- Software entitlement to the follow-on for Storage Discover
- With Program Unique Terms that restricts usage to the Data Cataloging Services only in IBM Storage Fusion
- No limit on TB usage for data cataloging

Approach:

- Entitle IBM Storage Scale DME/ECE customers to IBM Storage Fusion as a supporting program
- Shift from releasing Storage Discover as a standalone product to an Application Data Services within the global data platform framework
  - Scan legacy storage / data lakes using the data cataloging services
  - Leverage AFM virtualization use case to consolidate data in the global data platform
- Foundational component for the Unified Data Orchestration in the IBM Storage Scale roadmap

Storage Fusion Data Catalog Use Cases

Storage Optimization

- Decrease storage CAPEX by facilitating data movement to colder, cheaper storage

- Increase storage efficiency by identifying redundant, obsolete, and trivial data

- Migrate relevant data to IBM Storage Scale System

- Reduce storage OPEX by improving storage administrator productivity

Improve storage utilization

Data Governance and Privacy

- Data discovery and classification

- Data Fabric Integration

- Ensure data is consistent with governance policies

- Reduce risk buried in unstructured data stores

- Speed investigations for legal discovery & regulatory audits

Mitigate risk & improve data quality

For AI & Analytics

- Accelerate data identification for large-scale analytics

- Operationalize tasks to reduce the burden of data preparation

- Orchestrate ML/DL & MapReduce processes (Data Orchestrator)*

Reduce time to accuracy & results

Management Services – Cloudkit!

What is Storage Scale Cloudkit?

Create Storage Scale clusters on the cloud with

Bring Your Own License (BYOL) Model
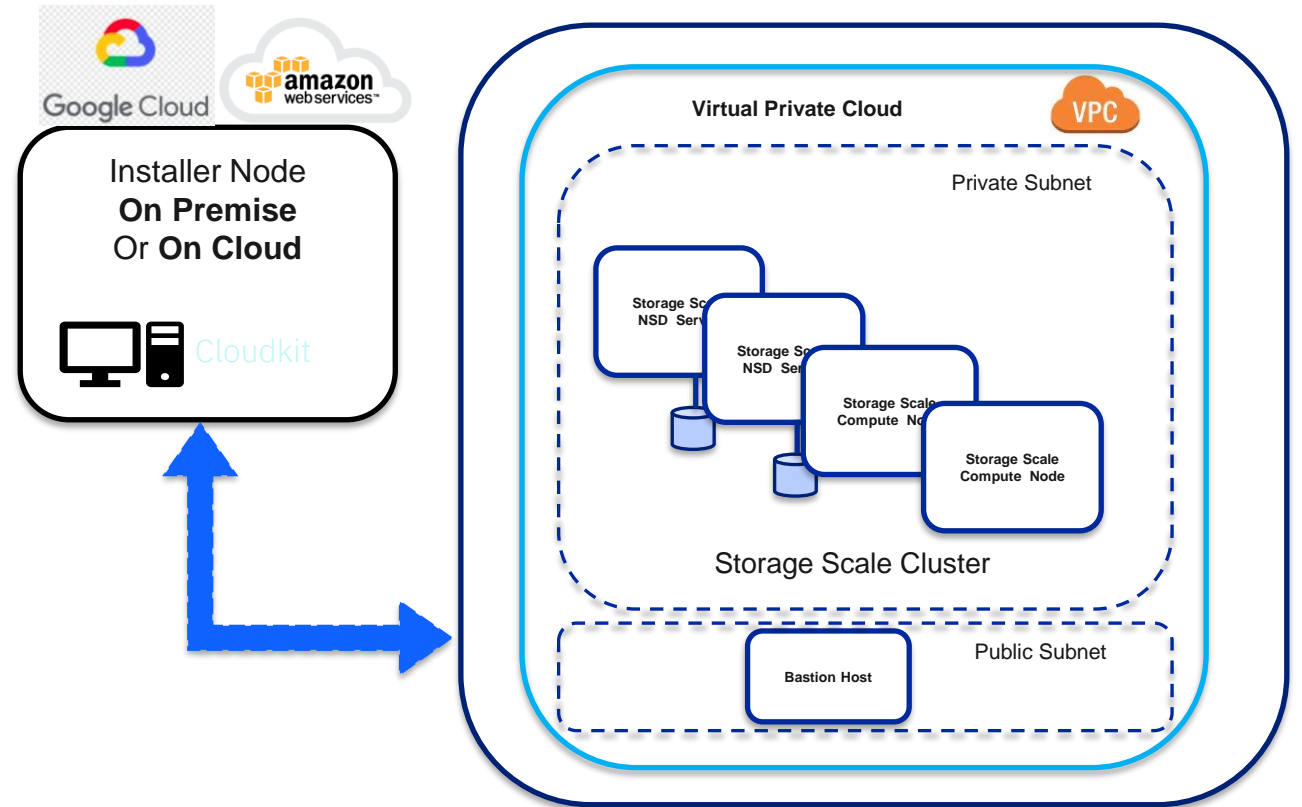Look in */usr/lpp/mmfs/VERSION/cloudkit*

Automates provisioning and deployment of
Storage Scale on the cloud

Applies Storage Scale best practices for
deploying on the cloud

Advantages

Support for major public clouds Amazon (AWS) and Google (GCP)
Support for scale-out on AWS cluster instances and upgrade

Google Cloud
amazon webservices™

Installer Node
On Premise
Or **On Cloud**

Cloudkit

Virtual Private Cloud
VPC

Private Subnet

Storage Sc
NSD Serv

Storage Sc
NSD Ser

Storage Scale
Compute No

Storage Scale
Compute Node

Storage Scale Cluster

Bastion Host

Public Subnet

Management Services – Cloudkit!
Cloudkit Fleet Performance Measurement

- Fleet (aka. Rapid expansion of compute/client nodes in a remote mount setup)

  - Traditional `mmaddnode` takes around `45sec` to add a single node (batching of 50 nodes is possible but it's is a sequential operation)

  - Quickly adding large nodes and deleting nodes before/after the burst situation is problematic.

  - Silent disappearance of nodes is a problem (as log recovery takes lot of time), which limited usage of spot instances (which comes at much cheaper price than an on-demand price).

Fleet Size	Minimum Time	Maximum Time	Average Time
1	42.69s	42.69s	42.69s
100	21.86s	32.33s	26.94s
200	20.98s	28.85s	23.74s
300	10.74s	1m7s	22.84s
400	19.38s	1m19s	24.91s

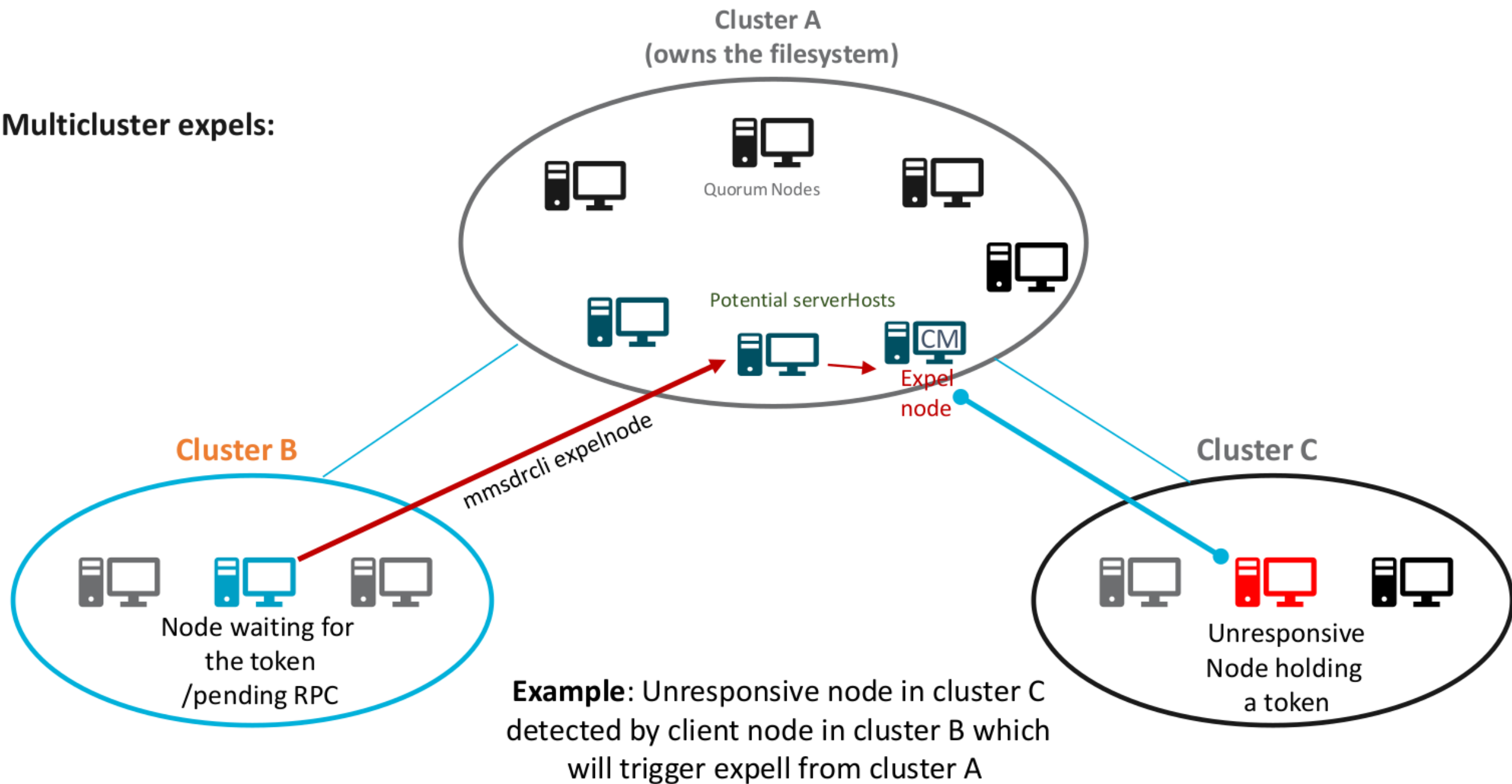# Management Services – mmxcp, mmapplypolicy, mmbackup

- **mmapplypolicy --silent-on-delete**

  - When running a cleanup policy using DELETE rule, it can happen that a file cannot be deleted for some reason

  - specified to tolerate directory not empty error during DELETE rule execution.

- **mmbackup**

  - difficult or impossible to determine the progress of an mmbackup job on Storage Scale and Protect

  - MMBACKUP_PROGRESS_CONTENT to instruct mmbackup to display total data size transferred

- **mmxcp --hardlinks** and **copy-attrs/verify**

  - added that executes an additional pass through the source files searching and copying hardlinked files as a single batch

  - appendonly and immutable, have been added which copies and verifies the appendonly and immutable attributes, if present.

Management Services –
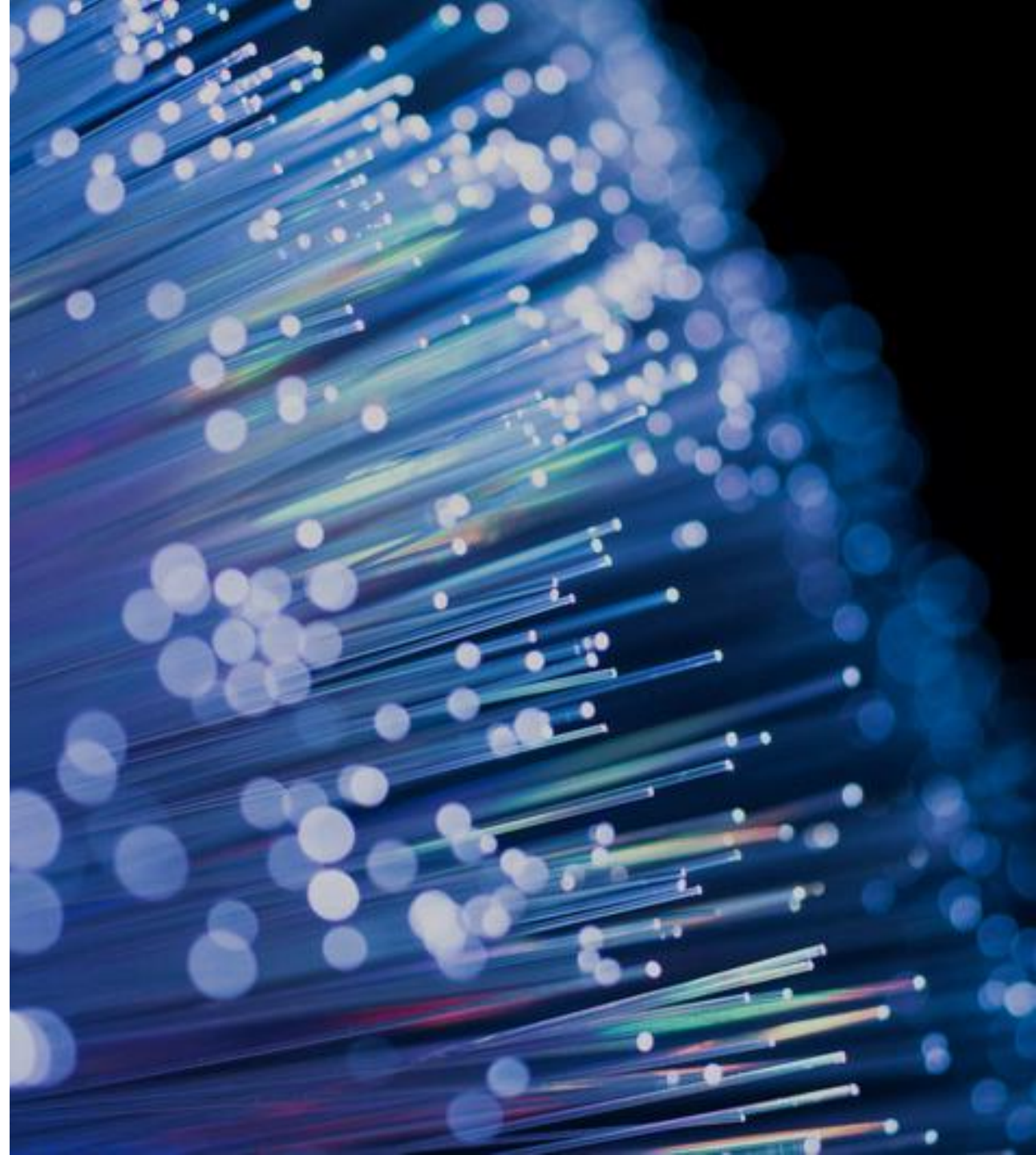# Monitoring, Availability & Proactive Services (MAPS) Updates

==Since 5.1.6.0 mmhealth detects pending RPCs for token revokes==

- Now an *rpc_waiters* event to inform the user about the misbehaving node
- Two mmchconfig options to control the feature:
  - mmhealthPendingRPCWarningThreshold (default = 180 seconds)
  - mmhealthPendingRPCExpelThreshold (default = 0 => disabled)

- mmhealth GPFS monitor runs "mmdiag --network -Y" every 2nd monitoring cycle (default 60s)
- Use mmchconfig to set a threshold; should be higher than the warning threshold:
  - mmchconfig mmhealthPendingRPCExpelThreshold=300

- If the mmdiag output contains some pending RPCs with an age > *mmhealthPendingRPCExpelThreshold*, it will:

  1. get the node to expel from the pending rpc output (victim)
  2. get the node (serverHost) where we send the expel request to from the connected node list, but matching the cluster id
  3. run mmsdrcli expelnode command with the victim and serverHost as parameters
  4. raise a rpc_waiters_expel warning event (non-state-change) for the particular node expel

Multicluster expels:

Cluster A
(owns the filesystem)

Quorum Nodes

Potential serverHosts

CM

Expel node

mmsdrcli expelnode

Cluster B

Node waiting for the token /pending RPC

Cluster C

Unresponsive Node holding a token

Example: Unresponsive node in cluster C detected by client node in cluster B which will trigger expell from cluster A

Resiliency Services

Security Services Ensure Assets are Safe
Following the NIST Standard

RECOVER

Near instant access to PB+ of data

CyberVault / Safeguarded Copy
IBM Storage Scale GUI/API
Integrated Tools

IDENTIFY

Metadata and services

IBM Security Assessment (CRAT)
IBM Storage Scale GUI/API

RESPOND

WORM Snapshots on threat detection

CyberVault / Safeguarded Copy
IBM Storage Scale GUI/API
Integrated Tools

Integrated Tools

Radar   >Splunk

IBM Cloud Pak for Security
IBM Storage Fusion Data Cataloging Service

PROTECT

WORM, encryption and air gapped data

Gov't Compliance FIPs/SEC
SED Drives and Software Encryption
CyberVault / Safeguarded Copy

IBM
Guardium

DETECT

Secure Audit Logs

Storage Scale GUI/API
Integrated Tools

Resiliency Services: Cyber Vault Framework

1) IBM STORAGE WITH INTEGRATED COPY MANAGEMENT

2) SAFEGUARDED COPIES

Protected PIT copies:
Immutable and Isolated with stringent RBAC's

Safeguarded copy
https://www.ibm.com/docs/en/Storage-scale/5.1.5?topic=administering-protecting-file-data-Storage-scale-safeguarded-copy

sudo-wrapper set up and configuration
https://community.ibm.com/community/user/storage/blogs/nils-haustein1/2020/12/17/Storage-scale-sudo-wrappers

4) SECURITY INTEGRATION

Detect and respond to threats real-time from a wide variety of data sources.

3) AUTOMATION

Automated data validation, data recovery and application integration
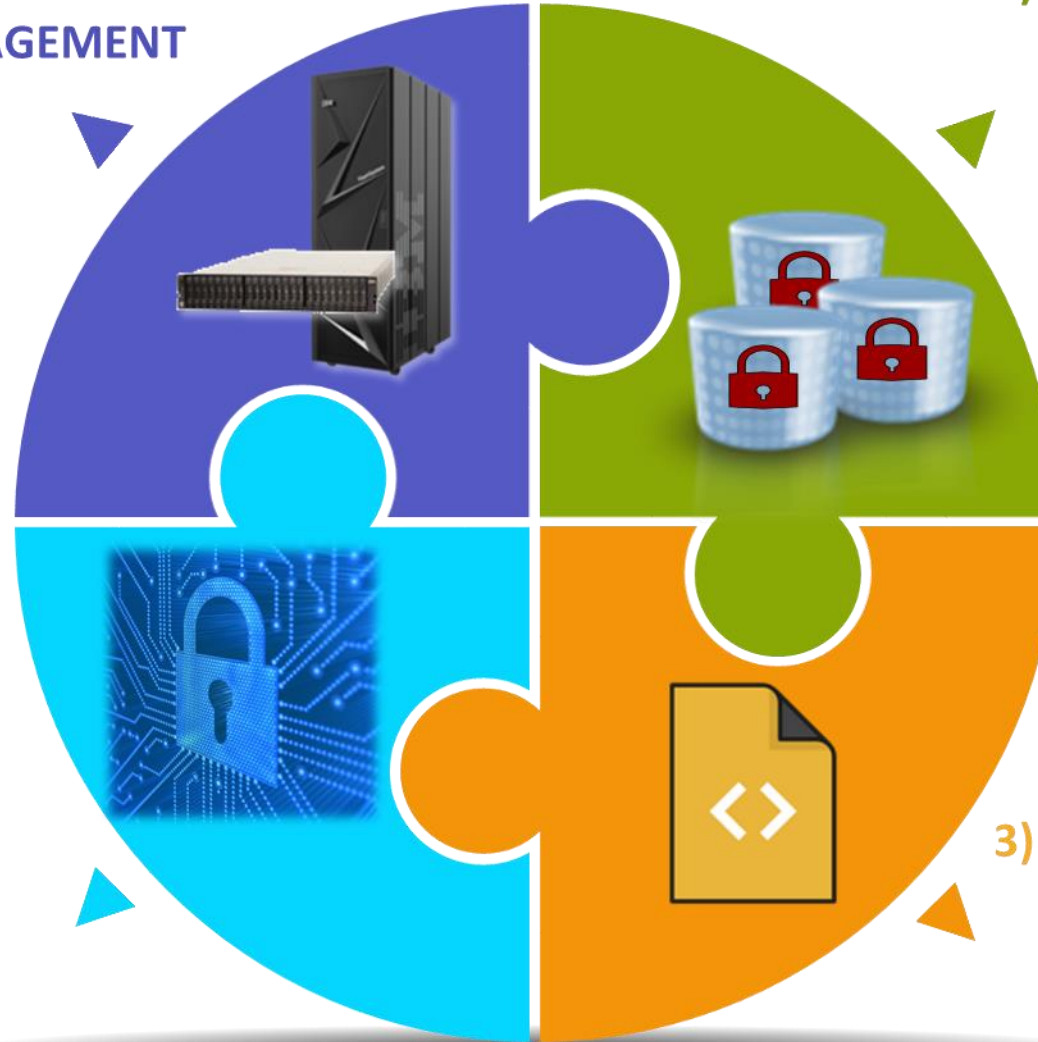
Resiliency Services: Native REST API
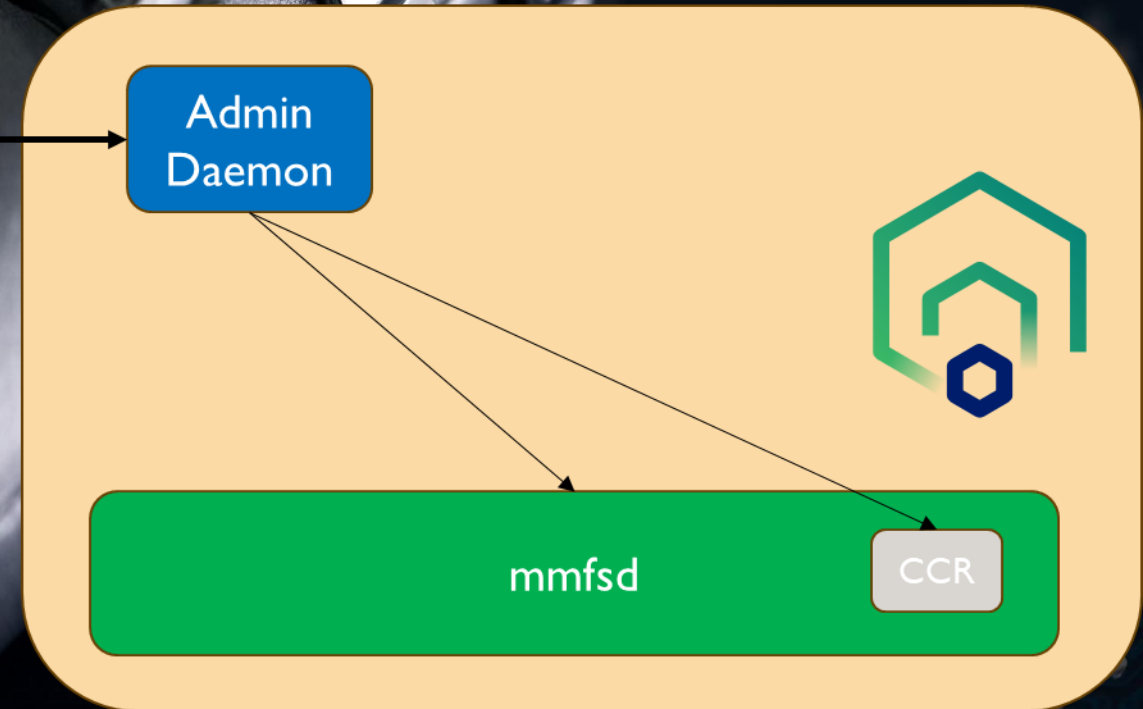
Technical Preview in Storage Scale 5.1.9.0
- Eliminates the need for passwordless SSH
- Introduces Role Based Access Control to the Scale control plane
- Allows administration of Scale by unprivileged users
- Allows remote administration of Storage Scale cluster
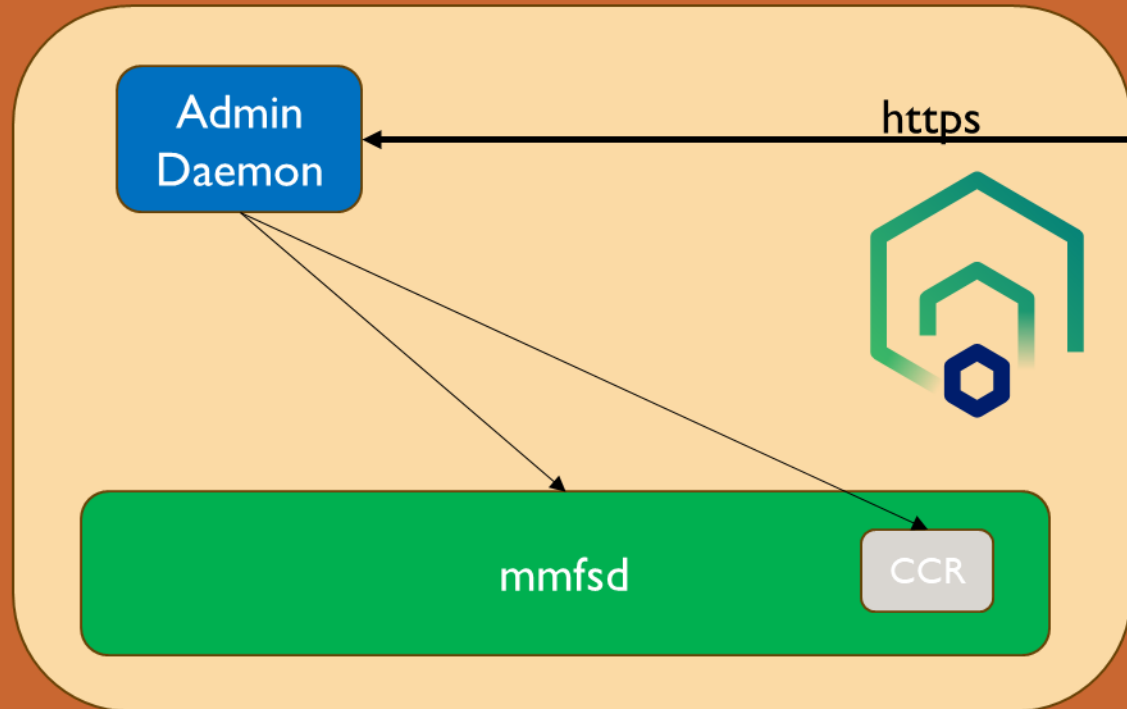
Removal of root requirement for control plane

Fine-Grained Access Control

API Server Node

Admin Daemon

https

Admin Daemon

mmfsd   CCR

mmfsd   CCR