IBM Spectrum Scale & ESS
Global Data Platform

IBM Storage Software Product Management
—

Matthew Geiser
IBM Storage for Data, AI & HPC
Slack: @Matt Geiser Email: geiser@ibm.com

The Need For A Global Data Platform

IBM Spectrum Scale

## What's driving the need for a global data platform?

Data centric application development

- Example: The emergence of AI/ML use cases
    - Data hungry apps and GPUs, need access to more data, faster
    - As new applications and use cases roll out, data silos occur
    - Need unified and consistent approach to accessing data throughout AI/ML Pipeline

- Data fabric initiatives with requirements to provide consistent services across diverse infrastructure

The diverse IT infrastructure options available

- Many choices, From edge to core data center to public cloud

- Containers to simplify hybrid cloud infrastructure choices

- Drives the need for a single source of truth across diverse infrastructure that facilitates secure access while eliminating data redundancy and inconsistencies.

| HPC | AI / ML | Analytics | Enterprise | Containers | Backup / Archive |

| Edge | Core | Cloud |
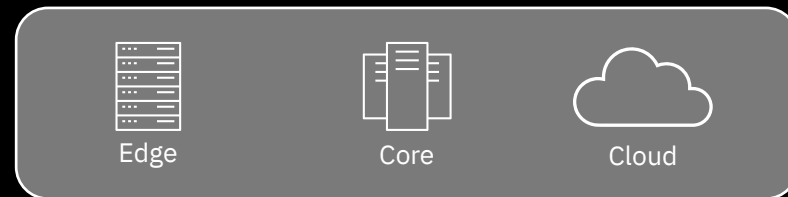
IBM Global Data Platform for Unstructured File & Object Data
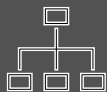Unstructured Data Services Framework

Applications and Workloads

Data Access Services

Data Caching Services

Data Management Services

Data Security Services

IBM's Global Data Platform for File & Object Data

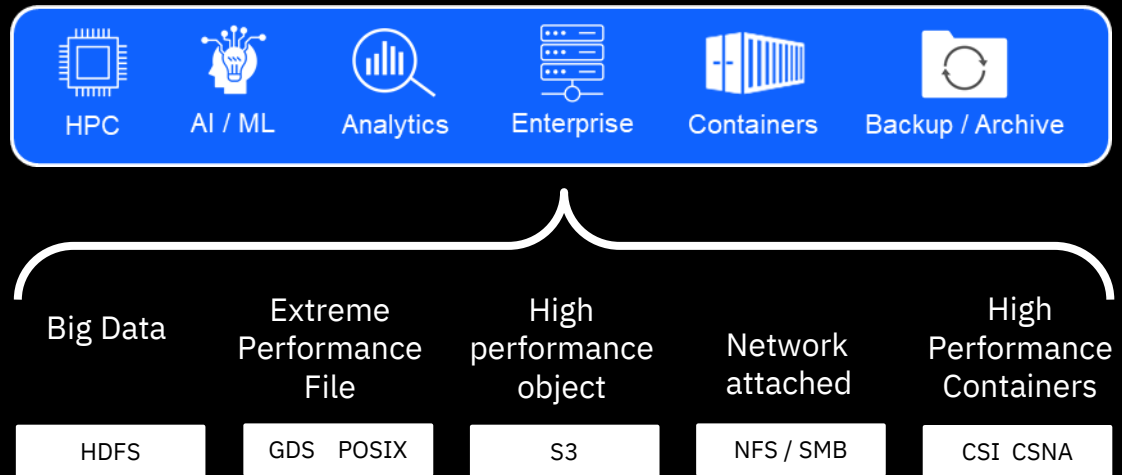
IBM Storage / © 2022 IBM Corporation

Data Access Services - Let your application requirements define the interface to use

Access to the same data, at the same time, using whichever protocol the workload requires

- **File Access** – POSIX, Linux/UNIX – NFS, Windows – SMB

- **Object Access** - S3, Hadoop/HDFS

- **Containers Access** – Container Storage Interface (CSI)

- **GPU-Direct Access** – GPU-Direct Storage (GDS)

Multiple interfaces access the same data

HPC	AI / ML	Analytics	Enterprise	Containers	Backup / Archive

Big Data	Extreme Performance File	High performance object	Network attached	High Performance Containers
HDFS	GDS POSIX	S3	NFS / SMB	CSI CSNA

Data Caching Services with active file/object management

Spectrum Scale Active File Management - Transparent data caching , enabling tiering and sharing of data across clusters

- **Investment protection** - Break down storage silos, easily leverage multi-vendor and multi-cloud resources

- **Increase application agility** - Accessing data from edge to core to cloud

- **Quickly scale your data** - From resources you choose with performance you require

- **Faster access to remote data** - transparently caching remote data locally when needed

Spectrum Scale AFM – Use Cases

Data Virtualization	Data Collaboration	Data Resilience	Hybrid cloud / Bursting
• Integrate legacy file and object data stores into a single file system to breakdown legacy data silos	• Geo-distributed collaboration on data transparently shared between data centers, the cloud and edge sites	• Provides a Disaster Recovery solution for business continuity	• Dynamically increase computation resources in the cloud or at another site
• Create a **High-Performance Tier** for analytics	• Consistent cache provides a single source of truth with no stale data copies	• Create an Active-Passive site relationship with failover and automatic data reconciliation on failback	• Burst site sees all data at home site and fetches data transparently on demand

Data Management Services to address distributed storage challenges and optimize time to result
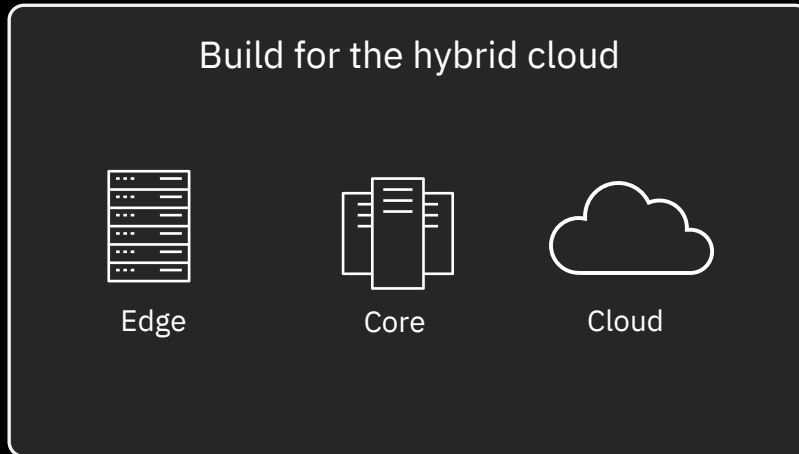
Data Cataloging and Orchestration Services

- Meta data catalog for file and object data

- Utilizes augmentation of metadata to enable dynamic, intelligent and automated data orchestration

Data Lifecycle Services

- Policy engine enables organizations to define rules for optimizes how long file and object is stored on differing classes of storage infrastructure

Data Retention and Archiving Services

- Services facilitating retention and archiving of file an object data

HPC · AI / ML · Analytics · Enterprise · Containers · Backup / Archive

Spectrum Scale · Spectrum Fusion · Spectrum Discover

Build for the hybrid cloud

Edge · Core · Cloud

Data Security Services - Active Protection for Cyber Resiliency

IDENTIFY

- Cyber Resiliency Assessment Tool, Probes 100s of different controls and best practices

PROTECT

- Multifactor Auth, RBAC, Privileged Access Monitoring (IBM Security Verify)
- Immutable snapshots, Logical air gap
- Scan snapshots for signs of ransomware
- Log all Admin & user actions

IBM Global Data Platform

RECOVER

- Instant Restore with Spectrum Scale AFM
- Spectrum Scale and Spectrum Protect – recover multi-petabyte filesystems in hours
- QRadar Incident Forensics

DETECT

- QRadar and Splunk SIEM integration
- File Audit Logging, Watch Folders
- Analyze backup data for signs of ransomware (Spectrum Protect)
- Reporting: QRadar User behavior analytics

RESPOND

- Automated action upon threat detection (QRadar)
  - Snapshot, Block Session , Etc..
- Alerts automatically prioritized based severity of the threat and criticality of the assets involved

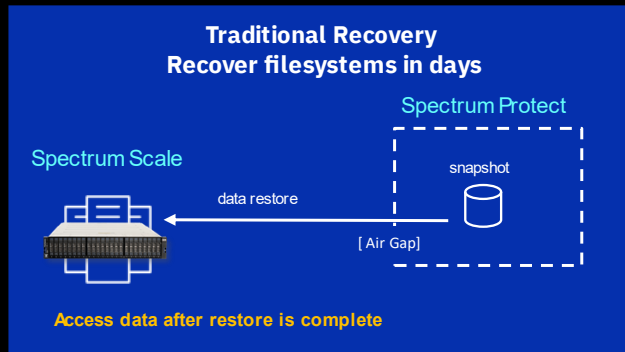# IBM Spectrum Scale – Active Protection for Cyber Resiliency

Recover: Quickly recover after a cyber attack
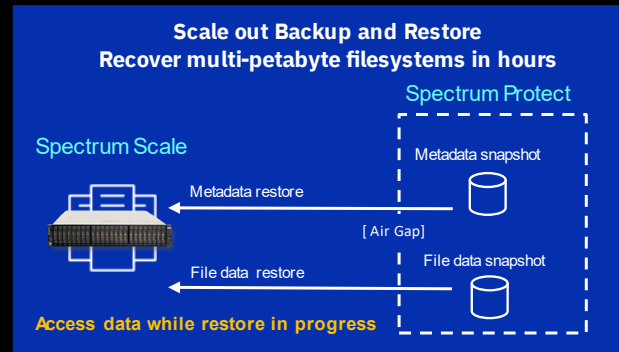
The evolution of recovery from ransomware attack
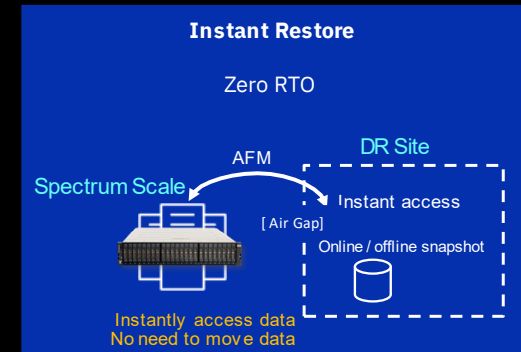
RTO - Days | RTO – Minutes to hours | RTO – Instant

Traditional Recovery
Recover filesystems in days

Spectrum Protect

Spectrum Scale

snapshot

data restore

[Air Gap]

Access data after restore is complete

- Spectrum Scale snapshot
- Recover from backup repository
- When restore complete can access data

Scale out Backup and Restore
Recover multi-petabyte filesystems in hours

Spectrum Protect

Spectrum Scale

Metadata snapshot

Metadata restore

[Air Gap]

File data restore

File data snapshot

Access data while restore in progress

- Snapshot metadata and file data separately
- Snapshot restore of file metadata
- Instant access to data stored in backup repository, while data is being restored

Instant Restore

Zero RTO

AFM

DR Site

Spectrum Scale

Instant access

[Air Gap]

Online / offline snapshot

Instantly access data
No need to move data

- Spectrum Scale AFM provides instant access to a read only snapshot
- Map a snapshot into a new volume so it is accessible without copying the data

IBM Elastic Storage System 3500

The simplest and fastest way to deploy a global data platform for AI and Hybrid Cloud workloads

Manage next generation and traditional workloads with simultaneous high-performance file and object data access services to the same data

Optimize local and remote access and simplify DR with global hybrid cloud data services

Speed access to critical data with Intelligent and automated data management services

Protect against cyber threats with Cyber-secure data services for unstructured data including end to end encryption and identification to recovery

Lower RTO times with proven data protection and data resiliency services

IBM Breaks Storage Performance Barriers for AI and Hybrid Cloud Workloads and Accelerates Recovery Times for Cyber Threats

IBM Spectrum Scale

up to **500+YB** per cluster
up to **30M IOPS** per rack
up to **91GB/s** per node
up to **1.8TB+**/s per rack

IBM

https://www.ibm.com/storage/artificial-intelligence