IBM

ADVANCED TECHNOLOGY GROUP (ATG)

Accelerate with ATG Webinar: IBM Storage Fusion – Data Foundation or Global Data Platform?

Shu Mookerjee

ATG Senior Storage Technical Specialist

Shu.Mookerjee@ibm.com

IBM

Accelerate with ATG Technical Webinar Series

Advanced Technology Group experts cover a variety of technical topics.

Audience: Clients who have or are considering acquiring IBM Storage solutions. Business Partners and IBMers are also welcome.

To automatically receive announcements of upcoming Accelerate with IBM Storage webinars, Clients, Business Partners and IBMers are welcome to send an email request to accelerate-join@hursley.ibm.com.

2024 Upcoming Webinars – Register Here!

IBM Storage Virtualize CDR 8.6.1 Technical Update - February 27th, 2024

IBM TS7700 Tape Solution Overview 101 - March 12th, 2024

IBM Storage Virtualize CDR 8.6.2 Technical Update – March 26th, 2024

Important Links to bookmark:

ATG Accelerate Site: https://ibm.biz/BdSUFN

ATG MediaCenter Channel: https://ibm.biz/BdfEgQ

Offerings

Client Technical Workshops

- IBM DS8900F Advanced Functions
- **IBM Cyber Resiliency with IBM Storage Defender:** Feb. 14-15 in Dallas, TX
- **IBM Storage Fusion:** Feb. 21-22 in Dallas, TX
- **IBM Storage Scale & Storage Scale System:** Feb. 28-29 in Costa Mesa, CA
- **IBM FlashSystem Deep Dive & Advanced Functions:** March 6-7 in Dallas, TX
- **IBM Storage Ceph:** March 20-21 in Costa Mesa, CA
- IBM FlashSystem and Storage Virtualize
- IBM Cloud Object Storage

TechZone Test Drive / Demo's

- IBM Storage Scale and Storage Scale System GUI
- IBM Storage Virtualize Test Drive
- IBM DS8900F Storage Management Test Drive
- Managing Copy Services on the DS8000 Using IBM Copy Services Manager Test Drive
- IBM DS8900F Safeguarded Copy (SGC) Test Drive
- IBM Cloud Object Storage Test Drive - (Appliance based)
- IBM Cloud Object Storage Test Drive - (VMware based)
- IBM Storage Protect Live Test Drive
- IBM Storage Ceph Test Drive - (VMware based)

Please reach out to your IBM Representative or Business Partner for more information.

***IMPORTANT* The ATG team serves clients and Business Partners in the Americas, concentrating on North America.**

Save the date

Storage @ IBM TechXchange Conference 2024

(Registrations open in March)

October 21-24, 2024

Mandalay Bay | Las Vegas

#IBMTechXchange

Key Learnings

- Practical how-to advice
- Patterns and best practices
- Success stories, IBM PoV, proven techniques

Featured Products

IBM Storage Defender | IBM Storage Fusion | IBM Storage Scale + IBM Storage Ceph | IBM Tape + IBM SAN | IBM Storage FlashSystem + IBM Storage DS8000

Collaborate. Learn. Play.

Community

IBM Champions | User Groups | Tech Peers | Business Partners

Sandbox

Network | Learn | Collaborate | Play

Roadmaps

Go deep with people in the know and set the stage for where IBM is going in the future

Breakout Sessions

Trends and Directions | User Groups | Product Deep Dives | Meet the Expert | Professional Development | Show the Code | Birds of a Feather | Academic/Research

Accelerate your Career

Labs (Instructor-Led, Self-paced) | IBM Certification Testing | Earn up to 25 hours in CPE credits

https://www.ibm.com/community/ibm-techxchange-conference/

Game On! →

Accelerate with ATG Survey

Please take a moment to share your feedback with our team!

You can access this 6-question survey via Menti.com with code 1708 6924 or

Direct link https://www.menti.com/alwhyze7z1gz

Or

QR Code

ADVANCED TECHNOLOGY GROUP (ATG)

Accelerate with ATG Webinar: IBM Storage Fusion – Data Foundation or Global Data Platform?

Shu Mookerjee

ATG Senior Storage Technical Specialist

Shu.Mookerjee@ibm.com

Meet the Speaker

Shu Mookerjee is a Level 2 Certified Technical Specialist with over twenty years at IBM, working in a variety of roles including sales, management and technology. For the last decade, he has focused exclusively on storage and has been the co-author of four (4) Redbooks. Currently, Shu is part of the Advanced Technology Group where he provides education, technical guidance, Proofs of Concept and Proofs of Technology to IBMers, business partners and clients.

Meet the Panelist

Lloyd Dean is an IBM Principal Storage Technical Specialist in IBM Storage Solutions. Lloyd has held numerous senior technical roles at IBM during his 22 plus years at IBM. Lloyd most recently is leading efforts in the Advanced Technology Group as the IBM Storage for Red Hat OpenShift focal and as a Hybrid Cloud storage solution SME covering IBM Block, File and Object storage solutions and their use cases supporting IBM Cloud Paks.

Agenda

- Goals and Objectives
- IBM Storage Portfolio
- IBM Fusion Overview
- Global Data Platform
- Data Foundation
- Summary

Goals and Objectives

Objective:
Explore the differences between Fusion Data Foundation (DF) and Global Data Platform (GDP)

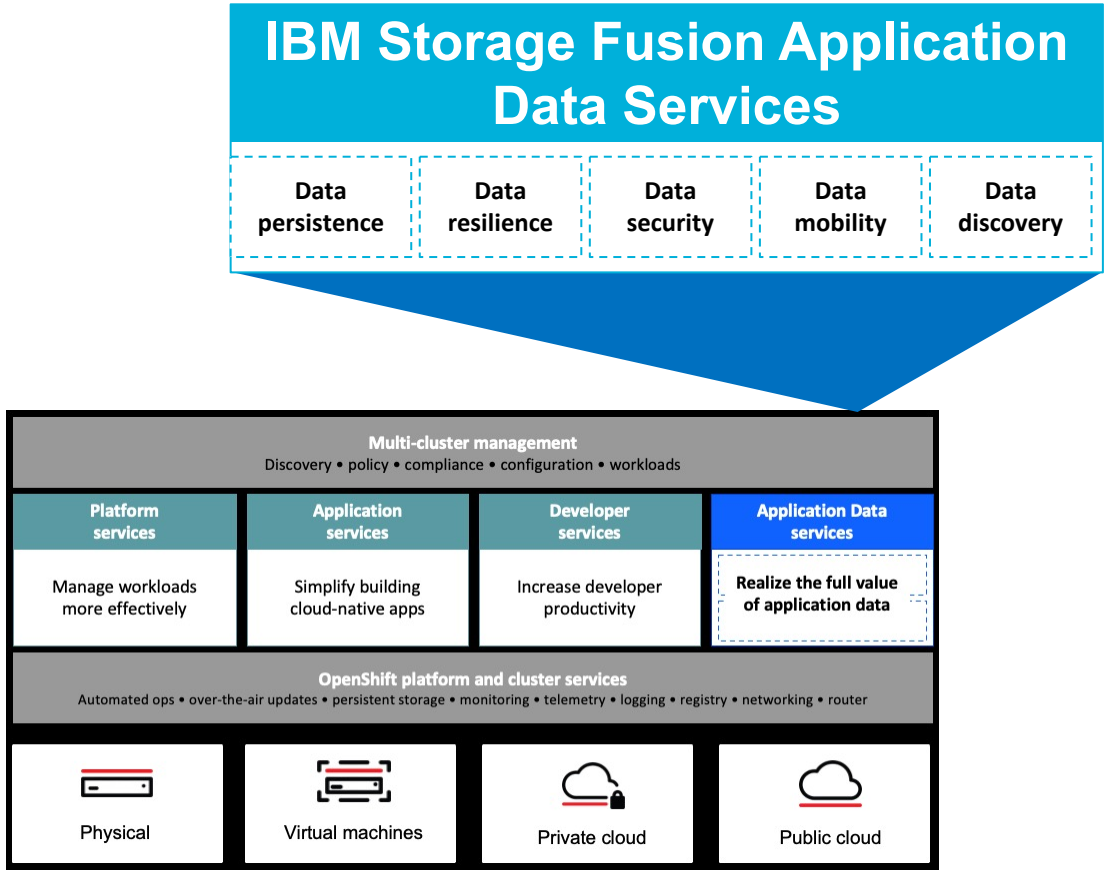
We WILL:
- Provide a brief overview of each storage service
- Cover their strengths
- Review their relevant replication and D/R technologies

We WILL Not:
- Do a deep dive into various Fusion configurations
- Cover implementation, installation or deployment processes
- Talk about competitive products or solutions

IBM Storage Portfolio

Providing solutions around	01 IBM Storage for App Modernization and Hybrid Cloud	02 IBM Storage for Data and AI	03 IBM Storage for Data Resiliency
	Drive innovation and scale application modernization with container-enabled enterprise storage that deploys seamlessly across hybrid infrastructures with a simple and consistent user experience.	Accelerate business results and innovation and unlock the latent value of unstructured data across the data ecosystem by eliminating data silos, advancing data discovery and classification.	Reduce the threat exposure window from days to hours and proactively safeguard data with a multi-faceted and scalable data resiliency approach that defends against cybervulnerabilities from detection to recovery.
Designed for	Chief Technology Officer | IT Director | VP OpenShift Engineering | Director Open Infrastructure | Cloud Architect | Data Scientist	VP of Engineering | VP of Development | Chief Data Officer | Data Architect | HPC Specialist	IT Architect | Storage Architect | Chief Information Security Officer | IT Director
Leading with	**IBM Storage Fusion**	**IBM Storage Scale** **IBM Storage Ceph**	**IBM Storage Defender**
Delivered on	Edge – to – Core – to – Cloud		
That run on sustainable infrastructure	**IBM Storage Fusion HCI System**	**IBM Storage Scale System**	**IBM Storage FlashSystem** **IBM Storage DS8K** **IBM Storage Tape**

IBM Storage Fusion Overview – Data Services



- **Data Persistence**
  - Persistent Volumes
  - File (RWX/RWO), block, S3 object. CSI compliant
- **Data Resilience**
  - Snapshots, backup / restore, recipes
  - Application consistent backups, lightweight spoke deployment
- **Data Security**
  - Storage Defender integration
  - Metro DR, Regional DR, 3-Zone deployments
- **Data Mobility**
  - Tiering, replication
  - Accelerate access to remote data
- **Data Discovery**
  - Data Cataloging
  - Scan and tag data residing anywhere in the organization for use in data science projects

IBM Storage Fusion Overview – Deployment Options

Customer Apps

IBM Cloud Paks

- Data
- Integration
- Security
- Cloud Satellite
- Business automation
- **Watson AIOps**
- Network automation

Databases

- Cassandra
- MongoDB
- RabbitMQ
- Elastic search
- PostgreSQL
- Spark

Off the shelf

- Pega
- Mulesoft
- TIBCO

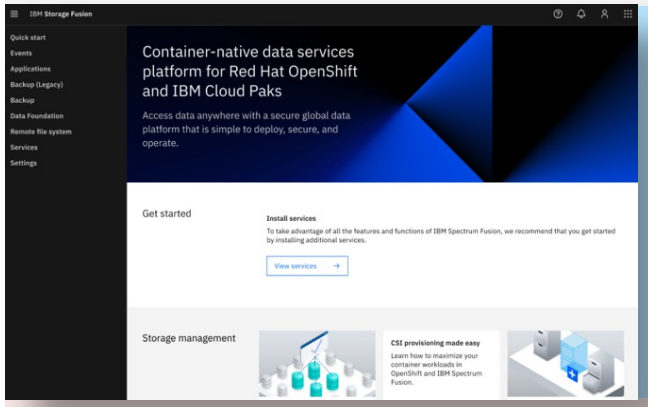
Custom apps

- Home grown

Offering

Fusion software

Data services for stateful OpenShift applications

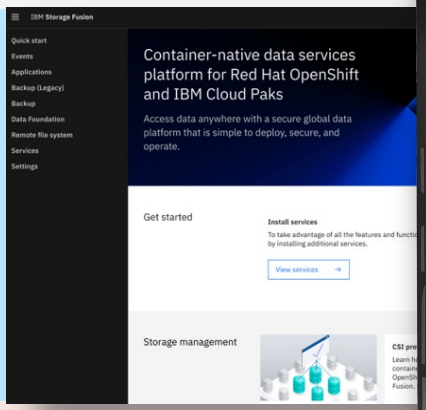


Consistent experiences

- APIs
- Fusion console
- Data protection
- Disaster Recovery
- Fusion Data Foundation (FDR)

Fusion HCI System

Integrated Application Platform for OpenShift



Built on

OPENSHIFT

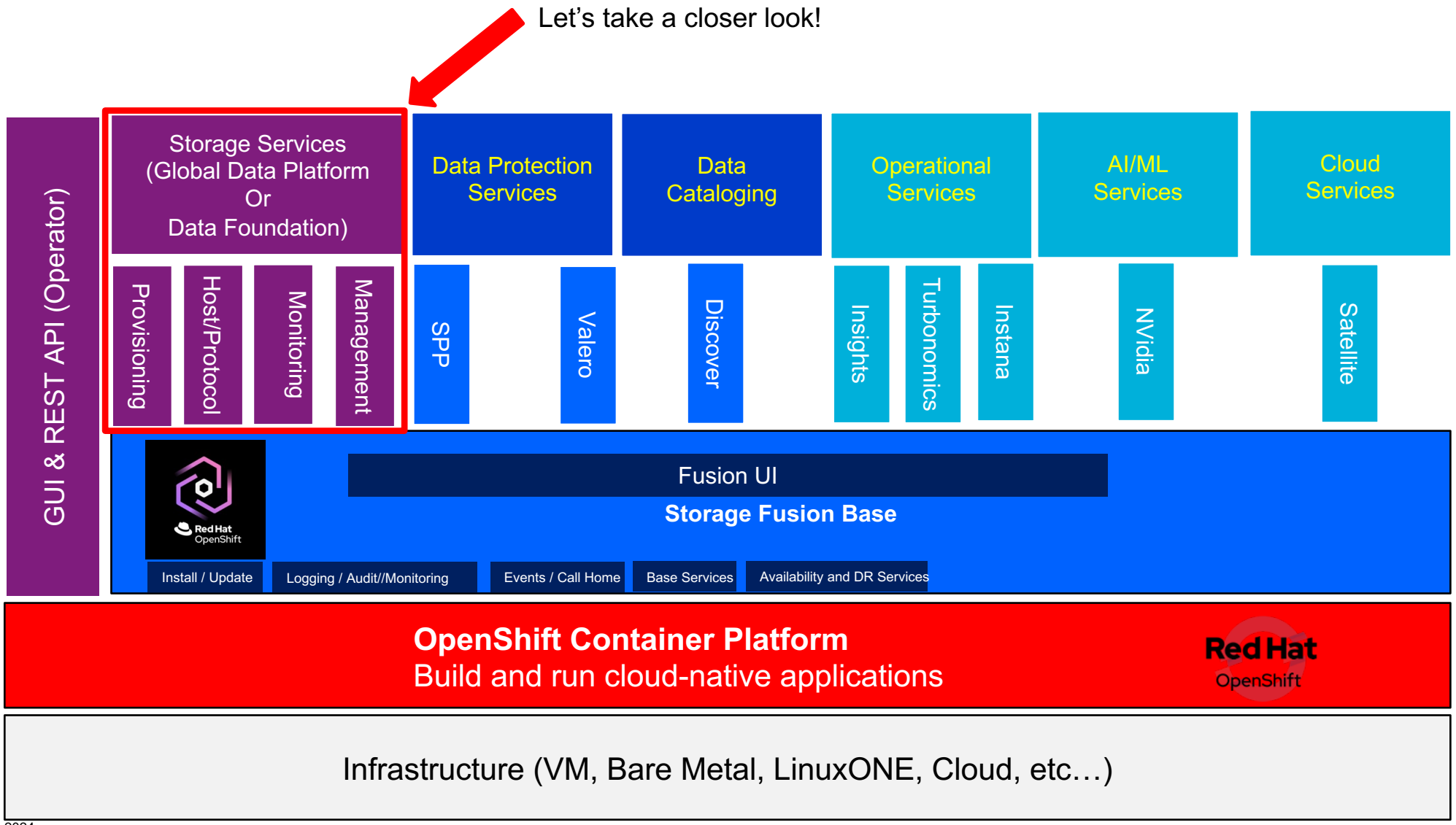
Deployments

AWS EBS	Azure block	Persistent disk	IBM block	SAN, vSAN	DS8K	FlashSystems	Bare metal Server-attached drives

Hyper-Converged Infrastructure for OpenShift

Switches, servers, storage

IBM Storage Fusion Overview – General Architecture



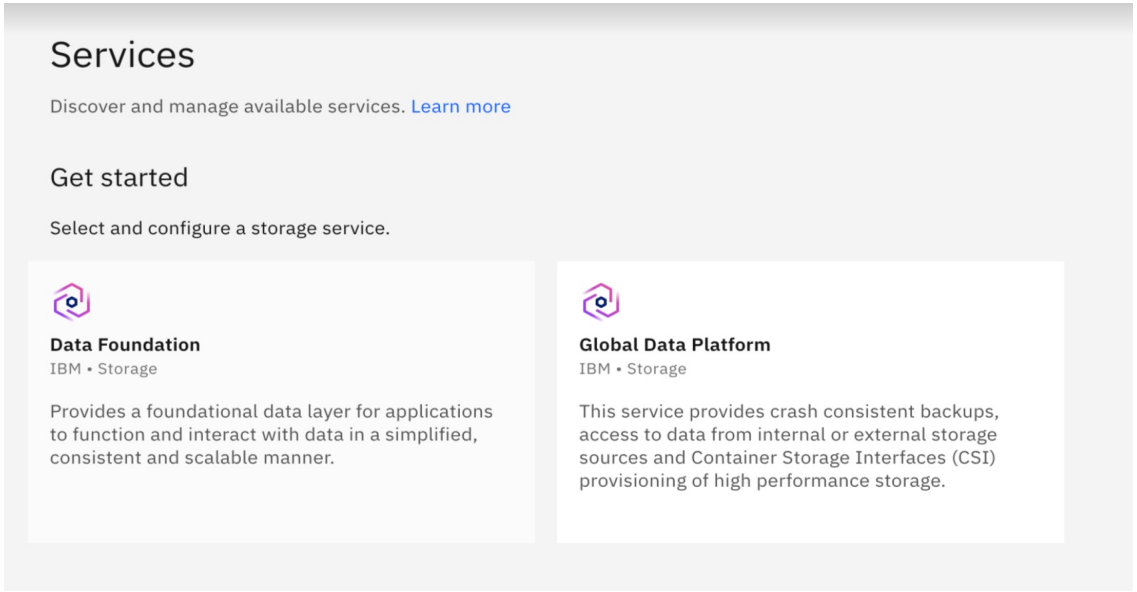
IBM Storage Fusion Overview – Storage Services Summary

Global Data Platform - Containerized IBM Storage Scale

- Storage efficiency with Storage Scale RAID
- Active File Management for geographic distance extension
- Scale out parallel file system
- Potential Use Cases:
  - Watsonx
  - Db2 Warehouse
  - Metro D/R or Regional D/R

Data Foundation - Ceph along with other containerized services

- Block, file and object storage (CephRBD, CephFS, RGW)
- Consistent architecture and storage classes across multiple infrastructures
- Simple lifecycle management
- Potential Use Cases:
  - Everything else!



Global Data Platform – Under the Covers

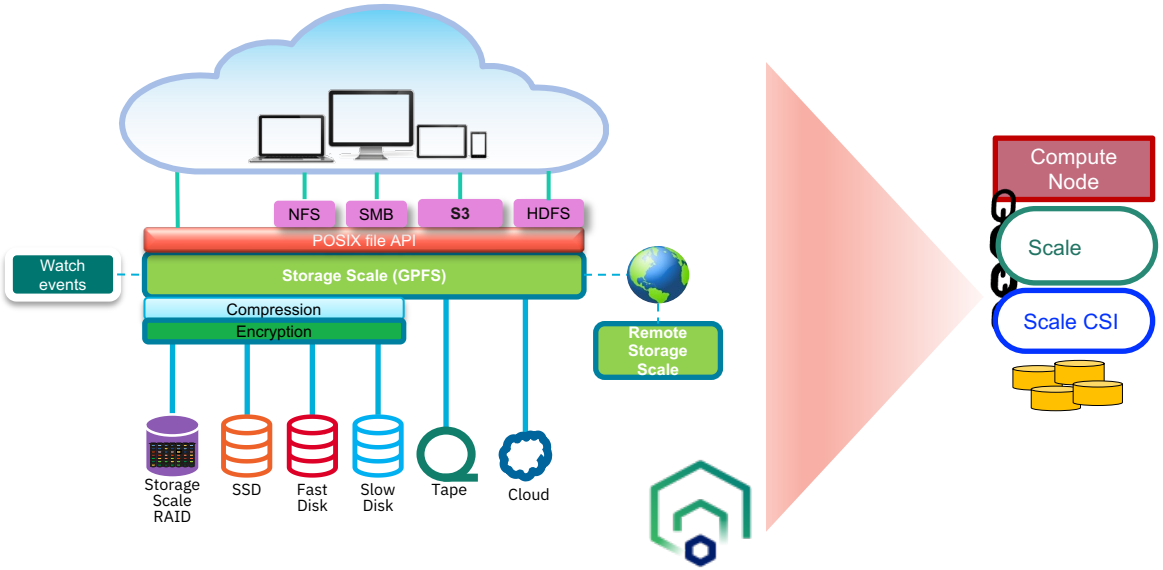
- **Global Data Platform** uses a containerized version of IBM Storage Scale Erasure Code Edition (Scale ECE) in HCI and Scale CSNA in Fusion software

- Storage Scale is a highly scalable distributed parallel POSIX ***file*** system

- Multiple Data Management features:
  - Built-in encryption and compression
  - Information Lifecycle Management for tiering and data placement
  - Support for multiple protocols (SMB,NFS, Object, HDFS)

- Cross Communication with existing Scale deployments
  - Remote cluster
  - Part of the same cluster
  - AFM target

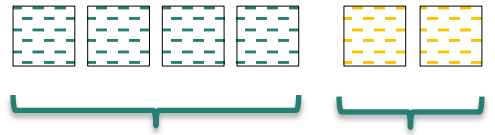
- Two unique technologies
  - Storage Scale RAID
  - Scale Replication (AFM and Stretch Cluster)



Global Data Platform – Under the Covers – Storage Scale RAID

- Storage Scale RAID is a data fault tolerance technology unique to ESS/SSS and Scale ECE
- Leverages a combination of erasure coding and declustured RAID
- Data integrity features
  - Checksum
  - Disk Hospital
- Erasure Coding
  - Divide blocks in to “strips”
  - Appends parity bits to each strip
  - Data blocks are then distributed via DRAID
- Declustured RAID (DRAID)
  - Utilizes all the drives in the array for data and parity
  - Improves fault tolerance
  - Greatly decreases the rebuild time

Reed-Solomon Erasure Coding 4+2p

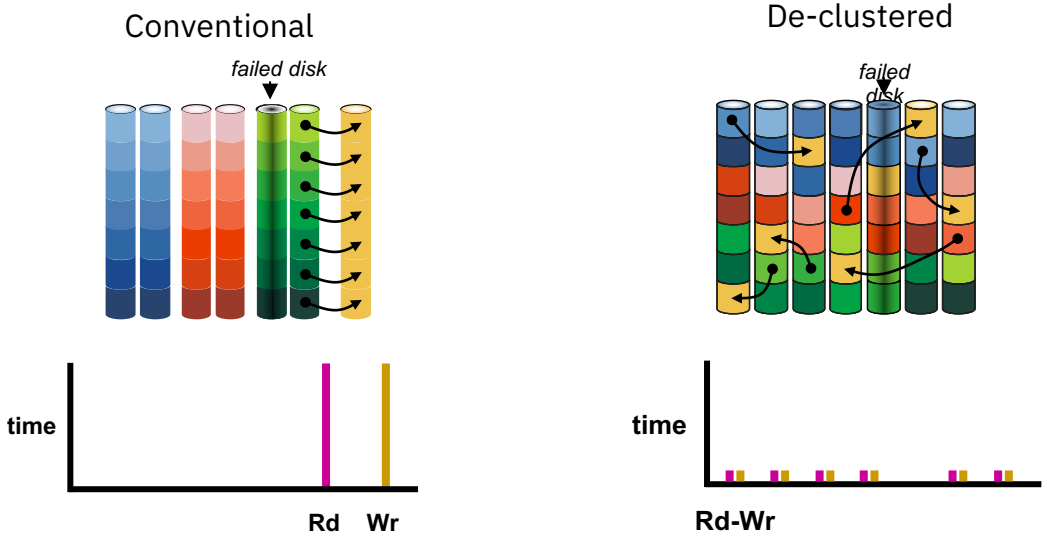


4 strips
(Scale block)

2 parity
strips

Tolerates
- 2 node failures, or
- 2 drive failures, or
- 1 node and 1 drive failure

Declustered RAID



Global Data Platform – Under the Covers – AFM and Scale Replication

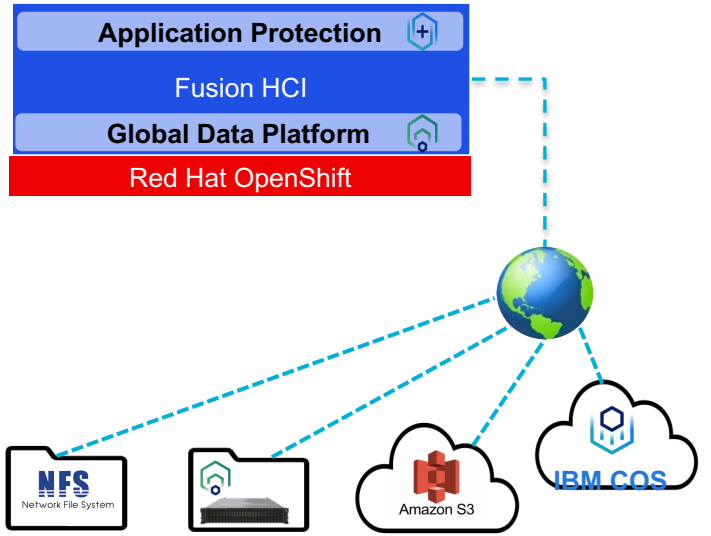
- Active File Management (AFM) is how Storage Scale extends across geographies

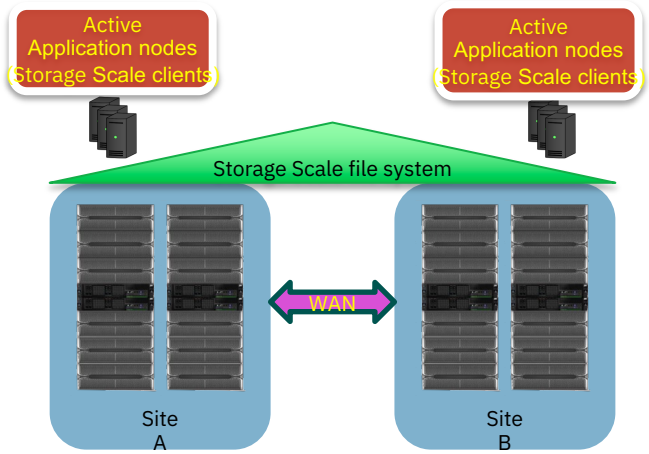
- High performance caching service
  - Mask WAN latencies
  - Ensure consistent read/write performance
  - No locking between caches
  - Enables failover/fail back

- AFM-DR is a specific deployment of AFM
  - Turns caching sideways!
  - Leverages asynch replication between sites
  - Active/passive

- AFM Alternative is Scale Stretch Cluster
  - Leverages synchronous replication
  - Active/active inter-site communication

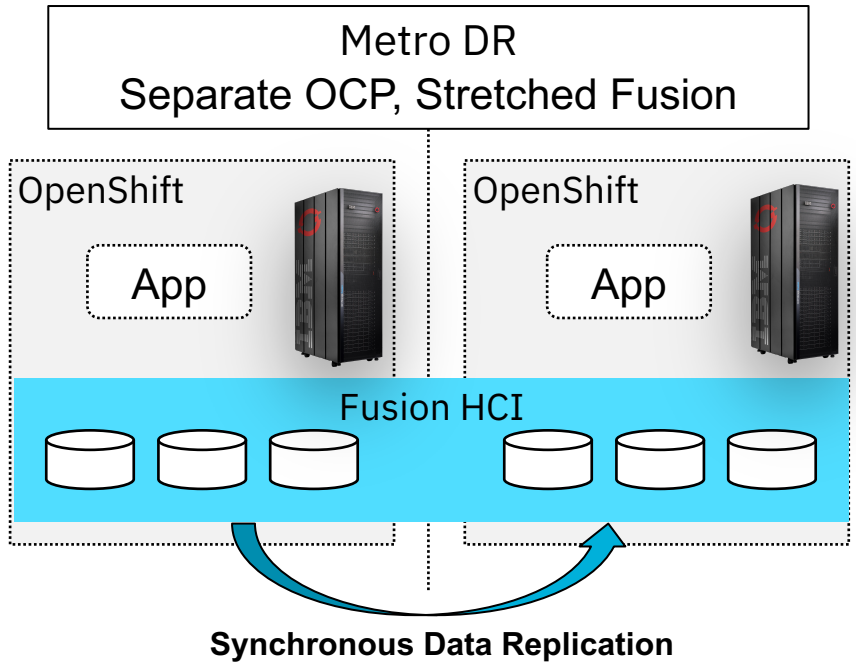
- GDP HCI Data Protection is based on these technologies





Global Data Platform – Metro-D/R (Fusion HCI)

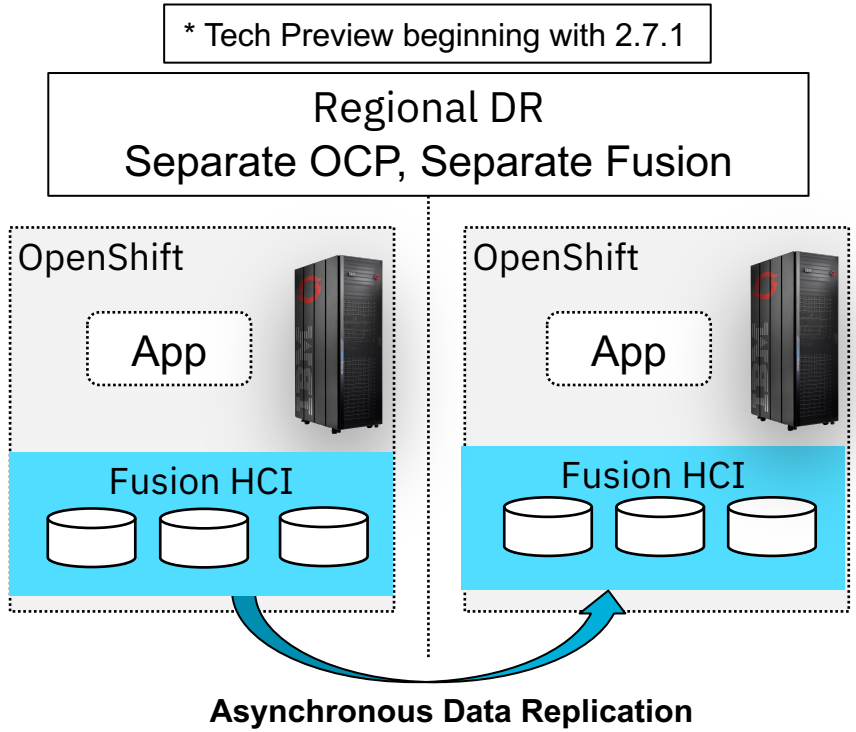
- Leverages Scale Stretched Cluster technology
- Use for distances of around 150 km/90 mi and roundtrip latencies of 40ms
- Data (volume and metadata) is synchronously replicated
- Enables:
  - Data protection at application granularity
  - Failover and failback applications without data loss (RPO = 0)
- Requirements:
  - Low link latency
  - Active/active deployment
  - Tie-breaker/quorum node (not shown)
  - Jumbo frame and proper network configuration
- Configuring Metro D/R with GDP



(1) To achieve RTO =0, the application must be able to accommodate active-active deployment

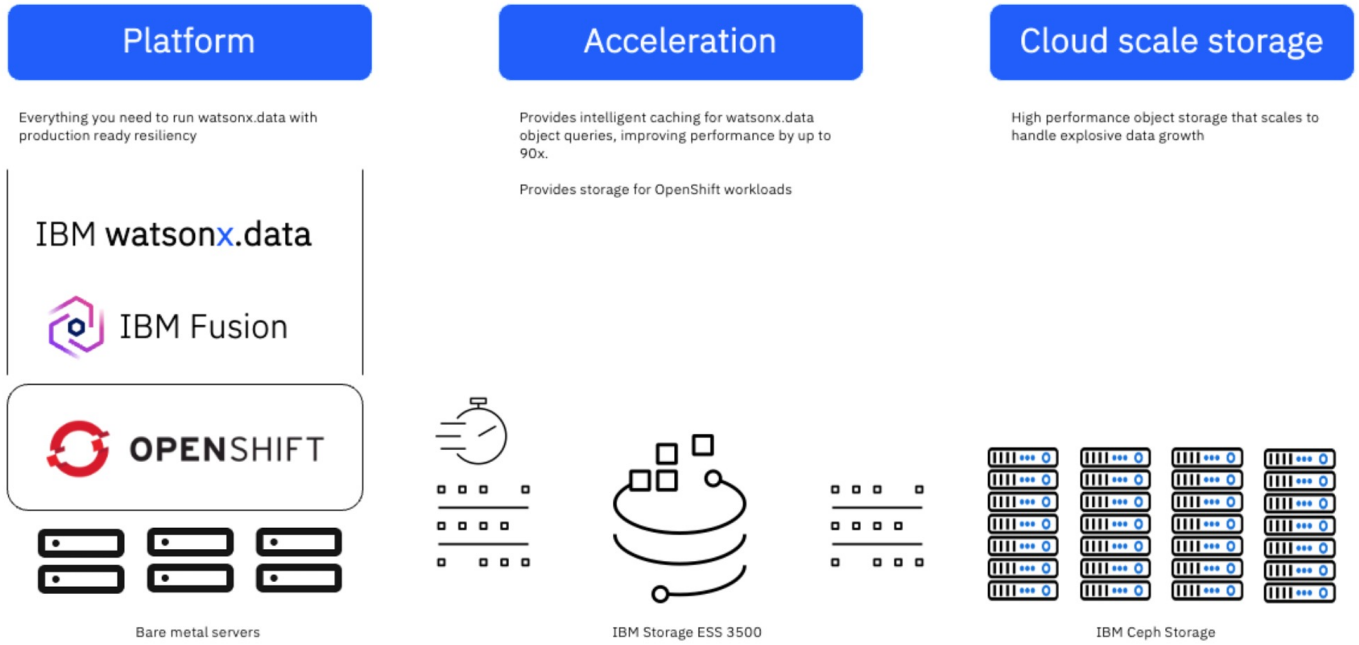
Global Data Platform – Regional D/R (Fusion HCI)

- Leverages Scale AFM-DR technology
- Use for distances greater than 150 km/90 mi
- Active (“cache”) site asynchronously replicates to passive read-only site
- RPO will be greater than zero
- Enables:
  - High network link latency tolerance
  - Long geographic distances
  - Failover/fail back between primary and secondary sites
- Requirements:
  - Fusion 2.7.1 or higher
  - Must have AFM and AFM gateways at both sites
- Configuring Regional D/R with GDP



Global Data Platform – AFM S3 Acceleration (Fusion HCI)

- Leverages Scale AFM-DR to accelerate cloud/S3 workloads
- Leverage external Scale as a high-speed cache tier
- Accelerate queries up to 90x
- Access data locally as file or S3 object
- Configurable policies:
  - Pre-fetch data or fetch data on demand
  - Configure policies to purge stale data
  - Read-only or read-write
- Perfect for AI/ML applications such as watsonx.data



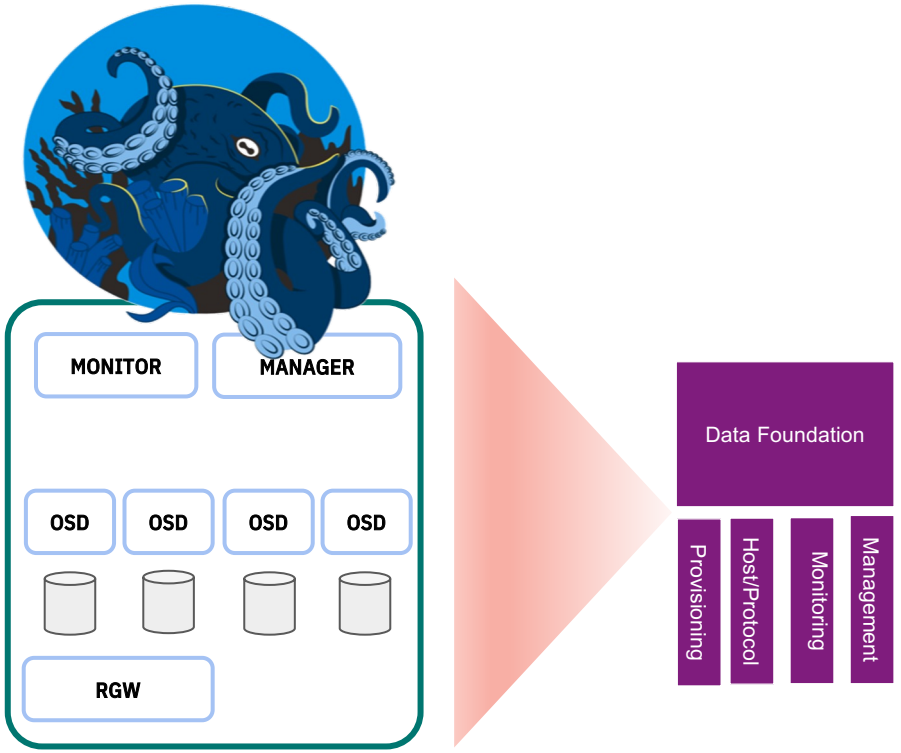
Data Foundation – Under the Covers

- **Data Foundation** consists of containerized storage services packages (including Ceph)

- Ceph is an open source distributed multi-protocol software-defined storage solution

- Multiple Data Management features:
  - Compression
  - Supports block, file and object storage
  - Policy based data lifecycle management
  - 3x Replication, (Erasure code supported on RGW only)

- Replication technologies available on *software deployment only*
  - Stretch cluster
  - Metro-DR (requires external IBM Storage Ceph)
  - Regional-DR (requires external IBM Storage Ceph)



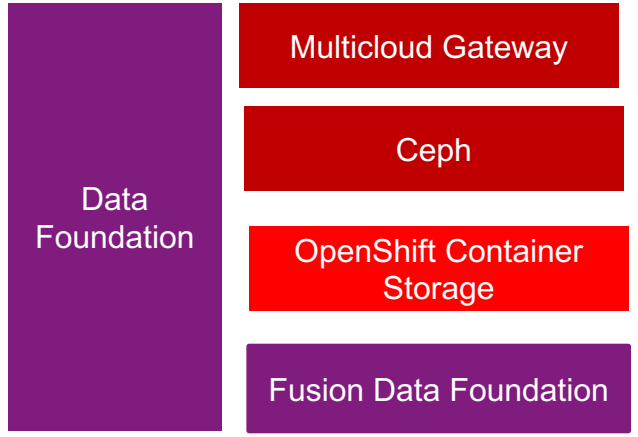
Data Foundation – Under the Covers – Architecture

- Ceph for the main storage services

- DF leveraging the following software packages:
  - Ceph – Provides main multiprotocol storage services
  - Ceph CSI – Manages provisioning and lifecycle of persistent volumes and claims
  - NooBaa - Provides a Multicloud Object Gateway (MCG)
  - Fusion Data Foundation Operators –Initialize and manage Fusion Data Foundation services

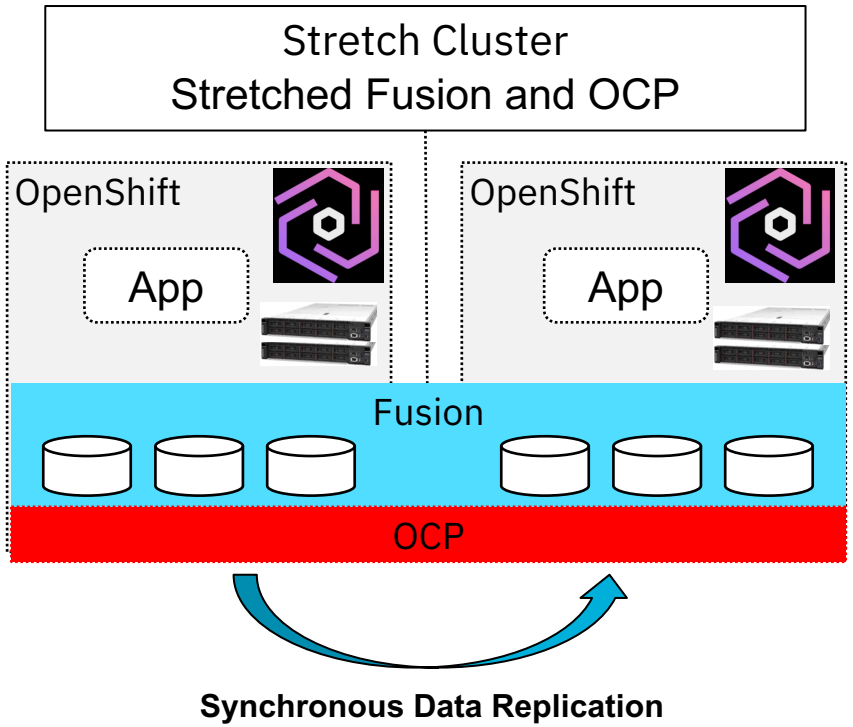
- Comprised of three (3) Operator Lifecycle Manager (OLM) bundles

- These deploy four (4) operators
  - odf-operator - Meta-operator. Deploys the Fusion Data Foundation storage services
  - ocs-operator – Meta operator. Creates resource definitions for Ceph and MCG
  - rook-ceph-operator – Deploys Ceph and Rook storage packages
  - mcg-operator – deploys the Multicloud Gateway



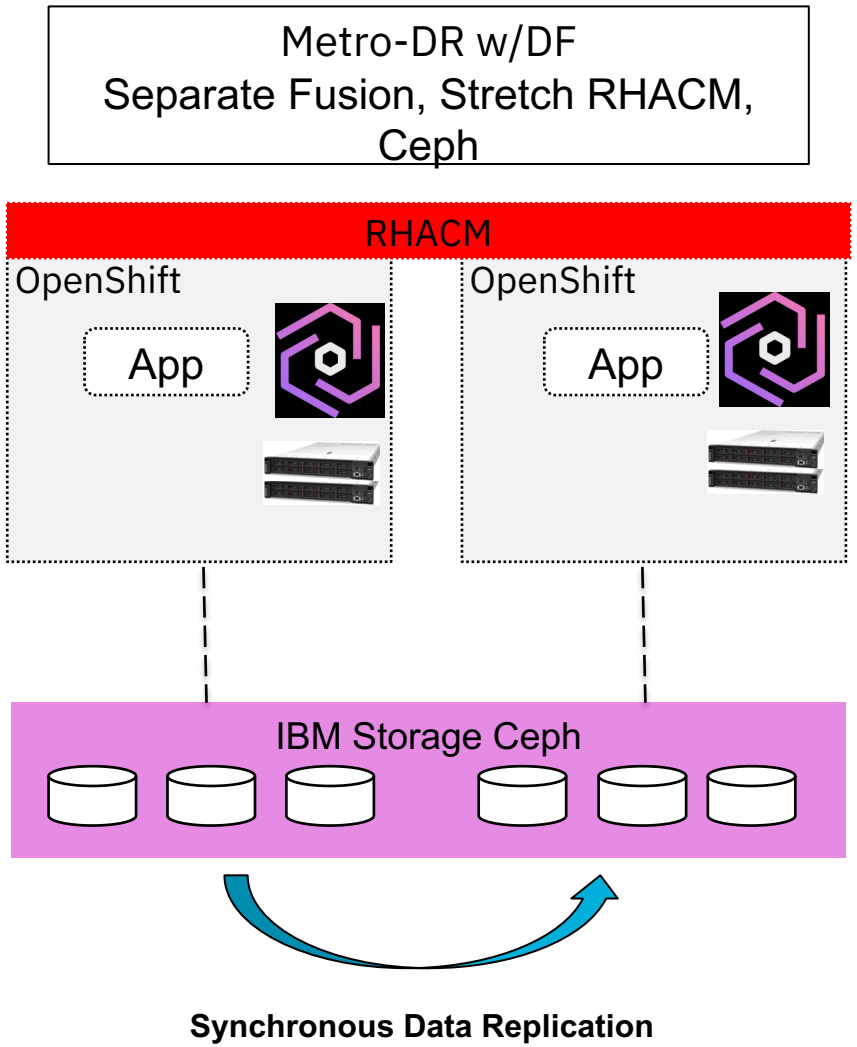
Data Foundation – Stretch Cluster (Software Only)

- Requires stretched Fusion and OCP
- Use for campus distances with roundtrip latencies of 10ms
- Enables:
  - HA/DR strategy with existing Fusion software
  - Failover and failback applications without data loss (RPO = 0)
- Requirements:
  - Three (3) master nodes/site
  - Four (4) worker nodes/site
  - SSDs for root drives if using bare-metal server deployments
  - Topology needs to be defined on Fusion AND OpenShift
- Configuring Stretch Cluster with Data Foundation



Data Foundation – Metro-DR w/Data Foundation (Software Only)

- Additional components are required in addition to DF
  - Red Hat Advanced Cluster Manager (RHACM)
  - OpenShift DR
  - External IBM Storage Ceph
- Use for campus distances roundtrip latencies of 10ms
- Enables:
  - Interop with existing IBM Ceph deployment
  - Failover and failback applications without data loss (RPO = 0)
- Requirements:
  - RHACM Hub cluster
  - Primary and secondary managed clusters
  - Ceph stretch cluster with arbiter (and related hardware/software)
- Configuring Metro-DR with Data Foundation
- NOTE: Regional-DR is similar except:
  - IBM Storage Replication is Asynchronous
  - Data Foundation Multicluster Orchestrator required
  - Supported with DF 4.14 and RHACM 2.9
- Configuring Regional-DR with Data Foundation



Summary - IBM Storage Fusion Storage Services for HCI – At-A-Glance

	Provisioning modes		DR replication		File		S3 Object	Raw Block [2]	HPC scale-out file system [3]	S3 Acceleration	Erasure coding
	Dynamic	Static	Metro	Regional [1]	RWX	RWO [2]					
Global Data Platform	✓	✓	✓	✓	✓	✓	✓	✗	✓	✓	✓
Data Foundation	✓	✓	✗	✗	✓	✓	✓	✓	✗	✗	✗
NFS	✗	✓	✗	✗	✓	✗	✗	✗	✗	✗	✗
vSphere Volume	✓	✓	✗	✗	✓	✓	✓ (3rd party)	✗	✗	✗	✗

1) Fusion GDP Regional DR is tech-preview beginning with 2.7.1 release. DF Regional DR is roadmap 1H24
2) Most applications that specify a need for block storage need RWO file storage, and not NFS storage, to avoid performance limitations in NFS. Relatively few applications need raw block storage.
3) High Performance Computing scale-out parallel file system

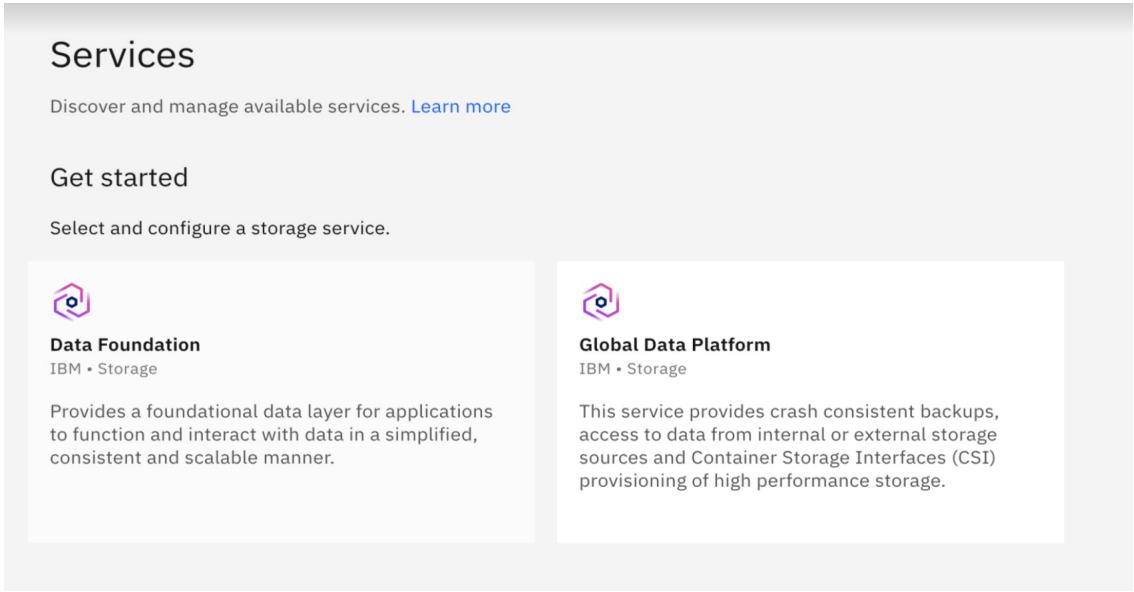
Summary – So Which One Do I Use?

Global Data Platform - Containerized IBM Storage Scale

- Fusion HCI Deployment (Scale ECE)
- Existing Scale environment (Scale CNSA)
- Large file systems/Databases
- HA/DR or replication requirements
- AI/ML

Data Foundation - Ceph along with other containerized services

- Fusion Software Deployment
- Existing Ceph environment
- Multi-protocol workloads, particularly block or object
- Multi-cloud infrastructures
- DevOps



“SHUPINION”

Thank you!

Accelerate with ATG Survey

Please take a moment to share your feedback with our team!

You can access this 6-question survey via Menti.com with code 1708 6924 or

Direct link https://www.menti.com/alwhyze7z1gz

Or

QR Code