A person is standing in a long, brightly lit server room aisle, looking at a laptop. The room is filled with rows of server racks on both sides, and the floor is made of metal grates. The lighting is cool and blue-toned, creating a high-tech atmosphere.





AFM NFS, S3 and S3 Glacier – San Jose UG 2025

Karrthik K G (kgkarrthik@in.ibm.com)
AFM Development

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.

Data Caching Services (Active File Management) Use cases

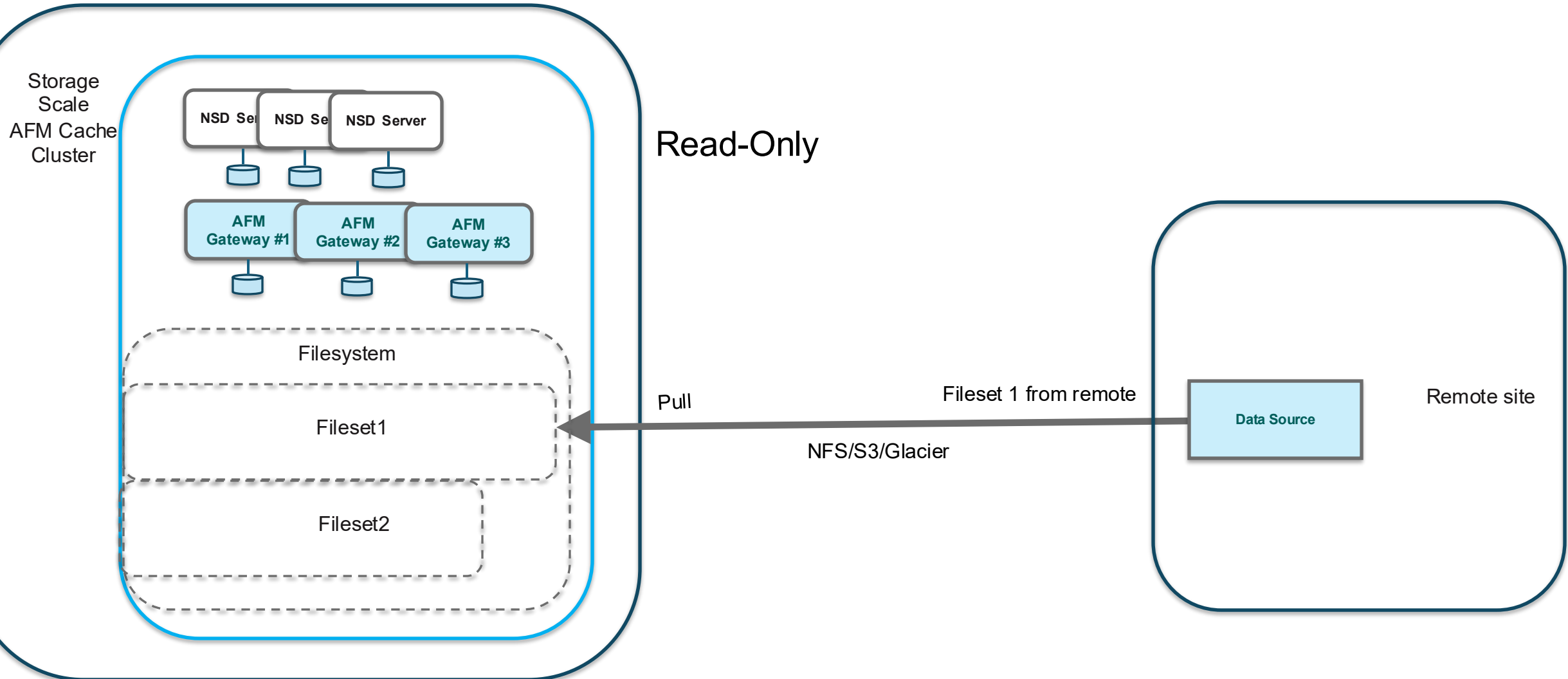
| | | | |
|---|--|--|---|
|  |  |  |  |
| Data Virtualization | Data Collaboration | Data Resilience | Hybrid cloud / Bursting |
| <ul style="list-style-type: none"> • Integrate legacy file and object data stores into single file system to breakdown legacy data silos • Migrate data to new storage or continue to use legacy stores • Create a High-Performance Tier for analytics for legacy data with transparent data access | <ul style="list-style-type: none"> • Geo-distributed collaboration on data transparently shared between data centers, the cloud and edge sites • Coalesce data to a home site from the edge and redistribute it to all sites | <ul style="list-style-type: none"> • Provide an asynchronous Disaster Recovery solution for business continuity over WAN distances • Supports analytics and archival access to passive data | <ul style="list-style-type: none"> • Dynamically increase computation resources in the cloud and optimally make required data available for Cloud bursting • Process data consolidated on S3 Cloud Storage on with high performance tier in the Cloud Compute Cluster • Archive data to S3 Object storage |
| Public Cloud Services Use case: Enables end user service to upload large amount of data via Object interface that can be analysed on high performance file system | Research / University Use case: Generate 100's of TB per day across multiple silos, leveraged AFM to provide common namespace with transparent multiprotocol data access | Multinational financial services Use case: Disaster Recovery, retention and compliance data with FileNet and ESS | Research Biopharmaceutical Use case: Multi site / public cloud bursting for collaboration |

AFM Concepts

- Node types
 - **Application node**
 - Writes/reads data based on application request to the GPFS filesystem at cache cluster
 - Can be Linux/AIX/Windows – NFS/SMB mounts
 - **Gateway node(s)** is the node that connects to the home cluster
 - Queue of pending operations is in memory at the gw node
 - Selected based on hash of filesetid or user-defined.
 - Reads/writes data from the home cluster to the cache cluster
 - Checks connectivity with the home cluster and changes to disconnected mode on connection outage
 - Triggers recovery on failure
 - Only Linux supported
- Sites
 - Home cluster (Object Store)
 - Exports a fileset that can be cached
 - Cache cluster
 - Runs AFM and “connects” a local fileset with the home fileset.
- Transport Protocol
 - NFS, NSD, S3 and Azure Blob, Glacier/Deep Archive

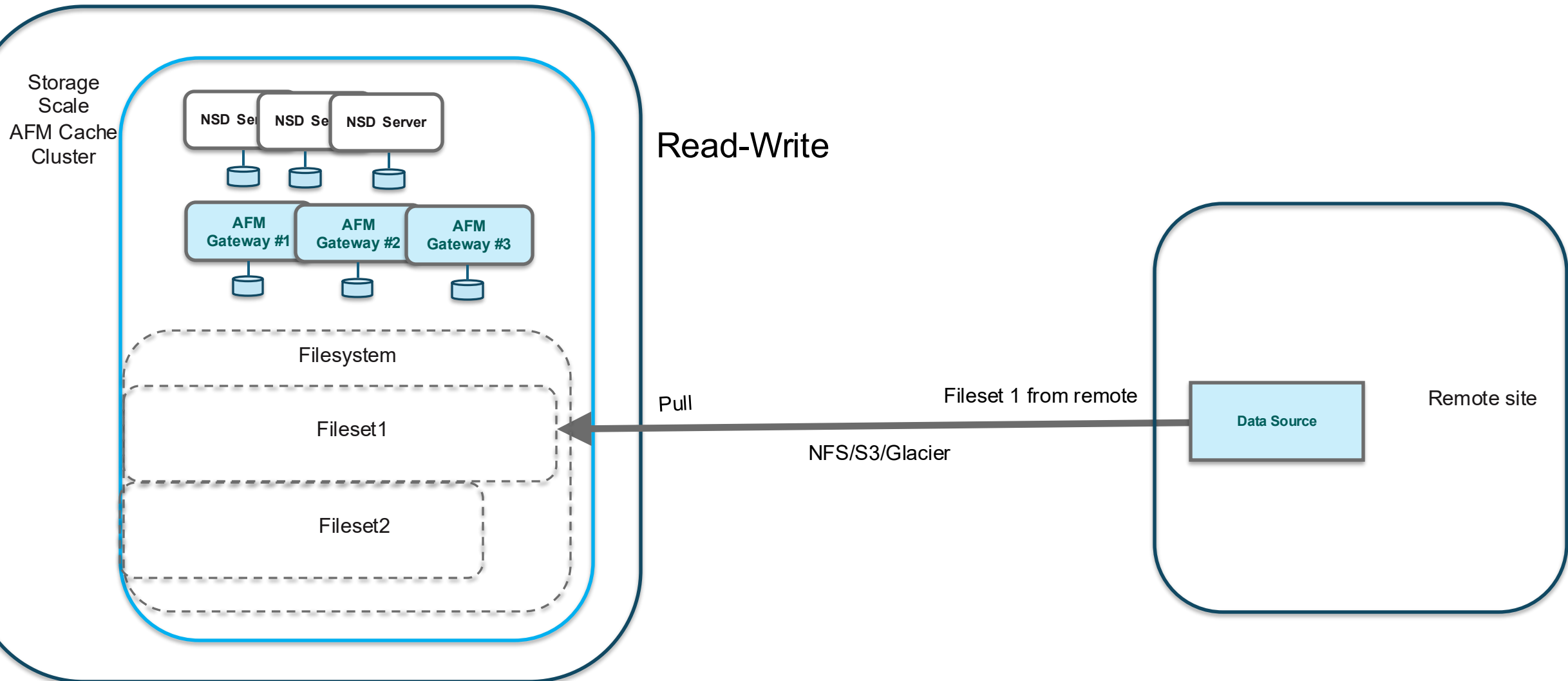
AFM Modes

Read-Only Mode



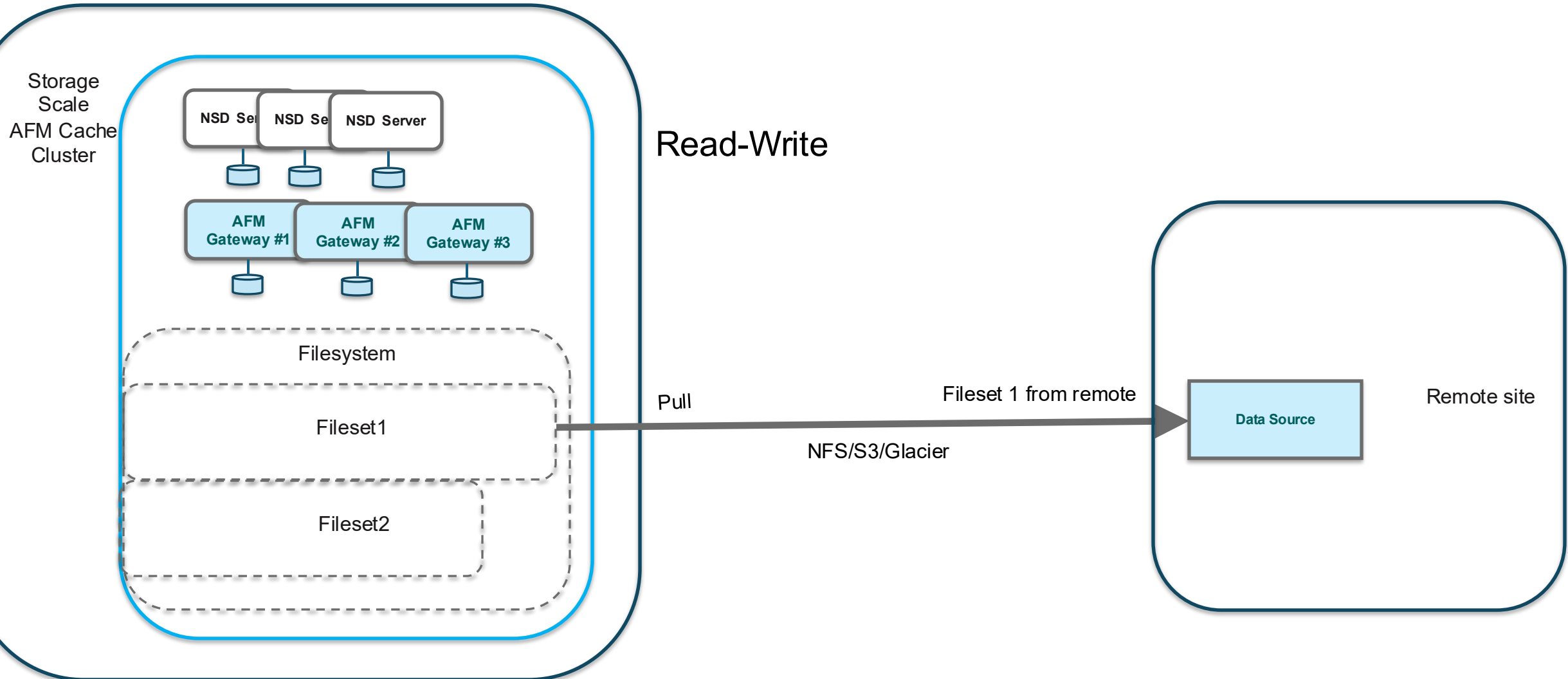
AFM Modes

Local-Updates Mode

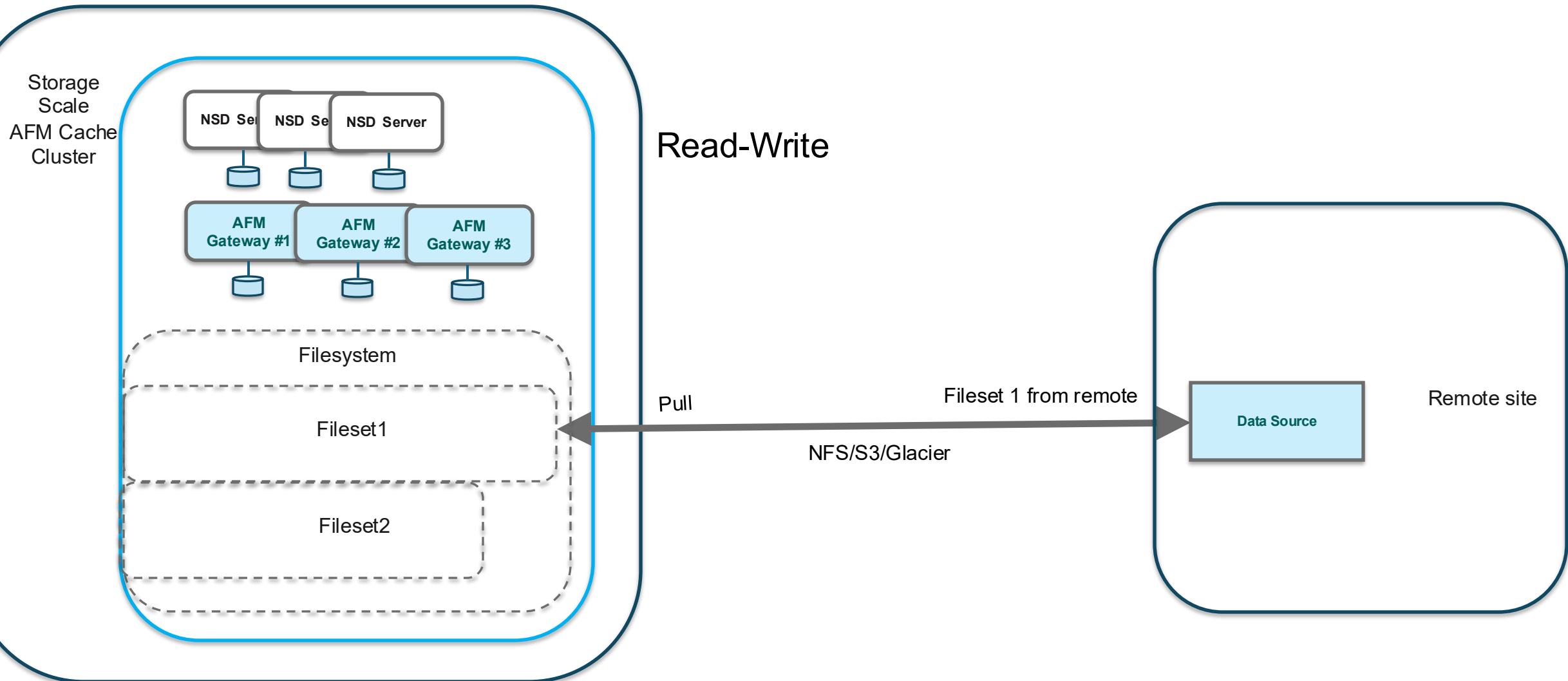


AFM Modes

Single-Writer Mode



AFM Modes Independent-Writer Mode



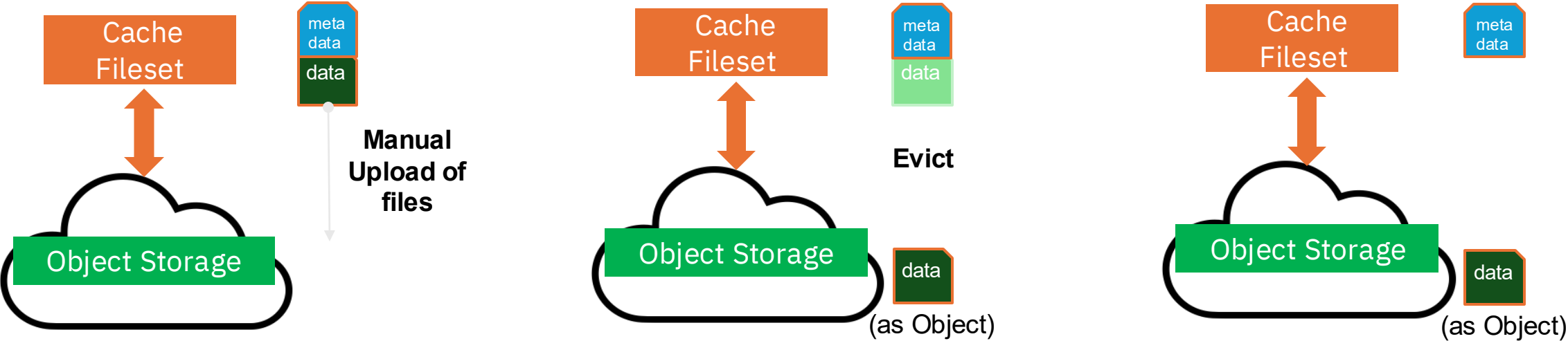
AFM Modes

AFM for archive use case with Manual Update (mu) mode

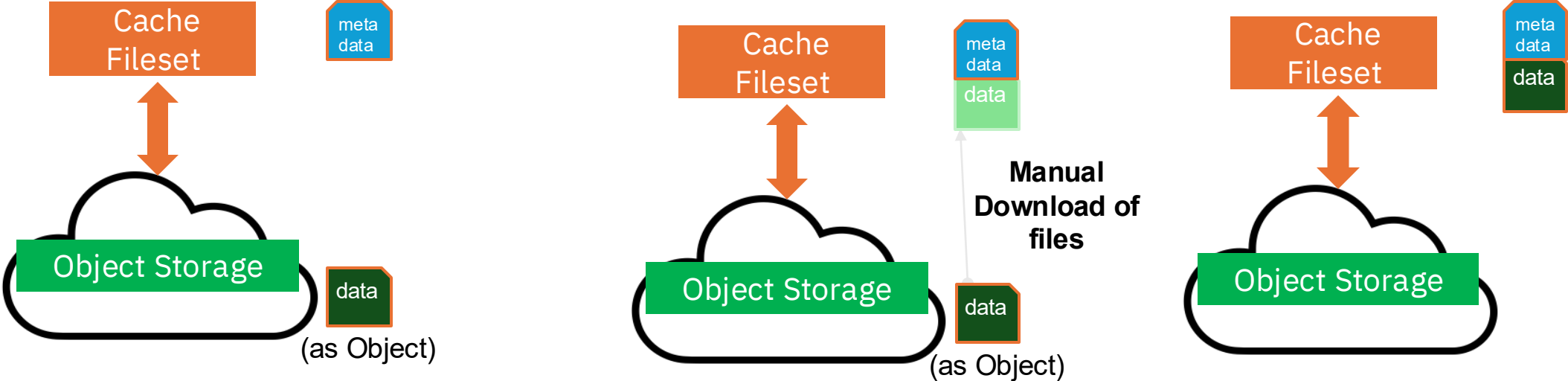
- Supports manual upload/ download of objects using ILM policies or object list and avoids automatic upload/ download
- All changes are local. Manual action is required to upload or download files to cloud object storage
- Metadata is refreshed only once when the MU fileset is created pointing to non-empty bucket
- Manual control over deletion from cache
- Manual deletion from Cloud Object Storage
- An independent fileset can be converted to a MU mode AFM fileset
- Auto removal of files/ objects from Cache and Object Storage using fileset parameter 'afmMuAutoRemove'
- Specify policy to be used with `mmafmcosctl reconcile -policy` command



AFM for archive use case with Manual Update (mu) mode



AFM for caching use case with Manual Update (mu) mode

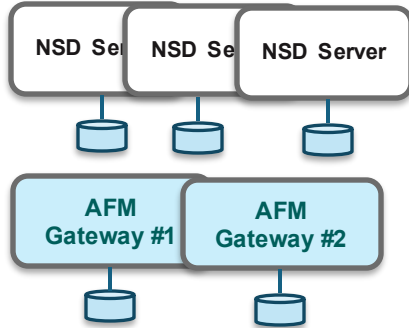


Multisite Replication with MU Mode

Virtual Private Cloud

Private Subnet

Storage Scale
AFM Cache
Cluster



Filesystem

Fileset1

Map1

Push/Pull

Push

Push

Fileset 1 to AWS S3

Fileset 1 to IBM COS

Fileset 1 to MINIO

Map1 → AWS/GW1,
ICOS/GW2,
MINIO/GW1

AWS S3
Storage

multitargetdemo

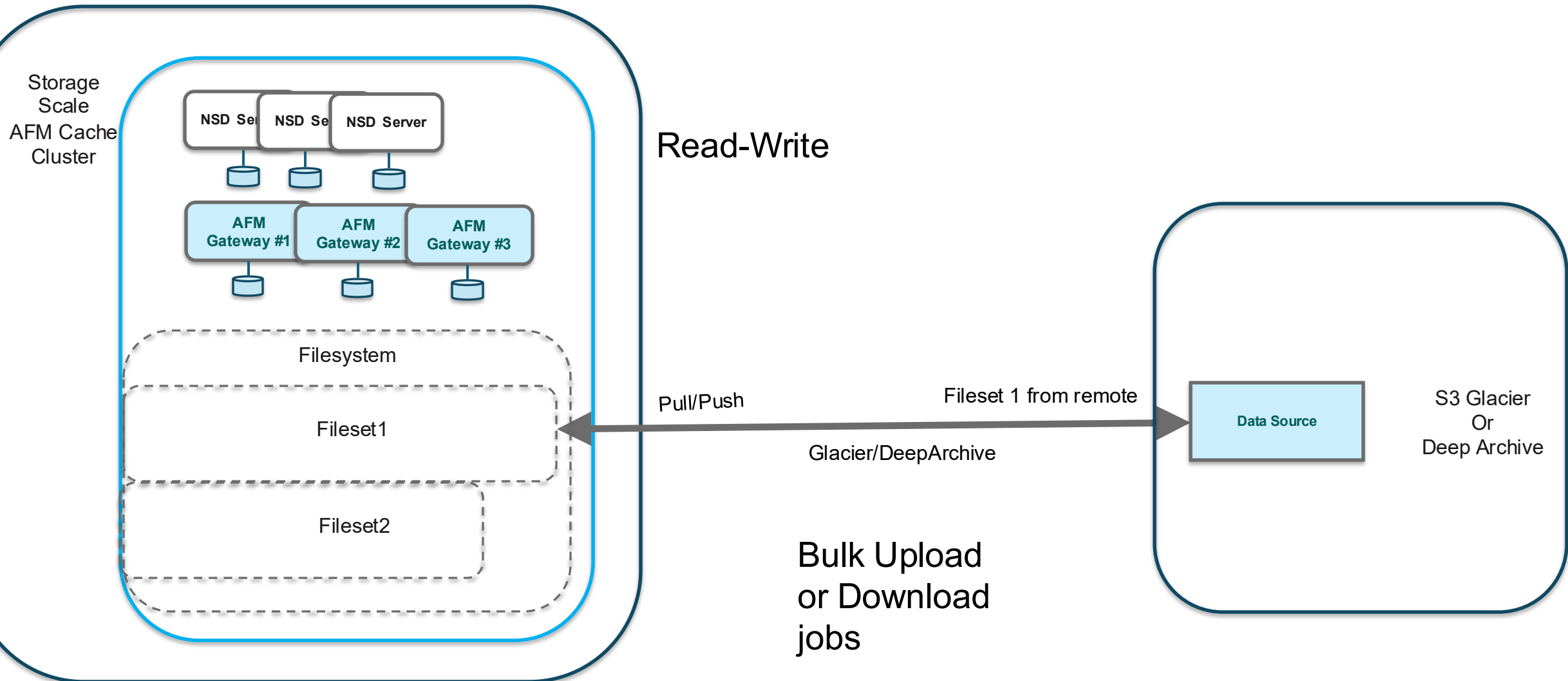
IBM COS

multitargetdemo

Minio

multitargetdemo

AFM \leftrightarrow S3 Glacier/Deep Archive



AFM Cloud Object Storage Operation modes

ObjectFS mode

- Behaves like normal AFM modes fileset.
- Objects are downloaded on read or on access
- IW and SW modes push files to cloud object storage
- RO, LU, IW automatically pulls objects from the cloud object storage and stores as files.

ObjectOnly mode

- Default for object operation mode
- No on-demand refresh on read
- Need to manually download metadata/data from COS.
- Objects are uploaded automatically (IW and SW)
- Avoids frequent trips and reduce network contention by selective download/uploads.

| | ObjectFS | ObjectOnly |
|-------------------------|---|--|
| Read Only (RO) | Upload - NA Download – On access (Auto) | Upload - NA Download - On demand |
| Local Update (LU) | Upload – NA (only On demand) Download – On access (Auto) | Upload – NA (only On demand) Download - On demand |
| Single Writer (SW) | Upload - Auto Download - On access / On demand | Upload - Auto Download – On demand |
| Independent Writer (IW) | Upload - Auto Download – On access (Auto) | Upload - Auto Download - On demand |
| Manual Update (MU) | Upload - On demand Download - On demand | |

Recent AFM S3 Object storage features

-
- Support of adding user defined prefix in AFM to cloud object storage fileset.
 - **# mmafmcosconfig fs1 afmbktprefix1 --endpoint https://region@endpoint --object-fs --xattr--prefix dir1 --bucket bkt1 --acls--mode sw**
 - Support of replicating more than 2K metadata in AFM to cloud object storage fileset.
 - Support for outband download of objects.
 - Immutability/ IAM modes support
 - Asynchronous Prefetch for CAS – 5.2.3
 - Support for S3 Glacier, IBM DeepArchive – 6.0.0
 - Porting AFM COS' V1 SDK base to V2 (Security item) – 6.0.0
 - Notification based cache revalidation from AWS S3 – 6.0.0

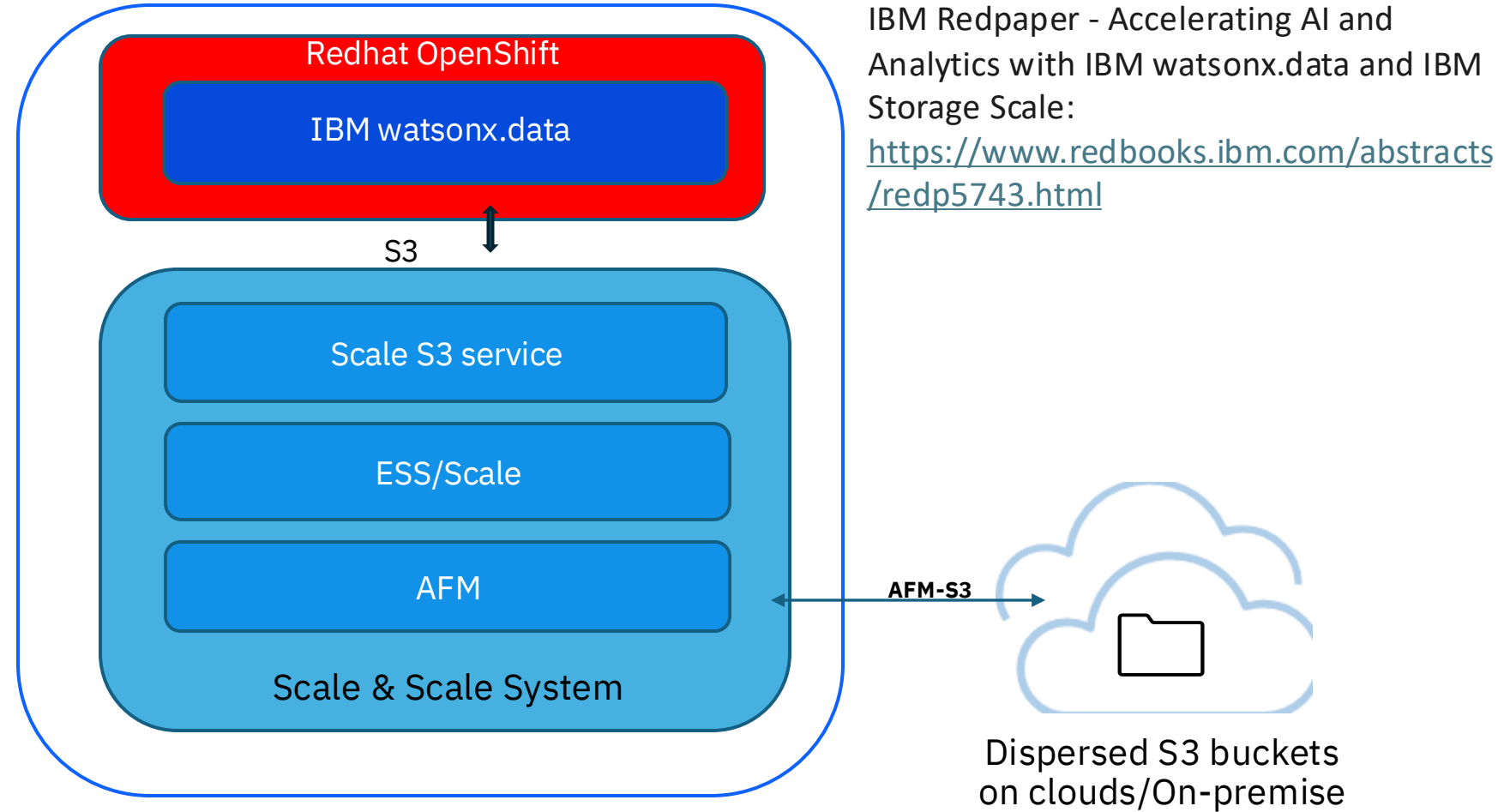
AFM – multi-site replication use cases

- Create multiple copies of the data in different regions. Process the data wherever the capacity/compute is available.
- AI workloads: Move only specific data-sets required for AI training based on the region.
- Move to multiple clouds, retrieve from the cheaper storage (minimal egress charges)

Analytics Use case : IBM watsonx.data + IBM Storage Scale (with AFM & CES S3)

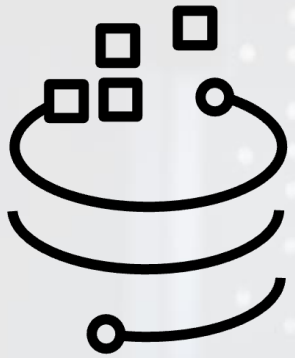
Data buckets could reside locally on the IBM Storage Scale file system, Or could be resident on remote locations e.g. – on various public clouds.

Storage Scale AFM provides a persistent cache layer to deliver high performance access to remote/slower/legacy data buckets



****IBM watsonx.data – Hybrid, open data lakehouse to power AI and analytics with all your data, anywhere.**

Thank you for using



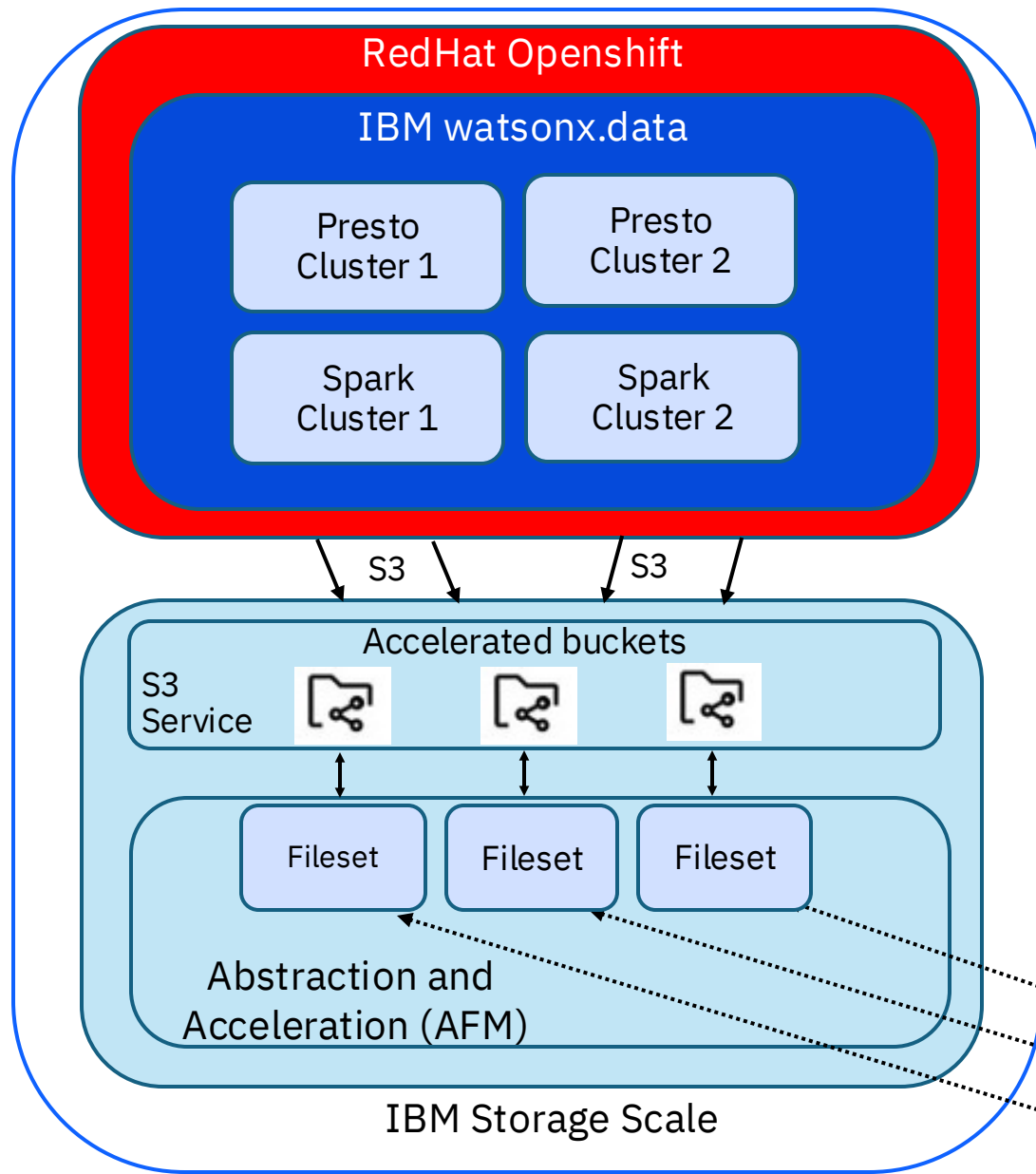
Storage Scale



Storage Scale
System

AFM S3/Azure limitations

- Nonempty directory renames are not supported
- Hard links can be created, but only one object is replicated



Analytics Use case : IBM watsonx.data + IBM Storage Scale (with AFM & CES S3)

AFM can virtualize remote S3 buckets to make them look like local buckets and act like a High Performance Storage Tier for your “Hot Data”, significantly accelerating query performance

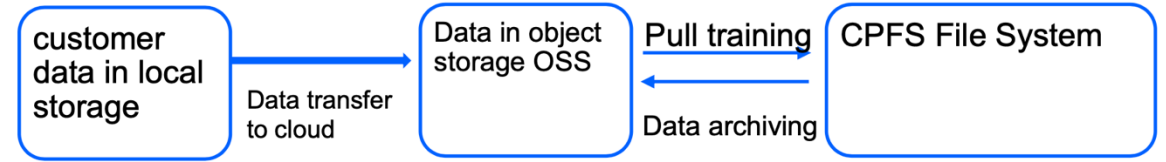
Cache available to all the engines and worker nodes in IBM watsonx.data. Cache survives engine restarts, available even for newly provisioned engines.

Cache is available to multiple protocols and not just to S3.

Huge reduction in public cloud data egress costs

Dispersed S3 buckets
on clouds/On-premise

Acceleration for S3 data on Cloud

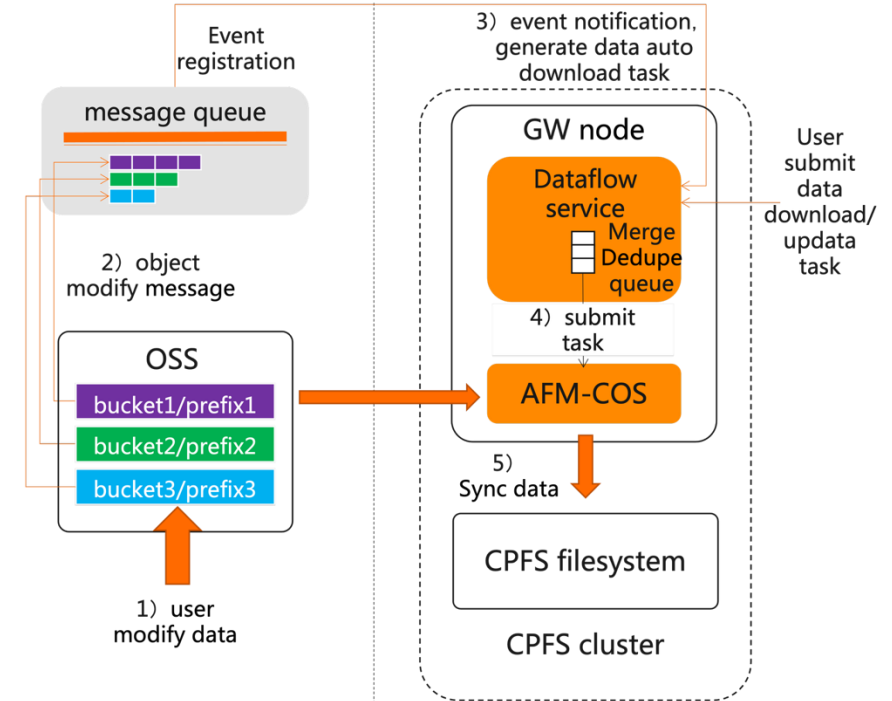


Solution for Data Analytics on Cloud with AFM + S3

- Data is ingested into Cloud S3 bucket by customer, end users
- AFM Caching / Acceleration is setup with S3 as backend storage
- Data is processed via Posix on CPFS (GPFS) filesystem

Use cases – ADAS, Media & Entertainment, Financials, HPC

Other customers considering similar solution: Other cloud providers in China, Financial customer on Azure (US), Top 10 Bank in USA on IBM Cloud



Solution offered by leading Cloud Provider in China

Active File management: Research example

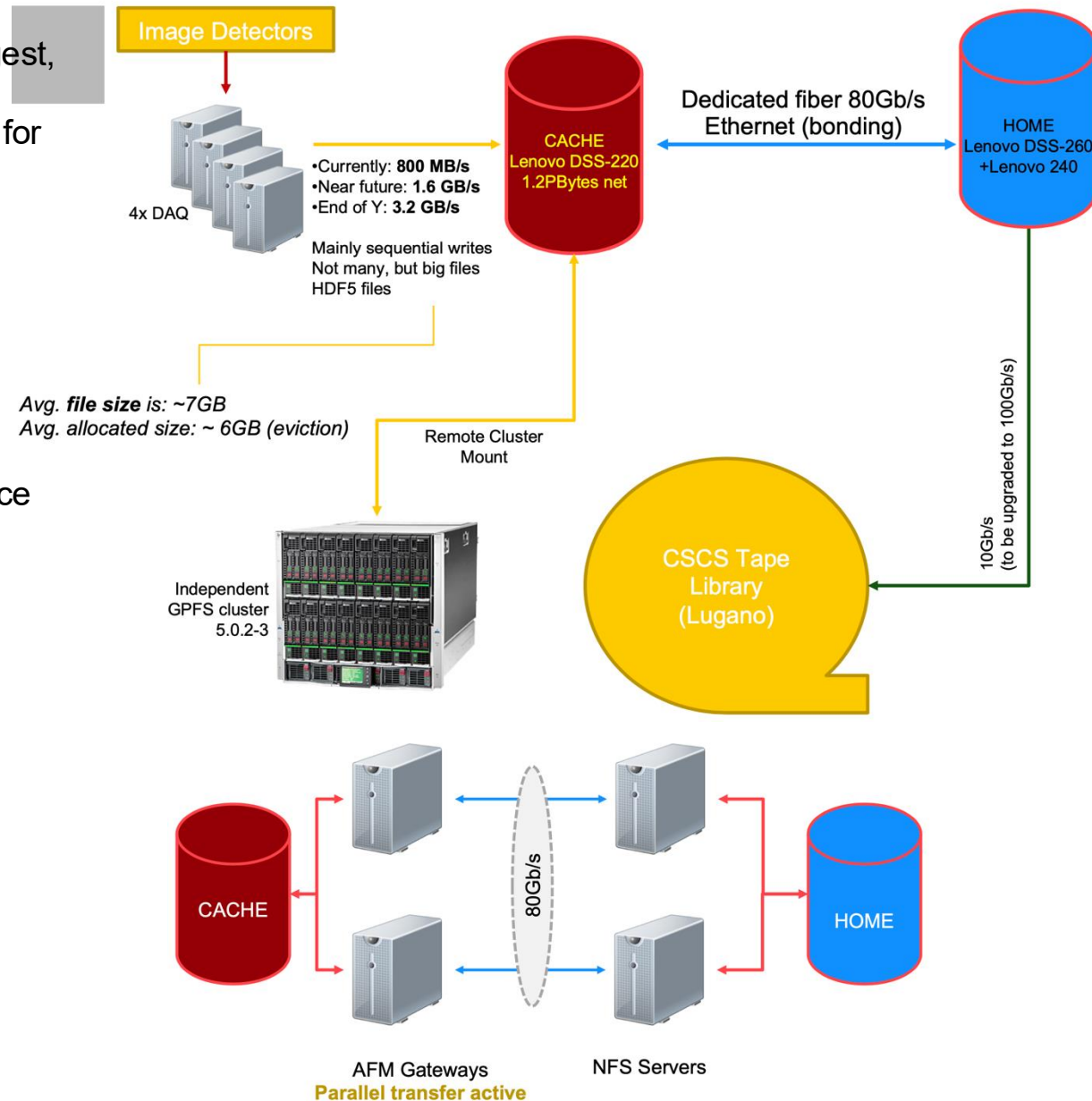
Research Institute in Switzerland

- Online Storage for large images processed by compute clusters for ingest, initial analysis & research
- Images once processed need to be stored offsite as long term storage for further analysis
- Needs data to be transferred promptly or real time from Online to Long term storage
- Produces ~1 PBytes of data (images) per year

Solution

- Online storage configured as AFM Cache; Long term storage as AFM Home
- Cache has eviction-enabled, to virtually extend its real fast-access space
- Eviction is automatic, based on filesets-level quota
- Spectrum Scale + GNR, GPFS 5.0.2.3, 2x AFM gateways
- NFS export

More: [Spectrum Scale user group presentation](#)



Research + HPC environment: University of Queensland, Australia

Overview

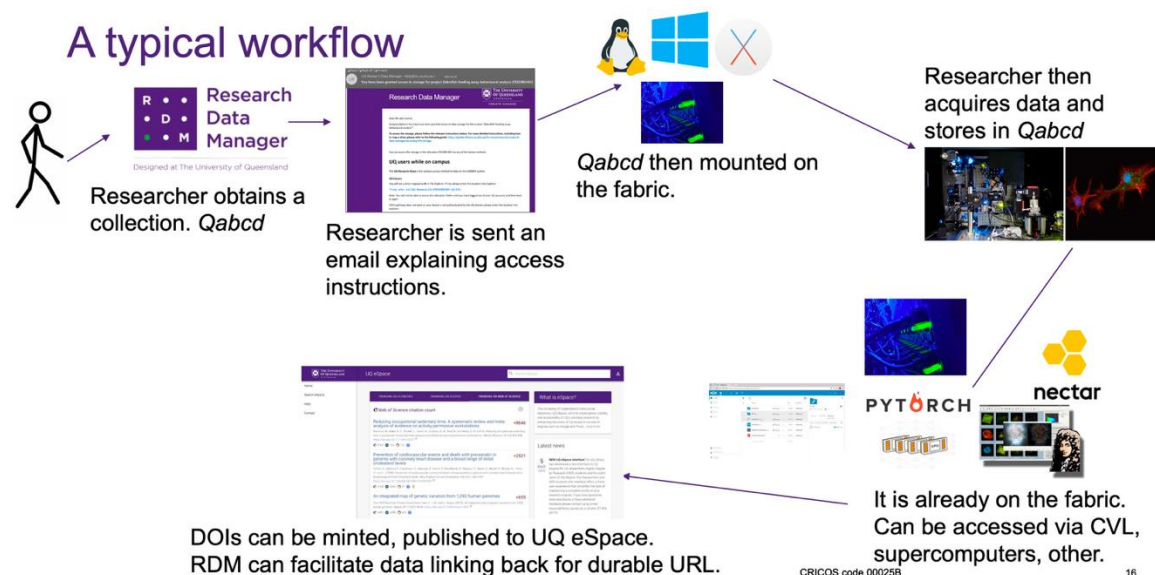
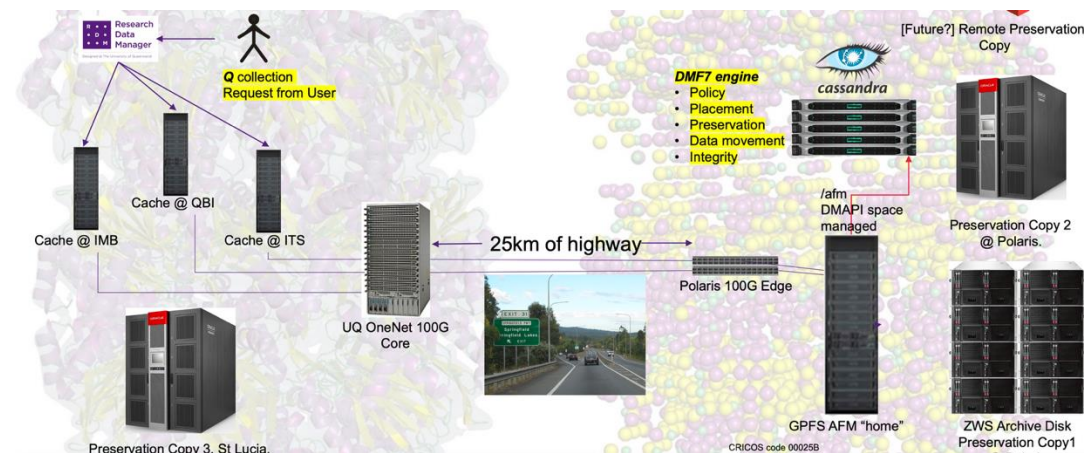
- Vocal reference customer for Spectrum Scale + AFM. Have delivered multiple user group presentations. Reference [Story](#).
- Signed joint agreement with IBM for Spectrum Scale/ESS in 2020

Use case – Single Data Fabric for research data and DR

- Medical, scientific instruments collect and ingest data via SMB, NFS at AFM cache sites
- Data is shared with home at Supercomputing Center
- Data needs to be accessed by Supercomputers, research stations, other departments spread across multiple campus locations for analysis and sharing results
- Certain data needs to be stored at DR location

Configuration Details

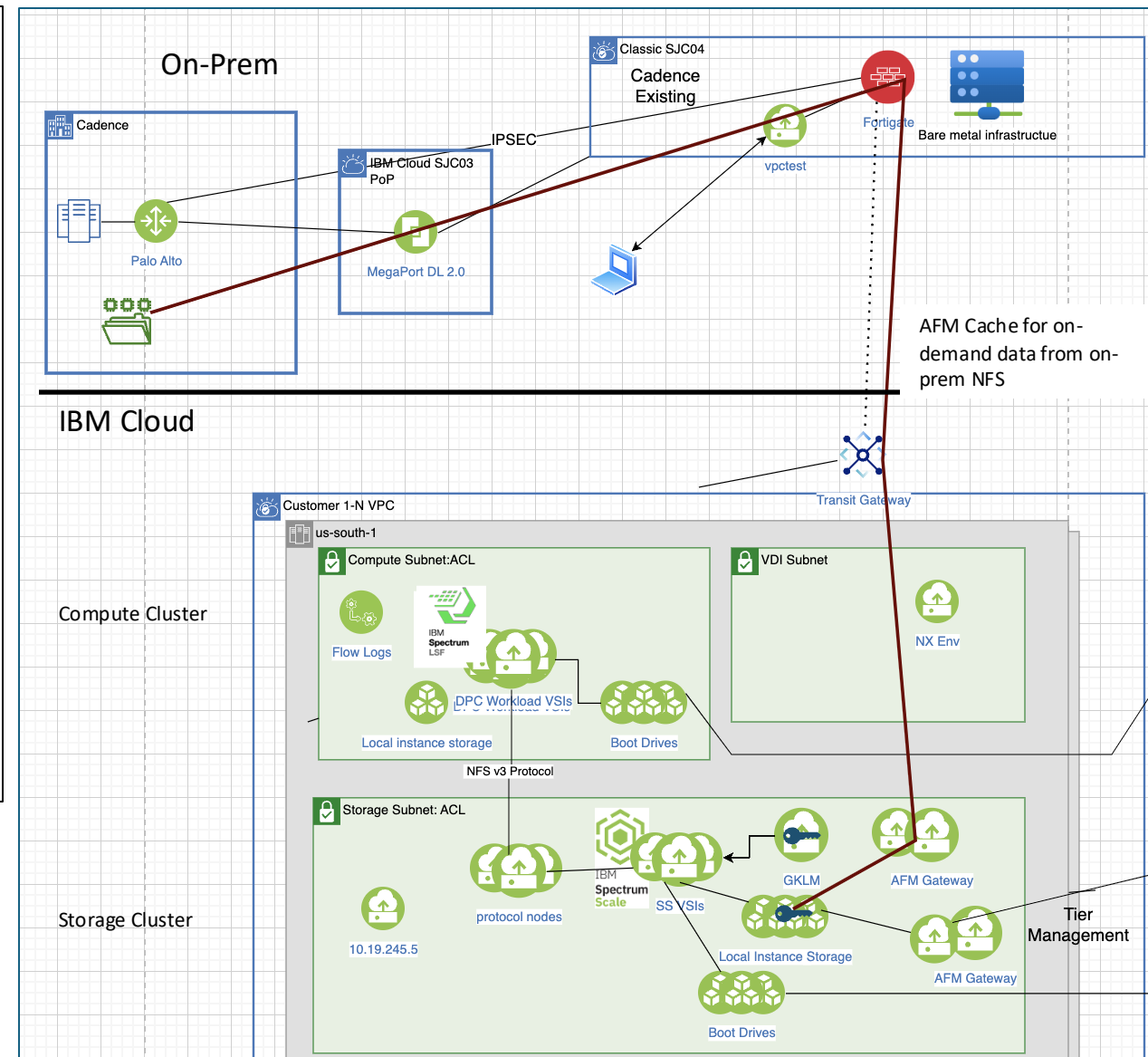
- Unique configuration that involves Data sharing + DR configuration (cascade)
- One of the largest sequential write performance over SMB for data ingest
- Customer uses 1500-2000 independent filesets (over published limits)
- Daily ingest of 100-290 TB data
- 2x ESS GH14S at Home connected to Cache over 100Gbps – 25Kms apart
- Use Scale Tiering to an external data pool at Home



Cache Solution with NFS backend: Leading EDA provider

- Caching between multiple compute locations
 - High performance – throughput upto 10GB/s
 - Metadata and data read/write on multiple file in parallel
 - Compute and IO intensive, non-containerized
- How AFM is used in the environment
 - Data source (Home) is on-prem data centers on Netapp NFS
 - EDA jobs execution on another site (IBM cloud)
 - AFM Cache for on-demand data from Netapp
 - Read only Cache setup on IBM Cloud for jobs to access data
 - Job results are stored in AFM cache in independent writer mode to push results back to on-prem Netapp Filer

Reference: [Storage scale user group presentation](#)



AFM Cache solution: Medical research example

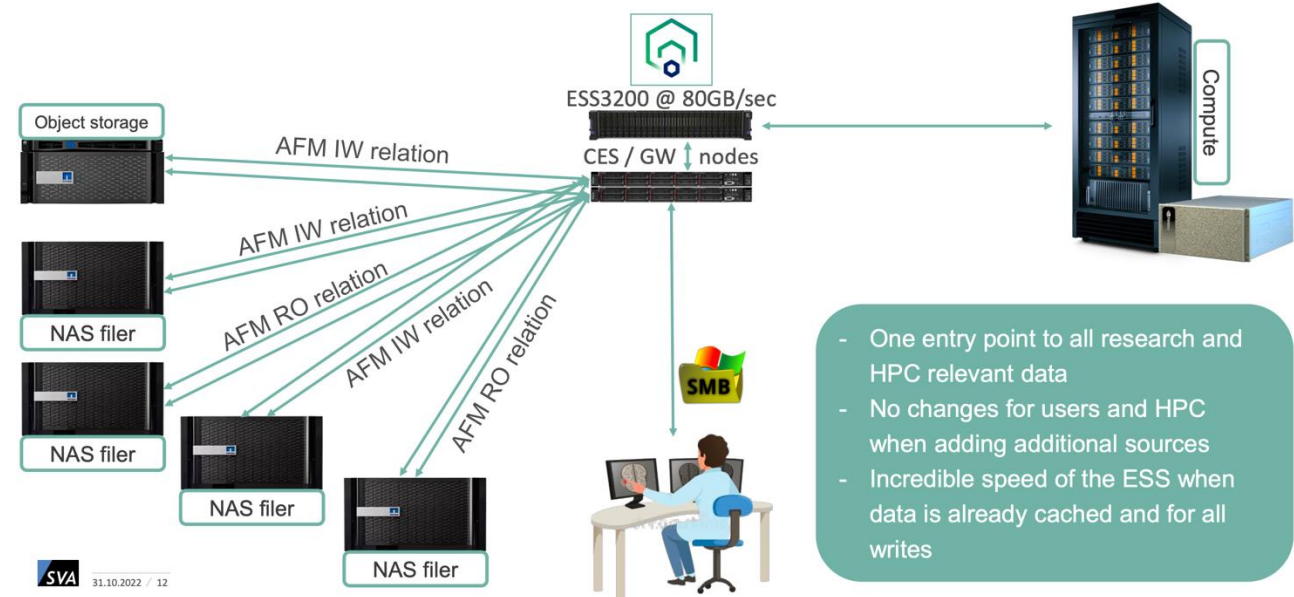
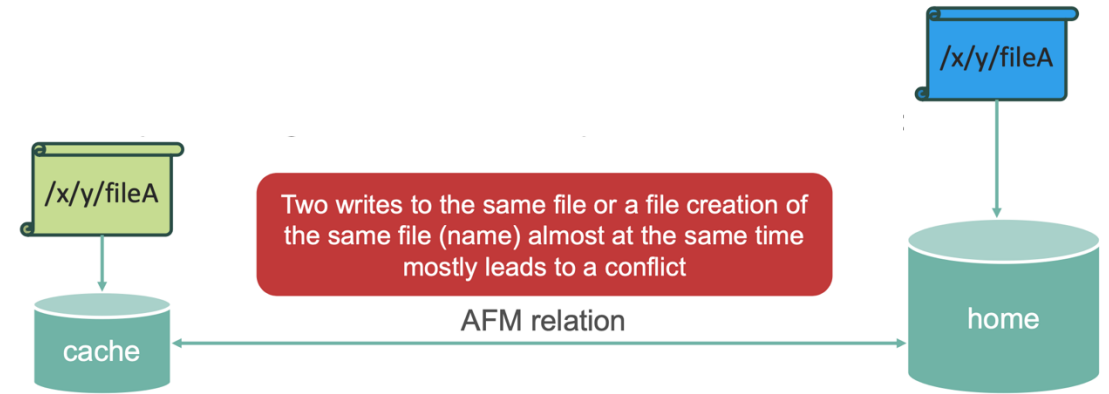
Before AFM

- Storages are distributed across org
- Medical researcher goes to all storages to get access to data – admin has to manage all access
- Legacy storages and hence slower access to data
- Data not available Centrally

After AFM

- Data available on fast storage at single, central location
- Investment in legacy storage is still protected making it much faster at additional costs
- ESS Storage capacity 10% of overall storage in the enterprise

Reference: [Scale user group reference](#)



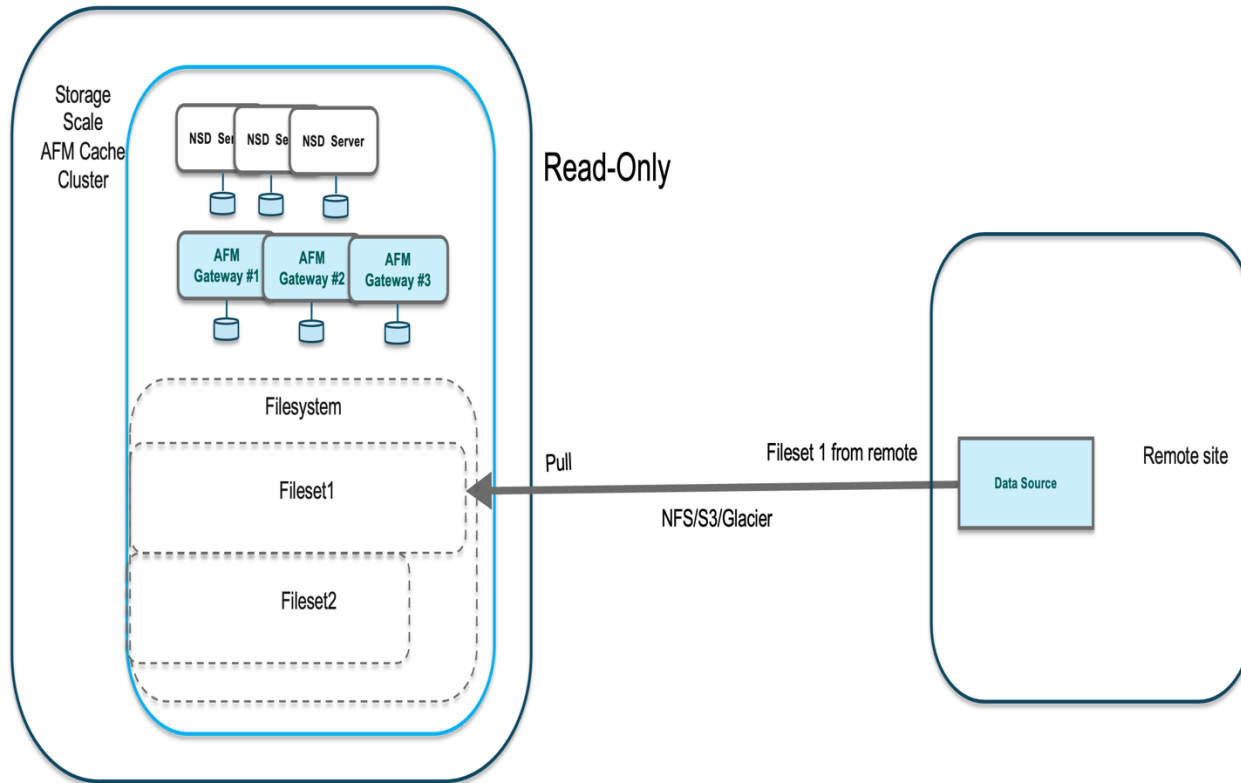
Important AFM S3 Object configuration parameters

- **afmMUAutoRemove**
 - Enabling this for a MU fileset will queue file remove operations to the COS automatically
- **afmObjectDirectoryObj**
 - Enabled at fileset level to support synchronisation of all directories with/without object with COS
- **afmObjectFastReaddir**
 - Enabled at the fileset level to skip the fetching of extended attributes and acl from COS for readdir operations leading to faster listing operations
- **afmObjKeyExpiration**
 - Specifies the COS key expiration timeout value which allows to reload the access/secret keys after the defined timeout value. Default is 36000 seconds. Set at the cluster level
- **afmParallelReadThreshold**
 - Defines the threshold beyond which parallel reads become effective and is enabled by default . Default is 1024MB.
- **afmParallelReadChunkSize**
 - Defines the minimum chunk size of the read that needs to be distributed among the gateway nodes during parallel read. Default is 128MB . A value of zero disables parallel reads
- **afmPrefetchThreshold**
 - Controls partial file caching. Default is 0 which caches the entire file by pulling all the blocks when 3 blocks are read at the cache. Values are in the range 1-100 and specified the percentage of file that must be read to cache the whole file. A value of 100 disables full file prefetching.
- **afmAsyncDelay**

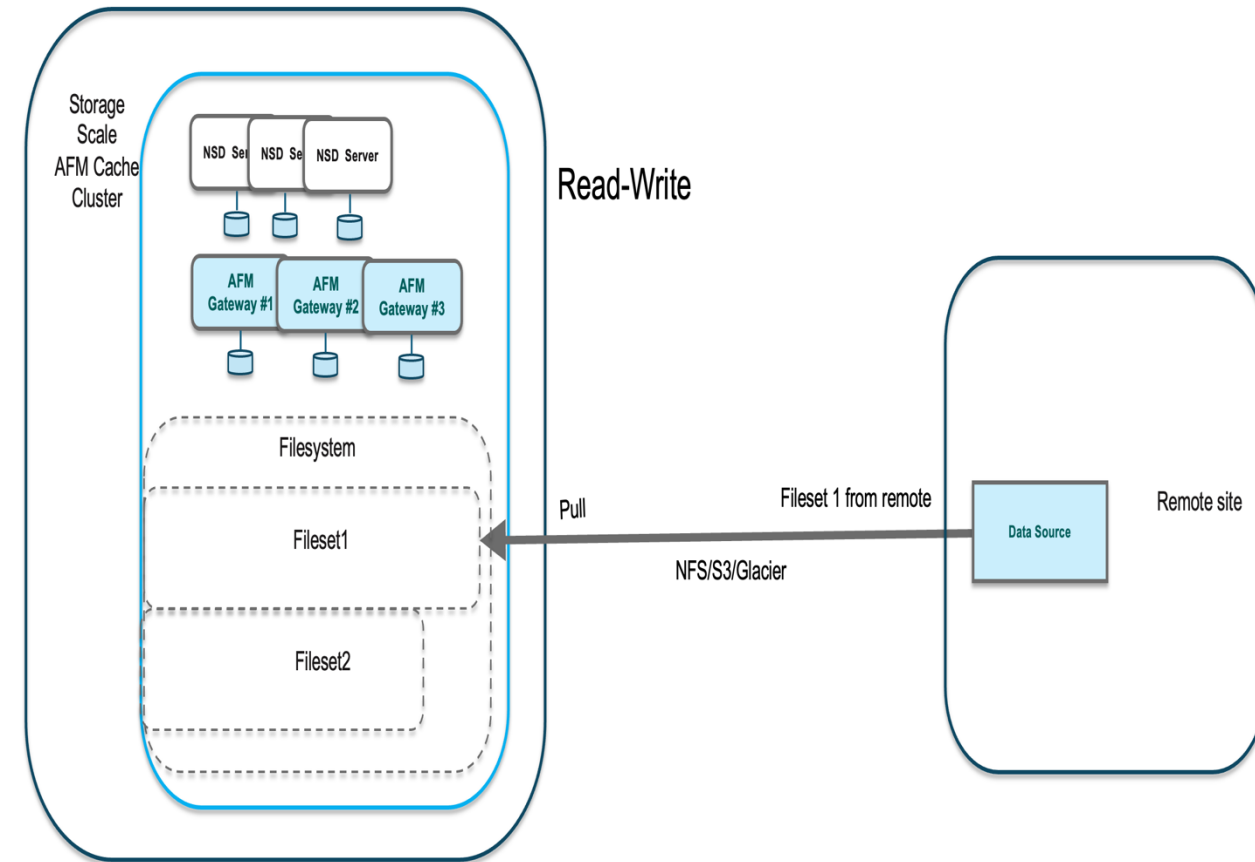
Time by which the asynchronous operations are delayed to home

AFM Modes

Read-Only Mode

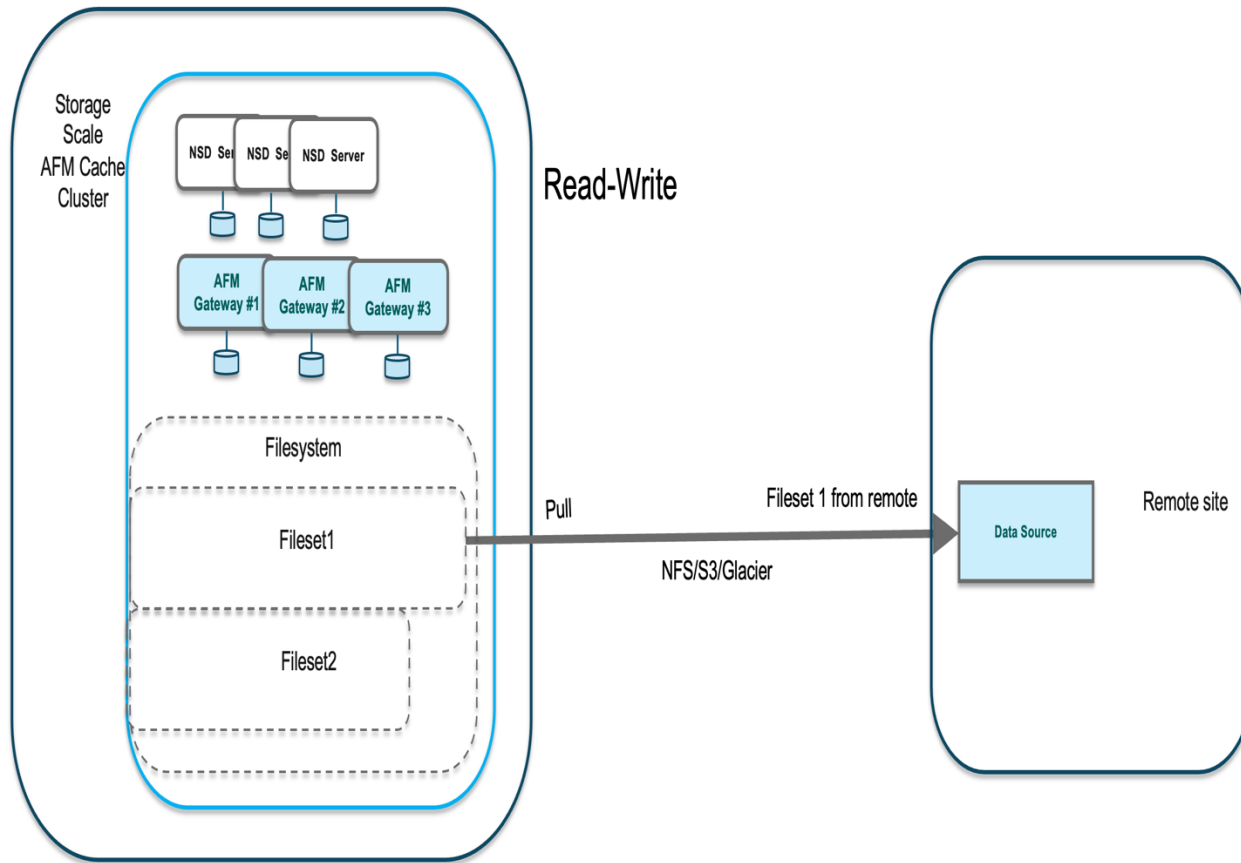


Local-Updates Mode



AFM Modes

Single-Writer Mode



Independent-Writer Mode

