Hewlett Packard Enterprise

EXPANDING DATA MANAGEMENT SERVICES BEYOND TRADITIONAL PARALLEL FILE SYSTEMS WITH HPE DATA MANAGEMENT FRAMEWORK

Kirill Malkin
Director, HPC Storage Engineering
Cray User Group Meeting, May 2022

CONFIDENTIAL DISCLOSURE AGREEMENT

- The information contained in this presentation is proprietary to Hewlett Packard Enterprise and is offered in confidence, subject to the terms and conditions of a binding Confidential Disclosure Agreement (CDA)
- HPE requires customers and partners to have signed a CDA in order to view this training
- The information contained in this training is HPE confidential
- This presentation is NOT to be used as a 'leave behind' for customers and information may only be shared verbally with HPE external customers under NDA
- This presentation may be shared with Partners under NDA in hard-copy or electronic format for internal training purposes only
- Do not remove any classification labels, warnings or disclaimers on any slide or modify this presentation to change the classification level
- Do not remove this slide from the presentation
- HPE does not warrant or represent that it will introduce any product to which the information relates
- The information contained herein is subject to change without notice
- HPE makes no warranties regarding the accuracy of this information
- The only warranties for HPE products and services are set forth in the express warranty statements accompanying such products and services
- Nothing herein should be construed as constituting an additional warranty
- HPE shall not be liable for technical or editorial errors or omissions contained herein
- Strict adherence to the HPE Standards of Business Conduct regarding this classification level is critical

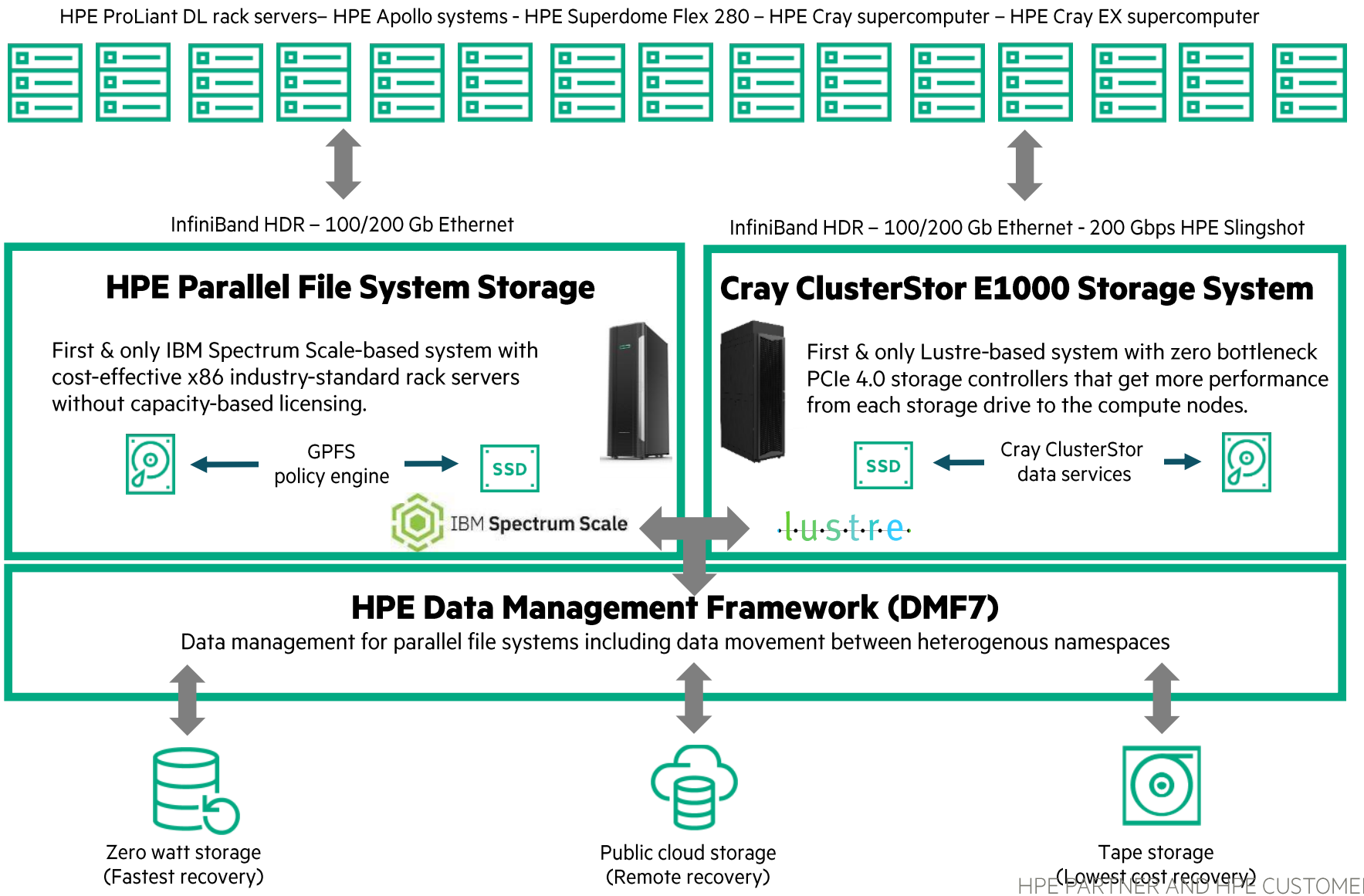
THIS PRESENTATION

- Expanding data management services beyond traditional parallel file systems with HPE Data Management Framework

- Lustre and Spectrum Scale are the most popular parallel file systems used for scratch storage in supercomputing today. Traditionally, these islands of storage have required supplemental solutions to prevent over-utilization and migrate data Traditional parallel file systems have long been the high-performance storage option of choice for HPC systems and applications. Centralized data management systems like DMF use connectors and tools provided by Lustre and Spectrum Scale to perform hierarchical storage management, protection, scalable search and other valuable services. But what about other types of storage like non-parallel file systems, Network Attached Storage and object storage that are proprietary and/or don't have the necessary connectors and tools for the central data management system? These storage systems create silos of data management and every HPC center is looking for better ways to manage them.

- Cray users who have been using RobinHood, HPSS, and other tools to supplement their chosen PFS – and sometimes both of In this session, Kirill Malkin, Engineering Director for HPE, will review new capability in DMF that extends centralized data management to non-parallel file systems, NAS filers and object stores. He will illustrate how DMF uses file stubs to reduce storage utilization on filers while still leaving a path to recover migrated files back into the original location. Also, he will review an innovative enhancement that enables DMF to add object storage to its list of front-end namespaces, and he'll walk through the workflow that uses S3 and the DMF API to protect objects in DMF's back end storage.

ONLY HPE DMF KNOWS WHERE ALL DATA IS



DATA MANAGEMENT WITH DMF

DMF Knows Where Data Is

Protect

- Continuous Deep Protection of Primary Data
  - Forever incremental file-based backup
  - Rapid namespace recovery

Move

- Managed Horizontal Data Movement
  - Safely migrate data among managed namespaces
  - Background media management

Scale

- Unlimited Oversubscription of Primary Storage
  - Transparently & seamlessly expand namespace to low-cost storage

DMF BACKEND STORAGE

Unlimited Scalability

Tape

- Lowest cost storage
- Largest libraries supported

Zero Watt

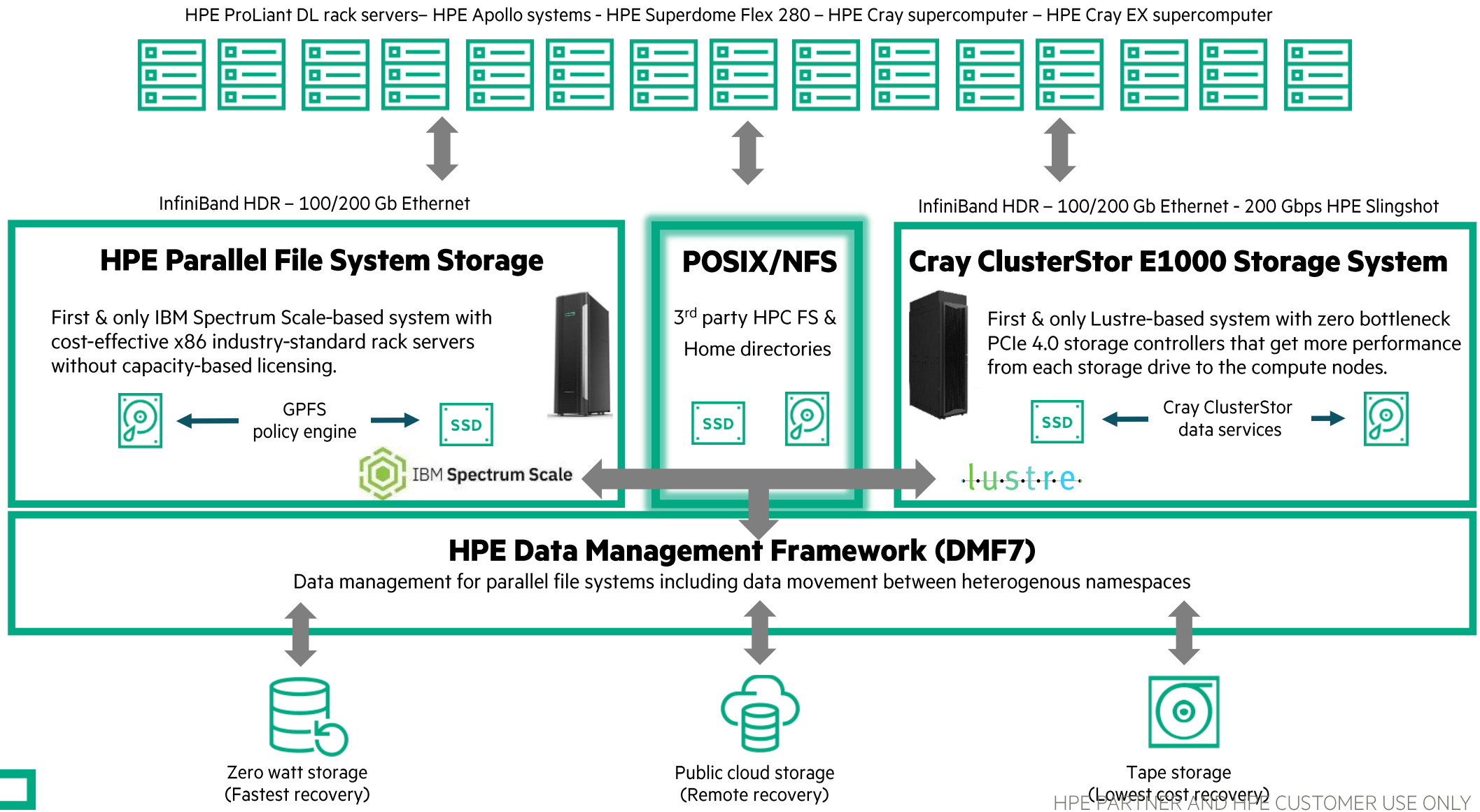
- Nearline JBODs
- Power optimized

Cloud

- Multi-cloud capable
- On-prem option

SCALITY MINIO ceph S3 ECS

HPE DMF KNOWS WHERE ALL DATA IS – *NEW IN 7.5!*



SUPPORT FOR GENERAL POSIX/NFS

- We want to:
  - Overprovision these storage resources and drive down effective cost & TCO
  - Facilitate data mobility between general NFS and HPC file systems
  - Protect data and integrity of the file system
  - Standardize data management tools and minimize silos
- But there are challenges
  - 3rd party file systems don't provide standard mechanism for integration with data management systems

- DMF enhancements for POSIX file system support
  - Use just like any application or user, i.e. via POSIX
  - Mimic HSM and archiving behavior in the file system
  - Do this without relying on integration tools or proprietary APIs
- Benefits
  - DMF reduces storage utilization
  - DMF captures changes to files and the file system so admins can automate their policies from one pane of glass
  - DMF versions files, snaps the namespace, & recovers both

MANAGED FILE SYSTEM REFLECTION WITH DMF

Managed file systems with native integration with DMF

- These file systems use API for communicating file system events and changes to external data management system – like DMAPI, Changelog, etc.

- How it works with DMF
  - Initial scan of the file system to build internal namespace reflection in DMF
  - Then update the namespace reflection using native file system tools:
    - Changelog (Lustre)
    - Streaming events (Spectrum Scale, XFS)
    - This is a valuable benefit of native integration

Managed file systems with no native integration

- General POSIX file systems that do not have API for communicating with external data management system can still be managed

- How it works with DMF
  - Initial scan of the file system to build internal namespace reflection in DMF
  - Then update the namespace reflection via:
    - Subsequent full scans
    - Optimized scanning practices (e.g. directory level)
      - Scan directories only first
      - Fully rescan only the ones that changed

CAPACITY MANAGEMENT WITH DMF

Managed file systems with native integration with DMF

- These file systems provide HSM API that is used by external data management system – like DMF - to manage storage tiering & utilization

- How it works with DMF
  - Run policy against the namespace reflection
    - Migrate file(s) into DMF
    - File state is DUL
  - Identify files that meet policy and can be released
  - Leave file stub using native file system functionality:
    - HSM coordinator (Lustre)
    - DMAPI (Spectrum Scale, XFS)
    - File state is OFL
  - Recall OFL files on access, via DMF policy or API
    - Copy file data from backend to original file
    - File state is DUL again

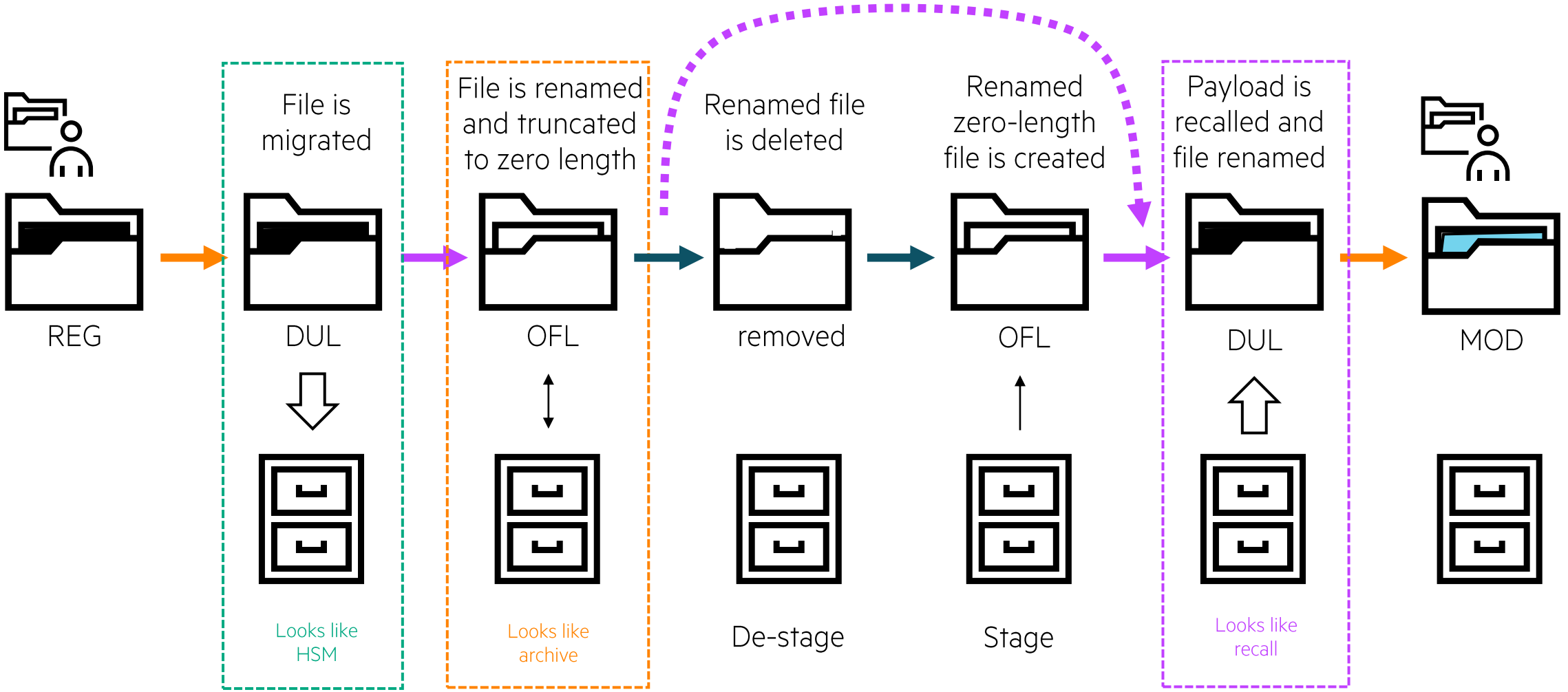
Managed file systems with no native integration

- These are general POSIX file systems without HSM API – storage utilization management is simulated

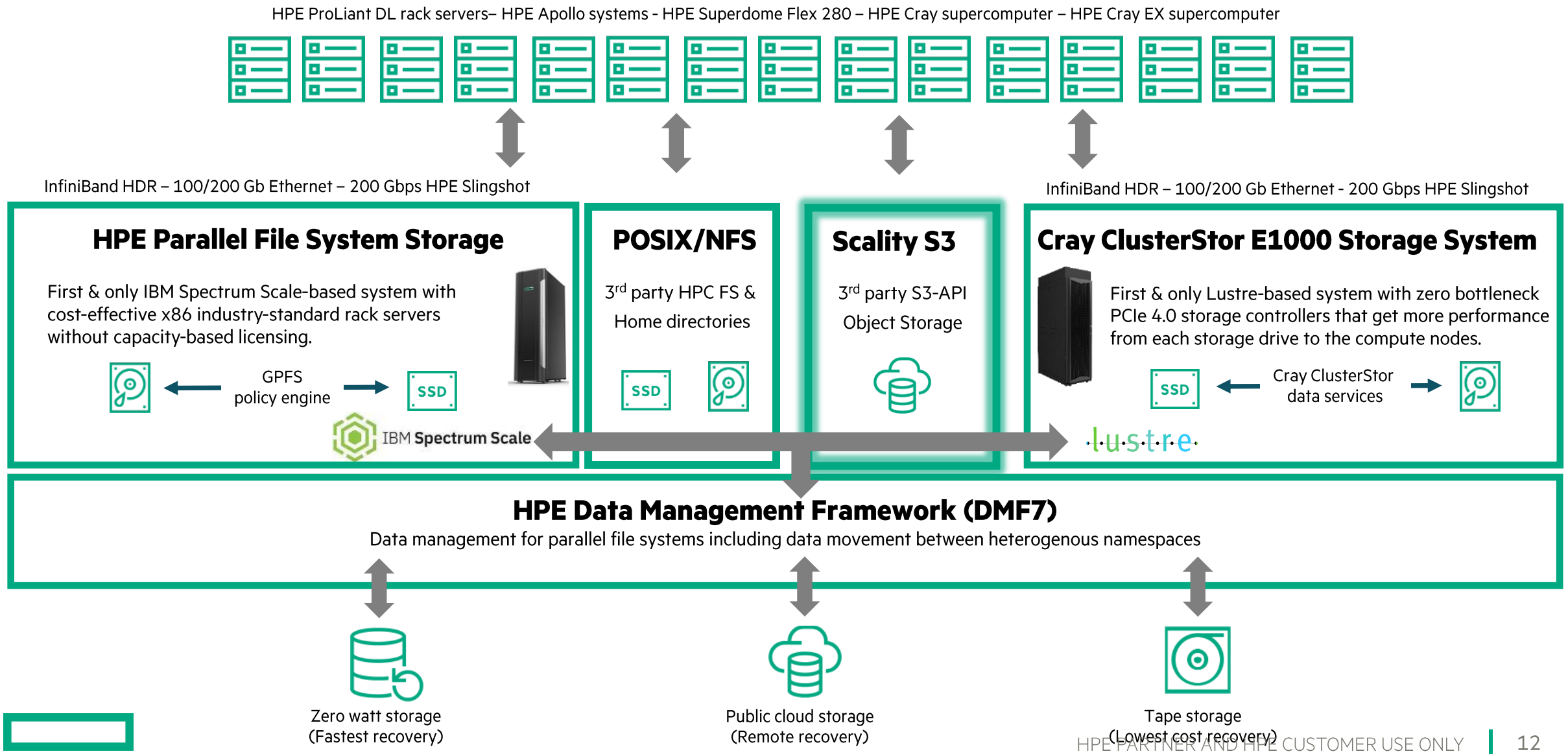
- How it works with DMF:
  - Run policy against the namespace reflection
    - Migrate file(s) into DMF
    - File state is DUL
  - Identify files that meet policy and can be released
    - Rename the file(s) in file system
      - Maintains position of the inode
      - Original file appears to be archived
      - Renamed file is effectively hidden, avoiding accidental access
    - Truncate renamed file to zero length
      - File state is OFL
      - Reduces storage utilization
      - *This effectively mimics HSM = inode with a stub*
  - Recall OFL files via DMF policy or API
    - Copy file data from backend to renamed, truncated file
    - Rename back to original
    - File state is DUL again

MANAGED FILE LIFECYCLE IN DMF

POSIX Namespace Support

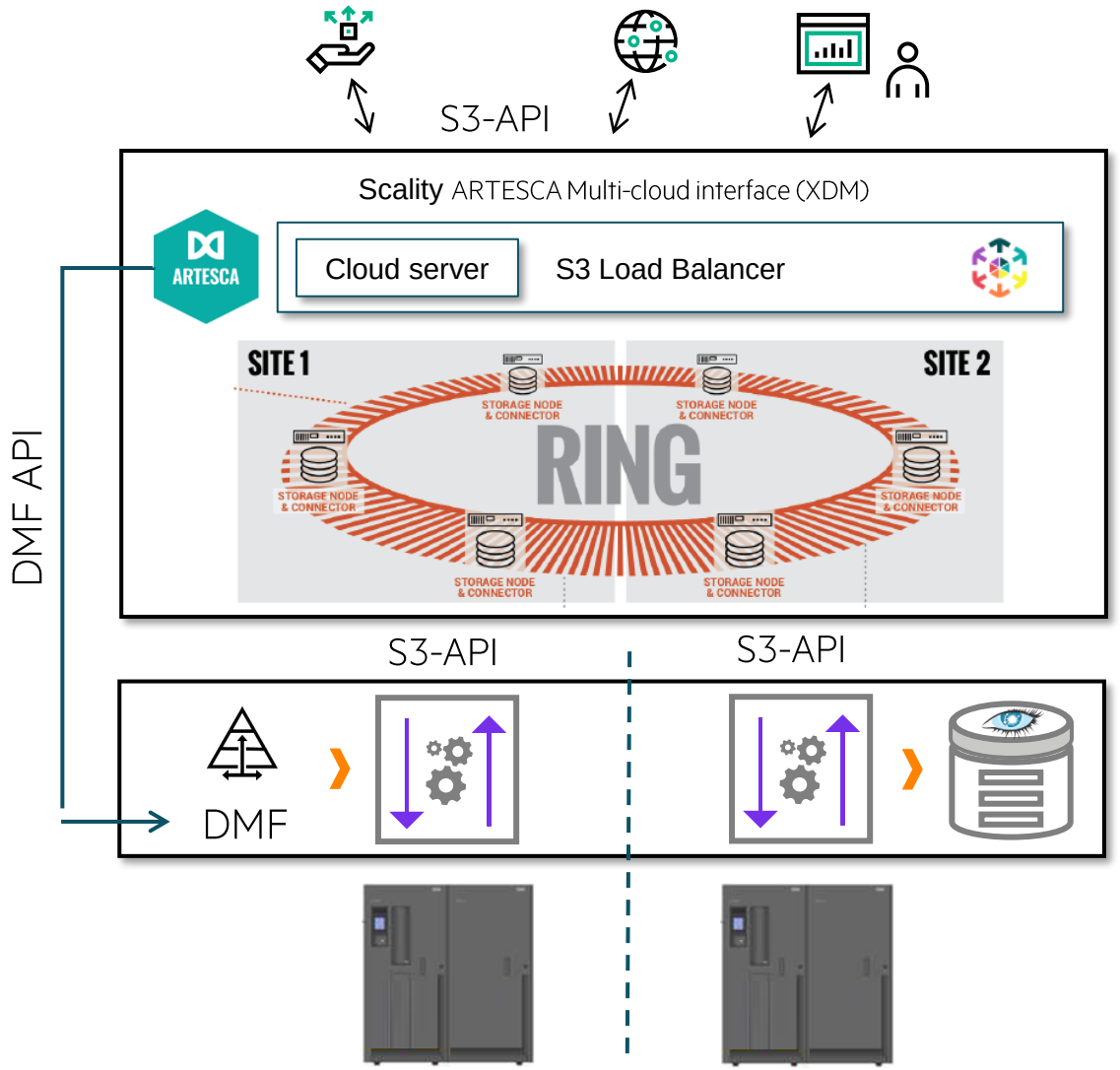


HPE DMF KNOWS WHERE ALL DATA IS – *COMING SOON!*



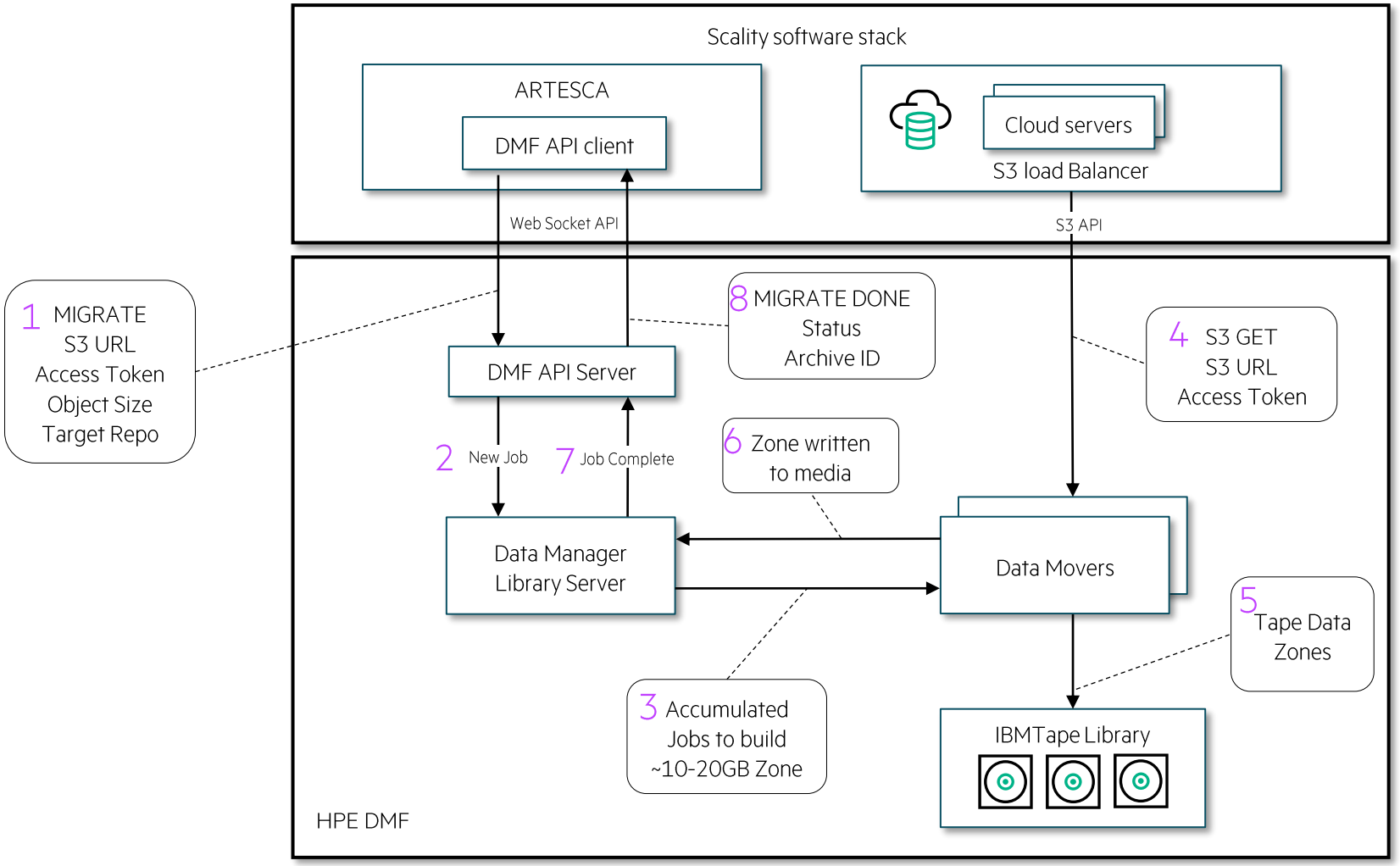
SCALITY S3 NAMESPACE SUPPORT WITH DMF

- All application access goes directly to Scality Ring Object Storage via S3-API
- Scality product (ARTESCA) manages offline objects
  - When object is qualified for archiving, ARTESCA calls DMF API to migrate object via S3-API
  - Once migrated, the object is stubbed
  - When application attempts to access offline object, ARTESCA calls DMF API to recall object via S3-API
  - Once recalled, stub is replaced with real object
- No S3 namespace reflection is created in DMF
  - Migration policy is driven externally by ARTESCA
- DMF fully manages tape libraries and object lifecycle
  - Maintains required number of copies
- Two site implementation with asymmetric protection requirements



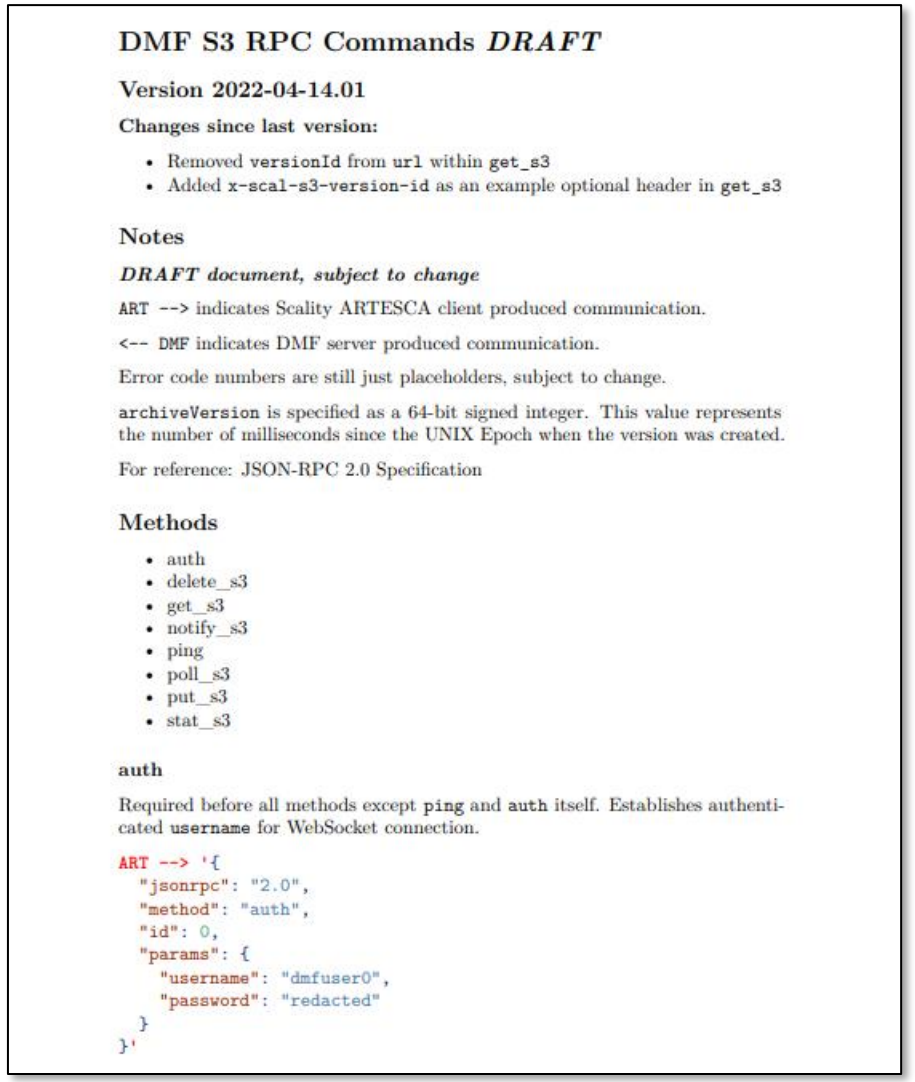
DMF & SCALITY INTEGRATION

Migration Request Path



DMF API S3 RPC COMMANDS

- JSON-RPC 2.0 compatible
- Methods
  - auth
    – Establishes connection
  - delete_s3
    – Delete archived object
  - get_s3
    – Recall archived object
  - notify_s3
    – Migrate/Recall/Delete completion notification
  - ping
    – API server check
  - poll_s3
    – Check job status
  - put_s3
    – Migrate object
  - stat_s3
    – Check for object presence in archvie



DMF S3 RPC Commands *DRAFT*

Version 2022-04-14.01

Changes since last version:

- Removed `versionId` from `url` within `get_s3`
- Added `x-scal-s3-version-id` as an example optional header in `get_s3`

Notes

DRAFT document, subject to change

`ART -->` indicates Scality ARTESCA client produced communication.

`<-- DMF` indicates DMF server produced communication.

Error code numbers are still just placeholders, subject to change.

`archiveVersion` is specified as a 64-bit signed integer. This value represents the number of milliseconds since the UNIX Epoch when the version was created.

For reference: JSON-RPC 2.0 Specification

Methods

- auth
- delete_s3
- get_s3
- notify_s3
- ping
- poll_s3
- put_s3
- stat_s3

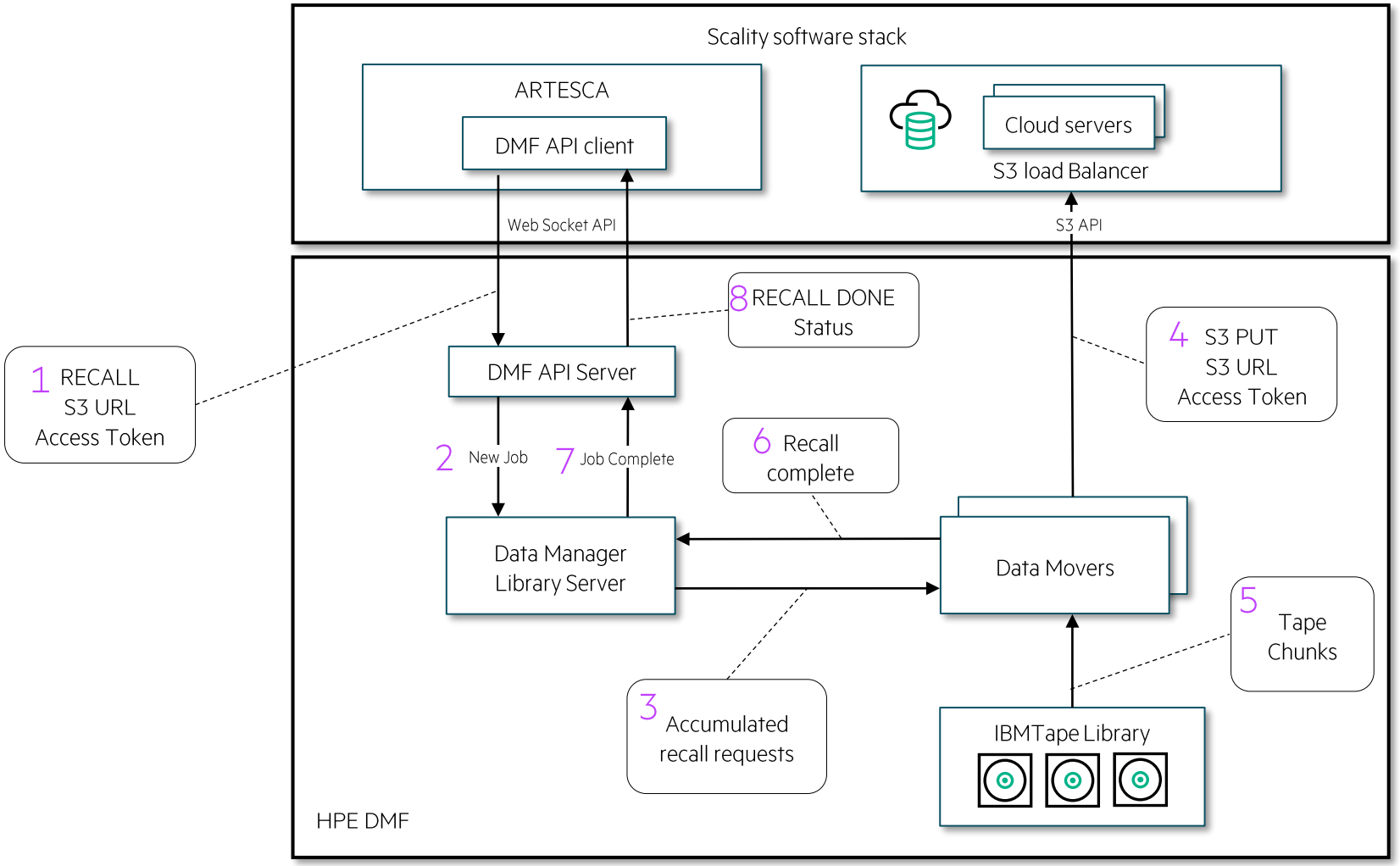
auth

Required before all methods except `ping` and `auth` itself. Establishes authenticated `username` for WebSocket connection.

```
ART --> '{
  "jsonrpc": "2.0",
  "method": "auth",
  "id": 0,
  "params": {
    "username": "dmfuser0",
    "password": "redacted"
  }
}'
```

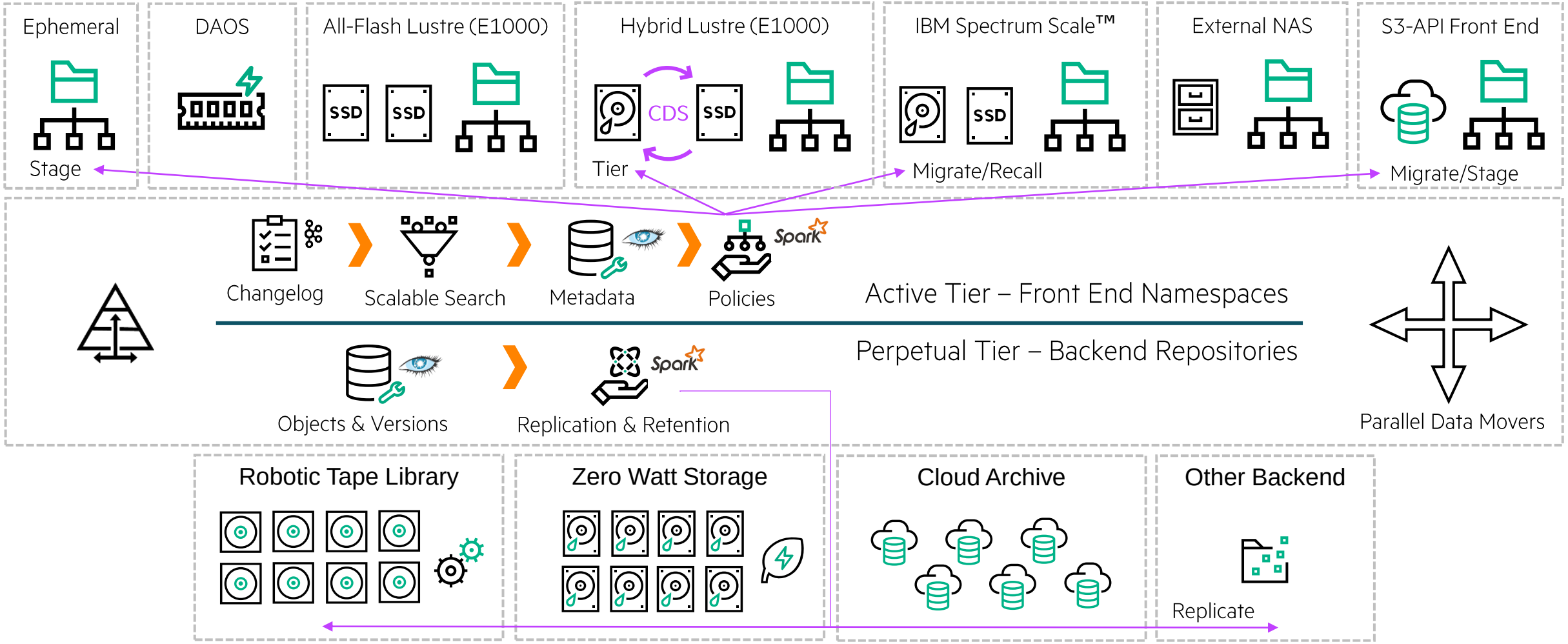
DMF & SCALITY INTEGRATION

Recall Request Path



BROADER VISION FOR DATA INFRASTRUCTURE

Pluggable Front/Back End, Metadata Reflection, Vertical & Horizontal Data Movement




Hewlett Packard
Enterprise
THANK YOU
kirill.malkin@hpe.com