



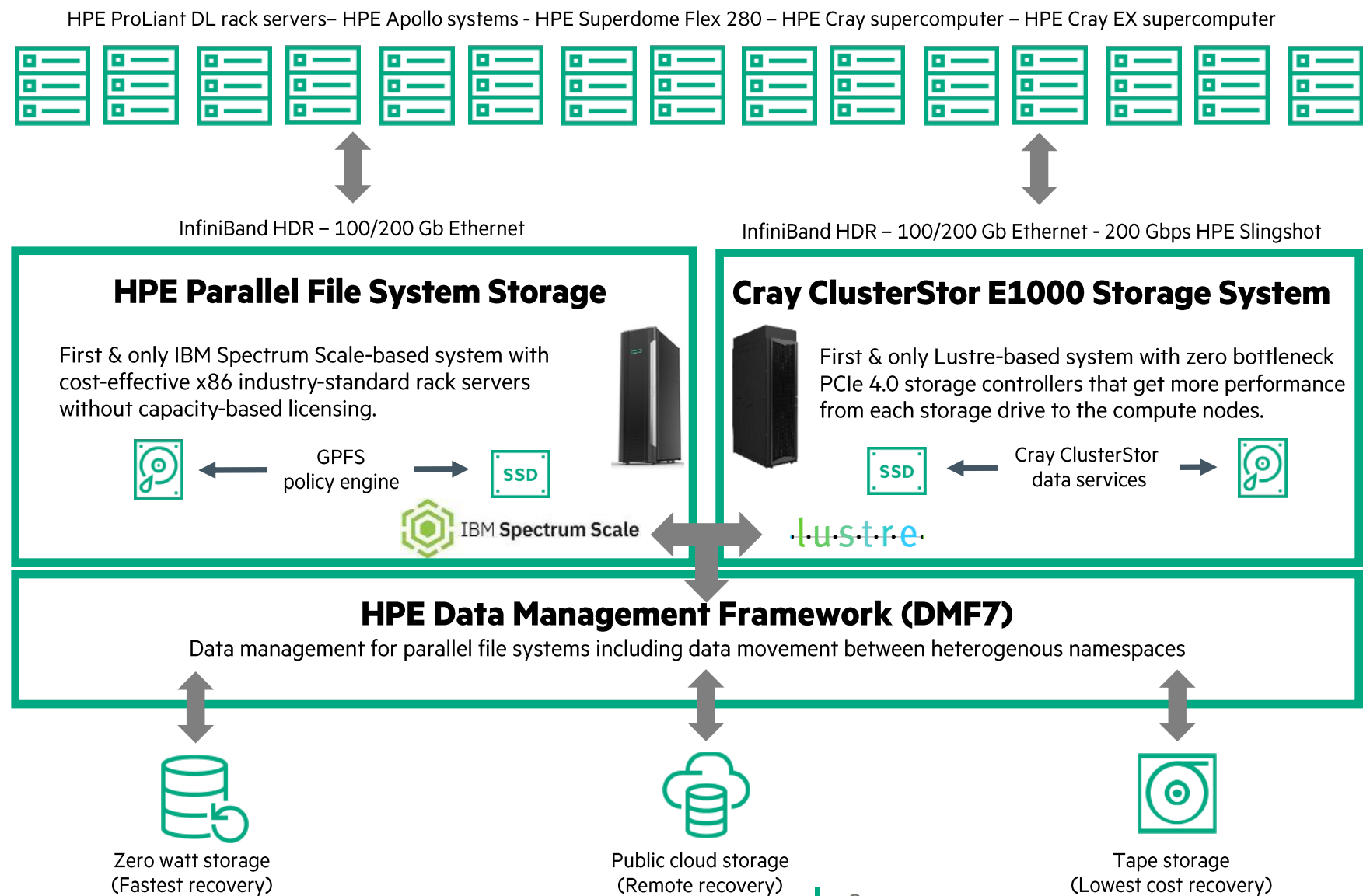
**Hewlett Packard**  
Enterprise

# **LUSTRE & SPECTRUM SCALE: SIMPLIFY PARALLEL FILE SYSTEM WORKFLOWS WITH HPE DMF**

Kirill Malkin – Director of Development  
Zsolt Ferenczy – Solutions Architecture

May 3, 2021

# ONLY HPE DMF KNOWS WHERE ALL DATA IS



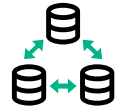
# DATA MANAGEMENT WITH DMF

DMF Knows Where Data Is



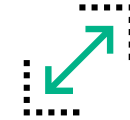
## Protect

- 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



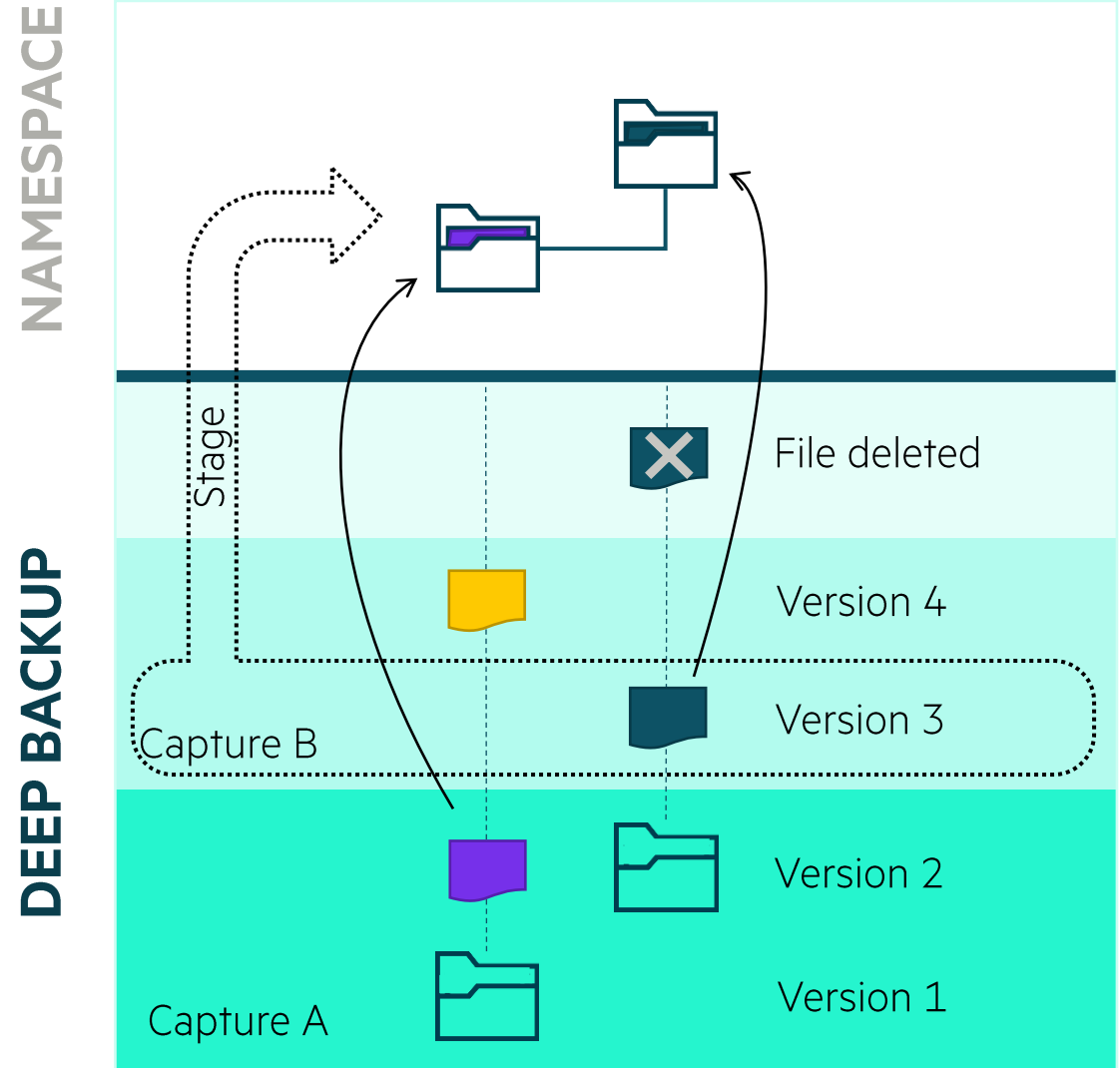
## Scale

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



# CONTINUOUS DEEP PROTECTION

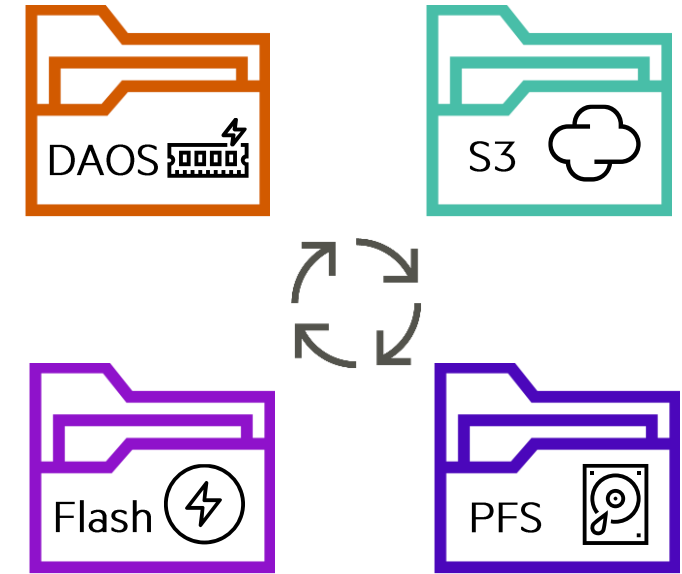
- Forever-incremental, file-granular file system backup
  - Files are automatically copied to append-only backend media as they stabilize
  - Tape Libraries, Nearline disk, and Multi-Cloud are supported as backend
  - As files are modified, change is new backend versions are created
  - Namespace state is captured at scheduled intervals
  - Older namespace captures are removed over time
  - Backend copies no longer referenced by captures are released and media is scheduled for compaction
- Rapid namespace recovery
  - Point-in-time restore of any file or entire file system
  - Recovery point capture, including directory structure and file stubs, is staged to namespace
  - Namespace becomes available to users and applications before all content is restored
  - File content is recalled just-in-time on access
  - Remaining content is rehydrated in background



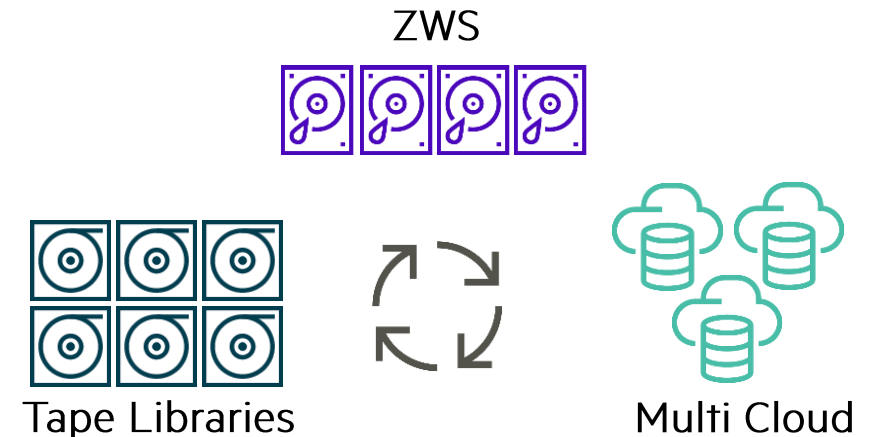
# HORIZONTAL DATA MOVEMENT

- Safely migrate data among managed namespaces
  - Movement granularity is a dataset that may contain a single file or an entire namespace
  - Data is moved in parallel across multiple nodes and between diverse filesystem types
  - Number of copy threads are scaled dynamically to target throughput, including bandwidth and files per second
  - Changes to source file are detected and transfer is restarted
  - Access to destination file is intercepted and I/O is delayed until completion
  - Incomplete transfers are automatically resumed
  - Movement is bidirectional with conflict resolution via backend versioning
- Manage media lifecycle in background
  - Policy-driven, unattended backend technology migration
  - Automatic backend media compaction
  - Asymmetric media migration, such as tape to cloud
  - Single versioned namespace for backend data ocean

## PRIMARY STORAGE

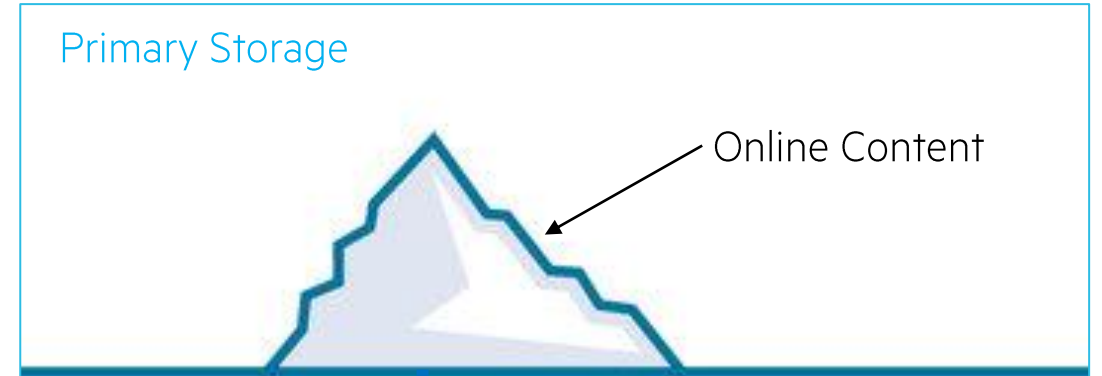


## BACKEND MEDIA



# UNLIMITED OVERSUBSCRIPTION

- Large-scale data lakes, such as parallel file systems, have fixed capacity limits
- Rapidly growing unstructured data causes depletion of capacity leading to performance degradation, application errors, and data loss
- DMF automatically identifies large, infrequently accessed data files and transparently copies them to low-cost media, such as nearline disk, tape or cloud
- Once data is copied and secured, DMF automatically releases space consumed by these files, while retaining file metadata inside namespace
- When application opens released file, DMF seamlessly restores file content to primary storage and delivers it to application

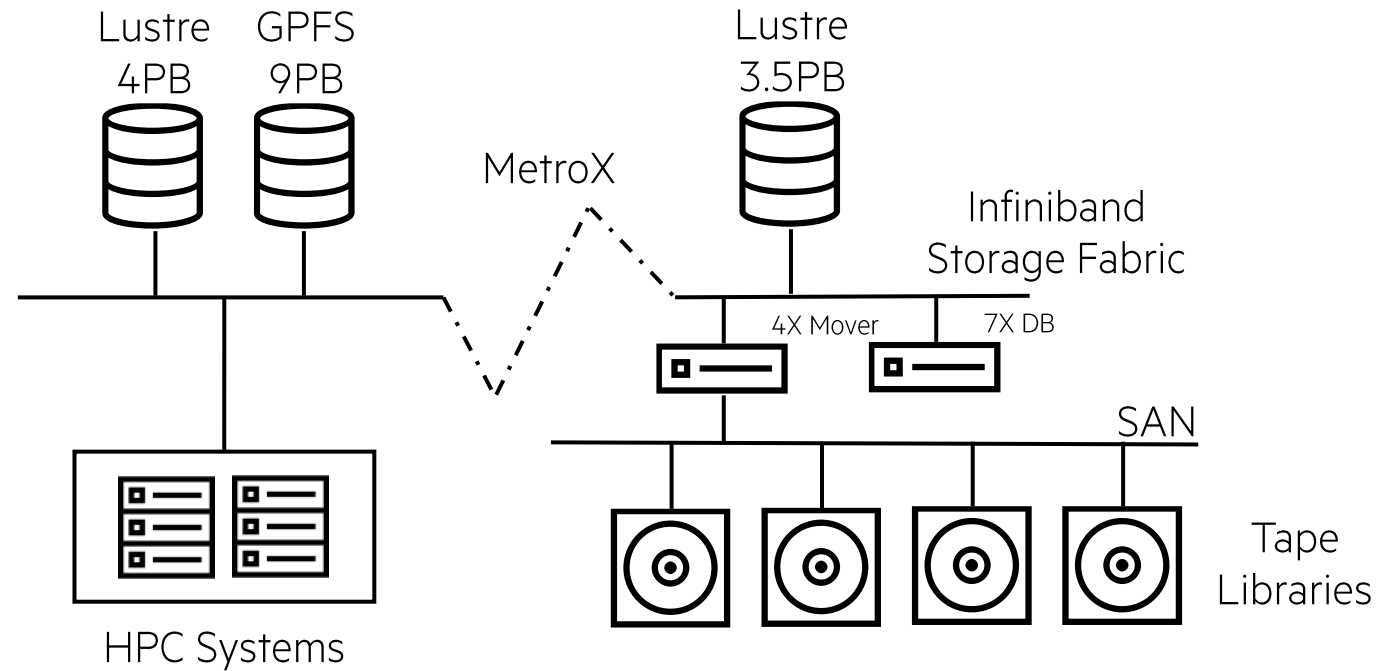




# DMF 7 CUSTOMER EXAMPLE

## Data Protection for Lustre & Spectrum Scale

- Multiple, highly active managed filesystems
- Home & Project use cases
- Lustre & Spectrum Scale
- DMF consolidated view of all data sets
- Moving datasets across managed filesystems
- Copying files & directories to backend for disaster recovery
- Parallel streaming to large number of high-speed tape drives





**Hewlett Packard**  
Enterprise

**THANK YOU**

And now it's time for Q&A

