

Enabling Object Storage Systems for High-Latency Media

Harald Seipp

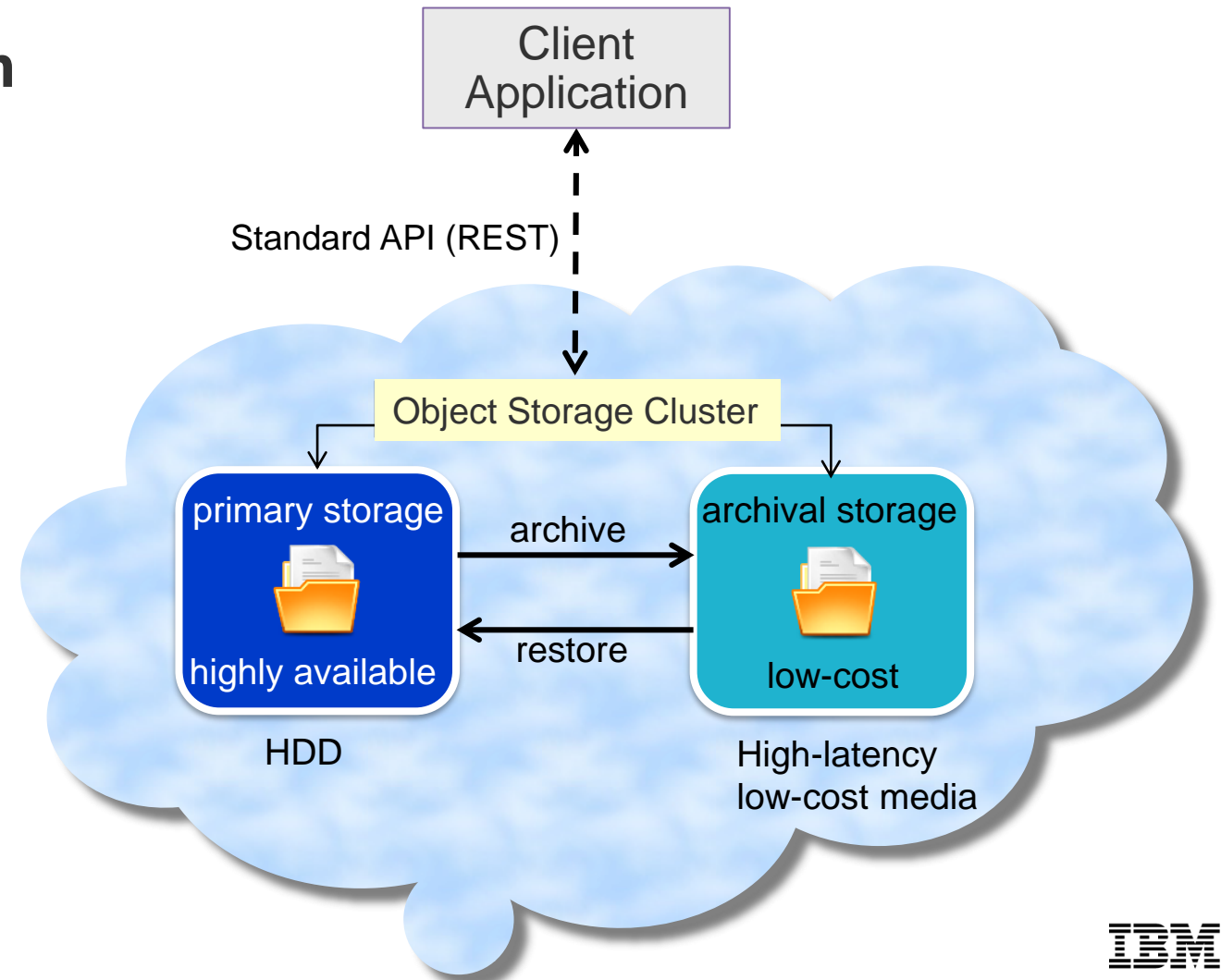
Leader, Center of Excellence for Cloud Storage
Part-time worker for IBM Research Zurich

Swiss Conference 2018



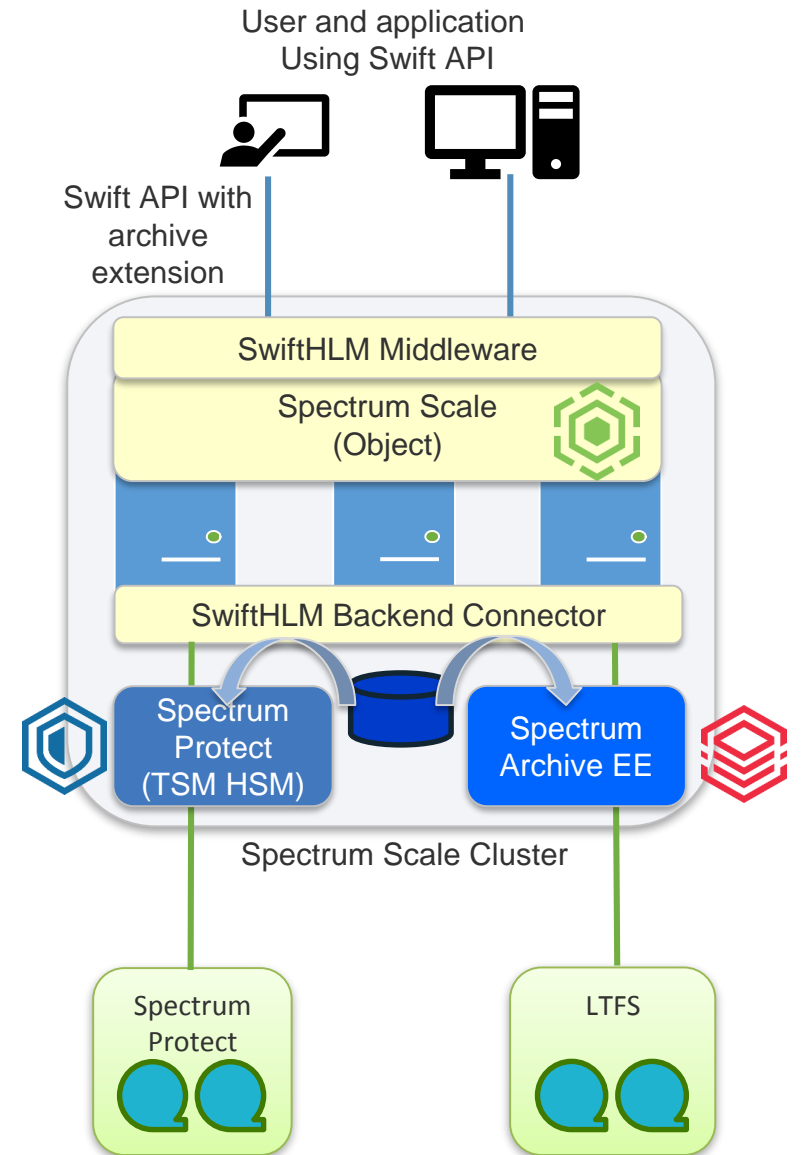
IceTier: Object storage on tape (or other high-latency media)

- **Augment cloud object storage with a low-cost, cold storage tier**
 - Tape, optical, MAID
 - Archive/backup use cases
- **Reduced cost**
 - E.g. tape up to 6x cheaper than disk (current HW/media specs)
 - Future projections in favor of tape
- **Reduced availability**
 - Minutes, 10s of minutes, or hours (depending on use case and SLA)



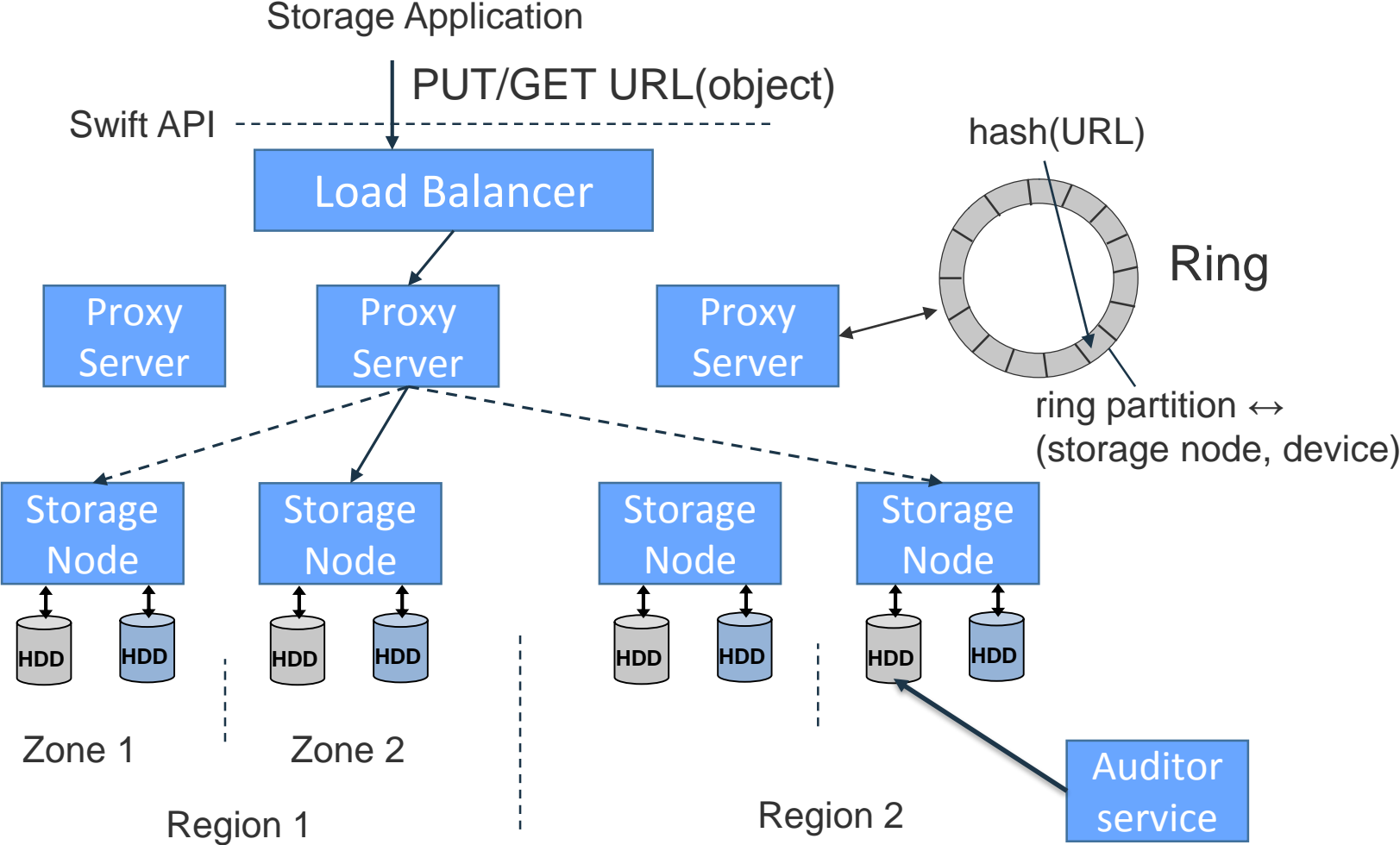
IceTier/SwiftHLM high level overview

- Shortcoming of traditional HSM solutions: limited control over data movement from and to High-latency media (HLM)
- SwiftHLM integrates tape-awareness into OpenStack Swift
 - Gives users and applications control over object and container movement from and to tape
 - Consolidates tape operations and collocates objects for better data access performance
- Benefits:
 - Integration of Tape into Cloud Storage environments
 - On-premises alternative to existing off-premises cloud archive storage offerings
 - Tape-optimized operations for efficient data access and scalability
 - Gives data movement control to users and applications
 - Integrating IBM SDS products (Scale, Archive, Protect)
 - Open for additional backends



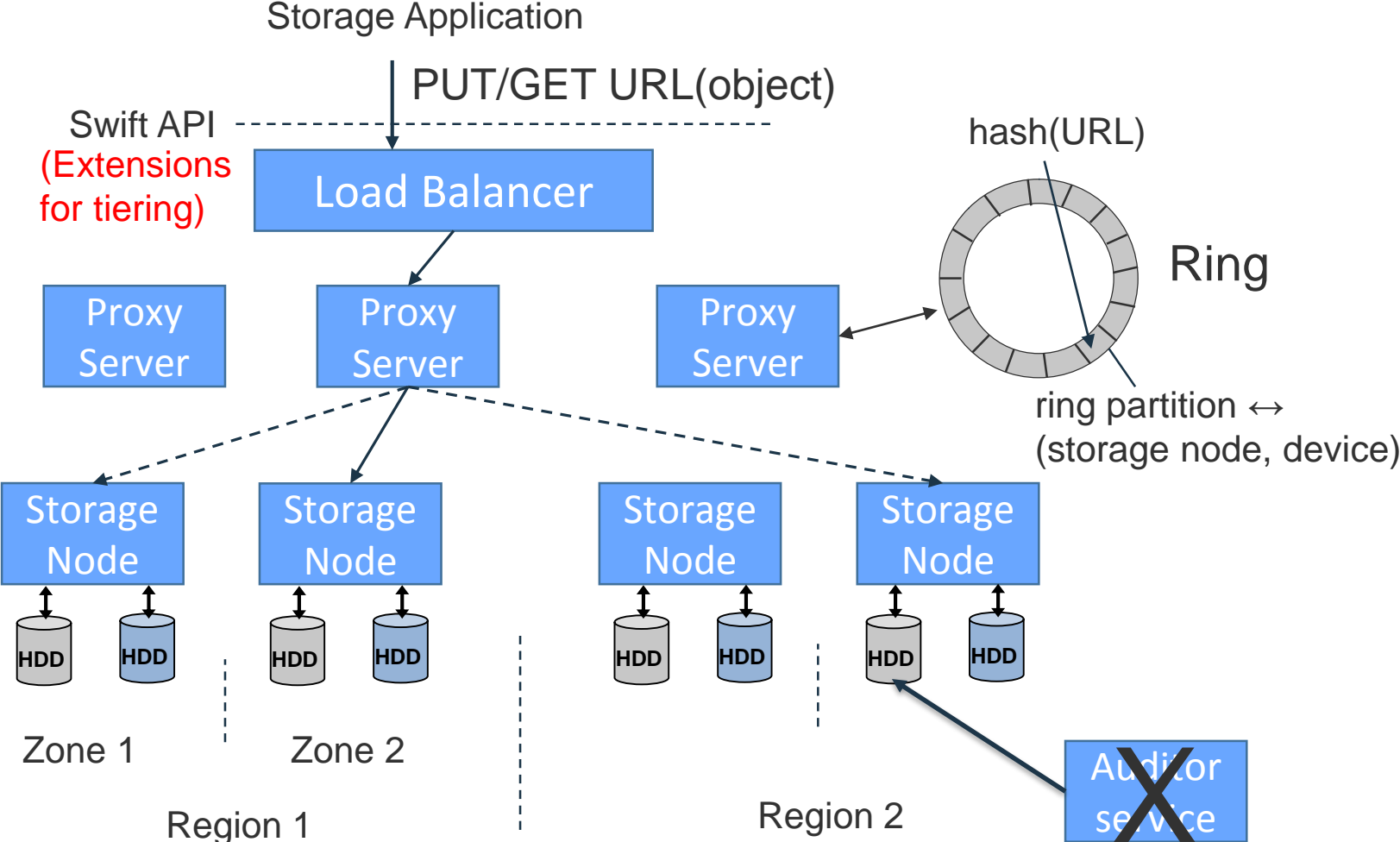
OpenStack Swift object storage on disk

- **Open Source**
 - Increasing adoption
 - Client side solutions
- **Simple REST interface**
 - Swift native
 - Amazon S3
- **Extreme Scalability**
 - Hash-based Data Rings:
 - Hash(URL) -> storage nodes, devices
 - One ring per storage policy (replication scheme, device set/type)
- **High Availability/Durability**
 - Replication
 - Erasure coding
 - Regular data health checks (auditing)



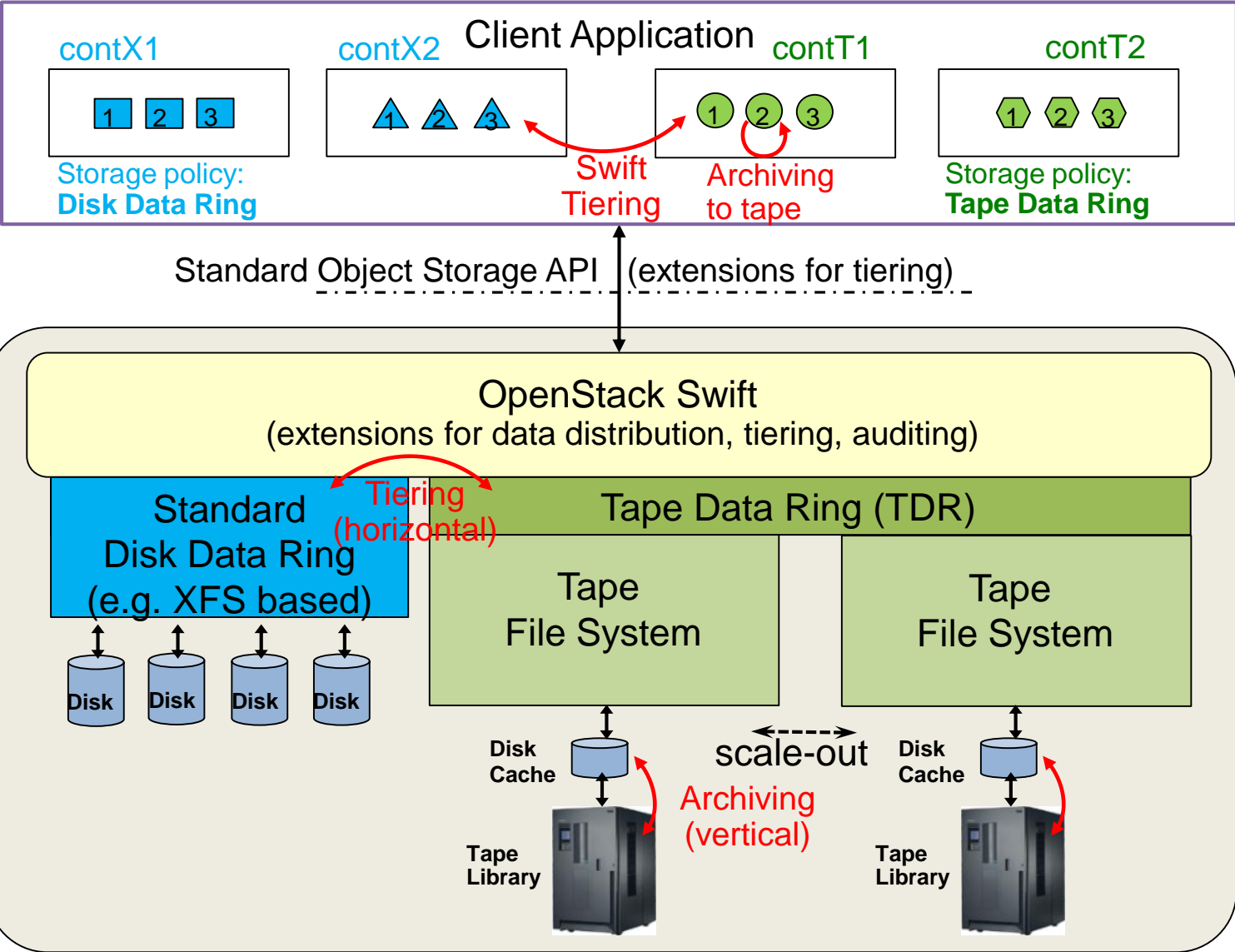
OpenStack Swift object storage – HLM extensions

- **Extend API**
 - Enable explicit archiving operations
 - **(Bulk)** migrate/**recall**/status/requests
 - Avoid timeouts
 - Cost-efficient use of drives
- **Modify health check (auditing)**
 - to not often recall tape data
- **Customize object distribution**
 - Avoid container spread over too many tapes through collocation
 - Lowers number of mounts and drives usage
 - Low cost => #drives << #tapes
- **Reuse Swift Replication**
 - Zones/Regions



OpenStack Swift object storage – HLM extensions (cont.)

- Introduce / add Tape Data Ring**
 - Single namespace for disk and tape
- Leverage Swift storage policies and ring-to-ring tiering**
 - Move data between disk and tape data rings



SwiftHLM architecture

SwiftHLM consists of

- **SwiftHLM Middleware** (Proxy nodes)
 - Proxy middleware exposing the enhanced API
- **SwiftHLM Dispatcher** (Swift node)
 - Background daemon creates a list of objects, identifies the Storage Node for each object, and dispatches asynchronously to the appropriate Swift Storage Node
- **SwiftHLM Handler** (Storage nodes)
 - provides/invokes generic interface toward SwiftHLM backend storage. Maps objects to files and submits the mapped list to the backend (via Connector)

SwiftHLM requires a **backend-specific Connector module**

- Supplied by the vendor of the backend software/hardware
 - IBM Spectrum Archive EE
 - IBM Spectrum Protect
 - others
- Note that the Connector is not part of the SwiftHLM packaging

SwiftHLM user/application API (extension of Swift API)

- Migrate/Recall

- POST `http://<host>:<port>/hlm/v1/<action>/<account>/<cont>/<obj>`
- POST `http://<host>:<port>/hlm/v1/<action>/<account>/<cont>`

`<action>` is MIGRATE or RECALL (case insensitive)

return code: 202 (ok), or an error code

- Status of submitted requests (query pending/non-completed requests)

- GET `http://<host>:<port>/hlm/v1/REQUESTS/<account>/<cont>/<obj>`
- GET `http://<host>:<port>/hlm/v1/REQUESTS/<account>/<cont>`

return code: 200 (ok), or a standard error

return value: JSON-encoded list of pending requests for object or container

- Status of objects (query status of object or container)

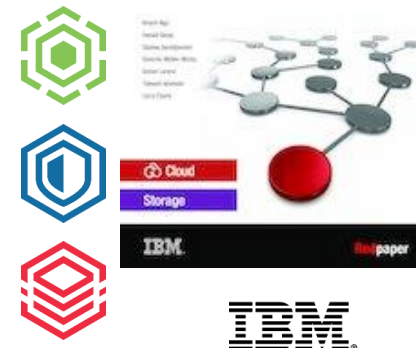
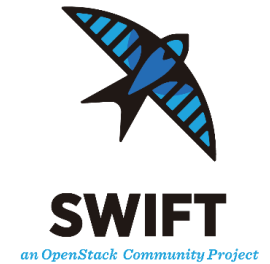
- GET `http://<host>:<port>/hlm/v1/STATUS/<account>/<cont>/<obj>`
- GET `http://<host>:<port>/hlm/v1/STATUS/<account>/<cont>`

return code: 200 (ok), or a standard error

return value: JSON-encoded list of objects and their states

SwiftHLM current status

- SwiftHLM extends the OpenStack Swift API to support high-latency media
 - provides explicit archiving and prefetching bulk operations
 - is available as Open Source Software as a Swift Associated Project
 - on the IBM Research github
 - significantly enhanced by recent distributed in-memory status cache implementation
- Generated a wide array of supporters and interested parties in the Swift community
 - BDT, RedHat, SuSE, NTT (Dev & Research), NTT-Data (storage service), Fujitsu (optical storage & tape), Panasonic (optical storage), Amethystum (optical storage)
- Active collaboration between SwiftHLM team and IBM Spectrum Software
 - Spectrum Archive bundling of SwiftHLM Connector (proprietary part) into EE V1.2.4+
 - Spectrum Protect Connector available externally at [TSM FTP site](#)
 - SwiftHLM Redpaper [published](#)



LTFS Data Management – file HSM for open filesystems

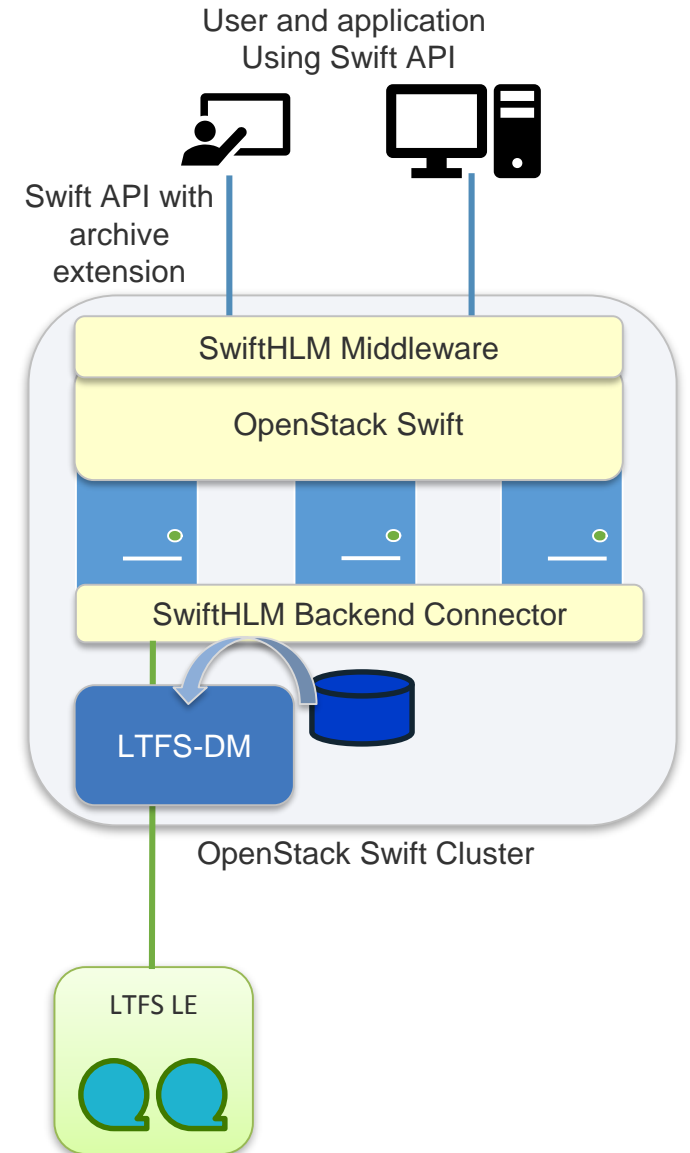
- Research project, soon-to-be Open Source
- Seamless integration of POSIX filesystems with Tape
- File migration, on-access or bulk recall
- Storage pools, replication, collocation, reclamation
- Implemented as overlay filesystem
 - exposes a unified namespace over both disk & tape



LTFS Data Management

Goal

Collaborate with the consortium to make sure that LTFS DM supports tape **hardware across vendors**, including **HPE** and **Quantum**.



IceTier/SwiftHLM outlook

- Exploring S3 API support (through Swift3 / S3API)
 - S3 lifecycle management API
- About to move to OpenStack github
 - Improving automated tests
 - Enabling additional 3rd party contributions
- Large Object support (DLO/SLO) awareness
- ACL support
- Swift versioning awareness
- Memcache footprint optimizations
- Concurrency with Swift releases

Live Demo

OpenStack Swift object storage w/ archiving to tape – demo setup

End user laptop



Swiftbrowser – Mozilla Firefox

Swiftbrowser 9.4.207.55:8000

test:tester

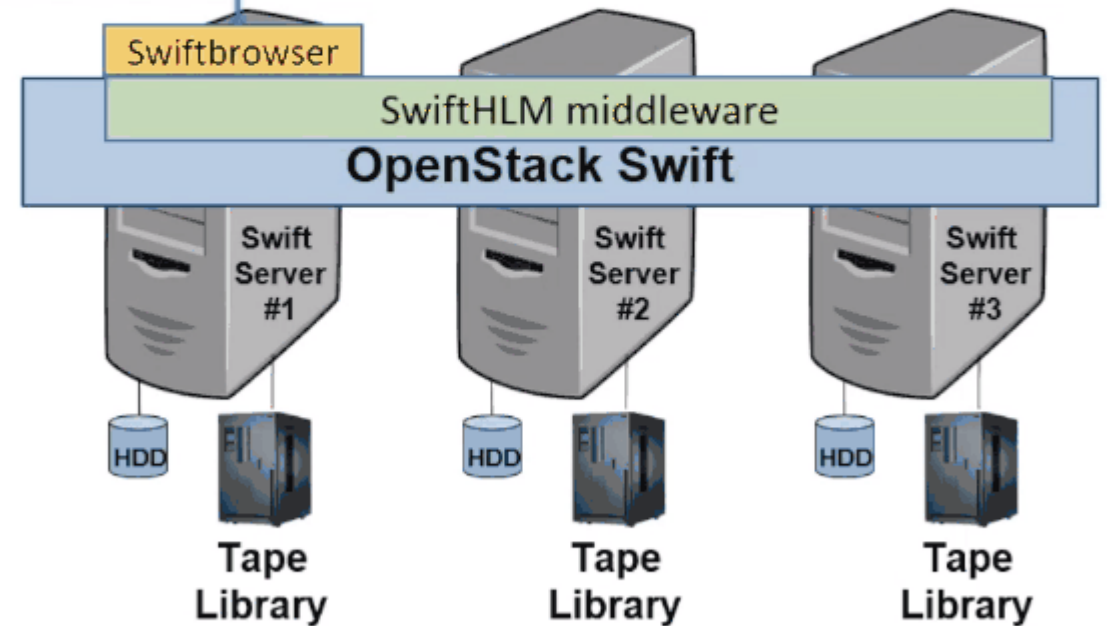
Containers

Name	Objects	Archived	Size	On disk	On tape	
cont0	10	10	10,0 MB	0 Bytes	10,0 MB	+
cont1	2	0	2,0 MB	2,0 MB	0 Bytes	-

Total size on disk: 2,0 MB

Total size on tape: 10,0 MB

Tape-equipped remote Swift object storage



SwiftHLM – Swift High Latency Media middleware (open source)

Swiftbrowser – Django Swiftbrowser with SwiftHLM API support



THANK YOU!
ibm.com/tape
seipp@de.ibm.com



Notice and disclaimers

- Copyright © 2018 by International Business Machines Corporation (IBM). No part of this document may be reproduced or transmitted in any form without written permission from IBM.
- **U.S. Government Users Restricted Rights — use, duplication or disclosure restricted by GSA ADP Schedule Contract with IBM.**
- Information in these presentations (including information relating to products that have not yet been announced by IBM) has been reviewed for accuracy as of the date of initial publication and could include unintentional technical or typographical errors. IBM shall have no responsibility to update this information. **This document is distributed “as is” without any warranty, either express or implied. In no event shall IBM be liable for any damage arising from the use of this information, including but not limited to, loss of data, business interruption, loss of profit or loss of opportunity.** IBM products and services are warranted according to the terms and conditions of the agreements under which they are provided.
- IBM products are manufactured from new parts or new and used parts. In some cases, a product may not be new and may have been previously installed. Regardless, our warranty terms apply.”
- **Any statements regarding IBM's future direction, intent or product plans are subject to change or withdrawal without notice.**
- Performance data contained herein was generally obtained in a controlled, isolated environments. Customer examples are presented as illustrations of how those customers have used IBM products and the results they may have achieved. Actual performance, cost, savings or other results in other operating environments may vary.
- References in this document to IBM products, programs, or services does not imply that IBM intends to make such products, programs or services available in all countries in which IBM operates or does business.
- Workshops, sessions and associated materials may have been prepared by independent session speakers, and do not necessarily reflect the views of IBM. All materials and discussions are provided for informational purposes only, and are neither intended to, nor shall constitute legal or other guidance or advice to any individual participant or their specific situation.
- It is the customer’s responsibility to insure its own compliance with legal requirements and to obtain advice of competent legal counsel as to the identification and interpretation of any relevant laws and regulatory requirements that may affect the customer’s business and any actions the customer may need to take to comply with such laws. IBM does not provide legal advice or represent or warrant that its services or products will ensure that the customer is in compliance with any law.

Notice and disclaimers continued

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products in connection with this publication and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products. IBM does not warrant the quality of any third-party products, or the ability of any such third-party products to interoperate with IBM's products. **IBM expressly disclaims all warranties, expressed or implied, including but not limited to, the implied warranties of merchantability and fitness for a particular, purpose.**

The provision of the information contained herein is not intended to, and does not, grant any right or license under any IBM patents, copyrights, trademarks or other intellectual property right.

IBM, the IBM logo, ibm.com, AIX, BigInsights, Bluemix, CICS, Easy Tier, FlashCopy, FlashSystem, GDPS, GPFS, Guardium, HyperSwap, IBM Cloud Managed Services, IBM Elastic Storage, IBM FlashCore, IBM FlashSystem, IBM MobileFirst, IBM Power Systems, IBM PureSystems, IBM Spectrum, IBM Spectrum Accelerate, IBM Spectrum Archive, IBM Spectrum Control, IBM Spectrum Protect, IBM Spectrum Scale, IBM Spectrum Storage, IBM Spectrum Virtualize, IBM Watson, IBM z Systems, IBM z13, IMS, InfoSphere, Linear Tape File System, OMEGAMON, OpenPower, Parallel Sysplex, Power, POWER, POWER4, POWER7, POWER8, Power Series, Power Systems, Power Systems Software, PowerHA, PowerLinux, PowerVM, PureApplica- tion, RACF, Real-time Compression, Redbooks, RMF, SPSS, Storwize, Symphony, SystemMirror, System Storage, Tivoli, WebSphere, XIV, z Systems, z/OS, z/VM, z/VSE, zEnterprise and zSecure are trademarks of International Business Machines Corporation, registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the Web at "Copyright and trademark information" at: www.ibm.com/legal/copytrade.shtml.

Linux is a registered trademark of Linus Torvalds in the United States, other countries, or both. Java and all Java-based trademarks and logos are trademarks or registered trademarks of Oracle and/or its affiliates.

Backup

SwiftHLM components

