



ACADEMIC
PRESERVATION
TRUST

Evolving a Multi-Cloud Strategy for Preservation

Nathan Tallman, Executive Director

Andrew Diamond, Lead Developer and Technical Architect

Flavia Ruffner, Lead Engineer for DevSecOps

Library of Congress, Designing Storage Architectures

April 15, 2024

Introduction

- APTrust is a cloud-native distributed digital preservation storage consortium, in production since December 2014.
- The original technical vision was a portable service that could be lifted and moved to any Infrastructure-as-a-Service provider.
- Expanding cloud boundaries

Initial Wasabi Exploration

- Hot storage, no egress, predictable costs: great.
- Some retention requirements.
 - 90 days? Totally acceptable for a long-term preservation repository.
- Options for ingress from AWS?
 - Pay full AWS egress costs? Expensive.
 - Direct Connect? Only saves money if you pass a certain transfer threshold and we don't know if we'll reach that.

Implementation

- Code changes:
 - Use Minio S3 client
 - Point it at any S3-compatible service and it works
 - Config changes to our services tell code that Wasabi objects are eligible for fixity checks and are preferred restoration sources
- New storage option settings for our depositors
 - Want copies in Wasabi? Just set the right storage option parameter
- Hooray for S3 protocol getting such wide support!
 - Geographic and vendor diversity are important for long-term preservation
 - Adding new storage providers in future will be trivial (on the technical side - lawyers are a different story)

Goals of the Wasabi Proof-of-Concept Test

- Primary was to accurately assess live data migration egress costs end to end from AWS to Wasabi in the modernized APTrust infrastructure.
- Analyze what types of data and networking rates AWS was actually charging.
- Provide membership options for architectural changes and their impact based on POC analysis.
- Have a basis to review other data migration connection options.
 - Direct Connect - AWS and Wasabi
 - 3rd Party options -Acembly/Equinix.

Design of the Wasabi Proof-of-Concept Test

- Criteria
 - A few volunteer members, no other uploads in the test window.
 - Upload 100-250GB of data to Wasabi VA and OR buckets
 - Only to Demo environment.
- Timeline
 - 1 month
- Cost tracking
 - Using the cost explorer for analyzing specific data flows and activities.

Lessons Learned from the Wasabi Proof-of-Concept Test

- Direct Connect requires a 3rd Party,(their data centers) and expensive for the amount of data we migrate.
- But now we know the threshold at which direct connect makes sense, and we can watch depositor behavior to see if we approach that threshold
- NAT Gateways are expensive for security provided, and other strategies should be considered when moving lots of data .
- Additional work for logging aggregation audit capabilities.

Putting it in Production

- Removing NAT Gateway and related rearchitecting
- Determining cost for members
- Integrating Wasabi logs into the centralized logging service
- Updating policies, guidelines, and documentation

Conclusion

- A multi-cloud strategy can vary
- Digital preservation in the multi-cloud
- Standardizing the S3 API would facilitate multi-cloud development