



# Cloud Native Preservation

Digital Preservation 2019





# Bill Branan

- LYRISIS - Sr Engineering Lead
- DuraCloud Technical Lead
  - Open Source project - est 2009
  - Preservation utilizing cloud storage
  - Preservation utilizing academic DDPs
  - Hosted by DuraSpace, now LYRISIS

# Cloud object storage

A data storage architecture that manages data as objects that are free of hierarchy, allowing all data objects to exist at the same level in a storage pool. Each object includes the data, a unique identifying name, and metadata. Cloud providers offer object storage as a massively scalable and low cost storage solution.

## **Examples**

Amazon S3, Microsoft Azure Blob Storage, Google Cloud Storage,  
Rackspace Cloud Files, Wasabi Cloud Storage

# Cloud object storage benefits

- Virtually unlimited capacity
- Low cost
- Greater institutional / departmental / individual control
- Access to content for processing
- Ability to use tools built against cloud APIs
- Reduced maintenance
- Access / availability
- Security controls



Are cloud features sufficient for  
preservation?

# Features of a Digital Preservation solution

- Create multiple distributed copies of content
- Maintain inventory of all content and storage locations
- Maintain audit log of all content stored and all changes made
- Verify file fixity / integrity
- Provide an administrative dashboard

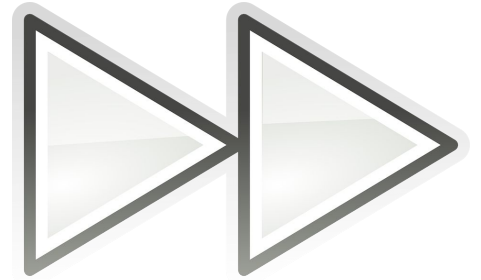
**Partial  
Support**

- Perform metadata extraction
- Perform format identification
- Perform virus scans
- Perform format migrations

**No  
Support**

# The Future

- Support all preservation functions
- Work in a cloud environment
  - To take advantage of cloud capabilities
  - To benefit from new cloud features and functions
  - To avoid egress charges
- Shared approach, tools, and best practices
- Shared development overhead



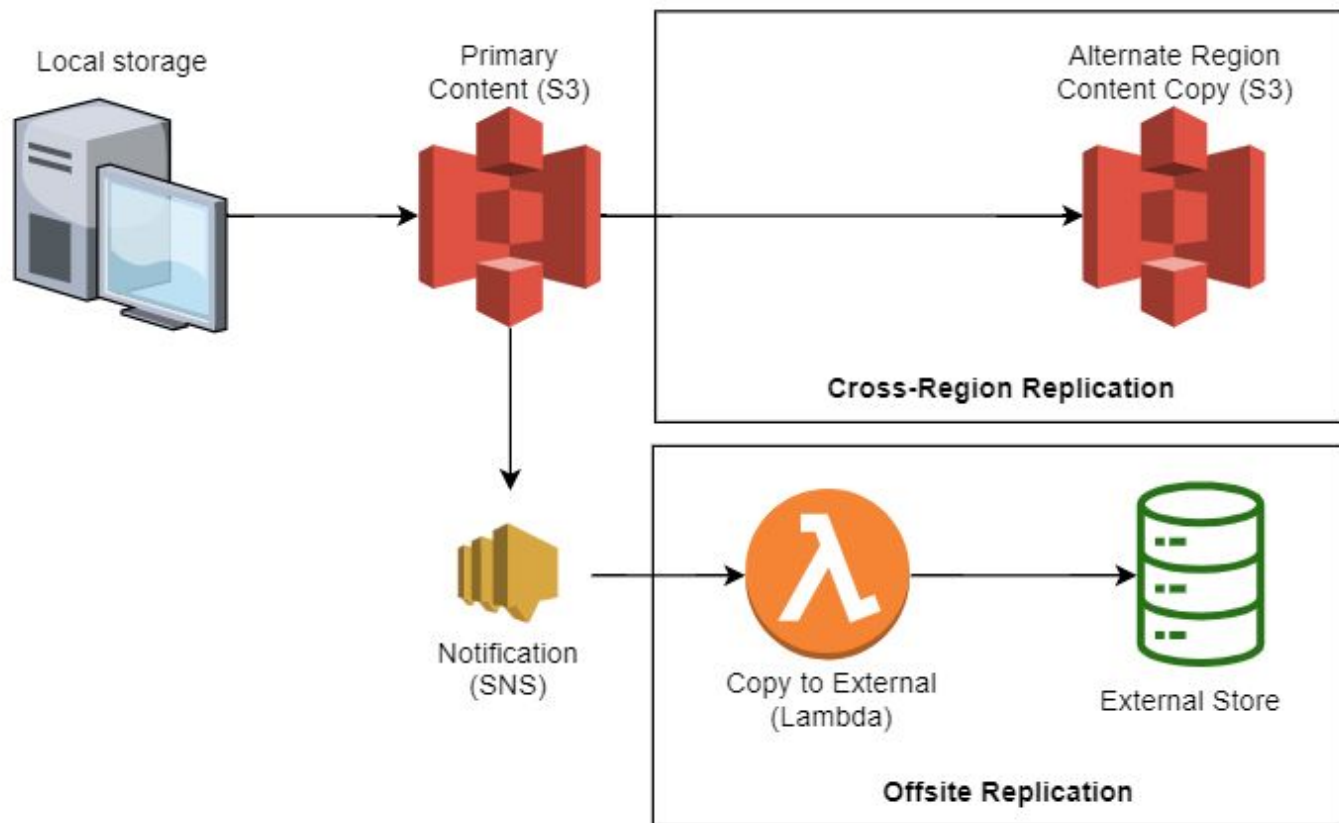
# Cloud Preservation Toolkit

- A set of tools to perform preservation functions
  - Independent deployment - use only what you need
  - Consistent, repeatable deployment using configuration management tools
  - Primarily serverless
- A community of developers to maintain and improve tools
  - Open source function code
  - Open source configuration management scripts
  - A community of practice

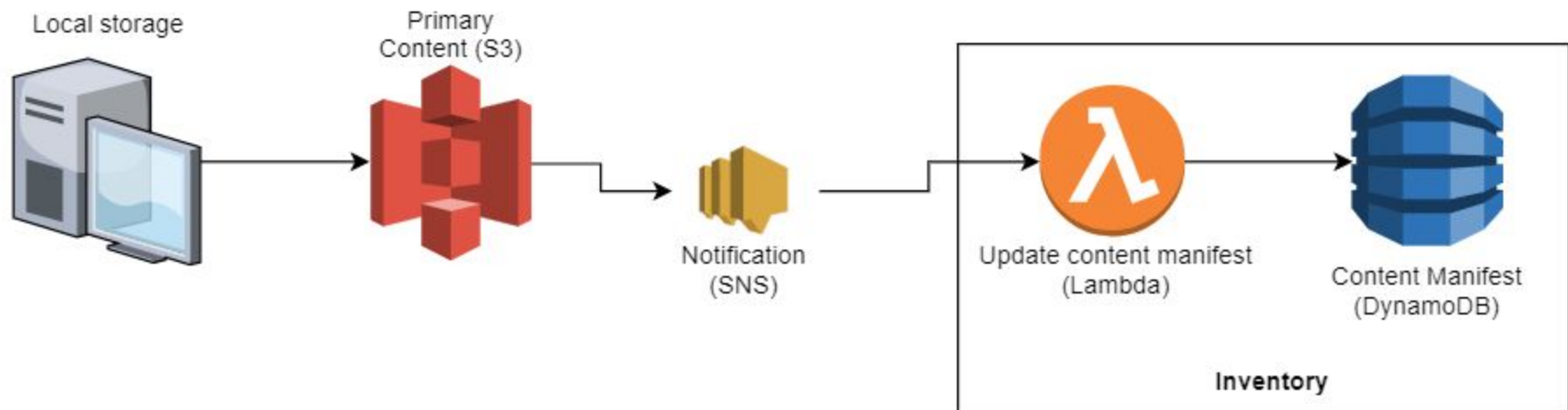




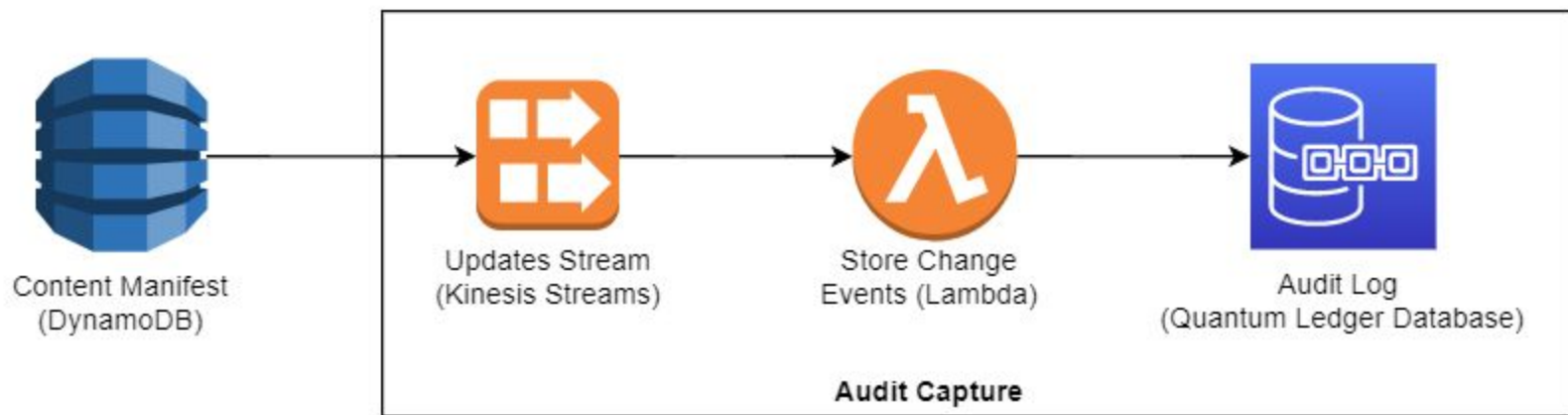
# Multiple copies



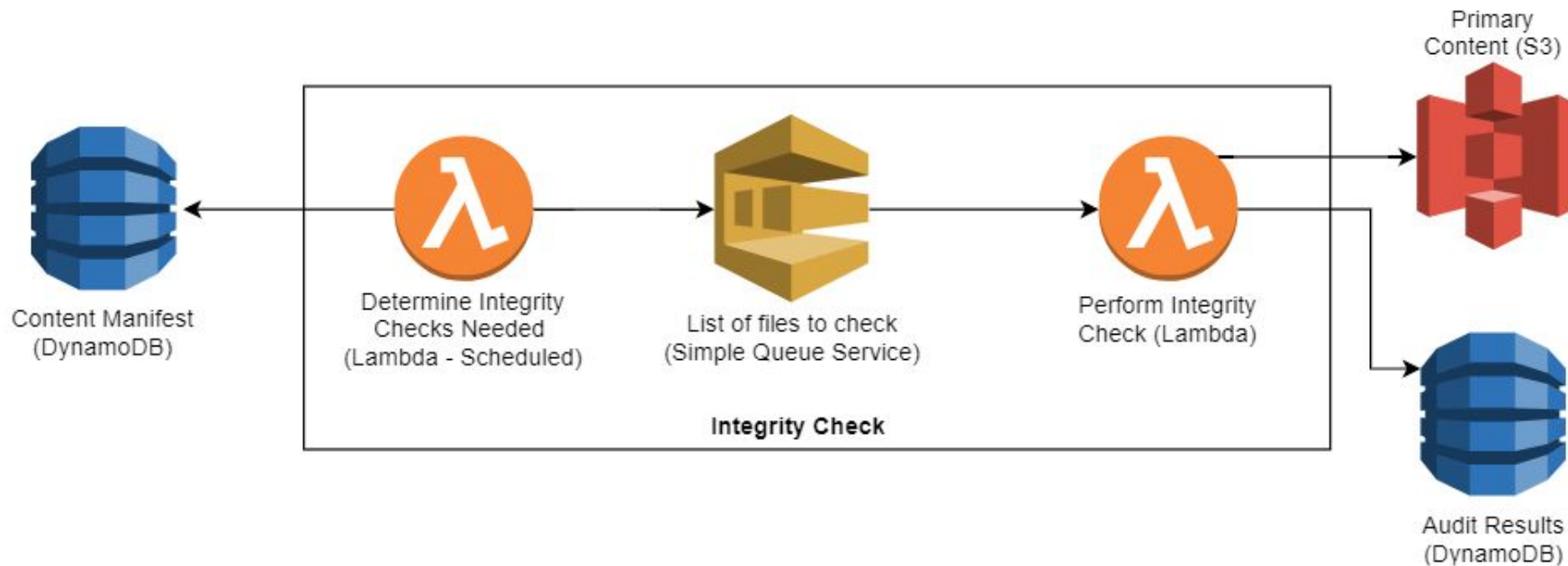
# Inventory



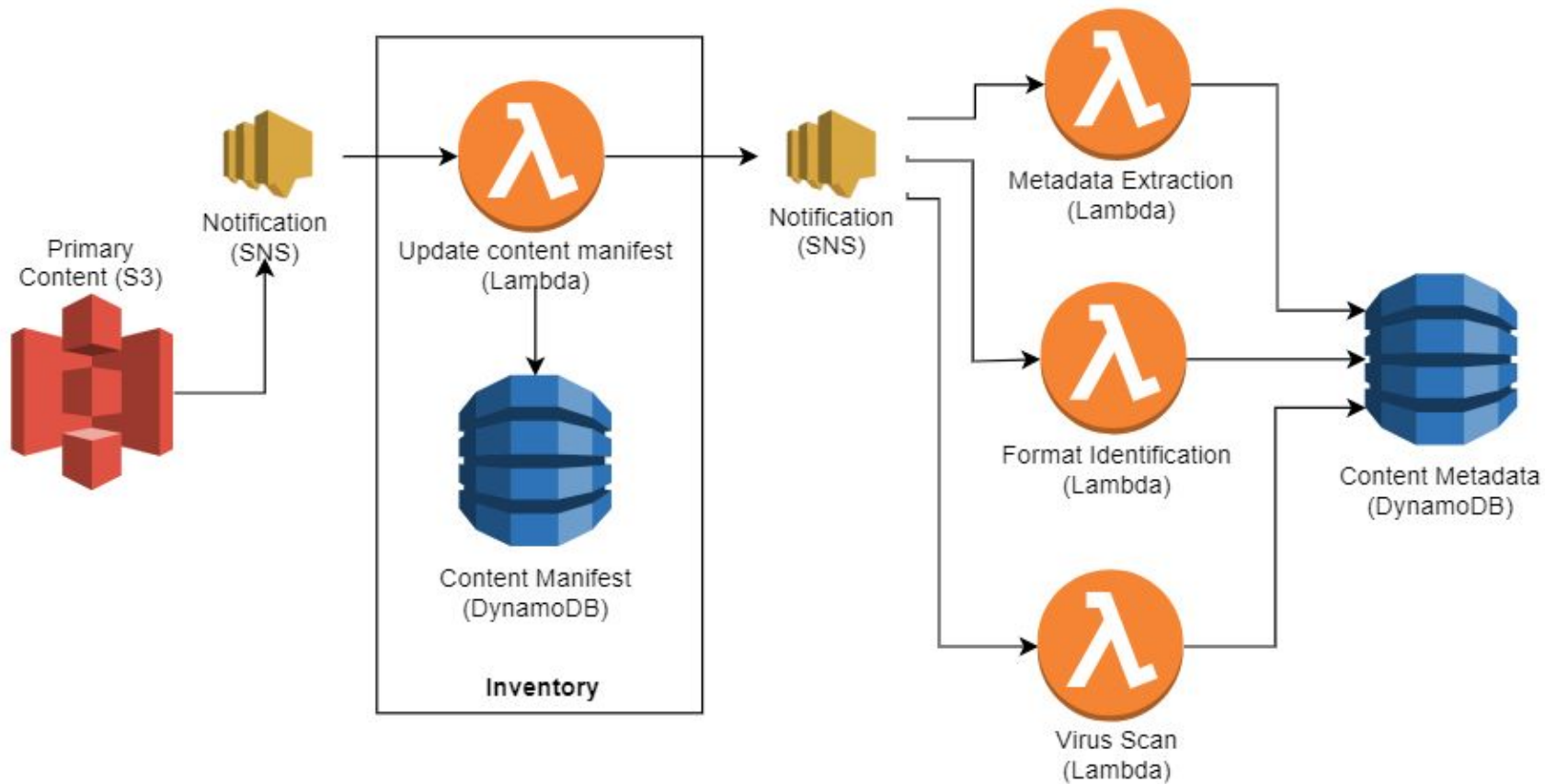
# Audit Log



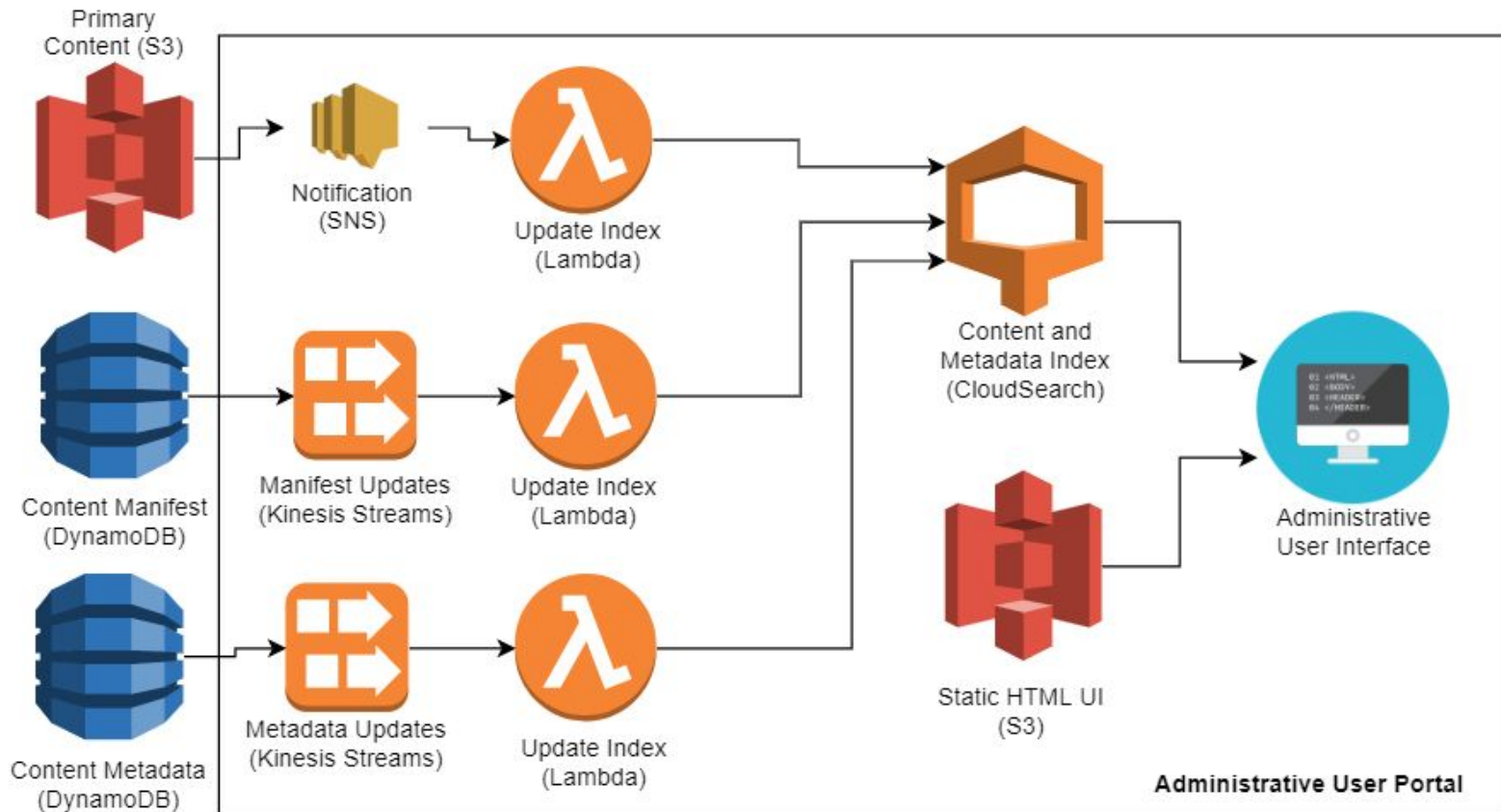
# Integrity Checks



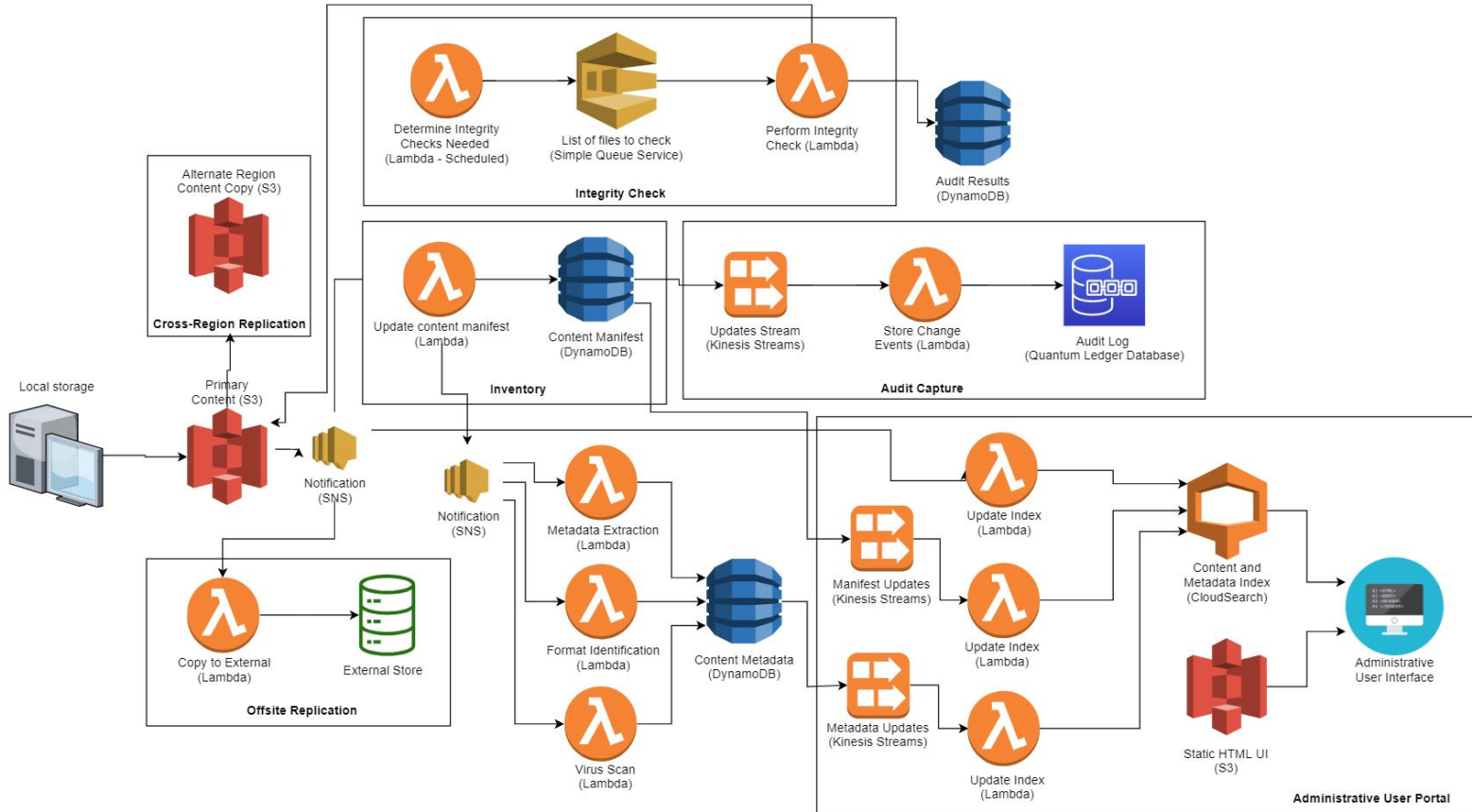
# File Processing



# Administrative UI



# Putting it all together



# Benefits

- Convenient - Allows the continued use of cloud storage for high-value content
- Efficient - Preservation actions occur within the cloud environment
- Shared - Consistent set of tools used and maintained across institutions
- Configurable - Use only the components that are relevant to you
- Manageable - Each function is limited in scope
- Scalable - Serverless tools allow for broad scaling
- Low cost - Serverless tools ensure charges only apply when work is done
- Automated - Deployment and updates managed by CM tools

Confidence that data is preserved!



Is this the future you want to see?



**[bill.branan@lyrasis.org](mailto:bill.branan@lyrasis.org)**