Staging Server

- 1 Setting up an Amazon S3 bucket as a MetaArchive staging server
  - 1.1 Pre-Ingest AWS S3 Bucket Configuration
    - 1.1.1 Properties
      - 1.1.1.1 Static website hosting
      - 1.1.1.2 Server access logging
    - 1.1.2 Permissions
      - 1.1.2.1 Access Control List
      - 1.1.2.2 Bucket Policy

A staging server is a location where archival units are placed to await ingest into the MetaArchive LOCKSS network. A plugin is used to "show" the LOCKSS network where the incoming AUs are located.

Staging Server/Web Server/Harvest Server - A server that is open to the public or controlled access for the MA network to be used for crawling the content. MetaArchive members add content to a staging server where it is harvested first for testing, and, once a test crawl is successful, the content is harvested from the staging server into 5 of the MetaArchive network's LOCKSS servers.

Setting up an Amazon S3 bucket as a MetaArchive staging server

This documentation describes the process for setting up an Amazon S3 bucket as a MetaArchive staging server. MetaArchive member representatives can share this document with their institutional IT staff and coordinate with them to implement the necessary configurations. MetaArchive members that have successfully configured and ingested content via this S3 method are also available to consult. MetaArchive member representatives, their admins, and IT staff should verify that the policy configurations documented herein do not conflict with other institutional policies on access to computing resources. Note that the costs associated with this are for storage and data transfer and if you currently do not pay for either in your local environment it may not be cost effective over setting up a local server.

Pre-Ingest AWS S3 Bucket Configuration

Properties

Static website hosting

Create the S3 bucket. On the Properties tab for the bucket choose “Static website hosting” and click “Use this bucket to host a website”. For Index document put "index.html". Note the value of the Endpoint which will be something like http://your-bucket-name.s3-website-your-region.amazonaws.com, i.e., http://your-bucket-name.s3-website-us-west.amazonaws.com.

Server access logging

Turning on “Server access logging” is recommended to be able to monitor and troubleshoot any issues. To do this select “Enable logging” under “Server access logging” and choose a previously created S3 bucket to store your logs.

Permissions

Access Control List

On the Permissions tab under “Public Access” select “Everyone” and check “List objects”. There will be a warning that, “This bucket will have public access.” Notify the system admin because Amazon will email a warning that the bucket is public.
Bucket Policy

The current Bucket Policy is on GitHub. Replace your-bucket-name, and cidr-for-local-testing, and replace IPs with the MetaArchive Allow List (credentials are listed on this wiki page).