Technical session 2: Data

Friday, October 11, 9:40 - 10:20 AM

Session overview

Lead = Josh Rhoades, Axiom Data Science, Inc.

Format = Short presentations followed by open discussions (interactive discussions with slides)

Topics (in order of interest from our survey):

1. Camera integration: Costs, time, maintenance for installations vs. Cameras of opportunity
2. Overall data management pipeline and workflow
2. Technical requirements for different product applications
4. Data access methods
8. Website use and improvements
9. Integrating code from our existing WebCOOS algorithms into new cameras
10. Camera metadata

Axiom Data Science

Who we are:

- Headquartered in Anchorage, Alaska, with data center in Portland, Oregon
- Support a variety of federal, private, academic and non-governmental organizations managing and conducting research in the ecological, geological and ocean sciences.
- Dr. Karina Khazmutdinova, Project Manager (karina@axds.co)
- Josh Rhoades, Software Engineer (josh@axds.co)

https://axiomdatascience.com/

Camera integration challenges (installed vs. CoO)

New, standardized installs (via Surfline):

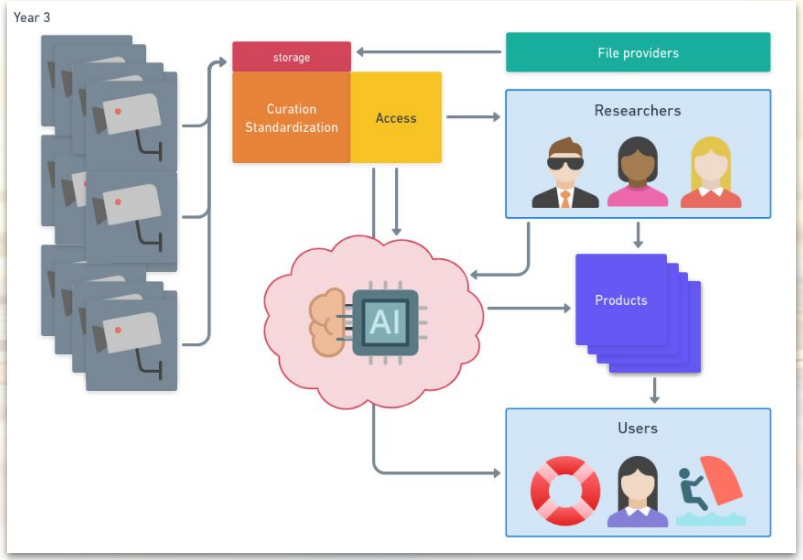
- Up front cost for installation, and likely also for maintenance (camera upgrades) and remote troubleshooting.
- Standard camera hardware and options allow for streamlined ingests and fewer unexpected behaviors.
- (Hopefully) easier hardware replacements or upgrades over time.
- Better control over hardware (NDAA security compliance)

Cameras of Opportunity (CoO):

- Network, power, and access methods may exist, but have to be discovered (and likely, debugged)
- Tend to be labor intensive, less reliable, and can exhibit unexpected behaviors.
- CoO can’t often be remotely managed.
- Non-standard or non-compliant hardware (NDAA security compliance)

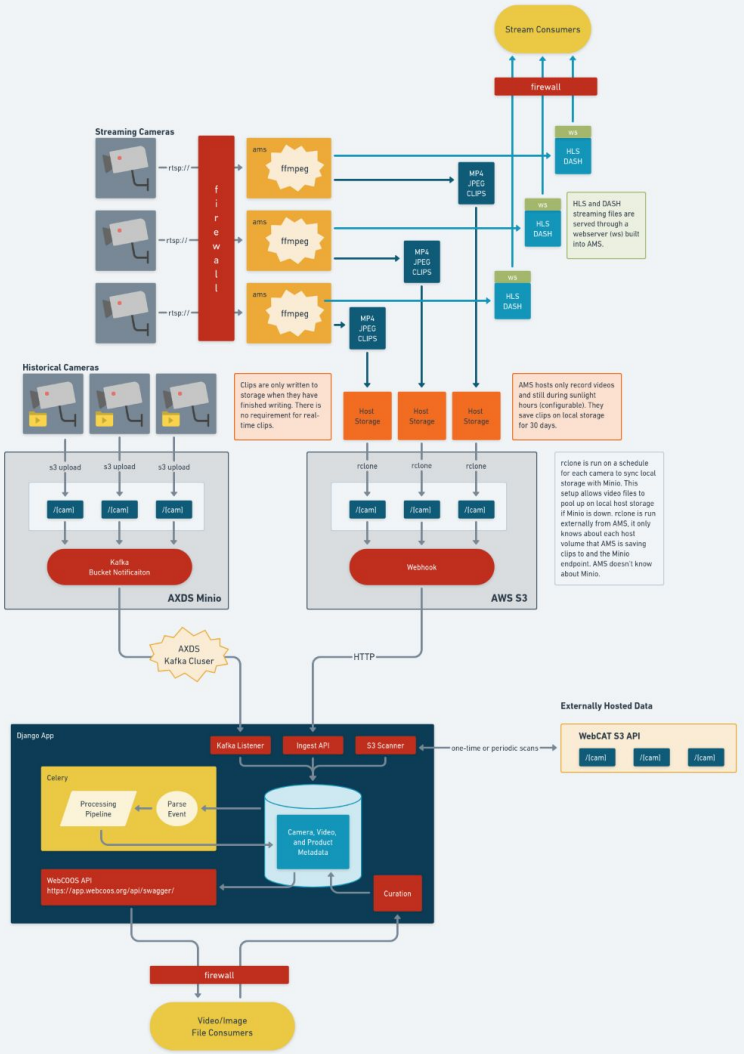
Overall data management pipeline and workflow

- Real-time video stream ingests over the network, or externally provided media (S3 upload)
- Post-processing for:
  - Media metadata indexing, and…
  - Secondary product generation (detections, time averages, brightest pixel)
- Website and HTTP/REST API access for results



Overall data management pipeline and workflow (cont.)

- Includes “live” streaming products (HLS/DASH streams over HTTP)
- Multiple S3 storage options
- Ingest orchestration via Ansible and Docker.
- Live stream and downstream product monitoring via Prometheus/Grafana
- Post-processing via custom Python scripts and web services wrapping researchers’ code provided to Axiom.



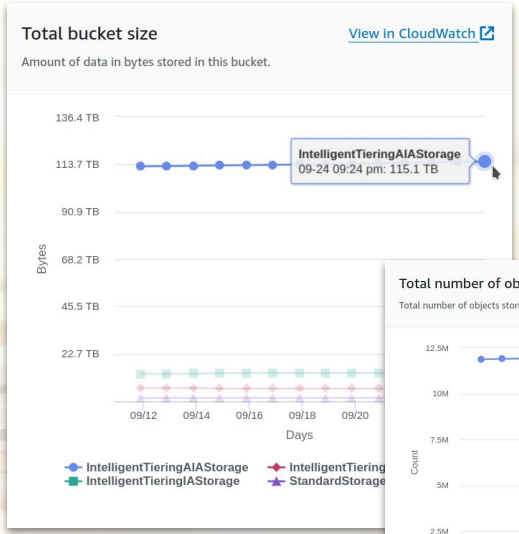
Integrating code from existing algorithms into new cameras

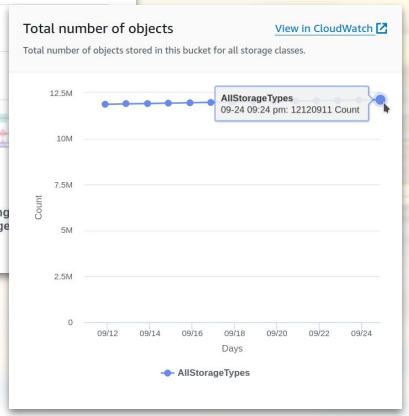
- Existing:
  - ML detections (2 rip current models, 1 seal model, 1 general-purpose object detection model)
  - Shoreline Detection (with brightest pixel and time averaged pixel products)
- Uses Axiom compute resources (CPU, GPU, storage, network)
- Lots of compute, but also, very labor intensive (code collaboration, adapting existing code to HTTP services, testing, and performance profiling)

5 minute load average over 48 hours (9/25/2024)

Technical requirements for different product applications

- CPU and GPU compute
  - ML model detection tends to require GPU support (both GPU compute and VRAM)
- Scaling: number of cameras over time
  - ~200GB a day
  - ~300-700 media files per asset per day (that potentially need some form of processing)
- Currently: 9 physical hosts, 3 GPUs





136.2TB among 12 million media files

Data access methods

Incoming (methods of ingress to AXDS):

- RTSP/RTMP (“push” or “pull”, depending on network situation)
  - RTSP/RTMP “push” method similar to what is used for Youtube, Twitch, etc.
- HLS/DASH streams (via HTTP)
- S3 upload

Outgoing (methods of egress from AXDS):

- webcoos.org (website)
- WebCOOS API (HTTP/REST)
- HLS/DASH streams (via HTTP)
- S3 download
- HTTP indexed directories (less used)

Website use & improvements

- Not technically broken, but definite room for improvement.
- Josh would like:
  - deep-links (link to a specific time, or a specific media element)
  - Fix UI timezone issues
  - Revisit the existing deployment workflow (very labor intensive)
  - Integrate better into Axiom’s other portal products (better time series integration, better map integration)

Camera metadata

OpenAPI spec:https://app.webcoos.org/api/swagger/

- Media files with metadata indexed in Postgres (available via WebCOOS API request)
  - Time extent (duration)
  - Size
  - Location (lat/lon)
- JPG and MP4 files also updated with EXIF metadata (lat/lon, camera title, with UTC timestamp in basename)
- JSON files also serve to further annotate ML detection and shoreline detection results.