Agenda

- living room

SOW

1. Discussion and validate direction
2. Next Steps
3. Brian Killough - role? organization,
    - access to our buckets
    - should we make them public readable
4. Research Questions
    - What use cases does a stock ARD (tar) directory cover/serve?
    - Simple Untar and Deduplication - how much better is this configuration?
    - Is reworking the geotiff to be a COG worth the effort?
        - Does the tile size of 4000 x 4000 impeed the utility of a COG?
            * should it be 128 * 32 = 4096 instead?
    - What use cases will HDF 5 (distributed s3) support? and when?
    - What are the strengs and weaknesses of AWS S3
    - How mature is s3fs?
    - What are all of the access methods for S3?; What use cases do they support?
        - http
        - gdal and vertical derivatives
        - static web page urls
        - aws s3 cli and aws s3api cli access
        - python boto3 bindings
        - vsicurl - vsis3 - vsitar ... etc
    - What security best practices will be needed to strike the right ballance of access versus availability?
    - Region and failure domain strategy for USGS EROS?
        - US ARD
        - Global ARD
        - level1 tier 1 collection 1
        - lessor tiers and legacy data
5. Start to document findings ina white paper
    - partner with the DAAC
    - partner with Gacke

Deliverables

1. This week a full COG bucket with ingestor geotiff settings
    - ga-odc-eros-co3-west

- ga-odc-eros-un3-west
- ga-odc-eros-ard-west

2. Next week
   - terraform docker ship creation and distrcution - IaC Infra as code
3. Feb 14 2018 - presentation and findings - eng labs - pandas - rasterio - dc.load etc.
   - s3 geotiff ard findings paper
     - s3 good for this
   - follow on researh - how about processing in the cloud
   - transcient plus cassandra
   - AWS all in strategy
   - other cloud providers; generic approaches

Approach

1. Create an engineering strategic filter
   - must be open source
   - must use python
   - must be cattle; must not be pet
   - must be able to learn and deploy in a single day
   - must be agile - not blocked on anything - $, politics, procurements, people etc
   -
   -
2. Develop ninja skills in necessary technologies
   - terraform
   - python
   - gis, web mapping,
   - data science
   - nosql - key value
   - orchestration
   - message queueing and job control
   - AWS stuff
     - s3
     - ec2
     - ebs
     - ecs
     - rds; maybe GA here
     - lambda
     - ems; maybe lsrd unicorns here

Benefits of the Collaboration for EROS

1. No longer the ones not coming to the table

2. Ability to collaborate with an experienced and diverse group
    - even if its only creaping on the slack channels
    - general awarness
    - better defines the remote sensing landscape and ways to exploit
3. Likely to accelerate USGS EROS participation in disruptive technologies = most notably the AWS cloud
4. Ultimately lowers the cost of many of the government's objectives
5. Move us back to an open systems development paradigm in the face of ...
6. Consistent with teh direction of cloud first

Tour

Slack

Ideas

Brian's questions and suggestions - the Sauer Hour