

Early adopters of the Earthdata Cloud are in a unique position to take advantage of cloud-hosted NASA earth science datasets and in-place analysis, the future of data access and science. However, early adoption comes with the burden of encountering bugs, inefficiencies and other challenges associated with a shifting paradigm.

Obstacles and barriers encountered by NASA Earthdata Cloud early adopters are presented here, where **we encourage you to concur, push back, or suggest solutions**. Feedback collected here will be shared with cross-DAAC groups who are working toward collaborative solutions for Earthdata Cloud users.

Special acknowledgement of the organizations and people who provided the space and support for our early cloud adopters. We thank the NASA Openscapes mentors, Openscapes leadership, and International Interactive Computing Collaboration (2i2c).

Alexis Hunzinger^{1,2}, Christopher Battisto^{1,2}, Binita KC^{1,2}, Allison Alcott^{1,2}

¹NASA Goddard Space Flight Center, ²ADNET Systems, Inc.

alexis.hunzinger@nasa.gov

Learning Curve

"I don't have the time or energy to learn so many new concepts" -UWG member



- Those teaching are also continuously learning
- Must keep using new knowledge in order to retain it (opportunity, space, discipline)
- New jargon and concepts in computing, data format

Cloud Optimization

"[The cloud] is faster, but only by a factor of 2. Our setup and data granules aren't structured to take advantage of a faster cloud."

– Openscapes Champion



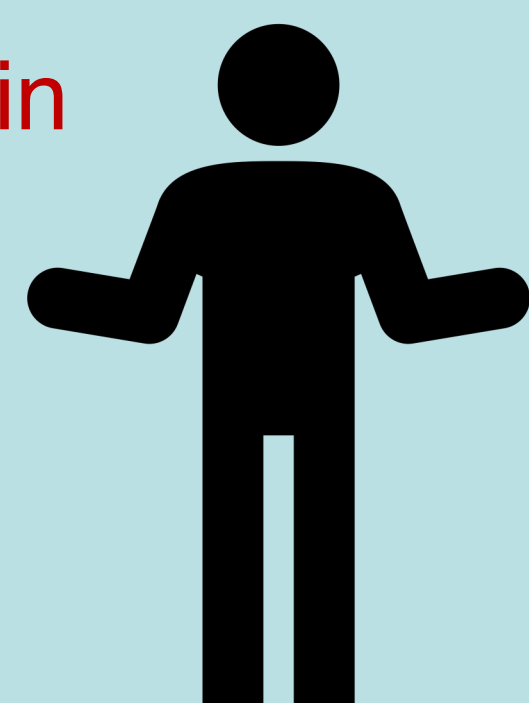
- Compute and monetary costs of bringing non-optimized workflows to the cloud
- Learning what parts of workflow are unchanged, what requires alternative solutions

"We were unable to open the HDF data directly through S3 access, so had to download them in our [EC2] instance to be able to process them."

– Openscapes Champion

Cost Estimation and Overhead

"Major challenge: budgeting for cloud in a proposal" – Openscapes Champion



- Cost model is new and obscured
- Costly to run trials in cloud (personal credit card?)
- Institutional support falls short or reluctant to adopt cloud resources

When To Use The Cloud

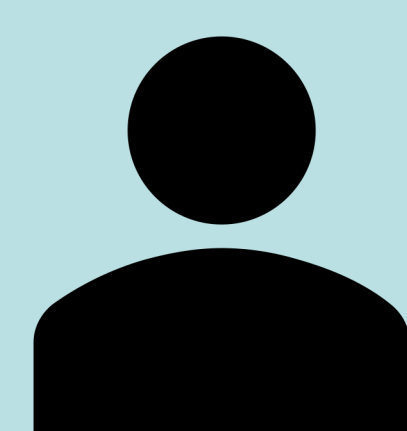


"My workflow isn't parallelizable, the file format isn't cloud-optimized, I'd rather download everything I need and perform my analysis locally."

– Openscapes Champion/UWG member

- Unoptimized data and code can be costly (\$ and energy)
- Services that are/are not available in the cloud
- Balancing data volume with download speed
- Team compute needs (shared space? Compute power?)

Cloud Knowledge and Management

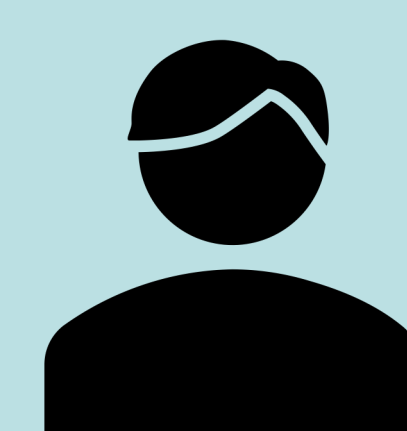


"How does using data from AWS (Earthdata Cloud) work with GEE? Advantages and disadvantages between the two?" – Openscapes Champion

- Data storage cost and locations (personal data, NASA Earthdata, inter-cloud interactions)
- Advanced computing concepts that were previously obscured

"Worried about doing things 'on my own', outside of a cloud-hosted JupyterHub. How do I utilize the 'spot market' effectively?"

– Openscapes Champion



- Transition out of supported, easy entry to cloud environment (i.e. Hub to your own AWS account)
- Available options for pre-determined environments

How Can We Use Open Science to Overcome Obstacles?

TRANSPARENCY

ACCESSIBILITY

INCLUSIVITY

REPRODUCIBILITY