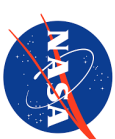




GESDISC



Summer ESP 2024

Leaving the Nest: Have users incorporated cloud computing into their work after finishing supportive cloud-focused programs?

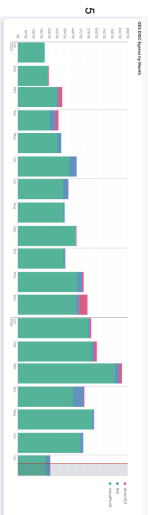
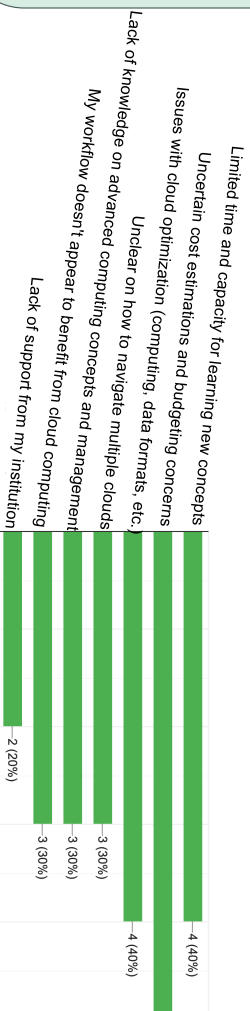
NASA OpenScapes and NASA DAAC staff publish and distribute educational resources on cloud data access and computing, and provide access to a cloud computing platform free of cost to hundreds of scientists, what we call *onboarding*. While engagement and usage is high during programs and one-on-one use cases, not all scientists move away from the provided platform and set up their own cloud compute afterwards, what we call *offboarding* or *fledging*.

Based on feedback we have received in the past 3 years and responses to a survey on cloud usage after programs end, we present the primary reasons users struggle to effectively use the cloud on their own.

About These Users

- 15 of 125 past participants responded to our survey.
- Each gained access to a cloud-based JupyterHub between 2021 and 2024.
- Primarily from **Academia & Government**. Their interest in the cloud was triggered by:
 - Government agencies moving data to the cloud
 - Curiosity about alternatives to on-premises servers and maintenance
 - More efficient computing for model analyses and increasingly larger datasets

What parts of cloud computing are you struggling with?



Evidence from one of NASA's data archives:
How can we tell if users are working in the cloud?

Bytes downloaded or accessed from the NASA GES DISC data archive between January 2023 and July 2024.

What's your experience?

What do you or your users still struggle with?

What solutions have you or your users found? Have you tried?

Cost Estimation and Justification

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

- Uncertainty of cost when we scale up processing or reprocess
- Balance budget for maintaining on-premises servers and including cloud computing
- Institutional support falls short or reluctant to adopt cloud resources

Knowing 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/IWG member*

- My workflow doesn't appear to benefit from the advantages cloud computing provides.
- I have a workstation that meets my needs, no cloud needed.
- Is there a service that transforms data so you don't have to use additional compute?
- After transforming 4D datasets into 2D datasets, I haven't been able to save them and then download for local analysis.

Cloud Knowledge & 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)
- Working between cloud vendors is challenging
- Spinning up processing environments that aren't Python-centric

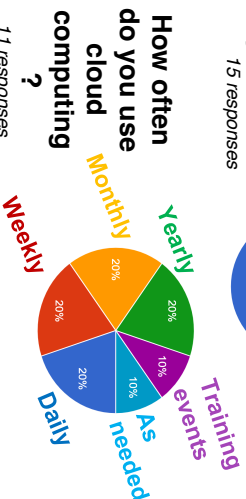
"Worried about doing things 'on my own', outside of a cloud-hosted JupyterHub. How do I utilize the 'spot market' effectively?" – *OpenScapes Champion*

- Transition out of supported, easy entry to cloud environment (i.e. Hub to your own AWS account)
- Finding available options for pre-determined environments
- I don't know how to get started on my own



Do you use cloud computing in your work?

15 responses



How often do you use cloud computing?

11 responses

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*

- Issues with cloud optimization (computing, data formats, etc.)
- I think that basic guidelines for data producers, and side-car files (e.g. Kerchunk) will be helpful.
- Only using cloud computing with cloud-native formats

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 (212c), as well as all of our participants.