Pangeo for training: Uses, Lessons Learned, and a User Perspective

Jessica Scheick, PhD
23 July 2020
ESIP Summer Meeting: “Pangeo Progress”
Pangeo is first and foremost a community. It also enables easy creation and management of computing environments. This flexibility can be leveraged across scales, institutions, and applications. Thus, Pangeo provides a critical resource that enables researchers (as individuals or groups) to leverage computing advances without needing to become experts in computational infrastructure and/or building and maintaining their own cluster.

We will review a recent custom deployment of a Pangeo hub for the Virtual ICESat-2 Cryospheric themed Hackweek that occurred in June 2020. We will share the value of using Pangeo in this training setting, as well as review some limitations and lessons learned. These successes and limitations will also be discussed in the context of using Pangeo in other research and training settings. Lastly, we will briefly review - from the user perspective - some exciting new developments, upcoming projects being worked on using Pangeo, and ways that the Pangeo team is expanding and diversifying engagement.
What is Pangeo?

“Pangeo is first and foremost a *community* promoting open, reproducible, and scalable science.”
Pangeo is built on Python’s scientific computing ecosystem.
What is a Hackweek?

A welcoming learning environment designed to build an open and collaborative research community while introducing participants to new software tools.
A typical hackweek includes...

- Community building activities
- Hands on tutorials
- Interactive peer-to-peer learning
- Project time to “hack” on something of interest to you
- Access to a team of experts in your field AND open-source software
- … and much more!
• June 2020
• First 100% virtual hackweek!
• Only ~2 months to transition to virtual
• A great success!
What about Pangeo?

The Cryospheric themed ICESat-2 Hackweek had its own custom Pangeo deployment!
Why use Pangeo for a Hackweek?

- Uniform computing environment
- Expanded computing resources
- Reduces need for individual troubleshooting (especially challenging virtually)
What are the challenges?

• Ephemeral
• Requires time to scale up resources
• Influence of novice users with shared compute resources (may affect performance)
Where else is Pangeo useful?

- Hackweeks
- Research groups
- Collaborations/projects
- Individuals
- Training centers
- Workshops
- Data and computation-intensive tasks
• A great, welcoming community committed to diversity!
• Establishing partnerships to increase diversity of open-source contributors (in progress)
• Creating spaces and resources/projects to help new contributors get on board
• Expanding gallery of examples
Exciting new developments (2 of 2)

or, why you should get involved from a user’s perspective!

- Fun, new features
  - VSCode within a Pangeo Hub
  - Kubernetes/Terraform now built in
- Exploring potential for contract-based offering of Pangeo as a service (along with formalization of the Hackweek learning model)
Resources, collaborators, funding, etc.

- [http://pangeo.io/](http://pangeo.io/)
- [https://icesat-2hackweek.github.io/learning-resources/](https://icesat-2hackweek.github.io/learning-resources/)
- NASA Project Number: 17-ACCESS17-0003
- Presenter: Jessica Scheick, [jbscheick@gmail.com](mailto:jbscheick@gmail.com), JessicaS11 (GitHub)