



CyberGIS-Compute: Geospatial Middleware for Simplifying Access to High Performance Computing

Alex Michels

**CyberGIS Center for Advanced Digital and Spatial Studies
Informatics Programs, University of Illinois Urbana-Champaign
Department of Geography and GIS, University of Illinois Urbana-Champaign**

*AAG Annual Meeting, 2023
March 25th, 2023*



Overview of Workshop

- Introduction
- Using the CyberGIS-Compute SDK (Hands-On)
- Making a CyberGIS-Compute Model (Hands-On)
- Wrap-Up/Next Steps
- Q&A



Motivation

- Geospatial discovery and innovation are increasingly computation and data intensive
- Personal computing environments are limited to resolve such computational intensity
- High-performance computing (HPC) environments are needed to enable computation- and data-intensive geospatial scientific workflows



But ...

- The learning curve to access and use HPC is very steep!

```
erichsiao — ssh -i cigi-gisolve@keeling.earth.illinois.edu — /Users/erichsiao/D...
Last login: Tue Sep 28 14:38:36 2021 from vpnpool-10-251-14-211.near.illinois.edu

 University of Illinois at Urbana-Champaign
College of Liberal Arts and Sciences
School of Earth, Society and Environment (SESE)

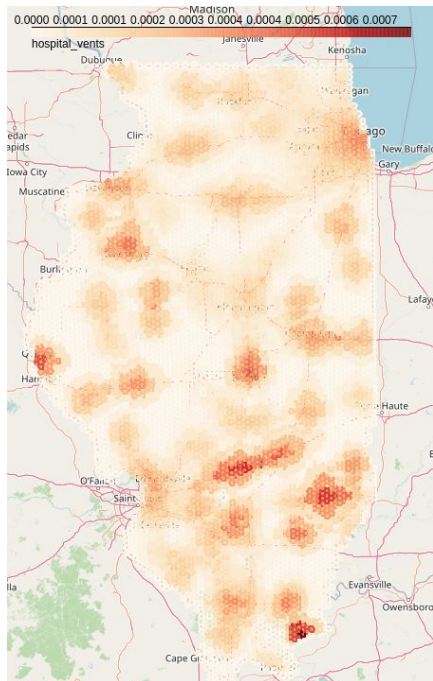
KEELING

This is the CentOS 7.9 deployment of the School of Earth,
Society, and Environment high-performance compute facility.

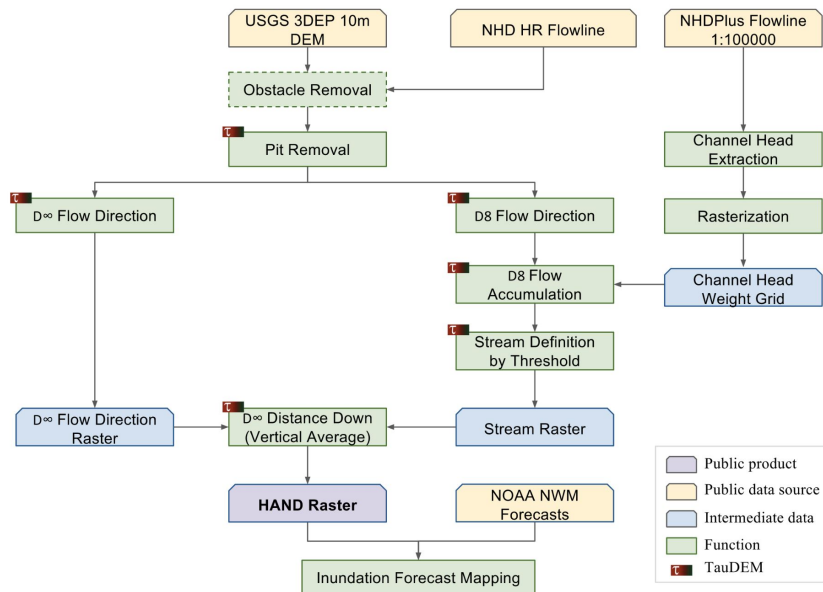
REMINDER:
  This is the login node for keeling, and it is shared among multiple
  users. Do not run parallel programs requiring more than seven compute
  threads on this machine; please use the batch system for such programs
  instead.
```



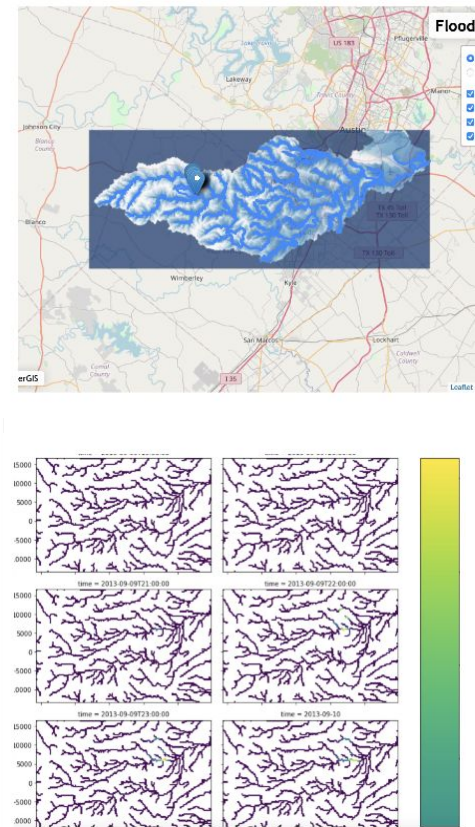
Examples



Spatial Accessibility Calculation



Estimate Height Above Nearest Drainage



WRFHydro Model



What is CyberGIS-Compute?

- Simplify access to HPC
- Bridge the gap between interactive computing environments (e.g. CyberGIS-Jupyter) and HPC
- Enable computation- and data-intensive geospatial workflows




User Interface

```
erichsiao — ssh -i cigi-gisolve@keeling.earth.illinois.edu — /Users/erichsiao/D...
Last login: Tue Sep 28 14:38:36 2021 from vpnpool-10-251-14-211.near.illinois.edu
```



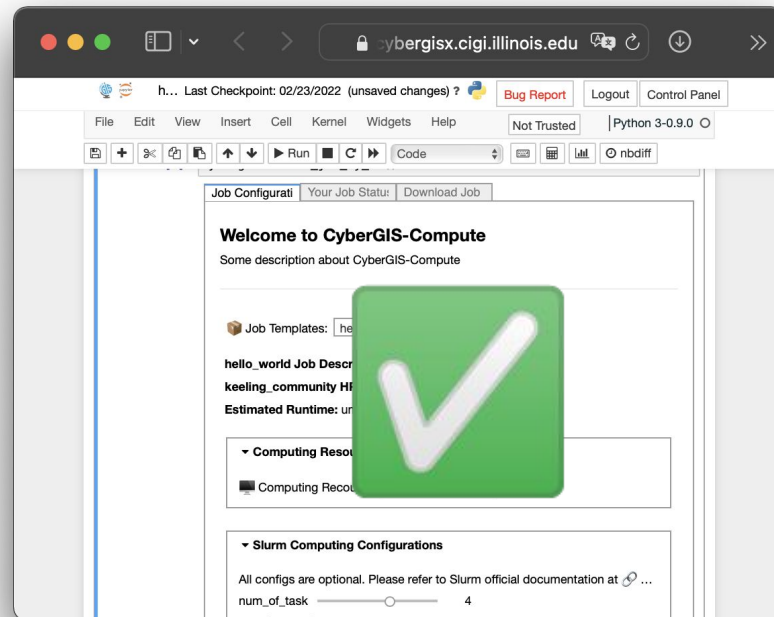
University of Illinois at Urbana-Champaign
College of Liberal Arts and Sciences
School of Earth, Society and Environment (SESE)

KEELING



This is the CentOS 7.9 deployment of the School of Earth, Society, and Environment high-performance computing facility.

REMINDER:
This is the login node for keeling, and it is shared among multiple users. Do not run parallel programs requiring more than seven compute threads on this machine; please use the batch system for such programs instead.





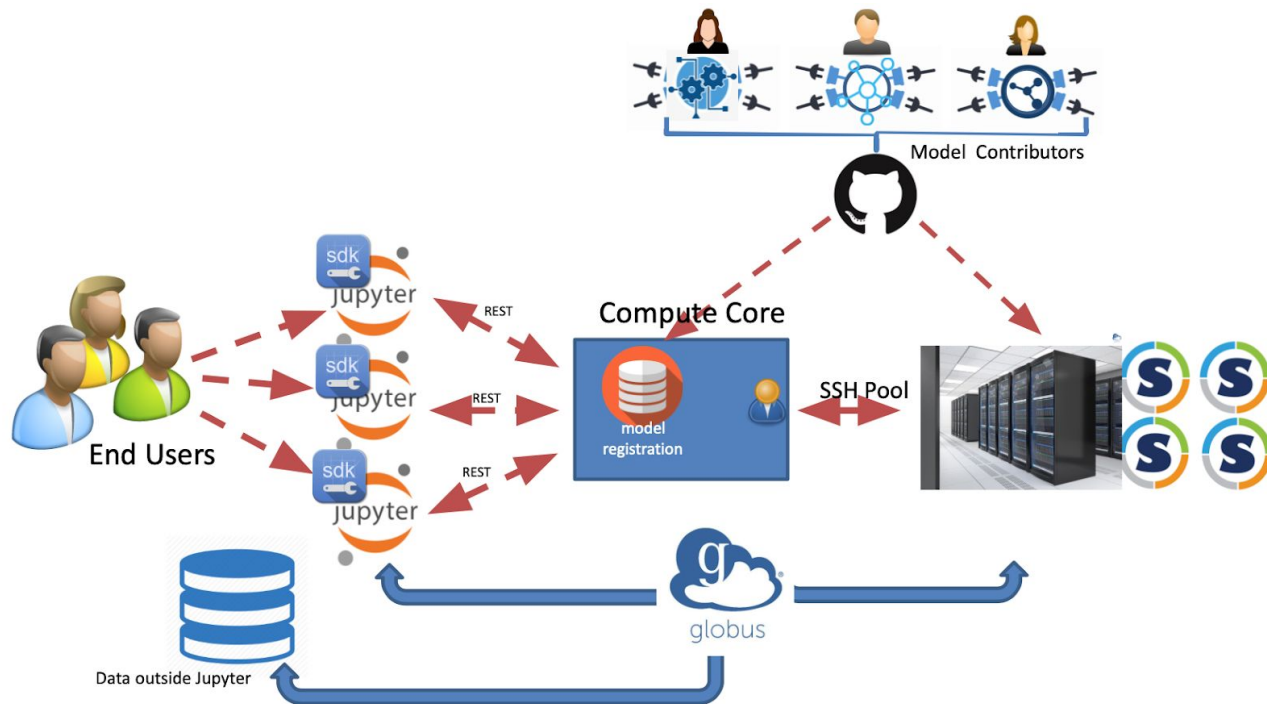
Bridging Ease of Use with Powerful Computing



- Padmanabhan, A., Vandewalle, R. C., Xiao, Z., Baig, F., Michels, A., Li, Z., and Wang, S. (2021) "CyberGIS-Compute for enabling computationally intensive geospatial research". In: *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Geospatial Data Access and Processing APIs*, <https://doi.org/10.1145/3486189.3490017>.
- Yin, D., Liu, Y., Hu, H., Terstriep, J., Hong, X., Padmanabhan, A., and Wang, S. (2018) "CyberGIS-Jupyter for Reproducible and Scalable Geospatial Analytics". *Concurrency and Computation: Practice and Experience*. <https://doi.org/10.1002/cpe.5040>



Geospatial Middleware



A **scalable middleware framework** for enabling high-performance and data-intensive geospatial research and education



Key Components

- **Core:** middleware server that automates job submission to HPC
- **SDK:** interactive client for Jupyter Notebook with code-less UI support
- **Contribution:** developer API that enables workflow contribution with little to no modification of existing code



CyberGIS-Compute
Core



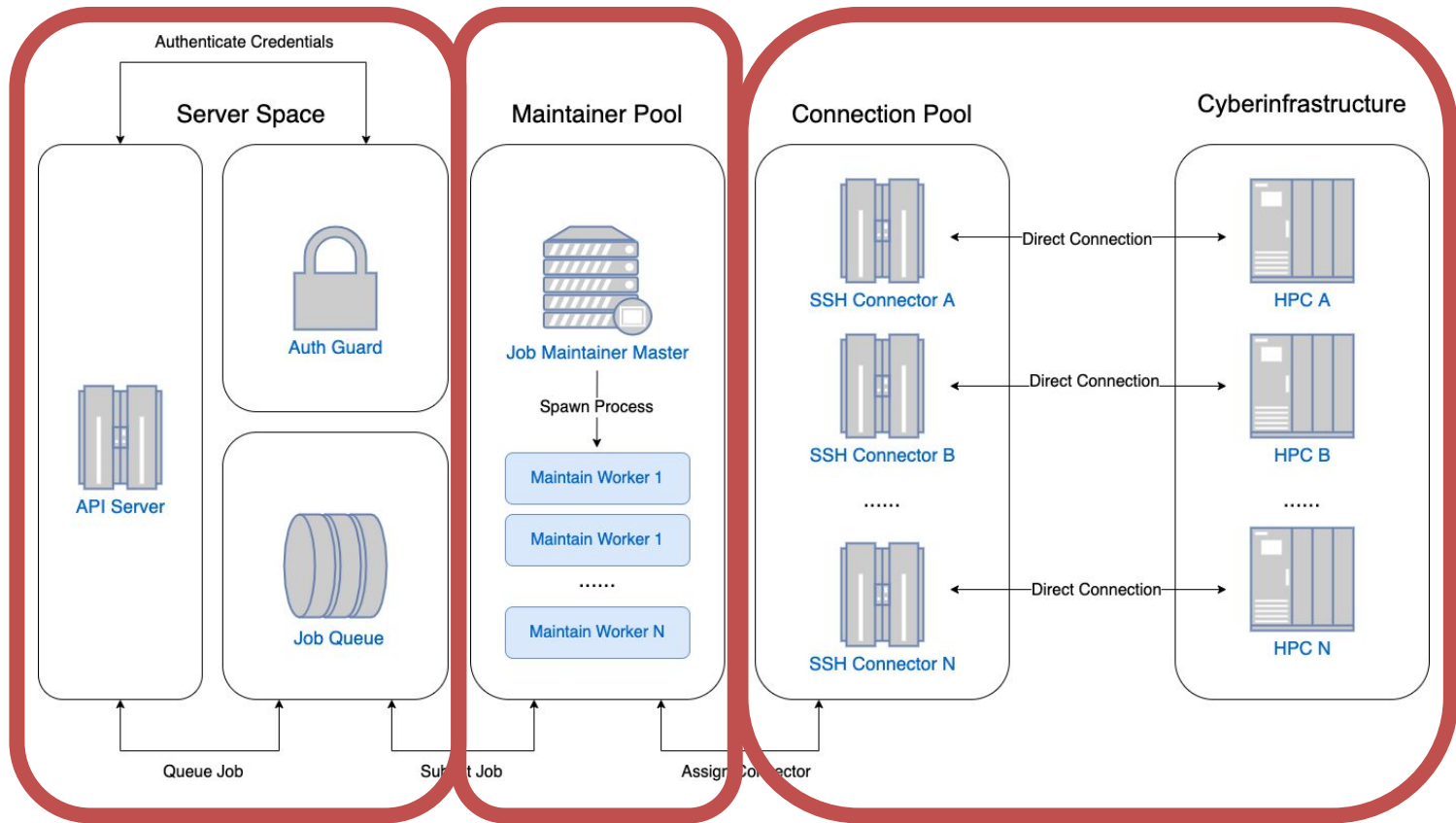
CyberGIS-Compute
SDK



Contribution

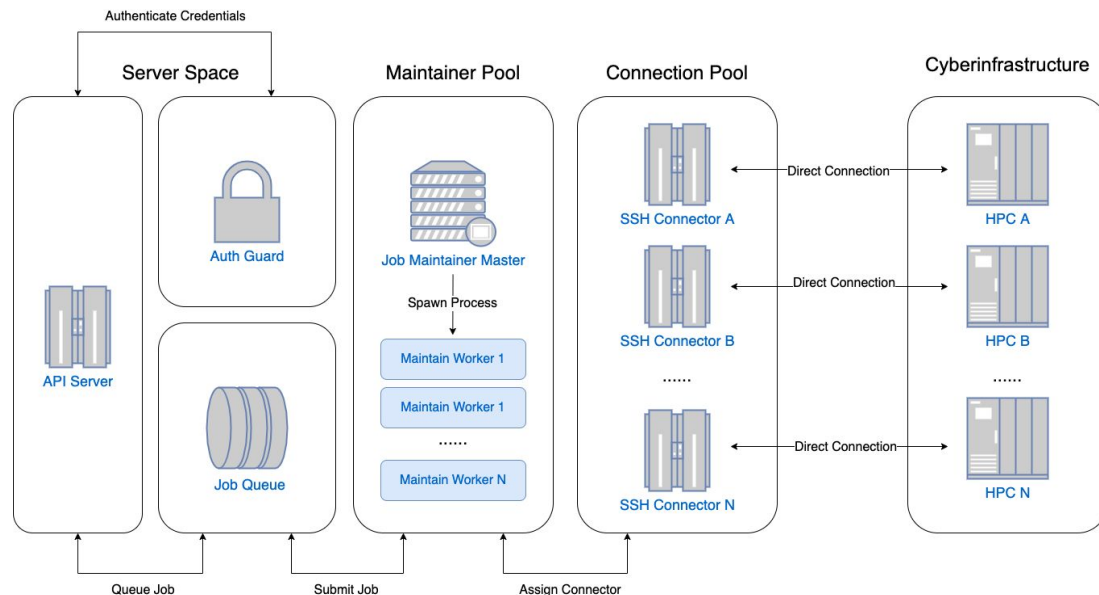


Architecture





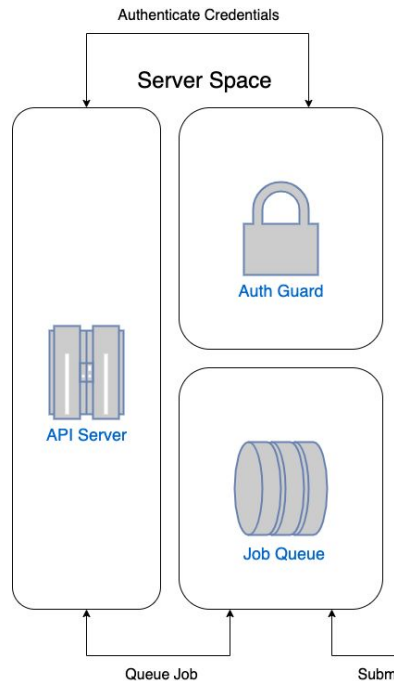
Middleware Server



-  Manages job submission to HPC
- Three layers
 - API Server Space
 - Maintainer Pool
 - Connection Pool



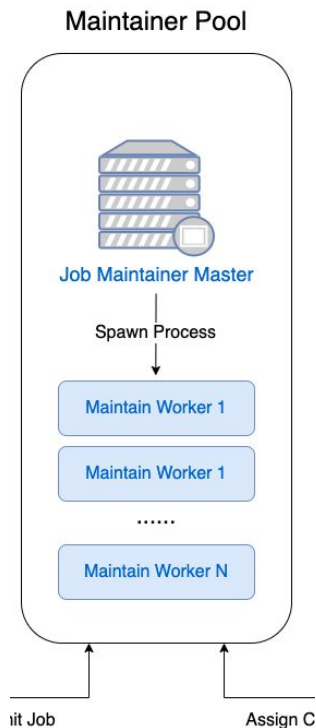
API Server



- Provides a **user-facing RESTful web interface** (API Server) for authentication and interaction with internal components of CyberGIS-Compute
- Pushes computation jobs into an internal job queue that the Maintainer Pool consumes



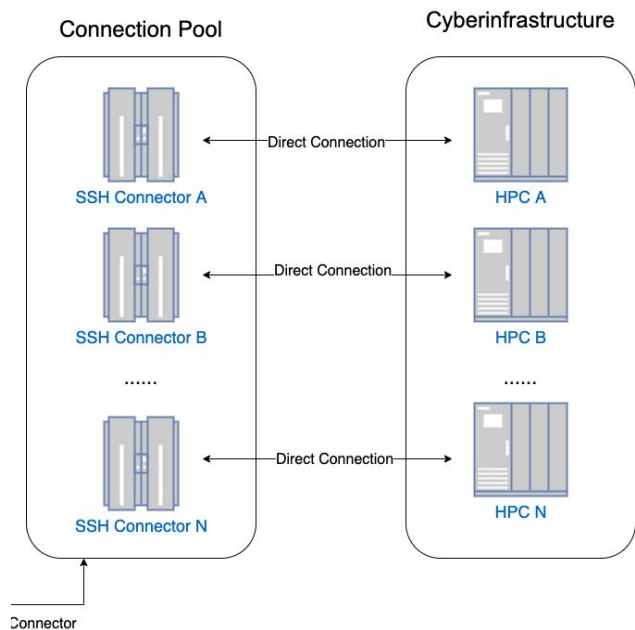
Maintainer Pool



- Maintainer Pool has a multithreaded life cycle that spawns and oversees worker processes called **Maintainer Workers** that contain **logic** on submitting, stopping, resuming, and ending a task executing on a remote HPC resource
- A Maintainer Worker can use a **Connector** to interact with remote HPC via SSH



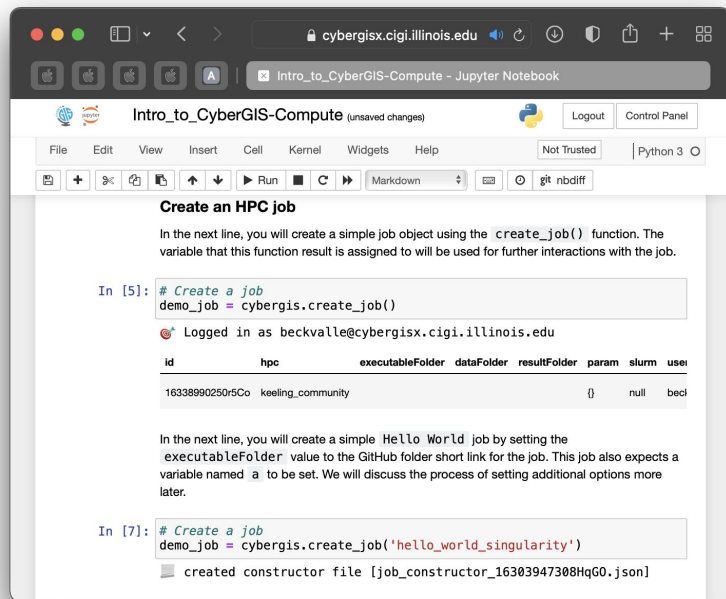
Connectors



- A Connector is a **long-living SSH connection** between the CyberGIS-compute core and HPC
- To reduce the connection rate, connectors are shared between Maintainer Workers and a **Mutex** is implemented to avoid race conditions in executing commands



SDK: Client Package



- A Python-based Jupyter Notebook client that integrates CyberGIS-Compute Core functionalities into CyberGIS-Jupyter
- Provides **seamless** interaction with HPC
- Provides **code-less** interactive UI

<https://cybergis.github.io/cybergis-compute-python-sdk/index.html>



Hands-On Part 1: Using the SDK

Please open this link:

<https://github.com/cybergis/AAG2023-ComputeTutorial>

Navigate to CyberGISX
(create an account if you
don't have one)

Open the repo on
CyberGISX

CyberGIS-Compute: Geospatial Middleware for High-Performance Computing

The CyberGIS-Compute workshop is a part of the [AAG 2023 Symposium on Harnessing the Geospatial Data Revolution for Sustainability Solutions](#). Check out the link for more information and instructions to register.

Overview of the Workshop

- Introduction to CyberGIS-Compute (Slides)
- Using the CyberGIS-Compute SDK (Hands-On) - [CyberGISX](#), [This repo on CyberGISX](#)
- CyberGIS-Compute Three Examples (Hands-On) - [Git Repo](#)
- Wrap-Up/Next Steps
- Q&A

Additional Resources

- [CyberGIS-Compute Documentation](#)



Seamless Access to HPC

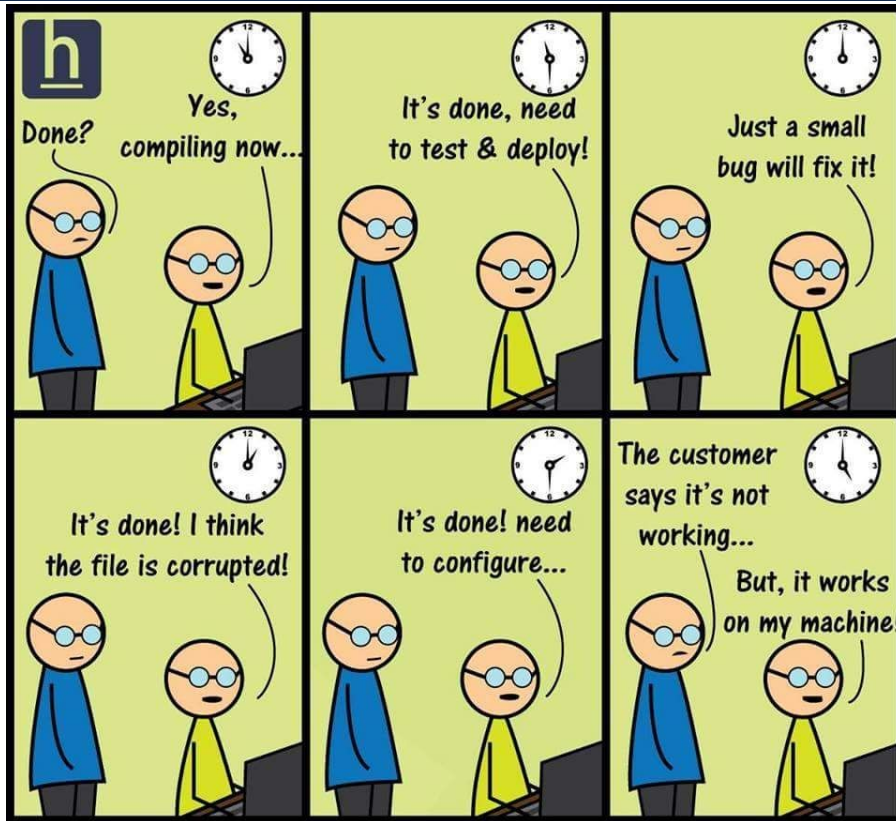


Containerization: Run code in familiar environments



Transparently interfaces with batch systems (e.g. Slurm):

Manage Slurm on behalf of developers



Why We use Containers



CyberGIS-Compute Contribute

CyberGIS-Compute Contribute allows users to **submit workflow code hosted on GitHub repositories** to be executed on HPC resources

Submissions are verified through a checking process

Provides configurations, system environment, and developer API

Git commit version lock for security



Contribution Process

```
1 import json
2 import os
3
4 print('running main...\n')
5
6 print('./job.json\n')
7 job = json.load(open('./job.json',))
8
9 print('SLURM_NODEID\n')
10 print(os.environ['SLURM_NODEID'])
11
12 print('SLURM_PROCID\n')
13 print(os.environ['SLURM_PROCID'])
14
15 print('job_id')
16 print(os.environ['job_id'])
```

+



manifest.json

=



ready for HPC



Hands-On Part 2: Creating a Model

Please open this link:

<https://github.com/cybergis/AAG2023-ComputeTutorial>

Go to the “cybergis-compute-examples” repo and follow the directions on the README.md

CyberGIS-Compute: Geospatial Middleware for High-Performance Computing

The CyberGIS-Compute workshop is a part of the [AAG 2023 Symposium on Harnessing the Geospatial Data Revolution for Sustainability Solutions](#). Check out the link for more information and instructions to register.

Overview of the Workshop

- Introduction to CyberGIS-Compute (Slides)
- Using the CyberGIS-Compute SDK (Hands-On) - [CyberGISX](#), This repo on [CyberGISX](#)
- CyberGIS-Compute Three Examples (Hands-On) - [Git Repo](#)
- Wrap-Up/Next Steps
- Q&A

Additional Resources

- [CyberGIS-Compute Documentation](#)



Next Steps for Contributing a Model

- Work with someone at the CyberGIS-Center (or other organization hosting a Core server) to have your model added.
- Test and debug your model just to make sure everything works.
- You're done and everyone else can now easily run your model on HPC!



Documentation

As you develop models, be sure to check out the CyberGIS-Compute documentation!

 CyberGIS-Compute Python SDK

Search

 cybergis-compute-python-...
4 Stars · 5 Forks

democratizing HPC for geospatial developers

CyberGIS-Compute Python SDK v0.2.4 documentation

CyberGIS-Compute Python SDK

- About
- Using CyberGIS-Compute
- Hello World Notebook
- Model Gallery
- Release Notes
- Related Projects
- Getting Help
- Contributing Models
- Developer Documentation
- Contributing to CyberGIS-Compute

Welcome to CyberGIS-Compute Python SDK's documentation!

Welcome to the documentation for CyberGIS-Compute! CyberGIS-Compute is designed to increase accessibility to super-computing resources by lowering the technical barriers, with a focus on geospatial problem solving. Our goal is to provide a simple and easy-to-use interface for users and model developers alike! This is important to ensure that everyone can use and benefit from CyberGIS-Compute regardless of their technical skills. We hope that by lowering the technical barriers, we can empower domain experts, policy makers, students, and everyone in between to harness the power of today's advanced cyberinfrastructure!

Why use CyberGIS-Compute?

Have you ever tried to run code or some analysis, but your laptop didn't have enough storage/memory/it was too slow on your computer? That's a pretty common problem, especially with the growth of big data and AI in geospatial problem solving and we are here with the solution!

There are a variety of High-Performance Computing (HPC) resources which can handle the analysis/model/code you're trying to run, although there are usually some technical hurdles to overcome to use them. We are here to make things a bit easier! All you need to do is make a few small changes to your code (like put the results in the results folder), add some metadata to your code, put it in Github, and then you can run the code on HPC from a Jupyter notebook/lab interface! This also means that distributing your code is much easier: anyone can run the job themselves without needing direct access to the HPC resources! For an example of the CyberGIS-Compute interface, see below:

Contents

- Welcome to CyberGIS-Compute Python SDK's documentation!
- Why use CyberGIS-Compute?
- Demo Hello World
- Table of Contents
- Indices and tables
- Show Source



Contributing Models Docs

Contributing Models

⚠ Attention

These guides are still a work in progress. Please let us know if you spot any issues or errors!

CyberGIS-Compute is built on *models*. Models in this context are simply code that performs some action. Models can do the thing every time (i.e. our Hello World model which just prints "Hello World"), but CyberGIS-Compute gives model developers a users the ability to customize and experiment with models through uploading data and passing parameters. This allows mod developers to build more flexible and reusable models and makes models much more interesting and valuable to those using models.

Model Contribution Guides

[Developing a Model with CyberGIS-Compute](#)

[What are the stages of your model?](#)

[Notebook to Script/Nbconvert](#)

[Getting Results from Compute](#)

[Crash Course in the Manifest](#)

[Crash Course in SLURM Parameters](#)

[Supported HPC](#)

[Providing Input Data](#)

[Installing Packages](#)

[Contributing Your Model](#)

[Advanced Topic: Manifest Options](#)

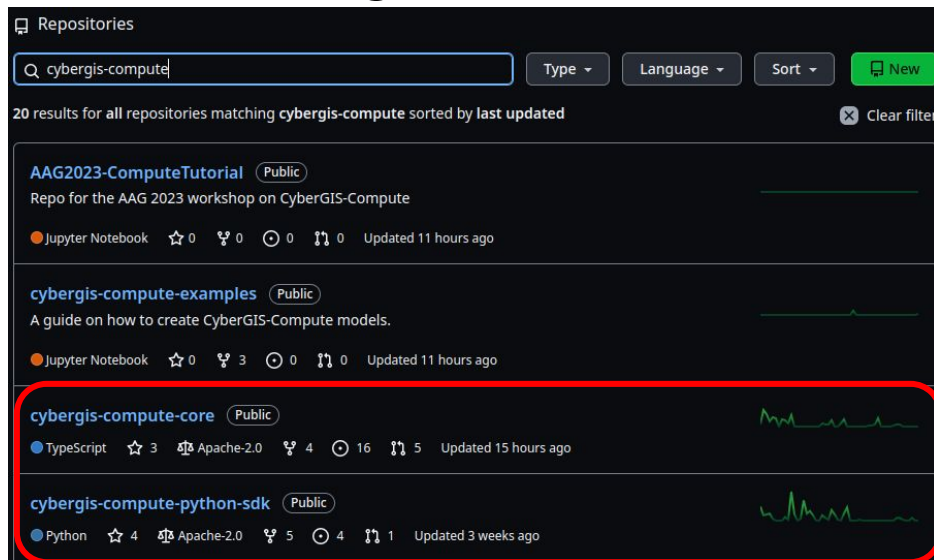
We have documentation on our site to guide you through creating a model!

It also details advanced options and answers FAQs!



Developers Welcome!

Both the Core and SDK are open-sourced and available on Github for anyone interested in contributing to the codebase!





Works in Progress

- **Eliminating containers.** You just use CyberGIS-Jupyter kernels!
- **Simplified UI.** A simplified UI for model users to avoid overwhelming them.

We hope to have these changes rolled out over the next year.



CyberGIS-Compute Team

- Mit Kotak
- Alexander Michels
- Anand Padmanabhan
- John Speaks
- Shaowen Wang
- Zimo Xiao
- Taylor Ziegler



Links



Project Development doc:

<http://github.com/cybergis/cybergis-compute-core>



SDK doc:

<http://github.com/cybergis/cybergis-compute-python-sdk>



Hello World doc:

<http://github.com/cybergis/cybergis-compute-hello-world>



CyberGIS Center for Advanced Digital and Spatial Studies

[Overview](#) ▾[Stories & Outreach](#) ▾[Initiatives & Resources](#) ▾[Contact Us](#)

Summer School 2023

[Summer School 2023](#)[Apply](#)[Logistics](#)[Sponsors](#)

Convergence Science in Action

UCAR campus in Boulder, Colorado
August 7-11, 2023

The Institute for Geospatial Understanding through an Integrated Discovery Environment, I-GUIDE, will lead a week-long Summer School in August 2023 on *Convergence Science in Action*. Certain complex and compelling societal problems require a convergent research approach, when knowledge, tools, and modes of thinking from multiple disciplines are strategically integrated and merged. About 25 graduate students and early career scholars will collaborate with project members of I-GUIDE to develop novel solutions to complex problems that rely on computation- or data-intensive geospatial data science. The participants will experience the collaborative and professional interactions that are key to comprehensively working on convergence science problems, including geoethics, geo-enabling reproducible and open science, geovisualization, and geoAI and spatial data science via cloud and high-performance computing.

This year our application areas will include topics such as climate change, biodiversity, water security, sustainable development, and implications of these as studied via social science data.

The Summer School will be hosted on the UCAR campus in Boulder, Colorado, from Monday August 7 to Friday, August 11, 2023. If you are a graduate student or early-career scholar new to geospatial data science and want to learn more about integrating this into your research, or are already working with data-intensive geospatial science approaches, this Summer School will offer new and exciting opportunities for your professional development, and will help you develop interdisciplinary skills and build new connections with others in related fields.

Learn more:

<https://iguide.illinois.edu/summer-school-2023/>

Acknowledgments

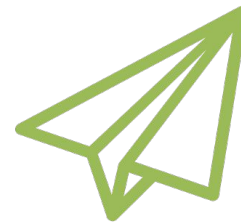
- National Science Foundation
 - 2118329
 - XSEDE
- Virtual ROGER



Thanks !



Comments / Questions?



Email:

michels9@illinois.edu