



# **CyberGIS-Compute: Simplifying Access to High Performance Computing for your Geospatial Computation**

**Furqan Baig and Alexander Michels**

CyberGIS Center for Advanced Digital and Spatial Studies  
Informatics Program  
Department of Geography and Geospatial Information Science  
University of Illinois at Urbana-Champaign

*ACES Workshop 2023  
July 15th, 2023*

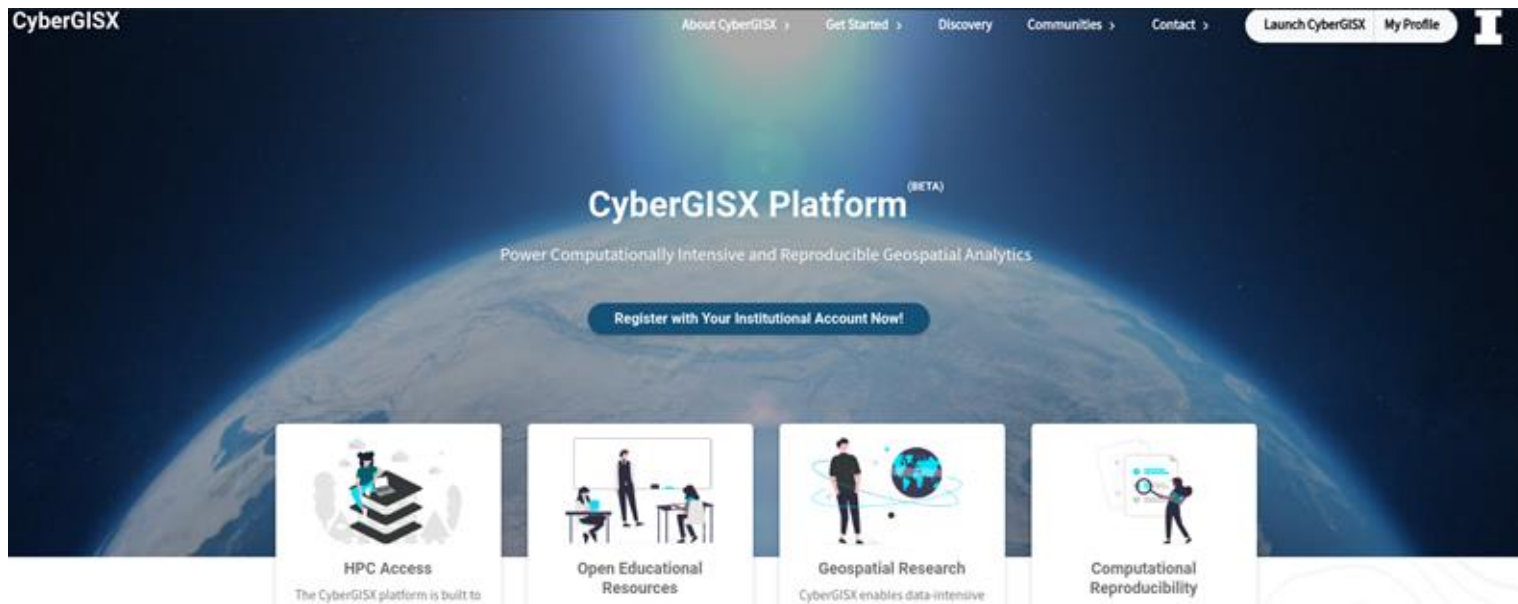


Award #: 2118329



# Hands-On: CyberGISX

Please sign up for a CyberGISX account if you have not already done so: <https://cybergisxhub.cigi.illinois.edu/>



# 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

  I

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

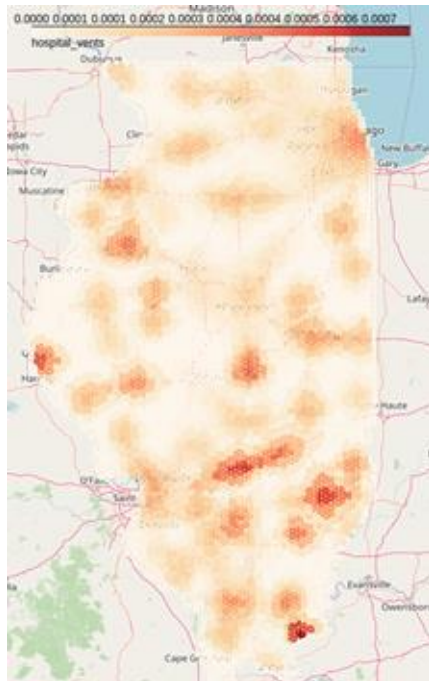
KEELING

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

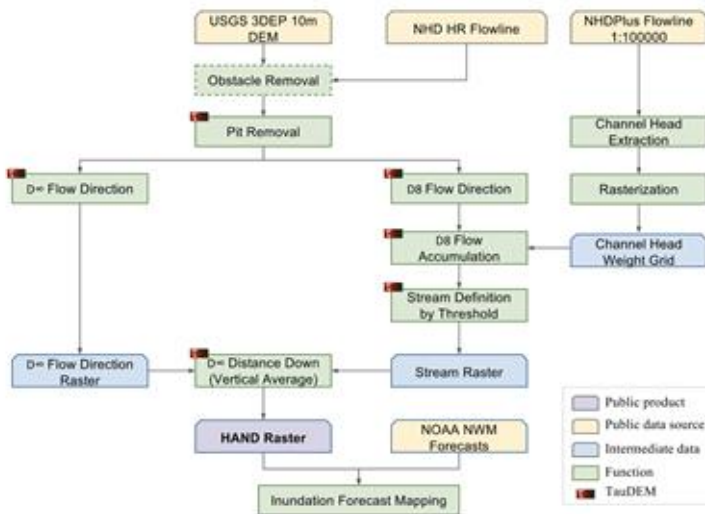
REMINDER:
  This is the login node for keeling, and it is shared among multiple
  users. Do not run parallel programs requiring more than seven compute
  threads on this machine; please use the batch system for such programs
  instead.
```



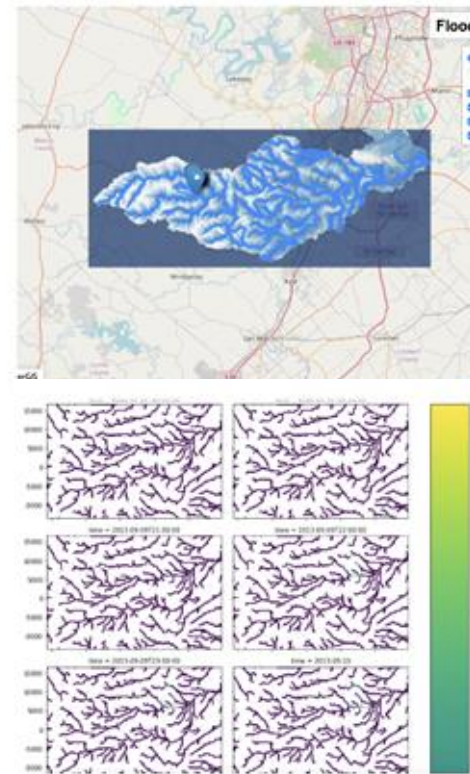
## Examples



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
```



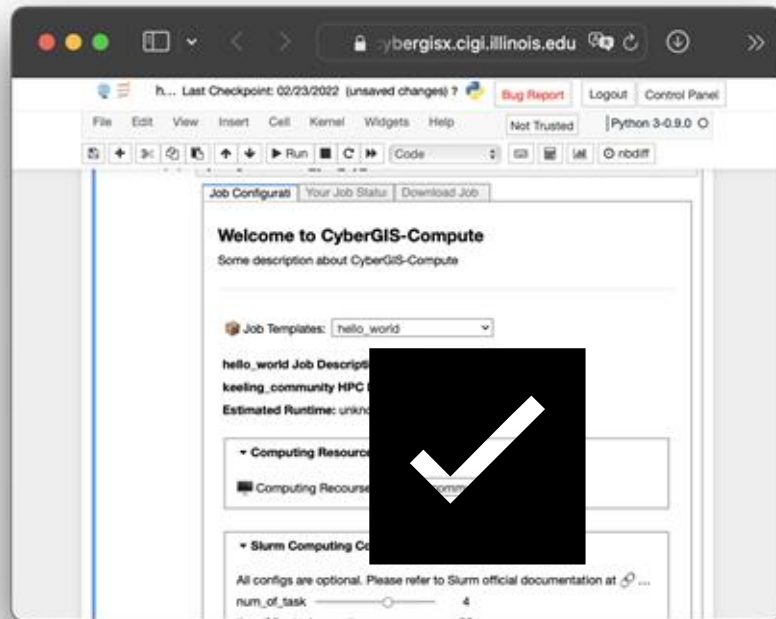
University of Illinois at Urbana-Champaign  
College of Liberal Arts and Sciences  
School of Earth, Society and Environment (SESE)



KEELING

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



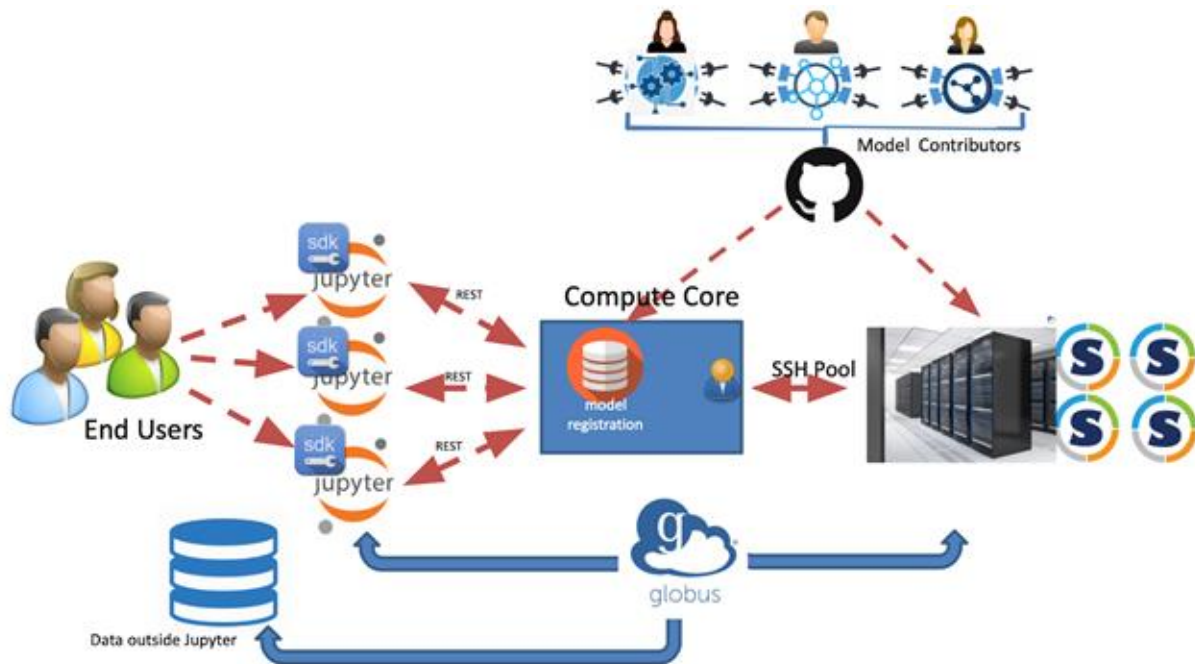
# Bridging Ease of Use with Powerful Computing



- 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



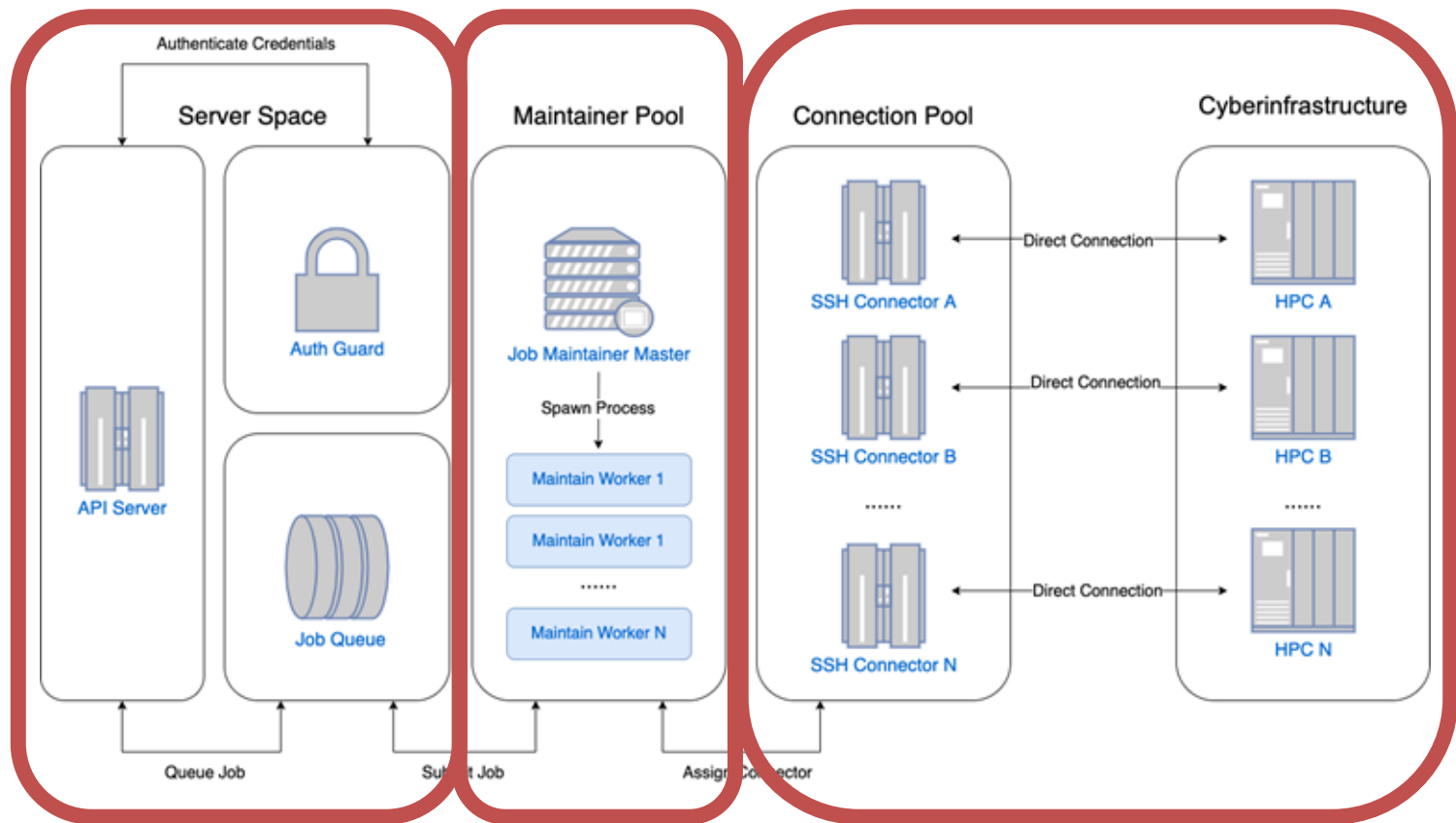
CyberGIS-Compute  
SDK



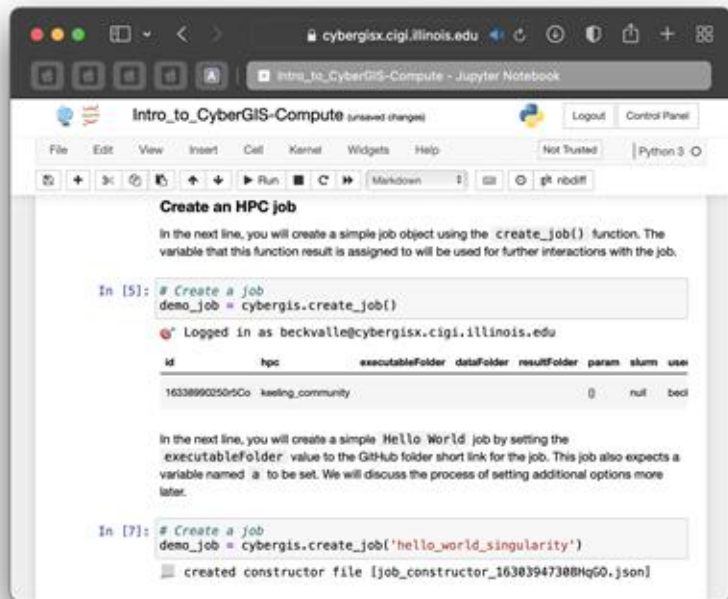
Contribution



## CyberGIS-Compute Core Middleware



# SDK: Client Package



cybergisx.cigi.illinois.edu

Intro\_to\_CyberGIS-Compute - Jupyter Notebook

Intro\_to\_CyberGIS-Compute (unsaved changes)

File Edit View Insert Cell Kernel Widgets Help Not Trusted Python 3

### Create an HPC job

In the next line, you will create a simple job object using the `create_job()` function. The variable that this function result is assigned to will be used for further interactions with the job.

```
In [5]: # Create a job
demo_job = cybergis.create_job()
```

Logged in as beckvalle@cybergisx.cigi.illinois.edu

id	hpc	executableFolder	dataFolder	resultFolder	param	shum	user
163389902505Co	keeling_community					null	bed

In the next line, you will create a simple 'Hello World' job by setting the `executableFolder` value to the GitHub folder short link for the job. This job also expects a variable named `a` to be set. We will discuss the process of setting additional options more later.

```
In [7]: # Create a job
demo_job = cybergis.create_job('hello_world_singularity')
```

created constructor file [job\_constructor\_16383947388HgG0.json]

- 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>

# Large Datasets

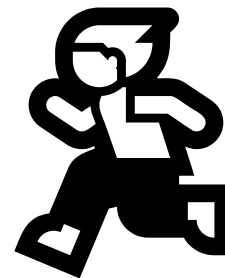


Big Data

+



=



Fast & Reliable Transfer



# Seamless Access to HPC



**Containerization:** Run code in familiar environments



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

Manage Slurm on behalf of developers



## 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 Demo !



**Comments / Questions?**

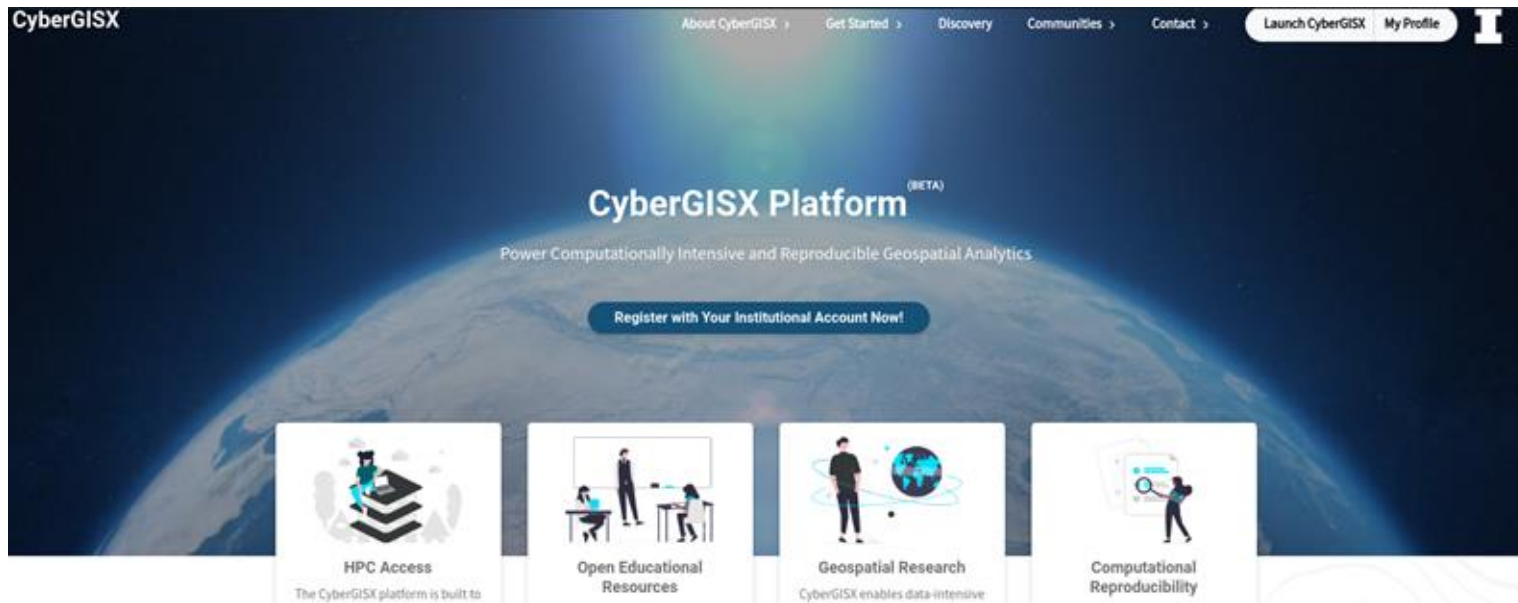


**Email:**

[fbaig@illinois.edu](mailto:fbaig@illinois.edu)  
[michels9@illinois.edu](mailto:michels9@illinois.edu)

# Hands-On: CyberGISX

Please sign up for a CyberGISX account if you have not already done so: <https://cybergisxhub.cigi.illinois.edu/>





# Hands-On Outline

0. Getting everyone to <https://cybergisx.cigi.illinois.edu>
  1. Very simple “hello world” job
  2. More involved spatial accessibility application
  3. Overview of how you can create your own models!



# Future Work and Documentation



# CyberGIS-Compute on ACES

We are currently working with ACES to get CyberGIS-Compute connected to ACES.

This will allow model developers to take advantage of the cutting-edge resources on ACES.

# Simplified UI

We've received feedback that while the verbose UI is appreciated by model developers, it can confuse less technical users.

Our goal is to create a more streamlined and end-user oriented UI in addition to the existing one.

```
(1) from cybergis_compute_client import CyberGISCompute
cybergis = CyberGISCompute(url="cgsjobshop.cgl.illinois.edu", isJupyter=True, protocol="HTTPS", port=443, suffix="v2")
(2) cybergis.show_all()
```

The screenshot displays the CyberGIS-Compute web interface. At the top, there are tabs for 'Job Configuration', 'Your Job Status', 'Download Job Result', and 'Your Jobs'. The 'Job Configuration' tab is active. The main content area is titled 'Welcome to CyberGIS-Compute' and includes a subtitle: 'A scalable middleware framework for enabling high-performance and data-intensive geospatial research and education on CyberGIS-Jupyter'. Below this, there is a 'Job Template' dropdown menu set to 'hello\_world'. Underneath, the job details are listed: 'hello\_world Job Description: none', 'kneeling\_community\_HPC Description: none', and 'Estimated Runtime: unknown'. There are three expandable sections: 'Computing Resource', 'Slurm Computing Configurations', and 'Input Parameters'. At the bottom, there is a checkbox for 'receive email on job status?' with an email input field containing 'example@illinois.edu'. A 'Submit Job' button is located at the bottom right.



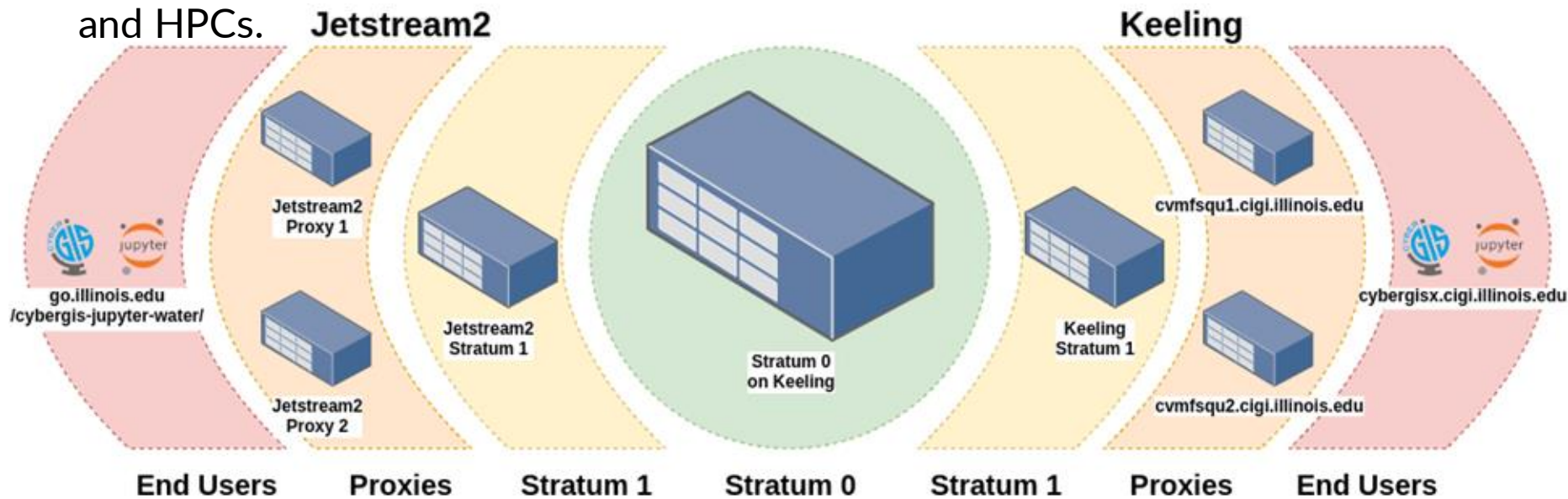
## Private Repos

Our codebase currently assumes that model repositories are public, but many researchers do not want to make their code public before publication acceptance.

We are testing a system to allow private Github repositories to run on CyberGIS-Compute.

# CVMFS Integration

Our JupyterHubs (CyberGISX, CJW) currently depend on a software stack in CVMFS and we are working on a way to provide this environment on HPCs. This would provide a consistent software environment between JupyterHubs and HPCs.





# Cloud Backends

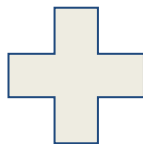
Motivation: Public datasets available on cloud storage (AWS Open Data, GCP Public Datasets etc.)

Move **Computation** to **Data**

Write code - Specify resources - run on cloud



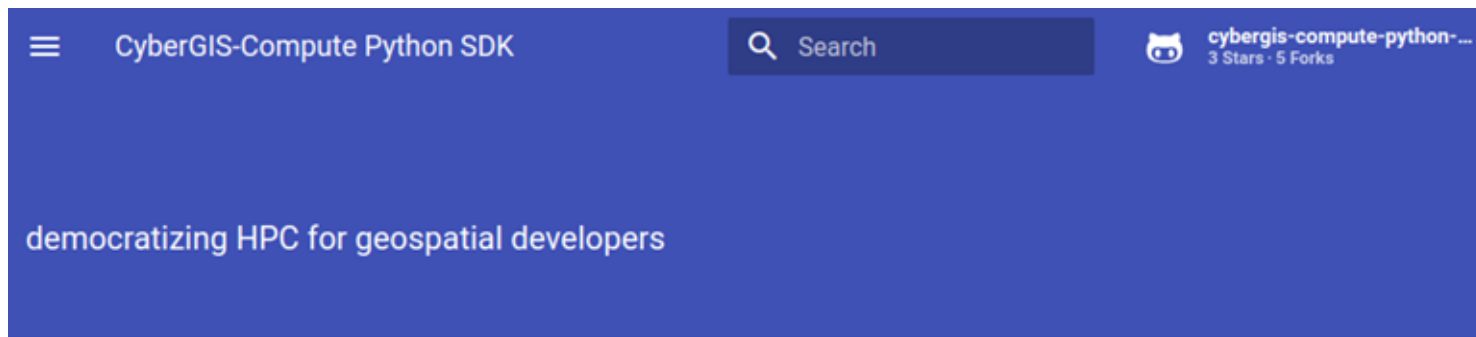
CyberGIS-Compute





## Documentation

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



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

### Contents

Welcome to CyberGIS-Compute Python SDK's documentation!

Why use CyberGIS-Compute?

Demo Hello World

Table of Contents

Indices and tables

# Model Gallery

We have a growing gallery of models that have utilized CyberGIS-Compute on our website.

## Remote Sensing Image Fusion Model

Remote sensing data fusion (data\_fusion) has been extensively studied to generate high spatial- and temporal- resolution data products by integrating high-spatial-low-temporal resolution (such as Landsat) and high-temporal-low-spatial-resolution images (such as MODIS). In this model, we utilize a robust hybrid deep learning model based on a super-resolution convolutional neural network (SRCNN) and long short-term memory (LSTM) for spatiotemporal data fusion to integrate Landsat images with MODIS. The work uses code from the following paper:



Input MODIS



Predicted Result

Fangzheng Lyu, Zijun Yang, Zimo Xiao, Chunyuan Diao, Jinwoo Park, and Shaowen Wang. 2022. CyberGIS for Scalable Remote Sensing Data Fusion. In Practice and Experience in Advanced Research Computing (PEARC '22). Association for Computing Machinery, New York, NY, USA, Article 35, 1-4. <https://doi.org/10.1145/3491418.3535145>

- Github Repository: [cybergis/data\\_fusion](https://github.com/cybergis/data_fusion)
- DOI for Paper: [10.1145/3491418.3535145](https://doi.org/10.1145/3491418.3535145)
- Related Publication: [A Robust Hybrid Deep Learning Model for Spatiotemporal Image Fusion](#)

## SUMMA

The Structure for Unifying Multiple Modeling Alternatives (SUMMA) hydrological modeling framework can be run with CyberGIS-Compute:

- Github Repository: [cybergis/cybergis-compute-v2-summa](https://github.com/cybergis/cybergis-compute-v2-summa)
- Hydroshare Resource: [Implementation of SUMMA model using CyberGIS-Compute V2](#)
- Notebook on CyberGISXHub: [Run ensemble SUMMA 3.0 model on HPC with CyberGIS-Compute Service on CJW](#)



# CyberGIS Center for Advanced Digital and Spatial Studies

The screenshot shows the GitHub repository page for **cybergis-compute-core**. The repository is public and has 5 stars, 4 forks, and 3 watchers. It is currently on version **v2**. The repository has a README, an Apache-2.0 license, and 5 stars. The repository is being watched by 3 people and has 4 forks. The repository is reported by 1 person. The repository is a Python project. The repository is a core component of the CyberGIS Center for Advanced Digital and Spatial Studies. The repository is a core component of the CyberGIS Center for Advanced Digital and Spatial Studies. The repository is a core component of the CyberGIS Center for Advanced Digital and Spatial Studies.

File	Commit Message	Time
.github/workflows	Update build-core.yml	8 months ago
archive	housekeeping	last year
configs	Merge pull request #64 from cyber...	3 months ago
data	major update	2 years ago
docker	Merging refactor_file_system into v2 (#48)	last year
docs	added singcvms note	3 months ago
examples/helloworld	developer notes	3 years ago

The screenshot shows the GitHub repository page for **cybergis-compute-python-sdk**. The repository is public and has 3 stars, 5 forks, and 2 watchers. It is currently on version **v2**. The repository has a README, an Apache-2.0 license, and 3 stars. The repository is being watched by 2 people and has 5 forks. The repository is reported by 1 person. The repository is a Python project. The repository is a SDK component of the CyberGIS Center for Advanced Digital and Spatial Studies. The repository is a SDK component of the CyberGIS Center for Advanced Digital and Spatial Studies. The repository is a SDK component of the CyberGIS Center for Advanced Digital and Spatial Studies.

File	Commit Message	Time
.github	Minor Fixes to Site and Docstrings (#59)	9 months ago
cybergis_compute_...	Updated tests and fixed bug in Zip.py (#63)	5 months ago
docs	releasing v0.2.4 (#64)	5 months ago
test	Updated tests and fixed bug in Zip.py (#63)	5 months ago
.gitignore	working on the docs and updated hello w...	last year
LICENSE	update license	2 years ago
README.md	updated README.md, fixed license in set...	last year

## Core and SDK on Github



## 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>