I-GUIDE

Institute for
Geospatial Understanding
through an Integrative
Discovery Environment

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

Anand Padmanabhan and Alexander Michels

University of Illinois Urbana-Champaign

I-GUIDE Virtual Consulting Office

Award #: 2118329

Date July 27, 2022

CyberGIS-Compute Team

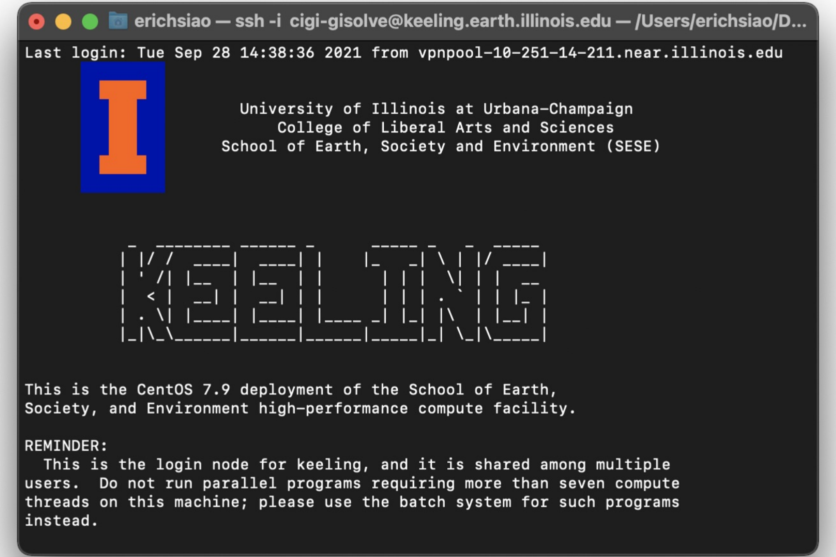
- Mit Kotak
- Zhiyu Li
- Alexander Michels
- Anand Padmanabhan
- Shaowen Wang
- Zimo Xiao
- Taylor Ziegler

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 …

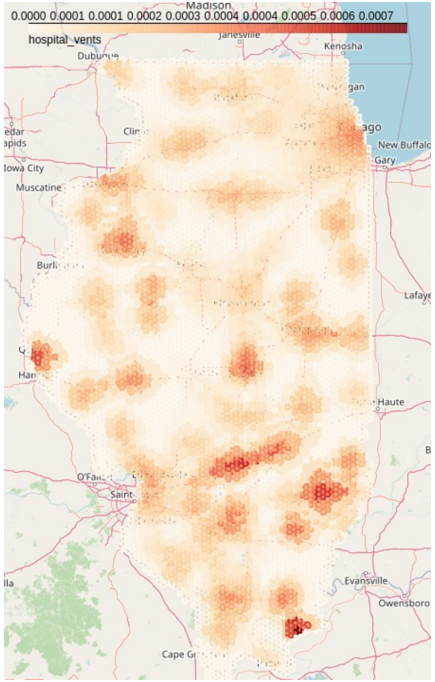
- The learning curve to access and use HPC is very steep!

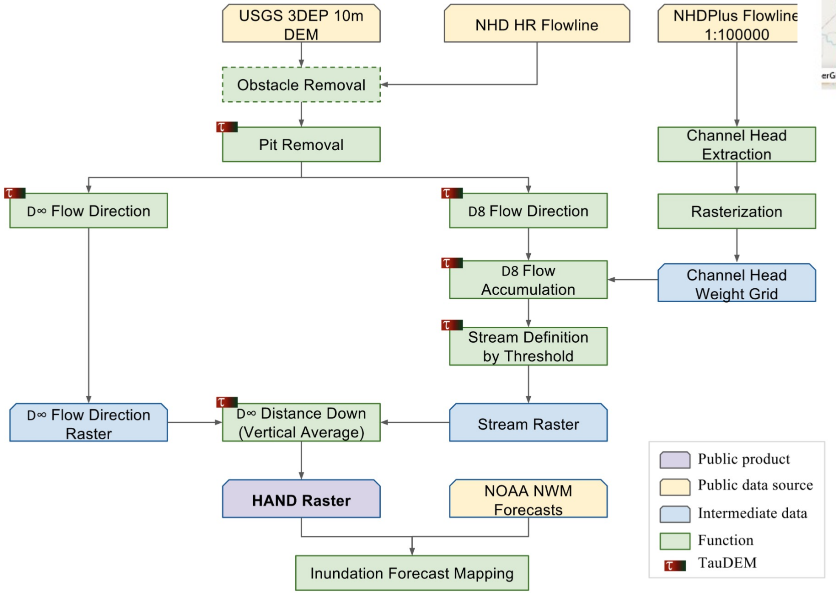


erichsiao — ssh -i cigi-gisolve@keeling.earth.illinois.edu — /Users/erichsiao/D...

```
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)



          _  ________ ______ _      _____ _   _  _____
         | |/ /  ____|  ____| |    |_   _| \ | |/ ____|
         | ' /| |__  | |__  | |      | | |  \| | |  __
         |  < |  __| |  __| | |      | | | . ` | | |_ |
         | . \| |____| |____| |____ _| |_| |\  | |__| |
         |_|\_\______|______|______|_____|_| \_|\_____|


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





USGS 3DEP 10m DEM

NHD HR Flowline

NHDPlus Flowline 1:100000

Obstacle Removal

Pit Removal

Channel Head Extraction

D∞ Flow Direction

D8 Flow Direction

Rasterization

D8 Flow Accumulation

Channel Head Weight Grid

Stream Definition by Threshold

D∞ Flow Direction Raster

D∞ Distance Down (Vertical Average)

Stream Raster

HAND Raster

NOAA NWM Forecasts

Inundation Forecast Mapping

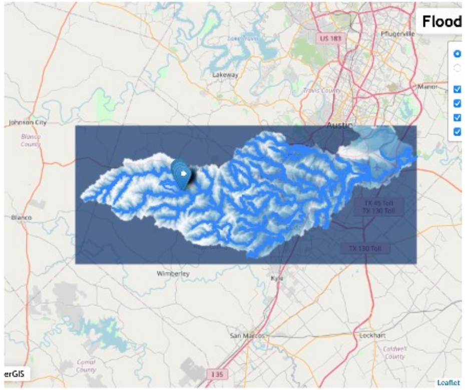
Public product

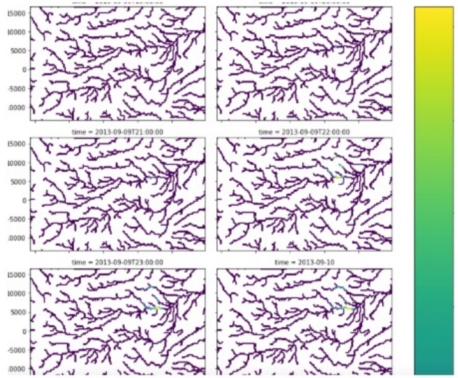
Public data source

Intermediate data

Function

τ TauDEM



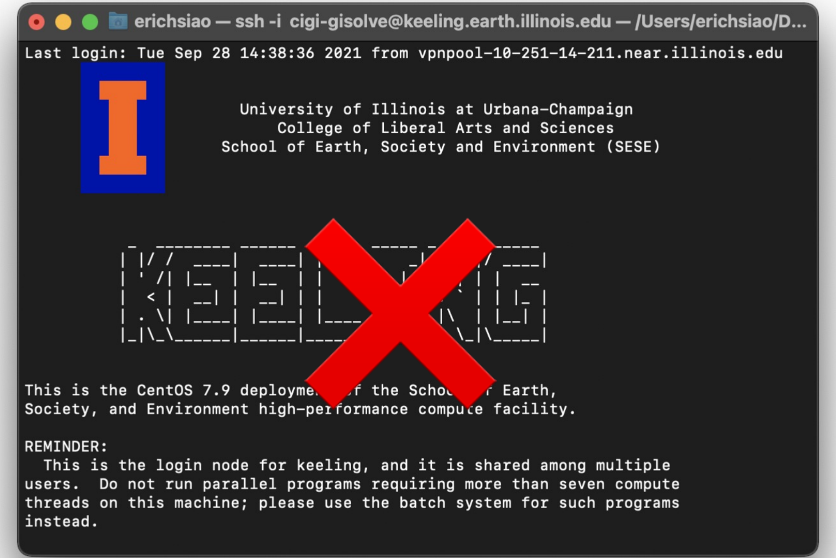


Spatial Accessibility Calculation   Estimate Height Above Nearest Drainage   WRFHydro Model

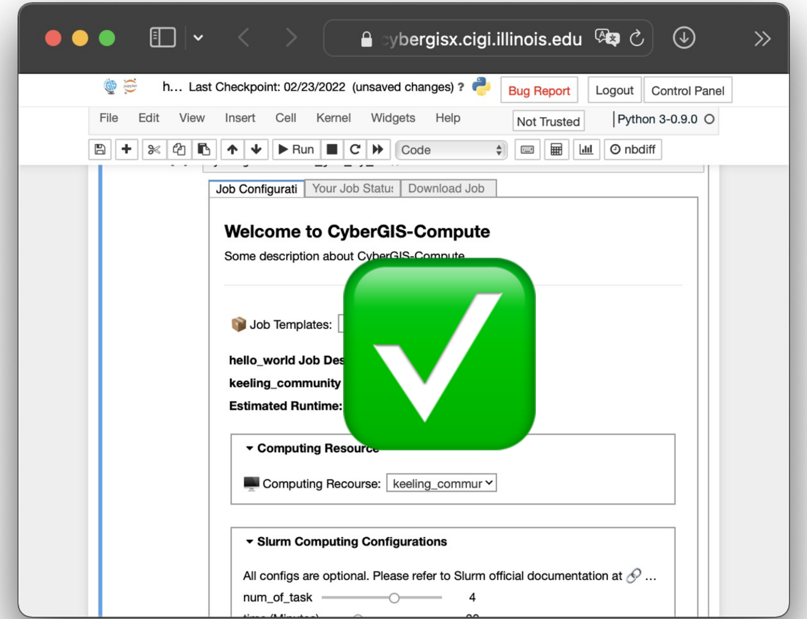
What is CyberGIS-Compute?

- 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)

This is the CentOS 7.9 deployme    f the Scho    f Earth,
Society, and Environment high-pe formance compu e 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.
```



Welcome to CyberGIS-Compute

Some description about CyberGIS-Compute

Job Templates:

hello_world Job Des

keeling_community

Estimated Runtime:

▾ **Computing Resource**

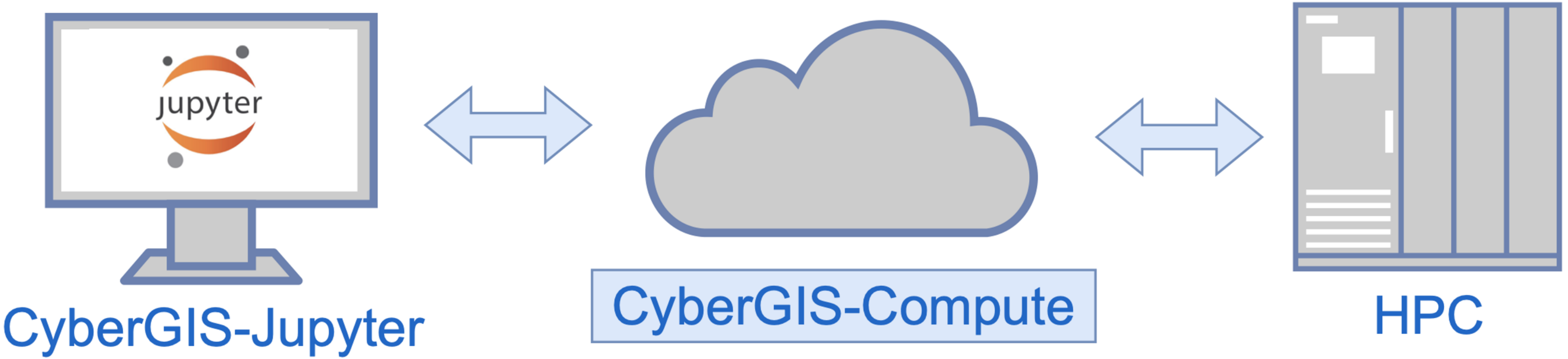
Computing Recourse: keeling_commur

▾ **Slurm Computing Configurations**

All configs are optional. Please refer to Slurm official documentation at …

num_of_task 4

Bridging Ease of Use with Powerful Computing



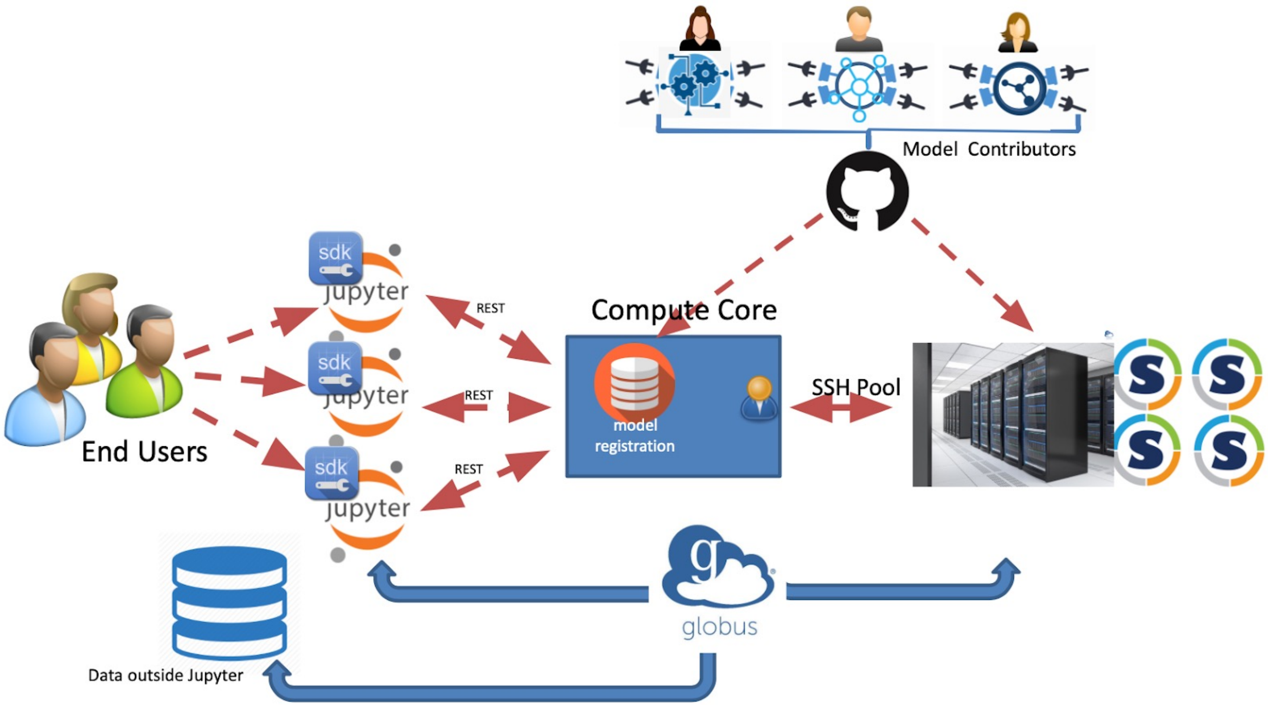
CyberGIS-Jupyter

CyberGIS-Compute

HPC

- 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

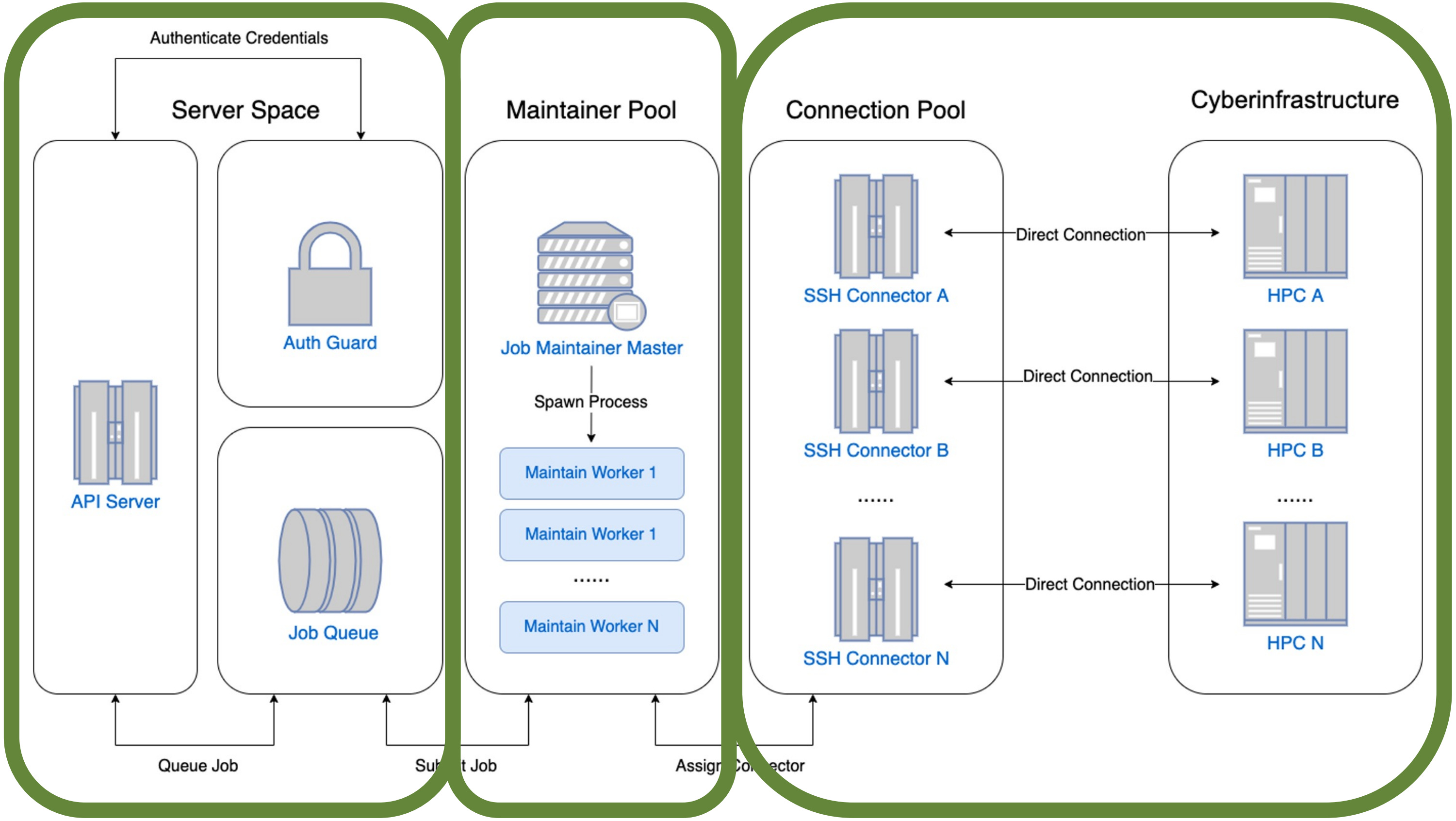
- **Core**: middleware server that automates job submission to HPC
- **SDK**: interactive client for Jupyter Notebook with code-less UI support
- **Contribution**: developer API that enables workflow contribution with little to no modification of existing code

CyberGIS-Compute Core

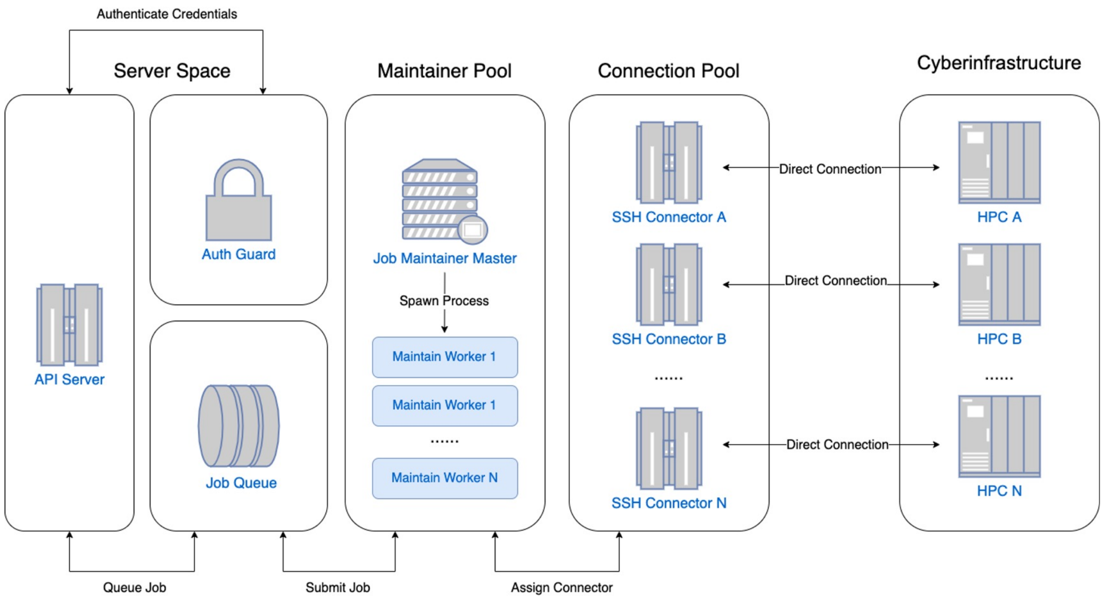
CyberGIS-Compute SDK

Contribution

Architecture



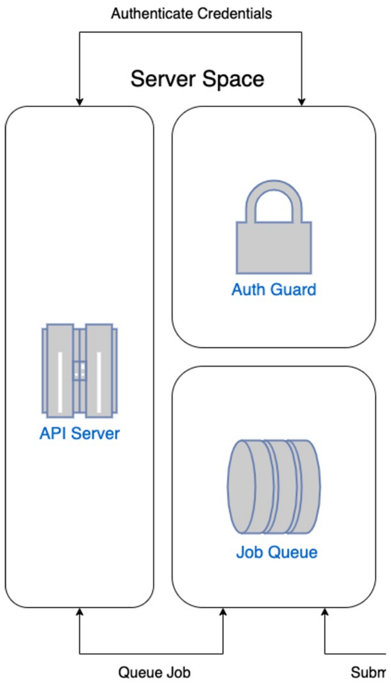
Middleware Server



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

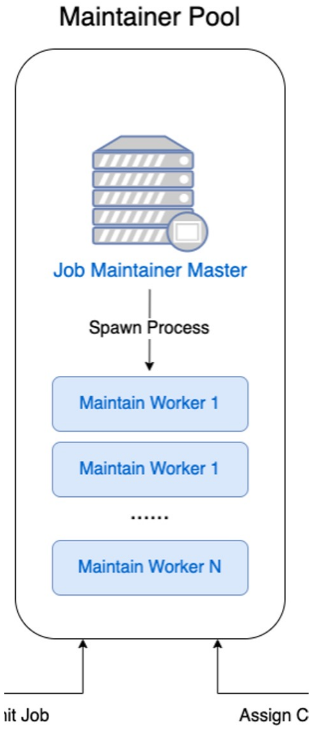
Padmanabhan, A., Xiao, Z., Vandewalle, R, C., Michels, A., and Wang, S. (2021) ”Enabling computationally intensive geospatial research on CyberGIS-Jupyter with CyberGISCompute”. In: *Proceedings of the Gateways 2021 Conference*, October 19 – 21, 2021, DOI: 10.5281/zenodo.5570056

API Server



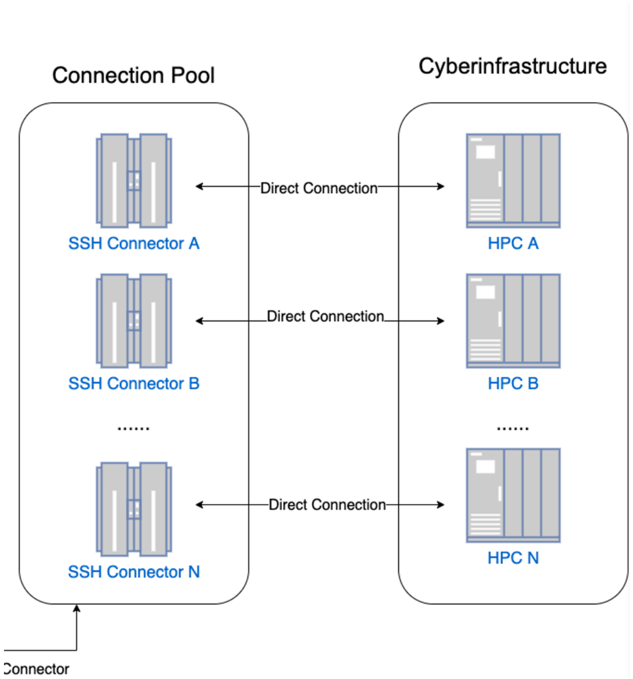
- Provides a **user-facing RESTful web interface** (API Server) for authentication and interaction with internal components of CyberGIS-Compute
- Pushes computation jobs into an internal job queue that the Maintainer Pool consumes

Maintainer Pool



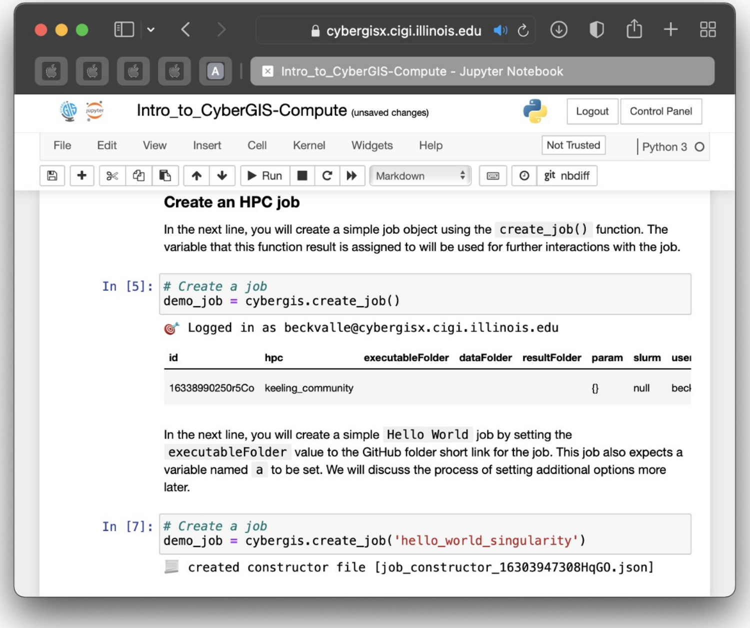
- Maintainer Pool has a multithreaded life cycle that spawns and oversees worker processes called **Maintainer Workers** that contain **logic** on submitting, stopping, resuming, and ending a task executing on a remote HPC resource
- A Maintainer Worker can use a **Connector** to interact with remote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


cybergisx.cigi.illinois.edu
Intro_to_CyberGIS-Compute - Jupyter Notebook
Intro_to_CyberGIS-Compute (unsaved changes)
Logout
Control Panel
File
Edit
View
Insert
Cell
Kernel
Widgets
Help
Not Trusted
Python 3
Run
Markdown
nbdiff

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	slurm	usei
16338990250r5Co	keeling_community				{}	null	beck

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_16303947308HqGO.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

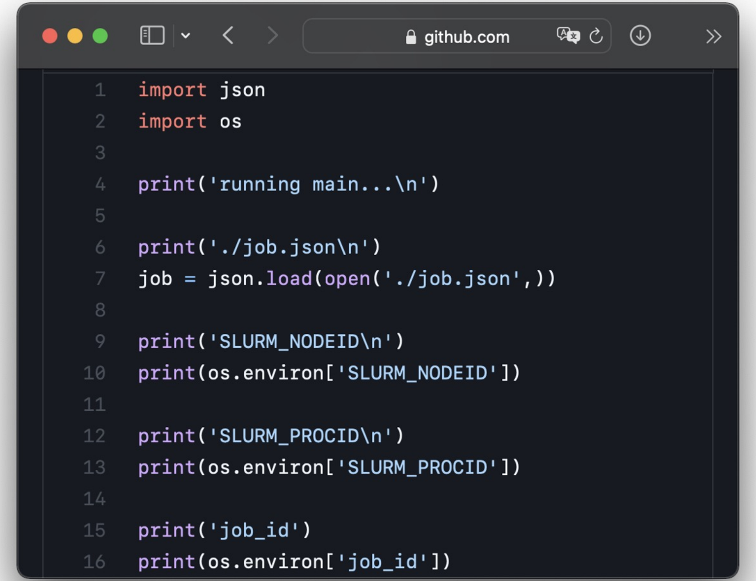
CyberGIS-Compute Contribute allows users to **submit workflow code hosted on GitHub repositories** to be executed on HPC resources

Submissions are verified through a checking process

Provides configurations, system environment, and developer API

Git commit version lock for security

Contribution Process



```python
import json
import os

print('running main...\n')

print('./job.json\n')
job = json.load(open('./job.json',))

print('SLURM_NODEID\n')
print(os.environ['SLURM_NODEID'])

print('SLURM_PROCID\n')
print(os.environ['SLURM_PROCID'])

print('job_id')
print(os.environ['job_id'])
```

+

manifest.json

=

ready for HPC

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

Acknowledgments

- **National Science Foundation**
  - 2118329
  - XSEDE
- Virtual ROGER

Thanks !

Comments / Questions?

Email:

apadmana@illinois.edu