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
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!
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
Bridging Ease of Use with Powerful Computing


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 Center for Advanced Digital and Spatial Studies

CyberGIS-Compute Core Middleware
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

CyberGIS Center for Advanced Digital and Spatial Studies

Large Datasets

Big Data + globus online = Fast & Reliable Transfer
Seamless Access to HPC

**Containerization**: Run code in familiar environments

**Transparencyly 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...
')
print('./job.json
')
job = json.load(open('./job.json',))
print('SLURM_NODEID
')
print(os.environ['SLURM_NODEID'])
print('SLURM_PROCID
')
print(os.environ['SLURM_PROCID'])
print('job_id')
print(os.environ['job_id'])
```

+ manifest.json = ready for HPC
Hands-on Demo!

Comments / Questions?

Email:

fbaig@illinois.edu
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.
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
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 democratizing HPC for geospatial developers.

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.
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