ESIP Lab Project

**Geoweaver: a web-based system for managing compound geospatial workflows of large-scale distributed deep networks**

September Progress Report

Prepared by

Ziheng Sun, Liping Di

Center for Spatial Information Science and Systems

George Mason University

20180930

Contents

[Executive Summary 2](#_Toc523305779)

[Project Actions 3](#_Toc523305780)

[Project Kick-off 3](#_Toc523305781)

[System Architecture 3](#_Toc523305782)

[Development Environment Setup 3](#_Toc523305783)

[Web wrapper for deep learning and HPC library/software 3](#_Toc523305784)

[Geoweaver programming 3](#_Toc523305785)

[LSTM Experiment 4](#_Toc523305786)

[Geoweaver GitHub repository 5](#_Toc523305787)

[Evaluation 6](#_Toc523305788)

[Next Steps 7](#_Toc523305789)

Executive Summary

This month Geoweaver team has kicked off the project and started the development lifecycle as scheduled. Our progresses in August are listed in the following table:

|  |  |  |
| --- | --- | --- |
| **Milestones** | **Progress** | **Actions** |
| July 31: kick off, set up the development environment, develop web wrapper on top of open sourced deep learning/high performance computing library | 100% |  |
| Sep 30: develop workflow designer and data producer, complete bridge assembly between Geoweaver and data/function resources | 50% | * An initial interface of workflow designer has been developed * The web ssh in bridge assembly has been developed. |
| Nov 30: complete module integration, create and conduct LSTM experiment | 20% | * LSTM experiment is ready in the virtual machines. |
| Jan 31: complete source code wrap-up, upload demonstration video, snapshot cloud instance, finish the GitHub final report and demonstrate it in ESIP winter 2019 | 20% | * Geoweaver repo is transferred to ESIPFed organization on GitHub. * The proposal and report are uploaded to Github. |

Project Actions

System Architecture

The architecture design is put here to remind us to follow the original design and remember our goal during the development.

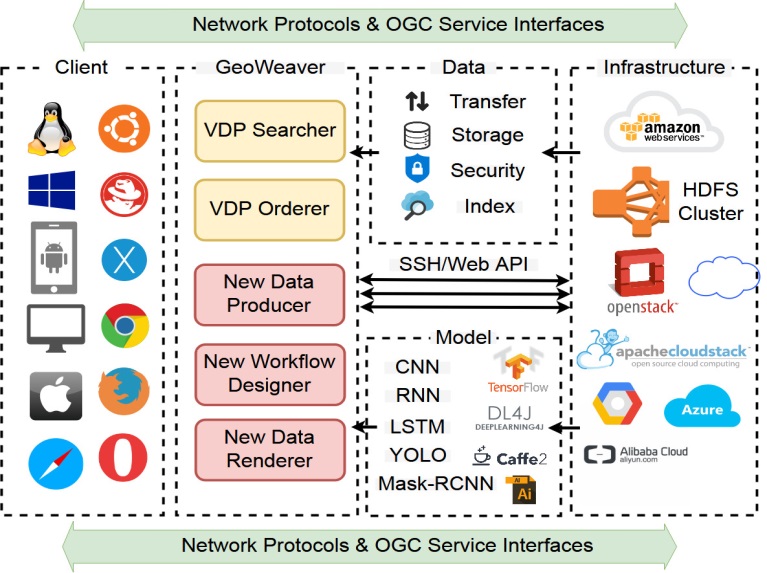


Figure 1. Architecutre design of Geweaver

Geoweaver workflow designer

A new workflow designer based on D3.js is being developed. The initial interface is as shown in Fig. 2. It contains two major areas: workplace and side menu.

The side menu has three resource categories: host, process, and workflow. A host is an infrastructure which could be a PC, a linux, a rack server, a virtual machine or even a cluster. A process is an executable which could be a Shell script, a c/c++ program, a python code, a web service, etc. Process represents the atomic unit in business logic. Process must run on host and could run on various host as long as the environment was already set up. Workflow is a chain or just a graph of composition of processes.

The workspace is a graphic intuitive interface for users to visualize the three kinds of virtual resources. In Fig. 2, each circle with a name inside denotes a process. The connections among them mean the control flow which controls which and when the process is executed. The workflows are sharable and reusable knowledge which are the key achievements of this project to foster the innovation of deep learning by simplifying the learning curve, facilitating the management of massive resources and eliminating the confusion within process execution.

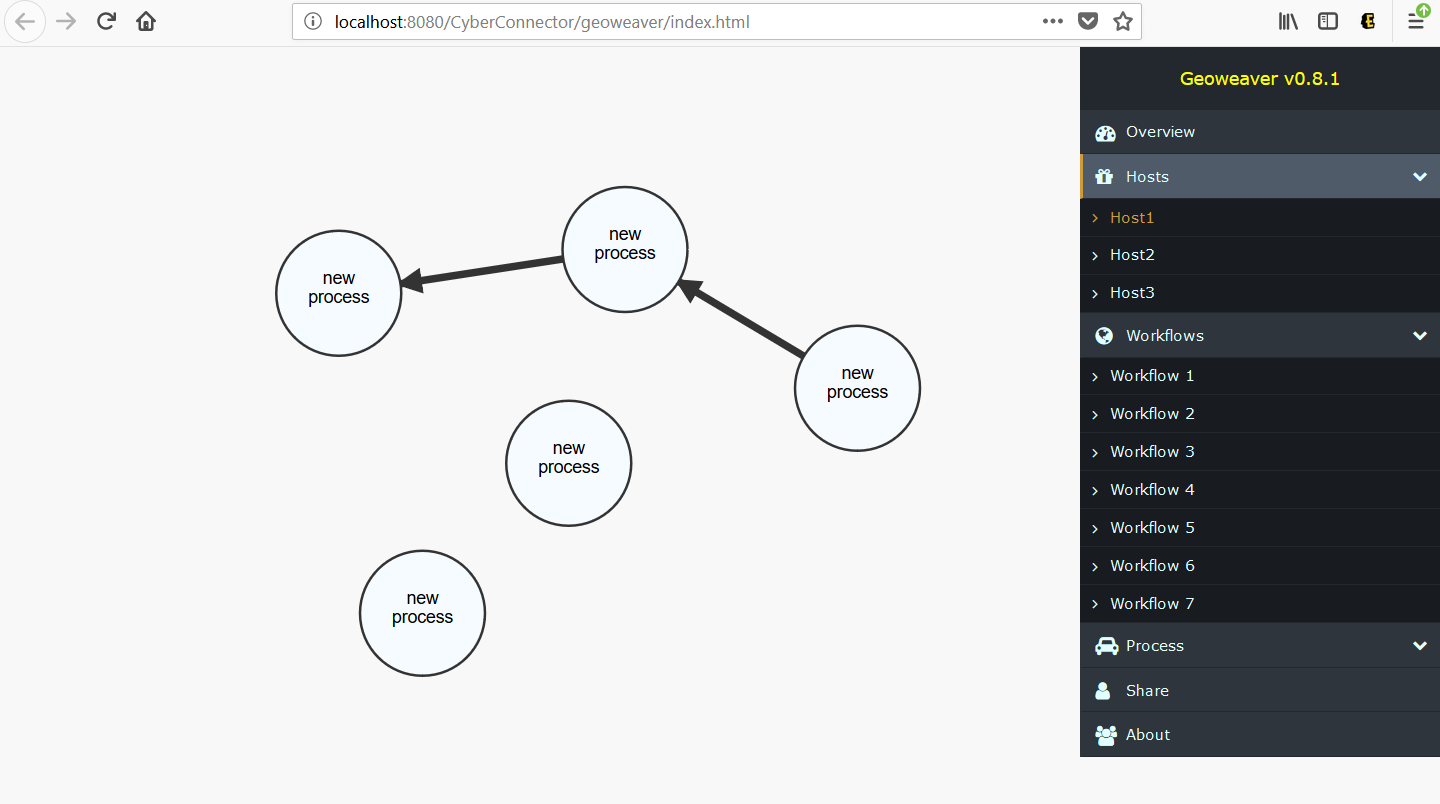


Figure 2. The initial interface of Geoweaver

Geoweaver bridge assembly

To communicate with the involved hosts in geoweaver, we established a web ssh within Geoweaver. Scientists can use this entry to access their servers like they did via SSH secure client or putty. One advantage of having this web ssh is that scientists can easily capture their routine data preprocessing scripts and turn them into managed processes in Geoweaver. Geoweaver will take care of these steps and allow scientists to accomplish data processing steps by only one single click.

Fig. 3 shows the login interface of Geweaver web ssh. Fig. 4 shows the command line interface and an example result after executing a command.

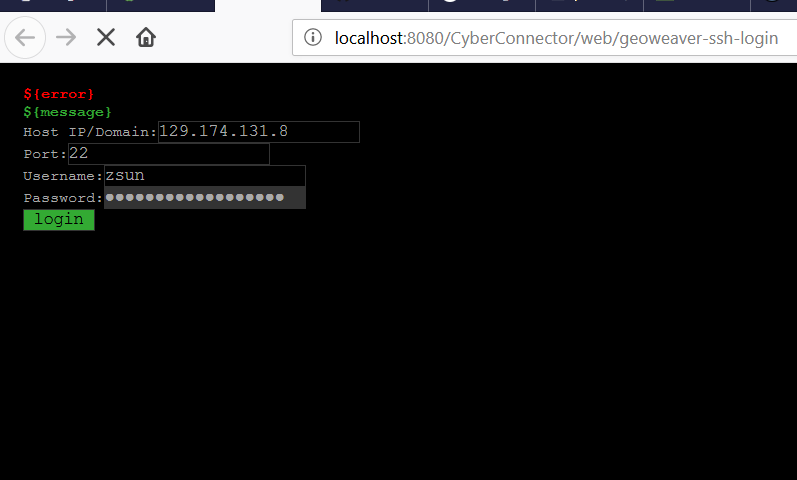


Figure 3. Geoweaver web SSH login interface

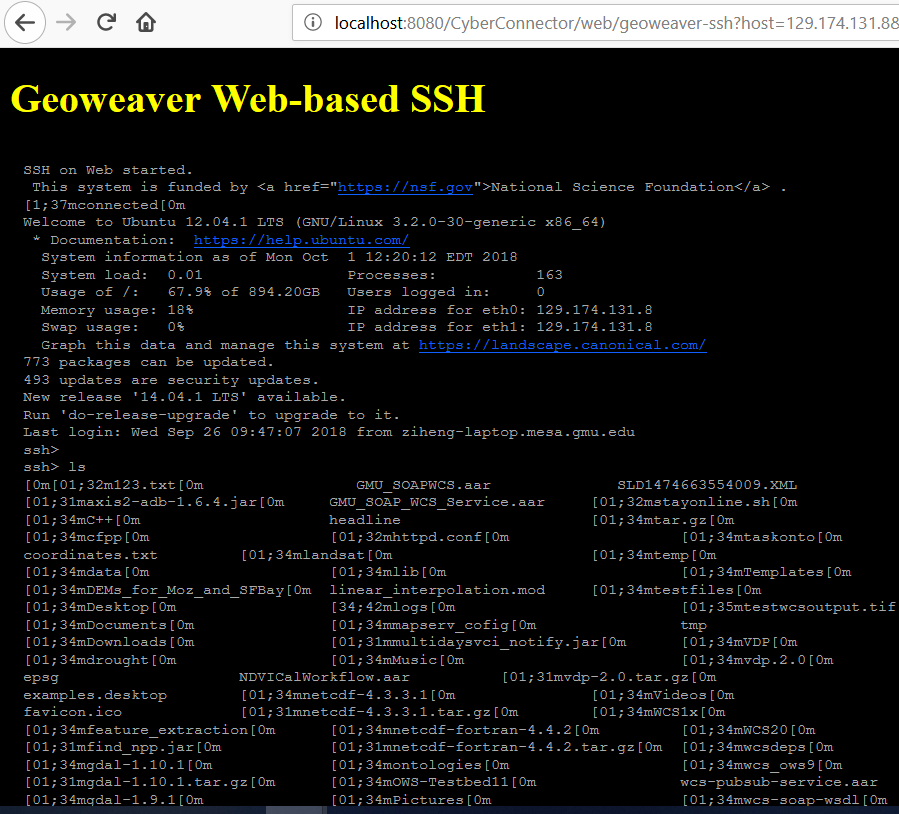


Figure 4. Geoweaver web SSH command line interface

Geoweaver Data Renderer

Rendering module will reuse the module from CyberConnector.

Geoweaver GitHub repository

The github repository has been transferred to ESIPFed (https://github.com/ESIPFed/Geoweaver). All the documents have been uploaded to the repository. I am organizing the source code and hope the code can be pushed there soon.

Evaluation

Not yet.

Next Steps

* Continue the development of Geoweaver workflow designer and bridge assembly.
* Complete the development of Geoweaver web ssh module to collect user command lines into processes.
* Use Geoweaver to manage our LSTM experiments.
* Evaluate Geoweaver.