CyberGIS Center for Advanced Digital and Spatial Studies

University of Illinois at Urbana-Champaign

A Sustainable and Reproducible Geospatial Analytics with CyberGIS-Jupyter

Anand Padmanabhan, Alexander Michels, Shaohua Wang, and Shaowen Wang

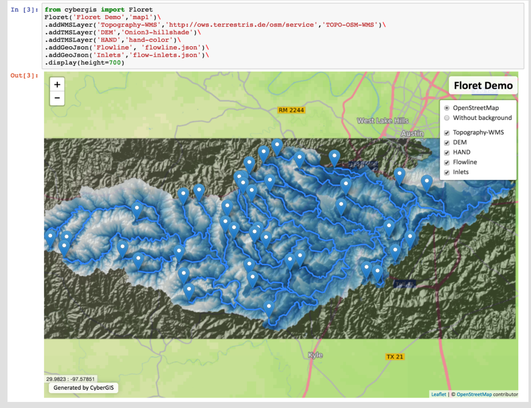
Goal: Implement a scalable platform that conduct sustainable and reproducible geospatial analytics at scale: (a) geospatial analytics of CyberGIS-Jupyter; and (b) batch job submission of computationally intensive models.

CyberGIS-Jupyter

CyberGIS-Jupyter is an innovative cyberGIS framework for achieving data-intensive, sustainable, reproducible, and scalable geospatial analytics using Jupyter Notebook.

Purposes

- Achieves data-intensive, reproducible, and scalable geospatial analytics using Jupyter Notebooks
  - Provides a holistic solution
  - Makes sharing codes and workflows easy
- Perform sustainable geospatial research
  - Cloud Native GIS (Kubernetes)
- Reduces the barrier to accessing advanced cyberinfrastructure and cyberGIS capabilities
  - Exploits JupyterHub, cloud and high performance computing resources



Use Floret to Visualize Geospatial Data

Geovisualization

- Interactive map generation inside notebooks
- Support multiple formats of geospatial data
- Layer management, transparency and styles

Capabilities

- Provides notebook servers with cyberGIS libraries and many geospatial software packages installed
  - Built-in cyberGIS capabilities to accelerate gateway application development
    - E.g., HAND Application
  - Geospatial data, analytics, algorithms, and workflow runtime environments are encapsulated into application packages
- Deployment can be elastically scaled to accommodate the computational needs of cyberGIS users
  - Straightforward management and maintenance of computational infrastructure
  - Seamless scaling between Virtual ROGER and XSEDE JetStream

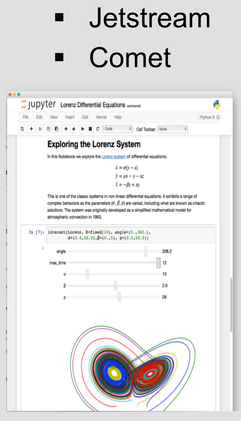
Architecture and Implementation

HPC and Cloud Resources

- Virtual ROGER
  - HPC
  - VMWare Cloud
- XSEDE
  - Jetstream
  - Comet

Architecture layers

- User layer
- Application layer
- Cloud resources
- HPC resources

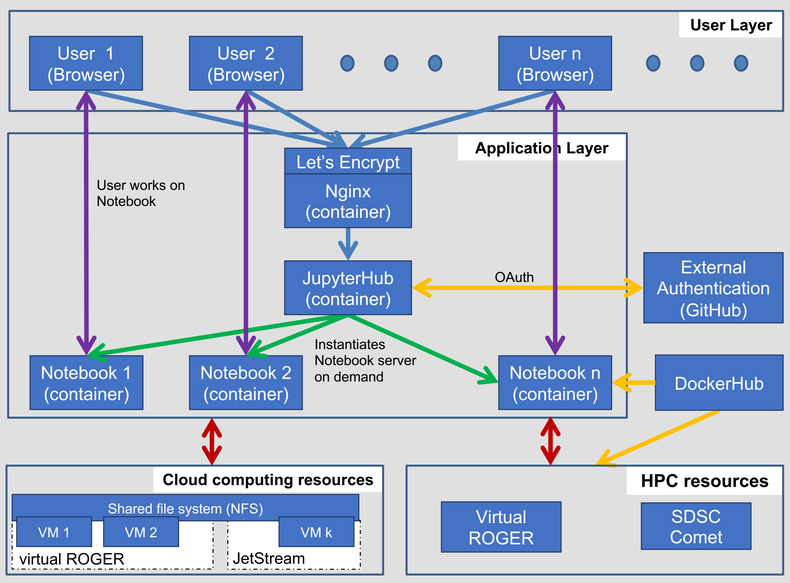


Batch Job management

Cloud-based architecture

Data/storage synchronization

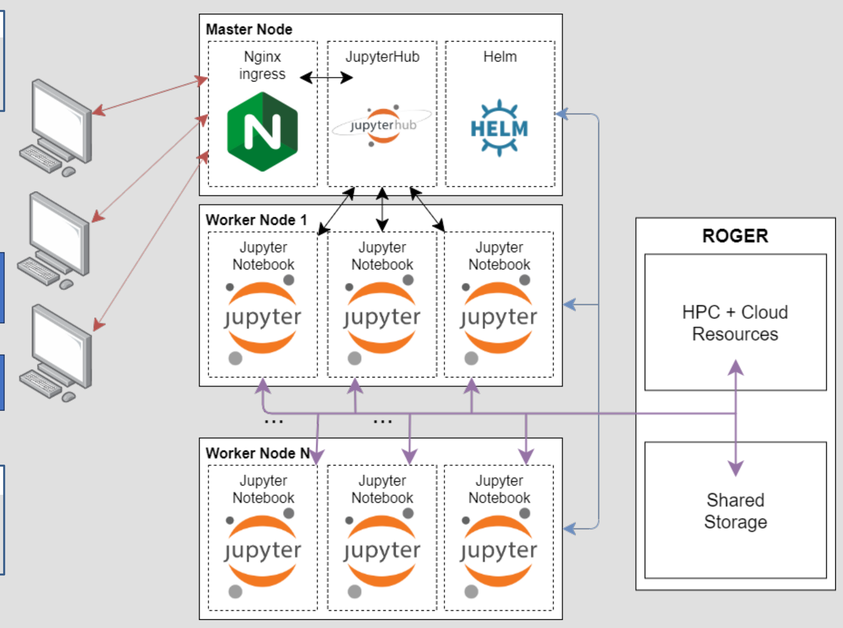
CyberGIS-Jupyter Technologies



Containers might run on any of the VMs. Shared file system allows data to be persisted between sessions.

HPC jobs to submitted from the Jupyter Notebook to batch system on virtual ROGER and Comet using CyberGIS libraries

Architecture of Docker Swarm CyberGIS-Jupyter Deployment



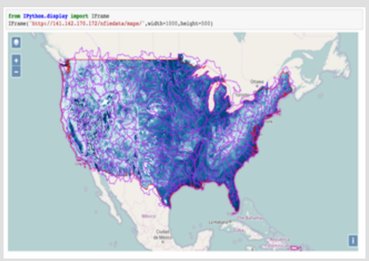
Architecture of Kubernetes CyberGIS-Jupyter Deployment

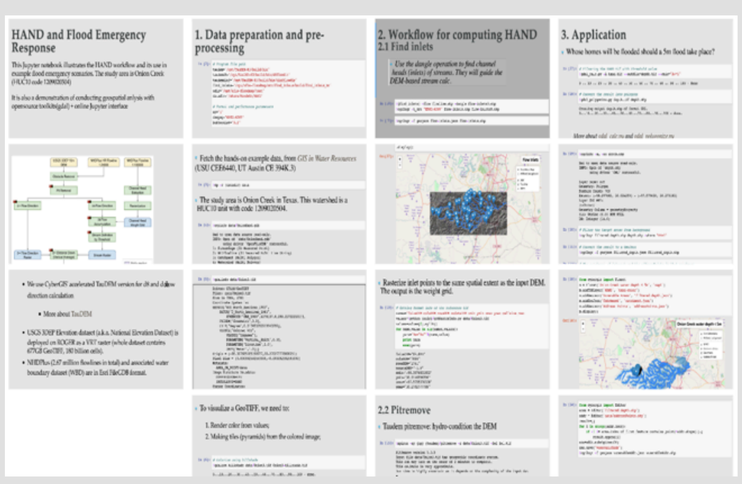
Case Studies and Results

Map Flood Inundation at Continental Scale

Calculate **Height Above Nearest Drainage (HAND)** at 10m for continental US

- Flood analysis map derived from 10m USGS 3DEP national elevation dataset (180 billion cells) and National Hydrography Dataset (2.67 million stream reaches, raw data size: 5.2TB)





HAND Notebook in CyberGIS-Jupyter

Goals

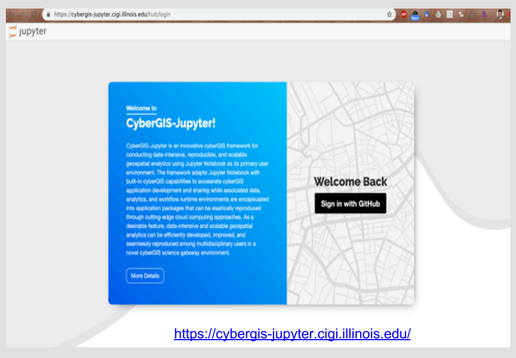
- Collaborative methodology development
- Scalable data analytics
- Deliver methodology and data products to different user communities
  - Collaborators
  - Researchers
  - Decision makers
  - Students

AAG-UCGIS Summer School

- Week-long summer school on Reproducible Problem Solving with CyberGIS and Geospatial Data Science
- Held between July 8 and 13, 2019 at CyberGIS Center at UIUC
- Over 35 students and 10 mentors engaged for intensive problem solving using CyberGIS-Jupyter
- Over 50 Jupyter notebooks were simultaneous running as docker container using a distributed computing infrastructure

Summer School students presenting their work using CyberGIS-Jupyter

User Interface



https://cybergis-jupyter.cigi.illinois.edu/

CyberGIS-Jupyter Login Page

Concluding Discussion

- Ability to bridge a local cyberinfrastructure with XSEDE
- CyberGIS-Jupyter demonstrated ability to support reproducible problem solving at scale
- Provides a valuable educational environment
- Kubernetes-based deployment has been developed and is being tested to support automatic scaling

Acknowledgements: These materials are based upon work supported in part by the National Science Foundation under Grant No. 1443080, 1664119, and 1743184. This work used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation under grant number ACI-1548562. We greatly appreciate the XSEDE allocation support through two awards: SES170023 and EAR190007 and the support provided by the XSEDE Extended Collaborative Support Service (ECSS) program.

Contact: Anand Padmanabhan • University of Illinois at Urbana-Champaign • apadmana@illinois.edu • http://cybergis.illinois.edu