CyberGIS-Jupyter for Sustainable and Reproducible Geospatial Analytics
Anand Padmanabhan, Alexander Michels, Shaohua Wang, and Shaowen Wang

**Goal:** Implement a scalable platform that conduct sustainable and reproducible geospatial analytics.

CyberGIS-Jupyter

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

**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

**CyberGIS-Jupyter Technologies**
- JupyterHub, cloud and high performance computing resources
- Virtual ROGER and Comet

**CyberGIS-Jupyter Deployment**
- User works on Jupyter Notebook
- Cloud computing resources
- Virtual machines
- HPC resources
- Containerization
- Docker
- Networking
- Security

**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)

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

**AAG-UUCGIS 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

**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

**User Interface**

**Acknowledgements**: These materials are based upon work supported in part by the National Science Foundation under Grant No. 1443080, 1684119, 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 education support through two awards: SES170023 and EAR160007 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