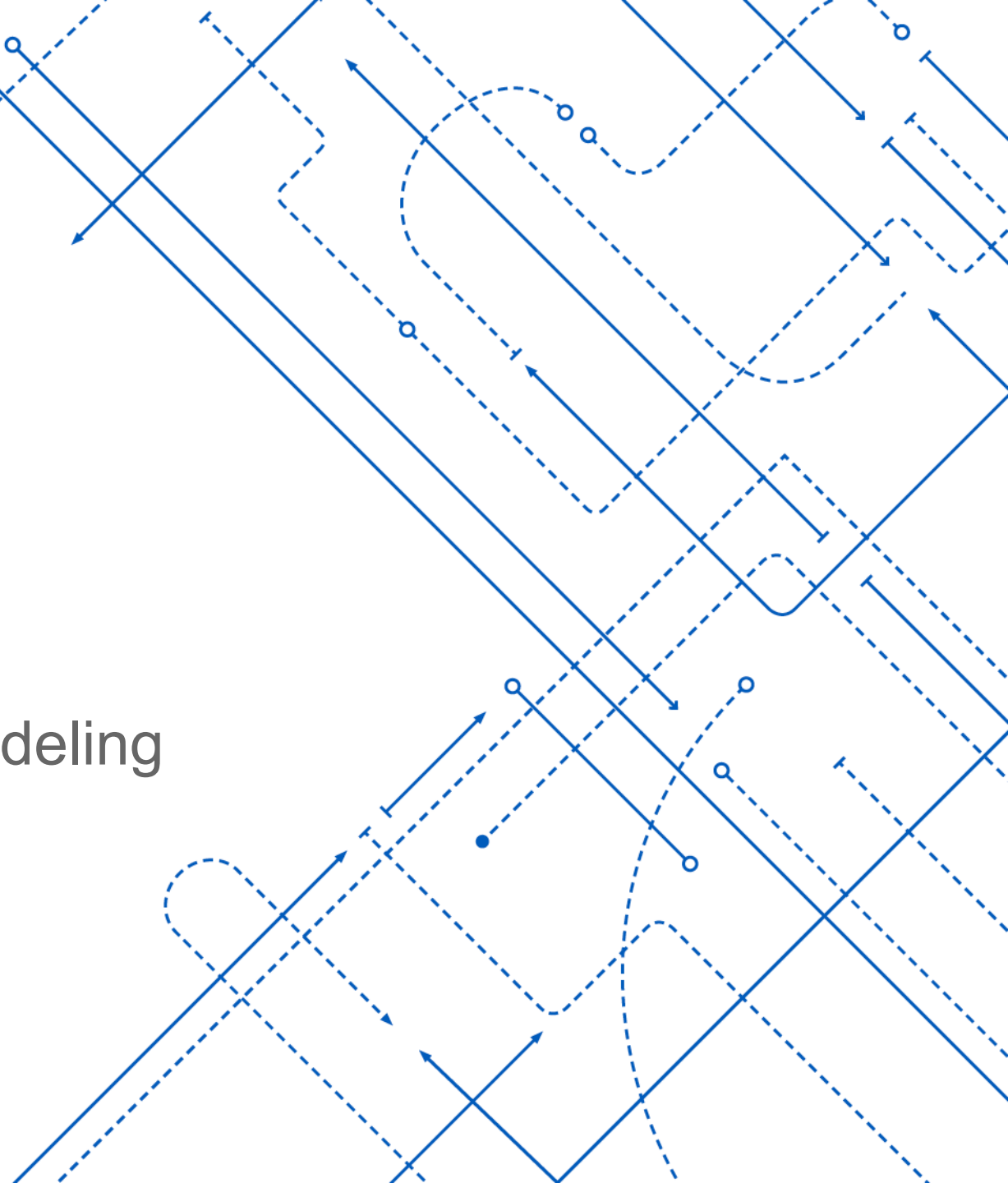
CLOUD COMPUTING

GLY 606, Water Data Analysis & Modeling

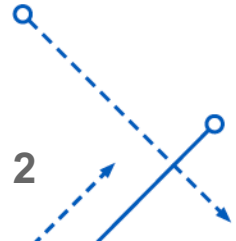
Aug 30th 2024



University at Buffalo The State University of New York

Recap – hydrological modeling

- Coding languages
  - What language is most used for GCMs and hydrological models?
- Important concepts
  - Water balance equation
    - $P = Q + E + \Delta S$
  - How do we define the boundary of watersheds?
    - Based on topography
  - Lumped *versus* distributed hydrological models?
  - Process-based *versus* ML hydrological models?

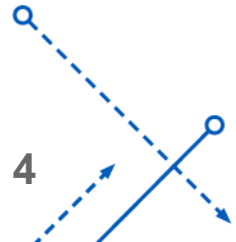


What platforms can we use to run hydrologic models?

PC / Laptops

Challenges to run hydrologic models on PCs

- Software / module dependencies
  - Software installation and system configuration can be tricky and time-consuming
- Data accessibility
  - Voluminous input (such as meteorological forcing) for distributed hydrologic models
- Limited computing resources
  - Basic concept: core-hour (A unit of measurement for the amount of computational activity that occurs on a single core for a period of one hour)



What platforms can we use to run hydrologic models?

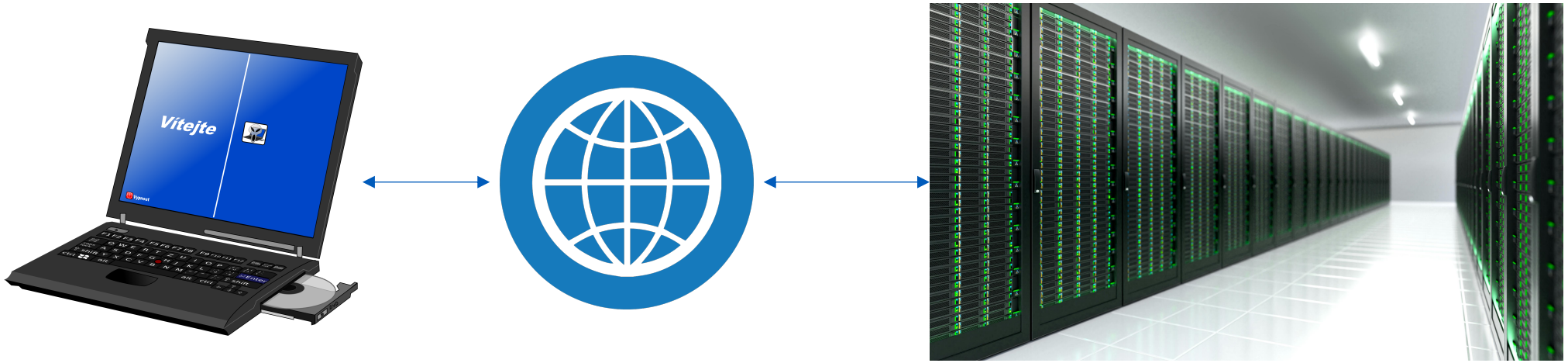
High Performance Computing / Servers

- Major research institutions have their own HPC centers or servers

University at Buffalo Center for Computational Research

- Great technical support
- Large data storage
- Limited to researchers (Usually not accessible to the general public)

What platforms can we use to run hydrologic models?



Cloud computing

- Cloud computing has more flexibilities than HPC centers
- However, the cloud computing can be more expensive per core-hour than HPC.

Many available resources

- Google Colab and Azure Notebooks are great resources to start learning Python.
- They both have free versions.
- However, in the free versions, the users cannot customize the virtual environment for Python (i.e., you cannot install the desired packages) and have no access to terminals

Many available resources

CUAHSI
allied for water science
www.cuahsi.org

Google colab
GETTING STARTED WITH GOOGLE COLAB

PANGEO

Azure Notebooks

Google Earth Engine

Google Earth Engine

- Widely used in remote sensing community
  - Hosts satellite imagery and stores it in a public data archive that includes historical earth images going back more than forty years
  - Landsat and Sentinel-2
- The available API for Google Earth Engine is Python and JavaScript.
  - Customization of Python environment is supported
- GEE is usually not used to run hydrologic models

PANGEO

PANGEO

- Pangeo is first and foremost a community promoting open, reproducible, and scalable science.
  - This community provides documentation, develops and maintains software, and deploys computing infrastructure to make scientific research and programming easier.
- The Pangeo software ecosystem involves open source tools such as **xarray**, iris, dask, **jupyter**, and many other packages.

DASK

Parallel computing

jupyter

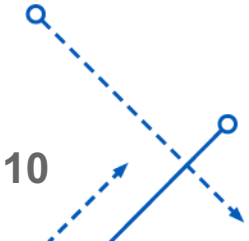
Interactive computing

xarray

Geospatial datasets

Iris

Analyzing and visualizing meteorological and oceanographic data sets



CUAHSI

CUAHSI
allied for water science
www.cuahsi.org

Consortium of Universities for the Advancement of Hydrologic Science, Inc. (Sponsored by National Science Foundation)

- Support water science through education and data services
- Provide free and open source software for managing, archiving, sharing, discovering, publishing, and analyzing all types of water data
  - Hydroshare
  - Jupyterhub
  - MATLAB Online
  - Hydrologic Information System (HIS)
  - Model Domain Subsetter

CUAHSI

Consortium of Universities for the Advancement of Hydrologic Science, Inc. (Sponsored by National Science Foundation)

- Support water science through education and data services
- Provide free and open source software for managing, archiving, sharing, discovering, publishing, and analyzing all types of water data
  - **Hydroshare**
  - **Jupyterhub**
  - MATLAB Online
  - Hydrologic Information System (HIS)
  - Model Domain Subsetter

CUAHSI
allied for water science
www.cuahsi.org

HYDROSHARE

- **Discover** multiple types of water data published by others.
- **Share** data with colleagues and groups.
- **Formally publish** data with a Digital Object Identifier (DOI) so your work can be easily cited.

Demo time

- Invited speaker: Clara Cogswell (CUAHSI)

Github

- Please create a Github account