2024 Unidata Community Equipment Awards Proposal

Building Capacity and Reducing Barriers for Geoscience Students at Northern Illinois University

Date: March 25, 2024

Principal Investigator Name: Allison C. Michaelis
Title: Assistant Professor
Institution: Northern Illinois University
Telephone number: 336-558-4293
Street Address: 218 Normal Road, 118 Davis Hall, DeKalb, IL 60115

Email address: amichaelis@niu.edu

Signature of PI:_____

Co-Principal Investigator Name: Vittorio (Victor) A. Gensini
Title: Associate Professor
Institution: Northern Illinois University

Email address: vgensini@niu.edu

Signature of co-PI:_____

Co-Principal Investigator Name: Alex M. Haberlie
Title: Assistant Professor
Institution: Northern Illinois University
Email address: ahaberlie1@niu.edu

Signature of co-PI:_____

Name of Institution Official: Yvonne Harris
Title: Vice President for Research and Innovation Partnerships
Telephone number: 815-753-1217

Email address: yharris@niu.edu

Signature of University Official: _____

Building Capacity and Reducing Barriers for Geoscience Students at Northern Illinois University

Project Summary

We propose a classroom-dedicated JupyterHub server for geoscience courses at Northern Illinois University (NIU) in the Department of Earth, Atmosphere, and Environment (EAE). Many of our core required courses, especially for Meteorology, include Python-based laboratory assignments involving viewing, downloading, and/or analyzing Unidata-hosted products (e.g., satellite and radar data, forecast model output). Yet, students often have trouble compiling their own Python environments and properly installing the necessary software, thus creating a "barrier to entry" most critically impacting students with limited domain knowledge in computing. JupyterHub mitigates the formation of this barrier, allowing students to write code in classroom-consistent, managed environments set up by the Instructor or Teaching Assistant. In fact, the current EAE JupyterHub server (https://triton.niu.edu) has become so popular (both for undergraduate teaching and for graduate-level research), that we are in desperate need of converting Triton to a research-only server and establishing a new server solely dedicated to undergraduate teaching. This server will, therefore, be a vital component for undergraduate teaching, aiding in the instruction of approximately 50 undergraduate students each semester, and ensuring that we streamline and improve our curriculum to: (1) continue training our students on state-of-the-art computing techniques, thus producing STEM/computer-literate graduates that remain competitive on the job market, (2) increase accessibility and create a more uniform computing environment for our students, and (3) expand computing access to more courses within EAE.

Project Description

Goals of the Project

Established in 1895, NIU is now a public Carnegie R2 institution and emerging Hispanic-Serving Institution (HSI) located ~70 miles west of Chicago in DeKalb, IL. As an emerging HSI, NIU draws many students from populations that are traditionally underrepresented in STEM. For example, out of NIU's 12,000+ undergraduate students, more than 40% are Pell Grant eligible, more than 50% identify as female, more than 50% identify as a racial minority, and more than 50% are first-generation students. Thus, as the longest-standing Meteorology program in the state of IL, and the only degree-granting institution for Meteorology north of I-80, we attract a diverse student body wishing to study Meteorology and other geoscience disciplines (e.g., Climate Science, Geography, Geology). Within the broader EAE department, we have four full-time Meteorology faculty members who regularly engage with undergraduate students through course work and undergraduate research projects. Outside of Meteorology, we 14 full-time faculty across other earth and environmental disciplines.

In recent years, computer programming and data analysis have become a vital part of our geoscience curriculum within EAE, a trend that is expected to continue well into the future given job market demands in STEM fields. Under a previously proposal awarded to co-PI Gensini in 2019, EAE has adapted to this

growing need by increasing our capacity to display and process data through https://weather.niu.edu, an LDM that facilitates viewing and downloading of GFS forecast products. Additionally, several of our required Meteorology courses involve Python-based laboratory assignments involving downloading, viewing, and/or analyzing real and/or idealized datasets. However, with such a diverse student body, significant barriers exist regarding equitable access to computing resources. For one, not all students are able to provide their own personal laptops, and secondly, even if/when computers are provided, not all students enter our curriculum with the same level of computing skills. The requested JupyterHub server is, therefore, critical to providing consistent environments for our students that are managed by the classroom Instructor and/or Teaching Assistant, thus mitigating these barriers to ensure our courses remain inclusive and equitable for any student aiming to develop, or enhance, their computer programming skills. Furthermore, with the requested server, we can increase the accessibility to computational, Python-based course work to other classes within EAE.

In short, integrating the requested JupyterHub server into our Meteorology and geoscience curricula will help meet the diverse and growing computing and data technology needs of our students. By increasing our capacity to provide uniform classroom computing environments, we will support equitable training of undergraduate students across Meteorology and EAE disciplines on state-of-the-art computing techniques consistent with the job demands of STEM-related industries.

Project Activities

The requested server will be an invaluable asset for several courses in both the Meteorology and EAE degree paths, providing computing resources in the form of a Jupyterhub server and interface to undergraduate students who are on campus or virtually connected to our VPN. One such course that will greatly benefit from the requested server is *EAE 493: Computer Programming for the Geosciences*. In EAE 493, students are introduced to computer programming and computer science topics using the Python programming language, often for the first time. Examples of use-cases in EAE 493 include running introductory programming assignments (e.g., defining variables, using data types, and control structures), intermediate programming assignments (e.g., plotting data, reading and analyzing a csv file, and reading and analyzing geospatial data), and a semester-long term project that requires students to apply what they have learned to a climate science problem. This server will significantly simplify the introduction of computer programming to students, allowing them to focus more on logic and problem solving and less on maintaining and/or fixing their programming environment. Furthermore, for the more advanced assignments, the new hardware will allow for a more seamless learning experience that will reduce the time students spend debugging their programs.

In *MET 421: Synoptic Meteorology*, our students apply their programming skills to synoptic-scale data visualization and analysis. Unidata software and data platforms like MetPy, netCDF, Siphon, THREDDS, and LDM are used by our students to either visualize or access data for several laboratory assignments, case studies, and semester projects. Examples of use-cases in MET 421 include manipulating data arrays using NumPy, calculating geostrophic wind from real data using MetPy, generating simple 1-D plots with Matplotlib to visualize the relationship between geostrophic wind and latitude or vertical changes in pressure with height, and creating more sophisticated 2-D contour plots to visualize real and idealized datasets. With the requested server providing all students with access to consistent computing resources,

significantly less time will be needed to configure (and trouble-shoot) each student's individual environment, thus increasing the time spent teaching and learning core course concepts.

With the expanded resources provided by the new server, we can extend our computer programming offerings beyond these two courses. For example, our Physical Meteorology course ("MET 300"), Dynamic Meteorology courses ("MET 410 and 411"), and Regional Climatology course ("EAE 370") would greatly benefit from the implementation of computer programming examples and assignments. In MET 300, a use-case for the server would be tasking students to create a time series of temperature data for two locations, convert temperature from ºC to ºF, and find the maximum and minimum values. By comparing the two graphs, students can use real data to practice a core concept from class: identifying physical reasons (e.g., altitude, proximity to water) for differing temporal signatures between two locations. In MET 410/411—also a new class prep for PI Michaelis for the next academic year—a use-case for the server would be to first derive equations by hand, then code the final equations into Python to answer word problems. For example, tasking students to derive the hypsometric equation and then calculate the geopotential height at various pressure levels under certain temperature conditions. In EAE 370, a use-case for the server would be demonstrating climate classification approaches through techniques like k-means clustering using the scikit-learn package. The visual interface used by JupyterHub would be extremely useful for demonstrating how classification is performed on a dataset. In addition, students could learn about the various approaches to gridding point observations using the MetPy interpolation tools. The impressive computational capabilities of the server will reduce the amount of time that the instructor or students wait for the interpolation results, which can be characteristically slow.

Resources Requested

We request funds to acquire the hardware needed to support a classroom-dedicated JupyterHub server for undergraduate teaching. The proposed server configuration is based on previous experience by co-PI Gensini. The server specifications are designed to accommodate the increasing size of geoscience datasets and number of users in the future. Below is the hardware request with the primary specifications itemized; see Equipment Quotes for a full itemized list of the server specifications provided by the manufacturer.

Threadripper PRO WRX90 T140-4U Computer System
AMD Threadripper PRO 7995WX 2.5 GHz (up to 5.1 GHZ max) [96 total cores]
512 GB (64 x 8 GB) DDR5-5600 ECC Reg. Memory
2x 2 TB NVMe PCIe Gen4 M.2 SSD (Primary and Secondary Drive; RAID 10 configuration)
2x 18 TB Seagate IronWolf Pro (Data Storage Drives; RAID 10 configuration)

Information Technology Support Available

The Triton server has been a huge success and asset for EAE, particularly for our students, and this success is owed to an experienced and diverse support team. This team will also be in place to assure that the requested system is similarly operating at full capacity to best serve our student body. EAE has a dedicated lab manager (Josh Schwartz) and two faculty members who will maintain and fix hardware and software issues on the server (Co-PI Gensini and Alex Haberlie). Co-PI Gensini also maintains servers that use Unidata solutions like LDM to provide daily data for various projects and is well-versed in the best practices for maintaining operational systems. We also work closely with experts in the College of Liberal Arts and

Sciences (CLAS Techs) that assist with general computing issues (Rob Carlson) and acquisitions (Aleksandar Kosoric), assuring that any downtime due to NIU network or infrastructure issues will be extremely limited.

Benefits for Education or Research

Deploying the proposed server to integrate programming and data science topics throughout the EAE curriculum will create a host of educational and professional advantages that are pivotal in keeping the NIU Meteorology and broader EAE curriculum aligned with the evolving scientific computing and data analysis needs of STEM-related industries. The new server will not only accommodate our current needs, but will also increase computing access to other geoscience courses within EAE, including our senior-level undergraduate and graduate-level seminar courses. Additionally, the changing demographics at NIU, and more broadly in STEM courses, has increased the number of students who have little-to-no computational knowledge nor the financial ability to purchase a modern laptop. These shifts present unprecedented equity concerns—namely, are students with extensive computing experience and/or expensive computers more likely to succeed in courses that involve computer programming? With this server, we can assure equity across our courses by providing all students with the identical access to powerful computing capabilities. Our current EAE server, Triton, has been severely strained by the concurrent connection of more than 50 undergraduate and graduate students running scripts that range from "Hello World" to parallelized research analyses on 10s of TB of data, and therefore, will soon be converted to a research-only server, thus removing our capacity for providing classroom-based computing resources. With the requested equipment, we can ensure that our courses will remain diverse, inclusive, and equitable for all students who want to develop their computer programming skills.

Potential Community Benefit

The requested server will be available through the intuitive Jupyterhub web interface for students at NIU and interested community members. The server will host and/or facilitate materials generated by NIU students and faculty, including curated datasets and guided computer programming and data science tutorials. While the main goal of this system is to empower capacity-building in the classroom by removing the barrier of hardware and system maintenance, a secondary goal is to provide educational and technical resources for the Unidata community. A particular focus will be placed on how to facilitate educating students in our unique areas of strength at NIU, which include severe local storms, regional climate change, climate analytics, and seasonal-to-subseasonal predictability. We will include educational material (i.e., data, examples, and scripts) used in courses that employ the server, as well as guides on how to configure and manage Jupyterhub servers for use at other institutions in the Unidata community. Guided tutorials will be delivered through Github (e.g., "How to use for loops in Python"), the server website (PDF of "Steps to set up and manage a Jupyterhub at your institution"), or a THREDDS server ("Preprocessed and reorganized data for a synoptic case study on April 27th, 2011"). In summary, we will not only provide the typical information related to "setting up a JuyterHub server", but also our unique perspectives related to our student body demographics and faculty interests.

Budget

The proposed equipment will be purchased for a total of $19,573.54:

Quantity	Item	Unit Cost	Total Cost
1	Threadripper PRO WRX90 T140-4U Computer System	$19,573.54	**$19,573.54**

While NIU will not provide direct cost sharing or matching funds toward the purchase of the requested equipment, faculty and staff time needed for the installation, configuration, and maintenance of the server as well as space to house the server is provided by NIU. As an Equipment item only, this budget is exempt from university overhead and indirect funds. See the Equipment Quote at the end of this proposal for the specific manufacture quote.

Project Milestones

Vendor quotes have already been obtained (see Equipment Quotes below), permitting immediate purchase of the equipment as soon as this proposal is accepted, and funds disbursed to NIU. For example, a proposal acceptance in early June 2024 would permit purchase of the new server by late June/early July 2024, followed by installation of the server during summer 2024 while campus activity is at a minimum. Following this timeline, the server would be up-and-running and ready for classroom use by the start of Fall semester (mid-August) 2024. The actual timeline of installation will depend on the final timing of proposal acceptance and fund disbursement.

Equipment Quotes

See attached.

Puget Systems

Puget Sound Systems, Inc.
2707 West Valley Highway N
Auburn, WA 98001

Tax ID 20-0056154
Cage Code 31YA3

QUOTE

Quote # 000256684
Quote Date 03/24/24
Rep Wilson

Bill To

Northern Illinois University
Victor Gensini
43W545 Scott Rd.
Sugar Grove, IL 60554 US
Phone: 8157538696
Fax: 815 303 2381

Ship To

Northern Illinois University
Victor Gensini
43W545 Scott Rd.
Sugar Grove, IL 60554 US
Phone: 8157538696
Fax: 815 303 2381

Item	Qty	Unit Price	Total
Puget Rackstation Threadripper PRO WRX90 T140-4U Computer System	1	$19,573.54	$19,573.54

Puget CX4250i 4U
Asus Pro WS WRX90E-SAGE SE
AMD Threadripper PRO 7995WX
Asetek 360mm AIO CPU Liquid Cooler
DDR5-5600 ECC Reg. 64GB (8)
AMD Radeon Pro W7500 8GB
Onboard Sound
Integrated Ethernet
2TB NVMe PCIe Gen4 M.2 SSD Primary drive.
2TB NVMe PCIe Gen4 M.2 SSD Secondary drive.
18TB Seagate IronWolf Pro Tertiary drive.
18TB Seagate IronWolf Pro Quaternary drive.
Super Flower LEADEX VII Gold 1000W
Case Fans PWM Upgrade Kit
Puget CX4250i Rackmount Rail Kit
Paperless Packet
Lifetime Labor and Tech Support, 3 Year Parts Warranty
Ubuntu 22.04 LTS w/ Gnome Desktop Installation (64-bit) [LIMITED SUPPORT]
Shipping: unselected

Pricing is valid for 30 days from the QUOTE date. Payment or an approved purchase order is required upon placement of your order. This quote is based upon the acceptance of the Terms and Conditions found at pugetsystems.com/terms

$19,573.54

These commodities, technology, or software are subject to the United States Export Administration Regulations. Diversion contrary to U.S. law is prohibited.

For questions concerning this quote, please contact sales@pugetsystems.com or (425) 458-0273