2023 SIParCS Technical Project Descriptions
Summer Internships in Parallel Computational Science (SIParCS)

Undergraduate

- Project 1. Software Engineering: Applying Search Techniques for Scientific Data Discovery and Exploration
- Project 2. Python data analysis & visualization for unstructured grids data
- Project 3. Optimizing GeoCAT Computational Routines for High Performance Computing
- Project 4. Investigating holographic images of clouds with machine learning
- Project 5. Improving the Accessibility of Open IoTwx: Open Hardware Protocol and Documentation
- Project 6. An interactive website to support the co-design process

Undergraduate and Graduate

- Project 7. Using Machine Learning Uncertainty Estimates to Aid Scientific Analysis

Graduate

- Project 8. Reproducible and scalable analysis of remote sensing data in the cloud using Xarray
- Project 9. Optimization of an ocean biogeochemistry model
- Project 10. Just-in-time compilation of a chemistry solver for GPU
- Project 11. Interactive Visualizations of Climate Data
- Project 12. Interactive Visualization of Uncertainty using Data Assimilation and High-Resolution Earth System Ensembles
- Project 13. Explore Native C++ Parallelization Abilities on an Compression Algorithm
- Project 14. Creating bias corrected global climate model output to support regional climate research
- Project 15. Continuous Integration for ASAP applications (CPU and GPU applications)
- Project 16. Containerization of simulation applications for frequently re-run

Note: Please apply to no more than two (2) SIParCS projects.

Non-Technical Project

Graduate

- CISL Outreach Diversity Education (CODE) Intern

Project 1. Software Engineering: Applying Search Techniques for Scientific Data Discovery and Exploration

Areas of Interest in order of relevance: Software Engineering, Digital Asset Management

Description: NCAR's diverse scientific data holdings have historically been difficult for external scientists and users to search across and find the data they need to do their science. While we have a current search system that aggregates these data holdings, we are experimenting with a simpler approach.

This project is focused on continued development of a new scientific data search Java-based web application. This summer we're enhancing our current software by including but not limited to the following: search facets, improving the usability of our web front-end, making our front end more responsive to both large and small screen internet enabled devices, and storing and reading more metadata from our Solr search platform, scientific metadata validation, and email notifications if incomplete metadata is found.

This project does NOT involve artificial intelligence or machine learning.

Students: The project is open to undergraduate students.

Skills and Qualifications: Basic understanding of software development programming. Basic experience in languages such as Java, Javascript, XML, HTML, CSS, and a query language like SQL. Basic understanding of web services and web based user interfaces. Basic understanding of controlled vocabularies and metadata schemas. Ability to interact with diverse mentors and peers in a friendly, professional manner that supports collaboration and inquiry. Good problem solving skills. Good oral and written communication skills. Willingness to learn and use software development tools and programs. Curiosity to explore new things.

Project 2. Python data analysis & visualization for unstructured grids data

Areas of Interest in order of relevance: Visualization, Software Engineering, Data Science

Description: Project Raijin was awarded by NSF EarthCube in order to develop community-owned, sustainable, scalable tools for data analysis and visualization that can operate on unstructured (i.e. not regular lat/lon grids) climate and global weather data at global storm resolving resolutions. The development of Project Raijin leverages the Scientific Python Ecosystem (SPE), particularly the open development Xarray and Dask packages, and the Pangeo community. Our work is conducted under an open development model that encourages participation of the community in all aspects of the project's development. As a result of this, the brand new Python package, UXarray, has been created.

This internship is part of a long-term effort as Raijin had a SIParCS 2022 internship program for the same project. During the SIParCS 2023 summer internship, the intern will have the opportunity of working in a novel research and development project in the SPE by helping implement various data analysis and visualization workflows for unstructured grids such as example plotting scripts for different unstructured grid datasets, Jupyter Notebook based training modules, UXarray usage examples, computational functions, etc. Throughout these efforts, the student will explore and learn about data visualization and analysis in geosciences using commonly used Python tools such as Matplotlib, Cartopy, and Holoviews. The student will also learn high performance computing (HPC) principles through use of NCAR's HPC clusters as well as using parallelization and optimization packages such as Dask, Numba, Datashader, etc. Most or all of the student's work will be made publicly available through our open development model, which will in turn help the intern create a strong Python portfolio in data analysis and visualization. The Project Raijin team is excited to provide the student with an in-depth experience of working within a professional software engineering team; therefore, the student will participate in all project development activities such as regular team meetings, morning standups, hackathons, cross-team discussion/collaboration meetings, debugging and bug-fixing, pair programming, documentation, etc.

Students: The project is open to undergraduate students.

Skills and Qualifications:
- Experience with Python programming
- Familiarity with Jupyter Notebooks
- Ability and willingness to work with a team
- Good communication and writing skills

Optional Skills and Qualifications:
- User-level familiarity with Linux and Unix-based tools for scripting and file manipulation
- Experience with Xarray, NumPy, Matplotlib, Cartopy, Dask, Numba, Datashader, Holoviews
- Familiarity with unstructured grids/meshes

Project 3. Optimizing GeoCAT Computational Routines for High Performance Computing

Areas of Interest in order of relevance: Software Engineering, Application Optimization/Parallelization, High Performance Computing

Description: The GeoCAT project is an open development, community-owned effort, managed by the National Center for Atmospheric Research (NCAR). Its primary goal is to produce Python-based tools that help make sense of geoscience data on both laptops and on supercomputers. Python can be computationally slower than the previous NCAR Command Language (NCL) standard; however, Python is easier to learn, has a wide selection of scientific analysis packages, has a massive open source geoscience community, and can be parallelized. The goal of this project is to optimize existing GeoCAT computational routines, so they run more quickly and efficiently.

During this project, the student will:
- Test the performance of software on laptops and NCAR's Cheyenne Supercomputer
- Explore how the software can be optimized and improved
- Compare the performance of the optimized Python code to the equivalent NCL code with benchmarking techniques
- Document and report on their findings
- Summarize and justify the optimization techniques they used

Students: The project is open to undergraduate students.

Skills and Qualifications: Required skills:
- Be in an undergraduate program for computer science, data science, or related field
- Proficiency with Python 3
- Familiarity with algorithm analysis
- Basic understanding of GitHub
- Strong motivation to learn new skills

Desired but not required skills:
- Familiarity with Linux and Unix-based tools
- Experience with package managers and virtual environments (i.e. Conda)
- Experience with Xarray, Numpy, Dask
- Experience with HPC systems
- Familiarity with physics, atmospheric science, scientific computing, or computational mathematics

Project 4. Investigating holographic images of clouds with machine learning

Areas of Interest in order of relevance: Machine Learning, Application Optimization/Parallelization, Data Science

Description: This project aims to improve the performance of a neural network processor for holographic images of cloud particles obtained using the HOLODEC instrument (https://arxiv.org/abs/2203.08898). HOLODEC is an airborne cloud particle imager developed at NCAR which captures holographic images of liquid and ice cloud particles. A "U-Net" style neural network is used to recognize particles in the holograms after they have been computationally refocused. The objective of this project is to modify the neural net to reduce over-prediction and reduce data preprocessing requirements.

Over the summer, the successful applicant will work with scientists in the Computation and Information Systems Lab (CISL) and the Earth Observing Lab (EOL) toward developing a new training dataset that utilizes multiple depth layers as additional "input channels". Upon successful generation of the training dataset, the applicant will then modify the neural net to leverage this new data and train it. Depending on time, there may be opportunities to explore other potential solutions including mixed recurrent/computer vision modeling approaches as well as improving processing performance on the HPC systems at NCAR.

This project is developed jointly by CISL and EOL.

Students: The project is open to undergraduate students.

Skills and Qualifications: The candidate needs to have prior experience with Python, Jupyter notebooks, and at least basic UNIX/LINUX experience. HPC and GPU experience is desired but is not required.

Prior experience with machine learning, data analysis, and Git/Github is desired, as well as pyTorch.

A willingness to learn!

Project 5. Improving the Accessibility of Open IoTwx: Open Hardware Protocol and Documentation Development for Enhanced Community Engagement

Areas of Interest in order of relevance: Internet-of-Things (IoT), Software Engineering, Data Science

Description: The goal of the Open IoTwx project is to transform equitable accessibility to open source Internet-of-Things (IoT) instrumentation across diverse communities. The aim is to lower barriers and broaden community access to low cost observational instrumentation networks (including design, development, deployment, and data distribution) and to empower a diverse community of citizen scientists to co-design sensornets that best meets their needs.

The student will modify, assemble, and deploy Open IoTwx in both “low data” and “big data” configurations. Low data configurations include standard measurement nodes such as digital rain, wind, and air (T,P,RH, air quality, etc). Big data mode includes a lidar that can produce 100,000 cloud point data per second and high-resolution images from various camera options. This data will be stored locally and a subset of the data will be transmitted via station communications protocols. While this project will not focus on the data, it will focus on open extensibility protocols to allow for a wide range of sensor options.

The student will expand and develop open documentation to allow for the growth of an open source community around the project. In the process of developing this open documentation system the student will apply their hands-on expertise (Arduino, 3D Cad, data communications protocols) to deploy a local station to improve accessibility of the entire build and deployment process. This specifically involves recommendation and design changes to create standard extensible protocols for hardware and IoT software components to promote interoperability while reducing costs and increasing accessibility.

Students: The project is open to undergraduate students.

Skills and Qualifications: Internet-of-things, strong arduino and python, ESP32, MQTT, RS485 protocols, ipfs, robotics, IoT power budgeting, 3D modelling (i.e. CAD, fusion, etc), 3D Printing, github, git pages, Jupyterbook. Interest in the expansion, design, and development of open educational manuals.

Project 6. An interactive website to support the co-design process

Areas of Interest in order of relevance: Software Engineering, website design

Description: At the National Center for Atmospheric Research (NCAR) we perform world-class research in support of our goal of advancing our understanding of our Earth system. A powerful tool that supports our scientific objectives is high-performance computing. We have observed that the uptake of innovative computing technologies has traditionally been slow. We believe that a fundamental reason behind the slow uptake is that it typically takes expert knowledge to determine if a scientific objective is amenable to new technology. While we have created a set of questions that supports this evaluation, these questions are not particularly accessible. We, therefore intend to develop an interactive website that can be used to make this kind of evaluation accessible to a broader audience.

The goal of this 2023 summer internship is to develop an interactive learning website that will simplify the evaluation of the suitability of new technology, in this case, Graphics Processing Unit (GPU) computing for a particular science objective. The student's primary focus will be developing the website and testing it on suitable volunteers.

Students: The project is open to undergraduate students.

Skills and Qualifications: Familiarity with graphical website development. Interest in interactive learning. Strong motivation to learn new skills and resolve issues in a team is required.

Project 7. Using Machine Learning Uncertainty Estimates to Aid Scientific Analysis

Areas of Interest in order of relevance: Data Science, Visualization, Machine Learning

Description: Both weather forecasters and researchers want to know when they should trust the guidance from their models as well as when not to. Being able to understand the sources of uncertainty in their guidance can enable them to convey to decision makers when to wait for further updates versus taking immediate protective actions. However, traditional machine learning models can only provide limited estimates of uncertainty within the realm of their training experience, so they tend to be overconfident in their predictions, especially when being applied to unseen data. A new class of machine learning models called evidential models can estimate the total uncertainty of a single prediction by making strong prior assumptions about how the data and model are expected to behave. In this project, the intern will work with the NCAR AIML group to develop new ways to analyze and visualize uncertainty estimates and explanations of the uncertainty from these evidential models for multiple weather forecasting use cases, including winter precipitation type and severe storm hazard (e.g., tornadoes, hail) prediction.

Students: The project is open to undergraduate and graduate students.

Skills and Qualifications: Prior classwork and/or research experience in statistics and/or machine learning. Scientific Python programming experience, especially with libraries like scipy, matplotlib, Tensorflow, PyTorch, or xarray. Scientific visualization experience preferred. Some background in atmospheric, oceanic, or climate science preferred.

Project 8. Reproducible and scalable analysis of remote sensing data in the cloud using Xarray

Areas of Interest in order of relevance: Data Science, Software Engineering, Reproducible Science

Description: Science today requires software that enables expressive and easily-parallelized workflows on gigabyte to petabyte sized datasets. Xarray is an actively developed open source library that provides scientists with a powerful interface for parallelized computation with multi-dimensional raster datasets (e.g. image stacks), which are prevalent today across all scientific domains. A paradigm shift in scientific computing is underway and modern workflows can leverage Xarray to analyze massive cloud-hosted archives such as NASA's Earth observation archive. Come work with us to transition your research workflows to the commercial cloud, learn new technical skills, and become an open source contributor!
Over the summer, you will learn skills that are key to practicing open science and that are transferable to both academic and non-academic career paths:

- Collaborate with a team of research scientists, data scientists, and software developers to produce publicly-accessible tutorials that leverage Xarray for scientific analysis of Cloud-hosted remote sensing data.
- Contribute to multiple open source geoscientific Python projects (particularly Xarray and RioXarray) as well as general open source tools such as JupyterBook.
- Effectively use cloud-based datasets and computational resources.
- Gain experience with collaborative software development workflows via GitHub.
- Understand the technical components of reproducible computational workflows including testing, continuous integration, and efficiently sharing data and results

We welcome applications from any intern interested in cloud-hosted, geospatial data, and open science. For examples from a prior round of this work see here and here. There is considerable room to tailor this project to your preferences, including an opportunity to attend SciPy 2023!

Students: The project is open to graduate students.

Skills and Qualifications: Experience with basic Python programming.
Familiarity with Jupyter Notebooks.
Ability and willingness to work with a team.

Project 9. Optimization of an ocean biogeochemistry model

Areas of Interest in order of relevance: Data assimilation / parameter estimation, Data Science, Geostatistics

Description: Ocean ecosystems sustain marine fisheries and mediate fluxes of carbon that are important for maintaining the ocean's vast inventory of carbon dioxide. Earth system models represent these ocean ecosystems using numerical models based on empirical constraints from ocean observations and understanding of fundamental biogeochemical processes. However, these models include many parameters specifying, e.g., interactions among trophic levels and between physical and biological processes. In many cases, the "correct" values of these parameters are poorly understood, and these uncertainties translate into errors in model representations of ocean ecosystems and make it difficult to compare simulations to real-world observations and make useful ocean forecasts.

In this project, we are seeking a student to develop a parameter optimization framework for the Marine Biogeochemistry Library (MARBL), which is the ocean ecosystem model coupled to the Community Earth System Model, using the Data Assimilation Research Testbed (DART). DART [dart.ucar.edu] is a sequential ensemble DA tool that has been heavily tested and used for weather forecasting, ocean prediction, climate projections, flood prediction, parameter estimation, and other applications. Here we will exploit simplified (one-dimensional) ocean model configurations to enable rapid prototyping and iteration and consider optimizing a collection of these across a set of "testbed" sites sampling different oceanographic settings. The student will join a dynamic team of researchers with diverse expertise in oceanography and data assimilation; the student will have the opportunity to learn about ocean biogeochemistry and ecology, Earth system modeling, data assimilation, and high performance computing.

Students: The project is open to graduate students.

Skills and Qualifications: Familiarity with Python, Jupyter Notebooks, Linux environment, and Git version control. Experience with differential equations, statistics, physical science, and/or machine learning. Strong motivation to solve problems and learn new skills is essential.

Project 10. Just-in-time compilation of a chemistry solver for GPU

Areas of Interest in order of relevance: Software Engineering, Application Optimization/Parallelization, Numerical Methods

Description: At the National Center for Atmospheric Research (NCAR) we perform world-class research in Earth system science with a particular focus on the interaction between the atmosphere and other components like the oceans and land surface. One of the many critical pieces in our understanding of the Earth system is the way in which chemical reactions impact the atmosphere. Atmospheric chemistry is typically solved by a numerical solver in a weather or climate model. Currently a user executes a “preprocessor” to generate a specialized version of the source code for a specific chemical mechanism. This preprocessing step both negatively impacts the user experience and software maintenance costs. These problems can potentially be overcome using a just-in-time (JIT) compilation approach. The JIT compilation can build the necessary chemistry solver at runtime. This means that once a researcher changes the chemical mechanism, the JITed solver is able to generate the appropriate chemistry solver automatically.. However, several open questions remain: (1) Does the JITed code yield a competitive performance against the one generated by the “preprocessor”? (2) Is the JITed code portable between different platforms such as CPU and Graphics Processing Units (GPU)? Addressing these concerns is likely to greatly enhance the attraction and adoption of the JIT compilation in this research community.

The goal of this 2023 summer internship is to develop a GPU version of an existing JIT based chemistry solver written in C++. The student’s primary focus will be developing the JITed chemistry solver for GPU and documenting the procedures, success or known issues. The student will also run various chemistry solvers on different linux clusters at NCAR to verify their accuracy and evaluate their performance.

Students: The project is open to graduate students.

Skills and Qualifications: Familiarity with Linux and C++ is required. Strong motivation to learn new skills and resolve issues in a team is required. Experience with either JIT compilation or GPU programming is desirable.

Project 11. Interactive Visualizations of Climate Data

Areas of Interest in order of relevance: Visualization, Data Science, Geostatistics

Description: The recent advancements in both climate modeling and modern computational infrastructures enable us to run climate simulations at higher resolutions. Effective visualizations of climate model outputs and climate data can help communicate climate change issues to the general public so that these issues can be easily understood by non–climate experts. Furthermore, advanced and interactive visualizations enable climate scientists to detect patterns, time-evolving features, and trends in complex datasets and model outputs that might not be obvious from looking at the raw data alone.

In this project, we are going to create a user-friendly dashboard for reading large climate data and visualizing these dataset in an easy to understand and intuitive way. Next, we are going to host this application on a commercial platform and study the performance of Xarray/Dask for reading large volumes of climate datasets on different architectures. These data visualizations dashboards will be used to communicate scientific findings to domain experts, policy makers, and the general public.

Over the summer, you will have opportunities to: (1) Collaborate with a team of research scientists, data scientists, and software developers to produce publicly-accessible interactive dashboards that leverage Xarray/Dask, scientific python stack, and interactive visualization libraries such as Bokeh and Holoviews for visualization and analysis of climate data. (2) Learn about and contribute to open source geoscientific Python projects and open source tools. (3) Gain experience to effectively access and read large datasets on commercial platforms. (4) Gain hands-on experience with version control software for collaborative software development via Git and GitHub. (5) Investigate the performance of GPU-native analytics with Xarray.
We welcome applications from any intern interested in creating dashboards for cloud-hosted, geospatial data and will tailor the project and specific learning objectives to the intern's interests and experience.

Students: The project is open to graduate students.

Skills and Qualifications: - Experience with Python programming - Xarray - Pandas - Numpy
- Experience with interactive visualization tools such as Python Bokeh or experience with Javascript libraries for advance visualization such as D3.js
- Experience with Linux environment , and version control with Github
- Ability and willingness to work within a multidisciplinary team
- Good communication and writing skills.
- Familiarity with running applications on commercial clouds.

Highly desirable skills:
- Experience with Dask parallelization in Python
- Experience with GPU programming would be a plus.
- Experience working with atmospheric and geosciences datasets would be a plus

Project 12. Interactive Visualization of Uncertainty using Data Assimilation and High-Resolution Earth System Ensembles

Areas of Interest in order of relevance: Visualization, Software Engineering, Software Engineering

Description: Quantifying uncertainty through analysis of ensemble forecasts in numerical weather prediction or ocean modeling remains a challenge. The massive number of observations in addition to the large dimension of atmospheric and ocean models makes it difficult to properly assess the quality and uncertainty of the prediction. Consequently, deriving risk measures and informative solutions may become strenuous and borderline impossible.

Ensemble data assimilation (DA) provides a flexible ensemble framework to estimate the state of an earth system. The Data Assimilation Research Testbed (DART) at NCAR [dart.ucar.edu] is a sequential ensemble DA tool that has been heavily tested and used for weather forecasting, ocean prediction, climate projections, flood prediction, etc. The main goal of the project is to interface the ensemble visualization package OVIS with DART. OVIS [ovis.thomashollt.com] is an interactive visualization framework that allows for an efficient and easy analysis of ocean forecasts and their uncertainties. By utilizing data on the fly, OVIS can help enable the users to dive into the data, change parameters, select subsets of the ensemble, and instantly visualize the results. Various risk measures can be also computed based on the statistics of the ensemble. While DART is written in Fortran, OVIS is implemented in Objective C and OpenGL. The project also entails extending the scope of OVIS to support atmospheric and other earth system models. This will be conducted while potentially exploring CISL's own VAPOR [vapor.ucar.edu] for depicting uncertainty. Overall, the objective is to make it possible for DART's large userbase to have access to a state-of-the-art diagnostic package that is modern, highly informative, and easy to use.

Overall, the project offers the opportunity for interested applicants to work on exciting real-world problems while collaborating with experts in science fields such as computer science, geosciences, and data assimilation. Working with an advanced visualization software package in addition to DART will help applicants gain extensive experience in uncertainty analysis, visualization, and ensemble forecasting.

Students: The project is open to graduate students.

Skills and Qualifications: - Graduate program such as computer science, geoscience, or related field
- Good oral and written communication skills
- Ability to read, write and understand code
- Open to work with people from diverse backgrounds
- Ability to work in a team environment

Project 13. Explore Native C++ Parallelization Abilities on an Compression Algorithm

Areas of Interest in order of relevance: Application Optimization/Parallelization, Software Engineering

Description: Modern C++ (17 and 20) has brought great parallelization capabilities to the language itself. These capabilities include parallel algorithms in C++17 and coroutines in C++20, among others. Though showing great potentials, these new language features have not seen wide use in high-performance computing (HPC). This SIParCS project aims to explore the viability of applying modern C++ parallelization features to HPC programs. Specifically, this project investigates parallelization opportunities in a scientific data compressor, SPERR, and how to exploit them using modern C++ language features.

Students: The project is open to graduate students.

Skills and Qualifications: Strong C++ programming skills, some parallelization experience.

Project 14. Creating bias corrected global climate model output to support regional climate research

Areas of Interest in order of relevance: Software Engineering, Digital Asset Management, Numerical Methods

Description: Are you interested in supporting scientific research to assess the regional and local impact of predicted climate change?

The goal of this project is to work with a NCAR scientist to update and improve existing software to create a new version of an existing global bias-corrected climate dataset that is built from the NCAR Community Earth System Model (CESM) output. The previous version of this dataset was produced under phase 5 of the Coupled Model Intercomparison Project (CMIP5), and the new updated dataset produced in this internship will be built from CESM output as part of the newer CMIP6 initiative and which supports the Intergovernmental Panel on Climate Change Sixth Assessment Report (IPCC AR6). Since all global climate models contain regional scale biases due to insufficient spatial resolution and limited representation of some physical processes, it is common to bias correct the climate model output before using it as input to regional scale models.

This new dataset will provide a valuable community resource for regional climate researchers to run numerical simulation experiments based on the most recent climate model predictions, and use the results to determine the expected regional and local impacts from a range of future climate change scenarios. Additionally, the intern will gain experience in data curation, management, and preservation by archiving the new dataset in the NCAR Research Data Archive.

The successful candidate will:
1. Become familiar with:
   -The format and data structure of CESM CMIP6 output that is stored on the NCAR HPC system
   - Existing software that was previously used to create the previous version of this dataset
   - The description and structure of the previous version of this dataset
2. Update and improve the existing software and workflows needed to produce a new version of the bias corrected dataset from CESM CMIP6 model output.
3. Preserve the updated software in a public GitHub repository to support reuse by others
4. Run the updated workflow and software on CISL's HPC Data Analysis System
5. Verify that the output dataset meets format and content expectations
6. Create and curate a new data collection in the NCAR Research Data Archive to host the updated dataset for use by the research community
7. Present the project results at the conclusion of the internship

Students: The project is open to graduate students.

Skills and Qualifications: Desired Skills and Qualifications

1. Working skill in Python and compiled languages such as C, C++, or Fortran
2. Familiarity with scientific Python libraries, such as Xarray and NumPy
3. Experience working with large geoscience gridded datasets, such as the ECMWF ERA5, in NetCDF format
4. User-level familiarity with the Linux operating system
5. Familiarity with git and Github software collaboration tools
6. Good oral and written communication skills
7. Ability and willingness to work in a team environment

Project 15. Continuous Integration for ASAP applications (CPU and GPU applications)

Areas of Interest in order of relevance: Software Engineering, System Integration, Containers, DevOps, Verification and Reproducibility

Description: An Open Source and community driven Weather and climate modeling code receive regular updates from scientists and software engineers all over the world. These are often related to science, portability, and performance updates. These updates are often submitted as a Pull Request to the repository and the admins/owners of the repository usually perform the code review for syntax, standard, verification, code norms, etc. before accepting the Pull Request. This process is even more time consuming when you have to check using multiple hardware architectures, compilers, and systems.

Code updates, changes in the software stack, compilers, and even changes in the system introduce bugs regularly. It is very important to identify these bugs before committing the changes to the code. Continuous integration is an application development practice that integrates tests, syntax checks, etc. into source code changes as part of a CI/CD pipeline or DevOps process. This project aims to develop a process to make use of CI tools to develop automated code testing for an ASAP application on both CPU and GPU.

Students: The project is open to graduate students.

Skills and Qualifications: Skills and Qualifications:
Experience with Python programming or any scripting language
Experience with Github actions or any other CI/CD framework
Experience with Linux environment
Experience with cloud computing platforms
Ability and willingness to work within a multidisciplinary team
Good communication and writing skills

Highly desirable skills:
Software documentation
Experience with Docker or any other container technology
Knowledge of GPU computing
Knowledge of Parallel programming

Project 16. Containerization of simulation applications for frequently re-run

Areas of Interest in order of relevance: Software Engineering, Application Optimization/Parallelization

Description: Container technology is rapidly developing, and we wonder if it can help to reduce data storage. What the containers can do is that we run the scientific simulation applications on containers with the required OS, software stacks, configurations, initialization/grid files, and input data on current supercomputers. Then we rebuild containers with one version older MPI (Message Passing Interface) or OS to check if it still can run successfully. In this way, we can validate the re-run capability of containers in the long term. If this workflow works, we can use this way to re-run the old simulation whenever we need it. Thus when scientists try to save large amounts of data from simulations, we can urge them to use this container strategy to re-run their applications at any time and only save some small amount of data instead.

The project will focus on using singularity containers to automate the compilation process of a scientific simulation CM1 with various MPI versions. Basically, students will first manually build CM1, then create container images to automatically build it and test it on Casper, furthermore change MPI versions to see if the container images will still be able to run on Casper. The final step will confirm the re-run capability of the containers.

Students: The project is open to graduate students.

Skills and Qualifications: Familiarity with high performance computing (HPC) cluster, linux environment and make/cmake is required. Strong motivation to learn new skills and resolve issues in a team is required. Experience with container runtimes is preferred.

Non-Technical Project:
CISL Outreach, Diversity, and Education (CODE) Intern

Areas of Interest in order of relevance: Higher Education Administration / Student Affairs, Science Education, Education Policy, Social Work, or related program.

Description: *This is a paid student intern position.* As the CODE Intern, you will provide administrative support to the SIParCS Program office and affiliated programs. You will also assist with planning and preparation for education and outreach programs to occur during the 2023 -2024 school year. This is a full-time (40 hours/week) student intern position that runs from May 15 - August 4, 2023. End date may be flexible.

Responsibilities
Student intern support:

- Be an active participant on the SIParCS team to provide support and mentoring for students.
- Live-in at the suite-style apartments with the interns, and plan and participate in after-hours team building activities.
- Keep program leadership informed of any issues that arise.
- May assist students/participants with special needs.
- May travel to assist with intern recruitment during fall months.
- Attend the Rocky Mountain Advanced Computing Consortium (RMACC) with the SIParCS program.

Summer program logistics:

- Assist with summer program support including planning and running events. Events include orientation, professional development workshops, field trips, and other learning opportunities for interns.
- Assist with apartment move-in and move-out logistics, distributing and collecting student supplies, and coordinating with apartment administration.
- Help write and edit SIParCS Annual Report.
- Update SIParCS program alumni tracking documents for program assessment and evaluation purposes.

General administrative support:

- Maintain program databases and updating web pages.
- Edit and compile information for documents, posters, brochures and newsletters.
- Provide multimedia support at program events (photos, video) and keep inventory of program pictures.
- Provide meeting support and prepare meeting materials.

School-year program support:

- Update presentation slides and other educational materials for school-year programs.

- May write and or edit educational materials such as text updates for HPC-TV and/or Meteo-AR.
- May test and evaluate new educational materials.
- May provide input to the new exhibit space at NWSC
- The above statements describe the primary work being performed. Additional tasks and professional development opportunities may be assigned according to the intern's skill set and interests.

Students: The project is open to graduate students in Higher Education Administration / Student Affairs, Science Education, Education Policy, Social Work, or related program.

Skills and Qualifications:
What you need:

- Must have attained a bachelor's degree at the time of application.
- Must be enrolled in a graduate level college, university or accredited professional program during the normal school year.
- Ability to work 40 hours/week from May 15 - August 4, 2023 (dates to be confirmed). End date may be flexible.
    - NCAR has 3 unpaid holidays for interns during the summer internship (Memorial Day, Juneteenth, 4th of July).
- Experience with word processing, database, and spreadsheet applications in a Windows environment (Microsoft Word, PowerPoint, Excel, Access) as well as a Google Apps environment (Sheets, Docs, etc).
- Proven ability to organize, prioritize, and follow through on multiple tasks, with close attention to detail.
- Good written and verbal communication skills with the ability to convey information to interns, visitors, and partners in a welcoming and professional manner.
- Interest and sensitivity in working in diverse settings that include interns from a broad spectrum of underrepresented groups.
- Proven ability to work in a team and individually. A strong work ethic.
- Patience and adaptability.
- Discretion in handling confidential information.
- On-site participation in Boulder.

Preferred:

- Experience working with undergraduate and graduate students.
- Currently enrolled in masters level program in: Higher Education Administration / Student Affairs, Science Education, Education Policy, Social Work, or related program.

The above qualifications describe the ideal skill set we look for in candidates. We encourage you to apply even if you do not possess all of the listed qualifications.

More information:
https://www2.cisl.ucar.edu/siparcs/cisl-outreach-diversity-and-education-code-intern