HIGH PERFORMANCE COMPUTING MODERNIZATION PROGRAM

RESEARCH PROJECT #: HPCMP-HIP-24-001

HPC-Enabled Data Processing and Analysis Pipeline

About DEVCOM ARL:

The U.S. Army Research Laboratory Army Research Directorate, DoD Supercomputing Resource Center (DSRC) conducts research critical to the Army's assured land power dominance into the deep future. We work with emerging computational platforms and architectures, advanced algorithms, data-intensive analysis workflows, and immersive visualization tools, among other areas. Our facilities include high performance computers, emerging processors, development platforms, and a mixed-reality visualization lab.

RESEARCH LOCATION: Aberdeen Proving Ground, MD

PROJECT DESCRIPTION:

This project investigates the development of a distributed processing and analysis pipeline for a user to efficiently analyze test data. This stems from the need for users to explore large amounts of heterogeneous data, leveraging HPC resources to provide an enduring capability for data ingestion, exploration, analytics, and reporting.

The intern will enhance the capabilities of the tool stack by becoming familiar with the test data domain, conducting data exploration, determining a set of appropriate data elements and analytic techniques, developing explainable machine learning models, generalizing those models, containerizing services, prototyping various configurations for scaling up the analytical workflows on the HPC systems, and evaluating the approaches for effectiveness and performance. The intern will evaluate both model performance and runtime performance and suggest recommendations for future directions with model design and tools deployment.

This internship will provide the intern with exposure to a relevant Army use case for utilizing this technology, and the intern will gain first-hand experience performing research and development for the Army. This project will provide the intern the opportunity to experiment with state-of-the-art data analysis, processing, warehousing, and governance tools on HPC resources. As part of the summer intern experience at the ARL DSRC, students participate in a 2-day tour of the ARL Laboratories at Aberdeen Proving Ground and at Adelphi Center. Summer interns at the ARL DSRC are placed in a collaborative workspace, providing opportunities to interact, collaborate, and share their experiences. Mentors interact with students regularly in this workspace, and hence students learn from other mentors as well as their own.

The intern will be in a unique position to engage with the ARL DSRC Data Science team and the Data Analysis and Assessment Center (DAAC) team, along with ATEC testers and evaluators through collaboration.

ANTICIPATED START DATE:

June 2024 – Exact start dates will be determined at the time of selection and in coordination with the selected candidate.

QUALIFICATIONS:

Prospective candidates must at least have taken introductory CS courses and know how to program. Experience with Python, Linux, Data Science, and Machine Learning is preferred.

ACADEMIC LEVEL:

Degree received within the last 60 months or currently pursuing:

- Bachelor's
- Master's

DISCIPLINE NEEDED:

- Computer, Information, and Data Science
- Engineering
- Mathematics and Statistics
- Science & Engineering-related