CS25-310 Advancing High-Performance Computing

Project Proposal

Prepared for

Alberto Cano

VCU College of Engineering

By

Yunus Bidav, Steven Holcombe, James Jenkins, Amaka Odidika

Under the supervision of

Alberto Cano, Carlisle Childress, Romano Woodfolk

11 Oct 2024

Executive Summary

This document presents a comprehensive proposal for the senior capstone project spearheaded by the High Performance Research Computing (HPRC) center at Virginia Commonwealth University (VCU). Central to this initiative is the establishment and configuration of a state-of-the-art high-performance computing cluster, equipped with updated software and hardware. This cluster's setup will then be evaluated for its potential integration into the existing production clusters operated by the HPRC.

In this proposal, which forms the first major segment of the final report, several key sections outline the foundational planning stages of the project. The Problem Statement section provides essential background information and articulates the overarching purpose of the project, establishing a clear context for the proposed work. Following this, the Engineering Design Requirements section delineates the general goals, specific design objectives, technical specifications, and the relevant codes and standards that will guide our efforts. Further, Section C elaborates on the Scope of Work, presenting more concrete indicators of progress, such as detailed deliverables and critical milestones. This clarity will facilitate effective tracking of our advancements throughout the project lifecycle. Additionally, the proposal includes Appendices 1 and 2, which feature a Gantt chart timeline outlining project phases and a team contract that underscores our collective commitments and responsibilities. Please note that sections that fall outside the scope of this proposal are intentionally left blank and marked as such.

Table of Contents

Section A. Problem Statement

High Performance Computing, as described by IBM, uses clusters of machines that work in parallel to process large data sets and solve complex problems at much higher speeds than other methods [1]. This is in contrast to mainframes, which are large-scale, high-performance single computers. Large HPC clusters are often referred to as supercomputers. Virginia Commonwealth University's High Performance Research Computing Center runs two production HPC clusters, Athena and Apollo [2].

Many researchers at VCU use the HPRC's resources to do their intensive research, using compute power for research such as rendering frames, statistical support, DNA, RNA and proteomics pipelines, and fluid and molecular dynamics [3]. The HPRC has been continually operating and innovating since 2006 [4], and this year is no exception. The purpose of our advancement project of the HPRC is to help integrate and showcase the usefulness and power of the resources we have to students of VCU. The project at hand is a unique opportunity and an open ended one as well. We are given the opportunity to build a cluster from the ground up from hardware configuration to software setup. Our client is VCU and the students/users of athena and the other varying resources that the HPRC project provides.

We are led by Alberto Cano and Carlisle Childress. They have been and will continue to be instrumental in our efforts for the capstone project. Alberto Cano is our faculty advisor and our sponsor. He has been our first point of contact during the beginning of the project and helped lay the foundation for what is to come throughout the year. Carlisle Childress is our acting technical advisor and manager. Carlisle is meeting with our group weekly, showing us current updates for Athena, and continuing the setup and timeline for our fall winter semester goals.

We aim to install HPL, a ranking system for compute clusters to see where our system and configuration compares against other clusters around the globe. Another software we are to install and set up for ease of use is OpenHPC. This will allow a web interface for users who may not be familiar with the intricacies of connecting to a cluster and sending a compute workload using the terminal or other specialized software. This specific piece of the project is one that we aim to have fully operational come springtime for our demos. Most of the computer science programs at VCU have not used athena or may not even be familiar with what athena is. Installing OpenHPC will be a great first way to interact with the compute power at hand for first time users. We also will be using Rocky Linux 9, currently the cluster we are using for the project and other clusters are on an older version. OS migrations can be tricky and finicky at scale such as the one the HPRC center is on. We are focusing on aiding in helping this migration and ensuring it goes smoothly.

Previous HPRC capstones have followed similar guidelines as we have been given with the exception of implementing AI workloads for use with open HPC. This is a clear distinction and one that provides the challenge of running LLM's on a cluster for workloads that were not designed to run on such compute structure. Besides the above challenge we are presented with most of the guidelines and standards the previous teams before us have used we will be using and referencing them along the way to ensure we are on track and doing the work correctly and efficiently.

Section B. Engineering Design Requirements

B.1 Project Goals (i.e. Client Needs)

The project is aimed to have the following goals achieved. These objectives should be completed and ready for display before the capstone project exposition.

- Working cluster hardware
- Software stack being used
- List of software integrated with OpenHPC/Slurm
- Ranked performance of the HPC
- Access to OpenHPC for the project expo for student use
- Integration with LLMs and AI workloads
- Educate the student body about Athena and its abilities

B.2 Design Objectives

- The design will be capable of running a standard HPL benchmark.
- The design will be capable of being accessed remotely from within the VCU Network.
- The design will support scalability, allowing for future expansion of compute resources.
- The design will demonstrate improved performance over the existing HPRC clusters for specific benchmark tests.
- The design will implement effective resource allocation and scheduling through Slurm.
- The design will support a variety of scientific and research applications commonly used by VCU researchers.

B.3 Design Specifications and Constraints

- The design must perform communication between compute nodes using IEEE 802.3-compliant Ethernet cabling using the Message Passing Interface
- The design must consist of hardware mountable in a standard 19" wide rack space.
- The design must not require the use of software that is not free and open source.
- The design must make use of Slurm and Warewulf to provision resources.
- The design must incorporate other guidelines provided by OpenHPC's documentation.

B.4 Codes and Standards

When working in a HPC environment the following standards should be applied to ensure security, performance, modularity, and maintenance. These standards are applicable to all hardware and software setups we may use to integrate the above requirements and specifications.

- NIST SP 800-223 - the design and access of HPC environments are advised to adhere to the best practices listed in the standard's write up
- IEEE 802.3 - Ethernet standard which is used throughout the hardware connections between hardware racks
- MPI 4.1 - Message Passing Interface, the method of communication between the nodes of the cluster

- IEEE P2416 - Resource management standards for HPC workloads and service management
- NIST SP 800-53 - Provides best practices for security and privacy controls for cloud computing in general
- IEEE 802.1 - Standards for general local access networks and network management for servers on a hardware scale
- IEEE 2937-2022 - Benchmarking standards for intensive workloads related to AI workloads

Section C. Scope of Work

C.1 Deliverables

Below are all of the deliverables we are expected to produce for our project. All of the non-academic related deliverables were discussed with the team through various in person meetings with our technical manager, Carlisle Childress. These are to be completed in full in an almost sequential manner as the project itself builds together one step after the other (ignoring the academic deliverables).

In order of completion date the deliverables are listed:

- Team contract
- GitHub Repo Setup
- Project proposal
- Preliminary Design Report
- The first POC cluster is brought up and accessible
- Configure and set up Slurm
- HPL Ranking
- Fall Poster
- Open HPC integration completed
- Final Design Report
- Install interactive LLMs
- Fully working cluster with all software installed
    - Accessible through OpenHPC
- Capstone Expo poster
- Capstone Expo presentation

C.2 Milestones

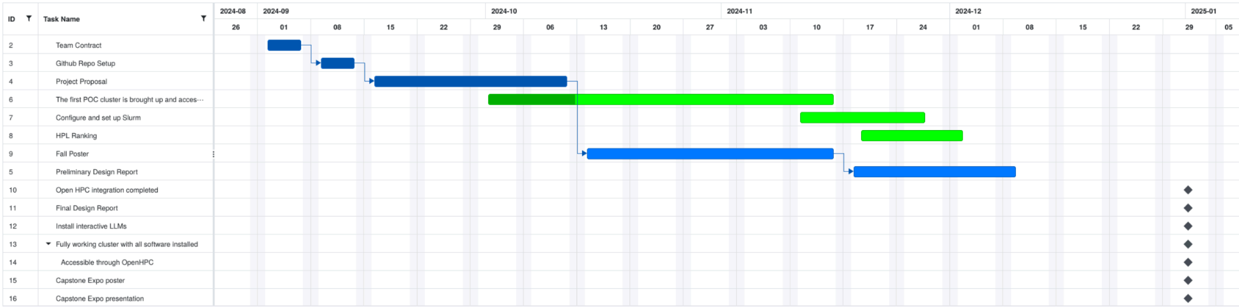
- Project Initiation (Completed)
    - Initial meetings with advisors
    - Familiarization with project scope and HPC environment
- Hardware Setup (In Progress)
    - Set up demo rack hardware
- Operating System Installation
    - Install Rocky Linux 8 on a test cluster
    - Migrate to Rocky Linux 9
- Software Implementation
    - Install and configure OpenHPC
    - Set up Slurm workload manager
    - Integrate Message Passing Interface (MPI)
- Performance Benchmarking
    - Install HPL (High-Performance Linpack)

        - Rank cluster performance
- User Interface Development
    - Implement OpenHPC web interface
- Documentation and Testing
    - Develop user guides and documentation
- Final Deliverables
    - Prepare the final design report
    - Create capstone expo poster and presentation
- Project Closure
    - The final demonstration of working cluster
    - Handover to HPRC for potential production integration

C.3 Resources

Hardware is far and away the most integral and potentially expensive piece of this project, and fortunately a robust inventory is already available at the HPRC center. Specifically, several servers, which are currently not integrated into the Athena or Apollo clusters, are poised to serve as the head and compute nodes for the capstone cluster. This pre-existing infrastructure will significantly streamline our efforts and enhance our operational capabilities. Furthermore, proficiency in Linux systems administration will be paramount for the successful setup and configuration of the cluster. This expertise will ensure that we can effectively manage the hardware resources, optimize performance, and facilitate a seamless integration of the necessary software tools. Lastly, all efforts will be made to use only free and open source software, or at least software with no cost. This means that this project will proceed without the need for funding from the capstone project budget, relying entirely on the resources and knowledge already present at the HPRC center.

Appendix 1: Project Timeline



Appendix 2: Team Contract (i.e. Team Organization)

Contents

Step 1: Get to Know One Another. Gather Basic Information.

Team Member Name	***Strengths each member bring to the group***	***Other Info***	***Contact Info***
Yunus Bidav	Team collaboration, project management, creative problem solving, and technical skills	Experience with linux on server environments. Experience managing users/clients. Experience with neural networks and interpretable ML models. In depth knowledge of discrete math and its applications in computer science. ACO *(Algorithms, Combinatorics, and Optimizations)* Enthusiast.	Bidavye@vcu.edu 929-330-7737
Steven Holcombe	Experience with linux systems and reliability and maintenance of large multi tenant applications hosted on aws and azure (for my job). I like to learn more about most topics in computer science so I am always up for a challenge.	Reachable by discord I never check my email. I work at my job MWF 1-5 so I may be unresponsive at those times but for the most part I am highly available. Familiar with AWS, myriad of SAP products and I am pretty proficient with github. I would be able to learn more about clustering and setup/maintenance of the software side of things.	holcombes@vcu.edu 571-224-4308
James Jenkins	Enthusiasm for and experience with Linux and server operations, more free time than I ought have.	Reachable by email or Discord, will likely not respond from 10pm-7am.	jenkinsjb2@vcu.edu (540) 621-6313
Amaka Odidika	The ability to be open-minded and learn from others, computer science has also taught me to pay attention to details.	Sometimes actual text message works better for me depending on the times, usually after 9pm. I also like organization and agendas.	odidikaa@vcu.edu 240-917-6910

Other Stakeholders	*Notes*	*Contact Info*
Alberto Cano	*Faculty Advisor*	acano@vcu.edu
Carlisle Childress	*Project Advisor*	cgchildr@vcu.edu
Romano Woodfolk	*Project Advisor*	rmwoodfolk@vcu.edu
James Davis	*Project Technical Advisor*	jmdavis1@vcu.edu

Step 2: Team Culture. Clarify the Group's Purpose and Culture Goals.

Culture Goals	*Actions*	*Warning Signs*
Submit deliverables before the night they are due.	- Set internal deadlines - Communicate frequently progress and expectations	- Unresponsiveness from team members - Infrequent progress updates on approaching deadlines
Making the effort to inform the team members if you will be unable to attend a meeting	- Communicate clearly what meeting dates the member will not be able to attend - Making an effort to modify the meeting time so all members may be present	- Missing meetings with little to no notice - Not making an effort to modify the meeting within a reasonable window of time before the meeting - Repeated failure to attend meetings with no notice will be brought up with advisor
Every member should do their part pertaining to their role(s) and responsibilities	- Every member will achieve their goals in a timely manner in order to complete all milestones on time - Making sure other members are offered help incase others are falling behind	- Parts of the project assigned to a member are incomplete or missing - If there are multiple missing parts pertaining to a member the advisor will be informed

Each team member should provide support while also holding one another accountable.	- If a team member is struggling there is an effort to assist or guide. - Ensuring each team member does their part and if otherwise accountability is taken to ensure proper support can be given.	- Responsibilities are not completed and there are no reasons given

Step 3: Time Commitments, Meeting Structure, and Communication

Meeting Participants	***Frequency Dates and Times / Locations***	***Meeting Goals Responsible Party***
Students Only	Whenever needed within discord chat or voice if members are available for a short voice chat.	Simple day to day responsibility updates sent in the updates text channel of our discord. *Managed by Steven Holcombe*
Students Only	Voice call through Discord, in-person if necessary. Thursdays 5-6pm	Set responsibilities and goals for the week, provide updates to team on progress and setbacks, work on project. *James will manage the flow of the call while Steven will scribe.*
Students + Faculty advisor	Weekly Meeting with Faculty Advisor 6-6:45pm Thursdays	Weekly check in for questions on work being performed or next steps as needed. *Steven Will scribe and Yunus will manage the meeting times and changes*

Step 4: Determine Individual Roles and Responsibilities

Team Member	***Role(s)***	***Responsibilities***
Steven Holcombe	*Reports Manager*	Take meeting notes, Update github weekly with status reports.
Yunus Bidav	*Logistics Manager*	Handle meeting times. Follow up on commitments.
Amaka Odidika	*Financial Manager*	Research technical purchases and justifications for the allocated budget .
James Jenkins	*Project Manager*	Coordinate communication with Faculty Advisors/Sponsor, plan meetings and establish goals for the project. Submit deliverables.

Step 5: Agree to the above team contract

Team Member:	*Steven Holcombe*	*Signature:* ______________________
Team Member:	*James Jenkins*	*Signature:* ______________________
Team Member:	*Yunus Bidav*	*Signature:* ______________________
Team Member:	*Amaka Odidika*	*Signature:* ______________________

References

[1] Susnjara, Susan. (2024, July 9). *What is High-Performance Computing (HPC)?*. Retrieved October 1, 2024.

[2] VCU Research and Innovation. (2022). *HPRC Services*. Retrieved October 1, 2024. research.vcu.edu/cores/hprc/facilities

[3] VCU Tech Services. (2022, April 26). *High Performance Computing: IT Catalog: VCU*. Retrieved September 29, 2024. catalog.ts.vcu.edu/service/high-performance-computing

[4] VCU Research and Innovation. (2022). *HPRC Services*. Retrieved September 29, 2024. research.vcu.edu/cores/hprc/services