

# Samuel Mucyo

sammucyo@college.harvard.edu  
sam-mucyo.github.io

## OBJECTIVE

---

Software engineer with experience in distributed systems and high-performance computing. Demonstrated track record of building scalable solutions at Amazon across multiple teams, with particular expertise in backend development and cloud architecture. Strong academic foundation in computer science with interest and focus in data systems and parallel computing.

## EDUCATION

---

**Harvard University**, Cambridge, MA

Graduation Date: May 2025

Bachelor of Arts in Computer Science

*Relevant Coursework:* High-Performance Computing (CS2050), Data Systems (CS1650), Systems Programming & Machine Organization (CS61), Data Structures & Algorithms (CS1240), Systems Development (CS107), Machine Learning (CS1810), Data Science I (CS109A)

**Teaching Fellow, Introduction to Computer Science (CS50)**

*Harvard University*

Fall 2022 and Fall 2023

- Conducted detailed code reviews for over 40 student projects and weekly problem sets through grading, providing feedback on code quality, memory management, and algorithm efficiency
- Mentored students during 6 weekly office hours, focusing on debugging issues across multiple programming languages: C, Python, and JavaScript
- Led weekly 2-hour hands-on lab sessions, covering topics such as introduction to data structures & algorithms in C, and web development with Flask, HTML, CSS, and JavaScript

## EXPERIENCE

---

**Amazon**

*Seattle, WA*

*Software Development Engineer Intern at Amazon.com*

May 2024 - August 2024

- Created a reusable A/B testing framework for dynamic profile badges using Java/Spring MVC, reducing similar A/B experiment deployment time from weeks to days
- Improved API usage and accuracy by identifying and resolving misuse of the internal A/B API, which caused excessive triggers and signal noise; Conducted a statistical analysis to support the redesign

*Software Development Engineer Intern at Amazon.com*

May 2023 - August 2023

- Architected APIs for a rule-based risk monitoring and alarming system to reduce fraudulent activity detection time by working collaboratively with software engineers, data engineers, and investigators
- Developed a new Python-based backend with AWS CDK, enabling rapid and consistent deployments; leveraged AWS Lambda and DynamoDB for scalability and resource utilization

*Software Development Engineer Intern at AWS - Redshift*

June 2022 - August 2022

- Designed a serverless data lake using Python, AWS S3, and Glue to enhance bottleneck detection in Redshift infrastructure testing
- Automated data pipeline integration with internal visualization tool to provide detailed insights, enabling fast, data-driven decisions

## PROJECTS

---

**Optimized Column-Store Database System**

*Individual Project: Data Systems (CS1650)*

Fall 2024

- Developed a high-performance columnar database engine for efficient data retrieval and analysis (select-project-join) with optimized storage techniques, including B-trees for indexing and memory-mapped files for persistence
- Optimized query execution and concurrency by implementing multi-threaded scan operators, achieving a top 3 placement in class benchmarks for index operation and skewed data handling

- Implemented parallel batch select query execution using POSIX threads in C, with cache-conscious design and minimal data movement

### **Parallelizing Urban Transit Construction with Minimum Spanning Trees**

*Group Project: High Performance Computing for Science and Engineering (CS2050)* Spring 2024

- Collaborated in a team of five to implement parallel version of Kruskal's algorithm using OpenMP and OpenMPI for distributed computing, for final project
- Conducted comprehensive performance analysis using PAPI for hardware counter measurements
- Performed memory profiling using Valgrind's massif tool to optimize resource utilization
- Automated strong and weak scaling analysis through parameterized bash scripts

### **Astrolibrary**

*Group Project: Systems Development for Computational Science (CS107)* Fall 2023

- Followed the Software Engineering Development Life Cycle to deliver a Python library for astronomical spectral analysis
- Collaboratively designed API contracts, ensuring alignment with project requirements
- Developed and tested core functionality, supplemented with a comprehensive test suite using *pytest* for validation and documentation using *Sphinx*, enabling ease of use and future maintainability
- Integrated an automated CI/CD pipeline for seamless builds, testing, and deployment using GitHub Actions
- Coordinated with team members through regular meetings, version control, and code reviews

## **SKILLS**

---

- **HPC & Systems Programming:** C, C++, POSIX threads, OpenMP, MPI, GDB, Valgrind
- **Cloud & Backend Development:** AWS (Lambda, S3, DynamoDB, CDK, Glue, Redshift), Java/Spring MVC, Python
- **Data Engineering:** SQL, database design, data pipelines, ETL, AWS Glue
- **Web Development:** JavaScript, React, Flask, HTML, CSS
- **Development Tools:** Git, Linux/Unix environments, CI/CD (GitHub Actions), Bash
- **Data Science & ML:** R, TensorFlow, Pandas, statistical analysis