

Caleb Carlson

inf0rmatiker.github.io
github.com/inf0rmatiker

ccarlson355@gmail.com

(512) 934-3355

linkedin.com/in/inf0rmatiker

SUMMARY

Ambitious and motivated Software Engineer with solid leadership and teamwork skills, with over 5 years' experience building robust solutions in industry. Always eager to learn new methods and technologies while promoting sustainable development practices and minimizing technical debt. Passionate about open-source, community-driven projects and sharing creativity through contributing to public codespaces.

INDUSTRY EXPERIENCE

• Hewlett Packard Enterprise

HPC Storage - Systems Software Engineer II

Fort Collins, CO

June 2022 – Present

- Implemented a cluster fabric configuration API and address pool manager with accompanying CLI commands
- Created an automated, user-friendly, and scalable software upgrade procedure for DAOS storage cluster
- Designed and implemented cluster bootstrapping process for DAOS storage product, using HPCM as a foundation
- Triaged and repaired Kubernetes management issues in customers' legacy storage products
- Contributed bugfixes and LNet kernel module tuning improvements to community Lustre filesystem project
- Benchmarked HPC Lustre filesystems against GPU Direct Storage and POSIX clients
- Created Jenkins pipeline that builds DKMS-enabled Lustre clients for a variety of distros and kernels
- Investigated Talos OS as platform for new storage product and enabled InfiniBand device passthrough
- Streamlined add-on data mover node discovery and software mapping process

• Cray, Inc.

HPC Storage - Software Development Intern

Longmont, CO

May 2020 – May 2022

- Automated Jenkins builds, Go unit testing, and Slack notifications for team's repositories
- Built Jenkins pipelines, Go unit tests, and Slack notifications for team's repositories
- Improved gRPC communication strategy between data path microservices
- Automated Lustre filesystem provisioning based on discovered NVMe/SCSI drives
- Added redundant resource persistence via both filesystem and Kubernetes' etcd keystore

• Data Ductus, Inc.

Software Development Intern

Longmont, CO

May 2019 – Aug 2019

- Implemented administrative RBAC for Verizon APIs
- Added audit log entries to API calls for user/action tracking
- Reduced end-to-end testing time from over 20 minutes to 6 minutes

EDUCATION AND RESEARCH

• Colorado State University

M.S. in Computer Science; GPA: 4.000

Fort Collins, CO

Aug 2020 – May 2022

• Colorado State University

B.S. in Computer Science, Math Minor; GPA: 3.984

Aug 2016 – May 2020

• Graduate Research Assistant

Pallickara Lab

Jan 2020 - May 2022

- Published two papers to ACM and IEEE conferences, see *Projects*
- Automated management of 150+ node clusters with Kubernetes and Docker
- Established procedures for ingesting and sharding petabyte-sized datasets into MongoDB
- Built on-request data analytics services and contributed to open-source, NSF-funded Urban Sustain project
- Helped develop LSTM recurrent neural networks for point-cloud datasets generated from agricultural LIDAR equipment

- **Teaching Assistantships**

Aug 2017 – Dec 2020

Intro to Operating Systems

- Taught recitations for over 70 students, publishing help videos and documentation for topics like thread safety, virtual memory, paging, and sockets programming
- Designed containerization-focused term project for the class with Kubernetes and Docker

Intro to Software Engineering

- Practiced Agile and Scrum master role, assisting teams with Sprint planning; automated DevOps
- Provided reference React/JSX web applet with RESTful API for students to test against
- Taught version control and Git project etiquette, clean code practices, REST API implementation with Java, and MySQL database creation/maintenance

Data Structures and Algorithms

- As lead TA, helped manage schedules and onboard new TAs
- Proctored exams and implemented automated grading/assignment submission system
- Facilitated learning inheritance, polymorphism, data structures, object-oriented design, recursion, and various algorithm implementations

PROJECTS

- **Omniscient** Distributed resource monitoring for memory pressure, CPU usage, storage I/O, Ethernet and InfiniBand throughput. [github.com/inf0rmatiker/omniscient]
- **Validation Service** Validates spatiotemporal models on large datasets in a distributed fashion and visualizes metrics as a geospatial heatmap. [github.com/Project-Sustain/validation-service]
- **Storm Topics** Detects most popular topics from live twitter message streams using the lossy counting algorithm with Apache Storm and Zookeeper. [github.com/inf0rmatiker/stormtopics]
- **Docusum** Finds sentences that best summarize a Wikipedia document with Hadoop MapReduce. [github.com/inf0rmatiker/docusum]
- **Model Service** A service providing federated model training for spatially-segregated data. [github.com/inf0rmatiker/model-service]
- **Research Papers**
 - ACM BDCAT 2021 — *Distributed Orchestration of Regression Models Over Administrative Boundaries*
 - IEEE ICCBE 2022 — *Resource Efficient Profiling of Spatial Variability in Performance of Regression Models*

SKILLS

- **Languages:** Go, Bash, C, C++, Python, Java, Scala
- **Technologies:** Docker, Kubernetes, Helm, Slurm, Jenkins, Git, GitHub, Jira, Confluence, Apache Spark, Apache Storm, Hadoop, Nginx, Flask, Numpy, Pandas, Anaconda, gRPC, RESTful APIs, PyTorch, Scikit-Learn, Spring Boot, Matplotlib, Jupyter Notebook, Sockets/RDMA, Linux (RHEL/SLES/Ubuntu), RPMs, Autotools