Kaleb Coberly

📍 Bellingham, WA, USA | ✉ kaleb.coberly@gmail.com | 🌐 kalebcoberly.com

Summary

Data engineer, backend software engineer, quality engineer with a background in research and analysis. Builds, tests, and deploys tools, analytics, and pipelines that help others reliably use data better. Supports users and developers with documentation, dev tools, issue triage, hands-on training, and debugging. Improves systems with a focus on clarity and maintainability. Prioritizes data quality, reproducibility, validity, and meaningful results.

Skills

- *Programming & Scripting:* Python, R, Bash, SQL, NoSQL, C++.
- *DevOps & Tooling:* Git, Make, Docker, GitHub Actions, Jenkins, CI/CD, Python packaging, debugging, QE and testing frameworks.
- *Data Engineering:* ETL pipelines, RDBMS, database design, data modeling, data cleaning, data validation, methodological implementation, data versioning, MySQL, MariaDB, MongoDB, ElasticSearch, HTML/XML/JSON/YAML/CSV.
- *ML Engineering:* ML pipelines, feature engineering, performance evaluation, model validation.
- *Analysis & Visualization:* EDA, Jupyter Notebook, RStudio, R Markdown, Mermaid, Excel, Kibana, Tableau.
- *Documentation & Support:* Sphinx, reStructuredText, user training, contributor onboarding, issue triage.
- *Platforms & Environments:* MacOS, Windows, Linux, HPC (Slurm, UGE), conda, Kubernetes (k8s), GitHub, Bitbucket/Stash, VS Code, JIRA, Asana.

Experience

Research Engineer, IHME – Central Computation (GBD), Nov 2022–Oct 2024

- www.healthdata.org/research-analysis/gbd
- Developed and maintained internal Python and dev tools for data modeling, versioning, access, and dev workflows.
- Contributed to a large in-house Python ecosystem: features, bugfixes, major test coverage increases, and documentation.
- Maintained CI/CD pipelines with Jenkins builds and integrated testing, monitored build health across ~100 packages.

- Debugged issues across DBs, Python APIs/CLIs, Docker, k8s, and HPC clusters.
- Directly supported research teams: Investigated failed model runs, data access issues, and tooling gaps.
- Analyzed system usage and tool performance via ElasticSearch; ran ad-hoc reports and built automated reports.

Data Analyst, IHME – Pandemics Team, Nov 2021–Dec 2022
- Owned multiple data pipelines for COVID-19 production; reduced key runtime by 65%.
- Refactored and extended vaccine and booster modeling pipeline.
- Built diagnostics, with automated reporting for data QA/QC/QE; inspected and cleaned data.
- Led early UGE-to-Slurm transition for production jobs.
- Built ETL pipeline in R of national data into crosswalked, versioned output.
- Produced executive reports for world leaders, running ETL into R Markdown for PDFs; adapted to changing model outputs and feature requests:
  - www.healthdata.org/research-analysis/diseases-injuries/covid/policy-briefings

Advisory Board Member and Citizen Science Volunteer Engineer, Cascade STEAM, April 2025–Present
- cascadesteam.org
- Leading the Citizen Science group's development of computer-vision ETL tool for stormwater monitoring field observation datasheets:
  - github.com/crickets-and-comb/stormwater_monitoring_datasheet_extraction
- Integrating DevOps tools and workflows with Cascade STEAM infrastructure.
- Consulting on selection of collaborative and project-management platforms.

Founder and Lead Engineer, Crickets and Comb, Nov 2024–Present
- cricketsandcomb.org
- Implemented and maintaining reusable GitHub Actions CI/CD pipelines and dev flows with automated testing and QE for local + Docker development and open-source Python package releases.
- Onboard developers by using documentation, presentations, and hands-on support of testing conventions, CI steps, and Git Flow.
- Automated part of Bellingham Food Bank's delivery operations, reducing several staff-hours with each weekly use:
  - Designed and implemented an ETL CLI integrated with staff workflows and 3rd-party API.
  - Achieved high test coverage, by line count, problem space, and abstraction level.
  - Led incremental design, rollout, and user training.
  - Wrote and deployed technical and non-technical documentation.
  - Maintain ongoing user support and dependency management.
  - Released as open-source project; lead collaborative development at all skill levels.

Education

B.S., Database Management and Data Analytics – Western Governors University

B.A., Critical Studies and Pedagogy – The Evergreen State College

Certifications

- Quantum Computing (Qiskit & Python) – Udemy, Apr 2025
- Mathematics for ML & Data Science – DeepLearning.AI, Jan 2025
- Oracle Database SQL Certified Associate – Oracle, Jul 2020
- CompTIA A+, Network+, Project+ – CompTIA, 2019 / 2020

References

Available upon request.