

PAVEL YAKOVLEV

yvpavel@gmail.com
github.com/paulyakovlev

EXPERIENCE

General Motors

February 2021 - Present

Data Engineer, Connected Vehicle Data Engineering (April 2025 - Present)

Austin, TX

- Developed and configured automated data collection for connected vehicles, enabling real-time monitoring of embedded vehicle systems across production fleets.
- Conducted vehicle testing and validation in sandbox environments to ensure data collection accuracy before production deployment.
- Implemented ETL data pipelines for vehicle configuration metadata management.

Software Engineer, Big Data Engineering & Infrastructure (February 2021 - April 2025)

- Led migration from Cloudera Data Analytics Studio to Tez UI, eliminating closed-source dependencies and reducing licensing costs.
- Responded to Log4Shell zero-day vulnerability by developing automated remediation script that scanned and patched production Hadoop clusters, then integrated solution into standard cluster startup process to ensure ongoing protection.
- Contributed to Apache Ambari and Apache Bigtop open-source projects by implementing Livy service integration and service management (Ambari PR#3790, Bigtop PR#1282).
- Integrated open-source Hadoop components (Apache Ambari, Livy, Tez UI) into GM's custom Hadoop platform, collaborating with platform engineers to migrate from Cloudera Hadoop Platform and reduce annual infrastructure costs by \$1.2M.
- Developed Oozie ETL pipeline for Ranger audit logs to Parquet format, enabling PowerBI analytics across production clusters.
- Enhanced Docker-based Hadoop development environment with local RPM repos and auto-build + deployment of Ambari, significantly improving setup time for developers.

PROJECTS

Long Marine Stranding Map

<https://github.com/lmlstrandingnetwork/lml-stranding-map>

- Built full-stack geospatial web application for UCSC Long Marine Lab scientists to digitize field documentation of marine mammal strandings, replacing manual paper forms with structured data collection including location tracking, photo uploads, and filterable search capabilities.
- Implemented interactive mapping enabling scientists to filter stranding records by species, year, age, and location with real-time visualization.
- Established separate dev/prod environments with CI/CD automation and conducted code reviews to maintain code quality throughout development.
- Led regular stakeholder meetings with lab technicians to gather requirements and iterate on features based on feedback.
- Mentored next cohort of engineering students on codebase, development practices, and technical implementation.
- Technologies: React, Node.js, Mapbox, Algolia, Firebase, Heroku, Git

EDUCATION

University of California, Santa Cruz

Class of 2020

Bachelor's Degree in Computer Science

SKILLS

Languages: Python, Java, Bash, SQL, Javascript

DevOps and Automation: Docker, CI/CD Pipelines, deployment automation, configuration management, scripting

Build Tools: Maven, Gradle, Git, RPM, Make