PAVEL YAKOVLEV

yvpavel@gmail.com · github.com/paulyakovlev

EXPERIENCE

General Motors April 2025 - Present
Data Engineer, Connected Vehicle Data Engineering *Austin, TX*

- Implemented connected vehicle telemetry collection by configuring data elements and event triggers, deploying to production fleets and validating accuracy using SQL, Kafka, and Databricks
- Resolved production data collection failures for rental fleet partner (100,000+ vehicles) by implementing direct read method to bypass queue bottleneck, ensuring accurate fuel level and odometer telemetry
- Developed Python ETL pipeline migrating CSV/XML configuration files from Git repository into Databricks, enabling PowerBI analytics on production configurations

General Motors February 2021 - April 2025
Software Engineer, Big Data Engineering & Infrastructure *Austin, TX*

- Developed GM's custom Hadoop distribution (GMDP) by integrating open-source components including Tez UI, Ambari, and Livy, migrating from Cloudera and reducing annual costs by $1.2M
- Contributed to Apache Ambari and Bigtop open-source projects by implementing Livy service integration (Ambari PR#3790, Bigtop PR#1282)
- Automated Docker-based Hadoop development environment, reducing developer setup time from 35+ minutes to 10 minutes by eliminating manual build and deployment steps
- Responded to Log4Shell vulnerability by developing automated remediation script for production Hadoop clusters, integrating into standard startup process for ongoing protection
- Built automated Oozie ETL pipeline converting cluster audit logs to Parquet and ingesting into Hive tables, enabling PowerBI analytics for cloud migration planning

PROJECTS

Long Marine Lab Stranding Map https://github.com/lmlstrandingnetwork/lml-stranding-map

- Built full-stack geospatial web application with Node.js/Express REST API backend and React frontend, enabling scientists to digitize marine mammal stranding records with location tracking, photo uploads, and filterable search
- Established separate dev/prod environments with CI/CD automation and implemented Git-based code review process to maintain code quality throughout development
- Mentored next cohort of engineering students on codebase, development practices, and technical implementation

EDUCATION

University of California, Santa Cruz 2020
Bachelor's Degree in Computer Science

SKILLS

Languages: Python, Java, Bash, SQL, JavaScript

Backend & Data: Node.js, Databricks, Azure, Oracle, Hadoop, PowerBI, CI/CD

Development Tools: Git, Docker, Maven, Gradle, RPM, Linux