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.
- Develop 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.

Lawrence Livermore National Lab June 2019 - September 2019

Software Test Engineering Intern *Livermore, CA*

- Improved automated testing by contbuting to JUnit test suite and providing technical documentation

remote.it May 2018 - June 2019

Software Engineering Intern *Palo Alto, CA*

- Prototyped user-end web portals for device monitoring and management
- Wrote a library of user-end Python and Shell scripts to manage IoT edge devices
- Provided technical documentation for internal and end-user software tools

EDUCATION

University of California, Santa Cruz Class of 2020

Bachelor's Degree in Computer Science

SKILLS

Languages: Python, Java, Bash, SQL

DevOps and Automation: Docker, CICD Pipelines, deployment automation, configuraiton management, scripting

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