REGEX-BASED LOG ANALYZER

GROUP 11

Project Overview

The Regex-Based Log Analyzer is a Python-based tool designed to ingest server or application logs and extract insights using configurable regular expressions. Our primary goal was to build a Minimum Viable Product (MVP) within one week as part of the Advanced Python Capstone Project. The tool enables users to upload a log file, apply regex filters, and view summary statistics such as counts of errors, warnings, and top requested endpoints. We implemented this project in Python 3.10 using pipenv for dependency management and pytest for testing.

System Design / Implementation

The application consists of three main components:

1. Log Parser: Reads the uploaded log file and applies user-defined regex patterns.
2. Regex Config Module: Validates and compiles regex patterns to handle malformed inputs gracefully.
3. Dashboard UI: Displays the extracted insights, including counts of errors/warnings and the most frequently requested endpoints.

We also created two pytest tests to ensure the core parsing logic works reliably. The repository includes a README with setup instructions, a Pipfile, and a link to our demo video.

Challenges Faced

Our biggest challenges were handling diverse log formats and testing complex regex patterns. We also learned to coordinate commits across multiple contributors using GitHub and to use meaningful commit messages for clear project history.

Ideas for Future Work

Potential improvements include real-time log tailing, saving regex presets for repeated use, and adding interactive charts or graphs for better visualization. Exporting results to CSV or PDF formats would also enhance usability.

Conclusion

This project helped us strengthen our Python skills, especially in regular expressions, testing, and collaborative development. The MVP demonstrates how regex can quickly turn raw logs into actionable insights.