Web Enabling Analytics Workflows: Automated Core Logging

Samuel Bradley

CSIRO, Kensington, WA, Australia, sam.bradley@csiro.au

ABSTRACT

As a Software Engineer in the research sector I am often being presented with a lot of code written by researchers. Much of this code is executing resource intensive processes and is targeted towards people with high levels of domain knowledge. A motived researcher may even create complex codebases with a number of classes and lots of state, which cannot feasibly be re written or rearchitected. It may also be necessary or beneficial for a researcher to be able to continue contributing to the code used by any downstream solutions.

Stateful analytics workflows present a lot of challenges for web development, particularly in a world that has moved heavily towards light weight, scalable, "microservices". The challenge extends beyond architecture, the workflow also needs to be made user friendly, responsive and scalable to moderate numbers of simultaneous users.

During the development of Data Mosaic, a web application for automated drill core logging I have been able to overcome all these challenges in order to create a user-friendly, web enabled workflow where the core scientific code is still being developed by the primary researcher (Dr. June Hill, CSIRO).

Researcher contributions were able to be maintained through frequent contact between Dr. Hill and myself, isolation of the core code from the API, education on the software tools (git & vscode) and a continuous integration pipeline to automatically package and deploy changes.

I have also been able to achieve a stateless backend through the use of fast serialisation of the stateful classes, this ensures the application can be scaled through the use of a multi-worker web server and containers deployed in a compute cluster.

DESCRIPTION

My lighting talk will focus on challenges and solutions for web application development in the Research Domain using the development of the automated core logger, Data Mosaic, as a case study. I believe this fits in well with the overall eResearch theme.