

15-716-954

Date: 28.04.2022

Name: Michael Ziörjen

Keywords: Scaling / Load-Balancing of Services, Build-time improvements, Performance improvements

Overall Progress: 4

Activities completed:

- US5: Setup and configured nginx as a load-balancer running on a separate docker container. Scaled both annotation and prediction services to use two instances. Research on how to automate (re-)starting docker containers for next sprint
- Improved build stability and time for annotation-service
- Improved performance of services by disabling linux emulation on docker (was previously required because of M1 MacBook processor architecture)
- Project Management for current Sprint
- US12: Improved code-fetching tool and added command line options
- Added simple performance measurements to web-api and code-fetcher
- Added simple load-testing tool to test different workloads on our API
- Research potential deployment options on Azure

Current activities:

- Investigate python language bug (empty python files)

Goals or Activities for next deliverable:

- Perform thorough performance analysis and identify potential bottlenecks
- Implement switching of model without downtime (potentially by orchestrating the restarting of the prediction service in docker)
- Document the annotation service (Code, Endpoints, Architecture)
- Improve test coverage of annotation service

Immediate attention: -

