

ROB: Benchmarking on the cloud

Ajay Rawat (University of Washington) [Speaker]

Heiko Mueller (New York University)

Shih-Chieh Hsu (University of Washington)

Reproducible Open Benchmark:

A platform developed to help evaluate data analysis workflows in a controlled competition-style environment designed to:

- Reduce the amount of time required to organize and evaluate such benchmarks
- Ensure reproducibility of benchmark results.

REANA

To achieve reproducibility of benchmark results we integrate ROB with the Reproducible research data analysis platform (REANA).

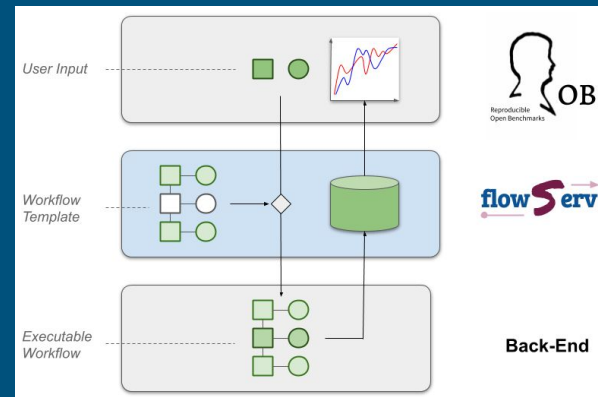
REANA allows users to run their containerised workflows and archive the results on a remote serve¹.

Using REANA as the execution backend for ROB makes it easier to archive a workflow run and reproduce the results since REANA archives all the required files and workflow logs for reproducing the results on the cloud. By using REANA, we also achieve a significant scalability improvement for ROB over the previously used native workflow engine.

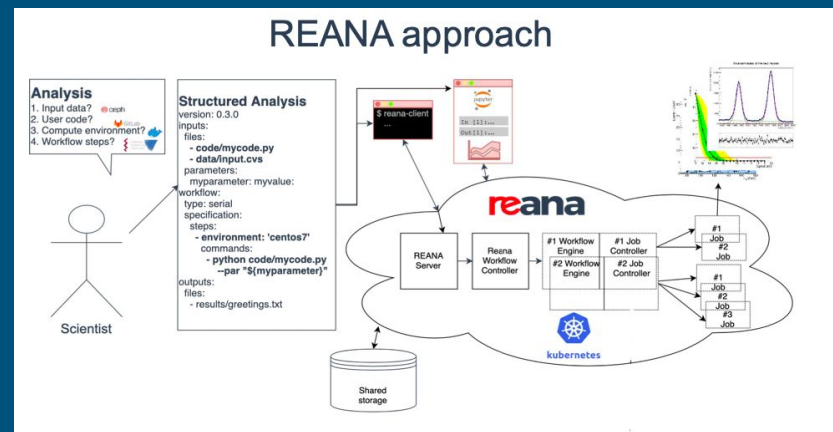
References:

[1] Scallin. "Reproducible and Reusable Data Analysis Workflow Server." GitHub, github.com/scallin/flowserv-core

[2] REANA Poster <http://cds.cern.ch/record/2695179/files/Poster-2019-942.pdf>



ROB framework [1]



Running workflows on REANA [2]