



Contribution ID: 332

Type: Talk

## Benchmarking massively-parallel Analysis Grand Challenge workflows using Snakemake and REANA

Thursday 24 October 2024 17:45 (18 minutes)

We have created a Snakemake computational analysis workflow corresponding to the IRIS-HEP Analysis Grand Challenge (AGC) example studying ttbar production channels in the CMS open data. We describe the extensions to the AGC pipeline that allowed porting of the notebook-based analysis to Snakemake. We discuss the applicability of the Snakemake multi-cascading paradigm for running massively-parallel RECAST-compatible physics analysis workflows where the analysis process may run over numerous independent data samples with large number of independent individual data files in a fully concurrent manner. The created Snakemake workflow example was run on the REANA reproducible analysis platform. We describe the improvements brought to the REANA job scheduling, tracking and termination processes for massively-parallel Snakemake workflows. We present results of several numerical experiments running the same workflow on the Kubernetes cluster with increasing number of identical nodes. We infer on the feasibility of REANA to schedule numerous concurrent jobs from the same Snakemake workflow rule, study the importance of cluster node size from the point of view of the job memory requirements, as well as estimate the overhead of dispatching workload to many cluster nodes. The results demonstrate the applicability of Snakemake for even massively-parallel RECAST-compatible physics analysis workflows.

**Primary authors:** DONADONI, Marco (CERN); FEICKERT, Matthew (University of Wisconsin Madison (US)); HELD, Alexander (University of Wisconsin Madison (US)); POVSTEN, Andrii (Princeton University (US)); SHADURA, Oksana (University of Nebraska Lincoln (US)); SIMKO, Tibor (CERN)

**Presenter:** DONADONI, Marco (CERN)

**Session Classification:** Parallel (Track 9)

**Track Classification:** Track 9 - Analysis facilities and interactive computing