

Dynamic hybrid workflows for DL on HPC infrastructure

Iacopo Colonnelli¹

¹*Università degli Studi di Torino, Italy*

Hybrid workflow abstractions [1] allow users to quickly design and orchestrate cross-facility workloads, decoupling tasks from environment-specific technical details to reduce complexity and increase reusability. Plus, workflow descriptions help ensure the reproducibility of scientific experiments through prospective and retrospective provenance collection.

This module has been designed to provide a hands-on exploration of scientific workflows from various angles, from the initial design phase to their orchestration at extreme scales. We will use the practical example of the Common Workflow Language (CWL) open standard [2] to demonstrate how workflows can be written, and the StreamFlow workflow system [3] to execute them seamlessly on the CINECA HPC facility. We will also delve into the integration between scientific workflows and Jupyter Notebooks [4], which aims to give data scientists a familiar interface to scientific workflows.

In this module, students will gain a comprehensive understanding of scientific workflows. They will learn how to use these workflows to model and orchestrate Machine Learning and Deep Learning pipelines. Additionally, they will explore how modern workflow management systems can efficiently scale data-oriented workloads from a researcher's laptop to an entire HPC facility.

- [1] I. Colonnelli, "Workflow models for heterogeneous distributed systems," in *Proceedings of the 2nd Italian Conference on Big Data and Data Science (ITADATA 2023), Naples, Italy, September 11-13, 2023* (N. Bena, B. D. Martino, A. Maratea, A. Sperduti, E. D. Nardo, A. Ciaramella, R. Montella, and C. A. Ardagna, eds.), vol. 3606 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2023.
- [2] M. R. Crusoe, S. Abeln, A. Iosup, P. Amstutz, J. Chilton, N. Tijanic, H. Ménager, S. Soiland-Reyes, and C. A. Goble, "Methods included: Standardizing computational reuse and portability with the Common Workflow Language," *Communication of the ACM*, 2022.
- [3] I. Colonnelli, B. Cantalupo, I. Merelli, and M. Aldinucci, "Streamflow: cross-breeding cloud with hpc," *IEEE Transactions on Emerging Topics in Computing*, vol. 9, no. 4, p. 1723–1737, 2021.
- [4] I. Colonnelli, M. Aldinucci, B. Cantalupo, L. Padovani, S. Rabellino, C. Spampinato, R. Morelli, R. D. Carlo, N. Magini, and C. Cavazzoni, "Distributed workflows with jupyter," *Future Generation Computer Systems*, vol. 128, p. 282–298, 2022.