ClimateDT Workflow: a containerized climate workflow

Francesc Roura-Adserias[1] (francesc.roura@bsc.es), Aina Gaya-Avila[1], Leo Arriola i Meikle[1], Iker Gonzalez-Yeregui[1], Bruno De Paula Kinoshita[1], Jaan Tollander de Balsch[2] , Ivan Alsina Ferrer[1] and Miguel Castrillo[1]

[1] Barcelona Supercomputing Center (BSC), Earth Sciences, Barcelona, Catalonia, Spain

[2] CSC – IT Center for Science, Espoo, Finland

DESTINATION EARTH

Climate Adaptation Digital Twin (DE340)

Earth Sciences Department

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Overview

- The Climate Adaptation Digital Twin aims to provide **operational climate projections** and sector-relevant climate information **globally at an unprecedented temporal and spatial resolution** [1].
- **The ClimateDT workflow** is the tool that is used to **orchestrate all its components**. It is a highly modular software that allows multiple configurations (Figure 1)[2] and enables running climate simulations and impact sectors synchronously by using **data streaming** [3].
- The ClimateDT workflow has to support **operational simulations in non-operational environments**, in different EuroHPC machines around Europe.

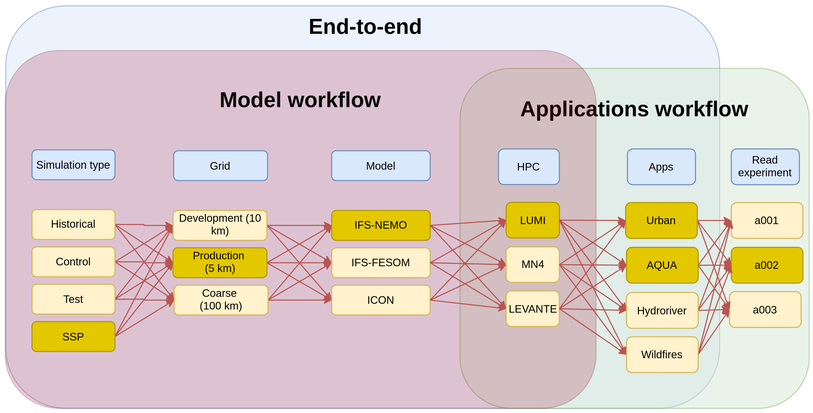

Figure 1: combinations that the great modularity of the workflow allows.

What have we gained?

- The workflow and their components are now much more **portable** and its **logic complexity has been** greatly **reduced**.
- We have achieved **seamless execution** of components across different HPCs.
- We obtain a clear **traceability** of application source code and their needed environment.
- Easy deployment in offline machines
- We have **decoupled the deployment of the components** (which can be done more directly by the component developers) **from the workflow development**.
- This solution enables the inclusion of an arbitrary number of components by avoiding conflicts between software dependencies, thereby improving **scalability**.

Issues

Issues to be solved

- The workflow (and its components) **has to be easily portable**. We have to deal with an environment that should be used for all the components, without having **system administration permissions**.
- We also have to support the installation of some components that need an **internet connection** (e.g. *pip* (pypi.org) installable Python applications), which is not available in all HPCs.

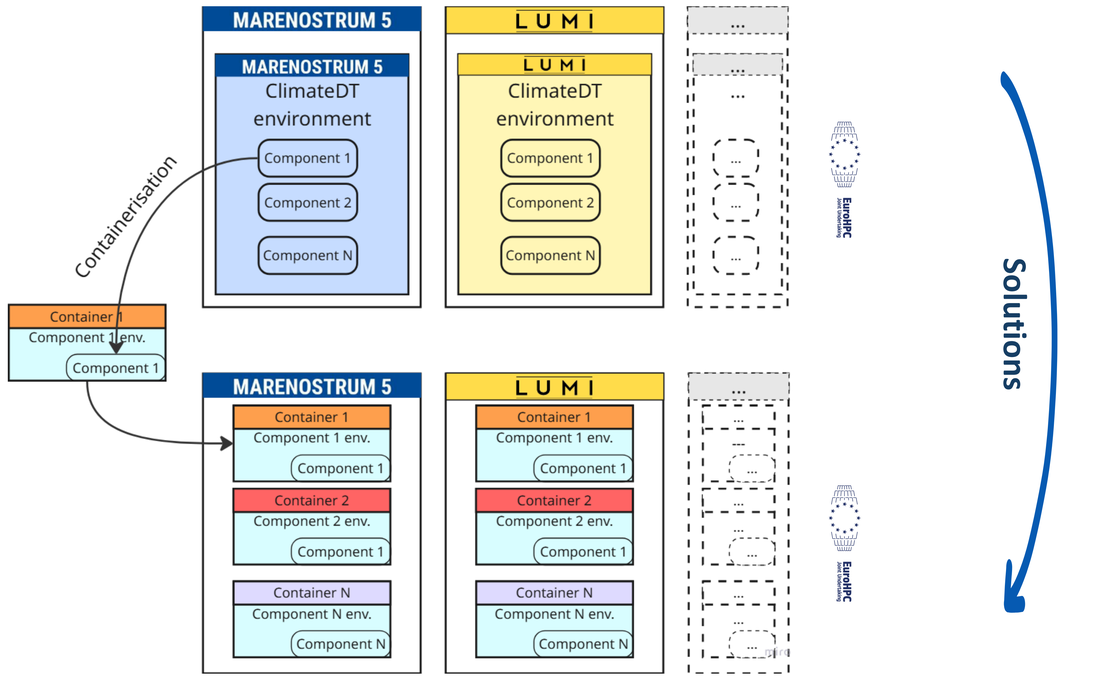

Figure 2: Deployment of containers in the different HPCs.

Solutions

Link to abstract

How do we solve them?

- **Containerize** all the needed software (including dependencies)(Figure 2). We use Docker to build the container and then we convert them to ***Singularity*** (sylabs.io/singularity/) images. Singularity containers can be run in HPC machines without administrator rights.
- We link (bind) the container to the minimum set of variables and directories in the HPC that the component needs.
- **Automatic** deployment (build and uploading to HPCs)

Findings

References

1. Hoffmann, J., Bauer, P., Sandu, I., Wedi, N., Geenen, T., and Thiemert, D.: Destination Earth – A digital twin in support of climate services, Climate Services, 30, 100394, https://doi.org/10.1016/j.cliser.2023.100394, 2023.

2. Gaya-Àvila, A., Arriola i Meikle, L., Roura-Adserias, F., De Paula Kinoshita, B., Beltrán Mora, D., Ahmed, R., Andrés-Martínez, M., and Castrillo, M.: A workflow for the Climate Digital Twin, EGU General Assembly 2024, Vienna, Austria, 14–19 Apr 2024, EGU24-2533, https://doi.org/10.5194/egusphere-egu24-2533, 2024.

3. Roura-Adserias, F., Gaya i Avila, A., Arriola i Mikele, L., Andrés-Martínez, M., Beltran Mora, D., Gonzalez Yeregui, I., Grayson, K., De Paula Kinoshita, B., Ahmed, R., Lacima-Nadolnik, A., and Castrillo, M.: The data streaming in the Climate Adaptation Digital Twin: a fundamental piece to transform climate data into climate information, EGU General Assembly 2024, Vienna, Austria, 14–19 Apr 2024, EGU24-2164, https://doi.org/10.5194/egusphere-egu24-2164, 2024.

Acknowledgements

Destination Earth is a European Union funded initiative launched in 2022, with the aim to build a digital replica of the Earth system by 2030. The initiative will be jointly implemented by three entrusted entities: the European Centre for Medium-Range Weather Forecasts (ECMWF) responsible for the creation of the first two 'digital twins' and the 'Digital Twin Engine', the European Space Agency (ESA) responsible for building the 'Core Service Platform', and the European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT), responsible for the creation of the 'Data Lake'.

Funded by the European Union Destination Earth implemented by ECMWF esa EUMETSAT