eFlows4HPC

Enabling dynamic and Intelligent workflows in the future EuroHPC ecosystem

D7.1 Training Plan

Version 1.0

Documentation Information

Contract Number	9555558
Project Website	www.eFlows4HPC.eu
Contractual Deadline	30.06.2021
Dissemination Level	PU
Nature	R
Author	Nikoleta Kiapidou (BSC)
Contributors	Renata Giménez (BSC), Rosa M. Badia (BSC), Stéphane Zeng (Atos), Gianluigi Rozza (SISSA), Donatello Elia (CMCC)
Reviewer	Riccardo Rossi (CIMNE)
Keywords	Training, workshops, hackathon

This project has received funding from the European High-Performance Computing Joint Undertaking (JU) under grant agreement No 955558. The JU receives support from the European Union's Horizon 2020 research and innovation programme and Spain, Germany, France, Italy, Poland, Switzerland, Norway.

Change Log

Version	Description Change
V0.1	First draft
V0.2	Comments by contributors
V0.3	Second draft
V0.4	Comments by contributors
V0.5	Third draft sent to internal reviewer
V1.0	Final version

Table of Contents

1. Executive Summary 3
2. Introduction 3
3. Training goals and target audiences 3
4. Training timeline 4
5. Training courses 4
5.1. Hackathon 5
5.2. SiSSA Summer School 6
5.3. Additional plans and resources 6
6. Key Performance Indicators 7

List of Tables & Figures

Table 1: eFlows4HPC trainings overview 4

Figure 1: Group image from SISSA Summer School 2019 6

Table 2: eFlows4HPC training KPIs 7

1. Executive Summary

A sustained, high-quality training program is important to guarantee that the eFlows4HPC project will obtain the expected results as part of the training and education strategy. This document describes the training-related activities.

2. Introduction

WP7 includes task 7.1 which aims at organising training courses focusing on the eFlows4HPC users and developers.

This document presents an initial training plan for eFlows4HPC project that will be updated during the project's lifetime, if needed. It includes the training goals and timeline, as well as detailed information on the specific training activities that the eFlows4HPC partners will organise during the project.

Apart from D7.1, the D7.4 Report of the organization of community workshops, due at the end of the project, will report on the facilitation and outcomes of the external workshops, which are described in the D8.1 Communication and dissemination plan.

3. Training goals and target audiences

The main aim of the eFlows4HPC training activities is to educate internal users and developers (and when possible, external) in relationship with the main technological components of the project and the three Pillars. In particular, the overall training goals are:

- To properly demonstrate the project solutions internally among all partners, including HPC users and developers
- To promote the efficient use of platforms and architectures best suited for the targeted use cases and applications to the communities linked to the three Pillars (manufacturing, climate, and urgent computing for natural hazards)
- To transfer the project methodologies and results to the relevant CoEs and communities, contributing to the reduction of skill gaps related to HPC in Europe

The target audiences of the trainings are defined according to the internal and external nature of the courses, as follows:

- Internal users and developers within the eFlows4HPC project
- Users from the three project Pillars that will employ the eFlows4HPC technologies in their related domains:
    - Manufacturing, Climate, and Urgent computing for natural hazards
- External users, software engineers, and developers (in need of efficient workflows), who:
    - Use parallel programming models combining High Performance Computing (HPC), Data Analytics (DA), and Machine Learning technologies for e-science applications
    - Use computational resources, software environments and programming models for execution of large-scale e-science applications, i.e., with the objective of generating predictions of real processes, such as weather forecasting or protein interaction modelling

4. Training timeline

A description of the training activities of eFlows4HPC is described in this section, including an overview of each training course, their potential date, place and format (online/physical), their type (internal/external), and target audience.

The training activities overview can be seen in Table 1, while a detailed analysis of each course can be found in the following part.

Table 1: eFlows4HPC trainings overview

Training	Date & Place	Format	Type	Target audience
Initial training: Basic technologies	Several sessions across multiple months, starting in M6 (June 2021)	Online	Internal	Internal users and developers
Second training: eFlows4HPC platform (first version)	M17 (May/June 2022)	Online	Internal	Internal users and developers
Hackathon across all Pillars	M17 (May 2022)	Online or physical (in Lecce, Italy)	External	Users from the three Pillars
Final training: eFlows4HPC software (second version)	M30 (September 2023)	Online or physical (in Grenoble, France)	Internal (+potentially external attendees TBC)	Internal users and developers, Pillar users (+ potentially external users, software engineers and developers)

5. Training courses

As shown in Table 1, several training courses will be organized during the eFlows4HPC lifecycle. The idea is to organise a first training course at the beginning of the project (First training: Basic technologies), a second training course (eFlows4HPC platform, first version) along with a hackathon across all Pillars with the software stack release 1, and a final training course (eFlows4HPC platform, second version) for software stack release 2. In principal, these trainings will be internal, targeted to the developers and users of the eFlows4HPC consortium and related users from the three Pillars. However, on a one-to-one basis, and particularly for the hackathon and the final training, it will be considered to open them to a wider audience of external users and developers who would benefit from an efficient workflow system to combine multiple technologies for e-science applications.

More specifically:

1. An **initial internal training** course on the eFlows4HPC basic technologies, addressed to internal users and developers, will be planned during the first phase of the project.

Given the pandemic situation, this first internal training has been converted into multiple online trainings that will be recorded and made available to all project participants. This initial training aims to enhance the partners' skills in in the tools available for the workflow implementation. It will be divided in multiple parts, covering aspects of the workflow tools (PyCOMPSs, Ystia orchestrator), machine learning libraries (dislib, HeAT, EDDL), Data Analysis (Ophidia, ParSoDA, DMCF) and others.

2. A **second training course and a hackathon across all Pillars**, addressed to internal users, developers and Pillar experts, is planned along with the first software release towards the end of the second phase of the project around M17. These courses will focus on the use of the first version of the eFlows4HPC platform for the development of the workflows. As the Pillars will perform integration and development activities for the first version of the Pillar workflows, these trainings will be useful to educate the Pillar experts. A detailed account of the hackathon can be found in section 5.1 below.
3. **A final training course** is planned for M30 when the second software release 2 will have been completed. As with the training of the first version of the eFlows4HPC platform, this is addressed to the projects' internal users as well as Pillars' and domain experts who will be educated on the second version of the design and implementation of the eFlows4HPC software stack interfaces. The training could potentially be opened to external audiences that include users, software engineers, and developers in need of efficient workflows, as explained in section 3 above.

The first training will be held as multiple virtual meetings, while the others could be held as physical ones in Lecce, Italy and Grenoble, France respectively, depending on the developments of the Covid-19 pandemic, travelling restrictions, and practical considerations. BSC has the tools and platforms needed to host online training courses and will use them to offer the support needed to the speakers and attendees. The technical experts from the consortium will also help to identify the content and speakers of each training, validate the agenda, and deliver the actual training courses.

5.1. Hackathon

The hackathon is planned to be co-located within the second training in M17 (May 2022) either online or physically at the premises of CMCC in Lecce, Italy. The main organisers of the hackathon will be CMCC and BSC, who will arrange the content of the training with the technical experts, offer logistical support, and also promote the course to external audiences. This course will have a longer duration than the normal trainings (possibly one or two full day course) and a practical approach in order to train the attendees in a hands-on environment.

Apart from Pillar-related experts and eFlows4HPC internal users and developers, the hackathon is open to external participants including selected stakeholders and domain-related communities. The trainers will be experts from WP1 (Workflow interfaces for the integration of HPC, data analytics, and machine learning), WP2 (Optimization of runtime and libraries, workflow deployment, resource management, and data management), and WP3 (Co-design aspects between applications, software stack and actual hardware), who will separate the attendees in small working groups and will closely instruct them with real time tutoring and problem-solving.

As a reinforcing training activity of the eFlows4HPC software release 1 at the end of the second project phase, the hackathon's content will focus on educating the trainees on the integration and development activities of the first version of the Pillar workflows, including the quality metrics and their evaluation, as previously defined during the first project phase. The Pillar experts, as trainees,

will have the opportunity to be educated on the eFlows4HPC software workflows, practise on their usage, and solve any queries with the help of hands-on examples and use cases. They will also be able to provide feedback to WP1, WP2, and WP3, so that the latter can update their requirements for the second software release.

5.2. SiSSA Summer School

The eFlows4HPC partner SISSA organises a week-long thematic summer school in Trieste, Italy, addressed to computer science students: SISSA Summer School. Although this initiative was put on hold during the Covid-19 pandemic, there are plans to re-initiate it in the following years. After the release of the first version of the eFlows4HPC platform (meaning after M17, May 2022), the SISSA summer school plans to include hands-on sessions on the eFlows4HPC technologies and help to spread the developed methodologies to a broader audience of users and developers, especially of a younger age. The target audience will be master and Ph.D. students as well as early-stage researchers working in the field of reduced-order models. The participation of industry research scientists is also encouraged.

Figure 1: Group image from SISSA Summer School 2019

5.3. Additional plans and resources

Additional efforts will be made in order to enhance the training experience of the eFlows4HPC members and fully support their educational process. These efforts will include the following:

- A dedicated trainings resources folder in the internal repository (B2DROP hosted by BSC) of the project where all training resources and tools will be gathered and deposited for all partners to access at any time.
- In collaboration with WP8 Dissemination and Communication, all external trainings will be promoted via the eflows4HPC and partner channels to engage with the potential target audience is addressed to. All eflows4HPC partners will also disseminate them among the partners' communication channels.
- Existing video recordings, presentations, and documents conducted by consortium members that explain the eFlows4HPC architecture, which partners can watch offline at their own pace.
- Recordings, demos, training material and presentation slides from the training courses described above. As the training courses are planned to be recorded, the full recordings

will be deposited in the project's repository and shared internally. For any trainings that will be opened to an external audience, the recordings will be published on the project website, as well as the established BSC YouTube channel and, possibly, the PRACE Training Platform.

- A dedicated GitHub channel for all eFlows4HPC software architecture and workflows components. The GitHub resources will be open source, whenever possible, and include the EC acknowledgment sentence in order to comply with the EC Open Access requirements. The link to this GitHub channel will also be posted on the project website and disseminated externally.
- Additional mini internal trainings will be considered, depending on the educational needs of the Pillar-related partners and the feedback obtained after each session. These sessions can be easily organised in a virtual format.

6. Key Performance Indicators

The following Key Performance Indicators (KPIs) have been identified for the WP7 training task in order to monitor the progress of the training activities and measure their impact. These indicators may be adjusted during the project's lifetime depending on the project developments and the Covid-19 pandemic.

Table 2: eFlows4HPC training KPIs

Dissemination channel	KPI	Measure
Training courses	Number of training courses	Three training courses
Hackathon	Number of Hackathon attendees	One hackathon across Pillars Over 25 Hackathon attendees
Training material	Recordings from training courses	5 internal recordings in internal B2DROP repository 2 external recordings on the project website/BSC YouTube channel