D3.1 First update on the assessed applications/codes
Version 1.0

Document Information

Contract Number	101143931
Project Website	www.pop-coe.eu
Contractual Deadline	M6, June 2024
Dissemination Level	PU - Public
Nature	R - Report
Author	Brian Wylie (JUELICH)
Contributor(s)	Marta Garcia (BSC)
Reviewer	Xavier Teruel (BSC)
Keywords	assessments, application codes, systems

This project has received funding from the European High-Performance Computing Joint Undertaking (JU) under grant agreement No 101143931. The JU receives support from the European Union's Horizon Europe research and innovation programme and Spain, Germany, France, Portugal, and the Czech Republic.

Change Log

Version	Author	Description of Change
V0.1	Brian Wylie	Initial Draft
V0.2	Brian Wylie	Updated Draft
V0.3	Marta Garcia	Reviewed content
V0.4	Brian Wylie	Included review updates
V1.0	Elena Markocic	Final version, formatted for submission

Table of Contents

Executive Summary

This deliverable reports on the application codes for which services have been requested from the Services work-package (WP3) of the POP3 HPC Centre of Excellence (CoE) during its first half-year of operation. These services are provided with the goal of helping developers and users of parallel codes to quantify their execution and scaling inefficiencies, identify current bottlenecks and promote performance analysis as best practice.

POP3 has a particular emphasis on services for the other HPC CoEs and use of HPC computer systems provided by EuroHPC Joint Undertaking.

16 application codes are characterised, with 14 of them originating from other HPC CoEs and the others of interest to CoEs and the HPC community in Europe. 13 of the assessments were done using EuroHPC computer systems, with 6 of these using accelerated partitions with Nvidia or AMD GPUs. In combination with MPI message-passing, most of the codes include multithreading with OpenMP or pthreads and/or offload of computational kernels to GPUs via OpenACC, CUDA or HIP.

This report will be updated every half year and included in expanded analysis of completed assessments produced at the end of every year.

1. Introduction

This report characterises the application codes for which performance assessment, or second-level services were requested from POP3 and undertaken by WP3. It also considers the computer systems used for the assessments and the scale of execution, i.e., the number of compute nodes, CPU cores and/or GPU devices used. It refers to the first half-year of POP3.

Primary assessments of parallel application execution efficiency and scaling are undertaken by Task 3.1, supplemented by a smaller number of secondary assessments of execution correctness, energy utilisation, and proof-of-concept prototype implementations of potential remedies done by Task 3.2.

In addition to tracking the demand for our assessment services and their timely delivery, it allows us to ensure their relevance to the HPC application CoEs and EuroHPC computer systems. Service quality indicated by customer satisfaction is independently measured by our customer advocate Task 2.4 following each service and reported in the associated deliverables. Other WP2 tasks (dissemination, training and business development) assist with promoting completed services and acquiring new service requests. WP4 tasks 4.1 and 4.2 study completed assessments to identify appropriate kernels for their co-design activities, while Tasks 4.3 and 4.4 develop tool and methodology improvements prioritised by issues analysts encountered when doing assessments.

70% of POP3 services are targeted to the other HPC application CoEs, both via ongoing engagements with them or newly established via the associated Coordination and Support Action (CSA) CASTIEL2. In addition to individual service requests, assessment campaigns are offered to CoEs where they can have all of their application codes assessed collectively.

CoE	Codes identified	#
bioExcel-3	GROMACS	1
CEEC	FLEXI, Alya, Nek5000/NekRS, Neko, waLBerla	5
ChEESE2	**Elmer/ICE, ExaHyPE2, FALL3D, HySEA, LaMEM, Open-PDAC, pTatin3D, SeisSol, SPECFEM3D, TANDEM, xSHELLS**	11
EoCoE	(not specified)	5
ESiWACE	EC-Earth, IFS, NEMO, ***ICON***	6
EXCELLERAT2	Neko, Alya, M-AIA, VISTLE	6
HiDALGO2	OpenFOAM, EULAG, waLBerla	3
MaX	Quantum ESPRESSO, SIESTA, BigDFT, Fleur, Yambo	5
MultiXscale	**ESPResSo**	3
Plasma-PEPSC	BIT1, BIT3, GENE-X, PIConGPU, Vlasiator	4
SPACE	**iPiC3d, OpenGADGET**, PLUTO, RAMSES	3
(BrainScoEPE)	Arbor, ATLaS, ExTract, MiMiC, ***NEST***	6
PerMedCoE	PhysiBoSS	3
TREX	(not specified)	2

Table 1: HPC application CoE codes

Prior to POP3 commencement, 14 running and proposed CoEs documented 64 of their codes for a potential total of 82 studies. Some of these studies are expected to depend on CoE priorities and funding, however, two additional CoEs are expected to start mid-2024 and will be contacted to determine their service needs. Table 1 summarises the identified codes per CoE, showing in bold the codes for which assessments have been done.

2. Assessment services

This deliverable covers 16 performance assessments and 1 second-level correctness check service performed as part of one of those performance assessments during the initial half-year of service.

BSC, IT4I, JUELICH, RWTH and USTUTT provided these services, with INESC-ID and UVSQ contributing to service review discussions and expected to deliver services in future.

6 performance assessments have been completed to date, 5 assessments have preliminary results reported, whereas the remaining 5 are still in earlier stages. (One further assessment service request for a commercial code is pending a Non-Disclosure Agreement and not included in this report.)

11 assessment requests were received as part of an assessment campaign for the ChEESE CoE, comprising all of their 11 flagship codes. 2 assessment requests were also received for SPACE CoE flagship codes, along with another for MultiXscale CoE. The additional 2 assessment requests, while not received from HPC CoEs, are also large-scale application codes of interest to CoEs and broader HPC community.

Figure 1 summarises the assessment request origin:

- 11 ChEESE CoE
- 2 SPACE CoE
- 1 MultiXscale CoE
- 2 other

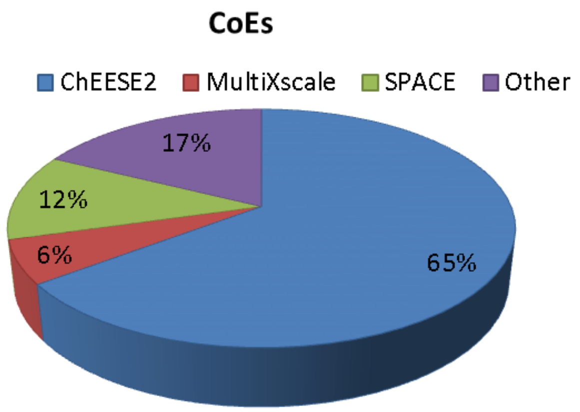

Figure 1: Assessment request origin

13 of the assessments used EuroHPC computer systems, with 3 using other systems (including those available at POP member institutions). 6 of these assessments used accelerated partitions with Nvidia or AMD GPUs.

Figure 2 summarises the computer systems (partitions) used for assessments:

- 5 Leonardo-Booster [Leo-B] (4 GPUs per node)
- 1 LUMI-G (8 GPUs per node)
- 4 LUMI-C (128 CPU cores per node)
- 2 Leonardo-DCGP [Leo-C] (116 CPU cores per node)
- 1 Vega-CPU (128 CPU cores per node)
- 3 other

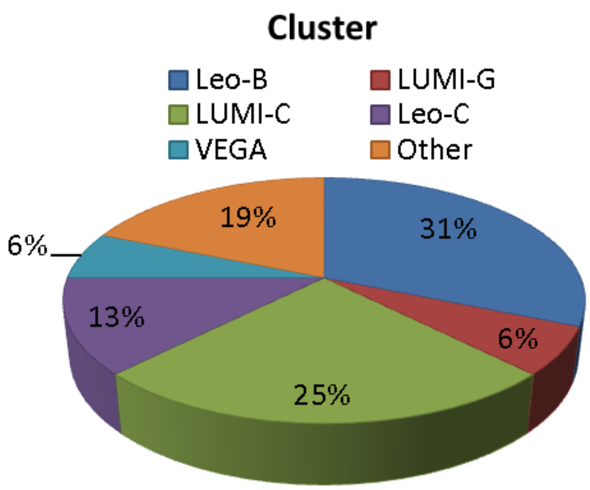

Figure 2: Computer systems (partitions) used for assessments

3. Application code characterisation

C++ was the most commonly employed application programming language, followed by Fortran and C, but often a combination of languages was used. Python was additionally used for some C++ codes, often as a driver or pre/post-processing step.

Figure 3 summarises the programming languages employed in the assessed applications:

- 4 C++
- 4 C++, Python
- 1 C
- 3 C, C++
- 2 Fortran
- 1 Fortran, C
- 1 C, C++, Fortran

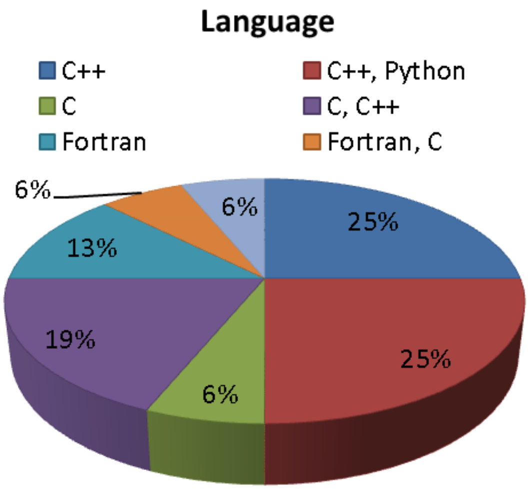

Figure 3: Programming languages employed in assessed applications

All of the assessed codes included MPI message-passing, with only 5 considering exclusively MPI. 3 have included multithreading (OpenMP or pthreads) and 5 included accelerator offload (CUDA, HIP or OpenACC). Several codes have accelerated versions in development, which were not ready for assessment at this point in time but expected to be the subject of a subsequent assessment.

Figure 4 summarises the application parallelization paradigms analysed:

- 5 exclusively MPI message-passing
- 2 MPI+OpenMP multithreading
- 1 MPI+pthreads+CUDA
- 1 MPI+OpenACC
- 2 MPI+CUDA
- 1 MPI+HIP
- (4 cases of MPI with OpenMP, OpenACC or CUDA not yet determined)

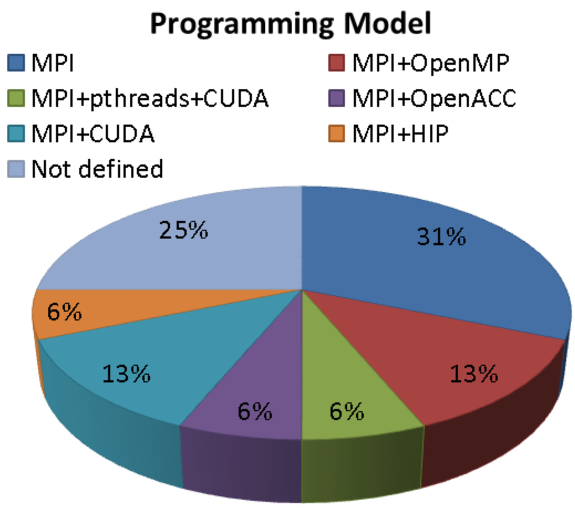

Figure 4: Parallelization paradigms analysed

Figure 5 shows that 10 assessments examined strong scaling (constant total problem size), whereas weak scaling (fixed problem size per CPU or GPU) was specified in 6 assessments: a few cases considered both forms of scaling. The largest execution configuration analysed used 8192 CPU cores (64 compute nodes), whereas accelerated configurations used up to 512 GPUs: in all cases each GPU was controlled from a dedicated corresponding MPI process, such that additional CPU cores on the node were generally not used.

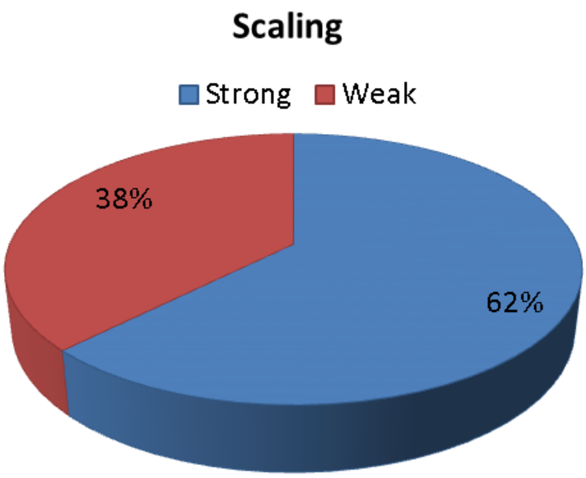

Figure 5: Type of scaling assessed

4. Results and impacts

POP3 targets 120 completed services (at M36), of which 82 are for codes from CoEs. For each six-month period, 20 services are therefore expected of which 14 are CoE codes. Most services are expected to be (initial) performance assessments, with a smaller proportion being follow-on second-level services (proof-of-concept prototypes, correctness checks, energy-efficiency audits, etc.).

During the first six-month period of service, 16 performance assessments have been undertaken (of which 6 were completed in time for this report), along with a single second-level service (correctness check) in conjunction with one of them. 14 of these assessments are flagship codes of HPC CoEs, with another code also relevant to an HPC CoE, meeting our expectation.

Although service requests originating from the ChEESE CoE campaign provided an initial batch of performance assessments, these were subject to a number of delays relating to the readiness of the application codes and associated testcases for the assessments, and access to the EuroHPC computer systems where measurements were desired. Other delays relate to Non-Disclosure Agreements with customers required before services can start. These delays are familiar from prior instances of POP, and were expected to contribute to a service lag during the initial period.

It is too early to determine performance and efficiency improvements for the services just completed: these will become clearer in follow-ups and ongoing engagements with CoEs. The correctness check service identified and provided remedies for five distinct race conditions in the code which have subsequently been incorporated.

Most of this initial set of assessments was done on EuroHPC computer systems, confirming that the CoE application codes and POP performance tools can already be used on these platforms. However, the application codes (and particularly GPU-enabled versions) are still in development and not yet ready to execute extreme-scale computations. This manifests in the performance assessments which revealed low execution efficiencies and poor scaling (already at modest scale).

Load-balance is a common parallel inefficiency factor, particularly for strong scaling test cases that must be partitioned over increasing numbers of compute nodes (and their CPU cores and GPU accelerator devices). Where GPUs provide the bulk of computational power, it is common to focus on their effective utilisation while underutilising (or idling) associated CPU cores that are also available on the accelerated compute nodes. While some assessments found that the GPUs were generally well utilised, various others identified significant under-utilisation often in conjunction with overheads associated with the accelerated kernels being too small.

Generally, it has been straightforward to port and install our tools on this first subset of EuroHPC and other HPC computer systems in Europe. Installations have been local to individual projects, with effort ongoing to deploy the tools for general use in WP4 Task 4.3.

Various limitations of the POP performance tools and methodology have been identified from this initial set of assessments which are being considered in the associated tasks of WP4. The tools need improved support for certain parallelisation paradigms (particularly HIP) and combinations of paradigms which complicate measurement and analysis. The analysis methodology needs to be adapted to cases where GPUs are being used in different ways, with combinations of paradigms (where the use of associated CPU cores for computation is of minimal importance), and streamlined to improve ease of use. A common issue encountered is the huge size of execution traces that are prohibitively large for analysis, which can sometimes be worked around by limiting measurements to a small number of steps/iterations.

5. Conclusion

The first six-month period of POP3 has completed assessments of 6 codes with 10 more in progress. All but one application code is relevant to HPC CoEs, and most assessments have used EuroHPC computer systems, with roughly half using GPU-accelerated partitions. While all codes relied on MPI message-passing, a wide variety of multithreading and/or accelerator-offload paradigms were exploited, motivating associated improvements to be developed and incorporated within POP3 tools and methodology.

Acronyms and Abbreviations

- CA – Consortium Agreement
- CAdv – Customer Advocate
- CoE – Centre of Excellence
- CUDA – Compute Unified Device Architecture
- CPU – Central Processing Unit
- D – deliverable
- DoA – Description of Action (Annex 1 of the Grant Agreement)
- EC – European Commission
- GA – General Assembly / Grant Agreement
- GPU – Graphics Processing Unit
- HIP – Heterogeneous-Compute Interface for Portability
- HPC – High Performance Computing
- IPR – Intellectual Property Right
- KPI – Key Performance Indicator
- M – Month
- MPI – Message Passing Interface
- MS – Milestones
- NDA – Non-Disclosure Agreement
- PEB – Project Executive Board
- PM – Person month / Project manager
- POP – Performance Optimisation and Productivity
- R – Risk
- RV – Review
- WP – Work Package
- WPL – Work Package Leader

<u>POP3 Beneficiaries</u>
- BSC: BARCELONA SUPERCOMPUTING CENTER - CENTRO NACIONAL DE SUPERCOMPUTACION
- FZJ: FORSCHUNGSZENTRUM JÜLICH GMBH
- RWTH: RHEINISCH-WESTFAELISCHE TECHNISCHE HOCHSCHULE AACHEN
- IT4I@VSB: TECHNICAL UNIVERSITY OF OSTRAVA
- INESC ID: INSTITUTO DE ENGENHARIA DE SISTEMAS E COMPUTADORES, INVESTIGACAO E DESENVOLVIMENTO EM LISBOA
- TERATEC: TERATEC
- UVSQ: UNIVERSITE DE VERSAILLES SAINT-QUENTIN-EN-YVELINES
- USTUTT: UNIVERSITY OF STUTTGART FOR ITS HIGH PERFORMANCE COMPUTING CENTER STUTTGART

List of Figures

List of Tables