E-Infrastructures
H2020-INFRAEDI-2018-1

INFRAEDI-2-2018: Centres of Excellence on HPC

EoCoE-II

Energy oriented Center of Excellence : toward exascale for energy

Grant Agreement Number: INFRAEDI-824158

D5.3
Melissa-DA: Final Code Release

Project and Deliverable Information Sheet

EoCoE	Project Ref:	INFRAEDI-824158
	Project Title:	Energy oriented Centre of Excellence
	Project Web Site:	http://www.eocoe.eu
	Deliverable ID:	D5.3
	Lead Beneficiary:	INRIA
	Contact:	Bruno Raffin
	Contact's e-mail:	Bruno.Raffin@inria.fr
	Deliverable Nature:	Report and Code
	Dissemination Level:	PU*
	Contractual Date of Delivery:	M34 01/05/2022
	Actual Date of Delivery:	M38 4/08/2022
	EC Project Officer:	Evangelia Markidou

* - The dissemination level are indicated as follows: PU – Public, CO – Confidential, only for members of the consortium (including the Commission Services) CL – Classified, as referred to in Commission Decision 2991/844/EC.

Document Control Sheet

Document	Title :	Melissa-DA: Final Code Release
	ID :	D5.3
	Available at:	http://www.eocoe.eu
	Software tool:	LaTeX
Authorship	Written by:	Bruno Raffin, Sebastian Friedemann, Yen-Sen Lu, Kai Keller
	Contributors:	Bibi Naz, Harrie-Jan Hendricks Franssen, Anna Sekula, Bartłomiej Pogodziński, Christoph Conrads, Leonardo Bautista-Gomez
	Reviewed by:	Sebastian Lührs, Mathieu Lobet

Contents

1. Overview

1.1 General Objectives of WP5

The WP5 is one of EoCoE technical challenges called *Ensemble Runs*. The goal is to efficiently run large simulation ensembles on coming pre-exascale and exascale systems, in particular for data assimilation. The objective pursued here is to provide a flexible and maintainable way of executing simulation ensembles in multiple jobs on a given supercomputer while enabling communication between ensemble members to allow ensemble management and data assimilation processes. For that purpose WP5 targets developing an elastic exascale-ready framework for ensemble runs extending the Melissa approach developed at INRIA [3].

Two out of five EoCoE-II Scientific Challenges (Meteorology for Energy and Water for Energy) integrate some support for ensemble runs for data assimilation or sensitivity analysis, usually relying on one monolithic big MPI job or file-based solutions, like ESIAS developed during EoCoE-I. WP5 relies on a novel developed framework, called Melissa-DA, to empower the Meteorology and Water EoCoE-II applications to benefit from next generation exascale machines for large ensemble runs. WP5 builds on top of WP4 (I/O) work re-using the PDI interface for data access and the FTI library for fault tolerance.

1.2 WP5 Work Progress

The WP5 is structured in 3 tasks:

- Task 5.1: Ensemble Run Framework Development, M1-M30, Task leader INRIA,
- Task 5.2: Ensemble Runs for Meteorology. M6-M36, Task leader: FZJ,
- Task 5.3: Ensemble Runs for Water, M6-M36, Task leader: FZJ,

with three deliverables:

- D5.1 - Architecture Specification and prototype codes for the framework and the 2 applications with initial documentation as well as an early usability and performance evaluation. (M18) (Report + code DEM, PU) (INRIA),
- D5.2 - M24 - First stable version of the framework code and the two applications adapted to the framework. Report will include documentation as well as performance evaluation. (M24) (Report + code DEM, PU) (FZJ),
- D5.3 - M34 - Final code release for the framework and the three applications, with the associated documentation. Report on testing at large scale. (M36) (Report + code DEM, PU) (CEA),

The WP5 contributions include:

- First period (deliverable D5.1):
  - A first functional code implementing **Melissa-DA**, the Melissa extension for data assimilation.
  - A toy use case based on Lorenz equation, as well as an advanced use case based on the ParFlow hydrology simulation code.
  - Integration of the PDAF data assimilation engine into Melissa server, enabling to support different assimilation algorithms.

  - Experiments on supercomputers with ParFlow simulations, EnKF assimilation, running up to 1024 members.
  - Identification of ambitious use cases with code and datasets for the Weather and Hydrology applications.
  - Submission of two research papers related to Melissa-DA results

- Second period (deliverable D5.2):
  - Following remarks on D5.1, switch to a single (public) code repository for Melissa-DA: `https://gitlab.inria.fr/melissa/melissa-da`)
  - Consolidation of the Melissa-DA code
  - Large scale Melissa-DA experiment with ParFlow simulations, EnKF assimilation, propagating 16,384 members on 16,240 cores.
  - Development of a variation of Melissa-DA specifically targeting data assimilation with a Particle Filter
  - Support of the WRF (Weather Research and Forecasting Model) for enabling assimilation with a Particle Fitler using Melissa-DA
  - Large scale Melissa-DA experiment with WRF simulations, Particle Filter assimilation, running up to 2,555 members on 20,442 cores.
  - Communications:
    * Revision of paper under submission at International Journal of High Performance Computing Applications.
    * Two paper submissions at IEEE Cluster 2021 and HPCS 2021.
    * Oral presentation at EnKF Workshop 2021
    * Joint organisation with WP4 of the HPCDA workshop at ISC High Performance 2021

- **Final period (this deliverable D5.3)**:
  - Consolidation of the Melissa-DA (EnKF and particle filters).
  - Production run comparing PDAF and Melissa with Parflow simulation code. Melissa is 210% faster than PDAF at 2500 members using 500 nodes. Details available in [1].
  - Production run comparing a file-based approach and Melissa with WRF simulations, Particle Filter assimilation up to 1025 members. Melissa is about 3× more efficient than the file-based solution at 1024 members (Detailed results in preparation for article submission at Journal of Computational Science).
  - Porting of Melissa-DA on Fugaku supercomputer(#2 at top500.org), with extreme scale tests (49952 cores) using a Lorenz96 model and particle filter.
  - Kai Keller public PhD Defense, BSC. Towards zero-waste recovery and zero-overhead checkpointing in ensemble data assimilation. 1/7/2022.
  - Sebastian Friedemann public PhD Defense, INRIA. 1/7/2022. Ensemble-based Data Assimilation for Large Scale Simulations.
  - Communications:
    * Journal article: An elastic framework for ensemble-based large-scale data assimilation, Journal of High Performance Computing Applications [1].

- ∗ Conference article: Towards Zero-Waste Recovery and Zero-Overhead Checkpointing in Ensemble Data Assimilation, 2021 IEEE 28th International Conference on High Performance Computing, Data, and Analytics (HiPC) [2]
- ∗ Oral presentation at International Symposium on Data Assimilation (2022): Melissa-DA: An Elastic and Fault-Tolerant Large-Scale Online Data Assimilation Framework.
- ∗ Oral presentation at 19th Workshop on high performance computing in meteorology (2021). Elastic Large Scale Ensemble Data Assimilation with Particle Filters for Continental Weather Simulation. video.
- ∗ Participant to Round table "Complex application workflows in the exascale era". EuroHPC week, march 2022.
    - · Andrea Ferretti, MAX, Researcher at CNR
    - · Anthony Scemama, Trex, Engineer of research at CNRS
    - · Bruno Raffin, EoCoE, Director of research at INRIA,
    - · Miguel Vázquez, PerMed, Head of genome informatics team at BSC
    - · Peter Coveney, CompBioMed coordinator, Director of the center for computational science, UCL
- ∗ 13th JLESC (Joint Laboratory for Extreme Scale Computing) Workshop:
    - · Chair of Break Out Session *In Situ Processing at Large*. Bruno Raffin (INRIA)
    - · Talk *Fault Tolerant and Elastic Particle Filter for Extreme Scale Weather and Climate Data Assimilation*. Kai Keller, BSC.

WP5 developments are done by INRIA (lead and Melissa-DA core development), PSNC (core Melissa-DA development), FZJ (use cases) and CEA (PDI and Melissa-DA integration). BSC is contributing to Melissa-DA fault-tolerance mechanism with FTI (work attached to WP4).

2. Code Repositories

This deliverable is mainly providing the final code of Melissa-DA developed during EoCoE-II. Code repository, including documentation: `https://gitlab.inria.fr/melissa/melissa-da`. Instrumented version of Eocoe-II applications to run with melissa-DA:

- WRF: `https://gitlab.inria.fr/melissa/WRF`
- Parflow and Parflow+CLM: `https://gitlab.inria.fr/melissa/parflow-melissa-da`

3. Scalability Targets

Note that a one-to-one direct comparison between Pre-EoCoE-II and Melissa based EoCoE-II target runs should be made very carefully as the numbers given below are focused on scalability only and do not reflect the gains in efficiency and resilience. *Hero run* denotes a large scale run pushing the scalability in number of members of the system. *Production run* denotes a realistic run involving coupled codes and full observation datasets. Notice that we are facing difficulties to have the necessary access rights to reserve very large number of cores for super large scale tests.

Meteorology - ESIAS (WRF+EURAD-IM)

- Pre-EoCoE-II status: production runs with 4096 ensemble with 262,144 CPU cores
- **EoCoE-II results.** Data assimilation with Melissa using a particle filter for WRF, cloud coverage as observations. All runs are performed on the JUWELS machine:
  - Hero run (WRF): 2,555 members on 20,442 cores
  - Production runs. Compare a Melissa run versus the file-based approach developed for EocoE-II. We provide here the average times per cycle:

Members	Cores	seconds (Melissa)	Seconds (File-based)	Melissa versus File-based Speed-up
128	384	1062	785.44	0.73958569
256	768	1062	803.87	0.75693974
512	1536	1068	1209.43	1.1324251
1024	3072	1071	1234.64	1.1527918

Water - TerrSysMP (ParFlow-CLM)

- Pre-EoCOE-II status:
  - Hero run (ParFlow, CLM and TSMP-PDAF): 256 members on 132,768 CPU cores
  - Production runs :
    - Neckar use-case (Parflow, CLM and TSMP-PDAF): 64 ensemble members on 4,608 CPU cores.
    - Europe use-case (CLM and TSMP-PDAF): 20 ensemble members on 1,920 CPU cores.
- **EoCoE-II results**:
  - Hero run (Parflow, Melissa): 16,384 members on 16,240 cores.
  - Production runs comparing PDAF and Melissa on the Neckar catchment in Germany. The simulation is performed with ParFlow. Data assimilation is performed with a EnKF filter implemented by PDAF in both cases. The experiments ran on the Jean-Zay machine and we report average times per cycle:

Members	Nodes	Seconds (PDAF)	Seconds (Melissa)	Speed-up (PDAF/Melissa)
250	20	18	19.6	0.91836735
1000	200	24	21	1.1428571
2500	500	56	26	2.1538462

4. Summary

We developed an operational first version of Melissa-DA with two flavours, one for data assimilation relying on the PDAF assimilation engine, and one for particle filters. Code and documentation are publicly available on a gitlab repository. We performed extended performance tests, hero and production runs, with particle filters and the WFR code, and with EnKF and the Parflow code. These results demonstrate Melissa capabilities to push the boundaries of DA at scale, with 2x and more gains beyond 1000 members compared to other approaches. In addition to these performance achievements, Melissa also support fault-tolerance and elasticity, features that we believe will be essential when pushing the scale even further on the coming Exascale machines.

Beyond the planned tasks, we managed to port, test and validate Melissa on a large variety of supercomputers (Juwels, Jean-Zay, MareNostrum, Fugaku - MareNostrum and Fugaku porting were done by BSC). Fugaku is particularly interesting as being today the second largest supercomputer in the world with a very specific architecture relying on an ARM processor and a custom high performance network. This effort helped to consolidate Melissa code and prepare it for next generation machines.

WP5 has tightly collaborated with WP4, in particular on PDI to develop a PDI plugin for Melissa and on checkpointing, FTI being a key component for particle filtering with Melissa.

Melissa development efforts will be pursued beyond EoCoe-II through EU (REGALE) and french fundings (NUMPEX).

5. Acknowledgments

This work was granted access to the HPC resources of IDRIS under the allocation 2020-A8 A0080610366 attributed by GENCI (Grand Equipement National de Calcul Intensif). The authors gratefully acknowledge the Gauss Centre for Supercomputing e.V. (www.gauss-centre.eu) for funding this project by providing computing time through the John von Neumann Institute for Computing (NIC) on the GCS Supercomputer JUWELS at Jülich Supercomputing Centre (JSC). We also acknowledge PRACE for awarding us access to JUWELS at Jülich Supercomputing Centre (JSC), Germany. Kai Keller, BSC, was granted a student Intership from Riken to work on the Fugaku machine. In case of PSNC the scientific/academic work is co-financed from financial resources for science in the years 2019 - 2022 granted for the realization of the international project co-financed by Polish Ministry of Science and Higher Education (5057/H2020/19/2020/2).

6. References

[1] Sebastian Friedemann and Bruno Raffin. An elastic framework for ensemble-based large-scale data assimilation. *The international journal of high performance computing applications*, 36(4):543–563, June 2022.

[2] Kai Keller, Adrian Cristal Kestelman, and Leonardo Bautista-Gomez. Towards zero-waste recovery and zero-overhead checkpointing in ensemble data assimilation. In *2021 IEEE 28th International Conference on High Performance Computing, Data, and Analytics (HiPC)*, pages 131–140. IEEE, 2021.

[3] Théophile Terraz, Alejandro Ribes, Yvan Fournier, Bertrand Iooss, and Bruno Raffin. Melissa: Large scale in transit sensitivity analysis avoiding intermediate files. In *International Conference for High Performance Computing, Networking, Storage and Analysis (SC'17)*, Denver, 2017.