PDI and DEISA to decouple I/O concerns towards in-situ analysis

J. Morice, J. Auriac, B. Martin, Y. Wang, J. Bigot
Maison de la Simulation, CEA / CNRS / UVSQ / Université Paris-Saclay

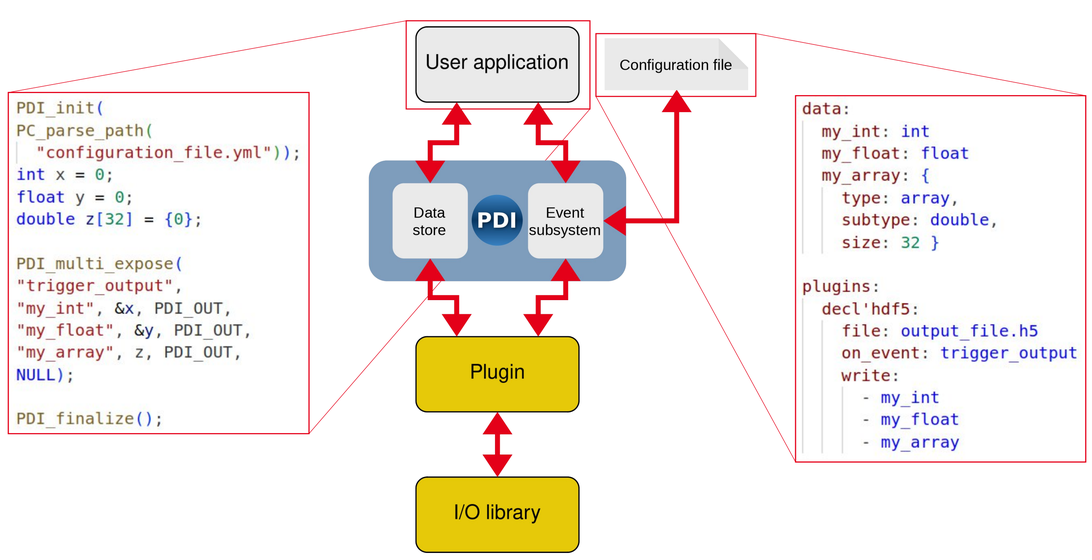
Summary

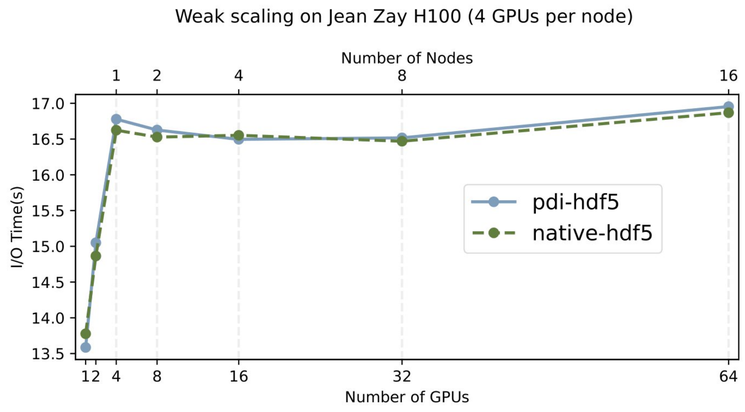
High-performance exascale simulations increasingly face **I/O bottlenecks** due to the growing gap between computational speed and I/O bandwidth. PDI addresses these challenges by **decoupling simulations from I/O concerns**, offering a declarative API and supporting libraries like HDF5 and NetCDF through a plugin system. It enables **in-situ analysis** by processing data in real time and reducing disk I/O. DEISA builds on PDI, integrating MPI-parallel simulations with Dask-based analytics for seamless, scalable, and efficient in-situ data analysis.

PDI

The PDI data interface is designed to support loose coupling of simulation codes with external I/O and data analysis libraries.

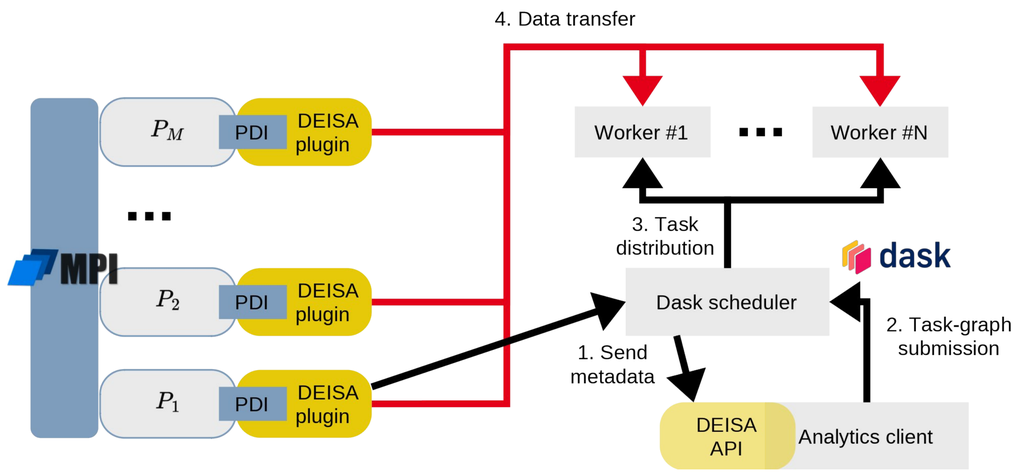
- Light and easy to use API
- No added I/O from the interface, no significant overheads
- A simulation code is annotated using the PDI declarative API, and the I/O libraries are specified without recompiling the simulation code. These I/O libraries are available via different plugins
- Integrated in multiple large scale HPC production simulation codes: Gysela, ParFlow, SHEMAT, TRUST
- Open support & feedback through a dedicated Slack channel
- Main plugins :
  - HDF5, NetCDF
  - Python, user_code: call custom Python/C code
  - trace: logging and debugging
  - DEISA: in-situ analysis, to prevent I/O bottlenecks





I/O time, weak scaling study with and without PDI for the HDF5 strategy [1, 2].

DEISA



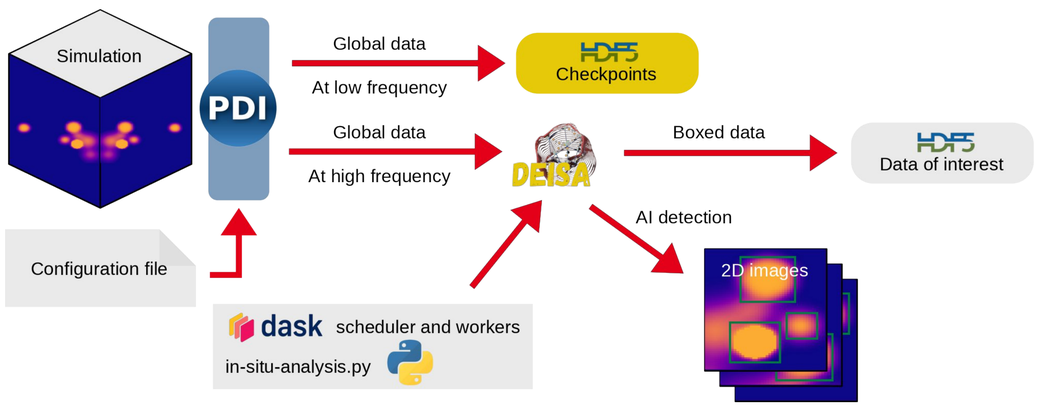
Coupling an MPI application with M processes to a Dask instance running N workers [3].

The arrows and numbers illustrate the different steps of the internal workflow of DEISA.

DEISA (Dask-Enabled In-Situ Analytics) [4] is a library and PDI plugin which enables coupling MPI simulation codes with Dask framework for in-situ analytics.

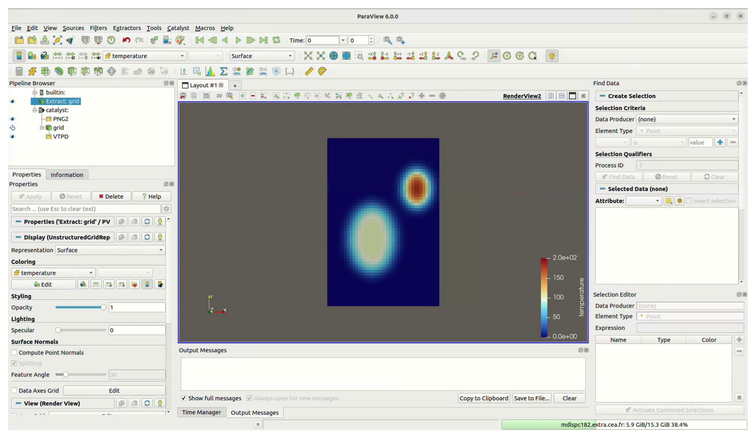
- Based on Dask, a Python framework for parallel and distributed computing
- Leverage from a robust ecosystem of tools for data analysis and visualization, such as NumPy, Pandas, scikit-learn, and matplotlib
- Fine-grain control over data and timestep selection
- Transition from post-hoc to in-situ analysis

Use case



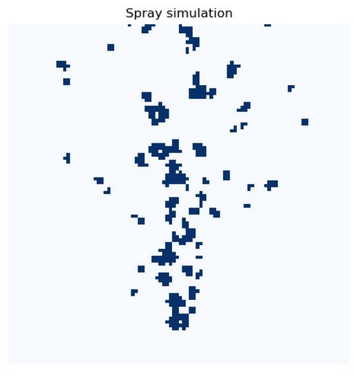
Example of advanced data analysis using PDI and DEISA in an AI workflow.

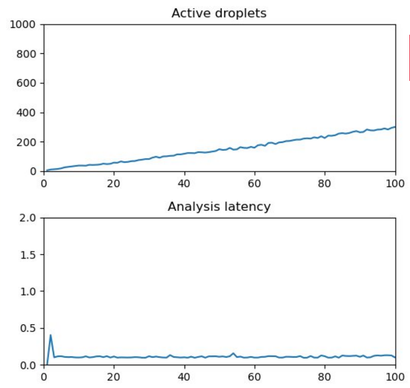
Perform periodic HDF5 checkpoints and detect zones of interest using AI [4, 5].



Catalyst plugin for PDI [6].

In-situ visualization of a heat equation simulation handled through PDI and the Catalyst plugin.





Spray simulation with data analysis by DEISA.

In-situ plot of the spray simulation (left).
Total number of droplets computed by DEISA (top right).
Wall time of analysis by DEISA (bottom right).

Acknowledgments

As part of the "France 2030" initiative, this work has benefited from a State grant managed by the French National Research Agency (Agence Nationale de la Recherche) attributed to the Exa-DoST project of the NumPEx PEPR program, reference: ANR-22-EXNU-0004.

[1] C. Roussel, K. Keller, M. Gaalich, L. B. Gomez, J. Bigot. PDI, an approach to decouple I/O concerns from high-performance simulation codes. 2017. hal-01587075.
[2] R. Bourgeois. Contributions to the study of MHD convection : theoretical framework, finite volume methods, and pre-exascale simulations. Numerical Analysis [cs.NA]. Université Paris-Saclay, 2024. English. ⟨NNT : 2024UPASP044⟩.
[3] A. Gueroudji, J. Bigot, B. Raffin. DEISA: dask-enabled in situ analytics. HiPC 2021 - 28th International Conference on High Performance Computing, Data, and Analytics, Dec 2021, virtual, India. pp.1-10. hal-03509198.
[4] M.Mancip, Y. Wang, B. Martin, M. Bourenane, Myriam Hamidou, SPOT : Simulation Pattern Observation Tool.
[5] M. Bourenane. Optimisation In-Situ de la Détection de Phénomènes par Computer Vision et Reconstruction par Graphes dans les Simulations Numériques LWFA. Université Paris Dauphine - CFA AFIA, 2023-2024.
[6] U. Ayachit, A. C. Bauer, B. Boeckel, B. Geveci, K. Moreland, P. O'Leary, and T. Osika. 2021. Catalyst Revised: Rethinking the ParaView in Situ Analysis and Visualization API. In High Performance Computing, Springer International Publishing, Cham, 484–494.

Join our Slack channel for tutorial information