



Adaptive multi-tier intelligent data manager for Exascale



admire-eurohpc.eu

An Introduction to ADMIRE

Rocco Sedona

Iceland, 28/08/2023

Grant Agreement number: 956748 — ADMIRE — H2020-JTI-EuroHPC-2019-1

- ❑ Introduction to ADMIRE
 - Tools
- ❑ Applications
- ❑ Outlook:
 - Possible integrations
- ❑ Others



unsplash.com

TO BE ADDED

TO BE ADDED

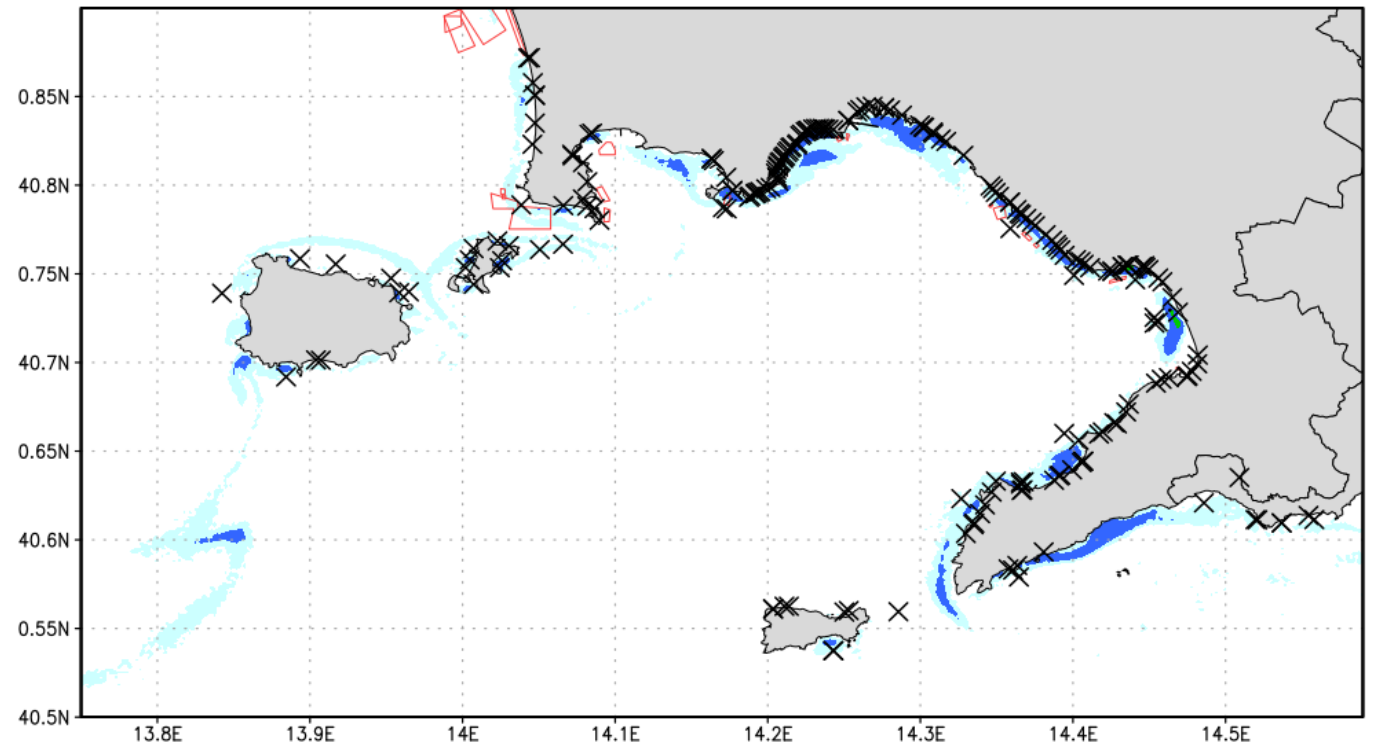
- ❑ Coordinate the work of all application case studies
- ❑ Application co-design
 - Malleability
 - ❖ Flex MPI for efficient load-balancing
 - ❖ Other existing solutions (f.e., Elastic Horovod)
 - Integration of ad hoc file systems
- ❑ Streamline the porting of ADMIRE technologies

- Environment (A1 - ENV): Marine and weather forecasts and simulations
- Molecule Simulation - Quantum-Espresso (A2 - QE): Describes complex electronic interactions
- Turbulence Simulation - Nek5000 (A3 - NEK): Simulations of large-scale turbulent flows
- Remote sensing (A4 - RS): Classification of remotely sensed images
- Life Sciences - SRRF (A5 - SRRF): Live-cell Super-Resolution Microscopy
- Software Heritage Analytics (A6 - SHA): Software stack that enables Data Analytics on Software Heritage

- ❑ Use case:
 - WaComM++, a Lagrangian model that simulates marine pollutants' transport and diffusion processes.
- ❑ Work done:
 - Ported applications to Torino cluster (WRF – ROMS – WaComM++)
 - Improved workflow engine "DagOnStar" (parallel pattern)
- ❑ On going:
 - Support to ad-hoc filesystem integration: Hercules File System
 - CPU malleability in WaComM++ application based on FlexMPI

<http://meteo.uniparthenope.it>

Forecast: 07Z05JUN2023 Golfo di Napoli (VET0130/wcm3)



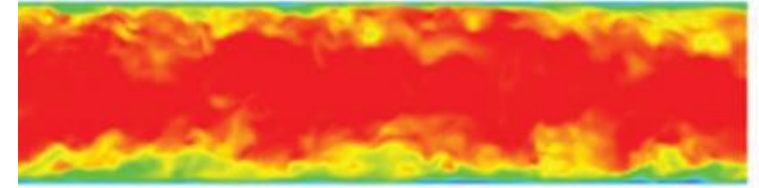
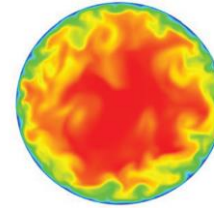
WaComM++ output

- ❑ Use case: Simulation of material modelling at the nanoscale
 - molecular dynamics with CP (Car-Parrinello)
 - vibrational properties with PH (PHonon)

- ❑ Work done:
 - Suite installed on Torino cluster for CPUs
 - Benchmark with PH on 1440 cores (40 nodes)
 - Support to lossless compression (synchronous-asynchronous) with ADIOS

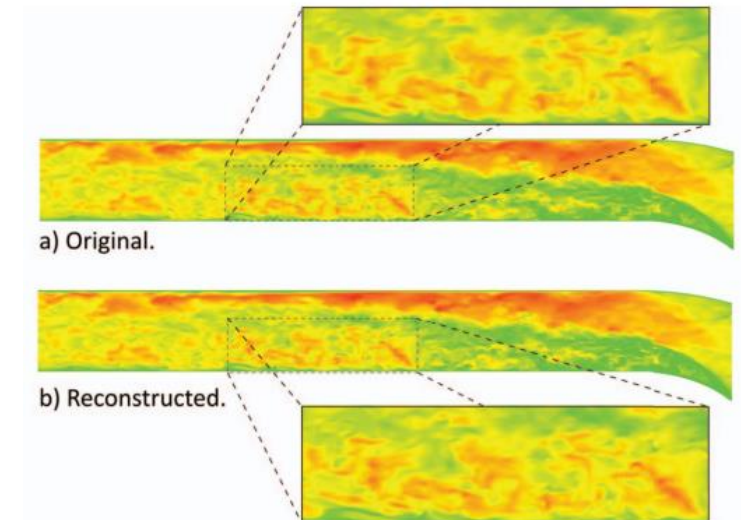
- ❑ On going:
 - ❑ In situ workflow based on QEHeat alongside CP (wip)
 - ❑ Support to ad-hoc filesystem integration : GekkoFS and Hercules (discussed)
 - ❑ CPU malleability in PH based on FlexMPI (discussed)

- ❑ Use case:
 - Simulations of turbulent flows



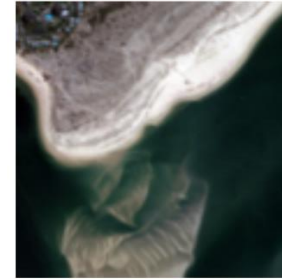
- ❑ Work done up to now
 - Application ported to Turin
 - In situ data compression (with MPCDF)
 - Support for AdHoc file system integration
 - ❖ GekkoFS (Mainly work by Marc Vef and Mainz)
 - ❖ Hercules (Mainly work by Francisco Garcia and UC3M)

- ❑ Ongoing activities
 - Integration with IC and compute malleability (Evaluation)



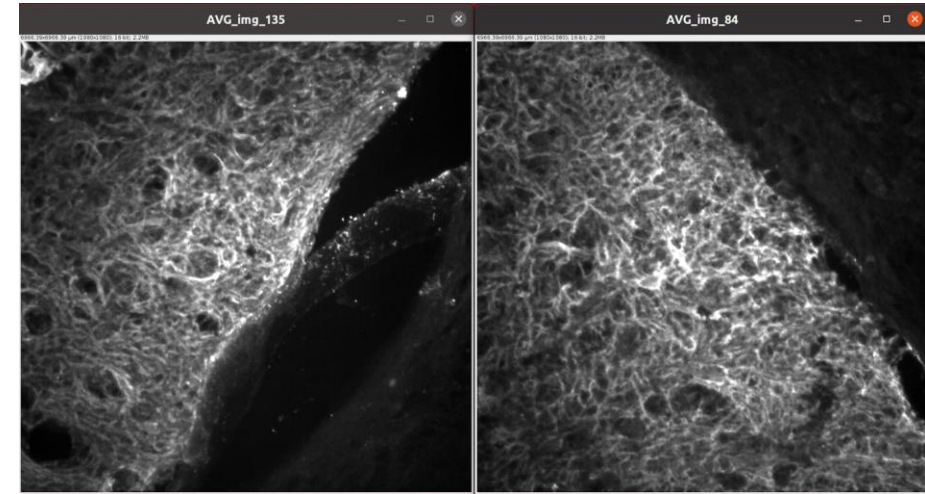
Y. Ju, A. Perez, S. Markidis, P. Schlatter and E. Laure, "Understanding the Impact of Synchronous, Asynchronous, and Hybrid In-Situ Techniques in Computational Fluid Dynamics Applications," 2022 IEEE 18th International Conference on e-Science (e-Science), Salt Lake City, UT, USA, 2022, pp. 295-305, doi: 10.1109/eScience55777.2022.00043.

- Use case:
 - Land cover classification
 - BigEarthNet
 - ❖ Patch based
 - ❖ Multi-class problem
 - Training a Convolutional Neural Network
 - Scaling with Horovod
- Ported application to Turin cluster
- Ongoing activities:
 - Adopt Elastic Horovod
 - ❖ Horovod with Gloo backend installed on DEEP and Turin cluster
 - ❖ Problem: network interface issues on DEEP
 - ❖ Strong collaboration with LS application



R. Sedona, G. Cavallaro, J. Jitsev, A. Strube, M. Riedel and J. A. Benediktsson, "Remote Sensing Big Data Classification with High Performance Distributed Deep Learning", journal of Remote Sensing (MDPI), 2019

- Use case:
 - Polyglutamine disease detection in mice cerebellums
 - Opera microscope images
 - Binary classification
 - Scaling job using Horovod library
- Work done:
 - Port application to Turin cluster
 - Integrating Elastic Horovod into LS application (on PSNC premises)
- Ongoing:
 - Migrating solutions to Turin cluster
 - ❖ Problems : Setting whole environment from scratch (Python, Cmake, Venv)



❑ Use case: Software Heritage Analytics Framework

A framework that allows the execution of distributed data analytics application on projects archived in the Software Heritage repository

❑ Work done:

- Porting the modules Cachemire, Spark of the framework on Turin Cluster
- Write trivial test application

❑ Ongoing:

- Porting the modules Web Interface, Orchestrator of the framework on Turin Cluster
- Try to solve integration problem with ad hoc file system (Hercules, GekkoFS)
- Develop a non-trivial demo application

TO DO

- ❑ Two main goals:
 - Malleability
 - Integration of ad hoc file system
- ❑ Porting all applications on UNITO cluster
 - Mattermost channel with UNITO cluster system admins
- ❑ Provide additional Darshan traces

Thanks!