Accelerating Earth System Workflows with In Situ Workflow Task Management

Manuel G. Marciani[1], Mario Acosta[1], Gladys Utrera[2], Miguel Castrillo[1], and Mohamed Wahib[3]

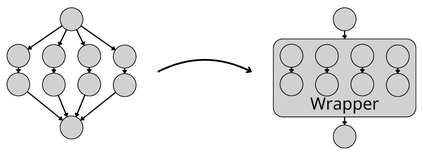
1 Barcelona Supercomputing Center (BSC), Barcelona, Spain 2 *Universitat Politècnica de Catalunya* (UPC), Barcelona, Spain 3 *RIKEN* Center for Computational Science, Kobe, Japan

Introduction

Recently, the Earth Sciences community has begun to consider the performance of the entirety of the execution of simulations, considering waiting time and restarting failed jobs [1], that are executed in HPC platforms.

To minimize the impact of queue time, Autosubmit [2] developers came up with the idea of bundling multiple subsequent and/or independent jobs into a single submission. But the current implementation is limited.

Therefore, aiming to increase the flexibility and performance—while maintaining the portability—we explored in this work three workflow managers to handle the execution of workflow tasks inside the allocation. We call these workflow managers "in situ."



In situ managers and applications

In this work, we explored COMPSs [3], HyperQueue [4], and Flux [5] and measured their computational performance and configurability in MareNostrum 5 and Fugaku.

We tested the scalability of the in situ workflow managers by performing six experiment configurations running NICAM-DC [6], Stress-NG [7], and CMOR [8]. We chose these applications because they are memory, CPU, and I/O bound, respectively.

Finally, we tested if the in situ workflow manager could distribute multi-node MPI tasks.

Workflows tasks

For the IO-bound experiment, we executed two tasks, post-processing ocean, running on ten cores, and atmosphere, running on 16 CPUs, data using CMOR.

For the memory-bound experiment, each task ran NICAM-DC using ten cores simulating a single month of the baroclinic wave test [10] with 240 km of resolution.

For the CPU-bound experiment, Stress-NG executed in ten cores with a timeout of ten minutes.

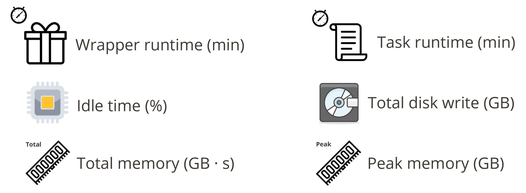
For the task scaling experiment, we used Stress-NG again, but running on a single core for a minute, and executed ten subsequent tasks.

Finally, we performed a ten-member ensemble of the same NICAM-DC test running for six years and with a ten-kilometer resolution needing 320 cores each.

Testing methodology

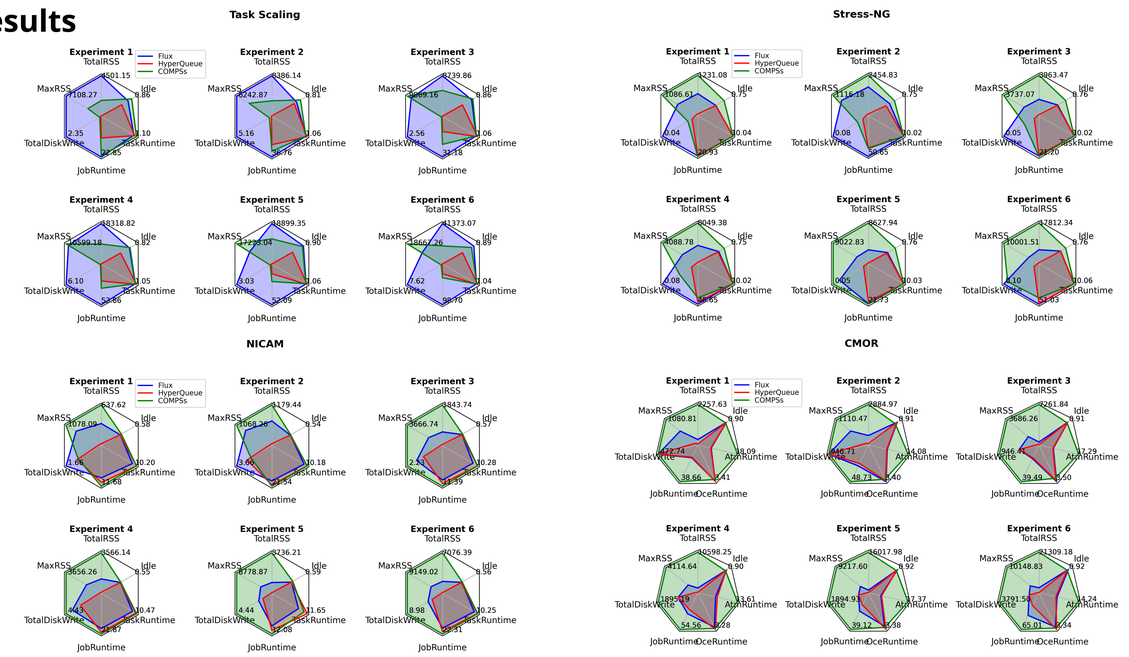
For each of the workflows, we tested how each of the in situ scaled by increasing both the number of workflow tasks and the resources.

For each application, we compute a "workload," which is the total number of tasks necessary to fill up a node in MareNostrum 5, which has 112 cores.



Experiment	Exp 1	Exp 2	Exp 3	Exp 4	Exp 5	Exp 6
Workload	1	2	2	4	4	8
Nodes	1	1	2	2	4	4

Results



Conclusions

HyperQueue has the lowest memory and makespan across all experiments. Moreover, it successfully managed to run multiple node tasks. Flux followed it closely, but it did not perform as well with the task scaling experiment. COMPSs is the only in situ workflow manager to offer a standardized way of provenance tracking.

References

[1] Acosta, M. C. et al. "The computational and energy cost of simulation and storage for climate science: lessons from CMIP6."
[2] Marciani, M. G. et al. "Evaluating the impact of task aggregation in workflows with shared resource environments: use case for the MONARCH application."
[3] Lordan, F. et al. "Servicess: An interoperable programming framework for the cloud".
[4] Beránek, J. et al. "HyperQueue: Efficient and ergonomic task graphs on HPC clusters".
[5] Ahn, D. H. et al. "Flux: Overcoming scheduling challenges for exascale workflows".
[6] Kodama, C. et al. "The Nonhydrostatic ICosahedral Atmospheric Model for CMIP6 HighResMIP simulations (NICAM16-S): experimental design, model description, and impacts of model updates".
[7] https://github.com/ColinIanKing/stress-ng
[8] https://cmor.llnl.gov/
[9] Döscher, R. et al. "The EC-Earth3 Earth system model for the Coupled Model Intercomparison Project 6".
[10] Jablonowski, C. and Williamson, D.L. "A baroclinic instability test case for atmospheric model dynamical cores."

Acknowledgements

We acknowledge the EuroHPC Joint Undertaking (JU) for awarding this project access to the EuroHPC supercomputer LUMI and MareNostrum5 through a EuroHPC JU Special Access call.

Funded by the European Union Destination Earth

implemented by ECMWF esa EUMETSAT

This project has received funding from the grant CEX2021-001148-S-20-5 funded by MICIU/AEI/10.13039/501100011033 and ESF+.

MINISTERIO DE CIENCIA, INNOVACIÓN Y UNIVERSIDADES
Cofinanciado por la Unión Europea
AGENCIA ESTATAL DE INVESTIGACIÓN

Earth Sciences Department

BSC Barcelona Supercomputing Center Centro Nacional de Supercomputación