Geophysical Research Abstracts
Vol. 14, EGU2012-8381, 2012
EGU General Assembly 2012

NEMO Oceanic Model Optimization

I. Epicoco (1,2), S. Mocavero (2), A. Murli (2,3), G. Aloisio (1,2)
(1) University of Salento, Lecce, Italy, (2) Euro Mediterranean Center for Climate Change, (3) SPACI Consortium

NEMO is an oceanic model used by the climate community for stand-alone or coupled experiments. Its parallel implementation, based on MPI, limits the exploitation of the emerging computational infrastructures at peta and exascale, due to the weight of communications. As case study we considered the MFS configuration developed at INGV with a resolution of 1/16° tailored on the Mediterranenan Basin.
The work is focused on the analysis of the code on the MareNostrum cluster and on the optimization of critical routines.
The first performance analysis of the model aimed at establishing how much the computational performance are influenced by the GPFS file system or the local disks and wich is the best domain decomposition. The results highlight that the exploitation of local disks can reduce the wall clock time up to 40% and that the best performance is achieved with a 2D decomposition when the local domain has a square shape.
A deeper performance analysis highlights the obc_rad, dyn_spg and tra_adv routines are the most time consuming routines.
The obc_rad implements the evaluation of the open boundaries and it has been the first routine to be optimized. The communication pattern implemented in obc_rad routine has been redesigned. Before the introduction of the optimizations all processes were involved in the communication, but only the processes on the boundaries have the actual data to be exchanged and only the data on the boundaries must be exchanged. Moreover the data along the vertical levels are "packed" and sent with only one MPI_send invocation. The overall efficiency increases compared with the original version, as well as the parallel speed-up. The execution time was reduced of about 33.81%.
The second phase of optimization involved the SOR solver routine, implementing the Red-Black Successive-Over-Relaxation method. The high frequency of exchanging data among processes represent the most part of the overall communication time. The number of communication is triggered by the size of an extra-halo region assigned to each parallel processes. To efficiently exploit the implementation of the SOR algorithm, the selection of the optimal value of the extra-halo size, to trade-off computation increasing with communication decreasing, is required. A performance analytical model has been designed and used to analytically set the optimal value for the extra-halo size. The use of best value for the extra-halo reduced the total execution time of about 5% when the number of MPI tasks increases. Moreover an optimized version, overlapping computation and communication of the solver, has been implemented and tested.
The third phase of optimization regarded the introduction of a further level of parallelism along the vertical levels. An hybrid approach, using the OpenMP has been implemented with a considerable improvement of the parallel speed-up. The benefits derived from the hybrid parallelization is more evident when the number of MPI processes exceeds 30. The minimum wallclock time happens on 396 cores: it is reduced of about 13.78%.