Funded by the European Union

Co-ordinated by ECMWF

ESCAPE 2

ECMWF
DKRZ Deutsches Klimarechenzentrum
Max-Planck-Institut für Meteorologie
MeteoSwiss
Barcelona Supercomputing Center – Centro Nacional de Supercomputación
cea
Loughborough University
RMI
Politecnico Milano 1863 – MOX
DMI
cmcc – Centro Euro-Mediterraneo sui Cambiamenti Climatici
Bull – atos technologies

ESCAPE 2

Funded by the European Union

Weather and climate benchmarks: HPCW

Final Dissemination Workshop
September 03, 2021

Erwan Raffin, David Guibert – Atos

Ralf Mueller – DKRZ

Michiel Van Ginderachter – RMI

Andreas Mueller, Olivier Iffrig – ECMWF

Daniel Beltran – BSC

Outline

- HPCW overview
- Presentation of the HPCW framework
- Definition of quality measures across hardware and software
- Workflow analysis with Kronos
- Results and Perspectives
- Link with the Centre of Excellence in Simulation of Weather and Climate in Europe - ESiWACE2
- Questions

Funded by the European Union

ESCAPE 2

Funded by the European Union

High Performance Climate and Weather benchmark suite

Overview

Funded by the European Union

HPCW v1 overview

- High Performance Climate and Weather (HPCW) benchmark for (pre)-exascale application of climate and weather codes
  - Set of relevant and realistic, near-operational weather forecast workloads
  - Climate science workflows
  - Same usage as HPL or HPCG

- HPCW components
  - Models
    - ICON Ocean and Atmosphere
    - ICON Atmosphere GPU (from DSL implementation)
    - IFS (RAPS)
    - NEMO
  - Dwarfs
    - IFS atmosphere FV dwarf (IFS-FVM)
    - Radiation dwarf (ACRANEB2)
    - ICON ocean advection dwarf
    - ecRad
  - Workload Simulator
    - Kronos

- Test configurations
  - ICON Ocean: Small (160km), Medium (40km) and big (10km), 3-points on a strong scaling line
  - ICON Atmosphere: Small (160km)
  - ICON Atmosphere GPU: Small (160km)
  - ICON coupled atmosphere and ocean (160km)
  - ICON Ocean advection dwarf: Small
  - IFS (RAPS): Small (TL159), Medium (TCo639), Big (TCo1999)
  - NEMO: Small BENCH1 (ORCA1 like) Medium (ORCA0,25)
  - IFS-FVM: Small (O160), Medium (O640), Big (O1280)
  - ACRANEB2: Small
  - Kronos: Single-serial, single-parallel, multi-serial-events, external-job
  - ecRad: Small

ESCAPE 2

Funded by the European Union

Dwarfs in HPCW

components:	options:	ocean	atmosphere	global	regional	D1.7	D1.8	HPCW	Component
discretisation	spectral transform*		✓	✓	✓			✓	**IFS**
	finite volume	✓	✓	✓	✓			✓	**IFS-FVM**
	discontinuous Galerkin	✓	✓	✓	✓		✓	✓ (grey)	
time-stepping	multigrid elliptic solver	✓	✓	✓	✓	✓			
	fault tolerant elliptic solver	✓	✓	✓	✓		✓		
	horizontal explicit, vertical implici	✓	✓	✓	✓			✓	**IFS / ICON**
advection	semi-Lagrangian		✓	✓	✓	✓		✓	**IFS / ICON**
	MPDATA*	✓	✓	✓	✓	✓		✓	**IFS-FVM**
	MUSCL	✓	✓	✓	✓	✓		✓	**NEMO**
physics	CLOUDSC microphysics*		✓	✓	✓	✓		✓	**IFS**
	ecRad radiation		✓	✓	✓			✓	**ecRad**
	ACRANEB2 radiation*		✓	✓	✓	✓		✓	**ACRANEB2**
	machine learned radiation		✓	✓	✓		✓		

grey: work in progress, *from ESCAPE 1

work-steps for each dwarf: isolation into self-contained prototype, documentation, adaptation to different hardware, maintenance of repo

ESCAPE 2

Funded by the European Union

High Performance Climate and Weather benchmark suite

Framework overview

ECMWF DKRZ Deutsches Klimarechenzentrum Max-Planck-Institut für Meteorologie MeteoSwiss Barcelona Supercomputing Center Centro Nacional de Supercomputación cea Loughborough University RMI Politecnico Milano 1863 MOX DMI cmcc Centro Euro-Mediterraneo sui Cambiamenti Climatici Bull atos technologies

Framework overview

Agnostic, customizable and parallel

- HPCW is a CMake-based framework
  - able to compile all the components on top of their own build system
  - agnostic to each code build system (autotools, Makefile, CMake, …)
  - agnostic to each cluster environment (compilers/libraries version…)
  - agnostic to each scheduler system (slurm, …)

- Cluster specifics are defined in a “toolchain” file:
  - Compilers
  - Default flags
  - Template job file

- Parallel compilation

ECMWF DKRZ Max-Planck-Institut für Meteorologie MeteoSwiss Barcelona Supercomputing Center cea Loughborough University RMI POLITECNICO MOX DMI cmcc Bull atos technologies

Framework overview

Multi model and multi dependencies management

- Dependencies management
  - Can be provided by the system
  - Will be built with different versions of the same dependencies handled
  - SPACK / others to provide dependencies

- Support both online and offline cluster transparently
  - Dependencies are searched into the folder “downloads”
  - If not already there, they are downloaded
    - from upstream server unless “OFFLINE=ON”
    - optionally as a package hosted in the ECMWF FTP server
    - from SVN server unless “DISABLE_SVN_CHECKOUT=ON”
  - And their integrity checked

- To mirror the required source files/directory, a convenient script has been developed
  - cmake -P populate-downloads.cmake

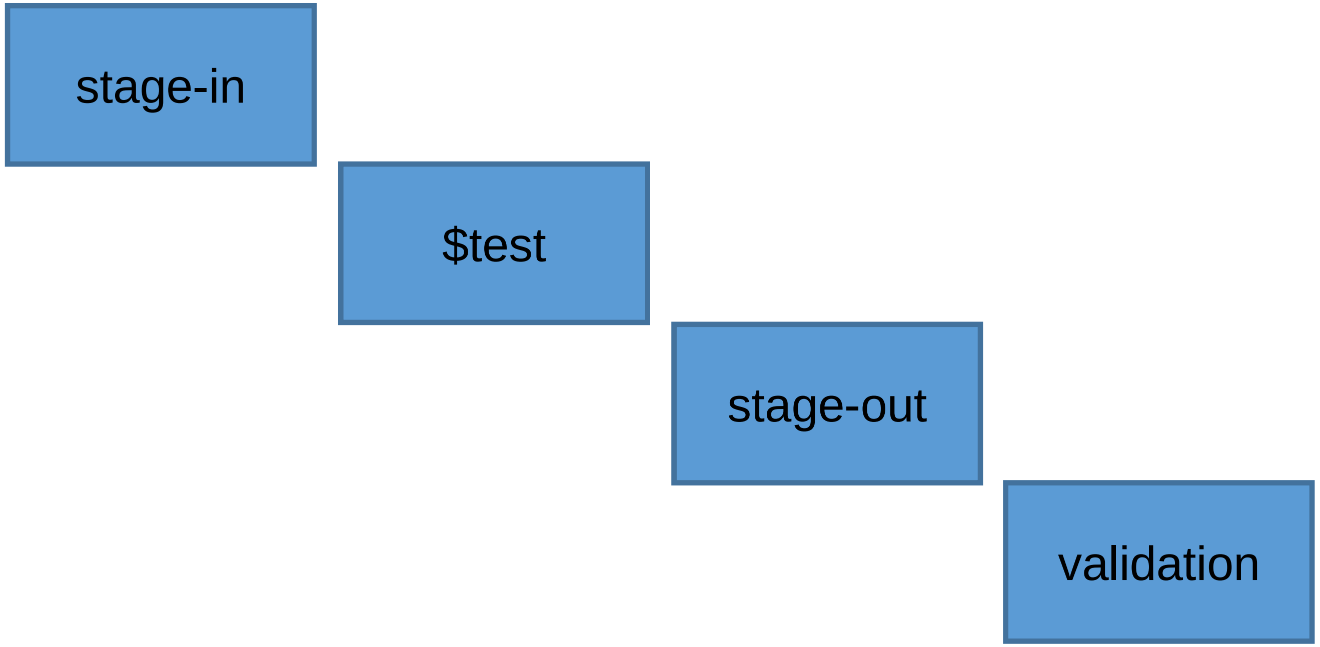
Funded by the European Union

Framework overview

Test case steps

- A test may require additional steps
  - Stage-in: preparation of input files (symlinks/copies)
  - Stage-out: removing of unneeded output files
  - Verification/validation

  - The steps dependencies are handled by CMake (CTest)



Framework overview

Customization

- Environment, toolchain and job launcher are setup in 3 different files

- The environment file (like *bull-intel.env.sh*) is used to enter within a configured shell to compile HPCW:
    - Set the compiler paths
    - Define CC, CXX, FC
    - Define common flags
        - CFLAGS/FFLAGS/LDFLAGS…
        - cmakeFlags
    - Load modules or set PATH/CPATH/LIBRARY_PATH/…

- The toolchain file (like *bull-intel.cmake*) defines CMake variables according to the environment
    - compilers and associated flags
    - any CMake-variables needed on your configuration could be defined here…
    - job launcher

- Job launcher file (like bull-job-launcher.sbatch) defines for each supported cluster/test case
    - Launcher: srun/mpirun/…/profiler/…
    - Options and runtime environment variables: OMP_NUM_THREADS…

ESCAPE 2

Funded by the European Union

High Performance Climate and Weather benchmark suite

Definition of quality measures across hardware and software

ECMWF DKRZ Max-Planck-Institut für Meteorologie MeteoSwiss Barcelona Supercomputing Center cea Loughborough University RMI POLITECNICO MILANO MOX DMI cmcc Bull atos technologies

Funded by the European Union

Verification routine

- Ensemble-based verification routine based on existing literature was implemented and tested in the context of HPCW
- Method proved to be useful for existing test cases regarding quality and energy efficiency
- Preparation of verification data is very specifc to the test case and a considerable investment during the development, ideally this should be done in collaboration with the model developers
- Future versions of HPCW may include verification for more test cases esp. high resolution because energy-to-solution becomes critical
- In-depth analysis D3.2

ESCAPE 2

Funded by the European Union

High Performance Climate and Weather benchmark suite

Workflow analysis with Kronos

ECMWF DKRZ Max-Planck-Institut für Meteorologie MeteoSwiss Barcelona Supercomputing Center cea Loughborough University RMI POLITECNICO MILANO MOX DMI cmcc Bull atos technologies

Workflow analysis with Kronos

- Objective:
  - Provide a procedure to synthesize the Tier-1 models using Kronos and replace components of the synthesized workload by dwarfs
  - Describe and evaluate the first versions of such workloads

- Work performed:
  - Kronos has been extended to
    - execute dwarfs in its workloads
    - report and analyze the runtime statistics that the dwarfs may report
  - A synthetic workload has been generated for IFS, with two configurations:
    - the first fully relying on Kronos' synthetic application, and
    - the second using a mixture of synthetic application and dwarfs mimicking the work of some of the IFS components

Funded by the European Union

High Performance Climate and Weather benchmark suite

Results and Perspectives

Funded by the European Union

Results and Perspective

- Many components were build+run on multiple systems: Atos, DKRZ, BSC, ECMWF, …
- Continuous Integration (CI) is already running for HPCW components at DKRZ – very important for a stable development in the future within ESiWACE2 for example
- Next Steps
  - Enable testing on DKRZ facilities for a fully automated development cycle
  - Extend to all HPCW components
  - Include more HPC system to ensure portability
  - Work on validation of more components
  - Add new HPCW components, especially using GPU, and experiments

Funded by the European Union

Results and Perspective

- Atos/Bull: all components
- BSC: all components
- DKRZ: build and run ICON, NEMO, IFS-FVM, ACRANEB2
- ECMWF: ICON Ocean, IFS-FVM
- Next Steps
  - Port HPCW to upcoming HPC system at DKRZ and ECMWF
  - Port HPCW to pre-exascale EuroHPC systems
- License setup ongoing

Results and Perspective

- Definition of quality measures across hardware and software
  - Tested models come with quite different test scenarios
    - ICON: established scientific ocean model test (OMIP)
    - IFS-FVM: analytic setup for testing the dynamical core
  - In both applications the ensemble-based verification routine could successfully be applied
  - This proves that the technique although developed in the context of climate science is also useful in NWP

ESCAPE 2

Funded by the European Union

Time and Energy to solution

- Current results at Atos Bull
  - Hardware setup
    - BullSequana XH2000
    - CPU: 2x AMD EPYC 7763 64-Core Processor (MILAN)
    - + GPU: 4x Nvidia A100 GPU when needed
    - Memory: 256GB RAM DDR4 3200 MT/s
    - Interconnect: HDR 100 (fat tree)
  - Software setup
    - OS: RHEL 8.3
    - Environments:
      - bull-intel.env.sh
      - bull-intel+mpi+mkl.env.sh
      - bull-intel+openmpi+mkl.env.sh
      - bull-intel19+openmpi+mkl.env.sh
      - bull-spack-icon-nvhpc.env.sh
    - Toolchain:
      - bull-intel.cmake
      - bull-nvhpc.cmake
    - Job launcher
      - bull-job-launcher.sbatch
    - Energy measurement: Bull Energy Optimizer (BEO)

BullSequana X

Direct liquid cooling

BullSequana X2410/5 AMD blades

AMD

ECMWF DKRZ Max-Planck-Institut für Meteorologie MeteoSwiss Barcelona Supercomputing Center cea Loughborough University RMI Politecnico Milano MOX DMI cmcc Bull atos technologies

Time and Energy to solution

- Current results at Atos Bull
  - HPCW help tools
    - Build wrapper (*build-wrapper.sh*) to ease reproducibility and log management
    - Result analysis (*analyse.sh)* to generate the status
    - Retrieve energy to solution with BEO and dedicated script
  - Early baseline results without any optimization

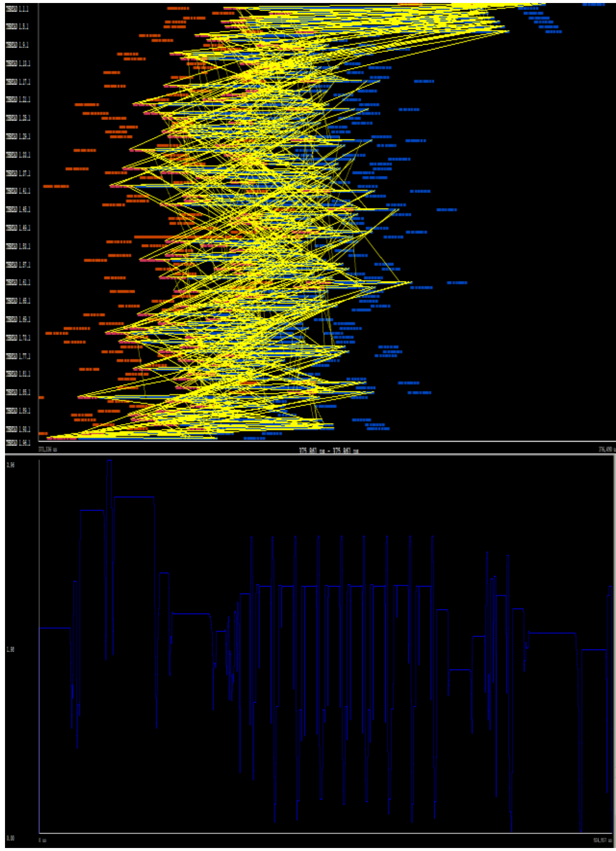
HPCW_component-test_case-type	revision	status	time (s)	time_app	energy (Wh)	Nb Nodes (Cores)
dwarf-p-radiation-acraneb2-lonlev-0.91-small	v0.2-8-g8a1873c43	OK	0,17	-	1,95	1 (128)
dwarf-p-radiation-acraneb2-lonlev-0.9-small	v0.2-8-g8a1873c43	OK	12,28	-	3,14	1 (128)
ecrad-small	v0.2-8-g8a1873c43	OK	0,52	0,29	0,48	1 (128)
icon-atmo-small	v0.2-8-g8a1873c43	OK	53,00	-	7,19	1 (128)
icon-coupled-small-n24	v0.2-8-g8a1873c43	OK	369,76	-	67,52	2 (48)
icon-ocean-advection-dwarf	v0.2-8-g8a1873c43	OK	18,16	-	2,09	1 (128)
icon-ocean-big	v0.2-8-g8a1873c43	OK	1436,46	-	9341,90	50 (6400)
icon-ocean-medium	v0.2-8-g8a1873c43	OK	812,63	-	1186,60	10 (1280)
icon-ocean-small	v0.2-8-g8a1873c43	OK	18,39	-	3,09	1 (128)
icon-atmo-gpu-small	v0.2-8-g8a1873c43	OK	31,71	-	9,47	1 (128)
ifs-fvm-big	v0.2-8-g8a1873c43	OK	10359,26	-	104048,93	60 (7680)
ifs-fvm-medium	v0.2-8-g8a1873c43	OK	3816,17	-	12787,74	20 (2560)
ifs-fvm-small	v0.2-8-g8a1873c43	OK	1024,87	-	167,09	1 (128)
ifs-tco1999-big	v0.2-8-g8a1873c43	OK	1231,45	1182,06	18971,49	120 (15360)
ifs-tco639-medium	v0.2-8-g8a1873c43	OK	349,19	343,17	467,85	10 (1280)
ifs-tl159-small	v0.2-8-g8a1873c43	OK	19,96	16,14	4,52	1 (128)
kronos-single-serial	v0.2-8-g8a1873c43	OK	-	55,47	5,61	1 (1)
kronos-single-parallel	v0.2-8-g8a1873c43	OK	-	542,36	38,46	1 (8)
kronos-multi-serial-events	v0.2-8-g8a1873c43	OK	-	254,83	18,27	1 (1)
kronos-external-job	v0.2-8-g8a1873c43	OK	-	132,89	9,78	1 (1)
nemo-orca25-medium	v0.2-8-g8a1873c43	OK	1415,48	-	554,70	3 (384)
nemo-bench-orca1-like-small	v0.2-8-g8a1873c43	OK	56,89	-	6,70	1 (9)

Funded by the European Union

Deployment at BSC and IFS-FVM profile results

- Current results at BSC marenostrum4
  - All models compiled
  - Hardware setup
    - CPU: 2 sockets Intel Xeon Platinum 8160
    - Total: 3456 nodes and 165888 cores
    - Memory: 12x 8GB RAM DIMM 2667 MT/s
    - Interconnect: 100 Gbit/s Intel Omni-Path
  - Software setup
    - OS: Linux 4.4.120-92.70-default
    - Details:
      - Branch -> db/bsc-config
      - Intel version -> intel/2018.4 (Kronos: 2018.1)
      - IFS-FVM: Profiled with only MPI
    - Environments:
      - bsc-intel_marenostrum4.env.sh
      - bsc-intel_marenostrum4+kronos.env.sh
    - Toolchain:
      - bsc-intel_marenostrum4.cmake
    - Modifications:
      - kronos-slurm.bsc.py
      - Icon wrapper: marenostrum4.intel-18.0.4
    - Job launcher
      - bsc-job-launcher.sbatch

- [Summer school] Profile analysis of HPCW included model: IFS-FVM
  - Software used:
    - Extrae: A package to instrument the model which gives trace-files with HW counters, MPI messages and other info.
    - Paraver: a powerful trace browser.
    - Dimemas: Allows to simulate a trace in an optimal environment.
  - IFS-FVM traces configuration:
    - MPI + Papi counters + Call stacks
    - 1node*48tasks + 2nodes*48 tasks + 4*48 tasks
  - IFS-FVM results:
    - Load balance issues
    - MPI overhead may affect IPC



ESCAPE 2

Funded by the European Union

High Performance Climate and Weather benchmark suite

Link with ESiWACE2

Link with ESiWACE2

Funded by the European Union

- Presentation at 15:45 "**Uptake of ESCAPE-2 by ESiWACE-2**" by Florian Ziemen and Joachim Biercamp

ESCAPE 2 → HPCW → esiwace

CENTRE OF EXCELLENCE IN SIMULATION OF WEATHER AND CLIMATE IN EUROPE

ESCAPE 2

Funded by the European Union

Questions?

The ESCAPE-2 project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 800897.
This material reflects only the author's view and the Commission is not responsible for any use that may be made of the information it contains.