EUROPEAN COMMISSION

Community research

is-enes

INFRASTRUCTURE FOR THE EUROPEAN NETWORK FOR EARTH SYSTEM MODELLING

IS-ENES2 DELIVERABLE (D -N°: 6.4)

Report on Collaboration between ENES and HPC companies

{File name: IS-ENES2_D6.4.pdf}

Authors: ***Kerstin Fieg***
Joachim Biercamp

Reviewers: ***Reinhard Budich***
Marie- Alice Foujols

Reporting period: ***01/04/2016 – 31//03/2017***

Release date for review: ***17/03/2017***

Final date of issue: ***30/03/2017***

Revision table			
Version	**Date**	**Name**	**Comments**
0.1	21.02.2017	K. Fieg	1st Draft
0.2	03.03.2017	R. Budich / K. Fieg	1st Review / 2nd Draft
0.3	09.03.2017	J. Behrens / K. Fieg	2rd Review / 3rd draft
0.9	21.03.2017	J. Biercamp / K. Fieg	3rd Review / 4th draft
1.0	17.03.2017	M.-A. Foujols	Review

Abstract

The deliverable documents interactions of the IS-ENES community with European and international hardware vendors concentrating on two events, taken place in 2014 and 2016 / 2017. We summarize the findings from the discussions and try to draw conclusions on how we think this would influence the strategy of building and using the next generation of Earth System Models for future generation (Exascale) hardware systems.

We conclude that most Earth System Model codes as they are in use today will not be able to efficiently leverage the potential performance gain of next generation HPC systems. The bottlenecks, which we think will require refactoring and rewriting of code, are discussed shortly.

Project co-funded by the European Commission's Seventh Framework Programme (FP7; 2007-2013) under the grant agreement n°312979

Dissemination Level		
PU	Public	**X**
PP	Restricted to other programme participants including the Commission Services	
RE	Restricted to a group specified by the partners of the **IS-ENES2** project	
CO	Confidential, only for partners of the IS-**ENES2** project	

Table of contents

Executive Summary

Representatives of seven European and international hardware vendors (BULL, CRAY, INTEL, IBM, NEC, NVIDIA, DDN) gave overview talks (partly under NDA) on how they expect the next generation hardware landscape to look like.
Starting from this prognosis this deliverable aim at summarizing the impacts the foreseeable hardware development would likely have on the ESM codes in use today.
It is clear, that using today's codes without any adaption does not allow exploiting future performance potential. On the other hand comprehensive code refactoring or rewriting would require a massive investment of money and manpower for every code while today it is still unclear how to leverage performance of next generation hardware with our applications. Thus to make this investment sustainable, we suggest to address the foreseeable bottlenecks and to solve them under consideration of the concept "separation of concerns".
Not only with respect to the new design of the hardware, but equally important on the non-hardware side the general focus of the hardware vendors will increasingly shift to serve large and still growing market segments with applications that differ significantly from ours. From this development, we expect a reduced engagement in support of specific tools, compilers and libraries used for Earth System Modelling, or at least higher support costs and slower response times related to maintenance.

1. Introduction

One target of IS-ENES2 is to follow the technical innovations in general happened during project runtime of IS-ENES2 and discuss, if and how they may influence the climate modelling community. Concrete there is the task “Technology tracking” led by CMCC in WP2 / NA2 that is looking at activities regarding exascale issues and WP6 / NA5 that intents to foster collaboration between HPC manufacturers and IS-ENES2. The target is to link the findings resulting from this collaboration to the community and influence the work on the update of the ENES infrastructure strategy.

Three pillars of collaboration within IS-ENES2 are related to this:

- Liaison of climate research community and ICT companies.
- Liaison of climate research community and consultancies and corporates
- Liaison of climate research community and Climate Services

The present report (Deliverable D6.4) reports on interactions with HPC manufacturers / ICT companies.

In section 2 we will first draft technical and strategic developments the vendors announced in their roadmaps in 2014 and 2017, in section 3 we will sum up the current state of discussion on the implications this roadmaps will have for climate modelling strategies, finally in section 4 we will briefly describe possible approaches for future collaboration with the HPC manufacturers developed during the two dedicated workshops.

2. Roadmaps on Technical Innovations – the vendors view

While IS-ENES2 Deliverable 3.1 summarized the (still on – going) discussion on what next generation compute software, hardware and architecture would imply for the designs of next generation global climate models, D6.4 aims to give an overview – and thus update D3.1 - on what relevant hardware vendors plan in their roadmaps for the near future. For this the IS-ENES proposal scheduled two opportunities of interaction with ICT companies, first a dedicated session in the 3rd ENES HPC workshop and second a series of talks with vendors given under NDA.

2.1 HPC Hardware challenges & solutions for the climate community, 19.3.2014

First session of the workshop described future trends in climate science & related HPC challenges. Session seven of the "*Workshop on Exascale Technologies & Innovation in HPC for Climate Models"* (Agenda see Annex A) was dedicated to "*HPC Hardware challenges & solutions for the climate community*".

Talks were given by high level technical staff from Intel, DDN, BULL/Atos, IBM, Cray, NEC and NVIDIA. The topics focused on hardware developments planned for the following couple of years, on co-design and cooperation with climate scientists or on companies' general perspectives on HPC and extreme scale computing.

Slides of the presentations can be found at IS-ENES2 portal

https://verc.enes.org/ISENES2/events/ws3

This workshop encouraged us to continue to follow hardware evolution in HPC systems , to stay open in term of software development and to share information, in particular via the ENES HPC task force. .

2.2 Series of Talks on Innovative Technologies, Nov. 2016 – Jan. 2017

As a follow-on of the 2014 presentations above, Deutsches Klimarechenzentrum (DKRZ) and Max-Planck-Institute for Meteorology (MPI-M) invited NEC, Cray Ltd., Bull / Atos, IBM, Intel and NVIDIA to present an update on their roadmaps.

The presentations took place between November 2016 – January 2017. This time the companies were explicitely asked to talk about planned developments for a longer timeframe - until 2020 – 2025.

A core element of the workshop was the intensive interaction of scientists and vendors to get an bidirectional understanding of the following points:

1. The vendors hardware and system strategies
2. The consequences this may have for software development and the design of next generation climate models
3. To make the vendors aware of the special requirements that the weather and climate modelling community has
4. To encourage the discussion about co-design projects

Dates

- 30.11.2016 NEC
- 01.12.2016 Cray
- 10.01.2017 Bull -> postponed
- 11.01.2017 IBM
- 12.01.2017 NVIDIA
- 13.01.2017 INTEL

The audience, which included representatives of the ENES HPC task force, signed an NDA (Non Disclosure Agreement), to allow the vendors to present non-public information. Therefore, here only a general conclusion on the non-disclosed parts of the events can be reported and slides will not be available on the IS-ENES2 portal for the public.

2.3 Changing focus during IS-ENES2 runtime

At the beginning of IS-ENES2, in April 2012, a paper was written and presented to the community (https://enes.org/community/about-enes/the-future-of-enes/the-future-of-enes) that discussed infrastructure strategy for Earth System modelling and can be understood as starting point for the discussion on infrastructure changes.
At the end of the runtime of the IS-ENES2 project this strategy paper will be updated and supplemented with some of the discussion points from the perspective of innovation technology to integrate them into a broader strategy perspective.

HPC Computing / Exascale Systems:
Already at beginning of project runtime of IS-ENES2 vendors, EC and scientific community discussed about Exascale computing systems and speculate how such a system would look like. Now, the roadmaps presented by the vendor in 2016 / 2017 can be seen as blueprints of what the community has to expect.

From the vendors' presentations it seems to be clear, that the computational complexity will increase (left column of table). This will lead to following software bottlenecks for ESM codes (right column of table).

Vendors' perspective	**Consequences for strategy in ESM**
Heterogeneity of systems (mixture of CPU / GPU / accelerators) will increase	Increasing number of abstraction in memory levels have to be expected and will require serious code adaption to use the systems efficiently. This could include delegating system adaption to tools like OpenMP, DSL and/or subdivide models into software layers (abstraction) on top of optimized kernels (tuning level).
Increasing number of cores per socket with decreasing amount of local memory per core	Performance improvements can only be achieved by a higher level of parallelism, which then will increase code complexity, will substantially impede code development and efficient use of the
Decreasing Memory Bandwidth and	

network bandwidth per core	new systems. Higher parallelism with at the same time minimal communication will be required and might narrow the choice of algorithms available depending on respective communication patterns.
Clock rate of cores can not be expected to increase in the same rate as in the past. On the contrary, it will probably decrease.	
Hardware development will increasingly focus the mass market (e.g. mobile phones) and emerging new market segments (e.g. deep learning) and decreasingly serve classical application areas with special requirements (e.g. climate science)	This may lead to reduced engagement in supporting special toolkits, compiler, libraries, higher support costs or to slower response times when it comes to bug fixing etc.

3. Implications for future climate modeling strategies

3.1 Consequences

The science complexity is continuing to increase and weather and climate research demands still more compute power to improve ensemble size, time or spatial resolution or the approximation of physical processes.

In their presentations the vendor's impressively illustrated the technical potential of future Exascale computer systems. Now, for specialized communities like climate science the challenge will be to gain the benefit from these new developments – which will only be possible, when we accept that the general computational complexity will increase and find a concept to deal with it.

Deliverable D3.1 (written in September 2014) discussed several technical challenges, which have to be solved to fully benefit from next generation computer systems (e.g. scalability, need to improve performance of codes, load imbalance), and one of the challenges mentioned there has to be highlighted again:

- Next generation compute systems will integrate different types of processors, and will often be equipped with co-processors and / or accelerators (e.g. GPGPUs).

This has implications for the design of weather and climate codes, because it will result in ever more levels of concurrency to meet performance requirements with the anticipated hardware. Currently we expect at least seven types of parallelism that need to be handled by the applications.

- Vectorisation
- Distributed memory concurrency across nodes / sockets (MPI)
- Shared memory concurrency within nodes / sockets (OpenMP)
- Hardware accelerators (GPUs, used with, e.g., OpenACC, OpenMP)
- External coupling of independent models (Oasis MCT, Cpl, YAC)
- Asynchronous I/O (XIOS, I/O servers)
- Executing model components concurrently (ESMF, FMS)
- Execution in parallel of different members of ensembles

Note, that leveraging massive concurrency proved to be very difficult and often not possible with climate codes on some hardware (IBM BlueGene for example).

3.2 What does this mean in praxis?

ESM codes are complex and their lifetime is much larger (+20 years) than the lifetime of typical hardware architectures (3 – 6 years). Thus the question of the practicability of restructuring an ESM code with millions of lines of code and hundreds of subroutines to make it run on a special hardware with maximum performance has to be discussed carefully.
On the other hand the hardware landscape is changing and codes have to be adapted to a hybrid programming model – e.g based on the use of the message passing combined with a shared memory programming model - to fully exploit emerging hybrid architectures and all the above mentioned levels of parallelism.
This porting work will be a huge investment, which for some codes might e.g enable the efficient use of accelerators or co- processors but surely will need to allocate a reasonable amount of manpower and time to rewrite codes. However, this strategy is in sharp contrast to the traditional view to fulfil the continuous need to improve the model's scientific performance concurrently with increasing the real time execution performance.

4. *Separation of Concerns* as recommendation for future ENES HPC strategy

One solution discussed within the HPC community and also in IS-ENES2 is to follow the design principle called "separation of concerns (SoC)". SoC describes the general strategy of splitting a problem into distinct subsections or modules in a way, that every subsection solves a well-defined set of issues and communicates with the context by well-defined interfaces. This strategy of encapsulation hides the low level complexity from the user, who then only has to pass information via given interfaces to solve his or her problem and treats the details of problem solving as a black box. An example for adopting this principle to code in use was presented at the "*Workshop on Exascale Technologies & Innovation in HPC for Climate Models"* in the talk of Th. Schulthess (ETH Zürich)[1] during session 1 (see annex).
This strategy reflects the traditional way of code development and modular programming.
In our case SoC can support to structure the overall workflow into defined layers of hard- and software dependencies and solve them separately. If then one of the elements in the workflow (e.g. hardware, compiler, program language) is changed, only the corresponding module has to be updated. The assumption is, that this leads to more flexible, portable, hardware and software environment independent workflows. It is hoped that vendors would contribute by maintaining a manageable amount of lower level modules allowing to efficiently interface higher level applications to new hardware architectures without having to change the application itself.

Identified core elements of this strategy

- Separate science code (and science code development) from code parallelisation and code optimisation
- Enable parallelisation and optimisation for portable performance
- Use domain specific expert knowledge for parallelisation and optimisation

[1] https://is.enes.org/documents/Talks/WS3HH/session-1-future-trends-in-climate-science-related-hpc-challenges/thomas-schulthes

  - Higher level specifications of structures / interfaces
  - Domain specific language (DSL) / abstraction libraries
- Use scientific expert knowledge for physical code development
  - Keep single science source
- Define fundamental algorithm building blocks for co-design with vendors / different institutions for benchmark and co-develop code pieces for next generation ESM models and next generation HPC architectures

5. Co – Design strategies between ENES and HPC manufacturers

The science community is optimising the codes in use on their respective local machines at the institutions. The vendors get the codes – or critical parts of the codes - as procurement benchmarks. Because benchmarking is very expensive for vendors, they usually look into full applications only for procurement purposes and then they have to invest a lot of time to make the (mostly unknown) codes run on the machines, they want to sell.
With this respect we think co-design based on a well-defined representative set of application might be of benefit for both: the research institutes and the HPC vendors:

Vendors	Scientific community
Vendors have to propose usable computers for climate simulation. (avoid a BlueGene like machine)	Users have to describe their needs: languages, libraries, concurrent execution specification, ...
Well tested vendor software (e.g., compiler) that can handle scientific requirements	Users do not have to debug the given software environment
Contact person on science side to get access to code and support	Contact person at vendors side to address upcoming requirements
Continuous work on the codes would reduce the time to get an optimal benchmarking result	A short cut to experts on the vendors side to address compiler bugs, features and establish interaction
Having access to real in use code, allows bug fixing and compiler updating in an early development stage - which might reduce cost by cross dependencies	Access most recent compiler version would allow optimising code in an early development stage and feed back bugs to vendors.
Vendors could follow the community experience with advanced technics like DSL – a method that could also be interesting for other science communities	Scientists can test innovative technics on innovative hardware and make it fit for future hardware produces
Innovative hardware can be offered more tailored	Better integration of innovative hardware products into scientific workflow

First steps to establish application based co-design:

1. Revive RAPS (Real Application on Parallel Systems) idea

- The goal is setting up portable benchmarks of real applications with proper documentation on how to run the benchmarks
- RAPS could be open for codes with versioning system, available for all vendors
- Provides applications not only in procurements, but to allow vendors to have a much more elaborate insight on the developments going on

2. Share lessons learned (among institutes) about how to do procurements
    - On strategies, methods, procedures, outcomes

3. Continue the cooperation between ENES participants through ESIWACE a Center Of excellence project founded with support of the ENES HPC task force.

6. Conclusions and Summary

The primary conclusion of the findings above is, that it will be difficult to benefit from the increase of compute power of the next generation HPC machine with new architecture with the software we have today: to a reasonable extent our software is not fit to use the new systems.

It is foreseeable, that pressure on the community to develop concepts to solve the problem is growing over time and up to now possible solutions are still under discussion.
Nevertheless, a point where we have a broad consensus is that a structure like ENES will be increasingly important during the time of discussion and afterwards during the time of implementation of the decided strategy. We think ENES will help to address the problems of the community and consequences discussed here with the European Commission.

A second point of action discussed here is the question of co-design with vendors. We think the climate community should intensify the efforts with HPC industry to make sure, that they are aware of our special needs and understand the consequences of the technical hardware changes. We think co-design can help to ensure, that future compute hardware environments and software fulfils at least the basic needs of the community in a way that libraries, compilers and program languages as FORTRAN are continuously developed and supported.
Finally, the community will continue this work together in particular through the activities of the ENES HPC task force and supported by the ESIWACE CoE..

Glossary

DSL	Domain Specific Language
ENES	European Network for Earth System Modelling
ESM	Earth System Modelling
HPC	High Performance Computing
ICT	Information- and Communication Technologies
NDA	Non Disclosure Agreement
RAPS	Real Application on Parallel Systems
SoC	Separation of concerns
WP	Work Package

Annex

Schedule: Exascale Technologies & Innovation in HPC for Climate Models, 17. - 19. 3. 2014

ENES Workshop on

Exascale Technologies* & "*Innovation in HPC

for Climate Models

C ap San Diego, Überseebrücke, 20459 Hamburg

	Monday March 17th	speakers, chairs & moderators
10:00-11:00	***Registration***	
11:00-11:30	**Welcome session**	
	Welcome and opening remarks	Joachim Biercamp (DKRZ, DE)
	Introduction to (IS)ENES. Review of workshops 1 & 2 (Lecce / Toulouse)	Sylvie Joussaume (IPSL, FR)
11:30-15:00	**Session 1 – Future trends in climate science & related HPC challenges**	Reinhard Budich (MPIMet, DE)
11:30-12:00	Scientific challenges in climate modeling	Jochem Marotzke (MPIMet, DE)
12:00-12:30	Scalable Software Developement for Climate Models (tbc)	Thomas Schulthess (ETH & CSCS, CH)
12:30-13:00	Overview from US	Venkatramani Balaji (Princeton Univ. & GFDL, US)
13:00-14:00	***Lunch***	
14:00-14:30	Refactoring CESM for exascale	Rich Loft (NCAR, US)
14:30-15:00	The Upscale project	Pier-Luigi Vidale (Univ Reading & NCAS,UK)
15:00-17:30	**Session 2 - Status of EU Exascale projects**	Marie Alice Foujols (IPSL, FR)
15:00-15:25	CRESTA, Collaborative Research Into Exascale Systemware, Tools & Applications	Erwin Laure (KTH Royal Institute of Technology, SE)
15:25-15:50	DEEP, Dynamical Exascale Entry Platform	Hendrik Merx (MPIC & CYI, DE)
15:50-16:15	***Break***	
16:15-16:40	MONT-BLANC, European scalable and power efficient HPC platform based on low-power embedded technology	Paul Carpenter (BSC, SP)
16:40-17:05	EESI 2, European Exascale Software Initiative	Phillipe Ricoux (Total, FR)
17:05-17:30	EXA2CT, EXascale Algorithms & Advanced Computational Techniques	Marie-Christine Sawley (Intel FR)
17:30-18:30	***General Discussion on Session 1 & 2***	Sylvie Joussaume (IPSL, FR)
18:30	**End of the 1st Day**	

	Tuesday March 18th	speakers, chairs & moderators
8:30-9:30	**Session 3 - Status of EU G8 projects**	Giovanni Aloisio (CMCC, IT)
08:30-08:50	G8 ESC, Enabling Climate Simulations at Extreme Scale	Rich Loft (NCAR, US)
08:50-09:10	ICOMEX, ICOsahedral-grid Models for EXascale Earth system simulations	Julian Kunkel (DKRZ,DE)
09:10-9:30	EXARCH, climate analytics on distributed EXascale data ARCHives	Martin Juckes (BADC, UK)
9:30-14:00	**Session 4 - HPC Software challenges & solutions for the climate community**	Uwe Fladrich (SMHI, SE)
9:30-9:50	The use of GPU in Climate models	Will Sawyer (CSCS, CH)
9:50-10:10	Porting the COSMO model to GPUs	Xavier Lapillonne (MeteoSuiss, CH)
10:10-10:40	***Break***	
10:40-11:00	Experiences with XEON PHI in the Max-Planck-Society	Markus Ramp (RZG, DE)
11:00 -11:20	Results on XEON PHI at GFDL	Christopher Kerr (GFDL, US)
11:20-11:40	Experiences with MIC at the MetOffice	Christopher Maynard (MetOffice, UK)
11:40-12:00	Why Compilers and workflow matter	Luis Kornblueh (MPIM, DE)
12:00 -12:20	Numerical Libraries and Framework (PETSc)	Jed Brown (Argonne,US)
12:20-13:20	***Lunch***	
13:20-13:40	Performance tools (Paraver/Dimemas)	Jesus Labarta (BSC, SP)
13:40-14:00	The ECMWF Scalability Project	Peter Bauer (ECMWF,UK)
14:00-15:30	**Session 5 - New Parallel Approaches at Exascale**	Graham Riley (UniMan, UK)
14:00-14:30	Hybrid Programming	William Gropp (UIUC,US)
14:30-15:00	Communication-Avoiding Algorithms	Laura Grigori (INRIA, F)
15:00-15:30	Talk on space-time parallelization	Yvon Maday (UPMC, F)
15:30-16:00	***Break***	
16:00-17:00	***General discussion on sessions 3, 4 & 5***	Reinhard Budich (MPIMet)
17:00 - 18:30	**Session 6 – Working session on performance intercomparisons of climate models**	Jean-Claude Andrè (F)
17:00-17:10	Introduction (output from the Toulouse workshop)	Jean-Claude André
17:10-17:20	Performance measurements of HPC-applications at LRZ	Gilbert Brietzke (LRZ, DE)
17:20-17:30	A metric for computational performance based on SYPD	Balaji (GFDL, US)
17:30-18:15	***Open Discussion***	
18:15-18:30	Summary	Sylvie Joussaume
18:30	**End of the 2nd Day**	
19:30	**Dinner at Gröniger Brauhaus (Microbrewery)**	

	Wednesday March 19th	speakers, chairs & moderators
08:30-11:30	**Session 7 - HPC Hardware challenges & solutions for the climate community**	Joachim Biercamp (DKRZ, DE)
08:30-08:55	Weather and Climate roadmap to extreme scale: the Intel perspective	Marie-Christine Sawley (Intel, F)
08:55-09:20	Designing a Highly Redundant Maintenance and Distribution System for Critical Research Data	Dave Fellinger (DDN, US)
09:20-09:45	A use case (based on NEMO)	Damien Declat & Franck Vigilant (Bull SA)
09:45-10:10	Load unbalance : a major bottleneck for climate applications on exascale systems	Francois Thomas & Christoph Pospiech (IBM)
10:10-10:40	***Break***	
10:40-11:05	Architectures for Extreme Scale Earth System Modeling	Per Nyberg (Cray)
11:05-11:30	SX-ACE technology and future visions	Rudi Fischer (NEC)
11:30-11:50	NVIDIAS perspective on EXASCALE	Stan Posey (NVIDIA)
11:50-14:35	**Session 8 - Porting Climate Codes on top-of-the edge machines**	Sophie Valcke (CERFACS, FR)
11:50-12:10	From Gung Ho to LFRic - replacing the Met Office Unified Model	Steve Mullerworth (Met Office, UK)
12:10-12:30	ICON for HD(CP)2 (High definition clouds and precipitation for climate prediction)	Panagiotis Adamidis (DKRZ, DE)
12:30-12:50	The SPRUCE Project	Eric Maisonnave (CERFACS, F)
12:50-13:50	***Lunch***	
13:50-14:10	The HiResClim and SPECS Projects	Francisco J. Doblas-Reyes (IC3)
14:10-14:30	DYNAMICO	Yann Meurdesoif (IPSL, FR)
14:30-15:15	***General discussion on sessions 7 & 8***	Bryan Lawrence (NCAS, UK)
15:15-17:00	**Session 9 - Center of Excellence on Climate**	Sylvie Joussaume (IPSL, FR)
15:15-15:45	Status of Commission initiatives	Sylvie Joussaume (IPSL, FR) and participants from task force
15:45-16:45	***Open Discussion***	
16:45-17:00	*Summary and further actions*	Sylvie Joussaume (IPSL, FR)
17:00	**End of the Joint Workshop**	