

PROCESS

PROviding Computing solutions for ExaScale ChallengeS

A hybrid platform for exascale
multidisciplinary scientific applications

Marian Bubak, Piotr Nowakowski, Jan Meizner, Martin Bobák, Ondrej Habala, Ladislav Hluchý, Viet Tran, Adam S. Z. Belloum, Reginald Cushing, Maximilian Hüb, Dieter Kranzlmüller, Jan Schmidt

Supercomputing Frontiers Europe 2020, 23-25 March 2020, Warsaw, PL

- **Vision**

- As the result of research we will deliver a comprehensive set of mature services prototypes and tools specially developed to enable extreme scale data processing in both scientific research and advanced industry settings

- **Principles**

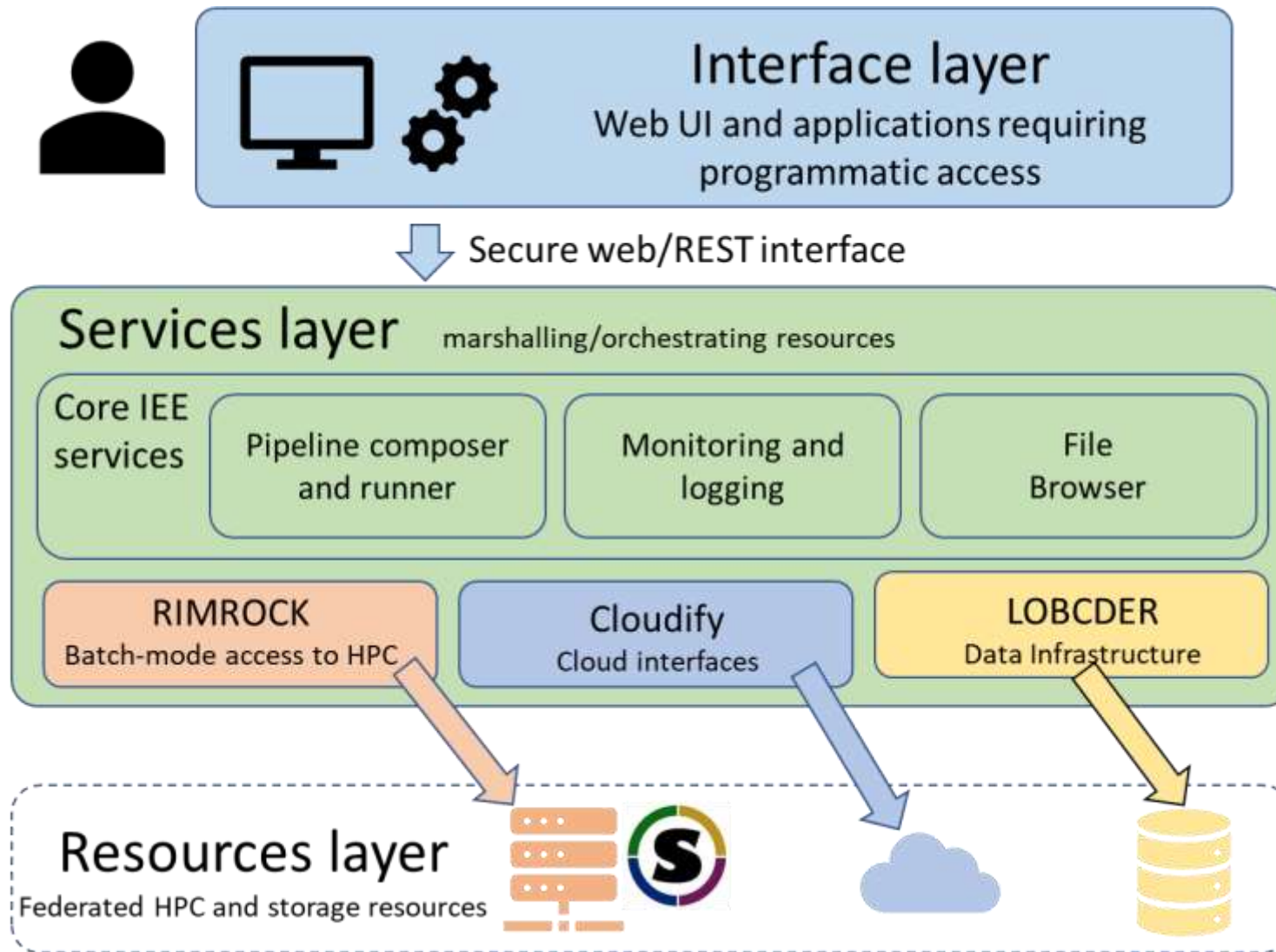
- Leapfrog beyond the current state of the art
- Ensure broad research and innovation impact
- Support the long tail of science and broader innovation

- **Concept**

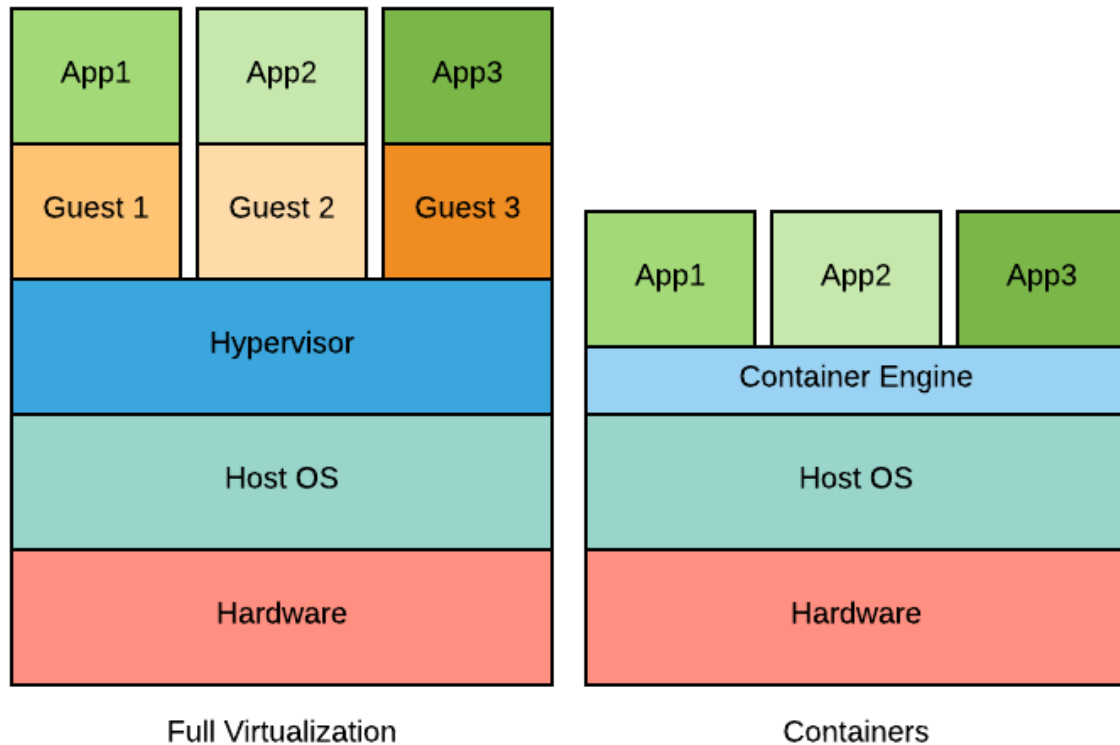
- A user-friendly modular exascale service platform to combine data and computational services on top of European research infrastructures – HPC and Clouds alike

- **Goals**

- Mature, modular, generalizable Open Source solutions for user friendly exascale data
 - Exascale learning on medical image data
 - Low Frequency Array / Square Kilometer Array
 - Supporting innovation based on global disaster risk data
 - Ancillary pricing/airline revenue management
 - Agro-Copernicus (correlating data between simulation and observation)



- “Focus on services and forget about infrastructures” idea
 - Analysis, data mining, pattern recognition, etc.
 - Heterogeneous research datasets
 - Support HPC and cloud-based computations
 - Utilize Docker, Singularity or Charliecloud
-
- Follows simplified workflow:
 - User defines pipeline/steps in IEE via UI
 - IEE reads a computation for execution or starts it automatically (as requested)
 - Pipeline steps are executed – containers are scheduled as HPC jobs via Rimrock
 - HPC queues and executes jobs (SLURM)
 - Jobs run embedded containers with appropriate options
-
- Data transfers are also handled automatically by the LOBCDER



Future work

- streamlining data transfers across technologies and sites
- concurrent execution of applications on multiple sites
- supporting a multi-node MPI on Singularity (with Horovod)
- scaling up pipelines to run on SuperMUC-NG (TOP500 #9)
- improving interface between IEE and the LOFAR portal

Current status of the PROCESS platform

- integrated solution
- running use cases on heterogeneous e-infrastructures
- support for a range of core technologies (HPC, Cloud)
- applications containerized for portability
- multiple geographical locations
- Containers: small footprint, low overhead, quick launch, manageable images
- Singularity: built for HPC, support for SLURM, unprivileged, support for MPI and GPUs

References

1. M. Bobak, A. S. Z. Belloum, P. Nowakowski, J. Meizner, M. Bubak, M. Heikkurinen, O. Habala, and L. Hluchý, "Exascale computing and data architectures for brownfield applications," in Fuzzy Systems and Knowledge Discovery (FSKD), 2018 14th International Conference on. IEEE, 2018, pp. 461 – 468.
2. M. Bobak, A. S. Z. Belloum, R. Cushing, J. Meizner, P. Nowakowski, V. Tran, O. Habala, J. Maassen, B. Somosköi, M. Graziani, M. Heikkurinen, M. Höb, J. Schmidt, L. Hluchý, "Reference exascale architecture", 15th International Conference on eScience, San Diego, USA
3. Docker, <https://www.docker.com/>
4. Singularity, <https://sylabs.io/singularity/>
5. Charliecloud, <https://hpc.github.io/charliecloud/>
6. The PL-Grid National Computing Infrastructure, www.plgrid.pl

PROCESS

PROCESS A hybrid platform for future exascale scientific applications

Horizon 2020

Marcin Bielecki¹, Piotr Bielecki^{2,3}, Jan Bielecki⁴, Martin Bielecki⁵, Dariusz Hadas⁶, Jarosław Hucy⁷, Wit Tys⁸, Adam S. Z. Bielecki⁹, Reginald Cushing¹⁰, Marcin Kubiś¹¹, Dieter Kraussbacher¹², and Jan Sabharwal¹³

¹ICT 4 Pharma, AGH University of Science and Technology, Kraków, Poland; ²Scala Centre for Computational Personalised Medicine, Institute of Informatics, Massey University, Auckland, New Zealand; ³Informatics Institute, University of Amsterdam, Amsterdam, Netherlands; ⁴MIMM Team, Leibniz-Materialienforschung, Aachen, Germany; ⁵Leibniz Supercomputing Centre, Garching, Germany

Goals

- Provide exascale ready computational and data services that will accelerate innovation
- Address the services in real-world settings of scientific research and in industry pilots:
 - Computational using data from the ESA satellite, technology preparation for the ESA challenges:
 - Medical use-case utilizing AI on CPUs and GPUs for cancer diagnosis based on imaging data
 - Supporting innovation based on global disease risk data
 - Business use-case for building models enabling resource optimization and real-time recalculations of prices
 - Agricultural simulations utilizing data from the part of Copernicus dataset collected from the Sentinel satellites

Extreme Large Scale Computing Services

Interface layer
Web UI and applications requiring programmatic access

Services layer
Secure web/REST interface

Resources layer
Heterogeneous HPC and storage resources

EMROC (exascale ready MPI on GPU)
LORENE (exascale ready MPI on CPU)

Why containers? Why Singularity?

- Small footprint
- Less overhead
- Quick launch
- Manageable images
- Built for HPC
- SLURM support
- Unprivileged
- MPI and GPUs

Challenges and future work

- The current status of the PROCESS platform:
 - Integrated solution
 - running use cases on heterogeneous infrastructures
 - support for a range of core technologies (MPI, Cloud)
 - applications containerized for portability
 - multiple geographical locations
- In the future we will work on:
 - Further stabilization of the solution
 - making it as user friendly as possible
 - streamlining data transfer across technologies and sites
 - more secure execution of applications on multiple sites
 - supporting a multi-node MPI on Singularity (with hardware)
 - adding up options to run on SuperMUC-AI (OS/POWER BI)
 - improving interface between BE and the SDRS portal

Acknowledgements

This work is supported by the "PROviding Computing solutions for ExaScale ChallengeS" (PROCESS) project that received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 777533.

References

[1] M. Bielecki, P. Bielecki, J. Bielecki, M. Bielecki, D. Hadas, J. Hucy, W. Tys, A. S. Z. Bielecki, R. Cushing, M. Kubiś, D. Kraussbacher, and J. Sabharwal, "A hybrid platform for future exascale scientific applications," in *2020 IEEE/ACM International Conference on Big Data (Big Data)*, pp. 1045–1054, 2020.

[2] M. Bielecki, P. Bielecki, J. Bielecki, M. Bielecki, D. Hadas, J. Hucy, W. Tys, A. S. Z. Bielecki, R. Cushing, M. Kubiś, D. Kraussbacher, and J. Sabharwal, "A hybrid platform for future exascale scientific applications," in *2020 IEEE/ACM International Conference on Big Data (Big Data)*, pp. 1045–1054, 2020.

[3] M. Bielecki, P. Bielecki, J. Bielecki, M. Bielecki, D. Hadas, J. Hucy, W. Tys, A. S. Z. Bielecki, R. Cushing, M. Kubiś, D. Kraussbacher, and J. Sabharwal, "A hybrid platform for future exascale scientific applications," in *2020 IEEE/ACM International Conference on Big Data (Big Data)*, pp. 1045–1054, 2020.

[4] M. Bielecki, P. Bielecki, J. Bielecki, M. Bielecki, D. Hadas, J. Hucy, W. Tys, A. S. Z. Bielecki, R. Cushing, M. Kubiś, D. Kraussbacher, and J. Sabharwal, "A hybrid platform for future exascale scientific applications," in *2020 IEEE/ACM International Conference on Big Data (Big Data)*, pp. 1045–1054, 2020.

[5] M. Bielecki, P. Bielecki, J. Bielecki, M. Bielecki, D. Hadas, J. Hucy, W. Tys, A. S. Z. Bielecki, R. Cushing, M. Kubiś, D. Kraussbacher, and J. Sabharwal, "A hybrid platform for future exascale scientific applications," in *2020 IEEE/ACM International Conference on Big Data (Big Data)*, pp. 1045–1054, 2020.

[6] M. Bielecki, P. Bielecki, J. Bielecki, M. Bielecki, D. Hadas, J. Hucy, W. Tys, A. S. Z. Bielecki, R. Cushing, M. Kubiś, D. Kraussbacher, and J. Sabharwal, "A hybrid platform for future exascale scientific applications," in *2020 IEEE/ACM International Conference on Big Data (Big Data)*, pp. 1045–1054, 2020.

[7] M. Bielecki, P. Bielecki, J. Bielecki, M. Bielecki, D. Hadas, J. Hucy, W. Tys, A. S. Z. Bielecki, R. Cushing, M. Kubiś, D. Kraussbacher, and J. Sabharwal, "A hybrid platform for future exascale scientific applications," in *2020 IEEE/ACM International Conference on Big Data (Big Data)*, pp. 1045–1054, 2020.

[8] M. Bielecki, P. Bielecki, J. Bielecki, M. Bielecki, D. Hadas, J. Hucy, W. Tys, A. S. Z. Bielecki, R. Cushing, M. Kubiś, D. Kraussbacher, and J. Sabharwal, "A hybrid platform for future exascale scientific applications," in *2020 IEEE/ACM International Conference on Big Data (Big Data)*, pp. 1045–1054, 2020.

[9] M. Bielecki, P. Bielecki, J. Bielecki, M. Bielecki, D. Hadas, J. Hucy, W. Tys, A. S. Z. Bielecki, R. Cushing, M. Kubiś, D. Kraussbacher, and J. Sabharwal, "A hybrid platform for future exascale scientific applications," in *2020 IEEE/ACM International Conference on Big Data (Big Data)*, pp. 1045–1054, 2020.

[10] M. Bielecki, P. Bielecki, J. Bielecki, M. Bielecki, D. Hadas, J. Hucy, W. Tys, A. S. Z. Bielecki, R. Cushing, M. Kubiś, D. Kraussbacher, and J. Sabharwal, "A hybrid platform for future exascale scientific applications," in *2020 IEEE/ACM International Conference on Big Data (Big Data)*, pp. 1045–1054, 2020.

[11] M. Bielecki, P. Bielecki, J. Bielecki, M. Bielecki, D. Hadas, J. Hucy, W. Tys, A. S. Z. Bielecki, R. Cushing, M. Kubiś, D. Kraussbacher, and J. Sabharwal, "A hybrid platform for future exascale scientific applications," in *2020 IEEE/ACM International Conference on Big Data (Big Data)*, pp. 1045–1054, 2020.

[12] M. Bielecki, P. Bielecki, J. Bielecki, M. Bielecki, D. Hadas, J. Hucy, W. Tys, A. S. Z. Bielecki, R. Cushing, M. Kubiś, D. Kraussbacher, and J. Sabharwal, "A hybrid platform for future exascale scientific applications," in *2020 IEEE/ACM International Conference on Big Data (Big Data)*, pp. 1045–1054, 2020.

[13] M. Bielecki, P. Bielecki, J. Bielecki, M. Bielecki, D. Hadas, J. Hucy, W. Tys, A. S. Z. Bielecki, R. Cushing, M. Kubiś, D. Kraussbacher, and J. Sabharwal, "A hybrid platform for future exascale scientific applications," in *2020 IEEE/ACM International Conference on Big Data (Big Data)*, pp. 1045–1054, 2020.

<http://www.process-project.eu>

<http://dice.cyfronet.pl>



This work is supported by the "PROviding Computing solutions for ExaScale ChallengeS" (PROCESS) project that received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 777533.

This research was supported in part by PL-Grid Infrastructure. The authors also want to thank Mara Graziani, Balázs Somosköi, Jason Maassen and Henning Müller for their valuable insight on the presented subject.