

OpenCUBE: Building an Open Source Cloud Blueprint with EPI Systems*

Ivy Peng^{**1}, Martin Schulz², Utz-Uwe Haus³, Craig Prunty⁴, Pedro Marcuello⁵, Emanuele Danovaro⁶, Gabin Schieffer¹, Jacob Wahlgren¹, Daniel Medeiros¹, Philipp Friese², and Stefano Markidis¹

¹ KTH Royal Institute of Technology, Sweden

² Technical University of Munich, Germany

³ Hewlett Packard Enterprise, Switzerland

⁴ SiPearl, France

⁵ Semidynamics, Spain

⁶ ECMWF, Italy

Abstract. OpenCUBE aims to develop an open-source full software stack for Cloud computing blueprint deployed on EPI hardware, adaptable to emerging workloads across the computing continuum. OpenCUBE prioritizes energy awareness and utilizes open APIs, Open Source components, advanced SiPearl Rhea processors, and RISC-V accelerator. The project leverages representative workloads, such as cloud-native workloads and workflows of weather forecast data management, molecular docking, and space weather, for evaluation and validation.

Keywords: Open-source · Converged HPC and Cloud · Computing continuum · EPI · RISC-V

1 Introduction

OpenCUBE is a project funded by the European Commission and initiated in January 2023. Its primary aim is to create, implement, and validate a full software stack for enabling a European Cloud computing blueprint deployed on European hardware infrastructure and cater to industrial and consumer cloud workloads. Additionally, the project aims to prioritize power and energy efficiency by incorporating power awareness at all levels. OpenCUBE will support the diverse requirements of the entire computing continuum, spanning from edge to cloud and high-performance computing (HPC).

OpenCUBE involves designing and installing a prototype hardware infrastructure composed of SiPearl processors and Semidynamics RISC-V accelerators, which are outcomes of the European Processor Initiative (EPI). Heterogeneous compute nodes will be interconnected with a high-performance Ethernet network to support the exploration of emerging memory and storage disaggregation. OpenCUBE will create a unified software stack encompassing various best

* OpenCUBE is funded by European Commission Horizon Project 101092984.

** Corresponding author ipeng@acm.org.

practices and open-source tools at the operating system, middleware, and system management levels. The OpenCUBE software stack for cloud services will be open-source and leverage industry-standard Open APIs and Open Source components.

In line with the European Green Deal initiative, OpenCUBE is designed to enable energy awareness as a fundamental feature across the entire stack. Through software-hardware co-design, the OpenCUBE software stack will provide API access to various site levels, from core, socket, node, and even to the electricity grid. The project will utilize representative workloads, such as weather forecast data management, molecular docking, and space weather workflows, to inform the design and deployment of the OpenCUBE system.

2 Approach

OpenCUBE takes an approach focusing on close hardware and software interfaces and is organized into four thrusts – hardware platform, middleware, heterogeneous data center (DC), and applications (Figure 1).

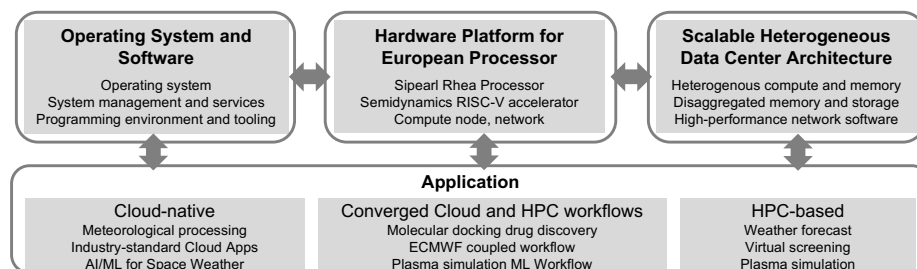


Fig. 1. An overview of the OpenCUBE approach with a focus on close interface between software and hardware.

- **The Hardware Platform** is designed for hosting EPI systems, e.g., SiPearl Rhea and Semidynamics RISC-V processors. The design is adapted for catering to both cloud-native and HPC workloads’ requirements. Also, the hardware platform provides inputs to the design of the operating system and middleware for a heterogeneous data center. As in the early phase, a prototype platform is built to exploit ARM-based processors available in commercial cloud systems, high-performance Ethernet-based interconnects, and the EPAC RISC-V accelerator.
- **Operating Systems and Software** will leverage and extend open-source components to support newly released hardware features for monitoring and resource management on the hardware platform. Extensions to system management services and OS will be designed to utilize the high-performance Ethernet network and enable power awareness at various system levels to achieve energy efficiency and performance. Open-source profiling tools will be adapted to the new CPU architecture to improve performance tuning and debugging.

- **A Scalable Heterogeneous Data Center** architecture will be deployed through open-source middleware specifically adapted to the hardware platform and targeted workloads. In particular, the OpenCUBE middleware will focus on improving the utilization of heterogeneous compute and memory resources on node. Also, extensions are to be developed to enable memory disaggregation over the fast network for either memory expansion or data staging in workflows. High-performance network software is designed to enable the efficient execution of applications in cloud-based containerized environments. For instance, MPI is the dominant communication API for traditional HPC applications. However, its relatively static model, which was developed for conventional HPC setups and schedulers, needs to be revised for and adapted to cloud-based application deployment, for example, using the recently introduced MPI Sessions concept [2].
- **Driver Applications & Workflows** will interact with hardware, software, and DC middleware thrusts to input workload requirements. The OpenCUBE stack will be validated and evaluated through these applications, including cloud-native workloads and workflows targeting the computing continuum from cloud, HPC, and edge. The applications will continuously provide feedback throughout the development of the stack. For instance, ECMWF [1]’s IFS will drive customization of general data storage middleware. A workflow with integrated ML-based analytics and iPIC3D [3]-based simulation will guide the software stack for converged HPC and cloud. A virtual screening workflow based on AutoDock [7], a widely used molecular docking software for drug discovery, will be used for validating cloud-based workflows on heterogeneous resources.

3 Preliminary Results and Roadmap

In the first phase, we are deploying a prototype hardware platform of one rack of four HPE ProLiant RL300 servers with 256 GB DDR4 memory and a minimum of 1 TB NVMe SSD for provisioning fabric-attached memory. An FPGA emulator is employed to integrate EPAC RISC-V processors for acceleration. The platform is equipped with Slingshot interconnects, including one Cassini Switch GB Ethernet and the Cassini network interface card.

To enable cloud-native workflows, we investigate a popular workflow management software, Apache Airflow. As a case study, we used a virtual screening software in drug discovery, AutoDock, to enable a workflow of automatic elastic molecular docking on the cloud [4]. Our preliminary results confirm the feasibility of deployment into the containerized environment. We investigated the state-of-the-art disaggregated memory technologies such as Compute Express Link (CXL) for the scalable heterogeneous data center architecture. We developed a memory-centric profiling tool and a software emulation framework to explore design space [9] quantitatively.

To enable converged computing between Cloud and HPC, we developed a framework atop the open-source container orchestrator Kubernetes that enables the reuse of already provisioned infrastructure. This capability enables automatic horizontal scaling for tightly coupled MPI-based applications, which is

cumbersome to realize on traditional HPC systems [5]. In analyzing application requirements for input to OpenCUBE stack design, we also identified a scalability bottleneck in reduction operation due to a large number of synchronization points. We proposed a matrix-based multi-dimensional reduction algorithm for accelerating the local search of the scoring function and explored a tensor-based implementation for optimizing the molecular docking process [8]. The results show an over 25% improvement in average docking time for a real-world docking scenario.

On the roadmap towards an open-source cloud blueprint, the OpenCUBE project employs an approach focusing on close interaction between software and hardware development to create an open-source cloud blueprint on EPI systems. The prototype implementation in OpenCUBE will be validated and evaluated with industrial and consumer cloud applications. Development in enabling adaptive MPI communication setup, e.g., session, will also provide feedback to standardization bodies [6]. OpenCUBE’s roadmap aligns the major upgrade of the prototype hardware infrastructure with chips produced from EPI. Meanwhile, insights and findings learned during the design and development of the OpenCUBE stack based on the Sipearl Rhea and Semidynamics RISC-V processors are feedback to the Open Source, EPI, and Computing Continuum communities.

Acknowledgment: OpenCUBE is funded by EU Horizon Project 101092984.

References

1. Bauer, P., Thorpe, A., Brunet, G.: The quiet revolution of numerical weather prediction. *Nature* **525**(7567), 47–55 (2015)
2. Holmes, D., Mohror, K., Grant, R.E., Skjellum, A., Schulz, M., Bland, W., Squyres, J.M.: MPI sessions: Leveraging runtime infrastructure to increase scalability of applications at exascale. In: *Proceedings of the 23rd European MPI Users’ Group Meeting*. pp. 121–129 (2016)
3. Markidis, S., Lapenta, G., et al.: Multi-scale simulations of plasma with iPIC3D. *Mathematics and Computers in Simulation* **80**(7), 1509–1519 (2010)
4. Medeiros, D., Schieffer, G., Wahlgren, J., Peng, I.: A GPU-accelerated molecular docking workflow with kubernetes and apache airflow. In: *International Conference on High Performance Computing*. pp. 193–206. Springer (2023)
5. Medeiros, D., Wahlgren, J., Schieffer, G., Peng, I.: Kub: Enabling elastic hpc workloads on containerized environments. In: *35th International Symposium on Computer Architecture and High Performance Computing (SBAC-PAD)*. IEEE (2023)
6. the MPI Forum: The MPI Standard, v4.0 (June 2021)
7. Santos-Martins, D., Solis-Vasquez, L., Tillack, A.F., Sanner, M.F., Koch, A., Forli, S.: Accelerating AutoDock4 with gpus and gradient-based local search. *Journal of chemical theory and computation* **17**(2), 1060–1073 (2021)
8. Schieffer, G., Peng, I.: Accelerating drug discovery in AutoDock-GPU with tensor cores. In: *Euro-Par 2023: Parallel Processing: 29th International Conference on Parallel and Distributed Computing, Proceedings*. Springer (2023)
9. Wahlgren, J., Schieffer, G., Gokhale, M., Peng, I.: A quantitative approach for adopting disaggregated memory in HPC systems. In: *International Conference for High Performance Computing, Networking, Storage and Analysis (2023)*, to appear