

Project Partners



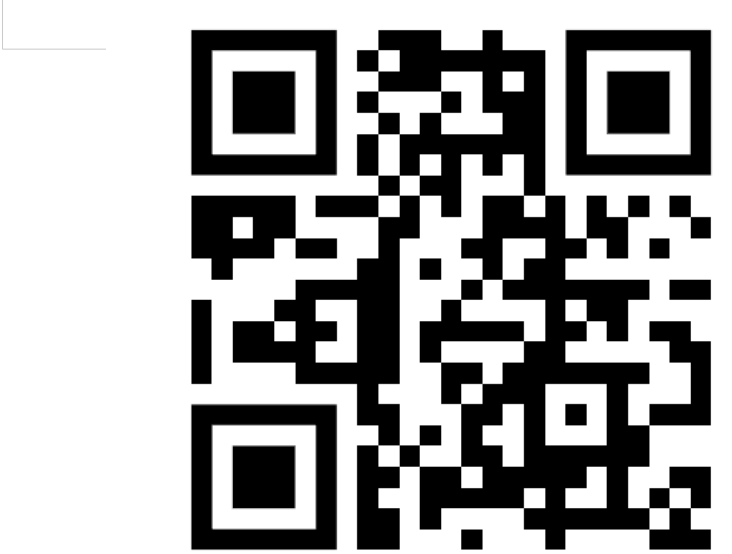
Friedrich-Alexander-Universität Erlangen-Nürnberg



High-Performance Computing Center Stuttgart



Software



Publications

J. Eitzinger, T. Gruber, A. Afzal, T. Zeiser, and G. Wellein: **ClusterCockpit – A web application for job-specific performance monitoring**. Published at HPCMASPA 2019, Albuquerque, NM, USA. Held in conjunction with IEEE Cluster 2019. DOI: 10.1109/CLUSTER.2019.8891017

T. Röhl, J. Eitzinger, G. Hager, and G. Wellein: **LIKWID Monitoring Stack: A flexible framework enabling job-specific performance monitoring for the masses**. Published at HPCMASPA 2017. Held in conjunction with IEEE Cluster 2017, Honolulu, HI, September 5, 2017. DOI: 10.1109/CLUSTER.2017.115.

J. Treibig, G. Hager, and G. Wellein: **LIKWID performance tools**. In: C. Bischof et al. (eds.), *Competence in High Performance Computing 2010*. Springer, ISBN 978-3-642-24025-6 (2012), 165-175. DOI: 10.1007/978-3-642-24025-6_14

Funding

SPONSORED BY THE



EE-HPC Website



EE-HPC – A framework for energy efficient HPC system operation

J. Eitzinger, T. Gruber, Rafael Ravedutti
Friedrich-Alexander Universität Erlangen-Nürnberg
Erlangen National High-Performance Computing Center
(NHR@FAU)

C. Terboven, R. Liem
RWTH Aachen University
IT Center

J. Gracia, K. Haldar
High Performance Computing
Center Stuttgart (HLRS)

J.F. Engels, P. Giesselmann
German Climate Computing Center
DKRZ

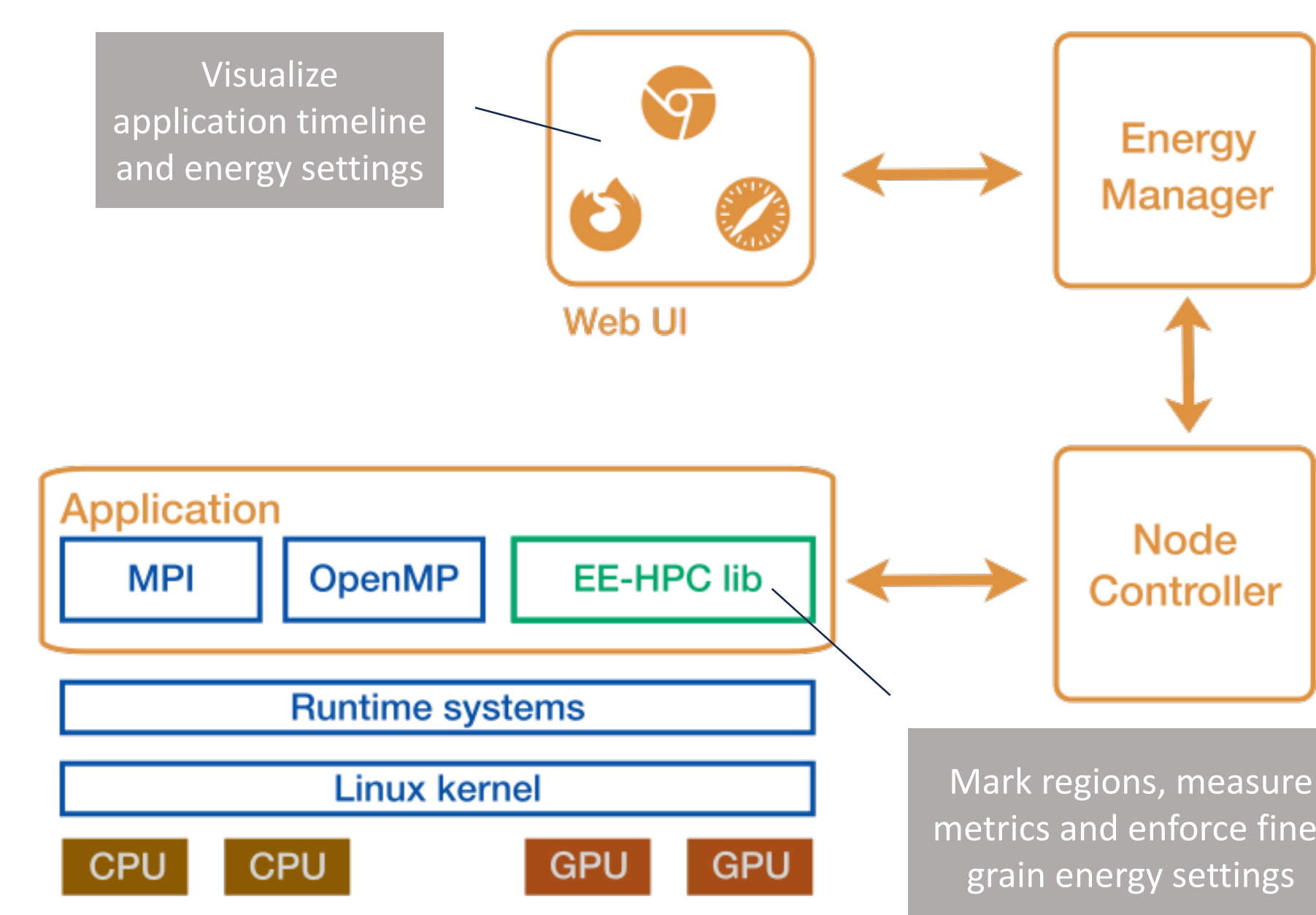
**T. Wilde, C. Simmendinger,
M. Marquardt, J. Maeder, D. Brayford**
HPE Germany

Overview and Partners

The primary goal of the EE-HPC project is to achieve more efficient energy use in HPC systems. This is crucial because energy consumption is a significant factor in the procurement and operation costs of HPC data centers. EE-HPC focuses on developing a solution for monitoring and controlling energy consumption in HPC systems. The project places importance on making the monitoring and energy control solution easy to use, reliable, safe, and capable of handling large-scale systems. The project's framework will be tested and evaluated using a set of relevant HPC applications from various fields, including molecular dynamics, engineering, and climate research. The consortium consists of the two TIER2 HPC centers NHR@FAU and IT Center (RWTH Aachen University), the TIER0/1 HPC center HLRS, the Climate Research computing center DKRZ, and HPE Germany as industry partner.

Instrumentation Library for Fine-Grain Control and automatic Phase Detection

In addition to the default energy settings, there is an optional mechanism for achieving fine-grained optimization of applications with repetitive phases. To accomplish this, an instrumentation library is used. This library interacts with applications to gather information about their characteristics, particularly focusing on identifying repetitive phases within the applications' execution. The instrumentation library leverages interfaces such as the PMPI and OMPT profiling interfaces to collect this data.



The library provides this information to the node controller, which can then adjust system settings or resource allocation to reduce energy consumption during these phases. This approach can be particularly beneficial for applications with complex and dynamic resource usage patterns.

Data Structure Format and Job Archive

We propose a set of generic data structure specifications for job meta-data, job metric data, and a cluster description. The data structures are specified based on a JSON Schema description. Other data formats as file formats, DB schemas, and API payloads are derived from these data-structure descriptions. The generic data-structure specification is used in the form of a JSON file-based job archive as a versatile approach for storing and sharing job data.



Performance and Energy Monitoring: ClusterCockpit

ClusterCockpit is developed at NHR@FAU and offers the following components:

- cc-backend:** A web-backend providing REST and GraphQL APIs
- cc-frontend:** A web UI implemented in Svelte
- cc-metric-collector:** A node agent for collecting metric data
- cc-metric-store:** A simple time-series metric data in-memory cache

All components can also be used standalone. ClusterCockpit is stable and used in production at several large HPC centers.

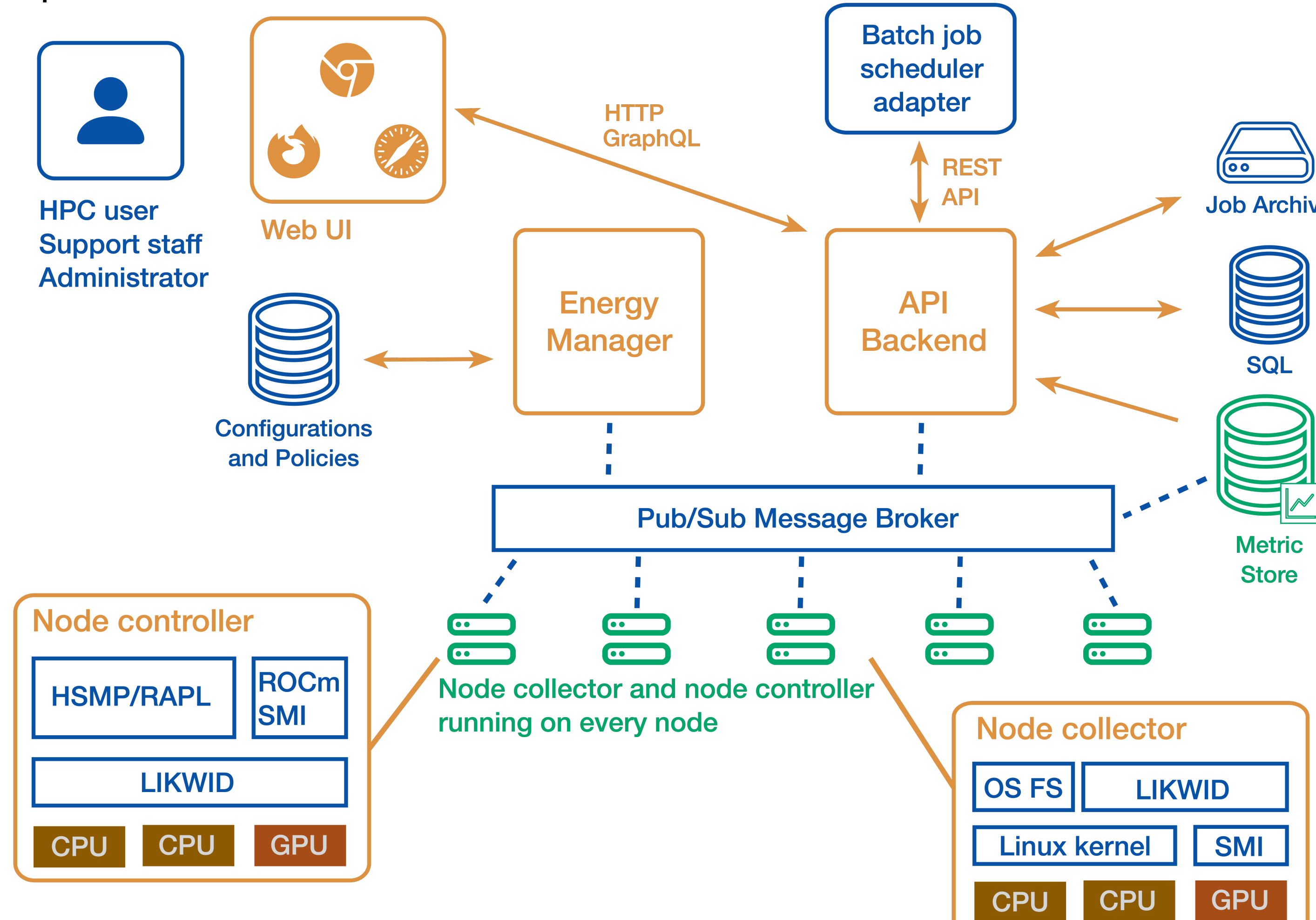


<https://github.com/ClusterCockpit>

Architecture of EE-HPC

Node metrics are collected continuously by node collectors and sent to a metrics store at fixed intervals. Job details are provided by an external adapter for the batch job scheduler and sent to the API backend via a REST API. For running jobs, the API backend queries the metrics store to collect the required time series data. Once a job is finished, it is persisted to a JSON file-based job archive that contains all job metadata and metrics data. Finished jobs are then loaded from the job archive. The metrics store uses cyclic buffers and stores data only for a limited duration.

For a description of the Energy Manager and Node controller refer to the separate description box.



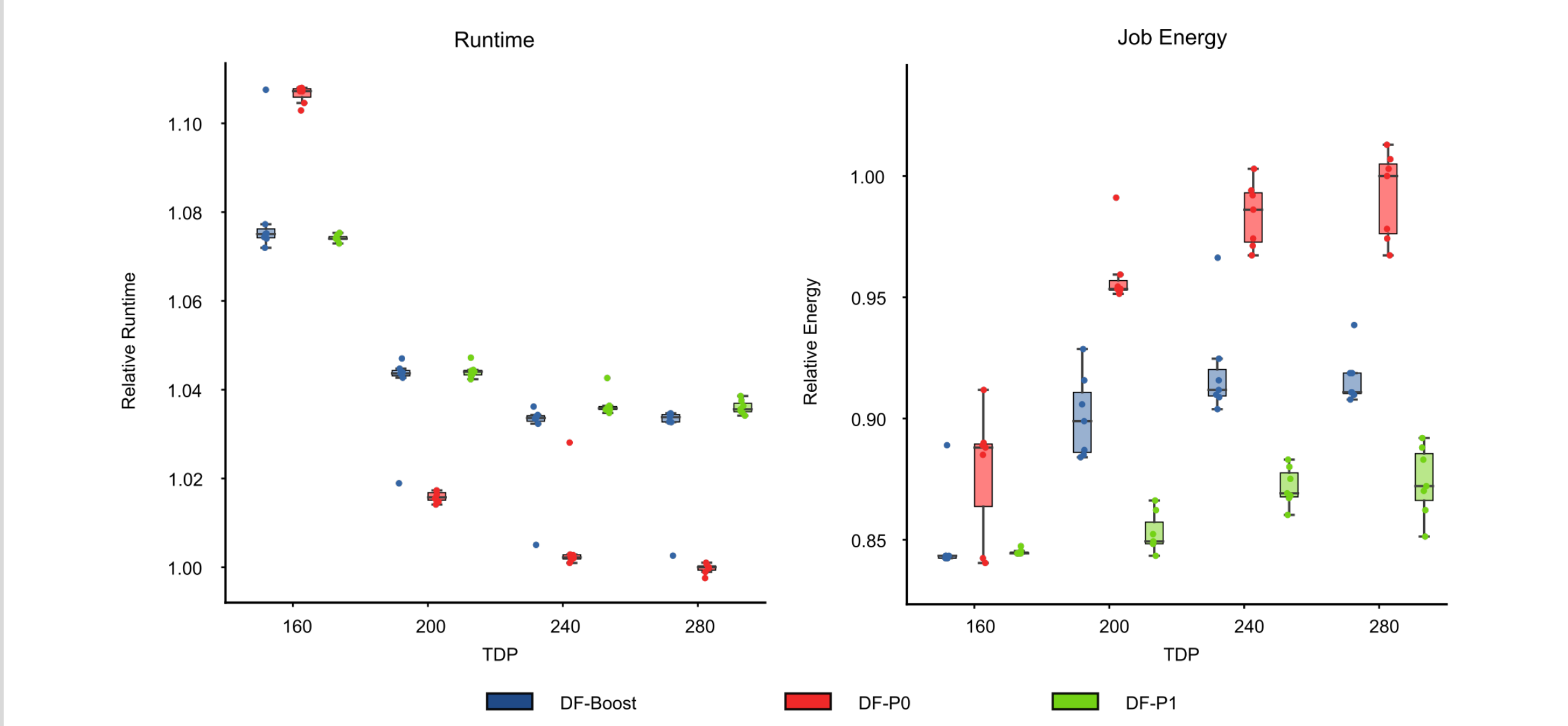
Validation - ICON climate Code

The ICON weather and climate model consists of atmosphere, ocean, land and chemistry components.



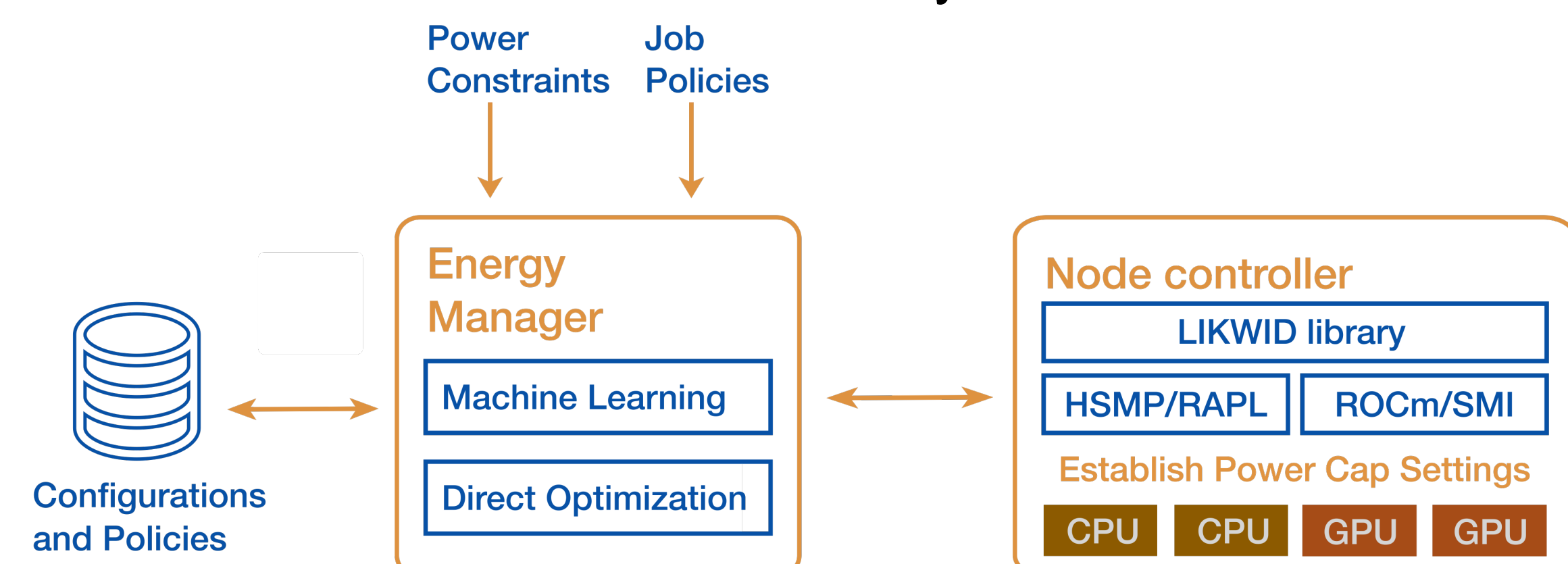
Over the course of more than two decades, it has evolved with dual objectives: serving as a scientific research tool and a real-time operational forecasting model.

Originally coded primarily in Fortran and leveraging MPI/OpenMP for parallelization, the model has grown to encompass over two million lines of code. More recently, it has been adapted to harness the computational power of GPUs through OpenACC. Presently, the code is undergoing a substantial redesign aimed at creating a more modular software structure. The goal is to release it as an open-source project by 2024. Due to its frequent use by many users, optimizing its runtime energy efficiency could play a substantial role in mitigating its carbon footprint.



Description of Energy Manager and Node controller

Job specific power distribution and energy consumption is guided by a central component, the Energy Manager (EM). The EM is aware of external power constraints and job policies. The node controllers gather required metrics and enforce the decisions from the EM. During runtime, jobs undergo dynamic adjustments, guided by specific job policies and the real-time characteristics of the application workload. These application workloads are continually assessed and characterized in a dynamic manner.



To implement these adjustments, the framework establishes power caps for the hardware energy management unit on both CPUs and GPUs, effectively enforcing the desired power and performance settings.

| App | Use case | Runtime difference | Energy difference |
|----------|---------------------------------|--------------------|-------------------|
| LBC | Lid-driven Cavity, 32 Nodes | 3,80% | -17,54% |
| NS3DNeo | Multi-Phase Flow, ZPG, 8 Nodes | 3,30% | -12,90% |
| OpenFoam | FixedTol, Single-Phase, 8 Nodes | -3,50% | -12,90% |
| StarCCM+ | Open Water Propeller, 16 Nodes | 1,00% | -11,11% |
| XHPL | HPL-2.3, 8 Nodes | 2,00% | -1,45% |

Conclusion and Future work

The EE-HPC project introduces an integrated monitoring and power management framework for hybrid HPC clusters. The framework leverages Hardware Monitoring (HPM) metrics provided by a job-specific monitoring infrastructure and a central energy manager responsible for making node configuration decisions. These decisions can be based on either machine learning techniques or direct optimization targeting specific performance metrics. To implement this framework, two distinct daemons operate on each node. The first is a node collector responsible for measuring and transmitting performance metrics, while the second is a node controller that assesses the current energy policies, relays this information to the energy manager, and ensures that the recommended settings are enforced. Optionally, an energy efficiency library is available, offering adaptive fine-grained phase detection and the enforcement of energy policies. To assess the framework's effectiveness, a comprehensive evaluation is carried out using a diverse set of large HPC application codes from various domains. These evaluations encompass multiple clusters from various vendors, illustrating the project's versatility and applicability across different computing environments.

Multimedia Content

ClusterCockpit Demo



LIKWID Tools Intro Video

