The Art of Process Pinning: Turning Chaos into Core Harmony

T. Breuer, W. Frings, J. H. Göbbert, F. S. M. Guimarães, C. Himmels, C. Paschoulas

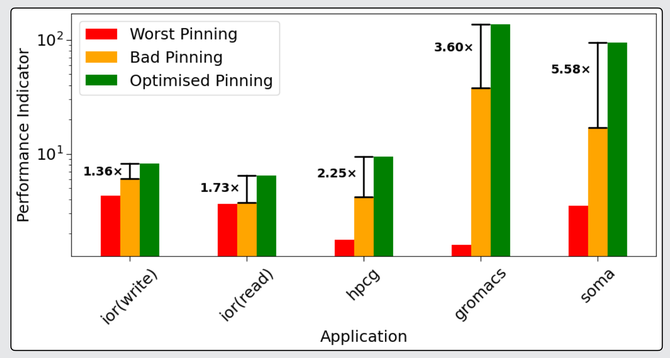
Jülich Supercomputing Centre, Forschungszentrum Jülich

Motivation

Optimising node usage in High-Performance Computing (HPC) systems is challenging, as each node has numerous cores with non-uniform access speeds to memory and connected hardware, such as GPUs. Efficient pinning of processes to cores is not straightforward; no single configuration fits all cases. It requires coordination across:

- the **system side**, to ensure the expected behaviour from supercomputers is achieved,
- the **user side**, to optimise application resource usage,
- the **operator side**, to monitor and resolve issues proactively.

To manage these challenges, we have developed and tested a comprehensive pinning infrastructure for HPC centres, and brought it into everyday production. Proper process pinning increases the throughput of an HPC system, achieving more results in the same amount of time.



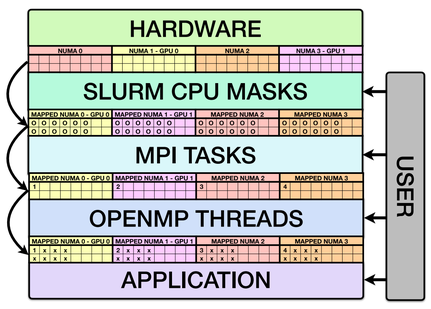
Measurement of the influence of pinning on the performance of selected HPC applications on a single JUWELS Booster node.

Challenges

Achieving optimal system usage with respect to process pinning is challenging at a variety of levels.

System side

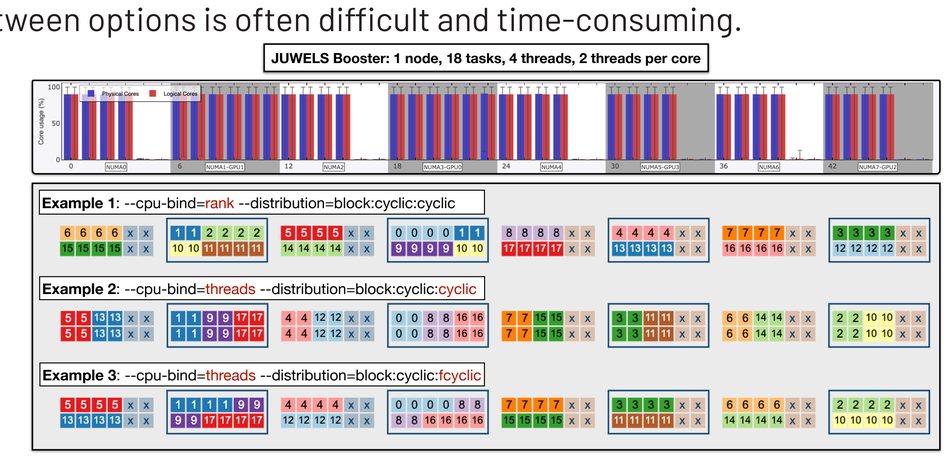
Ensuring HPC systems perform as expected requires managing their complexity, including diverse architectures and configurations. A deep understanding of the interplay between hardware and software is essential, especially as default settings can shift with software updates.



Example of a hybrid MPI-OMP app on a Slurm-managed cluster.

User side

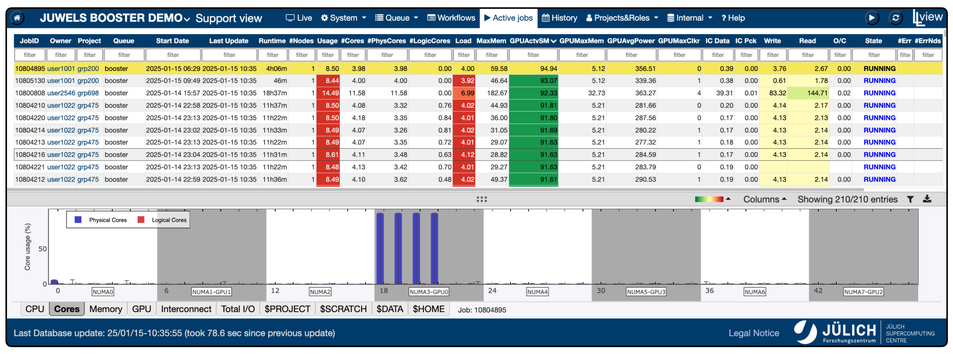
Users need to select the best pinning options depending on application characteristics, input data and system architecture, but the many available choices make this challenging. Testing and understanding the differences between options is often difficult and time-consuming.



Example of the effect varying pinning parameters has on task and thread distribution with the same core usage pattern (shown in the top row). This can adversely affect performance.

Operator side

Once jobs conclude, we should verify that the system resources were efficiently utilised, as sub-optimal pinning choices increase job runtime and reduce system throughput. Proactive monitoring and support are essential to quickly identify and address any issues.



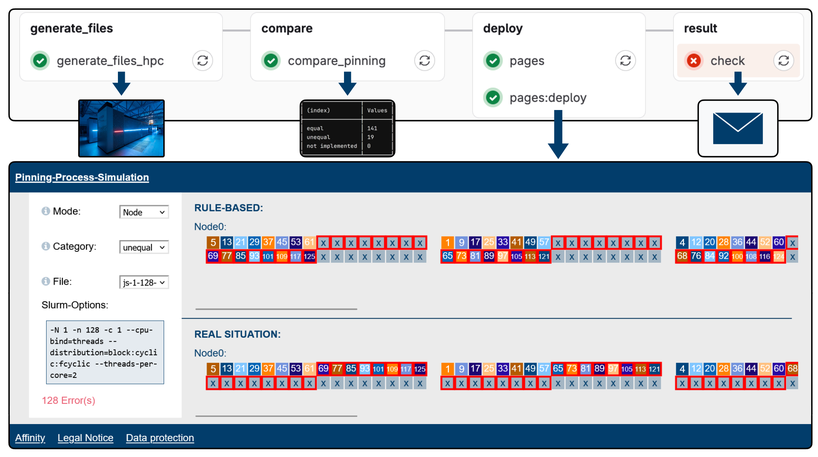
Monitoring core usage patterns with LLview reveals inefficient jobs.

Solutions

Resulting from in-depth analysis of the needs faced by users, admins, and support teams, we propose workflows and tools to target these pinning challenges.

System side

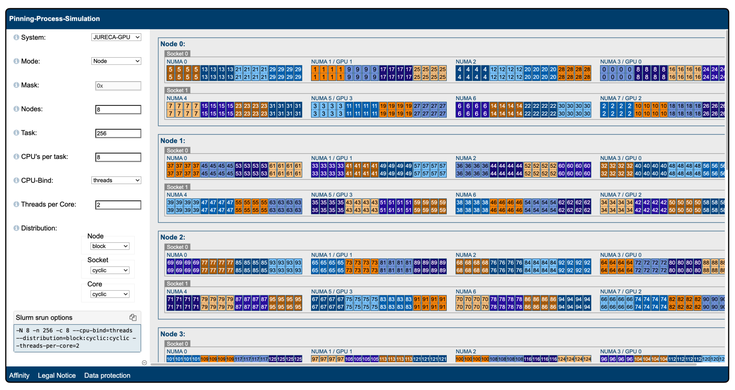
By running jobs periodically on HPC test systems with various pinning options using GitLab CI, we capture actual job pinning. The **JuPin** tool then compares them with the expected results, highlighting discrepancies so issues can be identified and resolved.



Automatic tests of pinning options and visual representation of deviations from expected results.

User side

JuPin provides an interactive web interface to explore and evaluate pinning options on HPC systems, offering visual representations of task and thread distributions.

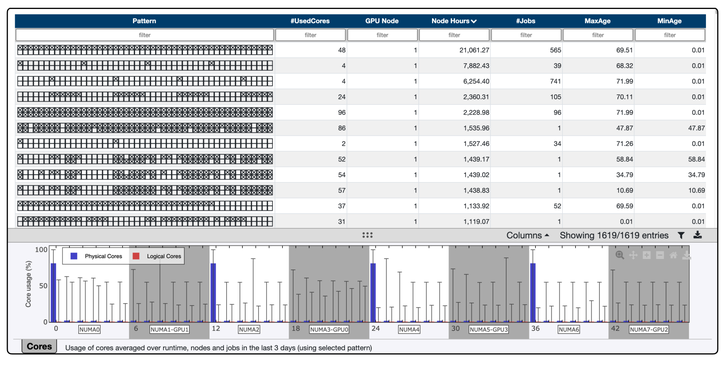


Visualising and comparing pinning options with JuPin.

While it does not perform the pinning itself, it assists users in understanding and optimising configurations and generates Slurm-compatible job submission options.

Operator side

LLview monitors CPU usage patterns for each job and generates statistics across all jobs, helping identify issues and enabling proactive user support.



Core usage patterns and detailed statistics provided by LLview.

Conclusion and Outlook

Effective process pinning is a long-standing, challenging, and critical issue in optimising HPC overall system usage. We have developed, tested and deployed **JuPin** and **LLview** as infrastructure that simplifies this process, providing users and HPC system operators with the insights needed to fine-tune configurations and identify issues proactively. These advancements are also vital in preparing for future challenges, such as the JUPITER exascale supercomputer. By building robust tools and workflows now, we pave the way for seamless operation in the exascale era.

References and Acknowledgements

JuPin go.fz-juelich.de/pinning

LLview go.fz-juelich.de/llview

Federal Ministry of Education and Research

Ministry of Culture and Science of the State of North Rhine-Westphalia

GCS Gauss Centre for Supercomputing

JUPITER